Using Neural Network Technique to Enhance Assimilating Surface Satellite Observation into Numerical Models.

Vladimir Krasnopolsky[1], Carlos J. Lozano[1], Hae-Cheol Kim[2], Bin Li[2]
[1]NWS/NCEP/EMC, [2]IMSG at NWS/NCEP/EMC
Email: vladimir.Krasnopolsky@noaa.gov; Phone: 301-683-3742

A generic approach that allows extracting functional nonlinear dependencies and mappings between atmospheric or ocean state variables in a relatively simple form is presented. These dependencies and mappings between the 2- and 3-D fields of the prognostic and diagnostic variables are implicitly described by the highly nonlinear coupled partial differential equations of an atmospheric or ocean dynamical model. They also are implicitly contained in the numerical model output. For example, when 2-D observations like surface wind, surface currents, or sea surface elevation are assimilated in the atmospheric or oceanic data assimilation system (DAS), the impact of this information in the atmosphere state or ocean state is conveyed by the observation operator and model and observations error covariances. Several attempts have been made to extract simplified linear dependencies of such a kind from observed data [1] or model simulations [2] for the use in an ocean DAS. However, these simplified and generalized linear dependencies that are often derived from local data sets do not properly represent the complicated nonlinear relationship between the model variables. If we were able to extract/emulate these dependencies in a simple, but not overly simplified, and yet adequately nonlinear analytical form, they could be used in the DAS to facilitate a more effective 3-D assimilation of the 2-D surface data. These analytical functions and mappings could also be used for efficient model output compression, archiving, and dissemination, and for sensitivity studies and error analysis. It is only recently that steps are being taken to use the NN technique to accomplish this objective. Here we introduce a generic NN technique using a particular application to the NN emulation for sea surface height. This new and generic NN application requires a reasonable quality of the Jacobian of the NN emulation. The Jacobian analysis and an ensemble approach to improve the quality of the NN emulation and NN Jacobian are presented in [3].

Sea surface height (SSH), η, is one of the prognostic variables in ocean circulation models. The particular ocean model that was used in this study is the Real Time Ocean Forecast System (RTOFS-Global), NCEP's "ocean weather" forecasting system. Its ocean modeling component is eddy-resolving HYbrid Coordinate Ocean Model (HYCOM). This model is a primitive equation model that uses a generalized hybrid coordinate (isopycnal/terrain following (σ)/z-level) in the vertical. The hybrid coordinate extends the geographic range of applicability of traditional isopycnal coordinate (coordinates that follow the selected levels of constant water density), toward shallow coastal seas and unstratified parts of the ocean. The particular version of HYCOM used in this study covers the global domain with an average horizontal resolution of 1/12°; and 32 vertical levels, and the potential density is referenced to 20 MPa with thermobaric correction [4].

An approach based on using neural network techniques is developed here to extract the inherent nonlinear relationship between the sea surface height anomaly and the other physically related variables of an ocean model. Specifically, numerically generated grid point fields from the global HYCOM model of NCEP (National Centers for Environmental Prediction) are used for training and validating this relationship. Accurate determination of such relationships is an important first step to enhance the assimilation of the sea surface height measurements into an ocean model by propagating the signal to other dependent variables through the depth/height of the model. Since the reduced physics model has mainly a 1-D vertical structure, we assumed, in this initial attempt, that SSH, or η, at a particular model grid point (i.e., at a particular horizontal location lat/lon) depends only on the vector of state variables X at the same horizontal location. Therefore, this dependence (a target mapping) can be written as,

$$\eta = \varphi(X) \quad (1)$$

where φ is a nonlinear continuous function and X is a vector that represents a complete set of state variables, which determines SSH. In this particular case the vector X was selected as,

$$X = \{sin(\tau), \cos(\tau), \sin(lon), \cos(lon), \sin(lat), uavg, vavg, p_{bot}, temp, dp, th3d, u, v\}$$

where $\tau = 2\pi t/365$, t is the day of the year, p_{bot} is bottom pressure, dp is a profile of interfaces (vertical coordinates used in HYCOM), *temp* is a profile of temperature, *th3d* - a profile of potential density, and u and v are profiles of internal x- and y-velocities, and uavg and vavg are vertically x- and y- averaged velocities. This set of variables represents (or is used as proxy for) the physics of the deep ocean. Therefore, the mapping (1) with this particular selection of components for the vector X will not be applicable in coastal areas (depth less than ~450 m). For the coastal areas a different set of state variables should be selected. All statistics presented below in this section were calculated using the test set where coastal areas were excluded. The NN technique is applied to derive an analytical NN emulation for the relationship between model state variables, X, and SSH, or η,

$$\eta_{NN} = \varphi_{NN}(X) \quad (2)$$

using the simulated model fields which are treated as error free data. Table 1 shows statistics calculated on validation data set for SSH (simulated data), SSH_{NN} (calculated using eq. (2)), and for the difference between them.

Table 1. Validation/Interpolation Accuracy for SSH(in m). Here STD is for standard deviation of SSH, SSH_{NN}, or RMSE for their difference (last row in the table).

	Min	Max	Mean	STD/RMSE	CC
SSH	-1.848e-1	1.859e-1	3.143e-3	4.757e-2	-
SSN_{NN}	-1.021e-1	1.595e-1	1.188e-3	4.684e-2	-
$SSH – SSH_{NN}$	-1.517e-1	1.394e-1	-5.512e-4	5.563e-3	0.993

The Jacobian of (2) has been estimated. The derivative ∂SSH/∂dp dominates the Jacobian, it is about two orders of magnitude larger than other derivatives, constituting Jacobian. *The accuracy of Jacobian (derivatives) in terms of SSH obtained using this Jacobian is of order of several cm.* As the next step, this Jacobian will be introduced into ocean DAS to propagate SSH to the thickness of vertical layers *dp*, using equation,

$$SSH = \sum_{k=1}^{IN} \left(\frac{\partial SSH}{\partial dp}\right)_k \cdot \Delta dp_k$$

where IN is the number of vertical layers in the model.

[1] S. Guinehut, P.Y. Le Traon, G. Larnicol, S. Philipps, “Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields – a first approach based on simulated data”, J Marine Systems, 46, pp. 85-98, 2004

[2] Mike Cooper, Keith Haines, “Altimetry assimilation with water property conservation”, J Geophys. Res., 101, pp. 1059-1077, 1996

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

[4] E.P. Chassignet, L. T. Smith, G. R. Halliwell, R. Bleck, “North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity”. J. Phys. Oceanogr., 33, 2504–2526.