Filtering the Noise from Time Series and Spatial Data

Olaf Berke

Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany

berke@omega.statistik.uni-dortmund.de

ABSTRACT

Noisy observations form the basis for almost every scientific research and especially in environmental monitoring. The Noise is often an effect of imprecise instruments which cause measurement errors. If the noise variance is known it is possible to filter out the contaminating noise from the observations and then to predict the latent signal process. Solutions for this problem exist for time series application and will be briefly reviewed. In the geostatistical literature, i.e. for the analysis of spatial data, similar methods have been foreshadowed in the literature and will be outlined in this work.

KEY WORDS: Geostatistics, Kalman Filter, Kriging, Prediction, Signal, Time Series Analysis.

This research was supported by the German Research Council (DFG) through the Collaborative Research Centre at the University of Dortmund (SFB 475): Reduction of complexity for multivariate data structures.

1. INTRODUCTION

It is widely accepted that observations from scientific experiments are contaminated by measurement errors. Furthermore, in engineering applications the interesting process may be corrupted by an interference with other processes. This motivated the notions of *signal* and *noise*. In this setting the observational process $Z(\cdot)$ is the sum of the interesting process or signal $S(\cdot)$ and the contaminating process or noise $\varepsilon(\cdot)$

$$Z(\cdot) = S(\cdot) + \varepsilon(\cdot).$$

Given the measurements it is generally of interest to predict the signal and not the observational process. Because the signal process is assumed to be disturbed by a zero mean white noise process it follows, that the expectations of the observational process and the signal process are equivalent. Thus predicting the signal and not the observables will often result in smaller mean squared prediction errors due to the extra variability of the noise.

The prediction of the signal presumes knowledge of the distribution of the noise process. With restriction to linear prediction only the first and second order moments are needed to calculate a linear prediction. When the noise is modelled by a zero-mean white-noise process, knowledge of the noise variance parameter σ_ε^2 is required only. Generally the instrumental variance, i.e. the noise variance, is known from previous calibration experiments.

To handle the noise-filtering problem in spatial and temporal situations the following stochastic processes for the observational process are considered. First let

$$\mathcal{Z}_{ts} = \{Z_t\}_{t \in \mathcal{T} \subset I\!N}$$

denote a discrete parameter time series and second

$$\mathcal{Z}_{sp} = \{Z(s) : s \in \mathcal{D} \subset I\!R^2\}$$

be a continuous parameter spatial process.

The noise-filtering problem has been solved for time series applications by the development of the Kalman filter (cf. Kalman, 1960). In geostatistics a modified version of the kriging predictor to filter the noise was foreshadowed by Cressie (1988) and later on by Christensen, Johnson and Pearson (1992).

The outline of the paper is as follows. Sections 2 reviews the noise-filtering solution for time series data, i.e. temporal processes. Section 3 is concerned with outline of the solution to the filter problem for geostatistical data, i.e. spatial processes. Section 4 concludes with a discussion of further aspects.

2. TEMPORAL FILTERING

The Kalman filter given by Kalman (1960) is a recursive least squares method (cf. Duncan and Horn, 1972) which can be given a Bayesian interpretation (cf. Meinhold and Singpurwalla, 1983). In the later, the predictor and the corresponding mean squared prediction errors approximate the first and second order moments of the predictive distribution. Generally the Kalman filter is used in time series applications to filter the noise form the observational process to enable the prediction of the signal or state of nature. To do so a state space formulation for the observational process is needed. Some authors use the term state space model others prefer the term dynamic linear model (cf. Harvey, 1989; West and Harrison, 1989).

2.1 Dynamic Linear Models

The *dynamic linear model* consists of two equations. The first equation is the observation equation (or measurement equation)

$$\mathbf{Z}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \boldsymbol{\Sigma}_t).$$

Then the evolution equation (also called the state-, system- or transition equation) models the state parameter linearly and autoregressive

$$\boldsymbol{\beta}_t = \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\xi}_t, \qquad \boldsymbol{\xi}_t \sim (\mathbf{0}, \mathbf{Q}_t).$$

In this model, $\boldsymbol{\beta}_t$ is an (unobservable) random parameter vector that describes the state of the dynamic linear model at time t, hence, it is called the state parameter, $\mathbf{Z}_t$ is the vector of observations related to the state parameter by the observation (or regression) matrix $\mathbf{X}_t$, $\mathbf{G}_t$ is the evolution matrix, and $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\xi}_t$ are the observational and evolutional noise vectors assumed to be independently distributed with zero means and covariance matrices $Cov(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_t$ and $Cov(\boldsymbol{\xi}_t) = \mathbf{Q}_t$, respectively.

To complete the model initial values must be specified defining the distribution of the state parameter at time $t = 0$. If these initial values are not given by prior knowledge, the approximate moments of a non-informative prior $\boldsymbol{\beta}_0 \sim (\mathbf{0}, \kappa\mathbf{I})$ with $\kappa \to \infty$ may be used. This is equivalent to view the initial state parameter as fixed but unknown, $\boldsymbol{\beta}_0 \sim (\mathbf{b}_{0|0}, \mathbf{0})$. However, the initial state parameter $\boldsymbol{\beta}_0$ is assumed to be uncorrelated with the noise vectors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\xi}_t$.

The covariance matrices are assumed to be known but have to be estimated. This is the problem of model specification. Further the observation and evolution matrices are assumed to be known. This states the problem of model selection in practical applications. The dynamic linear model is said to be time invariant if the covariance matrices as well as the observation and evolution matrices are constant in time, i.e. $\boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma}$, $\mathbf{Q}_t \equiv \mathbf{Q}$, $\mathbf{X}_t \equiv \mathbf{X}$ and $\mathbf{G}_t \equiv \mathbf{G}$.

Stationary time series are known to have an autoregressive moving average (ARMA) representation that can be given a state space formulation as well. Further autoregressive integrated moving average (ARIMA) models, i.e. non-stationary time series models, can be represented by dynamic linear models. It is also possible to include exogenous variables into the dynamic linear model to represent the class of ARMAX models. Furthermore, dynamic linear models may represent single or multiple time series. In what follows, almost all widely used linear time series models may be represented by time invariant dynamic linear models.

In dynamic linear models the observational errors $\boldsymbol{\varepsilon}_t$ form the noise and the linear combination of the state parameters $\mathbf{S}_t = \mathbf{X}_t\boldsymbol{\beta}_t$ represents the signal while evolutional errors $\boldsymbol{\xi}_t$ form model inadequacies at time t.

2.2 The Kalman Filter Recursions

The Kalman filter is useful since it offers a unique method to filter the noise from the signal for many classes of time series models. Further it is a powerful method for on-line prediction since the filter algorithm works recursively using the last prediction of the state parameter and the current observation only. The mean squared prediction errors (MSPE) are by-products when the Kalman filter is run. The predictions from Kalman filtering are optimal in the sense that the MSPE is minimised within the class of linear predictors. To be more precise, the optimality depends on the initialisation (cf. Tsimikas and Ledolter, 1994). Using a point prior with unknown state parameter or equivalently a non-informative prior results in a mixed model and hence the predictors are best linear unbiased predictors (BLUP). Otherwise, the dynamic linear model is just a random effect model and the predictors from Kalman filtering are best linear predictors (BLP).

The Kalman filter recursions proceed in two steps. These are the prediction and updating steps to be performed before and after the new observation becomes available.

Before the Kalman filter recursions can be presented the following notation is introduced. The time indices $t|t-1$ and $t|t$ denote the predictors for time t based on observations up to and including time $t-1$ and t, respectively. Let $\mathbf{b}_{t|t-1}$ and $\mathbf{b}_{t|t}$ denote the predictors of the state parameter. Further $\mathbf{R}_{t|t-1}$ and $\mathbf{R}_{t|t}$ are the corresponding covariance matrices of the predictors.

With this set-up the prediction step for the state parameter is given by

$$\begin{aligned} \mathbf{b}_{t|t-1} &= \mathbf{G}_t \mathbf{b}_{t-1|t-1} \\ \mathbf{R}_{t|t-1} &= \mathbf{G}_t \mathbf{R}_{t-1|t-1} \mathbf{G}_t' + \mathbf{Q}_t \end{aligned}$$

and the updating step is

$$\begin{aligned} \mathbf{b}_{t|t} &= \mathbf{b}_{t|t-1} + \mathbf{K}_t(\mathbf{Z}_t - \mathbf{X}_t \mathbf{b}_{t|t-1}) \\ \mathbf{R}_{t|t} &= (\mathbf{I} - \mathbf{K}_t \mathbf{X}_t) \mathbf{R}_{t|t-1}, \end{aligned}$$

where

$$\mathbf{K}_t = \mathbf{R}_{t|t-1}\mathbf{X}_t'(\mathbf{X}_t\mathbf{R}_{t|t-1}\mathbf{X}_t' + \boldsymbol{\Sigma}_t)^{-1}$$

is the so called Kalman gain matrix, which is actually a vector in single time series applications.

The predictor of the signal $\mathbf{S}_t = \mathbf{X}_t\boldsymbol{\beta}_t$ follows straightforward from the predictor of the state parameter $\boldsymbol{\beta}_t$. Accordingly to the prediction and updating step one may distinguish between the prior and posterior predictor of the signal. Sometimes these predictors are called the forecast and filtering predictors, respectively. So let

$$\widehat{\mathbf{S}}_{t|t-1} = p(\mathbf{S}_t|\mathbf{Z}^{t-1}) = \mathbf{X}_t\mathbf{b}_{t|t-1}$$

and

$$\widehat{\mathbf{S}}_{t|t} = p(\mathbf{S}_t|\mathbf{Z}^{t}) = \mathbf{X}_t\mathbf{b}_{t|t}$$

denote the linear Bayesian prior and posterior predictors of the signal. Here $\mathbf{Z}^t = (\mathbf{Z}_t, \mathbf{Z}^{t-1})$ with $\mathbf{Z}_0 = (\mathbf{b}_{0|0}, \mathbf{R}_{0|0})$ is used to denote the sample and prior information up to time t. Theirs MSPE's are as follows

$$MSPE(\widehat{\mathbf{S}}_{t|t-1}) = E(\mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{X}_t\mathbf{b}_{t|t-1})^2 = Cov(\mathbf{X}_t\mathbf{b}_{t|t-1}) = \mathbf{X}_t\mathbf{R}_{t|t-1}\mathbf{X}_t'$$

and

$$MSPE(\widehat{\mathbf{S}}_{t|t}) = \mathbf{X}_t\mathbf{R}_{t|t}\mathbf{X}_t'.$$

It is worth to face the predictors of the signal with the corresponding predictors of the observational process now. The forecast predictor for the observational process $\mathbf{Z}_t = \mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$, i.e. the signal plus the noise, is the same like the corresponding signal forecast predictor

$$\widehat{\mathbf{Z}}_{t|t-1} = p(\mathbf{Z}_t|\mathbf{Z}^{t-1}) = \mathbf{X}_t\mathbf{b}_{t|t-1}.$$

However the MSPE is surely larger due to the variability of the noise

$$\begin{aligned} MSPE(\widehat{\mathbf{Z}}_{t|t-1}) &= E(\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t - \mathbf{X}_t\mathbf{b}_{t|t-1})^2 \\ &= \mathbf{X}_t Cov(\boldsymbol{\beta}_t - \mathbf{b}_{t|t-1})\mathbf{X}_t' + Cov(\boldsymbol{\varepsilon}_t) \\ &= \mathbf{X}_t\mathbf{R}_{t|t-1}\mathbf{X}_t' + \boldsymbol{\Sigma}_t. \end{aligned}$$

Therefore the gain in predicting the signal instead of the observational process in forecast situations measured in terms of the MSPE is quantified by the noise covariance matrix $\boldsymbol{\Sigma}_t$ that is assumed to be known and being positive definite.

This result does not hold for the filtering predictor of the observational process since in this case the optimal predictor is based on and given by the sample variables itself, viz.

$$\widehat{\mathbf{Z}}_{t|t} = p(\mathbf{Z}_t|\mathbf{Z}^t) = \mathbf{Z}_t.$$

With this the MSPE becomes zero

$$MSPE(\widehat{\mathbf{Z}}_{t|t}) = E(\mathbf{Z}_t - \mathbf{Z}_t)^2 = \mathbf{0}.$$

Thus the MSPE of the unobservable signal process is larger than the MSPE of the predictor for the observational process in the situation of temporal filtering

$$MSPE(\widehat{\mathbf{Z}}_{t|t}) \;=\; \mathbf{0} \;<\; \mathbf{X}_t\mathbf{R}_{t|t}\mathbf{X}_t' \;=\; MSPE(\widehat{\mathbf{S}}_{t|t}).$$

These are the one step ahead predictions from the Kalman filter. The h-step ahead predictions will be calculated in a similar way (cf. Harvey, 1989). Note that here forecasting means the prediction of the process at time t from time $t-1$ or later. This is in contrast with the general time series literature where forecasting means prediction from time t to time $t+1$.

3. SPATIAL FILTERING

The universal kriging predictor proposed by Matheron (1969; see Cressie, 1993, p. 151) is the best linear unbiased predictor (BLUP) in geostatistical applications. The predictor is often defined in terms of the variogram, since the spatial process needs then just to be intrinsic stationary rather then weak stationary. However, in practise the process is assumed to be ergodic, which is a stronger assumption than the different types of stationarity. With this it is equivalent to consider the universal kriging method using the variogram or the covariogram.

3.1 Spatial Linear Models

Consider the random sample $\mathbf{Z} = (Z(s_1), ..., Z(s_n))'$ to be taken at the locations $s_1, .., s_n$. The *spatial linear model* for these sample variables is of the following form

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}, \qquad \boldsymbol{\delta} \sim (\mathbf{0}, \boldsymbol{\Sigma}).$$

In this model, $\boldsymbol{\beta}$ is an unobservable but fixed parameter that determines the mean function of the spatial process, $\mathbf{X}$ is the (deterministic) spatial regression matrix generally depending on functions of the sample site co-ordinates, i.e. $\mathbf{x}(s)$, and $\boldsymbol{\delta}$ is an unobservable random vector with zero-mean and covariance matrix $\boldsymbol{\Sigma}$. The matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ depends on the relative position of the sample sites and is structured according to a small set of spatial structure parameters $\boldsymbol{\theta}$. However, dependence on $\boldsymbol{\theta}$ will be suppressed often.

The spatial linear model is of the same form like the Aitken or general linear model. From knowledge about $\boldsymbol{\theta}$ and spatial regression functions $\mathbf{x}(\cdot)$ the model for any other spatial location s_0 in the sampling domain $\mathcal{D}$ is given by

$$Z(s_0) = \mathbf{x}'\boldsymbol{\beta} + \delta(s_0), \qquad \delta(s_0) \sim (0, \sigma_\delta^2),$$

with $\mathbf{x} = \mathbf{x}(s_0)$ and $\sigma_\delta^2 = \sigma_\delta^2(\boldsymbol{\theta})$.

To account for observational noise the spatial linear model may be extended to the *spatial linear noise model* by splitting the $\boldsymbol{\delta}$ component of the spatial linear model into noise $\boldsymbol{\varepsilon}$ and the zero mean spatial random component $\boldsymbol{\eta}$. From this follows

$$\mathbf{Z} = \mathbf{S} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon},$$

where $\mathbf{Z} = (Z(s_1), ..., Z(s_n))'$ denotes the spatial sample vector to be observed at locations $s_1, ..., s_n \in \mathcal{D}$ with

$$\begin{aligned} \boldsymbol{\varepsilon} &\sim (\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}) \\ \boldsymbol{\eta} &\sim (\mathbf{0}, \mathbf{V}) \\ \mathbf{Z} &\sim (\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} = \mathbf{V} + \sigma_\varepsilon^2\mathbf{I}). \end{aligned}$$

This means that the spatial sample vector $\mathbf{Z}$ is the sum of unobservable and uncorrelated signal and noise components. The signal component of the sample vector is given by $\mathbf{S} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$ and the noise component is denoted by $\boldsymbol{\varepsilon}$. Since $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ it is clear that $\mathbf{V}$ also depends on $\boldsymbol{\theta}$ and particularly σ_ε^2 is part of the spatial structure parameter $\boldsymbol{\theta}$.

Further, the model is valid for any other location in the sampling domain, i.e. for any $s_0 \in \mathcal{D}$. Therefore the following model will be used to describe the spatial process at site s_0

$$Z(s_0) = S(s_0) + \varepsilon(s_0) = \mathbf{x}'\boldsymbol{\beta} + \eta(s_0) + \varepsilon(s_0),$$

with

$$\begin{aligned} \eta(s_0) &\sim (0, \sigma_\eta^2) \\ \varepsilon(s_0) &\sim (0, \sigma_\varepsilon^2) \end{aligned}$$

as well as $\mathbf{x} = \mathbf{x}(s_0)$ introduced above and $\sigma_\eta^2 = \sigma_\eta^2(\boldsymbol{\theta})$. The noise variance parameter is $\sigma_\varepsilon^2 = \sigma_\varepsilon^2(\boldsymbol{\theta})$. Note that the sum $\sigma_\eta^2 + \sigma_\varepsilon^2$ gives σ_δ^2 which is called the nugget effect in geostatistics. Besides σ_η^2 and σ_ε^2 the spatial structure parameter $\boldsymbol{\theta}$ generally contains the so called range and sill parameters.

3.2 Universal Kriging and the Spatial Filtering Equations

Within the context of spatial linear models using the notation $\boldsymbol{\sigma}' = Cov(Z(s_0), \mathbf{Z})$, the universal kriging predictor for $Z(s_0)$, i.e. the spatial BLUP is (cf. Goldberger, 1962)

$$\widehat{Z}_{UK}(s_0) = p(Z(s_0)|\mathbf{Z}) = \mathbf{x}'\widehat{\boldsymbol{\beta}}_{GLSE} + \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}),$$

where $\widehat{\boldsymbol{\beta}}_{GLSE}$ denotes the general least squares estimate (GLSE) of $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}}_{GLSE} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}.$$

The MSPE of the universal kriging predictor also called universal kriging variance is

$$MSPE(\widehat{Z}_{UK}(s_0)) = \sigma_\delta^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} + (\mathbf{x}' - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\mathbf{X})(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{x}' - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\mathbf{X})'.$$

The universal kriging predictor is known and sometimes criticised to be a direct interpolator. This means that the predictor for the sample variables is given by themselves, i.e.

$$\widehat{\mathbf{Z}}_{UK} = p(\mathbf{Z}|\mathbf{Z}) = \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}) = \mathbf{Z}.$$

So the MSPE becomes in this situation

$$MSPE(\widehat{\mathbf{Z}}_{UK}) = E(\mathbf{Z} - \mathbf{Z})^2 = \mathbf{0}.$$

However this predictor can be modified to solve the noise-filtering problem in geostatistical applications within the framework of the spatial linear noise model (cf. Cressie, 1988; Christensen, Johnson and Pearson, 1992).

To filter the noise in spatial or geostatistical situations using the method of best linear unbiased prediction first note that the BLUP for a random quantity, say $Z(s_0)$, given the sample $\mathbf{Z}$ is generally of the form

$$p(Z(s_0)|\mathbf{Z}) = E(\widehat{Z(s_0)}) + Cov(Z(s_0), \mathbf{Z})[Cov(\mathbf{Z})]^{-1}(\mathbf{Y} - \widehat{E(\mathbf{Z})}).$$

Here $E(\widehat{Z(s_0)})$ and $\widehat{E(\mathbf{Z})}$ denote the optimal linear estimates for the corresponding mean parameters that are given by theirs generalised least squares estimates.

From the spatial linear noise model introduced above with known spatial structure parameter $\boldsymbol{\theta}$, i.e. second order moments, it follows explicitly

$$\begin{aligned}
E(\mathbf{S}) &= E(\mathbf{Z}) = \mathbf{X}\boldsymbol{\beta} \\
Cov(\mathbf{Z}) &= Cov(\mathbf{S}) + Cov(\boldsymbol{\varepsilon}) = \mathbf{V} + \sigma_\varepsilon^2\mathbf{I} = \boldsymbol{\Sigma} \\
Cov(\mathbf{S}, \mathbf{Z}) &= Cov(\mathbf{S}, \mathbf{S} + \boldsymbol{\varepsilon}) = Cov(\mathbf{S}) = \mathbf{V} \\
&= \boldsymbol{\Sigma} - \sigma_\varepsilon^2\mathbf{I} = Cov(\mathbf{Z}) - Cov(\boldsymbol{\varepsilon}).
\end{aligned}$$

The universal kriging predictor for the signal at the sampling locations $s_1, ..., s_n$ is given by

$$\begin{aligned}\widehat{\mathbf{S}}_{UK} = p(\mathbf{S}|\mathbf{Z}) &= \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE} + \mathbf{V}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}) \\ &= \mathbf{Z} - \sigma_\varepsilon^2 \mathbf{I}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}).\end{aligned}$$

To find the MSPE connected with the predictor of the signal at the sampling locations first note that the predictor $\widehat{\mathbf{S}}_{UK}$ can be written as

$$\widehat{\mathbf{S}}_{UK} = \{\mathbf{V} + (\mathbf{X} - \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{X})(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\}\boldsymbol{\Sigma}^{-1}\mathbf{Z} = \boldsymbol{\lambda}'\mathbf{Z},$$

which shows that the predictor is unbiased and linear in the sample variables. With this the MSPE follows to be of the form

$$\begin{aligned}MSPE(\widehat{\mathbf{S}}_{UK}) &= Cov(\mathbf{S} - \boldsymbol{\lambda}'\mathbf{Z}) = Cov((\mathbf{I} - \boldsymbol{\lambda}')\mathbf{S}(\mathbf{I} - \boldsymbol{\lambda}')' - \boldsymbol{\lambda}'\boldsymbol{\varepsilon}) \\ &= (\mathbf{I} - \boldsymbol{\lambda}')\mathbf{V}(\mathbf{I} - \boldsymbol{\lambda}')' + \sigma_\varepsilon^2 \boldsymbol{\lambda}'\boldsymbol{\lambda} \\ &= \mathbf{V} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\mathbf{V}.\end{aligned}$$

Since the spatial linear noise model is assumed to hold for any location in the sampling domain, the BLUP for the signal at any location $s_0 \in \mathcal{D}$ is given by

$$\widehat{S}_{UK}(s_0) = p(S(s_0)|\mathbf{Z}) = \mathbf{x}'\widehat{\boldsymbol{\beta}}_{GLSE} + \mathbf{v}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}),$$

where $\mathbf{v}' = \mathbf{v}(\boldsymbol{\theta})' = Cov(S(s_0), \mathbf{Z})$ represents the vector of covariance's between the signal at location s_0 and the sampling variables depending on $\boldsymbol{\theta}$.

Further investigations lead to the result that the BLUP for the signal coincides with the BLUP of the observational process at unsampled locations, i.e.

$$\widehat{S}_{UK}(s_0) = \widehat{Z}_{UK}(s_0), \qquad s_0 \in \mathcal{D} \setminus \{s_1, ...s_n\}.$$

This is similar to the result $\widehat{\mathbf{S}}_{t|t-1} = \widehat{\mathbf{Z}}_{|t-1}$ in Kalman filtering and is based on the fact that the noise is modelled through white noise, i.e. a family of uncorrelated random variables. Thus, for $s_0 \in \mathcal{D} \setminus \{s_1, ...s_n\}$ follows

$$\boldsymbol{\sigma}' = Cov(Z(s_0), \mathbf{Z}) = Cov(S(s_0) + \varepsilon(s_0), \mathbf{Z}) = Cov(S(s_0), \mathbf{Z}) = \mathbf{v}'$$

and hence the equivalence of the predictors.

Similar to the derivation of $MSPE(\widehat{\mathbf{S}}_{UK})$ the MSPE of $\widehat{S}_{UK}(s_0)$ will be shown to be of the form

$$\begin{aligned} MSPE(\widehat{\mathbf{S}}_{UK}(s_0)) &= Cov(S(s_0) - \tilde{\boldsymbol{\lambda}}'\mathbf{Z}) \\ &= Cov(S(s_0)) + Cov(\tilde{\boldsymbol{\lambda}}'\mathbf{Z}) - 2\tilde{\boldsymbol{\lambda}}'Cov(S(s_0), \mathbf{Z}) \\ &= \sigma_\eta^2 + \tilde{\boldsymbol{\lambda}}'\boldsymbol{\Sigma}\tilde{\boldsymbol{\lambda}} - 2\tilde{\boldsymbol{\lambda}}'\mathbf{v}, \end{aligned}$$

with

$$\tilde{\boldsymbol{\lambda}}' = \{\mathbf{v}' + (\mathbf{x}' - \mathbf{v}'\boldsymbol{\Sigma}^{-1}\mathbf{X})(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\}\boldsymbol{\Sigma}^{-1}.$$

The following Corollary will be stated to summarise the results developed in the preceding text.

Corollary:

In the spatial linear model the best linear unbiased predictor for the signal component for all locations $s_0 \in \mathcal{D}$, is given by

$$\widehat{S}_{UK}(s_0) = \mathbf{x}'\widehat{\boldsymbol{\beta}}_{GLSE} + \mathbf{v}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE})$$

with mean squared prediction error

$$MSPE(\widehat{\mathbf{S}}_{UK}(s_0)) = \sigma_\eta^2 + \tilde{\boldsymbol{\lambda}}'\boldsymbol{\Sigma}\tilde{\boldsymbol{\lambda}} - 2\tilde{\boldsymbol{\lambda}}'\mathbf{v}.$$

Proof: To proof that $\widehat{S}_{UK}(s_0)$ is the BLUP for the signal at any location $s_0 \in \mathcal{D}$ needs just to notice that $\mathbf{v}' = Cov(S(s_0), \mathbf{Z})$ and $\boldsymbol{\Sigma} = Cov(\mathbf{Z})$. Hence, the predictor has the form

$$\widehat{S}_{UK}(s_0) = \mathbf{x}'\widehat{\boldsymbol{\beta}}_{GLSE} + Cov(S(s_0), \mathbf{Z})(Cov(\mathbf{Z}))^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE})$$

which gives the BLUP according to the theorem about best linear unbiased prediction (cf. Goldberger, 1962; Christensen, 1996, p. 266). The form of the MSPE was derived above.

$\square$

For completeness the MSPE's of the predictors for the observational process and the signal process will be compared now.

The results are similar to that in Kalman filtering. First, for the observational process at the sampling locations $s_1, ..., s_n$ the MSPE is zero, i.e. $MSEP(\widehat{\mathbf{Z}}_{UK}) = \mathbf{0}$. So the MSPE is smaller than for the predictor of the signal at the same locations

$$\mathbf{0} = MSEP(\widehat{\mathbf{Z}}_{UK}) < MSPE(\widehat{\mathbf{S}}_{UK}) = \mathbf{V} + \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\mathbf{V}.$$

This follows from

$$\widehat{\mathbf{S}}_{UK} = \widehat{\mathbf{Z}}_{UK} - \sigma_\varepsilon^2 \mathbf{I}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}).$$

So the MSPE of the predictor for the signal becomes

$$\begin{aligned} MSPE(\widehat{\mathbf{S}}_{UK}) &= E(\widehat{\mathbf{S}}_{UK} - \mathbf{S})^2 \\ &= E(\boldsymbol{\varepsilon} - \sigma_\varepsilon^2 \mathbf{I}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{GLSE}))^2. \end{aligned}$$

And this is positive definite, since this is the covariance matrix of the prediction error which is almost surely unequal to zero.

Lastly the $MSPE(\widehat{Z}_{UK}(s_0))$ is compared to $MSPE(\widehat{S}_{UK}(s_0))$ for unsampled locations $s_0 \in \mathcal{D} \setminus \{s_1, ..., s_n\}$. As shown above the predictors of interest are equivalent, i.e. $\widehat{Z}_{UK}(s_0) = \widehat{S}_{UK}(s_0)$. So this gives

$$\begin{aligned} MSPE(\widehat{Z}_{UK}(s_0)) &= E(\widehat{Z}_{UK}(s_0) - Z(s_0)\,)^2 \\ &= E(\,(\widehat{S}_{UK}(s_0) - S(s_0)) + \varepsilon(s_0)\,)^2 \\ &= MSPE(\widehat{S}_{UK}(s_0)) + \sigma_\varepsilon^2, \end{aligned}$$

or vice versa

$$MSPE(\widehat{S}_{UK}(s_0)) < MSPE(\widehat{S}_{UK}(s_0)) + \sigma_\varepsilon^2 = MSPE(\widehat{Z}_{UK}(s_0)).$$

4. DISCUSSION

The results for noise filtering in time series and geostatistical data are very similar and may be summarised as follows. There are two cases to be distinguished: predicting at time or site where an observation is sampled (case 1) or not (case 2). In case 1 the predictors for the signal and the observational processes are different and the MSPE of the signal predictor is larger than the MSPE of the predictor for the observational variable. In case 2 the predictors for the signal and the observation coincide but the MSPE of the signal predictor is smaller and the gain in predicting the signal is given by the noise variance.

To compare the predictors for different processes, i.e. the signal and the observational process, by use of the MSPE's makes only sense for unbiased predictors. Then the predictors are centred around the same value. The mean of the observable is given by the mean of the signal since the noise is modelled by zero mean white noise. So in practise the Kalman filter will be started with a non-informative prior resulting in BLUP's like universal kriging gives BLUP's, i.e. unbiased predictors.

Lastly note that the subject of spatial statistics is divided into three parts: point processes, lattice or regional data, and geostatistics. The outline of spatial filtering applies to the geostatistical frame work, however, extension to the regional data set-up is also possible.

REFERENCES

Christensen, R. (1996) *Plane Answers to Complex Questions: The Theory of Linear Models.* Springer, New York.

Christensen, R., Johnson, W. and Pearson, L.M. (1992) Prediction diagnostics for spatial linear models. *Biometrika*, 79, 583-591.

Cressie, N. (1988) Spatial prediction and ordinary kriging. *Mathematical Geology*, 20, 405-421.

Cressie, N. (1993) *Statistics for Spatial Data*, revised edition. Wiley, New York.

Duncan, D.B. and Horn, S.D. (1972) Linear dynamic recursive estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association*, 67, 815-821.

Goldberger, A.S. (1962) Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57, 369-375.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 83, 35-45.

Matheron, G. (1969) Le Krigeage Universel. *Cahiers du Centre de Morphologie Mathematique*, No. 1, Fontainebleau, France.

Meinhold, R.J. and Singpurwalla, N.D. (1983) Understanding the Kalman filter. *The American Statistician*, 37, 123-127.

Tsimikas, J. and Ledolter, J. (1994) REML and best linear unbiased prediction in state space models. *Communications in Statistics – Theory and Methods*, 23, 2253-2268.

West, M. and Harrison, J. (1989) *Bayesian Forecasting and Dynamic Models*. Springer, New York.