

# Simple procedures for data analysis based on continuous-time Brownian motion

G.K. Robinson

*CSIRO Mathematical and Information Sciences*

This paper discusses techniques for analysis of sequential data from variable processes, particularly ones that can still be used when the data are not equally-spaced in time. The techniques are based on models which use continuous-time Brownian motion or its integrals to describe the pattern of variation in an underlying physical process and use white noise to describe variation due to measurement processes. The paper concentrates on making statements about what has been happening in the region covered by the data rather than on making predictions about regions far from the data, and argues that variograms are a useful way to describe process variability for such purposes. It is argued that the continuous-time integrated Brownian motion plus white noise model is always preferable to its discrete-time analogue, the local linear trend model. Further integrals of Brownian motion can be fitted as an alternative to using splines. All of these continuous-time Brownian motion models can be fitted to data associated with a time interval (interval data) as well as to data associated with a single time (spot data). They can be fitted using mixed model methodology as well as by using Kalman filtering and smoothing.

Keywords: *Integrated Brownian motion; mixed models; statistical process control; variograms; Kalman filtering*

## 1 Introduction

My personal interest in the topic of this paper has arisen largely from consulting work in the mining and mineral processing industries where sequential data are often unequally-spaced and of variable precision. Data-spacing is commonly unequal because samples are collected at equally-spaced times but the mass flow rate is very variable, with occasional complete stoppages. Data precision is often variable because sampling variance changes with particle size distribution. Measurement variance may also be variable because batches of material of different importance are measured differently. Another source of complexity is that the samples are often composite samples that are representative of the material flowing over a time interval, rather than at a point in time.

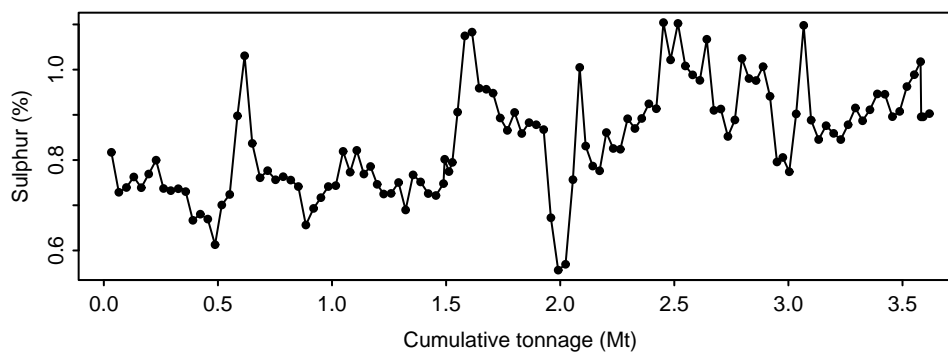


Figure 1: Sulphur grade (percentage sulphur by weight) of composite samples taken from ore entering a metallurgical plant

Figure 1 shows the sulphur grade of some daily composite samples taken from ore used as input to a metallurgical plant. Some other data sets with unequally-spaced or variable-precision data are the following.

- Grades of shipments of iron ore. These data are unequally-spaced, whether the measure of spacing is taken to be time or tonnage. They have unequal precision because different sampling and testing regimes are followed for shipments of different sizes.

- In the wine industry, samples of grapes are tested about once or twice a week in order to predict when the grapes will be ready to be harvested.
- Resistance data from aluminium reduction cells is generally recorded at regular intervals, but its variance can increase because of factors relevant to a particular cell or because of “anode effects” occurring in other cells on the same potline.
- Measurements of water quality at two different sites in a reticulation system might show similar patterns over time, but the displacement in time between the patterns will generally be smaller at times of high water usage than at times of low water usage.

One model which is often useful for data sets like these has two components of variation: continuous-time Brownian motion for the wandering of an underlying process mean and white noise for measurement variation. The sample variograms are approximately linear as a function of the lag, at least for small lags. (See section 6.2 for definition of a population variogram and section 5 for definition of a sample variogram.) Section 2 of this paper reviews methodologies for the fitting of this model. Most of these methodologies are already known, though not as widely known as I believe they ought to be.

Section 3 discusses models involving integrated forms of continuous-time Brownian motion. These models are less well known. They can be fitted using either Kalman filtering and smoothing or mixed model methodology.

Section 4 discusses the differences between interval data and spot data. Applications of these models often ignore uncertainty about the sizes of variance components. Section 5 discusses some issues about estimating variance components.

Section 6 gives some practical advice and value judgements about choosing models and estimating variance components.

## 2 Brownian motion plus white noise

For a Brownian motion process the change over a time interval has variance proportional to the length of that time interval, and changes over non-overlapping time intervals are independent. The increment variance per unit time will be denoted by  $V_B$ .

Brownian motion which is regarded as a function of continuous time but is only observed at equally-spaced times is equivalent to discrete-time Brownian motion. However, the cumulative sum of discrete-time Brownian motion is not equivalent to the integral of continuous-time Brownian motion and discrete-time Brownian motion cannot be fitted as naturally to unequally-spaced data as can continuous-time Brownian motion.

A very simple and particularly useful model is commonly called the “Brownian motion plus white noise model” or the “local level model”. In this model there is an unobservable time-varying mean level for a process; this mean level varies over time like a Brownian motion process with known increment variance per unit time; and observations are normally distributed with mean equal to the mean level for the process at the time of the observation and known variance.

This may be written as a state space model with observation equation

$$y_i = \mu_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma_i^2)$$

and transition or state equation

$$\mu_{i+1} = \mu_i + \eta_i \quad \eta_i \sim N(0, Q_i)$$

where the white noise and Brownian motion disturbances,  $\varepsilon_i$  and  $\eta_i$  are independent. The index  $i$  increases between the time points of interest,  $t_i$ , which are not necessarily equally-spaced. The variance,  $Q_i$  of the Brownian motion disturbances is  $V_B(t_i - t_{i-1})$  if the unobservable mean level behaves like continuous-time Brownian motion with increment variance of  $V_B$  per unit time. Sometimes context-specific information suggests different choices for  $Q_i$ . In some applications the variance of changes in the local level,  $\eta_i$  depends on the tonnage of material passing between data points or on some other measure of the opportunity for

the level to change. For time intervals during which a measurement system has been recalibrated or there has been a dramatic change in the source of material it may be desirable to use large values for  $Q_i$ .

Given observations  $y_i$  taken at times  $t_i$  for  $i = 1, 2, \dots, n$ , the best linear unbiased estimates of the unobservable true states and their estimation precisions can be found using a Kalman filter.

### Example of Kalman filter: Ripening of grapes

Consider modelling the ripening of grapes. Samples of grapes are taken once or twice a week from each of many vineyards. The problem is to predict when grapes will be ready to be harvested. Predictions up to two weeks into the future are of primary interest.

Suppose that grapes are expected to ripen at an average of 0.08 units of Baumé per day, that the white noise standard deviation is 0.4 units, and that the Brownian motion increment variance is 0.0324 per day. For one vineyard the Baumé was measured on days 7, 14, 17 and 21 with the measurements being 8.1, 10, 11.1 and 10.9. What is a reasonable estimate of the Baumé on days 22–31 when harvesting of the grapes might be scheduled?

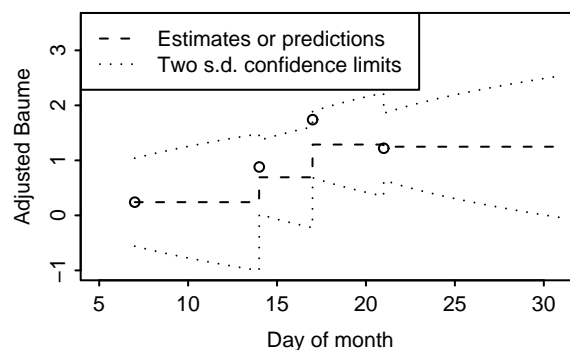


Figure 2: **Estimates and predictions of adjusted Baumé, computed by subtracting  $8 + 0.08 \times \text{Day}$  from the raw Baumé.**

A trend with slope of 0.08 units of Baumé per day is being treated as known, so it is subtracted from the data before the local level model is fitted. The adjusted data is shown as four plotted points on Figure 2.

The Kalman filter estimate of adjusted Baumé using the local level model starts off at day 7 with the same value as the first adjusted data point, no other relevant information being available. It is the best estimate in the least squares sense. The variance of this estimate is equal to the variance of the first data point at that time, but increases at 0.0324 per day until the time of the second data point. At the time of the second data point, the best estimate of the local level changes abruptly. The new estimate is a weighted average of the previous estimate and the new data point, with weights proportional to the precisions of these pieces of information, and the new precision is the sum of the precisions. The Kalman filter consists of repeating this computational process for all adjusted data points in time order.

On Figure 2, the dashed line shows the best estimate of adjusted Baumé as a function of time. The dotted lines show two-standard-deviation confidence limits about that best estimate. Note that after the last data point the best estimate does not change but the variance of the estimate increases linearly with time. The standard deviation increases like the square root of a linear function of time.

Figure 3 shows the same information as Figure 2, but after adding the trend back on to the estimates. It also shows a line fitted by least-squares to the four points. The predictions of such a line are usually quite different from the predictions made by fitting a local level model.

The term “filter” is used for such estimates which use all data up to that point in time but do not take subsequent data into account. Estimates that do take subsequent data into

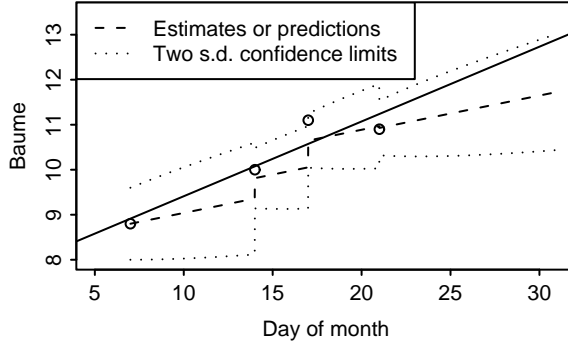


Figure 3: Estimates and predictions of Baumé.

account are referred to as “smoothers”. Smoothers can be computed for state space models using algorithms such as “fixed interval smoothing”. Strictly speaking, such smoothers should not be referred to as “Kalman smoothers”, since they were not known to Kalman, but the term is sometimes used. Many people ascribe the fixed interval smoothing algorithm to Rauch, Tung and Striebel (1965). Others give much credit to Anderson and Moore (1979).

Brownian motion plus white noise is a particularly useful model because it is robust or conservative, in the sense that the estimate of the local trend depends on the data much more than it depends on assumptions implicit in the model. For this reason, it is common to describe the result of fitting the local level model as “smoothing the data” rather than describing it as an estimate of some aspect of a model.

### 3 Integrated Brownian motion plus white noise

The model with white noise and the integral of continuous-time Brownian motion is referred to as the “continuous analogue of the local linear trend model” by Durbin and Koopman (2001, page 60) and as the “continuous time local linear trend model” by Harvey (1989, page 487). In the state space formulation of this model, the state vector  $\alpha_i$  has two components: a velocity  $\beta_i$  and a position  $\mu_i$ . Taking the Brownian motion increment variance to be  $V_B$ , over a time interval of length  $h$  the random fluctuations of the state space vector have covariance matrix

$$\begin{bmatrix} \frac{1}{3}V_B h^3 & \frac{1}{2}V_B h^2 \\ \frac{1}{2}V_B h^2 & V_B h \end{bmatrix}.$$

Intuitively, the Brownian motion affecting the velocity also affects the position. These effects are highly correlated, but they are not perfectly correlated. A change in the velocity early in an interval will cause a larger change in position than the same change in the velocity late in the interval.

#### 3.1 Problems with the discrete-time “local linear trend model”

The corresponding discrete-time model, called the discrete local linear trend model, is often used for processes which are expected to vary in a fairly smooth fashion, with the local slope behaving like Brownian motion. A commonly-used mathematical description for equally-spaced times is

$$\begin{aligned} y_i &= \mu_i + \varepsilon_i & \varepsilon_i &\sim N(0, \sigma_\varepsilon^2) \\ \mu_i &= \mu_{i-1} + \beta_{i-1} + \eta_i & \eta_i &\sim N(0, \sigma_\eta^2) \\ \beta_i &= \beta_{i-1} + \zeta_i & \zeta_i &\sim N(0, \sigma_\zeta^2) \end{aligned}$$

where the white noise, level and slope disturbances,  $\varepsilon_i$ ,  $\eta_i$  and  $\zeta_i$  are independent. Intuitively, we can think of  $\beta_i$  as a true velocity and  $\mu_i$  as true position and  $y_i$  as measured position. In the case when  $\sigma_\eta^2 = 0$ , the model is referred to as the *smooth trend* model.

Even for equally-spaced data, the discrete-time model is not a good approximation to the continuous-time model. Consider the artificial set of “data” computed as  $\sin(t)$  and given in Table 1. The standard deviation of the data was taken to be 0.03 and the Brownian motion variance in the velocity was taken to be 1.0 per unit time for this example.

Table 1: **Table of position estimates obtained by smoothing “data” with values given by  $\sin(t)$**

Data		Discrete smooth trend model		Continuous smooth trend model	
$t$	$\sin(t)$	$\hat{\mu}$	s.e.	$\hat{\mu}$	s.e.
0.0	0.0000	0.0008	0.0289	0.0008	0.0286
0.2	0.1987	0.1988	0.0253	0.1989	0.0238
0.4	0.3894	0.3893	0.0250	0.3894	0.0233
0.6	0.5646	0.5637	0.0250	0.5637	0.0233
0.8	0.7174	0.7150	0.0253	0.7145	0.0238
1.0	0.8415	0.8440	0.0289	0.8443	0.0286

Table 2: **Table of velocity estimates obtained by smoothing “data” with values given by  $\sin(t)$**

True velocity		Discrete smooth trend model		Continuous smooth trend model	
$t$	$\cos(t)$	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.
0.0	1.0000			0.9963	0.3260
0.1	0.9950	0.9897	0.1767	0.9918	0.2192
0.2	0.9801			0.9782	0.2098
0.3	0.9553	0.9524	0.1530	0.9544	0.1993
0.4	0.9211			0.9194	0.2096
0.5	0.8776	0.8723	0.1530	0.8734	0.1990
0.6	0.8253			0.8165	0.2096
0.7	0.7648	0.7562	0.1530	0.7540	0.1993
0.8	0.6967			0.6911	0.2098
0.9	0.6216	0.6450	0.1767	0.6439	0.2192
1.0	0.5403			0.6281	0.3260
1.1	0.4536	0.6450	0.4809	0.6281	0.4542

The first two columns of Table 1 give the data: the time and the value of the position variable,  $\sin(t)$ . The next two columns give the position estimated using the discrete formulation of the smooth trend model and its estimated standard error. The final two columns give the same quantities estimated using the continuous formulation of the smooth trend model.

Table 2 gives the true velocity,  $\cos(t)$  in its second column. The next two columns give the velocity estimated using the discrete formulation of the smooth trend model and its estimated standard error. For the discrete smooth trend model, the estimates of velocity are often reported as being associated with the data points, but they are more accurately regarded as being estimates of the velocity at times midway between the points at which the position is observed. They have been presented this way in Table 2. The final two columns give the same quantities estimated using the continuous formulation of the smooth trend model. These would usually only be computed for the times for which position data was used as input, but have also been computed for the intermediate times for the purpose of comparison. (The model was fitted using time increments of 0.1, but regarding every second data point as having very large measurement variance.)

The estimates obtained using the two models are fairly similar, once allowance has been made for the fact that the velocity estimates for the discrete model are applicable to times intermediate between the data items. However, the standard errors are substantially different.

A reason for preferring the continuous-time version of the smooth trend model is that it gives the same results when data of very large measurement variance are added at intermediate times. This internal-consistency property does not hold for the discrete smooth trend model. In the limit as more such intermediate points are added, the estimates and standard errors given by the discrete-time smooth trend model tend to those for the continuous-time smooth trend model.

It should also be noted that the continuous-time model does not require much more computational effort than the discrete-time model. It is generally more realistic than the discrete-time model, so I believe that the continuous-time versions of the smooth trend model and the local linear trend model should always be preferred to the discrete-time versions.

### 3.2 Further integrated forms of Brownian motion

Processes for which a derivative is modelled by Brownian motion can be useful when the underlying true trend is expected to be smooth. Such models are an alternative to splines and have the advantage of not requiring choices to be made about the number and locations of knots.

Figure 4 shows some data from an experiment in which fluorescence was measured at a regular series of temperatures. One model fitted to this data assumed that there was white noise plus a trend with third derivative which behaves like Brownian motion. The increment variance for the Brownian motion was taken to be equal to the white noise variance.

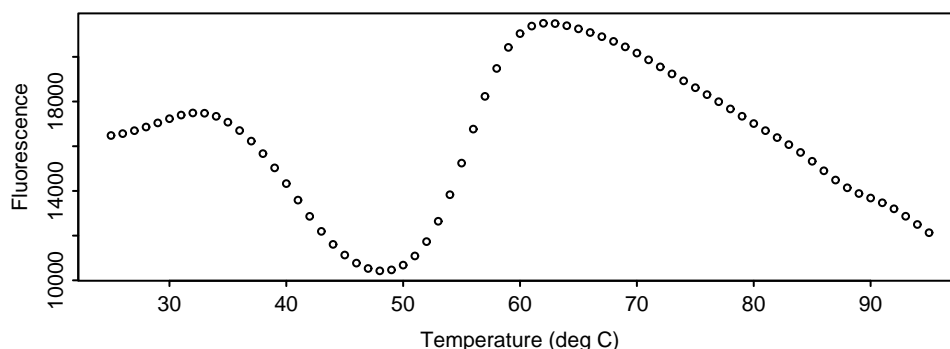


Figure 4: **Data to be smoothed**

A Brownian motion plus white noise model was also fitted to provide a comparison. The increment variance for this model was chosen to be such that the two models have the same sums of absolute residuals.

The residuals for the two models are shown in Figure 5. The residuals for the Brownian motion plus white noise model which are shown by a continuous line seem to have some trend left in them, while those for the third integral of Brownian motion plus white noise model seem to consist mainly of very high frequency noise. That the residuals for the Brownian motion plus white noise model appear to contain some low-frequency trend suggest that this model is inadequate.

## 4 Interval data

The variable that describes the spacing between data points is often time, but may be cumulative tonnage or some other measure of the opportunity for changes to occur. It will be referred to as a “spacing variable”. Most of the data items discussed in previous sections of this paper are “spot” values with respect to a spacing variable. This means that they are associated with a single value of the spacing variable. On the other hand, some data items are naturally associated with an interval in the set of possible values for the spacing variable.

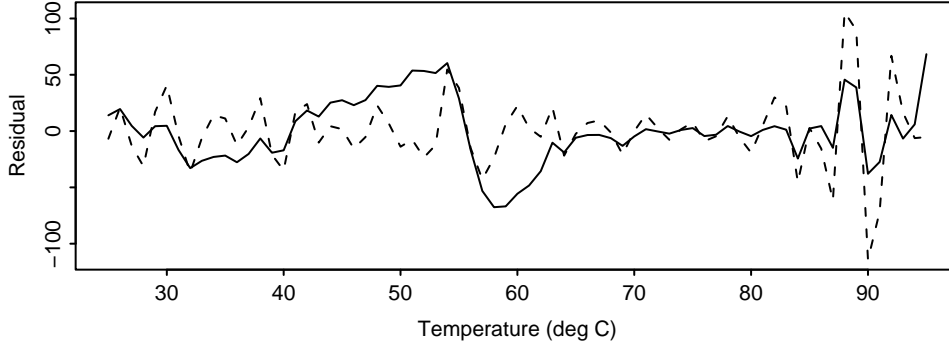


Figure 5: **Residuals for Brownian motion plus white noise model (shown as a solid line) and for third integral of Brownian motion plus white noise model (shown as a dashed line).**

In the processing of physical materials, “composite samples” are samples aggregated so as to be representative of the material over an interval. Routine samples in mining and mineral processing are often composite samples. For instance, the data illustrated in Figure 1 were interval data from testing of composite samples.

In econometrics, the terms “stocks” and “flows” are often used corresponding to spot samples and composite samples.

In Geostatistics, most data is regarded as interval data. The term “support” is used to describe the region of space, time or tonnage for which an item of data is representative. See David (1977) and Matheron (1965).

Sometimes interval data are reported as totals over a range of the spacing variable. For instance sales might be reported as monthly totals. In other circumstances, interval data are reported as averages over a range of the spacing variable. For instance the chemical composition of a composite sample is indicative of the average chemical composition of material over the period during which the composite sample was accumulated. The current discussion will assume that interval data are reported as averages. Data reported as totals can be translated into averages by dividing them by the lengths of the ranges of the spacing variable, so this does not represent any loss of generality.

In order to apply the Kalman filter to interval data, it is crucial to use a version of the Kalman filter like that presented in Koopman (1997) which allows the random components in the state equation and the observations to be correlated. Provided that such a version of the Kalman filter is used, fitting models to interval data is straightforward.

The two previous models will now be discussed for the case when the measured data are interval data. The intervals for which the data are averages will be assumed to adjoin one another, with the end of one interval being the start of the next. However, cases where intervals do not adjoin could easily be handled by adding intermediate intervals for which averages were measured with zero precision.

#### 4.1 Interval data for the Brownian motion plus white noise model

This combination of model and data may be written as a state space model with observation equation

$$y_i = \mu_i + G_i \varepsilon_i$$

and transition or state equation

$$\mu_{i+1} = \mu_i + H_i \varepsilon_i$$

where there is a three-dimensional vector of standardized disturbances,  $\varepsilon_i$  which are common to the state space model with observation equations. Three components of noise is the minimum number that can be used to explain the measurement error, the portion of the Brownian motion that equals  $\mu_{i+1} - \mu_i$ , and the integral of the Brownian motion that equals  $y_i - \mu_i$ . Note that the state space variables  $\mu_i$  are associated with starts of intervals but the

observations  $y_i$  are associated with the entire intervals. If the Brownian motion increment variance is  $V_i$  over the time period from  $t_i$  to  $t_{i+1}$  and the measurement error variance is  $M_i$  then the vectors  $G_i$  and  $H_i$  are

$$G_i = [ \sqrt{V_i/4} \quad \sqrt{V_i/12} \quad \sqrt{M_i} ],$$

$$H_i = [ \sqrt{V_i} \quad 0 \quad 0 ].$$

The coefficients of  $\sqrt{V_i}$  in these formulae come from the fact that the variance-covariance matrix of  $\mu_{i-1} - \mu_i$  and  $y_i$  is  $V_i$  times that of standard Brownian motion  $B_0(t)$  and its integral  $B_1(t)$ , which is shown in the Appendix to be

$$V_i \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

and the lower Cholesky factor of this is

$$\begin{bmatrix} \sqrt{V_i} & 0 \\ \sqrt{V_i/4} & \sqrt{V_i/12} \end{bmatrix}.$$

The three components of noise can be interpreted as follows.

1. Noise which affects the change in the level of the process between times  $t_i$  and  $t_{i+1}$ .
2. Noise which affects the difference between the average of the levels of the process at times  $t_i$  and  $t_{i+1}$  and the average level of the process over the continuous range of times from  $t_i$  to  $t_{i+1}$ .
3. Measurement error, which is the difference between the observed data item and the average level of the process over the continuous range of times from  $t_i$  to  $t_{i+1}$ .

## 4.2 Interval data for the integrated Brownian motion plus white noise model

This combination of model and data may be written as a state space model with observation equation

$$y_i = Z_i \alpha_i + G_i \varepsilon_i$$

and transition or state equation

$$\alpha_{i+1} = T_i \alpha_i + H_i \varepsilon_i.$$

There is a two-dimensional state space vector and a four-dimensional vector of standardized disturbances,  $\varepsilon_i$ . Denote  $t_{i+1} - t_i$  by  $h_i$ . The matrix  $Z_i$  has components 1 and  $\frac{1}{2}h_i$ , since the average position over an interval is affected by the starting position and starting velocity with these coefficients. The matrix  $T_i$  is

$$T_i = \begin{bmatrix} 1 & h_i \\ 0 & 1 \end{bmatrix}.$$

If the Brownian motion increment variance is  $V_i$  over the time period from  $t_i$  to  $t_{i+1}$  and the measurement error variance is  $M_i$  then

$$G_i = [ \sqrt{V_i/36} \quad \sqrt{V_i/48} \quad \sqrt{V_i/720} \quad \sqrt{M_i} ],$$

$$H_i = \begin{bmatrix} \sqrt{V_i} & 0 & 0 & 0 \\ \sqrt{V_i/4} & \sqrt{V_i/12} & 0 & 0 \end{bmatrix}.$$

The variance-covariance matrix of changes to the first two state space variables and  $y_i$  is  $V_i$  times that of  $B_0(t)$ ,  $B_1(t)$  and  $B_2(t)$ , which is shown in the Appendix to be

$$V_i \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{8} \\ \frac{1}{6} & \frac{1}{8} & \frac{1}{20} \end{bmatrix}$$

and its lower Cholesky factor is

$$\begin{bmatrix} \sqrt{V_i} & 0 & 0 \\ \sqrt{V_i/4} & \sqrt{V_i/12} & 0 \\ \sqrt{V_i/36} & \sqrt{V_i/48} & \sqrt{V_i/720} \end{bmatrix}.$$

The four components of noise can be interpreted as follows.

1. A component which primarily affects the slope of the process between times  $t_i$  and  $t_{i+1}$ .
2. A component which allows the change in the level of the process between times  $t_i$  and  $t_{i+1}$  to not be perfectly consistent with the average slope.
3. A component which allows the average level of the process between times  $t_i$  and  $t_{i+1}$  to not be perfectly consistent with the average slope and the values at the endpoints.
4. The measurement error, which is the difference between the observed data item and the average level of the process over the continuous range of times from  $t_i$  to  $t_{i+1}$ .

### 4.3 Does the difference between spot and interval data matter?

When analysing interval data from physical processes, it is important to remember that current plant settings should not be based on the best estimate of what has happened on average over the most recent interval for which data is available, but on an estimate of what was happening at the end of that interval or a prediction of the average over an interval subsequent to the most recent data (and which is not necessarily of the same length as the most recent interval for which data is available). Similarly, actions or policies based on the analysis of econometric interval data should be based on an estimate of what was happening at the end of the last interval or on a prediction of the average over an interval subsequent to the last data point.

As a simple example of how much difference it makes to treat data as interval data rather than as spot data, consider three data items 5, 5 and 4 occurring at times 1, 2 and 3, respectively. Suppose that the Brownian motion increment variance is 1 per unit time (the spacing between the data items) and that the measurement variance is 0.1. The local level model and the integrated Brownian motion model have been fitted to these data, first regarding the data as spot data and second regarding the data as interval data.

Figure 6 shows the best estimates and one-standard-deviation confidence limits for these four situations. The horizontal axes give the time, with times 1, 2 and 3 corresponding to the data points and times 4 and 5 corresponding to predictions of the level of the process at the next two yet-unobserved data points.

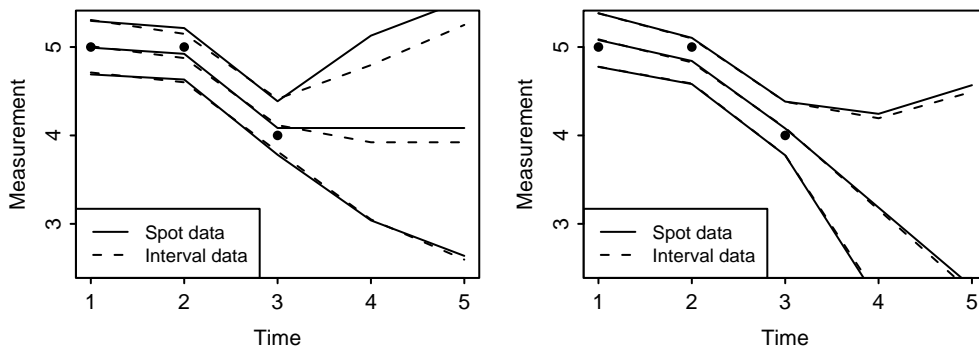


Figure 6: Smoothers for the local level model (on the left) and for the integrated Brownian motion model (on the right) for cases where the data are spot data or interval data.

For the local level model, if the data are spot data then the best prediction of the next data point is 4.084. This is above the level of the last data point because it is more likely

that the last measurement error was positive than negative. However, if the data are interval data then the best prediction of the next data point is 3.923. This is below the level of the last data point because that data point indicates that the process mean over an interval was approximately 4 and since the process level was approximately 5 at the start of that interval it was probably below 4 at the end of the interval.

The standard errors of prediction are smaller if the data are interval data than if they are spot data. This is consistent with the fact that, for the same variance parameters, the variogram for interval data is slightly less than for spot data. For standard Brownian motion described as  $B_0(t)$  in the Appendix, the variogram at lag  $t_2 - t_1$  is

$$\frac{1}{2}\text{Var}[B_0(t_2) - B_0(t_1)] = \frac{1}{2}(t_2 - t_1).$$

The time average over the period from  $t$  to  $t + h$  is  $(B_1(t + h) - B_1(t))/h$ . The variogram for lag  $t_2 - t_1$  for interval data where the intervals are of width  $h$  is

$$\frac{1}{2}\text{Var}\left[\frac{B_1(t_2 + h) - B_1(t_2)}{h} - \frac{B_1(t_1 + h) - B_1(t_1)}{h}\right].$$

For lags larger than  $h$  this is

$$\frac{1}{2}(t_2 - t_1 - \frac{1}{3}h)$$

and for lags smaller than  $h$  (i.e. overlapping intervals) it is

$$\frac{1}{2}(t_2 - t_1 - \frac{1}{3}h) + \frac{1}{6h^2}(t_1 + h - t_2)^3.$$

Suppose the true Brownian motion increment variance is  $V_B$  and the true measurement variance is  $V_M$ . The population variogram for spot data at lag  $h$  is  $hV_B + V_M$  and the population variogram for interval data at lag  $h$  is  $(h - \frac{1}{3})V_B + V_M$ . If a large amount of regularly-spaced interval data were incorrectly analysed as if it were spot data then provided that  $V_M - \frac{1}{3}V_B \geq 0$ , the white noise variance would be estimated to be approximately  $V_M - \frac{1}{3}V_B$  and the predictions from treating the data as spot data will be close to those which would be obtained from a more correct analysis. However, if  $V_M - \frac{1}{3}V_B < 0$  then the usual restriction  $V_M \geq 0$  on values for the dispersion parameters will increase the discrepancy between the predictions from treating the data as spot data and the correct analysis.

For the integrated Brownian motion model, Figure 6 shows that the data suggests that the process is generally decreasing. Such a general trend outside the range of the data would never be suggested by fitting the local level model. The difference between interval and spot data is less substantial than for the local level model.

### Subset of metallurgical plant data

A practical example where the difference between interval and spot data is important is illustrated in Figure 7. It shows some data from Figure 1.

The data points are actually interval data, so they have been plotted at the midpoints of the tonnage intervals to which they apply. Cumulative tonnage has been marked in kilo-tonnes, rather than mega-tonnes as in Figure 1.

The measurement error variance has been taken to be 0.001 and the Brownian motion increment variance has been taken to be 0.0004 per kT. A special feature of this data analysis is that the source of ore was changed at approximately 622 kT. This means that a large jump in grade could reasonably be expected at that point on the cumulative tonnage scale.

The continuous line on Figure 7 shows the estimated grade taking all of this information into account. The best estimates of the spot grades have been computed at the times marking boundaries between the intervals corresponding to the interval data. The points corresponding to these best estimates have been joined by straight lines, except that two separate horizontal line segments have been drawn for the interval including the stockpile change.

The estimates could be computed by Kalman filtering and smoothing for a model with a three-dimensional state vector containing spot grades for each of the stockpiles as well the average grade over an interval. They were actually computed using a mixed model.

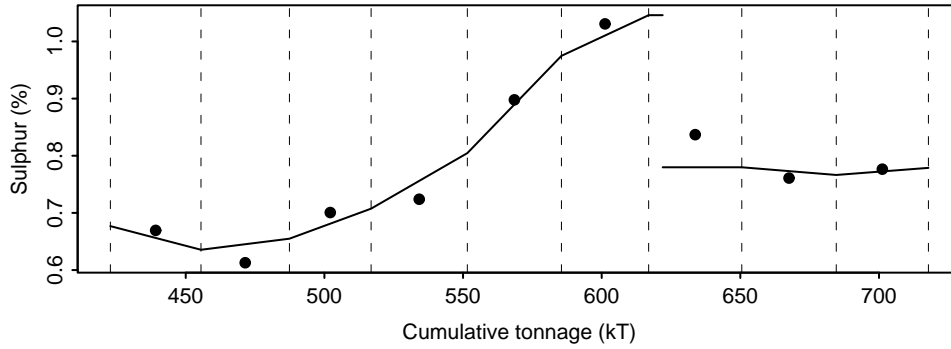


Figure 7: Interval data from a metallurgical plant together with an estimate of how the sulphur grade has varied with cumulative tonnage

Contributions to the inverse matrix for a pair of spot grades and the intervening interval grade are generally proportional to

$$\mathbf{V} = \begin{bmatrix} 4 & -6 & 2 \\ -6 & 12 & -6 \\ 2 & -6 & 4 \end{bmatrix}$$

as discussed in the Appendix. However for the case where a proportion  $p$  of the interval is from the first stockpile, the contribution to the inverse matrix is instead

$$\mathbf{V} - \mathbf{V}\mathbf{p}(\mathbf{p}^T\mathbf{V}\mathbf{p})^{-1}\mathbf{p}^T\mathbf{V}$$

where the vector  $\mathbf{p}$  has components 1,  $p$  and 0.

### Horticultural field trials

Another situation where the distinction between spot and interval data should be considered is in the spatial analysis of field experiments where the trend in soil fertility between neighbouring plots is considered. For instance, see Gleeson and Cullis (1987). For most field trials, the data are affected by averages of fertility over intervals, not spot fertilities.

One of the most-analysed data sets is from a Rothamstead experiment on control of mildew in barley. The data appear in Draper and Guttman (1980). The experiment is discussed more fully in Jenkyn *et al* (1979). The residuals from a simple linear model in which the yields depend only on the treatments are shown in Figure 8. Their sample variogram is shown in Figure 9, with the ordinary sample variogram shown as a continuous line and the robust variogram shown as a dashed line.

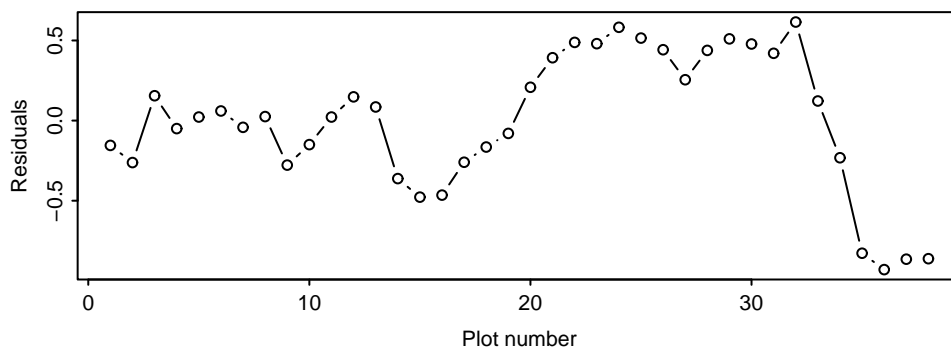


Figure 8: Residuals for Rothamstead mildew control experiment

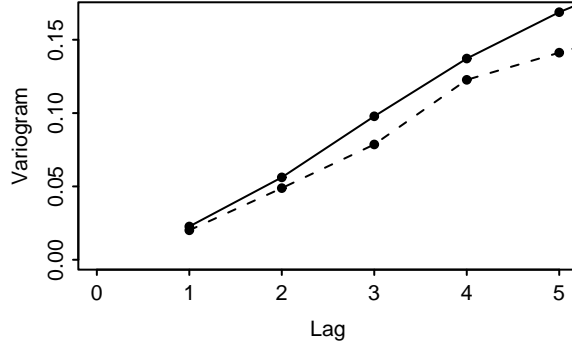


Figure 9: **Sample variogram of residuals for Rothamstead mildew control experiment**

The ordinary sample variogram was 0.023, 0.056, 0.098, 0.137 and 0.169 for lags 1 to 5 and the values for the robust variogram were 0.020, 0.049, 0.079, 0.122 and 0.141. These values are reasonably consistent with the algebra above which showed that the time average of standard Brownian motion has variogram  $\frac{2}{3}h, \frac{5}{3}h, \frac{8}{3}h, \dots$  at lags 1, 2, 3,  $\dots$ . However, the dispersion parameters for any model which fails to consider that the data are interval data rather than spot data might be very constrained by the fact that the variogram at lag zero cannot be negative.

## 5 Estimation of variance components

To fit models in practice, we must estimate variance components or variograms. Methods of doing this have been largely ignored in earlier portions of this paper.

Maximum likelihood and REML (restricted or residual maximum likelihood) are the most efficient methods for estimating variance components. However, simple methods based on sample variograms may sometimes be preferred because they are more robust. This preference is particularly likely when the amount of data is very large, because efficiency of estimation is less important and examining outliers is more onerous.

For independent observations  $\{Y_1, Y_2, \dots, Y_n\}$ , the sample variance

$$\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

is more efficient but less robust as an estimate of the population variance than an alternative based on the mean absolute deviation which is often used in Statistical Process Control, namely

$$\frac{\pi}{2} \left[ \frac{1}{n-1} \sum_{i=1}^n |Y_i - \bar{Y}| \right]^2.$$

For regularly-spaced data  $Y_1, Y_2, \dots, Y_n$  at times 1, 2,  $\dots, n$ , the sample variogram at lag  $h$  defined by

$$\frac{1}{2(n-h)} \sum_{i=1}^{n-h} (Y_i - Y_{i+h})^2$$

is an efficient estimate of the population variogram. By analogy to the robust estimate of the variance of independent observations, a robust form of the sample variogram for lag  $h$  based on absolute deviations is

$$\frac{\pi}{4} \left\{ \frac{1}{n-h} \sum_{i=1}^{n-h} |Y_i - Y_{i+h}| \right\}^2.$$

This is less sensitive to occasional very large jumps in the data than is the sample variogram. In the case where a single jump of size  $\delta$  dominates the estimates, the ordinary sample

variogram at lag 1 will be approximately  $\delta^2/[2(n-1)]$  and the robust variogram will be approximately  $\pi\delta^2/[4(n-1)^2]$ . The ordinary sample variogram will be larger by a factor of  $2(n-1)/\pi$ , which will generally be much larger than unity.

Cressie and Hawkins (1986) recommended the variogram estimate

$$\frac{\left\{ \sum_{i=1}^{n-h} |Y_i - Y_{i+h}|^{1/2} / (n-h) \right\}^4}{0.914 + 0.988/(n-h) + 0.090/(n-h)^2}$$

having shown it to be robust with respect to contamination of distributions by outliers. However, it is sensitive to rounding of data because the square roots of small differences are very sensitive to the rounding of those differences. It should not be used when a non-trivial proportion of pairs of nearby data items appear to be identical. The robust variogram estimate based on absolute differences is not sensitive to rounding of data and is preferred.

When estimating variance components, it is sometimes important to aim to use estimates of variance components which are appropriate for the times in the future when the most critical decisions will be made, rather than using estimates which are appropriate on average. For instance, when monitoring changes to the rate of survival from a relatively new surgical procedure, it may be sensible to use a variance component to describe Brownian motion increments based on opinions about the rates of change to survival rates that we would like to be able to detect. The alternative of estimating the historical variance per unit time of changes to the rate of survival is often more difficult, and is not necessarily indicative of future rates of change to the rate of survival.

The robust sample variogram should only be used when occasional substantial jumps in historical data are thought not to be indicative of the changes likely to occur in the future. For a situation like that indicated in Figure 17 where jumps in coal grade associated with stockpile changes are expected to occur in the future, the ordinary sample variogram should be preferred for assessing the precision achieved by acceptance control sampling.

## 5.1 Information about the white noise variance

The white noise variance can be estimated by fitting a variogram model and taking the value of the fitted variogram at zero lag. Saunders *et al* (1989) suggested that it is often useful to fit a linear variogram model by weighted linear regression to the values of the sample variogram for a few small lags. The simplest version of this procedure is to fit a straight line through the sample variograms at lags 1 and 2, denoted by  $V(1)$  and  $V(2)$ , giving  $2V(1) - V(2)$  as an estimate of the white noise variance.

Such estimates of the white noise variance can be quite misleading, even for regularly-spaced data. This possibility is illustrated here using simulated data.

Figure 10 shows a sample variogram calculated using every tenth of 10000 simulated data points, as if only those data points had been observed. This simulation first calculated a velocity-like quantity

$$X_i = 0.97X_{i-1} + \varepsilon_i$$

where  $X_0 = 0$  and  $\varepsilon_i \sim N(0, 0.01^2)$ . Then

$$Y_i = 0.93Y_{i-1} + X_i$$

where  $Y_0 = 0$  and

$$Z_i = Y_i + \eta_i$$

where  $\eta_i \sim N(0, 0.16^2)$ .

It appears on Figure 10 that the sample variogram is approximately linear at short lags. The white noise variance might be estimated by fitting a straight line to the sample variogram at the shortest available lags, 10 and 20, giving an estimated white noise variance of 0.006. This is very much smaller than the white noise variance  $0.16^2 = 0.0256$  which was used to generate the simulated data, and which is shown as a dashed line on Figure 10.

Figure 11 which shows the sample variogram calculated using all of the simulated data. The estimate of the white noise variance obtained by fitting a straight line to the variogram

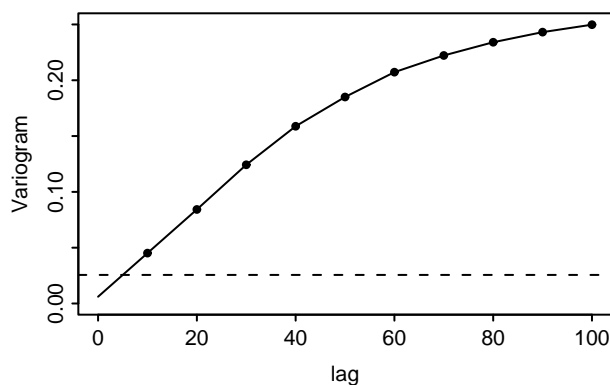


Figure 10: Variogram for simulated data using only every 10th data point, with extrapolation to lag zero. The dashed line shows the true white noise variance.

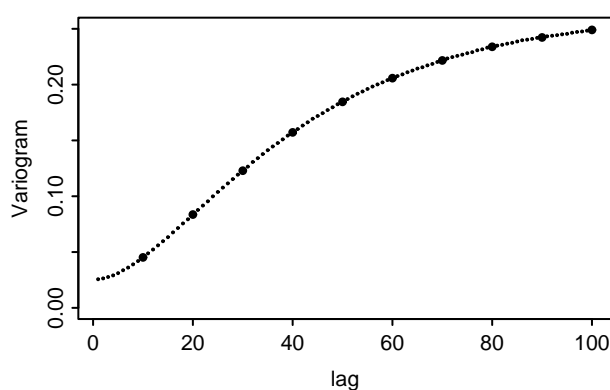


Figure 11: Variogram using all of a set of simulated data. The points for lags which are multiples of 10 (but which are not precisely the same as the values in Figure 10 because they are based on more data) are plotted more boldly than the intermediate values.

at the shortest available lags, 1 and 2, is 0.0249, which is close to the true white noise variance of 0.0256.

A small amount of the data at unit lags would have been sufficient to indicate that the variogram at short lags is nearer to 0.0256 than to 0.006. For instance, 30 observations at unit lags would have provided enough information to over-ride the impression from Figure 10 that the white noise variance was approximately 0.006.

This simulation illustrates the common possibility that the white noise variance can be estimated more reliably if a small proportion of the available data-collection effort is expended on replicating some observations, obtaining data at smaller than usual time lags, obtaining data at smaller than usual spatial separations, or conducting trials to estimate sampling and testing precision. The argument often expressed with words like “We cannot afford to replicate everything!” implicitly ignores such data-acquisition options.

Whenever possible, information about the white noise variance should be sought from sources other than the sequential observational data which is the primary data. Sometimes such information can be obtained without making extra measurements. Two practical situations where calculations using the binomial distribution provided useful estimates of white noise variance were estimating the proportion of potato chips that are of at least a prescribed length and monitoring the survival rates from a class of surgical procedures in a set of hospitals.

For some data sets, white noise variance should not be assumed to be constant. For instance, a measuring instrument might be replaced with a different instrument of different precision. Data on grades of shipments of iron ore is generally more precise for large

shipments than for small shipments, because the sampling and testing is done with more replication.

A more complicated example of changing white-noise variance occurs in signal analysis for aluminium smelting. The precision of measuring the resistance of a reduction cell usually decreases when there is an anode effect taking place for some other cell in the same line. The first of the events which cause this is that the voltage across the cell with an anode effect increases very rapidly. Automatic devices then reduce the amperage for the entire potline so that the total power drawn is approximately constant. This decrease in amperage reduces the voltage drop across a cell not experiencing an anode effect. The amperage and voltage both change very quickly, so any slight difference between the time responses of the voltage-measuring device and the amperage-measuring device is much more important than it was when both voltage and amperage were changing less rapidly.

## 5.2 Example — Monitoring air temperature

Consider the 1297 items of air temperature data collected at spacings of 10 minutes over 9 days starting on Sunday 28 March 1999 which are shown in Figure 12. The temperature was measured inside a box containing some water-measuring equipment. This example illustrates a non-standard way of estimating white noise variance. It also illustrates practical use of an integrated Brownian motion model.

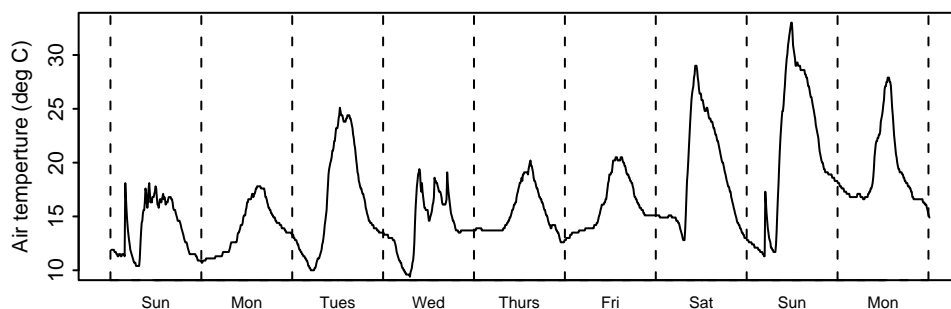


Figure 12: **Air temperature data for CityWest site 4A starting Sunday 28 March 1999. The vertical dashed lines correspond to midnights.**

The temperature fluctuations visible in the data are generally fairly smooth. Exceptions are upward spikes on both Sunday mornings and some irregular fluctuations around the middle of the Wednesday. On each of the Sunday mornings, there was one data item which was approximately  $6^{\circ}\text{C}$  higher than the preceding data item. These fluctuations were due to the automatic flushing of copper coils used for measuring the pH of reticulated water at 4:30 on Sundays and 16:30 on Wednesdays. The flushing rapidly changed the measured air temperature to very close to the temperature of the reticulated water, approximately  $19^{\circ}\text{C}$ . Another complication is that the clock on the logger was adjusted from Daylight Saving Time to East Australian Standard Time at approximately 13:20 on the Wednesday, and the copper coils were flushed at this time also.

A lower bound on the white noise variance was calculated as the variance due to rounding which occurred because temperature was estimated by measuring the voltage across a thermocouple and recording the voltage using a low-precision digital data logger. The average rounding variance calculated in this way is 0.005662. It was considered to be a useful approximation to the white noise variance even though there may be some additional white noise variance due to other sources.

It is expected that temperature inside the box will behave like integrated Brownian motion rather than like ordinary Brownian motion. When the external air temperature changes abruptly due, say, to a wind change, this will affect the rate at which heat flows into or out of the box, but will not affect the temperature very quickly. Therefore we look at the variogram of temperature differences in order to estimate the variance per unit time of increments for fitting an integrated Brownian motion model.

Figure 13 shows the robust variogram computed using as data the differences in air temperature between adjacent readings which are 10 minutes apart. In this case differences across the three times (4:30 on Sundays and 16:30 on Wednesday) at which the box had been opened were not included in the averages used for computation of the sample variogram. The robust form of variogram was preferred because we suspect that there may have been other jumps in temperature which might not be indicative of future trends in air temperature.

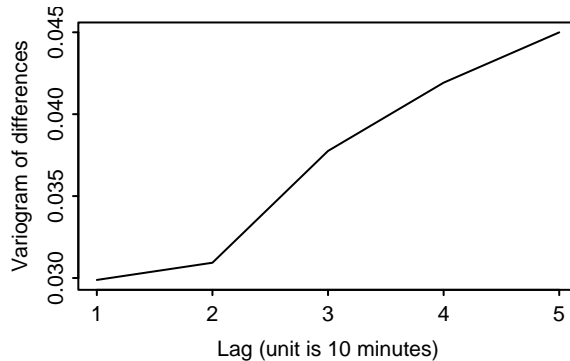


Figure 13: **Robust variogram of differences in air temperature over intervals of 10 minutes**

From the plotted variogram it appears that the variogram at lag zero is approximately 0.03. The variogram plotted is of differences between adjacent temperature readings, so the white noise associated with a single temperature reading is approximately half of this: 0.015. This is larger than the rounding error variance estimated earlier, which suggests that the rounding errors are not a critical problem.

The slope of the plotted variogram is approximately  $0.015/4$  per unit lag. If we fit an integrated Brownian motion model to this data then this analysis of historical data suggests that we use an increment variance of  $2 \times 0.015/4 = 0.0075$ .

Another issue to be considered is that the fluctuations in air temperature during the night are much slower than the fluctuations during the daylight hours. Since we are much more interested in developing data analysis procedures which perform well during daytime we should use a larger increment variance. Say 0.012.

## 6 Value judgements and practical advice

### 6.1 Basing computations on mixed models rather than Kalman filtering

Mixed models provide a way of analysing sequential data which gives the same estimates and standard errors for the last data point as the Kalman filter. Paige and Saunders (1977, page 192) suggest that it is easier to see the equivalence by noting that both methods give best linear unbiased estimates than by demonstrating the equivalence algebraically. This logic shows that mixed models give answers which are the same as smoothers with diffuse prior information.

For simple models based on continuous-time Brownian motion, fitting of mixed models requires a similar amount of computational effort to the combination of Kalman filtering and fixed interval smoothing, despite the impression that might otherwise be gained from sources such as Piepho and Ogutu (2007), Harvey (1989) and Durbin and Koopman (2001). If standard errors are not required then the mixed model approach seems often to be computationally cheaper.

Inverses of variance-covariance matrices for the simple models being discussed can be constructed by adding contributions of the types listed in the Appendix for each pair of successive time points. The matrices formed in this way are symmetric positive definite with an overlapping block diagonal structure. Computational procedures for dealing with banded matrices of this specialized class can be used as the basis for efficient data analysis.

These remarks about comparisons of computational effort are based on the proviso that inverses of variance-covariance matrices required for the mixed models are constructed directly, rather than first assembling the variance-covariance matrices and then inverting them. The remarks are somewhat imprecise because computer code for fitting of mixed models could be made computationally faster than mine by writing specialized versions for particular overlapping block structures; and computer code for the combination of Kalman filtering and fixed interval smoothing could be made computationally faster than mine by writing specialized versions for particular models, taking advantage of knowledge that some matrices are diagonal or nearly diagonal, and that some numbers are known to be zero or unity.

In notation like that of Robinson (1991) except that  $\sigma^2$  is taken to always be unity and  $Z$  is a unit matrix, the matrices assembled in this way are versions of  $G^{-1}$ . A diagonal matrix of measurement precisions is denoted by  $R^{-1}$ . The size of this matrix corresponds to the total number of random effects of interest. Many of these random effects are not measured and their measurement precisions are zero.

In the case where there are no fixed effects to be estimated, the matrix form of the *mixed model equations* for the best estimates is

$$(R^{-1} + G^{-1})\hat{u} = R^{-1}y$$

and their variance-covariance matrix of estimation errors is

$$(R^{-1} + G^{-1})^{-1}.$$

The portion of the inverse matrix which would be found if Kalman filtering and fixed interval smoothing were used can be found by a procedure due to Takahashi, Fagan and Chen (1973) which is explained in a more accessible manner by Erisman and Tinney (1975). This was pointed out to me by Frank de Hoog. Consider a symmetric matrix  $A$  which has been factored as  $A = LDL^T$  where  $L$  is unit lower triangular and  $D$  is diagonal. Suppose  $Z$  is the inverse of  $A$ . The relationship

$$Z = D^{-1}L^{-1} + (I - L^T)Z,$$

which can be checked by multiplying both sides by  $A$ , can be used as a recurrence relationship between the elements of  $Z$  at locations corresponding to the portion of the inverse matrix which is found by using Kalman filtering and fixed interval smoothing.

For the simple models discussed in this paper, this alternative computing procedure will not seem attractive to those readers who are already familiar with Kalman filtering. However for complicated models such as ones involving large numbers of fixed effect parameters, iterative methods for solving the mixed model equations provide a computational alternative which is sometimes very attractive relative to Kalman filtering. This may be useful as a part of recursive techniques to deal with non-normal error distributions.

The mixed model equations deal in a natural way with the case of no prior information, but they can be modified to deal with prior information by treating the prior estimates as additional data which has the prior variance-covariance matrix as its measurement variance. This is more straightforward than modifying Kalman filtering and smoothing algorithms to deal with the special case of no prior information.

## 6.2 When are variograms useful?

The mathematical definition of the population variogram of a random variable  $Y_t$  at lag  $h$  is

$$\gamma(h) = \frac{1}{2}\text{Var}(Y_t - Y_{t+h}),$$

provided that this is independent of  $t$ . If the process mean is independent of  $t$  and the autocovariance function  $C(h) = \text{cov}(Y(t+h), Y(t))$  is finite and independent of  $t$  then the population variogram is related to the autocovariance function by

$$\gamma(h) = C(0) - C(h)$$

where the process variance is  $C(0)$ .

The factor of  $\frac{1}{2}$  is included in the definition of a variogram so that the variogram is equal to the variance for a white noise process. The limiting value of the variogram for very small lags can be interpreted as the white noise variance. This limit is sometimes referred to as the “nugget effect” in mining applications. The name arose because when the grade of gold ore is measured for samples taken from very close together it is common to get substantially different measured grades. Some of the large differences are due to testing variation, but most are due to differences between samples in amounts of gold present as small nuggets.

The factor of  $\frac{1}{2}$  sometimes leads to confusion when interpreting the slope of a variogram. For instance, for a Brownian motion plus white noise process the slope of the variogram is half the increment variance per unit time for the Brownian motion.

The population variogram and the autocovariance function have a complementary relationship, so it might be thought that the variogram is an unnecessary concept. However, the variogram is much more useful than the autocovariance function when two conditions are satisfied.

First, it may be much easier to estimate a variogram than an autocovariance function. One such situation was illustrated in Robinson (1990). Intuitively, this occurs when the mean of a process is difficult to estimate from available data but the amount of short-term variation is relatively easy to estimate.

Second, the variogram is generally adequate for making estimates and statements about their precision for regions where there is plenty of data. For regions where there is little or no data, estimates and statements about their precision are referred to as predictions. Predictions are sensitive to the entire autocovariance function, not merely the variogram.

An example supporting this remark is illustrated in Figures 14, 15 and 16. Figure 14 shows two population autocovariance functions. The one shown as a dashed line has equation  $1 + \exp(-.5h)$  and the one shown as a dotted line has equation  $1 + 4.1347 \exp(-.1h)$ . The corresponding variograms are  $2 - \exp(-.5h)$  and  $5.1347 - 4.1347 \exp(-.1h)$  and are illustrated in Figure 15. The number  $4.1347 = (1 - \exp(-.5))/(1 - \exp(-.1))$  has been chosen so that the variograms are the same at lags 0 and 1. A linear variogram  $1 + 0.39347h$  with the same value at lags 0 and 1 is shown as a solid line on Figure 15. There is no corresponding autocovariance function because this is the variogram for a non-stationary model.

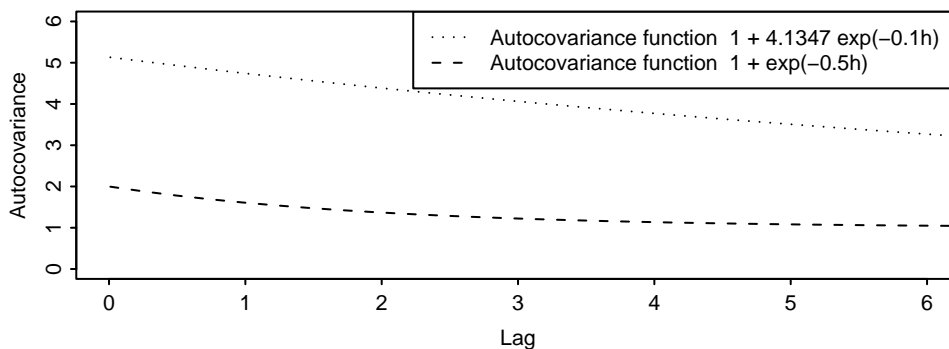


Figure 14: **Two autocovariance functions.**

Figure 16 shows estimates of the underlying trend based on a set of eight data points equally-spaced in time for each of the three variograms. The point estimates and two-standard-deviation limits are all displayed. (The quantities plotted have only been computed for discrete times at the spacing of the data points, and these points have been joined by straight line segments. More detailed computations would show that the standard deviations of estimates vary between the times of data points in a more complicated way, but this is not relevant for the present discussion.)

In the region where there is a moderate amount of data the estimates based on the three different variograms are similar. However, predictions for regions where there is little or no data are not similar.

Most books on Time Series give little attention to variograms. Diggle (1990) is an exception. I would encourage greater use of them. They are easy to interpret intuitively for

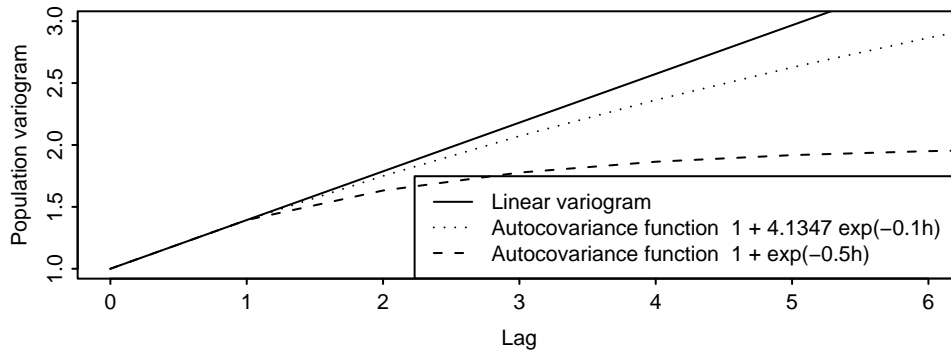


Figure 15: The variograms corresponding to the autocovariance functions given in Figure 14 and a linear variogram.

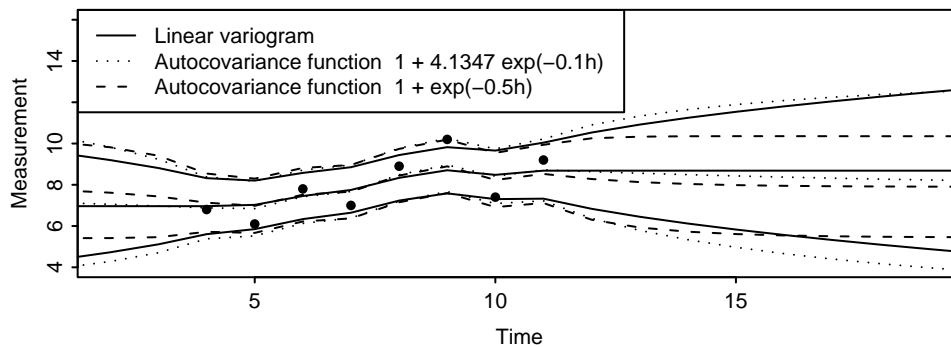


Figure 16: Data, trend estimates and confidence bounds based on the three variograms given in Figure 15.

most processes of practical interest. However, they should not be used for making predictions outside the region where there is plenty of data.

### 6.3 Process monitoring

Following Woodall (2000) and the associated discussion, I consider that “Statistical Process Control” is of surprising little use, given how broadly it is taught. SPC data-analysis techniques are based on describing the usual pattern of variation using only one component of noise. In my experience, a better starting point is to use two components: Brownian motion for the wandering of the process mean and white noise for measurement variation. These two components can be easily estimated and plotted separately. The variances of these components can be estimated using the slope and intercept of a straight line fitted by eye to a variogram for small lags.

There are many processes for which a small amount of wandering of the process mean occurs and is acceptable. Standard SPC rhetoric tends to suggest that such processes are “not in statistical control” and that no measure of process capability may be computed. However, when a Brownian motion plus white noise model is fitted it is possible to assess the likely maximum amount of drift in the process mean as well as the likely amount of future scatter about the process mean, and thereby make useful computations as to the probability that limits will be violated.

I have heard reports of people increasing the interval between data points until the data obtained suggest that “statistical control” has been achieved. This was mentioned by Hoerl and Palm (1992, p.271). Such practice would tend to mean that data were being collected too infrequently to provide information that would be useful for adjusting that process. It should be discouraged.

## 6.4 Acceptance control

When bulk commodities are traded, acceptance control does not check that every item in a lot is within acceptance limits. Rather, it checks that the average characteristics of lots are within acceptance limits. For instance, cargoes of iron ore will contain some lumps of quartz that contain virtually no iron, but the cargoes are acceptable provided that the average concentration of iron is greater than a specified lower limit, and the amounts of moisture, alumina, silica, phosphorus and undersize material are not greater than specified upper limits.

Note that the lot average is not the same as the long-run process average. Consider the ash percentage of coal from 100 sample increments taking during the loading of a ship, shown on Figure 17.

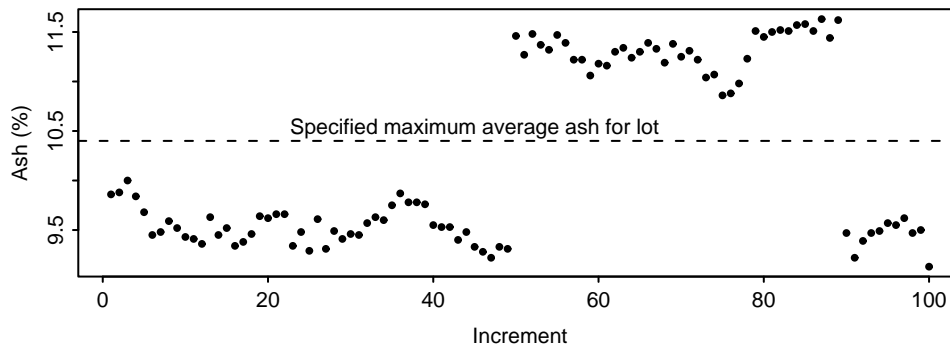


Figure 17: Simulated ash grade of increments for a shipment of coal

The precision of the lot average (10.2403%) as an estimate of the long-term process mean is extremely poor. From Figure 17 we can see that there are two jumps of about 2% ash. However, we have very little idea about the frequency of such jumps or the distribution of their sizes. The long-term process mean might easily be 9% ash or 15% ash.

Luckily, acceptance control only requires us to estimate the precision of the lot average as an estimate of the true lot mean. One way to estimate this precision of the lot average is to use the sample variogram. If the variogram of ash grades is linear with white noise variance  $A$  and slope  $B$ , then as shown by Matheron (1965, p.195) and Saunders *et al* (1989) the estimation error variance is approximately

$$\frac{1}{n} \left( A + \frac{1}{6} B d \right)$$

where  $n$  is the number of increments taken at systematic spacings during the loading of the shipload, and  $d$  is the spacing between increments using the same unit of spacing as was used to determine the variogram slope,  $B$ .

This example is unrealistic in that it is not common practice to separately assay all increments taken during shiploading. A moderately common practice is to prepare interpenetrating samples which are separately assayed. For instance, increments 1, 3, 5, 7 and 9 from the shipment could be collected into a single container, and a composite sample prepared and tested. Similarly, increments 2, 4, 6, 8 and 10 could be collected, a composite sample prepared, and a test result produced. Differences between such pairs of test results for many shipments can be used to get an adequate indication of the estimation error variance. See ISO 3084 (1998), for instance, for a description of options for routine monitoring of precision for acceptance control purposes.

## 6.5 Continuous-time models

In many situations continuous-time models can be interpreted more naturally than discrete-time models. For econometrics, Koopmans (1950) expressed the opinion that continuous-time models are likely to be superior to discrete-time models. He argued that “limitations to the usefulness of ... new methods ... arise from ... the treatment of time as a discrete

variable". He preferred to consider "disturbances ... as generated by a stochastic process with a continuous time variable". However, "the method of observation is ... a discrete procedure." "One method of observation is to make readings ... at equidistant points in time. Price variables are sometimes observed in this way. Quantities of goods and flows of money are usually observed through averages ... over a period of observation." "Although the mathematical difficulties involved may be considerable, a model of this kind would provide a means of studying the estimates of lag distributions and the choice of the most economic unit of observation." "... the most important advantage of a continuous treatment of this kind [is that it provides] the best way to study fully all aspects of the identification problem of relations between economic time series."

In this paper, discrete-time models have only been used where they can be interpreted as continuous-time models which have been observed at discrete points, and the interpretation would still work for unequally-spaced data and for making statements about what might have been happening at times between the observations. Some discrete-time models and techniques which are commonly used in Time Series theory do not have this property so I would discourage their use. The techniques include differencing of time series and the taking of moving averages. The models include the local linear trend model and the smooth trend model, as discussed in Section 3.1.

In discrete-time Time Series, the integrated moving average model IMA(0,1,1) is described as the "reduced form" of the random walk plus noise model. This means that once the parameters of the two models are appropriately matched they have the same autocorrelation function and give the same forecasts of future data. However, this equivalence does not extend to making statements about what might have been happening at times between the observations.

Continuous-time models allow clear differentiation between situations where observations correspond to averages over time intervals and situations where observations are spot data, unlike discrete-time models.

Continuous-time Brownian motion models are non-stationary, but are tractable from mathematical and computational points of view because Kalman filtering works for them and the inverses of covariance matrices are well-defined and easy to compute.

## 7 Key ideas

The main conclusions that I hope readers will retain from this paper are the following. They are listed in the same order as the subsections of Section 6 of this paper.

- The mixed model equations are generally a computationally efficient alternative to Kalman filtering and smoothing, provided that inverses of variance-covariance matrices are calculated directly rather than being calculated from the variance-covariance matrices. In particular, handling of situations with no prior information is generally easier if the mixed model equations are used.
- Variograms are a useful alternative to autocovariance functions when it is only desirable to make conclusions about what has been happening near to the data.
- The models used in Shewhart "Statistical Process Control" assume that there is only one component of noise. Models based on two components of noise are likely to be much more useful.
- Acceptance control situations where the average grade of a lot must be acceptable should be differentiated from situations when all items in a lot need to be acceptable. And when estimating the average grade of a lot it is important not to confuse this with the long-term process mean.
- Continuous-time models should be preferred to discrete-time models for modelling time series, particularly because they allow for a clear distinction between spot data and interval data.

## References

- Anderson, B.D.O. and Moore, J.B. (1979). *Optimal Filtering*. Prentice-Hall.
- Cressie, N. and Hawkins, D.M. (1980). Robust estimation of the variogram. *Journal of the International Association for Mathematical Geology* **12**, 115–125.
- David, M. (1977). *Geostatistical Ore Reserve Estimation*. Elsevier.
- Diggle, P.J. (1990). *Time Series. A Biostatistical Introduction*. Oxford University Press.
- Draper, N.R. and Guttman, I. (1980). Incorporating overlap effects from neighbouring units into response surface models. *Appl. Statist.* **29**, 128–136.
- Durbin, J. and Koopman, S. J. (2001) *Time Series Analysis by State Space Methods*. Oxford University Press. Matheson Mos 517.23 D953T
- Erismann, A.M. and Tinney, W.F. (1975). On computing certain elements of the inverse of a sparse matrix. *Communications of the ACM* **18** 177–179.
- Gleeson, A.C. and Cullis, B.R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43** 277–288.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Hoerl, R.W. and Palm, A.C. (1992). Discussion: integrating SPC and APC. *Technometrics* **34** 268–272.
- International Organization for Standardization (1998). *International Standard ISO 3084: Iron ores—Experimental methods for evaluation of quality variation*
- Jenkyn, J.F., Bainbridge, A., Dyke, G.V. and Todd, A.D. (1979). An investigation into inter-plot interactions, in experiments with mildew on barley, using balanced designs. *Ann. Appl. Biol.* **92**, 11–28.
- Koopman, S.J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* **92** 1630–1638.
- Matheron, G. (1965). *Les Variables Regionalis ees et leur Estimation*. Masson, Paris.
- Martin, R.S., Peters, G. and Wilkinson, J.H. (1965). Symmetric decomposition of a positive symmetric matrix. *Numerische Mathematik* **7** 362–383.
- Paige, C.C. and Saunders, M.A. (1977). Least squares estimation of discrete linear dynamic systems using orthogonal transformations. *SIAM Journal on Numerical Analysis* **14** 180–193.
- Piepho, H-P. and Ogutu, J.O. (2007). Sample state-space models in a mixed model framework. *The American Statistician* **61** 224–232.
- Rauch, H., Tung, F. and Striebel, C. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal* **3** 1445–1450.
- Robinson, G.K. (1990). A role for variograms. *Australian Journal of Statistics* **32** 327–335.
- Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science* **6** 15–51.
- Saunders, I.W., Robinson, G.K. Lwin, T. and Holmes, R.J. (1989). A simplified variogram method for determining the estimation error variance in sampling from a continuous stream. *International Journal of Mineral Processing* **25** 175–198.
- Takahashi, K. Fagan, J. and Chin, M-S. (1973). Formation of a sparse bus impedance matrix and its application to short circuit study. *8th PICA Conf. Proc.* June 4–6, Minneapolis, Minn.
- Taylor, H.M. (1982). Brownian motion. *Encyclopedia of Statistical Sciences*. Volume 1, 319–322. Wiley.
- Woodall, W.H. (2000). Controversies and contradictions in statistical process control. (with discussion). *Journal of Quality Technology* **32** 341–350.

## Appendix: Covariances of standard Brownian motion and variants

Consider continuous-time Brownian motion which is *standard* in the sense that changes in position of a particle over unit time have zero mean and unit variance. The process starts from position zero at time zero. A series of discrete approximations to this process can be generated using independent normally distributed steps of variance  $1/n$  for every time step of size  $1/n$ . According to Taylor (1982), these approximations converge weakly to standard Brownian motion. This series of approximations can be used to calculate moments of Brownian motion and its integrals.

Let  $B_0(t)$  denote standard continuous time Brownian motion at time  $t$  and let its successive integrals be denoted by  $B_1(t), B_2(t), \dots$ . Let  $Z(i)$  for  $i = 1, 2, \dots$  denote a sequence of independent standard normal variates. Then a discrete approximation with  $n$  steps per unit time to  $B_0(t)$  is

$$\sum_{i=1}^{i=nt} n^{-\frac{1}{2}} Z(i)$$

provided that  $t$  is an integer multiple of  $1/n$  so that  $nt$  is an integer.

These discrete approximations can be used to calculate variances and covariances, using the fact that the  $Z(i)$  are uncorrelated. The approximation to  $\text{Cov}(B_0(t_1), B_0(t_2))$  is

$$\text{Cov} \left( \sum_{i=1}^{nt_1} n^{-\frac{1}{2}} Z(i), \sum_{i=1}^{nt_2} n^{-\frac{1}{2}} Z(i) \right) = \sum_{i=1}^{n \min(t_1, t_2)} n^{-1}$$

As  $n \rightarrow \infty$  the discrete approximation tends to standard Brownian motion and the covariance tends to  $\min(t_1, t_2)$ .

### Variance calculations for standard continuous-time Brownian motion

Formulae for covariances for integrals of standard Brownian motion can be derived starting from the relationships  $\text{Cov}(B_0(t_1), B_0(t_2)) = \min(t_1, t_2)$ ,  $B_1(t) = \int B_0(t) dt$ ,  $B_2(t) = \int B_1(t) dt$ , etc. and swapping orders of integration. Two simple formulae are the following.

$$\begin{aligned} \text{Cov}(B_1(t_1), B_0(t_2)) &= \int_0^{t_1} \text{Cov}(B_0(t), B_0(t_2)) dt = \begin{cases} \frac{1}{2} t_1^2 & \text{if } t_1 \leq t_2 \\ t_1 t_2 - \frac{1}{2} t_2^2 & \text{if } t_1 \geq t_2 \end{cases} \\ \text{Cov}(B_1(t_1), B_1(t_2)) &= \int_0^{t_1} \text{Cov}(B_0(t), B_1(t_2)) dt = \begin{cases} -\frac{1}{6} t_1^3 + \frac{1}{2} t_1^2 t_2 & \text{if } t_1 \leq t_2 \\ \frac{1}{2} t_1 t_2^2 - \frac{1}{6} t_2^3 & \text{if } t_1 \geq t_2 \end{cases} \end{aligned}$$

In general, using  $\binom{j}{k}$  to denote the number of possible selections of  $k$  out of  $j$  objects,

$$\text{Cov}(B_i(t_1), B_j(t_2)) = \begin{cases} \frac{1}{(i+j+1)!} \sum_{k=0}^j (-1)^{j-k} \binom{i+j+1}{k} t_1^{i+j+1-k} t_2^k & \text{if } t_1 \leq t_2 \\ \frac{1}{(i+j+1)!} \sum_{k=j+1}^{i+j+1} (-1)^{j+1-k} \binom{i+j+1}{k} t_1^{i+j+1-k} t_2^k & \text{if } t_1 \geq t_2 \end{cases}$$

This will be proved by induction over the smaller of  $i$  and  $j$ .

For  $i = 0$  and  $j = 0$ ,  $\text{Cov}(B_0(t_1), B_0(t_2)) = \min\{t_1, t_2\}$ . When  $j > 0$ , assume that the relationship holds for smaller  $j$  and use the relationship

$$\text{Cov}(B_i(t_1), B_j(t_2)) = \int_0^{t_2} \text{Cov}(B_i(t_1), B_{j-1}(t)) dt.$$

To deal with the case  $t_1 \leq t_2$ , break the region of integration at  $t = t_1$

$$\int_0^{t_1} \text{Cov}(B_i(t_1), B_{j-1}(t)) dt = \int_0^{t_1} \frac{1}{(i+j)!} \sum_{k=j}^{i+j} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} t^k dt$$

$$\begin{aligned}
&= \frac{1}{(i+j)!} \sum_{k=j}^{i+j} (-1)^{j-k+1} \binom{i+j}{k} t_1^{i+j-k} \frac{1}{k+1} t_2^{k+1}. \\
\int_{t_1}^{t_2} \text{Cov}(B_i(t_1), B_{j-1}(t)) dt &= \int_{t_1}^{t_2} \frac{1}{(i+j)!} \sum_{k=0}^{j-1} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} t^k dt \\
&= \frac{1}{(i+j)!} \sum_{k=0}^{j-1} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} \frac{1}{k+1} t_2^{k+1} - \frac{1}{(i+j)!} \sum_{k=0}^{j-1} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} \frac{1}{k+1} t_1^{k+1}.
\end{aligned}$$

Now the terms in  $t_1^{i+j+1}$  from these two contributions have coefficients which add to

$$\frac{(-1)^{j+1}}{(i+j)!} \sum_{k=0}^{i+j} (-1)^k \binom{i+j}{k} \frac{1}{k+1} = \frac{(-1)^{j+1}}{(i+j+1)!} \sum_{k=0}^{i+j} (-1)^k \binom{i+j+1}{k+1} = \frac{(-1)^j}{(i+j+1)!}$$

since

$$\sum_{k=0}^{i+j+1} (-1)^k \binom{i+j+1}{k+1} = (1-1)^{i+j+1} = 0.$$

Hence the total of the two contributions is

$$\begin{aligned}
&\frac{(-1)^j}{(i+j+1)!} t_1^{i+j+1} + \frac{1}{(i+j)!} \sum_{k=0}^{j-1} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} \frac{1}{k+1} t_2^{k+1} \\
&= \frac{1}{(i+j+1)!} \sum_{k=0}^j (-1)^{j-k} \binom{i+j+1}{k} t_1^{i+j+1-k} t_2^k.
\end{aligned}$$

The case when  $t_1 \geq t_2$  is simpler.

$$\begin{aligned}
\int_0^{t_2} \text{Cov}(B_i(t_1), B_{j-1}(t)) dt &= \int_0^{t_2} \frac{1}{(i+j)!} \sum_{k=j}^{i+j} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} t^k dt \\
&= \frac{1}{(i+j)!} \sum_{k=j}^{i+j} (-1)^{j-k} \binom{i+j}{k} t_1^{i+j-k} \frac{1}{k+1} t_2^{k+1} \\
&= \sum_{k=j}^{i+j} (-1)^{j-k} \frac{1}{(i+j-k)!(k+1)!} t_1^{i+j-k} t_2^{k+1} \\
&= \sum_{m=j+1}^{i+j+1} (-1)^{j+1-m} \frac{1}{(i+j+1-m)!m!} t_1^{i+j+1-m} t_2^m.
\end{aligned}$$

This completes the proof.

A useful special case of this general formula is found by taking  $t_1 = t_2 = t$  in the general expression for covariances:

$$\text{Cov}(B_i(t), B_j(t)) = \frac{1}{(i+j+1)!} \sum_{k=0}^j (-1)^{j-k} \binom{i+j-1}{k} t^{i+j-1} = \frac{1}{(i+j+1)!i!j!} t^{i+j-1}.$$

For instance

$$\text{Var} \begin{bmatrix} B_0(t) \\ B_1(t) \\ B_2(t) \\ B_3(t) \end{bmatrix} = \begin{bmatrix} t & \frac{1}{2}t^2 & \frac{1}{6}t^3 & \frac{1}{24}t^4 \\ \frac{1}{2}t^2 & \frac{1}{3}t^3 & \frac{1}{8}t^4 & \frac{1}{30}t^5 \\ \frac{1}{6}t^3 & \frac{1}{8}t^4 & \frac{1}{20}t^5 & \frac{1}{72}t^6 \\ \frac{1}{24}t^4 & \frac{1}{30}t^5 & \frac{1}{72}t^6 & \frac{1}{252}t^7 \end{bmatrix}.$$

This is useful for fitting models using Kalman filtering, because it tells us how much change in a multiply integrated Brownian motion process is likely to occur over a given time interval.

## Inverses

General rules for assembling the inverses of variance-covariance matrices for fitting the models which have been discussed in this paper are as follows, where the Brownian motion has increment variance  $V_B$  per unit time. The limit of the inverse of the variance-covariance matrix of  $B_0(t)$  and  $B_0(t+h)$  as  $t \rightarrow \infty$ , which could be used for fitting the Brownian motion plus white noise model to spot data using a mixed model approach, is

$$V_B^{-1}h^{-1} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The limit of the inverse of the variance-covariance matrix of  $B_0(t)$ ,  $[B_1(t+h) - B_1(t)]/h$  and  $B_0(t+h)$ , which could be used for fitting the Brownian motion plus white noise model to interval data using a mixed model approach, is

$$V_B^{-1}h^{-1} \begin{bmatrix} 4 & -6 & 2 \\ -6 & 12 & -6 \\ 2 & -6 & 4 \end{bmatrix}.$$

The limit of the inverse of the variance-covariance matrix of  $B_0(t)$ ,  $B_1(t)/h$ ,  $B_0(t+h)$  and  $B_1(t+h)/h$ , which could be used for fitting the integrated Brownian motion plus white noise model to spot data using a mixed model approach, is

$$V_B^{-1}h^{-1} \begin{bmatrix} 4 & 6 & 2 & -6 \\ 6 & 12 & 6 & -12 \\ 2 & 6 & 4 & -6 \\ -6 & -12 & -6 & 12 \end{bmatrix}.$$

The limit of the inverse of the variance-covariance matrix of  $B_0(t)$ ,  $B_1(t)/h$ ,  $[B_2(t+h) - B_2(t)]/h^2$ ,  $B_0(t+h)$  and  $B_1(t+h)/h$ , which could be used for fitting the integrated Brownian motion plus white noise model to interval data using a mixed model approach, is

$$V_B^{-1}h^{-1} \begin{bmatrix} 9 & 36 & -60 & -3 & 24 \\ 36 & 192 & -360 & -24 & 168 \\ -60 & -360 & 720 & 60 & -360 \\ -3 & -24 & 60 & 9 & -36 \\ 24 & 168 & -360 & -36 & 192 \end{bmatrix}.$$

The next three matrices could be used for fitting further integrated forms of Brownian motion to spot data. The limit of the inverse of the variance-covariance matrix of  $B_0(t)$ ,  $B_1(t)/h$ ,  $B_2(t)/h^2$ ,  $B_0(t+h)$ ,  $B_1(t+h)/h$  and  $B_2(t+h)/h^2$  is

$$V_B^{-1}h^{-1} \begin{bmatrix} 9 & 36 & 60 & -3 & 24 & -60 \\ 36 & 192 & 360 & -24 & 168 & -360 \\ 60 & 360 & 720 & -60 & 360 & -720 \\ -3 & -24 & -60 & 9 & -36 & 60 \\ 24 & 168 & 360 & -36 & 192 & -360 \\ -60 & -360 & -720 & 60 & -360 & 720 \end{bmatrix}.$$

The limit of the inverse of the variance-covariance matrix of  $B_0(t)$ ,  $B_1(t)/h$ ,  $B_2(t)/h^2$ ,  $B_3(t)/h^3$ ,  $B_0(t+h)$ ,  $B_1(t+h)/h$ ,  $B_2(t+h)/h^2$  and  $B_3(t+h)/h^3$  is  $V_B^{-1}h^{-1}$  times

$$\begin{bmatrix} 16 & 120 & 480 & 840 & 4 & -60 & 360 & -840 \\ 120 & 1200 & 5400 & 10080 & 60 & -840 & 4680 & -10080 \\ 480 & 5400 & 25920 & 50400 & 360 & -4680 & 24480 & -50400 \\ 840 & 10080 & 50400 & 100800 & 840 & -10080 & 50400 & -100800 \\ 4 & 60 & 360 & 840 & 16 & -120 & 480 & -840 \\ -60 & -840 & -4680 & -10080 & -120 & 1200 & -5400 & 10080 \\ 360 & 4680 & 24480 & 50400 & 480 & -5400 & 25920 & -50400 \\ -840 & -10080 & -50400 & -100800 & -840 & 10080 & -50400 & 100800 \end{bmatrix}.$$

The limit of the inverse of the variance-covariance matrix of  $B_0(t)$ ,  $B_1(t)/h$ ,  $B_2(t)/h^2$ ,  $B_3(t)/h^3$ ,  $B_4(t)/h^4$ ,  $B_0(t+h)$ ,  $B_1(t+h)/h$ ,  $B_2(t+h)/h^2$ ,  $B_3(t+h)/h^3$  and  $B_4(t+h)/h^4$  is  $V_B^{-1}h^{-1}$  times

$$\begin{bmatrix} 25 & 300 & 2100 & 8400 & 15120 & -5 & 120 & -1260 & 6720 & -15120 \\ 300 & 4800 & 37800 & 161280 & 302400 & -120 & 2760 & -27720 & 141120 & -302400 \\ 2100 & 37800 & 317520 & 1411200 & 2721600 & -1260 & 27720 & -267120 & 1310400 & -2721600 \\ 8400 & 161280 & 1411200 & 6451200 & 12700800 & -6720 & 141120 & -1310400 & 6249600 & -12700800 \\ 15120 & 302400 & 2721600 & 12700800 & 25401600 & -15120 & 302400 & -2721600 & 12700800 & -25401600 \\ -5 & -120 & -1260 & -6720 & -15120 & 25 & -300 & 2100 & -8400 & 15120 \\ 120 & 2760 & 27720 & 141120 & 302400 & -300 & 4800 & -37800 & 161280 & -302400 \\ -1260 & -27720 & -267120 & -1310400 & -2721600 & 2100 & -37800 & 317520 & -1411200 & 2721600 \\ 6720 & 141120 & 1310400 & 6249600 & 12700800 & -8400 & 161280 & -1411200 & 6451200 & -12700800 \\ -15120 & -302400 & -2721600 & -12700800 & -25401600 & 15120 & -302400 & 2721600 & -12700800 & 25401600 \end{bmatrix}$$

For simple cases, these limits of inverses of variance-covariance matrices can be computed to low precision simply by trying various values of  $t$ . An approach which gives much better precision is based on noting that the effect of increasing  $t$  on such matrices is equivalent to increasing uncertainty about the location of the Brownian motion process at time zero and increasing uncertainty about some of its derivatives at time zero.

Suppose that  $V$  is a variance-covariance matrix of some  $B_i(t)$  for finite  $t$ . Let  $Z$  denote the matrix of contributions to those  $B_i(t)$  which would be caused by changes to the origin of the Brownian motion. The matrix identity

$$(R + ZGZ^T)^{-1} = R^{-1} - R^{-1}Z(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}.$$

is used to compute the inverse of the limit of the variance-covariance as the variance,  $G$ , of such contributions becomes infinity large by taking the limit as  $G^{-1}$  tends to the zero matrix.