Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Multivariate Approach for Modeling Spatio-Temporal Agrometeorological Variables

Sandra De Iaco[1] | Claudia Cappello[1,2] | Monica Palma[1] | Klaus Nordhausen[3]

[1]DES-Sect. of Mathematics and Statistics, University of Salento, Lecce, Italy | [2]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria | [3]Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

Correspondence: Sandra De Iaco (sandra.deiaco@unisalento.it)

Received: 12 May 2023 | **Revised:** 4 October 2024 | **Accepted:** 6 November 2024

Funding: This work was supported by National Biodiversity Future Center Austrian Science Fund and Research Council of Finland.

Keywords: multiple correlation | space–time coregionalization model | space–time prediction cokriging

ABSTRACT

One of the main issues facing agrometeorological studies involves measuring and modeling the evolution of different environmental variables over time; this often requires a dense monitoring network. Spatio-temporal geostatistics has the potential to provide techniques and tools to estimate the spatio-temporal multiple covariance function and define an appropriate multivariate correlation function capable of reliable predictions. This paper presents a spatio-temporal multivariate geostatistical modeling approach based on the joint diagonalization of the empirical covariance matrix evaluated at different spatio-temporal lags. The possibility to consider a reduced number of uncorrelated variables (lower than the number of observed variables) and separately model the spatio-temporal evolution of these uncorrelated components represents a substantial simplification for multivariate modeling. A space–time linear coregionalization model (ST-LCM) with appropriate parametric models for the latent components was fitted to the matrix-valued covariance function estimated for five relevant agrometeorological variables, including evapotranspiration, minimum and maximum humidity, maximum temperature, and precipitation. The analyses highlight how to identify space–time components and choose the corresponding model by evaluating some characteristics of these components, such as symmetry, separability, and type of non-separability. The predictive results of this multivariate study will be of interest for agriculture, in particular for addressing drought emergencies.

1 | Introduction

Modern agrometeorology and related scientific fields require appropriate techniques to exploit the information given by multivariate spatial or spatio-temporal data. To achieve this, a significant role is played by multivariate geostatistics tools, which are used to describe and interpolate the studied variables in an accurate and reliable manner (Alidoost, Stein, and Su 2018; Gnann et al. 2018; Silva and Deutsch 2018). In this context, research efforts have focused on the estimation and modeling of the matrix-valued covariance function, which explains the direct and cross-linear dependence in space or space–time among the variables.

Multivariate spatio-temporal data modeling has been developed since the early nineties (Berrocal, Gelfand, and Holland 2010; Calculli et al. 2015; Choi et al. 2009; De Iaco, Myers, and Posa 2003; Fassò and Finazzi 2011; Goovearts and Sonnet 1993; Ip and Li 2016, 2017a, 2017b; Krupskii and Genton 2017; Rouhani and Wackernagel 1990). Two recent papers of Chen, Genton, and

Sun (2021); Porcu, Furrer, and Nychka (2021) proposed a comprehensive review on spatio-temporal covariance models, as well as a comparative analysis presented in the study by Otto et al. (2024) and new classes of multivariate spatio-temporal models extending the Gneiting family were discussed in the studies by Allard, Clarotto, and Emery (2022) and Dörr and Schlather (2023). However, the linear coregionalization model (LCM) developed in space and in space–time (ST-LCM) was the first to be sufficiently computationally flexible to be applied in a large range of fields of application (Babak and Deutsch 2009; Bevilacqua, Hering, and Porcu 2015; Emery 2010; Genton and Kleiber 2015; Gneiting, Kleiber, and Schlather 2010; Li, Genton, and Sherman 2008).

It is worth highlighting that other approaches to model a matrix-valued covariance function were provided in the literature, as substitute ways to the linear coregionalization-based modeling. Evidently each of them presents its own advantages and drawbacks. For instance, in the LCM the smoothness of any component of the multivariate random field is that of the roughest underlying univariate random field (the latent component), as recalled by Porcu, Furrer, and Nychka (2021). On the other hand, by comparing the bivariate LCM with the bivariate Matérn class developed by Genton and Kleiber (2015) and Bevilacqua, Hering, and Porcu (2015), a large flexibility of the former model with respect to the latter one, in terms of the permissible range for the co-located correlation coefficient, was pointed out. Lately, a recent contribution based on the blind source separation technique offered an alternative model a multivariate space–time random field (Muehlmann, De Iaco, and Nordhausen 2023). In this case, the latent independent variables are estimated directly from the study data.

This paper analyses spatio-temporal dependence among five agrometeorological indicators, that is, evapotranspiration, minimum and maximum humidity, maximum temperature, and precipitation, with the objective of offering insights into their joint evolution while constructing a valuable spatio-temporal geostatistical model for predicting evapotranspiration levels (primary variable) through the use of the ST-LCM. Although data quality and resolution depend on the density of the meteorological monitoring network and on the satellite maps used, it is important to produce a model for these indicators and to generate spatio-temporal maps of their evolution, which addresses the needs of ecologists and agronomists. In particular, since agrometeorology is most commonly applied in agriculture, evapotranspiration is a key variable: it is among the primary factors influencing many natural processes, such as plant growth, the life cycle of fauna in the soil, their distribution, and their abundance (Hurtado-Uria et al. 2013).

A first attempt to use geostatistical interpolation methods to obtain spatial estimates of potential evapotranspiration values can be found in the study by Martínez-Cob (1996), where annual reference evapotranspiration was interpolated over Aragón, NE Spain. In recent years, a number of spatial interpolation methods for evapotranspiration mapping have been adopted, such as the inverse distance weighted interpolation and kriging methods used by Hodam (2017), as well as simple kriging and simple cokriging in the study by Gentilucci et al. (2021). More recently, thanks to improvements in information technology, machine learning methods, such as artificial neural networks (Abrishami, Sepaskhah, and Shahrokhnia 2019) and support vector machines (Chia, Huang, and Koo 2020) have also been applied to estimate evapotranspiration. Despite numerous attempts to estimate evapotranspiration over spatial domains, limited research has been conducted to predict evapotranspiration within a spatio-temporal context. Furthermore, predicting evapotranspiration in a multivariate spatio-temporal framework by incorporating secondary information from correlated meteorological variables remains an under-explored area that warrants further investigation.

A valuable contribution to the multivariate spatio-temporal analysis of evapotranspiration data was made by De Iaco, Palma, and Posa (2019a), where the relationship between evapotranspiration and air temperature were modeled without providing predictions for the variable of interest. The present paper provides an extension of the study referred to five significant agrometeorological indicators, namely potential evapotranspiration, maximum temperature, minimum and maximum humidity, and precipitation recorded over Veneto Region (North-eastern Italy), the application of the ST-LCM is developed for prediction purposes. Apart from the practical importance of this study, it is noteworthy that this is the first demonstration of modeling the matrix-valued covariance structure using the ST-LCM with more than two or three variables. This advancement is made possible through an approach based on the joint diagonalization of sample covariance matrices at different lags and enables analysts to overcome the complexity of fitting the ST-LCM, particularly when the number of variables to be analyzed increases, because it does not require modeling all direct and cross-covariance functions. Furthermore, through the aforementioned procedure, analysts can more easily identify the basic components of the ST-LCM and model each of which has its own empirical characteristics; this may necessitate the use of different classes of covariance models featuring various types of non-separability (De Iaco, Palma, and Posa 2019a).

The paper is structured in six sections. The following section presents the multivariate spatio-temporal geostatistical framework and introduces the ST-LCM selection procedure (Section 2). Then, a multivariate study is presented for five agrometeorological indicators collected weekly over a 23-year span (2000–2022) at various monitoring stations located in the Veneto Region (North-eastern Italy) and the fitted ST-LCM is described and validated (Section 3). In Section 4, a comparison of the fitted ST-LCM with respect to the Gneiting–Matérn covariance model proposed by Allard, Clarotto, and Emery (2022) is discussed and the respective predictive performances are evaluated. Finally, spatio-temporal predictions for the primary variable under study (the evapotranspiration) are produced (Section 5) and some comments on the usefulness of these results for supporting actions to prevent drought emergency scenarios are considered (Section 6).

2 | The ST-LCM and the Selection Procedure

Observations of a given vector of variables taken at different sample locations and time points can be assumed to reflect a multivariate space–time random function (MSTRF) $\{\mathbf{X}(\mathbf{s},t),(\mathbf{s},t)\in D\times T\subseteq\mathbb{R}^d\times\mathbb{R}\}$, where $\mathbf{X}(\mathbf{s},t)=\left[X_1(\mathbf{s},t),\ldots,X_m(\mathbf{s},t)\right]^T$, $m\geq 2$. Under the assumption of second-order stationarity, the first and second order moments of the MSTRF are defined as follows:

$$\boldsymbol{\mu} = \left[\mu_1, \ldots, \mu_m\right]^T, \tag{1}$$

where μ_i is the expected value of $X_i(\mathbf{s}, t), \forall(\mathbf{s}, t)$, $i = 1, \ldots, m$, and

$$\mathbf{C}(\mathbf{u}, v) = \left[C_{ij}(\mathbf{u}, v)\right], \tag{2}$$

with

- $(\mathbf{u}, v) \in \mathbb{R}^d \times \mathbb{R}$, with $\mathbf{u} = \left(\mathbf{s} - \mathbf{s}'\right)$ and $v = \left(t - t'\right)$, for any $(\mathbf{s}, t)$ and $\left(\mathbf{s}', t'\right) \in D \times T$;
- $C_{ij}(\mathbf{u}, v) = E\left[\left(X_i(\mathbf{s} + \mathbf{u}, t + v) \cdot X_j(\mathbf{s}, t)\right)\right] - \mu_i \mu_j$, which is
  - the cross-covariance function for any X_i and X_j, $i, j = 1, \ldots, m$, if $i \neq j$,
  - the direct covariance function of X_i if $i = j, i, j = 1, \ldots, m$.

In space–time multivariate geostatistics, several applications make use of an ST-LCM modeling approach, where the model is constructed by the linear combination of the $L \leq m$ basic scalar covariance functions.

In particular, the covariance matrix $\mathbf{C}$ is modeled as follows:

$$\mathbf{C}(\mathbf{u}, v) = \sum_{l=1}^{L} \mathbf{B}_l \; c_l(\mathbf{u}, v), \tag{3}$$

where $c_l(\mathbf{u}, v)$ are the above-mentioned basic scalar covariances associated with the hidden variables of the random field $\mathbf{X}$ and $\mathbf{B}_l = \left[b_{ij}^l\right]$, $l = 1, \ldots, L$, are the $(m \times m)$ positive definite matrices of coregionalization. Note that the structure of the above covariance model can be derived from the assumption that the MSTRF is defined as a linear combination of uncorrelated latent components. For further details, the readers can refer to Myers (1995) and De Iaco, Palma, and Posa (2005).

The model in Equation (3) can be fitted on the basis of the steps given below:

I. Selection of the uncorrelated components and computation of the empirical basic covariance functions;

II. Modeling the empirical basic covariances using appropriate model classes (according with the empirical characteristics of each basic component);

III. Computation of admissible coregionalization matrices.

The first step begins with the estimation of the matrix-valued covariance function, that is, the m direct covariances and $m(m-1)/2$ symmetric cross-covariances for K space–time lags. These are selected by users according to the geometry of the sample points. Thus, a symmetric $(m \times m)$ matrix $\widehat{\mathbf{C}}(\mathbf{u}, v)_k = \left[\widehat{C}_{ij}(\mathbf{u}, v)_k\right]$ is obtained for each lag $(\mathbf{u}, v)_k = \left(\mathbf{u}_k, v_k\right)$, with $k = 1, \ldots, K$, taking into account all observations at any pair of sample points $(\mathbf{s}, t)$ and $\left(\mathbf{s}', t'\right)$ such that $\left(\mathbf{s} - \mathbf{s}'\right) \approx \mathbf{u}_k$ and $\left(t - t'\right) \approx v_k$, with a spatial separation tolerance $Tol\left(\mathbf{u}_k\right)$ and a temporal tolerance $Tol\left(v_k\right)$ (Wackernagel 2003).

After estimating the matrix-valued covariance function, joint diagonalization is applied to the sample covariance matrix evaluated at different space–time lags with the aim of detecting the basic components as follows:

$$\boldsymbol{\Psi}\, \widehat{\mathbf{C}}(\mathbf{u}, v)_k\, \boldsymbol{\Psi}^T = \boldsymbol{\Lambda}(\mathbf{u}, v)_k, \qquad k = 1, \ldots, K, \tag{4}$$

where $\boldsymbol{\Psi}$ is a $(m \times m)$ orthogonal matrix and $\boldsymbol{\Lambda}_k$ are the diagonal $(m \times m)$ matrices. Joint diagonalization with respect to the lags implies that the matrix $\boldsymbol{\Psi}$ does not depend on the lags. Many algorithms exist for the purpose of joint diagonalization; see Illner et al. (2015) for an overview of these approaches. In this study, we will use an algorithm based on Jacobi rotations (Cardoso and Souloumiac 1996), which is available, for example, in the `R` package `JADE` (Miettinen, Nordhausen, and Taskinen 2017). Then, given the $(m \times m)$ matrix $\boldsymbol{\Psi}$, the latent components are obtained as a matrix multiplication between the $(n \times m)$ data matrix $\mathbf{x}$ and $\boldsymbol{\Psi}$:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\Psi}. \tag{5}$$

Considering the uncorrelated variables in Equation (5), the corresponding m spatio-temporal sample covariance functions are computed and only the $L \leq m$ basic components $\widehat{c}_l$, $l = 1, \ldots, L$, are selected as those which exhibit distinct spatio-temporal scales of variability (corresponding to the lag at which the surface decays). In particular, by the joint visual inspection of the covariance surfaces of the latent components and the corresponding marginals in space and in time, the distinct lags where the surfaces decay can be detected and consequently the scales of spatio-temporal variability can be fixed. Then, the selected basic components are used in the following step of the procedure, where they are modeled.

A reasonable class of models to be fitted to each component can be assessed, according to characteristics such as full symmetry $\left(c_l(\mathbf{u}, v) = c_l(\mathbf{u}, -v) = c_l(-\mathbf{u}, v)\right)$ and separability $(c_l(\mathbf{u}, v) = c_l(\mathbf{u}, 0)c_l(\mathbf{0}, v)/c_l(\mathbf{0}, 0))$, which may be satisfied by sample covariance surfaces. If the separability assumption is not likely, the type of non-separability can be studied through the computation of non-separability ratios of the sample, as in the study by De Iaco and Posa (2013):

$$\widehat{r}_l(\mathbf{u}, v) = \widehat{c}_l(\mathbf{0}, 0)\frac{\widehat{c}_l(\mathbf{u}, v)}{\widehat{c}_l(\mathbf{u}, 0)\widehat{c}_l(\mathbf{0}, v)}, \qquad l = 1, \ldots, L, \tag{6}$$

which imply:

- uniform positive non-separability, as in the case of the ratios in Equation (6), which are much greater than 1 for the fixed lags;
- uniform negative non-separability, as in the case of the ratios in Equation (6), which are much smaller than 1 for the fixed lags;
- non uniform non-separability, in all other cases.

As comprehensively detailed in De Iaco, Posa, and Myers (2013), spatio-temporal interactions determine the type of non-separability, in other words, the discrepancy between non-separable and separable covariance functions (i.e., the product of the spatial and temporal marginals) establishes the type of non-separability. Box plot representations of ratios in Equation (6), grouped by spatial and temporal lags, enable the easy identification of the type of non-separability exhibited by each basic component. However, in the study by Cappello, De Iaco, and Palma (2022), some statistical tests were also proposed

to determine the statistical significance of hypotheses using the above-mentioned characteristics.

The coregionalization matrices $\mathbf{B}_l, l = 1, \ldots, L$, of the model in Equation (3) are thus estimated. Starting from the $m(m+1)/2$ direct and cross-covariances $\widehat{C}_{ij}(\mathbf{u}, v)$, $i, j = 1, \ldots, m$, the elements b_{ij}^l of $\mathbf{B}_l, l = 1, \ldots, L$, correspond to the ratio between the contributions of $\widehat{C}_{ij}$ at the lth scale of variability, by $c_l(\mathbf{0}, 0)$, that is,

$$b_{ij}^l = \frac{\widehat{C}_{ij}(\mathbf{u}, v)_{l-1} - \widehat{C}_{ij}(\mathbf{u}, v)_l}{c_l(\mathbf{0}, 0)}, \qquad l = 1, \ldots, L, \tag{7}$$

where $\widehat{C}_{ij}(\mathbf{u}, v)_0 = \widehat{C}_{ij}(\mathbf{0}, 0)$, with $i, j = 1, \ldots, m, i \leq j$.

The positive definiteness condition of the matrices $\mathbf{B}_l, l = 1, \ldots, L$, is verified by confirming that their eigenvalues are non-negative. Then, by performing spectral decomposition of these matrices,

$$\mathbf{B}_l = \mathbf{V}_l \mathbf{\Lambda}_l \mathbf{V}_l^T, \qquad l = 1, \ldots, L,$$

and computing their eigenvector matrix $\mathbf{V}_l$ and the diagonal matrix of the eigenvalues $\mathbf{\Lambda}_l$, one can confirm whether there are any negative eigenvalues and set these to zero. In this case, the transformed coregionalization matrix $\mathbf{B}_l'$ is derived through the following expression:

$$\mathbf{B}_l' = \mathbf{V}_l \mathbf{\Lambda}_l' \mathbf{V}_l^T, \qquad l = 1, \ldots, L, \tag{8}$$

where the diagonal matrix of the eigenvalues $\mathbf{\Lambda}_l'$ is modified with respect to the original $\mathbf{\Lambda}_l$ since zeros are used in place of negative eigenvalues.

3 | Space–Time Multivariate Agrometeorological Indicators

In the present analysis, the space–time behavior of evapotranspiration, together with other variables, that is, minimum and maximum humidity, maximum temperature, and precipitation, will be jointly investigated. The matrix-valued covariance of the above mentioned environmental variables will be estimated and modeled using the ST-LCM. It is worth noting that (i) the number of uncorrelated components should be reduced with respect to the number of the original variables and (ii) models for the hidden components will be selected on the basis of their symmetry and separability characteristics. This multivariate spatio-temporal analysis will offer an opportunity to better understand the way in which the selected agrometeorological indicators jointly evolve in space and time, as well as to select a model of coregionalization which can be applied for predictive purposes and for better understanding evapotranspiration.

3.1 | Data Set and Study Area

The Veneto Region, located in the northeastern part of Italy, was chosen as the study area. This region extends for approximately 18,400 km^2 mostly occupied by the Venetian plain (55%). A portion of the Po Valley is included in the Veneto Region; this area is bounded by the pre-Alpine hills to the northwest, the Adriatic Sea to the southeast, the Tagliamento river to the northeast, and the Po river to the south.

In 2022, the Italian Government declared a state of emergency (environmental crisis) in the Veneto Region as a result of the lack of rain and warmer temperatures during the previous 3 years. The water crisis in Veneto is the result of complex interactions between multiple agrometeorological variables, hence it is important to analyze the spatio-temporal relationships among five variables. In particular, the dataset under study consists of weekly averages of evapotranspiration (ET_0, in mm), minimum (H_m) and maximum (H_M) humidity (%), maximum temperature (T_M, in°C), and weekly precipitation (P, in mm) measured at 72 stations managed by the Veneto weather monitoring network (Figure 1). Measurements for the previous 23 years (2000–2022) were used, totaling 1180 weekly temporal values (from the first week in 2000 to the fourteenth week in 2022) for each location.

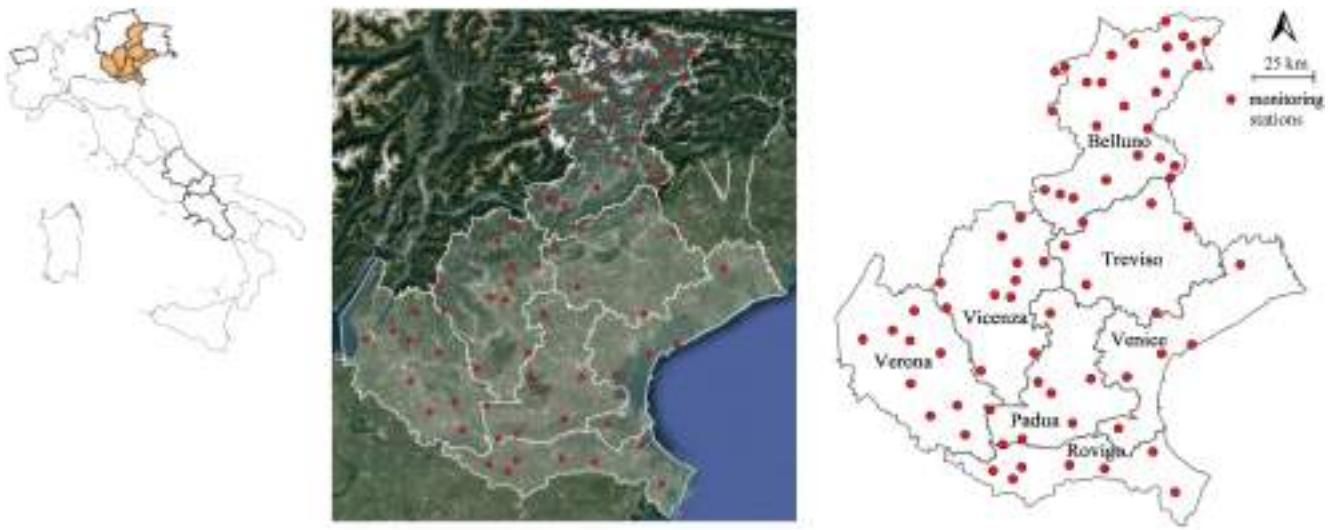


FIGURE 1 | Left panel: map of Italy showing the study area (indicated in orange). Middle and right panels: the 72 sampling locations in the study area.

The data were recorded by a network of well-distributed meteorological stations belonging to the Venetian Regional Agency for Environmental Protection (https://www.arpa.veneto.it/) and have been recorded at regular time intervals; missing values are sparse for all five analyzed variables.

3.2 | Exploratory Data Analysis

Figure 2 provides a graphical representation of the spatial profile of the data and in particular illustrates the color maps of the average values for the first week (winter season) and the 28th week (summer season) for the 23-year span considered. Lower values of evapotranspiration and maximum temperature have been registered in the northern part of the region in winter and in the northwest area during the summer. These parts of the region are characterized by a wet climate with higher humidity and precipitation, especially during the winter.

To evaluate temporal changes, Figure 3 shows the box plots of the weekly values registered at the 72 meteorological stations studied, highlighting seasonal behavior. In particular, evapotranspiration, maximum temperature, and maximum humidity present higher values from May to August and lower values during the other periods. For evapotranspiration, temperature, and maximum humidity, the highest mean values occur in July and August, whereas the lowest occur in January and December. On the other hand, minimum humidity and precipitation are characterized by an opposite temporal evolution throughout the year, with a decrease during spring and the summer, and an increase during winter. The highest (lowest) mean values have been recorded during the period October to January (June to July) for minimum humidity and in October (June) for precipitation, whereas the lowest mean values of minimum humidity occurred in June and July. Note that for most weeks the rainfall levels were equal to zero and the distribution of precipitation was highly skewed; hence, the logarithmic transformation of the values (after adding 1) has been considered.

Seasonal effect, that is, temporal non stationarity has to be taken into account by removing weekly averages. However, Figure 3 shows that the variance also has seasonal effects. Moreover, Veneto is a highly non-stationary region, with the presence of the Alps in the North and the Adriatic Sea to the East. Thus, spatial non-stationarity has to be also taken into account. The overall statistical model is reasonably $\mathbf{Y}(\mathbf{s},t) = \boldsymbol{\mu}(\mathbf{s},t) + \sigma(\mathbf{s},t)\mathbf{X}(\mathbf{s},t)$, where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are vectors of non stationary means and standard deviations, respectively. Then the geostatistical analysis has been conducted on $\mathbf{X}(\mathbf{s},t)$.

As illustrated below, the spatio-temporal direct and cross covariance functions of the standardized residuals of evapotranspiration, maximum temperature, humidity (maximum and minimum), and the logarithm of precipitation have been analyzed and an appropriate ST-LCM has been detected according to the fitting procedure described in Section 2.

3.3 | Selection of Basic Latent Components

To identify a multivariate model describing spatio-temporal correlations among the five variables studied, a preliminary estimate of the direct and cross-covariances is required.

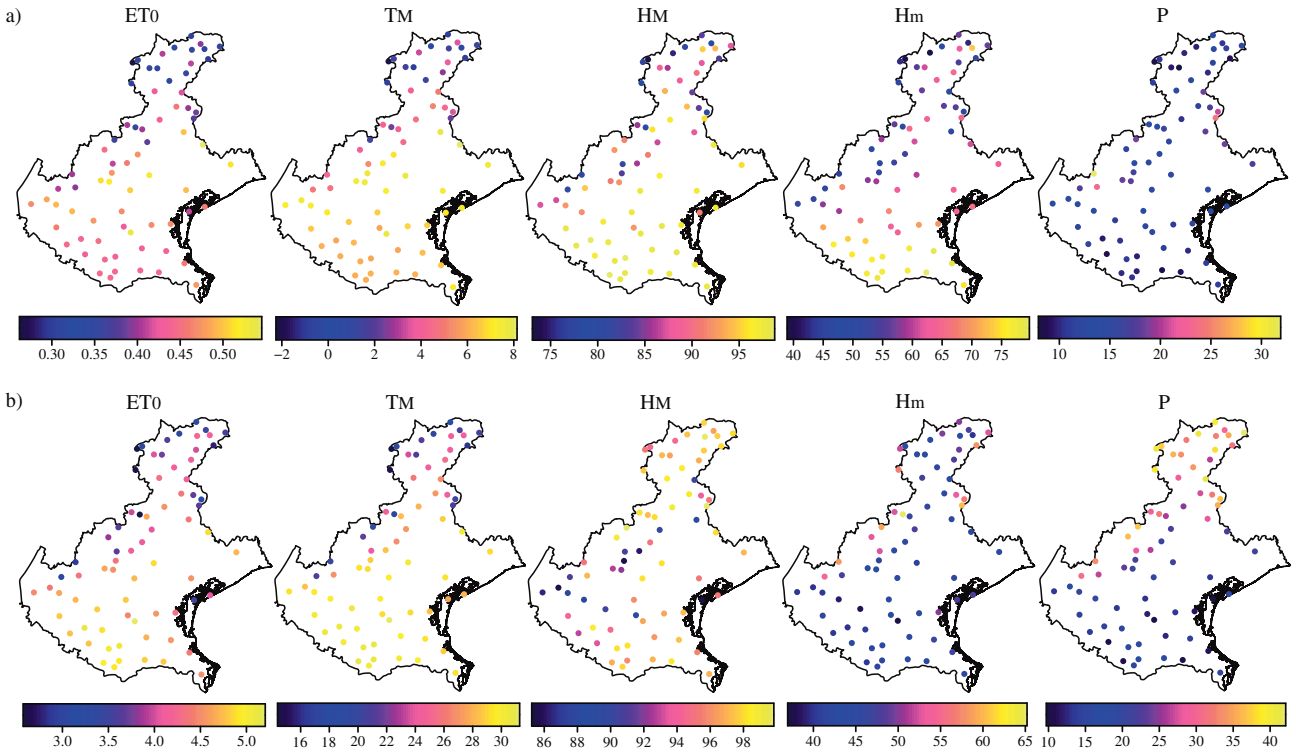
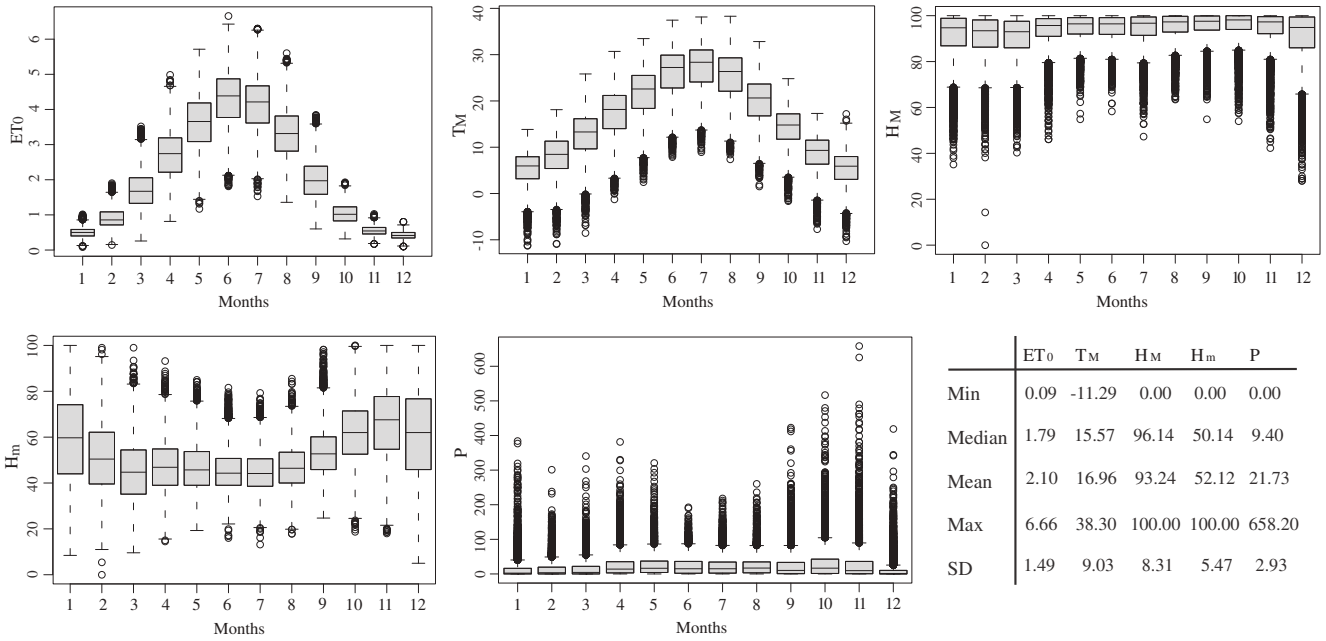


FIGURE 2 | Maps visualizing the values of evapotranspiration (ET_0), maximum temperature (T_M), maximum and minimum humidity (H_M and H_m), and precipitation (P) during (a) the first week (winter season) and (b) the 28th week (summer season) for all measurement stations in this study.



	ET_0	T_M	H_M	H_m	P
Min	0.09	-11.29	0.00	0.00	0.00
Median	1.79	15.57	96.14	50.14	9.40
Mean	2.10	16.96	93.24	52.12	21.73
Max	6.66	38.30	100.00	100.00	658.20
SD	1.49	9.03	8.31	5.47	2.93

FIGURE 3 | Box plots showing the five environmental variables grouped by month and some descriptive statistics (all stations have been pooled).

Firstly, an appropriate number of spatio-temporal lags has been set up by considering the spatio-temporal geometry of the points. In particular, eight lags in space (with a lag spacing of 5 km) and seven in time (with a lag interval of 1 week) have been fixed according to the geometry of the sample points and covariance decay rate, hence $K = 56$ lags in space–time.

In Figure 4, the 3D plots for the fifteen spatio-temporal sample covariances (five direct and ten symmetric cross-covariances) are provided for the deseasonalized variables. The sample direct covariance surfaces show a well-structured positive correlation for all lags; on the other hand, the sample cross-covariance surfaces for $ET_0 - H_m$, $ET_0 - \ln P$, $T_M - H_m$, and $T_M - \ln P$ highlight the presence of negative correlations in space–time.

Then, the joint diagonalization (Cardoso and Souloumiac 1996) of the empirical covariance matrices, that is, 56 matrices with dimensions of (5×5), has been carried out to identify five uncorrelated latent components.

The goodness of the diagonalization results was evaluated by computing the following relative index, which was constructed to compare the diagonal versus off diagonal entries of $\mathbf{\Lambda}_k$.

$$\psi_k = \frac{\sum_{i=1}^{m}\sum_{j=1;j\neq i}^{m}\lambda_{ij,k}^2}{\sum_{i=1}^{m}\lambda_{ii,k}^2}, \qquad k = 1, \ldots, K, \tag{9}$$

where $\lambda_{ij,k}, i, j = 1, \ldots, m, k = 1, \ldots, K$, are the entries of $\mathbf{\Lambda}_k$ at the K lags chosen by the analyst. At values closer to zero, the performance of the diagonalization is improved. An analysis of the index values for each spatio-temporal lag indicates that 75% of the computed indexes are less than 0.164 (median value of 0.062 and standard deviation of 0.183). Hence, these results confirm that most of the 56 diagonalized matrices are close to diagonal matrices, such that the uncorrelated components can be obtained as clarified in Equation (5).

Then, visual inspection of the covariance surfaces of the five uncorrelated components shows that only two of them ($L = 2$) exhibit different levels of variability in space–time (Figure 5), that is:

1. the small distance component, which decays at 14.5 km and 2 weeks;
2. the long distance component, which decays at 25 km and 4 weeks.

The remaining three uncorrelated components obtained from the diagonalization process exhibit the same variability in space and time, hence they have been discarded because they carry no spatio-temporal information and can be considered as white noise. In particular, it is worth pointing out that a joint visual inspection of the surfaces of the basic components allows the analyst to detect the $L \leq m$ distinct basic components characterized by different scales of variability; in other words, by the 3D plots of all the covariance surfaces and the respective marginals in space and in time it is easy to find the different lags where the surfaces of the basic covariances decay, that is, the scales of spatio-temporal variability.

In the following, the procedure for fitting the ST-LCM by considering the components retained will be discussed. Note that the use of only two basic components represents a significant simplification of the multivariate modeling stage.

3.4 | Modeling Basic Components

The sample spatio-temporal covariance surfaces of the retained basic components, as well as at the corresponding purely spatial and purely temporal covariances (Figure 5) have shown a linear behavior at the origin. Note that the sample covariance functions of all the variables under study have presented the same behavior

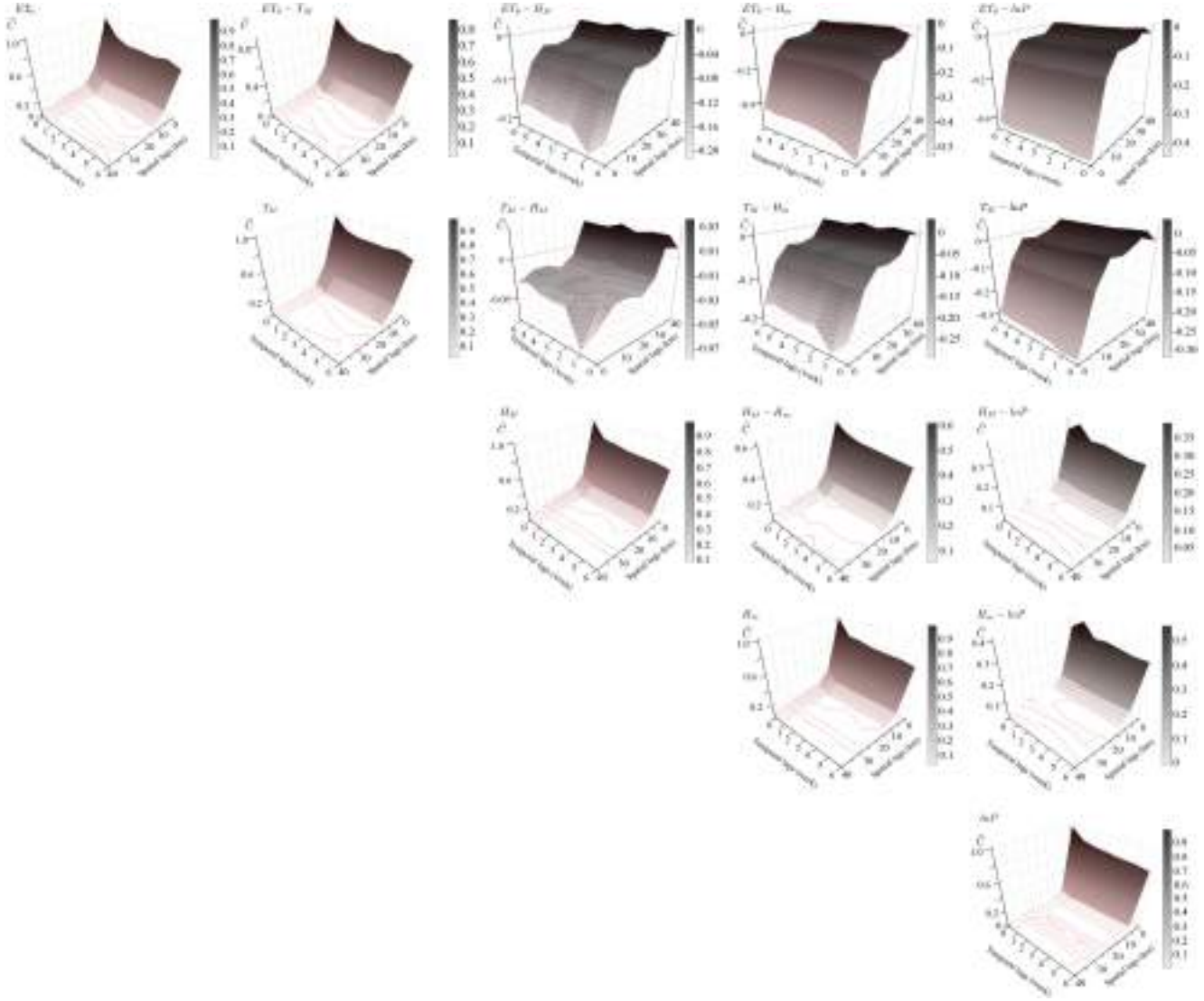


FIGURE 4 | 3D plots of the spatio-temporal direct (diagonal) and cross (off diagonal) covariance surfaces and their contour lines of ET_0, T_M, H_M, H_m, and lnP standardized residuals, for fixed spatial and temporal lags.

near the origin (Figure 4), thus in terms of smoothness there will be no impact on the posterior analysis.

In order to choose a model reflecting the empirical characteristics of each uncorrelated component, tests of symmetry and separability, which are based on the asymptotic joint normality of the sample space–time covariance estimators (Li, Genton, and Sherman 2007), have been performed according to the procedure proposed by Cappello, De Iaco, and Posa (2018, 2020). In the same studies, all details regarding the test statistics, denoted with TS_1 and TS_2, and the corresponding probability distribution are given.

For both the small and the long distance components, the full symmetry hypothesis has been tested on a set of spatio-temporal lags with a strong correlation. In particular, ten pairs of spatial locations, at distances of 6 and 12 km, and four temporal lags (± 1 and ± 2 weeks) have been considered for TS_1.

At the 5% significance level, there is no evidence for rejecting the null hypothesis of full symmetry, since the test statistics were 2.0081 ($p = 0.999$, df = 20) and 3.109 ($p = 0.999$, $df = 20$) for the small and the long components, respectively. Next, the separability condition has been tested using the same spatial and temporal lags previously fixed to test the symmetry, with the exception of the negative temporal lags. At the 5% significance level, the separability assumption is rejected for both the small and long distance components, since the test statistics were 41.02 ($p = 0.004$, $df = 20$) and 39.43 ($p = 0.006$, $df = 20$), respectively.

The rejection of the hypothesis of separability implies that a non-separable class of models is preferable for modeling the retained uncorrelated components. Hence, the type of non-separability used must be carefully evaluated by (1) computing the sample non-separability ratios Equation (6) and (2) displaying the estimated ratios using box and whiskers plots classified for spatial and temporal lags.

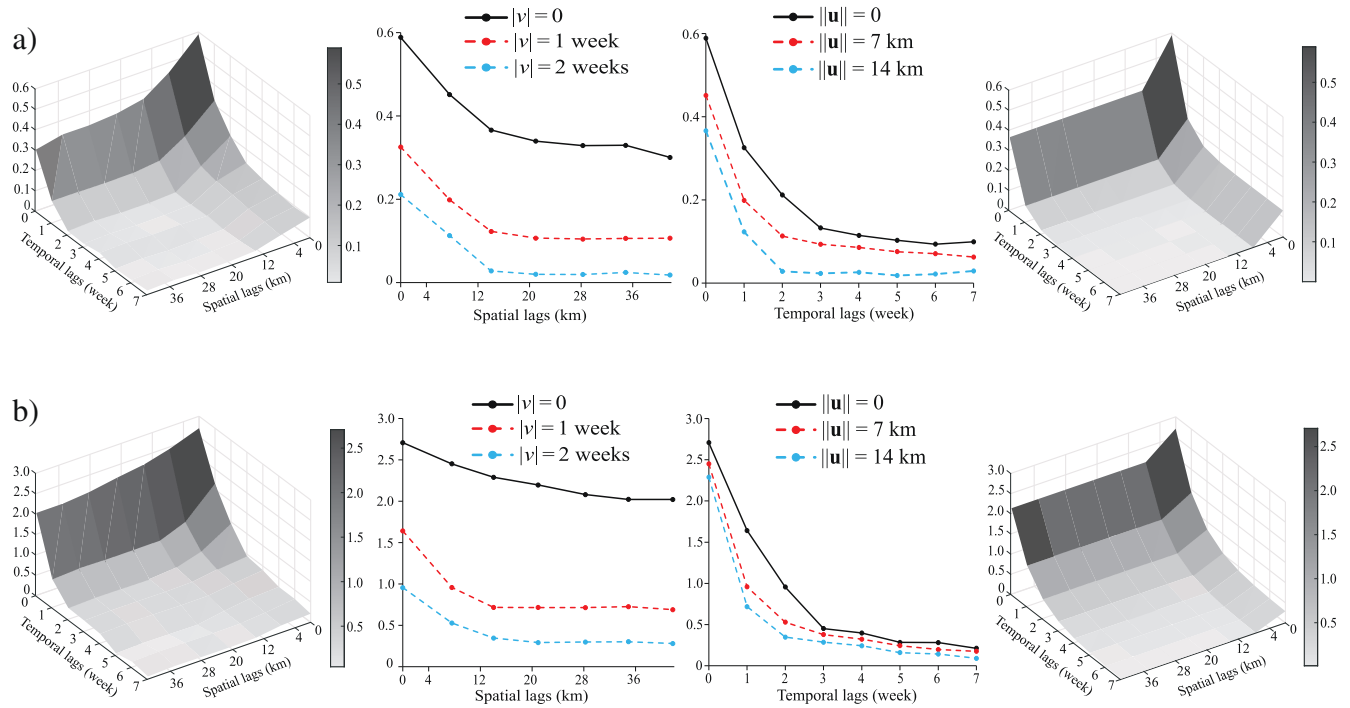


FIGURE 5 | Sample covariance surfaces (on the left), spatial profiles versus $|v|$ and temporal profiles versus $\|\mathbf{u}\|$ (in the middle), theoretical covariance surfaces (on the right), for: (a) the small distance basic component at 14.5 km and 2 weeks, (b) the long distance basic component at 25 km and 4 weeks.

Consequently, the non-separability index has been computed for all spatio-temporal lags and then classified with respect to the spatial and temporal profile. Through an accurate graphical evaluation of the plots shown in Figure 6, it is evident that the ratios are smaller than, or equal to, 1 for all spatio-temporal lags. Hence, for the short and long distance components, uniform negative non-separability can be reasonably assumed and tests of the type of non-separability have been conducted on the right tail of the standard normal distribution for both components, as detailed in the study by Cappello, De Iaco, and Posa (2018). For this aim, two spatial lags (i.e., 6 and 12 km) and two temporal lags (i.e., 1 and 2 weeks) have been used. The null hypothesis of negative non-separability is not rejected at the 5% significance level: the test statistics have been equal to -0.812 ($p = 0.792$) and -0.776 ($p = 0.781$) for the short and long distance components, respectively.

The above results provide statistical support for the use of a fully symmetric and non-separable, that is, uniformly negative non-separable, space–time covariance function to model uncorrelated components at small and large scales of variability. In this case, both the classes of the product-sum and integrated product-sum covariance models can be suitable to describe the main characteristics of the covariance functions estimated for the two uncorrelated components (De Iaco and Posa 2013; De Iaco, Palma, and Posa 2016). However small differences in the models which are characterized by the same a priori properties in terms of behavior near the origin and at infinity as well as in terms of type of non-separability, do not impact on the predictions of the variable(s) of interest, thus the product-sum class of covariance function has been chosen in the present case study, that is,

$$c_l(\mathbf{u}, v) = k_l\left[k_{1_l} C_{s_l}(\mathbf{u}) C_{t_l}(v) + k_{2_l} C_{s_l}(\mathbf{u}) + k_{3_l} C_{t_l}(v)\right], \qquad l = 1, 2, \tag{10}$$

where k_l is the sill value, $k_{1_l} > 0, k_{2_l} \geq 0, k_{3_l} \geq 0$, (parameterized by constraining that $\left(k_{1_l} + k_{2_l} + k_{3_l}\right) = 1$ with $c_l(\mathbf{0}, 0) = k_l$) C_{s_l} is the spatial covariance in $\mathbb{R}^d$ and C_{t_l} is the temporal covariance in $\mathbb{R}$, $l = 1, 2$. In the following the unit sill models for the basic components are denoted with $c_l^*(\mathbf{u}, v) = k_{1_l} C_{s_l}(\mathbf{u}) C_{t_l}(v) + k_{2_l} C_{s_l}(\mathbf{u}) + k_{3_l} C_{t_l}(v)$. Note that given the covariance model in Equation (10), no assumption on the specific functional form for the random field has been considered. Conversely, a discussion about the explicit model for random field that leads to the product-sum can be found in De Iaco et al. (2019b, 2020).

The estimated values for the parameters of the product-sum models Equation (10), are reported in Table 1, for the small and long distance components.

After fitting the ST covariance model for each basic component, the ST-LCM is as follows:

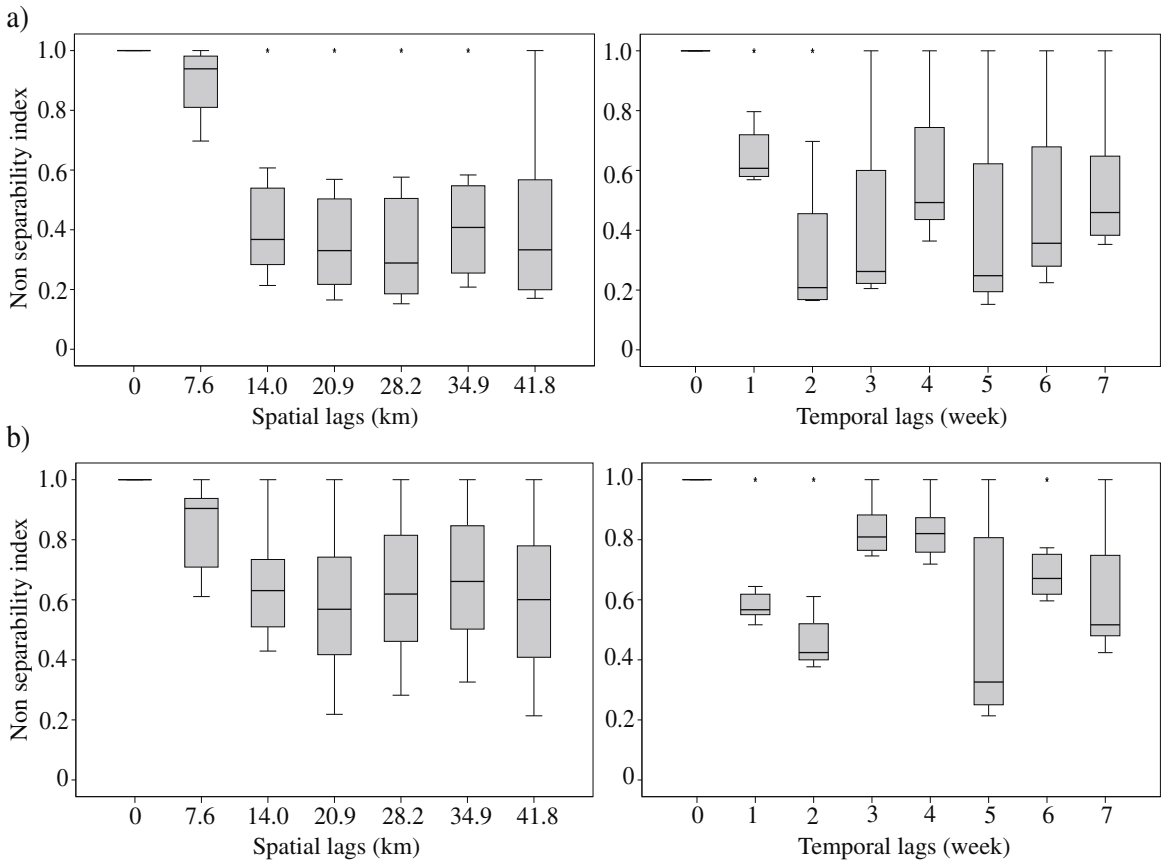


FIGURE 6 | Box and whisker plots showing sample non-separability ratios classified by spatial and temporal lags, computed for (a) the small distance component and (b) the long distance component.

TABLE 1 | Estimated parameters for the product-sum covariance models in Equation (10).

Small distance component ($l = 1$)	Long distance component ($l = 2$)
$k_1 = 0.589$	$k_2 = 2.707$
$k_{1_1} = 0.158$	$k_{1_2} = 0.111$
$k_{2_1} = 0.224$	$k_{2_2} = 0.102$
$k_{3_1} = 0.618$	$k_{3_2} = 0.787$
$C_{s_1}(\mathbf{u}) = Exp(\|\|\mathbf{u}\|\|; 14.5)$[a]	$C_{s_2}(\mathbf{u}) = Exp(\|\|\mathbf{u}\|\|; 25)$[a]
$C_{t_1}(v) = Exp(\|v\|;\ 2)$[a]	$C_{t_2}(v) = Exp(\|v\|;\ 4)$[a]

[a] $Exp(\cdot; a)$ is the exponential covariance model, with range a.

$$\mathbf{C}(\mathbf{u}, v) = \mathbf{B}_1\ c_1^*(\mathbf{u}, v) + \mathbf{B}_2\ c_2^*(\mathbf{u}, v), \tag{11}$$

where the unit sill covariances of both the short ($l = 1$) and long distance ($l = 2$) structures are modeled using the product-sum covariance model in Equation (10).

In the next section, the procedure to estimate the coregionalization matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ is discussed.

3.5 | Computation of Coregionalization Matrices

The computation of the elements of the (5×5) coregionalization matrices $\mathbf{B}_1$ and $\mathbf{B}_2$, as explained in Equation (7), required the knowledge of the following information:

- $\widehat{C}_{ij}(\mathbf{0}, 0), i, j = 1, \ldots, 5$, of the sample direct and cross-covariances of the deseasonalized and standardized variables under study (where $i = 1$ refers to evapotranspiration, $i = 2$ refers to maximum air temperature, $i = 3$ refers to maximum humidity, $i = 4$ refers to minimum humidity, and $i = 5$ refers to log-precipitations),
- the sample values of the direct and cross-covariances at the two scales of spatio-temporal variability, that is, $\widehat{C}_{ij}(\mathbf{u}, v)_1$ and $\widehat{C}_{ij}(\mathbf{u}, v)_2$ with $i, j = 1, \ldots, 5$, where $(\mathbf{u}, v)_1 = (14.5\,\text{km}, 2\,\text{weeks})$ and $(\mathbf{u}, v)_2 = (25\,\text{km}, 4\,\text{weeks})$.

Note that, given the model in Equation (11) based on unit sill basic components, the $\mathbf{B}_l, l = 1, 2$, clearly highlight the contribution of the components. Thus, the elements b_{ij}^1, with $i, j = 1, \ldots, 5$, of the matrix $\mathbf{B}_1$ can be easily determined as follows:

$$
\begin{array}{lllll}
b^1_{11} = 1 - 0.142 & b^1_{12} = 0.865 - 0.070 & b^1_{13} = -0.170 + 0.042 & b^1_{14} = -0.542 + 0.069 & b^1_{15} = -0.445 + 0.043 \\
b^1_{21} = b^1_{12} & b^1_{22} = 1 - 0.178 & b^1_{23} = -0.043 + 0.021 & b^1_{24} = -0.297 + 0.052 & b^1_{25} = -0.313 + 0.018 \\
b^1_{31} = b^1_{13} & b^1_{32} = b^1_{23} & b^1_{33} = 1 - 0.256 & b^1_{34} = 0.610 - 0.111 & b^1_{35} = 0.351 - 0.053 \\
b^1_{41} = b^1_{14} & b^1_{42} = b^1_{24} & b^1_{43} = b^1_{34} & b^1_{44} = 1 - 0.216 & b^1_{45} = 0.522 - 0.058 \\
b^1_{51} = b^1_{15} & b^1_{52} = b^1_{25} & b^1_{53} = b^1_{35} & b^1_{54} = b^1_{45} & b^1_{55} = 1 - 0.202
\end{array}
$$

and the elements of the matrix $\mathbf{B}_2$ reflect the residual variability. Hence, the following coregionalization matrices $\mathbf{B}_l, l = 1, 2$, are obtained:

$$
\mathbf{B}_1 = \begin{bmatrix}
0.858 & 0.795 & -0.128 & -0.473 & -0.402 \\
0.795 & 0.822 & -0.022 & -0.245 & -0.295 \\
-0.128 & -0.022 & 0.744 & 0.499 & 0.298 \\
-0.473 & -0.245 & 0.499 & 0.784 & 0.464 \\
-0.402 & -0.295 & 0.298 & 0.464 & 0.798
\end{bmatrix}, \quad
\mathbf{B}_2 = \begin{bmatrix}
0.142 & 0.070 & -0.042 & -0.069 & -0.043 \\
0.070 & 0.178 & -0.021 & -0.052 & -0.018 \\
-0.042 & -0.021 & 0.256 & 0.111 & 0.053 \\
-0.069 & -0.052 & 0.111 & 0.216 & 0.058 \\
-0.043 & -0.018 & 0.053 & 0.058 & 0.202
\end{bmatrix}, \tag{12}
$$

for which it is required that the positive definiteness condition is satisfied in order to guarantee the admissibility of the fitted ST-LCM.

3.6 | Specification of the ST-LCM

After computing the entries of the two coregionalization matrices $\mathbf{B}_l, l = 1, 2$, the appropriate ST-LCM for the analyzed data is specified as follows:

$$
\mathbf{C}(\mathbf{u}, v) = \mathbf{B}_1 \; c_1^*(\mathbf{u}, v) + \mathbf{B}_2 \; c_2^*(\mathbf{u}, v). \tag{13}
$$

The basic model parameters are listed in Table 1, $\mathbf{B}_1$ and $\mathbf{B}_2$ specified in Equation (12).

In the following section, the model in Equation (13) will be used to predict evapotranspiration levels by including secondary variables which complement the primary variable of interest. It is important to note that, in the multivariate ST approach, both direct and cross-correlation factors between the primary and secondary variables contribute to determining the weights of the estimator used for prediction. Empirical spatio-temporal correlation (Figure 4) shows that temperature residuals are positively correlated with ET_0 residuals, whereas the correlation is negative with respect to humidity and precipitation residuals; this influences the accuracy of prediction of the target variable.

The steps involving the detection of uncorrelated components, identification of an appropriate covariance model for each component, and the computation of coregionalization matrices can be executed with the support of properly defined functions in the R environment (R Core Team 2023). The code is available upon request from the corresponding author.

3.7 | Validation of Spatio-Temporal Models

At this stage of the analysis, it is interesting to evaluate the reliability of the fitted ST-LCM; thus, the adequacy of the model Equation (13) has been evaluated by comparing the direct and cross-covariances estimated for the study variables with the respect to the direct and cross-covariance models specified by the ST-LCM in Equation (13). Then, the mean absolute error (MAE) and root mean square error (RMSE) have been computed and the results have been summarized in Table 2, where the obtained error metrics have been averaged for all direct and cross-covariance functions, or exclusively for the direct covariance functions, as well as for the cross-covariance functions. In the Appendix of the paper, the 3D plots of the theoretical direct and cross-covariance functions computed by model (13) have been displayed (Figure A1); for each of them the error metrics (MAE and RMSE) measured between the sample and the theoretical covariance functions have been also reported. These statistics are characterized by very low values (the greatest error value

TABLE 2 | Error metrics for models' evaluation.

	ST-LCM (13)							
				ST Jackknife				
	All	Direct covariance	Cross-covariance	LOOC	15th week	16th week	17th week	18th week
MAE	0.036	0.055	0.026	0.303	0.228	0.166	0.240	0.373
RMSE	0.049	0.070	0.037	0.425	0.299	0.217	0.279	0.414
	Multivariate Gneiting–Matérn covariance model (14)							
				ST Jackknife				
	All	Direct covariance	Cross-covariance	LOOC	15th week	16th week	17th week	18th week
MAE	0.034	0.056	0.024	0.297	0.258	0.181	0.243	0.344
RMSE	0.044	0.072	0.030	0.420	0.317	0.210	0.289	0.384

is the RMSE averaged for all the direct covariance functions) and confirm the adequacy of the fitted ST-LCM to describe the spatio-temporal correlation among the variables.

Successively, a more accurate analysis on the reliability of the ST-LCM in Equation (13) has been carried out through the following techniques:

(a) the traditional leave-one-out cross-validation (LOOC), which has been computed to evaluate the performance of the model in interpolation mode; indeed, at each spatio-temporal point, the available measurement has been removed and the fitted multivariate model (13) has been used to estimate the primary variable of interest (deseasonalized and standardized values of evapotranspiration) by considering the other study variables (secondary variables) available in the neighborhood of the removed point;

(b) a jackknife-based validation, which has been implemented by removing the whole time series of the primary variable at some locations and then predicting the primary variable for 4 weeks of 2022 (from the 15th to the 18th week of 2022) after the last available week in the data set, by using measurements of the primary variable not at these locations, but exclusively the ones in the neighborhood and/or the values recorded at these selected locations for the secondary variables.

Note that for the above points (a) and (b), cokriging estimates have been obtained through the "COK2ST" *GSLib* routine, proposed by De Iaco et al. (2010), which has been customized to include the ST-LCM in Equation (13) based on the product-sum model at both small and large scales of variability.

In Table 2, the error metrics MAE and RMSE from the cross-validation and jackknife estimates of the primary variable obtained one time ahead, namely for each week from the 15th to the 18th week of 2022, are provided. These error metrics encourage the use of the fitted multivariate model, especially if compared with the same indexes computed for the univariate case. Indeed, the univariate model for the primary variable ET_0, obtained by excluding the correlations with the other variables, has produced an MAE equal to 0.481 and an RMSE equal to 0.668 on the cross-validation results, which correspond to a worsening of 58.7% and 57.2%, respectively, with respect to the multivariate case.

In the next section, the performance of the fitted ST-LCM (13) has been evaluated through a comparison with respect to the findings from the above-listed validation techniques (a) and (b) applied in the case of the fully non-separable multivariate Gneiting–Matérn covariance model, proposed by Allard, Clarotto, and Emery (2022).

4 | A Comparative Analysis Using a Multivariate Spatio-Temporal Model

In this section, the performance of the proposed ST-LCM based on the product-sum model has been compared with respect to the one of the fully non-separable multivariate Gneiting–Matérn covariance model, proposed by Allard, Clarotto, and Emery (2022). This recent model can be adopted in case of both non-negative or non-positive non-separability, according to its parameters α and β; thus, given the non-positive non-separability detected in Section 3.4, β has been set equal to zero, while α has been fixed equal to one. Consequently, within the class provided by Allard, Clarotto, and Emery (2022), the following model for the direct and cross covariance functions has been considered:

$$C_{ij}(\mathbf{u},v)=\frac{1}{\left(1+c|v|^{2a}\right)^{\delta}-A_iA_j\left(1+r|v|^{2\lambda}\right)^{-\delta}}\frac{\rho_{ij}b_{ij}^{-d}\sqrt{\tau_{ii}\tau_{jj}}}{\left[\left(1+c|v|^{2a}\right)^{b}-A_iA_j\left(1+r|v|^{2\lambda}\right)^{-b}\right]^{\nu_{ij}}}\mathcal{M}\left(\mathbf{u};\frac{\left[\left(1+c|v|^{2a}\right)^{b}-A_iA_j\left(1+r|v|^{2\lambda}\right)^{-b}\right]^{0.5}}{b_{ij}},\nu_{ij}\right)$$
$$(\mathbf{u},v)\in\mathbb{R}^d\times\mathbb{R}, \tag{14}$$

where ν and $\mathbf{b}^2$ are $(m\times m)$ symmetric conditionally negative semidefinite matrices all with positive entries and are assumed to be $a-$ separable, that is, $\nu_{ij}=0.5\left(\nu_{ii}+\nu_{jj}\right)$ and $b_{ij}^2=$

$0.5\left(b_{ii}^2 + b_{jj}^2\right), i = 1, \ldots, m$, $\mathcal{M}(\cdot)$ is a Matérn covariance function (Matérn 1966), τ is a $(m \times m)$ symmetric real matrix such that $\tau_{ii} = \sigma_i^2 b_{ii}^2 \left(1 - A_i^2\right)^{(1+\nu_{ii})}, i = 1, \ldots, m$ and $\tau_{ij} = \rho_{ij}\sqrt{\tau_{ii}\tau_{jj}}$, with $\rho_{ij}, i = 1, \ldots, m$, being a correlation matrix. As detailed in Theorem 2 and Corollary 1 of the paper of Allard, Clarotto, and Emery (2022), the admissibility condition of the model is that $\tau e^{-\nu}/\Gamma(\nu)$ is a positive semidefinite matrix. Note that, the scale parameters in the model are such that $c, r > 0$, the shape parameters (a, λ) are $0 < a \leq 1, 0 < \lambda < 1$, and the smoothness parameters $\nu_{ij} > 0$. Furthermore, the parameter $\delta \in [0, 1]$, and the coefficients which describe the multivariate aspect of the temporal part are $0 \leq A_i < 1, i = 1, \ldots, m$.

As suggested by the authors, the parameters estimation has been obtained through the optimization of the pairwise marginal Gaussian likelihoods computed on all pairs of data (for distance less than 21 km in space and 3 weeks in time, beyond which the direct correlations decay) (Bourotte, Allard, and Porcu 2016). Because of the high-dimensional parameter space, it has been maximized sequentially and iteratively by considering blocks of parameters, while fixing all other parameters to the previously estimated values. Regarding the separability parameter b, which ranges between 0 (separability case) and 1 (full spatio-temporal non-separability case), the optimization procedure, performed by calling the R function `nlminb`, has been conducted for different fixed values of b (equal to 0,0.1, …, 1), then the value of b which guarantees the maximum likelihood has been picked.

Furthermore, for each iteration $\widehat{\rho}$ has been computed using the estimated parameters $\widehat{\mathbf{b}}$, $\widehat{\nu}$, and $\widehat{\mathbf{A}}$ and enforcing $\widehat{\nu}$ and $\widehat{\tau} e^{-\widehat{\nu}}/\Gamma(\widehat{\nu})$ to be admissible, as clarified in the study by Allard, Clarotto, and Emery (2022). Table 3 reports the estimated parameters maximizing the pairwise likelihood (PL) for the multivariate spatio-temporal covariance model in Equation (14), corresponding to $\widehat{b} = 0.5$, which is the separability parameter value associated to the highest maximized PL.

The error indices MAE and RMSE have been computed in order to measure, on one hand the discrepancy between the empirical and the theoretical covariance surfaces, and on the other hand the deviation between the observed and the estimated values at some points over the domain. As regards the former aspect, the above error metrics are given for the theoretical covariance functions, classified by direct and cross, in Table 2, and for each direct or cross-covariance model in Figure A2 of the Appendix.

On the other hand, to evaluate the predictive performance of model (14) the same validation approaches described at points (a) and (b) of Section 3.7 have been recalled in this context, so that the error metrics related to the two classes of models (ST-LCM and Gneiting–Matérn covariance model) have been also compared (Table 2).

By analyzing these statistics, it is evident that the selection of the Gneiting–Matérn covariance model has determined, in some cases, a slightly worse fitting with respect to the ST-LCM. This is also confirmed in terms of the results obtained from cross-validation and jackknife. Thus, it cannot be declared the supremacy of one model with respect to the other. However, it is worth to underline the complexity of the fitting step of the Gneiting–Matérn covariance model, under a computational point of view, especially when the number of variables is greater than 3 and the number of sample points are not in the order of a few tens in space or a few tens of consecutive points in time. Indeed, in these cases, the pairwise marginal Gaussian likelihoods computed on all pairs of data, as described in Bourotte, Allard, and Porcu (2016), is extremely time-consuming.

The previous results have confirmed the goodness of both multivariate models to describe spatio-temporal correlations between evapotranspiration and meteorological conditions. Thus, in the following, spatio-temporal predictions of the primary variable (ET_0) based on the ST-LCM in Equation (13) will be provided and some comments on the usefulness of these results for supporting actions to prevent drought emergency scenarios are considered.

5 | Spatio-Temporal Predictions

Previous models validations support the use of the ST-LCM in Equation (13) to make predictions over the study area and for some time points after the last available data.

In this section, spatio-temporal cokriging has been performed to predict ET_0 values for 4 weeks after the last observed week. In particular, by using the ST-LCM in Equation (13) and the standardized residuals for the primary and secondary variables available from the first week of 2000 to the fourteenth week of 2022, the ET_0 residuals could be predicted from the fifteenth (between April 11th and April 17th) to the eighteenth (between May 2nd and May 8th) week of 2022. Subsequently, the predicted values have been rescaled and the ET_0 seasonal component, previously estimated as discussed in Section 3.2, has been added to obtain the ET_0 weekly levels.

TABLE 3 | Estimated parameters for the multivariate spatio-temporal model in Equation (14).

Temporal parameters		**Spatial parameters**		**Correlation parameters**	
$\mathbf{c = 1.109}$	$\mathbf{A_1 = 0.035}$	$\mathbf{b_1 = 3.68 \cdot 10^{-5}}$	$\mathbf{\nu_1 = 0.205}$	$\mathbf{\rho_{1,2} = -0.214}$	$\mathbf{\rho_{2,4} = 0.357}$
$a = 0.583$	$A_2 = 0.010$	$b_2 = 1.54 \cdot 10^{-5}$	$\nu_2 = 0.250$	$\rho_{1,3} = -0.592$	$\rho_{2,5} = -0.071$
$r = 97.074$	$A_3 = 0.004$	$b_3 = 2.46 \cdot 10^{-5}$	$\nu_3 = 0.205$	$\rho_{1,4} = -0.597$	$\rho_{3,4} = 0.566$
$\lambda = 0.999$	$A_4 = 0.041$	$b_4 = 1.68 \cdot 10^{-5}$	$\nu_4 = 0.453$	$\rho_{1,5} = 0.845$	$\rho_{3,5} = -0.358$
$\delta = 1.000$	$A_5 = 0.025$	$b_5 = 4.54 \cdot 10^{-5}$	$\nu_1 = 0.204$	$\rho_{2,3} = 0.687$	$\rho_{4,5} = -0.486$

Note: The indices 1, 2, 3, 4, and 5 refer to ET_0, H_M, H_m, lnP, and T_M, respectively.

The color maps displayed in Figure 7 show the spatial distributions of the ET_0 values for the 15th to 18th weeks of 2022. Estimated crop evapotranspiration levels are high across the entirety of the Veneto Region, since the data have been estimated for the time of the year in which moderate to almost warm air temperatures occur and when ET_0 values tend to assume increasing levels, as already highlighted through the results from the exploratory data analysis discussed in Section 3.2.

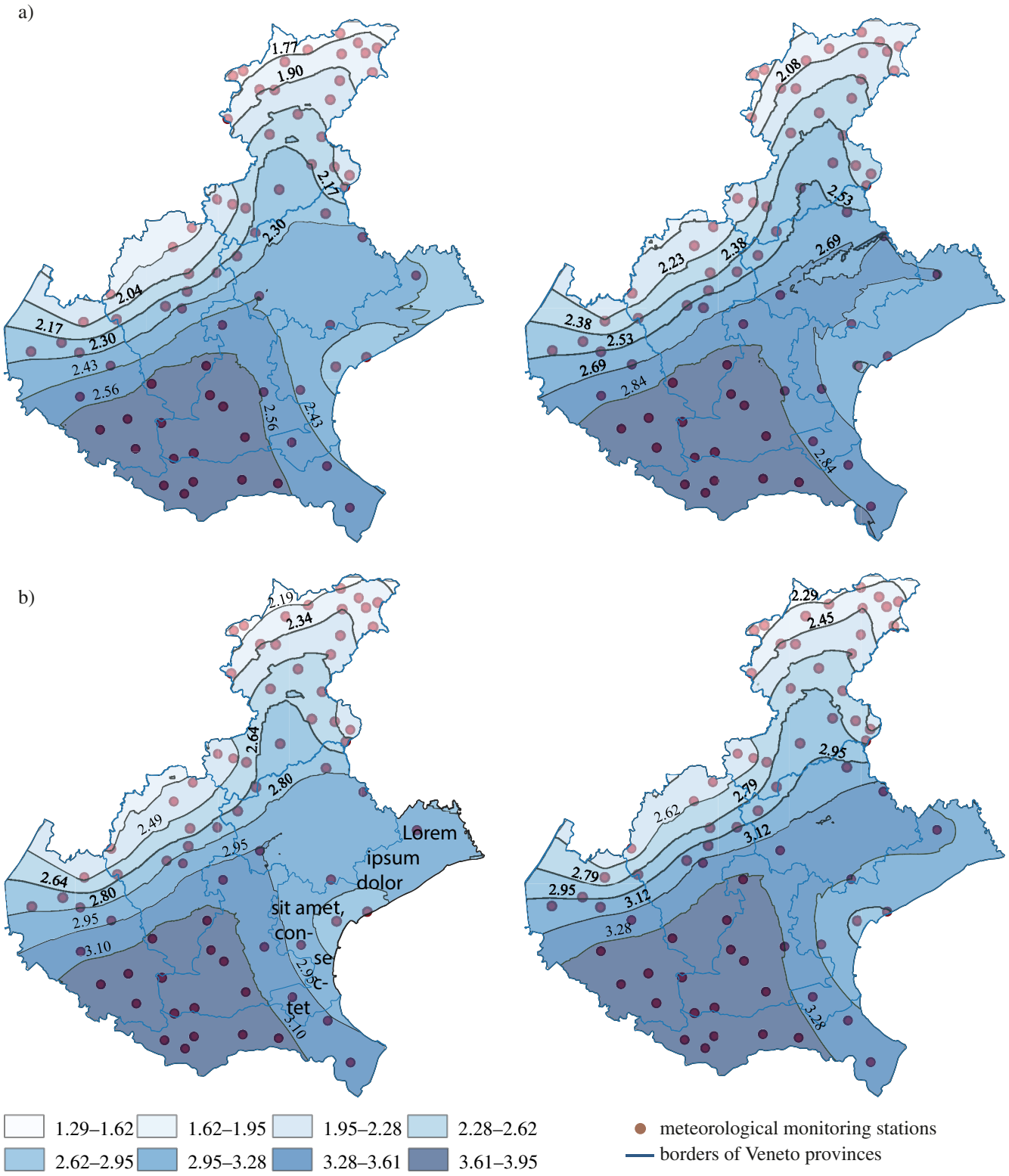


FIGURE 7 | Color maps showing the ET_0 predictions obtained by fitting the ST-LCM and using all available data for (a) the 15th, (b) the 16th, (c) the 17th and (d) the 18th weeks of 2022.

In particular, large values are estimated in the southwestern area (Verona, Rovigo, and Padua provinces), close to the border with Emilia-Romagna. This area is primarily characterized by sandy soils which, under the same climatic conditions, allow more intensive evapotranspiration than areas with fine-textured soils composed of clay.

On the other hand, low values of evapotranspiration have been systematically estimated in the northern part of Veneto, that is, the pre-Alpine area and Alpine hills (Belluno province), where the altitude determines the air temperature during the warm season; consequently, the evapotranspiration is lower than in the other parts of the region, which feature wet heat.

It is important to note that, when predicting the ET_0 levels, climatic conditions, as well as the pedological structure of the soil, are crucial factors and their interrelationships in space and time with evapotranspiration processes cannot be neglected. Moreover, it is worth underlining that in 2022 the spring season in Veneto has been characterized by a strong reduction in rainfall and increasing values of ET_0 and temperature, determining the decrease in the availability of water resources, which has caused suffering and slowed down the development of agricultural crops. The proposed multivariate ST-LCM aims to provide reliable predictions of the ET_0 at unsampled spatio-temporal points. This model can also be a useful decision-making tool for water resource management, accounting for the climatic characteristics within the Veneto region as well as the temporal evolution of meteorological variables. Hence, policy makers would therefore benefit from multivariate geostatistical modeling and prediction of ET_0 levels, as discussed in this paper, to define supporting measures for the agricultural sector, especially during the driest periods of the year.

6 | Discussion and Conclusions

In this study, the spatio-temporal linear coregionalization model (ST-LCM) fitting procedure was applied to describe spatio-temporal correlations between evapotranspiration and meteorological weekly conditions (air temperature, maximum and minimum humidity, and precipitation) measured over the Veneto Region between 2000 and 2022. The basic uncorrelated components were chosen based on their primary features assessed in terms of symmetry, separability/non-separability, and type of non-separability. Following model validation and a comparative analysis using the multivariate spatio-temporal Gneiting–Matérn covariance model (Allard, Clarotto, and Emery 2022), the goodness of the ST-LCM in predicting evapotranspiration was quantified.

In contrast to various methods reported in the literature for computing evapotranspiration values at the same sparse meteorological stations where climatic parameters are recorded (Di Nunno et al. 2023; Hargreaves 1974), the procedure described in this paper enables the construction of a suitable multivariate model for making predictions at unsampled spatial and temporal points. Furthermore, the direct and cross spatio-temporal correlations of primary and secondary variables contribute to determining the weight of the estimators used for prediction. In particular, the spatio-temporal predictions of ET_0 using ST-LCM yielded the identification of parts of the Veneto Region characterized by large values of evapotranspiration, namely along the coastal area, ranging from the Po Delta to the Venetian Lagoon, and over the areas bordering Friuli-Venezia Giulia in the East and Emilia Romagna in the South-West. Conversely, lower values were estimated in the northeastern part, which is a mountainous area. Monitoring the spatial and temporal evolution of ET_0, according to the meteorological variables, might be of great interest in the light of climate change and the impact of evapotranspiration on water management resources in agriculture. Indeed, Veneto is one of the most productive Italian regions in the agriculture sector, and it is devoted to produce grain crops, in the southern and eastern areas, as well as fruits and wine, in the plains and foothills, where higher evapotranspiration levels have been predicted. Hence, the results obtained in this paper might be particularly valuable for water managers who require reliable estimates of spatio-temporal evapotranspiration and predictions for water resources distribution planning, optimal irrigation scheduling, and sustainable water resources management. Although the ST-LCM produced appreciable results, future work will be required to compare this model against other recent methods, as the ones based on blind source separation ideas (Muehlmann, De Iaco, and Nordhausen 2023), which allows a similar latent modeling approach as presented here with the ST-LCM. Moreover, remote sensing satellite data from air temperature, net radiation, soil heat flux density, wind speed, and vapor pressure could be utilized to enhance reference evapotranspiration estimates.

Acknowledgments

The work of Claudia Cappello and Klaus Nordhausen was partly supported by the Austrian Science Fund P31881-N32 and Klaus Nordhausen also acknowledges support from the Research Council of Finland, grant 363261. The work of Sandra De Iaco was supported by the National Biodiversity Future Center-NBFC, Spoke 4, Activity 4.1., sub-activity 4.1.1. Funder: Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4—Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union—Next GenerationEU. Award Number: Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP F87G22000290001, Project title "National Biodiversity Future Center—NBFC."

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from Arpa Veneto. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of Arpa Veneto.

References

Abrishami, N., A. Sepaskhah, and M. Shahrokhnia. 2019. "Estimating Wheat and Maize Daily Evapotranspiration Using Artificial Neural Network." *Theoretical and Applied Climatology* 135: 945–958. https://doi.org/10.1007/s00704-018-2418-4.

Alidoost, F., A. Stein, and Z. Su. 2018. "Copula-Based Interpolation Methods for Air Temperature Data Using Collocated Covariates."

Spatial Statistics 28: 128–140. One world, one health. https://doi.org/10.1016/j.spasta.2018.08.003.

Allard, D., L. Clarotto, and X. Emery. 2022. "Fully Nonseparable Gneiting Covariance Functions for Multivariate Space–Time Data." *Spatial Statistics* 52: 100706. https://doi.org/10.1016/j.spasta.2022.100706.

Babak, O., and C. V. Deutsch. 2009. "An Intrinsic Model of Coregionalization That Solves Variance Inflation in Collocated Cokriging." *Computers & Geosciences* 35, no. 3: 603–614. https://doi.org/10.1016/j.cageo.2008.02.025.

Berrocal, V., A. Gelfand, and D. Holland. 2010. "A Bivariate Space–Time Downscaler Under Space and Time Misalignment." *Annals of Applied Statistics* 4, no. 4: 1942–1975. https://doi.org/10.1214/10-AOAS351.

Bevilacqua, M., A. Hering, and E. Porcu. 2015. "On the Flexibility of Multivariate Covariance Models: Comment on the Paper by Genton and Kleiber." *Statistical Science* 30, no. 2: 167–169. https://doi.org/10.1214/15-STS516.

Bourotte, M., D. Allard, and E. Porcu. 2016. "A Flexible Class of Non-Separable Cross-Covariance Functions for Multivariate Space–Time Data." *Spatial Statistics* 18: 125–146. https://doi.org/10.1016/j.spasta.2016.02.004.

Calculli, C., A. Fassò, F. Finazzi, A. Pollice, and A. Turnone. 2015. "Maximum Likelihood Estimation of the Multivariate Hidden Dynamic Geostatistical Model With Application to Air Quality in Apulia, Italy." *Environmetrics* 26, no. 2: 406–417. https://doi.org/10.1002/env.2345.

Cappello, C., S. De Iaco, and M. Palma. 2022. "Computational Advances for Spatio-Temporal Multivariate Environmental Models." *Computational Statistics* 37: 651–670. https://doi.org/10.1007/s00180-021-01132-0.

Cappello, C., S. De Iaco, and D. Posa. 2018. "Testing the Type of Non-Separability and Some Classes of Spacetime Covariance Function Models." *Stochastic Environmental Research and Risk Assessment* 32: 17–35. https://doi.org/10.1007/s00477-017-1472-2.

Cappello, C., S. De Iaco, and D. Posa. 2020. "Covatest: An R Package for Selecting a Class of Space-Time Covariance Functions." *Journal of Statistical Software* 94, no. 1: 1–42. https://doi.org/10.18637/jss.v094.i01.

Cardoso, J., and A. Souloumiac. 1996. "Jacobi Angles for Simultaneous Diagonalization." *SIAM Journal on Matrix Analysis and Applications* 17: 161–164. https://doi.org/10.1137/S0895479893259546.

Chen, W., M. Genton, and Y. Sun. 2021. "Space-Time Covariance Structures and Models." *Annual Review of Statistics and Its Application* 8: 191–215. https://doi.org/10.1146/annurev-statistics-042720-115603.

Chia, M., Y. Huang, and C. Koo. 2020. "Support Vector Machine Enhanced Empirical Reference Evapotranspirationestimation With Limitedmeteorological Parameters." *Computers and Electronics in Agriculture* 175: 105577. https://doi.org/10.1016/j.compag.2020.105577.

Choi, J., B. Reich, M. Fuentes, and J. Davis. 2009. "Multivariate Spatial-Temporal Modeling and Prediction of Speciated Fine Particles." *Journal of Statistical Theory and Practice* 3, no. 2: 407–418. https://doi.org/10.1080/15598608.2009.10411933.

De Iaco, S., D. Myers, M. Palma, and D. Posa. 2010. "FORTRAN Programs for Space–Time Multivariate Modeling and Prediction." *Computers & Geosciences* 36: 636–646. https://doi.org/10.1016/j.cageo.2009.10.004.

De Iaco, S., D. Myers, and D. Posa. 2003. "The Linear Coregionalization Model and the Product-Sum Space-Time Variogram." *Mathematical Geosciences* 35: 25–38. https://doi.org/10.1023/A:1022425111459.

De Iaco, S., M. Palma, and D. Posa. 2005. "Modeling and Prediction of Multivariate Space-Time Random Fields." *Computational Statistics and Data Analysis* 48, no. 3: 525–547. https://doi.org/10.1016/j.csda.2004.02.011.

De Iaco, S., M. Palma, and D. Posa. 2016. "A General Procedure for Selecting a Class of Fully Symmetric Space-Time Covariance Functions." *Environmetrics* 27: 212–224. https://doi.org/10.1002/env.2392.

De Iaco, S., M. Palma, and D. Posa. 2019a. "Choosing Suitable Linear Coregionalization Models for Spatio-Temporal Data." *Stochastic Environmental Research and Risk Assessment* 33: 1419–1434. https://doi.org/10.1007/s00180-021-01132-0.

De Iaco, S., and D. Posa. 2013. "Positive and Negative Non-Separability for Space-Time Covariance Models." *Journal of Statistical Planning and Inference* 143: 378–391. https://doi.org/10.1016/j.jspi.2012.07.006.

De Iaco, S., D. Posa, C. Cappello, and S. Maggio. 2019b. "Isotropy, Symmetry, Separability and Strict Positive Definiteness for Covariance Functions: A Critical Review." *Spatial Statistics* 29: 89–108. https://doi.org/10.1016/j.spasta.2018.09.003.

De Iaco, S., D. Posa, C. Cappello, and S. Maggio. 2020. "On Some Characteristics of Gaussian Covariance Functions." *International Statistical Review* 89: 36–53. https://doi.org/10.1111/insr.12403.

De Iaco, S., D. Posa, and D. Myers. 2013. "Characteristics of Some Classes of Space-Time Covariance Functions." *Journal of Statistical Planning and Inference* 143: 2002–2015. https://doi.org/10.1016/j.jspi.2013.06.006.

Di Nunno, F., M. De Matteo, G. Izzo, and F. Granata. 2023. "A Combined Clustering and Trends Analysis Approach for Characterizing Reference Evapotranspiration in Veneto." *Sustainability* 15, no. 14: 11091. https://doi.org/10.3390/su151411091.

Dörr, C., and M. Schlather. 2023. "Covariance Models for Multivariate Random Fields Resulting From Pseudo Cross-Variograms." *Journal of Multivariate Analysis* 197: 105–199. https://doi.org/10.1016/j.jmva.2023.105199.

Emery, X. 2010. "Iterative Algorithms for Fitting a Linear Model of Coregionalization." *Compututers & Geosciences* 36, no. 9: 1150–1160. https://doi.org/10.1016/j.cageo.2009.10.007.

Fassò, A., and F. Finazzi. 2011. "Maximum Likelihood Estimation of the Dynamic Coregionalization Model With Heterotopic Data." *Environmetrics* 22, no. 6: 735–748.

Gentilucci, M., M. Bufalini, M. Materazzi, et al. 2021. "Calculation of Potential Evapotranspiration and Calibration of the Hargreaves Equation Using Geostatistical Methods Over the Last 10 Years in Central Italy." *Geosciences* 11, no. 8: 348. https://doi.org/10.3390/geosciences11080348.

Genton, M. G., and W. Kleiber. 2015. "Cross-Covariance Functions for Multivariate Geostatistics." *Statistical Science* 30, no. 2: 147–163. https://doi.org/10.1214/14-STS487.

Gnann, S., M. Allmendinger, C. Haslauer, and A. Bárdossy. 2018. "Improving Copula-Based Spatial Interpolation With Secondary Data." *Spatial Statistics* 28: 105–127. https://doi.org/10.1016/j.spasta.2018.07.001.

Gneiting, T., W. Kleiber, and M. Schlather. 2010. "Matérn Cross-Covariance Functions for Multivariate Random Fields." *Journal of the American Statistical Association* 105, no. 491: 1167–1177. https://doi.org/10.1198/jasa.2010.tm09420.

Goovearts, P., and P. Sonnet. 1993. "Study of Spatial and Temporal Variations of Hydrogeochimical Variables Using Factorial Kriging Analysis." In *Geostatistics tróia '92*, edited by A. Soares, vol. 1, 745–756. Dordrecht: Springer Netherlands.

Hargreaves, G. 1974. "Estimation of Potential and Crop Evapotranspiration." *Transactions of the American Society of Agricultural and Biological Engineers* 17: 701–704. https://doi.org/10.13031/2013.36941.

Hodam, S. 2017. "Spatial Interpolation of Reference Evapotranspiration in India: Comparison of Idw and Kriging Methods." *Journal of the Institution of Engineers (India): Series A* 98: 511–524. https://doi.org/10.1007/s40030-017-0241-z.

Hurtado-Uria, C., D. Hennessy, L. Shalloo, D. O'Connor, and L. Delaby. 2013. "Relationships Between Meteorological Data and Grass Growth

Over Time in the South of Ireland." *Irish Geography* 46, no. 3: 175–201. https://doi.org/10.1080/00750778.2013.865364.

Illner, K., J. Miettinen, C. Fuchs, et al. 2015. "Model Selection Using Limiting Distributions of Second-Order Blind Source Separation Algorithms." *Signal Processing* 113: 95–103. https://doi.org/10.1016/j.sigpro.2015.01.017.

Ip, R., and W. Li. 2016. "Matérn Cross-Covariance Functions for Bivariate Spatio-Temporal Random Fields." *Spatial Statistics* 17: 22–37. https://doi.org/10.1016/j.spasta.2016.04.004.

Ip, R., and W. Li. 2017a. "A Class of Valid Matérn Cross-Covariance Functions for Multivariate Spatio-Temporal Random Fields." *Statistics & Probability Letters* 130: 115–119. https://doi.org/10.1016/j.spl.2017.07.019.

Ip, R., and W. Li. 2017b. "On Some Matérn Covariance Functions for Spatio-Temporal Random Fields." *Statistica Sinica* 27: 805–822. https://doi.org/10.5705/ss.202015.0037.

Krupskii, P., and M. G. Genton. 2017. "Factor Copula Models for Data With Spatio-Temporal Dependence." *Spatial Statistics* 22: 180–195. https://doi.org/10.1016/j.spasta.2017.10.001.

Li, B., M. G. Genton, and M. Sherman. 2007. "A Nonparametric Assessment of Properties of Space-Time Covariance Functions." *Journal of the American Statistical Association* 102, no. 478: 736–744. https://doi.org/10.1198/016214507000000202.

Li, B., M. G. Genton, and M. Sherman. 2008. "Testing the Covariance Structure of Multivariate Random Fields." *Biometrika* 95, no. 4: 813–829. https://doi.org/10.1093/biomet/asn053.

Martínez-Cob, A. 1996. "Multivariate Geostatistical Analysis of Evapotranspiration and Precipitation in Mountainous Terrain." *Journal of Hydrology* 174, no. 1: 19–35. https://doi.org/10.1016/0022-1694(95)02755-6.

Matérn, B. 1966. *Spatial Variation; Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Stockholm: University of Sweden.

Miettinen, J., K. Nordhausen, and S. Taskinen. 2017. "Blind Source Separation Based on Joint Diagonalization in R: The Packages JADE and BSSasymp." *Journal of Statistical Software* 76, no. 2: 1–31. https://doi.org/10.18637/jss.v076.i02.

Muehlmann, C., S. De Iaco, and K. Nordhausen. 2023. "Blind Recovery of Sources for Multivariate Space-Time Random Fields." *Stochastic Environmental Research and Risk Assessment* 37: 1593–1613. https://doi.org/10.1007/s00477-022-02348-2.

Myers, D. E. 1995. "The Linear Coregionalization and Simultaneous Diagonalization of the Variogram Matrix Function." *Sciences de la Terre* 32: 125–139.

Otto, P., A. Fusta Moro, J. Rodeschini, et al. 2024. "Spatiotemporal Modelling of PM2.5 Concentrations in Lombardy (Italy): A Comparative Study." *Environmental and Ecological Statistics* 31: 245–272. https://doi.org/10.1007/s10651-023-00589-0.

Porcu, E., R. Furrer, and D. Nychka. 2021. "30 Years of Space–Time Covariance Functions." *WIREs Computational Statistics* 13, no. 2: e1512. https://doi.org/10.1002/wics.1512.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing [Computer Software Manual]*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rouhani, S., and H. Wackernagel. 1990. "Multivariate Geostatistical Approach to Space-Time Data Analysis." *Water Resources Research* 26, no. 4: 585–591. https://doi.org/10.1029/WR026i004p00585.

Silva, D., and C. Deutsch. 2018. "Multivariate Data Imputation Using Gaussian Mixture Models." *Spatial Statistics* 27: 74–90. https://doi.org/10.1016/j.spasta.2016.11.002.

Wackernagel, H. 2003. *Multivariate Geostatistics: An Introduction With Applications*. Berlin: Springer.

Appendix A

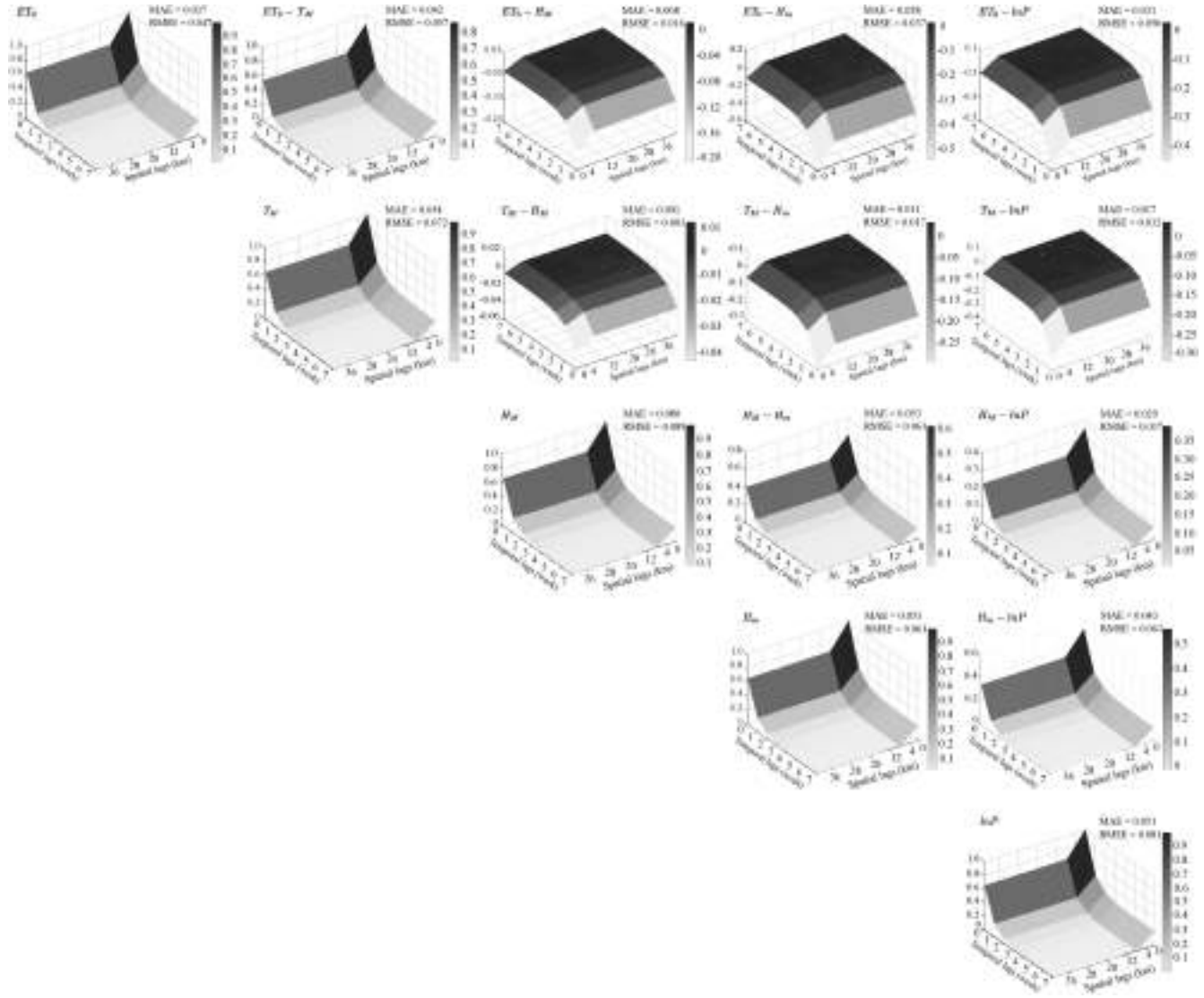


FIGURE A1 | 3D plots of the spatio-temporal direct (diagonal) and cross (off diagonal) covariance models in Equation (13) of ET_0, T_M, H_M, H_m, and ln P standardized residuals.

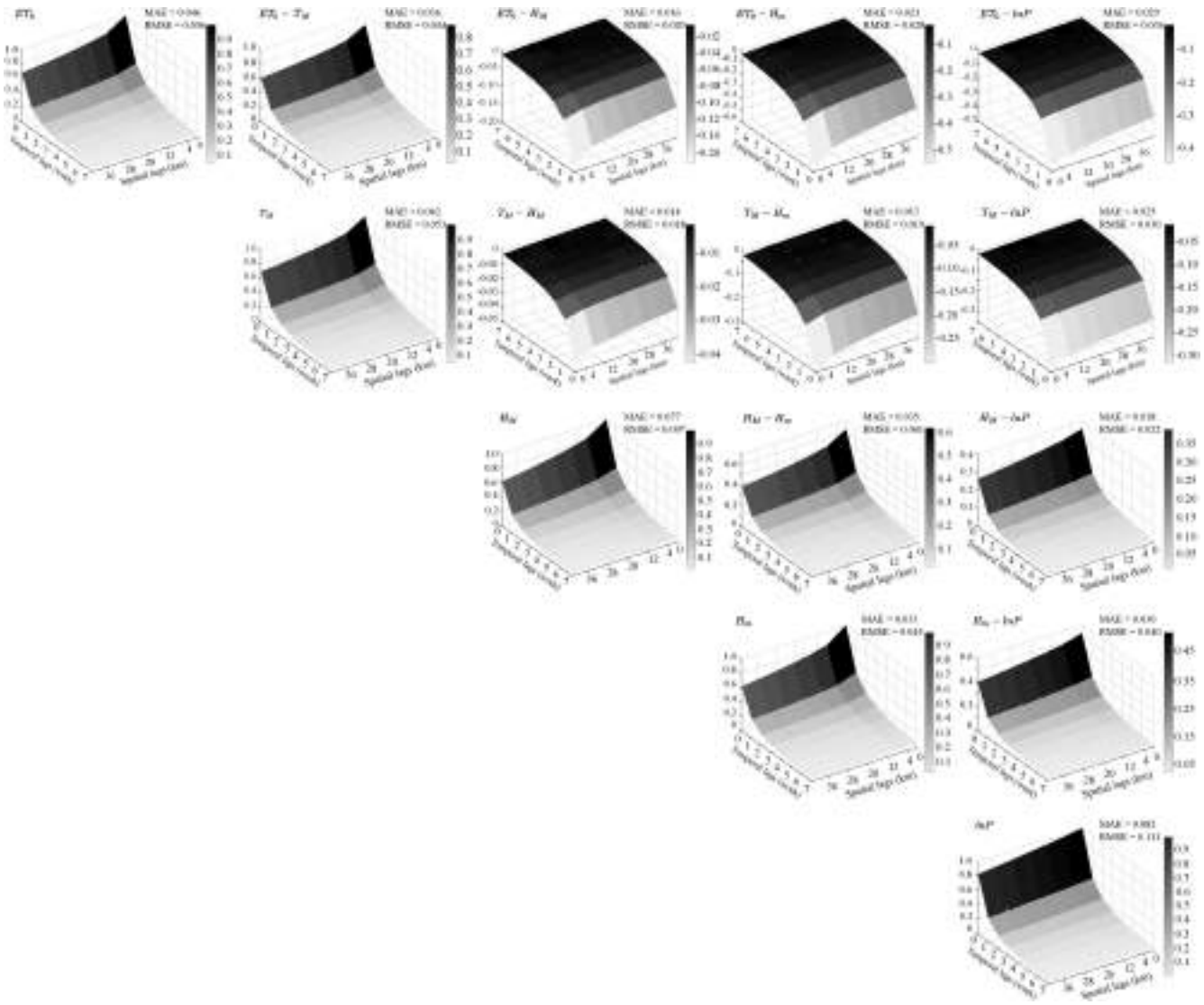


FIGURE A2 | 3D plots of the spatio-temporal direct (diagonal) and cross (off diagonal) covariance models in Equation (14) of ET_0, T_M, H_M, H_m, and ln P standardized residuals.