



Bayesian Latent Variable Co-kriging Model in Remote Sensing for Quality Flagged Observations

Bledar A. KONOMI[✉], Emily L. KANG, Ayat ALMOMANI, and Jonathan HOBBS

Remote sensing data products often include quality flags that inform users whether the associated observations are of good, acceptable or unreliable qualities. However, such information on data fidelity is not consistently considered in remote sensing data analyses. Motivated by observations from the atmospheric infrared sounder (AIRS) instrument on board NASA's Aqua satellite, we propose a latent variable co-kriging model with separable Gaussian processes to analyze large quality-flagged remote sensing data sets together with their associated quality information. We augment the posterior distribution by an imputation mechanism to decompose large covariance matrices into separate computationally efficient components taking advantage of their input structure. Within the augmented posterior, we develop a Markov chain Monte Carlo (MCMC) procedure that mostly consists of direct simulations from conditional distributions. In addition, we propose a computationally efficient recursive prediction procedure. We apply the proposed method to air temperature data from the AIRS instrument. We show that incorporating quality flag information in our proposed model substantially improves the prediction performance compared to models that do not account for quality flags.

Supplementary materials accompanying this paper appear online.

Key Words: Co-kriging; Gaussian process; Markov chain Monte Carlo; Remote sensing; Separable covariance function.

1. INTRODUCTION

Remote sensing technology provides a wealth of information for understanding geophysical processes with unprecedented spatial and temporal coverage. Remote sensing data provide indirect information on these geophysical quantities of interest, which are typically estimated or inferred from instrument spectra (Susskind et al. 2003). The heterogeneous

B. A. Konomi (✉) · E. L. Kang · A. Almomani
Division of Statistics and Data Science, Department of Mathematical Sciences, University of Cincinnati, Cincinnati, USA (E-mail: alex.konomi@uc.edu).
J. Hobbs, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA.

© 2023 International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics
<https://doi.org/10.1007/s13253-023-00530-9>



24 nature of the Earth's atmosphere and surface contributes to remote sensing data records
 25 with variable quality, which is often documented by the product development teams and
 26 included as quality flags associated with observations in remote sensing data products (e.g.,
 27 [Thrastarson et al. 2021](#)). These quality flags, instead of providing quantitative uncertainty
 28 measures for the observations, indicate whether data are of good, acceptable or unreliable
 29 qualities. When the remote sensing data products are used in a wide range of downstream
 30 analyses, these quality flags are usually treated in a dichotomous way: Data flagged as unre-
 31 liable are removed, while the remaining data, no matter how they may be flagged differently
 32 (e.g., good or acceptable), are combined directly (e.g., [Zhu et al. 2015](#); [Ma and Kang 2020b](#);
 33 [Tian et al. 2020](#); [Waliser et al. 2020](#)). This practice ignores the delicate quality difference
 34 between observations with different quality flags.

35 Our work is motivated by data products from the atmospheric infrared sounder (AIRS)
 36 instrument on board NASA's Aqua satellite. The AIRS instrument collects radiance spectra,
 37 termed Level 1 data products, across the globe in the infrared portion of the spectrum. A
 38 retrieval algorithm is used to infer atmospheric quantities such as temperature and humid-
 39 ity from the spectra at $45\text{km} \times 45\text{km}$ spatial resolution, called the Level 2 data products
 40 ([Susskind et al. 2003](#)). Note that the AIRS Level 2 data products include a quality flag
 41 (QF) variable: QF value of 0 indicates observations of very good quality, QF value of 1
 42 indicates observations of acceptable quality, and QF value of 2 is deemed to be bad or unre-
 43 liable. AIRS data products have been used by weather prediction centers around the world
 44 to improve weather forecasts ([Lu and Zhang 2019](#)). They are also used to assess the skill of
 45 climate models and in applications ranging from volcanic plume detection ([Prata and Lynch](#)
 46 [2019](#)) to drought forecasting ([Behrangi et al. 2016](#)). Although data with QF value 2 are
 47 removed from such analyses, AIRS data with QF values of 1 and 0 are combined without
 48 further consideration of their quality difference indicated by their QF values (e.g., [Tian et al.](#)
 49 [2020](#); [Waliser et al. 2020](#)). In this work, we focus on AIRS Level 2 air temperature data
 50 at 100 different vertical pressure levels which we refer to as 3D (longitude, latitude and
 51 vertical atmospheric pressure level) air temperature observations. Previous work ([Cohen](#)
 52 [2003](#); [Wunch et al. 2011](#)) has mainly focused on determining the quality flag from the
 53 observed satellite data and then analyzing the populations separately. Instead, our work here
 54 focuses on using the information from quality-flagged remote sensing data to build a unified
 55 statistical model that allows for inference from all available data.

56 Since the quality flags indicate variable fidelity associated with remote sensing obser-
 57 vations, a natural statistical model to explore the dependencies of successive fidelity lev-
 58 els is the autoregressive co-kriging model ([Kennedy and O'Hagan 2000](#)). Essentially, this
 59 autoregressive co-kriging method considers a scalar and an additive discrepancy to model
 60 observations with a sequential fidelity order. Several important variations of this model
 61 have been proposed. [Qian and Wu \(2008\)](#) consider the scale discrepancy as a function of
 62 the input space by casting it as a Gaussian process (GP) which produces nonstandard con-
 63 ditional posteriors. [Perdikaris et al. \(2017\)](#) relax the autoregressive structure by using deep
 64 learning ideas to introduce nonlinear relationships, but the computational cost to train the
 65 model is significantly increased. [Konomi and Karagiannis \(2021\)](#) propose a Bayesian aug-
 66 mented hierarchical co-kriging procedure which makes possible the analysis of non-nested
 67 input and non-stationary output. These methods rely on Gaussian processes and do not scale

well with large data sets. Meanwhile, these methods are initially designed for deterministic computer models and thus do not include a term for measurement errors which is highly recommended in spatial statistics (Cressie 1993; Stein 1999).

Spatial statistical methods for big data have been evolving over the past two decades due to emergence of massive spatial data sets. Banerjee (2017) and Heaton et al. (2019) provide a good overview of these methods with comparisons. Briefly we distinguish between low-rank approximation methods (Banerjee et al. 2008; Cressie and Johannesson 2008), approximate likelihood methods (Stein et al. 2004; Gramacy and Apley 2015), sparse structures (Lindgren et al. 2011; Nychka et al. 2015; Ma and Kang 2020a; Peruzzi et al. 2020), multiple-scale approximation (Sang and Huang 2012; Katzfuss 2017) and lower-dimensional conditional distributions (Vecchia 1988; Stein et al. 2004; Datta et al. 2016; Katzfuss and Guinness 2021). Directly applying these methods within the autoregressive co-kriging framework for multi-fidelity remote sensing data is complicated. In particular, when the locations of data at different fidelity levels are not nested, the likelihood function no longer benefits from the Markovian property assumed in the model, making it complex to apply the methods aforementioned for massive spatial data (Konomi and Karagiannis 2021). Recently, Cheng et al. (2021) have extended the nearest neighbor Gaussian process (NNGP) to analyze multi-level data sets by introducing a nested NNGP reference set for each level. Although this method can handle multi-fidelity data, its way to define the reference sets makes it more appropriate when low-fidelity data are observed homogeneously over the spatial domain, but not suitable for the AIRS data described in Sect. 2 when the observations with the same QF values are clustered spatially.

In this paper, we propose a latent variable co-kriging model with separable Gaussian processes, which is able to account for multi-fidelity remote sensing data with measurement errors. Noticing that for AIRS Level 2 data products the latitude and longitude of observation locations are fixed across different vertical pressure levels, we adopt a multiplicative (separable) covariance function for the horizontal and vertical dimensions. Such a multiplicative (separable) covariance function is widely used in spatial statistics (Banerjee et al. 2014) and uncertainty quantification for computer experiments (Gramacy 2020). For AIRS air temperature data, the resulting covariance matrix can be decomposed into a Kronecker product of a purely horizontal (i.e., latitude and longitude) correlation matrix and a purely vertical (i.e., pressure level) correlation matrix. This can alleviate the computational bottleneck related to Gaussian process likelihood evaluation and spatial prediction (Genton 2007; Rougier 2008; Bilonis et al. 2013; Guillas et al. 2018; Ma et al. 2019). However, the introduction of the different QF values as well as missing data destroys this Kronecker product representation of the separable covariance structure. To facilitate efficient inference, we introduce an imputation mechanism within the Markov chain Monte Carlo (MCMC) procedure to take advantage of the latent variable representation and the data structure, which enables us to decompose the large covariance matrices into two separate computationally efficient components. Moreover, we propose a computationally efficient Monte Carlo recursive prediction procedure to make spatial prediction at high fidelity level. The proposed prediction procedure relies on a Student t process where we integrate uncertainty regarding the unknown “missing data” and parameters. We apply this proposed method to analyze level 2 AIRS air temperature data. Extensive numerical results demonstrate that compared

to methods that ignores the different QF values of data, our method provides more accurate predictions but also remains computationally efficient.

The rest of the paper is organized as follows. In Sect. 2, we introduce the quality-flagged AIRS 3D air temperature data studied in this work. Section 3 presents the latent variable co-kriging model. In Sect. 4, we give details on Bayesian inference including spatial prediction by constructing an augmented posterior using imputation of latent variables. In Sect. 5, we apply the proposed method to analyze the AIRS air temperature data and demonstrate its inferential advantages compared to methods ignoring the quality flags. We offer conclusions and discussion in Sect. 6.

2. DESCRIPTION OF THE AIRS AIR TEMPERATURE DATA

The atmospheric infrared sounder (AIRS) instrument on board NASA's Aqua satellite measures radiance spectra in infrared channels along the satellite's polar orbit. These infrared channels are sensitive to thermal emission from the atmosphere. The AIRS retrieval algorithm first obtains cloud properties and effective cloud-cleared radiance (CCR) and then uses CCR to further infer atmospheric properties including temperature and humidity at different vertical pressure levels at $45 \text{ km} \times 45 \text{ km}$ spatial resolution. A single $45 \text{ km} \times 45 \text{ km}$ areal unit is known as a field of regard (FOR) in [Susskind et al. \(2003\)](#). AIRS Level 2 data are further sectioned into pieces called granules. Each granule is roughly $2250 \text{ km} \times 1650 \text{ km}$ and thus contains $n_s = 1350$ FORs horizontally at each pressure level. There are a total of $n_p = 100$ pressure levels in AIRS support data products. Therefore, within a granule, there are potentially $1350 \times 100 = 135,000$ in the three-dimensional domain (horizontally and vertically).

It has been noted that heterogeneity of clouds yields substantial variability in the radiance, resulting in varying quality of the retrievals. Therefore, most retrieved remote sensing data for atmospheric properties including air temperature are accompanied by a quality flag (QF) variable. The AIRS convention is that QF values of 0 indicate best quality data, mostly retrieved under a clear-sky condition, values of 1 indicate acceptable quality, and values of 2 are deemed bad quality. The AIRS Level 2 products we analyze in this paper are extracted from the AIRS Version 6 retrieval support product ([Kahn et al. 2014](#)). Further details on the data products have been recently documented for the AIRS Version 7 products ([Thrastarson et al. 2021](#)).

In this paper, we focus on the 3D (longitude, latitude and pressure level) Level 2 air temperature data product from AIRS, and we choose to focus on a specific granule over the subtropical eastern Pacific Ocean. This region's weather variability spans multiple cloud regimes that present multiple challenges for remote sensing. Therefore, the region corresponding to this particular granule has been chosen to be the study region in works assessing retrievals from AIRS and other instruments such as the MAGIC validation campaign ([Zhou et al. 2015](#); [Kalmus et al. 2015](#)). Figure 1 displays the air temperature in our study region at two different pressure levels, together with their associated QF values. Note that the pressure levels are ordered in increasing values of pressure, with higher pressures closer to the surface, while lower pressure levels correspond to higher altitudes in the atmosphere. The

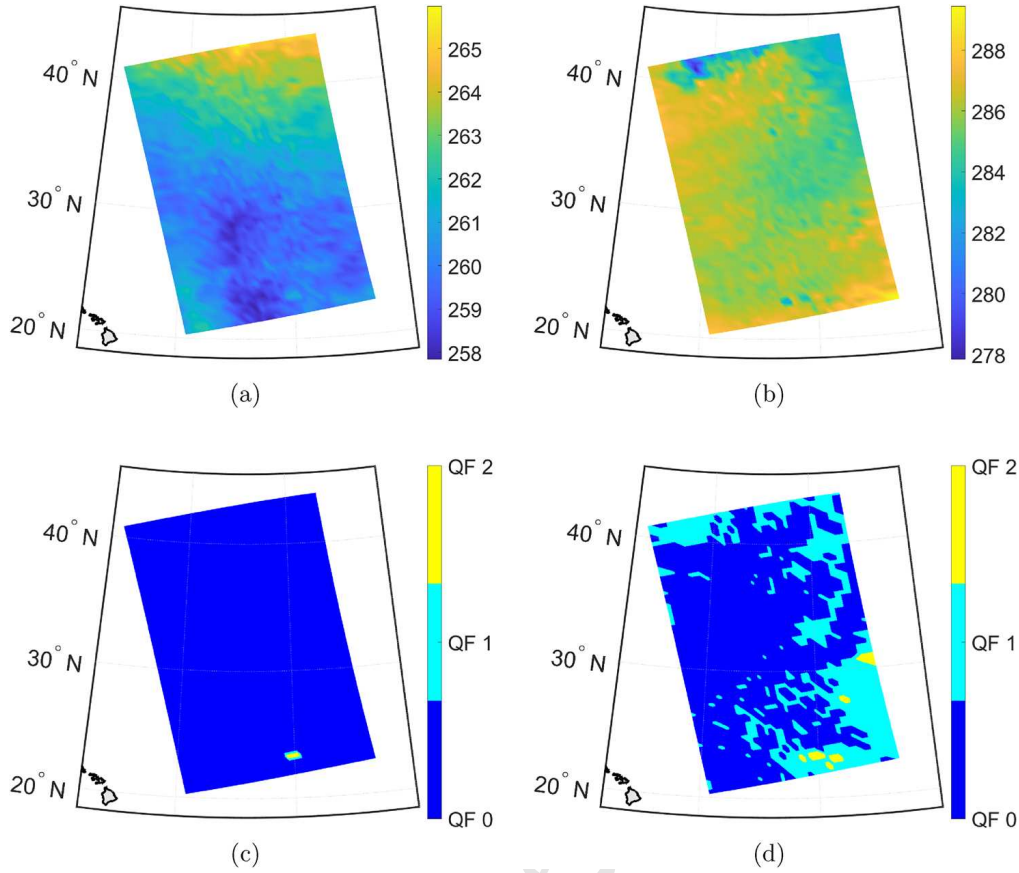


Figure 1. Top: AIRS retrieved temperature (units are Kelvin) on two different pressure levels, **a** pressure level 10 ($P = 1.2972$ hPa) and **b** pressure level 90 ($P = 958.5911$ hPa). Bottom: Retrieved temperature quality flag on two different pressure levels, **c** pressure level 10 ($P = 1.2972$ hPa) and **d** pressure level 90 ($P = 958.5911$ hPa).

number of observations with QF value of 1 (i.e., acceptable quality) and 2 (i.e., bad quality) increases as we move from low to high pressure levels. Panels (c) and (d) in Fig. 1 show that there are a lot more observations with QF value 1 and 2 at pressure level 90 (pressure level value $P = 958.5911$ hPa) compared to those at pressure level 10 (pressure level value $P = 1.2972$ hPa). Meanwhile, it is clear that observations with the same QF values form spatial clusters instead of being distributed uniformly in the region, which is expected as lower-quality observations do not occur randomly but are related to heterogeneous cloud properties at medium or fine spatial scales. Because observations with QF value of 2 are not reliable data, we treat them as missing values in our analysis. For simplicity, we will refer to air temperature data with QF value of 1 as low-fidelity observations and those with QF value of 0 as high-fidelity observations.

3. THE LATENT VARIABLE CO-KRIGING MODEL

Let $\mathbf{x} = (\mathbf{s}, p) \in \mathcal{X} = \mathcal{S} \times \mathcal{P}$ represent the input space with the $\mathbf{s} = (s_1, s_2) \in \mathcal{S} \subset \mathbb{R}^2$ denoting the longitude and latitude of the FOR centers and $p \in \mathcal{P} \subset \mathbb{R}$ denoting the pressure level values. Let $\{(z_L(\mathbf{x}), z_H(\mathbf{x})) : \mathbf{x} \in \mathcal{X}\}$ represent the observation of low fidelity (QF value 1, subscript L) and high fidelity (QF value 0, subscript H). We assume the data observed at input $\mathbf{x}$ is contaminated by additive random noise: $z_L(\mathbf{x}) = y_L(\mathbf{x}) + \epsilon_L(\mathbf{x})$ and $z_H(\mathbf{x}) = y_H(\mathbf{x}) + \epsilon_H(\mathbf{x})$. Specifically, we model the measurement error of the low fidelity as $\epsilon_L(\mathbf{x}) \sim N(0, \tau_L^2)$ and the measurement error of the high fidelity as $\epsilon_H(\mathbf{x}) \sim N(0, \tau_H^2)$.

For the latent low- and high-fidelity processes $y_L(\mathbf{x})$ and $y_H(\mathbf{x})$, we assume that they are linked via the latent variable autoregressive co-kriging model as follows:

$$\begin{aligned} y_L(\mathbf{x}) &= \mathbf{h}_L(\mathbf{x})^T \boldsymbol{\beta}_L + w_L(\mathbf{x}), \\ y_H(\mathbf{x}) &= \rho y_L(\mathbf{x}) + \delta_H(\mathbf{x}) \\ &= \rho y_L(\mathbf{x}) + \mathbf{h}_H(\mathbf{x})^T \boldsymbol{\beta}_H + w_H(\mathbf{x}), \end{aligned} \quad (1)$$

where ρ represents a constant scalar discrepancy and $\delta_H(\mathbf{x})$ represents the additive discrepancy between $y_L(\mathbf{x})$ and $y_H(\mathbf{x})$. Here, $\mathbf{h}_L(\cdot)$ and $\mathbf{h}_H(\cdot)$ are p_H and p_L known covariates to describe the mean structure of $y_L(\mathbf{x})$ and $\delta_H(\mathbf{x})$, respectively. Utilizing the structure of the observations, we use a tensor product of separate basis functions for the spatial locations and the pressure level values. Suppose, we wish to use $m_{L,p}$ basis functions to capture the pressure dependence of the lower fidelity mean: $h_{L,p} = \{h_{L,p,1}(p), \dots, h_{L,p,m_{L,p}}(p)\}$. We choose also $m_{L,s}$ basis functions to capture the spatial dependence of the mean: $h_{L,s} = \{h_{L,s,1}(\mathbf{s}), \dots, h_{L,s,m_{L,s}}(\mathbf{s})\}$. These can be, for example, any polynomial basis function of the model or any suitable functions. The global basis functions for the low and high fidelity level are formed from the tensor products: $h_L = h_{L,s} \otimes h_{L,p}$ and $h_H = h_{H,s} \otimes h_{H,p}$. Thus, the number of bases function is now $m_L = m_{L,s}m_{L,p}$ and $m_H = m_{H,s}m_{H,p}$ for the low and the high level, respectively. This strategy allows the construction of separate basis functions, in lower dimensions, with the help of the tensor product. Correspondingly, $\boldsymbol{\beta}_L$ and $\boldsymbol{\beta}_H$ are the unknown vectors of coefficient vectors. Finally, we model the latent $w_H(\cdot)$ and $w_L(\cdot)$ as mutually independent Gaussian processes: $w_L(\cdot) \sim GP(0, \sigma_L^2 R_L(\cdot, \cdot; \boldsymbol{\theta}_L))$ where σ_L^2 is the variance and $R_L(\cdot, \cdot; \boldsymbol{\theta}_L)$ is a correlation function with parameters $\boldsymbol{\theta}_L$ at fidelity level L . Similarly, $w_H(\cdot) \sim GP(0, \sigma_H^2 R_H(\cdot, \cdot; \boldsymbol{\theta}_H))$, but independent of $w_L(\cdot)$.

Note that we can generalize the scalar discrepancy to be a polynomial function of pressure and space. However, this may increase the parametric space significantly and we have not observed accuracy differences in our application. The simple Markovian condition with constant scalar parameter translates to the conditional representation of the coregionalization covariance function (Banerjee et al. 2014).

We define $\boldsymbol{\sigma}^2 = (\sigma_L^2, \sigma_H^2)$, $\boldsymbol{\tau}^2 = (\tau_L^2, \tau_H^2)$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_L, \boldsymbol{\beta}_H)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_L, \boldsymbol{\theta}_H)$. A direct Bayesian inference can be computationally costly when the number of the observations is large. The computational complexity of the likelihood is $\mathcal{O}(n_s^3 n_p^3)$ flops, and the parameters of the model cannot be expressed in closed form. For more details, see Supplementary Material S₁.

To overcome this computational issue, we assume that the covariance function can be expressed in a multiplicative (separable) form $R_L(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}_L) = R_{L,s}(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}_{L,s})R_{L,p}(p, p'; \boldsymbol{\theta}_{L,p})$ and $R_H(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}_H) = R_{H,s}(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}_{H,s})R_{H,p}(p, p'; \boldsymbol{\theta}_{H,p})$. This multiplicative (separable) model requires the specification of correlation functions for space and atmospheric pressure for each of the two fidelity levels: L and H . To ensure the well conditioning of the correlation matrices involved in the calculations, we add nugget effects for the spatial component and for the atmospheric pressure component. In the literature, these values are usually fixed based on a cross-validation exploration and they are expected to be typically small (of the order of 10^{-6}). In practice, considering these parameters as unknown and estimating them from the data not only improves the stability of the computations but also can lead to better predictive accuracy as shown in our application.

Based on a preliminary study on AIRS data, for both levels of fidelity, we choose a product exponential correlation function:

$$R_{.,s}(\mathbf{s}, \mathbf{s}' | \boldsymbol{\theta}_{.,s}) = \exp \left(- \sum_{i=1}^2 \frac{|s_i - s'_i|}{\phi_{.,i,s}} \right) + g_{.,s}^2 \delta(\mathbf{s}, \mathbf{s}') \quad (2)$$

$$R_{.,p}(p, p' | \boldsymbol{\theta}_{.,p}) = \exp \left(- \frac{|p - p'|}{\phi_{.,p}} \right) + g_{.,p}^2 \delta(p, p'), \quad (3)$$

where $\boldsymbol{\theta}_{.,s} = (\boldsymbol{\phi}_{.,s}, g_{.,s}^2)$, $\boldsymbol{\theta}_{.,p} = (\phi_{.,p}, g_{.,p}^2)$, $\phi_{.,1,s}$ controls the spatial dependence strength in longitude, $\phi_{.,2,s}$ controls the spatial dependence strength in latitude, and $\phi_{.,p}$ controls the spatial dependence strength in the pressure level values. Also $g_{.,s}^2$ accounts for the nugget effect in the spatial component and $g_{.,p}^2$ accounts for the nugget effect in the atmospheric pressure component. More intricate covariance functions such as the Matérn family (Cressie 1993; Stein 1999; Banerjee et al. 2014) or non-stationary models (Paciorek and Schervish 2006; Konomi et al. 2014) can also be used within the proposed latent variable co-kriging model.

Despite this flexible and general representation of the separable covariance function, the covariance matrix of the observations or the latent variables cannot be represented as a Kronecker product of separate components. This is simply because none of the input combinations of low- and high-fidelity data can be represented as a tensor product. The presence of missing data for both quality flags destroys the Kronecker product representation of the covariance matrices and makes the computations impossible in practice.

4. BAYESIAN INFERENCE

We make computations and practical implementation possible by constructing an augmented posterior which is based on the imputation of latent variables. The augmented posterior takes advantage of the tensor product of the locations and the pressure level values in a granule. Based on this augmented posterior, we are also able to construct a MCMC procedure with mostly closed-form conditional distributions for parameter inference as well as a computationally efficient recursive prediction procedure.

4.1. AUGMENTED POSTERIOR

Let $\{\mathbf{D}^H, \mathbf{D}^L, \mathbf{D}^M\}$ be the observed input location set of QF 0, QF 1 and QF 2, respectively. The union of these three input $\tilde{\mathbf{D}} = \mathbf{D}^H \cup \mathbf{D}^L \cup \mathbf{D}^M$ represents a granule and is defined in a grid format. In addition, assume the $\{z_H, z_L, z_M\}$ be the observed output set of quality flag 0, quality flag 1 and quality flag 2, respectively. Assume sets of points $\tilde{D}^L$ such that $\tilde{\mathbf{D}}^L = \mathbf{D}^H \cup \mathbf{D}^M$ is defined as the relative complement of $\mathbf{D}^L$ in $\tilde{\mathbf{D}}$. Further, assume sets of points $\tilde{D}^H$ such that $\tilde{\mathbf{D}}^H = \mathbf{D}^L \cup \mathbf{D}^M$ is defined as the relative complement of $\mathbf{D}^H$ in $\tilde{\mathbf{D}}$. Let $\tilde{z}_L = z_L(\tilde{D}^L)$ and $\tilde{z}_H = z_H(\tilde{D}^H)$ be the missing output values of the temperature at the corresponding input points in the low-fidelity (quality flag 1) and high-fidelity (quality flag 0) data, respectively. We refer to $\{\tilde{z}_L, \tilde{\mathbf{D}}^L\}$ and $\{\tilde{z}_H, \tilde{\mathbf{D}}^H\}$ as the missing data set. We also refer to $\{\tilde{z}_L, \tilde{\mathbf{D}}\}$ and $\{\tilde{z}_H, \tilde{\mathbf{D}}\}$ as the complete data set of the low and high fidelity level, respectively.

Assume that $\tilde{\mathbf{w}}_L$ is the latent variables of the low fidelity level (QF 1) obtained in $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{w}}_H$ is the latent variables of the high fidelity level (QF 0) obtained in $\tilde{\mathbf{D}}$. The covariance function of both $\tilde{\mathbf{w}}_L$ and $\tilde{\mathbf{w}}_H$ is defined in a grid where we can take advantage of the Kronecker product representation of the separable covariance function. For simplicity in the formulation below, let $\Theta = (\theta, \beta, \rho, \sigma^2, \tau^2)$, $\tilde{\mathbf{w}} = (\tilde{\mathbf{w}}_L, \tilde{\mathbf{w}}_H)$, and $\tilde{\mathbf{z}} = (\tilde{z}_L, \tilde{z}_H)$. We assign a prior distribution on the parameter $(\Theta, \tilde{\mathbf{w}})$, such as:

$$p(\Theta, \tilde{\mathbf{w}}) = p(\beta_L, \sigma_L^2) p(\theta_L) p(\rho, \beta_H, \sigma_H^2) p(\theta_H) p(\tau_L^2) p(\tau_H^2) p(\tilde{\mathbf{w}}_L | \theta_L) p(\tilde{\mathbf{w}}_H | \theta_H).$$

The GP parameters are assumed a priori independent of each other for different fidelity level data. Within each fidelity level, we choose non-informative priors for hyper-parameters (β_L, σ_L^2) and $((\beta_H, \rho), \sigma_H^2)$ and $\phi_{\cdot,s}, \phi_{\cdot,p}, g_{\cdot,s}^2$, and $g_{\cdot,p}^2$ are considered to be independent truncated *Gamma* distributions.

The joint posterior distribution of the above model given observations $\mathbf{z} = (z_L, z_H)$ is:

$$\begin{aligned} p(\Theta, \tilde{\mathbf{w}}, \tilde{\mathbf{z}} | \mathbf{z}) &\propto p(\tilde{\mathbf{z}} | \Theta, \tilde{\mathbf{w}}, \mathbf{z}) p(\mathbf{z} | \Theta, \tilde{\mathbf{w}}) p(\tilde{\mathbf{w}} | \Theta) p(\Theta) \\ &= f(\tilde{z}_L | \theta_L, \beta_L, \tau_L^2, \tilde{\mathbf{w}}_L) f(\tilde{z}_H | \theta_H, \beta_H, \rho, \tau_H^2, \tilde{\mathbf{w}}_H, \tilde{\mathbf{w}}_L) \\ &\quad \times p(\Theta_L) p(\tilde{\mathbf{w}}_L | \theta_L) p(\Theta_H) p(\tilde{\mathbf{w}}_H | \theta_H), \end{aligned} \quad (4)$$

where the conditional distributions $f(\tilde{z}_L | \cdot)$ and $f(\tilde{z}_H | \cdot)$ are independent multivariate Gaussian probability density functions. The joint posterior distribution of $\tilde{\mathbf{w}}$ can be factorized in two parts, because the proposed augmentation artificially creates the same design for both latent variables. The Markovian condition induces the required conditional independence. Given the independent specification of the prior, it is easy to see that the augmented posterior is also factorized into two conditionally independent parts. Note that $p(\Theta, \tilde{\mathbf{w}}, \tilde{\mathbf{z}} | \mathbf{z})$ admits the posterior of interest $p(\Theta, \mathbf{w} | \mathbf{z})$ as marginal by construction, and hence leads to the same Bayesian analysis.

4.2. MCMC SAMPLER

We design a MCMC sampler, targeting the augmented posterior, that involves a random permutation scan of blocks updating $[\tilde{z}|\mathbf{z}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\tau}^2]$, $[\tilde{\mathbf{w}}|\tilde{z}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\tau}^2]$, $[\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2|\tilde{z}, \tilde{\mathbf{w}}, \boldsymbol{\theta}]$ from a direct Gibbs sampler with closed-form distributions, and $[\boldsymbol{\theta}|\tilde{z}, \tilde{\mathbf{w}}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2]$ from a Metropolis–Hastings (M–H) step. We are not only reducing computational complexity based on the covariance structure, but also are able to find closed-form conditional distributions for $\sigma_L^2|\boldsymbol{\theta}_L, \tau_L^2, \tilde{\mathbf{w}}, \sigma_H^2|\boldsymbol{\theta}_H, \tau_H^2, \tilde{\mathbf{w}}, \tau_L^2|\sigma_L^2, \boldsymbol{\theta}_L, \tilde{z}, \tilde{\mathbf{w}}$ and $\tau_H^2|\sigma_H^2, \boldsymbol{\theta}_H, \tilde{z}, \tilde{\mathbf{w}}$ and do not need a high-dimensional M–H that can slow down convergence. Details regarding the MCMC blocks are explained below.

The full conditional posterior of $\tilde{\mathbf{w}}_L$ is normal $\tilde{\mathbf{w}}_L|\dots \sim N(\boldsymbol{\mu}_{\tilde{\mathbf{w}}_L}, \mathbf{R}_{\tilde{\mathbf{w}}_L})$ and $\tilde{\mathbf{w}}_H$ is normal $\tilde{\mathbf{w}}_H|\dots \sim N(\boldsymbol{\mu}_{\tilde{\mathbf{w}}_H}, \mathbf{R}_{\tilde{\mathbf{w}}_H})$, where

$$\boldsymbol{\mu}_{\tilde{\mathbf{w}}_L} = \mathbf{R}_{\tilde{\mathbf{w}}_L} \left[\frac{1}{\tau_L^2} \mathbf{I}(\tilde{\mathbf{z}}_L - \mathbf{H}_L \boldsymbol{\beta}_L) + \frac{1}{\tau_H^2} \mathbf{I}(\rho^{-1}(\tilde{\mathbf{z}}_H - \mathbf{H}_H \boldsymbol{\beta}_H - \tilde{\mathbf{w}}_H) - \mathbf{H}_L \boldsymbol{\beta}_L) \right] \quad (5)$$

$$\mathbf{R}_{\tilde{\mathbf{w}}_L} = \left(\frac{1}{\tau_L^2} \mathbf{I} + \frac{1}{\tau_H^2} \mathbf{I} + \frac{1}{\sigma_L^2} \mathbf{R}_{L,s}^{-1} \otimes \mathbf{R}_{L,p}^{-1} \right)^{-1}, \quad (6)$$

$$\boldsymbol{\mu}_{\tilde{\mathbf{w}}_H} = \mathbf{R}_{\tilde{\mathbf{w}}_H} \frac{1}{\tau_H^2} \mathbf{I}_n(\tilde{\mathbf{z}}_H - \mathbf{H}_H \boldsymbol{\beta}_H - \rho(\mathbf{H}_L \boldsymbol{\beta}_L + \tilde{\mathbf{w}}_L)), \quad (7)$$

$$\mathbf{R}_{\tilde{\mathbf{w}}_H} = \left(\frac{1}{\tau_H^2} \mathbf{I} + \frac{1}{\sigma_H^2} \mathbf{R}_{H,s}^{-1} \otimes \mathbf{R}_{H,p}^{-1} \right)^{-1}. \quad (8)$$

The availability of $\tilde{\mathbf{w}}_L$ and $\tilde{\mathbf{w}}_H$ makes the computation of the missing variables easy since both $[\tilde{z}_L|\tilde{\mathbf{w}}_L, \boldsymbol{\beta}_L, \boldsymbol{\tau}_L]$ and $[\tilde{z}_H|\tilde{\mathbf{w}}_H, \boldsymbol{\beta}_H, \boldsymbol{\tau}_H]$ follow independent normal distribution with constant diagonal variance. The conditional posterior $\pi(\boldsymbol{\beta}, \rho, \sigma^2, \boldsymbol{\tau}^2|\tilde{z}, \tilde{\mathbf{w}}, \boldsymbol{\theta})$ has the form:

$$\boldsymbol{\beta}_L|\tilde{z}, \tilde{\mathbf{w}}, \sigma^2, \boldsymbol{\theta} \sim N(\hat{\boldsymbol{\beta}}_L, \hat{\mathbf{V}}_{b,L}^*) \quad (9)$$

$$(\rho, \boldsymbol{\beta}_H)|\tilde{z}, \tilde{\mathbf{w}}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\beta}_L \sim N((\hat{\rho}, \hat{\boldsymbol{\beta}}_H), \hat{\mathbf{V}}_{b,H}^*) \quad (10)$$

$$\sigma_L^2|\tilde{\mathbf{w}}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\beta} \sim IG(\hat{a}_{\sigma_L}, \hat{b}_{\sigma_L}) \quad (11)$$

$$\sigma_H^2|\tilde{\mathbf{w}}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\beta} \sim IG(\hat{a}_{\sigma_H}, \hat{b}_{\sigma_H}) \quad (12)$$

$$\tau_L^2|\tilde{z}, \tilde{\mathbf{w}}, \boldsymbol{\beta} \sim IG(\hat{a}_{\tau_L}, \hat{b}_{\tau_L}) \quad (13)$$

$$\tau_H^2|\tilde{z}, \tilde{\mathbf{w}}, \boldsymbol{\beta} \sim IG(\hat{a}_{\tau_H}, \hat{b}_{\tau_H}), \quad (14)$$

where the hatted quantities are given in Supplementary Section S₂ Eqs. (3)–(14). The conditional posterior $p(\boldsymbol{\theta}_L|\tilde{\mathbf{w}}_L, \sigma_L^2)$ and $p(\boldsymbol{\theta}_H|\tilde{z}, \tilde{\mathbf{w}}, \sigma_H^2)$ cannot be sampled directly. Conditional independence in (4) implies that $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_H$ can be simulated by running in parallel two

Metropolis–Hastings algorithms each of them targeting distributions. These full conditional distributions are known up to some normalizing constant:

$$p(\theta_L | \tilde{\mathbf{w}}_L, \sigma_L^2) \propto p(\mathbf{g}_L^2) p(\phi_L) \sigma_L^{-1} |\mathbf{R}_{L,s}|^{-1/2} |\mathbf{R}_{L,p}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma_L^2} \tilde{\mathbf{w}}_L^T (\mathbf{R}_{L,s}^{-1} \otimes \mathbf{R}_{L,p}^{-1}) \tilde{\mathbf{w}}_L \right\},$$

$$p(\theta_H | \tilde{\mathbf{w}}_H, \sigma_H^2) \propto p(\mathbf{g}_H^2) p(\phi_H) \sigma_H^{-1} |\mathbf{R}_{H,s}|^{-1/2} |\mathbf{R}_{H,p}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma_H^2} \tilde{\mathbf{w}}_H^T (\mathbf{R}_{H,s}^{-1} \otimes \mathbf{R}_{H,p}^{-1}) \tilde{\mathbf{w}}_H \right\}.$$

In general, the whole sampler can be described as a Metropolis-within-Gibbs sampler. This is a computationally efficient sampler since avoids the inversion of big matrices and also most of the parameters are updated via a closed-form conditional distribution.

4.3. PREDICTION

Assume there is available a MCMC sample $\mathcal{S}^N = (\tilde{\mathbf{w}}, \sigma^2, \theta, \tau^2, \beta)$ generated from the MCMC sampler in Sect. 4.2. The central limit theorem can be applied to facilitate inference as the proposed sampler is aperiodic, irreducible, and reversible (Roberts and Rosenthal 2004). The proposed latent variable co-kriging model with separable covariance structure (LVCS) allows inference to be performed for the “missing” output $\hat{\mathbf{y}}_H$ at input points in $\hat{\mathcal{D}}^H$. Inference on $\hat{\mathbf{y}}_H$ can be particularly useful when the retrieval algorithm has been unable to generate reliable output due to the presence of clouds ($QF = 2$). The marginal posterior distribution of $\hat{\mathbf{y}}_H$, along with its expectations, can be approximated via standard Monte Carlo (MC) using the generated samples $\{\hat{\mathbf{y}}_H\}$. Alternatively, point estimates of $\hat{\mathbf{y}}_H$ at $\hat{\mathcal{D}}^H$ can be approximated by the more accurate Rao–Blackwell MC estimator $E(\hat{\mathbf{y}}_H | \mathbf{z}_L, \mathbf{z}_H) \approx \frac{1}{N} \sum_{j=1}^N \hat{\boldsymbol{\mu}}_H^{(j)}$, where $\{\hat{\boldsymbol{\mu}}_H^{(j)}\}$ is the j th mean MCMC realization for $\hat{\mathbf{y}}_H$.

To retrieve the temperature values $y_H(\mathcal{D}^*)$ at unmeasured input points $\mathcal{D}^*$, we propose a Monte Carlo recursive prediction procedure which is able to facilitate fully Bayesian predictive inference on the output. We first obtain $\mathbf{y}_L(\mathcal{D}^*)$ and then use it to obtain $\mathbf{y}_H(\mathcal{D}^*)$. A direct prediction of $y_H(\mathcal{D}^*)$ can be computationally not feasible since the prediction distribution of the quality flag 0 cannot be simplified as it happens for the joint prediction distribution. The conditional distribution $[(y_L(\cdot), y_H(\cdot)) | \tilde{\mathbf{w}}_L, \tilde{\mathbf{w}}_H, \theta, \sigma^2, \beta, \rho]$ inherits a conditional independence similar to the likelihood due to the augmentation of the latent variable $(\mathbf{w}_L, \mathbf{w}_H)$. Hence, after integrating out β, ρ and σ^2 we have Student T processes (STP) for the conditional representation as:

$$y_L(\cdot) | \tilde{\mathbf{w}}_L, \theta_L \sim \text{STP}(\mu_L^*(\cdot | \tilde{\mathbf{w}}_L, \theta_L), \hat{\sigma}_L^2 \hat{R}_L^*(\cdot, \cdot | \theta_L), 1 + \tilde{n}); \quad (15)$$

$$y_H(\cdot) | y_L(\cdot), \tilde{\mathbf{w}}, \theta_H \sim \text{STP}(\mu_H^*(\cdot | \tilde{\mathbf{w}}, \theta_H), \hat{\sigma}_H^2 \hat{R}_H^*(\cdot, \cdot | y_L(\cdot), \theta_H), 1 + \tilde{n}), \quad (16)$$

where

$$\hat{\mu}_L^*(x | \tilde{\mathbf{w}}_L, \theta_L) = \mathbf{H}_L(x; \tilde{\mathbf{w}}_L) \hat{\boldsymbol{\beta}}_L + \mathbf{R}_L(x, \tilde{D}) \mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) \tilde{\mathbf{w}}_L$$

$$\hat{\mu}_H^*(x | \tilde{\mathbf{w}}_H, \theta_H) = \hat{\mu}_L^*(x | \tilde{\mathbf{w}}_L, \theta_L) + \mathbf{H}_H(x; \tilde{\mathbf{w}}_H) \hat{\boldsymbol{\beta}}_H + \mathbf{R}_H(x, \tilde{D}) \mathbf{R}_H^{-1}(\tilde{D}, \tilde{D}) \tilde{\mathbf{w}}_H$$

and

$$\begin{aligned}
 \hat{R}_L^*(\mathbf{x}, \mathbf{x}' | \theta_L) &= \mathbf{R}_L(\mathbf{x}, \mathbf{x}') - \mathbf{R}_L(\mathbf{x}, \tilde{D}) \mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) \mathbf{R}_L^\top(\mathbf{x}', \tilde{D}) \\
 &\quad + \left[\mathbf{H}_L(\mathbf{x}) - \mathbf{R}_L(\mathbf{x}, \tilde{D}) \mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) \mathbf{H}_L(\tilde{D}) \right] \hat{\mathbf{A}}_L \\
 &\quad \times \left[\mathbf{H}_L(\mathbf{x}') - \mathbf{R}_L(\mathbf{x}', \tilde{D}) \mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) \mathbf{H}_L(\tilde{D}) \right]^\top \\
 \hat{R}_H^*(\mathbf{x}, \mathbf{x}' | y_L, \theta_H) &= \mathbf{R}_L(\mathbf{x}, \mathbf{x}') - \mathbf{R}_H(\mathbf{x}, \tilde{D}) \mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) \mathbf{R}_H^\top(\mathbf{x}', \tilde{D}) \\
 &\quad + \left[\mathbf{L}_H(\mathbf{x}; y_L) - \mathbf{R}_H(\mathbf{x}, \tilde{D}) \mathbf{R}_H^{-1}(\tilde{D}, \tilde{D}) \mathbf{L}_H(\tilde{D}; y_L) \right] \hat{\mathbf{A}}_H \\
 &\quad \times \left[\mathbf{L}_H(\mathbf{x}'; y_L) - \mathbf{R}_H(\mathbf{x}', \tilde{D}) \mathbf{R}_H^{-1}(\tilde{D}, \tilde{D}) \mathbf{L}_H(\tilde{D}; y_L) \right]^\top
 \end{aligned}$$

for $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, and $\hat{\mathbf{A}}_L = (\mathbf{H}_L^\top \mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) \mathbf{H}_L)^{-1}$, $\mathbf{L}_H(\cdot; y_L) = [\mathbf{H}_H(\cdot), y_L(\cdot)]$, $\hat{\mathbf{A}}_H = (\mathbf{L}_H^\top \mathbf{R}_H^{-1}(\tilde{D}, \tilde{D}) \mathbf{L}_H)^{-1}$, $\mathbf{R}_L^{-1}(\tilde{D}, \tilde{D}) = \mathbf{R}_{L,s} \otimes \mathbf{R}_{L,p}$, $\mathbf{R}_H^{-1}(\tilde{D}, \tilde{D}) = \mathbf{R}_{H,s} \otimes \mathbf{R}_{H,p}$, $\mathbf{R}_L(\mathbf{x}, \tilde{D}) = (\mathbf{R}_{L,s}(s, \tilde{D}) \otimes \mathbf{R}_{L,p}(p, \tilde{D}))$, and $\mathbf{R}_H(\mathbf{x}, \tilde{D}) = (\mathbf{R}_{H,s}(s, \tilde{D}) \otimes \mathbf{R}_{H,p}(p, \tilde{D}))$.

The proposed prediction procedure integrates uncertainty regarding the unknown “missing data” and parameters. It is computationally preferable compared to a one-step prediction, based on the prediction distribution of the high fidelity only, because it allows the parallel inversion of smaller covariance matrices with sizes $n_s \times n_s$ and $n_p \times n_p$ while the others require the inversion of a large covariance matrix of size $n_s n_p \times n_s n_p$. Moreover, it is able to recover the whole predictive distribution and its moments.

5. DATA ANALYSIS AND RESULTS

This section conducts a full analysis of the AIRS data set described in Sect. 2 using the proposed latent variable co-kriging model with separable covariance structure (LVCS). We compare LVCS with two alternatives that ignores the flag values: (a) the latent variable separable Gaussian process (SGP) model and (b) the non-separable additive approximate Gaussian process (AAGP) (Ma et al. 2019). The SGP is a special case of our model, assuming only one level and ignoring the quality flag information. SGP is an extended version of the separable model proposed in Bilionis et al. (2013), since we account for spatial error. For both SGP and AAGP, data with QF values 0 and 1 are combined to fit the corresponding model. Analyses are performed in MATLAB R2020a, on a computer with specifications (intelR i7-3770 3.4GHz Processor, RAM 8.00GB, MS Windows 64bit). In addition, we consider two variations for each of the three models: a) fix the variance of the nugget effects ($\mathbf{g}^2$) within the correlation functions based on a cross-validation exploration and b) sample $\mathbf{g}^2$ within the MCMC procedure as explained in Sect. 4.2.

Each granule has 135, 000 estimated temperature values of three different quality flags. Since the vertical pressure levels for atmospheric variables increase in an exponential order, we take a logarithmic transformation of the pressure level values. In each specific FOR, the temperature is taken vertically at 100 distinct pressure level values. Pressure level values with non-available (NA) data are removed prior to the analysis. These pressure level values

are below the surface pressure and correspond to structurally missing/undefined data due to physical constraints.

We start by testing our proposed model in a granule of August 1 of 2013 within the MAGIC validation campaign (Zhou et al. 2015). The last 4 of the distinct pressure levels have only NA data, so we consider only 96 distinct pressure levels in our analysis. Figure S1a shows the temperature as a function of the log of pressure level values for a collection of FORs with only QF 0, and Figure S1b shows the temperature as a function of pressure level values for FORs with variable quality flags. From these plots as well as Figure 1, one can observe that the QF 1 and 2 data are more common in higher vertical pressure levels (closer to the sea surface) within the vertical profile.

To simulate a realistic missing data (i.e., with QF 2) scenario, we randomly select testing data in a block of input. Since missing observations are more common in higher vertical atmospheric pressure values (close to the sea surface), we only consider vertical atmospheric pressure values higher than 114.0070 hPa (pressure level greater than 47). Evaluation of predictive performance is based on mean squared prediction errors (MSPE), coverage probability of the 95% equal tail credible interval (CVG(95%)), average length of the 95% equal tail credible interval (ALCI(95%)) and continuous rank probability score (CRPS) (Gneiting and Raftery 2007).

Based on the relationship of the log pressure values with the data in Figure S1, we chose polynomial basis functions of degree three for the log pressure level. Since we do not observe a spatial structure of the mean of temperature given a pressure level, we chose a constant mean for the spatial basis. The basis functions of the complete data are created based on the Kronecker product of these two bases. Based on an empirical study between three different covariance functions, the square exponential, exponential and the 5/2 Matérn, we chose the exponential covariance function. For all the tested days and models, the exponential covariance function gave the best prediction performance.

We apply the same specifications for all three models, the proposed LVCS, the separable model, and the AAGP. To speed up computations in the AAGP, we specify a grid of knots: 15 knots for pressure level, 8 knots for longitude and 6 for latitude. In total, we use 720 knots for the predictive process part and the same specification for the separable part. For the Bayesian inference of the three models on the unknown parameters β_L , β_H , we assigned independent Normal prior distributions with zero mean and large variances. We used inverse Gamma priors for the spatial and noise variances σ_L^2 , σ_H^2 , τ_L^2 , τ_H^2 as $IG(2, 1)$. The range correlation parameters in space are assigned a uniform prior $U(0, 100)$, and the range correlation parameters in logarithmic pressure level value a uniform prior $U(0, 20)$. For all three models, we ran the MCMC sampler as described in Sect. 3 with 25,000 iterations where the first 5000 iterations were discarded as burn-in. For all the parameters of the three models, the MCMC converges within the first 2,000 iterations. The convergence of the MCMC sampler for each parameter was assessed from their associated trace plots.

Table 1 shows the results with the prediction performances for each model when we assume an unknown nugget effect (g^2) within the correlation functions as described in Sect. 3 and when we fix it to 10^{-6} . Based on the MSPE and CRPS, the proposed LVCS kriging gives better prediction results than the separable kriging and the AAGP for both fixed and random nugget effects cases. Making the nugget effect random improves the predictability

Table 1. Method comparisons

Model	Nugget	MSPE	CRSP	CVG 95%	ALCI 95%	Time (h)
SGP	Fixed	0.2714	0.3011	84.9%	1.8113	3.973
	Random	0.2503	0.2974	89.2%	1.9213	5.977
AAGP	Fixed	0.2783	0.3361	91.7%	2.089	65.402
	Random	0.2482	0.3361	92.3%	2.139	96.419
LVCS	Fixed	0.1341	0.2323	92.9%	1.58	7.801
	Random	0.1081	0.2270	96.1%	1.63	10.432

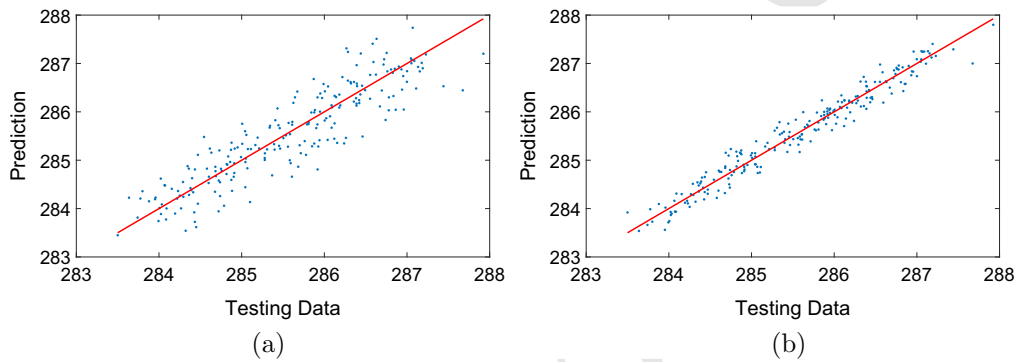


Figure 2. Scatter plot of predicted values against held-out testing data for the last three pressure levels values using two different models: **a** Separable model, and **b** the proposed latent variable co-kriging model with separable covariance structure (LVCS) .

for all models. The proposed LVCS has smaller MSPE and CRPS than both the separable kriging and the AAGP kriging. Specifically, the MSPE of the proposed model is less than half of both AAGP and separable model. The LVCS has nominal coverage probability close to 0.95 and short interval length for the 95% credible interval. The separable model and the AAGP kriging have very similar MSPE which indicates that AIRS data have a separable covariance structure. The computational time for the AAGP is almost 10 times slower. It is worth pointing out that we can improve the computational time of AAGP by decreasing the number of predictive process knots. However, this may result in missing a potential non-separable small-scale variation.

To better demonstrate the prediction benefit of accounting for the quality flag into our model, we plot the predicted temperature values against held-out temperature values for pressure level values between 750 hPa and 850 hPa using: (a) Figure 2a, the model which ignores the quality flag of the data, and (b) Fig. 2b, the proposed LVCS model which accounts for different quality flags. From these scatter plots, it is clear that accounting for the different quality of the retrievals improves predictions. Specifically the predicted temperature values of the proposed LVCS scattered in a narrower interval around the 45 degree straight red line. The fact that the predicted values when ignoring the fidelity level are still scattered around the 45 degree straight red line indicates that the flag one observations may be unbiased estimates of the temperature with bigger variance.

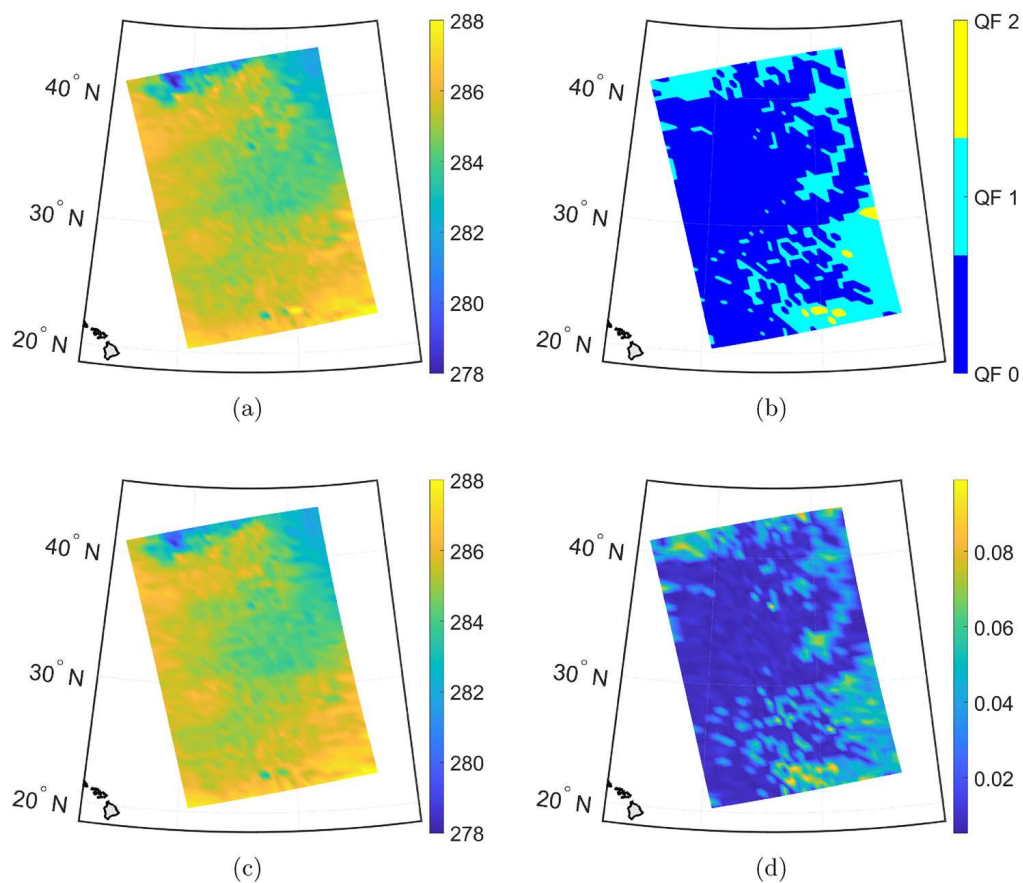


Figure 3. **a** AIRS air temperature data (units in Kelvin), **b** corresponding quality flag values, **c** predicted means of the temperature with quality flag value 0, **d** prediction standard error .

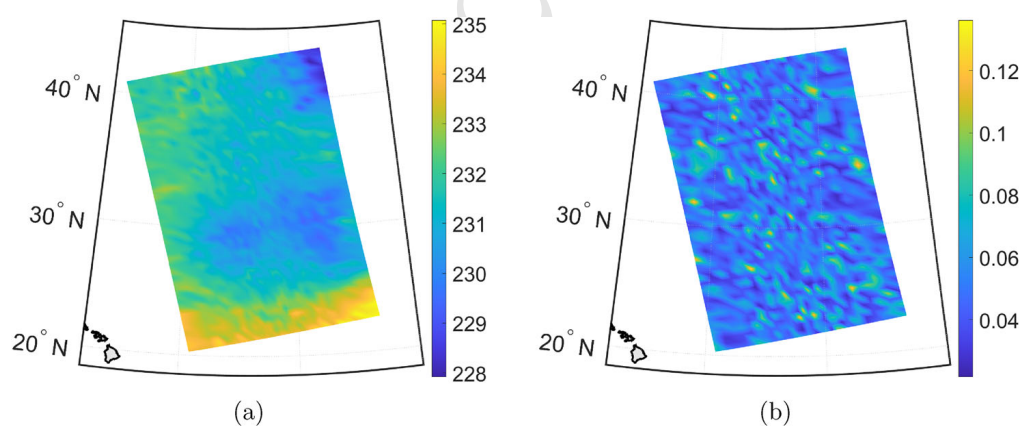


Figure 4. **a** Predicted means of the temperature (units are Kelvin) at 265 hPa with quality flag value 0, **b** prediction standard error .

Table 2. MSPE (in Kelvin) for multiple dates of two different models (a) ignoring the quality flag and (b) the proposed LVCS (which accounts for the quality flag)

Model	Aug. 1	Aug. 4	Aug. 6	Aug. 8	Aug. 9
Ignoring quality flag SGP	0.2503	0.2963	0.6080	0.2668	0.2309
Proposed LVCS	0.1081	0.1603	0.2901	0.1203	0.0985

We can reconstruct the 3D observations with the high fidelity (quality flag 0). Given the MCMC sample from fitting the model, the reconstruction based on the quality flag 0 data is automatic and does not require additional computational burden. Figure 3a shows the observed temperature values of mixed quality flags at vertical pressure level 827 hPa (pressure level 89). Figure 3b shows the corresponding quality flag values of the data based on the locations. Figure 3c shows the quality flag 0 predicted mean values of the temperature at vertical pressure level 827 hPa using the proposed recursive prediction procedure. The two images have a lot of similarities with more significant differences when the quality flag values are equal to 2 (Fig. 3b). Also the proposed procedure can take into account the parameter uncertainty as well as the nugget effects. Figure 3d shows the predictive variance of the predicted values. There is a clear association of the variance with the quality flag values and the spatial proximity as they are shown in Fig. 3b.

Within the boundaries of a granule, we can also use the proposed procedure to predict the air temperature values at an unmeasured vertical pressure level different from those 100 levels in AIRS data products, i.e., interpolating vertically and horizontally. Given the MCMC values obtained in fitting of the model, we can find the predictive distribution for any location and vertical pressure value. With 20,000 posterior samples of model parameters, the total computing time to make prediction for a vertical pressure level at the specified granule locations is about 145.5 seconds based on MATLAB 2020 on a laptop with specifications described above. For example, Fig. 4a shows the predicted temperature and Fig. 4b shows the associated standard deviation for unmeasured vertical pressure level value 265 hPa which corresponds to pressure level between 60 and 61.

Within the study region of the subtropical eastern Pacific Ocean, we repeat the proposed model and inference for five different granules on different days in August 2013. Two different models are considered: (a) the SGP which ignores the quality flag and (b) the proposed LVCS which accounts for the quality flag. Because the AAGP gives similar results to the SGP and it is computationally more expensive, we discard it from this analysis. This is also a strong evidence of the separable covariance structure between coordinates and atmospheric pressure for the AIRS observation. We follow the same computational strategy as explained above where the MCMC is run in a sequence for both models. Table 2 shows that the MSPE applying the LVCS is consistently almost half to the MSPE if we ignore the different flag quality. These results strengthen our modeling approach to account for different quality flags when we analyze AIRS temperature data sets.

6. CONCLUSIONS AND DISCUSSION

In this paper, we propose a latent variable co-kriging model with separable Gaussian processes (LVCS) to analyze 3D AIRS air temperature data with different quality flags from NASA's AIRS instrument. We propose an MCMC procedure with an imputation mechanism which takes advantage of the 3D input structure to facilitate efficient computation. By applying our method to the AIRS data, we demonstrate that incorporating the quality-flagged information into statistical modeling and data analysis can provide substantial inferential benefits. Unlike other methods for NASA's AIRS mission in the literature (e.g., Tian et al. 2020; Waliser et al. 2020), we allow for different fidelity and missing data. Our methodology provides a coherent framework for the combined use of remote sensing retrievals with variable quality in both the horizontal and vertical directions. This capability can bolster the utility of observations from AIRS and next-generation infrared sounder instruments, particularly in challenging observing conditions such as the subtropical ocean regions illustrated in this work.

Our LVCS model can be generalized for larger spatial data sets over multiple granules. For example, we may apply a computationally efficient method such as low-dimensional conditional approximation (e.g., Datta et al. 2016; Katzfuss and Guinness 2021) for both horizontal and vertical components in the separable covariance function. It is also possible to extend our method for multi-fidelity spatiotemporal remote sensing data by adding the temporal dependence and considering 4D data (longitude, latitude, vertical pressure level and time). Extending the LVCS modeling framework for multivariate multi-fidelity data could also prove to be fruitful.

The current LVCS model assumes a linear relationship between low and high fidelity levels. One direction in future research is to extend this work by borrowing the idea of deep Gaussian process in Perdikaris et al. (2017) and Ming et al. (2021) so that we can incorporate nonlinear dependence structure into the model. Future directions include extending the proposed method into a multivariate setting by using similar ideas to parallel partial autoregressive co-kriging Ma et al. (2022). In addition, the LVCS modeling framework can be applied in observing system uncertainty experiments (Hobbs et al. 2017; Turmon and Braverman 2019; Braverman et al. 2021; Ma et al. 2021), in particular when surrogates with different fidelity levels and computational efficiency are used to perform forward uncertainty propagation and inverse calibration for remote sensing data products.

SUPPLEMENTARY MATERIAL

Code for accessing the data used in this manuscript can be found in: https://github.com/L2UQ/airs_products. MATLAB code for the proposed method: The "JABES_LVCS" file contains the main code (written in MATLAB) to perform the proposed latent variable co-kriging methods described in the article.

ACKNOWLEDGEMENTS

The research of Konomi and Kang was partially supported by National Science Foundation Grant NSF DMS-2053668. Kang was also partially supported by Simons Foundation's Collaboration Award (#317298 and #712755), and the Taft Research Center at the University of Cincinnati. Part of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. Support was provided by the Atmospheric Infrared Sounder (AIRS) mission. The authors thank Amy Braverman and Hai Nguyen for valuable discussions during the course of this work.

[Received December 2021. Revised November 2022. Accepted January 2023.]

REFERENCES

- Banerjee S (2017) High-dimensional Bayesian geostatistics. *Bayesian Anal* 12:583–614. <https://doi.org/10.1214/17-BA1056R>
- Banerjee S, Carlin BP, Gelfand AE (2014) *Hierarchical modeling and analysis for spatial data*, 2nd edn. Chapman and Hall/CRC, New York
- Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial data sets. *J R Stat Soc Ser B (Stat Methodol)* 70:825–848
- Behrangi A, Fetzer EJ, Granger SL (2016) Early detection of drought onset using near surface temperature and humidity observed from space. *Int J Remote Sens* 37:3911–3923
- Bilionis I, Zabarar N, Konomi BA, Lin G (2013) Multi-output separable Gaussian process: towards an efficient, fully Bayesian paradigm for uncertainty quantification. *J Comput Phys* 241:212–239
- Braverman A, Hobbs J, Teixeira J, Gunson M (2021) Post hoc uncertainty quantification for remote sensing observing systems. *SIAM/ASA J Uncert Quant* 9:1064–1093. <https://doi.org/10.1137/19M1304283>
- Cheng S, Konomi BA, Matthews JL, Karagiannis G, Kang EL (2021) Hierarchical Bayesian nearest neighbor co-kriging gaussian process models; an application to intersatellite calibration. *Spat Stat* 44:100516
- Cohen WB, Maier-Sperger TK, Yang Z, Gower ST, Turner DP, Ritts WD, Berterretche M, Running SW (2003) Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in North America: a quality assessment of 2000/2001 provisional MODIS products. *Remote Sens Environ* 88:233–25. <https://www.sciencedirect.com/science/article/pii/S0034425703001810>
- Cressie N (1993) *Statistics for spatial data*. Wiley, New York
- Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. *J R Stat Soc Ser B (Stat Methodol)* 70:209–226
- Datta A, Banerjee S, Finley AO, Gelfand AE (2016) Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J Am Stat Assoc* 111:800–812
- Genton MG (2007) Separable approximations of space-time covariance matrices. *Environmetrics* 18:681–695
- Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc* 102:359–378
- Gramacy RB (2020) *surrogates: Gaussian process modeling, design and optimization for the applied sciences*. Chapman Hall/CRC, Boca Raton. <http://bobby.gramacy.com/surrogates/>
- Gramacy RB, Apley DW (2015) Local Gaussian process approximation for large computer experiments. *J Comput Gr Stat* 24:561–578
- Guillas S, Sarri A, Day SJ, Liu X, Dias F (2018) Functional emulation of high resolution tsunami modelling over Cascadia. *Ann Appl Stat* 12:2023–2053. <https://doi.org/10.1214/18-AOAS1142>
- Heaton MJ, Datta A, Finley A, Furrer R, Guhaniyogi R, Gerber F, Gramacy RB, Hammerling D, Katzfuss M, Lindgren F, Nychka D, Sun F, Zammit-Mangion A (2019) A case study competition among methods for analyzing large spatial data. *J Agric Biol Environ Stat* 24:398–425

- Hobbs J, Braverman A, Cressie N, Granat R, Gunson M (2017) Simulation-based uncertainty quantification for estimating atmospheric CO₂ from satellite data. *SIAM/ASA J Uncert Quant* 5:956–985
- Kahn BH, Irion FW, Dang VT, Manning EM, Nasiri SL, Naud CM, Blaisdell JM, Schreier MM, Yue Q, Bowman KW, Fetzner EJ, Hulley GC, Liou KN, Lubin D, Ou SC, Susskind J, Takano Y, Tian B, Worden JR (2014) The atmospheric infrared sounder version 6 cloud products. *Atmos Chem Phys* 14:399–426
- Kalmus P, Wong S, Teixeira J (2015) The pacific subtropical cloud transition: A MAGIC assessment of AIRS and ECMWF thermodynamic structure. *IEEE Geosci Remote Sens Lett* 12:1586–1590
- Katzfuss M (2017) A multi-resolution approximation for massive spatial datasets. *J Am Stat Assoc* 112:201–214
- Katzfuss M, Guinness J (2021) A general framework for Vecchia approximations of Gaussian processes. *Stat Sci* 36:124–141
- Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. *Biometrika* 87:1–13
- Konomi BA, Karagiannis G (2021) Bayesian analysis of multifidelity computer models with local features and nonnested experimental designs: application to the WRF model. *Technometrics* 63:510–522
- Konomi BA, Sang H, Mallick BK (2014) Adaptive Bayesian nonstationary modeling for large spatial datasets using covariance approximations. *J Comput Graph Stat* 23:802–829
- Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc Ser B (Stat Methodol)* 73:423–498
- Lu Y, Zhang F (2019) Toward ensemble assimilation of hyperspectral satellite observations with data compression and dimension reduction using principal component analysis. *Mon Weather Rev* 147:3505–3518
- Ma P, Kang EL (2020) A fused gaussian process model for very large spatial data. *J Comput Graph Stat* 29:479–489
- Ma P, Kang EL (2020) Spatio-Temporal data fusion for massive sea surface temperature data from MODIS and AMSR-E instruments. *Environmetrics* 31:e2594 (**E2594 env.2594**)
- Ma P, Karagiannis G, Konomi BA, Asher TG, Toro GR, Cox AT (2022) Multifidelity computer model emulation with high-dimensional output: an application to storm surge. *J R Stat Soc Ser C (Appl Stat)* 71:861–883
- Ma P, Konomi BA, Kang EL (2019) An additive approximate Gaussian process model for large spatio-temporal data. *Environmetrics* 30:e2569
- Ma P, Mondal A, Konomi BA, Hobbs J, Song JJ, Kang EL (2021) Computer model emulation with high-dimensional functional output in large-scale observing system uncertainty experiments. *Technometrics* 0:1–15
- Ming D, Williamson D, Guillas S (2021). Deep Gaussian process emulation using stochastic imputation
- Nychka D, Bandyopadhyay S, Hammerling D, Lindgren F, Sain S (2015) A multiresolution Gaussian process model for the analysis of large spatial datasets. *J Comput Graph Stat* 24:579–599
- Paciorek C, Schervish M (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* 17:483–506
- Perdikaris P, Raissi M, Damianou A, Lawrence N, Karniadakis GE (2017) Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proc R Soc A Math Phys Eng Sci* 473:20160751
- Peruzzi M, Banerjee S, Finley AO (2020) Highly scalable Bayesian geostatistical modeling via meshed gaussian processes on partitioned domains. *J Am Stat Assoc* 0:1–14
- Prata F, Lynch M (2019) Passive earth observations of volcanic clouds in the atmosphere. *Atmosphere* 10:199
- Qian PZ, Wu CJ (2008) Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics* 50:192–204
- Roberts GO, Rosenthal JS et al (2004) General state space Markov chains and MCMC algorithms. *Probab Surv* 1:20–71
- Rougier J (2008) Efficient emulators for multivariate deterministic functions. *J Comput Graph Stat* 17:827–843
- Sang H, Huang JZ (2012) A full scale approximation of covariance functions for large spatial data sets. *J R Stat Soc Ser B (Stat Methodol)* 74:111–132
- Stein ML (1999) Interpolation of spatial data: some theory for kriging, 2nd edn. Springer, New York
- Stein ML, Chi Z, Welty LJ (2004) Approximating likelihoods for large spatial data sets. *J R Stat Soc Ser B (Stat Methodol)* 66:275–296

- 606 Susskind J, Barnet CD, Blaisdell JM (2003) Retrieval of atmospheric and surface parameters from
607 AIRS/AMSU/HSB data in the presence of clouds. *IEEE Trans Geosci Remote Sens* 41:390–409
- 608 Thrastarson HT, Manning E, Kahn B, Fetzner EJ, Yue Q, Wong S, Kalmus P, Payne V, Olsen ET, Wilson RC, Blaisdell
609 J, Iredell L, Susskind J, Warner J, Cady-Pereira K (2021) AIRS/AMSU/HSB version 7 Level 2 product user
610 guide. https://docserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/V7_L2_Product_User_Guide.pdf
- 611 Tian B, Manning E, Roman J, Thrastarson H, Fetzner E, Monarrez R (2020) AIRS version 7 Level 3 product user
612 guide. https://docserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/V7_L3_Product_User_Guide.pdf
- 613 Turmon M, Braverman A (2019) Uncertainty quantification for JPL retrievals. Technical report, Pasadena. Jet
614 propulsion laboratory, national aeronautics and space administration, CA <http://hdl.handle.net/2014/45978>
- 615 Vecchia AV (1988) Estimation and model identification for continuous spatial processes. *J R Stat Soc Ser B*
616 (Methodol) 50:297–312
- 617 Waliser D, Gleckler PJ, Ferraro R, Taylor KE, Ames S, Biard J, Bosilovich MG, Brown O, Chepfer H, Cinquini
618 L, Durack PJ, Eyring V, Mathieu P-P, Lee T, Pinnock S, Potter GL, Rixen M, Saunders R, Schulz J, Thépaut
619 J-N, Tuma M (2020) Observations for model intercomparison project (Obs4MIPs): status for CMIP6. *Geosci*
620 *Model Dev* 13:2945–2958
- 621 Wunch D, Wennberg PO, Toon GC, Connor BJ, Fisher B, Osterman GB, Frankenberg C, Mandrake L, O'Dell C,
622 Ahonen P, Biraud SC, Castano R, Cressie N, Crisp D, Deutscher NM, Eldering A, Fisher ML, Griffith DWT,
623 Gunson M, Heikkinen P, Keppel-Aleks G, Kyrö E, Lindenmaier R, Macatangay R, Mendonca J, Messerschmidt
624 J, Miller CE, Morino I, Notholt J, Oyafuso FA, Rettinger M, Robinson J, Roehl CM, Salawitch RJ, Sherlock
625 V, Strong K, Sussmann R, Tanaka T, Thompson DR, Uchino O, Warneke T, Wofsy SC (2011) A method for
626 evaluating bias in global measurements of CO₂ total columns from space. *Atmos Chem Phys* 11:12317–12337
- 627 Zhou X, Kollias P, Lewis ER (2015) Clouds, precipitation, and marine boundary layer structure during the MAGIC
628 Field campaign. *J Clim* 28:2420–2442
- 629 Zhu Y, Kang EL, Bo Y, Tang Q, Cheng J, He Y (2015) A robust fixed rank kriging method for improving the spatial
630 completeness and accuracy of satellite SST products. *IEEE Trans Geosci Remote Sens* 53:5021–5035

631 **Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published
632 maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.