

General Kriging for spatial-temporal processes with random ARX–regression parameters

J. Franke and B. Gründer

April 11, 1997

1 The data and the model

In the following, we discuss a procedure for interpolating a spatial-temporal stochastic process. We stick to a particular, moderately general model but the approach can be easily transferred to other similar problems. The original data, which motivated this work, are measurements of gas concentrations (SO_2 , NO , O_3) and several meteorological parameters (temperature, sun radiation, precipitation, wind speed etc.). These data have been and are still recorded twice every hour at several irregularly located places in the forests of the state Rheinland-Pfalz as part of a program monitoring the air pollution in forests. Let $\zeta(t, x_j)$, $j = 1, \dots, N$, $t = 0, \dots, T$, denote the observations of e.g. SO_2 concentration which we model as part of a spatial-temporal stochastic process $\zeta(t, x)$, $t \in \mathbb{Z}$, $x \in \mathbb{R}^2$. A particular feature is a large amount of data in the time direction (T very large), but only few locations in the plane where data are available (N small). A more detailed description of the data has been given in Franke and Gründer (1992) and Gründer (1992).

One of the goals which had to be achieved by modelling the data was a procedure for interpolating the gas concentration, i.e. for $x \notin \{x_1, \dots, x_N\}$, $\zeta(t, x)$ should be estimated from $\zeta(s, x_j)$, $0 \leq s \leq t$, $j = 1, \dots, N$. The procedure had to be adaptive with respect to new incoming data, and it should allow for the information contained in the meteorological observations. We start from the following model for the gas concentration (or some monotone normalizing transformation of it):

$$\zeta(t, x) = f(x)' \beta(t) + \eta(t, x), \quad t \in \mathbb{Z}, \quad x \in \mathbb{R}^2 \quad (1)$$

where $f(x)^T$ is a known vector of simple functions of x which allows for systematic differences between the various locations due, e.g., to the topography of the country. The vector $\beta(t)$ of regression coefficients forms a multivariate random time series which is independent of the residual

spatial-temporal process $\eta(t, x)$. We assume $\mathcal{E}\eta(t, x) = 0$ and that $\eta(t, x)$ is stationary in t and homogeneous in x , i.e. the joint distribution of $\eta(t_1, x_1), \dots, \eta(t_k, x_k)$ is invariant against a common translation of the arguments. The main assumption of model (1) is the presence of a global time-varying effect represented by $\beta(t)$ which influences the data at all locations x . Due to the size of the region and due to the specification of the places where the data come from, this assumption is satisfied for our data. $f(x)' \beta(t)$ explains even the major part of variability whereas $\eta(t, x)$ takes care of smaller local fluctuations only.

We assume that $\beta(t)$ is a m -variate autoregressive process of order p with an exogenous part of order r or an $\text{ARX}(p, r)$ -process:

$$\beta(t) = \sum_{j=1}^p A_j \beta(t-j) + \delta(t) + \sum_{j=1}^r C_j \xi(t-j) \quad (2)$$

$\delta(t)$ are i.i.d. m -dimensional Gaussian random vectors with mean 0 and covariance matrix Σ_δ . $\xi(t)$ is the l -dimensional vector of exogenous variables representing temperature, precipitation etc. at time t . The seasonality of these variables also takes care of the well-known seasonality in gas concentrations. We assume that for

$$A(z) = I_m - \sum_{j=1}^p A_j z^j, \quad C(z) = \sum_{j=1}^r C_j z^j$$

$\det A(z) \neq 0$ for all $|z| \leq 1$, and that $A(z)$, $C(z)$ have no common left-divisor (up to unimodular matrices).

Due to prevalent wind directions in the area under consideration, we cannot assume that the fluctuation process $\eta(t, x)$ is isotropic with respect to its spatial coordinates, i.e. the covariance $\text{cov}(\eta(t, x), \eta(t, y))$ does not only depend on the distance $\|y - x\|$. Looking at the rather scarce information on spatial dependence we cannot consider general homogeneous processes. We therefore follow Vecchia (1988) and consider so-called (α, λ) -isotropy, i.e., for fixed t , the covariance between $\eta(t, x)$ and $\eta(t, y)$ depends only on the norm of the suitably rotated and dilated difference $y - x$:

$$\text{cov}(\eta(t, x), \eta(t, y)) = c_\eta(\|S_\lambda R_\alpha(y - x)\|)$$

with

$$R_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \text{ and } S_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \text{ for } 0 \leq \alpha < \pi, \lambda > 0.$$

In practice, c_η would be a given function up to few unknown parameters which have to be estimated from the data.

With respect to the temporal dependence structure of $\eta(t, x)$, we assume that it is a *temporal Markovian process*, i.e. for all $(s_1, y_1), (s_2, y_2), \dots \in \{t, t-1, \dots\} \times \mathbb{R}^2$, $(s_j, y_j) \neq (t, x)$ for all j , we have

$$\mathcal{E}\{\eta(t, x) | \eta(s_j, y_j), j \geq 1\} = \mathcal{E}\{\eta(t, x) | \eta(s_j, y_j), s_j = t-1 \text{ or } s_j = t\}.$$

We even assume that $\eta(t, x)$ has an autoregressive structure of order 1 if we consider it only at a finite number of locations x_1, \dots, x_N :

$$Y(t) = L Y(t-1) + \varepsilon(t) \text{ with } Y(t)' = (\eta(t, x_1), \dots, \eta(t, x_N)). \quad (3)$$

$\varepsilon(t)$ is N -variate, mean-zero Gaussian white noise with covariance Σ_ε , and L satisfies the stationarity condition $\det(I_N - Lz) \neq 0$ for all $|z| \leq 1$.

2 Estimating the random regression coefficients

To interpolate $\zeta(t, x)$ using the past and present data at x_1, \dots, x_N we have to estimate $\beta(t)$. For this purpose, we consider a state-space representation of this ARX-process. Let $Z(t)' = (\zeta(t, x_1), \dots, \zeta(t, x_N))$ be the observations at time t , $F = (f(x_1), \dots, f(x_N))$ the $m \times N$ -matrix of regression functions evaluated at x_1, \dots, x_N , and $Y(t)$ be, as above, the fluctuation process at time t evaluated at x_1, \dots, x_N . Then we have from (1) the observation equation

$$Z(t) = F' \beta(t) + Y(t) \quad (4)$$

The ARX-equation (2) has the state-space representation (compare, e.g., Priestley, 1981)

$$b(t+1) = \mathbf{A} b(t) + \mathbf{D} \delta(t+1) + \mathbf{C} \xi(t) \quad (5)$$

with coefficient matrices \mathbf{A} , \mathbf{D} , \mathbf{C} of dimensions $mq \times mq$, $mq \times m$, $mq \times l$:

$$\mathbf{A} = \left(\begin{array}{c|ccc} A_1 & I_m & \dots & 0_m \\ A_2 & 0_m & & 0_m \\ \vdots & \vdots & \ddots & \vdots \\ A_{q-1} & 0_m & \dots & I_m \\ \hline A_q & 0_m & \dots & 0_m \end{array} \right), \quad \mathbf{D} = \begin{pmatrix} I_m \\ 0_m \\ \vdots \\ 0_m \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} C_1 \\ \vdots \\ \vdots \\ C_q \end{pmatrix}$$

Here, $q := \max(p, r)$, and we set $A_j = 0_m$ if $p < j \leq q$ and $C_j = 0$ if $r < j \leq q$. The first m coordinates of the state vector $b(t) \in \mathbb{R}^{mq}$ coincide with $\beta(t)$: $b_j(t) = \beta_j(t)$, $j = 1, \dots, m$. The remaining part of $b(t)$ can be easily written down recursively from (5), starting with

$$\begin{aligned} (b_{(m-1)q+1}(t), \dots, b_{mq}(t))' &= A_q \beta(t-1) + C_{q-1} \xi(t) \\ (b_{(m-2)q+1}(t), \dots, b_{(m-1)q}(t))' &= A_{q-1} \beta(t-1) + A_q \beta(t-2) \\ &\quad + C_{q-2} \xi(t) + C_{q-1} \xi(t-1) \end{aligned}$$

and so on. Setting $\mathbf{H} = F'(I_m 0_m \dots 0_m)$ we get from (4) the equation

$$Z(t) = \mathbf{H} b(t) + Y(t). \quad (6)$$

For some s let $b(t|s)$ denote the best (in mean-square sense) estimate for $b(t)$ based on the observations $Z(k)$ and the exogenous variables $\xi(k)$, $0 \leq k \leq s$. $b(t|s)$ may be calculated recursively in time using the Kalman filter, though the linear system given by (5), (6) is non-standard due to the presence of exogenous variables and the serial dependence of the observational noise $Y(t)$. We assume that we observe the system from time 0 onwards, that the initial state $b(0)$ is a Gaussian vector independent of $\rho(t)$, $t \geq 1$, and $Y(t)$, $t \geq 0$. Following Hannan and Deistler (1988, ch. 3.2), we treat the exogenous variables $\xi(t)$ as deterministic, i.e. all calculations are done conditionally on these data.

To formulate the algorithm, we need some notation: For $s < t$, $Z(t|s)$ is the best estimate for $Z(t)$ given $Z(k)$, $\xi(k)$, $k \leq s$.

$$d_b(t|s) = b(t) - b(t|s), \quad d_z(t|s) = Z(t) - Z(t|s)$$

are the corresponding estimation errors with covariance and cross covariance matrices

$$P(t|s) = \mathcal{E} d_b(t|s) d_b'(t|s), \quad R(t|s) = \mathcal{E} d_z(t|s) d_z'(t|s), \quad Q(t|s) = \mathcal{E} b(t) d_z'(t|s).$$

Theorem 2.1 *Let $b(0|-1) = \mathcal{E} b(0)$ and $P(0|-1)$ be the covariance matrix of $b(0)$. Under the conditions stated above for the linear system (5), (6), we have for $t \geq 1$*

$$\begin{aligned} b(t|t-1) &= \mathbf{A} b(t-1|t-1) + \mathbf{C} \xi(t-1) \\ b(t|t) &= b(t|t-1) + K(t) d_z(t|t-1) \end{aligned}$$

$$\begin{aligned} P(t|t-1) &= \mathbf{A} P(t-1|t-1) \mathbf{A}' + \mathbf{D} \Sigma_\delta \mathbf{D}' \\ P(t|t) &= P(t|t-1) - K(t) R(t|t-1) K(t)' \end{aligned}$$

$$d_z(t|t-1) = Z(t) - (\mathbf{H} \mathbf{A} - L \mathbf{H}) b(t-1|t-1) - L Z(t-1) - \mathbf{H} \mathbf{C} \xi(t-1)$$

$$\begin{aligned} K(t) &= Q(t|t-1) R(t|t-1)^{-1} \\ R(t|t-1) &= (\mathbf{H} \mathbf{A} - L \mathbf{H}) P(t-1|t-1) (\mathbf{H} \mathbf{A} - L \mathbf{H})' \\ &\quad + \mathbf{H} \mathbf{D} \Sigma_\delta \mathbf{D}' \mathbf{H}' + \Sigma_\varepsilon \\ Q(t|t-1) &= \mathbf{A} P(t-1|t-1) (\mathbf{H} \mathbf{A} - L \mathbf{H})' + \mathbf{D} \Sigma_\delta \mathbf{D}' \mathbf{H}' \end{aligned}$$

For starting the algorithm the relations

$$\begin{aligned} d_z(0|-1) &= Z(0) - \mathbf{H}b(0|-1), \quad R(0|-1) = \mathbf{H}P(0|-1)\mathbf{H}' + \Sigma_0, \\ Q(0|-1) &= P(0|-1)\mathbf{H}' \end{aligned}$$

may be used, where Σ_0 denotes the covariance matrix of $Y(0)$.

Proof: Let \mathcal{P}_s denote the orthogonal projection onto the span of $Z(0), \dots, Z(s), \xi(0), \dots, \xi(s)$. Using (3) and (6), we have

$$\begin{aligned} d_y(t|t-1) &:= Y(t) - \mathcal{P}_{t-1}Y(t) \\ &= Y(t) - \mathcal{P}_{t-1}(LZ(t-1) - L\mathbf{H}b(t-1) + \varepsilon(t)) \\ &= Y(t) - LZ(t-1) + L\mathbf{H}b(t-1|t-1) \\ &= -L\mathbf{H}d_b(t-1|t-1) + \varepsilon(t) \end{aligned}$$

Using (5) we immediately have

$$b(t|t-1) = \mathbf{A}b(t-1|t-1) + \mathbf{C}\xi(t-1)$$

and therefore the following three relations, using also (6),

$$d_b(t|t-1) = \mathbf{A}d_b(t-1|t-1) + \mathbf{D}\delta(t) \quad (7)$$

$$\begin{aligned} Z(t|t-1) &= \mathbf{H}b(t|t-1) + \mathbf{Y}(t|t-1) \\ &= \mathbf{H}(\mathbf{A}b(t-1|t-1) + \mathbf{C}\xi(t-1)) + LZ(t-1) \\ &\quad - L\mathbf{H}b(t-1|t-1) \\ &= \mathbf{\Delta}b(t-1|t-1) + \mathbf{H}\mathbf{C}\xi(t-1) + LZ(t-1) \\ d_z(t|t-1) &= \mathbf{\Delta}d_b(t-1|t-1) + \mathbf{H}\mathbf{D}\delta(t) + \varepsilon(t), \end{aligned} \quad (8)$$

where $\mathbf{\Delta} = \mathbf{H}\mathbf{A} - L\mathbf{H}$. (8) and the independence of $d_b(t-1|t-1)$, $\delta(t)$ and $\varepsilon(t)$ imply immediately the recursions for $R(t|t-1)$ and $Q(t|t-1)$. Now, the rest of the algorithm follows analogously as in Theorem 3.2.3 of Hannan and Deistler (1988) which includes our result for the special case $L = 0$. \blacksquare

In Gründer (1993) it is shown that the error covariance matrices $P(t|t)$ converge to some limit P_∞ in the strong sense:

$$\sum_{t=0}^{\infty} \|P(t|t) - P_\infty\| < \infty,$$

and the same type of limit behaviour is shared by the Kalman gain $K(t)$ and by the matrices $R(t|t-1)$, $Q(t|t-1)$. As a consequence of these limit results it follows that the spectral radius, i.e. the maximum of the absolute values of eigenvalues, of the matrices $\mathbf{A} - K(t)$ ($\mathbf{H}\mathbf{A} - L\mathbf{H}$) is less than 1 for all t large enough. This implies the asymptotic stability of the algorithm.

The following result allows for fixed interval smoothing, i.e. for calculating $b(t|s)$ for all $t \leq s$, s fixed. It is analogous to Theorem 3.2.2 of Hannan and Deistler (1988) with some slight modification necessary due

to the serial dependence of $Y(t)$. As there, $P(t+1|t+1)^{-1}$ is a generalized inverse of the error covariance matrix.

Theorem 2.2 *Let $\Delta = \mathbf{H}\mathbf{A} - L\mathbf{H}$, and let the assumption of Theorem 2.1 be satisfied. Then, for $s > t$*

$$\begin{aligned} b(t|s) &= b(t|t+1) + F(t)\{b(t+1|s) - b(t+1|t+1)\} \\ F(t) &= P(t|t)\{\mathbf{A} - K(t+1)\Delta\}' P(t+1|t+1)^{-1} \\ b(t|t+1) &= b(t|t) + P(t|t)\Delta'R(t+1|t)^{-1}dz(t+1|t). \end{aligned}$$

Proof: We give the proof only for the case $C \equiv 0$. Then, the general case follows exactly as in the proof of Theorem 3.2.3 of Hannan and Deistler (1988). As abbreviations, we use $e(t) = d_z(t|t-1)$ with the corresponding covariance matrix $R_t = R(t|t-1)$. From Theorem 2.1, (7) and (8) we have

$$\begin{aligned} d_b(t+1|t+1) &= d_b(t+1|t) - K(t+1)e(t+1) \\ &= \{\mathbf{A} - K(t+1)\Delta\}d_b(t|t) + \{I - K(t+1)\mathbf{H}\}\mathbf{D}\delta(t+1) \\ &\quad - K(t+1)\varepsilon(t+1) \end{aligned}$$

Using this relation and the orthogonality of $\delta(t+1)$, $\varepsilon(t+1)$ and $b(t)$ we get

$$\mathcal{E}b(t)d_b'(t+1|t+1) = \mathcal{E}b(t)d_b'(t|t)\{\mathbf{A} - K(t+1)\Delta\}' = P(t|t)\{\mathbf{A} - K(t+1)\Delta\}'$$

As, by definition, $\mathcal{E}b(t+1)d_b'(t+1|t+1) = P(t+1|t+1)$ we have

$$\mathcal{E}\{b(t) - F(t)b(t+1)\}d_b'(t+1|t+1) = 0$$

By iteration, using the above equation for $d_b(t+1|t+1)$ repeatedly,

$$\mathcal{E}\{b(t) - F(t)b(t+1)\}e'(j) = 0 \quad \text{for all } j > t+1. \quad (9)$$

As $e(t+j)$ spans the linear space, representing the new information at time $t+j$ and orthogonal to the past, we have for $t < s$

$$b(t|s) - b(t|t+1) = \sum_{j=t+2}^s \mathcal{E}\{b(t)e'(j)\}R_j^{-1}e(j)$$

and, using (9),

$$\begin{aligned} F(t)\{b(t+1|s) - b(t+1|t+1)\} &= F(t)\sum_{j=t+2}^s \mathcal{E}\{b(t+1)e'(j)\}R_j^{-1}e(j) \\ &= b(t|s) - b(t|t+1). \end{aligned}$$

Finally, again using (8),

$$\begin{aligned} b(t|t+1) &= b(t|t) + \mathcal{E} \{b(t) e'(t+1)\} R_{t+1}^{-1} e(t+1) \\ &= b(t|t) + P(t|t) \Delta' R_{t+1}^{-1} e(t+1) \end{aligned}$$

■

In practice, we do not know the parameters of the linear system given by (3), (5) and (6), but we have to work with estimates. Let us assume that we use the algorithm of Theorem 2.1 with approximations $\overline{\mathbf{A}}, \overline{L}, \overline{\mathbf{C}}, \overline{\Sigma}_\delta, \overline{\Sigma}_\varepsilon$ replacing the true system matrices. Then, the additional error in estimating the state $b(t)$ is of the same size as the approximation errors $\|\overline{\mathbf{A}} - \mathbf{A}\|, \|\overline{L} - L\|$ etc. To make this statement precise let us assume that we have sequences $\overline{\mathbf{A}}_n, \overline{L}_n, \overline{\mathbf{C}}_n, \overline{\Sigma}_{\delta,n}$ and $\overline{\Sigma}_{\varepsilon,n}$ converging for $n \rightarrow \infty$ to $\mathbf{A}, L, \mathbf{C}, \Sigma_\delta$ and Σ_ε at least with the rate $O(\nu_n)$ for some sequence $\nu_n \rightarrow 0$. Let $\overline{P}_n(t|t)$ be the error covariances matrices resulting in the use of $\overline{\mathbf{A}}_n$ etc. instead of \mathbf{A} etc. in the Kalman filter. Then:

Theorem 2.3 $\lim_{t \rightarrow \infty} \|P(t|t) - \overline{P}_n(t|t)\| = O(\nu_n)$

3 Spatial-temporal Kriging

Knowing how to estimate the random regression coefficients $\beta(t)$ of (1), we return to the original interpolation problem. We want to estimate $\zeta(t, x)$ from $\xi(s), \zeta(s, x_j), j = 1, \dots, N, s = 0, \dots, T$. Let \mathcal{A}_T denote the σ -algebra generated by the latter random variables. Furthermore, we use the following notation:

$$\begin{aligned} \mathbf{Z}(t) &= \begin{pmatrix} Z(0) \\ \vdots \\ Z(t) \end{pmatrix} = (\zeta(0, x_1), \dots, \zeta(0, x_N), \dots, \zeta(t, x_N))' \\ \mathbf{Y}(t) &= \begin{pmatrix} Y(0) \\ \vdots \\ Y(t) \end{pmatrix} = (\eta(0, x_1), \dots, \eta(0, x_N), \dots, \eta(t, x_N))' \\ \beta(t) &= \begin{pmatrix} \beta(0) \\ \vdots \\ \beta(t) \end{pmatrix} = (\beta_1(0), \dots, \beta_m(0), \dots, \beta_m(t))' \end{aligned}$$

$$\mathbf{F}_t = \begin{pmatrix} F & 0 & \dots & 0 \\ 0 & F & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & F \end{pmatrix} \text{ as a } m(t+1) \times N(t+1) \text{ - matrix}$$

In particular, we have from (4) the relation

$$\mathbf{Z}(t) = \mathbf{F}'_t \boldsymbol{\beta}(t) + \mathbf{Y}(t) .$$

Under the normality assumptions which we have imposed in paragraphs 1 and 2, the best estimate of $\zeta(t, x)$ is by (1)

$$\hat{\zeta}_T(t, x) = \mathcal{E}\{\zeta(t, x)|\mathcal{A}_T\} = f(x)' \mathcal{E}\{\boldsymbol{\beta}(t)|\mathcal{A}_T\} + \mathcal{E}\{\eta(t, x)|\mathcal{A}_T\}$$

where with \mathbf{S}_t denoting the covariance matrix of $\mathbf{Z}(t)$

$$\begin{aligned} \hat{\eta}_T(t, x) = \mathcal{E}\{\eta(t, x)|\mathcal{A}_T\} &= \mathcal{E}(\eta(t, x) \mathbf{Z}'(T)) \mathbf{S}_T^{-1} \mathbf{Z}(T) \\ &= \mathcal{E}(\eta(t, x) \mathbf{Y}'(T)) \mathbf{S}_T^{-1} \mathbf{Z}(T) \\ &=: g'_T(t, x) \mathbf{S}_T^{-1} \mathbf{Z}(T) \end{aligned}$$

Analogously, we have for the vector of sample residuals

$$\begin{aligned} \hat{\mathbf{Y}}(T) &= (\hat{\eta}_T(0, x_1), \dots, \hat{\eta}_T(T, x_N))' = \mathcal{E}\{\mathbf{Y}(t)|\mathcal{A}_T\} \\ &= \mathbf{G}_T \mathbf{S}_T^{-1} \mathbf{Z}(T) \end{aligned}$$

where \mathbf{G}_t denotes the covariance matrix of $\mathbf{Y}(t)$. Therefore, $\hat{\eta}_T(t, x)$ can also be written as a linear combination of the sample residuals

$$\hat{\eta}_T(t, x) = g'_T(t, x) \mathbf{G}_T^{-1} \hat{\mathbf{Y}}(T)$$

Otherwise, as

$$\hat{\eta}_T(t, x) = \mathcal{E}\{\zeta(t, x) - f'(x)\boldsymbol{\beta}(t)|\mathcal{A}_T\}$$

we have

$$\begin{aligned} \hat{\mathbf{Y}}(T) &= \mathcal{E}\{\mathbf{Z}(T) - \mathbf{F}'_T \boldsymbol{\beta}(T)|\mathcal{A}_T\} \\ &= \mathbf{Z}(T) - \mathbf{F}'_T \hat{\boldsymbol{\beta}}(T) \end{aligned}$$

where $\hat{\boldsymbol{\beta}}(T)$ denotes the best estimate of $\boldsymbol{\beta}(T)$ given the information in \mathcal{A}_T . Finally, we get as the desired interpolator for $\zeta(t, x)$

$$\begin{aligned} \hat{\zeta}_T(t, x) &= f(x)' \hat{\boldsymbol{\beta}}(t|T) + \hat{\eta}_T(t, x) \\ &= f(x)' \hat{\boldsymbol{\beta}}(t|T) + g'_T(t, x) \mathbf{G}_T^{-1} \hat{\mathbf{Y}}(T) \\ &= f(x)' \hat{\boldsymbol{\beta}}(t|T) + g'_T(t, x) \mathbf{G}_T^{-1} (\mathbf{Z}(T) - \mathbf{F}'_T \hat{\boldsymbol{\beta}}(T)) \quad (10) \end{aligned}$$

(10) generalizes the interpolation procedure known in geostatistics as "Kriging" (compare, e.g., Ripley, 1982). If our model (1) is reduced to

$$\zeta(x) = f(x)' \boldsymbol{\beta} + \eta(x)$$

without any time dependence and with deterministic regression parameter $\boldsymbol{\beta}$ then (10) is just the well-known "Universal Kriging". If $f(x)' \boldsymbol{\beta} \equiv \mu$ even

reduces to a constant, then (10) is the original "Kriging".

For the original practical problem we are mostly interested in estimating $\zeta(t, x)$ using all data up to time t , i.e. in calculating $\hat{\zeta}_t(t, x)$. For this purpose, the first summand $f(x)' \hat{\beta}(t|t)$ on the right-hand side of (10) can be calculated recursively in t using the Kalman filter of Theorem 2.1. In the second summand $\hat{\beta}(t)$ appears which consists of the subvectors $\hat{\beta}(0|t), \dots, \hat{\beta}(t|t)$. They can be calculated efficiently using Theorem 2.2. The resulting procedure works quite well, and it is fast enough to allow real-time calculations of $\hat{\zeta}(t, x)$, $t = 0, 1, 2, \dots$ for all knots x of an equispaced lattice in the plane such that the spatial-temporal evolution of the process can be studied as an animated graphic in detail.

For sake of illustration we consider one particular example with simulated data. Figure 1a and 1b show $c_0(y - x) = \mathcal{E} \eta(t, x) \eta(t, y)$ and $c_1(y - x) = \mathcal{E} \eta(t, x) \eta(t + 1, y)$. This corresponds to a nonisotropic dependency where the peak of $c_1(y - x)$ is shifted in a direction corresponding to the prevailing wind direction. Observations were taken at 16 irregularly spaced locations, and as a function representing general topographic trends, only linear functions were allowed, i.e. $f(x) = (1, x_1, x_2)'$. Figure 2a and 2b show two snapshots of the spatial-temporal interpolations $\hat{\zeta}(t, x)$ for $t = 4$ and $t = 18$.

4 Some remarks on estimation of model parameters

Up to now we have pretended to know the model parameters like the ARX-coefficient matrices A_1, \dots, A_p , C_0, \dots, C_r of (2), the AR(1)-matrix L of (3) or the noise covariance matrices Σ_δ and Σ_ε . As in the known procedures for Kriging of purely spatial data, these parameters have to be estimated (compare, e.g., Ripley, 1982). We do not want to go into details but point out some special features which follow from our modelling approach.

Hannan and his coworkers have given over the years an extensive theoretical treatment of maximum likelihood estimates for the parameters of multivariate Gaussian ARMA- and ARMAX-systems. A compilation of the relevant results is provided by Hannan and Deistler (1988). In our case, we cannot use this estimation theory directly as we do not observe the ARX-process $\beta(t)$. However, let F^- denote a pseudo inverse of F' , e.g. $F^- = (FF')^{-1}F$ for $m \leq N$ and FF' invertible, and let U^{-1} be the one time unit backshift operator: $U^{-1}\beta(t) = \beta(t-1)$. Then, we have from (2) and (4):

$$\begin{aligned} A(U^{-1}) F^- Z(t) &= A(U^{-1}) F^- (F' \beta(t) + Y(t)) \\ &= A(U^{-1})(\beta(t) + F^- Y(t)) \\ &= A(U^{-1}) F^- Y(t) + \delta(t) + C(U^{-1})\xi(t) \end{aligned}$$

If L^s decreases fast enough for $s \rightarrow \infty$ such that $Y(t) \approx \sum_{s=0}^q L^s \varepsilon(t-s)$, then we have for the observable time series $Z^*(t) = F^- Z(t)$

$$A(U^{-1}) Z^*(t) \approx A(U^{-1}) F^- \sum_{s=0}^q L^s \varepsilon(t-s) + \delta(t) + C(U^{-1}) \xi(t)$$

As the first two summands of the right-hand side form a $(p+q)$ -dependent time series we have for suitable white noise $\psi(t)$ and matrix coefficients B_0, \dots, B_{p+q}

$$A(U^{-1}) Z^*(t) \approx B(U^{-1}) \psi(t) + C(U^{-1}) \xi(t), \quad B(z) = \sum_{j=0}^{p+q} B_j z^j,$$

i.e. $Z^*(t)$ is approximately an ARMAX($p, p+q, r$)-process with the same autoregressive and exogenous part $A(U^{-1})$, $C(U^{-1})$ as the ARX-process $\beta(t)$. Therefore, fitting an ARMA($p, p+q, r$)-model to $Z^*(0), \dots, Z^*(T)$ as described in Hannan and Deistler (1988) provides estimates of A_1, \dots, A_p , C_1, \dots, C_r .

In a similar manner, the parameters of the AR(1)-model (3) may be estimated. If we could observe the time series $Y(t)$ directly, we could estimate the autocovariances

$$\Sigma_0 = \mathcal{E} Y(t) Y(t)' \quad \text{and} \quad \Sigma_1 = \mathcal{E} Y(t) Y(t+1)'$$

directly by the corresponding sample autocovariances, and then, using

$$L = \Sigma_1 \Sigma_0^{-1} \quad \text{and} \quad \Sigma_\varepsilon = \Sigma_0 - \Sigma_1 \Sigma_0^{-1} \Sigma_1$$

we would have estimates of L and Σ_ε , too. Now, let

$$\begin{aligned} \tilde{Z}(t) &= (I_N - F' F^-) Z(t) = (I_N - F' F^-)(F' \beta(t) + Y(t)) \\ &= (I_N - F' F^-) Y(t) \end{aligned}$$

as $F' F^- F' = F'$. Therefore, we may consider the sample autocovariances of the observable time series $\tilde{Z}(t)$ and calculate from them estimates of Σ_0 , Σ_1 . Here, the properties of $I_N - F' F^-$ have to be exploited, in particular the fact that it has eigenvalues 1 and 0 only and that its rank is $N - m$. The details of the somewhat involved procedure have been given by Grnder (1993).

Acknowledgement: This work has been funded by the Volkswagenstiftung as part of the Center for Practical Mathematics. We thank A. Roeder of the Forest Research Center of Rheinland-Pfalz for providing the problem, the data and continuous advice.

References

- [1] Franke, J., and B. Grnder (1992). *Stochastic modelling for analyzing immission data in forests*. In: Tagungsberichte der Arbeitsgruppe

Biometrie in der Ökologie der Deutschen Region der Internationalen Biometrischen Gesellschaft, Wuppertal.

- [2] Gründer, B. (1992). *Gasdistribution as a space-time model with regard to meteorological data*. In: Mathematical Modelling of Forest Ecosystems, ed. J. Franke and A. Roeder, J.D. Sauerländer's Verlag, Frankfurt a.M. .
- [3] Gründer, B. (1993). *Stochastische Modelle fr die Luftschadstoffverteilung in Waldgebieten*. Ph.D. Thesis, University of Kaiserslautern.
- [4] Hannan, E.J., and M. Deistler (1988). *The Statistical Theory of Linear Systems*, Wiley, New York.
- [5] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Engr., Series D*, **82**, 35-45.
- [6] Priestley, M.B. (1981). *Spectral Analysis and Time Series*, vol. 2, Academic Press, London, New York.
- [7] Ripley, R.D. (1981). *Spatial Statistics*. Wiley, New York.
- [8] Vecchia, A.V. (1988). Estimation and model identification for continuous spatial processes. *J.R. Statist. Soc. B*, **50**, 297-312.

Corresponding author's address:

Prof. Dr. J. Franke
Department of Mathematics
University of Kaiserslautern
P.O. Box 3049
D-67653 Kaiserslautern
F.R. of Germany