Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-156-RC1, 2019

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Efficient multi-scale Gaussian process regression for massive remote sensing data with satGP v0.1" *by* Jouni Susiluoto et al.

Anonymous Referee #1

Received and published: 24 September 2019

This manuscript describes a model to analyze large spatio-temporal data. Although analyzing remote sensing data of enormous sizes is no double important and challenge, the manuscript fails to describe the model and its computation details and properties sufficiently or clearly. Please see below my comments that are not necessarily ordered chronologically or by importance:

1. This manuscript suggest using the mean function of a particular form when analyzing OCO-2 data: $m(x;\beta,\delta) = \tilde{f}(x^t;\delta(x^s))'\beta(x^s)$ as given in Equation (3). This mean function is not a linear form of unknown parameters $\{\delta(x^s),\beta(x^s)\}$, noting that they are both dependent (i.e., varying) across locations. I find the description

Printer-friendly version

Discussion paper

on how to estimate $\delta(x^s)$ and $\beta(x^s)$ extremely confusing.

- In Lines 10-20 of Page 6, it states that β will be estimated using the formula of generalized least squared as given in Equation (6), and δ will be calibrated, but no explanation is given on how δ will be *calibrated*. In addition, the authors did not explain the dimension of the matrices F and K in Equation (6). Are they large so that K^{-1} or $(F^T K^{-1} F)^{-1}$ difficult to compute?
- How is $\beta(x^s)$ estimated for a location x^s? For a location $x^{s,test}$ *without* data/observation, can we estimate $\beta(x^{s,test})$ and how?
- Although the authors have included Section 2.4 on learning $\beta(x^s)$ as a Markov random field, this section is not connected to other parts of the manuscript but only adds confusion. It is unclear what the authors meant by modeling $\beta(x^s)$ as a Marko random field. Does this mean that the authors no longer use Equation (6) to estimate $\beta(x^s)$? What are the assumptions of this Markov random field (MRF)? What are the parameters in this MRK and how is this MRK fitted?
- It is also confusing how the parameters $\delta(x^s)$ are estimated.
- Line 14 of Page 8: " ... finding $\hat{\beta}$ with Eq. (9) and (10), ..." Is this a typo? Should it be Eq. (6) and (7)?
- Page 8 Line 15: The objective function $\sum_{j=1}^{n}(m(x_\nu;\beta_\nu,\delta_\nu) - \psi_j)^2 + \sum_{j' \in \partial\nu}(\delta_\nu - \delta_{j'})^2$ and the optimization procedure are poorly explained. It should be noted that the mean function $m(\cdot;\cdot,\cdot)$ involved δ and β. It is very confusing how or why this function is used to estimate δ or β individually or both of them jointly, and why it should be used this way.

2. The notations in this manuscript are very confusing overall. For example, the authors sometimes use $\beta(^s)$ and later use β_ν. The covariance parameters are even more confusing. There are l, l_c, and l_I. Even the definition of I is not consistent: It is originally stated $I \subset \{x^s, x^t\}$, but later used as $I = ST$ or I_S,

and I_{ST}. Also, the authors used Δ_{year} in Equation (11) and stated Δ_{year} is the duration of one year, does this mean $\Delta_{year} = 365$? Similarly, in Equation (15), the authors used Δ_{period}; is it 365 as well?

3. The authors suggest the multi-scale covariance function given in Equation (18): $k(x, x'; \theta) = \delta_{x,x'}\sigma_x^2 + k_{per}(x, x'; \theta, I_S) + k_M(x, x'; \theta) + k_{exp}(x, x'; \theta, I + ST) + k_W(x, x'; \theta, I)$.
   - First, I am not sure multi-scale is an accurate way to describe this covariance function. I feel this function is to add different *types* of covariance functions together, but these components not necessarily differ in terms of scales.
   - The authors did not explain clearly the component $k_W(x, x'; \theta, I)$. Although Equation (16) states it is equal to $k_{exp}(x_W, x'_W; \theta^W, ST)$, the authors fail to explain x_W or the quantifies in Equation (17) especially, l, l_t, $l^{||}$, and $l^{\perp}$, and how these parameters are chosen/estimated.
   - What will happen if there are missing data in wind velocity?
   - Why isn't there an I involved in the Matérn component $k_M(\cdot, \cdot; \cdot)$?
   - For the exponential component, the definition given in Equations (12) and (13) are not clear. At least there are two ways to define this component:

     $$k_{exp}(x, x'; \theta, I_{ST}) = \tau^2 exp(-|\frac{x - x'}{l_{ST}}|^{\gamma})$$

     or

     $$k_{exp}(x, x'; \theta, I_{ST}) = \tau^2 exp(-|\frac{x^s - x^{s\prime}}{l_S}|^{\gamma^s}) exp(-|\frac{x^t - x^{t\prime}}{l_T}|^{\gamma^t})$$

     dependent on whether the spatial and temporal components share the scale or exponent parameters. I don't know what the authors have used, and there is no justification of their choice.

- The authors need to provide a better description of these components in the covariance function and explain why they are identifiable based on their formulations and definitions. Also, it is necessary to clarify whether some parameters are the same or vary across these components, such as τ^2, γ, and l.

4. I find Sections 2.6 and 2.7 quite difficult to understand. It seems that the authors use local kriging, that is, using a subset of data close to a prediction location x^* to estimate the covariance parameters and to make prediction. Furthermore, it appears that the authors use different subsets of data to estimate the components in the covariance function. Why not using a single subset data to estimate the entire covariance function? Or, were the authors trying to avoid identifiability issue by using different data sets to estimate different covariance components? If a subset of data are used, I assume the size of this chosen subset is not too large, but why is there a need to use a block diagonal matrix $\tilde{K}_i$ as in Equation (22)? This approximation is not clearly explained, neither is E_{ref} in Equation (22). Moreover, in Equation (19), should it be $> \sigma^2_{min}$ rather than $< \sigma^2_{min}$?

5. The authors mentioned the nearest neighbor Gaussian process, but did not cite the reference correspondingly.

6. It is unclear where or why MCMC is needed and how it is implemented (prior specification etc.). The authors described optimization in Section 2 and also in the first paragraph of Page13. However, later in Page 17, the authors stated that MCMC is used instead. Section 2 does not describe MCMC.

7. It should be Matérn covariance function, instead of Matern.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-156, 2019.