From the Academy

Data assimilation and its applications

Bin Wang*[†], Xiaolei Zou[‡], and Jiang Zhu*

*Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; and [‡]Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520

In data assimilation, one prepares the grid data as the best possible estimate of the true initial state of a considered system by merging various measurements irregularly distributed in space and time, with a prior knowledge of the state given by a numerical model. Because it may improve forecasting or modeling and increase physical understanding of considered systems, data assimilation now plays a very important role in studies of atmospheric and oceanic problems. Here, three examples are presented to illustrate the use of new types of observations and the ability of improving forecasting or modeling.

A geophysical fluid system, whose past, current, and future behaviors are of great interest, is governed by a set of partial differential equations. The governing equations are solved numerically to obtain future states starting from an initial condition describing the current state of the system. Before the solution, however, this initial condition must be provided, which, along with the model equations, controls the evolution of the solution trajectory in space and time. How to prepare initial conditions with high quality and accuracy is receiving an increasing attention in the field of geophysical fluid dynamics.

The atmosphere and the ocean are two typical fluids on the Earth. Their behaviors directly affect the life and activities of human beings. To efficiently predict their future behaviors is very important. Since the advent of more powerful computers, higher-resolution atmospheric models and oceanic models have been developed. These models have shown a remarkable capability to predict some key phenomena and to simulate some important characteristics of the atmosphere or ocean. All of the models, however, require a complete and accurate specification of the three-dimensional (3D) structure of the initial state of a considered system. Besides conventional data (radiosonde, surface, and dropsonde data, mainly), many new sources of data, such as satellite data, radar, profilers, and other remote-sensing devices, have become available. However, observations are still sparse, and it is still impossible to measure all of the model's degrees of freedom at a given time. In addition, the observations are irregularly distributed in space and time, and they have different structures of random error. Therefore, an efficient data assimilation method is needed to combine these irregular observations to generate the initial conditions that are distributed on regular model grids.

The development of data assimilation methodology has mainly experienced three stages: simple analysis, statistical or optimum interpolation, and variational analysis. Simple analysis methods were mostly used in 1950s, when computers were unavailable or at the beginning stage. Simple analysis methods were the earliest bases of data assimilation. In the 1960s and 1970s, statistical considerations were introduced into the atmospheric data assimilation. Based on these considerations, some forms of optimum interpolation were used to assimilate observations into forecast models. These optimum interpolation analysis methods were used in many operational centers worldwide. In the 1980s and 1990s, atmospheric data assimilation switched to variational methods, in particular the three- and four-dimensional variational data assimilation (3D-Var/4D-Var) by using adjoint techniques. The 3D-Var/4D-Var approaches attempt to combine observations and background information in an optimal way to produce the best possible estimate of the model initial state. This technique not only has broad applications for the assimilation of atmosphere and ocean, but also can be used for many other applications in numerical weather prediction.

Data assimilation plays a more and more important role in numerical weather prediction, and it is considered as a frontier branch of atmospheric and oceanic sciences. In this paper, three examples are presented to illustrate (*i*) the use of new types of observations, and (*ii*) the ability to improve the forecast skill of numerical weather prediction.

Global Positioning System (GPS)/Meteorology (MET) Data Assimilation

With the advent of the GPS by using high performance transmitters in high orbits and low earth orbiting satellites equipped with GPS receivers, it is now possible to remotely sound the Earth's atmosphere by using radio occultation techniques (1–2). A prototype demonstration of this capability has been provided by the GPS/MET experiment. Although it was shown that high vertical resolution profiles of atmospheric refractivity, temperature, and geopotential height at constant pressure levels can be derived from the GPS measurements, with high accuracy under many circumstances, many issues remain. These issues include the existence of multipath propagation, the ambiguity between water vapor and temperature in moist regions of the atmosphere, and the difficulty in retrieving an accurate refractivity profile from the GPS refraction angle measurements over regions where the horizontal gradient of the refractivity is large. For this reason, a methodology for incorporating the GPS "raw" measurements (refraction angles) directly into numerical weather analysis and/or prediction systems is required and has been developed (3). It includes a ray-tracing observation operator that converts the atmospheric state variables to the GPS refraction angle measurements, and its tangent linear and adjoint operators. These three operators are required for the direct use of GPS refraction angle measurements in a variational data analysis system.

A twin least-square-fit experiment incorporating simulated GPS refraction angles is then conducted. The cost function converges to a known minimum, which shows that the develop-

FROM THE ACADEMY

This paper is a summary of a session presented at the second annual Chinese–American Frontiers of Science symposium, held August 20–22, 1999, at the New Century Hotel, Beijing, People's Republic of China.

Abbreviations: 3D, three-dimensional; GPS, global positioning system; MET, meteorology; NCEP, National Centers for Environmental Prediction; BDA, bogus data assimilation.

[†]To whom reprint requests should be addressed at: Building 932, Zhong Guan Cun, Apartment 512, Beijing 100086, China. E-mail: wab@lasgsgi4.iap.ac.cn.

ment of the tangent linear and adjoint versions of the ray-tracing observation operator and the linkage to the National Centers for Environmental Prediction (NCEP) global analysis are carried out correctly in this idealized case. Also, an analysis of all of the real GPS refraction angle measurements available from the GPS/MET experiment over a selected 12-h period is carried out. The changes made to the NCEP temperature, specific humidity, and pressure fields as a result of the minimization of two cost functions measuring the distance between the model-simulated and the observed GPS/MET refraction angles in two 6-h time windows suggest (*i*) a positive adjustment in the NCEP temperature field in the upper atmosphere between 15 km and 30 km; (*ii*) a negative adjustment of the moisture field in the lower troposphere below 500 hPa; and (*iii*) a negative adjustment of pressure field throughout the entire atmosphere up to about 20 km. This study also indicated that use of GPS refractivity might produce errors in the lower troposphere, where the refractivity gradient is large.

Recently, the GPS/MET data were included in a version of the Spectral Statistical Interpolation (SSI) analysis system at the NCEP in the United States by using the GPS ray-tracing operator and its adjoint operator (4). This GPS data assimilation system is tested by incorporating 30 actual GPS/MET observations of refraction angle obtained during the GPS/MET experiment. Insight into the impacts of the GPS/MET observations on the analyses is obtained by comparing the analyses with several combinations of GPS/MET observations and conventional (mostly radiosonde) data. A use of mixed GPS/MET refraction angles and derived refractivity, where the mix is based on the height and magnitude of the difference between GPS refractivity and atmospheric refractivity, produces a result similar that obtained from the use of refraction angles, along with a significant saving in computational cost. The results from all of the experiments suggest that the continuous assimilation of many (hundreds to thousands) GPS/MET refraction angle observations globally will soon become possible and that the impact of GPS data on global analysis and forecast could be assessed through an extended period of assimilation cycles.

Hurricane Initialization by Data Assimilation

Accurate prediction of hurricane track and intensity change accurately is still a challenging problem. Because of the lack of data over tropical regions, one of the major difficulties in numerical prediction of hurricanes is hurricane initialization. Initial vortices provided by large-scale analysis from operational centers are often ill-defined, too weak, and sometimes misplaced. This may be one of the main reasons causing the big average position errors for NCEP official hurricane track forecasts for seven Atlantic hurricanes (5), which were about 160 km in 24 h and 250 km in 48 h. Therefore, it is necessary to find an initialization procedure to augment a more realistic initial vortex.

One usual method of initialization is to implant a bogus vortex, a vortex specified based on the size of the hurricane, the position, and the intensity, into the model initial state. Many successful predictions of hurricane movement and structure have been made by using such a scheme. However, detailed procedures of these bogus methods vary from one form to another. Often, the nonlinear balance equation, gradient wind relation, geostrophic relation, hydrostatic relation, and so on are used to derive one variable from another in these early bogus schemes. How to generate all model fields of the initial vortex more objectively with dynamic and physical consistency is still an unresolved problem.

Recently, a bogus data assimilation (BDA) scheme was developed to generate the initial structure of a tropical cyclone for hurricane prediction (6). It was tested on Hurricane Felix (1995) in the Atlantic Ocean during its mature stage. The results show that the hurricane prediction of Felix was improved greatly. The dramatic improvements are shown in the track forecast and intensity forecast. The 24-h, 48-h, and 72-h forecast track errors with BDA were 76 km, 76 km, and 84 km, respectively, compared with the track errors of 93 km, 170 km, and 193 km without BDA, while the mean error of the central sea-level pressure during the entire 72-h forecast period reduced from 25.9 hPa without BDA to less than 2.1 hPa with BDA. Also, the initial fields of model variables describing the BDA initial vortex are well adapted to the forecast model, and the structure of the storm is captured by the model reasonably well. Note that the Penn State/National Center for Atmospheric Research (NCAR) nonhydrostatic Mesoscale Model Version 5 (MM5) was used for both the data assimilation and prediction in this study.

Estimation of Ocean Mixed Layer by Data Assimilation

Ocean mixed layer is a well-mixed zone of 50–100 m thick immediately below the sea surface where temperature and salinity are generally uniform. Among many facets of the ocean's role in the climate system, the influence of the sea surface temperature is most essential, and an accurate ocean mixed layer estimation is required. It is also indispensable to determine the sea surface temperature. Determination and forecast of the ocean mixed layer are also important for fisheries and underwater telecommunication. Traditionally there are two ways that determine the ocean mixed layer: collect observations and carry out numerical prediction by using a forecasting model. Increasingly, it is being recognized that an alternative way called oceanic data assimilation could provide us a better view of the ocean mixed layer. The basic concept of data assimilation is to combine real observations with numerical models. However, observations about the ocean mixed layer usually are sea surface temperature images from space. To solve the problem, some reduced order methods could be used (7). Because of the strong nonlinearities and turbulent nature in the ocean mixed layer models, the traditional data assimilation methods, like Kalman Filtering and 4D-VAR, faced some technical difficulties that prevent them from practical applications. Strong nonlinearities and threshold processes made the variational data assimilation problems nonsmooth. Some non-smooth methods are discussed. Other potential problems with data assimilation in ocean mixed layer include numerical instabilities in tangent linear models (8).

B.W. was supported in part by the National Natural Science Foundation of China (Grants 49825109 and 49823002), the National Key Development Planning Project for Basic Research (Grant G1999032801), and the Institute of Atmospheric Physics/Chinese Academy of Sciences Key Innovation Project (Grant 8-1205). J.Z. was supported in part by National Natural Science Foundation of China (Grant 49776283) and the ninth five-year key project of the Chinese Academy of Sciences (Contract K2952-51-434).

1. Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P. & Hardy, K. R. (1997) *J. Geophys. Res.* **102,** 23429–23465.
2. Rocken, C., Anthes, R., Exner, M., Hunt, D., Sokolovskiy, S., Ware, R., Gorbunov, M., Schreiner, W., Feng, D., Herman, B., *et al*. (1997) *J. Geophys. Res.* **102,** 29849–29866.
3. Zou, X., Vandenberghe, F., Wang, B., Gorbunov, M. E., Kuo, Y.-H., Sokolovskiy, S., Chang, J. C., Sela, J. G. & Anthes, R. A. (1999) *J. Geophys. Res.* **104,** 22301–22318.
4. Zou, X., Wang, B., Liu, H., Anthes, R. A., Matsumura T. & Zhu, Y.-J. (2000) *Q. J. R. Meteorol. Soc.*, in press.
5. Kurihara, Y., Bender, M. A. & Ross, R. J. (1993) *Mon. Weather Rev.* **121,** 2030–2045.
6. Zou, X. & Xiao, Q. (1999) *J. Atmos. Sci.* **57,** 836–860.
7. Zhu, J. & Kamachi, M. (2000) *Tellus* **52A,** 265–279.
8. Zhu, J. & Kamachi, M. (2000) *Mon. Weather Rev.* **128,** 1562–1572.