

Evaluating a prototype ensemble water quantity and quality forecasting system for the Fitzroy River Basin

L. E. Neumann^a, D. E. Robertson^a, B. Robson and R. Searle^a

^a CSIRO Land and Water

Email: luis.neumann@csiro.au

Abstract: The eReefs initiative is developing a series of marine hydrodynamic and biogeochemical models that will model and provide forecasts of rainfall and flooding impacts on the Great Barrier Reef. These models require real-time predictions and forecasts of riverine inflows and associated concentrations of fine sediments, speciated nutrients and carbon at each time step. This paper describes and evaluates one possible approach to the generation of water quantity and quality predictions and forecasts by linking ensemble streamflow forecasts and empirical Generalised Additive Models (GAMs). Forecasts of daily sediment and nutrient concentrations are generated by forcing GAMs with hourly streamflow forecasts that have been aggregated to daily totals. The streamflow and water quality forecasts are evaluated for over a 24-month period concluding in December 2013. The ensemble streamflow forecasts have considerably lower errors than simple persistence, which is used as input for the prototype marine models in forecasting mode. This suggests that marine modellers can potentially improve their simulations by using the streamflow forecasts in place of simple persistence. The ensemble forecasts of nutrient concentrations however display large errors, often significantly overestimating the observed values, which may limit their value for marine modelling. Errors in sediment and nutrient concentration forecasts, and the forecast uncertainties tend to be largest when the GAMs are extrapolating beyond the range of observations used to fit the GAMs model. Therefore improvements in the performance of sediment and nutrient concentration forecasts are most likely to be realised by fitting the GAMs to a larger set of either modelled or observed data.

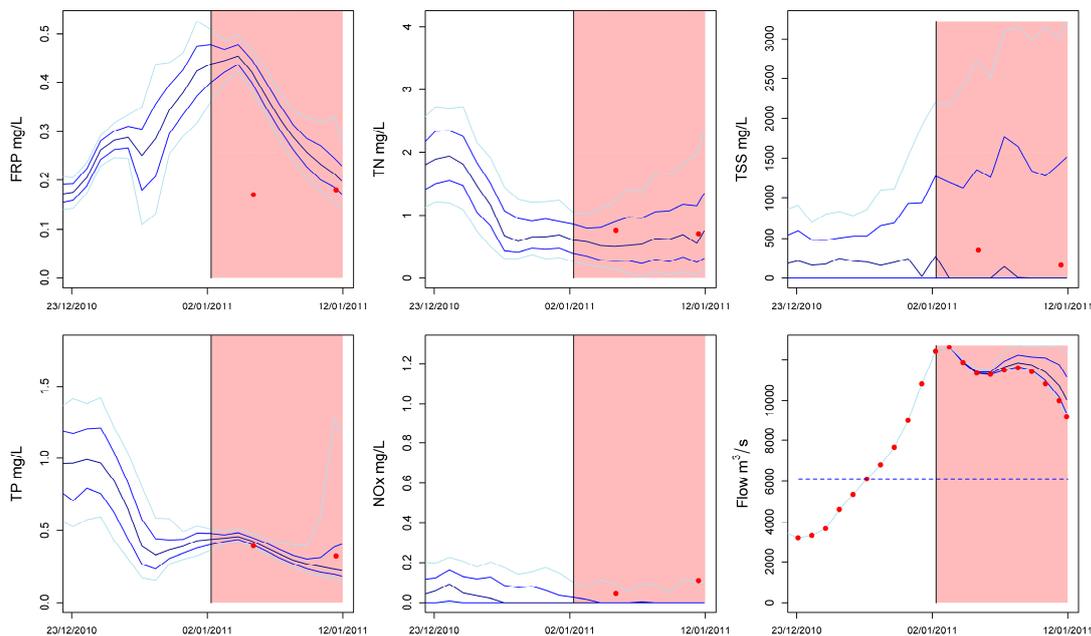


Figure 1. Forecasts of streamflow, sediment (TSS) and nutrient concentrations for Fitzroy River at The Gap issued at 2100 UTC on 02/01/2011, corresponding to a period of high streamflow flow. The 10 day forecast period is shaded, while the unshaded area are the predictions made using observed streamflows in the 10 days prior to issue of the forecast. The black line is the prediction/forecast median, mid blue lines and pale blue lines represent the [0.25,0.75] and [0.05,0.95] predictions/forecasts quantile ranges, respectively. Red points represent the observations. The dashed blue line on the flow plot is the maximum streamflow used in fitting the GAMs.

Keywords: Short term flood forecasting, ensemble forecasting, general additive models, water quality

1. INTRODUCTION

The eReefs initiative is developing a series of marine hydrodynamic and biogeochemical models that are intended to be used in near-real time or in forecasting mode (i.e. predicting conditions a few days ahead). These marine models require riverine inflows and associated concentrations of fine sediments, speciated nutrients and carbon at each time step as boundary conditions. For forecasting and near-real time applications, the use of monitoring data is not possible and a combination of models will be required to produce input series of streamflow and associated sediments and nutrients.

One possible combination of models is a semi-distributed hydrological model to simulate streamflows coupled with a set of generalised additive models (GAMs) to predict water quality constituents. Such a system would provide providing on-the-fly modelling and forecasting capability. The catchment forecasting system used here generates forecasts of water quantity and concentrations of nutrients and sediments using a two-step process. Firstly, ensemble forecasts of streamflow volumes are produced using the System for Continuous Hydrological Ensemble Forecasting [SCHEF; *Bennett et al.*, 2014], which forms the basis of the Bureau of Meteorology's current deterministic short-term streamflow forecasting service. A statistical model (specifically, a GAM) is then used to relate observed and forecast streamflows to concentrations of nutrients and sediments [*Robson and Dourdet*, 2015].

The aim of this paper is to describe and assess the retrospective performance of a prototype catchment forecasting model for the Fitzroy River generating predictions of water volumes, nutrient and sediment concentrations required by marine forecasting models. The first section introduces the approach to streamflow modelling and constituents modelling, followed by results of the flood forecasting from SCHEF and finally the estimations of nutrient and sediment concentrations using the GAMs and flood forecasts.

2. MODEL COMPONENTS

2.1. Data

Rainfall, potential evapotranspiration and streamflow observations are required for calibrating and applying models for forecasting. Hourly rainfall data for rain gauges within the Fitzroy basin were provided by the Bureau of Meteorology for the period 1/1/2007 until 20/05/2013. These data are from operational databases and therefore we apply a quality control algorithm to remove implausibly large observations and incorrectly recorded zero observations. To generate an extended record of hourly rainfall data to support adequate warming up of the hydrological models, we also extract daily rainfall from the Australian Water Availability Project [*Raupach et al.*, 2008] for the gauge locations and disaggregate these data to hourly time steps using simple averaging. Rainfall for model subareas are calculated from rainfall observations using inverse-distance-squared-weighting based on the distance between rainfall gauges and catchment subarea centroids. Potential evaporation (PE) is taken from the monthly gridded Australian Water Availability Project (AWAP) dataset [*Raupach et al.*, 2008] by disaggregating to daily values. The disaggregation assumes the monthly mean PE occurs at the middle of the month, and values for other days are obtained by linear interpolation between those values. Hourly streamflow observations are from the Queensland Department of Natural Resources and Mines online water monitoring portal and Bureau of Meteorology operational databases.

To ensure the forecasts were independent of any hydrological and error-correction model fitting, for the Fitzroy basin we calibrated the model over the period 01/01/2007 – 30/08/2010. To reduce the influence of the initial values of the state variables, the model was warmed up for a period 5 years using disaggregated AWAP rainfall data. To understand the performance of the hydrological and error models for an independent period, we validated for the period 31/08/2010 to 20/05/2013.

2.2. Hydrological modelling and forecasting

SCHEF uses a hydrological model to convert precipitation to runoff, and a nonlinear Muskingum channel routing algorithm to route the streamflow through a stream network. At each gauge location, an error correction model is applied to improve the streamflow forecasts generated by SCHEF. The hydrological component of SCHEF is applied in semi-distributed form [*Bennett et al.*, 2014], with runoff calculated using GR4H, the hourly time step version of the GR4J model described by Perrin *et al.* [2003]. Hydrological model predictions and forecasts are updated using an error correction model based on the dual-pass error correction model of Pagano *et al.* [2011]. For a more detailed description of the SCHEF model and its application to 9 Australian catchments, see *Bennett et al.* [2014]. The objective function used in the calibration of the SCHEF is a multi-objective function [*Bennett et al.*, 2014] using an average of the well-known Nash Sutcliffe

efficiency [NSE, *Nash and Sutcliffe*, 1970], a log transformed NSE to reduce bias towards large streamflow values flows [*Legates and McCabe*, 1999] and symmetrical measure of overall mean error, first introduced by Wang *et al.* [2011].

In forecast mode, SCHEF requires rainfall forecasts to generate streamflow predictions. The raw rainfall forecasts used in SCHEF are generated by the global instance of the Australian Parallel Suite 1 version of the Australian Community Climate and Earth-System Simulator (ACCESS) NWP model [*Bureau of Meteorology*, 2012]. The ACCESS-G NWP forecasts used by SCHEF are available at 3 hour intervals a horizontal resolution of ~40 km, and are produced for the next 10 days from the issue time. As the NWP model results are susceptible to spin-up issues for the first few hours, an initial period of 9 hours is neglected, and therefore SCHEF is able to produce forecasts for up to 9.5 days [*Bennett et al.*, 2014]. NWP rainfall forecasts were available for the period 1 January 2011 to 1 January 2013. The rainfall forecasts from the NWP model are post processed to correct bias and quantify uncertainty using a Bayesian rainfall post-processor [*Robertson et al.*, 2013] which combines a simplified version of the Bayesian joint probability (BJP) modelling approach [*Wang and Robertson; Wang et al.*] with the Schaake shuffle [*Clark et al.*, 2004]. The BJP model is then used to generate a 500 member ensemble of unbiased rainfall forecasts for each subarea and lead time. As the period in which NWP forecasts are available is short, a leave-one-month-out cross-validation procedure is used to generate the rainfall forecasts for this study, therefore maximizing the available number of forecasts used in the evaluation. The cross validation procedure ensures the post processed forecast is independent of observations for that period.

The forecasts are evaluated for the period 1 January 2011 to 1 January 2013. Three error measures are used to estimate the reliability of forecasts: the Nash Sutcliffe efficiency, relative bias and the Continuous Ranked Probability Score (CPRS), which measures the error of all ensemble members with respect to observations by integrating the squared distance between forecast and observed cumulative distribution functions [*Bennett et al.*, 2014]. The performance of SCHEF is also compared against a reference forecast, in this case persistence, in which the observed streamflow at the forecast issue time is assumed to continue for the duration of the forecast. Although it has been demonstrated elsewhere [*Bennett et al.*, 2013; *Bennett et al.*, 2014] that the use of persistence as a reference forecast is not a high benchmark for short term flow forecasts, currently marine modellers are using persistence streamflow forecasts and therefore it is a relevant reference forecasts for this study. In addition to persistence, the model outputs are also compared to raw NWP forecasts (i.e. before post processing) and for “perfect rainfall”, in which the model is forced using observed rainfall which show the best possible forecast taking into account the errors in the hydrological model.

Table 1. Summary of models predicting sediment and nutrient concentrations for the Fitzroy River at The Gap

Model Name	Modelled parameter	GAM terms
TSS	Total Suspended Solids	Log(Flow), s(DF,d=0.9)
DON	Dissolved Organic Nitrogen	s(DF,d=0.95), s(Nogoa)
DOP	Dissolved Organic Phosphorus	s(DF,d=0.95), s(log(Flow)), s(Nogoa)
FRP	Filterable Reactive Phosphorus	s(log(Flow)), s(Nogoa), s(Isaac)
NH4	Ammonium	s(DF,d=0.95), s(log(Comet+0.0001))
NOx	Nitrate	s(Flow), s(DF,d=0.95), s(log(Comet+0.0001))
PN	Particular Nitrogen	s(Flow), s(DF,d=0.95)
PP	Particulate Phosphorus	s(Flow), s(DF,d=0.95)
Flow is the streamflow in the Fitzroy River at The Gap, Nogoa is streamflow in the Nogoa River at Duck Ponds, Isaac is streamflow in the Isaac River at Yatton, Comet is streamflow in the Comet River at Comet Weir, DF is discounted streamflow at The Gap with a discount factor of d.		

2.3. Water quality modelling and forecasting

Generalised Additive Models (GAMs) are used to relate streamflows within the Fitzroy Basin to sediment and nutrient concentrations. GAMs are generalised regression models which use smoothing functions based on various covariates and they require significant less input data and have lower computational costs than process-based models and can be a powerful tool for the prediction of sediment loads. The smoothing

functions do not have a pre-determined form, but rather aim to capture the main features of the data. The basic methodology of the GAMs developed for the Fitzroy as described in detail in Robson and Dourdet [2015]. In deriving their model, Robson and Dourdet [2015] considered a large range of possible covariates to estimate sediment and nutrient concentrations at The Gap in the Fitzroy River. A variety of different covariates were found to provide some predictive power, but the most significant variable was found to be flow at different locations, and those variables are used here as shown in Table 1.

Predictions and forecasts of sediment and nutrient concentrations are generated by conditioning the GAMs on the simulated and forecast streamflows produced by the hydrologic and error correction models. The GAMs are fitted using daily observations of streamflow and the available measured concentrations of sediment and nutrient concentrations. We therefore aggregated hourly streamflow observations and forecasts to daily values and generated forecasts of daily sediment and nutrient concentrations. The GAMs can estimate not only the expected sediment and nutrient concentrations, but also the prediction uncertainty. We generated ensemble predictions and forecasts of sediment and nutrient concentrations by sampling one stochastic realisation from the GAM for each streamflow ensemble member using the following equation.

$$C_i = f(Q_i) + \sigma_{f(Q_i)} + \sigma_{res} \quad (1)$$

Where C_i is the sediment or nutrient concentration for ensemble member i , $f(Q_i)$ is the expected sediment or nutrient concentration from the GAM generated using streamflow predictors from ensemble member i , $\sigma_{f(Q_i)}$ is a sample from $N(0, \sigma_{f(Q_i)})$, where $\sigma_{f(Q_i)}$ is the standard error of the expected sediment or nutrient concentration, and

characterises the uncertainty in the expected sediment or nutrient concentration. σ_{res} is a sample from $N(0, \sigma_{res})$, where σ_{res} is the standard deviation of the model residual errors, and characterises the residual error of the model (pers. comm. Simon Wood, Professor Statistics

University of Bath, author of GAM R package). Only 13 water quality observations are available for the forecast evaluation period. This limits the ability to produce a robust, quantitative assessment of the performance of the sediment and nutrient concentration forecasts. We therefore assess the forecast performance using visual comparisons between observed and forecast sediment and nutrient concentrations.

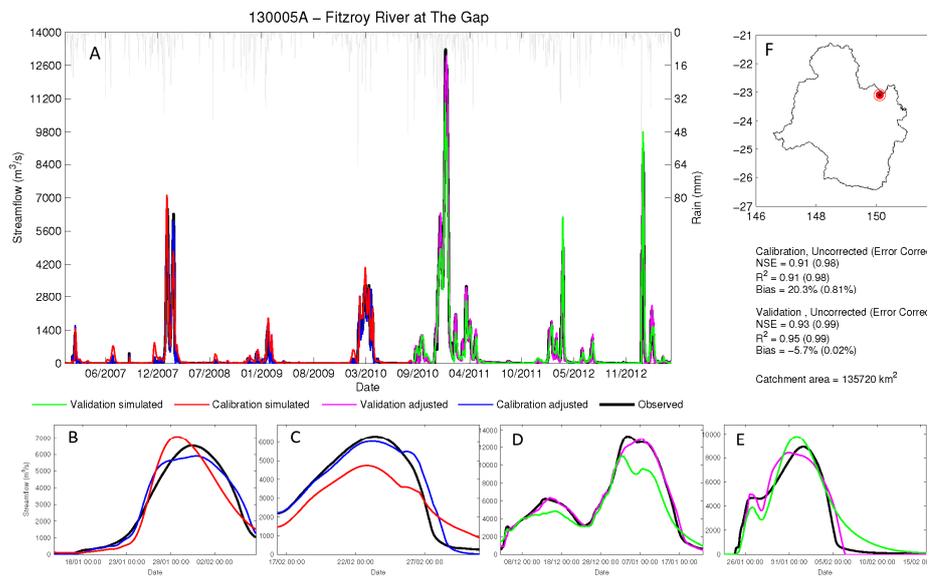


Figure 2. Calibration and validation results for gauge 130005A located at The Gap; A: hydrographs for both calibration and validation periods including raw (“simulated”) and error corrected (“adjusted”) model results; B-E: two largest events for calibration (B, C) and validation (D, E); F: location of gauge.

3. RESULTS

The Fitzroy model using SCHEF was calibrated at 39 gauging stations throughout the catchment with results at The Gap shown in Figure 2. In general, the performance of the model is satisfactory at most stations with $NSE > 0.7$ for the majority of the catchments and a similar performance between the calibration and validation periods. However, for several catchments the performance during the calibration period is rather poor but the performance during validation is much better. This is mainly due to the increase in availability of hourly rainfall from the calibration to the validation period as more hourly gauges are available. In terms of bias, there is a larger variability for both the validation and calibration period without obvious correlation as the NSE. Across most gauges, the error correction (adjusted results) significantly improves the calibration and validation results, with the model being able to predict the timing and magnitude of the peak reasonably

well at most locations. An example of the post processed rainfall and streamflow forecasts for all lead times up to 9 days for the Fitzroy River Catchment at The Gap using SCHEF are shown in Figure 3. The streamflow forecast from the SCHEF performs well, with the observations contained within the [0.75, 0.25] interval for most of forecast with the exception of the period around 3 days lead time. In particular, the model correctly predicts the magnitude of the peak flow with a lead time of 7 days; however the timing of the peak flow is offset by 24 hours.

An estimate of the accuracy of the streamflow forecasts generated using SCHEF is shown in Figure 4 using the CRPS score, and it is clear that forecasts using raw and post processed rainfall showing better CPRS scores than simple persistence, with similar improvement for NSE and bias (not shown). Although the results for the Fitzroy River at Gap shown in

Figure 4 shown very little difference CPRS scores between forecasts using raw and post-processed NWP rainfall, results for other locations in the catchment show locations where the use of post-processed NWP rainfall results in significant better results for lead times > 3 days, and for bias in particular, as already shown in other studies [Bennett *et al.*, 2014]. This difference in performance may be due do different causes including the high frequency of near zero rainfall and streamflow in headwaters and errors in observed rainfall.

Figure 1 present forecasts, and predictions for the preceding 10 days, for a set of forecasts issued during a period of high flow, when the GAMs are extrapolating beyond the range of streamflows data used to train the model. Total Phosphorus (TP) and Total Nitrogen (TN) concentrations tend to display an inverse relationship with streamflow with relatively lower flows tending to be associated with higher concentrations. This apparent inverse relationship will primarily arise due to a lagged response to streamflow dominated by the discounted streamflow predictors. Filterable reactive phosphorus (FRP) concentrations on the other hand tend to be positively correlated with streamflows. Total suspended solids (TSS) concentration predictions and forecasts display a weak inverse relationship with streamflow and have large prediction uncertainties. During low flow periods (not shown) nutrient and sediment concentrations tend to be relatively consistent with time.

Observations are plotted on the forecasts as red points. In some instances, the observations lie within the plotted prediction uncertainty intervals, particularly for the high flow forecast depicted in Figure 1. However for the majority of nutrient concentration forecasts, the observations lie well outside the plotted prediction uncertainty intervals. This suggests that the out-of-sample mean predictions contain significant errors and the uncertainty estimates produced by the GAMs may be under-dispersed. However, there are insufficient observations available to provide conclusive evidence. Errors in the mean prediction and uncertainty estimates may potentially arise from model over-fitting or changes in catchment conditions which may

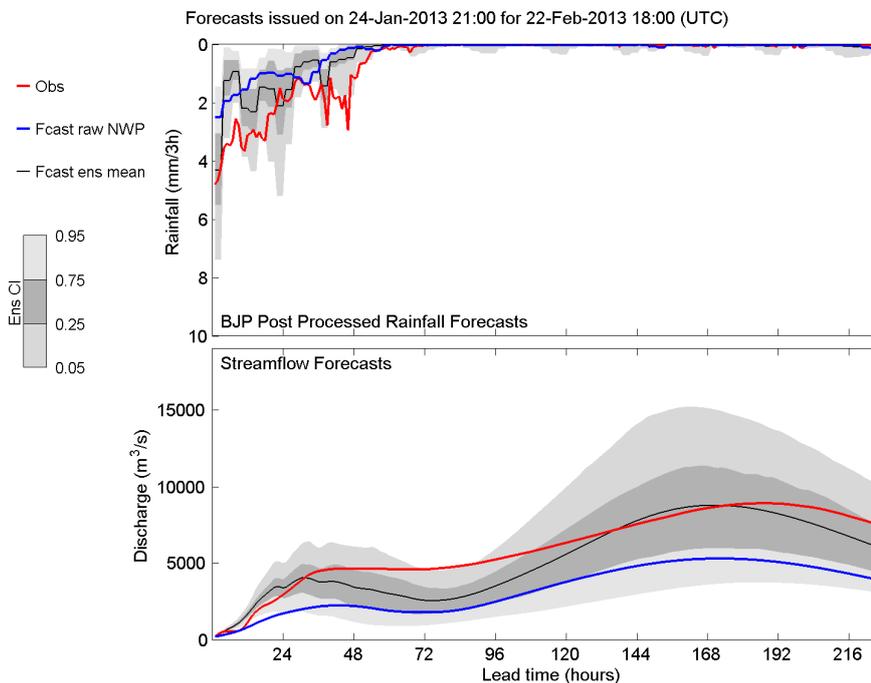


Figure 3. Example forecast generated by SCHEF for at the Fitzroy River at the Gap issued at 21:00 on 22 February March 2013. The top panel shows time series of post-processed ACCESS-G rainfall forecast for all lead times, and the bottom panel show ensemble streamflow forecasts and observations. Red lines are observations, blue lines are forecast using the NWP model rainfall, while the black lines show the ensemble mean. The light and dark grey shades show the [0.95, 0.05] and [0.75, 0.25] prediction intervals respectively.

invalidate the fitted model [Robson and Dourdet, 2015]. The forecasting period was generally wetter than the model training period, and statistical modelling approaches are not suited to this kind of extrapolation.

Forecasts of sediment concentrations appear to have smaller errors than for nutrient concentrations. However, where the GAMs are extrapolating the prediction uncertainty intervals are much larger than for the nutrient concentrations. There are many instances where the median forecast TSS concentration is equal to zero, while there are no instances of the observed TSS concentrations is zero. This artefact may arise due to the assumption that model residuals are normally distributed and suggests that further investigation into the residual error distribution may be required.

4. CONCLUSIONS AND RECOMMENDATIONS

The eReefs initiative is developing a series of marine hydrodynamic and biogeochemical models requiring real-time predictions and forecasts of riverine inflows and associated concentrations of fine sediments, speciated nutrients and carbon at each time step. This paper describes and evaluates one possible approach to the generation of water quantity and quality forecasts. The approach generates ensemble streamflow forecasts by integrating semi-distributed hydrological modelling, recent rainfall and streamflow observations, numerical weather predictions and rainfall and streamflow post-processing methods. Generalised Additive Models are developed to relate concurrent and lagged daily streamflow observations to daily sediment and nutrient concentrations. GAMs are established for Total Suspended Solids, Dissolved Organic Nitrogen, Dissolved Organic Phosphorus, Filterable Reactive Phosphorus, Ammonium, Nitrate, Particulate Nitrogen and Particulate Phosphorus concentrations. Forecasts of daily sediment and nutrient concentrations are generated by forcing statistical GAMs with daily streamflow forecasts that are produced by aggregating the hourly streamflow forecasts.

The streamflow and water quality forecasts are evaluated for over a 24-month period concluding in December 2013. The ensemble streamflow forecasts have considerably lower errors than simple persistence, which is the current input used by marine modellers. This suggests that marine modellers can potentially improve their simulations by using the streamflow forecasts in place of simple persistence. The ensemble forecasts of sediment and nutrient concentrations all display errors, often overestimating the observed values. Errors in nutrient concentration forecasts tend to be larger than those for sediment concentration forecasts. The largest forecast errors tend to occur when the GAMs are extrapolating beyond the range of streamflows data used to train the model.

Estimates of the uncertainty of the streamflow forecasts tend to be under dispersed, particularly at short lead times. The reason for this limitation is related to uncertainties in the hydrological modelling not being described by the current error correction model. This short-coming is currently being addressed by WIRADA research. The poor performance of the forecasts of nutrient and sediment concentrations may be related to several factors, including: limited number of observations available to fit the models (particularly at higher flows), model over fitting and changes

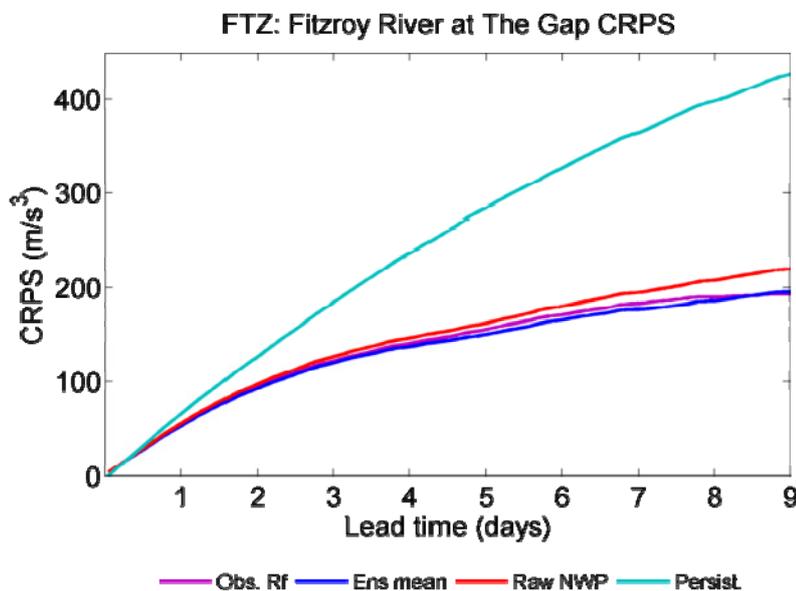


Figure 5. CRPS for different reference forecasts calculated at each lead time for gauge 130005A located at the Fitzroy River at the Gap. Results shown are: hydrological model forced with perfect (observed) rainfall forecasts (purple); hydrological model forced with post-processed NWP rainfall forecasts (blue); hydrological model forced with raw NWP rainfall forecasts (red); and persistence (green).

in the catchment processes that influence the nutrient and sediment generation.

Limitations in the available of observations can potentially be overcome by increased frequency of nutrient and sediment sampling. Work currently being undertaken as a part of eReefs is collecting water quality data at very high temporal resolutions. High frequency turbidity monitoring can be a good proxy for direct monitoring of TSS, though this requires the relationship between TSS and turbidity to be established for local conditions. These data were not available for the work described in this report, but could potentially lead to significant improvements the performance of the GAMs and water quality forecasts. Future work should investigate how best to harness the high temporal resolution water quality observations for forecasting applications.

An alternative strategy to improve the performance of the water quality forecasts may be to fit the GAMs to the output of a process-based model, such as the Source catchment models held by the Queensland Department of Natural Resources and Mines. While this strategy would increase the range of flow conditions over which the GAMs could be fitted and would ensure that near-real time and forecasting inputs to the marine model were consistent with inputs to be used in hindcasting, it presupposes that the output of the process-based models adequately describe the observations well at daily or sub-daily time scales.

ACKNOWLEDGMENTS

The assistance of the Extended Hydrological Prediction section of the Bureau of Meteorology in providing hourly rainfall and streamflow data, and supporting the catchment delineation is gratefully acknowledged.

REFERENCES

- Bennett, J., D. Robertson, D. Shrestha, and Q. Wang (2013), Selecting reference streamflow forecasts to demonstrate the performance of NWP-forced streamflow forecasts, paper presented at MODSIM 2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Adelaide.
- Bennett, J., D. Robertson, D. Shrestha, Q. Wang, D. Enever, P. Hapuarachchi, and N. Tuteja (2014), A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days, *In press*.
- Bureau of Meteorology (2012), APS1 upgrade of the ACCESS-G Numerical Weather Prediction system *Rep.*, NMOO Operations Bulletin Number 93, Bureau of Meteorology, Melbourne, Australia.
- Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby (2004), The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields, *Journal of Hydrometeorology*, 5(1), 243–262.
- Legates, D. R., and G. J. McCabe, Jr. (1999), Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35(1), 233–241.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10(3), 282–290.
- Pagano, T. C., Q. J. Wang, P. Hapuarachchi, and D. Robertson (2011), A dual-pass error-correction technique for forecasting streamflow, *Journal of Hydrology*, 405(3–4), 367–381.
- Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, 279(1–4), 275–289.
- Raupach, M. R., P. R. Briggs, V. Haverd, E. A. King, M. Paget, and C. M. Trudinger (2008), Australian Water Availability Project (AWAP), final report for Phase 3 *Rep.*, CSIRO Marine and Atmospheric Research, Canberra, Australia.
- Robertson, D. E., D. L. Shrestha, and Q. J. Wang (2013), Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, *Hydrol. Earth Syst. Sci.*, 17(9), 3587–3603.
- Robson, B. J., and V. Dourdet (2015), Prediction of sediment, particulate nutrient and dissolved nutrient concentrations in a dry tropical river to provide input to a mechanistic coastal water quality model, *Environmental Modelling & Software*, 63(0), 97–108.
- Wang, Q. J., and D. E. Robertson (2011), Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, *Water Resources Research*, 47(2), W02546.
- Wang, Q. J., D. E. Robertson, and F. H. S. Chiew (2009), A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resources Research*, 45(5), W05407.
- Wang, Q. J., T. C. Pagano, S. L. Zhou, H. A. P. Hapuarachchi, L. Zhang, and D. E. Robertson (2011), Monthly versus daily water balance models in simulating monthly runoff, *Journal of Hydrology*, 404(3–4), 166–175.