

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



***National Institute for Applied Statistics Research
Australia***

University of Wollongong, Australia

Working Paper

05-23

**Spatial Statistical Prediction of Solar-Induced Chlorophyll
Fluorescence (SIF) from Multivariate OCO-2 Data**

Josh Jacobson, Noel Cressie, and Andrew Zammit-Mangion

*Copyright © 2023 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia T: +61 2 42215076. E: karink@uow.edu.au



Article

Spatial Statistical Prediction of Solar-Induced Chlorophyll Fluorescence (SIF) from Multivariate OCO-2 Data

Josh Jacobson , Noel Cressie and Andrew Zammit-Mangion

School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia

* Correspondence: jj829@uowmail.edu.au

Abstract: Solar-induced chlorophyll fluorescence, or SIF, is a part of the natural process of photosynthesis. SIF can be measured from space by instruments such as the Orbiting Carbon Observatory-2 (OCO-2), making it a useful proxy for monitoring gross primary production (GPP), which is a critical component of Earth's carbon cycle. The complex physical relationship between SIF and GPP is frequently studied using OCO-2 observations of SIF since they offer the finest spatial resolution available. However, measurement error (noise) and large gaps in spatial coverage limit the use of OCO-2 SIF to highly aggregated scales. To study the relationship between SIF and GPP across varying spatial scales, de-noised and gap-filled (i.e., Level 3) SIF data products are needed. Using a geostatistical methodology called cokriging, which includes kriging as a special case, we develop coSIF: a Level 3 SIF data product at a 0.05-degree resolution. As a natural secondary variable for cokriging, OCO-2 observes column-averaged atmospheric carbon dioxide concentrations (XCO_2) simultaneously with SIF. There is a suggested lagged spatio-temporal dependence between SIF and XCO_2 , which we characterize through spatial covariance and cross-covariance functions. Our approach is highly parallelizable and accounts for non-stationary measurement errors in the observations. Importantly, each datum in the resulting coSIF data product is accompanied by a measure of uncertainty. Extant approaches do not provide formal uncertainty quantification, nor do they leverage the cross-dependence with XCO_2 .

Keywords: multivariate spatial prediction; uncertainty quantification; cokriging; full bivariate Matérn model; photosynthesis; solar-induced chlorophyll fluorescence; column-average carbon dioxide; Orbiting Carbon Observatory-2



Citation: Jacobson, J.; Cressie, N.; Zammit-Mangion, A. Spatial Statistical Prediction of Solar-Induced Chlorophyll Fluorescence (SIF) from Multivariate OCO-2 Data. *Remote Sens.* **2023**, *15*, 4038. <https://doi.org/10.3390/rs15164038>

Academic Editor: Meredith Franklin

Received: 30 June 2023

Revised: 2 August 2023

Accepted: 8 August 2023

Published: 15 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quantifying the spatio-temporal patterns of gross primary production (GPP), which is the total amount of carbon dioxide (CO_2) absorbed through photosynthesis, is an important aspect of studying ecosystem function and the carbon cycle. While GPP cannot be directly measured on global scales, remote sensing observations of solar-induced chlorophyll fluorescence (SIF) provide a reliable proxy for quantifying GPP globally [1–4]. SIF occurs when chlorophyll molecules return from excited to non-excited states and dissipate the light energy absorbed during photosynthesis, in part, through re-emission of fluorescent photons in the spectral range between 650 nm (red) and 800 nm (far red) with varying intensity [5]. Hence, there is a complex physical relationship between SIF and GPP [6]. An accurate characterization of this relationship relies on the availability of high-quality, spatially complete SIF data [7].

Measurements of SIF have been examined in laboratory and field studies, from the sub-cellular scale up to the leaf scale, for decades [8], although more recently it has become possible to measure SIF from satellite instruments at the global scale [9–12]. Different satellites collect global remote-sensing observations of SIF at different spatial and temporal resolutions. Available records of these observations, sometimes called Level 2 data products (e.g., [13]), are of limited use in analyzing the physical relationship with GPP due to large measurement errors, coarse spatial resolutions, and/or sparse spatial observations [14].

These issues are less prevalent in the fine-resolution, high-density observations recently available from the Tropospheric Monitoring Instrument (TROPOMI), but gaps in coverage are still present [15,16]. To examine the relationship between SIF and GPP across different spatial scales, de-noised and gap-filled (i.e., Level 3) data products with quantified uncertainties are required.

Several studies have proposed methods to produce Level 3 SIF data products with enhanced spatial resolution, typically at 0.05 degrees and as fine as 500 m in longitude and latitude. These methods are based on a variety of Level 2 SIF data products, but they are categorized into two general approaches: (1) physiologically motivated approaches that leverage relationships between SIF and other explanatory variables to downscale SIF from coarser resolutions [17–21], and (2) machine learning approaches that either downscale coarse-resolution SIF using ancillary datasets available at a finer resolution [22–24], or gap-fill fine-resolution but spatially sparse SIF observations [25–27]. Though these Level 3 data products have demonstrated overall capability in capturing the spatial patterns in SIF [7], none are accompanied by formally quantified uncertainties of each gap-filled SIF value. As a consequence, validation procedures for estimated Level 3 SIF fields are lacking.

Spatial statistical prediction (e.g., kriging) is another approach for producing Level 3 data products. As highlighted by [28], a key advantage of a statistical approach is the full quantification of uncertainty that is immediately available with the spatial predictions. Locally constructed kriging models were developed in [29,30] to produce coarse-resolution Level 3 data products for multiple remotely sensed processes, including SIF. However, a problematic feature of local kriging is that the spatial union of multiple local models can result in an incoherent global probability structure and, hence, in invalid prediction uncertainties [28]. Instead, spatial-statistical methods that consider satellite observations as realizations from an underlying spatial stochastic process that covers the domain of interest result in coherent uncertainty quantifications of all upscaled SIF estimates.

When observations of multivariate processes are available, as is often the case in remote sensing, the additional information can be leveraged for prediction of a primary process of interest through cokriging [31,32]. For instance, NASA's Orbiting Carbon Observatory-2 (OCO-2) satellite observes both SIF and column-averaged atmospheric carbon dioxide concentrations (XCO_2). Since CO_2 is a reactant in the process of photosynthesis, for which SIF is a proxy, dependence between SIF and XCO_2 is expected.

In this article, we propose a cokriging-based method that leverages any dependence between SIF and XCO_2 to produce de-noised and gap-filled predictions for SIF with quantified uncertainties. This reduces to a kriging-based method when SIF has no dependence on XCO_2 , which can happen during non-productive seasons. Our cokriging method is highly parallelizable, directly accounts for non-stationary measurement errors, and offers flexible spatial resolution through aggregation of the fine-resolution Level 3 data product. Here, we apply the cokriging method to multivariate OCO-2 data over a large region in North America to illustrate coSIF, which is our 0.05-degree, monthly resolution Level 3 SIF data product that includes coherent uncertainty quantification of each SIF estimate. Compared to kriging with SIF observations only, we find that our cokriging method can reduce prediction uncertainty and improve prediction accuracy when there is dependence between SIF and XCO_2 .

The remainder of this article is organized as follows. In Section 2, we describe how OCO-2 observations of SIF and XCO_2 are used in the construction of a bivariate spatial-statistical model, and how this model is used to produce cokriging predictions with their uncertainties quantified. Validation strategies are described in Section 2.5, and in Section 3 the strategies are applied to the coSIF data product. We give conclusions and discuss future research directions in Section 4.

2. Materials and Methods

Our multivariate spatial-statistical-prediction framework involves four major steps: (1) Obtain multivariate, monthly remote sensing datasets that are spatially indexed (Sections 2.1 and 2.2). (2) Use the multivariate spatial data to estimate covariance and cross-covariance functions of a multivariate spatial-statistical model (Section 2.3). (3) Use the multivariate spatial (cross-) covariance functions to predict the process of interest along with uncertainties (root-mean-squared prediction errors) on a fine-resolution ($0.05^\circ \times 0.05^\circ$) regular grid (Section 2.4). (4) Evaluate the predictions and their uncertainties using a variety of statistical validation metrics (Section 2.5). When applied to SIF and XCO₂, we obtain an estimated, de-noised, gap-filled SIF data product, coSIF, which includes uncertainty quantification. These methods and the specific datasets used to create coSIF are detailed in the following subsections.

2.1. Datasets Used

Production of the coSIF data product requires both observational and auxiliary datasets.

Observational Datasets: OCO-2 SIF and XCO₂

OCO-2 is a NASA satellite that launched in July 2014 and actively observes two main quantities: SIF and XCO₂ [3,33]. Currently, OCO-2 yields the finest spatial resolution of all spaceborne SIF observations [13], and it is often used as a benchmark for the validation of SIF data products (e.g., [14,20,24]). The OCO-2 satellite has a sun-synchronous, 98.8-min polar orbit with roughly a 13:36 local crossing time and a 16-day ground-track repeat cycle [34]. Observations are made in one of three modes: “nadir”, “glint”, and “target”. In nadir mode, the viewing zenith angle (VZA) is near zero, and the instrument views the ground directly beneath the satellite, while the glint mode generally has higher VZAs as it tracks high reflectance from Earth’s surface. Target mode has VZAs that can vary substantially, since various locations on Earth’s surface are targeted when in this mode. Each mode operates at a fine spatial resolution; for example, swath widths are approximately 10 km with eight measurements across the track, each about 1.3 km \times 2.25 km (across \times along track) in nadir mode [33].

For use among the broader scientific community, NASA publishes OCO-2 datasets as collections of “Lite” files. Lite files contain bias-corrected, daily Level 2 (non-gridded) data at the same spatio-temporal resolution as that of the original observations. We obtained version 10r Lite files for SIF [35] and XCO₂ [36], which are available through NASA’s Goddard Earth Sciences Data and Information Services Center (GES DISC), each from September 2014 to February 2022 inclusive. SIF measurements are retrieved in Watts per square meter, per steradian, and per micrometer (hereafter $\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$) at wavelengths of 757 nm and 771 nm. From these wavelengths, estimates of SIF are produced at a wavelength of 740 nm. As recommended in [13], we use the SIF estimates at 740 nm, which have a lower error overall than the measurements at the 757 nm and 771 nm wavelengths [3]. Further, we only retain observations associated with quality flags 0 and 1 for the Level 2 SIF data. With regard to the Level 2 XCO₂ data, which are available in parts per million (ppm), we only retain those associated with quality flag 0. Estimates of measurement error standard deviations accompany both the SIF measurements and the XCO₂ measurements; we square these measurement error standard deviations to obtain the corresponding measurement error variances.

Auxiliary Dataset: MODIS LCC

From 2001, the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Climate Modeling Grid (MCD12C1) provides yearly land-cover classification (LCC) at a 0.05-degree spatial resolution over the globe. We obtained MCD12C1 Version 6.1 [37] for 2021 to produce a binary mask of land-based prediction locations. Following the International Geosphere-Biosphere Programme (IGBP) classification scheme, we use values not marked as “water” to create the land mask over North America.

2.2. Data Preparation

We first group the daily, spatially irregular SIF and XCO₂ measurements and their corresponding measurement error variances by month and average both the measurements and the measurement error variances onto a regular grid with a 0.05-degree spatial resolution (around 5.6 km × 5.6 km at the Equator), which aligns with the MODIS Climate Modeling Grid. This resolution is the same order of magnitude as that of the OCO-2 footprint and, by averaging multiple measurements, we effectively reduce the measurement error at this resolution [1]. For illustration purposes, we restrict the gridded data used in our analysis to a large region over North America within latitude and longitude boundaries of 22 to 58 degrees North and −125 to −65 degrees East, respectively. This region, which includes the conterminous United States, is similar to that considered by [24], and features heterogeneous land-cover types over which to evaluate the coSIF data product.

As an indicator of photosynthetic activity, SIF is expected to exhibit an inverse relationship with XCO₂. Indeed, the monthly time series in Figure 1 show that SIF and XCO₂ follow strong seasonal patterns that are shifted relative to one another, with the annual maxima in SIF occurring about one month before the annual minima in XCO₂. To leverage this dependence, we use both SIF data and XCO₂ data one month ahead in a multivariate spatial statistical model. We present our results for four months in 2021 to represent the four seasons in North America. For SIF, we use February, April, July, and October 2021, and hence we group these with the XCO₂ data from March, May, August, and November 2021, respectively. For each pair of months, the columns of Figure 2 offer histograms of the OCO-2 data at a 0.05-degree resolution, along with the number of data points available in each month. As in Figure 1, Figure 2 shows that larger XCO₂ values follow months in which SIF values are small, and that XCO₂ decreases after months in which SIF values are larger. Additionally, Figure 2 shows that there is more variability in the summer months: July (SIF) and August (XCO₂). This variability is due to increased SIF activity in certain productive regions, which in turn leads to additional spatial variability in XCO₂. Figure 3 shows the spatial distribution of the OCO-2 data at a 0.05-degree resolution for the summer months. Notice the spatial variability in both SIF and XCO₂, as well as the large gaps where there are no data.

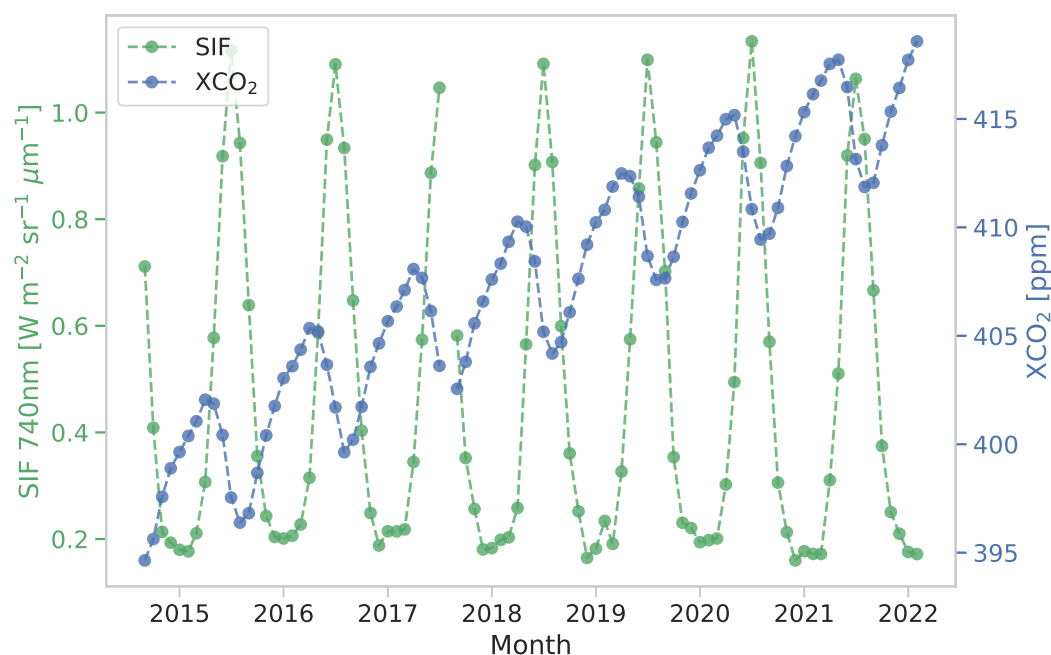


Figure 1. Time series of OCO-2 SIF and XCO₂ averaged by month for data available between September 2014 and February 2022 over North America. The gap in both datasets near the middle of 2017 is due to a time period when the OCO-2 satellite was offline.

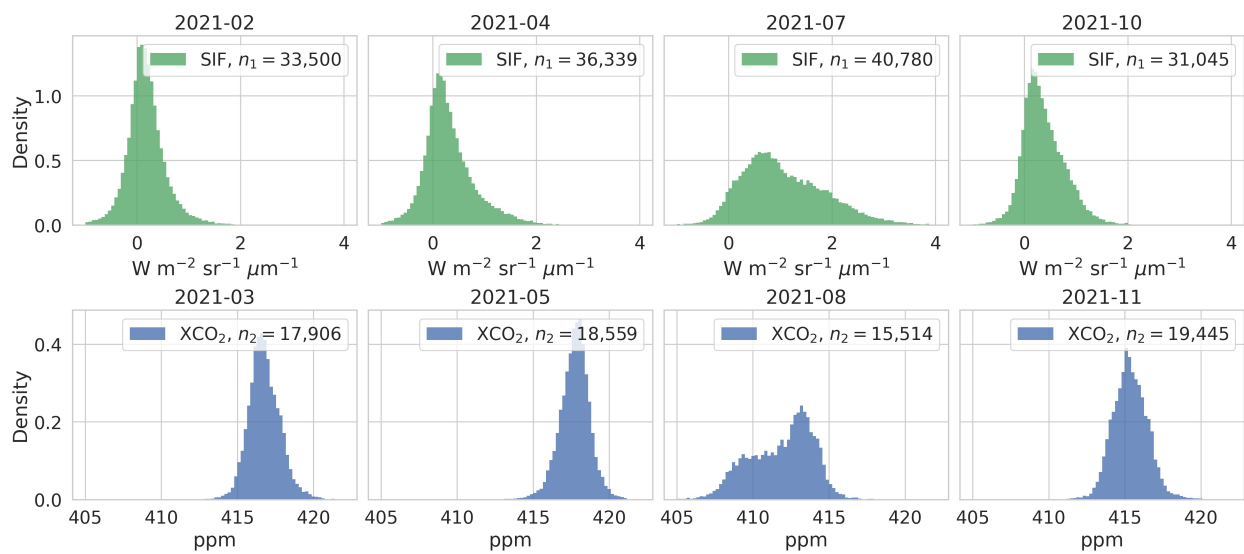


Figure 2. Histograms of OCO-2 data over North America averaged over a month and gridded to 0.05-degree resolution. **Top panels:** SIF in February, April, July, and October 2021. **Bottom panels:** XCO₂ in March, May, August, and November 2021. Panel legends indicate the number of data points available in each month.

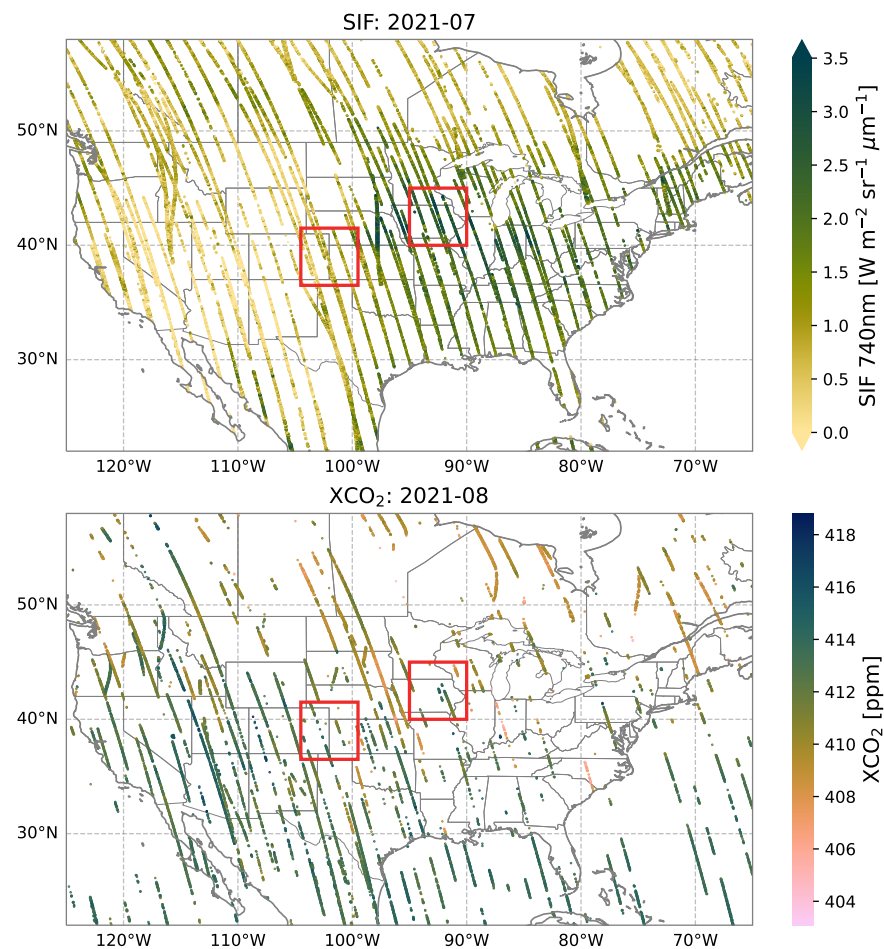


Figure 3. OCO-2 data over North America averaged over a month and gridded to 0.05-degree resolution. **Top:** SIF in July 2021. **Bottom:** XCO₂ in August 2021. The 0.05-degree cells containing the data are enlarged for visibility. The red squares outline two validation blocks considered in Section 3.3.

2.3. Modeling Multivariate Spatial Dependence between SIF and XCO₂

In what follows, we discuss the modeling approach for a single pair of months (e.g., SIF in July and XCO₂ in August). Let D be the spatial region of interest shown in Figure 3 and described in Section 2.2, tessellated by 0.05-degree-resolution grid cells. Then, $\mathbf{s} \in D$ refers to the centroid \mathbf{s} of a grid cell in D . The true SIF process averaged over a given month is $\{Y_1(\mathbf{s}) : \mathbf{s} \in D\}$, and the true XCO₂ process averaged over the following month is $\{Y_2(\mathbf{s}) : \mathbf{s} \in D\}$, but the OCO-2 satellite observations are imperfect and incomplete on D . We denote the gridded and monthly averaged SIF and XCO₂ data (subsequently simply referred to as “data”) by the vectors \mathbf{Z}_1 and \mathbf{Z}_2 , respectively, with $\mathbf{Z}_i \equiv (Z_{i1}, \dots, Z_{in_i})^T$, $i = 1, 2$, where there are n_1 grid cells that contain SIF data and n_2 grid cells that contain XCO₂ data. In Figure 3 (Top), $n_1 = 40,780$, and in Figure 3 (Bottom), $n_2 = 15,514$; see also Figure 2. In the spatial domain D , we model the noisy data value Z_{il} at grid cell location $\mathbf{s}_{il} \in D$ as

$$Z_{il} = Y_i(\mathbf{s}_{il}) + \epsilon_{il}; \quad l = 1, \dots, n_i, \quad i = 1, 2, \quad (1)$$

where $\{\epsilon_{il}\}_{il}$ is a set of (non-stationary) independent measurement errors, with mean zero and variances $\{\sigma_{\epsilon,il}^2\}_{il}$ obtained from the Lite files and averaged as described in Section 2.2. The latent processes $Y_1(\cdot)$ and $Y_2(\cdot)$ are defined on all $\mathbf{s} \in D$ and are modeled as

$$Y_i(\mathbf{s}) = \mu_i(\mathbf{s}) + \psi_i(\mathbf{s}); \quad i = 1, 2. \quad (2)$$

Here, $\mu_i(\cdot) \equiv E(Y_i(\cdot))$ represents the deterministic trend component, which can be thought of as *large-scale variation*, while $\psi_i(\cdot)$ is a random process with zero mean that accounts for any remaining spatial variation. For $\mathbf{s} \in D$, we write

$$\psi_i(\mathbf{s}) = \eta_i(\mathbf{s}) + \omega_i(\mathbf{s}); \quad i = 1, 2, \quad (3)$$

where $\eta_i(\cdot)$ is a mean-zero Gaussian process that captures the *smooth small-scale variation*, and $\omega_i(\cdot)$ is a process uncorrelated at the grid scale with mean zero and constant variance $\sigma_{\omega,i}^2$ that accounts for *micro-scale variation*.

Because our objective is spatial prediction, we take a relatively parsimonious approach for modeling $\mu_i(\cdot)$ to avoid over-fitting the data. At each location $\mathbf{s} \in D$, estimates for the deterministic trend components, $\hat{\mu}_i(\mathbf{s})$, $i = 1, 2$, are obtained through regression on a vector of covariates, $\mathbf{x}_i(\mathbf{s})$, as follows:

$$\hat{\mu}_i(\mathbf{s}) = \mathbf{x}_i(\mathbf{s})^T \hat{\boldsymbol{\beta}}_i; \quad i = 1, 2, \quad (4)$$

where $\hat{\boldsymbol{\beta}}_i$ is a vector of ordinary-least-squares-estimated regression coefficients obtained from fitting $\{\mathbf{x}_i(\mathbf{s}_{il})^T \boldsymbol{\beta}_i\}_l$ to $\{Z_{il}\}_l$ separately for $i = 1, 2$. For each $i = 1, 2$, the elements of $\mathbf{x}_i(\cdot)$ include an intercept term, and k_i spatial bisquare basis functions (e.g., [38]), here used to express a spatial trend. We obtained the bisquare basis functions from the package *FRK* in R [39] and chose them such that their spatial supports cover the data region with centers arranged in a regular, 6×10 grid. Some basis functions over water were not appropriate for the land-based SIF data; three were identified and withheld from $\mathbf{x}_1(\cdot)$, the vector of covariates for SIF. Hence, $k_1 = 57$ and $k_2 = 60$. The locations of the basis functions and the extent of their spatial supports are displayed in Figure 4.

After the estimated spatial trend is removed, the behavior of the small-scale spatial variation is modeled based on the vectors of spatial residuals $\mathbf{R}_i = (R_{i1}, \dots, R_{in_i})^T$, with elements computed from the data as

$$R_{il} = Z_{il} - \hat{\mu}_i(\mathbf{s}_{il}); \quad l = 1, \dots, n_i, \quad i = 1, 2. \quad (5)$$

To mitigate bias and trend contamination often found in the residuals ([40], Section 2.2.5), we standardize each set of values using the empirical mean, $\hat{\mu}_{R,i}$, and standard deviation, $\hat{\sigma}_{R,i}$, as follows:

$$R_{il}^* = \frac{R_{il} - \hat{\mu}_{R,i}}{\hat{S}_{R,i}}, \text{ for } \hat{\mu}_{R,i} = \frac{1}{n_i} \sum_{l=1}^{n_i} R_{il}, \text{ and } \hat{S}_{R,i} = \sqrt{\frac{1}{n_i - 1} \sum_{l=1}^{n_i} (R_{il} - \hat{\mu}_{R,i})^2}, \quad (6)$$

$l = 1, \dots, n_i$, and $i = 1, 2$. We then model the standardized spatial residuals as

$$R_{il}^* = \eta_i^*(\mathbf{s}_{il}) + \omega_i^*(\mathbf{s}_{il}) + \epsilon_{il}^*; \quad l = 1, \dots, n_i, \quad i = 1, 2, \quad (7)$$

where $\eta_i^*(\cdot)$ is the smooth small-scale variation and $\omega_i^*(\cdot)$ is the micro-scale variation with mean zero and variance $\sigma_{\omega,i}^2 / \hat{S}_{R,i}^2$ on the standardized scale for the standardized residuals $\{R_{il}^*\}_{il}$, and $\epsilon_{il}^* = \epsilon_{il} / \hat{S}_{R,i}$ are spatially varying independent measurement errors on the standardized scale, obtained from the spatially varying independent measurement errors $\{\epsilon_{il}\}_{il}$ as defined for Equation (1).

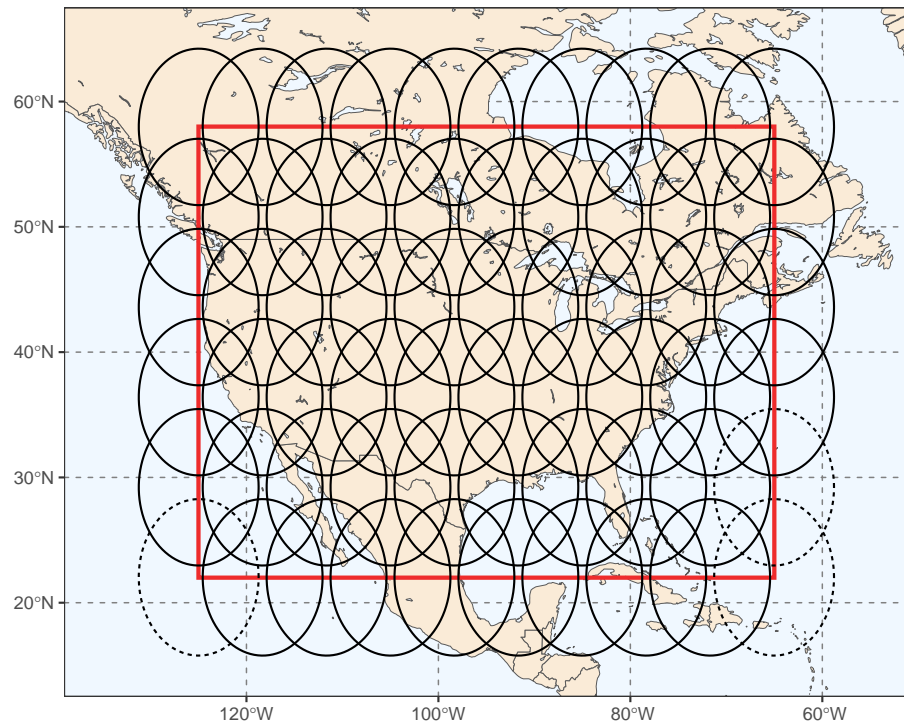


Figure 4. Solid lines outline the spatial supports of the bisquare basis functions for the spatial trend that we use for the large-scale variation of both SIF and XCO₂, while dashed lines outline the spatial supports of additional basis functions that we use for XCO₂. Note that the spatial supports are circular (the bisquare function is radially symmetric), but they appear as ovals due to the mapping projection. The red rectangle outlines the spatial domain from which OCO-2 data are taken.

The smooth small-scale-variation processes $\eta_1^*(\cdot)$ (for SIF) and $\eta_2^*(\cdot)$ (for XCO₂) are the focus of our multivariate spatial modeling efforts. Between any two spatial locations $\mathbf{s}, \mathbf{u} \in D$, we model the underlying spatial dependence and cross-dependence in the standardized small-scale-variation processes with a bivariate covariance function $C_{ij}^*(\mathbf{s}, \mathbf{u}) \equiv \text{cov}(\eta_i^*(\mathbf{s}), \eta_j^*(\mathbf{u}))$, for $i, j \in \{1, 2\}$. For $C_{ij}^*(\cdot, \cdot)$, we selected the full bivariate Matérn model [41,42], popular for its flexible parameterization of spatial smoothness. For spatial lag $h = \|\mathbf{s} - \mathbf{u}\|$, the model is symmetric and isotropic:

$$C_{ij}(h) = C_{ji}(h) = \rho_{ij} \sigma_{ii} \sigma_{jj} M(h; v_{ij}, \ell_{ij}),$$

$$M(h; v_{ij}, \ell_{ij}) = \frac{2^{1-v_{ij}}}{\Gamma(v_{ij})} \left(\frac{h \sqrt{2v_{ij}}}{\ell_{ij}} \right)^{v_{ij}} K_{v_{ij}} \left(\frac{h \sqrt{2v_{ij}}}{\ell_{ij}} \right), \quad (8)$$

for $i, j \in \{1, 2\}$, where $\sigma_{ii} > 0$ is a scale parameter and ρ_{ij} is a (cross-) correlation coefficient, with $\rho_{ij} = 1$ when $i = j$ and $\rho_{ij} \in [-1, 1]$ otherwise. The function $M(\cdot; \nu_{ij}, \ell_{ij})$ is the Matérn correlation function [43,44], with $\nu_{ij} > 0$ and $\ell_{ij} > 0$ denoting the smoothness and correlation length, respectively. Further, $\Gamma(\cdot)$ is the gamma function and $K_{\nu_{ij}}$ is a modified Bessel function of the second kind of order ν_{ij} ([45], Section 10.2). As detailed in [41], the values of $\{\nu_{ij}\}_{ij}$, $\{\ell_{ij}\}_{ij}$, and $\{\rho_{ij}\}_{ij}$ that result in a positive-semidefinite bivariate covariance function belong to a constrained parameter space. Finally, note that since \mathbf{s} and \mathbf{u} are locations on Earth's surface, we avoid Euclidean distance and let $\|\mathbf{s} - \mathbf{u}\|$ be the chordal distance between \mathbf{s} and \mathbf{u} (e.g., [46]).

To estimate the parameters of the full bivariate Matérn covariance model, we use the standardized residuals, $\mathbf{R}_i^* = (R_{i1}^*, \dots, R_{im_i}^*)^T$, $i = 1, 2$, given by Equation (6). We first compute empirical semivariograms and cross-semivariograms, and then we fit the parameters of the covariance and cross-covariance functions simultaneously using a multivariate extension of the weighted-least-squares approach developed in [47]. This process is described in Appendix A.1; a check for the validity of the fitted parameters is also established there.

2.4. Generating the Spatially Contiguous coSIF Data Product with Quantified Uncertainties Based on Cokriging

From the noisy and relatively sparse spatial data \mathbf{Z}_1 (SIF) and \mathbf{Z}_2 (XCO₂), we seek the optimal spatial prediction, $\hat{Y}_1(\cdot)$, and its root-mean-squared prediction error, $RMSPE(\hat{Y}_1(\cdot))$, for the latent SIF process, $Y_1(\cdot)$, at each of the 0.05-degree land-based grid cell locations $\{\mathbf{s}_1, \dots, \mathbf{s}_m\}$, $m = 591,119$, in the domain D spanning North America, shown in Figure 3. In generating our SIF data product, we produce spatial predictions at the 0.05-degree resolution for two main reasons: (1) 0.05 degrees is emerging as the standard for Level 3 SIF data products (e.g., [20,23,25–27]); (2) predictions are at the same resolution as the pre-processed data (Section 2.2). Since the spatial covariance model is formulated at the 0.05-degree resolution, upscaling to coarser resolutions is straightforward (e.g., [48]).

The practice of leveraging interdependence between multiple related processes to optimally predict one of those processes is known in geostatistics as cokriging (e.g., [32]). Recall from Equation (2) that we model the spatial process $Y_1(\cdot)$ as the combination of large-scale variation and any remaining spatial variation through the decomposition $Y_1(\mathbf{s}) = \mu_1(\mathbf{s}) + \psi_1(\mathbf{s})$, for $\mathbf{s} \in D$. Having standardized the SIF and XCO₂ data as in Equation (6), our task in cokriging is to optimally predict $\psi_1^*(\cdot) = (Y_1(\cdot) - \hat{\mu}_1(\cdot) - \hat{\mu}_{R,1}) / \hat{\sigma}_{R,1}$, which is the standardized version of $\psi_1(\cdot)$ defined in Equation (3). From predictions of $\psi_1^*(\cdot)$, we can obtain predictions of $Y_1(\cdot)$ on the original scale by reversing the standardization with a simple linear transformation.

As in Section 2.3, information about the behavior of $\psi_1^*(\cdot)$ (for SIF) and $\psi_2^*(\cdot)$ (for XCO₂) is taken from the standardized spatial residuals \mathbf{R}_1^* and \mathbf{R}_2^* described in Equation (6). For a given set of months, \mathbf{R}_1^* and \mathbf{R}_2^* have a combined size on the order of 50,000 elements (see Figure 2), so that traditional cokriging is computationally intractable due to a required matrix inversion that is shown in Equation (11) below. However, spatial prediction is a relatively local operation ([40], Section 3.4), so for a given prediction location \mathbf{s}_0 , we use a local neighborhood defined by the 150 values whose corresponding locations are closest in chordal distance to \mathbf{s}_0 , respectively, in \mathbf{R}_1^* and \mathbf{R}_2^* . Let \mathbf{R}_{01}^* and \mathbf{R}_{02}^* be these 150-dimensional vectors of local residuals, and define $\mathbf{R}_0^* = (\mathbf{R}_{01}^{*T}, \mathbf{R}_{02}^{*T})^T$. The bivariate spatial variation across the locations associated with \mathbf{R}_{0i}^* and \mathbf{R}_{0j}^* , $i, j \in \{1, 2\}$, is described by the variance-covariance matrix,

$$\text{cov}(\mathbf{R}_0^*) \equiv \Sigma_0 = \begin{pmatrix} \mathbf{C}_{0,11} & \mathbf{C}_{0,12} \\ \mathbf{C}_{0,21} & \mathbf{C}_{0,22} \end{pmatrix} + \begin{pmatrix} \frac{\sigma_{\omega,1}^2}{\hat{\sigma}_{R,1}^2} \mathbf{I} + \text{diag}(\sigma_{\epsilon,01}^*) & \mathbf{0} \\ \mathbf{0} & \frac{\sigma_{\omega,2}^2}{\hat{\sigma}_{R,2}^2} \mathbf{I} + \text{diag}(\sigma_{\epsilon,02}^*) \end{pmatrix}, \quad (9)$$

where each block $\mathbf{C}_{0,ij}$ is a 150×150 (cross-) covariance matrix having elements given by $C_{ij}(\cdot)$ from Equation (8) with estimated parameters substituted in; \mathbf{I} is the 150×150

identity matrix; and $\sigma_{\epsilon,0i}^*$ has elements $\{\sigma_{\epsilon,0il}^2 / \hat{S}_{R,i}^2 : l = 1, \dots, 150\}$ from the definition of ϵ_{il}^* in Equation (7), corresponding to locations associated with \mathbf{R}_{0i}^* . Similarly, we obtain the covariance vector $\mathbf{c}_0 = (\mathbf{c}_{01}^T, \mathbf{c}_{02}^T)^T$, with elements of the j -th sub-vector ($j = 1, 2$) given by $\mathbf{c}_{0j} \equiv (\text{cov}(\psi_1^*(\mathbf{s}_0), R_{0jl}^*) : l = 1, \dots, 150)^T$, where

$$\text{cov}(\psi_1^*(\mathbf{s}_0), R_{0jl}^*) = C_{1j}(h_{0l}) + \frac{\sigma_{\omega,1}^2}{\hat{S}_{R,1}^2} \mathbf{1}\{j = 1, h_{0l} = 0\}; \tag{10}$$

$h_{0l} = \|\mathbf{s}_0 - \mathbf{s}_l\|$; \mathbf{s}_l is the spatial location associated with R_{0jl}^* ; $\|\cdot\|$ denotes chordal distance; and $\mathbf{1}\{\cdot\}$ is the indicator function. Notice the absence of measurement error variances in Equation (10), since we are predicting the (noiseless) residual process.

Then, by minimizing the mean squared prediction error (MSPE) over all linear predictors of $\psi_1^*(\mathbf{s}_0)$ given \mathbf{R}_0^* , we obtain the cokriging prediction and minimized MSPE (Equations (6) and (7) in [32]):

$$\hat{\psi}_1^*(\mathbf{s}_0) = \mathbf{c}_0^T \Sigma_0^{-1} \mathbf{R}_0^*, \tag{11}$$

$$\text{MSPE}(\hat{\psi}_1^*(\mathbf{s}_0)) = \frac{\sigma_{\omega,1}^2}{\hat{S}_{R,1}^2} + C_{11}(0) - \mathbf{c}_0^T \Sigma_0^{-1} \mathbf{c}_0. \tag{12}$$

The estimator given in Equation (11) is an unbiased linear predictor for $\psi_1^*(\mathbf{s}_0)$ under squared error loss, and the uncertainty measure we use is the square root of the MSPE given in Equation (12); notice that $\hat{\psi}_1^*(\mathbf{s}_0)$ is an approximation to the optimal predictor, $E(\psi_1^*(\mathbf{s}_0) \mid \mathbf{R}^*)$, since it is based on a subset of the available data. To obtain predictions and uncertainty quantification on the original scale of the latent SIF process, we reverse the standardization to yield

$$\hat{Y}_1(\mathbf{s}_0) = \hat{\mu}_{R,1} + \hat{\mu}_1(\mathbf{s}_0) + \hat{\psi}_1^*(\mathbf{s}_0) \hat{S}_{R,1}, \tag{13}$$

$$\text{RMSPE}(\hat{Y}_1(\mathbf{s}_0)) = \sqrt{\text{MSPE}(\hat{\psi}_1^*(\mathbf{s}_0)) \hat{S}_{R,1}^2}. \tag{14}$$

Operationally, these predictions and their root-mean-squared prediction errors, given by Equations (13) and (14), respectively, can be produced at each location in a highly parallelizable manner. Hence, for each month, our 0.05-degree resolution coSIF data product consists of $\{\hat{Y}_1(\mathbf{s}_0) : \mathbf{s}_0 \in D\}$ and $\{\text{RMSPE}(\hat{Y}_1(\mathbf{s}_0)) : \mathbf{s}_0 \in D\}$.

2.5. Statistical Validation of the coSIF Data Product

Cross-validation is a widely used method for evaluating the predictive performance of statistical models [49]. The method seeks to evaluate model predictions by withholding some subset of the available data when fitting the model and making predictions, using the withheld subset for validation, and repeating with other subsets. Often, one does leave-one-out cross-validation, where data are withheld one at a time, or K -fold cross-validation, where data are withheld in “folds”. However, there are two important considerations in the spatial setting. First, fitting multiple spatial models to different subsets of data can quickly become computationally prohibitive if the number of subsets is large. Second, the spatial arrangement of the withheld validation data is important. For example, the locations of the withheld data may be chosen at random, or they may be chosen so that a region or block of data are missing. The latter reflects a common scenario in remote-sensing applications where a satellite’s observations are hindered by blocks of clouds or retrieval malfunctions. In a spatial context, missing blocks of data also lead to a more rigorous validation test, since spatial prediction over large spans of unobserved regions is generally a harder problem than that over small unobserved regions [50].

Given the challenges of cross-validation in the spatial setting, we instead evaluate the performance of geostatistical methods (cokriging and kriging) for producing our

data product through validation (rather than cross-validation) within blocks of withheld SIF data. Specifically, for a given block with spatial support A , the SIF validation data $\mathbf{Z}_{1A} = (Z_{11}, \dots, Z_{1n_A})^T$ associated with spatial locations $\{\mathbf{s}_1, \dots, \mathbf{s}_{n_A}\} \subset A$ are withheld from the available data, leaving \mathbf{Z}^{-A} , of dimension $(n_1 - n_A) + n_2$, for use in the prediction framework. Because the withheld validation data, \mathbf{Z}_{1A} , contain measurement error, we seek to predict \mathbf{Z}_{1A} rather than the noiseless latent vector \mathbf{Y}_{1A} in the validation setting [28]. It is easy to see that the predicted values, $\hat{\mathbf{Z}}_{1A} = (\hat{Z}_{11}, \dots, \hat{Z}_{1n_A})^T$, are identical to $\hat{\mathbf{Y}}_{1A}$ given by Equation (13). However, the RMSPEs of the elements of $\hat{\mathbf{Z}}_{1A}$ are different from those of $\hat{\mathbf{Y}}_{1A}$. Specifically,

$$RMSPE(\hat{Z}_{1a}) = \sqrt{RMSPE(\hat{Y}_1(\mathbf{s}_a))^2 + \sigma_{\epsilon,1a}^2}; \quad a = 1, \dots, n_A. \quad (15)$$

To assess the advantage of leveraging geostatistical dependence between SIF and XCO₂, we use both a cokriging and a kriging version of this procedure to produce predictions of the withheld data. To assess the advantage of incorporating spatial dependence within SIF in the cokriging and kriging models, we compare the validation predictions from these geostatistical models to those from a trend-surface-only model (a special case of kriging) that does not leverage spatial dependence. Kriging and trend-surface predictions are based on SIF data alone, whereas cokriging predictions are based on SIF and XCO₂ data. Further details are given in Appendix A.2.

We select a number of criteria to evaluate these predictions at validation locations. To characterize the marginal accuracy of the predictions, we consider average prediction error, or bias (BIAS), and root-average-squared prediction error (RASPE), given by

$$BIAS \equiv \frac{1}{n_A} \sum_{a=1}^{n_A} (\hat{Z}_{1a} - Z_{1a}), \quad (16)$$

$$RASPE \equiv \sqrt{\frac{1}{n_A} \sum_{a=1}^{n_A} (\hat{Z}_{1a} - Z_{1a})^2}. \quad (17)$$

A benefit of a geostatistical prediction framework is that, at each prediction location $\mathbf{s}_a \in A$, we obtain a predictive distribution, say F_{1a} , rather than a predicted value alone. Because we assume the data are Gaussian, each F_{1a} is a Gaussian distribution with mean $E(Z_{1a} | \mathbf{Z}^{-A}) \approx \hat{Z}_{1a}$ and variance $\text{var}(Z_{1a} | \mathbf{Z}^{-A}) \approx RMSPE(\hat{Z}_{1a})^2$; see Equations (13) and (15), respectively. Recall that the approximations are due to the predictor \hat{Z}_{1a} using only the nearest 150 data points for SIF and for XCO₂, and that the 150 data points for SIF are outside of the validation block A .

Through the predictive distribution, F_{1a} , we can quantify the uncertainty in our Level 3 coSIF data product. We compare these quantified uncertainties to analogous ones obtained from the univariate kriging model and the trend-surface-only model using scoring rules, which measure how well a probabilistic prediction aligns with what is actually observed. Here, we consider the interval score which assesses calibration and sharpness by favoring narrow prediction intervals and penalizing those that do not contain the true observation. With a 95% two-sided prediction interval given by $[p, q] \subset \mathbb{R}$ for Z_{1a} at validation location $\mathbf{s}_a \in A$, the interval score (INT) is given by

$$INT(p, q; Z_{1a}) \equiv (q - p) + \frac{2}{\alpha}(p - Z_{1a})\mathbf{1}\{Z_{1a} < p\} + \frac{2}{\alpha}(Z_{1a} - q)\mathbf{1}\{Z_{1a} > q\}, \quad (18)$$

for $a = 1, \dots, n_A$, where $\alpha = 0.05$ and $\mathbf{1}\{\cdot\}$ is the indicator function [51]. Another scoring rule that assesses both the first and second moments of a predictive distribution is the Dawid-Sebastiani score (DSS; [52]) defined as

$$DSS(F_{1a}, Z_{1a}) \equiv \frac{(Z_{1a} - E(Z_{1a} | \mathbf{Z}^{-A}))^2}{\text{var}(Z_{1a} | \mathbf{Z}^{-A})} + 2 \log \sqrt{\text{var}(Z_{1a} | \mathbf{Z}^{-A})}, \quad (19)$$

for $a = 1, \dots, n_A$, where recall that F_{1a} is a distribution with mean $E(Z_{1a} | \mathbf{Z}^{-A}) \approx \widehat{Z}_{1a}$ and variance $\text{var}(Z_{1a} | \mathbf{Z}^{-A}) \approx \text{RMSPE}(\widehat{Z}_{1a})^2$. Because F_{1a} is Gaussian, Equation (19) is functionally equivalent to the logarithmic score, which is one of the most-used proper scoring rules [53]. The INT and DSS are computed pointwise, for each validation location, and typically the average of each collection of scores is used to compare the different prediction frameworks, without any regard to the spatial dependence between them.

The validation summaries introduced above focus on an average of pointwise predictions, but since the predictions are spatially correlated within a given block A , it is also important to consider multivariate scoring rules. These scoring rules assess $\widehat{\mathbf{Z}}_{1A}$ as a single multivariate prediction of \mathbf{Z}_{1A} through the corresponding multivariate predictive distribution denoted as F_{1A} . Again, because the process model in Equation (2) is Gaussian, F_{1A} is a multivariate Gaussian distribution with conditional mean vector $E(\mathbf{Z}_{1A} | \mathbf{Z}^{-A})$ and conditional variance-covariance matrix $\text{var}(\mathbf{Z}_{1A} | \mathbf{Z}^{-A}) \equiv \mathbf{V}_{1A}$. Hence, F_{1A} can be assessed using the multivariate DSS:

$$MDSS(F_{1A}, \mathbf{Z}_{1A}) \equiv \left(\mathbf{Z}_{1A} - E(\mathbf{Z}_{1A} | \mathbf{Z}^{-A}) \right)^T \mathbf{V}_{1A}^{-1} \left(\mathbf{Z}_{1A} - E(\mathbf{Z}_{1A} | \mathbf{Z}^{-A}) \right) + \log \det \mathbf{V}_{1A}. \quad (20)$$

The MDSS provides a scoring metric to assess multivariate predictive distributions (e.g., [51], Equation (25)). Our evaluation of Equation (20) in Section 3.3 uses $\widehat{\mathbf{Z}}_{1A}$ to approximate $E(\mathbf{Z}_{1A} | \mathbf{Z}^{-A})$, and we obtain an approximation of the diagonal elements of \mathbf{V}_{1A} , which are needed for the pointwise scores, from the square of Equation (15). The whole variance-covariance matrix \mathbf{V}_{1A} is approximated in Appendix A.3.

3. Results and Discussion

In this section, we provide an overview of our coSIF data product by considering a large region over North America that is within the latitude and longitude boundaries of 22 to 58 degrees North and -125 to -65 degrees East, respectively. We illustrate coSIF for North America in this article, but our framework applies equally to other large continental regions, albeit with different estimates of spatial dependence parameters. In Section 3.1, we present fitted bivariate spatial dependence models for February, April, July, and October 2021 (representing the four seasons), and we find strong spatial cross-dependence between SIF and XCO₂ in July, the month with the most prominent SIF activity. In Section 3.2, we consider the spatial and seasonal patterns of the coSIF-derived predictions and their corresponding prediction uncertainties, with a focus on July 2021. Section 3.3 presents validation metrics for July 2021 where bivariate cokriging, univariate kriging, and trend-surface-only predictions of SIF are compared.

We produce all results on a high-end server with 56 physical cores and 376 gigabytes of memory. For reference, compute times using this architecture are as follows. Fitting the bivariate spatial model takes roughly 10 min for each pair of months. Production of the coSIF data product takes on the order of 12 h for each pair of months; however, the procedure is highly parallelizable and this compute time could be reduced by using multiple servers. In a given validation block, procuring validation predictions and approximating their corresponding multivariate variance-covariance matrix (see Appendix A.3) takes 15 to 30 min, depending on the amount of data in the block.

3.1. Evaluation of Fitted Spatial Models

Recall from Section 2.2 that we consider bivariate spatial models for SIF and XCO₂ (with XCO₂ one month ahead of SIF) for four months in 2021 to represent the four seasons in North America. These models are constructed for SIF in February, April, July, and October 2021, and grouped with the secondary variable XCO₂ in March, May, August, and Novem-

ber 2021. Figure 5 shows the empirical and fitted (cross-) semivariogram models based on the standardized spatial residuals \mathbf{R}_1^* and \mathbf{R}_2^* , for each of the four months; the fitted model parameters for all four months are available in the Supplementary Material.

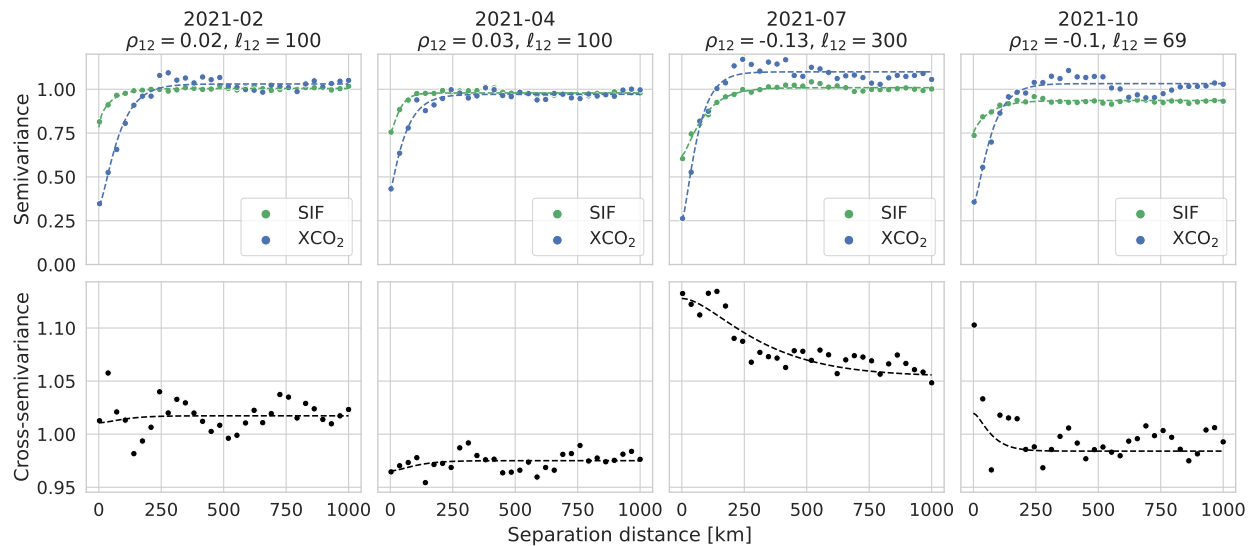


Figure 5. Solid dots show the empirical (cross-) semivariograms and dashed lines show the corresponding fitted models. **Top panels:** Semivariograms for standardized SIF residuals (February, April, July, October) and standardized XCO₂ residuals (March, May, August, November) in 2021. **Bottom panels:** The corresponding cross-semivariograms. Column titles give the estimated cross-correlation coefficient (ρ_{12}) and cross-correlation length (ℓ_{12}) for each model. Chordal distance is shown on the horizontal axes (in km).

In Figure 5, the empirical semivariograms clearly indicate the presence of small-scale spatial dependence in SIF (and XCO₂) not captured by the basis function trend surface. The fitted semivariogram models exhibit relatively small correlation lengths (on the order of 50 to 100 km in all cases). In July 2021, the empirical cross-semivariogram indicates pronounced spatial cross-dependence between the small-scale SIF and XCO₂ latent processes, and this is captured by the fitted cross-semivariogram model with cross-correlation coefficient, $\rho_{12} \approx -0.13$, and cross-correlation length, $\ell_{12} \approx 300$ km. The smaller cross-correlation coefficients and/or cross-correlation lengths in the other three months considered suggest that the cross-dependence between SIF and XCO₂ decreases when GPP, and hence SIF, is small. This suggests that when SIF activity is not prominent (e.g., when SIF values are small in Figure 1), it may be sufficient to de-noise and gap-fill SIF using the special case of kriging. Still, for many applications (e.g., agricultural planning and yield estimation), seasons of prominent SIF activity, such as the late spring and summer, are of primary interest. For months when SIF activity is prominent, we suggest operational decisions between cokriging and kriging be made based on the presence of dependence in the empirical cross-semivariogram for SIF and XCO₂. In what follows, we consider summer SIF activity through a focus on July 2021 where we have identified cross-dependence between SIF and XCO₂; results for the other three months shown in Figure 5 are provided in the Supplementary Material.

3.2. Level 3 coSIF Predictions and Quantified Uncertainties

Next, we consider the spatial characteristics of coSIF across D , the North American region of interest. Figure 6 shows the de-noised and gap-filled coSIF data product in July 2021, with SIF predictions given at a 0.05-degree resolution in the top panel, and the corresponding RMSPEs (i.e., prediction standard errors) in the bottom panel. The spatial predictions show that coSIF is prominent across the eastern half of the region, particularly in the North American Corn Belt, and less so across the western half of the region. The striped

diagonal pattern of the prediction standard errors are representative of the OCO-2 orbit tracks (e.g., Figure 3). This is expected from the local behavior of cokriging, which leads to smaller prediction standard errors at grid locations in areas where the data are dense and larger prediction standard errors at grid locations where the data are sparse or non-existent.

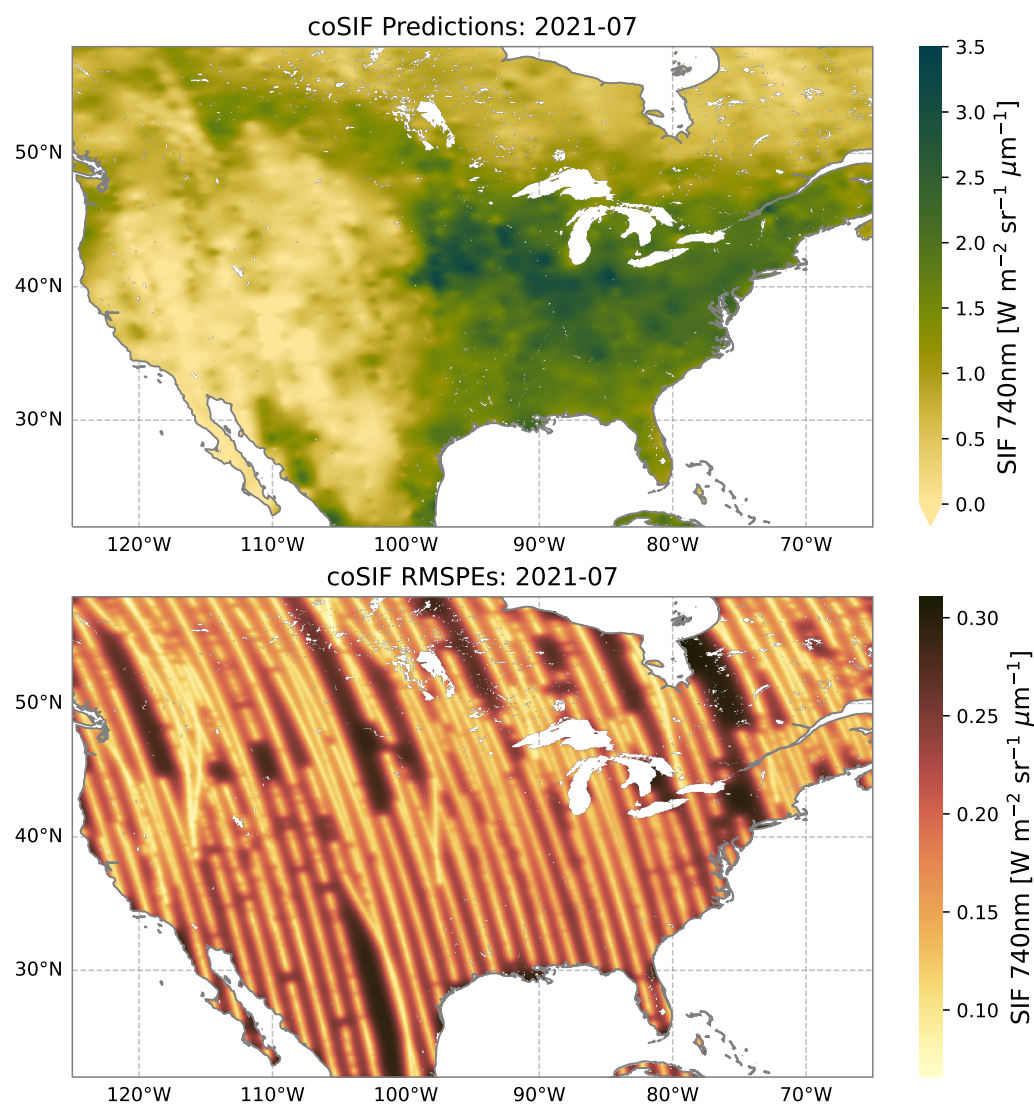


Figure 6. The coSIF data product for July 2021 at 0.05-degree resolution. **Top:** The coSIF predictions given by Equation (13). **Bottom:** The corresponding root-mean-squared prediction errors given by Equation (14). All units are $\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$. The color white indicates the absence of predictions over large bodies of water.

The observed seasonal variation of SIF in Figure 1 is reproduced in the coSIF data product. When comparing Figure 6 to similar figures for different months (given in Figures S1–S3 in the Supplementary Material), coSIF predictions are lower in the spring (April), fall (October) and winter (February), and higher in the summer (July). This seasonal variation is also reflected in the prediction standard errors, which scale with the coSIF predictions; there is more variability in the residuals given by Equation (5) in the summer (July), so the prediction standard errors are larger. Still, for all months considered here, the coSIF RMSPEs are on average about four times smaller than the gridded measurement-error standard deviations $\{\sigma_{e,1l}\}_l$ obtained from the Level 2 Lite files as described for Equation (1) (see Table S1 in the Supplementary Material for uncertainty reduction estimates in all four months).

3.3. Validation of coSIF and Comparison with Simpler Methods

We evaluate the cokriging method for producing the coSIF data product by performing validation on data withheld in spatially contiguous blocks. As discussed in Section 2.3, the bivariate spatial covariance model used for cokriging is able to leverage cross-spatial dependence between SIF and XCO₂, beyond the spatial dependence within SIF alone. Hence, we perform the same validation procedure for the special case of kriging predictions of SIF (where no relationship between SIF and XCO₂ is assumed). Prior to obtaining kriging predictions, we re-fit the univariate spatial covariance model for SIF using the SIF-only semivariogram. We also compare these geostatistical predictions to the special case of trend-surface-only predictions (where it is assumed further that there is no spatial dependence in the SIF residuals); see Appendix A.2.

As opposed to leave-one-out cross-validation, or K -fold cross-validation with subsets of validation data selected at random, validation with subsets of data missing in blocks is especially challenging in the geostatistical setting, since spatial predictions depend heavily on the quantity, proximity, and precision of nearby data. Here, we separately consider two 5-degree by 5-degree validation blocks outlined by the red squares in Figure 3. One of these blocks has extents [−95°E, −90°E, 40°N, 45°N] and covers a large section of Iowa in the center of the North American Corn Belt where SIF activity is especially prominent in July. The other block has extents [−104°E, −99°E, 36.5°N, 41.5°N] and consists predominately of cropland in the plains of Colorado and Kansas where SIF activity is more moderate in July. The spatial supports of these two validation blocks (Corn Belt and Cropland) are representative of large unobserved regions that realistically can occur in OCO-2 datasets (e.g., Figure 3).

Recall from Section 2.5 that we produce a prediction, \hat{Z}_{1a} , and prediction standard error, $RMSPE(\hat{Z}_{1a})$, at each validation location \mathbf{s}_a , for $a = 1, \dots, n_A$, in a given validation block with spatial support A where data are withheld from predictions. The validation standardized residuals are given by

$$\left\{ \frac{\hat{Z}_{1a} - Z_{1a}}{RMSPE(\hat{Z}_{1a})} : a = 1, \dots, n_A \right\}. \quad (21)$$

As a diagnostic check, it is important to explore whether these residuals are standard-Gaussian distributed, which is an indicator that the model is a good fit to the data in our setting. Quantile-quantile (Q-Q) plots of validation standardized residuals in July 2021 are shown in Appendix B (Figure A1) for the two validation blocks. For both blocks, the Q-Q plots that result from using the cokriging, kriging, and trend-surface-only methods are all quite similar and show that the residuals are approximately Gaussian, although each set of residuals exhibits slightly heavy-tailed behavior relative to the standard Gaussian distribution. This initial diagnostic result indicates that more sensitive validation metrics are needed to distinguish between the different prediction methods.

We now evaluate the validation metrics outlined in Section 2.5 (BIAS, RASPE, INT, DSS, and MDSS). Table 1 presents these metrics for the validation predictions and prediction standard errors produced for July 2021 using the cokriging, kriging, and trend-surface-only methods in both validation blocks. With the exception of average prediction error (BIAS), which should be as close to zero as possible, all metrics are negatively oriented scores (lower is better). We find that BIAS is reasonably negligible for all methods in both validation blocks. The root-average-squared prediction error (RASPE) also exhibits similar results across the different prediction methods. The RASPE for cokriging and kriging is slightly lower than that for the trend-surface-only method, suggesting that leveraging the spatial dependence identified in Section 3.1 yields slightly improved prediction accuracy.

While prediction accuracy is important, prediction uncertainty must also be considered when performing probabilistic prediction. We summarize the distribution of our predictions using the average interval score (INT) and the average Dawid-Sebastiani score (DSS), which

offer different assessments on the quality of the prediction uncertainties. Between the two validation blocks, INT differs slightly as follows: the 95% prediction intervals are best calibrated when using cokriging for the Corn Belt validation block and when using kriging for the Cropland validation block. However, cokriging is found to offer the best characterization of uncertainty with respect to DSS in both blocks. While cokriging and kriging are similar with respect to INT and DSS, the performance of the trend-surface-only method is noticeably worse.

Table 1. Validation metrics for the Corn Belt and Cropland $5^\circ \times 5^\circ$ validation blocks in July 2021. Bold typeface indicates the best performing method according to each metric. Locations of the validation blocks are shown in Figure 3.

| Block | Method | BIAS | RASPE | INT | DSS | MDSS |
|-----------|---------------|--------------|-------------|-------------|--------------|----------------|
| Corn Belt | Cokriging | −0.06 | 0.58 | 2.85 | −0.13 | −311.94 |
| | Kriging | −0.01 | 0.59 | 2.93 | −0.12 | −308.63 |
| | Trend surface | 0.03 | 0.62 | 3.03 | 0.03 | 21.32 |
| Cropland | Cokriging | 0.01 | 0.56 | 2.95 | −0.22 | −356.18 |
| | Kriging | 0.03 | 0.56 | 2.93 | −0.20 | −355.44 |
| | Trend surface | 0.07 | 0.60 | 3.05 | −0.08 | −57.77 |

In different ways, the BIAS, RASPE, INT, and DSS are all summary statistics of the validation standardized residuals given in Equation (21), averaged over all the residuals in the validation block. Thus, these metrics assess the individual predictive distributions at each validation location, and they weight each result equally. However, because spatial dependence is present in the residuals (see Figure A2 in Appendix B), we also consider a multivariate extension of the Dawid-Sebastiani score (MDSS) to evaluate the multivariate predictive distribution at all validation locations simultaneously. The MDSS metric clearly favors the methods that leverage spatial dependence (cokriging and kriging), which provide superior multivariate predictive distributions. Compared to kriging, cokriging achieves a slightly better MDSS in both validation blocks, suggesting there is an advantage to leveraging the dependence between SIF and XCO₂ identified in Section 3.1.

The validation metrics presented in Table 1 suggest that proper characterization of uncertainty requires that the prediction method accounts for spatial dependence in the data. Cokriging and kriging both account for spatial dependence, but cokriging offers smaller prediction uncertainty when there is cross-dependence between SIF and XCO₂. The validation metrics in Table 1 are available for each of the other three months considered in this study in Tables S2–S4 in the Supplementary Material. There, we find that when there is weak dependence between SIF and XCO₂, there is little advantage in cokriging. Instead, modeling the univariate SIF covariance function and predicting using kriging is more efficient and provides comparable results to cokriging when there is weak dependence between SIF and XCO₂.

4. Conclusions, Limitations, and Future Research

Satellite observations of SIF provide a viable proxy for GPP at the global scale [3,6]. However, these observational data are noisy and spatially incomplete so scientists need denoised and gap-filled Level 3 data products to better characterize the relationship between SIF and GPP. Here, we use a geostatistical methodology called cokriging to produce coSIF, a Level 3 SIF data product at a 0.05-degree, monthly resolution with uncertainty quantification. Our highly parallelizable approach leverages OCO-2 observations of both SIF and XCO₂, which we show helps to reduce the uncertainty of the Level 3 data product in months of high GPP. It achieves this by modeling the cross-dependence between SIF and XCO₂ one month later and, moreover, it accounts for non-stationary measurement errors in the observations. Due to its fine resolution, coSIF can be aggregated to coarser resolutions, along with coherent uncertainty quantification, which enables statistical comparison with other SIF products at varying resolutions.

We limit our analysis of coSIF to a large region over North America, but by splitting global land into contiguous regions where geostatistical stationarity assumptions and the use of chordal distances are appropriate, our approach can yield a global data product, albeit with different estimates of spatial dependence parameters in different regions. Though we do not formally incorporate temporal dependence in the bivariate covariance model, this is the subject of future research. Incorporating dependence with additional environmental processes in a higher-dimensional multivariate covariance model is another promising extension. The cokriging method could also be used to produce improved predictions of XCO₂; one would simply swap the roles of $Y_1(\cdot)$ and $Y_2(\cdot)$ to produce a de-noised, gap-filled data product for XCO₂ with SIF as the secondary variable, and with uncertainty quantification of XCO₂.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs15164038/s1>, Dataset S1: Fitted parameters of the bivariate spatial models for February, April, July, and October 2021; Table S1: Average ratio of the coSIF RMSPE to the measurement error standard deviation for February, April, July, and October 2021; Figures S1–S3: The coSIF data product (predictions and corresponding RMSPEs) for February, April, and October 2021; Tables S2–S4: Validation metrics for cokriging, kriging, and trend-surface-only prediction in February, April, and October 2021.

Author Contributions: Conceptualization, N.C., J.J. and A.Z.-M.; methodology, J.J., N.C. and A.Z.-M.; software, J.J.; validation, J.J., N.C. and A.Z.-M.; formal analysis, J.J.; investigation, J.J.; resources, A.Z.-M. and N.C.; data curation, J.J.; writing—original draft preparation, J.J.; writing—review and editing, J.J., N.C. and A.Z.-M.; visualization, J.J.; supervision, N.C. and A.Z.-M.; project administration, J.J., N.C. and A.Z.-M.; funding acquisition, N.C. and A.Z.-M. All authors have read and agreed to the published version of the manuscript.

Funding: J.J. was supported by a University Postgraduate Award from the University of Wollongong, Australia. N.C. was supported by Australian Research Council Discovery Project DP190100180. A.Z.-M was supported by an Australian Research Council Discovery Early Career Research Award (DECRA) DE180100203 and by Discovery Project DP190100180. All authors were additionally supported by NASA ROSES grant 20-OCOST.

Data Availability Statement: The coSIF data product for February, April, July, and October 2021 is openly available from Zenodo at: <https://doi.org/10.5281/zenodo.8078592>. The supplementary dataset of fitted model parameters is openly available from Zenodo at: <https://doi.org/10.5281/zenodo.8078560>. The input data presented in this study are collected in a compressed file that is openly available from Zenodo at: <https://doi.org/10.5281/zenodo.8078476>. These input data are a subset of the openly available data in [35–37]. The code repository can be found at: <https://github.com/joshhjacobson/coSIF>.

Acknowledgments: We are grateful for helpful comments from two referees and the Academic Editor. We thank Yi Cao for his assistance with maintaining the high-performance computing architecture on which the results in this article were produced. We would also like to acknowledge Fabio Cramerini for developing the Scientific Colour Maps package [54], which was used for Figures 3, 6, and A2.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-----------------|---|
| BIAS | Average Prediction Error |
| CMG | Climate Modeling Grid |
| CO ₂ | Carbon Dioxide |
| DSS | Dawid-Sebastiani Score |
| GES DISC | Goddard Earth Sciences Data and Information Services Center |
| GPP | Gross Primary Production |
| IGBP | International Geosphere-Biosphere Programme |

| | |
|------------------|---|
| INT | Interval Score |
| LCC | Land-Cover Classification |
| MDSS | Multivariate Dawid-Sebastiani Score |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MSPE | Mean Squared Prediction Error |
| NASA | National Aeronautics and Space Administration (United States) |
| OCO-2 | Orbiting Carbon Observatory-2 |
| RASPE | Root-Average-Squared Prediction Error |
| RMSPE | Root-Mean-Squared Prediction Error |
| SIF | Solar-Induced Chlorophyll Fluorescence |
| TROPOMI | Tropospheric Monitoring Instrument |
| XCO ₂ | Column-Averaged Atmospheric Carbon Dioxide Concentrations |

Appendix A. Methodological Detail

Appendix A.1. Parameter Estimation: Fitting (Cross-) Semivariograms

Let $\psi_i^*(\cdot) = (Y_i(\cdot) - \hat{\mu}_i(\cdot) - \hat{\mu}_{R,i}) / \hat{S}_{R,i}$, $i = 1, 2$, be the standardized version of the de-trended random process $\psi_i(\cdot)$ defined in Equation (3). The empirical estimator of a stationary (cross-) covariance function, $\text{cov}(\psi_i^*(\mathbf{s} + \mathbf{h}), \psi_j^*(\mathbf{s}))$, for $i, j \in \{1, 2\}$, is known as the covariogram for $i = j$ and the cross-covariogram for $i \neq j$. The (cross-) covariogram is a method-of-moments estimator that is highly susceptible to bias and trend contamination ([40], Section 2.4.1). Instead, another measure of multivariate spatial dependence is advocated in [32], namely the (cross-) semivariogram function given by

$$\gamma_{ij}^*(\mathbf{h}) \equiv \frac{1}{2} \text{var}(\psi_i^*(\mathbf{s} + \mathbf{h}) - \psi_j^*(\mathbf{s})); \quad i, j \in \{1, 2\}, \quad (\text{A1})$$

for any $\mathbf{s}, \mathbf{s} + \mathbf{h} \in D$. This is sometimes called the pseudo cross-semivariogram when $i \neq j$ [31]. To study spatial variation in $\psi_1^*(\cdot)$ and $\psi_2^*(\cdot)$, the empirical (cross-) semivariogram can be found in terms of residuals calculated from the data as in Equations (5) and (6) ([40], Section 3.4.3). Under isotropy, the empirical (cross-) semivariogram is given by

$$\hat{\gamma}_{ij}(h) = \frac{1}{2|N_{ij}(h; \lambda)|} \sum_{l, k: (\mathbf{s}_{il}, \mathbf{s}_{jk}) \in N_{ij}(h; \lambda)} \left((R_{il}^* - \bar{R}_i^*) - (R_{jk}^* - \bar{R}_j^*) \right)^2, \quad (\text{A2})$$

where $h = \|\mathbf{h}\| \geq 0$, $l = 1, \dots, n_i$, $k = 1, \dots, n_j$, and $i, j \in \{1, 2\}$. Here, subtracting $\bar{R}_i^* = \frac{1}{n_i} \sum_{l=1}^{n_i} R_{il}^*$ removes any remaining constant-mean variation, and $|N_{ij}(h; \lambda)|$ is the cardinality of $N_{ij}(h; \lambda)$, for $\lambda \geq 0$. The set $N_{ij}(h; \lambda)$ represents all pairs of points $(\mathbf{s}_{il}, \mathbf{s}_{jk})$ whose chordal distance, $h = \|\mathbf{s}_{il} - \mathbf{s}_{jk}\|$, is in the spatial bin $[h - \lambda, h + \lambda]$.

We obtain a bivariate semivariogram model to fit to the empirical estimator given in Equation (A2) from a convenient relationship between semivariogram functions and covariance functions given in Equation (A3) below. Under second-order stationarity assumptions, it is shown in [32] that (cross-) semivariogram functions at spatial lag h can be obtained from (cross-) covariance functions. The relationship is modified here to account for information about the stochastic micro-scale variation and measurement error that are also contained in the residuals:

$$\gamma_{ij}(h; \theta) = \frac{1}{2} \left((C_{ii}(0; \theta^\circ) + \tau_i^2) + (C_{jj}(0; \theta^\circ) + \tau_j^2) \right) - C_{ij}(h; \theta^\circ); \quad i, j \in \{1, 2\}, \quad (\text{A3})$$

for $h \geq 0$, where $\tau_i^2 = (\sigma_{\omega,i}^2 + \sigma_{\epsilon,i}^2) / \hat{S}_{R,i}^2$ is the variability due to micro-scale variation and measurement error, with $\hat{S}_{R,i}$ given in Equation (6); $\sigma_{\omega,i}^2$ is as described for Equation (3); and $\sigma_{\epsilon,i}^2 \equiv \text{med}(\{\sigma_{\epsilon,il}^2; l = 1, \dots, n_i\})$ is the median of the measurement error variances given as described for Equation (1). Here, $\theta^\circ \equiv \{\rho_{12}, \sigma_{11}, \sigma_{22}, \nu_{11}, \nu_{12}, \nu_{22}, \ell_{11}, \ell_{12}, \ell_{22}\}$ is the set of parameters that completely characterizes $C_{ij}(\cdot)$, $i, j \in \{1, 2\}$, the full bivariate Matérn covariance model given in Equation (8), and $\theta \equiv \{\theta^\circ, \tau_1^2, \tau_2^2\}$ is the set of parameters that

completely characterizes the spatial dependence model. The result in Equation (A3) follows because the two sets of standardized measurement errors, $\{\epsilon_{il}^* : l = 1, \dots, n_i\}, i = 1, 2$, are independent of one another; and they are each independent of the standardized small-scale-variation processes $\eta_1^*(\cdot)$ and $\eta_2^*(\cdot)$, and of the standardized micro-scale-variation processes $\omega_1^*(\cdot)$ and $\omega_2^*(\cdot)$.

Weighted least squares is a popular approach for fitting univariate semivariogram models to their empirical estimates [47] since the method automatically favors lags where the (cross-) spatial dependence is strongest and downweights those lags associated with the fewest spatial pairs. An extension to fitting bivariate semivariogram models is now given. Let Θ be a set containing the valid parameter space of the full bivariate Matérn covariance model, as established in [41]. Taking the sum of the weighted squared differences between the empirical and model (cross-) semivariograms as the argument to be minimized, we minimize over $\theta \in \Theta$. For a fixed set of spatial bins centered on $\{h_1, \dots, h_r\}$ and a fixed tolerance $\lambda \geq 0$, a multivariate weighted-least-squares estimate of θ is given by

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left\{ \sum_{i=1}^2 \sum_{j=i+1}^2 \sum_{l=1}^r |N_{ij}(h_l; \lambda)| \left(\frac{\hat{\gamma}_{ij}(h_l) - \gamma_{ij}(h_l; \theta)}{\gamma_{ij}(h_l; \theta)} \right)^2 \right\}, \tag{A4}$$

where r is the number of spatial bins, $\hat{\gamma}_{ij}(\cdot)$ is given in Equation (A2), and $\gamma_{ij}(\cdot; \cdot)$ is given in Equation (A3). Note that Equation (A4) can be easily generalized from the bivariate to the multivariate setting. For our purposes, each set $N_{ij}(\cdot; \cdot)$ contains all pairs of points in one of the 30 equally spaced bins centered on $\{h_1, \dots, h_{30}\}$, which span between 0 km and a maximum distance of 1000 km. The tolerance λ is determined such that the sets $\{N_{ij}(h_l; \lambda) : l = 1, \dots, 30\}, i, j \in 1, 2$, are contiguous and non-overlapping. This approach yields numbers of pairs that are well beyond the recommended minimum number of 30 pairs in each set $N_{ij}(\cdot; \cdot)$; the importance of this is emphasized in [40], Section 2.4. To preserve total variation in prediction using Equations (11) and (12), each micro-scale-variation component, $\sigma_{\omega,i}^2$ is estimated as

$$\hat{\sigma}_{\omega,i}^2 = \left(\hat{\tau}_i^2 \hat{S}_{R,i}^2 - \hat{\sigma}_{\epsilon,i}^2 \right) \mathbf{1} \left\{ \hat{\tau}_i^2 \hat{S}_{R,i}^2 > \hat{\sigma}_{\epsilon,i}^2 \right\}; \quad i = 1, 2, \tag{A5}$$

where recall that $\hat{\tau}_i^2$ is an element of $\hat{\theta}$, $\hat{S}_{R,i}$ is given in Equation (6), and $\mathbf{1}\{\cdot\}$ is the indicator function, which ensures that the estimated variance is non-negative.

Recall that the values of $\{v_{ij}\}_{ij}$, $\{\ell_{ij}\}_{ij}$, and $\{\rho_{ij}\}_{ij}$ that result in a positive-semidefinite bivariate covariance function belong to a constrained parameter space [41]. Accordingly, we check a necessary condition for validity of the spatial model given by Equation (8). At each prediction location, \mathbf{s}_0 , in the spatial domain, D , we check that the joint covariance matrix

$$\text{cov} \begin{pmatrix} \psi_1^*(\mathbf{s}_0) \\ \mathbf{R}_0^* \end{pmatrix} = \begin{pmatrix} C_{11}(0) & \mathbf{c}_0^T \\ \mathbf{c}_0 & \Sigma_0 \end{pmatrix} \tag{A6}$$

is positive-semidefinite. Here, $C_{11}(\cdot)$ is given by Equation (8); Σ_0 is a 300×300 covariance matrix given by Equation (9); and \mathbf{c}_0 is a 300-dimensional vector with elements given as described by Equation (10).

Appendix A.2. Kriging and Trend-Surface-Only Prediction for Validation Comparisons

The cokriging method described in Sections 2.3 and 2.4 is able to leverage dependence between the latent SIF and XCO₂ processes, respectively, $Y_1(\cdot)$ and $Y_2(\cdot)$, for prediction and uncertainty quantification using the noisy data \mathbf{Z}_1 and \mathbf{Z}_2 . Now, through two successive simplifications of cokriging, we describe two special cases for producing predictions and corresponding prediction standard errors. In the first simplification, we assume there is no cross-dependence between $Y_1(\cdot)$ and $Y_2(\cdot)$, and hence we seek to model $Y_1(\cdot)$ using \mathbf{Z}_1

alone; we refer to this model as a univariate spatial dependence model. Recall that at each data location $\mathbf{s}_l \in D$, we model the elements of \mathbf{Z}_1 as

$$Z_{1l} = Y_1(\mathbf{s}_l) + \epsilon_{1l} = \mu_1(\mathbf{s}_l) + \psi_1(\mathbf{s}_l) + \epsilon_{1l}; \quad l = 1, \dots, n_1, \tag{A7}$$

where $\{\epsilon_{1l}\}_l$ is a set of (non-stationary) independent measurement errors, assumed to have mean zero and variances $\{\sigma_{\epsilon_{1l}}^2\}_l$ given as described for Equation (1). Estimates for the large-scale trend-surface, $\mu_1(\cdot)$, are obtained identically to Equation (4) when $i = 1$. The process $\psi_1^*(\cdot) = \eta_1^*(\cdot) + \omega_1^*(\cdot)$ is a combination of smooth small-scale variation and micro-scale variation *on the standardized scale* and is the focus of the univariate spatial dependence model, from which the optimal linear predictor known as the kriging predictor is derived.

As in the cokriging method, the micro-scale-variation term, $\omega_1^*(\cdot)$, is a process uncorrelated at the grid scale with mean zero and variance $\sigma_{\omega,1}^2 / \hat{S}_{R,1}^2$, where $\hat{S}_{R,1}$ is given in Equation (6) when $i = 1$. To model small-scale spatial variation in $\eta_1^*(\cdot)$ only, the spatial dependence captured in Equation (8) can be simplified to

$$C_1(h) = \sigma_1^2 M(h; \nu_1, \ell_1), \tag{A8}$$

for chordal distance $h \geq 0$, scale parameter $\sigma_1 > 0$, and $M(\cdot; \nu_1, \ell_1)$ the Matérn correlation function [43,44] given in Equation (8) with smoothness parameter $\nu_1 > 0$ and correlation length $\ell_1 > 0$. For inference, we obtain an empirical semivariogram using the standardized spatial residuals, $\mathbf{R}_1^* = (R_{11}^*, \dots, R_{1n_1}^*)^T$, given by Equation (6) when $i = 1$. Using the (univariate) relationship between semivariograms and covariance functions given in Equation (A3) when $i = j = 1$, we obtain an estimate, $\hat{\theta}$, for $\theta = \{\sigma_1, \nu_1, \ell_1, \tau_1^2\}$ through fitting the empirical semivariogram by weighted least squares [47].

As a simplification of cokriging, the kriging prediction at validation location $\mathbf{s}_a \in A$, \hat{Z}_{1a}^{kr} , and corresponding root-mean-squared prediction error, $\text{RMSPE}(\hat{Z}_{1a}^{\text{kr}})$, use only the noisy SIF data \mathbf{Z}_1 . Let $\mathbf{R}_{a1}^{*(-A)}$ be the 150-dimensional subset of \mathbf{R}_1^* whose values correspond to the 150 data locations found in D but not in A that are closest in chordal distance to \mathbf{s}_a . We model the spatial variation across the locations associated with $\mathbf{R}_{a1}^{*(-A)}$ by the covariance matrix

$$\Sigma_{a1}^{\text{kr}} = \mathbf{C}_{a1} + \frac{\sigma_{\omega,1}^2}{\hat{S}_{R,1}^2} \mathbf{I} + \text{diag}(\sigma_{\epsilon,a1}^*), \tag{A9}$$

where \mathbf{C}_{a1} is a 150×150 covariance matrix having elements given by $C_1(\cdot)$ from Equation (A8) with estimated parameters $\hat{\theta}$ substituted in; \mathbf{I} is the 150×150 identity matrix; $\sigma_{\epsilon,a1}^*$ is a 150-dimensional vector with elements $\{\sigma_{\epsilon,a1l}^2 / \hat{S}_{R,1}^2 : l = 1, \dots, 150\}$ corresponding to locations associated with $\mathbf{R}_{a1}^{*(-A)}$; and $\text{diag}(\sigma_{\epsilon,a1}^*)$ is a diagonal matrix with the elements of $\sigma_{\epsilon,a1}^*$ along the diagonal. Similarly, we define the vector $\mathbf{c}_{a1}^{\text{kr}} \equiv \left(\text{cov}(\psi_1^*(\mathbf{s}_a), R_{a1l}^{*(-A)}) : l = 1, \dots, 150 \right)^T$, where

$$\text{cov}(\psi_1^*(\mathbf{s}_a), R_{a1l}^{*(-A)}) = C_1(h_{al}) + \frac{\sigma_{\omega,1}^2}{\hat{S}_{R,1}^2} \mathbf{1}\{h_{al} = 0\}, \tag{A10}$$

for $h_{al} = \|\mathbf{s}_a - \mathbf{s}_l\|$, \mathbf{s}_l the spatial location associated with $R_{a1l}^{*(-A)}$; $\|\cdot\|$ denotes chordal distance; and $\mathbf{1}\{\cdot\}$ is the indicator function. After updating through (see also Equations (11) and (12))

$$\hat{\psi}_1^{*(kr)}(\mathbf{s}_a) = \left(\mathbf{c}_{a1}^{kr}\right)^T \left(\boldsymbol{\Sigma}_{a1}^{kr}\right)^{-1} \mathbf{R}_{a1}^{*(-A)}, \tag{A11}$$

$$MSPE\left(\hat{\psi}_1^{*(kr)}(\mathbf{s}_a)\right) = \frac{\sigma_{\omega,1}^2}{\hat{\Sigma}_{R,1}^2} + C_1(0) - \left(\mathbf{c}_{a1}^{kr}\right)^T \left(\boldsymbol{\Sigma}_{a1}^{kr}\right)^{-1} \mathbf{c}_{a1}^{kr}, \tag{A12}$$

we obtain the kriging validation prediction, \hat{Z}_{1a}^{kr} , and its root-mean-squared prediction error, $RMSPE\left(\hat{Z}_{1a}^{kr}\right)$, using Equations (13)–(15). It should be noted that the kriging predictor, \hat{Z}_{1a}^{kr} , can be considered a special case of cokriging where the XCO₂ residuals, \mathbf{R}_{a2}^* , are given zero weight; hence, kriging is a *simplification* of cokriging.

As a further simplification, the trend-surface-only prediction at validation location $\mathbf{s}_a \in A$, \hat{Z}_{1a}^{tr} , and the corresponding root-mean-squared prediction error, $RMSPE\left(\hat{Z}_{1a}^{tr}\right)$, use only the SIF data \mathbf{Z}_1 but assume no spatial dependence in $\psi_1^*(\cdot)$. In particular, the trend-surface-only prediction is simply given by $\hat{Z}_{1a}^{tr} = \hat{\mu}_1(\mathbf{s}_a)$, for $a = 1, \dots, n_A$, where $\hat{\mu}_1(\cdot)$ is defined in Equation (4) when $i = 1$. Since the trend-surface-only predictions do not account for spatial dependence, their associated prediction errors depend only on the micro-scale variation $\omega_1(\cdot)$ and the measurement errors $\{\epsilon_{1a}\}_a$, and consequently the covariance as a function of distance from prediction location to data locations is identically zero. Recall that $\omega_1(\cdot)$ is a process uncorrelated at the grid scale with mean zero and variance $\sigma_{\omega,1}^2$, and that $\{\epsilon_{1a}\}_a$ is a set of (non-stationary) independent measurement errors with mean zero and variances $\{\sigma_{\epsilon,1a}^2\}_a$; see Equation (1). To obtain an estimate of $\sigma_{\omega,1}^2$, we take an average of the squared residuals given in Equation (6). Define

$$\hat{\tau}_1^2 \equiv \frac{1}{n_1} \sum_{l=1}^{n_1} R_{1l}^{*2}, \tag{A13}$$

which is an estimate of the total variation at data locations. We then use the median of the measurement error variances, $\tilde{\sigma}_{\epsilon,1}^2 = \text{med}(\{\sigma_{\epsilon,1l}^2; l = 1, \dots, n_1\})$, to obtain $\hat{\sigma}_{\omega,1}^2$ from $\hat{\tau}_1^2$, according to Equation (A5). Hence, Equation (15) simplifies to

$$RMSPE\left(\hat{Z}_{1a}^{tr}\right) = \sqrt{\hat{\sigma}_{\omega,1}^2 + \sigma_{\epsilon,1a}^2}; \quad a = 1, \dots, n_A. \tag{A14}$$

Appendix A.3. Multivariate Predictive Covariance Matrix Approximation

Recall that for spatial locations $\{\mathbf{s}_1, \dots, \mathbf{s}_{n_A}\}$ in a validation block A , F_{1A} is the multivariate Gaussian predictive distribution with conditional mean vector $E(\mathbf{Z}_{1A} | \mathbf{Z}^{-A})$ and conditional variance-covariance matrix $\text{var}(\mathbf{Z}_{1A} | \mathbf{Z}^{-A}) \equiv \mathbf{V}_{1A}$ (Section 2.5). In what follows, we approximate $E(\mathbf{Z}_{1A} | \mathbf{Z}^{-A})$ by the localized predictor $\hat{\mathbf{Z}}_{1A} \equiv \left(\hat{Z}_{11}, \dots, \hat{Z}_{1n_A}\right)^T$, with elements given by the right-hand-side of Equation (13), and \mathbf{V}_{1A} by the matrix $\tilde{\mathbf{V}}_{1A}$ whose elements we now derive. The diagonal elements of $\tilde{\mathbf{V}}_{1A}$ can be obtained via the square of Equation (15); here, we derive the off-diagonal elements of $\tilde{\mathbf{V}}_{1A}$. Each element is given by

$$\tilde{V}_{1A,ab} \equiv \text{cov}\left(Z_{1a}, Z_{1b} | \mathbf{Z}_{ab}^{-A}\right); \quad a, b \in \{1, \dots, n_A\}, \tag{A15}$$

where the vector \mathbf{Z}_{ab}^{-A} is composed of sub-vectors $\mathbf{Z}_{ab,1}^{-A}$ and $\mathbf{Z}_{ab,2}$. Here, the sub-vector $\mathbf{Z}_{ab,1}^{-A}$ is a subset of the SIF values in \mathbf{Z}_1 corresponding to the 150 locations not in A that are closest to \mathbf{s}_a , and the 150 locations not in A that are closest to \mathbf{s}_b . Note that there may be overlap in each set of 150 locations, resulting in a vector of dimension smaller than 300. The sub-vector $\mathbf{Z}_{ab,2}$ is defined similarly, but it does *not* exclude the locations in the validation region A

because its entries are XCO₂ data, not SIF. Using the definition of the noisy data value, Z_{1a} , from Equation (A7), Equation (A15) becomes

$$\tilde{V}_{1A,ab} = \hat{S}_{R,1}^2 \text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right) + \sigma_{\epsilon,1a}^2 \mathbf{1}\{a = b\}; \quad a, b \in \{1, \dots, n_A\}, \quad (\text{A16})$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and the variances $\{\sigma_{\epsilon,1a}^2\}_a$ are given as described for Equation (1). The vector $\mathbf{R}_{ab}^{*(-A)}$ has sub-vectors $\mathbf{R}_{ab,1}^{*(-A)}$ and $\mathbf{R}_{ab,2}^{*(-A)}$, with elements obtained, respectively, from $\mathbf{Z}_{ab,1}^{-A}$ and $\mathbf{Z}_{ab,2}$, by the standardization process involving $\hat{\mu}_{R,1}$, $\hat{\mu}_{R,2}$, $\hat{S}_{R,1}$, and $\hat{S}_{R,2}$ as described in Equation (6).

Because the joint distribution of \mathbf{R} is Gaussian, $\text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right) = \text{E}\left(\text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right)\right)$. Then, by the law of total expectation,

$$\begin{aligned} &\text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right) \\ &= \text{E}\left(\text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right)\right) \\ &= \text{E}\left(\text{E}\left\{\left(\psi_1^*(\mathbf{s}_a) - \text{E}\left\{\psi_1^*(\mathbf{s}_a) \mid \mathbf{R}_{ab}^{*(-A)}\right\}\right)\left(\psi_1^*(\mathbf{s}_b) - \text{E}\left\{\psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right\}\right) \mid \mathbf{R}_{ab}^{*(-A)}\right\}\right) \\ &= \text{cov}\left(\psi_1^*(\mathbf{s}_a) - \text{E}\left\{\psi_1^*(\mathbf{s}_a) \mid \mathbf{R}_{ab}^{*(-A)}\right\}, \psi_1^*(\mathbf{s}_b) - \text{E}\left\{\psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right\}\right), \end{aligned} \quad (\text{A17})$$

where the final equality uses that $\text{E}\left(\psi_1^*(\mathbf{s}_a) - \text{E}\left\{\psi_1^*(\mathbf{s}_a) \mid \mathbf{R}_{ab}^{*(-A)}\right\}\right) = 0$, for $a = 1, \dots, n_A$. By the properties of the multivariate Gaussian distribution (e.g., [55], Chapter 4),

$$\text{E}\left(\psi_1^*(\mathbf{s}_a) \mid \mathbf{R}_{ab}^{*(-A)}\right) \approx \mathbf{c}_a^T \Sigma_a^{-1} \mathbf{R}_a^{*(-A)} = \hat{\psi}_1^*(\mathbf{s}_a), \quad (\text{A18})$$

with components as defined for Equation (11). The vector $\mathbf{R}_a^{*(-A)}$ has 150-dimensional sub-vectors $\mathbf{R}_{a1}^{*(-A)}$ (for SIF residuals) and $\mathbf{R}_{a2}^{*(-A)}$ (for XCO₂ residuals), and since each sub-vector is determined with respect to \mathbf{s}_a only, $\mathbf{R}_a^{*(-A)}$ is potentially a subset of $\mathbf{R}_{ab}^{*(-A)}$, in which case Equation (A18) is an approximation. Denote the cokriging weights in Equation (A18) by $\alpha_a^T \equiv \mathbf{c}_a^T \Sigma_a^{-1}$. Then, from Equation (A17) and the approximation in Equation (A18), we obtain

$$\begin{aligned} &\text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) \mid \mathbf{R}_{ab}^{*(-A)}\right) \\ &\approx \text{cov}\left(\psi_1^*(\mathbf{s}_a) - \hat{\psi}_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b) - \hat{\psi}_1^*(\mathbf{s}_b)\right) \\ &= \text{cov}\left(\psi_1^*(\mathbf{s}_a) - \alpha_a^T \mathbf{R}_a^{*(-A)}, \psi_1^*(\mathbf{s}_b) - \alpha_b^T \mathbf{R}_b^{*(-A)}\right) \\ &= \text{cov}\left(\psi_1^*(\mathbf{s}_a), \psi_1^*(\mathbf{s}_b)\right) - \text{cov}\left(\psi_1^*(\mathbf{s}_a), \mathbf{R}_b^{*(-A)}\right) \alpha_b \\ &\quad - \alpha_a^T \text{cov}\left(\mathbf{R}_a^{*(-A)}, \psi_1^*(\mathbf{s}_b)\right) + \alpha_a^T \text{cov}\left(\mathbf{R}_a^{*(-A)}, \mathbf{R}_b^{*(-A)}\right) \alpha_b \\ &= C_{11}(h_{ab}) + \frac{\sigma_{\omega,1}^2}{\hat{S}_{R,1}^2} \mathbf{1}\{a = b\} - \mathbf{c}_{ab}^T \alpha_b - \alpha_a^T \mathbf{c}_{ba} + \alpha_a^T \Sigma_{ab} \alpha_b, \end{aligned} \quad (\text{A19})$$

where $C_{11}(\cdot)$ is the covariance function given in Equation (8); $h_{ab} = \|\mathbf{s}_a - \mathbf{s}_b\|$ is the chordal distance between validation locations \mathbf{s}_a and \mathbf{s}_b ; $\sigma_{\omega,1}^2$ is the variance of the micro-scale process $\omega_1(\cdot)$; $\hat{S}_{R,1}$ is given in Equation (6) when $i = 1$; and $\mathbf{1}\{\cdot\}$ is the indicator function. The vector \mathbf{c}_{ab} is made up of sub-vectors $\mathbf{c}_{ab,l} \equiv \left(\text{cov}\left(\psi_1^*(\mathbf{s}_a), R_{bl}^{*(-A)}\right)\right) : l = 1, \dots, 150)^T$ and $\mathbf{c}_{ab,2} \equiv \left(\text{cov}\left(\psi_1^*(\mathbf{s}_a), R_{b2l}^*\right) : l = 1, \dots, 150\right)^T$, where

$$\begin{aligned} \text{cov}(\psi_1^*(\mathbf{s}_a), R_{b1l}^{*(-A)}) &= C_{11}(h_{ab}) + \frac{\sigma_{\omega,1}^2}{\hat{S}_{R,1}^2} \mathbf{1}\{h_{ab} = 0\}, \\ \text{cov}(\psi_1^*(\mathbf{s}_a), R_{b2l}^*) &= C_{12}(h_{ab}). \end{aligned} \tag{A20}$$

We obtain the vector $\mathbf{c}_{ba} = (\mathbf{c}_{ba,1}^T, \mathbf{c}_{ba,2}^T)^T$ similarly. The covariance matrix Σ_{ab} in Equation (A19) is given by

$$\Sigma_{ab} \equiv \text{cov}(\mathbf{R}_a^{*(-A)}, \mathbf{R}_b^{*(-A)}) = \begin{pmatrix} \mathbf{C}_{ab,11} & \mathbf{C}_{ab,12} \\ \mathbf{C}_{ab,21} & \mathbf{C}_{ab,22} \end{pmatrix} + \begin{pmatrix} \mathbf{Q}_{ab,11} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ab,22} \end{pmatrix}, \tag{A21}$$

where each block $\mathbf{C}_{ab,ij}$, $i, j \in \{1, 2\}$, is a 150×150 (cross-) covariance matrix having elements $\{C_{ij}(h_{lk}) : l = 1, \dots, 150; k = 1, \dots, 150\}$ for $C_{ij}(\cdot)$ given in Equation (8) with estimated parameters substituted in, and $h_{lk} = \|\mathbf{s}_l - \mathbf{s}_k\|$ the chordal distance between data locations \mathbf{s}_l and \mathbf{s}_k . The locations $\{\mathbf{s}_l : l = 1, \dots, 150\}$ used in the predictor $\hat{\psi}_1^*(\mathbf{s}_a)$ are associated with the values of sub-vector $\mathbf{R}_{a1}^{*(-A)}$ (when $i = 1$) or \mathbf{R}_{a2}^* (when $i = 2$); and the locations $\{\mathbf{s}_k : k = 1, \dots, 150\}$ used in the predictor $\hat{\psi}_1^*(\mathbf{s}_b)$ are associated with the values of sub-vector $\mathbf{R}_{b1}^{*(-A)}$ (when $j = 1$) or \mathbf{R}_{b2}^* (when $j = 2$). Additionally, each block $\mathbf{Q}_{ab,ii}$, $i = 1, 2$, is a 150×150 matrix with elements

$$\left\{ \left(\frac{\sigma_{\omega,i}^2 + \sigma_{\epsilon,a il}^2}{\hat{S}_{R,i}^2} \right) \mathbf{1}\{h_{lk} = 0\} : l = 1, \dots, 150; k = 1, \dots, 150 \right\}, \tag{A22}$$

where the measurement error variances $\{\sigma_{\epsilon,a il}^2\}_l$ correspond to locations associated with $\mathbf{R}_{a1}^{*(-A)}$ (when $i = 1$) or \mathbf{R}_{a2}^* (when $i = 2$).

The matrix $\tilde{\mathbf{V}}_{1A}$ has elements given by Equation (A16), which we approximate using the result in Equation (A19). In fact, we need only compute the upper triangular elements since $\tilde{\mathbf{V}}_{1A}$ is symmetric. Note that together, $\tilde{\mathbf{Z}}_{1A}$ and $\tilde{\mathbf{V}}_{1A}$ fully characterize \tilde{F}_{1A} , the multivariate Gaussian predictive distribution we use for the cokriging prediction at each prediction location $\mathbf{s}_a \in A$ based on the nearest 150 values of SIF (not in A) and the nearest 150 values of XCO₂.

In the simpler case of kriging prediction, Equation (A18) (and hence the weight vector $\alpha_a^T \equiv \mathbf{c}_a^T \Sigma_a^{-1}$) is updated using Equation (A11). Then, to obtain the variance-covariance matrix for kriging prediction, $\tilde{\mathbf{V}}_{1A}^{\text{kr}}$, the result in Equation (A19) is updated by replacing \mathbf{c}_{ab} with $\mathbf{c}_{ab,1}^{\text{kr}}$, \mathbf{c}_{ba} with $\mathbf{c}_{ba,1}^{\text{kr}}$, and Σ_{ab} with $\Sigma_{ab,1}^{\text{kr}} \equiv \mathbf{C}_{ab,1}^{\text{kr}} + \mathbf{Q}_{ab,1}^{\text{kr}}$. The superscript (kr) is used to highlight that the elements of these components are obtained using the univariate covariance model given by Equation (A8), rather than the bivariate covariance model given by Equation (8).

In the simplest case, trend-surface-only prediction assumes no spatial dependence in $\psi_1^*(\cdot)$, so the off-diagonal elements of Equation (A15) are identically zero. Hence, the variance-covariance matrix for trend-surface-only prediction, $\tilde{\mathbf{V}}_{1A}^{\text{tr}}$, is a diagonal matrix with diagonal elements given by the square of Equation (A14).

Appendix B. Model Diagnostics

Graphical diagnostics are an important consideration in any modeling framework; they offer an initial evaluation of whether modeling assumptions have been upheld and can provide a general assessment of the quality of the fitted model. Here, we consider diagnostics for the validation standardized residuals given by Equation (21). Quantile-quantile (Q-Q) plots for the validation standardized residuals in July 2021 are shown in Figure A1 for the Corn Belt and Cropland validation blocks. For the Cropland validation block, the Q-Q plots that result from using the cokriging, kriging, and trend-surface-only methods are all slightly heavy-tailed relative to the standard Gaussian distribution. Still, for both validation blocks, the validation standardized residuals are reasonably Gaussian distributed.

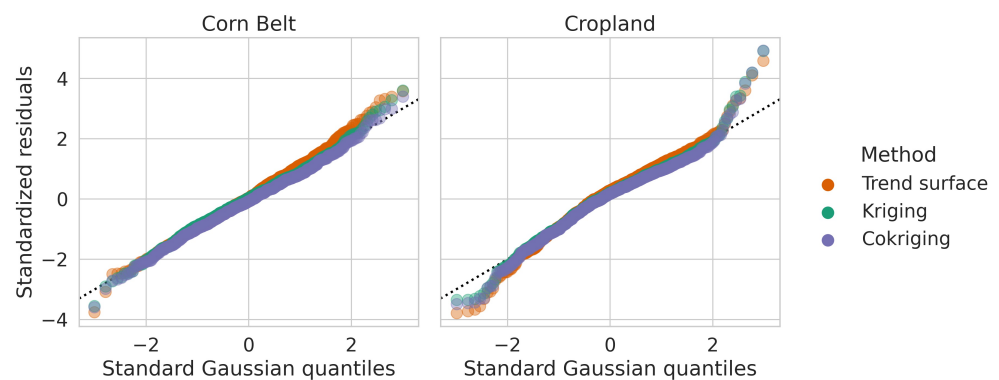


Figure A1. Quantile-quantile plots for validation standardized residuals in the Corn Belt and Cropland $5^\circ \times 5^\circ$ validation blocks for each prediction method in July 2021. The dotted reference line has unit slope and is constrained to pass through the origin. Locations of the validation blocks are shown in Figure 3.

Figure A2 shows the spatial distribution of the validation standardized residuals given in Equation (21) produced from the cokriging, kriging, and trend-surface-only methods for both validation blocks in July 2021. Values in the range of $[-1, 1]$ are withheld to reduce clutter and illustrate spatial patterns; hence, the depicted residuals represent the upper and lower tails of each distribution. The percentages of retained residuals (those depicted) are given in the panel titles. In both validation blocks, there are fewer residuals (on the order of five percentage points fewer) plotted for the cokriging and kriging methods than for the trend-surface-only method, suggesting that by accounting for spatial dependence, cokriging and kriging lower the dispersion of the residuals. Nonetheless, spatial dependence is present in each set of residuals, as seen through the clustering of positive or negative values. This is expected given that the SIF validation data, Z_{1A} , are spatially correlated, and highlights the need for the MDSS validation metric in Table 1.

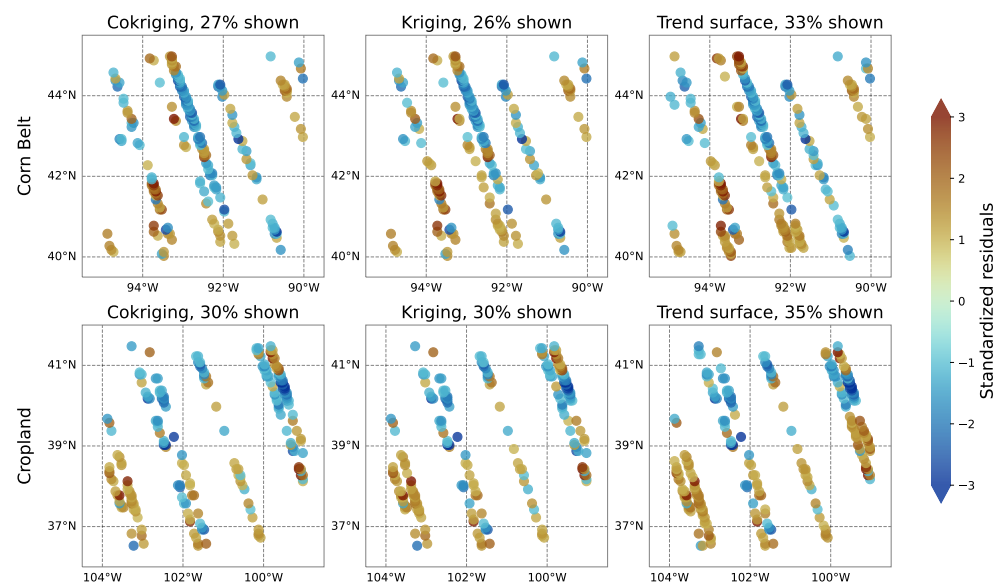


Figure A2. Spatial patterns of standardized residuals given by Equation (21) for each prediction method. **Top panels:** Standardized residuals for the $5^\circ \times 5^\circ$ Corn Belt validation block in July 2021. **Bottom panels:** Standardized residuals for the $5^\circ \times 5^\circ$ Cropland validation block in July 2021. The 0.05 -degree cells containing the residuals are enlarged for visibility. To reduce clutter and illustrate spatial patterns, values in the range of $[-1, 1]$ are not shown. Panel titles indicate the spatial prediction method used and the percentage of total values displayed. Locations of the validation blocks are shown in Figure 3.

References

1. Frankenberg, C.; O'Dell, C.; Berry, J.; Guanter, L.; Joiner, J.; Köhler, P.; Pollock, R.; Taylor, T.E. Prospects for Chlorophyll Fluorescence Remote Sensing from the Orbiting Carbon Observatory-2. *Remote Sens. Environ.* **2014**, *147*, 1–12. [[CrossRef](#)]
2. Sun, Y.; Frankenberg, C.; Wood, J.D.; Schimel, D.S.; Jung, M.; Guanter, L.; Drewry, D.T.; Verma, M.; Porcar-Castell, A.; Griffis, T.J.; et al. OCO-2 Advances Photosynthesis Observation from Space via Solar-Induced Chlorophyll Fluorescence. *Science* **2017**, *358*, eaam5747. [[CrossRef](#)]
3. Sun, Y.; Frankenberg, C.; Jung, M.; Joiner, J.; Guanter, L.; Köhler, P.; Magney, T. Overview of Solar-Induced Chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, Cross-Mission Comparison, and Global Monitoring for GPP. *Remote Sens. Environ.* **2018**, *209*, 808–823. [[CrossRef](#)]
4. Li, X.; Xiao, J.; He, B.; Altaf Arain, M.; Beringer, J.; Desai, A.R.; Emmel, C.; Hollinger, D.Y.; Krasnova, A.; Mammarella, I.; et al. Solar-Induced Chlorophyll Fluorescence Is Strongly Correlated with Terrestrial Photosynthesis for a Wide Variety of Biomes: First Global Analysis Based on OCO-2 and Flux Tower Observations. *Glob. Chang. Biol.* **2018**, *24*, 3990–4008. [[CrossRef](#)] [[PubMed](#)]
5. Frankenberg, C.; Berry, J. Solar Induced Chlorophyll Fluorescence: Origins, Relation to Photosynthesis and Retrieval. In *Comprehensive Remote Sensing*; Liang, S., Ed.; Elsevier: Oxford, UK, 2018; pp. 143–162. [[CrossRef](#)]
6. Porcar-Castell, A.; Tyystjärvi, E.; Atherton, J.; van der Tol, C.; Flexas, J.; Pfündel, E.E.; Moreno, J.; Frankenberg, C.; Berry, J.A. Linking Chlorophyll *a* Fluorescence to Photosynthesis for Remote Sensing Applications: Mechanisms and Challenges. *J. Exp. Bot.* **2014**, *65*, 4065–4095. [[CrossRef](#)]
7. Sun, Y.; Wen, J.; Gu, L.; Joiner, J.; Chang, C.Y.; van der Tol, C.; Porcar-Castell, A.; Magney, T.; Wang, L.; Hu, L.; et al. From Remotely-Sensed Solar-Induced Chlorophyll Fluorescence to Ecosystem Structure, Function, and Service: Part II—Harnessing Data. *Glob. Chang. Biol.* **2023**, *29*, 2893–2925. [[CrossRef](#)] [[PubMed](#)]
8. Baker, N.R. Chlorophyll Fluorescence: A Probe of Photosynthesis in Vivo. *Annu. Rev. Plant Biol.* **2008**, *59*, 89–113. [[CrossRef](#)] [[PubMed](#)]
9. Guanter, L.; Alonso, L.; Gómez-Chova, L.; Amorós-López, J.; Vila, J.; Moreno, J. Estimation of Solar-Induced Vegetation Fluorescence from Space Measurements. *Geophys. Res. Lett.* **2007**, *34*, L08401. [[CrossRef](#)]
10. Frankenberg, C.; Fisher, J.B.; Worden, J.; Badgley, G.; Saatchi, S.S.; Lee, J.E.; Toon, G.C.; Butz, A.; Jung, M.; Kuze, A.; et al. New Global Observations of the Terrestrial Carbon Cycle from GOSAT: Patterns of Plant Fluorescence with Gross Primary Productivity. *Geophys. Res. Lett.* **2011**, *38*, L17706. [[CrossRef](#)]
11. Joiner, J.; Yoshida, Y.; Vasilkov, A.P.; Yoshida, Y.; Corp, L.A.; Middleton, E.M. First Observations of Global and Seasonal Terrestrial Chlorophyll Fluorescence from Space. *Biogeosciences* **2011**, *8*, 637–651. [[CrossRef](#)]
12. Mohammed, G.H.; Colombo, R.; Middleton, E.M.; Rascher, U.; van der Tol, C.; Nedbal, L.; Goulas, Y.; Pérez-Priego, O.; Damm, A.; Meroni, M.; et al. Remote Sensing of Solar-Induced Chlorophyll Fluorescence (SIF) in Vegetation: 50 Years of Progress. *Remote Sens. Environ.* **2019**, *231*, 111177. [[CrossRef](#)]
13. Doughty, R.; Kurosu, T.P.; Parazoo, N.; Köhler, P.; Wang, Y.; Sun, Y.; Frankenberg, C. Global GOSAT, OCO-2, and OCO-3 Solar-Induced Chlorophyll Fluorescence Datasets. *Earth Syst. Sci. Data* **2022**, *14*, 1513–1529. [[CrossRef](#)]
14. Parazoo, N.C.; Frankenberg, C.; Köhler, P.; Joiner, J.; Yoshida, Y.; Magney, T.; Sun, Y.; Yadav, V. Towards a Harmonized Long-Term Spaceborne Record of Far-Red Solar-Induced Fluorescence. *J. Geophys. Res. Biogeosci.* **2019**, *124*, 2518–2539. [[CrossRef](#)]
15. Köhler, P.; Frankenberg, C.; Magney, T.S.; Guanter, L.; Joiner, J.; Landgraf, J. Global Retrievals of Solar-Induced Chlorophyll Fluorescence with TROPOMI: First Results and Intersensor Comparison to OCO-2. *Geophys. Res. Lett.* **2018**, *45*, 10456–10463. [[CrossRef](#)]
16. Guanter, L.; Bacour, C.; Schneider, A.; Aben, I.; van Kempen, T.A.; Maignan, F.; Retscher, C.; Köhler, P.; Frankenberg, C.; Joiner, J.; et al. The TROPISIF Global Sun-Induced Fluorescence Dataset from the Sentinel-5P TROPOMI Mission. *Earth Syst. Sci. Data* **2021**, *13*, 5423–5440. [[CrossRef](#)]
17. Duveiller, G.; Cescatti, A. Spatially Downscaling Sun-Induced Chlorophyll Fluorescence Leads to an Improved Temporal Correlation with Gross Primary Productivity. *Remote Sens. Environ.* **2016**, *182*, 72–89. [[CrossRef](#)]
18. Duveiller, G.; Filipponi, F.; Walther, S.; Köhler, P.; Frankenberg, C.; Guanter, L.; Cescatti, A. A Spatially Downscaled Sun-Induced Fluorescence Global Product for Enhanced Monitoring of Vegetation Productivity. *Earth Syst. Sci. Data* **2020**, *12*, 1101–1116. [[CrossRef](#)]
19. Turner, A.J.; Köhler, P.; Magney, T.S.; Frankenberg, C.; Fung, I.; Cohen, R.C. A Double Peak in the Seasonality of California's Photosynthesis as Observed from Space. *Biogeosciences* **2020**, *17*, 405–422. [[CrossRef](#)]
20. Chen, X.; Huang, Y.; Nie, C.; Zhang, S.; Wang, G.; Chen, S.; Chen, Z. A Long-Term Reconstructed TROPOMI Solar-Induced Fluorescence Dataset Using Machine Learning Algorithms. *Sci. Data* **2022**, *9*, 427. [[CrossRef](#)] [[PubMed](#)]
21. Wang, X.; Biederman, J.A.; Knowles, J.F.; Scott, R.L.; Turner, A.J.; Dannenberg, M.P.; Köhler, P.; Frankenberg, C.; Litvak, M.E.; Flerchinger, G.N.; et al. Satellite Solar-Induced Chlorophyll Fluorescence and near-Infrared Reflectance Capture Complementary Aspects of Dryland Vegetation Productivity Dynamics. *Remote Sens. Environ.* **2022**, *270*, 112858. [[CrossRef](#)]
22. Gentine, P.; Alemohammad, S.H. Reconstructed Solar-Induced Fluorescence: A Machine Learning Vegetation Product Based on MODIS Surface Reflectance to Reproduce GOME-2 Solar-Induced Fluorescence. *Geophys. Res. Lett.* **2018**, *45*, 3136–3146. [[CrossRef](#)] [[PubMed](#)]

23. Wen, J.; Köhler, P.; Duveiller, G.; Parazoo, N.; Magney, T.; Hooker, G.; Yu, L.; Chang, C.; Sun, Y. A Framework for Harmonizing Multiple Satellite Instruments to Generate a Long-Term Global High Spatial-Resolution Solar-Induced Chlorophyll Fluorescence (SIF). *Remote Sens. Environ.* **2020**, *239*, 111644. [[CrossRef](#)]
24. Gensheimer, J.; Turner, A.J.; Köhler, P.; Frankenberg, C.; Chen, J. A Convolutional Neural Network for Spatial Downscaling of Satellite-Based Solar-Induced Chlorophyll Fluorescence (SIFnet). *Biogeosciences* **2022**, *19*, 1777–1793. [[CrossRef](#)]
25. Zhang, Y.; Joiner, J.; Alemohammad, S.H.; Zhou, S.; Gentine, P. A Global Spatially Contiguous Solar-Induced Fluorescence (CSIF) Dataset Using Neural Networks. *Biogeosciences* **2018**, *15*, 5779–5800. [[CrossRef](#)]
26. Li, X.; Xiao, J. A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data. *Remote Sens.* **2019**, *11*, 517. [[CrossRef](#)]
27. Yu, L.; Wen, J.; Chang, C.Y.; Frankenberg, C.; Sun, Y. High-Resolution Global Contiguous SIF of OCO-2. *Geophys. Res. Lett.* **2019**, *46*, 1449–1458. [[CrossRef](#)]
28. Zammit-Mangion, A.; Cressie, N.; Shumack, C. On Statistical Approaches to Generate Level 3 Products from Satellite Remote Sensing Retrievals. *Remote Sens.* **2018**, *10*, 155. [[CrossRef](#)]
29. Tadić, J.M.; Qiu, X.; Yadav, V.; Michalak, A.M. Mapping of Satellite Earth Observations Using Moving Window Block Kriging. *Geosci. Model Dev.* **2015**, *8*, 3311–3319. [[CrossRef](#)]
30. Tadić, J.M.; Qiu, X.; Miller, S.; Michalak, A.M. Spatio-Temporal Approach to Moving Window Block Kriging of Satellite Data v1.0. *Geosci. Model Dev.* **2017**, *10*, 709–720. [[CrossRef](#)]
31. Myers, D.E. Pseudo-Cross Variograms, Positive-Definiteness, and Cokriging. *Math. Geol.* **1991**, *23*, 805–816. [[CrossRef](#)]
32. Ver Hoef, J.M.; Cressie, N. Multivariable Spatial Prediction. *Math. Geol.* **1993**, *25*, 219–240; Errata in *Math. Geol.* **1994**, *26*, 273–275. [[CrossRef](#)]
33. Eldering, A.; Wennberg, P.O.; Crisp, D.; Schimel, D.S.; Gunson, M.R.; Chatterjee, A.; Liu, J.; Schwandner, F.M.; Sun, Y.; O'Dell, C.W.; et al. The Orbiting Carbon Observatory-2 Early Science Investigations of Regional Carbon Dioxide Fluxes. *Science* **2017**, *358*, eaam5745. [[CrossRef](#)] [[PubMed](#)]
34. Crisp, D.; Pollock, H.R.; Rosenberg, R.; Chapsky, L.; Lee, R.A.M.; Oyafuso, F.A.; Frankenberg, C.; O'Dell, C.W.; Bruegge, C.J.; Doran, G.B.; et al. The On-Orbit Performance of the Orbiting Carbon Observatory-2 (OCO-2) Instrument and Its Radiometrically Calibrated Products. *Atmos. Meas. Tech.* **2017**, *10*, 59–81. [[CrossRef](#)]
35. OCO-2 Science Team; Gunson, M.; Eldering, A. OCO-2 Level 2 Bias-Corrected Solar-Induced Fluorescence and Other Select Fields from the IMAP-DOAS Algorithm Aggregated as Daily Files, Retrospective Processing V10r. 2020. Available online: https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_SIF_10r/summary (accessed on 1 March 2022). [[CrossRef](#)]
36. OCO-2 Science Team; Gunson, M.; Eldering, A. OCO-2 Level 2 Bias-Corrected XCO2 and Other Select Fields from the Full-Physics Retrieval Aggregated as Daily Files, Retrospective Processing V10r. 2020. Available online: https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_10r/summary (accessed on 1 March 2022). [[CrossRef](#)]
37. Friedl, M.; Sulla-Menashe, D. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V061. 2022. Available online: <https://lpdaac.usgs.gov/products/mcd12c1v061/> (accessed on 1 March 2022). [[CrossRef](#)]
38. Cressie, N.; Johannesson, G. Fixed Rank Kriging for Very Large Spatial Data Sets. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2008**, *70*, 209–226. [[CrossRef](#)]
39. Zammit-Mangion, A.; Cressie, N. FRK: An R Package for Spatial and Spatio-Temporal Prediction with Large Datasets. *J. Stat. Softw.* **2021**, *98*, 1–42. [[CrossRef](#)]
40. Cressie, N. *Statistics for Spatial Data*, revised ed.; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 1993. [[CrossRef](#)]
41. Gneiting, T.; Kleiber, W.; Schlather, M. Matérn Cross-Covariance Functions for Multivariate Random Fields. *J. Am. Stat. Assoc.* **2010**, *105*, 1167–1177. [[CrossRef](#)]
42. Apanasovich, T.V.; Genton, M.G.; Sun, Y. A Valid Matérn Class of Cross-Covariance Functions for Multivariate Random Fields with Any Number of Components. *J. Am. Stat. Assoc.* **2012**, *107*, 180–193. [[CrossRef](#)]
43. Matérn, B. *Spatial Variation—Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*; Meddelanden från Statens Skogsforskningsinstitut: Stockholm, Sweden, 1960; Volume 49.
44. Stein, M.L. *Interpolation of Spatial Data: Some Theory for Kriging*; Springer Series in Statistics; Springer: New York, NY, USA, 1999.
45. Abramowitz, M.; Stegun, I.A. (Eds.) *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, 9th ed.; Dover Books on Mathematics; Dover Publications: New York, NY, USA, 1965.
46. Jeong, J.; Jun, M.; Genton, M.G. Spherical Process Models for Global Spatial Statistics. *Stat. Sci.* **2017**, *32*, 501–513. [[CrossRef](#)]
47. Cressie, N. Fitting Variogram Models by Weighted Least Squares. *J. Int. Assoc. Math. Geol.* **1985**, *17*, 563–586. [[CrossRef](#)]
48. Cressie, N. Change of Support and the Modifiable Areal Unit Problem. *Geogr. Syst.* **1996**, *3*, 159–180.
49. Arlot, S.; Celisse, A. A Survey of Cross-Validation Procedures for Model Selection. *Stat. Surv.* **2010**, *4*, 40–79. [[CrossRef](#)]
50. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Aroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
51. Gneiting, T.; Raftery, A.E. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [[CrossRef](#)]
52. Dawid, A.P.; Sebastiani, P. Coherent Dispersion Criteria for Optimal Experimental Design. *Ann. Stat.* **1999**, *27*, 65–81. [[CrossRef](#)]

53. Gneiting, T.; Katzfuss, M. Probabilistic Forecasting. *Annu. Rev. Stat. Its Appl.* **2014**, *1*, 125–151. [[CrossRef](#)]
54. Crameri, F. Scientific Colour Maps. Zenodo, 2023. Available online: <https://zenodo.org/record/8035877> (accessed on 14 June 2023). [[CrossRef](#)]
55. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2007.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.