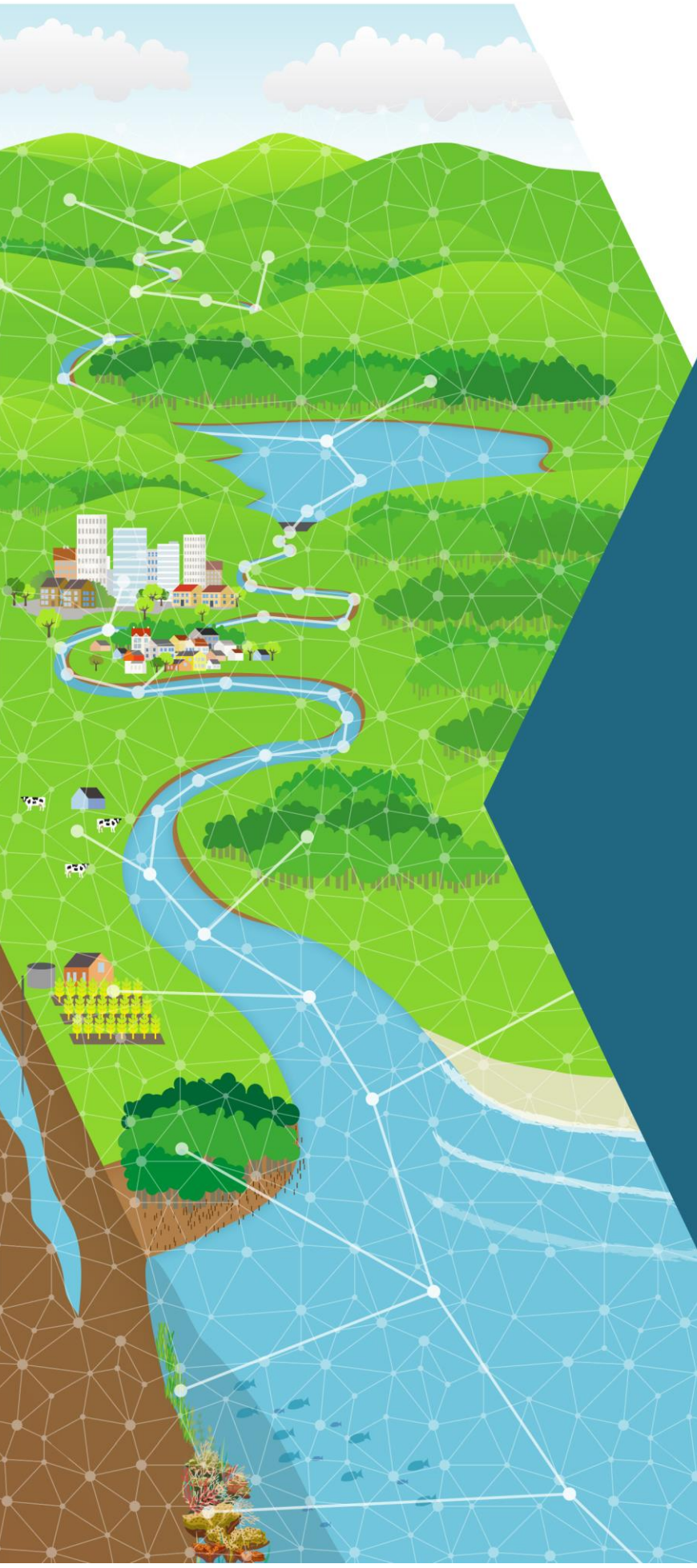
QUEENSLAND WATER MODELLING NETWORK



Water planning, integration and management

Addressing uncertainty in catchment models using machine learning techniques

Final Report

Report prepared by:
Aditya Singh, Dr. Daniel Botelho and Dr. Matthew Hipsey

For the Queensland Water Modelling Network

October 2020

BMT

Queensland Government

The Queensland Water Modelling Network (QWMN) is an initiative of the Queensland Government that aims to improve the state's capacity to model its surface water and groundwater resources and their quality. The QWMN is led by the Department of Environment and Science in partnership with the Department of Natural Resources, Mines and Energy and the Queensland Reconstruction Authority, with key links across industry, research and government.

This report documents research performed by BMT Commercial Australia Pty Ltd for the Queensland Water Modelling Network. The study investigated the potential for using machine learning techniques to model water quality and quantity in estuarine catchments. It should be noted that the methods and results presented in this study are for research purposes only and do not reflect current water modelling practices used by the Queensland Government.

Prepared by: Aditya Singh, Dr. Daniel Botelho and Dr. Matthew Hipsey

Citation:
BMT Commercial Australia. 2020. Addressing uncertainty in models using machine learning techniques. Report to the Queensland Water Modelling Network, Department of Environment and Science.

Acknowledgements

The authors would like to acknowledge the contributions of Dr. John Doherty and Dr. Frederick Bennett for their technical inputs into the project. Their commitment and availability to the project were timely and fundamental to achieve the project outcomes. Acknowledgements are also due to steering committee members Stephen Jeffrey, Department of Environment and Science and Dan Pagendam, CSIRO for their comments and review and the QWMN project management team, including Sarah Stevens, Jenny Riches and Rosie Hancocks. The authors are also grateful to Dr. Iris Tsoi from Healthy Land and Water, who provided all the observed water quality data used in the project.

Executive Summary

Catchment models are an important component of our environmental decision-making framework. They are used in a variety of contexts related to management of critical coastal ecosystems and landuse planning. Water quality predictions from catchments have typically been made using models that have been primarily developed for use in water resource planning and allocation applications. These models resolve hydrology at sub-catchment scale (few kms) and function on daily to monthly timesteps. This modelling approach is not fully compatible for a range of applications. An example of such an application is when the catchment models are coupled with three-dimensional receiving water quality models (RWQM) of estuaries for the assessment of estuarine environmental health. This incompatibility arises from the much finer spatial (in the order of metres) and temporal scales (in the order of seconds) of the RWQMs.

A distributed catchment modelling scheme was developed as part of this study to address issues around spatial and temporal resolution. This distributed model (with spatial scale in the order of 300 m) was comprised elements of existing models for rainfall-runoff and base flow prediction (AWBM) and routing (WBNM). These models were designed and executed to predict hourly flows, therefore providing much refined granularity to usual water modelling information available to water managers.

An integrated approach was taken to address cumulative model uncertainty. This approach involved using PEST as a tool for parameter optimisation with subsequent application of machine learning methods to model residuals from the parameter optimisation stage.

Most of the comparisons between observed and predicted flow data have resulted in high (>0.8) value for modelling efficiency. This indicated that the new model approach derived from a combination of distributed model and machine learning is capable of producing a good fit to observations, particularly considering that the model performance statistics were calculated against high-temporal resolution (i.e. hourly) data.

A simple empirical load generation model (based on a power law relationship) was used to estimate the load generation for a range of water quality constituents, including:

- total nitrogen (TN),
- total phosphorus (TP),
- total suspended solids (TSS),
- nitrogen oxides (NOx),
- ammonia (NH4),
- organic nitrogen (OrgN),
- filterable reactive phosphorus (FRP), and
- organic phosphorus (OrgP).

The resulting parameter-optimised output for the water quality constituents presented very low modelling efficiency. Machine Learning methods (GBM, DRF, DNN and an ensemble method) were subsequently used to model the residual between the model and observed data. Outputs from the machine learning models were able to improve the modelling efficiency significantly (in the region of 0.5 to 0.8) for most of the water quality constituents.

These modelling methods present a significant improvement in model performance over incumbent methods currently adopted for estuarine health assessments.

Contents

List of Figures

List of Tables

1 Introduction

Estuarine ecosystems are inherently complex, and their functions are influenced by a combination of oceans, climate, catchments and the region's anthropogenic activity. Ecosystems like the Great Barrier Reef and Moreton Bay are very sensitive to the nutrient loads flowing through small streams in the upper catchments, aggregating in the estuaries and finally making their way to the oceans. The estuaries are therefore a dynamic ecotone displaying the transition between the freshwater from the catchment and the coastal ocean, with intrinsic ecological processes at play.

Modelling this transition across complex receiving environments entails a nesting approach where outputs from one model (with different scale, complexity and functionality) are used to force the other. This trend is on the rise within the wider coastal ecosystems modelling community with models amalgamating outputs from large scale catchment models to force high resolution receiving models (Hipsey & Hamilton, 2015). Linking models of varying complexity and scope can lead to a cascade of uncertainties from each individual model and magnification in the uncertainty of final predictions.

In Southeast Queensland, the incumbent methodology to model catchment flows involved the use of e-Water SOURCE to generate daily predictions, which were linearly interpolated to sub-daily timesteps for linkage with a receiving water model - TUFLOW FV (as outlined in the Healthy Land and Water report cards process). SOURCE is a lumped, semi-distributed hydrology platform and the spatial outputs are available for each land use unit at a sub-catchment scale. TUFLOW FV is a hydrodynamic model that functions at a much smaller timestep (order of seconds) and higher spatial resolution (order of metres).

Rainfall events contribute with a higher proportion of pollutant loads through the river continuum into the estuaries, with temporal scales not longer than a few days, and with typical flow transitions occurring from minutes to hours. Noting these scales, water quality monitoring in the upper catchments is often carried out at sub-daily intervals during significant events. The daily timesteps (used by SOURCE) are rarely adequate to resolve storm events. As a result, performance to accurately predict catchment loads can be poor, comparisons with the monitored data can be problematic and simulated transformations across the estuarine system can be compromised to rectify the inaccuracies and uncertainty of the catchment model.

Despite the issues, the inherent treatment of uncertainty from the catchment model has been rudimentary. As a modelling platform, SOURCE offers internal functionality to use PEST (a parameter estimation tool, described later in section 4), but its application is limited, and more robust methods are required, especially in the context of modelling within South-east Queensland.

Process-based models like SOURCE use a combination of physical and empirical relationships to model complex catchment behaviour. In an ideal world, a complex physical model should be able to explain the whole gamut of processes acting within a system, but this is impractical. Data driven models (machine learning methods) explore complex, non-linear and intrinsic relationships within the data to predict otherwise unexplained system behaviour. A hybridisation approach of merging knowledge from physical and data driven methods could potentially exploit the benefits of each individual method. Errors in process-based model structure, parameter and data lead to both random and systematic error (Xu et al., 2012). In the context of catchment models, data driven models are

powerful tools for discovering complex non-linear relations from data, and therefore can be used to capture the systematic patterns in the error of process-based models.

With the above principles in mind, this study aimed to address this identified need for research, development and innovation (RD&I) in this area, entailing the development of a model that improves representation of sub-daily processes and can estimate and quantify its uncertainty for improved simulation of the linkages between catchments and the coastal ocean through the estuarine environment.

More specifically, this study proposes to meet the following objectives:

- Address structural inadequacy of the incumbent catchment model platform to meet the temporal and spatial resolution requirements.
- Evaluate the performance of parameter estimation methods and the overall quantification of parameter uncertainty.
- Develop machine learning techniques to model and improve model performance associated with residual uncertainty in the process-based model.

The following sections describe the site used for this study, the development of an alternative physical model from first principles, its performance against different storm events, application of parameter estimation methods (PEST) and the subsequent quantification of residual uncertainty using machine learning methods.

2 Study Site

The Bremer River catchment has been selected for this study. The catchment contributes with substantial flows, suspended sediment and nutrient loads to the Brisbane River estuary and subsequently into Moreton Bay. Healthy Land and Water (HLW) organise regular collection of event-based water quality data at key catchment locations through the Queensland Government as part of the South-east Queensland Catchment Loads Monitoring Program (SEQCLMP).

The Bremer River catchment spans an area of 203,000 Ha and is largely composed of Grazing and Natural Vegetation (71%), Conservation (10.5%) and Irrigated Agriculture (8.4%) as the key land uses. The catchment has two flow gauges operated by the Department of Natural Resources, Mining and Energy (DNRME), stations number 143110 (Adams Bridge) and 143107 (Walloon), with hourly flow data. The gauge at Walloon registers 'no flow' conditions through dry periods and measured flows ranging from 0.001 m^3/s to 2,055 m^3/s. The observed gauge data is 97% complete through the course of the whole period and is 99.8% complete during 2014-2019 (period of WQ observations). The following figure shows the location of the gauges and the extent of the catchment (Figure 2-1).

Water quality data is only available at station 143107 (Walloon) during specific events for the period spanning 2015 - 2019. The following water quality constituents (in mg/l) are available as part of the SEQCLMP:

- Total Suspended Solids (TSS)
- Total Nitrogen (TN)
- Nitrate and Nitrite as N (NO_x)
- Ammonium as N (NH_4)
- Organic Nitrogen as N (OrgN)
- Total Phosphorus (TP)
- Filterable Reactive Phosphorus as P (FRP)
- Organic Phosphorus as P (OrgP).

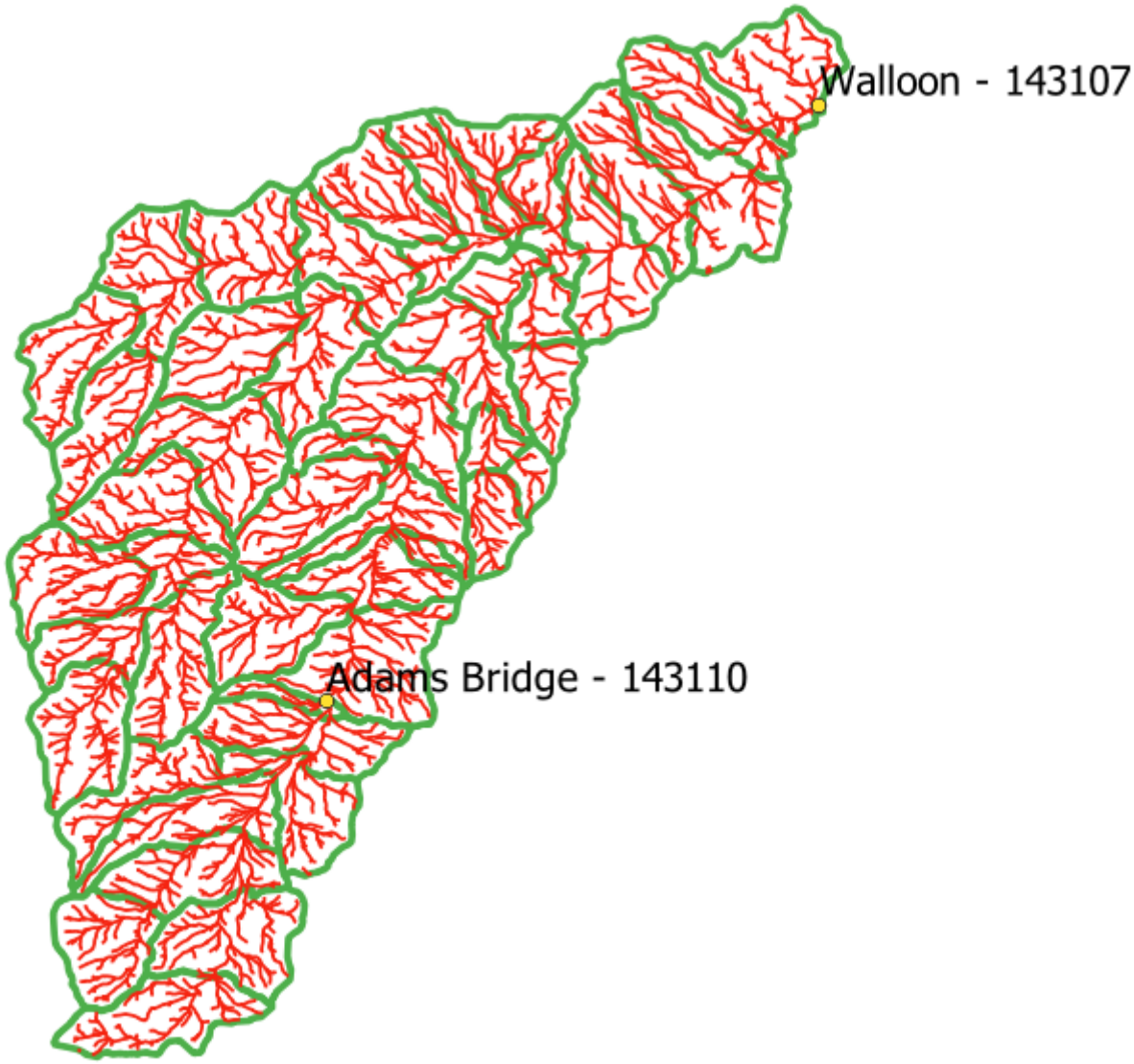


Figure 2-1 Locality Map

3 Development of a Physical Model

3.1 The case for a distributed hydrological model

Catchment modelling is a well-established discipline within the broader environmental modelling community (BMT, 2020). There is an increasing focus on the behaviour of catchments during rainfall events and the impact on the resulting sediment and nutrient loads. Like most natural systems, catchments are also viewed as a collection of largely complex (sometimes random) processes at a micro-scale that coalesce into a system whose behaviour can be explained at a macro-level using basic physical relations.

There are two major uses of hydrology models. The first involves using them to predict behaviour of catchments during extreme flow events (i.e. floods). These hydrology models are used to force more refined, 2-d models that predict inundation levels and flood extent downstream. This use-case involves running the models over a few days and focuses on modelling the flood hydrograph correctly. The second use case involves the analysis of long-term behaviour of the catchment (water resource allocation) and the export of nutrients and sediments through time scales of a year or longer. Both applications of models have a significant amount of overlap, yet different modelling tools are often used. Flood models have well developed methods to model flow routing between catchments because of the focus on getting a perfect fit on the hydrograph at sub-event time scales. Nutrient export models at the catchment scale function at a daily timestep and resolve functioning of a long-term water balance which is vaguely parameterised in flood models through initial and continuing losses. There is therefore merit in working with both classes of models to ensure they work at sub-daily (i.e. hourly) timescales and incorporate some of the complexities related to the water balance.

Most of the modelling tools available have operated on a reasonably large spatial scale of the orders of kilometres, being delineated at the sub-catchment level to comprise large extents associated with the entire catchments and watersheds. These models regionalise the hydrological properties over a significantly large area and are often referred to as lumped parsimonious hydrological models (Willems et al., 2014). Each sub-catchment behaves like a bucket that drains into the next and this whole system replicates the functioning of streams constrained by a localised water balance. The main drawback with this approach is the loss of non-homogeneity within the catchment leading to an inadequate representation of underlying processes.

SOURCE uses sub-catchments as its primary hydrological unit but gives an additional option of specifying land uses (namely functional units) as a proportion of the sub-catchment area, with the provision of model outputs discretised across the different land uses (Chiew & Siriwardena, 2020). This in fact makes SOURCE a semi-distributed hydrology model. The model (as used in Southeast Queensland) only operates at a daily time scale and, although it has provisions for it, the concept of flow routing between sub-catchments is neither well-developed nor adopted.

Unified River Basin Simulator (URBS), which is a common flood hydrology model, uses the sub-catchments as its basic computing unit. Muskingum routing is commonly used to route flows from one catchment to the other. Different options are available to model initial and continuing loss through rainfall events.

Soil Water Assessment Tool (SWAT) is a commonly used hydrology model across the world and combines some of the features of SOURCE in terms of a robust water balance model and also the ability to route flows (using Muskingum routing) from one catchment to the other at a sub-daily timescale. Like SOURCE, the weakness with the SWAT modelling approach is the regionalisation of parameters over a large area and the subsequent loss of information.

To overcome the issue with parameter regionalisation, a distributed hydrology model involving the use of structured mesh, with the mesh elements being the basic computational unit, can be adopted as an alternative to the models described above. The size and distribution of these elements can be controlled and therefore regionalisation of hydrological properties can happen at a more refined scale. Both fixed (square elements only) and flexible (quadrilateral and triangle elements) can be potentially used. The water balance can be computed on each model cell and drainage from one cell to the other can be specified using an appropriate routing scheme.

Distributed hydrology models have been used extensively with several instances available in scientific literature (Merritt et al., 2003). Adoption of these models within the practitioner community has been relatively slow, due to the relatively large computing requirements. This factor has greatly improved with greater computational power in recent times and the use of distributed models is likely to increase. The components of the distributed hydrology model developed as part of the present study are described below.

3.2 Model Mesh

A flexible mesh approach was adopted as part of this study (Figure 3-1). This approach was chosen over the more conventional fixed grid because of the ability to mesh around areas of interest and add/remove resolution in the flexible mesh. This type of mesh is also used by common hydrodynamic models like TUFLOW FV and so this gives the added ability to use existing pre and post processing tools for efficient model deployment.

The Bremer catchment model mesh has 19500 elements (both quadrilaterals and triangles), each with a nominal area of around 10 ha. The mesh elements were generally aligned with previously de-lineated sub-catchment boundaries (Figure 2-1).

The Digital Elevation Model (DEM) (DNRME, 2007) for the catchment area was interpolated on to the mesh cell centres to derive cell elevation values. A routing algorithm was written in Python to work out the order in which cells would drain from one to the other. As expected, these routes (links) between cells were observed to coincide with major overland flow paths (Figure 3-2).

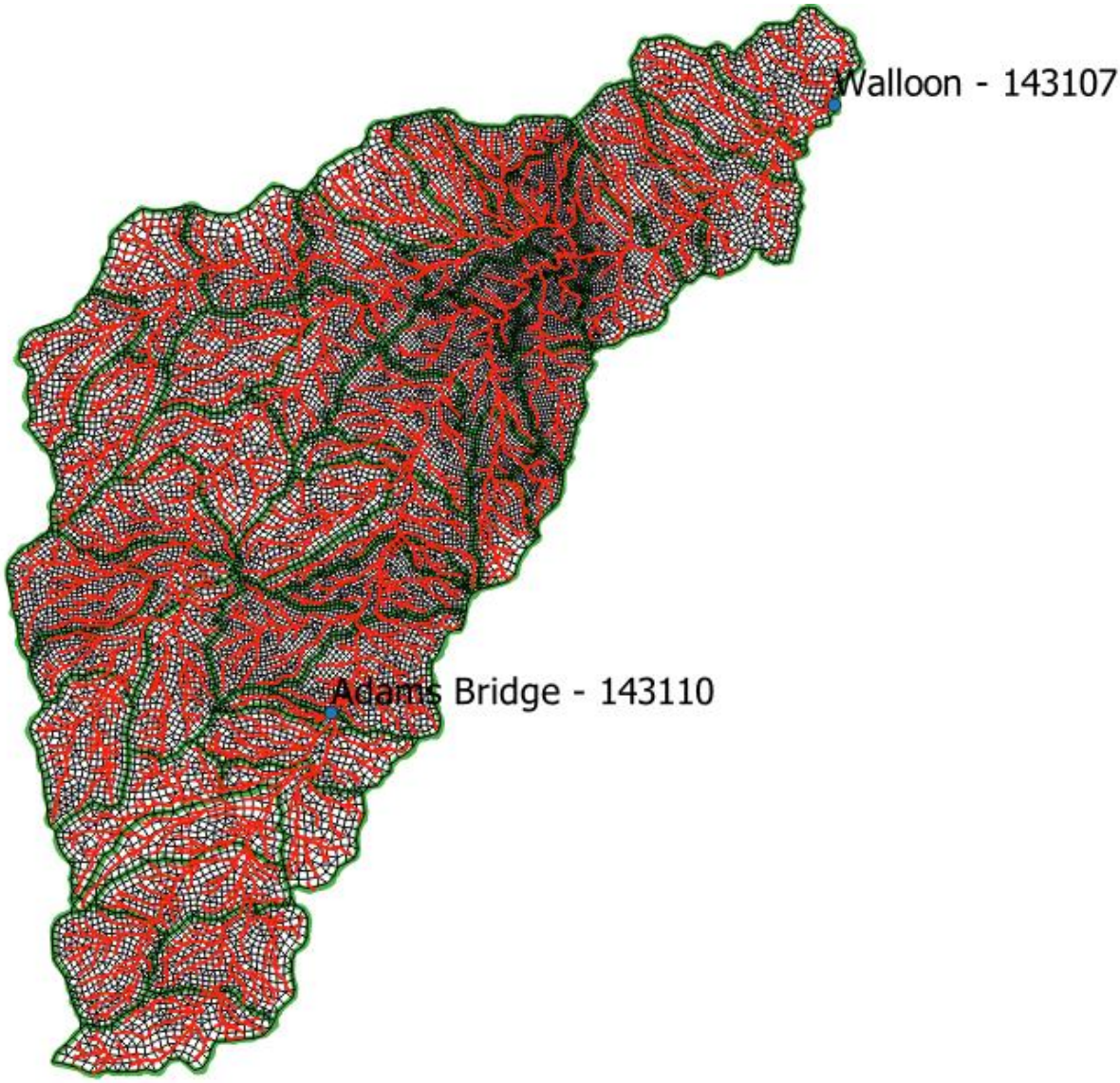


Figure 3-1 Flexible mesh

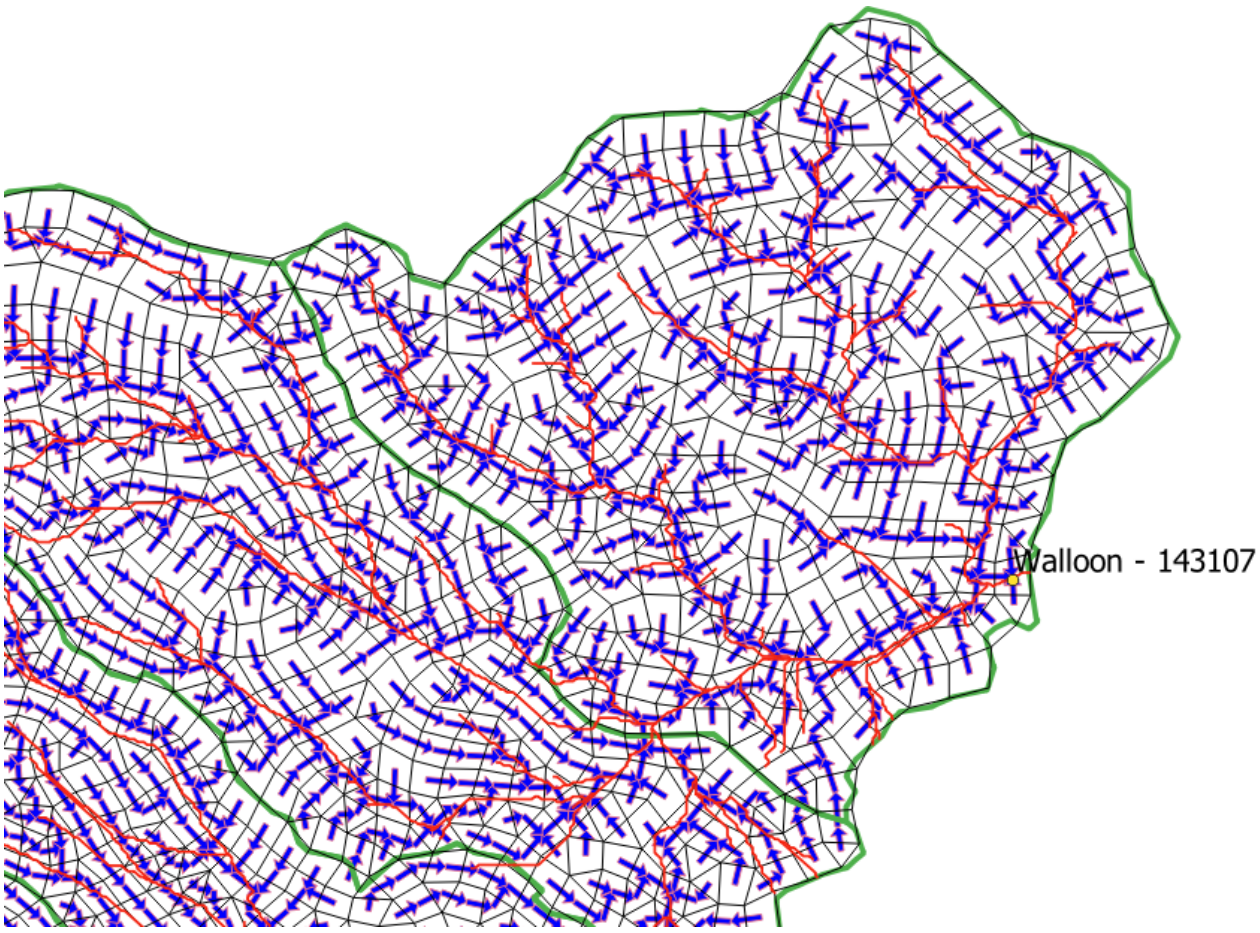


Figure 3-2 Flow routes

3.3 Water Balance Model

The Australian Water Balance Model (AWBM) was adopted for calculation of run-off and subsurface flow in this study. This model was developed as part of the work done in the erstwhile CRC-Hydrology and has been applied extensively to a wide range of Australian catchments (Yu & Zhu, 2015). AWBM is a key component of the SOURCE modelling platform and is often used as an alternative to another water balance model called SIMHYD (used in the incumbent model system).

AWBM models five storages - three surface stores to simulate partial areas of runoff, a base flow store and a surface runoff routing store (Figure 3-3).

The model was modified to run on an hourly timestep by forcing with hourly precipitation and rainfall data. AWBM was implemented to undertake calculations on each cell element separately.

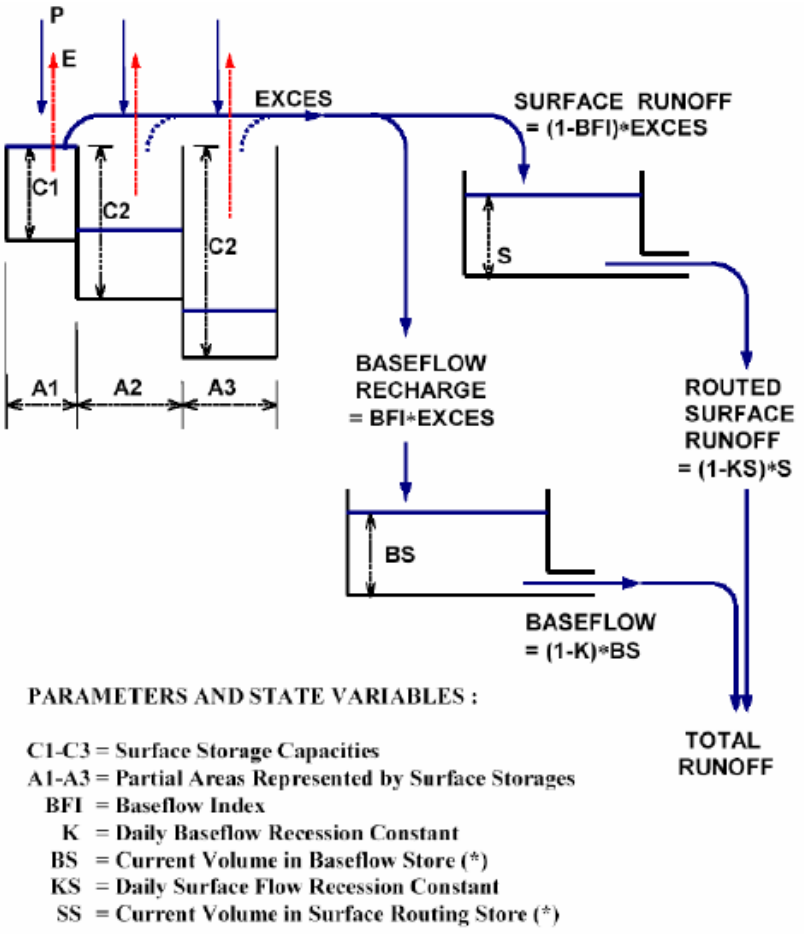


Figure 3-3 AWBM model structure (Yu & Zhu, 2015)

3.4 Flow Routing

The excess flow from each cell has to be routed through the network and a scheme was adapted from the Watershed Bounded Network Model (WBNM). WBNM calculates a delay separately for overland flow (Boughton & Askew, 1968) and stream channel (Boyd, 1978) based on the following empirical relationships:

$$D_o = L * (A_i)^{0.57} * (Q_j)^c$$

$$D_s = 0.6 * L * (A_i)^{0.57} * (Q_j)^c$$

D_o is the overland delay, D_s is the stream channel delay, L is the lag parameter, A_i is the element i area, Q_j is the flow through the link j and c is the non-linearity constant that links flow to the lag (in hours). Usually the non-linearity implies that larger flows pass through with less lag and smaller flows have larger lag.

This model is relatively simple (and empirical) when compared to other common schemes like the Muskingum routing, which models the propagation of the flood wave through a catchment with fixed storage. The Muskingum routing was extensively tested and was found to be numerically unstable

when applied to a distributed model (Barry & Bajracharya, 1995). Therefore, the Boughton & Askew (1968) and Boyd (1978) approximations were adopted.

3.5 Load Generation

In addition to flow, the model was developed to simulate eight different constituents – Total Suspended Solids (TSS), Total Nitrogen (TN), Nitrogen as Nitrates and Nitrites (NO_x), Nitrogen as Ammonium (NH_4), Organic Nitrogen (OrgN), Total Phosphorus (TP), Filterable Reactive Phosphorus (FRP) and Organic Phosphorus (OrgP).

The following empirical relationship was used to model the generation of different constituents.

$$E = \gamma * (Q_j)^{\delta}$$

E is the constituent concentration (in mg/l), γ is a constant of proportionality and δ is the exponential power used to scale the constituent concentration to the flow generated. This formulation is in line with the approximately linear relationship between the logarithms of flow and concentration (see Appendix A).

More complex models involving build-up and wash-off of pollutants have also been considered (e.g. Wijesiri et al., 2015), and although they can be potentially applied in the future, were not considered in this study.

3.6 Rainfall Distribution

Spatial and temporal distribution of rainfall form important and necessary components of the model. Rainfall data is sparse and available from an array of rain gauges within the catchment. There are 18 active rain gauge stations in the vicinity of the Bremer catchment (Figure 3-4). A raw time series of all the gauges was obtained from the Bureau of Meteorology (BoM) and hourly rainfall was calculated at each rain gauge.

The following steps were performed to distribute (i.e. interpolate) the rainfall on to the model mesh:

- At each timestep, rain gauges with available data were identified.
- The distance was calculated to each station for each cell.
- Weights were calculated for each station for each cell using the following inverse-distance relationship:

$$w = \frac{\left(\frac{1}{d_i}\right)^{\alpha}}{\sum\left(\frac{1}{d_i}\right)^{\alpha}}$$

where d_i is the distance of a particular station i from the cell and α is the power.

- The rainfall from each station was multiplied by its weight and added up to obtain the overall rainfall.

α controls the interpolation between stations and determines the spread of rainfall over the grid. Lumped catchment models tend to spread rainfall uniformly over the whole catchment area and this

can often be erroneous. While rainfall interpolation aims at diminishing this effect, the distribution of rainfall across the catchment is often a considerable source of uncertainty in the modelling.

Figure 3-4 Rainfall gauge locations

3.7 Evapotranspiration

Evapotranspiration is a major factor affecting the hydrology and ultimately the overall runoff from the catchment. Losses from evapotranspiration influence the amount of saturation in the soil water stores, which in turn affects the amount of runoff generated. Incumbent models like SOURCE often use long term datasets like SILO to obtain spatially variable evapotranspiration at a daily timescale.

In this study, SILO data was obtained at the BoM gauging station - Hattonvale O'Shea Road (40095). A daily timeseries was available and this was uniformly applied throughout the catchment after conversion to an hourly value. The same hourly value was used throughout the day. This is a very rudimentary assumption and more complex relationships like the Penman-Monteith equation can be used to calculate spatially variable, hourly evapotranspiration. Such a relationship was not pursued as part of the present study.

4 Application of PEST

Model calibration (i.e. adjustment to match observations) is an important aspect of hydrological models, for both flow and constituent generation. The process of 'model calibration' is essentially an ill-posed inversion problem involving observations, parameter estimates and model operation (Doherty, 1994).

$$h = Z[k] + \epsilon$$

where h is the set of observations/system state and ϵ is the error associated with them. Z represents the operation of a model (complex and non-linear), and k is the parameter set applied.

Regularisation is the process of achieving a unique solution to the ill posed matrix inversion. An ill-posed inversion problem does not have a unique solution and regularisation guarantees uniqueness (but not necessarily correctness). Estimates of the parameter set obtained can then be used to make future predictions and error minimisation is done by ensuring that the prediction is roughly at the center of the associated posterior probability distribution.

Manual Regularisation (Regionalisation) is often done as a first step. This often involves setting some of the parameters to known (best estimate) fixed values during the parameter estimation process. In spatial models (like catchment models) values can be fixed based on some known spatial relationships (like landuse, slope etc.). The fundamental assumption behind manual regularisation is that macro-scale system properties that are homogenous are estimable but micro-scale inhomogeneities are not.

Once the initial manual regionalisation is undertaken, Singular Value Decomposition (SVD) is often used as an approach to progress model regularisation and involves reduction of the parameter space dimensions to an extent which yields a unique solution to the problem (Doherty, 1994). This simplification is done in a mathematically optimal way with respect to calibration data available. The SVD framework involves subdivision of parameters that are estimable into one subspace and individual or parameter combinations that are not estimable into another subspace. The framework allows for mathematical and computational efficiency while also being able to compute the potential predictive and parameter error.

PEST was used as a parameter estimation tool to undertake the model regularisation and a global parameter optimisation. PEST HP (the high-performance computing optimised version of PEST) was used as part of this project, because of its ability to distribute model runs across different compute nodes and the availability of high-performance computing facility owned by BMT. As a result, the Jacobian matrix (a parameter ensemble that tests models sensitivity) was efficiently populated in parallel under practical run times.

4.1 Regionalisation of Parameters

As part of the model setup, there are twelve parameters that influence the flow calibration and sixteen others (8 constituents times 2 parameters) that influence the water quality calibration. Optimising about 28 parameters for each cell (19500 cells) can be an arduous task and hence a regionalisation approach (manual regularisation) was required to limit the total number of parameters.

A conventional approach has involved the application of land use practices to regionalise hydrology and pollutant export parameters. This method has been used extensively for catchment modelling within Southeast Queensland. A major drawback with this method is the fact that land use distribution within a catchment can be relatively homogenous (i.e. 71% of the Bremer catchment is covered under Grazing and Natural Vegetation, Figure 4-1). Regionalisation of hydrology parameters over such a large area is likely to result in an inadequate representation of catchment flow and transport properties.

An alternative regionalisation of parameters based on slope was amenable to homogeneous catchment land uses, and therefore applied in this study. A regionalisation based both on land uses and slope is likely a more complementary approach, however this strategy was not pursued in the present study. The cell-wise elevations obtained previously for determining the routing were used to calculate slope at each cell and a classification was made based on the percentile distribution to have nearly equal number of cells within 10 different slope zones (Figure 4-2). These 10 zones were used as a basis for regionalisation of parameters. This approach supplements the potential weakness of using a simplistic routing scheme (that ignores catchment slope) and its parameters can now incorporate changes due to slope.

α, which is the parameter influencing the interpolation of rainfall on the mesh was regionalised based on Voronoi polygons built around 10 different rain gauges to represent their respective areas of influence. This was kept separate to the regionalisation of flow parameters.

4.2 Flow Calibration

The following parameters were selected for the optimisation process in PEST:

Table 4-1 Summary of flow parameters

Parameter	Model Component	Upper Bound	Lower Bound	Default Value
A1	Water Balance	0.001	0.99	0.143
A2	Water Balance	0.001	0.99	0.5
C1	Water Balance	1	300	7
C2	Water Balance	1	1000	70
C3	Water Balance	1	2000	150
K	Water Balance	0.001	0.99	0.95
Ks	Water Balance	0.001	0.99	0.35
BFI	Water Balance	0.001	0.99	0.35
L (lag parameter)	Routing/Lag	0.001	10	1.7
c (Non-linearity)	Routing/Lag	0.001	0.99	0.23
α	Rainfall distribution	0.001	10	2

Different parameter variables were created to correspond with each of the 10 slope zones. The parameters were approximated to be statistically independent from each other and normally

distributed within the upper and lower bounds specified above. This is a simplified first pass assumption and more complex prior distributions can also be specified.

A multi-criteria objective function was setup to facilitate matching with hourly observation data from the two gauges, 143110 (Adams Bridge) and 143107 (Walloon). The objective function was weighted to match the following objectives:

- Identify and match the peaks in the dataset
- Identify and match the rising and falling limbs of the hydrograph
- Identify and match the number of high flow events in the dataset
- Identify and match a percentile distribution of flows through the model run period

The following weights were applied to each observation type:

Table 4-2 Summary of flow parameters

Observation Group	Relative Weight
Below the 90th percentile flows	1
Above the 90th percentile flows	3
Percentile Distribution of Flows	10
Maximum Flow	20
Number of Peaks	40

Flow calibration was done over a period between 18/01/2015 and 10/05/2015.

4.3 Water Quality Calibration

The water quality calibration was done on the parameters γ and δ for each of the constituents. Regionalisation of these parameters was done based on slope in this study.

Since, γ and δ have been specified for each constituent (eight in total) and the regionalisation was done for each of the 10 slope zones, there are 160 different parameter variables available for the optimisation process.

Similar assumptions adopted in the flow calibrations around the parameters being normally distributed within the upper and lower bounds were applied.

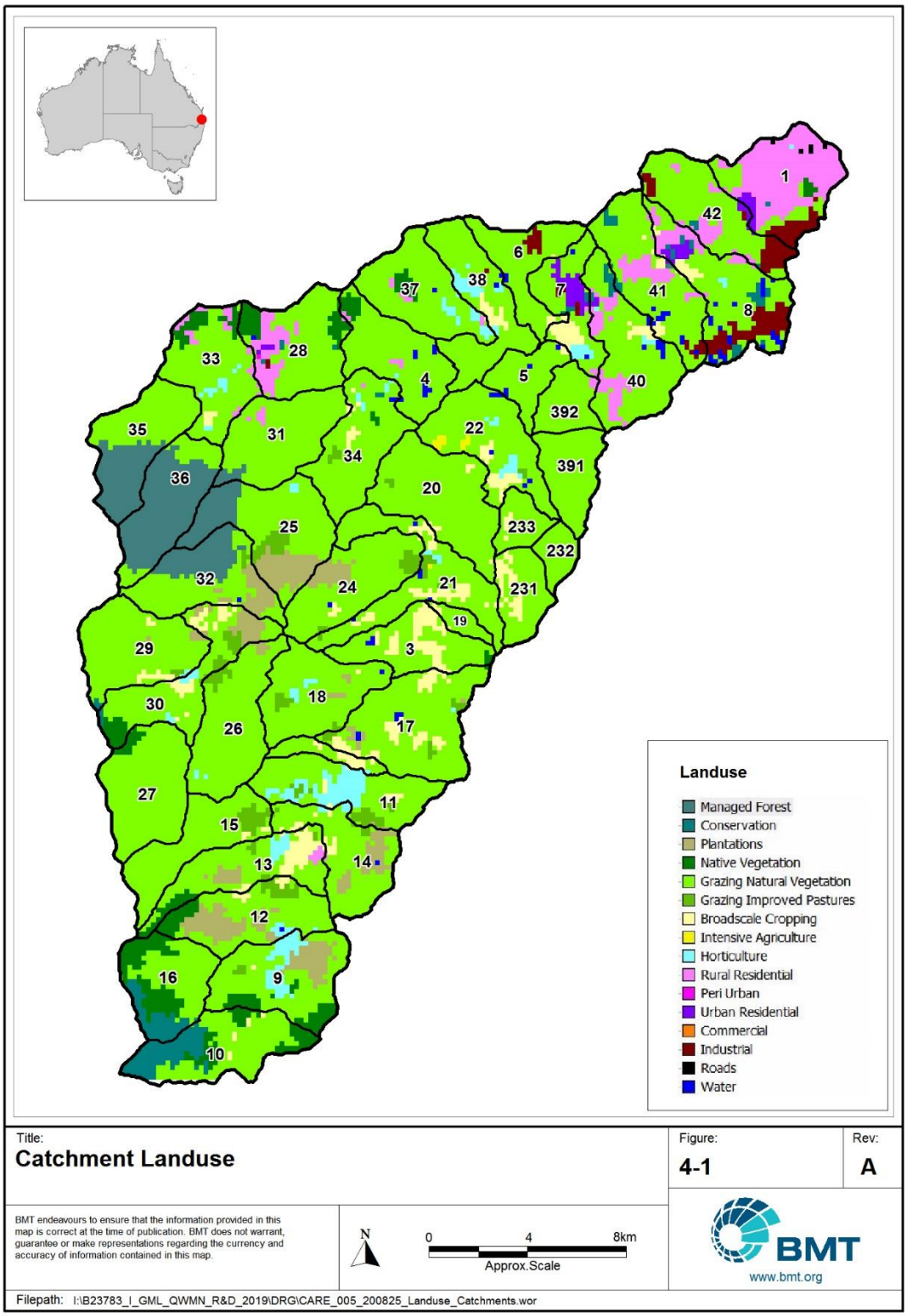


Figure 4-1 Catchment land use

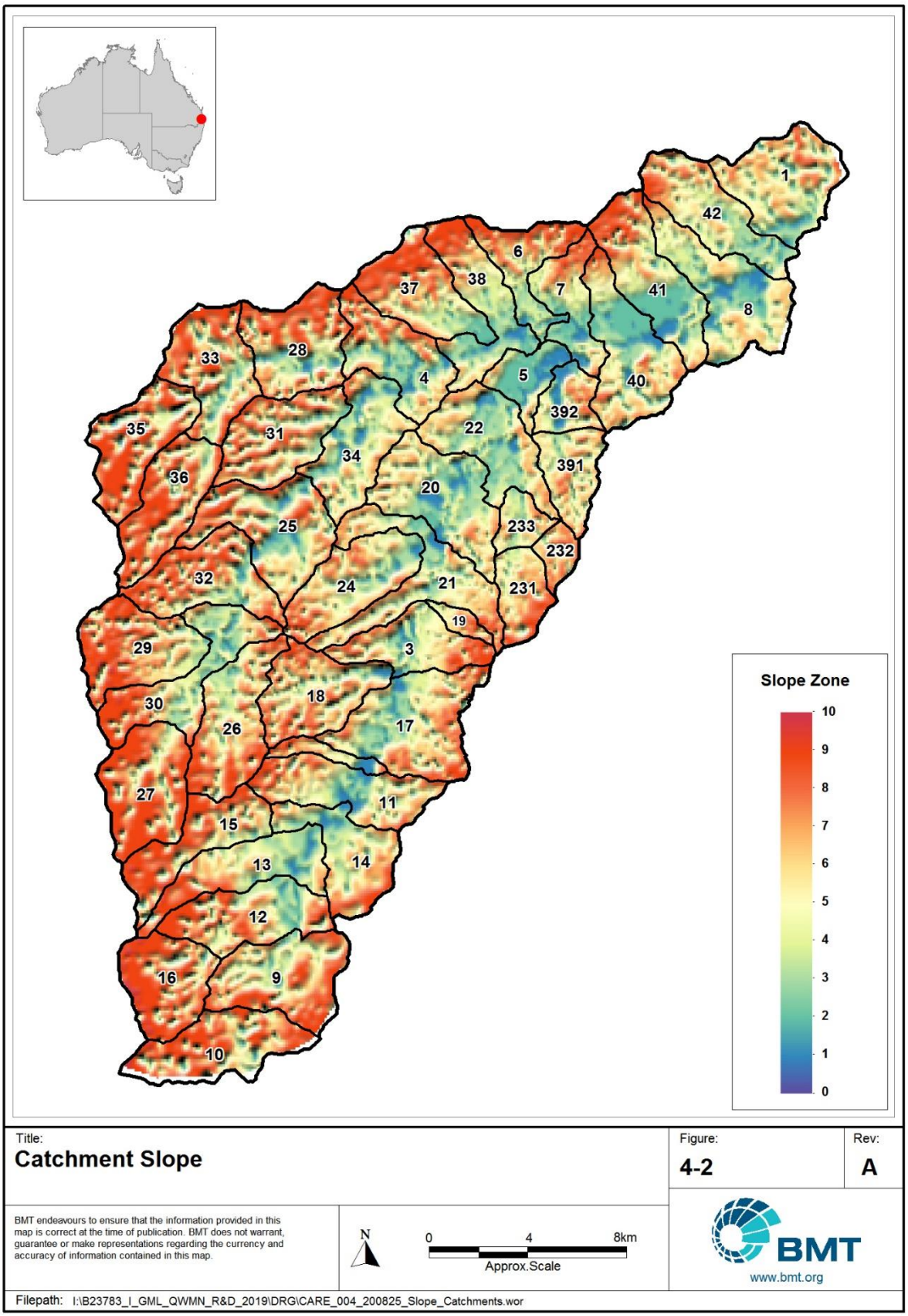


Figure 4-2 Catchment Slope

5 Machine Learning Methods

Machine learning methods exploit non-linear relationships within known data sources (predictors) to predict complex system behaviour. A hybridisation of a calibrated physical model (like the one described in the previous sections) with machine learning methods to predict model residuals could potentially improve the overall prediction quality.

The following four different machine learning methods were tested in the present study:

- Gradient Boosting Machine (GBM)
- Distributed Random Forest (DRF)
- Deep Neural Network (DNN)
- An ensemble method using a combination of the above three (Ens).

Distributed Random Forest (DRF) is an ensemble machine learning technique that can be used for both regression and classification tasks. DRF uses multiple decision trees whose predictions are aggregated using bootstrap aggregation, also commonly called bagging. This involves using a different data sample for training each tree with replacement from the original set (Breiman, 2017).

Gradient Boosting Machines (GBM) is a framework where decision trees or any form of base learners are recursively improved in terms of their prediction quality. The first step involves fitting the model to the data, then fitting a model to the residuals and getting the new model by combining the previous model and the residual model. The iteration process is described as:

$$F(x) = F_1(x) \rightarrow F_2(x) = F_1(x) + h_1(x) \ldots \rightarrow F_M(x) = F_{M-1}(x) + h_{M-1}(x)$$

$h_M(x)$ simply represents a model of the residuals while $F_M(x)$ is a model of the predictions from the base learner.

The underlying base learner used typically with DRF and GBM methods is the Classification and Regression Tree (CART). As the name suggests they can be used for both classification and regression tasks. Assuming R^d is the data space, the data can be split into K disjoint subspaces $\{R_1, R_2, \ldots, R_k\}$ where each $R_j \subset R^d$. The same decision/prediction is made for all $x \in R_j$ for each feature subspace. The following algorithm is used for generating regression trees (Breiman, 2017):

- Begin with the first feature subspace.
- For each feature $j = 1, \ldots, d$ and for each value $\mathrm{v} \in R$ on which a split is possible:
  - Split the dataset:

$$I_L = \{\mathrm{X1}, x_i^a < \mathrm{v}\} \text{and } I_R = \{\mathrm{X2}, x_j^a \geq \mathrm{v}\}$$

  - Estimate the average y for each node using $\overline{y_R}$ and $\overline{y_L}$.

$$\overline{y_L} = \frac{\sum_{i \in I_L} y_i}{|I_L|} \; and \; \overline{y_R} = \frac{\sum_{i \in I_R} y_i}{|I_R|}$$

  - Estimate the quality of the split by calculating the squared loss.

$$L^2 = \sum_{l=0}^{L} (y_l - \overline{y_L})^2 + \sum_{l=0}^{L} (y_r - \overline{y_R})^2$$

- Choose the split with minimal loss.
- Do the process recursively on both children.

Both DRF and GBM are ensemble methods and combine a number of regression trees to make predictions. The averaged values in the terminal node of each tree are treated as the prediction of that tree and the average of all the trees is taken as the final prediction.

Deep Neural Networks (DNN) are a group of algorithms modelled loosely on the functioning patterns of a human brain. The key idea is to recognise underlying patterns and in a way trace back raw data through these identifiable features in the data. These algorithms have been used in a variety of clustering and classification applications like image recognition and feature identification (Sutskever et al., 2013).

The stacked ensemble method within H2O determines the optimal combination of a collection of prediction algorithms (GBM, DRF and DNN) in this case using a process called stacking (van der Laan et al., 2007).

5.1 Key Predictors

Predictors are the key sets of data which machine learning methods employ to build non-linear relationships. A suite of predictors have been previously proposed in (Wang et al., 2019)and these include lagged hydrological and rainfall data.

The following table lists some of the key predictors:

Table 5-1 List of machine learning predictors

Key Predictor	Description
Q	Total Flow (raw, lagged by 4 days and 7 days)
Q_{BF}	Total Base Flow (raw, lagged by 4 days and 7 days)
Q_{QF}	Total Quick Flow (raw, lagged by 4 days and 7 days)
P	Total weighted precipitation (raw, lagged by 4 days and 7 days)
JD	Calculated Julian Day
$cos(JD)$	Cosine of the Julian Day
$sin(JD)$	Sine of the Julian Day
All constituents	Other constituents will also be used.

5.2 Modelling Process

H2O was used as the basic computation engine to implement the machine learning algorithms described earlier. A version of H2O compatible with Python was used.

The following steps were implemented as part of the modelling process:

- All the predictors were arranged in a H2O data frame and the predictors were clearly identified.

- The main aim of the models was to predict the error between observed and modelled values.
- The data frame was split into a ratio of 70:30 with 70% being used for training and the remaining 30% used for testing.
- A set of hyper-parameters were described for each of the three methods – GBM, DRF and DNN.
- The three methods were implemented with a random grid search to identify the best set of hyper parameters.
- Once the best model was identified for each, it was run in prediction mode to model the error for the whole WQ timeseries. The contribution of each individual variable to the predictive ability of the model was also output as a table.
- The best set of models amongst the three was used to create an ensemble.
- The ensemble was also run to model the error for the whole timeseries.
- The error was added back to the predictions to constitute the final predictions.

The following section compares the performance of these methods individually against the physical model.

6 Modelling Results

This section describes the modelled results obtained sequentially through each process. Here we note that machine learning was only applied to the water quality component of the modelling, noting flow is one of the predictors.

6.1 2011 Event with default values

The model was run for the major flow event in Jan 2011 using default hydrological values. Sheet plots were prepared to show the spatial distribution of rainfall and streamflow across the Bremer catchment. The spatial plots clearly indicate that the model can replicate some of the underlying physical characteristics related to flow connectivity.

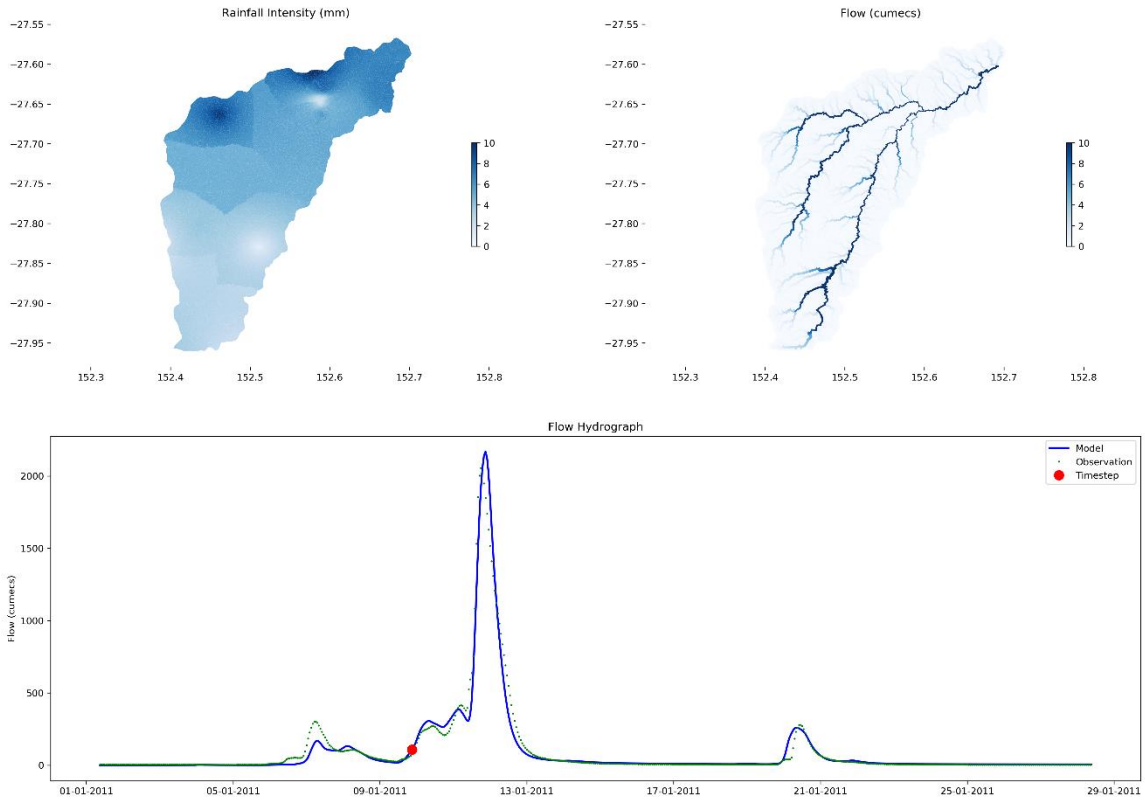


Figure 6-1 Spatial distribution of rainfall and flow

6.2 Performance of Flow Model

The model was calibrated to four small storm events between 18/01/2015 and 10/05/2015. Small storm events are particularly difficult to model given the rainfall spatial-temporal distribution is not as uniform as a large-scale event, and therefore the interpolation of rainfall between rain gauges is likely to be less accurate. Spurious data at the rain gauges is also another factor that was found to affect the calibration (not shown).

The following results were obtained at the end of the flow calibration (Figure 6-2).

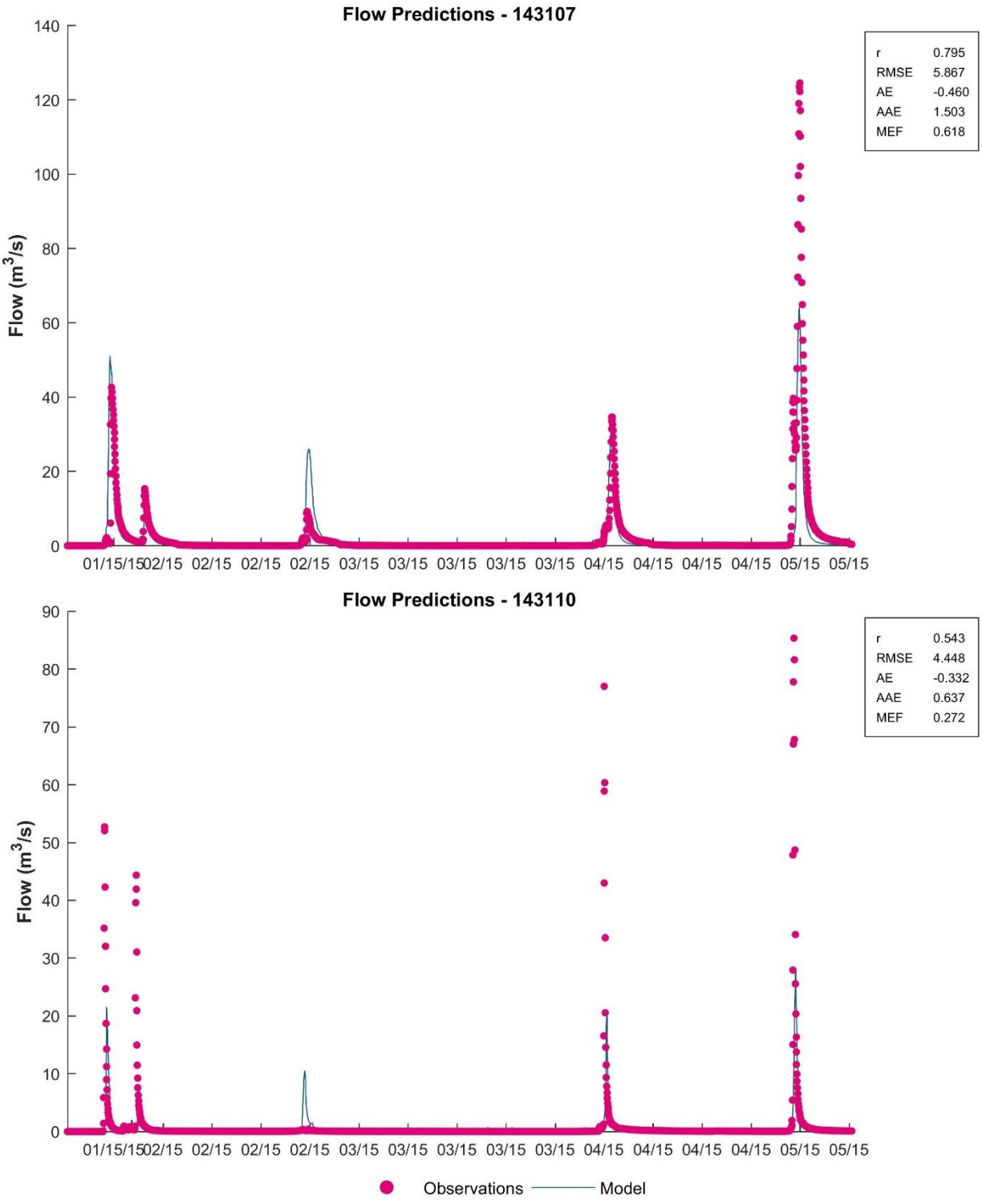


Figure 6-2 Flow calibration - 2015

The model was re-run over other periods of significant rainfall. Comparisons against observations in different periods are shown in Figure 6-3 to Figure 6-10. Note that the range of flow changes across the different figures to better illustrate the flow evolution.

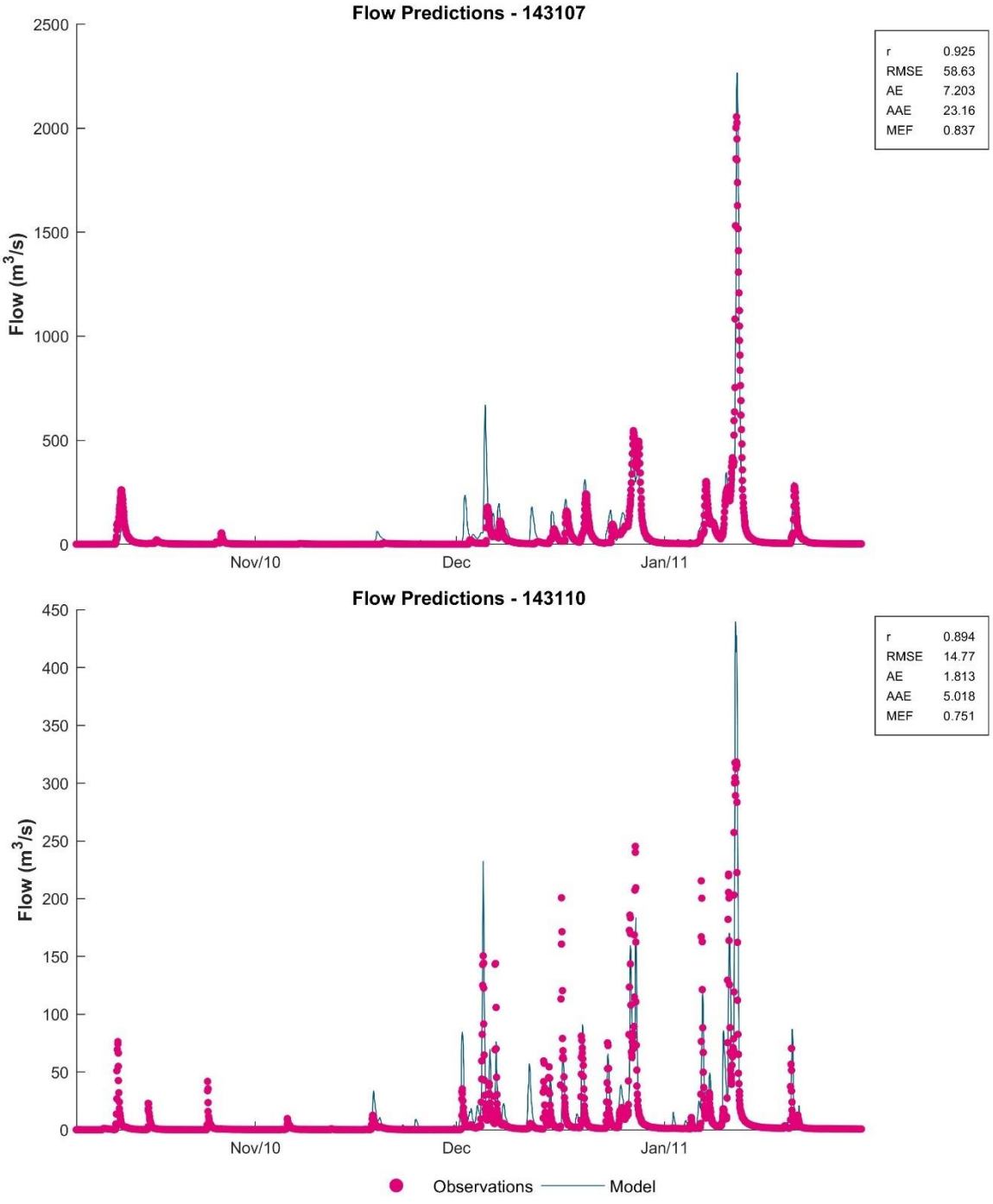


Figure 6-3 Flow comparison – 2011

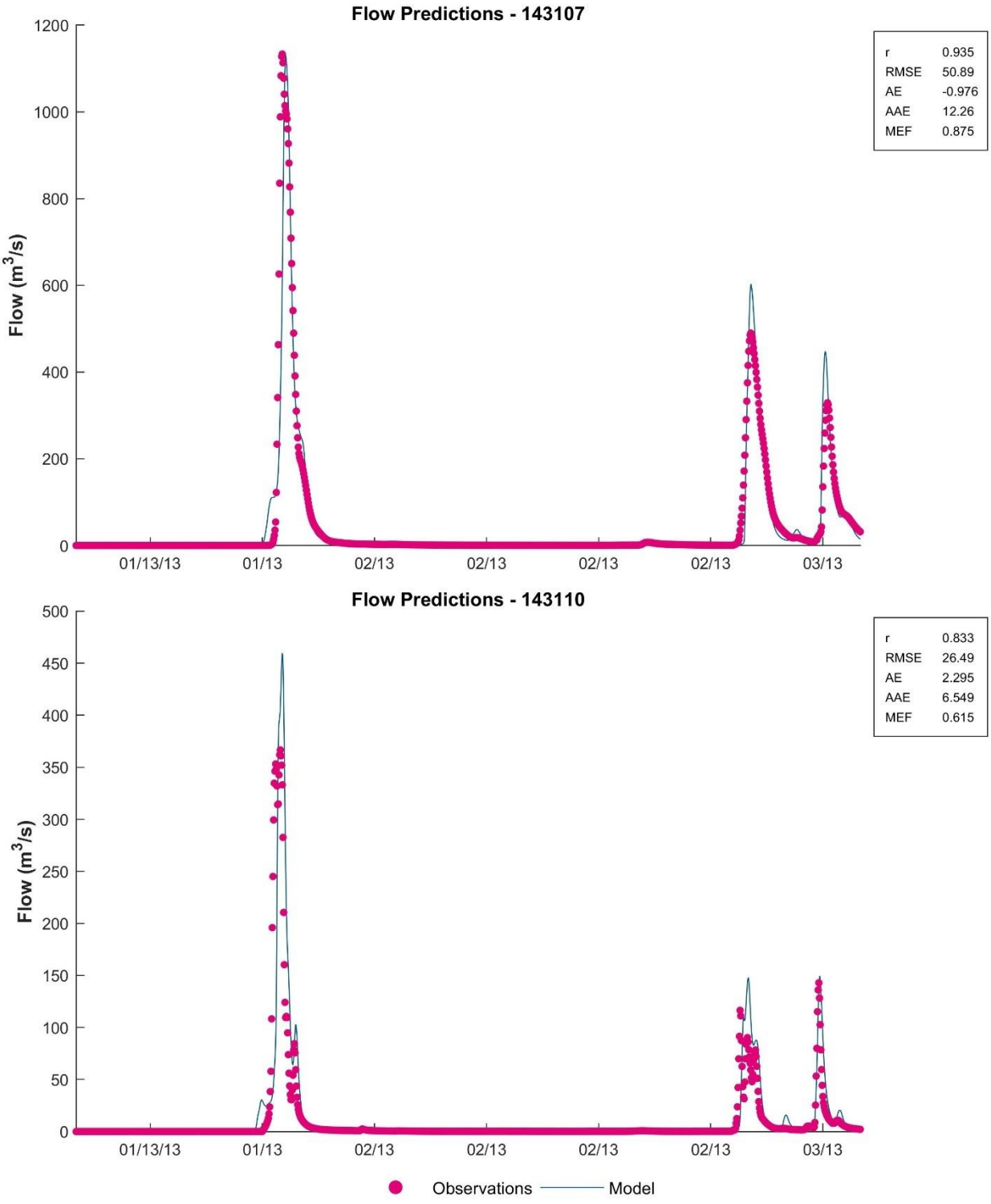


Figure 6-4 Flow comparison – 2013

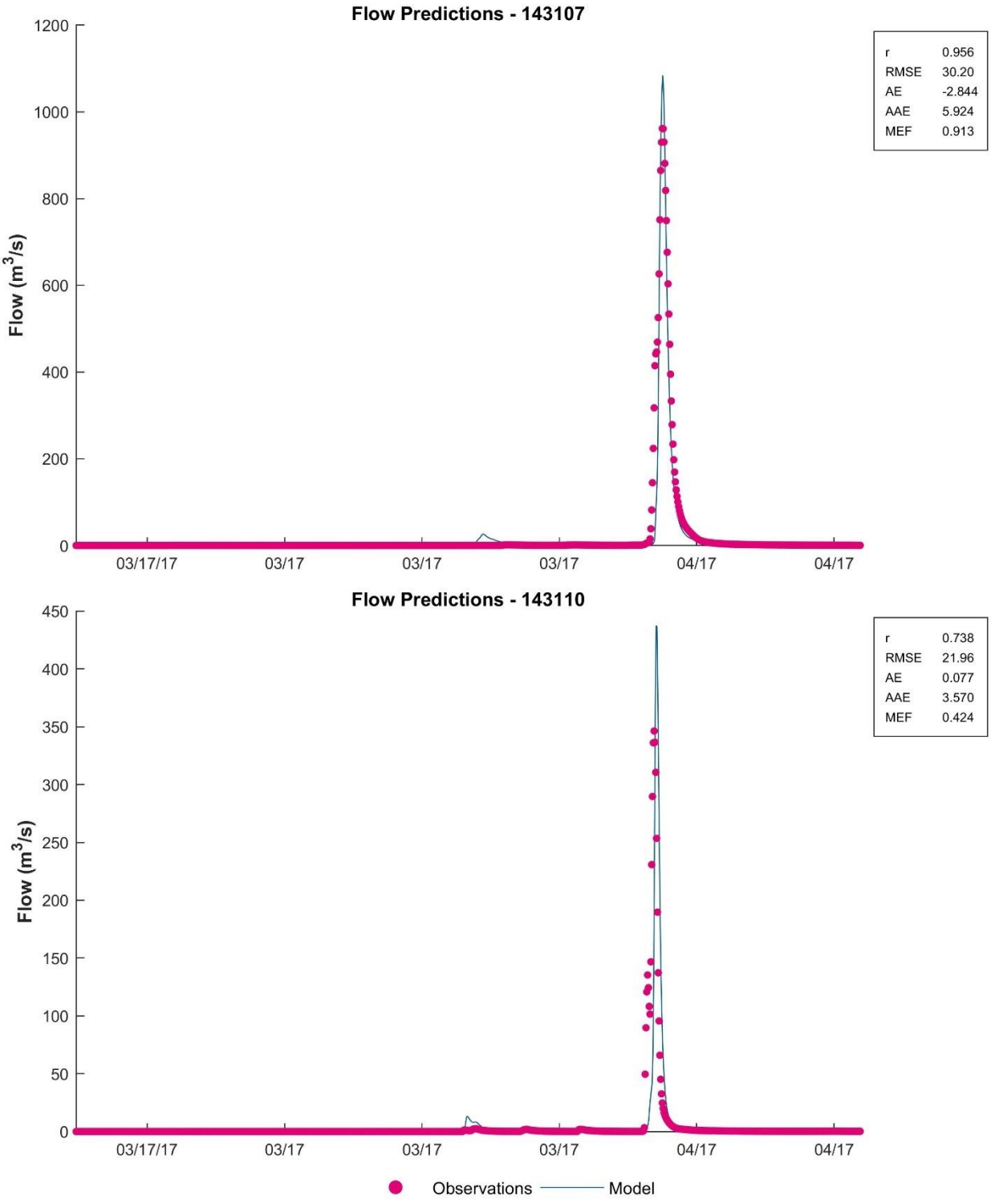


Figure 6-5 Flow comparison – 2017

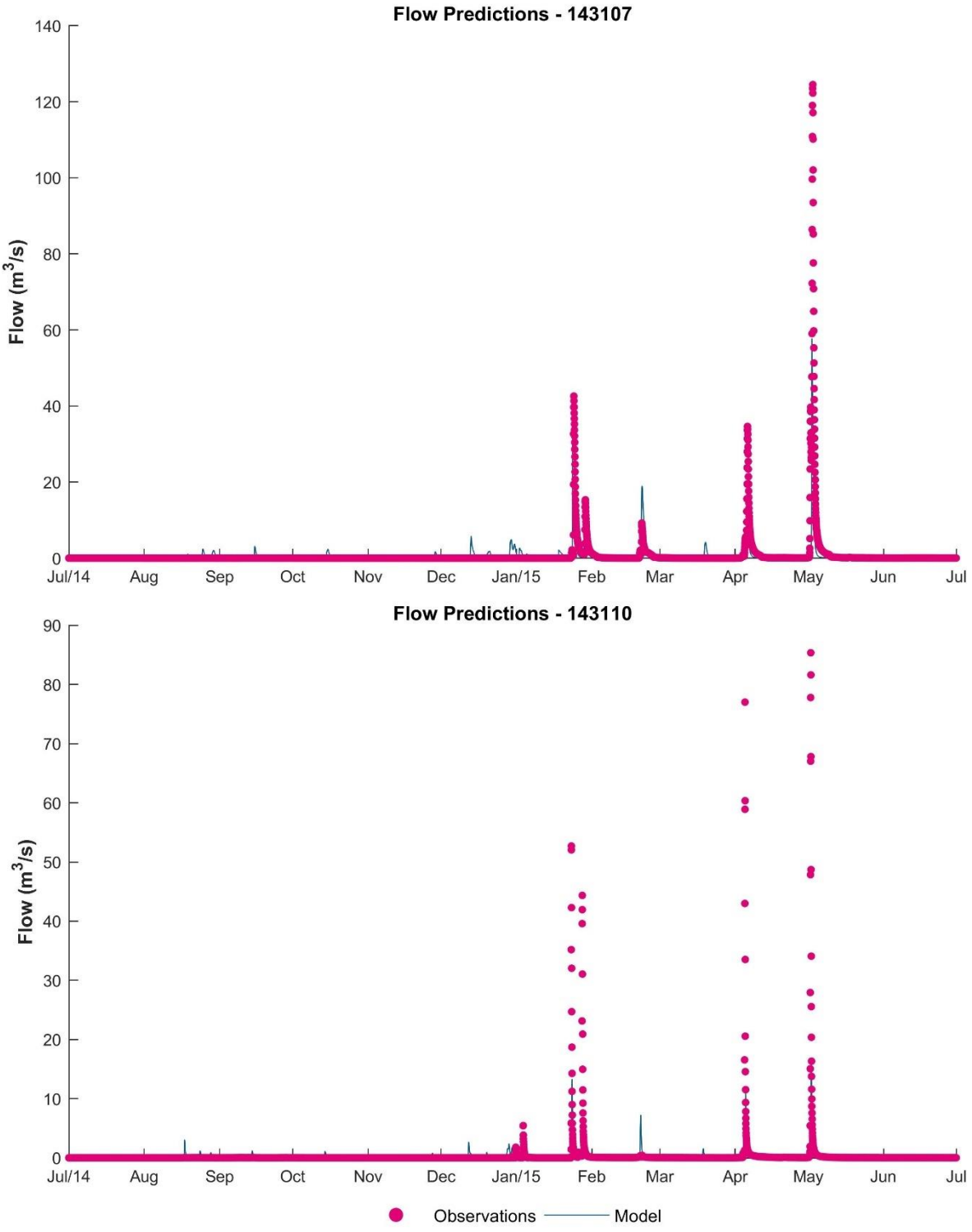


Figure 6-6 Flow comparison – 2014 to 2015

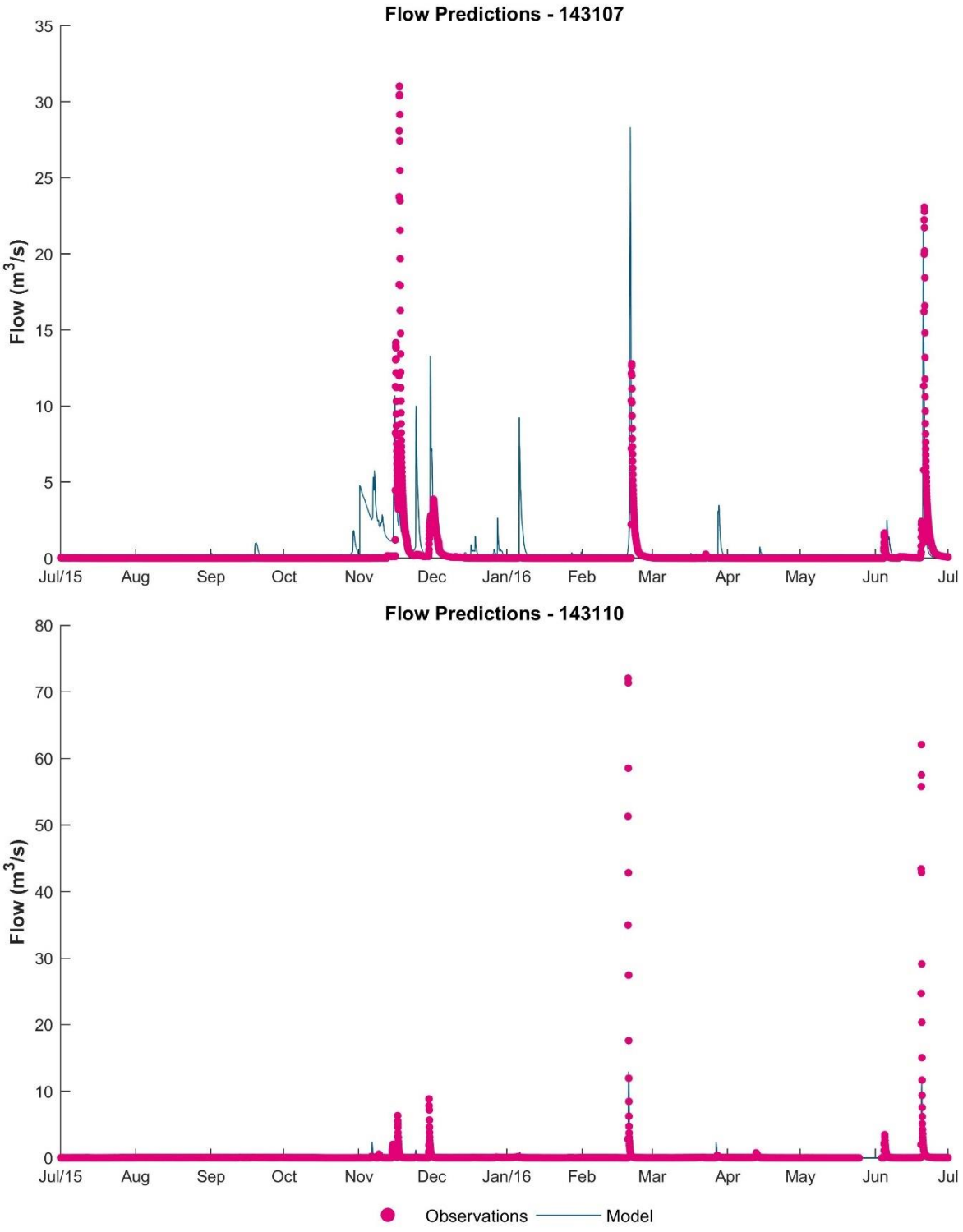


Figure 6-7 Flow comparison – 2015 to 2016

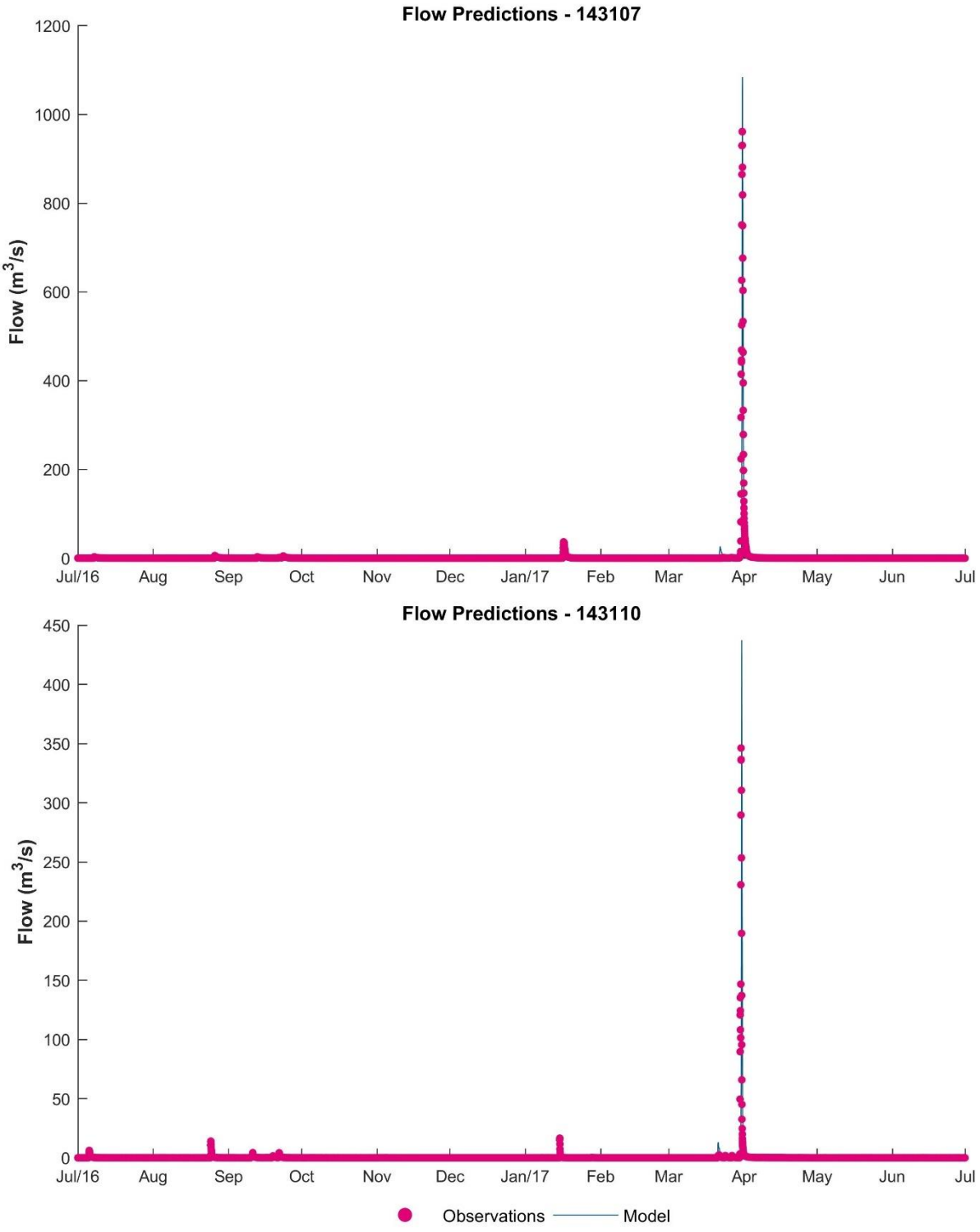


Figure 6-8 Flow comparison – 2016 to 2017

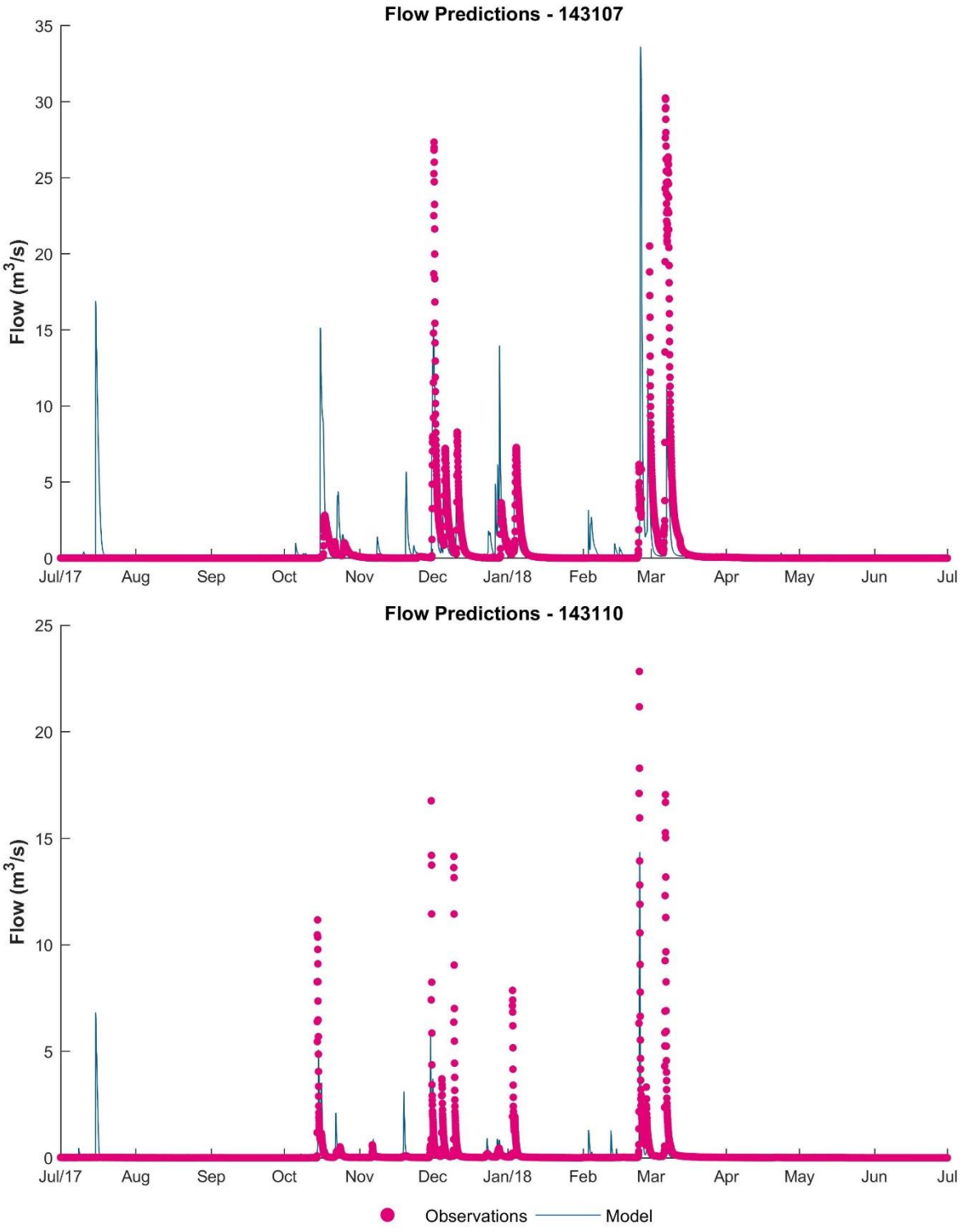


Figure 6-9 Flow comparison – 2017 to 2018

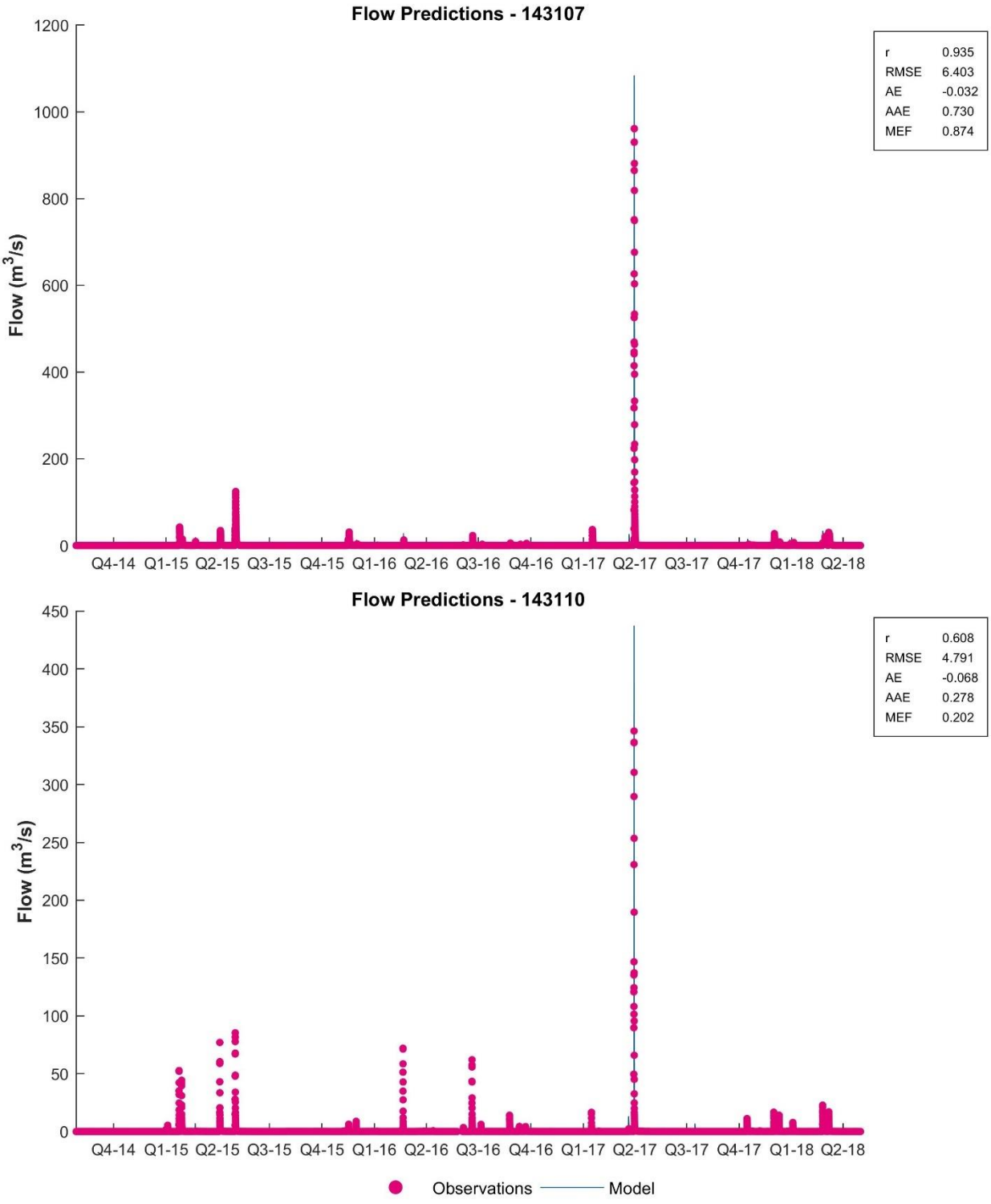


Figure 6-10 Flow comparison – 2014 to 2018 (combined with statistics)

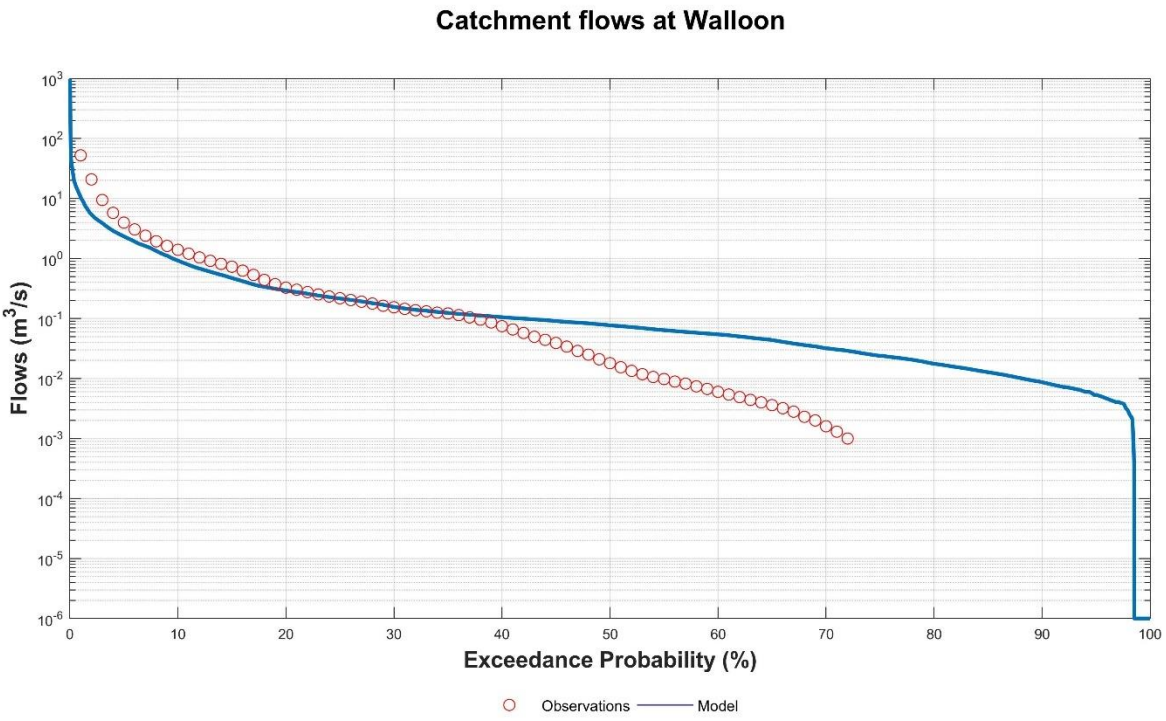


Figure 6-11 Flow Exceedance Probability at Walloon – 2014 to 2018

6.3 Additional Model Statistics

Moriasi et al.(2008) is commonly used as a benchmark to evaluate performance of catchment models. It has been recommended that common statistics like Nash-Sutcliffe efficiency (NSE), root mean square error divided by the standard deviation (RSR), the percent of bias (PBIAS) and the coefficient of linear determination (r) are calculated on monthly flows. Table 6-1 shows a comparison of the results were obtained for the model run from 2014 to 2018 against similar predictions by a SOURCE model for the same gauge.

The SOURCE model used for the comparison was developed as part of a previous modelling study for Queensland Government to set target loads for estuaries in south-east Queensland. The model was run between a period from 2014 to 2016 (BMT, 2018). The new model, in comparison to SOURCE, presented improved performance according to all Moriasi et al. (2008) statistical indicartors, particularly RSR, r and PBIAS.

Table 6-1 Summary of model monthly statistics

Statistic	Value for model predictions	Value for SOURCE predictions	Quality
RSR	0.25	0.76	As close to 0 as possible (considered 'very good' when less than 0.5)
NSE	0.94	0.92	As close to 1 as possible (considered 'very good' when greater than 0.75)
r	0.98	0.67	As close to 1 as possible
PBIAS	0.05	0.42	As low as possible (considered 'very good' when less than 0.1)

6.4 Water Quality Calibration

The water quality calibration was done over the same period as flow – 18/01/2015 to 10/05/2015. The model was run over a prolonged period covering the water quality observations. Comparisons with the machine learning methods are also shown (Figure 6-12 to Figure 6-19). Here it is noted that machine learning methods were only applied to water quality predictions and not to flow. Observations plotted in the figure show the division between the training and testing datasets.

The following statistics were examined:

- The r statistic
- Root mean squared error (RMSE)
- Average error (AE)
- Absolute average error (AAE)
- Modelling efficiency (MEF).

Univariate statistics are sensitive to phase errors and should be considered in concert with the timeseries plots for this reason. The suite of metrics should also be considered in their entirety, as some statistics may provide a misleading impression of the skill of the model. For example, a score of 1 for the r statistic indicates that the model varies perfectly in step with the observations, but it says nothing about any bias that may be present. Also, high scores for RMSE, AE and AAE may indicate a bias within the model, or may just be the result of one or two outlier observations that affect the overall score. The following provides some notes on interpreting each metric:

- r
  - Varies between -1 and 1, with a score of 1 indicating the model varies perfectly with the observations and a negative score indicating the model varies inversely with the observations. Model and observations do not need to match to provide a high score, as a consistent bias may be present.
- RMSE
  - Measures the mean magnitude, but not direction, of the difference between model data and observations. This accounts for the cancelling of positive and negative errors, but is weighted

towards large errors and is therefore sensitive to outliers. Values near zero indicate good model skill.

- AE
  - Measures the mean magnitude and direction of the difference between model data and observations, and hence can be used to measure bias. Values near zero are desirable but negative and positive errors cancel each other out so low scores can be misleading.
- AAE
  - Also measures the mean magnitude, but not direction, of the difference between model data and observations. AAE is always equal to or lower than the RMSE and the difference between the two is a measure of the variability of the errors. If the difference between AAE and RMSE is low, this indicates a consistent bias and low error variability; if the difference is large, this indicates a small number of outliers and high error variability. Values near zero indicate good model skill.
- MEF
  - Is a measure of how well a model predicts observations relative to the mean of the observations. A value near 1 suggests the model is skilful. A value near 0 suggests the model is no better at predictions than the average of the data. A value below 0 indicates that the mean of the observations would be a better predictor than the model.

All the model stats described above have been calculated for each model hybridization and separately for the training, testing and complete timeseries. A table showing all the stats for each model and each data subset is presented in Table 6-2 to Table 6-9. The plots also indicate the statistics for the whole timeseries, with the training statistics show in parentheses.

The plots have been shown below.

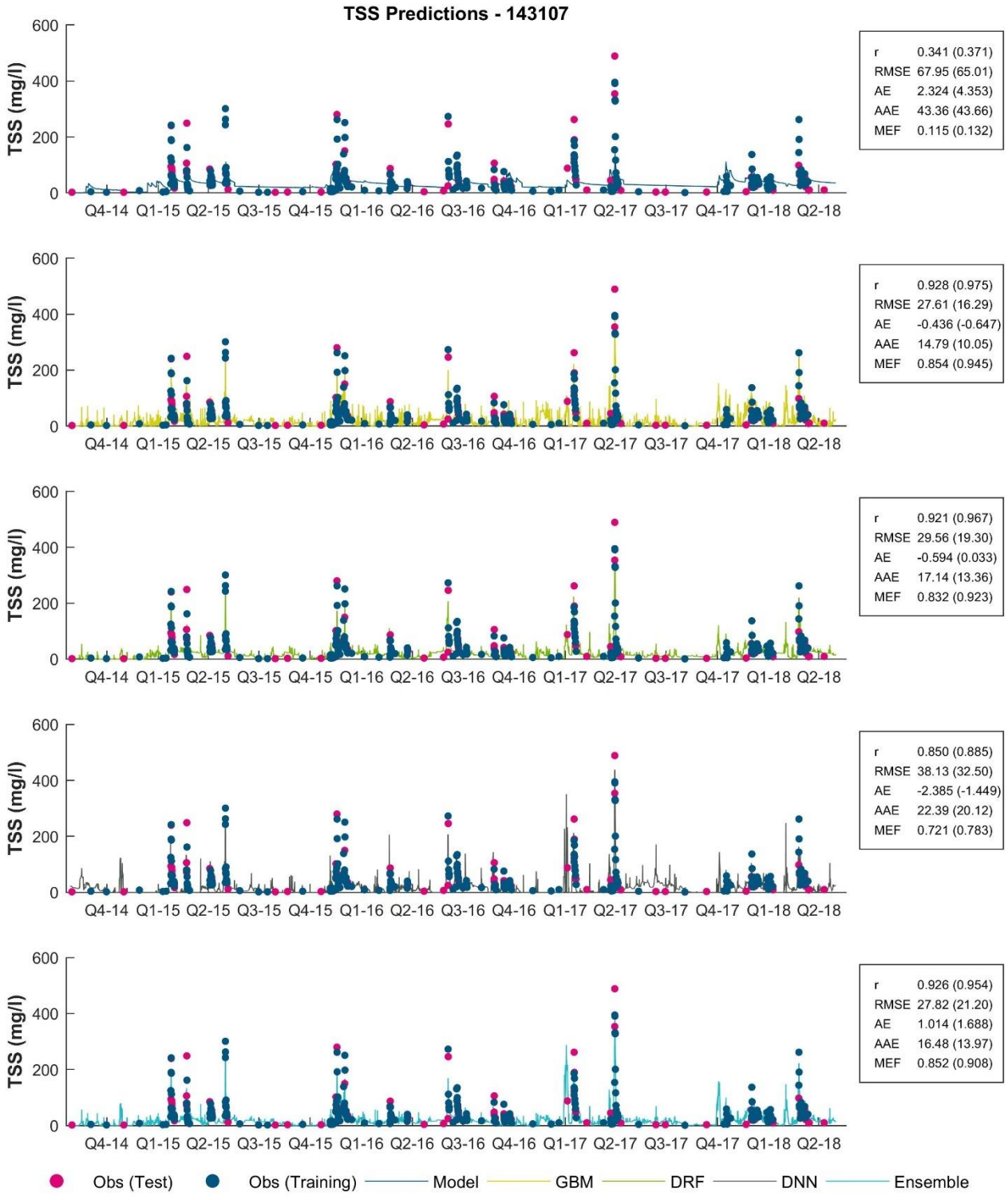


Figure 6-12 TSS comparison

Table 6-2 Summary of performance statistics for TSS predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.341
Model	RMSE	-	-	67.949
Model	AE	-	-	2.324
Model	AAE	-	-	43.364
Model	MEF	-	-	0.115
GBM	r	0.975	0.822	0.928
GBM	RMSE	16.286	45.353	27.606
GBM	AE	-0.647	0.112	-0.436
GBM	AAE	10.052	27.137	14.790
GBM	MEF	0.945	0.664	0.854
DRF	r	0.967	0.819	0.921
DRF	RMSE	19.301	46.690	29.559
DRF	AE	0.033	-2.227	-0.594
DRF	AAE	13.361	26.990	17.141
DRF	MEF	0.923	0.644	0.832
DNN	r	0.885	0.773	0.850
DNN	RMSE	32.497	49.903	38.129
DNN	AE	-1.449	-4.823	-2.385
DNN	AAE	20.116	28.310	22.388
DNN	MEF	0.783	0.593	0.721
Ens	r	0.954	0.866	0.926
Ens	RMSE	21.198	40.252	27.821
Ens	AE	1.688	-0.744	1.014
Ens	AAE	13.970	23.005	16.475
Ens	MEF	0.908	0.735	0.852

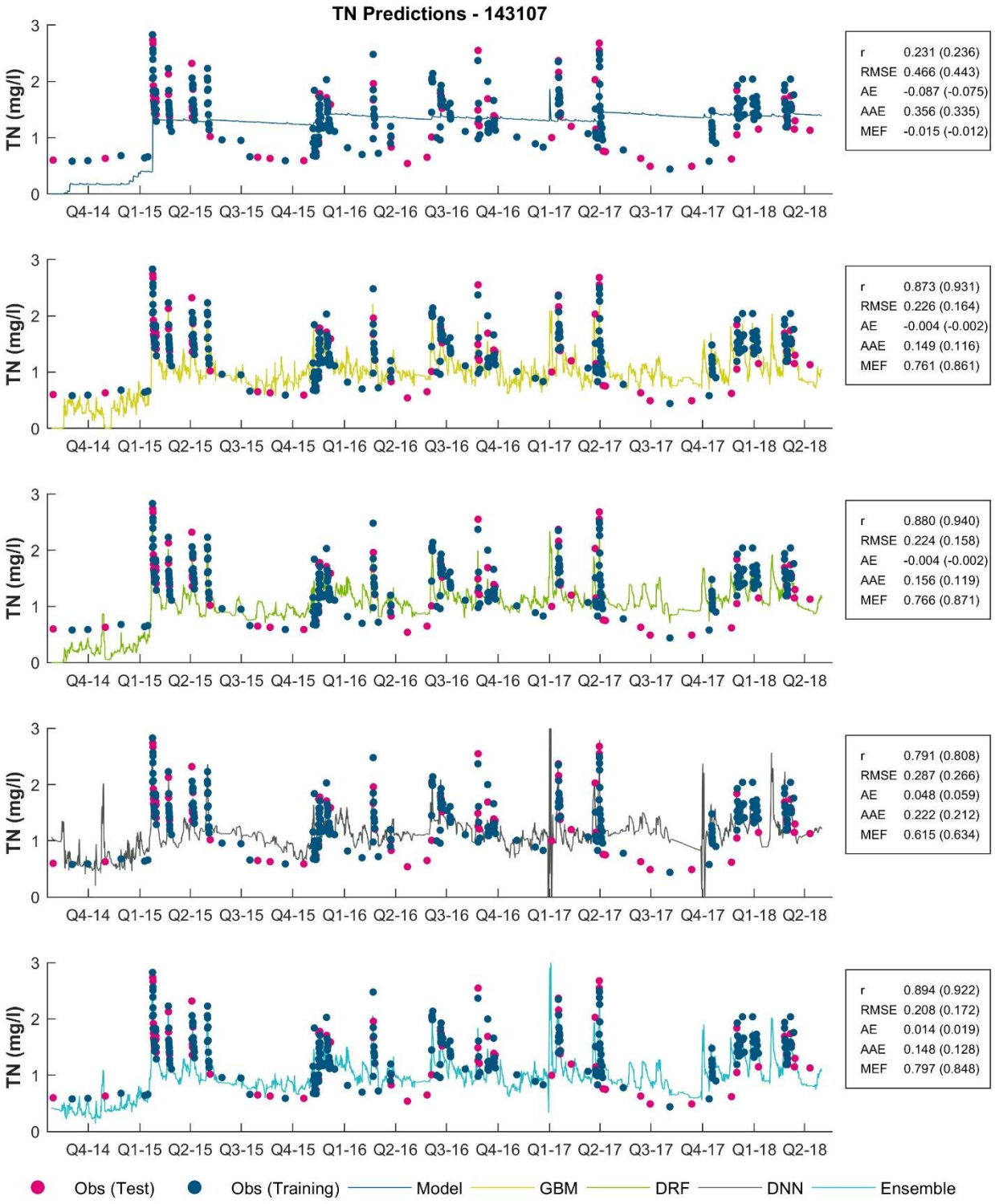


Figure 6-13 TN Comparison

Table 6-3 Summary of performance statistics for TN predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.231
Model	RMSE	-	-	0.466
Model	AE	-	-	-0.087
Model	AAE	-	-	0.356
Model	MEF	-	-	-0.015
GBM	r	0.931	0.756	0.873
GBM	RMSE	0.164	0.338	0.226
GBM	AE	-0.002	-0.007	-0.004
GBM	AAE	0.116	0.236	0.149
GBM	MEF	0.861	0.570	0.761
DRF	r	0.940	0.753	0.880
DRF	RMSE	0.158	0.339	0.224
DRF	AE	-0.002	-0.008	-0.004
DRF	AAE	0.119	0.251	0.156
DRF	MEF	0.871	0.565	0.766
DNN	r	0.808	0.760	0.791
DNN	RMSE	0.266	0.335	0.287
DNN	AE	0.059	0.019	0.048
DNN	AAE	0.212	0.246	0.222
DNN	MEF	0.634	0.576	0.615
Ens	r	0.922	0.839	0.894
Ens	RMSE	0.172	0.282	0.208
Ens	AE	0.019	0.003	0.014
Ens	AAE	0.128	0.201	0.148
Ens	MEF	0.848	0.700	0.797

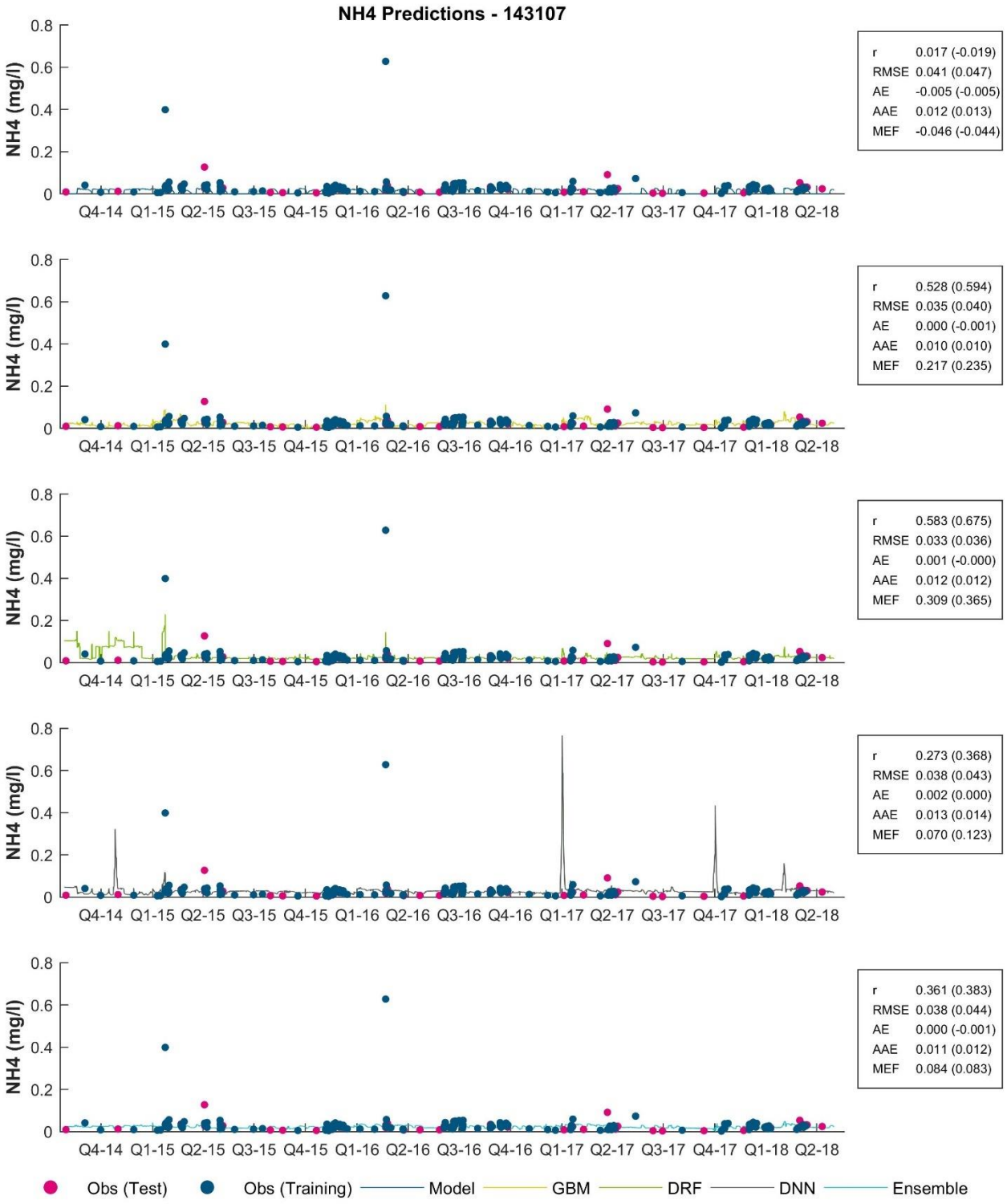


Figure 6-14 NH4 Comparison

Table 6-4 Summary of performance statistics for NH4 predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.017
Model	RMSE	-	-	0.041
Model	AE	-	-	-0.005
Model	AAE	-	-	0.012
Model	MEF	-	-	-0.046
GBM	r	0.594	0.202	0.528
GBM	RMSE	0.040	0.017	0.035
GBM	AE	-0.001	0.003	0.000
GBM	AAE	0.010	0.011	0.010
GBM	MEF	0.235	-0.183	0.217
DRF	r	0.675	-0.065	0.583
DRF	RMSE	0.036	0.022	0.033
DRF	AE	-0.0003	0.0031	0.0007
DRF	AAE	0.012	0.012	0.012
DRF	MEF	0.365	-0.886	0.309
DNN	r	0.368	-0.019	0.273
DNN	RMSE	0.043	0.023	0.038
DNN	AE	0.0001	0.0056	0.0017
DNN	AAE	0.014	0.013	0.013
DNN	MEF	0.123	-1.050	0.070
Ens	r	0.383	0.299	0.361
Ens	RMSE	0.044	0.016	0.038
Ens	AE	-0.001	0.002	0.000
Ens	AAE	0.012	0.010	0.011
Ens	MEF	0.083	0.066	0.084

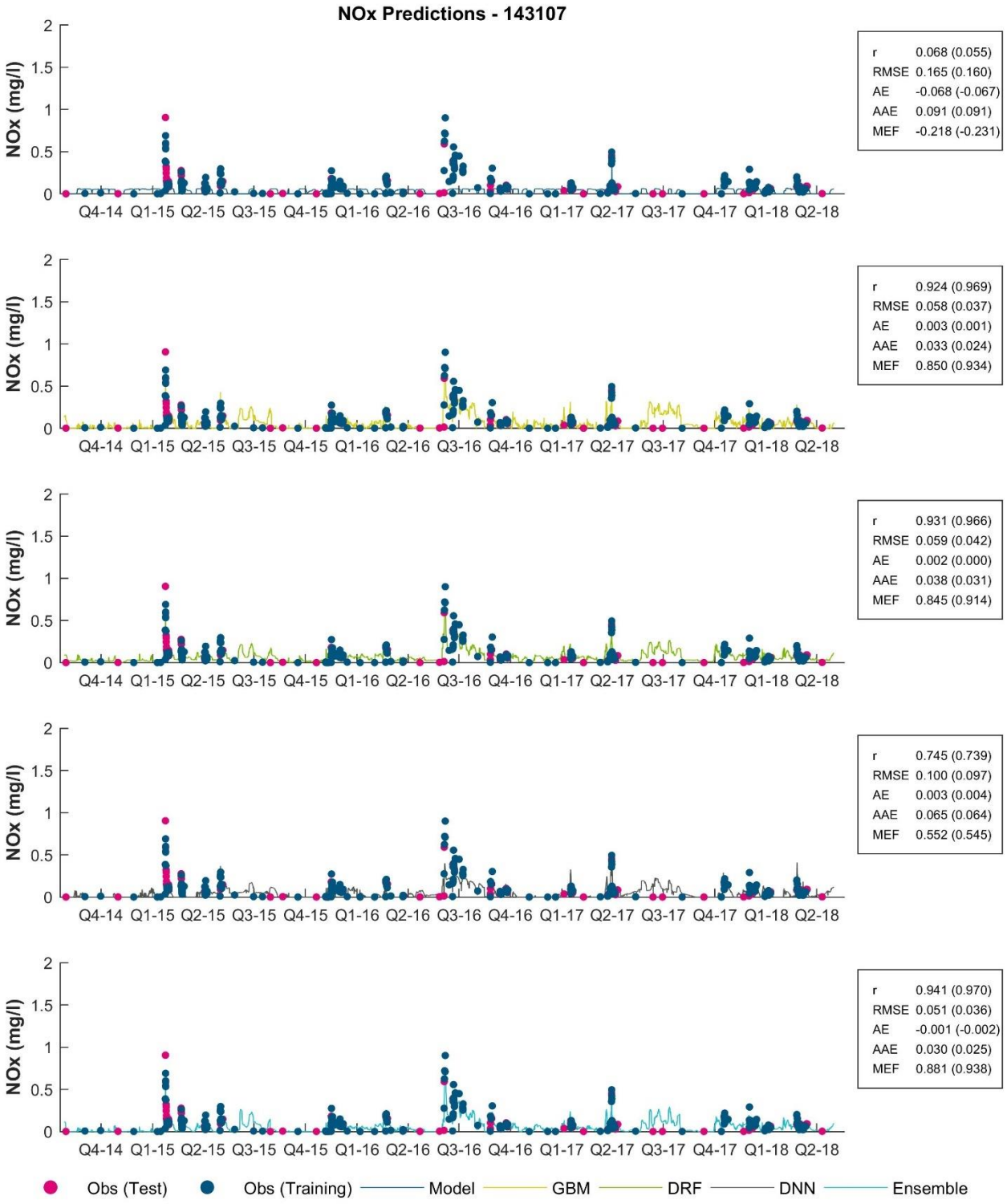


Figure 6-15 NOx Comparison

Table 6-5 Summary of performance statistics for NOx predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.068
Model	RMSE	-	-	0.165
Model	AE	-	-	-0.068
Model	AAE	-	-	0.091
Model	MEF	-	-	-0.218
GBM	r	0.969	0.824	0.924
GBM	RMSE	0.037	0.092	0.058
GBM	AE	0.001	0.009	0.003
GBM	AAE	0.024	0.055	0.033
GBM	MEF	0.934	0.673	0.850
DRF	r	0.966	0.853	0.931
DRF	RMSE	0.042	0.088	0.059
DRF	AE	0.0005	0.0075	0.0024
DRF	AAE	0.031	0.058	0.038
DRF	MEF	0.914	0.700	0.845
DNN	r	0.739	0.765	0.745
DNN	RMSE	0.097	0.106	0.100
DNN	AE	0.004	-0.001	0.003
DNN	AAE	0.064	0.069	0.065
DNN	MEF	0.545	0.565	0.552
Ens	r	0.970	0.881	0.941
Ens	RMSE	0.036	0.078	0.051
Ens	AE	-0.002	-0.001	-0.001
Ens	AAE	0.025	0.043	0.030
Ens	MEF	0.938	0.763	0.881

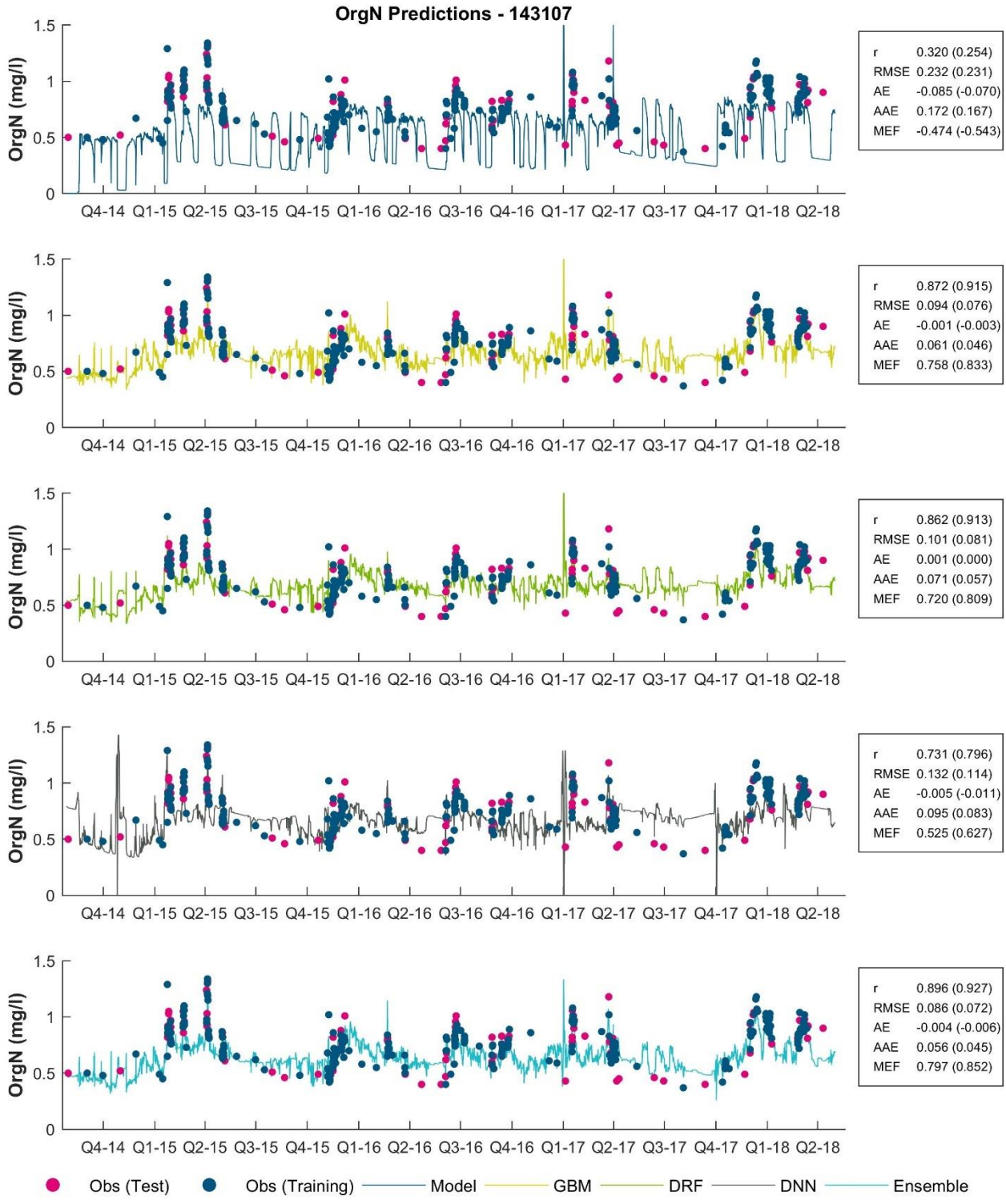


Figure 6-16 Organic Nitrogen Comparison

Table 6-6 Summary of performance statistics for Organic Nitrogen predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.320
Model	RMSE	-	-	0.232
Model	AE	-	-	-0.085
Model	AAE	-	-	0.172
Model	MEF	-	-	-0.474
GBM	r	0.915	0.773	0.872
GBM	RMSE	0.076	0.130	0.094
GBM	AE	-0.003	0.002	-0.001
GBM	AAE	0.046	0.098	0.061
GBM	MEF	0.833	0.594	0.758
DRF	r	0.913	0.740	0.862
DRF	RMSE	0.081	0.140	0.101
DRF	AE	0.000	0.002	0.001
DRF	AAE	0.057	0.108	0.071
DRF	MEF	0.809	0.527	0.720
DNN	r	0.796	0.581	0.731
DNN	RMSE	0.114	0.171	0.132
DNN	AE	-0.011	0.009	-0.005
DNN	AAE	0.083	0.128	0.095
DNN	MEF	0.627	0.301	0.525
Ens	r	0.927	0.825	0.896
Ens	RMSE	0.072	0.116	0.086
Ens	AE	-0.006	0.001	-0.004
Ens	AAE	0.045	0.084	0.056
Ens	MEF	0.852	0.677	0.797

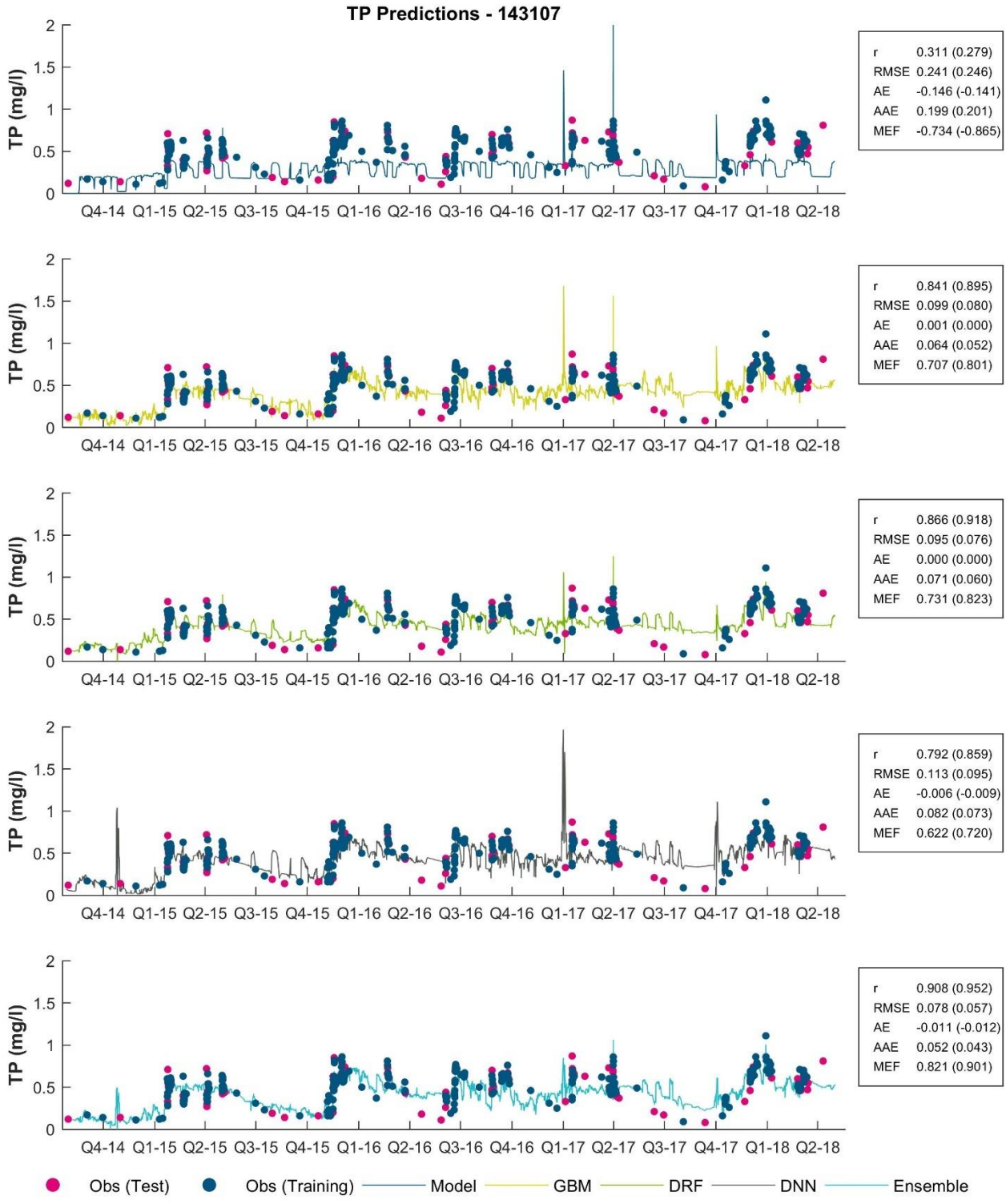


Figure 6-17 TP Comparison

Table 6-7 Summary of performance statistics for TP predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.311
Model	RMSE	-	-	0.241
Model	AE	-	-	-0.146
Model	AAE	-	-	0.199
Model	MEF	-	-	-0.734
GBM	r	0.895	0.700	0.841
GBM	RMSE	0.080	0.136	0.099
GBM	AE	0.0004	0.0044	0.0015
GBM	AAE	0.052	0.096	0.064
GBM	MEF	0.801	0.489	0.707
DRF	r	0.918	0.729	0.866
DRF	RMSE	0.076	0.133	0.095
DRF	AE	0.0001	0.0004	0.0002
DRF	AAE	0.060	0.100	0.071
DRF	MEF	0.823	0.517	0.731
DNN	r	0.859	0.630	0.792
DNN	RMSE	0.095	0.149	0.113
DNN	AE	-0.009	0.000	-0.006
DNN	AAE	0.073	0.107	0.082
DNN	MEF	0.720	0.392	0.622
Ens	r	0.952	0.798	0.908
Ens	RMSE	0.057	0.115	0.078
Ens	AE	-0.012	-0.009	-0.011
Ens	AAE	0.043	0.078	0.052
Ens	MEF	0.901	0.634	0.821

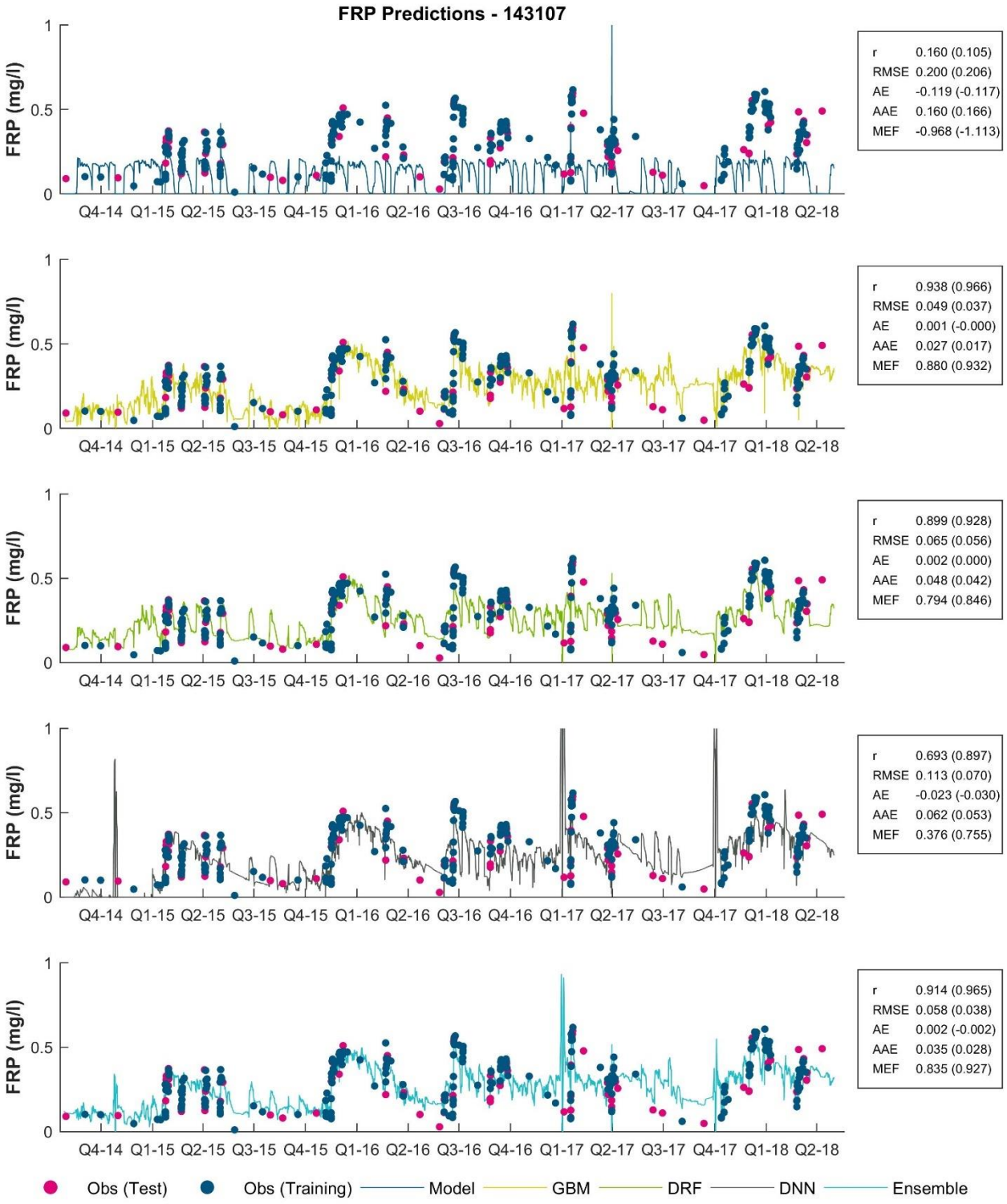


Figure 6-18 FRP Comparison

Table 6-8 Summary of performance statistics for FRP predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.160
Model	RMSE	-	-	0.200
Model	AE	-	-	-0.119
Model	AAE	-	-	0.160
Model	MEF	-	-	-0.968
GBM	r	0.966	0.863	0.938
GBM	RMSE	0.037	0.072	0.049
GBM	AE	-0.0002	0.0038	0.0009
GBM	AAE	0.017	0.051	0.027
GBM	MEF	0.932	0.743	0.880
DRF	r	0.928	0.818	0.899
DRF	RMSE	0.056	0.084	0.065
DRF	AE	0.0002	0.0075	0.0022
DRF	AAE	0.042	0.065	0.048
DRF	MEF	0.846	0.658	0.794
DNN	r	0.897	0.405	0.693
DNN	RMSE	0.070	0.181	0.113
DNN	AE	-0.030	-0.005	-0.023
DNN	AAE	0.053	0.085	0.062
DNN	MEF	0.755	-0.604	0.376
Ens	r	0.965	0.783	0.914
Ens	RMSE	0.038	0.091	0.058
Ens	AE	-0.002	0.010	0.002
Ens	AAE	0.028	0.053	0.035
Ens	MEF	0.927	0.595	0.835

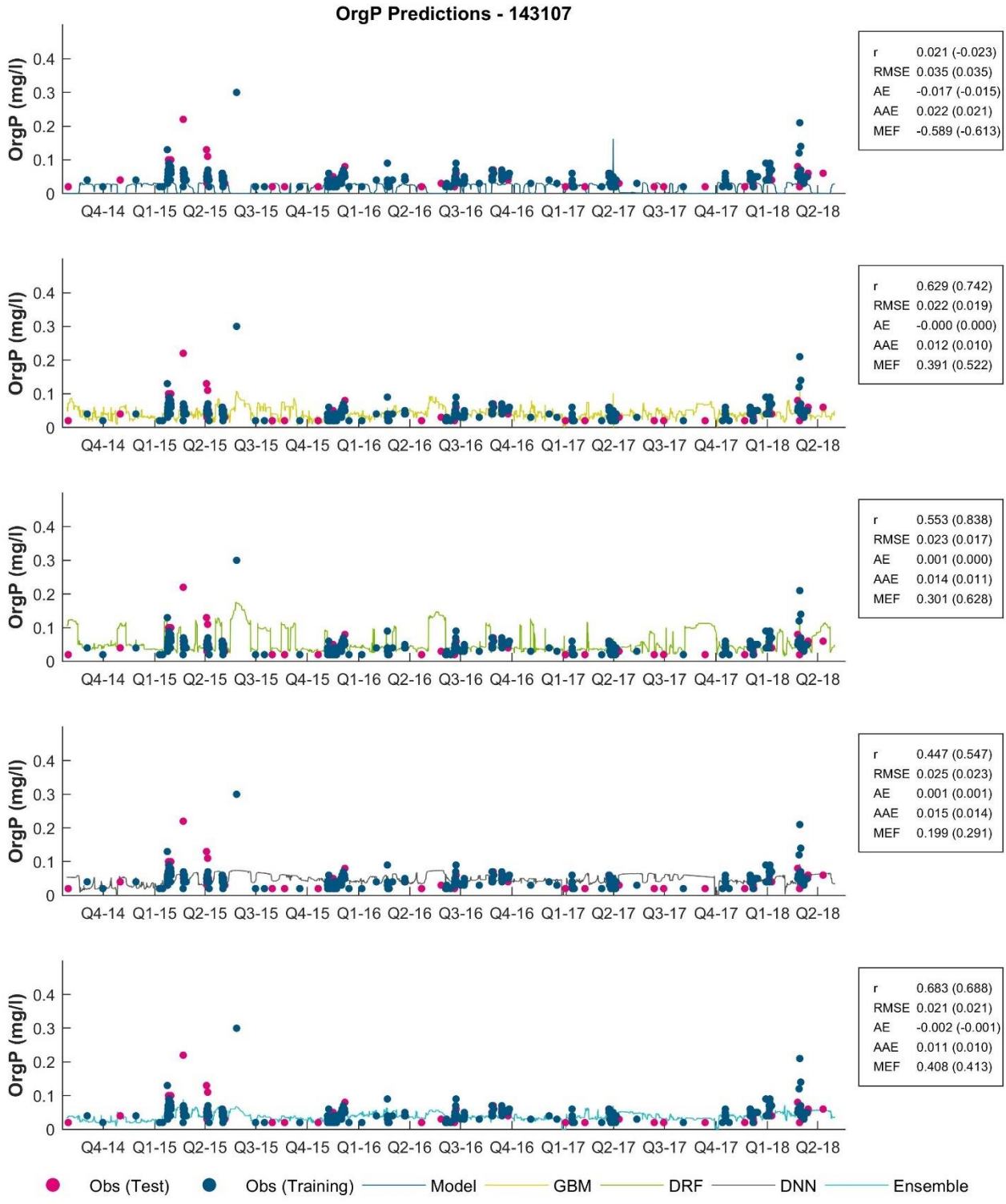


Figure 6-19 Organic Phosphorus Comparison

Table 6-9 Summary of performance statistics for Organic Phosphorus predictions

Model	Statistic	Training	Test	Complete
Model	r	-	-	0.021
Model	RMSE	-	-	0.035
Model	AE	-	-	-0.017
Model	AAE	-	-	0.022
Model	MEF	-	-	-0.589
GBM	r	0.742	0.349	0.629
GBM	RMSE	0.019	0.028	0.022
GBM	AE	0.0002	-0.0009	-0.0001
GBM	AAE	0.010	0.017	0.012
GBM	MEF	0.522	0.082	0.391
DRF	r	0.838	0.041	0.553
DRF	RMSE	0.017	0.035	0.023
DRF	AE	0.000	0.001	0.001
DRF	AAE	0.011	0.021	0.014
DRF	MEF	0.628	-0.462	0.301
DNN	r	0.547	0.206	0.447
DNN	RMSE	0.023	0.029	0.025
DNN	AE	0.001	-0.001	0.001
DNN	AAE	0.014	0.020	0.015
DNN	MEF	0.291	-0.023	0.199
Ens	r	0.688	0.667	0.683
Ens	RMSE	0.021	0.023	0.021
Ens	AE	-0.001	-0.002	-0.002
Ens	AAE	0.010	0.014	0.011
Ens	MEF	0.413	0.390	0.408

Figure 6-12 to Figure 6-19 indicate significant improvements in modelling efficiency between the parameter-optimised model output and the hybridised machine learning model. The following table summarises the improvements for each of the variables.

Table 6-10 Summary of machine learning model outputs

Variable	Best Hybrid	Improvement in MEF
TSS	GBM	0.12 to 0.85
TN	Ens	-0.02 to 0.8
NH4	DRF	-0.05 to 0.3
NOx	Ens	-0.22 to 0.88
OrgN	Ens	-0.47 to 0.8
TP	Ens	-0.73 to 0.81
FRP	GBM	-0.97 to 0.88
OrgP	Ens	-0.59 to 0.4

6.5 Parameter Identifiability

Calculations were made on the parameter identifiability and relative uncertainty reduction of each of the parameters. Different results were obtained for flow calibration and the water quality calibration.

107 of the 120 parameters related to the flow calibration were found to be identifiable and this could potentially be a result of the large amounts of hourly flow data used for the calibration.

30 of the 160 parameters were found to be identifiable as part of the water quality calibration and this could be attributed to the relatively fewer observations.

7 Discussion

7.1 Flow Predictions

The physical model developed as part of this study has used a combination of other known and commonly used conceptual models to enable discretization in both spatial and temporal scales. The results shown above clearly indicate that the combination of the distributed physical model and machine learning techniques can enhance the predictive capabilities of the model. This approach can be particularly useful when the model is being coupled with another model that has comparable spatial and temporal resolution. Enhanced discretization also enables the model to be used reliably for applications like the quantification of nutrient loads from the catchment, where temporal resolution around events is required to obtain more reliable predictions. This can be particularly useful within the context of nutrient management in coastal catchments (e.g., Great Barrier Reef, Moreton Bay).

Most of the comparisons between observed and predicted data have resulted in high (>0.8) value for MEF. This indicates a good fit especially considering that comparisons are against hourly data. A key contributor to the performance of this method is the availability of reasonably good rainfall data. The coverage of rain gauges is excellent around the catchment and this has a direct impact on model performance. Having said this, there still exists uncertainty in interpolation of this data on the model grid.

A key feature of model performance has been the resolution around small storm events. Incumbent models do not resolve these events and quite often they overestimate nutrient loads to the rivers' upper catchment. This conclusion was borne by the event-based, observed data provided by Healthy Land and Water.

7.2 Water Quality

A very simple load generation model was deployed as part of this study. This modelling approach is similar to the incumbent modelling approach and has significant potential for improvement. The load generation from the catchment by itself did not perform well against observed data. The modelling efficiency was very low and there was an over-dependence on flow as a key driver of load generation.

The parameter optimisation with PEST was done over a relatively small observation period due to runtime constraints. There could have been potential overfitting of the parameters to those data points. Total nitrogen predictions are an example of this behaviour, noting its concentration manifested as an almost constant dry weather concentration, occasionally peaking during events.

7.3 Machine Learning

Machine Learning was observed to be an effective tool in modelling some of the non-linear unexplained residuals between observed and modelled data. GBM and DRF were quite effective in modelling the residuals. Models like DNN can sometimes result in counter-intuitive model behaviour, for example, concentrations rising and falling, somewhat erratically during some flow events (Figure 6-18).

Data driven models hybridised with a physical model can significantly improve the predictive ability of the combined model. Combined model performance shows that there was a significant improvement in modelling efficiency across all water quality variables.

7.4 Future Work

The physical model has currently been prototyped in Python. While this has provided a flexible programming environment, runtimes and computing efficiency has been lagging. There is significant potential to improve computing efficiency using compiled code from C, FORTRAN etc. There is also significant opportunity in migrating some of the code base on to CUDA to exploit the efficiencies from running models on Graphical Processing Units (GPUs).

In the current setup, regionalisation of water quality parameters has been done based on catchment slope only. This approach potentially overlooks the importance of landuse in nutrient generation. A hybridised approach using parameter groups based on a mix of landuse and slope can be developed to improve model skill.

This study has demonstrated the improvement in model skill using a discretised spatial and temporal approach. Implementation of similar methods in other catchments will help build credibility and these methods can gradually replace incumbent methods.

The proposed modelling approach is likely to improve predictions for flood modelling as well. This could be another area of future work.

Interactions between surface water and ground water are still an area of research and this model can potentially be used in conjunction with MODFLOW to model the integrated system.

8 References

Barry, D. A., & Bajracharya, K. (1995). On the Muskingum-Cunge flood routing method. *Environment International*, *21*, 485–490. https://doi.org/10.1016/0160-4120(95)00046-N

BMT (2020). Strategic Review of Queensland Water Models

BMT (2018). EHP Target Loads Modelling. http://203.8.128.186/water/policy/pdf/sustainable-loads-modelling-seq-catchments.pdf

Boughton, W. C., & Askew, A. J. (1968). *Hydrological characteristics of Catchments and Lag time for natural catchments*. http://researcharchive.lincoln.ac.nz/handle/10182/5687

Boyd, M. J. (1978). A storage-routing model relating drainage basin hydrology and geomorphology. *Water Resources Research*, *14*(5), 921–928. https://doi.org/10.1029/WR014i005p00921

Breiman, L. (2017). *Classification and Regression Trees*. CRC Press. https://books.google.com.au/books?id=MGlQDwAAQBAJ

Chiew, F. H. S., & Siriwardena, L. (2020). Estimation of Simhyd parameter values for application in ungauged catchments. *MODSIM 2005 - International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, Proceedings*, 2883–2889.

Doherty, J. (1994). *PEST - Model Independent Parameter Estimation*. Watermark Numerical Computing.

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling and Software*, *18*(8–9), 761–799. https://doi.org/10.1016/S1364-8152(03)00078-1

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2008). *Model evaluation guidelines for systematic quantification of accuracy in watershed simulations*. *39*(3), 227–234. https://doi.org/10.1234/590

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. *30th International Conference on Machine Learning*, 8609–8613. https://doi.org/10.1109/ICASSP.2013.6639346

van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1). https://doi.org/https://doi.org/10.2202/1544-6115.1309

Wang, B., Hipsey, M. R., & Oldham, C. (2019). The prediction of daily surface water nutrient concentrations using a hybrid machine learning framework. *Under Review*.

Wijesiri, B., Egodawatta, P., McGree, J., & Goonetilleke, A. (2015). Process variability of pollutant build-up on urban road surfaces. *Science of the Total Environment*, *518–519*, 434–440. https://doi.org/10.1016/j.scitotenv.2015.03.014

Willems, P., Mora, D., Vansteenkiste, T., Taye, M. T., & Van Steenbergen, N. (2014). Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes - Part 2: Intercomparison of models and calibration approaches. *Journal of Hydrology*, *510*, 591–609. https://doi.org/10.1016/j.jhydrol.2014.01.028

Xu, T., Valocchi, A. J., Choi, J., & Amir, E. (2012). Improving groundwater flow model prediction using complementary data-driven models. *XIX International Conference on Computational Methods in Water Resources, Univ. of Ill., Urbana-Champaign, Ill*, *October 2015*, 1–8.

Yu, B., & Zhu, Z. (2015). *A comparative assessment of AWBM and SimHyd for forested watersheds A comparative assessment of AWBM and SimHyd for forested watersheds*. *6667*. https://doi.org/10.1080/02626667.2014.961924

Appendix A Analysis of observed data

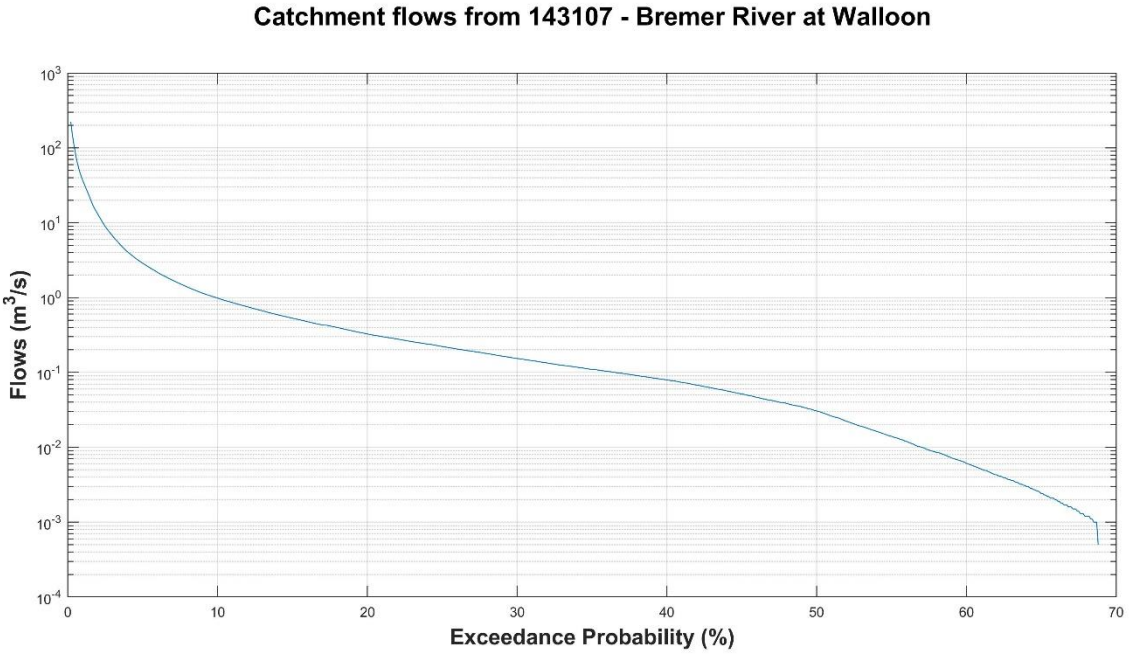


Figure A-1 Exceedance probability of gauged flow data – 143107

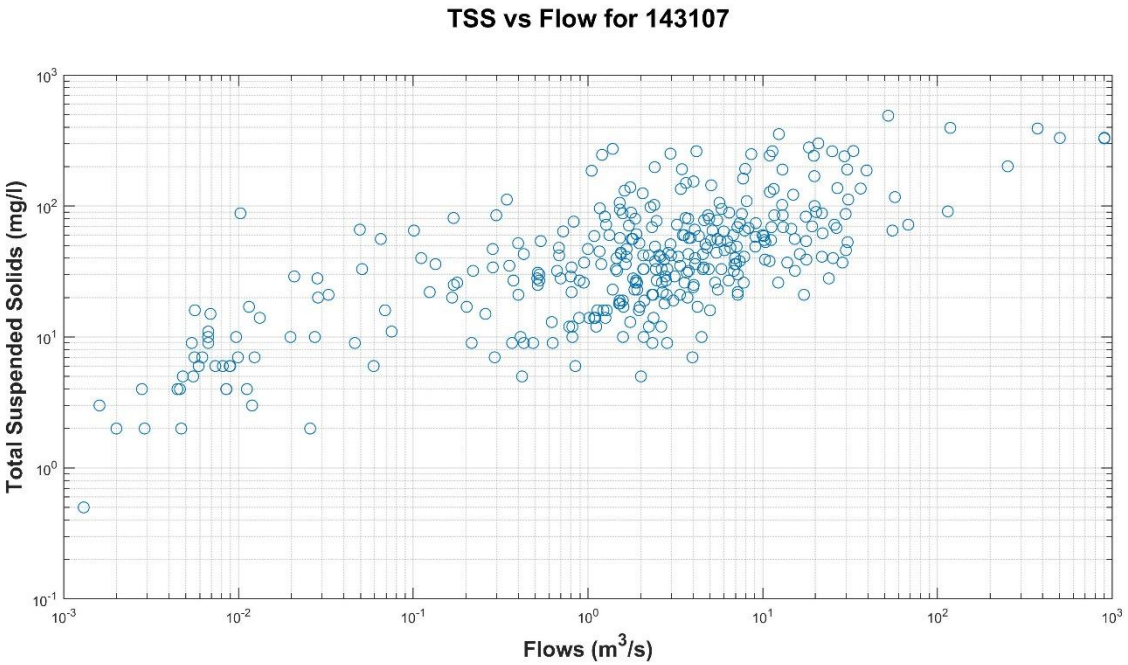


Figure A-2 Scatter plot of total suspended solids as a function of flow - 143107

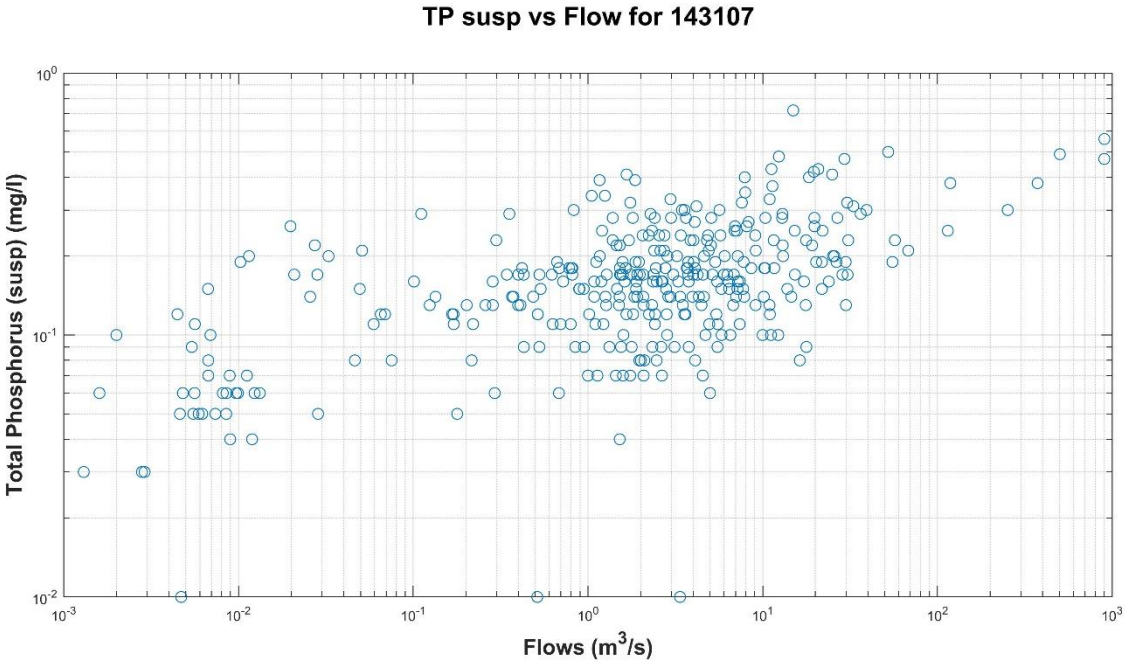


Figure A-3 Scatter plot of suspended total phosphorus as a function of flow - 143107

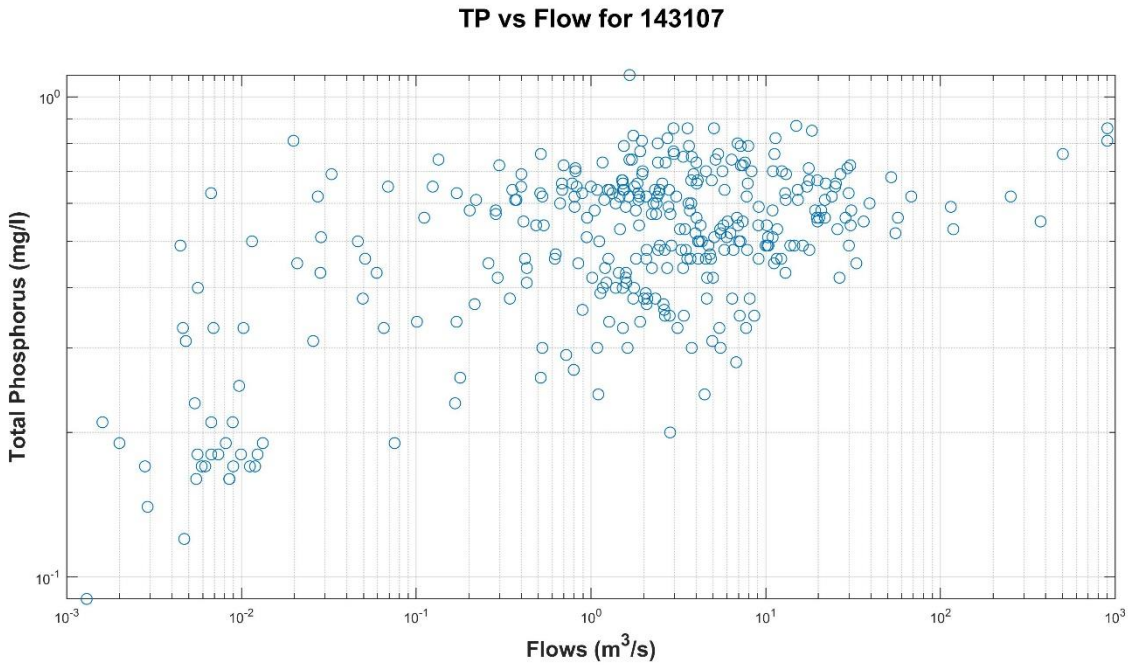


Figure A-4 Scatter plot of total phosphorus as a function of flow – 143107

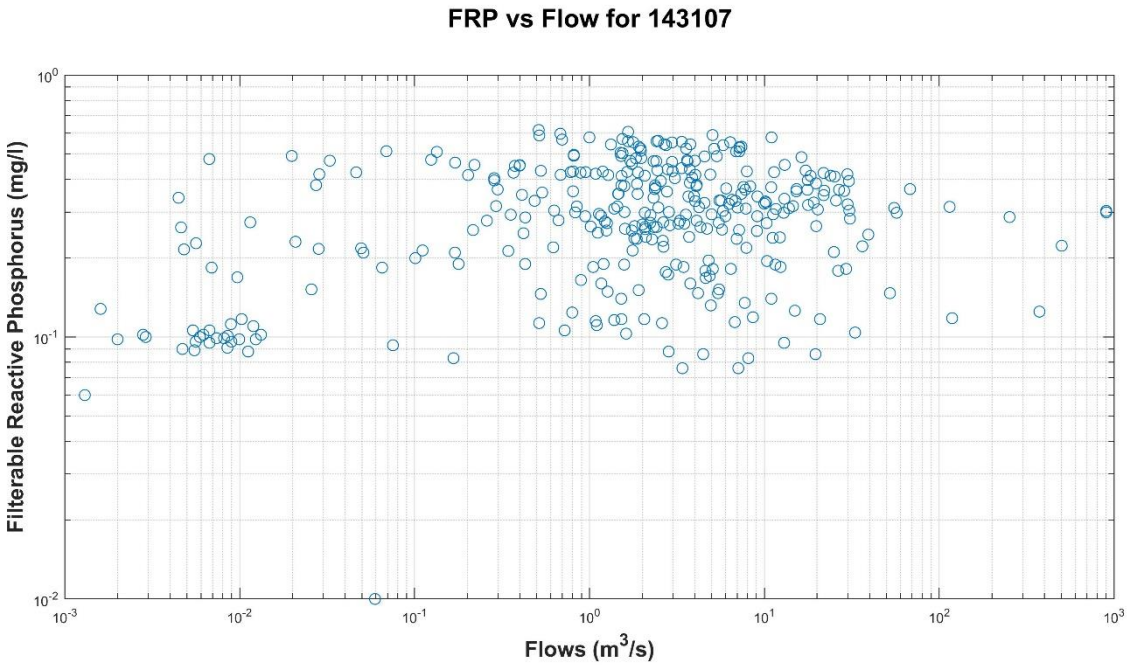


Figure A-5 Scatter plot of filterable reactive phosphorus as a function of flow – 143107

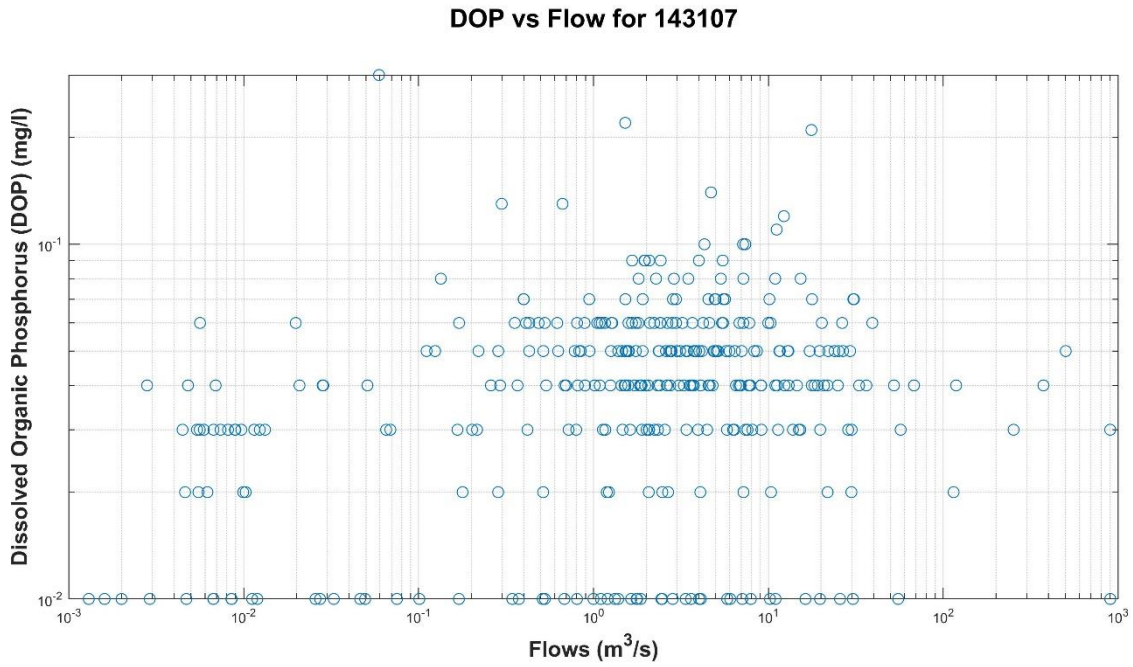


Figure A-6 Scatter plot of dissolved organic phosphorus as a function of flow – 143107

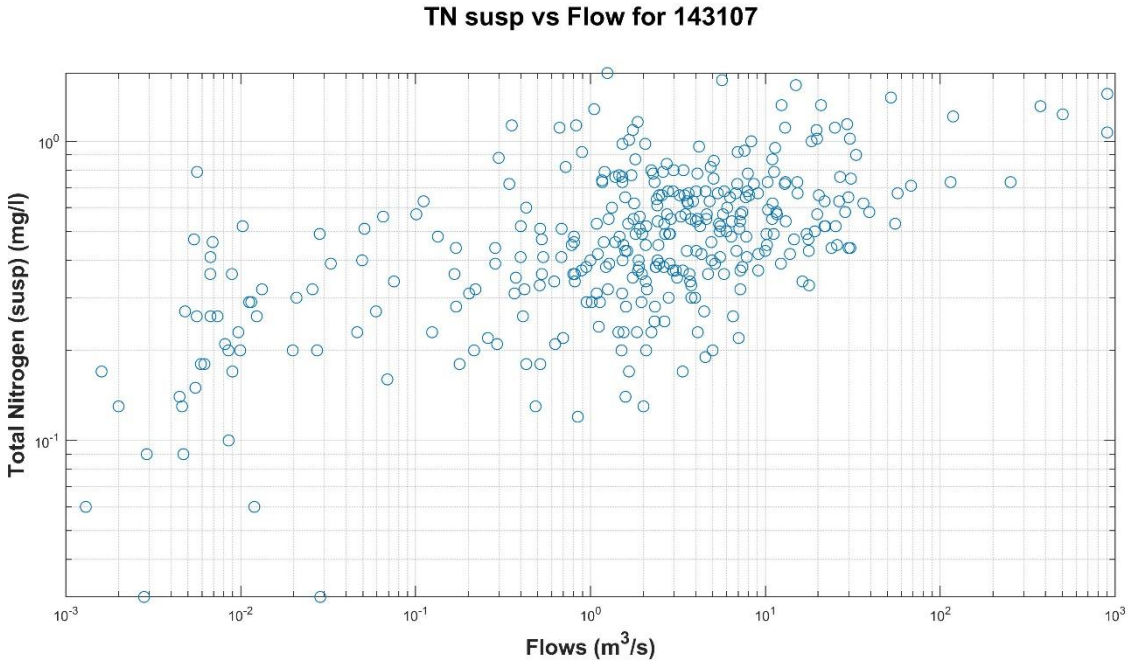


Figure A-7 Scatter plot of suspended total nitrogen as a function of flow - 143107

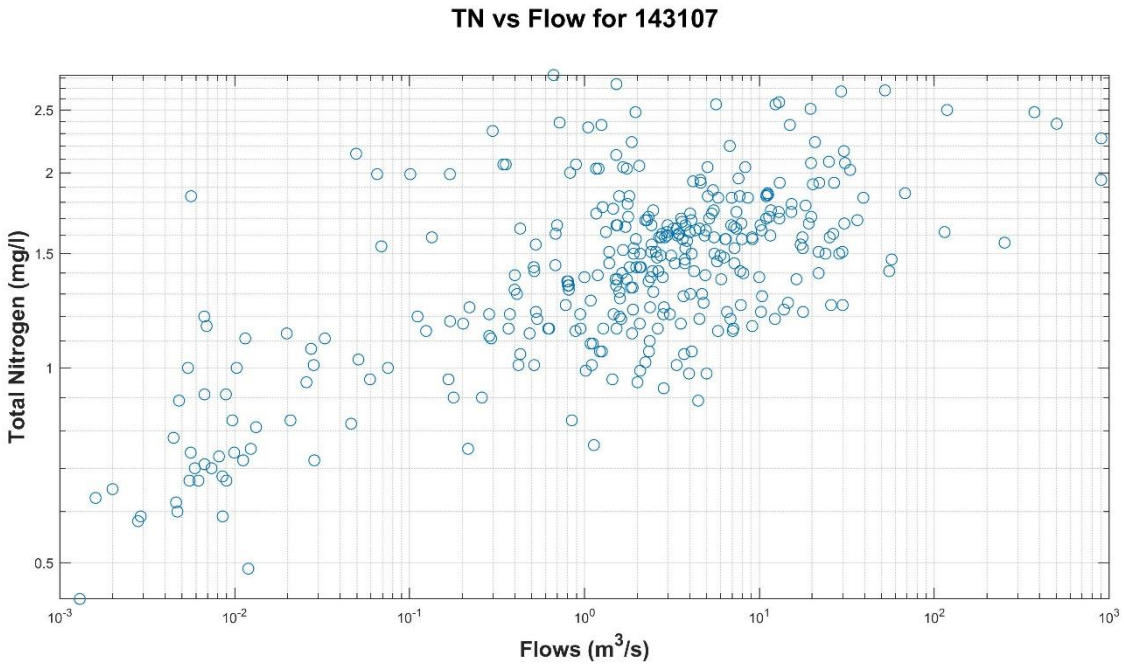


Figure A-8 Scatter plot of total nitrogen as a function of flow - 143107

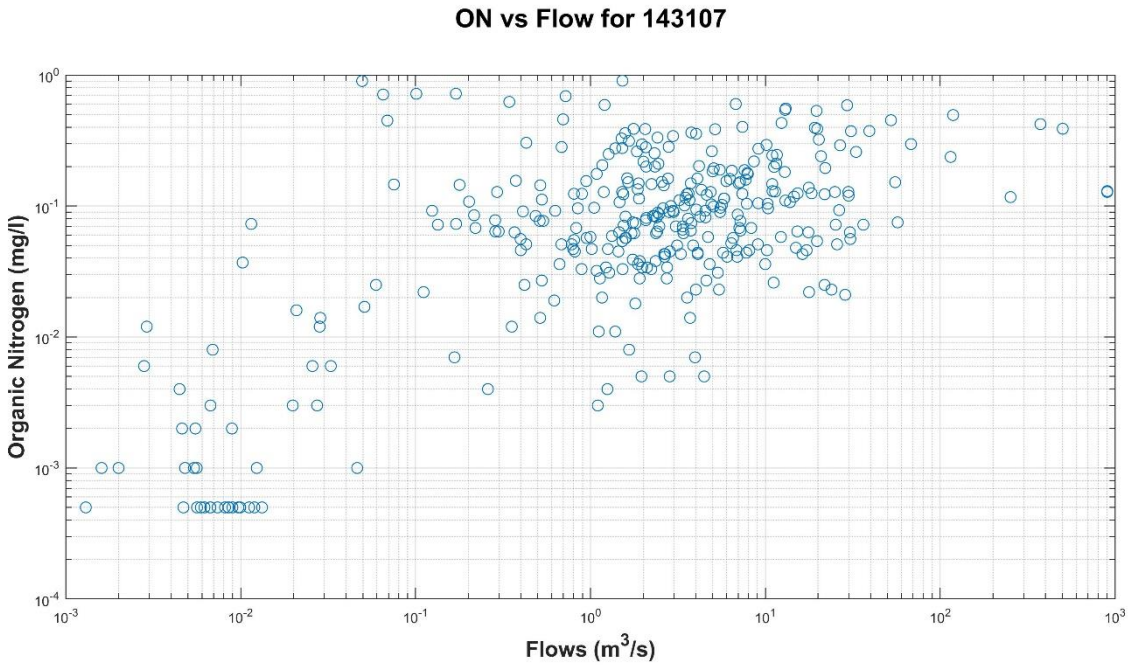


Figure A-9 Scatter plot of organic nitrogen as a function of flow - 143107

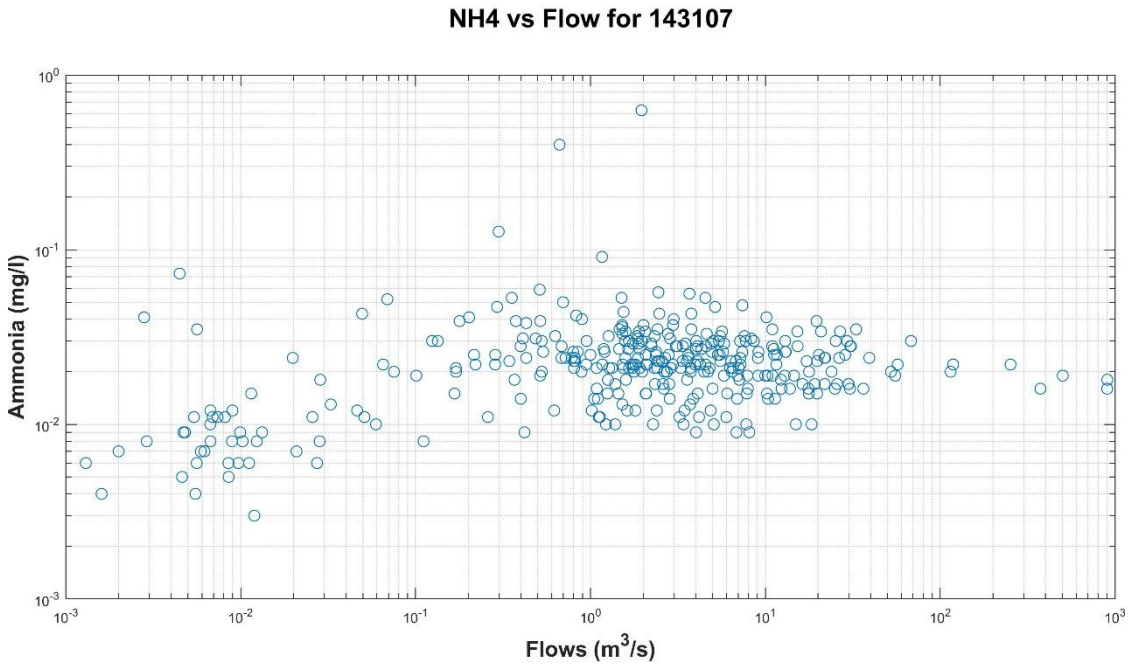


Figure A-10 Scatter plot of ammonia as a function of flow - 143107

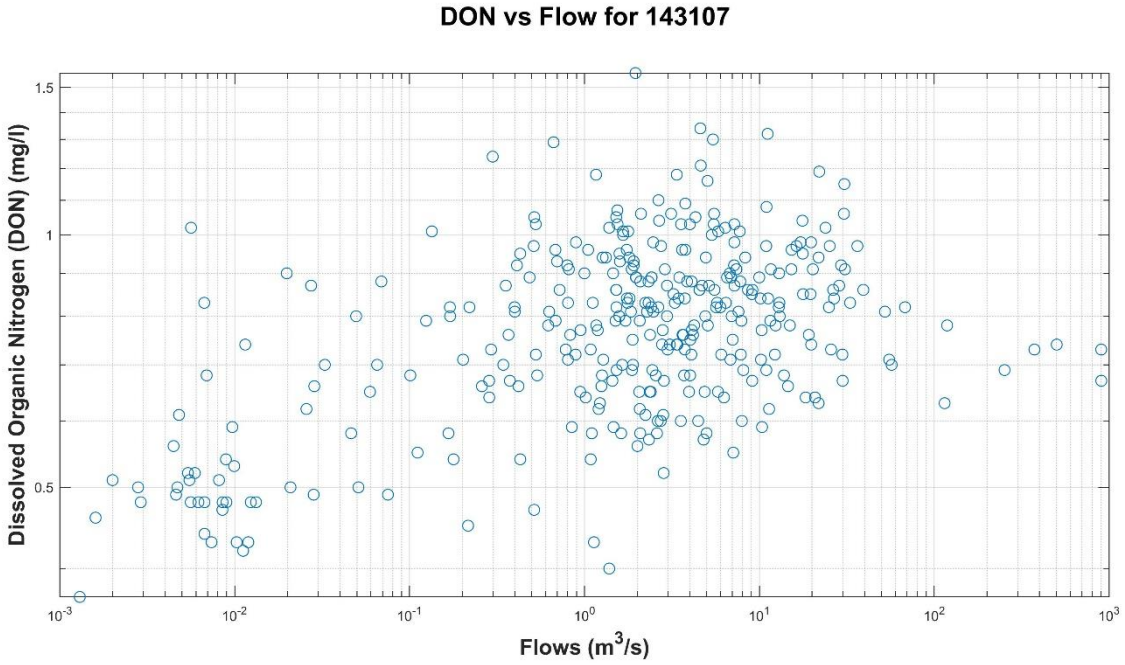


Figure A-11 Scatter plot of dissolved organic nitrogen as a function of flow – 143107

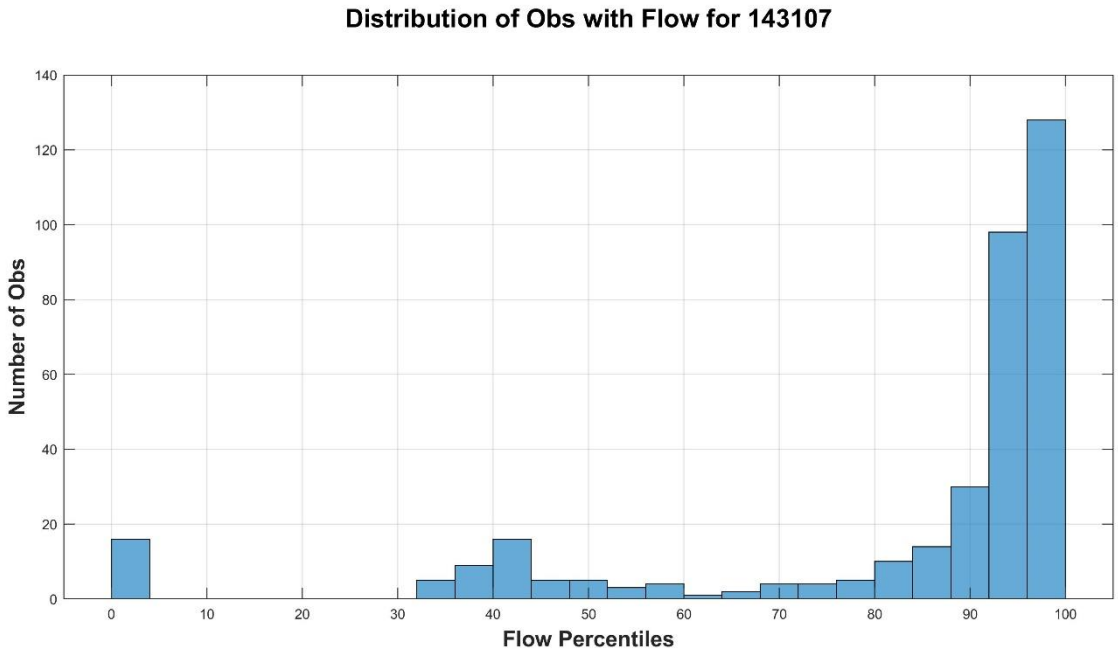


Figure A-12 Distribution of observations as a function of flow – 143107

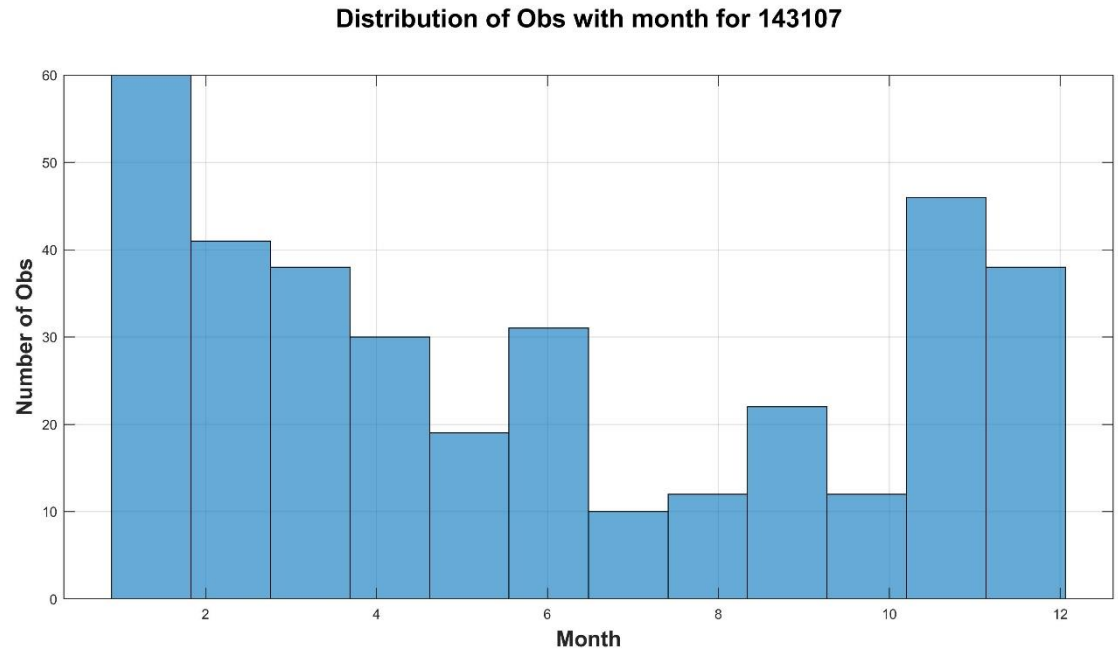


Figure A-13 Distribution of observations with month – 143107

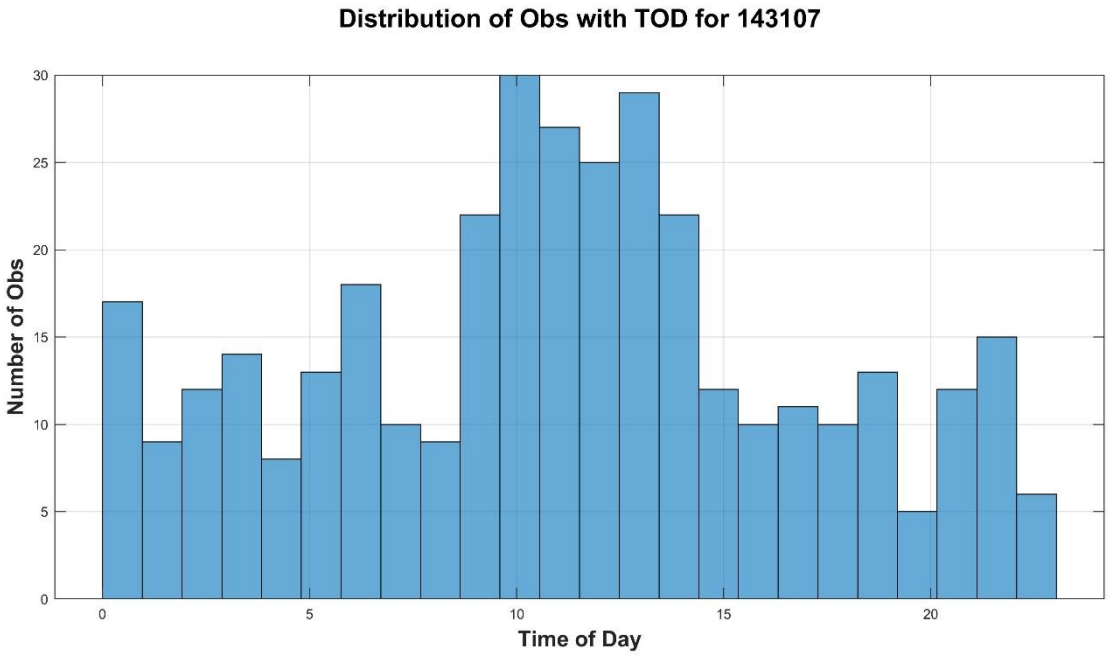


Figure A-14 Distribution of observations with time of day – 143107

Brisbane	Level 8, 200 Creek Street, Brisbane QLD 4000 PO Box 203, Spring Hill QLD 4004 Tel +61 7 3831 6744 Fax +61 7 3832 3627 Email brisbane@bmtglobal.com Web www.bmt.org
Denver	8200 S. Akron Street, #B120 Centennial, Denver Colorado 80112 USA Tel +1 303 792 9814 Fax +1 303 792 9742 Email denver@bmtglobal.com Web www.bmt.org
London	International House, 1st Floor St Katharine's Way, London E1W 1UN Tel +44 20 8090 1566 Fax +44 20 8943 5347 Email london@bmtglobal.com Web www.bmt.org
Melbourne	Level 5, 99 King Street, Melbourne 3000 Tel +61 3 8620 6100 Fax +61 3 8620 6105 Email melbourne@bmtglobal.com Web www.bmt.org
Newcastle	126 Belford Street, Broadmeadow 2292 PO Box 266, Broadmeadow NSW 2292 Tel +61 2 4940 8882 Fax +61 2 4940 8887 Email newcastle@bmtglobal.com Web www.bmt.org
Northern Rivers	5/20 Byron Street, Bangalow 2479 Tel +61 2 6687 0466 Fax +61 2 66870422 Email northernrivers@bmtglobal.com Web www.bmt.org
Perth	Level 4, 20 Parkland Road, Osborne, WA 6017 PO Box 2305, Churchlands, WA 6918 Tel +61 8 6163 4900 Email perth@bmtglobal.com Web www.bmt.org
Sydney	Suite G2, 13-15 Smail Street, Ultimo, Sydney, NSW, 2007 PO Box 1181, Broadway NSW 2007 Tel +61 2 8960 7755 Fax +61 2 8960 7745 Email sydney@bmtglobal.com Web www.bmt.org
Vancouver	Suite 401, 611 Alexander Street Vancouver, British Columbia V6A 1E1 Canada Tel +1 604 683 5777 Fax +1 604 608 3232 Email vancouver@bmtglobal.com Web www.bmt.org