Graph-Based Interpolation for Remote Sensing Data

Johanna Garcia Cardona
Dept. Electrical and Computer Engineering
University of Southern California
Los Angeles, CA, USA

Antonio Ortega
Dept. Electrical and Computer Engineering
University of Southern California
Los Angeles, CA, USA

Nereida Rodriguez-Alvarez
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA, USA

***Abstract*—Remote sensing technologies, such as radiometry radar or synthetic-aperture radar (SAR) provide satellite observations with different spatio-temporal resolutions. Some of these observations can be combined to enhance the resolution of satellite products. In this paper, we formulate the data interpolation problem as a signal reconstruction on a graph, where the weights of edges between nodes are chosen as a function of high resolution ancillary data, e.g., geospatial data or observations from an instrument providing high resolution imagery. Our method is initialized with a high resolution signal directly interpolated from coarse observations obtained from a first instrument. Then, we solve a quadratic Laplacian optimization problem to ensure that the obtained high resolution observations are consistent with our original estimates while being smooth on the constructed graph. Preliminary results show that our approach leads to better remote sensed data reconstruction when compared against traditional spatial filtering and interpolation techniques.**

***Index Terms*—Graph signal processing (GSP), data merging, interpolation methods, remote sensing**

I. Introduction

Advances in satellite sensing technology are leading to the deployment of a variety of instruments that observe land surface parameters. Earth-observing satellites independently provide information at various spatial and temporal resolutions. For instance, passive sensors such as microwave radiometers provide data at resolutions of tens of kilometers ($\sim$20-40 Km); while active sensors such as radars can generate higher resolution data ($\sim$0.5-10Km). Observations obtained from passive sensors with coarse resolutions usually cannot satisfy all regional science requirements. For this reason, it is important to develop techniques to generate detailed spatial data from coarse datasets using either ancillary terrain information or observations from other instruments. While different instruments are designed to provide specific types of information (e.g., different types of images), combining their data can significantly improve satellite observations for environmental monitoring.

Existing techniques to combine multiple sources of remote sensing data have some limitations. Some methods, such as the optimally localized average Backus-Gilbert [1], can combine observations at multiple resolutions, but they do not take into consideration ancillary data, e.g., terrain characteristics that affect observations obtained from passive sensors. Techniques from standard signal processing and geostatistics [2] can increase the quality and resolution of remote sensed data, but they cannot always be used to combine heterogeneous data (with different spatio-temporal resolutions or data types).

Our goal is to generate detailed spatial data from coarse datasets using additional sources of information, such as ancillary terrain data and observations from other instruments. To achieve this, we propose a graph signal processing (GSP) interpolation method where the graph construction makes use of these additional sources of information. The fundamental advantage of our proposed GSP approach is that it provides a general methodology for graph construction and interpolation that can be used in multiple settings and allows us to combine multiple types of data [3], e.g., measurements from multiple space-borne instruments, terrain attributes, and other overlapping imagery, while working with *irregular data* and *making use of ancillary datasets*. This is in contrast with standard signal processing techniques which assume that data are located on a regular spatial grid. Datasets of interest in our work have inherent *irregularity* due to the measurement system. For example, images can have large missing regions due to cloud cover, which can lead to distorted structures or blurry textures inconsistent with surrounding areas [4].

Products for land-atmosphere observations are sometimes retrieved from radiative transfer models and require other *ancillary data* such as vegetation properties and surface roughness [5]. Also, since optical/thermal remote sensing data are affected by the presence of clouds, many existing methods cannot be applied under non-clear-sky conditions [6]. While geostatistical methods, which form an estimate of a physical quantity at an unmeasured location based on estimates of spatial correlation [7], can deal with the previously mentioned challenge of data irregularity, they cannot easily incorporate auxiliary information from multiple diverse products and can also be sensitive to outlier deviations [8].

In this paper we propose an alternative to existing methods based on graph signal interpolation. Graphs are a good fit for processing signals that: i) lie on irregular domains and ii) are the result of physical processes where observed correlations can be attributed to the effect of multiple variables. Our novel GSP approach estimates high resolution observations by using graph-based interpolation on a graph constructed with weights that are a function of ancillary data such as altitude and temperature at high resolution.

While similar graph-processing methods have been proposed in other contexts, e.g., in bilateral image, where photometric distance and pixel distance are combined [9], graph construction for processing of remotely sensed data presents additional challenges. First, the data to be interpolated, or used

as ancillary variables are available at different spatio-temporal resolutions. Second, the spatial and temporal correlation between sampled signals and terrain or meteorological variables are complex and need to be derived from data. Thus, the strength of the connection between nodes as a function of local correlation between samples, is learned from data sources that influence the signal at the nodes. A further advantage of our method is that other modalities, e.g., in-situ measurements, can be incorporated into our graphs, which can include nodes corresponding to irregularly distributed sensors.

To illustrate the benefits of this methodology we present a case study where we combine information from two satellites to obtain high resolution soil moisture estimates from coarse radiometric observations. Our results show that by considering ancillary information in the interpolation task, better fine estimates of soil moisture for regional studies can successfully be obtained. The rest of the paper is organized as follows. We introduce basic GSP concepts in Section II, then present our proposed interpolation technique in Section III, with experimental results and discussion in Section IV and conclusions in Section V.

II. Preliminaries

An undirected graph $G = (\mathcal{V}, E)$ is a collection of nodes $\mathcal{V} = 1, 2, ..., N$ connected by edges $E = (i, j, w_{ij})$, $i, j \in \mathcal{V}$ where (i, j, w_{ij}) denotes the weight w_{ij} between node i and j. The $N \times N$ weighted adjacency matrix $\mathbf{W}$ of the graph is such that $\mathbf{W}(i, j) = w_{ij}$. The degree d_i for node i is the sum of the weights of edges connected to the node. The degree matrix is $\mathbf{D} = \text{diag}(d_1, d_2, ..., d_N)$, and the combinatorial Laplacian to be used throughout our experiments is defined by $\mathbf{L} = \mathbf{D} - \mathbf{W}$.

We represent remote sensed images as *graph signals* on a grid graph with $\sqrt{N} \times \sqrt{N}$ vertices, where the value at each node is a sensed measurement and neighboring nodes are connected with horizontal and vertical edges with different non-negative edge weights. The highest similarity corresponds to the largest weight, $w_{ij} = 1$, while $w_{ij} = 0$ represents lack of an edge between those nodes (minimum similarity). A low frequency graph signal is expected to vary slowly across nodes, so that strongly connected nodes will tend to have very similar signal values [10]. Signal variation is based on how different a sample $x(i)$ at node i is from values at the nodes in its neighborhood $\mathcal{N}_i$, i.e., $\mathcal{N}_i = \{k \in \mathcal{V} : w_{ki} \neq 0)\}$. Specifically, variation can be quantified using the Laplacian quadratic form:

$$\Delta_{\mathbf{L}}(\boldsymbol{x}) = \boldsymbol{x}^\top \mathbf{L} \boldsymbol{x} = \sum_{i \sim j} w_{ij} (x(i) - x(j))^2, \qquad (1)$$

where $i \sim j$ denotes that nodes i and j are connected. Given that edge weights are non-negative, the graph signal $\boldsymbol{x}$ is considered to be smooth if strongly connected nodes have similar values, so that $\Delta_{\mathbf{L}}(\boldsymbol{x})$ is small. Thus, to take into account the information provided by ancillary data and high resolution instruments, we construct a graph with weights chosen based on this information and optimize high resolution signal reconstruction by introducing a regularization term that favors minimizing (1).

III. Graph-based Interpolation Methodology

We propose a methodology to construct graphs that allow us to combine multiple observations (from different satellites and at different resolutions) with relevant geoinformation. In this formulation, coarse observations are considered to be graph signals at certain nodes, while graph edge weights are chosen to be a function of terrain information (e.g., distance, differences in altitude or similarity between neighboring observations from another instrument). This approach relies on the implicit assumption that nodes connected with larger edge weights will tend to have more similar values. Signals at the nodes $\boldsymbol{x}$ (coarse resolution) and ancillary information $\boldsymbol{y}$ (high resolution) are both interpolated to a (target) high resolution of interest, resulting in $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$, respectively. Then, $\tilde{\boldsymbol{y}}$ is used for graph construction, while $\tilde{\boldsymbol{x}}$ is chosen to initialize the reconstruction process.

Our methodology consists of four steps: a) *Graph topology estimation*: determining the graph structure that best fits the representation of data (Section III-A), b) *Graph signal initialization*: using the coarse signal to interpolate data at the higher resolution nodes following a kriging approach (Section III-B); c) *Edge weight selection*: determining from the high resolution information (profile) the edge weights between neighboring observations following a bilateral-like filter approach with data adaptive coefficients (Section III-C); and d) *Optimization*: estimating the final high resolution signal by optimizing the Laplacian quadratic form to enforce the local smoothness property on the graph (Section III-D).

A. Graph topology estimation

Selecting a graph is a fundamental step for graph signal analysis, since the definitions of frequency and node domain operations are dependent on this choice [10]. When data is exclusively obtained from satellite images and regularly sampled ancillary data, we choose a graph topology for the high resolution image where pixels are located on a *regular grid*. In this representation, each pixel corresponds to a node and the edge weights between nodes are chosen to capture information about pixel position and pixel similarity. The graph signal associated with each node is the intensity or color information at the corresponding pixel. There are two main reasons to choose a regular grid. First, there are graph filterbank designs for weighted grid graphs for which separable implementations (row-wise and column-wise filters) are available [11]. Second, the target high resolution images will also be displayed on a regular grid and thus it is convenient to define such grid graphs for interpolation. In scenarios where data from an in-situ sensor network needs to be combined with satellite imagery it is possible to use both the regular grid graph and additional nodes representing the sensors. Nodes corresponding to in-situ sensors are connected to the closest points on the regular grid. In this case we can assign edge weights based on the Euclidean distance between the sensor's position or on the statistical correlation between signal observations at the nodes.

B. Graph Signal Initialization

We initialize the graph signal at our target resolution using kriging interpolation. At each node in the target resolution we estimate a value based on all known pixels obtained from the coarse satellite image (or from in-situ sensor observations). In order to avoid oversmoothing effects when increasing resolutions, we adopted a kriging approach since it uses the distance between observations and a correlation model to predict values for unobserved locations. This baseline interpolated signal will be used as part of our optimization, as well as a reference for comparison against our proposed GSP technique.

In our experiments the initial interpolated observations were obtained by ordinary point kriging, where an unbiased weighted average of the neighboring points x_i is used to estimate the value of an unobserved point $\hat{x}_i$. We used the kriging interpolation method proposed on [12], where the estimated values at unobserved points are obtained by minimizing the kriging variance:

$$E((\hat{x}_i - x_i)^2) = 2\sum_{i=1}^{N} \beta_i \gamma(x_i, x_0) - \sum_{i=1}^{N}\sum_{j=1}^{N} \beta_i \beta_j \gamma(x_i, x_j), \tag{2}$$

where $\gamma(x_i, x_0)$ and $\gamma(x_i, x_j)$ are the exponential variograms between the data points and the unobserved points and between the data points x_i and x_j, respectively, and β_i and β_j are the weights of the *i*th and *j*th data points. Note that other choices of initialization are also possible. For instance, the exponential variogram model in (2) can be replaced by spherical or linear models if they provide a better fit [13].

C. Edge weight selection

Our graph construction method assigns edge weights based on Euclidean distance between nodes and statistical correlation among ancillary variables. Considering this type of information has the potential to incorporate both observed data behavior (correlation) and physical system characteristics (distance), and can take advantage of the fact that ancillary variables are often available at high resolution. We define the edge weights using a combination of a static variable and a time-varying variable. In general it would be possible to use more than two variables and any combination of static and time-varying variables.

The edge weights are computed using a Gaussian kernel so that both static and dynamic information can be used in a similar manner to the one adopted in bilateral filtering [14] in image processing:

$$w_{ij} = \exp\left(-\frac{(d_s(i,j))^2}{\sigma_s^2}\right) \cdot \exp\left(-\frac{(d_t(i,j))^2}{\sigma_t^2}\right), \tag{3}$$

where $d_s(i,j)$ and $d_t(i,j)$ represent the difference in terms of static and time-varying variables, respectively, for neighboring nodes $\boldsymbol{y}_i, \boldsymbol{y}_j$. The relative values of σ_s and σ_t control the relative importance associated with these two types of variables. Our method can combine, for instance, the differences in altitude (static) and in temperature (time-varying) between two nodes. Our method can use both spatial proximity between observations as well as measures of similarity based on the respective ancillary data in the two nodes. Note that for each variable of interest, an additional Gaussian kernel can be included in (3), following a multilateral filter approach.

In order to consider temporal analysis during our graph construction, the σ parameters were calculated using a temporal correlation analysis between the observed signal and the profile of preference. Thus, each σ is a constant that embeds correlation information between the signal at a node and the profile that determines the edges weights. For future work, the σ values can be obtained through the leave-one-out cross-validation approach, where each validation set has all the ancillary data for a node in the graph node and we iterate through nodes until most of them have served as a validation set. We could then evaluate the cross-validation scores to select the best choice of terrain parameters that determine the edge weights. Note that if $d(i,j)$, one of the distances in attribute domain, becomes much larger than the parameter σ the corresponding weight approaches zero, so that all nodes can become disconnected.

D. Optimization

Embedding ancillary information on the signal estimation while preserving smoothness on the graph is accomplished by the following optimization problem:

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}} \boldsymbol{x}^\top \mathbf{L} \boldsymbol{x} + \lambda \|\boldsymbol{y} - \boldsymbol{x}\|_2{}^2, \tag{4}$$

where $\lambda \geq 0$ is the penalty parameter for the initial interpolated estimation $\boldsymbol{y}$. Note that if $\lambda >> 0$, more relevance would be given to $\boldsymbol{y}$.On the other hand, if $\lambda \sim 0$, the Laplacian quadratic form $\boldsymbol{x}^\top \mathbf{L} \boldsymbol{x}$ would dominate. Solving the optimization of (4) leads to finding a graph Laplacian $\mathbf{L}$ that enforces smoothness of the observed signal and accounts for terrain or other auxiliary information represented on the edge weights.

IV. Case Study, Results and Discussion

Soil moisture, a fundamental variable for studies that involve land-atmosphere interactions for applications in weather and agriculture, is derived from coarse radiometric brightness temperature. Using the proposed GSP method, we obtain high resolution estimates of soil moisture using land surface temperature from an additional instrument.

A. Case study

We consider a practical remote sensing scenario of soil moisture measurement to test our approach. The remote sensed datasets analyzed on the graph are generated by the Soil Moisture Active Passive (SMAP) radiometer [15], and the Moderate-Resolution Imaging Spectroradiometer (MODIS) onboard of NASA satellites Terra and Aqua [16]. Our method improves coarse satellite soil moisture measurements from the SMAP mission, using a fine resolution land surface temperature (LST) obtained from MODIS. The signal at the nodes represents a measure of soil moisture on a 36Km grid, obtained from a global daily EASE-Grid soil moisture (0-5 cm)

in cm^3/cm^3 derived from brightness temperatures (January-April 2019) [15]. The edge weights are derived from the altitude of the terrain and from land surface temperature (LST) variations in a 5.6Km grid provided by the MODIS optical imaging sensor product projected on a 0.05 degree (5.6Km) geographic climate moderate grid (CMG) [17]. We chose LST variations as additional data because daily temperature changes and the presence of vegetation affect the soil moisture [18].

The original SMAP satellite product at 36Km has been interpolated using the Backus-Gilbert optimal interpolation algorithm to increase the resolution to 9Km [19]. This enhanced product, in addition to general geolocation and calibration, has also been corrected for atmospheric effects, Faraday rotation, and low-level RFI effects prior to regridding. For this reason, we have chosen the SMAP product available at 9Km from [19] as reference for validation in our experiments.

B. Experimental results

Our proposed interpolation was tested for the geographic region of Walnut Gulch experimental watershed (WGEW). Such area encompasses the 150 square kilometers in southeastern Arizona, U.S.A., that surrounds the historical western town of Tombstone. Satellite soil moisture estimates for this region are shown in Fig. 1.

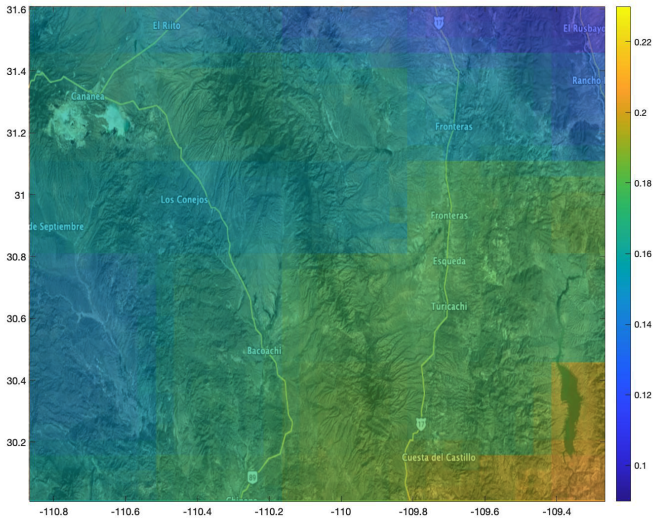

Fig. 1. Soil moisture (cm^3/cm^3) signal at a coarse 36Km resolution on Walnut Gulch study site. Note that the lake on the south-eastern corner is distinguished by the highest values of soil moisture from radiometric observations. The proposed GSP methodology allows the estimation of high resolution soil moisture using the terrain differences identified on the image.

The GSP interpolation algorithm was compared against a kriging interpolator using the enhanced SMAP product at 9Km as the validation set. The observations made in Fig. 2 demonstrate that graph-based interpolation leads to better spatial heterogeneity when the Laplacian quadratic form affects the variation between consecutive nodes; consequently, obtaining a better error performance when compared with the validation set at 9Km resolution. Our proposed GSP method presents a RMSE of 12.82% compared to a RMSE of 13.45% obtained from kriging.

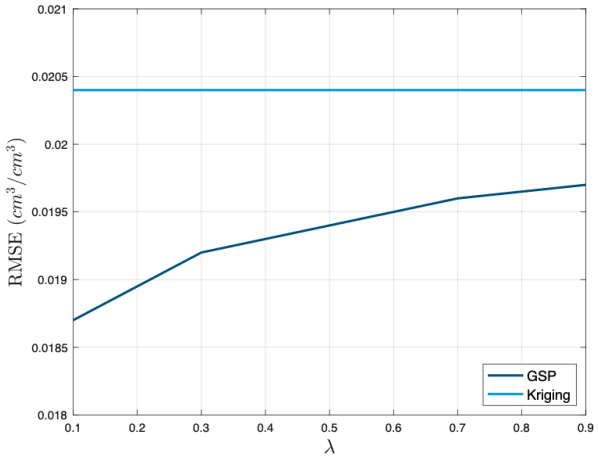

Fig. 2. Error comparison between geostatistic kriging approach and GSP using altitude and land surface temperature (LST). Remote sensed observations provide soil moisture measurements at 0-5cm depth. For such root zone, the typical soil moisture (SM) ranges are $0.02 - 0.5(cm^3/cm^3)$.

The frequency distribution histogram shown in Fig. 3 demonstrates that for small λ values, the estimated signal variation is low. On the other hand, following the optimization in (4), for a larger λ, the obtained signal would be closer to the estimates from kriging presenting higher global variation.

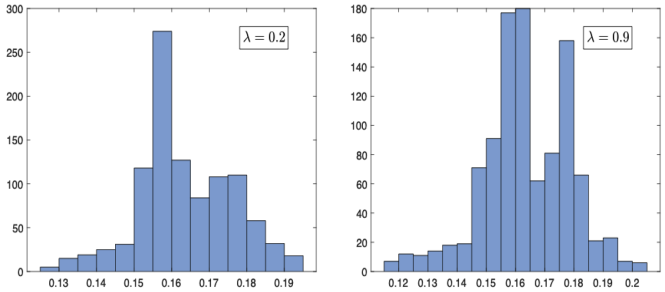

Fig. 3. Frequency distribution of estimated soil moisture at 5.6Km. Note from (4) that small λ values give more relevance to Laplacian quadratic form, thus smoothing the signal on the graph.

Rough estimates of soil moisture at 36km were downscaled at 5.6km using our graph interpolation method. Spatial results show that the GSP-based estimates present smoother transitions than a kriging approach, while incorporating terrain information in the edge weights (Fig. 4).

V. Conclusions

Given the ability of graphs to combine spatial and ancillary information, the explored GSP interpolation technique offers better estimates than traditional geostatistical interpolation methods for remote sensing applications. Signal estimation on graphs is a good alternative when simultaneously analyzing data from different spatio-temporal resolutions. GSP-based methods are also good candidates for validation of satellite

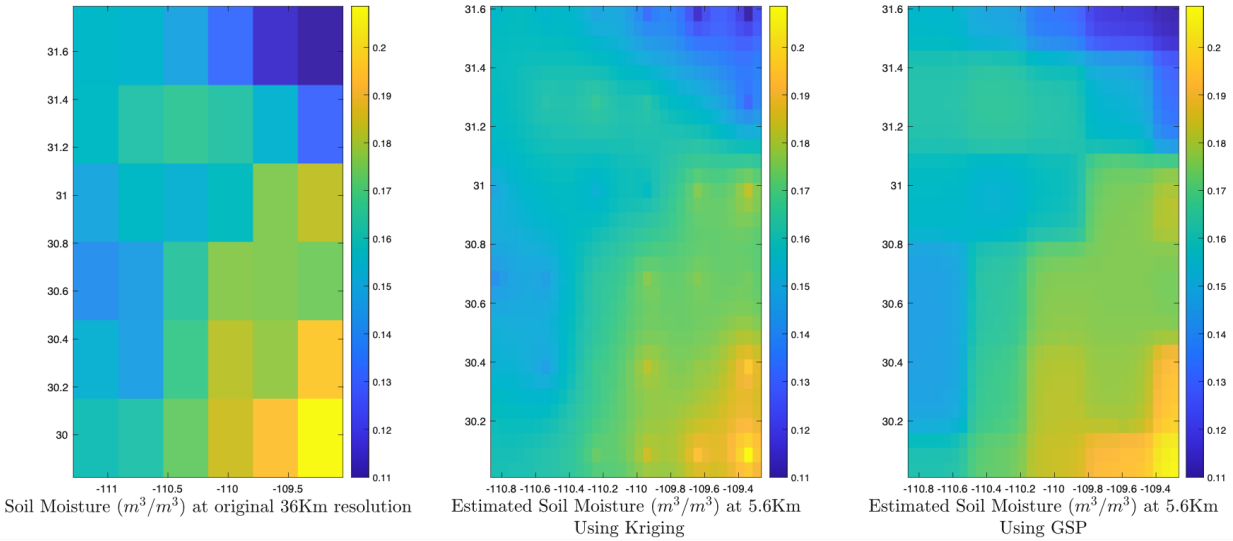

Fig. 4. Comparison of soil moisture observations. The original SMAP product at 36Km (left) is interpolated with kriging method at 5.6Km (center) and GSP-based approach at 5.6Km (right).

data using in-situ observations, since in-situ networks are often spatially distributed in an irregular manner.

VI. Acknowledgements

N. Rodriguez-Alvarez's contribution to this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

References

[1] J. Zhou and H. Yang, "Comparison of the remapping algorithms for the advanced technology microwave sounder (ATMS)," *Remote Sensing*, vol. 12, no. 4, 2020.

[2] V. Pawlowsky-Glahn, "Multivariate geostatistics, an introduction with applications, second, completely revised edition," *Computers & Geosciences - COMPUT GEOSCI*, vol. 26, pp. 1069–1070, 11 2000.

[3] J. Garcia Cardona, A. Ortega, and N. Rodriguez-Alvarez, "Downscaling SMAP soil moisture with ECOSTRESS products using a graph-based intepolation method," in *Proc. of IGARSS 2022*, (Kuala Lumpur, Malaysia), 2022.

[4] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," *CoRR*, vol. abs/1801.07892, 2018.

[5] E. G. Njoku and D. Entekhabi, "Passive microwave remote sensing of soil moisture," *Journal of hydrology*, vol. 184, no. 1-2, pp. 101–129, 1996.

[6] J. Peng, A. Loew, O. Merlin, and N. E. Verhoest, "A review of spatial downscaling of satellite remotely sensed soil moisture," *Reviews of Geophysics*, vol. 55, no. 2, pp. 341–366, 2017.

[7] N. Cressie, "Spatial prediction and ordinary kriging," *Mathematical geology*, vol. 20, no. 4, pp. 405–421, 1988.

[8] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*. Springer, 2003.

[9] P. Milanfar, "Symmetrizing smoothing filters," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 263–284, 2013.

[10] A. Ortega, *Introduction to graph signal processing*. Cambridge University Press, 2022.

[11] D. E. Tzamarias, E. Pavez, B. Girault, A. Ortega, I. Blanes, and J. Serra-Sagristà, "Orthogonality and zero DC tradeoffs in biorthogonal graph filterbanks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5509–5513, IEEE, 2021.

[12] M. H. Trauth, *Matlab recipes for earth sciences*. Springer, 2006.

[13] P. Kitanidis, *Introduction to Geostatistics. Applications in Hydrogeology*. Cambridge University Press, 1998.

[14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images.," *ICCV*, vol. 98, p. 2, 1998.

[15] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. Chaubell., "SMAP L3 radiometer global daily 36 km EASE-grid soil moisture, version 6."

[16] Z. Wan, S. Hook, and G. Hulley, "Myd11c2 modis/aqua land surface temperature/emissivity," 2015.

[17] K. Didan, "MOD13C1 MODIS/Terra vegetation indices 16-day L3 global 0.05deg cmg v006, NASA EOSDIS land processes DAAC."

[18] "An interpretation of methodologies for indirect measurement of soil water content," *Agricultural and Forest Meteorology*, vol. 77, no. 3, pp. 191–205, 1995. Thermal Remote Sensing of the Energy and Water Balance over Vegetation.

[19] J. Chaubell., "SMAP L1B enhancement radiometer brightness temperature data product."