Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-183-RC2, 2018

Geoscientific Model Development Discussions

Interactive comment on "A new method (M^{3}Fusion-v1) for combining observations and multiple model output for an improved estimate of the global surface ozone distribution" *by* Kai-Lan Chang et al.

Anonymous Referee #2

Received and published: 25 October 2018

General comments. Overall quality

The article proposes a method for combining measurements from 6 different global models with the aim of generating an improved estimation of the global surface ozone when compared to the estimation obtained by the simple average of these 6 different global models. Hence, this article proposes a method for estimating the weight factor to give to each global model within a weighted average of global models available, and also proposes a method for fusing this result with kriging estimates depending of closeness of locations to monitoring networks. The latter results in an estimated

global surface ozone which is a combination between interpolation-based kriging for areas near monitoring networks, and the proposed weighted average for areas far from monitoring networks. Notwithstanding, results from this article and the final surface for global ozone is estimated by a smoothing splines approach which is applied to the estimation either of the composite model or its fused version. Consequently, the smoothing splines step is the key for the method presented, however it is not explained in the article and authors only dedicate two to three lines to comment about its use to avoid discontinuities in the joints between continental regions.

The authors are trying to address three different problems in geostatistics. First, irregularly spaced sources of information or when the coordinates of the locations from different sources do not match. Second, the lack of information due to observations sparsely distributed or missing locations or almost no information in certain regions. And third, to obtain a better estimate of a surface or compare one estimate with others.

The first problem is more related with interpolation and this is explained well in sections 2.1 and 2.2 of the article.

The second problem is being addressed by the use of global ozone models to obtain a better guess of the non-observed ozone in certain locations. Here the authors propose the composite mean between the global models and its fused version with the interpolation depending on closeness of monitoring network stations which in practice is working as a method for "imputation" of ozone in non-observed locations. The description of the method is mostly well explained (although it is missing important details which I describe in the next section of this report) but the method is not a solution for this problem in areas like Africa or South America where there is not enough information and this is not solved by the composite nor fused method, but by having more measurements. This should be stated clearly in the article. It is difficult to believe that the weight coefficients estimated for Africa or South America would be good estimates given the little sample size available. Nevertheless, the authors could have taken advantage of the dense data available for North America, Europe and East Asia to perform cross-validation and then using the data from South America, Africa and Australia as validation sets of data. This would have provided a rough idea of the quality of the composite estimates to perform the imputation of the ozone level on areas with sparsely distributed observations.

The third problem is poorly or not explained in the article. As presented, the article gives the impression that the main modelling product is the composite mean or its fused version which is confusing since the results are based on a smoothed version and this is not explained in the article. One reader could think that in equation (1) the $\hat{y}(s_g)$'s are the "imputed" observations based on the interpolation technique while the composite mean is the proposed model for the ozone level. However, given that later in the article it is expressed that the results are obtained using smoothing splines over the fitted composite mean surface or its fused version, other readers can interpret that actually the fitted composite or fused mean are the imputed observations of the ozone level and the proposed model is the smoothing spline. This is extremely confusing and the article is poorly explained in all this part.

Specific comments. Individual scientific questions/issues

(a) There are important issues which are not addressed by the authors. What is the real role of the smoothing spline applied to the fussed estimation as described in page 8 lines 10-15 and the supplementary material Figure S-2? As the article is presented, this step seems to have a minor role for their proposed method, however it is a key step and the authors did not explain this in detail. In the abstract and along the presentation until section 2.2, the authors' product of this work is a method which relies on a fusion between a weighted average of the 6 global models and the interpolation/kriging step. Nevertheless, from the results it can be deduced that the final product of this work is actually the smoothing spline fit to the surface obtained either by the composite method or its fused version. Therefore, the results presented in Figure 8 (and therefore all related results) rather than being the surface obtained by the multi-model composite and multi-model composite plus bias correction, are respectively the smoothing spline

fit to the surface obtained by the multi-model composite and the smoothing spline fit to the surface obtained by the multi-model composite plus bias correction. This must be clearly stated.

(b) What is the interpretation of the parameters in every model discussed? For example:

-What is the practical interpretation of the parameter α_r in equation (1)? It is related to the general mean over the r-th continental region.

-What is the practical interpretation of the parameter β_{rk} in equation (1)? The cell-by-cell average model corresponds to assume $\beta_{rk} = 1/6$, thus the same weight is given to each model on each continental region. Then on the composite model β_{rk} can have a meaningful interpretation but the authors do not comment on this.

-What does it mean if $\beta_{rk} = 0$ (i.e. with respect to the cell-by-cell mean model)? Moreover, what does it mean if $\beta_{rk} \neq 0$? How to interpret if $\beta_{rk} < 0$? How to interpret if $\beta_{rk} > 0$? This is important and connects the cell-by-cell average with the composite mean model proposed. It tells us whether the composite model offers or not a better representation than the average model.

(c) Regarding comment (b), some statistical summaries are not presented in Table 2. For example, what is the significance of each weight/coefficient for the global models and their standard errors or confidence limits? Note that the proposed quadratic programming idea can be seen as a multiple linear regression within each continental region where the $\hat{y}(s_g)$'s are seen as the (imputed) observations, the global ozone models $\eta_{r1}(s_g), \ldots, \eta_{r6}(s_g)$ are seen as predictors or covariates, and the errors are assumed uncorrelated with constant variance. From this approach the authors can obtain variability estimates for the weight coefficients and test their significance.

(d) The authors comment that their composite and fused composite method is better than the simple average (or cell-by-cell average) method. It would be helpful if the authors presented p-values for a test comparing these two hypotheses.

(e) The authors do not mention the assumption for the mean nor covariance of the smoothing splines model, nor give any details about which type of splines they used (tensor products, thin-plate splines, regression splines, etc.). Did you use penalties? The authors only refer to mgcv R package (Wood, 2017) in line 20 of page 13 and we need to see the code to see what they did, however they should also explain their method, procedure and assumptions in the article. It is the most important modelling they are doing and their results depend on this smoothing splines step.

(f) We can see three different steps in this method. The initial interpolation using INLA, the determination of the weights, and the final smoothing splines using mgcv. INLA and the composite are imputing the ozone measurement on unobserved locations, and the gam function of mgcv package is performing the fit using smoothing splines. In practice, INLA and the smoothing splines are performing the same procedure: interpolation. The only difference is that INLA is based on a triangulation and finite element approach to find a solution. Besides, in both cases the authors are assuming a Matérn covariance function. Therefore, in practice they are fitting an interpolation model to the data (using INLA), and then fitting pretty much the same interpolation model (but using mgcv package) to the previous fit obtained by INLA. Thus, the INLA interpolation is "smoothing" the variation (Figure 3a, page 28), and then an additional smoothing using gam function is being performed (Figure 8a and 8b, page 33, and Figure S-4 in supplementary material). These two fits seem very similar and differences between them can be (visually) attributed mainly to the "variation" generated by the composite mean fit (Figure S-2).

-The first INLA smoothing imputes the ozone at unobserved locations but the resulting "smoothed" process has smaller variation than what we could expect from the original spatial process. Why did the authors not use resampling methods for the imputation step (either before applying the INLA interpolation and/or before applying the smoothing splines fit)? This would have allowed them to keep some spatial variability on the

imputed spatial process, and also evaluate assumptions regarding this spatial variability model.

-Examples of how to implement resampling methods can be found in Liang et al. (2013), and in a more practical presentation by Muñoz et al. (2010). Other approaches based on Expectation-Maximization (EM) algorithm are presented in Schneider (2001).

(g) Regarding the previous comment (f), the authors did not present results about the estimation of the Matérn semivariogram's parameters for the INLA and smoothing splines step. Given that they are performing a "pre-smoothing" of the variation using INLA, it would be expected that the Matérn semivariogram in the smoothing splines step would be modelling a significantly lower amount of spatial variation which might result in almost uncorrelated errors (except in areas where there are peaks or throughs in the process). Is that a good representation or assumption for the global ozone process?

(h) Regarding the modelling part, as presented the article overlooks the real role that the smoothing splines step (page 8 lines 8-15 and Figure S-2 in supplementary material) is playing in the resulting global ozone surface estimated. What is the main modelling technique of their method: the composite method with/without bias correction, or the smoothing splines in question? This is a key issue which cannot be disregarded.

(i) At moments, it is not clear whether the key goal of the article is to propose a novel method to estimate the weights to give to each global ozone model, or to propose an estimated global ozone surface (which is indeed being obtained based on the smoothing splines step). I suggest this is clarified.

(j) The authors mention that the success of their composite mean obtained via the quadratic programming approach depends on the existence of a global ozone model which reproduces correctly the correct curvature on the process (line 30 page 7 – line 5 page 8). This is not necessarily true since equation (1) is only selecting α_r, β_{rk} based on a least squares criterion and no regularization conditions on the solution are being specified, i.e. a curvature penalty. Besides, the curvature of a surface is defined throughout the two-dimensional space in question (the map), and requires the existence of first and second derivatives of the surface. None of these conditions are being established in the article, so that the weights of the composite mean are best only in terms of the "squared distance" between the (imputed) observations and the composite mean at the regular grid of locations being used, not throughout the continuous two-dimensional space.

Additional (tentative) references

Liang, Faming, Yichen Cheng, Qifan Song, Jincheol Park, and Ping Yang. "A resampling-based stochastic approximation method for analysis of large geostatistical data." Journal of the American Statistical Association 108, no. 501 (2013): 325-339.

Muñoz, Breda, Virginia M. Lesser, and Ruben A. Smith. "Applying Multiple Imputation with Geostatistical Models to Account for Item Nonresponse in Environmental Data." Journal of Modern Applied Statistical Methods 9, no. 1 (2010): 27.

Schneider, Tapio. "Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values." Journal of climate 14, no. 5 (2001): 853-871.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-183, 2018.