Article

Estimation of the Covariance Matrix in Hierarchical Bayesian Spatio-Temporal Modeling via Dimension Expansion

Bin Sun and Yuehua Wu *

Department of Mathematics and Statistics, York University, Toronto, ON M3J 1P3, Canada; bsun1010@gmail.com
* Correspondence: wuyh@yorku.ca

Abstract: Ozone concentrations are key indicators of air quality. Modeling ozone concentrations is challenging because they change both spatially and temporally with complicated structures. Missing data bring even more difficulties. One of our interests in this paper is to model ozone concentrations in a region in the presence of missing data. We propose a method without any assumptions on the correlation structure to estimate the covariance matrix through a dimension expansion method for modeling the semivariograms in nonstationary fields based on the estimations from the hierarchical Bayesian spatio-temporal modeling technique (Le and Zidek). Further, we apply an entropy criterion (Jin et al.) based on a predictive model to decide if new stations need to be added. This entropy criterion helps to solve the environmental network design problem. For demonstration, we apply the method to the ozone concentrations at 25 stations in the Pittsburgh region studied. The comparison of the proposed method and the one is provided through leave-one-out cross-validation, which shows that the proposed method is more general and applicable.

Keywords: entropy; environmental network design; dimension expansion; hierarchical Bayesian spatio-temporal modeling; nonstationary field; semivariogram

1. Introduction

Ozone concentrations are the daily maximum 8 h moving averages of hourly ozone concentration data recorded in micrograms per cubic meter, $\mu g/m^3$, which are key indicators of air quality. Monitoring the changes both spatially and temporally is very important for the assessment of air quality change, which has a great impact on our environment, society and economy. However, modeling the ozone concentrations is not an easy task since the ozone concentrations vary over space and time with complicated spatial structures, temporal structures and spatio-temporal interactions. Furthermore, the presence of missing data brings even more difficulties. As commented in [1], although we cannot escape the "curse of dimensionality", we can take advantage of recent developments in computing speed and numerical advances (e.g., Markov chain Monte Carlo) that allow us to implement Bayesian spatio-temporal dynamical models in a hierarchical framework. Such a framework provides simple strategies for incorporating complicated spatio-temporal interactions at different stages of the models' hierarchy, and the models are feasible to be implemented for high-dimensional data. Two popular hierarchical Bayesian spatio-temporal models can be found in [1,2], among others. The latter one was used in [3].

Ref. [3] studied the ozone concentrations within $-79°$ to $-81.5°$ longitude and $39.5°$ to $41.5°$ latitude around the Pittsburgh region ($-79.23°$, $43.39°$), in which all of the monitoring stations have missing data. That paper dealt with the missing problems in two steps. First, it filled in some of the missing measurements by using linear models so that the pattern of missing data became monotone (the monotone missing is also referred to as the staircase pattern). Second, it applied *h*ierarchical *B*ayesian *s*patio-*t*emporal (HBST) modeling proposed in [2] on this staircase of missing data to estimate the hyperparameters of the spatial-temporal model. Based on the estimated hyperparemeters, it estimated the

spatial correlation function for the monitoring stations. Then, it estimated the covariance matrix for all of the stations and derived the predictive distribution for the ungauged sites.

Generalized linear models can be used to accommodate non-Gaussian geostatistical data (e.g., see [4]). Ref. [3] selected the generalized linear model with the quasi-Poisson family as an appropriated spatial correlation function by examining the pattern of spatial correlations obtained via the hierarchical model in the plot. However, their link function is not appropriate if there are negative correlations. This is a strong restriction because negative correlations are common for the ozone concentrations and other spatial-temporal data. Moreover, choosing a model by examining the plots derived in terms of the observed data set is not rigorous enough and may only be suitable just for a particular data set.

In this paper, we propose a method to estimate the covariance matrix through a dimension expansion method for modeling the semivariograms in nonstationary fields based on the estimations from hierarchical Bayesian spatio-temporal modeling. For demonstration, we apply the proposed method on the same data as in Jin et al. [3]. Without any assumption on the correlation structure, the proposed method is more general than the method in [3] such that it is applicable to other spatio-temporal data sets. Using the covariance matrix estimated by the proposed method on the entropy criterion in the environmental network design problem, our study provides interesting findings, and the locations of the selected ungauged stations are more reasonable. We provide comparison of these two methods through leave-one-out cross-validation, which shows that the proposed method provides improved results.

The paper is arranged as follows. In Section 2, we briefly introduce hierarchical Bayesian spatio-temporal modeling. In Section 3, we describe the ozone concentrations in the Pittsburgh region and apply the hierarchical Bayesian spatio-temporal modeling techniques for filling in missing measurements following [3]. In Section 4, we model the ozone concentrations in the Pittsburgh region. We first introduce the method for estimating the covariance matrix through a dimension expansion method for modeling the semivariograms in nonstationary fields, and we then give spatial predictive distributions on the ungauged sites using the covariance matrix estimated by the proposed method. In Section 5, we present the results of the entropy of the predictive distributions and an optimality criterion for extending an environmental network. In Section 6, we provide the model evaluation through leave-one-out cross-validation. We conclude this paper with a conclusion in Section 7.

Throughout the rest of the paper, the L_1-norm of a vector c is denoted by $\|c\|_1$, a $p \times p$ identity matrix is denoted by I_p, the transpose of a matrix A is denoted by $A^\top$ and the trace of a square matrix B is denoted by $\text{tr}(B)$. In addition, '$\otimes$' represents the Kronecker product, $N_{k \times \ell}(\cdot, \cdot)$ refers to a matrix Gaussian distribution, $t_{k \times \ell}(\cdot, \cdot)$ denotes a matric-t distribution, $IW(\cdot, \cdot)$ stands for the inverted Wishart distribution (see (a) of the appendix for definitions of these distributions) and $GIW(\cdot, \cdot)$ denotes the generalized inverted Wishart distribution.

2. Hierarchical Bayesian Spatio-Temporal Modeling

We briefly describe HBST modeling in this section, which is the same as that given in [3] excluding Step 3 in the HBST modeling procedure. It is noted that this modeling is a special case of the HBST modeling presented in Chapter 10 of Le and Zidek (2006) excluding Step 3 in the HBST modeling procedure.

Define the following notations:

d = number of different type stations (e.g., agricultural, residential, commercial and industrial);

n = number of time points (e.g., number of days);

u = number of locations with no monitors (i.e., ungauged sites);

g = number of locations with monitors (i.e., gauged sites).

The stations are organized into k blocks where the g_j ($j = 1, 2, \cdots, k$) sites in the jth block have the same number of timepoints m_j at which no measurements are taken. These

blocks are numbered so that the measurements correspond to a monotone data pattern or a staircase structure, that is,

$$m_1 > m_2 > \cdots > m_k \geq 0.$$

The response variables are written as

$$Y = [Y^{[u]}, Y^{[g]}].$$

Here, $Y^{[u]}$ of dimension $n \times u$ denotes the unobserved responses at ungauged sites while $Y^{[g]}$ of dimension $n \times g$ is given by

$$Y^{[g]} = \left[Y^{[g_1]}, \cdots, Y^{[g_k]} \right] = \left[\begin{pmatrix} Y^{[g_1^m]} \\ Y^{[g_1^o]} \end{pmatrix}, \cdots, \begin{pmatrix} Y^{[g_k^m]} \\ Y^{[g_k^o]} \end{pmatrix} \right],$$

where $Y^{[g_j^m]}$ is an $m_j \times g_j$ matrix of missing measurements at the g_j gauged sites for the m_j time points and $Y^{[g_j^o]}$ is an $(n - m_j) \times g_j$ matrix of observed measurements at the g_j gauged sites for the $(n - m_j)$ time points.

We assume that the response matrix Y follows the Gaussian and generalized inverted Wishart model specified by

$$\begin{cases} Y|B, \Sigma & \sim & N(XB, I_n \otimes \Sigma), \\ B|\Sigma, B_0 & \sim & N(B_0, V_B \otimes \Sigma), \\ \Sigma & \sim & GIW(\Theta, \delta). \end{cases} \tag{1}$$

where B is an $l \times (g + u)$ coefficient matrix with the hyperparameter mean matrix B_0 and the variance components V_B, X is the matrix of covariates which is defined in (4) and $\{\Theta, \delta\}$ is a set of model parameters specified below.

We partition B corresponding to the l time-varying covariates in conformance with the block structure as

$$B = (B^{[u]}, B^{[g_1]}, \cdots, B^{[g_k]}).$$

By assuming an exchangeable structure across sites, B can be written as $B = \tilde{B}R$, where $\tilde{B}$ is the $l \times d$ hyperparameter matrix and $R = (r_{i,j})_{d \times (u+g)}$ with $r_{i,j} = 1$ for Station j under Class i and $r_{ij} = 0$ otherwise.

Likewise, we partition the $(u + g) \times (u + g)$ covariance matrix Σ over gauged and ungauged sites conformably as

$$\Sigma = \begin{pmatrix} \Sigma^{[u,u]} & \Sigma^{[u,g]} \\ \Sigma^{[g,u]} & \Sigma^{[g,g]} \end{pmatrix},$$

where $\Sigma^{[u,u]}$ is a $u \times u$ matrix being for the ungauged sites. Further, we partition the $g \times g$ covariance matrix $\Sigma^{[g,g]}$ for the gauged site blocks as follows:

$$\Sigma^{[g,g]} = \begin{pmatrix} \Sigma^{[g_1,g_1]} & \cdots & \Sigma^{[g_1,g_k]} \\ \vdots & \vdots & \vdots \\ \Sigma^{[g_k,g_1]} & \cdots & \Sigma^{[g_k,g_k]} \end{pmatrix}.$$

Similarly, for $j = 1, \cdots, k$, we put

$$\Sigma^{[g_j, \cdots, g_k]} = \begin{pmatrix} \Sigma^{[g_j,g_j]} & \cdots & \Sigma^{[g_j,g_k]} \\ \vdots & \vdots & \vdots \\ \Sigma^{[g_k,g_j]} & \cdots & \Sigma^{[g_k,g_k]} \end{pmatrix}.$$

We reparametrize the matrix Σ through the recursive one-to-one Bartlett transformation for the two blocks:

$$
\Sigma = \begin{pmatrix} \Gamma^{[u]} + (Y^{[u]})^\top \Sigma^{[g,g]} Y^{[u]} & (Y^{[u]})^\top \Sigma^{[g,g]} \\ \Sigma^{[g,g]} Y^{[u]} & \Sigma^{[g,g]} \end{pmatrix},
$$

where $\Gamma^{[u]} = \Sigma^{[u,u]} - \Sigma^{[u,g]} \left(\Sigma^{[g,g]} \right)^{-1} \Sigma^{[g,u]}$ and $Y^{[u]} = \left(\Sigma^{[g,g]} \right)^{-1} \Sigma^{[u,g]}$. Similarly, by applying the Bartlett decomposition, we can represent the submatrix $\Sigma^{[g_j, \cdots, g_k]}$, for $j = 1, \cdots, k-1$, as

$$
\Sigma^{[g_j, \cdots, g_k]} = \begin{pmatrix} \Gamma_j + Y_j^\top \Sigma^{[g_{j+1}, \cdots, g_k]} Y_j & Y_j^\top \Sigma^{[g_{j+1}, \cdots, g_k]} \\ \Sigma^{[g_{j+1}, \cdots, g_k]} Y_j & \Sigma^{[g_{j+1}, \cdots, g_k]} \end{pmatrix},
$$

where $\Gamma_k = \Sigma^{[g_k, g_k]}$ and for $j = 1, \cdots, k-1$,

$$
\Gamma_j : g_j \times g_j = \Sigma^{[g_j, g_j]} - \Sigma^{[g_j, (g_{j+1}, \cdots, g_k)]} \left(\Sigma^{[g_{j+1}, \cdots, g_k]} \right)^{-1} \Sigma^{[(g_{j+1}, \cdots, g_k), g_j]},
$$

$$
Y_j : (g_{j+1} + \cdots + g_k) \times g_j = \left(\Sigma^{[g_{j+1}, \cdots, g_k]} \right)^{-1} \Sigma^{[(g_{j+1}, \cdots, g_k), g_j]},
$$

with

$$
\Sigma^{[g_{j+1}, \cdots, g_k]} = \begin{pmatrix} \Sigma^{[g_{j+1}, g_j]} \\ \vdots \\ \Sigma^{[g_k, g_j]} \end{pmatrix}.
$$

Therefore, the GIW prior distribution for Σ in (1) is equivalently defined in terms of $(\Gamma^{[u]}, Y^{[u]})$ and $\{(\Gamma_1, Y_1), \cdots, (\Gamma_k, Y_k), \Sigma_k\}$ as follows:

$$
\begin{cases} Y^{[u]} | \Gamma^{[u]} & \sim & N(Y_{00}, H_0 \otimes \Gamma^{[u]}), \\ \Gamma^{[u]} & \sim & IW(\Lambda_0, \delta_0), \\ Y_j | \Gamma_j & \sim & N(Y_{0j}, H_j \otimes \Gamma_j), \quad j = \cdots, k-1, \\ \Gamma_j & \sim & IW(\Lambda_j, \delta_j), \quad j = 1, \cdots, k, \end{cases} \tag{2}
$$

where $Y^{[u]}$ is the slope of the optimal linear predictor of $Y^{[u]}$ based on $Y^{[g]}$ and $\Gamma^{[u]}$ is the residual covariance of the optimal linear predictor. Similar interpretations can be applied to Y_j and Γ_j, for $j = 1, \cdots, k-1$.

Let $\mathcal{H}$ be the set of the hyperparameters in (1) and (2), i.e., $\mathcal{H} = \{\Theta, \delta, V_B, B_0\}$, where $\Theta = \{(Y_{00}, H_0, \Lambda_0), (Y_{01}, H_1, \Lambda_1), \cdots, (Y_{0k-1}, H_{k-1}, \Lambda_{k-1}), \Lambda_k\}$ with degrees of freedom parameters $\delta = (\delta_0, \delta_1, \cdots, \delta_k)^\top$. Write $\mathcal{H} = [\mathcal{H}_u, \mathcal{H}_g]$. Here, $\mathcal{H}_g = \{V_B, B_0, (Y_{01}, H_1, \Lambda_1, \delta_1), \cdots, (Y_{0,k-1}, H_{k-1}, \Lambda_{k-1}, \delta_{k-1}), (\Lambda_k, \delta_k)\}$, which represents the hyperparameters involved in the marginal distribution of $Y^{[g^o]}$.

If a data matrix appears to be an ascending staircase, the HBST modeling procedure is given as follows:

Step 1. Compute the hyperparameter values that maximize the marginal distribution $f(Y^{[g^o]} | \mathcal{H}_g)$ using an empirical Bayesian approach (see (b) of Appendix A). The EM algorithm is used to obtain $\hat{\mathcal{H}}_g$.

Step 2. Obtain the predictive distributions $f(Y^{[g_k^m]} | Y^{[g^o]}, \hat{\mathcal{H}}_g)$ of missing measurements as in (c) of Appendix A. Fill in the missing data by using the predictive distributions.

Step 3. Obtain the estimate $\hat{\Sigma}^{[g,g]}$ from the estimate of $\hat{\mathcal{H}}_g$. In terms of $\hat{\Sigma}^{[g,g]}$, obtain the estimate of the covariance matrix by using a dimension expansion method given in Qin et al. [5] and the thin-plate spline method given in Wabba and Wendelberger (1980). The details are given in Section 4.1.

Step 4. Estimate the hyperparameters $\mathcal{H}_u$ and obtain the conditional predictive distribution $f(Y^{[u]}|Y^{[g]}, \hat{\mathcal{H}})$ (see Section 4.2).

3. Ozone Concentrations from the Monitoring Stations in Pittsburgh Region

The ozone concentrations were recorded within $-79°$ to $-81.5°$ longitude and $39.5°$ to $41.5°$ latitude around the Pittsburgh region ($-79.23°$, $43.39°$) for four consecutive summer months, June, July, August and September, over the period from 1995 to 2007. There were 25 monitoring stations in the region as shown in Figure 1, which is the same as Figure 1 in [3]. The original data set Y_0 was collected from 25 stations, and there were a total of 1586 (13 years $\times$ 122 days) measurements at each station. The number of missing data in Y_0 is shown by N1.Miss in Table 1, which is the same as Table 1 in [3]. In this section, we fill in missing measurements.

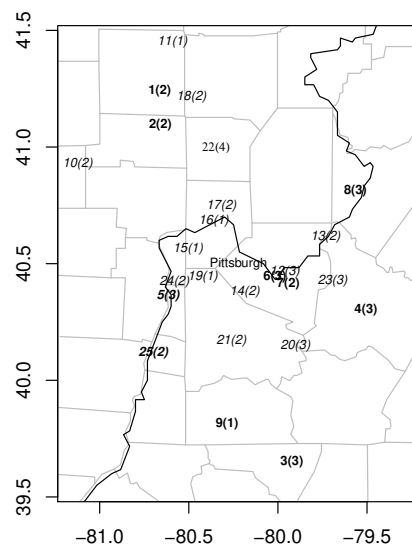**Agricultural(1) Residential(2) Commercial(3) Industrial(4)**

Figure 1. Monitoring stations in the Pittsburgh region.

Table 1. Location of the stations and number of missing data.

ID	Class	Lon	Lat	N1.Miss	N2.Miss	N3.Miss	ID	Class	Lon	Lat	N1.Miss	N2.Miss	N3.Miss
1	2	−40.24	80.66	855	854	854	14	2	−40.38	80.18	22	0	0
2	2	−41.09	80.65	610	610	610	15	1	−40.56	80.50	13	0	0
3	3	−39.64	79.92	618	610	610	16	1	−40.68	80.35	11	0	0
4	3	−40.30	79.50	488	488	488	17	2	−40.74	80.31	4	0	0
5	3	−40.36	80.61	858	854	366	18	2	−41.21	80.48	5	0	0
6	3	−40.44	80.01	370	366	366	19	1	−40.44	80.42	16	0	0
7	2	−40.41	79.94	370	366	366	20	3	−40.14	79.90	3	0	0
8	3	−40.81	79.56	328	318	318	21	2	−40.17	80.26	1	0	0
9	1	−39.81	80.28	278	244	244	22	4	−40.99	80.34	0	0	0
10	2	−40.93	81.12	12	0	0	23	3	−40.42	79.69	5	0	0
11	1	−41.45	80.59	1	0	0	24	2	−40.42	80.58	5	0	0
12	3	−40.46	79.96	2	0	0	25	2	−40.12	80.69	488	488	0
13	2	−40.61	79.73	8	0	0							

The numbers 1, 2, 3 and 4 under **Class** denote agricultural, residential, commercial and industrial, respectively.

3.1. Filling in the Missing Measurements for Each Monitoring Station within the Period of Monitoring Blocks

Since there are missing data in the dataset, we follow the steps in [3] in filling in some missing measurements occurred during the operation of each monitoring station, using the regression model as

$$
\begin{aligned}
y_{122(i-1)+j} &= a\sin\left(\frac{2(122(i-1)+j)\pi}{122}\right) + b\cos\left(\frac{2(122(i-1)+j)\pi}{122}\right) + c_i + \varepsilon_{122(i-1)+j} \\
&= a\sin\left(\frac{j\pi}{61}\right) + b\cos\left(\frac{j\pi}{61}\right) + c_i + \varepsilon_{122(i-1)+j},
\end{aligned}
\tag{3}
$$

for $i = 1,\ldots,13$, and $j = 1,\ldots,122$, where a and b are regression coefficients, c_i, for $i = 1,\ldots,13$, are the categorical factors and $\{\varepsilon_t\}$ is a sequence of independently and identically distributed Gaussian random variables with mean 0 and variance σ^2. The model (3) assigns different means to the years with a yearly cycle of 122 days. We re-express the 13 factors in the model via Helmert contrasts, which compare the first level of the factor with all later levels, the second level with all later levels, and so forth. The Helmert matrix, $Z_{13\times13}$, is defined as follows.

$$
Z = \begin{pmatrix}
1 & -1 & -1 & \cdots & -1 & -1 \\
1 & 1 & -1 & \cdots & -1 & -1 \\
1 & 0 & 2 & \cdots & -1 & -1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & 0 & 0 & \cdots & 11 & -1 \\
1 & 0 & 0 & \cdots & 0 & 12
\end{pmatrix}.
$$

Let X, the matrix of covariates, be

$$
X = \begin{pmatrix} S & Z\otimes\mathbf{1}_{122} \end{pmatrix}_{1586\times15},
\tag{4}
$$

where $\mathbf{1}_n = (1,1,\ldots,1,1)_{1\times n}^\top$ and

$$
S = \begin{pmatrix}
\sin(\pi/61) & \cdots & \sin(i\pi/61) & \cdots & \sin(1586\pi/61) \\
\cos(\pi/61) & \cdots & \cos(i\pi/61) & \cdots & \cos(1586\pi/61)
\end{pmatrix}_{2\times1586}^\top,
$$

and let $\mathbf{y} = (y_1, y_2, \ldots, y_{1586})^\top$, $\boldsymbol{\beta} = (a, b, d_1, \ldots, d_{13})^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{1586})^\top$ denote the response variables, regression coefficient vector and error variables, respectively. The model (3) is written as $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

We then fill in the missing measurements within the blocks by the least squares predictions plus errors and obtain a new data set Y_1 in which the unfilled missing measurements are either in the end of the time period or in the beginning of the time period. The number of missing data in Y_1 is shown in Table 1 by N2.Miss.

3.2. Filling in the Missing Measurements in Y_1

To fill in the missing measurements in Y_1, we can proceed as follows [3]:

(i)   Obtain a new data set Y_2 from Y_1 by filling in the 488 missing measurements at Stations 5 and 25 during the end of the time period by using the HBST modeling technique. N3.Miss in Table 1 displays the number of missing data in the data set Y_2, which shows that Y_2 has a staircase data structure, as all of the missing data are located in the beginning of the time period.

(ii) Put $d = 4$, $l = 15$, $n = 1586$, $k = 7$, $m_1 = 854$, $m_2 = 610$, $m_3 = 488$, $m_4 = 366$, $m_5 = 318$, $m_6 = 244$, $m_7 = 0$, $g_1 = 1$, $g_2 = 1$, $g_3 = 0$, $g_4 = 3$, $g_5 = 1$, $g_6 = 1$ and $g_7 = 16$. Fill in the remaining missing values in Y_2 by executing Steps 1–2 of the HBST modeling procedure.

4. Model the Ozone Concentrations in the Pittsburgh Region

To model ozone concentrations in the Pittsburgh region by spatial interpolation, we cover the region by the 100 grid boxes of a spatial resolution of latitude $0.2°\times$ longitude

$0.2°$. Thus, $u = 100$. The grid points are ungauged sites, and their classes are displayed in Figure 4. To derive the predictive distributions for these grid points, a key step is to estimate the covariance matrix.

4.1. Estimation of the Covariance Matrix

In this subsection, we introduce a method for estimating the covariance matrix through a dimension expansion method for modeling the semivariograms in nonstationary fields in terms of $\hat{\mathcal{H}}_g$ from Step 1 of the HBST procedure.

Let $\{Y(x) : x \in \mathcal{S}\}$, $\mathcal{S} \in \mathcal{R}^d$, be an environmental random process, where x is a d-dimensional spatial index that varies continuously throughout the region $\mathcal{S}$. At n spatial locations denoted by $\{x_i : i = 1, \ldots, n\}$, we observe realizations of the random process $Y(x)$, i.e., $\{Y(x_i) : i = 1, \ldots, n\}$. We are interested in learning the spatial dependency of the process through the observed data. The semivariogram function which describes the degree of spatial dependency of an intrinsic stationary random process is a cornerstone in spatial statistics. An intrinsic stationary random process satisfies the following two conditions (Cressie [6]):

1. $\mathbb{E}(Y(x)) = \mathcal{U}$, for $x \in \mathcal{S}$,
2. $\text{var}(Y(x_i) - Y(x_j)) = 2\gamma(x_i - x_j)$,

where a semivariogram is defined as $\gamma(x_i - x_j) = \frac{1}{2}\text{var}(Y(x_i) - Y(x_j))$ for two different locations, x_i and x_j, in the monitored region. The estimated covariance matrix of the monitoring stations $\hat{\Sigma}^{[g,g]}$ is based on the estimation of $\mathcal{H}_g$ from Step 1 of the HBST procedure. We estimate the semivariograms of the ozone concentrations from the monitoring stations by

$$\hat{\gamma}(x_i - x_j) = \frac{1}{2}\text{vâr}(Y(x_i)) + \frac{1}{2}\text{vâr}(Y(x_j)) - \text{côv}(Y(x_i), Y(x_j)). \tag{5}$$

From Figure 2, we notice that the estimated semivariograms related to Station 3 (marked by "×") are much higher than the other stations. We examine the location of Station 3 and notice that it was on the edge of the monitored region. Moreover, there were over ten airports around this station. According to Xue et al. [7], there is a great impact of high-altitude aircrafts on the ozone layer in the stratosphere. This becomes an influential factor in modeling the ozone concentrations. Next, we introduce how this factor is considered in the proposed modeling technique.
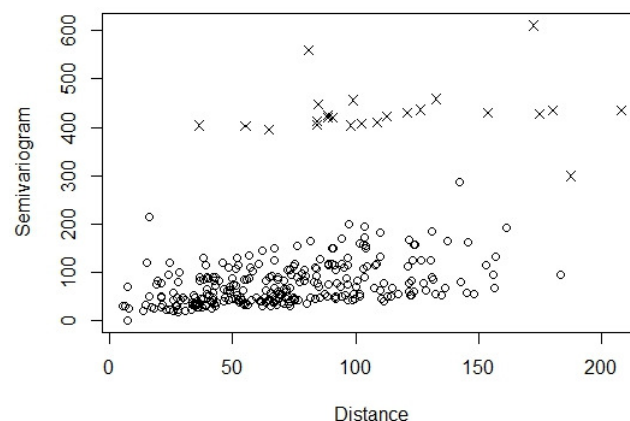

Figure 2. Empirical semivariograms of the ozone concentrations from the monitoring stations versus the Euclidean distances between monitoring stations based on the Bayesian hierarchical model. The semivariograms related to Station 3 are marked by "×".

It is obvious that this field is not stationary. Bornn et al. [8] proposed a novel approach to find the latent dimensions over which the nonstationary fields exhibit stationarity through dimension expansion. They justified that for a nonstationary Gaussian process

$Y(x)$, where $x \in R^d$, there exists a vector $z \in R^p$, $p > 0$, such that the expanded process $Y([x, z])$ is stationary under appropriate moment constraints. Note that $[x, z]$ is the concatenation of the vectors x and z. The stationary semivariogram with latent vectors can be expressed by

$$2\gamma([x_i, z_i] - [x_j, z_j]) = \mathbb{E}(Y([x_i, z_i]) - Y([x_j, z_j]))^2,$$

where $[x_i, z_i]$ is the expanded spatial index for the ith location. Qin et al. [5] improved the method in Bornn et al. [8] by considering the covariance structure of the $\hat{\gamma}_{i,j}$, for $j \neq i$, which are generally correlated. In our application, we use the lasso-penalized weighted least-squares criterion (WLS) in Qin et al. [5] as follows,

$$(\hat{\phi}, Z)_{WLS} = \underset{\text{Œ}, Z}{\text{argmin}} \sum_{j < i} \frac{1}{\gamma_\phi^2(d_{i,j}([X, Z]))} \{\hat{\gamma}_{i,j} - \gamma_\phi(d_{i,j}([X, Z]))\}^2 + \lambda \sum_{k=1}^{p} \|Z_{\cdot k}\|_1. \quad (6)$$

to estimate the parameters and the expanded dimensions. Here, $\hat{\gamma}_{i,j}$ is the estimated semivariogram by (5) and $d_{i,j}([X, Z])$ is the Euclidean distance between the locations $[x_i, z_i]$ and $[x_j, z_j]$ and $Z_{\cdot k}$ is the kth column of Z. $[X, Z]$ is the concatenation of the matrices X and Z. The tuning parameter λ in the group lasso is used to determine the number of latent dimensions and regularize the estimation of Z to prevent overfitting. $\gamma_\phi(d_{i,j}([X, Z]))$ is a parametric stationary semivariogram model with parameter ϕ. The most popular ones are the exponential model, the spherical model and the Gaussian model (see Journel and Huijbregts [9] and Cressie [6]), among others). For example, the exponential model is defined as

$$\gamma_\phi(d) = \phi_1(1 - \exp(-d/\phi_2)) + \phi_3,$$

where $\phi = (\phi_1, \phi_2, \phi_3)^\top$, $\phi_1 \geq 0$, $\phi_2 \geq 0$ and $\phi_3 \geq 0$.

The semivariogram plot with estimated expanded dimensions (Figure 3) of the monitoring stations shows that the field is in good agreement with the theoretical model, as most of the points are near the solid red line, the fitted exponential semivariogram model. Two extra dimensions are added to the original coordinate with $\lambda = 0.01$. Figure 3 shows that with the extra dimensions, Station 3 is pushed much further out of the two-dimensional plane, reflecting the impact of high-altitude aircrafts on the ozone layer in the stratosphere we have mentioned earlier.
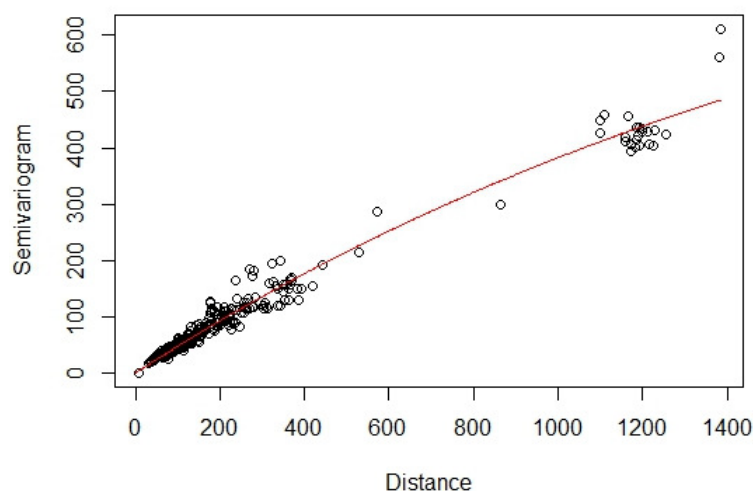

Figure 3. Semivariogram plot of the ozone concentrations from the monitoring stations over a larger range of distances than the range shown in Figure 2, owing to the application of dimension expansion. The fitted exponential semivariogram model is shown by the red solid line.

After the expanded dimensions for the monitoring stations are obtained, we use the thin-plate spline method [10] to estimate the hidden dimensions for the ungauged

sites. The semivariograms for the ungauged stations are estimated by the exponential model using the estimated parameter vector $\hat{\phi}$. Next, we estimate the semivariograms γ_{s_i,s_j} between stations s_i and s_j using the exponential model based on the distances over the space composed by the original and the expanded dimensions. Last, the covariance between any two sites can be estimated by

$$\hat{\Sigma}_{i,j} = \widehat{\mathrm{Cov}}(Y(s_i), Y(s_j)) = \frac{1}{2}\hat{\sigma}_{Y(s_i)} + \frac{1}{2}\hat{\sigma}_{Y(s_j)} - \hat{\gamma}_{s_i,s_j},$$

where $\hat{\sigma}_{Y(s_i)}$ and $\hat{\sigma}_{Y(s_j)}$ are estimates of $\sigma_{Y(s_i)}$ and $\sigma_{Y(s_j)}$ obtained by the thin-plate spline approach.

4.2. Prediction of the Daily Ozone Concentrations at the Grid Points

By Chapter 10 of Le and Zidek (2006), spatial predictive distributions at the grid points given the monitoring sites are as follows:

$$(Y^{[u]}|Y^{[g]}, \mathcal{H}) \sim t_{n \times u}\left(\mathcal{U}^{u|g}, \frac{\Phi^{[u|g]} \otimes \Psi^{[u|g]}}{\delta_0^*}, \delta_0^*\right), \tag{7}$$

where $\delta_0^* = \delta_0 - u + 1$, $\Psi^{[u|g]} = \Lambda_0$, $\mathcal{U}^{[u|g]} = XB_0^{[u]} + (Y^{[g]} - XB_0^{[g]})Y_{00}$ and $\Phi^{[u|g]} = I_n + XV_BX^\top + (Y^{[g]} - XB_0^{[g]})H_0(Y^{[g]} - XB_0^{[g]})^\top$ (see (a) of Appendix A for definition of the matric-*t* distribution).

We estimate the hyperparameters associated with the grid points Λ_0, Y_{00}, H_0 and δ_0 via

$$\hat{\delta}_0 = \frac{\hat{\delta}_1 + \cdots + \hat{\delta}_k}{k}, \quad \hat{H}_0 = \hat{\Lambda}^{[1,\cdots,k]}, \quad \hat{Y}_{00} = (\hat{\Sigma}^{[g,g]})^{-1}\hat{\Sigma}^{[g,u]},$$

$$\hat{\Lambda}_0 = \frac{\hat{\delta}_0 - u - 1}{1 + \mathrm{tr}(\hat{\Sigma}^{[g,g]}\hat{H}_0)}(\hat{\Sigma}^{[u,u]} - \hat{Y}_{00}^\top\hat{\Sigma}^{[g,g]}\hat{Y}_{00})$$

with

$$\hat{\Lambda}^{[j,\cdots,k]} = \begin{pmatrix} \hat{\Lambda}_j + \hat{Y}_{0j}^\top\hat{\Lambda}^{[j+1,\cdots,k]}\hat{Y}_{0j} & \hat{Y}_{0j}^\top\hat{\Lambda}^{[j+1,\cdots,k]} \\ \hat{\Lambda}^{[j+1,\cdots,k]}\hat{Y}_{0j} & \hat{\Lambda}^{[j+1,\cdots,k]} \end{pmatrix}, \quad j = 1,\ldots,k-1,$$

and $\hat{\Lambda}^{[k]} = \hat{\Lambda}_k$.

After all of the hyperparameters in the predictive distributions are estimated, we can predict the daily ozone concentrations at all the grid points in the time period of study by generating samples from the predictive distributions.

5. Environmental Network Extension

Assume that Y has the density function f. The total reduction in uncertainty of Y can be presented by the entropy of its distribution, i.e., $\mathfrak{H}(Y) = -E[\log f(Y)/h(Y)]$, where $h(\cdot)$ is a not necessarily integrable reference density (Jaynes [11]). According to the predictive distribution (7), the total entropy $\mathfrak{H}(Y^{[u]}|Y^{[g]})$ can be defined as

$$\mathfrak{H}(Y^{[u]}|Y^{[g]}) = \frac{1}{2}\log\left(\left|\Psi^{[u|g]}\right|\right) + c_u(u,q), \tag{8}$$

where $c_u(u,q)$ is a constant depending on the degree of freedom and the dimension of the ungauged sites.

The key step in expanding an environmental network is to find appropriate ungauged sites to add to the existing network that maximizes the corresponding entropy. We use the following optimality criterion as given in [3]:

$$\max_{add}\left(\frac{1}{2}\log\left|\Psi^{[u|g]}\right|\right)^{add} \tag{9}$$

The *add* sites, in a vector of dimension u_1, are selected to maximize the entropy in (8). In [3], the grid points $\{91, 92, 93\}$ were selected with the highest entropy 11.3774. The proposed method selects the grid points $\{41, 71, 100\}$ with entropy 12.1207. This selection is more reasonable, as they are not gathered in the southeast corner of the region like $\{91, 92, 93\}$. The selected sites among 100 grid points by the two methods are shown in Figure 4 below.
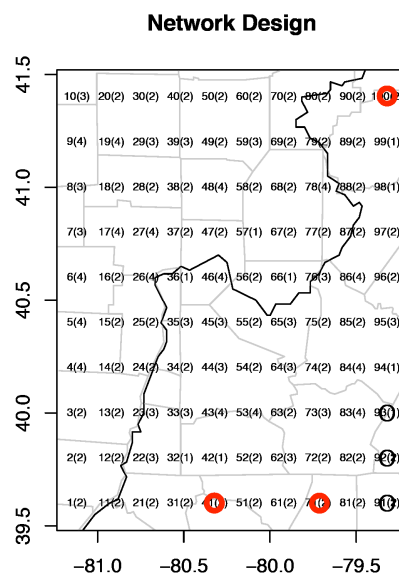

Figure 4. The selected sites among 100 grid points (black circled points by [3] and red circled points by our method).

6. Model Evaluation

In this section, we use the leave-one-out cross-validation to evaluate the accuracy of the predictive model derived using the proposed method and compare the proposed method with the one in [3]. We select the observations from one of the original 25 stations as validation data, and observations in the remaining 24 stations are treated as training data. We use the data from day 855 to day 1586 at the end of the study from each station to evaluate the prediction because during this period, none of the stations has missing data. By choosing this period, we avoid using the Bayesian hierarchical modeling technique for estimating the missing data in the training data set, which is time-consuming and not our intention for evaluating the proposed method on estimating the covariance matrix. Station 22 is excluded because it is the only industrial station in the study. For each of the 24 stations, we generate 100 samples from the predictive distribution with parameters estimated using observations from the rest of the 23 stations. We compute the average of relative absolute bias (ARAB) as $\sum_{j=1}^{100}\left|\left(y_{j,i,t} - y_{i,t}\right)/y_{i,t}\right|$, where $y_{j,i,t}$ is the jth sample generated from the predictive distributions and $y_{i,t}$ is the observation from Station i on time t. The results are given in Table 2.

In Table 2, "-" means that there is no prediction for the station because there are negative correlations and the method in [3] is not applicable to estimate the predictive distribution. The results in Table 2 show that the proposed method provides slightly more accurate predictions than the one in [3] for most of the stations. More important is that, when there are negative correlations obtained from the estimations of the hierarchical Bayesian spatio-temporal modeling technique, the method in [3] fails to estimate the covariance matrix, while the proposed method still provides accurate predictions except

for Station 3. This is expected because Station 3 is an influential station. Therefore, if we use observations at Station 3 as the validation data set, it has a great impact on estimating the covariance matrix.

Table 2. Mean and SD of the average of relative absolute bias.

ID	Our Method	Jin et al. (2012) [3]	ID	Our Method	Jin et al. (2012) [3]
1	0.0789 (0.0627)	0.8134 (0.0682)	13	0.1145 (0.1096)	0.2003 (0.1769)
2	0.1206 (0.1356)	0.1221 (0.1121)	14	0.1361 (0.1732)	0.2211 (0.2283)
3	0.8517 (0.8517)	0.1572 (0.1572)	15	0.1911 (0.2052)	-
4	0.1756 (0.1693)	-	16	0.1189 (0.1179)	0.1285 (0.1161)
5	0.1575 (0.1731)	0.1986 (0.1855)	17	0.1496 (0.1594)	0.1669 (0.1727)
6	0.1336 (0.1513)	0.1477 (0.1667)	18	0.1253 (0.1154)	0.1256 (0.1372)
7	0.1265 (0.1563)	0.1456 (0.1732)	19	0.1369 (0.1272)	0.1026 (0.0994)
8	0.0968 (0.0804)	0.1135 (0.1023)	20	0.1603 (0.1598)	0.1310 (0.1134)
9	0.1497 (0.1104)	0.1619 (0.1208)	21	0.1351 (0.1154)	0.1274 (0.1123)
10	0.1589 (0.1796)	-	23	0.1617 (0.1858)	-
11	0.6913 (0.6455)	-	24	0.1286 (0.1051)	-
12	0.1406 (0.1409)	0.1265(0.1416)	25	0.1583 (0.1701)	0.1722 (0.1675)

7. Conclusions

In this paper, we have derived a predictive model through the hierarchical Bayesian spatio-temporal modeling technique given in [12] at ungauged sites based on the covariance matrix estimated by a dimension expansion method for modeling semivariograms in nonstationary fields. Further, we have applied an entropy criterion (see [12] or [3] for details) based on the predictive model to decide if new stations need to be added. This entropy criterion helps to solve the environmental network design problem. For demonstration, we have applied the proposed method on ozone concentrations at 25 stations in the Pittsburgh region studied in [3]. The proposed method has provided satisfactory results. Moreover, the results have shown that the method is more general and applicable, as no assumption is imposed on the correlation structure.

Author Contributions: Conceptualization, B.S. and Y.W.; methodology, B.S. and Y.W.; software, B.S.; validation, B.S. and Y.W.; formal analysis, B.S.; investigation, B.S. and Y.W.; resources, Y.W.; data curation, B.S.; writing—original draft preparation, B.S. and Y.W.; writing—review & editing, Y.W.; supervision, Y.W.; project administration, Y.W. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Hierarchical Bayesian Spatio-Temporal Modeling

The following is mainly based on Chapter 10 of Le and Zidek (2006).

(a) Somedistributions

Matrix normal distribution. If, for an $n \times m$ matrix $\mathcal{U}$ and positive definite matrices $A : n \times n$ and $B : m \times m$, the density function of an $n \times m$ random matrix X has the form

$$f(X) = (2\pi)^{-\frac{nm}{2}} |A|^{-\frac{m}{2}} |B|^{-\frac{n}{2}} \exp\{-\frac{1}{2}\text{tr}[A^{-1}(X - \mathcal{U})B^{-1}(X - \mathcal{U})^{\top}]\},$$

then X is said to have a matrix normal distribution and is denoted by $X \sim N_{n \times m}(\mathcal{U}, A \otimes B)$.

Inverted Wishart distribution. If, for a $p \times p$ positive definite matrix Ψ and a positive constant $\delta \geq p$, the density function of a $p \times p$ random matrix S has the form

$$f(S) = \frac{1}{2^{\frac{\delta p}{2}} \Gamma_p(\frac{\delta}{2})} |\Psi|^{\frac{\delta}{2}} |S|^{\frac{-(\delta+p+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(S^{-1}\Psi)\right\},$$

then S is said to have an inverted Wishart distribution and is denoted by $S \sim IW(\Psi, \delta)$.

Matric-t distribution. If, for an $n \times m$ matrix $\mathcal{U}$, positive definite matrices $A : n \times n$ and $B : m \times m$ and a positive constant $\delta > 0$, the density function of an $n \times m$ random matrix X has the form

$$f(X) \propto |A|^{-\frac{m}{2}} |B|^{-\frac{n}{2}} \left| I_n + \delta^{-1} A^{-1}(X-U)B^{-1}(X-\mathcal{U})^\top \right|^{-\frac{\delta+n+m-1}{2}},$$

then X is said to have a matric-t distribution and is denoted by $X \sim t_{n \times m}(\mathcal{U}, A \otimes B, \delta)$.

(b) Estimation of $\mathcal{H}_g = \{V_B, B_0, (Y_{01}, H_1, \Lambda_1, \delta_1), \cdots, (Y_{0,k-1}, H_{k-1}, \Lambda_{k-1}, \delta_{k-1}), (\Lambda_k, \delta_k)\}$

Le and Zidek (2006) showed that the predictive distributions derived through the integrated framework above are completely characterized by their hyperparameters, which are estimated by an empirical Bayes approach, that is, to estimate them by maximizing the marginal likelihood of all the measured responses (conditional on those hyperparameters) evaluated at their observed values. This procedure is referred to as type-II maximum likelihood estimation (type-II MLE). To estimate $\mathcal{H}_g$, the following procedure can be employed: Compute the hyperparameter values that maximize the marginal distribution $f(Y^{[g^o]}|\mathcal{H}_g)$, where $Y^{[g^o]} = \{Y^{[g_1^o]}, \cdots, Y^{[g_k^o]}\}$. The subscript g indicates that not all the hyperparameters are involved in this marginal distribution. The response matrix Y follows the GIW distribution specified by (1) and (2). The marginal distribution $f(Y^{[g^o]}|\mathcal{H}_g)$ can be written as

$$Y^{[g^o]}|\mathcal{H}_g \sim \prod_{j=1}^{k} t_{(n-m_j) \times g_j}(\mathcal{U}_o^{[j]}, \Phi_o^{[j]} \otimes \Psi_o^{[j]}, \delta_o^{[j]}), \tag{A1}$$

where $\mathcal{U}_o^{[j]} = \mathcal{U}_{(2)}^{[j]}$, $\Phi_o^{[j]} = A_{22}^{[j]}$, $\Psi_o^{[j]} = \frac{\Lambda_j}{\delta_j - g_j + 1}$, $\delta_o^{[j]} = \delta_j - g_j + 1$ with $\tilde{\mathcal{E}}^{[g_{j+1}, \cdots, g_k]} = Y^{[g_{j+1}, \cdots, g_k]} - XB_0^{[g_{j+1}, \cdots, g_k]}$, for $j = 1, \cdots, k-1$,

$$\begin{pmatrix} \mathcal{U}_{(1)}^{[j]} \\ \mathcal{U}_{(2)}^{[j]} \end{pmatrix} : \begin{pmatrix} m_j \times g_j \\ (n-m_j) \times g_j \end{pmatrix} = XB_0^{[g_j]} + \tilde{\mathcal{E}}^{[g_{j+1}, \cdots, g_k]} Y_{0j},$$

and

$$\begin{pmatrix} A_{11}^{[j]} & A_{12}^{[j]} \\ A_{21}^{[j]} & A_{22}^{[j]} \end{pmatrix} \quad : \quad \begin{pmatrix} m_j \times m_j & m_j \times (n-m_j) \\ (n-m_j) \times m_j & (n-m_j) \times (n-m_j) \end{pmatrix}$$

$$= I_n + XV_BX^\top + \tilde{\mathcal{E}}^{[g_{j+1}, \cdots, g_k]} H_j (\tilde{\mathcal{E}}^{[g_{j+1}, \cdots, g_k]})^\top.$$

Although $f(Y^{[g^o]}|\mathcal{H}_g)$ can be written as a matric-t distribution as in (A1), direct maximization of this marginal density presents a challenge. The EM algorithm helps circumvent it.

(c) Predictive distributions of missing data

By Theorem 10.1 of Le and Zidek (2006), it follows that

$$Y^{[g_k^m]}|Y^{[g^o]}, \mathcal{H}_g \sim t_{m_k \times g_k}(\mathcal{U}_{(m|g)}^{[k]}, \Phi_{(m|g)}^{[k]} \otimes \Psi_{(m|g)}^{[k]}, \delta_{(m|g)}^{[k]}), \tag{A2}$$

$$Y^{[g_j^m]} | Y^{[g_{j+1}^m, \cdots, g_k^m]}, Y^{[g^o]}, \mathcal{H}_g \sim t_{m_j \times g_j}(\mathcal{U}_{(m|g)}^{[j]}, \Phi_{(m|g)}^{[j]} \otimes \Psi_{(m|g)}^{[j]}, \delta_{(m|g)}^{[j]}), \tag{A3}$$

where $\mathcal{U}_{(m|g)}^{[j]} = \mathcal{U}_{(1)}^{[j]} + A_{12}^{[j]}(A_{22}^{[j]})^{-1}(Y^{[g_j^o]} - \mathcal{U}_{(2)}^{[j]})$, $\Phi_{(m|g)}^{[j]} = \frac{\delta_j - g_j + 1}{\delta_j - g_j + n - m_j + 1}$ $[A_{11}^{[j]} - A_{12}^{[j]}(A_{22}^{[j]})^{-1}A_{21}^{[j]}]$, $\Psi_{(m|g)}^{[j]} = \frac{1}{\delta_j - g_j + 1}[\Lambda_j + (Y^{[g_j^o]} - \mathcal{U}_{(2)}^{[j]})^\top (A_{22}^{[j]})^{-1}(Y^{[g_j^o]} - \mathcal{U}_{(2)}^{[j]})]$ and $\delta_{(m|g)}^{[j]} = \delta_j - g_j + n - m_j + 1$.

References

1. Wikle, C.K.; Berliner, L.M.; Cressie, N. Hierachichical Bayesian space-time models. *Environ. Ecol. Stat.* **1998**, *5*, 117–154. [CrossRef]
2. Le, N.D.; Sun, W.; Zidek, J.V. Spatial prediction and temporal backcasting for environmental fields having monotone data patterns. *Can. J. Stat.* **2001**, *29*, 529–554. [CrossRef]
3. Jin, B.; Wu, Y.; Chan, E. Hierarchical Bayesian spatial-temporal modeling of regional ozone concentrations and respecitve network design. *J. Environ. Stat.* **2012**, *3*, 1–32.
4. Diggle, P.J.; Ribeiro, P.J., Jr. *Model-Based Geostatistics*; Springer: Berlin/Heidelberg, Germany, 2007.
5. Qin, S.; Sun, B.; Wu, Y.; Fu, Y. Generalized least-squares in dimension expansion method for nonstationary processes. *Environmetrics* **2021**, *32*, e2684. [CrossRef]
6. Cressie, N. *Statistics for Spatial Data, Revised Edition*; Wiley: New York, NY, USA, 2015.
7. Xue, X.T.; Guy, B.; Xing, L.; Pierre, F.; Claire, G.; Philip, R. The Impact of High Altitude Aircraft on the Ozone Layer in the Stratosphere. *J. Atmos. Chem.* **1994**, *18*, 103–128.
8. Bornn, L.; Shaddick, G.; Zidek, J.V. Modeling non-stationary processes through dimension expansion. *J. Am. Stat. Assoc.* **2012**, *107*, 281–289. [CrossRef]
9. Journel, A.G.; Huijbregts, C.J. *Mining Geostatistics*; Academic: London, UK, 1978.
10. Wabba, G.; Wendelberger, J. Some new mathematical methods for variational objective analysis using splines and cross-validation. *Mon. Weather Rev.* **1980**, *108*, 1122–1143. [CrossRef]
11. Jaynes, E.T. *Information Theory and Statistical Mechanics, Statistical Physics*, 3rd ed.; Ford, K.W., Ed.; Benjamin: New York, NY, USA, 1963.
12. Le, N.D.; Zidek, J.V. *Statistical Analysis of Environmental Space-Time Processes*; Springer: New York, NY, USA, 2006.