AN OVERVIEW OF NOISE IN SIGNAL ANALYSIS

I N Fabris-Rotelli[1] and P K Smith*[1]

[1]Department of Statistics, University of Pretoria, 0002, Pretoria, South Africa,
inger.fabris-rotelli@up.ac.za

ABSTRACT

We present an overview of common methods for modelling a signal contaminated by noise. Specifically discussed are the structure imposed on a model, the properties the noise assumed, some common statistical distributions for the noise and the Signal-to-Noise ratio as an indicator of the quality of the contaminated signal.

1. INTRODUCTION

A *signal* can be defined as 'a function that conveys information about the behaviour of a system or attributes of some phenomenon' (Priemer, 1991). The term *noise* is most commonly known in its original setting as acoustic interference, however can be generalized to a 'random error, in which there are unpredictable variations in the measured signal from moment to moment or from measurement to measurement' (O'Haver, 2012). Examples of signals are photographs, a radio emitting sound or the stock-price over time. Common examples of noise that will distort these examples of signals are random speckles, often called salt-and-pepper noise, in photographs, background static-hiss in radio and the small day to day fluctuations of the stock price which may hide an increasing trend. An understanding of the noise inherent in a measured signal is useful in the development of noise removal techniques. This paper presents common noise modelling techniques in signals as a precursor to future work in noise removal techniques and is intended as a starting point for the statistics of signal processing and noise modelling.

2. COMMON METHODS FOR MODELLING A NOISY SIGNAL

The following notation will be used: for an arbitrary space T , and for $t \in T$, the true signal will be denoted by $S(t)$, the noise contaminating the true signal by $N(t)$, and the observed signal by $X(t)$, that is, the raw data obtained containing both the signal and the noise components. The space T can be one-dimensional or multidimensional, discrete or continuous, in order that all of the models are defined in a general manner. The models described below all assume $S(t)$ and $N(t)$ are independent for each $t \in T$.

*I N Fabris-Rotelli is the corresponding author. P K Smith is the presenting author. Thanks to Department of Statistics, Statomet and SARChI at the University of Pretoria for funding.

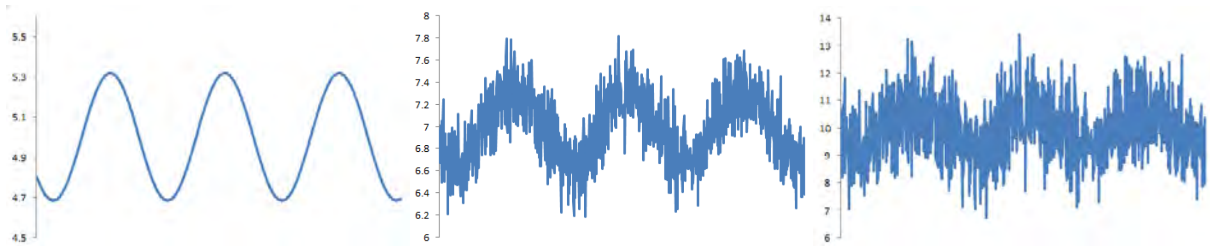

Fig. 1: Illustration of additive (center) and multiplicative (right) noise ($N(2, 0.04)$) models on a periodic signal (left). The multiplicative model distorts the signal more drastically than the additive noise model (both in vertical location and scale).

Additive noise: This is the simplest most common model for a contaminated signal where the contaminated signal X is the sum of the true signal S and the noise N, represented as, $X(t) = S(t) + N(t), t \in T$.
Multiplicative Noise: This type of noise is modelled ias $X(t) = S(t)N(t), t \in T$. In some applications (see for example Petrou & Petrou (2010)) it is convenient to transform multiplicative noise into additive noise by taking the logarithm so that $\log X(t) = \log S(t) + \log N(t)$. Notice that this transformation assumes that $N(t)$ and $S(t)$ are strictly positive values, a reasonable assumption since one would often expect multiplicative noise to distort the signal by random stretching or shrinking rather than to change its sign. Also note that $E[\log N(t)] = 0$ if and only if $E[N(t)] = 1$.
Figure 1 illustrates the additive and multiplicative noise models.
Impulse Noise: This noise describes the random occurrence of sudden sharp spikes (Yaroslavsky, 2000) and can be modelled in general as: $X(t) = (1 - e)S(t) + eN(t)$ where $e = 1$ with probability p and 0 with probability $1 - p$. Two important impulse-noise distributions are the Bernoulli-Gaussian and Poisson-Gaussian model where the Bernoulli and Poisson distributions govern the impulse occurrence in the two models respectively and the Gaussian distribution the impulse magnitude in both (Vaseghi, 2000). A common manifestation of impulse noise is salt-and-pepper noise in images. In an 8-bit gray-scale image, pixel values range from 0 to 255 (completely black to completely white) but impulse noise can be enough to spike a pixel value beyond these bounds in which case the pixel is entirely dominated by noise. Salt-and-pepper noise model can thus be modelled as 255 with probability αp, 0 with probability $(1 - \alpha)p$ and $S(t)$ with probability $1 - p$ where $0 \leq \alpha, p \leq 1$.

3. SOME DESCRIPTORS AND STATISTICAL PROPERTIES REGARDING NOISE

The type of noise found in measured raw data is generally categorized according to its specific properties. Several common properties are described below.
Independence: The noise $N(t)$ is said to be *independent* on T if, for every $a_i < b_i$ we have $P[a_1 \leq N(t_1) \leq b_1, \ldots, a_k \leq N(t_k) \leq b_k] = \prod_{i=1}^{k} P[a_i \leq N(t_i) \leq b_i]$, for distinct elements of T, $t_1, \ldots, t_k$ (Bain & Engelhardt, 1992), so that the value of the noise experienced at each $t_i \in T$ is unaffected by the noise at any other $t_j \in T$. This assumption is common, (Fabris-Rotelli et al., 2010; Rank et al., 1999).

Fig. 2: Illustration of impulse noise ($N(127, 100)$) (center) and salt-and-pepper noise (right). The parameters used were $\alpha = 0.5$ and $p = 0.1$.

Zero-mean (unbiased) noise: Noise is said to be *zero-mean* or *unbiased* if its expected value is zero, namely $E[N(t)] = 0$. This assumption is associated with additive noise. If the mean of the noise is known to be non-zero, then the noise can be transformed to be zero-mean simply by subtracting this mean from the noisy signal (Petrou & Petrou, 2010).

Correlation: Correlation is a statistical measure of dependence between two variables. The correlation between two random variables A and B is defined as $\text{corr}(A, B) = \frac{cov(A,B)}{\sigma_A \sigma_B}$ where σ_A^2, σ_B^2 denote the variance of A and B respectively although often in signal processing the denominator in this expression is omitted (see for example Leneman & Lewis (1966); Yaroslavsky (2000)). Correlation within an element occurs when $\text{corr}[N(t), S(t)] \neq 0$ for each value of t, implying that the noise is signal-dependent. An example is Poisson noise (Section 4). Autocorrelation, correlation between elements, occurs when $\text{corr}[N(t), N(t+r)] \neq 0$ where $t, t+r \in T$ are distinct. Certain methods, for example the 3-point moving-average smoother, of signal processing can introduce autocorrelated noise into the signal.

Stationarity: $N(\cdot)$ is said to be stationary on T if, for $t_1, t_2, \ldots, t_n \in T$, and $t_1+k, \ldots, t_n+k \in T$ for some k, the joint distribution of $N(t_1), \ldots, N(t_n)$ is the same as the joint distribution of $N(t_1+k), \ldots, N(t_n+k)$, see Esch et al. (2010). Leneman & Lewis (1966) assume weakly stationarity where only the first two moments of the noise are stationary on T . Under this assumption $\text{corr}(t, t+r)$ depends only on r , the 'geographical' relationship between the two elements (Goossens et al., 2008).

Colour: In electric signals a common method of describing noise is by its frequency power spectrum representing signal power at different frequencies (as opposed to different times), namely how much signal lies in each of the frequency bands (O'Haver, 2012). Any continuous periodic signal s_k with frequency ω_k can be written as a sum of possibly infinite terms of sinusoidal signals varying in frequency (Shatkay, 1995). The average power of the original signal can then be calculated as $P_k = \frac{1}{Q} \int_0^Q s_k^2(t)dt$ where Q denotes the period. The concept of seperating a signal into many sinusoidal components of different frequencies can be generalised to non-periodic, discrete and even random signals via the (Discrete) Fourier Transform. The Fourier transform of signal $X(t)$ at frequency ω is given by $F(\omega) = \int_{-\infty}^{\infty} e^{-2\pi i\omega t} X(t)dt$ (Shatkay, 1995) and will often result in a complex value $F(\omega) = a(\omega) + b(\omega)i$. The power spectrum is then $P(\omega) = |F(\omega)|^2$ for each ω. If the frequency spectrum of the noise $N(t)$ is known, the power is defined as $P = E[N^2(t)]$ so that the high power frequencies are those with high variance due noise. So these frequencies are expected to be more corrupted by noise. The colour of the noise describes power as being proportional to $1/f^\beta$ for varying values of β where f denotes frequency. White noise ($\beta = 0$) and pink noise ($\beta = 1$) are the most common noise colours. White noise assumes a constant power frequency spectrum implying equal noise strength at all frequencies (similarly white light is comprised of equal amounts the colours of the spectrum). While this type of noise is in practice impossible since it requires infinite energy (Carter, 2002), it is still very widely used (see Pauluzzi & Beaulieu (2000)) and occurs in statistics as i.i.d noise. Pink noise (Carter, 2002) exhibits a decrease in noise strength as frequency increases. In O'Haver (2012) pink noise is just defined with more power at the lower end of the spectrum and $1/f$ noise is defined as a subset of pink noise. Pink noise is commonly found in electronics (Ohguro et al., 2012) and other applications. Other noise colours, such as brown, grey or green, exist however they are less common.

4. COMMON DISTRIBUTIONS FOR NOISE AND SIGNALS USED IN MODELS

We now examine some commonly used distributions and the reasons for their usage.

Gaussian distribution: This is the most commonly used distribution when modelling noise (O'Haver (2012); Pauluzzi & Beaulieu (2000); Rank et al. (1999)) since observed signal noise is often the sum of many random events, the Central Limit Theorem dictates an approximate Gaussian distribution under certain conditions (O'Haver, 2012). Gaussian noise is used in the modelling of electric current noise - where it is typical to experience shot noise (random movement of electrons), thermal noise, flicker noise (from unknown origin) and burst noise (due to imperfections in semiconductors) (Carter, 2002).

Poisson distribution: The Poisson distribution occurs when counting the number of events at each $t \in T$, so that $X(\cdot) \sim Poisson(\cdot)$. An example of a noise commonly modelled by the Poisson distribution is shot noise (Petrou & Petrou, 2010). A common occurrence of this type of noise is in a grayscale image, where pixel signal intensity is as a result of the number of discrete photons observed at that pixel. Of course, it is well-known that the Poisson distribution converges to the Gaussian distribution as its parameter increases (Bain & Engelhardt, 1992) and hence if the signal has a reasonably high expected value, the Normal distribution is a reasonable approximation for noise of this type (reinforcing the common usage of the Gaussian model). However, use of the Normal distribution makes no assumption as to the relationship between signal strength and the magnitude of the noise, whereas the Poisson distribution implicitly does, since its variance is equal to its mean (which is, in this case, the pure-signal intensity) (Wentzell & Brown, 2000). Poisson noise is thus an example of signal-dependent noise.

Rayleigh distribution: The Rayleigh distribution is useful in describing the magnitude of signals in the complex plane. Suppose $X(t) = a_t + b_t i$ is a complex number where $i = \sqrt{-1}$ and $a_t,\ b_t$ are identically and independently $N(0, \sigma^2)$ distributed. Then, the magnitude of $X(t)$ at t is given by $||X(t)|| = \sqrt{a_t^2 + b_t^2}$

which has a $Rayleigh(\sigma)$ distribution. The Rayleigh distribution is commonly used in medical imaging to model speckle noise. MRI scanners measure the magnitude of a complex signal created by changes in the magnetic field and are hence often modelled using the Rayleigh or the more general, the Rice/Rician distribution (Wang et al., 2011). $R = \sqrt{a_t^2 + b_t^2}$ is said to have $Rice(\alpha, \sigma)$ distribution if $a_t \sim N(\alpha\cos(\theta), \sigma^2)$ and $b_t \sim N(\alpha\sin(\theta), \sigma^2)$ are independently distributed and θ is real. The expressions for the normal mean parameters are related to the polar form of a complex number, namely $c = \alpha(\cos(\theta) + i\sin(\theta))$.
Stable distributions: Random variable X is said to be *stable* if, for independent variables X_1, X_2, with the same distribution as X and for $a,b>0$, there exists some $c>0$ and some $d\in\mathbb{R}$ such that $aX_1+bX_2 \stackrel{d}{=} cX+d$ (Nolan, 2007). The importance of stable distributions in signal processing is primarily as a result of the Generalised Central Limit Theorem which states that if the sum of i.i.d. random variables with finite or infinite variance converges to a distribution by increasing the number of variables, the limit distribution must be stable (Scheaffer & Young, 2009). In most cases stable distributions lack a closed form density function, however they can be described by their characteristic function: X is stable if and only if X has characteristic function

$$E\left[\exp(iuX)\right] = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 - i\beta(\tan\frac{\pi\alpha}{2})\mathrm{sgn}(u)\right] + i\delta u\right) & \alpha \neq 1 \\ \exp\left(-\gamma|u|\left[1 + i\beta\frac{2}{\pi}\mathrm{sgn}(u)\log|u|\right] + i\delta u\right) & \alpha = 1 \end{cases}$$

where $\alpha \in (0, 2]$ determines the thickness of the tails, $\beta \in [-1, 1]$ is a skewness parameter, $\gamma \geq 0$ is a scale parameter, $\delta \in \mathbb{R}$ is a location parameter, and $\mathrm{sgn}(\cdot)$ is the sign function (Nolan, 2007). Note that this is only one of many parametrizations of the characteristic function. In probability theory the characteristic function is simply the Fourier transform of the probability density function (Papoulis, 1991).
Setting $\beta = 0$, we obtain the the symmetric α-stable distributions (SαS) class. Like the Gaussian distribution, SαS models are smooth, unimodal, symmetric with respect to the median and bell-shaped (Shao & Nikias, 1993) however the SαS distribution is more versatile due to the characteristic exponent α. The thickness of the tails increases as the value of α decreases (Tsakalides & Chrysostomos, 1998). Setting $\alpha = 2$ yields the Gaussian distribution (Nolan, 2007) hence the SαS distribution is always thicker-tailed than (or equivalent to) the Gaussian distribution. This thicker tail makes the SαS distribution more suitable than the Gaussian distribution for modelling signals prone to large spikes and outliers and in fact this is their primary use (Shao & Nikias, 1993). It should also be noted that, for $\alpha \neq 2$, the SαS distribution has infinite variance hence noise removal techniques relying on finite variance are not appropriate for stable distributions (Shao & Nikias, 1993). An interesting comparison between the Gaussian distribution and the SαS distribution in approximating real-world audio signals is done in (Georgiou et al., 1999).

5. SIGNAL TO NOISE RATIO (SNR)

The Signal to Noise Ratio (abbreviated SNR) is a measurement of the amount of meaningful information (the underlying true signal) compared to the amount of unwanted information (the noise) in an observed signal, that is, a measure of signal strength, and is commonly used to compare signal acquisition methods

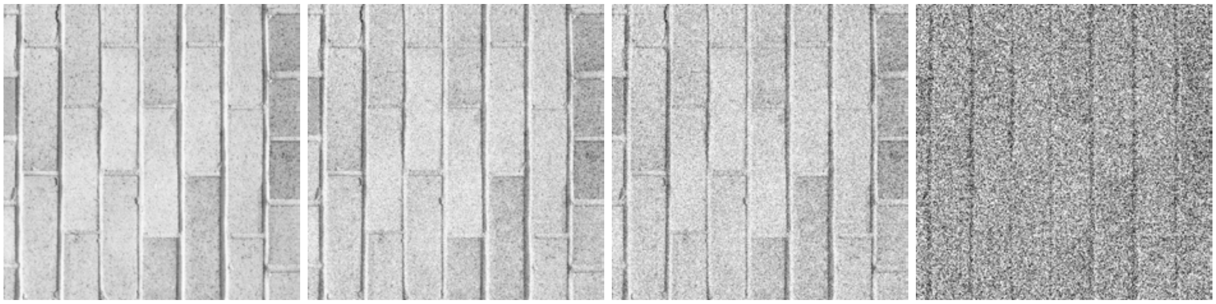

Fig. 3: Illustration of the Rose criterion: from left to right - original image, SNR = 9 (strong), 5(medium) and 1(weak)

or hardware (Parish et al., 2000). There are several different mathematical definitions for SNR but most have the common feature of being the ratio of some measure of location of the signal to some measure of spread of the noise. A definition of SNR that fits this description is $SNR = \frac{\mu_s}{\sigma_n}$ (Parish et al., 2000), where μ_s denotes the mean of the signal and σ_n^2 denotes the variance of the noise. This definition of SNR has the drawback that, if $\mu_s = 0$, we will always have $SNR = 0$. Hence this definition of SNR is usually applied to strictly non-negative signals, thus ensuring $\mu_s > 0$ unless $S(t) = 0$ for all $t \in T$, in which case we are only observing noise. An application where this definition is useful is in images where μ_s is the signal intensity (a strictly non-negative value) and σ_n^2 is the variance of the signal intensity (Watanabe et al., 2002). The Rose criterion states that, using this definition for the SNR, an SNR value of at least 5 is required for the human eye to distinguish features in an image with 100% certainty (Bushberg, 2002), in other words an SNR of less than 5 indicates the noise in the image overpowers the true signal. Figure 3 illustrates the Rose criterion. For a periodic signal oscillating around 0 (and hence interchanging between positive and negative values), we must find another way to measure the strength of the signal compared to that of the noise. Observing that in this case the amplitude of the signal is an indication of the signal strength, a proposed solution is to replace the mean of the signal in the formula with the amplitude of the signal, that is, $SNR = \frac{amplitude\ of\ signal}{\sigma_n}$ (Fabris-Rotelli et al., 2010). This definition can be modified for signals with varying amplitudes to $SNR = \frac{average\ amplitude\ of\ signal}{\sigma_n}$ (O'Haver, 2012).

For a deterministic signal, a definition which seems commonly used in electrical engineering is $SNR = \frac{P_s}{P_n}$ (Pauluzzi & Beaulieu, 2000) where P_s and P_n denote the average power of the signal and noise respectively. Power is the rate of energy transferal and is calculated for a signal as $P_s = \frac{1}{t_u - t_l} \int_{t_l}^{t_u} S^2(t)dt$ and similarly for noise. For discrete power, the integrals are simply replaced with summations. Note that this value can be interpreted as the average value for the signal squared calculated over the time period $[t_l, t_u]$. Bearing this interpretation in mind, when the signal is no longer deterministic but rather a stationary stochastic process, we have that P_s and P_n are unbiased estimators for $E[S^2(t)]$ and $E[N^2(t)]$ respectively. If the noise has a zero-mean (i.e. $E[N(t)] = 0$) then using the well known identity $var[N(t)] = E[N^2(t)] + E[N(t)]^2$, it follows that P_n is an unbiased estimator for σ_n^2. It should be noted at this point that all of the above definitions fail when the noise has infinite variance since $\lim_{\sigma \to \infty} \frac{a}{\sigma} = 0$ for all a, implying that all these definitions will have $SNR = 0$. This problem can be overcome by using some other measure of spread for the noise (for example, in the $S\alpha S$ distribution, the dispersion parameter γ plays a role analogous to variance (Shao & Nikias, 1993)).

6. CONCLUSION

This paper has been intended as an introduction into the methods in which noise is commonly handled in literature. A broad overview has been given of the models often associated with noise, some common characteristics of noise and some distributions with which noise is often associated with no emphasis being made on any one specific application. Furthermore, the Signal-to-Noise Ratio was discussed as an often used method of quantifying the strength of a signal versus amount of noise present. Several definitions of the Signal-to-Noise Ratio were presented for use in different applications. Future work will be the investigation of methods of modelling noise in more specific applications, such as images, as well as an investigation on the different methods of noise removal with comparisons being made as to their effectiveness under different noise-types.

References

Bain, L. & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics*. Duxbury.

Bushberg, J. (2002). *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins.

Carter, B. (2002). *Op Amps for Everyone*. Texas Instruments.

Esch, T., Rungeler, M., Heese, F., & Vary, P. (2010). Combined reduction of time varying harmonic and stationary noise using frequency warping. In *ASILOMAR Proceedings* (pp. 533 –537).

Fabris-Rotelli, I., van Oldenmark, K., & van Staden, P. (2010). Evaluation of noise removal in signals by LULU operators. In P. D. et al (Ed.), *Proceedings of SASA* (pp. 44–51). North-West University.

Georgiou, P., Tsakalides, P., & Kyriakakis, C. (1999). Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE T Multimedia*, 1(3), 291 –301.

Goossens, B., Pizurica, A., & Philips, W. (2008). EM-based estimation of spatially variant correlated image noise. In *ICIP Proceedings* (pp. 1744 –1747).

Leneman, O. & Lewis, J. (1966). Random sampling of random processes: Mean-square comparison of various interpolators. *IEEE T Automat Contr*, 11(3), 396 – 403.

Nolan, J. (2007). *Stable Distributions: Models for Heavy-Tailed Data*. Springer Verlag.

O'Haver, T. (2012). An introduction to signal processing with applications in chemical analysis. Available online [Accessed 25 June 2012].

Ohguro, T., Higashi, Y., Okano, K., Inaba, S., & Toyoshima, Y. (2012). The optimum device parameters for high rf and analog/ms performance in planar mosfet and finfet. In *VLSIT Proceedings* (pp. 149 –150).

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*, chapter Characteristic Functions. McGraw-Hill: New York, third edition.

Parish, T., Gitelman, D. R., LaBar, K. S., & Mesulam, M. M. (2000). Impact of signal-to-noise on functional MRI. *Magn Reson Med*, 44, 925–932.

Pauluzzi, D. R. & Beaulieu, N. C. (2000). A comparison of SNR estimation techniques for the AWGN channel. *IEEE T Commun*, 48(10), 1681–1691.

Petrou, M. & Petrou, C. (2010). *Image Processing: The Fundamentals*. John Wiley & Sons.

Priemer, R. (1991). *Introductory Signal Processing*. World Scientific.

Rank, K., Lendl, M., & Unbehauen, R. (1999). Estimation of image noise variance. *Vision, Image and Signal Processing, IEE Proceedings-*, 146(2), 80–84.

Scheaffer, R. & Young, L. (2009). *Introduction to Probability and Its Applications*. Cengage Learning.

Shao, M. & Nikias, C. (1993). Signal processing with fractional lower order moments: stable processes and their applications. *P IEEE*, 81(7), 986 –1010.

Shatkay, H. (1995). *The Fourier Transform - A Primer*. Technical report, Providence, RI, USA.

Tsakalides, P. & Chrysostomos, N. L. (1998). *A Practical Guide to Heavy Tails*, chapter 5, (pp. 379–404). Birkhauser.

Vaseghi, S. V. (2000). *Advanced Digital Signal Processing and Noise Reduction*, chapter 12, (pp. 355–377). John Wiley & Sons Ltd., 2nd edition.

Wang, G., Pan, Z., Xu, J., Zhang, Z., Liu, C., & Ding, J. (2011). Active contour model for images corrupted by multiplicative noise with rayleigh distribution. In *CIS Proceedings* (pp. 124 –127).

Watanabe, M., Williams, D. B., & Tomokiyo, Y. (2002). Comparison of detection limits for elemental mapping by EF-TEM and STEM-XEDS. volume 8 (pp. 1588–1589).

Wentzell, P. D. & Brown, C. D. (2000). *Signal Processing in Analytical Chemistry*, (pp. 9764–9800). John Wiley & Sons Ltd.

Yaroslavsky, L. (2000). Advanced image processing lab: A tutorial, EUSIPCO200, lecture 2. Available online [Accessed 25 June 2012].