Modeling irregularly spaced residual series as a continuous stochastic process

Jos R. von Asmuth

Water Resources Section, Faculty of Civil Engineering, Delft University of Technology, Delft, Netherlands

Marc F. P. Bierkens

Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, Netherlands

Received 11 October 2004; revised 22 August 2005; accepted 26 August 2005; published 1 December 2005.

[1] In this paper, the background and functioning of a simple but effective continuous time approach for modeling irregularly spaced residual series is presented. The basic equations were published earlier by von Asmuth et al. (2002), who used them as part of a continuous time transfer function noise model. It is shown that the methods behind the model are build on two principles: The first is the fact that the equations of a Kalman filter degenerate to a form that is equivalent to ''conventional'' autoregressive moving average (ARMA) models when the modeled data are considered to be free of measurement errors. This assumption, in comparison to the ''full'' Kalman filter, also yields a better prediction efficiency (Ahsan and O'Connor, 1994). The second is the mathematical equivalence between discrete time AR parameters and continuous exponentials and the point that continuous time models provide an elegant solution for modeling irregularly spaced observations (e.g., Harvey, 1989). Because simple least squares methods do not apply in case of modeling irregular data, a sum of weighted squared innovations (SWSI) criterion is introduced and derived from the likelihood function of the innovations. In an example application it is shown that the estimates of the SWSI criterion converge to maximum likelihood estimates for larger sample sizes. Finally, we propose to use the so-called innovation variance function as an additional diagnostic check, next to the well-known autocorrelation and cross-correlation functions.

Citation: von Asmuth, J. R., and M. F. P. Bierkens (2005), Modeling irregularly spaced residual series as a continuous stochastic process, *Water Resour. Res.*, *41*, W12404, doi:10.1029/2004WR003726.

1. Introduction

[2] One of the immediate consequences of the stochastic nature of natural processes is that the predictions of a deterministic model will never match a set of observations completely. The difference between the model predictions and the observed time series forms a time series of its own, called the ''residuals.'' Residuals are ''caused'' by errors in the observation process, errors in the model parameters, simplifications or errors in the model concept and/or numerical errors when evaluating the model equations. Furthermore, the value of a model residual at a certain time instant is often correlated with its value at earlier time instants, so residuals cannot simply be modeled as a set of independent Gaussian deviates. Explicitly modeling the behavior of the residuals of a transfer or deterministic model can have several purposes. First, when a noise model is fitted to the residuals, it can be used for stochastic simulation. This is especially useful in case one is interested in the probability of extremes, as such probabilities are underestimated when using only the deterministic model [e.g., *Knotters and Van Walsum*, 1997]. Second, because of the autocorrelation in the signal, the noise model can also be used to yield predictions of the residuals at unobserved time steps, either for smoothing, forecasting or updating purposes. Such applications, which are widely used in the meteorological sciences and are often referred to as data assimilation [*McLaughlin*, 1995; *Kalnay*, 2002], make optimal use of both model prediction and observations and can significantly improve the accuracy of the predictions of the combined model. Third, optimization algorithms for estimating the model parameters and their covariance matrix often assume that model errors are uncorrelated. Correlations between the input, predicted output and the residual series hamper an independent estimate of the model parameters, whereas autocorrelation in the residuals causes the variance of the parameters to be underestimated. ''Whitening'' of the residuals with a noise model thus improves the parameter estimates of transfer and other deterministic models [*Bryson and Henrikson*, 1965; *Te Stroet*, 1995].

[3] In light of the problem of adequately dealing with residuals, or more in general noise corrupted signals, one can discern two directions in the methods used. First, there are statistical time series analysis methods, which are popular in fields of science like econometry or social sciences. Their application developed rapidly after the publication of the comprehensive text book by *Box and Jenkins* [1970]. In the univariate version of these models, which are classified under the name autoregressive integrated

0043-1397/05/2004WR003726$09.00

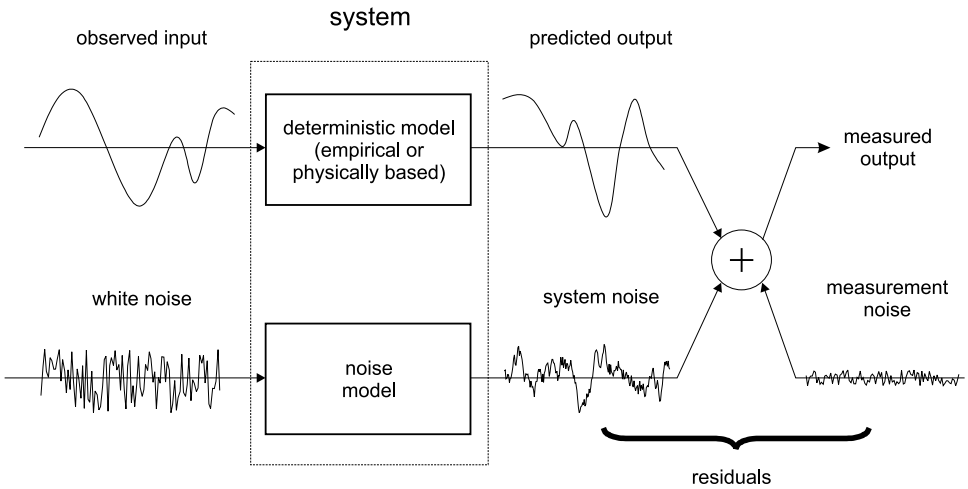

Figure 1. Schematical representation of a combined deterministic-stochastic model of a system under the influence of system noise, measurement noise, and a deterministic signal.

moving average or (X)ARIMA, a series of observations of a given variable is modeled using only the temporal correlation structures in the data itself. In the multivariate case, also data of other, explanatory variables is used (transfer part of the model), while the errors of the transfer model are modeled with a separate univariate or so-called noise model. The combined model is referred to as a transfer function noise (TFN) model. Second, there are the filtering methods, of which N. Wiener can be considered the founder [e.g., *Wiener*, 1949]. Such methods were firstly applied in the field of control, navigation and communication engineering, where the problem of extracting useful information from signals that are noise corrupted arises most naturally. Here the introduction of the Kalman filter [*Kalman*, 1960] was an important factor in the rapid spreading of its application across the more exact sciences, owing to its generality, simplicity, and ease of application. Nowadays, the Kalman filter is applied across a wide range of disciplines for modeling the behavior and uncertainty of phenomena that are not well explained by physical laws alone.

[4] The distinction made above, however, is somewhat artificial, because both methods are closely related. Apart from the fact that their mathematical treatment and notation differs, there are two important differences. First, in the Kalman filter a distinction is made between noise that perturbs the state of the system itself, and noise perturbing the process of measuring that state (Figure 1). In ARIMA models, on the other hand, there is only one noise term uniting all sources of uncertainty in the modeled signal. However, the Kalman filter is often also applied for pure prediction problems where measurement error is assumed to be absent. *Ahsan and O'Connor* [1994] show that in that case, the Kalman gain becomes redundant, and the filter equations reduce to a simpler form equivalent to that of an ARMA model. A second important difference, one that was not discussed by Ahsan and O'Connor, lies in dealing with irregular data. On this point, conventional discrete time ARMA models have an important practical drawback, as they cannot be readily applied to data with missing observations, while also the frequency of the input and output variables is coupled and has to be equal. Over the years, there has been a lot of effort in solving this problem, in the time series literature [*Jones*, 1980; *Harvey and Pierse*, 1984; *Little and Rubin*, 1987] but more recently also focused on hydrologic problems [*Bierkens et al.*, 1999; *Koutsoyiannis*, 2001; *Berendrecht et al.*, 2003; *Yi and Lee*, 2004]. Mostly, these solutions involve a state-space representation (SSR) of the model equations embedded in a Kalman filter. This way, missing observations can be easily handled by simply omitting the updating equations while retaining the prediction equations. Furthermore, to estimate the ARMA parameters, a likelihood function is constructed using the prediction errors and their variances, as simple least squares methods no longer apply.

[5] In this paper, we focus on dealing with irregular data, and argue that also in that case the Kalman filter can be reduced to a more simple form. Furthermore, we will solve the problem in a continuous time domain using a simple Ornstein-Uhlenbeck–based (OUB) noise model. Continuous time models are often considered to be more fundamental than their discrete time counterparts, and can provide an elegant solution for modeling irregularly spaced observations and data with mixed frequencies [*Harvey*, 1989]. For many variables, the process generating the observations can be regarded as a continuous one even though the observations themselves are only made at discrete intervals. In the economic sciences, already a good deal of the theory is based on continuous time models [*Khabie-Zeitoune*, 1982; *Bergstrom*, 1990; *Brockwell*, 2001]. The basic equations of the OUB noise model were published earlier by *von Asmuth et al.* [2002] where they were used as part of a continuous time transfer function noise (TFN) model. In that paper, we mainly restricted ourselves to describing how the transfer part of the model transforms irregularly spaced input series. The noise model, however, deals with prediction errors in the output, the parameter optimization process, and stochastic simulation applications. Because of its stochastic nature, its properties are fundamentally different from those of the transfer model, and certainly as important in light of dealing with irregular data. Therefore, its background and functioning are described separately in the current paper. The OUB model is equivalent to an AR(1) model, which is often used to model the residuals in hydrological applications, and we

will show that it also in this case suffices to effectively whiten the residuals.

[6] The paper is organized as follows: first, the forecasting mode of an AR(1) model, the equations of the time update in the Kalman filter, and the OUB model are discussed, and it is shown that they are mathematically equivalent. Next, specific attention is paid to the methods for estimating the model parameters. As an alternative for the maximum likelihood (ML) function, a sum of weighted squared innovations (SWSI) criterion is derived and introduced. In an example application, both optimization criteria are compared and it is shown that the SWSI criterion converges to the ML function for larger sample sizes. Finally, the way the prediction error or innovation variance varies with the time step is investigated, because that is a crucial part in handling data with irregular time steps correctly. In Appendix A the main derivations are given in detail.

2. Theory and Background

2.1. Irregular Data and the AR1 Model

[7] For a general treatment on the functioning of ARMA models we refer to *Box and Jenkins* [1970]. Here we will restrict ourselves to an AR(1) model and aspects that concern handling irregular data, because of the equivalency with the OUB model discussed later on. The AR(1) model is given in mathematical terms by

$$\tilde{n}_t = \phi \tilde{n}_{t-1} + a_t \tag{1}$$

where $\tilde{n}_t$ are the deviations of the residual series n_t from its mean $E(n_t)$, ϕ is an autoregressive parameter, and a_t is a discrete white noise process, with properties $E\{a_t\} = 0$, so that $a_t = \tilde{a}_t$, $E\{a_t^2\} = \sigma_a^2$, and $E[a_t a_{t+i}] = 0$, $i \neq 0$. In ARMA models in general, the white noise series a_t equals the series of one-step-ahead prediction errors, or innovations. For the AR(1) model, an estimate of a_t is obtained by subtracting the one step ahead prediction $\hat{\tilde{n}}_{t|\tilde{n}_{t-1}} = \phi \tilde{n}_{t-1}$ from the observed value of $\tilde{n}_t$. Because of this, a missing observation of $\tilde{n}$ at time t implies that a_t and a_{t+1} cannot be calculated. Prediction of $\tilde{n}_t$ using $\tilde{n}_{t-2}$ or more time steps back does not offer a straightforward solution, because the lead time influences the prediction error or variance of a_t, which should be stationary. Therefore, a time series must be complete and given at regular intervals in order to be able to fit an ARMA model on the data. The forecasting mode of ARMA models, however, does provide the necessary equations to predict over a variable lead time and to quantify the accompanying prediction error [see *Box and Jenkins*, 1970]. For the AR(1) model, the prediction and its variance are a function of the time lag l and given by

$$\begin{aligned} \hat{\tilde{n}}_{t+l} &= \phi^l \tilde{n}_t \\ \sigma^2_{e_{t+l}} &= \sigma_a^2 \frac{(1-\phi^{2l})}{(1-\phi^2)} \end{aligned} \tag{2}$$

where $\hat{\tilde{n}}$ is the unbiased prediction of $\tilde{n}_{t+l}$ given the available observations up to time step 1 and $\sigma^2_{e_{t+l}}$ is the variance of prediction error e_{t+l}. In other words, the main reason that ARMA models cannot cope with irregular data is that parameter estimation and estimation of missing values cannot be done simultaneously. *Brubacher and Wilson* [1976] already recognized this and devised a technique that regenerates the residuals using the forecasting and hindcasting mode of ARMA models, but according to [*Hipel and McLeod*, 1994, p. 695] in practice it was not very convenient.

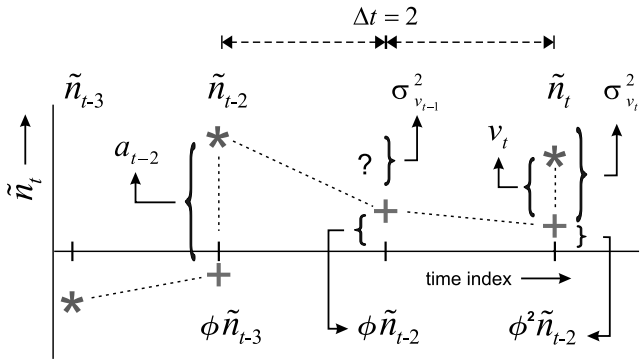

Figure 2. Schematical representation of the functioning of an AR(1) model combined with a "degenerate" Kalman filter. The star denotes a residual, whereas the cross denotes a predicted residual. The model predicts missing values of the residual series at every time step, along with the prediction error variance.

2.2. Combined AR(1) Model and "Degenerate" Kalman Filter

[8] We refer to *Ahsan and O'Connor* [1994] for an extensive treatment on how the Kalman filter can be "degenerated" to a simpler form in the pure prediction scenario, and how the state-space representation in general relates to ARMA models in standard notation. To facilitate comparison, here we will use standard notation. Figure 2 gives a schematic representation of a combined AR(1) model and the degenerate Kalman filter. When observations are missing or scarce, the value of a_t cannot be determined for every time step. Instead, an irregularly spaced innovation series ν_t is estimated. While σ_a^2 is a constant, $\sigma^2_{\nu_t}$ depends on the time lag between two observations. In this approach, fitting the model to an irregular series yields an estimate of ϕ, σ_a^2, ν_t, and a specific variance $\sigma^2_{\nu_t}$ for every ν_t. The parameters are estimated by optimizing a ML function, made up of the innovations and their variances. The equations of the Kalman filter are evaluated recursively, while different actions are taken depending on whether or not n_t is available at a time step. Starting with the initial conditions $\hat{\tilde{n}}_0$ and $\sigma^2_{e_0}$, the following equations are evaluated in the so-called time update [after *Bierkens et al.*, 1999]:

$$\begin{aligned} \bar{\tilde{n}}_t &= \phi \hat{\tilde{n}}_{t-1} \\ \sigma^2_{e_t} &= \sigma_a^2 + \phi^2 \sigma^2_{e_{t-1}} \end{aligned} \tag{3}$$

where $\bar{\tilde{n}}_t$ is the prediction of $\tilde{n}_t$ in the time update, $\hat{\tilde{n}}_t$ is the prediction of $\tilde{n}_t$ in the measurement update, and $\sigma^2_{e_t}$ is the variance of the error in the time update.

[9] When an observation is available, the measurement update is evaluated:

$$\begin{aligned} \nu_t &= \tilde{n}_t - \bar{\bar{n}}_t \\ \sigma^2_{\nu_t} &= \sigma^2_{e_t} \\ \hat{\tilde{n}}_t &= \tilde{n}_t \\ \sigma^2_{e_t} &= 0 \end{aligned} \tag{4}$$

If, however, no observation is available:

$$\hat{\tilde{n}}_t = \bar{\bar{n}}_t \tag{5}$$

The recursive application of equations (3), (4), and (5) mathematically equals the forecasting mode of an AR(1) model as described by (2) because after a time lag l, the error in the time update is

$$\sigma^2_{e_{t+l}} = \sigma^2_a\left(1 + \phi^2 + \ldots + \phi^{2l-2}\right) = \sigma^2_a \frac{\left(1 - \phi^{2l}\right)}{\left(1 - \phi^2\right)} \tag{6}$$

Hence in the SSR "the forecasting mode" of the AR(1) model that handles predictions over variable lead times is part of the basic model equations. However, even in the degenerate form, the recursive evaluation of the equations in SSR becomes computationally increasingly inefficient with an increasing model frequency. Even more so when it is used in combination with a deterministic model, because also that has to operate on the same frequency. Next to that, the autoregressive parameter ϕ becomes badly scaled as it asymptotically approaches the value of 1 when the time step Δt (in real time units) approaches 0. Finally, operating on a high frequency also poses a problem for ARMA type transfer models, because the number of MA parameters increases linearly with the frequency [*von Asmuth et al.*, 2002].

2.3. Ornstein-Uhlenbeck–Based Model

[10] Figure 3 gives a schematic representation of the functioning of the OUB noise model. In this approach all time series and functions, including the noise process, are regarded as being continuous. Equation (1) can be transformed to continuous time by writing it in moving-average form:

$$\tilde{n}_t = \sum_{i=0}^{\infty} \phi^i a_{t-i} \tag{7}$$

and replacing a_t by a continuous white noise process $\mathrm{d}W(t)$. The residual series $n(t)$ can then be modeled as a continuous stochastic process, which is given by

$$\tilde{n}(t) = \int_{-\infty}^{t} \varphi(t-\tau)\mathrm{d}W(\tau) \tag{8}$$

where $\varphi(t)$ is the noise impulse response function and $W(t)$ is the Wiener process [L] with properties, $\mathrm{E}\{\mathrm{d}W(t)\} = 0$, $\mathrm{E}[\{\mathrm{d}W(t)\}^2] = c\mathrm{d}t$, $\mathrm{E}[\{\mathrm{d}W(t_1)\}\mathrm{d}W(t_2)] = 0$, $t_1 \neq t_2$, where c is a constant [$\mathrm{L}^2\ \mathrm{T}^{-1}$]. As in the Kalman filter approach, the model does not yield an estimate of the white noise series $\mathrm{d}W(t)$, but instead gives an estimate of the innovation series $\nu(t)$. $\nu(t)$ is modeled as the irregularly sampled effect of the noise process on the residual series between time steps $t - \Delta t$ and t, which is given by

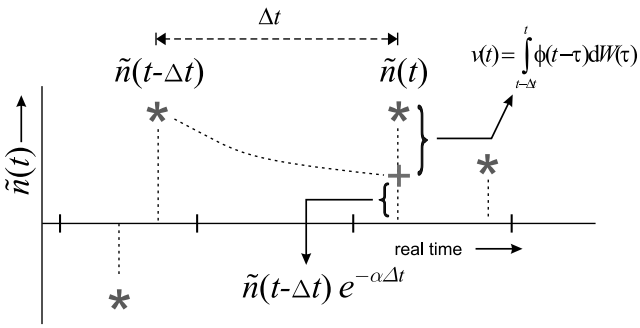

Figure 3. Schematical representation of the functioning of an exponential noise model in continuous time. The star denotes a residual, whereas the plus denotes a predicted residual. The model directly predicts the value of the residual series for the next observation available.

$$\nu(t) = \int_{t-\Delta t}^{t} \varphi(t-\tau)\mathrm{d}W(\tau) \tag{9}$$

For the noise impulse response (IR) function the following exponential is chosen, so that (8) reduces to an AR(1) model when it used on data with regular time steps:

$$\varphi(t) = \sqrt{\frac{2\alpha\sigma^2_n}{c}} e^{-\alpha t} \tag{10}$$

with α defining the decay rate of the noise, and σ^2_n denoting the variance of the residuals. With the choice of an exponential IR function, equation (8) can be written as [see, e.g., *Gardiner*, 1994]

$$\tilde{n}(t) = e^{-\alpha\Delta t}\tilde{n}(t-\Delta t) + \int_{t-\Delta t}^{t} \sqrt{\frac{2\alpha\sigma^2_n}{c}} e^{-\alpha(t-\tau)}\mathrm{d}W(\tau) \tag{11}$$

which is known as an Ornstein-Uhlenbeck process [*Uhlenbeck and Ornstein*, 1930]. By combining equation (9) and (11), the innovation series $\nu(t)$ can be calculated from the available residuals using simply

$$v(t) = \tilde{n}(t) - \mathrm{e}^{-\alpha\Delta t}\tilde{n}(t-\Delta t) \tag{12}$$

where the last term on the right side equals $\hat{\tilde{n}}_{t|\tilde{n}_{t-\Delta t}}$ (in Figure 3 illustrated by the dashed exponential). Because of its continuous formulation, (12) gives an exact solution and only needs to be evaluated once for every observation of n_t. This can reduce computation times substantially, compared

to a recursive algorithm which discretizes Δt in small time steps.

2.4. Parameter Estimation Method and Derivation of the SWSI Criterion

[11] Next to the equations discussed in the previous paragraph, an important aspect of dealing with irregular data lies in the optimization methods used. For fixed time steps, σ_a^2 is constant and minimizing the sum of squares of a_t yields exact ML estimates, under the assumption of a Gaussian distribution. This is, however, also conditional on the choice of the starting values a_0 and n_0, and is therefore called the conditional sum of squares [*Box and Jenkins*, 1970]. With a variable $\sigma_{v(t)}^2$, simple least squares methods are no longer straightforwardly applicable. Instead, a likelihood function is constructed and maximized using the Kalman filter estimates of $v(t)$ and $\sigma_{v(t)}^2$ at every time step [see, e.g., *Schweppe*, 1973; *Mélard*, 1984]. The likelihood function, however, is a rather complicated expression and flexible algorithms are needed in order to maximize it [*Hipel and McLeod*, 1994]. In the following, we will show that also for irregular data, a simple least squares criterion can be derived using the innovation variance function (IVF) that defines the general relationship between $\sigma_{v(t)}^2$ and Δt. For ARIMA models, when the set of parameters that is estimated is $\beta = \{\beta_1, \beta_2, .., \beta_p\}^T$, and assuming a_t is a time series of N random variables which are normally independently distributed (NID(0, σ_a^2)), then their joint pdf can be written as:

$$P(a_t|\beta, a_0, n_0) = \left(2\pi\sigma_a^2\right)^{-\frac{N}{2}} \exp\left(\sum_{i=1}^{N} \frac{-a_t^2}{2\sigma_a^2}\right) \tag{13}$$

in which a_0 and n_0 denote the initial conditions of a_t and n_t. Using (13) the log likelihood function of β is given by

$$L(\beta|a_t, a_0, n_0) = -0.5N\ln(2\pi) - 0.5N\ln\left(\sigma_a^2\right) - 0.5\left(\sum_{i=1}^{N} \frac{a_t^2}{\sigma_a^2}\right) \tag{14}$$

For the likelihood function of continuous time innovations, we first adapt the discrete notation to allow for discretely sampled, irregularly spaced observations. For this reason we use t in real time, and index that with i instead. A data set O of N observations of a continuous process like $n(t)$ is then given by

$$O = [n(t_1), n(t_2), \ldots\ldots\ldots, n(t_N)] \tag{15}$$

As stated earlier, the innovation variances $\sigma_{v(t)}^2$ that are otherwise individually estimated and stored for every time step, can be jointly described by the IVF as a function of the time step Δt (and the parameter set β). In the example application, we will pay more attention to the IVF as it plays a crucial function in handling irregular data. The likelihood function can now be written as [*Schweppe*, 1973]

$$L\{\beta|O\} = -0.5N\ln(2\pi) - 0.5\sum_{i=1}^{N} \ln\left\{\sigma_v^2(\Delta t_i, \beta)\right\} - 0.5\sum_{i=1}^{N} \frac{v^2(t_i, \beta)}{\sigma_v^2(\Delta t_i, \beta)} \tag{16}$$

In order to reduce the amount of parameters that has to be numerically optimized, we will start with eliminating $E(n_t)$. For sample sizes usually considered, the mean of a time series can be adequately estimated as [*Box and Jenkins*, 1970]:

$$E\{n(t)\} = \frac{\sum_{i=1}^{N} n(t_i)}{N} \tag{17}$$

Hence we can estimate $\tilde{n}(t)$ directly from data set O. Because $\sigma_v^2(\Delta t, \beta)$ is a function of the time step, it cannot be straightforwardly estimated from the available innovations. However, using (9) and (10), $\sigma_v^2(\Delta t, \beta)$ can be written as a function of the residual variance as [*Gardiner*, 1994]:

$$\sigma_v^2(\Delta t, \beta) = \left(1 - e^{-2\alpha\Delta t}\right)\sigma_n^2(\beta) \tag{18}$$

which in turn can yield an estimator for $\sigma_n^2(\beta)$ using the individual innovations (see Appendix A):

$$\sigma_n^2(\beta) = \frac{\sum_{i=1}^{N} \left(\frac{1}{1-e^{-2\alpha\Delta t_i}}\right) v^2(t_i, \beta)}{N} \tag{19}$$

With (18) and (19) in (16), we can now eliminate $\sigma_v^2(\Delta t, \beta)$ form the equations, and the likelihood function can be written as (see Appendix A)

$$L\{\beta|O\} = -0.5N\ln(2\pi) - 0.5N\ln\left[\frac{\sum_{j=1}^{N} \frac{\sqrt[N]{\prod_{i=1}^{N}\left(1 - e^{-2\alpha\Delta t_i}\right)}}{1 - e^{-2\alpha\Delta t_j}} v^2(t_j, \beta)}{N}\right] - 0.5N \tag{20}$$

Thus the only parameter that has to be numerically optimized for the noise model is α, defining the decay rate of the noise. Because the first and last terms in the likelihood function are constant, an estimate of β can be obtained by minimizing the following criterion:

$$S^2\{\beta|O\} = \sum_{j=1}^{N} \frac{\sqrt[N]{\prod_{i=1}^{N}\left(1 - e^{-2\alpha\Delta t_i}\right)}}{1 - e^{-2\alpha\Delta t_j}} v^2(t_j, \beta) \tag{21}$$

which may be referred to as the sum of weighted squared innovations (SWSI) criterion. The SWSI criterion is similar to other weighted least squares criteria, as also here the weights reflect the variances of the innovations. For optimization and parameter estimation, we can now use standard nonlinear least squares regression methods [see, e.g., *Snedecor and Cochran*, 1967]. First, the partial derivatives of (21) are obtained, either numerically or analytically, and used to construct a Jacobian matrix:

$$J = \begin{bmatrix} \frac{\partial S_1}{\partial \beta_1} & \cdots & \frac{\partial S_1}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial S_N}{\partial \beta_1} & \cdots & \frac{\partial S_N}{\partial \beta_p} \end{bmatrix} \tag{22}$$

Using J the model is calibrated on a set of observations with a Levenberg-Marquardt optimization algorithm [*Marquardt*, 1963] that adjusts β, while minimizing $S\{\beta|O\}$. Then, the covariance matrix of the parameters can be estimated with

$$\sigma_\beta^2 = \frac{\min(S^2)}{N-p}(J'J)^{-1} \tag{23}$$

Next, σ_n^2 is estimated using (19). The model results can be checked by examining the covariance matrix and the variance of the IR function. The variance of the IR function $\varphi(t)$ equals

$$\mathrm{var}(\varphi(t)) = \left(\frac{\sqrt{2\sigma_n^2}}{2\sqrt{\alpha}}e^{-\alpha t} - t\sqrt{2\alpha\sigma_n^2}e^{-\alpha t}\right)^2 \mathrm{var}(\alpha) \tag{24}$$

A confidence interval for $\varphi(t)$ can be plotted as $\pm 2\sigma$, when assuming a normal distribution for α. As in the ARIMA approach, serious model inadequacy can be detected by examining the autocorrelation function of the innovation series $\nu(t)$ (which gives an indication on whether the white noise assumption holds) and the cross-correlation function between $\nu(t)$ and the transfer model input series $p(t)$ (which indicates whether there are still patterns left in the innovation series that could be explained by the input series). The autocorrelation and cross-correlation functions at lag k are defined in the same way as in discrete TFN models, but because of the irregularity of the time steps, a tolerance around lag k of $\pm 0.5k$ is implied. A proof for the fact that, like AR(1) models, the innovations of the continuous model are not autocorrelated, nor cross correlated with observations of the process $n(t)$ itself, is given in Appendix A. This makes the model suitable for simulation purposes, and also is a prerequisite for the parameter estimation process.

3. Summary of Method

[12] When the parameters of a combined deterministic or transfer model and noise model are estimated simultaneously, the procedure is as follows (see also Figure 4). First, using the initial values of β, the deterministic model is evaluated in order to get a time series of residuals. Under the assumption of exponential noise decay (equation (10)), the innovation series is obtained from the residual series using equation (12). Next, the SWSI criterion is calculated by weighting the squared innovations according to their respective variances and relative to the geometrical average of the innovation variances for all time steps (equation (21)). Subsequently, the parameter set β is estimated by minimizing the criterion using a Levenberg-Marquardt algorithm. Finally the validity and performance of the model are checked using the autocorrelation and cross-correlation functions of the innovations, the correlation matrix of the model parameters and the variance of the IR function.

4. Example Application

4.1. Description of Setup and Data Set

[13] The effectiveness of a combined AR(1) model and Kalman filter in modeling irregularly observed hydrological data has already been shown rather elaborately [*Bierkens et al.*, 1999; *Berendrecht et al.*, 2003; *Yi and Lee*, 2004]. Therefore we will not focus on that here. The mathematical

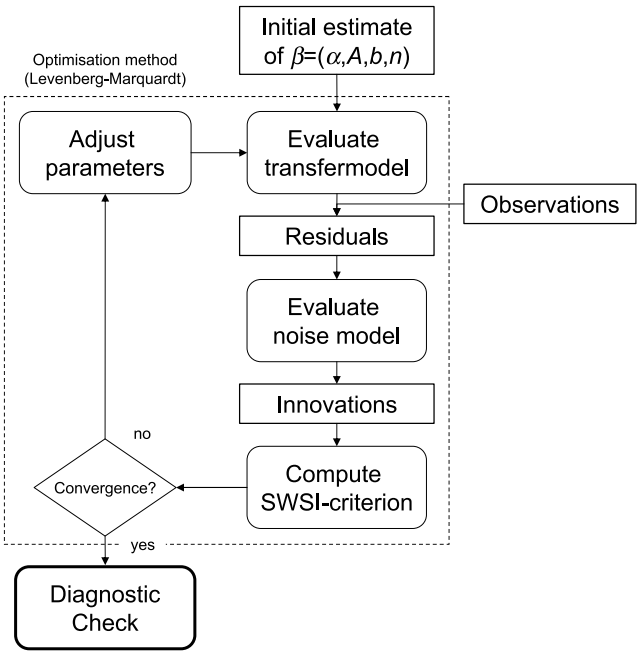

Figure 4. Procedure for applying the noise model in combination with a deterministic model.

equivalency of the OUB model and the AR(1) model are straightforward, so logically its results will also be equivalent. Instead, we will first focus on comparing the SWSI criterion with the likelihood function. Logically, if the minima of both criteria are identical, they will yield identical parameter estimates and therefore also identical model predictions. Second, we will examine the innovation variance function, which is crucial for handling irregular data in both the OUB and the Kalman filter approach. We will illustrate that also in this case, the model effectively whitens the residuals and the assumption of exponential noise decay is therefore valid. As a test case, we will use the noise model in conjunction with a continuous transfer function model in the context for which it was developed, i.e., modeling the residuals of groundwater level observations. In the simple case where groundwater level fluctuations are influenced by precipitation surplus only, the combined transfer function noise model is given by:

$$h(t) = \int_{-\infty}^{t} p(\tau)\theta(t-\tau)d\tau + n(t) + d \tag{25}$$

where

- $h(t)$ the observed groundwater level at time t [T], relative to some reference level [L];
- $p(t)$ the precipitation surplus at time t [L];
- $\theta(t)$ impulse response function (dimensionless), for which a Pearson type III distribution function is chosen, i.e., $\theta(t) = A\frac{b^n t^{n-1}\exp(-bt)}{\Gamma(n)}$;
- $n(t)$ the residual series [L], modeled with an Ornstein-Uhlenbeck based model conform (11);
- d the local drainage level, relative to some reference level [L].

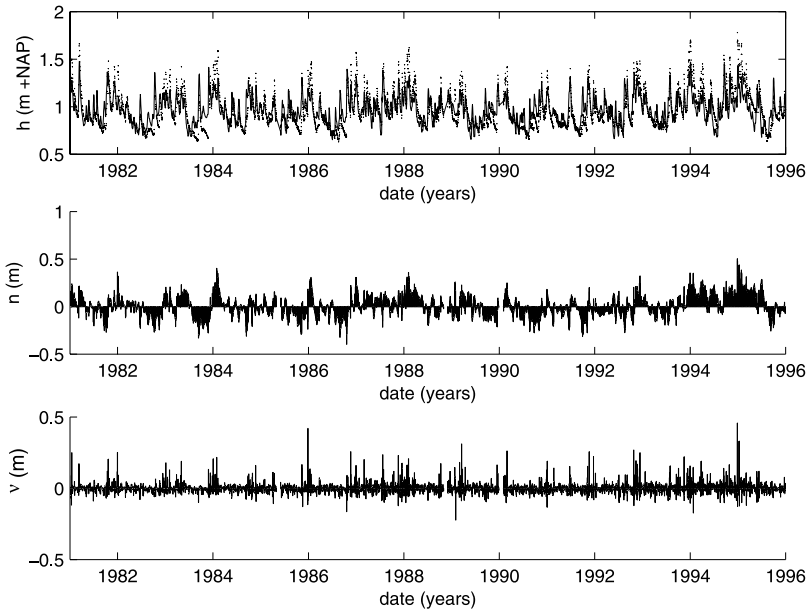

Figure 5. Time plot of the available groundwater level observations (dots) from piezometer 32cl0034 relative to the national reference level (h+NAP), the predictions of the transfer model (solid line), the model residuals (n), and the innovation series (ν).

More details on the background of the transfer function model are given by *von Asmuth et al.* [2002] and *von Asmuth and Knotters* [2004]. The TFN model is calibrated on a 15-year (1981–1996) groundwater level series observed with a daily frequency. However, this series is not totally complete but 8.75% of the observations are missing. The series originates from a piezometer located on the main meteorological field of the Royal Dutch Meteorological Institute at the town of De Bilt in the center of the Netherlands [see also *Bierkens et al.*, 1999]. The precipitation surplus is obtained from daily averaged observations of precipitation and potential evapotranspiration at the meteorological field. A time plot of the available groundwater level observations is given in Figure 5.

[14] The parameter set β of the combined model is integrally estimated using a numerically derived Jacobian matrix and the methods described in the previous paragraph. As d is eliminated from the equations, the parameters that have to be estimated are (A, b, n) from the transfer model, along with α from the noise model. For this purpose, all available observations in the period were used. The results of the TFN model are given in Figure 5, as time plots of the observations and predictions, and of the residuals and innovations. The parameter estimates and calibration results (percentage of variance accounted for (R^2_{adj}), root mean squared error (RMSE), root mean squared innovation (RMSI)) are listed in Table 1. The autocorrelation function of the innovations, for which a time lag increment of one month is chosen in order to reveal seasonal patterns in the autocorrelation, indicates that the white noise assumption holds (Figure 6). Because of the large number of available observations, the autocorrelation function is rather smooth and the accompanying confidence interval narrow.

4.2. Comparison of the Likelihood Function and the SWSI Criterion

[15] Plots of both functions were made by varying two of the parameters around their estimated value, while keeping the others constant. The parameters were varied ±0.5 half their estimated value. In order to make both functions comparable, they were normalized by respectively setting their maximum and minimum to zero, and dividing the result by the range. Figure 7 gives a contour plot of the normalized log likelihood function and SWSI criterion for the first two parameters of the transfer model (A and b). Figure 7 shows that the contours of both objective functions are almost identical, which confirms that the SWSI criterion yields almost exact ML estimates. This can be expected, as the only difference between the ML function and SWSI criterion is that in first case σ_a^2 (for the minimal time step) is part of the parameter set to be estimated, while in the SWSI criterion σ_n^2 is assumed known and implicitly approximated with Equation (19). This approximation converges to the true value of σ_n^2 for large N. Consequently, a comparison of σ_a^2 (β) estimated as a parameter with the likelihood function or estimated afterward using the SWSI approach will exemplify the likeness of both criteria for all parameters. Using the SWSI criterion, σ_a^2 (β) is estimated from the data in the following way:

$$\sigma_a^2(\beta) = \frac{\sum_{i=1}^{n} \left(\frac{1-e^{-2\alpha}}{1-e^{-2\alpha\Delta t_i}} \right) \nu^2(t_i, \beta)}{n} \tag{26}$$

Table 1. Calibration Results, Estimated Parameters, and Characteristics for Piezometer 32cl0034

	Value
R^2_{adj}	66.5%
RMSE	11.0 cm
RMSI	3.6 cm
A (±2σ)	61.6 (±4.2) day
b (±2σ)	0.28 (±0.02) day^{-1}
n (±2σ)	1.88 (±0.06)
α (±2σ)	18.2 (±3.2) day^{-1}

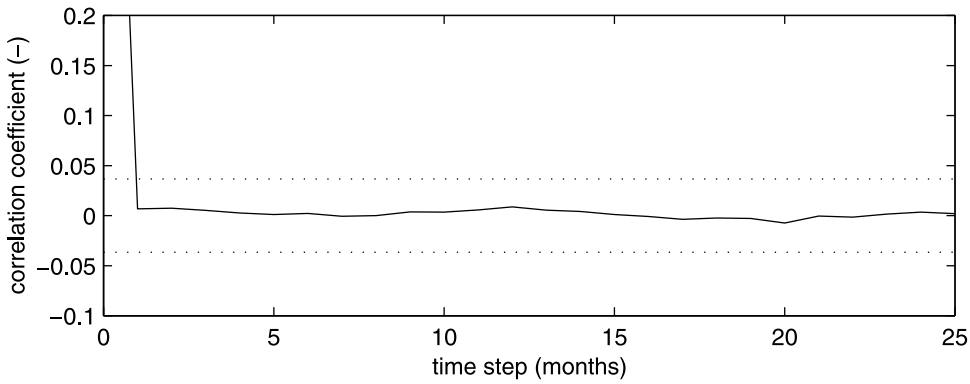

Figure 6. Autocorrelation function of the innovations of the OUB model. The dotted lines denote the 95% confidence interval. The autocorrelation function indicates that the white noise assumption holds.

which is equivalent to equation (6) for fixed Δt when $\frac{\sum_{i=1}^{n} \nu^2(t_i, \beta)}{n}$ is replaced by $\sigma_\nu^2(\beta)$, and $\phi = e^{-\alpha}$. In Figure 8 the likelihood function of the innovations is plotted and compared to $\sigma_a^2(\beta)$ obtained with equation (26) (i.e., 0.0013 m^2). From Figure 8 the estimate of $\sigma_a^2(\beta)$ proves to coincide well with the maximum of the likelihood function for this parameter.

4.3. A Check on the Innovation Variance Function

[16] Because the IVF(18) is a key factor in both the derivation of the SWSI criterion and in general for stochastic modeling at mixed frequencies, in the following we will examine it more closely using the data of paragraph 4.2. For this purpose, the groundwater level series was resampled at frequencies Δt = [1, 2, 3.....75] days. The transfer and noise model were evaluated using the estimate of β obtained with the daily series, which logically should be optimal. Thus a residual and innovation series were obtained for every sample frequency. As stated earlier, the daily series was not complete and therefore also the resampled series were not. Because of this, only those innovations and residuals were selected from the resampled series for which Δt_i equaled the sampling frequency exactly. Both the innovation and residual variance of the resulting series were estimated and plotted in Figure 9. In Figure 9, equation (18), being the theoretical IVF, is plotted together with the 95% confidence interval of the estimates. The latter is given by [*Snedecor and Cochran*, 1967]

$$\frac{\sigma_v^2(\Delta t_i, \beta)*(N-1)}{\chi^2_{0.975}} <= \sigma_v^2(\Delta t_i, \beta) <= \frac{\sigma_v^2(\Delta t_i, \beta)*(N-1)}{\chi^2_{0.025}} \quad (27)$$

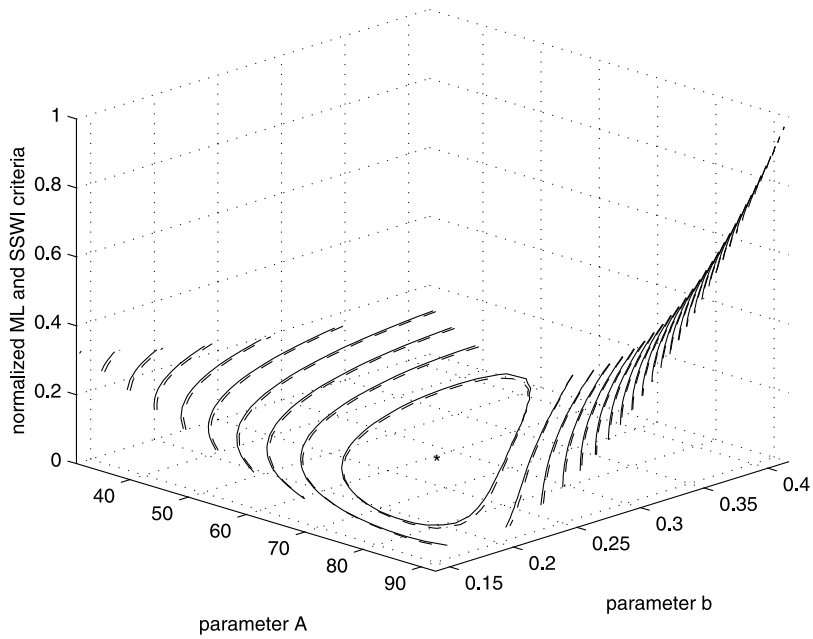

Figure 7. Contour plot of the normalized log likelihood function of the innovations (dashed lines) and the SSWI criterion (solid lines) as a function of the parameters *A* and *b* (all other things kept constant). The star denotes the estimated values. It can be seen that both functions are almost identical.

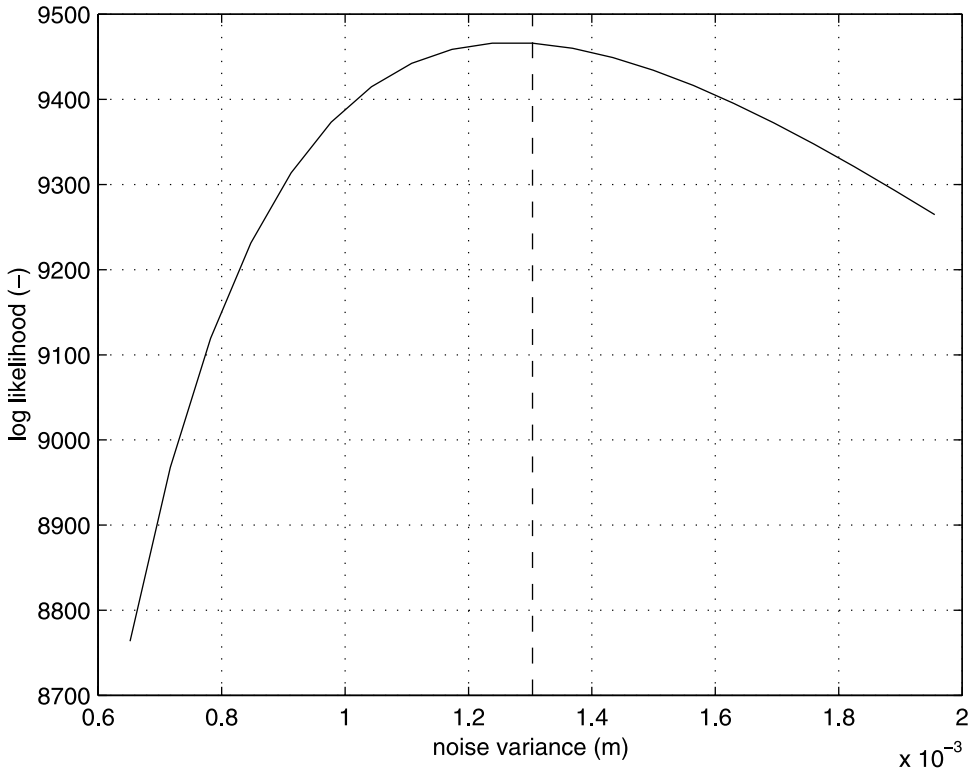

Figure 8. Plot of the log likelihood function of the innovations as a function of the parameter σ_a^2. The dashed line denotes the value of $\sigma_a^2(\beta)$ obtained with the SSWI criterion. The minimum of the log likelihood function proves to coincide well with the value obtained from the SSWI criterion.

Figure 9 thus allows for a comparison between the estimated and theoretical IVF. A check on whether the estimated variances differ significantly from the theoretical IVF could well be used as a further diagnostic check on the validity of the model, next to checking the autocorrelation and cross-correlation functions. From Figure 9 it is concluded that the estimates of the innovation variance σ_v^2 (Δt, β) do not significantly differ from the theoretical IVF, although the agreement is less for the intermediate time steps. This points to the fact that the parameter α, defining the decay rate of the noise model, could be sensitive to the sampling frequency. Therefore, special care has to be taken when the assumption of exponential decay for the noise model does not, or not well, hold and/or when the model is used for simulations at a different frequency than that of the data with which it was calibrated. A standard diagnostic check like the autocorrelation function of the innovations will not provide enough information on this point.

5. Discussion and Conclusions

[17] A comparison of the AR(1) model, conventional or embedded in a Kalman filter, and the Ornstein-Uhlenbeck–based noise model shows that their equations for respectively the forecasting mode in the AR(1) model, those handling the predictions and their variance in the time update of the degenerate Kalman filter, and the continuous time innovations are mathematically equivalent. The continuous equations, however, are more general, give an exact solution and are computationally more efficient as they are not evaluated recursively. Because of its simplicity, the OUB model can be easily implemented for modeling irregularly spaced errors of various deterministic models. The continuous time approach, however, makes it especially suited for combination with analytic models or continuous time transfer models. A restriction of the OUB model in its present form, is that it is limited to processes that show exponential decay. Work is needed to further generalize and test the approach, for example using continuous time

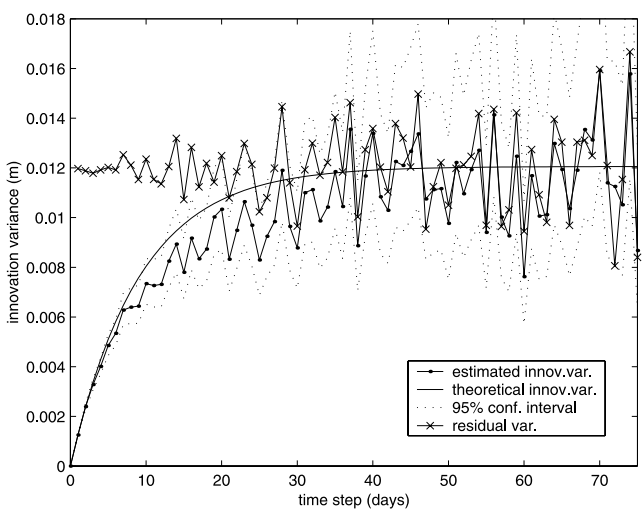

Figure 9. Plot of the estimated and theoretical IVF as a function of the time step and of the estimated residual variance. In general, the estimated innovation variances do not differ significantly from the theoretical IVF, although the agreement is better for smaller time steps.

ARMA models in general [e.g., *Brockwell*, 2001], non-Gaussian OUB models [*Barndorff-Nielsen and Shephard*, 2001] or by superposition of OU type processes [*Barndorff-Nielsen*, 1999].

[18] It was shown that from the likelihood function normally used for irregularly spaced innovations, a weighted least squares (SWSI) criterion can be derived. The main difference between both functions is that in the first case, the innovation variances are calculated by relating them to the innovation variance of the minimal time step, which itself is estimated as a parameter, and stored for every time step. In the latter case, the innovation variances are implicitly related to the variance of the residuals, and in turn approximated implicitly with equation (19). This approximation converges to the true value for large sample sizes, so parameter estimates that are based on the SWSI criterion should approximate the ML estimates well for sample sizes usually considered. With the aid of the SWSI criterion, the parameters and their covariance matrix can be estimated using standard nonlinear least squares regression methods. The use of such methods is more efficient than minimizing the log likelihood function with some global optimization algorithm, and a Monte Carlo approach to estimating the parameter variances. The SWSI criterion itself could also be used for optimizing AR(1) models, in the conventional sense or embedded in a Kalman filter.

[19] By comparing the estimated innovation variances for different time steps with the theoretical innovation variance function (IVF), the validity of exponential decay for the noise model can be checked. The impulse response function is reflected in the behavior of the IVF. Thus a plot of both functions and the accompanying confidence interval is useful as a diagnostic check. Such a check provides additional information on the validity of the model, especially in light of its application at mixed frequencies.

Appendix A: Derivations

A1. Equation for the Autocorrelation and Cross-Correlation Functions of the Residuals and Innovations

[20] The autocorrelation of the innovations equals (with $t'' < t' < t$)

$$E\{\nu(t)\nu(t')\} = E\left\{\int_{t'}^{t}\sqrt{\frac{2\alpha\sigma_n^2}{c}}e^{-\alpha(t-\tau)}\mathrm{dW}(\tau)\cdot\int_{t''}^{t'}\sqrt{\frac{2\alpha\sigma_n^2}{c}}e^{-\alpha(t-\tau)}\mathrm{dW}(\tau)\right\}=0 \tag{A1}$$

due to the properties of the Wiener process ($E\{\mathrm{dW}(t)\,\mathrm{dW}(t')\} = 0$ if $t \neq t'$). Using (11), the autocorrelation of the residuals can be written as (for $t' < t$)

$$E\{\tilde{n}(t)\tilde{n}(t')\} = E\left\{e^{-\alpha(t-t')}\tilde{n}(t')\tilde{n}(t') + \tilde{n}(t')\cdot\int_{t-t'}^{t}\sqrt{\frac{2\alpha\sigma_n^2}{c}}e^{-\alpha(t-\tau)}\mathrm{dW}(\tau)\right\} \tag{A2}$$

which gives

$$E\{\tilde{n}(t)\tilde{n}(t')\} = e^{-a(t-t')}\sigma_n^2 \tag{A3}$$

Using (12), the cross correlation between residuals and innovations equals, for any $t' < t$,

$$E\{\nu(t)\tilde{n}(t')\} = E\{\tilde{n}(t)\tilde{n}(t') - \mathrm{e}^{-\alpha\Delta t}\tilde{n}(t-\Delta t)\tilde{n}(t')\} \tag{A4}$$

which gives, using (A3),

$$E\{\nu(t)\tilde{n}(t')\} = e^{-a(t-t')}\sigma_n^2 - \mathrm{e}^{-\alpha\Delta t}e^{-a(t-\Delta t-t')}\sigma_n^2 = 0 \tag{A5}$$

A2. Equation Relating the Residual Variance to Individual Innovations

[21] Starting with (18), we can write the innovation variance as an expected value and get

$$\sigma_n^2(\beta) = \left(\frac{1}{1-e^{-2\alpha\Delta t_i}}\right)\mathrm{E}\{\nu^2(t_i,\beta)\} \tag{A6}$$

using every $\nu(t_i)$ individually, N single sample estimates of $\hat{\sigma}_n^2(t_i, \beta)$ can be obtained with

$$\hat{\sigma}_{n,t_i}^2(\beta) = \left(\frac{1}{1-e^{-2\alpha\Delta t_i}}\right)\nu^2(t_i,\beta) \tag{A7}$$

next, we can get a more accurate estimate if we average the N estimates of $\hat{\sigma}_n^2(t_i, \beta)$:

$$\hat{\sigma}_n^2(\beta) = \frac{\sum_{i=1}^{N}\left(\frac{1}{1-e^{-2\alpha\Delta t_i}}\right)\nu^2(t_i,\beta)}{N} \tag{A8}$$

A3. Derivation of the Sum of Squared Weighted Innovations Criterion

[22] Given the following likelihood function,

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5\sum_{i=1}^{N}\ln\{\sigma_\nu^2(\Delta t_i,\beta)\} - 0.5\sum_{i=1}^{N}\frac{\nu^2(t_i,\beta)}{\sigma_\nu^2(\Delta t_i,\beta)} \tag{A9}$$

we can replace $\sigma_\nu^2(\Delta t_i, \beta)$ by $\sigma_n^2(\beta)$ using (18) and get

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5\sum_{i=1}^{N}\ln\{(1-e^{-2\alpha\Delta t_i})\sigma_n^2\} - 0.5\sum_{i=1}^{N}\frac{\nu^2(t_i,\beta)}{(1-e^{-2\alpha\Delta t_i})\sigma_n^2} \tag{A10}$$

Next, we get, by placing $N/\sigma_n^2(\beta)$ in the last term outside the summation sign,

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5\sum_{i=1}^{N}\ln\{(1-e^{-2\alpha\Delta t_i})\sigma_n^2\} - 0.5\frac{N}{\sigma_n^2}\frac{\sum_{i=1}^{N}\left(\frac{1}{1-e^{-2\alpha\Delta t_i}}\right)\nu^2(t_i,\beta)}{N} \tag{A11}$$

and replace in both terms $\sigma_n^2(\beta)$ by (A8) to get

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5\sum_{i=1}^{N}\ln\left\{(1-e^{-2\alpha\Delta t_i})\frac{\sum_{j=1}^{N}\frac{1}{1-e^{-2\alpha\Delta t_j}}\nu^2(t_j,\beta)}{N}\right\} - 0.5N \tag{A12}$$

In the last term, several items have thus been eliminated. Because the sum of several logarithms is the logarithm of their product, (A12) equals

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5\ln\left[\prod_{i=1}^{N}\left\{\left(1-e^{-2\alpha\Delta t_i}\right)\cdot\frac{\sum_{j=1}^{N}\frac{1}{1-e^{-2\alpha\Delta t_j}}\nu^2\left(t_j,\beta\right)}{N}\right\}\right] - 0.5N \tag{A13}$$

As it is constant for any set of the model parameters $\beta|n(t_i)$, the summation term can be placed outside the product sign, giving

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5\ln\left[\left\{\frac{\sum_{j=1}^{N}\frac{1}{1-e^{-2\alpha\Delta t_j}}\nu^2\left(t_j,\beta\right)}{N}\right\}^N\cdot\prod_{i=1}^{N}\left\{\left(1-e^{-2\alpha\Delta t_i}\right)\right\}\right] - 0.5N \tag{A14}$$

which equals, by taking an *Nth* power root to the power of N, and placing N outside the logarithm

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5N\ln\left[\frac{\sum_{j=1}^{N}\frac{1}{1-e^{-2\alpha\Delta t_j}}\nu^2\left(t_j,\beta\right)}{N}\cdot\sqrt[N]{\prod_{i=1}^{N}\left(1-e^{-2\alpha\Delta t_i}\right)}\right] - 0.5N \tag{A15}$$

The *Nth* power root of the product term is the geometrical mean which is also constant given $n(t_i)$ and can be placed inside the summation term:

$$J\{\beta|\mathrm{O}\} = -0.5N\ln(2\pi) - 0.5N\ln\left[\sum_{j=1}^{N}\frac{\sqrt[n]{\prod_{i=1}^{N}\left(1-e^{-2\alpha\Delta t_i}\right)}}{1-e^{-2\alpha\Delta t_j}}\nu^2\left(t_j,\beta\right)\cdot\frac{1}{N}\right] - 0.5N \tag{A16}$$

As the first and last terms of the likelihood function are now constant, $J\{\beta|n(t_i)\}$ can be maximized by minimizing a sum of weighted squared innovations:

$$S^2\{\beta|\mathrm{O}\} = \sum_{j=1}^{N}\frac{\sqrt[N]{\prod_{i=1}^{N}\left(1-e^{-2\alpha\Delta t_i}\right)}}{1-e^{-2\alpha\Delta t_j}}\nu^2\left(t_j,\beta\right) \tag{A17}$$

[23] **Acknowledgments.** The authors wish to thank A.W. Heemink for his valuable comments on an earlier version of this paper. We also thank the Associate Editors, D. Koutsoyiannis, and two anonymous reviewers for their comments, which helped to improve the clarity of the paper considerably.

References

Ahsan, M., and K. M. O'Connor (1994), A reappraisal of the Kalman filtering technique, as applied in river flow forecasting, *J. Hydrol.*, *161*, 197–226.

Barndorff-Nielsen, O. E. (1999), Superposition of Ornstein-Uhlenbeck type processes, *Theory Probab. Appl.*, *45*(2), 175–194.

Barndorff-Nielsen, O. E., and N. Shephard (2001), Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics, *J. R. Stat. Soc., Ser. B*, *63*, 1–42.

Berendrecht, W. L., A. W. Heemink, F. C. Van Geer, and J. C. Gehrels (2003), Decoupling of modeling and measuring interval in groundwater time series analysis based on response characteristics, *J. Hydrol.*, *278*, 1–16.

Bergstrom, A. R. (1990), *Continuous Time Econometric Modeling*, Oxford Univ. Press, New York.

Bierkens, M. F. P., M. Knotters, and F. C. Van Geer (1999), Calibration of transfer function-noise models to sparsely or irregularly observed time series, *Water Resour. Res.*, *32*, 1741–1750.

Box, G. E. P., and G. M. Jenkins (1970), *Time Series Analysis: Forecasting and Control*, Holden-Day, Boca Raton, Fla.

Brockwell, P. J. (2001), Continuous-time ARMA processes, in *Stochastic Processes: Theory and Methods*, edited by D. N. Shanbhag and C. R. Rao, *Handb. Stat.*, *19*, 249–276.

Brubacher, S. R., and G. T. Wilson (1976), Interpolating time series with applications to the estimation of holiday effects on electricity demand, *J. R. Stat. Soc., Ser. C*, *25*(2), 107–116.

Bryson, A. E., and L. J. Henrikson (1965), Estimation using sampled data containing correlated noise, *J. Spacecr. Rockets*, *5*, 662–665.

Gardiner, C. W. (1994), *Handbook of Stochastic Methods*, Springer, New York.

Harvey, A. C. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge Univ. Press, New York.

Harvey, A. C., and R. G. Pierse (1984), Estimating missing observations in economic time series, *J. Am. Stat. Assoc.*, *79*, 125–131.

Hipel, K. W., and A. I. McLeod (1994), *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, New York.

Jones, R. H. (1980), Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, *22*, 389–395.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35–45.

Kalnay, E. (2002), *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge Univ. Press, New York.

Khabie-Zeitoune, E. (1982), Prediction in continuous time, in *Time Series Analysis: Theory and Practice 2*, edited by O. D. Anderson, pp. 7–24, Elsevier, New York.

Knotters, M., and P. E. V. Van Walsum (1997), Estimating fluctuation quantities from time series of water-table depths using models with a stochastic component, *J. Hydrol.*, *197*, 25–46.

Koutsoyiannis, D. (2001), Coupling stochastic models of different time-scales, *Water Resour. Res.*, *37*(2), 379–391.

Little, R. J., and D. B. Rubin (1987), *Statistical Analysis With Missing Data*, John Wiley, Hoboken, N. J.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, *11*(2), 431–441.

Mélard, G. (1984), Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models, *J. R. Stat. Soc., Ser. C*, *33*, 104–114.

McLaughlin, D. (1995), Recent developments in hydrologic data assimilation, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994*, *Rev. Geophys.*, *33*, 977–984.

Schweppe, F. C. (1973), *Uncertain Dynamic Systems*, Prentice-Hall, Upper Saddle River, N. J.

Snedecor, G. W., and W. G. Cochran (1967), *Statistical Methods*, Iowa State Univ. Press, Ames.

Te Stroet, C. B. M. (1995), Calibration of stochastic groundwater models: Estimation of system noise statistics and model parameters, Ph.D. thesis, Delft Univ. of Technol., Delft, Netherlands.

Uhlenbeck, G. E., and L. S. Ornstein (1930), On the theory of the Brownian motion, *Phys. Rev.*, *36*, 823–841.

von Asmuth, J. R., and M. Knotters (2004), Characterising spatial differences in groundwater dynamics based on a system identification approach, *J. Hydrol.*, *296*, 118–134.

von Asmuth, J. R., M. F. P. Bierkens, and C. Maas (2002), Transfer function noise modeling in continuous time using predefined impulse response functions, *Water Resour. Res.*, *38*(12), 1287, doi:10.1029/2001WR001136.

Wiener, N. (1949), *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, MIT Press, Cambridge, Mass.

Yi, M. J., and K. K. Lee (2004), TFN modeling of irregularly observed groundwater heads, *J. Hydrol.*, *288*, 272–287.

M. F. P. Bierkens, Department of Physical Geography, Faculty of Geosciences, Utrecht University, 3508 TB, Utrecht, Netherlands.

J. R. von Asmuth, Water Resources Section, Faculty of Civil Engineering, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, Netherlands. (jos.von.asmuth@kiwa.nl)