



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Perera, KC;Western, AW;Robertson, DE;George, B;Nawarathna, B

Title:

Ensemble forecasting of short-term system scale irrigation demands using real-time flow data and numerical weather predictions

Date:

2016-06-01

Citation:

Perera, K. C., Western, A. W., Robertson, D. E., George, B. & Nawarathna, B. (2016). Ensemble forecasting of short-term system scale irrigation demands using real-time flow data and numerical weather predictions. *Water Resources Research*, 52 (6), pp.4801-4822. <https://doi.org/10.1002/2015WR018532>.

Persistent Link:

<https://hdl.handle.net/11343/291377>

# Ensemble forecasting of short-term system scale irrigation demands using real time flow data and numerical weather predictions

Kushan C. Perera <sup>1</sup>, Andrew W. Western <sup>1\*</sup>, David E. Robertson <sup>2</sup>, Biju George <sup>3</sup>, Bandara Nawarathna <sup>4</sup>

<sup>1</sup> Department of Infrastructure Engineering, The University of Melbourne, Australia, <sup>2</sup> CSIRO Land and Water Flagship, Australia, <sup>3</sup> International Centre for Agricultural Research in the Dry Areas (ICARDA), Cairo, Egypt, <sup>4</sup> Hazards, Warnings and Forecasts Division, the Bureau of Meteorology, Australia,

Corresponding author: Andrew W. Western ([a.western@unimelb.edu.au](mailto:a.western@unimelb.edu.au))

## Key Points:

- This study forecasts probabilistic short-term system scale irrigation demand for lead time up to 5 days
- Measurement/estimation/forecast errors for flows, NWP forecasts & observed weather are integrated
- Rank histograms and other indices indicated that a reliable ensemble spread is achieved.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version record](#). Please cite this article as [doi:10.1002/2015WR018532](https://doi.org/10.1002/2015WR018532).

## Abstract

Irrigation demands fluctuate in response to weather variations and a range of irrigation management decisions, which creates challenges for water supply system operators. This paper develops a method for real-time ensemble forecasting of irrigation demand and applies it to irrigation command areas of various sizes for lead times of 1 to 5 days. The ensemble forecasts are based on a deterministic time series model coupled with ensemble representations of the various inputs to that model. Forecast inputs include past flow, precipitation, and potential evapotranspiration. These inputs are variously derived from flow observations from a modernized irrigation delivery system; short-term weather forecasts derived from numerical weather prediction models and observed weather data available from automatic weather stations. The predictive performance for the ensemble spread of irrigation demand was quantified using rank histograms, the mean continuous rank probability score (CRPS), the mean CRPS reliability and the temporal mean of the ensemble root mean squared error (MRMSE). The mean forecast was evaluated using root mean squared error (RMSE), Nash–Sutcliffe model efficiency (NSE) and bias. The NSE values for evaluation periods ranged between 0.96 (1 day lead time, whole study area) and 0.42 (5 days lead time, smallest command area). Rank histograms and comparison of MRMSE, mean CRPS, mean CRPS reliability and RMSE indicated that the ensemble spread is generally a reliable representation of the forecast uncertainty for short lead times but underestimates the uncertainty for long lead times.

**Keywords:** ensemble forecast, irrigation demand, numerical weather prediction, time series model

## 1 Introduction

Short-term system scale irrigation demand forecasts are extremely useful for system operators to make irrigation water distribution decisions, but they are subject to uncertainties resulting from input parameters and model structure. Parameter and structural uncertainties are inherent in models, as models try to simplify the complex reality. Input uncertainties are also important as inputs are subject to measurement, estimation or prediction uncertainties. From the irrigation demand forecasting prospective, uncertainties in the irrigation demand forecasts can result from observation, estimation or prediction uncertainties in biophysical (crop-soil-climate interaction), behavioral (farmers and system operators attitude that influencing management decisions) and supply (supply source, seasonal allocation, permanent entitlement) factors [[Zaman et al., 2007](#)]; as well as from the respective parameter uncertainties, depending on the specific models used. This research makes ensemble forecasts of irrigation demand using a multivariate time-series model of short-term (up to 5 days) daily irrigation demand forced by observed demands and observed and forecast weather integrating measurement errors, estimation errors, and weather forecast uncertainty.

Past research has used sophisticated and highly complex modelling architectures such as process-based (conceptual) and data-driven (statistical) approaches [[Alfonso et al., 2011](#); [Pulido-Calvo and Gutierrez-Estrada, 2009](#)] to derive irrigation demand forecasts. Typically, process-based approaches have been used at field scale and data-driven approaches have been used at system scale. Few of these studies have attempted to estimate irrigation demand forecast uncertainties. Studies at the system scale have mainly used data-based deterministic models to forecast irrigation demand [[Pulido-Calvo and Gutierrez-Estrada, 2009](#); [Pulido-Calvo et al., 2007](#); [Pulido-Calvo et al., 2003](#)]. A few of these studies have considered uncertainty, mainly focusing on parameter uncertainties. Both bootstrap methods [[Ticlavilca et al., 2011](#)] and Bayesian techniques [[Alfonso et al., 2011](#)] have been used to investigate the irrigation demand forecast uncertainties resulting from model parameters. Also, some univariate time series models based on previous irrigation flows [[Pulido-Calvo and Gutierrez-Estrada, 2009](#); [Pulido-Calvo et al., 2007](#); [Pulido-Calvo et al., 2003](#)] have captured the combined uncertainty from inputs and model structure by including the associated error model to represent the random error component in the irrigation demand forecasts.

Many models have used short-term weather forecasts to make irrigation decisions, especially precipitation [[Azhar and Perera, 2011](#); [Cai et al., 2011](#); [Gowing and Ejeji, 2001](#); [Wang and Cai, 2009](#); [Wilks and Wolfe, 1998](#)], temperature [[Ticlavilca et al., 2011](#); [Wilks and Wolfe, 1998](#)] and evapotranspiration [[Alfonso et al., 2011](#); [Tian and Martinez, 2014](#)]. Weather forecasts have

often been considered for field scale and a few studies that we are aware of have incorporated weather forecasts at the system scale [[Tian and Martinez, 2014](#)]. Field scale models have often used weather forecasts for irrigation scheduling (i.e. predict timing and its volume) and weather forecast uncertainties were manifest mainly in variations of the irrigation timing [[Cai et al., 2011](#); [Gowing and Ejieji, 2001](#); [Wang and Cai, 2009](#); [Wilks and Wolfe, 1998](#)]. These models only used limited forecast information as the quality of the available weather forecast at that time was poor at longer lead times (greater than 3 days lead times) [[Gowing and Ejieji, 2001](#)]. This was seen as a major impediment at the time. [[Alfonso et al., 2011](#)] and [[Tian and Martinez, 2014](#)] forecasted system scale irrigation demands, combining stochastic reference evapotranspiration forecasts derived from a machine learning algorithm and currently operational Global Ensemble Forecast System (GEFS), respectively. [[Ticlavilca et al., 2011](#)] also combined daily maximum and minimum temperature with system scale irrigation demand forecasts using a machine learning algorithm.

These system scale models were mainly developed for arid-zone agriculture and precipitation forecasts were not considered, as it is not important for irrigation decisions in those environments [[Alfonso et al., 2011](#); [Tian and Martinez, 2014](#); [Ticlavilca et al., 2011](#)]. These studies have generated system scale stochastic irrigation demand forecasts with prediction intervals. In terms of assessing the probabilistic forecasts, all studies provided a graphical comparison of observed time series overlain on the forecast prediction intervals without a quantitative evaluation. [[Tian and Martinez, 2014](#)] also provided relative operating characteristic (ROC) diagrams that suggested, they didn't provide any statistical analysis of the reliability or sharpness of the probabilistic forecasts. The maximum lead time for the most of these studies was 2 days or less and no study comprehensively integrates input uncertainties into the irrigation demand forecast uncertainties. While ensemble techniques have been used elsewhere, in hydrology [[Addor et al., 2011](#); [Li et al., 2015](#); [Shrestha et al., 2013a](#); [Zappa et al., 2011](#)] and hydrometeorology [[Brown et al., 2010](#); [Ebert et al., 2011](#); [Gneiting, 2013](#); [Robertson et al., 2013a](#); [Rossa et al., 2011](#)], no studies we are aware of have used both stochastic precipitation and reference evapotranspiration forecasts nor have there been studies that combine uncertainties in antecedent flows and observed and forecast weather to derive stochastic volumetric irrigation demand forecasts at system scale.

Ensemble techniques are commonly used to represent uncertainty in non-linear models. In the context of forecasting; ensemble forecasting is a form of Monte Carlo analysis that is used to characterize uncertainty in model outputs [[Toth and Kalnay, 1993](#)]. In principle, this technique can be applied for both input and parameter uncertainty depending on the context of the modelling. Ensemble forecasting techniques have been widely used in science, engineering, medicine and ecology, among other areas. It has been widely used in weather forecasting [[Ebert, 2001](#); [Ebert et](#)

[al., 2011](#); [Gneiting, 2013](#); [Gneiting and Raftery, 2005](#)] and species distribution modelling [[Araújo and New, 2007](#); [Buisson et al., 2010](#); [Grenouillet et al., 2011](#); [Thuiller et al., 2009](#)] to capture errors in the initial condition and model structure. In the field of water resources engineering, ensemble forecasting techniques have also been used extensively in forecasting stream flow [[Bennett et al., 2014](#)], short-term water demand [[Hutton and Kapelan, 2015](#)] and floods [[Alvarez-Garretón et al., 2014](#); [Cloke and Pappenberger, 2009](#); [Li et al., 2014](#); [Schaake, 2006](#); [Schaake et al., 2005](#)]. In these studies, sources of uncertainty included in forecast ensembles were uncertainty in inputs (mainly precipitation forecasts), state variable (soil moisture), and model structures through model parameters. We are unaware of any irrigation demand forecast model that has used ensemble forecasting techniques to derive stochastic irrigation demand forecasts.

The structure of the multivariate time series model used here, which has auto-correlated and cross-correlated multivariate inputs, some of which are non-linearly transformed, suggests ensemble forecasting techniques to derive output uncertainties would be useful. This technique has advantages in multivariate time series models compared with the bootstrap method, which can often distort the cross correlations between multivariate inputs and outputs [[Khaliq et al., 2009](#); [Rummel et al., 2010](#)], whereas ensemble forecasts can preserve the error covariance among input time series which is then transferred through the regression structure to the stochastic output. However, the outcomes from an ensemble forecasting technique is dependent on the quality of the input time series. Therefore, a good quality data set (i.e. fine scale, free from bias and complete) is important for generating reliable, unbiased output ensembles. Past impediments such as the lack of required data, the expense of data acquisition and low data quality are being reduced with the availability of irrigation flow data from fully automated irrigation distribution systems, short-term weather forecasts from numerical weather predictions (NWP) models and observed weather data from automatic weather stations. In particular, the accuracy of the weather forecasts is continuously improving and consistent real-time irrigation flow data are now available from modernized distribution systems. This provides an opportunity to apply ensemble forecasting techniques to generate stochastic irrigation demand forecasts.

This paper develops ensemble irrigation demand forecasts using the irrigation distribution system in the Goulburn-Murray Irrigation District (GMID) in Northern Victoria as a case study. This system has been modernized and automated connections provide consistent real-time flow data [[NVIRP, 2010a](#)]. Previously, [[Perera et al., 2015a](#)] developed a deterministic model and assessed its performances under perfect weather forecasts i.e. using observed weather data. This paper uses that model as a basis and both brings real weather forecasts into the analysis and develops an ensemble framework to incorporate uncertainties arising from the weather forecast data as well as

other inputs. In doing this we use data from the GMID and numerical weather predictions (NWP) from the Bureau of Meteorology (BOM) to develop ensemble irrigation demand forecasts. This is a relatively new and high level of data availability compared with many irrigation systems. The analysis includes input uncertainties associated with the flow measurement and weather measurements and forecasts. The remainder of the paper describes the area of study where the methodology has been applied and forecasting performance for command area driven stochastic irrigation demand forecasts and conclusions that has been drawn. This provides novel understanding about the irrigation demand prediction uncertainties like input uncertainties related to biophysical, behavioral and supply factors and in turn assists system operators to mitigate the risk associated with their routine irrigation distribution decisions.

## 2 Study area and data

### 2.1 Study area

The study area and data sources are described in detail in [[Perera et al., 2015a](#)]. The study area is located in the Central Goulburn Irrigation District (CGID), Victoria, Australia. Agriculture in the study area is dominated by irrigated dairy, pome and stone fruit production, with other agricultural activities related to sheep for wool, beef and dairy cattle [[RDV, undated](#)]. The main source of water supply for the CGID is from Lake Eildon, with delivery via the Goulburn River, Goulburn weir, and the Stuart Murray Canal. Water flows from Stuart Murray Canal to the CGID through six gravity irrigation distribution channels namely CG1, 2, 3, 4, 5 and 6. Any excess water in the Stuart Murray Canal is diverted to Waranga Basin (Fig. 1). The irrigation distribution system in the CGID is highly automated and a SCADA (supervisory control and data acquisition) system monitors levels and flows and controls the regulator gates and meter outlets across the district. Irrigation water takes about 4 days to travel from Lake Eildon to the farms in this area. The proposed ensemble forecasting methodology was applied to 287 km<sup>2</sup> of irrigated agricultural land supplied by CG 1, 2, 3 and 4. The characteristics for each channel are given in Table 1. The irrigation year starts on the 15 of August and continues to the 15 of May the following year. The study area is approximately 110 meters above the Australian height datum (AHD) and the climate is temperate with a hot summer ( $T_{hot} \geq 22^{\circ} C$ ) but without a dry season (Köppen climate type Cfa) [[Peel et al., 2007](#)].

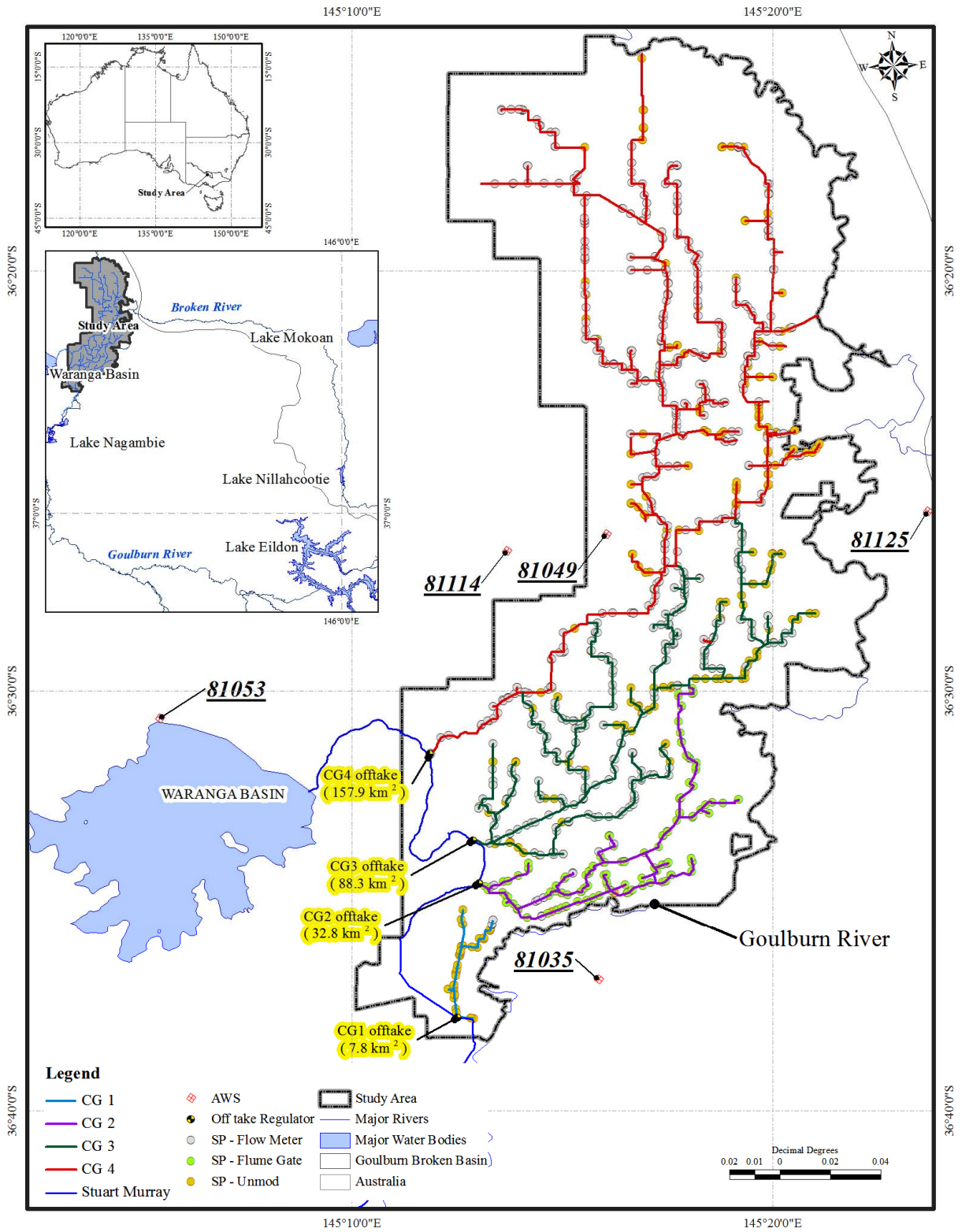


Figure 1 – Study area – Central Goulburn channel 1, 2, 3, and 4 [Perera et al., 2015a]

Table 1 – Characteristics of CG 1, 2, 3 and 4 channels

Channel	Length(km)				Area (km <sup>2</sup> )	Service points (Nos.)			Regulators (Nos.)			Degree of Mod. <sup>2</sup> (%)	
	L <sup>1</sup> -1	L <sup>1</sup> -2	L <sup>1</sup> -3	Total		Mod. <sup>2</sup>	Unm. <sup>3</sup>	Total	Mod. <sup>2</sup>	Unm. <sup>3</sup>	Total	SP <sup>4</sup>	Vol <sup>5</sup>
CG1	5.5	2.4		7.87	7.8	5	41	46	5	3	8	11	20
CG2	18.5	15.6	3.2	37.3	32.8	142	7	149	36	0	36	95	96
CG3	17.32	43.34	25.5	86.2	88.3	246	109	355	88	17	105	70	65
CG4	39.8	76.9	23.4	140.0	157.9	357	109	466	117	32	149	77	60
CG1234	81.1	138.2	52.1	271.4	286.9	750	266	1016	246	32	298	74	65

<sup>1</sup>Levels of distribution (L-1- backbone, L-2 – primary and L-2- secondary), <sup>2</sup>Modernised, <sup>3</sup>Unmodernised, <sup>4</sup>Service points, and <sup>5</sup>Volume through automated service points

## 2.2 Data sources and pre-processing

### 2.2.1 Irrigation flow data

The irrigation flow data related to the regulators and service points were collected from the operational SCADA (supervisory control and data acquisition) system known as Total Channel Control™ (TCC™) that is used by Goulburn-Murray Water (GMW). TCC™ is a fully automated open channel delivery system that captures flow measurements in real-time at all regulating structures and farm supply points. We used the flow data recorded at 1016 supply points for the period 15-August-2006 to 15-May-2012; which includes 6 irrigation years. The service point delivery data were aggregated to a daily time step. The SCADA system records the start and end times and flow rates of irrigation order deliveries. Service point flows were aggregated across all service points for each individual channel and the study area as a whole. This aggregation implicitly assumes that the travel time along the local channel or study area is significantly less than a day. The aggregated service point flows are denoted  $ID_{CGi, ASP}$ . Here “ASP” denotes the Aggregated Service Points, and “i” provides spatial aggregation area (where  $i=1, 2, 3, 4$  or 1234).

### 2.2.2 Observed climate data

The observed climate data were obtained from two Automatic Weather Stations (AWSs) and three daily-read rain gauge sites operated by the Australian Bureau of Meteorology (BoM) [BoM, 2005] located in and around the study area (Fig. 1). The characteristics of each weather station are given in Table 2. Daily precipitation records were collected from all five sites as the area often witnesses localised patchy precipitation events. These five sites were established well before

the channel automation and record start dates vary between 1883 and 1996. The hourly weather variables needed to estimate daily  $ET_0$  were collected from the AWSs (air temperature, dewpoint temperature, wind speed) and satellite imagery (solar radiation) as described by [Perera et al., 2014; Perera et al., 2015b]. The measurement ranges and accuracies of these observed weather variables are given in Table 3. The AWSs and rain gauges nominally provide continuous measured weather data; however, there were times within the study period when hourly/daily weather data were missing due to various reasons. For missing daily precipitation, the aggregated value recorded at the end of missing period was uniformly disaggregated over the missing period, while other weather variables were infilled from the neighbouring AWS.

Table 2 – Characteristics of automatic weather stations and rain gauges

N o.	AWS no.	Name	Latitude (degrees)	Longitude (degrees)	El. (m)	Precip. <sup>5</sup>	T <sup>6</sup> , WS <sup>8</sup> & SRad <sup>9</sup>	Dew <sup>7</sup> ,
1	81125	Shepparton Apt. <sup>1</sup>	-36°25'44"	145°23'41"	113.9	✓	✓	
2	81049	Tatura Inst. <sup>2</sup> sus. <sup>3</sup> ag <sup>4</sup>	-36°26'16"	145°16'02"	114.0	✓	✓	
3	81114	Thiess Service	-36°26'40"	145°13'40"	114.0	✓	✗	
4	81053	Waranga Reservoir	-36°30'39"	145°05'25"	121.0	✓	✗	
5	81035	Murchison	-36°36'53"	145°12'51"	115.0	✓	✗	

<sup>1</sup> Airport, <sup>2</sup> Institute, <sup>3</sup> Sustainable, <sup>4</sup> agency, <sup>5</sup> Precipitation <sup>6</sup> Daily temperature, <sup>7</sup> Dew point temperature, <sup>8</sup> Wind speed and <sup>9</sup> Solar Radiation

Table 3 – Measurement Range and Accuracy of Climate Variables from AWS [BoM, 2005]

Sensor	Range	Accuracy	Unit
Air Pressure	750 to 1060	0.3	hPa
Air temperature	-25 to +60	0.3	°C
Wet bulb temperature	-25 to +60	0.3	°C
Relative Humidity	2 to 100	3	%
Wind Speed	2 to 180	2	knot
Wind Direction	0 to 359	5	degree
Precipitation	0 to 999.8	2%	mm

### 2.2.3 Weather forecasts

The short-term weather forecasts relevant for the study area were collected from the Australian Bureau of Meteorology's operational NWP forecasts derived from the Australian Community Climate and Earth System Simulator – (ACCESS) [Puri et al., 2013]. The ACCESS systems are non-hydrostatic, hybrid vertical level structure, mesoscale assimilation & forecast

systems. They have been operational since 17 August, 2010. The temporal (lead time) and spatial resolution for different ACCESS systems vary from +1 to +240 hours and 5 to 80 km, respectively. The forecast performances for precipitation and mean sea level pressure of the ACCESS systems have been comprehensively evaluated [[BoM, 2010](#); [BoM, 2012](#); [Puri et al., 2013](#)] and also its outputs have been extensively used for short-term stream flow forecasting [[Pagano et al., 2010](#); [Shrestha et al., 2012](#); [Shrestha et al., 2013b](#)]. We selected the ACCESS-G system, which has the largest lead time (+240 hrs.) and the largest grid cells (80 km) in order to derive irrigation demand forecasts for longer lead times, where the next higher resolution model has a lead time of only 72 hours. ACCESS-G is run twice a day providing forecasts starting at 10.00 AM and 10.00 PM local time (i.e. 0000 and 1200 UTC). The outputs from these runs are available at 03:50 PM and 03:50 AM (the next day), respectively [[BoM, 2010](#)]. We used the 10:00 AM local time run and constructed 9 mid-night to mid-night daily reference evapotranspiration –  $ET_0$  and 09:00 AM to 09:00 AM daily precipitation forecasts using the 3 hourly NWP forecast outputs of precipitation, air temperature, dew point temperature, wind speed and incoming solar radiation from 17 August 2010 to 01 August 2012. The four grid points surrounding the station were linearly interpolated to the AWS location and biases in the forecast  $ET_0$  input variables were corrected following [[Perera et al., 2014](#)] and precipitation forecasts were post processed following [[Robertson et al., 2013b](#)]. The  $ET_0$  forecasts have been evaluated in detail by [[Perera et al., 2014](#)], who found that the root mean squared error (RMSE) and coefficient of determination ( $R^2$ ) values ranged over 0.73 -1.43 mm day<sup>-1</sup> and 0.91-0.01 for 1 and 9 day lead times, respectively. Precipitation forecasts have  $R^2$  values of 0.65 and 0.55 for 1 and 3 day lead times, respectively [[BoM, 2010](#)].

### 3 Methodology

In this paper, the multivariate time series model developed previously by [[Perera et al., 2015a](#)] is used to derive irrigation demand forecast ensembles for lead time up to 5 days. In principle, the ensemble forecasting approach can include all sources of uncertainties. We undertook some preliminary evaluations to decide which uncertainties to incorporate in this analysis (not presented). This involved perturbing model inputs and parameters individually and collectively and examining the output ensemble spread. It was found that including both model input and parameter uncertainty did not significantly change the ensembles compared with only including model input uncertainty. Including only parameter uncertainty led to very narrow output ensembles. Therefore, the ensemble forecast approach used here was simplified by omitting the parameter component.

Figure 2 provides the schematic diagram of the ensemble forecast approach that has been adopted and shows the flow of input data through to ensemble irrigation demand forecasts along the

time line. Steps 1-3 are about characterizing the statistical structure of each input time series, including measurement error characterisation; correction of NWP forecast bias and characterisation of NWP forecast uncertainty, followed by creation of the input ensembles for each model input. We used different perturbation methods in order to account for measurement, observation and forecast uncertainties given that the deterministic multivariate time series model is forced by observed demands and observed and forecasted weather. The selected perturbation methods are discussed in detail later in this section. We post-processed input ensembles using the Schaake Shuffle method [[Clark et al., 2004](#)] to achieve realistic cross-correlations and temporal persistence within each ensemble. Step 4 involves preparing past and future forcing variable (water supply deficit) and running the deterministic ARMAX models with the various input ensembles. A brief description of the deterministic model and its calibration is given in the next section and further details can be found in [[Perera et al., 2015a](#)]. Step 5 summarizes the daily ensemble irrigation demand forecasts into a probabilistic irrigation demand forecasts.

The various modelling steps above are described in more detail in the following methodology sections. The last part of the methodology describes the evaluation methods that have been used to quantify the forecast performance of ensemble irrigation demand forecast.

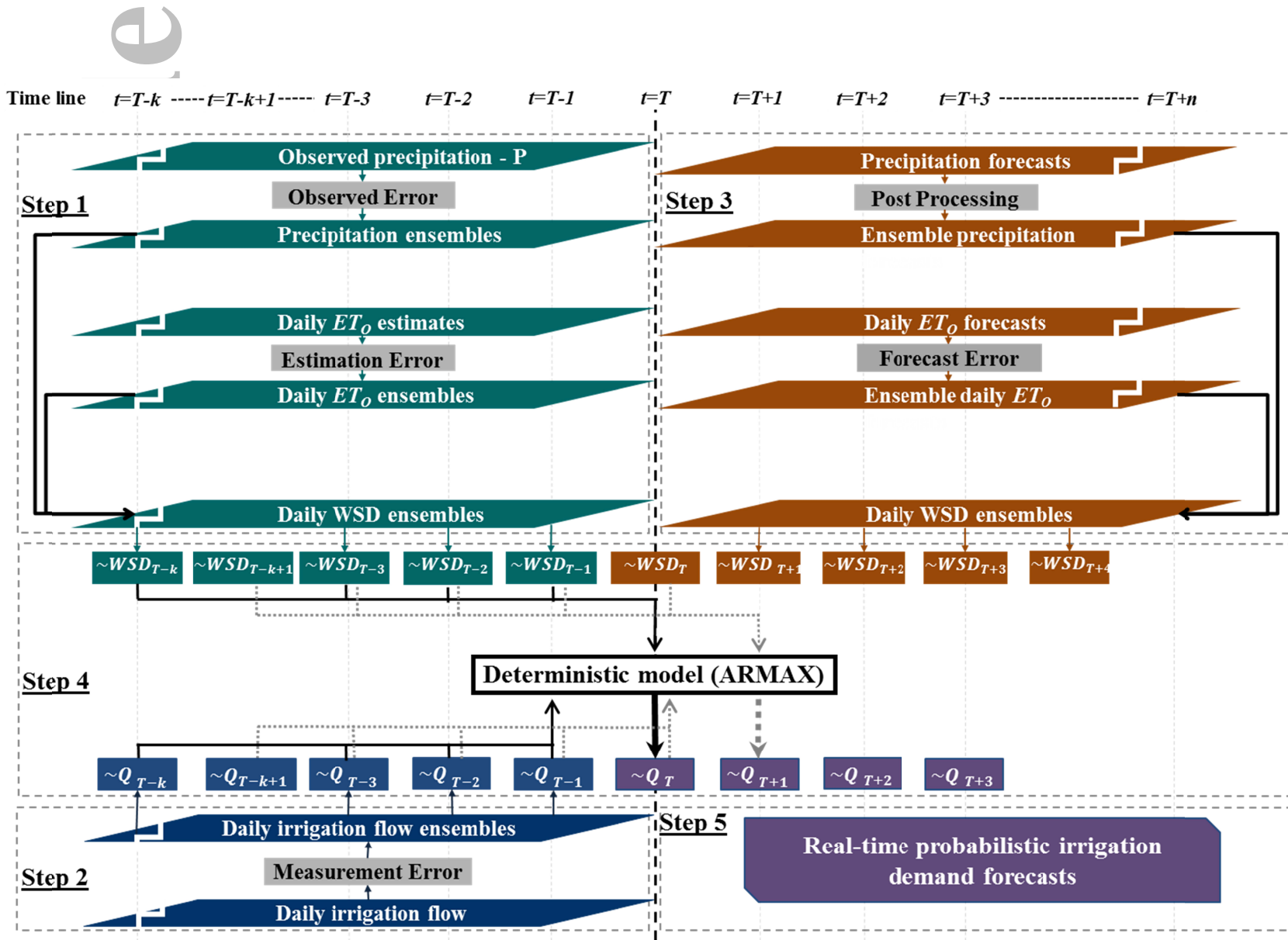


Figure 2 - The schematic diagram of the ensemble forecast approach

## 260 3.1 Deterministic model and its calibration

### 261 3.1.1 Deterministic model structure

262 The model used here is a multivariate time series model, which has been previously  
263 developed to derived irrigation demand forecasts for the same study area [[Perera et al., 2015a](#)].  
264 This time series model combines bio-physical factors (weather) as the exogenous variable with  
265 behaviour factors captured through the auto-regressive process dependent on immediate past  
266 irrigation demands. The model is an autoregressive moving average model with an exogenous  
267 variable (ARMAX) (Eqn. 01).

$$268 \quad ID(t) + a_1 \cdot ID(t - 1) + \dots + a_{n_a} \cdot ID(t - n_a) = b_1 \cdot u(t - n_k) + \dots + b_{n_b} \cdot u(t - n_k - n_b + 1) + \\ 269 \quad c_1 \cdot e(t - 1) + \dots + c_{n_c} \cdot e(t - n_c) + e(t) \quad (1)$$

270 where,  $ID(t)$  is the daily irrigation demand at time  $t$ ,  $u(t)$  is the daily exogenous variable at time  $t$   
271 ,  $n_a$  is the number of past irrigation demand terms included,  $n_b$  is the number of time points at  
272 which the exogenous term is specified plus 1,  $n_c$  is the number of autoregressive error terms,  $n_k$  is  
273 the number of input samples that occur before the input affects the output, also called the dead time  
274 in the system,  $ID(t) \dots ID(t - n_a)$  are the previous outputs on which the current output depends,  
275  $u(t - n_k) \dots u(t - n_k - n_b + 1)$  are the previous and delayed inputs on which the current output  
276 depends,  $e(t) \dots e(t - n_c)$  are the white-noise disturbance values, which is modelled as an  
277 independent and identically distributed (iid) Gaussian process, with zero mean and variance  $\sigma_e^2$ .

278 The exogenous variable ( $u$ ) is a water supply deficit ( $WSD$ ) index which reflects the effect  
279 of atmospheric forcing on irrigation demand as a combination of precipitation and  $ET_O$ . The  $WSD$   
280 is estimated using using a simple soil water bucket model with two parameters.

$$281 \quad WSD_t = ET_{O_t} \times \left[ 1 - \left( \frac{S_t}{S_{max}} \right)^\gamma \right] \quad (2)$$

282 The storage at time  $t$ ,  $S_t$ , is calculated using a soil water balance,

$$283 \quad S_t = \min \left( S_{t-1} + P_{t-1:t} - ET_{O_{t-1:t}} \left( \frac{S_t}{S_{max}} \right)^\gamma, S_{max} \right) \quad (3)$$

284 where,  $ET_O$  ( $\text{mm d}^{-1}$ ) is the daily reference crop evapotranspiration,  $S_t$  ( $\text{mm}$ ) is the soil moisture  
285 storage,  $S_{max}$  is the maximum soil moisture storage ( $\text{mm}$ ),  $P$  ( $\text{mm}$ ) is the precipitation. The bucket  
286 capacity ( $S_{max}$ ) and the nonlinearity ( $\gamma$ ) of actual evapotranspiration are determined in the model  
287 fitting.  $WSD$  is essentially a command area average index aiming to represent the pattern of  
288 supplemental water that would need to be supplied through irrigation.

### 289 3.1.2 Deterministic model calibration

290 A detailed description of the model fitting and selection through cross-validation is  
291 provided in [[Perera et al., 2015a](#)]. This included developing data transformation to remove  
292 seasonality and scaling differences in the raw data set. Then, the model parameters were estimated  
293 in two steps: the first of which determined the order of each transfer function and  $S_{max}$  and  $\gamma$ ; and the  
294 second of which determined the lag coefficients. Model selection was guided using the Bayesian  
295 information criterion (BIC)

$$296 \text{BIC} = N \cdot \log \text{MSE} + d \cdot \log N \quad (4)$$

297 where,  $\text{MSE}$  is the mean squared prediction error,  $d$  is the number of parameters and  $N$  is the  
298 number of observations.

299 This is essentially a mean squared error minimisation penalised by parameter number as a  
300 measure of model complexity. The deterministic model calibration process used a leave-one-year-  
301 out cross-validation (LOOCV) technique to select the best time series models based on the BIC  
302 calibration. In the original paper [[Perera et al., 2015a](#)] 6 years of data were used (calibration on 5  
303 years, validation on 1 year); however, the NWP weather forecasts from ACCESS-G are only  
304 available the 2010-11 and 2011-12 irrigation seasons due to changes in the Bureau's NWP systems.  
305 Therefore this study focusses on just two of the original models; the ones that were validated on the  
306 2010-11 and 2011-12 irrigation seasons. Cross-validations showed that the overall patterns between  
307 command areas and years were similar. The average performances for RMSE and NSE during  
308 calibration, among six cross-validation scenarios across 5 command areas ranged from 2.33 ML/d  
309 (CG 1) to 28.5 ML/d (CG 1234) and from 0.55(CG 1) to 0.93 (CG 1234) respectively, for one day  
310 lead time [[Perera et al., 2015a](#)].

## 311 3.2 Ensemble generation

### 312 3.2.1 Real-time flow data

313 The measurement uncertainty for the irrigation flows was estimated from the manufacture's  
314 specifications and in-field audits. The measurement uncertainty for automated off take regulators  
315 and meter outlets is  $\pm 2.5\%$  (95% confidence) under laboratory test conditions [[Rubicon Water,](#)  
316 [2014](#)]. However, under operational conditions, the Total Channel Control <sup>TM</sup> (TCC<sup>TM</sup>) SCADA  
317 system was assessed as being able to maintain a constant rate of flow as channel level varied.  
318 Specifically, the field assessment concluded that "*the modernised backbone will deliver a uniform*  
319 *flow within +/-5% more than 90% of the time*" [[NVIRP, 2010b](#)]. We have interpreted these two

320 sources of information as representing likely bounds on the in-field performance of the system and  
 321 we consider error scenarios between these bounds. The above statements imply measurement error  
 322 standard deviations (as a proportion of measured flow) of 0.0304 ( $\pm 5.0\%$  with 90% confidence) and  
 323 0.0128 ( $\pm 2.5\%$  with 95% confidence) respectively, assuming a normal distribution. A preliminary  
 324 evaluation (not presented) was carried out to select the measurement uncertainty standard deviation  
 325 for generating ensembles of observed flow. This evaluation considered standard deviations of  
 326 0.0304, 0.025, 0.02, 0.015 and 0.0128. The results shows the standard deviation for measurement  
 327 error for and meter outlets of 0.02 (i.e. 2% of measured flow) generated reasonably reliable one day  
 328 lead time forecast ensembles under calibration conditions. This measurement uncertainty is  
 329 assumed to be a constant proportional error across the full range of flows.

330 In addition, the measurement errors under laboratory conditions [[Rubicon Water, 2014](#)]  
 331 and supply error in the field [[NVIRP, 2010b](#)] need to be generalised across all the automated off  
 332 take regulators and meter outlets as regulators record times of flow rate change and meter outlets  
 333 record the start and end timings and flow rates of irrigation order deliveries. The off take regulator  
 334 and aggregated service point flow ensembles were created similarly, perturbing daily irrigation flow  
 335 time series with an additive percentage error. The additive error is assumed to be a white noise. The  
 336 variance is adjusted for the daily aggregated service point ( $ID_{CGi,ASP}$ ) data, given that it is the sum  
 337 of many individual service points with individual errors that were assumed to be independent of  
 338 each other. Therefore, the irrigation flow ensembles for daily aggregated service point ( $ID_{CGi,ASP}$ )  
 339 are derived using eqns. 5 and 6.

$$340 \quad ID_{CGi,ASP}^{ens} = ID_{CGi,ASP} + \xi_{CGi,ASP}^{ID} \quad (5)$$

$$341 \quad \xi_{CGi,ASP}^{ID} \sim \mathcal{N}(0, \sigma_{CGi,ASP}^2) \text{ and } \sigma_{CGi,ASP} = 0.02 \sqrt{\sum ID_{CGi,ASP}^2} \quad (6)$$

342 Where,  $ID_{CGi,ASP}^{ens}$  is the irrigation demand ensemble for aggregated service,  $ID_{CGi,ASP}$  is the daily  
 343 aggregated service point, and  $\xi_{CGi,ASP}^{ID}$  is the Gaussian additive noise representing measurement  
 344 error with zero mean and a variance of  $\sigma_{CGi,ASP}^2$ . The variance derives from assuming that the 99%  
 345 measurement confidence interval is  $\pm 5.0\%$  of the measured flow.

### 346 **3.2.2 Observed weather variables**

347 The errors in the atmospheric forcing variables for the *WSD* are a key component  
 348 contributing to uncertainties in the irrigation demand estimation and forecasting. Estimation  
 349 uncertainty for the exogenous variable *WSD* results from the precipitation measurement errors and

350 the estimation errors of  $ET_O$ . Given the various non-linearities in converting weather variables to  
351  $ET_O$  and  $WSD$ , a Monte Carlo method is applied to create ensembles for the  $WSD$ , where  
352 precipitation and  $ET_O$  are perturbed with known error parameters.

353 Precipitation,  $ET_O$  and irrigation flow ensembles were post-processed using the Schaake  
354 Shuffle method [[Clark et al., 2004](#)] and historical records. During the calibration, we used a  
355 moving window of 100 days (equivalent to the number for ensembles); either centred, forward  
356 shifted or back shifted, depending on the day of the year. These three types of moving window were  
357 necessary to given that the irrigation distribution operation is inactive for three months (15 May to  
358 15 August) of each year, meaning there is no irrigation flow data for that period. Similar moving  
359 windows were used to post-process the irrigation flow ensembles for the evaluation period, but the  
360 immediate past year's data were used because observations are unknown ahead of time in an  
361 operational forecasting context.

### 362 3.2.2.1 Precipitation

363 Lognormal multiplicative error models have been widely used to generate ensembles for  
364 precipitation and to simulate precipitation error in hydrologic data assimilation [[Alvarez-Garreton](#)  
365 [et al., 2014](#); [Li et al., 2014](#)]. Therefore, we used a lognormal multiplicative error to create the  
366 ensembles representing measured precipitation:

$$367 P_{ens} = \xi_p \times P_{obs} \quad (7)$$

$$368 \xi_p \sim LN(\mu_p, \sigma_p) \quad (8)$$

369 where,  $P_{ens}$  is the daily precipitation ensemble member,  $P_{obs}$  is observed daily precipitation,  $\xi_p$  is  
370 the multiplicative error, which follows a lognormal distribution with the mean of  $\mu_p$  and standard  
371 deviation of  $\sigma_p$ .

372 To create an unbiased precipitation ensemble,  $\mu_p$  is set to 1  $\sigma_p$  set to be 0.25 (25% error).  
373 This error represents both gauge errors and spatial variability. The assumption of the variance of the  
374 precipitation multiplier being 25% is agreed with various studies that have been investigated rain  
375 gauge representativeness errors [[Barancourt et al., 1992](#); [Ciach and Krajewski, 1999](#); [Villarini et](#)  
376 [al., 2008](#)] and also consistent with other recent studies [[Alvarez-Garreton et al., 2014](#); [DeChant and](#)  
377 [Moradkhani, 2012](#); [Li et al., 2014](#)]. The representativeness error of rain gauges for areas of order  
378 250 km<sup>2</sup> is quite variable but of this order. Perera et al., 2015a found the performance of the  
379 ARMAX model was insensitive to choice of individual or combinations of rain gauges in the area,  
380 indicating that the modelling is likely to have relative low sensitivity to this choice.

381 In accordance with the definition of the lognormal probability distribution model, the  
 382 natural logarithm of  $\xi_P$ , denoted as  $\xi_P^{LN}$ , follows a normal distribution with mean and variance  
 383 given by:

$$384 \quad \mu = \ln(\mu_P) - \frac{1}{2} \ln\left(1 + \frac{\sigma_P}{\mu_P}\right) \quad (9)$$

$$385 \quad \sigma = \sqrt{\ln\left(1 + \frac{\sigma_P}{\mu_P}\right)} \quad (10)$$

### 386 3.2.2.2 Reference evapotranspiration – $ET_O$

387  $ET_O$  was calculated based on the guidelines provided by the United Nations Food and  
 388 Agricultural Organization Irrigation and Drainage Paper No. 56 (FAO 56) [Allen et al., 1998]. The  
 389 daily  $ET_O$  ensembles were created by perturbing observed weather variables (daily mean air  
 390 temperature; daily mean dew-point temperature, daily mean wind speed and incoming shortwave  
 391 solar radiation) using zero mean additive noise. The equations used to calculate the inputs to  $ET_O$   
 392 are given in Table 3, along with the corresponding measurement accuracy [BoM, 2005] and the  
 393 standard deviations used in the perturbations. Measurement accuracies were assumed to represent a  
 394 3 standard deviation spread [BoM, 2005] and measurement errors were assumed to be independent  
 395 when calculating the perturbation standard deviations. The perturbation standard deviations  
 396 account for the number of measurements contributing to each  $ET_O$  input.

397 Table 3 – Perturbation methods for  $ET_O$  related weather variables (daily mean temperatures  
 398 ensembles –  $T_{mean_{ens}}$ , daily mean dew point temperatures ensembles –  $DewPt_{mean_{ens}}$ , daily mean  
 399 wind speed ensembles –  $Wndspd_{mean_{ens}}$  and daily solar radiation ensembles –  $Srad_{ens}$ )

Input estimator	Measurement accuracy ( $\pm 3\sigma$ )	Perturbation standard deviation
$T_{mean_{ens}} = \frac{T_{max} + T_{min} + \xi_{Temp}}{2}$	0.3 °C	$\sqrt{2} * \frac{0.3}{3}$
$DewPt_{mean_{ens}} = \frac{1}{24} \sum_{hr=1}^{24} Dewpt_{hr} + \xi_{Dewpt}$	0.3 °C	$\sqrt{24} * \frac{0.3}{3}$
$Wndspd_{mean_{ens}} = \frac{1}{24} \sum_{hr=1}^{24} Wndspd_{hr} + \xi_{Wndspd}$	1.03 ms <sup>-1</sup>	$\sqrt{24} * \frac{1.03}{3}$
$Srad_{ens} = Srad + \xi_{Srad}$	1.5 MJ m <sup>-2</sup>	$\frac{1.5}{3}$

### 400 3.2.3 Forecast weather variables

To forecast irrigation demand, forecast weather is required, which has errors resulting from the NWP weather forecasts. NWP models like ACCESS usually contain both systematic biases due to the NWP model structure, site elevation, temporal and spatial resolution and interpolation techniques, together with noise [Perera et al., 2014; Shrestha et al., 2013b]. In constructing irrigation demand ensembles, the NWP weather forecasts should be corrected for bias and the noise should also be represented. The error characteristics of the forecast precipitation are quite different to the forecast errors for  $ET_O$  and hence they are treated differently here.

Forecast precipitation ensembles were developed using the approach of Wang et al, 2009, which was developed for ACCESS NWP forecasts. This approach uses a simplified version of the Bayesian joint probability technique to derive forecast probability distributions for individual sites as well as for each lead time [Wang et al., 2009]. A joint probability distribution of observations and forecasts is fitted based on past forecast data. Ensemble forecasts are then generated from conditional probability distributions based on the joint probability distribution conditioned on the NWP forecast for that particular day. Space-time correlations are imposed by linking the samples from the forecast probability distributions using the Schaake shuffle [Clark et al., 2004]. Further detail of the post processing can be found in [Robertson et al., 2013b].

While the [Wang et al., 2009] method could potentially also be applied to  $ET_O$  forecasts, differences in the distribution characteristics meant this was unsuccessful. An alternative approach based on the analysis of [Perera et al., 2014] where ACCESS-G NWP weather forecasts were used to generate deterministic daily  $ET_O$  forecasts for lead times of 1 to 9 days. In essence the weather variables input to  $ET_O$  were first bias corrected using a regression between forecast and observed data following [Perera et al., 2014] and then used to make a deterministic  $ET_O$  forecast. The errors in that deterministic forecast were then examined against observed  $ET_O$  for the relevant climate station. This showed that the  $ET_O$  forecast error is multiplicative. Finally, bias-free ensemble daily  $ET_O$  forecasts were generated using the deterministic daily  $ET_O$  forecasts and forecast error realizations based on the observed multiplicative error and then daily  $ET_O$  ensemble forecasts were post-processed using the Schaake Shuffle method as described above.

### 3.3 Evaluation methods

The reliability of the ensemble in terms of bias and spread is the most important attribute of ensemble forecasts as it describes the capability of the ensemble spread to represent the real probabilistic uncertainty of the forecast. In this paper, the reliability of ensemble irrigation demand forecasts is quantified using the temporal mean of the ensemble root mean squared error (MRMSE), the mean continuous ranked probability score (CRPS) and one of its decomposition the mean CRPS

434 reliability [[Hersbach, 2000](#)], which are defined in Eqn. (11) – (13) respectively. The MRMSE is  
 435 derived by calculating the RMSE at each time step using all ensemble members and then averaging  
 436 the RMSEs for the entire period of interest [[Li et al., 2013](#); [Li et al., 2015](#)]. The mean CRPS and  
 437 mean CRPS reliability are calculated following step provided by Brown et al., 2010. The negative  
 438 orientation of these skill scores leads to 0 for a perfect spread and higher skill scores indicates that  
 439 forecasts become more biased or the spread less reliable (or both).

440 The rank histogram approach [[Hamill, 2001](#)] is also used to evaluate the reliability of  
 441 ensemble irrigation demand forecasts. Rank histograms are generated by repeatedly finding the rank  
 442 of the observation relative to values from the forecast ensemble sorted from lowest to highest. The  
 443 resulting ranks are then plotted into a histogram. If the ensemble is reliable, the observations should  
 444 be evenly spread across the various ensemble member ranks. Therefore, a flat rank histogram  
 445 indicates a reliable ensemble spread; a U-shaped (n-shaped) histogram indicates the ensemble  
 446 spread is too narrow (wide); and an asymmetric histogram is a sign of bias. Other attributes such as  
 447 accuracy, sharpness and skill are also important to determine the overall quality of forecasts.  
 448 Therefore, we further quantify the predictive performance of the mean ensemble daily irrigation  
 449 demand forecasts through three additional statistical indices, (1) the root mean squared error  
 450 (RMSE) of the ensemble mean, (2) the Nash-Sutcliffe efficiency coefficient (NS) of the ensemble  
 451 mean and (3) the mean error (BIAS), which are described in Eqn. (14) – (16) respectively. We also  
 452 looked at the ratio between MRMSE and RMSE. This ratio becomes one for a perfect spread, but  
 453 values less than or more than one indicate ensemble forecasts spreads are under or overestimated,  
 454 respectively.

455 All forecast verification scores are normalized using the size of the respective command  
 456 area to make the predictive performance indicators are independent from the size of the command  
 457 area (NSE is unaffected by this normalization). The uncertainties of all of the forecast verification  
 458 scores are evaluated using the bootstrapping technique [[Efron and Tibshirani, 1986](#)] and the 5<sup>th</sup> and  
 459 95<sup>th</sup> confidence intervals are provided along with score.

460 (a) The temporal mean of the ensemble root mean squared error, MRMSE is:

$$461 \text{MRMSE} = \frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{N} \sum_{i=1}^N (ID_{fcst,t}^i - ID_{obsv,t})^2} / CA_j \quad (11)$$

462 (b) The mean continuous ranked probability score (CRPS)

$$463 \text{Mean CRPS} = \frac{1}{T} \sum_{t=1}^T \int_{-\infty}^{\infty} (F_{fcst,t}(ID) - F_{obsv,t}(ID))^2 d ID / CA_j \quad (12)$$

$$464 \quad F_{obsv,t}(ID) = \begin{cases} 0 & (ID < ID_{obsv,t}) \\ 1 & (ID \geq ID_{obsv,t}) \end{cases} \quad (13)$$

465 (c) The root mean squared error, RMSE is:

$$466 \quad RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (\overline{ID_{fcst,t}^i} - ID_{obsv,t})^2} / CA_j \quad (14)$$

467 (d) Nash–Sutcliffe (NS) model efficiency coefficient (NSE) is:

$$468 \quad NSE = 1 - \frac{\sum_{i=1}^T (\overline{ID_{fcst,t}^i} - ID_{obsv,t})^2}{\sum_{i=1}^T (ID_{obsv,t} - \overline{ID_{obsv,t}})^2} \quad (15)$$

469 (e) The the mean error (BIAS) is:

$$470 \quad BIAS = \frac{1}{T} \sum_{i=1}^T (\overline{ID_{fcst,t}^i} - ID_{obsv,t}) / CA_j \quad (16)$$

471 In equations 11-16,  $ID_{fcst,t}^i$  is the  $i^{th}$  ensemble prediction at time,  $t$ , and  $ID_{obsv,t}$  is the  
 472 observed values at  $t$ ,  $N$  is the number of ensembles,  $T$  is the total number of time steps  $ID_{fcst,t}^i$  is the  
 473 predicted ensemble mean at  $t$ ,  $\overline{ID_{obsv,t}}$  is the mean of  $ID_{obsv,t}$  for the period of verification (POV)  
 474 and  $CA_j$  is the size of  $j^{th}$  command area.

## 475 **4 Results**

### 476 **4.1 Precipitation and $ET_O$ forecast uncertainties**

477 The exogenous variable,  $WSD$ , acts as a weather forcing variable for irrigation demand and  
 478 it is calculated using precipitation and reference evapotranspiration ( $ET_O$ ). The forecast and  
 479 observed uncertainties in these two weather variables contribute to irrigation demand forecast  
 480 uncertainties. The observation uncertainties for the weather variables are instrument and  
 481 measurement network dependent, while the forecast uncertainties depend on the forecast skill of the  
 482 NWP system. The observation uncertainties for weather variables were obtained from the  
 483 Australian Bureau of Meteorology [[BoM, 2005](#)]. For the weather forecasts, ensemble forecasts were  
 484 first constructed as outlined in section 3.2.3. Detailed evaluations of these ensemble forecasts for  
 485 the study region were then undertaken through comparisons with local observations. This section  
 486 briefly evaluates the forecast performance of the precipitation and  $ET_O$  ensemble forecasts derived  
 487 from ACCESS-G.

Table 4 summarizes the statistical indicators related to the forecast performance for precipitation and  $ET_o$  ensemble forecasts and Figures 3 and 4 summaries the results for precipitation and evapotranspiration, respectively. These stats are for the periods of 15-08-2010 to 15-05-2011 and 15-08-2011 to 15-05-2012 (551 days). The forecast performance for ensemble daily  $ET_o$  and precipitation forecasts declines with increasing lead time, as expected. For precipitation forecasts, the MRMSE, mean CRPS and RMSE increase approximately 42%, 32% and 27% respectively and mean CRPS reliability remain same, as the lead time increases from 1 to 5 days. For  $ET_o$  forecasts the MRMSE, mean CRPS, mean CRPS reliability and RMSE increase more than twice for lead time 5 days and indicated spread of  $ET_o$  ensemble forecasts is poor reliable for longer lead times. The NS efficiency coefficient for mean precipitation and  $ET_o$  ranged between 0.51 and 0.20 and between 0.90 and 0.77, respectively, for lead times of 1 to 5 days. The rank histogram and bias shows that the ensemble precipitation forecast was slightly negatively biased and that the ensemble  $ET_o$  forecast was slightly positively biased and the uncertainty was over-estimated for  $ET_o$ . The results also suggest that the forecast performance for  $ET_o$  was higher than for precipitation.

Table 4 – Statistical indicators related to prediction performance for ensemble precipitation and  $ET_o$  forecasts for study area between 15-08-2010 and 15-05-2011 and 15-05-2012 to 15-05-2012 (551 days). The range shown in brackets is the 5<sup>th</sup> -95<sup>th</sup> confidence interval from the bootstrapping analysis.

Lead time	Ensemble verification score	Variable	
		Precipitation	$ET_o$
1 Day	MRMSE mm day <sup>-1</sup>	1.89 (1.6-2.19)	1.00 (0.96-1.04)
	Mean CRPS mm day <sup>-1</sup>	1.01 (0.79-1.27)	0.37 (0.34-0.39)
	Mean CRPS Rel. mm day <sup>-1</sup>	0.06 (0.03-0.12)	0.02 (0.01-0.02)

	RMSE (mean) mm day <sup>-1</sup>	4.52 (3.48-5.53)	0.72 (0.67-0.78)
	NSE (mean)	0.51 (0.42-0.62)	0.90 (0.88-0.91)
	BIAS (mean) mm day <sup>-1</sup>	-0.69 (-0.98--0.43)	-0.09 (-0.13--0.04)
3 Day	MRMSE mm day <sup>-1</sup>	2.46 (2.12-2.84)	1.37 (1.32-1.43)
	Mean CRPS mm day <sup>-1</sup>	1.26 (0.99-1.58)	0.48 (0.45-0.51)
	Mean CRPS Rel. mm day <sup>-1</sup>	0.05 (0.03-0.1)	0.03 (0.03-0.04)
	RMSE (mean) mm day <sup>-1</sup>	5.67 (4.18-7.11)	0.9 (0.82-0.97)
	NSE (mean)	0.23 (-0.07-0.46)	0.84 (0.81-0.87)
	BIAS (mean) mm day <sup>-1</sup>	-0.63 (-0.99--0.31)	-0.09 (-0.15--0.04)
5 Day	MRMSE mm day <sup>-1</sup>	2.7 (2.34-3.07)	1.67 (1.61-1.74)
	Mean CRPS mm day <sup>-1</sup>	1.34 (1.05-1.64)	0.58 (0.55-0.62)
	Mean CRPS Rel. mm day <sup>-1</sup>	0.06 (0.03-0.12)	0.05 (0.04-0.06)
	RMSE (mean) mm day <sup>-1</sup>	5.75 (4.42-7)	1.07 (0.99-1.16)
	NSE (mean)	0.2 (0.04 -0.37)	0.77 (0.73-0.81)
	BIAS (mean) mm day <sup>-1</sup>	-0.66 (-1.02--0.34)	-0.14 (-0.21--0.08)

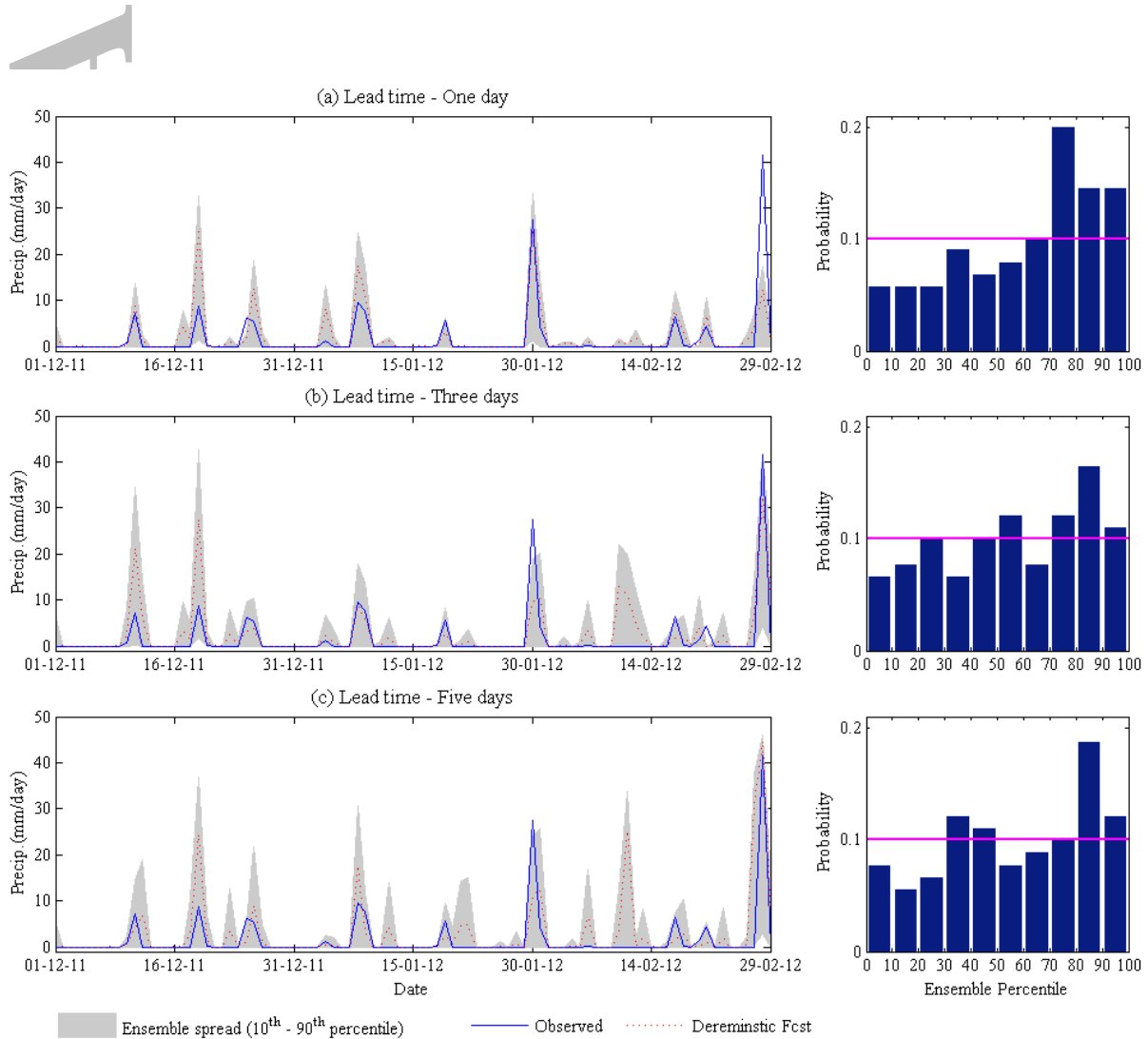


Figure 03 – Time series plot (91 days) of observed daily precipitation vs. post process ensemble precipitation forecast and respective rank histograms (551 days) for lead times of 1, 3 and 5 days.

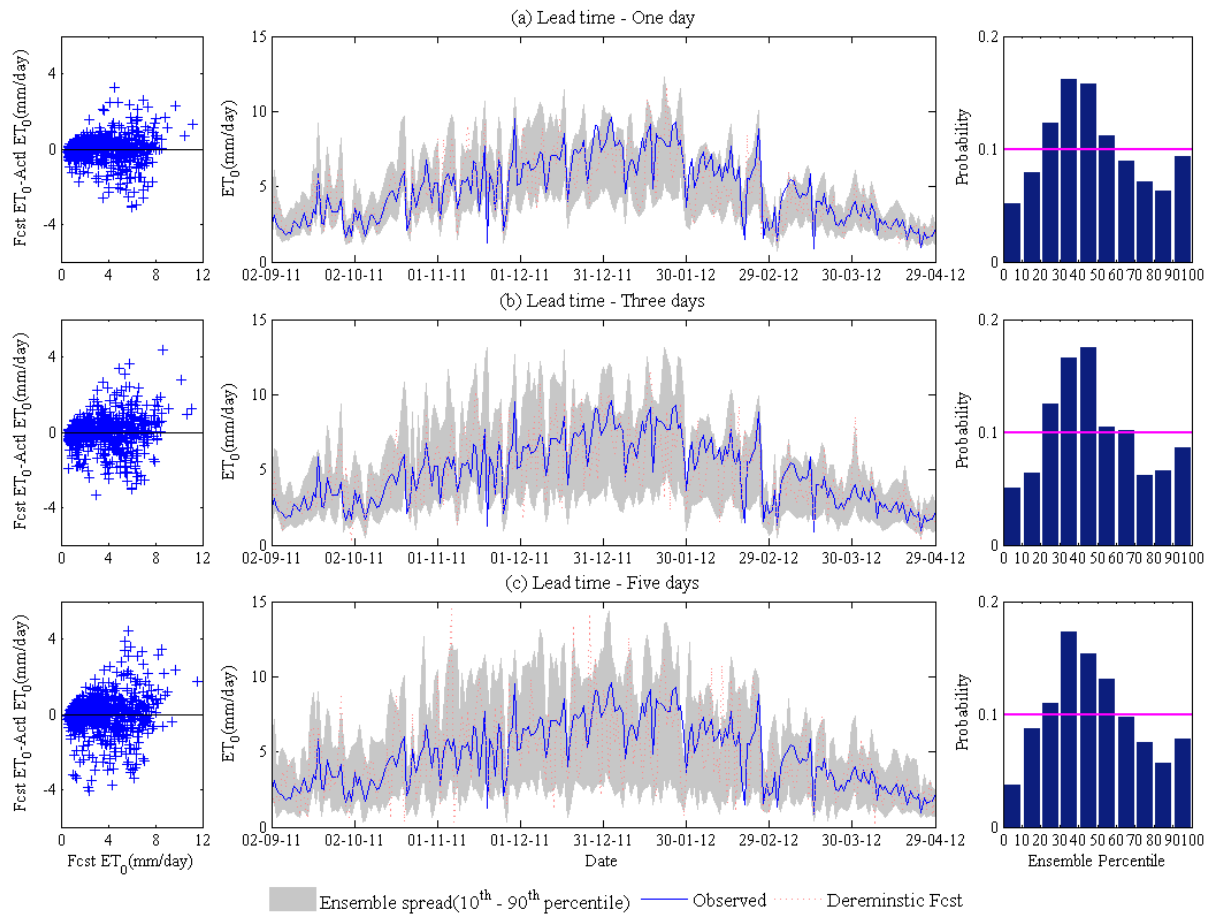


Figure 04 –The scatter plot (551 days) between the deterministic daily  $ET_0$  forecast and forecast error, time series plot (241 days) of observed daily  $ET_0$  vs. ensemble daily  $ET_0$  forecast with spread between 10<sup>th</sup> and 90<sup>th</sup> percentile and respective rank histograms (551 days) for lead times of 1, 3 and 5 days.

#### 4.2 Ensemble demand forecasts for the model calibration period

The model parameters reported in [Perera et al., 2015a] are used here. The following results are based on forecasts for the 2010-11 and 2011-12 irrigation seasons. The deterministic model was calibrated to the 2006-07, 2007-08, 2008-09 and 2009-10 seasons plus one of 2010-11 or 2011-12. Independent validation results were obtained for each of the 2010-11 and 2011-12 irrigation seasons. Tabulated values below were obtained by calculating the calibration (or validation) metrics for each of the two seasons and then averaging these. We note that the

“calibration period” results correspond to only one of the five years contributing to the original model fitting.

Table 5 provides the average statistical indicators between two calibration scenario for ensemble daily  $ID_{CG i, ASP}$  forecasts across all 5 command areas and the performance for each individual calibration scenario is in Table A-1 (Supplementary material Table A-1). These calibration periods have four years in common, with the fifth year being different between the two sets of results and the results in Table 5 only represent the average of those fifth year. Only the results relating to the best LOOCV calibration scenario and the full study area (i.e. LOOCV-S 1 for CG 1234) are shown graphically as time series (Figure 5) and rank histograms (Figure 6). The statistical indicators for the two scenarios corresponding to each command area vary significantly due to the differences in irrigation flows between years. We start by considering the ensemble mean forecast. The NSE between forecast and observed irrigation demand ranged between 0.91 (1 day lead time, CG 1234) and 0.31 (5 day lead time, CG 1). The highest and lowest NSE values were always found at CG 1 234 and CG 1, respectively due to differences in the area irrigated and number of supply points. For 5 day lead times compared with 1 day lead times, NSE decreased by more than 40 % for small command areas (CG 1, CG 2 and CG 3) and approximately 25% for larger command areas (CG 4 and CG 1234). The bias was highest at CG 1234 and lowest at CG 1 and it is proportional to the size of the area irrigated. The bias showed that the ensemble daily demand forecasts slightly over predicted the observed flows for all command areas and mostly remains approximately same with increased in lead time. .

The ensemble spread can be evaluated through the MRMSE, mean CRPS, mean CRPS reliability, RMSE and the rank histograms. The difference between MRMSE and RMSE increased with the increasing command area or decreasing lead times. The mean CRPS, mean CRPS reliability and difference between these two scores increased with the increasing command area or lead times. These facts suggest that reliability of ensemble spread declines with increasing command area or lead times.. The MRMSE was higher than the corresponding RMSE for all command areas and for all lead times. Across the five command areas, MRMSE, mean CRPS, mean CRPS reliability and RMSE increase on average by approximately 23 %, 53%, 8% and 73% respectively from 1 day to 5 days lead time. For most cases, these statistics and the rank histograms (Figure 06) show that the ensemble spread marginally overestimated the forecast error variability for the first day and then tend to be flat with increased in lead time, indicating a reliable ensemble spread. During the calibration periods, the observed daily  $ID_{CG 1234, ASP}$  values were within the ensemble daily  $ID_{CG i, ASP}$  forecast spread (10<sup>th</sup> -90<sup>th</sup> percentile) for 86%, 83% and 82% of the time with respect to lead time 1, 3 and 5 days.

Table 5 – Average forecast performance for ensemble daily  $ID_{CG i, ASP}$  forecasts related to the 2 cross validation scenarios for 4 channels and study area during the last year of calibration periods 15-08-2010-15-05-2011 (274 days) or 15-08-2011-15-05-2012 (275 days)). The range shown in brackets is the 5th -95th from the bootstrapping

Lead time	Ensemble verification score	Average Performances for Command Area				
		CG1	CG2	CG3	CG4	CG1234
1 Day	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.39 (0.38-0.41)	0.41 (0.39-0.42)	0.32 (0.31-0.33)	0.36 (0.34-0.37)	0.25 (0.24-0.26)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.21 (0.2-0.22)	0.19 (0.18-0.2)	0.12 (0.11-0.12)	0.11 (0.1-0.11)	0.08 (0.08-0.08)
	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.05 (0.04-0.06)	0.01 (0.01-0.02)	0.01 (0.01-0.01)	0.03 (0.02-0.03)	0.01 (0.01-0.01)
	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.38 (0.36-0.41)	0.34 (0.32-0.36)	0.2 (0.19-0.22)	0.15 (0.14-0.16)	0.12 (0.12-0.13)
	NSE (mean)	0.46 (0.37-0.52)	0.73 (0.69-0.75)	0.83 (0.8-0.85)	0.91 (0.9-0.92)	0.94 (0.93-0.94)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.1 (0.08-0.13)	0.06 (0.03-0.08)	0.04 (0.03-0.06)	0.05 (0.04-0.06)	0.03 (0.02-0.04)
	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.48 (0.46-0.5)	0.49 (0.47-0.51)	0.37 (0.36-0.38)	0.38 (0.36-0.39)	0.32 (0.31-0.33)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.27 (0.25-0.29)	0.25 (0.24-0.26)	0.17 (0.16-0.18)	0.14 (0.14-0.15)	0.13 (0.13-0.14)
	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.06 (0.06-0.08)	0.02 (0.02-0.03)	0.01 (0.01-0.01)	0.01 (0.01-0.01)	0 (0-0.01)
3 Day	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.48 (0.45-0.51)	0.44 (0.42-0.46)	0.3 (0.28-0.31)	0.25 (0.23-0.26)	0.24 (0.22-0.25)
	NSE (mean)	0.17 (0.07-0.22)	0.55 (0.51-0.59)	0.64 (0.6-0.67)	0.76 (0.73-0.78)	0.77 (0.74-0.8)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.1 (0.06-0.13)	0.04 (0.01-0.07)	0.03 (0.01-0.05)	0.05 (0.03-0.07)	0.04 (0.02-0.06)
	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.48 (0.46-0.5)	0.52 (0.5-0.53)	0.39 (0.38-0.4)	0.37 (0.36-0.38)	0.35 (0.34-0.36)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.27 (0.25-0.29)	0.27 (0.26-0.29)	0.19 (0.18-0.2)	0.15 (0.15-0.16)	0.16 (0.15-0.17)
5 Day	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.07 (0.06-0.08)	0.02 (0.02-0.03)	0.01 (0.01-0.02)	0 (0-0.01)	0.01 (0-0.01)
	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.48 (0.45-0.51)	0.48 (0.45-0.5)	0.34 (0.32-0.36)	0.28 (0.26-0.29)	0.29 (0.27-0.3)
	NSE (mean)	0.15 (0.06-0.21)	0.48 (0.44-0.51)	0.52 (0.48-0.56)	0.69 (0.66-0.72)	0.67 (0.63-0.71)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.1 (0.06-0.13)	0.03 (-0.01-0.06)	0.02 (0-0.05)	0.04 (0.02-0.06)	0.03 (0.01-0.05)

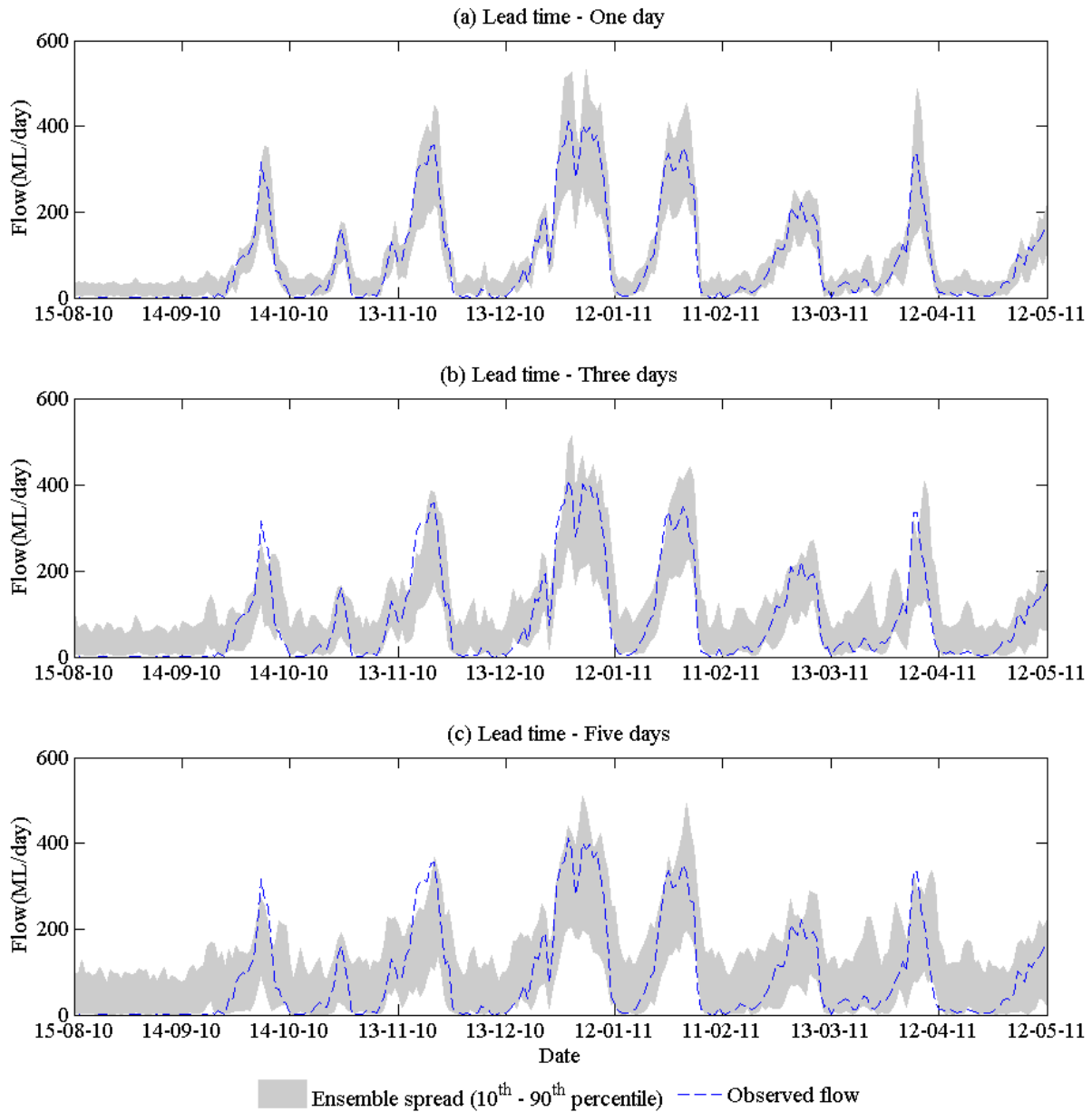


Figure 05 – Time series plot of observed vs. ensemble daily  $ID_{CG 1234, ASP}$  forecasts with the ensemble spread between 10<sup>th</sup> and 90<sup>th</sup> percentile for lead times of 1, 3 and 5 days for irrigation year 15-08-2010-15-05-2011(274 days - validation scenario 1 in the table 5)

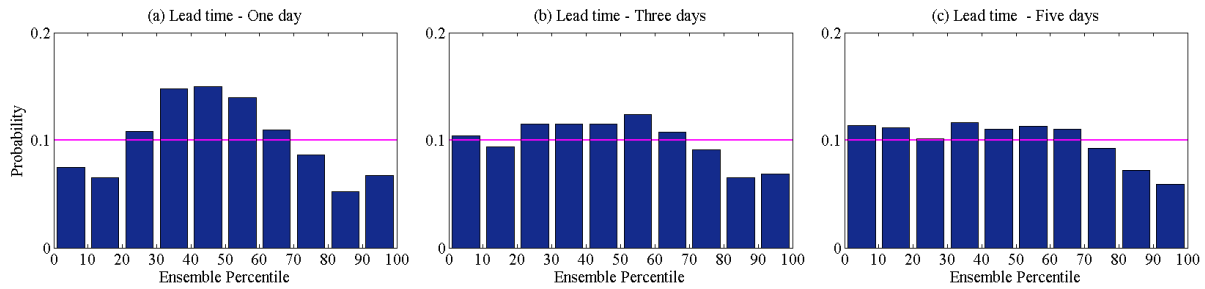


Figure 06 – Rank histograms for observed and ensemble daily  $ID_{CG\ 1234, ASP}$  forecasts for lead times of 1, 3 and 5 for irrigation years 15-08-2010 to 15-05-11(274 days - validation scenario 1 in the table 5)

### 4.3 Real-time forecasting with numerical weather prediction outputs

We now examine the results for validation conditions using a similar approach to the previous section. With the exception of the fitting of uncertainty parameters for the NWP ensembles, these demand forecasts are made under operational conditions; that is using observation data available from automatic weather stations, the SCADA network of supply points and NWP forecast data, all of which would be available at the time of making an operational forecast, and with the ARMAX model fitted to independent data. The aim is to test the forecast performance of both the ensemble mean and to assess how well the ensemble represents the forecast uncertainty.

Table 6 provides the average statistical indicators between two evaluation scenario for ensemble daily  $ID_{CG\ i, ASP}$  forecasts across all 5 command areas and the performance for each individual evaluation scenario is in Table A-2 (Supplementary material Table A-2). Figure 7 shows the time series plots of observed and ensemble daily demand forecast for evaluation year 2011/12, while Figure 8 shows the rank histograms and associated probabilities. The NSE values ranged between 0.97 (1 day lead time, CG 4) and 0.22 (5 day lead time, CG 1) and for the whole study area it ranged between 0.97 and 0.75 for 1 and 5 day lead times, respectively. The bias varied with the evaluation year, being positive for 2010/11 and negative for 2011/12. This is partly reflects the variability of irrigation flow within each year, where the highest variability occurred during year 2011/12.

The ensemble spread is again examined by using MRMSE, mean CRPS, mean CRPS reliability and RMSE and rank histograms. In general all MRMSE, mean CRPS, mean CRPS reliability and RMSE decrease with increasing irrigated area reflecting the higher forecast performances for larger command areas and the spread of forecast uncertainties is not reliable for smaller command areas. The pattern between the MRMSE, mean CRPS, mean CRPS reliability and RMSE was similar to calibration period, where MRMSE is higher for all command areas and all lead times. The

Accepted Article

difference between MRMSE and RMSE increased and the difference between mean CRPS and mean CRPS reliability decreased with the increasing command area and decreasing lead times. Typically MRMSE is larger than RMSE values indicating that the ensemble spread is somewhat overestimating the forecast uncertainty at system scale, when irrigation flow among distribution channels aggregated. The average ratio MRMSE/RMSE across two validation years varies between 1.10 and 4.20 for all command areas for one day lead time. Ideally this ratio should be one and significant deviations from one indicate that the ensemble spread is unreliable. For the small command areas (CG 1 and CG 2) MRMSE/RMSE varies between 1.16 and 1.44, indicating the ensemble spread is slightly over estimated for the smaller command areas and consistently overestimated for the larger command areas. The rank histograms for the whole study area show that the ensemble is marginally too wide and slightly negative biased for one day lead time. This negative bias increases with lead time and forecast ensemble spreads are slightly under-dispersive for lead times of three and five days. This is consistent with the indications from mean CRPS, mean CRPS reliability and MRMSE vs RMSE; however, those statistics are likely to be more sensitive to outliers than the rank histograms. During the evaluation periods, the observed daily  $ID_{CG\ 1234, ASP}$  values were within the ensemble daily  $ID_{CG\ i, ASP}$  forecasts spread (10<sup>th</sup> -90<sup>th</sup> percentile) for 86%, 80% and 72% times for lead times of 1,3 and 5 days respectively. Overall, these results suggest that the ensemble forecasts give good estimates of forecast uncertainty and reliable probabilistic irrigation demand forecasts, particularly for the large command areas, but that the ensemble spread does not grow quickly enough with time.

Table 6 – Average forecast performance for ensemble daily  $ID_{CG\ i, ASP}$  forecasts related to the 2 cross validation scenarios for 4 channels and study area during the evaluation periods (15-08-2010-15-05-2011 (274 days) or 15-08-2011-15-05-2012 (275 days)). The range shown in brackets is the 5<sup>th</sup> -95<sup>th</sup> from the bootstrapping

Lead time	Ensemble verification score	Average Performances for Command Area				
		CG1	CG2	CG3	CG4	CG1234
1 Day	MRMSE	0.35	0.39	0.31	0.35	0.23
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.34-0.36)	(0.38-0.4)	(0.3-0.33)	(0.33-0.37)	(0.22-0.25)
	Mean CRPS	0.17	0.15	0.1	0.1	0.07
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.16-0.17)	(0.14-0.16)	(0.1-0.11)	(0.1-0.11)	(0.07-0.07)
	Mean CRPS Rel.	0.03	0.02	0.02	0.04	0.02
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.03-0.04)	(0.01-0.02)	(0.02-0.02)	(0.03-0.04)	(0.02-0.02)
	RMSE (mean)	0.3	0.27	0.15	0.11	0.09
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.28-0.31)	(0.26-0.28)	(0.14-0.16)	(0.1-0.11)	(0.08-0.09)
3 Day	NSE (mean)	0.63	0.82	0.9	0.95	0.96
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.55-0.68)	(0.79-0.83)	(0.88-0.91)	(0.94-0.96)	(0.96-0.97)
	BIAS (mean)	0.1	0.01	0.02	0.02	0.01
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.08-0.11)	(0-0.03)	(0.01-0.03)	(0.02-0.03)	(0-0.02)
	MRMSE	0.41	0.46	0.35	0.36	0.28
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.4-0.42)	(0.44-0.47)	(0.34-0.36)	(0.35-0.37)	(0.28-0.29)
	Mean CRPS	0.2	0.21	0.13	0.12	0.11
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.19-0.21)	(0.2-0.22)	(0.13-0.14)	(0.12-0.13)	(0.1-0.11)
5 Day	Mean CRPS Rel.	0.04	0.04	0.02	0.02	0.01
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.04-0.05)	(0.03-0.05)	(0.01-0.02)	(0.02-0.02)	(0.01-0.02)
	RMSE (mean)	0.37	0.39	0.25	0.2	0.18
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.35-0.39)	(0.37-0.4)	(0.23-0.26)	(0.19-0.21)	(0.18-0.19)
	NSE (mean)	0.44	0.64	0.74	0.84	0.86
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.32-0.51)	(0.6-0.67)	(0.71-0.77)	(0.82-0.85)	(0.84-0.87)
	BIAS (mean)	0.09	-0.04	-0.02	-0.01	-0.01
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.07-0.12)	(-0.06--0.01)	(-0.03-0)	(-0.03-0)	(-0.03-0)
5 Day	MRMSE	0.41	0.48	0.37	0.37	0.31
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.4-0.43)	(0.46-0.5)	(0.35-0.38)	(0.36-0.39)	(0.3-0.32)
	Mean CRPS	0.21	0.23	0.15	0.14	0.13
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.2-0.22)	(0.22-0.24)	(0.15-0.16)	(0.14-0.15)	(0.13-0.14)
	Mean CRPS Rel.	0.05	0.05	0.03	0.03	0.02
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.04-0.06)	(0.04-0.06)	(0.02-0.03)	(0.02-0.03)	(0.02-0.03)
	RMSE (mean)	0.38	0.43	0.29	0.25	0.24
	(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.36-0.4)	(0.41-0.45)	(0.28-0.3)	(0.24-0.26)	(0.23-0.25)
NSE (mean)	0.42	0.57	0.65	0.75	0.76	
(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.31-0.49)	(0.52-0.6)	(0.61-0.68)	(0.73-0.77)	(0.74-0.79)	
BIAS (mean)	0.09	-0.07	-0.04	-0.05	-0.04	
(ML day <sup>-1</sup> Km <sup>-1</sup> )	(0.06-0.11)	(-0.09--0.04)	(-0.06--0.02)	(-0.07--0.04)	(-0.05--0.02)	

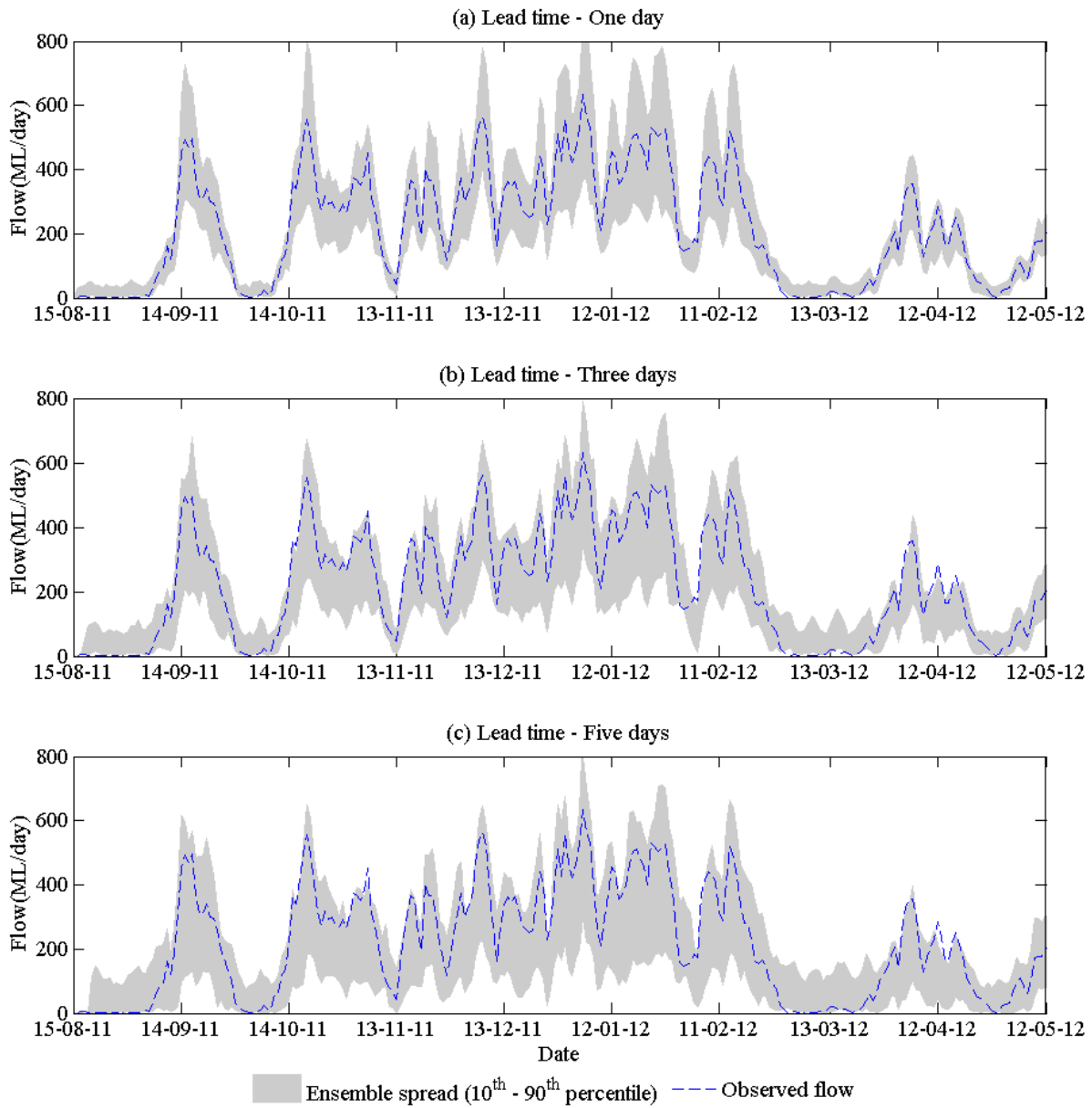


Figure 07 - Time series plot of observed vs. ensemble daily  $ID_{CG\ 1234, ASP}$  forecasts with the ensemble spread between 10<sup>th</sup> and 90<sup>th</sup> percentile for lead times of 1, 3 and 5 for irrigation year 15-08-2011 to 15-05-2012 (275 days)

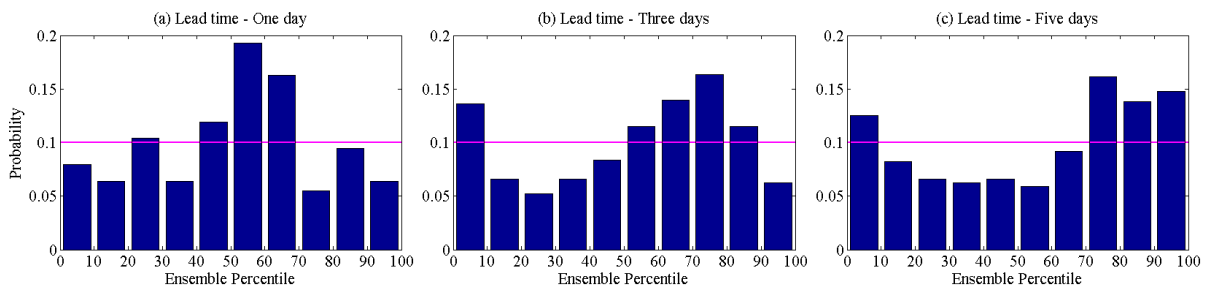


Figure 08 – Rank histograms for observed and ensemble daily  $ID_{CG\ 1234, ASP}$  forecasts for lead times of 1, 3 and 5 days during irrigation year 15-08-2011 to 15-05-2012 (275 days).

## 5 Discussion

The ensemble forecasting techniques used here build on those used in forecasting short runoff or stream flow [[Addor et al., 2011](#); [Bennett et al., 2014](#); [Robertson et al., 2013a](#); [Shrestha et al., 2013a](#); [Smiatek et al., 2012](#)]. Those studies have either used stochastic precipitation forecasts that have been post-processed or output from NWP models. While, the inputs and the techniques are similar for forecasting stream flows and irrigation demands, the dynamic responses of the systems are the opposite; with irrigation demand decreasing when precipitation increases and evapotranspiration decreases and streamflow doing the opposite. In catchments there is also likely to be more variation in antecedent conditions prior to precipitation events (antecedent moisture tends to be controlled by the application of irrigation in irrigated fields). This means that the forecast performance is not directly comparable between streamflow and irrigation demand.

The forecast performance for the ensemble irrigation demand predictions were evaluated for lead times of 1 to 5 days across the four command areas plus the full study area. Both the average forecasts (ensemble means) and the uncertainty estimates (ensemble spread) performed well overall. NSEs for forecast conditions were up to 0.97 for one day lead time and larger command areas and remained above 0.65 for 5 day lead times, with the exception of the two smallest command areas. There were clear dependencies on lead time and command area. The area dependence relates to the amount of averaging between individual irrigation farms while the dependence on lead time relates to accumulation of forecast errors over time.

From the system operators' perspective, these probabilistic short-term system scale irrigation demand forecasts could assist with planning operations, particularly those operations such as transferring water from the main storage at Lake Eildon to the irrigation command areas. The ensemble approach provides a reliable estimate of forecast uncertainty that could be used to inform operational risks. The modelling also provides a quantitative link to enable operational weather forecasts to be more easily utilized in system operation decisions.

From the forecasting context, this is the first study to generate probabilistic daily irrigation demand forecasts with lead times up to 5 days at the system scale. The results showed that the forecasting performance for ensemble daily irrigation demand forecasts were higher than most previous studies [[Alfonso et al., 2011](#); [Pulido-Calvo and Gutierrez-Estrada, 2009](#); [Tian and Martinez, 2014](#); [Ticlavilca et al., 2011](#)] and the performance of the forecasted ensemble mean is marginally lower than the deterministic forecasts derived by [[Perera et al., 2015a](#)] using the same

time series model. The rank histograms highlighted that the spreads were slightly over dispersive for lead time 1 and then change to slightly under dispersive as lead time increases. This implies the influence of other sources of uncertainty that are not being fully captured by input errors. There are a number of influences not explicitly included in the model. These include responses to regular (~monthly) adjustments in yearly water allocation, changes in the price of water in the market, variations in sowing date for crops, on-farm storages and stock and domestic (non-irrigation) water leading to additional errors. These are in general model structural uncertainties, which we have not dealt with in this paper.

There are some subtle differences in the evolution of ensemble spread in different situations. In general, where there is a low antecedent flow, the one day lead time spread is small and it increases with antecedent flow. The low initial spread is due primarily to the multiplicative error applied to the observed flows, which has the largest influence on the lead one ARMAX outputs. The evolution of the ensemble spread over time is influenced by both the initial ensemble spread and the contributions to spread from the error component of the ARMAX model and the weather forecasts. The error structure for the observed reference evapotranspiration is complicated by its propagation from measurements through the Penman-Monteith equation. The inputs to the observed  $ET_O$  calculation are assumed to have additive errors. The precipitation observation error model is multiplicative. These choices of measurement error structure reflect the nature of measurement errors for the different types of instruments involved. The ensemble evolution is influenced by the error structure in the ARMAX model itself and the weather forecasts. The ARMAX model has an additive error structure, which is justified by the forecast errors from the deterministic ARMAX model (Perera *et al.*, 2015a, Figs 8 and 9). The forecast  $ET_O$  also has an additive error structure, justified by error analysis in Perera *et al.* (2014), Fig 7. The precipitation forecast errors are determined by the structure of the Bayesian Joint Probability model [Robertson *et al.*, 2013a; Wang *et al.*, 2009]. Of these, the additive errors in the ARMAX model have the greatest effect on the ensemble evolution under different conditions and they lead to rapid ensemble widening when the initial spread is small but slow widening when the initial spread is larger.

As discussed in the previous paragraph, there are a variety of sources of uncertainty incorporated into the ensembles, including uncertainty in antecedent conditions (observed flow/demand, observed weather) and forecast weather uncertainty. To further understand the importance of the antecedent conditions in contributing to the ensemble variability, we undertook two additional sets of simulations. The first excluded uncertainty in the antecedent flow and the second excluded uncertainty in the antecedent weather. Figure 9 shows the resulting rank histograms. The top row of histograms indicates that the ensembles become biased when flow

uncertainty is excluded. Comparing the middle and lower rows in Figure 9 shows that including the impact of antecedent weather uncertainty also leads to subtle improvements in the ensemble (less under-dispersion, slightly less bias), but the improvements are smaller than the effect of antecedent flow. These results suggest that including the antecedent conditions as a source of uncertainty is important and that flow is more important than past weather. The difference in influence of past flow and past weather is as expected given that the ARMAX model weights past flows more heavily than past weather.

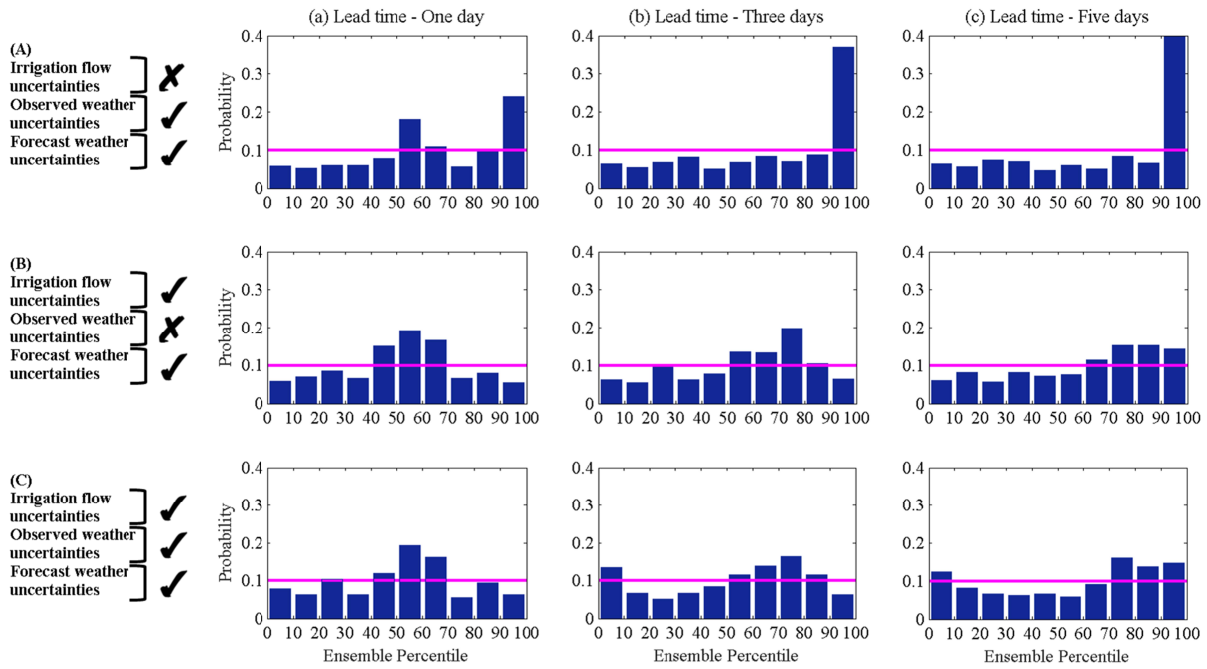


Figure 09 – Rank histograms (A) Excluding irrigation flow uncertainties (B) Excluding observed weather uncertainties and (C) Including irrigation flow and observed and forecast weather uncertainties, for observed and ensemble daily  $ID_{CG\ 1234, ASP}$  forecasts for lead times of 1, 3 and 5 days during the evaluation period (15-08-2011 to 15-05-2012 (275 days)).

In this paper, the input error terms are assumed to be independently and identically distributed. These assumptions are not valid in practice because the time series plots show a serial correlation of the irrigation demand forecast error. This might be due to structural errors in the model but could also be partly due to serial correlation of the  $ET_O$  forecast error as the serial correlation of the precipitation error was corrected using the post-processing approach following [Robertson et al., 2013b]. This serial correlation of irrigation demand forecast error is highly subjective for non-stationary time series like previous irrigation flows and accordingly, the systematic bias increased with the increase in lead time. However, in the face of higher auto-

correlation between consecutive observed irrigation flows, the serial correlation of  $ET_O$  forecast error is not significant. Nevertheless, ensemble forecasting scheme has provided sufficiently accurate probabilistic irrigation demand forecasts that can be useful to the system operators for their routine irrigation distribution decisions.

## 6 Conclusion

This paper developed an ensemble forecasting scheme to generate a real-time probabilistic short-term system scale irrigation demand forecasts. We used a deterministic multivariate time series model that was developed previously [[Perera et al., 2015a](#)], along with real-time data recorded at irrigation regulators; short-term weather forecast derived from an NWP model and observed weather variables recorded at automatic weather stations. This method was applied to four neighbouring irrigated agricultural areas operating under fully automated irrigation distribution system.

The observation, estimation and forecast uncertainties were guided by the manufacturer's specification, obtained from the literature or calculated. A range of perturbation methods were used to generate ensemble for each input that is necessary for the deterministic time series models. The spatial and temporal predictive performance for probabilistic daily irrigation demand forecast ( $ID_{CG\ i, ASP}$  and  $ID_{CG\ i, OTR}$ ) were evaluated against observed data recorded at 1016 service points. Averaged over the two evaluation periods, the NSE values for  $ID_{CG\ i, ASP}$  across the 5 command areas were in the range 0.96 (CG 1234) - 0.63 (CG 1) for 1 day lead time and between 0.76 (CG1234) and 0.42 (CG 1) for 5 days lead time.. The predictive performance improved as the irrigation area served by the channel increased and the temporal predictive performances declined with the lead time. The rank histograms showed that the ensemble spread was slightly over dispersive for shorter lead time (1-2 days) and slightly under dispersive for longer lead time (3 to 5 days), suggesting that additional sources of error that are not accounted for accumulate over time.

This study investigated input uncertainties in detail and overlooks the model structural and parameter uncertainties. While preliminary trials showed parameter uncertainties have little influence, it would be valuable to examine ways of including model structural uncertainty. Nevertheless, the ensemble forecasting scheme adopted provides useful information on demand forecast uncertainty that would be useful in helping system operator manage the risk associated with operational decisions.

## Acknowledgements

The authors wish to express their gratitude to Goulburn-Murray Water and the Bureau of Meteorology, Australia for providing irrigation flow data and observed and forecast (from the NWP

system ACCESS-G) weather data for research purposes and Dr Mark Bailey, Mr Mick Doherty and Mr John Weber, Goulburn-Murray Water, Australia and Dr Alan Seed and Dr Shaun Cooper, Bureau of Meteorology, Australia for their advice and assistance with data, especially during data collection. Irrigation system flow data are available upon request at Goulburn-Murray Water ([reception@gmwater.com.au](mailto:reception@gmwater.com.au)) and AWS and the NWP system data are available upon request at National Climate Centre, Bureau of Meteorology ([iSupportNotification@bom.gov.au](mailto:iSupportNotification@bom.gov.au)). Kushan C Perera was supported by an Australian Post-Graduate Award from the University of Melbourne

#### **Appendix A. Supplementary Material**

The following are the supplementary data to this article:

Accepted Article

754 Table A-1 – Forecast performance for ensemble daily  $ID_{CGi,ASP}$  forecasts related to the 2 cross validation scenarios for 4 channels and study area during the calibration periods (15-08-2010-15-05-2011 (274 days) or 15-08-2011-  
755 15-05-2012 (275 days)). The range shown in brackets is the 5<sup>th</sup> - 95<sup>th</sup> from the bootstrapping

Lead time	Ensemble verification score Scenario No.   Validation year	Command Area														
		CG1			CG2			CG3			CG4			CG1234		
		1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.
1 Day	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.32 (0.3-0.33)	0.46 (0.44-0.49)	0.39 (0.38-0.41)	0.32 (0.3-0.33)	0.5 (0.48-0.52)	0.41 (0.39-0.42)	0.22 (0.21-0.23)	0.42 (0.4-0.44)	0.32 (0.31-0.33)	0.24 (0.22-0.25)	0.48 (0.44-0.51)	0.36 (0.34-0.37)	0.16 (0.15-0.17)	0.34 (0.32-0.36)	0.25 (0.24-0.26)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.17 (0.15-0.18)	0.25 (0.23-0.28)	0.21 (0.2-0.22)	0.15 (0.14-0.17)	0.22 (0.21-0.24)	0.19 (0.18-0.2)	0.09 (0.08-0.1)	0.15 (0.14-0.16)	0.12 (0.11-0.12)	0.08 (0.07-0.08)	0.14 (0.13-0.14)	0.11 (0.1-0.11)	0.06 (0.05-0.06)	0.1 (0.1-0.11)	0.08 (0.08-0.08)
	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.06 (0.05-0.07)	0.04 (0.03-0.05)	0.05 (0.04-0.06)	0.02 (0.02-0.03)	0 (0-0.01)	0.01 (0.01-0.02)	0.01 (0-0.01)	0.01 (0.01-0.01)	0.01 (0.01-0.01)	0.01 (0.01-0.01)	0.04 (0.04-0.05)	0.03 (0.02-0.03)	0 (0-0.01)	0.02 (0.02-0.02)	0.01 (0.01-0.01)
	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.3 (0.27-0.33)	0.46 (0.42-0.5)	0.38 (0.36-0.41)	0.28 (0.26-0.31)	0.4 (0.36-0.42)	0.34 (0.32-0.36)	0.16 (0.15-0.17)	0.25 (0.23-0.27)	0.2 (0.19-0.22)	0.13 (0.12-0.14)	0.17 (0.16-0.19)	0.15 (0.14-0.16)	0.09 (0.09-0.1)	0.15 (0.14-0.17)	0.12 (0.12-0.13)
	NSE (mean)	0.36 (0.21-0.45)	0.56 (0.48-0.62)	0.46 (0.37-0.52)	0.7 (0.64-0.74)	0.75 (0.71-0.79)	0.73 (0.69-0.75)	0.82 (0.78-0.85)	0.83 (0.8-0.86)	0.83 (0.8-0.85)	0.9 (0.88-0.92)	0.92 (0.9-0.93)	0.91 (0.9-0.92)	0.94 (0.92-0.95)	0.93 (0.92-0.94)	0.94 (0.93-0.94)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.13 (0.11-0.16)	0.08 (0.03-0.12)	0.1 (0.08-0.13)	0.09 (0.06-0.12)	0.02 (-0.02-0.06)	0.06 (0.03-0.08)	0.06 (0.04-0.07)	0.03 (0-0.05)	0.04 (0.03-0.06)	0.05 (0.04-0.06)	0.05 (0.04-0.07)	0.05 (0.04-0.06)	0.03 (0.02-0.04)	0.03 (0.02-0.05)	0.03 (0.02-0.04)
	3 Day	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.38 (0.36-0.4)	0.57 (0.54-0.61)	0.48 (0.46-0.5)	0.37 (0.35-0.39)	0.61 (0.59-0.64)	0.49 (0.47-0.51)	0.27 (0.26-0.28)	0.47 (0.45-0.49)	0.37 (0.36-0.38)	0.27 (0.25-0.28)	0.49 (0.46-0.51)	0.38 (0.36-0.39)	0.22 (0.21-0.23)	0.42 (0.4-0.44)
Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.21 (0.19-0.23)	0.33 (0.3-0.36)	0.27 (0.25-0.29)	0.18 (0.17-0.2)	0.32 (0.3-0.34)	0.25 (0.24-0.26)	0.13 (0.12-0.14)	0.21 (0.19-0.22)	0.17 (0.16-0.18)	0.11 (0.1-0.12)	0.18 (0.17-0.19)	0.14 (0.14-0.15)	0.09 (0.09-0.1)	0.17 (0.16-0.18)	0.13 (0.13-0.14)
Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.08 (0.06-0.1)	0.05 (0.04-0.06)	0.06 (0.06-0.08)	0.02 (0.02-0.04)	0.02 (0.01-0.02)	0.02 (0.02-0.03)	0.01 (0.01-0.02)	0 (0-0.01)	0.01 (0.01-0.01)	0.01 (0-0.01)	0.01 (0.01-0.01)	0.01 (0.01-0.01)	0.01 (0-0.01)	0 (0-0.01)	0 (0-0.01)
RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.37 (0.34-0.41)	0.58 (0.53-0.64)	0.48 (0.45-0.51)	0.33 (0.3-0.36)	0.55 (0.52-0.59)	0.44 (0.42-0.46)	0.23 (0.21-0.24)	0.37 (0.34-0.4)	0.3 (0.28-0.31)	0.2 (0.19-0.22)	0.29 (0.27-0.31)	0.25 (0.23-0.26)	0.17 (0.16-0.18)	0.3 (0.28-0.32)	0.24 (0.22-0.25)
NSE (mean)		0.05 (-0.15-0.15)	0.29 (0.24-0.33)	0.17 (0.07-0.22)	0.58 (0.52-0.63)	0.52 (0.46-0.58)	0.55 (0.51-0.59)	0.63 (0.56-0.69)	0.64 (0.59-0.69)	0.64 (0.6-0.67)	0.74 (0.7-0.78)	0.77 (0.74-0.81)	0.76 (0.73-0.78)	0.81 (0.77-0.83)	0.74 (0.69-0.78)	0.77 (0.74-0.8)
BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.14 (0.11-0.18)	0.05 (-0.01-0.11)	0.1 (0.06-0.13)	0.09 (0.06-0.12)	-0.01 (-0.07-0.05)	0.04 (0.01-0.07)	0.06 (0.04-0.08)	0.01 (-0.03-0.04)	0.03 (0.01-0.05)	0.05 (0.03-0.07)	0.05 (0.02-0.08)	0.05 (0.03-0.07)	0.05 (0.03-0.06)	0.03 (0-0.06)	0.04 (0.02-0.06)
5 Day		MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.38 (0.37-0.41)	0.58 (0.55-0.61)	0.48 (0.46-0.5)	0.38 (0.37-0.4)	0.65 (0.62-0.68)	0.52 (0.5-0.53)	0.29 (0.28-0.3)	0.49 (0.47-0.52)	0.39 (0.38-0.4)	0.28 (0.26-0.29)	0.46 (0.44-0.48)	0.37 (0.36-0.38)	0.24 (0.23-0.25)	0.46 (0.44-0.47)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.21 (0.19-0.23)	0.33 (0.3-0.36)	0.27 (0.25-0.29)	0.2 (0.18-0.21)	0.35 (0.32-0.38)	0.27 (0.26-0.29)	0.14 (0.13-0.15)	0.24 (0.22-0.26)	0.19 (0.18-0.2)	0.12 (0.11-0.13)	0.19 (0.17-0.2)	0.15 (0.15-0.16)	0.11 (0.1-0.12)	0.21 (0.19-0.23)	0.16 (0.15-0.17)
	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.08 (0.07-0.1)	0.05 (0.04-0.06)	0.07 (0.06-0.08)	0.03 (0.02-0.04)	0.02 (0.02-0.03)	0.02 (0.02-0.03)	0.02 (0.01-0.02)	0.01 (0.01-0.01)	0.01 (0.01-0.02)	0.01 (0-0.01)	0 (0-0.01)	0 (0-0.01)	0.01 (0.01-0.01)	0 (0-0.01)	0.01 (0-0.01)
	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.38 (0.34-0.41)	0.59 (0.53-0.63)	0.48 (0.45-0.51)	0.36 (0.32-0.39)	0.6 (0.56-0.64)	0.48 (0.45-0.5)	0.26 (0.24-0.27)	0.43 (0.4-0.47)	0.34 (0.32-0.36)	0.23 (0.21-0.25)	0.32 (0.3-0.35)	0.28 (0.26-0.29)	0.2 (0.19-0.21)	0.37 (0.34-0.4)	0.29 (0.27-0.3)
	NSE (mean)	0.02 (-0.15-0.13)	0.28 (0.23-0.32)	0.15 (0.06-0.21)	0.53 (0.46-0.57)	0.43 (0.37-0.49)	0.48 (0.44-0.51)	0.54 (0.47-0.6)	0.51 (0.44-0.57)	0.52 (0.48-0.56)	0.66 (0.62-0.71)	0.72 (0.67-0.76)	0.69 (0.66-0.72)	0.73 (0.69-0.76)	0.61 (0.55-0.67)	0.67 (0.63-0.71)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.14 (0.11-0.18)	0.05 (-0.01-0.11)	0.1 (0.06-0.13)	0.08 (0.05-0.12)	-0.03 (-0.09-0.03)	0.03 (-0.01-0.06)	0.05 (0.03-0.08)	-0.01 (-0.05-0.04)	0.02 (0-0.05)	0.05 (0.02-0.07)	0.03 (0-0.06)	0.04 (0.02-0.06)	0.05 (0.03-0.07)	0.01 (-0.02-0.05)	0.03 (0.01-0.05)

Table A-2 – Forecast performance for ensemble daily  $ID_{CGi,ASP}$  forecasts related to the 2 cross evaluation scenarios for 4 channels and study area during the evaluation periods ( 15-08-2010-15-05-2011 (274 days) or 15-08-2011-15-05-2012 (275 days)). The range shown in brackets is the 5<sup>th</sup>-95<sup>th</sup> from the bootstrapping

Lead time	Ensemble verification score Scenario No.   Validation year	Command Area														
		CG1			CG2			CG3			CG4			CG1234		
		1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.	1   2011/12	2   2010/11	Avg. Perf.
1 Day	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.36 (0.34-0.37)	0.34 (0.33-0.35)	0.35 (0.34-0.36)	0.44 (0.42-0.46)	0.34 (0.33-0.35)	0.39 (0.38-0.4)	0.39 (0.36-0.42)	0.23 (0.22-0.24)	0.31 (0.3-0.33)	0.46 (0.42-0.49)	0.25 (0.23-0.27)	0.35 (0.33-0.37)	0.3 (0.28-0.32)	0.17 (0.16-0.18)	0.23 (0.22-0.25)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.18 (0.16-0.19)	0.16 (0.15-0.16)	0.17 (0.16-0.17)	0.16 (0.15-0.17)	0.14 (0.13-0.14)	0.15 (0.14-0.16)	0.12 (0.11-0.13)	0.08 (0.08-0.09)	0.1 (0.1-0.11)	0.13 (0.12-0.14)	0.08 (0.07-0.08)	0.1 (0.1-0.11)	0.09 (0.08-0.09)	0.05 (0.05-0.06)	0.07 (0.07-0.07)
	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.01 (0.01-0.01)	0.06 (0.04-0.07)	0.03 (0.03-0.04)	0.02 (0.01-0.03)	0.01 (0.01-0.02)	0.02 (0.01-0.02)	0.03 (0.02-0.03)	0.01 (0.01-0.01)	0.02 (0.02-0.02)	0.06 (0.05-0.07)	0.02 (0.01-0.02)	0.04 (0.03-0.04)	0.03 (0.03-0.03)	0.01 (0.01-0.01)	0.02 (0.02-0.02)
	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.33 (0.3-0.36)	0.27 (0.26-0.29)	0.3 (0.28-0.31)	0.3 (0.28-0.32)	0.24 (0.23-0.26)	0.27 (0.26-0.28)	0.17 (0.16-0.18)	0.14 (0.13-0.14)	0.15 (0.14-0.16)	0.11 (0.1-0.12)	0.1 (0.1-0.11)	0.11 (0.1-0.11)	0.1 (0.09-0.11)	0.08 (0.07-0.08)	0.09 (0.08-0.09)
	NSE (mean)	0.77 (0.74-0.8)	0.49 (0.33-0.59)	0.63 (0.55-0.68)	0.86 (0.84-0.87)	0.78 (0.72-0.81)	0.82 (0.79-0.83)	0.92 (0.91-0.94)	0.87 (0.84-0.89)	0.9 (0.88-0.91)	0.97 (0.96-0.97)	0.93 (0.92-0.94)	0.95 (0.94-0.96)	0.97 (0.97-0.98)	0.96 (0.95-0.97)	0.96 (0.96-0.97)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.02 (-0.01-0.05)	0.17 (0.15-0.19)	0.1 (0.08-0.11)	-0.09 (-0.11--0.06)	0.11 (0.09-0.13)	0.01 (0-0.03)	-0.03 (-0.05--0.02)	0.07 (0.06-0.08)	0.02 (0.01-0.03)	0 (-0.01-0.01)	0.05 (0.04-0.05)	0.02 (0.02-0.03)	-0.02 (-0.03--0.01)	0.03 (0.03-0.04)	0.01 (0-0.02)
	3 Day	MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.41 (0.39-0.44)	0.4 (0.39-0.42)	0.41 (0.4-0.42)	0.5 (0.47-0.53)	0.41 (0.4-0.43)	0.46 (0.44-0.47)	0.4 (0.38-0.43)	0.3 (0.29-0.31)	0.35 (0.34-0.36)	0.43 (0.41-0.46)	0.29 (0.27-0.3)	0.36 (0.35-0.37)	0.33 (0.32-0.35)	0.24 (0.23-0.25)
Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.22 (0.2-0.24)	0.19 (0.18-0.2)	0.2 (0.19-0.21)	0.24 (0.22-0.26)	0.18 (0.17-0.19)	0.21 (0.2-0.22)	0.15 (0.14-0.16)	0.12 (0.11-0.12)	0.13 (0.13-0.14)	0.14 (0.14-0.15)	0.1 (0.1-0.11)	0.12 (0.12-0.13)	0.12 (0.12-0.13)	0.09 (0.08-0.09)	0.11 (0.1-0.11)
Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.01 (0.01-0.02)	0.07 (0.06-0.09)	0.04 (0.04-0.05)	0.05 (0.03-0.07)	0.02 (0.01-0.03)	0.04 (0.03-0.05)	0.03 (0.02-0.04)	0.01 (0.01-0.02)	0.02 (0.01-0.02)	0.03 (0.03-0.04)	0.01 (0.01-0.01)	0.02 (0.02-0.02)	0.02 (0.01-0.03)	0.01 (0-0.01)	0.01 (0.01-0.02)
RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )		0.41 (0.37-0.45)	0.33 (0.32-0.35)	0.37 (0.35-0.39)	0.45 (0.42-0.48)	0.32 (0.31-0.34)	0.39 (0.37-0.4)	0.29 (0.26-0.31)	0.21 (0.2-0.21)	0.25 (0.23-0.26)	0.22 (0.21-0.24)	0.17 (0.16-0.18)	0.2 (0.19-0.21)	0.21 (0.2-0.23)	0.15 (0.14-0.16)	0.18 (0.18-0.19)
NSE (mean)		0.65 (0.63-0.66)	0.24 (0-0.37)	0.44 (0.32-0.51)	0.68 (0.65-0.71)	0.61 (0.53-0.65)	0.64 (0.6-0.67)	0.78 (0.76-0.81)	0.7 (0.64-0.74)	0.74 (0.71-0.77)	0.87 (0.85-0.88)	0.81 (0.77-0.83)	0.84 (0.82-0.85)	0.87 (0.85-0.89)	0.84 (0.8-0.87)	0.86 (0.84-0.87)
BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )		-0.02 (-0.06-0.02)	0.2 (0.18-0.23)	0.09 (0.07-0.12)	-0.2 (-0.24--0.16)	0.12 (0.09-0.15)	-0.04 (-0.06--0.01)	-0.12 (-0.14--0.09)	0.09 (0.07-0.1)	-0.02 (-0.03-0)	-0.08 (-0.1--0.06)	0.05 (0.03-0.07)	-0.01 (-0.03-0)	-0.09 (-0.1--0.07)	0.06 (0.04-0.07)	-0.01 (-0.03-0)
5 Day		MRMSE (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.42 (0.39-0.44)	0.41 (0.4-0.42)	0.41 (0.4-0.43)	0.53 (0.49-0.56)	0.43 (0.42-0.45)	0.48 (0.46-0.5)	0.42 (0.39-0.44)	0.32 (0.31-0.33)	0.37 (0.35-0.38)	0.44 (0.42-0.47)	0.3 (0.29-0.32)	0.37 (0.36-0.39)	0.36 (0.34-0.38)	0.27 (0.26-0.28)
	Mean CRPS (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.23 (0.21-0.25)	0.19 (0.18-0.2)	0.21 (0.2-0.22)	0.27 (0.25-0.29)	0.19 (0.18-0.2)	0.23 (0.22-0.24)	0.18 (0.16-0.19)	0.13 (0.13-0.14)	0.15 (0.15-0.16)	0.17 (0.16-0.18)	0.12 (0.11-0.12)	0.14 (0.14-0.15)	0.16 (0.15-0.17)	0.11 (0.1-0.11)	0.13 (0.13-0.14)
	Mean CRPS Rel. (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.02 (0.01-0.02)	0.08 (0.06-0.09)	0.05 (0.04-0.06)	0.07 (0.05-0.09)	0.02 (0.02-0.04)	0.05 (0.04-0.06)	0.04 (0.03-0.05)	0.01 (0.01-0.02)	0.03 (0.02-0.03)	0.05 (0.04-0.06)	0 (0-0.01)	0.03 (0.02-0.03)	0.03 (0.03-0.04)	0.01 (0-0.01)	0.02 (0.02-0.03)
	RMSE (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	0.42 (0.38-0.46)	0.34 (0.32-0.35)	0.38 (0.36-0.4)	0.51 (0.47-0.54)	0.35 (0.33-0.37)	0.43 (0.41-0.45)	0.34 (0.32-0.37)	0.24 (0.23-0.24)	0.29 (0.28-0.3)	0.29 (0.27-0.3)	0.21 (0.2-0.22)	0.25 (0.24-0.26)	0.28 (0.26-0.3)	0.19 (0.18-0.2)	0.24 (0.23-0.25)
	NSE (mean)	0.63 (0.61-0.65)	0.22 (-0.01-0.35)	0.42 (0.31-0.49)	0.6 (0.55-0.63)	0.54 (0.46-0.59)	0.57 (0.52-0.6)	0.69 (0.66-0.72)	0.61 (0.54-0.66)	0.65 (0.61-0.68)	0.78 (0.75-0.8)	0.72 (0.68-0.76)	0.75 (0.73-0.77)	0.78 (0.75-0.8)	0.75 (0.7-0.79)	0.76 (0.74-0.79)
	BIAS (mean) (ML day <sup>-1</sup> Km <sup>-1</sup> )	-0.02 (-0.06-0.02)	0.2 (0.17-0.23)	0.09 (0.06-0.11)	-0.26 (-0.3--0.21)	0.12 (0.09-0.15)	-0.07 (-0.09--0.04)	-0.16 (-0.19--0.13)	0.08 (0.06-0.11)	-0.04 (-0.06--0.02)	-0.14 (-0.17--0.11)	0.04 (0.01-0.05)	-0.05 (-0.07--0.04)	-0.14 (-0.17--0.12)	0.06 (0.04-0.08)	-0.04 (-0.05--0.02)

## References

- 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(7): 2327-2347. DOI:10.5194/hess-15-2327-2011
- Alfonso, F.T.R., Andres, M.T., Walker, W.R., Mac McKee, 2011. Bayesian data-driven models for irrigation water management, Utah State University, All Graduate Theses and Dissertations.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome, 300(9).
- Alvarez-Garreton, C., Ryu, D., Western, A., Crow, W., Robertson, D., 2014. The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *J. Hydrol.*
- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends in ecology & evolution*, 22(1): 42-47.
- Azhar, A.H., Perera, B.J.C., 2011. Prediction of rainfall for short term irrigation planning and scheduling - Case study in Victoria, Australia. *J. Irrig. Drainage Eng-ASCE*, 137(7): 435-445. DOI:10.1061/(asce)ir.1943-4774.0000317
- Barancourt, C., Creutin, J.D., Rivoirard, J., 1992. A method for delineating and estimating rainfall fields. *Water Resour. Res.*, 28(4): 1133-1144.
- Bennett, J.C., Robertson, D.E., Shrestha, D.L., Wang, Q.J., Enever, D., Hapuarachchi, P., Tuteja, N.K., 2014. A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days. *J. Hydrol.*, 519, Part D: 2832-2846. DOI:<http://dx.doi.org/10.1016/j.jhydrol.2014.08.010>
- BoM, A., 2005. Automatic Weather Stations for Agricultural and Other Applications. The Bureau of Meteorology, Australia., [http://www.bom.gov.au/inside/services\\_policy/pub\\_ag/aws/aws.shtml](http://www.bom.gov.au/inside/services_policy/pub_ag/aws/aws.shtml).
- BoM, A., 2010. Operational implementation of the ACCESS Numerical Weather Prediction systems, NMOC operations bulletin 83.

- BoM, A., 2012. APS1 upgrade of the ACCESS-G Numerical Weather Prediction system NMOC Operations Bulletin. The Bureau of Meteorology, Australia.
- Brown, J.D., Demargne, J., Seo, D.-J., Liu, Y., 2010. The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling & Software*, 25(7): 854-872.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16(4): 1145-1157.
- Cai, X., Hejazi, M.I., Wang, D., 2011. Value of Probabilistic Weather Forecasts: Assessment by Real-Time Optimization of Irrigation Scheduling. *Journal of Water Resources Planning & Management*, 137(5): 391-403. DOI:10.1061/(asce)wr.1943-5452.0000126
- Ciach, G.J., Krajewski, W.F., 1999. On the estimation of radar rainfall error variance. *Adv. Water Resour.*, 22(6): 585-595. DOI:[http://dx.doi.org/10.1016/S0309-1708\(98\)00043-8](http://dx.doi.org/10.1016/S0309-1708(98)00043-8)
- Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5(1): 243-262.
- Cloke, H., Pappenberger, F., 2009. Ensemble flood forecasting: a review. *J. Hydrol.*, 375(3): 613-626.
- DeChant, C.M., Moradkhani, H., 2012. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resour. Res.*, 48(4): W04518. DOI:10.1029/2011WR011011
- Ebert, E.E., 2001. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Monthly Weather Review*, 129(10): 2461-2480.
- Ebert, E.E., Turk, M., Kusselson, S.J., Yang, J., Seybold, M., Keehn, P.R., Kuligowski, R.J., 2011. Ensemble tropical rainfall potential (eTRaP) forecasts. *Weather and Forecasting*, 26(2): 213-224.
- Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*: 54-75.

- Gneiting, T., 2013. Probabilistic Weather Forecasting, 59th ISI World Statistics Congress, Hong Kong.
- Gneiting, T., Raftery, A.E., 2005. Weather forecasting with ensemble methods. *Science*, 310(5746): 248-249.
- Gowing, J.W., Ejeji, C.J., 2001. Real-time scheduling of supplemental irrigation for potatoes using a decision model and short-term weather forecasts. *Agric. Water Manage.*, 47(2): 137-153. DOI:10.1016/s0378-3774(00)00101-3
- Grenouillet, G., Buisson, L., Casajus, N., Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, 34(1): 9-17.
- Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3): 550-560.
- Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5): 559-570.
- Hutton, C.J., Kapelan, Z., 2015. A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting. *Environmental Modelling & Software*, 66: 87-97.
- Khaliq, M., Ouarda, T.B., Gachon, P., Sushama, L., St-Hilaire, A., 2009. Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. *J. Hydrol.*, 368(1): 117-130.
- Li, Y., Ryu, D., Western, A.W., Wang, Q., 2013. Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags. *Water Resour. Res.*, 49(4): 1887-1900.
- Li, Y., Ryu, D., Western, A.W., Wang, Q., 2015. Assimilation of stream discharge for flood forecasting: Updating a semidistributed model with an integrated data assimilation scheme. *Water Resour. Res.*
- Li, Y., Ryu, D., Western, A.W., Wang, Q.J., Robertson, D.E., Crow, W.T., 2014. An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood

NVIRP, 2010a. Business Case for Northern Victoria Irrigation Renewal Project Stage 1.

NVIRP, 2010b. Connections program, Information for farmers.

Pagano, T., Lan, C., Wang, Q., 2010. NWP-forced short-term streamflow forecasts for Australia, Third International Conference on QPE/QPF and Hydrology. World Weather Research Programme, Nanjing, China, pp. 4 pp.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology & Earth System Sciences*, 11(5): 1633-1644.

Perera, K.C., Western, A.W., George, B., Nawarathna, B., 2015a. Multivariate time series modelling of short-term system scale irrigation demand. *J. Hydrol.*

Perera, K.C., Western, A.W., Nawarathna, B., George, B., 2014. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. *Agricultural and Forest Meteorology*, 194(0): 50-63.  
DOI:<http://dx.doi.org/10.1016/j.agrformet.2014.03.014>

Perera, K.C., Western, A.W., Nawarathna, B., George, B., 2015b. Comparison of hourly and daily reference crop evapotranspiration equations across seasons and climate zones in Australia. *Agric. Water Manage.*, 148: 84-96.

Pulido-Calvo, I., Gutierrez-Estrada, J.C., 2009. Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosyst. Eng.*, 102(2): 202-218.  
DOI:10.1016/j.biosystemseng.2008.09.032

Pulido-Calvo, I., Montesinos, P., Roldán, J., Ruiz-Navarro, F., 2007. Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems. *Biosyst. Eng.*, 97(2): 283-293. DOI:10.1016/j.biosystemseng.2007.03.003

Pulido-Calvo, I., Roldán, J., López-Luque, R., Gutiérrez-Estrada, J.C., 2003. Demand Forecasting for Irrigation Water Distribution Systems. *Journal of Irrigation & Drainage Engineering*, 129(6): 422-431. DOI:10.1061/(asce)0733-9437(2003)129:6(422)

Puri, K., Dietachmayer, G., Steinle, P., Dix, M., Rikus, L., Logan, L., Naughton, M., Tingwell, C., Xiao, Y., Barras, V., 2013. Implementation of the initial ACCESS numerical weather prediction system. *Aust. Meteorol. Oceanogr. J.*, 63: 265-284.

RDV, A., undated. Victorias Food Bowl. In: Victoria, R.D. (Ed.).

Robertson, D., Shrestha, D., Wang, Q., 2013a. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences*, 17(9): 3587-3603.

Robertson, D.E., Shrestha, D.L., Wang, Q.J., 2013b. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrol. Earth Syst. Sci.*, 17(9): 3587-3603. DOI:10.5194/hess-17-3587-2013

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C., Krahe, P., 2011. The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems. *Atmos. Res.*, 100(2): 150-167.

Rubicon\_Water, 2014. FlumeGate data sheet In: Water, R. (Ed.), [https://www.rubiconwater.com/modules/prodcatalogue/files/28601/719/719\\_Rubicon%20Data%20Sheet%20FlumeGate%20\(English%20US\).pdf](https://www.rubiconwater.com/modules/prodcatalogue/files/28601/719/719_Rubicon%20Data%20Sheet%20FlumeGate%20(English%20US).pdf).

Rummel, C., Müller, M., Baier, G., Amor, F., Schindler, K., 2010. Analyzing spatio-temporal patterns of genuine cross-correlations. *Journal of Neuroscience Methods*, 191(1): 94-100. DOI:<http://dx.doi.org/10.1016/j.jneumeth.2010.05.022>

Schaake, J., 2006. Hydrologic Ensemble Prediction: Past, Present and Opportunities for the Future. *Ensemble Predictions and Uncertainties in Flood Forecasting*: 15.

Schaake, J., Restrepo, P., Seo, D., Demargne, J., Wu, L., Perica, S., 2005. Hydrologic ensemble prediction: challenges and opportunities, *International Conference on Innovative Advances and Implementation of Flood Forecasting Technology*,(October), Tromso, Norway.

Shrestha, D., Robertson, D., Wang, Q., Pagano, T., Hapuarachchi, H., 2013a. Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose. *Hydrology and Earth System Sciences*, 17(5): 1913-1931.

Shrestha, D.L., Robertson, D., Wang, Q., Pagano, T., Hapuarachchi, P., 2012. Evaluation of numerical weather prediction model rainfall forecasts for streamflow forecasting,

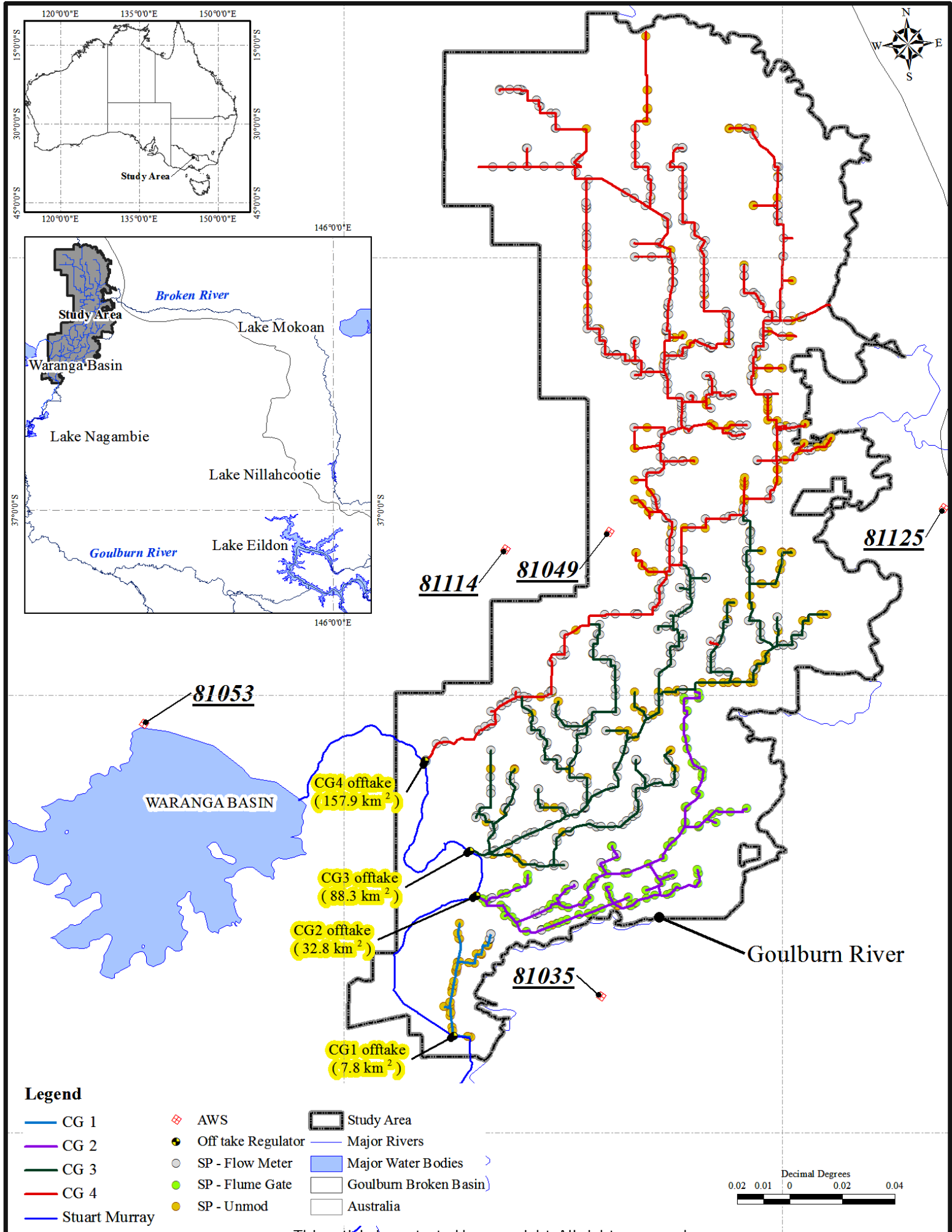
- Hydrology and Water Resources Symposium 2012. Engineers Australia, Sydney, pp. 856-863.
- Shrestha, D.L., Robertson, D., Wang, Q., Pagano, T.C., Hapuarachchi, P., 2013b. Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose. *Hydrology and Earth System Sciences*, 17: 1913-1931.
- Smiatek, G., Kunstmann, H., Werhahn, J., 2012. Implementation and performance analysis of a high resolution coupled numerical weather and river runoff prediction model system for an Alpine catchment. *Environmental modelling & software*, 38: 231-243.
- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD—a platform for ensemble forecasting of species distributions. *Ecography*, 32(3): 369-373.
- Tian, D., Martinez, C.J., 2014. The GEFS-based daily reference evapotranspiration forecast and its implication for water management in the southeastern United States. *Journal of Hydrometeorology*, 15(3): 1152-1165.
- Ticlavilca, A., McKee, M., Walker, W., 2011. Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model. *Irrig. Sci.*: 1-17. DOI:10.1007/s00271-011-0300-6
- Toth, Z., Kalnay, E., 1993. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteorol. Soc.*, 74(12): 2317-2330.
- Villarini, G., Mandapaka, P.V., Krajewski, W.F., Moore, R.J., 2008. Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research: Atmospheres*, 113(D11): n/a-n/a. DOI:10.1029/2007JD009214
- Wang, D.B., Cai, X.M., 2009. Irrigation Scheduling-Role of Weather Forecasting and Farmers' Behavior. *J. Water Resour. Plan. Manage.-ASCE*, 135(5): 364-372. DOI:10.1061/(asce)0733-9496(2009)135:5(364)
- Wang, Q.J., Robertson, D.E., Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resour. Res.*, 45(5): W05407. DOI:10.1029/2008wr007355

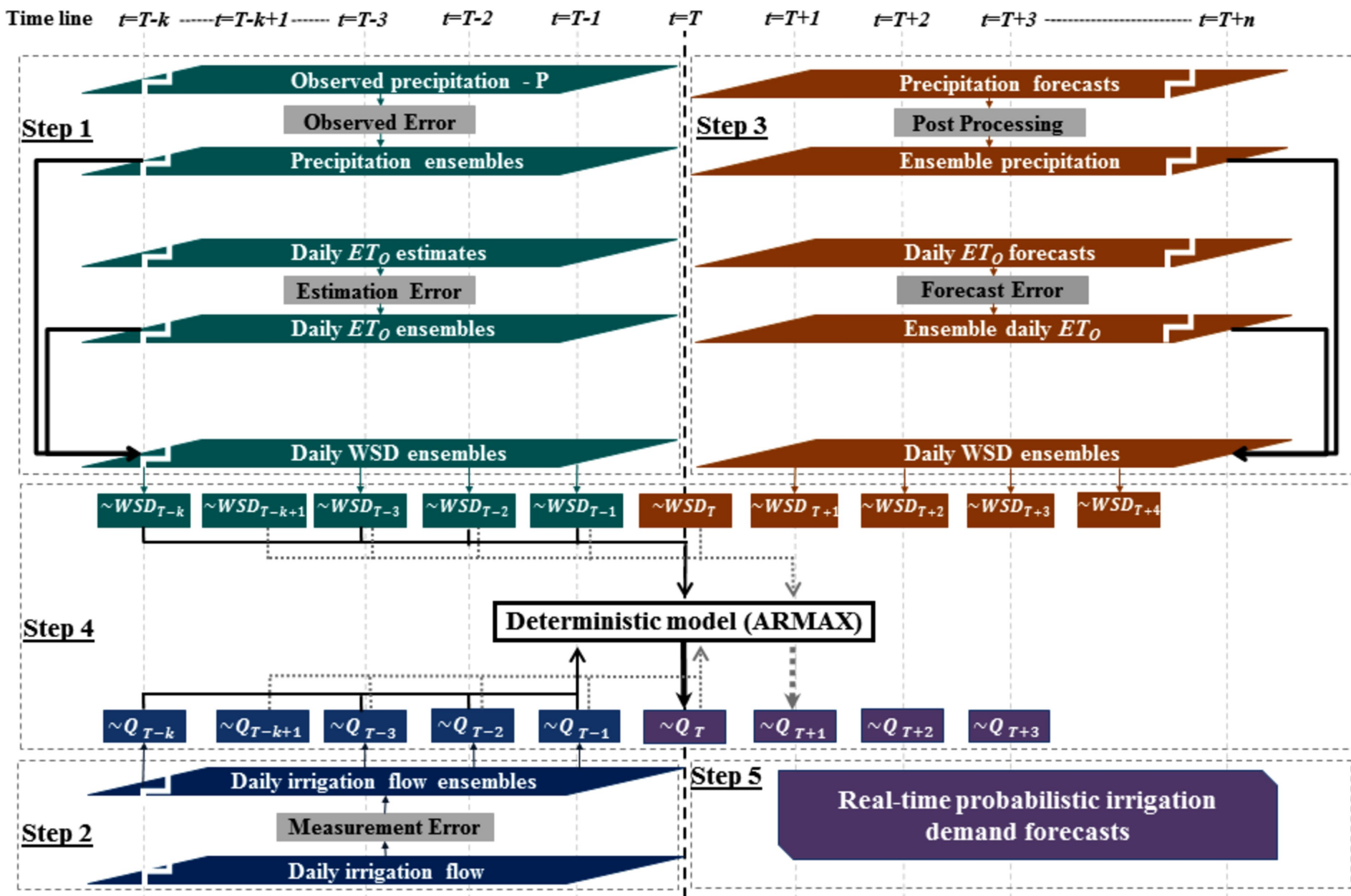
Wilks, D.S., Wolfe, D.W., 1998. Optimal use and economic value of weather forecasts for lettuce irrigation in a humid climate. *Agricultural and Forest Meteorology*, 89(2): 115-129. DOI:10.1016/s0168-1923(97)00066-x

Zaman, A.M., Etchells, T.M., Malano, H.M., Davidson, B., 2007. Towards the Next Generation of Rural Water Demand Modelling. *Modsim 2007: International Congress on Modelling and Simulation: Land, Water and Environmental Management: Integrated Systems for Sustainability*. Univ Western Australia, Nedlands, 2555-2561 pp.

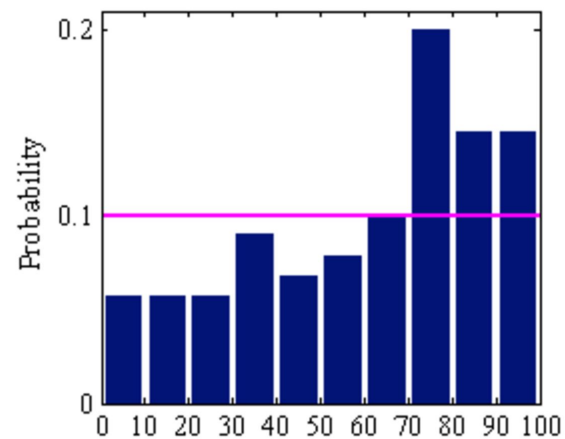
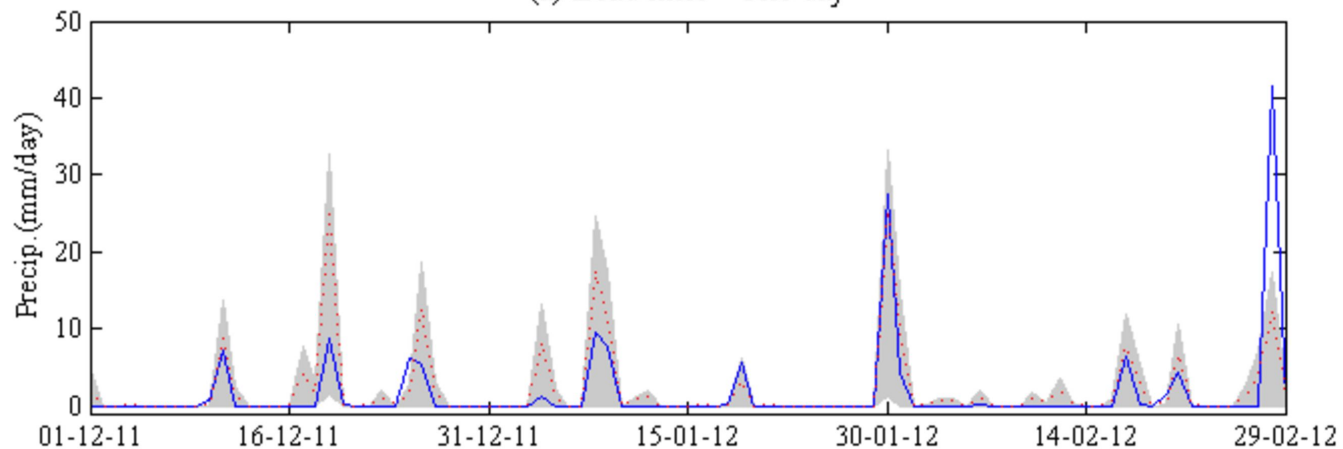
Zappa, M., Jaun, S., Germann, U., Walser, A., Fundel, F., 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmos. Res.*, 100(2): 246-262.

Accepted Article

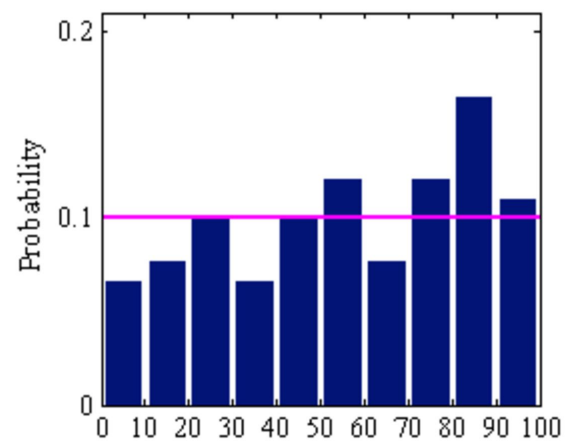
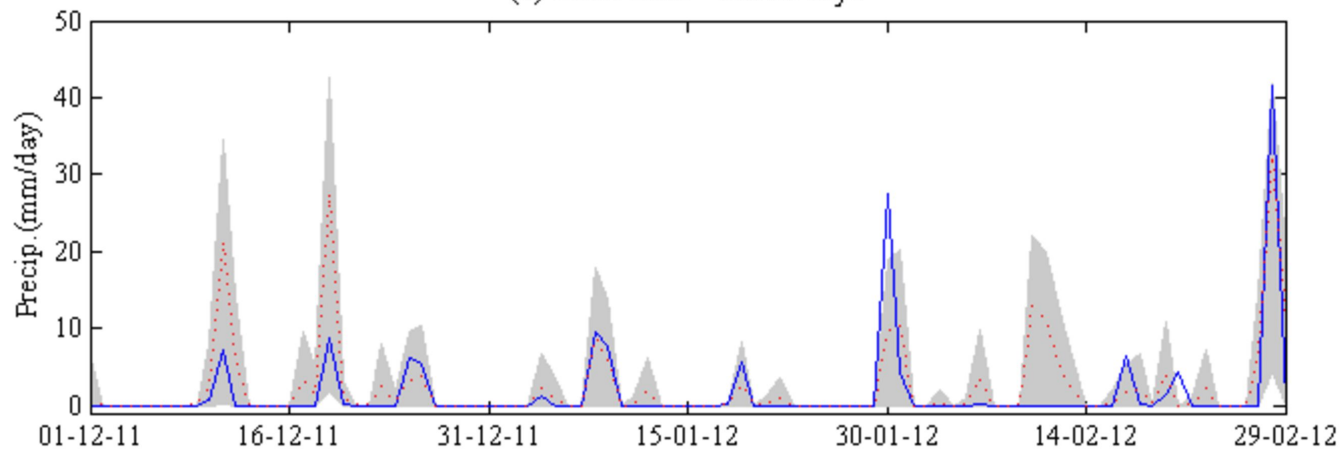




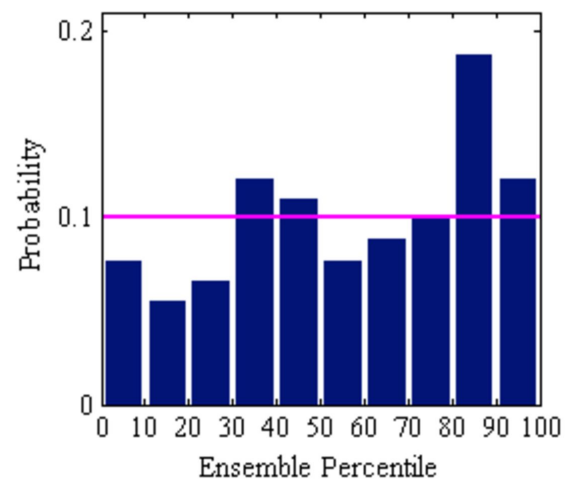
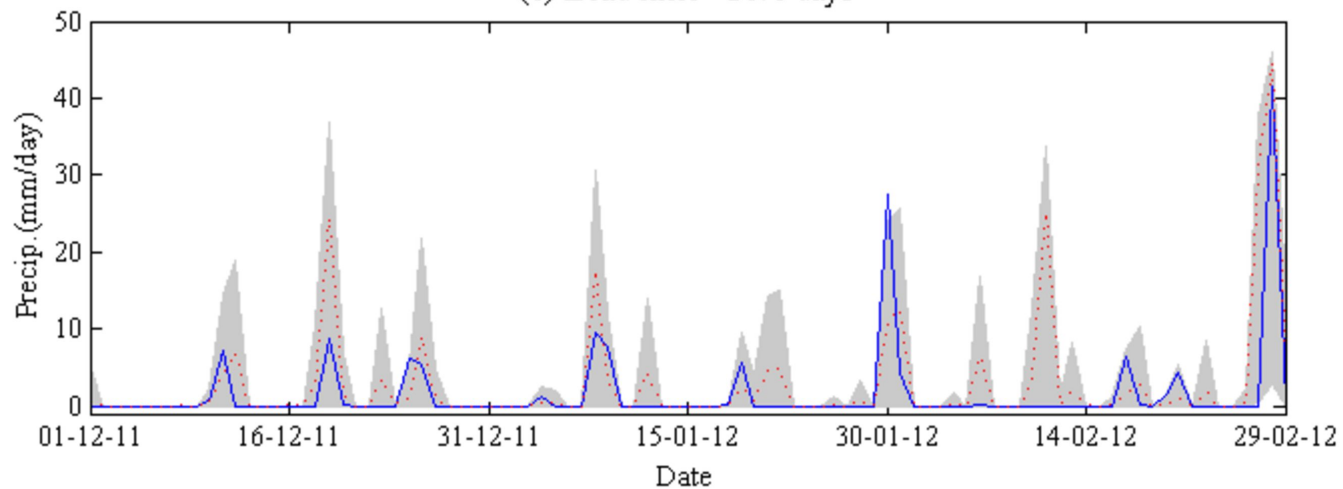
(a) Lead time - One day

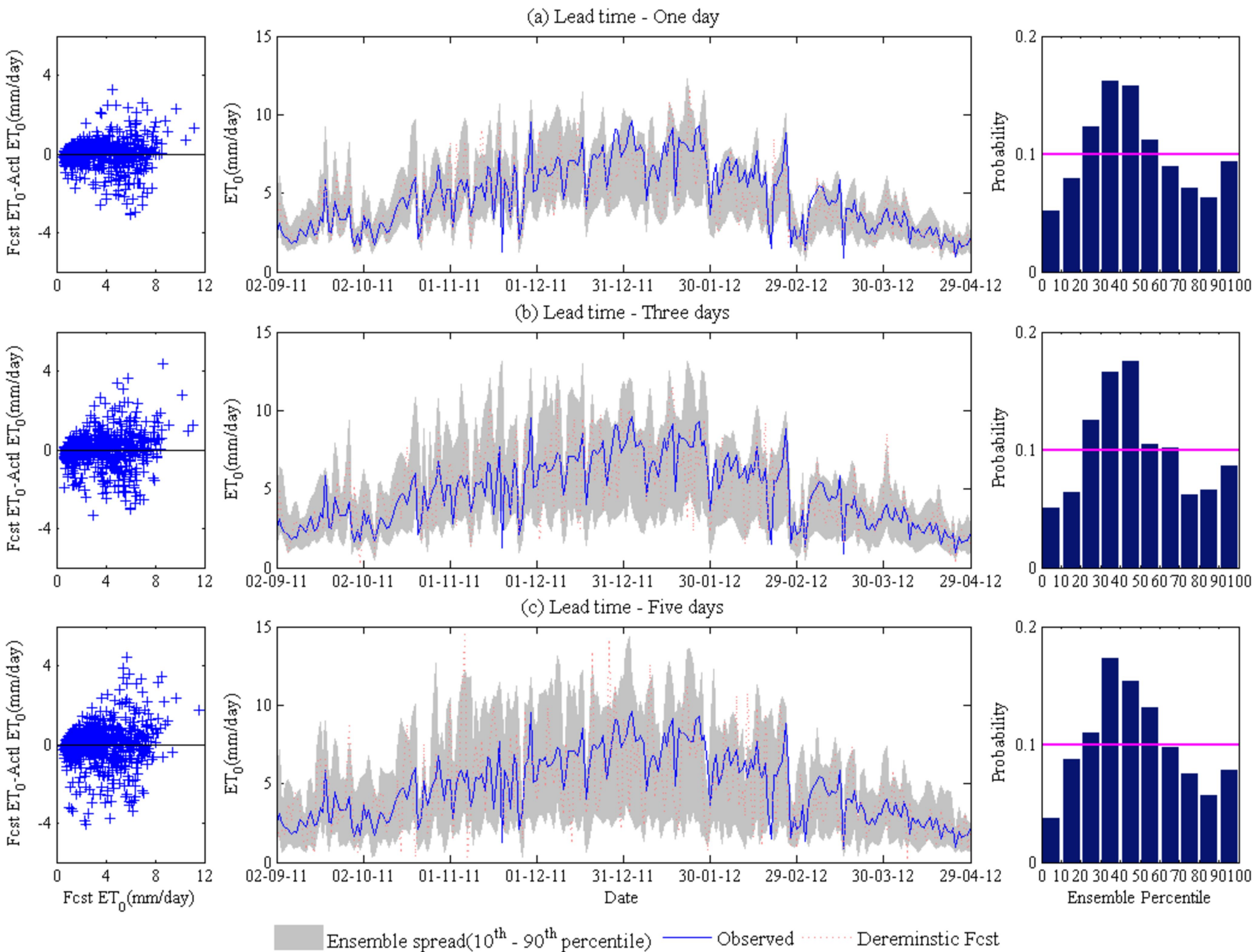


(b) Lead time - Three days

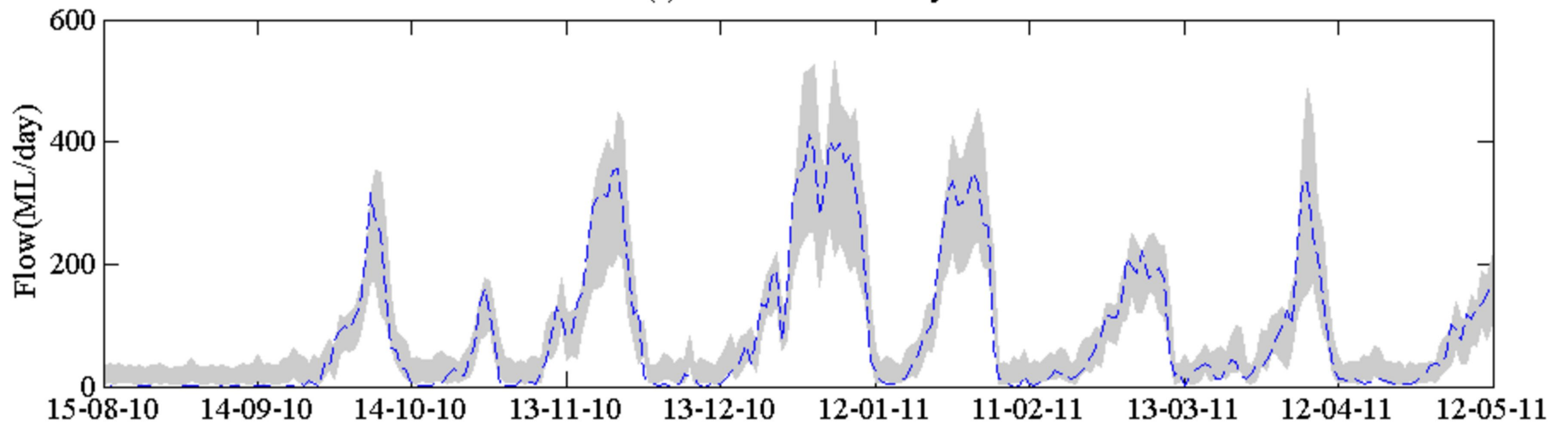


(c) Lead time - Five days

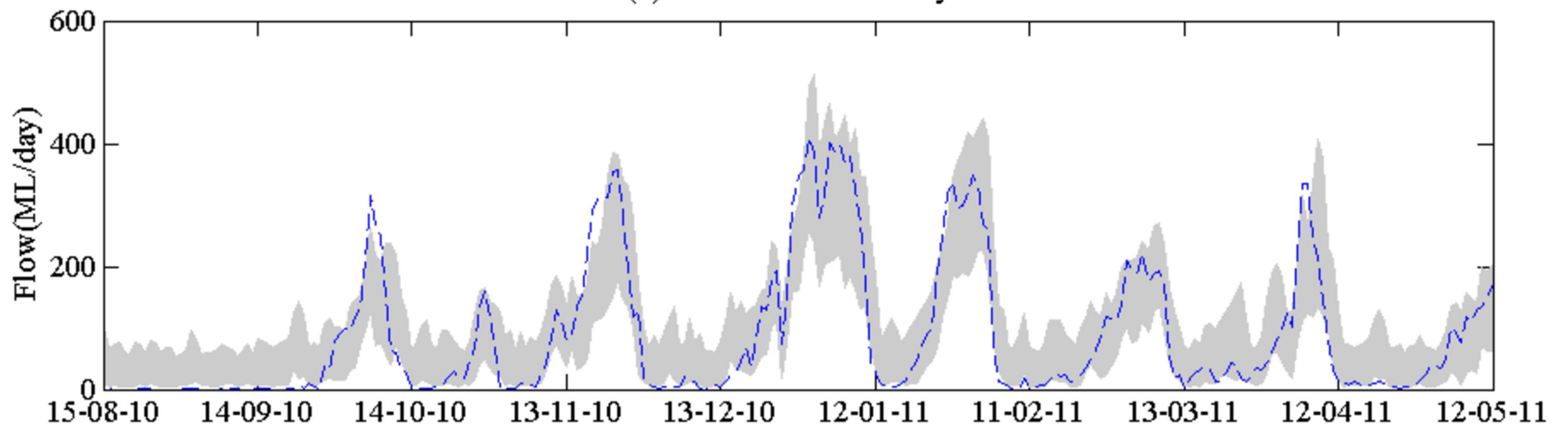




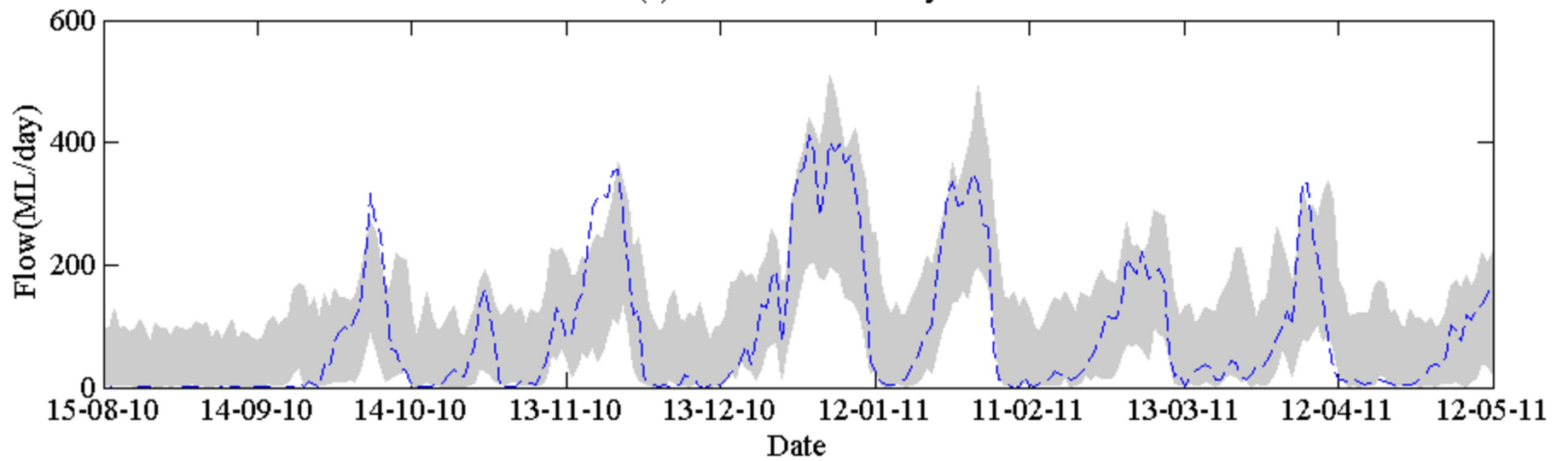
(a) Lead time - One day



(b) Lead time - Three days

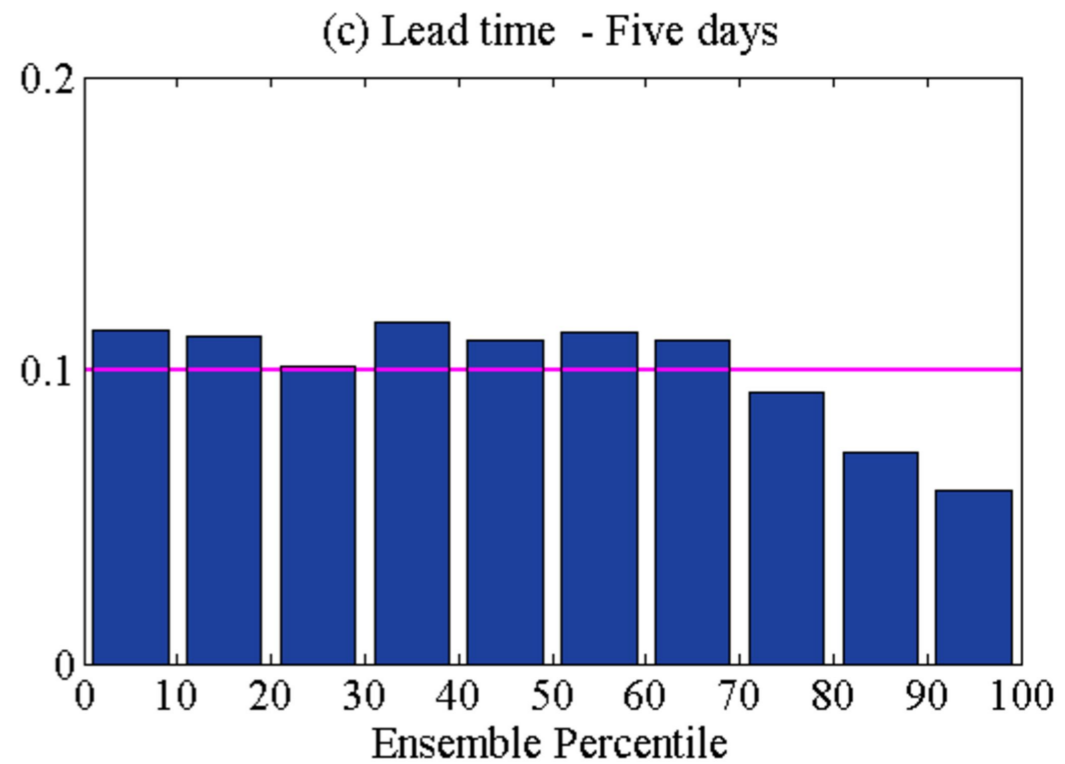
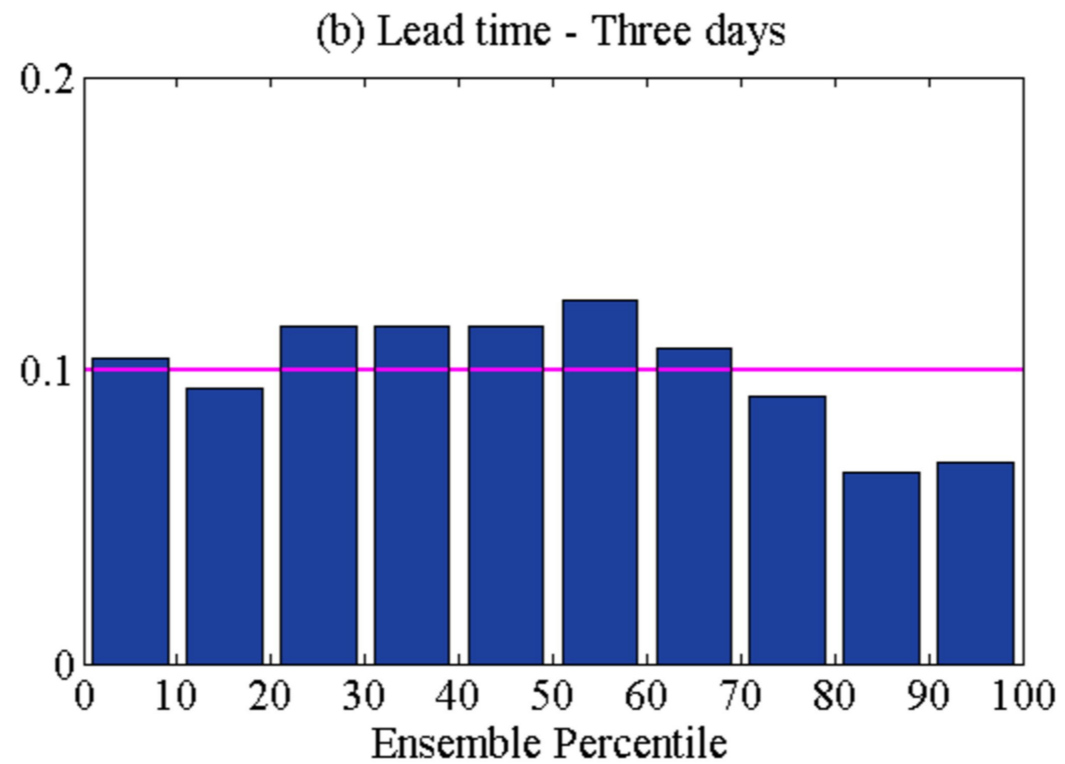
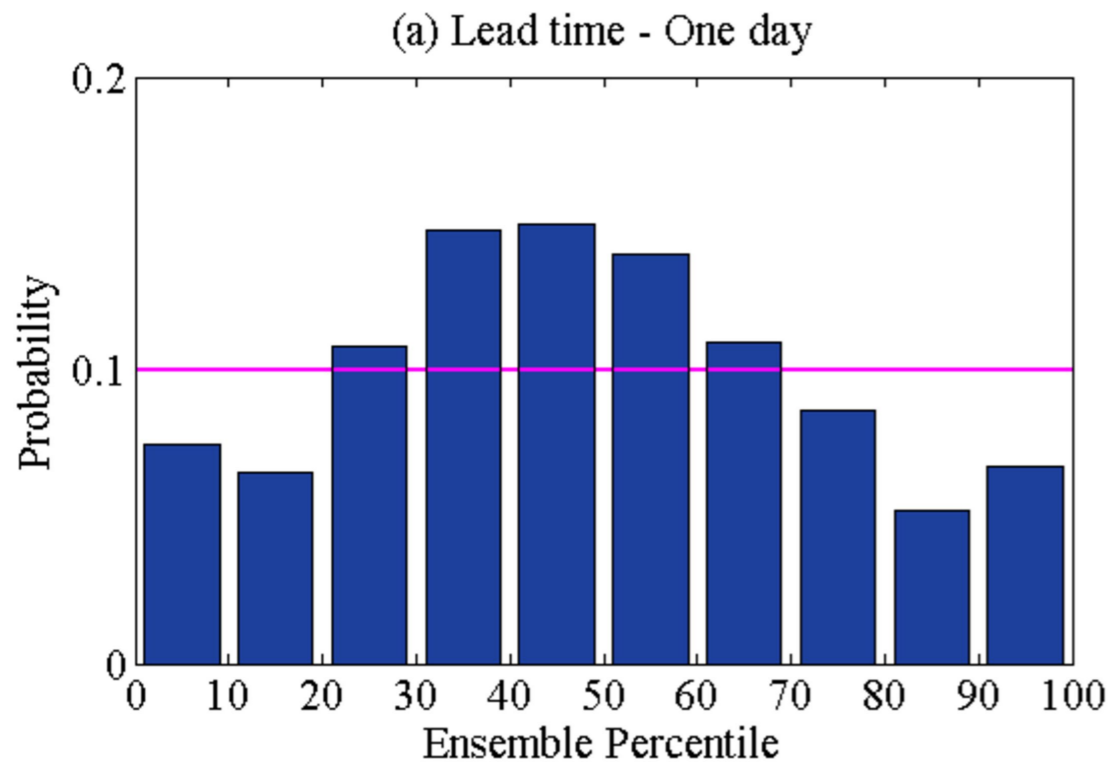


(c) Lead time - Five days

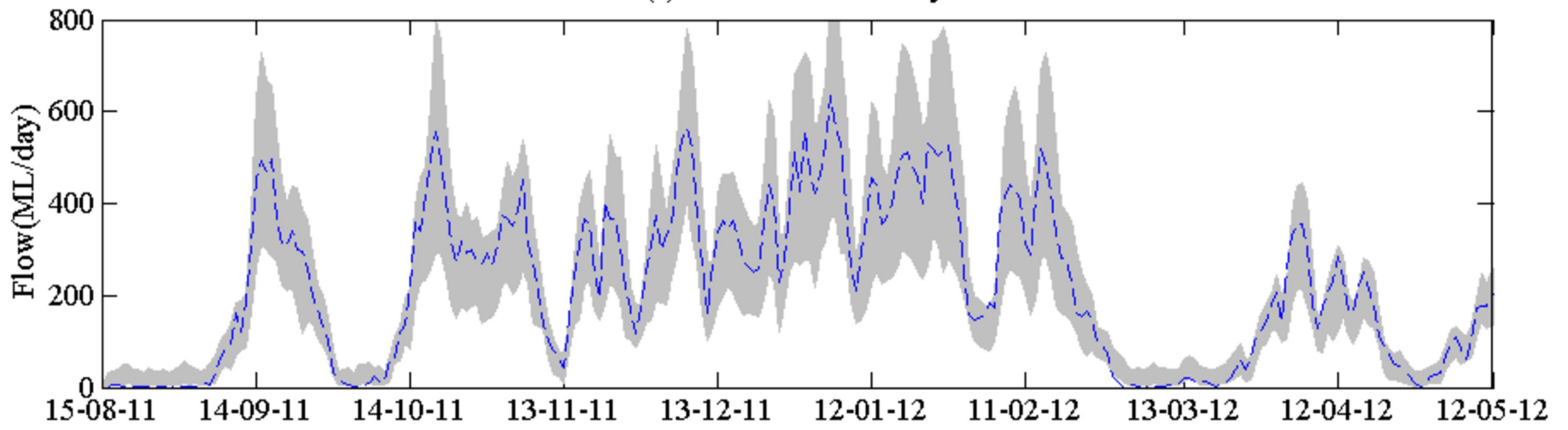


Ensemble spread (10<sup>th</sup> - 90<sup>th</sup> percentile) — — Observed flow

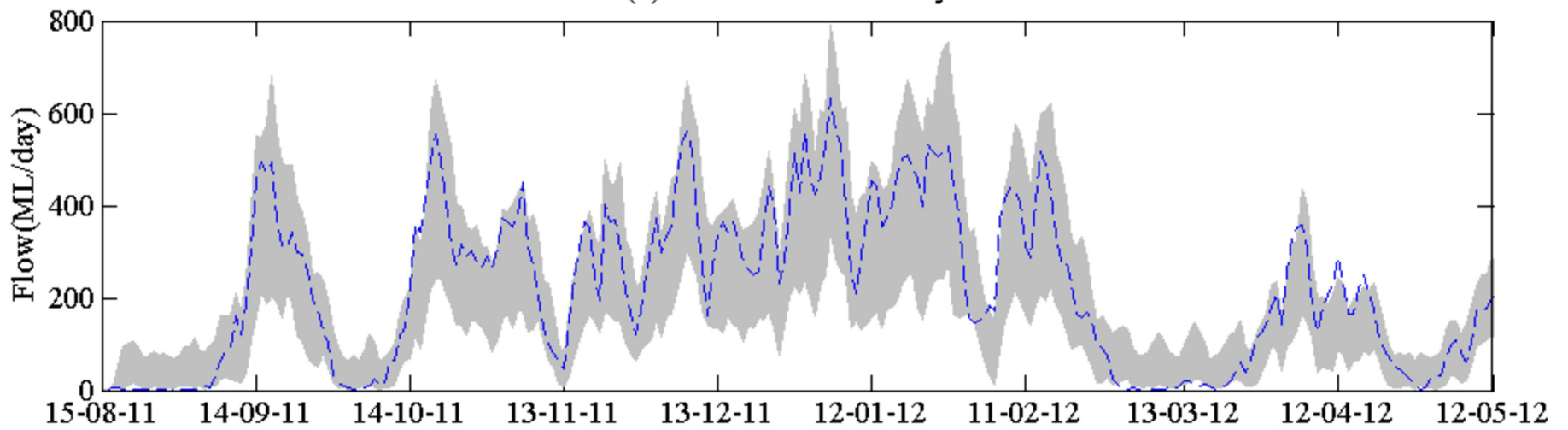
This article is protected by copyright. All rights reserved.



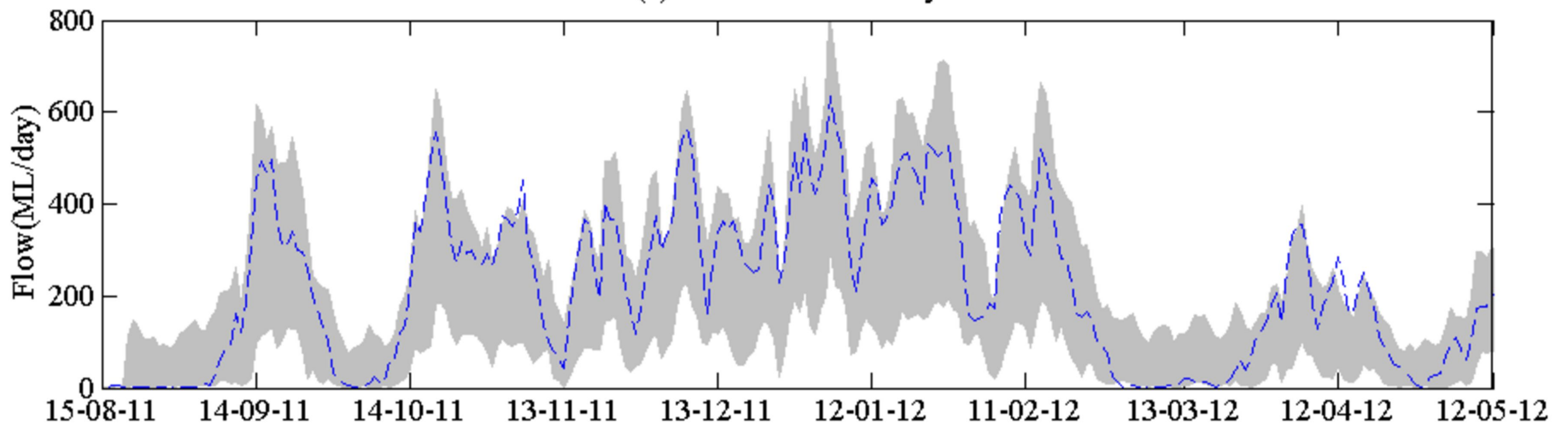
(a) Lead time - One day



(b) Lead time - Three days



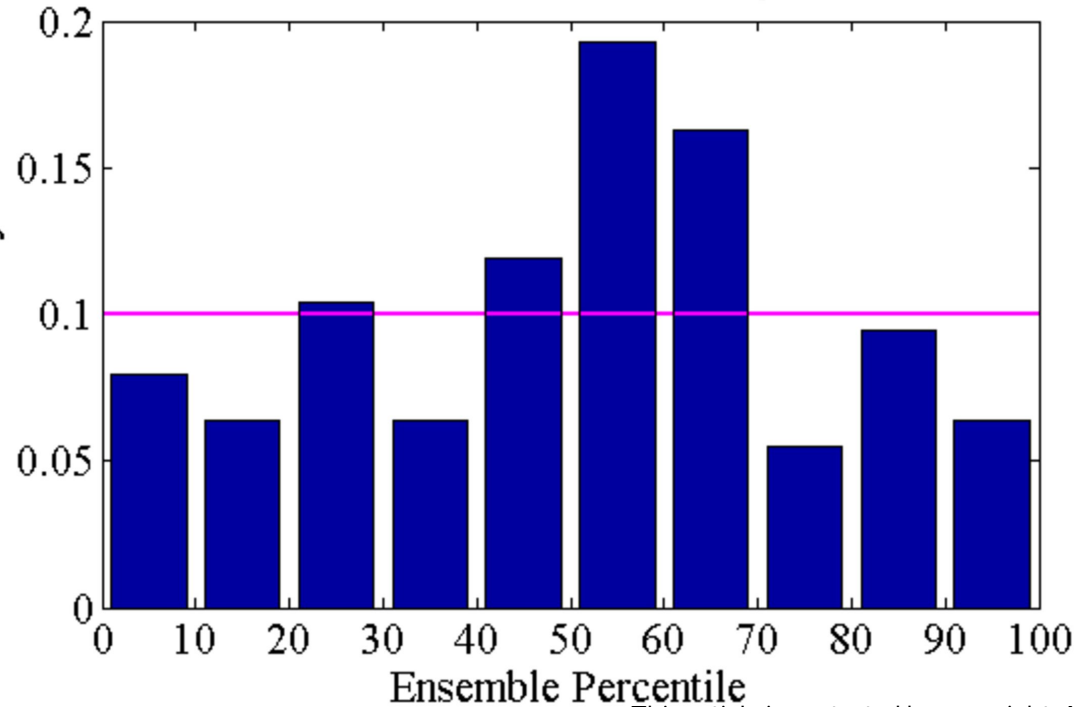
(c) Lead time - Five days



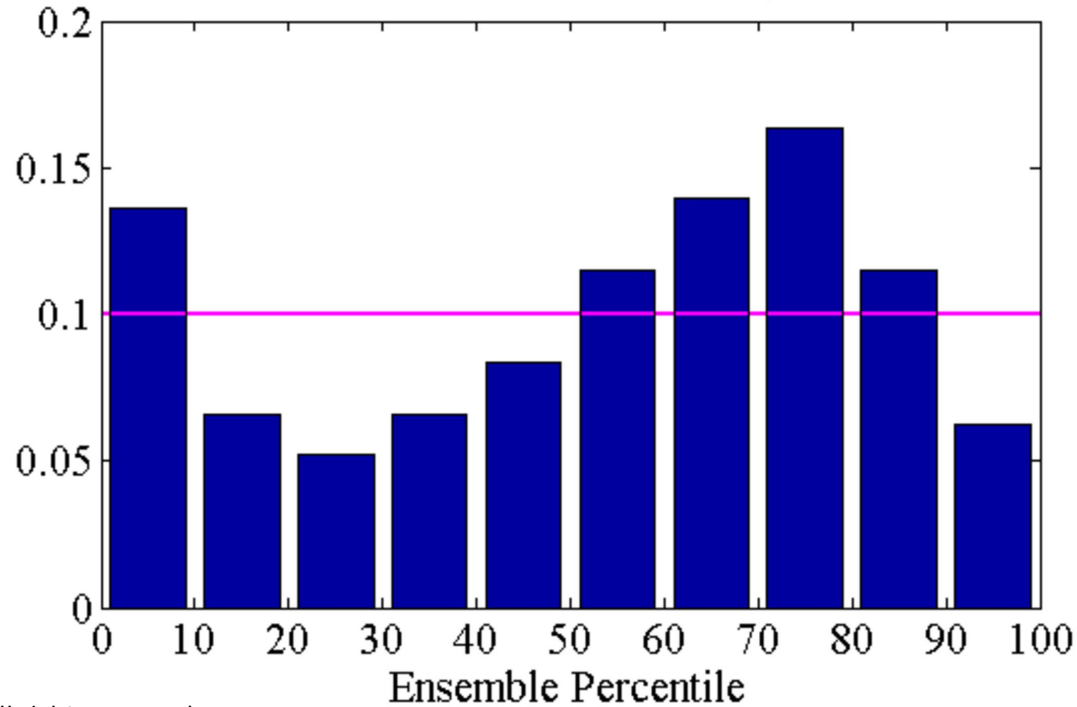
Ensemble spread (10<sup>th</sup> - 90<sup>th</sup> percentile) — — Observed flow

This article is protected by copyright. All rights reserved.

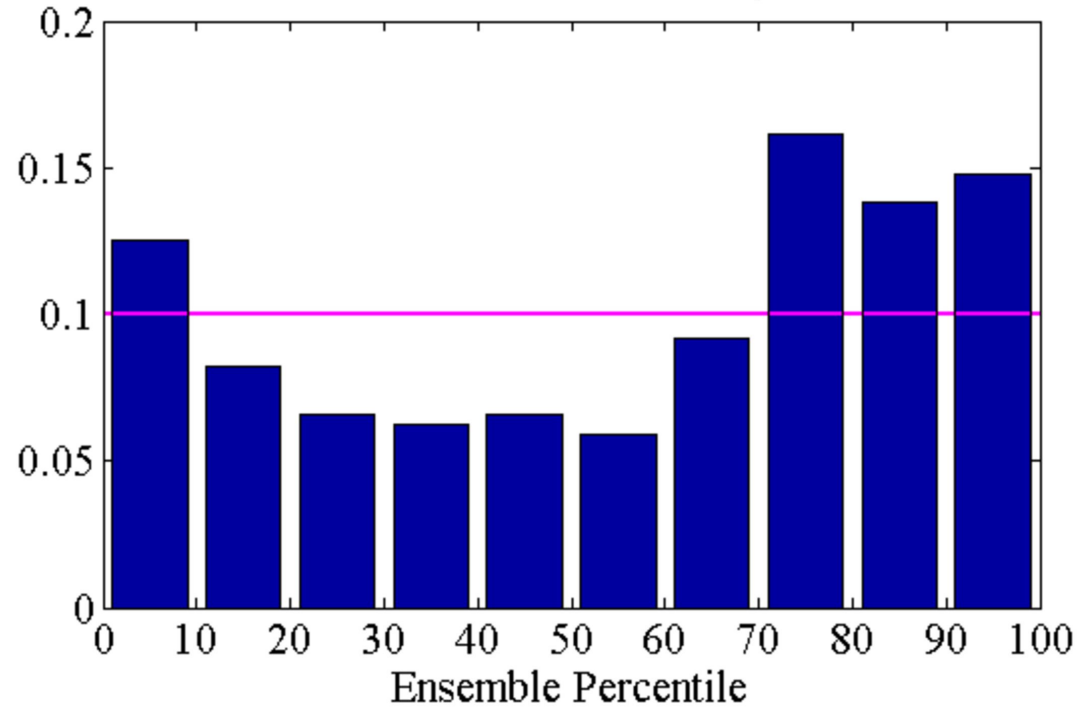
(a) Lead time - One day



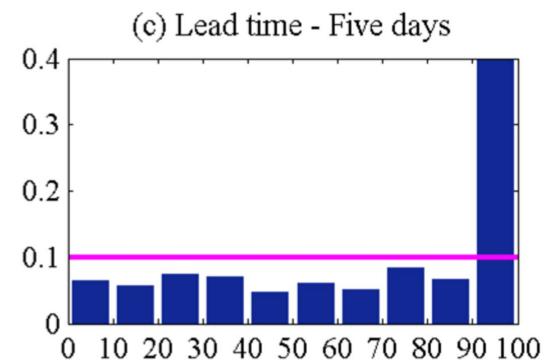
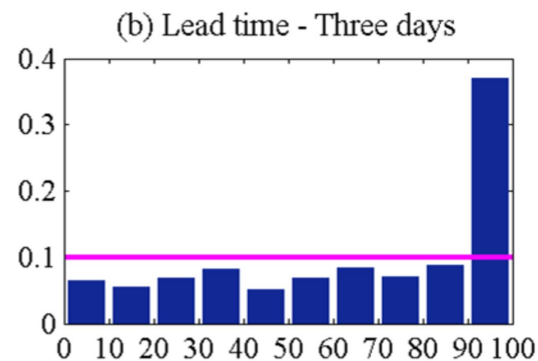
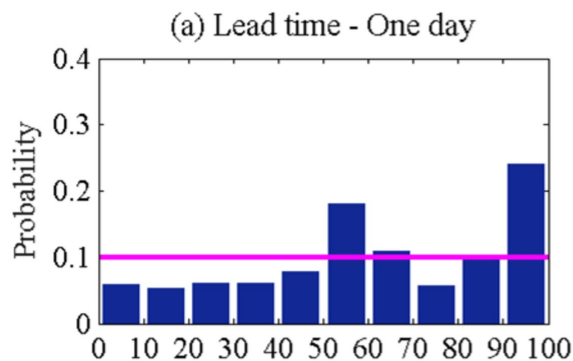
(b) Lead time - Three days



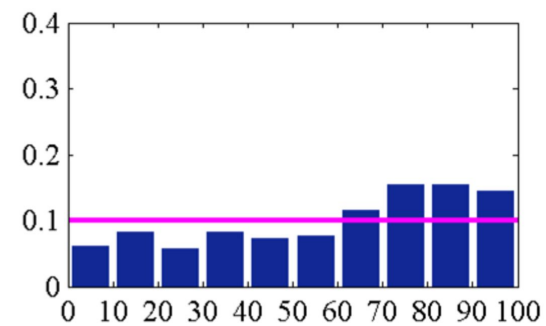
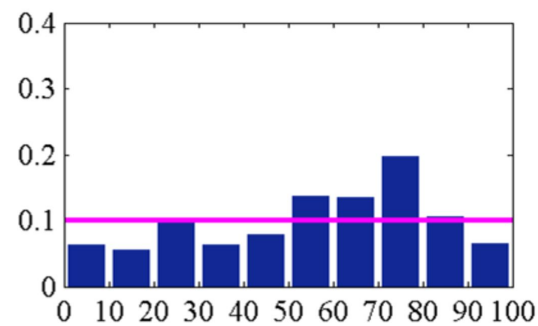
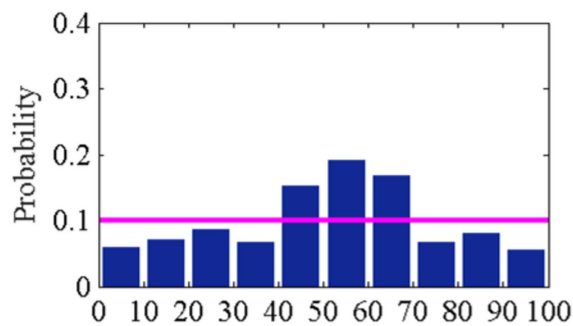
(c) Lead time - Five days



(A)  
 Irrigation flow  
 uncertainties } ✗  
 Observed weather  
 uncertainties } ✓  
 Forecast weather  
 uncertainties } ✓



(B)  
 Irrigation flow  
 uncertainties } ✓  
 Observed weather  
 uncertainties } ✗  
 Forecast weather  
 uncertainties } ✓



(C)  
 Irrigation flow  
 uncertainties } ✓  
 Observed weather  
 uncertainties } ✓  
 Forecast weather  
 uncertainties } ✓

