Comparison of cloud-filling algorithms for ocean color data

Andy Stock[a], Ajit Subramaniam[a], Kevin Arrigo[b], Fiorenza Micheli[c], Lisa Wedding[c], Mary Cameron[b] & Gert van Dijken[b]

a: Lamont-Doherty Earth Observatory, Columbia University, USA; b: Department of Earth System Science, Stanford University, USA; c: Center for Ocean Solutions, Stanford University, USA

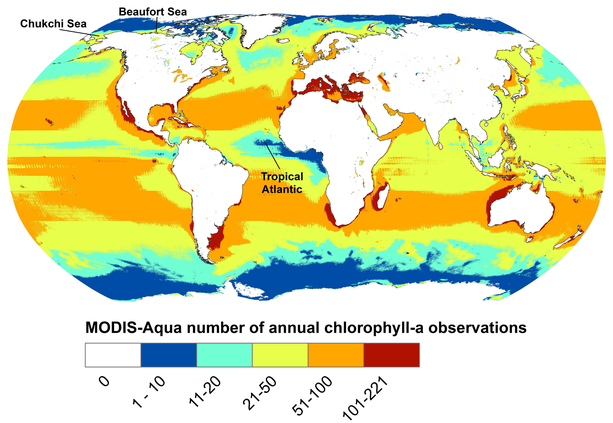

Fig. 1. Average number of valid MODIS-Aqua chlorophyll-a observations per year, 2003-2017.

Background. Ocean color remote sensing provides timely, temporally and spatially exhaustive descriptions of chlorophyll concentrations in surface waters, allowing for the monitoring and study of phytoplankton biomass and primary productivity at broad spatial scales. However, cloud cover is a major challenge in ocean color remote sensing. For example, MODIS-Aqua has a revisit time of 2 days, but nevertheless observes some locations less than 10 times per year (Fig. 1). Researchers have thus proposed various algorithms to fill data gaps “below the clouds”, but a comprehensive and quantitative comparison of the algorithms’ performance has not yet been conducted. Based on analyses in three study areas, this poster thus investigates how accurately the existing and various new, geostatistical and statistical-learning based cloud-filling algorithms predict chlorophyll-a concentrations “under the clouds”.

Methods. We used a 4-step approach (Fig. 2) to test how well various cloud-filling algorithms were able to reconstruct cloud-covered parts of MODIS-Aqua and SeaWiFS 8-day chlorophyll-a composites in three study areas. First, we identified images with less than 10% cloud cover in each study area. Second, we transplanted clouds from hundreds of other images onto the relatively cloud-free ones. Third, we used the cloud-filling algorithms to reconstruct the original images from the images with transplanted clouds. Fourth, we calculated two error measures: the expected root mean squared error (RMSE) of the reconstruction of individual pixels, and the expected difference between the regional mean chlorophyll concentrations in the original and the reconstructed images. The tested algorithms included data-interpolating empirical orthogonal functions (DINEOF; Alvera-Azcarate et al., 2005), self-organizing maps (Jouini et al., 2013), as well as general geostatistical and statistical learning methods, some of which had not been applied to cloud-filling before.

Results. The performance of the tested algorithms, in terms of errors in reconstructed pixels and spatial means, depended on the study area (Figs. 3 and 4). However, several algorithms worked consistently well: for example, random forests using features that summarize chlorophyll concentrations around a prediction location in different distance classes, ordinary Kriging, and DINEOF. Other algorithms also worked well, or even better, but only in specific study areas (e.g., temporal interpolation in the Beaufort Sea). Furthermore, cloud cover indeed led to major errors in regional means (Fig. 4). As expected, these errors increased as cloud cover increased. However, prior gap-filling by means of the best of the tested algorithms, depending on the study area, reduced errors in the calculation of spatial means by 40%-75%.

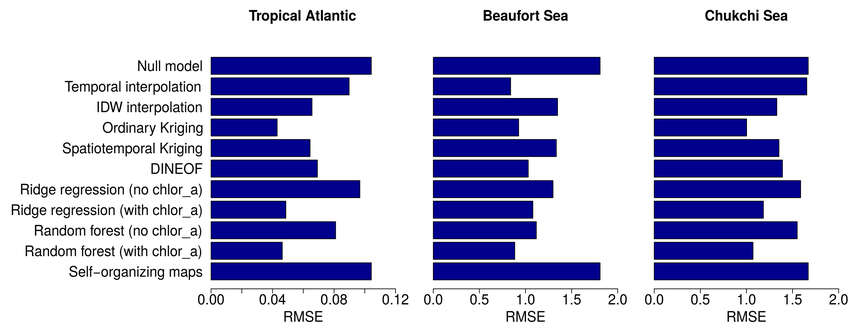

Fig. 3. Per-pixel RMSEs. The shorter the bar, the better the algorithm's prediction of chlorophyll concentrations in cloud-covered pixels. The null model simply fills data gaps with the mean of all visible pixels.

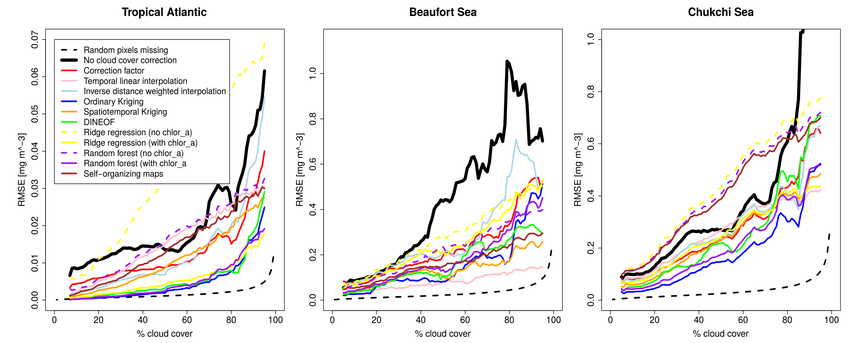

Fig. 4. RMSEs of the study regions' mean chlorophyll concentration as a function of cloud-cover, calculated without prior cloud-filling (solid black line) and using the different algorithms (colored lines).

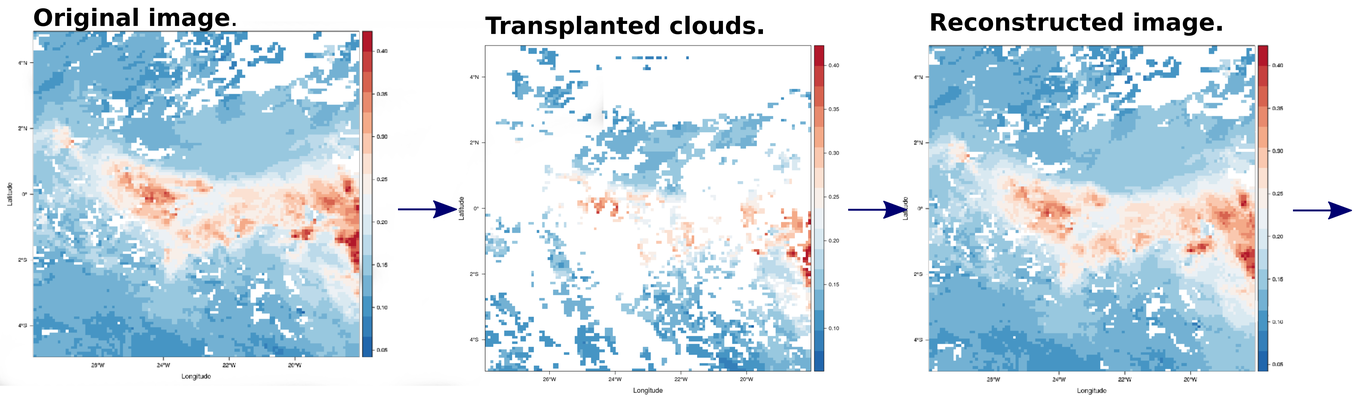

Fig. 2. We tested the algorithms by combining mostly cloud-free images with cloud masks transplanted from other images.

Summary & conclusions.

1. Which algorithms perform best depends on the study area. However, some algorithms such as ordinary Kriging and random forests mimicking interpolation worked consistently well.

2. The best algorithms focus on interpolation in space, time, or both.

3. Cloud cover can lead to major errors in the calculation of regional means of ocean color data products. These errors can, however, be substantially reduced by prior cloud-filling.

References

Alvera-Azcárate et al. (2005). Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. *Ocean Modelling*, 9(4), 325-346.

Jouini et al.(2013).Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method. *Remote Sensing of Environment*, 131, 232-246.

Acknowledgments

This work was supported by the Gulf of Mexico Research Initiative’s "Ecosystem Impacts of Oil and Gas Inputs to the Gulf" (ECOGIG) program. This is ECOGIG contribution #530. It was also supported by NASA OBB grant NNX16AAJ08G, and a grant from Stanford University's Catalyst for Collaborative Solutions Program.