

Space–time Kalman filter

Noel Cressie & Christopher K. Wikle

Volume 4, pp 2045–2049

in

Encyclopedia of Environmetrics

(ISBN 0471 899976)

Edited by

Abdel H. El-Shaarawi and Walter W. Piegorsch

© John Wiley & Sons, Ltd, Chichester, 2002

Space–time Kalman filter

The Kalman filter arose out of R.E. Kalman’s interest in applying the concept of state vectors to the Wiener filtering problem [14]. The success of this method was evident in early applications to trajectory estimation and control of spacecraft; it was so successful, in fact, that the Kalman filter quickly became an essential component of modern **control theory** and practice (e.g. [13]). This initial success led to the propagation of Kalman-filtering ideas to other scientific disciplines, within which the methodology was adapted to suit numerous state–space oriented problems (*see State–space methods*). For example, in the atmospheric sciences, the approach was discussed and tested in the context of numerical **weather prediction** data assimilation in the early 1980s, where the spatial dimension is handled by considering the state process as a vector of pixel values (e.g. [7]). Kalman filtering (of spatial data) is still an active area of research in the atmospheric and oceanic sciences [4, 18, 20]. More generally, the idea of dynamic space–time statistical models (*see Dynamic model*) (albeit from a **time series** perspective) can be found in earlier work in geography by Bennett [1].

Initial efforts at space–time modeling in statistics focused on what we would now call *linear mixed models*. These models typically included deterministic spatio-temporal trend plus various spatial, temporal, and spatio-temporal **random effects**. Such approaches necessarily require that one specifies explicitly the joint space–time covariance structure of the process of interest (*see Space–time covariance models*). In practice, when one considers high-dimensional processes with complicated dynamics, this approach is limiting. In the 1990s, several researchers independently began looking at the potential of having a dynamic temporal aspect in space–time statistical modeling. Commenting on Handcock and Wallis’ [11] Bayesian approach to geostatistical space–time modeling, Cressie [6] suggested that a Kalman filter incorporating space and time would be a powerful way to apply the Bayesian paradigm to modeling space–time phenomena. In an invited paper to the Seventeenth International Biometric Society Conference in 1994, Goodall and Mardia [8] sketched the idea for an approach to space–time Kalman filtering that resulted in the article by Mardia et al. [15]. Also in 1994, Guttrop

et al. [8] considered many of the notions of the dynamic space–time model, although they did not consider Kalman filtering. It appears that the first fully detailed, statistically motivated, continuous space, space–time Kalman filter (STKF) appeared in Huang and Cressie [12], there applied to the problem of predicting snow-water equivalent during the winter and spring months of the western US. The dynamic space–time structure was relatively simple, being separable in space and time. More complicated structures have recently been considered by Berke [3], Meiring et al. [17], and Wikle and Cressie [22].

A General Space–Time Dynamical Model

Before presenting the details of space–time Kalman filtering, it is useful to examine the general notions of dynamic space–time statistical modeling. We begin by considering a *data model* for our observations $z(\mathbf{s}; t)$, namely

$$z(\cdot, t) = f_d[y(\cdot, t); \theta_d(t)] \quad (1)$$

where f_d is a stochastic functional that depends on the true process $y(\cdot, t)$ and parameters $\theta_d(t)$. We then specify a *process model* to describe how the state process evolves with time; that is,

$$y(\cdot, t) = f_p[y(\cdot, t-1), \dots, y(\cdot, 1); \theta_p(t)] \quad (2)$$

where f_p is a stochastic functional of past values of the true process and parameters $\theta_p(t)$. Equation (2) models the space–time dynamics of the phenomenon, such as might be obtained by a discretization of stochastic partial differential equations derived from physical laws of temperature, pressure, and wind. Often, in the environmental sciences, we have incomplete knowledge of the physical mechanisms responsible for the dynamics, and we must resort to statistical estimation to describe the dynamics. We note that this modeling framework makes no distributional or linearity assumptions, and is flexible enough to consider both discrete and continuous spatial and temporal processes.

From a **hierarchical model** point of view, (1) gives the first-level probability distribution $[z|y, \theta_d]$, and (2) gives the second-level probability distribution $[y|\theta_p]$, where for notational convenience (only) we momentarily leave out dependence on t and the notation $[u|v]$ denotes the distribution of u given v .

2 Space–time Kalman filter

Then Bayes’ theorem gives the posterior probability distribution through

$$[y|z, \theta] \propto [z|y, \theta_d] \times [y|\theta_p] \quad (3)$$

where $\theta \equiv (\theta'_d, \theta'_p)'$ (see **Bayesian methods and modeling**). If we assume squared-error loss (as is usually the case), the optimal predictor of $y(\mathbf{s}_0, t_0)$ is

$$\hat{y}(\mathbf{s}_0, t_0) = E(y(\mathbf{s}_0, t_0)|z, \theta) \quad (4)$$

which is the mean of the **posterior distribution** (3). The posterior variance obtained from (3) is a measure of how precise the predictor $\hat{y}(\mathbf{s}_0, t_0)$ is.

To complete our discussion of general space–time statistical modeling, we discuss how θ can be handled. First, we note that ‘plugging in’ an estimate $\hat{\theta}$ into the posterior distribution $[y|z, \theta]$ or the predictor $E(y(\mathbf{s}_0, t_0)|z, \theta)$ is, in effect, an **empirical Bayes method** of analysis. This ignores the extra variability inherent in estimating θ . Alternatively, a fully hierarchical Bayesian implementation puts a **prior distribution** $[\theta]$ on the parameters θ and bases all inference on the resulting posterior distribution $[y, \theta|z]$. Waller et al. [21], Wikle et al. [23], and Berliner et al. [2] show this to be a very powerful methodology for including complicated space–time structure (see **Hierarchical Bayesian space–time analysis**). The drawback is a high computational overhead (see **Bayesian computation**), which may prove to be insurmountable if on-line filtering is required.

The STKF and Dimension Reduction

When (1) and (2) satisfy linear and Gaussian assumptions, a STKF can be obtained. Furthermore, as discussed below, Wikle and Cressie [22] show how a state–space dimension reduction can be achieved by featuring dominant components of variation in (2).

Consider the following three model equations:

$$z(\mathbf{s}, t) = y(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \quad (5)$$

$$y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \nu(\mathbf{s}, t) \quad (6)$$

$$\mu(\mathbf{s}, t) = \int \omega_s(\mathbf{u})\mu(\mathbf{u}, t-1) d\mathbf{u} + \eta(\mathbf{s}, t) \quad (7)$$

where the error terms ε , ν , and η are assumed Gaussian. In the general framework outlined above, (5) corresponds to the measurement equation (1), and (6) and (7) represent the process model (2). In (5), the

error term ε is white noise and represents measurement error. In (6), the error term ν represents a spatial structure that is independent from one time point to the next. In the terminology of Cressie [5, Chapter 2], ν represents the small-scale spatial structure that, while exhibiting temporal variability, has no temporal dynamics associated with it. Note that Mardia et al. [15] assume that $\nu \equiv 0$, resulting in space–time predictions that may oversmooth and mean squared prediction errors that do not account for the component ν . We also note that the model (5), (6), and (7) is continuous in space and discrete in time. Such a formulation is ideal for dynamic environmental processes, which we are often interested in predicting, and whose principle dynamics are governed by relatively large time scales.

The space–time dynamics are found in the other component of (6), namely μ . In (7), $\mu(\mathbf{s}, t)$ is seen to depend dynamically on $\mu(\mathbf{u}, t-1)$, where \mathbf{u} belongs to some ‘neighborhood’ of \mathbf{s} determined by the support of the interaction function $\omega_s(\mathbf{u})$. Thus, if rainfall is the variable of interest and weather patterns generally develop from west to east, then the rainfall tomorrow at location \mathbf{s} will depend on today’s rainfall at locations \mathbf{u} within a swath west of \mathbf{s} , up to a distance determined by the speed of the weather system. Importantly, the error term η in (7) is temporally white but spatially ‘colored’, which provides the independent shocks needed to keep the autoregressive process of **random fields** $\{\mu(\cdot, t): t = 1, 2, \dots\}$ moving forward in time in a stochastic manner.

Identifiability of the decomposition (6) is maintained by writing

$$\mu(\mathbf{s}, t) = \sum_{j=1}^K a_j(t)\phi_j(\mathbf{s}) \quad (8)$$

where $\{\phi_j(\cdot): j = 1, 2, \dots\}$ are a complete and orthonormal sequence of deterministic spatial functions and $\{a_j(t): t = 1, 2, \dots\}$ is a random time series for each $j = 1, 2, \dots, K$. The upper index K is the key to dimension reduction in the STKF and, once chosen, determines the dependence structure in the error term ν in (6). Now, because of completeness, we can write

$$\omega_s(\mathbf{u}) = \sum_{l=1}^{\infty} b_l(\mathbf{s})\phi_l(\mathbf{u}) \quad (9)$$

Truncating the infinite series in (9) and using the orthonormality of the basis functions allows us to

write (7) as

$$\boldsymbol{\phi}(\mathbf{s})' \mathbf{a}(t) = \mathbf{b}(\mathbf{s})' \mathbf{a}(t-1) + \eta(\mathbf{s}, t) \quad (10)$$

where $\boldsymbol{\phi}(\mathbf{s}) \equiv [\phi_1(\mathbf{s}), \dots, \phi_K(\mathbf{s})]'$, $\mathbf{a}(t) \equiv [a_1(t), \dots, a_K(t)]'$, and $\mathbf{b}(\mathbf{s}) \equiv [b_1(\mathbf{s}), \dots, b_K(\mathbf{s})]'$. This equation holds in particular for locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, which allows the vector $\mathbf{a}(t)$ to be isolated on the left-hand side of a first-order autoregressive equation. Thus, (5), (6), and (7) become

$$z(\mathbf{s}, t) = y(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \quad (11)$$

$$y(\mathbf{s}, t) = \sum_{j=1}^K a_j(t) \phi_j(\mathbf{s}) + v(\mathbf{s}, t) \quad (12)$$

$$\mathbf{a}(t) = H\mathbf{a}(t-1) + J\boldsymbol{\eta}(t) \quad (13)$$

where $\boldsymbol{\eta}(t) \equiv [\eta(\mathbf{s}_1, t), \dots, \eta(\mathbf{s}_n, t)]'$, $H = JB$, $J = (\Phi' \Phi)^{-1} \Phi'$, and the $n \times K$ matrices Φ and B are defined as $\Phi \equiv [\boldsymbol{\phi}(\mathbf{s}_1), \dots, \boldsymbol{\phi}(\mathbf{s}_n)]'$ and $B \equiv [\mathbf{b}(\mathbf{s}_1), \dots, \mathbf{b}(\mathbf{s}_n)]'$.

The goal of the STKF is to predict the unseen process y based on space–time data $\{z(\mathbf{s}_i, t_i)\}$. Assuming squared-error loss, the optimal predictor of $y(\mathbf{s}_0, t_0)$ is $E\{y(\mathbf{s}_0, t_0) | \{z(\mathbf{s}_i, t_i)\}\}$. Under Gaussian assumptions, this conditional expectation is linear in the data. And, if the Gaussian assumption is not appropriate, this linear predictor is still optimal, but now in the class of all linear predictors.

Given (11) and (12) and the state-process equation (13), the optimal predictor of $\mathbf{a}(t)$ given observations up to and including t is expressed recursively in terms of a Kalman filter [14, 16]:

$$\begin{aligned} \hat{\mathbf{a}}(t|t) &\equiv E[\mathbf{a}(t) | \mathbf{z}(t), \dots, \mathbf{z}(1)] \\ &= \hat{\mathbf{a}}(t|t-1) + G(t)[\mathbf{z}(t) - \Phi \hat{\mathbf{a}}(t|t-1)] \end{aligned} \quad (14)$$

for $t \geq 1$, with mean-squared prediction error

$$\begin{aligned} P(t|t) &\equiv E\{[\mathbf{a}(t) - \hat{\mathbf{a}}(t|t)][\mathbf{a}(t) - \hat{\mathbf{a}}(t|t)]'\} \\ &= P(t|t-1) - G(t)\Phi P(t|t-1) \end{aligned} \quad (15)$$

where $\mathbf{z}(t) \equiv [z(\mathbf{s}_1, t), \dots, z(\mathbf{s}_n, t)]'$; the Kalman gain $G(t)$ is given by

$$G(t) = P(t|t-1)\Phi'[R + V + \Phi P(t|t-1)\Phi']^{-1} \quad (16)$$

and R and V are defined below.

The one-step-ahead prediction equations are given by

$$\begin{aligned} \hat{\mathbf{a}}(t|t-1) &\equiv E[\mathbf{a}(t) | \mathbf{z}(t-1), \dots, \mathbf{z}(1)] \\ &= H\hat{\mathbf{a}}(t-1|t-1), \end{aligned} \quad (17)$$

$$\begin{aligned} P(t|t-1) &\equiv \text{var}[\mathbf{a}(t) | \mathbf{z}(t-1), \dots, \mathbf{z}(1)] \\ &= HP(t-1|t-1)H' + JQJ' \end{aligned} \quad (18)$$

where $R \equiv \text{var}[\varepsilon(t)]$, $V \equiv \text{var}[v(t)]$, $Q \equiv \text{var}[\boldsymbol{\eta}(t)]$,

$$\boldsymbol{\varepsilon}(t) \equiv [\varepsilon(\mathbf{s}_1, t), \dots, \varepsilon(\mathbf{s}_n, t)]' \quad (19)$$

$$\mathbf{v}(t) \equiv [v(\mathbf{s}_1, t), \dots, v(\mathbf{s}_n, t)]'$$

and $\boldsymbol{\eta}(t)$ was similarly defined earlier. To start the Kalman recursion, we assume that $\hat{\mathbf{a}}(0|0) \equiv \mathbf{0}$ and $P(0|0) \equiv JC_0^\mu J'$, where in practice an estimate \hat{C}_0^μ is substituted for $C_0^\mu \equiv \text{var}[\mu(\mathbf{s}_1, t), \dots, \mu(\mathbf{s}_n, t)]$ [22].

Now, consider prediction of the process $y(\mathbf{s}, t)$ based on the Kalman filter predictor $\hat{\mathbf{a}}(t|t)$. The optimal predictor is then

$$\hat{y}(\mathbf{s}, t|t) = \boldsymbol{\phi}(\mathbf{s})' \hat{\mathbf{a}}(t|t) + \mathbf{c}_v(\mathbf{s})'(C_0^z)^{-1} \mathbf{z}(t) \quad (20)$$

where

$$\begin{aligned} C_0^z &\equiv \text{var}[\mathbf{z}(t)], \quad \mathbf{c}_v(\mathbf{s}) \equiv E[v(\mathbf{s}, t)v(t)] \\ &= [c_v(\mathbf{s}, \mathbf{s}_1), \dots, c_v(\mathbf{s}, \mathbf{s}_n)]' \end{aligned} \quad (21)$$

and $c_v(\mathbf{s}, \mathbf{r}) \equiv E[v(\mathbf{s}, t)v(\mathbf{r}, t)]$. We note that the second term in (20) is a type of simple **kriging** (e.g. [5, p. 110]) applied to the spatial error term $v(\mathbf{s}, t)$. Thus, as the truncation integer K decreases, the optimal predictor of $y(\mathbf{s}, t)$ begins to look more and more like the simple-kriging predictor in the presence of measurement error. This formula is contrasted with that of Mardia et al. [15], which does not have the second term. The recognition of the component of variation v in (6) and its presence in the optimal predictor (20) is important.

The conditional prediction error variance for $y(\mathbf{s}, t)$ is

$$\begin{aligned} \text{var}[y(\mathbf{s}, t) - \hat{y}(\mathbf{s}, t|t) | \mathbf{z}(t), \dots, \mathbf{z}(1)] \\ = \boldsymbol{\phi}(\mathbf{s})' P(t|t) \boldsymbol{\phi}(\mathbf{s}) + \mathbf{c}_v(\mathbf{s}, \mathbf{s}) - \mathbf{c}_v(\mathbf{s})'(C_0^z)^{-1} \mathbf{c}_v(\mathbf{s}) \\ - 2\boldsymbol{\phi}(\mathbf{s})' \text{cov}[\hat{\mathbf{a}}(t|t), \mathbf{z}(t)](C_0^z)^{-1} \mathbf{c}_v(\mathbf{s}) \end{aligned} \quad (22)$$

The first term in (22) corresponds to the prediction error variance from the μ process; the second and third terms together correspond to the simple-kriging prediction variance of the v process; the last

4 Space–time Kalman filter

term is a correction resulting from the covariance between the Kalman-filter prediction of the μ process [through the Kalman-filter prediction of $\mathbf{a}(\cdot)$] and the simple-kriging predictor. As the truncation integer K decreases, the prediction error variance tends to behave more and more like the simple-kriging variance.

The one-step-ahead optimal predictor and prediction error variance for $y(\mathbf{s}, t + 1)$, based on the model (11), (12), and (13), can be found in a similar fashion. In this case, the kriging terms do not contribute to the one-step-ahead prediction since the ν process is assumed not to evolve in time. We also note that one can implement the backward ‘smoothing’ recursions to get a space–time Kalman smoother if desired (e.g. [19]).

Unknown Parameters in the STKF

The predictor presented above is optimal assuming the covariances R , V , and Q are known, as well as is the state matrix H (i.e. Φ and B are known). In our case, we can choose any class of ϕ s as long as they are complete and orthonormal, but we must estimate the R , V , Q , and B matrices. By substituting these estimates in place of the unknown matrices, we no longer obtain exactly the conditional expectation. This approach is completely analogous to what is done when Kalman filtering in time or kriging in space, where the covariance or variogram parameters are usually unknown and have to be estimated. A general discussion of the problems in obtaining predictive distributions given estimated parameters, can be found in [24]. As mentioned previously, this approach corresponds to viewing the Kalman filter as an empirical Bayesian technique [16]. Alternatively, fully Bayesian hierarchical approaches to space–time dynamic modeling can be implemented [2, 23], but there is a tradeoff between computational efficiency, with the empirical Bayesian approach, and statistical precision, with the fully Bayesian approach. Wikle and Cressie [22] give standard method-of-moments estimators for R , V , Q , and B . Although **maximum likelihood estimators** of model parameters are more efficient, the high-dimensional nature of many spatio-temporal problems makes for poorly behaved likelihood surfaces and iterative algorithms that are difficult to implement, thus defeating one of our goals

of being able to process efficiently large spatio-temporal datasets. However, if the dimension reduction is considerable, maximum likelihood estimation of the STKF can be considered using modifications of Gauss–Newton methods (e.g. [9]) or the **EM algorithm** (e.g. [19]).

References

- [1] Bennett, R.J. (1979). *Spatial Time Series*, Pion, London.
- [2] Berliner, L.M., Wikle, C.K. & Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling, *Journal of Climate* **13**, 3953–3968.
- [3] Berke, O. (1998). On spatiotemporal prediction for on-line monitoring data, *Communications in Statistics – Theory and Methods* **27**, 2343–2369.
- [4] Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C. & Busalacchi, A.J. (1996). Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter, *Journal of Geophysical Research* **101**, 22 599–22 617.
- [5] Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Revised Edition, Wiley, New York.
- [6] Cressie, N. (1994). Comment on ‘An approach to statistical spatial–temporal modeling of meteorological fields’ by M.S. Handcock and J.R. Wallis, *Journal of the American Statistical Association* **89**, 379–382.
- [7] Ghil, M., Cohn, S.E., Tavantzis, J., Bube, K. & Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction, in *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil & E. Kallen, eds, Springer-Verlag, New York, pp. 139–224.
- [8] Goodall, C. & Mardia, K.V. (1994). Challenges in multivariate spatio-temporal modeling, *Proceedings of the Seventeenth International Biometric Conference*, Hamilton, Ontario, Canada, 8–12 August.
- [9] Gupta, N.K. & Mehra, R.K. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations, *IEEE Transactions on Automatic Control* **19**, 774–783.
- [10] Guttorp, P., Meiring, W. & Sampson, P.D. (1994). A space–time analysis of ground-level ozone data, *Environmetrics* **5**, 241–254.
- [11] Handcock, M.S. & Wallis, J.R. (1994). An approach to statistical spatial–temporal modeling of meteorological fields (with discussion), *Journal of the American Statistical Association* **89**, 368–390.
- [12] Huang, H.C. & Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter, *Computational Statistics and Data Analysis* **22**, 159–175.
- [13] Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, San Diego.
- [14] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering, Series D* **82**, 34–45.

- [15] Mardia, K.V., Goodall, C.R., Redfern, E. & Alonso, F.J. (1998). The kriged Kalman filter (with discussion), *Test* **7**, 217–285.
- [16] Meinhold, J. & Singpurwalla, N.D. (1983). Understanding the Kalman filter, *The American Statistician* **37**, 123–127.
- [17] Meiring, W., Guttorp, P. & Sampson, P.D. (1998). Space–time estimation of grid-cell hourly ozone levels for assessment of a deterministic model, *Environmental and Ecological Statistics* **5**, 197–222.
- [18] Miller, R.N., Busalacchi, A.J. & Hackert, E.C. (1997). Applications of data assimilation to analysis of the ocean on large scales, *Journal of the Meteorological Society of Japan* **75**, 445–462.
- [19] Shumway, R.H. & Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis* **3**, 253–264.
- [20] Thiebaux, J.J. (1997). The power of the duality in spatial–temporal estimation, *Journal of Climate* **10**, 567–573.
- [21] Waller, L.A., Carlin, B.P., Xia, H. & Gelfand, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association* **92**, 607–617.
- [22] Wikle, C.K. & Cressie, N. (1999). A dimension-reduced approach to space–time Kalman filtering, *Biometrika* **86**, 815–829.
- [23] Wikle, C.K., Berliner, L.M. & Cressie, N. (1998). Hierarchical Bayesian space–time models, *Environmental and Ecological Statistics* **5**, 117–154.
- [24] Zimmerman, D.L. & Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters, *The Annals of the Institute of Statistical Mathematics* **44**, 27–43.

(See also **Gaussian process; Modeling, stochastic; Point processes, dynamic; Point processes, spatial–temporal**)

NOEL CRESSIE & CHRISTOPHER K. WIKLE