GEOPHYSICAL DATA ANALYSIS

Class Notes by Bob Parker

CHAPTER 1: STOCHASTIC PROCESSES

1. Introducing Ordered Data

In the statistical portion of this class the observations is far have been almost entirely of the kind that has no special order associated with it. From now on we will study data in which the order is important. Usually we will call the sequence x_1, x_2, $x_3 \cdots x_n$, a **time series**; the index n denotes the time $t = n\,\Delta t$ at which the observation was made with respect to some initial time; obviously we are sampling here uniformly in time at an interval Δt. Obviously we could equally well be treating a data set sampled in space or, much more rarely, some quite different independent variable. For theoretical purposes it is useful to be able to treat time as continuous sometimes, and to work with $x(t)$, but here when we introduce the concept of a random variable for the observation, things can become mathematically extremely difficult without some further quite severe simplifications. For ideal models we will often want to consider infinite sequences, maybe infinite in both directions.

So a **stochastic process** $\{X_n\}$ or $\{X(t)\}$ is a family of random variables indexed by the integer n, when it is a discrete process, or by the real number t, when it is a continuous process. (Random variables or functions will normally be denoted by upper case letters, but the converse is not true.) As with ordinary random variables, X_1 for example, has no definite value, but is to be thought of as an infinite collection of values from which a particular experiment will extract a value. A given data series is conceptually the result of an experiment that could be repeated as many times as we like. We usually have just one **realization,** drawn from an **ensemble** of alternative series, all generated by the underlying process fully described by its **probability distribution function** (PDF). When we think of operations such as taking the average or **expected value** at a particular n or t, in general this is an average over different realizations for the same n or t. See Figure 1 for a picture.

Even in discrete time the general stochastic process is a horribly complex affair. To specify it completely requires the joint PDF of every element X_n with every other element. The simplest case is the Gaussian model, which you already have met. Given N data we have the joint PDF is in the form

$$\phi(\mathbf{X}) = \frac{1}{(2\pi)^{½N} \det(C)} \exp\left[-½(\mathbf{X}-\bar{\mathbf{x}})^T C^{-1}(\mathbf{X}-\bar{\mathbf{x}})\right] \qquad (1.1)$$

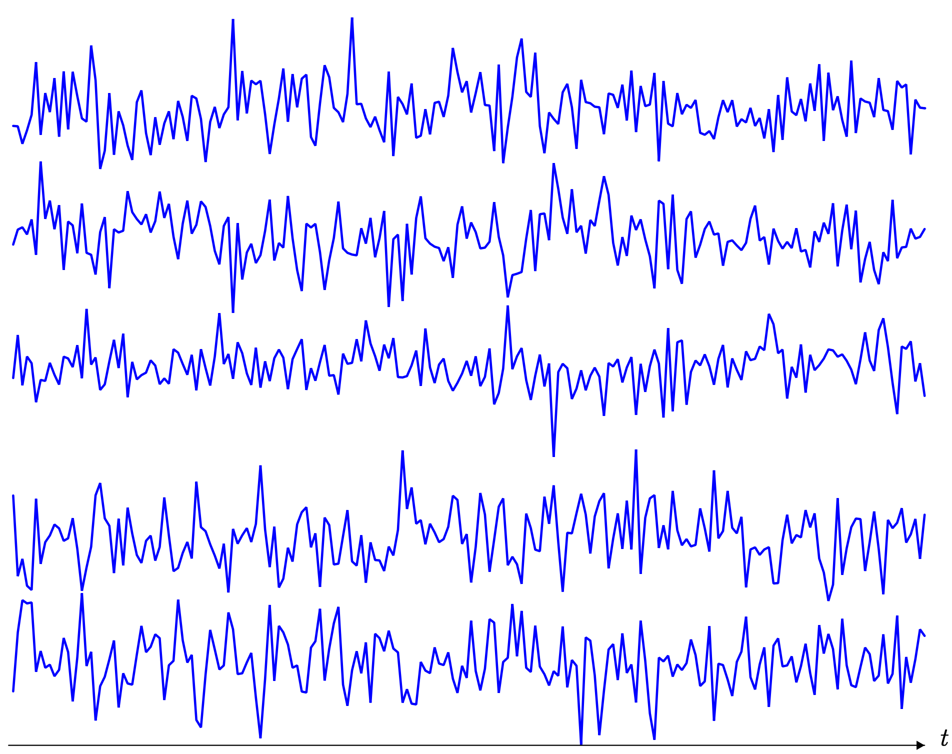
where $\mathbf{X} = (X_1, X_2, \cdots X_N)^T$, $\bar{\mathbf{x}} \in \mathrm{I\!R}^N$ is vector of mean values, and $C \in \mathrm{I\!R}^{N\times N}$ is the covariance matrix, a symmetric, positive definite matrix that describes the correlations among the random variables X_n. The reason this is complicated is that to describe completely these N random variables, we need a total of ½$N(N+3)$ parameters, far more than we are likely to have observations: if we have even a short time series of say 50 numbers we would need 1325 parameters; for a reasonably long data series of 2000 members, more than 2 million parameters. This is for the Gaussian distribution, the simplest kind of PDF. Now imagine what this means if we consider continuous time instead; the problem is almost completely intractable, and such a general treatment has no practical value because it would be impossible to make estimates of the necessary parameters, even if the mathematics were possible, which turns out to be very hard for the general case. See *Priestley,* Chapter 3 for a further discussion.

2. Stationary Processes and Autocovariance

The perfectly general stochastic process is too general to be useful, and so the conventional wisdom is to focus on a much more restrictive class of random processes called **stationary processes.** The idea here is that, while the actual observables vary in time, the *underlying statistical description is time invariant.* This can be weakened a bit, but we will consider only these so-called **completely stationary** processes. This has the gratifying effect of reducing the number of parameters needed to a manageable number. For example, the mean value of a stationary process:

$$\mathcal{E}\,[X_n] = \bar{x} \tag{2.1}$$

Figure 1: Five realizations of the same stochastic process.



is a constant independent of n, or of time t if the process is continuous. It is often assumed the mean is zero, since it is a trivial operation to add the mean value back into the data series, if necessary. Of course, many observational series do not look as if the mean is constant – there may be a **secular trend.** Stationarity is such a powerful and useful property that one often attempts to convert an evidently nonstationary series into a stationary process, for example, by fitting a straight line trend, or forming a new stationary sequence by differencing:

$$Y_n = X_{n+1} - X_n \tag{2.2}$$

Recall from the variance of single random variable the definition

$$\sigma_X^2 = \mathrm{var}\,[X_n] = \mathcal{E}\,[(X_n - \bar{x})^2]\,. \tag{2.3}$$

With stationarity this number must also be independent of n (or t). And the covariance between any two random variables in the sequence cannot depend on where we are in the series and therefore

$$\mathrm{cov}\,[X_m,\, X_n] = \mathcal{E}\,[(X_m - \bar{x})(X_n - \bar{x})] = R_X(m - n) \tag{2.4}$$

that is, a function R_X of the interval between the two points. The function R_X is called the **autocovariance.** For continuous processes this is usually written

$$\mathrm{cov}\,[X(t),\, X(t+\tau)] = R_X(\tau)\,. \tag{2.5}$$

Then τ is called the **lag**. Observe that by definition

$$R_X(0) = \sigma_X^2\,. \tag{2.6}$$

Also notice that, because of stationarity, one can set $s = t - \tau$ in (2.5) and the result will be the same, since the answer is independent of which time was selected. This yields:

$$R_X(-\tau) = R_X(\tau) \tag{2.7}$$

which shows that the autocovariance function is an even function of the lag. We see then that a stationary process does not contain information on which way time is flowing – it is the same process if time is reversed.

Returning to a stochastic process with a Gaussian PDF as in (1.1), we see that in place of the vector $\bar{\mathbf{x}}$ of mean values we have a single number. How about the covariance matrix? You may recall that the j-kth entry of C is

$$C_{jk} = \mathrm{cov}\,[X_j,\, X_k] = R_X(j - k)\,. \tag{2.8}$$

Hence the covariance matrix has the same values on all its diagonals

$$C = \begin{bmatrix} R_X(0) & R_X(1) & R_X(2) & R_X(3) & \cdots \\ R_X(1) & R_X(0) & R_X(1) & R_X(2) & \cdots \\ R_X(2) & R_X(1) & R_X(0) & R_X(1) & \cdots \\ R_X(3) & R_X(2) & R_X(1) & R_X(0) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{2.9}$$

This type of matrix is called a **Toeplitz** matrix. Now instead of $\frac{1}{2}N(N+3)$, there are only $N+1$ parameters in the Gaussian stationary PDF.

Just because the stochastic process is composed of random variables does not mean it is completely unpredictable. Recall the correlation coefficient of two random variables:

$$\rho_{XY} = \frac{\operatorname{cov}[X, Y]}{\sqrt{\operatorname{var}[X]\operatorname{var}[Y]}}. \tag{2.10}$$

Then from (2.5) we have the correlation coefficient between any two points in a sequence in continuous time is

$$\rho(\tau) = \frac{R_X(\tau)}{\sigma_X^2} \tag{2.11}$$

and a similar result for discrete processes. This function is called the **autocorrelation function.** Thus, unless the autocovariance function is exactly zero for the lag τ, one can predict something about the value further along in the sequence at $X(t+\tau)$ from value at t, because they are correlated.

So far we have considered the average of the process, and averages of the products $(X(t_1)-\bar{x})(X(t_2)-\bar{x})$. Such averages are examples of **moments** of the process. The second order moments concern the variance and autocovariance function, and nothing else. Higher order moments, those involving the product of three or more Xs can obviously be defined, but if the process is based on a Gaussian PDF, all the information about the PDF is contained in the second order moments and the higher moments can be predicted. Even if the PDF is not Gaussian a great deal can be learned about a process from its second order moments, and so the higher order moments are rarely investigated. We will follow this path.

3. White Noises and their Relatives

Let us look a few concrete examples of stationary stochastic processes. The simplest (but artificial) example is that of **white noise**. This is defined as a stationary process in which the random variable at any point is independent of every other variable. It is also common to assume the mean value of white noise is zero. In the discrete case we have

$$R_W(n) = \sigma_0^2\, \delta_{n0} \tag{3.1}$$

where δ_{jk} is the Kroenecker delta symbol. Hence in this case the random process is unpredictable to the extent that we learn nothing about the next or subsequent values from the current one. The results are still not completely unpredictable, because we can say something about the range of values to be expected on account of the known variance, σ_0^2.

For continuous processes, it turns out that a white noise is singular because the variance at any point must be infinite:

$$R_W(\tau) = s^2\, \delta(\tau) \tag{3.2}$$

But these definitions do **not** completely specify the stochastic process. At any particular time t (or index n) X is a random variable with a PDF; that PDF will be the same for every t, but so far we have no specified it. Of course, the most common choice is the Gaussian, so that

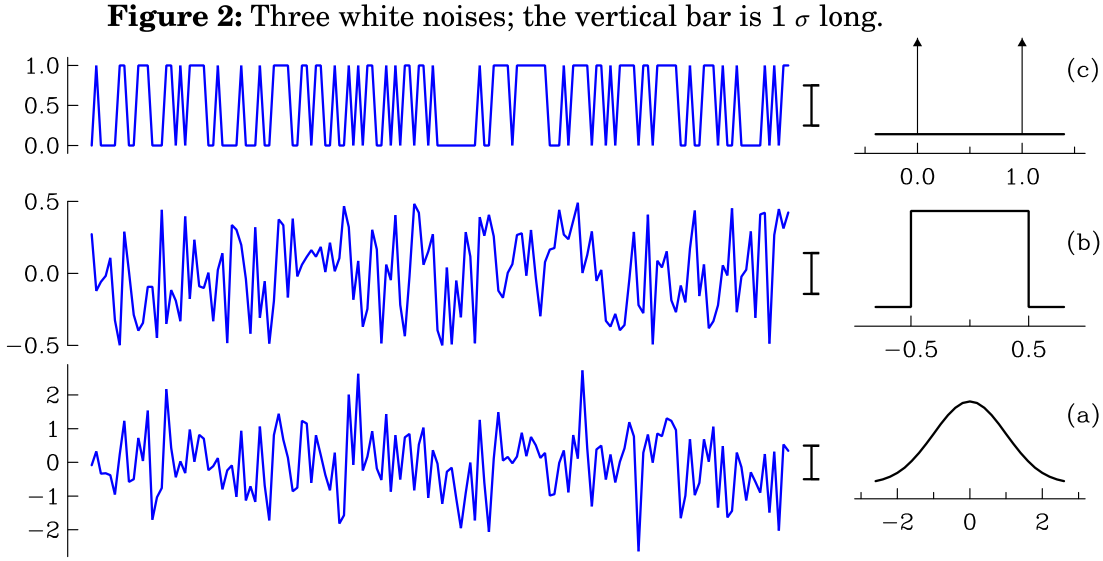
$$\phi(X) = \frac{1}{\sigma_0\sqrt{2\pi}}\, e^{-\frac{1}{2}(X/\sigma_0)^2}\ . \tag{3.3}$$

Because of the statistical independence, the joint PDF is

$$\psi(X_1, X_2, X_3, \cdots) = \phi(X_1)\cdot\phi(X_2)\cdot\phi(X_3)\cdot\cdots\ . \tag{3.4}$$

This is of course **Gaussian white noise;** shown by in Figure 2a.

Figure 2: Three white noises; the vertical bar is 1 σ long.



We allow any other PDF for X_n; suppose we choose a uniform distribution:

$$\phi(X) = b^{-1} \operatorname{box}(X/b) \tag{3.5}$$

where now $\operatorname{var}[X] = b^2/12$. This is a different kind of white noise, easily seen in Figure 2b. The joint PDF is given by (3.4) again, with the appropriate choice of ϕ. This kind of white noise turns up quite frequently in real measurements: **round-off noise.** Suppose a discrete time series is recorded and the number is rounded, say to 1 decimal. If the series varies by much more than ± 0.1 over each Δt, the value recorded is the true value of the measurement plus an unpredictable (ie random) amount that lies in $(-0.05, +0.05)$ with a uniform distribution. Thus the recorded series appears to be the true signal with a white noise added, since the consecutive values of the rounding error are uncorrelated. The noise is zero mean and its variance is in this case $0.1^2/12$. If the last significant digit doesn't change very often in the series, then the round-off noise added is no longer uncorrelated; but then it seems likely the signal is not being recorded with enough accuracy.

Another random white series with a limited range is the **random telegraph signal:** this one switches discontinuously between only two values, say, 0 and 1:

$$\phi(X) = \tfrac{1}{2}[\delta(X) + \delta(X-1)]\,. \tag{3.6}$$

Here the mean value is not zero but one half, and the variance is a quarter. The random telegraph signal is used as a calibration signal for seismometers or other instruments since it is easy to generate electronically. A zero mean version is sometimes suggested as a model for the Earth's dipole moment over time scales of 10^5 to 10^6 years, but this is actually very implausible because the moment is far from constant between reversals. See Figure 2c for a picture.

These three examples of white noise are clearly different to the eye. Somewhat remarkably to me at least, if they are normalized to the same variance and converted into a sound track, they sound identical – the ear doesn't have a very good density function discriminator.

These three sequences were obviously not observational; they were made with a random number generator in MATLAB. We can obtain other kinds of stochastic sequences by filtering white noise in various ways; the series in Figure 1 were generated that way. If one filters white noise, the result tends to be Gaussian in distribution as a consequence of the Central Limit Theorem. Many physical processes can be thought of as resulting from this kind of process, so Gaussian distributions are often assumed in a signal, and are quite often observed too, but not always. A common exception appears to be traces of marine magnetic anomalies, which are usually heavy in the tails compared with a Gaussian distribution.

Let us briefly consider the simplest kind of filter, a **FIR:** this stands for a **Finite Impulse Response filter** (aka MA or **Moving Average**). For a FIR filter we convolve the white noise sequence W_j with a finite number of weights w_k:

$$Y_n = \sum_{k=1}^{K} w_k W_{n-k} \, . \tag{3.7}$$

Then we can calculate the autocovariance of the new sequence using the definition (2.4); let's assume for simplicity the mean is zero. Then

$$R_Y(l) = \text{cov}\,[Y_{n+l}, Y_n] = \mathcal{E}\,[Y_{n+l} Y_n] \tag{3.8}$$

$$= \mathcal{E}\,[\sum_{j=1}^{K} w_j W_{n+l-j} \sum_{k=1}^{K} w_k W_{n-k}] \tag{3.9}$$

$$= \sum_{j=1}^{K} \sum_{k=1}^{K} w_j w_k \, \mathcal{E}\,[W_{n+l-j} \, W_{n-k}] \tag{3.10}$$

$$= \sum_{j=1}^{K} \sum_{k=1}^{K} w_j w_k \, \sigma_0^2 \delta_{l-j+k,0} \quad . \tag{3.11}$$

The delta symbol vanishes except when $l - j + k = 0$, namely when $j = l + k$; so

$$R_Y(l) = \sigma_0^2 \sum_{k=1}^{K} w_k \, w_{k+l} = \sigma_0^2 \, w_k * w_{-k} \, . \tag{3.12}$$

Of course we see that the previously uncorrelated (zero covariance) white noise W_k has become correlated. Notice that R_Y does go to zero once $|\,l\,| > K$.

The student is invited to verify that the same result is obtained for the continuous time version: if

$$Y = w * W \tag{3.13}$$

$$R_Y = s^2 \, w(t) * w(-t) \tag{3.14}$$

A lot of space is wasted (in my opinion) in books (eg, Priestley) applying various recursive filters (AR, ARMA filters) to white noise and looking at the consequences.

Let us move on to some examples of actual observations that might be realizations of stationary processes.

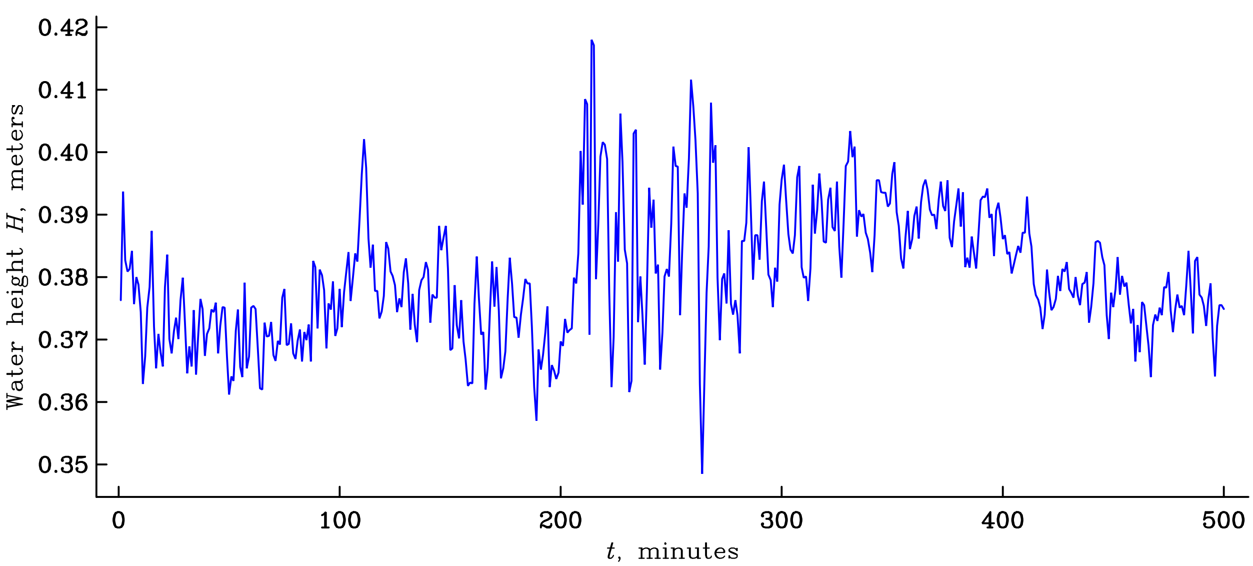
4. Examples from the Real World

Our first example is a recording of the water height at the edge of a lake over a period of several hours. The water is rising and falling in response to the wind, but because of standing waves in the basin, there are well-defined oscillations in time, known as a **seiche.** Below in Figure 3 we see a record of nearly 10 hours; I have data for the best part of a full day, 1301 minutes. First observe, the mean is clearly not zero. Next notice the oscillations, which are not completely regular, but none-the-less there is a suggestion of periodicity. This definitely looks like a stochastic process, but one might be skeptical that it is stationary, given the amplitude increase at about 210 minutes. None-the-less we will stick with that model because it is so useful.

We can ask if this looks like a good approximation to a white noise. Then values at one time would not be related, even on average, to values at other times. That looks improbable to the eye. We can also draw a scatter plot, for example, as shown in Figure 4 where the observed value at one time is plotted against the height 8 minutes later. A very clear correlation is visible. I estimate the correlation coefficient by (2.10) to be 0.626. In the sample there were 1293 data, and we can calculate the probability that ρ would be this big in an uncorrelated Gaussian sample by the t test (see Rice's book for details): we find $t_{N-2} = 28.8$, and the chances of random data exceeding this value are less than 10^{-180}.

How good is the Gaussian model? Ignoring the fact this is a sequence, and just treating the values as an ordered set, we can find a histogram, or better, the cumulative distribution function. This can be tested against the Gaussian model with the Kolmogorov-Smirnov test. Figure 5 shows the quantile-quantile plot on the Gaussian hypothesis: the fit is remarkably good. From the value if the D_N statistic we calculate

Figure 3: Water height on the shore of the Salton Sea; a seiche.



that the probability of a random sample exceeding the observed value is 0.83, so the fit to the Gaussian model is excellent, maybe even slightly too good.

Figure 4: Scatter plot of data in Figure 3 with a lag of 8 minutes.

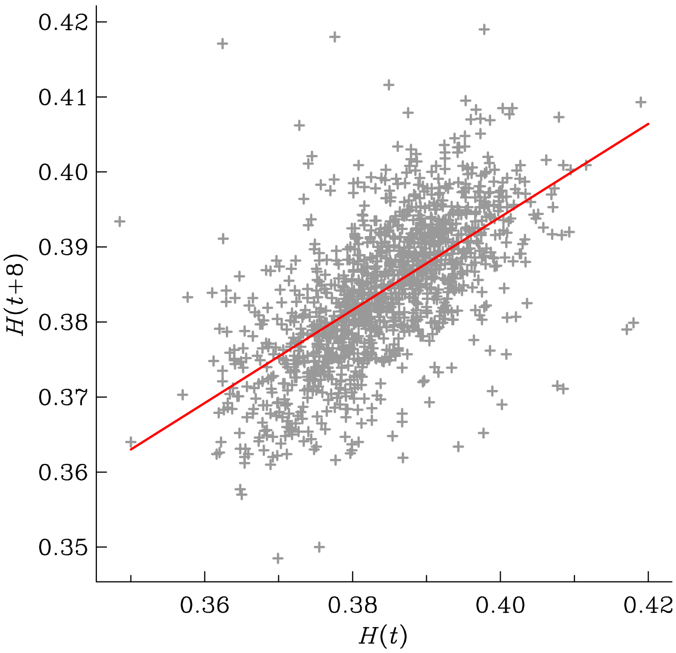
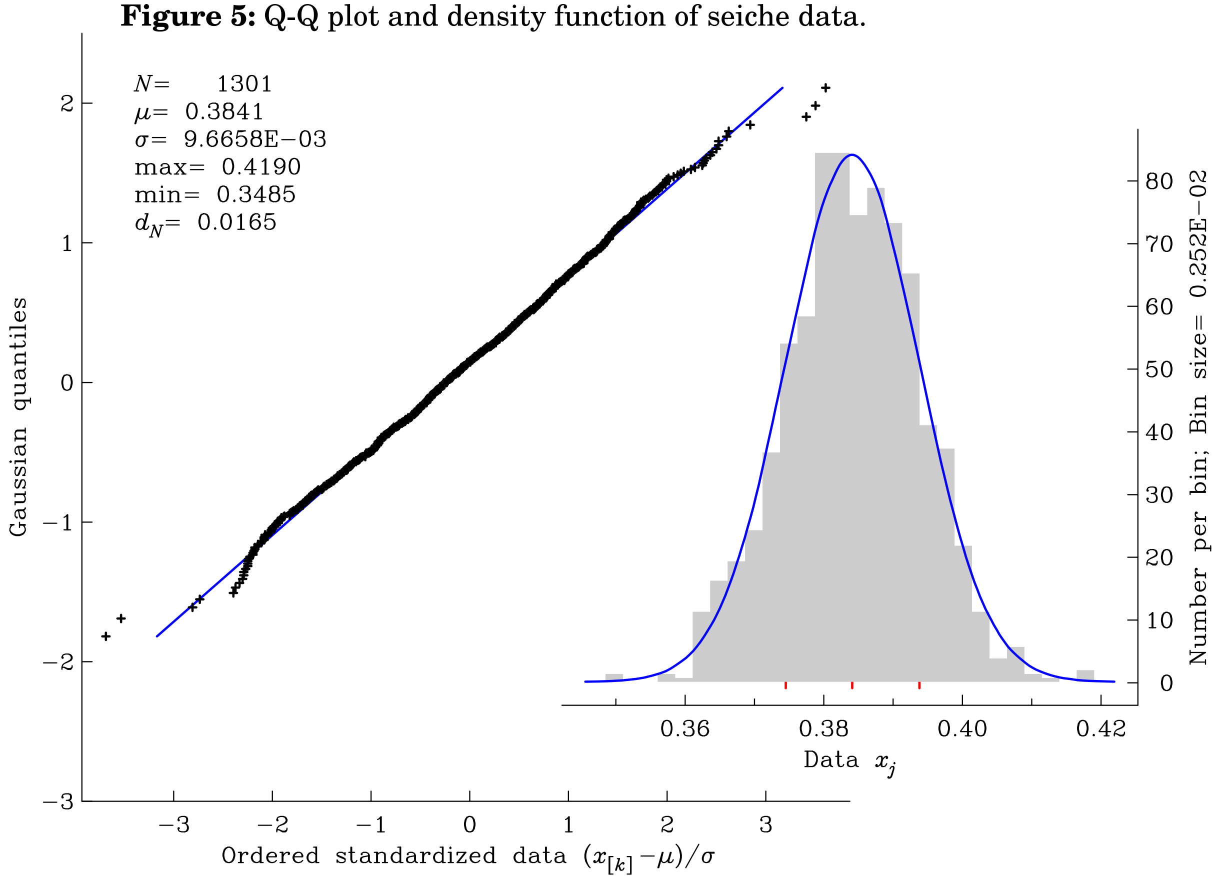


Figure 5: Q-Q plot and density function of seiche data.



The second geophysical data sequence is a series of values in space, the magnetic field measured on by a high flying aircraft (7,000 meters altitude) over the south-eastern Pacific Ocean. In Figure 6 I show only the vertical component Z, and the horizontal, along flight-path component X. Both components are plotted after removal of a standard geomagnetic model, so that their mean values should be nearly zero. The spacing of the sampling is uniform at 3.5 km. Again, as with the lake level data, we see an irregular line with a certain amount of order. A stochastic process seems like a good model, but here there seems to be little evidence of a regular oscillation, or even an irregular one. A feature to notice here is that the two components appear to be related, varying together in some not very obvious fashion: there is a phase lag and perhaps a suggestion of differentiation of Z to obtain X. We will discuss later how to look at pairs of time series for evidence of common variability of this kind.

Concentrating for the moment on the Z component, let us look at an estimate of the autocorrelation function, shown in Figure 10. I will not describe yet how that estimate was made because that will be the subject of a later lecture. For now notice how the R_Z dies away monotonically from one. This means that neighboring values are likely to be very similar indeed, but as one separates two samples in space, their correlation fades away so that by a lag of 30 samples, they are uncorrelated. It is easy to believe this series could be generated by smoothing white noise, that is, by applying a suitable FIR filter to white noise.

The Q-Q plot shows something I mentioned earlier. The magnetic anomaly cumulative plot does not follow the Gaussian model very well: it is rather asymmetric with a large positive tail and a compressed lower tail. The K-S test says that random Gaussian variables would generate such a large value for d_N only 18 percent of the time. This is not a resounding rejection of the Gaussian model, but tells us we should be suspicious of it. The reason for this commonly observed behavior is not understood.

Figure 6: Magnetic anomaly profile over the eastern Pacific Ocean.

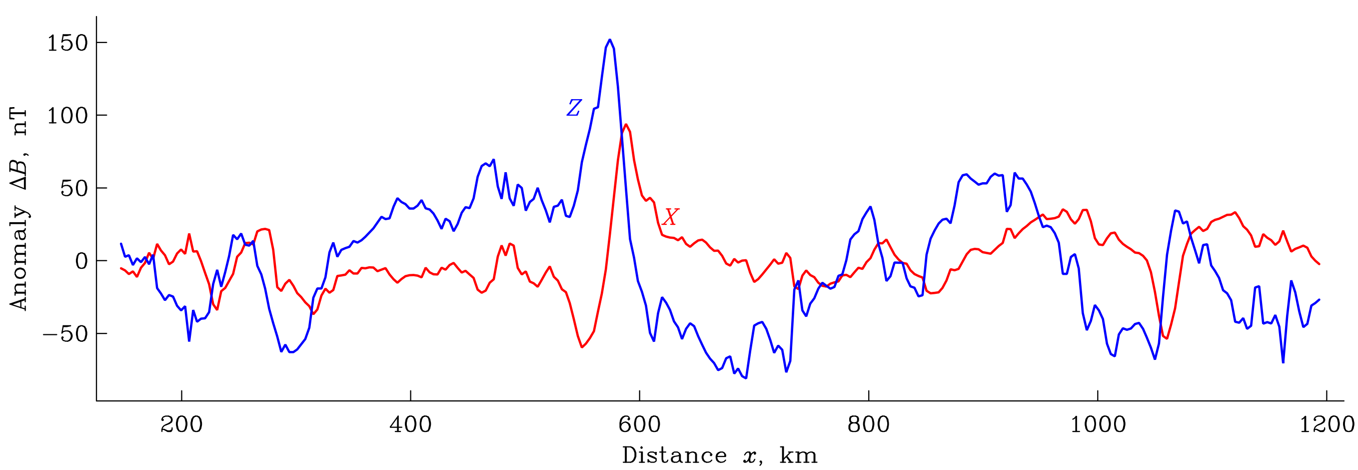


Figure 7: Estimated autocorrelation function for magnetic component Z.

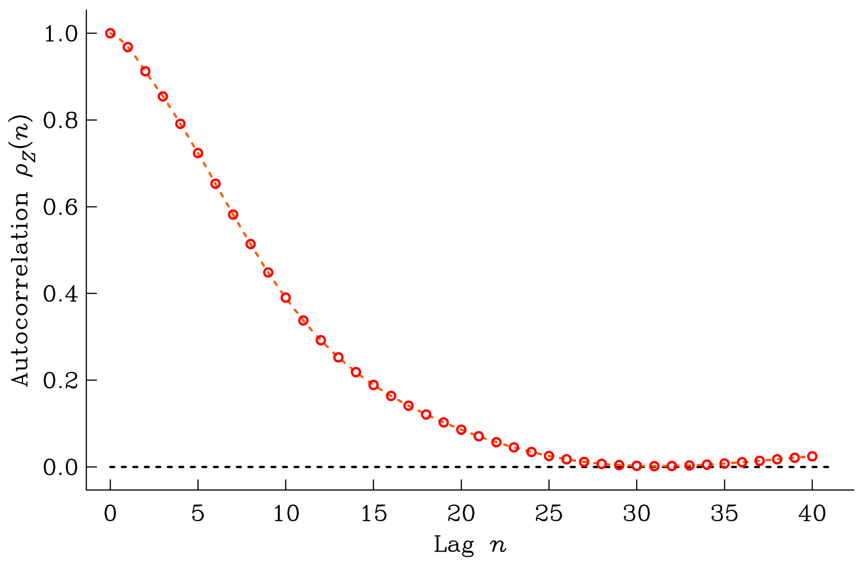
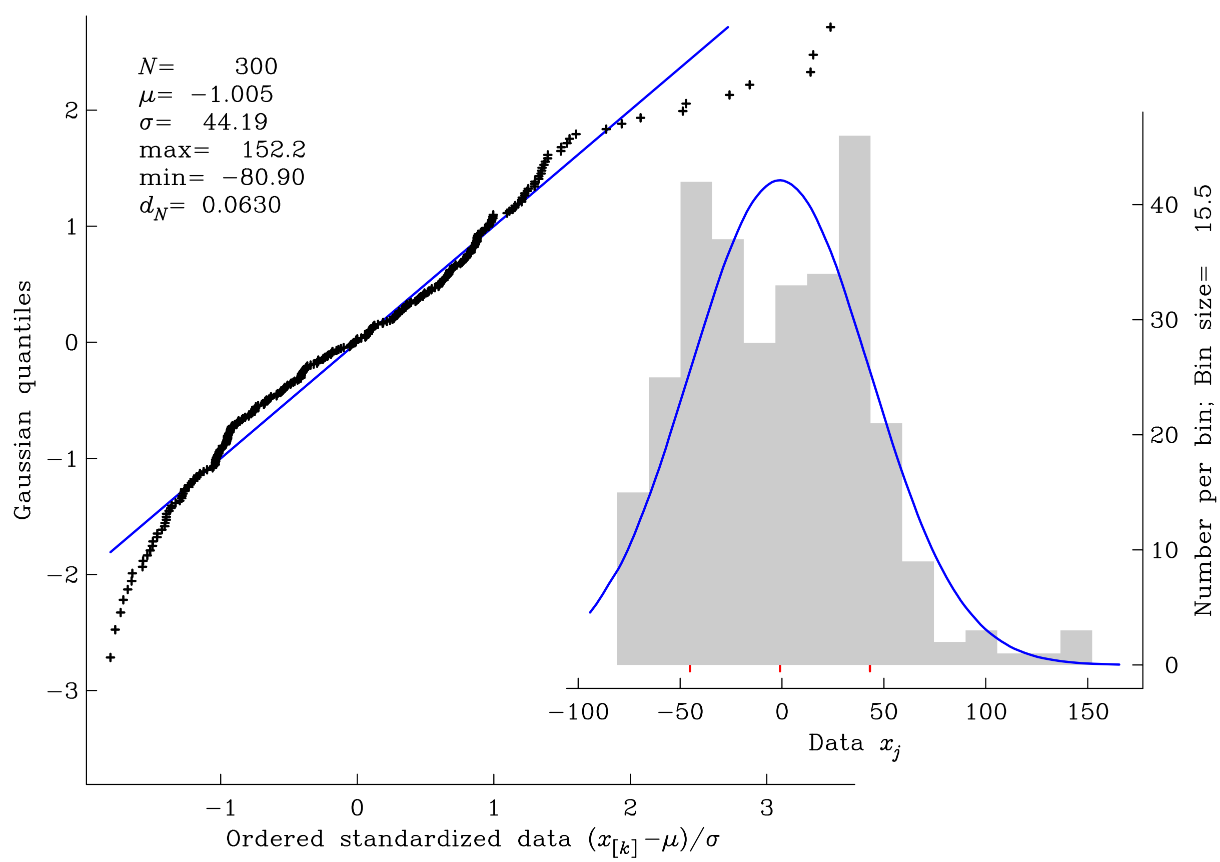
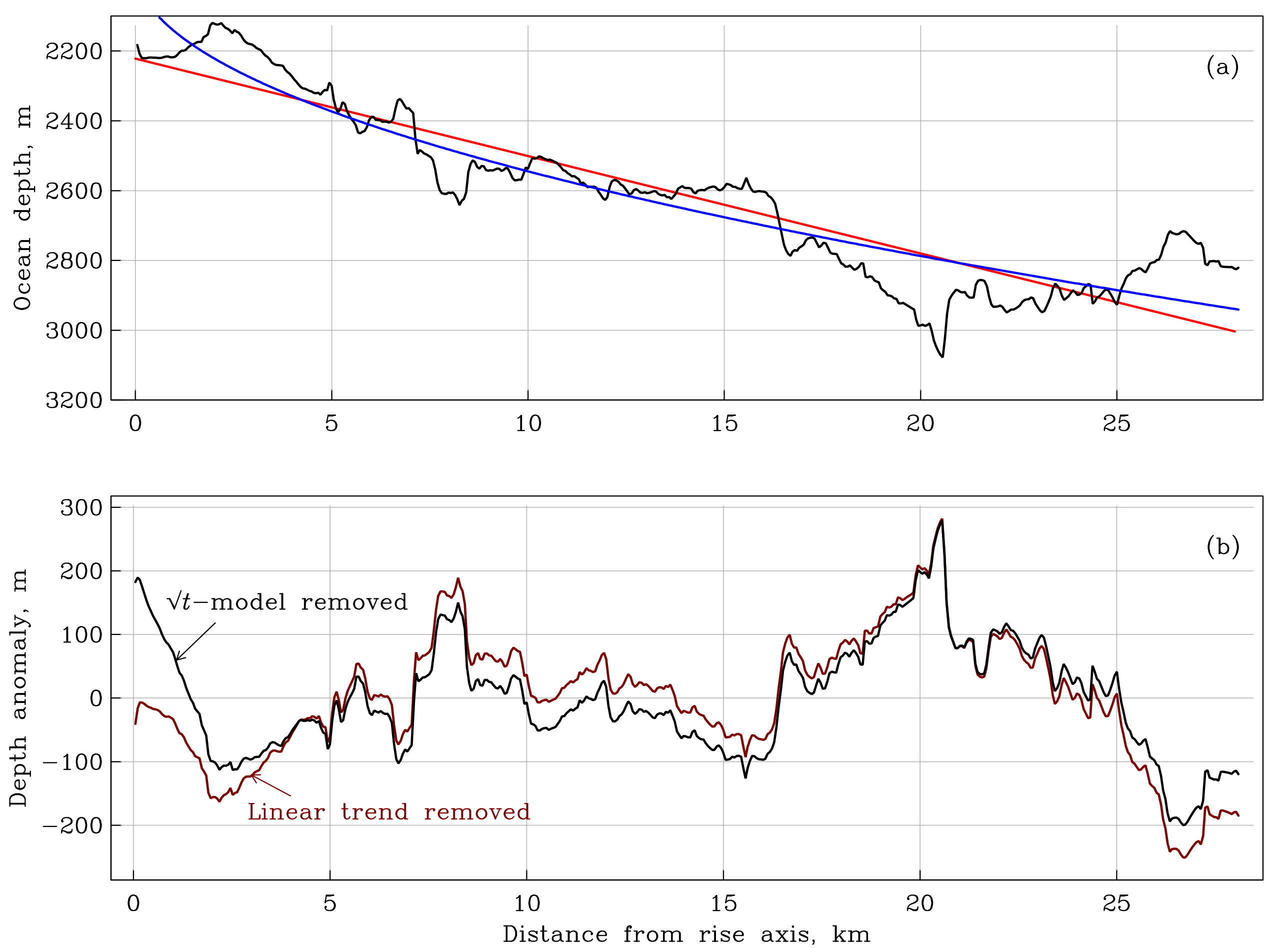


Figure 8: Q-Q plot and density function of Z data.



As a final example I show in Figure 9(a) a bathymetry profile across the East Pacific Rise, an example of a spatial series, which could be a true time series if I had calculated the age of seafloor from the spreading rate. Here the depth signal is very obviously not stationary, because as we expect from marine geology, the seafloor becomes deeper with age. As a realization of a stationary process, the depth curve is a miserable failure because the mean value is not independent of t. But if we remove an average trend, (the line shown), we get a much more satisfactory-looking approximation to a stationary series, as we see from the gray line in Figure 9(b). Now the geologists among the group will, know I should not have removed a straight line but a parabola, because of the famous $\sqrt{t}$ approximation. The least-squares best-fitting age curve is shown also, and it fits the observations slightly better than the straight line. The residual is the black curve in Figure 9(b).

Figure 9: EPR bathymetry, and a depth anomaly.



What we have done is to model the observations as a steadily evolving part, plus a random, stationary part. Unusually in this example, we have a good model for the evolving piece of the model; normally we would just take a straight-line trend to represent that, and as you can see in this case the difference is not very large.

The statistics here are not quite Gaussian again – the distribution is heavy tailed, but not at a high level of significance according to K-S.