

# Spatial-Temporal Modeling of Active Layer Thickness

Qian Chen

*Advisor:* Dr. Tatiyana Apanasovich

*Department of Statistics  
The George Washington University*

## Abstract

The objective of this study is to provide the methodology to model and estimate spatial-temporal variation in the active layer thickness (ALT) at the U1 Barrow site of the Circumpolar Active Layer Monitoring network, and to demonstrate its use in spatial-temporal interpolation. Specifically, we use 18 years of data (1995–2012) collected on 11 by 11 square grid of locations separated by 100 meters to build the model. Then, we use the data collected in 2013 to demonstrate the validity and predictive power of our methodology. In our study, we propose two models that provide a realistic description of space-time variability in ALT. At the same time, these models are feasible to efficiently estimate model parameters from available data. Specifically, we adopt linear modeling approach. The main modeling difficulties lie in defining a deterministic trend that represents the large scale spatial and temporal variation, and a realistic stochastic model that characterizes the space-time dependency of the residuals. Formulations that take into account interactions among spatial and temporal components are also developed. Fitting the space-time geostatistical model can be computationally demanding since the number of observations is large. Hence, we use a composite likelihood approach which is a criterion function based on the likelihood of marginal events. In our data analysis, we demonstrate that our models resemble the empirical patterns. Moreover, we compare our models to the naive one, which does not take the spatial and temporal correlation in residuals into consideration. The root mean squared error is reduced by 27 percent when our approach is taken.

## 1 Introduction

### 1.1 Significance of the study

In recent years, Arctic regions have witnessed significant environmental and climatic changes such as rise in temperature, melting of glaciers, and the degradation of permafrost. In the past 100 years, average Arctic temperatures have been increasing at almost twice the rate of the rest of the world [1]. The active layer, i.e., the portion of earth between the surface soil and permafrost, is an archetypal component of the Arctic climate system; it freezes in winter and thaws in summer [13] and is sensitive to the change of high-latitude environments. Previous studies have indicated that thickening of the active layer will have numerous negative impacts on the health of natural and human systems: for example, thickening of the layer may result in damage to roads and utilities, alter local hydrological patterns, cause profound ecological changes, and intensify the emission of greenhouse gases from the uppermost permafrost [3]. Since the active layer is closely related to climate change and natural as well as human activities, monitoring and modeling the spatial and temporal variations in active layer thickness (ALT) have become important. The goal of this study is to provide methodology to model and estimate spatial-temporal variations in ALT at the site U1 of the Circumpolar Active Layer Monitoring network and to demonstrate its use in spatial-temporal interpolation. Our model will provide insight into the large/small scale relative contributions of spatial and temporal variations along with their interaction. Such findings will be useful for optimizing spatial-temporal sampling and monitoring designs. Moreover, our efforts will contribute to understanding how ALT varies in space and time.

## 1.2 Description of data

Our study focuses on the Circumpolar Active Layer Monitoring (CALM) site U1 in Barrow, Alaska. The environment of this site is shown in Figure 1. The raw ALT data were collected over 121 locations with different vegetation, and soil and permafrost conditions. In order to have a group of homogeneous locations with similar landforms, we deleted the ALT data collected from rivers, ponds, and fox mound edges where the landforms are quite different from the other 105 locations. For the site U1 in Barrow, one square kilometer grid consists of a square array of surveyed permanent stakes separated by 100 meter (m), yielding an array of sampling nodes on each grid; 105 locations were considered with the local coordinates of the grid, ranging from 0 to 1000 m in the east and north directions. For each station, the active layer data was recorded yearly from 1995 to 2012 and measured by centimeter (cm). In order to visualize the spatial variability of ALT in the site U1, we calculated the average ALT for 18 years for each station and constructed 3D and contour plots. Figure 2 and 3 show that the average ALT values for 18 years for each station are not uniformly distributed. The area between approximately 200–300 m (E), 700–800 m (N) has relatively thicker active layer. The spatial heterogeneity at Barrow’s site U1 indicates that the spatial dependency pattern is nontrivial.



Figure 1: The environment of CALM site U1 in Barrow, Alaska. Image used with permission [4]

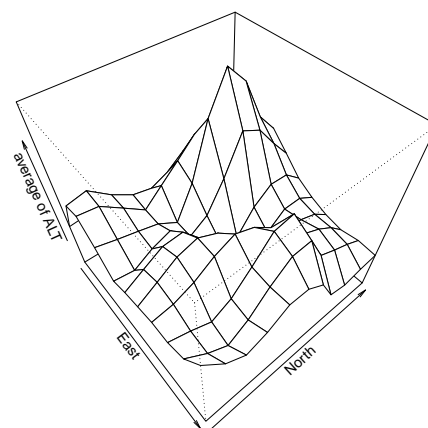


Figure 2: 3D plot of 18 years average ALT for each of 105 locations

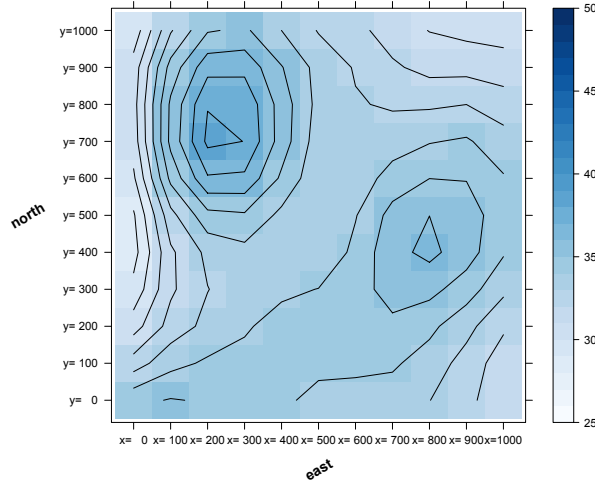


Figure 3: Contour plot of the 18 years average ALT for each of the 105 locations

## 2 Methods

### 2.1 Stationary and Isotropic Processes

A spatial stochastic process is called strictly stationary if the finite dimensional joint distributions are invariant under spatial shifts. It means that for vectors  $\mathbf{h} \in \mathbb{R}^d$  ( $d = 2$  in our study, indicating a surface in two-dimensional space), we get [8]

$$F(y_1, \dots, y_n; \mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_n + \mathbf{h}) = F(y_1, \dots, y_n; \mathbf{s}_1, \dots, \mathbf{s}_n), \quad (1)$$

where  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subseteq \mathbb{R}^d$  is the finite set of spatial locations,  $\{y_1, \dots, y_n\} \subseteq \mathbb{R}$ ,  $F(y_1, \dots, y_n; \mathbf{s}_1, \dots, \mathbf{s}_n) = \mathbb{P}(z(\mathbf{s}_1) \leq y_1, \dots, z(\mathbf{s}_n) \leq y_n)$  is the collection of the finite-dimensional joint distributions of the random vector  $(z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))^T$ .

In the case of Gaussian process  $z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d$ , where the finite dimensional distributions are determined by their second-order properties, we have [8]

$$\mathbb{E}(z(\mathbf{s})) = \mathbb{E}(z(\mathbf{s} + \mathbf{h})) \quad (2)$$

and

$$\text{cov}(z(\mathbf{s}), z(\mathbf{s} + \mathbf{h})) = \text{cov}(z(\mathbf{0}), z(\mathbf{h})) = C(\mathbf{h}), \quad (3)$$

where  $C(\mathbf{h}), \mathbf{h} \in \mathbb{R}^d$  is the covariance function.

For a random field on  $\mathbb{R}^d$ , isotropy can be considered as invariance under rotations and reflections. It is an important assumption for covariance structure, in which the covariance function between any two values only depends on the Euclidean distance between their corresponding locations [17]:

$$C(\mathbf{h}) = C(\|\mathbf{h}\|), \quad (4)$$

where  $\|\cdot\|$  indicates the Euclidean distance. In **Section 3.1**, we discuss why spatial stationarity and isotropy are good working assumptions in our data.

### 2.2 Spatial Variogram Modeling

The semivariogram is a method for providing a concise description of the extent of spatial or temporal variation. For a set of  $n$  observations, if we choose any pair of data values,  $z(\mathbf{s}_i)$  and  $z(\mathbf{s}_j)$ , the dissimilarity of the data pair can be calculated as half the square of the difference between two observed values [16]:

$$\gamma_{ij} = \frac{1}{2} \times [z(\mathbf{s}_i) - z(\mathbf{s}_j)]^2, \quad (5)$$

where  $\gamma_{ij}$  = the semivariance between  $\mathbf{s}_i$  and  $\mathbf{s}_j$ ,  $z(\mathbf{s})$ = attribute value, e.g.  $z(\mathbf{s})$  can be the residual of ALT values after removing the trend in the given year.

The semivariogram cloud is obtained using Formula (5), by plotting the difference  $\gamma_{ij}$  against the separation distance. However, the semivariogram cloud is diffuse and suffers from pointwise instability [5]. Hence, the process of binning, by which the semivariance data are averaged by distance and direction, is introduced. First, group pairs of sample points into lag classes. If lag ( $\mathbf{h}$ ) size is 500 m, the lag classes are: ( $\leq 500$ ,  $500 \sim 1000$ ,  $1000 \sim 1500$ , etc). Next, pairs are grouped by direction (e.g. using grid cells, where each grid cell will have pairs with similar direction and distance). Thus, we obtain the more stable semivariogram, denoted  $\gamma(\mathbf{h})$  [14]:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \times \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h})]^2, \quad (6)$$

where  $\gamma(\mathbf{h})$ = average semivariance between sample points separated by lag  $\mathbf{h}$ , and  $N(\mathbf{h})$  = the number of pairs sample points sorted by distance  $\mathbf{h}$ .

After binning, if the evidence of spatial autocorrelation exists, semivariance is expected to increase with distance. However, if the empirical variogram continues climbing steadily, it often indicates a significant spatial trend in the variable, resulting in a negative correlation between variable values separated by large lags.

In **R**, the package *gstat* contains *variogram* function to construct an empirical semivariogram, given each year's ALT value and the coordinates of locations. On the basis of the empirical semivariogram, we can establish the pattern of autocorrelation. For example, various properties of the autocorrelation such as range and sill can be inferred on the basis of the semivariogram. For the subsequent inference, the parametric model should fit to the empirical semivariogram. Some of the commonly used semivariogram models are powered exponential model, Gaussian model, spherical model and Matérn model. The package *gstat* includes semivariogram modeling function *vgm*, which can choose one of parametric models to fit the empirical semivariogram. In our study, we applied Matérn and powered exponential models to fit the empirical variograms for each of spatial models for the data for 18 years.

The Matérn isotropic correlation function is defined by Handcock and Stein (1993) [11] as follows:

$$K_{\theta}(\|\mathbf{h}\|) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{\|\mathbf{h}\|}{\theta_1'}\right)^{\theta_2} \mathcal{K}_{\theta_2}\left(\frac{\|\mathbf{h}\|}{\theta_1'}\right), \quad (7)$$

where  $\mathbf{h}$  is the separation lag;  $\theta_1' = \theta_1/2\sqrt{\theta_2}$ ;  $\mathcal{K}_{\theta_2}$  is a modified Bessel function of the second kind of order  $\theta_2$ ;  $\Gamma$  is the gamma function,  $\theta_1 > 0$  is a scale parameter controlling the range of correlation; and  $\theta_2 > 0$  is the smoothness parameter controlling the smoothness of random field.

Sometimes, the variation in the data is partly attributable to measurement error:

$$z(\mathbf{s}_i) = \xi(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad (8)$$

where  $\xi(\mathbf{s}_i)$  is the measurement error at location vectors  $\mathbf{s}_i$ ;  $\epsilon(\mathbf{s}_i)$  represents the remains of the residual at  $\mathbf{s}$  after subtracting the measurement error. The covariance for  $z(\mathbf{s}_i)$  is then given by:

$$C(\mathbf{h}) = \begin{cases} c_0 + K(\mathbf{h}) & \text{if } \mathbf{h} = 0 \\ K(\mathbf{h}) & \text{if } \|\mathbf{h}\| > 0, \end{cases} \quad (9)$$

where  $K$  is the covariance of  $\epsilon(\mathbf{s}_i)$ ;  $c_0$  is the nugget effect, which was used to interpret micro-scale variations causing a discontinuity at the origin in the semivariogram.

Hence, the Matérn semivariance with nugget effect is shown below [15]:

$$\gamma(\mathbf{h}) = c_0 + c_1 \left[ 1 - \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{\|\mathbf{h}\|}{\theta_1'}\right)^{\theta_2} \mathcal{K}_{\theta_2}\left(\frac{\|\mathbf{h}\|}{\theta_1'}\right) \right], \quad (10)$$

where  $c_0$  is the nugget variance, and  $c_0 + c_1$  is the sill variance.

The powered exponential model is given below:

$$\gamma(\mathbf{h}) = c_0 + c_1 \left[ 1 - \exp \left[ - \left( \frac{\|\mathbf{h}\|}{\phi} \right)^{\omega} \right] \right], \text{ where } 0 < \omega \leq 2 \quad \phi > 0 \quad (11)$$

The Matérn model shows better spatial kriging performance than the powered exponential model owing to the smoothness parameter ( $\theta_2$ ) of the former model; the higher the value of  $\theta_2$ , the smoother is the process [15]. Thus, the Matérn model can be used as a general model for ALT spatial variation. This model will be also useful for modeling the local variogram automatically for spatial kriging. We fitted two models to the data and described the results in **Section 3.3**.

### 2.3 Spatial Kriging

Kriging is a generalized least-square regression technique that allows one to account for the spatial dependence between observations, as revealed by the semivariogram, into spatial prediction [10]. This technique is widely used in geology, hydrology, and environmental monitoring to interpolate spatial data. Most of geostatistics is relied on the concept of a random function, whereby the set of unknown values is regarded as a set of spatially dependent random variables. Thus, each measurement  $z(\mathbf{s}_\alpha)$  is interpreted as a particular realization of a random variable  $Z(\mathbf{s}_\alpha)$ , where  $\mathbf{s}$ ,  $\mathbf{s}_\alpha$ = location vectors for estimation point and one of the neighboring data points, indexed by  $\alpha$ . Geostatistical interpolation amounts at estimating the unknown thickness of active layer  $z$  at the location  $\mathbf{s}$  as a linear combination of neighboring observations [2]:

$$Z^*(\mathbf{s}) = m(\mathbf{s}) + \sum_{\alpha=1}^{n(\mathbf{s})} \lambda_\alpha [Z(\mathbf{s}_\alpha) - m(\mathbf{s}_\alpha)], \quad (12)$$

where  $n(\mathbf{s})$ = number of data points used for estimation of  $Z^*(\mathbf{s})$ ;  $m(\mathbf{s})$ ,  $m(\mathbf{s}_\alpha)$ = expected values of  $Z(\mathbf{s})$  and  $Z(\mathbf{s}_\alpha)$ ;  $\lambda_\alpha$  = kriging weight assigned to datum  $Z(\mathbf{s}_\alpha)$  for estimation location  $\mathbf{s}$ ; the same datum will receive different weights for different estimation locations.

In our study, we implement one of kriging variants, ordinary kriging (OK) for optimal prediction of spatial processes. In the OK method, we assume that the mean  $m(\mathbf{s})$  is constant for each nearby value; that is  $m(\mathbf{s}_\alpha) = m(\mathbf{s})$ . Thus, the OK estimator  $Z_{OK}^*(\mathbf{s})$  is defined as [10]:

$$Z_{OK}^*(\mathbf{s}) = \sum_{\alpha=1}^{n(\mathbf{s})} \lambda_\alpha^{OK}(\mathbf{s}) Z(\mathbf{s}_\alpha), \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{s})} \lambda_\alpha^{OK}(\mathbf{s}) = 1, \quad (13)$$

The ordinary kriging weights  $\lambda_\alpha^{OK}(\mathbf{s})$  are determined such as to minimize the estimation variance,  $Var\{Z_{OK}^*(\mathbf{s}) - Z(\mathbf{s})\}$ , while ensuring the unbiasedness of the estimator,  $E\{Z_{OK}^*(\mathbf{s}) - Z(\mathbf{s})\} = 0$ . These weights are obtained by solving a system of linear equations which is known as ordinary kriging system [10]:

$$\begin{aligned} \sum_{\beta=1}^{n(\mathbf{s})} \lambda_\beta(\mathbf{s}) \gamma(\mathbf{s}_\alpha - \mathbf{s}_\beta) - \mu(\mathbf{s}) &= \gamma(\mathbf{s}_\alpha - \mathbf{s}), \quad \alpha = 1, \dots, n(\mathbf{s}); \\ \sum_{\beta=1}^{n(\mathbf{s})} \lambda_\beta(\mathbf{s}) &= 1, \end{aligned} \quad (14)$$

where  $\mu(\mathbf{s})$  is the Lagrange parameter accounting for the constraint on the weights. The only information required by the kriging system (14) are semivariogram values for different lags, and these are readily derived once a semivariogram model has been fitted to empirical values. **Section 2.2** shows two different types of permissible models that are combined with a nugget-effect model for the fitting of the empirical semivariogram of ALT.

In order to demonstrate whether the Matérn semivariogram model provides better prediction performance than powered exponential model, we do the following steps: first, randomly remove 10 known points from 105 locations; second, use remaining 95 points to estimate the ALT values at points removed; third, compare two models' estimated with true values by calculating the root mean square error (RMSE) and mean absolute error (MAE):

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (Z_{i,act} - Z_{i,OK.est})^2} \quad (15)$$

and

$$MAE = \frac{1}{n} \sum_{i=1}^n |Z_{i,act} - Z_{i,est}|, \quad (16)$$

where  $Z_{i,OK.est}$  are the estimated ALT values for removed locations;  $(Z_{i,act})$  are the true values of ALT for those locations.

Smaller values of RMSE and MAE will indicate the better performance of spatial prediction.

## 2.4 Space-Time Kriging

In the section, we discuss the space-time kriging. First, let  $\{z(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}\}$  be a real valued Gaussian space-time random field; this is based on the assumption that  $Var[z(\mathbf{s}, t)]$  is finite, for all  $\mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$ . The variable  $z$  can be observed at the spatial-temporal points such that  $\{[Z(\mathbf{s}_{1i}, t_i), \dots, Z(\mathbf{s}_{ni}, t_i)] : i = 1, \dots, m \in \mathbb{N}\}$ . Here,  $\mathbf{s}_{1i}, \dots, \mathbf{s}_{ni}$  are the  $n$  data locations at time  $i$ , and  $t_1 < t_2 < \dots < t_m$  are the times of observation. In our study,  $z$  is the value of ALT,  $n = 105(\text{locations})$  and  $m = 18(\text{years})$ . Thus, we have  $105 \times 18 = 1890$  spatio-temporal points of ALT observations. In general, the goal is to use the data to predict unobserved  $z(\mathbf{s}_0, t_0)$  at the points  $(\mathbf{s}_0, t_0)$  where  $\mathbf{s}_0 \in \mathbf{S}$  and  $t_0 \in \mathbf{T}$  ( $\mathbf{S}$  is the spatial domain and  $\mathbf{T}$  is the temporal domain), typically  $t_m \leq t_0$ . In our case,  $t_0 = 19$  (19th year: Year 2013).

Before we apply the ordinary space-time kriging methodology to our data, we have to perform the following steps: First step of our data analysis results in the deterministic trend that varies in the time only and has an interpretable parametric form; in second step we fit two families of parametric space-time covariance models. The importance of a covariance function in space-time kriging was shown by Gneiting, Genton and Guttorp (2007) [7]. In our study, we first assumed that the covariance structure is isotropic in space and time such that

$$cov\{Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)\} = C(\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|), \quad (17)$$

where  $C$  is the space-time covariance function, with  $\mathbf{s}_i, \mathbf{s}_j \in \mathbb{R}^d$  and  $t_i, t_j \in \mathbb{R}$ . Thereafter, we constructed isotropic separable and nonseparable covariance models to evaluate the space-time kriging. In the former model, we assume that the space-time covariance is a simply scaled product of the purely spatial and purely temporal covariances  $C_{space}$  and  $C_{time}$  such that

$$cov\{Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)\} \propto C_{space}(\|\mathbf{s}_i - \mathbf{s}_j\|) \times C_{time}(|t_i - t_j|), \quad (18)$$

where  $\mathbf{s}_i, \mathbf{s}_j \in \mathbb{R}^d$  and  $t_i, t_j \in \mathbb{R}$ .

We use Matérn model introduced in Section 2.2 Equation (7) and express the spatial isotropic covariance function as follows:

$$C_{space}(\mathbf{h}) = \sigma^2 K_\theta(\|\mathbf{h}\|) = \sigma^2 \frac{1}{2^{\theta_2-1} \Gamma(\theta_2)} \left( \frac{\|\mathbf{h}\|}{\theta_1'} \right)^{\theta_2} \mathcal{K}_{\theta_2} \left( \frac{\|\mathbf{h}\|}{\theta_1'} \right), \quad (19)$$

where  $\mathbf{h} \in \mathbb{R}^d$  is the space lag and  $\sigma \geq 0$ . Now, in order to parameterize the temporal covariance function, we use the correlation function of Cauchy class [9]:

$$c(x) = (1 + (\theta x)^\phi)^{-v}, \text{ where } x > 0, \phi \in (0, 2], v > 0, \theta > 0 \quad (20)$$

Gneiting and Schlather (2004) stated that the Cauchy class provides flexible power-law correlations and generalizes the stochastic models commonly dealt with and synthesized in geostatistics and time series analysis [9]. Therefore, in our study, we chose Cauchy class to construct the isotropic model of temporal covariance function:

$$C_{time}(\mu) = \sigma^2 c(|\mu|) = \sigma^2 (1 + (\theta|\mu|)^\phi)^{-v}. \quad (21)$$

Further, we consider the realistic scenario that space-time interaction exists in the case of nonseparable functions. The general nonseparable covariance function was defined by Gneiting (2001) [6] as

$$C(\mathbf{h}, \mu) = \frac{\sigma^2}{\psi(|\mu|^2)^{d/2}} f \left( \frac{\|\mathbf{h}\|^2}{\psi(|\mu|^2)} \right), \quad \mathbf{h} \in \mathbb{R}^d, \mu \in \mathbb{R}, \quad (22)$$

where  $f(x)$  is a completely monotone function on  $[0, \infty)$ , and  $\psi(x)$  is positive with a completely monotone derivative on  $[0, \infty)$ , and  $\sigma^2 > 0$ .

The model applied in our study is a combination of Cauchy and powered exponential covariance functions [6]:

$$C(\mathbf{h}, \mu) = \frac{\sigma^2}{(a|\mu|^{2\alpha} + 1)^{\beta d/2}} \exp\left(-\frac{c\|\mathbf{h}\|^{2\kappa}}{(a|\mu|^{2\alpha} + 1)^{\beta\kappa}}\right), \quad (23)$$

where  $a, c > 0$  are scale parameters;  $\alpha, \kappa \in [0, 1]$  are temporal and spatial smoothing parameters, respectively;  $\beta \in [0, 1]$  is a space-time interaction parameter; and  $\sigma^2$  is the variance of the random field. This model is referred as the "Gneiting Model", which is widely used by fixing some parameters and has been applied in the package of *CompRandFld* in **R**.

Next, we determine the kriging of unobserved  $z(\mathbf{s}_0, t_0)$  where  $z$  is assumed to be a realization of random function  $Z(\mathbf{s}, t)$  which holds a full statistical model, including its space-time dependence structure [12]. The spatial-temporal prediction of  $Z(\mathbf{s}_0, t_0)$  can be described as a global trend — the mean function  $m(\mathbf{s}, t)$  and a high frequency autocorrelated residual component obtained from a weighted linear combination of the residuals:

$$Z^*(\mathbf{s}_0, t_0) = m(\mathbf{s}_0, t_0) + \sum_{i=1}^n \lambda(\mathbf{s}_i, t_i) (Z(\mathbf{s}_i, t_i) - m(\mathbf{s}_i, t_i)), \quad (24)$$

where  $Z(\mathbf{s}, t)$  is assumed that  $Var(Z(\mathbf{s}, t)) < \infty$  for  $\mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$ , the mean function  $m(\mathbf{s}, t) = E[Z(\mathbf{s}, t)]$  and the covariance structure is stationary.  $\lambda(\mathbf{s}_i, t_i)$  is the weight of the individual observations in the prediction, which is determined by the covariance structure of the space-time process  $Z(\mathbf{s}, t)$  [6].

In **R**, the package of *CompRandFld* includes: the separable model with different combinations of functions and the non-separable model called "Gneiting Model", which were used in our study to fit the marginal spatial and temporal empirical variogram, respectively. We then accessed to the parameters of these models to compute the kriging and calculated the RMSE and MAE to compare the prediction performance of the models.

## 3 Results

### 3.1 The Spatial-Temporal Trends

The average ALTs for 18 years do not exhibit a trivial pattern in either directions, as shown in Figure 4. For instance, in North direction, the average ALTs for 18 years is concentrated in the range 30—40 cm. Since we do not see apparent and systematic patterns, we can assume that the variability is stochastic and that stationarity and isotropy are appropriately working assumptions.

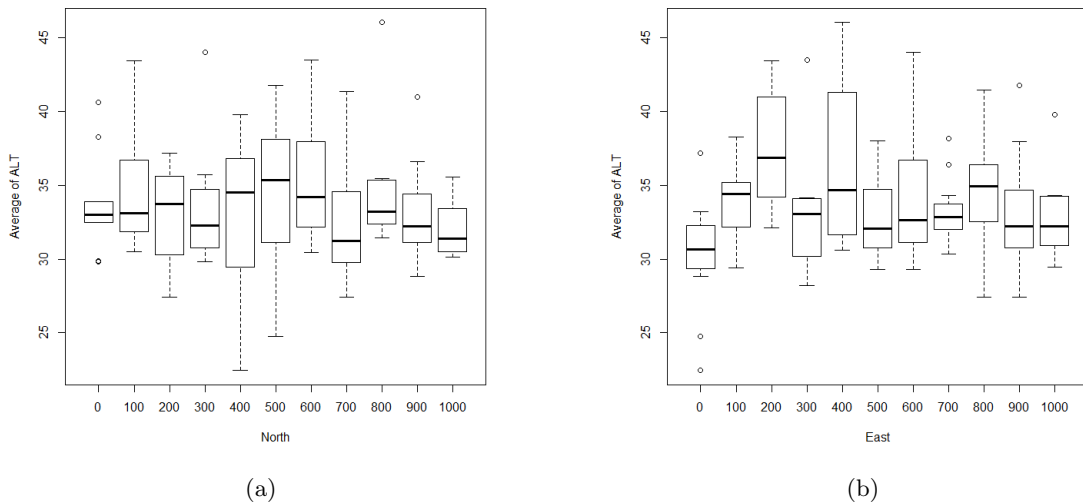


Figure 4: (a): Distribution of 18 years average ALTs for each of 105 locations along North direction, and (b): Distribution of 18 years average ALTs for each of 105 locations along East direction

**Appendix** contains the contour plots of ALT at 105 locations in 18 consecutive years from 1995 to 2012. These graphs show some consistent patterns as well as the evidence of spatial autocorrelation in ALT at 105 locations. If we compare the spatial variations at different locations in any given year, we observe that although ALT displays some spatial similarities, the similar features are not well pronounced; hence, we assume they are stochastic. However, if we compare the temporal variations in any given spatial region, we observe that the similar pattern. The obvious temporal pattern of ALT can be explained by the known climatic pattern in the Arctic area [18].

Therefore, we are interested in fitting the systematic temporal trend of ALT. Figure 5 compares the time series for southwestern corner\* (gray lines) of CALM site U1 in Barrow with those for northeastern corner† (red lines). As we can see, four neighboring locations in the same corner demonstrate more similar temporal trend of ALT variations than the locations in the other corner.

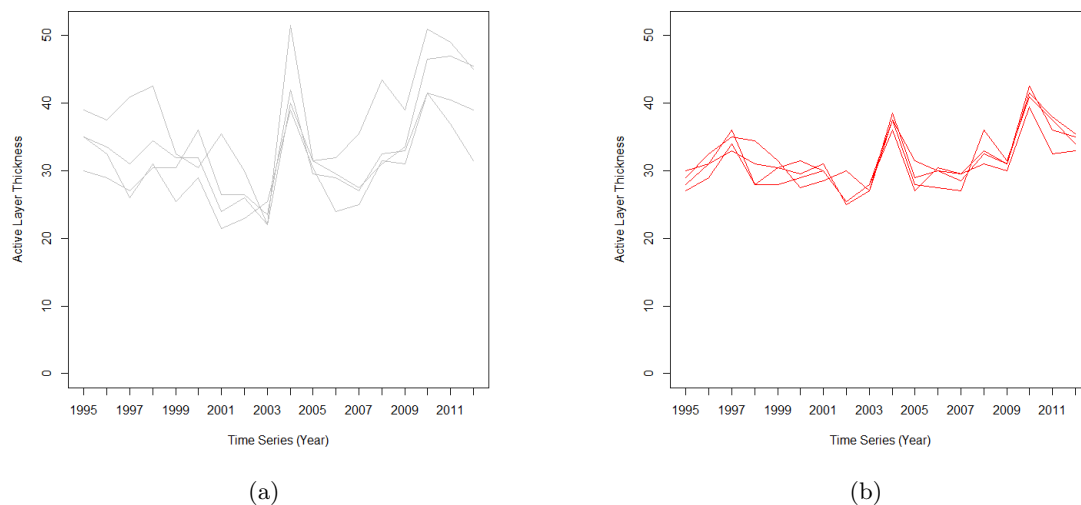


Figure 5: Time series for southwestern corner (a) and northeastern corner (b) of CALM site U1 in Barrow

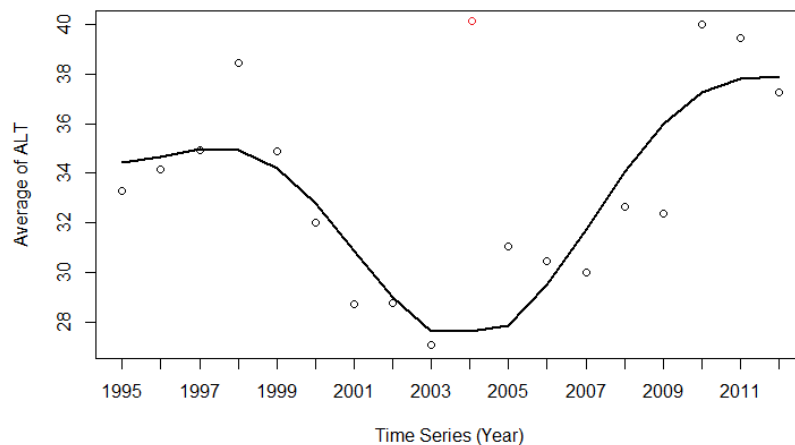


Figure 6: 105 locations' average ALTs for each of the 18 years (1995-2012) along with fitted model for time trend (excluding year of 2004)

From Figure 6, we can see the periodic pattern in the average ALT if an abnormally high value for

\*southwestern corner: four neighboring locations with local coordinates of the grid (0,0), (0,100),(100,0), (100,100)

†northeastern corner: four neighboring locations with local coordinates of the grid (1000,900), (1000,1000), (900,900), (900, 1000)



2004 (the red dot) is excluded. Hence, we considered including Fourier frequency  $\cos(\omega * year)$  to fit the time trend. We observed that average ALT has two peaks, which are in 1998 and 2010. This implies that the period is 11 ~ 12 years. We assume that a linear multiple regression model including  $\cos(\omega * year)$ ,  $year$  and  $year^2$  should be appropriate to describe time trend:

$$AverageALT = \beta_0 + \beta_1 \cos(2\pi/\omega * (year + \beta_2)) + \beta_3 year + \beta_4 year^2, \quad (25)$$

where  $year$  is the record years minus 1994, excluding 2004;  $year^2 = year \times year$  denotes a quadratic term of  $year$ , and  $\omega$  means the period of cosine term, which is either 11 or 12. Using the nonlinear regression function  $nls$  in **R**, we obtained the estimated parameters in Listing 1:

Listing 1: R Output of Estimated Parameters for time trend (remove 2004)

---

```

Parameters :
      Estimate Std. Error t value Pr(>|t|)
beta0  37.27146   1.93709  19.241 2.19e-10 ***
beta1   2.87983   0.92386   3.117 0.00890 **
beta2   6.41218   0.46134  13.899 9.26e-09 ***
beta3  -1.62333   0.46832  -3.466 0.00466 **
beta4   0.09046   0.02398   3.773 0.00266 **
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.175 on 12 degrees of freedom

```

---

Thus, the regression model of time trend is as follows:

$$AverageALT = 37.2715 + 2.8798 \times \cos\left(\frac{2\pi}{11}(year + 6.4122)\right) - 1.6233 \times year + 0.0905 \times year^2. \quad (26)$$

Since the estimated parameters are all significant in our model, we used this model to predict the average ALT for 105 locations in each year from 1995 to 2012 and plotted fitted (black) line in Figure 6. Further, based on the model of time trend regression, we computed the average ALT at 105 locations in 2013 by setting  $year=19$  into Equation (26).

Next, the stochastic residual  $R(s, t)$  was calculated by subtracting the fitted average ALT for each year from the 105 values of ALT in corresponding year (except for year of 2004). In this approach, uncertainties in the detrending procedure are not taken into account in any subsequent analysis.

### 3.2 Empirical Variograms and Fitted Model

In this section, we first consider the spatial empirical variograms along with fitted models for each year (except for 2004) using stochastic residuals. We use **R** package *gstat* and its default weighted least squares method to fit the variograms in Figure 7. Then, we plot marginal spatial and temporal variograms using the package of *CompRandFld* and add fitted lines using *geoR* package in Figure 8.

Figure 7 shows spatial empirical variograms by year (dots) and fitted Matérn model in solid line. We can observe that there is significant variability in the empirical variograms and fits. Specifically, *sill*, which is the variance of random field, varies from 20 ~ 50. And nugget effect, which is the measurement error, ranges from 0 to 30. Further, the range of each variogram, which is the distance at which data are no longer autocorrelated, varies between 200 and 600 m.

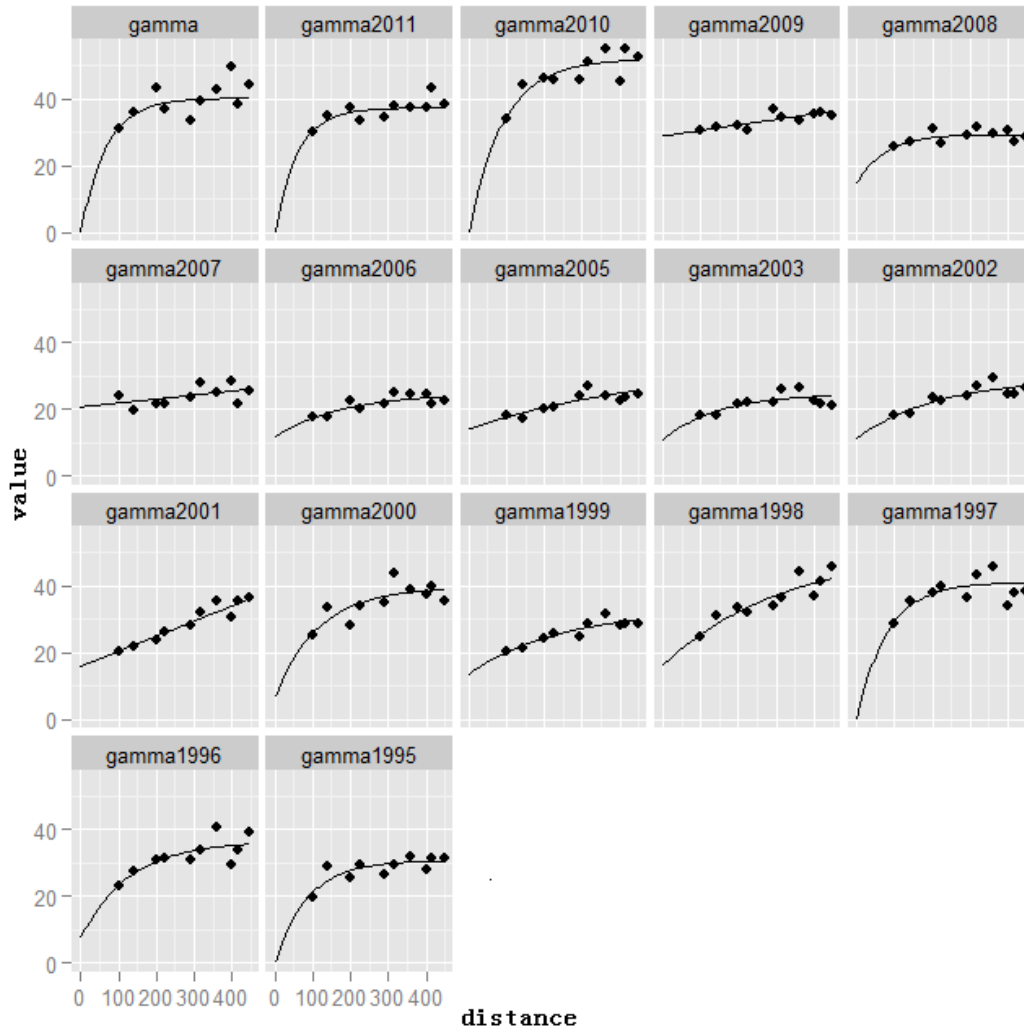


Figure 7: Empirical variograms with fitted variogram lines in 1995-2012 (except 2004): First plot labeled "gamma" represents the  $\gamma(h)$  in 2012

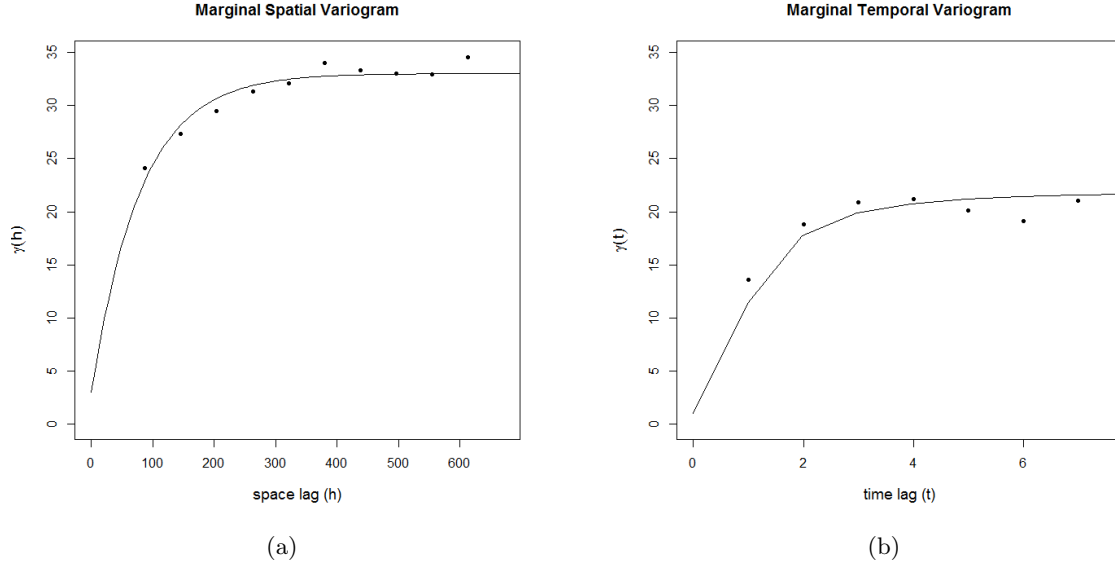


Figure 8: Marginal Empirical Variograms with fitted variogram lines

Table 1: Parameter Estimates for Marginal Spatial and Temporal Fitted Variogram Lines

| Marginal Variogram | Fitted Model | nugget | sill | range | smoothness parameter | minimized weighted sum of squared residuals |
|--------------------|--------------|--------|------|-------|----------------------|---|
| Space              | Matérn       | 3      | 30.4 | 238.8 | 0.5                  | 2378  |
| Time               | Cauchy       | 2      | 19.9 | 3.2   | 1                    | 2552  |

Figure 8 (a) and (b) shows the marginal empirical spatial and temporal variograms with their fitted variogram lines, respectively. Table 1 shows the estimated parameters of fitted variogram lines. Note that, we use Matérn model to fit marginal spatial variogram with smoothness parameter =0.5 ( $\theta_2$  in Eq.19). *sill* ( $\sigma^2$  in Eq.19) is approximately 30, *nugget* is about 3, and the autocorrelation goes to zero beyond the range of 239 meter. In the temporal part, we apply Cauchy model to fit marginal temporal variogram with  $\phi = 2$  (Eq.21) and smoothness parameter =1 ( $v$  in Eq.21 ). *sill* ( $\sigma^2$  in Eq.21) is approximately 20, which is lower than *sill* in marginal space, and the range is around 3 at which the semivariogram reaches *sill* value, indicating that temporal correlation no longer exists after time lag = 3. In addition, the nugget effect for marginal spatial variogram is higher than that for the temporal variogram, which is around 2.

### 3.3 Spatial Kriging

Here, we pick ALT data from the year of 2012 and compare predictive power of ordinary kriging in the Matérn model and the powered exponential model. Figure 9 shows 10 randomly removed locations and the color of the dots represents the prediction values for those locations using Matérn model.

Table 2: Performance of Spatial Prediction in Matérn Model and Powered exponential Model

| Model                 | RMSE   | MAE    |
|-----------------------|--------|--------|
| Matérn:               | 5.4018 | 4.3891 |
| Powered exponential : | 5.4021 | 4.3911 |

Table 2 shows the RMSE and MAE values between true values of ALT and the predicted values using the Matérn model and powered exponential model. We can clearly observe that RMSE and MAE for Matérn model is slightly less than those in the powered exponential model. Therefore, Matérn model seems a better option for spatial kriging.

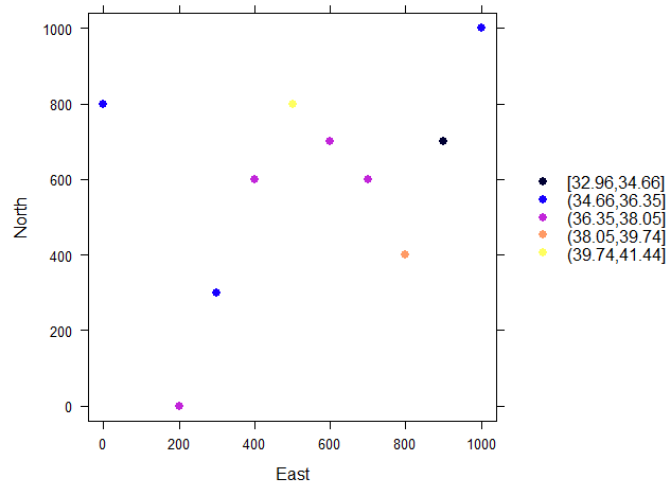


Figure 9: Spatial Kriging for 10 removed values of ALT in 2012 under Matérn model

### 3.4 Space-time Kriging

For Space-Time Kriging, we fitted two covariance models: the separable Matérn-Cauchy model and nonseparable model (called "Gneiting" model in **R** package). We used the function *FitComposite* in the package *CompRandFld* in **R** to get the fitted parameters of space-time covariance functions. We used Marginal likelihood, with time lag of 7 and space lag of 700 in both models. Figure 10 shows the Matérn-Cauchy model fit. 3D space-time fitted variogram is in the top right corner and the empirical space-time variogram is in the top left corner. In the lower-left corner of the figure, the empirical marginal temporal variogram (dots) with fitted variogram line is plotted. We can see the fit is good and the line is close to the dots for the range  $\leq 3$ . In the lower-right corner of the figure, the marginal spatial variogram (dots) with its fitted line is plotted. Here, the fit is not as good as the temporal one, and it seems that the spatial correlation is slightly underestimated. Overall, the Matérn-Cauchy fit is applicable.

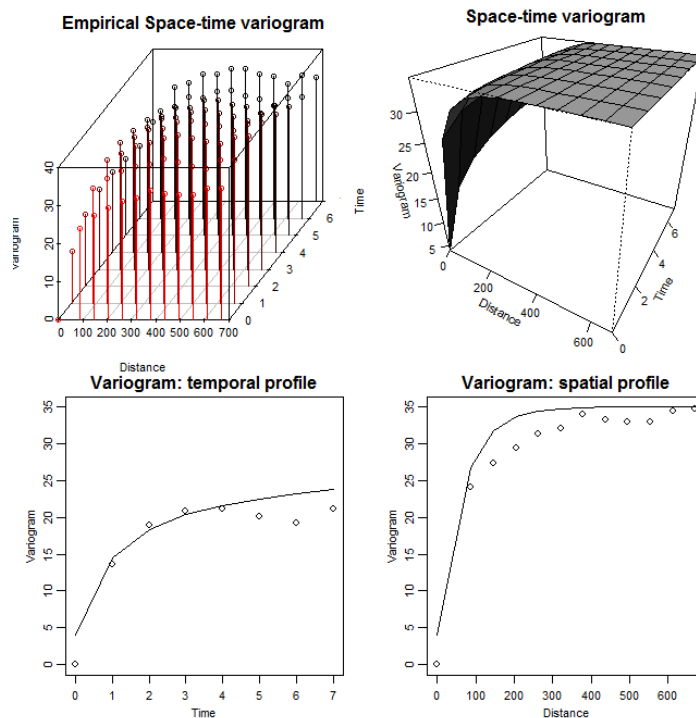


Figure 10: Matérn-Cauchy Separable Model

Figure 11 shows 3D space-time fitted variogram: (top right corner) as well as the empirical space-time variogram for the Gneiting model. The temporal and spatial marginal variograms are shown in the lower-left and the lower-right corners, respectively. The fits in the marginal variograms looks worse than the one for Matérn-Cauchy separable model. Both fits are underestimated the spatial and temporal correlation.

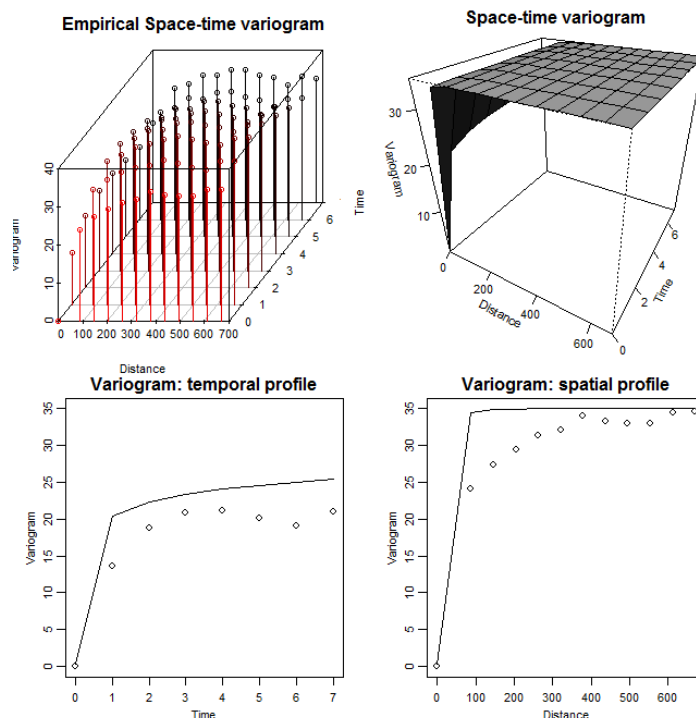


Figure 11: Gneiting Model

Table 3 shows the fitted parameters of the space-time covariance models (separable: Matérn-Cauchy and non-separable: Gneiting). Note that, for the Matérn-Cauchy, the parameters *power time* ( $\alpha$  in Eq.23), *power space* ( $\kappa$  in Eq.23) and *sep* ( $\beta$  in Eq.23), which are exclusive for the Gneiting model, do not exist. Similarly, the parameters *smooth space* ( $\theta_2$  in Eq.19) and *power2* ( $\phi$  in Eq.21), which are specific to the Matérn function and Cauchy function, respectively. We found that the separable parameter *sep* = 0.9145 in the nonseparable Gneiting model, which indicates strong interaction between space and time.

Table 3: Parameter Estimates of the Space-Time Covariance Models

| Model          | <i>nugget</i> | <i>sill</i> | <i>scale time</i> | <i>scale space</i> | <i>power2</i> | <i>smooth space</i> | <i>power time</i> | <i>power space</i> | <i>sep</i> |
|----------------|---------------|-------------|-------------------|--------------------|---------------|---------------------|-------------------|--------------------|------------|
| Matérn-Cauchy: | 0.01798       | 33.3739     | 0.3937            | 189.513            | 0.2559        | 0.0988              | 1                 | None               | 0          |
| Gneiting:      | 0.3145        | 34.6974     | 0.36              | 6.929              | 1             | 1                   | 0.3226            | 0.5525             | 0.9145     |

Next, we used both covariance models in space-time kriging for values of ALT in 2013. We compared our ALT predictions for 2013 with the true ALT values to examine the performance of space-time kriging.

Table 4 shows the summary statistics to compare the separable Matérn-Cauchy model, non-separable Gneiting model and Time Trend Regression (Eq. (26)). The RMSE and MAE values indicate that two space-time kriging models have better prediction performance for ALT in 2013 than the Time Trend Regression model. For instance, the RMSE of the Matérn-Cauchy model is less than that of Time Trend Regression by 26.75%, which is a significant improvement in prediction. If we compare the performances of Matérn-Cauchy model and Gneiting model, we observe that the RMSE and MAE values are very close. However, the former Matérn-Cauchy model yields a predicted average ALT closer to the true ALT in 2013, 38.38095.

Table 4: Models' Performance of Prediction for ALT in 2013

| Model                 | RMSE   | MAE    | Predicted Average ALT in 2013 |
|-----------------------|--------|--------|-------------------------------|
| Matérn-Cauchy:        | 5.1191 | 3.6864 | 38.1904                       |
| Gneiting:             | 5.0795 | 3.6974 | 37.9971                       |
| Time Trend Regression | 6.988  | 5.5264 | 38.0208                       |

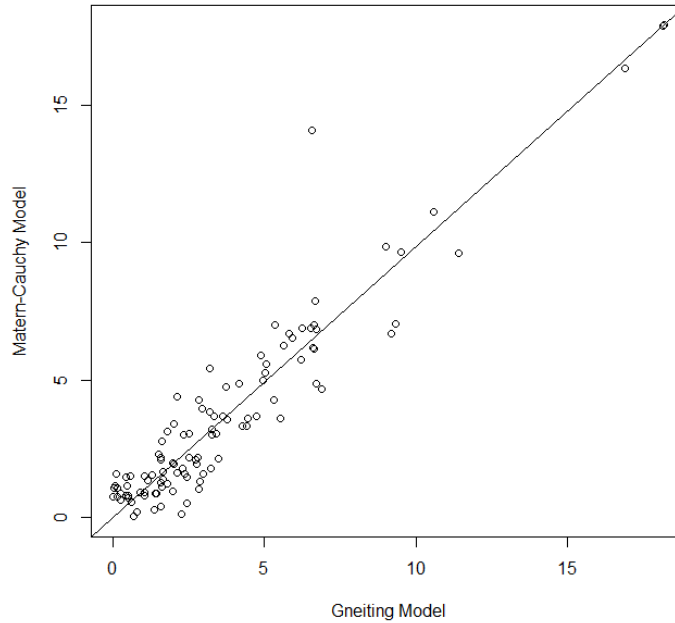


Figure 12: Absolute values of errors in two models

Figure 12 shows that the absolute values of residuals between the true values of 2013 ALT and predicted ALTs in two models are very similar, except for one point. Overall, space-time kriging under Matérn-Cauchy and Gneiting models both show significant improvements in the prediction of ALT in 2013, compared with the naive model, which does not take the spatial and temporal correlation in residuals into consideration.

## 4 Conclusions

The purpose of this study was two-fold. First, we described the methodology for modeling and estimation of spatiotemporal variation in ALT. Second, we demonstrated its use in the spatiotemporal interpolation. Specifically, we considered 19 years of data (years 1995–2013) collected at the site named Barrow (Lat. 71 deg. 19 min. N, Location Lon. 156 deg. 35 min. W). The data are on 11 by 11 grid, however for 16 grid locations, the observations were either missing for most of the years or exhibited systematically higher values due to proximity to water. Hence, for our analysis, we used ALT collected at 105 grid locations in years 1995–2012. We used year 2013 to demonstrate the validity and predictive power of our methodology. In our data analysis we first estimated the parametric trend and then examined the residuals for presence of spatial and temporal dependencies. The empirical marginal temporal and marginal spatial variograms indicated the presence of correlation, with the spatial range being approximately 500 meters and the temporal one is around 4 years. We used **R** package called *CompRandFld* to fit the parametric space-time covariance model to the residuals. Specifically, we tried two families of structures, "Matern-Cauchy" and "Gneiting". These models differ not only in their functional forms, but also in the way they account for interaction between spatial and temporal

components. "Matérn-Cauchy" is a separable model with Matérn and Cauchy spatial and temporal marginal covariance respectively. While the model called "Gneiting" is a nonseparable one with powered exponential and Cauchy spatial and temporal marginal covariance respectively.

The major limitation of our approach, however, is that we could not fit a nonseparable version of "Matérn-Cauchy" covariance structure because of the lack of readily available software to model it. Since marginally Matérn provides a better fit compared to powered exponential, we argue that a nonseparable "Matérn-Cauchy" should provide a better fit than "Gneiting" or powered exponential-Cauchy. We plan to investigate this in our next study. Nevertheless, the use of covariance structures for residual terms significantly improves predictive ability of the model. We compared our models to the naive one, which does not take the spatial and temporal correlation in residuals into consideration. The root mean squared error is reduced by 27 percent when our approach was taken. Even though our research focuses on one observation site—the U1 Barrow site, our methodology can be applied to other sites in the Arctic region.

The Circumpolar Active Layer Monitoring (CALM) represents the only coordinated program that provides data collection, standardization, open access, and dissemination services for active layer data world-wide. At present, the CALM database includes long-term active layer, soil temperature and moisture content (where available), and heave/subsidence data from over 260 sites distributed throughout Circumpolar Arctic. Despite the success of the CALM program at providing observations to the broader scientific community, greater potential exists in conducting comprehensive statistical analyses on the large amounts of data accumulated by the CALM program over its 20 year history. Primarily results from the application of the spatial interpolation technique conducted for CALM site U1 in Barrow, Alaska showed that the proposed technique is an effective tool for the estimation of active layer spatial variability. By applying the proposed methodology, we confirmed that the resulting fits resemble empirical patterns better than the naive alternatives. This shows that preliminary application of the model is very promising.

## 5 Acknowledgments

I would like to thank Dr. Nikolay Shiklomanov and Dr. Dmitry Streletskiy from the Department of Geography at The George Washington University, for collecting ALT data from 1995 to 2013 and providing relevant geographical knowledge for this study. A great debt of gratitude is owed to my advisor Dr. Tatiyana Apanasovich. Her encouragement, advising and patience have gone beyond the call of duty as mentor and friend. Thanks Dr. Apanasovich from the bottom of my heart.

## References

- [1] "Spatial Distribution of Changes in Temperature, Circulation and Related Variables - AR4 WGI Technical Summary," 2007. [Online]. Available: [http://www.ipcc.ch/publications\\_and\\_data/ar4/wg1/en/tssts-3-1-2.html](http://www.ipcc.ch/publications_and_data/ar4/wg1/en/tssts-3-1-2.html)
- [2] G. Bohling, "Kriging," pp. 1–20, 2005. [Online]. Available: <http://people.ku.edu/~gbohling/cpe940/Kriging.pdf>
- [3] Circumpolar Active Layer Monitoring, "CALM Program Description." [Online]. Available: <http://www.gwu.edu/~calm/about/program.html>
- [4] Circumpolar Active Layer Monitoring network, "U1 Barrow." [Online]. Available: [http://www.gwu.edu/~calm/data/CALM\\_Data/NorthAmerica/Alaska/NorthSlope/u01\\_barrow\\_grid/](http://www.gwu.edu/~calm/data/CALM_Data/NorthAmerica/Alaska/NorthSlope/u01_barrow_grid/)
- [5] P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger, *Analysis of Longitudinal Data*. New York: Oxford University Press, 2002.
- [6] T. Gneiting, "Nonseparable, Stationary Covariance Functions for Space-Time Data," *NRCSE-Technical Report Series*, no. 063, 2001.
- [7] T. Gneiting, M. G. Genton, and P. Guttorp, "Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry," in *Statistical Methods for Spatio-Temporal Systems*, L. H. Barbel Finkenstadt, Ed. Boca Raton, FL: Chapman & Hall/CRC, 2007, ch. 4, pp. 151–175.
- [8] T. Gneiting and P. Guttorp, "Continuous Parameter Stochastic Process Theory," in *Handbook of Spatial Statistics*, A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, Eds. Boca Raton, FL: Chapman & Hall/CRC, 2010, ch. 2, pp. 18–19.

- [9] T. Gneiting and M. Schlather, “Stochastic Models That Separate Fractal Dimension and the Hurst Effect,” *SIAM Review*, vol. 46, no. 2, pp. 269–282, Jan. 2004. [Online]. Available: <http://epubs.siam.org/doi/abs/10.1137/S0036144501394387>
- [10] P. Goovaerts, “Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall,” *Journal of Hydrology*, vol. 228, no. 1-2, pp. 113–129, Feb. 2000. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S002216940000144X>
- [11] M. S. Handcock and M. L. Stein, “A Bayesian Analysis of Kriging,” *Technometrics*, vol. 35, no. 4, pp. 403–410, 1993.
- [12] G. B. M. Heuvelink and D. a. Griffith, “Space-Time Geostatistics for Geography: A Case Study of Radiation Monitoring Across Parts of Germany,” *Geographical Analysis*, vol. 42, no. 2, pp. 161–179, Apr. 2010.
- [13] K. M. Hinkel and F. E. Nelson, “Spatial and temporal patterns of active layer thickness at Circumpolar Active Layer Monitoring (CALM) sites in northern Alaska, 1995-2000,” *Journal of Geophysical Research*, vol. 108, no. D2, p. 8168, 2003.
- [14] G. Matheron, “Les Variables Regionalisees et leur Estimation,” 1965.
- [15] B. Minasny and A. B. McBratney, “The Matérn function as a general model for soil variograms,” *Geoderma*, vol. 128, no. 3-4, pp. 192–207, Oct. 2005.
- [16] L. Spadavecchia, “Edinburgh Space-Time Geostatistics Users Guide,” pp. 172–227, 2008.
- [17] M. L. Stein, “Properties of Random Fields,” in *Interpolation of Spatial Data: Some Theory for Kriging*, M. L. Stein, Ed. New York, NY: Springer, 1999, pp. 1–4.
- [18] D. A. Streletskiy, N. I. Shiklomanov, F. E. Nelson, and A. E. Klene, “13 Years of Observations at Alaskan CALM Sites : Long-term Active Layer and Ground Surface Temperature Trends,” in *Ninth International Conference on Permafrost*, 2007.



# Appendix

