Adams R, Quinn PF, Perks M, Barber NJ, Jonczyk J, Owen GJ. [Simulating high frequency water quality monitoring data using a catchment runoff attenuation flux tool \(CRAFT\)](#). *Science of the Total Environment* 2016

DOI: <http://dx.doi.org/10.1016/j.scitotenv.2016.01.045>

Copyright:

© 2016. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

Date deposited:

12/07/2016

Embargo release date:

06 February 2017

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Simulating high frequency water quality monitoring data using a catchment runoff attenuation flux tool (CRAFT)

Authors: Russell Adams^{a*}, Paul F. Quinn^a, Matthew Perks^b, Nicholas J. Barber^c, Jennine Jonczyk^a, Gareth J. Owen^a

^a *School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK*

^b *School of Geography, Politics and Sociology, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK*

^c *Dept. of Geography, Science Laboratories, Durham University, South Rd, Durham, DH1 3LE UK*

**Corresponding Author: Email russelladams68@gmail.com*

Abstract

High resolution water quality data has recently become widely available from numerous catchment based monitoring schemes. However, the models that can reproduce time series of concentrations or fluxes have not kept pace with the advances in monitoring data. Model performance at predicting phosphorus (P) and sediment concentrations has frequently been poor with models not fit for purpose except for predicting annual losses. Here, the data from the Eden Demonstration Test Catchments (DTC) project have been used to calibrate the Catchment Runoff Attenuation Flux Tool (CRAFT), a new, parsimonious model developed with the aim of modelling both the generation and attenuation of nutrients and sediments in small to medium sized catchments. The CRAFT has the ability to run on an hourly timestep and can calculate the mass of sediments and nutrients transported by three flow pathways representing rapid surface runoff, fast subsurface drainage and slow groundwater flow (baseflow). The attenuation feature of the model is introduced here; this enables surface runoff and contaminants transported via this pathway to be delayed in reaching the catchment outlet. It was used to investigate some hypotheses of nutrient and sediment transport in the Newby Beck Catchment (NBC). Model performance was assessed using a suite of metrics including visual best fit and the Nash-Sutcliffe efficiency. It was found that this approach for water quality models may be the best assessment method as opposed to using a single metric. Furthermore, it was found that, when the aim of the simulations was to reproduce the time

series of total P (TP) or total reactive P (TRP) to get the best visual fit, that attenuation was required. The model will be used in the future to explore the impacts on water quality of different mitigation options in the catchment; these will include attenuation of surface runoff.

Keywords

Catchment modelling; diffuse pollution; nutrient pollution; phosphorus; sediment transport; high resolution data

1. Introduction

Much research has been carried out at the field scale particularly to investigate the cycling of phosphorus (P) and fine sediments during storm events and to identify generation and transformation processes that can be deduced from observations (e.g. hysteresis patterns and the connectivity of the riparian zone to the hillslope during runoff events) (e.g. Halliday et al., 2014; Mellander et al., 2012; Outram et al., 2014; Perks et al., 2015). These have been studied at different scales from plot to catchment (<1ha to 100+ km²) (e.g. Haygarth et al., 2005; Bowes et al., 2003) and in different climatic conditions to attempt to identify the transient, yet important drivers for high fluxes (e.g. antecedent conditions) without using models (e.g. Bilotta et al., 2007, 2010). Models are however necessary where predictions of the effects of land use and climate change on water quality are required (Wellen et al., 2015).

Water quality models however, have not kept pace with the proliferation of high resolution water quality monitoring networks in the past decade. The most widely used such as INCA, AGNPS/AnnAGNPS and SWAT (Wade et al., 2002; Binger et al., 2011; Gassman et al., 2007) operate on a daily or monthly timestep whereas observations of nutrient concentrations are now available at intervals as short as 30 minutes. Physically-based models such as SHETRAN and complex, lumped models like HSPF (Bicknell et al., 1996) are capable of simulating sub-daily fluxes of nutrients (e.g. from SHETRAN the Slapton nitrate study of Birkinshaw and Ewen, 2000; and the P modelling study in Ireland of Nasr et al., 2007)) but their data requirements are onerous in terms of parameterizing both the physical catchment (e.g. gridded elevation and soil property data) and the nutrient cycle (e.g. a complete representation of the nitrogen (N) and P cycles). Wellen et al (2015) have reviewed the performance of all the models listed above at predicting nutrients and sediments, emphasising poor performance both in terms of inaccurate predictions of concentrations and/or loads and substandard modelling practices (e.g. not performing a sensitivity analysis) even when running at a daily timestep. Jackson-Blake et al. (2015) have critiqued both the performance of the INCA-P model and the methods commonly used to assess performance (e.g. the Nash-Sutcliffe efficiency NSE) as being inadequate for water quality models. Beven (2009) has also critiqued the methods commonly used to evaluate hydrological models as not

being appropriate in many cases and has encouraged a full evaluation of model uncertainties to be made, using the limits of acceptability as pre-defined by the modeller (e.g. Holloway et al., *submitted*).

In the European Union the Water Framework Directive (WFD) (2000/60/EC) has prioritised the reduction of diffuse pollution of freshwaters from agricultural catchments (McGonigle et al., 2014). Barber and Quinn (2012) have suggested that tackling ‘incidental’ sources of N and P should be a priority in order to prevent high loads and concentrations of these agricultural pollutants entering surface water courses unchecked. In the Eden catchment, the 2012 WFD classification data (<http://data.gov.uk/dataset/wfd-surface-water-classification-status-and-objectives>) identified that 46% of the area had achieved this status, 41% was classified as “good” and 13% as “poor” or “bad”.

Several mitigation options have been studied in detail in terms of either spatially targeted (engineered) features (e.g. Barber and Quinn, 2012; Wilkinson et al., 2014) to remediate hotspots and Contributing Source Areas (CSAs; as identified by Heathwaite et al., 2005; Pionke et al., 2000), or policies such as seasonally-implemented management measures including winter breaks in slurry and fertilizer applications (e.g. the closed winter periods implemented under the Nutrients Action Programme (NAP) in Ireland, Jordan et al., 2012). An assessment has been made of the improvements caused by these options. In terms of the NAP this was made by soil P status testing and measuring of in-stream water quality using bankside analysers (Jordan et al., 2012). However, one conclusion of their study was that the “flashiness” of the hydrology was a better predictor of the P loss during the closed periods than other more traditional measures such as soil P status and fertilizer application rates. In terms of engineered features, preliminary studies have identified that concentrations of NO₃, P and SS have been reduced based on observations of concentrations by constructed runoff attenuation features: e.g. the Lady’s Well feature at Belford (Barber and Quinn, 2012; Wilkinson et al., 2014). However, further research is urgently required to quantify the benefits of mitigation projects in terms of reducing loads and fluxes at the catchment scale, although obtaining flow measurements and pollutant concentrations can be problematic in the field (Lloyd et al., 2015).

More information from models is required in order to assess their impacts on different nutrient fluxes and pathways and to better target mitigation strategies for nutrient and sediment reduction in the UK/EU and elsewhere in the world. If newly developed models are capable of successfully predicting changes to pollutant dynamics, modelling could add value to these studies and assist in their design and implementation by predicting the number and density of mitigation features that need to be constructed. It should then be possible to quantify the benefits of these improvements in terms of specific land management practices, required in order to meet specific water quality improvements (e.g. WFD) at the catchment scale by reducing diffuse pollution (McGonigle et al., 2014).

Jackson-Blake et al. (2015) have questioned how models lacking parameter values relating to different land uses (i.e. spatially lumped) can be applied to mitigation measures, especially given a lack of

additional spatial monitoring data to assess model performance at a sub-catchment level. The Catchment Runoff Attenuation Flux Tool (CRAFT model; Adams et al., 2015) can explore the key concept of attenuation (and thus buffering rates) and its impact on water quality dynamics. Therefore, it can be used to assess the importance of attenuation on firstly catchment fluxes, then secondly potential management options, a feature not incorporated into the models reviewed above. The model can then be run to test different runoff hypotheses and management scenarios. Hypotheses testing against the robustness of time series of data can help to indicate both the dynamics and reproducibility of processes by a model (Jakeman et al., 2006). The model can also help to test the data information content, perhaps leading on to further improvements in order to incorporate management options such as mitigation measures (e.g. Runoff Attenuation Features – RAFs). Lastly, the model can indicate how these observations and impacts may be scaled up from the plot scale (where features are installed) to the catchment scale.

This study aims to model the delivery of fine/suspended sediment and phosphorus concentrations in a small, intensively monitored headwater catchment over a period of one year. This will enable: (i) the identification of pathway(s) that contribute the major sources of TRP, TP and SS exported from the catchment; and (ii) an examination of the potential for land management or mitigation to reduce fine sediment and nutrient exports through pollution-swapping (cf. Adams et al., 2015). The CRAFT model has been used in this study since: (i) its use allows hypotheses surrounding the major flow pathways in the catchment to be investigated; (ii) it can run on an hourly timestep thereby capturing within-storm processes (unlike INCA-P and SWAT which can only run on a daily timestep); and (iii) it enables the catchment exports of nutrients and sediments to be disaggregated into the amounts exported through each flow pathway (Adams et al., 2015). Modelling flow, nutrients and sediment is a challenge at this high frequency and many questions may arise from a detailed analysis of the data. Having a model that can simulate all the flux processes simultaneously, using a common set of soil and hydrological parameters, is also revealing in terms of model accuracy and structure. Both calibration and validation were performed using the continuous monitoring dataset. The modelling will be discussed in more detail in Section 2.2.

2. Methodology

2.1 Description of Case Study

We used data collected from the Newby Beck catchment in Cumbria, NW England (the NBC) as described by Perks et al (2015). The NBC forms a subcatchment of the much larger Eden catchment. The 2012 WFD classification data indicated that the NBC catchment had achieved “moderate” ecological status and the intention is for it to achieve “good” ecological status by 2027. It is 12.5 km² in area and has continuous monitoring of flow (Q) and water quality at the catchment outlet since 2011

(Fig. 1). In addition there are two tipping bucket raingauges and one weather station in the catchment also shown on Fig. 1 and the rainfall data used in the study is the average of the three values from these.

Figure 1 Base map of Newby Beck catchment showing land use, raingauges and contours (m AOD). Catchment outlet indicated by black star. Inset shows Eden catchment (shaded)

Firstly in terms of land use, the catchment is 90% improved grassland supporting intensive grazing by livestock, a mixture of dairy and beef cattle (Outram et al., 2014), see Fig. 1. There are also areas of farmland that are being reseeded or with feed (fodder) crops. The NBC consists of three categories of soils. In the headwaters to the south of the catchment are locally deep and well drained fine loamy soils. In the middle reaches of the catchment there are slowly permeable and seasonally waterlogged acid loamy to clay soils, exhibiting some degradation in structure. Finally, in the north of the catchment near the outlet there are reddish fine and coarse loamy soils (Cranfield University, 2014). Bedrock consists of a mixture of steeply dipping fractured limestone and sandstone units of the Carboniferous Yoredale group interbedded with argillaceous rocks (shales and mudstones) of lower permeability (Allen et al., 2010). This is overlain over most of the NBC by glacial till deposits. The long term average annual rainfall recorded in the NBC is 1187mm (Met Office, 2009), and the regional climate is described as cool temperate maritime.

The continuous monitoring station at the catchment outlet measured TRP (total reactive phosphorus) and TP (total phosphorus) using Hach Lange combined Sigmatax sampling module and Phosphax Sigma analyser analysers at 30-min intervals (Owen et al., 2012; Perks et al., 2015). TP incorporates all phosphorus species, whilst TRP is an operationally defined measurement predominantly comprising of orthophosphate (PO_4 ; SRP) in the NBC catchment (Perks et al., 2015), although readily hydrolysable P species in the sample may also be present within this TRP fraction (Halliday et al., 2014).

Turbidity was measured at 15-min intervals using a YSI 6600 multi-parameter sonde. A strong relationship (from regression analysis) between turbidity and SS (suspended sediments; collected by an autosampler at the NBC outlet during events) was identified by Perks et al. (2015), enabling turbidity to act as a proxy for SSCs. These data are used for the calculation of an “observed” yield from the catchment for modelling purposes. Loads and yields for the NBC were also calculated from the observed TRP and TP data and are shown in Table 1. During 2011-2 the mean annual TRP concentration was $0.041 \text{ mgL}^{-1} \text{ P}$ and the mean TP concentration $0.076 \text{ mgL}^{-1} \text{ P}$. The mean annual SS concentration calculated through regression of SS against turbidity data was 10.1 mgL^{-1} . Perks et al. (2015) also regressed TRP against SRP concentrations (obtained from manual and automatic sampling) and estimated SRP to constitute 89% of the observed TRP concentrations. Inspection of the observed hourly time series of TRP and TP indicated that many peaks in concentration coincided with peaks in flow, and these events will subsequently be termed “Type 1” and “behavioural”. Perks et al. (2015) indicated that near stream mobilisation of P would have generated these high concentrations of P during events. The information from the hourly data will be lost if aggregated to daily mean values.

2.1.1 Hydrology

The first year of the continuous monitoring dataset commencing on 1st October 2011 was chosen over which to calibrate and run model simulations. Analysis of the observed flows identified both dry and

wet periods during the twelve months, including runoff events in all four seasons. Runoff (specific discharge) Q was calculated by dividing the hourly flows by the catchment area. The total precipitation for 2011-2 was 1207 mm generating 709.5 mm of runoff (a runoff coefficient of 0.59). During the winter of 2011-2 most of this precipitation fell as rainfall and snow was uncommon and the winter and early spring were unusually dry (See Table 1 for the water balance), so subsequently it is referred to as “rain”. A second period of data was available for validating the model from late 2012 to early 2013. Unfortunately due to gaps in the continuous monitoring dataset it was not possible to validate the model using a year of P concentration data. Table 1 also shows the observed rainfall and runoff over this period which was also fairly wet with a high runoff coefficient.

2.2 Description of CRAFT model and Scenarios

The CRAFT model is used in this study to simulate three different runoff pathways and the associated export of sediment and nutrients (N and P) via each. The model has been developed with the concept of runoff attenuation (pollution mitigation) in mind enabling the effects of altering one, or multiple water/pollutant pathways on the export of sediment and nutrients (N and P) to be determined. A schematic diagram of the model is shown in Fig. 2. There are three stores each associated with a flow pathway (see Section 2.2.1 and Fig. 2). Rainfall is the principal factor generating surface runoff from the Dynamic Surface store when the drainage capacity of this store ($S_{D\text{MAX}}$) is exceeded by the rainfall rate. Therefore, (i) selecting an appropriate value for this parameter and (ii) the choice of timestep (daily vs. hourly) are very important as both of these will influence the model’s performance in terms of generating surface runoff.

Readers are directed to Adams et al. (2015) for a more detailed overview of the CRAFT model including equations for the runoff component. A full description of the model with equations has also been included in Appendix A1 in the Supplementary Material and a description of the parameters can be found in Table 3. In this application, observed TRP (C) data were available for comparison with the model results. Loads are summed at the catchment outlet according to Eq. 1a, firstly for TRP (L_{TRP})

$$L_{\text{TRP}} = Q_{\text{SS}} \times C_{\text{SS}}(\text{TRP}) + Q_{\text{GW}} \times C_{\text{GW}}(\text{TRP}) + K_{\text{SR}}(\text{TRP}) \times Q_{\text{SR}} \quad (1a)$$

Secondly, the TP load (L_{TP}) is calculated using Eq. 1b, where the particulate P (PP) load transported by surface runoff is added to the TRP load calculated above (here the assumption is made that the PP is unreactive, insoluble P)

$$L_{\text{TP}} = L_{\text{TRP}} + K_{\text{SR}}(\text{PP}) \times Q_{\text{SR}} \quad (1b)$$

Where Q_{SS} and $C_{\text{SS}}(\text{TRP})$ are the flow and concentration of TRP in the fast subsurface pathway respectively and Q_{GW} and $C_{\text{GW}}(\text{TRP})$ are the flow and concentration of TRP in the slow groundwater pathway respectively.

Figure 2 Schematic diagram of CRAFT model with SR attenuation store added

$K_{SR}(PP)$ and $K_{SR}(TRP)$ are coefficients relating the concentration of PP and TRP respectively in surface runoff to the instantaneous surface runoff (Q_{SR}) assuming that a linear concentration vs. discharge relationship applies. Suspended sediment (SS) loads are also calculated using Eq. 1a but in this case with the K_{SR} coefficient and the concentrations in the deep groundwater C_{GW} and fast subsurface C_{SS} taking calibrated values for SS rather than P.

The CRAFT runs within a MS ExcelTM interface (allowing the user to calibrate the model manually and investigate different model structures; i.e. an “Expert” mode). The NBC was not delineated into smaller units e.g. sub-catchments for modelling purposes, since the land use was dominated by improved grassland (Fig. 1) and it was assumed that a spatially lumped representation would be adequate, as has been the case in previous applications of this model to the Frome catchment. In the Frome application the model simulated runoff and pollutant concentrations at a daily time-step, thereby precluding the need to employ a delay function with it being assumed that the flood peaks reached the outlet (where

the gauging station was located) within one day (Adams et al., 2015). For this application however, pollutants are measured at 30-min intervals and the model was run on an hourly time step (the interval of the rainfall and flow data) therefore channel routing effects could not be ignored even in this small 12.5 km² catchment.

2.2.1 Attenuation Features of the CRAFT Model

The attenuation features of the model are associated with each of the three flow pathways and are key to the performance of CRAFT. Attenuation can be adjusted by the modeller in the three flow pathways by the following:

- (i) Surface Runoff - Increasing the maximum drainage rate from the surface store S_{DMAX} to permit more drainage to the two subsurface stores, and decreasing K_{SURF} to reduce the magnitude of the peak flows. This can represent better management of the cultivated layer of the soil.
- (ii) Fast Subsurface flow – Decreasing the time constant K_{SS} to reduce the peak value of this flow. This can represent adding attenuation to the field drainage network.
- (iii) Slow Groundwater flow - Decreasing the time constant K_{GW} to lengthen recession periods in order to store water in the subsurface for a longer period of time.

It was imperative that the issue associated with routing flow through the drainage network was appropriately tackled, due to the potential for an incorrect choice of timestep introducing timing errors depending on the speed of propagation of the flood wave down the river system. It was necessary therefore to introduce an additional component that could attenuate surface runoff (as this was identified as the most important of the three flow pathways in terms of controlling the runoff dynamics during events and floods).

2.2.2 Surface Runoff Attenuation Component

Surface runoff (SR) attenuation has been achieved through the addition of a SR attenuation store, which uses a simple linear storage function to delay the surface runoff in reaching the catchment outlet based on that used by the AWBM model (Boughton, 2004). Nutrients and sediments that travel via the surface runoff pathway are also subject to attenuation, and here the assumption is made that the same delay (i.e. time lag) applies. The lagged surface runoff Q_{SRLAG} is calculated from the storage in the SR attenuation store (S_{ATT}) by Eq. (2)

$$Q_{SRLAG} = S_{ATT} \times (1 - K_{LAG}) \quad (2)$$

Where K_{LAG} is the time coefficient, representing the fraction of the SR attenuation store that drains at each timestep. Clearly, a value of unity will never permit any surface runoff, whereas a value of zero will not generate any lag (by draining the attenuation store in one timestep which is the model's default mode for a daily timestep e.g. Adams et al., 2015). The store is assumed to have infinite capacity and

is empty at the start of the simulation. A mass balance updates the value of S_{ATT} at the current times (t) from the value at the previous one ($t-1$)

$$S_{ATT}(t) = S_{ATT}(t-1) + Q_{SR} - Q_{SRLAG} \quad (3)$$

To simulate the attenuation of sediments and nutrients the same procedure is adopted where flows are replaced by loads, e.g. $L_{SR}(PP)$ representing the load of particulate P transported by surface runoff which is defined by Eqs. 1a and 1b (see above). The attenuation parameter K_{LAG} can either be calibrated using the same method (e.g. using a criteria such as maximising the Nash-Sutcliffe efficiency (NSE)) as the other runoff parameters, or alternatively by visually comparing the shapes of peaks in the observed and modelled water quality concentration time series (i.e. “chemigraphs” and “sedigraphs”) and their timing, as K_{LAG} has a negligible effect on total runoff volumes or nutrient loads. The hypotheses that firstly a lag is required and secondly that different lags may apply to runoff and the different water quality variables will be explored further through a series of hypothetical model simulations.

2.2.3 Runoff Attenuation Features

The attenuation capability of CRAFT can also be used to simulate runoff attenuation features (e.g. swales and wetlands) that are constructed in the catchment as part of mitigation schemes (e.g. Wilkinson et al. 2014, Barber, 2013, Barber and Quinn, 2012, Ockenden et al., 2012). In this case a removal (or trapping) efficiency e (for surface runoff) has to be specified by the user based on data collected in the field from monitoring flow and transport through mitigation features (Barber, 2013; Ockenden et al., 2012). Fast subsurface flow can be attenuated by decreasing the value of the K_{SS} parameter to reduce the flashiness of the flow, and the flow rate can be reduced by decreasing K_{SPLIT} , to represent improved soil management and potentially pollution swapping. It is assumed for modelling purposes that the surface runoff attenuation factor K_{LAG} is the same for flow and the modelled sediments and nutrients, however the efficiency may vary (e.g. be higher for particulate nutrients than dissolved nutrients) if the user has sufficient information for this. The basic form of the removal rate equation where $L_{SR(ATT)}$ is the load of sediment or nutrient transported by the surface runoff pathway into the attenuation store is thus written as

$$L_{SR(ATT)} = L_{SR} \times (1-e) \quad (4)$$

A further publication will investigate how the field data can be used to parameterize the values of efficiencies for N, P and sediment removal from different types of mitigation features based on data collected from field studies (e.g. Ockenden et al., 2012; Barber, 2013).

2.2.4 Model Calibration and Validation

For the manual calibration procedure used here there are 6 parameters (described in Table 3) for the runoff component that require calibrating plus K_{LAG} . These 6 are K_{SURF} , S_{DMAX} , S_{RMAX} , K_{SPLIT} , K_{GW} and K_{SS} (cf. Adams et al., 2015). The water quality component uses Equations 1a and 1b with the related

parameters for P and SS described above and in Table 3 to calculate loads, then concentrations are simply calculated by dividing the loads by the total flow. To simplify the model calibration process the ratios between $K_{SR}(TRP)$, $K_{SR}(PP)$ and $K_{SR}(SS)$ were assumed to be fixed, thus reducing by two the number that require calibration. Evidence from the observed concentration time series at Newby Beck outlet of TRP, TP and SS was used to verify that this interdependence was plausible (R^2 values between these time series > 0.7 were calculated). Perks et al. (2015) observed that peak event concentrations of TP were generally double that of TRP as measured at the catchment outlet.

The performance metrics used to assess the simulations were: (i) for runoff: mass balance error (MBE) and Nash–Sutcliffe efficiency (NSE); (ii) for water quality: load error (LE), NSE (SS only); normalised $1/RMSE$; (iii) visual fit (by eye) for runoff and water quality. These metrics were calculated on each model time series of flow and concentration by the MS ExcelTM interface, and the multi-criteria approach proposed by Jackson-Blake et.al (2015) was used rather than merely depending on a single metric to discriminate between different model runs.

Model validation was carried out using the observed data (See Section 2.1.1). The same suite of performance metrics were used to evaluate the model as were used in the calibration procedure but the model parameters themselves were not adjusted.

2.2.5 Alternative Model Structure Hypotheses

The CRAFT model was developed using the Minimum Information Required (MIR) philosophy (Quinn et al, 2008). In principle MIR models are based on the amount of information which is obtained from localised and experimental studies on nutrient and sediment losses, so that the most pertinent process components can be retained in the model. A series of simulations were first carried out to determine single, optimal parameter sets for the default structure plus each of the options discussed below. These structures were:

- (1) The default with no SR attenuation component, i.e. K_{LAG} was set to zero. The parameter values were optimised to give the best fit visually to flows whilst achieving the maximum possible NSE and smallest MBE values for flows.
- (2) Using the SR attenuation component to apply a lag to the default model simulation, and calibrating an optimum value of the parameter K_{LAG} for predicting runoff and concentration. This is referred to as the “lagged” model structure and the resulting simulation as the “lagged” simulation. A sensitivity analysis on the effects of varying K_{LAG} on the NSE for each model output (flow and concentrations) was carried out as part of the calibration procedure.
- (3) Testing the lagged model with the K_{SURF} , $K_{SR}(PP)$ and $K_{SR}(TRP)$ parameter values set to values intended to maximise the generation of PP and TRP in surface runoff, with the performance assessment being primarily made visually by inspecting the time series plots for goodness of fit (possibly at the expense of poorer NSE and MBE/LEs). Employing the SR attenuation

component was considered to be important here to “lump and route” P to the outlet, hence this model simulation will be referred to as “LR”.

The 1/RMSE metric was used in this study following the findings of a full uncertainty analysis of the INCA-P model by Dean et al. (2009). They found it difficult to obtain positive NSE values for TP when simulating the Lugg catchment, leading to the requirement for an alternative method to the NSE metric to assess water quality model performance. This metric has a positive value for all simulations, and its value increases as the model error decreases (i.e. for a “best fit” a high value is desirable). Here it is normalised by first dividing the RMSE by the observed mean concentration of the nutrient or sediment so that a higher value indicates better model performance when comparing different simulations. Wellen et al. (2015) also noted, from a review of several hundred modelling studies, that the NSEs were usually lower for water quality simulations than for runoff alone, and furthermore that the 25th percentile NSE values were lower for TP simulations than for phosphate (SRP).

The loads of SS, TRP and TP were also broken down by flow pathway (surface runoff, fast subsurface and deep groundwater) to enable the different model structures to be compared to see if there were any differences between the flow pathways and for comparison against the observed load.

3. Results

3.1 Different Model Structures

3.1.1 Calibration of Default model

The results obtained using the default model are shown firstly for the entire 2011-12 calibration period in Fig. 3 as time series plots of modelled and observed runoff with rainfall also shown by the blue line. Secondly, in Figs. 4a-d the runoff (Q – 4a) and modelled and observed concentrations (TRP - 4b, TP - 4c and SS -4d) at the NBC outlet are shown for August 2012 only (since this month had some interesting events which will be further discussed below). In Fig. 3, the time series plot of modelled and observed runoff indicates a good model fit over most of the year with a NSE value of 0.82 (Table 2). At higher flows, some of the peaks were under or over estimated by the model and this can be seen more readily in Fig. 4a along with a timing error of 1-2 hours between the modelled and observed flow peaks, which generally occur shortly after rainfall events. There was also a tendency for the model to generate runoff after every rainfall event, where peaks were not necessarily observed, particularly in the summer months. Perks et al. (2015) reported that a time-to-peak of three hours had been estimated from the observed runoff. Surface runoff (SR) accounted for 16% of the total runoff over the year according to the default model. The baseflow index (BFI) according to the gauged flow record was 0.39 according to Ockenden et al. (2016) which is in broad agreement with the partitioning of flows by the model (deep groundwater accounted for 50% of the total runoff). Figure 4 also shows daily mean flows and

concentrations for comparison with the hourly observed data, at a daily interval most of the information is lost from the dataset by averaging.

Figure 3 Time series plot of hourly modelled and observed runoff (Q) and rainfall (Rain) for entire modelled period 1/10/2011 – 30/9/2012, default model simulation

3.1.2 Validation of Default model

The model results for the validation period are summarised in Table 2. Additional time series plots of the modelled and observed Q, TRP, TP and SS concentrations can be found in the Supplementary Material (Fig 9). In summary, the results for the validation period for runoff were almost as good as for the calibration period with only a small decrease in the NSE (to 0.78) and an increase in the MBE indicating that the model was overpredicting runoff depth by nearly 5%. The performance in terms of reproducing the observed nutrient and SS concentrations (the NRMSE and NSE metrics for TRP and SS) was still acceptable, although the load errors (for TP and SS) were higher than desirable. The NSE values for both TRP and SS concentrations did not reduce significantly either from those obtained during the calibration period. Interestingly the NRMSE metric indicated (for Q and P) a slightly better fit over the validation period than over the calibration period.

3.1.3 Results from Alternative model structures

In order to assess visually the results obtained from the different model structures, some additional results from these runs are shown in Figs. 4a-d for Q, TRP, TP and SS. The parameter values, determined by manual calibration (evaluating both the default and the alternative model structures) are listed in Table 3. In the “LR” simulation the K_{SURF} parameter was increased so that SR accounted for 20% of the total runoff. However, it was not possible to distinguish any major differences in the runoff predictions compared to the default structure over a 1 year period by using a graphical method alone, although there were clear differences in the timing of the runoff peaks in the lagged and “LR” simulations (later than in the default simulation) which can be seen in Fig. 4a.. Therefore, the

performances of these alternative model structures (at predicting runoff) were compared against the default CRAFT model using the metrics MBE and NSE, which are also shown in Table 2.

Figure 4 Time series plots of Q (4a), TRP (4b), TP (4c) and SS (4d) during August 2012 indicating model predictions (red line) and observed hourly values (solid black line) from different model structures/simulations. Red ovals on TP and TRP panes indicate Type 2, i.e. “non behavioural” events. Hourly rainfall is also shown in Fig 4a by the blue line. The dashed black lines indicate observed daily mean data (denoted by the suffix “D” in the Legends).

The time series plots of flows, nutrient and sediment concentrations covering August 2012 illustrate several issues with the observed data and the different model structures. The plots in Figs. 4b and 4c

show that most of peaks in TRP and TP coincided with the peaks in surface runoff (in Fig. 4a) and are thus classed as Type 1 (“behavioural”) events. This finding supports the hypothesis that surface runoff is the major flow pathway for P export in the NBC also proposed by Perks et al. (2015). Visually, the simulations with a lag term applied have fitted the observed time series of TRP and TP better than the default with no lag, although there was still a tendency to underpredict or overpredict some of the observed event concentrations.

In summer 2012 in particular some peaks were observed in TRP and TP concentrations that were outside the predictive range of the calibrated model, thus these can be termed “non-behavioural” Type 2 events. One Type 2 event in particular is highlighted a by red circle (in early August 2012) in Figs 4b and 4c. In this period the observed peaks in TRP and TP concentrations were not associated with correspondingly high observed runoff (and also the modelled runoff from the SR component in the CRAFT). The “LR” simulation, which in Figs. 4b and 4c is shown to generate higher concentrations of both TRP and TP, was not able to reproduce this observed peak either. These results may also have implications for modelling the catchment in general terms and identifying the significant flow and transport pathways. This will be discussed further below in Section 4.2.

3.2 Load Breakdown

In order to evaluate the modelled flow pathways the loads transported by each pathway were calculated (i.e. from the time series of flow and concentrations). These are converted to annual exports (i.e. loads per unit area per year) to enable comparison to be made with other studies and the observed values (shown in the top left pane of Fig. 5 by the green bars). The observed SRP load was estimated to be 89% of the TRP load based on data analysed by Perks et al. (2015), and the dashed black line indicates this value. The observed total exports of TP, TRP and SS and the modelled exports from the default simulation over the calibration period are shown in Table 1. The yields of TRP and TP exported via the three model pathways are shown in Fig. 5 by bar charts, one from each of the different model structures. Note that the yields from the lagged and default models were identical so only two panes of results are shown. The modelled PP yields exported by the surface runoff (SR) pathway were the largest of the three flow pathways transporting P (at around 60-70% in all three simulations), with the “LR” having the largest of all the simulations. The modelled SS yields broken down by flow pathway from all three simulations are not shown graphically, since the fraction transported by the SR pathway accounted for around 70-75% of the total export from all three of the simulations.

Figure 5 Breakdown of modelled TRP and TP yields from 2011-2 by different flow and transport pathways for each model structure/simulation (Bottom pane), and observed TRP, SRP and TP yields (Top pane) for comparison.

4. Discussion

The choice of the CRAFT model to simulate runoff, sediment and nutrient generation in the NBC has been justified by the model results after calibration, in this case using an hourly timestep. The MBEs

of less than $\pm 2\%$ and NSE values of at least 0.7 in terms of predicting runoff, indicated that the performance of the model was satisfactory. In terms of predicting P and SS concentrations, positive NSE values and LEs of less than $\pm 10\%$ were also obtained from the default model structure for SS, TP and TRP. The model is therefore suitable for assessing future scenarios relating to the effects of mitigation measures in the NBC catchment. The model performance compares favourably with other distributed and physically based models widely used by the nutrient and sediment modelling community (Wellen et al., 2015). Jackson –Blake et al. (2015) have discussed the issue of assessing water quality model performance (specifically INCA-P) and suggested that a “weight of evidence” approach is used that includes a visual comparison of modelled and observed time series. They have also pointed out that performance that is considered acceptable in terms of predicting flows (e.g. a $NSE > 0.65$ being considered as “good”) may be difficult to achieve when predicting P concentrations from agricultural catchments where a much lower NSE may suffice. This study has achieved NSE values of > 0.2 for TRP and TP, > 0.3 for SS vs. a NSE value of circa 0.8 for runoff, which fits neatly into their evaluation of model performance at simulating flow, nutrients and sediments.

In terms of model validation, over a short period in winter 2012-3 the model results were evaluated and model predictions were still very good for runoff and acceptable for P and SS. It must be stressed that there may be unknown issues with using such a short time period to validate the water quality component of the model. Also, there may have been step changes in the concentration and turbidity data obtained by the bankside analysers which has not been picked up by the QA/QC procedure. Therefore, both the observed P concentrations and turbidity values (used to obtain the observed SS timeseries) may have shifted upwards and downwards respectively. In general, validation must be viewed cautiously when used with high-frequency water quality monitoring data due to these limitations and issues.

In terms of resolving the uncertainty issue, it is assumed here that expert judgement on what constitutes a “best fit” model is more important when simulating sediments and nutrients than attempting to quantify uncertainty directly as was attempted by Holloway et al. (*this issue*) using SWAT without fully resolving the issue.

4.1 Comparison of Different Model Structures

In terms of predicting runoff, Table 2 shows that there was little difference in the NSE and MBE values from the lagged and default model structures and that these metrics alone do not really discriminate between these. This reflected that it was relatively straightforward to calibrate the CRAFT model to the hourly flow data and achieve NSE values circa 0.75-0.8 both with and without a lag term. Higher NSE values were achievable in the default simulation (up to 0.83) at the expense of a poorer visual fit to the hydrographs, with a tendency to underpredict the peak runoff during events. In terms of the modelled outputs (P and SS), Figure 6 shows that the default model structure generally performed the best out of

all the model structures in terms of achieving the best NSE and LE values with one or two exceptions that will be discussed in Section 4.1.1.

Figure 6 Results of comparing different model structures as assessed by (a top) load error LE (b bottom) NSE performance metrics for TRP, TP and SS predictions

4.1.1 Introduction of SR Attenuation Component

The introduction of a SR attenuation (lag) component in the CRAFT has improved (according to the NSE values shown in Fig. 6) the prediction of flow and P but has had a detrimental effect on the model's prediction of SS (the NSE decreased from 0.37 to 0.3 as shown in Fig. 6). Fig. 7 explores the attenuation feature of the model further by varying K_{LAG} between 0.25 and 0.8 for each modelled variable in turn and calculating the increase or decrease in the NSE (an increase resulting in a positive y-axis value)

relative to the default simulation with a zero lag. It can be seen through the shape of the 4 different curves that each of these variables would benefit from different optimal values of K_{LAG} and that TP benefitted most from a high value of K_{LAG} with an optimal value of 0.75 (the value chosen for the lagged model simulation). Perks et al. (2015) analysed a series of events over 2012-3 that included the second half of the twelve month period analysed in our study. They found that there were different hysteresis patterns in the observed event SS and P dynamics. SS was mobilized by events with clockwise hysteresis indicating that there were near-stream sources of sediment that were readily mobilised by surface runoff during storms. The evidence from both Fig 4d, which shows that visually the peaks in late –August 2012 were in fact best predicted by the default model (compared to both the lagged and “LR” models), and Fig 7, which shows that setting K_{LAG} to less than 0.25 produced the highest NSE value for SS, supports this finding and is in agreement with the findings of Perks et al. (2015).

Figure 7 Graph showing the effect of varying the degree of SR attenuation (K_{LAG}) for flow (Q) , TP, TRP and SS. Positive values indicate an improvement in the NSE relative to the value achieved by the default simulation with $K_{LAG} = 0$

Visually, (Fig 4b and 4c respectively) the results from the “LR” simulation also indicated that the transport of TP and TRP by the SR pathway can be better modelled using this method of lumping a large load of P into this pathway, then attenuating its transport to the outlet, rather than restricting the routing of both TP and TRP to the outlet to take place over a single hour. The purpose of examining the “LR” version of the lagged model structure has also been to illustrate that the model can better reproduce some of the “Type 1 behavioural” peaks of P exhibited by the NBC using this technique. In terms of

the visual fit to the TP and TRP observed concentrations during events in summer 2012, the “LR” simulation performed the best of the three. Figs. 4b and 4c show that in August 2012 this model was able to fit the observations reasonably well except for one “non-behavioural” Type 2 event at the start of the month. These results were however obtained at the expense of a higher LE (i.e. model overprediction) compared to the default model structure. There may be an issue here in terms of the “fitness for purpose” of the model, and the best way of obtaining a “good fit” to the observed data.

Both simulations incorporating attenuation (lagged and “LR”) generated a significant delay in the time to peak (of the time series shown in Figs. 4b-d) of several hours compared to the default simulation (where runoff and nutrient peak concentrations were predicted to occur one hour after the rainfall peak). The importance of this behaviour may be that without any attenuation the modelled peaks in concentration are simply too “spiky” and also have been predicted to occur several hours before the observed peaks. Again, Fig. 7 shows that using the attenuation term with $K_{LAG} > 0.25$ improved the goodness of fit compared to the “default” simulation for both TP and TRP. This delay may be due to the anti-clockwise hysteresis patterns identified by Perks et al. (2015) from the observed flow and concentration patterns, which was attributed to soil water being the dominant flow pathway (equivalent to the fast subsurface flow pathway in the CRAFT).

The SR attenuation component does lower the concentrations but does not affect the load exported. The observed SS concentrations do not demonstrate the influence of attenuation and hence the predicted SS concentrations do not benefit from the SR attenuation component. The SS appears to be more dominated by ‘on-off’ dynamics that are very sensitive to the rainfall and surface runoff rates. Suspended sediment concentrations may well be reducing in surface runoff whilst P concentrations are still high in the near surface of the soil and in land drains, as stated by Perks et al. (2015).

4.1.2 Conceptual Model of P and SS Dynamics

It is postulated here that surface runoff also entrains high concentrations of P through runoff entering and leaving the upper soil layers that are heavily disturbed and compacted by agricultural activities (i.e. intensive grazing and ploughing). The overall impact on catchment load or yield of these different patterns and the model’s ability to capture them is probably less important, so calibrating the model to fit the peak concentrations is probably the key here if it is to be a useful tool for examining mitigation options.

Figure 8 below shows a conceptual sketch of fluxes (flow, P and SS) observed at the NBC outlet during an event. Evidence from the observed data (TP and TRP flux time series) suggests that a peak in TRP is observed after a peak in TP at the catchment outlet and it is assumed that the flow pathways that transport TRP and SRP are identical given no evidence to support a more complex mechanism (TRP flux is shown by the red line). The flux timeseries of PP is also shown by the dashed blue line and this

peak is coincident with the peaks in both surface runoff (Q_{SR}) (top pane) and suspended sediment (bottom pane). The TP flux (green line) is thus the sum of the PP and TRP fluxes.

Figure 8. Conceptual sketch showing fluxes during an event in NBC (top: flow; middle: P; bottom; SS).

We are thus implying (in Fig. 8) that the SS dynamics are sensitive to short bursts of fully connected surface runoff, however the P dynamics are not dominated by this flow pathway alone and a fast subsurface flow pathway including land drains or grips could account for the attenuation shown in the TRP time series (Dils and Heathwaite, 1999) apparent both in Fig. 8 and in the observed time series of concentration in Fig. 4 from which the conceptual model shown in Fig. 8 is derived. This “on-off” effect, generating and transporting PP and SS via surface runoff, is sensitive to rainfall rates and generates rapid, connected surface runoff (often termed “overland flow”) that may switch off and re-infiltrate when the rainfall rate declines, hence the surface runoff becomes rapidly disconnected from the watercourses. Conversely should the rainfall rates increase, near-surface flow could lead to surface runoff restabilising and generating increased SS losses. There is assumed to be a secondary peak in suspended sediment flux from the fast subsurface pathway (i.e. in conjunction with Q_{ss}) which may be due to rapid drainage of SS through the field drainage network. Suspended sediment fluxes via the slow groundwater pathway are assumed to be negligible.

Hence a model is needed that can capture both the “on-off” spiky nature of surface runoff and the more attenuated subsurface pathways for P. The data information content and current analysis do not reveal if the P losses are occurring in the near surface of the soil or through interactions with the land drains (or both). However, the soils in the area are prone to water logging and land drains are installed to alleviate this problem (see Dils and Heathwaite (1999) for an overview of the role of drains in transporting P in agricultural catchment) so the two processes are probably closely linked. The attenuation term is useful and the impact on attenuation on flow concentration is strong, and may be relevant to policy makers.

4.1.2 Using a suite of metrics to assess model performance

In comparing the two different metrics NSE and normalised 1/RMSE to assess model performance, it was clear that in this study that the use of a NSE was able to distinguish reasonably well between the performance of the different model structures at predicting both P and SS. The normalised 1/RMSE metric, as used by Dean et al. (2009), was less discriminatory for TRP than the NSE metric (values were within $\pm 16\%$ for TRP from the three simulations). The normalised 1/RMSE metric could be useful though where the NSE values from all runs are negative (as was reported by Dean et al. (2009)), however any model predictions made (using these parameter sets with NSE values below zero) will be worse than using the observed mean as the predictor of flow or concentration. Visual methods of assessing model performance at predicting concentrations of P were also deemed acceptable alongside using metrics such as NSE, correlation and bias by Jackson-Blake et al. (2015), and evidence from the results of this study suggests that using a suite of model assessment metrics may be the best method in the NBC as well.

4.2 Load breakdown by pathway

The default model can generate high fluxes of P immediately following events via the SS component (by employing high values of $C_{SS}(TRP)$) and high concentrations and fluxes during events in surface runoff (by employing high values of $K_{SR}(PP)$). In “behavioural” events it is assumed that peak Cs were positively correlated with peak flows and occurred at approximately the same time, as was observed in the other Eden subcatchments by Barber (2013), Mills and Bathurst (2015), Ockenden et al. (2016). Both mechanisms have a plausible physical basis in representing the wet areas of the catchment that become connected to the stream channel network during and after runoff events and have a readily mobilized source of TRP and PP (e.g. from farmyards and hardstandings) (Outram et al., 2014). A second source of P includes the drainage of enriched soil water via a subsurface field drain system (i.e. a fast subsurface pathway) (Perks et al., 2015) and the entrainment of particulate and colloidal sources of P from the degraded soil surface and tracks (Outram et al., 2014), which is classified as a surface runoff pathway.

The modelled load breakdown (in the default simulation) of P shown in Fig 5 also indicated that the surface runoff pathway accounted for 64% of the modelled TP export and 72% of the modelled SS export. Perks et al. (2015) suggested that physical runoff interception would be the best method of targeting the surface runoff pathway that is by far the largest source of SS according to the model. Observed TRP and TP yields (Fig.5 top pane) indicated that reactive P only constituted approximately 40% of the TP yield from the NBC. This indicates that there must be a large source of unreactive P, e.g. organic P being mobilised in the catchment which may be either particulate or soluble. Moreover, the lag observed in the observed TRP concentration due to attenuation could have implications when selecting mitigation options for this catchment.

4.3 Seasonality and non-behavioural events

Rainfall and runoff were fairly uniformly distributed during the 12 months in 2011-2 except for a dry spell in late winter and early spring, with the runoff coefficient for the entire period being high (0.59). Any seasonality in the year analysed (2011-2) was not especially pronounced in terms of runoff generation. The dry spell in between these periods may have reduced the runoff ratio somewhat in late spring-early summer as later on during summer 2012 it was above 0.7 due to the wet conditions. There were also at least two Type 2 non-behavioural events observed during summer 2012. These events were not significant in terms of TRP and TP export as the flows during the events were fairly low. The relatively high observed P concentrations (TP: $1.0 \text{ mgL}^{-1} \text{ P}$, TRP: $0.35 \text{ mgL}^{-1} \text{ P}$) were -likely a result of random agricultural activities in the catchment, which could be either point sources e.g. wash-down operations at dairy farms and piggeries, or diffuse sources such as badly-timed applications of slurry to the fields, and as such are almost impossible to model. Suspended sediment also exhibited a

correspondingly high C peak in early August (Fig. 4d) that was not reproduced by the model, which supports the hypothesis that slurry applications may have been the cause (Ockenden et al., 2016).

4.4 Comparison with Other Studies

A previous study of sediment export in the Eden catchment by Mills and Bathurst (2015) found that the subcatchments monitored by the CHASM project exported between 4 and 73 t km⁻² yr⁻¹ of SS with no clear relationship between size and yield. In comparison with their findings, the SS export of 38.2 t km⁻² yr⁻¹ (for the entire 12 month period) from the NBC (calculated from this study) was similar, when compared to a subcatchment of a similar size to Newby Beck (Helm Beck; 18 km²: 46 t km⁻² yr⁻¹), but higher than the export from Swindale Beck (16 km²: 26 t km⁻² yr⁻¹). The authors suggested that sediment yield varied considerably between the smaller subcatchments in the Eden due to heterogeneity in the rates of both sediment supply and transport.

Barber (2013) analysed a dataset comprising both TP and SRP samples collected by grab and automatic sampling (autosamplers) from the same Eden subcatchments monitored by the CHASM project. The mean export of TP from the upper Eden catchment was 42.4 kg km⁻² yr⁻¹, and the export of SRP was 14.5 kg km⁻² yr⁻¹. These totals were calculated for the period 2010-2011. The export rates from the 9 km² Blind Beck subcatchment were the highest out of all the subcatchments, with the 2011 totals being higher than 2010 (precipitation was higher at 1429 mm versus 779 mm in 2010). The NBC was not included within the nested CHASM subcatchments, however it has exported similar amounts of TP, TRP and SS for the period analysed here (Table 1) probably due to having similar land uses, soils and climate.

5. Conclusions

The NBC catchment in the upper Eden, NW England has been intensively monitored since 2011 with the objective of understanding nutrient and sediment pathways and transfer, with the ultimate aim of developing and field testing strategies to reduce diffuse pollution via mitigation. Export rates of P and sediments (SS) over the 12 months from October 2011 were not especially high compared to previous studies in the upper Eden catchment, where land use and climate are similar to this one, a dry spell in late winter may have contributed to this.

Modelling of this dataset is important in order to assess the impact of future mitigation plans on sediment and nutrient export. For this purpose the CRAFT model has been evaluated. The hypothesis testing reported above indicated that the default CRAFT model structure is appropriate at simulating both flow and water quality in the NBC for most applications, but with some drawbacks in terms of phosphorus (P) simulation in particular due to in-stream routing effects. Water quality models are often assessed in terms of their performance at predicting loads, and also over a daily or even longer time

period rather than by using hourly concentration data. In this study we showed that concentration data could also be acceptably reproduced by the CRAFT model using an hourly timestep, which has rarely been reported elsewhere by modelling studies.

Any seasonality over the 12 month period assessed in the NBC was not particularly evident as runoff was generated at all times of the year. The continuous monitoring of nutrients using bankside analysers has identified that there was probably more than one occasion (in summer 2012) when farming activities may have generated peak P concentrations that appear to be outside the predictive range of the default CRAFT model structure. This assumes that high concentrations of P and SS are generated from high surface runoff rates using a linear relationship between Q and C which may not always hold true, and this could be enhanced in future versions of the model if the observations support a more complex relationship.

We assessed a revised version of the CRAFT model structure incorporating a surface runoff attenuation component (i.e. lag) to delay the peaks in flow and concentrations (of P and SS) in reaching the outlet. These results were interesting. The revised model structure produced better results for flow than the default model in terms of the visual fit to the data (assessed via time series plots) since timing errors were addressed, however the NSE reduced slightly. In terms of modelling both TP and TRP, a better fit was achieved by adding attenuation to the model. A further model simulation, including attenuation, termed LR (standing for “lump and route”) generated much higher concentrations of both PP and TRP during runoff events and then routed these to the outlet using the attenuation component. The results obtained from this simulation were acceptable in terms of matching the peak TP and TRP concentrations visually, but as a result overpredicted the TP export from the NBC by 56%.

For simulating the future effects of mitigation features however, the attenuation capabilities of the CRAFT model will be highly useful and should be tested with a wide range of model parameter values including those obtained from the LR simulation. The attenuation of both surface runoff and fast subsurface flow should be considered as a result of these features introducing a lag to the system and also if improved farming practices have resulted in a decrease in surface runoff due to improved soil conditions. These measures will be required if the catchment is to achieve “Good” ecological status by 2027 as targeted. A conceptual understanding of both hydrology and diffuse pollution (sources and pathways) points towards improved management regimes. It is important that water quality models are transparent to end users and can be applied to the assessment of mitigation measures and the effect of pollution swapping (changing the dominant flow pathways) in the catchment. The breakdown by (modelled) load pathways in the NBC indicated that surface runoff will transport up to 90% of SS and 60% of P making this the key pathway that needs to be targeted by mitigation measures in this particular catchment. A further study will examine the use of the extended attenuation capabilities of the CRAFT model to simulate the behaviour of these mitigation features in the NBC.

6. Acknowledgements

This study was carried out using field data supplied by the Eden Demonstration Test Catchment (Eden DTC) research platform. The Eden DTC is funded by the Department for Environment, Food and Rural Affairs (Defra), project WQ0210, although the lead author was not funded by or involved directly with this Project.

7. References

- Adams, R., Quinn, P.F., Bowes, M.J. 2015. The Catchment Runoff Attenuation Flux Tool, A minimum information requirement nutrient pollution model. *Hydrol. Earth Syst. Sci.*, 19(4), 1641-1657.
- Allen, D.J., Newell, A.J., Butcher, A.S., 2010. Preliminary review of the geology and hydrogeology of the Eden DTC sub-catchments. BGS Groundwater Science Programme, Open Report OR/10/063, 45pp.
- Barber, N. J., Quinn, P. F., 2012. Mitigating diffuse water pollution from agriculture using soft-engineered runoff attenuation features. *Area* 44(4), 454-462.
- Barber N., 2013. Sediment, nutrient and runoff management and mitigation in rural catchments. Unpubl. PhD Thesis, Newcastle University, School of Civil Engineering and Geosciences. 447pp.
- Beven, K., 2009. Environmental modelling: an uncertain future? An introduction to techniques for uncertainty estimation in environmental prediction. Routledge, Abingdon, Oxon (UK).
- Bicknell, B. R., Imhoff, J. C., Kittle, Jr., J. L., Donigan, Jr., A. S., Johanson, R. C., 1996. Hydrologic Simulation Program - FORTAN (HSPF): User's Manual for Release 11; U.S. EPA Environmental Research Lab: Athens, GA, USA.
- Bilotta, G.S., Brazier, R.E., Haygarth, P.M., 2007. Processes affecting transfer of sediment and colloids, with associated phosphorus, from intensively farmed grasslands: erosion. *Hydrol. Process.* 21, 135–139.
- Bilotta, G.S., Krueger, T., Brazier, R.E., Butler, P., Freer, J., Hawkins, J.M.B., et al., 2010. Assessing catchment-scale erosion and yields of suspended solids from improved temperate grassland. *J. Environ. Monit.* 12, 731–739.
- Binger, R. L.; Theurer, F. D.; and Yuan, Y., 2011. AnnAGNPS Technical Processes. Version 5.2, <http://www.ars.usda.gov/research/docs.htm?docid=5222>. Last accessed 17/11/15
- Birkinshaw, S. J., Ewen, J., 2000. Modelling nitrate transport in the Slapton Wood catchment using SHETRAN. *J. Hyd.* 230(1), 18-33.
- Boughton, W., 2004. The Australian water balance model. *Environ. Model. Soft.* 19, 943–956.
- Bowes, M.J., House, W.A., Hodgkinson, R.A., 2003. Phosphorus dynamics along a river continuum. *Sci. Total Environ.* 313, 199–212.
- Cranfield University, 2014. The Soils Guide. Cranfield University.

- Dean, S., Freer, J., Beven, K., Wade, A.J. Butterfield, D., 2009. Uncertainty assessment of a process-based integrated catchment model of phosphorus. *Stoch. Environ. Res. Risk. Assess.* 23:991-1010.
- Dils R.M., Heathwaite A.L., 1999. The controversial role of tile drainage in phosphorus export from agricultural land. *Water Science Technology*, 39, 55–61.
- Gassman, P.; Reyes, M.; Green, C.; Arnold, J., 2007. The Soil and Water Assessment Tool: Historical development, applications, and future research directions. *Trans. ASABE*, 50 (4), 1211–1250.
- Halliday, S., Skeffington, R., Bowes, M., Gozzard, E., Newman, J., Loewenthal, M., et al., 2014. The water quality of the River Enborne, UK: observations from high frequency monitoring in a rural, lowland river system. *Water* 6, 150–180.
- Haygarth, P.M., Wood, F.L., Heathwaite, A.L., Butler, P.J., 2005. Phosphorus dynamics observed through increasing scales in a nested headwater-to-river channel study. *Sci. Tot. Env.* 344: 83-106. doi:10.1016/j.scitotenv.2005.02.007
- Heathwaite, A.L., Quinn, P.F., Hewett, C.J.M., 2005. Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. *J. Hyd.* 304: 446-461. doi:10.1016/j.jhydrol.2004.07.043.
- Holloway, M.J., Beven, K.J. et al., Evaluating a water quality model on a small UK headwater catchment: defining limits of acceptability. *Sci. Tot. Environ.* *Submitted*.
- Jackson-Blake, L. A., Dunn, S. M., Helliwell, R. C., Skeffington, R. A., Stutter, M. I., Wade, A. J., 2015. How well can we model stream phosphorus concentrations in agricultural catchments? *Environmental Modelling and Software*, 64, 31-46.
- Jakeman, A. J., Letcher, R. A., Norton, J. P. 2006., Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software*, 21(5), 602-614.
- Jordan, P., Melland, A. R., Mellander, P. E., Shortle, G., Wall, D., 2012. The seasonality of phosphorus transfers from land to water: Implications for trophic impacts and policy evaluation. *Sci. Tot. Environ.* 434, 101-109.
- Lloyd, C. E. M., Freer, J. E., Johnes, P. J., Coxon, G., Collins, A. L., 2015. Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams. *Hydrological Processes*.
- McGonigle, D.F., Burke, S.P., Collins, A.L., Gartner, R., Haft, M.R., Harris, R.C., et al., 2014. Developing Demonstration Test Catchments as a platform for transdisciplinary land management research in England and Wales. *Environ. Sci. Process. Impacts* 16, 1618–1628.
- Mellander, P.-E., Melland, A.R., Jordan, P., Wall, D.P., Murphy, P.N.C., Shortle, G., 2012. Quantifying nutrient transfer pathways in agricultural catchments using high temporal resolution data. *Environ. Sci. Pol.* 24, 44–57.
- Met Office, 2009. UKCP09: Average Annual Rainfall Dataset.
- Mills, C.F., Bathurst, J.C., 2015. Spatial variability of suspended sediment yield in a gravel-bed river across four orders of magnitude of catchment area. *Catena* 133, 14-24.

- Nasr, A., Bruen, M., Jordan, P., Moles, R., Kiely, G., Byrne, P., 2007. A comparison of SWAT, HSPF and SHETRAN/GOPC for modelling phosphorus export from three catchments in Ireland. *Wat. Res.* 41(5), 1065-1073.
- Ockenden et al. 2016. Potential effects of changing climate on catchment processes and nutrient transfers: Evidence from high temporal concentration-flow dynamics in headwater catchments. *Sci. Tot. Environ.* 548: 325-339.
- Ockenden, M. C., Deasy, C., Quinton, J. N., Bailey, A. P., Surridge, B., Stoate, C., 2012. Evaluation of field wetlands for mitigation of diffuse pollution from agriculture: sediment retention, cost and effectiveness. *Env. Sci. Pol.* 24, 110-119.
- Outram, F.N., Lloyd, C.E.M., Jonczyk, J., Benskin, C.M.H., Grant, F., Perks, M.T., et al., 2014. High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in England. *Hydrol. Earth Syst. Sci.* 18, 3429–3448.
- Owen, G. J., Perks, M. T., Benskin, C. M. H., Wilkinson, M. E., Jonczyk, J., Quinn, P. F. 2012. Monitoring agricultural diffuse pollution through a dense monitoring network in the River Eden Demonstration Test Catchment, Cumbria, UK. *Area* 44(4), 443-453.
- Perks, M.T., Owen, G.J., Benskin, C.M.H., Jonczyk, J., Deasy, C., Burke, S., Reaney, S.M. and Haygarth, P.M., 2015. Dominant mechanisms for the delivery of fine sediment and phosphorus to fluvial networks draining grassland dominated headwater catchments. *Sci. Tot. Environ.* 523(0): 178-190.
- Pionke, H.B., Gburek, W.J., Sharpley, A.N., 2000. Critical source area controls on water quality in an agricultural watershed located in the Chesapeake Basin. *Ecol. Eng.* 14: 325-335. doi:10.1016/S0925-8574(99)00059-2
- Quinn, P., Hewitt, C.J.M., Dayawansa, N.D.K. 2008. TOPCAT-NP: a minimum information requirement model for simulation of flow and nutrient transport from agricultural systems, *Hydrol. Process.* 22(14):2565-2580.
- Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., Butterfield, D., Rankinen, K., Lepisto, A., 2002. A nitrogen model for European catchments: INCA, new model structure and equations. *Hydrol. Earth Syst. Sci.* 6 (3), 559–582. DOI: 10.5194/hess-6-559-2002.
- Wellen, C., Kamran-Disfani, A. R., Arhonditsis, G. B. 2015. Evaluation of the Current State of Distributed Watershed Nutrient Water Quality Modeling. *Env. Sci. Tech.* 49(6), 3278-3290.
- Wilkinson, M.E., Quinn, P.F., Barber, N.J., Jonczyk, J., 2014. A framework for managing runoff and pollution in the rural landscape using a Catchment Systems Engineering approach. *Sci. Tot. Environ.* 468–469, 1245–1254.