Bulletin of the American Meteorological Society

The THORPEX Interactive Grand Global Ensemble (TIGGE) and its Achievements

--Manuscript Draft--

Manuscript Number:	BAMS-D-13-00191
Full Title:	The THORPEX Interactive Grand Global Ensemble (TIGGE) and its Achievements
Article Type:	Article
Corresponding Author:	Richard Swinbank, PhD Met Office Exeter, UNITED KINGDOM
Corresponding Author's Institution:	Met Office
First Author:	Richard Swinbank, PhD
Order of Authors:	Richard Swinbank, PhD
	Masayuki Kyouda
	Piers Buchanan
	Lizzie Froude
	Thomas Hamill
	Tim Hewson
	Julia Keller
	Mio Matsueda
	John Methven
	Florian Pappenberger
	Michael Scheuerer
	Helen Titley
	Laurie Wilson
	Munehiko Yamaguchi
Manuscript Classifications:	2.232: Extratropical cyclones; 2.272: Hurricanes; 7.012: Ensembles; 7.016: Forecast verification/skill; 7.048: Probability forecasts/models/distribution
Abstract:	TIGGE is a major component of the THORPEX (The Observing System Research and Predictability Experiment) research program, whose aim is to accelerate improvements in forecasting high-impact weather. By providing ensemble prediction data from leading operational forecast centers, TIGGE has enhanced collaboration between the research and operational meteorological communities and enabled research studies on a wide range of topics. The paper covers the objective evaluation of the TIGGE data. For a range of forecast parameters, it is shown to be beneficial to combine ensembles from several data providers in a Multi-model Grand Ensemble. Alternative methods to correct systematic errors, including the use of reforecast data, are also discussed. TIGGE data have been used for a range of research studies on predictability and dynamical processes. Tropical cyclones are the most destructive weather systems in the world, and are a focus of multi-model ensemble research. Their extra-tropical transition also has a major impact on skill of mid-latitude forecasts. We also review how TIGGE has added to our understanding of the dynamics of extra-tropical cyclones and storm tracks. Although TIGGE is a research project, it has proved invaluable for the development of products for future operational forecasting. Examples include the forecasting of tropical cyclone tracks, heavy rainfall, strong winds, and flood prediction through

	coupling hydrological models to ensembles. Finally the paper considers the legacy of TIGGE. We discuss the priorities and key issues in predictability and ensemble forecasting, including the new opportunities of convective-scale ensembles, links with ensemble data assimilation methods, and extension of the range of useful forecast skill.
Author Comments:	
Suggested Reviewers:	David Parsons Director , University of Oklahoma dparsons@ou.edu Former Chief of World Weather Research at WMO
	Eugene Poolman South Africa Weather Service eugene.poolman@weathersa.co.za Leader in establishment of SWFDP in Souther Africa; experienced user of ensemble products
	Sharan Majumdar University of Miami smajumdar@rsmas.miami.edu Expert in ensemble methods & application for tropical cyclone forecasting
	Zoltan Toth NOAA Zoltan.Toth@noaa.gov Expert in ensemble forecasting & its applications (former co-chair of GIFS-TIGGE WG)
	Roberto Buizza ECMWF Roberto.Buizza@ecmwf.int Ensemble forecasting expert (not directly involved in TIGGE)

The THORPEX Interactive Grand Global Ensemble (TIGGE) and its Achievements

Richard Swinbank[1], Masayuki Kyouda[2], Piers Buchanan[1], Lizzie Froude[3], Thomas M. Hamill[4], Tim Hewson[5], Julia H. Keller[6], Mio Matsueda[7,8], John Methven[3], Florian Pappenberger[5, 9, 10], Michael Scheuerer[4], Helen Titley[1], Laurence Wilson[11,] Munehiko Yamaguchi[12].

[1] Met Office, Exeter, UK; [2] Japan Meteorological Agency, Tokyo, Japan; [3] University of Reading, UK; [4] NOAA ESRL/PSD, Boulder, Colorado; [5] ECMWF, Reading, UK; [6] Deutscher Wetterdienst, Offenbach, Germany; [7] University of Tsukuba, Japan; [8] University of Oxford, UK; [9] University of Bristol, UK; [10] Hohai University, Nanjing, China; [11] Environment Canada, Montreal, Canada; [12] Meteorological Research Institute, Tsukuba, Japan. .

Corresponding author:

Richard Swinbank

Met Office

FitzRoy Road,

Exeter EX1 3PB

United Kingdom

Email: richard.swinbank@metoffice.gov.uk

Capsule

The TIGGE project has made a rich dataset of ensemble predictions available for research. It has supported a wide range of scientific studies and new products for forecasting severe weather.

Abstract

TIGGE is a major component of the THORPEX (The Observing System Research and Predictability Experiment) research program, whose aim is to accelerate improvements in forecasting high-impact weather. By providing ensemble prediction data from leading operational forecast centers, TIGGE has enhanced collaboration between the research and operational meteorological communities and enabled research studies on a wide range of topics.

The paper covers the objective evaluation of the TIGGE data. For a range of forecast parameters, it is shown to be beneficial to combine ensembles from several data providers in a Multi-model Grand Ensemble. Alternative methods to correct systematic errors, including the use of reforecast data, are also discussed.

TIGGE data have been used for a range of research studies on predictability and dynamical processes. Tropical cyclones are the most destructive weather systems in the world, and are a focus of multi-model ensemble research. Their extra-tropical transition also has a major impact on skill of mid-latitude forecasts. We also review how TIGGE has added to our understanding of the dynamics of extra-tropical cyclones and storm tracks.

Although TIGGE is a research project, it has proved invaluable for the development of products for future operational forecasting. Examples include the forecasting of tropical cyclone tracks, heavy rainfall, strong winds, and flood prediction through coupling hydrological models to ensembles.

Finally the paper considers the legacy of TIGGE. We discuss the priorities and key issues in predictability and ensemble forecasting, including the new opportunities of

convective-scale ensembles, links with ensemble data assimilation methods, and extension of the range of useful forecast skill.

1. Introduction

THORPEX – The Observing System Research and Predictability Experiment – is a decade-long international research and development program to accelerate improvements in the accuracy and benefits of high-impact weather forecasts up to two weeks ahead (WMO, 2005a, b; Shapiro & Thorpe, 2004). THORPEX was established in 2003 as part of the WMO World Weather Research Program (WWRP). It has three major foci: predictability & dynamical processes; data assimilation & observing systems; and ensemble forecasting. These are reflected by the three WWRP-THORPEX working groups: PDP (Predictability and Dynamical Processes), DAOS (Data Assimilation and Observing Strategies) and GIFS-TIGGE (Global Interactive Forecasting System – THORPEX Interactive Grand Global Ensemble). The execution phase of THORPEX started in 2005, so the ten-year program is due to finish at the end of 2014. It is thus an opportune time to take stock of the achievements of the THORPEX program. This paper is focused on achievements related to the TIGGE project, whilst subsequent articles will cover the broader achievements of THORPEX.

A major part of the original THORPEX vision was the design and development of a "global interactive forecasting system" (GIFS) including the use of ensemble prediction systems that would be configured interactively in response to varying weather situations and user needs. TIGGE was developed as a resource to support research and development of the GIFS concept, as well to provide data for research on predictability, dynamics and impacts. Although on-demand ensemble predictions are not yet an operational reality, TIGGE has enhanced cooperation between the academic and operational meteorological communities by providing ready access to ensemble prediction data from leading operational forecast centers,

The TIGGE database contains ensemble predictions from ten global NWP centers, and is available via three archive centers, ECMWF, NCAR, and CMA. (See Table 1 for a list of TIGGE partners and their acronyms, as used in this paper, and Box 1 for information on accessing the data). Since the basis of TIGGE is to support research, and not operations, the technical set-up (Worley *et al*, 2008) is not designed to support real-time exchange of data. Instead, the data are made available to users 48 hours after the initial time of each forecast. A "TIGGE-LAM" panel has also been established to apply TIGGE concepts to limited area model ensembles. Several European regional ensembles are now available from a TIGGE-LAM archive established at ECMWF during 2014.

Due to the huge data volume, it was not feasible to include a full range of model fields at all levels in the TIGGE database; instead fields were selected taking into account user requirements discussed at a workshop hosted by ECMWF (Richardson *et al*, 2005). Documentation of the archived fields is available on the TIGGE project website, http://tigge.ecmwf.int and in Bougeault *et al.* (2010). The TIGGE data are stored in GRIB2 format, the standard established by WMO for the storage of gridded binary data that was designed to cater for ensembles. The TIGGE partners agreed a series of standards and conventions, to enable users to read forecast data from any of the TIGGE partners using the same computer code. The TIGGE data portals include links to tools contributed by TIGGE users, which are designed to help new users to read and plot the TIGGE data, including tools to convert the GRIB2 data to NetCDF format if required.

Since it was launched on 1st October 2006, the usage of the TIGGE archive has increased steadily. During calendar year 2013, there were at least 110 active users of the archive each month, and about 800 Terabytes of data were accessed from the database over the year.

Bougeault *et al* (2010) described some early results from TIGGE and pointed out that multi-model grand ensemble systems – combining predictions from several TIGGE models – have been demonstrated to give additional skill for some types of forecast parameters. Section 2 of this paper reviews the result of recent research on that topic, plus other studies evaluating the quality of the TIGGE forecasts.

TIGGE has opened up the opportunity for researchers to use the ensemble data for a wide range of studies, particularly on predictability and dynamical processes. At the time of writing, around 120 TIGGE-related papers have been published. Highlights of studies of dynamics and predictability of both mid-latitude and tropical systems are presented in Section 3. A wide range of information about TIGGE is displayed on the "TIGGE Museum" website (see Box 2), and several examples of graphical products from the website are used to illustrate this article.

Despite the fact that the TIGGE database was not designed to cater for real-time use, Section 4 shows that TIGGE has proved invaluable for the development of products to support forecasts and warnings of high-impact weather, as part of the vision for GIFS. The final section of the paper looks beyond the THORPEX program and explores how the achievements of TIGGE should be built on in the future.

2. Verification, combination, and calibration of TIGGE forecasts

2.1. Verification

The TIGGE database is designed to facilitate comparative verification of the ensembles contained therein, and many examples have been published. Figure 1, from the TIGGE Museum website, compares root-mean square errors of 500 hPa geopotential height for the Northern Hemisphere in winter 2013/14. The relative ordering of skill is typical of many other cases: ECMWF has lower errors than other centers, with tight competition for second place. Hamill (2012) found a somewhat similar relative ordering for precipitation over the contiguous US. In a more extreme case of error differences, Hagedorn *et al.* (2012, Fig. 3) showed that, in 2008-2009, the 2-meter temperature forecasts at the 1-week lead from the ECMWF system were similar in quality to several of the least skillful forecast systems at the 1-day lead. Of course, each system has been upgraded during the course of TIGGE, so these results will not necessarily reflect the precise relative or absolute performance of these systems at the current time.

Though it is preferable when available to verify against observations, analyses are often used instead to provide information on forecast quality that includes observation-sparse areas. Unfortunately, the relative performance of various modeling systems can depend strongly on which analysis is used for verification. For example, for low-level tropical regions, the model whose analysis was used as the verification field appeared to be the most accurate (see Park *et al.* 2008, Fig. 14). The *yearly mean* analyzed 2-meter temperature from five of the TIGGE systems was shown to vary by almost 5K between the warmest and coldest analyses for a location in the Amazon basin (Fig. 2). Large differences were also commonplace for some upper-air variables and for data at other locations. Given these differences, verifying against more than one analysis is preferable, for if a given model is unambiguously higher in skill than another regardless of which analysis was used, this lends credence to the result. Alternatively, a consensus of the more skilful analyses might be used.

2.2 Combination.

Probabilistic forecast skill and reliability can be improved through the combination of TIGGE data, i.e., the generation of a multi-model grand ensemble by combining raw ensemble predictions from multiple centers. As mentioned in Hagedorn *et al* (2005), "*the key to the success of the multi-model concept lies in combining independent and skilful models, each with its own strengths and weaknesses.*" Two underlying assumptions behind the success of combination are that: (1) the modeling systems may have independent (or nearly so) systematic errors, thus providing some benefit through cancellation, and (2) the modeling systems collectively may provide more realistic estimates of event probabilities than individually. Several studies have demonstrated such improvement, including Matsueda and Tanaka (2008), Park *et al.* (2008), Johnson and Swinbank (2009), Candille (2009), Hagedorn *et al.* (2012), and Hamill (2012). Large benefits have been found for quantities relevant to weather impacts such as surface air temperature, surface wind, and precipitation. Hamill (2012) showed that multi-model combination improved the overall skill and reliability of precipitation forecasts over the contiguous US; similar results are shown in Fig. 3. There are both practical and theoretical considerations that will affect how much benefit users derive from multi-model ensemble combination. Practically, the global ensemble prediction systems in TIGGE contain forecasts with differing qualities.

Hagedorn *et al.* (2012) showed that the combination of the four highest-performing ensembles led to forecasts that were statistically significantly better than the raw ensemble guidance from the best-performing system. However, when the combination included data from all available TIGGE systems there was no unambiguous statistical advantage, showing that some account needs to be taken of relative quality. More theoretically, as ensemble prediction systems are upgraded (e.g., higher resolution, improved initialization procedures, and improved parameterizations), the systematic biases in each center's mean forecast will decrease. Should the prediction systems also incorporate more sophisticated methods for simulating the model uncertainty, then their spread will also become more consistent with the ensemble-mean error, as expected by theory. In this (desirable) situation, the simple combination of ensemble prediction data would become less beneficial, aside from the reduction in sampling error from the use of a larger ensemble. There will also be less benefit of combining data from current generation ensemble prediction systems if the combination is attempted after some post-processing to remove bias and calibrate spread. Online appendix A provides further discussion of how ensembles might be combined in the presence of correlated errors.

2.3. Calibration

Whether in combination or alone, the information provided by ensemble prediction systems often requires some statistical post-processing to reduce systematic errors, as well as to deal with sampling error. TIGGE has provided a rich set of data that has enabled research on a range of potential methods for calibration of ensemble predictions, contributing to the large body of literature on the subject (see, for example, Joliffe and Stephenson, 2012). Which approach works best often depends on the variable in question; a post-processing method that works well with temperature is probably not optimal for precipitation, because of the different characteristics of their probability distributions.

The accuracy and reliability of post-processed guidance may depend on the amount of training data available, particularly for more uncommon, high-impact events such as heavy precipitation. How does one obtain a sufficiently large sample when forecast models are updated every year or so, which may change the model's error characteristics? An ideal method is to use a *reforecast* dataset, incorporating a large number of forecasts of past cases that have been re-run with the current NWP system. The advantage of using training samples from a reforecast dataset for calibration of surface temperature data is clearly shown by Hagedorn *et al.* (2012), although the results for precipitation from Hamill (2012) are less clear cut. Ideally the retrospective forecasts will have the same error characteristics as the operational model. Should the forecast modeling system change significantly, a new reforecast dataset should be generated. Because of the computational expense, many centers seek to provide statistically adjusted guidance using shorter training data sets, such as the 30-day training period used in part for the calibration in Hagedorn *et al.* (2012) and the 40-day training period used in Wilson *et al.* (2007). Shorter training periods have been shown to produce acceptable results for shorter-range forecasts of variables such as surface temperature, but larger sample sizes are increasingly valuable for longer-range forecasts and for forecasts of more rare events such as heavy precipitation.

3. Dynamics and Predictability

3.1 Extratropical cyclones and storm tracks

Extratropical cyclones, and the associated baroclinic waves, are the primary cause of variability in weather across the mid-latitudes. Mesoscale features embedded within cyclones, such as fronts, can bring both damaging surface winds and heavy precipitation leading to impacts such as widespread flooding.

The regions where extratropical cyclones frequently occur are often called *storm tracks*: the most prominent storm tracks in the Northern Hemisphere span the Atlantic and Pacific Oceans. The heat and moisture fluxes associated with cyclones dominate the poleward transport of energy in the atmosphere and therefore have a crucial influence on climate. Using TIGGE data, individual cyclones were tracked and systematic errors diagnosed for the global ensemble forecasts from the ten centers (Froude 2010, 2011; Fig. 4). This methodology has revealed valuable information about the representation of cyclones in numerical weather prediction models, and their lower resolution cousins – climate models. The ECMWF ensemble was found to have the highest level of performance in predicting cyclone position, intensity and propagation speed. However, there may be some bias as all the ensembles were verified against the ECMWF analysis (as discussed in Section 2.1). Figure 4a also shows that the intensity of the cyclones was not predicted so well by the ensembles with lower spatial resolutions (NCEP, BoM and CPTEC), perhaps indicating some systematic errors in simulating the contraction and intensification of ascent into narrow regions as a result of latent heat release. An intriguing, but as yet unexplained forecast error, is that all the EPS were found to under-predict the propagation speed of cyclones (Fig. 4b). Froude (2011) also assessed this bias in the ECMWF high resolution forecast and the bias was found to be significantly smaller than the lower resolution EPS.

Ensembles of cyclone tracks can be displayed to illustrate uncertainty. This is illustrated for T+72 h forecasts in Fig. 5 for the high-impact St Jude's Storm case on 28 Oct 2013 (see also Hewson *et al*, 2014). This intense cyclone caused a trail of severe damage across highly populated areas including southeast England, the Netherlands and Denmark. Both the Met Office and ECMWF run the Hewson and Titley (2010) cyclonic feature identification and tracking methodology on their global ensembles and the results are used by operational forecasters. The cyclonic features are detected using a combination of vorticity maxima and pressure minima. In Fig 5(a, b), the dots locate the centers of cyclonic features with intensities indicated by the colors. The scatter provides an immediate visual impression of the uncertainty in feature locations represented by the ensembles. The analyzed storm center reached Denmark at about 1400 UTC. Approximately half the Met Office ensemble clustered toward Denmark,, the other solutions showing the cyclone nearer the UK. In contrast, all members of the ECMWF ensemble predicted the cyclone to move too slowly. The feature points and the associated values were used to create forecasts of strike probability in Fig. 5(c, d). They are somewhat analogous to the 'cone of uncertainty' plots employed in hurricane forecasts (e.g., Majumdar and Finnochio 2010). The marked difference between the ECMWF and Met Office probability forecasts illustrates that ensemble forecasting systems are not perfect and more research is required to transform ensemble predictions into accurate probability forecasts for weather events.

3.2 Jet stream variability: Large-scale flow regimes and blocking

TIGGE has facilitated studies of large-scale, low frequency variations in the jet stream. The jet stream is characterized by very large-scale meanders and the phenomenon of Rossby wave breaking. Low frequency variability is dominated by a few large-scale patterns (e.g., Cassou et al, 2004 identified four in the Euro-Atlantic sector). It might be anticipated that low resolution models would be able to simulate such patterns. However, Dawson *et al* (2012) showed that lower resolution (T159) simulations fail to capture the observed variability, while the free-running ECMWF model at the resolution of the ECMWF EPS (T511) captures the structure and variance of the large-scale patterns over the Atlantic. Doubling the resolution again to T1279 obtains similar results, indicating convergence in the ECMWF model representation of low frequency dynamics.

The TIGGE database was used by Matsueda (2009) to show that ensemble forecasts perform well in simulating the frequencies of Euro-Atlantic (EA) and Pacific (PA) blocking, even after a lead time of 9 days. However, probabilistic forecasts of blocking over the PA sector were more skilful than those for the EA sector. Frame *et al* (2011) took a different approach in quantifying the skill in the prediction of the probability of transition between 3 states of the North Atlantic jet stream (South, Mid and North). They showed that forecast centers (ECMWF, CMC, UKMO) exhibited consistent *flow-dependent predictability*: predictive skill is greatest when the jet is in the south state, linked to greater persistence of that state. Ensemble forecasts diverge most rapidly passing through the north jet state. The sensitivity to initial conditions, like the “butterfly” of the Lorenz model, is associated in this case with Rossby-wave breaking and split jet formation.

Patterns associated with persistent behavior have a major influence on regional weather extremes and their impacts. Matsueda (2011) used TIGGE to investigate the predictability of surface temperature in Eurasian blocking events such as the Russian heat-wave of 2010 (Dole *et al,* 2011). While the blocking in June-August of 2010 was predictable on average, even for a lead time of 9 days, there was little skill beyond 6 days in predicting the particular blocking event that brought the severe heat wave (30 July – 9 August). Most of the forecasts predicted a decay of the blocking earlier than that observed. At the same time a trough over Pakistan, downstream of the Russian blocking anticyclone, in conjunction with a monsoon depression brought extreme precipitation and flooding to Northwest Pakistan (e.g., Galarneau *et al,* 2012). A key lesson from this case study is that simultaneous extreme events can be linked via Rossby waves, but have differing predictability.

Gray *et al* (2014) have used TIGGE forecasts to quantify systematic errors in the representation of Rossby waves on the jet stream using diagnostics that were not sensitive to longitudinal phase displacements of waves: namely the total area occupied by ridges and the average horizontal potential vorticity (PV) gradient across the tropopause. Both ridge area and PV gradient decrease with lead time. None of the models can maintain a gradient as tight as the observed in the face of numerical dissipation, implying that the jet stream is weaker than observed. The decrease in ridge area points to a decline in wave activity in the forecasts. This may be because overly smooth PV gradients resulted in faster dispersion of Rossby wave activity, or because incorrect representations of the diabatic processes resulted in a loss of amplitude. Further dynamics research is required to identify the processes responsible for these systematic errors and their consequences for weather events downstream.

3.3 Madden-Julian Oscillation

The Madden-Julian Oscillation (MJO) is the dominant mode of intraseasonal variability in the tropics, and influences tropical weather and extratropical circulations via large-scale teleconnections. There is only a partial understanding the dynamics of the MJO and its interaction with convective processes and the surface layers of the ocean, and its prediction remains a major challenge. Although the forecast range of the TIGGE ensembles is shorter than the period of the MJO, the TIGGE data allows a good comparison of the MJO forecasts over about half a cycle of the oscillation.

Matsueda and Endo (2011) assessed the MJO forecast performance of operational medium-range ensemble forecasts by using the TIGGE data for the period of 1st January 2008 – 31st December 2010 (see the example forecast comparison in Fig. 6). Wheeler and Hendon (2004) defined a bivariate index of the amplitude and phase of the MJO which provides a convenient framework for evaluating the forecasts. Matuseda and Endo (2011) found that ECMWF and Met Office generally yield the best performances in predicting the MJO; however, they do not always show similar skill. ECMWF performs well in simulating the maintenance and onset of the MJO in phases 1 – 4 (where the region of enhanced convection progresses from east Africa, across the Indian Ocean to the Maritime Continent), whereas Met Office and NCEP perform well in phases 5 – 8 (where the enhanced convection progresses from the Maritime Continent across the Pacific and on to Africa). They also found that simulations of the MJO generally show a slower phase speed and a larger amplitude than that observed. Predicted amplitude over the Maritime Continent (phase 4 and 5), however, tends to be smaller than that observed, suggesting that most models still face the Maritime Continent predictability barrier (Seo *et al*, 2009). The quasi-real time MJO forecasts based on TIGGE data are available via the TIGGE museum (Box 2).

3.4 Tropical cyclones

Tropical cyclones (TCs) are one of the most destructive atmospheric disturbances on Earth and pose the greatest threat to life and property (King et al, 2010). Establishing effective warning systems and strengthening international cooperative frameworks are of fundamental importance for disaster risk reduction of TCs. This need is addressed both by improving the underlying TC predictions (discussed in this section) and by developing new informative forecast products (Section 4).

One of the great benefits of TIGGE is that it is now feasible to create and evaluate a multi-model grand ensemble of TC predictions (e.g., Majumdar and Finocchio 2010; Yamaguchi *et al*, 2012; Matsueda and Nakazawa 2014). Yamaguchi et al. (2012) demonstrated the objective statistical benefits of track forecasts based on a multi-model grand ensemble compared to a single-model ensemble for the western North Pacific basin. However, Majumdar and Finocchio (2010) pointed out that there are some circumstances where combination of ensembles does not improve track forecast skill. On most occasions the observed track should be well within the spread of forecast tracks, but, as shown in Fig. 7, there will be some occasions when the actual track falls on the edge of the forecast ensemble.

It is sometimes necessary to forecast the most likely TC track; in general this will be given by the ensemble mean track, but Qi *et al* (2013) and Tsai and Elsberry (2013) have developed some more sophisticated approaches. Tsai and Elsberry (2013) showed that, in situations where there was a track bifurcation (two clusters of forecast tracks), the track cluster with a percentage greater than 70% can be reliably selected as the better choice. For situations when later observations are available, Qi *et al* (2013) developed an approach by which larger weight are given to ensemble members that are closer to the observed TC locations.

For probabilistic predictions of TCs, it is important that the ensemble initial perturbations are a realistic representation of the uncertainties in the initial conditions. TIGGE has helped analyze and interpret the initial perturbations and their impact on TC forecasts (e.g., Hamill *et al*, 2011, Magnusson et al, 2014). Yamaguchi and Majumdar (2010) demonstrated that singular vector-based perturbations grow through a baroclinic energy conversion in a vortex, which amplifies the ensemble spread of TC tracks. TIGGE has also contributed to the analysis of the sensitivity of forecasts to initial condition perturbations, which can be used for the targeting of observations to improve TC forecasts (e.g., Majumdar *et al*, 2011).

TIGGE has facilitated studies on understanding TC dynamics and their prediction across TC basins worldwide. Majumdar and Torn (2014) showed that ensembles have potential for probabilistic prediction out to 5 days. Although the reliable prediction of TC formation is in its infancy, studies using TIGGE data demonstrate skill in predicting formation using multi-model grand ensembles (e.g., Belanger *et al*, 2012; Halperin *et al*, 2013). Given that TC intensity changes and genesis events are often affected by environmental influences such as wind vertical shear and tropical waves (e.g., Kepert 2010; Tory and Frank 2010), even relatively low resolution ensemble data could be beneficial.

3.5 Extratropical transition of tropical cyclones

TCs can also have a profound effect on the synoptic evolution in mid-latitudes. A poleward moving TC interacts with the mid-latitude Rossby wave guide and may undergo extratropical transition (ET), transforming from a tropical into an extra-tropical cyclone (Jones et al. 2003). The outflow and circulation of the TC may amplify or even trigger the development of a mid-latitude Rossby wave train, leading to the potential for high-impact weather in regions downstream of the TC itself. The difficulties in representing ET often leads to a decrease in forecast skill, which can be investigated using ensemble forecasts, as demonstrated by Harr *et al* (2008) and Anwender *et al* (2008).

TIGGE has opened up the possibility of using a range of ensembles to address the impact of transitioning TCs on predictability in downstream regions. Keller *et al* (2011) showed that TIGGE offers a broader range of possible forecast scenarios for ET events and the downstream impact than an ensemble generated by a single forecasting system. Whether these additional scenarios provide a reasonable representation of the uncertainty of the actual development requires further investigation. In a dynamical study using TIGGE data, Archambault et al (2014) investigated the role of transitioning TC Malakas on the amplification of a mid-latitude wave train, and the consequent high-impact weather over North America. Both studies highlight the use of TIGGE to further advance our knowledge of ET events and their impact on predictability.

4. Applications for the Forecast User Community

4.1 Tropical Cyclone Forecasting

During the THORPEX Pacific Asian Regional Campaign (T-PARC), several TIGGE partners started to exchange tropical cyclone track predictions in near real-time, using an XML (extensible markup language) based format that was developed for the purpose (Cyclone XML or CXML format, see http://www.bom.gov.au/cyclone/cxmlinfo/). Ensemble forecast products based on the CXML data proved invaluable for the North Western Pacific Tropical Cyclone Ensemble Forecast Project (NWP-TCEFP) which was launched in 2009. During TCEFP, the ensembles were utilized by forecasters from the ESCAP/WMO Typhoon Committee, and also the south-east Asia region of the WMO Severe Weather Forecast Demonstration Project (SWFDP, see Section 4.2 below).

Although TC track predictions have become significantly more accurate over the past few decades, there is room for improvement in quantifying and communicating uncertainty in the forecasts (e.g., Heming and Goerss, 2010). As discussed in Section 3.4, multi-model grand ensembles generally give objectively more skillful forecasts than single-model ensembles. These new TC products provide forecasters with additional information by summarizing the forecast uncertainty from the grand ensemble, and so increase the level of confidence in the forecasts.

Some examples of multi-model ensemble products are shown for the forecasts of hurricane Sandy in Fig. 8. Sandy developed in the Caribbean Sea, and was declared a hurricane on the 24th October 2012. During its lifetime Sandy underwent a complex evolution, making landfall in Jamaica, Cuba and the Bahamas. After tracking over the Atlantic, Sandy turned westward and made landfall unusually far north, near Atlantic City, New Jersey at 00UTC on the 30th October 2012, with sustained winds of 80 mph and a central pressure of 945hPa. Because of its huge size, Sandy caused a storm surge along the entire east coast, but particularly in New York and New Jersey – leading to around $50 billion damage and at least 147 fatalities. The NHC produced a comprehensive report on Sandy and its impact (Baker *et al,* 2013), while Magnusson *et al,* (2014) investigated the skill of medium-range forecasts of Sandy.

Figure 8 shows 5-day forecasts of strike probability, individual track and ensemble mean track based on three ensembles (ECMWF, NCEP and Met Office), giving an early warning of the landfall. These plots are produced from 96 equally-weighted ensemble members. In this case, the actual track of the storm sits within the areas of highest probability in the strike probability. The ensemble-mean tracks (right-hand side) are plotted for each individual center and the consensus of the 3 centers.

4.2 Early warning products

Using TIGGE data, Matsueda and Nakazawa (2014) have developed a prototype suite of ensemble-based early warning products for severe weather events, using both single-model (ECMWF, JMA, NCEP, and Met Office) and multi-model grand ensembles. These products estimate the forecast probability of the occurrence of heavy rainfall, strong winds and severe high/low temperatures, based on each model’s climatology, i.e., using information from the climatological probability density function to determine appropriate thresholds for severe weather events. The products are now routinely available as part of the TIGGE Museum.

Objective verification of these products demonstrates that the construction of multi-model grand ensembles by combining four single-model ensembles can improve the skill of probabilistic forecasts of severe events (Matsueda and Nakazawa, 2014). The grand ensemble provides more reliable forecasts than single-model ensembles for all lead times, although the grand ensemble is still overconfident, especially for lead times greater than 216 hours.

An example of this type of forecast product is shown in Fig. 9, for a heavy precipitation event in West Africa on 1st September 2009 that caused severe flooding in Ouagadougou, Burkina Faso. In all ensembles, there is an indication of the risk of heavy rainfall over West Africa 4-5 days ahead of the event. However,the location of the peak rainfall in Burkino Faso was captured only 2-3 days ahead (not shown). Mesoscale convective systems (MCSs), which lead to such events, are not well predicted by the current ensemble systems. The multi-model ensemble produces a smoother probability map, suggesting that the main benefit of combination for this region is achieved by increasing the ensemble size. Using TIGGE data, Hopsch *et al* (2014) showed that the link between large-scale circulation and MCSs could potentially be exploited to improve their prediction.

Since the skill of these TIGGE forecast products has been demonstrated, there is a strong incentive to implement them in real time, avoiding the 2-day delay in accessing data from the TIGGE archive. A system is currently being set up to supply these early warning products to the WMO SWFDP forecasters in real time.

The SWFDP (http://www.wmo.int/pages/prog/www/swfdp/) enables countries in some of the less developed regions of the world to benefit from state of the art numerical model predictions. The global NWP centers generate graphical products that are tailored to support regional SWFDP initiatives. The current SWFDP products will be supplemented both by the ensemble-based early warning products and the multi-model tropical cyclone products developed using TIGGE. Designated regional forecast centers disseminate these products, and associated forecast guidance, to neighboring national meteorological services. The first region to be covered by SWFDP was southern Africa, with Pretoria as the primary regional center; the project has since been extended to cover the South Pacific islands and, more recently, east Africa and south-east Asia

5. Discussion and future prospects

The TIGGE project has provided a valuable dataset to facilitate research on ensemble techniques, including demonstrating the benefit of combining predictions from several ensemble prediction systems – this conclusion also carries through to hydrological applications (see Box 3). Although combination has proved a pragmatic approach to improving probabilistic forecast skill, we expect less benefit from the technique in the future, as systematic errors in ensembles are reduced. TIGGE has also supported a wide range of research and on dynamics, the fundamental nature of predictability and development of forecast applications.

In view of TIGGE's success, it has been agreed that the project should continue for five further years beyond the completion of the THORPEX research program at the end of 2014. (Any extension beyond 2019 will be considered nearer the time.) The great majority of TIGGE partners will continue to participate, and provide ensemble predictions for use by the research community. Both ECMWF and CMA will continue to host TIGGE archive centers. To reflect the completion of THORPEX, it is planned to change the name of TIGGE to "The International Grand Global Ensemble"

Building on the success of THORPEX, the WWRP is establishing three THORPEX legacy projects: the Subseasonal to Seasonal Prediction (S2S) and Polar Prediction (PPP) projects are already underway, while the High-Impact Weather (HI Weather) has been approved by the WMO Executive Council, and will start in 2015. S2S explores the longer range prediction problem, when the interactions between the atmosphere and other elements of the earth system, especially oceans, are increasingly important. S2S is a joint initiative between WCRP (World Climate Research Program) and WWRP. The project will be underpinned by establishing an S2S database, which is expected to go live in 2015. The S2S database will be based closely on the TIGGE database, using similar data formats and conventions – which should facilitate research on seamless predictions ranging from 1 day to 2 months ahead, using both datasets. PPP is concerned with the prediction of weather in high latitudes and its link with lower latitudes (e.g., Jung and Matsueda, 2014). The main focus of PPP will be preparing for, coordinating and analyzing results from the “Year of Polar Prediction” (YOPP), a combined modeling and field campaign which is planned to take place between mid-2017 and mid-2019. TIGGE will play a key role in providing ensemble prediction data for PPP, and some enhancements to TIGGE may be implemented to support the requirements of PPP or other WWRP projects. The HI Weather project addresses the improvement of forecasts and warnings of high-impact weather, with a focus on five hazard areas: urban flooding, localized extreme wind, wildfire, urban heat & air quality, and disruptive winter weather. A key aspect of the project will be understanding vulnerability and risk, and improving the communication of warnings of high-impact weather.

Looking forward, increases in computer performance allow short-range convective-scale ensemble forecasts that will be a major step forward to the prediction of details of hazardous weather. Currently both DWD (Deutscher Wetterdienst) and the Met Office are running operational ensemble systems with around a 2-km grid, Météo-France and other centers also have high-resolution systems under development. The US Hazardous Weather Testbed project has been running for more than a decade, comparing experimental ensemble forecasts run at 4km resolution across the central

USA (Clark *et al,* 2012). The recent establishment of the European TIGGE-LAM archive means that forecasts from high-resolution ensembles will be more readily available to the research community, and it is hoped that similar facilities will, in future, be developed in other continents. These datasets will provide invaluable data to underpin focus on improving the detailed prediction of high-impact weather events at short timescales, and should prove especially valuable for the HI Weather project..

Ensemble methods are also being increasingly employed in data assimilation, in both purely ensemble approaches to data assimilation (e.g., Houtekamer et al, 2005) and hybrid ensemble-variational methods (e.g., Clayton et al, 2013). An ensemble of model states provides a good framework to specify the relationship between uncertainties in model variables, i.e., well-specified ensemble perturbations should be closely related to the background error covariance information that is used for data assimilation. A very large ensemble (over 100 members) is needed in order to satisfactorily represent the error covariance information in an ensemble data assimilation system, while hybrid techniques permit the use of fewer ensemble members, by combining flow-dependent information from an ensemble with static climatological error covariances.

A new WWRP working group on Predictability, Dynamics and Ensemble Forecasting (PDEF) will address the theoretical basis of ensemble forecasting and its relation to the dynamics of the atmospheric phenomena and coupled systems. Research with TIGGE has highlighted some key aspects of flow-dependent predictability on the large-scale and connections with high-impact weather events. The new generation of convective-scale ensembles raises many important issues: the suitability of data assimilation approaches developed for the synoptic scale, the construction of ensembles, the role of stochastic parameterization in representing model uncertainty and the fundamental nature of predictability itself on finer spatial and temporal scales. The PDEF working group will be scientifically responsible for the development of the TIGGE and TIGGE-LAM datasets, to promote and support ongoing scientific research and especially the WWRP projects. PDEF will bring dynamical expertise from the academic community to bear on these exciting new challenges, ultimately driving towards improved probabilistic prediction.

In recent years there has been a rapid growth in the utilization of probabilistic forecasts by both industry and government organizations to manage risks. The TIGGE project has been at the forefront of these developments, making a major contribution to the development of ensemble methods to provide these risk-based forecasts. The multi-year ensemble forecast dataset has been an unparalleled resource to the applied science research community. TIGGE has also provided a rich seam of data which has been used for a range of studies, covering research on atmospheric dynamics, improvement of predictive skill of models and development of ensemble techniques. Ensemble techniques are increasingly important for prediction at both short space and timescales, extending the limits of predictability and data assimilation. Looking forward, we expect TIGGE and TIGGE-LAM to support a range of exciting developments, underpinning further improvements to the use of ensemble techniques in both data assimilation and prediction, and also the developments of a rich collection of risk-based forecasting applications.

Acknowledgements

We are indebted to Steve Worley, Doug Schuster, Baudouin Raoult, Manuel Fuentes and their colleagues for setting up and managing the TIGGE archives at ECMWF, NCAR and CMA. The TIGGE project would not have been possible without support from the TIGGE partnership: the 3 archive centers, the 10 data providers, the National Climate Data Center, and the past and present members of the THORPEX GIFS-TIGGE working group. We are grateful for funding from the European Commission via the GEOWOW project (grant agreement no. 282915), which supported enhancements to the ECMWF TIGGE archive, and some of the development and evaluation of severe weather forecast products.

Box 1 – Accessing TIGGE data

An overview of TIGGE, with links to further information and documentation is given on the website http://tigge.ecmwf.int/

The TIGGE data are available from the following portals:

ECMWF http://apps.ecmwf.int/datasets/

CMA http://wisportal.cma.gov.cn/wis/

NCAR http://tigge.ucar.edu/ (until end of 2014)

The TIGGE-LAM archive enables researchers to have access forecasts from several European regional ensemble prediction systems. The forecasts are produced at high resolution (between 12km and 2km grid spacing) and provide detailed forecasts up to a few days head. TIGGE-LAM data are available via the ECMWF portal shown above.

Box 2 – The TIGGE Museum

The "TIGGE Museum" website was established by Mio Matsueda, with Tetsuo Nakazawa. The website is currently hosted by the University of Tsukuba, at http://gpvjma.ccs.hpcc.jp/TIGGE/. It displays a variety of graphical information based on the TIGGE dataset, including:

- Statistical verification of TIGGE forecasts;
- Ensemble-based forecasts of severe weather;
- Forecasts of the Madden-Julian Oscillation and blocking;
- Sample scripts to show how to download and plot TIGGE data.

The TIGGE Museum products are regularly updated with a 2-3 day delay, and are available for non-commercial use.

Box 3 - Hydrological forecasting

Hydrological models act as non-linear filters and integrators of rainfall predictions. They are therefore ideal for understanding the impact of deficiencies in the ensemble forecasts for downstream applications.

TIGGE was first used for hydro-meteorological forecasting when Pappenberger et al. (2008) demonstrated the potential of grand ensembles for early flood warning, applying the European Flood Awareness System (EFAS, Thielen *et al*, 2009) to a hindcasted flood event in Romania. Figure 10, from this study, shows forecasts of river level for a point on the river Jiu which was severely flooded in October 2007, based on 7 single-model ensembles and a multi-model grand ensemble. While all the ensembles predict the onset of the rising river level correctly, only 2 single-model ensembles and the multi-model ensemble correctly bracket the flood peak. The conclusion of the study was that, if grand ensemble forecasts had been used, flood warnings could have been issued 8 days before the event, whereas warnings based on a single ensemble system would only have allowed for a lead time of 4 days.

Several studies have now shown that a TIGGE-based approach increases lead time and skill across many climatic regions (e.g., Bao and Zhao, 2012; Pappenberger *et al.*, 2008). The information gain in applying TIGGE for hydrological forecasts has proven to be consistent in a way that is independent of the hydrological model applied. However, there is a clear sensitivity to catchment size: the smaller the catchment, the more important ensemble post-processing, calibration and combination becomes - as shown by He *et al*, 2009, for a mesoscale catchment area in the Midlands area of the UK. It is clear that the TIGGE archive has been of incredible value for furthering research in hydro-meteorological forecasting and demonstrating the potential of earlier flood warning.

Table 1 - TIGGE Project Partners

Center	***Country***	***Acronym***
Bureau of Meteorology	Australia	BoM
China Meteorological Administration	China	CMA
Canadian Meteorological Centre	Canada	CMC
Centro de Previsão de Tempo e Estudos Climáticos	Brazil	CTPEC
European Centre for Medium-range Weather Forecasts	Europe	ECMWF
Japan Meteorological Agency	Japan	JMA
Korea Meteorological Administration	Korea	KMA
Météo-France	France	MF
Met Office	UK	UKMO
National Center for Atmospheric Research	USA	NCAR
National Centers for Environmental Prediction	USA	NCEP
National Climate Data Center	USA	NCDC

References

Anwender, D., Harr, P. A. and Jones, S. C. 2008. Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: case studies. *Mon. Wea. Rev.,.,* **136**, 3226–3247.

Archambault, H.M., P.A. Harr and R.W. Moore, 2014: A diagnosis of the poorly forecast interaction between recurving tropical cyclone Malakas (2010) and the extratropical flow. Poster presentation, *AMS Hurricanes and Tropical Meteorology 2014*, San Diego, CA, USA.

Bao, H. and L. Zhao, 2012: "Development and application of an atmospheric-hydrologic-hydraulic flood forecasting model driven by TIGGE ensemble forecasts." *Acta Meteorologica Sinica,* **26**: 93-102.

Belanger, J.I., P.J. Webster, J.A. Curry, M.T. Jelinek, 2012: Extended prediction of north IndianI ocean tropical cyclones. *Wea. Forecasting*, **27**, 757–769. doi: http://dx.doi.org/10.1175/WAF-D-11-00083.1

Blake, E.S., T.B. Kimberlain, R.J. Berg, J.P. Cangialosi and J.L Beven II: 2013: Tropical Cyclone Report Hurricane Sandy. Report AL182012, National Hurricane Center, 157pp.

Bougeault, P., Z. Toth, C. Bishop, B. Brown, D. Burridge, D. Chen, E. Ebert, M. Fuentes, T. Hamill, K. Mylne, J. Nicolau, T. Paccagnella, Y.-Y. Park, D. Parsons, B. Raoult, D. Schuster, P. Silva Dias, R. Swinbank, Y. Takeuchi, W. Tennant, L. Wilson and S. Worley, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Met. Soc*., **91**, 1059–1072.

Candille, G, 2009: The multiensemble approach: the NAEFS example. *Mon. Wea. Rev.*,. **137**, 1655-1665. DOI: 10.1175/2008MWR2682.1

Cassou, C., L. Terray, J.W. Hurrell and C. Deser, 2004: North Atlantic winter climate regimes: spatial asymmetry, stationarity with time and oceanic forcing, *J. Climate*, **17**, 1055-1068.

Clark, A. J., and coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Clayton, A. M., Lorenc, A. C. and Barker, D. M. (2013), Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q.J.R. Meteorol. Soc.*, **139**: 1445–1461. doi: 10.1002/qj.2054

Dawson A., T. N. Palmer and S. Corti: 2012: Simulating Regime Structures in Weather and Climate Prediction Models. *Geophysical Research Letters*, **39**, L21805, doi:10.1029/2012GL053284

Dole, R., M. Hoerling, J. Perlwitz, J. Eischeid, P. Pegion, T. Zhang, X.-W. Quan, T. Xu, and D. Murray, 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, **38**, L06702, doi:10.1029/2010GL046582.

Frame, T., Ambaum, M.H.P., Gray, S.L. and Methven, J. (2011). Ensemble prediction of transitions of the North Atlantic eddy driven jet. *Q. J. Roy. Met. Soc.*, **137**, 1288-1297.

Froude, L. S. R., 2010: TIGGE: comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Weather and Forecasting*, **25**, 819-836.

Froude, L. S. R., 2011: TIGGE: comparison of the prediction of Southern Hemisphere extratropical cyclones by different ensemble prediction systems. *Weather and Forecasting*, **26**, 388-398.

Galarneau, T. J., T. M. Hamill, R. M. Dole, and J. Perlwitz, 2012:: A multiscale analysis of the extreme weather events over western Russia andnorthern Pakistan during July 2010. *Mon. Wea. Rev.*, **140**, 1639–1664.

Gray, S. L., Dunning, C., Methven, J., Masato, G. and Chagnon, J. (2014): Systematic model forecast error in Rossby wave structure. *Geophysical Res. Lett*., **41**, doi:10.1002/2014GL059282.

Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., and T. N. Palmer, 2012: Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart J. Royal Meteor. Soc.*, **138**, 1814-1827.

Halperin D. J. and co-authors, 2013: An evaluation of tropical cyclone genesis forecasts from global numerical models. *Wea. Forecasting*, in press.,

Hamill, T.M., J.S. Whitaker, M. Fiorino and S.G. Benjamin, 2011, Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter, *Mon. Wea. Rev*.,. **139**, 668-688. doi: http://dx.doi.org/10.1175/2010MWR3456.1

Hamill, T. M., 2012: Verification of TIGGE multi-model and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the conterminous US. *Mon. Wea. Rev.*, **140**, 2232-2252.

Harr, P. A., Anwender, D. and Jones, S. C.,., 2008:: Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: methodology and a case study of Typhoon Nabi (2005). *Mon. Wea. Rev*. **136**, 3205–3225.

Harr, P.A., 2010: The extratropical transition of tropical cyclones: structural characteristics, downstream impacts, and forecast challenges. *Global Perspectives on Tropical Cyclones*, J.C.L. Chan and J.D. Kepert, Ed., World Scientific, 149-174.

He, Y., F. Wetterhall, et al. 2009: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteor. Applications*, **16**: 91-101

Heming, J., and J. Goerss, 2010: Track and structure forecasts of tropical cyclones, *Global Perspectives on Tropical Cyclones*, World Scientific, 287-323.

Hewson, T. D. and Titley, H. A., 2010: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteor. Apps.*, **17**, 355–381.

Hopsch, S.B., S.C. Jones, J.H. Keller, 2014: Analysis of tropical high impact weather events using TIGGE data. Oral Presentation, 31st AMS Conference on Hurricanes and Tropical Meteorology, 2014, San Diego, CA, USA.

Houtekamer, P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.*, **133**, 604-620.

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

Jolliffe, I.T., and D.B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. 2nd Edition. Wiley and Sons Ltd, 274 pp.

Jones, S. C., Harr, P. A., Abraham, J., Bosart, L. F., Bowyer, P. J. and co-authors. 2003. The extratropical transition of tropical cyclones: forecast challenges, current understanding, and future directions. *Wea. Forecasting,*, **18**, 16–56.

Jung, T. and M. Matsueda, 2014: Verification of global numerical weather forecasting systems in polar regions using TIGGE data. *Quart. J. Roy. Meteor. Soc.*, doi: 10.1002/qj.2437.

Keller, J. H., S. C. Jones, J. L. Evans, and P. A. Harr, 2011: Characteristics of the TIGGE multimodel ensemble prediction system in representing forecast variability associated with extratropical transition, *Geophys. Res. Lett.*, **38**, L12802, *doi:*10.1029/2011GL047275.

Kepert, J. D., 2010: Tropical cyclone structure and dynamics, *Global Perspectives on Tropical Cyclones,* World Scientific, 3-53.

King, D., J. Davidson, and L. Anderson-Berry, 2010: *Disaster Mitigation and Societal Impacts, Global Perspectives on Tropical Cyclones*, World Scientific, 409-436.

Madden, R. A., and P. R. Julian, 1972:: Description of global-scale circulation cells in the tropics with a 40-50 day period, *J. Atmos. Sci.*, **29**, 1109-1123.

Magnusson, L., A. Thorpe, M. Bonavita, S. Lang, T. McNally, N. Wedi and M. Yamaguchi, 2014: Evaluation of forecasts for hurricane Sandy, *Mon. Wea. Rev.*, **142**,1962-1981).

Majumdar, S.J. and P.M. Finocchio, 2010: On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Wea. Forecasting*, **25**, 659-680. http://journals.ametsoc.org/doi/abs/10.1175/2009WAF2222327.1

Majumdar, S. J., Chen, S.-G. and Wu, C.-C., 2011, Characteristics of ensemble transform Kalman filter adaptive sampling guidance for tropical cyclones. *Quart. J. Roy. Meteor. Soc.*, **137**, 503-520. doi: 10.1002/qj.746

Matsueda, M., 2009: Blocking predictability in operational medium-range ensemble forecasts. *SOLA*, **5**, 113-116.

Matsueda, M., 2011: Predictability of Euro-Russian blocking in summer of 2010. *Geophys. Res. Lett.*, **38**, L06801, doi:10.1029/2010GL046557.

Matsueda, M. and T. Nakazawa, 2014: Early warning products for severe weather events derived from operational medium-range ensemble forecasts. *Meteor. Apps*. doi: 10.1002/met.1444

Matsueda, M. and H. Endo, 2011: Verification of medium-range MJO forecasts with TIGGE. *Geophys. Res. Lett.*, **38**, L11801, doi:10.1029/2011GL047480.

Matsueda, M., and Tanaka, H. L., 2008: Can MCGE outperform the ECMWF ensemble? *SOLA*, **4**, 77-80.

Majumdar, S. J., and R. D. Torn, 2014: Probabilistic verification of global and mesoscale ensemble forecasts of tropical cyclogenesis. *Wea. Forecasting*. (In press)

Qi, L., H. Yu, and P. Chen, 2013: Selective ensemble mean technique for tropical cyclone track forecast by using ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **140**, 805–813,

Pappenberger, F., J. Bartholmes, and others, 2008: New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophys. Res. Letters*, **35**, L10404, doi:10.1029/2008GL033837.

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

Richardson, D., R., Buizza and R. Hagedorn, 2005: *First Workshop on the THORPEX Interactive Grand Global Ensemble (TIGGE)*. WMO/TD-No. 1273, WWRP/THORPEX No. 5, 3939 pp.

Seo, K. -H., and others, 2009:,: Evaluation of MJO forecast skill from several statistical and dynamical forecast models, *J. Climate,Climate* **22**, 2372 – 2388.. doi:10.1175/2008JCLI2421.1.

Shapiro, M. and A. Thorpe, 2004: THORPEX: a global atmospheric research programme for the beginning of the 21st century. *WMO Bulletin*, July 2004.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System – part 1:concept and development, *Hydrol. Earth Syst. Sci.*, **13**, 125-140, doi:10.5194/hess-13-125-2009, 2009.

Tory K. J., and W. M. Frank, 2010: Tropical cyclone formation, *Global Perspectives on Tropical Cyclones*, World Scientific, 55-91.

Tsai, H.-C., and R. L. Elsberry (2013), Detection of tropical cyclone track changes from the ECMWF ensemble prediction system, *Geophys. Res. Lett.*, **40**, 797–801, doi:10.1002/grl.50172.

Wheeler, M.C. and H.H. Hendon, 2004, An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, *Monthly Weather Review*, **132**, 1917-1932.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences* (2^{nd} edition). Academic Press, 627 pp.

WMO, 2005a: *The Observing System Research and Predictability experiment (THORPEX): International Science Plan*. World Meteorological Organization, WMO/TD-No. 1246, 5757 pp. [Available online at http://www.wmo.int/pages/prog/arep/wwrp/new/documents/CD_ROM_international_science_plan_v3.pdf]

WMO, 2005b: *The Observing System Research and Predictability experiment (THORPEX): International Research Implementation Plan*. World Meteorological Organization, WMO/TD-No. 1258, 4848 pp. [Available online at http://www.wmo.int/pages/prog/arep/wwrp/new/documents/CD_ROM_implementation_plan_v1.pdf]

Worley, S. Schuster, D., Raoult, B, Chen, D. and Gong, J., 2008: Improving highhigh-impact weather forecasts. *EOS*, **89**330-331.

Yamaguchi, M., T. Nakazawa, and S. Hoshino, 2012: On the relative benefits of a multi-centre grand ensemble for tropical cyclone track prediction in the western North Pacific.*Quart. J. Roy. Meteor lSoc.*, **138**, 2019-2029.. doi: 10.1002/qj.1937.

Yamaguchi, M. and S. J. Majumdar, 2010: Using TIGGE data to diagnose initial perturbations and their growth for tropical cyclone ensemble forecasts. *Mon. Wea. Rev.*, **138**, 3634-3655.

Figure captions

Figure 1: Comparison of the skill of Northern Hemisphere 500 hPa forecasts from systems contributing to TIGGE, for December 2013 through February 2014. Each forecast is verified against its own analysis. Solid lines show the RMS error of the ensemble mean, and dashed lines the control member of each ensemble. Refer to Table 1 for forecast center abbreviations. The number following the center name indicates the number of ensemble members used.

Figure 2:Time series of daily (00 UTC) 2-meter temperature analyses from four different TIGGE analyses, here for a grid point in the Amazon basin (10°S, 60°W). The numbers associated with the legend indicate the yearly mean analyzed temperature. Thin, lighter-colored lines provide the daily analyses, and thicker, darker-colored lines provide the smoothed analysis, an average of +/- 15 days.

Figure 3: Reliability diagrams for T+48 to T+72 h accumulated precipitation forecasts on a 1-degree grid over the contiguous US, for individual ensemble prediction systems (panels a-d) and for the multi-model ensemble (e).This used the Jul-Oct 2011 ensemble data set as in Hamill (2012), but here the reliability diagrams were populated with forecasts from both 00 and 12 UTC initial conditions. Brier Skill Scores (BSS) computations were performed as in Wilks (2006). The inset histogram shows the frequency with which forecasts were issued; horizontal solid lines therein denote the frequency distribution of climatological forecasts.

Figure 4: Bias in (a) intensity and (b) propagation speed of extratropical cyclones tracked in forecasts from the different global centers contributing to TIGGE, as a function of lead time (Froude, 2010).

Figure 5: Ensemble forecasts for a high-impact extratropical cyclone crossing the UK and Denmark. The circle shows the observed location of the cyclone at 1200 UTC 28 Oct 2013. Top panels: "Dalmatian plots" representing cyclonic features in the T+72 ensemble forecast from (a) ECMWF and (b) Met Office. The features are colored by maximum wind speed (see scale in knots) within a 300km radius at 1km altitude. Note that the features from every ensemble member are overlain, so the location of the cyclone is indicated by 52 or 24 dots, for ECMWF and Met Office forecasts, respectively. Mean sea level pressure from the control run is also shown for both centers. Lower panels: Cyclonic feature strike probability estimated from (c) ECMWF and (d) Met Office ensemble forecasts (T+72) using cyclonic feature tracking. At each point the color represents the probability that a moving cyclonic centre associated with wind speeds over 60kts (at 1km altitude, within 300km of the centre) will at some point, within a centered 24-hour window, be less than 300km away.

Figure 6: (a) ECMWF analysis for the real-time multivariate MJO index for the 90 days prior to the initial date of the forecast. Real-time multivariate MJO index forecasts by (b) BoM, (c) CMA, (d) CMC, (e) CPTEC, (f) ECMWF, (g) JMA, (h) KMA, (i) NCEP, and (j) Met Office, initialized at 1200 UTC on 1st April 2009. The black circle and the black line with numbered circles correspond to each analysis (note that there are considerable differences between some of the analyses). The numbers in the colored circles indicate the number of days from the initial date. The colored lines indicate ensemble members. The color changes reflect the lead time of the forecast. Analyses and forecasts generally travel in a counterclockwise direction. (Figure from Matsueda and Endo (2011))

Figure 7: Track predictions (thin lines) by multi-model grand ensemble (a) for typhoon Megi initiated at 1200 UTC on 25 October 2010 and (b) for typhoon Conson initiated at 1200 UTC on 12 July 2010. The black line is the observed track, and blue, green, purple, orange and red denote prediction times of 1 to 5 days, respectively (after Yamaguchi *et al*., 2012).

Figure 8: 5-day forecasts of individual ensemble tracks strike probability and ensemble mean track forecasts for Hurricane Sandy from 12Z on the 25th October 2012. The strike probability is the probability that the center of the storm will pass within 75 miles (approximately 120km) during the forecast period. The observed track is indicated by a thicker black line with diamond symbols in 6 hourly increments, and with a grey line before the forecast period.

Figure 9: Occurrence probabilities of heavy rainfall on 1st September 2009, when there was severe flooding in Ouagadoudou (marked with an ‘X’). The shading indicates occurrence probabilities by the (a) multi-model grand ensemble, (b) ECMWF ensemble, (c) JMA ensemble, (d) NCEP ensemble, and (e) Met Office ensemble, initialized at 1200 UTC 27th August, and showing rainfall for 1200 UTC 31st August to 1200 UTC 1st September 2009. Contours in (b–e) indicate predicted sea level pressure in each control run. The climatological 90th percentiles of the models at each lead time were used to define the predicted extremes. (f) Observed rainfall from GSMaP (Global Satellite Mapping of Prediction) dataset, relative to observed climatology, and observed pressure (contours).

Figure 10: River discharge predictions for a point on the river Jiu, Romania, where flooding was observed. The 5th and 95th percentile of predictions are shown for the different forecasts with a 5-day lead time. The dashed horizontal lines show the four

EFAS warning thresholds. “Observed” discharges refer to simulations based on observed meteorological input. (Figure from Pappenberger *et al*, 2008.)

Figure 1

Click here to download Rendered Figure: BAMS_RMSE_Fig1.eps

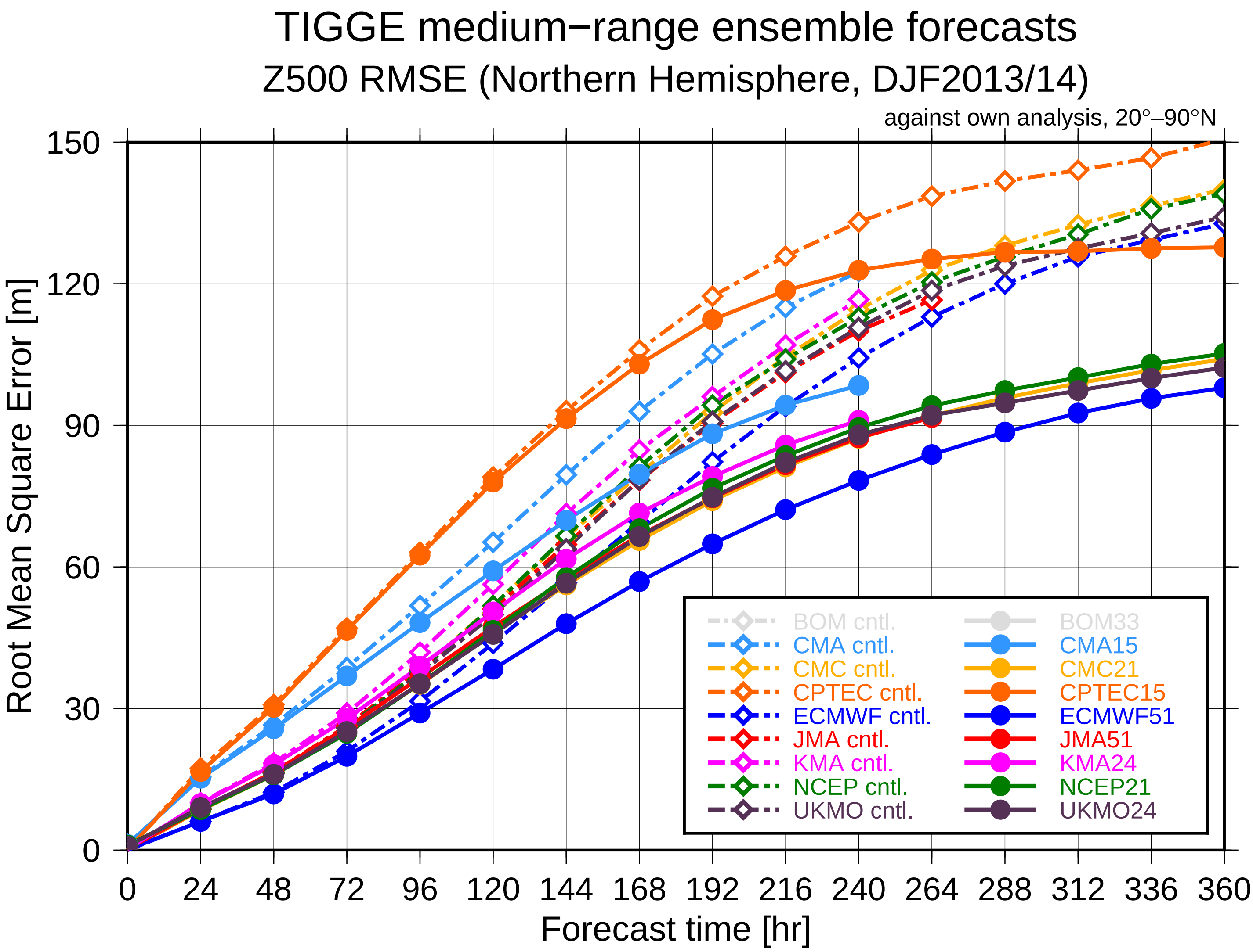


Figure 2

Click here to download Rendered Figure: Amazon_Fig2.pdf

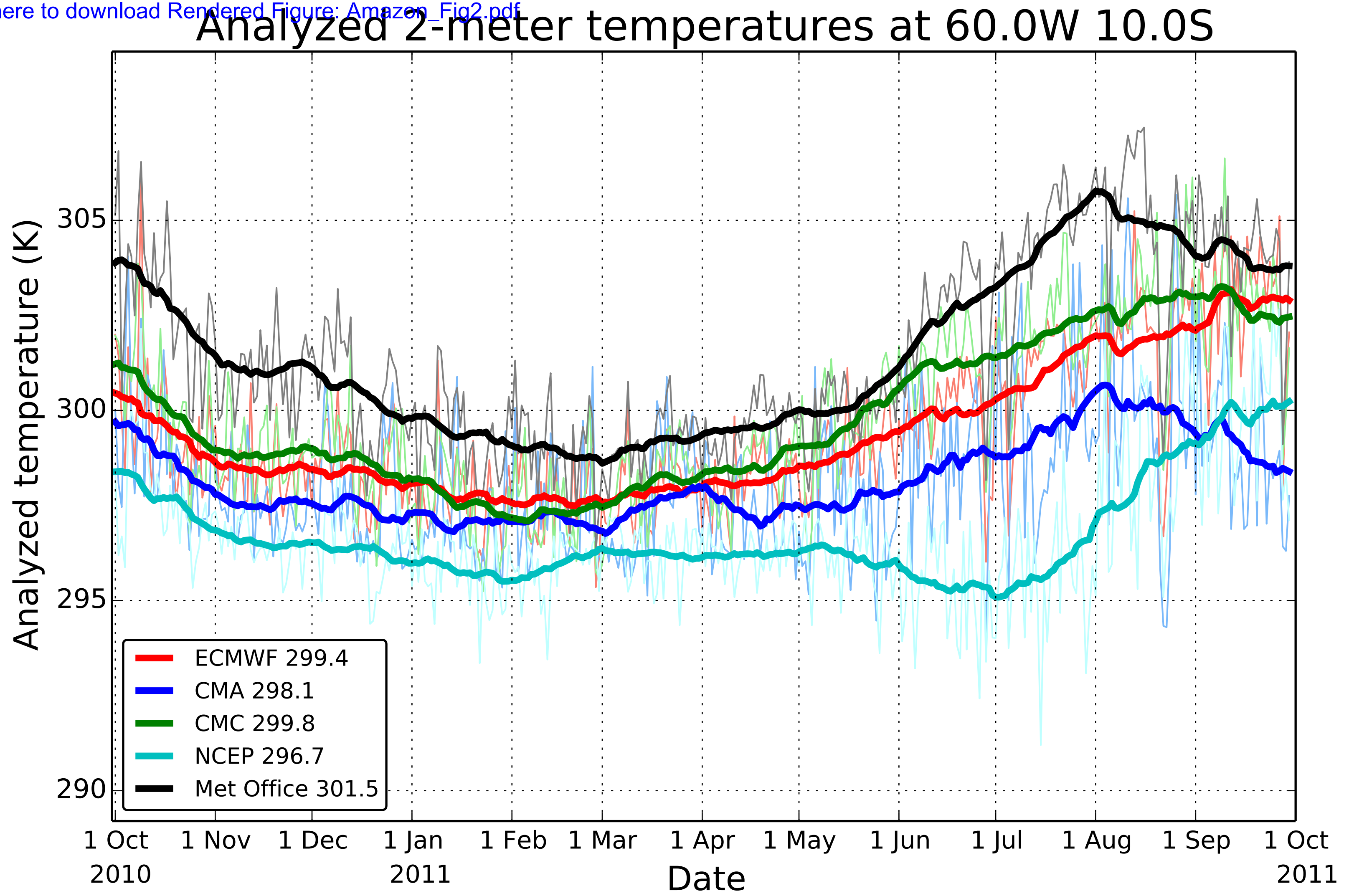


Figure 3

Click here to download high resolution image

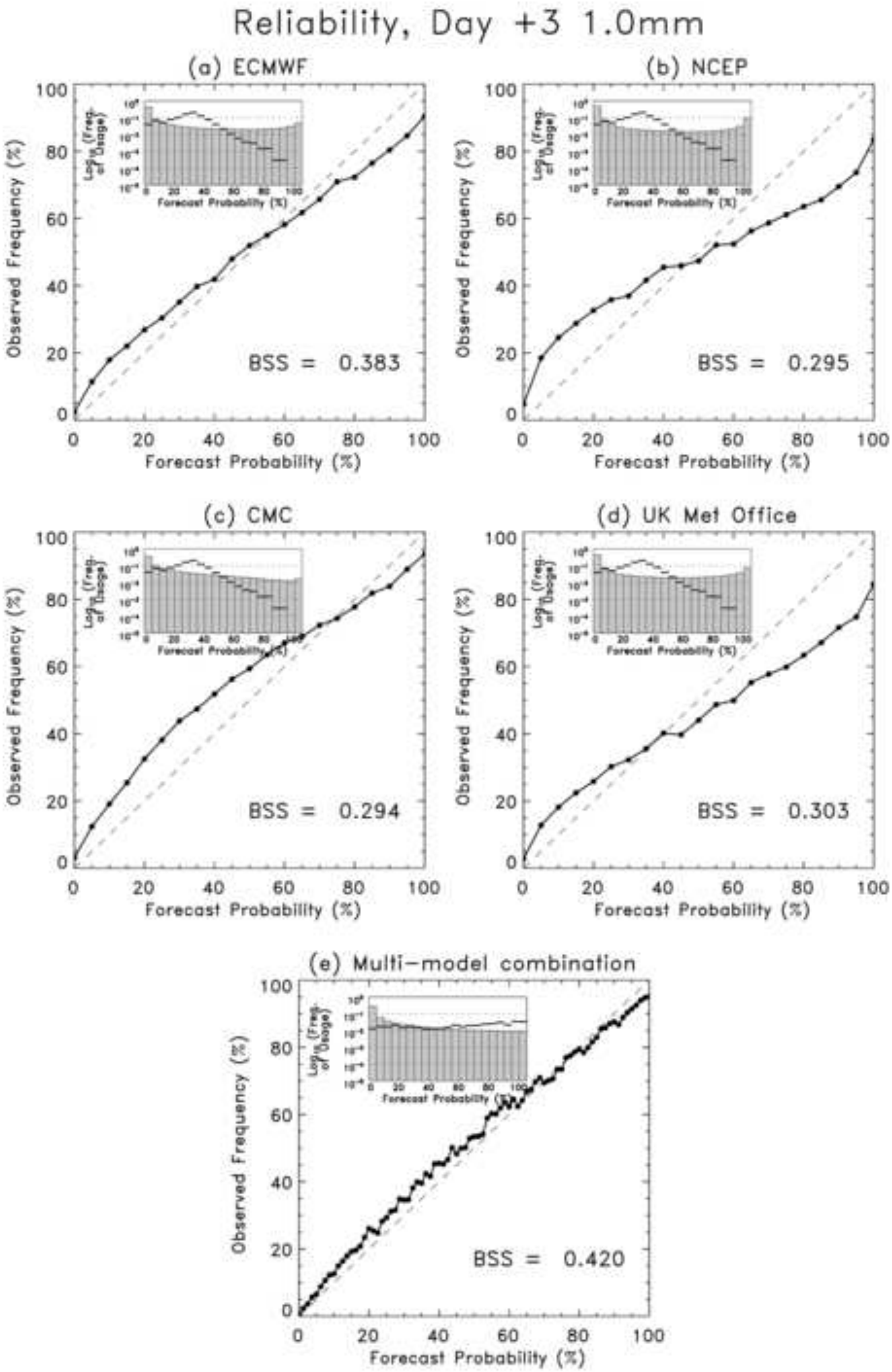


Figure 4
Click here to download Rendered Figure: Int_Prop_Bias_Fig4.eps

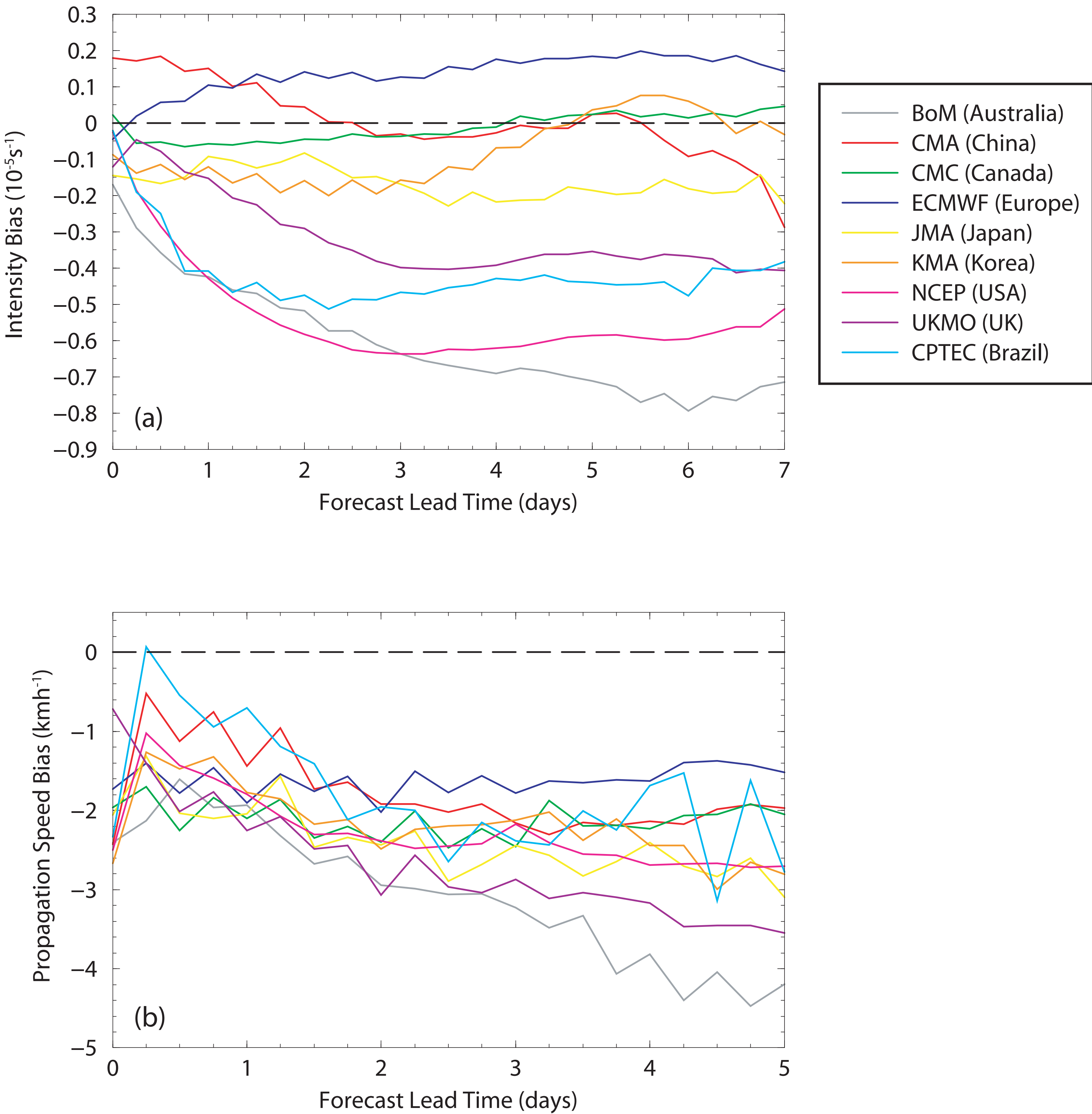


Figure 5
Click here to download high resolution image

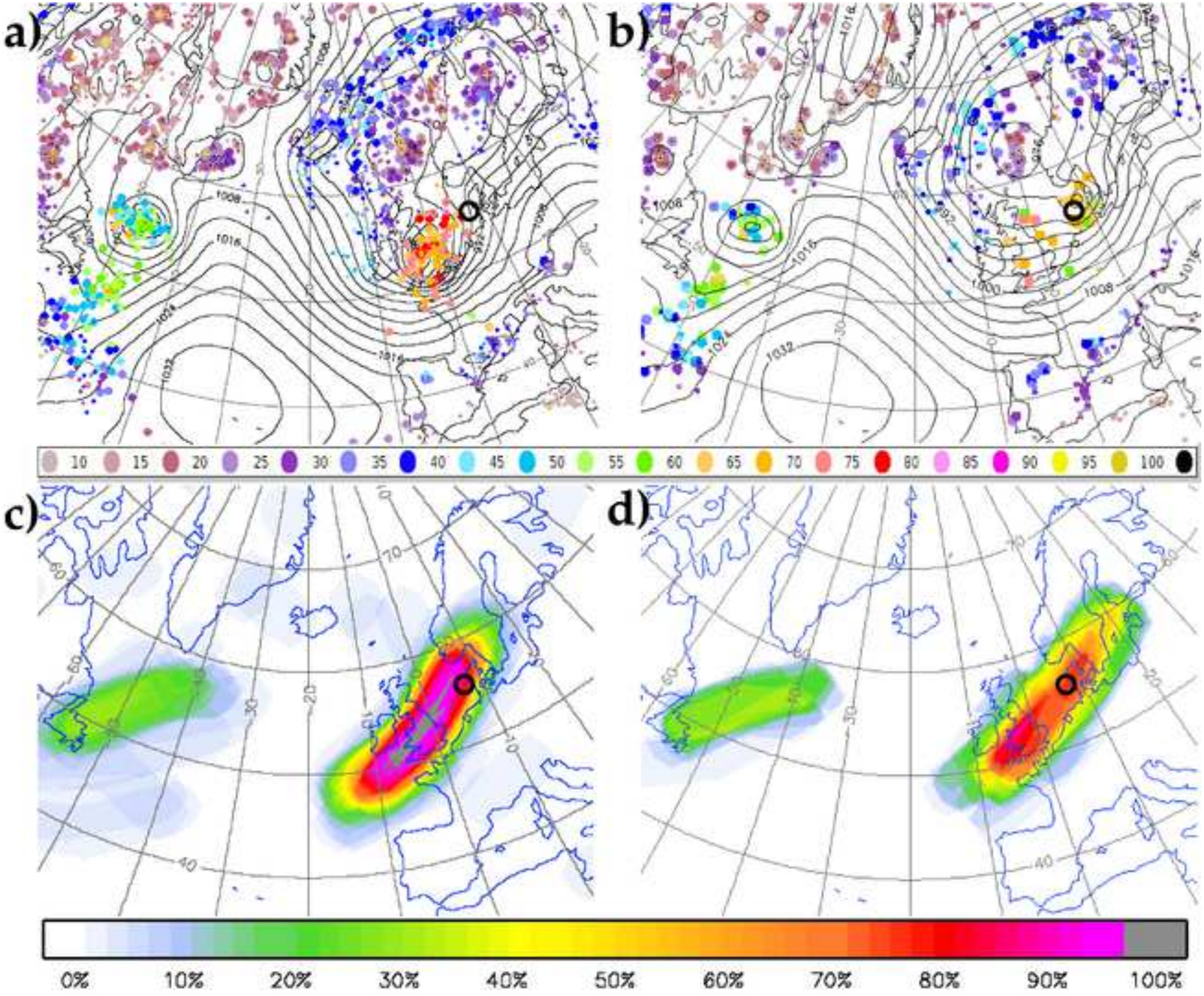


Figure 6

Click here to download Rendered Figure: MJO_index_Fig6.eps

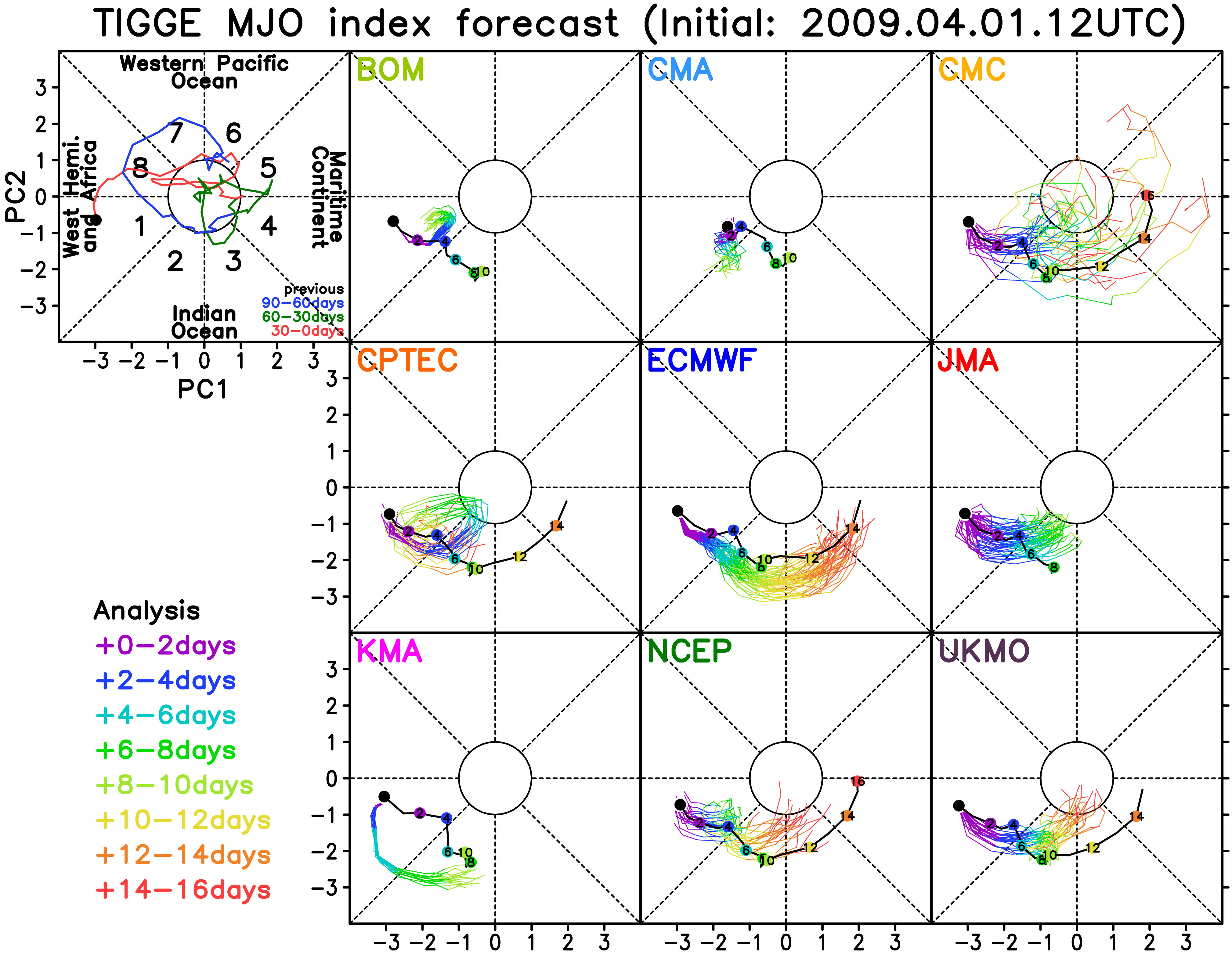
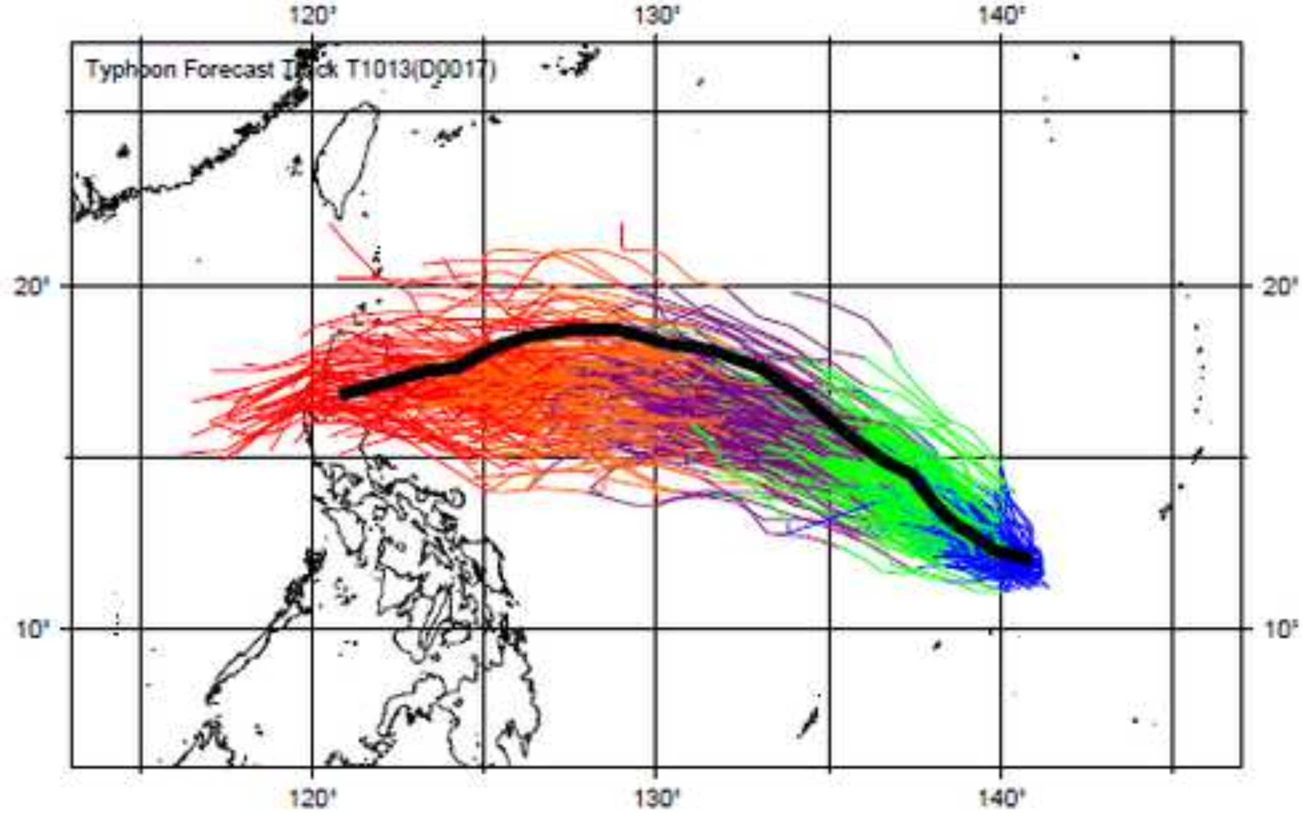


Figure 7

Click here to download high resolution image



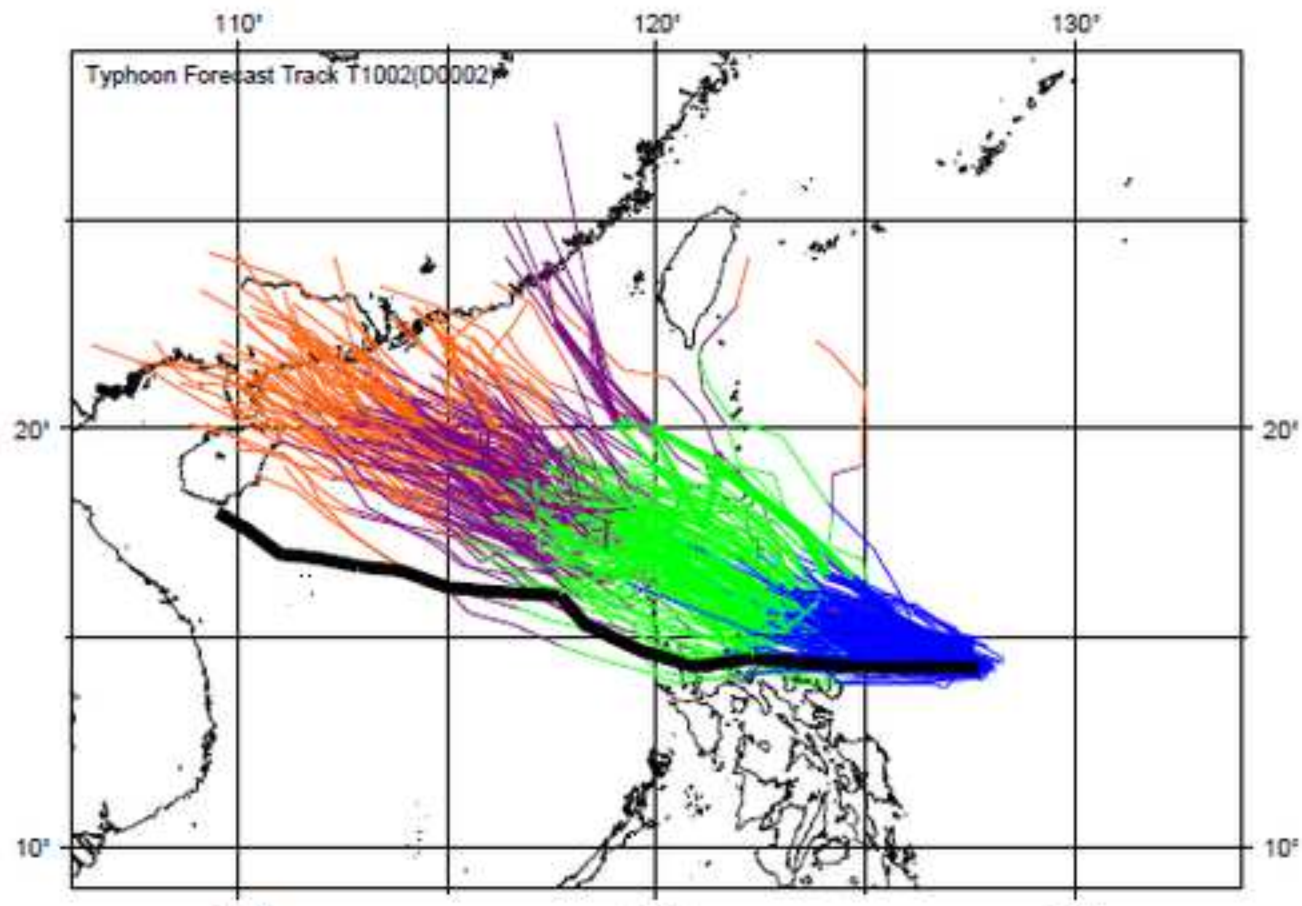


Figure 8
Click here to download high resolution image

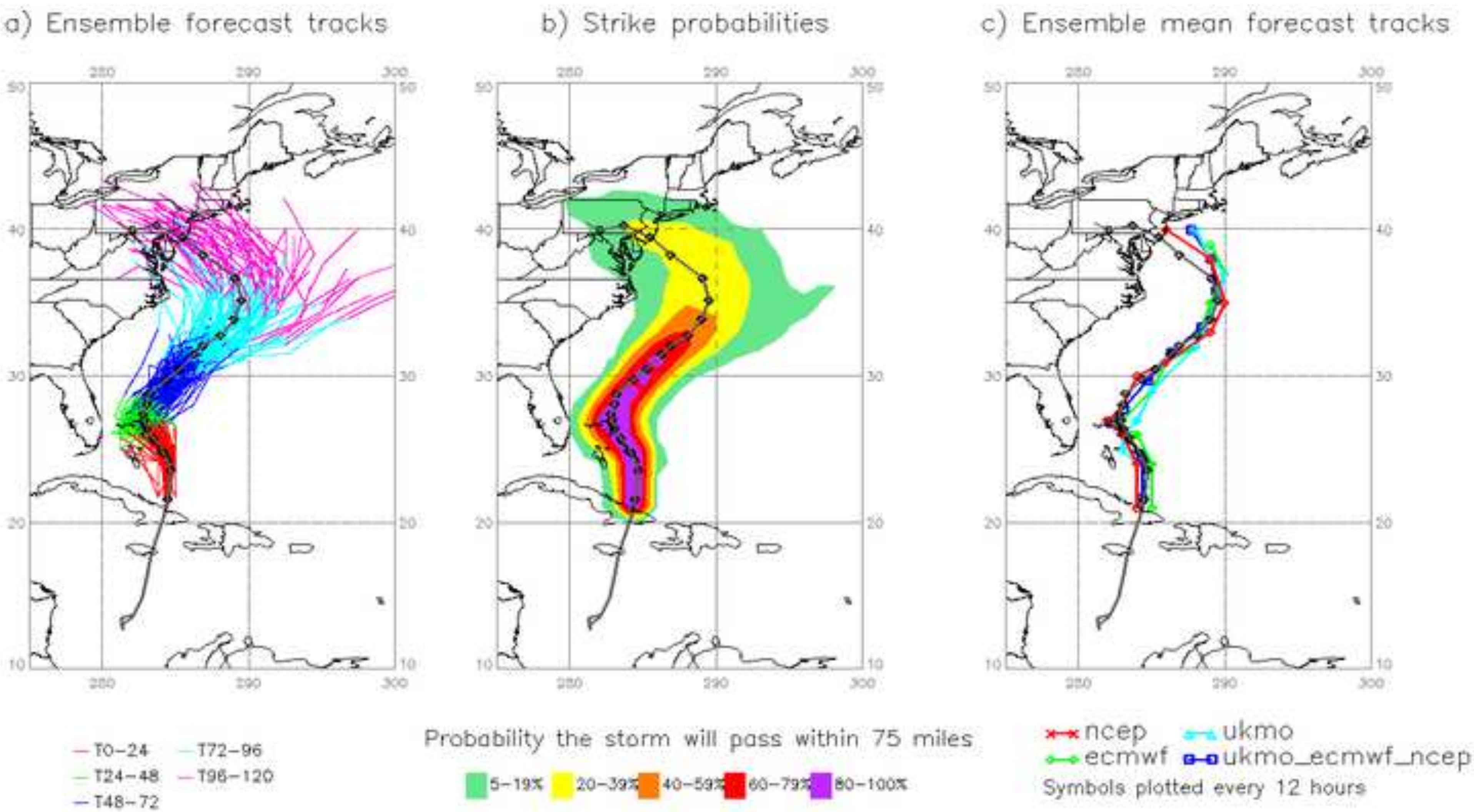


Figure 9
Click here to download Rendered Figure: Extreme_Fig9.eps

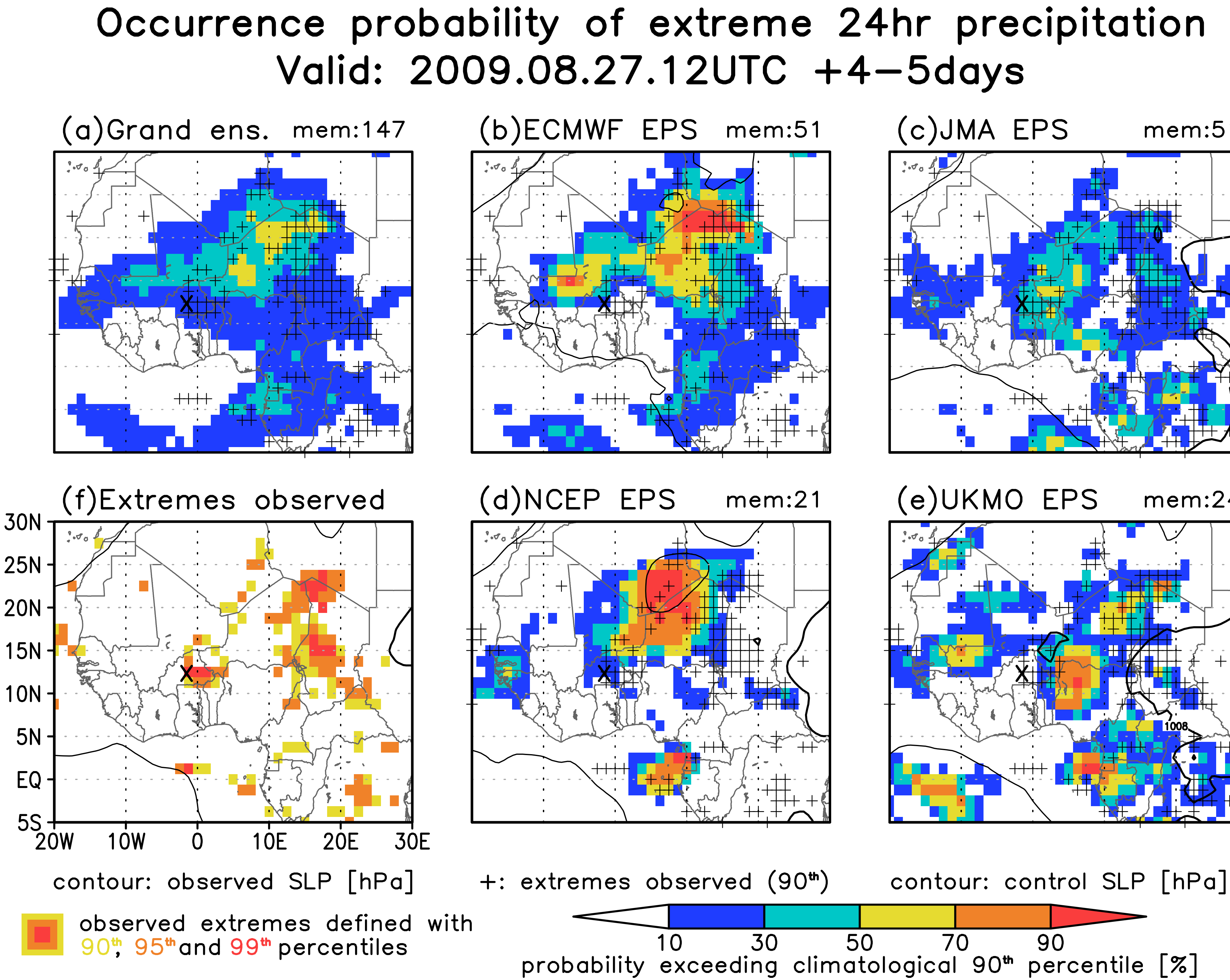


Figure 10
Click here to download high resolution image

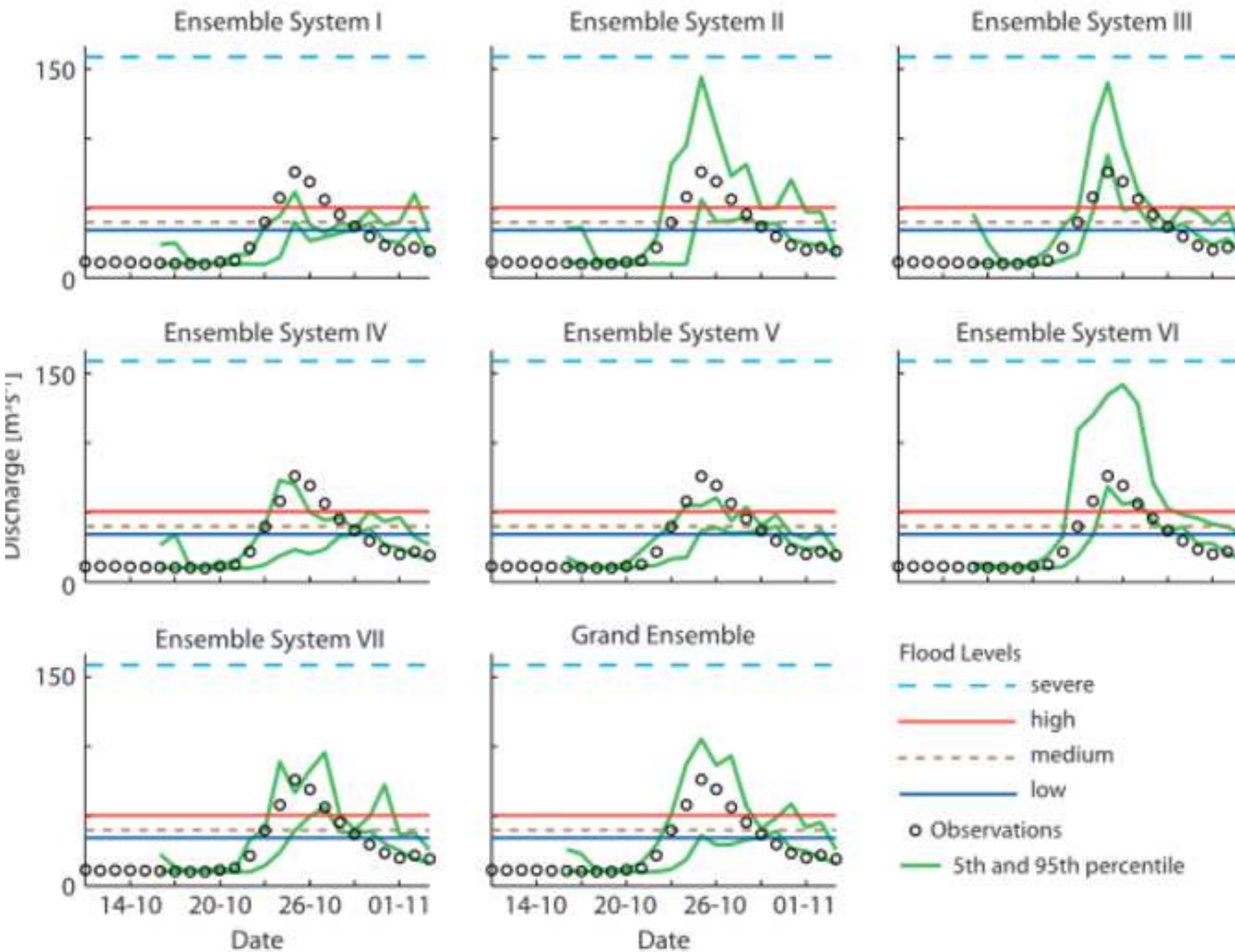
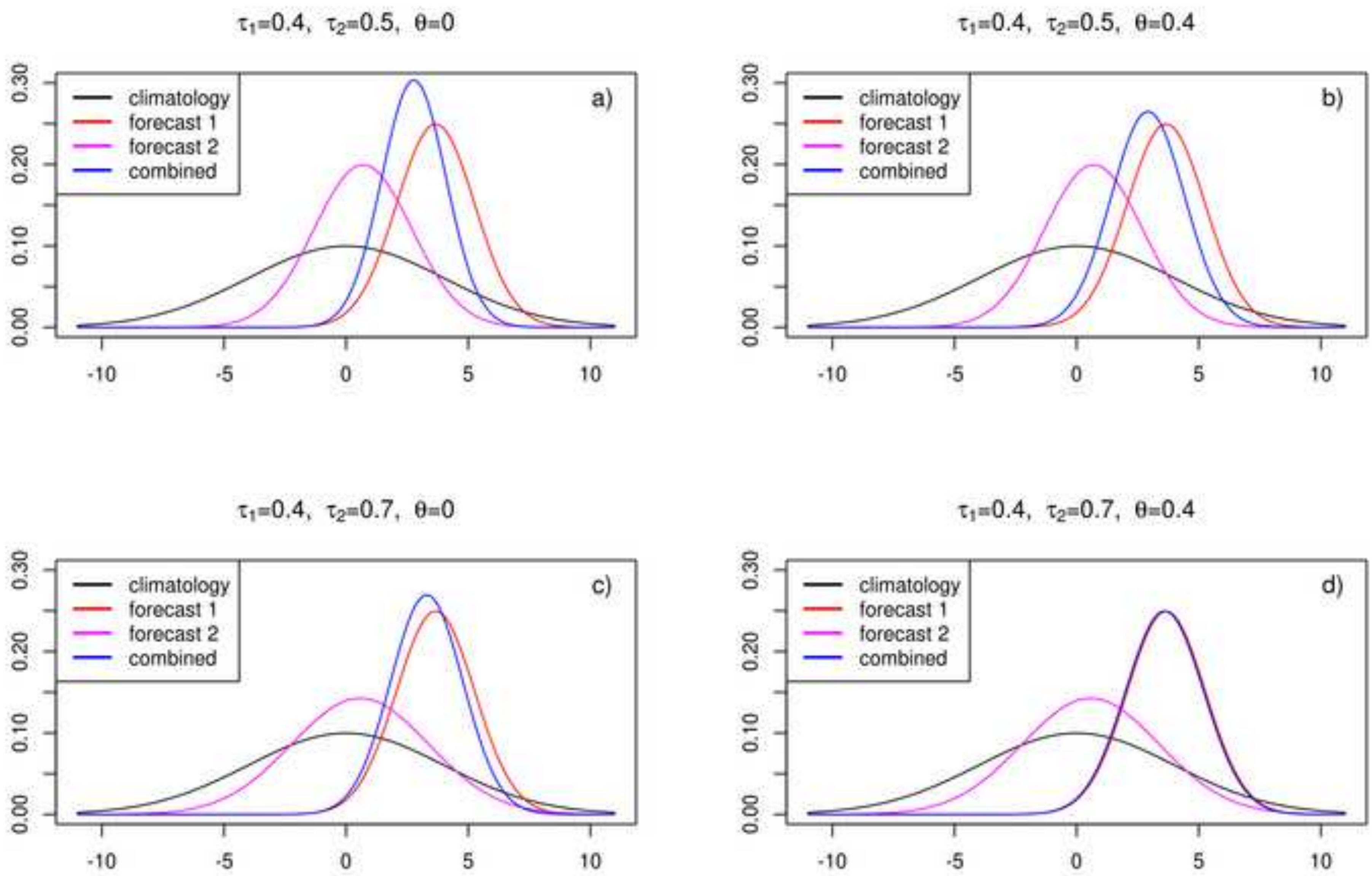


Figure A1

Click here to download high resolution image



General remarks

In addition to revisions that specifically address the reviewers' comments, please note that we have now merged the original sections 3 and 4 to make a new section 3 on "Dynamics and Predictability". We re-ordered the sections so that the sub-section on the Madden-Julian Oscillation now comes before the sub-sections on tropical cyclones, so that Figs 6 and 7 are now switched.

Reviewer 1 - Summary

The manuscript represents a review of the accomplishments of TIGGE since it was initiated in 2006. As the authors correctly stated, TIGGE has made a substantial and diverse contribution in advancing research, and in enhancing cooperation between the research and operational communities. The timing of publishing a paper on BAMS that describes these accomplishments is fitting as the THORPEX era winds down at the end of 2014, and it will also serve to raise further awareness of TIGGE in the community and suggest opportunities moving forward using TIGGE and/or TIGGE-LAM. The overall structure of the manuscript is reasonable. Several published studies that used TIGGE have been documented in the main body of the manuscript to convey the breadth of accomplishments. My overall view is that the manuscript will eventually be of sufficient quality and interest to be published in BAMS. However, it still feels like an early draft, with several areas that require improvement to be useful to the broader community. Therefore, I am suggesting a major revision, though I am enthusiastic about its publication after it has been revised.

Major Comments

1. The manuscript is not a smooth, flowing read. It appears as though each individual author gave their contribution, and then each contribution was pasted together as opposed to being woven into the text. Each sub-section has a different style of writing, which gives the overall manuscript a disjointed feel. The manuscript needs to be written in a uniform style of high-quality English such that it reads as a coherent document, with a common structure through each section. There is redundancy: some general (though correct) statements that were likely written independently by each author will need to be removed. The manuscript could probably convey the same message if the overall length is reduced by 25%.

We have rewritten large sections of the paper– we hope the revised version has a more uniform style, with sections woven, rather than pasted, together. We have also removed redundancy and repetition, such as most of the parts of the conclusion which summarized the earlier sections. However, in some cases, we needed to add further text to address other points raised by the reviewers. The main part of the text is now around 2 (doubled spaced) pages shorter than in the original version.

2. A suggestion for each sub-section is to structure them as follows: motivate the problem specific to that class of weather phenomenon, what the main research and/or operational challenges are, and how they were addressed by TIGGE. Instead of listing a brief summary of each paper in a matter-of-fact fashion (e.g. Section 3.2), the main contributions to the advancement of that field through TIGGE could be summarized. The key references are important, but with their contribution placed in the context of the challenges. In some sections, this is performed well (e.g. 2.2), whereas several other sections do not convey the message of how TIGGE contributed to advance the field.

The new Section 3 now highlights new findings enabled by the TIGGE database, as well as open opportunities in dynamics and predictability research identified through the analysis of TIGGE. Other changes have been made to better convey how TIGGE has contributed to advances in ensemble forecasting and its applications. A lot of the advances come from the fact that the ensemble prediction data is simply far more readily accessible to researchers than it was before TIGGE; there is *a limit to how many times that point can be reiterated. We have not rigidly followed the suggested structure of each sub-section, since that would run counter to the advice to shorten the article.*

3. The two major goals are listed on lines 26-29. In the conclusions, it would be useful to succinctly summarize the extent to these goals were met. The first goal was clearly met. The second goal was largely met in providing public access to the TIGGE database, and some subsequent research in predictability. However, how about societal and economic impacts?

In the interests of streamlining the introduction in response to the general tenor of the reviewers' comments, we have now omitted the specific goals from the THORPEX implementation plan. Whilst that has removed the imperative to discuss socio-economic impacts here, there will be benefits through the applications described in section 4. We anticipate that this issue will be better addressed in subsequent papers reviewing other aspects of THORPEX.

4. A more balanced, critical assessment is necessary, especially in the conclusions. In addition to the many accomplishments, what gaps in the field were not addressed, and why were they not addressed? Can we learn from this, for example, by adding more variables and/or levels, improving the efficiency of data transfer, enhancing more international collaborations, tying in more closely with other WMO/WWRP groups, promoting TIGGE-LAM, and by creating databases focused on a large number of case studies of high-impact weather?

Adding much more volume to the TIGGE data (such more variables and levels) would not be feasible without strong justification. However some minor enhancements to TIGGE may be possible, to support specific requirements of the legacy projects – this is now mentioned in the final section.

5. The summary and discussions are long and sprawling. It essentially condenses Sections 2-5 then discusses TIGGE-LAM and future changes within WWRP. It may be more useful not to condense Sections 2-5 but instead give a perspective on the broad contributions based on what is in Sections 2-5, together with the aforementioned critical assessment. A new final paragraph can give suggestions to the broad community on the way forward.

The final section has been re-written along these lines

6. Though some papers and projects are described in Section 5 [*now 4*], the specific contributions of TIGGE studies to the subsequent improvement of operational forecasting have not been clearly summarized.

Whilst we have mentioned a few aspects of the pull-through from TIGGE to operational forecasting, the overall review of the contributions of TIGGE and the rest of THORPEX to operations are closely linked. A broader treatment of the topic would be a significant widening of the scope of this paper, and it should be better covered by one of the planned THORPEX summary papers.

7. Figures 2, 8 and 10 could be improved. In Fig. 2, to save space, the daily values could be plotted in a lighter shade, with the 15-day smoothed lines on the same plot made thicker. Solid lines would be cleaner. In Fig. 8, the figures are too small for publication, and there is extraneous text. I could not resolve the diamond symbols and grey line. In Fig. 10, a more insightful illustration of how TIGGE fields are specifically used in hydrology through to warnings would be more helpful to the reader than the basic 'cartoon' schematic.

The figures have been changed. Figure 2 has been changed as suggested. Figure 8 has been improved to make it more legible. Figure 10 has been replaced by some results from a study of river level predictions based on TIGGE, from a case study of floods in Romania.

Minor Comments

1. Lines 2-3, abstract: WWRP and THORPEX will likely need to be spelled out.

2. Line 6, abstract: the non-academic research community is also an important user of TIGGE, yet they have been excluded. Best to replace "academic" with "research", which implicitly includes academia.

3. Line 13, abstract: a reference is needed to back up the statement that "tropical cyclones are most severe weather systems in the world", and in any case this statement does not offer anything to the BAMS audience. Furthermore, tropical cyclones are not "a major focus of the paper", only being explicitly used in Sections 4.1 and 5.1. It is fine to give this weighting to tropical cyclones, but the sentence about them being a "major focus" is misleading.

The first three comments refer to the abstract, which we were asked not to change once the original proposal to BAMS was accepted, although we might have preferred make some changes before the full paper was submitted. However, we agree with the logic behind the comments and have made the suggested changes to the revised manuscript. We have changed the description of the tropical cyclones from "most severe" to "most destructive", since strong tornados are more severe, but much more local; a reference to justify that statement is given at the start of section 3.4.

4. Introduction: roughly how many peer-reviewed papers have been published to date using TIGGE? That would be a powerful statement on the contribution of TIGGE to the field. Are there any other statistics (data downloads, number of users etc) that would also be useful?

We have now included these metrics in the introduction.

5. Lines 67-70 seem out of place. They disrupt the flow of the text that outlines the remainder of the paper. Furthermore, some more specific details of TIGGE-LAM and their recent successes in particular regions can be summarized.

The introduction has been restructured in a more logical order.

6. Section 2.2: although the informed reader may be aware that biases are likely to be reduced when ensembles from different models are combined, it may be useful to explicitly state this at the start of the section, to give the broader audience the main driver(s) for combining TIGGE data.

Section 2.2 has been largely re-written and now opens with a discussion of how combination can lead to more skilful forecasts.

7. Line 133: what is "reliable" guidance? From Figure 3, it appears to mean a better reliability diagram. In the text, "reliable" could be defined.

This part of the text has been re-written.

8. Line 144: why is the simple combination of ensemble system data for an improved ensemble more "suspect"? I would anticipate that it is less necessary, but am not clear on why it is suspect.

"more suspect" changed to "less beneficial"

9. Section 2.3: How did TIGGE contribute to the advancement of calibration methods?

The text has been amended to point out that the main way that TIGGE has contributed is by providing the data on which these studies depend.

10. Section 3.1: How did TIGGE therefore advance our understanding of extratropical cyclones and storm tracks? The section starts off well but degenerates into a matter-of-fact list of studies without the broad perspective.

Greater emphasis has been placed on the role of TIGGE early on in the section. A paragraph summarizing some studies, but without the broader perspective, has been cut.

11. Section 3.2: See major comments. It recites details at the expense of an overarching perspective about what we have learnt with TIGGE. The descriptions of some of the papers will need to be condensed.

Changes have been made to this section, first to clarify the discussion of the role of resolution, and later to remove the paragraph on interactions between weather systems and the eddy-driven jet – where it was not so clear what we have learned using TIGGE.

12. Section 4.1: the first sentence is controversial and needs a reference to confirm it.

We have added a reference to King et al, which reviews the impacts of tropical cyclones.

13. Line 302: Typo: "Finocchio". Also, the combination of ensembles in Majumdar and Finocchio (2008) actually degraded the forecast skill of TC track probabilities, compared with ECMWF alone.

We now mention the Majumdar and Finocchio finding that, in some cases, there can be degradation of forecast skill.

14. Line 131: "design and interpret…"

Done

15. Line 319: Two recent papers have used TIGGE to examine the predictability and predictive skill of tropical cyclogenesis:

a. Komaromi, W. A., and S. J. Majumdar, 2014: Ensemble-Based Error and Predictability Metrics Associated with Tropical Cyclogenesis. Part I: Basin-Wide Perspective. *Mon. Wea. Rev*., In Press

b. Majumdar, S. J. and R. D. Torn, 2014: Probabilistic Verification of Global and Mesoscale Ensemble Forecasts of Tropical Cyclogenesis. *Wea. Forecasting*, In Press.

We are aiming to minimize references to unpublished papers, but we have included the latter reference.

16. Line 320: Several leading scientists in the tropical cyclone community would dispute that "TC intensity changes and genesis events are often controlled by large scale synoptic features". If a statement this strong is to be made, a reference to confirm it would be necessary.

We have amended this statement and added references.

17. Lines 331-334: an example of some generic introductory text, which can be removed.

Done.

18. Line 339: do the individual constituent EPS contribute to only the synoptic scenarios, or the structure of the cyclone undergoing ET as well?

The study only addresses the synoptic scenarios.

19. Section 4.2: how have these studies served to advance our understanding in ET, compared with if TIGGE was not available? As with several other sections, this section merely reports on papers but is short on perspectives.

Simply put, the studies would not have been feasible without TIGGE making the data available.

20. Section 4.3: the structure of the first paragraph is jumbled. The MJO can be introduced first (third sentence), before describing how hard it is to predict it and its variability.

Section 4.3 has been re-structured more logically

21. Line 354: a medium-range (3-14 day?) forecast of a seasonal phenomenon?

It is made clearer that TIGGE can only be used to compare forecasts over a partial MJO cycle.

22. Line 359: what do phases 1-4 and 5-8 correspond to?

The MJO phases are now explained.

23. Line 365: the web link could be included as either a footnote, or in an electronic supplement listing websites.

It is now explained that the MJO forecast data can be accessed from the TIGGE Museum, which is now introduced in a separate box.

24. Line 376: “see below”. Where?

The bracket was misplaced – it is the SWFDP that is described below (in 4.2).

25. Section 5.1: what were the main results of the NWP-TCEFP?

The text now explains that the TCEFP successfully demonstrated the use of ensemble-based forecast products for TC forecasting.

26. Section 5.1: the Sandy example is unconvincing, since the single-model ensembles appear very similar in Figure 8, compared with average operational forecast errors of tropical cyclone track. Specifically how was the multi-model ensemble forecast better than the single-model ensemble forecast?

Figure 8 is an illustration of examples of how multi-model grand ensembles can be displayed. In itself, it does not prove that multi-model ensemble products are better. The basis for that assertion is the statistical studies described in section 3. The text has been re-worded to make that point more clearly.

27. Lines 402-408: this paragraph needs references. “Severe surface temperature” is a strange phrase.

The conclusions are drawn from the Matsueda and Nakazawa paper – the reference is repeated to make that clearer. “events” has been added to the “severe surface temperature” term.

28. Lines 418-420: this sentence is vague and gives no perspective on the paper.

This sentence has been changed to make the perspective clearer.

29. Lines 459-462: this sentence could be moved more upfront to suggest how hydrological models use meteorological predictors from ensembles. An improved Figure 10 would help.

The hydrological applications have now been moved to a separate box, as suggested by another reviewer. The box text now opens with that point. Figure 10 has been replaced.

30. Line 527: this paragraph may start better with PPP and S2S, which are better established and in the case of S2S will use a database similar to TIGGE as stated. Then HIWeather can be introduced, though it has not yet been formally approved by the WWRP.

As part of a bigger re-write of the last section, the legacy projects are now described in the suggested order. The HI Weather project has now been approved by WMO EC, though the research plan has not yet been formally approved by the WWRP SSC.

31. There are several grammatical errors through the manuscript. The style of writing is too casual in places, e.g. “ramping up”. The tenses are jumbled between present and past in places
We have attempted to address these general points during the revision process.

Reviewer #2: General Comments:

Overall, I find the manuscript to be a well-constructed review of the major results and achievements of the TIGGE project. It is generally written clearly and should be accessible to the broad audience of BAMS. There are a number of minor issues, particularly regarding the figures, that need to be addressed (see below). Hopefully these edits will not prove to be burdensome.

In terms of larger issues with the paper, the length is well above the BAMS guidelines (about 15 double-spaced manuscript pages). I think it would be prudent to shorten the main text somewhat, both by eliminating some text and using a sidebar or two. For example, I think Sec. 5.3 is well-suited to be a sidebar, focusing on a particular "downstream" application of TIGGE forecast data, as input into hydrological models. In Sec. 6, I think much of the material in lines 471-515 can be eliminated (it largely summarizes results presented earlier, which already were reviews of existing literature), shifting the focus of the final section much more heavily towards future directions.   I don't expect the paper to be shortened to 15 pages, but I do think it is a reasonable goal to eliminate 2-3 pages.

We have taken on board the reviewer's point about the length. By shortening some text from most of the sections of the paper, and removing much of the summarized material from the last section, we have reduced the page length of the main text by around 2 pages. We have taken up the suggestion of changing section 5.3 (as it was) to be a sidebar (or box). We have also shifted the focus of the last section to discuss future directions in more depth.

One thing I believe is missing from the paper is information concerning the continuation of existing TIGGE archiving activities. At the end of 2014, is the archiving of ensemble forecasts (available at the ECMWF TIGGE portal, for example) going to stop? Or will it continue under TIGGE or one of the successor programs (HI Weather, PPP, S2S) mentioned in Sec. 6? Users of the TIGGE ensemble products would likely be interesting in some brief comments on these matters if answers to such questions are known at this time.

We have now covered these issues in the first part of the last section

Detailed Comments:

Line 23: I do not see anything in the conclusions regarding "links with ensemble data assimilation methods".

We apologize for this oversight – the discussion was originally omitted to save space, but we have now added a paragraph in the conclusions.

Line 99: Should read "Park et al."

Corrected

Line 167: Perhaps "reforecast dataset" or "reforecast sample".

"dataset" added

Line 176: Is "stormtracks" really one word? I have always seen it as "storm tracks". The same goes for "jetstream" in Sec. 3.2. I've always seen "jet stream".

These terms have been split into 2 words.

Line 211: I assume 22M, 17M, and 6M are the populations of the respective three areas, in millions. This level of detail is not necessary, I think.

Population details omitted

Line 213: I don't believe I saw where MOGREPS was defined. Is that a regional ensemble?

MOGREPS refers to Met Office Global and Regional Ensemble; for consistency through the paper, we now just refer to the ensembles by data provider, not the specific ensemble designations.

Lines 306-307: "However, as shown in Figure 6, there are still some occasions when the actual track falls on the edge of the forecast ensemble". This suggests that it is due to some deficiency in the ensemble that the actual track sometimes falls on the edge of the ensemble forecast track distribution. If the ensemble is reliable, there should be cases in which the actual track falls near the edge of the ensemble forecast track distribution, and indeed some cases in which the actual track falls outside the envelope of ensemble member tracks (there just shouldn't be too many of such cases!). This is a common misconception concerning interpretation of ensemble forecasts, I believe.

The plots in Fig. 6 should be perhaps be presented simply as examples of multi-model ensemble tropical cyclone forecast tracks, or removed altogether.

We have corrected the text, and Fig 7 (was 6) is now presented to show contrasting examples of multi-model forecasts and actual tracks.

Line 359: Are the MJO phases referred to here the Wheeler and Hendon (2004) phases? Please be specific.

Yes – we have added a brief description of the Wheeler & Hendon phases.

Line 376: Should read "(SWFDP, see below)."

Done

Line 380: No capitalization need here for "hurricane".

Done

Lines 381-382: According to the National Hurricane Center Tropical Cyclone report for Sandy (http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf), the post-tropical Sandy made landfall with a central pressure of 945 hPa (see top of page 4 in the report).

Corrected

Line 385: Should read "comprised" instead of "comprising".

Done

Line 393: Should read "developed a prototype suite of ensemble-based early warning products".

Corrected

Line 405: Perhaps add "events" to the end of the sentence.

Done

Lines 414-416: Combining the ensembles clear provides a smoother probability field than any individual-model ensemble, as seen in Fig. 9. It is not clear from the figure, though, that the combined ensemble probabilities are more skillful than those of the individual-model ensembles, or that the combined ensemble reduces the impact of model error. These things may well be true, but they are not something we can conclude from one example forecast.

The text has been changed to make it clearer that this is one example – the statistical benefit of combining ensemble track forecasts is discussed in section 4.

Line 436: I don't think Figure 10 adds anything substantive to the manuscript. I would either remove it, or add something else concerning hydrology. Perhaps something from the Romanian flood case study would be interesting.

We have made these changes, replacing the original version of Fig 10.

Line 441: Should read "were" instead of "where".

This subsection has been re-written as box 3.

Line 447: Remove "and catchments".

This subsection has been re-written as box 3.

Line 464: Should read "is" instead of "us".

This subsection has been re-written as box 3.

Lines 525-526: Not sure what to do with this sentence, but it doesn't sound right in its present form.

This was changed as part of the general re-writing of the final section.

Box 1: Is there an NCAR portal for the TIGGE data? In the introduction, it states that an NCAR portal had been set up, but perhaps it is no longer maintained? Also, in line 10 replace "head" with "ahead".

NCAR was one of the original TIGGE archive centers, but is being discontinued – the original text was inconsistent about whether to include NCAR, but it is now included in Box 1, and the future plans are set out in the final section. "head" changed to "ahead".

Figure 1: What do the numbers mean in the legend (e.g. ECMWF51)? I assume they are the number of ensemble members. Please include this information in the figure caption (or remove the numbers from the legend).

They are – caption has been changed.

Figure 2: To save space, I would present one panel with the legend from existing panel (a) and the lines from existing panel (b).

The panels have been combined.

Figure 3: What are the horizontal black bars in the inset histograms?

Now explained in caption.

Figure 5: It is very difficult to see the mean sea level pressure field and coastal outline in Fig. 5; please make these lines darker. Also there are two periods at the end of the caption.

The pressure& coast lines are now darker

Figure 8: Please make these plots larger, perhaps by a factor of two. Replace "SANDY" with "Sandy" in the caption.

Plot replaced by an improved version.

Figure 9: There are no letters labeling the individual panels of the figure.

The figure is now updated

Reviewer #3:

This paper represents an important update to the community on the research efforts associated with the TIGGE archive. I recommend that this paper be published with minor revisions. The paper needs some improvement before it is publishable in BAMS. My suggested changes are generally minor in nature and include:

More major concerns:

1) The authors make an important point regarding the issues that arise from comparing multiple models against one archive (pg. 9) The authors then conclude that a consensus analysis might be a better approach. The inference is to narrow. What about also comparing the models directly against observations?

We now discuss the issue of verification data in more depth.

2) To me, the discussion on pg 10 and 11 regarding the future of ensemble models is an oversimplification. I think the conclusions reached by the authors, strongly hinge on greatly advancing model physics given the parameters where the multi-model performs better than the best ensemble (e.g., surface temperature, rainfall etc). Thus, the improvement probably hinges both on efforts to improve the characteristics of uncertainty in the ensembles, but also to improve the physical parameterizations in deterministic and ensemble systems. What I mean is better paramterizations and then go for stochastic physics around those better parameterizations. Please improve the text.

We have improved the discussion of these points in the text.

3) If I am correct, on page 13, 15 and elsewhere, the authors are drawing inferences from comparing models at different resolutions. The Dawson conclusion in the text checks the results by comparing all models at a similar resolution. I do not think that the statement on pg. 13 regarding intensity in different models is similarly strong. When looking at intensity, one needs to take into account the resolution of the obs and the different models.

The Dawson et al paper examined the impact of varying resolution (of a single model) on the ability to simulation jet stream variability; we have clarified the text to make it clearer what inferences we are drawing from that study. The TIGGE models have other differences, of course, but other results (e.g. the Froude studies) indicate that resolution is a significant common factor in the quality of storm track simulations, though Gray et al results show that even the highest resolution models are unable to maintain a high PV gradient.

4) The statement is made on pg 26 that the TC intensity forecasts are limited by resolution. Please read Hakim (2013), which describes the low inherent predictability in tropical cyclone intensity. Resolution is only part of the story.

This text is now omitted

Minor issues that also need to be rectified

1) The paper is overly vague. One example is the section on the combination of different TIGGE EPS members where the authors state that "the largest benefit" and "marginal benefit". It would be good to quantify the improvements (e.g., was the benefit 12 h of additional skill, 2 days in week 2, an improved representation of ensemble spread etc????). You are trying to educate and motivate the reader so you can not be vague. Another is the ET process as "mid-latitude flows" is pretty vague. Don't you mean the Rossby wave guide or at least the westerlies to get the strong downstream errors. Look the paper over for these types of statements and conclusions that are preceded by "maybe", then correct.

We have attempted to address these issues and be more specific when we can.

2) The writing style needs some improvement.

a) The general style of "Figure 1 shows......" , "Figure 2 shows that the yearly mean analyzed 2-meter temperature from five of the TIGGE systems...." etc is not a compelling style for the general readership of BAMS. I would urge the authors to adopt a style such as "The yearly mean analyzed 2-meter temperature from five of the TIGGE systems (Fig. 2) can vary by almost 5 K etc".

b) The authors seem to have an aversion to commas that grew as the paper went on.

Again, we have attempted to improve these stylistic aspects.

3) The article should be written in a more compelling style informing the reader about TIGGE and the associated research if the paper is to appear in BAMS. The intro is something that needs improving. Too much is spent on the goals in the implementation plan and not enough on the real accomplishment of furthering collaboration between operational centers and between centers and researchers.

We have tightened up the introductory section, and skipped over the specifics of addressing the original goals of the THORPEX implementation plan.

4) The paper misses some opportunities to inform the reader of advances regarding TIGGE. I did not know what the TIGGE museum was since it suddenly appeared in the text. So I went on line. The museum looks like a great tool for those interested in weather forecasting and dynamics, yet the concept is just mentioned without any background. The TIGGE museum is a development of interest, yet almost no background information is provided on the why, what and future of this concept. Students, for example, would find this site to be valuable. Please provide some background.

We have raised the profile of the TIGGE museum by introducing it earlier and providing a box to summarize its "exhibits".

5) The summary section is very disappointing. One would think that the authors would take this opportunity to discuss future topics that could be addressed with this archive and, well, the future of the EPS, in general. The limitations of the data set could also be addressed. However, the authors instead talk about THORPEX legacy projects. Remove or reduce the legacy information and stick to the TIGGE theme and be bold and creative on future directions. You are a group of leaders and developing leaders in the ensemble prediction field so something visionary to the BAMS readers is appropriate in the conclusions. Currently, you have a few paragraphs of summary, then one of the future and then you jump to the legacy projects. I think that the reader would like to know "What the most important issues that TIGGE will address in the next 5-10 years?", "How will the archive change as storage and communication of data improves (e.g., more levels, more model tendencies etc)? How will ensemble research change in the next decade and what is the role, if any, for TIGGE?

We have added more discussion of the future directions of TIGGE and ensemble research. Much of the role of TIGGE is to support the THORPEX legacy projects, so we disagree with the reviewer's recommendation to reduce or remove that information. There would need to be a very persuasive case made to convince the archive centers before we could significantly increase the scope of TIGGE to more levels, or adding tendency information.

6) The paper is uneven in its treatment of events. For example, the St. Jude storm is given a reference and a description of the impact and Sandy is treated in a secondary manner without a reference nor a mention of the impact. BAMS articles are suppose to be somewhat news worthy with a general audience appeal. By skipping over Sandy, you missed a real chance to engage the reader. Similarly, the background info is also quite small on the MJO.

We have added further background information on both Sandy and the MJO.

7) The paper makes a number of broad, partially correct statements that need rewording. These statements include: i) Tropical cyclones are the most severe weather system in the world (tell that to someone that has seen the damage from a strong tornado) - please reword;

We agree “most severe” may not be the best term to use, but damage from tropical cyclones is much more widespread, so it is more precise to say “more destructive” – we have quoted an appropriate reference at the start of section 4.

ii) (pg. 12) Mesoscale features can bring both damaging surface winds and heavy precipitation leading to flooding --- please reword ---some countries have snow and freezing rain;

We now say “impacts such as widespread flooding”

iii) TC intensity changes are often controlled by large-scale synoptic features (pg. 18) --- the literature also suggests that aspects like eyewall replacement and regional SST factors are also important and can dominate -- please reword;

We have modified this wording.

iv) Multi-model approaches in hydrology are well established and "popular"--- please provide a citation, not a figure to support this point.

We have quoted references to support this point.

v) By providing data, TIGGE has enhanced cooperation between academic and operational communities (pg 5) -- true but cooperation is from both academic use of operational data and direct operational-academic research partnerships that have grown out of TIGGE - please clarify;

We have focused on the former as a direct benefit of TIGGE. It is more debatable whether direct operational-academic partnerships stem from TIGGE, although by contributing to bridging the gap between the operational and academic communities, TIGGE has probably helped the development of these partnerships (and also partnerships between operational centers).

vi) GRIB2 is a WMO format, but I thought the development of GRIB2 was necessary because of TIGGE -- pleae clarify was GRIB2 ensemble exchange format necessary because of TIGGE?;

GRIB2 was brought in to support ensemble prediction data, not just for TIGGE. TIGGE and NAEFS were the first projects to use the format in a major way. We now mention that GRIB2 supports ensemble data, but we consider that going into further detail in the introduction would be inappropriate.

vii) The intro and other text leads one to believe that TIGGE has simply followed the THORPEX implementation plan, but one focus of the plan was a truly interactive system that included targeting -- please correct, by this oversimplification you are misleading the audience.

It is true that a truly interactive system, including data targeting, was part of the original THORPEX vision, though it is outside the scope of this paper on TIGGE. We now mention that on-demand ensemble predictions are not yet a reality. By omitting some specifics on the original THORPEX goals, we should avoid potentially misleading the readers. This discussion would be better left to the planned papers surveying the overall achievements of THORPEX,

8) In looking at Fig. 2 of data in Brazil, I am surprised to see that the data from the Brazilian center is not included. It seems appropriate to include, does it not?

For that matter, several other centers' data were also omitted – e.g., Australia, Korea Japan. Rather than being a slight against Brazil, this should be interpreted as being due to the already large amount of data on the plot.

****

Page Charge Estimate Form
Click here to download Page Charge Estimate Form: IEPC_BAMS.pdf