Model Parameters

0.1 Models Identified to Approximate Systems A and B

It is important to note that the discretisation step size of these models is 1 second.

0.1.1 Model Parameters of Models Approximating System A

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.05540618850	-0.01350287933	-0.67977994584	-0.66261165389
A_2	0.36145144555	0.01152582209	-0.480671584105	0.21914276246
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	0.00824535052	-0.00897154910	0.42843696718	-0.13875198407
B_2	0.02766332279	0.00550953274	1.03591283249	1.12482710546

Table 1: Model parameters for the linear ARX model obtained from experiment OL 1

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.98803735585	-0.01339715658	0.87189300156	-0.56884905581
A_2	0.32557832900	0.01246544591	-1.23352869196	0.20667995323
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	0.00137498349	-0.00407464899	-0.056466380618	-0.05001750421
B_2	0.00976029726	-0.00632513248	0.40323968391	0.40384123462

Table 2: Model parameters for the linear ARX model obtained from experiment OL 2

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.04725363270	-0.01991174643	3.63872504485	-1.12231384235
A_2	0.30488686275	0.01837536135	-5.41880898503	0.58727066482
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.04507952931	0.00175120057	-1.68128605887	-1.20257462913
B_2	0.11172667138	-0.00999264889	4.14518594857	3.17228632894

Table 3: Model parameters for the linear ARX model obtained from experiment OL 3

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.11618977808	-0.01678707218	1.74331094047	-1.05829562657
A_2	0.34867654023	0.01419167518	-2.69228563106	0.44167155386
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.02476003502	0.01114333986	-0.78795193771	-0.28051263685
B_2	0.11621952787	-0.02309544480	3.85644953305	2.72964955942

Table 4: Model parameters for the linear ARX model obtained from experiment OL 4

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	0.05523177779	-0.03543249650	-0.58945493244	-0.99350050417
A_2	0.00475310885	-0.00041287532	-1.92754636587	0.41326721907
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	0.01797593532	0.01797593532	1.42823790120	1.42823790120
B_2	0.01797593532	0.01797593532	1.42823790120	1.42823790120

Table 5: Model parameters for the linear ARX model obtained from experiment OL 5

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.35492725283	-0.03205333056	-0.76767349350	-1.01888022414
A_2	-0.21124380817	0.01902878664	0.46639025457	0.36494806209
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.03797338165	-0.01257285935	-0.21203542351	0.01356643406
B_2	0.17540547059	-0.0781157085	3.52057257027	2.50370104312

Table 6: Model parameters for the linear ARX model obtained from experiment CL 1

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.29158542253	-0.03308850631	0.78558247897	-1.05570683351
A_2	-0.27146854979	0.01918839699	-0.704484451347	0.36967441771
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	0.01540798511	-0.01129893304	0.04951126814	0.27016980585
B_2	0.11671001706	-0.08472506853	3.22952289179	2.05624642341

Table 7: Model parameters for the linear ARX model obtained from experiment CL 2

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.37481553195	-0.03406319568	0.07543412214	-0.97765987478
A_2	-0.18062615515	0.02019540197	-0.50759865194	0.34537199610
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.04408712203	0.04229286782	-0.83604824191	-0.80961473621
B_2	0.18641803658	-0.13455876988	4.30101531358	3.49670506002

Table 8: Model parameters for the linear ARX model obtained from experiment CL 3

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.36241828238	-0.03558465153	0.16092107278	-0.98040994093
A_2	-0.18255538514	0.02075544302	-0.57501455813	0.34817187889
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.05833634116	0.03913103449	-1.06599691533	-0.87318248954
B_2	0.19969780541	-0.13669599174	4.57456384942	3.58232873383

Table 9: Model parameters for the linear ARX model obtained from experiment CL 4

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.31737645559	-0.03778810386	0.15511207256	-0.99037923461
A_2	-0.19456718004	0.02031090071	-0.68134176748	0.35762777170
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.03329086741	0.05859050629	-0.80372217829	-1.01426549559
B_2	0.16837378551	-0.17454838483	4.20347753086	3.69426106453

Table 10: Model parameters for the linear ARX model obtained from experiment CL 5

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.35597078532	-0.03394470302	-0.06706438060	-0.96533891454
A_2	-0.18155993015	0.02070207998	-0.17423175403	0.34938782342
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.08064033406	0.04238127740	-1.27576423589	-0.95363698793
B_2	0.23127994614	-0.13640923209	4.79012484624	3.59425592527

Table 11: Model parameters for the linear ARX model obtained from experiment CL 6

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.21388974368	-0.04312707303	-0.20714040318	-1.00281235174
A_2	-0.26472603850	0.0245876209	-0.39490587534	0.35331703669
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.02477733288	0.06542217177	-0.05365451593	-0.59412680517
B_2	0.16409569229	-0.19334911713	3.26198783686	3.16929229716

Table 12: Model parameters for the linear ARX model obtained from experiment CL 7

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.18274773057	-0.0438467930	-0.14737253859	-0.98326358916
A_2	-0.26440975071	0.02509333961	-0.60550551843	0.34470357206
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.04530406143	0.07985957392	0.13037431809	-0.83325340885
B_2	0.20082128177	-0.21069981058	3.06745682700	3.46617053125

Table 13: Model parameters for the linear ARX model obtained from experiment CL 8

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.22720255465	-0.03132446634	-0.24092659818	-1.05096906795
A_2	-0.27416350698	0.01616883815	-0.21451716194	0.38970194695
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.03216084559	0.02783765963	-0.00710434710	-0.17452495416
B_2	0.18849169454	-0.13332033407	3.21192718092	2.68156100767

Table 14: Model parameters for the linear ARX model obtained from experiment CL 9

0.1.2 Model Parameters of Models Approximating System B

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.00779060540	-0.01982430460	-0.00669698192	-1.00577974178
A_2	-0.03090753047	0.06964146528	-0.01732037318	-0.07741226132
A_3	0.00687481794	0.01024726841	0.09060220336	-0.02142178838
A_4	0.20817842329	0.14650916294	-0.17720295660	-0.10861437316
A_5	-0.11325642850	-0.22954059695	0.09991475086	0.23126386876
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.23308196048	0.18380485365	-0.32987917406	0.30951867644
B_2	0.01385168928	-0.01249929151	0.00393552599	-0.00691405350
B_3	-0.00534224498	0.01425939680	0.04115511761	-0.03965053817
B_4	-0.01698373939	0.01906607760	0.03282468741	-0.03196380393
B_5	-0.11071224602	0.10006083631	0.10451415651	-0.10020629644

Table 15: Model parameters for the linear ARX model obtained from experiment NOL 1

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.09205646210	-0.18009788354	0.08264130209	-0.84547063939
A_2	0.03032499066	0.31154799908	-0.10702718914	-0.33144788177
A_3	0.17929310379	-0.06191085872	-0.06687894759	0.06088284265
A_4	-0.30945294275	-0.48675354903	0.21572881023	0.34260777524
A_5	0.22038633001	0.39963504877	-0.12635301022	-0.20360813067
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.24083758187	0.19315807102	-0.32655045591	0.31429987833
B_2	0.08567978640	-0.07659926150	-0.06751431906	0.06100119373
B_3	-0.02650209212	0.02525239705	0.07364486578	-0.06685751335
B_4	-0.03639510878	0.04927427012	0.06369340042	-0.06666177072
B_5	0.18159646002	-0.17460032386	-0.09450130419	0.09277011940

Table 16: Model parameters for the linear ARX model obtained from experiment NOL 2

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.85847488142	0.02398926872	0.05064682937	-1.86104337547
A_2	0.76071913241	-0.07165965222	-0.04903257334	0.87693294007
A_3	0.14164279285	0.01657898854	-0.00073789704	-0.01444070950
A_4	0.02833518620	0.03841570957	0.0346071982	0.06928001689
A_5	-0.06755011097	-0.00894925099	-0.02990706294	-0.06777619681
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.25270285779	0.18362466469	-0.279860093881	0.25469002880
B_2	0.22379585464	-0.16068563397	0.22837499210	-0.21106921228
B_3	0.03248578061	-0.01796706829	-0.00061501183	-0.00008163213
B_4	0.00048611405	-0.01301990321	0.00177894993	-0.00113481136
B_5	-0.02142288127	0.01663316895	-0.02498038644	0.02395966480

Table 17: Model parameters for the linear ARX model obtained from experiment NOL 3

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.92006860924	-0.04836966715	0.04797748177	-1.91329927264
A_2	0.93480053777	0.05329017815	-0.02077891324	0.99123552070
A_3	0.11200660762	-0.04271546579	-0.03229052836	-0.06736710805
A_4	-0.07506103206	0.08021146912	-0.00865060862	0.04167168465
A_5	-0.04143788657	-0.04464535638	0.01865384835	-0.04404738701
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.18991783638	0.17633543959	-0.25193472608	0.214726214771
B_2	0.20714354503	-0.18700305562	0.21397343843	-0.1841853698
B_3	-0.01792465783	0.01230407290	-0.01001059473	0.00883565498
B_4	0.00048611405	-0.01301990321	-0.00853642447	-0.00199924156
B_5	-0.02499396562	0. 02485112960	-0.00229233417	0.00392408965

Table 18: Model parameters for the linear ARX model obtained from experiment NOL 4

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.53375927843	0.22658160946	0.18708355418	-1.26741655361
A_2	0.67269874579	-0.29813721338	-0.29072400940	0.20424038412
A_3	-0.13025058335	0.17703085018	0.13036884804	0.03994323238
A_4	0.05424920199	-0.17336974243	-0.03659238368	-0.00429562156
A_5	-0.02929166414	0.05835009926	0.00542674510	0.03843579365
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.22750814662	0.17481022400	-0.33066693627	0.31817344528
B_2	0.04903639448	-0.02321399278	0.04707786896	-0.05435029888
B_3	-0.00044057631	0.00052566465	0.05128284697	-0.04310349554
B_4	-0.03635823383	0.03821547699	0.00305825300	-0.00652497254
B_5	0.00504566617	-0.02146965031	0.01504309738	-0.01254326425

Table 19: Model parameters for the linear ARX model obtained from experiment NOL 5

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.02330230763	0.00827106824	-0.01596992339	-1.01084852107
A_2	0.08336739801	-0.05116322064	0.00580034430	-0.01284217518
A_3	0.03814134902	0.04860700089	0.00659677193	0.04681396839
A_4	-0.03759125290	-0.04913099710	-0.01846024838	-0.03709246571
A_5	-0.02919306781	0.03292734173	0.03016913505	0.04061838117
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.00613141996	-0.00613141996	-0.00088006321	-0.00088006321
B_2	-0.00613141996	-0.00613141996	-0.00088006321	-0.00088006321
B_3	-0.00613141996	-0.00613141996	-0.00088006321	-0.00088006321
B_4	-0.00613141996	-0.00613141996	-0.00088006321	-0.00088006321
B_5	-0.00613141996	-0.00613141996	-0.00088006321	-0.00088006321

Table 20: Model parameters for the linear ARX model obtained from experiment NOL 6

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.114002892253	-0.05338478793	0.02738858062	-1.11883741108
A_2	0.14242078005	-0.04119998167	-0.03775028957	0.15251875982
A_3	0.03737727205	0.13319563205	0.00819033672	0.02630111623
A_4	-0.03250354385	-0.09500906209	-0.05345777889	-0.08732148267
A_5	-0.01505587345	0.05547319421	0.05742538562	0.04765961770
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.25006848782	0.2032518720	-0.32637138472	0.34474505317
B_2	0.04701722955	-0.02984412197	-0.00144704355	-0.03923475136
B_3	0.00543629234	-0.03035793950	0.00594319006	-0.00738877413
B_4	-0.03784746424	0.03286285141	-0.01025684615	0.01077415319
B_5	0.01566220987	-0.00568094039	0.03599468129	-0.02187573308

Table 21: Model parameters for the linear ARX model obtained from experiment NCL 1

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.11301160385	-0.0269242806	0.02420332459	-1.13461008448
A_2	0.11082500672	-0.08829302484	-0.03149642513	0.17811888597
A_3	0.08700195529	0.15154080440	0.00707647950	0.01438629105
A_4	-0.04232323370	-0.10931823652	-0.02979242783	-0.08513532261
A_5	-0.02360280313	0.07017770224	0.03067990719	0.04792701672
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.24749965791	0.19856873862	-0.31721937074	0.34699736010
B_2	0.02593127353	-0.01859365879	-0.00156039279	-0.04961429516
B_3	0.02890661862	-0.03919153271	0.01542599627	-0.00353180637
B_4	-0.02796951034	0.03516337797	-0.00720827905	0.00544434531
B_5	0.00553670580	-0.01369541195	0.02797600272	-0.02124670983

Table 22: Model parameters for the linear ARX model obtained from experiment NCL 2

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.11048787557	-0.04910182179	0.02820812890	-1.09180124782
A_2	0.13925235680	-0.03258808435	-0.05903603215	0.10254917922
A_3	0.05063460569	0.11010954468	0.04840348839	0.03295365450
A_4	-0.04622293483	-0.09391538879	-0.06231211329	-0.07321257208
A_5	-0.01195410298	0.06164038033	0.04545463819	0.04971441892
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.25711024863	0.20527831025	-0.31768826480	0.33323817006
B_2	0.04482470782	-0.02923171392	-0.00738136942	-0.03386423336
B_3	0.01281678569	-0.02531327670	0.02845830280	-0.01559315292
B_4	-0.03746834477	0.02819231404	-0.01167069670	0.00648723898
B_5	0.01180581676	-0.01152382941	0.03315737592	-0.02180329025

Table 23: Model parameters for the linear ARX model obtained from experiment NCL 3

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.01416966062	0.04509901196	-0.03866205513	-1.10186684370
A_2	0.01337859527	-0.06214108119	0.06066911029	0.03483680617
A_3	0.10197010802	0.04443914686	-0.04142016484	0.05334283952
A_4	-0.06380595939	-0.09660356136	0.02047993830	-0.07307137246
A_5	-0.01659875871	0.06640434037	0.00350599621	0.10353275520
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.23517021985	0.16350455419	-0.32994580983	0.32279683143
B_2	0.01637752456	0.00198399547	0.05216310095	-0.05002251207
B_3	0.01774128257	-0.02135827170	0.02579396474	-0.03651885061
B_4	-0.02387709330	0.02294364304	0.02815259858	-0.01768267312
B_5	0.02424959043	-0.01486247169	0.03137581454	-0.03520650800

Table 24: Model parameters for the linear ARX model obtained from experiment NCL 4

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.02931255332	0.00084461575	0.00110838085	-1.03692535303
A_2	0.06244715449	-0.03438255431	-0.01631473280	0.00763517615
A_3	0.05848848471	0.04886955197	0.03259090868	0.04317709933
A_4	-0.05028460864	-0.07105286076	-0.03976670585	-0.04193639550
A_5	-0.01707555813	0.05207151184	0.02356904503	0.04769740460
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.26075498721	0.20743209687	-0.31663609708	0.31827484135
B_2	0.01355561066	-0.00926026995	-0.00482322729	-0.02477900749
B_3	0.01365931909	-0.00905043179	0.04448628736	-0.01811339944
B_4	-0.04427381617	0.02496001105	-0.01551096542	-0.00035871237
B_5	0.01206321833	-0.01306626099	0.02081923837	-0.01176121848

Table 25: Model parameters for the linear ARX model obtained from experiment NCL 5

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.03552640782	-0.00113780594	0.00412227327	-1.02024268343
A_2	0.07161380813	-0.02454558154	-0.02346160964	-0.01180710915
A_3	0.05232078972	0.03695795056	0.03522385674	0.05893381383
A_4	-0.05212992860	-0.06740053155	-0.03550764402	-0.05100333392
A_5	-0.01219339149	0.05223406242	0.02106515806	0.04419856198
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.26116735104	0.20923804169	-0.31601390318	0.31618232917
B_2	0.01513231568	-0.01352949756	-0.00862673957	-0.01571908717
B_3	0.01019532289	-0.00394311546	0.04458602931	-0.02340267683
B_4	-0.04370543358	0.02316510909	-0.01618131762	0.00575329305
B_5	0.01762419333	-0.01763236910	0.01290994084	-0.00658906165

Table 26: Model parameters for the linear ARX model obtained from experiment NCL 6

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.11412930875	-0.14843607870	-0.05384951677	-1.31859151072
A_2	0.16325534664	0.42086398649	0.04881465447	0.3636052359
A_3	0.11066473944	-0.81490133460	-0.01757343457	0.12237954521
A_4	-0.09606272568	0.88265711502	-0.02261657167	-0.12226928979
A_5	-0.04120095383	-0.33666986541	0.04210523402	-0.03294412830
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.32176538676	0.25317419031	-0.32963506273	0.39207125237
B_2	0.06555836406	-0.13220927012	0.043024394452	-0.11755562439
B_3	0.05328119194	0.21914980801	0.00564490995	-0.041583073015
B_4	-0.06678759180	0.00360410188	0.00360410188	0.01967268392
B_5	-0.01027242782	0.10613904921	0.02652363194	0.00065973465

Table 27: Model parameters for the linear ARX model obtained from experiment NCL 7

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.09637569095	-0.03719963793	0.00360760749	-1.12512807010
A_2	0.07634517430	-0.04930291617	0.01176656140	0.14345780046
A_3	0.11066473944	0.11232067051	-0.01971221878	0.05376633125
A_4	-0.06180884789	-0.08047449626	-0.02826609492	-0.11343889001
A_5	-0.00630199195	0.05302841259	0.03356025367	0.06074848913
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.25153915975	0.20779334120	-0.32203352329	0.34323645660
B_2	0.02492646361	-0.02319253574	0.001376166722	-0.04656522642
B_3	0.02657878330	-0.0335880554	0.01366988815	-0.007979365715
B_4	-0.04292529022	0.02960344702	-0.01069412114	0.01270466562
B_5	0.01800769921	-0.00620082379	0.03237426749	-0.02345541112

Table 28: Model parameters for the linear ARX model obtained from experiment NCL 8

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.12363912109	-0.02816070424	0.03309928905	-1.13283271525
A_2	0.14270896356	-0.07937491109	-0.04942400280	0.16402705546
A_3	0.05123692721	0.14775520446	0.032316084758	0.02802288937
A_4	-0.04554734244	-0.10400686663	-0.05558731353	-0.08983342837
A_5	-0.00649570282	0.06193937781	0.04050178388	0.05094823329
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.252985851103	0.19880252416	-0.31693311316	0.34786286978
B_2	0.04048649763	-0.02071829655	-0.009784038312	-0.04637454715
B_3	0.01419194622	-0.03468224577	0.02116314560	-0.00833607974
B_4	-0.03927986459	0.03452383848	-0.01432664047	0.01076940004
B_5	0.01766723148	-0.00936094837	0.03134706616	-0.02193866773

Table 29: Model parameters for the linear ARX model obtained from experiment NCL 9

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.11411510047	-0.05657354173	0.00518687502	-1.11754722212
A_2	0.11340740639	-0.04953561887	-0.02223744167	0.16049782513
A_3	0.06780488525	0.13887805862	0.015558234158	0.00740081567
A_4	-0.01961492456	-0.11583603673	-0.04546445817	-0.09499475662
A_5	-0.02921021765	0.07922007107	0.04848808338	0.06543981160
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.254097456773	0.19934151027	-0.32221743601	0.34543093461
B_2	0.05218691052	-0.02766784525	-0.005976138802	-0.04532532367
B_3	0.01693313559	-0.03973569230	0.00915254294	-0.00277966215
B_4	-0.02689078042	0.03080987844	0.00465904885	0.00809017966
B_5	0.01295809081	-0.01597382096	0.03306420379	-0.02969961766

Table 30: Model parameters for the linear ARX model obtained from experiment NCL 10

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.12735809499	-0.03239482610	0.01572290021	-1.10451721833
A_2	0.14644813331	-0.08091017494	-0.02939100554	0.13669108103
A_3	0.07226560457	0.15538950379	0.015003501318	0.02642546894
A_4	-0.05529500844	-0.10851902993	-0.03345714221	-0.07130878024
A_5	-0.01559493971	0.06318563190	0.03165161786	0.03390567802
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.266253745323	0.20047125032	-0.31780098694	0.34588274284
B_2	0.05645247703	-0.01788148886	-0.002121736172	-0.03908157110
B_3	0.00982864098	-0.03981194765	0.00751772314	-0.01073813721
B_4	-0.03762579719	0.04047255858	-0.00602885133	0.00721699005
B_5	0.00676751332	-0.00978865712	0.02629735059	-0.01518604603

Table 31: Model parameters for the linear ARX model obtained from experiment NCL 11

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.10720072914	-0.05673161879	0.02492721261	-1.08495455404
A_2	0.13213526240	-0.00625994581	-0.05887883272	0.07049251323
A_3	0.05515001871	0.08857139227	0.052330369968	0.03806720625
A_4	-0.04408303037	-0.08570606040	-0.06631912262	-0.06514521887
A_5	-0.01417025601	0.05612835291	0.04822927681	0.06099235942
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.260523887003	0.20887047973	-0.31861275899	0.32489651486
B_2	0.04680612546	-0.03337486351	-0.000462660792	-0.031924726559
B_3	0.01350282526	-0.02085232344	0.03614610764	-0.02294534346
B_4	-0.03460215898	0.02455817550	-0.00643749536	0.00084837681
B_5	0.00915946870	-0.01132288716	0.03705053704	-0.02577863379

Table 32: Model parameters for the linear ARX model obtained from experiment NCL 12

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.04871126093	0.03795250741	-0.03530447684	-1.12107314963
A_2	0.02122868488	-0.03756415066	0.06440042499	0.01052885887
A_3	0.10589212362	0.02849460049	-0.046435271458	0.06876296267
A_4	-0.06836981432	-0.08628458431	0.02132579066	-0.06547575776
A_5	0.00893484647	0.05570728309	0.00002790269	0.12105556536
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.237227059243	0.17050977233	-0.33101405789	0.31442497739
B_2	0.02030753339	-0.00754208838	0.065664975092	-0.05407509470
B_3	0.02212758844	-0.01738680776	0.03456920387	-0.04743135366
B_4	-0.02388545461	0.01997765520	0.03540012251	-0.02451168309
B_5	0.02467941722	-0.01605595050	0.03624111265	-0.03896281875

Table 33: Model parameters for the linear ARX model obtained from experiment NCL 13

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.05150503819	0.00275228775	0.00100936625	-1.05521730333
A_2	0.06910123122	-0.02526841537	0.05757548532	0.04430822733
A_3	0.05757548532	0.04675334297	0.03197556655	0.06876296267
A_4	-0.04698709556	-0.07031525249	-0.04082889916	-0.03785429897
A_5	-0.00515556068	0.04294618057	0.02499598787	0.06587403651
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.265267574673	0.21168008449	-0.32009547396	0.31147068462
B_2	0.01859912284	-0.01543428954	0.008243031952	-0.02752213870
B_3	0.01844277308	-0.00905364672	0.05151053402	-0.02560675990
B_4	-0.03961871492	0.02172078896	-0.00659727021	-0.00759606455
B_5	0.01358231687	-0.01500906007	0.02612260725	-0.01812213031

Table 34: Model parameters for the linear ARX model obtained from experiment NCL 14

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.05385637088	0.00605945815	-0.00214685229	-1.03337385640
A_2	0.07745154073	-0.03270220848	-0.02346946846	-0.00633580150
A_3	0.05151141681	0.04055084910	0.03358495922	0.06050956842
A_4	-0.04492298298	-0.06293607016	-0.03564005419	-0.04290434077
A_5	-0.00721629029	0.04539525057	0.02753412840	0.04198206075
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.26751812283	0.20869874886	-0.31792275377	0.31611660340
B_2	0.02306029153	-0.01445152723	-0.00059347846	-0.01952386269
B_3	0.01698424453	-0.00674943044	0.04631495022	-0.02859207779
B_4	-0.03762647497	0.01793306890	-0.01030024883	0.00035806418
B_5	0.01608192473	-0.01791729845	0.01745773018	-0.0085982361

Table 35: Model parameters for the linear ARX model obtained from experiment NCL 15

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.07653936026	-0.08577367408	-0.05274428659	-1.21382717130
A_2	0.11568317366	0.24128866916	0.037884040263	0.23189535666
A_3	0.09290207093	-0.49710279904	-0.01274444526	0.08819374944
A_4	-0.06203556356	0.54011557574	-0.02081904203	-0.09068725626
A_5	-0.04676733328	-0.19673110252	0.04475660718	0.00055511083
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.30409179640	0.23716901217	-0.326258476817	0.36098213137
B_2	0.04164287888	-0.08374360998	0.04008309321	-0.08440790850
B_3	0.04084527622	0.13337716124	0.01138576608	-0.03587818898
B_4	-0.04515588081	-0.14798349498	0.00907493622	0.00840545243
B_5	-0.01108775242	0.06757701895	0.02760952111	-0.00946298157

Table 36: Model parameters for the linear ARX model obtained from experiment NCL 16

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.11373017990	-0.05737590102	0.05319841125	-1.16373553639
A_2	0.12518172393	-0.04125307507	-0.05858720138	0.19656205438
A_3	0.06948540813	0.12258736324	0.00961452786	0.01044793555
A_4	-0.05673127121	-0.07831754025	-0.05515629765	-0.10335960379
A_5	-0.00706097638	0.05123262260	0.05271534205	0.07866783558
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.24019694380	0.19939997825	-0.32056446491	0.35239971161
B_2	0.03547295451	-0.03380831234	-0.01074747620	-0.05181964327
B_3	0.02568743102	-0.03120224851	0.01723257148	-0.00782832480
B_4	-0.03908499040	0.02826817735	-0.00415623646	0.00505771832
B_5	0.013483716492	-0.00716646356	0.04445815366	-0.03321700533

Table 37: Model parameters for the linear ARX model obtained from experiment NCL 17

0.2 Models Identified to Approximate the Pilot Scale Distillation Column

It is important to note that the discretisation step size of these models is 10 seconds.

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.64226253762	0.04386878063	0.14191751213	-0.91590198231
A_2	0.79110206508	-0.12145038113	-0.03986093134	0.04786096310
A_3	-0.14202958224	0.02046402331	0.04212848184	-0.09333211826
A_4	0.02756155956	0.11146970816	-0.22952954342	-0.09023278471
A_5	-0.10501517278	-0.12843701332	0.12576498416	0.02059519201
A_6	0.09794104179	0.12414002523	-0.09254680973	0.08119184072
A_7	0.00365694080	-0.10241297107	0.10374837974	0.02107932502
A_8	-0.02009837627	0.04275679039	-0.02720402710	-0.01869916421
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.04004651467	0	-0.00121942216	0.00433566399
B_2	0.00660432404	0	-0.01182523719	0.02492642536
B_3	0.02908581241	0.00045717707	-0.01298954445	0.00536921903
B_4	-0.00773245465	0.00306474281	-0.01097624598	0.00318133343
B_5	0.00240040914	-0.00532951121	-0.00049305569	-0.00597835830
B_6	0.002971553312	-0.00139781606	-0.00772280631	-0.00683025422
B_7	0.00110354397	-0.00139913805	0.00847554225	-0.00523924563
B_8	0.00036153449	0.00131508928	0.01372125291	-0.00202618118
B_9	0	0.00193401670	0	0
B_{10}	0	-0.00120400267	0	0

Table 38: Model parameters for the linear ARX model obtained from experiment DOL 1

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.32739822224	0.14864943311	0.04692083970	-0.68387567791
A_2	0.67598917475	-0.03283379840	-0.10716446036	-0.14122396444
A_3	-0.2547372076	-0.06189077444	0.02800139637	-0.21659117290
A_4	0.05481041256	0.02257495630	0.07513827964	0.10258369581
A_5	-0.10427751513	-0.07301546741	-0.11611063475	-0.08691222848
A_6	-0.07750742015	-0.02050423096	0.04351194751	-0.01550065805
A_7	-0.02190625396	-0.01819378944	-0.01698983443	0.06617591185
A_8	0.08072738619	0.01110396287	0.04241541419	0.014883204474
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.0567833756	0	-0.01134708982	0.00247048766
B_2	-0.00771854441	0	0.01494759336	0.02084047380
B_3	0.01537373656	-0.00077893790	0.01144510508	0.00341830618
B_4	-0.02204413225	0.00413891291	-0.02505321861	-0.00108043383
B_5	0.01564529096	0.00552055302	0.00693788846	-0.00434503956
B_6	0.02094011488	0.00135685050	-0.01031289677	-0.00851896190
B_7	0.00017735557	0.00377741850	0.03105878373	-0.00775707277
B_8	0.02593325391	-0.00121619018	-0.01333672847	-0.00250144808
B_9	0	-0.00674814722	0	0
B_{10}	0	-0.00510315207	0	0

Table 39: Model parameters for the linear ARX model obtained from experiment DCL 1

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-1.03509467472	-0.02038863879	0.13967008092	-0.72809863803
A_2	0.30722473049	0.03510090506	-0.2757522994	0.07944438510
A_3	-0.14180755160	-0.06642868548	0.21022381736	-0.32235583778
A_4	-0.11476570779	-0.00480259269	-0.13764522287	0.11392406161
A_5	0.08619376088	0.04274411055	0.07015117041	-0.07337318275
A_6	-0.06414386327	-0.01669625230	0.00490367732	0.01290153671
A_7	0.02531270102	-0.00297931647	-0.01698983443	0.06617591185
A_8	-0.04300507468	0.02487854825	0.08813789358	-0.02674937310
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.08703476242	0	-0.02111664438	0.00618452960
B_2	-0.01807136818	0	0.03288578973	0.01376425899
B_3	0.04077365902	-0.00519381920	-0.02647845080	-0.00210561914
B_4	-0.01556834220	0.00880381991	0.00611460398	0.00275146928
B_5	0.01921162480	0.00627827603	0.00898031289	-0.00244696755
B_6	0.02361350641	0.00423420216	-0.02040432214	-0.00358004095
B_7	-0.02726660783	0.00038021325	-0.02977485254	-0.00353951398
B_8	0.05488545845	-0.00380990356	0.04043511672	-0.00335217107
B_9	0	-0.00363316179	0	0
B_{10}	0	-0.00413281802	0	0

Table 40: Model parameters for the linear ARX model obtained from experiment DCL 2

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_0	1	0	0	1
A_1	-0.45884675950	-0.06320329927	0.46479163808	-0.75177031383
A_2	-0.18404270255	0.16330901072	-0.44320168347	-0.01900154285
A_3	0.24665581214	-0.10143586768	0.10393677207	-0.16468915843
A_4	-0.02333115412	-0.10777608572	-0.10322483613	0.03601071966
A_5	-0.20023817026	0.16954240234	0.07950244572	-0.08279080657
A_6	0.01204728160	-0.04482029811	-0.04498715983	-0.02933835196
A_7	0.08318314049	-0.07271697329	-0.05709992482	0.12459628636
A_8	-0.11903945519	0.03523598038	0.01489475627	-0.02788312122
B	b_{11}	b_{12}	b_{21}	b_{22}
B_0	0	0	0	0
B_1	-0.12352688561	0	-0.09101296532	-0.00553615279
B_2	-0.05037248407	0	0.02257359158	0.02387416365
B_3	0.02339795732	0.00215803760	-0.05844375438	0.00488654781
B_4	0.10344138459	0.00460681511	0.15045443536	0.00962286357
B_5	-0.09254994127	0.01283884431	-0.07280729473	-0.00003420248
B_6	0.09414164413	0.00016623467	-0.03504722112	-0.00615814174
B_7	-0.11305132830	0.00168267534	0.00043619575	-0.00582530786
B_8	0.12357695305	-0.00708186350	0.04354877817	-0.00680767009
B_9	0	-0.00584745615	0	0
B_{10}	0	-0.00167879767	0	0

Table 41: Model parameters for the linear ARX model obtained from experiment DCL 3

Open and Closed-loop Model Identification and Validation

Hernan Guidi

Open and Closed-loop Model Identification and Validation

by

Hernan Guidi

A dissertation submitted in partial fulfillment
of the requirements for the degree

Master of Engineering (Control Engineering)

in the

Department of Chemical Engineering

Faculty of Engineering, the Built Environment and Information Technology

University of Pretoria
Pretoria

5th December 2008

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Open and Closed-loop Model Identification and Validation

Author:	Hernan Guidi
Date	5th December 2008
Supervisor:	Professor P. L. de Vaal
Department:	Department of Chemical Engineering
	University of Pretoria
Degree:	Master of Engineering (Control Engineering)

Synopsis

Closed-loop system identification and validation are important components in dynamic system modelling. In this dissertation, a comprehensive literature survey is compiled on system identification with a specific focus on closed-loop system identification and issues of identification experiment design and model validation. This is followed by simulated experiments on known linear and non-linear systems and experiments on a pilot scale distillation column. The aim of these experiments is to study several sensitivities between identification experiment variables and the consequent accuracy of identified models and discrimination capacity of validation sets given open and closed-loop conditions. The identified model structure was limited to an ARX structure and the parameter estimation method to the prediction error method.

The identification and validation experiments provided the following findings regarding the effects of different feedback conditions:

- Models obtained from open-loop experiments produced the most accurate responses when approximating the linear system. When approximating the non-linear system, models obtained from closed-loop experiments were found to produce the most accurate responses.

- Validation sets obtained from open-loop experiments were found to be most effective in discriminating between models approximating the linear system while the same may be said of validation sets obtained from closed-loop experiments for the non-linear system.

These finding were mostly attributed to the condition that open-loop experiments produce more informative data than closed-loop experiments given no constraints are imposed on system outputs. In the case that system output constraints are imposed, closed-loop experiments produce the more informative data of the two. In identifying the non-linear

system and the distillation column it was established that defining a clear output range, and consequently a region of dynamics to be identified, is very important when identifying linear approximations of non-linear systems. Thus, since closed-loop experiments produce more informative data given output constraints, the closed-loop experiments were more effective on the non-liner systems.

Assessment into other identification experiment variables revealed the following:

- Pseudo-random binary signals were the most persistently exciting signals as they were most consistent in producing models with accurate responses.
- Dither signals with frequency characteristics based on the system's dominant dynamics produced models with more accurate responses.
- Setpoint changes were found to be very important in maximising the generation of informative data for closed-loop experiments

Studying the literature surveyed and the results obtained from the identification and validation experiments it is recommended that, when identifying linear models approximating a linear system and validating such models, open-loop experiments should be used to produce data for identification and cross-validation. When identifying linear approximations of a non-linear system, defining a clear output range and region of dynamics is essential and should be coupled with closed-loop experiments to generate data for identification and cross-validation.

Keywords: Closed-loop system identification; LTI approximations; Cross-validation; Prediction error method; Identification experiment design, ARX

Acknowledgements

I would like to give my sincere thanks to Professor Philip de Vaal for his guidance and support on not only this dissertation but through out my academic career.

A mis padres y a mi hermano. Me siento eternamente endeudado y les agradezco de mi corazn por todo su apoyo incesante y sin condiciones que ha sido inestimable para mi. Los quiero mucho.

"*Joy in looking and comprehending is nature's most beautiful gift*" – Albert Einstein

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Roman Symbols

$\bar{u}$	Steady State Input
$\bar{y}$	Steady State Output
a, b, c, d	Regression Coefficients
D_c	The set into which $\hat{\theta}_N$ converges
E	Expectation
e	White Noise
$f_e(\cdot)$	Probability Density Function of Noise
G	System Transfer Function
g	Unit Impulse Response
$G(q, \theta)$	Parameterised System Transfer Function
$G_{0,OE}$	OE-LTI-SOE transfer function
G_0	True System Transfer Function
H	Noise Transfer Function
P	Probability
P_θ	Asymptotic Covariance Matrix of θ
q	Time Shift Operator

R_{SW}	Cross-Correlation Function between Signals S and W
R_S	Auto-Correlation Function of Signal S
s	Laplace Operator
S_0	True Sensitivity Function
t	Time
T_N	Total Experiment Time
u	System Input
v	System Disturbance Variable
V_N	Criterion function to be minimised
W	Weighting function
y	System Output
Z_N	Data Set

Acronyms

ACF	Auto-Correlation Function
AIC	Akeike's Information Criterion
ARMA	Auto-Regressive with Moving Average
ARMAX	Auto-Regression with Moving Average Exogenous
ARX	Auto-Regressive with External Input
CCF	Cross-Correlation Function
ETFE	Empirical Transfer Function Estimate
ETFE	Empirical Transfer-function Estimate
FIR	Finite Impulse Response
FPE	Akeike's Final Prediction Error
IID	Independant and Identically Distributed
IV	Instrumental Variable

LS	Least-Squares
LTI	Linear Time Invariant
MA	Moving Average
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
NARX	Non-linear auto-regression with external inputs
OE-LTI-SOE	Output Error LTI Second Order Equivalent
PDF	Probability Distribution Function
PEM	Prediction Error Method
PRBS	Pseudorandom Binary Sequence
PRBS	Pseudorandom Multilevel Sequence
RBS	Random Binary Signal
SNR	Signal-to-Noise Ratio

Greek Symbols

ϵ	Prediction Error
ϵ_F	Filtered Prediction Error
λ_0	Variance of $e(t)$
ω	Frequency
Φ_S	Power Spectrum of Signal S
Φ_{SW}	Cross-Power Spectrum between signal S and W
ψ	The gradient of $\hat{y}$ with respect the θ
σ	Signal Magnitude or Step Size
τ	Operator Index
θ	Parameter Vector
θ_0	Parameter vector reflecting true system

$\varphi(t)$	Regression Vector
$\varphi(t,\theta)$	Pseudo-Linear Regression Vector
$\zeta(t)$	Instrumental Vector

Subscripts

0	The True System
ϵ	Prediction Error
k	Operator Index
N	Discrete Data Index
t	Time
u	System Input
v	System Disturbance
ML	Maximum Likelihood

Superscripts

cl	Closed Loop
m	Measured Value
T	Matrix or Vector Transpose
IV	Instrumental Variable
LS	Least-Squares

Others Symbols

$\bar{R}_u$	Covariance matrix of signal u
$\bar{V}_N$	Limit of the Criterion Function
$\ell(\cdot)$	Prediction Error Norm
$\hat{\theta}$	Estimated Parameter Vector
$\hat{\hat{G}}_(N)(e^{j\omega t})$	Empirical Transfer Function Estimate

$\hat{\theta}_N$	Estimate of θ_0 based on data set Z^N
$\hat{y}(t \mid \theta)$	Output prediction based on parameterised model
$\hat{y}(t \mid t-1)$	Prediction of output y at time t based on past data
$\mathcal{F}$	Fourier Transform
$\mathcal{L}$	Laplace Transform
$\mathcal{M}$	Model Structure
$\mathcal{M}(\hat{\theta}_N)$	Model member out of model set defined by θ estimate
$\mathcal{S}$	The True System
$\mathbf{Cov(x)}$	Covariance Matrix of vector x
$\mathbf{R}^d$	Euclidian d-dimensional space
w.p. 1	With Probability 1

CHAPTER 1

Introduction

This chapter presents the introduction to this work. A concise presentation of the problem statement and investigative approach is made followed by some introductory concepts and theory that further detail the investigative direction taken. The last section of this chapter presents the structure of this dissertation by briefly indicating what the subsequent chapters entail and contain.

1.1 Introduction

System identification deals with the problem of building mathematical models of dynamical systems from observed data. The generation of such observed data is dependent on the identification experiment (Ljung, 1987). Given the system to be identified normally operates as part of a closed-loop configuration, there are various experimental approaches that may be used to generate identification data. An obvious approach is an open-loop experiment where controller feedback loops are opened, the system is disturbed, and data accumulated. Alternatively the identification experiment may be conducted under closed-loop conditions where the controllers maintain their normal operating condition and disturbances are introduced via setpoint changes and/or superimposed input signals.

Literature does indicate open-loop experiments to allow for more persistent excitation of the system and generation of more informative data (Goodwin & Payne, 1977). The question at hand however is what are the advantages and disadvantages of each approach and how do they manifest themselves in the quality of the identified model. From this, the importance of model validation is revealed since it is through validation that models are discriminated and measures of accuracy are determined. Furthermore, there are several other experiment design variables, namely those concerned with the characteristics of disturbance signals. It is not so well understood how the sensitivities between these

experimental variables and model accuracy change given different feedback conditions. The question of how these sensitivities change further extends itself when considering the effects of approximating a non-linear system with a linear model.

The objective of this work may thus be given by the following points:

- Compile an extensive literature survey on system identification approaches with a specific focus on closed-loop system identification, model validation techniques and identification experiment design.
- Through the use of known linear and non-linear models, simulate open and closed-loop system identification and validation experiments to obtain and validate linear approximations of these known models. From this, obtain an understanding with regard to the sensitivities of the key identification variables, specifically the disturbance signal characteristics, and how they change given a non-linear system.
- Specifically concentrate on cross-validation approaches and assess how the different experimental conditions used to generate validation data affect the discrimination and bias of validation sets.
- Use a pilot scale distillation column to obtain an understanding regarding pragmatic issues of implementing identification and validation techniques on a real system that would not necessarily arise in simulated experiments.

While the literature survey will attempt to cover most of the pertinent methods and techniques, the identification efforts through simulated experiments and those performed on the distillation column will use the prediction error method as the identification method and the direct approach for the closed-loop identification experiments. Furthermore, the model type and structure used will be limited to parametric ARX structures. These choices will be explained and justified in the sections and chapters to come.

1.2 The Philosophy and Purpose Behind Modelling and System Identification

It is appropriate at this point to mention a maxim accredited to statistician George Box: "All models are wrong, but some are useful" (Box, 1979). From this one may gather that pursuing the perfect model to perfectly represent the system is baseless and that, defining the purpose for a model is the first step in obtaining a useful one. Dynamical models have found application in areas as diverse as engineering and the hard sciences, economics, medicine and the life sciences, ecology and agriculture (Norton, 1986). The same few basic purposes underlie identification in all these fields.

The next few sections further present some of the key model applications and purposes for system identification.

1.2.1 Models to Satisfy Scientific Curiosity

Norton (1986) states that a principal characteristic of science is its use of mathematical models to extract the essentials from complicated evidence and to quantify the implications. This is the essence of scientific modelling, to increase an understanding of some mechanism by finding the connections between observations relating to it.

1.2.2 Modelling for Diagnosis of Faults and Inadequacies

A great benefit of system identification is its ability to uncover shortcomings and anomalies. In this application models are used in the disclosure of unexpected or atypical behaviour of a system. It is generally understood that modelling for fault detection may require efforts that are typically different to those required for other purposes. The main reason for this is that, in order to assure a robust diagnostic system which is aptly able to isolate faults, modelling of abnormal regions of operation is essential (Frank, 2000)

1.2.3 Models for Simulation

Mathematical models allow for the possibility to explore situations which in actuality are hazardous, difficult or expensive to set up. Aircraft and space vehicle simulators are good examples of this. Through such applications of models valuable insight may be obtained with regard to the dynamics of the system being studied. Model accuracy and comprehensiveness are essential for such purposes (Norton, 1986).

1.2.4 Models for Prediction and Control

The desire to predict is a common and powerful motive for dynamical modelling. From a strictly pragmatic point of view, a prediction model should be judged solely on the accuracy of its predictions. Norton (1986) states that the plausibility and simplicity of the prediction model and its power to give insight are all incidental. Section 1.9 further elaborates on the issues of system identification and control.

1.3 Model Structures and Model Types

Provided a purpose for modelling, the next most definitive issue is to find a suitable model type and structure that would adequately represent the system keeping in mind its intended purpose. Ljung (1987) states a basic rule in estimation, namely that, one

should not estimate what is already known. In other words, one should utilise prior knowledge and physical insight about the system when selecting the model structure and type to be used. Literature customarily distinguishes between three types of models based on the prior knowledge at hand. These are as follows (Sjöberg et al.):

White Box Models : This is the case when a model is perfectly known and has been possible to construct it from prior knowledge and physical insight.

Grey Box Models : This is the case when some physical insight is available, but several parameters remain to be determined from observed data. It is useful to consider two sub-cases

- Physical Modelling: A model structure can be built on physical grounds, which has a certain number of parameters to be estimated from data. This could for example be a state space model of given order and structure.
- Semi-physical Modelling: Physical insight is used to suggest certain non-linear combinations of measured data signal. These new signals are then subjected to structures of black box character.

Black Box Models : No physical insight is available or used, but the chosen model structure belongs to families that are known to have good flexibility and have been successful in the past.

From these categories it is established that the type of model to be used in an identification task greatly depends on the information at hand. On the one end of the scale are white box models, which are used when the physical laws governing the behaviour of the system are known. On the other end are black box models where no physical insight on the system is known.

It has been found that despite the quite simplistic nature of many black-box models, they are frequently very efficient for modelling dynamical systems. In addition to this, such models typically require less engineering time to construct than white-box models (Forssell, 1999).

An additional model characteristic that is of importance is whether it is defined as linear or non-linear. While it is understood that most systems in "real-life" to some degree are non-linear, linear models are used more in practice. It is true that they provide idealised representations of the "real-life" non-linear systems, even so, they are justified in that they usually provide good results in many cases (Ljung, 1987).

More details on different models and modelling approaches are given in chapter 2

1.4 System Identification Approach and Concepts

As has been mentioned earlier, system identification deals with the construction of models from data. The components that characterise the procedure of such model constructions are now further introduced, with figure 1.1 showing the procedural interactions between the key components (Norton, 1986):

Problem Formulation : This initial step in identifying a system is to first establish all that is necessary in order to determine the type of identification effort. That is, whether a priori knowledge is available and what the purpose of the model will be.

Experimental Design : This section is perhaps the most important. Here the system to be identified is defined in terms of the inputs, outputs and assumptions. Typically the system to be identified is part of a larger system, it is thus necessary to make assumptions and define which components will be identified. From these definitions and assumptions, an experiment is designed. The key design variables are typically the characteristics of the disturbance signal and whether the experiment is done under open or closed-loop conditions (further discussed in section 1.5).

Data Preparation : Before the data obtained from the identification experiment may be used to generate a model, it needs to be prepared. This preparation involves cleaning the data of drifts, outliers and trends.

Estimation : This identification component involves the mapping of information from the experimental data set to the model coefficients. This typically involves optimisation routines of functions with norms that are to be maximised or minimised.

Model Validation : This section is essential in determining the success of any identification efforts. While section 4.1 details the different approaches to model validation, it is simply understood as the determination of whether the model is "good enough" for its intended use.

More detail regarding each of these identification components is provided in the subsequent chapters.

1.5 Closed-Loop System Identification

A central focus point of this work is closed-loop system identification. Closed-loop identification results when the identification experiment is performed under closed-loop conditions, this is, where a signal feedback mechanism is incorporated into the system so as to control a variable via manipulation of another. Figure 1.2 illustrates the closed and

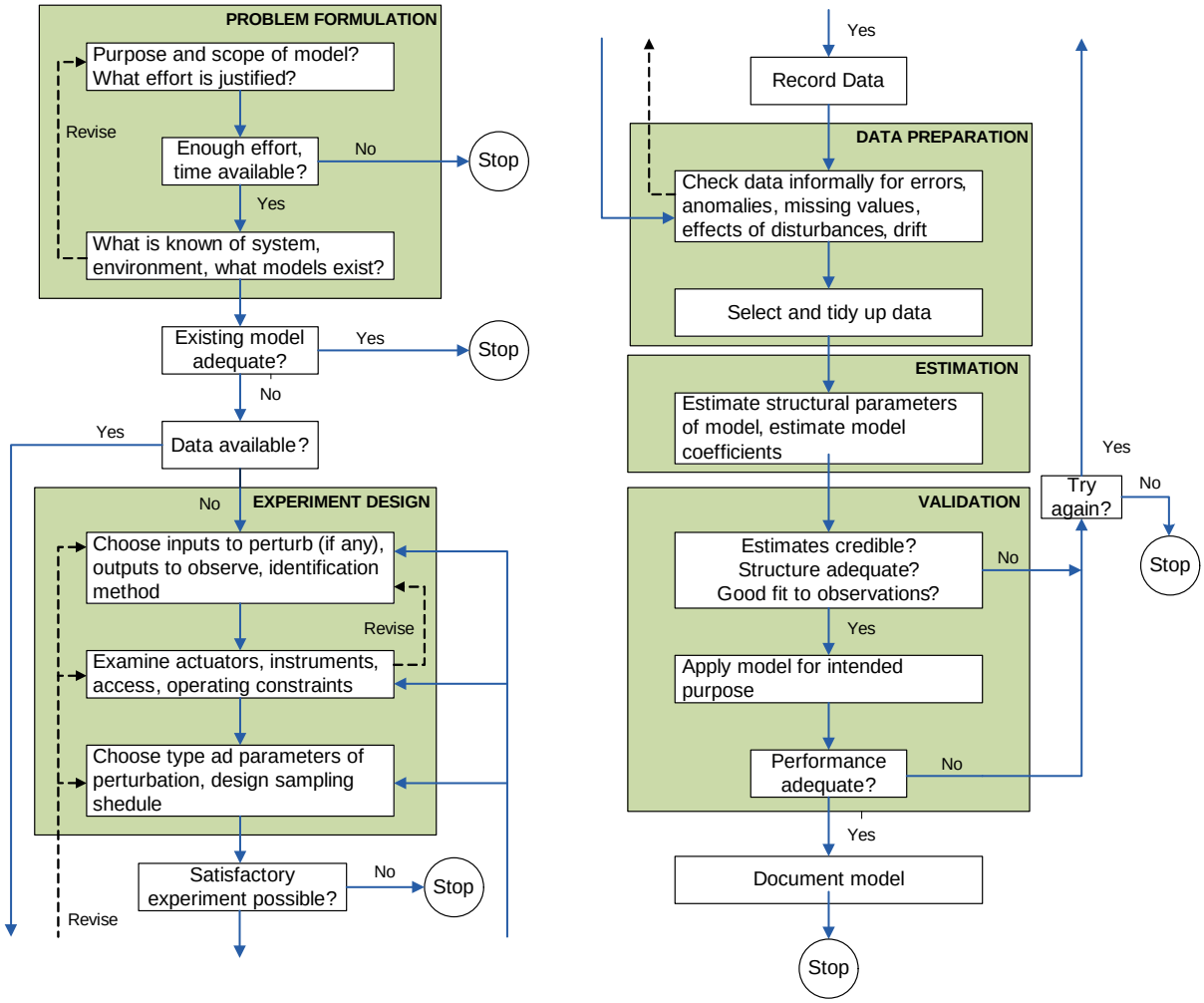


Figure 1.1: Flow diagram illustrating identification and validation logic sequence

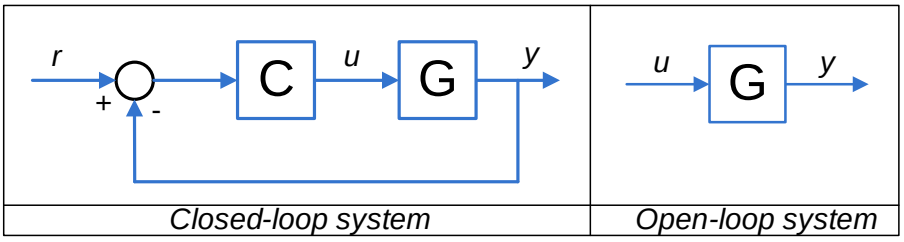


Figure 1.2: Comparison of closed and open-loop configurations

open-loop configurations next to each other with u being the system input, y the system output, r the reference or setpoint signal, C the controller and G represents the system.

Due to possible unstable behaviour of a plant, required safety and/or efficiency of operation, many industrial processes need to be identified under closed-loop conditions. This is especially the case when attempting to identify biological and economic systems that inherently contain feedback loops that can not be removed or opened.

As it will be demonstrated in chapter 3, closed-loop system identification has some advantages and disadvantages. The primary advantages may be given as (Forssell & Ljung, 1999):

- Due to the feedback loop being closed, the disturbance to the system is reduced.
- Allows for better identification of the system given output constraints.
- Allows for better identification of control relevant dynamics.

the disadvantages may be stated as follows (Forssell, 1999):

- The feedback loop causes correlations between the input signals and the system noise which results in the failure of many identification methods. Additionally several model structures become more difficult to use specifically the non-parametric structures (Ljung, 1987).
- By definition a closed-loop system contains a controller which has a purpose of reducing system disturbances. This reduces the excitation of the system and thus the information content in the data used to identify models from.

Further details of these advantages and disadvantages will be given in the theoretical chapters that follow. Of specific interest will be how these characteristics extend to linear estimations of non-linear systems.

1.6 Identification Experiment Design

The primary principle behind identification experiment design is that the choice of experimental conditions has a significant effect on the achievable accuracy of the identified model. The following are the primary identification experiment design variables (Goodwin & Payne, 1977):

1. Inputs signal characteristics.
2. Total experimental time.
3. Total number of samples.

4. Sampling rate

The objective behind optimising these variables is to obtain the most informative data of the system with the model objective in mind while keeping within specified constraints. The primary variable of those mentioned is undoubtedly the input signal. This is the signal which disturbs the system in a specific manner so as to obtain informative data. There are several characteristics of the signal that may be specified so as to maximise the persistent excitation of the system. These are typically the signal type, amplitude constraints, frequency characteristics and power constraints.

It is important to note that the success of accurately identifying a system depends on the generation of informative data. For open-loop experiments the only criteria for informative data is that the input signal be sufficiently exciting. For the closed-loop experiments this is not sufficient (Forssell, 1999).

1.7 Model Validation and Discrimination

Model validation can loosely be defined as the determination of whether a model is good enough for its intended use. While there are several statistical tools available that validate models based on confidence intervals and hypothesis testing, this work will focus on cross-validation approaches. Cross-validation implies the condition where the identified model response is simulated using different measured input data to that used for identification and the output is compared to the measured output (Forssell, 1999). From these simulation results, the differences between the model output and the measured output, defined residuals, may be analysed in the time domain and in the frequency domain. Further more the residuals may be tested for correlations to allow for further insight into the extent of model inaccuracy.

An important aspect of this work will be to test the validation results. That is, to determine how accurate are the validation results and whether certain validation techniques are more biased than others. This effectively means assessing the ability and conditions by which validation sets discriminate between models. Since cross-validation relies on experimental data, the experimental conditions under which the validation data are generated undoubtedly should affect the validation results. Determining validation results sensitivity to validation data and the conditions they are generated under will be an essential component of this work.

1.8 Linear Approximations of Non-linear systems

A central investigative aspect of this work will be identification and validation of linear approximations of non-linear systems. This delves into assessing what experimental

characteristics improve linear approximations of non-linear systems. Additionally, model validation takes on some different challenges to the typical validation of linear models of non-linear systems. It becomes important to determine to what extent does the linear system approximate the non-linear system.

In studying linear estimation of non-linear systems convergence theory, Enqvist et al. (2007) states that one may establish the "best" linear representation of a non-linear system and assess the factors that affect the convergence of linear estimations towards this best linear representation. So the question becomes, can one determine what is the best linear approximation of the non-linear system from experimental data and how does one validate a linear model in view of this best linear approximation? While this issue of establishing the best possible linear approximation and determining how close linear models are to this best approximation is not directly addressed in this work, it is of relevance and discussed in the validation theory presented and assessment of model validations.

1.9 Identification and Control

Gevers (1986) proclaims that few people would object to the assertion that control engineering has been the key component in the development of identification theory. This may be specifically attributed to the fast development and implementation of model predictive controllers (MPC) and the extension of such controllers to be adaptive. This implies that models are to be continuously generated, validated and implemented in such controllers with the objective of optimising controller performance. This while maintaining the system under closed-loop conditions.

As stated earlier, one of the advantages of closed-loop system identification is its ability to identify better models for control. Zhu & Butoyi (2002) attributes this to two reasons:

1. The identification experiment efforts in information generation may be concentrated over strictly defined output ranges due to feedback control.

2. Closed-loop experiments have been shown to be particularly accurate at identifying frequency regions that are relevant for control.

3. Closed-loop experiments have been prominent features in identifying regions where uncertainty is not tolerated. Reduced model uncertainty is essential in producing robust model based control.

These points indicating the central role closed-loop system identification plays in identification for control alone justify the importance of studying closed-loop identification and validation.

1.10 Scope and Outline of Dissertation

This document comprises of four theoretical chapters, two experimental chapters, one discussions and conclusions chapter and a recommendations chapter.

It is noted that a relatively detailed literature survey of relevant system identification theory is one of the primary outcomes of this work. The theoretical chapters may be described as follows:

Chapter 2 - System Identification Theory : Presents the fundamental theory and background of system representation and system identification. While a general basis is established that covers most of the different approaches, the later sections presenting the asymptotic and convergence theories focus more on the parametric model structures.

Chapter 3 - Closed-loop System Identification Theory : Here the closed-loop system identification problem is presented along with the key concepts and approaches. As with the general system identification theory, theory focuses on parametric approaches.

Chapter 4 - Model Validation Theory : This chapter presents the fundamentals behind model validation, specifically residual analysis. Additionally the chapter delves into convergence theory of linear approximations of non-linear systems and how these approximations of such systems may be validated.

Chapter 5 - Experimental Design Theory : Presents the theory regarding designing identification experiments. Basic theory is presented regarding the conditions for informative data and how the different design variables affect the experiment outcomes.

The chapters following these are those concerned with the simulation and experimentation work. The approach taken was such that simulation efforts in system identification and validation were conducted before attempting to identify a pilot scale distillation column. These chapters may be further detailed as follows:

Chapter 6 - Identification of Simulated models : This chapter presents simulated identification and validation efforts with a purpose of gaining insight into sensitivity relationships between experimental design variables and identification and validation results. Two systems were simulated, one via a linear ARX model and the other via a non-linear ARX model. Several identification experiments were designed for both and models were identified from all the data sets generated by these experiments. In a similar fashion several different experimental designs were used to generate validation sets. All the identified models were validated against all the validation sets for both the linear and non-linear systems.

Chapter 7 - Identification of a Distillation Column : This chapter presents efforts in identifying and validating models of a pilot scale distillation column. Findings from the previous chapter were used to design identification and validation experiments under the knowledge that the column is highly non-linear.

Chapter 8 - Discussion and Conclusions : Presents discussions and conclusions based on the simulations and distillation column identification and validation efforts and results.

Chapter 9 - Recommendations and Further Investigation : This chapter presents recommendations regarding general approaches to system identification and validation based on the study of literature and experimental findings presented in this work. Additionally, suggestions are made on how to further work in the identification of the pilot scale distillation column.

CHAPTER 2

System Identification Theory

This chapter stresses the theoretical background and principles defining system identification. The general fundamentals, methods and approaches to representing linear and non-linear systems are surveyed. The later sections in this chapter focus on parametric estimation methods. Specific attention is given to the prediction error method and the convergence and asymptotic properties of this method. Theoretical background chapters 3 and 4, on closed loop system identification and model validation respectively, grow upon this chapter.

2.1 Background and History

Up until the 1950's most of control design relied on Bode, Nyquist and Nichols charts or step response analyses. These methods where mostly limited to SISO systems. In the early 1960's Kalman introduced the state-space representation and through this established state-space based optimal filtering and optimal control theory with Linear Quadratic optimal control as a cornerstone for model-based control design (Gevers, 2006).

It was on the heels of this introduction and rise of model-based control that system identification developed. Literature shows how two landmark papers, Åström et al. (1965) and Ho & Kalman (1965), gave birth to the two primary streams of research in system identification in the mid 1960's. Ho & Kalman (1965) gave the first solutions to the state-space realisation theory. This went on to establish stochastic realisation and later to subspace identification. The contribution by Åström et al. (1965) was such that it laid the foundations for the highly successful Prediction Error Identification framework. These two mainstream identification techniques, subspace and prediction error, still dominate the identification field today.

For a more elaborate extension of the advances made in the field of system identifi-

cation over the more recent years and the origin of these advances, the reader is referred to literature. Specifically (Raol & Sinha, 2000) and Gevers (2006).

While there have been, and will continue to be, many sources and contributions to the ever growing compendium of System Identification techniques, present day contributions made by Lennart Ljung must be commended. The book "System Identification: Theory for the user", Ljung (1987), is said to have had a major impact on the engineering community of system identifiers and has established system identification as a design problem that plays a central role in many engineering problems. Much of the literature presented in the forthcoming sections will be from this text.

2.2 Linear System Representation and Fundamentals

A very well understood and common approach to identification is to introduce known input disturbances to the system and record the system's response. From knowledge of the input and the type of response it yields, system dynamics may be extracted and represented in an appropriate form. For such an identification approach typical inputs are the impulse, unit pulse, step and sinusoidal-waves with figures 2.1 and 2.2 illustrating (Norton, 1986: 24). Figure 2.2 illustrates the important connotation of a linear system and assumption that facilitates the extraction of a system's dynamics from a response to an input. That is, a feature in the dynamics of a linear system is the additivity of output responses to separate inputs.

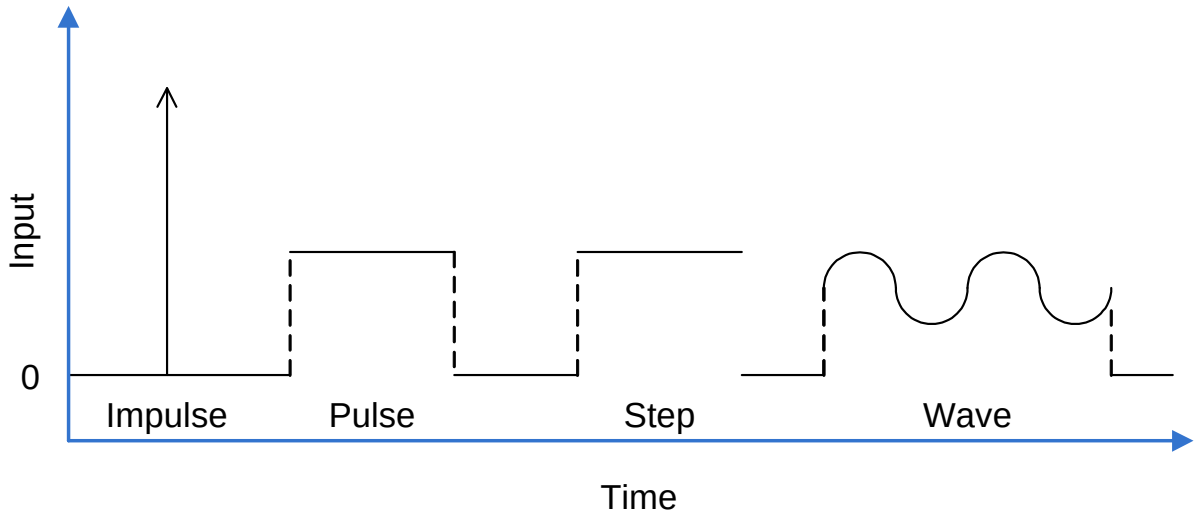


Figure 2.1: Possible Selection of Input Disturbances

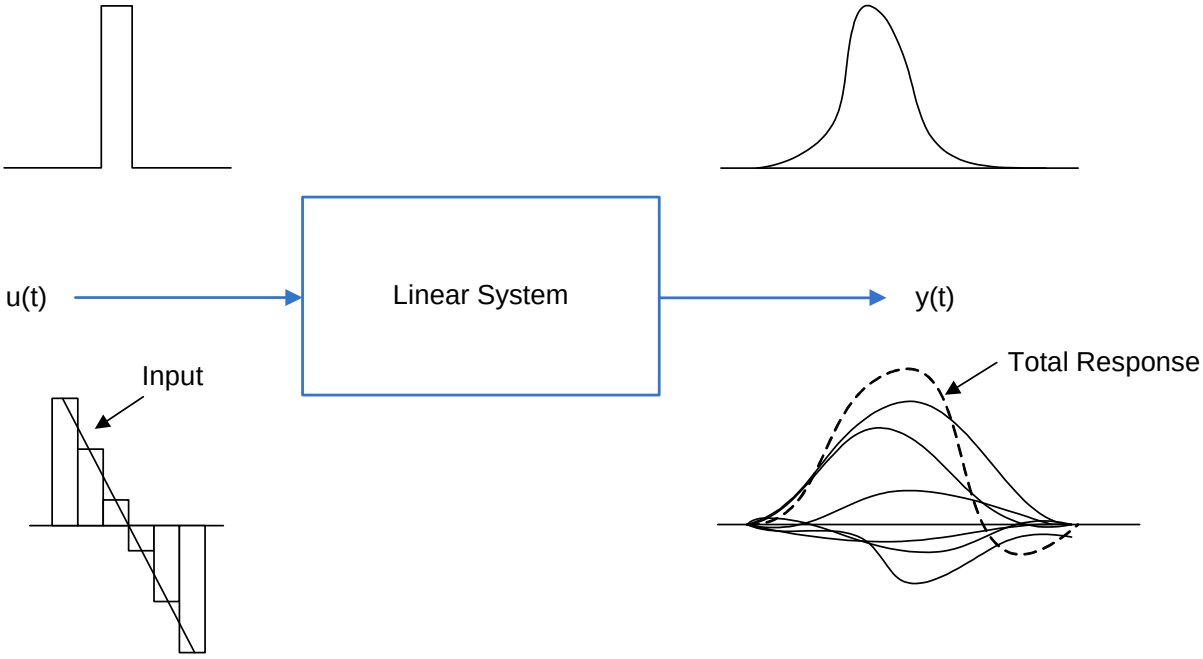


Figure 2.2: Linear System Response

2.2.1 Linear Response Modelling

Impulse, Pulse and Step Responses

Common types of identification approaches based on recorded system responses are those of the impulse and unit pulse response. An impulse, which is a pulse covering an infinitesimally short time, $\Delta\tau$, may be used to obtain working knowledge of the system. In other words, the total response of a system may be determined by summing the contributions from infinitesimally short sections of the input. With the impulse being precisely such an input, the total response of a system may be given via the convolution or superposition integral (Norton, 1986: 24),

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau \tag{2.1}$$

where g is the response to the impulse. A unit-pulse response can be used in the same manner. The advantage this approach has over the impulse response is that in practise it is easier to implement a unit pulse than an infinitesimally short pulse. This superposition of an isolated response may also be represented in discrete time form,

$$y_t = \sum_{k=0}^{\infty} g_k u_{t-k} \tag{2.2}$$

this is a useful representation as typically recorded data is obtained in discretised form. A system representation in the form of a convolution sum based on a step response is

also obtained by extension of these concepts.

Sinusoidal-Wave Responses

Although a larger component of this project is centred on discrete-time approaches it is important to note the sine-wave response based system representation. Once again the convolution integral is used to represent the system response to a sine-wave. Algebraically this is stated as the response to $e^{j\omega t}$, applied from $-\infty$,

$$y(t) = \int_{-\infty}^{\infty} g(t-\tau)e^{j\omega\tau}d\tau \tag{2.3}$$

This form of representation is central to frequency domain based system representations. This is further discussed in the next section.

2.2.2 Transforms and Transfer Functions

Fourier and Laplace transforms form the basis of classical control design. The Laplace and Fourier transform definitions are recalled as (Norton, 1986: 30),

$$\mathcal{L}\left[f(t)\right] \equiv F(s) \triangleq \int_0^{\infty} f(t)e^{-st}dt \tag{2.4}$$

$$\mathcal{F}\left[f(j\omega)\right] \equiv F(j\omega) \triangleq \int_0^{\infty} f(t)e^{-j\omega t}dt \tag{2.5}$$

and have the major advantage in that through the transform they simplify the input-output convolution relation given in equations (2.1) and (2.3) to a multiplication of rational transforms,

$$Y(s) = G(s)U(s) \tag{2.6}$$

$$Y(j\omega) = G(j\omega)U(j\omega) \tag{2.7}$$

$G(s)$ being the Laplace transform of the impulse response $g(t)$ and $G(j\omega)$ being the frequency transform of the impulse response. This makes solving convolution equations much easier as the appropriate transform inversion of the transform solution provides the time domain solution (Norton, 1986: 35). All these characteristics make using a transfer function very practical in terms of quantitative descriptions a system.

At this point it is convenient to introduce the basic notation that will be used from here on to represent the essential components that describe a system. Recalling the discretised form of the convolution integral, equation (2.2), this may be given in shorthand notation by the introduction of the time shift operator q,

$$qu(t) = u(t+1), \qquad q^{-1}u(t) = u(t-1)$$

this allows equation (2.2) to be represented in the following form:

$$y(t) = \left[\sum_{k=1}^{\infty} g(k)q^{-k}\right] u(t) \tag{2.8}$$

A complete linear model, illustrated in figure 2.3, may thus be defined as follows (Ljung, 1987: 69):

$$\begin{gathered} y(t) = G(q)u(t) + v(t), \qquad v(t) = H(q)e(t) \\ f_e(\cdot), \text{ the PDF of } e \end{gathered} \tag{2.9}$$

with

$$G(q) = \sum_{k=1}^{\infty} g(k)q^{-k}, \qquad H(q) = 1 + \sum_{k=1}^{\infty} h(k)q^{-k} \tag{2.10}$$

$G(q)$ is the transfer function of the linear system subject to an input $u(t)$ resulting in the output response $y(t)$. The system disturbance variable is represented by $v(t)$, which is further defined through the noise model $H(q)$. $H(q)$ is assumed monic, i.e. $H(q) = \sum_{k=0}^{\infty} h(k)q^{-k}$, and inversely stable, i.e. $(H(q))^{-1}$ is stable. Furthermore, $e(t)$ is an unpredictable white noise signal with zero mean, a variance of λ_0 and a probability density function (PDF) of $f_e(\cdot)$. This probability density function allows for a description of the probability that the unpredictable values of $e(t)$ fall with in certain ranges, this is given by:

$$P(a \leq e < b) = \int_a^b f_e(x)dx$$

This representation of a system in a discrete form is advantageous since this is the typical data acquisition mode (Ljung, 1987: 14).

2.2.3 Prediction and Parameterisation

Given the description expressed by equation (2.9) and the assumption that $y(s)$ and $u(s)$ are known for $s \leq t-1$, since

$$v(s) = y(s) - G(q)u(s) \tag{2.11}$$

$v(s)$ is also known for $s \leq t-1$, this means that it is possible to obtain a prediction for

$$y(t) = G(q)u(t) + v(t)$$

given by the following one-step-ahead prediction expression (Ljung, 1987: 56):

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + \left[1 - H^{-1}(q)\right] y(t) \tag{2.12}$$

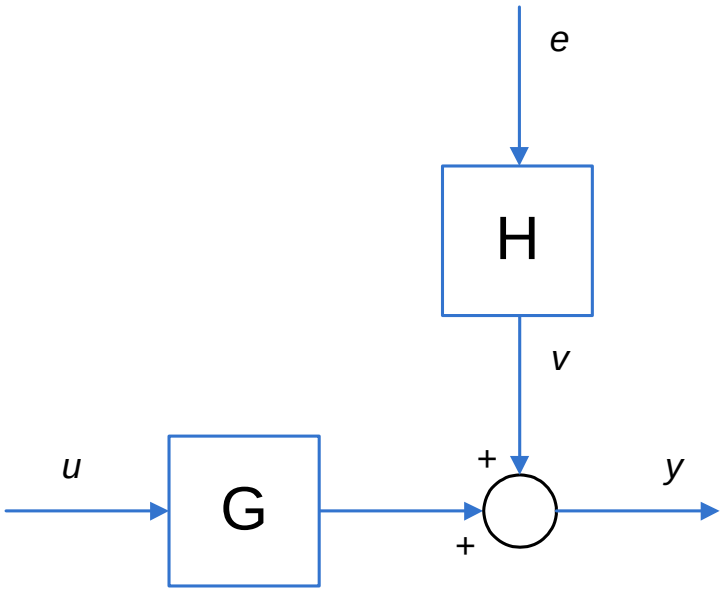


Figure 2.3: Open-loop Transfer Function Representation

With some further manipulation of equation (2.12) the predictor may be extended to a k-step-ahead predictor (Ljung, 1987: 57),

$$\hat{y}(t|t-k) = W_k(q)G(q)u(t) + [1 - W_k(q)]\, y(t) \tag{2.13}$$

with

$$W_k(q) = \bar{H}(q)H^{-1}(q), \qquad \bar{H}_k(q) = \sum_{l=0}^{k-1} h(l)q^{-l} \tag{2.14}$$

From equation (2.9) and (2.12) it is found that the prediction error $y(t) - \hat{y}(t|t-1)$is given by

$$y(t) - \hat{y}(t|t-1) = -H^{-1}(q)G(q)u(t) + H^{-1}(q)y(t) = e(t) \tag{2.15}$$

This means that variable $e(t)$ represents the component of output $y(t)$ that cannot be predicted from past data.

From this we can establish that through representation of the system transfer function, $G(q)$, and disturbance transfer function, $H(q)$, we have a working model of the system. However, representing these transfer functions by enumerating the infinite sequences contained in equation (2.10) would render the task of prediction through equation (2.12) very impractical. Instead, one chooses to work with structures that permit the specification of $G(q)$ and $H(q)$ in terms of a finite set of numerical values.

This representation of a system through a finite number of numerical values, or coefficients, has a very important consequence. Quite often it is not possible to determine these coefficients a priori from knowledge or some physical interpretation of the mechanisms that govern the dynamics of a system (Ljung, 1987: 70). This means that the determination of these values enter the system description as parameters to be determined by

some estimation procedure. These parameters are denoted by the vector θ, and thus we have a system description given as

$$\begin{gathered} y(t) = G(q,\theta)u(t) + v(t), \qquad v(t) = H(q,\theta)e(t) \\ f_e(x,\theta), \text{ the PDF of } e \end{gathered} \tag{2.16}$$

with f_e being the probability density function (PDF) of e which is required for the specification of the modelled disturbance. The parameter vector θ ranges over a subset of $\mathbf{R}^d$, where d is the dimension of θ (Ljung, 1987: 70):

$$\theta \in D_{\mathcal{M}} \subset \mathbf{R}^d$$

At this point it is important to note that equation (2.16) is no longer a model but a set of black-box models defined through the range of θ. It is through the estimation procedure that the vector θ is obtained and the model defined from the set. Using equation (2.12) we can compute the one-step-ahead prediction for equation (2.16).

2.2.4 A Class of Model Structures

Linear Regression

Ljung (1987: 71) states that the most immediate way of parametrizing $G(q)$ and $H(q)$ is to represent them as rational functions with parameters entering as the coefficients of the numerator and denominator. This parameterized representation of the relationship between the input and output may be described as a linear difference equation (Ljung, 1987: 71):

$$\begin{aligned} &y(t) + a_1 y(t-1) + \ldots + a_{n_a} y(t-n_a) \\ &= b_1 u(t-1) + \ldots + b_{n_b} u(t-n_b) + e(t) \end{aligned} \tag{2.17}$$

where the choice of coefficients a and b and model order, n_a and n_b, define the system representation. The adjustable parameters may be represented as

$$\theta = [a_1 \ a_2 \ \ldots \ a_{n_a} \ b_1 \ \ldots \ b_{n_b}]^T \tag{2.18}$$

Such a linear difference equation is written in the transfer function form expressed in equation (2.16) through the use of the time shift operators as follows:

$$\begin{aligned} A(q) &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}, \\ B(q) &= b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \end{aligned} \tag{2.19}$$

It must be noted that in some cases the dynamics from $u(t)$ to $y(t)$ contains a delay of

n_k samples. This means that some of the leading coefficients of $B(q)$ would be zero, thus according to Ljung (1987):

$$B(q) = b_{n_k} q^{-n_k} + b_{n_k+1} q^{-n_k-1} + \ldots + b_{n_{b_k}+n_b-1} q^{-n_k-n_b+1} \tag{2.20}$$

Applying this transfer function expression described in equations (2.17) and (2.19) to equation (2.16) allows $G(q,\theta)$ and $H(q,\theta)$ to be defined as

$$G(q,\theta) = \frac{B(q)}{A(q)}, \qquad H(q,\theta) = \frac{1}{A(q)} \tag{2.21}$$

This model is called an Auto-Regressive with External Input (ARX) model. If the case were such that n_a in equation (2.19) was zero, the model would be come a Finite Impulse Response (FIR) model.

The prediction function given in equation (2.12) can now be given as follows:

$$\hat{y}(t|\theta) = B(q)u(t) + [1 - A(q)]\, y(t) \tag{2.22}$$

This may be re-written as a linear regression (Ljung, 1987: 72),

$$\hat{y}(t|\theta) = \theta^T \varphi(t) = \varphi(t)^T \theta \tag{2.23}$$

with

$$\varphi(t) = [-y(t-1) \ldots - y(t-n_a)\ u(t-1) \ldots u(t-n_b)]^T \tag{2.24}$$

This shows the predictor to be the scalar product between a known data vector $\varphi(t)$, also known as the regression vector, and the parameter vector θ.

A General Family of Model Structures

Consider the following generalised model structure (Forssell, 1999):

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t), \tag{2.25}$$

where $C(q)$ and $D(q)$ are polynomials similar to $A(q)$ and $B(q)$ defined by equation 2.19. The structure is also depicted in figure 2.4. The advantage of this type of system description is that there is more freedom in describing the properties of the disturbance term. The introduction of polynomial $C(q)$ does precisely this as it introduces flexibility in describing the equation error as a moving average (MA) of white noise.

Table 2.1 shows the different structures and their names based on how one defines the model by use of the different combinations of polynomials. Of particular value is the ARMAX structure. This structure has become a standard tool in control and econometrics

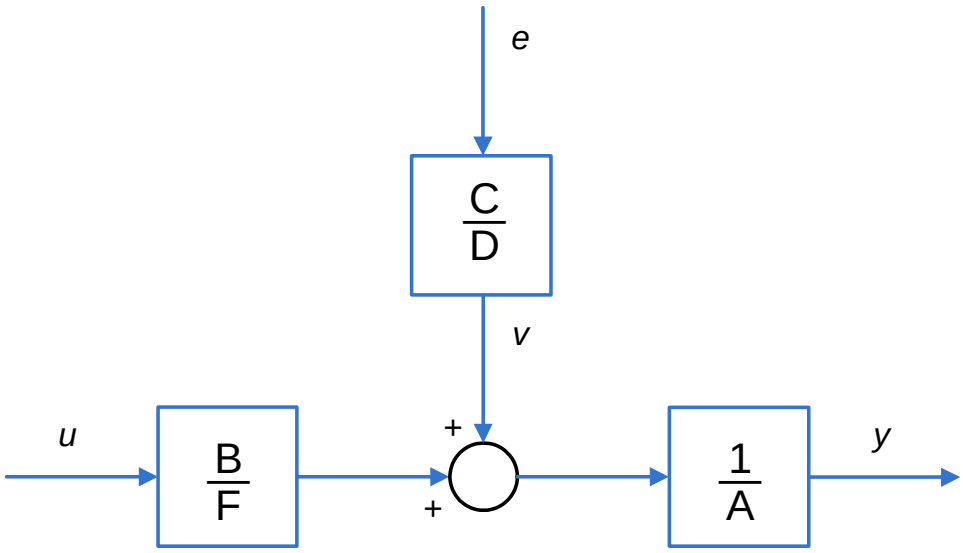


Figure 2.4: General Structure Transfer Function Representation

Polynomials Used in eq (2.25)	Name of Model Structure
B	FIR (finite impulse response)
AB	ARX
ABC	ARMAX
AC	ARMA
ABD	ARARX
ABCD	ARARMAX
BF	OE (output error)
BFCD	BJ (box-jenkins)

Table 2.1: Common Model Structures
(Ljung, 1987: 78)

for both system description and control design. A version with an enforced integration in the system description is the ARIMAX model which has become useful to describe systems with slow disturbances (Box & Jenkins, 1970).

It is noted that for model structures other than ARX and FIR, the regression is not strictly linear. For example, given an ARMAX structure, the regression vector becomes:

$$\varphi(t,\theta) = [-y(t-1)\ldots -y(t-n_a)\ u(t-1)\ldots u(t-n_b)\ \epsilon(t-1,\theta)\ldots\epsilon(t-n_c,\theta)]^T \quad (2.26)$$

with the resulting predictor being

$$\hat{y}(t|\theta) = \varphi^T(t,\theta)\theta \quad (2.27)$$

which is a pseudolinear regression, due to the non-linear effect of θ on the vector $\varphi(t,\theta)$ (Ljung, 1987: 74). Additionally, literature occasionally refers to the FIR structures as non-parametric due to fact that FIR structures require much more parameters to describe

the same dynamic behaviour .

2.3 Non-Linear Time-Varying System Identification

While non-linear system identification methods are not the primary focus of this work some key approaches and concepts are worth discussing as they provide insight into the non-linear framework which is relevant.

2.3.1 Non-Linearity and System Memory

Non-Linearity

Norton (1986: 246) explains the case of non-linearity very appropriately in stating that in contrast to the tight constraints on system behaviour imposed by linearity and time-invariance, non-linear behaviour is vastly diverse and complex. We can thus expect non-linear system identification to be more difficult

A system L is said to be linear if the following properties hold true for the system:

$$\text{Additive Property} \qquad L[u_1 + u_2] = L[u_1] + L[u_2] \tag{2.28}$$

$$\text{Homogenous Property} \qquad L[cu] = cL[u] \tag{2.29}$$

If a system or model does not satisfy the properties given by equations (2.28) and (2.29) it is non-linear (Bendat, 1991: 2).

In the case of a linear system, the system output, or response, to a random Gaussian input sequence will have a Gaussian probability density function, while non-linear systems will tend to have non-Gaussian outputs in response to Gaussian inputs (Bendat, 1991: 3).

It is understood that most of the dynamics that models try to approximate are non-linear, that is, most of "real-life" is non-linear.

System Memory

At this point it is appropriate to define system memory. System memory is a characteristic that is used in defining some of the common approaches to non-linear system identification. The memory of a system refers to the effect that past inputs have on the present outputs of a system and may be classified as follows (Bendat, 1991: 3):

- Infinite Memory,
- Finite Memory,
- Zero Memory.

A system with infinite memory has an output that is influenced by all past inputs. For most systems the inputs exerted on the system have negligible effects after a certain time, such a system has a fading memory or a finite memory. The system defined as having zero memory is a system that is not affected by the past inputs. The output of such a system may be predicted from the current input alone.

2.3.2 Identification Approaches and Methods for Non-linear Systems

While research in the field of non-linear system identification has yielded many useful methods and approaches, a few of the more common and established are now briefly introduced.

Linear Approximation

It is common practice to approximate non-linear systems via linear models as they are typically simpler and are usually accurate enough for the intended application (Enqvist et al., 2007). In addition to this, or in assertion of this, it is noted that while most industrial processes are non-linear, some processes are sufficiently linear around a given operating point. It is through this approach that the fields of non-linear dynamics detection and validation of linear approximations has established themselves. The former being a tool used to ascertain the extent of non-linearities that govern the system allowing one to determine whether or not a linear approximation would suffice. The latter being a tool that establishes whether a linear approximation is a valid enough approximation of the non-linear system. These fields are further discussed in section 4.2 and are important topics in this work.

Volterra-Series Models

As introduced in section 2.2.1, the response of a system may be represented through the superposition of a recorded response to a known input which may be given in the form of an input-output convolution. The non-linear generalisation of this convolution equation is the Volterra series (Norton, 1986: 247),

$$\begin{aligned} y(t) &= \int_0^\infty g_1(\tau)u(t-\tau)d\tau + \int_0^\infty \int_0^\infty g_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2)d\tau_1 d\tau_2 + \dots \\ &\quad + \int_0^\infty \dots \int_0^\infty g_r(\tau_1,\tau_2\dots,\tau_r)\prod_{i=1}^{r} u(t-\tau_i)d\tau_i + \dots\infty \end{aligned} \tag{2.30}$$

The Volterra series is generally seen as a higher order extension of the linear impulse response model where the g_r coefficients are called Volterra kernels. While its use and

application in non-linear identification is extensive, it does have drawbacks. The primary drawback of this approach is its impracticality and computational intensity of an nth order constrained Volterra series when $n > 2$. Thus in most applications reduced order Volterra series are used with some of the coefficients constrained to zero (Norton, 1986: 247).

Norton (1986) mentions that there are complications in calculating the Kernel coefficients. There is, however, extensive literature on the different estimation methods that could be used to obtain such coefficients. Schetzen (1980) surveys some of these methods with a focus on correlation approaches.

Block-Oriented Models

Block-oriented models are represented by a combination of static non-linearities and linear dynamic components. A non-linearity is defined as static when it has zero memory dynamics, that is, it does not depend on past input values (Ljung, 1987: 132).

This means that a system that exhibits memoryless non-linearities may be represented as cascades of linear dynamic and non-linear instantaneous sub-systems. Figure 2.5 illustrates two possible block-oriented models of this sort. Literature shows that physical insight into the system does help in identifying the different blocks but there are alternative methods that explore the property of separability via cross-correlation functions that allow one to determine $f(u(t))$ (Norton, 1986: 248).

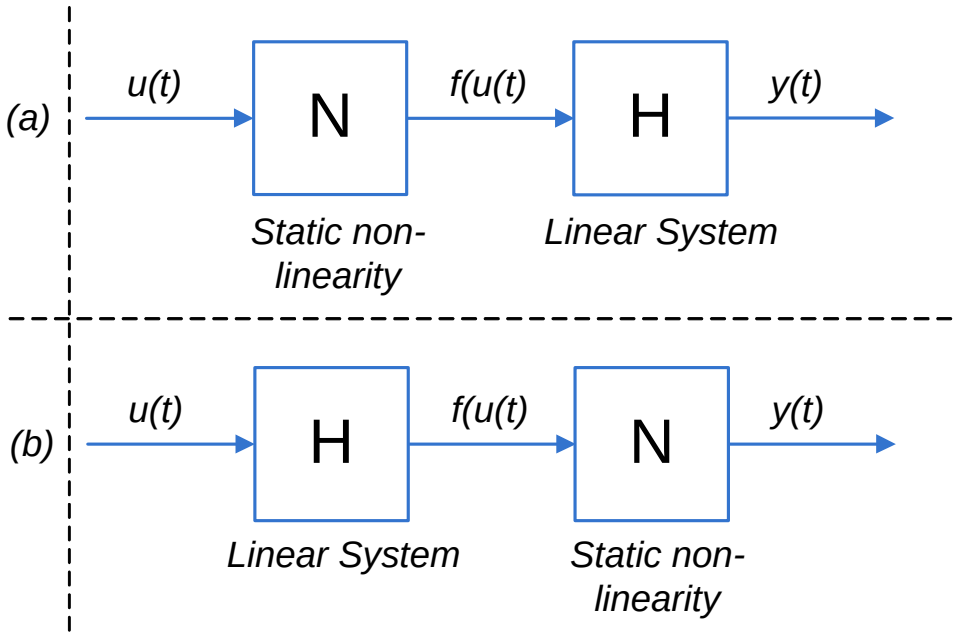


Figure 2.5: (a) Hammerstein Model (b) Wiener Model

Two common or well known extensions of the block-oriented models are the Hammerstein and Weiner models with the former illustrated above the latter in figure 2.5 .

The figure shows how each is a combination of a zero memory non-linear block, N, and a linear dynamic element with a steady state gain of 1, represented by H. While the two different models contain the same blocks the difference in order has the significance in that their steady state behaviour will be the same but their dynamic behaviour is different (Pearson & Pottmann, 2000).

Regression Models

Ljung (1987) states how, if one were to recall the linear regression model structure defined by equation (2.23), the regressor vector, $\varphi(t)$, is a vector of the lagged input and output values. How this regressor vector forms or is defined is however immaterial, that is, $\varphi(t)$ need not be consecutive lagged outputs and inputs, eg.

$$\varphi(t) = [u(t-1), u(t-5)]$$

All that matters is that it is a known quantity at time t. Thus one could write

$$\hat{y}(t|\theta) = \theta_1\varphi_1(u^t, y^{t-1}) + \ldots + \theta_d\varphi_d(u^t, y^{t-1}) = \varphi^T(t)\theta \tag{2.31}$$

with arbitrary functions, φ_i, of past data. The structure of equation (2.31) could be regarded as a finite-dimensional parametrisation of a general, unknown, non-linear model (Ljung, 1987: 130). The key is to determine the functions $\varphi_i(u^t, y^{t-1})$ and this usually requires physical insight into the system, however, determination in terms of input-output data only is also possible (Sjöberg et al.).

This approach extends itself to very large sets of possible regression options which may ultimately be described through

$$\hat{y}(t|\theta) = g(\varphi(t), \theta) \tag{2.32}$$

where g is a non-linear function that serves to map the regressor space to the output space. This translates the regression problem into two partial problems for dynamical systems (Sjöberg et al.):

1. How to choose the regression vector $\varphi(t)$ from past inputs and outputs
2. How to choose the non-linear mapping function $g(\varphi)$.

While there are several possible ways to define the regression vector, those defined in section 2.2.4, summarised in table 2.1, are commonly used. In the case of the ARX structure, under the non-linear form given by equation (2.32), it becomes a non-linear auto-regression with external inputs (NARX).

There are other choices for regressors, typically based on physical insight of the system (van den Hof & Boker, 1994).

Defining the non-linear mapping, $g(\varphi, \theta)$, is typically approached by defining the mapping as parameterised function expansions (Sjöberg et al.):

$$g(\varphi, \theta) = \sum \alpha_k \kappa(\beta_k(\varphi - \gamma_k)) \tag{2.33}$$

where κ is a non-linear function, α_k, β_k and γ_k are parameters of the non-linear estimator.

The non-linear function κ is some times termed the *mother basis function*. There are various forms and structures of this function that allow for certain approaches to solving the parameters defining the function. Sjöberg et al. details some of these structures of which the more common ones are:

- Wavelet Networks
- Kernel Estimators
- Sigmoid Neural Networks

2.4 Non-Parametric - Frequency Domain Methods

This section will explore some of the different identification techniques available that do not explicitly employ a finite-dimensional parameter vector in the search of a best description of the system.

2.4.1 Co-Correlation, Cross-Correlation and Spectral Density

Definitions

Before establishing some of the frequency domain based identification approaches found in literature it is essential to define some of the fundamental signal analysis tools. Frequency analysis is all about the analysis of signals in terms of their frequency content and/or contribution. A signal frequency analysis tool that is widely used throughout identification theory is the power spectrum - or spectral density. In order to define the spectral density of a signal, or cross-spectral density between two signals, it is necessary to define the auto-correlation function (ACF) - also termed the covariance - of a signal, S, and cross-correlation function (CCF) - also known as the cross-covariance - between signals S and W. These are respectively given as (Ljung, 1987: 27-28):

$$R_S(\tau) = \bar{E}[s(t)s(t-\tau)] \tag{2.34}$$

$$R_{SW}(\tau) = \bar{E}[s(t)s(t-\tau)] \tag{2.35}$$

with

$$\bar{E}[f(t)] = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[f(t)]$$

The notation $E[f(t)]$ is the mathematical expectation of $f(t)$ and may be interpreted as the mean of a signal if the signal were a stationary stochastic signal. That is, $E[f(t)]$ is analogous to the mean in that if $f(t)$ were a deterministic signal the mean, $\bar{f}$, would be determined by the limit of the normalised sum,

$$\bar{f} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} f(t)$$

Broch (1990) explains how the correlation function and power spectrum reveal certain characteristic properties of the signals or processes they are applied to, but goes on to mention how such characteristics are not easily interpreted. Nevertheless, Broch (1990: 33) explains the auto-correlation function of signal S, R_S, as the description of how a particular instantaneous amplitude value of the signal S, at time t, depends upon previously occurring instantaneous amplitude values at time $t - \tau$. While the cross-correlation function may be interpreted as the dependence between the magnitude value of a signal S at an arbitrary instant of time, t, and the magnitude value of the observed signal, W, at time $t - \tau$. Equations (2.34) and (2.35) reflect this mathematically.

With the correlation functions now defined this allows for the definition of the spectral density, or power spectrum, of a signal S as the Fourier transform of the auto-correlation function, R_S, given in discretised form by

$$\Phi_S(\omega) = \sum_{\tau=-\infty}^{\infty} R_S(\tau) e^{-i\tau\omega} \tag{2.36}$$

while the cross-power spectrum, or cross-spectral density, between signals S and W is similarly defined as the Fourier transform of the cross-correlation function (CCF), R_{SW}, and given by

$$\Phi_{SW}(\omega) = \sum_{\tau=-\infty}^{\infty} R_{SW}(\tau) e^{-i\tau\omega} \tag{2.37}$$

Note that by definition of the inverse Fourier transform we have

$$\bar{E}s^2 = R_S(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_S(\omega) d\omega \tag{2.38}$$

The relationship between the auto-correlation and spectral density is known as the Wiener-Khintchine relation which is a special case of Parseval's theorem (Rayner, 1971).

It must be noted at this point that the correlation function and power spectrum may be used to qualitatively identify characteristics within signals and determine relationships

amongst signals. Figures 2.6 and 2.7 illustrate this well. Each illustrates the auto-correlation (b), power spectrum (c), of the the corresponding signal (a).

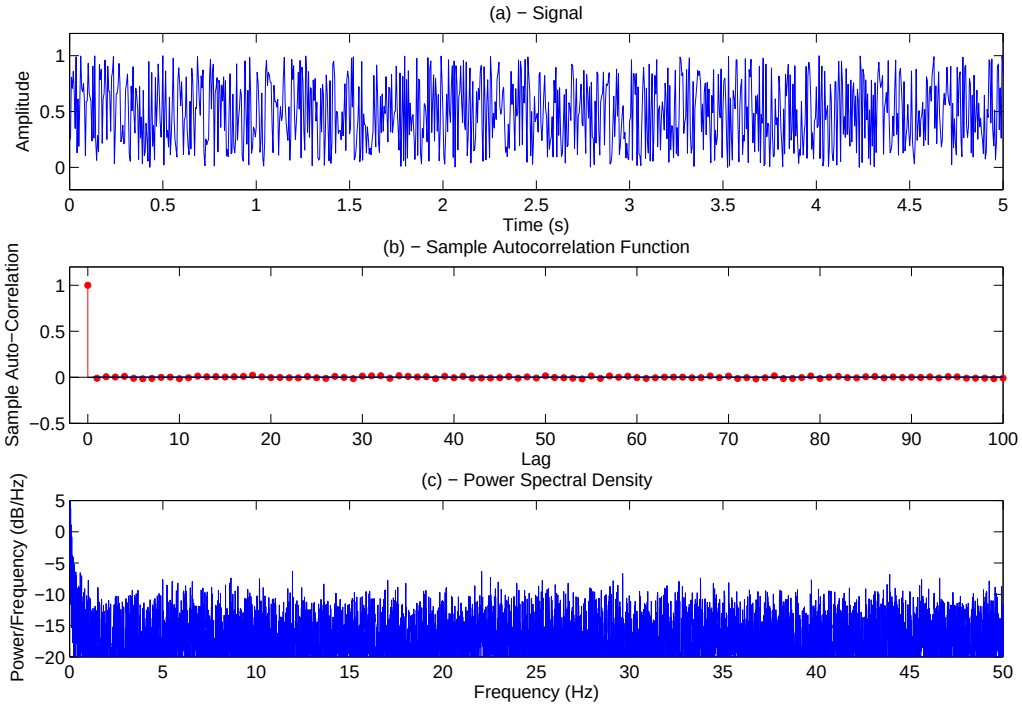


Figure 2.6: Auto-correlation Function and Power Spectrum of a white noise signal

The signal in figure 2.6(a) is a signal of white noise. Broch (1990) states that if one were to consider an "ideal" stationary random signal the auto-correlation function would consist of an infinitely narrow impulse at $\tau = 0$ as all instantaneous signal magnitudes are independent of each other. This is reflected in figure 2.6(b) for the noisy signal. In terms of the power spectrum of white noise, literature points out that the profile should be flat as there is no signal frequency that contributes substantially more than the another, figure 2.6(c) reflects this.

In contrast, the signal assessed in figure 2.7 and given in 2.7(a) is composed of two sinusoidal functions superimposed on top of each other. The auto-correlation function is effective in recognising the governing relationship within the signal while the power spectrum shows two peaks at frequencies corresponding to each sinusoidal function. The peak magnitudes are proportional to the corresponding sine functions contribution to the signal in terms of functions amplitude magnitude.

These applications of frequency domain analysis tools will be further emphasised in chapter 4.1 when theory concerning model validation methods is introduced.

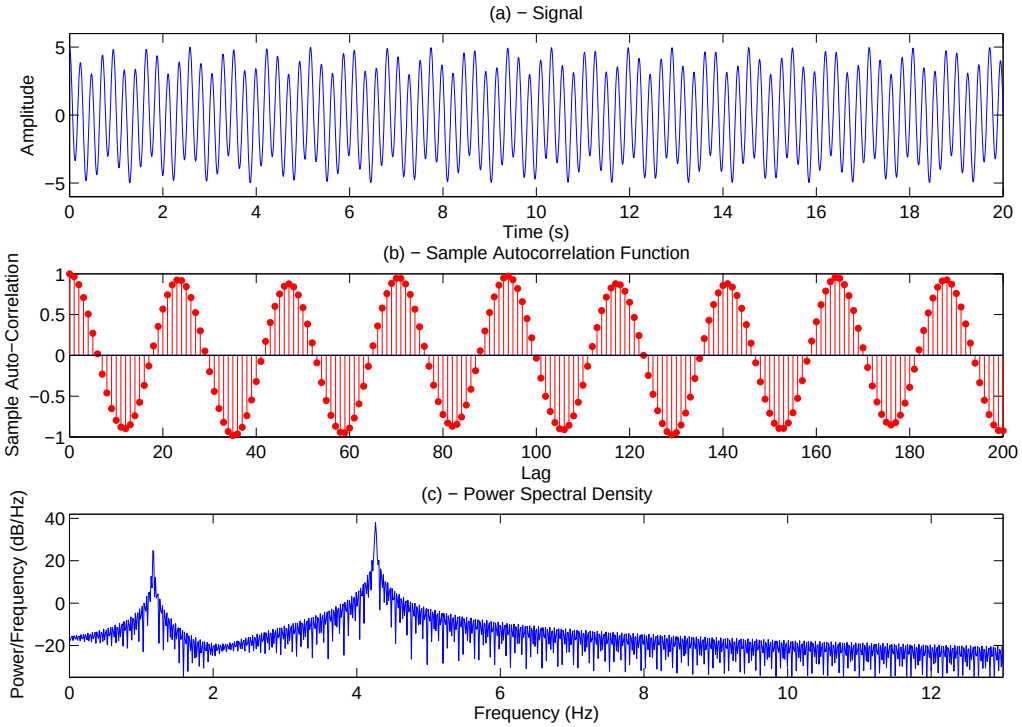


Figure 2.7: Auto-correlation Function and Power Spectrum of a sinusoidal signal

Spectral Density Transformations

Let signal S be defined by the following linear system:

$$s(t) = G(q)w(t)$$

Ljung (1987: 43) shows that

$$\begin{aligned} \Phi(\omega)_S &= |G(e^{j\omega})|^2 \Phi_w \\ \Phi(\omega)_{SW} &= G(e^{j\omega})\Phi_w \end{aligned}$$

The multivariable extension of this is given in section 2.7

2.4.2 The Empirical Transfer-function Estimate

In section 2.2.2 it was shown that through use of the Fourier transform one can transform a system description into the frequency domain. Extending this transformation, in discrete form, to output and input values of a system we have (Ljung, 1987: 146):

$$\hat{\hat{G}}_N(e^{j\omega}) \triangleq \frac{Y_N(\omega)}{U_N(\omega)} = \frac{\sum_{t=1}^{N} y(t)e^{-j\omega t}}{\sum_{t=1}^{N} u(t)e^{-j\omega t}} \tag{2.39}$$

This estimate is usually referred of as the Empirical Transfer-function Estimate (ETFE), and is considered the basic building block for most of the existing non-parametric frequency function estimates (Ljung, 1987).

It is well understood that the ETFE is a very crude estimate of the true transfer frequency function, $G_0(e^{i\omega})$ (Stenman et al., 2000). This is mostly due to the fact that the observed output, $y(t)$, can be corrupted by measurement noise, $v(t)$, which propagates into the ETFE via the Fourier transform. However, it is noted that typically for a fixed input $u(t)$ and sufficiently large N, the ETFE allows for an unbiased estimate of the true transfer frequency function. This condition is clouded by the fact that the variance associated with the ETFE does not decay to zero as N increases. Instead it tends to the noise-to-input-signal ratio at the frequency ω.

As a result of these conditions, and the understanding that in some cases the estimate $\frac{Y_N(\omega)}{U_N(\omega)}$ is undefined for some frequency ω such that $U_N(\omega) = 0$, the ETFE is not very smooth. This is problematic as the true transfer frequency function, $G_0(e^{i\omega})$, is a smooth continuous function of ω. For this reason much of the literature presented on the ETFE focuses on the smoothing out of the ETFE.

Amongst the approaches established are simple but not so effective solutions like assuming that the values of the ETFE can be estimated via a local averaging procedure which assumes that neighbouring ETFE values are related. This is typically done with some weighting function (Stenman et al., 2000). More elaborate successful approaches to smoothing out the ETFE are available. Stenman et al. (2000) suggests an adaptive parametric approach while Ljung (1987: 153) discusses established Blackman-Tukey (Blackman & Tukey, 1958) methods that use spectral estimates.

2.5 Parametric Estimation Methods

2.5.1 Principles Behind Parameter Estimation

In section 2.2.4 a parameterised family of model structures was introduced where a parameter vector, θ, was used to define the model with in a model structure. This section introduces the different methods that literature presents to obtaining a parameter vector estimate $\hat{\theta}$ and the governing principles.

Given the information or data set,

$$Z^N = [y(1), u(1), y(2), u(2), \ldots, y(N), u(N)] \tag{2.40}$$

the objective, or problem to resolve, is how to use this information to select a vector $\hat{\theta}_N$, and hence the proper model member, $\mathcal{M}(\hat{\theta}_N)$, amongst the possible sets of models, $\mathcal{M}^*$. Strictly speaking one wants to determine a mapping from the data set, Z^N, to $D_{\mathcal{M}}$, the

set of $\hat{\theta}_N$ values in the model structure $\mathcal{M}$ (Ljung, 1987: 170):

$$Z_N \rightarrow \hat{\theta}_N \in D_{\mathcal{M}} \tag{2.41}$$

This forms the very essence of the identification problem in terms of parametric models and will be addressed next.

2.5.2 The Prediction Error Estimation Framework

In the prediction error framework one uses the difference between the predictor and the data set to optimise the estimate. Additionally there are parameters or functions in the form of norms and filters within the optimisation procedure that allow a "truer fit" in the face of unrepresentative data. This does essentially mean that these norms and filters are variable functions or scalars that could be optimised with respect to minimising the prediction error.

It is further noted that the prediction error framework is amongst the broader of parameter estimation families and can be applied to quite arbitrary model parameterisations (Ljung, 2002). For these reasons and more, most of the work presented here will be relative to the prediction error estimation method being the primary estimation candidate.

The Prediction Error

At this point it is convenient recall section 2.2.3 and to further introduce the prediction error between the measured output, $y(t)$, and the predicted output, $\hat{y}(t|\theta)$,

$$\epsilon(t, \theta) = y(t) - \hat{y}(t|\theta) \tag{2.42}$$

Together with the data set Z^N, containing the system's measured response $y(t)$ and inputs $u(t)$, the prediction errors can be computed for $t = 1, 2, 3, \ldots N$.

The prediction error sequence defined by equation (2.42) can be seen as a vector. The size of this vector may be measured using any norm, be it quadratic or non-quadratic. It is through the minimisation of the size of the prediction error sequence, i.e. the norm, that defines the procedure for determining the estimate for a vector θ that best represents the system at hand. Thus, let the prediction error be filtered through a stable linear filter $L(q)$ (Ljung, 1987: 171):

$$\epsilon_F(t, \theta) = L(q)\epsilon(t, \theta), \qquad 1 \leq t \leq N \tag{2.43}$$

applying a norm:

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \ell(\epsilon_F(t,\theta)) \tag{2.44}$$

where $\ell(\cdot)$ is the norm usually given by a positive scalar-valued norm function. The standard choice is the quadratic norm,

$$\ell(\epsilon(t,\theta)) = \frac{1}{2}\epsilon_F^2(t,\theta) \tag{2.45}$$

Recalling and rearranging equation (2.15), and implementing some linear, monic, parameterised prefilter, $L(q,\theta)$ we have

$$\epsilon(t,\theta) = L(q,\theta)H^{-1}(q,\theta)(y(t) - G(q,\theta)u(t)) \tag{2.46}$$

Thus the effect of the prefilter can be included in the noise model and $L(q,\theta) = 1$ can be assumed without loss of generality.

The estimate $\hat{\theta}_N$ is then defined by the minimisation of equation (2.44),

$$\hat{\theta}_N = \hat{\theta}_N(Z^N) = \arg\min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N) \tag{2.47}$$

The mapping given by (2.41) is thus consolidated by equations (2.43) through (2.47). The general term *prediction-error identification method* (PEM) is defined as the family of approaches that correspond to this set of equations (Ljung, 1987: 171).

Numerical Issues and the Least-Square Estimate

The actual calculation of the minimising argument presented by equation (2.47) can be complex with substantial computations, and possibly a complicated search over a function with several local minima. Literature shows the typical numerical search is carried out using the *damped Gauss-Newton* method (Ljung, 2002) apart from this other frequently used methods for solving this optimisation problem are documented in Evans & Fischl (1973) and Steiglitz & Mcbride (1965). There are however exceptions to the case.

Given the case where the model structure used to estimate $y(t)$ is a linear regression

$$\hat{y}(t|\theta) = \varphi(t)^T\theta$$

a unique feature is established via the choice of a quadratic norm, equation (2.45), in that a least-squares (LS) criterion is ensued (Forssell, 1999). This means that equation (2.47) can be minimised analytically, forgoing the necessity for a complex numerical search,

provided the inverse exists (Ljung, 1987: 176),

$$\hat{\theta}_N^{LS} = \arg\min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N) = \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) \tag{2.48}$$

This is the well known Least-Squares Estimate.

2.5.3 Other Common Estimation Frameworks

Instrumental Variable

Analogous to the least-squares estimate, Forssell (1999) states that through the standard instrumental variable method, the parameter vector estimate, $\hat{\theta}_N$, may be obtained through

$$\hat{\theta}_N^{IV} = \left[\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\zeta(t)y(t) \tag{2.49}$$

where $\zeta(t)$ is known as the instrumental variable vector. It must be noted that the least-squares method is obtained if $\zeta(t) = \varphi(t)$. The elements of vector $\zeta(t)$ are termed instruments and are usually obtained from past inputs by linear filtering and can be conceptually written as

$$\zeta(t) = \zeta(t, u^{t-1}) \tag{2.50}$$

For θ_N^{IV} to tend to θ_0 at large values of N, the instrumental vector and hence the elements that form the vector should allow for the following conditions (Forssell, 1999):

$$\bar{E}\zeta(t)\varphi^T(t) \textbf{ be nonsignular} \tag{2.51}$$

$$\bar{E}\zeta(t)e(t) = 0 \tag{2.52}$$

In loose terms, the instrument vector should be well correlated with lagged inputs and outputs but uncorrelated with the noise sequence. The natural choice for an instrument vector that satisfies these conditions would be a vector similar to the linear regression in equation (2.17) so as to secure condition (2.51) but different in that it is not influenced by $e(t)$ (Ljung, 1987: 193),

$$\zeta(t) = K(q)[-x(t-1) - x(t-2)\ldots - x(t-n_a)u(t-2)\ldots u(t-n_b)]^T \tag{2.53}$$

where K is a linear filter and x(t) is generated from the input through a linear system

$$N(q)x(t) = M(q)u(t)$$

with

$$N(q) = 1 + n_1 q^{-1} + \cdots + n_{n_n} q^{-n_n}$$

$$M(q) = m_0 + m_1 q^{-1} + \cdots + m_{n_m} q^{-n_m}$$

This further shows how $\zeta(t)$ can be generated from past inputs.

Maximum Likelihood

In contrast to the prediction error approach previously mentioned, the maximum likelihood method is to some extent very similar. The difference being that the ML (Maximum likelihood) method appeals to statistical arguments for the estimation of θ (Forssell, 1999). More specifically it uses the probability density functions of the observed variable y^N, given by $f_y(\theta; x^N)$ such that

$$P(y^N \in A) = \int_{x^N \in A} f_y(\theta; x^N) dx^N$$

The ML estimation function is such that probability, P, that the observed value, y^N, is equal to the estimated value, y_*^N is maximised. This is said to be proportional to $f_y(\theta; x^N)$ and thus the ML parameter estimation seeks

$$\hat{\theta}_{ML}(y_*^N) = \textbf{arg max}\ f_y(\theta; y_* N) \tag{2.54}$$

this function is called the *maximum likelihood estimator* (MLE) and is further discussed in Ljung (1987)

2.6 Convergence and Asymptotic Properties

It has been previously established that the central nature of the identification process is the mapping from the data set, Z^N, to the estimation vector, $\hat{\theta}_N$. The previous section introduced some common mapping or estimation approaches with the most pertinent method being that of the prediction error framework.

This section continues by characterising the convergence and asymptotic properties of the prediction error estimation method.

It is noted that the property of convergence bias is discussed in section 3.2 since this property is best discussed after introducing the closed-loop identification condition.

2.6.1 Convergence and Identifiability

Informative Data Set

Identifiability of model structures is concerned with whether different parameter vectors may describe the same model in the model set $\mathcal{M}^*$. This is directly related to whether the data set Z^N allows for the distinction between different models in the model set. If a data set does allow for this distinction then the data set is said to be *informative enough* (Ljung, 1987).

This concept of informative data is related to the concept of persistent signal excitation. Ljung (1987) details the definition of an input signal that is persistently excited. He states that such a signal allows for the identifiability of the model structure that uniquely describes the data set if, and only if, the input is persistently excited to a sufficiently high order. More of this concept is addressed in chapter 5.2.

A Frequency Domain Description of Convergence

Consider the prediction error method with a quadratic norm (Forssell, 1999),

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\epsilon^2(t,\theta) \tag{2.55}$$

under moderate conditions

$$V_N(\theta, Z^N) \to \bar{V}(\theta) = \bar{E}\frac{1}{2}\epsilon^2(t,\theta) \text{ w.p. 1 as } N \to \infty \tag{2.56}$$

and (Ljung, 1978)

$$\hat{\theta}_N \to D_c = \arg\min_{\theta \in D_{\mathcal{M}}} \bar{V}(\theta) \text{ w.p. 1 as } N \to \infty \tag{2.57}$$

with w.p. signifying *with probability*, D_c signifying the model set into which $\hat{\theta}_N$ converges and $\bar{V}_N(\theta)$ the convergence limit of $V_N(\theta, Z^N)$. This result is quite general and can be applied to other norms and non-linear time varying systems. It states that the estimate will converge to the best possible approximation of the system that is available in the model set (Ljung, 1987: 216).

Recalling Parseval's theorem, the relationship between auto-correlations and spectral density given by

$$\bar{E}s^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_S(\omega)d\omega$$

applying this to equation (2.56) we have

$$\bar{V}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\Phi_\epsilon(\omega)d\omega \tag{2.58}$$

where $\Phi_\epsilon(\omega)$ is the power spectrum of the prediction errors. Given the true system is described by,

$$y(t) = G_0(q)u(t) + v_0(t) \tag{2.59}$$

this is substituted into the prediction error equation previously given by equation (2.15) where a linear estimation is used,

$$\begin{aligned} \epsilon(t,\theta) &= H^{-1}(q,\theta)\left[y(t) - G(q,\theta)u(t)\right] \\ &= H^{-1}(q,\theta)\left[(G_0(q) - G(q,\theta))u(t) + v_0(t)\right] \end{aligned} \tag{2.60}$$

This allows for the definition of Φ_ϵ,

$$\Phi_\epsilon(\omega,\theta) = \frac{|G_0(e^{j\omega}) - G(e^{j\omega},\theta)|^2\Phi_u(\omega) + \Phi_v(\omega)}{|H(e^{j\omega},\theta)|^2} \tag{2.61}$$

which may now be used in the convergence limit definition,

$$\bar{V}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}|G_0(e^{j\omega}) - G(e^{j\omega},\theta)|^2\frac{\Phi_u(\omega)}{|H(e^{j\omega},\theta)|^2}d\omega + \int_{-\pi}^{\pi}\frac{\Phi_v(\omega)}{|H(e^{j\omega},\theta)|^2}d\omega \tag{2.62}$$

This definition of the convergence limit allows for the specification of the estimation vector in equation (2.57) (Ljung, 2002),

$$\begin{aligned} \hat{\theta}_N \to D_c &= \arg\min_{\theta\in D_\mathcal{M}}\frac{1}{4\pi}\int_{-\pi}^{\pi}|G_0(e^{j\omega}) - G(e^{j\omega},\theta)|^2\frac{\Phi_u(\omega)}{|H(e^{j\omega},\theta)|^2}d\omega \\ &\quad + \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\Phi_v(\omega)}{|H(e^{j\omega},\theta)|^2}d\omega \ \textbf{w.p. 1 as } N\to\infty \end{aligned} \tag{2.63}$$

$\frac{\Phi_u}{|H(e^{j\omega},\theta)|^2}$ is termed the quadratic frequency norm (Ljung, 1987: 227). It is interesting to consider this norm as a frequency weighting function in that:

- If $H(e^{j\omega},\theta)$ has primarily high frequency components, a larger weight is put on low frequency misfits between G_0 and $G(\theta)$.
- If $\Phi_u(\omega)$ has more high frequency components than low, a larger weight is put on high frequency misfits between G_0 and $G(\theta)$.
- The ratio between $\Phi_u(\omega)$ and $H(e^{j\omega},\theta)$ is known as the model signal-to-noise ratio.

It is through these points that it is understood that the extent of signal excitation through the input $u(t)$ and definition of the noise model play important roles in the estimate convergence. This being said, equation (2.63) shows that if the true system, $\mathcal{S}$, is represented in the model set, $\mathcal{M}^*$, the estimate will converge towards the true system.

It must be noted that this stated property of convergence for a Least-Squares Prediction Error Method (PEM) approach is bound to the condition that N tends to ∞. This

is obviously strictly theoretical as all applications are bound to finite limitations in N Ljung (1976).

2.6.2 Asymptotic Distribution

A General Expression for Asymptotic Variance

With convergence covered in the previous section it is now appropriate to present the asymptotic properties of the estimation approach specific to the PEM. Consideration of the asymptotic distribution deals with the rate at which the convergence limit is approached (Ljung, 2002).

Let it be assumed that the true model of the system is defined within the model set. The basic outline of this property is such that if the prediction error, $\epsilon(t,\theta)$, is Gaussian with variance λ_0, then the distribution of the random vector $\sqrt{N}(\hat{\theta}_N - \theta^*)$ converges to a normal distribution with zero mean. Ljung (1987: 242) states that the covariance matrix of $\hat{\theta_N}$, **Cov** $\hat{\theta}_N$, is related to the asymptotic covariance matrix P_θ of θ through

$$\textbf{Cov}\ \hat{\theta}_N \sim \frac{1}{N} P_\theta \tag{2.64}$$

where P_θ is directly given by

$$\begin{aligned} P_\theta &= \lambda_0 R^{-1} \\ R &= \left[\bar{E}\psi(t,\theta^*)\psi^T(t,\theta^*)\right] \end{aligned} \tag{2.65}$$

and

$$\lambda_0 = \bar{E}\epsilon^2(t,\theta^*) \tag{2.66}$$

$$\psi(t) = -\frac{d}{d\theta}\epsilon(t|\theta)|_{\theta=\theta^*} = \frac{d}{d\theta}\hat{y}(t|\theta)|_{\theta=\theta^*} \tag{2.67}$$

Ljung (1987: 243) explains that equation (2.65) has a natural interpretation. As equation (2.67) shows, ψ, is the gradient of $\hat{y}$ with respect the the estimation parameter, that is, ψ shows how much the estimation error, ϵ, changes as the parameter vector estimation, $\hat{\theta}$, changes. This is also termed the sensitivity derivative. Additionally it is noted that $\bar{E}\psi(t)\psi^T(t)$ is by definition the covariance of ψ. Therefore equation (2.65) naturally defines the asymptotic covariance of the estimate as being proportional to the inverse of the covariance matrix of the sensitivity derivative.

Ljung (1995) further elaborates that the variability of the model expressed through the covariance matrix is an indication of the model uncertainty. It is through model uncertainty that an understanding of how model parameters will differ if the estimation procedure is repeated if another data set, Z^N (with the same input sequences as the original data set), was used.

Asymptotic Black-box Variance

While the previous section shows an analysis of the covariance of the parameter vector, it is sometimes useful to study the covariance of the transfer function itself.

Ljung (1985) shows that when given a black box parameterisation of linear models, the transfer function estimates, $G(e^{j\omega}, \hat{\theta}_N)$ and $H(e^{j\omega}, \hat{\theta}_N)$, will, as N and model order n (i.e. the dimensionality of θ_N) tend to infinity, have a normal distribution with covariance matrix

$$\mathbf{Cov} \begin{bmatrix} G(e^{j\omega}, \hat{\theta}_N) \\ H(e^{j\omega}, \hat{\theta}_N) \end{bmatrix} \approx \frac{n}{N} \Phi_v(\omega) \begin{bmatrix} \Phi_u & \Phi_{ue} \\ \Phi_{eu} & \lambda_0 \end{bmatrix}^{-1} \tag{2.68}$$

Here Φ_{ue} is the cross power spectrum between input u and noise source e. This expression of the covariance matrix of the transfer function estimates is a general expression applicable to both open and closed loop conditions.

Further discussion on this will be done in section 3.2.1

2.7 Multivariable System Identification

So far the theory presented has been of systems with a scalar input and a scalar output. This section deals with the extension or necessary adaptations of such theory to systems with p outputs and m inputs. Such a system is termed multivariable and is depicted in figure 2.8. Ljung (1987: 35) states that the work required in this adaptation may be broken into two parts:

1. **The easy part:** Mostly notation changes, keeping track of transposes and that certain scalars become matrices and might not commute.

2. **The difficult part:** Multivariable models have richer internal structures, this has consequences that result in non-trivial parameterisations and estimations.

Figure 2.8: Possible Selection of Input Disturbances

Note that multiple input and single output (MISO) systems do not have the problems stated in part 2 (Ljung, 1987: 35).

2.7.1 Notation and System Description

Representing a multivariable system is done by collecting the p output components and m input components into respective p- and m- dimensioned column vectors, y(t) and u(t) respectively. Let the disturbance $e(t)$ also be represented as a p-dimensional column vector and we have,

$$y(t) = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} \qquad u(t) = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \qquad e(t) = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix}$$

This allows for the basic multivariable system to be represented in a similar manner as the scalar system description (Ljung, 1987: 37):

$$y(t) = G(q)u(t) + H(q)e(t)$$

In this case $G(q)$ is a transfer function matrix of dimension $p \times m$ and $H(q)$ has the dimension $p \times p$. The sequence $e(t)$ is a sequence of p-dimensional vectors with zero mean values and covariance matrix $Ee(t)e^T(t)$.

This translates into a system description that is formulated by subsystem transfer functions,

$$G(q) = \begin{bmatrix} G_{11}(q) & G_{12}(q) & \cdots & G_{1m}(q) \\ G_{21}(q) & G_{22}(q) & \cdots & G_{2m}(q) \\ \vdots & \vdots & \ddots & \vdots \\ G_{p1}(q) & G_{p2}(q) & \cdots & G_{pm}(q) \end{bmatrix} \tag{2.69}$$

Where each subsystem in G(q) is denoted with the indices, i and j, allowing for the scalar transfer function from the input number j to the output number i to be denoted as $G_{ij}(q)$.

The adaptation of the system description for a multivariable system extends to definitions of multivariable correlations and power spectrum transformation. That is

$$\begin{aligned} R_S(\tau) &= \bar{E}[s(t)s^T(t-\tau)] \\ R_{SW}(\tau) &= \bar{E}[s(t)w^T(t-\tau)] \end{aligned}$$

with R_S and R_{SW} now being matrices and $s(t)$ and $w(t)$ being column vectors.

The definitions of the power spectra remain unchanged while the manner in which they are transformed given a multivariable system

$$s(t) = G(q)w(t)$$

becomes

$$\Phi(\omega)_S = G(e^{j\omega})\Phi_w G^T(-e^{j\omega})$$

2.7.2 Parameterisation and Estimation Methods

Parametrisation

Representing a multivariable system as a parametric model structure is done in a similar way to the mostly straightforward scalar counterparts. In terms of a linear regression ARX structure, equations (2.17) through (2.19), the regression is no longer scalar polynomials but matrix polynomials (Ljung, 1987: 80),

$$\begin{aligned} &y(t) + A_1 y(t-1) + \ldots + A_{n_a} y(t-n_a) \\ &= B_1 u(t-1) + \ldots + B_{n_b} u(t-n_b) + e(t) \end{aligned} \tag{2.70}$$

using the time shift operator

$$\begin{aligned} A(q) &= I + A_1 q^{-1} + \ldots + A_{n_a} q^{-n_a}, \text{ and} \\ B(q) &= B_1 q^{-1} + \ldots + B_{n_b} q^{-n_b} \end{aligned} \tag{2.71}$$

where A_i are $p \times p$ matrices and B_i are $p \times m$ matrices.

This allows for multivariable parametric rational transfer functions of q,

$$G(q,\theta) = A^{-1}(q)B(q), \qquad H(q,\theta) = A^{-1} \tag{2.72}$$

The derivation of the parameter vector, θ, and representation of the multivariable linear regression equation is a fairly subtle issue and is slightly more complex than what these given representations initially connote. This is primarily as the multivariable condition results in a higher degree or order specification. That is, the choice of structure used for representation requires the specification of the matrix fraction description.

Specifically, this entails determining how each subsystem's regression order is defined. If one was to consider the coefficient matrix $A(q)$, given in equation (2.71), and expand for a system with p outputs and m inputs the following is obtained (Ljung, 1987: 116):

$$A(q) = \begin{bmatrix} A_{11}(q) & A_{12}(q) & \ldots & A_{1p}(q) \\ A_{21}(q) & A_{22}(q) & \ldots & A_{2p}(q) \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1}(q) & A_{p2}(q) & \ldots & A_{pp}(q) \end{bmatrix} = A^{(0)} + A^{(1)}z^{-1} + \ldots + A^{(na)}z^{-na} \tag{2.73}$$

with subsystem entries as follows:

$$A_{ij}(q) = a_{ij}^{(0)} + a_{ij}^{(1)}z^{-1} + a_{ij}^{(2)}z^{-2} + \ldots + a_{ij}^{(na_{ij})}z^{-na_{ij}} \tag{2.74}$$

This means that the order of the multivariable ARX system representation is done by specifying the orders na_{ij} and nb_{ij} (i.e. $p(p+m)$ integers). This results in a staggering amount for possible model structures that allow for different parameterisations.

Certain parameterisations allow for completely independent subsystems which means that each subsystem is parameterised as if it were a SISO system. Other parameterisations result in MIMO system representations that are compiled by partial models that are constructed from all the inputs and one output, i.e. multiple MISO models. Zhu & Butoyi (2002) shows that these types of parameterisation do typically result in bias models as they do not completely incorporate variable interactions especially in ill-conditioned systems. The most suitable MIMO system representation results from parameterisations where all inputs and all outputs are parameterised and estimated simultaneously (Ljung (1987: 80-81, 115) and Johnson & Wichern (2007: 360)).

In any case, an immediate analogy of the scalar parameter vector is possible. Suppose all the matrix entries in equation (2.70) are included in θ, we may now define the $[n_a \cdot p + n_b \cdot m] \times p$ matrix as

$$\boldsymbol{\theta} = [A_1 A_2 \cdots A_{n_a} B_1 \cdots B_{n_b}]^T \tag{2.75}$$

and the $[n_a \cdot p + n_b \cdot m]$ dimension column vector

$$\varphi = \begin{bmatrix} -y(t-1) \\ \vdots \\ -y(t-n_a) \\ u(t-1) \\ \vdots \\ u(t-n_b) \end{bmatrix} \tag{2.76}$$

allowing for

$$y(t) = \boldsymbol{\theta}^T \varphi(t) + e(t) \tag{2.77}$$

It is noted that not all structure specifications allow for this representation to be used.

Parameter Estimation

The multivariable counterpart of the prediction error frame work is very similar to those represented by equations (2.43) through (2.47) (Ljung, 1987: 175). The quadratic norm and the prediction error method once again lend themselves to the often-used least-squares estimation method given by

$$\ell(\epsilon) = \frac{1}{2} \epsilon^T \Lambda^{-1} \epsilon \tag{2.78}$$

and

$$\epsilon_f(t,\theta) = L(q,\theta)\epsilon(t,\theta) \tag{2.79}$$

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \epsilon_F^T(t,\theta)\Lambda^{-1}\epsilon_F(t,\theta) \tag{2.80}$$

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N) \tag{2.81}$$

with Λ being a symmetric, positive definite $p \times p$ weighting matrix and L once again being a monic filter that can be incorporated into $H(q,\theta)$, i.e. $L = I$ (Johnson & Wichern, 2007: 364).

2.7.3 Convergence and Asymptotic Properties

Ljung (1987: 223) states that the convergence and asymptotic results for multivariable systems are entirely analogous to the scalar case with the only major difference being notation changes.

2.8 Model Validation

Model validation should form the final stage of any identification procedure (Billings & Voon, 1986). While the parameter estimation procedure picks the "best" model within the chosen structure, determination of the quality of the model is essential. The different quantitative measures of model quality and under which circumstances certain definitions are used, is the central aspect of model validation.

The literature presented in this work will focus on two primary aspects of model validation. The first being the various techniques and approaches and the second being the necessity for higher order validation techniques in the case of linear estimates of non-linear systems. The chapter of Model validation (Chapter 4.1) presents these topics in detail.

2.9 Chapter Summary

In overview of this chapter it must be noted that the focus began on general modelling and system representation approaches and gradually moved towards parametric model structures. Through the surveying of literature it was found that the simple structure of such models to lend themselves to facilitated estimation procedures. Parameter estimation convergence and asymptotic theory showed the importance of system excitation via system inputs and how the correct definition of noise models and pre-filters may serve as frequency weighting functions.

CHAPTER 3

Closed-Loop System Identification Theory

Chapter 2 presented the basis and fundamental theory of system identification. This chapter extends this theory to the condition where the data set, Z_N, used to generate the parameter vector, $\hat{\theta}_N$, is obtained under closed-loop conditions. The problems and issues concerned with closed-loop system identification are introduced along with the three primary approaches to successfully identifying a model from closed-loop data. Once again the convergence and asymptotic properties are presented relative to the prediction error framework along an assessment into the role controller complexity has in system identification.

3.1 Basic Concepts, Issues and Method Classifications

3.1.1 Closed-Loop System Description, Notation and Conventions

A closed-loop system, represented by figure 3.1, is a system that is under feedback control. In general, this work will assume that the controller is a linear controller with a mechanism that is not always known. Recalling the representation of the true system (denoted by the 0 subscript),

$$\begin{aligned} y(t) &= G_0(q)u(t) + v_0(t) \\ &= G_0(q)u(t) + H_0(q)e(t) \end{aligned}$$

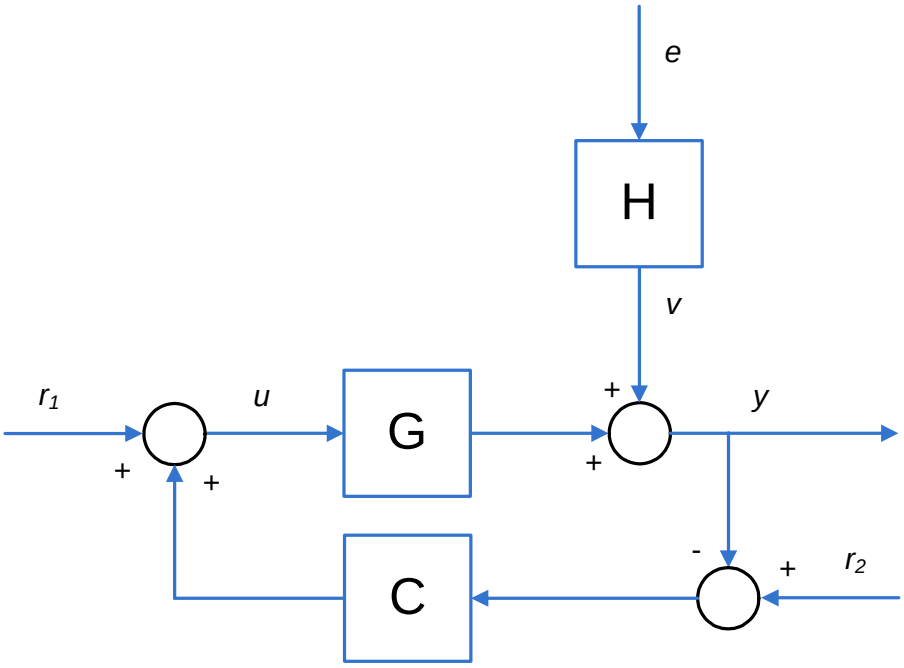


Figure 3.1: Closed-loop system representation

defining the input signal as

$$u(t) = r_1(t) + C(q)(r_2(t) - y(t)) \tag{3.1}$$

where $r_1(t)$ is regarded as the dither or excitation signal while $r_2(t)$ is the set-point (Forssell & Ljung, 1997). For purposes of this work it is sufficient to consider these as the reference signal defined by

$$r(t) = r_1(t) + C(q)r_2(t) \tag{3.2}$$

where $r(t)$ is assumed independent of the noise $e(t)$. The input signal may now be redefined by

$$u(t) = r(t) - C(q)y(t) \tag{3.3}$$

Allowing for the closed-loop system representation through

$$y(t) = G_0(q)S_0(q)r(t) + S_0(q)H_0(q)e(t) \tag{3.4}$$

with $S_0(q)$ termed the sensitivity function and defined by

$$S_0(q) = \frac{1}{1 + C(q)G_0(q)} \tag{3.5}$$

We may further define closed loop transfer functions as

$$\begin{aligned} G_0^{cl}(q) &= G_0(q)S_0(q) \\ H_0^{cl}(q) &= S_0(q)H_0(q) \end{aligned}$$

this allows equation (3.4) to be re-written,

$$\begin{aligned} y(t) &= G_0^{cl}(q)r(t) + v^{cl}(t) \\ &= G_0^{cl}(q)u(t) + H_0^{cl}(q)e(t) \end{aligned} \tag{3.6}$$

with the input further defined as

$$u(t) = S_0(q)r(t) - C(q)v^{cl}(t) \tag{3.7}$$

From the above equation it is noted that the input signal has two components, one sourced in the reference signal and the other in the system disturbance. It is thus convenient to establish the power spectrum for the input

$$\Phi_u(\omega) = |S_0(e^{j\omega})|^2\Phi_r(\omega) + |C(e^{j\omega})|^2|S_0(e^{j\omega})|^2\Phi_v(\omega) \tag{3.8}$$

where $\Phi_r(\omega)$ and $\Phi_v(\omega)$ are the power spectrums for the reference signal and the disturbance signal respectively. It follows that it is possible to denote

$$\Phi_u^r(\omega) = |S_0(e^{j\omega})|^2\Phi_r(\omega) \tag{3.9}$$

$$\Phi_u^e(\omega) = |C(e^{j\omega})|^2|S_0(e^{j\omega})|^2\Phi_v(\omega) \tag{3.10}$$

with $\Phi_u^r(\omega)$ and $\Phi_u^e(\omega)$ interpreted as the part of the input spectrum that originates from the reference signal r and noise e respectively.

3.1.2 Issues in Closed-Loop Identification

The Input-Noise Correlation

The primary problem or difficulty to overcome in closed-loop system identification is the correlation between unmeasurable noise and the input. Considering figure 3.1 it becomes clear that if the feedback controller, $C(q)$, is not zero, the input and the noise will be correlated. This correlation typically results in a biased estimate (Forssell, 1999).

This condition of a biased estimate due to input-noise correlation is revealed if we were to consider a true system representation given by,

$$y(t) = B(q)u(t) + v(t) = b_1u(t-1) + \ldots + b_n(t-n) + v(t)$$

which may be re-written as a linear regression,

$$y(t) = \varphi^T(t)\theta_0 + v(t)$$

with

$$\theta_0 = [b_1 \dots B_n]^T \quad and \quad \varphi(t) = [u(t-1) \dots u(t-n)]^T$$

From the above we note that the chosen model structure, $\mathcal{M}$, contains the true system, $\mathcal{S}$, that is, $\mathcal{S} \in \mathcal{M}$. Additionally θ_0 is the true parameter vector. Thus, given N data points, the least-squares estimate may be solved for analytically using equation (2.48) (Forssell, 1997),

$$\begin{aligned} \hat{\theta}_N^{LS} &= \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) \\ &= \theta_0 + \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) \end{aligned} \tag{3.11}$$

It is known that, under mild conditions, $\hat{\theta}_N \to \theta^*$ **w.p. 1** (Forssell, 1997), thus

$$\theta^* = \lim_{N\to\infty} E\hat{\theta}_N = \theta_0 + \left[\bar{E}\varphi(t)\varphi^T(t)\right]^{-1}\bar{E}\varphi(t)v(t) \tag{3.12}$$

From this it is quite clear that if there is an input-noise correlation, $\bar{E}\varphi(t)v(t) \neq 0$, θ^* will result in a biased estimate. In closed loop systems this is the case, and this typically why many open-loop identification methods fail when applied to closed-loop data (Forssell, 1997).

Signal Excitation and Information

An underlying assumption in system identification is that the data set used, Z^N, is sufficiently rich so as to allow conditions that make it possible to uniquely determine the system parameters. This is a much studied subject in closed-loop identification and open-loop identification (Anderson & Gevers, 1982)(Gustavsson et al., 1977).

In section 2.6.1 the concept of persistent signal excitation allowing for informative data was established for the open-loop case. For the closed-loop system the condition is not the same. Ljung (1987: 365) and Forssell (1999) show that the condition of informative data in a closed-loop system is not guaranteed by sufficient signal excitation. Gustavsson et al. (1977) goes on to say that this is natural as one of the purposes of the feedback controller is to minimise the output variance and hence the consequent minimisation of output information content.

The methods or aspects concerned with guarantying informative conditions, and hence identifiability, will be discussed in section 3.2 and again in section 5.2. This will be done

relative to the different methods commonly available for closed-loop system identification.

3.1.3 Problematic Identification Methods

Due to the issues mentioned in the previous section, specifically the problem on input-noise correlation, most of the well-known identification methods, which work well when used with open-loop data, fail when applied directly to input-output data obtained from a system operated under feedback (Forssell, 1999).

A few of the better known identification methods that typically fail are now mentioned together with a brief explanation as to why they fail. In some cases references to literature sources that investigate adaptations to the respective method allowing for identification under closed-loop conditions are made.

Instrumental Variable (IV) Methods

Under open-loop conditions the typical approach taken when using the IV method is to let the instruments defining the regression vector be filtered delayed versions of the input. This results in biased estimates due to the conditions, given by equations (2.51) and (2.52), required for consistent identification not being met as a result of the input-noise correlation (Forssell, 1999).

Söderström et al. (1987) indicates that through constructing instruments from the reference signal which are uncorrelated to noise or choosing delayed versions of the regressor with large delays, the conditions for unbiased closed-loop identification can be met. It must be said that it is not completely straightforward to apply the IV method to closed loop conditions.

Subspace Method

The problem with subspace method is very much similar to that of the IV method. Here the problem is finding a suitable multiplication matrix, which is usually obtained from delayed inputs, which is uncorrelated to noise allowing for the elimination of the noise term in the Hankel Matrix equation.

More detail on this problem is given in Forssell (1999), while alternative solutions or adaptations to the method that overcome the input-noise correlation problem are surveyed in Lin et al. (2005) and Ljung & McKelvey (1996).

Correlation and Spectral Analysis Methods

The identification principle behind these non-parametric methods is based on time domain correlations (covariance) or frequency contributions (power spectrum) between the input and the output data. These methods obviously encounter problems when presented with

closed-loop data due to the input-noise correlation. Forssell (1999) further details this problem while possible solutions have been proposed by Akaike (1967)

3.1.4 Classification of Closed-Loop Identification Approaches

The classification of approaches is based on the distinction of the following assumptions when dealing with the feedback (Forssell & Ljung, 1999):

1. Assume no knowledge of the nature of the feedback and use only measurements of the system input,
2. Assume the feedback to be known and use possible reference signals and set-points
3. Assume the regulator the be unknown but of a certain structure

In assumptions (a) and (b) it is common to assume that the controller is linear. However, if the controller were non-linear the same principles would apply with the price that the estimation procedure becomes more involved. Unless otherwise stated most of the theory presented here will assume a linear controller. It is however noted that most controllers in industry are non-linear. This is as no controller has a simple form, most have various delimiters, anti-windup functions and other components that would instill even in a strictly linear controller structure some type of non-linearity (Forssell & Ljung, 1999).

From these assumptions the closed-loop identification methods correspondingly fall into approach groups as follows (Forssell & Ljung, 1999):

1. **The Direct Approach :** Ignore the feedback and identify the system using only in input and output measurements
2. **The Indirect Approach :** Identify the closed-loop transfer function and determine the open-loop parameters using knowledge of the (linear) controller.
3. **The Joint Input-Output Approach :** Regard the input and output jointly as the output from the system driven by some reference or set-point signal and noise. Then use a method to determine open loop parameters from this augmented system.

3.2 Closed-Loop Identification in the Prediction Error Framework

As the previous section established, most non-parametric system identification methods do not directly work when using closed-loop data (Ljung, 1987). Literature also repeatedly indicates the prediction error method (PEM) as the best approach under closed-loop conditions (Forssell & Ljung, 1999).

Ljung (2002) states that this is mostly due to the PEM being a very robust method, applicable to broad data conditions and parameterisation structures. Huang & Shah (1997) additionally supports the PEM approach primarily based on its ability to achieve a minimum asymptotic variance near to that achieved under open-loop conditions.

These different approaches, and their respective properties, will be further detailed in the next sections. The properties concerning the approach of directly applying the PEM to closed-loop data, in accordance with literature, will firstly be established followed by those of the Indirect and Joint Input-Output methods. This detailing of the convergence and asymptotic properties of these closed-loop approaches complement theory presented for the general identification condition discussed in section 2.6. For a complete and more detailed elaboration of the convergence, asymptotic and covariance properties of the different approaches see Forssell & Ljung (1997) and Ljung (1985).

3.2.1 The Direct Method

As pointed out in the previous section, the direct approach ignores the feedback and directly uses the input and output measurements. This entails an identification procedure that coincides with the open loop approach. Thus, the identification methodology detailed in section 2.5.2 is directly applied.

Literature consistently indicates that this direct identification approach should be seen as the natural approach to closed-loop data analysis (Forssell, 1999) (Forssell & Ljung, 1997). The main reasons given for this are (Forssell & Ljung, 1997):

- It requires no knowledge of the regulator or character of the feedback and works regardless of the complexity of the regulator.
- No special algorithms or software are required.
- Consistency and optimal accuracy is obtained if the model structure contains the true system (including the noise properties).

There are two primary drawbacks with this direct approach (Forssell & Ljung, 1997):

- Unlike open-loop identification, an accurate noise model, $\hat{H}(q, \theta)$, is required.
- The second problem is a consequence of the first. When a simple model is sought that should approximate the system dynamics, the noise model deviation from true noise characteristics will introduce bias.

Convergence and Identifiability

Let us recall equation (2.60),

$$\begin{aligned}\epsilon(t,\theta) &= H^{-1}(q,\theta)\left[y(t) - G(q,\theta)u(t)\right] \\ &= H^{-1}(q,\theta)\left[(G_0(q) - G(q,\theta))u(t) + (H_0(q) - H(q,\theta))e(t)\right] + e(t) \quad (3.13)\end{aligned}$$

and further re-arrange this equation such that

$$\epsilon(t,\theta) = H^{-1}(q,\theta)\tilde{T}^T(q,\theta)\chi_0(t) + e(t) \quad (3.14)$$

with

$$\tilde{T}(q,\theta) = T_0(q) - T(q,\theta) \quad (3.15)$$

and

$$T(q,\theta) = \begin{bmatrix} G(q,\theta) \\ H(q,\theta) \end{bmatrix} \qquad T_0(q) = \begin{bmatrix} G_0(q) \\ H_0(q) \end{bmatrix} \qquad \chi_0(t) = \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} \quad (3.16)$$

Assuming that $G(q,\theta)u(t)$ and $G_0(q)u(t)$ depend only on $e(s)$ for $s < t$ and that $H(q,\theta)$ (and $H_0(q)$) is monic, then the last term in equation (3.13) is independent of the rest and in reflection of equations (2.60) through (2.63) we have (Forssell, 1999),

$$\Phi_\epsilon(\omega) = \frac{1}{|H(e^{j\omega},\theta)|^2}\tilde{T}^T(e^{j\omega},\theta)\Phi_{\chi_0}(\omega)\tilde{T}(e^{-j\omega},\theta) + \lambda_0 \quad (3.17)$$

with

$$\Phi_{\chi_0}(\omega) = \begin{bmatrix} \Phi_u & \Phi_{ue} \\ \Phi_{eu} & \lambda_0 \end{bmatrix} \quad (3.18)$$

this, **w.p 1 as** $N \to \infty$, translates into,

$$\hat{\theta}_N \to D_c = \arg\min_{\theta \in D_\mathcal{M}} \frac{1}{4\pi}\int_{-\pi}^{\pi} \tilde{T}^T(e^{j\omega})\Phi_{\chi_0}(\omega)\tilde{T}(e^{-j\omega})\frac{1}{|H(e^{j\omega},\theta)|^2}d\omega \quad (3.19)$$

In equation (3.19) we thus have an understanding that the prediction error estimate $T(q,\hat{\theta}_N)$ will converge to the true transfer function $T_0(q)$ if the parameterisation is flexible enough so that $\mathcal{S} \in \mathcal{M}$ and if the structure $\tilde{T}^T(e^{j\omega},\theta)$ does not lie in the left null space of $\Phi_{\chi_0}(\omega)$ (Forssell, 1999). This latter condition of $\tilde{T}^T(e^{j\omega},\theta)$ not lying in the left null space of $\Phi_{\chi_0}(\omega)$ is directly related to an informative enough data set. An informative data set is achieved when the matrix $\Phi_{\chi_0}(\omega)$ is positive definite for almost all frequencies.

It is important to note that equation (3.19) is equation (2.63) factorised differently in order to allow for a better interpretation of the convergence properties under closed loop conditions. Note that under open-loop conditions the off-diagonal values in equation

(3.18) will be zero.

Consider the factorisation

$$\Phi_{\chi_0}(\omega) = \begin{bmatrix} 1 & \Phi_{ue}(\omega)/\lambda_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Phi_u^r(\omega) & 0 \\ 0 & \lambda_0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \Phi_{eu}(\omega)/\lambda_0 & 1 \end{bmatrix} \tag{3.20}$$

It follows that $\Phi_{\chi_0}(\omega)$ is positive definite for almost all frequencies if $\Phi_u^r(\omega)$ is positive for almost all frequencies. Recalling equation (3.9), stating that $\Phi_u^r(\omega)$ is part of the input power spectrum that originates from the reference signal r and, under a linear feedback law, is defined as $\Phi_u^r(\omega) = |S_0(j\omega)|^2\Phi_r(j\omega)$. Since $S_0(e^{j\omega})$ can only have finitely many zeros, the informative data condition becomes $\Phi_r(\omega) > 0$ for almost all frequencies. In other words, the reference signal, $r(t)$, should be persistently exited for closed loop identifiability (Forssell, 1999).

Bias Distribution

Inserting the alternative factorisation of $\Phi_{\chi_0}(\omega)$,

$$\Phi_{\chi_0}(\omega) = \begin{bmatrix} 1 & 0 \\ \Phi_{eu}(\omega)/\Phi_u(\omega) & 1 \end{bmatrix} \begin{bmatrix} \Phi_u(\omega) & 0 \\ 0 & \Phi_e^r(\omega) \end{bmatrix} \begin{bmatrix} 1 & \Phi_{ue}(\omega)/\Phi_u(\omega) \\ 0 & 1 \end{bmatrix} \tag{3.21}$$

into equation (3.19), gives the following characterisation of the limit estimate **w.p. 1 as** $N \to \infty$ (Forssell, 1999):

$$\begin{aligned} \hat{\theta}_N \to D_c &= \arg\min_{\theta \in D_{\mathcal{M}}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big[|G_0(e^{j\omega}) + B(e^{j\omega}, \theta) - G(e^{j\omega}, \theta)|^2 \Phi_u(\omega) \\ &+ |H_0(e^{j\omega}) - H(e^{j\omega}, \theta)|^2 \Phi_e^r(\omega) \Big] \frac{1}{|H(e^{j\omega}, \theta)|^2} d\omega \end{aligned} \tag{3.22}$$

where

$$B(e^{j\omega}, \theta) = (H_0(e^{j\omega}) - H(e^{j\omega}, \theta))\Phi_{eu}(\omega)/\Phi_u(\omega) \tag{3.23}$$

B is termed the *Bias Pull* and is an indicator of the bias inclination in the system estimate. It follows that this bias inclination will be small over frequency ranges where either the noise model is good, i.e. $(|H_0(e^{j\omega}) - H(e^{j\omega}, \theta)|)$ is small, or the input power spectrum, Φ_u, dominates the cross power spectrum, $\Phi_{eu}(\omega)$. It is also noted that the bias term, $B(e^{j\omega}, \theta)$ is always identically zero in open-loop since $\Phi_{eu}(\omega)$ is zero.

Asymptotic Properties

Recalling the definition of the black-box transfer function covariance matrix given by equation (2.68) and the definition given by equation (3.18) we have,

$$\mathbf{Cov}\begin{bmatrix} G(e^{j\omega},\hat{\theta}_N) \\ H(e^{j\omega},\hat{\theta}_N) \end{bmatrix} \approx \frac{n}{N}\Phi_v(\omega)\Phi_{\chi_0}^{-1}$$

This expression indicates that the asymptotic covariance of the transfer function is given by the noise-to-signal ratio (v, noise; χ_0, signal) multiplied by the model order, n, to number of data, N, ratio.

Additionally this expression holds regardless of the nature of the feedback. Under open-loop conditions the off diagonal values of Φ_{χ_0} are zero and we have (Ljung, 1987: 251)

$$\mathbf{Cov}\ G(e^{j\omega},\hat{\theta}_N) \approx \frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u(\omega)} \tag{3.24}$$

$$\mathbf{Cov}\ H(e^{j\omega},\hat{\theta}_N) \approx \frac{n}{N}|H_0(e^{j\omega})|^2 \tag{3.25}$$

The above formulations reflect the points discussed in section 2.6.1. It is clear that the signal-to-noise ratio and hence the extent of input signal excitation have a large role to play in model estimation.

Under closed-loop conditions equation (3.24) becomes (Forssell & Ljung, 1997)

$$\mathbf{Cov}\ G(e^{j\omega},\hat{\theta}_N) \approx \frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u^r(\omega)} \tag{3.26}$$

This illustrates precisely the closed-loop informative data condition for accurate estimation in contrast to the open-loop condition. Here it is required that the input signal spectral component that originates in the reference signal r be excited.

3.2.2 The Indirect Method

Given the structure and characteristics of the feedback controller, $C(q)$, are known, together with the reference signal, the indirect approach may be used. This approach may be broken down into two components (Forssell & Ljung, 1997):

1. Identify the closed-loop system from the reference signal r to the output y defined by G_0^{cl}.

2. Compute the open-loop model, $\hat{G}_N$, from the closed-loop model using knowledge of the regulator.

The advantage this method has is that any identification technique may be used since estimating the closed-loop system from measurements of r and y is an open-loop problem. This means that methods such as spectral analysis, IV and subspace methods may be used (Forssell & Ljung, 1997).

In the prediction error method, which allows for arbitrary parameterisations, it is natural to parameterise the model of the closed-loop system in terms of the open-loop model (Forssell, 1999),

$$y(t) = \frac{G(q,\theta)}{1 + C(q)G(q,\theta)} r(t) + H_*(q)e(t) \tag{3.27}$$

Here a fixed noise model H_* is used which is common in this approach. This is mostly due to the fact that since the identification problem has become an open-loop one, there is no loss in consistency if a fixed noise model is chosen to shape the bias distribution. Through this parameterisation, computing the open loop estimate is avoided and the open-loop model is delivered directly making the second step superfluous.

An alternative parameterisation approach is to use the so called dual Youla parameterisation that parameterises all systems that are stabilised by a regulator $C(q)$. See Vidyasagar (1985)

Bias Distribution

Given the system description in equation (3.27) the following may be established (Forssell, 1999):

$$\hat{\theta}_N \to D_c = \arg\min_{\theta \in D_{\mathcal{M}}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \left| \frac{G_0(e^{j\omega}) - G(e^{j\omega},\theta)}{1 + C(e^{j\omega})G(e^{j\omega},\theta)} \right|^2 \frac{|S_0(e^{j\omega})|^2 \Phi_r(\omega)}{|H_*(e^{j\omega})|^2} d\omega \tag{3.28}$$

This indicates that the indirect method can give consistent estimates of $G_0(e^{j\omega})$ provided $\mathcal{S} \in \mathcal{M}$ even with a fixed noise model. Unlike the direct approach it is difficult to directly quantify the bias. The estimation will however be a compromise between making $G(e^{j\omega},\theta)$ close to G_0 and making the model sensitivity function, $1/(1 + C(e^{j\omega})G(e^{j\omega},\theta))$, small. There is thus a bias-pull towards a transfer function that gives small sensitivity for the given regulator (controller) C. Literature (Gevers, 1986) shows this to be advantageous if the model is to be used for control.

Asymptotic Properties

From equation (3.27) the asymptotic variance of $G^{cl}(e^{j\omega}, \hat{\theta}_N)$ is

$$\mathbf{Cov}\ G^{cl}(e^{j\omega}, \hat{\theta}_N) \approx \frac{n}{N} \frac{|S_0(e^{j\omega})|^2 \Phi_v(\omega)}{\Phi_r(\omega)} \tag{3.29}$$

regardless of the noise model H_* (Forssell & Ljung, 1997). Forssell & Ljung (1997) goes on to verify that from the above expression the following is obtained:

$$\textbf{Cov}\ G(e^{j\omega}, \hat{\theta}_N) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u^r(\omega)} \tag{3.30}$$

which is the same asymptotic variance obtained under the direct approach

3.2.3 The Joint Input-Output Method

The third of the primary approaches used in closed-loop system identification is the so called joint input-output approach. As mentioned before the primary assumption here is that the input is generated using a regulator of a known form. Exact knowledge of the regulator is not required. Note this is an advantage over the indirect method where complete knowledge of the regulator is necessary.

There have been many adaptations and variations of this approach. The three most discussed in literature are the Coprime, Two-Stage and Projection methods. It is noted however that each of these methods do fall under a unifying framework as Forssell & Ljung (1999) shows for the multivariable case.

Of these three joint input-output approaches, focus will be put on the projection method. This is due to the fact that the projection method's characteristic dissimilarities from the other methods are what allow it to be more applicable to systems with arbitrary, non-linear feedback mechanisms.

The projection method consists of the following two steps (Forssell & Ljung, 2000):

1. Estimate the parameters s_k in the non-causal FIR model

$$u(t) = S(q)r(t) + e(t) = \sum_{k=-M_1}^{M_2} s_k r(t-k) + e(t)$$

   where M_1 and M_2 are chosen so large that any correlation between $u(t)$ and $r(s)$ for $t - s \notin [-M_1, \ldots, M_2]$ can be ignored, and simulate the signal $\hat{u}(t) = \hat{S}(q)r(t)$

2. Identify the open-loop system using a model of the kind

$$y(t) = G(q, \theta)\hat{u}(t) + H_* e(t)$$

The first step in this approach may be viewed as a least squares projection of u onto r, hence the name *the projection method*. Conceptually it is best understood as an identification procedure where the system input and the output maybe be considered as outputs of a system driven by the reference input $r(t)$ and unmeasured noise $v(t)$. Knowledge of the system and the controller is recovered from this joint model.

The two-stage method can essentially be explained with the same two steps. What distinguishes the projection method from the two-stage method is that a non-causal model is used in the first step instead of a causal model. In the two-stage method causal high-order FIR or ARX models in are implemented in the first step (Forssell & Ljung, 2000).

The properties of the projection method will be discussed in the next sections. For further detail on the Coprime, Two-stage and Projection methods see van den Hof et al. (1995), van den Hof & Schrama (1993) and Forssell & Ljung (2000) respectively.

Bias Distribution

Restating that for the projection method we have the following definitions,

$$y(t) = G(q,\theta)\hat{u}(t) + H_* e(t) \qquad \hat{u}(t) = \hat{S}(q)r(t) \qquad \tilde{u}(t) = u(t) - \hat{u}(t) \tag{3.31}$$

this results, **w.p. 1 as** $N \to \infty$, in

$$\hat{\theta}_N \to D_c = \arg\min_{\theta \in D_{\mathcal{M}}} \frac{1}{4\pi} \int_{-\pi}^{\pi} |G_0(e^{j\omega}) + \tilde{B}(e^{j\omega},\theta) - G(e^{j\omega},\theta)|^2 \frac{\Phi_u(\omega)}{|H_*(e^{j\omega})|^2} d\omega \tag{3.32}$$

where

$$\tilde{B}(e^{j\omega}) = G_0 \Phi_{\tilde{u}\hat{u}}(\omega) \Phi_{\hat{u}}^{-1}(\omega) \tag{3.33}$$

This holds true for both the two-stage and the projection methods for arbitrary feedback mechanisms (Forssell, 1999).

Equation (3.33) shows that $\Phi_{\tilde{u}\hat{u}}(\omega)$ needs to be zero (i.e. $\tilde{u}$ and $\hat{u}$ be uncorrelated) for consistency to prevail for both the two-stage and the projection methods. By construction the projection method asymptotically achieves this regardless of the nature of the feedback . This condition which is more applicable to the projection method is due to the non-causal model used in the first step of the approach (Forssell & Ljung, 2000).

Asymptotic Properties

(Forssell & Ljung, 1997) shows that the transfer function asymptotic variance may be given by

$$\mathbf{Cov}\ G(e^{j\omega},\hat{\theta}_N) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u^r(\omega)}$$

which is the same as that established for the direct and indirect approaches. This hold for the projection, two-stage and coprime adaptations of the approach.

3.2.4 Overview on Closed-Loop Approaches

The following overview is compiled with reference to Forssell & Ljung (1999), Forssell & Ljung (1997) and Forssell (1999):

The Direct Approach

- If $\mathcal{S} \in \mathcal{M}$ and the noise model contains the true noise model, then consistency and optimal accuracy will be achieved regardless of the feedback.
- Requires a correct noise model to avoid incurring bias in the system transfer function.
- Does not require any knowledge of the feedback mechanism or controller.

The Indirect Approach

- Requires perfect knowledge of the controller but can be applied with fixed noise models and still allow for consistency.
- Allows for other open-loop identification approaches to be applied that would otherwise fail under closed-loop conditions.
- Does typically give the best results in terms of identification for control.

The Joint Input-Output Approach

- Gives consistent estimations of the open-loop system regardless of the noise model used.
- With exception of the projection method, only knowledge or specification on the controller is that it has a certain linear structure.
- The projection method is capable of consistent identification under non-linear controller structures.

Other Comments

- Given a linear feedback mechanism, as the model order tends to infinity the asymptotic variance expressions are the same for all methods .
- Given a finite model order, the indirect approach generally gives the worst accuracy.
- The improved accuracy achieved by the direct approach can be attributed to the fact that a constrained noise model that contains the true noise dynamics is used.

3.3 Multivariable identification and Controller Specific Effects

Most literature concerned with the fundamentals of closed-loop system identification assumes that the system is scalar and may be accurately represented through a linear model and is controlled by a linear controller. It must be said however that the majority of systems are to some extent non-linear and multivariable and in some instances are controlled by non-linear controllers.

3.3.1 Multivariable Closed-Loop Identification

In section 2.7 it was mentioned the extension of system identification methods from those applicable to scalar systems to those applicable to multivariable systems, is mostly a task or notation changes and some slight non-trivialities in the parameterisation and estimation approaches. Under the condition where the data set, Z^N, is closed-loop data, the same can generally be said.

Forssell & Ljung (1999) produced a survey paper which presents the entire closed-loop identification problem and all its aspects concerning the prediction error framework, parameter estimation and asymptotic properties for the multivariable condition. The paper affirms the understanding that all methods presented for the SISO case are analogously applicable to the MIMO case with a few exceptions that are mostly concerned with the properties and type of controller used.

The next sections introduce some aspects on controller specific effects on identification. This will allow for some insight into the reasoning behind the theory presented in section 3.2 and additionally illustrate some of the multivariable closed-loop identification issues .

3.3.2 Linear Decentralised Controllers

Sections 3.1.2 and 3.2 showed how introducing the feedback can reduce the generation of informative data via. The extent of this can be controller specific. The tuning of the controller determines the strength of the "feedback path". A very detuned controller means the control actions will be sluggish resulting in the "forward path" dynamics dominating the data (Doma et al., 1996). This implies that the controller parameters do have effects on identifiability.

3.3.3 Non-linear Controllers and MPC

As stated in section 3.2, Forssell (1999) discussed how the indirect and joint input-output approaches become very involved and complex in the face of non-linear controllers with the possible exception of the projection method. The problem with this is that

many multivariable systems are dynamically complex and are consequently controlled by centralised or model-based controllers which are non-linear. Furthermore, in some cases even the linear PID controllers exhibit some non-linear traits due to the necessary delimiters and anti-windup functions (Forssell & Ljung, 1999) (note that the extent of this is dependant on input saturation which is dependent on controller tuning). Thus it is relatively important to have an understanding of the effect of a non-linear controller on the closed-loop identification process.

In the previous section it was mentioned that controller tuning parameters have effects on the extent of identifiability. van den Hof (1998) takes this further and establishes that switching controller parameters contributes to providing informative data that allows for consistent identification. In essence a variable parameter controller is a non-linear controller. This means that the introduction of a non-linear controller allows for the data to be more informative while additionally complicating the identification methodology. This is with exception of the direct approach as this approach is completely independent of the feedback and the type of controller used.

From this we gather that the direct approach is the most suitable for complex, non-linear controllers. This is directly applicable to model predictive control (MPC) which is very much a non-linear controller since each control action is computed based on an optimisation routine every iteration (Zhu & Butoyi, 2002).

Special attention has been given in literature to closed-loop identification for MPC or on a system using MPC. Zhu & Butoyi (2002) and Zhu (1998) have concentrated on a method termed the asymptotic method, which uses automated multivariable tests to obtain specified desired input signal power spectra. This and other approaches have also been extended into recursive online model predictive control identification where the model is iteratively updated (Shouche et al., 1998)(Shouche et al., 2002).

3.4 Non-Linear Closed-Loop System Identification

While the focus of this work is primarily on linear estimation methods, it is of value to address some common non-linear estimation methods found in literature that have been applied to closed-loop conditions. It is noted however, that non-linear identification approaches that do not rely on physical insight are rare and this is more so the case when using closed-loop data. The more established method of the Youla-Kucera approach will briefly be introduced followed by some other general approaches.

3.4.1 The Youla-Kucera Parameter Approach

This approach is related to the Dual-Yuala approach (the indirect method in section 3.2) where specific tailor-made parameterisations are used. It was initially investigated

in Hansen (1989) and further investigated in a non-linear framework in Desgupta & Anderson (1996). Fujimoto et al. (2001) explains this approach as one which relies on the ability to parameterise the unknown system in terms of a known nominal model and the known controller along with the so-called Youla-Kucera parameter (Vidyasagar, 1985) associated with the system through the use of coprime factorisations. This turns the identification problem into an open-loop one.

While the given description of the approach is brief it must be noted that this approach has been extensively studied in literature. Additionally there are many adaptations of this approach to study non-linear controllers (Linard et al., 1999) and kernel form representations (Fujimoto et al., 2001).

3.4.2 Other Approaches

While the method identified in the previous section does seem to be the more widely covered non-linear closed-loop system identification method in literature, many other approaches have been established. In section 2.3.2 the Block-Oriented models were introduced as an approach to modelling non-linear systems. There has been some effort in applying this to the closed loop data, in particular Eskinat et al. (1991). In this particular instance it was shown that the non-linear Hammerstein Block-Oriented model did generally better represent the non-linear system than a linear model.

Again recalling a non-linear modelling method introduced in section 2.3.2, the Volterra-Series has been successfully applied to closed loop-data. Zheng & Zafiriou (1994) does this with a control-relevant approach. A Volterra-series model obtained from closed-loop data was combined with an uncertainy model and implemented in a high performance model based controller.

Neural Networks models trained on closed-loop data, have also been used successfully (Atsushi & Goro, 2001). Although this approach is far from the typical parametric or convolution approaches discussed in this work, its success illustrates the extent of diverse approaches and continuous effort and justification in furthering the field of closed-loop identification.

3.5 Chapter Summary

This chapter presented the closed-loop system identification problem as being one where feedback causes a lesser degree of information content in identification data and input-noise correlations cause bias estimations. Three approaches to closed-loop system identification were presented:

- The direct approach.

- The indirect approach.
- The join input-output approach.

with a comparative overview of these approaches given in section 3.2.4. Of these three approaches the direct approach was found to be most suitable due to its simple application and its ability to readily identify a system regardless of the controller complexity. It was however found that in order to reduce model inaccuracy and bias using this method, it is important to persistently excite the system and assure an accurate noise model.

CHAPTER 4

Model Validation Theory

This chapter deals with the final component in system identification, model validation. Several essential validation techniques are presented, with a specific emphasis on linear approximations of non-linear systems and the validation of such approximations. The last sections further study non-linear dynamics detection and the determination of the extent of unmodelled non-linear dynamics via higher order residual correlation analysis.

4.1 General Linear Model Validation

Several methods have been developed for linear model validation. Sections 4.1.1 through 4.1.5, that follow, present the more common approaches to model validation or gaining insight into the validity of a model.

4.1.1 Statistical Tools, Hypothesis Testing and Confidence Intervals

There are several chapters in statistical texts dedicated to defining validation criteria, confidence intervals and hypothesis testing, specifically for linear regression models.

Literature shows hypothesis testing to have wide applicability in modelling, specifically parametric modelling. Ljung (1987: 422) illustrates the use of statistical hypothesis testing for comparing model structures and rejecting model structures and parameterisations based on defined confidence intervals. Montgomery et al. (2001: 319) presents a section on hypothesis testing in linear regressions where tests for regression significance and tests on individual regression coefficients are presented. The former allowing for verification of the linear relationship between the true response variable, $y(q)$, and the

regression coefficients, the latter allowing for an investigation into the potential value of each of the regression coefficients in the model. This is very useful in determining the order of the model as it essentially statistically investigates the value in reducing or increasing the model order.

There are several other statistically based goodness of fit testing methods. An example that has found wide application in system identification is the Chi-squared (χ^2) goodness of fit test based on the Chi-squared distribution. An overview on this method is found in Cressie & Read (1989).

These tests together with the many well developed procedures to obtain confidence intervals in coefficients themselves, allow for a very practical understanding of the goodness of fit of the model and give insight into the uncertainty of the model accuracy (Montgomery et al., 2001: 326).

4.1.2 Model-comparison Based Validation

This validation approach involves applying statistical tests that generate quantitative values that may be used to compare models pairwise and to select the best model with the minimum or maximum statistical value. A commonly used criterion that falls under this approach is Akaike's Final Prediction Error (FPE) and the closely related Information Theoretic Criterion (AIC). Once several possible models of a system have been computed, they may be compared using these criteria. According to Akaike's theory, the most accurate model has the smallest FPE and AIC (Ljung, 1995).

Akaike's FPE is defined as follows (Ljung, 1987: 420):

$$FPE = \left(\frac{1 + d_{\mathcal{M}}/N}{1 - d_{\mathcal{M}}/N}\right) V_N(\hat{\theta}_N, Z^N) \tag{4.1}$$

with

$$d_{\mathcal{M}} = \dim \theta^{\mathcal{M}}$$

signifying the number of estimation parameters. For a quadratic norm this becomes

$$FPE = \left(\frac{1 + d_{\mathcal{M}}/N}{1 - d_{\mathcal{M}}/N}\right) \cdot \frac{1}{2}\epsilon^2(t, \hat{\theta}_N) \tag{4.2}$$

This criterion may be understood as a reflection of the prediction-error variance obtained, on average, between the model defined by $\hat{\theta}_N$, and a data set other than that used in the identification procedure.

Akaike's AIC is presented as (Ljung, 1995):

$$AIC = \log V_N(\hat{\theta}_N, Z^N) + \frac{2d_{\mathcal{M}}}{N} \tag{4.3}$$

For $d_{\mathcal{M}} << N$

$$AIC = \log\left(V_N(\hat{\theta}_N, Z^N) + \left(1 + \frac{2d_{\mathcal{M}}}{N}\right)\right) \tag{4.4}$$

this stems out of general identification theory closely related to the maximum-likelihood approach, the advantage being that this criterion is applicable to all model structures. Ljung (1987: 189) shows how through Akaike's AIC, a form of mean log-likelihood identification criterion is obtained and that minimisation of this criterion is a reflection of the goodness of fit.

It is noted however, that AIC and FPE comparison criteria are mostly geared towards model structure comparisons than comparisons between models of the same structure.

4.1.3 Model Reduction

A procedure or test that indicates whether the estimate is a simple and appropriate description of the true system is to reduce is model order reduction. Ljung (1987: 426) states that if the model order can be reduced without affecting the input-output properties very much, then the original model was unnecessarily complex.

4.1.4 Simulation and Consistent Model Input-Output Behaviour

While it is good practise to compare system and simulated responses to common inputs, Ljung (1987: 425) states that it is essential to compare system responses and model responses in Bode Plots. Especially useful is the comparison between spectral analysis estimates, such as those rooted in smoothed out frequency domain Empirical Transfer-Function Estimates, and Bode plots as they are formed from different underlying assumptions. This Bode plot comparison approach is especially useful when comparing different models obtained by PEM (Prediction Error Method) with different structures. This typically does allow for insight into whether essential dynamic features of the system have been captured in a specific model structure.

Additionally, while most validation and model comparison approaches investigate the accuracy of the model's ability to simulate, it is also important to assess the model's prediction capacity, specially if the model is to be used for control purposes. The primary difference between simulation and prediction is that the former uses *previously calculated* outputs generated by the model to generate a next value while the latter uses *previously measured (from the data set) and previously calculated* values to generate the next value (Ljung, 1995). From the model validation point of view, the difference is that simulation typically emphasises low frequency behaviour while predictions emphasise mid-range frequency behaviour.

Validation through simulation is typically termed cross-validation. This is as the model is validated against an independent data set of measurements containing system

inputs and outputs of the true system. The model is disturbed (simulated) using the measured inputs and response is compared against the measured outputs. Figure 4.1 illustrates.

4.1.5 Residual Analysis

Residual (estimation errors) analysis is a very well developed field of model diagnosis and validation founded mostly in the 1960's by Anscombe & Tukey (1963) and is very much a strong component of cross-validation. An enormous amount has been written about this field and many text books have been published providing its different applications and adaptations, an example is Cook & Weisberg (1982). As Box & Jenkins (1970) states, studying the difference between the model output and the output of the true systems - residual analysis - allows for the study of the existence and nature of model inadequacies, thus its place in model validation. Figure 4.1 illustrates. Specifically in the

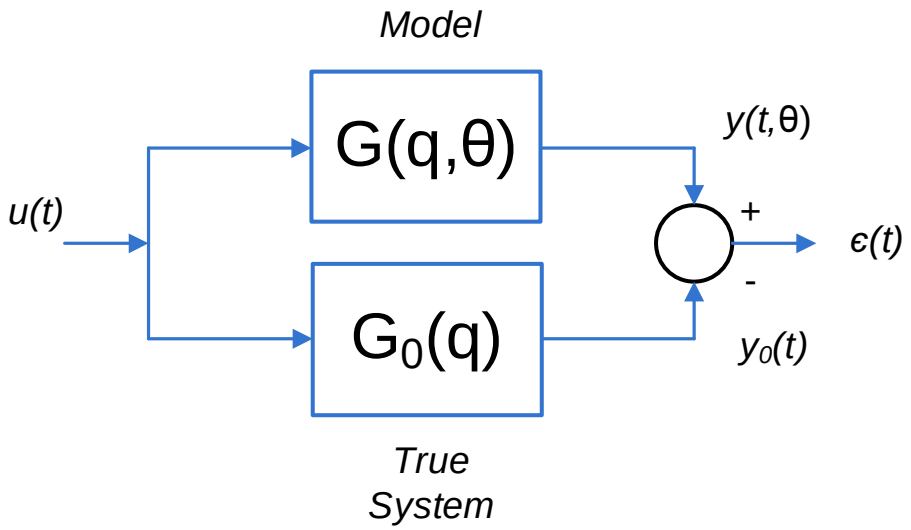


Figure 4.1: Model Residual

case of regression models, there are many well known methods of residual analysis where statistical tools are applied to extract information about the model and ascertain as to how good a representation of the true system it really is.

In this work concise literature on two approaches will be presented. The first is standard residual plotting and the second is residual correlation analysis. Section 4.2.2 discusses an extension of the second approach.

Standard Residual Plotting

Standard residual plots are those in which the residuals, $\epsilon_N(t,\theta)$, are plotted against the estimated values, $\hat{y}_N(t,\theta)$, or other functions of $\varphi(t)$ (Cook & Weisberg, 1982). Figure 4.2 is an example of such a plot. Draper & Smith (1998: 63) reveals that a plot of

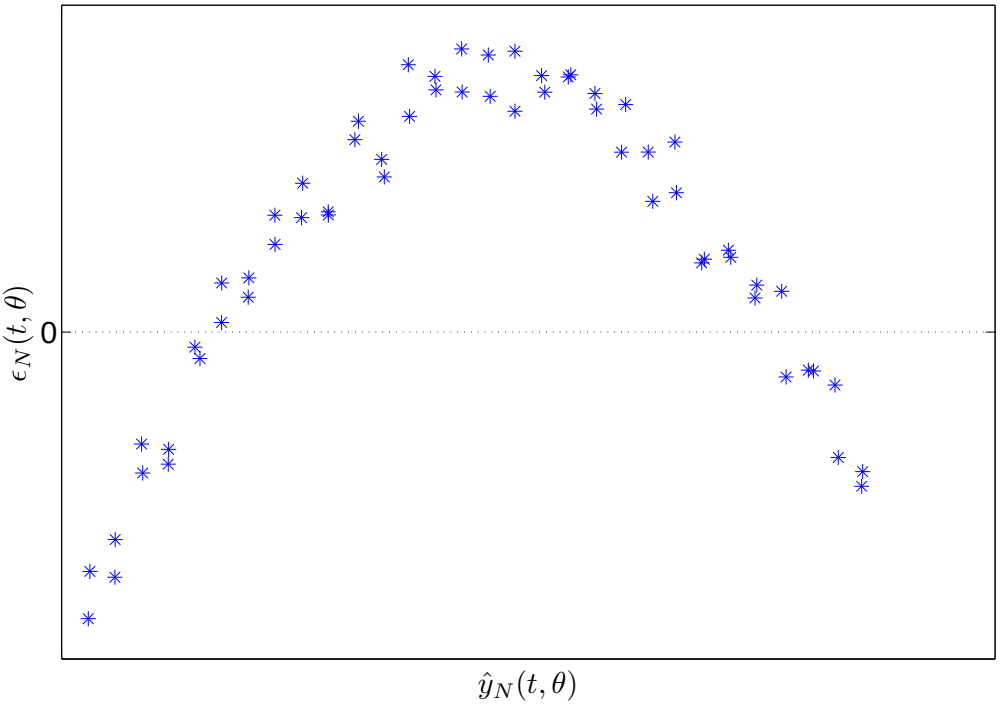


Figure 4.2: Residual plot

residuals from a satisfactory or correctly specified model would resemble a horizontal band of haphazard points such as figure 4.3a. In contrast to this there are many possible unsatisfactory plots that indicate nonconstant error variances of which the most typical forms are given in figures 4.3b through 4.3d. It is noted that figure 4.2 resembles the residual plot indicated in figure 4.3d revealing an unsatisfactory model. Anscombe (1973) makes the interesting argument in motivation for residual plotting in that through the indication of nonconstant error variances, unmodelled non-linearites are implied.

The haphazardness or randomness of data presented in residual plots that reflect a correctly specified model does have reason behind it. Recalling equation (2.15)

$$y(t) - \hat{y}(t|t-1) = -H^{-1}(q)G(q)u(t) + H^{-1}(q)y(t) = e(t)$$

and the concept implied that the prediction error represents the component of the output, $y(t)$, that cannot be predicted from past data and is formulated by $e(t)$, an unpredictable white noise signal. It is thus understood that for an accurate model the residuals should be white - or random - and fit the assumption that states:

$$\epsilon(t) \approx e(t) \tag{4.5}$$

Draper & Smith (1998) is referred to as a further source for more detailed literature

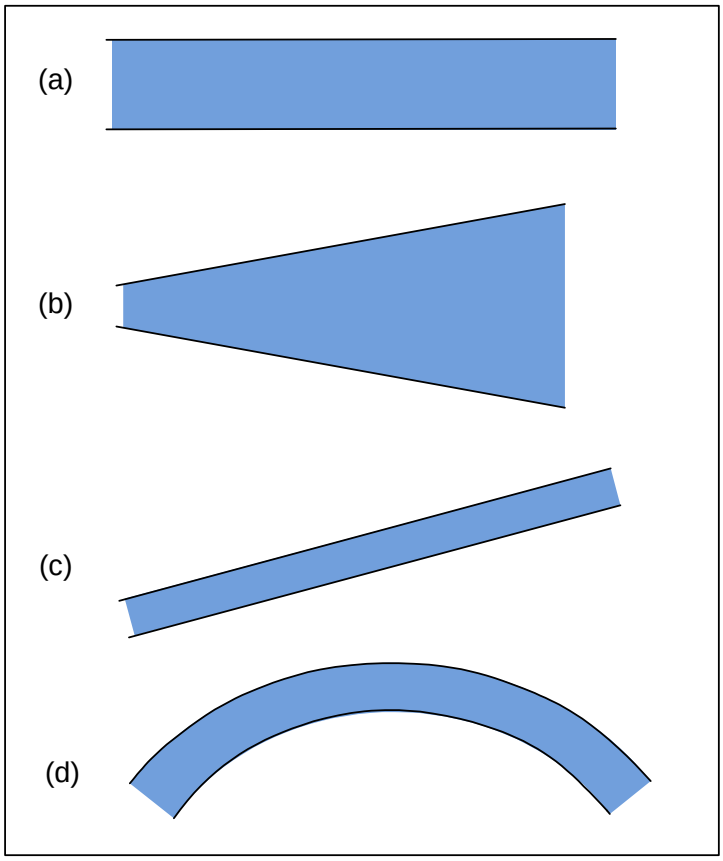


Figure 4.3: (a) Acceptable residual plot (b) - (d) Unacceptable residual plots

on residual plots, specifically on gathering model diagnostic interpretations from plots.

Residual Correlations

The analysis of correlations amongst estimation residuals (estimation errors) and between the estimation residuals and the system inputs are commonly used linear validation approaches. This analysis may be considered as a direct quantification of the concept behind residual plotting and consists of the following two tests (Ljung, 1995):

The Whiteness Test : This test is based on the condition that the residuals, or the prediction errors between the model and the system, of a good model should be independent of each other and of past data. Therefore a good model has residuals that are uncorrelated, i.e confirmation of the condition assumed through equation (4.5). The whiteness test is thus obtained through the residual auto-correlation (Ljung, 1987: 428),

$$R_\epsilon(\tau) = \bar{E}[\epsilon(t)\epsilon(t-\tau)] \tag{4.6}$$

The Independence Test : Is a measure of the correlation between the residuals and the corresponding inputs. A good model has residuals uncorrelated to past inputs. A peak in the residual-input correlation function at lag τ_k indicates that the output

$y(t)$ that originates from input $u(t-\tau_k)$ is not properly described in the model. Thus the model has unmodelled dynamics. This correlation is given by the cross-correlation function as follows (Ljung, 1987: 429):

$$R_{\epsilon u}(\tau) = \bar{E}[\epsilon(t)u(t-\tau)] \tag{4.7}$$

Through the assertion of these two tests the following conditions for model adequacy are formed (Billings & Voon, 1986):

$$\left.\begin{array}{l} R_{\epsilon}(\tau) = \delta(\tau) \\ R_{\epsilon u}(\tau) = 0 \ \forall \tau \end{array}\right\} \tag{4.8}$$

where $\delta(\tau)$ is the auto-correlation function of a white signal, as discussed in section 2.4.1 (see figure figure 2.6(b)), which shows no correlations patterns but an infinitely narrow impulse at $\tau = 0$,

$$\delta(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \textbf{otherwise} \end{cases} \tag{4.9}$$

Box & Jenkins (1970) further notes that if the system model is correct but the noise model is incorrect, the residual auto-correlation function shows marked correlation patterns such that $R_{\epsilon}(\tau) \neq \delta(\tau)$ but they will be uncorrelated with the input such that $R_{\epsilon u}(\tau) = 0 \ \forall \tau$. Alternatively, if both the noise model and the system model are inadequate, both the residual auto-correlation and residual-input cross-correlation will show correlation patterns.

4.1.6 Model Uncertainty Considerations

An important consideration in modelling is the extent of uncertainty in the model. There are many texts that delve into the details of accounting for model uncertainty in control, specifically where model based controllers are implemented. Skogestad & Postlethwaite (1997: 255) details that the various sources of model uncertainty may be grouped as follows:

Parametric Uncertainty : Here the structure of the model is known, but some of the parameters are uncertain.

Neglected and Unmodelled Dynamics Uncertainty : Here model uncertainty is rooted in unmodelled dynamics.

Skogestad & Postlethwaite (1997: 255) further details the different manners in which these model uncertainties may be represented in the control system description.

In terms of the identification - validation procedure, there have been efforts to incorporate uncertainty. Hjalmarsson (1993) presented work where considerations on the effects

of incorporating model uncertainty and ill-defined model terms in the model validation process where studied. Furthermore as sections 2.6.2 and 3.2 revealed, the description of model uncertainty through the covariance of the estimations is an established approach to viewing mode uncertainty information (Ljung, 1995).

4.2 Linear Time Invariant Approximations of Non-linear Systems

In this section we look linear model validation under the condition where the dynamics of the system being modelled might be non-linear. This introduces issues such as being able to determine whether the residual is due to bad identification rooted in unfavourable model parameterisation or structure selection or due to unmodelled non-linearities. Additionally if a system is non-linear it is useful to determine how well the linear approximation fits the linear components of the non-linear system and under what conditions do the non-linearities distort this linear approximation.

4.2.1 LTI Aprroximation Sensitivity to System Non-Linearity

The LTI Second Order Equivalent

LTI approximations of non-linear systems have been studied in many frameworks. Related material on these different frameworks can be found in Schoukens et al. (2003), Pintelon & Schoukens (2001) and Mäkilä & Partington (2003). This work focus on framework presented by Enqvist & Ljung (2004), where under the PEM the input and output signals are assumed to have the following properties:

1. The input $u(t)$ and output $y(t)$ are real-valued stationary stochastic processes with $E[u(t)] = E[y(t)] = 0$.

2. There exists $K > 0$ and α, $0 < \alpha < 1$ such that the second order moments, $R_u(\tau)$, $R_{yu}(\tau)$ and $R_y(\tau)$ satisfy

$$\begin{aligned} |R_u(\tau)| &< K\alpha^{|\tau|} \qquad \forall \tau \in \mathbb{Z} \\ |R_{yu}(\tau)| &< K\alpha^{|\tau|} \qquad \forall \tau \in \mathbb{Z} \\ |R_y(\tau)| &< K\alpha^{|\tau|} \qquad \forall \tau \in \mathbb{Z} \end{aligned}$$

3. The z-spectrum, $\Phi_u(z)$ (the z-transform of $R_u(\tau)$) has a canonical spectral factorisation

$$\Phi_u(z) = L(z) r_u L(z^{-1})$$

where $L(z)$ and $1/L(z)$ are causal transfer functions, $L(+\infty) = 1$ and r_u is a positive constant. This implies the power spectrums, or spectral densities, $\Phi_u(e_{j\omega})$ and $\Phi_{yu}(e^{j\omega})$, are well defined for all $\omega \in [-\pi, \pi]$.

In addition to these assumptions it is implied that a OE (output-error) structure is used, i.e. $H(q, \theta) = 1$, which is illustrated in figure 4.4.

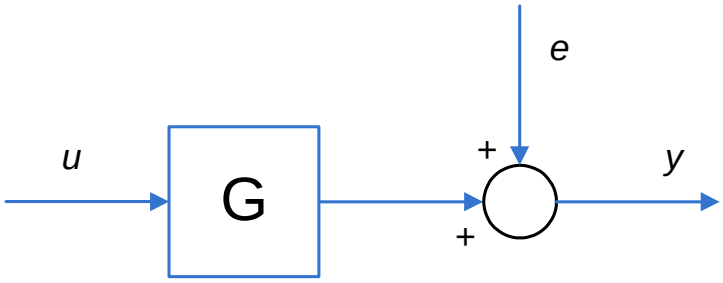


Figure 4.4: Output-Error Model Structure

These assumptions and the afore mentioned structure selection allow for the definition of the *Output Error LTI Second Order Equivalent* (OE-LTI-SOE), as the cuasal, stable, LTI approximation of a non-linear system that minimises the mean-square error, $E[(y(t) - G(q)u(t))^2]$. i.e. a convergence theorem is presented such that

$$G_{0,OE}(q) = \arg\min_{G \in \mathcal{Y}} E[(y(t) - G(q)u(t))^2] \tag{4.10}$$

where

$$G_{0,OE}(e^{j\omega}) = \left[\frac{\Phi_{yu}(e^{j\omega})}{\Phi_u(e^{j\omega})}\right]_{causal} \tag{4.11}$$

given a non-linear system. Note $\mathcal{Y}$ denotes the set of all stable and causal LTI models (Enqvist & Ljung, 2004). Further elaboration and proof for equation (4.11) is given in Ljung (1987).

From this the following convergence features are noted:

- Φ_{yu}/Φ_u may correspond to a non-causal function even if the true system is causal, it is thus essential to assure the causal part is taken. The condition of non-causality is realised under feedback conditions.
- THE OE-LTI-SOE will be typically approached with the rate $1/\sqrt{N}$ and the path taken to the limit will depend on the input.

Properties of OE-LTI-SOE for Slightly Non-Linear Systems

With the OE-LTI-SOE being defined as the linear, time-invariant model to which the least-squares optimisation converges given a non-linear system, its properties are extremely pertinent in understanding linear approximations of non-linear systems and the key factors involved.

Enqvist & Ljung (2004) shows that a slightly non-linear system may be represented as a function of the linear component and the non-linear component as follows:

$$y(t) = G_l(q)u(t) + \alpha y_n(t) + c(t) \tag{4.12}$$

where $G_l(q)u(t)$ is the linear part and αy_n is the non-linear component defined as a non-linear function of $u(t)$. The term α defines the size of the non-linear component of the system. Given the assumption that the input signal is generated by filtering white, possibly non-Gaussian, noise $e(t)$ through a minimum phase filter,

$$u(t) = L_m(q)e(t)$$

the z-transform form of equation (4.11) may be expressed as

$$G_{0,OE}(z) = \frac{\Phi_{yu}(z)}{\Phi_u(z)} = \frac{\Phi_{ye}(z)}{L_m(z)R_e(0)} \tag{4.13}$$

where

$$R_e(0) = E[(e)^2]$$

Defining the minimum phase input filter by the constant c, (eg. $L_m(q,c) = (1 - cq^{-1})^2$ **for** $0 < c < 1$) we have:

$$G_{0,OE}(z,\alpha,c) = \frac{\Phi_{yu}(z,\alpha,c)}{\Phi_u(z,c)} \tag{4.14}$$

$$= G_l(z) + \alpha\frac{\Phi_{y_n e}(z,c)}{L_m(z,c)R_e(0)} \tag{4.15}$$

This description of the OE-LTI-SOE is very useful in that it allows for the an investigation of the effects of the input signal used and the extent of non-linearity of the system. Enqvist & Ljung (2004) concluded the following points through such an investigation:

- The larger α is, the size of the non-linear component of the system, the further OE-LTI-SOE will be from the linear component, $G_l(q)$, of the system. Small values of α allowed for an OE-LTI-SOE near to $G_l(q)$
- Regardless of how linear or not the system is, there is always a bounded input signal such that its OE-LTI-SOE is far from $G_l(e^{j\omega})$ for $\omega = 0$.
- The distribution of the input signal components affects the results significantly. While this is not the case for linear approximation of linear systems, when approximating a non-linear system a non-Gaussian input signal distribution can result in the OE-LTI-SOE being far form the linear component of the system.

- Gaussian input distributions reduced the sensitivity of the OE-LTI-SOE to non-linaerities.

It is clear from the above that Enqvist & Ljung (2004) reveals the importance of input signal design in system identification. This importance of using the correct input signal to curve the effects of non-linearities on the linear estimate together with condition of informative data generation through persistent input signal excitation has had much attention in literature. Chapter 5 introduces some aspects presented in literature concerning input signal design in the context of experimental design theory.

4.2.2 Validation of LTI Approximations of Non-linear Systems

Categorisation of Non-linearity Detection Tests

The following is a categorisation of the different Non-linear detection tests as per Enqvist et al. (2007):

1. Methods based on data from more than one experiment.
2. Methods based on comparisons with an estimated non-linear model.
3. Methods based on the assumption that the true system is a second order linear system
4. Methods based on the assumption that the mean-square error optimal linear approximation of the system is noncausal
5. Methods based on second order moments of the input and output signals.
6. Methods for detection of non-linearities in time series.
7. Methods based on the use of special multi-sine input signals.
8. Methods based on higher order moments, either in time or in frequency domain.

For a more extensive and detailed survey of methods for nonlinearity detection in identification see Haber & Keviczky (1999).

Enqvist et al. (2007) states that categories 1 and 2 are most probably the most common approaches to non-linear detection and linear model validation. However, these methods are not ideal for the task as they have limited applicability in that category 1 assumes the input signals are unconstrained and category 2 requires a non-linear model before hand.

Categories 3 and 4 are very much pertinent to the OE-LTI-SOE tool described in the previous section. Enqvist et al. (2007) mentions that while these methods are useful in particular instances, they are not applicable to the standard non-linear system. A Linear

system or the linear component of a system can have a higher order than two, this means that these methods would indicate false non-linearities in such condiations. Additionally causality is generally implied through these approaches which is not always the case, specifically if feedback is used.

The use of second order moments of input and output data in the 5th category, detailed in section 4.1.5, has problems distinguishing unmodelled non-linearities from measurement noise. Billings & Voon (1986) found that auto-correlation functions and cross-correlation functions of residuals does not detect all unmodelled non-linear dynamics making it relatively ineffective.

Category 6 and 7 approaches are labelled as only being able to detect non-linearities in some special cases. This means they cannot by considered as a general all purpose approach to non-linear dynamics detection.

Enqvist et al. (2007) and Billings & Zhu (1994) do however reveal that category 5 methods, correlation-based model approaches, have an advantage compared with model-comparison based methods due to the fact that it is possible to diagnose directly if an identified model is adequate or not without testing all the possible model sets. These authors further mention that it is possible to increase the diagnosis power of this method by considering higher order moments of the data. This approach is categorised as category 8.

Linear models are said to be able to describe the second order moments in a data set with input and output measurements but in general not the higher order moments. This is the underlying principle behind the fact that linear validation techniques fail in the face on non-linear dynamics even if a linear approximation is attempted. This is additionally the driving factor behind increasing the diagnosis power of correlation approaches in category 8. Higher order spectral densities or correlation functions are said to be the better of the non-linear detection/linear model validation approaches. Furthermore, an advantage with these methods is that they give only statistically significant model inadequacies. This makes it easy to control the risk of false detections. The section that follows, further details higher order correlation tests.

Higher Order Correlation Test

In section 4.1.5 it was established that correlation functions, $R_{\epsilon u}(\tau)$ and $R_{\epsilon}(\tau)$, could be used to determine whether a linear model has unmodelled linear dynamics. These correlation functions may be extended to higher order correlation functions that allow for the determination of whether a system is linear or non-linear. Billings & Voon (1986) shows that if,

$$R_{yy^2}(\tau) = \bar{E}[y(t-\tau)y^2(t)] = 0 \qquad \forall\tau \tag{4.16}$$

$R_{u^2\epsilon^2}(\tau)$	$R_{u^2\epsilon}(\tau)$	$R_{u\epsilon}(\tau)$	**Significance**
$= 0$	$= 0$	$= 0$	System model is adequate and unbiased
$\neq 0$	$= 0$	$= 0$	Inadequate noise model, even or odd terms missing
$\neq 0$	$\neq 0$	$\neq 0$	System model is inadequate or System and noise model is inadequate
$\neq 0$	$\neq 0$	$= 0$	Even powered unmodelled non-linearity in system model and possibility of inadequate noise model
$\neq 0$	$= 0$	$\neq 0$	Odd powered unmodelled non-linearity in system model and possibility of inadequate noise model

Table 4.1: Significance of higher order correlation results

where $y(t)$ is the system's response to a gaussian signal or any other signal with zero third-order moments of $u(t)$, then the system that produced $y(t)$ is linear.

In addition to this, there have been several extensions of these higher order correlation functions that allow for the determination of whether a linear estimation has unmodelled non-linear dynamics. Of the many contributions to this field of study, Billings & Zhu (1994) and Enqvist et al. (2007) have been found to be the most accessible. This work will focus on the approach presented by Billings & Voon (1986) for its pragmatic form.

Enqvist & Ljung (2004) states that correlation tests may be used to look for dependencies between $u(t)^2$ and $\epsilon(t)^2$, $u(t)^2$ and $\epsilon(t)$ in addition to the already introduced pair, $u(t)$ and $\epsilon(t)$. From these correlations, the following conditions are presented that indicate the model to be unbiased and adequate are:

$$\left.\begin{aligned} R_{u^2\epsilon^2}(\tau) &= 0 \ \forall\tau \\ R_{u^2\epsilon}(\tau) &= 0 \ \forall\tau \\ R_{u\epsilon}(\tau) &= 0 \ \forall\tau \end{aligned}\right\} \tag{4.17}$$

with table 4.1 summarising the significance of the different combinations of these conditions. The interpretations of these correlation results show that these test to not only be model adequacy tests but also tests that give insight into the type of unmodelled non-linear dynamics. This allows for further model discrimination.

These higher order correlation functions are straight forward extensions of the corre-

lation functions defined in section 2.4.1,

$$\begin{aligned} R_{u^2\epsilon^2}(\tau) &= \bar{E}[u^2(t)\epsilon^2(t-\tau)] \\ R_{u^2\epsilon}(\tau) &= \bar{E}[u^2(t)\epsilon(t-\tau)] \\ R_{u\epsilon}(\tau) &= \bar{E}[u(t)\epsilon(t-\tau)] \end{aligned} \tag{4.18}$$

Billings & Voon (1986) illustrates through simulations the importance of using these higher order correlation techniques for model validation. The lower order validation techniques have been found to give false model validations that would otherwise be detected by these higher order correlations functions.

4.3 Multivariable Model Validation

It is understood that with the scope of this work primarily focusing on regression modelling, the model validation problem becomes one regression analysis. Multivariable regression analysis is a well developed field and deals the the fact that for a system with p outputs and m inputs there can be up to $p \times m$ regressions to fit. This can make the extension of single variable validation techniques more complex or less effective (Draper & Smith, 1998).

The extension of most validation techniques to the multivariate condition is mostly a matter of notation changes. In the case of the higher order correlation functions, along with the notation change of input and output variables to column vectors, the residual variable becomes

$$\epsilon(t) = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{bmatrix}$$

allowing for the correlation matrix function. In the case of $R_{u^2\epsilon}(\tau)$ (Billings & Zhu, 1995),

$$R_{u^2\epsilon}(\tau) = \bar{E}[u^2(t)\epsilon^T(t-\tau)] = \begin{bmatrix} R_{u_1^2\epsilon_1}(\tau) & R_{u_1^2\epsilon_2}(\tau) & \cdots & R_{u_1^2\epsilon_p}(\tau) \\ R_{u_2^2\epsilon_1}(\tau) & R_{u_2^2\epsilon_2}(\tau) & \cdots & R_{u_2^2\epsilon_p}(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ R_{u_m^2\epsilon_1}(\tau) & R_{u_m^2\epsilon_2}(\tau) & \cdots & R_{u_m^2\epsilon_p}(\tau) \end{bmatrix} \tag{4.19}$$

With respect to the conditions for model adequacy inferred on these correlation matrices, they are direct extensions of the scalar conditions expressed by equation (4.17) but in matrix form.

Billings & Zhu (1995) shows how the application of linear validation tests, given by equation (4.8), to a multivariable system with p outputs and m inputs, essentially translates into $p \times m + p \times p + 2$ local tests.

In efforts to reduce this number, Billings & Zhu (1995) further introduces a global test formed from the local auto-correlations among all the submodel residuals and cross-correlations among all the inputs and submodel residuals using,

$$R_\eta(\tau) = \bar{E}[\eta(t)\eta(t-\tau)] \tag{4.20}$$
$$R_{\beta\eta}(\tau) = \bar{E}[\beta(t)\eta(t-\tau)] \tag{4.21}$$

where

$$\eta(t) = \epsilon_1(t) + \epsilon_2(t) + \ldots + \epsilon_p(t) \tag{4.22}$$
$$\beta(t) = u_1(t) + u_2(t) + \ldots + u_p(t) \tag{4.23}$$

with the model adequacy conditions for R_η and $R_{\beta\eta}$ respectively being those of the scalar R_ϵ and $R_{u\epsilon}$.

4.4 Chapter Summary

In overview of this chapter, several model validation techniques were presented with a specific concentration on cross-validation based residual analysis techniques. Residual correlation tests were shown to be able to determine whether the identified noise model or the dynamics model are the source of inaccuracy. In addition to this, literature revealed higher order residual correlation tests to improve the efforts of model discrimination, specifically when validating linear estimates of non-linear systems. A study of convergence theory of linear approximation of non-linear systems showed that certain characteristics do facilitate the identification of the best linear fit of a non-linear system. It was revealed that Gaussian input signals are most favourable in producing informative data for linear approximation generations.

CHAPTER 5

Experimental Design Theory

In the preceding chapters, specifically chapters 2 and 3, an overview of open and closed-loop system identification techniques was presented. In this chapter the pragmatic nature of system identification is addressed through the presentation of the theory behind experimental design. The concepts of informative data and signal excitation are explored, with the later sections focusing on input signal design, other key experimental variables and data preparation.

5.1 General Considerations

With the concepts of persistent excitation and informative data already introduced in chapters 2 and 3, it is at this stage appropriate to study the extension of these principles into practical experimental implications. These implications are realised through experimental design variables that together form an identification experiment or test.

Before the primary design variables are refined it is important to properly characterise the experimental problem. This involves clarifying and defining which variables are to be considered as measured system outputs and which variables are system inputs. Ljung (1987) stresses an important point in that in a system there are signals that rightly are to be considered as inputs (in the sense that they have an effect on the system), even though it is not possible or feasible to manipulate them. If such variables are measured it is still highly desirable to include them among the measured input signals even though, from an operational point of view, they are considered as measurable disturbances.

Furthermore, it is important to understand the purpose of the model being identified in the experiment. This being said, the experimental conditions need to resemble the situation for which the model is to be used, specifically when approximating a system that might be non-linear with linear model (Norton, 1986).

Let all the design variables associated with the experiment be denoted by $\mathcal{X}$. These variables, in practise, are limited by the conditions incurred by the system and the operation of such system. As much of this work has detailed, many instances require closed-loop system operation and determining whether this is required/necessary or not must be done early in the identification procedure as there are implied constraints on the design variables. Goodwin & Payne (1977) lists the typical constraints as follows:

1. Amplitude constraints on inputs, outputs or internal variables.
2. Power constraints on inputs, outputs or internal variables.
3. Total time available for the experiment.
4. Total number of samples that can be taken or analysed.
5. Maximum sample rate.
6. Availability of transducers and filters.
7. Availability of hardware and software for analysis.

5.2 Informative Experiments

Creating conditions for informative experimental data and optimum input signals is a matter that may be divided into two areas of consideration. One is concerned with higher order information of $u(t)$ such as $R_{ue}(\tau)$ and $\Phi_u(\tau)$ and the other is the shape of the input signal. Aspects concerned with the former are studied through the correlation matrix, covariance matrix and information matrix.

5.2.1 Informative Data

In section 2.6.1 it was mentioned that for an experimental data set to be informative enough to allow for model discrimination, the input signal used needs to be persistently excited to a sufficiently high order. Ljung (1987: 365) proves that this excitation order, of a quasi-stationary signal, may be defined by the correlation (also termed covariance) matrix such that

$$\bar{R}_u(n) = \begin{bmatrix} R_u(0) & R_u(0) & \cdots & R_u(n-1) \\ R_u(1) & R_u(0) & \cdots & R_u(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_u(n-1) & R_u(n-2) & \cdots & R_u(0) \end{bmatrix} \tag{5.1}$$

where the input $u(t)$ is persistently exciting of order n, if and only if $\bar{R}_u(n)$ is nonsingular. This is a direct translation of the definition of persistent excitation where a signal is persistently excited of order n, if $\Phi_u(\omega)$ is different from zero at at least n points in the interval $-\pi < \omega \leq \pi$.

It is at this point important to recall that persistent excitation of a signal is not sufficient to guarantee informative data from a closed-loop system (see section 3.1.2). Section 3.2 details the different conditions pertinent to the different closed-loop identification approaches that result in informative data. In addition to this section 3.3.3 established that literature indicates non-linear controllers or the shifting between controller parameters as important aspects of creating informative data from closed-loop experiments.

5.2.2 Asymptotic Covariance and Information Matrices

Ljung (1987: 369) states that the requirement of informative experiments still leaves considerable freedom in input signal optimisation. The pursuit for the optimum input signal requires a better objective function or optimisation problem. Much literature presents the minimisation of the asymptotic covariance matrix as an optimisation function,

$$\min_{\mathcal{X} \in X} \alpha(P_\theta(\mathcal{X})) \tag{5.2}$$

where $\alpha(P)$ is a scalar measure of how large the asymptotic covariance matrix P_θ is and X is the set of admissible designs as per the constraints on the design variables. This covariance matrix function is further used to set the condition where if P_θ satisfies

$$P_\theta \geq M^{-1} \tag{5.3}$$

then the covariance matrix is that of an unbiased estimation of $G_0(q)$. M is termed the *Fisher information matrix*, Ljung (1987: 370) shows that this information matrix may be defined as

$$M_N = \frac{1}{\kappa_0} \cdot \sum_{t=1}^{N} E[\psi(t, \theta_0)\psi^T(t, \theta_0)] \tag{5.4}$$

with the average information matrix per sample, $\bar{M}$, being

$$\bar{M}(\mathcal{X}) = \lim_{N \to \infty} \frac{1}{N} M_N = \frac{1}{\kappa_0} \sum_{t=1}^{N} \bar{E}[\psi(t, \theta_0)\psi^T(t, \theta_0)] \tag{5.5}$$

where

$$\frac{1}{\kappa_0} = \int_{-\infty}^{\infty} \frac{[f_e'(x)]^2}{f_e(x)} dx$$

It is noted that if $e(t)$ were a Gaussian innovation, $\kappa_0 = \lambda_0$

This condition, expressed by equation (5.3) and termed the *Cramer-Rao lower bound*

inequality, is not only a sufficient condition for unbiased estimates but also for estimation efficiency. From this, the idea to choose $\mathcal{X}$ so that the scalar transformations of the asymptotic covariance matrix $P_\theta(\mathcal{X})$ and the information matrix $\bar{M}(\mathcal{X})$ are minimised and maximised respectively, is established.

There has been much analysis of these matrices for both the open and closed-loop conditions. From these analyses, information is extracted concerning input properties. Ljung (1987: 371) indicates that a frequency analysis on the information matrix shows how at frequencies where there is large parametric variance (i.e. large covariance matrix) input power spectra should be more prominent to allow for informative data and the consequent larger information matrix. Goodwin & Payne (1977) presented a study on how constraints on input and output signal energy content - power spectra - affect the information content and consequently the relative performance between open-loop and closed-loop experiments. It was revealed that under constraints on input energy, open-loop experimentation is preferable. However, under output energy constraints closed-loop experimentation is preferable.

5.3 Input Design

Until now second order properties (information matrix, covariance matrix, power spectrum) and their characteristics with respect desirable input sequences have been discussed. At this point it is appropriate to present some typical input sequence types that are used to generate data for system identification along with the signal-to-noise ratio (SNR) characteristic. It must however be noted that while these are generic signals, the functions and properties mentioned in the previous section, specifically the information matrix, are shown in literature as tools to design the optimal input signal for a specific system.

5.3.1 Step Signals

Step inputs and other deterministic type signals (see figure 2.1) are widely used in the identification of linear models. Step signals may be defined as follows (Söderstrom & Stoica, 1989):

$$u(t) = \begin{cases} 0, & t < 0 \\ \sigma, & t \geq 0 \end{cases} \tag{5.6}$$

where σ is the the step size. Step signals do allow for informative data about the system regarding system gains, dead time and dominant time constant, however, revelation of non-linear dynamics via such a signal does depend on the size of the step. This means under constrained input conditions, step signals are not the most practical (Sung & Lee, 2003).

This is reflected through the covaraince function ($R_u(n)$) in that given a step size of σ, $R_u(n) = \sigma^2$ for all τ. Thus the covariance matrix (or correlation matrix) $\bar{R}_u(n)$ is non-singular if and only if $n = 1$, which means a step input is persistently exciting of order 1. Such a low order of excitation means that the consistent estimation of more than one parameter in a noisy environment cannot be achieved with a single step input. (Söderstrom & Stoica, 1989).

5.3.2 Gaussian White Noise

Gaussian white noise signals are white noise signals with Gaussian distributions. These signals are understood to have interesting consequences when the system being identified is non-linear. As was discussed in section 4.2.1, if a signal is non-Gaussian then it is likely that a linear approximation of a non-linear system will deviate from the best linear approximation of this system. That is, Gaussian signals disturb the system so as to allow for data that is more likely to provide better linear estimates. On the other hand, Gaussian signals are understood as being poor at identifying non-linearities as the signal operates about the origin and hence the full operating range is not used (Nowak & Veen, 1994).

This means that Gaussian signals are favourable for linear modelling, even if the system is non-linear, and non-Gaussian signals are favourable for non-linear modelling. Figure 5.1 illustrates a white Gaussian signal. The covariance matrix of a Gaussian white noise sequence of variance λ_0 is $\bar{R}_u(n) = \lambda_0 I_n$ with I_n being an identity matrix of size n. Since this matrix is invertible for all n, Gaussian white noise sequences are persistently exciting of all orders (Söderstrom & Stoica, 1989: 121).

5.3.3 Pseudorandom Sequences

Binary

In the presence of constraints, the use of bounded (or amplitude constrained) signals that have desirable second order properties that allow for informative data is important. In satisfying this condition it is often better to let the input signal be *binary*:

$$u(t) = u_1 \quad \textbf{or} \quad u_2$$

where u_1 and u_2 are permissible input levels.

Two binary signal forms have been widely used throughout identification experiments. One is the random binary signal (RBS) and the other is the pseudorandom binary signal (PRBS). A pseudorandom binary signal is typically generated by a random sequence of length Ps and repeating this sequence as many times as required, as figure 5.2 illustrates (Söderstrom & Stoica, 1989). These sequences are deterministic signals that are designed

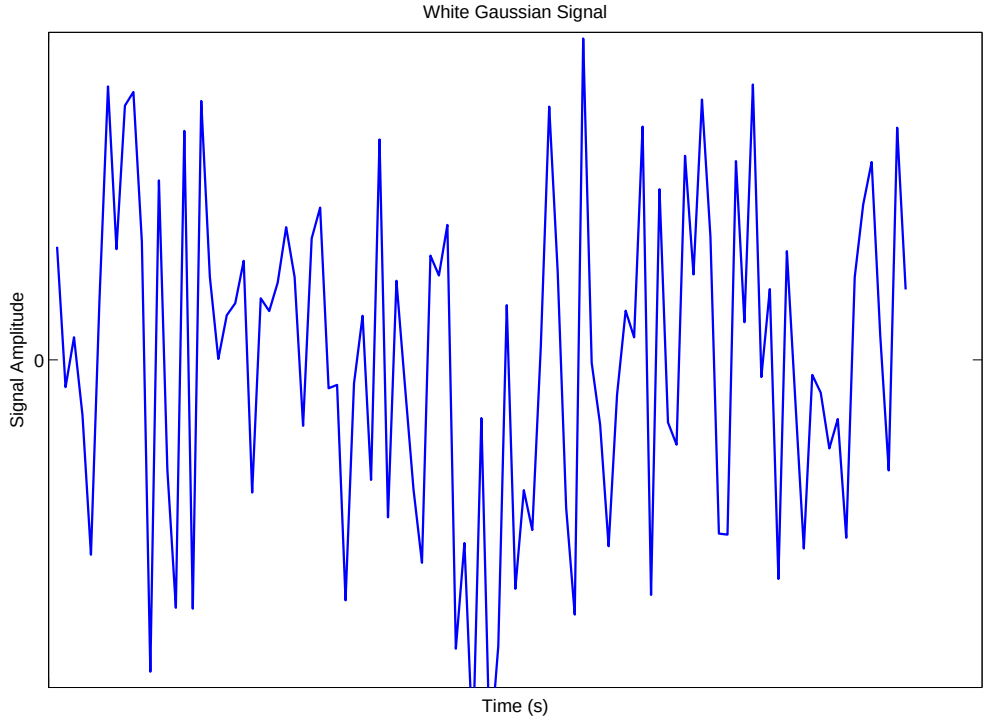


Figure 5.1: A white Gaussian signal

to have the same covariance function, independent of signal strength, as white noise, see figure 5.2(b) (horizontal lines are 95 percent confidence intervals), it can be shown that a covariance function for a PRBS of period Ps is given by

$$R_u(\tau) = \begin{cases} \sigma^2, & \tau = 0, \pm Ps, \pm 2Ps \ldots \\ -\frac{\sigma^2}{Ps}, & \textbf{otherwise} \end{cases} \tag{5.7}$$

since $R_u(\tau) = R_u(1)$ for all $\tau \neq 0, \pm Ps, \pm 2Ps \ldots$, the covariance matrix for a PRBS is given by

$$\bar{R}_u = \begin{bmatrix} \sigma^2 & -\frac{\sigma^2}{Ps} & \cdots & -\frac{\sigma^2}{Ps} \\ -\frac{\sigma^2}{Ps} & \sigma^2 & \cdots & -\frac{\sigma^2}{Ps} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\sigma^2}{Ps} & -\frac{\sigma^2}{Ps} & \cdots & \sigma^2 \end{bmatrix} \tag{5.8}$$

(Söderstrom & Stoica, 1989) additionally reveal that optimum time between each sequence switch is about 20% of the largest system response time and that the best length of signal perturbation period is several times the largest time constant.

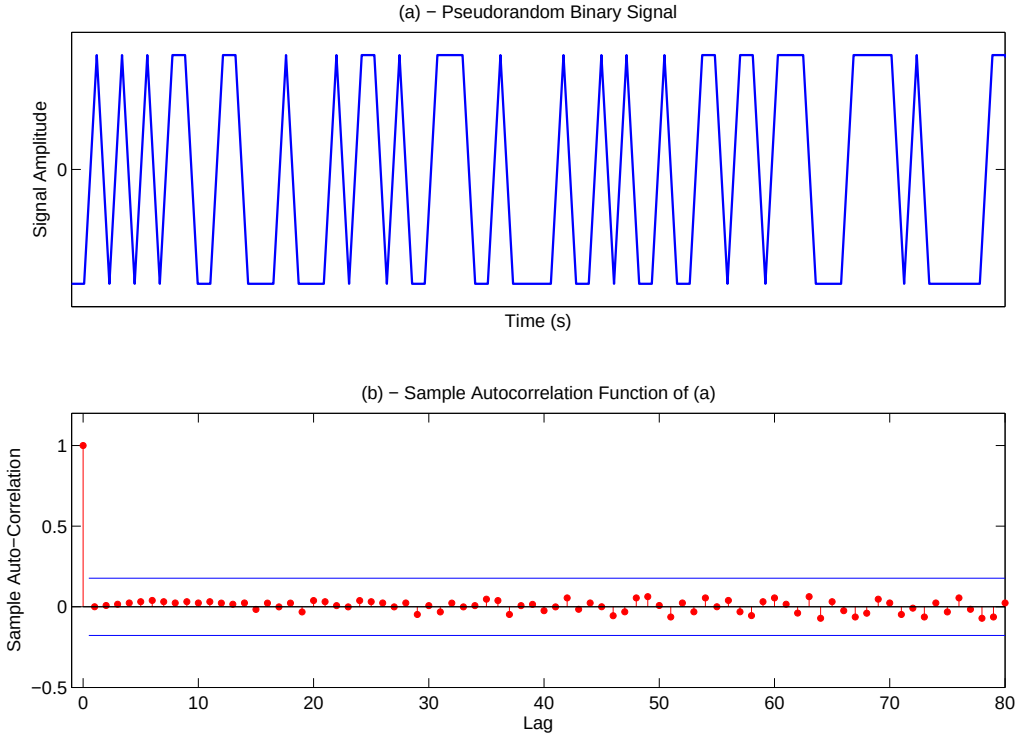


Figure 5.2: A pseudorandom binary signal

Multilevel

Pseudorandnom binary sequences have been found to be persistently exciting for linear systems, when used on non-linear systems, the extent of persistent excitation is less (Toker & Emara-Shabaik, 2004). Pseudorandom Multilevel Sequences (PRMLS), illustrated in figure 5.3, are similar to PRBSs except that the signal magnitudes randomly vary between binary signal magnitudes. The second order and higher order properties of PRMLSs approach those of white noise. Figure 5.3(b) illustrates the covariance function, compared to that of the PBRS in figure 5.2(b), it is noted that the 95 percent confidence intervals are not completely satisfied. Toker & Emara-Shabaik (2004) does however show that through determination of the number of different levels used and the signal switching probability, the closeness of the properties of PRMLSs to white noise can be improved. These signals have been found to be more persistently exciting for non-linear systems than PRBSs.

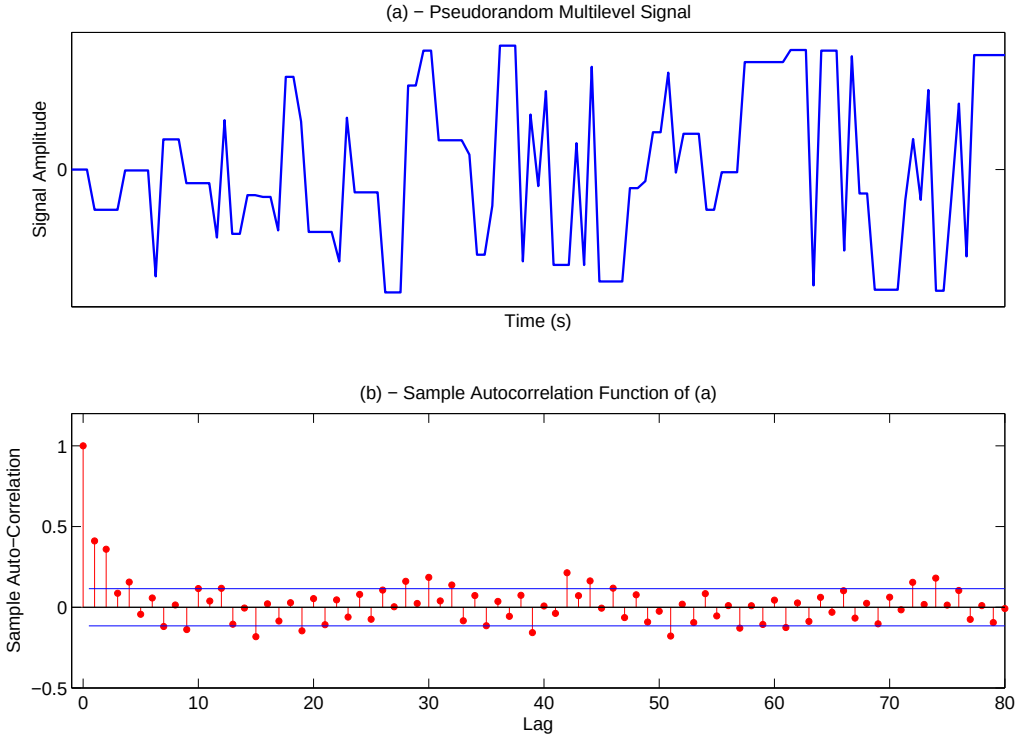


Figure 5.3: A pseudorandom multilevel signal

5.4 Other Design Variables

5.4.1 Model Structure and Order Selection

Generally there are two approaches to model structure and order selection. The first is based on analysing preliminary data for indications of the true structure of the system. The second is a comparative approach where different structures are used and compared through validation techniques. The former, also termed data-aided model structure selection, appears to be an underdeveloped field. However, the importance of a priori knowledge of the system is always useful in facilitating the determination of which structure is most representative (Ljung, 1987: 413).

An exception to this underdevelopment of data-aided selection approaches is model order estimation. There have been several developments in different techniques to systematically select model orders. Literature specifically does this with emphasis on pole-zero cancellation under conditions of over-specifying model orders. The following are categories for order selection methods (Ljung, 1987: 413):

1. Examination of spectral analysis estimates of the transfer function.
2. Testing the rank in sample covariance matrices.
3. Correlating variables

4. Examining the information matrix.

5.4.2 Bias Considerations and Sampling Intervals

Recalling section 2.6.1, specifically equation (2.63), and assuming a monic or fixed noise model we have,

$$\hat{\theta}_N \to D_c = \arg \min_{\theta \in D_{\mathcal{M}}} \frac{1}{4\pi} \int_{-\pi}^{\pi} |G_0(e^{j\omega}) - G(e^{j\omega}, \theta)|^2 Q(\omega, \theta) d\omega \tag{5.9}$$

where

$$Q(\omega, \theta) = \frac{\Phi_u(\omega)}{|H(e^{j\omega}, \theta)|^2}$$

As was previously established, Q may be considered as a weighting function that would allow for the correction of model bias by selecting the following (Ljung, 1987: 351):

- Input Spectrum, $\Phi_u(\omega)$
- Noise model, $H_1(q)$
- Prefilter, $L(q)$
- Prediction horizon, k

where $H_1(q)$ is a component of the noise model that is usually fixed to give the noise model a frequency weighting.

The sampling interval is an additional variable that also contributes to the bias of the model. Before this model bias due to sampling interval is presented, it is valuable to introduce the role of the sampling interval.

Due to the discretised nature of data-acquisition systems there is an inevitable loss of data associated with sampling intervals. This makes it important to select a sampling interval, T, that reduces or makes information loss insignificant (Norton, 1986).

If the total experiment time, $0 \leq t \leq T_N$, is limited and the acquisition of data within this time is costless, it is clearly advantageous from an theoretical point of view to sample as fast as possible. However, there are aspects that are undesirable when doing so. Firstly, creating a model based on data obtained using a high sample rate can become computationally intensive, secondly, model fitting may concentrate on the high-frequency band. The former problem is becoming less and less relevant with ever increasing availability of computing power. The latter problem may be dealt with by determining the best sampling interval relative to the bandwidth over which the concerned dynamics needs to be captured. This however is not the typical approach. This problem is usually dealt with by prefiltering the data so as to redistribute the bias (Ljung, 1987: 378).

(Ljung, 1987) shows where the T-dependence enters in parameter convergence theory through equation 5.9. As sample interval, T, tends to zero, the frequency range over which the fit is made in equation (5.9) increases. However, system dynamics are normally such that $G_0(e^{j\omega}) - G(e^{j\omega}, \theta)$ is well damped for high frequencies so that the contribution from higher values of ω in equation (5.9) will be insignificant even if the input is has a broad bandwidth.

An important exception is the case where the noise model is coupled with the system model. Such is the case of the ARX structure where $H(e^{j\omega T}) = 1/A(e^{j\omega T})$. Thus the product

$$\frac{|G_0(e^{j\omega}) - G(e^{j\omega}, \theta)|^2}{|H(e^{j\omega T}, \theta)|^2}$$

does not tend to zero as ω increases and the fit in equation (5.9) is pushed into very high frequency bands as T decreases. This means small sample intervals are undesirable as in effect it results in putting a higher weighting on higher frequency prediction errors and in turn providing a model that is biased to higher frequency dynamics (Ljung, 1987: 383).

This effect can be countered in two ways, firstly the effects in the bias may be counteracted by a pre-filter that essentially functions as a frequency counter-weight to that created by the short sampling interval. Secondly the prediction horizon in the prediction error calculation may be increased.

5.5 Closed-Loop Experiments

In a similar fashion as open-loop experiments, covariance matrices and information matrices may be derived for the closed-loop case where the reference signal is perturbed and optimised. Goodwin & Payne (1977) does make it clear that the only condition where closed-loop experimentation allows for larger information matrices is under output energy constraints. Additionally it is established that white setpoint perturbations of variance λ are the optimal inputs for experiments with feedback where the λ may be derived from an information matrix analysis.

In addition to this, as illustrated in figure 3.1 and explained in section 3.1.1, besides the setpoint signal, r_2, there is the additional signal r_1 termed the dither or excitation signal. Perturbing the system via the dither signal is typically the preferred approach as it disturbs the system directly via the input and not through the controller. This means that the perturbation is superimposed on the controller efforts to maintain the setpoint. This implies to some extent the identification process is occurring in the background of the control system (Norton, 1986).

Furthermore, as has already been mentioned, designing a closed-loop experiment that is formed by sub-experiments where the feedback law is varied improves the information content of the experiment.

5.6 Multivariable Inputs

There are essentially three basic approaches to multivariable testing and the signals involved (McLellan, 2005),

- Perturb inputs sequentially and estimate SISO models (SISO)
- Perturb all inputs simultaneously and estimate models for a given output (MISO)
- Perturb all inputs simultaneously and estmate models for all outputs simultaneously (MIMO)

Each has advantages and disadvantages. It may generally be said that while the SISO and MISO approaches do allow for easy identification of the true structure and dynamics between individual variables, the loss of information due to not accounting for all the variable interactions results in a poorer model in comparison the MIMO approach (Zhu, 1998).

This makes automatic simultaneous multivariable signal perturbation the best option for multivariable system identification. In terms of the signals used to induce such perturbations, all those previously mentioned and all their advantages and disadvantages apply to the multivariable case. McLellan (2005) goes further to suggest that cross-correlated input signals should be used.

5.7 Data Preparation

After data has been collected via the identification experiment, it is likely that the data is not ready for immediate use in identification algorithms. Once data collection is completed it is typical to plot the data to inspect it for deficiencies that could cause problems later in the identification procedure. The following are several typical deficiencies that should be attended to (Ljung, 1987: 386):

1. High-frequency disturbances in the data record, above the frequencies of interest to the system dynamics.
2. Occasional bursts and outliers.
3. Drift and offset, low-frequency disturbances, possibly of periodic character.

This chapter presents methods and approaches on how to polish the data and avoid problems later in the identification procedure due to these deficiencies.

5.7.1 High-frequency Disturbances

As mentioned earlier when discussing sampling intervals, if a sampling interval is too small the resulting data set could contain higher energy components at bandwidths that are not of concern. A typical solution to treating data that has this characteristic is to re-sample this original data set, at a different sampling rate thus effectively increasing the sampling interval (Ljung, 1987: 386).

5.7.2 Bursts and Outliers

Isolated erroneous or highly disturbed values of measured data may have substantial effects on the resulting estimate. Protecting the estimate from such bad data is thus very necessary. A robust identification procedure, specifically a robust identification norm, does accomplish this. See Ljung (1987: 396) for a more elaborate discussion on the selection of a robust norm. It is however noted that bad data is typically easy to detect and remove or replace with some type of interpolation (Ljung, 1987: 387).

5.7.3 Slow Disturbances: Offset

Low-frequency disturbances, offsets, trends, drift and periodic (seasonal) variations are typically common in data and usually stem from external sources that we may or may not prefer to include in the system model. There are basically two different approaches to deal with these issues (Ljung, 1987: 387):

1. Removing the disturbances by explicit pre-treatment of data.

2. Letting the noise model take care of the disturbances.

The first approach involves removing trends and offsets by direct subtraction while the second relies on accurate noise models. These two approaches will now be discussed with respect to the offset problem.

The offset problem is best understood as the result of estimation procedure constraint in modelling the static behaviour of a system. This means that it is typically necessary to specify how the estimation algorithm is going to treat the initial state of data.

Consider the standard linear model:

$$A(q)y(t) = B(q)u(t) + v(t) \tag{5.10}$$

This model describes both the dynamics properties of the system and the static relationships between the steady state input, $\bar{u}$, and the steady state output, $\bar{y}$, given by

$$A(1)\bar{y} = B(1)\bar{u} \tag{5.11}$$

In practice, the raw input-output data used for identification are the measurements, $u^m(t)$ and $y^m(t)$. These measurements are recorded in physical units, the values of which may be quite arbitrary. This being said, the constraint on equation (5.10), and the estimation algorithm, to satisfy the fitting of the arbitrary values in equation (5.11) is unnecessary. There are five ways to resolve this problem (Ljung, 1987: 387):

Let $u(t)$ and $y(t)$ be deviations from equilibrium: The natural approach is to determine the steady state values at the desired operating point and define :

$$y(t) = y^m(t) - \bar{y} \tag{5.12}$$
$$u(t) = u^m(t) - \bar{u} \tag{5.13}$$

as the deviations from the steady state. This approach emphasises the physical values in the dynamic models as linearisations around an equilibrium. Furthermore this approach allows for the definition of the arbitrary physical values while not imposing a constraint on the estimation procedure.

Subtract sample means: This approach is closely related to the first approach. Here we define

$$\bar{y} = \frac{1}{N}\sum_{t=1}^{N} y^m(t), \qquad \bar{u} = \frac{1}{N}\sum_{t=1}^{N} u^m(t) \tag{5.14}$$

and use equation (5.12). If an input is used that varies around $\bar{u}$, and the resulting output varies around $\bar{y}$, the system is likely to be near an equilibrium point.

Estimate the offset explicitly: One could also model the system using variables in the original physical units and add a constant that corrects the offsets,

$$A(q)y^m(t) = B(q)u^m(t) + \alpha + v(t) \tag{5.15}$$

Comparing equation(5.10) to equation (5.12) we find that α corresponds to $A(1)\bar{y} - B(1)\bar{u}$. The value α is then included in the parameter vector θ and estimated from data.

Using a noise model with integration In equation (5.15) the constant α could be viewed as a constant disturbance,

$$\frac{\alpha}{1-q^{-1}}\delta(t)$$

where $\delta(t)$ is the unit pulse at time zero. The resulting model is thus

$$y^m(t) = \frac{B(q)}{A(q)}u^m(t) + \frac{1}{(1-q^{-1})A(q)}w(t) \tag{5.16}$$

where $w(t)$ is the combined noise source $\alpha\delta(t) + v(t) - v(t-1)$. The offset α can thus be described by changing the noise model from $1/A(q)$ to $1/[(1-q^{-1})A(q)]$. this is equivalient to prefiltering the data through the filter $L(q) = 1 - q^{-1}$, that is, differencing the data (Ljung, 1987: 388):

$$\begin{aligned} y_F^m(t) &= L(q)y^m(t) = y^m(t) - y^m(t-1) \\ u_F^m(t) &= L(q)u^m(t) = u^m(t) - u^m(t-1) \end{aligned} \tag{5.17}$$

Extending the noise model: It is noted that the mode described in equation (5.16) becomes a special case of the model described by equation (5.10) if the polynomial orders of A and B are increased by 1. A common factor can then be included in A and B which implies a higher-order model, when applied to the raw measured data, y^m and u^m, will converge to a model like in equation (5.16).

Of these several approaches to dealing with offset, the first one is recommended. In conditions where a steady state condition in the experiment in not feasible, then the second approach of explicitly estimating the offset is recommended (Ljung, 1987: 389).

It is important to note that in the case of non-linear model fitting it is generally better not to remove offsets.

5.7.4 Slow Disturbances: Drift, Trends and Seasonal Variations

Coping with the other slow disturbances such as drifts, trends and seasonal variations, are very much analogous to the approaches taken for the offset slow disturbance. Drifts and trends may be interpreted as time varying steady states, or time varying means. High pass filters have also been used to compensate for these conditions. As for the seasonal variations, Box & Jenkins (1970) have developed several useful techniques.

5.8 Chapter Summary

This chapter presented a concise account of system identification experiment design and the involved variables. The more important variable was found to be the disturbance signal used to persistently excite the system so as to generate informative data. Step signals were found to be least informative while the PRBS and the White Gaussian signals were found to be most exciting with the White Gaussian signal being more favourable for non-linear information extraction. The importance of bias manipulation via input signal properties and noise model frequency weighting was established. The data sampling rate was found to be of particular value in manipulating the frequency weighting of the estimation procedures especially when using an ARX model structure.

CHAPTER 6

Identification of Simulated Systems

This chapter is the first of two chapters establishing the investigative experimental work done. In this first chapter, the experimental work presents itself through simulation. That is, open and closed-loop experiments are conducted on known mathematical models through the simulation of system disturbances (inputs) and the consequent responses (ouputs). The data generated from these experiments is then used for model identification and validation. The identification framework and validation methods used is to be defined over the next few chapters together with the objectives of these experiments. It is important to note that the results obtained from these experiments will form the basis of the experimental design efforts for the identification of the pilot scale distillation column. This is presented in the next chapter.

6.1 Investigative Approach and Model Description

6.1.1 Investigative Approach

As stated in chapter 1, the primary experimental objectives of this work are to:

- Investigate, within a specific framework, how well different identification and validation experiment designs perform.
- Specifically assess the effects of disturbance signal characteristics and feedback conditions (open/closed-loop).
- Focus the validation assessment on cross-validation methods and study how the different experimental conditions under which validation data is generated affect

the validation results. A specific focus is again made on the effects of disturbance signals and feedback conditions (open/closed-loop).

- Investigate the effect of different identification and validation experiment conditions on the linear approximation of non-linear systems.

This chapter pursues these objectives through the use of known true systems given by known mathematical models where system disturbances and responses are simulated. These true systems are those to be identified. The direct objectives of this chapter may thus be further refined as follows:

1. To obtain suitable linear and non-linear multivariate mathematical models which will be the true known systems that need to be identified. The linear and non-linear models used to represent the true systems are from here on referred to as system A and system B respectively.

2. To design a broad range of open and closed-loop experiments to be conducted on system A and system B to generate data for identification of these known systems and validation of the identified models.

3. To generate linear approximations of system A and system B using the data generated during the identification experiments.

4. To validate the linear approximations of each system using different validation data sets and techniques.

5. To assess the differences between the validation results obtained from the different validation techniques.

With the investigative approach now generally defined, it is at this point important to reiterate that identification and validation methods and approaches are generally defined by the purpose of identification, be it simulation or control (prediction). The experiments and the techniques used were carried out in such a way so as to yield information for both perspectives of identification. It is however important to recall that the ultimate model validation test with respect to control purposes is the extent of successful implementation of such a model in a model based controller. Since this is not done the conclusions reached in this work are more suited to identification for simulation.

The next sections introduce the linear and non-linear systems to be identified. That is, the mathematical models used to represent system A and system B respectively. Furthermore the identification and validation framework used is defined.

6.1.2 Linear Model - System A

System Description

The mathematical model used to represent the linear system to be identified was obtained from work by Nooraii et al. (1999). In this work the authors investigated uncertainty characterisation and robustness of a pilot scale distillation column (illustrated in figure 6.1) separating water and ethanol. Their work consequently lead to the generation of the higher order parametric ARX model to be used in this work. Nooraii et al. (1999) details that a flow-sheeting package, SPEEDUP, was used together with Matlab's identification toolbox to identify the model.

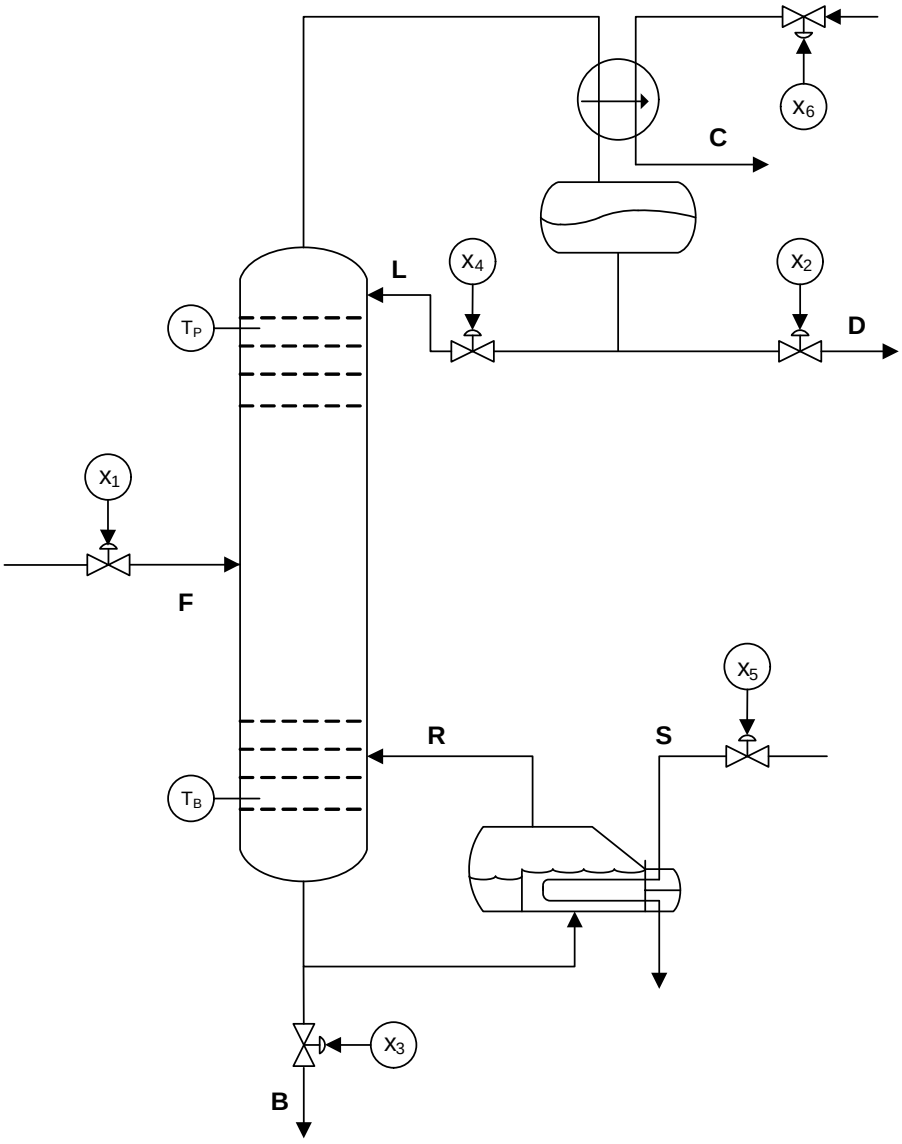


Figure 6.1: General distillation column

Referring to figure 6.1, the authors modelled the distillation column such that the multivariable ARX structure was selected with two outputs and two inputs. The outputs,

y_1 and y_2, being the temperature of plate 12 and 5 respectively (T_P and T_B on the figure) with the inputs, u_1 and u_2, being the reboiler duty (energy transfered to R by S) and the reflux flow rate (L) respectively. The authors scaled the model to be a deviation model. For experimental consistency the normal operating range of both model inputs was taken to be -0.5 to 0.5 while the output range for both open and closed-loop simulations was taken as -0.5 to 0.5.

The model generated was documented as reflecting the column's high levels of variable interaction and ill-conditioned nature. Figure 6.2 illustrates the open-loop step responses of the column while table 6.1 contains the parameters defining the model. In addition to this a system pole-zero analysis revealed that the system does contain poles out side the unit circle on the z-plane (Nooraii et al., 1999). This implies an inherent instability in the system.

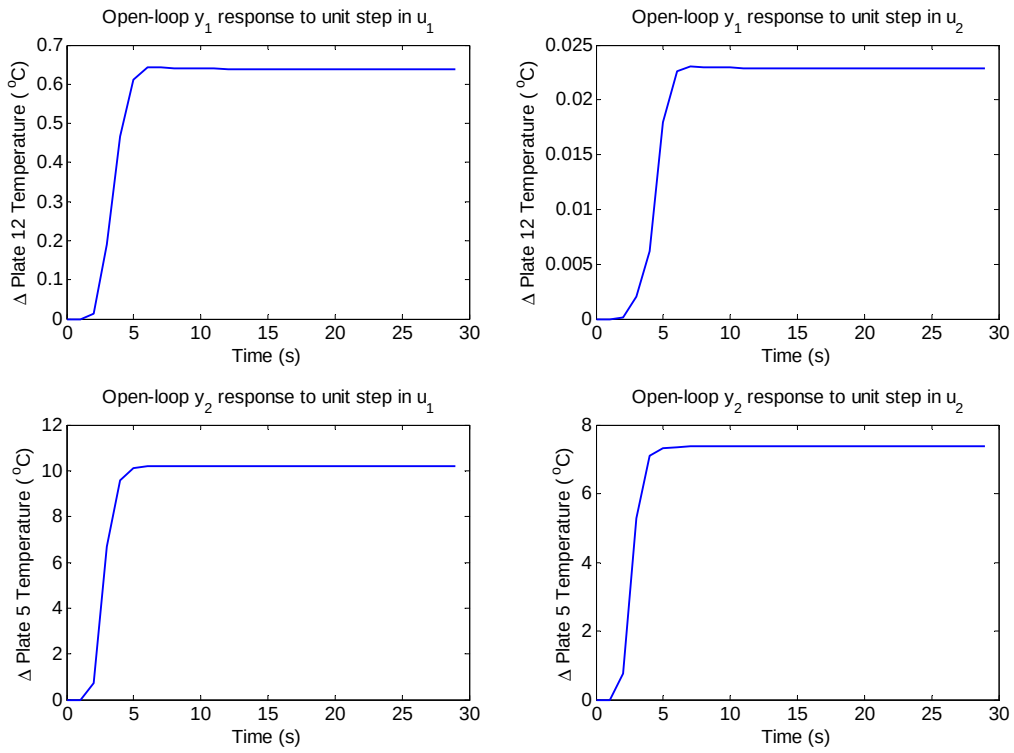


Figure 6.2: Open-loop step response of linear system - system A

System Control

The input variables, u_1 and u_2, were directly manipulated to control the model's outputs under closed-loop conditions. While it is evident that the best controller, in the face of such conditions, would be a model predictive controller, a PI controller was used. This was mostly due to its easier application and the fact that, while not the best, the PI controller was found to provide adequate control.

Coefficient	Parameter Values $\times 10^4$			
A	a_{11}	a_{12}	a_{21}	a_{22}
A_1	10000	0	0	10000
A_2	40.2	-3.26	446.8	-583
A_3	113.5	-24.8	205.9	-57
A_4	62.4	-0.305	53.05	-8.63
A_5	14.8	0	8.63	-0.56
A_6	1.33	0	0.83	-0.018
A_7	0.068	0	0.037	0
A_8	24.6	0	18.5	0
B	b_{11}	b_{12}	b_{21}	b_{22}
B_1	0	0	0	0
B_2	135	1	7309.3	7714
B_3	1765	16.65	58936	44790
B_4	2720	8.015	25766	154.57
B_5	1332	0	3544	840
B_6	301.3	0	423.3	100
B_7	11.26	0	16.04	0
B_8	1	0	1	0

Table 6.1: Model parameters for the linear ARX model representing system A

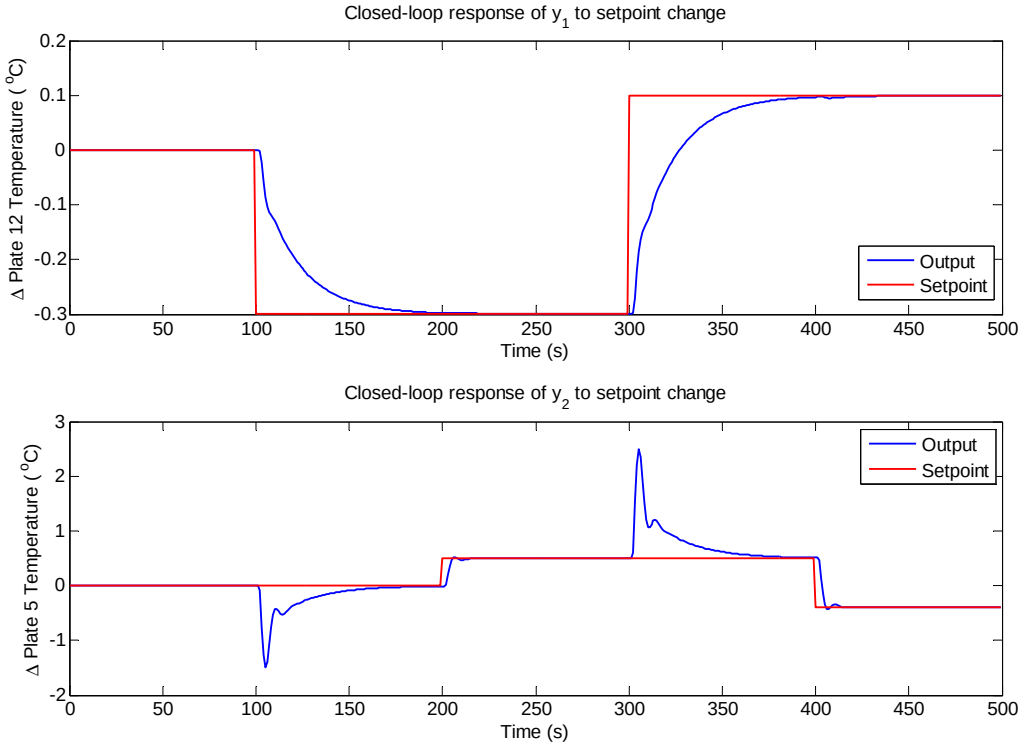


Figure 6.3: Closed-loop response to setpoint changes for linear system - system A

A controllability analysis conducted by Nooraii et al. (1999) revealed that the system has substantial levels of interaction and is ill-conditioned. The step responses given in figure 6.2 reflect the ill-conditioned state of the system in that the steady state gains are much larger for output y_2. It was found that the best variable pairing for the decentralised controller is; y_1-u_1 and y_2-u_2. The controller parameters were found by trail and error. Figure 6.3 shows the closed-loop response to setpoint changes.

6.1.3 Non-linear Model - System B

System Description

In a similar fashion as the linear system, the non-linear system was obtained from efforts by Chien (1996) to identify a non-linear model of a distillation column separating water and ethanol. The column was characterised as a high purity column and being highly non-linear.

Through open-loop experimentation on the distillation column, Chien (1996) used disturbance and disturbance response data to identify the non-linear approximation of the column. The model obtained from these efforts and used in this work is thus a multivariable non-linear ARX model (NARX) defined by:.

$$\begin{aligned} y_1(1+k) &= C_1 + C_2 y_1(k) + C_3 u_1(k) + C_4 u_2(k) + C_5 y_1(k) u_1(k) \\ &+ C_6 y_1(k) u_2(k) + C_7 y_1^2(k) u_1(k) + C_8 y_1^3(k) \\ &+ C_9 y_1^2(k-1) u_1(k) + C_{10} y_1^2(k-3) u_2(k) + C_{11} y_1^2(k-2) \end{aligned} \tag{6.1}$$

and

$$\begin{aligned} y_2(1+k) &= D_1 + D_2 y_2(k) + D_3 u_1(k) + D_4 u_2(k) + D_5 y_1(k) u_1(k) \\ &+ D_6 y_1(k) u_2(k) + D_7 y_2^2(k) u_2(k) + D_8 y_2(k) u_2^2(k) \\ &+ D_9 y_2(k-1) u_2^2(k) \end{aligned} \tag{6.2}$$

with the parameters given in table 6.2. The outputs, y_1 and y_2, were modelled as the temperatures on tray 21 and 7 respectively (T_P and T_B on the figure 6.1), while the inputs, u_1 and u_2, are the reflux flow rate and the boilup rate respectively (L and R). The model was scaled to be a model of deviation from a specific steady state.

In assessing the model and its extent of non-linearity, the size of the regions over which non-linear dynamics were small enough to be negligible were quickly identified. Figure 6.4 shows the response of the system to a 2 percent and 5 percent change in the inputs. If the system were linear there would be a direct proportionality between the change in input and the change in output. This is clearly not the case. While the changes in response to input u_1 do not so clearly indicate this deviation from direct proportionality in response changes, the differences in responses to magnitude changes in u_2 do. This is

Output 1 (y_1)		Output 2 (y_2)	
Coefficient	Parameter Values	Coefficient	Parameter Values
C_1	0	D_1	0
C_2	0.98	D_2	0.99
C_3	-0.27	D_3	-0.32
C_4	0.21	D_4	0.31
C_5	-1.91	D_5	0.76
C_6	1.44	D_6	-1.02
C_7	0	D_7	-1.22
C_8	0	D_8	0
C_9	5.96	D_9	-12.1
C_{10}	-3.71		
C_{11}	-0.11		

Table 6.2: Model parameters for non-linear ARX model representing system B

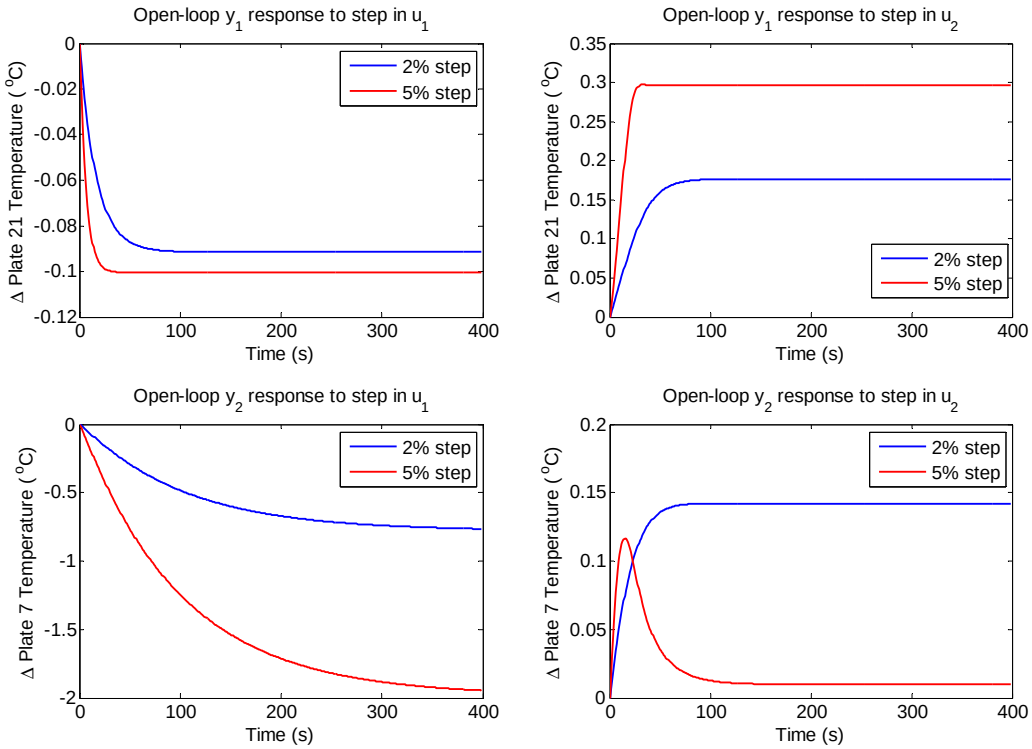


Figure 6.4: Response to 2 and 5 percent step disturbances for the non-linear system - system B

especially the case for the y_2 response, one notes that the response to a 5 % input step presents an inverse response that is not present at in the response to a 2 % input step.

This dynamic behaviour may be accounted for by the final term in equation (6.2). This term is an exponential function of input u_2, compounding this is coefficient D_9, which is negative and the largest of all coefficients.

In addition to this, while further perusing an understanding of the dynamics of the system, particularly the extent of the non-linearity, the higher order correlation test for non-linear dynamics defined by equation 4.16 was used. The result is given in figure 6.5 together with the result obtained for the linear model given white Guassian input signals. This clearly shows a deviation from linear dynamics.

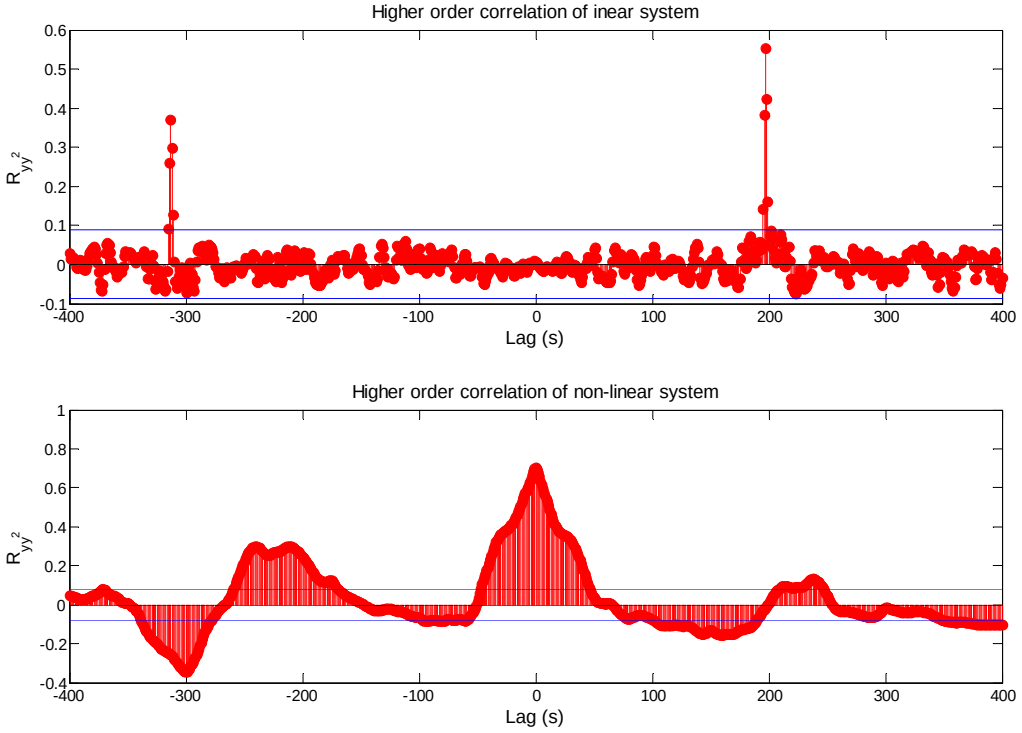


Figure 6.5: Higher order correlation test for non-linear dynamics

From this understanding of the effect of input signal magnitude on the types of dynamics exhibited and the regions and boundaries over which the system may be expected to exhibit smaller or greater extents of non-linear dynamics, it was made possible to design experiments to study the effect of input signal properties on linear model identification and validation of the non-linear system.

For experimental consistency the normal operating range of both model inputs was taken to be -0.03 to 0.03 while the normal output range for both outputs was taken as -0.03 to 0.03. In addition to these normal operating ranges over which most of the simulations took place, larger ranges were also chosen for investigative purposes, these

are; -0.06 to 0.06 and -0.05 to 0.05 for the input and output respectively.

System Control

The same approach was taken for the closed-loop simulation of the non-linear model (system B) as the linear model (system A). The closed-loop system was controlled by a decentralised controller in the form of a PI controller which was tuned by trial and error. An extensive study into controller performance given different variable pairing showed the best pairing configuration to be y_1-u_1 and y_2-u_2. Figure 6.6 shows the closed-loop response of the system.

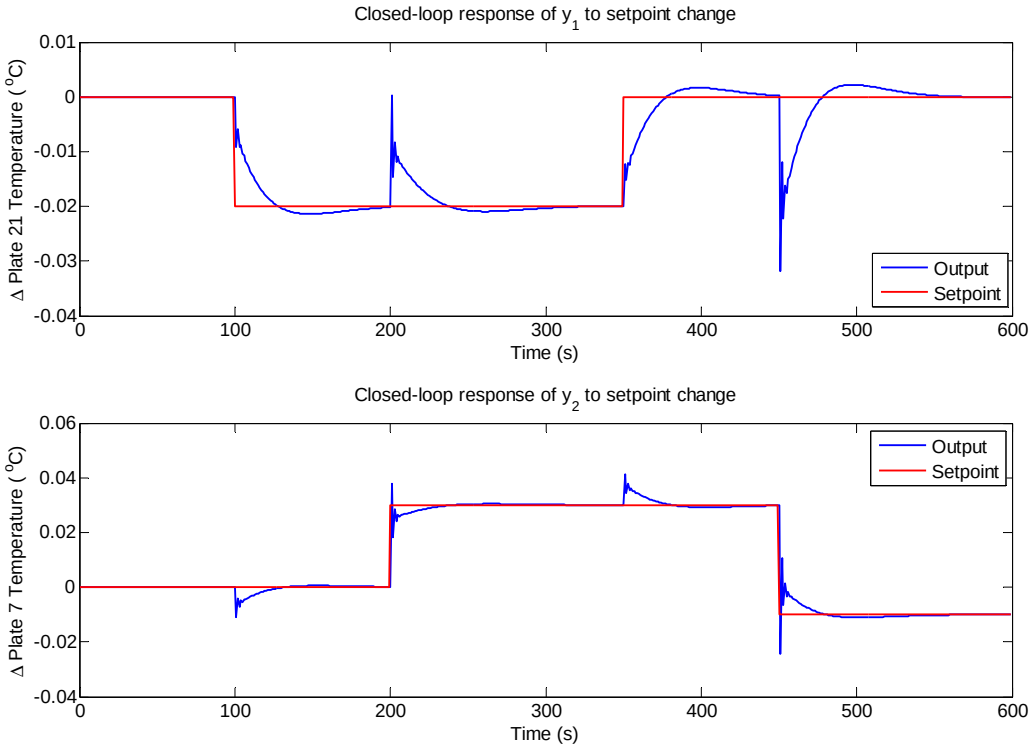


Figure 6.6: Closed-loop response to setpoint changes for non-linear system - system B

It Is noted that there is a substantial amount of controller interaction, this and the non-linear nature of the system resulted in a de-tuned controller being the only controller (not including advanced controllers) that could adequately control the system without introducing instability under different variations in setpoint changes.

6.2 Experimental Method : Framework and Design

6.2.1 General Identification Method

The identification method used to identify system A and system B was that of the prediction error method (PEM) together with an unfiltered, unweighted quadratic norm making the identification approach a least squares estimate defined by equation (2.48).

When applied to closed-loop data this was done via the direct approach. As established in section 3.2, the direct approach is the more robust closed-loop identification method, specifically since its adaptation to complex or non-linear controllers is the least involved. This, together with the consideration that modern-day controllers are ever increasing in complexity and tending towards non-linear forms such as MPC controllers, justifies the value in its investigation.

It is understood that prior knowledge of a system's controllability problems or unique characteristics, such as being ill-conditioned, should be incorporated into the identification method through proper weighting functions or norm adjustments. These adjustments are forgone in this work, however, the consequences and effects will be discussed in later sections.

6.2.2 Model Structure and Order Selection

It is understood that for an approximation of a system to converge to the true system, the model structure must contain the true system. It is further understood that there is value in studying identification cases where it is unknown whether the true structure is contained in the model structure or not. This understanding and the fact that the focus of this work is on other identification variables, lead to a selection method for model structure and order that was based on reducing complexity and calculation considerations.

As one may deduce from the choice of identification method, parametric model structures were used, specifically the multivariable auto regression with external input (ARX) structure, given by equations (2.17) through (2.24).

Although the ARX structure has the advantages of being compact and conducive to simpler parameter estimation routines, there is a disadvantage in terms of its ability to independently model the system noise from the system dynamics. Since the ARX model structure is defined by

$$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{1}{A(q)}e(t) \tag{6.3}$$

it is clear that the disturbance model, $1/A(q)$, is directly coupled to the dynamics model, $B(q)/A(q)$. This is specifically problematic for closed-loop identification since it has greater demand on an accurate noise model, due to input - noise correlations, in order to minimise model bias.

In terms of order selection and delay considerations, while the true orders (and structure) of both system A and system B are known, reduced orders were typically used for the estimates. While several orders were briefly assessed in an effort to obtain an understanding on structure selection and its effects, the selection of model order is not dealt with in depth in this work. The selection of the orders used for the estimates of both the linear system (system A) and the non-linear system (system B) was done based on a brief simulation fit analysis. The final order selections were two for system A estimates and five for system B estimates. In terms of the delay considerations, the known system delays were used and incorporated into the identification procedure.

Thus, to reiterate, the model structure used to approximate both system A and B is defined in equation (6.3), with its expanded form given as

$$\begin{aligned} &y(t) + A_1 y(t-1) + \ldots + A_{n_a} y(t-n_a) \\ &= B_1 u(t-1) + \ldots + B_{n_b} u(t-n_b) + e(t) \end{aligned} \tag{6.4}$$

where

$$A_i(q) = \begin{bmatrix} a_{i_{11}} & a_{i_{12}} \\ a_{i_{21}} & a_{i_{22}} \end{bmatrix} \tag{6.5}$$

$$B_i(q) = \begin{bmatrix} b_{i_{11}} & b_{i_{12}} \\ b_{i_{21}} & b_{i_{22}} \end{bmatrix} \tag{6.6}$$

with $n_a = 2$ and $n_b = 2$ for the system A estimates while for the system B estimates $n_a = 5$ and $n_b = 5$. Note that with both system A and B being multivariate systems of size 2×2, the correct order expression, in the case of the system A approximation, is

$$n_i = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

however, since the same order is chosen for all submodels, the scalar representation is used to represent the multivariate representation.

6.2.3 Identification Experiment Conditions

It is at this point appropriate to state that the two primary identification experiment design variables are the feedback condition, whether the experiment is operated under open or closed-loop conditions, and the signal used to disturb the system in order to generate informative data for identification and validation purposes.

Recalling figures 2.3 and 3.1, the open and closed-loop block diagrams respectively. In the case of the open-loop experiments, these disturbance signals were implemented via the input signal, $u(t)$, while for the closed-loop case the disturbance signal entered via

the dither signal, $r_1(t)$, with the setpoint signal, $r_2(t)$, being values of typical operation.

It is important to note that the open-loop experiments were designed by specifying the input signal magnitude ranges, thus open-loop experiments were operated under input constraints. The closed-loop experiments on the other hand were designed by specifying output signal magnitude ranges (setpoint ranges), thus they were operated under output constraints.

The following types of disturbance signals were used:

- Step
- White Gaussian
- PRBS
- PRMLS

with several variations of each signal in terms of switching probabilities for the pseudo-random signals. Since both the linear (system A) and non-linear (system B) systems have two inputs and two outputs, two sets of disturbance signals are required to disturb the system. For the white Gaussian, PRBS and PRMLS signals, this was done by creating two signals that are maximally shifted from each other.

In addition to varying the disturbance signal characteristics the following conditions were implemented so as to create different experimental conditions, these are mostly relevant to the closed-loop experiments:

Control Parameter Changes : Controller parameters were varied for some experiments so as to introduce a slightly non-linear controller.

Input Constraints : Constraints were implemented on the system inputs for some experiments.

Disturbance Signal Implementation : For some experiments no dither signal disturbances were implemented and for others no setpoint changes were made.

The different identification experiment cases given by the different disturbance signals, setpoint variations and controller settings used on system A (the linear system) under open and closed-loop conditions are given in table 6.3 and 6.4 respectively. While those used for identification of system B (the non-linear system) are similarly given in tables 6.6 and 6.5. Note that the case tags were created by firstly describing whether the experiment was done under open or closed-loop conditions (OL or CL) followed by the experiment number. The same is done for all experiments, with the only difference being that experiments concerned with system B (tables 6.5 and 6.6) are supplemented with the prefix "N" for the purpose of distinction, "N" signifying non-linear.

Case Tag	Input Signal ($u(t)$)		
	Signal Type	Range	B
OL 1	White Gaussian	[-0.5, 0.5]	1
OL 2	PRBS	[-0.5, 0.5]	1
OL 3	PRBS	[-0.5, 0.5]	2
OL 4	PRMLS	[-0.5, 0.5]	2
OL 5	Step	[0, 0.5]	-

Table 6.3: Open-loop experimental conditions for identification of the linear model

Case Tag	Dither Signal ($r_1(t)$)			Setpoint Signal ($r_2(t)$)
	Signal Type	Range	B	SP Values $[SP_1]$,$[SP_2]$
	SNR = 2.5			
CL 1	White Gaussian	[-0.0125, 0.0125]	1	[0.3, -0.1],[-0.5,0.4]
CL 2	PRBS	[-0.0125, 0.0125]	1	[0.3, -0.1],[-0.5, 0.4]
CL 3	PRBS	[-0.0125, 0.0125]	2	[0.3, -0.1],[-0.5, 0.4]
CL 4	PRBS NL	[-0.0125, 0.0125]	2	[0.3, -0.1],[-0.5, 0.4]
CL 5	PRBS CNL	[-0.0125, 0.0125]	2	[0.3, -0.1],[-0.5,0 .4]
CL 6	PRBS no Δ SP	[-0.0125, 0.0125]	2	[0, 0],[0, 0]
CL 7	PRMLS	[-0.0125, 0.0125]	2	[0.3, -0.1],[-0.5, 0.4]
CL 8	PRMLS sync Δ SP	[-0.0125, 0.0125]	2	[0.3, -0.1],[0.5,-0.4]
CL 9	No Dither Signal	-	-	[0.3, -0.1],[-0.5, 0.4]

Table 6.4: Closed-loop experimental conditions for identification of the linear model

Case Tag	Dither Signal ($r_1(t)$)			Setpoint Signal ($r_2(t)$)
	Signal Type	Range	B	SP Values $[SP_1]$,$[SP_2]$
	SNR = 2.5			
NCL 1	White Gaussian	[-0.00125, 0.00125]	1	[-0.02, 0],[0.03, -0.01]
NCL 2	PRBS	[-0.00125, 0.00125]	1	[-0.02, 0],[0.03, -0.01]
NCL 3	PRBS	[-0.00125, 0.00125]	10	[-0.02, 0],[0.03, -0.01]
NCL 4	PRBS L	[-0.00125, 0.00125]	10	[-0.04, 0],[0.035, -0.05]
NCL 5	PRBS NL	[-0.00125, 0.00125]	10	[-0.02, 0],[0.03, -0.01]
NCL 6	PRBS CNL	[-0.00125, 0.00125]	10	[-0.02, 0],[0.03, -0.01]
NCL 7	PRBS no Δ SP	[-0.00125, 0.00125]	10	[0, 0],[0, 0]
NCL 8	PRMLS	[-0.00125, 0.00125]	10	[-0.02, 0],[0.03 ,-0.01]
NCL 9	No Dither Signal	-	-	[-0.02, 0],[0.03, -0.01]
	SNR = 4			
NCL 10	White Gaussian	[-0.002, 0.002]	1	[-0.02, 0],[0.03, -0.01]
NCL 11	PRBS	[-0.002, 0.002]	1	[-0.02, 0],[0.03, -0.01]
NCL 12	PRBS	[-0.002, 0.002]	10	[-0.02, 0],[0.03, -0.01]
NCL 13	PRBS L	[-0.002, 0.002]	10	[-0.04, 0],[0.035, -0.05]
NCL 14	PRBS NL	[-0.002, 0.002]	10	[-0.02, 0],[0.03, -0.01]
NCL 15	PRBS CNL	[-0.002, 0.002]	10	[-0.02, 0],[0.03, -0.01]
NCL 16	PRBS no Δ SP	[-0.002, 0.002]	10	[0, 0],[0, 0]
NCL 17	PRMLS	[-0.002, 0.002]	10	[-0.02, 0],[0.03 ,-0.01]

Table 6.5: Closed-loop experimental conditions used for identification of the non-linear model

Case Tag	Input Signal ($u(t)$)		
	Signal Type	Range	B
NOL 1	White Gaussian	[-0.03,0.03]	1
NOL 2	PRBS	[-0.03,0.03]	1
NOL 3	PRBS	[-0.03,0.03]	10
NOL 4	PRBS L	[-0.06,0.06]	10
NOL 5	PRMLS	[-0.03,0.03]	10
NOL 6	Step	[-0.03,0.03]	-

Table 6.6: Open-loop experimental conditions for identification of the non-linear model

The remainder of this section further elaborates on how the identification experiments were designed and clarifies the content of the tables detailing the experiments.

The following can be said regarding the identification experiment conditions:

- Values of all variables are initially zero, thus all the input and disturbance signal magnitudes are specified as deviation values
- All inputs of both the linear (system A) and non-linear (system B) multivariate systems were disturbed simultaneously, thus making the experiments multivariable experiments.
- The amplitude of the White Gaussian signal superimposed on the output signals used to simulate output noise was taken as 1% of the full normal operating range of each system.
- The same noise realisation was used for all the identification cases unless otherwise stated (i.e. where sensitivity to different noise realisations was investigated - this is discussed in the following sections)
- The amplitude of the dither signal (r_1) superimposed on the system input for closed-loop experiments was taken as 2.5% the normal operating range in the case of system A. For system B two dither signal amplitudes were used, 2.5% and 4%. Figures 6.7(a) and 6.7(b) illustrate the differences in output excitation between the open and closed-loop condition on the linear system. The effect of the dither signal is additionally illustrated.
- With the exception of the step signal which is only defined by the range, all other

signals are defined by the type, range and average time of constant signal interval (B), given in seconds.

- Signals with $B = 1$ are white signals, the other signals have B values selected as 30% of the largest time constant. Theory dictates this approach to selecting B values as optimal for generating response data reflecting the dominant dynamics of the system (Norton, 1986).

- The identification experiment duration was taken as several times the system's closed-loop response in order to allow for several setpoint changes.

- Two setpoint changes were made to both controlled variables during closed-loop experiments. As mentioned earlier, the initial setpoint is zero. The changes made to each setpoint were done in different directions and a different times.

- In the case of system A the setpoint changes for output y_1 and y_2 were made at 100 and 300 seconds and 200 and 400 seconds respectively. In the case of system B the setpoint changes for output y_1 and y_2 were made at 100 and 350 seconds and 200 and 450 seconds respectively. Figures 6.6 and 6.3 respectively illustrate these changes for system A and system B.

The following further elaborates on the information presented in tables 6.3, 6.4, 6.6 and 6.5:

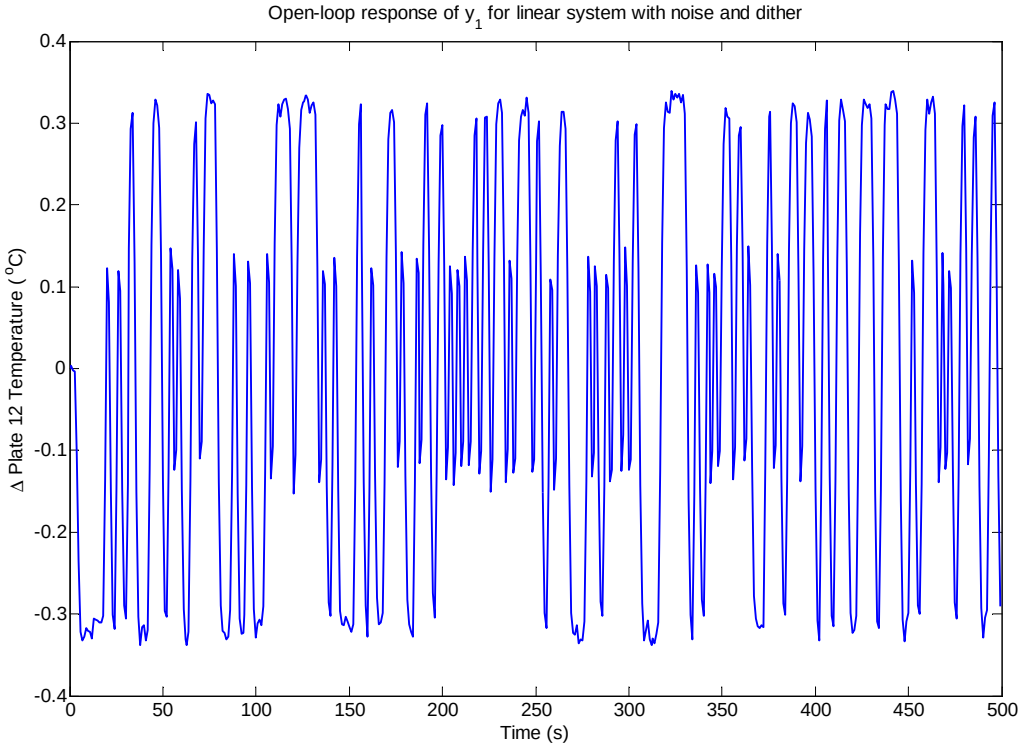
Range : Specifies the range of the input and dither signals used in the open and closed-loop experiments respectively. Since both input u_1 and u_2 are similarly scaled for both system A and B, the range of these input ($u(t)$) and the dither ($r_1(t)$) signals are only given once e.g. [-0.5, 0.5] specifies the lower and upper bound of both signal, u_1 and u_2.

SP Values : Specifies the changes in setpoint values made for each output. [SP_1] are the setpoint changes made to setpoint 1, [SP_2] are the setpoint changes made to setpoint 2.

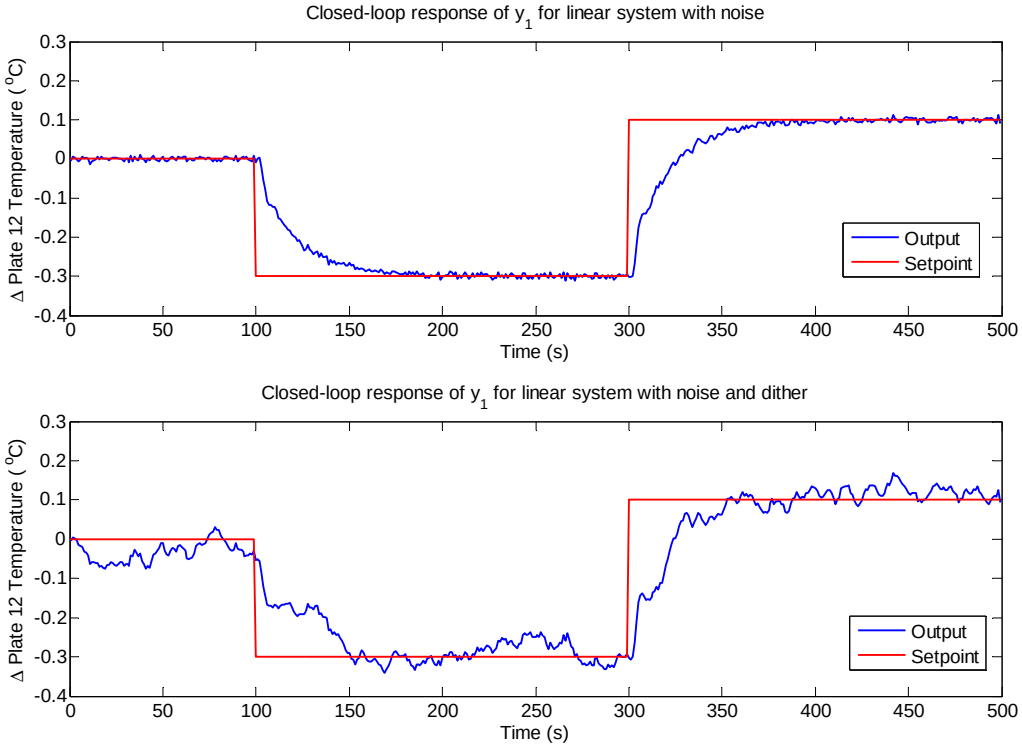
NL : Implies a non-linear controller, which was achieved by varying the controller parameters between four sets of parameters throughout the closed-loop experiment period.

CNL : Implies the same NL condition mentioned above but with the input signals constrained between specified limits.

no ΔSP : is the condition where no setpoint changes were incurred.



(a) Open-loop output excitation of system A using PRMLS input signals and superimposed noise



(b) Closed-loop output excitation of system A indicating the effect of introducing noise and PRBS dither signals

Figure 6.7: Open and closed-loop output excitation of system A

sync ΔSP : is the condition where setpoint changes were made on both controlled variables at the same time and in the same direction.

No Dither Signal : is the case where no dither signal was used, only setpoint changes.

L : For the experiments on system B, the indicator L, was assigned to label experiments which operated over a larger range than the typical experiments. This larger range was selected so as to assure operation in regions of more pronounced non-linear dynamics. The normal identification range was selected to be closer to the lower extents of non-linearity. It is important to note that when referring to a larger range for the open-loop experiment, this is specified by larger input signal magnitudes, while for the closed-loop experiment this is achieved by larger setpoint changes.

Thus 14 (5 open-loop and 9 closed-loop) experimental conditions for system A (the linear system) and the 23 (6 open loop and 17 closed-loop) for system B (the non-linear system) were used to generate 14 and 23 models for each system respectively.

As will be revealed in the discussion of the simulation results in section 6.3, the process of identification and model generation was repeated 50 times for each case for both system A and system B. The primary purpose of this was to obtain an understanding of the effects of the random realisations of noise on the models generated by specific signals. From this parameter variances and basic validation result variances were obtained.

It is however important to note that while several models where generated due to the different noise realisations and a basic assessment was made into each of these models, the models obtained by the first realisation of noise were taken as the primary models for investigation.

6.2.4 Validation Techniques and Experiment Conditions

The validation of a model may be considered as the essence of the identification procedure since it determines whether the model generated is useful or not. As established in section 4.1, there are several techniques available for model validation. The following techniques were used to assess the validity and accuracy of the models generated (further detail on the execution of these techniques is presented in the following section):

- Model Variance and Uncertainty
- Percentage Fit it for simulations and predictions
- Frequency Content Analysis
- Residual Correlations
- Higher Order Residual Correlations

In addition to these, two other techniques were implemented. These are model pole-zero stability and frequency response analysis (Bode plot). Both of these techniques were only applied in validation efforts of the linear system (system A).

From the list of validation methods being used it must be noted that most of them are cross-validation techniques. These techniques involve validating the identified model against a validation data set. Figure 6.8 illustrates the process of cross-validation in the context of generating data and identifying models.

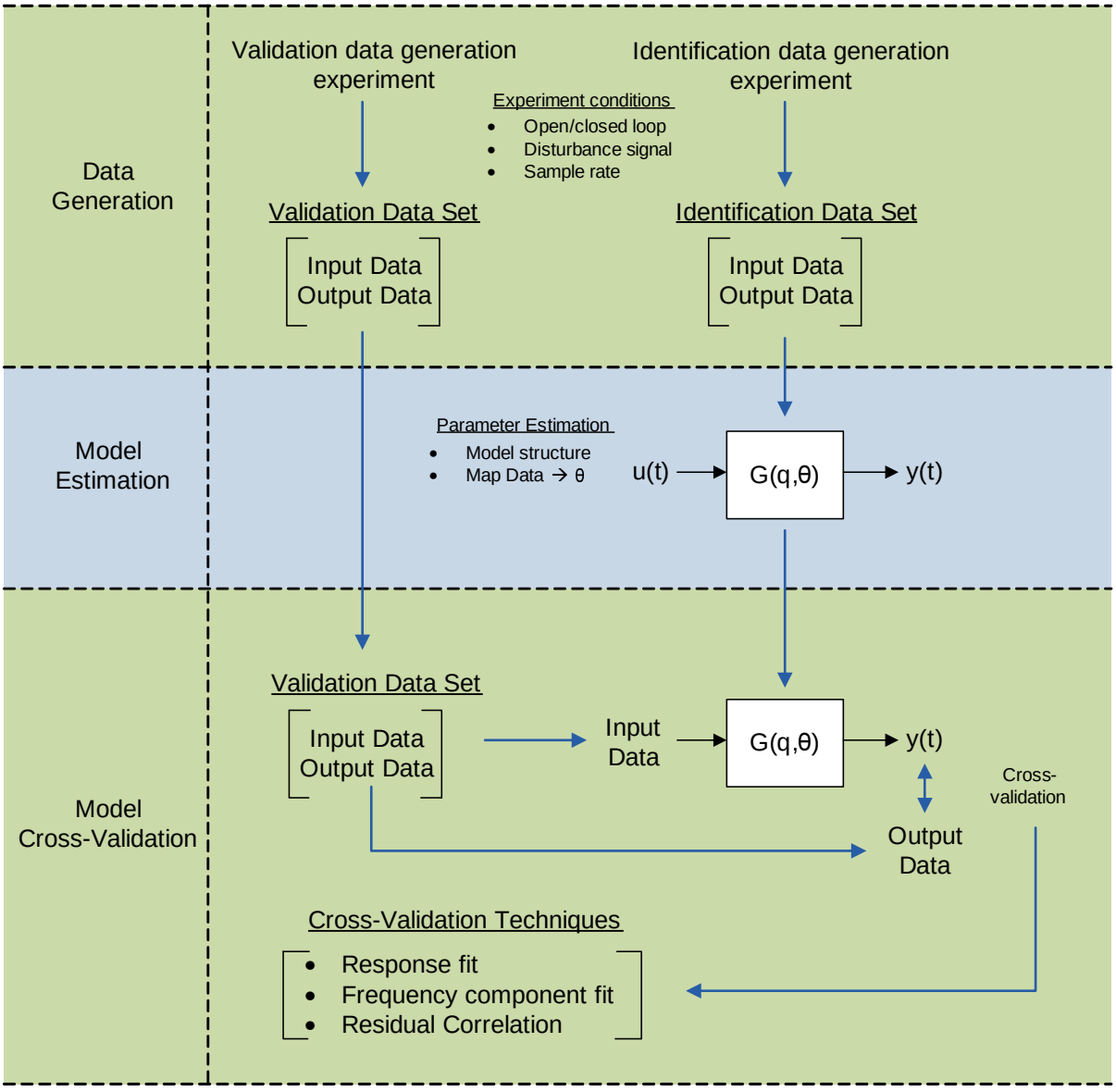


Figure 6.8: Illustration of data cross-validation in the context of data generation and model identification

Thus, given the primary focus on cross-validation techniques, attention was given to generating several validation data sets from different experimental conditions in order to obtain an understanding into the effects of experimental conditions on the accuracy of cross-validation techniques. The primary experiment variables studied in terms of their

effects were disturbance signal characteristics and feedback conditions.

Thus several validation data sets were generated from experiments using PRBS and step signals under both open and closed-loop conditions on both systems A and B. In generating validation data for the non-linear system, system B, data was generated over normal and larger operating ranges.

Tables 6.7 and 6.8 show the different experimental conditions used to create different validation data sets for the linear and non-linear system experiments. In a similar fashion as the experimental condition used to generate identification data, these are given experimental case tags. The same rules apply in that "N" is used to signify the experiment was done on system B, "OL" and "CL" indicate open-loop and closed-loop conditions respectively and following this is the experiment number. In addition to this is the phrase "Val" signifying the experiment is for generating validation data.

The following can be said about the generation of data for validation:

- Even though some validation data sets were generated from step disturbance experiments, due to the fact that research into literature clearly showed the low information content of step response data, more emphasis was put into assessing validation results from data generated by PRBS disturbance signals.

- The setpoint changes made during the closed-loop experiments were slightly different to those implemented for the identification experiments, however, in a similar fashion as identification experiments, setpoints were changed twice for each validation experiment.

- The average constant signal interval values (B) of the disturbance signals used to generate validation data were slightly larger than those of the disturbance signals used to generate data for identification. However, they were not so much larger so as to generate data with frequency characteristics outside of the frequency band of interest for modelling.

- Noise was not always superimposed on the output data of the validation data sets. This is clarified later.

- Just as with the identification data generation experiments for system B, two experimental regions were created for validation data generation. One over a smaller region of operation, and another over a larger region of operation. This was done in hope of creating sets of validation data with different dynamic contents in terms of non-linear dynamics.

From the tables describing the experimental conditions used to generate validation data (tables 6.7 and 6.8), it can be said that 3 validation data sets (2 open-loop and 1

Open-Loop Validation Data			
Case Tag	Input Signal ($u(t)$)		
	Signal Type	Range	B
ValOL 1	PRBS	[-0.5, 0.5]	2.5
ValOL 2	Step	[0, 0.5]	-

Closed-Loop Validation Data				
Case Tag	Dither Signal ($r_1(t)$)			Setpoint ($r_2(t)$)
	Signal Type	Range	B	SP Values $[SP_1]$,$[SP_2]$
ValCL 3	PRBS	[-0.0125, 0.0125]	2.5	[0.3, 0.5, -0.1],[0.5, -0.4]

Table 6.7: Conditions for generation of validation data for linear model

Open-Loop Validation Data			
Case Tag	Input Signal ($u(t)$)		
	Signal Type	Range	B
NValOL 1	PRBS	[-0.02, 0.02]	5
NValOL 2	PRBS L	[-0.05, 0.05]	5
NValOL 3	Step	[0, 0.02]	-

Closed-Loop Validation Data				
Case Tag	Dither Signal ($r_1(t)$)			Setpoint ($r_2(t)$)
	Signal Type	Range	B	SP Values $[SP_1]$,$[SP_2]$
NValCL 4	PRBS	[-0.00125, 0.00125]	5	[0.02, 0, -0.01],[-0.01, 0.01]
NValCL 5	PRBS L	[-0.00125, 0.00125]	5	[0.05, 0, 0.04],[0.04, -0.055]

Table 6.8: Conditions for generation of validation data for non-linear model

closed) were generated for system A and 5 validation data sets (3 open-loop and 2 closed-loop) were generated for system B. Thus in terms of summarising the cross-validation efforts, 14 identified models were cross-validated against 3 validation sets for system A, for sytem B, 23 models were cross-validated against 5 validation sets.

Recalling that all the identified models have ARX structures given by equation (6.3), which may be broken down into a dynamics model, $\frac{B(q)}{A(q)}$, and a noise model, $\frac{1}{A(q)}$. It is at this point appropriate to note that most of the cross-validation techniques validated only the dynamics model. That is, no noise signals were superimposed on the validation data outputs allowing for direct validation of the identified dynamics model against the true known system. The only validation techniques that assess the validity of the noise model together with the dynamics model are the residual correlation assessments.

It is acknowledged that the topic of model stability as a validation technique is important, specifically when modelling for control. This validation tool is not used to great depth in this work and is limited to the pole-zero stability comparisons between linear system, system A, and the identified linear approximations of the system.

Is is noted that model variance and uncertainty play very large roles in assessing how well a model represents a system and that ultimately, knowledge of this uncertainty should be included in all validation assessments. The approach taken by this work was such that instead of imposing variance and uncertainty measures on all the validation techniques and results, the variance and uncertainty measures were directly assessed on their own.

6.2.5 General Execution and Software

The generation of system disturbance signals and the simulated responses of both systems A and B, together with the identification and validation of models from data comprising of these disturbances and responses was all done through Matlab. The regression models representing systems A and B were simulated in discrete form, the controllers were thus discrete controllers in the velocity form. The following can be said about the functions used and generated and how they were executed:

- The function *idinput* was used to create all the disturbance signals except for the PRMLS, a function was written to generate this signal based on the *idinput* function.
- The function *ARX* was used to generate multivariable ARX models with specified orders and delays given input and output data.
- The *compare* function was used to create simulations and predictions to generate percentage fits between the identified model responses and validation data sets.

Matlab's percentage fit is defined as follows:

$$\textbf{percentage fit} = 100 \times \left(1 - \frac{|y - \hat{y}|}{|y - \bar{y}|}\right) \tag{6.7}$$

where y is the 'measured' (simulated) output, $\hat{y}$ is the output generated by the identified model given the input used to generate the 'measured' output. $\bar{y}$ is the mean of y.

- The *fft* function, a discrete Fourier transformation function, was used to convert discrete time data to discrete frequency data. This, combined with the *compare* function, allowed for the assessment of the cross-validation of the identified model response fits in the frequency domain.
- The function *frqresp* was used to extract frequency response data from linear parametric models.
- The *simsd* function was used to assess each identified model's covariance matrix and the consequent indication of model uncertainty. This is achieved through Matlab's Monte Carlo simulation engine which expresses the model's covariance matrix (obtained from the model parameter estimation procedure) as simulation response variance.
- The function *pzmap* was used to determine the poles and zeros of models and consequently their stability.
- For residual auto-correlations and input-residual cross-correlations the *corr* and *autocorr* functions where used respectively.
- A function was created called *crossco* for the higher order correlations and cross-correlations. This function was based on the higher order correlation functions found in the Neural Network based system identification toolbox (Norgaard, 2000).

See Appendix A for a complete description of the functions and programs used.

6.3 Identification and Validation Results for System A - The Linear System

This section presents and discusses the results of the identification of models approximating system A and the validation of such models. Note that all the parameters of each identified model are presented in appendix A.2. Each of the 14 identified models (5 generated from open-loop data, 9 generated from closed-loop data) and the experimental

conditions that generated the respective data used to identify them will be referred to by the concerned experimental case label. That is, the case labels given in table 6.3 for the models obtained from open-loop experiments and table 6.4 for the models obtained from closed-loop. Note however that a brief mention of the conditions characterising the experiment used to generate the model or validation set being discussed is done where necessary and relevant to the point being made.

Section 6.3.1 presents the assessment of variance and uncertainty of the identified models while section 6.3.2 presents and assessment of the identified model stability. These model validity assessments are not cross-validation techniques, they thus do not use the validation data sets generated. However, in assessing the variance and uncertainty of the identified models it was necessary to invoke responses. Normal PRBS signals confined to the normal input ranges were used as inputs to produce such responses.

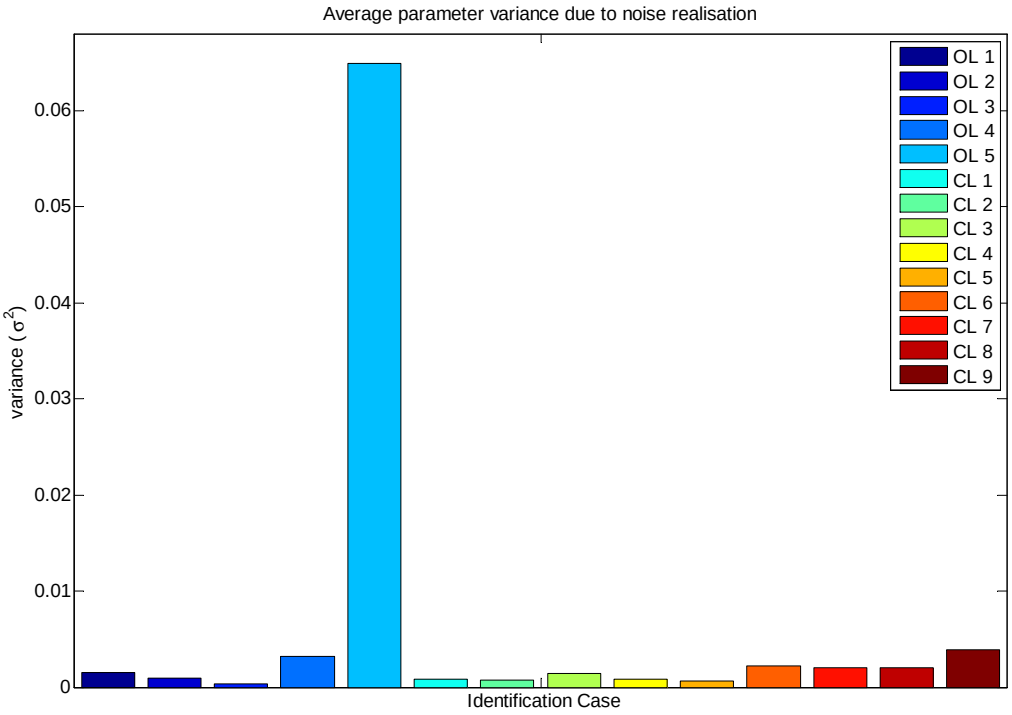
Sections 6.3.3 through 6.3.5 primarily focus on the validation results and findings based on cross-validation techniques as presented in figure 6.8. These techniques are based on comparing simulated responses from the identified models to responses from the true known system, system A. This is done using the validation data sets generated. These are given in table 6.7, recall that there are two sets generated under open-loop conditions (ValOL 1 and 2) and one under closed-loop conditions (ValCL 3). Each validation set contains input and output data. The inputs are used to disturb the identified model so as to allow for a simulated output response which is to be compared to the response contained in the validation data set. Thus effectively comparing the simulated response accuracy of the identified models.

6.3.1 Model Variance and Uncertainty

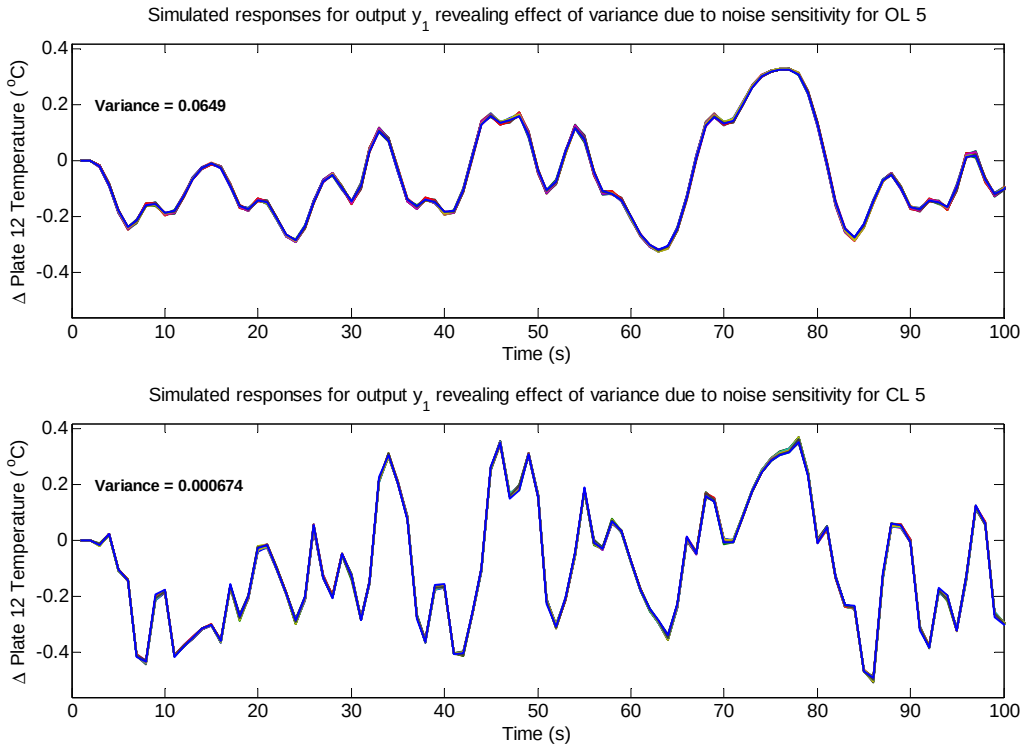
The assessment of model variance and uncertainty was done on two levels. The first being model parameter sensitivity to random noise realisations in identification data. The second being parameter uncertainty as per the covariance matrix obtained from the parameter estimation procedure expressed via simulated response variances (Monte Carlo simulations).

Variance Due to Sensitivities to Random Noise Realisations

The investigation into experimental design sensitivity to noise realisations was done by repeating all the identification and consequent model estimation experiments 50 times. The number of repetitions was selected based on the observation that using fewer than 50 repetitions did not produce consistent results. The execution conditions were such that all variables, besides the random realisation of the White Guassian noise superimposed on the output signals, were kept constant. The resulting model parameters of the models identified from these repeated experiments were then compared to each other to



(a) Parameter variance due to random noise realisations



(b) Model variance shown through simulation response variance

Figure 6.9: Parameter and response variance due to random noise realisations

assessed for variance. The average variance of all the parameters was used as the variance measurement. Figure 6.9(a) illustrates the findings.

It is clear from this figure that case OL 5, defined as an open-loop step disturbance experiment, results in a significantly larger parameter variance due to random noise realisations. This could reflect the experiment's inability to produce informative data which consequently resulted in the estimation procedure modelling the random noise realisations. While the significance of the differences in parameter variance measurements between the other cases might be questionable, since they are relatively small, it must be noted that, of all the other experimental cases, CL 9 was found to have the largest variance. Note that CL 9 is the closed-loop identification experiment where no dither signal was used.

It is interesting to note that while there is significant distinction when comparing the parameter variance due to noise experienced by case OL 5 to that of the other cases, such distinction is not observed in the output responses. This is indicated in figure 6.9(b). Here the 50 responses of output y_1 to the same input (PRBS), generated by each of the 50 models identified from experiments OL 5 and CL 5 given the 50 different noise realisations, are plotted together. Note that cases OL 5 and CL 5 are the cases with the largest and smallest parameter variances respectively, as per figure 6.9(a). The difference in observable variance in response y_1 is relatively insignificant. The same was found for output y_2. This does imply that while the parameter variance due to noise sensitivity in case OL 5 is clearly larger than the others, it is still an insignificant amount of variance.

Model Uncertainty

The response variance of each identified model due to parameter uncertainty was measured by standard deviations obtained from Monte Carlo simulations. Using the covariance matrices generated during the estimation procedure of each identified model, the simulations produced 50 random response simulations to a PRBS input signal reflecting the parameter uncertainties. From these simulated responses, standard deviations at each sample point (i.e. at each time measure) were obtained. Figure 6.10(a) illustrates the 50 response simulations of output y_1 for the model identified from experiment CL 9 and figure 6.10(b) shows those of output y_2 for the same case. Figure 6.11 shows the averaged response standard deviation values over time of each identified model.

From these averaged response standard deviations, it is evident that there is no experimental case that produced a model with significantly larger extents of uncertainty than the others. No clear trend may be observed besides that the model identified from the open-loop step response experiments, OL 5, has the lowest average response standard deviations due to parameter uncertainty for both outputs.

The results for output y_1 are particularly less distinguishable in comparison to those

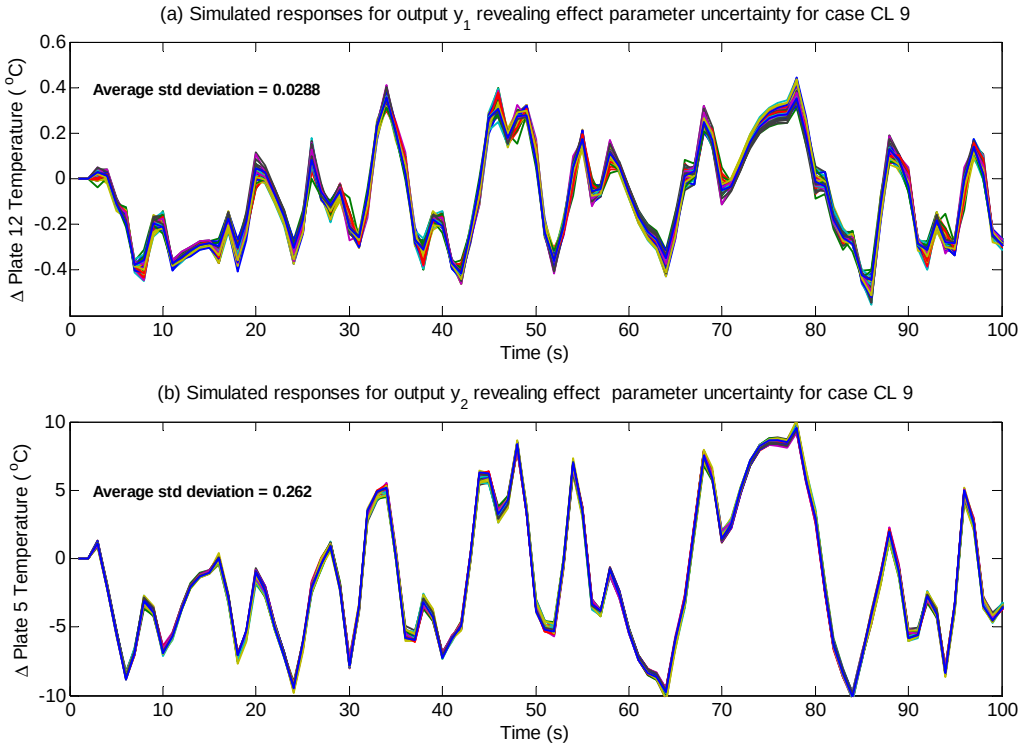


Figure 6.10: Monte Carlo Simulation for case CL 6

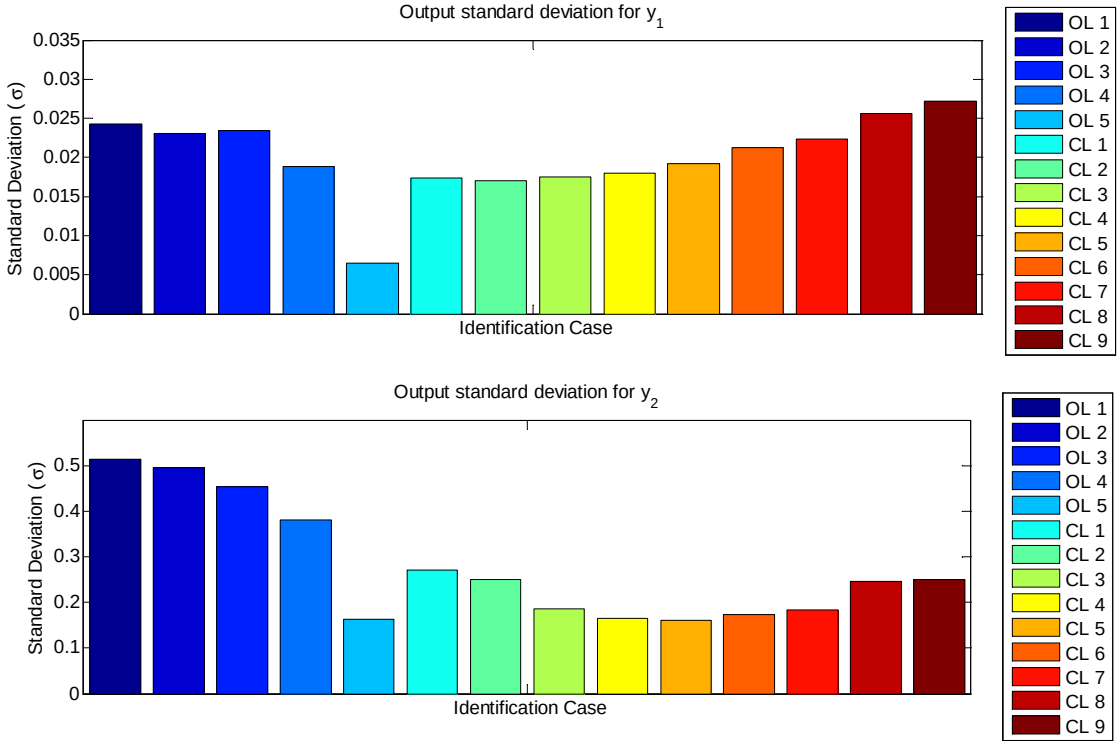


Figure 6.11: Average output response standard deviations due to parameter uncertainty

of y_2. The standard deviation values for the models obtained from open-loop experiments (the OL cases) show larger values for output y_2 relative to those of models obtained from closed-loop experiments. Looking at the results for output y_1, it may however be observed that the models generated by closed-loop, cases CL 7-9, revealed slightly larger extents of model uncertainty than most of the other closed-loop cases. Cases CL 7-9 being the experiments defined by PRMLS dither signals (CL 7-8) and the lack of a dither signal (CL 9).

Note that the standard deviation values for output y_2 are larger than those of y_1. This may be justified by the larger system gains between the system inputs and output y_2.

6.3.2 Model Stability

The brief assessment made into model stability was based on the understanding that the true system is characterised as containing unstable poles.

An assessment into the system poles and zeros of each identified model revealed that the model obtained from case OL 5, the open-loop step response experiment, was the only model with no pole situated outside the unit circle on the z-plane. This implies that, unlike the true system which had several poles outside the unit circle making it an unstable system, the model obtained from case OL 5 is stable. This does reflect the experimental design's inability to obtain informative data on the system and furthermore implies that its incorporation into a model based controller will not only result in poor control but in an unstable controller.

All the other experiments generated models which had unstable poles. However, as mentioned earlier, since the structure of the identified models is different to that of true system, the unstable poles were not located in the same regions as those of the true system. Given the multivariable nature of the system, this implies that the identified models will be unstable in different directions in comparison to the true system.

6.3.3 Simulation and Prediction

This section introduces the first and primary component of the cross-validation techniques. Here percentage fit values are used to validate the accuracy of each identified model's simulated response in comparison to the true system's response given by the validation set. This is done for all 14 models against all 3 validation sets. Since all the models were generated under different experimental conditions and all the validation data sets were generated under different experimental conditions, the cross-validation results will allow for insight into experimental effects on model accuracy and validation results.

Figures 6.12 and 6.13 show the response percentage fit values for each model validated against the open and closed-loop validation sets respectively. Figure 6.14 shows an example of simulated responses using validation set ValOL 1, which is characterised

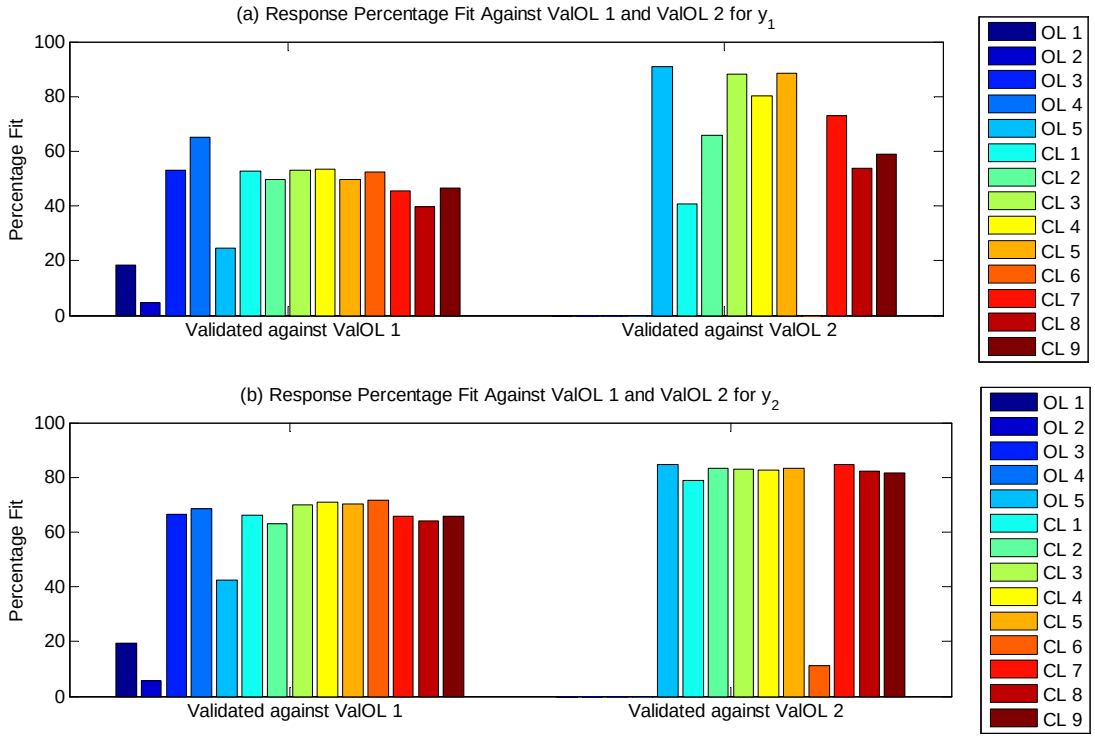


Figure 6.12: Percentage fit values for simulation validation against open-loop validation sets

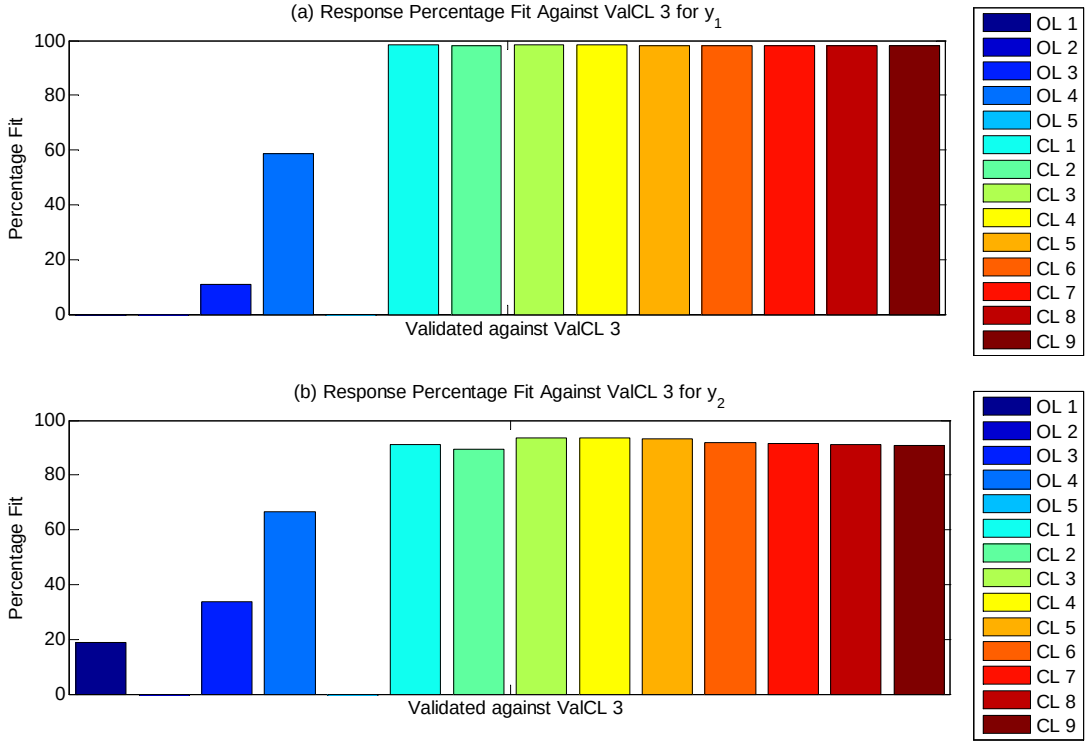


Figure 6.13: Percentage fit values for simulation validation against closed-loop validation sets

by open-loop experimental conditions and PRBS disturbance signals (see table 6.7 for more details on the validation set). It is noted that figure 6.14(a) plots the response of models obtained from open-loop experiments while figure 6.14(b) plots those obtained from closed-loop experiments.

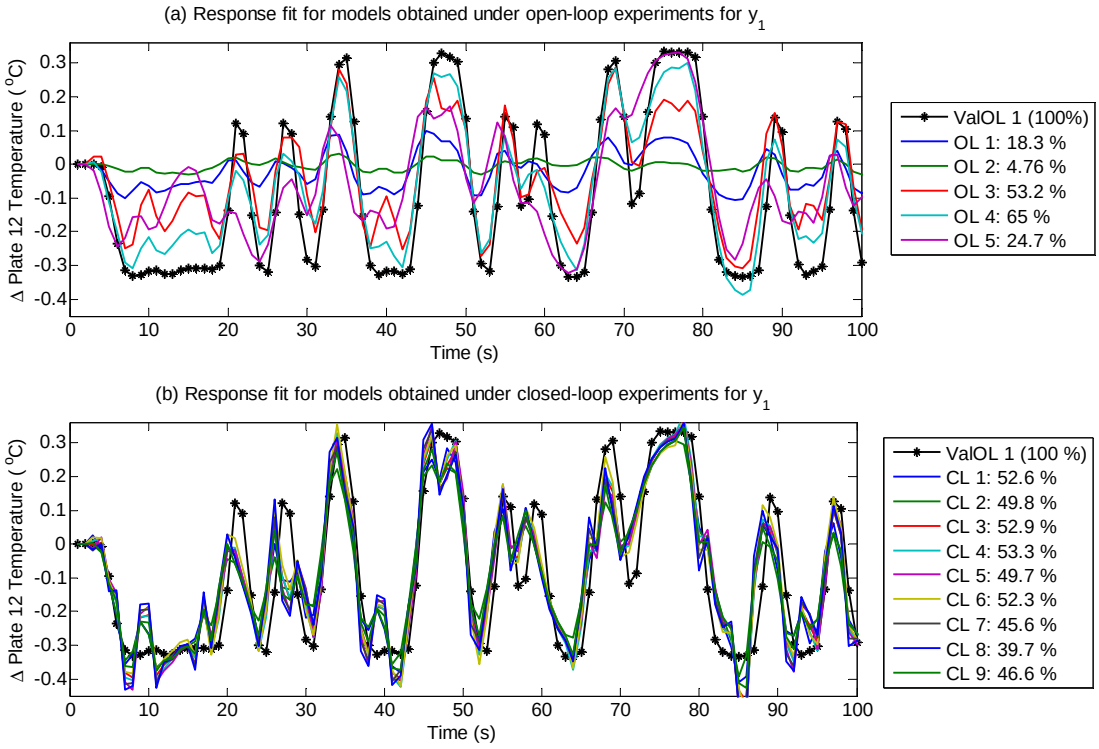


Figure 6.14: Percentage fit simulations validated against ValOL 1

Assessment of Validation Sets - Model Discrimination and Bias

Looking at the percentage fit values in figures 6.12 and 6.13, it may be said that the percentage fit values generally indicate worse fits for the identified models when using validation sets obtained from open-loop experiments (ValOL 1 and 2). Furthermore the models obtained from open-loop experiments (OL 1 -5) did very poorly when validated against closed-loop validation data while the models obtained under closed-loop conditions (CL 1 -9) all showed very large percentage fit values. It is additionally noted that when validating the models against validation set ValOL 2, obtained from an open-loop step disturbance experiment, the difference in percentage fit values between output y_1 and y_2 for each model was more significant than when using the other validation sets. This lack of consistent results between the two outputs made it difficult to determine any clear trend regarding how the different models performed relative to each other when validated against ValOL 2.

Further analysing the differences in percentage fit values when using ValOL 2 in

comparison to those when using the other validation sets, it is noted that validation set ValOL 2 is the only set that produced a relatively large percentage fit value for case OL 5. Case OL 5 referring to the model identified from the experimental case OL 5 defined as an open-loop experiment where the disturbance signals were step disturbances (i.e. ValOL 2 and case OL 5 are both rooted in open-loop step disturbance experiments). In fact, it may be said that the model produced by experiment OL 5 produced the most accurate response when validated against ValOL 2, while it was amongst the worst when validated against ValOL 1 and ValCL 3. On the other hand the model produced by case CL 6, characterised by a closed-loop experiment with no setpoint changes, performed very poorly when validated against ValOL 2. In fact, it was the worst performing case of all the closed-loop cases against ValOL 2 by a significant amount. However, when validated against the other validation data sets, this distinction was not found.

Observing the differences in percentage fits when using the validation data sets obtained under open-loop conditions to those obtained under closed-loop conditions, it is clear that the closed-loop validation data set, produced by case ValOL 3, did not produce informative enough data to allow for the discrimination between the models generated by the closed-loop cases. On the other hand, validation set ValOL 2, which was expected to be very uninformative since it was generated from step disturbances, did a better job than the closed-loop validation set ValCL 3 at discriminating between the models obtained from these closed-loop experiments. However, validation set ValOL 2 found the model obtained from the least informative disturbance signal (the step disturbance, i.e. OL 5) to be the most accurate since it gave the best fit.

From these findings it is clear that the closed-loop validation set is bias towards models generated from closed-loop experiments while the open-loop step disturbance validation set is especially bias towards models that where obtained from similar open-loop step disturbance experiments and generally bias towards experimental data that contain the dynamics revealed by step disturbances and low frequency setpoint changes.

Another way of interpreting this is that these validation sets were highly prejudice against models obtained from experimental conditions different to their own.

The validation data set produced by PRBS under open-loop conditions, ValOL 1, allowed for the best discrimination between models and revealed the most consistent trends amongst the two outputs.

Assessment of Identified Models - Identification Experiment Condition Sensitivities

The following observations and comments are made regarding the response percentage fit results of the different models obtained from the different experiment conditions when validated against open-loop validation set ValOL 1, which was found to be least biased

and most discriminating (see figure: 6.12 - ValOL 1 results)

- It is observed that the only difference between cases OL 2 and OL 3 is that OL 3 had a larger input signal average constant interval value, B (which may infer frequency characteristics). This larger value of B for case OL 3, was selected considering disturbance signal characteristics that more closely resemble the dominant dynamic characteristics of system A. The much larger percentage fit value for case OL 3 does indicate the effect of designing the disturbance signal in accordance with the dominant dynamics of the system being identified. However, this larger value of B used in case OL 3, relative to OL 2, is also nearer to that used to generate the validation set, thus the improved response fit might not necessarily be due to better system identification but to the validation set's bias towards the model obtained from disturbance signals with B values nearer to its own.

- Interestingly enough, the model produced by case OL 5 (defined by step disturbance input signals), showed better validation results than the models obtained from OL 1 and OL 2. Recalling that cases OL 1 and OL 2 are defined as experiments that used higher frequency PRBS input signals (smaller B values) than that used to generate validation set ValOL 1. This could imply that input signals that are not very persistently exciting produce models with better cross validation results than models produced by persistently exciting signals with characteristics that significantly differ form those used to generate the validation data.

- It is clear that of all the open-loop cases, case OL 4, which is characterised as using a PRML input signal, produced the model that best fits the validation set ValOL 1.

- With respect to the closed-loop cases, it is clear that even when validated against the validation set thought to be most informative (ValOL 1), there is little to distinguish and discriminate between them. This may be accounted for by the controller's efforts in dampening the effects of the excitation signal used to identify the system. This would thus reduce the effective differences between the identification experiment designs.

- An alternative interpretation of this lack of response fit distinction among the models obtained from closed-loop experiments is that the dither signals were not as effective as expected. When comparing the results of the models obtainted from cases CL 1-6 there is little significant difference in simulation results. This implies that the differences in dither signal characteristics might have done little to differently excite the system.

- The differences between results of the models obtained from experiments CL 3, 6 and 9 are of significant interest and address the previous point made. All three are defined as closed-loop experiments with case CL 3 using a PRBS dither signal and two setpoint changes to disturb the system. Case CL 6 is further characterised as having the same dither signal as CL 3 but no setpoint changes. While case CL 9 is further defined as having the same setpoint changes as CL 3 but no dither signal. Of the three models obtained from these cases, that produced by case CL 3 had the largest percentage fit values with those of case CL 6 closely following. The model generated by case CL 9 clearly had the lower percentage fit values of the 3. From this one may deduce that the excitation in output caused by servo tracking seems to contribute little since its elimination did little to worsen the fit, while the effect of not exciting the system via the dither signal was more significant. This implies that the dither signals where valuable in producing informative data.

- It is interesting to note that, while the previous observation claims the effect on identification due to excitation from setpoint change is minimal, the difference between the percentage fit values of cases CL 7 and 8 states otherwise. With the only difference in experimental design between these 2 cases being setpoint direction and time of setpoint change, it is important to note that, even though the percentage fit is significantly different only for y_1, this difference does imply that setpoint direction and controller interaction play important roles in multivariable closed-loop system identification. Figure 6.15 shows the difference in identifed model response caused by setpoint direction changes in the identification experiments.

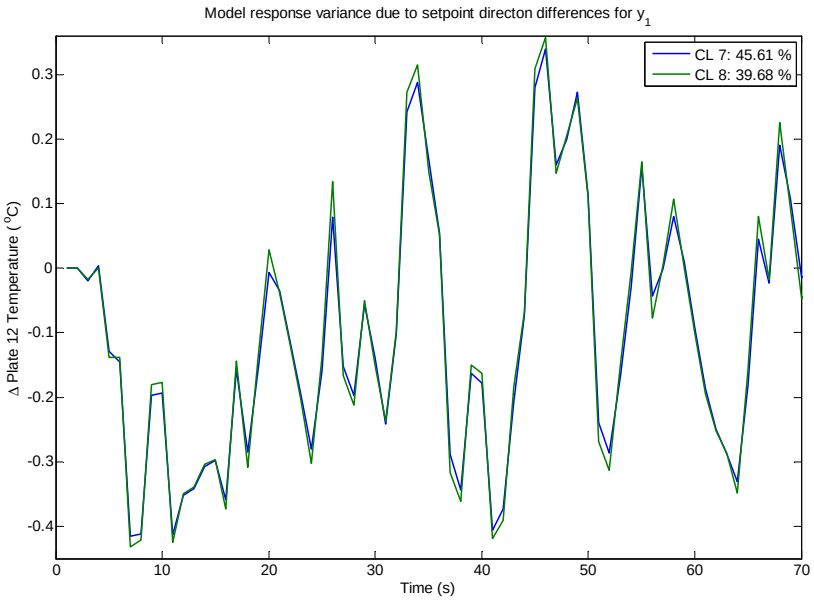


Figure 6.15: Model simulation revealing closed-loop identification sensitivity to setpoint direction

- Observing that there is little difference in percentage fit values between the two identification cases defined by non-linear and constrained non-linear controller conditions, CL 6 and 7 respectively, and those of cases CL 1-5, this indicates that the effect of varying controller parameters and constrained inputs was not significant.

In addition to the observations made, it must be noted that there was no significant difference found between the validation results obtained from simulation and those obtained from prediction. This does suggest that the results obtained may be used to infer indications of model validity in terms of control.

6.3.4 Frequency Analyses

The approach taken in using the frequency domain to analyse the identified models and the validation data was done on two fronts. The first concerns comparing the identified models obtained by the different experimental cases with the true known system in the frequency domain. This involves comparing the frequency responses (bode plots) of the identified model to that of the known mathematical model which represents system A.

The second involves the same cross-validation technique as the previous section, but in the frequency domain. That is, the simulated response of the identified model to the input signals from a validation set, is compared against output response contained in that validation set but in the frequency domain. This is done by taking the validation set response data and the identified model's simulated response data and representing them in the discrete frequency domain. This is done through the use of discrete Fourier transforms and allows for the frequency content of the simulated response fits against each validation set to be compared. Furthermore, the assessment of the frequency content of the response data in the validation sets can be compared to the frequency response of the true system. This allows for an investigation into how well the validation set represents the true system.

Frequency Response Analysis of Identified Models

The analyses of the different frequency responses of the identified models did allow for a clearer understanding of the effects of different identification experiment designs on the identified model's accuracy.

Figure 6.16 illustrates the effect of changing the frequency characteristic of the disturbance signals used to generate information from system A for identification. With figure 6.16(a) being the frequency responses of models identified from open-loop experiments OL 2 and OL 3 and 6.16(b) being those of closed-loop experiments CL 2 and CL 3. The only difference between the experiment conditions used to generate to models that produce the two responses in each plot ((a) and (b)) being the average constant

signal time interval (B) of the disturbance signal used. Cases OL 2 and CL 2 have B values of 1 which characterises their disturbance signals as white signals, while OL 3 and CL 3 have B values of 2 which is more representative of the dominant time constant of system A. It is observed that increasing the B value (effectively decreasing the frequency) of the input signal used in the open-loop identification experiments produced a model with a more accurate frequency response. Figure 6.16(a) shows this clearly in that the model identified from experiment OL 3 has a frequency response that much better fits the frequency response of the true system, system A.

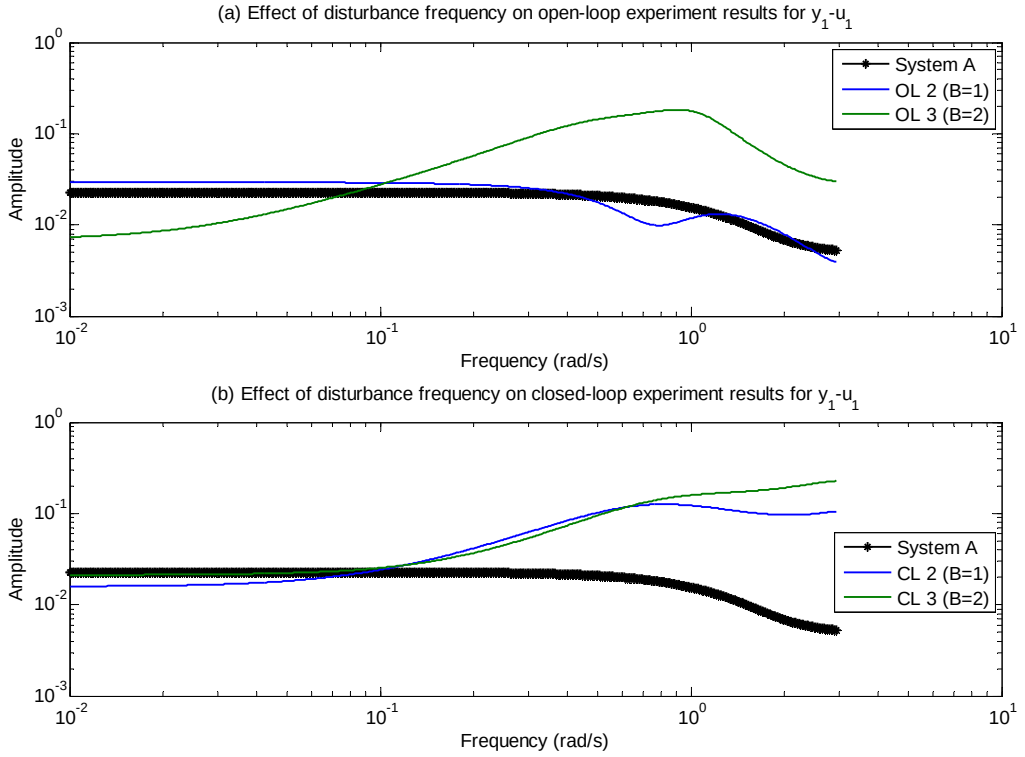


Figure 6.16: Frequency response comparing experimental cases of different identification signal frequency characteristics

Figure 6.16(b) shows the frequency response fits of the two models identified from closed-loop experiments. The same is seen here in that increasing the dither signal's average constant signal time, B, from 1 to 2, produces a model with a slightly improved the frequency fit. This improvement is however less significant than that for the open-loop experiment. This may be accounted for by the closed-loop controllers efforts to reduce dither signal's disturbance effects and the smaller magnitude of the dither signal in comparison to the open-loop disturbance signal.

This finding proves to be very valuable. In the previous section, assessing simulated response fits showed that this increase in B value does provide a better fit of the validation data. It was however uncertain whether this was due to the validation set's bias towards a frequency characteristic (B value) nearer to that of its own, or due to an inher-

ently better representation of the true system. These results establish that the better fit was due to B being more representative of the dominant dynamics of the system being identified, not due to the validation set's bias towards a model obtained from similar experimental conditions. This additionally attests to validation set ValOL 1 being a good representation of the true system.

While most of the findings in analysing the frequency responses of the identified models did reflect the response simulation findings in the previous section, some interesting cases did reveal themselves.

Figure 6.17 compares the frequency responses of models obtained from closed-loop experiments CL 3, 6 and 9. As was done in the previous section, this compares the model obtained from case CL 3 defined as an experiment that used both setpoint changes and dither signal disturbances to excite the system with that from case CL 6, that used only dither signal disturbances, and case CL 9, that used only setpoint changes to disturb the system. The figure further reveals how the frequency responses between output y_1 and input u_1 (figure (a)) differ from that between y_2 and u_2 (figure (b)). Specifically it is noted how not using certain disturbances signals affected the accuracy of the different output responses of the models in different ways. The response between y_1 and u_1 shows that experimentation without setpoint changes, case CL 6, produced a submodel with reduced accuracy at lower frequencies, while removing the dither signal (case CL 9) revealed little effect. The frequency response between y_2 and u_2 shows that there is no effect in not changing setpoints while the lack of dither signal in the closed-loop identification experiment caused a larger submodel mismatch at higher frequencies.

Figure 6.18 shows the frequency responses of both outputs to both inputs for the models identified from cases CL 3, CL 4, CL 5, OL 3 and OL 4. All of these models showed little distinction between each other when validated against ValOL 1 in terms of response fit accuracy in the previous section. That is, validation of simulated response fits against validation set ValOL 1, which is expected to be the most informative set, did little to distinguish between the models generated by these cases. Analysing figure 6.18 it can be said that frequency response analysis does not discriminate between the closed-loop cases since the differences in responses are not significant enough.

However, the responses of the models identified from open-loop experiments were significantly different to those of the models obtained from closed-loop experiments. The open-loop cases were found to be less accurate at lower frequencies for most of the responses. These findings better reflect the response fit results when validated against the closed-loop validation set, ValCL 3, than against the open-loop validation set, ValOL 1. It is noted (in figure 6.18) that case OL 4 has a smaller deviation from the true model frequency response in comparison to OL 3. This discrimination is found in the response fit results when validated against the closed-loop validation set, ValCL 3 (in figure 6.13), but not when validated against the open-loop validation set, ValOL 1 (in figure 6.12).

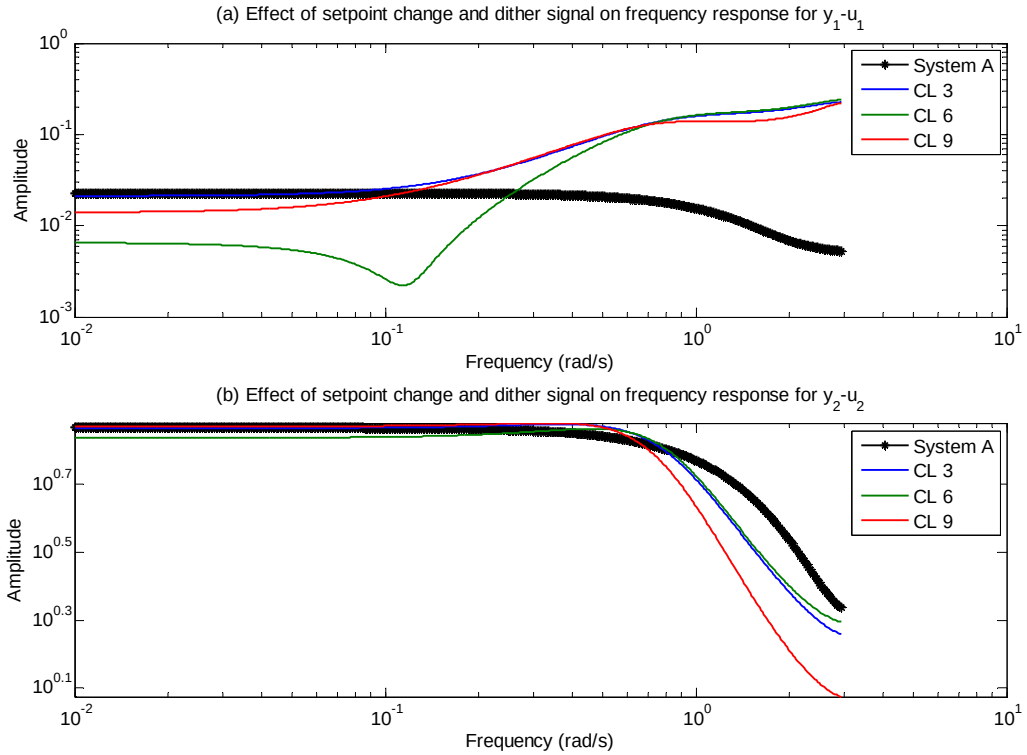


Figure 6.17: Frequency response comparing submodel responses to removal of identification signals

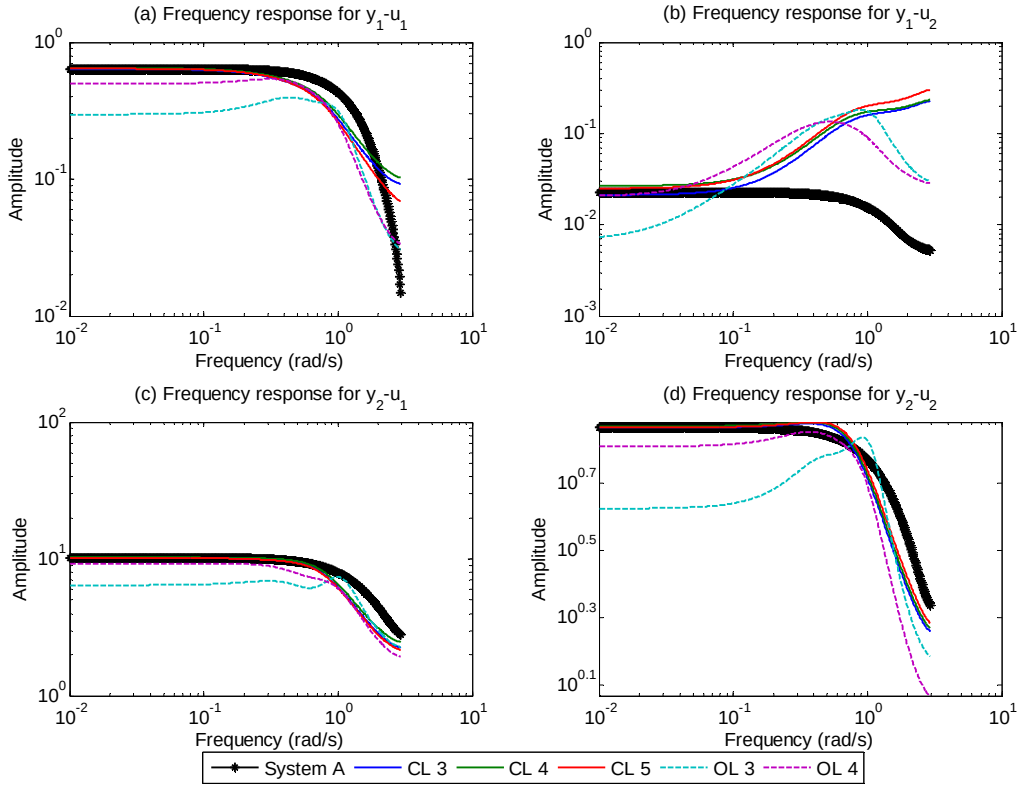


Figure 6.18: Frequency responses of cases with indistinguishable simulation fits against ValOL 1

Once again these findings suggest that the inability of the closed-loop validation set (ValCL 3) to discriminate between models obtained from closed-loop experiments through response fit validation (see figure 6.13) may not have been due to bias towards models obtained from experiments with the same feedback condition. This is since the frequency response analysis is not a cross-validation technique in that it is dependent only on the characteristics of the models themselves, independent of validation sets. Yet the frequency response results given in figure 6.18 more closely resemble the cross-validation results when using the closed-loop validation set than any other in that there is little to distinguish between the responses of the models obtained from closed-loop experiments. This suggests the closed-loop validation data set may more accurately reflects the true system.

Further analysing the frequency responses of the models obtained from open-loop experiments in figure 6.18 (cases OL 3 and 4), it is observed that there is a general mismatch between OL 3 and OL 4 at the lower frequencies with OL 3 producing the worse fit. The only difference between the two experiments is the type of input signal used. OL 3 used PRBS disturbances signals while OL 4 used PRMLS. Also, 6.18(b) shows how the open-loop cases generated models with better fits at higher frequencies than the closed-loop cases.

This better fit at higher frequencies by the models identified from open-loop experiments, while not absolutely evident in all the responses, might be accounted for by the fact that the identified models all have ARX structures, which do not adequately allow for independent modelling of noise and dynamics. This is an established problem of the ARX structure as has been explained in section 5.4.2. This issue of dynamic model corruption due to noise is expected to be more of a problem for the closed-loop experiments as the feedback condition further increases this inability to independently model noise and dynamics. This might justify the better high frequency fit for model obtained from open-loop experiments.

Frequency Content of Response Fits

Through the previous assessment into each identified model's frequency response and how they compare to the true system's frequency response, an understanding was obtained regarding the true accuracy of the different models over certain frequencies without the use of validation data. Following this, it is appropriate to assess how the different validation sets reveal the model accuracies and how this compares to the true accuracy assessment done via the frequency response. Thus in essence, while the previous assessment presented model accuracy based on inherent model characteristics independent of validation sets, here the presented accuracy of the identified model is dependent on the validation set. As explained earlier, this is done by analysing response fits between the identified

model and the validation data set via discrete frequency representation and serves the purpose of testing the credibility of the validation sets.

The assessment is done against validation sets ValOL 1 and ValCL 3 (the sets generated by PRBS under open and closed-loop conditions respectively) for cases: OL 2, OL 3, OL 4 and CL 3, Cl 4, CL 5. This is as these cases yielded the most information and discussion on the simulated response fits and model frequency response analyses (figure 6.18).

The comparisons of frequency component fits between each of the models obtained from these cases and validation set ValOL 1 for output y_1 are shown on figure 6.19. From comparison with the frequency response results show in figure 6.18, it is deduced that the validation data set ValOL 1 does to some extent allow for accurate validation and model discrimination that reflects the differences between the identified models and true model. Figure 6.19 shows how responses from models obtained from closed-loop experiments (figures 6.19(d),(e) and (f)) validated against ValOL 1 closely fit the validation data at lower frequencies while slightly drifting at the higher frequencies. In comparison, the frequency content analysis of the responses from models obtained from open-loop experiments (figures 6.19(a),(b) and (c)) have worse fits at the lower frequencies. This is exactly what was found in the frequency response analysis in figure 6.18.

This may once again be explained by the fact that the closed-loop experiments used setpoint changes and dithers signal disturbances to generate information on the system, whereas the open-loop experiments only used input disturbances. The condition of using low frequency setpoint changes may explain the better fit at lower frequencies for the models generated from closed-loop data.

Looking at figure 6.19 it is observed that at the higher frequencies the signals become noisy. This is thought to be since the response data of the validation set (ValOL 1) and the simulated model responses may not have sufficient and accurate information at those frequencies since they were not sufficiently disturbed at such frequencies. It is however strange to find that the model obtained from case OL 2 in figure 6.19(a), produced a response trend that is much smoother at these high frequencies than the others, even though it was disturbed by the same signal. What distinguishes this model from all the other models being assessed is that the experiment used to generate data for its identification is characterised by an input disturbance signal with a much smaller constant signal interval value (B). This means the PRBS input signal disturbed the system at a higher frequency. Thus this model should be more accurate at higher frequencies, however, this accuracy was not expected to manifested its self since the validation set did not disturb the system at such frequencies and thus was not expected to produce consistent information at such frequencies.

Figure 6.20 shows the frequency component representations of the same models as 6.19 (those obtained from experiment cases OL 2, OL 3, OL 4 and CL 3, Cl 4, CL

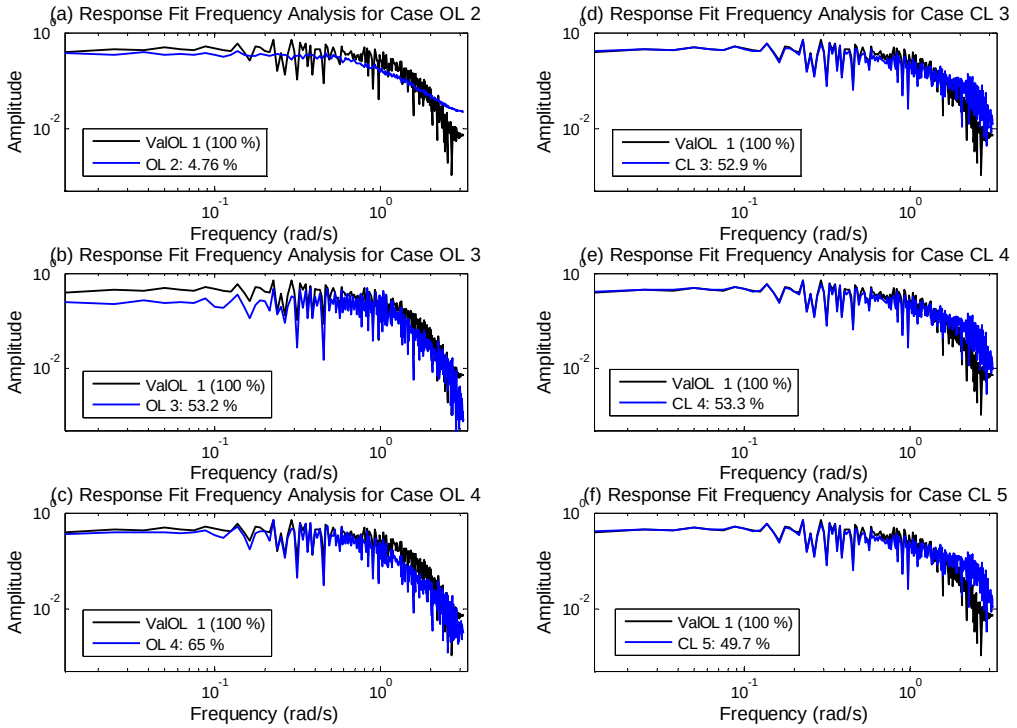


Figure 6.19: Frequency analysis of simulation responses validated against ValOL 1

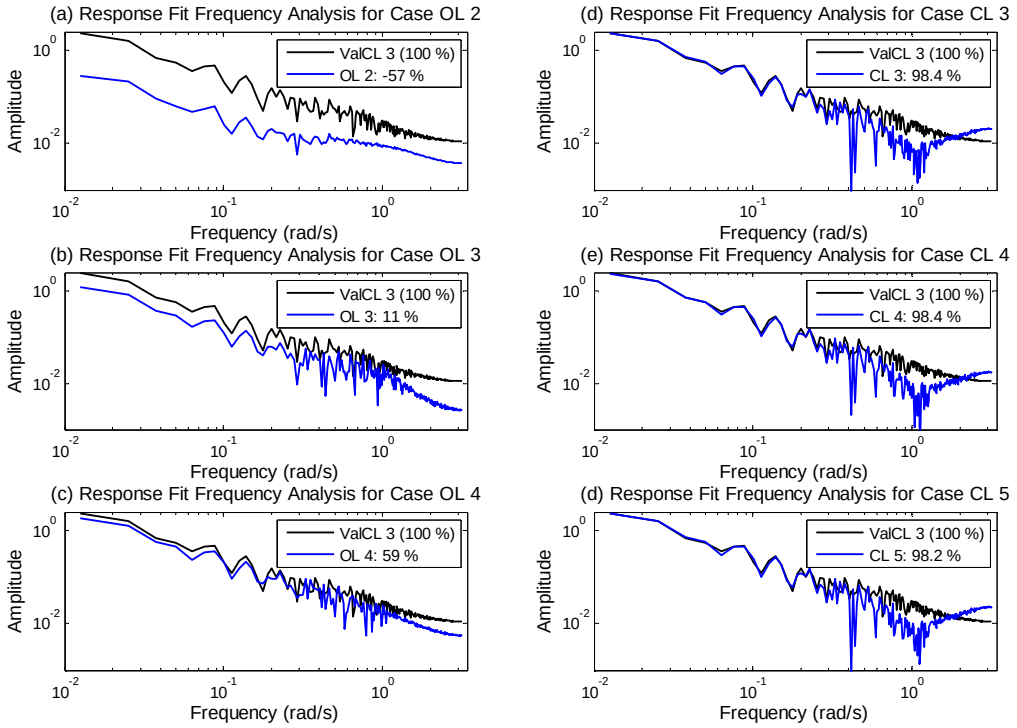


Figure 6.20: Frequency analysis of simulation responses validated against ValCL 3

5) but validated against ValCL 3, the data set obtained under closed-loop conditions. The first thing that is noticed is the earlier decay in amplitude in comparison to the frequency analysis against ValOL 1 (figure 6.19). With the frequency profile having larger magnitudes persisting at low frequencies of small bandwidths and small magnitudes with larger persistence at the higher frequencies, this reveals the closed-loop identification condition. Where the large disturbances are implemented by setpoint changes and at very low frequencies leaving small magnitude dither signals to extract most of the information at higher frequencies.

Recalling a previous observation made from figure 6.18(b) where the models identified from open-loop experiments (OL 3 and OL 4) produced slightly improved frequency response fits at the higher frequencies than the closed-loop cases. That is, the models obtained from closed-loop experiments were found to have a high frequency respons misfit when compared to the true system. Looking at figures 6.19 and 6.20, this misfit at higher frequencies of the models obtained from closed-loop experiments, characterised by the rise in magnitude where the validation set falls, is made more evident when using validation set ValCL 3 (figures 6.20(d),(e) and (f)) than ValOL 1 (figures 6.19(d),(e) and (f)). This means that the closed-loop validation set, ValCL 3 did a better job at revealing these high frequency inaccuracies of the models obtained from closed-loop experiments than the validation set obtained from the open-loop experiment.

This finding reveals an interesting conditions. Previous assessments suggested that the models produced by closed-loop experiments are less accurate at the higher frequencies than those obtained from open-loop experiments. While this last finding reveals that the validation set obtained from closed-loop experiments better discriminates at higher frequencies, which implies more accurate information, than the validation set obtained from open-loop conditions.

6.3.5 Residual Correlation Analysis

Both the whiteness and independence correlation tests were used to assess the model residuals. With the residuals being generated by cross-validation, the same cross-validation approach is used here as in the simulation and prediction section. The inputs used to generate responses from the identified models were those of validation sets ValOL 1 and ValCL 3 (the open and closed-loop experiments conditions respectively with both using PRBS signals). ValOL 2 (the open-loop step response experiment case) was not used since its inability to persistently excite the system did not yield informative enough data for effective correlation analyses. In essence this section conveys the findings from the implementation of correlation tests on the residuals generated when assessing the response simulation percentage fits in section 6.3.3.

The important difference, however, is that the simulations that provided the responses

used to assess percentage fit values where done without the noise model. This was done so as to compare the identified dynamics model and the true system (system A). These residual correlation tests are done on simulated responses of identified models include the noise model. Noise was accordingly also superimposed on the validation output data. Thus, as was established in section 4.1.5, if the identified model was to perfectly represent system A, the residuals would only be white random realisations due to the difference between realisations of noise used in simulating responses from the identified model to that superimposed on the validation data set output.

The reader is referred to section 4.1.5 for the residual correlation result interpretation rules and is reminded that ϵ_1 and ϵ_2 signify the residuals for output 1 and 2 respectively.

Whiteness Test Results

The whiteness test, assessed by residual auto-correlation plots, of the models generated by open-loop experiments gave some debatable results. It was found that the results heavily depended on the validation data by which they were disturbed by and simulated against. Figure 6.21(a) shows the residual correlations for the model identified from case OL 2 when simulated against validation set ValOL 1 (the open-loop set) while figure 6.21(b) shows the result against ValCL 3 (the closed-loop set).

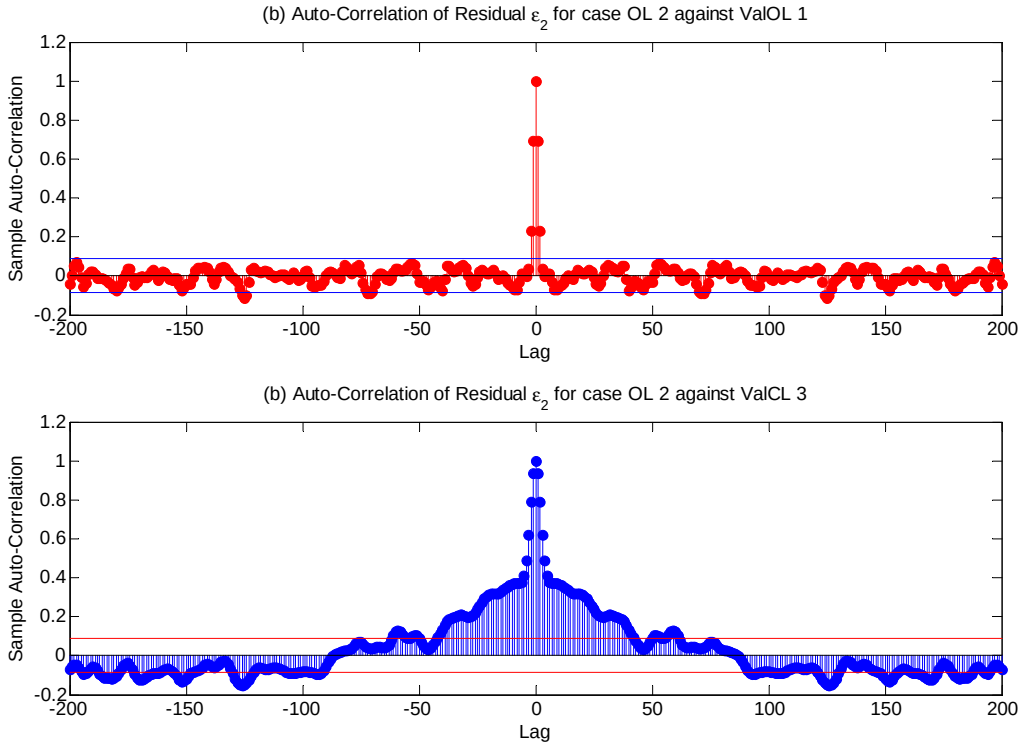


Figure 6.21: Illustration of sensitivity of correlation tests to validation data for OL cases

The correlation against ValOL 1 does not show perfect whiteness, however, in com-

parison to the correlation against ValCL 3 these correlations are relatively insignificant. Interpretation of this correlation dependence on validation set suggests that the closed-loop validation set, ValCL 3, allows for the detection of unmodelled dynamics where ValOL 1 did not. This trend is seen throughout the models obtained from open-loop experiments for both residuals, ϵ_1 and ϵ_2.

If the simulated response percentage fit results were a reflection of these correlation results, then all the open-loop cases would have generated models that performed worse when validated against ValCL 3 than when validated against ValOL 1. The fact is that to some extent this is the case, recalling figures 6.12 and 6.13, one observes that all the models obtained from open-loop experiments (the OL cases) showed worse percentage fit values when validated against ValCL 3 than against ValOL 1.

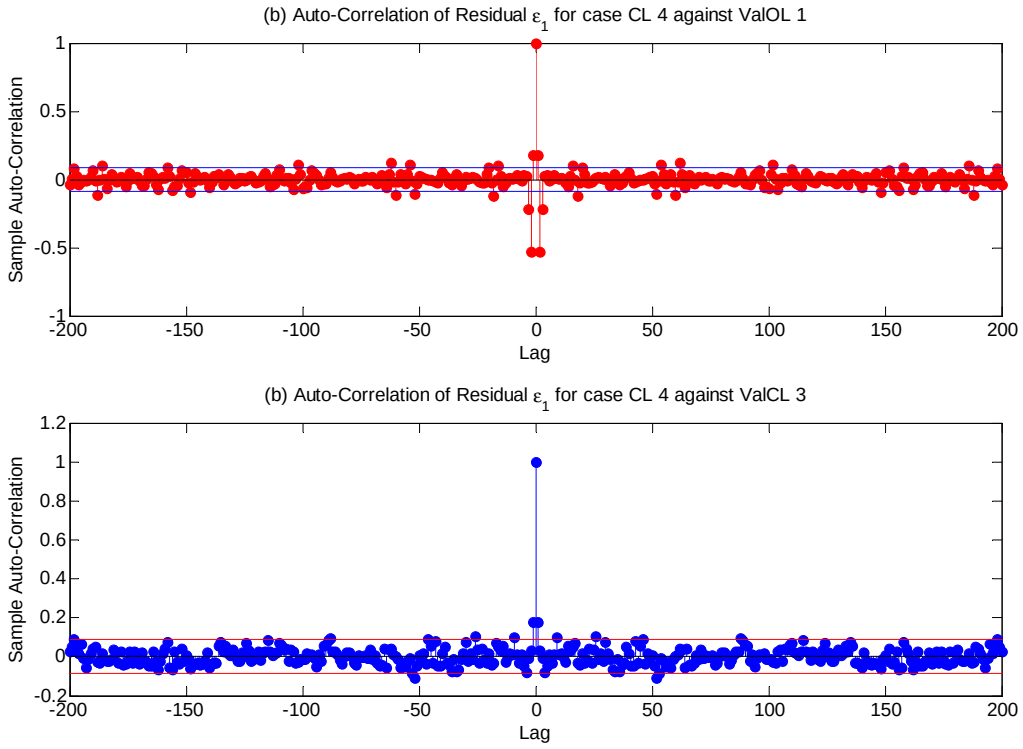


Figure 6.22: Illustration of sensitivity of correlation tests to validation data for CL cases

The residual correlation analyses of the models generated from closed-loop experiments showed much more consistency and independence of the validation set against which they were validated. All the correlation plots resembled those shown in figure 6.22 illustrating the results for case CL 4. The results indicate very little significant correlation regardless of the validation set against which they were validated and suggest little differences in indications of unmodelled dynamics. This is contrary to simulated response percentage fit results for the models obtained from closed-loop experiments (CL cases) as they implied large extents of umodelled dynamics when validated against ValOL 1 and

almost none when validated against ValCL 3.

In terms of the differences in residual correlations between the two outputs, the same trends were shown for both, that is, there where no differences in correlation results between ϵ_1 and ϵ_2.

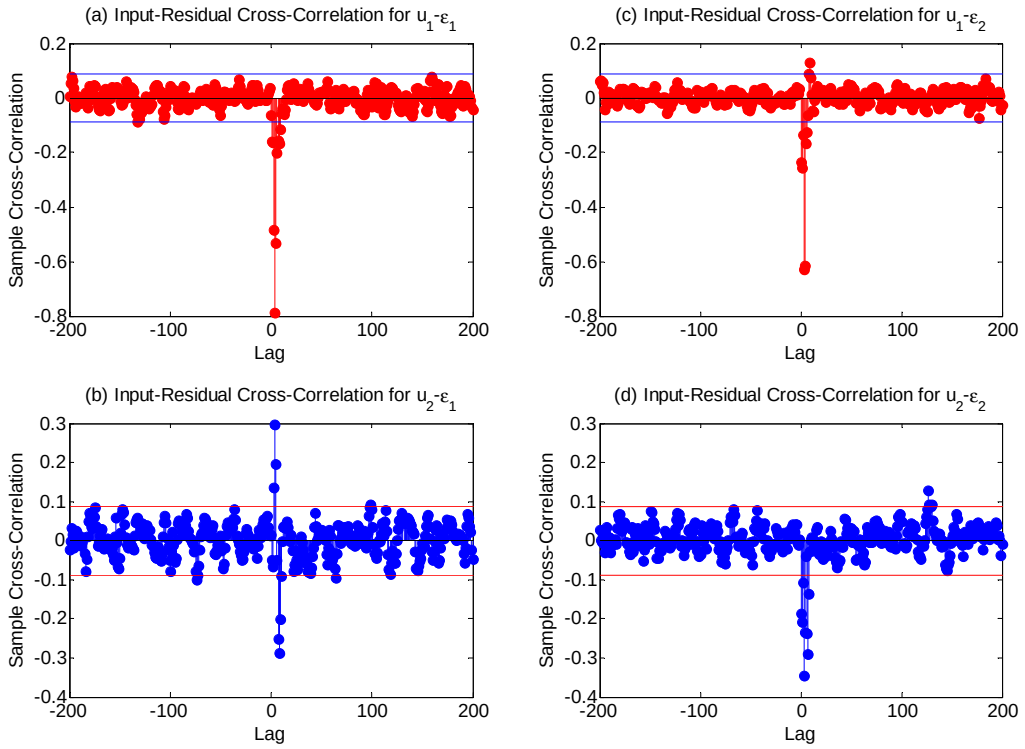
Independence Test Results

The results on residual independence tests reflected those of the whiteness test. The models generated from open-loop data showed larger sensitivities to the validation set used in comparison to those generated from closed-loop experiments. Figure 6.23(a) shows how the model obtained from case OL 3 reveals little residual dependence on input signals when validated against the open-loop validation set, ValOL 1. In contrast to this, figure 6.23(b), shows a much larger extent of input dependence of the same model when validated against the closed-loop validation set, ValCL 3. This larger dependence implies a larger component of the model that wrongly fits the true system. This was found for all the models obtained from open-loop experiments.

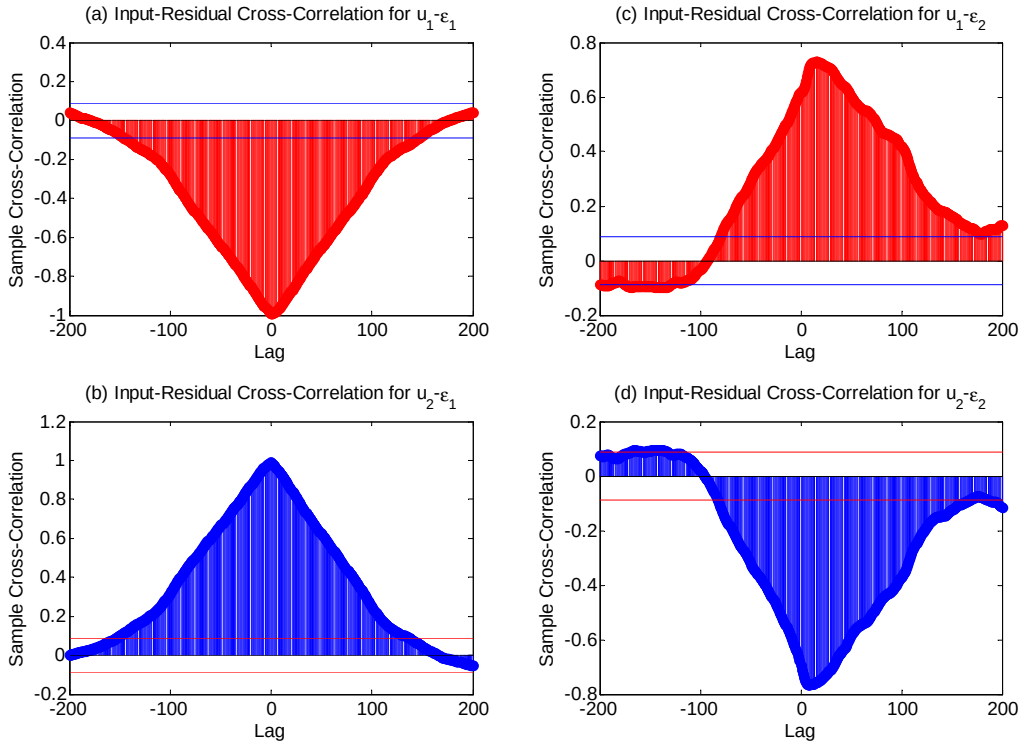
Almost all the models obtained from closed-loop experiments showed little residual dependence to input signals regardless of the validation set used. Only case CL 6 showed larger dependencies that suggested model dynamics that wrongly describe the true system, the rest of the closed-loop cases showed input-residual correlations that closely resembled that of case CL 4 shown by figure 6.24(a) validated against the open-loop validation set, ValOL 1, and figure 6.24(b) against the closed-loop validation set, ValCL 3.

It must be noted that case CL 6, the only closed-loop case that generated a model that revealed some significant input-residual correlation, is characterised as the closed-loop experiment where no setpoint changes where implemented. This means only the dither signals were used to excite the system, this resembles the open-loop experiments but with lower magnitude disturbance signals and controller action reducing the disturbance effects. From this one may gather that the signal structure changes caused by the setpoint changes are beneficial in assuring independence between inputs and residuals.

In trying to interpret the differences in input-residual correlations obtained when using closed-loop validation sets to those when using open-loop validations sets, it is important to recall that the input-residual correlation results validated against closed-loop experiments are not to be completely trusted. Closed-loop experiments generate data with definite correlations between the inputs and output signals due to the feedback loop. This undoubtedly means there is a correlation between the input and the residual values when using closed-loop data. From this, it is implied that these correlations cannot purely signify incorrectly modelled dynamics.

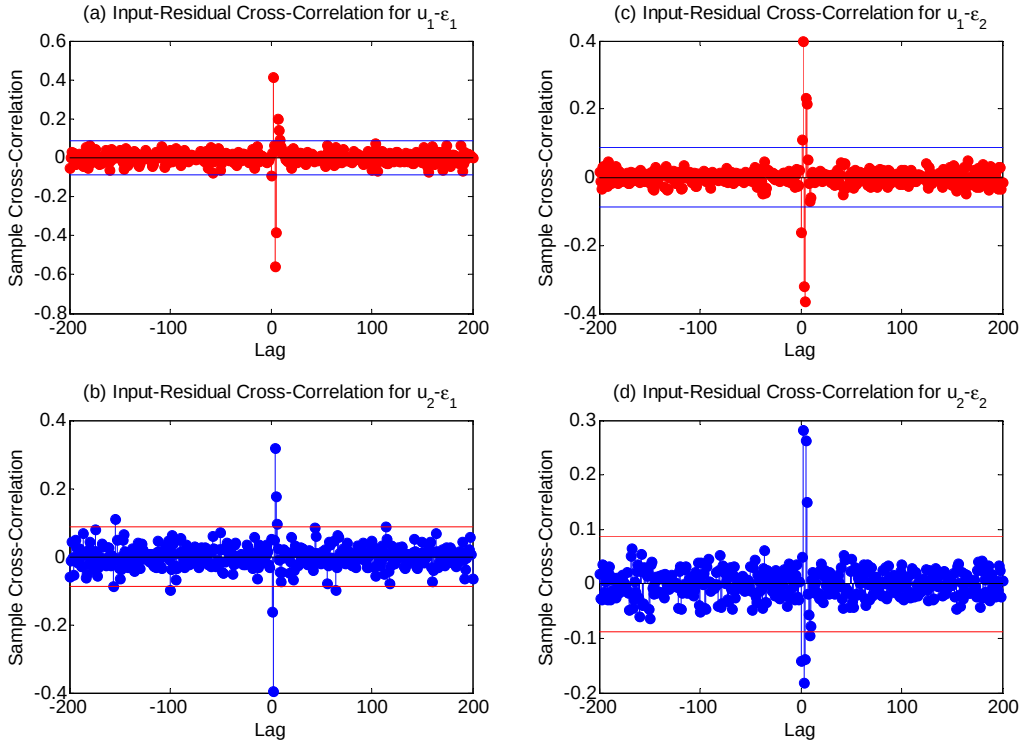


(a) Input-Residual correlations correlated against ValOL 1 for OL 3

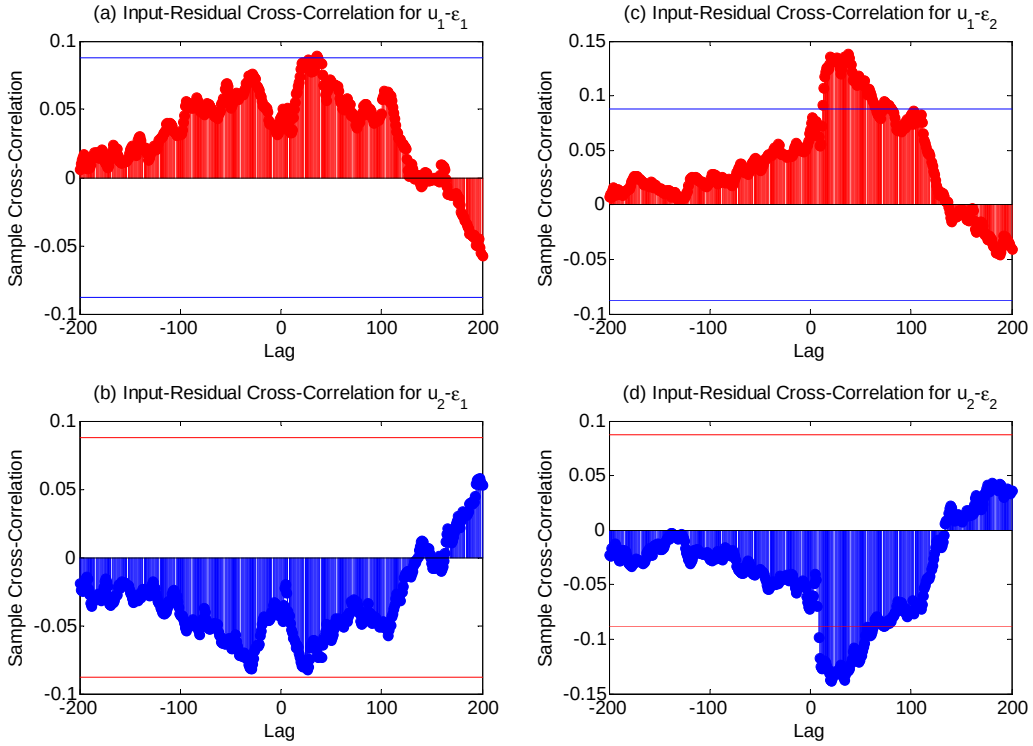


(b) Input-Residual correlations correlated against ValCL 3 for OL 3

Figure 6.23: Input-Residual correlations for OL 3 correlated against ValOL 1 and ValCL3



(a) Input-Residual correlations correlated against ValOL 1 for CL 4



(b) Input-Residual correlations correlated against ValCL 3 for CL 4

Figure 6.24: Input-Residual correlations for CL 4 correlated against ValOL 1 and ValCL3

6.3.6 Results Overview

The identification and validation results for system A, the linear system, allowed for some interesting findings. These may be summarised as follows:

Model Variance, Uncertainty and Stability

Variance due to Noise Sensitivity:

- The information generated from the open-loop step disturbance experiment (OL 5) was found to be most sensitive to different noise realisations. This implies a low capacity to generate dynamics information by a step disturbance.

Model Uncertainty:

- The model obtained from the open-loop step disturbance experiment (OL 5) was found to have the smallest measure of model uncertainty. This was however not by a significant amount.

Stability:

- The model obtained from the open-loop step disturbance experiment (OL 5) was found to be the only model that did not have a pole outside the unit circle on the z-plane as the true system did.

Simulation and Prediction

Bias and Discrimination:

- The worst response fits from the identified models were obtained when validated against the open-loop validation ValOL 1 (characterised by PRBS disturbance signals).
- The response fit results when validated against the closed-loop validation set, ValCL 3, suggested the validation set was biased towards models obtained from closed-loop experiments.
- The response fit results when validated against the open-loop step disturbance validation, ValOL 2, suggested that validation set ValOL 2 was biased towards models obtained from experiments with similar signal disturbances i.e. step disturbances and setpoint changes.
- The above point was emphasised by the fact that validation set ValOL 2, was the only set to reveal a terrible fit for the model obtained from closed-loop experiments with no setpoint changes.

Experimental Condition Sensitivity:

- It was established that models generated by experiments with disturbance signal frequency characteristics chosen in consideration of the dominant time constant of system A had better fits. These disturbance signal frequency characteristics were also closer to that of the validation set being used. This meant it was not possible to distinguish whether the better fit was due to better inherent model accuracy or validation set bias.
- The PRMLS disturbance signal (closely followed by the PRBS) proved to produce the model with the most accurate responses from those that were generated from open-loop experiments while the PRBS proved to be most consistent for the closed-loop experiments.
- The models obtained from closed-loop experiments were found to be less sensitive to changes in experimental design variables. This was thought to be due to the controller dampening the effects of disturbances signals.
- The previous point means that given disturbance signals that were relatively similar in terms of persistent excitation, the open-loop experiments produced models with more accurate responses. However, given poorly designed disturbance signals that were not very persistently exciting, the open-loop experiments were affected more in that they generated models that produced less accurate responses compared to those generated from closed-loop experiments.
- When validated against the open-loop PRBS disturbed validation set, it was revealed that models identified from experiments that did not use dither signal disturbances did worse than those identified from experiments that did not use setpoint changes.
- The direction and timing of setpoint changes made during identification experiments were found to have effects on the identified model. It was found that setpoint changes in opposite direction and at different times were most informative.

Frequency Analysis

Frequency Response Analysis:

- Frequency response analysis allowed for more insight into the simulate response fits results.
- It was found that the better simulated response fits of model obtained from experiments disturbed by signals with frequency characteristics chosen in consideration

of the dominant time constant of system A was in fact not due to validation set bias.

- It was shown that not using a dither signal disturbance can reduce model accuracy at higher frequencies and that implementing setpoint changes can drastically improve model accuracy at low frequencies.
- It was shown that experiment variable changes can have different effects on the response accuracy of different outputs of the same model.
- Assessing how the frequency response of each identified model varied relative to each other and the response of the true system, system A, it was revealed that the results indicated by the frequency responses best resemble the response fits obtained when validated against the closed-loop validation set.
- Frequency response analyses suggested that the models obtained from open-loop experiments have better model accuracy at higher frequencies. This was thought to reflect the condition of dynamic model corruption by noise which closed-loop models are more susceptible to.

Frequency Content of Response fits:

- Analysing the simulated response fits in the frequency domain and comparing the finding to the frequency response analysis it was found that both the open-loop PRBS disturbed (ValOL 1) and the closed-loop (ValCL 3) validation sets both did well in representing the true system in terms of information content.
- However, the validation set obtained from closed-loop conditions was found to better represent the true system at higher frequencies.

Residual Correlation

Whiteness Test:

- The whiteness test reflected much of the findings made from the simulated response fit analyses.
- The closed-loop validation set was found to better discriminate between the models obtained from open-loop experiments sets than the open-loop validation set.
- The whiteness tests revealed that the models obtained from closed-loop experiments have less unmodelled dynamics.

Independence Test:

- All the models obtained from open-loop experiments showed little residual-input correlations when validated against the open-loop validation set but large correlations against the closed-loop validation set.
- All but one of the models obtained from closed-loop experiments showed little residual-input correlation regardless of the validation set used.
- The model obtained from experiment CL 6, characterised as a closed-loop experiment with not setpoint changes, was the only one obtained from closed-loop experiments that showed significant residual-input correlations. This was the case against both open and closed-loop validation sets. This result clearly revealed the value of setpoint changes in closed-loop experiments in assuring model residuals were not correlated to inputs.
- Emphasis was placed on the use of closed-loop validation data in assessing input-residual correlations. Closed-loop data inherently contain correlations between the inputs and outputs via the feedback, thus introducing correlations between the input and residuals that don't necessarily imply incorrectly modelled dynamics.
- Thus, disregarding the input-residual correlation results obtained when validated against closed-loop validation sets, the only model that showed input-residual correlation was that of CL 6, obtained from the closed-loop experiment that did not implement setpoint changes.

6.4 Identification and Validation Results for System B - The Non-Linear System

This section presents and discusses the results of the identification of linear models approximating the non-linear system, system B, and the validation of such models. Note that all the parameters of each identified model are presented in appendix A.2 Each of the 23 identified models (6 generated from open-loop experiments and 17 from closed-loop experiments) are first assessed for variance sensitivities and model uncertainty in section 6.4.1 in a similar fashion as was done for the models approximating system A (section 6.3.1). Following this the identified models are cross-validated against the 5 validation sets (3 obtained from open-loop experiments and 2 obtained from closed-loop experiments) generated from system B. The validation approach is similar to that of the previous section in that cross-validation techniques (see figure 6.8) are primarily used.

It is at this point appropriate to recall tables 6.6 and 6.5 for the open and closed-loop experimental conditions respectively used to identify models from. Table 6.8 details the experimental conditions used to generate validation data. It is important to note

the case tags of each identification experiment and validation experiment as they will be continuously referred to in this section. As with the previous section, a brief mention of the conditions characterising the experiment used to generate the model or validation set being discussed is done where necessary and relevant to the point being made.

6.4.1 Model Variance and Uncertainty

The assessment of sensitivity to noise and the model uncertainties of the models identified from each identification experiment in efforts to approximate system B, was done in the same manner as that of the linear system, system A. That is, 50 repetitions of each identification experiment was done to generate 50 different noise realisation and assess the effects thereof. Also, Monte Carlo simulations were used to generate 50 simulated responses reflecting the response variance due to parameter uncertainty.

Variance Due to Sensitivities to Random Noise Realisations

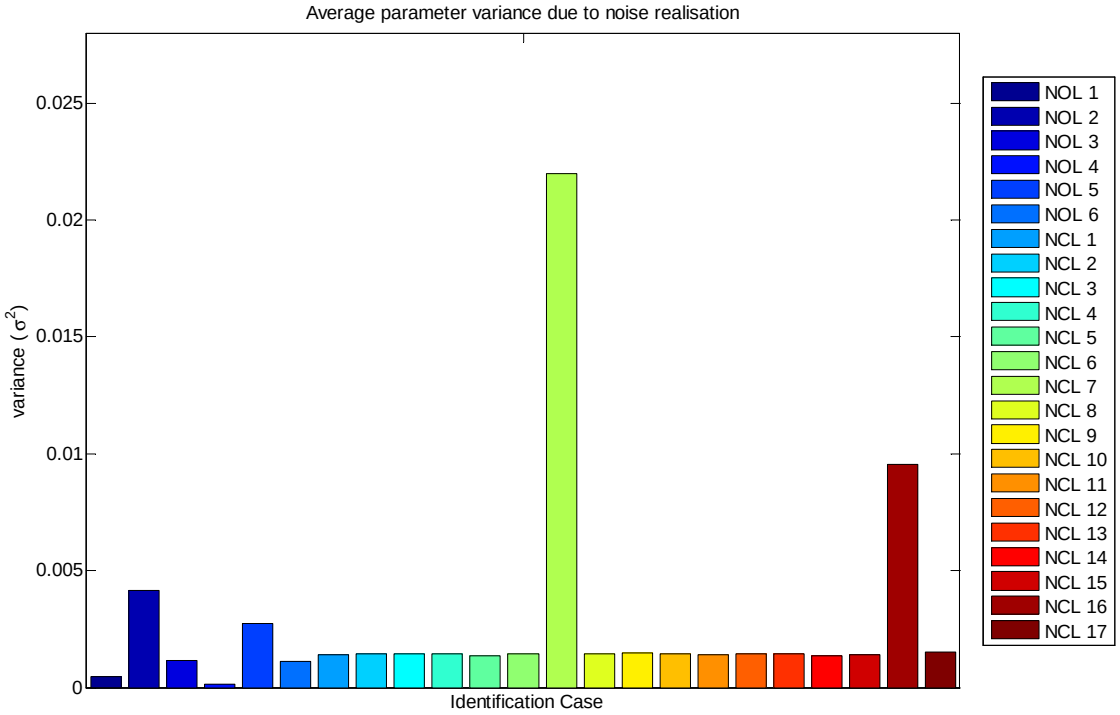


Figure 6.25: Parameter variance of non-linear system estimates due to random noise realisations

Figure 6.25 shows the resulting average parameter variances of the models identified for the respective identification experiments given the 50 repetitions of each. It is interesting to note that the two cases that yielded the largest parameter variances due to noise sensitivities were cases NCL 7 and NCL 16, these are both closed-loop experiments with

no setpoint changes. While the variance magnitudes of these two cases are significantly larger than the others, they are also significantly different to each other. That is, case NCL 16 which is precisely the same as NCL 7 but with a larger signal-to-noise ratio has the smaller parameter variance due to noise sensitivity.

Furthermore, all the other closed-loop experiments showed the same variance while the open-loop experiments showed no trend or allowed for any logical deductions with case NOL 2, characterised as a white PRBS disturbance signal, having the largest variance.

It must however be noted that even though the variances experienced through cases NCL 7 and NCL 16 might be considered significantly larger than those of the other cases, none of the variance magnitudes were large enough to effectively show significant variance in simulated responses to PRBS input disturbances.

Model Uncertainty

The average standard deviations obtained via Monte Carlo simulations reflecting the model uncertainties are shown in figure 6.26. Once again, as with the parameter variance due to noise sensitivities, the closed-loop cases with no setpoint changes (NCL 7 and NCL 16) gave the largest average simulated response standard deviations. This implies that the model generated from these experimental conditions had the largest extents of model uncertainty.

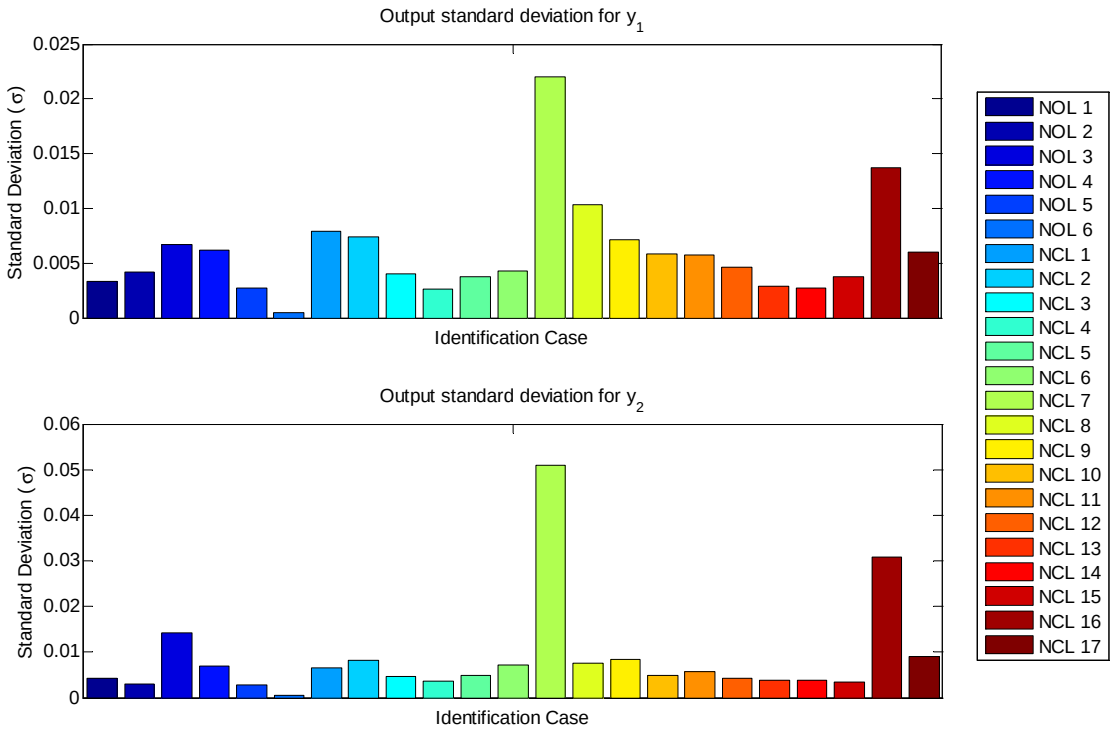


Figure 6.26: Standard deviations of output responses found through Monte Carlo simulations showing model uncertainty

The trends for the different cases are very much the same for both output y_1 and y_2. The model uncertainty does however seem to be greater in components associated with output y_2. This is most likely due to fact that the larger system gains are between both inputs and output y_2.

6.4.2 Simulation and Prediction

General Observations

Figure 6.27 shows the percentage fit values for the identified model responses when validated against the open-loop validation data sets (NValOL 1, NValOL 2, NValOL 3). Comparing the results between the 3 open-loop validation sets it is quite evident that they vary immensely between the different sets. Figure 6.28 shows the percentage fits for the identified model responses when validated against the closed-loop sets (NValCL 4 and NValCL 5). These results do not show a clear trend when comparing the results between the two validation sets. This is mostly on account of the difference in fit values between y_1 and y_2 for NValCL 5.

The response fit results validated against the closed-loop sets do however show greater variance within the sets. That is, it is evident that the closed-loop validation sets allowed for greater discrimination between the identified model responses than the open-loop validation sets.

Bias and Discrimination by PRBS Disturbance Validation Sets - Disturbance Signal Magnitude Sensitivities

Looking at the results validated against the open-loop data sets, figure 6.27, the drastic reduction in percentage fit values between validation set NValOL 1 and NValOL 2 introduces the characteristic problem behind estimating a non-linear system with a linear model. With the only difference between the two validation sets being that set NValOL 2 was generated using larger disturbance signals than NValOL 1, it may be said that these larger system disturbances reveal the non-linear nature of the system to a greater extent and consequently the linear models, identified from smaller disturbance magnitudes and their responses, performed worse.

This is further emphasised when looking at the results of the models identified from experiments that used larger input signals similar to those used to generate validation set NValOL 2. These cases, NOL 4, NCL 4 and NCL 13, are characterised as the dips in the percentage fit profile validated against NValOL 1 and the peaks when validated against NValOL 2. This is more evident for the y_2 fits.

As with the open-loop validation sets just discussed, the only difference between the two closed-loop validation sets, NValCL 4 and NValCL 5, is that set NValCL 5 was allowed

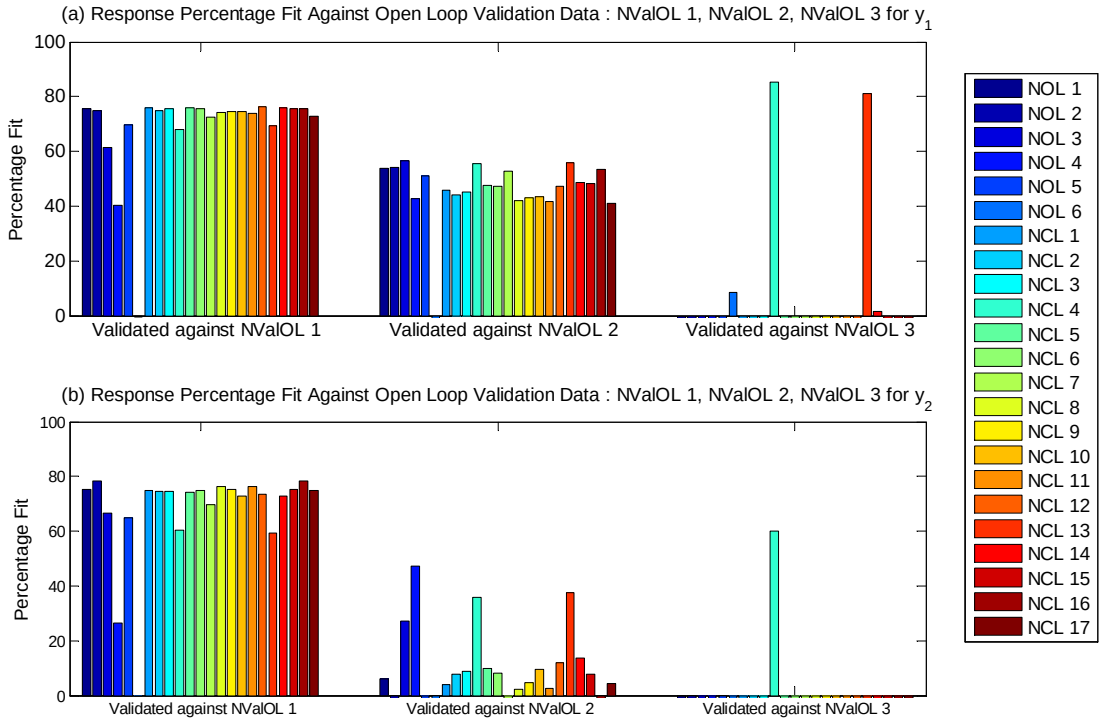


Figure 6.27: Simulation percentage fit values validated against open-loop validation data sets

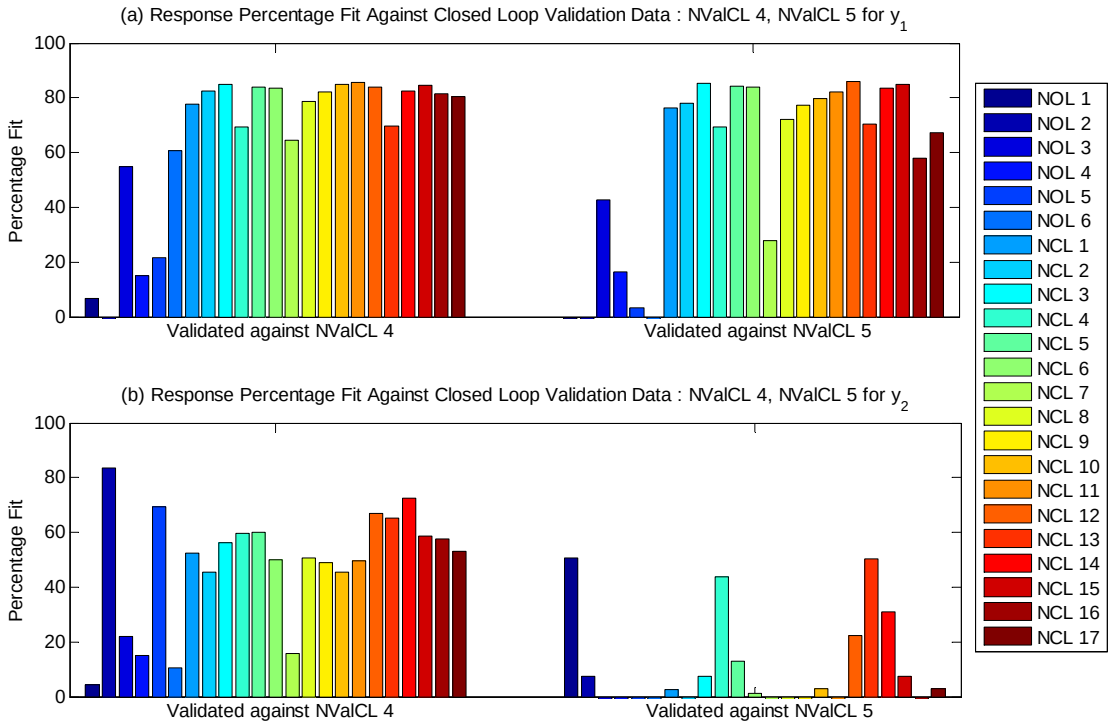


Figure 6.28: Simulation percentage fit values validated against closed-loop validation data set

larger setpoint changes and consequently operated over a larger output range. Figure 6.28 shows the response percentage fits of each identified model validated against these two closed-loop validation sets. The effect of varying the magnitude of the disturbance signals, and hence the dynamic region over which data was generated, on the experiments used to generate to closed-loop validation sets was not as evident as when using the open-loop validation sets. As mentioned earlier, this may be attributed to inconsistencies in results between the results for output y_1 and y_2. Looking at the results for output y_1 (figure 6.28(a)) it is noted that when validating the models against the closed-loop validation set generated by larger disturbances signals (NValCL 5), the change in response fit, relative to the results validated against NValCL 4, is not evident for most models. In contrast to this the results for output y_2 (figure 6.28(b)) do more clearly indicate that most of the models do have significantly worse fits when validated against NValCL 5 in comparison to results when validated against NValCL 4.

This is very interesting. This condition that the accuracy of the response fits for ouput y_2, in comparison to y_1, are more sensitive to the magnitude of the disturbance signals and output ranges used in generating models and validation data sets affirms the larger extent of non-linear dynamics associated with output y_2. This is a definitive characteristic of the system and was discussed in section 6.1.3.

It must however be said that differences between response fit results validated against NValOL 1 and NValOL 2 to those between NValCL 4 and NValCL 5 are not entirely comparable. This is since the difference in dynamic region, over which the data was generated, between NValOL 1 and NValOL 2, was specified by differences in the magnitudes of the input signals disturbing the system (since they were open-loop experiments). In the case of the closed-loop experiments, the difference in dynamic region was specified by setpoint ranges, thus the experimental range was defined by the output range. This means that the effect of increasing the disturbance signal magnitudes in the open-loop experiments is not directly comparable to that of increasing the magnitude of disturbance signals of the closed-loop experiments since the former is done via the input signal and the latter is done via the setpoint signal. However, looking at figure 6.29, which compares the input signals used to generate the open-loop validation sets, NValOL 1 and 2 (figure 6.29(a)), and the input signals used by the controllers in generating the closed-loop validation sets, NValCL 4 and 5 (figure 6.29(b)), the ranges are very similar.

Bias and Discrimination by Step Disturbance Validation Sets - Disturbance Signal Magnitude Sensitivities

The results generated by validation set NValOL 3 were extremely interesting. Looking at the results when using validation set NValOL 3 (figure 6.27), characterised as the open-loop step disturbance experiment, all the identified models performed very poorly.

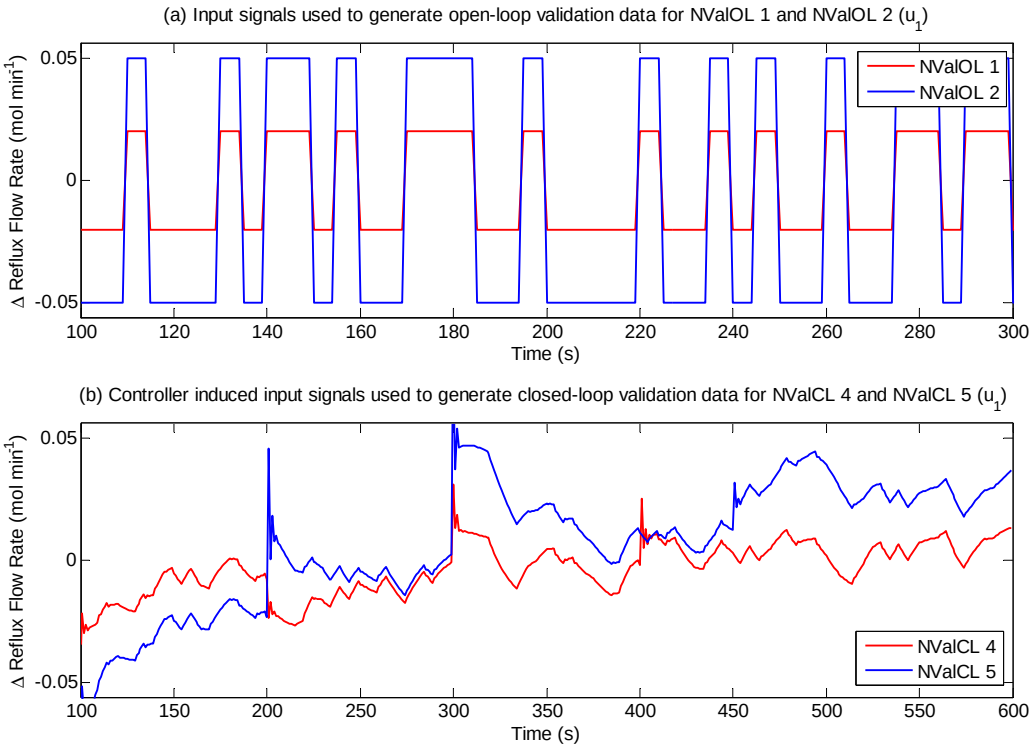


Figure 6.29: Comparison between input signals generated under open and closed-loop conditions for normal and larger disturbance ranges

The exception to this statement being the fits produced by the models generated from experiments NCL 4 and NCL 13 for output y_1 and NCL 4 for y_2. It is interesting to note that these experiments are those characterised as having larger disturbance signals.

This greater discrimination between models based on the region over which they were identified, and consequently the exposure to non-linear dynamics, proved to make validation set NValOL 3 very informative and revealing in terms of allowing insight into the non-linear system identification problem. This is specifically the case for response fits for output y_2 (figure 6.27(b)). Output y_2 is of is of particular interest as it exhibits the larger sensitivity to input magnitude changes. Section 6.1.3, in describing the dynamics of the true system - system B, characterised the dynamic behaviour of output y_2 as having an inverse response given sufficiently large disturbances. Looking at the simulated responses for output y_2 of the identified models, figure 6.30, when validated against validation set NValOL 3, it is evident that most of the identification experiments generated models that did not identify the inverse response. This does indicate that the identification experiments designed to disturb system B so as to generate data for identification, did not all force the inverse response.

The experiments that did allow for the identification of the inverse response were NOL 3, NOL 6, NCL 4 and NCL 6 with NCL 4 having the best fit. Note how NCL 4 is the only experiment of the 4 characterised by larger disturbance signal ranges. Additionally,

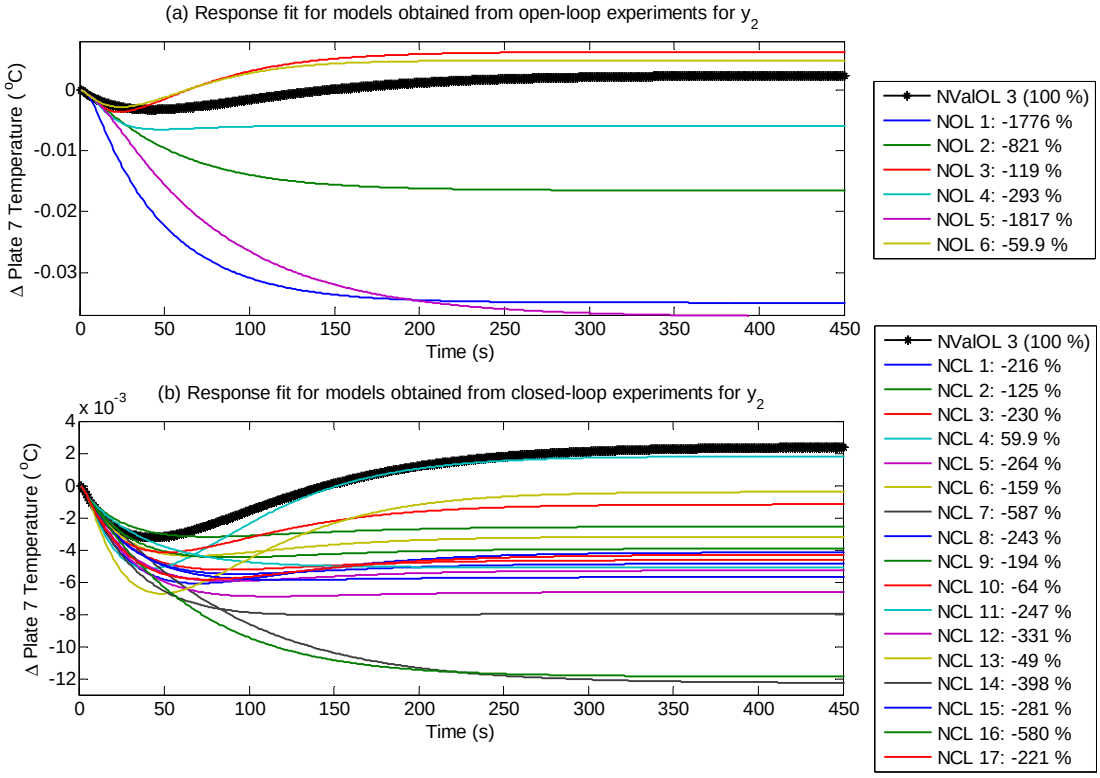


Figure 6.30: Simulation validated against open-loop step response validation data set NValOL 3

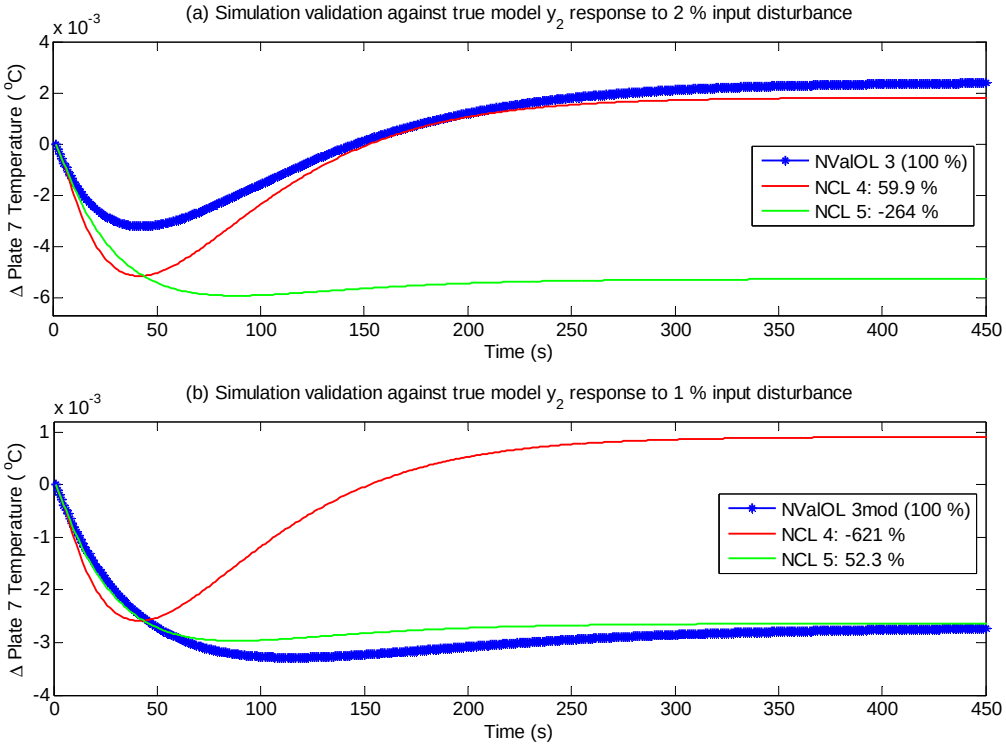


Figure 6.31: Illustration of the characteristic problem in linear model approximation of a non-linear system

looking at tables 6.6 and 6.8 one notes the experimental conditions of NOL 6 are the same open-loop step disturbance conditions as those used to generate validation set NValOL 3 but with a larger step disturbance magnitude. Figure 6.30(a) shows how the model generated by experiment NOL 6 produced a more pronounced inverse response than the validation set response. Recalling section 6.1.3, it was established that the larger the step disturbance, the more pronounced the inverse response in output y_2 for system B. Since experiment NOL 6 used a larger step disturbance magnitude than that used in validation set NValOL 3, the inverse response used to identify the model was more pronounced, thus the result in figure 6.30(a).

Figure 6.31(a) shows the same simulated responses from the models obtained from experiments NCL 4 and NCL 5 given in figure 6.30, validated against the same NValOL 3, isolated from the others for better visibility. Here it is clearly seen that the experimental conditions of NCL 4 allowed for the identification of the inverse response, while those of NCL 5 did not. Below figure 6.31(a), figure 6.31(b) shows the simulated responses generated by the same identified models but validated against a modified version of validation set NValOL 3, NValOL 3mod. The only difference being that the while NValOL 3 was generated using a 2% step disturbance magnitude, NValOL 3mod was generated using a 1% step disturbance magnitude. It is noted that the reduction in step magnitude of the signal disturbing the true non-linear system, system B, produced a much less pronounced inverse response. More importantly, it is noted that the linear models identified from experiments NCL 4 and NCL 5 maintain the same dynamic characteristics in their responses (as per definition of a linear system). The model generated from NCL 4 performed very well when validated against NValOL 3 but poorly against NValOL 3mod and vice-versa for NCL 5. .

This shows how a poor model may accurately fit a validation set response if that validation set is not appropriately chosen and obtained from experiments appropriately designed so as to generate informative data. Through this the importance of generating validation data, and hence designing experiments, based on the defined dynamic region over which the identified model must satisfy its purpose is emphasised, especially when approximating a non-linear system with a linear model

Other Experiment Condition Sensitivities

In analysing the sensitivities of the identified models to experimental variables other than disturbance signal magnitudes, it must be said that the open-loop validation sets showed little ability in discriminating based on the variances of such variables. The closed-loop validation sets however seemed to be able to better discriminate between models based on the effects of such variances on these model. This is clearly seen when comparing the response fit profiles validated against the open-loop validation set NValOL 1 in figure6.27

to the profile when validated against the closed-loop validation set NValCL 4 in figure 6.28. The larger variance in response fit results when validated against the closed-loop validation set is clearly evident.

This implies the closed-loop validation set better discriminated between the models identified from different experimental conditions. Thus the following observations regarding the identification experiment variable sensitivities are predominantly made relative to the response fits validated against the closed-loop validation set NValCL 4 (figure 6.28):

- Looking at the experimental conditions of identification experiments NCL 10 - NCL 16, it is noted that they are respectively precisely the same as those of NCL 1 - 8 but with larger dither signal magnitudes, effectively giving these experiments larger signal-to-noise ratios. The effect of increasing the signal-to-noise ratio only presented itself in the models identified from the experiments where no setpoint changes were incurred. That is, the difference between experiments NCL 7 and NCL 16. Experiment NCL 16, different to NCL 7 only in the larger signal-to-noise ratio, generated a model with significantly better response fits when validated against the closed-loop validation set NValCL 4 in comparison to NCL 7. This further reveals the importance of setpoint changes in closed-loop identification experiments since it alleviates the strain off the dither signal in exciting the system. That is, without setpoint changes, dither signals need to be much more persistently exciting to allow for the same level of informative data generation.

- With respect to the different types of disturbance signals and their effect on the accuracy of the identifed model. The White Gaussian signals were expected have greater capacity to persistently excite non-linear systems (see section 5.3.2), this was not the case as the model identified from the open-loop experiment which used white Gaussian disturbance signals, case NOL 1, performed worst of all the modes when validated against the closed-loop validation set. Looking at the other models obtained from open-loop experiments, no clear trend regarding the model's performance relative to the disturbance signal may be made since there was little consistency in results between the fits for output y_1 and y_2.

- In terms of the models generated from closed-loop experiments and their sensitivity to the disturbance signals used, the relevant experiments are those of NCL 3 and NCL 8. Comparing the response fits of the models generated from experiment NCL 3, which used a PRBS disturbance signal, to that of NCL 8, which used a PRMLS disturbance signal, it is noted that the PRBS generated the slightly better fits.

- Looking at the effect on model response accuracy when not using a dither disturbance signal (case NCL 9) and comparing this to the effect of not implementing

setpoint changes (case NCL 7), it is found that the dither signal has a smaller contribution in persistently exciting system B. This is as the model identified from case NCL 7 had a much worse response fit in comparison to that of the model generated from NCL 9.

- Experiment NCL 5 is characterised as having varied controller parameters and experiment NCL 6 is characterised as the same condition as NCL 5 but with constrained inputs. It is found from their insignificant difference in response percentage fit values (specially for output y_1) relative to each other and to that of case CL 3 (no controller parameter variance and input constraints) that the variance of controller parameters and implementation of input constraints had little effect on the extent of informative data generation.

In addition to the observations made, it must be noted that there was no significant difference found between the validation results obtained from simulation and those obtained from prediction.

6.4.3 Frequency Content Analysis

In the case of the of system A, the linear system, the frequency analysis presented in section 6.3.4 was done such that the true model's frequency response was assessed (bode plots) and compared to the frequency responses of the identified models. In addition to this the frequency content of the validation data was compared to the true frequency response, this allowed for an indication of content of the validation data and the extent at which the validation set represented the true system. In the case of the frequency analysis of system B, presented in this section, an assessment is made into the frequency content of the validation data and the simulated responses when validated against each validation set.

As was established in the previous section, models generated from experiments where larger disturbance signals were used had more accurate response fits than when validated against the validation sets that were generated from similarly large signals, these being sets NVaOLl 2 and NValCL 5. This condition was interpreted as an indication of the nonlinearity of the system being identified (system B). Figures 6.32(a)-(b) and 6.32(c)-(d) respectively show the differences in the input signals used to generated validation data for the open and closed-loop experiments. It clearly confirms the designed experiment conditions in that the only difference between the signals for the respective open and closed-loop experiments is the signal magnitude. Figure 6.33 on the other hand shows the frequency content of the response of the true system, system B, to these input disturbances, these are thus the frequency representations of the outputs generated for the validation data sets.

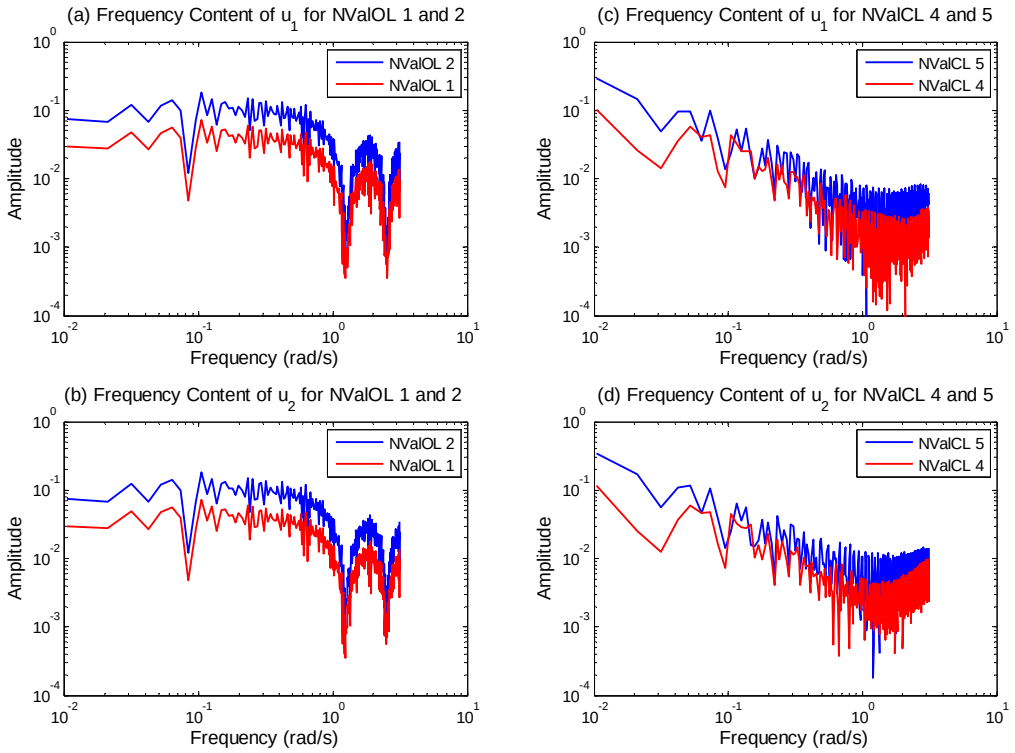


Figure 6.32: Comparison of input signal frequency content between normal input range and larger input range validation sets.

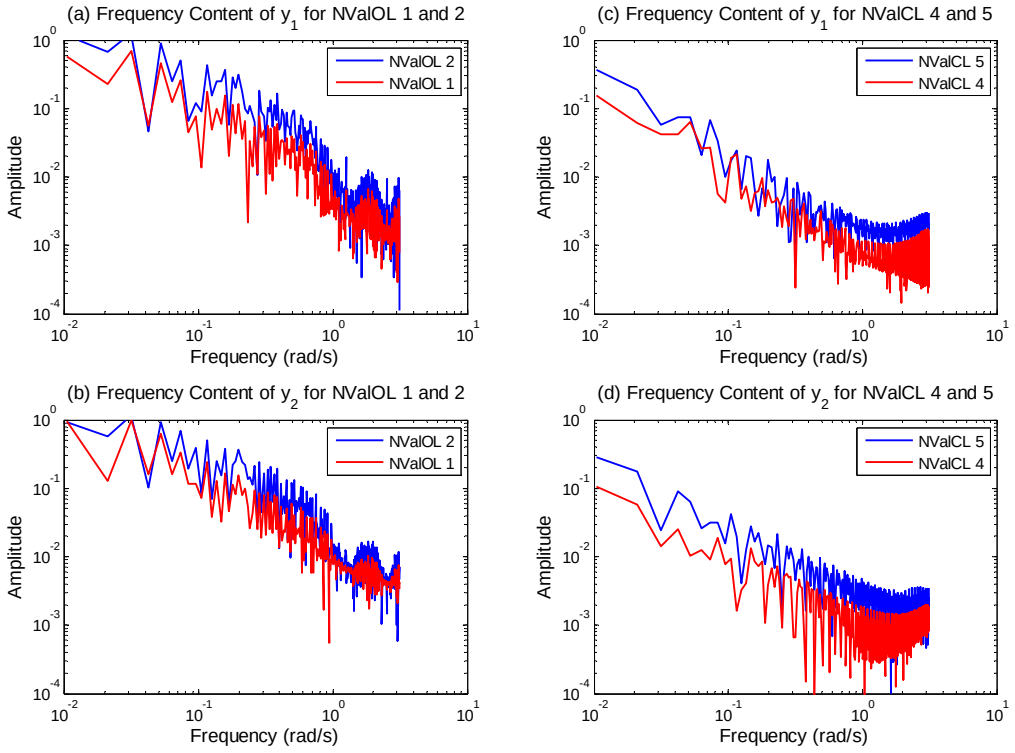


Figure 6.33: Comparison of output signal frequency content between normal input range and larger input range validation sets.

Analysing the differences between these responses, the first observation is that the frequency content of responses for the open-loop experiments (figures 6.33(a)-(b)) are less distinct than those of the closed-loop experiments (figures 6.33(c)-(d)). That is, the effect of disturbing system B with a larger input signal, showed very little effect on the frequency content of the response for the open-loop experiments. Looking at figures 6.33(c)-(d) it is observed that the frequency content of the responses obtained from the closed-loop experiments are slightly more distinct from each other. This signifies that the increase in disturbance signal magnitude under closed-loop conditions produced a more distinct response than under open-loop conditions.

These findings imply that the validation sets obtained from closed-loop experiments should do a better job at discriminating between models based on the range of data that each model was identified from. Recalling the response percentage fit results for the identified models when validated against the open-loop data sets and the closed-loop data sets given by figures 6.27 and 6.28 respectively, this is exactly what is seen. This is best noted when comparing the response fit results when validated against the open-loop validation set NValOL 1 to those when validated against closed-loop set NValCL 4. As has been discussed before, the dips in the response fit profiles the fits of models obtained from experiments that used larger signal disturbance signals in comparison to the others. The corresponding experiments are NOL 4, NCL 4 and NCL 13. It is noted that these dips are more pronounced when looking at the profile when validated against the closed-loop set NValCL 4.

From this apparent improved ability to discriminate between models based on the range of data that they were identified from, it may be stated that the closed-loop experiments did a better job at generating data that better discriminated between models based the region of data that they were identified from and consequently how well they approximate the non-linear dynamics presented in the validation set.

6.4.4 Residual Correlation Analysis

The residual correlations from the identified models approximating the non-linear system (system B) were assessed in the same manner as was done for the linear system (system A) in section 6.3.6. Correlation tests were done against all the validation sets besides that generated from the open-loop step response, NValOL 3. In addition to this, higher order correlation test were implemented on the identified models in efforts to obtain higher levels of model discrimination based on how well they approximate the non-linear dynamics of system B.

Whiteness Test Results

The primary observation made while testing for whiteness was that all the models generated showed large extents of residual correlations regardless of the validation set used. Neither of the validation sets obtained from open or closed-loop experiments could be attributed with showing for better model discrimination by producing different correlation profiles for the different identified models.

The residual correlation profiles of each model when validated against the closed-loop validation sets (NValCL 4 andOL 5) almost all resembled that shown in figure 6.34 revealing the results for the model identified from experiment NCL 8.

When validated against the open-loop validation set the residual correlation profiles obtained for most of the models differed very little from that presented in figure 6.34. Some however did seem to indicate slightly different residual correlation profiles. These were those of the models obtained from experiments NOL 6, NCL 7 and NCL 16. With experiment NOL 6 being characterised as the open-loop step disturbance experiment, NCL 7 and NCL 16 being characterised as the closed-loop experiments where no setpoint changes were imposed with NCL 16 having the larger signal-to-noise ratio.

Figure 6.35 illustrates the residual correlation plots for the model obtained from the open-loop step disturbance experiment, NOL 6, validated against open-loop validation sets NValOL 1 and NValOL 2. It is interesting to note that validation set NValOL 1 did not reveal any difference in residual correlations for output y_1 to that obtained from set NValOL 2 (figure 6.35(a)). In contrast to this a slight difference between the residual correlation profiles obtained between the two validation sets was revealed for output y_2 (figure 6.35(b)). This suggests that the two validation sets detected the same extent of unmodelled dynamics for output y_1 but different extents of unmodelled dynamics for y_2. This is also seen in figure 6.34. Given that the model being tested was obtained from an open-loop step disturbance experiment, this attests to the step disturbance input signal's ability to show different dynamic responses for output y_2.

It must be said that while some slight differences in residual correlations where found, these differences were very small. Almost all of the residual correlation profiles obtained were very similar to each other, regardless of the validation set used. Thus the whiteness test shows that all the identified models had large extents of unmodelled dynamics. Furthermore the lack of distinction in residual correlation profiles indicates that all the validation sets did a bad job at discriminating between the models and consequently could not verify whether some models better approximated the non-linear system (system B) than others.

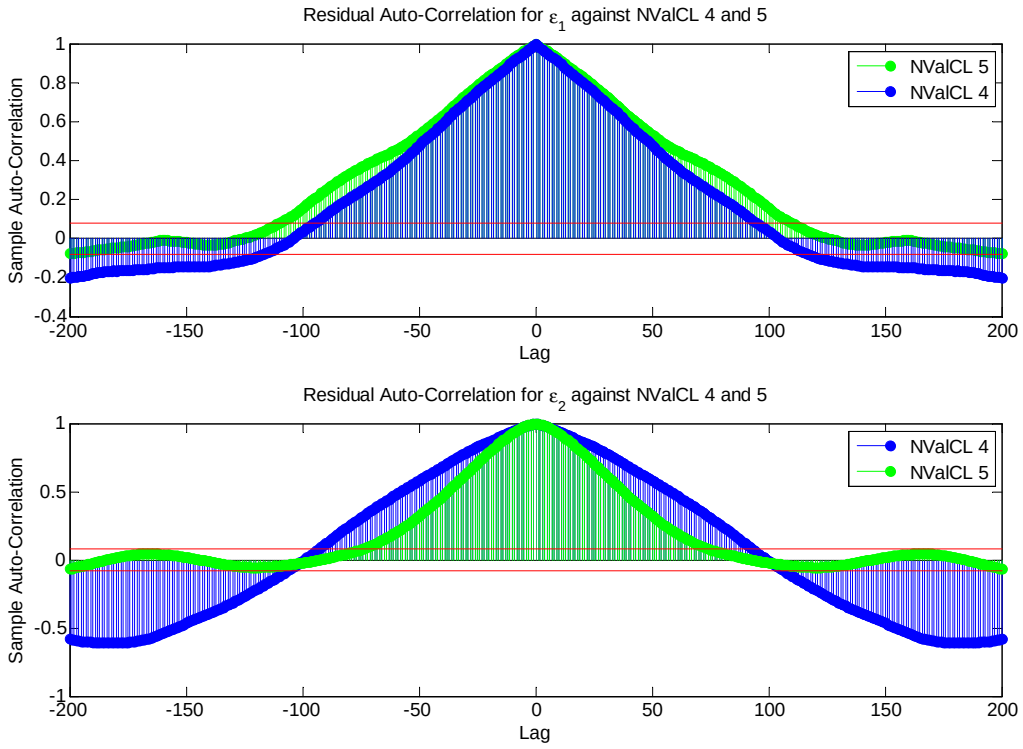


Figure 6.34: Residual correlation results for case NCL 8 using NValCL 4 and NValCL 5 as validation data

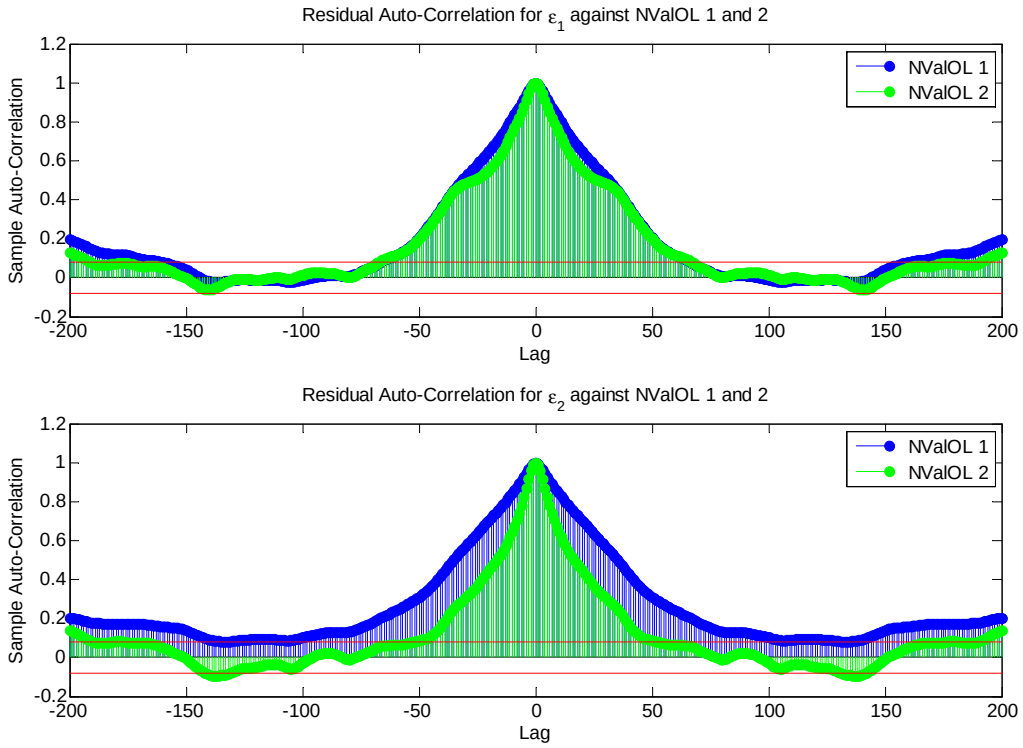


Figure 6.35: Residual correlation results for case NOL 6 using NValOL 1 and NValOL 2 as validation data

Independence Test Results using Closed-Loop Validation Sets

In testing for the independence between inputs and residuals for the different models generated by the different experiments and validated against the different validation sets, it was noted that the results closely reflected the findings for the whiteness tests. All the models generated showed high levels of residual-input correlations regardless of the validation data set used. In addition to this it was found that most correlation profiles changed very little between cases when validated against the different validation sets. This suggests all the models showed large extents of incorrectly modelled dynamics and that the validation sets did little to discriminate between most of the models based on differences in the incorrectly modelled dynamics.

Figures 6.36(a), (b) and (c) (left column) show the input-residual correlations for the models obtained from open-loop experiments NOP 1, 3, 5 and figures 6.36(d), (e) and (f) (right column) show correlations of those obtained from closed-loop experiments NCL 1, 2, 11 between residual ϵ_1 and input u_1 validated against the two closed-loop validation sets (NValCL 4 and 5). While there are some small differences in the input-residual correlation profiles, the figure clearly illustrates the point made regarding the little difference between profiles when validated against the closed-loop data. To be more specific however, there are differences in correlation profiles between models identified from open-loop cases (left column) and those identified from closed-loop cases (right column) but within the models generated under the same feedback condition there isn't much to discriminate between. This suggests that all the models obtained from the same feedback condition contain the same incorrectly modelled dynamics.

While it has been established that there is little to discriminate between the correlation results of most models when validated against the closed-loop data, figure 6.37 shows the correlation profiles of the models that did distinctly show different correlation results against the closed-loop validation sets. It is interesting to note that even though the correlatin profiles presented here are different to the general trends they are very similar to each other. These correlation profiles are those of the models obtained from the open-loop step disturbance experiments (NOL 6 - figure 6.37(a)) and the two experiments generated from closed-loop experiments where no setpoing changes were incurred (NCL 7 and 16 - figure 6.37(b) and (c)). Recall that the only difference between NCL 7 and 16 it that NCL 16 had a larger signal-to-noise ration. Note how the lack of distinction in correlation profile suggest that the difference in signal-to-noise ratio did not affect the extent of incorrectly modelled dynamics.

It is at this point appropriate to recall that input-residual correlations may not be completely trusted since some correlations may be due to the feedback loop being closed and not so much due to incorrectly modelled dynamics.

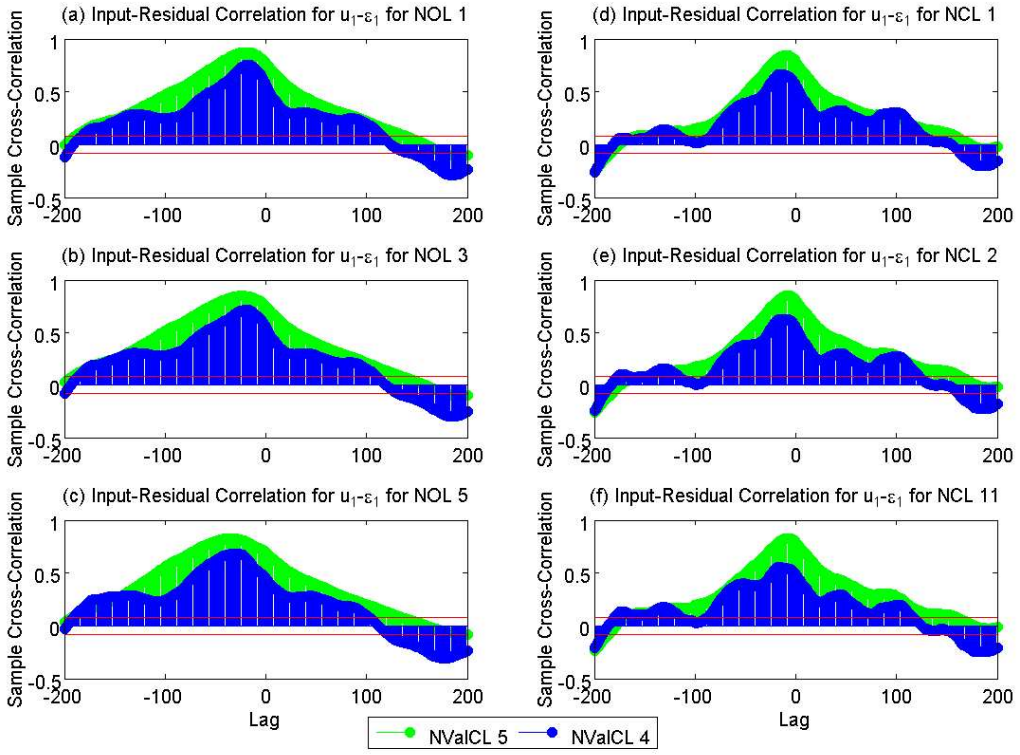


Figure 6.36: Input-Residual correlation between u_1 and ϵ_1, validated against closed-loop data, for some of the cases showing similar results

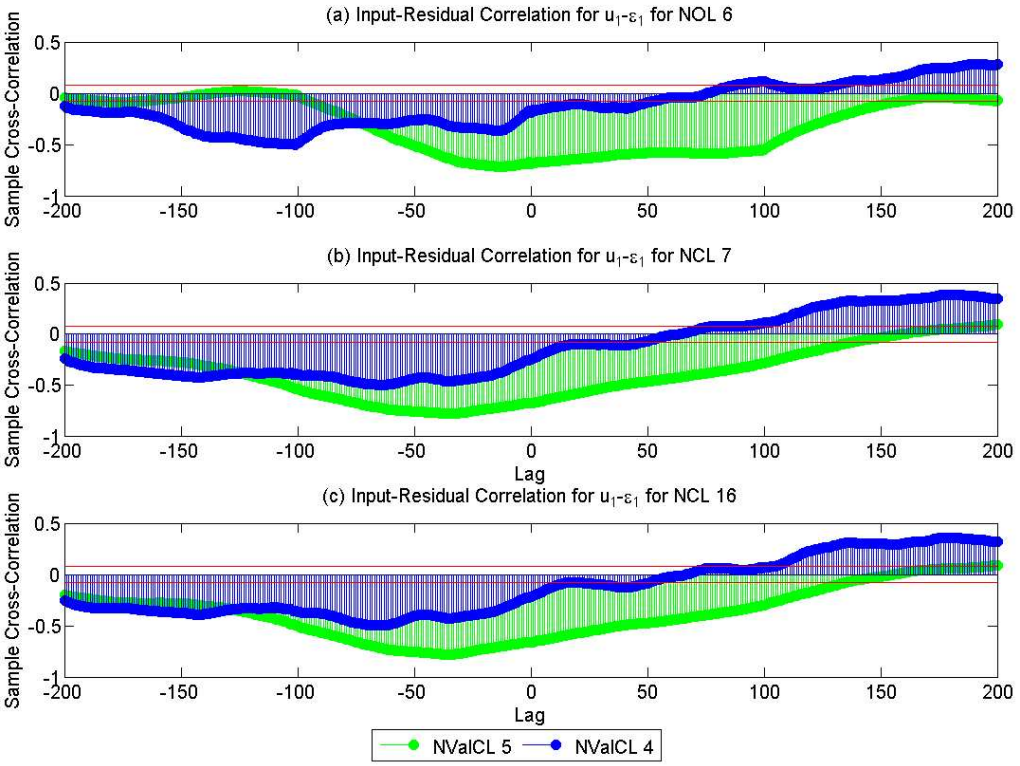


Figure 6.37: Input-Residual correlations between u_1 and ϵ_1, validated against closed-loop data, for the cases showing results different correlations to the general trend

Independence Test Results using Open-Loop Validation Sets

The residual-input correlation results obtained when using open-loop validation sets indicated the open-loop validation sets' slightly better ability to discriminate between models based on incorrectly modelled dynamics through input-residual correlation analysis.

Figures 6.38 and 6.39 are those very same cases assessed against the closed-loop validation sets in figures 6.36 and 6.37 but validated against the open-loop validation sets. Looking at figure 6.38, it is first observed that the magnitude of the correlations is smaller, additionally it is evident that the models generated from the open-loop experiments are better discriminated between when validated against the open-loop validation sets than the closed-loop sets (shown in figure 6.36(a)-(c)). The correlations of models obtained from closed-loop cases remain however indistinguishable.

Comparing the results shown in figure 6.39 to 6.37, the model obtained from open-loop step responses, case NOL 6 figure (6.39(a)), shows completely different correlations when validated against the open-loop sets to that obtained when validated against the closed-loop validation sets. The same extent of incorrectly modelled dynamics with respect to the two open-loop validation sets is revealed in contrast to two different correlation profiles in figure 6.37(a).

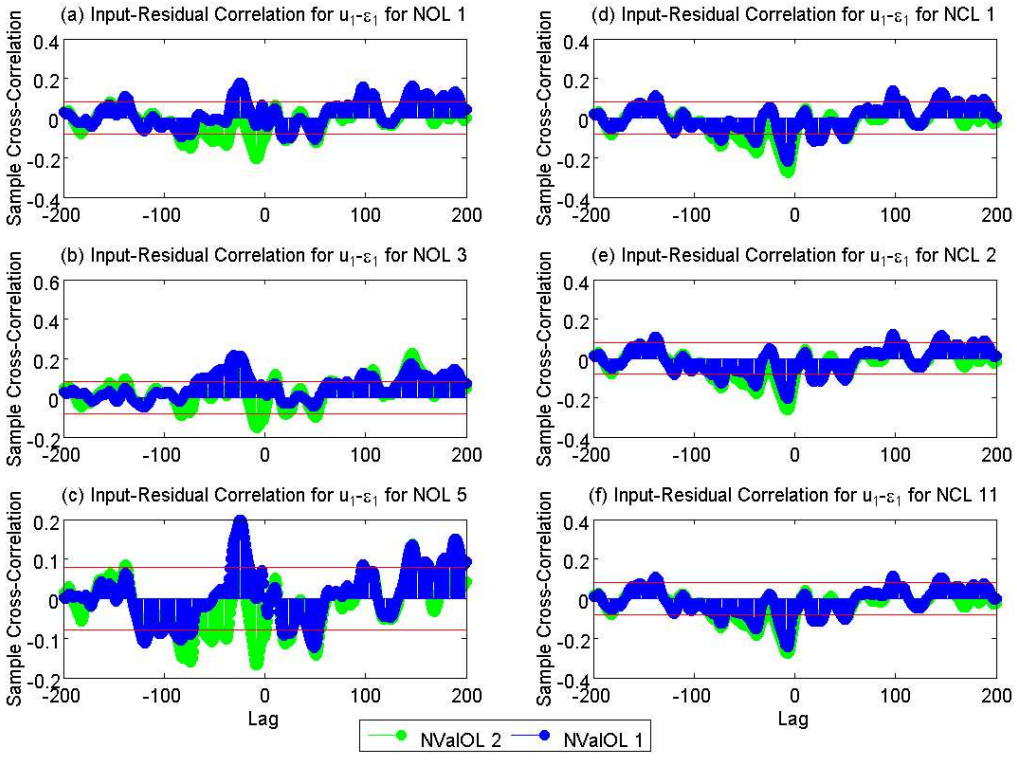


Figure 6.38: Input-Residual correlation between u_1 and ϵ_1, validated against open-loop data

The results of cases NCL 7 and 16 (figures 6.39(b) and (c)), revealed correlation profiles very similar to the general profiles obtained from models generated from closed-

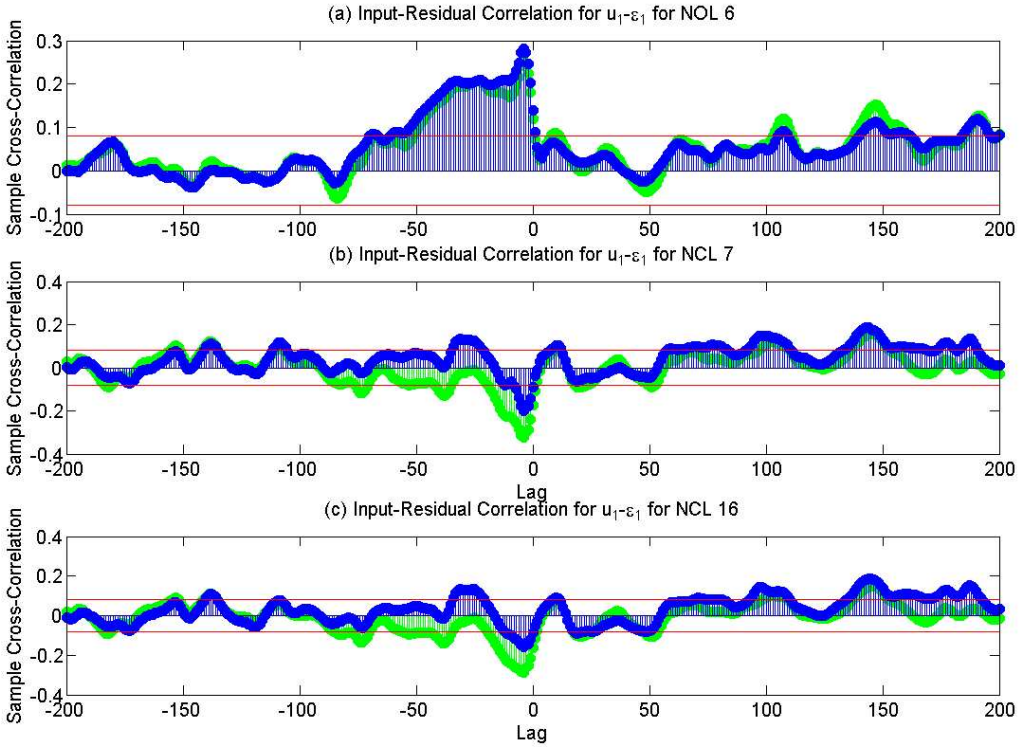


Figure 6.39: Input-Residual correlations between u_1 and ϵ_1, validated against open-loop data

loop experiments (figures 6.38(d), (e) and (f)). In contrast to the correlations obtained when using the closed-loop validation sets, it is implied that the open-loop validation sets did not detect different extents of incorrectly modelled dynamics from these to models in comparison to the other models. However, in a similar fashion as the closed-loop validation sets, the open-loop validation sets did not show an distinction in correlations that suggests different dynamics identification due to different signal-to-noise ratios.

Independence Test Results - Other Observations and Comments

In investigating the differences in input-residual correlations between the different pairs of residuals and inputs ($u_1 - \epsilon_1$, $u_2 - \epsilon_1$, $u_1 - \epsilon_2$, $u_2 - \epsilon_2$). It was observed that all but one of the models generated results with the correlation trends being near identical between the different residuals and the same input (eg. $u_1 - \epsilon_1$ and $u_1 - \epsilon_2$) but dissimilar between the different inputs and a residual (eg. $u_1 - \epsilon_1$ and $u_2 - \epsilon_1$).

Figure 6.40 shows all the input-residual correlations for the model identified from experiment NOL 5 illustrating this point. In contrast, the only model to show correlation result trends that were very similar for all input-residual pairs was that obtained from experiment NOL 6, shown in figure 6.41, characterised by the open-loop step disturbance experiment. The same was observed when using the open-loop validation set.

This observation is difficult to explain. Its implication is that the open-loop step

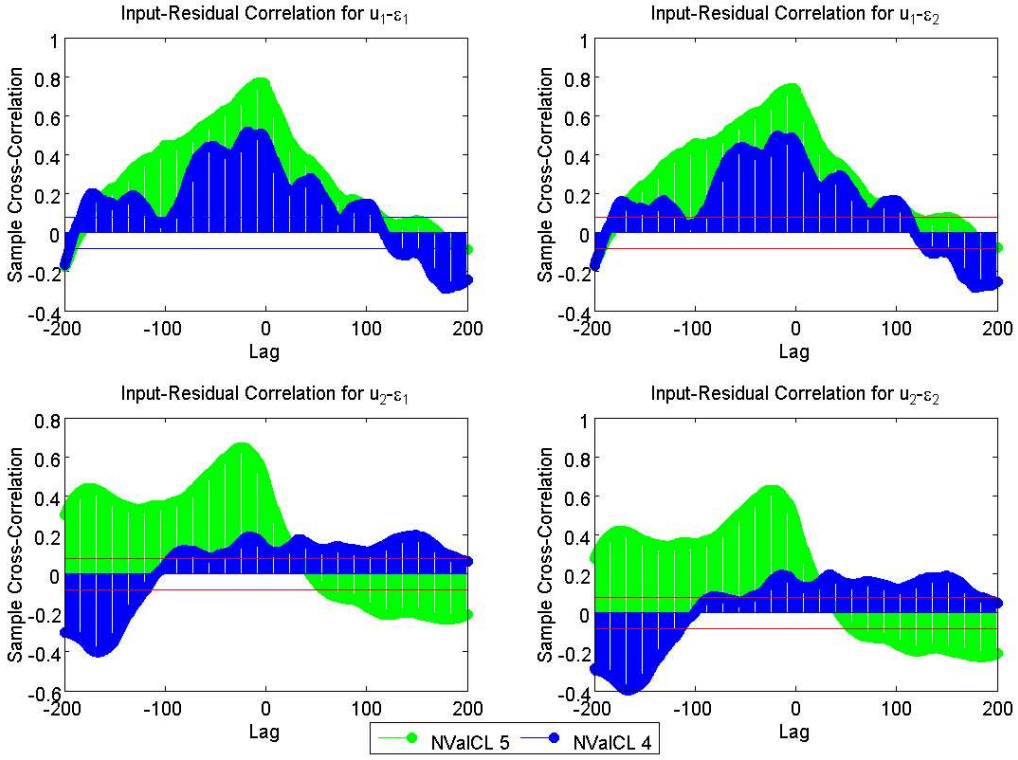


Figure 6.40: Input-Residual correlations for case NOL 5 showing no common trend between correlations

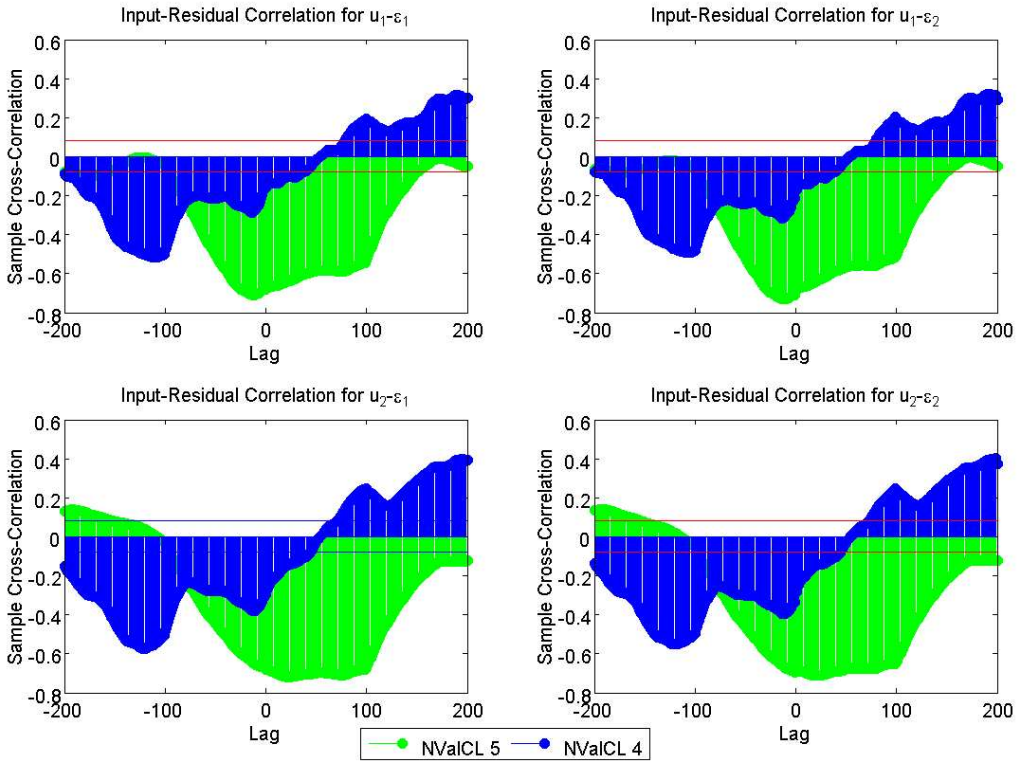


Figure 6.41: Input-Residual correlations for case NOL 6 showing common trends between correlations with between residuals and the same inputs

disturbance experiment, NOL 6, was the only experiment to generate a model with the same incorrectly modelled dynamics for both outputs while all the other models had incorrectly specified dynamics dependent on the output.

Higher Order Correlation Results

As explained in section 4.2.2, the higher order correlation tests are expected to provide a higher order of model discrimination since each test discriminates the model according to a different type of non-linear characteristic. Application of these tests to the identified models and validated against the different validation sets gave some interesting results.

It is appropriate at this point to refer to table 4.1 which indicates the significance of the correlation results. With the previous assessment into the correlations between inputs, u, and the model residuals, ϵ, showing that all the cases have non-zero input-residual correlations ($R_{u\epsilon}(\tau) \neq 0$), it is evident that $R_{u^2\epsilon^2}$ will be non zero for all cases (as per table 4.1), with only $R_{u^2\epsilon}$ to test for indications of whether the entire model is inadequate or whether the model is only inadequate at representing the odd powered non-linear dynamics.

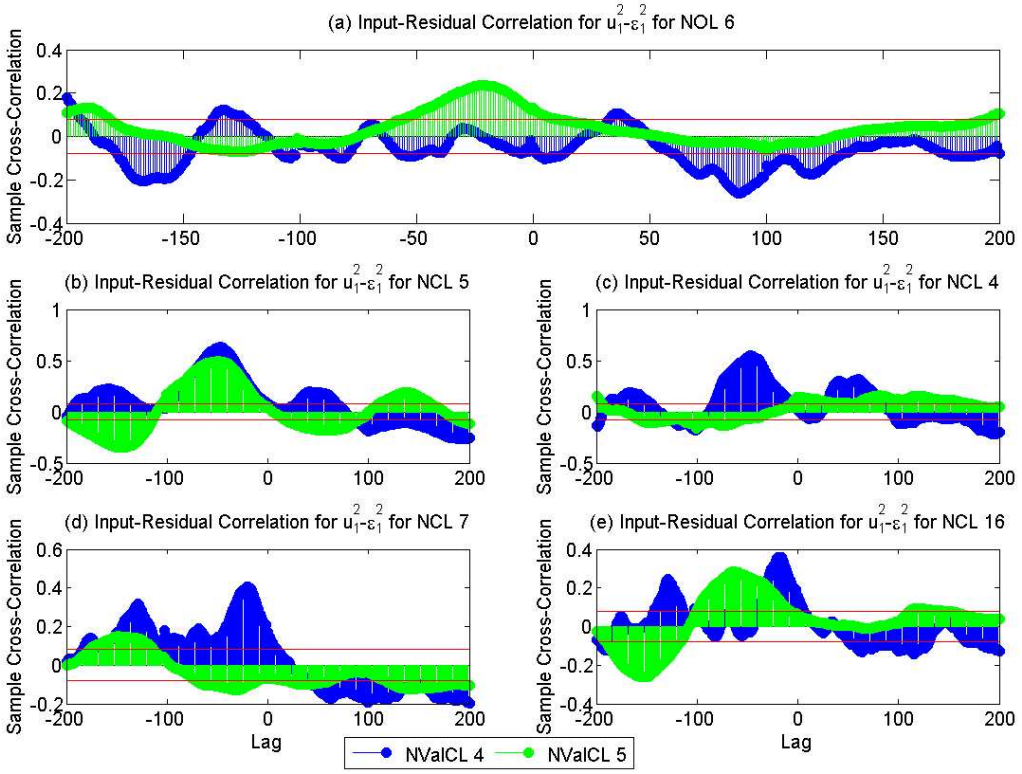


Figure 6.42: Higher order correlation: $R_{u^2\epsilon^2}$, of a select group of cases validated against closed-loop data

Figure 6.42 shows the higher order cross-correlation, $R_{u^2\epsilon^2}$, of representative models validated against closed-loop validation sets. It is observed that the results are non-zero

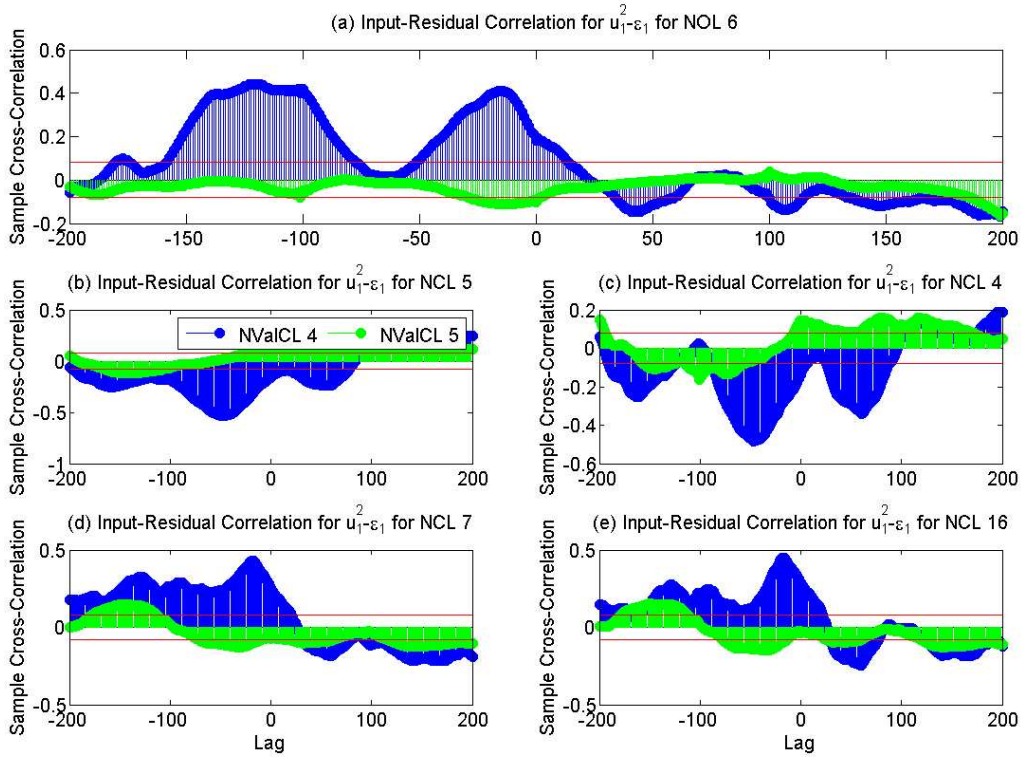


Figure 6.43: Higher order correlation: $R_{u^2\epsilon}$, of a select group of cases validated against closed-loop data

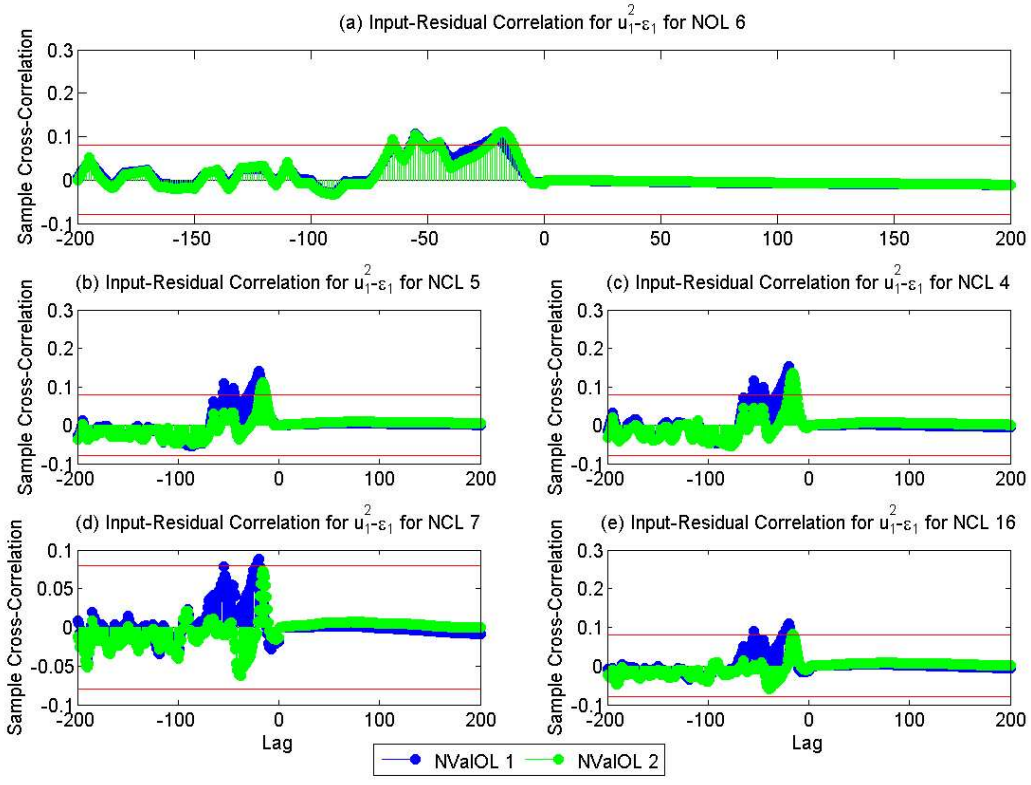


Figure 6.44: Higher order correlation: $R_{u^2\epsilon}$, of a select group of cases validated against open-loop data

for all the cases and that the two closed-loop validation sets differently discriminated between the models. Figure 6.43 shows the results of $R_{u^2\epsilon}$ similarly validated against the closed-loop validation sets, it is noted that while there are significant correlations when using validation set NValCL 5, the correlations are slightly less pronounced when using NValCL 4. This suggests that validation set NValCL 5 finds the models generated by these cases to be completely inadequate while NValCL 4 finds the models to be inadequate more so in accurately modelling odd non-linearities. This is as per the rules given in table 4.1. From this one may deduce that the two different validation sets discriminate between models on different terms.

When assessing the higher order correlations using open-loop data for the same cases, it is found that the correlations are not as pronounced. Figure 6.44, shows the results for the $R_{u^2\epsilon}$ assessments, the results obtained for $R_{u^2\epsilon^2}$ show very similar profiles. While the results show smaller correlation magnitudes in comparison to the correlations obtained when using closed-loop validation sets determining whether they are small enough to be considered negligible is difficult

From these results it must be said that both higher order correlations tests, $R_{u^2\epsilon}$ and $R_{u^2\epsilon^2}$, revealed significant correlations with the results obtained from the open-loop validation sets being slightly debatable. When using the closed-loop validation sets these correlations where larger in magnitude and more distinct correlation profiles were produced for each model than when using the open-loop validation sets. This implies that the closed-loop validation sets better discriminated between models and indicated larger extents of incorrectly modelled dynamics.

However, since the closed-loop validation sets are of questionable accuracy in input-residual correlation tests due to the inherent correlations caused via the feedback condition, the results may not be entirely trusted. Looking at the results obtained when using the open-loop validation sets, it must be said that all the models did exhibit correlations (although small in magnitude) implying that the models are inadequate in both the noise model and the dynamics model. Furthermore, the open-loop validation did not produce higher order correlation results for each model that were very distinct form each other. This means that the higher order correlation tests did not improve the ability to discriminate between models or give insight into how well each model approximated the non-linear system.

6.4.5 Results Overview

The results and findings concerning the identification and validation efforts on system B, the non-linear system, may be summarised as follows:

Model Variance and Uncertainty

Variance due to Noise Sensitivity

- Models generated from closed-loop experiments where no setpoint changes were incurred were found to be most sensitive to random realisations of noise.
- Increasing the signal-to-noise ratio proved to reduce this model parameter sensitivity to noise. This was however only noticed in the closed-loop experiments that did not induce setpoint changes.

Model Uncertainty

- The model that revealed the largest measure of model uncertainty was that obtained from the closed-loop experiments where no setpoint changes were incurred.

Simulation and Prediction

General Observations:

- Response percentage fit results showed that all the models were very sensitive to the validation set used.
- Closed-loop validation sets showed indications of allowing for better discrimination of models than the open-loop validation sets since they provided larger variances in response fits amongst models and outputs (y_1 and y_2)

Bias and Discrimination by PRBS Disturbance Validation Sets - Disturbance Signal Magnitude Sensitivities:

- The disturbance signal magnitude proved to be a very important variable.
- It was found that validation sets produced by disturbance signals of certain magnitudes were extremely biased towards models that were generated from disturbance signals of similar magnitudes. This was thought to reflect the non-linear nature of the dynamics of the system being identified
- The accuracy of response fits of output y_2 was found to be the most sensitive of the two outputs to disturbance signal magnitudes. This was thought to be due to output y_2 exhibiting larger extents of non-linear dynamics.

Bias and Discrimination by Step Disturbance Validation Sets - Disturbance Signal Magnitude Sensitivities:

- Validating the response fit results against the step disturbance validation set revealed the step response validation set to be extremely biased towards models identified from experiments that used disturbance signals with similar magnitudes.

- The step disturbance validation set allowed for the confirmation that the increased sensitivity of response fits of output y_2 to disturbance signal magnitudes may be attributed to larger extent on non-linear dynamics associated with output y_2.
- This validation set further revealed how using different disturbance signal magnitudes generates different response dynamics and how this can affect the validation of a model.
- This further emphasised importance of validating models against validation sets containing the dynamic behaviour most representative of the models purpose or intended use, especially when approximating a non-linear system with a linear model.

Other Experiment Condition Sensitivities:

- The closed-loop validations sets were found to better discriminate between models based on differences in identification experiment conditions used to generate the models.
- The changes in signal-to-noise ratio were only found effective in closed-loop experiments where no setpoint changes were incurred. This shows the importance of setpoint changes in closed-loop experiments.
- From investigating the effect of removing disturbance signals and not inducing setpoint changes, the greater value in setpoint changes was further established in terms of its contribution to persistently exciting the system.
- PRBS disturbance signals were found to be most consistent for closed-loop identification efforts while no disturbance signal type given open-loop conditions could clearly be stated as being best.
- No clear trend was observed in the effects of varying controller parameters and implementing input constraints.

Frequency Content Analysis

- Analysing the frequency content of the validation data sets it was found that the closed-loop validation sets were better suited for model discrimination based on the magnitude of the disturbances signals used to generate the data from which the models were identified than the open-loop validation sets.
- Since it was been established that disturbance signal magnitude greatly affects the dynamics revealed by the non-linear system, the previous point implies that the closed-loop validation sets are better at discriminating between models identified from different dynamic responses from the non-linear system.

- These finding, that the frequency characteristics of the closed-loop validation data sets indicate them to better discriminate models, were found to reflect the simulated response results.

Residual Correlation Analysis

Whiteness Test:

- Large residual correlations were found for all the models regardless of the validation set used. This implies large extents of unmodelled dynamics.
- In addition to this the correlation profiles were mostly similar for all the models. This indicates that the validation sets did little to discriminate between models based on differences in unmodelled dynamics.

Independence Test Results Using Closed-Loop Validation Sets

- In a similar fashion as the whiteness test, all the models showed large input-residual correlations indicating large extents of incorrectly modelled dynamics.
- It was found that the closed-loop validation sets only discriminated between models based on the feedback condition of the experiment from which the model was identified. Otherwise little discrimination was evident.

Independence Test Results Using Open-Loop Validation Sets

- The input-residual correlation results obtained when validated against the open-loop validation sets suggest the open-loop validation sets slightly better discriminated between models than the closed-loop validation sets.
- The input residual correlation tests could not detect any differences in modelled dynamics that may be attributed to larger or smaller signal-to-noise values.

Higher Order Correlation Tests

- The higher order correlation tests did not produce a greater capacity to discriminate between models based on how well the models approximated the non-linear system.
- When using closed-loop validation sets, higher order correlation results did indicate a greater capacity to discriminate between models relative to previous validation assessments, however, the results could not be trusted due to correlations created via the feedback that are indistinguishable from correlations due to incorrectly modelled dynamics.
- The results obtained when using the open-loop validation sets where inconclusive in that the significance of correlation magnitudes were difficult to assess. The correlations did however indicate that there open-loop validation sets did little to discriminate between models.

6.5 Discussion

This section aims to discuss the findings of the identification and validation efforts on systems A and B relative to each other.

6.5.1 Identification

Determining which experiments generated better approximations of system A, the linear system, was easier than for system B, the non-linear system. This was primarily due to the added intricacies in validating the non-linear system approximations.

With respect to disturbance signal characteristics, it can generally be said that the with the benefits or improvements in model accuracy from using signals other than PRBS being relatively little to none for both open and closed-loop experiments, it must be stated that the simple nature of PRBS disturbance signal is most applicable. This however is only the case when the frequency characteristics of the signal are near to that of the dominant time constant of the system

In terms of the effect of the feedback condition, both the linear and non-linear system analyses showed how the accuracy of the models generated from open-loop conditions were more sensitive to the properties of the disturbance signals used. While the accuracy of the models generated from closed-loop data was not so sensitive to the disturbance signal properties. This was thought to be accounted for by the fact that the feedback loop dampens the effects of the disturbance signals consequently dampening the effects of varying signal characteristics.

From this it must be said that this implies that the models identified from closed-loop experiments are less sensitive to signals that are less persistently exciting or badly designed in terms of frequency characteristics, however such models are also less capable of producing more accurate models from better designed experiments. Thus closed-loop experiments are found to allow for models that perform slightly worse but are more robust, while open-loop experiments can produce models that perform better but are less robust in terms of sensitivities to badly designed identification experiments.

6.5.2 Validation

Comparing the model cross-validation findings, specifically the effect of experimental conditions on bias and discrimination characteristics of the validation sets, is not easily done between the efforts on the linear system (system A) and non-linear system (system B).

In general it may be stated that open-loop experiments produced validation sets that better discriminated between models for the linear system while the closed-loop validation set was found to best represent the system A. With respect to the non-linear system the

validation sets obtained from closed-loop experiments were found to be most useful in discriminating between models.

In terms of the different validation techniques, it was found for both the linear and non-linear systems that response fit validations were the most sensitive to the differences between validation data thus making the results extremely relative to the validation set used. The application of residual correlation analysis on model approximating the non-linear system was found to be very ineffective. While when applied to the models approximating the linear system the results did contribute to better understanding on the discrimination abilities of the validation sets based on different extents of unmodelled dynamics.

It was quickly found from the linear system that the validation sets can seem to be biased towards models obtained from similar experimental conditions as the models. Validation sets obtained from closed-loop experiments seemed biased towards models obtained from closed-loop experiments. The same can be said about the other validation sets. This was found to become much more complex when validating linear approximations of the non-linear system. Primarily since the disturbance signal magnitude and dynamic region effects are introduced into the bias and prejudice of the validation sets.

From these issues it was make very evident that the question may not necessarily be how representative is the validation set of the true system and is it discriminating between models based on deviations from dynamic representation of the true system, but how representative is the validation set of the intended use of the model. This is especially the case if the model is of a different structure and order of the true system.

6.5.3 Other Issues

An important issue, while difficult to resolve but important to present, is that determining whether these findings are characteristic results of identification and validation efforts of linear systems and non-linear systems or characteristics of these specific systems being identified. These results may not be attributed to being absolute findings, that is, while it was evident that closed-loop experiments produced models that were less likely to produce completely inaccurate responses than the open-loop experiments, this is not necessarily a characteristic of closed-loop experiments but perhaps of the system being identified.

When assessing how sensitive each experiment and the respective identified models were to noise in sections 6.3.1 and 6.4.1, for the linear and non-linear system respectively, it was found that the experiment characterised by open-loop step disturbances was most sensitive for system A while the closed-loop experiments where no setpoint changes were incurred were most sensitive for system B. This means that the data generated from these experiments was not informative enough. Consequently the estimation procedures that used these data sets were more susceptible to modelling the noise, thus the identification

sensitivity to random noise realisations.

This is very closely related to the fact that ARX model structures were used and how these structures are prone to dynamic model corruption by noise due to an inability to independently model noise and dynamics. Recalling that literature states that closed-loop experiments are meant to be more susceptible to this problem. This makes it interesting that the assessments on the linear system found the dynamics model obtained from an open-loop experiment to most sensitive to noise. While assessments on the non-linear system found the dynamics models obtained from closed-loop experiments to be most sensitive. This might indicate that the greater sensitivity of closed-loop experiments to noise might be more so when attempting to approximate a non-linear system.

CHAPTER 7

Identification of a Distillation Column

In the previous chapter, known mathematical models and their simulated responses to simulated disturbances allowed for an investigation into aspects of identification experiment design variable sensitivities and cross-validation effectiveness. This chapter aims to extent this investigation to a 'real' system, a pilot scale distillation column. A smaller scope of experimental designs, based on the findings of the previous chapter, was used to generate data from the column for identification and validation.

7.1 Investigative Approach and System Description

7.1.1 Investigative Approach

The investigative approach of the work presented in this chapter aims to study the same concepts as those presented in the previous chapter but extended to a 'real' system. That is, the investigation into system identification and validation is now focused on a 'real' system in search of pragmatic issues and difficulties that would not have been encountered up to this point due to the controlled environment of simulated investigations. Thus the investigative approach may be stated as follows:

- Investigate how accurately the 'real' system may be identified using identification experiments and the sensitivity of such accuracy to disturbance signal and feedback conditions.
- Cross-validate the identified models using validation sets obtained from various experimental conditions. The focus is again made on the effects of varying disturbance signals and feedback conditions.

- Assess the results obtained from the different validation sets so as to obtain indications of how the different experimental conditions affect the identified model accuracy and validation set discrimination.

These objectives are pursued through the use of a pilot scale distillation column with unknown dynamics. The direct experimental objectives of this chapter may thus be further refined as follows:

1. Study the pilot scale distillation column, and establish an operating range which is to be identified. This range is to be established in consideration of normal operation of the distillation column.

2. Design identification experiments to generate data sets around this established region of operation under different experimental conditions, with a primary focus on varying feedback conditions.

3. Design experiments to generate data sets around and beyond the established region of optimal operation to serve as validation data sets.

It is important to note that objective number 1 above distinguishes the identification efforts on the distillation column from those on the simulated systems. This point implies that a specific output range is to be identified and that consequently both the open and closed-loop experiments are output constrained. This is unlike the identification experiments on the simulated systems where open-loop experiments were input constrained and closed-loop experiments were mostly only output constrained. This approach is taken in view of the results obtained regarding identification of the non-linear system in the previous chapter, system B. That is, it was found that clearly establishing the range over which a model is to approximate a non-linear system, is critical.

The following sections describe the distillation column used and further detail the identification and validation experimental conditions.

7.1.2 Description of the Pilot Scale Distillation Column

The distillation column used, represented by the detailed diagram presented in figure 7.1, is a ten plate glass distillation column which is not insulated, approximately 5 meters tall and 15 centimetres in diameter. The column is run near atmospheric pressure and is used to separate mixtures of water and ethanol. As figure 7.1 indicates, the column has two feed points, one above plate 7 and the other above plate 3. Furthermore, the system is a closed system in that the distillate and bottoms products are mixed and fed back into the system.

The column boiler is fed with saturated steam from electric boilers with a supply pressure between 600 and 700 kPa (gauge) and is throttled down to 40 - 100 kPa (gauge).

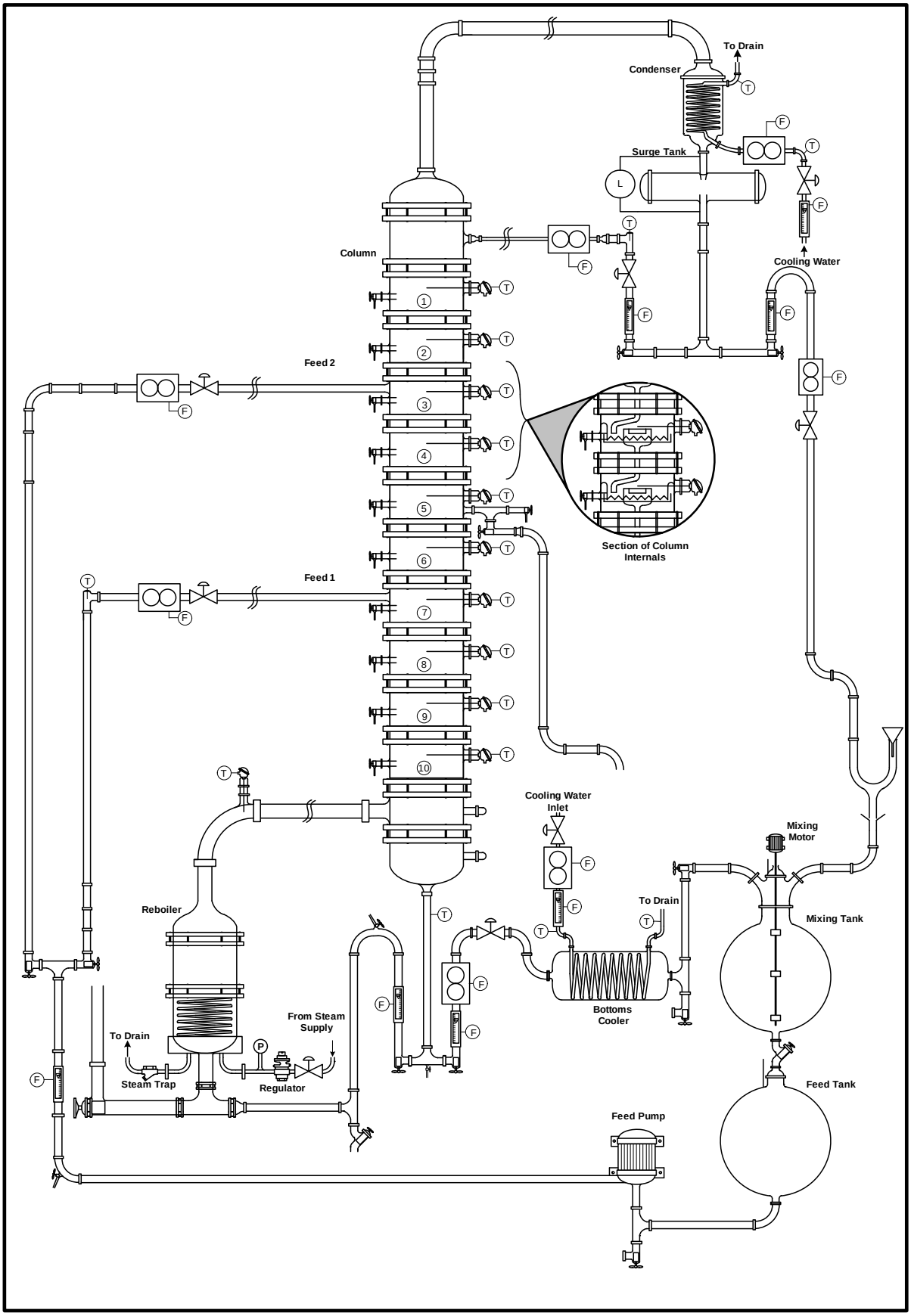


Figure 7.1: Detailed illustration of the pilot scale distillation column

The throttling is done by a pneumatic control valve. The condenser, which is made of glass, is fed with municipal water at ambient temperature. The reflux drum level, measured by means of differential pressure of the liquid head, supplies the pressure head for the reflux and distillate streams.

Equipment and Instrumentation Description

Figure 7.2 illustrates the piping and instrumentation of the column. The following may be said regarding the instrumentation and measurements:

- Pneumatic control valves are placed on all the streams entering and leaving the column.
- Temperature measurements are taken from Foundation Field Bus instruments on each plate as well as on all feed and product streams.
- Gauge pressure measurements are taken for the steam in the boiler after throttling.
- Level measurements are taken for the boiler and reflux drum from pressure differential measurements of the liquid head.
- Mass flowrate measurements are taken for the boilup, bottoms product, reflux, cooling water, feed and top products streams via Micro Motion Coriolis flowmeters.
- A gauge pressure measurement is made at the top of the column

It is important to note that the flowrate measurement devices cannot accurately measure any flowrate smaller than 3 kg/hr. In addition to this, on occasions when the steam pressure is increased very quickly, the pressure in the column increases momentarily, this causes the pressure in the reflux drum and boiler to increase causing the values of the pressure differentials used to infer level to increase momentarily.

All the instrumentation and equipment are connected and operated via a DeltaV distributed control system. It is important to note that the sample rate of all the measuring devices is once every second and that all data is stored in a DeltaV continuous historian.

Operation and Operating Ranges

A key issue in operating the distillation column is maintaining the top plate temperature above the water-ethanol mixture azeotrope which is at 78^oC. This introduces constraints or limits on the amounts of energy addition and removal from the system via the boiler and condenser. Given that the feed is sub-cooled as it is fed from a tank kept at atmospheric conditions, the energy introduced via the boiler is closely related to the feed flowrate.

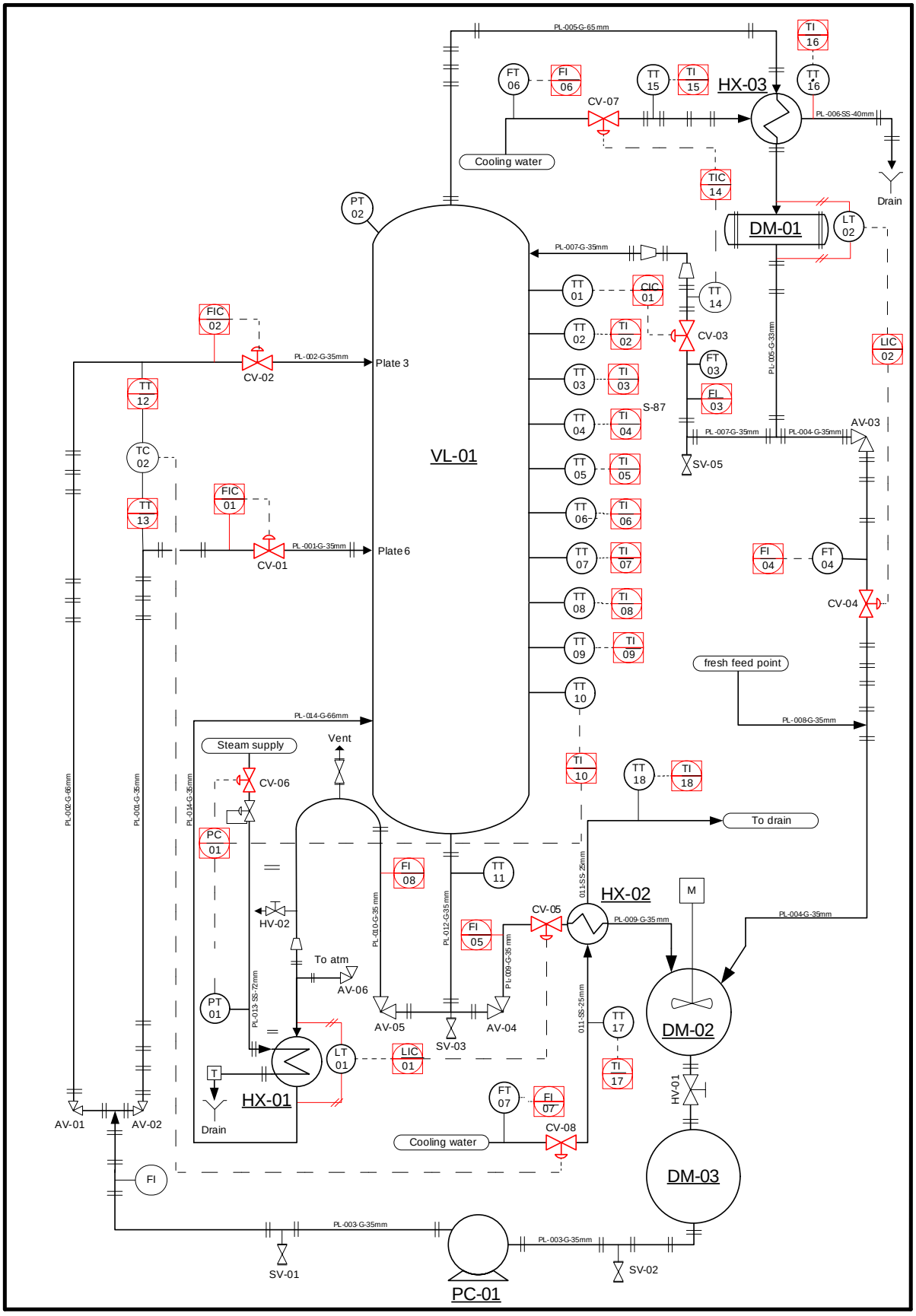


Figure 7.2: Diagram showing piping and Instrumentation of the pilot scale distillation column

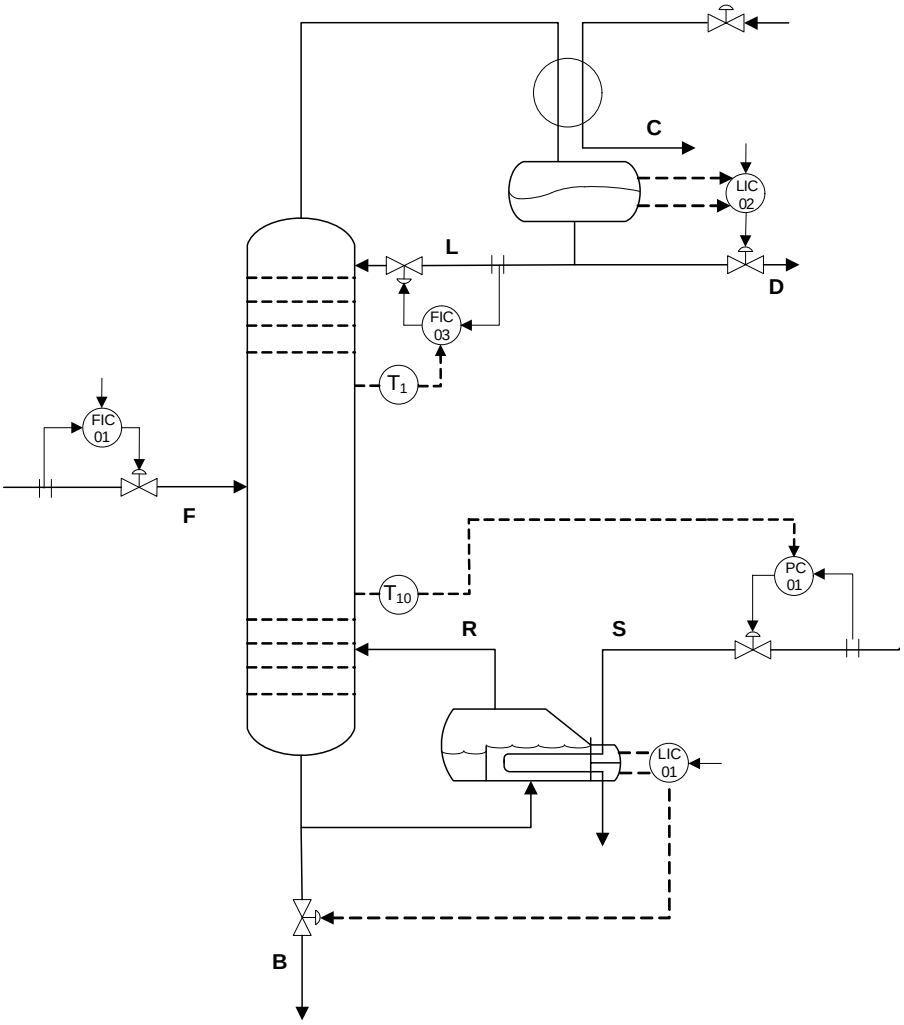


Figure 7.3: Simplified illustration of relevant control loops

Table 7.1 indicates the normal operating ranges of the different variables concerned with the operation of the column. Note that all the pressure values indicated are gauge pressure values.

The column is almost always operated with only one feed, typically the one feeding onto plate 7. While it has been stated that the system is a closed system in that the products are mixed and recycled back into the system via the feed tank, it has been established that given the feed tank has a surplus volume of feed, the effects of varying conditions or properties of the recycle streams (composition, temperature etc.) on the feed stream are negligible. Thus for all purposes of this work the feed stream is considered independent from the product streams.

Control and Dynamic Behaviour

The base layer control setup is primarily defined by the following control loops as illustrated in figure 7.2 and more clearly shown by figure 7.3:

Variable	Label	Units	Normal Range
Steam Pressure	P_S	kPa	55-90
Cooling Water Valve	X_C	% Open	50-100
Reflux Level	L_R	kPa	70-85
Boiler Level	L_B	kPa	24-30
Bottoms Flowrate	F_B	kg/hr	0-30
Distillate Flowrate	F_D	kg/hr	0-30
Reflux Flowrate	F_R	kg/hr	0-20
Feed Flowrate	F_F	kg/hr	0-40
Feed Temperature	T_F	^oC	20-38
Feed Composition	C_F	% mol	40-60
Top Plate Temperature	T_P	^oC	77-82
Bottom Plate Temperature	T_B	^oC	88-96

Table 7.1: Distillation column principle variables and normal operation ranges

LIC-01 : The boiler level is controlled by directly manipulating the bottoms product stream valve.

LIC-02 : The relux drum level is controlled by directly manipulating the distillate product stream valve.

FIC-01 : The feed flow rate is controlled my manipulating the valve in this stream.

FIC-03 : This is a cascade control loop. The secondary loop is that which controls the reflux flowrate by adjusting the valve in the reflux stream while the primary loop controls the top plate temperature by adjusting the setpoint reflux flowrate value for the secondary loop.

PC-01 : This is a cascade control loop with the secondary loop controlling the steam pressure by manipulating the throttling valve in the steam feed line while the primary loop controls the bottom plate temperature by adjusting the setpoint steam pressure value for the secondary loop.

These control configurations were chosen as they were found to perform better than any other. While it is understood that typical distillation control theory dictates that the condenser cooling water flowrate is a valuable manipulated variable as it allows for reflux temperature control, early controllability studies found the effect of this variable to saturate at a very low flowrate. This made it very ineffective as a manipulated variable.

All the control loops used PID controllers with most of the parameter tuning done by trial and error and the occasional assistance by the DeltaV Tune Tool Kit. Control loop FIC was the only loop of the five to use differential control action.

It must be said that improving the performance of the different controllers was a difficult task. This problem was largely rooted in the feed pressure oscillations caused by the electric boilers. The supply pressure before throttling was found to oscillate between 650 kPa and 700 kPa at a frequency of 0.18 Hz, the controller adjusting the throttling valve in order to control the steam pressure in the boiler was unable to regulate these disturbances. This consequently resulted in the steam pressure in the boiler oscillating by 5 kPa at a similar frequency. Figure 7.4 illustrates the performance obtained by the steam pressure controllers. Figure 7.4(a) shows the achievable steam pressure control while figure 7.4(b) shows the primary controllers efforts in manipulating the steam pressures setpoint to control the bottom plate temperature.

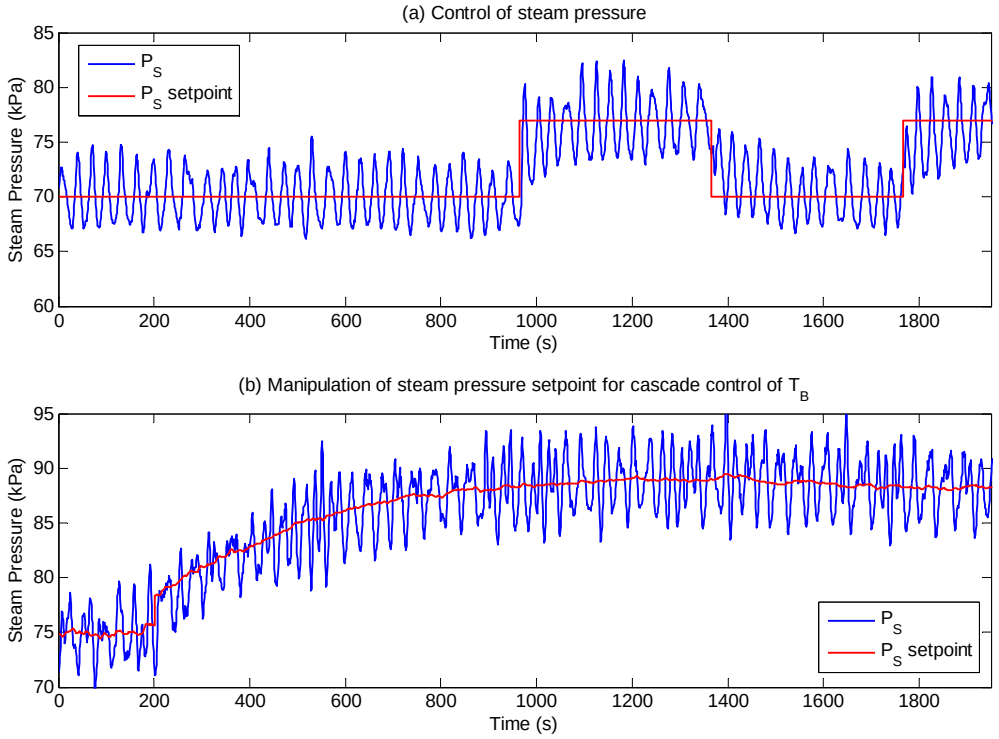


Figure 7.4: Steam presssure control and manipulation

The effects of the pressure supply oscillation extended well beyond the boiler steam pressure. Figure 7.5 shows the top and bottom plate temperature responses to steps in reflux flowrates and steam pressure with the primary cascade loops open. Here the control problem and multivariable nature of the column reveals itself in that the sensitivity of the top plate temperature to the oscillations varies according to region of operation and magnitude of system variables. Figure 7.5(a) shows a T_P response to a 50 % step in reflux flowrate magnitude, it is observed that the trend is smoother than when responding to a 50 % step in boiler steam pressure, 7.5(b).

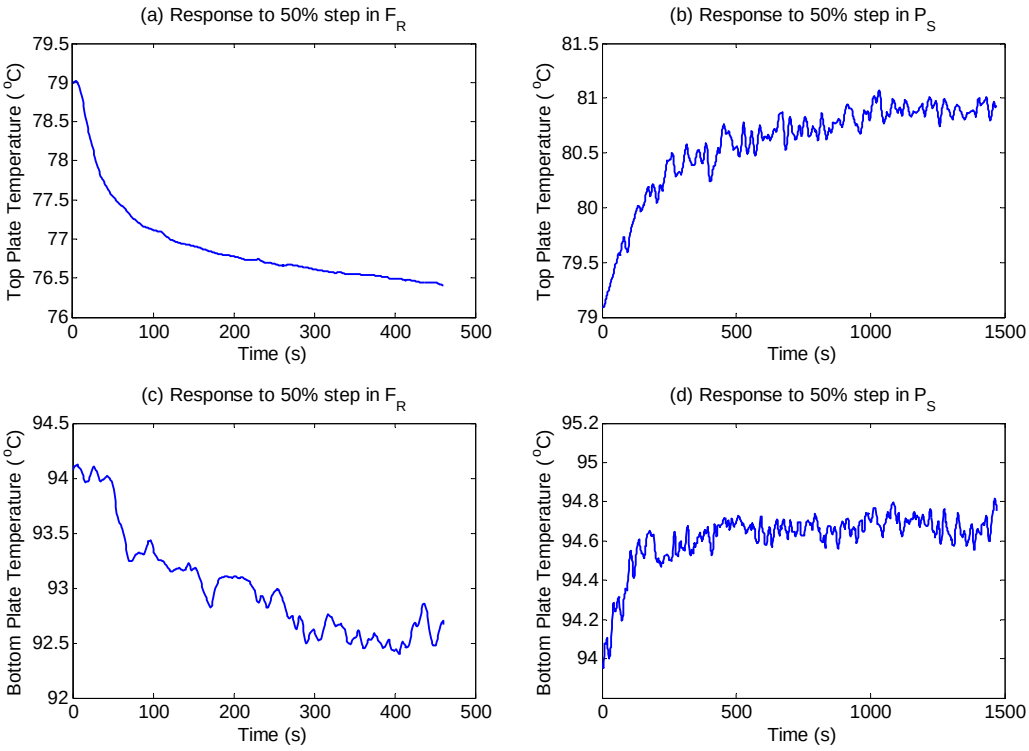


Figure 7.5: Open-loop step response

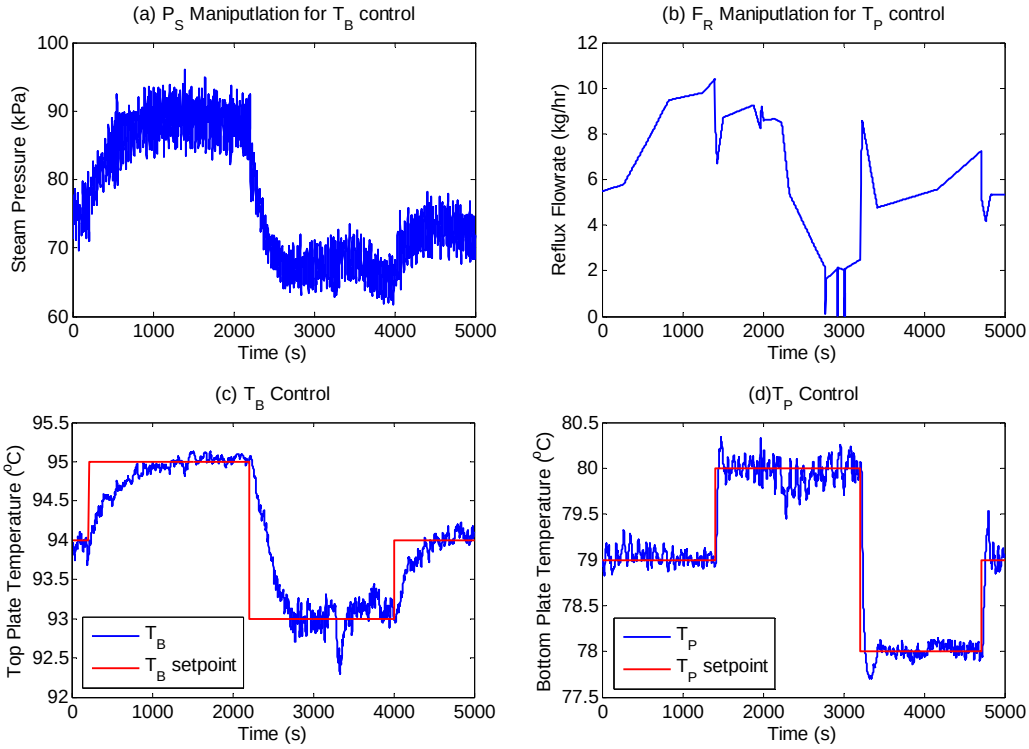


Figure 7.6: Distillation column setpoint tracking

This phenomenon is seen again in the closed-loop responses in figure 7.6, where the top plate temperature control (figure 7.6(d)) shows much larger oscillations when controlled at 80 ^oC than at 78 ^oC. This dynamic behaviour is most likely due to fact that at larger reflux flowrates the effects of the steam pressure oscillations are dampened. This type of dynamic interaction between system variables and control loops made it difficult to control for robustness and performance and was made more evident when operating the column over larger temperature ranges.

It is clear that the column dynamics are such that the system variables display varying dependencies to each other at different magnitudes. These observations reflect the non-linear nature of the column. Higher order correlation tests for non-linear dynamics (as done in the previous section, figure 6.5) could not be conducted as disturbing the column under open loop conditions with white Gaussian disturbance signals resulted in extremely unstable conditions.

Furthermore, it is clear that the column has several control problems. The use of controller gain scheduling would do very well in linearising the controller and system variable interactions. An MPC controller would do even better and feed forward logic may have been used to reduce the problems caused by the pressure supply fluctuations. However, the problems that these solutions might resolve are characteristic of 'real' systems and it is with this perspective that the column is operated and identified.

7.2 Experimental Method : Framework and Design

7.2.1 System and Identification Problem Definition and Assumptions

As has been described in recent sections, the distillation column being used has 12 principal variables and 5 control loops all interacting and affecting each other. In defining the identification problem presented for the distillation column, the key variables are defined as follows:

Input 1 - u_1 **:** Reflux flowrate - F_R

Input 2 - u_2 **:** Boiler steam pressure - P_S

Output 1 - y_1 **:** Top plate Temperature - T_P

Output 2 - y_2 **:** Bottom plate temperature - T_B

This implies a 2×2 multivariable structure defines the distillation column to be identified which is justified given the following assumptions:

Variable	Label	Units	Steady State Value
Steam Pressure	P_S	kPa	75
Cooling Water Valve	X_C	% Open	63
Reflux Level	L_R	kPa	79
Boiler Level	L_B	kPa	27
Bottoms Flowrate	F_B	kg/hr	8
Distillate Flowrate	F_D	kg/hr	15
Reflux Flowrate	F_R	kg/hr	7
Feed Flowrate	F_F	kg/hr	20
Feed Temperature	T_F	^oC	36
Feed Composition	C_F	% mol	50
Top Plate Temperature	T_P	^oC	79
Bottom Plate Temperature	T_B	^oC	94

Table 7.2: Steady state values used as initial states of experiments

- The other principal system variables not included in the model structure are the reflux drum level, boiler level, feed properties and environmental conditions. These variables will be maintained constant at relevant magnitudes (given in table 7.2) by the appropriate controllers. The controllers for most of these variables perform very well; it is thus assumed that there is no controller interaction through these variables that adversely affect the variables in the model structure.

- Effects of recycle streams on the feed are dampened by a large feed surplus and are thus assumed negligible.

- Both system inputs - F_R and P_S - are regulated by control loops manipulating the reflux line valve and the steam supply line valve respectively to obtain the specified setpoint values. It is understood that due to the steam pressure control difficulties this input is to an extent continuously disturbed.

- It is assumed that the top and bottom plate temperatures can be used to infer composition values of the distillate and bottoms streams respectively. This is however not expected to be so at temperature measurements bellow the azeotrope at 78 ^oC.

From this it is possible to define the condition and dynamic operating region of the distillation column to be identified. This condition may be characterised as an optimal region within the normal operating range. The region is defined by a 1 ^oC span above and bellow the optimal steady state values of the top and bottom plate temperatures with all else but the afore-mentioned inputs being constant and maintained at the steady state values given in table 7.2.

7.2.2 Identification Framework and Validation Methods

The very same identification framework used in the simulated identification experiments in chapter 6 was used on the distillation column. That is, the prediction error method together with an unfiltered unweighted quadratic norm. These are detailed in section 2.5.2. When applied to closed-loop data the direct approach was used (see section 3.2.1).

With respect to the validation methods used to validate the models generated, once again the cross-validation technique (as illustrated in figure 6.8) was the primary model validation approach. As before, several approaches to cross-validation were used, these may be categorised as follows :

- Simulation and Prediction percentage fits.
- Frequency Content Analysis
- Residual correlations
- Higher Order Residual Correlations

7.2.3 Model Structure and Order Selection

The same model structure and approach to selecting mode orders was taken in identifying models of the distillation column as for the simulated systems, systems A and B . That is, ARX structures defined by equations 6.3 through 6.5 were used with the order being selected by trial and error considering the cost of introduced model complexity for the improved response fit. The final order selected was

$$n_i = \begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}$$

In terms of the response delays used in the models, step disturbances responses were used to determine the response time delays. The delays used maybe given as follows (in seconds):

$$d = \begin{bmatrix} 5 & 32 \\ 8 & 6 \end{bmatrix}$$

However, as will be mentioned in next section, all data for identification (and validation) was sampled at intervals of 10 seconds, since the identified model structure is of discrete from, this means the response delays need to be represented as factors of the sampling interval. Thus, rounding off the time delays to the nearest product of the sampling interval, the sample delays of the identified models are given as

$$d = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$$

With respect to the scaling of the estimated models, all data generated was scaled around the steady states given in table 7.2, thus allowing for deviation models.

7.2.4 Identification and Validation Experiment Conditions

As with the simulated experiments, open and closed-loop experiments were done to generate data for both identification and cross-validation. Recalling figures 2.3 and 3.1, signals $u(t)$ and $r_1(t)$ were again used as disturbance signals for the open and closed-loop cases respectively with the occasional setpoint changes through $r_2(t)$ for the closed-loop experiments. With the previous chapter establishing PRBS disturbance signals to be the more consistent of the disturbance signals, this was the sole disturbance signal used in efforts to identify models of the distillation column. Figure 7.7 illustrates the input signal disturbances used to generate identification data for open-loop experiments.

It is at this point appropriate to reiterate that both the open-loop and closed-loop experiments were output constrained so as to allow for the identification of a specific region of dynamics. This means that while the closed-loop experiments where designed in accordance to output ranges (setpoint ranges) the input ranges defining the open-loop experiments were designed in consideration of constraints on the outputs or responses to such inputs.

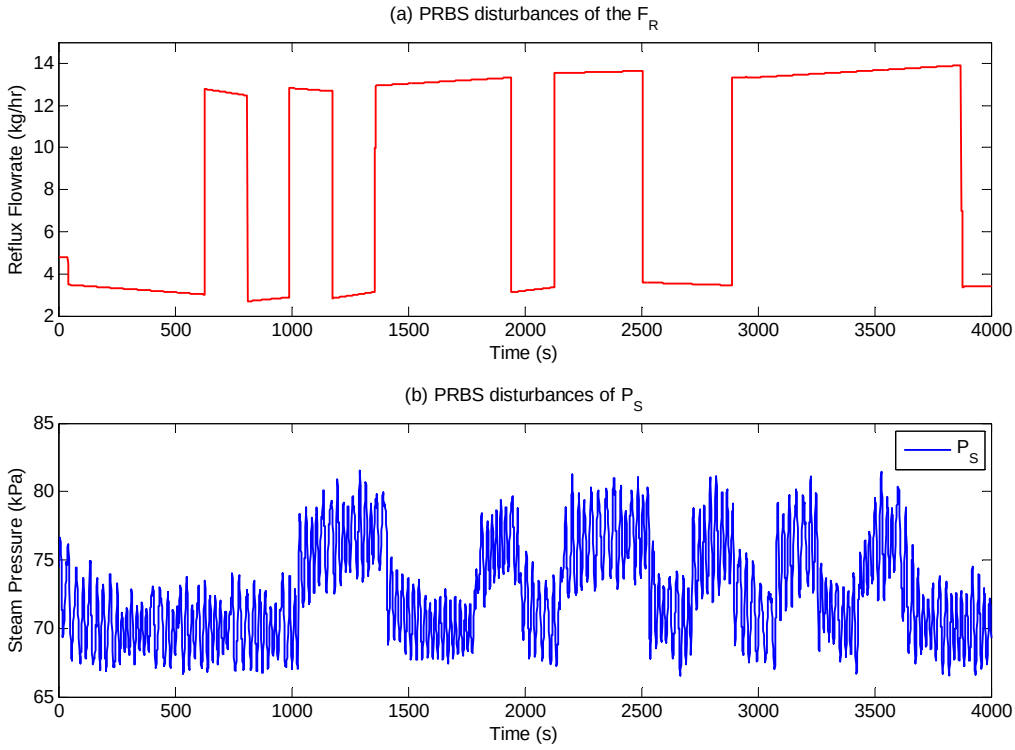


Figure 7.7: Input disturbance for open-loop distillation column identification experiments

As discussed earlier, the distillation column does produce dynamic conditions that

were not seen in the simulated experiments in the previous chapter. This is primarily the case in the continuous disturbance of one of the inputs, the boiler steam pressure (u_2). This introduces an interesting challenge in identifying models of the column in that the column is expected to exhibit non-linear dynamics, making it important to identify models over certain dynamic regions. With one of the inputs continuously disturbing the column at a certain frequency and over a span of magnitudes, specifying a region over which to generate data for identification and validation becomes more difficult.

The conditions defining the different experiments used to generate data for identification and validation are defined in tables 7.3 and 7.4 which detail the open and closed-loop identification experiments respectively, and tables 7.5 and 7.6 which detail the open and closed-loop experiments used to generate validation data. It must be said that the same approach used in designing the identification and validation experiments for the simulated systems was taken. That is, the experimental design common practises presented in chapter 5 are used to determine the average constant signal intervals (B values), experiment duration and disturbance signal magnitudes. The remainder of this section details the experimental conditions and further elaborates on the information presented in tables 7.3, 7.4, 7.5 and 7.6.

Case Tag	Input Signal ($u(t)$)		
	Signal Type	Range $[u_1]$,$[u_2]$	B
		[kg/hr],[kPa]	[s]
DOL 1	PBRS	[4, 14],[70,77]	200

Table 7.3: Open-Loop cases used to generated data for identification of the distillation column

Case Tag	Dither Signal ($r_1(t)$)			Setpoint Signal ($r_2(t)$)
	Signal Type	Range $[r_{1_1}]$,$[r_{1_2}]$	B	SP Values $[SP_1]$,$[SP_2]$
		[kg/hr],[kPa]	[s]	$[^oC]$,$[^oC]$
DCL 1	PRBS	[-0.6, 0.6],[-1.2, 1.2]	200	[78,80,79],[93,95,94]
DCL 2	PRBS No Dith	[-0.6, 0.6],[-1.2, 1.2]	200	[78,80,79],[93,95,94]
DCL 3	PRBS no Δ SP	[-0.6, 0.6],[-1.2, 1.2]	200	[79,79,79],[94,94,94]

Table 7.4: Closed-Loop cases used to generated data for identification of the distillation column

Case Tag	Input Signal ($u(t)$)		
	Signal Type	Range $[u_1]$,$[u_2]$	B
		[kg/hr],[kPa]	[s]
DVOL 1	PBRS	[4, 14],[70,77]	180
DVOL 2	PBRS L	[5, 18],[62,77]	180

Table 7.5: Open-Loop cases used to generated data for validation of the distillation column

Case	Dither Signal ($r_1(t)$)			Setpoint Signal ($r_2(t)$)
	Signal Type	Range $[r_{1_1}]$,$[r_{1_2}]$	B	SP Values $[SP_1]$,$[SP_2]$
		[kg/hr],[kPa]	[s]	$[^oC]$,$[^oC]$
DVCL 1	PRBS	[-0.6, 0.6],[-1.2, 1.2]	180	[78,80,79],[93,95,94]
DVCL 2	PRBS	[-0.6, 0.6],[-1.2, 1.2]	180	[80,78,79],[95,93,94]
DVCL 3	PRBS L	[-0.6, 0.6],[-1.2, 1.2]	180	[77,76,79],[94,92,93]

Table 7.6: Closed-Loop cases used to generated data for validation of the distillation column

The following further elaborates on the information presented in tables 7.3, 7.4, 7.5 and 7.6:

Range : Specifies the range of the input and dither signals used in the open and closed-loop experiments respectively. In the case of the open-loop experiments the ranges of the input signals themselves are given, while in the case of the closed-loop experiments the ranges of the dither signals disturbing each input are given.

SP Values : Specifies the changes in setpoint values made for each output. $[SP_1]$ are the setpoint changes made to output y_1, $[SP_2]$ are the setpoint changes made to output y_2.

no ΔSP : is the condition where no setpoint changes were incurred.

No Dither Signal : is the condition where no dither signal was used, only setpoint changes.

L : The indicator L, was assigned to show where the experiment was operated over a larger input range (open-loop experiments) or larger output range (closed-loop experiments) than the typical identification experiments.

The following can be said regarding the operation of the distillation column in conducting identification experiments and experiments to generate validation data:

- The column was always brought to the same steady state before any experiment was conducted. All experimental cases were generated around this steady state with the primary region of interest being 1^oC around the steady state temperatures of the top and bottom plates. Table 7.2 shows the steady state values.

- All data was sampled from the continuous historian every 10 seconds. This sample rate was chosen as it was found to be most optimal in terms of information content and the consequent computation intensity during model estimation procedures.

- Data pre-treatment for use as validation and identification data involved de-trending by subtracting the steady state values from the measured values, thus creating deviation from steady-state data.

- As was established in previous chapter, identification theory dictates that the average constant signal time interval (B) of the disturbance signal be taken as 30% of the largest time constant. Step tests found the largest time constant to be 650 seconds, thus the PRBS disturbance signals used in identification experiments had B values of 200 seconds. The B value for the disturbance signals used generate data for validation were taken as 180 seconds.

- The duration of each of the experiments was taken as several times the largest time constant, 5000 seconds.

- Both input variables were disturbed simultaneously but not at the exact same moment in time. Figure 7.7 illustrates this.

- All experiments besides those designed to be carried out over larger ranges (experiments indicated with "L") where maintained

- The input signal magnitudes for the open-loop experiments (not including the "L" experiments - larger range) were designed to disturb the system such that the output responses maintained themselves within 1^oC of the steady state values. Figure 7.8 illustrates this for input u_1 and output y_2.

- The magnitude of the dither signals used during closed-loop experiments was taken as 10 % of the input range that maintains the temperature values within 1^oC around the steady state values.

- The setpoint changes for output y_1 and y_2 ($[SP_1]$,$[SP_2]$), were made at 1400, 3200 and 4700 seconds and 200, 2200 and 4000 seconds respectively.

- Four data sets were generated within a tight range of operation considered to be optimal in terms of maximising the temperature differences between the top and bottom plate. One of these sets was generated under open-loop conditions (DOL

1) while the other three were under closed-loop conditions (DOL 1-3). These data sets were used for identification.

- Five data sets were generated as validation sets, two open-loop (DVOL 1-2) and three closed-loop (DVCL 1-3). All but one closed-loop and open-loop experiment were done over the normal operating range (1^oC around the steady state temperatures). The experiments that were not done over the normal operating ranges were operated over a range about 2^oC around the steady state values. Note that with the azeotrope at 78^oC and the steady state temperature at 79^oC the the azeotrope tempaerature was crossed for these experiments implying that equilibrium conditions were not always maintained.

- The only difference between validation experiments DVCL 1 and DVCL 2 is that the setpoint changes made for DVCL 1 are the same as those made for identification experiment DCL 1 while the setpoint changes for case DVCL 2 are not.

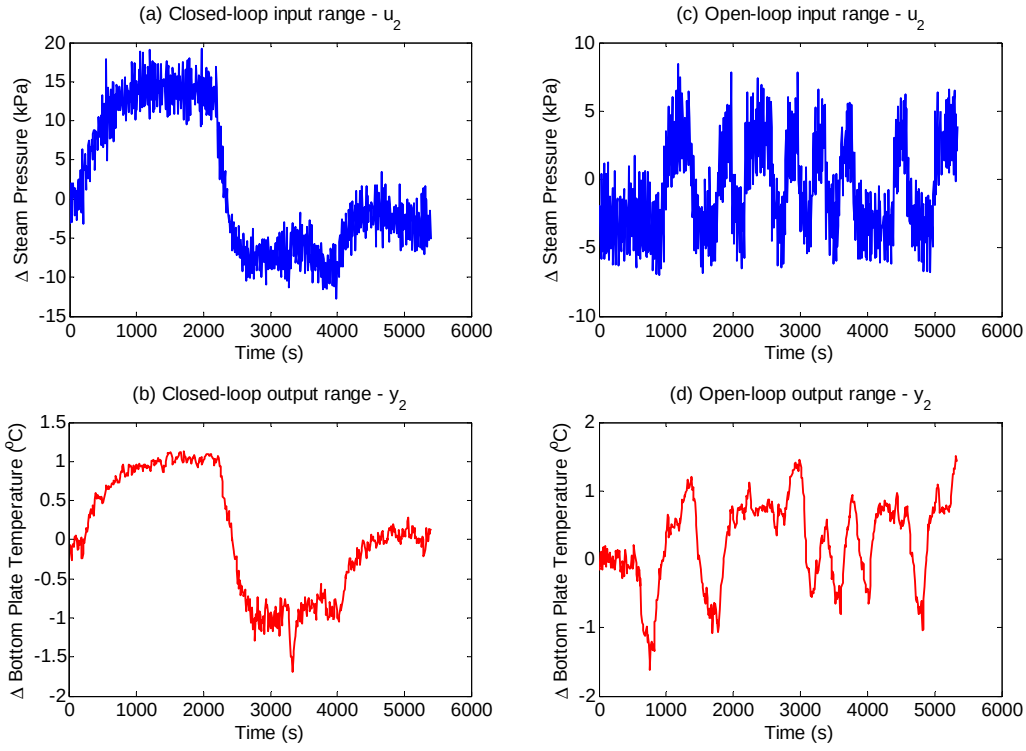


Figure 7.8: (a)-(b) u_2 and y_2 values generated from closed-loop identification experiment DCL 2. (c)-(d) u_2 and y_2 values generated from open-loop identification experiment DOL 1

7.2.5 General Execution

As mentioned before, the column was operated via the DeltaV operating system with the data being accessed via the continuous historian. While the setpoint changes and disturbance signal scheduling were done using the sequential flow functions in DeltaV, all other aspects of the generating signals, models and validation were done through Matlab.

The same points made in section 6.2.5 regarding the execution and code used in the identification of models for the simulated systems and the validation of said models extend to the identification and validation efforts for the distillation column.

7.3 Identification and Validation Results

This section presents and discusses the results of the identification of linear models approximating the pilot scale distillation column and the validation of such models. Note that all the parameters of each identified model are presented in appendix A.2.

7.3.1 Simulation and Prediction

The results presented here regarding the model response simulation and prediction analyses were generated in a very similar manner to the simulation and prediction results presented for the simulated systems, systems A and B, in sections 6.3.3 and 6.4.2 respectively. That is, each identified model was cross-validated against each validation set. This was done by disturbing each identified model by the input contained in each validation set, from this the model response to this input is compared against the response contained in the validation set (i.e. the response of the distillation column). Figure 6.8 illustrates this cross-validation approach. The accuracy of the model response relative to the distillation column response (validation set response) is expressed as a percentage fit value (see equation 6.7).

The percentage fit values of the simulated responses of each model when validated against open-loop validation sets and closed-loop validation sets are given in figures 7.9 and 7.10 respectively.

General Observations

In comparing the differences between figures 7.9 and 7.10, it is clear that all the models generated extremely inaccurate responses when validated against the open-loop validation sets (figure 7.9). In fact most of the response percentage fit values are negative. In contrast, the response fit values obtained by each model when validated against the closed-loop validation sets showed much better percentage fits in that the values were generally much larger. Furthermore it is noted that both open and closed-loop validation

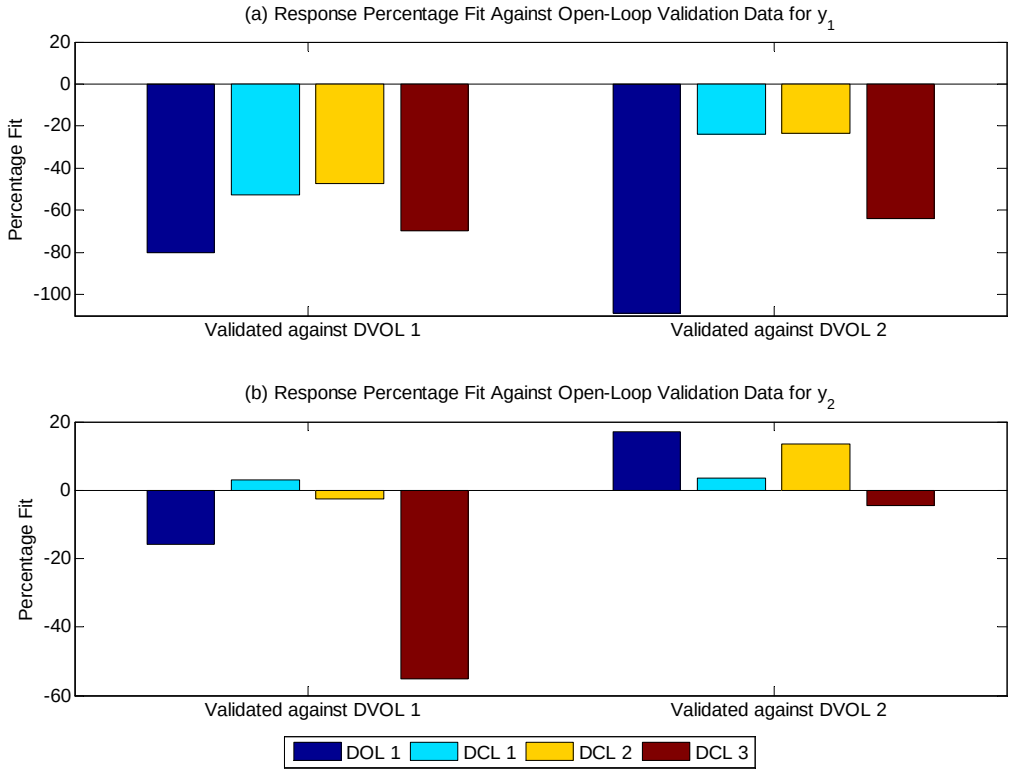


Figure 7.9: Percentage fit values when validating against open-loop data sets

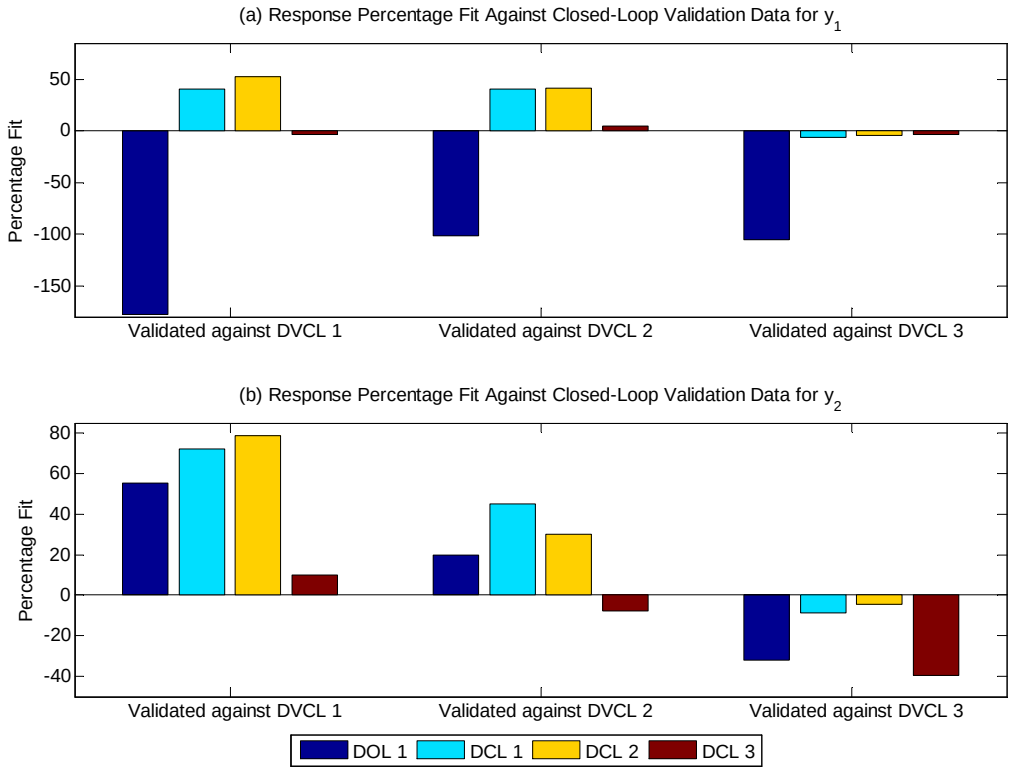


Figure 7.10: Percentage fit values when validating against closed-loop data sets

sets reveal that all the models generated more accurate response fits for output y_2 in comparison to those for output y_1. As will be discussed later, this may be attributed to the continuous disturbance of input u_2

Validation Set Model Discrimination and Bias - Closed-loop Validation Sets

The closed-loop validation sets proved to be more informative in terms of providing insight into the identification and validation problem. This is due to the larger contrast in results in comparison to those obtained when using the open-loop results which simply indicated that all the models produced terribly inaccurate responses.

In interpreting the significance of the general observations it is important to consider the different discrimination capacities by the different validation sets given certain experimental conditions. As was established in section 5.5, closed-loop experiments are expected to produce more informative data in comparison to open-loop experiments only under the implementation of output constraints. This is exactly the case here, figure 7.8 shows this effect of implementing output constraints clearly. The closed-loop conditions allowed for a larger input signal magnitude range while maintaining the output with in bounds.

Looking at the response percentage fit values for the different identified models when validated against the closed-loop validation sets (figure 7.10), it is very clear that the models obtained from closed-loop experiments (DCL 1-3) almost always generated more accurate responses to that obtained from the open-loop experiment, DOL 1. This discrimination against the model obtained from open-loop experiments is certainly due to the differences in information content between open-loop and closed-loop experiments based on the afore mentioned condition caused by output constraints.

In observing the differences in response fits when validating the models against closed-loop validation set DVCL 1 and DVCL 2 some interesting observations were made. With the only difference between the two validation sets being that set DVCL 1 was generated from experiments where the setpoint changes were precisely the same as those used to generate the identification data for cases DCL 1-2, while set DVCL 2 used a different sequence of setpoint changes (see tables 7.4 and 7.6). It is observed that the models all produced slightly less accurate responses when validated against DVCL 2. This implies that the closed-loop validation sets are slightly biased towards models obtained from similar setpoint direction changes.

Validation Set Model Discrimination and Bias - Signal Magnitude

The differences in response fit results when using validation sets obtained over the normal operating region to those obtained when using validation sets generated by larger input signals and output ranges are of particular interest. This is as the larger the effect

of signal magnitude on the results the more the column dynamics may be attributed as being non-linear. Additionally, response fit sensitivity to validation set signal magnitudes also imply the validation sets are discriminating based on differences in dynamic region exposure caused by different signal magnitudes.

Relative to validation sets obtained from closed-loop experiments, set DVCL 3 differs from DVCL 1 only in that the setpoint magnitudes are larger. Relative to the validation sets obtained from open-loop experiments, set DVOL 2 differs from DVOL 1 only in that larger input signals magnitudes were used.

Looking at the closed-loop validation results in figure 7.10 it is clear that all the models performed worse when validated against the validation set generated over a larger dynamic region (DVCL 3). That is, the accuracy of the model responses greatly diminished when validated against DVCL 3 in comparison to DVCL 1. The difference between figures 7.11 and 7.12 further shows this difference in response accuracy as figure 7.11 displays each model response when validated against DVCL 1 and figure 7.12 shows the responses validated against DVCL 3.

The results obtained from the open-loop validation sets regarding the effects of larger disturbance signals were very interesting. It may generally be stated that the response accuracy of the models obtained from closed-loop experiments (DCL 1-3) improved when validated against the open-loop validation set generated from larger disturbance signals (DVOL 2) in comparison to the respons accuracy obtained when using set DVOL 1. This while the model obtained from the open-loop experiment (DOL 1) generally revealed worse response fits. This may be justified by the point made earlier regarding the closed-loop experiments ability to use a larger range of the input signal while maintaining its output constraint (figure 7.8). This means that the closed-loop experiments identified a larger range of dynamics in comparison to the open-loop experiments given the same output constraints. Thus its improved fit when validated against the open-loop validation set obtained from larger input signal magnitudes.

Experimental Condition Sensitivities

In determining which experimental condition generated the model that produced the most inaccurate responses it was found that the results depended on the output. It was found that the model obtained from the open-loop experiments, DOL 1, produced the least accurate (smallest percentage fits) responses regardless of the validation set used for output y_1. While the same may be said for the model obtained from closed-loop experiments where no setpoint changes were made, DCL 3, for output y_2.

In actuality, as mentioned earlier, it may be said that all the models produced more accurate responses for output y_2 in comparison to y_1. These improved response fits for output y_2 may be explained by the fact that input u_2 in consistently being disturbed by

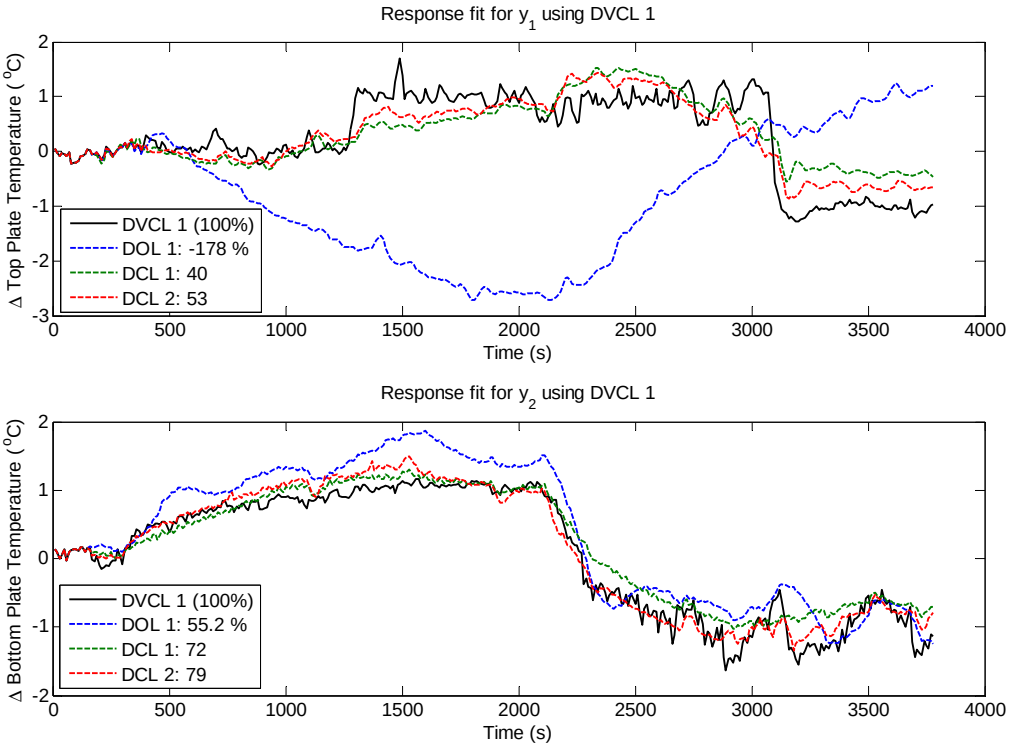


Figure 7.11: Validation simulations against set DVCL 1

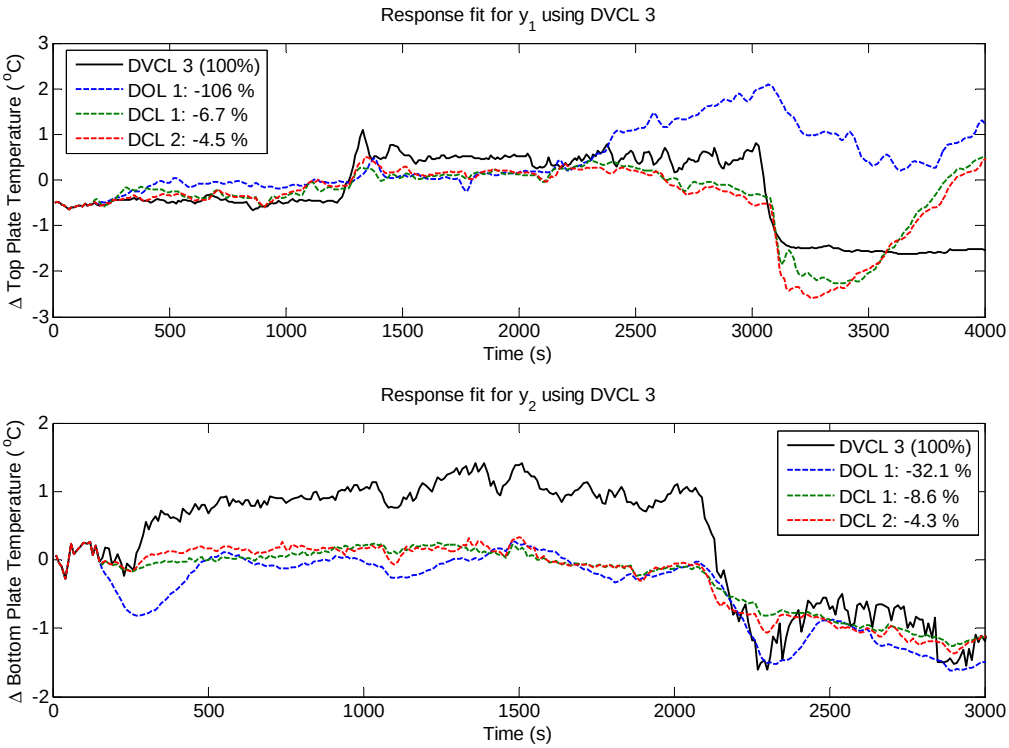


Figure 7.12: Validation simulations against set DVCL 3

an upstream condition. This, and the fact that input u_2 has a larger relative gain with output y_2 than y_1, means that output y_2 is being persistently excited to a greater extent. This condition is further discussed in the frequency analyses section that follows.

Experiments DCL 1 and DCL 2 are both closed-loop experiments with the only difference between the two being that DCL 2 did not use dither signal disturbances and DCL 1 did. Observing the differences in response fits of the models obtained from these experiments it is noted that there is no clear distinction trend between the two. In some cases, such as output y_2 when validated against DVCL 2 (figure 7.10(b)), the results suggest that the use of dither signal disturbances allowed for improved system excitation and information generation and the identification of a model with a more accurate response. However, most cases the difference in response fits were not so pronounced, in fact in some cases, such as output y_1 when validated against DVCL 1 (figure 7.10(a)), DCL 2 showed larger response fit values than DCL 1 suggesting that not using a dither signal allowed for more informative data and a more accurate model.

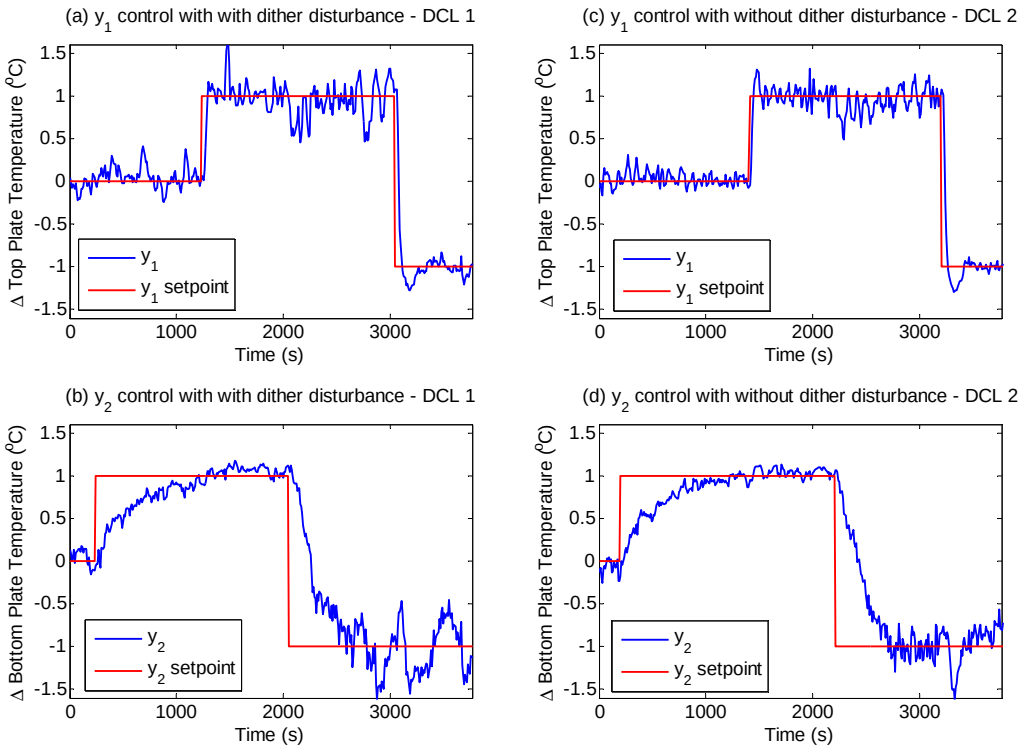


Figure 7.13: Experimentation with and without dither signals

This variance in effectiveness of the dither signal may be explained by the fact that the dither signal did little to additionally disturb the system on top of the disturbances caused by the continuous disturbance of input u_2 by an upstream condition. Figure 7.13 shows the output response differences between experimental case DCL 1 which used dither signals and case DCL 2 which did not. It is noted that the dither signals did

little to further excite or disturb the system. The persistent disturbance of input u_2 by the upstream condition was such that the contrast of the effects of the dither signals were diminished. This explains the small difference in model accuracy since both cases produced similarly informative data.

Prediction

Analysing the predictive accuracies of each model against each validation set, it was found that no significant difference in percentage fit could be found using prediction horizons larger than 20 steps. These 20 steps translate to 200 seconds given the 10 second data sampling rate and the consequent step size defining the discretised models identified. It is at horizons smaller than 200 seconds that the model fits begin to improve and the differences between the results generated by the different validation sets begin to diminish. Figure 7.14 shows the prediction percentage fits using a 100 second horizon when validated against the closed-loop validation sets.

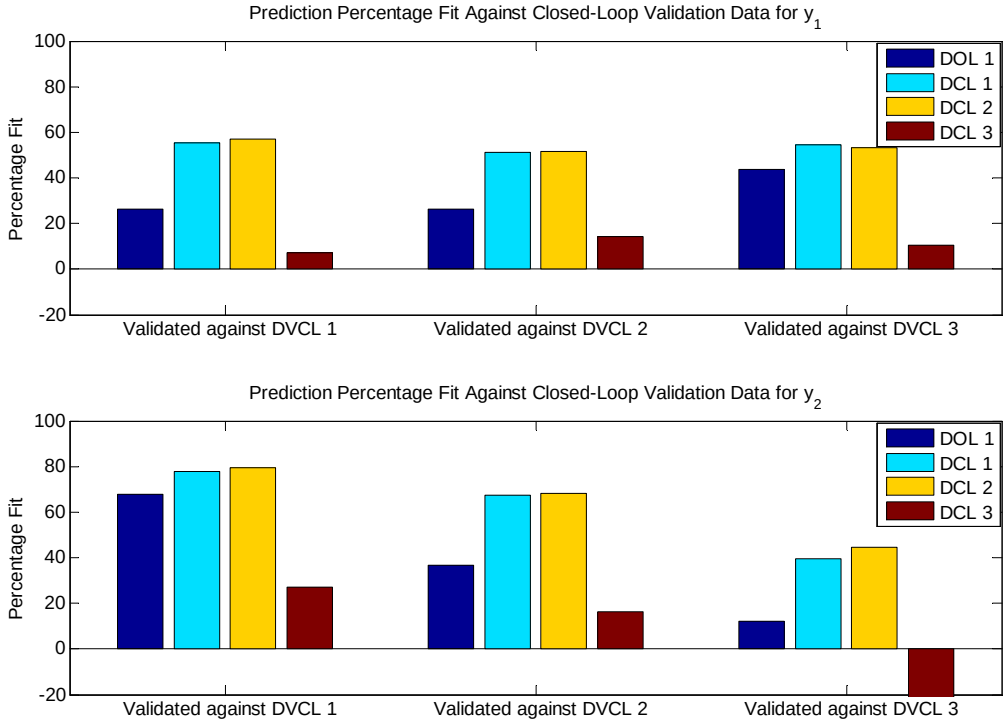


Figure 7.14: Validation predictions against the closed-loop validation sets

In comparing the prediction results given in figure 7.14 to the simulation results given in figure 7.10 it must be said that the models are more accurate in predictions than in simulations given small enough prediction horizons are used. Furthermore the effect of increasing the signal magnitudes and output range has a much smaller effect on the prediction accuracy than the simulation accuracy. This is seen by the much improved

results for the prediction when validated against validation set DVCL 3 in comparison to the simulation result. Furthermore the model obtained from experiment DCL 3, characterised as a closed-loop experiment where no setpoint changes were incurred, clearly produced the least accurate predictions of all the models.

7.3.2 Frequency Content Analyses

Analyses of the identified models approximating the distillation column in the frequency domain were done from the perspective of cross-validation. That is, the models themselves were not assessed but their responses and how they fit the validation data responses. In other words, each model's simulated response was assessed in the frequency domain and compared to that of the validation set response. Thus the frequency content of the response fit is presented in this section.

It was found that most of the frequency content analyses of the response fits did generally reflect the simulated response fit results given in the previous section. Some interesting observations were however made.

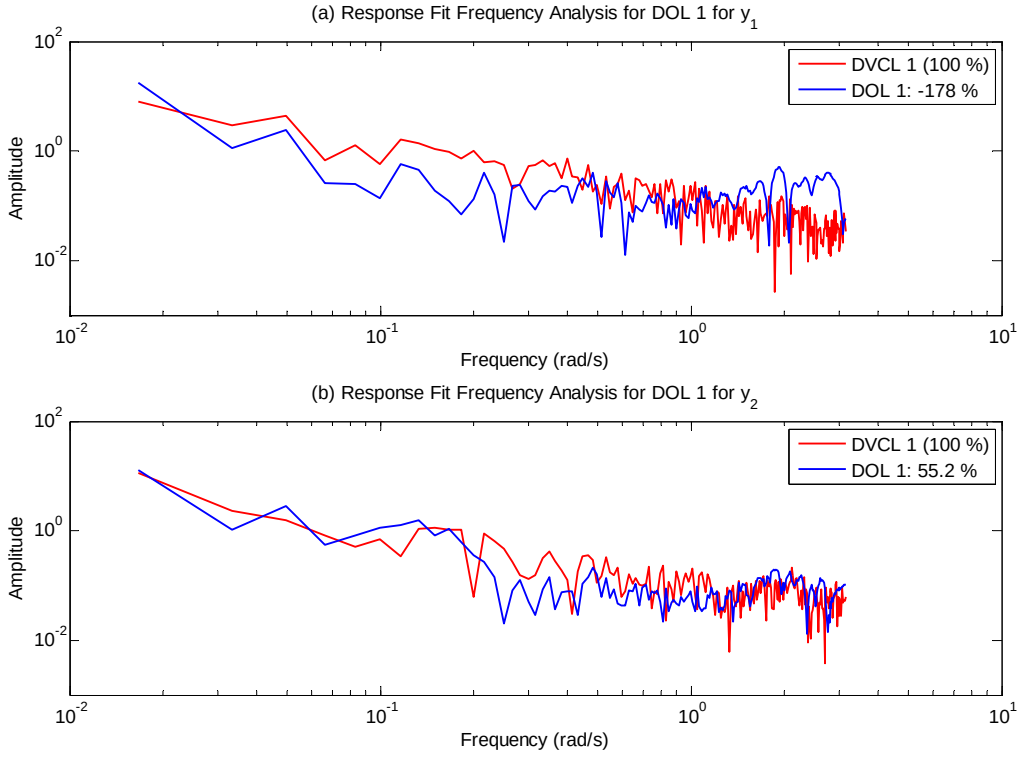


Figure 7.15: Frequency analysis of simulation validation for DOL 1 against DVCL 1

Recalling that the identified models showed slightly improved response fits for output y_2 in comparison to output y_1 and that this improved fit was thought to be due to the continuous disturbance of input u_2. This contrast in response fit for the different outputs was most evident for the model obtained from the open-loop identification experiment,

DOL 1 (see figure 7.10), thus it was most fitting to use this model and its responses to investigate these conditions in the frequency domain. In assessing this case and its response fits in the frequency domain it was found that output y_2 has improved fits at certain frequencies in comparison to output y_1. This is shown in figure 7.15. Figure 7.15(a) shows the frequency fit between the response generated by DOL 1 for output y_1 and the validation set response given by DVCL 1. Figure 7.15(b) shows the same but for output y_2. It is noted that the frequency content fit shows output y_2 to have a slightly better fit at both the lower and higher end of the frequency band.

It is interesting to note that the always present disturbance of input u_2 occurred at an approximate frequency of 0.18 Hz (1.13 rad/s), one might postulate from this that the improved fit for output y_2 could be due to the generation of more informative data due to these disturbances and the fact the relative gain is larger between this input-output pair.

7.3.3 Residual Correlation Analysis

As was done in sections 6.3.5 and 6.4.4 for the models approximating systems A and B respectively, the residual correlation analyses presented here was done so with the purpose of obtaining indications of unmodelled and incorrectly modelled dynamics in each identified model. Additionally, the differences between correlation profiles obtained when using different validation sets were studied to provide insight into how the validation sets differently discriminate. Furthermore higher order residual correlation tests were used to gain insight into how well the linear models approximated the non-linear dynamics.

Whiteness Test Results

In assessing the residual independence of each model when validated against each validation set it was found that the results did not clearly reflect the simulated response fit findings. Figure 7.16 shows the residual auto-correlation plots obtained for the model generated by closed-loop experiment DOL 1 validated against validation sets DVCL 1, DVCL 3 and DVOL 1. In a similar fashion figure 7.17 shows the correlation plots for the model generated by the open-loop experiment DCL 1.

It is at this point appropriate to recall the response fit results obtained by these two models (DOL 1 and DCL 1) in figures 7.9 and 7.10. As has been established, one of the key observations made regarding the response fits was that all the models produced more accurate response fits for output y_2, this was especially the case for the model obtained from the open-loop experiments (DOL 1). In addition to this, all the models produced extremely inaccurate response fits for both outputs when validated against the open-loop validation set (DVOL 1) in comparison to the response fits found when validated against the closed-loop validation set (DVCL 1). The last critical observation made was that all

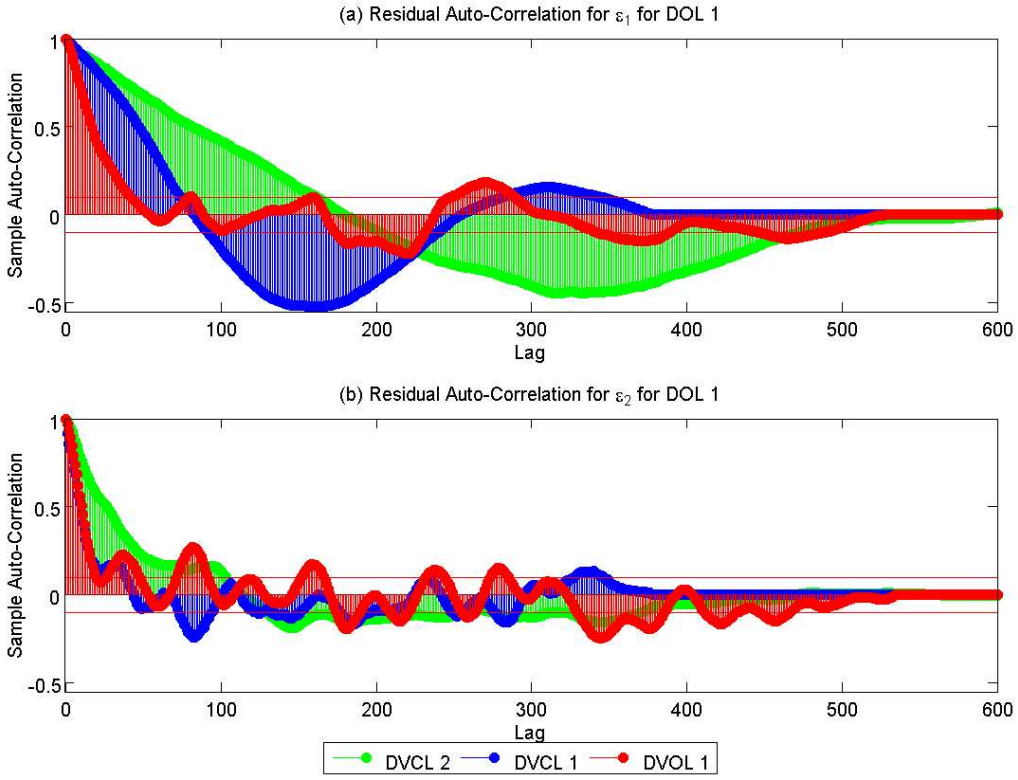


Figure 7.16: Residual auto-correlation for a case DOL 1 against validation sets DVCL 1, DVCL 3 and DVOL 1

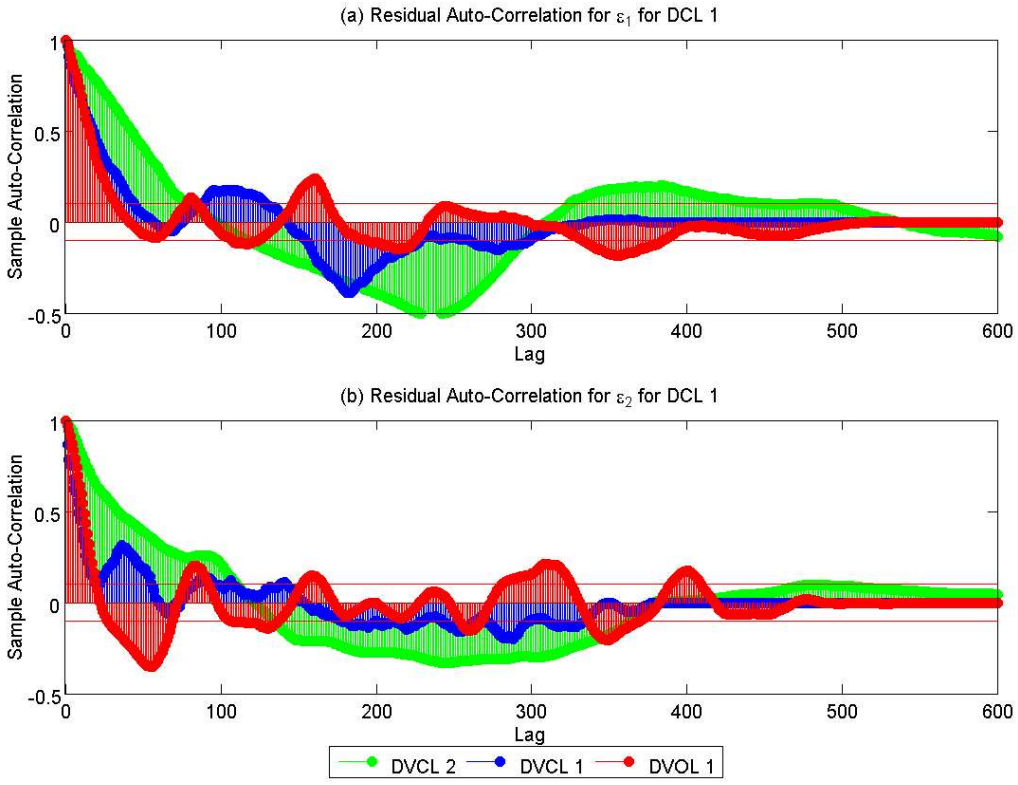


Figure 7.17: Residual auto-correlation for a case DCL 1 against validation sets DVCL 1, DVCL 3 and DVOL 1

the models produced very inaccurate responses when validated against the closed-loop validation set obtained by larger disturbance signals (DVCL 3).

Looking at the residual correlation profiles of cases DOL 1 and DCL 1 in figures 7.16 and 7.17 respectively, the first observation made is the greater distinction in correlations between output y_1 and y_2 for case DOL 1 in comparison to DCL 1. This reflects the findings that the model obtained from open-loop experiments did show the greatest differences in simulated response fits between output y_1 and y_2 of all the identified models.

Further observing these figures it is noted that while the response fits obtained when validated against the open-loop validation set DVOL 1 (figure 7.9) show extremely inaccurate fits that suggest large extents of unmodelled dynamics. The correlation profiles obtained when using DVOL 1 do not suggest significantly different correlation profiles compared to the others. Figure 7.18 shows this point clearly in that figure 7.18(a) shows the correlations profiles obtained when validating the models obtained from experiments DCL 1, DCL 3 and DOL 1 against DVOL 1, while figure 7.18(b) shows the same models validated against DVCL 1.

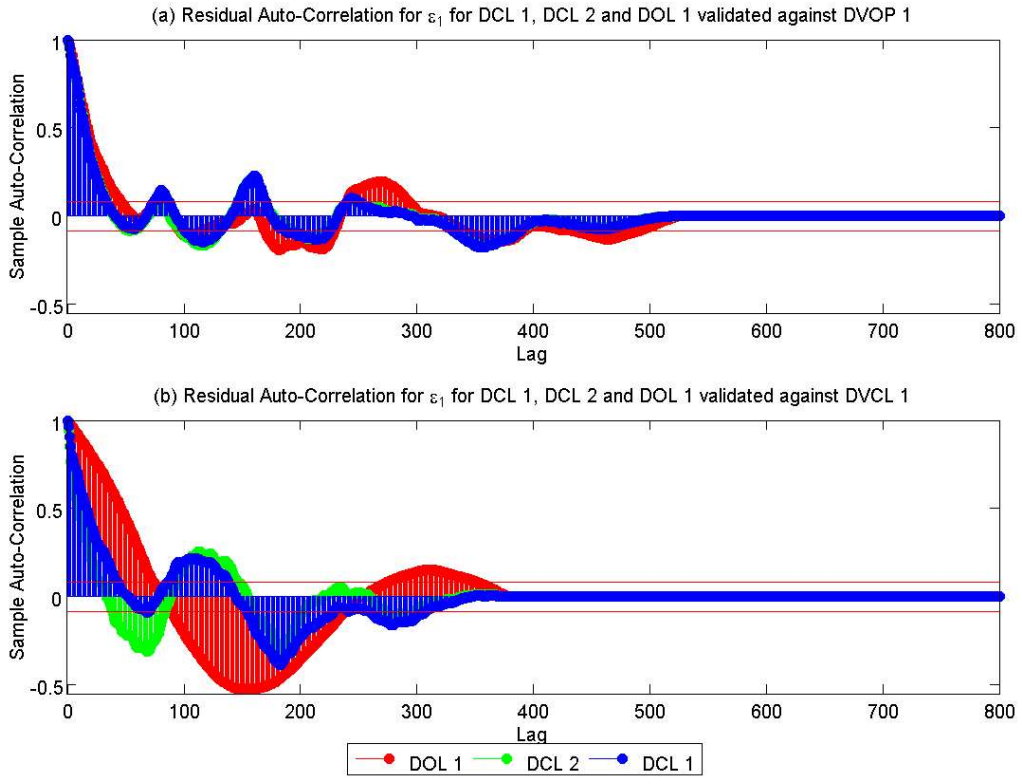


Figure 7.18: Residual auto-correlation for a cases DCL 1, DCL 2 and DOL 1 against validation set DVOL 1

It is additionally observed, looking at figure 7.18, that the correlation profiles obtained for the different models are more distinct when validated against the closed-loop validation set, figure 7.18(b), in comparison to the correlations obtained by the open-loop validation set, figure 7.18(b). This suggest that the closed-loop validation set did a better job at

discriminating between models based on differences in unmodelled dynamics.

Independence Test Results

The input-residual independence test results were found to be more agreeable with the simulated response fit results. Specifically in showing little differences in the correlation profiles between the models generated from experiments DCL 1 and DCL 2 further suggesting the redundancy of the dither disturbance signal due to the continuous disturbance of u_2. It must however be said that all the models showed relatively large amounts of input-residual correlations regardless of the validation set, suggesting incorrectly modelled dynamics.

Figures 7.19(a)-(c) show the input-residual correlation profiles for the model generated by closed-loop experiment DCL 1 validated against validation sets DVCL 1, DVCL 2 and DVOL 1 respectively. The correlations strongly indicate large extents of incorrectly modelled dynamics in that several peaks where found regardless of the validation set. The same extent of input-residual correlations were found for the other input-residual pairs suggesting that no input-output dynamic relationship was better modelled than the other.

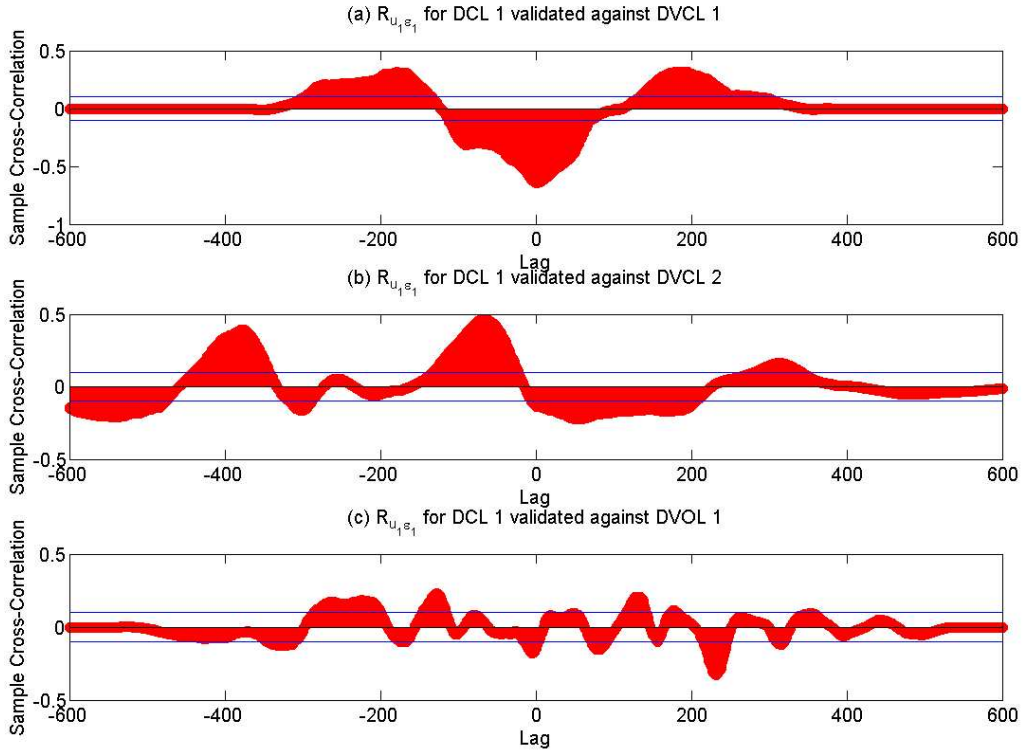


Figure 7.19: Input-residual correaltions for DCL 1

Figures 7.20(a)-(c) show the input-residual correlation profiles in the same fashion as figure 7.19(a)-(c) but for the model generated by the open-loop experiment DOL 1. It

is observed that profiles obtained when validated against the open-loop validation set (DVOL 1 -figure 7.20) are very similar to those obtained when using the closed-loop validation set (DVCL 1 -figure 7.20) except when validated against DVCL 2 (figures 7.19(b)-7.20(b)). This implies that closed-loop validation set DVCL 2 better discriminated between the models based on differences in incorrectly modelled dynamics.

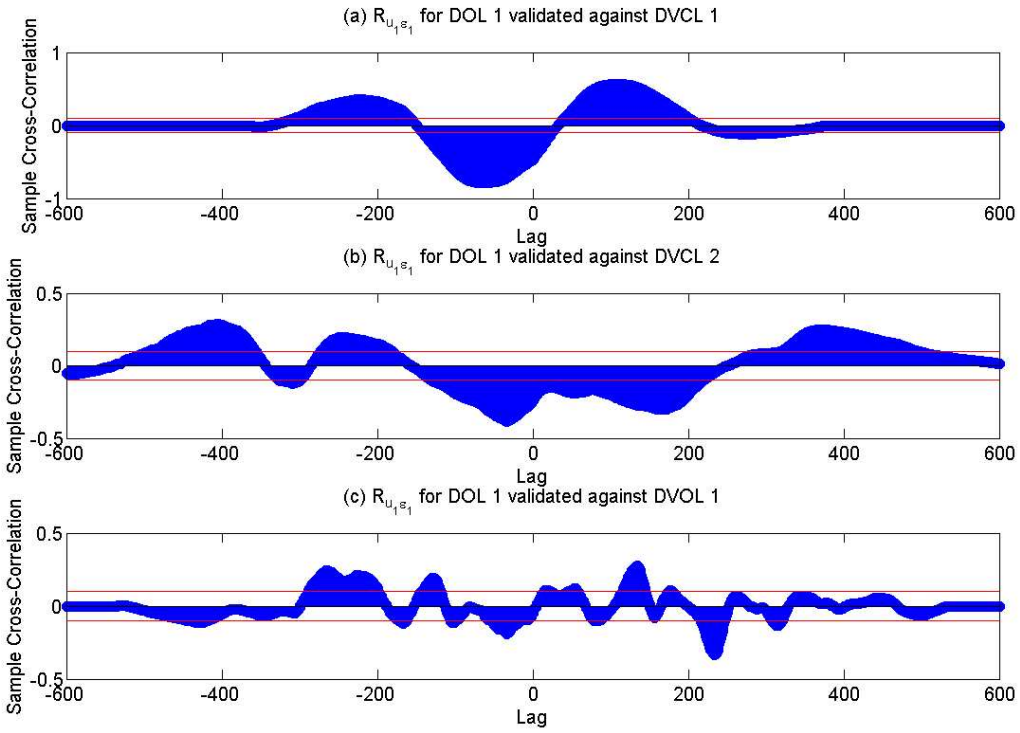


Figure 7.20: Input-residual correlations for DOL 1

It is additionally observed that the correlation profiles using validation data obtained from closed-loop experiments indicate larger correlation magnitudes in comparison to the correlations obtained when using the open-loop validation sets. As has been mentioned before, the closed-loop validation sets are prone to produce input-residual correlations that are rooted in input-output correlations caused by the feedback and not unmodelled dynamics. Determining whether these difference correlation magnitudes are due to differences in models discrimination or due to the correlations caused by feedback is difficult to determine.

Higher Order Correlation Results

The higher order correlation results were relatively uninformative. All the results for all the models against all the validations sets revealed significant higher order correlations between most the residuals - inputs pairs for both $R_{u^2\epsilon}$ and $R_{u^2\epsilon^2}$. This implies, by table 4.1, that both the system and noise models are inadequate and completely failed

to estimate both even and odd non-linear dynamics.

An interesting result was however found in comparing the correlation profiles for $R_{u^2\epsilon}$ to that of $R_{u^2\epsilon^2}$ for the model obtained from the open-loop experiment DOL 1 when validated against the closed-loop validation set DVCL 1 and the open-loop validation set DVOL 1. Figures 7.21 and 7.22 show the correlations profiles obtained for $R_{u^2\epsilon}$ for the model obtained from DOL 1 validated against DVCL 1 and DVOL 1 respectively. It is noted that the correlation profiles concerned with input u_2 are less pronounced when using the open-loop validation set DVOL 1 (figure 7.22(b) and (d)) than when using the closed-loop validation set DVCL 1 (figure 7.21(b) and (d)). This is not seen when analysing $R_{u^2\epsilon^2}$ which, according to table 4.1, implies that the open-loop validation set DVOL 1 indicates the unmodelled dynamics concerned with output u_2 for the model identified from case DOL 1 to be attributed to an inaccurate noise model more so than an inaccurate dynamics model. This observation suggests that the dynamics concerned

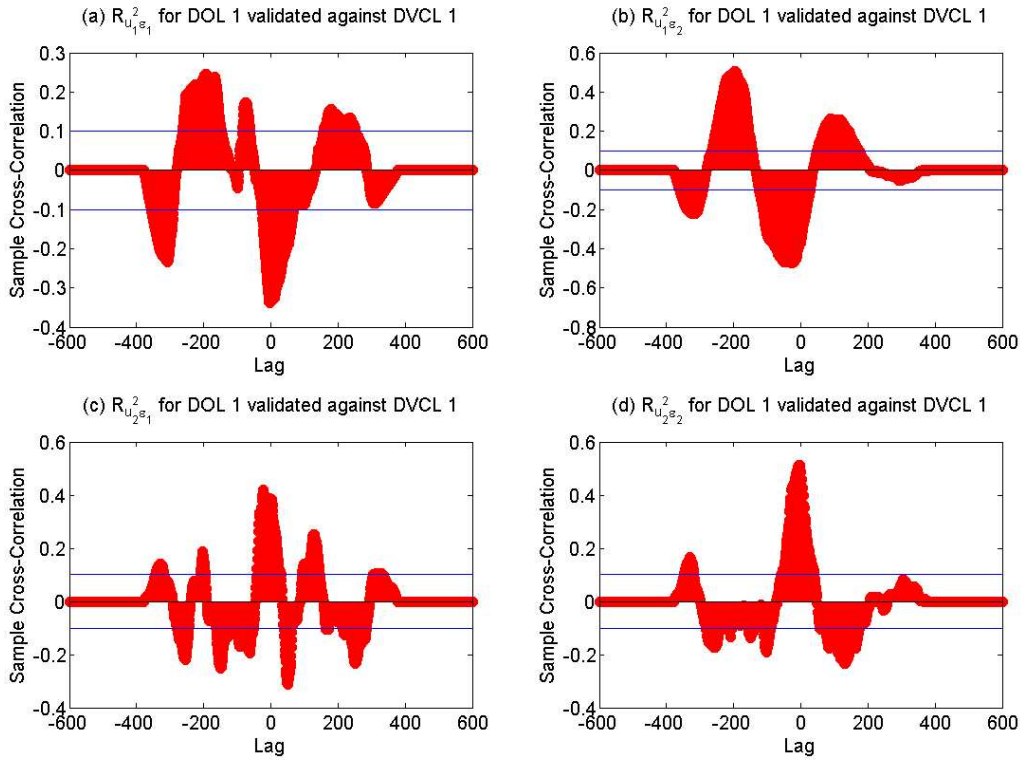


Figure 7.21: Higher order correlation profile, $R_{u^2\epsilon}$, for DOL 1 against DVCL 1

with input u_2 has been more accurately modelled. If this is so then it implies that the open-loop validation set is more informative as it has better discriminated between models so as to detect this condition. Additionally it is very likely that the closed-loop validation set did not detect this conditions in the higher order input-residual correlation due to correlation corruption through the correlations caused by the feedback loop.

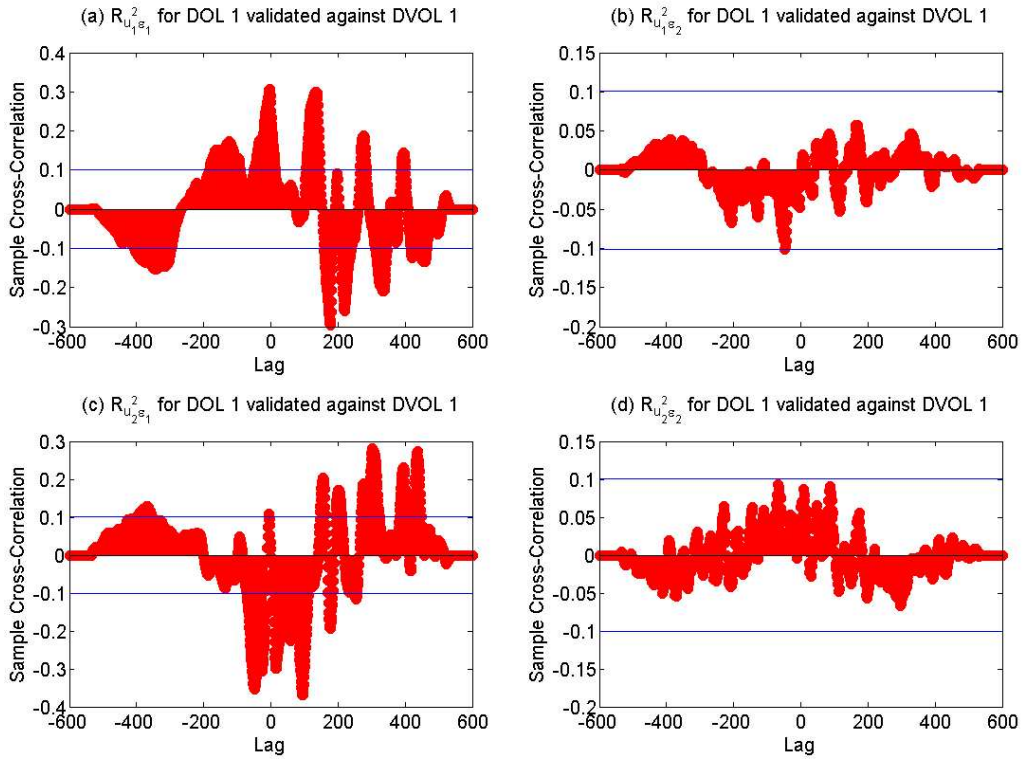


Figure 7.22: Higher order correlation profile, $R_{u^2\epsilon}$, for DOL 1 against DVOL 1

7.3.4 Results Overview

Simulation and Prediction

General Observations:

- The responses of the identified models were found to be very inaccurate when validated against open-loop validation sets while when validated against the closed-loop validation sets the responses were much more accurate.
- The responses of the identified models were additionally found to always be more accurate for output y_2.

Validation Set Model Discrimination and Bias - Closed-loop Validation Sets:

- The closed-loop validation sets were found to be bias towards models obtained from the same feedback condition. This was expected as the output constraints implemented on experiments resulted in a significant difference between open and closed-loop experiments in that the closed-loop experiments were able to use larger input signal magnitudes. This suggested the closed-loop experiments generated more informative data.
- The closed-loop validation sets additionally indicated bias towards models obtained

from experiments that used a similar sequence of setpoint changes to used to generate the validation set.

Validation Set Model Discrimination and Bias - Signal Magnitude:

- When validating the model responses against the closed-loop validation set obtained over a larger output range than that used to generate the normal closed-loop validation set, the accuracy of all the model responses worsened.
- When validating the model responses against the open-loop validation set obtained using larger input signal magnitudes than the normal open-loop validation set, the response accuracy of the model obtained from the open-loop experiment worsened while the response accuracy of the models obtained from closed-loop experiments improved.
- This reduced sensitivity to the signal magnitudes used to generate the open-loop validation set by the models obtained from closed-loop experiments was found to be due to the fact that the closed-loop validation sets used a larger input signal range than the open-loop experiments. This implies that the closed-loop validation experiments identified a larger range of dynamics in comparison to the open-loop validation experiments given the same output constraints.

Experimental Condition Sensitivities:

- As mentioned earlier, there were significant differences in response accuracy between the two outputs. That is, all the models produced responses that were more accurate in y_2 than y_1. The increased excitation of input u_2 by an upstream condition and the consequent generation of more informative data for the output with which this input shared a larger relative gain with (y_2) explains the difference in response fits.
- Further studying this effect however, lead to the finding that the dither signal was to some extent made redundant by this continuous disturbance. Thus experiments that did disturb the system via the dither signals did no not produce models with response accuracies that distinguished themselves from those experiments that did not use dither signal disturbances.

Prediction:

- Prediction results indicated that the identified models produced more accurate responses through prediction than simulation given the prediction horizon was smaller than 200 seconds.
- Prediction response accuracy was addition found to be less sensitive to the experiment conditions used to produce the different validation sets.

Frequency Content Analysis

- Analysing the model response fits validated against each validation set in the frequency domain reflected what the simulated response fit results indicated. The analysis was how ever useful in gaining insight into some differences in response accuracy.

- It was found that the better response fit for output y_2 generated by all the models may be attributed to a better fit of a high frequency component in comparison to the fit for output y_1. This higher frequency corresponds with the frequency of the continuous disturbance of input u_2. Thus affirming the contribution of the continuous disturbance of the system by an upstream condition.

Residual Correlation Analysis

Whiteness Test:

- All the identified models produced responses that generated significant residual auto-correlations regardless of the validation set used. This implies that all the models had large extents of unmodelled dynamics.

- Correlation profile differences were found that further indicate distinctions in response accuracy between output y_1 and y_2 and that suggest that the dynamics concerned with output y_2 were better modelled.

- Differences between correlation profiles when validated against the different validation sets suggested that the closed-loop validation sets better discriminated between models than the open-loop validation sets.

Independence Test:

- As with the whiteness test, the independence tests revealed that all the models identified had residuals with significant input dependencies regardless of the validation set used. This implies that all the models have incorrectly modelled dynamics.

- Differences between correlation profiles obtained by the different validation sets again suggested that the closed-loop validation sets better discriminated between models. Confidence in this finding however is uncertain due to the corruption of the correlation profiles by correlations caused by the feedback loop.

Higher Order Correlation Tests

- The higher order correlation results were found to be less informative than expected.

- All the models generated responses with significant higher order correlations. This implies that all the identified models had inadequate noise and dynamic models.

- The open-loop validation set did however produce correlation profiles that suggest that output y_2 was more accurately modelled.

7.4 Discussion

7.4.1 A 'Real' Identification Problem

Considering the results and findings of the response accuracy, frequency and residual correlation assessments it must be said that the distillation column did serve its purpose in providing an understanding of a few 'real' problems that are not always evident in simulated identification experiments. These 'real' problems predominantly manifested themselves as the effects of the continuous disturbance of input u_2, the steam pressure, by an upstream condition. This disturbance had large effects on the information being generated by the different identification and validation experiments.

It is important to note that while validation results and interpretations suggested these continuous disturbances may have allowed for the identification of models with more accurate responses, it cannot be stated with absolute confidence that these disturbances contributed positively or negatively to the successful identification and validation of the distillation column. This is mostly due to the fact that the validation data sets used to validate these models may have been bias towards the conditions caused by these disturbances and thus not representative of the dominant dynamics of the column.

This is perhaps the greater lesson in studying a 'real' system. Since the system dynamics are unknown, there is no way of comparing the identified models directly with the dynamics of the distillation column, only the validation sets. It is perhaps in this notion that the concept of a useful model being a good model is most applicable. Perhaps it is not necessary to know how well the model represents the true dynamics of the distillation column. Perhaps the understanding that certain models did well to represent certain validation sets is adequate, regardless of how bias the validation set is. That is, given that the column is continuously characterised by these continuous disturbances, then the models found to accurately represent the validation sets will be useful.

7.4.2 Experimental Conditions

In terms of the effects of the different experimental conditions there is a point that has been made relatively clear. That is, the result of implementing constraints on the outputs so as to allow for the identification of a dynamic region most representative of normal operation gave closed-loop experiments the advantage. The models identified from closed-loop experiments were found to produce much more accurate responses while the validation sets generated from closed-loop experiments were found to better discriminate

between models.

7.4.3 Validation Techniques

It must be said that the residual correlation assessments were not very informative. While this was expected of the whiteness and independence tests due to the non-linear nature of the system being identified, it was not expected of the higher order correlation tests. A greater ability of discrimination between models was expected form the higher order tests, a discrimination that would have allowed for insight into how well each model approximated the non-linear dynamics.

The response fit analysis was relatively informative as it clearly indicated how well models represented the different validation sets. As has be stated, it is very clear that the response fit results were very sensitive to the validation sets and the experimental conditions used to generate these validation sets.

CHAPTER 8

Discussion and Conclusions

This chapter presents key discussions relevant to the work presented and the results obtained in preceding chapters. Additionally the general conclusions gathered from the simulated and pilot scale distillation column identification and validation efforts are presented.

8.1 Discussion

8.1.1 Sensitivity to Noise and Model Uncertainty

The topic of model sensitivity to noise and model uncertainty while very briefly addressed in this work is acknowledged as being relatively important and worthy of further discussion. The manner in which both noise sensitivity and model uncertainty were assessed was through the simulated experiments on systems A and B.

In the case of noise sensitivity, identification experiments were repeated several times generating different identification data sets with different noise realisation from which several models were identified. The parameters of such models were assessed for variance, the experiments that produced models with the largest measures of parameter variance were labelled as most sensitive to noise. This has a serious implication. If an experiment was found to be most sensitive to noise then it implies that such an experiment produced identification data that was less informative of system dynamics and consequently more susceptible to corruption by noise. The experiments found to be most sensitive to noise were the open-loop step disturbance experiment and the closed-loop experiment where no setpoint changes were incurred. Both of which were characterised with conditions that were expected to be less informative than other experiments.

In terms of the model uncertainty, this measure was obtained from the parameter

estimation procedure, to be more specific, it was derived from a property of the asymptotic parameter convergence variance, the asymptotic covariance matrix. This parameter was used to infer parameter uncertainty in the identified models. It is important to note that this measure of uncertainty is independent of the actual accuracy of the identified model and solely dependent on the asymptotic variance of parameter convergence. This essentially makes this property a measure of how conducive the data generated from a specific experiment is to efficient parameter convergence. It is important to note however that in the case of identifying models approximating the non-linear system, system B, these uncertainty measures did coincide with models that were expected to be uninformative and did produce inaccurate responses relative to certain validation sets. These models being those obtained from closed-loop experiments where not setpoint changes were incurred.

8.1.2 Closed-loop Vs. Open-Loop

With one of the primary focal points of this work being the difference between open-loop identification and closed-loop identification it is at this point appropriate to discuss the insight obtained during this work with regard to such differences.

The simulated experiments on the linear system (system A) were quick to show that the closed-loop experiments produced informative data such that the models identified from them were more robust than those obtained from open-loop experiments.

To further elaborate on this it is recalled that in identifying the simulated linear system, experimental design variables were varied substantially so as to allow for an idea of their effects on the accuracy of the identified models and discrimination of the validation sets. The primary variables being the type of disturbance signal and the frequency characteristics of the disturbance signal. These variables were varied between values expected to produce informative data, to those expected to produce very uninformative data.

It was found that the models generated by the open-loop experiments were much more sensitive to these design variables while models generated from closed-loop experiment were less sensitive. This meant that optimising the design variables so as to allow for the most informative data had a larger effect on the open-loop cases, they thus produced better models and more discriminative validation sets. However, this also meant that using design variables expected to produce uninformative experiments affected the models obtained from open-loop experiments more than those obtained from closed-loop experiments.

Thus, given the same variance of quality in experimental design variables, open-loop experiments were found to generate the most accurate models and the least accurate models while the closed-loop experiments maintained relatively consistent. It must be

noted however that this is assuming that setpoint changes are made during closed-loop experiments.

The key point from this finding being that given uncertainty in experimental design variables and how optimal they are, closed-loop experiments are more likely to produce more accurate models.

8.1.3 Linear Approximation of Non-linear Systems

In efforts to identify the non-linear system, system B, it was established that model accuracy and validation set discrimination was not only more sensitive to the experimental conditions used to generate them, but also more relative. That is, designing identification and validation experiments was less about generating data that was informative about the true dynamics of the system and more about generating data that reflected the relative purpose of the model. This was primarily due to the fact that the non-linear nature of the system meant it was impossible to truly represent it with a linear model. Additionally, the increased sensitivity to experimental design variables meant that the condition mentioned in the previous section, where models obtained from open-loop experiments were more sensitive to experimental design variances, was compounded.

These conditions characterising non-linear dynamics identification were thus found to create an environment that was best suited for closed-loop experiments. A dominant concept through out the assessment of non-linear system identification was how varying disturbance signal magnitudes and output ranges affected the dynamics exposed during the experiments. Thus, given a linear approximation is to be obtained, there is the clear requirement to generate informative data over a constrained region so as to consistently expose the model estimation procedures to only the relevant dynamics. This meant that closed-loop experiments had the advantage. Closed-loop experiments could more effectively use the input signal to disturb the system while maintaining the output within its constraints.

This was exactly what was found when identifying approximations of the distillation column. Models identified from closed-loop experiments produces responses that more accurately approximated the distillation column responses. Furthermore, the validation sets obtained from closed-loop validation sets were found to better discriminate between models based on the dynamic regions approximated.

8.1.4 Identification and Control

An issue that has had limited attention in this work is that of the direct effects of the controllers themselves during closed-loop experiments. Some interesting considerations did however arise from the work done.

It was established in section 3.3 that literature suggests a non-linear controller would allow for more persistent disturbances of the system to be identified. In efforts to investigate this, non-linear controller action was implemented in specific simulated experiments on system A and system B through varying the PID controller parameters over the duration of the identification experiments. Validation techniques failed to detect any indications or account for any improved accuracy in the models identified from such experiments.

It is important to note that a repeated observation in this work has been that the closed-loop experiments have reduced sensitivity to variances in experimental variables and that this is most likely due to the controller doing its job in dampening the effects of system disturbances. It is interesting to consider the role played by controller performance with respect to this effect. This is of specific relevance to the experiments done on the distillation column. Recalling that the column's steam pressure was being continuously disturbed by an upstream condition, much effort went into reducing the effects of this disturbance in tuning the controllers to perform better. In doing so the controllers were made slightly less robust. This lack of robustness meant a slight variance in controller performance over the 2 ^oC range over which the respective outputs were being identified.

This variable controller performance effectively means variable extents of output disturbance which implies variances in the information generated over the identification range. This means, especially in the case where non-linear dynamics is concerned, that the model approximated from this data will be biased towards the regions where the controllers performed worst as they were most disturbed.

Taking this a step further, implementing such a model in a controller would thus result in very poor robustness since the model is most accurate over an even smaller region than initially designed. From this it can be understood that controller robustness and consistent performance are important in producing unbiased models and validation data.

8.1.5 Simulation Vs. Experimentation and Beyond

The contrast between investigating system identification and validation through simulated experiments to experiments on a real pilot scale system was found to be very large. While the simulated experiments allowed for consistent and unfaltering specification of all the variables, the pilot scale system presented many more variables that were beyond control of the experiment.

This is exemplified in the continuous disturbance of the steam pressure supply to the distillation column which nullified the effects of designed dither signals. Furthermore in order to define the identification problem regarding the distillation column it was necessary to assume any dynamic interactions with variables external to those included

in the model structure being identified were negligible.

This points to the key difference found between simulation and real experimentation, that is, in simulation the system being identified was isolated and closed, while the pilot scale system had external disturbances, some known and possibly some unknown. Determining whether an identified model response inaccuracy is due to unmodelled dynamics between accounted for variables or between unaccounted for variables it difficult. From this it is clear in the efforts presented here to an industrial scale system, which is part of a large process, isolating the dynamics and variables and defining the system becomes a key component in the identification problem.

8.2 Conclusions

The following were successfully achieved as a means to satisfy the objectives of this work:

- Different system identification approaches, specifically those concerned with closed-loop system identification, were surveyed together with different model validation techniques and identification experiment design.
- Identification and validation experiments were simulated on known linear and non-linear systems so as to obtain insight into the sensitivities between identified model accuracy and experimental variables. A specific focus was made on feedback conditions and disturbance signal characteristics.
- Different validation techniques were assessed relative to each other through these simulated experiments, a specific focus was made on cross-validation techniques using data sets generated under open and closed-loop conditions.
- A pilot scale distillation column was used to obtain understanding regarding pragmatic issues of implementing identification and validation techniques on a real system.

The following sections detail some of the key conclusions made considering literature survey findings and the simulated and pilot scale identification and validation experiments.

8.2.1 Model Accuracy Sensitivity to Identification Experiment Conditions

Feedback Condition

- Given no output constraints, models identified from open-loop experiments produced more informative data and consequently more accurate models.

- Open-loop experiment were however found to be more sensitive to experimental conditions. Thus the improved model accuracy is subject to assuring the correct specification of identification experiment variables.
- Given output constraints, closed-loop experiments produced more informative data and identified a larger range of dynamics.
- Closed-loop experiments were found to produce models with improved accuracy at lower frequencies.

Disturbance Signal Type and Characteristics

- Step disturbance signals under open-loop conditions were found to produce the models most sensitive to noise when identifying a linear system
- No specific signal type or characteristic was found to allow for significantly larger measures of model uncertainty when identifying a linear system.
- Experiments where setpoint changes were not incurred were found to produce models with the largest measure of model uncertainty and sensitivity to noise when idenitfying a non-linear system.
- PRMLS disturbance signals were found to be most informative (closely followed by PRBS) for open-loop experiments when identifying a linear system.
- PRBS disturbance signals were found to be most informative for closed-loop experiments when identifying a linear system.
- PRBS disturbance signals were found to be the most consistent and informative for closed and open-loop experiments when identifying a non-linear system.
- Step disturbance signals were found to be the least informative and produced models with the worst response accuracy when validated against validation sets generated from any experiments but open-loop step disturbance experiments.
- Step disturbance signals were however useful in extracting information regarding the extent of non-linear responses and their sensitivity to signal magnitudes.
- Dither signal disturbances were found to be slightly more informative than setpoint changes for closed-loop experiments when identifying a linear system.
- Effects from increased dither signal magnitudes (signal-to-noise ratio) were mostly only revealed in closed-loop experiments were no setpoint changes were incurred when identifying the non-linear system. Increasing the signal magnitude in these cases increased the model response accuracy.

- Setpoint changes were found to be vastly more informative than dither signals for closed-loop experiments when identifying a non-linear system.
- The accuracy of models identified from closed-loop experiments was found to be dependent on the sequence and direction of setpoint changes made.
- The better lower frequency accuracy of models identified from closed-loop experiments was attributed to information generated by setpoint changes.
- Designing the frequency characteristic of the disturbance signal in accordance with dominant time constant (i.e. 30 % the dominant time constant) was found to allow for the identification of models accurate over a larger frequency bandwidth.

8.2.2 Validation Set Bias and Discrimination Sensitivity to Validation Experiment Conditions

Feedback Condition

- Open-loop experiments were found to be most discriminative of models approximating the linear system.
- Closed-loop experiments were found to be most discriminative of models approximating a system revealing non-linear dynamics.
- Validation sets produced from both open and closed-loop experiments indicated bias towards models obtained from similar feedback conditions.

Disturbance Signal Type and Characteristics

- Step disturbance signals for open-loop experiments were found to produce the most biased validation sets.
- Designing the frequency characteristic of the disturbance signal in accordance with dominant time constant (i.e. 30 % the dominant time constant) was found to allow for the generation of validation data more representative of the true system.
- The bias and prejudice of the validation sets obtained from closed-loop experiments was found to be sensitive to the sequence and direction of setpoint changes made.

8.2.3 Validation Techniques

The dominant approach used in this work to validate the models was that of cross-validation. While other methods were used in the case of the linear system, they were

mostly in efforts to validate the validation results to obtain indications of how bias or prejudice they were.

Measuring the cross-validation results as response percentage fits was useful in that quantitative measures were generated representing model accuracy. In some cases however these measures did not allow for sufficient discrimination between models as the response fits were very similar.

The use of discrete Fourier transforms did prove very valuable in translating the simulated response fit assessments into the frequency domain. From these assessments model accuracy at specific frequencies and validation set discrimination at specific frequencies were made evident.

Model validation through residual correlation was found to be useful in discriminating between models based on unmodelled and incorrectly modelled dynamics. However, this was only the case for approximations of the linear system and when using open-loop validation sets. Closed-loop validation sets did produce residual correlations that suggested a greater capacity to discriminate between models compared to the open-loop validation sets but this could not always be trusted due to correlation profile corruption by correlations rooted in the feedback loop.

Of all the validation approaches the higher order correlation tests were expected to be most informative and most discriminative based on how each linear model approximated the non-linear system. Unfortunately this was not the case to the extent that this approach was most uninformative.

8.2.4 Linear Approximation of Non-linear Dynamics

As the points made in the previous sections indicate, the closed-loop experiments were quick to establish themselves as the better alternative over open-loop experiments when attempting to identify an approximation of a non-linear system. This was found to be due to the fact that disturbance signal magnitudes and output ranges had large effects on the dynamics revealed from non-linear systems. This is such that in order to successfully approximate non-linear dynamics with a linear model, the output signal magnitudes and ranges and the consequent dynamic regions to be approximated must be clearly stated and the data for identification and validation definitively generated within this range. The closed-loop experiment's capacity to generate more informative data over open-loop experiments given output constraints make it the best candidate for non-linear system approximation.

8.2.5 Identification and Validation of a real System

The primary conclusions obtained from identification and validation efforts on the pilot scale distillation column are such that there are many pragmatic and real problems in

identifying a real system. First and foremost system isolation and defining the identification problem and all the respective variables concerned is not so easily done. Real systems will always by disturbed by some upstream condition of have some unforeseen interaction with an unaccounted for variable.

8.2.6 General Conclusions

For the sake of generalising it may be said that the improved persistence of system excitation of open-loop experiments was found to produce the most accurate models given well defined and designed disturbance signals. This was especially the case if the system was well represented by a linear approximation and not restricted by system output constraints.

In examining the identification of non-linear systems, it was found that if the system has a high degree of non-linear dynamics and thus can not be well represented by a linear model, closed-loop experiments were most effective in that they allowed for accurate implementation of bounds on outputs allowing for data generation over strict regions where linear approximations accurately represented the non-linear system.

In terms of cross-validation efforts, it is concluded that when identifying a system that may accurately be represented by a linear model, there is merit in assuring the validation set is representative of the true system. However, if approximating a non-linear system with a linear model it is more valuable to determine whether the validation set is representative of the model purpose.

CHAPTER 9

Recommendations and Further Investigation

In this chapter recommendations are made with respect to system identification and validation approaches based on results and conclusions presented in this work. Additionally suggestions into further studies are made in terms of advancing the successful identification of the pilot scale distillation column and other topics of relevance that would be worthy of further investigation.

9.1 Recommendations

Given the conclusions stated in the previous chapter, an understanding towards an approach to system identification and validation is evident.

It is recommended that in cases where the linearity of the system is well known and the purpose of the model is well defined, open-loop experimentation be used to generate data for validation and identification of models. This must however be coupled with an extensive design of the characteristics of the disturbance signals with respect to the model's intended use.

In cases where the system is expected to be non-linear, or the extent of linearity is unknown, closed-loop experiments are suggested given an accurate definition of the dynamic range of interest. It is further recommended that when using closed-loop experimental conditions that setpoint changes be made where possible.

In terms of the validation of identified models, given that the identification and validation of any 'real' system will more often that not imply that the true dynamics of the system are unknown, cross-validation is recommended. In which case the same recommendations to identification experiments are made. If the system may be accurately represented by linear dynamics then use open-loop experiments. If the system being approximated is non-linear or of an unknown extent of non-linearity, then it becomes

important to establish the purpose of the model and designed closed-loop validation experiments that reflects such a purpose.

9.2 Further Investigation

9.2.1 Identification of Linear Approximations of Non-linear Systems

While this work attempted to uncover some issues of identifying linear models of non-linear systems and the validation of such systems, there is much to investigate further.

This work concentrated on experimental design variables, specifically feedback conditions and disturbance signals and how these variables affected the approximation of non-linear systems. There are however many other variables and conditions to further investigate. Of specific interest would be variables defining the structure of the linear model used. Investigating the effect of independence between the noise model and the dynamic model on the accuracy of the linear approximation would probably produce some interesting results.

In addition to looking at other model structures and orders assessing the model accuracy sensitivities to different model parameter estimation routines would be useful. Investigating the optimisation and sensitivities of different prediction error minimisation norms and frequency weighting functions could produce greater understandings of the ideal identification and estimation conditions that would allow for the convergence towards the best linear approximations.

Furthermore, while section 4.2.1 presented the LTI second order equivalent as the best linear approximation of a non-linear system on a theoretical basis, investigating the incorporation of this concept of the best possible linear approximation of a non-linear system into validation techniques would be extremely valuable.

9.2.2 Identification of the Pilot Scale Distillation Column

With respect to the efforts in modelling the pilot scale distillation column, it must be made evident that the less intensive, easily adaptable parametric black box model structures can be most accommodating provided a purpose for the model is kept in mind. Deriving an accurate model of the column based on first principles will be difficult and would in any case most likely require the incorporation of adjustable parameters to account for the many inefficiencies and unique dynamic characteristics.

It is noted however that a black box parametric model does have its limitations in identifying the column. That is, parametric estimates would be limited to a certain range of initial conditions. Attempting to develop a parametric model that accurately describes

the dynamics of the column given a mixture besides water and ethanol is being separated would be very difficult, if at all possible.

In terms of using linear models to represent the column's dynamics, establishing the different ranges that are accurately represented by different linear approximations would be very valuable. However, a non-linear model would do well in identifying the column.

9.2.3 MPCI and Other Recommendations

From the perspective of control, the ultimate model validation would be how well the model predictive controller performs with an identified model. Extending the understanding of experimental design for models to be used in MPC is extremely relevant to closed-loop system identification and validation. A more in depth study would also investigate adaptive model predictive control, or MPCI, where system identification and validation techniques are implemented online in efforts to continuously update the model being used.

From this further investigation into MPCI it is almost certain that model validation will need to be emphasised. Validating the different models being generated from a continuous stream of data while varying the identification variables is an interesting problem, specifically if using a linear approximation of a non-linear system.

Last but not least, studying the effects of closed-loop identification experiments on controller performance and further establishing clear boundaries beyond which closed-loop identification would no longer be successful would allow for an understanding of the limitations of implementing this approach.

APPENDIX A

Software and Identified Models

A.1 Description of Software Used and Compiled

All the programs written and used in this dissertation are provided in the attached DVD in the 'SOFTWARE' folder. Within the folder there are two subsequent folders, 'FUNDAMENTAL' and 'EVERYTHING'. The folder entitled 'EVERYTHING' contains all the matlab m-file and data files used and generated while the 'FUNDAMENTAL' folder contains all the matlab m-files and data files that form the fundamental components of the work.

The 'EVERYTHING' folder is further divided into three folders containing files used and generated when working on linear, non-linear and distillation column respectively.

To illustrate the use of the different functions and software, the fundamental components contained in the 'FUNDAMENTAL' folder are illustrated by figure A.1 and further described in following sections.

A.1.1 Signal Generation

The file Inputgen.m was the function used to generate the disturbance signals. The function calls on the Matlab function *idinput* to generate the various disturbance signals given specified signal characteristics i.e. type, magnitude, frequency, etc.

A.1.2 Identification and Validation Data Generation

Once the disturbance signals were identified they were used to generate data for identification and validation. While no programming was required to do this when using the pilot scale distillation column, the simulated experiments on systems A and B did.

Function	Signal Generation	Data Generation	Model Generation	Model Validation
File	InputGen	ArxModel/ArxControl/ NarxModel/NarxControl	ModelGen	SimuFit/Frequency/ Correl/HighCorrel
Description	Used to Generate Disturbance Signals: • White Gaussian • PRBS • PRML	Used to Simulate Open and Closed-loop responses to disturbances for the linear and non-linear systems, systems A and B respectively.	Used to estimate the ARX models approximating the system being identified from the identification data	Used to Conduct the various validation techniques: • Response Simulation and Prediction % fit • Frequency Response Analysis • Response fit frequency content analysis • Residual correlations

Figure A.1: Illustration of the fundamental software components and the associated matlab m-files.

With system A being an ARX model and system B being a NARX model, the files used to generate responses for the two models under open and closed-loop conditions are defined as follows:

ARXmodel.m : Generates open-loop responses for system A given specifications of disturbances signals (inputs) and noise.

NARXcontrol.m : Generates closed-loop responses for system A given specifications of disturbances signals (inputs), noise, setpoint changes, time of setpoint changes, controller parameters.

ARXmodel.m : Generates open-loop responses for system B given specifications of disturbances signals (inputs) and noise.

NARXcontrol.m : Generates closed-loop responses for system B given specifications of disturbances signals (inputs), noise, setpoint changes, time of setpoint changes, controller parameters..

A.1.3 Model Generation

The m-file ModelGen.m illustrates the use of the *arx* function used to generate the parametric model given input and output data, model structure specifications and output delays.

A.1.4 Model Validation

Simulation and Prediction Fit

The m-file SimuFit.m illustrates the use of the *compare* function used to generate model responses and response percentage fit values for simulated responses and prediction responses. The last parameter in the function, n, specifies the prediction horizon in number of simples, if this is omitted the prediction horizon is made very large and thus the response is a simulation.

Frequency Analysis

Recalling that two types of frequency analysis were used. One being the analysis of the frequency response of the identified model in the same fashion as that generated by bode diagrams. The other being the analysis of the simulated response in the frequency domain via the use of discrete Fourier transforms.

The m-file FreqAnal.m illustrates both frequency analysis types.

Residual Correlations

Recalling that the both normal residual correlations and higher order correlations were done, the m-file Correl.m and HighCorrel.m illustrate respectively. Both correlation m-files call the correlation generation function *crossco*.

A.2 Model Parameters of Identified Models

See attached DVD, model parameter tables are given in ModelParameterTables.pdf found in the 'DOCUMENTATION' folder.

BIBLIOGRAPHY

Akaike, H. (1967) *Spectral Analysis of Time Series*, Wiley, .

Anderson, B. and Gevers, M. (1982) "Identifiability of Linear stochastic Systems Operating Under Linear Feedback", *Automatica, 18*, 195–213.

Anscombe, F. (1973) "Graphs in statisical analysis", *Amer. Statistician, 27*, 17–21.

Anscombe, F. and Tukey, J. (1963) "The Examination and Analysis of Residuals", *Technomentrics, 5, No. 2*, 141–160.

Åström, K.; Bohlin, Y. and James, J. "Numberical identification of linear dynamic systems from normal operating records", in *Proceedings IFAC Symposium on Self-Adaptive Systems*, (1965).

Atsushi, N. and Goro, O. (2001) "Closed-Loop Identification and Control Using Neural Networks", *Jido Seigyo Rengo Koenkai Maezuri (Japan), 44*, 368–369.

Bendat, J. (1991) *Nonlinear system analysis and identification from random data*, Jon Wiley and Sons, .

Billings, S. A. and Voon, W. S. F. (1986) "Correlation based model validity tests for non-linear models", *International Journal of Control, 44*, 235–244.

Billings, S. and Zhu, Q. (1994) "Nonlinear model validation using correlation tests", *International Journal of Control, 60*, 1107–1120.

Billings, S. and Zhu, Q. (1995) "Model validation tests for multivariable nonlinear models including neural networks", *International Journal of Control, 62*, 749–766.

Blackman, R. and Tukey, J. (1958) "The measurement of power spectra from the point of view of communication engineerin", *Bell Syst. Tech Journal, 37*, 183–282.

Box, G. E. P. and Jenkins, G. M. (1970) *Time series analysis*, Holden-Day, .

Box, G. (1979) *Robustness in the strategy of scientific model building, in Robustness in Statistics*, Academic Press: New York, .

Broch, J. (1990) *Principles of experimental frequency analysis*, Elsevier Applied Science, .

Chien, I. (1996) "Simple empirical nonlinear model for temeprature-based high-purity distillation columns", *AICHE Journal, 42*, 2692–2697.

Cook, D. and Weisberg, S. (1982) *Residuals and Influence in Regression*, Chapman and Hall, New York, .

Cressie, N. and Read, T. (1989) "Pearson's X and the loglikelihood ratio statistic G: a comparative review", *International Statistial REview, 57*, 19–47.

Desgupta, S. and Anderson, D. (1996) "A parametrization for the closed-loop identification of nonlinear time-varying systems", *Automatica, 32*, 1349–1360.

Doma, M.; Taylor, P. and Vermeer, P. (1996) "Closed loop identification of MPC models for MIMO processes using genetic algorithms and dithering one variable at a time: Application to an industrial distillation tower", *Computers and Chemical Engineering, 20*, 1035–1040.

Draper, N. and Smith, H. (1998) *Applied Regression Analysis*, Wiley, New York, .

Enqvist, M. and Ljung, L. "LTI approximations of slightly nonlinear systems: some intriguing examples", in *Proc. NOCLOS 2004-IFAC*, (2004).

Enqvist, M.; Schoukens, J. and Pintelon, R. "Detection of Unmodeled Nonlinearities Using Correlation Methods", in *Instrumentation and Measurement Technology Conference Proceedings*, (2007).

Eskinat, E.; Johnson, S. and Luyben, W. (1991) "Use of Hammerstein models in identification of nonlinear systems", *AIChE Journal, 37* (2), 255–268.

Evans, A. and Fischl, R. (1973) "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters", *IEEE Translations on Audion and Electro-Acoustics, AU-21*, 61–65.

Forssell, U. (1997) "Properties and Usage of Closed-loop Identification Methods", Master's thesis, Department of Electrical Engineering, Linköping University,.

Forssell, U. (1999) "Closed-loop identification: Methods, Theory, and Applications", Master's thesis, Department of Electrical Engineering, Linköping University,.

Forssell, U. and Ljung, L. (1997) "Issues in closed-loop Identification", Technical report, Dept of EE. Linköping University,.

Forssell, U. and Ljung, L. (1999) "Closed-loop identification revisited", *Automatica, 35*, 1215–1241.

Forssell, U. and Ljung, L. (2000) "A projection method for closed-loop identification", *Automatic Control, IEEE Transactions on, 45* (11), 2101–2106.

Frank, P. (2000) "Modelling for fault detection and isolation versus modelling for control", *Mathematics and computers in simulation, 52*, 259–271.

Fujimoto, K.; Anderson, B. and De Bruyne, F. (2001) "A parameterization for closed-loop identification of nonlinear systems based on differentially coprime kernel representations", *Automatica, 37*, 1893–1907.

Gevers, M. "Identificaiton for control", in *Proc. 5th IFAC symposium on Adaptive Control and Signal Processing*, (1986).

Gevers, M. (2006) *System identification without Lennart Ljung : what would have been different ?*, pages 61–85 Studentlitteratur - T. Glad and G. Hendeby.

Goodwin, G. and Payne, R. (1977) *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, .

Gustavsson, I.; Ljung, L. and Söderström, T. (1977) "Identification of Processes in Closed-Loop - Identifiablity and Accuracy Aspexts", *Automatica, 13*, 59–75.

Haber, R. and Keviczky, L. (1999) *Nonlinear System Identification: Input-output Modeling Approach, Volume 2*, Springer, .

Hansen, F. (1989) *A fractional representation aproach to closed-loop system identification and experiment design*, PhD thesis, Stanford University, CA, USA,.

Hjalmarsson, H. "Detecting asymptotically non-vanishing model uncertainty", in *12 IFAC World Congress Proceedings, 5,63-68*, (1993).

Ho, B. and Kalman, R. (1965) "Effectice construction of linear state-variable models from input-output functions", *Regelungstechnik, 12*, 545–548.

Huang, B. and Shah, S. (1997) "Closed-loop identification: a two step approach", *Journal of Process Control, 7*, 425–438.

Johnson, R. and Wichern, D. (2007) *Applied Multivariable Statistical Analysis*, Pearson Prentice Hall, .

Lin, W.; Qin, S. and Ljung, L. "Comparison of subspace identification methods for systems operating in closed loop", in *Proc. 16th IFAC World Congress*, (2005).

Linard, N.; Anderson, B. and de Bryne, F. (1999) "Identification of a nonlinear plant under nonlinear feedback using left coprim fraction based representations", *Automatica, 35*, 655–667.

Ljung, L. (1976) "Consistency of the least-squares identification method", *Automatic Control, IEEE Transactions on, 21* (5), 779–781.

Ljung, L. (1978) "Convergence analysis of parametric identification methods", *IEEE Transactions of Automatic control, 23*, 770–783.

Ljung, L. (1985) "Asymptotic variances expressions for identified black-box transfer function models", *IEEE Trans. Automatic Control,* pages pg834–844.

Ljung, L. (1987) *System Identification, Theory for the user*, Prentice Hall, Englewood Cliffs, NJ.

Ljung, L. (1995) *System identification toolbox: for use with MATLAB*, The Math Works Inc., Mass, USA, .

Ljung, L. (2002) "Prediction error estimation methods", *Circuits, Systems, and Signal Processing, 21* (1), 11–21.

Ljung, L. and McKelvey, T. (1996) "Subspace identification from closed loop data", *Signal Processing, 52*, 209–215.

Mäkilä, P. and Partington, J. (2003) "On linear models for nonlinear systems", *Automatica, 39*, 1–13.

McLellan, J. (2005) "Dynamic Experiments - Maximizing the information content for control applications", Technical report, Department of Chemical Engineering, Queen's University, Kingston, Ont. Canada,.

Montgomery, D.; Runger, G. and Hubele, N. (2001) *Engineering Statistics*, John Wiley and Sons, .

Nooraii, A.; Romagnoli, J. and Figueroa, J. (1999) "Process identification, uncertainty characterisation and robustness analysis of a pilot-scale distillation column", *Journal of Process Control, 9*, 247–264.

Norgaard, M. (2000) "Neural Network Based System Identification Toolbox", Technical report, Technical Univeristy of Denmark, Department of Automation,.

Norton, J. (1986) *An introduction to identification*, Harcourt Brace Jovanovich, .

Nowak, R. and Veen, B. (1994) "Random and pseudorandom inputs for volterra filter identification", *IEEE transaction on signal processing, 42(8)*, 2124–2144.

Pearson, R. and Pottmann, M. (2000) "Gray-box identification of block-oriented nonlinear models", *Journal of Process Control, 10*, 301–315.

Pintelon, R. and Schoukens, J. (2001) *System Identification - A Frequency Domain Approach*, IEEE Press, .

Raol, J. and Sinha, N. (2000) "Advances in Modelling, System Identification and Parameter Estimation", *Sadhana, 25 (2)*.

Rayner, J. (1971) *An Introduction to spectral analysis*, Pion limited, .

Schetzen, M. (1980) *The Volterra And Wiener Theories of Nonlinear Systems*, John Wiley and Sons, .

Schoukens, J.; Pintelon, T.; Dobrowiecki, T. and Rolain, Y. "Identification of linear systems with nonlinear distortions", in *13th IFAC Symposium on System Identification*, (2003).

Shouche, M.; Genceli, H.; Vuthandam, P. and Nikolaou, M. (1998) "Simultaneous Constrained Model Predictive Control and Identification of DARX Processes", *Automatica, 34*, 1521–1530.

Shouche, M.; Genceli, H. and Nikolaou, M. (2002) "Effect of online optimization techniques on model predictive control and identification (MPCI)", *Computer and Chemical Engieering, 26*, 1241–1252.

Sjöberg, J.; Zhang, Q.; Ljung, L.; Benveniste, A.; Deylon, B.; Glorennec, P.-Y.; Hjalmarsson, H. and Juditsky, A.) "Nonlinear Black-Box Modeling in System Identification: a Unified Overview", Technical report, Dept of EE. Link .

Skogestad, S. and Postlethwaite, I. (1997) *Multivariable Feedback Control*, John Wiley and Sons, .

Söderstrom, T. and Stoica, P. (1989) *System Identification*, Prentice Hall, New York, .

Söderström, T.; Stoica, P. and Trulsson, E. "Intrsumental variable methods for closed loop systems", in *In Preprints IFAC 10th World Congress on Automatic Control, pg 364 -369*, (1987).

Steiglitz, K. and Mcbride, L. (1965) "A Technique for Identification of Linear Systems", *IEEEE Transactions on Automatic control, AC-10*, 467–464.

Stenman, A.; Gustafsson, F.; Rivera, D.; Ljung, L. and McKelvey, T. November (2000) "On adaptive smoothing of empirical transfer function estimates", *Control Engineering Practice, 8* (11), 1309–1315.

Sung, S. and Lee, J. (2003) "Pseudo-random binary sequence design for finite impulse response identification", *Control Engineering Practice, 11*, 935–947.

Toker, O. and Emara-Shabaik, H. (2004) "Pseudo-random multilevel sequences: Spectral properties and identification of Hammerstein Systems", *IMA Journal of Mathematical Control and Information, 21*, 183–205.

van den Hof, P. and Boker, J. "Identification with generelized orthonormal basis functions - statistical analysis and error bounds", in *10th IFAC Symposium on System Identification*, , 3 , 3 pages 3.207–3.212 (1994).

van den Hof, P. (1998) "Closed-loop issues in system identification", *Annual Reviews in Control, 22*, 173–186.

van den Hof, P. and Schrama, R. (1993) "An indirect method for transfer function estimation fro closed-loop data", *Automatica, 29*, 1523–1527.

van den Hof, P.; Schrama, R.; de Callafon, R. and Bosgra, O. (1995) "Identification of normalized coprime factors from closed-loop experimental data.", *European Journal of Control, 1*, 62–74.

Vidyasagar, M. (1985) *Control system synthesis: A factorisation approach*, Cambirdge, MA: MIT press, .

Zheng, Q. and Zafiriou, E. "Control-relevant Identification of Volterra Series Models", in *Proc. American Control Conference*, (1994).

Zhu, Y. (1998) "Multivariable process identification for MPC: the asymptotic method and its applications", *Journal of Process Control, 8* (2), 101–115.

Zhu, Y. and Butoyi, F. (2002) "Case studies on closed-loop identification for MPC", *Control Engineering Practice, 10* (4), 403–417.