Hydrology and
Earth System
Sciences
Discussions

Modelling and monitoring nutrient pollution at the large catchment scale: the implications of sampling regimes on model performance

R. Adams[1], P. F. Quinn[2], and M. J. Bowes[3]

[1]Visiting Scientist, School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK
[2]School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK
[3]Centre for Ecology and Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK

10161

Abstract

Daily and sub daily nutrient data are now becoming available to support nutrient research and which will help underpin policy making. It is vital that water quality models that utilize these high-frequency data sets are both appropriate and suitably accurate. Here we address the capability of process based models applied at larger catchment scales (size 100–500 km^2) and show what phenomena can be simulated by exploiting high frequency data for larger catchments. Hence we can suggest the dominant processes that underpin the fluxes observed in larger catchment and thus what can be simulated, and to what accuracy. Thus the implications of new sampling frequency and model structure can be addressed and the implication to catchment management is discussed. Here we show a case study using the Frome catchment (414 km^2), Dorset UK, which demonstrates:

1. The use of process based model of nutrient flow and nutrient flux (TOPCAT) for use in larger catchments.

2. Simulations of high frequency data at weekly and sub daily time steps, thus reflecting the simulations' strengths and weaknesses.

3. Cumulative distributions of observed and simulated fluxes – as an effective means of communicating the catchment dynamics in larger catchments.

1 Introduction

In recent years there has been a trend for water quality data to be collected more frequently as technology improves and monitoring agencies have become aware of the need to collect more data (Wheater and Peach, 2004; Wade et al., 2012). In the European Union member states, this has been partly driven by the introduction of the Water Framework Directive (WFD) (EC 2000/60/EC), requiring that all surface water bodies meet exacting water quality and ecological targets, and adopt river basin management

plans (Withers and Lord, 2002) forcing all dischargers to similarly adopt best practice to reduce point loads of nutrients. One consequence of improved data is that nutrient loads can be estimated with greater accuracy than was previously possible (Cassidy and Jordan, 2011), since the time intervals between water quality and flow samples have now converged to be of a similar order of magnitude. Up until a few years ago, hydro-meteorological and runoff data were typically recorded at a sub-hourly interval but water quality samples were often collected on a monthly, or at best, a weekly interval (Johnes, 2007). This led to a plethora of load estimation techniques (reviewed first by Kronvang and Bruhn, 1996, and revisited by Cassidy and Jordan, 2011) which attempted to overcome the uncertainties inherent in estimating loads from sparse concentration data where the "true" load was impossible to measure directly.

Most of the water quality models used with these observed data sets, such as INCA-N and INCA-P in the UK and EU (see Wade et al., 2002, 2006), SWAT on a global basis (Arnold et al., 1994), JAMS/J-2000 in Germany (Krause et al., 2006), and the family of DSS-based models developed in Australia: commencing with E2 (Argent et al., 2009), then WaterCAST and finally SourceCatchments (Storr et al., 2011; Bartley et al., 2012) rely on a daily simulation timestep to predict sediment and nutrient concentrations (C), and fluxes (i.e. $C \times$ daily flow). The use of a daily timestep was probably driven in part by: (i) scarcity of sub-daily data and (ii) problems with obtaining meaningful parameters for physical processes at a sub-daily timestep, for nutrient (solute) transport in particular (Ewen et al., 2000). Some models such as the spatially distributed PSYCHIC (Davison et al., 2008) operate on a monthly timestep, and PSYCHIC's strengths appear to be in the identification of flow pathways that transport sediment and P to watercourses and its ability to predict seasonally varying P loads from diffuse sources. The above models tend to be semi-distributed in space, using spatial units such as: (i) Hydrological Response Units (HRUs), in SWAT (Arnold et al., 1994) and JAMS (Krause et al., 2006); (ii) Functional Units (FUs), in E2 and subsequent developments of this model (Argent et al., 2009), or (iii) subcatchments in INCA (Wade et al., 2002, 2006), to simulate spatially varying properties. FUs and HRUs are typically related to the dominant

10163

land use with each land use type being allocated to a discrete unit in the model with set of unique input parameters.

Physically-based, distributed hydrological models (PBDHMs) that solve numerical equations of water flow and solute transport at a fine (e.g. $1\,km^2$) spatial resolution such as SHETRAN have been adapted to model solutes including nutrients on an hourly timestep (Ewen et al., 2000; Lunn et al., 1996; Birkinshaw and Ewen, 2000), the latter two describing nitrate simulation components. However, PBDHM adoption in water quality modelling has not been widespread due to model complexity, excessive numbers of parameters (see Beven, 1996, for a critique), and (at the time of their development) onerous demands on computing resources, so it has been limited to a small number of research applications. INCA-N (Wade et al., 2006) and INCA-P (Wade et al., 2002) are also physically-based but only semi-distributed, and contains many parameters to describe the processes of phosphorus (P) generation and nitrate (N) transport. In a sensitivity analysis using the GLUE framework, Dean et al. (2009) found that the INCA-P model performed barely adequately in estimating total P (TP) concentrations and poorly in reproducing total reactive phosphorus (TRP) for the Lugg catchment, UK, when based on monthly timestep data, suggesting that higher frequency P concentration data was required. INCA-N (Wade et al., 2006) was subjected to a similar analysis by McIntyre et al. (2005) who found that the model performed satisfactorily in predicting nitrate and ammonium concentrations in the Kennett (UK) catchment based on one year of data. The scale of application is also important, as temporal fluctuations in runoff and water quality observed in (nested) headwater catchments may not necessarily be observed at the outlet of the larger catchment area (Haygarth et al., 2005; Storr et al., 2011). As a rule therefore, the smaller the catchment the more detail is required in the model to define processes, but as the catchment size increases then in-stream processes associated with channel routing and the effect of point sources (especially of P) will tend to take over from nutrient generation processes in influencing the signal observed at the outlet of a large catchment (Haygarth et al., 2005).

10164

Most widely-used models such as SWAT and INCA aim to include physical and chemical processes which operate either in the soil, groundwater or in-stream portions of the catchment, but these processes cannot always be easily parameterised, and there may be issues with parameter equifinality (e.g. McIntyre et al., 2005). However, these more complex models offer the potential to replicate all the variability shown in high resolution data if available, but only after considerable expense in data collection. If the aim of the modelling study is to determine a total export of nutrients from the catchment outlet then simulating all the processes within the catchment may not be required and an export coefficient model (e.g. Johnes, 1996; Hanrahan et al., 2001) may be used. If predicting maximum concentrations are important (Bowes et al., 2009b) then process representation may be more appropriate. Hence the number of processes to be represented must be balanced against the model's goals and the data that can be used to test those models. A later publication will address the sensitivity of the model parameters and the implications to simulating future land management scenarios.

1.1 The MIR modelling approach

The Minimum Information Required (MIR) approach was developed as a response to a perceived excessive number of parameters in the established water quality and sediment transport models (Quinn et al., 1999; Quinn, 2004). MIR provides a framework for the evaluation of existing models and the selection of key generation and transport processes (e.g. nitrate leaching, Quinn et al., 1999, and sediment-attached P entrainment, Quinn et al., 2008) from these. The modelling of runoff is also kept as simple as possible to avoid excessive computation, although key runoff processes that influence nutrient and sediment are retained. By creating a meta model of more complex process based models, a minimum number of processes are retained in the model structure that are required to satisfy a model goal: in this case the simulation of catchment scale diffuse pollution. A series of simple equations are implemented in MIR models with a parsimonious number of parameters. The TOPCAT family of models (Quinn, 2004; Quinn et al., 2008) were developed using this approach to simulate various sources of

sediments and nutrients, and TOPCAT-NP (referred to subsequently as "TOPCAT") will be examined in more detail below.

1.2 High frequency monitoring data

High-frequency water quality monitoring has become achievable over the last decade, firstly with the availability of automatic water samplers, and, more recently, with the development of nutrient auto-analysers (Bowes et al., 2012; Evans and Johnes, 2004; Jordan et al., 2007, 2005; Palmer-Felgate et al., 2008; Wade et al., 2012). These highly-detailed data sets have enabled accurate estimations of nutrient export from catchments to be made for the first time, and allowed estimates of error associated with traditional monthly or weekly water quality records to be estimated (Bowes et al., 2009b; Johnes, 2007). In addition, the very high-frequency (hourly/sub-hourly) data sets have provided insights into the complex dynamics of both nutrient supply and within-channel processing. Diurnal nutrient cycling has often been observed, associated with the daily variation in P and N inputs from sewage treatment works (Palmer-Felgate et al., 2008; Wade et al., 2012). Phosphorus delivery and transport during storm events, from a combination of diffuse agricultural inputs and within-channel sediment remobilisation, can produce extremely complex hysteresis patterns which are often related to the size of the storm, the season, and the antecedent flow conditions (Bowes et al., 2009b; Wade et al., 2012). Such data greatly increase our understanding of catchment nutrient sources and behaviour, and provide a valuable resource for event scale and seasonal modelling.

The key aim of this paper is to assess the value of high frequency data in water quality models at larger scales. The assessment of the value of collecting high frequency nutrient data are important because there is a significant cost overhead in both in-situ (bankside – viz. Cassidy and Jordan, 2011; Wade et al., 2012) and traditional laboratory analysis, compared to the automated collection of high frequency runoff (or water level) data. The intended use(s) of the models themselves must also be considered.

Two examples are given here. Firstly, if models are to be used to examine the impact of policy changes such as removing P in wastewater treatment plant (WWTP) discharges (e.g. Hanrahan et al., 2001; Bowes et al., 2010), then daily or weekly monitoring may be sufficient since the load of soluble reactive P (SRP) in the WWTP discharge only change slightly from day to day. Secondly, high frequency nutrient time series data can be used to improve the process representation in the models such as when investigating the role of storm events in transporting P to watercourses from diffuse sources (Haygarth et al., 2005; Sharpley et al., 2008) where concentrations changed by an order of magnitude during the course of an event. A failure to simulate this event behaviour could under predict the load estimates made from the data (Sharpley et al., 2008) and/or fail to predict the high concentrations that impact on aquatic ecosystems (e.g. causing algal blooms to develop; Bowes et al., 2009b). However, it may only be possible to detect these processes at certain scales, usually smaller research studies (plot to small catchment scale). At the larger scale, the effects of mixing and randomly generated data spikes created by a local activity may be more prominent. The component processes that generate the fluxes are still present in the larger scale catchments but may be more difficult to detect from the observed data alone, regardless of the sampling frequency.

2 Methods

A case study based on an extensively-monitored large sized catchment in southwest England is used to investigate the ability of models to simulate high resolution data sets. For simplicity, a single parsimonious model (TOPCAT; Quinn, 2004; Quinn et al., 2008) is evaluated below. The MIR structure of this model lends itself to catchment applications as the simple model structure allows modifications to be made improve simulations by adding or removing processes as required. However, this adaptation does presuppose that the model is being modified to fit the observed data.

2.1 Study area

The 414.4 km^2 Frome catchment (Fig. 1) flows into Poole Harbour with its headwaters in the North Dorset Downs (Bowes et al., 2011; Marsh and Hannaford, 2008; Hanrahan et al., 2001). Nearly 50 % of the catchment area is underlain by permeable Chalk bedrock, the remainder consists of sedimentary formations either older than the chalk or more recent such as tertiary deposits along the valleys of the principal watercourses (including sand, clay and gravels). There are some areas of clay soils in the lower portion of the catchment. However, most of the soils overlaying the chalk bedrock are shallow and well drained. The land use breakdown is dominated by improved grassland (ca. 37 %, comprising hay meadows, areas grazed by livestock and areas cut for garden turf production), and ca. 47 % tilled (i.e. arable crops primarily cereals) usage (Hanrahan et al., 2001; Marsh and Hannaford, 2008).

The mean annual catchment rainfall from 1965 to 2005 was 1020 mm and mean runoff 487 mm (Marsh and Hannaford, 2008). The major urban area in the catchment is the town of Dorchester (2006 population over 26 000, Bowes et al., 2009b) otherwise the catchment is predominantly rural in nature. At East Stoke the UK Environment Agency (EA) has recorded flows since 1965. The Centre for Ecology and Hydrology (CEH) and Freshwater Biological Association have collected water quality samples at this same location at a weekly interval from 1965 until 2009 (Fig. 1) (Bowes et al., 2011). Hanrahan et al. (2001) presented both export coefficients for diffuse sources of TP, and load estimates for diffuse and point sources (comprising: WWTPs (serving Dorchester plus other towns); septic systems; and animal wastes). The total annual TP export from diffuse sources in the catchment was estimated to be 16.4 t P yr^{-1}, a yield of 0.4 kg P ha^{-1} yr^{-1}. Point source loads from WWTPs, septic systems and animals added an extra 11.5 t P yr^{-1} (from the data in Table 2 in Hanrahan et al., 2001) to the catchment export, giving a total load of 27.9 t P yr^{-1}.

The annual average nitrate-N concentration of the River Frome has steadily increased by between 0.08 and 0.11 mg yr^{-1} since the 1940s (Bowes et al., 2011; Casey

and Clarke, 1979; Casey et al., 1993; Howden and Burt, 2009), with groundwater across the catchment showing similar rises (Smith et al., 2010). This has been attributed to historic fertiliser application. Reductions in fertiliser application rates in the mid-1980s have led to a possible levelling off of nitrate increases in both river and groundwater in the late 2000s.

A report by the Environment Agency from their "Making Information available for Integrated Catchment Management" project (EA, 2007) provided spatial predictions of N in addition to diffuse P and sediment yield, on a 1 km grid covering the entire catchment using the models: PSYCHIC (for P) and NEAPN (for N; EA, 2007). Based on these predictions, N export varied from 0 to 63.4 kg ha^{-1} yr^{-1} (similar to the figure quoted above from Bowes et al., 2009b), and TP export varied from 0 to 2 kg ha^{-1} yr^{-1} P (which is lower than the range of TP export coefficients quoted in Hanrahan et al., 2001 for their baseline land use and management scenario). In March 2002, tertiary treatment of raw sewage to remove phosphorus by precipitation with iron chloride (referred to subsequently as "P-stripping") was implemented at Dorchester WWTP, by which time the total population served by the plant had increased to 24 200 (Bowes et al., 2009b). There was also a decline in SRP concentrations at East Stoke in early 2001 probably due to a Foot and Mouth livestock disease outbreak in the region. This outbreak led to a decline in the number of livestock in the catchment (Bowes et al., 2009b). This study also estimated a 52 % reduction in soluble (filterable) reactive P (SRP) point source load following the introduction of P-stripping at Dorchester WWTP, using Load Apportionment modelling.

2.1.1 Hydrological data

Forcing data (precipitation) was supplied by the EA (P. Hulme, personal communication, 2007) for the period 1997 to 2006 which was therefore chosen as the modelling period. Daily mean flow was also provided from East Stoke gauging station for the same time period. Potential Evapotranspiration (PET) was derived using an algorithm developed to estimate daily PET based on monthly temperature patterns, to estimate

a daily PET which when totalled for the year would match the known annual PET. The annual PET used in the model was 930 mm.

Daily rain gauge data was obtained from Kingston Maurwood (ST718912) located ca. 4 km downstream of Dorchester. Earlier studies have noted some spatial variation in precipitation across the catchment (Bowes et al., 2011), and Smith et al. (2010) reported that between 1993 and 2008 there were 3–5 gauges operational in the catchment. Therefore, model errors sourced from rainfall are likely to be significant and may influence predictions of surface runoff (where rainfall is an important factor) and the associated nutrient transport by this pathway. Variability in true rainfall patterns is a major concern for larger catchment scale studies and the impact on higher frequency data sets must be accounted for.

2.1.2 Water quality datasets

Two water quality data sets were used in this study (Table 1 below shows the statistics relating to long term concentrations).

(1) The CEH/Freshwater Biological Association long-term dataset (LTD) of water quality for the River Frome (Bowes et al., 2011; Casey, 1975; open access via https://gateway.ceh.ac.uk) which was collected from 1965 to 2009 at a weekly interval (with some gaps) and thus represents one of the longest (relatively) high frequency datasets on water quality in existence from the UK. In this study we used nitrate-N (nitrate), total phosphorus (TP) and soluble reactive phosphorus (SRP) data from 1997 to 2006. Between November 2004 and March 2006 there was a gap in the TP and SRP datasets. The LTD dataset was subdivided in our analysis with the SRP data split into pre and post March 2002 periods (with the time series analysis curtailed at 31 December 2001; Table 1) to coincide with the introduction of P-stripping measures at Dorchester WWTP (see above for a full description of these).

(2) Bowes et al. (2009b) describes a second high frequency data set (HFD) also collected at East Stoke gauging station between 1 February 2005 and 31 January 2006, using a stratified sampling approach using an automatic sampler. The entire dataset

which spans a slightly longer period (extending back to early 2004 for TP) was made available for this study (Table 1). The frequency of the water samples varied between two to four times daily during dry period with up to eight samples per day during rainfall events. The average number of samples was 3.7 per day. Also in the dataset were river flow values taken from the 15 min interval gauging data. In this study we used the TON, TP and SRP data. It is assumed that nitrite concentrations are negligible (Bowes et al., 2011) so TON data can be directly compared against both modelled nitrate and observed LTD weekly nitrate data. Moreover, ammonium concentrations from the long-term LTD dataset were less than 2 % of total nitrogen values from the same time series indicating that nitrate is the dominant form of inorganic N in the Frome (Bowes et al., 2011).

2.2 Model description

2.2.1 Hydrology

The modified version of TOPCAT used here to simulate runoff and nutrient generation in the catchment is based on the original TOPCAT model (Quinn, 2004). Here, "runoff" is actually specific discharge which simply refers to flow divided by catchment area. An additional slow baseflow component was used in the original model as a visual inspection of hydrograph from the East Stoke gauging station, indicated slower groundwater-driven recession period of several months following wet periods (Fig. 3a). The original TOPCAT fast baseflow term (Q_b) is based on TOPMODEL theory (Beven and Kirkby, 1987) representing drainage from a saturated subsurface store with exponentially-decreasing transmissivity with depth, and has a recession period (defined by parameter m) typically of less than one month. QUICKFLOW represents overland flow entering the channel directly, which is made up of local Critical Source Area Runoff (a small % of the land) and Washoff Overland flow that occurs only during very large storm events, which allows 100 % runoff for short time periods of time. Flow pathways in TOPCAT are shown in Fig. 2.

The newly added slower baseflow component represents the groundwater store as a linear reservoir, which is assumed to be of infinite capacity. This method is similar to the baseflow component of the Australian lumped rainfall-runoff model SimHYD (Chiew et al., 2002). Only two parameters are required to generate a time varying slow baseflow (Q_{gw}) (m d^{-1}): SPLIT (–) will apportion active drainage towards either the fast baseflow store of the slow baseflow store (Quinn, 2004); and C_g, a recession rate constant (d^{-1}). Therefore Q_{gw} at time t, is given by

$$Q_{gw}(t) = C_g S_g(t-1) \tag{1}$$

Where S_g is the groundwater storage (in m), and C_g is defined above.

The initial storage S_{g_0} is set by the user by specifying the initial value of Q_{gw} (Q_{gw_0}). It is convenient to commence the TOPCAT simulation during a dry spell, where the baseflow component is relatively constant and most of the runoff consists of baseflow. Therefore, rearranging Eq. (1) gives

$$S_{g_0} = Q_{gw_0}/C_g \tag{2}$$

Where $Q_{gw_0} \equiv$ observed runoff on first day of simulation (m d^{-1}), following the assumption above.

The performance of TOPCAT in reproducing observed flows has been usually assessed (Quinn, 2004) by a combination of visual inspection of the modelled against observed runoff and the use of standard evaluation metrics (Nash–Sutcliffe Efficiency NSE). Model calibration aimed to maximise the value of NSE whilst ensuring that the MBE (mass balance error) was less than 10 %. The parameters C_g, SPLIT, m and SR-MAX (the latter 2 described in Quinn, 2004) were adjusted iteratively to enable this. Flow duration curves were also used to visually assess the model performance.

2.2.2 Nutrients

Nitrogen

The basis of the nitrogen cycling model used in TOPCAT (referred to originally as TOPCAT-N) was described in Quinn (2004) and simulates nitrate N only. Figure 2 shows the fluxes and stores in the conceptual nitrate N model in this version of TOPCAT. Note that N_{back} is the nitrate-N concentration in the slower groundwater store in this version of the model and is assumed to be constant over time.

Phosphorus

The basis of the phosphorus model in TOPCAT is found in Quinn et al. (2008). It simulates SRP and PP separately (TP is the sum of both species, i.e. organic species of P are not simulated and assumed to be negligible). Note that P_{back} is the SRP concentration in the slow groundwater in this version of the model and is assumed to be constant over time. A conceptual model of the fluxes and storages in the model is shown in Fig. 2.

Land Use and Nutrient Loading

In Quinn et al. (2008), it was proposed that the nutrient model parameters, particularly the nutrient application terms, $P_{initial}$ and $N_{initial}$ should be tied to the dominant land use/farming system in a catchment. The input rates are based on existing published evidence such as export coefficient studies that link land use to nutrient loadings. This essentially sets up the concentration of the leachate from the upper soil layers into the baseflow, store which is believed to be dynamic in nature as depletion of the store is allowed (see Quinn et al., 2008). This function was switched off in this application as the catchment was relatively large compared to previous TOPCAT applications, so spatial and temporal fluctuations in leaching would not be observed at the downstream monitoring points. Hence the soil leachate from the fast baseflow component is also

constant with time. Thus the final flow weighted mixing of overland flow, fast baseflow and slowflow components are used to simulate the observed flows in time.

2.3 Nutrient model, baseline scenarios

We simulated runoff and nutrients for a ten year baseline period, 1 January 1997 to 31 December 2006. This period was based on available hydrological and meteorological data. The model parameters were assumed to be constant over space and time except for P_{back}. It was necessary to reduce P_{back} from February 2002, due to improvements in Dorchester WWTP by 40 % based on the estimated reduction in the SRP load described above. Two sets of observed nutrient data were available to assess the model performance in two baseline scenarios:

(i) Weekly LTD sample data from East Stoke (data from 1 January 1997 to 31 December 2006), in a baseline scenario (SBW). In this scenario comparison of the model performance at predicting SRP and TP concentrations was curtailed at the end of 2001, just before the improvements to the Dorchester WWTP. However, for nitrate the model performance over the full 10 yr period was assessed.

(ii) Sub-daily (up to 8 per day) sample data from the HFD. The raw sub-daily sample data were compared to the daily modelled data in a baseline scenario (SBHR). The HFD covered a period after the WWTP improvements so the lower P_{back} value was used in this simulation. Therefore, only LTD and HFD nitrate/TON data overlapped in time.

A daily timestep was used in TOPCAT to model the SBHR nutrient therefore it was not possible to investigate event dynamics or hysteresis effects in the sub daily data (Bowes et al., 2009a). In the case of multiple samples taken on one day, the predicted nutrient concentration on that day was compared against each of the sample Cs. This is a limitation of the modelling approach, but higher resolution meteorological data and flows were not available to develop a sub-daily version.

2.3.1 Nutrient model parameter calibration

Model parameters were calibrated by assessing the performance of the model in the following metrics:

- Visually comparing the time series of nitrate, SRP and TP against the observed data and adjusting the most sensitive nutrient model parameters to obtain a best fit between modelled and observed time series. If possible the same parameters were used in both SBW and SBHR simulations. More weight was given in this study to the SBHR simulation since it was assumed a-priori that the higher resolution dataset would be more representative of the range of observed concentrations in the Frome.

- Calculating the distribution (Concentration-Duration) functions for the modelled outputs from the three nutrient species, and comparing against the observed distributions visually by plotting these together.

- Optimising the errors between modelled and observed mean and 90th percentile concentrations with the aim of reducing these below 10 % if possible. The mean and 90th percentile concentrations were chosen as these represent the concentrations over the range of flows (mean) and events (90th percentile), and therefore allow the model performance under all flow regimes to be assessed.

- In the case of SBHR, calculating the modelled loads of TON and P and comparing against the loads in Bowes et al. (2009a).

If satisfactory nutrient model outputs were not obtained by adjusting the nutrient parameters in the first step then it was necessary to adjust the hydrology model parameters, particularly QUICK and SPLIT, to increase or decrease the proportions of surface runoff, fast baseflow and slow baseflow (Quinn et al., 2008).

10175

3 Results

3.1 Comparing the LTD and HFD datasets

Here we show a comparison of the raw input data for the LTD and the HFD datasets. More detailed interpretation of the data is carried during the simulations sections below. Essentially we can compare the data set core statistics directly (Table 1). The patterns of the runoff are also captured in Fig. 3, which shows the concentration-duration plots of the observed data. Flow (runoff) data (Fig. 3a) are often shown as a flow duration curve when studying the catchment scale patterns for the flow. Equally, TP, SRP and TON can all be shown as a cumulative distribution, which is useful when assessing both: (a) the range of observed concentrations; (b) the model performance in reproducing this range. Note that the TP and SRP data were collected over different time periods (Table 1).

Figure 3 is discussed in more detail below. The flow duration curves (TL pane) indicate the observed daily flow on the sampling day (in the case of the HFD, samples taken on the same day are assigned identical daily flow values since sub-daily flow data were not available). The concentration-duration curves show the distribution of observed samples from the LTD and HFD on the same axes. Interpretation of the results is shown in the discussion.

3.2 Model calibration

The hydrology model parameters from the final calibration are shown in Table 2. The model results from the modified TOPCAT were as follows: the NSE for the baseline hydrology simulation was 0.75. The mass balance error was +9.2 % (over prediction), less than the 10 % limit that we considered acceptable for assessing the model performance as "satisfactory". The results in terms of matching the observed flow duration curves are shown in Fig. 4 (note that the same modelled flows were used in both SBHR and SBW model simulations). In the Frome catchment the percentage of surface overland

10176

flow (runoff) according to the calibrated model was very small (2 % of the total runoff of 462.5 mm yr^{-1}). The calibrated SPLIT parameter was 0.67 which meant that 1/3 of the excess soil water (i.e. excess of SRMAX) was recharge to the slower baseflow store.

The nutrient model results are also assessed visually using concentration-duration plots shown in Fig. 5 for TP (top panels), SRP (centre panels) and TON (bottom panels). The left panel shows SBHR results tested against HFD and the right panel shows SBW results tested against the LTD. Timeseries plots are shown (Fig. 6) for the SBW results tested against the LTD datasets, and in Fig. 7 for the SBHR results tested against the HFD datasets (figure captions explain the ordering of the panels). The model N and P parameters are shown in Table 3. Calibrated parameter values are shown from the baseline scenarios SBW and SBHR. The nitrogen model was very sensitive to the values of $N_{initial}$ and the soil texture parameter Φ (as noted by Quinn, 2004), which accounts for the water holding capacity of the soil (dimensionless). In this version the nitrate concentration in surface runoff NSR was calibrated at 1 mg L^{-1} N. The nitrate concentration in the slow groundwater N_{back} was calibrated using samples collected during low-flow periods (i.e. dominated by slow groundwater and/or WWTP discharges). The phosphorus model contains more parameters but only the sensitive ones, $P_{initial}$, and P_{back}, were calibrated. The soil leaching parameter Φ influences both nitrate and SRP concentrations in the baseflow component, so a value that produced acceptable results for both nutrient species had to be determined by calibration to achieve an optimal value.

It proved possible to use the same N and P parameters in both the SBW and SBHR simulations apart from (i) the parameter Φ, which affects the predicted nitrate and SRP concentrations (ii) P_{back} for reasons relating to the WWTP modifications discussed above. In the SBHR simulations a smaller value of Φ (0.26) was found to improve the nitrate predictions (see below) by increasing the leaching into the fast baseflow component. It was thus possible to test the SBW results against the HFD dataset, this procedure will form part of the discussion below.

10177

Table 4 shows the modelled catchment loads and model prediction errors along with the modelled concentration statistics and prediction errors in the mean and 90th percentile concentrations from the SBHR simulations. The model was calibrated to minimise the error between modelled and observed mean and 90th percentile concentrations, with the objective of achieving errors of less than 10 % where possible. Please note the SBHR model parameters were not optimised to reduce the error between modelled and observed loads.

4   Discussion

4.1   Interpretation of results

4.1.1   Runoff

Figure 3 (Top left panel) shows the distribution function (i.e. flow duration curve) of observed flows on sampling days only. The results are somewhat surprising since the range of flows sampled by the LTD dataset were higher than the HFD dataset. However, the time period of the latter was quite short and coincided with a relatively dry spell in the catchment. Peak flows were therefore lower than those recorded between 1997 and 2005 which were as high as 4.5 mm d^{-1}. Comparing sample days with low flows (i.e. samples taken during baseflow conditions) in both datasets, the 99th percentile flow values (i.e. flows exceeded 99 % of the time) were equal (0.44 mm d^{-1}, approximately 2 m^3 s^{-1}). The 99th percentile of the daily observed flow data (3652 values) was also very similar (Fig. 3). The consented discharge from Dorchester WWTP was 0.09 m^3 s^{-1} in 2012 (termed Dry Weather Flow – DWF) (Source: Wessex Water).

4.1.2   Nutrients

The concentration-duration plots for the LTD and HFD datasets are compared in Fig. 3 to assess the differences between (i) *the observed nutrient distributions* (here we will

10178

use the terms "weekly" and "high resolution" to refer to observed datasets from (i) LTD and (ii) HFD respectively). The weekly SRP timeseries has clearly missed some of the peaks measured by the high resolution sampling. This is even more pronounced in the TP distribution plots where the high resolution TP concentrations increased to nearly $2\,\mathrm{mg\,L^{-1}}$ P. These peaks were completely missed by the weekly sampling. It is important to note that: (i) weekly sampling days contained some higher observed flows than those on the high resolution sample days (Fig. 3 top left), due to higher runoff in the earlier (weekly monitoring) period (ii) TP and SRP data in the weekly dataset are from an earlier (pre-2002) monitoring period and this period did not overlap with the high resolution monitoring period.

A plot from the weekly dataset of C vs. Q (not shown) did not establish strong correlations between concentration and flow. Most of the higher TP concentrations tended to be associated with low flows in the weekly dataset, indicating a dominance of WWTP effluent source of TP in the catchment (Bowes et al., 2009a; Jarvie et al., 2006). Less than five samples measured TP concentrations > $0.2\,\mathrm{mg\,L^{-1}}$ P on days with high flows ($> 15\,\mathrm{m^3\,s^{-1}}$) that may be associated with PP transport, indicating that runoff events were probably of secondary importance in the catchment. As most phosphorus entering the groundwater component will be precipitated within the Chalk (House, 2003), the majority of the baseflow SRP load will be from WWTP effluent.

The weekly and high resolution nitrate/TON timeseries were quite similar indicating that the weekly monitoring data were probably sufficient to estimate the range of nitrate/TON concentrations in the catchment in order to assess compliance with EU WFD quality standards. Unlike the TP and SRP timeseries the monitored periods overlapped (Fig. 6, top panel). The correlation between C and Q (not shown) was weak, so it would not be possible to develop a Q vs. C rating curve to estimate loads from this dataset using the methods in Cassidy and Jordan (2011). We assumed that TON was predominantly nitrate, similar concentrations where the two timeseries overlapped tends to support this.

10179

Resampling the HFD timeseries to produce weekly or monthly timeseries of observed nutrient data was carried out and the resulting smoothed time series (not shown for brevity) were similar to the LTD weekly data. Again, peaks and troughs in the TP and SRP timeseries were largely removed by the resampling, especially in the monthly timeseries of SRP. In the case of TON, the resampled weekly and monthly timeseries still preserved the temporal variability of TON concentrations in the catchment fairly well, with important implications for optimal sampling intervals.

4.2 Model performance: baseline scenarios SBW and SBHR

4.2.1 Hydrology

It is possible of course to optimise the model parameters to generate a smaller mass balance error or a larger value of the NSE, both these criteria having separate parameter sets, but this was not attempted in this case, rather the MBE was constrained to be < 10 % with the corresponding optimal NSE value achieved through calibration (0.75).

The model overprediction (+9.2 %) was partly due to the simulations retaining a rapid surface runoff component (controlled by parameter QUICKCSA) from critical source areas (CSAs) in the catchment (e.g. farm yards, hard standings, feed lots; Edwards et al., 2008; Heathwaite et al., 2005). Simulating the surface runoff impacts on P dynamics is important and should not be lost during the aggregation and smoothing processes. The overprediction is seen as spikes on the simulated timeseries data and correspondingly high (i.e. > $5\,\mathrm{mm\,d^{-1}}$) values on the modelled flow duration curve. McIntyre et al. (2005) found that the INCA-N model performed best when the runoff (flow) parameters were calibrated at the same time as the nutrient parameters (in terms of minimising the RMSE error between observed and predicted concentrations) and this was also true in our study. The model generated only $9\,\mathrm{mm\,yr^{-1}}$ runoff (2 % of the total) from surface runoff, which was the sum of saturation excess and CSA runoff. Information on the area of CSAs in the catchment was unknown, although urban land use was only 1 %

10180

of the catchment area. These urban areas will contain impervious surfaces that can generate rapid flow during events (i.e. storms) and may act as CSAs for nutrients as well if they are associated with farm structures (Edwards et al., 2008), assuming that these structures are classified as "urban".

4.2.2 Nitrate N

The proportion of nitrate loads generated by surface runoff was negligible (0.4 %) in both baseline scenarios (where surface runoff was simulated). The nitrate loads were split between the fast and slow baseflow components in the model in the baseline scenarios. Surprisingly, the nitrate loads in the slow baseflow only contributed around 5 % of the total load in the baseline scenarios despite the fact that 30 % of the modelled runoff came from this component. This implies that the $N_{initial}$ parameter is the most important one, when calibrating the model to reproduce observed nitrate concentrations followed by Φ. N_{back} is far less important. In terms of the flow model parameters, SPLIT is obviously important since it controls the proportions of slow and fast baseflow in the total runoff. Otherwise, the nitrate model was not particularly sensitive to the flow model parameters. The model error in the SBHR simulation was only −2.3 % (underprediction), based on 1 yr of load data from Bowes et al. (2009b).

4.2.3 Phosphorus

In the SBW simulation the proportion of TP (i.e. PP) generated by surface runoff was about 40 % which is quite high considering only 2 % of the modelled runoff was from this pathway. Bowes et al. (2009a) estimated that between 1991 and 2003, SRP provided 65 % of the TP load in the river, which is close to the figure obtained by the SBW simulation (Table 5). In SBW the slow baseflow generated ten times the amount of soluble P than the fast baseflow component. This seems reasonable as "slow" baseflow included the WWTP discharges (in addition to the SRP originating from groundwater in the catchment ≈ 12.5 % in the model based on a groundwater concentration of

10181

0.04 $mg\,L^{-1}$ P). The SRP concentrations in the fast baseflow were most sensitive to the $P_{initial}$ parameter, however the load component from this pathway was surprisingly low (around 5 %). Again, the SPLIT parameter in the water flow model also had an influence on modelled SRP, in adjusting the ratio between fast and slow SRP baseflow loads. QUICK and QUICKCSA influenced the PP generated by surface runoff, and in fact had more effect on mean and 90th percentile P concentrations than on the runoff calibration metrics (NSE and MBE), which were insensitive to small changes in these parameters.

In the SBHR simulation, the model errors in predicting both the SRP and TP loads (from Bowes et al., 2009b) were greater than 10 % underprediction (Table 5, around 16 % underprediction). This was surprising because the errors in predicting mean and 90th percentile concentrations of TP species indicated that the model overpredicted these and in the case of the 90th percentile TP concentrations by a fairly large degree (43 %). The distribution plots in Fig. 5 indicate that very low SRP concentrations were not simulated by the model and also clearly show that 80th to 95th percentile TP concentrations were overpredicted by the model (although this would have a very small effect on the TP load). The time series plots in Fig. 7 indicate that some events in the HFD generating significant loads of TP (indicated by black rectangles) were missed by the model, but that the model overpredicted TP concentrations in autumn 2004, and that low SRP concentrations between February and June 2005 were overpredicted by the model. Another reason could be that the time periods over which concentrations were compared were slightly longer than the 365 day period used to estimate "observed" loads in Bowes et al. (2009b) and may have contained periods where the model was overpredicting these concentrations.

4.3 Comparison of nutrient model performance: SBW and SBHR

The results shown in both the distribution (Fig. 5) and time series plots (Fig. 7) show that the model (in SBHR) was not capable of predicting all the higher TP concentrations

10182

in the HFD dataset even after calibration. The model performance in SBHR indicated that it underestimated TP concentrations above $1.7\,\text{mg}\,\text{L}^{-1}\,\text{P}$ which was the maximum modelled value (maximum observed values were just under $2\,\text{mg}\,\text{L}^{-1}\,\text{P}$). However, the results from SBW (the distribution plot in Fig. 5 top right pane) indicate that high Cs ($> 0.4\,\text{mg}\,\text{L}^{-1}\,\text{P}$) predicted by the model were not replicated in the observed data for reasons discussed above (i.e. those high Cs not associated with runoff events). The model parameters (apart from P_{back}) were the same in both simulations, and values of $P_{initial}$ giving the best fit to the high resolution data were used in SBW as well. It is apparent that if these parameters were calibrated specifically on the LTD dataset alone then the peaks in the high resolution data would not be simulated, so TP model performance was clearly dependent on the temporal resolution of the observed data used to calibrate it.

The SRP concentrations were reproduced reasonably well in the SBW simulation with errors of less than 10 % between modelled and observed mean and 90th percentile values (Figs. 5 and 6). In the SBHR simulation the model underpredicted the observed peaks in SRP ($> 0.15\,\text{mg}\,\text{L}^{-1}\,\text{P}$) simulation (Figs. 5 and 7). This is in contrast to the performance of the LAM model (Bowes et al., 2009a) which overpredicted concentrations of SRP in winter and spring. This was partly due to an overestimation of the SRP load from diffuse sources which may have reduced following the Foot and Mouth disease epidemic in early 2001. In our model the reasons are discussed below. The nitrate model was also able to reproduce both the SBW and SBHR observed data sets reasonably well with Fig. 5a and b (lower panes) indicating that the difference between modelled and observed concentrations was generally $< 0.2\,\text{mg}\,\text{L}^{-1}\,\text{N}$. Low (samples $<= 4\,\text{mg}\,\text{L}^{-1}\,\text{N}$) nitrate concentrations observed in SBW and SBHR were slightly overpredicted by the model.

The increase in nitrate concentrations over time discussed by Smith et al. (2010) can be weakly observed in the SBW dataset over the 10 yr simulated here, however the model appears to have been able to reproduce the concentrations reasonably well despite not having a time-varying N_{back} parameter to represent this. The SBHR results

10183

were slightly better than the SBW results in terms of matching the observed distribution functions (Fig. 5), probably because the Φ parameter was reduced slightly to 0.26 increasing the nitrate concentrations in baseflow. The higher Φ value was found to improve the P model results in SBW. Comparing the two distribution curves from the model results indicates the model sensitivity to this parameter. Both the observed SBW concentrations and the underlying trend in the SBHR concentrations could be fitted reasonably well by TOPCAT indicating that the higher resolution data was probably of less value in model fitting.

The timeseries plots of modelled and observed (HFD) concentrations of all 3 nutrients from the SBHR simulations (Fig. 7) further emphasize the above points. The TON timeseries show that the model can simulate observed drops in TON concentration when there were runoff events ($\text{NSR} = 1\,\text{mg}\,\text{L}^{-1}\,\text{N}$) suggesting that this dilution process has been correctly represented in TOPCAT. In these simulations the model generated a negligible load of SRP in surface runoff. It may be that this component of the SRP load needs to be adjusted upwards to reproduce the higher peaks in the observed HFD SRP timeseries (up to $0.5\,\text{mg}\,\text{L}^{-1}\,\text{P}$). The mechanism generating these peaks is unclear from these observed HFD data, surface runoff appears unlikely as some peaks did not appear to coincide with rainfall events (e.g. the observed spikes on 2 May 2005 and 16 May 2005). For both the SBW and SBHR simulations TOPCAT has not simulated short term higher concentrations of SRP very well. This reflects an assumption that SRP would be less flashy and be dominated by slower subsurface events. Clearly, even at the large catchments scale, quicker SRP processes are operating and may need to be simulated in future versions of the model.

The overall seasonal trend in the TON and SRP data was picked up well by the model in the SBHR simulations. In the HFD TP timeseries the dashed black rectangles indicate periods where no rainfall data was present (Fig. 7 middle pane). Since the PP component of TP in TOPCAT is generated by surface runoff entraining sediment it was not possible to reproduce these high concentration spikes with the existing model and rainfall data. The background concentrations (i.e. the soluble component of TP) were

probably controlled more by the WWTP discharges than groundwater in this catchment (based on the ratio of point: diffuse sources estimated by Bowes et al., 2011). Therefore, they tended to follow a similar pattern to the SRP concentrations. The modelled TP concentrations were much more constant over time than the observed concentrations which again may suggest that some physical processes controlling TP, such as in-channel remobilization of sediment P (Bowes et al., 2005) were not included. The difference between modelled and observed TP concentration was typically $\pm 1 \, \text{mg} \, \text{L}^{-1} \, \text{P}$ or less so the overall effect on catchment loads may be insignificant, however predicting peak nutrient concentrations is still important as they may promote algal blooms (Bowes et al., 2009b).

4.4 Testing SBW against the HFD dataset

In addition, a SBW model run (not shown graphically) that attempted to reproduce the range of concentrations (and therefore nutrient loads) observed in the HFD using a model calibrated using the weekly data from the LTD, simulated: (i) nitrate reasonably well with no deterioration in predictive accuracy; (ii) TP, where concentrations observed in the HFD were underpredicted by the SBW model, so the loads were also underestimated; (iii) SRP, where the performance of both the SBW and SBHR model simulations assessed against HFD concentrations were similar (high concentrations were not reproduced by either run), and SRP loads that were slightly underestimated by the SBHR model run (Table 2) were also underestimated by the SBW model run.

4.5 Testing SBHR against resampled data

The HFD dataset was resampled to create additional weekly and monthly timeseries. A simple approach was used based on starting the resampled weekly and monthly timeseries on the start date of the HFD (14 January 2004), then calculating the next sampling date in the series (7 days or 1 calendar month later respectively) and extracting the sample on or closest to this date for the next value in the timeseries. The

10185

SBHR model run was not recalibrated against the resampled data. Figure 8 shows the modelled (SBHR) concentration plotted with the weekly and monthly resampled HFD concentration data.

The resampled HFD TP data has omitted most of the high peaks in TP concentration, so the modelled and observed concentrations look quite similar except for some spikes in the modelled concentrations that were not picked up by the resampled observed data. A TP model calibrated on the resampled lower resolution data would therefore underpredict the "true" peaks in concentration identified by the HFD dataset (Fig. 7; backing up the point in the previous section about high resolution sampling being most important when measuring TP). The resampled HFD SRP data has kept some of the temporal variability shown in the high-resolution original dataset. The resampled HFD TON data also resembles the original HFD data, indicating that choosing a lower frequency dataset to calibrate or validate the SBHR nitrate or SRP model would not detrimentally affect the model performance.

It is stressed that collecting the longest possible high resolution dataset particularly for all forms of P is of the utmost importance for effective water quality monitoring and identifying the full range of observed concentrations (see Fig. 3 TR pane for an example here where the LTD TP dataset has missed the peak TP concentrations). Recent examples of long term monitoring at a high temporal frequency are the DTC (Demonstration Test Catchments) project in the UK, based in the Eden catchment in Cumbria (Owen et al., 2012), and the monitoring in the Blackwater catchment, Ireland (Cassidy and Jordan, 2011).

5 Conclusions

– The LTD low-frequency data have proven to be very useful in showing long terms trends in flow and nutrient pattern as well as seasonal fluxes. The data clearly show the impact of the improvement to the treatment processes at the Dorchester WWTP in the early 2000s (see point below relating to WWTP discharges).

10186

– The HFD high-frequency data have revealed much more detail in the seasonal and event driven fluxes. The data also show quite a lot of "noise" probably driven but inadequate rainfall data and unknown localised land use activities. The impact of individual storms was therefore clear in the HFD. Nevertheless the data are quite revealing and show some of the component inputs to the longer term nutrient loss patterns and reveals the larger scale impacts of these losses. The ability to move to regular hourly sampling using new technologies will continue to improve this situation.

– Observed data is important to the actual MIR model when choosing what process based components to include or to remove. Adding a dynamic slow flow store with constant nutrient concentration was imperative, but the use of the dynamics nutrient fluxes in the fast baseflow store was switched off. Mixing of event runoff with fast and slow subsurface component was the minimum requirement for the model to simulate observed patterns at this scale. The requirement for a more "flashy" SRP component is also seen in the data.

– Simulating the LTD dataset with a process based daily model has shown a number of dominant patterns that can be picked up by the model. Hence some justification for the causes of those patterns can be made. A small modification to the model parameter values was needed to simulate the HFD dataset, but essentially the model for both LTD and HFD was the same. Indicating a sound basis for the model structure.

– The TOPCAT model performed adequately at simulating nitrate and SRP in both LTD and HFD datasets. The performance in simulating TP was acceptable with the weekly monitoring data, but was not so good visually compared to the high frequency data. The ability to unravel processes and noise from this dataset may prove to be difficult, but it is important know that the pattern is there.

10187

– The spikes in phosphorus (both SRP and TP) concentration in the high-frequency monitoring data were not evident in the LTD weekly samples. The current TOP-CAT model cannot reproduce these spikes without drastically recalibrating the parameters in the SRP and PP models. The spikes in the TOPCAT model for the LTD were deliberately left in as expert knowledge from localised studies suggests that fast runoff is associated with TP spikes.

– The SRP concentration and loads in the Frome were strongly related to discharges from WWTPs. These changed greatly during the gathering of the calibration datasets, due to the implementation of P-stripping technology in the early 2000s. This change in phosphorus point source load throughout the 2000s was a common occurrence in most large UK catchments, due to WFD implementation. Therefore, a model with the potential for adjustable background nutrient load/flow (such as TOPCAT) is required to model such catchments.

– The TP results in the Frome were highly sensitive to the temporal resolution of the observed data. If low resolution data are used for calibration then a fitted model will underpredict the peaks in concentration not picked up by the monitoring. The value of high resolution TP monitoring can clearly be shown where this has picked up these peaks. However, the SBHR model failed to predict the observed TP and SRP loads in this dataset within 10 % (although the model was not calibrated to do so as HFD was acting as a validation dataset).

– Nitrate concentrations were observed to be rising in the Frome since the 1940s, however over the simulation period the rate of increase was fairly small and the model could predict the time series reasonably well. However it does require the mixing of both a soil dominated faster flow pathway and slower groundwater dominated baseflow component. Nitrate in the Frome may also have been affected by rural policy changes, and the Foot and Mouth disease outbreak appeared to cause levels to fall in the early 2000s. Using high resolution nitrate did not improve

10188

the predictions, as the same trendline could be fitted to SBHR using a model fitted to weekly data.

– The model currently runs on a daily timestep so some modifications and high frequency forcing data (observed flow and precipitation) would be necessary to fully evaluate its performance against sub-daily observed nutrient data. The ability to run this model at a higher frequency time step is possible but would require an advanced routing function in the model. The rewards of doing this may not be high as much of the observed HFD may be too "noisy" to simulate. It may be more important to run on a daily timestep with simple mixing equations and an argument of the process-based origins of this model to be made, based on observation acquired in local scale studies.

– There may be some evidence here that collecting higher resolution data is an advantage in order to understand extreme values and addressing the issue of "noise" in the datasets. It may still be beneficial to aggregate sub-daily data to daily data as a compromise between the capabilities of this process based model and information actually contained in the HFD data.

– Further work will now address the implications of the model parameter sensitivity and will look at the implication of managing land use that influences hydrological flow pathways, and nutrient loading.

– Higher resolution datasets recorded at multiple scales would allow more information to be built up on the actual fluxes in larger catchments. The MIR tools could then run alongside more physically-based model being applied at the "research scale". The more parsimonious, process-based models being more suitable for management purposes at larger scales as suggested in the on-going DTC study.

10189

References

Argent, R. M., Perraud, J.-M., Rahman, J. M., Grayson, R. B., and Podger, G. M.: A new approach to water quality modelling and environmental decision support systems, Environ. Model. Software., 24, 809–818, 2009.

Arnold, J. G.: SWAT (Soil and Water Assessment Tool), Grassland, Soil and Water Res. Laboratory, USDA, Agricultural Research Service, Temple, TX, USA, 1994.

Bartley, R., Speirs, W. J., Ellis, T. W., and Waters, D. K.: A review of sediment and nutrient concentration data from Australia for use in catchment water quality models, Mar. Pollut. Bull., 65, 101–116, 2012.

Beven, K. J.: A Discussion of Distributed Hydrological Modelling, Springer, Dordrecht, the Netherlands, 255–278, 1996.

Beven, K. J. and Kirby M. J.: A physically-based variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43–69, 1979.

Birkinshaw, S. J. and Ewen, J.: Nitrogen transformation component for SHETRAN catchment nitrate transport modelling, J. Hydrol., 230, 1–17, 2000.

Bowes, M. J., Leach, D. V., and House, W. A.: Seasonal nutrient dynamics in a chalk stream: the river Frome, Dorset, UK, Sci. Tot. Environ., 336, 225–241, 2005.

Bowes, M. J., Smith, J. T., Jarvie, H. P., Neal, C., and Barden, R.: Changes in point and diffuse source phosphorus inputs to the River Frome (Dorset, UK) from 1966 to 2006, Sci. Tot. Environ., 407, 1954–1966, 2009a.

Bowes, M. J., Smith, J. T., and Neal, C.: The value of high-resolution nutrient monitoring: a case study of the River Frome, Dorset, UK, J. Hydrol., 378, 82–96, 2009b.

Bowes, M. J., Neal, C., Jarvie, H. P., and Davies, H. N.: Predicting phosphorus concentrations in British rivers resulting from the introduction of improved phosphorus removal from sewage effluent, Sci. Tot. Environ., 408, 4239–4250, 2010.

Bowes, M. J., Smith, J. T., Neal, C., Leach, D. V., Scarlett, P. M., Wickham, H. D., Harman, S. A., Armstrong, L. K., Davy-Bowker, J., Haft, M., and Davies, H. N.: Changes in water quality of the River Frome (UK) from 1965 to 2009: is phosphorus mitigation finally working?, Sci. Tot. Environ, 409, 3418–3430, 2011.

Bowes, M. J., Palmer-Felgate, E. J., Loewenthal, M., Jarvie, H. P., Wickham, H. D., Harman, S. A., and Carr, E.: High-frequency phosphorus monitoring of the River Kennet, UK:

10190

are ecological problems due to intermittent sewage treatment works failures?, J. Environ. Monitor., 14, 3137–3145, 2012.

Casey, H.: Variation in chemical composition of River Frome, England, from 1965 to 1972, Freshw. Biol., 5, 507–514, 1975.

Casey, H. and Clarke, R. T.: Statistical analysis of nitrate concentrations from the River Frome (Dorset) for the period 1965–76, Freshw. Biol., 9, 91–97, 1979.

Casey, H., Clarke, R. T., and Smith, S. M.: Increases in nitrate concentrations in the River Frome (Dorset) catchment related to changes in land use, fertiliser applications and sewage input, Chem. Ecol., 8, 105–117, 1993.

Cassidy, R. and Jordan, P.: Limitations of instantaneous water quality sampling in surface-water catchments: comparison with near-continuous phosphorus time-series data, J. Hydrol., 405, 182–193, 2011.

Chiew, F. H. S., Peel, M. C., Western, A. W., Singh, V. P., and Frevert, D.: Application and testing of the simple rainfall-runoff model SIMHYD, in: Mathematical Models of Small Watershed Hydrology and Applications, Water Resources Publications, Highlands Ranch, CO, USA, 335–367, 2002.

Davison, P.S, Withers, P. J. A., Lord, E. I., Betson, M. J., and Strömqvist, J.: PSYCHIC – a process-based model of phosphorus and sediment mobilisation and delivery within agricultural catchments, Part 1: Model description and parameterisation, J. Hydrol., 350, 290–302, 2008.

Dean, S., Freer, J., Beven, K., Wade, A. J., and Butterfield, D.: Uncertainty assessment of a process-based integrated catchment model of phosphorus, Stoch. Environ. Res. Risk Assess., 23, 991–1010, 2009.

Edwards, A. C., Kay, D., McDonald, A. T., Francis, C., Watkins, J., Wilkinson, J. R., and Wyer, M. D.: Farmyards, an overlooked source for highly contaminated runoff, J. Enviro. Manag., 87, 551–559, 2008.

Environment Agency (EA): Making Information Available for Integrated Catchment Management Science Report – SC060035/SR2007, Bristol, UK, 114 pp., 2007.

Evans, D. J. and Johnes, P.: Physio-chemical controls on phosphorus cycling in two lowland streams, Part 1 – the water column, Sci. Tot. Environ., 329, 145–163, 2004.

eWater Cooperative Research Centre: Source Catchments Scientific Reference Guide, eWater Cooperative Research Centre, Canberra, ACT, Australia, 2010.

10191

Ewen, J., Parkin, G., and O'Connell, P. E.: SHETRAN: distributed river basin flow and transport modeling system, J. Hydrol. Eng., 5, 250–258, 2000.

Hanrahan, G., Gledhill, M., House, W. A., and Worsfold, P. J.: Phosphorus loading in the Frome Catchment, UK: seasonal refinement of the coefficient modelling approach, J. Env. Qual., 30, 1738–1746, 2001.

Haygarth, P. M., Wood, F. L., Heathwaite, A. L., and Butler, P. J.: Phosphorus dynamics observed through increasing scales in a nested headwater-to-river channel study, Sci. Tot. Environ., 344, 83–106, 2005.

Heathwaite, A. L., Quinn, P. F., and Hewett, C. J. M.: Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation, J. Hydrol., 304, 446–461, 2005.

House, W. A.: Geochemical cycling of phosphorus in rivers, Appl. Geochem., 18, 739–748, 2003.

Howden, N. J. K. and Burt, T. P.: Statistical analysis of nitrate concentrations from the Rivers Frome and Piddle (Dorset, UK) for the period 1965–2007, Ecohydrology, 2, 55–65, 2009.

Jarvie, H. P., Neal, C., and Withers, P. J. A.: Sewage-effluent phosphorus: a greater risk to river eutrophication than agricultural phosphorus?, Sci. Tot. Environ., 360, 246–253, 2006.

Johnes, P. J.: Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: the export coefficient modelling approach, J. Hydrol., 183, 323–349, 1996.

Johnes, P. J.: Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density, J. Hydrol., 332, 241–258, 2007.

Jordan, P., Arnscheidt, J., McGrogan, H., and McCormick, S.: High-resolution phosphorus transfers at the catchment scale: the hidden importance of non-storm transfers, Hydrol. Earth Syst. Sci., 9, 685–691, doi:10.5194/hess-9-685-2005, 2005.

Jordan, P., Arnscheidt, A., McGrogan, H., and McCormick, S.: Characterising phosphorus transfers in rural catchments using a continuous bank-side analyser, Hydrol. Earth Syst. Sci., 11, 372–381, doi:10.5194/hess-11-372-2007, 2007.

Krause, P., Bäse, F., Bende-Michl, U., Fink, M., Flügel, W., and Pfennig, B.: Multiscale investigations in a mesoscale catchment – hydrological modelling in the Gera catchment, Adv. Geosci., 9, 53–61, doi:10.5194/adgeo-9-53-2006, 2006.

10192

Kronvang, B. and Bruhn, A. J.: Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams, Hydrol. Process., 10, 1483–1501, 1996.

Lunn, R. J., Adams, R., Mackay, R., and Dunn, S. M.: Development and application of a nitrogen modelling system for large catchments, J. Hydrol., 174, 285–304, 1996.

Marsh, T. J. and Hannaford, J. (Eds.): UK Hydrometric Register, Hydrological data UK series, Centre for Ecology and Hydrology, Wallingford, 210 pp., 2008.

McIntyre, N., Jackson, B., Wade, A. J., Butterfield, D., and Wheater, H. S.: Sensitivity analysis of a catchment-scale nitrogen model, J. Hydrol., 315, 71–92, 2005.

Palmer-Felgate, E. J., Jarvie, H. P., Williams, R. J., Mortimer, R. J. G., Loewenthal, M., and Neal, C.: Phosphorus dynamics and productivity in a sewage-impacted lowland chalk stream, J. Hydrol., 351, 87–97, 2008.

Quinn, P.: Scale appropriate modelling: representing cause and effect relationships in nitrate pollution at the catchment scale for the purpose of catchment scale planning, J. Hydrol., 291, 197–217, 2004.

Quinn, P. F., Anthony, S., and Lord, E.: Basin scale nitrate modelling using a minimum information requirement approach, in: Water Quality: Processes and Policy, edited by: Trudgill, S., Walling, D., Webb, B., Wiley, Chichester, 101–117, 1999.

Quinn, P., Hewitt, C. J. M., and Dayawansa, N. D. K.: TOPCAT-NP: a minimum information requirement model for simulation of flow and nutrient transport from agricultural systems, Hydrol. Process., 22, 2565–2580, 2008.

Sharpley, A. N., Kleinman, P. J. A., Heathwaite, A. L., Gburek, W. J., Folmar, G. J., and Schmidt, J. P.: Phosphorus loss from an agricultural watershed as a function of storm size, J. Env. Qual., 37, 362–368, 2008.

Smith, J. T., Clarke, R. T., and Bowes, M. J.: Are groundwater nitrate concentrations reaching a turning point in some chalk aquifers?, Sci. Tot. Env., 408, 4722–4732, 2010.

Storr, E., Adams, R., and Western, A.: How can data from headwater catchments be used to improve runoff and nutrient predictions at larger scales?, in: MODSIM2011, 19th International Congress on Modelling and Simulation, edited by: Chan, F., Marinova, D., and Anderssen, R. S., Modelling and Simulation Society of Australia and New Zealand, 1652–1658, 2011.

Wade, A. J., Whitehead, P. G., and Butterfield, D.: The Integrated Catchments model of Phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heteroge-

10193

neous river systems: model structure and equations, Hydrol. Earth Syst. Sci., 6, 583–606, doi:10.5194/hess-6-583-2002, 2002.

Wade, A. J., Butterfield, D., and Whitehead, P. G.: Towards an improved understanding of the nitrate dynamics in lowland, permeable river-systems: applications of INCA-N, J. Hydrol., 330,185–203, doi:10.1016/j.jhydrol.2006.04.023, 2006.

Wade, A. J., Palmer-Felgate, E. J., Halliday, S. J., Skeffington, R. A., Loewenthal, M., Jarvie, H. P., Bowes, M. J., Greenway, G. M., Haswell, S. J., Bell, I. M., Joly, E., Fallatah, A., Neal, C., Williams, R. J., Gozzard, E., and Newman, J. R.: Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring, Hydrol. Earth Syst. Sci., 16, 4323–4342, doi:10.5194/hess-16-4323-2012, 2012.

Wheater, H. S. and Peach, D.: Developing interdisciplinary science for integrated catchment management: the UK lowland catchment research (LOCAR) programme, Int. J. Wat. Res. Develop., 20, 369–385, 2004.

Withers, P. J. A. and Lord, E. L.: Agricultural nutrient inputs to rivers and groundwaters in the UK: policy environmental management and research needs, Soil Use Manage., 14, 186–192, 2002.

10194

Table 1. Long term nutrient concentration statistics in the LTD and HFD datasets.

Dataset/Nutrient (time period)	Number of Observations	10th Percentile concentration (mg L^{-1})	Median concentration (mg L^{-1})	Mean concentration (mg L^{-1})	90th Percentile concentration (mg L^{-1})
LTD Nitrate (7 Jan 97–21 Nov 06)	384	4.6	5.6	5.6	6.9
LTD TP (7 Jan 97–20 Dec 01)	168	0.13	0.22	0.21	0.30
LTD SRP (7 Jan 97–20 Dec 01)	174	0.08	0.14	0.14	0.20
HFD TON (12 Dec 04–31 Jan 06)	1454	4.5	5.4	5.5	6.7
HFD TP (14 Jan 04–31 Jan 06)	2290	0.09	0.15	0.17	0.24
HFD SRP (1 Feb 05–31 Jan 06)	1340	0.06	0.09	0.09	0.14

Table 2. Hydrological model calibrated parameters and performance metrics. Parameter definitions can be found in Quinn (2004).

SRMAX (m)	m (d^{-1})	QUICK (–)	QUICKCSA (–)	SPLIT (–)	C_g (d^{-1})	NSE (–)	MBE (%)
0.037	0.045	2	0.12	0.67	0.0024	0.75	+9.2

Table 3. Nutrient modelling parameters, baseline scenarios.

Scenario Name	NSR $(mg\,L^{-1}\,N)$	N_{back} $(mg\,L^{-1}\,N)$	P_{back} $(mg\,L^{-1}\,P)$	$N_{initial}$ $(kg\,ha^{-1}\,N)$	$P_{initial}$ $(kg\,ha^{-1}\,P)$
Baseline SBW	1	2.3	0.32[a]	19.9	0.9
Baseline SBHR	1	2.3	0.2	19.3	0.9

[a] Reduced to 0.2 after 1 Feb 2002.

10197

Table 4. Nutrient modelling results, baseline scenarios. Note also that $a +$ ve error indicates model over prediction.

Dataset	Mean C_{mod} $(mg\,L^{-1})$	Error (%)	90th percentile C_{mod} $(mg\,L^{-1})$	Error (%)	Modelled[b] Annual Load (tonnes yr^{-1})	Error[b] (%)
SBW Nitrate	5.4	−4.6	6.3	−8.6	N/A	
SBW TP[a]	0.20	−5.1	0.30	+0.3	N/A	
SBW SRP[a]	0.13	−5.4	0.19	−2.9	N/A	
SBHR TON	5.7	−3.8	6.7	−0.2	818	−2.3
SBHR TP	0.18	+6.0	0.34	+43	21.8	−16.7
SBHR SRP	0.09	−2.8	0.12	−15	11.0	−16.9

[a] Until 31 Dec 2001;
[b] from observed loads in Bowes et al. (2009b); total load over the period 1 Feb 2005 to 31 Jan 2006.

Table A1. Nomenclature.

HFD	High Frequency data set of nitrogen and phosphorus, recorded several times per day
LTD	Long term data set of weekly nitrogen and phosphorus measurements in River Frome
MBE	Mass balance error
NSE	Nash–Sutcliffe Efficiency (model performance metric)
SBHR	Baseline model scenario simulating the HFD on a daily timestep
SBW	Baseline model scenario simulating the LTD (weekly data) on a daily timestep
SRP	Soluble reactive phosphorus (measured values filtered using 0.45 μm paper)
TON	Total oxidised nitrogen (nitrate + nitrite)
TP	Total phosphorus (soluble + insoluble forms)
WFD	Water Framework Directive
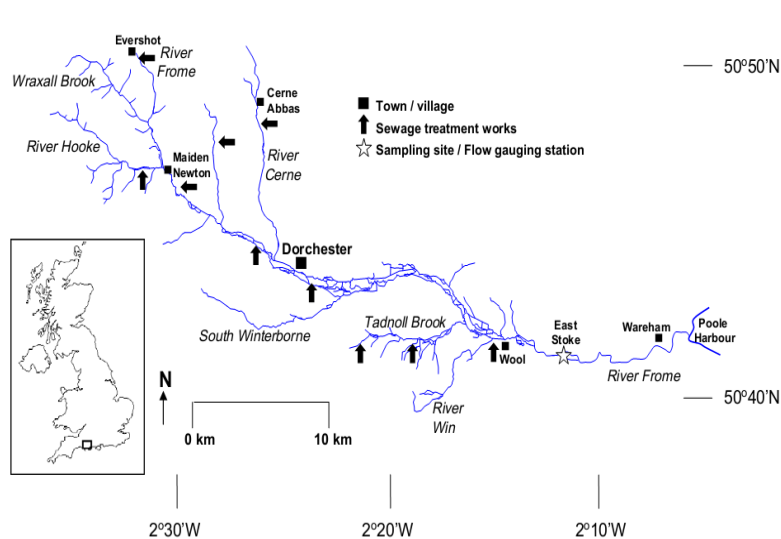
Fig. 1. Schematic map of Frome Catchment showing monitoring points (from Bowes et al., 2009a).
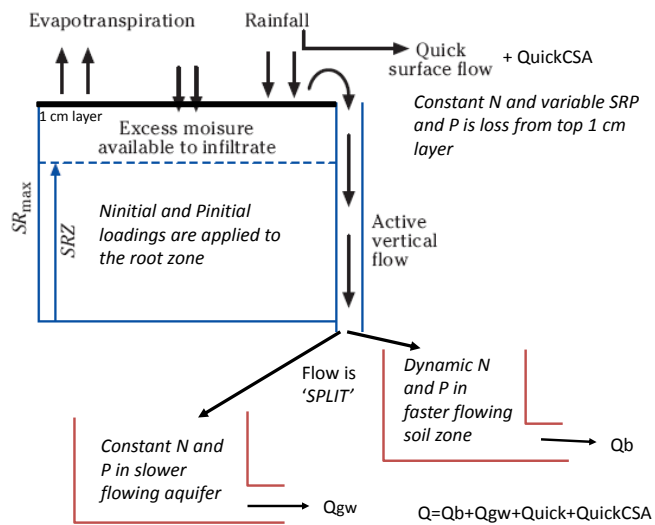
Fig. 2. Conceptual model of TOPCAT, showing the dominant flow pathways.
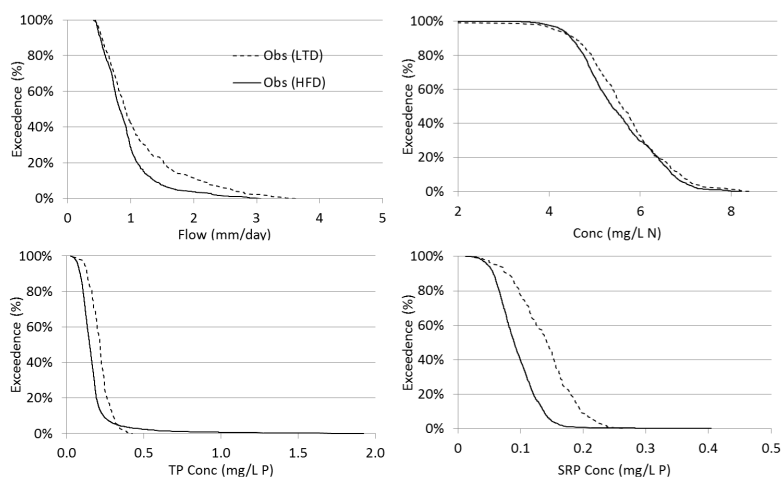
Fig. 3. Cumulative distributions of observed data for the LTD and HFD datasets: **(a)** (TL), Q; **(b)** (TR), TP, **(c)** (BL), SRP, **(d)** (BR) TON (HFD) and nitrate (LTD).
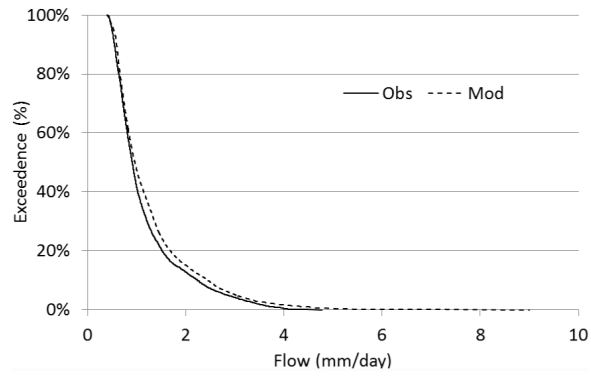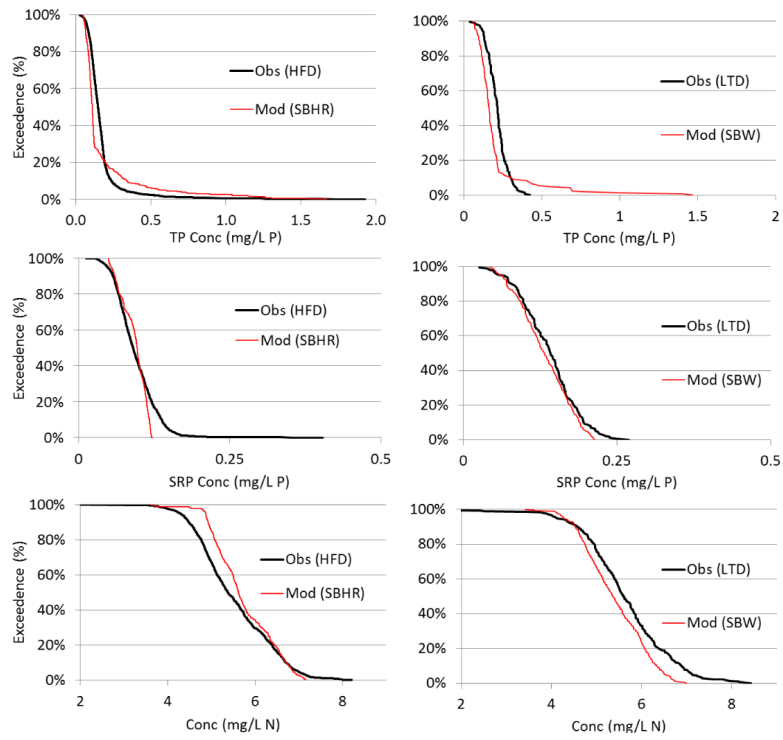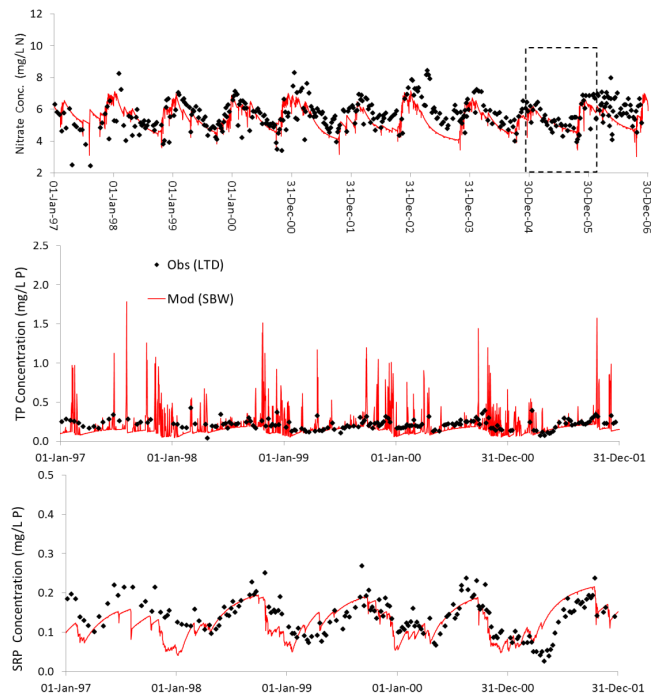
Fig. 4. Modelled and observed flow duration curves.

**Fig. 5.** Modelled and observed concentration duration curves for the SBW and SBHD simulations.

Fig. 6. Timeseries plots of modelled and observed nutrient data: nitrate using SBW shown with LTD (top panel), TP using SBW shown with LTD (middle panel), SRP using SBW shown with LTD (bottom panel). The dashed black rectangle in the top panel illustrates the period of the SBHR TON monitoring data.

10205
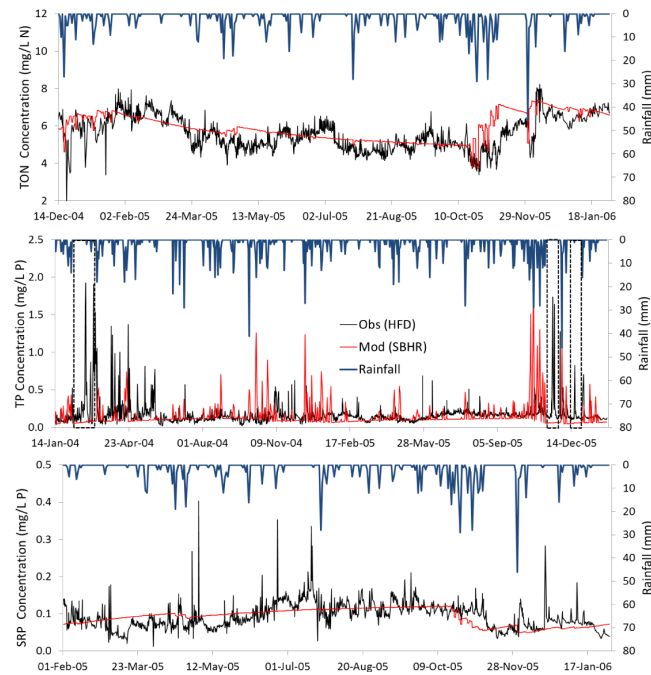
Fig. 7. Timeseries plots of modelled and observed nutrient data: **(a)**TON using SBHR shown with HFD observed data, **(b)** TP using SBHR shown with HFD, **(c)** SRP using SBHR shown with HFD. The dashed black rectangles indicate dry periods with unexplained spikes in concentration.

10206
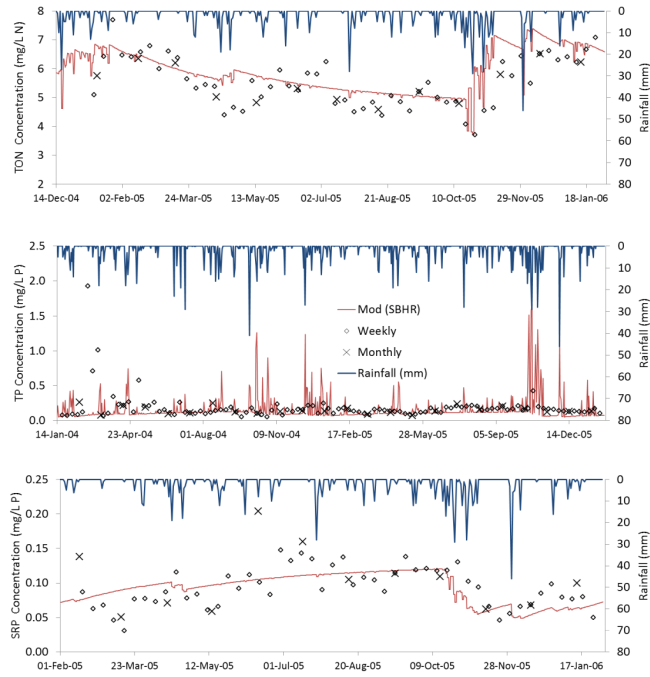
Fig. 8. Timeseries plots of SBHR model performance tested against resampled weekly and monthly HFD data, for: **(a)** TON/Nitrate; **(b)** TP; **(c)** SRP.

10207