Gfdnavi is a web-based data and knowledge server program for geophysical fluid data that constructs databases, provides analysis and visualization tools, and shares knowledge documents. A new Gfdnavi user interface for analyzing and visualizing data on web browsers is developed to improve the user experience by providing seamless analysis and visualization operations, multiple diagram editing, a layer function, and so on. An experimental data handling system for ensemble numerical weather prediction data is constructed using Gfdnavi to address such issues as data processing and transfer between weather centers and decision makers in various sectors, including that of disaster management. Special tools to analyze and visualize ensemble numerical weather prediction data are implemented as user-defined Gfdnavi plug-ins. An interactive document that provides basic ideas of how to utilize probabilistic ensemble data information is written with the Gfdnavi knowledge documentation system, in which hyperlinks enable users to edit diagrams in the document.

Keywords: analysis, visualization, database, web application, ensemble numerical weather predictions

1. Introduction

In the last century, data types and data sizes in meteorology and related fields have been increasing. The output of numerical models, reanalysis data, and satellite data, for example, are commonly used in recent years. Due to a rapid increase in computational resources and advances in satellite observations, the spatiotemporal resolution of these datasets is increasing, as is the types of physical quantities. Data handling is thus getting harder and downloading whole datasets to local computers is almost impossible. Data formats vary from data to data, which requires extra cost to learn how to handle them. In the field of disaster management, huge amounts of data of various types are needed. This makes the problem worse.

Ensemble numerical weather prediction (NWP), which has been conducted operationally for nearly two decades [1], has enabled us to quantify the uncertainty of prediction. This kind of probabilistic information is thought to be useful for both appropriate disaster management and efficient economic activities. The problem of data size becomes more serious, however, because an ensemble prediction system performs a lot of simulation (10-100 runs) at the same time.

Another problem in ensemble prediction is that the handling of ensemble data is complicated compared with classical deterministic forecasts. For data transfer from weather centers to decision makers in various sectors, including that for disaster management, ensemble prediction data must be processed into appropriate formats that are easily understood by people who are not NWP specialists. In recent years, this issue is being addressed by the World Meteorological Organization under the Observing System Research and Predictability Experiment (THORPEX) [2].

In this paper, the problem in handling large data is addressed by using Gfdnavi (Geophysical fluid data navigator1), which is a web-based data and knowledge server program for geophysical fluid data [3, 4]. A brief introduction of Gfdnavi and details of a new user interface of Gfdnavi for the analysis and visualization functions will be described in section 2.

To solve the problem of processing ensemble data, we develop tools for easy handling of ensemble prediction data. These tools are implemented as user-defined plug-ins of Gfdnavi. A user’s guide for the tool has also been prepared using the knowledge documentation system of Gfdnavi. This will be described in section 3.

2. Gfdnavi

2.1. Overview

Gfdnavi can be operated by large data centers, small groups, and personal users. End users who want to ana-
An Experimental Data Handling System for Ensemble Numerical Weather Predictions Using a Web-Based Data Server and Analysis Tool "Gfdnavi"

Lyze data need to access Gfdnavi servers. A typical flow of operation by end users is as follows:

1. Open a web browser and access a server that operates Gfdnavi.
2. Find datasets in Finder or Search windows. When datasets are selected, an Analysis window opens.
3. Analyze and visualize data in the Analysis window. Users obtain image files as a result of visualization.
4. Write a knowledge document for results in Knowledge window.
5. Write a Ruby script (advanced users) to automate the (2)-(4) process.

On Gfdnavi servers, Gfdnavi scans selected directories to make a database of metadata in data files of binary formats that are commonly used in the field of meteorology, such as NetCDF, GRIB, and GrADS. The database is used for a web-based viewer. The user interface over web browsers is intuitive and a web service programming interface is also provided for advanced users.

Gfdnavi provides server-side data analysis and visualization tools. As long as one uses server-side tools, including user-defined plug-ins, one does not need to download original data, and only results, i.e., image files, are transferred from the server to local machines. For those who need advanced analysis and visualization on local machines, downloading a minimum set of analyzed (processed) data is available.

Gfdnavi further provides a knowledge documentation system that stores documents into the same database. Users make documents for diagrams that are created with Gfdnavi. Hyperlinks that enable users to reproduce the same diagrams on Gfdnavi are automatically created in documents. The database of knowledge documents enables users to share knowledge. Gfdnavi is implemented using the Ruby interface of GFD-Dennou library and Ruby on Rails. See details in Horinouchi et al. (2010) [3] and Nishizawa et al. (2010) [4].

In the next subsection, the new features introduced to Gfdnavi version 2.2 will be described.

2.2. Analysis and Visualization User Interface

In the analysis of meteorological data, it is necessary to compare multiple quantities and multiple cross-sections at the same time. In the new user interface of Gfdnavi, parameters for visualization, which were shared by all of the diagrams in the previous version, can be set for each diagram to achieve this. Functions are implemented by object-oriented programming in which parameters for a diagram are defined as attributes of the object for the diagram, relating the user interface and the internal expression directly.

Figure 1 shows the user interface for analysis and visualization. When this window is opened for the first time after the selection of data, a diagram automatically appears with default parameters for visualization. Users need to change parameters to get what they want to draw. The four buttons atop each diagram – Input, Method & Axis, Options, Action – open tab windows to tune analysis and visualization parameters.

Users edit multiple diagrams at the same time (Fig. 1), play animation, and overlay diagrams. The green panel on the left shows selected variables, axes of data, and so on. The button at the top left corner is used to create a knowledge document with currently editing diagrams.

Figure 2 shows the Input tab window used to select input data of a diagram in which multiple variables are used as input to an analysis method, i.e., a subroutine used to apply mathematical operations. A zonal mean of $\sqrt{u^2 + v^2}$ is obtained in the case of Fig. 2. Data flow in analysis is displayed in this window. Users edit variables and types of operations directly in window, which enables users to operate intuitively. Users perform seamless anal-
Fig. 3. Method & Axis tab window for a diagram with two layers with the tab user interface for Layer 0 selected.

Fig. 4. Options tab window for the left diagram in Fig. 1.

Fig. 5. Action tab window for the left diagram in Fig. 1.

ysis and visualization operations through this user interface.

The Method & Axis tab window provides the user interface for selecting the Draw method, subroutines used to draw various types of diagrams. In the case of Fig. 3, tone_contour is selected to draw color tones with contour lines. The Method & Axis tab window also provides the user interface for specifying areas to be drawn for two-dimensional diagrams, variable range for one-dimensional diagrams, and slice points of data. Axes that are used for the abscissa and ordinate of the diagram are also selected in this window.

The Options tab window provides parameter settings for visualization (Fig. 4). The General options category contains parameters, such as diagram size and map projections, whereas the Specific options category contains options for each draw method, i.e., line types of contours, tone colors, and so on.

Figure 5 shows the Action tab window for the left figure in Fig. 1. When editing of a diagram is completed, users take various actions for the diagram. Users download Ruby scripts that reproduce the same diagrams when executed on the console. By editing downloaded scripts, users perform advanced analysis using Ruby scripting language [4].

Overlaying different diagrams is available through layer functionality. Fig. 3 shows the Method & Axis tab window for a diagram with two layers. Although it is not easy to see the main diagram, which is located beneath the tab window, in Fig. 3, the bottom layer, Layer 0, shows temperature by color tone and contour, whereas the top layer, Layer 1, shows wind vectors. Visualization parameters for each layer can be edited separately through the tab user interface in parameter windows. Layers of diagrams can be added, removed, and swapped, like the layer function of graphic editing programs.

3. Application to Ensemble NWP Data

The exponential growth of computer power allows us to perform high-resolution forecasts and increase the number of ensemble members. Accurate prediction by ensemble NWP systems would help risk management. Probabilistic weather information also helps decision making in other sectors. Specifically, commercial decisions are often made, not on the basis of events which are likely to occur, but on the basis of events which are unlikely to occur, but which, if they did occur, would involve serious financial loss [5].

As mentioned in section 1, raw output of ensemble NWPs is not suitable for dissemination to the public be-
cause of the larger sizes compared to deterministic NWPs and difficulties in interpretation. This requires data conversion to a user-friendly form before dissemination. Data types and formats required by users vary depending on their purpose. An ensemble-mean horizontal distribution might, for example, be the best in some cases, whereas statistics on extreme values at a particular point might be helpful in other cases. We have to consider not only the development of ensemble prediction systems but also the development of decision support systems to utilize output from ensemble NWPs. The decision support system conveys information from the NWP sector to decision makers in social, economic, and environmental sectors in an appropriate way. This idea is shown as a schematic diagram of the end-to-end forecast system in the THORPEX framework (Fig. 6) [2].

For this purpose, we started to develop an experimental data handling system for decision support. We have developed tools to store ensemble NWP data in a database and handle probabilistic information obtained from ensemble NWPs, depending on the purpose. Here, Gfdnavi is suitable to build such a system because Gfdnavi has a database system with analysis, visualization, and knowledge-sharing tools.

In recent years, it has become possible to install ensemble NWP systems in developing countries and regions [6]. As mentioned above, decision support systems for ensemble NWP data are also required at the same time. One of the purposes of the current study is to demonstrate how to develop such a system with free open-source programs.

This section briefly introduces the data handling system for ensemble NWP data.

3.1. Test Case: Cyclone Nargis

Cyclone Nargis hit Myanmar in May 2008, killing more than 13,800 people and causing economic loss reported to be more than 10 billion dollars. This is the worst natural disaster in Myanmar’s recorded history [7, 8].

The Japan Meteorological Agency (JMA) Meteorological Research Institute has conducted experimental ensemble NWPs and storm surge simulation on Cyclone Nargis and used the output as test data for tools to analyze and visualize ensemble NWP data, as detailed in the subsections that follows. Data is also used to create an interactive document on how to utilize probabilistic information obtained from ensemble NWP data.

The atmospheric model we use is the JMA nonhydrostatic model (NHM) [9]. The period of the experiment is from April 30, 2008, 1200 UTC, to May 3, 1200 UTC. The horizontal resolution is 10 km. Initial and boundary conditions for the NHM control run are taken from the JMA’s high-resolution global analysis and forecast. Initial and boundary perturbations are produced from the JMA’s one-week global ensemble prediction system. In total, the ensemble size is 21 – control run and 20 perturbed runs. The Princeton ocean model (POM) [10] is used for storm surge simulation. Input to POM is output from NHM, i.e., wind at 10 m and sea level pressure. The horizontal resolution in POM is 3.5 km. See Saito et al. (2010) [11] and Kuroda et al. (2010) [12] for simulation details.

3.2. Analysis and Visualization of Ensemble NWP Data

Gfdnavi is designed to accept user-defined plug-ins. Here, plug-ins of analysis and visualization tools for ensemble NWP data are implemented. Fig. 7 shows a drop-down menu for selecting visualization methods on Gfdnavi. Visualization tools for ensemble prediction data, ensemble_1D and ensemble_2D, are highlighted by red ellipses. Users do not need to specify details for mathematical operations to visualize ensemble prediction data because they are operated within these visualization tools. This enables a quick look at ensemble NWP data.2

2. The latest versions of tools are available from the web page: http://www-metec.kugi.kyoto-u.ac.jp/otsuka/gfdnavi_ensemble_tools/ (the address is subject to change).
3.2. 1D Diagrams of Ensemble NWP Data

In this section, 1D visualization methods for ensemble NWP data are demonstrated using sea surface elevation simulated by POM.

Figure 8 shows the horizontal distribution of maximum sea surface elevation for all of the time steps and all of the ensemble members. The highest sea surface elevation is obtained at 95.07°E, 16.1°N, which is the mouth of the Irrawaddy River (Ayeyarwady River). Although the maximum expected sea surface elevation is valuable for disaster management, it is more valuable for obtaining probabilistic information from the same data. The time evolution of sea surface elevation at the mouth of the Irrawaddy River will be shown next to quantify probability.

The top-left panel in Fig. 9 is a plume diagram of sea surface elevation at the mouth of the Irrawaddy River. A plume diagram consists of superimposed 1D line plots. Typically, time series of all of the ensemble members are superimposed. At the initial time, differences among ensemble members are small. As time evolves, differences become larger – the reliability of the forecast becomes lower –, which makes superimposed line plots spread away from each other. (This looks like the plume rising from a chimney.) In this figure, the ensemble member that shows the largest sea surface elevation is highlighted in red. The value and timing of the maximum sea surface elevation vary much from member to member.

The top-right panel in Fig. 9 shows the ensemble mean and spread (standard deviation) for the same data. The spread becomes larger as the sea surface elevation becomes higher.

The bottom-left panel in Fig. 9 shows a box plot for the same data. A box plot or box-and-whisker plot is introduced by Tukey (1977) [13]. The original box plot shows five quantities: minimum, lower quartile, median, upper quartile, and maximum. There are many variations [14]. In the bottom-left panel in Fig. 9, the maximum, minimum, and standard deviation are shown. With this plot, it is possible to see the distribution of extreme values (outliers) that are not shown in the spread.

The bottom-right panel in Fig. 9 shows percentiles of distribution for the same data. For example, most of the members show a sea surface elevation of about 1 m at around May 2, 0000 UTC, whereas some members show a sea surface elevation of nearly 4 m. After May 2, 0000 UTC, in contrast, ensemble members show a variety of sea surface elevations.

As seen above, we can extract many kinds of information from 1D diagrams of the same ensemble NWP data by using different plotting methods.

3.2.2. 2D Diagrams of Ensemble NWP Data

In this section, 2D visualization methods for ensemble NWP data are demonstrated using the sea level pressure simulated by NHM.

Three kinds of 2D plots for ensemble NWP data are implemented: ensemble mean and spread, so-called spaghetti plot, and distributions of probability exceeding a threshold value. The spaghetti plot shows the same isolopleth lines for all of the ensemble members in a single plot. The higher density of contour lines means the higher reliability of the forecast. The lower density of lines, which shows patterns similar to spaghetti on a plate, means lower reliability. A spaghetti plot of the geopotential height at 500 mb with isolopleth lines of 5640 m, for example, is shown as a cover illustration of Kalnay (2003) [15].

Figure 10 shows a spaghetti plot of sea level pressure with contour lines of 1000 hPa at the initial time, 24 hours, and 48 hours. As time evolves, the low-pressure region, which corresponds to Cyclone Nargis, becomes larger and
An Experimental Data Handling System for Ensemble Numerical Weather Predictions Using a Web-Based Data Server and Analysis Tool “Gfdnavi”

Fig. 9. Time series of sea surface elevation at the Irrawaddy River mouth simulated by POM. Diagrams are produced by the tool for one-dimensional visualization of ensemble prediction data. Plume (top left), mean and spread (top right), box and whisker (bottom left), and percentile (bottom right).

3.2.3. Importance of Advanced Utilization of Ensemble NWP Data

From Fig. 9, we point out the importance of utilizing information from all of the ensemble members. The difference between different members becomes larger. Specifically, the eastward movement speed of the cyclone differs from member to member, which makes differences in value and the timing of the maximum sea surface elevation shown in Fig. 9.

If one makes decisions based only on a single forecast, the forecast may show minor damage by chance. Some ensemble members, for example, show a maximum sea surface elevation of about 2 m (Fig. 9, top left). Shibayama et al. (2009) [16] reported that water level deviation due to the storm surge was estimated at 3–4 m from a field survey along the Yangon River. If a single forecast that predicts the maximum sea surface elevation of 2 m is used for decision making, decision makers may fail to make appropriate decisions.
If only the ensemble mean is available for decision making, the maximum sea surface elevation becomes 1.6 m at May 2, 0100 UTC (Fig. 9, top right), which is lower than maximum sea surface elevations of individual ensemble members and much lower than estimation from a field survey. Although the ensemble mean is often used as a product of ensemble predictions, this clearly shows that the ensemble mean is not suitable for the current purpose. If information on all of the ensemble members can be utilized, a sea surface elevation of nearly 4 m is predicted on May 1, 2100 UTC, and May 2, 0100 UTC (Fig. 9, top left), which is comparable to the value obtained by a field survey. It is possible, in addition, to say that the probability that the sea surface elevation exceeds 2 m is more than 31% (Fig. 9, bottom right).

As demonstrated here, it is important to fully utilize ensemble prediction data to provide meaningful information for decision making.

3.3. User’s Guide of Tools for Ensemble NWP Data

We created an interactive tutorial document on how to analyze ensemble NWP data by using the experimental analysis system built into Gfdnavi to obtain probabilistic information that is useful for the prevention and mitigation of meteorological disasters.

The user’s guide consists of systematic explanations of visualization functions of the system. Generic visualization functions include 1D plots (line, marker, and bar), 2D plots (contour, tone, and vector arrow), and statistical plots (scattergram and histogram). Specifically, how to extract information by slicing multidimensional data along each axis is demonstrated for five-dimensional ensemble NWP data as a function of latitude, longitude, height, time, and ensemble member.

Visualization functions specific to ensemble NWP data include 1D diagrams such as plume diagrams and 2D diagrams such as the spaghetti plot are described. How to compute the ensemble mean and spread by combining analysis tools in Gfdnavi is explained as analyses with simple mathematical operations.

Sample diagrams shown in the tutorial document are also provided in an interactive document on the experimental analysis system using the knowledge documenting system of Gfdnavi (Fig. 11). A hyperlink just below each diagram, Redraw this image, opens the Analysis window. When the window opens, the same diagram with the same parameters appears. This enables us to modify parameters of sample diagrams and redraw them.

This also enables us to save diagrams on the database and write documents on diagrams. By using this tool, it is possible to archive and share knowledge that is needed to analyze and interpret NWP data. Comment functionality also helps communication with other users. By archiving many case studies, the system will provide useful infor-
4. Summary

We have developed a new Gfdnavi user interface for analysis and visualization that will improve the user experience by seamless operations of analysis and visualization. The new user interface enable users to edit multiple diagrams at the same time and also enables overlaying multiple diagrams like a layer function of graphic editing programs.

We have also introduced an experimental system developed on Gfdnavi to provide analysis and visualization tools for ensemble numerical weather prediction data. These tools will help data processing and the transfer of data from weather centers to decision makers in various sectors. An interactive document that provides basic ideas of how to utilize probabilistic information on ensemble data is written with the knowledge documentation system of Gfdnavi in which automatic hyperlinks in the document enable users to edit diagrams in the document themselves.

Although Gfdnavi provides a useful platform, it is necessary to provide plug-in functionality depending on the needs of each user. The analysis and visualization tool for ensemble prediction data is an example of such plug-ins. We would like to improve the basic functions of Gfdnavi as well as plug-in tools for various needs. Improvement of the documentation and user interface to edit plug-ins is also important for enabling users to create plug-ins for them.

Acknowledgements

This work was supported in part by MEXT Special Coordination Funds for Promoting Science and Technology for FY 2007-09 “International Research for Prevention and Mitigation of Meteorological Disasters in Southeast Asia” and Kyoto University’s Global COE Program for FY 2009-13 “Sustainability/Survivability Science for a Resilient Society Adaptable to Meteorological Disasters in Southeast Asia” and Kyoto University’s Global COE Program for FY 2009-13 “Sustainability/Survivability Science for a Resilient Society Adaptable to Extreme Weather Conditions.” We thank Dr. Kazuo Saito and Dr. Tohru Kuroda for providing experimental ensemble numerical prediction data.

References:


Name: Seiya Nishizawa
Affiliation: Research Scientist, RIKEN Advanced Institute for Computational Science
Address: 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
Brief Career:
2005- Researcher, Kyoto University
2008- Researcher, Kobe University
2009- Assistant Professor, Kobe University
2012- Research Scientist, RIKEN Advanced Institute for Computational Science
Selected Publications:

Academic Societies & Scientific Organizations:
• Meteorological Society of Japan (MSJ)
• The Japanese Society for Planetary Sciences (JSPS)

Name: Takeshi Horinouchi
Affiliation: Division of Earth System Science, Graduate School of Environmental Science, Hokkaido University
Address: Kita-10, Nishi-5, Kita-ku, Sapporo 060-0810, Japan

Name: Shigeo Yoden
Affiliation: Division of Earth and Planetary Sciences, Graduate School of Science, Kyoto University
Address: Kitashirakawa-Oiwake-cho, Sakyo, Kyoto 606-8502, Japan
Brief Career:
1983 Research Associate, Kyoto University
1985-1987 JSPS Fellow for Research Abroad, University of Washington
1987 Associate Professor, Kyoto University
2002- Professor of Meteorology, Kyoto University
Selected Publications:
Academic Societies & Scientific Organizations:
• The Meteorological Society of Japan (MSJ)
• The American Meteorological Society (AMS)
• International Commission on the Middle Atmosphere, IAMAS/IUGG