THE THORPEX INTERACTIVE GRAND GLOBAL ENSEMBLE

BY Philippe Bougeault , Zoltan Toth, Craig Bishop, Barbara Brown, David Burridge, De Hui Chen, Beth Ebert, Manuel Fuentes, Thomas M. Hamill, Ken Mylne, Jean Nicolau, Tiziana Paccagnella, Young-Youn Park, David Parsons, Baudouin Raoult, Doug Schuster, Pedro Silva Dias, Richard Swinbank, Yoshiaki Takeuchi, Warren Tennant, Laurence Wilson, and Steve Worley

Leading NWP centers have agreed to create a database of their operational ensemble forecasts and open access to researchers to accelerate the development of probabilistic forecasting of high-impact weather.

OBJECTIVES AND CONCEPT. During the past decade, ensemble forecasting has undergone rapid development in all parts of the world. Ensembles are now generally accepted as a reliable approach to forecast confidence estimation, especially in the case of high-impact weather. Their application to quantitative probabilistic forecasting is also increasing rapidly. In addition, there has been a strong interest in the development of multimodel ensembles, whether based on a set of single (deterministic) forecasts from different systems, or on a set of ensemble forecasts from different systems (the so-called superensemble). The hope is that multimodel ensembles will provide an affordable approach to the classical goal of increasing the hit rate for prediction of high-impact weather without increasing the false-alarm rate.

This is being taken further within The Observing System Research and Predictability Experiment (THORPEX), a major component of the World Weather Research Programme (WWRP) under the World Meteorological Organization (WMO). A key goal of THORPEX is to accelerate improvements in

AFFILIATIONS: Bougeault, Fuentes, and Raoult—European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom; Toth—National Centers for Environmental Prediction, Camp Springs, Maryland; Bishop—Navy Research Laboratory, Monterey, California; Brown, Schuster, and Worley—National Center for Atmospheric Research, Boulder, Colorado; Burridge and Parsons—World Meteorological Organization, Geneva, Switzerland; Chen—Chinese Meteorological Administration, Beijing, China; Ebert—Centre for Australian Weather and Climate Research, Melbourne, Victoria, Australia; Hamill—NOAA/Earth System Research Laboratory, Boulder, Colorado; Mylne, Swinbank, and Tennant—Met Office, Exeter, United Kingdom; Nicolau—Météo-France, Toulouse, France; Paccagnella—Agenzia Regionale Prevenzione e Ambiante dell'Emilia-Romagna, Bologna, Italy; Park—Korean Meteorological Administration, Seoul, South Korea; Silva Dias—National Laboratory of Scientific Computing, and University of São Paulo, São Paulo, Brazil; Takeuchi—Japan Meteorological Agency, Tokyo, Japan; Wilson—Meteorological Service of Canada, Montreal, Quebec, Canada

CORRESPONDING AUTHOR: Philippe Bougeault, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
E-mail: philippe.bougeault@meteo.fr

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/2010BAMS2853.1

In final form 8 December 2009

the accuracy of 1-day to 2-week high-impact weather forecasts for the benefit of humanity. It is therefore not surprising that a key component of THORPEX is the THORPEX Interactive Grand Global Ensemble [TIGGE; see, e.g., the THORPEX Implementation Plan (TIP); TIP 2005].

TIGGE was initiated in 2005 at a workshop hosted by the European Centre for Medium-Range Weather Forecasts (ECMWF). A full report of this event was prepared by Richardson et al. (2005).

The following objectives of TIGGE were agreed to at the workshop:

i) enhance collaboration on ensemble prediction, both internationally and between operational centers and universities;
ii) develop new methods to combine ensembles from different sources and to correct for systematic errors (biases, spread over-/underestimation);
iii) achieve a deeper understanding of the contribution of observation, initial, and model uncertainties to forecast error;
iv) explore the feasibility and the benefit of interactive ensemble systems responding dynamically to changing uncertainty;
v) enable evolution toward an operational system, the Global Interactive Forecast System (GIFS).

To meet these objectives, it was agreed that ensemble forecasts generated by a number of NWP centers (hereafter "data providers") would be accumulated in real time in databases operated by three TIGGE "archive centers" (see Table 1) and made accessible to the scientific community for research and education with only a slight (2 day) time delay. The highest-priority data accumulated in the TIGGE archive are the ensemble forecasts generated routinely (operationally) at major forecast centers around the world. These core data stored in the TIGGE archive are accumulating at a daily rate of approximately 245 GB from 10 providers from around the world (see Table 1). Additional special datasets may be added in the future for specific research and application areas. Ensemble forecasts from a number of limited-area systems are being considered for addition to the archive.

As implied by its title, there is a concept of "interactivity" in TIGGE. Different kinds of interactivity may be invoked in building a multimodel ensemble; for example, the choice of the components or the weights attributed to the components may vary with time, domain, and weather situations, . . . In the future, decisions about these aspects may be entirely automated or supervised by a human forecaster. Interactivity may also exist in the observations used in the data assimilation system or in the decision to activate a specific high-resolution system when the weather situation demands it. The general architecture of TIGGE was defined in such a way as to allow for the exploration of these various possibilities. Research and practical considerations will ultimately dictate which of the above approaches is more beneficial, and the optimal configuration will probably be different in different parts of the world.

The TIGGE project has been developed under the leadership of the THORPEX GIFS-TIGGE Working Group, to which most of the authors belong. The WMO Working Group on Numerical Experimentation (WGNE)/WWRP Joint Working Group on Forecast Verification Research (JWGFVR) advises the project on verification methodol-

TABLE 1. TIGGE portals and data providers.

TIGGE archive centers, main Web pages, and data portals	
CMA	http://wisportal.cma.gov.cn/tigge/
NCAR	http://tigge.ucar.edu
ECMWF	http://tigge-portal.ecmwf.int
TIGGE	http://tigge.ecmwf.int
TIGGE-LAM	www.smr.arpa.emr.it/tiggelam
Centers supplying daily forecasts to the TIGGE archive	
ECMWF	
NCEP	
MSC	
CAWCR	
CMA	
Brazilian Centro de Previsão de Tempo e Estudos Climático (CPTEC)	
JMA	
Korea Meteorological Administration (KMA)	
Météo-France (MF)	
UKMO	

ogy. In addition, the WMO Expert Team on ensemble prediction systems (EPSs) advises the project on a number of issues, for instance, metadata formulation. TIGGE has strong links with the North American Ensemble Forecast System (NAEFS; see Toth et al. 2005), which synthesizes ensemble products from the National Centers for Environmental Prediction (NCEP) and the Meteorological Service of Canada (MSC). Although NAEFS uses data from only two centers and produces real-time operational products, TIGGE and NAEFS share many technical aspects, and NAEFS plans to implement results from TIGGE. It is believed that TIGGE and NAEFS will ultimately evolve into a single operational system. TIGGE is also registered as Task WE-06-03 of the Group on Earth Observations (2007). It has general relevance to the Group on Earth Observations's (GEO's) societal benefit areas that will benefit from access to advanced multimodel global weather forecasts and the derived products, especially in areas related to risk management, disaster mitigation, energy, agriculture, water, the environment, and health.

BUILDING THE TIGGE DATABASES. The implementation of TIGGE has been quite challenging. Data must be collected from 10 different centers and redistributed to a potentially large number of users very rapidly, using only readily available communication technologies, such as the Internet. The content of the database must be as homogeneous and have as few gaps as possible. The archive centers must operate user-friendly interfaces, enabling researchers to obtain subsets of ensemble data, especially over geographic regions of their choice. This postprocessing of archived data, done at the archive centers, typically includes grid conversions, format conversions, and the extraction of subareas, parameters, and levels. Archive centers must also provide links to associated regional and user-specific observational datasets.

Content and format of the archive. As a starting point, all partners have agreed on a common way of referencing data within the TIGGE dataset. Fields are described using the following attributes: *analysis date, analysis time, forecast time step, origin center, ensemble member number, level,* and *parameter*. In this context "parameter" refers to the physical quantity represented by the field, for example, temperature and pressure. Furthermore, all partners have agreed to provide data in the same units and with the same period of accumulation (when applicable). This led to the definition of the TIGGE core dataset to which all data providers must adhere (Table 2).

When the first data transfers were being set up between the partners, it became clear that most data providers could not contribute to the full agreed list of products, mainly because these products were not produced by their models. It was decided that waiting for all of the partners to upgrade their systems to produce the missing fields was an unnecessary delay in the building of the archive. Because all data providers were producing the most important fields (the usual surface parameters and upper-air data on pressure levels), a staged approach was adopted. Data providers would join the project by sending currently available parameters, and would add more parameters during the course of the project. The actual data accumulation started between October 2006 and January 2008, depending on the parameter and data provider. The TIGGE database now contains most requested data from all of the data providers, and holds more than 180 TB of data (1.1 billion fields; see Table 3). Forecast data have now been archived for more than 2 yr for several parameters.

To guarantee the best precision, original model grids and resolutions are preserved whenever possible. Data providers supply data on a horizontal grid of their choice, as close as possible (identical if possible) to the computational grid of their model. These data are stored in the database without any modification. On the other hand, users generally want data interpolated on common regular grids of their choice. The archive centers offer this interpolation service. Before delivery, data may be interpolated to a single point or to a regular, limited-area, or global latitude–longitude grid specified by the user. To respect the unique features of each model, data providers are encouraged to supply and regularly update the interpolation software used by the archive centers. Alternatively, the archive centers can use other available interpolation software.

As a common archive data format, it was decided to use Gridded Binary (GRIB) edition 2; it is the only WMO standard that supports ensemble data without the need for local extensions (see the WMO manual on codes, Vol. I.2, Part B, FM-92 GRIB edition 2). Moreover, the NAEFS community is committed to using it. Data providers are requested to provide data to archive centers directly in the archive format.

Data transfers, operational aspects, and quality control. After extensive testing, it was shown that Internet Data Distribution (IDD) system/Local Data Manager (LDM), an Internet-based distribution system developed by Unidata, suits TIGGE requirements. This was therefore defined as the preferred solution

Table 2. Agreed list of parameters and units to be delivered to the TIGGE database. Note that temperature, *u* velocity, *v* velocity, and specific humidity are provided on the following isobaric surfaces: 1,000, 925, 850, 700, 500, 300, 250, and 200 hPa. The geopotential height is provided on the same surfaces plus 50 hPa. All parameters have to be provided 6 hourly, included the initial time of the forecast. All of the fluxes are accumulated since the beginning of the forecast.

Parameter	Unit
Surface level parameters	
Mean sea level pressure	Pa
Surface pressure	Pa
10-m *u* velocity	m s^{-1}
10-m *v* velocity	m s^{-1}
Surface temperature	K
Surface dewpoint temperature	K
Surface max temperature	K
Surface min temperature	K
Skin temperature	K
Soil moisture	kg m^{-3}
Soil temperature	K
Total precipitation (liquid + frozen)	kg m^{-2}
Snowfall water equivalent	kg m^{-2}
Snow depth water equivalent	kg m^{-2}
Total cloud cover	0%–100%
Total column water	kg m^{-2}
Time-integrated surface latent heat flux	W m^{-2} s
Time-integrated surface sensible heat flux	W m^{-2} s
Time-integrated surface net solar radiation	W m^{-2} s
Time-integrated surface net thermal radiation	W m^{-2} s
Time-integrated outgoing longwave radiation	W m^{-2} s
Sunshine duration	s
Convective available potential energy	J kg^{-1}
Convective inhibition	J kg^{-1}
Orography (geopotential height at the surface)	m
Land–sea mask	0–1
Parameters on isobaric surfaces	
Temperature on eight isobaric surfaces	K
Geopotential height on nine isobaric surfaces	m
U velocity on eight isobaric surfaces	m s^{-1}
V velocity on eight isobaric surfaces	m s^{-1}
Specific humidity on eight isobaric surfaces	kg kg^{-1}
Parameters on potential temperature surfaces	
Potential vorticity on θ = 320-K surface	K m^2 kg^{-1} s^{-1}
Parameters of potential vorticity unit (PVU) surfaces	
Potential temperature on 2-PVU surface	K
U velocity on 2-PVU surface	m s^{-1}
V velocity on 2-PVU surface	m s^{-1}

Table 3. Parameter availability and configuration of ensemble for each data provider.

	CAWCR	CMA	MSC	CPTEC	ECMWF	JMA	KMA	MF	NCEP	UKMO
Standard fields (Out of 73 requested)	55	60	56	55	70	61	46	62	69	70
Ensemble members	33	15	21	15	51	51	17	11	21	24
Forecast length (day)	10	10	16	15	15	9	10	3	16	15
Forecast cycles per day	2	2	2	2	2	1	2	1	4	2

for the exchange of data between TIGGE partners (see additional information in Fig. 1a).

The available network bandwidth between Europe, the United States, and China is sufficient to meet the needs of TIGGE. Nevertheless, this would become a limiting factor if TIGGE partners decided to engage in real-time exchange for operational products.

The archive centers are in charge of the technical coordination of the project. For day-to-day operations, tools have been created to monitor the data transfers. Each archive center maintains a Web page showing volumes, the date of data, and the date of receipt from each data provider. Every effort is made to ensure that data series are complete and of the highest quality. Detailed information on the quality control procedures is given in Raoult and Fuentes (2008).

ACCESS TO TIGGE DATA FOR RESEARCH AND EDUCATION.

Access to TIGGE data is provided for research and education through a simple electronic registration process, which requires a valid e-mail address and acknowledgment of the conditions of supply. Under the simple registration process, access is given with a delay (48 h) after the initial time of each forecast. Real-time access is granted (subject to bandwidth limitations) in some cases, for example, for field experiments and projects of special interest to THORPEX. Registration for this real-time access is handled via the THORPEX International Project Office.

Data access is operated via the three TIGGE data portals operated by the National Center for Atmospheric Research (NCAR), ECMWF, and the China Meteorological Administration (CMA; see the URL for each portal in Table 1). The current functionalities of the data portal are i) registration; ii) search, discover, and download files; iii) select data by initialization date/time, data provider, file type, and forecast time; iv) interpolate data on a regular, limited-area, or global latitude–longitude grid specified by the user; and v) check volume and download data.

All three archive centers are currently able to distribute data in GRIB2 format. Network Common Data Format (NETCDF) is also available from NCAR and should soon become available from the other centers. Plans to expand the services available include, inter alia, the possibility of setting up standing data requests (e.g., order specific data to be sent routinely every day to interested users).

At the beginning of 2009, the three data portals had a total of about 230 registered users, of which one-third were active. Figure 1b shows the country of origin of the registered users.

EARLY RESULTS FROM RESEARCH BASED ON TIGGE.

A list of research papers based on TIGGE data is continuously updated online (see http://tigge.ecmwf.int/references.html). Only a few of them are being reviewed here.

Performance of individual systems. Park et al. (2008) have investigated the performance of various single- and multimodel ensemble systems available from TIGGE up to December 2007 (thus, their results reflect the performance of the various systems only up to this time). This study focused on 500-hPa geopotential height and 850-hPa temperature and was the first extensive comparison of the global operational ensemble prediction systems. Each system was verified primarily against its own analysis, but the sensitivity to the choice of the verification analysis was also investigated. This highlighted large differences in the forecast quality of the various contributed systems, both for the deterministic forecasts based on the control runs or on the ensemble mean, and, even more, for the probabilistic forecasts. Differences in the accuracy of probabilistic forecasts were shown to be due to both model error characteristics and to the quality of the spread–error relationship. Ideally, the spread of an ensemble should be equal to the RMSE of the ensemble mean throughout the forecast range, for all of the forecast parameters. This turns out to be a very challenging goal to attain. The best calibrated ensemble systems have now reached this optimal calibration for upper-air parameters such as the geopotential height at 500 hPa or the temperature at 850 hPa. For other parameters

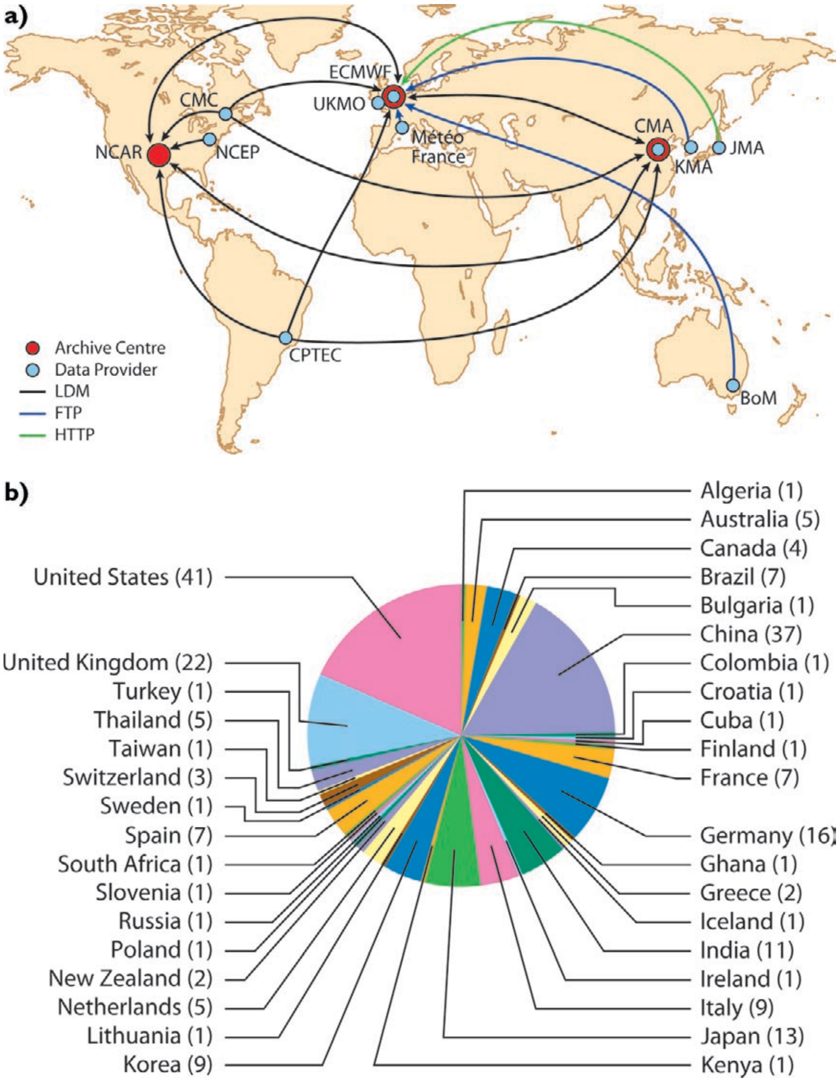

Fig. 1. (a) Protocols for exchange of data between data providers and archive centers. The preferred solution, LDM, is a broadcasting system, based on a subscription mechanism: a "downstream" LDM can subscribe to "products" from an "upstream" LDM. When a product is inserted in the upstream LDM, it is automatically sent to all of the downstream LDMs that have subscribed to this product. Unfortunately, such a broadcasting system does not guarantee that products will be received by all downstream LDMs, particularly if some are temporarily not running. To overcome this problem, a protocol has been defined on top of LDM to exchange fields by specifying a file name convention and a series of messages to request retransmission of missing fields. A complete description of the protocol is available on the TIGGE Web site (http://tigge.ecmwf.int). Although LDM is the preferred solution for the exchange of data between the TIGGE partners, it was not always possible for data providers to install an LDM server at their site. Some decided to use either FTP or hypertext transport protocol (HTTP) to transfer the data to one of the archive centers, which would in turn relay it to the two others. Figure 1a shows the various transfer protocols used between the data providers and archive centers. (b) Number of registered TIGGE users (by country).

(e.g., surface temperatures and precipitation) this has not been yet reached, and some systems are still quite far from it for all parameters. This, on top of model error differences, was shown to result in differences of up to 3 days in forecast skill between the various systems. Another result worth mentioning is that in the tropics, all systems (in 2007) were substantially underestimating the spread compared to the RMSE of the ensemble mean. This finding formed a strong incentive for several data providers to address more vigorously the issue of improving the quality of ensemble forecasts in the tropics.

The choice of the verification analysis was shown to have a relatively small impact for upper-air parameters in the midlatitudes as long as one of the best analyses was used. On the other hand, in the tropics, or generally for the near-surface parameters, despite considerable work at NWP centers, there are still large differences between analyses from various systems, and therefore the forecasts from most systems verify significantly better when scored against their own analysis than when scored against the analysis of a different system. This must be kept in mind when working on multimodel systems (see further discussion below).

To complement the above results, a more recent assessment of the spread–error relation in TIGGE systems is shown in Figs. 2 and 3, based on forecasts from December 2008. Figure 2 shows how the spread in sea level pressure develops with forecast range as a function of the latitude. It can be readily compared to Fig. 3, where the RMS errors of the ensemble means are shown with the same units and color code. Note that in order to obtain a fairer comparison, subensembles of 10 members have been used for each system, resulting in some degradation of the results for the largest ensembles. The spread in this recent period is still often smaller than the RMSE of the ensemble mean. This is especially true in the Southern Hemisphere, and in the tropics. For some systems, this situation is actually expected because they do not use initial perturbations in these regions [e.g., the Japan Meteorological Agency (JMA) system in the Southern Hemisphere]. Even in the Northern Hemisphere there are large differences from system to system, showing that beyond the size of the ensemble, the methods used to represent initial and model uncertainty are important.

Skill of multimodel systems. Park et al. (2008) have also compared the performances of various single- and multimodel systems, both with and without bias correction. They assessed several methods to compute the bias correction and showed that this is a sensitive issue. One particular result is reproduced here in Fig. 4 (cf. Fig. 17 of Park et al.). It compares the performance of the single ECMWF ensemble, with and without bias correction, and two bias-corrected multimodel ensembles [ECMWF + Met Office (UKMO) and ECMWF + UKMO + JMA + CMA]. Both the root-mean-square-error of the ensemble means and the ranked probability skill score (RPSS) are shown. The RPSS computation was based on 10 climatologically equally likely categories. The results cover 86 cases from June to August 2007. It can be seen that the performance for the geopotential height at 500 hPa over the Northern Hemisphere benefits very little from either the bias correction or the addition of the extra members. On the other hand, for temperature at 850 hPa over the tropics, bias correction has a large positive impact on the quality of ECMWF-only ensemble. The addition of extra members from other systems also has a positive impact, although the authors note that some saturation effect can be seen when many systems are used. Qualitatively similar results were found with other combinations of models and other periods; for example, multimodel forecasts

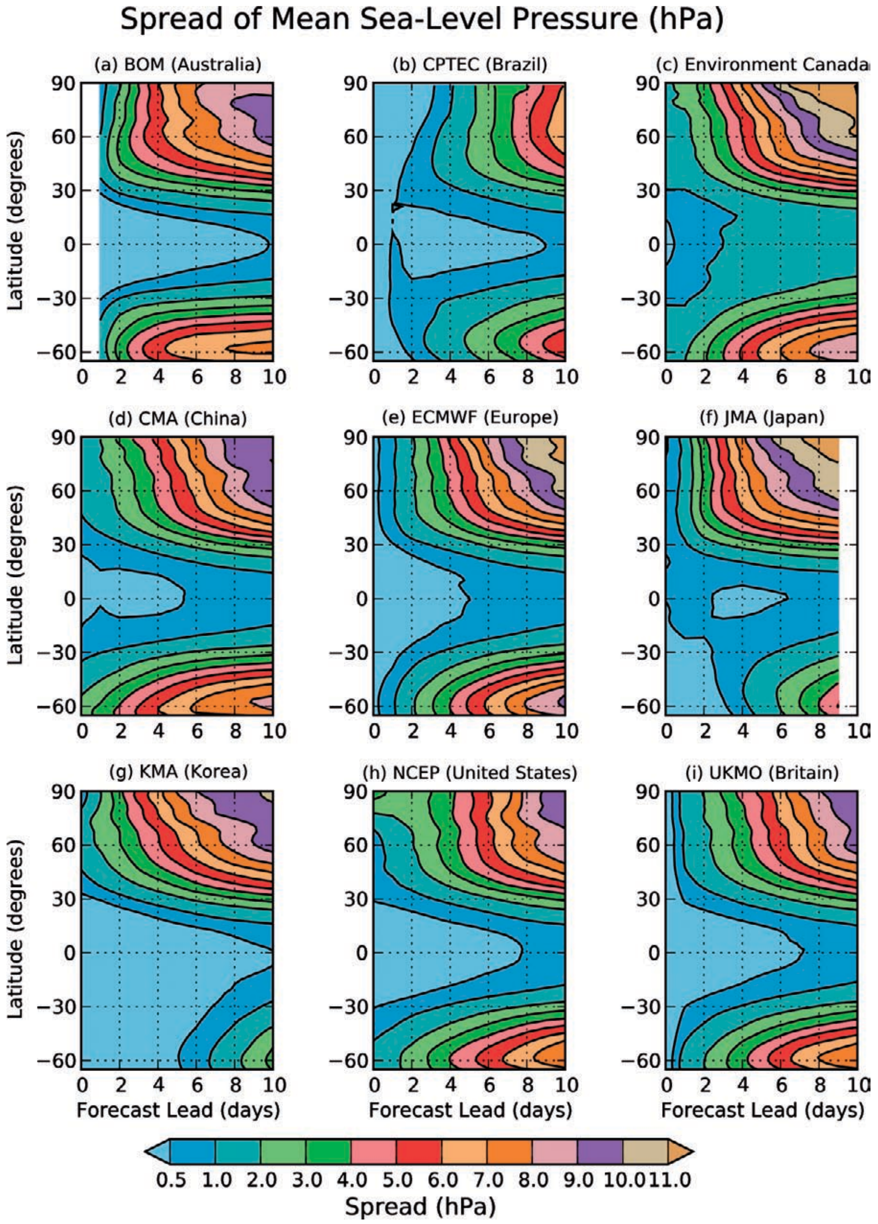

Fig. 2. Spread of the mean sea level pressure for the various TIGGE ensembles as a function of the forecast range and the latitude. For a fair comparison, only the first 10 members of each ensemble have been used. The period covered is Dec 2008.

only gave small benefits for forecasts of NH 500-hPa geopotential height, but gave generally better results for tropical 850-hPa temperatures. The results of Park et al. (2008) are generally confirmed by the independent work of Matsueda and Tanaka (2008).

A possible weakness of both Park et al.'s (2008) and Matsueda and Tanaka's (2008) results lies in their common choice of ECMWF analysis as the verification for all of the above systems. As discussed above, the choice of the optimal verification analysis is both a difficult and a sensitive one, and additional work is needed before drawing final conclusions about the relative merits of the various systems. Some fairer ways to compare ensembles or to evaluate multimodel ensembles with respect to analyses have been discussed by the GIFS-TIGGE group. They include the following:

i) Consider the analyses from all of the models under consideration as an ensemble, and use, for example, the rank probability skill score to compare the forecast and analysis distributions. This approach, however, has been criticized on the basis that the quality of the analyses from some centers is on average higher than that from other centers. One could account for objectively known accuracy differences by some sort of weighting scheme among the analyses. The basis of the weighting scheme would have to be determined independently of all of the models.

ii) Choose the verifying analysis at random for each case in the verification sample, with all of the candidate analyses having an equal chance of being chosen.

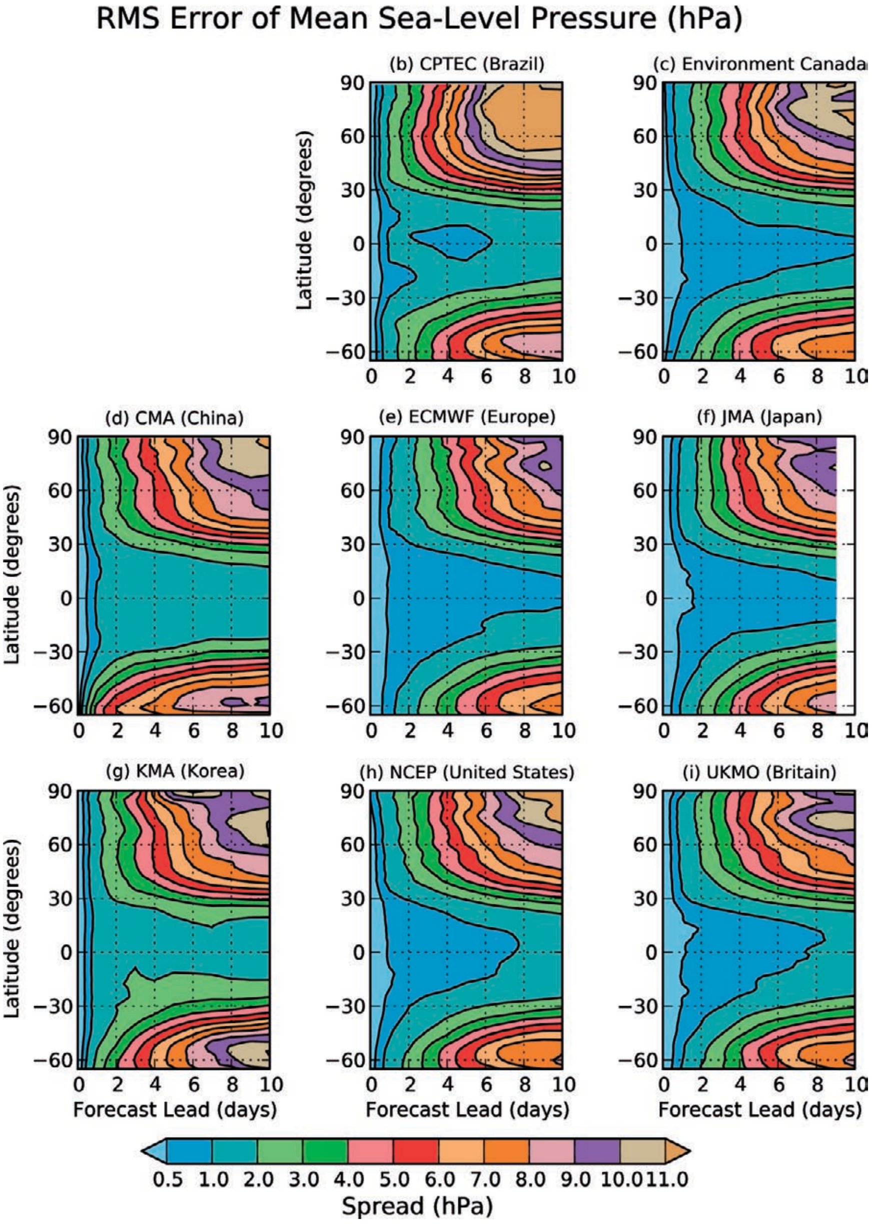

Fig. 3. RMSE of the ensemble mean (sea level pressure) for the various systems (the computation was not possible for the Australian system because no verification analysis is available for this period).

iii) Use an analysis that does not use any model forecast as a trial field. In general, this would be restricted to areas with reasonable data coverage, and would lead to verification over regional rather than global domains, requiring regional subsets of the TIGGE data. However, data-dense areas are often those areas where it is most important to know the ensemble performance.

It is clear that direct verification against observations needs to be done. Not only would this be fair, because all ensembles would be verified against the same model-independent data, but verification against observations is relevant to a wide variety of users. Verification against observations is, however, more difficult to carry out than verification against analysis, and is just beginning for the TIGGE archive.

Johnson and Swinbank (2009) investigated the benefit of a three-model ensemble, using ECMWF, NCEP, and UKMO ensembles. Figure 5 shows Brier skill scores for mean sea level pressure and surface (2 m) air temperature, verifying the skill of categorical probabilistic forecasts, with category boundaries set as the climatological quantiles defined using 40-yr ECMWF Re-Analysis (ERA-40) data. Each forecast was bias corrected, and forecasts were verified against a multimodel analysis (the mean of the three analyses). Three variations of multimodel ensemble were assessed: first, each ensemble was weighted equally; second, each ensemble was weighted to take account of its estimated RMS error; and third, both the weights and variance of each ensemble were adjusted. Figure 5a shows that the skill in forecasting sea level pressure greater than the climatological mean is very similar for both the ECMWF and multimodel ensembles. Figure 5b compares scores for forecasts of 2-m temperature, relative to the mean; in this case, all three multi-model ensembles give a significant improvement over any single ensemble. The largest benefit of multimodel ensembles is shown for forecasts of 2-m temperature greater than the 90th percentile (Fig. 5c). The results show relatively small impacts from varying the ensemble weighting, consistent with earlier results (e.g., Peña and Van den Dool 2008).

These statistical studies of the benefits of multimodel ensembles have been complemented by case studies of high-impact weather events (e.g., Titley et al. 2008). In late July 2005, a heat wave affected southeast Europe; from 21 to 25 July, temperatures reaching or exceeding 45°C affected most parts of Greece, Bulgaria, Romania, and Serbia. More than 500 deaths in Hungary were attributed to the heat wave, while major and widespread wildfires destroyed large areas of forest across the region. Figure 6 (taken from Titley et al.) shows forecast probabilities of the mean temperature exceeding the 95th percentile (based on ERA-40 climatology) for 20–25 July. The probabilities are calculated from three of the TIGGE models (Met Office, ECMWF, and NCEP), and from a multimodel ensemble composed of the same three models. At the longest lead time (10–15 days ahead), the Met Office ensemble gives a good indication of the affected area. This is supported by ECMWF and, to a lesser extent, NCEP. The multimodel ensemble combines these probabilities and shows a significant risk of heat wave through most of the affected area. As the lead time reduces, the individual forecasts generally home in better on the area. By 19 July, the Met Office forecasts shows a 100% probability of exceeding the 95th percentile for most of the affected area, supported in part of the area by ECMWF, NCEP, and the resulting multimodel ensemble.

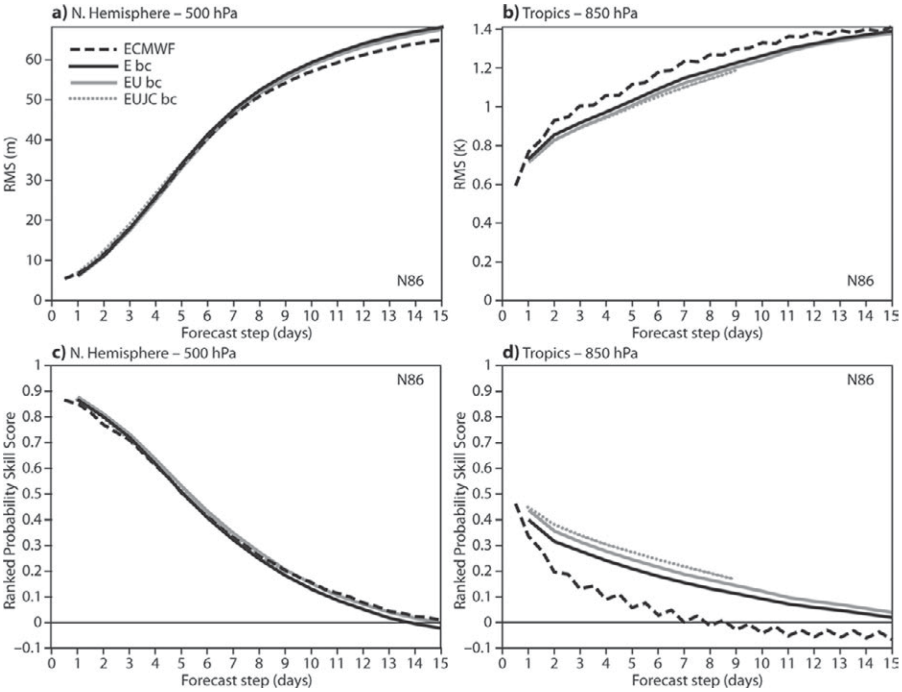

Fig. 4. RMSE of the ensemble mean and RPSS for four different ensemble systems, as a function of the forecast lead time (days): ECMWF alone and nonbias corrected (dashed), bias-corrected ECMWF (solid black), ECMWF + UKMO bias corrected (solid gray), ECMWF + UKMO + JMA + CMA bias corrected (dotted). Results on 86 cases are from Jun–Aug 2007.

Titley et al. (2008) carried out a series of case studies in which they compared the forecasts of several high-impact weather events, based on diagnostics from different ensemble prediction systems. As illustrated by the July 2005 heat wave, having access to different ensemble forecasts was valuable at both the short and medium ranges. There is value in the multimodel ensemble approach, both in cases where there is agreement between models (increasing confidence in the forecast) and where there are significant differences (giving a better representation of uncertainties). Different case studies had a different "best" model. There were several cases where a significant signal of the high-impact weather was forecast well into week 2 of the forecast, justifying running the ensemble forecast models out to 15 days.

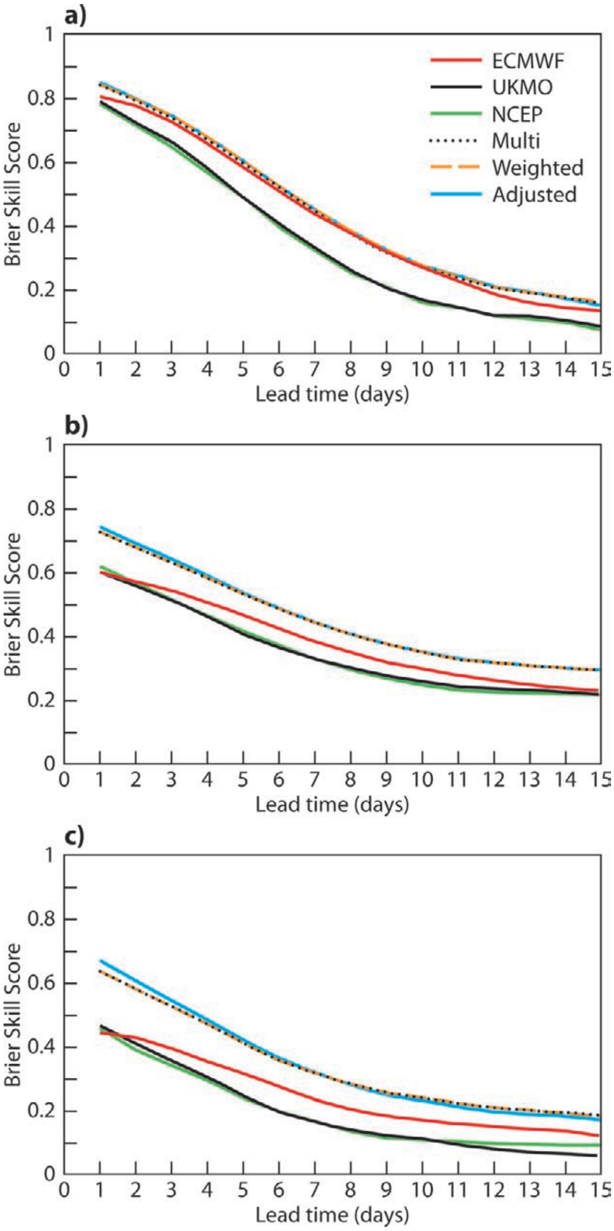

Fig. 5. Brier skill scores for (a) mean sea level pressure greater than the climatological mean, (b) 2-m temperature greater than the climatological mean, and (c) 2-m temperature greater than 90th percentile. In addition to the individual systems (ECMWF, UKMO, and NCEP), three almost equivalent variants of the multimodel system are shown (multiple, weighted, and adjusted). The data are globally averaged over 120 days, ending on 29 Apr 2008. [From Johnson and Swinbank (2009).]

In summary, TIGGE has shown promising results regarding improvement of the 2-m-level temperature forecasts, especially in the case of heat waves. Results for all parameters in the tropics also appear quite promising. In contrast, forecasts of 500-hPa height and sea level pressure in the midlatitudes seem to benefit less from the multimodel approach. One possible explanation is that large-scale, midtropospheric dynamical fields are generally consistently predicted by current NWP models. There is less consistency among models for near-surface variables, because these forecasts are more dependent on details of physical parameterizations and are thus affected by different model biases. The results are also consistent with the notion that benefits from multimodel combination are more significant when ensembles with comparable skill are combined, while the benefits are less clear when poorer-performing ensembles are added to a better-performing system. The verification statistics do seem to be sensitive to the verification data and climate reference data. Although we have only shown examples of one type of score from each study, all studies showed clearer benefits of multimodel ensembles for probabilistic scores than for deterministic scores. More work is needed to confirm the above conclusions on longer time series and by direct comparison to observations. There is also an urgent need to explore the forecast skill for other parameters, such as 10-m winds, rainfall, and clouds. Above all it is necessary to explore the impact of multimodel systems on severe weather forecasts more actively. It is likely that the benefits of multimodel systems vary depending on the weather parameter, lead time, and user. They may also vary rapidly in time, resulting from variations in the quality of component systems. It is important to fully document these aspects because the cost of maintaining operational multimodel systems is likely to be significant, and must not exceed the benefits.

Applications of TIGGE. Beyond the derivation of probabilistic weather forecasts, ensembles have a wide variety of applications. They can be used in decision support systems to explore the sensitivity of user-relevant consequences of weather conditions.

For example, Pappenberger et al. (2008) applied both single- and multimodel ensembles to the prediction of a particular flood event in Romania in October 2007. Results reveal that, in this case, warnings could have been issued as early as 8 days before the event. A comparison of 5-day forecasts, shown in Fig. 7, illustrates the positive impact of the multimodel approach at this lead time. The subsequent forecasts provided increasing insight into the range of possible flood conditions. This case study illustrates the potential value of the TIGGE archive and the multimodel ensembles approach to raise preparedness and reduce the negative socioeconomic impact of floods. He et al. (2009) present another application of TIGGE ensemble forecasts to flood forecasting.

Finally, the TIGGE database is opening the possibility of more upstream studies on how various (including multimodel) systems treat some features of the atmosphere. For example, Froude (2010) investigated the representation of extratropical cyclones in medium-range forecasts present in the database. A notable result is that models generally underestimate the speed of propagation, although in different proportions. Champion (2008) compared the different methods used for defining initial perturbations. He found that these result in large differences in initial amplitude of the perturbations and subsequent growth rates. Significant differences were found even between systems using similar methods, which points to the different behavior of the data assimilation systems. In particular, he found that singular-vector-based methods create perturbations with a westward tilt with height at initial time, experiencing a rapid baroclinic growth. On the other hand, perturbations based on the ensemble transform Kalman filter method have no tilt with height initially and progress to having an eastward tilt with height, which is consistent with decay.

Those few examples are just meant to show how TIGGE-based research will help understand the behavior of the various current approaches to ensemble forecasting.

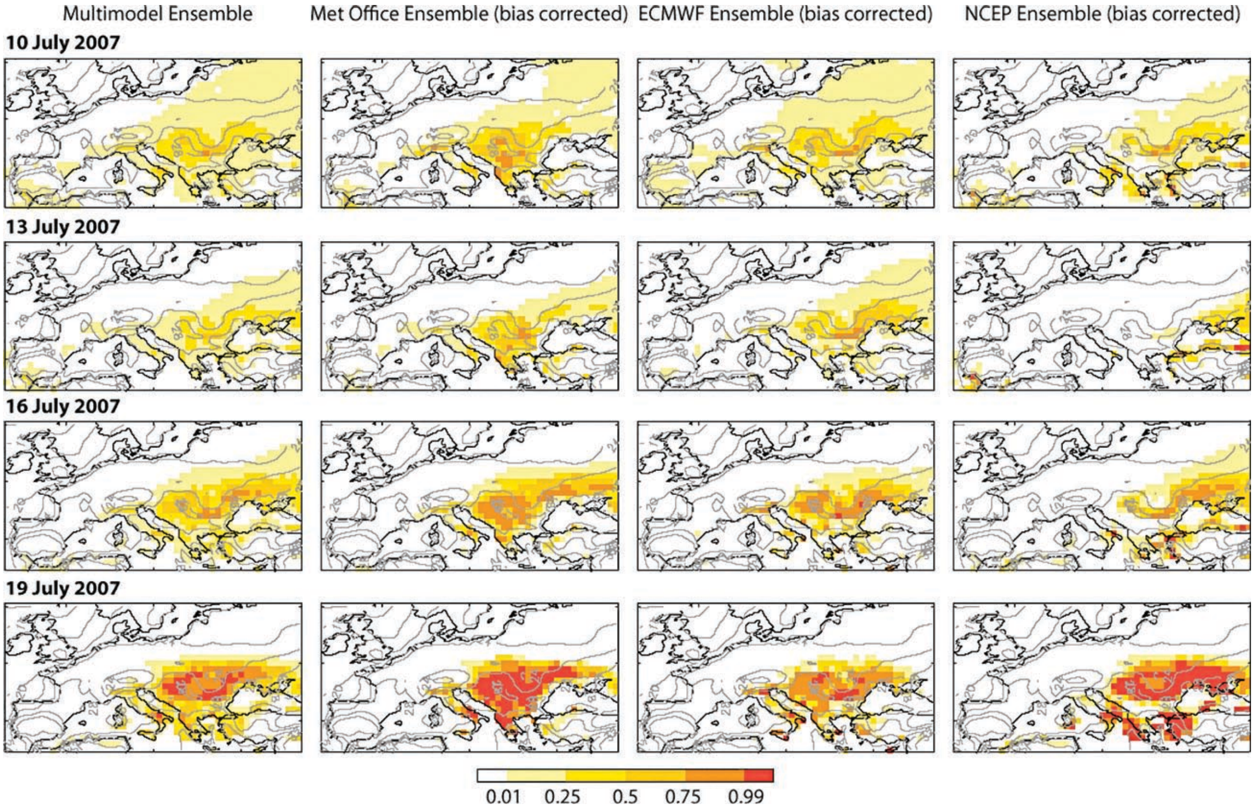

Fig. 6. Probability of mean temperatures, averaged over both 0000 and 1200 UTC 20–25 Jul 2007, that are greater than the 95th percentile of the ERA-40 climatology. The probabilities are calculated from a multimodel ensemble and its three component models (ECMWF, NCEP, and UKMO). The 95th percentile climatology data are overlaid in gray. Four sets of forecasts are shown with initial times (from top to bottom): 0000 UTC 10 Jul 2007 (averaged over 20–24 Jul, because the 25th is outside the 15-day forecast range), 13 Jul 2007, 16 Jul 2007, and 19 Jul 2007.

TOWARD THE FUTURE: TIGGE-LAM AND THE GLOBAL INTERACTIVE FORECASTING SYSTEM. Because of the large data volumes involved, an archive of the full forecast model output fields was not possible in TIGGE; consequently, the archive does not include all of the fields that are necessary for providing lateral boundary conditions to run limited-area models. More recently, an expert group (the TIGGE-LAM panel) was formed to coordinate the contribution of Limited Area Ensemble systems to TIGGE and, in a longer perspective, to the Global Interactive Forecast System (see below). Thus far the group has been focusing on the following three topics: i) creating a database of limited-area ensemble products, similar to the global TIGGE database; ii) making the various global and regional systems "interoperable"; and iii) relocating existing LAM EPS systems, already implemented and tested on specific regions, in other areas not covered by analogous forecasting systems. These activities will be planned and carried out in close cooperation with the WWRP Working Group on Mesoscale Weather Forecasting Research (WG-MWFR), with the WWRP/WGNE JWGFVR, with the local contact people and especially with the THORPEX Regional Committee representatives, who are in the right position to stress the relevant regional issues and to set priorities. (For more information on TIGGE-LAM, see www.smr.arpa.emr.it/tiggelam/.)

The GIFS is central to the THORPEX vision of accelerating the improvement of 1-day to 2-week forecasts, focusing on high-impact weather (see TIP 2005). The objective of the GIFS is the production of internationally coordinated advance warnings and forecasts for high-impact weather to mitigate the loss of life and property and to contribute to the welfare of all WMO nations, with a particular emphasis on the least-developed and developing countries. Ensemble predictions will play a critical role in assessing and mitigating weather- and climate-related risks by quantifying

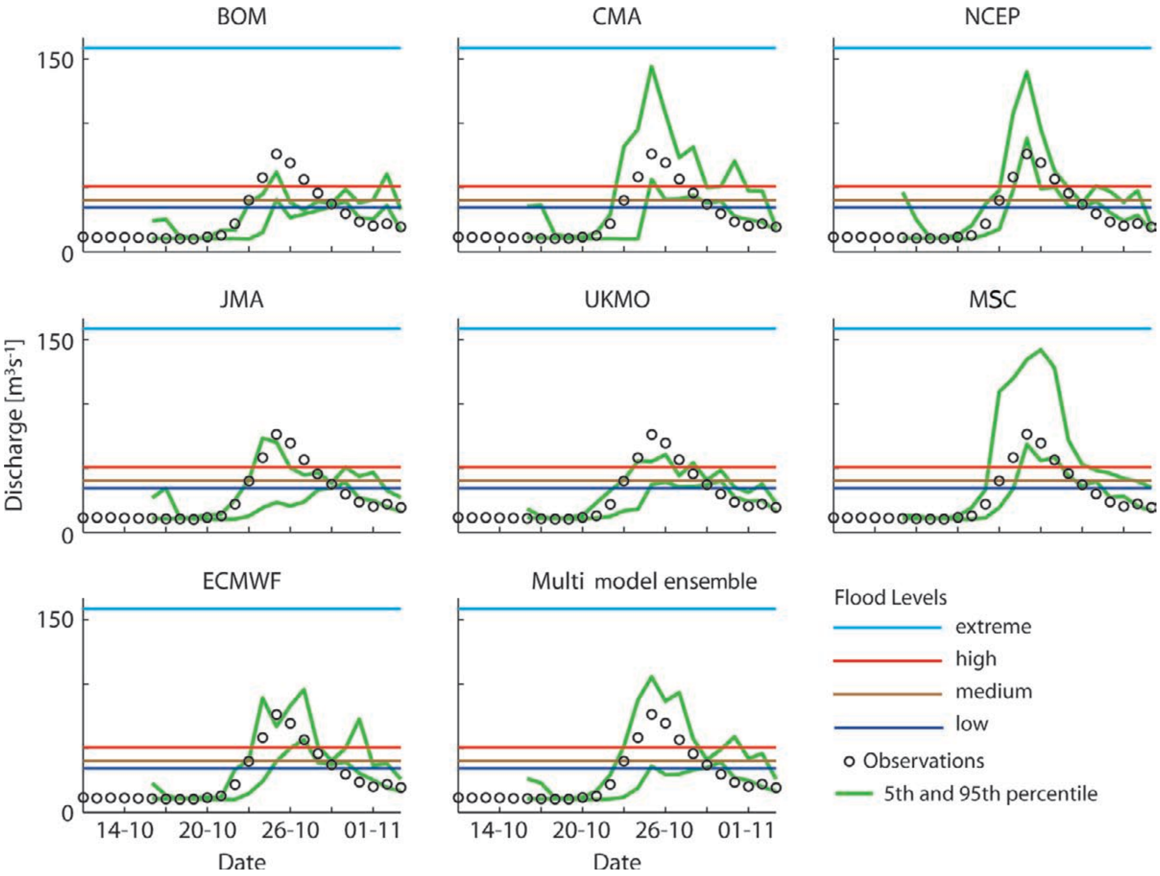

FIG. 7. Flow discharge that is "observed" and predicted by several TIGGE systems (called here systems I–VII) and one multimodel system (the grand ensemble) for a point on the river Jiu (in Romania) where flooding was observed. The 5th and 95th percentile of river discharge predictions are shown for the different forecasts with a 5-day lead time. The dashed horizontal lines show four classic flood-warning thresholds. Observed discharges in fact refer to simulations forced by observed rainfall. [From Pappenberger et al. (2008).]

forecast uncertainty. GIFS will be based on forecast products and services contributed voluntarily by NWP centers and other providers around the globe.

As its name indicates, the GIFS-TIGGE Working Group is in charge of developing concepts for the GIFS and fostering discussions with other THORPEX and WMO groups. The following issues have been identified:

- Science and applications: Additional research is strongly encouraged to further demonstrate the benefits of multimodel systems. The GIFS-TIGGE Working Group especially welcomes studies on high-impact weather and direct verification against observations. Demonstrations of applications of multimodel systems to, for example, hydrology, health, and civil protection are also strongly encouraged.
- Resource: Much hardware and manpower will be needed to develop reliable exchange mechanisms for real-time production. This requires advanced planning.
- Operational continuity: It will be a challenge to manage operational changes occurring at different times for the various component systems, to guarantee a smooth progress of the multimodel system skill, and to supply proper information on system upgrades to the users.
- Data policy: Several TIGGE providers will want to protect their commercial revenues from probabilistic forecasts. Negotiations will be needed to agree on a scheme that satisfies all partners.

As a way forward, the GIFS-TIGGE Working Group decided to develop pilot products that are clearly related to severe weather. In relation with the THORPEX Pacific Asian Regional Campaign (T-PARC) experiment of THORPEX, an exercise of real-time exchange of tropical cyclone tracks predicted by the various TIGGE systems has been defined and monitored by the Centre for Australian Weather and Climate Research (CAWCR; Australia). A special easy-to-read format for academic partners [the Cyclone Extensible Markup Language (CXML; XML) format; see Ebert et al. (2008)] was defined, and the TIGGE data providers were requested to provide tropical cyclone tracks on FTP sites in real time for the duration of the T-PARC experiment. These data will also be distributed by the TIGGE archive centers in addition to the usual TIGGE data.

Figure 8 shows an example of multimodel tropical cyclone tracks and strike probability charts generated from track data distributed in CXML format. This example takes data from only two ensembles (ECMWF and UKMO), but the technique can easily be extended to more. In this case there was a large overlap between the spreads of the two individual ensembles, but the ECMWF EPS showed a larger probability of a more southerly track, while the UKMO EPS gave a higher

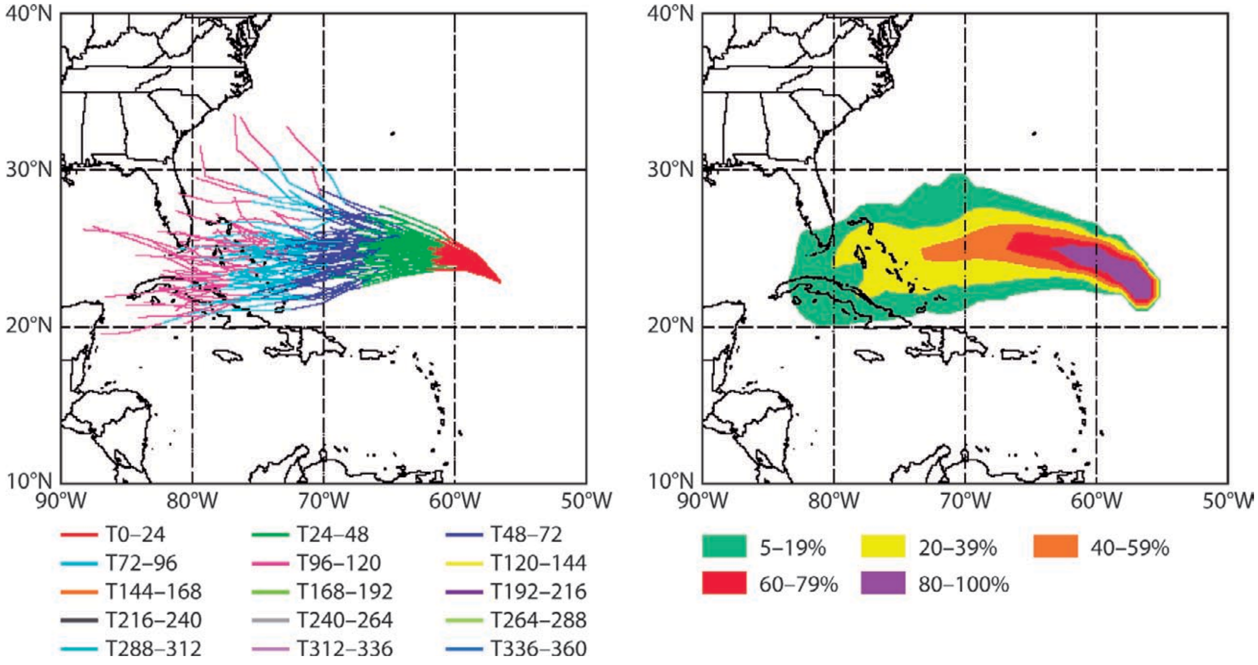

Fig. 8. Multimodel ensemble forecast tracks (left) and strike probabilities (right) for Hurricane Ike initiated at 1200 UTC 4 Sep 2008, combining outputs from the ECMWF and UKMO ensembles. These charts were generated at the UKMO using track data distributed using the CXML format.

probability to a more northerly track. Research continues into the optimal combination of ensembles in this way, for example, whether the contributions from individual ensembles should be weighted according to ensemble size or past performance.

CONCLUSIONS. The TIGGE project has attracted a high level of interest from both operational centers and the research community. TIGGE has already reached two key targets: first, it has led to the agreement of a data format to be used by all partners for exchanging forecasts, facilitating comparisons, and combining forecasts from different systems; second, it has let to an increased level of communication between the communities developing and using the ensemble forecasts. This will certainly promote the use of probabilistic forecasts.

We are convinced that the TIGGE databases will constitute a key resource for reaching the objective of THORPEX: the acceleration of the progress of the forecast skill for severe weather events from 1 day to 2 weeks ahead. This will be reached by a robust combination of research on the scientific basis of ensemble prediction, experimentation with new products, and development of new protocols and policies for data exchange across WMO Member States and across the science and application communities.

ACKNOWLEDGMENTS. We are indebted to Gordon Brent, Piers Buchanan, Roberto Buizza, O. Champion, Steve Chiswell, Luca Cinquini, Lizzie Froude, Renate Hagedorn, Christine Johnson, Honglian Lang, Mio Matsueda, Florian Pappenberger, Yves Pelletier, Louis Poulin, David Richardson, Kelvyn Robertson, Kiyo Sato, Peiliang Shi, Dave Stepaniak, Simon Thompson, Hao Tian, Helen Titley, Joerg Urban, Hannah Wilcox, Nathan Wilhelmi, Xin Yang, Shintaro Yokoi, Tom Yoksas, and many other colleagues from the data providers and archive centers for contributing to the TIGGE project and/or providing material to this paper.

REFERENCES

Champion, A., 2008: Properties of ensemble forecasts from different weather centres. M.S. thesis, University of Reading.

Ebert, B., Z. Toth, M. Charles, and G. Ross, 2008: Cyclone XML specification, version 1.0. Centre for Australian Weather and Climate Research, 6 pp. [Available online at www.cawcr.gov.au/bmrc/projects/THORPEX/CXML/CXMLspecification.doc.]

Froude, L. S. R., 2010: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting,* **25,** 819–836.

Group on Earth Observations, 2007: GEO 2007–2009 work plan toward convergence. Group on Earth Observations Work Plan 2007–2009, 30 pp. [Available online at www.earthobservations.org/documents/wp0709_v6.pdf.]

He, Y., F. Wetterhall, H. L. Cloke, F. Pappenberger, M. Wilson, K. Freer, and G. McGregor, 2009: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteor. Appl.,* **16,** 91–101.

Johnson, C., and R. Swinbank, 2009: Medium-range multimodel ensemble combination and calibration. *Quart. J. Roy. Meteor. Soc.,* **135,** 777–794.

Matsueda, M., and H. L. Tanaka, 2008: Can MCGE outperform the ECMWF ensemble? *Sci. Online Lett. Atmos.,* **4,** 77–80, doi:10.2151/sola.2008-020.

Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and A. de Roo, 2008: New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophys. Res. Lett.,* **35,** L10404, doi:10.1029/2008GL033837.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.,* **134,** 2029–2050.

Peña, M., and H. van den Dool, 2008: Consolidation of multimodel forecasts by ridge regression: Application to Pacific sea surface temperature. *J. Climate,* **21,** 6521–6538.

Raoult, B., and M. Fuentes, 2008: Implementation of TIGGE phase 1. *ECMWF Newsletter,* No. 116, ECMWF, Reading, United Kingdom, 10–16.

Richardson, D., R. Buizza, and R. Hagedorn, 2005: First Workshop on the THORPEX Interactive Grand Global Ensemble (TIGGE): Final report. World Meteorological Organization WMO/TD-No. 1273, WWRP/THORPEX No. 5, 39 pp. [Available online at http://tigge.ecmwf.int/references_pdf_files/TIGGE_report_WS1_WMO_TD1273_2005.pdf.]

TIP, 2005: THORPEX International Research Implementation Plan, version 1. World Meteorological Organization, WMO/TD-No. 1258, WWRP/THORPEX No. 4, 104 pp. [Available online at www.wmo.int/pages/prog/arep/wwrp/new/documents/CD_ROM_implementation_plan_v1.pdf.]

Titley, H., N. Savage, R. Swinbank, and S. Thompson, 2008: Comparison between Met Office and ECMWF medium-range ensemble forecast systems. Met Office Meteorology R&D Tech. Rep. 512.

Toth, Z., and Coauthors, 2005: The North American Ensemble Forecast System (NAEFS). *Proc. First THORPEX Int. Science Symp.,* Montreal, QC, Canada, WMO.