# CHAPTER CONTENTS

*Page*

# CHAPTER 2. SAMPLING METEOROLOGICAL VARIABLES

## 2.1 GENERAL

The purpose of this chapter is to give an introduction to this complex subject, for non-experts who need enough knowledge to develop a general understanding of the issues and to acquire a perspective of the importance of the techniques.

Atmospheric variables such as wind speed, temperature, pressure and humidity are functions of four dimensions – two horizontal, one vertical, and one temporal. They vary irregularly in all four, and the purpose of the study of sampling is to define practical measurement procedures to obtain representative observations with acceptable uncertainties in the estimations of mean and variability.

Discussion of sampling in the horizontal dimensions includes the topic of areal representativeness, which is discussed in Part I, Chapter 1, in other chapters on measurements of particular quantities, and briefly below. It also includes the topics of network design, which is a special study related to numerical analysis, and of measurements of area-integrated quantities using radar and satellites; neither of these is discussed here. Sampling in the vertical is briefly discussed in Part I, Chapters 12 and 13 and Part II, Chapter 5. This chapter is therefore concerned only with sampling in time, except for some general comments about representativeness.

The topic can be addressed at two levels as follows:

(a) At an elementary level, the basic meteorological problem of obtaining a mean value of a fluctuating quantity representative of a stated sampling interval at a given time, using instrument systems with long response times compared with the fluctuations, can be discussed. At the simplest level, this involves consideration of the statistics of a set of measurements, and of the response time of instruments and electronic circuits;

(b) The problem can be considered more precisely by making use of the theory of time-series analysis, the concept of the spectrum of fluctuations, and the behaviour of filters. These topics are necessary for the more complex problem of using relatively fast-response instruments to obtain satisfactory measurements of the mean or the spectrum of a rapidly varying quantity, wind being the prime example.

It is therefore convenient to begin with a discussion of time series, spectra and filters in sections 2.2 and 2.3. Section 2.4 gives practical advice on sampling. The discussion here, for the most part, assumes digital techniques and automatic processing.

It is important to recognize that an atmospheric variable is actually never sampled. It is only possible to come as close as possible to sampling the output of a sensor of that variable. The distinction is important because sensors do not create an exact analogue of the sensed variable. In general, sensors respond more slowly than the atmosphere changes, and they add noise. Sensors also do other, usually undesirable, things such as drift in calibration, respond non-linearly, interfere with the quantity that they are measuring, fail more often than intended, and so on, but this discussion will only be concerned with response and the addition of noise.

There are many textbooks available to give the necessary background for the design of sampling systems or the study of sampled data. See, for example, Bendat and Piersol (1986) or Otnes and Enochson (1978). Other useful texts include Pasquill and Smith (1983), Stearns and Hush (1990), Kulhánek (1976), and Jenkins and Watts (1968).

### 2.1.1 Definitions

For the purposes of this chapter the following definitions are used:

*Sampling* is the process of obtaining a discrete sequence of measurements of a quantity.

A *sample* is a single measurement, typically one of a series of spot readings of a sensor system. Note that this differs from the usual meaning in statistics of a set of numbers or measurements which is part of a population.

An *observation* is the result of the sampling process, being the quantity reported or recorded (often also called a measurement). In the context of time-series analysis, an observation is derived from a number of samples.

The ISO definition of a *measurement* is a "set of operations having the object of determining the value of a quantity". In common usage, the term may be used to mean the value of either a sample or an observation.

The *sampling time* or *observation period* is the length of the time over which one observation is made, during which a number of individual samples are taken.

The *sampling interval* is the time between successive observations.

The *sampling function* or *weighting function* is, in its simplest definition, an algorithm for averaging or filtering the individual samples.

The *sampling frequency* is the frequency at which samples are taken. The *sample spacing* is the time between samples.

*Smoothing* is the process of attenuating the high frequency components of the spectrum without significantly affecting the lower frequencies. This is usually done to remove noise (random errors and fluctuations not relevant for the application).

A *filter* is a device for attenuating or selecting any chosen frequencies. Smoothing is performed by a *low-pass* filter, and the terms *smoothing* and *filtering* are often used interchangeably in this sense. However, there are also *high-pass* and *band-pass* filters. Filtering may be a property of the instrument, such as inertia, or it may be performed electronically or numerically.

### 2.1.2 Representativeness in time and space

Sampled observations are made at a limited rate and for a limited time interval over a limited area. In practice, observations should be designed to be sufficiently frequent to be representative of the unsampled parts of the (continuous) variable, and are often taken as being representative of a longer time interval and larger area.

The user of an observation expects it to be representative, or typical, of an area and time, and of an interval of time. This area, for example, may be "the airport" or that area within a radius of several kilometres and within easy view of a human observer. The time is the time at which the report was made or the message transmitted, and the interval is an agreed quantity, often 1, 2 or 10 min.

To make observations representative, sensors are exposed at standard heights and at unobstructed locations and samples are processed to obtain mean values. In a few cases, sensors, for example transmissometers, inherently average spatially, and this contributes to the representativeness of the observation. The human observation of visibility is another example of this. However, the remaining discussion in this chapter will ignore spatial sampling and concentrate upon time sampling of measurements taken at a point.

A typical example of sampling and time averaging is the measurement of temperature each minute (the samples), the computation of a 10 min average (the sampling interval and the sampling function), and the transmission of this average (the observation) in a synoptic report every 3 h. When these observations are collected over a period from the same site, they themselves become samples in a new time sequence with a 3 h spacing. When collected from a large number of sites, these observations also become samples in a spatial sequence. In this sense, representative observations are also representative samples. In this chapter we discuss the initial observation.

### 2.1.3 The spectra of atmospheric quantities

By applying the mathematical operation known as the Fourier transform, an irregular function of time (or distance) can be reduced to its spectrum, which is the sum of a large number of sinusoids, each with its own amplitude, wavelength (or period or frequency) and phase. In broad contexts, these wavelengths (or frequencies) define "scales" or "scales of motion" of the atmosphere.

The range of these scales is limited in the atmosphere. At one end of the spectrum, horizontal scales cannot exceed the circumference of the Earth or about 40 000 km. For meteorological purposes, vertical scales do not exceed a few tens of kilometres. In the time dimension, however, the longest scales are climatological and, in principle, unbounded, but in practice the longest period does not exceed the length of records. At the short end, the viscous dissipation of turbulent energy into heat sets a lower bound. Close to the surface of the Earth, this bound is at a wavelength of a few centimetres and increases with height to a few metres in the stratosphere. In the time dimension, these wavelengths correspond to frequencies of tens of hertz. It is correct to say that atmospheric variables are bandwidth limited.

Figure 2.1 is a schematic representation of a spectrum of a meteorological quantity such as wind, notionally measured at a particular station and time. The ordinate, commonly called energy or spectral density, is related to the variance of the fluctuations of wind at each frequency $n$. The spectrum in Figure 2.1 has a minimum of energy at the mesoscale around one cycle per hour, between peaks in the synoptic scale around one cycle per four days, and in the microscale around one cycle per minute. The smallest wavelengths are a few centimetres and the largest frequencies are tens of hertz.

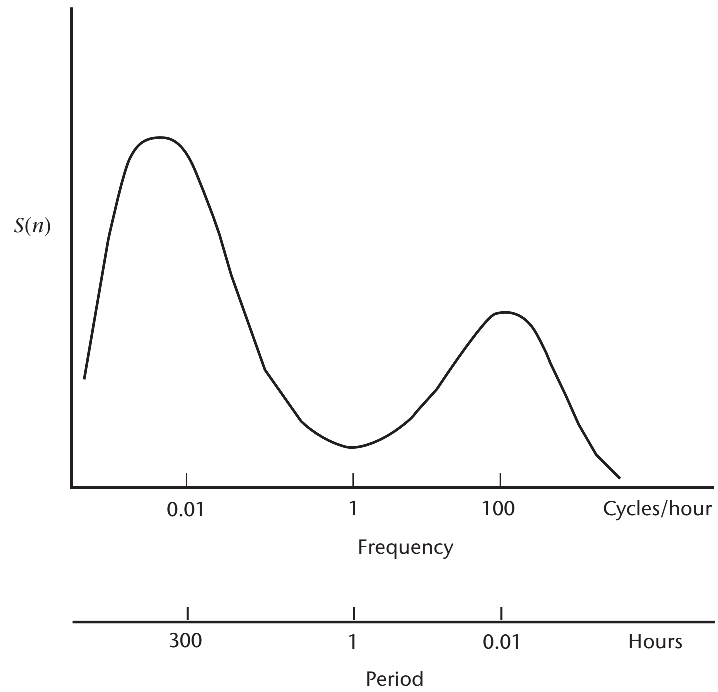


**Figure 2.1. A typical spectrum of a meteorological quantity**

## 2.2 TIME SERIES, POWER SPECTRA AND FILTERS

This section is a layperson's introduction to the concepts of time-series analysis which are the basis for good practice in sampling. In the context of this Guide, they are particularly important for the measurement of wind, but the same problems arise for temperature, pressure and other quantities. They became important for routine meteorological measurements when automatic measurements were introduced, because frequent fast sampling then became possible. Serious errors can occur in the estimates of the mean, the extremes and the spectrum if systems are not designed correctly.

Although measurements of spectra are non-routine, they have many applications. The spectrum of wind is important in engineering, atmospheric dispersion, diffusion and dynamics. The concepts discussed here are also used for quantitative analysis of satellite data (in the horizontal space dimension) and in climatology and micro-meteorology.

In summary, the argument is as follows:

(a) An optimum sampling rate can be assessed from consideration of the variability of the quantity being measured. Estimates of the mean and other statistics of the observations will have smaller uncertainties with higher sampling frequencies, namely, larger samples;

(b) The Nyquist theorem states that a continuous fluctuating quantity can be precisely determined by a series of equispaced samples if they are sufficiently close together;

(c) If the sampling frequency is too low, fluctuations at the higher unsampled frequencies (above the Nyquist frequency, defined in section 2.2.1) will affect the estimate of the mean value. They will also affect the computation of the lower frequencies, and the measured spectrum will be incorrect. This is known as aliasing. It can cause serious errors if it is not understood and allowed for in the system design;

(d) Aliasing may be avoided by using a high sampling frequency or by filtering so that a lower, more convenient sampling frequency can be used;

(e) Filters may be digital or analogue. A sensor with a suitably long response time acts as a filter.

A full understanding of sampling involves knowledge of power spectra, the Nyquist theorem, filtering and instrument response. This is a highly specialized subject, requiring understanding of the characteristics of the sensors used, the way the output of the sensors is conditioned, processed and logged, the physical properties of the elements being measured, and the purpose to which the analysed data are to be put. This, in turn, may require expertise in the physics of the instruments, the theory of electronic or other systems used in conditioning and logging processes, mathematics, statistics and the meteorology of the phenomena, all of which are well beyond the scope of this chapter.

However, it is possible for a non-expert to understand the principles of good practice in measuring means and extremes, and to appreciate the problems associated with measurements of spectra.

### 2.2.1 Time-series analysis

It is necessary to consider signals as being either in the time or the frequency domain. The fundamental idea behind spectral analysis is the concept of Fourier transforms. A function, $f(t)$, defined between $t = 0$ and $t = \tau$ can be transformed into the sum of a set of sinusoidal functions:

$$f(t) = \sum_{j=0}^{\infty} \left[ A_j \sin(j\omega t) + B_j \cos(j\omega t) \right] \tag{2.1}$$

where $\omega = 2\pi/\tau$. The right-hand side of the equation is a Fourier series. $A_j$ and $B_j$ are the amplitudes of the contributions of the components at frequencies $n_j = j\omega$. This is the basic transformation between the time and frequency domains. The Fourier coefficients $A_j$ and $B_j$ relate directly to the frequency $j\omega$ and can be associated with the spectral contributions to $f(t)$ at these frequencies. If the frequency response of an instrument is known – that is, the way in which it amplifies or attenuates certain frequencies – and if it is also known how these frequencies contribute to the original signal, the effect of the frequency response on the output signal can be calculated. The contribution of each frequency is characterized by two parameters. These can be most conveniently taken as the amplitude and phase of the frequency component. Thus, if equation 2.1 is expressed in its alternative form:

$$f(t) = \sum_{j=0}^{\infty} \alpha_j \sin\left(j\omega t + \phi_j\right) \tag{2.2}$$

the amplitude and phase associated with each spectral contribution are $\alpha_j$ and $\phi_j$. Both can be affected in sampling and processing.

So far, it has been assumed that the function $f(t)$ is known continuously throughout its range $t = 0$ to $t = \tau$. In fact, in most examples this is not the case; the meteorological variable is measured at discrete points in a time series, which is a series of $N$ samples equally spaced $\Delta t$ apart during a specified period $\tau = (N–1)\Delta t$. The samples are assumed to be taken instantaneously, an assumption which is strictly not true, as all measuring devices require some time to determine the value they are measuring. In most cases, this is short compared with the sample spacing $\Delta t$. Even if it is not, the response time of the measuring system can be accommodated in the analysis, although that will not be addressed here.

When considering the data that would be obtained by sampling a sinusoidal function at times $\Delta t$ apart, it can be seen that the highest frequency that can be detected is $1/(2\Delta t)$, and that in fact any higher frequency sinusoid that may be present in the time series is represented in the data as having a lower frequency. The frequency $1/(2\Delta t)$ is called the Nyquist frequency, designated here as $n_y$. The Nyquist frequency is sometimes called the folding frequency. This terminology comes from consideration of aliasing of the data. The concept is shown schematically in Figure 2.2. When a spectral analysis of a time series is made, because of the discrete nature of the data, the contribution to the estimate at frequency $n$ also contains contributions from higher frequencies, namely from $2\,jn_y \pm n$ ($j = 1$ to $\infty$). One way of visualizing this is to consider the frequency domain as if it were folded, in a concertina-like way, at $n = 0$ and $n = n_y$ and so on in steps of $n_y$. The spectral estimate at each frequency in the range is the sum of all the contributions of those higher frequencies that overlie it.

The practical effects of aliasing are discussed in section 2.4.2. It is potentially a serious problem and should be considered when designing instrument systems. It can be avoided by minimizing, or reducing to zero, the strength of the signal at frequencies above $n_y$. There are a couple of ways of achieving this. First, the system can contain a low-pass filter that attenuates contributions at frequencies higher than $n_y$ before the signal is digitized. The only disadvantage of this approach is that the timing and magnitude of rapid changes will not be recorded well, or even at all. The second approach is to have $\Delta t$ small enough so that the contributions above the Nyquist frequency are insignificant. This is possible because the spectra of most meteorological variables fall off very rapidly at very high frequencies. This second approach will, however, not always be practicable, as in the example of three-hourly temperature measurements, where if $\Delta t$ is of the order of hours, small scale fluctuations, of the order of minutes or seconds, may have relatively large spectral ordinates and alias strongly. In this case, the first method may be appropriate.

### 2.2.2 **Measurement of spectra**

The spectral density, at least as it is estimated from a time series, is defined as:

$$S\left(n_j\right) = \left(A_j^2 + B_j^2\right) \big/ n_y = \alpha_j^2 \big/ n_y \tag{2.3}$$

It will be noted that phase is not relevant in this case.

The spectrum of a fluctuating quantity can be measured in a number of ways. In electrical engineering it was often determined in the past by passing the signal through band-pass filters and by measuring the power output. This was then related to the power of the central frequency of the filter.

There are a number of ways of approaching the numerical spectral analysis of a time series. The most obvious is a direct Fourier transform of the time series. In this case, as the series is only of finite length, there will be only a finite number of frequency components in the transformation. If there are $N$ terms in the time series, there will be $N/2$ frequencies resulting from this analysis. A direct calculation is very laborious, and other methods have been developed. The first development was by Blackman and Tukey (1958), who related the auto-correlation function to estimates of various spectral functions. (The auto-correlation function $r(t)$ is the correlation coefficient calculated between terms in the time series separated by a time interval $t$). This was appropriate for the low-powered computing facilities of the 1950s and 1960s, but it has now been generally superseded by the so-called fast Fourier transform (FFT), which takes advantage of the general properties of a digital computer to greatly accelerate the calculations. The main limitation of the method is that the time series must contain $2^k$ terms, where $k$ is an integer. In general, this is not a serious problem, as in most instances there are sufficient data to conveniently organize the series to such a length. Alternatively, some FFT computer programs can use an arbitrary number of terms and add synthetic data to make them up to $2^k$.

As the time series is of finite duration ($N$ terms), it represents only a sample of the signal of interest. Thus, the Fourier coefficients are only an estimate of the true, or population, value.

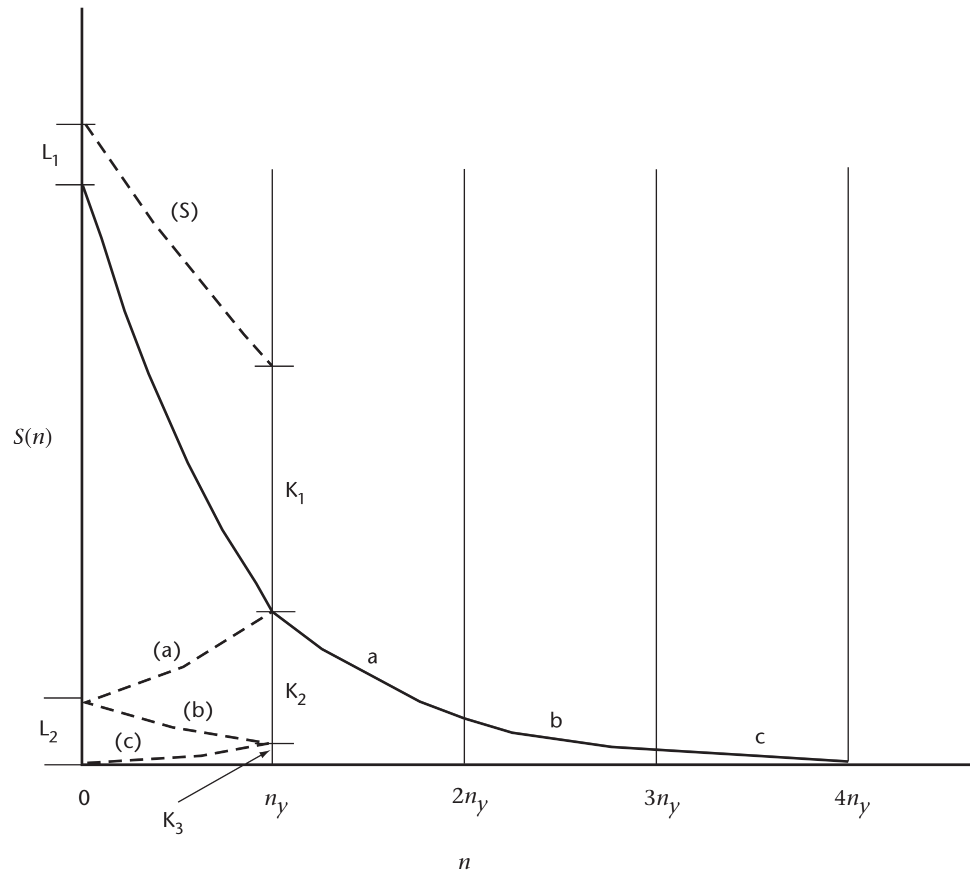


**Figure 2.2. A schematic illustration of aliasing a spectrum computed from a stationary time-series. The spectrum can be calculated only over the frequency range zero to the Nyquist frequency $n_y$. The true values of the energies at higher frequencies are shown by the sectors marked a, b and c. These are "folded" back to the $n = 0$ to $n_y$ sector as shown by the broken lines (a), (b), (c). The computed spectrum, shown by the bold broken line (S), includes the sum of these.**

To improve reliability, it is common practice to average a number of terms each side of a particular frequency and to assign this average to the value of that frequency. The confidence interval of the estimate is thereby shrunk. As a rule of thumb, 30 degrees of freedom is suggested as a satisfactory number for practical purposes. Therefore, as each estimate made during the Fourier transform has 2 degrees of freedom (associated with the coefficients of the sine and cosine terms), about 15 terms are usually averaged. Note that 16 is a better number if an FFT approach is used as this is $2^4$ and there are then exactly $2^k/2^4$ (= $2^{k-4}$) spectral estimates; for example, if there are 1 024 terms in the time series(so $k = 10$), there will be 512 estimates of the $A$s and $B$s, and 64 (= $2^{10-4}$) smoothed estimates.

Increasingly, the use of the above analyses is an integral part of meteorological systems and relevant not only to the analysis of data. The exact form of spectra encountered in meteorology can show a wide range of shapes. As can be imagined, the contributions can be from the lowest frequencies associated with climate change through annual and seasonal contributions through synoptic events with periods of days, to diurnal and semi-diurnal contributions and local mesoscale events down to turbulence and molecular variations. For most meteorological applications, including synoptic analysis, the interest is in the range minutes to seconds. The spectrum at these frequencies will typically decrease very rapidly with frequency. For periods of less than 1 min, the spectrum often takes values proportional to $n^{-5/3}$. Thus, there is often relatively little contribution from frequencies greater than 1 Hz.

One of the important properties of the spectrum is that:

$$\sum_{j=0}^{\infty} S\left(n_j\right) = \sigma^2 \tag{2.4}$$

where $\sigma^2$ is the variance of the quantity being measured. It is often convenient, for analysis, to express the spectrum in continuous form, so that equation 2.4 becomes:

$$\int_0^{\infty} S\left(n\right) dn = \sigma^2 \tag{2.5}$$

It can be seen from equations 2.4 and 2.5 that changes caused to the spectrum, say by the instrument system, will alter the value of $\sigma^2$ and hence the statistical properties of the output relative to the input. This can be an important consideration in instrument design and data analysis.

Note also that the left-hand side of equation 2.5 is the area under the curve in Figure 2.2. That area, and therefore the variance, is not changed by aliasing if the time series is stationary, that is if its spectrum does not change from time to time.

### 2.2.3 **Instrument system response**

Sensors, and the electronic circuits that may be used with them comprising an instrument system, have response times and filtering characteristics that affect the observations.

No meteorological instrument system, or any instrumental system for that matter, precisely follows the quantity it is measuring. There is, in general, no simple way of describing the response of a system, although there are some reasonable approximations to them. The simplest can be classified as first and second order responses. This refers to the order of the differential equation that is used to approximate the way the system responds. For a detailed examination of the concepts that follow, there are many references in physics textbooks and the literature (see MacCready and Jex, 1964).

In the first order system, such as a simple sensor or the simplest low-pass filter circuit, the rate of change of the value recorded by the instrument is directly proportional to the difference between the value registered by the instrument and the true value of the variable. Thus, if the true value at time $t$ is $s(t)$ and the value measured by the sensor is $s_0(t)$, the system is described by the first order differential equation:

$$\frac{ds_0(t)}{dt} = \frac{s(t) - s_0(t)}{T_I} \tag{2.6}$$

where $T_I$ is a constant with the dimension of time, characteristic of the system. A first order system's response to a step function is proportional to exp($-t/T_I$), and $T_I$ is observable as the time taken, after a step change, for the system to reach 63% of the final steady reading. Equation 2.6 is valid for many sensors, such as thermometers.

A cup anemometer is a first order instrument, with the special property that $T_I$ is not constant. It varies with wind speed. In fact, the parameter $s_0T_I$ is called the distance constant, because it is nearly constant. As can be seen in this case, equation 2.6 is no longer a simple first order equation as it is now non-linear and consequently presents considerable problems in its solution. A further problem is that $T_I$ also depends on whether the cups are speeding up or slowing down; that is, whether the right-hand side is positive or negative. This arises because the drag coefficient of a cup is lower if the airflow is towards the front rather than towards the back.

The wind vane approximates a second order system because the acceleration of the vane towards the true wind direction is proportional to the displacement of the vane from the true direction. This is, of course, the classical description of an oscillator (for example, a pendulum). Vanes, both naturally and by design, are damped. This occurs because of a resistive force proportional to, and opposed to, its rate of change. Thus, the differential equation describing the vane's action is:

$$\frac{d^2\phi_0(t)}{dt^2} = k_1\left[\phi_0(t) - \phi(t)\right] - k_2\frac{d\phi_0(t)}{dt} \tag{2.7}$$

where $\phi$ is the true wind direction; $\phi_0$ is the direction of the wind vane; and $k_1$ and $k_2$ are constants. The solution to this is a damped oscillation at the natural frequency of the vane (determined by the constant $k_1$). The damping of course is very important; it is controlled by the constant $k_2$. If it is too small, the vane will simply oscillate at the natural frequency; if too great, the vane will not respond to changes in wind direction.

It is instructive to consider how these two systems respond to a step change in their input, as this is an example of the way in which the instruments respond in the real world. Equations 2.6 and 2.7 can be solved analytically for this input. The responses are shown in Figures 2.3 and 2.4. Note how in neither case is the real value of the element measured by the system. Also, the choice of the values of the constants $k_1$ and $k_2$ can have great effect on the outputs.

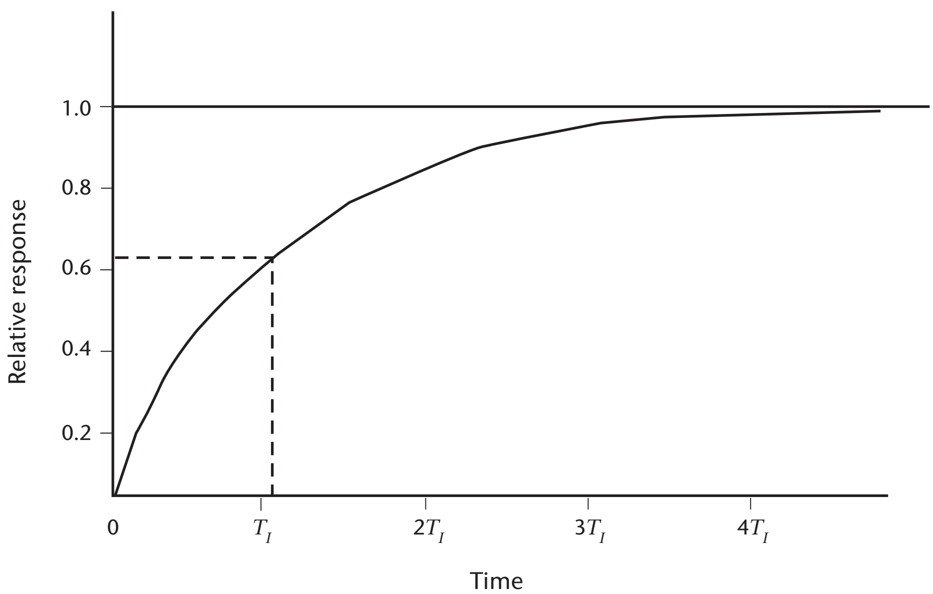


**Figure 2.3. The response of a first order system to a step function. At time $T_I$ the system has reached 63% of its final value.**

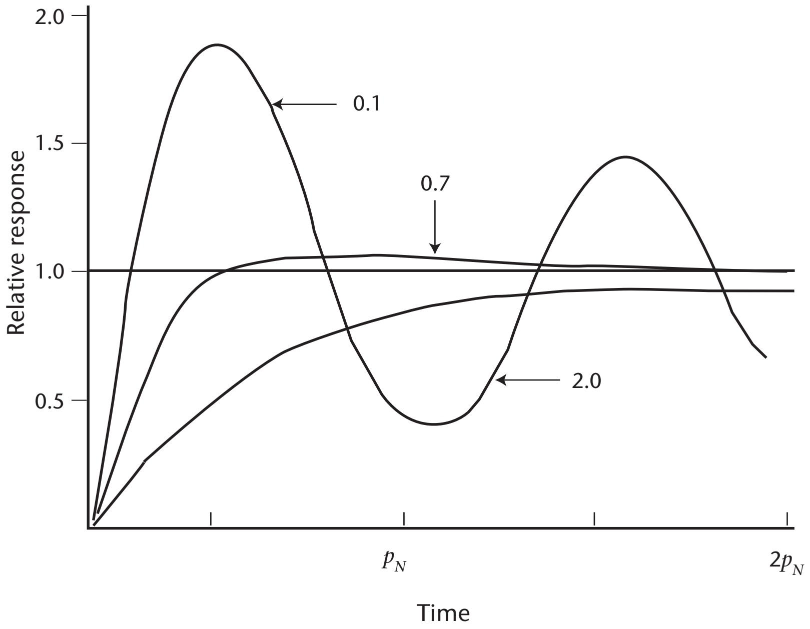


**Figure 2.4. The response of a second order system to a step function. $p_N$ is the natural period, related to $k_1$ in equation 2.7, which, for a wind vane, depends on wind speed. The curves shown are for damping factors with values 0.1 (very lightly damped), 0.7 (critically damped, optimum for most purposes) and 2.0 (heavily damped). The damping factor is related to $k_2$ in equation 2.7.**

An important property of an instrument system is its frequency response function or transfer function $H(n)$. This function gives the amount of the spectrum that is transmitted by the system. It can be defined as:

$$S(n)_{\text{out}} = H(n)S(n)_{\text{in}} \tag{2.8}$$

where the subscripts refer to the input and output spectra. Note that, by virtue of the relationship in equation 2.5, the variance of the output depends on $H(n)$. $H(n)$ defines the effect of the sensor as a filter, as discussed in the next section. The ways in which it can be calculated or measured are discussed in section 2.3.

### 2.2.4 **Filters**

This section discusses the properties of filters, with examples of the ways in which they can affect the data.

Filtering is the processing of a time series (either continuous or discrete, namely, sampled) in such a way that the value assigned at a given time is weighted by the values that occurred at other times. In most cases, these times will be adjacent to the given time. For example, in a discrete time-series of $N$ samples numbered 0 to $N$, with value $y_i$, the value of the filtered observation $\bar{y}_i$ might be defined as:

$$\bar{y}_i = \sum_{j=-m}^{m} w_j \, y_{i+j} \tag{2.9}$$

Here there are $2m + 1$ terms in the filter, numbered by the dummy variable $j$ from $-m$ to $+m$, and $\bar{y}_i$ is centred at $j = 0$. Some data are rejected at the beginning and end of the sampling time. $w_j$ is commonly referred to as a weighting function and typically:

$$\sum_{j=-m}^{m} w_j = 1 \tag{2.10}$$

so that at least the average value of the filtered series will have the same value as the original one.

The above example uses digital filtering. Similar effects can be obtained using electronics (for example, through a resistor and capacitor circuit) or through the characteristics of the sensor (for example, as in the case of the anemometer, discussed earlier). Whether digital or analogue, a filter is characterized by $H(n)$. If digital, $H(n)$ can be calculated; if analogue, it can be obtained by the methods described in section 2.3.

For example, compare a first order system with a response time of $T_I$, and a "box car" filter of length $T_s$ on a discrete time-series taken from a sensor with much faster response. The forms of these two filters are shown in Figure 2.5. In the first, it is as though the instrument has a memory which is strongest at the present instant, but falls off exponentially the further in the past the data goes. The box car filter has all weights of equal magnitude for the period $T_s$, and zero beyond that. The frequency response functions, $H(n)$, for these two are shown in Figure 2.6.

In the figure, the frequencies have been scaled to show the similarity of the two response functions. It shows that an instrument with a response time of, say, 1 s has approximately the same effect on an input as a box car filter applied over 4 s. However, it should be noted that a box car filter, which is computed numerically, does not behave simply. It does not remove all the higher frequencies beyond the Nyquist frequency, and can only be used validly if the spectrum falls off rapidly above $n_y$. Note that the box car filter shown in Figure 2.6 is an analytical solution for $w$ as a continuous function; if the number of samples in the filter is small, the cut-off is less sharp and the unwanted higher frequency peaks are larger.

See Acheson (1968) for practical advice on box car and exponential filtering, and a comparison of their effects.

A response function of a second order system is given in Figure 2.7, for a wind vane in this case, showing how damping acts as a band-pass filter.

It can be seen that the processing of signals by systems can have profound effects on the data output and must be expertly done.

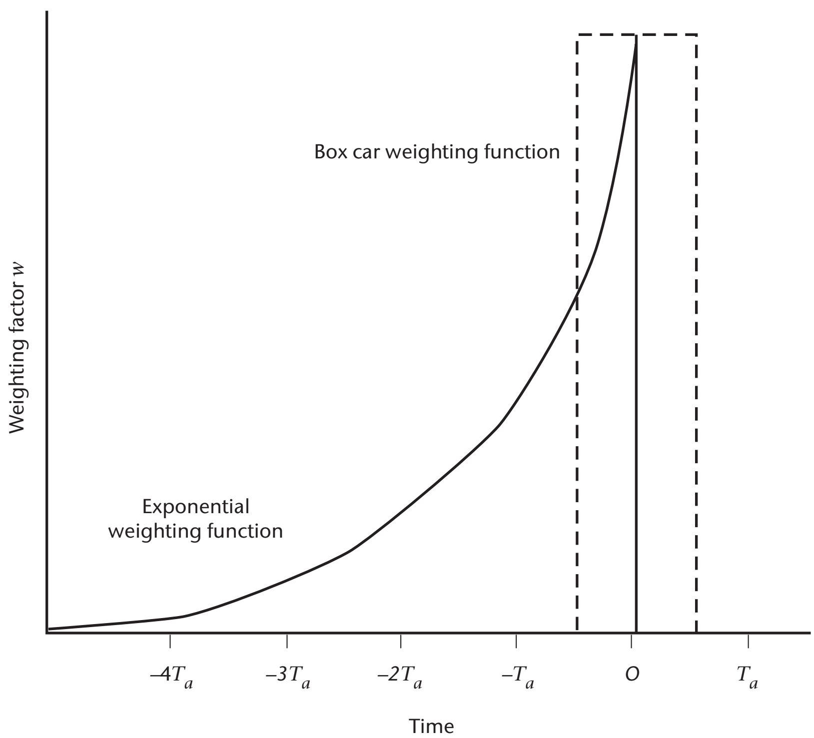


**Figure 2.5. The weighting factors for a first order (exponential) weighting function and a box car weighting function. For the box car $T_a$ is $T_s$, the sampling time, and $w = 1/N$. For the first order function $T_a$ is $T_I$, the time constant of the filter, and $w(t) = (1/T_I) \exp(-t/T_I)$.**

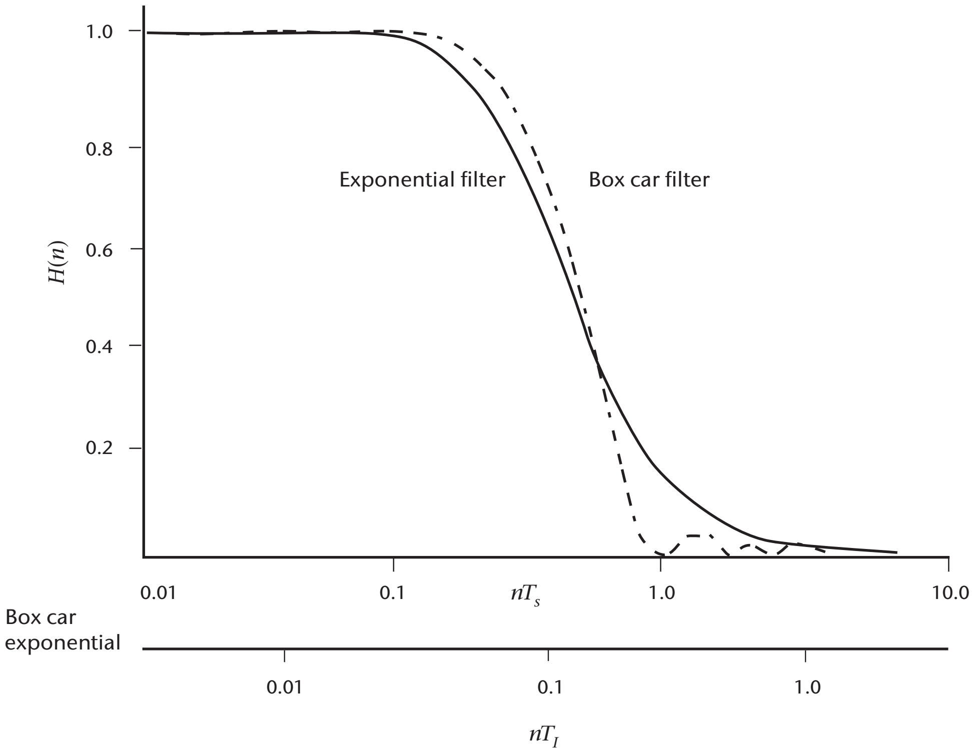


**Figure 2.6. Frequency response functions for a first order (exponential) weighting function and a box car weighting function. The frequency is normalized for the first order filter by $T_I$, the time constant, and for the box car filter by $T_s$, the sampling time.**

Among the effects of filters is the way in which they can change the statistical information of the data. One of these was touched on earlier and illustrated in equations 2.5 and 2.8. Equation 2.5 shows how the integral of the spectrum over all frequencies gives the variance of the time series, while equation 2.8 shows how filtering, by virtue of the effect of the transfer function, will change the measured spectrum. Note that the variance is not always decreased by filtering. For example, in certain cases, for a second order system the transfer function will amplify parts of the spectrum and possibly increase the variance, as shown in Figure 2.7.

To give a further example, if the distribution is Gaussian, the variance is a useful parameter. If it were decreased by filtering, a user of the data would underestimate the departure from the mean of events occurring with given probabilities or return periods.

Also, the design of the digital filter can have unwanted or unexpected effects. If Figure 2.6 is examined it can be seen that the response function for the box car filter has a series of maxima at frequencies above where it first becomes zero. This will give the filtered data a small periodicity at these frequencies. In this case, the effect will be minimal as the maxima are small. However, for some filter designs quite significant maxima can be introduced. As a rule of thumb, the smaller the number of weights, the greater the problem. In some instances, periodicities have been claimed in data that only existed because the data had been filtered.

An issue related to the concept of filters is the length of the sample. This can be illustrated by noting that, if the length of record is of duration $T$, contributions to the variability of the data at frequencies below $1/T$ will not be possible. It can be shown that a finite record length has the effect of a high-pass filter. As for the low-pass filters discussed above, a high-pass filter will also have an impact on the statistics of the output data.

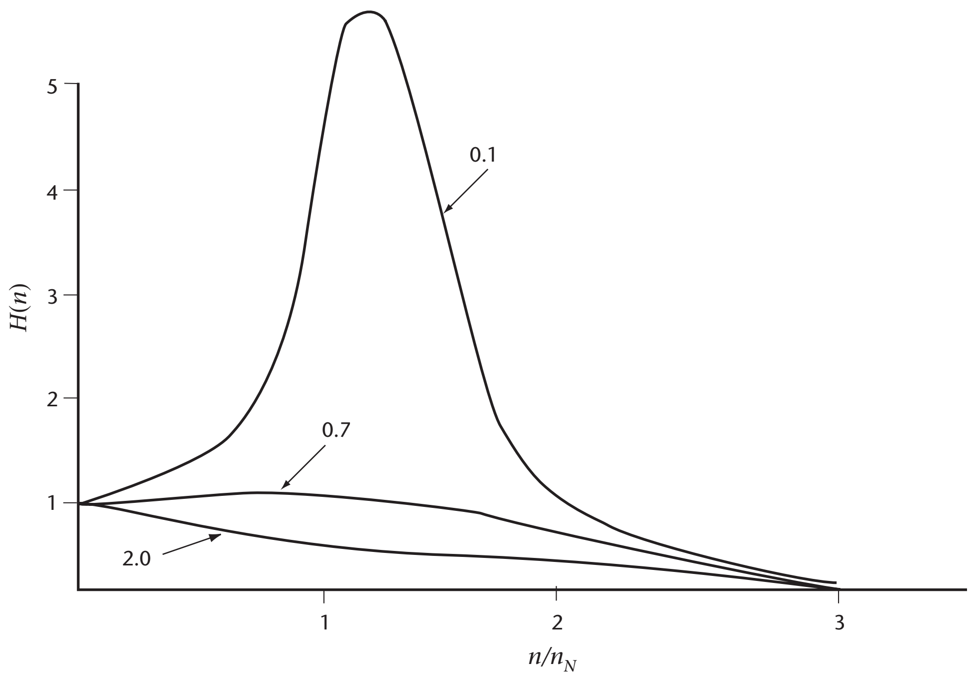


**Figure 2.7. Frequency response functions for a second order system, such as a wind vane. The frequency is normalized by $n_N$, the natural frequency, which depends on wind speed. The curves shown are for damping factors with values 0.1 (very lightly damped), 0.7 (critically damped, optimum for most purposes) and 2.0 (heavily damped).**

## 2.3 DETERMINATION OF SYSTEM CHARACTERISTICS

The filtering characteristics of a sensor or an electronic circuit, or the system that they comprise, must be known to determine the appropriate sampling frequency for the time series that the system produces. The procedure is to measure the transfer or response function $H(n)$ in equation 2.8.

The transfer function can be obtained in at least three ways – by direct measurement, calculation and estimation.

### 2.3.1 Direct measurement of response

Response can be directly measured using at least two methods. In the first method a known change, such as a step function, is applied to the sensor or filter and its response time measured; $H(n)$ can then be calculated. In the second method, the output of the sensor is compared to another, much faster sensor. The first method is more commonly used than the second.

A simple example of how to determine the response of a sensor to a known input is to measure the distance constant of a rotating-cup or propeller anemometer. In this example, the known input is a step function. The anemometer is placed in a constant velocity air-stream, prevented from rotating, then released, and its output recorded. The time taken by the output to increase from zero to 63% of its final or equilibrium speed in the air-stream is the time “constant” (see section 2.2.3).

If another sensor, which responds much more rapidly than the one whose response is to be determined, is available, then good approximations of both the input and output can be measured and compared. The easiest device to use to perform the comparison is probably a modern, two-channel digital spectrum analyser. The output of the fast-response sensor is input to one channel, the output of the sensor being tested to the other channel, and the transfer function automatically displayed. The transfer function is a direct description of the sensor as a filter. If the device whose response is to be determined is an electronic circuit, generating a known or even truly random input is much easier than finding a much faster sensor. Again, a modern, two-channel digital spectrum analyser is probably most convenient, but other electronic test instruments can be used.

### 2.3.2 Calculation of response

This is the approach described in section 2.2.3. If enough is known about the physics of a sensor/filter, the response to a large variety of inputs may be determined by either analytic or numerical solution. Both the response to specific inputs, such as a step function, and the transfer function can be calculated. If the sensor or circuit is linear (described by a linear differential equation), the transfer function is a complete description, in that it describes the amplitude and phase responses as a function of frequency, in other words, as a filter. Considering response as a function of frequency is not always convenient, but the transfer function has a Fourier transform counterpart, the impulse response function, which makes interpretation of response as a function of time much easier. This is illustrated in Figures 2.3 and 2.4 which represent response as a function of time.

If obtainable, analytic solutions are preferable because they clearly show the dependence upon the various parameters.

### 2.3.3 Estimation of response

If the transfer functions of a transducer and each following circuit are known, their product is the transfer function of the entire system. If, as is usually the case, the transfer functions are low-pass filters, the aggregate transfer function is a low-pass filter whose cut-off frequency is less than that of any of the individual filters.

If one of the individual cut-off frequencies is much less than any of the others, then the cut-off frequency of the aggregate is only slightly smaller.

Since the cut-off frequency of a low-pass filter is approximately the inverse of its time constant, it follows that, if one of the individual time constants is much larger than any of the others, the time constant of the aggregate is only slightly larger.

## 2.4 SAMPLING

### 2.4.1 Sampling techniques

Figure 2.8 schematically illustrates a typical sensor and sampling circuit. When exposed to the atmosphere, some property of the transducer changes with an atmospheric variable such as temperature, pressure, wind speed or direction, or humidity and converts that variable into a useful signal, usually electrical. Signal conditioning circuits commonly perform functions such as converting transducer output to a voltage, amplifying, linearizing, offsetting and smoothing. The low-pass filter finalizes the sensor output for the sample-and-hold input. The sample-and-hold and the analogue-to-digital converter produce the samples from which the observation is computed in the processor.

It should be noted that the smoothing performed at the signal conditioning stage for engineering reasons, to remove spikes and to stabilize the electronics, is performed by a low-pass filter; it reduces the response time of the sensor and removes high frequencies which may be of interest. Its effect should be explicitly understood by the designer and user, and its cut-off frequency should be as high as practicable.

So-called "smart sensors", those with microprocessors, may incorporate all the functions shown. The signal conditioning circuitry may not be found in all sensors, or may be combined with other circuitry. In other cases, such as with a rotating-cup or propeller anemometer, it may be easy to speak only of a sensor because it is awkward to distinguish a transducer. In the few cases for which a transducer or sensor output is a signal whose frequency varies with the atmospheric variable being measured, the sample-and-hold and the analogue-to-digital converter may be replaced by a counter. But these are not important details. The important element in the design is to ensure that the sequence of samples adequately represents the significant changes in the atmospheric variable being measured.

The first condition imposed upon the devices shown in Figure 2.8 is that the sensor must respond quickly enough to follow the atmospheric fluctuations which are to be described in the observation. If the observation is to be a 1, 2 or 10 min average, this is not a very demanding requirement. On the other hand, if the observation is to be that of a feature of turbulence, such as peak wind gust, care must be taken when selecting a sensor.

The second condition imposed upon the devices shown in Figure 2.8 is that the sample-and-hold and the analogue-to-digital converter must provide enough samples to make a good observation. The accuracy demanded of meteorological observations usually challenges the sensor, not the electronic sampling technology. However, the sensor and the sampling must be matched to avoid aliasing. If the sampling rate is limited for technical reasons, the sensor/filter system must be designed to remove the frequencies that cannot be represented.

If the sensor has a suitable response function, the low-pass filter may be omitted, included only as insurance, or may be included because it improves the quality of the signal input to the sample-and-hold. As examples, such a filter may be included to eliminate noise pick-up at the end of a long cable or to further smooth the sensor output. Clearly, this circuit must also respond quickly enough to follow the atmospheric fluctuations of interest.

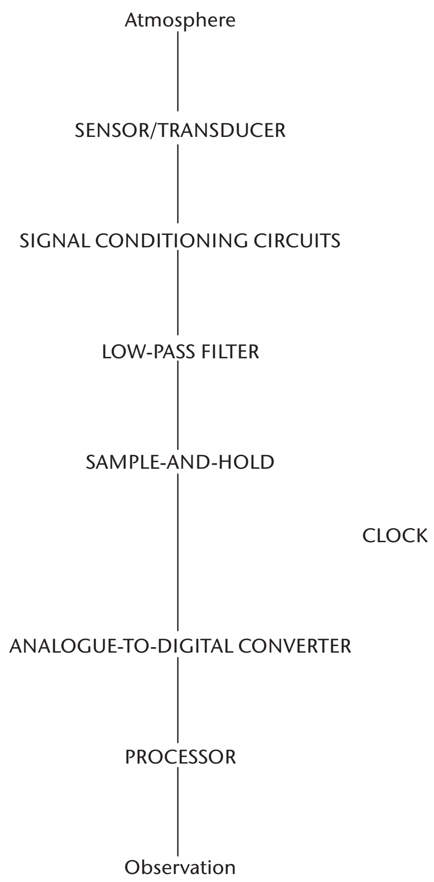


**Figure 2.8. An instrument system**

### 2.4.2 Sampling rates

For most meteorological and climatological applications, observations are required at intervals of 30 min to 24 hours, and each observation is made with a sampling time of the order of 1 to 10 min. Part I, Chapter 1, Annex 1.E gives a recent statement of requirements for these purposes.

A common practice for routine observations is to take one spot reading of the sensor (such as a thermometer) and rely on its time constant to provide an approximately correct sampling time. This amounts to using an exponential filter (Figure 2.6). Automatic weather stations commonly use faster sensors, and several spot readings must be taken and processed to obtain an average (box car filter) or other appropriately weighted mean.

A practical recommended scheme for sampling rates is as follows:[1]

(a) Samples taken to compute averages should be obtained at equispaced time intervals which:

  (i) Do not exceed the time constant of the sensor; or

  (ii) Do not exceed the time constant of an analogue low-pass filter following the linearized output of a fast-response sensor; or

  (iii) Are sufficient in number to ensure that the uncertainty of the average of the samples is reduced to an acceptable level, for example, smaller than the required accuracy of the average;

(b) Samples to be used in estimating extremes of fluctuations, such as wind gusts, should be taken at rates at least four times as often as specified in (i) or (ii) above.

For obtaining averages, somewhat faster sampling rates than (i) and (ii), such as twice per time constant, are often advocated and practised.

Criteria (i) and (ii) derive from consideration of the Nyquist frequency. If the sample spacing $\Delta t \le T_I$, the sampling frequency $n \ge 1/T_I$ and $nT_I \ge 1$. It can be seen from the exponential curve in Figure 2.6 that this removes the higher frequencies and prevents aliasing. If $\Delta t = T_I$, $n_y = 1/2T_I$ and the data will be aliased only by the spectral energy at frequencies at $nT_I = 2$ and beyond, that is where the fluctuations have periods of less than $0.5T_I$.

Criteria (i) and (ii) are used for automatic sampling. The statistical criterion in (iii) is more applicable to the much lower sampling rates in manual observations. The uncertainty of the mean is inversely proportional to the square root of the number of observations, and its value can be determined from the statistics of the quantity.

Criterion (b) emphasizes the need for high sampling frequencies, or more precisely, small time-constants, to measure gusts. Recorded gusts are smoothed by the instrument response, and the recorded maximum will be averaged over several times the time constant.

The effect of aliasing on estimates of the mean can be seen very simply by considering what happens when the frequency of the wave being measured is the same as the sampling frequency, or a multiple thereof. The derived mean will depend on the timing of the sampling. A sample obtained once per day at a fixed time will not provide a good estimate of mean monthly temperature.

For a slightly more complex illustration of aliasing, consider a time series of three-hourly observations of temperature using an ordinary thermometer. If temperature changes smoothly with time, as it usually does, the daily average computed from eight samples is acceptably stable. However, if a mesoscale event (a thunderstorm) has occurred which reduced the temperature by many degrees for 30 min, the computed average is wrong. The reliability of daily averages

[1] As adopted by the Commission for Instruments and Methods of Observation at its tenth session (1989) through Recommendation 3 (CIMO-X).

depends on the usual weakness of the spectrum in the mesoscale and higher frequencies. However, the occurrence of a higher-frequency event (the thunderstorm) aliases the data, affecting the computation of the mean, the standard deviation and other measures of dispersion, and the spectrum.

The matter of sampling rate may be discussed also in terms of Figure 2.8. The argument in section 2.2.1 was that, for the measurement of spectra, the sampling rate, which determines the Nyquist frequency, should be chosen so that the spectrum of fluctuations above the Nyquist frequency is too weak to affect the computed spectrum. This is achieved if the sampling rate set by the clock in Figure 2.8 is at least twice the highest frequency of significant amplitude in the input signal to the sample-and-hold.

The wording “highest frequency of significant amplitude” used above is vague. It is difficult to find a rigorous definition because signals are never truly bandwidth limited. However, it is not difficult to ensure that the amplitude of signal fluctuations decreases rapidly with increasing frequency, and that the root-mean-square amplitude of fluctuations above a given frequency is either small in comparison with the quantization noise of the analogue-to-digital converter, small in comparison with an acceptable error or noise level in the samples, or contributes negligibly to total error or noise in the observation.

Section 2.3 discussed the characteristics of sensors and circuits which can be chosen or adjusted to ensure that the amplitude of signal fluctuations decreases rapidly with increasing frequency. Most transducers, by virtue of their inability to respond to rapid (high-frequency) atmospheric fluctuations and their ability to replicate faithfully slow (low-frequency) changes, are also low-pass filters. By definition, low-pass filters limit the bandwidth and, by Nyquist’s theorem, also limit the sampling rate that is necessary to reproduce the filter output accurately. For example, if there are real variations in the atmosphere with periods down to 100 ms, the Nyquist sampling frequency would be 1 per 50 ms, which is technically demanding. However, if they are seen through a sensor and filter which respond much more slowly, for example with a 10 s time constant, the Nyquist sampling rate would be 1 sample per 5 s, which is much easier and cheaper, and preferable if measurements of the high frequencies are not required.

### 2.4.3 Sampling rate and quality control

Many data quality control techniques of use in automatic weather stations depend upon the temporal consistency, or persistence, of the data for their effectiveness. As a very simple example, two hypothetical quality-control algorithms for pressure measurements at automatic weather stations should be considered. Samples are taken every 10 s, and 1 min averages computed each minute. It is assumed that atmospheric pressure only rarely, if ever, changes at a rate exceeding 1 hPa per minute.

The first algorithm rejects the average if it differs from the previous one by more than 1 hPa. This would not make good use of the available data. It allows a single sample with as much as a 6 hPa error to pass undetected and to introduce a 1 hPa error in an observation.

The second algorithm rejects a sample if it differs from the previous one by more than 1 hPa. In this case, an average contains no error larger than about 0.16 (1/6) hPa. In fact, if the assumption is correct that atmospheric pressure only rarely changes at a rate exceeding 1 hPa per minute, the accept/reject criteria on adjacent samples could be tightened to 0.16 hPa and error in the average could be reduced even more.

The point of the example is that data quality control procedures that depend upon temporal consistency (correlation) for their effectiveness are best applied to data of high temporal resolution (sampling rate). At the high frequency end of the spectrum in the sensor/filter output, correlation between adjacent samples increases with increasing sampling rate until the Nyquist frequency is reached, after which no further increase in correlation occurs.

Up to this point in the discussion, nothing has been said which would discourage using a sensor/filter with a time constant as long as the averaging period required for the observation is taken as a single sample to use as the observation. Although this would be minimal in its demands upon the digital subsystem, there is another consideration needed for effective data quality control. Observations can be grouped into three categories as follows:

(a) Accurate (observations with errors less than or equal to a specified value);

(b) Inaccurate (observations with errors exceeding a specified value);

(c) Missing.

There are two reasons for data quality control, namely, to minimize the number of inaccurate observations and to minimize the number of missing observations. Both purposes are served by ensuring that each observation is computed from a reasonably large number of data quality-controlled samples. In this way, samples with large spurious errors can be isolated and excluded, and the computation can still proceed, uncontaminated by that sample.

## REFERENCES AND FURTHER READING


Acheson, D.T., 1968: An approximation to arithmetic averaging for meteorological variables. *Journal of Applied Meteorology*, 7:548–553.

Bendat, J.S. and A.G. Piersol, 1986: *Random Data: Analysis and Measurement Procedures*. Second edition, John Wiley and Sons, New York.

Blackman, R.B. and J.W. Tukey, 1958: *The Measurement of Power Spectra*. Dover Publications, New York.

Jenkins, G.M. and D.G. Watts, 1968: *Spectral Analysis and its Applications*. Holden-Day, San Francisco.

Kulhánek, O., 1976: *Introduction to Digital Filtering in Geophysics*. Elsevier, Amsterdam.

MacCready, P.B. and H.R. Jex, 1964: Response characteristics and meteorological utilization of propeller and vane wind sensors. *Journal of Applied Meteorology*, 3(2):182–193.

Otnes, R.K. and L. Enochson, 1978: *Applied Time Series Analysis. Volume 1: Basic techniques*. John Wiley and Sons, New York.

Pasquill, F. and F.B. Smith, 1983: *Atmospheric Diffusion*. Ellis Horwood, Chichester.

Stearns, S.D. and D.R. Hush, 1990: *Digital Signal Analysis*. Prentice Hall, New Jersey.