Neural Network Technique for Gap-Filling in Satellite Observation Streams

Vladimir Krasnopolsky[1], Sudhir Nadiga[2], Eric Bayler[3], Avichal Mehra[1], David Behringer[1]
[1]NWS/NCEP/EMC, [2]IMSG at NWS/NCEP/EMC, [3]NOAA/NESDIS/STAR
Email: vladimir.krasnopolsky@noaa.gov; Phone: 301-683-3742

Operational integration/assimilation of satellite observations into operational models has three fundamental requirements/conditions: 1) gaps in the observations need to be addressed, both in the current instance and for extended gaps; 2) the data being assimilated must be for a predicted parameter; and 3) the data being assimilated must have a long data record to facilitate compilation of a robust statistical database spanning multiple seasons. Thus, integrating satellite data fields into operational models requires robust techniques to address potential gaps in the observations. In this study, we investigate one such robust gap-filling methodology — a Neural Network (NN) technique. As a test bed for this methodology we consider the use of NNs for integrating satellite ocean color fields (chl_a, Kd490, KPAR) into NOAA's operational ocean models (MOM and HYCOM). In this case NN links the ocean color variability, which is primarily driven by biological processes, with the local and remote upper ocean physical processes. We use satellite-derived surface variables — sea surface temperature (SST), sea surface height (SSH), sea surface salinity (SSS) fields, and upper layers of Argo profiles for temperature and salinity — as signatures of upper ocean dynamics in this study. This method of correlating satellite ocean color fields with other satellite observations that are currently being assimilated in the operational ocean model has the advantages of (a) being less likely to instigate assimilation errors due to dynamic imbalance, and (b) not relying on sparse *in situ* observations of ocean color.

Neural networks are very generic, accurate, and convenient mathematical models that are able to emulate complicated nonlinear input/output relationships through statistical learning algorithms [1]. Neural networks can approximate the transfer functions between a large number of possibly-interconnected inputs and multiple outputs, even when the relationships between the outputs and inputs are nonlinear and not well known. Neural networks employ adaptive weights that are tuned through training with past data sets, providing robustness with respect to random noise and fault-tolerance. Multilayer perceptrons (MLP) NNs are a generic tool for approximating such mappings. They use a family of functions like:

$$y_q = a_{q0} + \sum_{j=1}^{k} a_{qj} \cdot \phi(b_{j0} + \sum_{i=1}^{n} b_{ji} \cdot x_i); \quad q = 1, 2, \ldots, m$$

where x_i and y_q are components of the input and output vectors X and Y, respectively, a and b are fitting parameters. The activation function ϕ is usually a hyperbolic tangent, n and m are the numbers of inputs and outputs, respectively, and k is the number of neurons in the hidden layer. While NN training is a complicated and a time-consuming nonlinear optimization task, NN training needs to be done only once for a particular application. The trained NN is repeatedly applied to new data, providing accurate and fast emulations.

The satellite observations used in this study are well studied and available (or will be available soon) in near-real-time. The ocean color fields from the Visible Imaging Infrared Radiometer Suite (VIIRS) mission [2] and the SSS fields from the Aquarius mission were obtained from NASA, while the SSH and SST fields are from NOAA. The period covered in this study is 2012-2014. It is expected that the correlations between the ocean color data and current fields of SSH/SST/SSS will be dependent on location and season. Thus, lat, lon, and day of the year is included as additional inputs for NN. The NN technique is trained for the first two years and tested on the remaining year. However, by rotating the time series, we can test for each of the three years of the data. Also, to test robustness, the input stream is varied by withholding various inputs (e.g, SSH or SSS). Finally, the root-mean-square difference (RMSD) between the observed ocean color fields and the NN output is computed and plotted and the results are analyzed.

Preliminary results of application of this methodology are shown in Figure 1. Here, daily VIIRS chl-a data (~ 20,000,000 data points) and MOM simulated fields for January 2012 through November 2013 were selected for NN training and validation. The data were split into training and validation sets by selecting every second point for training, with the other alternating points reserved for validation. NNs, with 23 inputs (satellite SSH, SST, SSS, upper levels of gridded monthly ARGO temperature and salinity, plus year, time of year (sine, cosine), longitude (sine, cosine), latitude (sine)) and one output (chlorophyll-a), but with different numbers of hidden neurons (HID = 3, 5, 10, 15, 20, 25, 30, 35, 40) were trained. The trained NN was then applied to 2014 input data for final NN validation. Figure 1 shows results for three days, December 30, 2013 (one-month projection), May 11, 2014 (six-month projection), and September 23, 2014 (9-month projection), show that the NN is able to skillfully capture the large-scale patterns of Chl-a observed in the VIIRS data, but the performance deteriorates as the projection period increases. In operational applications, this deterioration can be mitigated by using on-line adjustment of NN parameters as soon as new data become available.

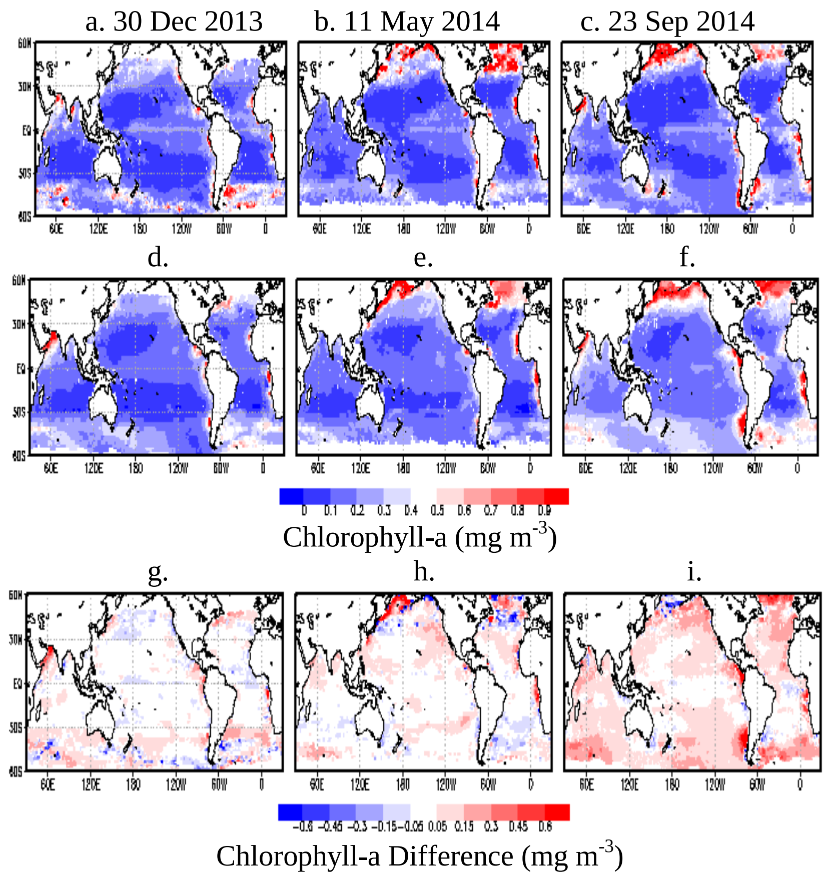

Figure 1. VIIRS chlorophyll-a (mg m^{-3}): a) 30 Dec 2013; b) 11 May 2014; and c) 23 Sep 2014; (d-f) are the same as (a-c) but are NN-predicted Chl-a; (g-i) are the difference (NN minus VIIRS).

These preliminary results demonstrate significant benefits of introducing NN approach for gap-filling in satellite observations for operational models.

[1] Krasnopolsky V., 2013, "The Application of Neural Networks in the Earth System Sciences. Neural Network Emulations for Complex Multidimensional Mappings", Springer, 200 pp.

[2] Nadiga, S., Bayler, E., Behringer, D., and Mehra, A., Towards Using Near-real-time VIIRS Ocean Color Data for NOAA's Operational Seasonal-Interannual Forecasting, 12th Annual JCSDA workshop, College Park, MD, May 21, 2014.