



ELSEVIER

Computational Statistics & Data Analysis 22 (1996) 159–175

**COMPUTATIONAL  
STATISTICS  
& DATA ANALYSIS**

# Spatio-temporal prediction of snow water equivalent using the Kalman filter

Hsin-Cheng Huang, Noel Cressie\*

*Department of Statistics, Iowa State University, Ames, IA 50011, USA*

Received February 1995; revised July 1995

---

## Abstract

Consider a spatio-temporal stochastic process  $\{Z(s; t): s \in D; t = 1, 2, \dots\}$  and suppose it is of interest to predict  $\{Z(s; t_0): s \in D\}$  at some fixed time point  $t_0$ . Purely spatial methods use data  $Z(s_1; t_0), \dots, Z(s_n; t_0)$  to construct a spatial predictor (e.g., kriging). But, when data  $\{Z(s_i; t): i = 1, \dots, n; t = 1, 2, \dots, t_0\}$  are available, it is advantageous to treat the problem as one of spatio-temporal prediction. The US National Weather Service now use current snow water equivalent (SWE) data and a purely spatial model to predict SWE at sites where no observations are available. To improve SWE predictions, we introduce a spatio-temporal model that incorporates the SWE data from the past, resulting in a Kalman-filter prediction algorithm. A simple procedure for estimating the parameters in the model is developed and an example is presented for the Animas River basin in southwest Colorado.

**Keywords:** Cross-validation; Kriging; Second-order stationary; Spatio-temporal model

---

## 1. Introduction

The increased demands of water for industry, agriculture, and urban living make the reliable forecasting of water resources an extremely important problem. As a result, the US National Weather Service has developed a set of hydrologic simulation models used to generate extended streamflow predictions, water supply

---

\*Corresponding author.

outlook, and flood forecasts. These models are intended to answer the three questions: How much snow has fallen in a particular winter? How much water will that snow turn into? Where will that water go?

One of the most important parts of the hydrologic simulation models is the snow accumulation and ablation model that uses observed temperature and precipitation data to simulate snow cover conditions. This model is updated during the snow season through estimates of water content in snow cover (called snow water equivalent, abbreviated here as SWE) so as to maintain accurate streamflow and water supply forecasts. Castruccio et al. (1980) state that the benefit of a 6% improvement in streamflow predictions could be as high as 10 million dollars for hydropower and 28 million dollars for irrigation annually in the western USA. Hence precise estimation of SWE becomes an attractive goal.

To estimate SWE accurately, snow data from the snow course sites have been manually collected during the snow seasons for as many as 50–60 years. Also, snow data from the automated sensor SNOTEL sites and remotely sensed airborne estimates of SWE have recently become available. Carroll et al. (1995) have shown how, at a given date, to use current spatial data in a statistical model to predict SWE at sites where no observations are available. In this article, we shall construct a spatio-temporal model that can incorporate information from past dates in the same year in order to obtain more reliable SWE predictions.

In Section 2, the proposed spatio-temporal model is described, a special case of which is the purely spatial model that is currently being used by the US National Weather Service. Section 3 considers parameter estimation for the models given in Section 2. Section 4 applies the prediction equations derived in Section 3 to SWE in the Animas River basin of Colorado and compares their performance through cross-validation. Discussion of the results is given in Section 5.

## 2. The proposed spatio-temporal model

To illustrate our approach, we concentrate on the Animas River basin in southwest Colorado. SWE data used in this article are collected from 13 SNOTEL sites in the basin. The site locations are displayed in Fig. 1. Data from the 13 sites have been observed daily during the snow season, anywhere between three and fifteen years. Because the US National Weather Service is interested in updates at weekly intervals, we consider only the data observed at the six time points, each one week apart, from February 25 through April 1 (inclusive), for each year (1987–1993), and at each site in the study area.

### 2.1. Exploratory data analysis

Fig. 2 shows the plot of average SWE as a function of weeks, for years 1987–1993, where the average is taken over the 13 SNOTEL sites. Fig. 3 shows the plot of average SWE as a function of weeks, for the 13 SNOTEL sites, where the average is taken over years. Fig. 4 shows the plot of average SWE as a function of years, for

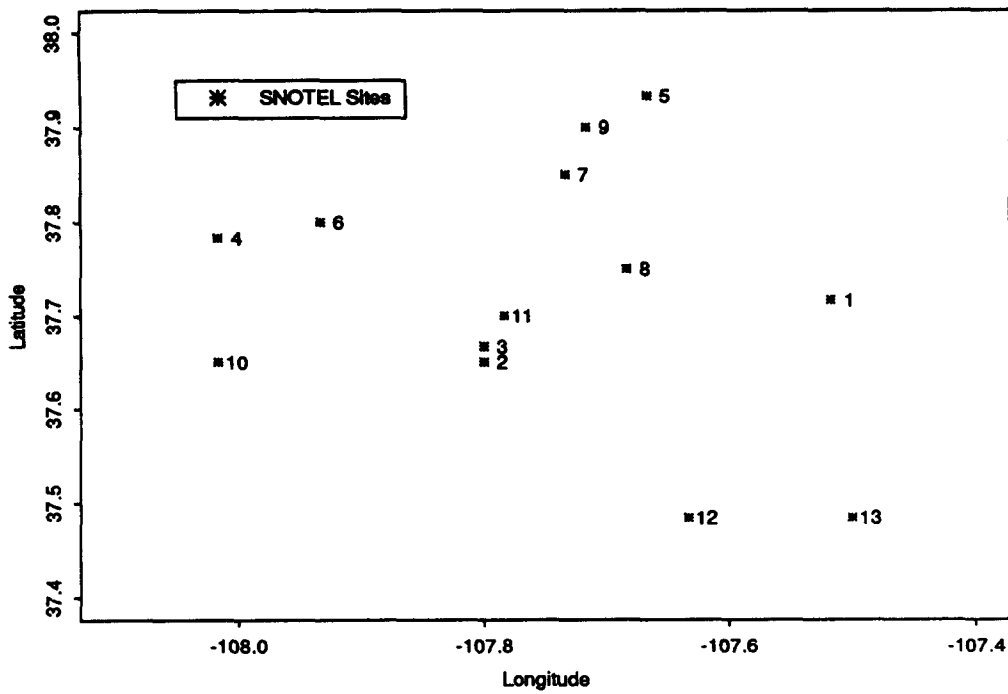


Fig. 1. Animas river basin and locations of SNOTEL sites in southwest Colorado.

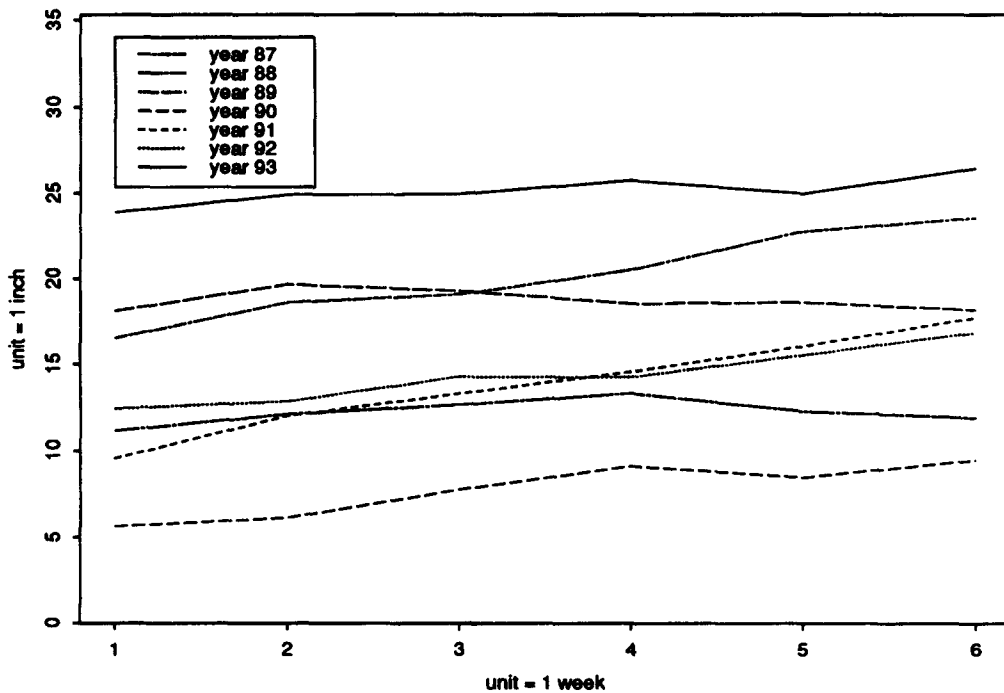


Fig. 2. Average of SWE over 13 sites from February 25 to April 1 for 1987 to 1993.

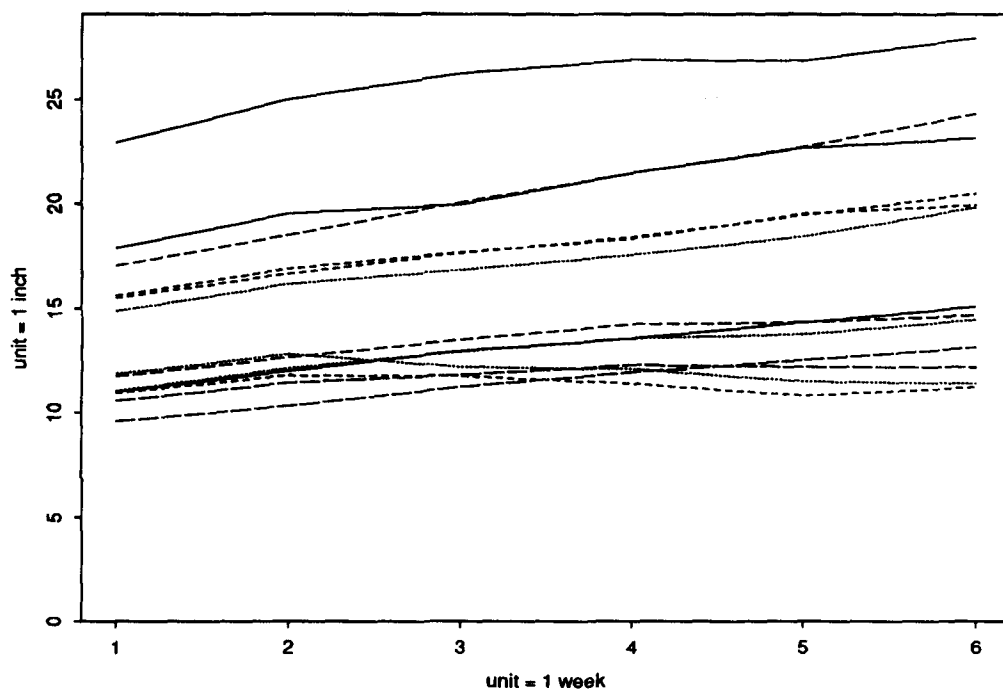


Fig. 3. Average of SWE over years 1987 to 1993 from February 25 to April 1 for 13 SNOTEL sites. Each line indicates one site.

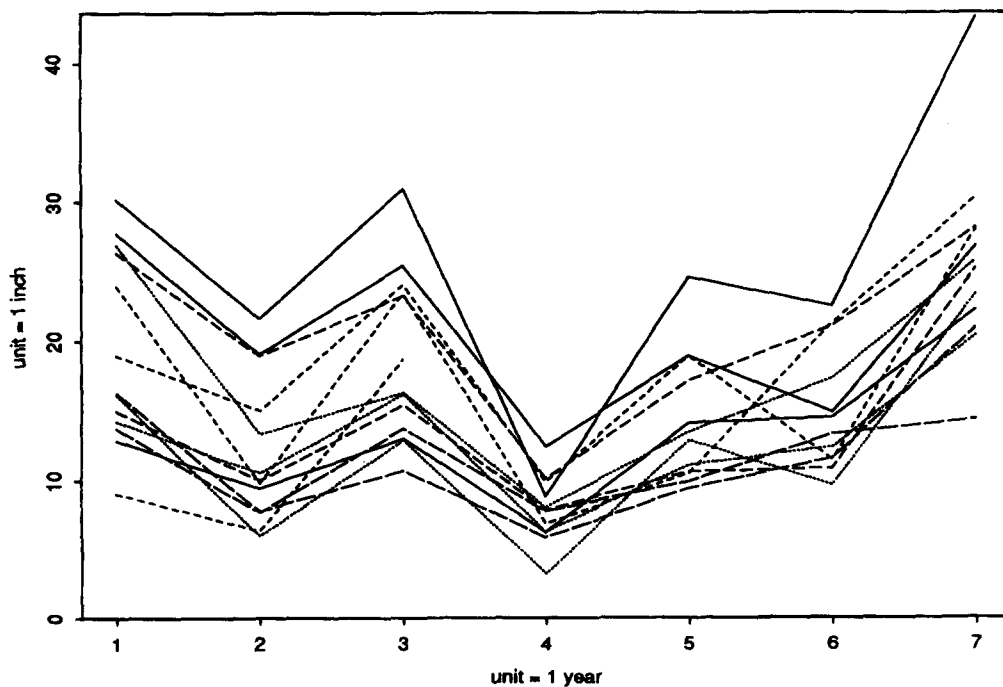


Fig. 4. Average of SWE over six weeks (February 25 to April 1) from 1987 to 1993 for 13 SNOTEL sites. Each line indicates one site.

the 13 SNOTEL sites, where the average is taken over weeks. From these plots, strong location effects, year effects, week effects and some possible interaction can be seen. Consequently, the large-scale variation among sites, among years, and among weeks should be considered. It should be noted that the National Weather Service currently uses spatial prediction (kriging) methods that account for this large-scale variation by kriging residuals from a mean (over years) map.

Furthermore, the standard deviations of SWE tends to be larger for larger mean SWE values. This can be seen in Fig. 5, the plot of the standard deviation versus mean SWE, where the means and standard deviations are taken over years for each SNOTEL site and week. These sorts of relations are used to standardize the data to have zero mean and unit variance.

We standardize the historical data at SNOTEL sites by subtracting the means and dividing by the standard deviations at those sites. For locations where no observations are available, and so kriging is required, mean maps produced by the National Weather Service on a weekly basis (Day, 1990) are used for mean SWE and the standard deviation is modeled as a function of the mean (e.g. Fig. 5). Details are given below.

Let  $\{s_i; i = 1, \dots, n\}$  denote the locations of  $n$  SNOTEL sites. The standardized SWE  $Z_{jt}(s)$ , for week  $t$  in year  $j$  at location  $s$ , derived by the National Weather

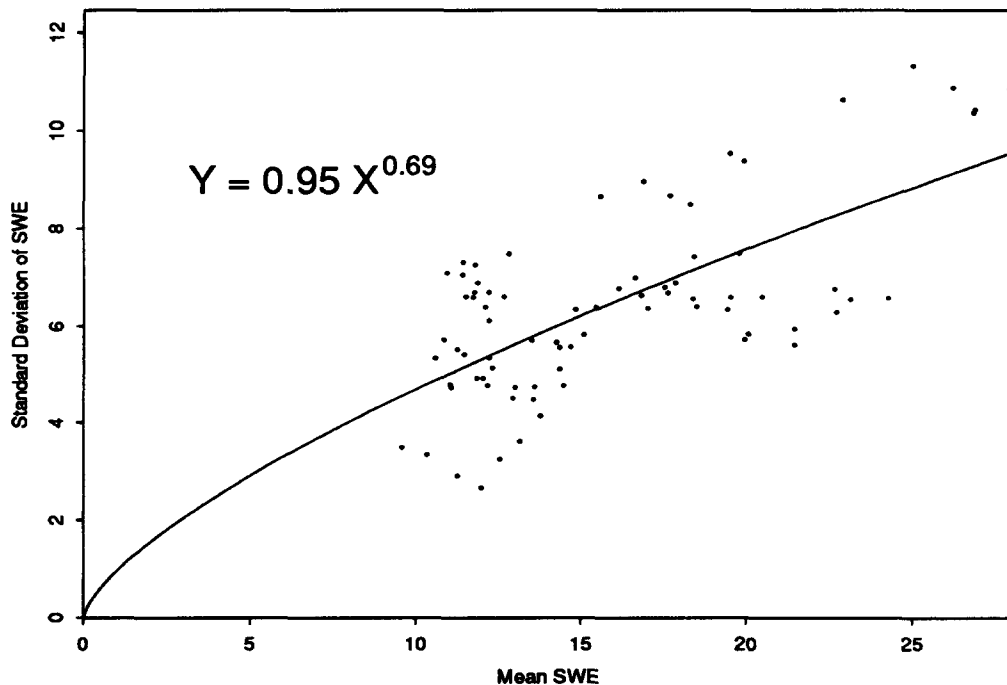


Fig. 5. Sample standard deviation versus sample mean (over years) of SWE. The super-imposed line shows the least-squares fit.

Service is as follows:

$$Z_{jt}(s) = \frac{Y_{jt}(s) - \mu_t(s)}{\sigma_t(s)}, \quad (2.1)$$

where  $t = 1, \dots, t_0$ ,  $\mu_t(s_i)$  is the sample mean (over years) of observed SWE  $Y_{jt}(s_i)$  on week  $t$  at site  $s_i$  and  $\sigma_t(s_i)$  is the corresponding sample standard deviation. For locations where no observations are taken,  $\mu_t(s)$  is obtained from mean maps and  $\sigma_t(s)$  is estimated by  $C_1(\mu_t(s))^{C_2}$ . The estimates of  $C_1$  and  $C_2$  obtained from nonlinear least squares are 0.95 and 0.69, respectively (see Fig. 5).

Notice that the standardization (2.1) used by the National Weather Service does not consider the year effect, which could be important (see Fig. 2). To adjust for the year effect, we replace the standardized SWE  $Z_{jt}(s)$  by the adjusted standardized SWE  $Z_{jt}^*(s)$  as follows:

$$Z_{jt}^*(s) = Z_{jt}(s) - \bar{Z}_j, \quad (2.2)$$

where

$$\bar{Z}_j = \frac{1}{nt_0} \sum_{i=1}^n \sum_{t=1}^{t_0} Z_{jt}(s_i).$$

## 2.2. Spatio-temporal prediction

For the purposes of SWE prediction (as distinct from estimation of model parameters), only 1993 data will be used for SWE predictions in this study; hence we replace subscript  $jt$  with subscript  $t$  in the following to simplify the notation. To begin with, let us consider the purely spatial model currently used by the National Weather Service.

Suppose (e.g. Cressie, 1993, Section 3.1)

$$Z_t(s) = S_t(s) + \varepsilon_t(s); \quad s \in D \subseteq R^2, \quad (2.3)$$

where  $S_t(\cdot)$  is a zero-mean,  $L_2$ -continuous (i.e.,  $E[S_t(s + \delta) - S_t(s)]^2 \rightarrow 0$  as  $\delta \rightarrow 0$ ), second-order stationary process, and  $\varepsilon_t(\cdot)$  is a zero-mean, white-noise process, independent of  $S_t(\cdot)$ , with  $\text{var}[\varepsilon_t(s)] = \sigma_\varepsilon^2$ . In (2.3), it is assumed that  $\{S_t(\cdot): t = 1, 2, \dots\}$  and  $\{\varepsilon_t(\cdot): t = 1, 2, \dots\}$  are all mutually independent.

To obtain the estimate of  $S_t(s_0)$ , at location  $s_0$ , we could use ordinary kriging. Let the best linear unbiased predictor (BLUP) of  $S_t(s_0)$ , given  $\{Z_t(s_i): i = 1, \dots, n\}$ , be

$$\hat{S}_t(s_0) = \sum_{i=1}^n \lambda_i Z_t(s_i),$$

where  $\sum_{i=1}^n \lambda_i = 1$ . If  $s_0 \neq s_i; i = 1, \dots, n$ , then  $\lambda_1, \dots, \lambda_n$  can be obtained from (e.g. Cressie, 1993, pp. 121–122)

$$\lambda_0 = A_0^{-1} \gamma_0,$$

where

$$\begin{aligned}\lambda_0 &\equiv (\lambda_1, \dots, \lambda_n, m)', \\ \gamma_0 &\equiv (\gamma_Z(s_0 - s_1), \dots, \gamma_Z(s_0 - s_n), 1)', \\ A_0 &\equiv \begin{cases} \gamma_Z(s_i - s_j); & i = 1, \dots, n, \quad j = 1, \dots, n, \\ 1; & i = n+1, \quad j = 1, \dots, n, \\ 0; & i = n+1, \quad j = n+1, \end{cases} \\ \gamma_Z(h) &\equiv \frac{1}{2} \text{var}[Z_t(s+h) - Z_t(s)],\end{aligned}$$

and  $A_0$  is symmetric. The kriging (or prediction) variance is given by  $\Sigma(s_0) = \gamma_0' A_0^{-1} \gamma_0$ .

If  $s_0 \in \{s_i: i = 1, \dots, n\}$ , then  $\lambda_1, \dots, \lambda_n$  can be obtained from (Cressie, 1993, pp. 128)

$$\lambda_0 = A_0^{-1} \gamma_0^*,$$

where  $\gamma_0^*$  is the same as  $\gamma_0$  except the  $k$ th component is replaced by  $\sigma_\varepsilon^2$ . The kriging (or prediction) variance is given by  $\Sigma(s_0) = \gamma_0^{*'} A_0^{-1} \gamma_0^* - \sigma_\varepsilon^2$ ;  $s_0 \in \{s_i: i = 1, \dots, n\}$ .

Notice that for ordinary kriging, only data obtained in the current week  $t$  is used to predict the unknown SWE,  $S_t(s_0)$ , at location  $s_0$ . In what follows, we use a spatio-temporal model and a Kalman filter to incorporate past data as well as current data into the prediction of SWE. Recall Eq. (2.3); this yields a vector relation given by

$$Z_t = S_t + \varepsilon_t; \quad t = 0, 1, 2, \dots, \quad (2.4)$$

where all the vectors are  $n \times 1$  corresponding to the  $n$  SNOTEL sites. The unobserved state processes  $\{S_t(s): s \in D, t = 0, 1, 2, \dots\}$  are assumed to be mean zero, spatially and temporally stationary Gaussian processes. Suppose they evolve in a Markov manner, which says that only recent weeks' SWE values influence the current week's value. This has proved to be a very effective assumption for modeling precipitation in general. Linearizing this temporal dependence leads to a  $p$ th order autoregressive process,

$$S_t(s) = \alpha_1 S_{t-1}(s) + \alpha_2 S_{t-2}(s) + \dots + \alpha_p S_{t-p}(s) + \eta_t(s), \quad s \in D, \quad (2.5)$$

where the coefficients  $\alpha_1, \dots, \alpha_p$  are chosen so that all (possibly complex) roots of  $\lambda^p - \sum_{i=1}^p \alpha_i \lambda^{p-i} = 0$  are less than 1 in absolute value in order to achieve temporal stationarity. Then the spatio-temporal model we shall consider is given by (2.4) and (2.5). In (2.4) and (2.5),

$Z_t \equiv (Z_t(s_1), \dots, Z_t(s_n))'$  is the vector of the standardized SWE observations for the  $n$  SNOTEL sites at time  $t$ ;

$S_t \equiv (S_t(s_1), \dots, S_t(s_n))'$  is the corresponding vector of the state variable for the  $n$  SNOTEL sites;

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2 I)$  is a white-noise process, independent of  $S_t$ ;

$\{\eta_t(s): s \in D, t = 0, 1, 2, \dots\}$  is a zero mean, spatially, and temporally stationary Gaussian process such that  $\eta_{t_1}(\cdot)$  and  $\eta_{t_2}(\cdot)$  are independent for all  $t_1 \neq t_2$ ;  $\eta_t(\cdot)$  and  $S_{t-1}(\cdot)$  are independent for all  $t$ ;  $\varepsilon_{t_1}$  and  $\eta_{t_2}(\cdot)$  are independent for all  $t_1, t_2$ ; and  $\alpha_1, \dots, \alpha_p$  are constants.

Notice that (2.5) can be written as the following first-order Markov process,

$$\begin{pmatrix} S_{t-p+1}(s) \\ S_{t-p+2}(s) \\ \vdots \\ S_t(s) \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & & 1 \\ \alpha_p & \alpha_{p-1} & \cdots & \alpha_1 \end{pmatrix} \begin{pmatrix} S_{t-p}(s) \\ S_{t-p+1}(s) \\ \vdots \\ S_{t-1}(s) \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \eta_t(s) \end{pmatrix}. \quad (2.6)$$

Due to temporal stationarity, we see that upon post-multiplying (2.5) by  $S_{t-j}(\mathbf{r})$  and taking expectations, we have

$$C_j^{(S)}(s, \mathbf{r}) - \sum_{m=1}^p \alpha_m C_{j-m}^{(S)}(s, \mathbf{r}) = \delta_{j0} C^{(n)}(s, \mathbf{r}), \quad j = 0, 1, 2, \dots, \quad (2.7)$$

where  $C_h^{(S)}(s, \mathbf{r}) \equiv \text{cov}[S_t(s), S_{t+h}(\mathbf{r})]$ , for all  $h = \dots, -1, 0, 1, \dots$ ,  $C^{(n)}(s, \mathbf{r}) \equiv \text{cov}[\eta_t(s), \eta_t(\mathbf{r})]$ ; and

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Set

$$\Gamma_j^{(S)}(s, \mathbf{r}) \equiv \begin{pmatrix} C_j^{(S)}(s, s) & C_j^{(S)}(s, \mathbf{r}) \\ C_j^{(S)}(\mathbf{r}, s) & C_j^{(S)}(\mathbf{r}, \mathbf{r}) \end{pmatrix}$$

and

$$\Gamma^{(n)}(s, \mathbf{r}) \equiv \begin{pmatrix} C^{(n)}(s, s) & C^{(n)}(s, \mathbf{r}) \\ C^{(n)}(\mathbf{r}, s) & C^{(n)}(\mathbf{r}, \mathbf{r}) \end{pmatrix}.$$

Then  $\Gamma_h^{(S)}(s, \mathbf{r})$ ;  $h = \dots, -1, 0, 1, \dots$ , can be determined by solving the (Yule-Walker) equations,

$$\Gamma_j^{(S)}(s, \mathbf{r}) - \sum_{m=1}^p \alpha_m \Gamma_{j-m}^{(S)}(s, \mathbf{r}) = \delta_{j0} \Gamma^{(n)}(s, \mathbf{r}), \quad j = 0, 1, 2, \dots, \quad (2.8)$$

obtained from (2.7). Note that  $\Gamma_{-h}^{(S)}(s, \mathbf{r}) = [\Gamma_h^{(S)}(s, \mathbf{r})]'$ . Hence, the first  $(p+1)$  of Eqs. (2.8) can be solved for  $\Gamma_0^{(S)}(s, \mathbf{r}), \dots, \Gamma_p^{(S)}(s, \mathbf{r})$ . The remaining equations then give  $\Gamma_{p+1}^{(S)}(s, \mathbf{r}), \Gamma_{p+2}^{(S)}(s, \mathbf{r}), \dots$ , recursively.

### 2.2.1. Derivation of the Kalman-filter procedure

The Kalman filter is a recursive algorithm for inference about the state variables of a model written in a state space form (e.g., the  $S$  variables defined according to the model (2.6)). Since it can be shown that

$$\hat{S}_{t'|t}(s) \equiv E[S_{t'}(s) | \mathbf{Z}_1, \dots, \mathbf{Z}_t]$$

is a linear function of  $\mathbf{Z}_t$  and  $\hat{S}_{t'-1|t-1}(s)$  for all  $t'$ , the Kalman filter enables the state variable to be updated once new observations are available. In the derivation below, we compute the optimal spatio-temporal predictor of  $S_t(s_0)$  (i.e.,  $E[S_t(s_0) | \mathbf{Z}_1, \dots, \mathbf{Z}_t]$ ), not only for sites where observations are taken but also for locations where no observations are available.



Assume the models (2.4) and (2.6) hold. We establish the following notation:

$$\hat{Z}_{t'|t}(\mathbf{s}) \equiv E[Z_{t'}(\mathbf{s}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t],$$

$$\hat{\mathbf{Z}}_{t'|t} \equiv E[\mathbf{Z}_{t'} | \mathbf{Z}_1, \dots, \mathbf{Z}_t],$$

$$\Sigma_{t_1, t_2|t}(\mathbf{s}, \mathbf{r}) \equiv \text{cov}[S_{t_1}(\mathbf{s}), S_{t_2}(\mathbf{r}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t],$$

$$\Sigma_{t'|t}(\mathbf{s}, \mathbf{r}) \equiv \text{cov}[S_{t'}(\mathbf{s}), S_{t'}(\mathbf{r}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t],$$

$$\Sigma_{t'|t}(\mathbf{s}, \mathbf{r}) \equiv \text{cov}[(S_{t'-p+1}(\mathbf{s}), \dots, S_{t'}(\mathbf{s}))', (S_{t'-p+1}(\mathbf{r}), \dots, S_{t'}(\mathbf{r}))' | \mathbf{Z}_1, \dots, \mathbf{Z}_t],$$

$$\mathbf{F}_t \equiv \text{var}[\mathbf{Z}_t | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}],$$

$$\mathbf{G}_t(\mathbf{s}) \equiv \text{cov}[\mathbf{Z}_t, (S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}))' | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}],$$

$$\boldsymbol{\alpha} \equiv (\alpha_p, \alpha_{p-1}, \dots, \alpha_1)'$$

and

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{e}'_2 \\ \vdots \\ \mathbf{e}'_p \\ \boldsymbol{\alpha}' \end{pmatrix} \equiv \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & & & 1 \\ \alpha_p & \alpha_{p-1} & \cdots & \alpha_1 \end{pmatrix}.$$

All joint distributions are multivariate Gaussian. Therefore, the conditional distribution of  $(S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}), S_{t-p+1}(\mathbf{r}), \dots, S_t(\mathbf{r}), \mathbf{Z}'_t)'$ , conditional on  $(\mathbf{Z}_1, \dots, \mathbf{Z}_{t-1})$ , is multivariate Gaussian with mean

$$(\hat{S}_{t-p+1|t-1}(\mathbf{s}), \dots, \hat{S}_{t|t-1}(\mathbf{s}), \hat{S}_{t-p+1|t-1}(\mathbf{r}), \dots, \hat{S}_{t|t-1}(\mathbf{r}), \hat{\mathbf{Z}}'_{t|t-1})'$$

and covariance matrix

$$\begin{pmatrix} \Sigma_{t|t-1}(\mathbf{s}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) & (\mathbf{G}_t(\mathbf{s}))' \\ \Sigma_{t|t-1}(\mathbf{r}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{r}, \mathbf{r}) & (\mathbf{G}_t(\mathbf{r}))' \\ \mathbf{G}_t(\mathbf{s}) & \mathbf{G}_t(\mathbf{r}) & \mathbf{F}_t \end{pmatrix}.$$

Consequently, the distribution of  $(S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}), S_{t-p+1}(\mathbf{r}), \dots, S_t(\mathbf{r}))'$ , conditional on  $(\mathbf{Z}_1, \dots, \mathbf{Z}_t)$ , is multivariate Gaussian with mean

$$(\hat{S}_{t-p+1|t-1}(\mathbf{s}), \dots, \hat{S}_{t|t-1}(\mathbf{s}), \hat{S}_{t-p+1|t-1}(\mathbf{r}), \dots, \hat{S}_{t|t-1}(\mathbf{r}))' \\ + (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r}))' \mathbf{F}_t^{-1} (\mathbf{Z}_t - \hat{\mathbf{Z}}_{t|t-1}),$$

and covariance matrix

$$\begin{pmatrix} \Sigma_{t|t-1}(\mathbf{s}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) \\ \Sigma_{t|t-1}(\mathbf{r}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{r}, \mathbf{r}) \end{pmatrix} - (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r}))' \mathbf{F}_t^{-1} (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r}))$$

(e.g. Searle, 1971, p. 46). Hence, we have

$$\begin{pmatrix} \hat{S}_{t-p+1|t}(s) \\ \vdots \\ \hat{S}_{t|t}(s) \end{pmatrix} = \begin{pmatrix} \hat{S}_{t-p+1|t-1}(s) \\ \vdots \\ \hat{S}_{t|t-1}(s) \end{pmatrix} + (G_t(s))' F_t^{-1} (Z_t - \hat{Z}_{t|t-1}), \quad (2.9)$$

$$\Sigma_{t|t}(s, r) = \Sigma_{t|t-1}(s, r) - (G_t(s))' F_t^{-1} G_t(r), \quad (2.10)$$

where

$$\begin{aligned} \begin{pmatrix} \hat{S}_{t-p+1|t-1}(s) \\ \vdots \\ \hat{S}_{t|t-1}(s) \end{pmatrix} &\equiv E \left[ \begin{pmatrix} S_{t-p+1}(s) \\ \vdots \\ S_t(s) \end{pmatrix} \middle| Z_1, \dots, Z_{t-1} \right] \\ &= E[A(S_{t-p}(s), \dots, S_{t-1}(s))' + (0, \dots, 0, \eta_t(s))' | Z_1, \dots, Z_{t-1}] \\ &= A(\hat{S}_{t-p|t-1}(s), \dots, \hat{S}_{t-1|t-1}(s))', \end{aligned} \quad (2.11)$$

$$\begin{aligned} \hat{Z}_{t|t-1}(s) &\equiv E[Z_t(s) | Z_1, \dots, Z_{t-1}] \\ &= E[S_t(s) | Z_1, \dots, Z_{t-1}] \\ &= E[\alpha'(S_{t-p}(s), \dots, S_{t-1}(s))' | Z_1, \dots, Z_{t-1}] \\ &= \alpha'(\hat{S}_{t-p|t-1}(s), \dots, \hat{S}_{t-1|t-1}(s))'; \end{aligned} \quad (2.12)$$

$$\begin{aligned} \Sigma_{t|t-1}(s, r) &\equiv \text{cov} \left[ \begin{pmatrix} S_{t-p+1}(s) \\ \vdots \\ S_t(s) \end{pmatrix}, \begin{pmatrix} S_{t-p+1}(r) \\ \vdots \\ S_t(r) \end{pmatrix} \middle| Z_1, \dots, Z_{t-1} \right] \\ &= \text{cov} \left[ A \begin{pmatrix} S_{t-p}(s) \\ \vdots \\ S_{t-1}(s) \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ \eta_{t-1}(s) \end{pmatrix}, \right. \\ &\quad \left. A \begin{pmatrix} S_{t-p}(r) \\ \vdots \\ S_{t-1}(r) \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ \eta_{t-1}(r) \end{pmatrix} \middle| Z_1, \dots, Z_{t-1} \right] \\ &= A \Sigma_{t-1|t-1}(s, r) A' + \begin{pmatrix} 0 & 0 \\ 0 & C^{(n)}(s, r) \end{pmatrix}, \end{aligned} \quad (2.13)$$

$$\begin{aligned} F_t &= \text{var}[Z_t | Z_1, \dots, Z_{t-1}] \\ &= \text{var}[S_t | Z_1, \dots, Z_{t-1}] + \sigma_\varepsilon^2 I \end{aligned}$$

$$= \begin{bmatrix} \Sigma_{t|t-1}(s_1, s_1) & \cdots & \Sigma_{t|t-1}(s_1, s_n) \\ \vdots & \ddots & \vdots \\ \Sigma_{t|t-1}(s_n, s_1) & \cdots & \Sigma_{t|t-1}(s_n, s_n) \end{bmatrix} + \sigma_\varepsilon^2 I \quad (2.14)$$

$$= \begin{bmatrix} \alpha' \Sigma_{t-1|t-1}(s_1, s_1) \alpha & \cdots & \alpha' \Sigma_{t-1|t-1}(s_1, s_n) \alpha \\ \vdots & \ddots & \vdots \\ \alpha' \Sigma_{t-1|t-1}(s_n, s_1) \alpha & \cdots & \alpha' \Sigma_{t-1|t-1}(s_n, s_n) \alpha \end{bmatrix} + \sigma_\varepsilon^2 I \quad (2.15)$$

$$G_t(s) \equiv \text{cov} \left[ Z_t, \begin{pmatrix} S_{t-p+1}(s) \\ \vdots \\ S_t(s) \end{pmatrix} \middle| Z_1, \dots, Z_{t-1} \right] \\ = \text{cov} [S_t, (S_{t-p+1}(s), \dots, S_t(s))' | Z_1, \dots, Z_{t-1}] \\ = \begin{bmatrix} \Sigma_{t,t-p+1|t-1}(s_1, s) & \cdots & \Sigma_{t|t-1}(s_1, s) \\ \vdots & \ddots & \vdots \\ \Sigma_{t,t-p+1|t-1}(s_n, s) & \cdots & \Sigma_{t|t-1}(s_n, s) \end{bmatrix} \quad (2.16)$$

$$= \begin{bmatrix} \alpha' \Sigma_{t-1|t-1}(s_1, s) e_2 & \cdots & \alpha' \Sigma_{t-1|t-1}(s_1, s) \alpha \\ \vdots & \ddots & \vdots \\ \alpha' \Sigma_{t-1|t-1}(s_n, s) e_2 & \cdots & \alpha' \Sigma_{t-1|t-1}(s_n, s) \alpha \end{bmatrix}. \quad (2.17)$$

Thus, we have established the following theorem.

**Theorem 2.1.** Assume that models (2.4) and (2.5) hold. For any point  $s_0 \in D$  and for spatio-temporal data  $Z_1, Z_2, \dots, Z_{t_0}$ , let  $(\hat{S}_{-p+1|0}(s_0), \dots, \hat{S}_{-1|0}(s_0), \hat{S}_{0|0}(s_0))'$  be the initial predictor of  $(S_{-p+1}(s_0), \dots, S_{-1}(s_0), S_0(s_0))'$ , and  $\Sigma_{0|0}(s, r)$  be the covariance matrix between  $(S_{-p+1}(s), \dots, S_{-1}(s), S_0(s))' - (\hat{S}_{-p+1|0}(s), \dots, \hat{S}_{-1|0}(s), \hat{S}_{0|0}(s))'$  and  $(S_{-p+1}(r), \dots, S_{-1}(r), S_0(r))' - (\hat{S}_{-p+1|0}(r), \dots, \hat{S}_{-1|0}(r), \hat{S}_{0|0}(r))'$ . Then the minimum mean squared error predictor (MMSEP) of  $(S_{t_0-p+1}(s_0), \dots, S_{t_0}(s_0))'$  and its matrix of mean squared prediction errors,  $\Sigma_{t_0|t_0}(s_0, s_0)$ , can be obtained recursively from the updating equations, (2.9)–(2.15).

Notice that  $\Sigma_{t|t}(s, r)$  does not depend on the observations, which is to be expected because of the Gaussian assumptions. Thus, for all  $t_1, t_2 = t - p + 1, \dots, t$ , we have

$$\begin{aligned} \Sigma_{t_1, t_2|t}(s, r) &= \text{cov} [S_{t_1}(s), S_{t_2}(r) | Z_1, \dots, Z_t] \\ &= E[(S_{t_1}(s) - \hat{S}_{t_1|t}(s))(S_{t_2}(r) - \hat{S}_{t_2|t}(r)) | Z_1, \dots, Z_t] \\ &= E[(S_{t_1}(s) - \hat{S}_{t_1|t}(s))(S_{t_2}(r) - \hat{S}_{t_2|t}(r))] \\ &= \text{cov} [S_{t_1}(s) - \hat{S}_{t_1|t}(s), S_{t_2}(r) - \hat{S}_{t_2|t}(r)]. \end{aligned}$$

This means that  $\Sigma_{t|t}(\mathbf{s}, \mathbf{r})$  is actually the matrix of covariances between the prediction errors at locations  $\mathbf{s}$  and  $\mathbf{r}$ .

Recall that the process has mean zero and so a reasonable choice for the initial predictor is

$$\hat{S}_{-p+1|0}(\mathbf{s}), \dots, \hat{S}_{-1|0}(\mathbf{s}), \hat{S}_{0|0}(\mathbf{s})' = E[(S_{-p+1}(\mathbf{s}), \dots, S_{-1}(\mathbf{s}), S_0(\mathbf{s}))'] = \mathbf{0}.$$

It follows that

$$\Sigma_{0|0}(\mathbf{s}, \mathbf{r}) = \text{cov}[(S_{-p+1}(\mathbf{s}), \dots, S_{-1}(\mathbf{s}), S_0(\mathbf{s}))', (S_{-p+1}(\mathbf{r}), \dots, S_{-1}(\mathbf{r}), S_0(\mathbf{r}))'],$$

which is a  $p \times p$  matrix whose  $(i, j)$ th entry is  $C_{i-j}^{(S)}(\mathbf{s}, \mathbf{r})$ .

**Theorem 2.2.** *The one-step MMSEP for  $S_{t_0+1}(\mathbf{s}_0)$  is given by*

$$\hat{S}_{t_0+1|t_0}(\mathbf{s}_0) = \alpha_1 \hat{S}_{t_0|t_0}(\mathbf{s}_0) + \alpha_2 \hat{S}_{t_0-1|t_0}(\mathbf{s}_0) + \dots + \alpha_p \hat{S}_{t_0-p+1|t_0}(\mathbf{s}_0), \quad (2.18)$$

where  $\hat{S}_{t_0|t_0}(\mathbf{s}_0), \dots, \hat{S}_{t_0-p+1|t_0}(\mathbf{s}_0)$  are obtained from Theorem 2.1. Further, its mean squared prediction error is

$$\Sigma_{t_0+1|t_0}(\mathbf{s}_0, \mathbf{s}_0) = \alpha' \Sigma_{t_0|t_0}(\mathbf{s}_0, \mathbf{s}_0) \alpha + C^{(n)}(\mathbf{s}_0, \mathbf{s}_0),$$

where  $\Sigma_{t_0|t_0}(\mathbf{s}_0, \mathbf{s}_0)$  is obtained from Theorem 2.1.

**Proof.**

$$\begin{aligned} \hat{S}_{t_0+1|t_0}(\mathbf{s}_0) &= E[S_{t_0+1}(\mathbf{s}_0) | \mathbf{Z}_1, \dots, \mathbf{Z}_{t_0}] \\ &= E[\alpha'(S_{t_0-p+1}(\mathbf{s}_0), \dots, S_{t_0}(\mathbf{s}_0))' + \eta_{t_0+1}(\mathbf{s}_0) | \mathbf{Z}_1, \dots, \mathbf{Z}_{t_0}] \\ &= \alpha_1 \hat{S}_{t_0|t_0}(\mathbf{s}_0) + \alpha_2 \hat{S}_{t_0-1|t_0}(\mathbf{s}_0) + \dots + \alpha_p \hat{S}_{t_0-p+1|t_0}(\mathbf{s}_0). \\ \Sigma_{t_0+1|t_0}(\mathbf{s}_0, \mathbf{s}_0) &= \text{var}[S_{t_0+1}(\mathbf{s}_0) | \mathbf{Z}_1, \dots, \mathbf{Z}_{t_0}] \\ &= \text{var}[\alpha'(S_{t_0-p+1}(\mathbf{s}_0), \dots, S_{t_0}(\mathbf{s}_0))' + \eta_{t_0+1}(\mathbf{s}_0) | \mathbf{Z}_1, \dots, \mathbf{Z}_{t_0}] \\ &= \alpha' \text{var}[(S_{t_0-p+1}(\mathbf{s}_0), \dots, S_{t_0}(\mathbf{s}_0))' | \mathbf{Z}_1, \dots, \mathbf{Z}_{t_0}] \alpha + \text{var}[\eta_{t_0+1}(\mathbf{s}_0)] \\ &= \alpha' \Sigma_{t_0|t_0}(\mathbf{s}_0, \mathbf{s}_0) \alpha + C^{(n)}(\mathbf{s}_0, \mathbf{s}_0). \quad \square \end{aligned}$$

For the case where the data are corrected by year effects (see (2.2)), the models and the procedures remain the same except the standardized SWE data  $\{\mathbf{Z}_t\}$  are replaced by the adjusted standardized SWE data  $\{\mathbf{Z}_t^*\}$ .

The assumption of joint Gaussianity above can be easily relaxed without changing the result. Instead of looking for MMSEPs that are best in the class of all predictors, we look for the minimum mean squared error linear predictors (MMSELPs) that are best in the more restrictive class of linear predictors. Then the predictors given by (2.9) and (2.16) are MMSELPs in non-Gaussian settings. They are also unbiased. More discussion of this can be found in Harvey (1989, Chapter 3) or Duncan and Horn (1972).

### 3. Parameter estimation

For the purely spatial model given by (2.3), let

$$2\gamma_Z(\mathbf{h}) \equiv \begin{cases} 0 & \text{if } \mathbf{h} = \mathbf{0}, \\ \theta_1 + \theta_2(1 - \exp(-\theta_3 \|\mathbf{h}\|)) & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

This is a flexible three-parameter variogram that exhibits nugget effect ( $\theta_1$ ), sill ( $\theta_1 + \theta_2$ ), and spatial dependence ( $\theta_3$ ). Since the data from the SNOTEL sites were measured electronically, the potential for measurement error is much less than for data from snow course sites. Based on an analysis of repeated measurements of instruments over short time periods, we found it is reasonable to assume that  $\sigma_\epsilon^2$  is zero. The parameters that must be estimated are  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ , where  $\theta_i \geq 0$ ;  $i = 1, 2, 3$ . The empirical robust variogram estimator suggested by Cressie and Hawkins (1980) is given by

$$2\hat{\gamma}(\mathbf{h}(l)) \equiv \frac{[\text{ave}\{|Z_t(s_i) - Z_t(s_j)|^{1/2}: (i, j) \in N(\mathbf{h}(l))\}]^4}{0.457 + 0.494/|N(\mathbf{h}(l))|},$$

where  $N(\mathbf{h}(l)) = \{(i, j, t): s_i - s_j = \mathbf{h}, \mathbf{h} \in \text{Tol}(\mathbf{h}(l)), i, j = 1, \dots, n, t = 1, \dots, t_0\}$ ,  $\text{Tol}(\mathbf{h}(l))$  is some specified “tolerance” region around  $\mathbf{h}(l)$ ;  $l = 1, \dots, k$ , and  $|N(\mathbf{h}(l))|$  is the number of distinct elements in  $N(\mathbf{h}(l))$ ;  $l = 1, \dots, k$ . The parameter  $\boldsymbol{\theta}$  can then be estimated by fitting this empirical variogram to  $2\gamma_Z(\mathbf{h}; \boldsymbol{\theta})$ . We use the weighted least-squares method (e.g. Cressie, 1993, pp. 96–97) to estimate  $\boldsymbol{\theta}$ , by minimizing

$$\sum_{l=1}^k |N(\mathbf{h}(l))| \left\{ \frac{\hat{\gamma}(\mathbf{h}(l))}{\gamma(\mathbf{h}(l); \boldsymbol{\theta})} - 1 \right\}^2$$

over all possible  $\boldsymbol{\theta}$ .

Now we consider parameter estimation for the spatio-temporal model given by (2.4) and (2.5). Let the spatially stationary error process be described by the flexible three-parameter covariance function,

$$C^{(n)}(s, \mathbf{r}) \equiv \begin{cases} \phi_1 \exp(-\phi_2 \|s - \mathbf{r}\|) & \text{if } s \neq \mathbf{r} \\ \phi_1 + \phi_3 & \text{if } s = \mathbf{r}, \end{cases}$$

and again assume  $\sigma_\epsilon^2$  to be zero. From (2.4) and (2.5), the parameters that must be estimated are  $\alpha_1, \dots, \alpha_p, \phi_1, \phi_2$ , and  $\phi_3$ , where  $\phi_i \geq 0$ ;  $i = 1, 2, 3$ .

From (2.7), we have

$$\begin{pmatrix} C_1^{(S)}(s, s) \\ C_2^{(S)}(s, s) \\ \vdots \\ C_p^{(S)}(s, s) \end{pmatrix} = \begin{pmatrix} C_0^{(S)}(s, s) & C_1^{(S)}(s, s) & \cdots & C_{p-1}^{(S)}(s, s) \\ C_{-1}^{(S)}(s, s) & C_0^{(S)}(s, s) & \ddots & \\ \vdots & \ddots & \ddots & C_1^{(S)}(s, s) \\ C_{1-p}^{(S)}(s, s) & & C_{-1}^{(S)}(s, s) & C_0^{(S)}(s, s) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}.$$

Now  $C_k^{(S)}(s, s)$ ;  $k = 0, 1, \dots, p$ , can be estimated by a consistent estimator that averages over  $n$  locations,  $J$  years, and up to  $t_0$  weeks (Brockwell and Davis, 1991,

p. 220):

$$\hat{C}_k^{(S)} \equiv \frac{1}{nJt_0} \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^{t_0-k} \{Z_{jt}(s_i) - \bar{Z}\} \{Z_{j,t+k}(s_i) - \bar{Z}\},$$

where

$$\bar{Z} \equiv \frac{1}{nJt_0} \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^{t_0} Z_{jt}(s_i).$$

Therefore,  $(\alpha_1, \dots, \alpha_p)'$  can be estimated by

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} \hat{C}_0^{(S)} & \hat{C}_1^{(S)} & \cdots & \hat{C}_{p-1}^{(S)} \\ \hat{C}_{-1}^{(S)} & \hat{C}_0^{(S)} & \ddots & \\ & \ddots & \ddots & \hat{C}_1^{(S)} \\ \hat{C}_{1-p}^{(S)} & & \hat{C}_1^{(S)} & \hat{C}_0^{(S)} \end{pmatrix}^{-1} \begin{pmatrix} \hat{C}_1^{(S)} \\ \hat{C}_2^{(S)} \\ \vdots \\ \hat{C}_p^{(S)} \end{pmatrix}.$$

Notice that by solving the first  $(p+1)$  of Eqs. (2.8), we have

$$C_0^{(S)}(s, r) = f_p(\alpha_1, \dots, \alpha_p) C^{(n)}(s, r)$$

for a particular  $f_p(\alpha_1, \dots, \alpha_p)$  that is easy to calculate. Hence the variogram of  $Z_t(\cdot)$  between  $s_i$  and  $s_j$ , given by

$$\begin{aligned} \text{var}[Z_t(s_i) - Z_t(s_j)] &= \text{var}[S_t(s_i) - S_t(s_j)] \\ &= \text{var}[S_t(s_i)] + \text{var}[S_t(s_j)] - 2\text{cov}[S_t(s_i), S_t(s_j)] \\ &= C_0^{(S)}(s_i, s_i) + C_0^{(S)}(s_j, s_j) - C_0^{(S)}(s_i, s_j), \end{aligned}$$

can be estimated by

$$\begin{cases} 0 & \text{if } i = j, \\ 2\{\hat{C}_0^{(S)} - f_p(\hat{\alpha}_1, \dots, \hat{\alpha}_p) \phi_1 \exp(-\phi_2 \|s_i - s_j\|)\} & \text{if } i \neq j. \end{cases} \quad (3.1)$$

Thus,  $\phi_1$  and  $\phi_2$  can be estimated by fitting an empirical variogram based on  $\{Z_{jt}(s_i): i = 1, \dots, n, j = 1, \dots, J, t = 1, \dots, t_0\}$  to this two-parameter variogram model. The fitted values of  $\phi_1$  and  $\phi_2$  are denoted by  $\hat{\phi}_1$  and  $\hat{\phi}_2$ .

Because

$$C_0^{(S)}(s, s) = (\phi_1 + \phi_3) f_p(\alpha_1, \dots, \alpha_p),$$

it follows that  $\phi_3$  can be estimated by

$$\hat{\phi}_3 \equiv \min \left\{ \frac{\hat{C}_0^{(S)}}{f_p(\hat{\alpha}_1, \dots, \hat{\alpha}_p)} - \hat{\phi}_1, 0 \right\}. \quad (3.2)$$

For the case where the data are corrected by year effects (see (2.2)), the parameter estimation procedures remain the same except the standardized SWE data  $\{Z_t\}$  are replaced by the adjusted standardized SWE data  $\{Z_t^*\}$ .

#### 4. Application to snow water equivalent data

In this section, the results of fitting the data to a purely spatial model and to the spatio-temporal model are illustrated. Six weekly observations from February 25 through April 1 are taken for each year (1987–1993) at each SNOTEL site in the Animas River basin. The data were first standardized by either (2.1) or (2.2). For the purely spatial model, we use only the standardized SWE data on 1 April 1993 to obtain SWE predictions at any location on 1 April 1993. For the spatio-temporal model, we use all standardized SWE data (adjusted or not adjusted for year effects) to estimate the parameters and use only the standardized SWE data (adjusted or not adjusted, respectively) from the three most recent dates, 18 March 1993 through 1 April 1993, to obtain SWE predictions on 1 April 1993.

Three cases are considered in this study and the parameters estimated for these cases are given as follows.

*Case 1.* For the purely spatial model (2.3) and the standardized SWE obtained from (2.1), the estimated variogram parameters are  $\hat{\theta}_1 = 0.2585$ ,  $\hat{\theta}_2 = 1.4426$ , and  $\hat{\theta}_3 = 0.0010$ .

*Case 2.* For AR(1) spatio-temporal model (2.4) and (2.5), and the standardized SWE obtained from (2.1),  $\hat{\alpha}_1 = 0.82267$ ,  $\hat{\phi}_1 = 0.2462$ ,  $\hat{\phi}_2 = 0.0141$ ,  $\hat{\phi}_3 = 0.0285$ .

*Case 3.* For AR(2) spatio-temporal model (2.4) and (2.5), and the adjusted standardized SWE obtained from (2.2) (i.e., adjusted for year effect),  $\hat{\alpha}_1 = 0.8497$ ,  $\hat{\alpha}_2 = -0.1090$ ,  $\hat{\phi}_1 = 0.0067$ ,  $\hat{\phi}_2 = 0.0501$ , and  $\hat{\phi}_3 = 0.0419$ .

Notice that it is not necessary to consider the year-effect adjustment in the purely spatial model, because ordinary kriging used therein already takes the year effect into account.

We use cross-validation techniques to compare these three cases. Cross-validation is implemented by removing data  $\{Z_{jt}(s_i): t = 1, \dots, t_0\}$  for each  $i$ , and predicting  $Z_{jt_0}(s_i)$  from the remaining data. The resulting predicted value and prediction variance of the standardized SWE for April 1 on year  $j$  at site  $i$  are denoted by  $\hat{Z}_{jt_0}^{(-i)}(s_i)$  and  $\hat{\Sigma}_{jt_0}^{(-i)}(s_i)$ , respectively. Then the predicted value and the prediction variance of unstandardized SWE for April 1 on year  $j$  at site  $i$  are

$$\hat{Y}_{jt_0}^{(-i)}(s_i) \equiv \hat{Z}_{jt_0}^{(-i)}(s_i) \sigma_{t_0}(s_i) + \mu_{t_0}(s_i),$$

$$\hat{\Sigma}_{Y_{jt_0}}^{(-i)}(s_i) \equiv \hat{\Sigma}_{Z_{jt_0}}^{(-i)}(s_i) \sigma_{t_0}^2(s_i).$$

For the case where the data are corrected by year effects (see (2.2)), the predicted value and the prediction variance of the adjusted standardized SWE for April 1 on year  $j$  at site  $i$  are denoted by  $\hat{Z}_{jt_0}^{*(-i)}(s_i)$  and  $\hat{\Sigma}_{Z_{jt_0}^*}^{(-i)}(s_i)$ , respectively. Then the predicted value and the prediction variance of unstandardized SWE for April 1 on year  $j$  at site  $i$  are

$$\hat{Y}_{jt_0}^{(-i)}(s_i) \equiv \{\hat{Z}_{jt_0}^{*(-i)}(s_i) + \bar{Z}_j\} \sigma_{t_0}(s_i) + \mu_{t_0}(s_i),$$

$$\hat{\Sigma}_{Y_{jt_0}}^{(-i)}(s_i) \equiv \hat{\Sigma}_{Z_{jt_0}^*}^{(-i)}(s_i) \sigma_{t_0}^2(s_i).$$

The cross-validation statistics suggested by Carroll and Cressie (1996) are as follows:

$$\begin{aligned} \text{CRV}_1 &\equiv \frac{1}{n} \sum_{i=1}^n \frac{Y_{j_{t_0}}(s_i) - \hat{Y}_{j_{t_0}}^{(-i)}(s_i)}{\{\hat{\Sigma}_{Y_{j_{t_0}}}^{(-i)}(s_i)\}^{1/2}}, \\ \text{CRV}_2 &\equiv \left\{ \frac{1}{n} \sum_{i=1}^n \frac{[Y_{j_{t_0}}(s_i) - \hat{Y}_{j_{t_0}}^{(-i)}(s_i)]^2}{\hat{\Sigma}_{Y_{j_{t_0}}}^{(-i)}(s_i)} \right\}^{1/2}, \\ \text{CRV}_3 &\equiv \left\{ \frac{1}{n} \sum_{i=1}^n [Y_{j_{t_0}}(s_i) - \hat{Y}_{j_{t_0}}^{(-i)}(s_i)]^2 \right\}^{1/2}, \end{aligned}$$

where here we compute them for  $j = 1993$ . The quantity  $\text{CRV}_1$  checks the unbiasedness of the predictor and should be approximately equal to zero;  $\text{CRV}_2$  checks the accuracy of the standard deviation of the prediction error and should be approximately equal to one; and  $\text{CRV}_3$ , the quantity we focus on, is a measure of goodness of prediction. The results of the cross-validation statistics are shown in Table 1.

Inspection of Table 1 indicates that the values of  $\text{CRV}_1$  and  $\text{CRV}_2$  are all reasonable, with case 3 showing worse bias than the others. However, case 3 performs better than the other two cases by having the smallest  $\text{CRV}_3$  value. On average, there is a 20% improvement in root mean squared prediction errors using the spatio-temporal approach with year-effect adjustment.

When we looked at the individual values of  $\{Y_{t_0}(s_i) - \hat{Y}_{t_0}^{(-i)}(s_i); i = 1, \dots, 13\}$  for each case, we found that the predicted SWE of site 1 is always substantially different from the observed SWE of site 1 in each case. This is because the standardized SWE of site 1 is much smaller than the standardized SWE of the others. Therefore, we may consider site 1 as an outlier. The cross-validation statistics with site 1 removed are shown in Table 2. From Table 2, we can see that the values of  $\text{CRV}_1$  and  $\text{CRV}_2$  are still reasonable for each case and case 3 again has a smaller  $\text{CRV}_3$  value than the other two cases. The improvement in root mean squared prediction error using case 3 increases to 29% when the outlying site 1 is removed.

Table 1

Cross-validation statistics. Case 1: purely spatial model; case 2: AR(1) spatio-temporal model; case 3: AR(2) spatio-temporal model adjusted for year effects.

Model	CRV <sub>1</sub>	CRV <sub>2</sub>	CRV <sub>3</sub>
Case 1	0.08818	0.99209	2.51406
Case 2	-0.02104	1.03040	2.31370
Case 3	0.26646	1.06048	2.08728

Table 2

Cross-validation statistics with site 1 removed

Model	CRV <sub>1</sub>	CRV <sub>2</sub>	CRV <sub>3</sub>
Case 1	-0.07941	0.83608	2.14515
Case 2	-0.13997	0.91231	1.93585
Case 3	0.10928	0.91225	1.66593



## 5. Discussion

In this article, we have proposed a spatio-temporal model and showed how to use the Kalman filter algorithm to obtain SWE prediction for locations where no observations are taken. For SWE data from the Animas River basin of southwest Colorado, the spatio-temporal model performs better than the purely spatial model, even when the standardized SWE are not adjusted for year effects. A further reduction of the  $CRV_3$  value is observed when there is an adjustment for year effects. Although the spatio-temporal model presented has been developed for the analysis of SWE data, this model should also prove useful for many other spatio-temporal prediction problems.

## Acknowledgements

We would like to thank Dr. Steven S. Carroll of Arizona State University for his assistance in obtaining and interpreting the data and for many helpful discussions on snow water equivalent. This research was supported by the National Weather Service, the National Science Foundation (DMS-9204521) and the Environmental Protection Agency (CR 822919–01–0).

## References

- Brockwell, P.J. and R.A. Davis, *Time Series: Theory and Methods*, 2nd edn. (Springer, New York, 1991).
- Carroll, S.S. and N. Cressie, Spatial modeling of snow water equivalent using covariances estimated from spatial and geomorphic attributes, *Journal of Hydrology*, tentatively accepted (1996).
- Carroll, S.S., G.N. Day, N. Cressie and T.R. Carroll, Spatial modeling of snow water equivalent using airborne and ground-based snow data, *Environmetrics*, **6** (1995) 127–139.
- Castruccio, P.A., H. Loats, Jr., D. Lloyd and P.A.B. Newman, Cost/benefit analysis for the operational applications of satellite snowcover observations (OASSO), *Proc. Final Workshop on Operational Applications of Satellite Snowcover Observations*, Washington, DC NASA CP-2116 (1980) 239–254.
- Cressie, N., *Statistics for Spatial Data*, revised edn. (Wiley, New York, 1993).
- Cressie, N. and D.M. Hawkins, Robust estimation of the variogram, I. *J. Int. Assoc. Math. Geology*, **12** (1980) 115–125.
- Day, G.N., A methodology for updating a conceptual snow model with snow measurements, NOAA Technical Report, NWS 43 (Silver Spring, MD, 1990).
- Duncan, D.B. and S.D. Horn, Linear dynamic regression from the viewpoint of regression analysis, *J. Amer. Statist. Assoc.* **67** (1972) 815–821.
- Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge, 1989).
- Searle, S.R., *Linear Models* (Wiley, New York, 1971).