Journal of Hydrology: Regional Studies 4 (2015) 196–227

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil

Fernando Mainardi Fan[a,b,*], Dirk Schwanenberg[b,c], Walter Collischonn[a], Albrecht Weerts[b,d]

[a] Instituto de Pesquisas Hidráulicas (IPH) – Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
[b] University of Duisburg-Essen (UDE), Essen, Germany
[c] Deltares, Delft, The Netherlands
[d] Wageningen University and Research Centre, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 23 April 2015
Accepted 9 May 2015
Available online 9 July 2015

Keywords:
Ensemble forecasting
TIGGE database
Inflow forecasting

ABSTRACT

Study region: This paper describes a major ensemble-forecasts verification effort for inflows of three large-scale river basins of Brazil: Upper São Francisco, Doce, and Tocantins Rivers.

Study focus: In experimental scenarios, inflow forecasts were generated forcing one hydrological model with quantitative precipitation forecasts (QPF) from three selected models of the TIGGE database. This study provides information on the regional ensemble performance and also evaluates how different QPF models respond for the different basins and what happens with the use of combined QPF in a greater ensemble.

New hydrological insights for the region: This work presents one of the first extensive efforts to evaluate ensemble forecasts for large-scale basins in South America using TIGGE archive data. Results from these scenarios provide validation criteria and confirm that ensemble forecasts depend on the particular EPS used to run the hydrological model and on the basin studied. Furthermore, the use of the Super Ensemble seems to be a good strategy in terms of performance and robustness. The importance of the TIGGE database is also highlighted.

* Corresponding author at: Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Caixa Postal 15029, Av. Bento Gonçalves, 9500, 91501-970 Porto Alegre, RS, Brazil. Tel.: +55 51 3308 7511.

E-mail address: fernando.fan@ufrgs.br (F.M. Fan).

http://dx.doi.org/10.1016/j.ejrh.2015.05.012

1. Introduction

Forecasts of streamflow and river stage can be used to anticipate flood events in exposed urban centres, to improve the safety of navigation, and to optimize the operation of reservoirs and other hydraulic structures. In many applications, the benefits of the forecasts increase with lead-time, as people have more time to prepare for the coming flood or drought. In numerous cases, this implies that forecasts of river discharge have to be based on forecasts of future rainfall. Unfortunately, those quantitative precipitation forecasts are far from perfect and show high uncertainty.

During the last two decades, the importance of considering uncertainties in hydrological forecasting has increasingly been recognized. The preferred method for considering uncertainty is to use ensemble forecasts (Georgakakos and Krzysztofowicz, 2001; Cloke and Pappenberger, 2009; Pappenberger and Brown, 2013). In this type of prediction, inferences about future scenarios are made by considering, for example, multiple possible trajectories of atmospheric variables to represent the meteorological uncertainty, which when applied to a hydrological model, result in distributions of streamflow trajectories (Cuo et al., 2011; Pappenberger et al., 2011a,b; Demeritt et al., 2007; Cloke and Pappenberger, 2009; Xuan et al., 2009). In some cases, the uncertainty related to the hydrological model is considered by applying multiple hydrological models, for example (Abebe and Price, 2003; Fraley and Raftery, 2010; Velázquez et al., 2011; Franz and Hogue, 2011; He et al., 2012; Demirel et al., 2013; Bourdin and Stull, 2013). The growing use of ensemble forecasts is justified by the increased performance over deterministic forecasting in the case of floods (Boucher et al., 2011; Younis et al., 2008; Schellekens et al., 2011; Verkade and Werner, 2011; Roulin, 2007; Bartholmes et al., 2009; Buizza, 2008; Golding, 2009) and in the case of reservoirs operation (Boucher et al., 2012; Zhao et al., 2011, 2012; Bourdin and Stull, 2013).

This paper describes a major ensemble-forecasts verification effort for inflows of three large scale basins reservoirs, located in the tropical regions of Brazil. The inflow forecasts are generated by forcing one hydrological model with quantitative precipitation forecasts, (QPFs), from three selected models of the TIGGE database (Bougeault et al., 2010). The objective of this study is not only to help fill the knowledge gap on ensemble performance in the region, but also to understand how different model's QPFs behave for different basins in comparison to each other, and what happens when using combined QPFs in a greater ensemble.

Verification investigations are important because they help to show the potential benefits of tested techniques and products, and allow better understanding of the modelled systems. Based on results, developments can be planned that aim to improve accuracy (Alfieri et al., 2014). Also, in the context of operational use, verification investigations help to highlight the current system fragilities, that need to be taken into account when forecast results are to be used for decision making (Pagano et al., 2014).

Some examples of streamflow ensemble forecasting studies focusing on verifications can be found in recent literature: Renner et al. (2009) introduces a performance evaluation of forecasts at various stations in the Rhine basin (central Europe) using a hydrological model forced with the weather forecasts from the Ensemble Prediction System (EPS) of the *European Centre for Medium-Range Weather Forecasts* (ECMWF) and from COSMO-LEPS (Marsigli et al., 2005). Velázquez et al. (2009) evaluate the performance of a hydrological model forced with the meteorological ensemble forecasts produced using the global model of the Meteorological Service of Canada for five watersheds in Quebec (Canada). Bergh and Roulin (2010) show an analysis of the hydrological ensemble prediction system (HEPS) developed by the Royal Meteorological Institute (RMI) of Belgium for the Meuse and Scheldt rivers basins, using one hydrological model forced with the ECMWF-EPS. Thiemig et al. (2010) made a feasibility study of ensemble flood forecasting in the African Juba–Shabelle river basin, generated using the same methodology used in the European Flood Alert System (EFAS – Bartholmes et al., 2009). After this study, Thiemig et al. (2014) presented the *African Flood Forecasting System* (AFFS) for medium- to large-scale African river basins, including verification analysis. In the work of Schellekens et al. (2011), the authors evaluate the performance of the MOGREPS (*Met Office Global and Regional Ensemble Prediction System*) EPS within one hydrological model for flood forecasting in the Thames River Region (UK). Bao et al. (2011) used five EPS from the TIGGE database to force one hydrological model calibrated to the river Xixian (China) and verified results for a flood event that occurred in July 2007. More recently, Pagano (2014) shows an evaluation of the operational flood forecasts issued by the Mekong River commission, and Alfieri et al. (2014) introduce a verification of the EFAS ensembles performance for multiple locations in the European river network. In addition to these described studies, other research that presents verifications of case studies can be found in Roulin (2007), Olsson and Lindstrom (2008), Verbunt et al. (2007), Pappenberger et al. (2008), Younis et al. (2008), Jaun and Ahrens (2009), Xuan et al. (2009), Demargne et al. (2010), Addor et al. (2011), Boucher et al. (2011), Bourdin and Stull (2013), and Fan et al. (2014a,b,c). In general, all these studies point to the benefits of using ensemble forecasts over deterministic forecasts, as well as providing useful insights into performances, comparisons of models and opportunities for developments.

Within the present study, we intend to contribute to the current knowledge about ensemble forecasting performance, especially for the studied region. Here, better forecasts can be translated into improved decisions relating to energy generation, dam safety and flood control through hydraulic structures operation. Our main objective is to investigate the ensemble forecasting system's performance between models and basins, and to discuss the implication of these results.

Our objectives are in compliance with the objectives of the TIGGE database initiative, because we primarily used the information available in the TIGGE database to run our tests. The database creation effort include aims related to; allowing a better understanding of combinations of ensembles from different sources, a deeper understanding of model uncertainties in forecast errors, and improvement of collaboration on ensemble prediction research. Also, this is the first study where TIGGE information is used for hydrological forecasting purposes in South America. Another scientific initiative related to this work is the HEPEX (Hydrological Ensemble Prediction Experiment; Schaake et al., 2006). In this initiative, forecast verification is listed as one of the main topics of interest. Finally, we expect this work to be a first step in the development of a continental forecasting system called SAFAS (South America Flood Awareness System) that would run using configurations similar to the ones tested here, taking into account the improvements based on the insights verified with the presented tests.

1.1. Operational hydrological forecasting in Brazil

Here, flow forecasts are mainly used for two purposes: (i) programming the operation of hydroelectric power plant (HPP) reservoirs; and (ii) flood forecasting at vulnerable locations. The predominant use is however, related to reservoir operation. In Brazil, the National System Operator (*Operador Nacional do Sistema* or ONS) is the agency responsible for coordinating and controlling the generation and transmission of electricity in the National Interconnected System (*Sistema Interligado Nacional* or SIN), under the supervision and regulation of the National Electric Energy Agency (*Agência Nacional de Energia Elétrica* or ANEEL).

Under normal flow conditions, ONS uses forecasts of daily average flows with 14 days lead time to schedule the generation of hydroelectricity in the power grid. These predictions are generated by the ONS itself for some HPPs, and the utilities operating the HPPs (operators of power plants) are responsible for preparing them in other cases (ONS, 2011, 2012a,b, 2014; Zambon et al., 2012, 2014a,b; Costa et al., 2014; Oliveira et al., 2014). The ways in which forecasts are generated in these cases vary widely. Many utilities still use forecasting models with little physical basis, such as PREVIVAZH (Guilhon et al., 2007; ONS, 2011), which is a model based on a rescaling of weekly forecasts based on trends of recent past flows and natural flows (Guilhon et al., 2007; ONS, 2011, 2012a,b, 2014).

To improve the forecast skills, ONS organized a performance evaluation of different alternatives for streamflow forecasting to the SIN in 2007 (Guilhon et al., 2007). The considered test sites were located in the basins of Iguaçu, Paraná and Parnaíba Rivers. Several forecasting methods were tested, ranging from physical modelling with lumped or distributed conceptual models, hybrid methodologies, stochastic models to artificial intelligence techniques. In the end, ONS came to the conclusion that the evaluated models had a higher performance than PREVIVAZH. Also, it was concluded that information of forecasted rainfall resulted in an increased skill of streamflow forecasting (Guilhon et al., 2007).

This work led to a greater diversity of models for predicting inflow into Brazilian HPPs. It has also increased the use of Numerical Weather Prediction (NWP) models. Examples are the models for Parnaíba River basin and for Uruguay River basin, shown by Collischonn et al. (2007a,b) and Fan et al. (2014a,b,c).

In cases where a reservoir operation is not considered normal, but as a state of attention, alert or emergency (defined by volumes and hydraulic constraints), the operation of the reservoirs is no longer controlled by ONS, but by the local utilities, following certain guidelines established by ONS. This is critical in the Brazilian reservoirs context, because many of them are multi-purpose and have operational constraints such as maximum outflow (to avoid downstream floods), maximum forebay elevation (to avoid upstream floods and guarantee the dam safety), and maximum rate of change of outflow and level (ONS, 2011, 2012a,b, 2014; Zambon et al., 2012, 2014a,b; Costa et al., 2014; Oliveira et al., 2014).

In these circumstances, the inflow forecasting information is particularly useful to support decision making, especially when dealing with situations of high flow where each extra anticipated hour of an approaching event can help to manoeuvre the reservoir into a better state, for example by pre-releasing water to create additional storage. Precisely for these circumstances, many Brazilian hydroelectric utilities invest into the development and application of dedicated flow prediction systems that can provide detailed streamflow forecasts to support the operational decision making, and mitigate the risk of constraint violations. Examples of forecasting systems developed for these purposes are described in Fan et al. (2014b).

In terms of reservoir inflow forecasting technical scenarios in Brazil, one of the still underexplored issues is the use of ensemble streamflow forecasts, as currently almost all of the systems in use work with deterministic predictions.

In the context of Brazilian reservoir systems, very few applications are found related to hydrological ensemble forecasts, despite the high local dependency on water for energy generation and intensive use of forecasts for reservoir operation programming, security and flood control. Experiments of ensemble forecasts for the region can be found only in the works of Collischonn et al. (2013) and Meller et al. (2014), Calvetti and Pereira Filho (2014), and Fan et al. (2014c). The first three works mentioned do not focus on medium-range forecasts in large scale basins, where usually the larger HPPs reservoirs are located. Instead, they focus on short-range forecasts (up to three to five days) for flood anticipation on relatively smaller basins by Brazilian standards (from 12,000 km^2 to 20,000 km^2).

2. Case studies

In this paper, three large-scale Brazilian basins are used as test cases, the Upper São Francisco River Basin, the Doce River Basin and the Tocantins River basin, see Fig. 1. The Upper São Francisco river basin (delimited until the confluence with Velhas River) is located in the southeast part of Brazil, has an area of approximately 50,000 km^2 and includes a HPP called Três Marias on its downstream region. The reservoir has a volume of almost 21 billion cubic metres and serves multiple purposes, including flood protection. The travel time of a flood from the most upstream regions of the basin to the HPP Três Marias reservoir is usually between 24 h and 48 h. The Doce river basin is also located in the southeast part of Brazil, near the coast, within a region of high-elevation rugged relief. The Aimorés HPP is located in the lower portion of the Doce river basin, where forecasts are important not only for power generation and flood control, but also for the sediment management of the reservoir by a flushing procedure. The travel time of a flood from the most upstream regions of the Doce river basin to the HPP Aimorés reservoir is usually between 24 h and 48 h. The Tocantins river basin is located in the central-north part of Brazil, has an area of approximately 350,000 km^2 and is extensively used for hydropower generation, with six major HPPs in the main river (from upstream to downstream: Serra da Mesa, Cana Brava, São Salvador, Lajeado, and Estreito). The travel time of a flood from the most upstream regions of the Tocantins river basin to the HPP Estreito reservoir is usually around 72–96 h.

3. Methodology

3.1. Telemetric observed data

Fig. 2 shows a detailed view of the test-beds, including the telemetric gauges (with hourly data) used in the hydrological models setup and forecasting experiments. These gauging systems are mainly maintained by the HPP operators in the river basins under the supervision of the Brazilian National

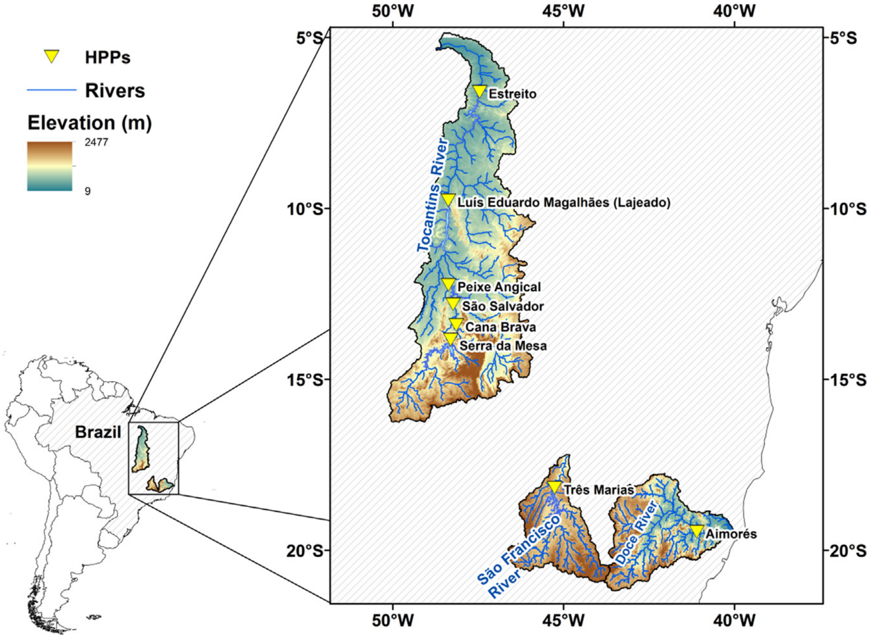

Fig. 1. Location of the studied basins.

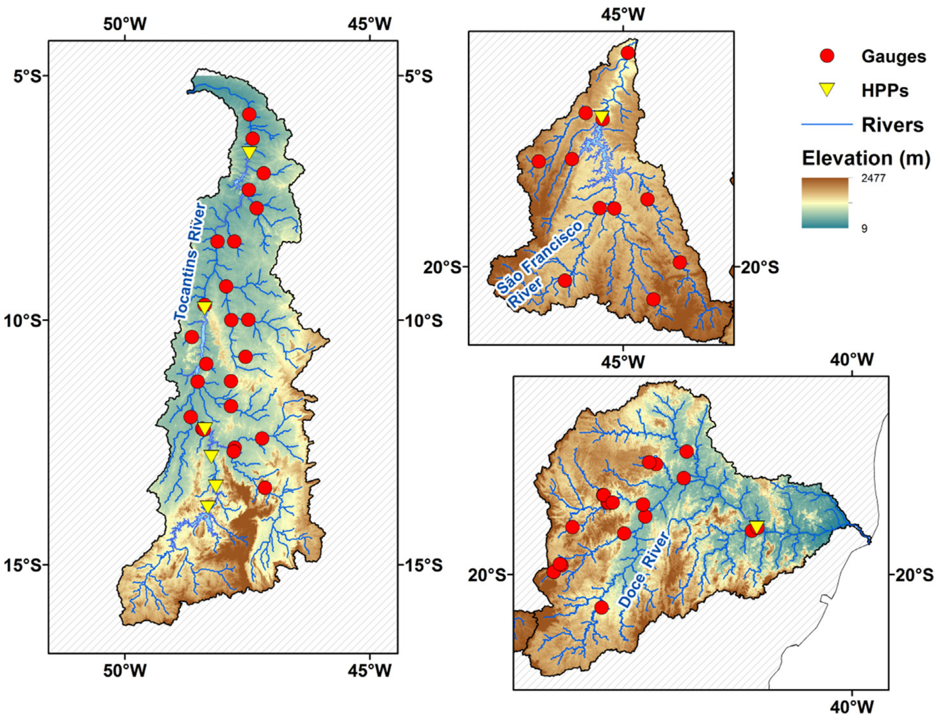

Fig. 2. Detailed view of the test-beds, including the telemetric gauges (with hourly data) used in the hydrological models setup and forecasting experiments.

Water Agency (*Agência Nacional de Águas* or ANA), and serve for both precipitation and streamflow measures, as a double function transmission system to avoid extra costs. The gauge data is usually submitted to ANA, who make it available through an Internet portal (www.ana.gov.br/telemetria).

In terms of coverage (Fig. 2), none of the basins have a very dense gauging network (density is around 4000 km^2 per gauge in the Upper São Francisco and Doce basins and approximately 7000 km^2 per gauge in the Tocantins basin). This is not exactly desirable for modelling and forecasting purposes, since heavy concentrated rainfalls that might induce floods cannot be captured by the gauges, or precipitation can be overestimated over large areas by the lack of surrounding information, as discussed by Fan et al. (2014c).

The Upper São Francisco basin can be considered the basin with the best distribution of gauging stations, with gauges over the entire region, with the exception of the upstream region of the west tributaries. In the Tocantins river basin, the most upstream portion has practically no information, and in the Doce River Basin the largest factor in data unavailability is due the lack of information in the north and south upstream areas.

Although this streamflow and rain gauge information does not have a desirable density, it is currently in use at operational forecasting systems developed for the basins. It is, therefore, the best information available, and so it was used in the presented work.

It is important to mention that the configuration used in the experiments is intended to be exactly the same as the operational systems developed for the basins, and was therefore not considered an ideal-world data availability (i.e., information that would not be available at real-time). This configuration reproduces the information currently available for operational forecasting and in use at the test-beds.

Finally, as the Tocantins river basin has part of the streamflow controlled by reservoirs, one extra consideration was taken into account. All the outflows of the reservoirs were considered as completely known until day 3 of the forecast. From this day, the outflows were considered as "run of the river" (e.g. inflows match outflows). This consideration comes from the practice of forecasting in the basin, where usually the expected outflow of reservoirs is known between one to five days in advance. This way, we adopted three days as an average value for the experiments. It is important to have in mind that this consideration affects the results of forecasts verification for the first days of the forecasts, usually indicating better performances in terms of metrics, since the knowledge about the future is greater.

3.2. Hydrological modelling

We use the MGB-IPH (*Modelo de Grandes Bacias – Instituto de Pesquisas Hidráulicas*) model (Collischonn et al., 2005, 2007a,b; Paz et al., 2007; Paiva et al., 2013) to conduct ensemble streamflow forecasts based on meteorological forcing. MGB-IPH is a large-scale distributed hydrological model that calculates streamflow from precipitation data. It uses a Hydrological Response Unit (HRU) approach (Kouwen et al., 1993), combining information from soil types and vegetation cover. In the model, the basin is divided into small catchments based on topography from digital elevation models (Paiva et al., 2013). Every catchment is further divided into different HRUs. Streamflow is generated within each HRU of every catchment using soil water storage and runoff generation approaches similar to the models Arno (Todini, 1996) and LARSIM (Ludwig and Bremicker, 2006). Water is routed within catchments to the main rivers using simple linear reservoirs, while river routing can be computed using a Muskingum–Cunge approach (used in the present application), a full hydrodynamic model (Paiva et al., 2013), or simplified hydraulic approaches such as the inertial model (Bates et al., 2010).

Although some processes are represented empirically, the hydrological model has a strong physical basis, which strengthens the relationship between the parameters and the physical characteristics of the modelled basin. The model has been applied in several different South American river basins, like the Amazon (Paiva et al., 2012, 2013), the Grande River (Tucci et al., 2008; Bravo et al., 2009) and the Uruguay River (Collischonn et al., 2005). The model is also currently being used operationally for streamflow forecasts in the Paranaíba river basin (Collischonn et al., 2007a,b), the Pelotas river basin (Fan et al., 2014a), the Canoas river basin (Fan et al., 2014a), and for the three basins used as test-beds in the present study.

Table 1
Performance indicators obtained in the calibration at some representative stations located in each basin.

Basin	Station name	Lat (S, W)	Long (S, W)	Area (km^2)	NS	NS log	dV (%)
Tocantins	Goiatins	−7.71	−47.31	10,101	0.60	0.71	−9
	Tupiratins	−8.39	−48.11	244,477	0.78	0.74	−4
	Porto Real	−9.31	−47.93	44,321	0.80	0.81	5
	Porto Jeronimo	−11.76	−47.84	10,219	0.70	0.81	−2
	Ponte Paranã	−13.43	−47.14	30,111	0.84	0.88	−1
	Fazenda Santana	−12.64	−47.88	40,742	0.81	0.86	4
	UHE Estreito	−6.59	−47.46	287,800	0.80	0.88	10
	UHE Lajeado	−9.76	−48.37	184,219	0.73	0.84	10
	UHE Cana Brava	−13.40	−48.14	50,975	0.67	0.84	−9
	UHE Serra da Mesa	−13.84	−48.30	57,777	0.65	0.84	−8
	UHE Peixe Angical	−12.25	−48.35	125,687	0.72	0.79	8
	UHE São Salvador	−12.74	−48.24	61,298	0.69	0.85	−10
Upper São Francisco	Ponte Nova do Paraopeba	−19.95	−44.31	5784	0.77	0.89	−10
	Porto Mesquita	−19.17	−44.70	10,450	0.81	0.93	−4
	Ponte dos Vilelas	−20.40	−44.63	2619	0.80	0.86	8
	Porto Pará	−19.29	−45.11	11,358	0.56	0.63	−15
	Iguatama	−20.17	−45.72	5485	0.88	0.90	−2
	Porto Andorinhas	−19.28	−45.29	14,244	0.86	0.90	−7
	Porto Indaiá	−18.68	−45.63	2208	0.50	0.79	6
	UHE Três Marias	−18.19	−45.25	51,098	0.84	0.74	3
	Major Porto	−18.71	−46.04	1216	0.64	0.28	−31
Doce	Fazenda Ouro Fino	−19.17	−42.83	6360	0.65	0.48	−31
	Fazenda Meloso	−19.08	−42.88	2190	0.57	0.47	−11
	Ponte Nova Jusante	−20.38	−42.90	6230	0.66	0.36	−23
	Mário de Carvalho	−19.52	−42.64	5270	0.55	0.24	−30
	Cenibra	−19.33	−42.40	24,200	0.85	0.63	−23
	Salto Grande	−19.17	−42.78	6530	0.63	0.49	−36
	Governador Valadares	−18.88	−41.95	40,500	0.75	0.61	−30
	Vila Matias Montante	−18.58	−41.92	9770	0.41	0.20	20
	UHE Aimorés	−19.46	−41.10	62,400	0.90	0.79	2

For the river basins described here the model was calibrated considering hourly time-steps using the rainfall and streamflow data of the observation networks briefly described before. Also, for the Tocantins river basin, we used naturalized inflow data in the HPPs locations to compare results. The selected period for calibration is December 2006 to June 2011 for Upper São Francisco and Doce basins, and January 2008 to May 2013 for Tocantins basin. Table 1 shows the Nash–Sutcliffe model efficiency coefficient (NS), the Nash–Sutcliffe model efficiency coefficient for flows logarithms (NS log), and the volume error (dV) of the model after calibration for some representative stations located on each basin.

The model calibration in the streamflow gauges and HPPs locations shows a performance considered acceptable for forecasting purposes in all three basins, if one takes into account the low data availability. In the Tocantins basin, NS and NS log values varied generally between 0.6 and 0.85, where volume errors around −10% and +10%. In the Upper São Francisco basin the results were pretty much the same, with exception of the Porto Indaiá and Major Porto gauges, where we believe that the poor performance is related to some missing events in the sub-basins upstream. For the Doce basin, the performance was below the other basins, with NS generally varying from 0.57 to 0.85, NS log from 0.2 to 0.79, and some volume errors up to −30%. We mainly attribute these performances to the lack of information in the basins upstream. However, in the Aimorés HPP, which is the focus point for the model development and flow forecasting, the performance of the model was better (NS = 0.9; NS log = 0.79; dV = 2%) what makes the calibration acceptable for the final purposes.

In the context of real-time forecasting, we employed the MGB-IPH standard data assimilation procedure presented by Paz et al. (2007) to assimilate the gauging stations information. Reservoir inflow data was not, however, assimilated into the hydrological model. Therefore, to obtain inflow forecasts to the reservoir, we applied an output correction to the model results, based on an Auto Regressive (AR) model, correcting the forecasts values based on the last observed values prior to the forecast start.

3.3. Quantitative precipitation forecast (QPF) data

The QPF data used in this study was derived from the TIGGE (THORPEX Interactive Grand Global Ensemble) database. This initiative is described by Bougeault et al. (2010) and consists of a large effort to assemble a database of ensemble medium-range forecasts issued by different centres around the world, so that they stay available for conducting scientific research. Currently, the TIGGE portals have EPS data from about 10 forecasting centres, and are a great resource for research about ensemble forecasting benefits, such as the experiments presented here.

Among all EPS available on TIGGE, three were selected: ECMWF (European Centre for Medium-Range Weather Forecasts), GEFS (Global Ensemble Forecasting System issued by NOAA), and CPTEC (*Centro de Previsão de Tempo e Estudos* Climáticos global model, from Brazil). As a deterministic reference forecast for comparison we adopted the high resolution ECMWF also available in the database of TIGGE. It was selected for its known good performance compared to other forecasting systems (Buizza et al., 2005).

The ECMWF-EPS forecasts, described by Buizza et al., 2007, consist of 50 members of perturbed precipitation of approximately 0.5° resolution for the whole globe considering initial uncertainties by using a singular vectors technique, and model uncertainties due to physical parameterizations by a stochastic scheme. The data becomes available twice a day at 00:00 UTC and 12:00 UTC with a forecast horizon of 15 days and time steps of 6 h. The ECMWF deterministic forecast is available in the same condition, except to that the total lead-time is 10 days, instead of 15 days.

The GEFS forecasts, described by Toth et al. (2003), Toth and Kalnay (1993, 1997) and Wei et al. (2008), consist of 20 members of perturbed precipitation of approximately 1° resolution for the whole globe considering initial uncertainties by using the Ensemble Transform with Rescaling (ETR) technique (Cui et al., 2012). The data becomes available four times a day at (00:00, 06:00, 12:00, 18:00 UTC) with a forecast horizon of 15 days and time steps of 6 h. For comparison purposes with the two other models used in this study, only the forecasts produced at 00:00 and 12:00 UTC were used. It is also important to say that Hamill et al. (2013) presented a new version of the model that is not in the TIGGE database, and is not used in this study.

The CPTEC EPS forecasts, described by Cunningham and Bonatti (2011), Mendonça and Bonatti (2009) and Cavalcanti et al. (2002), consist of 14 members of perturbed precipitation of approximately 1° resolution for the whole globe considering initial uncertainties by using EOF-based perturbation (Zhang and Krishnamurti, 1999) and denominate. The data becomes available twice a day at 00:00 UTC and 12:00 UTC with a forecast horizon of 15 days and time steps of 6 h.

For the use in the hydrological model, all the model data were spatially downscaled to the watersheds by Thiessen polygons and disaggregated to hourly time steps.

Forecasts from the ECMWF and GEFS models were chosen because they are available for the entire globe, widely used in many operational forecasting systems (Thielen et al., 2009; Demargne et al., 2014; Alfieri et al., 2013; Thiemig et al., 2014), and have a good performance in predicting rains as shown by several recent studies (Hamill et al., 2000; Wei and Toth, 2003; Buizza et al., 2005; Bowler, 2006; Leutbecher and Palmer, 2008; Park et al., 2008a,b; Bougeault et al., 2010).

CPTEC forecasts were chosen because they are generated by the Brazilian centre, and thus one of the first options to be considered for hydrological forecasting systems to be developed in the national scene.

We do not apply bias correction to the NWP models rainfall. The most important restriction for bias correction in the present case was the data availability. Good observations from the telemetric network only are available recently starting in the year 2005, and data from the NWP models is only available from October 2006, making the periods short for an adequate analysis.

3.4. Hindcasting experiments setup

The experiment we conducted first consisted of setting up the calibrated hydrological models to be fed with QPF from the EPS obtained in the TIGGE database. This creates a Hydrological Ensemble Prediction System (HEPS) that runs based on meteorological uncertainties. These HEPS were run retrospectively for a period of the past defined by the data available for each basin as hindcasting experiments.

Besides the EPS forecasts, we also processed two reference forecasts in order to compare the ensembles; the ECMWF deterministic and the perfect knowledge of observed rainfall. This means that we processed hindcasts using the ECMWF deterministic reference forecast and hindcasts using observed rainfall in the forecast horizon, and both were used as reference for comparisons. In these later tests, the rainfall predicted for the next 15 days, (forecast horizon), was assumed to be equal to the rainfall that was effectively captured by the telemetric network. Often this type of test is also called “perfect” rainfall, because observations are used instead of precipitation predicted by models. With it, one can evaluate the system’s response to their predictions considering what actually precipitated over the forecast horizon. However, it is important to note that even the observed rainfall is not actually perfect, since sparse telemetric networks are often not able to perfectly capture the correct spatial distribution of rainfall or its correct volume throughout the basin.

Finally, a sixth option for issuing forecasts was evaluated in the present study. This option consisted of the use of all EPS data together, as a *grand ensemble*, or a *super ensemble* as was denoted here.

For the composition of the Super Ensemble we assigned equal weights to all forecast members, independent of the centre or the original number of members provided by each centre, resulting in an 84 member ensemble. This was the same composition procedure adopted by He et al. (2010) and is one of the suggestions of Park et al. (2008a,b). With this assumption, it is expected that the ensemble with more members will have a great influence on the results, but the investigation about multiple ways of combining forecasts is beyond the scope of the present work, and so we adopted the same procedure already tested by the cited authors.

We used the following terminology to identify the forecasts on our tests:

- ECMWF ensemble forecasts: ECMWF-pf
- ECMWF deterministic forecasts: ECMWF-fc
- GEFS ensemble forecasts: GEFS
- CPTEC ensemble forecasts: CPTEC-pf
- Forecasts using observed precipitation: Observed Precipitation
- Combination of ensembles forecasts from the three centres: Super Ensemble.

3.5. Forecasts assessment

Forecast results were assessed for the three major dams located in the downstream area of each studied watershed. In this case, the evaluated locations were Três Marias HPP (Upper São Francisco), Aimorés HPP (Doce river), and Estreito HPP (Tocantins River).

The first analysis conducted was a visual assessment of the forecasts, where the major flood in the rainy season of 2011/2012 was selected as an example to be shown to the three basins, since it was the major flood in the period analyzed for the three cases.

The second analysis conducted was an overlook of the ensembles spread throughout the entire investigated period, to show how the theoretical uncertainties given by the different EPS data behave in the basins. This assessment was based on the simple measure of standard deviation of ensembles.

The third assessment was based on common ensemble evaluation metrics. We assessed the hydrological forecasts generated to the case studies using six different metrics. The ensemble average was evaluated using Mean Absolute Error (MAE). Regarding ensemble distribution and spread, we computed the Mean Continuous Ranked Probability Score (Mean CRPS) and the Rank Histogram. To compare errors relative to issuing discrete events we computed the Brier Score (BS). The model calibration was evaluated using Reliability Diagrams, and the consistency of forecasts was evaluated using the Forecast Convergence Score (FCS).

For each of the catchment case studies, we adopted a different exceedance threshold, based on the HPPs operational constraints and attention limits. For the Três Marias HPP, the value was 1400 m^3/s, for the Aimorés HPP the value was 1600 m^3/s and for the Estreito HPP the value was 7200 m^3/s.

Details about metrics interpretation are given hereafter. Further descriptions of the metrics computation and its mathematical basins can be found in Brown et al. (2010), Bradley and Schwartz (2011), Hersbach (2000), Jolliffe and Stephenson (2012), Stanski et al. (1989) and Wilks (2006).

Mean absolute error: In the case of ensemble forecasts, to calculate the MAE, the mean average value of ensemble members at each lead-time has to be computed first. Then, the absolute difference between this average value and the corresponding observation is calculated. The value of MAE from a perfect model would be equal to zero.

Mean continuous ranked probability score: The CRPS is a score that summarizes the quality of a continuous probability forecast into a number by comparing the integrated square difference between the cumulative distribution function of forecasts and observations. The average CRPS across all pairs of forecasts and observations leads to the used Mean CRPS. It is usually considered the probabilistic equivalent of the MAE, since it reduces the mean absolute error for deterministic forecasts, and allows the comparison of results between probabilistic and deterministic forecasts. Lower values of mean CRPS correspond to better results.

Rank histogram: Is used to evaluate the spread of ensembles. It consists in a simple percentage computing of cases where observed values are placed between the ensemble forecast members within all forecasts and lead-times. Each position between each forecast member is denominated a bin, and the number of bins is equal to the number of members in the ensemble forecast plus one. In the end, the resulting histograms give a measure of the forecast spread. A perfect spread set of forecasts would produce a flat uniform rank histogram. High probabilities in both tails ("U" shape) of the rank histogram are an indicative of lack of spread. An inverted "U" shape rank histogram is an indicative of excessive spreading.

Brier score: The BS measures the average square error of a probability forecast for a dichotomous event, defined by a flow threshold exceedance, for example. Error units are given in probabilities. A perfectly sharp set of forecasts will have resulting BS values equal to zero.

Reliability diagrams: The reliability diagram also gives information about the set of forecasts performance for a discrete event, such as a flow threshold exceedance. In the *x*-axis of the diagram, the probabilities that this event is issued by the ensemble forecasting system are plotted. Usually, the probabilities are given by discrete pre-defined classes (0–20%, 20–40%, . . ., 80–100%, for example). In the *y*-axis of the diagram, the conditional probabilities of the real event occurrence are plotted, given the forecasted probabilities defined in the *x*-axis. According to the theory behind the reliability diagram, a defined event should be observed to occur with the same probability as its forecast probability of occurrence over a large sample. This means that a perfect reliability diagram should be a forty five degrees line. Deviations from this line represent different kinds of inaccuracy in the forecast. It is noteworthy that the reliability diagram results are subject to sampling uncertainty. We believe that in the present study the five-year sample used is adequate for a metric result. In addition to the reliability diagram it is common to display a histogram with the number of forecasts that fall within each of the pre-defined classes of forecast probabilities. This histogram is called a sharpness histogram and a desirable format would be U-shape, indicating little uncertainty in the threshold occurrence indication, although this does not specifically mean a better performance.

The Forecast Convergence Score based on the Brier Score (FCSbs) describes the consistency of two sequential forecasts in terms of threshold detection (Pappenberger et al., 2011a,b). Values near to zero indicate more consistent decisions. The FCSbs, as discussed by Pappenberger et al. (2011a,b), is therefore not a performance metric, as it does not use comparisons with observations. It is a descriptive metric that addresses one useful characteristic of forecasts, which is the capability of issuing consistent consecutive forecasts. This characteristic is especially important in reservoir short-term operations when the reservoir is also utilized for flood control, which is the case for the Três Marias reservoir. More consistent forecasts permit an easier decision making process when related to opening spillway gates or increasing generation to prevent damages.

4. Results assessments

Illustrative examples of forecasts issued for each verification point are given in Figs. 3–5. These figures show examples of forecasts issued at flood situations in the basins.

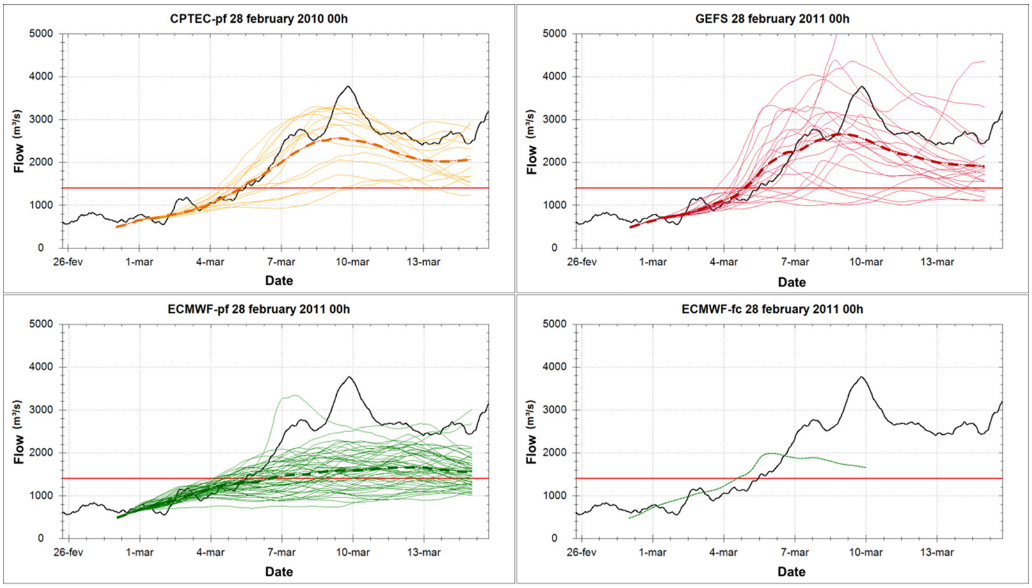

Fig. 3. Forecasts for Três Maras HPP (Upper São Francisco river) issued, respectively with CPTEC-pf, GEFS, ECMWF-pf, and ECMWF-fc on 28 February 2011 00 h. The dark line shows the observations and coloured lines show the forecasts. Ensemble mean is highlighted with a dashed line and the 1400 m^3/s threshold with a horizontal red line.

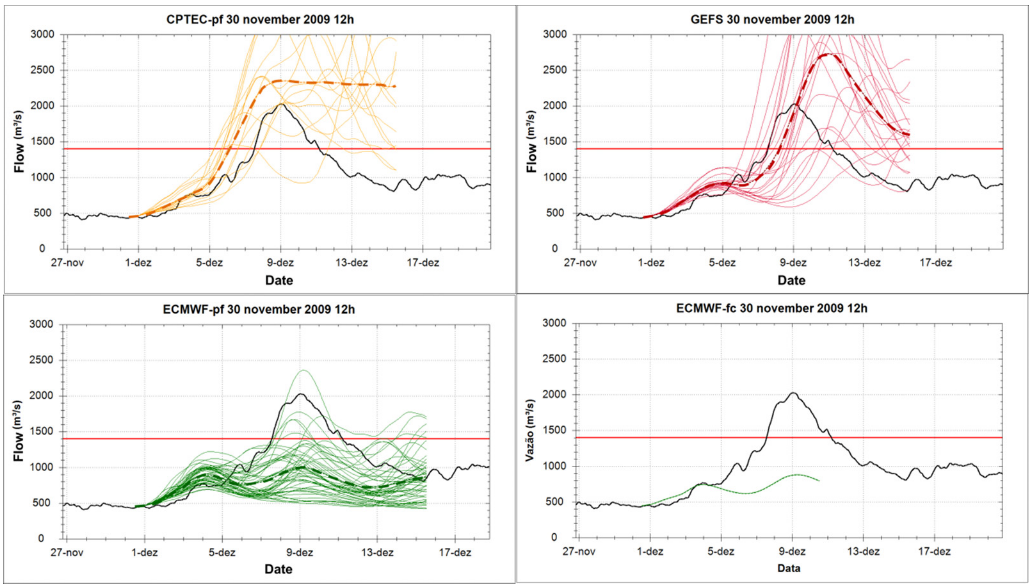

Fig. 4. Forecasts for Aimorés HPP (Doce river) issued, respectively with CPTEC-pf, GEFS, ECMWF-pf, ECMWF-fc on 30 November 2009 12 h. The dark line shows the observations and coloured lines show the forecasts. Ensemble mean is highlighted with a dashed line and the 1600 m^3/s threshold with a horizontal red line.

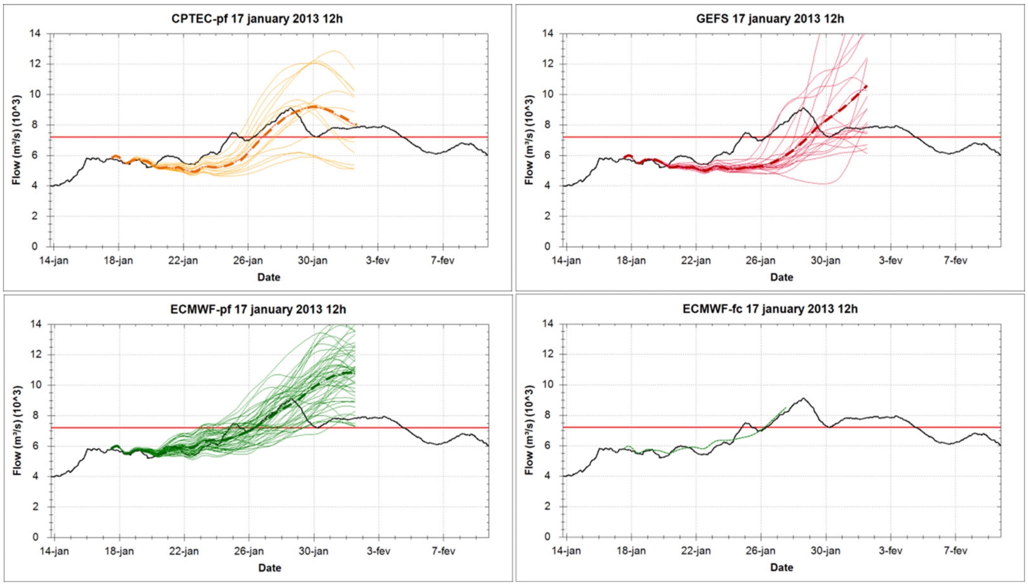

Fig. 5. Forecasts for Estreito HPP (Tocantins river) issued, respectively with CPTEC-pf, GEFS, ECMWF-pf, ECMWF-fc on 01 January 2013 12 h. The dark line shows the observations and coloured lines show the forecasts. Ensemble mean is highlighted with a dashed line and the 7200 m^3/s threshold with a horizontal red line.

The first sequence (Fig. 3) presents a forecast for Três Marias HPP, when the flow reached almost 2000 m^3/s on 08 March 2010. On 28 February, when the forecast was issued, the inflow was near 400 m^3/s. CPTEC-f was not very successful in predicting the 1400 m^3/s threshold and the peak, since all members under forecasted the event. GEFS missed the earlier peak and threshold overcoming, but from 5 March it performs well with several members crossing above a threshold of 2000 m^3/s, matching the observed. ECMWF-pf encompassed the observations well and indicated the surpassing of the threshold with the upper members of the ensemble, although some members indicated higher flows than observed. The deterministic reference forecast (ECMWF-fc) indicated the increase in the flow, but did not reach the threshold limits and missed the greater observed peak. In this sense, it is possible to say that GEFS and ECMWF-pf forecasts gave more information about the event than the deterministic forecast.

The second sequence (Fig. 4) presents a forecast for Aimorés HPP in an event that inflow reached 2000 m^3/s on 09 December 2010. On 30 November, when the forecast was issued, the inflow was 450 m^3/s. In this forecast, CPTEC-pf clearly indicated a rise in streamflow with all its members, but the threshold exceedance was 2 days early and peak inflows were overestimated by most members. GEFS also clearly indicated an increasing flow and the threshold exceedance was identified in time by some members of the ensemble, but the peak flow was overestimated by most of the ensemble members and the timing was shifted to around 2 days ahead. ECMWF-pf, on the other hand, did not clearly indicate the inflow increase, but some upper members issued the threshold exceedance and peak inflow with the correct timing. ECMWF-fc deterministic forecast did not follow observations after 5 December, giving no information about the threshold exceedance and flow peak.

The third sequence (Fig. 5) presents a forecast for Estreito HPP in an event that inflow reached values near 9000 m^3/s on 29 January 2013. On 17 January, when the forecast was issued, the inflow was 6000 m^3/s. CPTEC-pf forecast upper members indicated the occurrence of the threshold exceedance in time, and peak flow was well forecasted by the central tendency of the ensemble. GEFS also clearly indicated an increasing flow, but the threshold exceedance was a little late for most of the members, and the forecasted peak was greater and later than the observations. ECMWF-pf shows a good agreement with observations from all members until 29 January, after which the majority of members

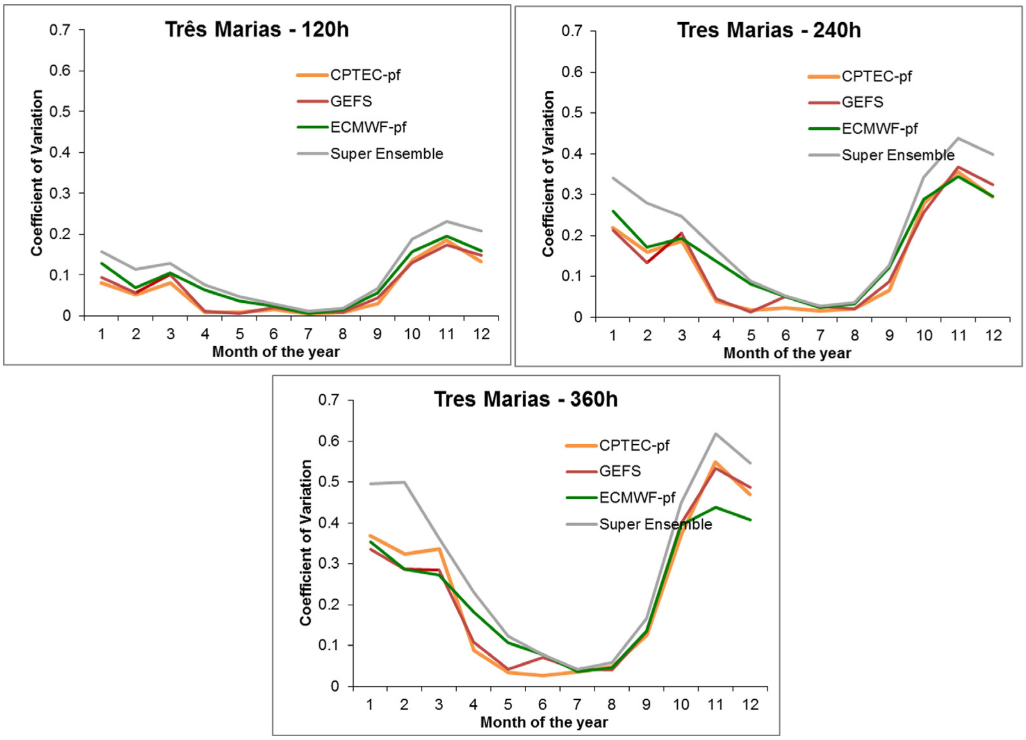

Fig. 6. Monthly analysis of the coefficient of variation (standard deviation divided by mean) among the members of the EPS for Upper São Francisco river.

show over forecasted inflows. The ECMWF-fc deterministic forecast for this example shows an almost perfect forecast, following observations nearly.

In Figs. 6–8, an analysis of the coefficient of variation (standard deviation divided by mean), among the members of the EPS according to the month in which the forecast was issued, is presented. The values shown for each month represent the mean coefficient of variation computed for that month over the whole tested period at each location.

For the three basins, the “U” shape of the curves shows that for predictions for the months of May to September (dry months of the tropical climate), the coefficient of variation is smaller, possibly because all ensemble members generally predict more rainfall near to zero. From October to April, the rainy season in the watersheds, values are however, higher. This, in a broader sense, may also indicate that for the dryer months the quality of forecasts is more dependent on initial conditions and the hydrological model, and less on the EPS. Further studies about the relationship of initial conditions and meteorological forcing for large watersheds can be found in the works of Paiva et al. (2012) and Candogan-Yossef et al. (2013).

All figures also show an increase in the coefficient of variation with the increase of the forecasts horizon. For example, at lead time 120 h and January, values for all basins stood around 0.1–0.2. And at the lead times of 360 h for January deviations ranged from 0.2 to more than 0.5. This increase is expected due to the increased uncertainty of the predictions with increasing forecast lead time.

For Upper São Francisco river (Três Marias HPP), the coefficient of variation between the EPS was very similar at all lead times. Exceptions are the values from ECMWF-pf, which are relatively higher in April and May, and the greater values of GEFS and ECMWF-pf from October to December at lead time 360 h. The Super Ensemble always had a greater coefficient than the different EPS from which it

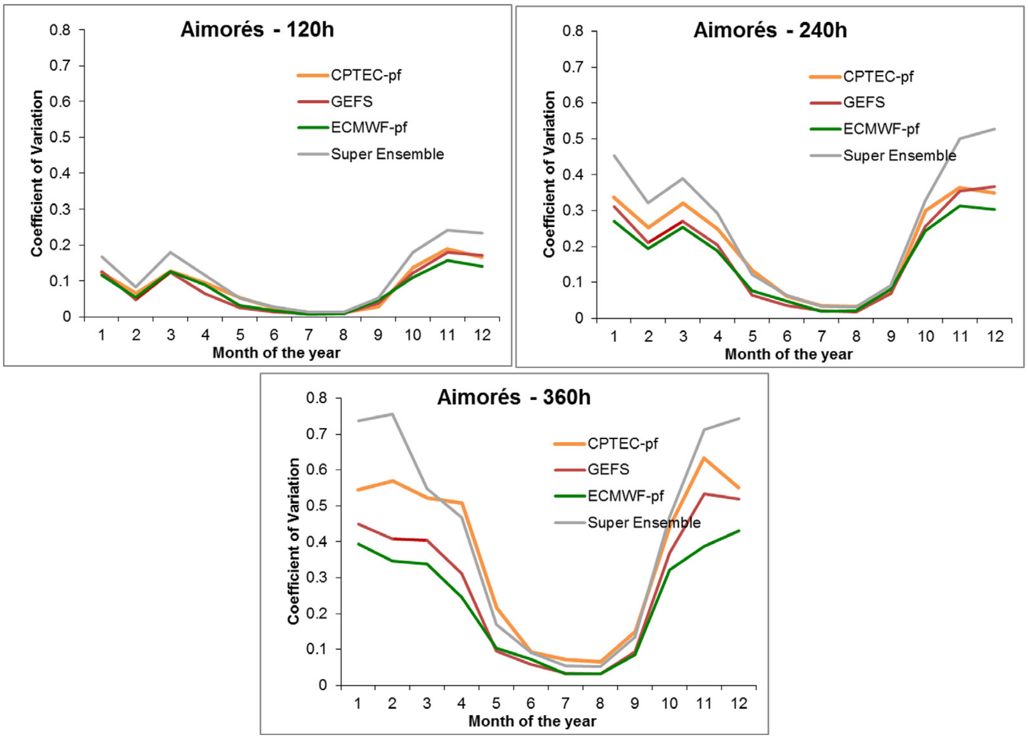

Fig. 7. Monthly analysis of the coefficient of variation (standard deviation divided by mean) among the members of the EPS for Doce river.

is composed. This suggests that the composition creates a wider forecast with a greater spread. This is also to be expected, since more uncertainties are covered by the Super Ensemble.

For Doce River (Aimorés HPP), the coefficient of variation between the EPS was similar at 120 h lead time. With increasing lead time, CPTEC-pf showed greater values, followed by GEFS and then ECMWF-pf. Again, the Super Ensemble suggested a greater coefficient than the EPS that compose it at almost all lead times and months (except relative to CPTEC-pf at some months and 360 h). We believe that this happens because the composition creates a wider forecast with a greater spread, and that more uncertainties are therefore covered by the Super Ensemble.

For Tocantins River (Estreito HPP) at 120 h lead time, the coefficients of variation of all EPS are very near zero, indicating a small spread in the forecasts. At the two greater lead times, CPTEC-pf showed relatively smaller values from October to April, followed by GEFS and ECMWF-pf depending on the month. Super Ensemble values suggested slightly greater coefficient than the individual EPS in the majority of lead times and months.

The general analysis of results between basins also indicates that the larger watershed (Tocantins) usually generates forecasts with a narrower spread than the relatively smaller watersheds, (Upper São Francisco and Doce).

Assessments of ensembles performance using metrics are shown hereafter. First, results of Mean Absolute Error (MAE) for the three tested locations are shown in Fig. 9.

As expected, all MAE results increase with the lead time as uncertainties also increase. In the initial lead times (between 24 h and 48 h) values of MAE are practically equal between the different tested data. This happens because at these lead times results are more dependent on observed values than on forecasted ones, due to the watershed concentration time, data assimilation, and error correction.

After these lead times, results using the different rain inputs start to differ. In the Três Marias HPP, the EPS have comparable results until lead time 120 h. After this lead time, CPTEC-pf shows a greater

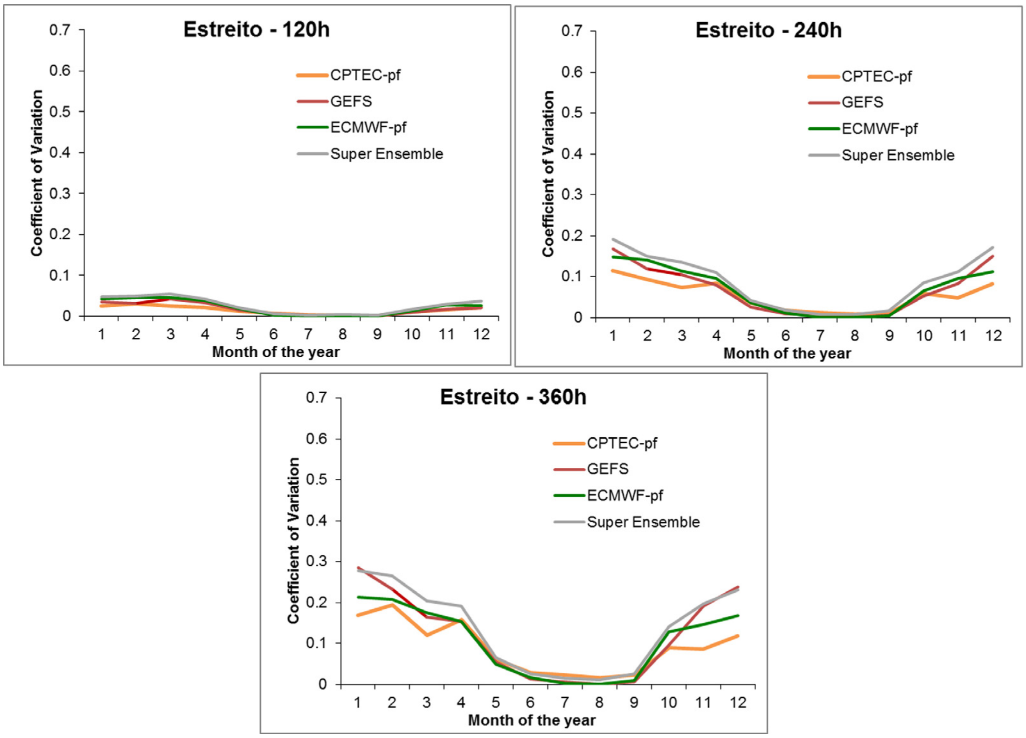

Fig. 8. Monthly analysis of the coefficient of variation (standard deviation divided by mean) among the members of the EPS for Tocantins river.

error than the other EPS, reaching 450 m^3/s (360 h), which is twice the error obtained with observed precipitation 150 m^3/s. The EPS with a better performance for Três Marias is ECMWF-pf, with an error reaching maximum 200 m^3/s, suggesting a curve near to the one with observed precipitation.

Also for Três Marias HPP, the comparison between the EPS data and the Super Ensemble data suggests that MAEs not lower than ECMWF-pf can be obtained by the union of all EPS. And the comparison between the EPS and the deterministic reference forecast indicated that the ensemble mean of ECMWF and the Super Ensemble lead to lower errors, but the ensemble means of GEFS and CPTEC-pf tend to have larger errors than the reference until lead time 240 h.

In the Aimorés HPP, the CPTEC-pf shows a greater error than the other EPS from lead time 120 h, reaching 500 m^3/s (360 h), which is twice the error obtained with other data sources. The EPS with a better performance for Aimorés is ECMWF-pf, with an error reaching a maximum of 180 m^3/s, where errors obtained using observed precipitation stayed around 100 m^3/s. The comparison between the EPS data and the Super Ensemble data showed that lower MAEs can be obtained by the union of all EPS in comparison to the individual ones, probably because the union compensates for errors in the ensemble mean. The comparison between the ensembles and the deterministic reference forecast indicates values very similar to GEFS errors, which are lower than CPTEC-pf, but higher than ECMWF-pf and the Super Ensemble.

In the Estreito HPP (Tocantins River), similar errors are observed between the EPS until lead time 120 h. After 120 h lead time the errors increase faster (in a 3-step shape). This shape occurs due to the consideration that upstream reservoirs perfect outflow is only known for three days ahead at every forecast, and for the remaining lead times the outflows were considered as “run of the river”.

From lead time 120 h onwards, CPTEC-pf shows a larger error than the other EPS, reaching values greater than 1200 m^3/s (after lead time 288 h). Other EPS, the deterministic forecast and

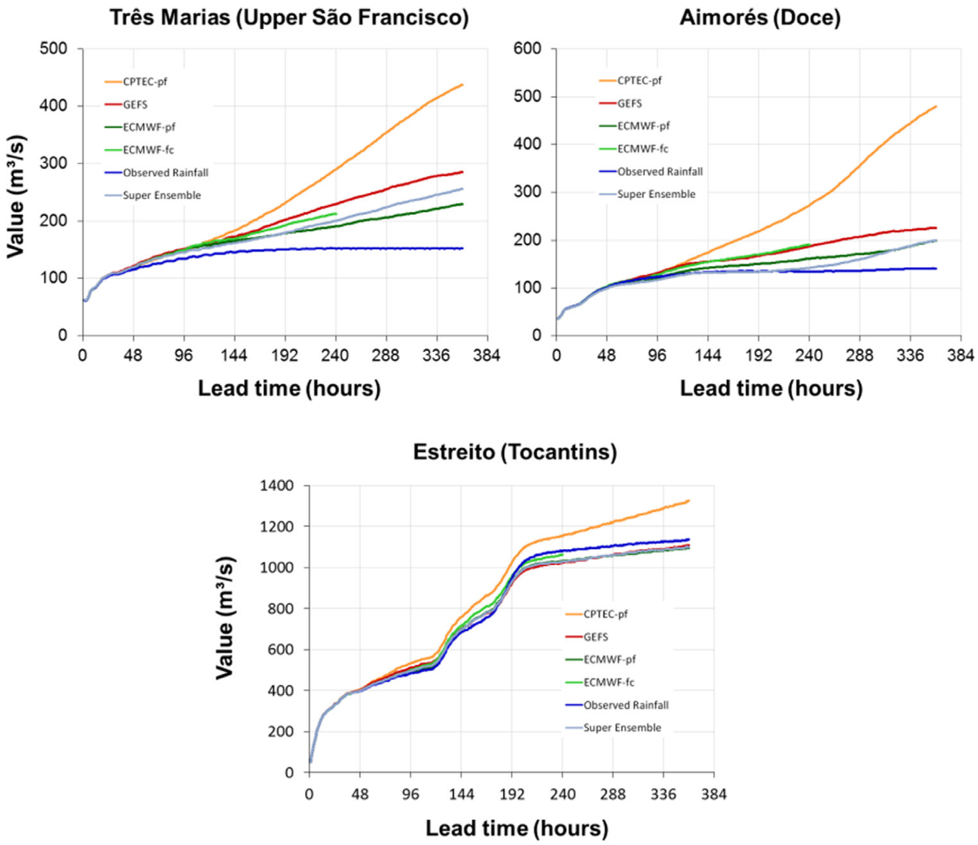

Fig. 9. Mean Absolute Error (MAE) results for the three tested locations.

Super Ensemble showed a very similar performance, including some smaller error using the EPS in comparison to forecasts using the observed precipitation on greater lead times.

Results of mean CRPS for the three tested places are shown in Fig. 10. The Mean CRPS is sometimes considered the MAE of ensembles (in fact it is equivalent to the MAE for a deterministic forecast), and its results also increase as expected with lead time and uncertainties. In the initial lead times (again between 24 h and 48 h) values of mean CRPS are very similar for the different tested data. And this happens due the greater dependency to observed values than to forecasted ones.

After the initial lead times cited above, mean CRPS results using the different rain inputs start to differ. In the Três Marias HPP, CPTC-pf has the lower value between the EPS until lead time 120 h. However, from this horizon it shows a greater error than the others. ECMWF-pf was the EPS with lower errors from lead time 120 h, including a better performance than the observed precipitation until lead times near 260 h. The comparison between the EPS data and the Super Ensemble data suggests that the union of all EPS has the best performance. In addition, the comparison between the EPS and the deterministic reference forecast always indicated a better performance for all ensembles.

At Aimorés HPP, all EPS had a very similar performance until lead time 120 h. From lead time 120 h onwards, the CPTEC-pf shows a greater error than the other EPS, reaching values near to 300 m^3/s (360 h), which is twice the error obtained with ECMWF-pf and observed precipitation. The EPS with a better performance for Aimorés is ECMWF-pf, but with results very similar to GEFS, including performances better than perfect forecasts values until lead times around 192 h. The Super Ensemble again indicated that the union of all EPS has the best performance of all, as the comparison between

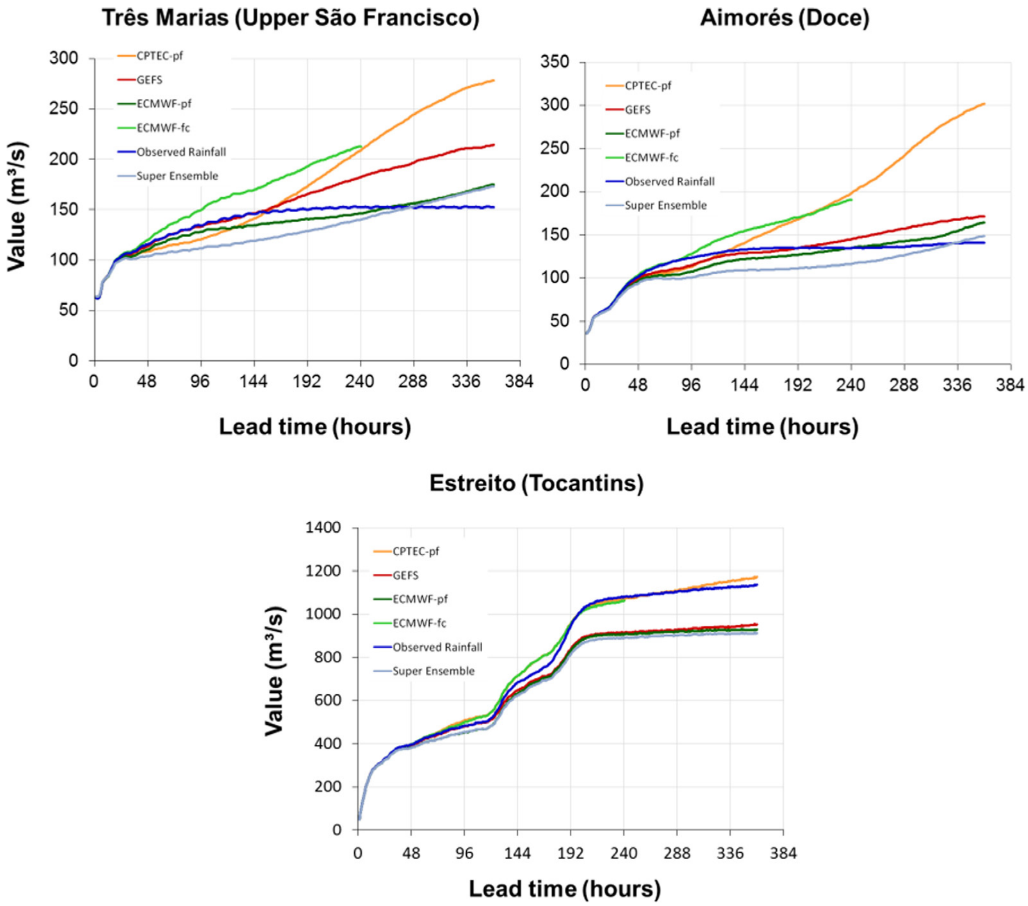

Fig. 10. Mean Continuous Ranked Probability Score (CRPS) results for the three tested locations.

the ensembles and the deterministic reference forecast always indicates a better performance for ensembles (except for CPTEC-pf after 192 h).

At Estreito HPP, results are split into two regions from lead times near 120 h. Results from GEFS, ECMWF-pf and Super Ensemble showed smaller errors below 1000 m^3/s at lead time 360 h. On the other hand, the deterministic forecast (ECMWF-fc) and CPTEC-pf showed greater errors, reaching, respectively, 1050 m^3/s (240 h) and 1200 m^3/s (360 h). Forecasts using observed rainfall also showed errors in the same order of magnitude as ECMWF-fc and CPTEC-pf, which is considered a worse result than for GEFS, ECMWF. We believe that this is an effect directly correlated to the raw number of observations in the basin, which induces the non-detection of events.

Rank Histograms for the three tested locations are shown in Figs. 11, 12 and 13, for selected lead times of 120 h, 240 h, 360 h. Since the EPS have different number of members (that led a different number of bins), all the EPS bins were resampled into a total of ten bins, for comparison purposes between the models.

In all cases, the Rank Histograms from the tested EPS have a “U” shape. This is indicative of a lack of spread in the ensembles, as it is common for the observation to be located near the lowest or the highest members of the ensembles. In the Tocantins basin (Estreito HPP) case, the histograms are also affected by the upstream reservoirs operation, which lead to greater frequencies at bin number ten, indicating forecasts below observed inflows.

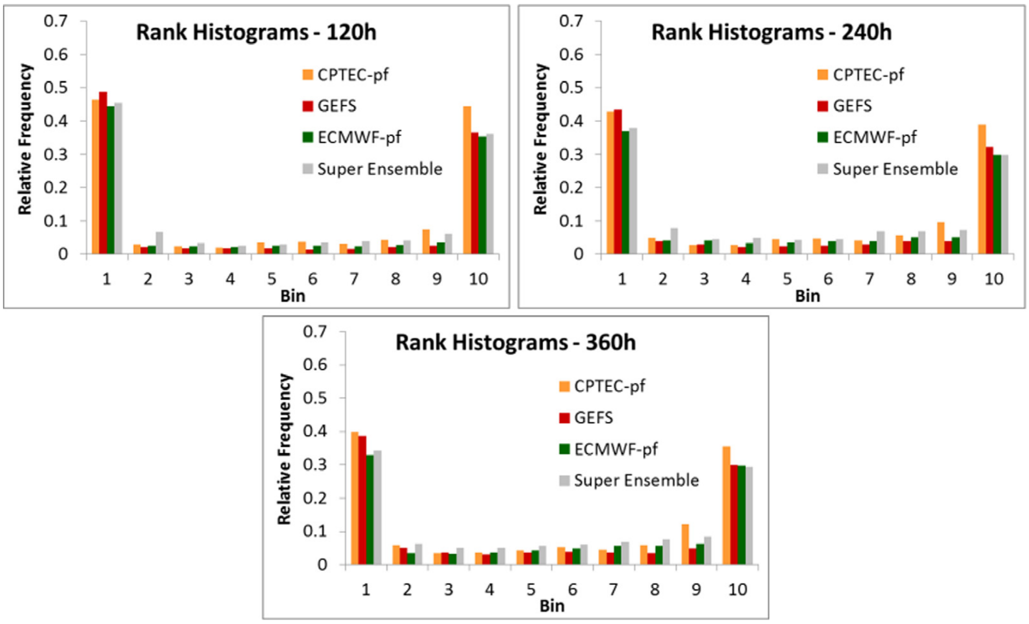

Fig. 11. Rank Histograms for three sampled lead times at Três Marias HPP.

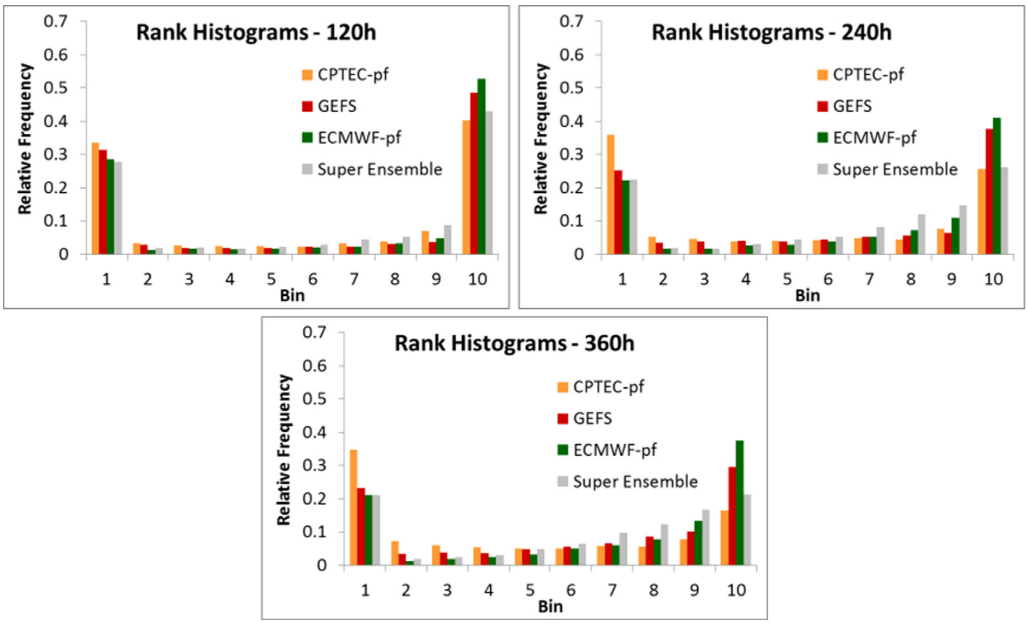

Fig. 12. Rank Histograms for three sampled lead times at Aimorés HPP.

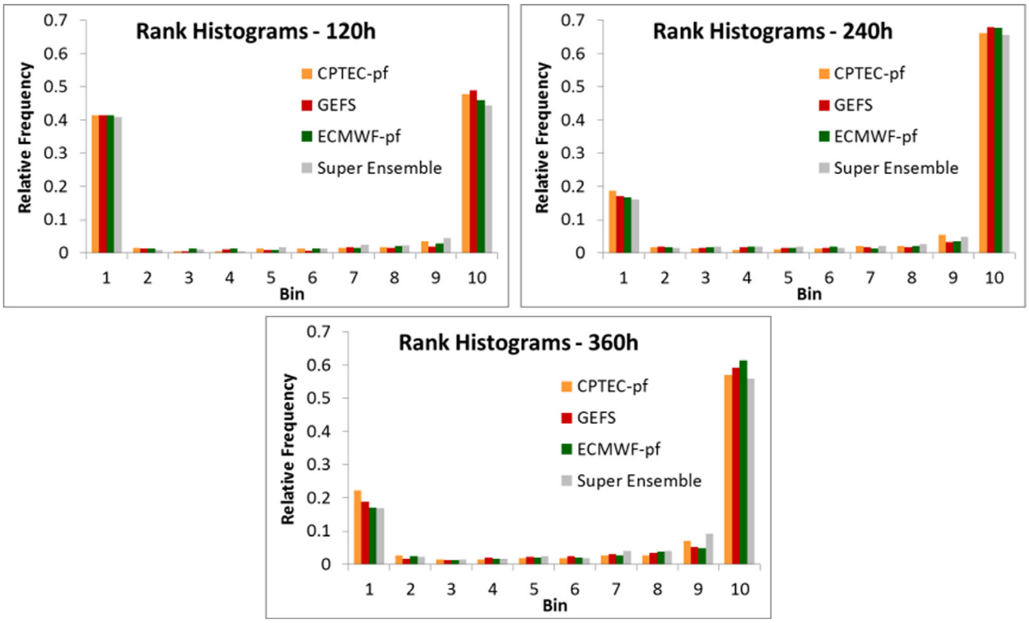

Fig. 13. Rank Histograms for three sampled lead times at Estreito HPP.

Results of Brier Score for the three tested locations are shown in Fig. 14. As in the two previous metrics, errors increase with lead time and uncertainties. In the initial lead times values of BS are similar between the different tested data.

At Três Marias HPP, all the EPS have similar BS values until lead time 144 h. Then, CPTEC-pf presents greater vales, near to 0.10 in the last lead time. ECMWF-pf shows the best performance amongst the EPS, very similar to the perfect forecast until lead time 240 h. The comparison between the EPS data and the Super Ensemble data suggests that the union of all EPS have a lower BS than the individual ones, even better than the ECMWF-pf. Also, the comparison between the EPS and the deterministic reference forecast indicated a better performance for ensembles until lead time near to 192 h, where after results of the deterministic forecast are better than the CPTEC-pf.

At Aimorés HPP results are similar until lead times near 120 h. From lead time 120 h onwards, the CPTEC-pf shows greater values, near to 0.09 in the last lead time, which is more than twice as big as the errors obtained with the others EPS (around 0.04) and the perfect forecast (around 0.02). The ECMWF-pf and the GEFS forecasts had a very similar performance for Aimorés HPP. In relation to the Super Ensemble, it is again indicated that the union of all EPS has the best performance of all. Also, the comparison between the ensembles and the deterministic reference forecast indicate a performance generally better for ensembles (exception for CPTEC-pf after 144 h).

At Estreito HPP, among the tested EPS, at early lead times (48–144 h) ECMWF-pf showed a performance slightly better than the others, with values around 0.035. GEFS showed the best performance at greater lead times (over 192 h) with values around 0.07, and CPTEC-pf the poorest one, which was also very similar to the deterministic forecast performance, with values over 0.08. The Super Ensemble showed a performance similar to ECMWF-pf at early lead times and between GEFS and ECMWF-pf at greater lead times. Surprisingly, the worse performance at greater lead times was the one obtained with forecasts driven by observed rainfall. This is a clear indication that flow peaks were missed by the observation network.

The Reliability Diagrams for the three tested locations are shown in Figs. 15, 16 and 17, for selected lead times of 120 h, 240 h, 360 h. The diagrams were generated using five classes 0–20%, 20–40%, 40–60%, 60–80%, and 80–100%.

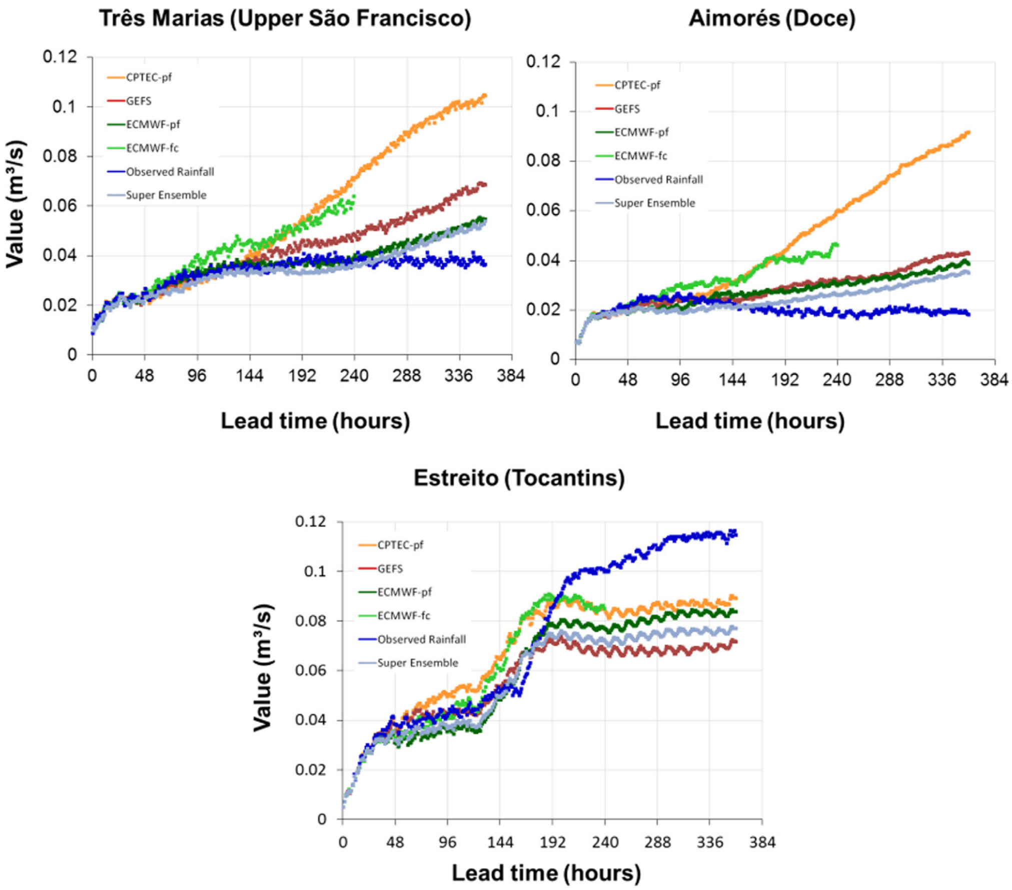

Fig. 14. Brier Score (BS) results for the three tested locations.

Três Marias HPP calibration diagrams suggest a positive bias in the calibration for the ECMWF-pf forecasts at the first lead time (120 h), as the points with greater probability of occurrence appear distinctly below the forty-five degrees line. For the greater lead times (240 h and 360 h), the model shows a good calibration, with its line always near forty-five degrees. In the CPTEC-pf and GFS calibration diagrams, results at all lead-times also point to positive type II bias in the forecasts, increasing with lead time. This calibration bias is more evident in the CPTEC-pf forecasts.

Results obtained with the Super Ensemble were very similar to the ones obtained with the ECMWF-pf forecasts, which is the EPS with the greatest number of members. In terms of sample counting, usually the first class of the forecasted probabilities appeared with a greater frequency than the others, but the greater classes usually also showed a number of samples between 50 and 100. Having fewer samples in the higher bins is not surprising when one consider that the lower probabilities are more likely to occur, since it is just a matter of some members indicating the threshold occurrence. Also, fewer samples occur in the higher bins due to the fact that almost all of the members would have to exceed this bin threshold.

Aimorés HPP calibration diagrams show a small negative type I bias for ECMWF-pf forecasts, with the curves above the forty five degrees lines, suggesting that the forecasting system underestimates the probability of occurrence of the threshold in comparison to its relative occurrence frequency. For the CPTEC-pf and GFS forecasts, the calibration diagrams results at all lead-times again pointed to

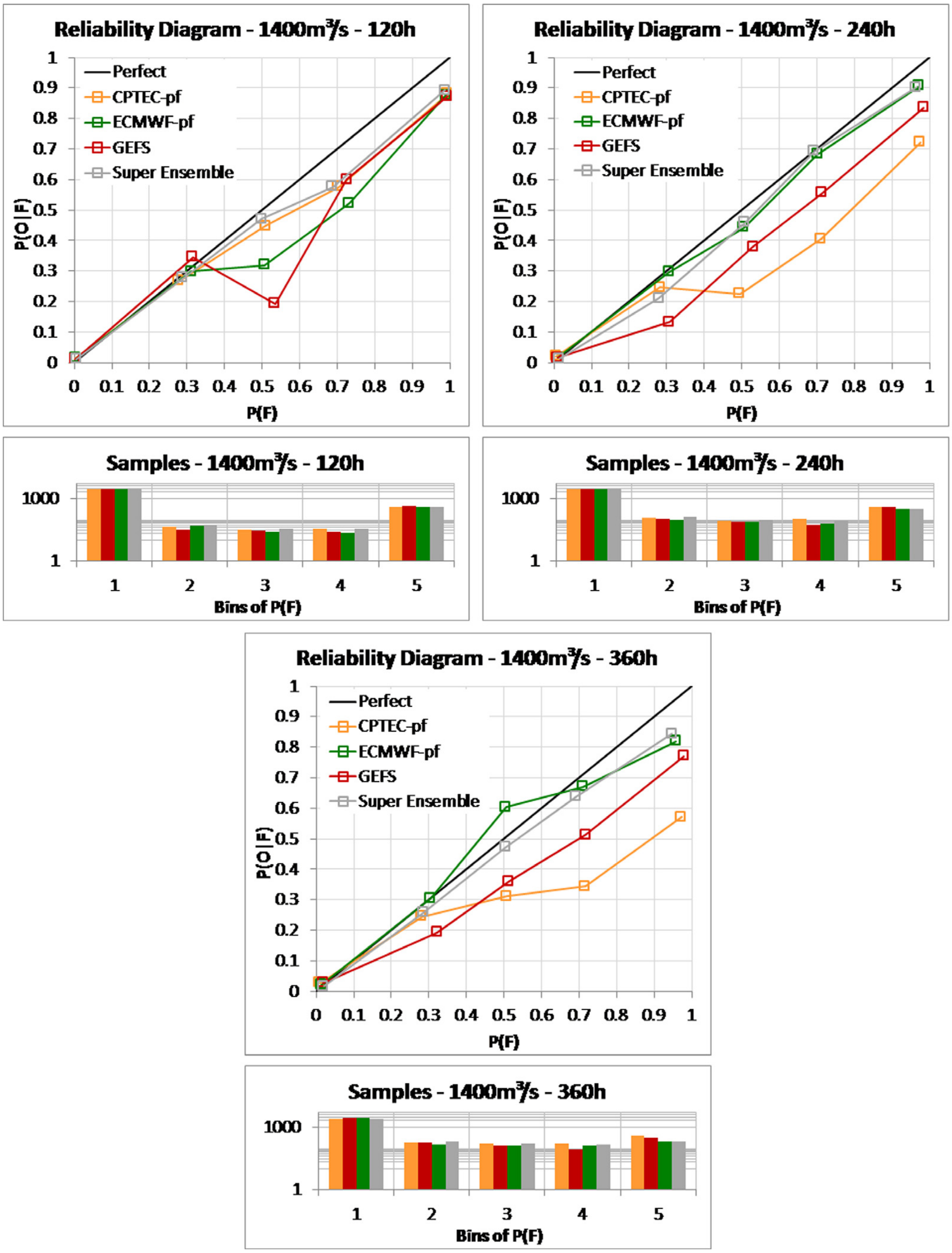

Fig. 15. Reliability diagrams for three sampled lead times at Três Marias HPP.

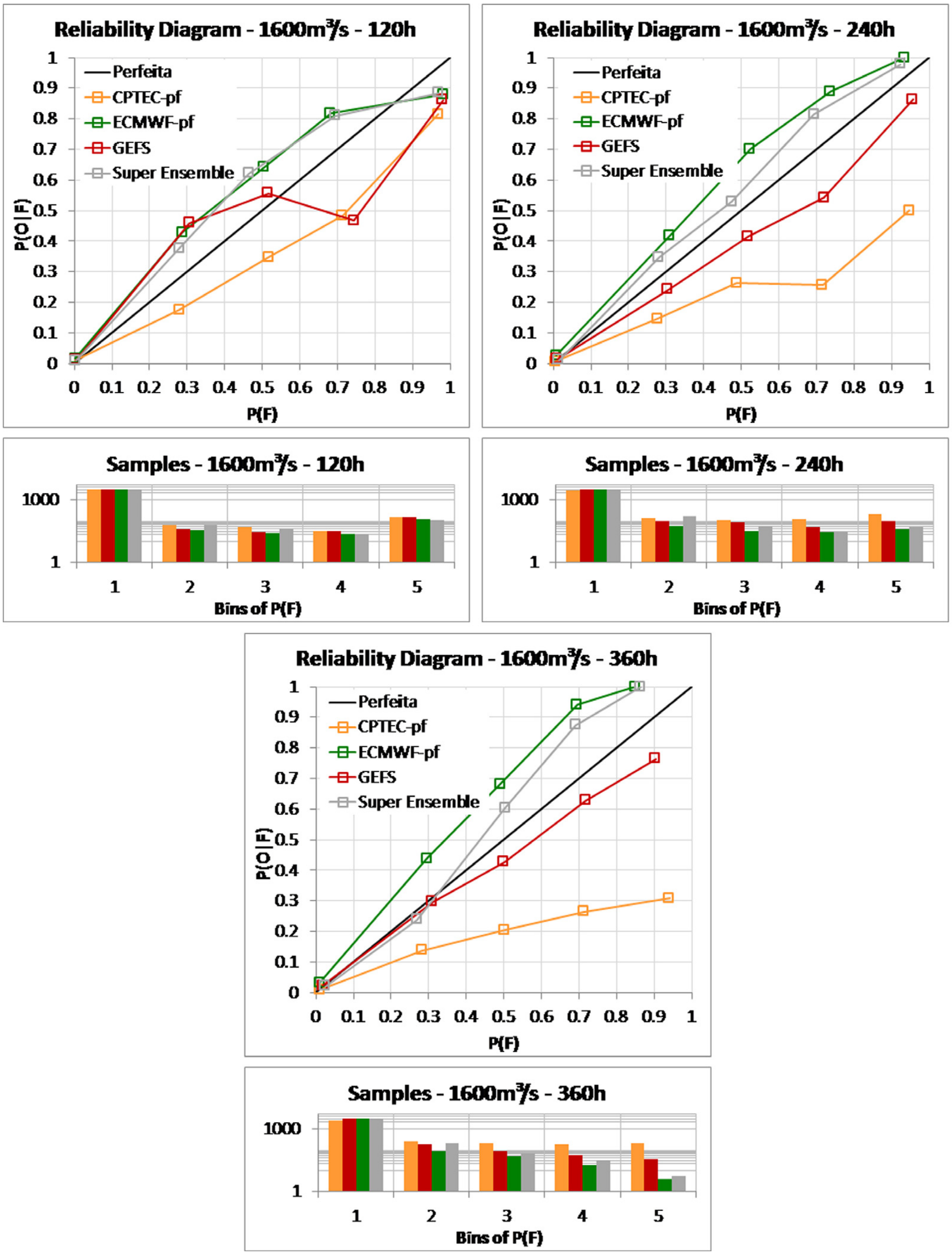

Fig. 16. Reliability diagrams for three sampled lead times at Aimorés HPP.

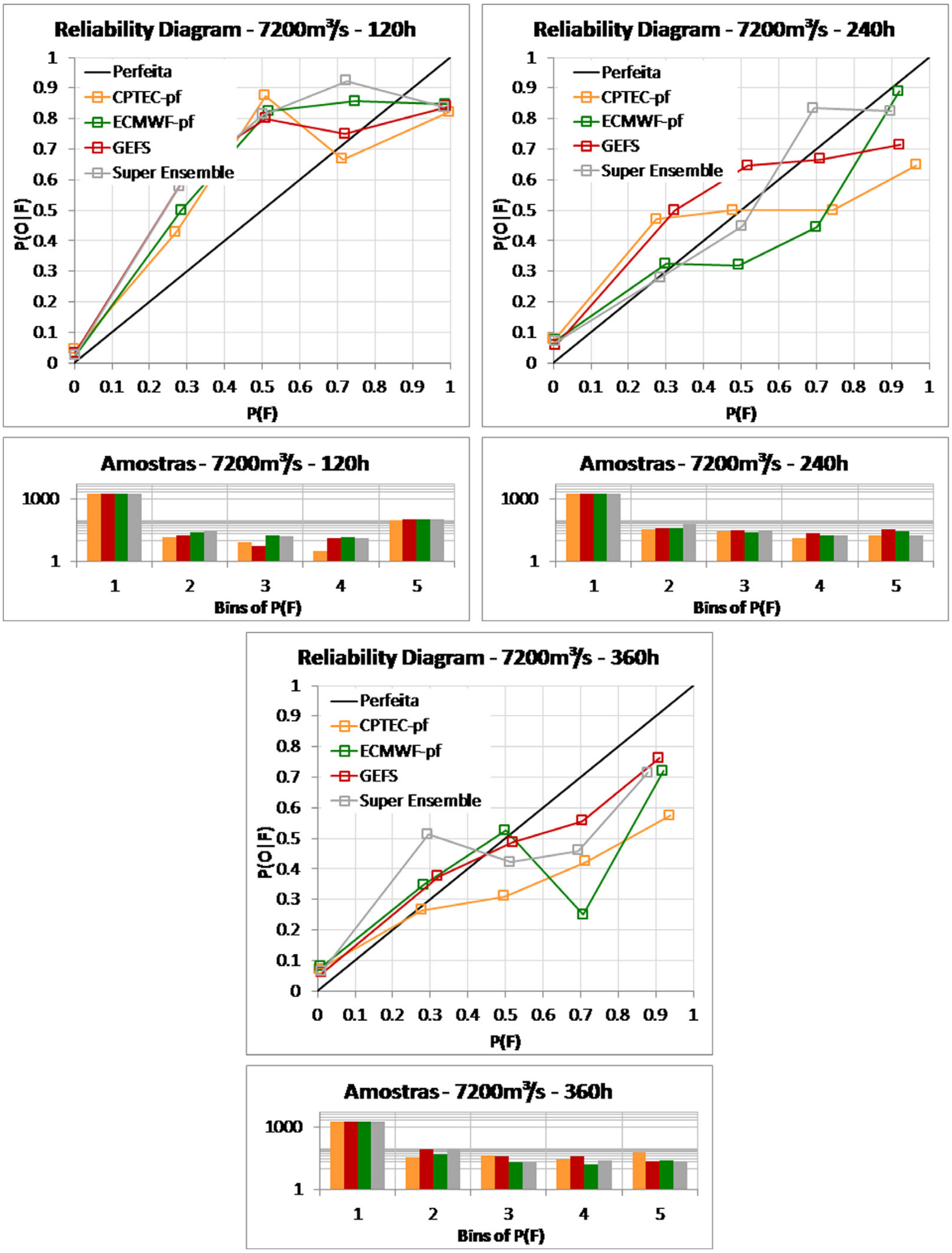

Fig. 17. Reliability diagrams for three sampled lead times at Estreito HPP.

positive type II bias in the forecasts, with this bias increasing with lead time. This calibration bias is more pronounced in the CPTEC-pf forecasts.

Super Ensemble results for Aimorés HPP had the same shape as the one obtained with the ECMWF-pf forecasts, but with smaller positive bias. In terms of sample counting, the first class of the forecasted probabilities appeared with a greater frequency than the others, but the greater classes usually showed a number of samples between 10 and 100, except ECWMF-pf forecasts at 360 h, which showed a smaller frequency.

Estreito HPP calibration diagrams for the earlier lead time (120 h) suggest a negative type I bias in the calibration of all models, although with an inverse tendency for forecasts forecasted with higher probabilities. At the mid lead time (240 h), ECMWF-pf indicated an overall positive type I bias tendency, while GEFS and CPTEC-pf indicated a positive tendency for forecasts issued with lower probabilities, and a negative tendency for forecasts emitted with lower probabilities. At the greater lead time, the EPS based forecasts mainly indicated a positive calibration bias for forecasts issued with probabilities higher than 0.2.

Furthermore, for Estreito HPP, results obtained with the Super Ensemble tended to follow intermediate values among the EPS. This, in general, allows for the Super Ensemble calibration to be considered slightly better than the individual EPS. In terms of sample counting, the first class of forecasted probabilities appeared more often with a greater frequency than the others, but the higher classes usually showed a number of samples greater than 50 (except for some intermediate classes at the first lead time).

Forecast convergence scores for the test-beds are shown in Fig. 18. On the vertical axis, the score values are shown and on the horizontal axis the paired lead times values. It is important to remark that this score measures a characteristic, and not the quality of forecasts, and therefore, lower values do not automatically indicate better performance.

For Três Marias HPP, the FCSbs results indicate increasing values for the deterministic forecast, GEFS forecasts, and for the CPTEC-pf forecasts, and decreasing values for ECMWF-pf forecasts with lead time. Also, values of ECMWF-pf were the lower ones among the EPS, the GEFS the higher ones, and the deterministic forecasts showed the highest FCSbs values among all tested data. This means that in general, the decision about the threshold exceedance changes less among the EPS (and less using ECMWF-pf among the EPS), than when using the deterministic one.

The Super Ensemble FCSbs values at Três Marias HPP were very similar to the ECMWF-pf ones, which decrease in time and indicates more consistency. Also the comparison with forecasts using observations indicates that ECMWF-pf and Super Ensemble were as consistent as the observations about the decisions.

For Aimorés HPP, increasing FCSbs values were verified for the deterministic forecast and for CPTEC-pf forecasts, and decreasing values with lead time were found for GEFS and ECMWF-pf forecasts. Again, ECMWF-pf values were the lowest ones among the EPS. The highest ones were from CPTEC-pf among the EPS, and the deterministic forecasts showed the largest FCSbs values among all the tested data. This again means that more often than not, the decision about the threshold occurrence changes less among the EPS (and less using ECMWF-pf), than when using the deterministic reference.

Super Ensemble results at Aimorés HPP were very close to ECMWF-pf in terms of FCSbs, indicating consistency between the consecutive decisions. In addition, the comparison with forecasts using observations indicates that ECMWF-pf (and perhaps also GEFS) are as consistent as forecasts driven by observations.

At Estreito HPP, the first important information is the one given by the shape of the metric results. At mid lead times (100–180 h), peaks occur in the metric values. Those peaks are a result of the consideration in the experiment of the reservoir's operation. As mentioned previously, at three days the outflow is switched from known upstream releases to 'run of the river', which caused the resulting peak in the metric graph. In addition, the deterministic forecast FCSbs values always indicate lower consistency among consecutive forecasts than any of the used EPS. Among the EPS, ECMWF-pf pointed to the highest consistency between forecasts, and this behaviour was followed by the Super Ensemble. In the comparison with observations, again ECMWF-pf and Super Ensemble can be considered to be as consistent as forecasts driven by observations.

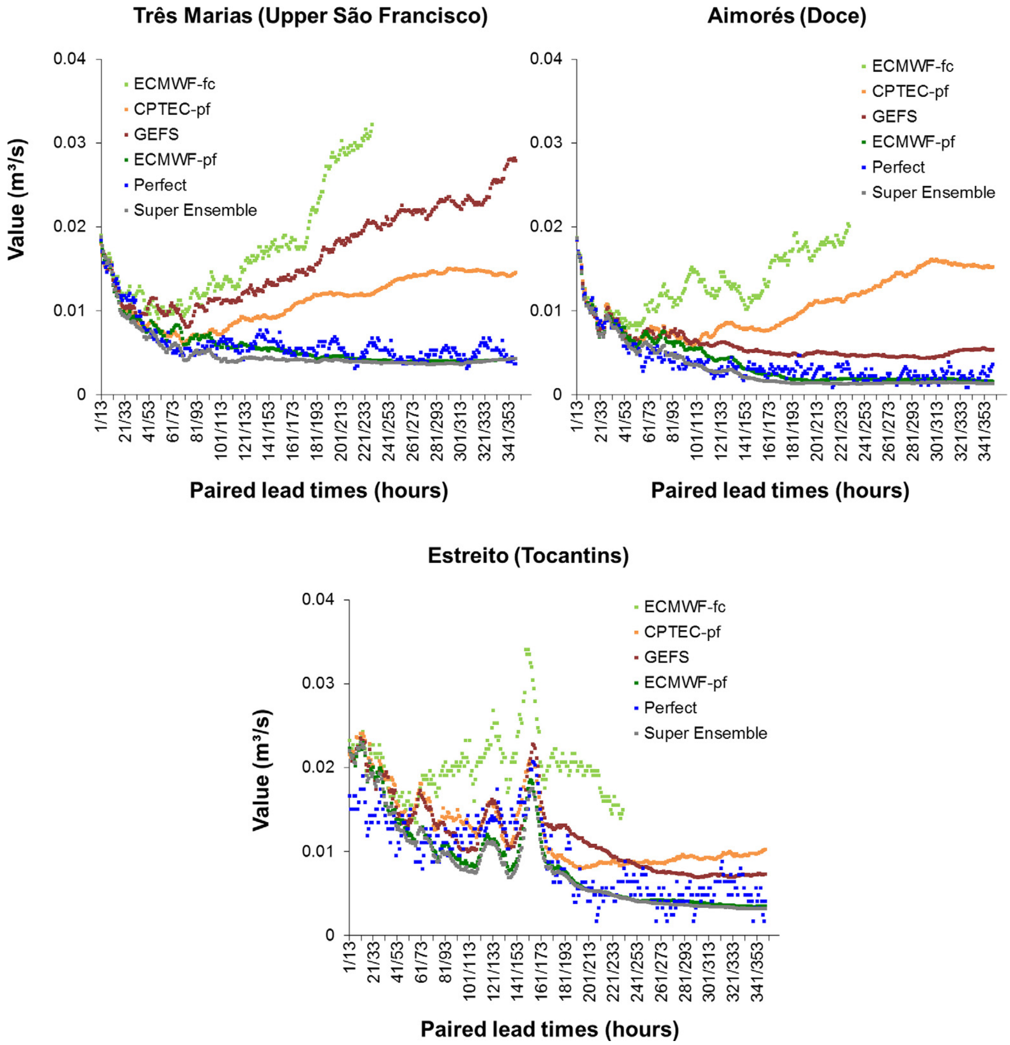

Fig. 18. Forecast Convergence Score (FCSbs) results for the three tested places.

5. Discussions

Metrics show that for small lead times, results are less dependent on the meteorological forcing and more related to observed conditions in the basin, with almost no spread in the ensemble. For lead times close to 5 days, results at all basin's and EPS differs, but differences are generally smaller compared with the ones verified on greater lead-times. At larger lead times, results indicate that ensemble forecasts issued using ECMWF-pf and GEFS, respectively, usually outperform the deterministic reference in terms of direct errors (MAE, CRPS) and the detection of threshold exceedance (BS). This includes some cases for which ECMWF-pf performs better than forecasts using observed rainfall. Forecasts using CPTEC-pf sometimes show poor performance in terms of MAE, CRPS and BS for larger lead-times, being outperformed by the other EPS and by the deterministic reference.

The assessment of the forecast's spread using the coefficient of variation does not give any evidence of a relationship between forecasts quality and spread. It does, however, provides three remarkable conclusions about the simulated system: (i) uncertainties are concentrated in the wet period; (ii) using Super Ensemble composition generates wider ensembles, which may sample more uncertainties; (iii) in the largest watershed (Tocantins) the coefficient (and also spread), was smaller than in the smaller ones, what can be related to a scale damping effect, and also to the basin flow control by dams.

In terms of calibration (reliability diagram), ECMWF-pf shows a good calibration for the Upper São Francisco river basin, a negative conditional type I bias in the Doce river basin, and a calibration varying with lead time in the Tocantins basin. Both CPTEC-pf and GEFS forecasts showed calibrations related to positive conditional type I bias in the Upper São Francisco and Doce basins, showing that with these models the exceedance of the threshold level had a bigger probability than its conditional observed frequency of occurrence. CPTEC-pf generally presented a decrease in the calibration with lead-time, while this was not so remarkable for GEFS. In the Tocantins basin the calibration was also shown to be dependent on lead-time, showing a negative type I bias at earlier horizons, which changes to a positive one in greater horizons.

In terms of the forecast consistency, two consecutive forecasts of the EPS are always more consistent than those of the deterministic reference forecast in terms of FCS and BS. This means that the ensembles promote fewer changes in the decisions about the threshold exceedance in comparison to the deterministic forecast. Also, of the tested EPS, the forecasts using ECMWF-pf were usually more consistent, followed consecutively by GEFS and CPTEC-pf in the Upper São Francisco and in the Tocatins rivers basins, and by CPTEC-pf and GEFS in the Doce basin.

It is important to mention that, in terms of BS and FCS, ensemble forecasts can issue for intermediate probabilities of occurrences, while deterministic forecasts can only issue for binary decisions (occurrence or not occurrence). Also, despite results already suggesting that full ensembles have better performance and are more consistent, there is another additional value in the probabilistic forecasts. Since decision making based on deterministic forecasts is also deterministic, the availability of probabilistic forecasts enable an application of stochastic techniques which consider both the expected value as well as its probability distribution. In other words, it is possible to make a more stable, weighted decision that would allow a satisfactory operation considering all possible futures.

An explanation for the different performances between the models may be found in various factors, such as different methods for generating initial conditions, models physics, spatial resolution, parameterizations, and the number of members in the ensemble. In terms of spatial resolution, ECMWF-pf is the EPS with the highest resolution, and also the one with the better performance in Três Marias HPP (Upper São Francisco) and Aimorés HPP (Doce) basins, but it was not always the best model for the Estreito HPP (Tocantins) basin. This last consideration is interesting, because Tocantins basin has the greater area (310,000 km^2) in comparison to the other two basins (50,000–65,000 km^2). This can be an indication of a relationship between watershed scale and necessary meteorological forecast resolution trade-off. This was, however, not the case for CPTEC-pf and deserves further investigations.

The number of ensemble members' influence in the metrics results may be also relevant, because the results usually show that the ensembles with the greater number of members (ECMWF-pf and Super Ensemble) presented good performance. In the works of Buizza and Palmer (1998), Richardson (2001), Müller et al. (2005), Weigel et al. (2007a,b), and Ferro et al. (2008) the authors discuss the importance of the number of members and suggest that ensembles with fewer members can influence some metrics results. Although, that was not the case in the verified results, which showed CPTEC-pf (14 members) as the best performance between 48 h and 120 h in terms of CRPS in the Três Marias basin. Also for the Tocantins basin, GEFS results (20 members) were better than the other EPS at greater lead times. Finally, the number of members by itself is a characteristic of the EPS, and the ability to produce more distinct forecasts is a merit of the centre that produces it.

When the used EPS were combined to compose a grand-ensemble (named here Super Ensemble), with 84 members, the results mostly showed better performance for the Super Ensemble than the individual EPS for the used metrics. The curves obtained for the results had always a shape similar to ECMWF-pf that has the greatest number of members, but with metric values considered better. We believe that these results are related to the larger range of uncertainties being sampled in the forecast, where the Super Ensemble also accounts for the meteorological models structure deficiencies.

Also, we believe that the composition of the Super Ensemble is a more robust strategy in comparison to the use of one single EPS. Not only due to its good performance, but also related to operational security, when one model forecast fails or results do not seem reliable, one can account for others results. This option about model results interpretation can be based on assessments such as the ones shown here, of EPS individual performances.

When comparing the basins, the results provide evidence that the tested EPS have a distinct performance at the different watersheds. For example, ECMWF-pf seemed to have the best performance in Três Marias HPP, but in Aimorés HPP performances from ECMWF-pf and GEFS were closer to each other, and at Estreito HPP, GEFS outperformed the other models after initial lead times. This reinforces the idea that, although all basins are located in the same climatic tropical region, performances obtained at one watershed cannot be directly transposed to another.

One cannot discard that bias may play a role in the comparison of results, since no bias removal technique was applied. This is an assessment that could be better addressed in the future, when databases are longer and consolidated enough for this purpose.

The Upper São Francisco and Doce rivers basins had problems related to the few observations in the region, but at Tocantins river basin this showed to be an issue of greater importance. Forecast runs using observed precipitation were outperformed by GEFS and ECMWF-pf in terms of CRPS and BS statistics. In addition, another characteristic that makes Tocantins a special case is the presence of the Dams cascade. This setting is very usual in Brazilian large scale basins, on rivers such as Paraná (where Itaipu HPP is located), Iguaçu (southeast), and Uruguay (southern region). Therefore, the model setup and assessment of results for these kinds of basins is of particular importance for the development of further forecasting systems, either for energy generation or for anticipating floods.

Finally, we would like to mention that we are aware that some of the models tested here are constantly under improvement. For example, GEFS received updates that were not included in the TIGGE archive yet, as described by Hamill et al. (2013). The CPTEC global model has also been improved recently, but the data from the new model was not yet in the TIGGE archive during the development of the present work. As our focus was to demonstrate the importance and use of TIGGE as an enabler of research on ensemble forecasting, we did not test these new versions, despite the fact that they are already in use. Results with new GEFS information, for example, can be found in Fan et al. (2014a,c,d).

6. Conclusions

This work presents one of the first extensive efforts to evaluate ensemble forecasts for large-scale basins in South America using TIGGE archive data. We have taken into the account the location of hydro power plants reservoirs, which are important users of forecasts.

We believe that the first remark of interest from the study is the importance of TIGGE, where data from multiple centres can be obtained, and thus the possible benefits of ensemble forecasts and tests on multiple EPS applications can be made. This can help in accelerating the development of research in regions where these products would not be so easily available.

In general, ensemble forecasts, especially those from ECMWF-pf and GEFS show a superior performance in terms of metrics in comparison with the deterministic prediction. Moreover, ensembles proved to be more consistent in terms of sequential decisions than the deterministic forecasts, which may be beneficial in reservoir operation decision making, for example, if the forecasts also appear to have fewer errors.

The performance, the calibration and the consistency of results among the test sites was distinct. This means that one cannot assume results to be similar to other basins, even though the entire study area lies within a tropical climate zone. In addition, for smaller basins a greater spreading of the ensemble was observed, and the opposite for the bigger watershed, possibly because the larger area has a damping effect on rainfall-runoff generation and routing processes.

The combined forecast, (denominated here Super Ensemble), showed to be a good option. Firstly, it was observed to span more uncertainty, and the resulting performance was always among the best of the options tested. Second, it can be a robust alternative for forecast generation. What could be visually inferred from the ensemble is that the individual models do not always give a unanimous prediction on the occurrence of events. However, indication by more than one model in the combined forecast gives more reliability to a user, and is less subjected to problems such as operational failures. Despite this, it is important to mention that the operation of this type of system can be affected by computational issues linked to preprocessing large amounts of data and the time required to run multiple models, which usually increases with the scale of the basins. We also suggest other forms of Super Ensemble composition to be tested, because in our case we attach great importance to the EPS with more members.

In all basins, results from the different EPS differ for lead times near to 5 days. However, these differences were not highly sensitive in comparison to the ones verified for greater lead times. Therefore, in the development of further forecasting systems for these large scale basins, different EPS or Super Ensembles may not be required, if five days is enough for decision support. Another implication of the tests is related to the rainfall observed, primarily a problem in the Tocantins River basin. The station's set-up in the basin was shown to be unable to capture all the rainfall occurring in the area. For the development of a forecasting system targeted to larger areas in South America, for example, this is certainly a challenge to overcome. In this case, it would be necessary to include some strategies for hydrological forecasting that combine point observed rainfalls with satellite rainfall estimations, as presented by Vila et al. (2009), Rozante et al. (2010), and Fan et al. (2014a).

For hydropower generation and flood protection in large-scale basins, the implications of the results shown here are that forecasts that can be generated are comparable in terms of quality and consistency to forecasts generated using observed rainfall in the horizon (even though observations are not at all perfect), and this good performance can certainly assist in systems operation. As ensemble forecasts are not as simple to interpret as deterministic forecasts, we believe that the best strategy to do this is through the use of forecasts in mathematical optimization models.

An extra operational challenge emerges regarding the consideration of the operation and the uncertainty of the effect of reservoir operation cascade, as was the case for the Tocantins river basin shown here, which is a common configuration of large Brazilian river basins. Finally, we believe that the techniques employed, including the pending improvement opportunities, proved to be suitable for employment in a planned "SAFAS" system, promoting the issue of large scale results in an hourly time-step.

We hope that this work will be considered an expansion of the knowledge about the application of ensemble forecasts verification all over the world, contributing to the composition of knowledge about the studied aspects.

Conflict of interest

There is no actual or potential conflict of interest in this publication.

Acknowledgments

We would like to acknowledge Brazilian CAPES, CNPQ and CEMIG for their funding of the projects that together made this research possible. Also, we would like to acknowledge the reviewers of this paper for the valuable suggestions given during the review process. And finally we would like to acknowledge Alex Curran and Na'ama Schweitzer for the precious English review.

References

Abebe, A.J., Price, R.K., 2003. Managing uncertainty in hydrological models using complementary models. Hydrol. Sci. 48 (5), 679–692.

Addor, N., Jaun, S., Fundel, F., Zappa, M., 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrol. Earth Syst. Sci. 15, 2327–2347, http://dx.doi.org/10.5194/hess-15-2327-2011.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS – global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci. 17, 1161–1175, http://dx.doi.org/10.5194/hess-17-1161-2013.

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., Salamon, P., 2014. Evaluation of ensemble streamflow predictions in Europe. J. Hydrol. 517 (September), 913–922, http://dx.doi.org/10.1016/j.jhydrol.2014.06.035, ISSN: 0022-1694.

Bao, H.-J., Zhao, L.-N., He, Y., Li, Z.-J., Wetterhall, F., Cloke, H.L., Pappenberger, F., Manful, D., 2011. Coupling ensemble weather predictions based on TIGGE database with Grid-Xinanjiang model for flood forecast. Adv. Geosci. 29, 61–67.

Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S., 2009. The European Flood Alert System EFAS. Part 2. Statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol. Earth Syst. Sci. 13 (2), 141–215.
Bates, P.D., Horritt, M.S., Fewtrell, T.J., 2010. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. J. Hydrol. 387, 33–45, http://dx.doi.org/10.1016/j.jhydrol.2010.03.027.
Bergh, Van Den J., Roulin, E., 2010. Hydrological ensemble prediction and verification for the Meuse and Scheldt basins. Atmos. Sci. Lett. 11, 64–71.
Boucher, M.-A., Anctil, F., Perreault, L., Tremblay, D., 2011. A comparison between ensemble and deterministic hydrological forecasts in an operational context. Adv. Geosci. 29, 85–94, http://dx.doi.org/10.5194/adgeo-29-85-2011.
Boucher, M.-A., Tremblay, D., Delorme, L., Perreault, L., Anctil, F., 2012. Hydroeconomic assessment of hydrological forecasting systems. J. Hydrol. 416–417, 133–144.
Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D., Ebert, E., Fuentes, M., Hamill, T., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Silva Dias, P., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., Worley, S., 2010. The THORPEX Interactive Grand Global Ensemble (TIGGE). Bull. Am. Met. Soc. 91, 1059–1072.
Bourdin, D.R., Stull, R.B., 2013. Bias-corrected short-range member-to-member ensemble forecasts of reservoir inflow. J. Hydrol. 502, 77–88, http://dx.doi.org/10.1016/j.jhydrol.2013.08.028, ISSN: 0022-1694.
Bowler, N.E., 2006. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. Tellus 58A, 538–548.
Bradley, A.A., Schwartz, S.S., 2011. Summary verification measures and their interpretation for ensemble forecasts. Mon. Weather Rev. 139 (9), 3075–3089.
Bravo, J., Paz, A., Collischonn, W., Uvo, C., Pedrollo, O., Chou, S., 2009. Incorporating forecasts of rainfall in two hydrologic models used for medium-range streamflow forecasting. J. Hydrol. Eng. 1 (5), 435–445.
Brown, J.D., Demargne, J., Seo, D.-J., Liu, Y., 2010. The ensemble verification system (EVS): a software tool for verifying ensemble forecasts of ydrometeorological and hydrologic variables at discrete locations. Environ. Model. Softw. 25 (7), 854–872.
Buizza, R., Palmer, T.N., 1998. Impact of ensemble size on ensemble prediction. Mon. Weather Rev. 126, 2503–2518.
Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, P., Wei, M., Zhu, Y., 2005. A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. Mon. Weather Rev. 133, 1076–1097.
Buizza, R., 2008. The value of probabilistic prediction. Atmos. Sci. Lett. 9, 36–42.
Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., Vitart, F., 2007. The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). Q. J. R. Meteorol. Soc. 133, 681–695.
Candogan-Yossef, N., Winsemius, H., Weerts, A., Van Beek, R., Bierkens, M.F.P., 2013. Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. Water Resour. Res., 49, http://dx.doi.org/10.1002/wrcr.20350.
Calvetti, L., Pereira Filho, A.J., 2014. Ensemble hydrometeorological forecasts using WRF hourly QPF and topmodel for a middle watershed. Adv. Meteorol. 2014, 1–12.
Cavalcanti, I.F.A., Marengo, J., Satyamurty, P., Nobre, C.N., Trosnikov, I., Bonatti, J.P., Manzi, A.O., Tarasova, T., Pezzi, L.P., Dalmeida, C., 2002. Global climatological features in a simulation using the CPTEC/COLA AGCM. J. Climate 15 (21), 2965–2988.
Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. J. Hydrol. 375 (3–4), 613–626.
Collischonn, W., Tucci, C.E.M., Clarke, R.T., Chou, S.C., Guilhon, L.G., Cataldi, M., Allasia, D.G., 2007a. Medium-range reservoir inflow predictions based on quantitative precipitation forecasts. J. Hydrol., Amsterdam 344, 112–122.
Collischonn, W., Tucci, C.E.M., Clarke, R.T., Delgado, M.C., Silva, B.C., Collischonn, B., Allasia, D.G., Paz, A.R., 2007b. Modelo hidrológico distribuído para previsão de vazões incrementais na bacia do rio Paranaíba entre Itumbiara e São Simão. Rev. Bras. Recur. Hídr. 12, 43–56.
Collischonn, W., Meller, A., Fan, F., Moreira, D.S., Silva Dias, P.L., Buarque, D., Bravo, J.M., 2013. Short-term ensemble flood forecasting experiments in Brazil. Geophys. Res. Abstr. 15, 11910.
Collischonn, W., Tucci, C.E.M., Haas, R., Andreolli, I., 2005. Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. J. Hydrol., Amsterdam 305, 87–98.
Costa, F.S., Raupp, I.P., Damazio, J.M., Oliveira, P.D., Guilhon, L.G.F.,2014. The methodologies for the flood control planning using hydropower reservoirs in Brazil. In: 6th International Conference on Flood Management, 2014, São Paulo. ICMF6. ABRH, São Paulo.
Cui, B., Toth, Z., Zhu, Y., Hou, D., 2012. Bias correction for global ensemble forecast. Weather Forecast. 27, 396–410.
Cunningham, C.A., Bonatti, J.P., 2011. Local and remote responses to opposite Ross Sea ice anomalies: a numerical experiment with the CPTEC/INPE AGCM. Theor. Appl. Climatol., 1–22.
Cuo, L., Pagano, T.C., Wang, Q.J., 2011. A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. J. Hydrometeorol. 12, 713–728.
Demargne, J., Brown, J., Liu, Y., Seo, D.-J., Wu, L., Toth, Z., Zhu, Y., 2010. Diagnostic verification of hydrometeorological and hydrologic ensembles. Atmos. Sci. Lett. 11, 114–122, http://dx.doi.org/10.1002/asl.261.
Demargne, J., Wu, L., Regonda, S.K., Brown, J.D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H.D., Fresch, M., Schaake, J., Zhu, Y., 2014. The science of NOAA's operational Hydrologic Ensemble Forecast Service. Bull. Am. Meteorol. Soc. 95, 79–98, http://dx.doi.org/10.1175/BAMS-D-12-00081.1.
Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., Ramos, M.-H., 2007. Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. Environ. Hazards 7, 115–127.
Demirel, M.C., Booij, M.J., Hoekstra, A.Y., 2013. Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. Water Resour. Res. 49, 4035–4053, http://dx.doi.org/10.1002/wrcr.20294.
Fan, F.M., Collischonn, W., Quiroz, K., Sorribas, M.V., Buarque, D.C., Siqueira, V.A., 2014a. Flood forecasting on the Tocantins River using ensemble rainfall forecasts and real-time satellite rainfall estimates. In: IFCM-6 proceedings, São Paulo, Brazil.
Fan, F.M., Pontes, P.R.M., Beltrame, L.F.S., Collischonn, W., Buarque, D.C., 2014b. Operational flood forecasting system to the Uruguay River Basin using the hydrological model MGB-IPH. In: IFCM-6 proceedings, São Paulo, Brazil.

Fan, F.M., Collischonn, W., Meller, A., Botelho, L.C.M., 2014c. Ensemble streamflow forecasting experiments in a tropical basin: The São Francisco river case study. J. Hydrol. 519 (Part D), 2906–2919, http://dx.doi.org/10.1016/j.jhydrol.2014.04.038, ISSN 0022-1694.

Fan, F.M., Collischonn, W., Quiroz, K., Sorribas, M.V., Buarque, D.C., Siqueira, V.A., 2014d. Ensemble flood forecasting on the Tocantins River – Brazil. Geophys. Res. Abstr. 16, 1818.

Ferro, C.A.T., Richardson, D.S., Weigel, A.P., 2008. On the effect of ensemble size on the ranked probability and multi-category Brier scores. Met. Appl. 15, 19–24.

Fraley, C., Raftery, A.E., 2010. Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. Mon. Weather Rev. 138, 190–202.

Franz, K.J., Hogue, T.S., 2011. Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community. Hydrol. Earth Syst. Sci. Discuss. 8, 3085–3131.

Georgakakos, K.P., Krzysztofowicz, R., 2001. Probabilistic and ensemble forecasting. J. Hydrol. 249, 1.

Golding, B.W., 2009. Long lead time flood warnings: reality or fantasy? Met. Appl. 16, 3–12, http://dx.doi.org/10.1002/met.123.

Guilhon, L.G.F., Rocha, V.F., Moreira, J.C., 2007. Comparação de métodos de previsão de vazões naturais afluentes a aproveitamentos hidroelétricos. Rev. Bras. Recur. Hídr. 12 (3), 13–20.

Hamill, T.M., Bates, G.T., Whitaker, J.S., Murray, D.R., Fiorino, M., Galarneau, T.J., Zhu Jr., Y., Lapenta, W., 2013. NOAA's second-generation global medium-range ensemble reforecast data set. Bull Am. Meteorol. Soc. 94, 1553–1565, http://dx.doi.org/10.1175/BAMS-D-12-00014.1.

Hamill, T.M., Snyder, C., Morss, R.E., 2000. A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. Mon. Weather Rev. 128, 1835–1851.

He, M., Hogue, T.S., Margulis, S.A., Franz, K.J., 2012. An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions. Hydrol. Earth Syst. Sci. 16, 815–831.

He, Y., Wetterhall, F., Bao, H., Cloke, H., Li, Z., Pappenberger, F., Hu, Y., Manful, D., Huang, Y., 2010. Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: a case study. Atmos. Sci. Lett. 11, 132–138, http://dx.doi.org/10.1002/asl.270.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast. 15, 559–570.

Jaun, S., Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrol. Earth Syst. Sci. 13, 1031–1043, http://dx.doi.org/10.5194/hess-13-1031-2009.

Jolliffe, I.T., Stephenson, D.B. (Eds.), 2012. Forecast Verification: A Practitioner's Guide in Atmospheric Science. , second ed.

Kouwen, N., Soulis, E.D., Pietroniro, A., Donald, J., Harrington, R.A., 1993. Grouped response units for distributed hydrologic modeling. J. Water Resour. Plan. Manage. 119 (3), 289–305.

Leutbecher, M., Palmer, T.N., 2008. Ensemble forecasting. J. Comput. Phys. 227, 3515–3539.

Ludwig, K., Bremicker, M., 2006. The water balance model LARSIM – design, content and applications. In: Freiburger Schiriften zür Hydrologie, Band 22. Institut fur Hydrologie der Universität Freiburg.

Meller, A., Collischonn, W., Fan, F.M., Buarque, D.C., Paiva, R.C.D., Dias, P., Moreira, D., 2014. Previsão de Cheias por Conjunto em Curto Prazo. Rev. Bras. Recur. Hídr. 19, 33–49.

Mendonça, A.M., Bonatti, J.P., 2009. Experiments with EOF-based perturbation methods and their impact on the CPTEC/INPE Ensemble Prediction System. Mon. Weather Rev. 137, 1438–1459.

Müller, W., Appenzeller, A.C., Doblas-Reyes, F.J., Liniger, M.A., 2005. A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. J. Climate 18, 1513–1523, http://dx.doi.org/10.1175/JCLI3361.1.

Marsigli, C., Boccanera, F., Montani, A., Paccagnella, T., 2005. The COSMO–LEPS ensemble system: validation of the methodology and verification. Nonlin. Proc. in Geophys. 12, 527–536.

Oliveira, P.D., Guilhon, L.G.F., Costa, F.S., Raupp, I.P., Damazio, J.M., 2014. The operation of Flood Control in Large Hydroelectric Power Systems – The Brazilian Experience. In: 6th International Conference on Flood Management, 2014, São Paulo. ICMF6.

Olsson, J., Lindstrom, G., 2008. Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. J. Hydrol. 350, 14–24.

ONS – Operador Nacional do Sistema, 2012a. Diretrizes para as Regras de Operação de Controle de Cheias – Bacia do Rio São Francisco (Ciclo 2012–2013). ONS RE 3/166/2012., pp. 158.

ONS – Operador Nacional do Sistema, 2014. Diretrizes para as Regras de Operação de Controle de Cheias – Bacia do Rio Iguaçu (Ciclo 2013–2014). ONS RE 3/0064/2014., pp. 48.

ONS – Operador Nacional do Sistema, 2012b. Inventário das Restrições Operativas Hidráulicas dos Aproveitamentos Hidrelétricos (Revisão 1 de 2012). ONS RE 3/0105/2012., pp. 154.

ONS – Operador Nacional do Sistema, 2011. Procedimentos de Rede Submódulo 9.5: Previsão de Vazões e Geração de Cenários de Afluências, vol. 2., pp. 9.

Pagano, T.C., Wood, A.W., Ramos, M.-H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., Verkade, J.S., 2014. Challenges of operational river forecasting. J. Hydrometeor. 15, 1692–1707, http://dx.doi.org/10.1175/JHMD-13-0188.1.

Pagano, T.C., 2014. Evaluation of Mekong River commission operational flood forecasts, 2000–2012. Hydrol. Earth Syst. Sci. 18, 2645–2656, http://dx.doi.org/10.5194/hess-18-2645-2014.

Paiva, R.C.D., Paiva, R.C.D., Collischonn, W., Bonnet, M.-P., de Gonçalves, L.G.G., Calmant, S., Getirana, A., Santos da Silva, J., 2013. Assimilating in situ and radar altimetry data into a large-scale hydrologic–hydrodynamic model for streamflow forecast in the Amazon. Hydrol. Earth Syst. Sci. 10, 2879–2925, Discussions (Online).

Paiva, R.C.D., Collischonn, W., Bonnet, M.P., De Gonçalves, L.G.G., 2012. On the sources of hydrological prediction uncertainty in the Amazon. Hydrol. Earth Syst. Sci. 16, 3127–3137.

Pappenberger, F., Brown, J.D., 2013. HP today: on the pursuit of (im) perfection in flood forecasting. Hydrol. Processes HEPS Spec. Issue 27, 162–163.

Pappenberger, F., Bogner, K., Wetterhall, F., He, Y., Cloke, H.L., Thielen, J., 2011a. Forecast convergence score: a forecaster's approach to analysing hydro-meteorological forecast systems. Adv. Geosci. 29, 27–32, http://dx.doi.org/10.5194/adgeo-29-27-2011.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H.L., Buizza, R., De Roo, A., 2008. New dimensions in early flood warning across the globe using grand-ensemble weather predictions. Geophys. Res. Lett. 35, http://dx.doi.org/10.1029/2008GL033837.
Pappenberger, F., Cloke, H.L., Persson, A., Demeritt, D., 2011b. HESS opinions "On forecast (in)consistency in a hydro-meteorological chain: curse or blessing?" Hydrol. Earth Syst. Sci. Discuss. 8, 1225–1245, http://dx.doi.org/10.5194/hessd-8-1225-2011.
Park, Y.-Y., Buizza, R., Leutbecher, M., 2008a. TIGGE: preliminary results on comparing and combining ensembles. Q. J. R. Meteorol. Soc. 134, 2029–2050, http://dx.doi.org/10.1002/qj.334.
Park, Y.-Y., Buizza, R., Leutbecher, M., 2008b. TIGGE: preliminary results on comparing and combining ensembles. Q. J. R. Met. Soc. 134, 2029–2050.
Paz, A.R., Collischonn, W., Tucci, C., Clarke, R., Allasia, D., 2007. Data Assimilation in a Large-scale Distributed Hydrological Model for Medium Range Flow Forecasts. IAHS Press, Wallingford, UK, pp. 471–478, IAHS Publication, n. 313.
Renner, M., Werner, M., Rademacher, S., Sprokkereef, E., 2009. Verification of ensemble flow forecasts for the River Rhine. J. Hydrol. 376, 463–475, http://dx.doi.org/10.1016/j.jhydrol.2009.07.059.
Richardson, D.S., 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. Q. J. R. Meteorol. Soc. 127, 2473–2489.
Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrol. Earth Syst. Sci. 11, 725–737, http://dx.doi.org/10.5194/hess-11-725-2007.
Rozante, J.R., Moreira, D.S., De Goncalves, L.G.G., Vila, D.A., 2010. Combining TRMM and surface observations of precipitation: technique and validation over South America. Weather Forecast. 25, 885–894.
Schaake, J., Franz, K., Bradley, A., Buizza, R., 2006. The Hydrologic Ensemble Prediction EXperiment (HEPEX). Hydrol. Earth Syst. Sci. Discuss. 3, 3321–3332.
Schellekens, J., Weerts, A.H., Moore, R.J., Pierce, C.E., Hildon, S., 2011. The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales. Adv. Geosci. 29, 77–84.
Stanski, H., Wilson, L., Burrows, W., 1989. Survey of Common Verification Methods in Meteorology. World Meteorological Organization, Geneva.
Thielen, J., Bartholmes, J., Ramos, M.-H., De Roo, A., 2009. The European flood alert system. Part 1. Concept and development. Hydrol. Earth Syst. Sci. 13 (2), 125–140.
Thiemig, V., Pappenberger, F., Thielen, J., Gadain, H., De Roo, A., Bodis, K., Del Medico, M., Muthusi, F., 2010. Ensemble flood forecasting in Africa: a feasibility study in the Juba–Shabelle river basin. Atmos. Sci. Lett. 11, 123–131, http://dx.doi.org/10.1002/asl.266.
Thiemig, V., Bisselink, B., Pappenberger, F., Thielen, J., 2014. A pan-African flood forecasting system. Hydrol. Earth Syst. Sci. Discuss. 11, 5559–5597, http://dx.doi.org/10.5194/hessd-11-5559-2014.
Todini, E., 1996. The ARNO rainfall-runoff model. J. Hydrol. 175 (1–4), 339–382, http://dx.doi.org/10.1016/S0022-1694(96)80016-3.
Toth, Z., Kalnay, E., 1997. Ensemble forecasting at NCEP and the breeding method. Mon. Weather Rev. 125, 3297–3319.
Toth, Z., Kalnay, E., 1993. Ensemble forecasting at NMC: the generation of perturbations. Bull. Am. Meteorol. Soc. 74, 2317–2330.
Toth, Z., Talagrand, O., Candille, G., Zhu, Y., 2003. Probability and ensemble forecasts. In: Jolliffeand, I.T., Stephenson, D.B. (Eds.), Forecast Verification: A Practitioner's Guide in Atmospheric Science. John Wiley and Sons, pp. 137–163.
Tucci, C.E.M., Collischonn, W., Clarke, R.T., Paz, A.R., Allasia, D., 2008. Short- and long-term flow forecasting in the Rio Grande watershed (Brazil). Atmos. Sci. Lett. 9, 53–56.
Velázquez, J.A., Anctil, F., Ramos, M.H., Perrin, C., 2011. Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. Adv. Geosci. 29, 33–42.
Velázquez, J.A., Petit, T., Lavoie, A., Boucher, M.-A., Turcotte, R., Fortin, V., Anctil, F., 2009. An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting. Hydrol. Earth Syst. Sci. 13, 2221–2231, http://dx.doi.org/10.5194/hess-13-2221-2009.
Verbunt, M., Walser, A., Gurtz, J., Montani, A., Schär, C., 2007. Probabilistic flood forecasting with limited-area ensemble prediction system: selected case studies. J. Hydrometeorol. 8, 897–909.
Verkade, J.S., Werner, M.G.F., 2011. Estimating the benefits of single value and probability forecasting for flood warning. Hydrol. Earth Syst. Sci. 15, 3751–3765, http://dx.doi.org/10.5194/hess-15-3751-2011.
de Goncalves, V.D.A., Luis Gustavo, G., Toll, D.L., Rozante, J.R., 2009. Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. J. Hydrometeorol. 10, 533.
Wei, M., Toth, Z., Wobus, R., Zhu, Y., 2008. Initial perturbations based on the Ensemble Transform (ET) technique in the NCEP global operational forecast system. Tellus 60A, 62–79.
Wei, M., Toth, Z., 2003. A new measure of ensemble performance: Perturbations versus Error Correlation Analysis (PECA). Mon. Weather Rev. 131, 1549–1565.
Weigel, A.P., Liniger, M.A., Appenzeller, C., 2007a. Generalization of the discrete Brier and ranked probability skill scores for weighted multi-model ensemble forecasts. Mon. Weather Rev. 135, 2778–2785.
Weigel, A.P., Liniger, M.A., Appenzeller, C., 2007b. The discrete Brier and ranked probability skill scores. Mon. Weather Rev. 135, 118–124.
Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences, 2a ed. Academic Press, pp. 627.
Xuan, Y., Cluckie, I.D., Wang, Y., 2009. Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilising short-range rainfall prediction. Hydrol. Earth Syst. Sci. 13, 293–303.
Younis, J., Ramos, M.-H., Thielen, J., 2008. EFAS forecasts for the March–April 2006 flood in the Czech part of the Elbe River Basin – a case study. Atmos. Sci. Lett. 9, 88–94.
Zambon Renato, C., de Barros, M.T.L., Yeh, W.W.-G.,2014a. Brazilian hydrothermal system operation: interconnected large system or isolated subsystems? In: World Environmental and Water Resources Congress 2014, Portland. American Society of Civil Engineers, Reston, p. 1926.
Zambon, R.C., de Barros, M.T.L., Gimenes, M., Bozzini, P.L., Yeh, W.W.-G., 2014b. Flood control and energy production on the Brazilian Hydrothermal System. In: 6th International Conference on Flood Management, São Paulo.

Zambon, R.C., de Barros, M.T.L., Lopes, J.E.G., Barbosa, P.S.F., Francato, A.L., Yeh, W.W.-G., 2012. Optimization of large-scale hydrothermal system operation. J. Water Resour. Plan. Manage. 138, 135–143.
Zhang, Z., Krishnamurti, T.N., 1999. A perturbation method for hurricane ensemble predictions. Mon. Weather Rev. 127, 447–469.
Zhao, T., Yang, D., Cai, X., Zhao, J., Wang, H., 2012. Identifying effective forecast horizon for real-time reservoir operation under a limited inflow forecast. Water Resour. Res. 48, W01540, http://dx.doi.org/10.1029/2011WR010623.
Zhao, T., Ximing, C., Dawen, Y., 2011. Effect of streamflow forecast uncertainty on real-time reservoir operation. Adv. Water Resour. 34 (4), 495–504.