Chapter 3

Modelling

THIS chapter describes the use of system identification techniques to derive a relationship between variables in the mould and surface defects. In essence, this relationship is a model with mould variables as inputs and defects as outputs. The model can be used to predict when defects will occur, hence it is called a predictor.

First principles methods would be an ideal way to derive a model to predict defects (see §2.2.1.10). A first principles model requires metallurgical modelling at a micro-structure level; a process that is not fully understood in the continuous casting process today. It involves variables such as dendritical arm growth, chemical composition, strain and stress, to name only a few. Except for this vast number of variables which have to be formulated into a mathematical model, the training and testing of such a model require vast amounts of data from several different types of sensors which do not exist.

The use of artificial neural networks (ANN) [17, 18] or expert systems [15] to predict defects has been implemented in practice, though very little information on such systems is available in the literature, with most designs being proprietary. The neural network modelling approach does not easily allow physical meaning of variables to be worked into the model *i.e.* first-principle information that could have been used is lost. Such black box type models are also undesired since control system analysis techniques do not truly exist for neural networks. This implies that controller design using ANN is largely limited to empirical tuning methods. The usefulness of the neural network or expert system does not go beyond it being a predictor of defects.

System identification (*i.e.* empirical modelling) techniques have to be used, because a first principles model is beyond the scope of this study, and AI methods have either been used or are undesirable. This method of modelling is also a black box approach, but the selection of the structure of the model can be facilitated by an understanding of first-principle models as described in Chapter 2. The system identification technique used in this chapter is in essence a Box-Jenkins type approach known as auto-regression with exogenous input. It operates as a linear discrete filter, with past values of inputs and outputs weighed to give current outputs. The weights (regressors) mentioned are determined using regression with plant input and output data, and are tested on plant input and output data sets which are the same or differ from the original training set. The modelling technique is simple, effective and manageable given the large model structures and sets of data to work with, and can be used for analysis and design using standard control system techniques.

This chapter starts with a description of the mould variables (inputs) and defect (outputs) data. The data are then perused using statistical hypothesis testing and correlation analysis to determine if some variables can be excluded from the model. An interesting result that will be covered is that the model can be split into two sub-models, making feedback control possible. The models are then derived using system identification techniques. Finally, some concluding remarks are made.

3.1 Data

This section describes the mould variable data and defect data. The data were gathered at a South African stainless steel producer over a period of six months from May to September, 1999. A validation set of data were gathered in June, 2002. The data can be categorised by the inputs[a] which are the mould variables such as casting speed, thermocouple temperatures *etc.* and the outputs which are defect data such as transversal cracks, inclusions, depressions *etc.* These data are required to derive a model to predict the occurrence of defects based on variation of parameters in the mould. The concept is illustrated in Fig. 3.1. The following sections describe the mould variable data, defect data and processing of the data.

3.1.1 Mould variable data

There are numerous variables[b] that are measured within the mould. The data are usually gathered on the level 1 system of the company, and stored on the level 2 system. Altogether

[a]Note that disturbances also act on the system and these will also be described in the following sections.
[b]see *e.g.* Fisher and Mesic [150] for a description of the database structures at a continuous casting plant.

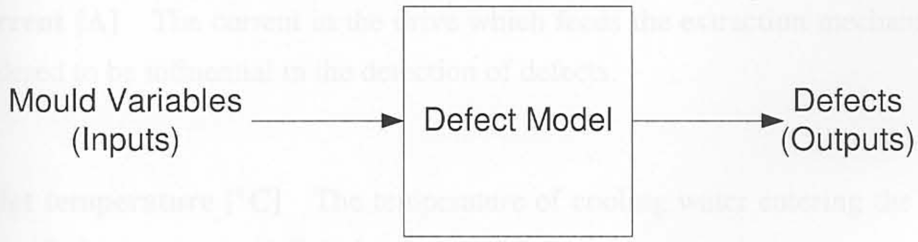

Figure 3.1 Block diagram of the total transfer from mould variables (inputs) to defects (outputs).

800 slabs were inspected for defects over the 6 month time period, but data for only about 500 slabs were available due to errors in the data gathering system which were caused by down-time or maintenance of the system. This is a small percentage of actual cast product, because slab inspection of every slab that was cast was not possible due to man-power constraints. About 3.3GB of mould variable data were collected.

3.1.1.1 Mould variables

The following variables characterise the operation of the mould. They pertain to the inputs in Fig. 3.1. Data for each of these variables were obtained from the level 2 system of the plant. These data are sampled at two second intervals.

Casting speed [mm/min] The casting speed is the speed at which the strand moves through the caster. It can be classified as a manipulated variable.

Mould level [mm] This indicates the fluctuations of the mould level at the meniscus and is measured by an Eddy-current sensor. The general consensus is that mould level should not fluctuate greatly, and should remain constant at a predefined value. At the industrial partner, mould level is controlled using the slide gate of the tundish as an actuator. The mould level can be classified as a measured disturbance.

Mould level controller activation [binary 0 or 1] The mould level controller activation is a binary switch which indicates whether the automatic mould level controller is switched on or whether mould level is manually controlled. The effect of the switch's position is seen in the variation of the mould level.

Drive current [A] The current in the drive which feeds the extraction mechanism and is not considered to be influential in the detection of defects.

Water inlet temperature [°C] The temperature of cooling water entering the mould. It can be classified as a measured disturbance.

Negative strip [mm] The length by which the mould moves faster downward relative to the strand during the oscillation of the mould. This value remains constant for all casts.

Oscillation frequency [Hz] The frequency at which the mould oscillates. The oscillation is sinusoidal. It is considered to be a manipulated variable.

Water flow rate [l/min] The flow-rate at which cooling water flows into each of the four mould faces. It should remain at the maximum possible value. It is therefore uncontrollable and is considered to be a measured disturbance.

Thermocouple temperatures [°C] These temperatures are measured through thermocouples situated in two rows with 8 on each wide face at both the top and bottom rows and 2 on each narrow face for both the top and bottom rows. Fig. 3.2 shows the location and naming conventions for the thermocouples. No classification can be awarded to this variable since it does not fall within the category of a manipulated variable or a disturbance.

Heat fluxes $[kW/m^2]$ The heat flux in each of the four faces of the mould. This variable also remains unclassified since it is not a manipulated variable or a disturbance.

Delta T [°C] A variable which describes the overall temperature distribution of the mould in each of the four faces. This variable also remains unclassified since it is not a manipulated variable or a disturbance.

Longitudinal temperature differences [°C] The temperature difference between the top row of thermocouples and the bottom row of thermocouples. These variables have tradi-

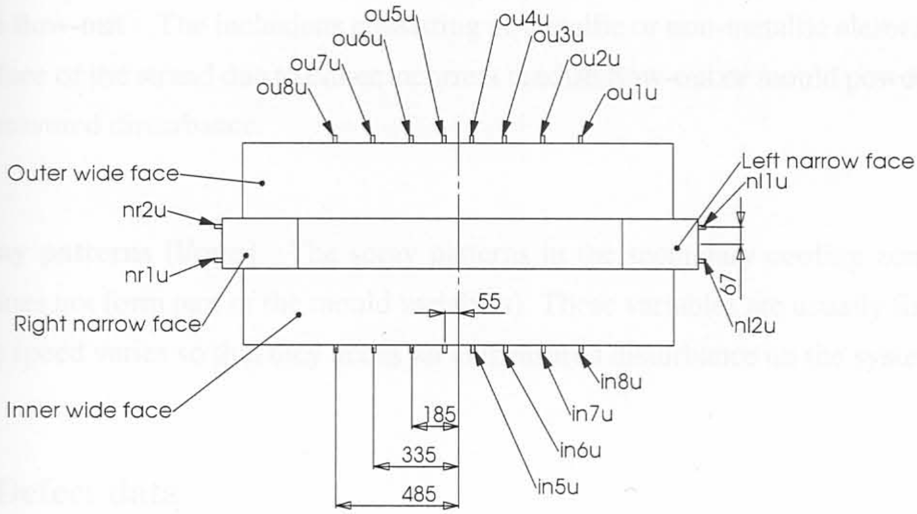

Figure 3.2 Top view of the mould depicting location and naming conventions of the thermocouples. Note that in this view, the thermocouples shown are the top row of thermocouples. The bottom row of thermocouples have names "ou8*l*", "nr1*l*", *etc.*

tionally been used in the detection of break-outs. No classification can be awarded to this variable since it does not fall within the category of a manipulated variable or a disturbance.

Steel flow-rate [l/min] The amount of molten steel flowing into the mould from the tundish. This variable is not measured and is considered to be an unmeasured disturbance. The position of the stopper rod (or slider gate) in the submerged entry nozzle is not measured and therefore the steel flow-rate can not be inferred.

Superheat [°C] The temperature of the in-flowing steel above the liquidus temperature. This variable is not measured and is considered to be an unmeasured disturbance.

Mould powder The mould powder is added at the top of the mould to aid in the lubrication of the strand so that it does not stick to the copper faces. This variable is unmeasured and is therefore considered to be an unmeasured disturbance.

Roller bulging [mm] As the strand moves through the secondary cooling zone, the rollers compress the delicate strand so that liquid metal within moves upward to affect the mould level. The effect is in essence difficult to detect and is considered to be a unmeasured disturbance.

Inclusion flow-out The inclusions consisting of metallic or non-metallic elements trapped at the surface of the strand due to either incorrect tundish flow-out or mould powder addition is an unmeasured disturbance.

SCZ spray patterns [l/min] The spray patterns in the secondary cooling zone (*i.e.* this variable does not form part of the mould variables). These variables are usually fixed or vary as casting speed varies so that they act as an unmeasured disturbance on the system.

3.1.2 Defect data

3.1.2.1 Defect measurement

The cost of an automatic electronic defect measurement system is extremely high, and hence they are very rarely found in practice[c]. Therefore, another method to gather defect data from cast slabs had to be found. For this a Human Measurement System (HMS) was used (see Hague and Parlington [105] for a similar idea). Three grinding plant operators with many years of experience on defects were instructed to investigate the slabs for defects during their (separate) shifts. (The operators inspect the slabs and mark defects which have to be grinded as part of the grinding process.) The idea is simple. Human operators use a schematic representation of the slabs to indicate positions where specific defects occur. They also award—based on their experience—a value of the severity of the defect (see *e.g.* Brockhoff, Hücking, Wagener, and Reichelt [154] for an index describing the severity of some defects). These values are termed as follows.

- None *i.e.* no defect occurred.
- Very slight *i.e.* the defect is very slight in the opinion of the operator.
- Slight.
- Medium *i.e.* the defect is considered to be a standard severity of the occurring defect.
- Bad.
- Very bad.

[c]see *e.g.* Mayos, Turon, Alexandre, Salon, Depeyris, and Rios [151], Knox [152] and Foster [153]

The date, slab number, grade (type), width and length are also indicated on the slab. Each slab inspection report has four slab faces depicted on it. They are for slabs that are inspected before grinding and after grinding (about 3mm off) and for the top and the bottom of the slab. A typical slab inspection report as designed by the author is shown in Fig. 3.3. At the bottom of the slab inspection report is a description of the different defects which the operator must consider together with a numbering scheme to identify each defect's severity on the specific slab face.

The example shows that an inclusion occurred 3 metres from the top of the slab on the left portion of the slab with medium (m) severity. After grinding, the defect was still present, but now only had a severity of "very slight". A longitudinal crack also formed on the bottom part of the slab at the centre location. The defect severity was bad (b). After grinding the defect was removed. For the purposes of this study, only defects on the wide faces were considered, since internal defects constitute a lesser problem and is directly proportional to the composition of the steel.

Each slab on the slab inspection report is divided (in length) into one metre intervals. This means that the average distance within which the operator would be able to indicate a defect would be 1/2 metres because the operator can indicate a defect on the separating line or on the space between two separating lines (see Fig. 3.3). Each slab is further divided into three segments along the transversal axis, *i.e.* a left side, right side and the center. This further restricts the area within which the operator can indicate the defect—these concepts are illustrated in Fig. 3.4. The figure also illustrates the naming conventions for the location of the defects. The casting direction is into the page and the top side is associated with the front or short wide face of the mould and the bottom side is associated with the back or long wide face.

Once all the slab inspection reports had been gathered, the data had to be converted into electronic format for manipulation on a personal computer.

3.1.2.2 Defuzzyfication

The slab inspection report data are read into a file for computer use. Since the slab was divided into 1/2 m segments, the defect files are said to be sampled at 1/2 metre intervals. The fuzzy levels of severity of each defect are then quantised into discrete numerical values, and this was done as follows.

SLAB INSPECTION REPORT	DATE:	2	2	0	6	1	9	9	9

SLAB No.:	3	1	7	4	0	9	3	TYPE:	3	0	4	3	1	

WIDTH:	1	2	8	5	LENGTH:	1	1	1	0	4	INSP:	John

TOP BEFORE GROUND■

1	L	C	R
2	L	C	R
3	L 2b (m)	C	R
4	L	C	R
5	L	C	R
6	L	C	R
7	L	C	R
8	L	C	R
9	L	C	R
10	L	C	R
11	L	C	R
12	L	C	R

TOP AFTER GROUND■

1	L	C	R
2	L	C	R
3	L 2b (vs)	C	R
4	L	C	R
5	L	C	R
6	L	C	R
7	L	C	R
8	L	C	R
9	L	C	R
10	L	C	R
11	L	C	R
12	L	C	R

□BOTTOM BEFORE GROUND

1	L	C	R
2	L	C	R
3	L	C	R
4	L	C 1b (b)	R
5	L	C	R
6	L	C	R
7	L	C	R
8	L	C	R
9	L	C	R
10	L	C	R
11	L	C	R
12	L	C	R

□BOTTOM AFTER GROUND

1	L	C	R
2	L	C	R
3	L	C	R
4	L	C	R
5	L	C	R
6	L	C	R
7	L	C	R
8	L	C	R
9	L	C	R
10	L	C	R
11	L	C	R
12	L	C	R

1a: transversal cracks, 1b: longitudinal cracks, 2a: casting powder entrapment, 2b other inclusions, 3: sticker, 4: bleeder, 5a: deep oscillation marks, 5b: uneven oscillation marks, 6: stopmarks, 7: footprints, 8: depressions.
vs: very slight, s: slight, m: medium, b: bad, vb: very bad

Figure 3.3 Slab inspection report (SIR).

- None: 0.
- Very slight: 1.
- Slight: 2.
- Medium: 3.
- Bad: 4.
- Very bad: 5.

There are several problems associated with this method of defect measurement. Firstly, there may be noise on the measurements because the value systems of the different operators can be different *i.e.* a very slight defect may be considered to be a slight defect by one operator and a medium defect by another. Secondly, the data are not necessarily ordinal, *i.e.* the distance between 1 and 2 is not necessarily the same as the distance between 2 and 3. This is a consequence of the value system of the different instruments (humans) in the HMS. The third problem is that the defect may not be seen by the operator and the fourth problem is that the defects are only measured several hours after they actually occurred, because the slab must cool down substantially before grinding can take place and before the operator

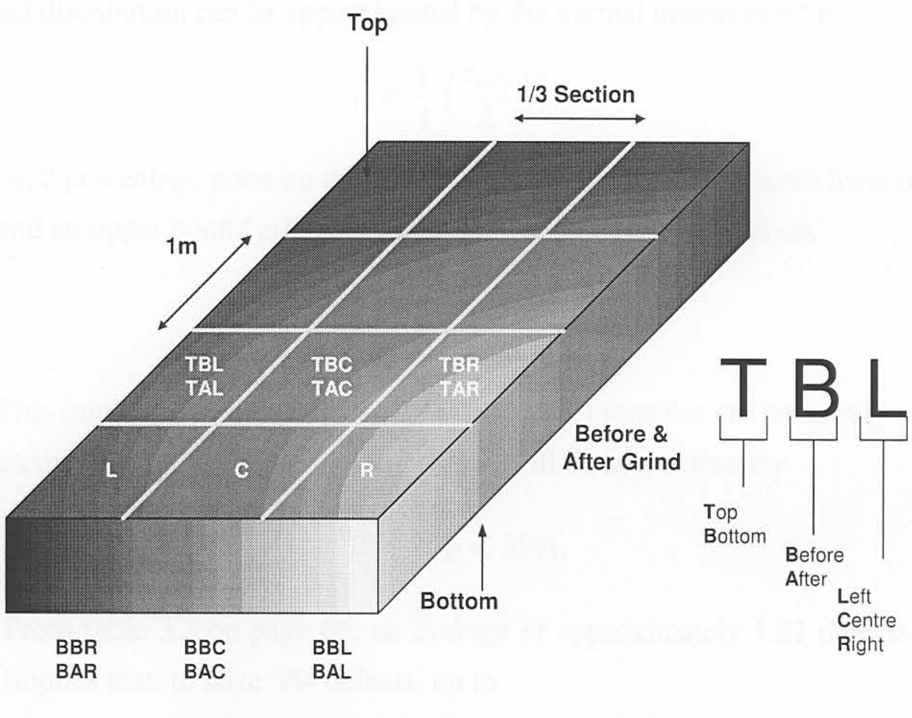

Figure 3.4 Sub-division of the slabs into top and bottom; left, right and center; and before and after grinding. The casting direction is into the page.

can inspect the slab (*i.e.* casuing a large delay). Lastly, not all slabs can be inspected. This delay and the non-inspection of many slabs make a feedback control scheme very difficult to implement.

3.1.2.3 Accuracy of the HMS

To quantify the accuracy of the HMS, several methods exist, but experiment design procedures will show that the quantification of the HMS constitutes a separate study.

One sample accuracy The first method that can be followed is to determine how accurate one person of the HMS is. This can be done statistically. Assume that n defects occur over some trial period in which one person of the HMS is instructed to investigate the slabs. This person may detect x number of the n defects. A point estimator of the population proportion (p) of positively[d] identifying a defect is then given by $\hat{p} = \frac{x}{n}$. n and p are parameters of a binomial distribution. The question is to determine how many defects have to be investigated before a "good" inference can be made on the population proportion, p. This question is simply answered in most statistical texts (see Montgomery *et al.* [144]). For a specified error E, an upper bound (conservative estimate) on the required sample size for estimating p when the binomial distribution can be approximated by the normal distribution[e] is

$$n = \frac{1}{4}\left(\frac{z_{\alpha/2}}{E}\right)^2. \tag{3.1}$$

$z_{\alpha/2}$ is the $\alpha/2$ percentage point on the normal distribution at a significance level of α. Using $\alpha = 0.05$ and an upper bound error of $E = 0.05$, the above equation gives

$$n = \frac{1}{4}\left(\frac{1.96}{0.05}\right)^2 = 384, \tag{3.2}$$

samples. This implies that if *e.g.* 296 of 384 ($p = 0.77$) samples are positively seen by the person, then one is 95% confident that the person will be approximately

$$72\% < p < 82\%, \tag{3.3}$$

accurate. From table 3.2 on page 66, an average of approximately 1.81 defects occur per slab. This implies that, to have 384 defects, up to

$$\frac{384}{1.81} \approx 213, \tag{3.4}$$

[d] assume that the same procedure can be followed for when the person sees a defect which is not truly there.

[e] $np > 5$ and $n(1-p) > 5$

slabs have to be inspected (conservative) to be 95% confident that an error of less than 5% in the accuracy estimate of the person is made.

During the validation data gathering experiment, estimates of p were determined (n=116 and x=113) to be in the region of 92 to 99% over all operators involved. These values are high because an undetected defect causes problems in post casting treatment, and operators are experienced in detecting surface defects and well aware of the importance of detecting a surface defect before grinding.

Repeatability Define one person in the HMS as one defect measurement instrument. Repeatability [155] has two possible meanings in this work. The first is the repeatability of one instrument within the HMS to measure defects consistently. This is defined as the repeatability "within" an instrument. The second is the repeatability among instruments in the HMS to measure defects consistently. This is defined as the repeatability "among" instruments.

To determine the repeatability within an instrument requires that an experiment be designed whereby an instrument measures the same set of defects a number (v) of times (see previous paragraph). Because the instrument may have memory, the defects should be randomized before each experiment run, and the instrument should not know that the defects that are being measured are the same defects as for all the experiment runs. The accuracy of that instrument in each of the v experiment runs can give an indication of the repeatability within the instrument to measure defects.

The repeatability among instruments can be measured by allowing w different instruments to measure the same defects. The accuracy of the various w instruments can then be used to determine a measure of repeatability among instruments.

To quantify the repeatability of the HMS within certain bounds is not a trivial task, since many instruments have to perform many runs on many defects ($n \times v \times w$ defects). Rather, hypothesis testing can be used in the form of a (two-way) analysis of variance (ANOVA) [156] to determine whether there is a significant difference between the runs of the experiment and whether there is a significant difference among the instruments. The following example will illustrate the concept. Assume that $w = 3$ instruments are used to measure the same randomized set of $n = 384$ defects $v = 4$ times. A table describing (example) data is given in table 3.1. The entries represent the fraction of correctly identifying defects, *e.g.* for the first instrument on the first run, a mean proportion of 81% of the defects were identified, *i.e.* 311 of the 384 defects were seen by the human (instrument). The result of the ANOVA

Table 3.1 ANOVA table for the experiment example.

	Run 1	Run 2	Run 3	Run 4
Instrument 1	.81	.83	.80	.85
Instrument 2	.90	.92	.91	.92
Instrument 3	.85	.86	.86	.86

calculation ($\alpha = 0.05$) is that, among columns (repeatability within an instrument), the F-statistic has a value of 2.98 which is less than the critical value of 4.76 implying that the instrument is repeatable. The F-statistic among rows *i.e.* among instruments is 67.4 which is higher than the critical value of 5.14 implying that the instruments vary amongst each other *i.e.* are not repeatable among each other.

The result of the example is that humans are consistent in their measurements but among the humans there is inconsistency.

The main question that has to be asked is: how many instruments (w) must perform how many runs (v) of the experiment so that the result of the test is significant. This can be calculated using the power (*e.g.* $\beta = 0.3$) of the ANOVA test together with a preliminary estimate of the overall variance of the instruments [157, 158]. Unfortunately, this was not possible, since only *one* human was available per shift to perform the inspections (one data point allows no variation and can not be used in a statistical test). The re-inspection of the slabs (runs) was also not possible since this would result in financial losses for the company, because the inspected slabs would have to be taken out of production until all inspections were done.

An initial estimate for v and w was performed for this thesis based on the assumption that $\alpha = 0.5$ and $n = 384$, by R.J. Grimbeek of STATOMET, University of Pretoria. It was assumed that a power of 70% was adequate and it was assumed that variation among the rows (instruments) and columns (runs) varied as

$$p = \bar{p} - (i-1)\epsilon^{i-1}, \tag{3.5}$$

where $\bar{p}$ is an initial estimate of the best performing candidate out of i runs or instruments. ϵ is an initial guess on the variability among or within instruments. It was assumed that $\bar{p} = 0.95$ *i.e.* the best performing instrument had a detection rate of 95% on one of the runs. The variability was chosen as $\epsilon = 0.1$. For three candidates on one run the detection rates (p) would then be 0.95, 0.93 and 0.85. Using these assumptions in the ANOVA with the desired power, delivered $v = 2$ and $w = 3$. This implies that three instruments (people) would have to inspect 384 defects twice, so that a power level of 70% can be reached. Only one person was available per shift to perform the inspections, so that an initial estimate of the variance in the system could not be calculated, and hence sample sizes could not be determined. The HMS accuracy can thus not be quantified at this stage.

The ANOVA test only gives an outcome based on whether a present defect was detected or not, and further complexities such as the severity of the defects and probabilities of detecting a defect when there is not one can also be worked into the test, resulting in larger sample sizes. These matters can be considered for future research.

For now, it is assumed that the humans are accurate in detecting a defect, due to training, knowledge and experience in the detection of defects. Any errors that occur on their part forms an uncertainty in the model, which can only be disregarded once the model validation procedure has proven that the modelling techniques work.

Precision For any particular defect, the precision [155] is one, *i.e.* a severity level in increments of one are allowable. The precision for the instrument is also difficult to quantify, since the values are measured by humans and are therefore not truly ordinal, and for the same reasons stated in the previous paragraph.

Resolution The resolution is defined [155] as the smallest change in the input (defect severity) which would cause a change in the instrument's output (severity level indicated by human). The resolution is also difficult to measure accurately due to the lack of instruments and defects as stated under the paragraph with the heading "repeatability".

3.1.2.4 Practical defect occurrence

Defects occurred at the rates depicted in Table 3.2 during the training period, and can be seen as a generalisation of the occurrence of defects at the plant. The number of defects is also known as the index of the defect and is simply a count of the specific defect. The average defects per slab is an indication of the frequency of the defect. The number of defected slabs is the number of slabs that have a specific defect present and the number of equivalent slabs is the amount of slabs that have the same grade and widths as the defected slabs. The percentage of defected slabs is the ratio of the number of defected slabs to the total number of slabs with equivalent widths and grades expressed as a percentage. From the table it is clear that other inclusions are the most prevalent defect that occurs (28.09%) with depressions second

Table 3.2 Summary of defect rate (before and after grinding) based on a sample of 502 slabs.

	Number of defects	Average defects per slab (N=502)	Number of defected slabs	Number of equivalent slabs	Percent defected slabs
Transversal cracks	4	7.968×10^{-3}	2	71	2.82
Longitudinal cracks	4	7.968×10^{-3}	3	104	2.88
Casting powder entrapment	4	7.968×10^{-3}	4	78	5.13
Other inclusions	328	653.3×10^{-3}	109	388	28.09
Bleeders	9	17.93×10^{-3}	3	151	1.99
Deep oscillation marks	43	85.66×10^{-3}	8	320	2.50
Uneven oscillation marks	26	51.79×10^{-3}	5	243	2.06
Stopmarks	243	484.1×10^{-3}	19	156	12.18
Depressions	247	492.0×10^{-3}	82	335	24.48

(24.48%). Transversal cracks, longitudinal cracks, bleeders, and both types of oscillation marks occur on the least slabs ($< 3\%$). Note that casting powder entrapment has the lowest index of defects (7.968×10^{-3}) but does not have the lowest occurrence on slabs. This means that the defect occurs only on small areas on the surface of a particular slab, but the defect occurs on many slabs.

3.1.3 Auxiliary data

The use of auxiliary data such as lengths and weights of slabs, cast numbers, MPO numbers, front and tail crop lengths *etc.* are also required. This data are obtained from the heat summaries and slab reports provided by the company.

3.1.3.1 Heat summary

The heat summaries correlate the cast numbers to the heat numbers. During each cast, several ladles may be used. Each heat number represents a ladle and naturally a specific steel grade (type). For each cast there is a fixed width. Widths cannot be changed during casting. For each cast there is a front crop and tail crop which are the slabs at the start (front) or end (tail) of the cast which are scrapped. An example of the cast number is N0008434 and the heat number is 418495.

3.1.3.2 Slab report

The slab reports contain information on the actual cast. Each heat number of the specific cast is subdivided into MPO numbers. Each MPO number represents a specific slab. Example of the MPO numbers are 3174953 and 3204950. The first slab in a specific heat is always denoted by xxxxxx1 and is never the front crop. The last slab in a heat is always denoted by xxxxxx0 and is never the tail crop. The slab report contains the actual length and weights of each slab in the specific cast. The slab report also contains the start time and end time of the cast.

3.1.4 Mould variable-defect reconciliation

This section gives a basic overview of the procedure that was followed to extract the relevant data from the mould data using the various components in §3.1.3.

3.1.4.1 Discussion

Since the defect data are given as a function of position (see §3.1.2) and the mould data are time stamped, one is required to find the relationship of position to time. To transform the defect data to time data—so that it can be used together with the time data of the mould variables—would be very difficult because there is no recorded relation between time and position on the slab inspection reports[f]. The easier method would be to re-sample the mould time data to position data.

3.1.4.2 Time-position relationship

The time and position relation is found using the cast number, sampling period[g], casting speed and numerical integration.

Cast number The cast number is found from the mould parameter file[h], and once the start of the cast has been determined from the acceleration of the strand, the relevant heat number is found from subsequent accelerations. This location is the start of the heat.

Sampling period The sampling period is two seconds.

Casting speed The casting speed is used in an integration algorithm to find the corresponding position.

[f] The defect would have to be identified at the meniscus: a very difficult task.

[g] in *seconds*

[h] Also referred to in some cases as the mould variable file.

Numerical integration The relationship between the time of sampling and the position of the slab is achieved through the use of the Galileo transform:

$$x(t) = \int_0^t v(\tau)d\tau, \tag{3.6}$$

where $x(t)$ is the position of the cast as a function of time and $v(t)$ is the casting speed as a function of time. The above equation gives the *hot* position of the strand, while the defect data are cold slab data. The actual cold slab position data—which are needed to compare with the defect data—is found by dividing x by an expansion factor[i], e. e was determined by the manufacturers of the caster to be 1.012 for ferritic stainless steels and 1.018 for austenitic stainless steel. This form of the equation is continuous and a discrete (numerical) method of integration *must* be used for our purposes. This is done through the use of a trapezoidal transformation of Eq. 3.6:

$$x_{k+1} = x_k + \frac{1}{2}(t_{k+1} - t_k)v_{k+1} + \sum_{i=0}^{k+1} \epsilon_i, \quad x_0 = 0. \tag{3.7}$$

In Eq.3.7, the position at the x+1-th instance is computed by using the k-th position and adding the trapezoid of the x+1-th instance. The error up to that stage is the sum of all errors made (ϵ). Once again, the cold position, x_{cold} is determined by dividing the hot position (x) by the expansion factor e.

3.1.4.3 Data extraction

The position versus time relation is now available. With this information—and using the heat summary together with the slab reports—one is now in a position to extract the mould data from the relevant mould parameter files. At the end of each heat there is a slowdown in casting speed which clearly defines the boundaries within the file between heats. So, if *e.g.* the data are required for the third slab in the heat (317xxx3), the lengths of the front crop and first two slabs are added to determine the start position of the third slab in the mould parameter file. All mould data from this position to the end of the slab is then extracted from the mould parameter file for further processing.

3.1.4.4 Interpolation

The position data sampling points do not correspond to the defect data sampling points. To ensure that the two sets of data correspond, linear interpolation is performed on the defect

[i] *i.e.* $x_{cold}(t) = \frac{x(t)}{e}$

data at points where position data are available from the mould data, because the mould data are sampled much more frequently than the defect data.

3.1.5 Automation of the procedure

This procedure can be automated if the slab defect data, mould data, slab reports and heat summaries are all incorporated. The implementation of such a system is described in Fig. 3.5. CASTS_MPO.txt contains the cast number vs. MPO number lookup table. The mould variable data are extracted from slabs with MPO numbers given in CASTS_MPO.txt. Starting from the first entry in the CASTS_MPO.txt file, the cast number is determined from the same row as the MPO number in the file. The relevant cast files are then opened. These are the files which contain the relevant mould variable data for the specific cast (*.mld) and the auxiliary file which contains the lengths of each of the slabs in the cast (*.aux). The file which will contain the specific slab's time and position mould variable data are also opened for writing (*.mpf). The next step is to determine the starting and ending lengths of the specific slab in question. This is done by adding the lengths of the slabs up to the beginning of the slab in question and then making this the "start" length. The length of the specific slab in question is then added to this total to give the end length. The time vector in the *.mld file is then integrated until the length obtained is the same as the start length determined earlier. Mould data are then logged to the *.mpf file until the end length is reached. This means that all the data logged in the *.mpf file pertains to that specific slab. Once the end length has been reached, the *.mpf file is closed and the next slab data are created until all slabs in the CASTS_MPO.txt file have been processed. Once this has been done, the position dependent data in the defect files can be interpolated to correspond to the position dependent data in the mould variable file for the specific slab.

3.1.6 General model structure

This section described the data that is involved when trying to train a model to predict defects in the continuous casting process. Many input variables are measured and these have been explained in this section. It has been seen that the treatment of input (mould variable data) involves the utilisation of much information. The extraction of relevant data from the plant data is a tedious task which requires considerable data processing.

Slab inspection reports were used by the operators to log defects that may occur. The specific

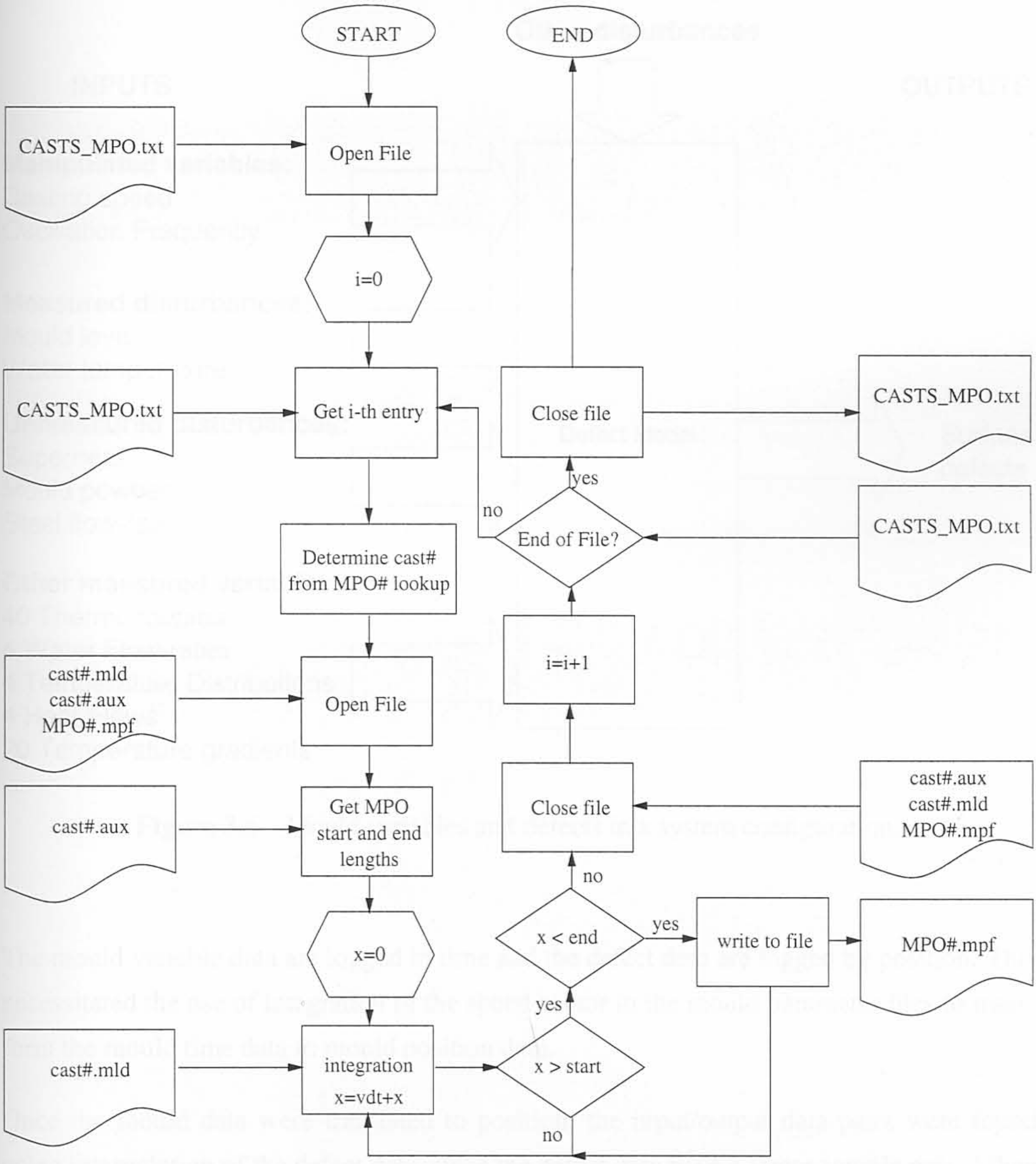

Figure 3.5 Flow chart describing the procedure to extract the mould data from the relevant files.

defect, location, position and severity of the defect was logged on the slab inspection reports. These slab inspection reports were then digitised.

The general input / output model structure is given in Fig. 3.6.

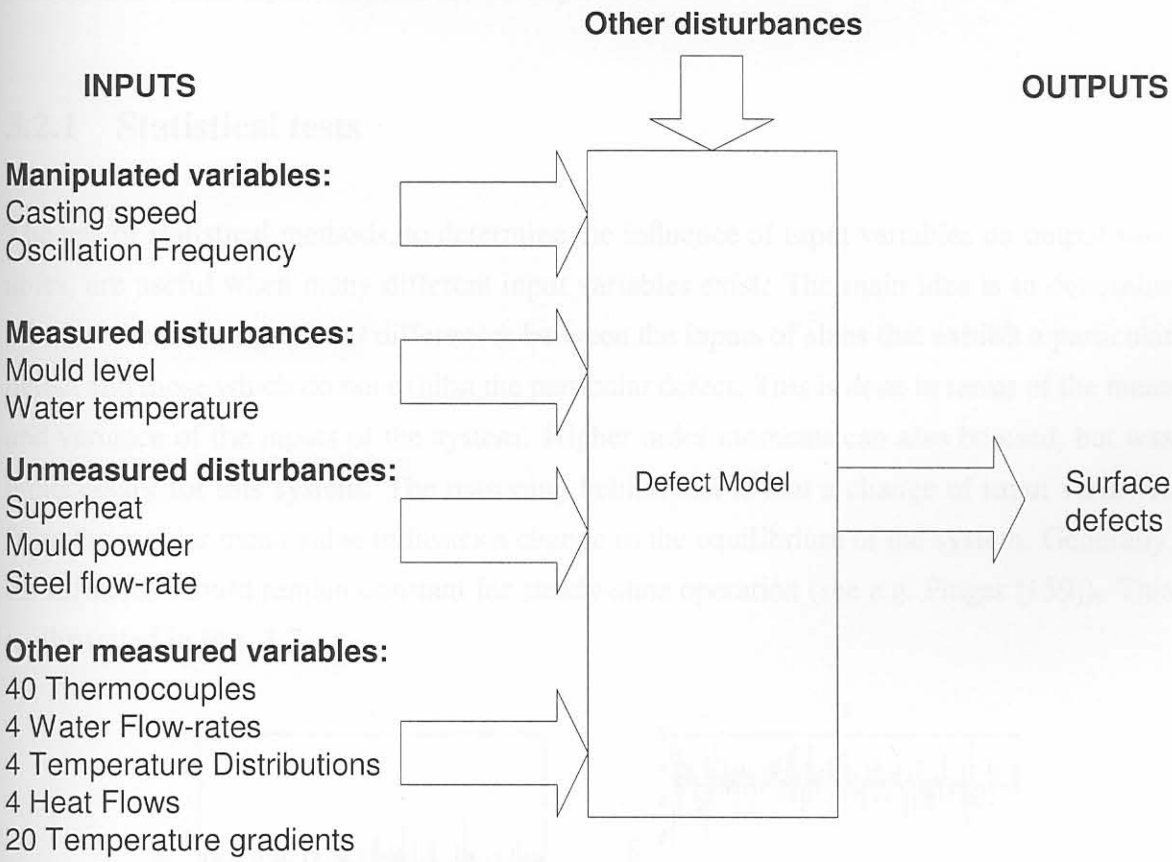

Figure 3.6 Mould variables and defects in a system configuration.

The mould variable data are logged in time and the defect data are logged by position. This necessitated the use of integration of the speed vector in the mould parameter files to transform the mould time data to mould position data.

Once the mould data were translated to position, the input/output data pairs were found using interpolation of the defect data, since the defect data have a larger sample period than the mould data.

With this new data in hand, the variables can be investigated and models can be trained to predict the occurrence of defects.

3.2 Statistical analysis

This section describes the use of statistical analysis to determine which inputs affect the defects and which input variables are closely correlated.

3.2.1 Statistical tests

The use of statistical methods, to determine the influence of input variables on output variables, are useful when many different input variables exist. The main idea is to determine whether there are *significant* differences between the inputs of slabs that exhibit a particular defect and those which do not exhibit the particular defect. This is done in terms of the mean and variance of the inputs of the system. Higher order moments can also be used, but was unnecessary for this system. The reasoning behind this is that a change of input variables from the regular mean value indicates a change in the equilibrium of the system. Generally, all variables should remain constant for steady-state operation (see *e.g.* Pinger [159]). This is illustrated in Fig. 3.7.

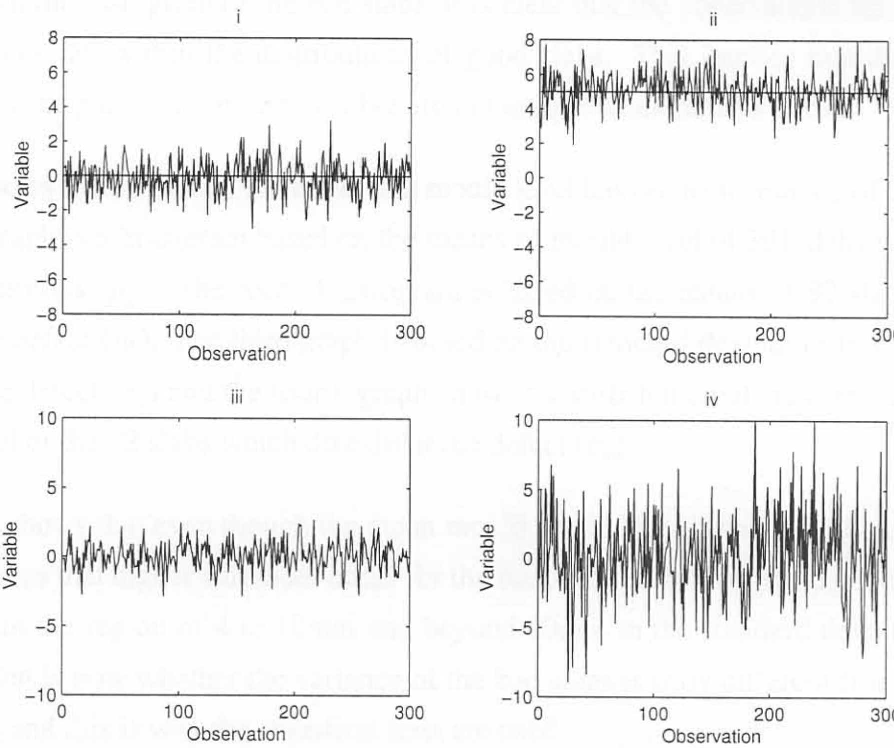

Figure 3.7 Depiction of the effect of different variances and means of a specific variable. i and ii have similar variances but different means while iii and iv have similar means but different variances.

Graph i has a mean value for the variable on the specific slab which is lower than the mean of the variable on a different slab (graph ii). The variances are similar in both cases. Graph iii has a variance lower than that of graph iv, indicating that there was more variation in the variable on the slab in graph iv. The means remain approximately the same for both cases. This reasoning can be extended to the case where both the variance and the mean are different for a variable on specific defected slabs, indicating that there is a significant change in both the location and scatter of the specific variable in question.

By examining the means of a specific input variable of all the slabs which do not have a particular defect contained on the slab (good slabs), and comparing these distributions of means with the distributions of means on slabs which do contain the specific defect (bad slabs), some conclusions can be drawn about the effect of a specific variable on the defect. The same reasoning follows for the variances/standard deviations. This is useful to indicate which variables change when defects occur.

Fig 3.8 shows the concept of means and variances of good slabs versus means and variances of bad slabs for any particular input variable. In this case, many good slabs are used to form the theoretical distribution of a particular variable. Few data exist for the bad slabs (3 in this case). From the histogram of the bad slabs, it is clear that the observations for both means and variances fall within the distributions of good slabs. This implies that the mean and variance for the particular mould variable are not unique when defects occur.

Fig. 3.9 shows histograms of the effect that mould level has on the formation of depressions. The first graph is a histogram based on the means of mould level of 391 slabs which do not have depressions (μ_g). The second histogram is based on the means of 82 slabs which do exhibit the defect (μ_b). The third graph is based on the standard deviations of the 391 slabs without the defect (σ_g) and the fourth graph shows the distribution of standard deviations in mould level of the 82 slabs which do exhibit the defect (σ_b).

The figure shows that even though the mean mould level is the same for both good and bad slabs, it seems that higher variances occur for the bad slabs compared to the good slabs. This is evident in the region of 4 to 10mm and beyond 20mm in the standard deviation graphs. The question is now whether the variance of the bad slabs is truly different from that of the good slabs and this is why the statistical tests are used.

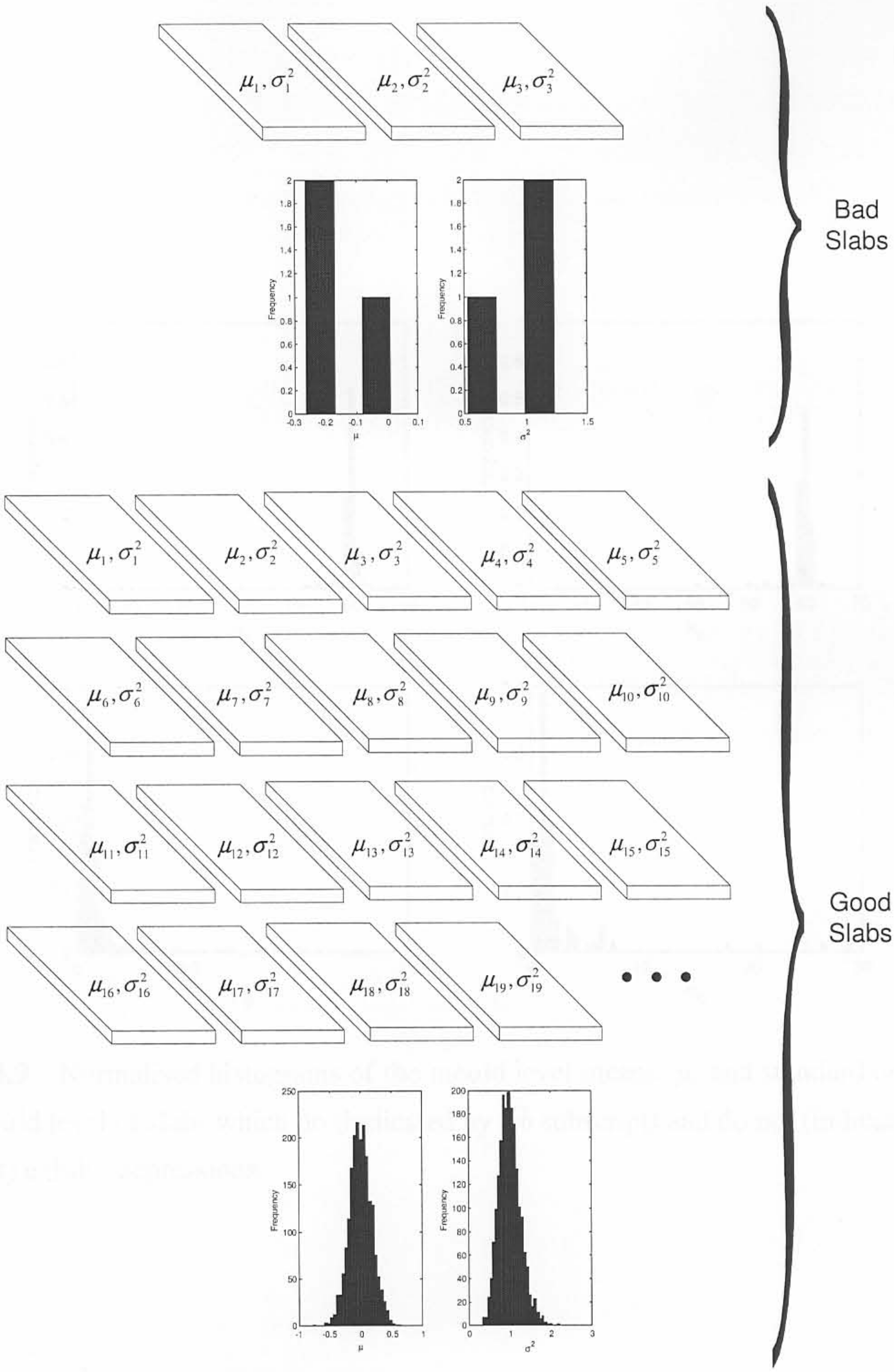

Figure 3.8 Depiction of the distributions of means and variances of bad slabs versus distributions of means and variances of good slabs for a specific input variable.

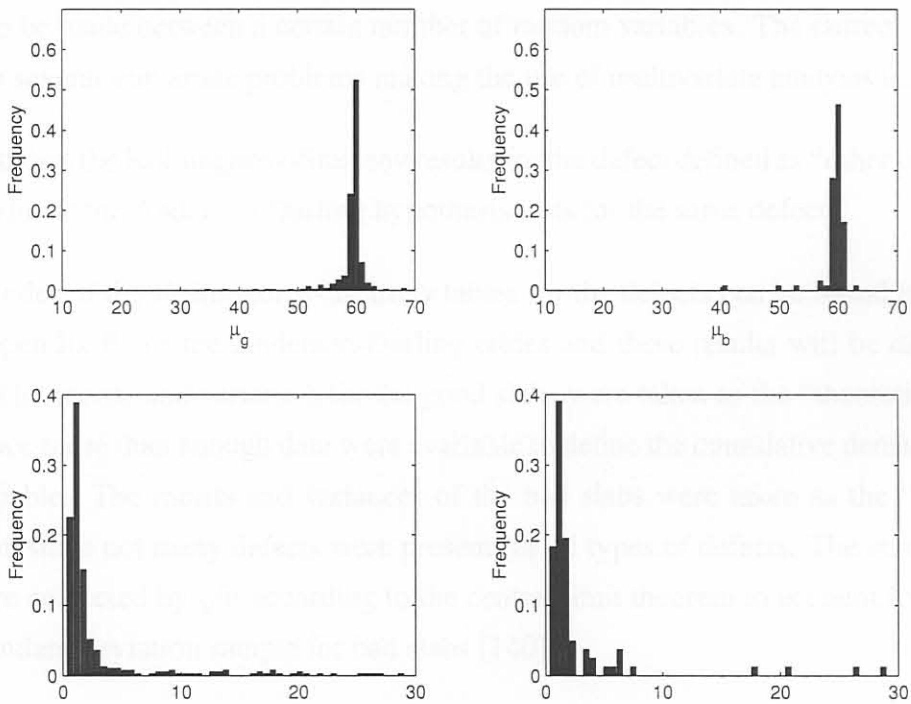

Figure 3.9 Normalised histograms of the mould level means, μ, and standard deviations, σ, of mould level of slabs which do (indicated by a b subscript) and do not (indicated by a g subscript) exhibit depressions.

3.2.1.1 Results

Kolmogorov-Smirnov and Anderson-Darling tests were performed for each defect and the corresponding input variables. These are univariate methods to determine whether a random variable belongs to a given distribution. These methods are preferred (for this investigation) over multivariate methods such as principal component analysis [158] because it is cumbersome to formulate the problem as a multivariate problem, and it appears that not much would be gained as the results of section 3.2.1.2 shows. In a multivariate problem, comparisons are made between a number of random variables. For the current problem, there are two comparisons to be made between a certain number of random variables. The current problem is reduced to several univariate problems making the use of multivariate analysis unnecessary.

Table 3.3 shows the Kolmogorov-Smirnov results for the defect defined as "other inclusions". Table 3.4 shows the Anderson-Darling hypothesis tests for the same defect.

The remainder of the Kolmogorov-Smirnov tables for the defects can be found in appendix A; and appendix B for the Anderson-Darling tables and these results will be discussed in §3.2.1.3. The means and variances for the good slabs were taken as the "theoretical" distribution, since more than enough data were available to define the cumulative density function of the variable. The means and variances of the bad slabs were taken as the "empirical" distribution since not many defects were present for all types of defects. The standard deviations were corrected by $\sqrt{n}$ according to the central limit theorem to account for the small size of standard deviation sample for bad slabs [140].

The $n_{2b} = 114$ in the caption of the table indicates the number of slabs with the specific defect that was used for the test; *i.e.* 114 means and variances were used in two respective Kolmogorov-Smirnov tests and 114 means and variances were used in two respective Anderson-Darling tests (this is a large sample); for each input variable. $n_g = 364$ indicates that 364 good slabs of the same dimensions and steel grades as the bad slabs that do not exhibit the specific defect were used as to generate the theoretical distribution.

$d_c(114) = 0.127$ is the critical value for the test. If the D statistic is less than $d_c(114) = 0.127$, the null hypothesis is accepted and it is concluded that *e.g.* the means are the same for good and bad slabs. If the statistic is rejected (*i.e.* D statistic is higher than the critical value), it means that the conclusion is that *e.g.* the means of good and bad slabs are different and that the input variable is a likely cause of the specific defect.

$1 - \alpha$ is the confidence level for the test, and is usually selected at 95% [140]. $\Delta_{0.95}$ is

Table 3.3 Kolmogorov-Smirnov hypothesis tests for Other Inclusions (defect 2b). n_{2b}=114. n_g =364. $d_c(114)$=0.127. α=0.05. $\Delta_{0.95}$=0.206. $\Delta_{0.5}$=0.122.

Variable	$D_{\mu_{2b}}$	$H_0(\mu_{2b})$	$D_{\sigma^2_{2b}}$	$H_0(\sigma^2_{2b})$	$H_0(\mu_{2b}) \cup H_0(\sigma^2_{2b})$
in1u	0.175	Reject	0.067	Accept	Reject
in1l	0.0977	Accept	0.0697	Accept	Accept
in2u	0.198	Reject	0.127	Accept	Reject
in2l	0.141	Reject	0.14	Reject	Reject
in3u	0.212	Reject	0.153	Reject	Reject
in3l	0.0942	Accept	0.102	Accept	Accept
in4u	0.171	Reject	0.173	Reject	Reject
in4l	0.0769	Accept	0.162	Reject	Reject
in5u	0.0833	Accept	0.207	Reject	Reject
in5l	0.0605	Accept	0.168	Reject	Reject
in6u	0.095	Accept	0.142	Reject	Reject
in6l	0.0967	Accept	0.168	Reject	Reject
in7u	0.215	Reject	0.157	Reject	Reject
in7l	0.143	Reject	0.178	Reject	Reject
in8u	0.145	Reject	0.0473	Accept	Reject
in8l	0.122	Accept	0.0817	Accept	Accept
nl1u	0.261	Reject	0.17	Reject	Reject
nl1l	0.26	Reject	0.0738	Accept	Reject
nl2u	0.31	Reject	0.0743	Accept	Reject
nl2l	0.249	Reject	0.166	Reject	Reject
ou1u	0.199	Reject	0.136	Reject	Reject
ou1l	0.155	Reject	0.0818	Accept	Reject
ou2u	0.129	Reject	0.122	Accept	Reject
ou2l	0.169	Reject	0.0956	Accept	Reject
ou3u	0.228	Reject	0.109	Accept	Reject
ou3l	0.191	Reject	0.149	Reject	Reject
ou4u	0.154	Reject	0.164	Reject	Reject
ou4l	0.221	Reject	0.146	Reject	Reject
ou5u	0.223	Reject	0.133	Reject	Reject
ou5l	0.2	Reject	0.169	Reject	Reject
ou6u	0.36	Reject	0.253	Reject	Reject
ou6l	0.233	Reject	0.214	Reject	Reject
ou7u	0.217	Reject	0.136	Reject	Reject
ou7l	0.204	Reject	0.1	Accept	Reject
ou8u	0.267	Reject	0.126	Accept	Reject
ou8l	0.154	Reject	0.0804	Accept	Reject
nr1u	0.199	Reject	0.119	Accept	Reject
nr1l	0.0899	Accept	0.155	Reject	Reject
nr2u	0.0831	Accept	0.113	Accept	Accept
nr2l	0.133	Reject	0.102	Accept	Reject
Casting Speed	0.193	Reject	0.105	Accept	Reject
Mould Controller Status	0.0605	Accept	0.0736	Accept	Accept
Mould level	0.127	Reject	0.0717	Accept	Reject
Inlet Temperature	0.136	Reject	0.174	Reject	Reject
Flowrate WL	0.0596	Accept	0.0797	Accept	Accept
Flowrate WF	0.117	Accept	0.2	Reject	Reject
Flowrate NL	0.1	Accept	0.0888	Accept	Accept
Flowrate NR	0.0835	Accept	0.0602	Accept	Accept
Delta T WL	0.155	Reject	0.171	Reject	Reject
Delta T WF	0.257	Reject	0.151	Reject	Reject
Delta T NL	0.1	Accept	0.133	Reject	Reject
Delta T NR	0.107	Accept	0.131	Reject	Reject
Oscillation Frequency	0.258	Reject	0.139	Reject	Reject
Drive Current	0.117	Accept	0.111	Accept	Accept
Heat Flux WL	0.117	Accept	0.177	Reject	Reject
Heat Flux WF	0.185	Reject	0.186	Reject	Reject
Heat Flux NL	0.1	Accept	0.128	Reject	Reject
Heat Flux NR	0.107	Accept	0.145	Reject	Reject
in1	0.188	Reject	0.0839	Accept	Reject
in2	0.174	Reject	0.0935	Accept	Reject
in3	0.188	Reject	0.119	Accept	Reject
in4	0.178	Reject	0.13	Reject	Reject
in5	0.181	Reject	0.179	Reject	Reject
in6	0.186	Reject	0.158	Reject	Reject
in7	0.129	Reject	0.119	Accept	Reject
in8	0.0887	Accept	0.0931	Accept	Accept
nl1	0.213	Reject	0.0979	Accept	Reject
nl2	0.211	Reject	0.0491	Accept	Reject
ou1	0.108	Accept	0.0851	Accept	Accept
ou2	0.0891	Accept	0.0861	Accept	Accept
ou3	0.1	Accept	0.0865	Accept	Accept
ou4	0.139	Reject	0.162	Reject	Reject
ou5	0.099	Accept	0.179	Reject	Reject
ou6	0.244	Reject	0.232	Reject	Reject
ou7	0.122	Accept	0.115	Accept	Accept
ou8	0.134	Reject	0.0773	Accept	Reject
nr1	0.265	Reject	0.118	Accept	Reject
nr2	0.127	Accept	0.114	Accept	Accept

Table 3.4 Anderson-Darling hypothesis tests for Other Inclusions (defect 2b). n_{2b}=114. n_g =364. $a_c^2(114)$=2.492. α=0.05.

Variable	$A^2_{\mu_{2b}}$	$H_0(\mu_{2b})$	$A^2_{\sigma^2_{2b}}$	$H_0(\sigma^2_{2b})$	$H_0(\mu_{2b}) \cup H_0(\sigma^2_{2b})$
in1u	4.28	Reject	0.629	Accept	Reject
in1l	1.95	Accept	0.791	Accept	Accept
in2u	∞	Reject	∞	Reject	Reject
in2l	∞	Reject	2.67	Reject	Reject
in3u	9.62	Reject	∞	Reject	Reject
in3l	∞	Reject	∞	Reject	Reject
in4u	∞	Reject	6.33	Reject	Reject
in4l	∞	Reject	5.47	Reject	Reject
in5u	2.34	Accept	7.86	Reject	Reject
in5l	0.652	Accept	6.76	Reject	Reject
in6u	1.67	Accept	4.63	Reject	Reject
in6l	∞	Reject	6.72	Reject	Reject
in7u	7.94	Reject	∞	Reject	Reject
in7l	5.38	Reject	∞	Reject	Reject
in8u	3.9	Reject	0.306	Accept	Reject
in8l	3.6	Reject	0.995	Accept	Reject
nl1u	15.7	Reject	5.06	Reject	Reject
nl1l	15.5	Reject	1.59	Accept	Reject
nl2u	22.8	Reject	1.26	Accept	Reject
nl2l	13.7	Reject	3.94	Reject	Reject
ou1u	5.99	Reject	3.39	Reject	Reject
ou1l	4.82	Reject	1.35	Accept	Reject
ou2u	∞	Reject	2.9	Reject	Reject
ou2l	∞	Reject	1.07	Accept	Reject
ou3u	∞	Reject	1.98	Accept	Reject
ou3l	∞	Reject	3.14	Reject	Reject
ou4u	∞	Reject	6.66	Reject	Reject
ou4l	∞	Reject	6.28	Reject	Reject
ou5u	∞	Reject	2.07	Accept	Reject
ou5l	11.3	Reject	6.9	Reject	Reject
ou6u	32.7	Reject	14	Reject	Reject
ou6l	21.3	Reject	15.6	Reject	Reject
ou7u	∞	Reject	∞	Reject	Reject
ou7l	∞	Reject	∞	Reject	Reject
ou8u	14.5	Reject	2.39	Accept	Reject
ou8l	6.94	Reject	0.757	Accept	Reject
nr1u	8.74	Reject	2.84	Reject	Reject
nr1l	1.8	Accept	∞	Reject	Reject
nr2u	0.772	Accept	∞	Reject	Reject
nr2l	2.32	Accept	∞	Reject	Reject
Casting Speed	∞	Reject	2.09	Accept	Reject
Mould Controller Status	1.62e+003	Reject	90.1	Reject	Reject
Mould level	2.68	Reject	1.09	Accept	Reject
Inlet Temperature	1.8	Accept	3.75	Reject	Reject
Flowrate WL	0.424	Accept	0.988	Accept	Accept
Flowrate WF	2.32	Accept	6.62	Reject	Reject
Flowrate NL	2.03	Accept	1.55	Accept	Accept
Flowrate NR	0.945	Accept	0.406	Accept	Accept
Delta T WL	4.76	Reject	4.65	Reject	Reject
Delta T WF	12.8	Reject	2.93	Reject	Reject
Delta T NL	2.18	Accept	2.77	Reject	Reject
Delta T NR	1.67	Accept	2.27	Accept	Accept
Oscillation Frequency	∞	Reject	19.3	Reject	Reject
Drive Current	1.52	Accept	∞	Reject	Reject
Heat Flux WL	4.82	Reject	6.2	Reject	Reject
Heat Flux WF	5.7	Reject	4.13	Reject	Reject
Heat Flux NL	2.2	Accept	2.79	Reject	Reject
Heat Flux NR	1.66	Accept	2.24	Accept	Accept
in1	∞	Reject	0.69	Accept	Reject
in2	6.44	Reject	1.99	Accept	Reject
in3	∞	Reject	2.06	Accept	Reject
in4	7.21	Reject	∞	Reject	Reject
in5	5.29	Reject	6.72	Reject	Reject
in6	18.8	Reject	5.16	Reject	Reject
in7	3.2	Reject	3.61	Reject	Reject
in8	2.01	Accept	0.855	Accept	Accept
nl1	∞	Reject	1.47	Accept	Reject
nl2	∞	Reject	0.324	Accept	Reject
ou1	1.53	Accept	1.07	Accept	Accept
ou2	2.2	Accept	1.17	Accept	Accept
ou3	3.81	Reject	1.24	Accept	Reject
ou4	9.08	Reject	4.62	Reject	Reject
ou5	5.86	Reject	∞	Reject	Reject
ou6	12.8	Reject	11.7	Reject	Reject
ou7	2.24	Accept	2.72	Reject	Reject
ou8	4.09	Reject	1.12	Accept	Reject
nr1	16.2	Reject	3.23	Reject	Reject
nr2	3.24	Reject	∞	Reject	Reject

the minimum difference between the theoretical and empirical distributions which will be detected with a power of 95%. Similarly, $\Delta_{0.5}$ is the minimum difference between the theoretical and empirical distributions which will be detected with a power of 50%.

$D_{\mu_{2b}}$ is the Kolmogorov-Smirnov test statistic value for bad slabs that exhibit defect 2b ("other inclusions") for a specific input variable. $H_0(D_{\mu_{2b}})$ is the outcome of the test. "Accept" means that the mean of a specific input variable for good and bad slabs is not different. "Reject" means that the means are different.

$D_{\sigma^2_{2b}}$ is the test statistic for the variances of bad slabs, and $H_0(D_{\sigma^2_{2b}})$ is the outcome of the test: "accept" means that the variances of a specific input variable is similar for both good and bad slabs and "reject" means that the variances are not the same.

$H_0(\boldsymbol{\mu}_{2b}) \cup H_0(\boldsymbol{\sigma}^2_{2b})$ is the final result outcome and means that if either the means test or the variances test is rejected, the input quite possibly has an influence on the occurrence of the defect in question. Both tests must be passed before it is decided that the variable has no influence on the defects.

The variables on which tests were performed are in column 1 of tables 3.3 and 3.4.

The Anderson-Darling tests (see table 3.4) have similar hypothesis conclusions as the Kolmogorov-Smirnov tests. The table shows the outcomes for the "other inclusions" defect. $a^2_c(114) = 2.492$ is the critical value for the test with 114 slabs. Note that all mean data are near-normal and all variance data are near χ^2 (skewed normal) so that the critical value of normal distributions of 2.492 reported by Johnson [140] was used for the Anderson-Darling tests, since precise critical values for the distributions are unknown. $A^2_{\mu_{2b}}$ is the test statistic for means and $A^2_{\sigma^2_{2b}}$ is the test statistic for variances. $H_0(\boldsymbol{\mu}_{2b})$ and $H_0(\boldsymbol{\sigma}^2_{2b})$ are the outcomes of the means and variances tests respectively. $H_0(\boldsymbol{\mu}_{2b}) \cup H_0(\boldsymbol{\sigma}^2_{2b})$ is the final outcome of the tests.

3.2.1.2 Interpretation

The Kolmogorov-Smirnov result (table 3.3) shows that most of the thermocouples[j] had an hypothesis rejection on both the means and variances, *i.e.* different from good slabs; implying that the thermocouple temperatures indicate that a (other inclusions) defect is likely to occur. The only outliers in this group are "in1l", "in3l", "in8l" and "nr2u", probably

[j]Thermocouples "ou6l" & "ou6l" are deemed inappropriate because these thermocouples were non-operational for most of the study.

due to bad contact within the mould[k], or simply that the defect never passed close to these thermocouples.

In some instances one of the tests, *e.g.* mean test, is accepted and the other, *e.g.* variance test, not (see "nr1l"). This implies that the scatter of the variable is larger than usual, but it still has a mean similar to that of good slabs. In some instances the variance test was accepted and the mean test not (*e.g.* "in1u"). This means that the scatter is similar to good slab cases but that the mean is different from the good slabs.

Other variables of note that are rejected in the Kolmogorov-Smirnov tests are

- the mould level, with a different mean value from that of good slabs,
- the casting speed, with a mean value different from that of the good slabs,
- water inlet temperature, with a rejection on both the mean and variance,
- wide front face water flow-rate,
- temperature differences,
- all heat fluxes with rejections on the variance tests, and
- most of the longitudinal temperature differences.

The Anderson-Darling tests have similar results for "other inclusions" as the Kolmogorov-Smirnov tests. The test, however, rejects all thermocouple temperatures with the exception of "in1l". The test also rejects variables similar to that of the Kolmogorov-Smirnov test such as casting speed, water inlet temperature, wide front face water flow-rate, heat fluxes and longitudinal temperature differences. In addition, mould level controller status is also rejected. Heat flux on narrow right side is accepted.

The results compare well with published results in the case of mould level, mould oscillation and casting speed (see table 2.1). These three variables were also considered in the published literature to be instrumental in inclusion formation [69]. The literature did however not state that thermocouple temperature or inlet temperature was an influential factor, but the statistical results show otherwise. The test that is probably more accurate is the Kolmogorov-Smirnov test, since it is equipped to deal with such a large number of values from the empirical distribution.

[k]a very common problem

Note that ∞ means that the test statistic was very high, *i.e.* fell completely outside the theoretical distribution.

3.2.1.3 Other defects

The statistical tests for other defects can be found in appendices A & B.

Transversal cracking The Kolmogorov-Smirnov and Anderson-Darling tests have a higher null hypothesis acceptance rate for transversal cracks variables than *e.g.* "other inclusions". This is because less data (only one slab was affected) were available for the tests and the possibility of making a type 1 error was higher (weak conclusions). Four thermocouples, oscillation frequency[1], and some of the longitudinal temperature differences were rejected in the Kolmogorov-Smirnov tests (see table A.1). The Anderson-Darling tests for transversal cracking (see table B.1) rejected seven thermocouples (including the four of the Kolmogorov-Smirnov tests), mould level, casting speed, oscillation frequency, drive current, and many of the longitudinal temperature gradients. From the literature review, transversal cracking is also affected by temperature variation in the mould [71]. This is underlined by the fact that the "in8u" and "in8l" thermocouples were rejected by the Kolmogorov-Smirnov and Anderson-Darling tests respectively. These thermocouples are situated near one of the transversal cracks occurring on the top left side after grinding has taken place. Casting speed and oscillation frequency are also noted from the literature study to be instrumental in the formation of the defect [77], which concurs with the result obtained from the statistical tests. In the defect literature, mould level is not generally noted as an influencing force (contrary to the mould level control literature), but was detected as influential by the statistical tests. Note that the Anderson-Darling test is probably more accurate than the Kolmogorov-Smirnov test because of the low number of samples from the empirical (bad) distribution.

Longitudinal cracking Three slabs contained longitudinal cracks. The Kolmogorov-Smirnov tests (see table A.2) rejected only one thermocouple variable. Other variables that were rejected were the inlet temperature and drive current. The Anderson-Darling tests (table B.2) rejected 6 thermocouple temperatures, mould controller status, drive current and several of

[1]Note that it will be shown later that casting speed and oscillation frequency are linearly dependent. This should have resulted in a rejection for casting speed on the Kolmogorov-Smirnov test. Noting from table A.1 the mean statistic for casting speed (0.971) is extremely close to the critical value (0.975), implying that the weak conclusion could be changed to a rejection of the null hypothesis.

the longitudinal temperature gradients. Mould level is considered based on the literature study to be an influencing factor in the formation of longitudinal cracks. However, the statistical tests show that mould level is not an influencing factor. Saito *et al.* [76] state that mould level has an influence on the formation of longitudinal *corner* cracks. As will be seen later, table 3.8 shows that no longitudinal corner cracks occurred and that is probably why the statistical tests show otherwise. In the Anderson-Darling test mould controller status is rejected, indicating that there is more manual control of mould level during the formation of the defect. Mould oscillation is also excluded as an influencing factor by the statistical tests but not by the literature [87]. This is probably because casting speed and mould oscillation frequency are linearly dependent (as will be seen from cross-correlation tests in §3.2.2) at the plant in question. If casting speed is shown to be an influencing factor, then oscillation frequency should also be an influencing factor. Casting speed is also not rejected by the statistical tests, but is considered based on the literature to be influential in the formation of the defect [88]. In the literature overview (page 21) it is said that casting speed variation or high casting speed is a factor in the formation of the defect. This is however not the case with the three defected slabs in question (graphs of casting speed are not shown).

Casting powder entrapment 7 thermocouple temperatures, wide front face flow-rate and 4 longitudinal thermocouple gradients were rejected for this defect with the Kolmogorov-Smirnov test (table A.3). 16 thermocouple temperatures, casting speed, mould controller status, oscillation frequency, and 12 longitudinal temperature differences were rejected with the Anderson-Darling tests (table B.3). This correlates well with the published results in table 2.1, except for mould level [69].

Bleeders 2 thermocouple temperatures and one thermocouple gradient were rejected by the Kolmogorov-Smirnov tests for the "bleeder" defect. See tables A.5 & B.5. Note that the result for "ou6l" can not be considered because the thermocouple was broken for most of the time period of data gathering. The thermocouple "in8u" is in the vicinity of the occurring defects at the top left location on the slab. No thermocouples were instrumental in the detection of the defects, based on the Anderson-Darling test. This is probably because the sample size is so small, and the Kolmogorov-Smirnov test should therefore be more accurate. The literature overview of table 2.1 showed that temperature was influential in the occurrence of the defect, similar to the result of the statistical tests [99].

Deep oscillation marks For deep oscillation marks, 5 thermocouple temperatures, 2 water flow-rates, 3 temperature differences and all heat fluxes were rejected. 6 longitudinal temperature differences were rejected. The above were all rejected by the Kolmogorov-Smirnov tests (table A.6). 6 thermocouple temperatures, mould controller status, 2 flow-rates, 3 temperature differences, and all the heat fluxes together with some longitudinal temperature differences were rejected by the Anderson-Darling tests (table B.6). The literature shows that mould oscillation [103] and casting speed [91, 160] also have influences, but the statistical tests did not consider this to vary much compared to good slabs.

Uneven oscillation marks These defects saw the rejection of 17 thermocouple temperatures, 3 flow-rates and 10 longitudinal temperature differences (table A.7). The Anderson-Darling test rejected 15 thermocouple temperatures, mould controller status, inlet temperature, 2 flow-rates and 10 longitudinal temperature differences (table B.7). The literature shows that mould oscillation and casting speed also have influences [91, 160], but the statistical tests did not consider this to vary much compared to good slabs.

Stopmarks All thermocouple temperatures were rejected for stopmarks by the Kolmogorov-Smirnov test. Casting speed, mould controller status and mould level were also rejected. All flow rates, heat fluxes and temperature differences were rejected. All longitudinal temperature differences were rejected. Similar results hold for the Anderson-Darling tests (table B.8). Only inlet temperature was not rejected and is thus supposed not to be influential in the formation of the defect. The huge rejection ratio implies that much goes wrong when this defect occurs, which is logical since it usually coincides with an abrupt slow-down of the strand and could probably be predicted by only considering the casting speed.

Depressions 25 thermocouple temperatures, casting speed, inlet temperature, temperature differences, oscillation frequency, heat fluxes and some of the longitudinal temperature differences were rejected for depressions using the Kolmogorov-Smirnov tests (table A.9). Similarly, all thermocouple temperatures except one were rejected using the Anderson-Darling tests (table B.9). Casting speed, mould controller status, inlet temperature, temperature differences, oscillation frequency, and 3 heat fluxes were rejected as well as nearly all the longitudinal temperature differences. Temperatures variations as a cause of depressions agrees with the literature review but all other variables are supposed to be non-influential [110, 111]. The high rejection ratio is probably due to the large sample size. Interestingly enough, mould level was accepted and is thus supposed to be uninfluential.

3.2.1.4 Summarised result

Table 3.5 shows a summary of the statistical results for both the Kolmogorov-Smirnov and Anderson-Darling tests. MCS is the mould controller status, OF the oscillation frequency, and LTD the linear temperature differences.

From the table it is evident that thermocouple temperatures play an important role in the detection or control of the respective defects, since their means and variances are considerably different for bad slabs compared to good slabs. All nine defects continually show that there are differences between good and bad slabs. Casting speed seems to differ from the good slabs only in three cases using the Kolmogorov-Smirnov test and in 5 cases using the Anderson-Darling test. The mould controller status, which describes whether a mould controller was used during operation, is influential only for bleeders using the Kolmogorov-Smirnov test; but is influential in all defects, except transversal cracks, using the Anderson-Darling test. Inlet temperature is influential in three defects, and the flow rates have an influence in 5 defects using the Kolmogorov-Smirnov test and 4 defects using the Anderson-Darling tests. The temperature differences are influential in 4 defects, and the oscillation frequency varies between 4 and 5 defects for the respective tests. Drive current affects three and four defects respectively, and the heat fluxes influence 4 of the defects. The longitudinal temperature differences affect all defects except longitudinal cracks as tested by the Kolmogorov-Smirnov test.

The results indicate that thermocouples are heavily influential in the detection or control of the defects, with some other variables such as casting speed, oscillation frequency, flow rates, temperature differences, drive currents and heat fluxes not far behind.

3.2.2 Correlation

This section describes the use of correlation to determine whether there is linear dependence between variables so that some variables can be dropped from possible modelling strategies.

Tables C.1, C.2, C.3, C.4 and C.5 in appendix C show correlation variables of the time-series inputs split into columns over five tables.

Note that multivariate techniques such as principal component analysis [158] on the correlation table can be used to determine influential variables. These methods were however not considered because enough information is contained in the table itself to make accurate

Table 3.5 Summarised results for Kolmogorov-Smirnov and Anderson-Darling tests. ○ indicates that the variable is a cause of the defect based on the literature survey and ● indicates that the variable is influential based on the goodness-of-fit tests. T is the total number of defects influenced by a variable based on the goodness-of-fit tests.

Variable	Kolmogorov-Smirnov									
	1a	1b	2a	2b	4	5a	5b	6	8	T
Thermocouple	●○	●○	●	●	●○	●○	●○	●	●○	9
Casting Speed	?○	○	○	●○		○	○	●○	●	4
MCS								●		1
Mould Level		○	○	●○				●		2
Inlet Temperature		●		●					●	3
Flow Rate			●	●		●	●	●		5
Delta T				●		●		●	●	4
OF	●○	○	○	●○		○	○	●	●	4
Drive Current	●	●						●		3
Heat Flux				●		●		●	●	4
LTD	●		●	●	●	●	●	●	●	8
	Anderson-Darling									
Thermocouple	●○	●○	●	●	○	●○	●○	●	●○	8
Casting Speed	●○	○	●○	●○		○	○	●○	●	5
MCS		●	●	●	●	●	●	●	●	8
Mould Level	●	○	○	●○				●		3
Inlet Temperature				●			●		●	3
Flow Rate				●		●	●	●		4
Delta T				●		●		●	●	4
OF	●○	○	●○	●○		○	○	●	●	5
Drive Current	●	●		●				●		4
Heat Flux				●		●		●	●	4
LTD	●	●	●	●	●	●	●	●	●	9

judgements on the cause and effect as well as the cross-correlation of the variables. Furthermore, interactive terms are also not considered because it is assumed that the effects are linearly dependant. Sections 3.3 and 3.4 will show that this is an accurate assumption.

The following important observations are made based on the correlation table:

1. The correlation parameter between casting speed (CS) and oscillation frequency (OF) in Table C.5 is one. This indicates that oscillation frequency is linearly dependent on casting speed. This is a common practice in industry [161]. Because of this, oscillation frequency does not provide additional information and can be excluded from the model.

2. Table C.4 shows that the heat fluxes (HWL, HWF, HNL and HNR) are perfectly positively correlated with the temperature differences (DWL, DWF, DNL and DNR respectively). This implies that the heat fluxes can be excluded from the model.

3. The thermocouple temperatures, in turn, have medium ($0.5 < \rho_{XY} \leq 0.75$) to strong ($0.75 < \rho_{XY} \leq 1$) correlations to the temperature differences which indicates that the temperature differences are computed from thermocouple temperatures. This was also validated with plant personnel[m] and thus the temperature differences should also be scrapped from the model.

4. The longitudinal temperature differences also correlate well with the thermocouple temperatures[n]. Therefore they should also be ignored from any model.

5. The drive current (DC) is not related to the mould section of the system and can also therefore be neglected from a model. This is underlined by the fact that correlations with other input variables are weak.

6. The flow rates (FWL, FWF, FNL and FNR) are weakly correlated ($0 \leq \rho_{XY} \leq 0.25$) with all other input variables. This is probably because the flow rates are usually maintained at a maximum with no changes (maximum changes of about 1.5%); implying that they have no effect on the output. The flow rates should then be dropped from the model.

[m]The dependence is not necessarily linear.

[n]These were in fact computed by the author from the thermocouple temperatures with the top row values delayed by the same amount of time as it would take a narrow strip of steel to move from the top row to the bottom row. The correlation with the top row was always strong.

7. The inlet temperature (IT) have a weak to medium correlation with the thermocouple temperatures. It is also very weakly correlated with casting speed. The inlet temperature has a medium correlation to mould level which is probably due to the stochastic nature of mould level.

8. Mould level (ML) is weakly correlated with the thermocouple temperatures. There exists a medium correlation between casting speed and mould level. This could also be due to the stochastic nature of mould level or because there may be some output feedback, *i.e.* as the speed increases, the mould level should drop momentarily. Note however that mould level is controlled at this plant with a slide gate. The effect of mould level may not be linear on other variables, but may have some non-linear effect or a delay may be present. Since it is considered to be an important variable to control, mould level will not be excluded from any model.

9. Mould controller status (MCS) only takes on values of zero or one and can therefore not be included in the linear model.

10. Casting speed has medium to strong correlations with the thermocouple temperatures and should therefore be included in the model as a manipulated variable.

3.2.3 Conclusion

Based on the above results, and the results in §3.2.1, the following conclusions can be made regarding the model structure. Due to the strong variation when defects occur compared to the case when defects do not occur (as indicated by the goodness-of-fit-tests), the thermocouple temperatures are influential (and/or carry much information) in the detection of defects. Note, however, that this does not imply that mould temperature is necessarily a cause in the formation of all the defects. The mould temperature may be influential in the formation of some defects as the literature shows, but the formation of the defects can be detected from temperature readings.

The strong correlation of casting speed with temperature gives an indication that casting speed can be used as a manipulated variable° to control temperature. This is also intuitive because a basic energy, mass and momentum balance of the system given by $\frac{\partial H}{\partial t} + v\frac{\partial H}{\partial z} + \rho v\frac{\partial v}{\partial t} = \nabla \cdot K\nabla T$ shows that an increase in casting speed must be balanced by an increase

°see Langer and Moll [162], Lally, Biegler, and Henein [163] and Lally, Biegler, and Henein [164] for optimal steady casting speed design to improve quality.

in temperature [35]. Lastly, measured disturbances that have an impact on the mould temperature are inlet temperature and mould level. These are the only variables that remain after the correlation result in table 3.5. These will then be used as measured disturbances, because they should remain constant and can therefore not be controlled. All other variables are left out because they are either non-influential on the defects or they are strongly correlated to the thermocouple temperatures.

It is proposed that the model structure be broken down into two sub-models (compare Fig. 3.6 on p.72). The first sub-model will be called the manipulated variable (MV) to intermediate variable (IV) model and the second sub-model will be called the intermediate variable (IV) to output variable (OV) model. The (only) MV is casting speed and the IVs are the thermocouple temperatures. The OVs are defects. The structure is depicted in Fig. 3.10. The measured disturbances in the MV to IV model are mould level and inlet temperature and

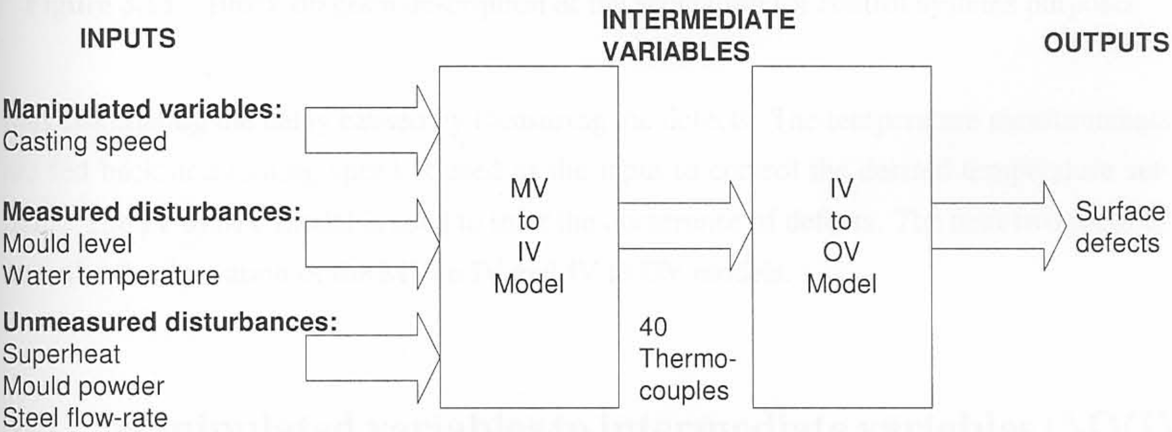

Figure 3.10 Separation of system into two models, manipulated variable to intermediate variable and intermediate variable to output variable.

the unmeasured disturbances are, amongst others, superheat, mould powder addition and flow-rate of steel into the mould.

Inlet (water) temperature was chosen to form part of the MV to IV model because a change in the inlet temperature has a direct effect on the thermocouple temperatures as the correlation showed. However, the variable could also have been made part of the IV to OV model, if desired, but this would be unwise since the variable is influential on the thermocouple temperatures and repetition of information could result. The mould level forms part of the MV to IV model because it is also already a controlled variable. Placing it in as part of the IV to OV model defeats the purpose because the inputs to the IV to OV model must ideally be changeable, and change in the mould level set-point is undesired, though this matter may be explored later.

The usefulness of this separation of the model into two sub-models is that the delay that occurs before the defects are measured is outside the feedback control loop. The IV to OV model can be used to find an optimal set-point for the temperatures and the MV to IV model can be used to design a controller to follow the temperature set-points so that the effect that temperature variation has on the formation of defects is negated. A block diagram of the structure for control is given in Fig.3.11. The temperatures are measured instantaneously,

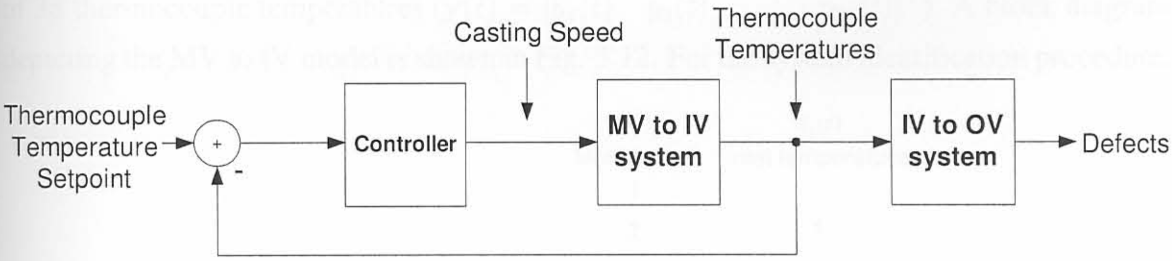

Figure 3.11 Block diagram description of the separation for control systems purposes.

thus eliminating the delay caused by measuring the defects. The temperature measurements are fed back and casting speed is used as the input to control the desired temperature set-point. The IV to MV model is used to infer the occurrence of defects. The next two sections describe the derivation of the MV to IV and IV to OV models.

3.3 Manipulated variables to intermediate variables (MVIV) model

This section describes the derivation of the manipulated to intermediate variables (MV to IV) model using system identification. Specifically, the ARX model structure will be used (see §2.2.4).

Note that in all cases it was verified (but not shown) that the residuals in the models are not auto-correlated (*i.e.* they are Gaussian) and there is no cross correlation between the input and the residuals (*i.e.* no output feedback or delays).