Hydrology and
Earth System
Sciences

Uncertainty assessment of a dominant-process catchment model of dissolved phosphorus transfer

Rémi Dupas[1], **Jordy Salmon-Monviola**[1], **Keith J. Beven**[2], **Patrick Durand**[1], **Philip M. Haygarth**[2], **Michael J. Hollaway**[2], and **Chantal Gascuel-Odoux**[1]

[1]INRA, Agrocampus Ouest, UMR1069 SAS, 35000 Rennes, France
[2]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

Correspondence to: Rémi Dupas (remi.dpas@gmail.com)

Abstract. We developed a parsimonious topography-based hydrologic model coupled with a soil biogeochemistry sub-model in order to improve understanding and prediction of soluble reactive phosphorus (SRP) transfer in agricultural headwater catchments. The model structure aims to capture the dominant hydrological and biogeochemical processes identified from multiscale observations in a research catchment (Kervidy–Naizin, 5 km^2). Groundwater fluctuations, responsible for the connection of soil SRP production zones to the stream, were simulated with a fully distributed hydrologic model at 20 m resolution. The spatial variability of the soil phosphorus content and the temporal variability of soil moisture and temperature, which had previously been identified as key controlling factors of SRP solubilization in soils, were included as part of an empirical soil biogeochemistry sub-model. The modelling approach included an analysis of the information contained in the calibration data and propagation of uncertainty in model predictions using a generalized likelihood uncertainty estimation (GLUE) "limits of acceptability" framework. Overall, the model appeared to perform well given the uncertainty in the observational data, with a Nash–Sutcliffe efficiency on daily SRP loads between 0.1 and 0.8 for acceptable models. The role of hydrological connectivity via groundwater fluctuation and the role of increased SRP solubilization following dry/hot periods were captured well. We conclude that in the absence of near-continuous monitoring, the amount of information contained in the data is limited; hence, parsimonious models are more relevant than highly parameterized models. An analysis of uncertainty in the data is recommended for model calibration in order to provide reliable predictions.

1 Introduction

Excessive phosphorus (P) concentrations in freshwater bodies result in increased eutrophication risk worldwide (Carpenter et al., 1998; Schindler et al., 2008). Eutrophication restricts economic use of water and poses a serious hazard to ecosystems and humans (Serrano et al., 2015). In western countries, reduction of point-source P emissions in the last 2 decades has resulted in a proportionally increasing contribution of diffuse sources, mainly from agricultural origins (Alexander et al., 2008; Grizzetti et al., 2012; Dupas et al., 2015a). Of particular concern are dissolved P forms, often measured as soluble reactive phosphorus (SRP) because they are highly bioavailable and therefore a likely contributor to eutrophication.

To reduce SRP transfer from agricultural soils, it is important to identify the spatial origin of P sources in agricultural landscapes, the biogeochemical mechanisms causing SRP solubilization in soils and the dominant transfer pathways, as well as the potential P resorption during transit. Research catchments provide useful data to investigate SRP transport mechanisms: typically, the temporal variations in water quality parameters at the outlet, together with hydro-climatic variables, are investigated to infer spatial origin and dominant transfer pathways of SRP (Haygarth et al., 2012; Outram et al., 2014; Dupas et al., 2015b; Mellander et al., 2015; Perks et al., 2015). Hypotheses drawn from analysis of water quality time series can be further investigated through hillslope monitoring and/or laboratory experiments (Heathwaite and Dils, 2000; Siwek et al., 2013; Dupas et al., 2015c). When dominant processes are considered reasonably known,

it is possible to develop computer models, for two main purposes. First, to validate scientific conceptual models one can test whether model predictions can produce reasonable simulations compared to observations. Of particular interest is the possibility of testing the capability of a computer model to upscale P processes observed at fine spatial resolution (soil column, hillslope) to a whole catchment. Second, if the models survive such validation tests, they can be useful tools to simulate the response of a catchment system to a future perturbation such as changes in agricultural management and climate changes.

However, process-based P models generally perform poorly compared to, for example, nitrogen models (Wade et al., 2002a; Dean et al., 2009; Jackson-Blake et al., 2015). This is of major concern because poor model performance suggests poor knowledge of dominant processes at the catchment scale, and poor reliability of the modelling tools used to support management. The origin of poor model performance might be conceptual misrepresentations, structural imperfection, calibration problems, irrelevant model evaluation criteria and difficulties in properly assessing the information content of the available data when it is subject to epistemic error. All five causes of poor model performance are intertwined; e.g. model calibration strategy depends on model performance evaluation criteria, which depend on the way the information contained in the observation data is assessed (Beven and Smith, 2015).

A key issue in environmental modelling is the level of complexity one should seek to incorporate in a model structure. Several existing P transfer models, such as INCA (Wade et al., 2002a), SWAT (Arnold et al., 1998) and HYPE (Lindstrom et al., 2010) seek to simulate many processes, with the view that complex models are necessary to understand processes and to predict the likely consequences of land use or climate changes. However, these complex models include many parameters that need to be calibrated, while the amount of data available for calibration is often low. An imbalance between calibration requirement and the amount of available observation data can lead to equifinality issues, i.e. when many model structures or parameter sets lead to acceptable simulation results (Beven, 2006). A consequence of equifinality is the risk of unreliable prediction when an "optimal" set of parameters is used (Kirchner, 2006), and large uncertainty intervals when Monte Carlo simulations are performed (Dean et al., 2009). In this situation, it will be worth exploring parsimonious models that aim to capture the dominant hydrological and biogeochemical processes controlling SRP transfer in agricultural catchments. For example, Hahn et al. (2013) used a soil-type-based rainfall–runoff model (Lazzarotto et al., 2006) combined with an empirical model of soil SRP release derived from rainfall simulation experiments over soils with different P content and manure application level/timing (Hahn et al., 2012) to simulate daily SRP load from critical sources areas.

A second key issue, linked to the question of model complexity, concerns model calibration and evaluation. Both calibration and evaluation require assessing the fit of model outputs with observation data. However, observation data are generally not directly comparable with model outputs, because of incommensurability issues and/or because they contain errors (Beven, 2006, 2009). Typically, predicted daily concentrations and/or loads are evaluated against data from grab samples collected on a daily or weekly basis. The information content of these data must be carefully evaluated to propagate uncertainty in the data into model predictions (Krueger et al., 2012). Uncertainty in grab sample data might stem from (i) sampling frequency problems or (ii) measurement problems (Lloyd et al., 2016). Grab sample data represent a specific point in the stream cross section, which can differ from the cross section mean concentration (Rode and Suhr, 2007), and a snapshot of the concentration at a given time of the day, which can differ from the flow-weighted mean daily concentration (McMillan et al., 2012). This difference between observation data and simulation output can be large during storm events in small agricultural catchments, as P concentrations can vary by several orders of magnitude during the same day (Heathwaite and Dils, 2000; Sharpley et al., 2008). Model evaluation can be severely hindered by this difference, because many popular evaluation criteria such as the Nash–Sutcliffe efficiency (NSE) are sensitive to extreme values and errors in timing (Moriasi et al., 2007). During baseflow periods, it is more likely that grab sample data are comparable to flow-weighted mean daily concentrations, as concentrations vary little during the day and they are usually low in the absence of point sources. However, measurement errors are expected to occur at low concentrations, either due to too long storage times or laboratory imprecision when concentrations come close to detection/quantification limits (Jarvie et al., 2002; Moore and Locke, 2013). Uncertainty in the data can also relate to discharge measurement and input data (e.g. maps of soil P content and rainfall data). In this paper we strive to identify and quantify the different sources of uncertainty in the data when the required quality check tests have been performed (on the discharge and SRP concentration data). A generalized likelihood uncertainty estimation (GLUE) "limits of acceptability" approach (Beven, 2006; Beven and Smith, 2015) is used to calibrate/evaluate the model.

This paper presents a dominant-process model that couples a topography-based hydrologic model with a soil biogeochemistry sub-model able to simulate daily discharge and SRP loads. The dominant processes included in the hydrologic and soil biogeochemistry sub-models have been identified in previous analyses of multiscale observational data, which have demonstrated, on the one hand, the control of groundwater fluctuation on connecting soil SRP production zones to the stream (Haygarth et al., 2012; Jordan et al., 2012; Dupas et al., 2015b, d; Mellander et al., 2015), and, on the other hand, the role of antecedent soil moisture and
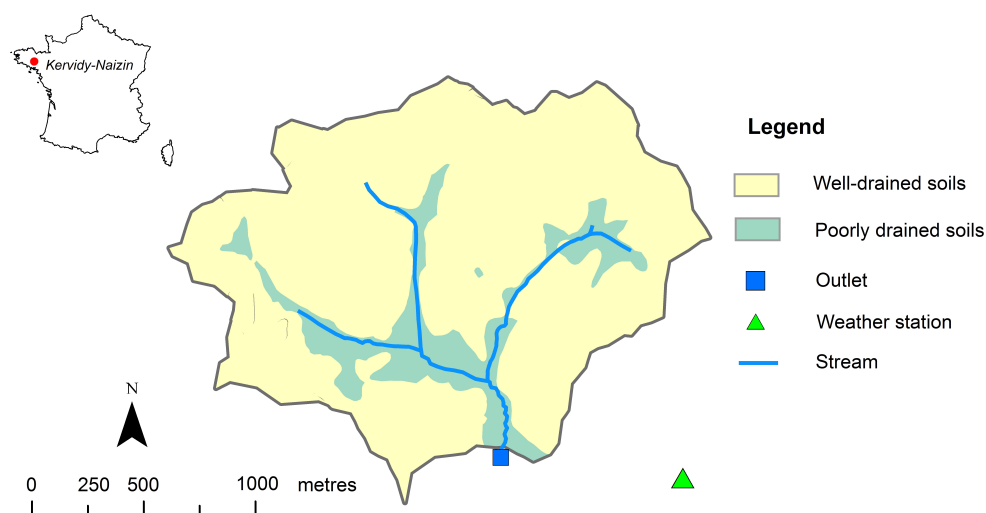
Figure 1. Soil drainage classes in the Kervidy–Naizin catchment (Curmi et al. 1998).

temperature conditions on SRP solubilization in soils (Turner and Haygarth, 2001; Blackwell et al., 2009; Dupas et al., 2015c). Model development and application were performed in the Kervidy–Naizin catchment in western France with the objectives of (i) testing if the model was capable of capturing daily variation of SRP load, thus confirming hypotheses on dominant processes; and (ii) developing a methodology to analyse and propagate uncertainty in the data into model prediction using a "limits of acceptability" approach.

2 Material and methods

2.1 Study catchment

2.1.1 Site description

Kervidy–Naizin is a small ($4.94\,\mathrm{km}^2$) agricultural catchment located in central Brittany, western France (48° N, 3° W). It belongs to the AgrHyS environmental research observatory (http://www6.inra.fr/ore_agrhys_eng), which studies the impact of agricultural activities and climate change on water quality (Molenat et al., 2008; Aubert et al., 2013; Salmon-Monviola et al., 2013; Humbert et al., 2015). The catchment (Fig. 1) is drained by a stream of second Strahler order, which generally dries up in August and September. The climate is temperate oceanic, with mean ± standard deviations of annual cumulative precipitation and specific discharge of 854 ± 179 and $290 \pm 106\,\mathrm{mm}$ respectively, from 2000 to 2014. Mean annual ± standard deviation of temperature is $11.2 \pm 0.6\,°\mathrm{C}$. Elevation ranges from 93 to 135 m above sea level. Topography is gentle, with maximum slopes not exceeding 5 %. The bedrock consists of impervious, locally fractured Brioverian schists and is capped by several metres of unconsolidated weathered material and silty, loamy soils. The hydrological behaviour is dominated by the de-

velopment of a water table that varies seasonally along the hillslope. In the upland domain, consisting of well-drained soils, the water table remains below the soil surface throughout the year, varying in depth from 1 to > 8 m. In the wetland domain, developed near the stream and consisting of hydromorphic soils, the water table is shallower, remaining near the soil surface generally from October to April each year. The land use is mostly agriculture, specifically arable crops and confined animal production (dairy cows and pigs). A farm survey conducted in 2013 led to the following land use subdivisions: 35 % cereal crops, 36 % maize, 16 % grassland and 13 % other crops (rapeseed, vegetables). Animal density was estimated as high as 13 livestock units ha^{-1} in 2010. Estimated soil P surplus was $13.1\,\mathrm{kg\,P\,ha^{-1}\,yr^{-1}}$ (Dupas et al., 2015b) and soil extractable P in 2013 (Olsen et al., 1954) was $59 \pm 31\,\mathrm{mg\,P\,kg^{-1}}$ ($n = 89$ samples). A survey targeting riparian areas highlighted the legacy of high soil P content in these currently unfertilized areas (Dupas et al., 2015c). No point-source emissions were recorded, but scattered dwellings with septic tanks were present in the catchment.

2.1.2 Hydroclimatic and chemical monitoring

Kervidy–Naizin was equipped with a weather station (Cimel Enerco 516i) located 1.1 km from the catchment outlet. It recorded hourly precipitation, air and soil temperatures, air humidity, global radiation, wind direction and speed, which are used to estimate Penman evapotranspiration. Stream discharge was estimated at the outlet with a rating curve and stage measurements from a float-operator sensor (Thalimèdes OTT) upstream of a rectangular weir.

To record both seasonal and within storm dynamics in SRP concentration, two monitoring strategies complemented each other from October 2013 to August 2015: a daily man-

ual grab sampling at approximately the same time (between 16:00 and 18:00 local time) and automatic high-frequency sampling during 14 storm events (Teledyne autosampler ISCO 6712 Full-Size Portable Sampler; 24 1 L bottles filled every 30 min). The water samples were filtered on-site, immediately after grab sampling and after 1–2 days in the case of autosampling. They were analysed for SRP (ISO 15681) within a fortnight. To assess uncertainty in daily SRP concentration related to sampling time, storage and measurement errors, a second grab sample was taken at a different time of the day (between 11:00 and 15:00 local time) in 36 instances during the study period. The second sample was analysed within 24 h with the same method; this second data set is referred to as verification data set, as opposed to the reference data set. Among the 36 pairs of comparable daily samples, 12 were taken during storm events and 24 during baseflow periods. To assess uncertainty in high-frequency SRP concentration during storm events due to delayed filtration of autosampler bottles, five grab samples were taken during the course of four distinct storms and were filtered immediately. The same lab procedure was used to analyse SRP.

2.1.3 Identification of dominant processes from multiscale observations

Observations in the Kervidy–Naizin catchment have highlighted that the temporal variability in stream SRP concentrations could not be related to the calendar of agricultural practices but rather to hydrological and biogeochemical processes (Dupas et al., 2015b). The primary control of hydrology on SRP transfer has also been evidenced in several other small agricultural catchments (e.g. Haygarth et al, 2012; Jordan et al., 2012; Mellander et al., 2015). In the Kervidy–Naizin catchment, the groundwater fluctuation in valley bottom areas was identified as the main driving factor of SRP transfer, through the hydrological connectivity it creates when the saturated zone intercepts shallow soil layers (Dupas et al., 2015b).

In situ monitoring of soil pore water at four sites (15 and 50 cm depths) in the Kervidy–Naizin catchment has shown that mean SRP concentration in soils is a linear function of Olsen P (Olsen et al., 1954). This reflects the current knowledge that a soil P test, or alternatively estimation of a degree of P saturation, can be used to assess solubilization in soils (Beauchemin and Simard, 1999; McDowell et al., 2002; Schoumans et al., 2015). This linear relationship derived from the data contrasts however with other studies, where threshold values above which SRP solubilization increases greatly have been identified (Heckrath et al., 1995; Maguire and Sims, 2002).

SRP solubilization in soil varies seasonally according to antecedent conditions of temperature and soil moisture. Dry and/or hot conditions are favourable to the accumulation of mobile P forms in soils, while water-saturated conditions lead to their flushing (Turner and Haygarth, 2001; Blackwell et al., 2009; Dupas et al., 2015c).

2.2 Description of the Topography-based Nutrient Transfer and Transformation – Phosphorus model (TNT2-P)

TNT2 was originally developed as a process-based and spatially explicit model simulating water and nitrogen fluxes at a daily time step (Beaujouan et al., 2002) in meso-scale catchments ($<50 km^2$). TNT2-N has been widely used for operational objectives, to test the effect of mitigation options proposed by local stakeholders or public policymakers (Moreau et al., 2012; Durand et al., 2015), on nitrate fluxes and concentrations in rivers.

TNT2-P uses a modified version of the hydrological sub-model in TNT2-N, to which a P biogeochemistry sub-model was added to simulate SRP solubilization in soils.

2.2.1 Hydrological sub-model

The assumptions in the hydrological sub-model are derived from TOPMODEL, which has previously been applied to the Kervidy–Naizin catchment (Bruneau et al., 1995; Franks et al., 1998). (1) The effective hydraulic gradient of the saturated zone is approximated by the local topographic surface gradient ($\tan\beta$). It is calculated in each cell of a digital elevation model (DEM) at the beginning of the simulation. (2) The effective downslope transmissivity (parameter T) of the soil profile in each cell of the DEM is a function of the soil moisture deficit (S_d). Hydraulic conductivity is assumed to decrease exponentially with depth (parameter m, Fig. 2). Hence, water fluxes (q) are computed as

$$q = T \cdot \tan\beta \cdot \exp(-\frac{S_d}{m}). \tag{1}$$

Based on these assumptions, TNT2 computes an explicit cell-to-cell routing of fluxes, using a D8 algorithm.

To simulate SRP fluxes, the hydrological sub-model is used to compute water fluxes from each soil layer by integrating Eq. (1) between the maximum depth of the soil layer considered and either

– estimated groundwater level, if the groundwater table is within the soil layer considered,

or

– the minimum depth of the soil layer considered, if the groundwater table above the soil layer considered.

In this application of the TNT2-P model, five soil layers with a thickness of 10 cm are considered. Hence, seven flow components are computed in the model:

– overland flow on any saturated surfaces;

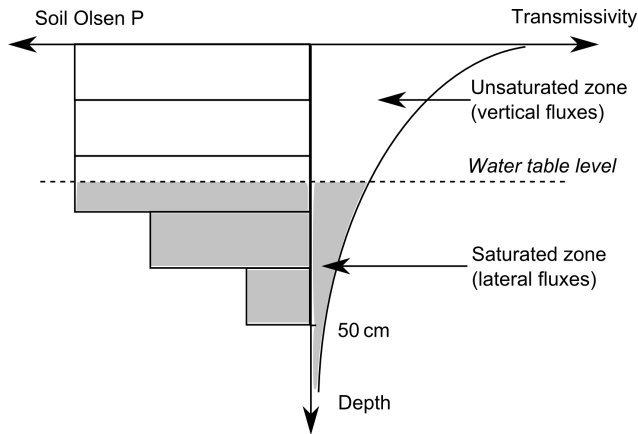– five sub-surface flow components, one for each soil layer;

Figure 2. Description of soil hydraulic properties and phosphorus content with depth.

– deep flow, i.e. flow below the five soil layers.

2.2.2 Soil P sub-model

The soil P sub-model is empirically derived from soil pore water monitoring data (Dupas et al., 2015c), specifically assuming that

– the background SRP concentration in the soil pore water of a given layer is proportional to soil Olsen P;

– seasonal increases in P availability compared to background conditions are determined by biogeochemical processes, controlled by antecedent temperature and soil moisture. Data show that SRP availability in the soil pore water increases following periods of dry and hot conditions (Dupas et al., 2015c).

Hence, SRP transfer is modelled with parameters that describe both mobilization and transfer to the stream. A different parameter is used to simulate transfer via overland flow and sub-surface flow.

$$F_{\text{SRP overland}} = \text{Coef}_{\text{SRP overland}} \cdot P_{\text{Olsen}} \cdot q_{\text{overland}}, \quad (2)$$

$$F_{\text{SRP sub-surface}} = \text{Coef}_{\text{SRP sub-surface}} \cdot P_{\text{Olsen}} \cdot q_{\text{sub-surface}}, \quad (3)$$

where $F_{\text{SRP overland}}$ and $F_{\text{SRP sub-surface}}$ are SRP transfer via overland flow and sub-surface flow for a given soil layer respectively, q_{overland} and $q_{\text{sub-surface}}$ are water flows from the same pathways. $\text{Coef}_{\text{SRP overland}}$ and $\text{Coef}_{\text{SRP sub-surface}}$ are coefficients that vary according to antecedent temperature and soil moisture conditions, such as

$$\text{Coef}_{\text{SRP}} = \text{Coef}_{\text{background}} \cdot (1 + F_T \cdot F_S), \quad (4)$$

where Coef_{SRP} is either $\text{Coef}_{\text{SRP overland}}$ or $\text{Coef}_{\text{SRP sub-surface}}$, and F_T and F_S are temperature and soil moisture factors, respectively. F_T and F_S are expressed as

$$F_T = \exp(\frac{\text{mean}\,(\text{temperature},\ i\,\text{days}) - T1}{T2}), \quad (5)$$

$$F_S = 1 - \left(\frac{\text{mean}\,(\text{water content},\ i\,\text{days})}{\text{maximum water content}}\right)^{S1}, \quad (6)$$

where $T1$, $T2$ and $S1$ are parameters to be calibrated. The antecedent condition time length consists in a period of $i = 100$ days. Both soil temperature and soil moisture are estimated by the TNT2 soil module (Moreau et al., 2013). Because soil moisture in the deep soil layers can differ significantly from that of shallow soil layers, two values of F_S are calculated for two soil depth ranges: 0–20 and 20–50 cm. The temperature factor F_T was calculated as an average value for the entire 0–50 cm soil profile. Contrary to the water fluxes, SRP fluxes are not routed cell-to-cell because we lack knowledge of the rate of SRP re-adsorption in downslope cells and of the long-term fate of re-adsorbed SRP. Hence, all the SRP emitted from each cell through overland flow and sub-surface flow reaches the stream on the same day. For deep flow, only the immediate riparian flux is used in determining SRP inputs to the river.

No long-term depletion of the different P pools was modelled, because annual P export from the catchment was small compared to the size of soil and sub-soil P pools.

2.2.3 Input data and parameters

Spatial input data required for TNT2-P include

– A DEM in raster format. Here, a 20 m resolution DEM was used; hence, model calculations were made in 12 348 grid cells covering a 4.94 km^2 catchment.

– A map of soil units that could be assumed to have homogeneous hydrological parameter values, in raster format. Here, two soil classes were considered by differentiating well-drained (86 %) and poorly drained soils (14 %) according to Curmi et al. (1998) (Fig. 1). Experimental determination of saturated hydraulic conductivity (29 soil cores) by Curmi et al. (1998) showed significantly different values for soils classified as well-drained and poorly drained in the Kervidy–Naizin catchment. The two units were treated as homogeneous, lacking information about the detailed variability in soil hydraulic characteristics at the model grid scale.

– A map of surface Olsen P in raster format and description of decrease in the Olsen P with depth for five soil layers between 0 and 50 cm. Here, the map of Olsen P in the 0–15 cm soil layer was obtained from statistical modelling with the rule-based regression algorithm CUBIST (Quinlan, 1992) using data from 198 soil samples (2013) in an area of 12 km^2 encompassing the 4.94 km^2 catchment (Matos-Moreira et al., 2015). To describe how Olsen P decreases with depth, land use

information was used. In tilled fields, i.e. all crop rotations including arable crops, Olsen P was assumed to be constant between 0 and 30 cm and to decrease linearly with depth between 30 and 50 cm. In no-till fields, i.e. permanent pasture and woodland, Olsen P was assumed to decrease linearly with depth between 0 and 50 cm. An exponential decrease with depth is more commonly adopted in untilled land (e.g. Haygarth et al., 1998; Page et al., 2005), but a specific sampling in currently untilled areas in the Kervidy–Naizin catchment (Dupas et al., 2015c) has shown that a linear function is more appropriate, probably because of these areas having been ploughed in the past. A previous study has shown that soil Olsen P was the most important factor controlling SRP solubilization in soils of the Kervidy–Naizin catchment (see Sect. 2.1.3); therefore, other parameters in the soil P sub-model (Sect. 2.2.2) were treated as homogeneous in the catchment (the soil classification into well-drained and poorly drained soils only concerned hydrological parameters).

A 20 m resolution was chosen for the DEM and the soil Olsen P raster map to allow for a detailed representation of the interaction of the groundwater table (as simulated by the hydrological model) and the soil Olsen P (as given by the soil Olsen P map). Indeed the soil saturation and soil Olsen P can be very different in a narrow zone close to the stream compared to upslope due to the presence of a 5 to 50 m unfertilized buffer zone with lower Olsen P compared to fertilized fields. The Olsen P value close to the stream has a determining influence on SRP transfer, because this area is the most frequently connected to the stream; therefore, a coarser resolution of the raster maps would degrade representation of the system.

Climate input data include minimum and maximum air temperature, precipitation, potential evapotranspiration and global radiation on a daily basis. The TNT2 model allows for several climate zones to be considered, in which case a raster map of climate zones must be provided to the model. Here, only one climate zone is considered.

In total, the TNT2-P model includes 15 parameters for each soil type, i.e. 30 parameters in total if two soil drainage classes are considered. To reduce the number of model runs necessary to explore the parameter space using Monte Carlo simulations, several parameters were given fixed values, or a constant ratio between the two soil types was set (Table 1). In the hydrological sub-model, the parameters to vary were identified in a previous sensitivity analysis (Moreau et al., 2013). In the soil sub-model, all the parameters were varied.

Finally, only 12 parameters were varied independently (see Table 1). Initial parameter ranges for the hydrological sub-model were based on values from several previous studies in western France (Moreau et al., 2013) and those for the soil sub-model were based on a preliminary manual trial and error procedure. The SRP concentration for deep flow water
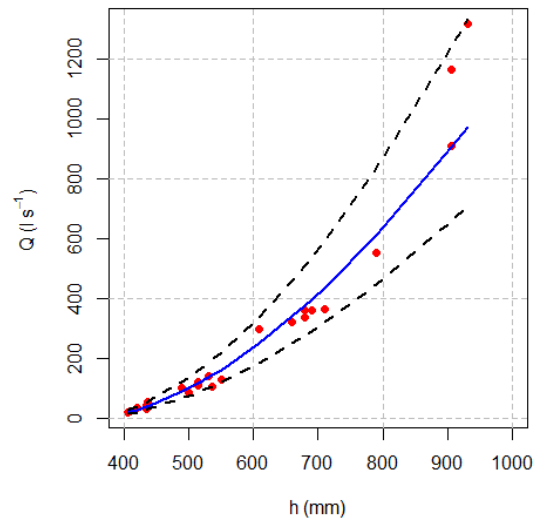

Figure 3. Rating curve in Kervidy–Naizin; acceptability bounds derived from 90 % prediction interval (blue line: fitting regression; black dashes: 90 % prediction interval). Red dots represent the original discharge measurements used to calibrate the stage–discharge rating curve (Carluer, 1998).

was based on actual measurement of SRP in the weathered schist (Dupas et al., 2015c). A constant flux value for domestic sources was set at the 1st percentile of the daily flux between 2007 and 2013 (Dupas et al., 2015b).

2.3　Deriving limits of acceptability from data uncertainty assessment

The Monte Carlo-based GLUE methodology has been widely used in hydrology and is described elsewhere (Beven and Freer, 2001b; Beven, 2006, 2009). Briefly, the rationale of GLUE is that many model structures and parameter sets can give "acceptable" results, according to one or several performance measures. Hence, GLUE considers that all models that give acceptable results should be used for prediction. A key issue in GLUE is to decide on a performance threshold to define acceptable models; typically, modellers set a threshold value of a measure such as the Nash–Sutcliffe efficiency based on their subjective appreciation of data uncertainty or on previously used values. To allow for a more explicit justification of the performance threshold values used, the limits of acceptability approach outlined by Beven (2006) relies on an assessment of uncertainty in the calibration/evaluation data. According to this approach, all model realizations that fall within the limits of acceptability are used for prediction, weighted by a score calculated based on overall performance.

Details on how the limits of acceptability for daily discharge and daily SRP load were derived from uncertainty assessment of the observational data are presented below. Input data, such as weather and soil Olsen P data, also contained uncertainties that were not accounted for explicitly in

Table 1. Initial parameter ranges in the hydrological and soil phosphorus sub-models.

	Abbreviation	Unit	Hydrological (H) or phosphorus (P) model	Range poorly drained soils (min–max)	Range well-drained soils (min–max)
Lateral transmissivity at saturation	T	$m^2\,day^{-1}$	H	4–8	$\rightarrow \times 1.5$
Exponential decay rate of hydraulic conductivity with depth	m	$m^2\,day^{-1}$	H	0.02–0.2	0.02–0.2
Soil depth	ho	m	H	0.3–0.8	$\rightarrow \times 1$
Drainage porosity of soil	po	$cm^3\,cm^{-3}$	H	0.1–0.4	$\rightarrow \times 1$
Regolith layer thickness	h1	m	H	5–10	$\rightarrow \times 4$
Exponent for evaporation limit	A	–	H	8 (fixed)	$\rightarrow \times 1$
kRC parameter for capillary rise	kRC	–	H	0.001 (fixed)	$\rightarrow \times 1$
n parameter for capillarity rise	N	–	H	2.5 (fixed)	$\rightarrow \times 1$
Drainage porosity of regolith layer	p1	$cm^3\,cm^{-3}$	H	0.01–0.05	$\rightarrow \times 1$
Background P release coefficient for sub-surface flow	$Coef_{SRP\ overland}$	–	P	0–0.015	$\rightarrow \times 1$
Background P release coefficient for overland flow	$Coef_{SRP\ sub\text{-}surface}$	–	P	0–0.25	$\rightarrow \times 1$
Temperature coefficient 1	T1	–	P	5–10	$\rightarrow \times 1$
Temperature coefficient 2	T2	–	P	2–10	$\rightarrow \times 1$
Soil moisture coefficient	$S1$	–	P	0–2	$\rightarrow \times 1$
SRP concentration in deep flow	SRP_deep	$mg\,L^{-1}$	P	0–0.007	$\rightarrow \times 1$

the limits of acceptability due to a lack of data to quantify them.

2.3.1 Discharge

Error in discharge measurement data was assessed from the original discharge measurements used to calibrate the stage–discharge rating curve (Carluer, 1998). The rating curve used in this study was

$$Q = a \cdot (h - h_0)^b, \tag{7}$$

where Q is discharge, h is stage reading, h_0 is stage reading at zero discharge, a and b are calibrated coefficients. Limits of acceptability were defined as the 90 % prediction interval of log–log linear regression (Fig. 3). The acceptability range estimated in this way was $\pm 39\%$ on average. This uncertainty interval is in the higher range of values found in other studies, e.g. Coxon et al. (2015), who found that mean discharge uncertainty was generally between 20 and 40 % in 500 catchments of the United Kingdom. This relatively large uncertainty interval is due to the fact that it was derived from a prediction interval rather than a confidence interval (the 90 % confidence interval of the log–log linear regression would be 14 % of the mean discharge value during the study period). A prediction interval is an interval in which future observations will likely fall, whereas a confidence interval is an interval in which the mean of repeated observation will likely fall. Because in the TNT2-P model's evaluation we want each observation to fall in the acceptability interval (Sect. 2.3.3), a prediction interval was more appropriate. For daily discharge

values below $2\,mm\,day^{-1}$, fixed acceptability limits were set at the 90 % prediction interval for a stage measurement corresponding to $2\,mm\,day^{-1}$.

2.3.2 SRP load

Uncertainty in "observed" daily load includes uncertainty in discharge (see Sect. 2.3.1) and uncertainty in SRP concentration. The acceptability limit for daily load was estimated by the sum of relative uncertainty assessed for discharge and SRP concentration (in percentage). Uncertainty in SRP concentration stems from sampling frequency problems as one grab sample collected on a specific day is incommensurable with the mean daily concentration or load simulated by the model. Further, measurement errors exist that include the effect of storage time (Haygarth et al., 1995). During baseflow periods, measurement error was expected to be the main source of uncertainty because relative measurement error was large for low concentrations, especially when sample storage time exceeded 48 h (Jarvie et al., 2002), whereas concentrations vary little. During storm events, sampling frequency was expected to be the main source of uncertainty because SRP concentration can vary by 1 order of magnitude within a few hours. Therefore, different acceptability limits were set for both flow conditions. We considered storms as events with $> 20\,L\,s^{-1}$ increase in discharge and the following 24 h.

During baseflow periods, the acceptability limits were derived from the 90 % prediction interval of a linear regression model ($y = a \cdot x + b$) linking pairs of data points sampled on the same day (reference sample between 16:00 and 18:00,
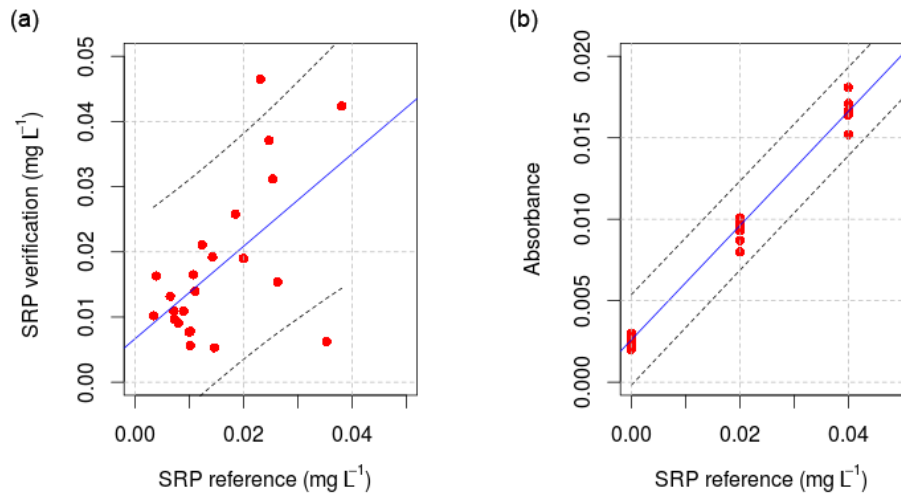
Figure 4. (a) Linear regression model linking the reference data and a verification data set; **(b)** measurement error as estimated from a repeatability test performed by the laboratory in charge of producing reference data (blue line: fitting regression; black dashes: 90 % prediction interval).

verification sample between 11:00 and 15:00) and analysed independently (within a fortnight for the reference sample and within 1–2 days for the verification sample). It was assumed that there was no systematic bias between the two data sets due to different sampling time. The reference SRP concentrations were on average 13 % lower than the verification value but this difference was not statistically significant (Mann–Whitney rank sum test, $p > 0.05$). This method encompasses all various sources of uncertainty, which results in prediction intervals much wider than what would result from a mere repeatability test; at the median concentration ($0.02 \, \text{mg L}^{-1}$), the estimated prediction interval was 166 % with this method vs. 57 % with a repeatability test (Fig. 4). As for discharge estimates, the high percentage represents a small absolute value ($0.03 \, \text{mg L}^{-1}$) during baseflow periods.

During storm events, acceptability limits were derived from the 90 % prediction interval of concentration discharge statistical models ($C = a \cdot Q^b$) using high-frequency autosampler data. Two reasons led us to use a statistical model (which also implies the assumption that errors are aleatory and temporally independent): (i) the measurement uncertainty as assessed by the laboratory repetition test was an underestimate of the real uncertainty of autosampler data, because it does not include other major sources of error such as delayed filtration and sample decay during storage; (ii) it was necessary to extrapolate the sub-daily observation to the daily resolution of the model. The limits of this choice will be discussed in Sect. 4.3. An empirical model was used to fit each storm event monitored separately and a delay term was introduced manually in the empirical model when a time lag existed between concentration and discharge peaks. The empirical models were then applied to extrapolate concentration estimation during 2 days at 10 min resolution, for each of the 14 storm events monitored. Finally the 2-day mean
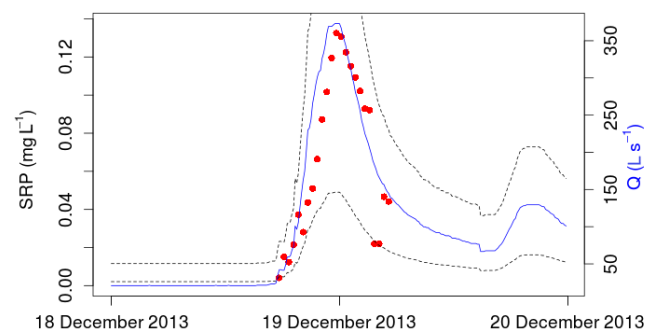

Figure 5. Example of an empirical concentration – discharge model; acceptability bounds derived from 90 % prediction interval. Red circles represent the SRP measurements.

"observed" load was estimated as the mean of 10 min loads and uncertainty limits were derived from the 90 % prediction interval. In model evaluation, the mean of simulated loads during 2 consecutive days was evaluated against the 2-day mean "observed" load for which prediction intervals have been calculated. A 2-day acceptability limit enables all the storm events to be covered (Fig. 5 and Supplement). A 2-day aggregation was necessary here because increased SRP load as a response to each storm event could occur either mainly during the day of the rainfall (if the rainfall occurred early in the morning) or mainly during the day following the rainfall (if the rainfall occurred late in the evening), and with the daily resolution of the input data and model simulation, the information about the timing of the rainfall event was not available to the model.

When comparing autosampler data with data from immediately filtered samples, the ratio obtained had the range 1–1.6 (mean = 1.3); hence, autosampler data were underesti-

Table 2. Starting and ending dates of periods studied.

Name	Starting date	Ending date
Autumn 2013	1 October 2013	31 December 2013
Winter 2014	1 January 2014	31 March 2014
Spring 2014	1 April 2014	31 July 2014
Autumn 2014	1 October 2014	31 December 2014
Winter 2015	1 January 2015	31 March 2015
Spring 2015	1 April 2015	31 July 2015

mates of the true concentration, arguably through adsorption or biological consumption. We used the mean ratio to correct all storm acceptability intervals by 30 % and the range values to extend the upper limit by 60 %. During days with a storm event not monitored at high frequency with an autosampler, we considered that the grab sample data did not contain enough information to derive an acceptability interval for daily SRP load; hence, simulated load was not evaluated for events not monitored at high frequency.

2.3.3 Model runs and selection of acceptable models

To explore the parameter space, 20 000 Monte Carlo realizations were performed to simulate daily discharge and SRP load during the water years 2013–2014 and 2014–2015. The number of Monte Carlo realizations was constrained by the computation time required to run a spatially explicit model in this catchment. A 7-month initialization period was run to reduce the impact of initial conditions on simulated results during the study period, from 1 October 2013 to 31 July 2015.

To be considered acceptable, model runs must fall within the acceptability limits defined in Sect. 2.3.1 and 2.3.2. More specifically, 100 % of simulated daily discharge, 100 % of simulated baseflow SRP load and 100 % of simulated storm SRP load have to fall within the acceptability limits. Thus, 572 acceptability tests were performed for discharge, 378 for baseflow SRP load and 14 for storm SRP loads, i.e. 964 evaluation criteria.

To evaluate the model performance in more detail, normalized scores were calculated during six periods (Table 2). To calculate the scores, a difference was calculated between each of the daily simulated discharge, baseflow SRP load and 2-day storm SRP loads and the corresponding observation. This difference was then normalized by the width of the acceptability limit defined for that day; therefore, the score has a value of 0 in the case of a perfect match with observation, -1 at the lower limit and $+1$ at the upper limit (Fig. 6a). Finally, the median of this ratio was calculated for each of the six periods to investigate whether the model tended to underestimate or overestimate discharge and loads at different moments of the year and between the two years.

Model runs were successively evaluated for discharge, baseflow SRP load and storm SRP load. To use the models

Table 3. Sensitivity analysis of the model to 18 model parameters (insignificant ".", important "*", critical "***"). Parameters significations are detailed in Table 1.

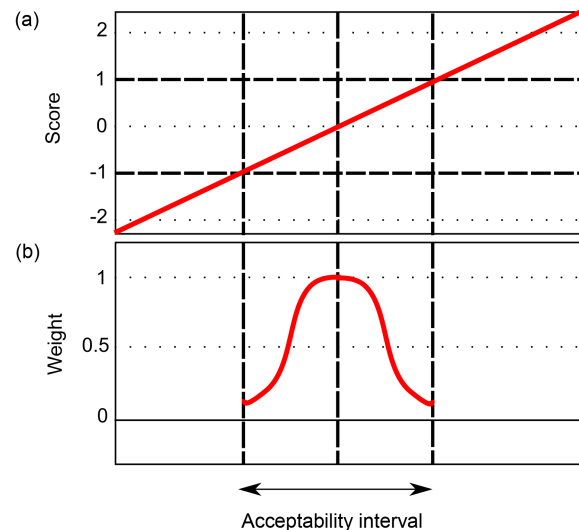|	Discharge	Baseflow SRP load	Storm SRP load
T (poorly drained soils)	.	***	***
m (poorly drained soils)	***	***	***
ho (poorly drained soils)	***	***	.
po (poorly drained soils)	***	***	***
h1 (poorly drained soils)	***	***	.
p1 (poorly drained soils)	***	***	***
T (well-drained soils)	.	***	***
m (well-drained soils)	***	***	***
ho (well-drained soils)	***	***	.
po (well-drained soils)	***	***	***
h1 (well-drained soils)	***	***	.
p1 (well-drained soils)	***	***	***
Coef_sub-surface	.	***	.
Coef_overland	.	***	***
SRP_deep	.	.	.
$S1$	.	***	***
$T1$	.	***	***
$T2$	.	***	***

Figure 6. Normalized scores **(a)** and weighting function **(b)**.

for prediction, each accepted model was given a likelihood weight according to how well it has performed for each of the 964 evaluation criteria. Here the statistical deviation weight was used (truncated to 90 % prediction interval) (Fig. 6b). To "combine" the weights derived from the rating curve and the SRP concentration statistical models, a kernel density estimate (with Gaussian smoothing kernel) was computed to fit 10 000 realizations of the multiplied error models. Calculated weights were then averaged for discharge, baseflow SRP load and storm SRP load respectively, and the final likelihood was calculated as the product of all three averages.

The model's sensitivity to each hydrological and soil parameter was performed with a Hornberger–Spear–Young Generalized Sensitivity Analysis (HSY GSA; Whitehead and Young, 1979; Hornberger and Spear, 1981). For each evaluation criteria (daily discharge, daily baseflow SRP load and 2-day storm SRP load), the model runs were split into acceptable and non-acceptable runs according to the above-mentioned acceptability limits. Then a Kolmogorov–Smirnov test was performed to assess whether the distribution of each of the three evaluation criteria differs between acceptable and non-acceptable models for each parameter. Because the Kolmogorov–Smirnov test might suggest that small differences in distribution are very significant when there are larger number of runs, this method is a qualitative guide to relative sensitivity. The p value of the Kolmogorov–Smirnov test is used to discriminate whether the model is critically sensitive ($p < 0.01$, "***"), importantly sensitive ($p < 0.1$, "*") or insignificantly sensitive ($p > 0.1$, ".") to each parameter and for each of the three evaluation criteria.

In addition to the acceptability limit approach, a NSE (Moriasi et al., 2007) was calculated for daily discharge and daily load and concentration to allow for a comparison with other modelling studies where it has been taken as an evaluation criterion.

3 Results

3.1 Presentation of observation data and calculation of acceptability limits

The two water years studied were highly contrasted in terms of hydrology and SRP loads. The water year 2013–2014 was the wettest in the last 10 years, with cumulative rainfall of 1289 mm and cumulative runoff of 716 mm. The water year 2014–2015 was an average year (fifth wettest in the last 10 years), with cumulative rainfall of 677 mm and cumulative runoff of 383 mm. Annual SRP load was 0.35 kg P ha^{-1} yr^{-1} in 2013–2014 and 0.17 kg P ha^{-1} yr^{-1} in 2014–2015, i.e. a difference 10 % higher than that of discharge. Observed mean SRP concentration during the study period was 0.024 mg L^{-1}.

Figure 7a and b show acceptability limits for daily discharge and daily SRP loads. Note that acceptability limits for discharge were calculated every day, while the acceptability limits for SRP load was calculated on a daily basis during baseflow periods and on a 2-day basis during storm events monitored at high frequency. No SRP load acceptability limit was calculated during storm events when no high-frequency autosampler data were available.

3.2 Model evaluation

First, model runs were evaluated against acceptability limits defined for discharge (Fig. 7c). A total of 5479 out of 20 000 models fulfilled the selection criterion for discharge; i.e. they

had 100 % of simulated daily discharge within the acceptability limits. The NSE estimated for these models ranged from 0.75 to 0.93. The normalized scores calculated seasonally (Fig. 8a) show that simulated discharge is often overestimated in autumn and spring, and underestimated in winter.

Then, model runs were evaluated against acceptability limits defined for SRP loads (Fig. 7d). During baseflow periods, 4964 out of 20 000 models fulfilled the selection criterion for SRP loads; i.e. they had 100 % of simulated daily SRP load within the acceptability limits. Among them, 1595 also fulfilled the previous selection criterion for discharge. Normalized scores for baseflow SRP load showed the same trend as for discharge (Fig. 8b), i.e. overestimation in autumn and spring, and underestimation in winter. During storm events, only seven models fulfilled the selection criterion for SRP loads; i.e. they had 14 out of 14 of simulated 2-day storm SRP loads within the acceptability limits, but none of them fulfilled the selection criteria for discharge and baseflow SRP loads. Two storm events were particularly difficult to simulate (number 2 and number 9, Fig. 8c), probably because their acceptability interval was very narrow as a result of only small changes in discharge and concentration. To obtain a reasonable number of acceptable models, we relaxed the selection criterion so that the acceptable models had to simulate 12 out of 14 of storm loads within the acceptability limits, in addition to the selection criteria defined for discharge and baseflow SRP load; 539 models were then accepted. Estimated NSE of these 539 models ranged from 0.09 to 0.81 for daily load and from negative values to 0.53 for daily concentrations (this includes all data from the regular sampling).

3.3 Sensitivity analysis and prediction results

According to the HSY generalized sensitivity analysis, simulated discharge was critically sensitive to 10 out of the 12 hydrological parameters varied (Table 3). Simulated SRP load was critically sensitive to the sub-surface and overland flow parameters during baseflow periods and to the overland flow parameter during storm events. During baseflow periods, SRP load was insignificantly sensitive to the parameter associated with deep flow load. Both baseflow and storm SRP loads were critically sensitive to the parameter related to soil moisture and soil temperature-dependent SRP solubilization ($S1$, $T1$ and $T2$), in addition to 12 and 8 hydrological parameters respectively. This identification of sensitive parameters can be used in future application of the TNT2-P model in the study catchment, as suggested by Whitehead and Hornberger (1984) and Wade et al. (2002b).

Figure 9 shows the daily discharge, SRP load and concentration as simulated by the acceptable models. Simulated SRP load during the water year 2013–2014 ranged from 0.81 to 3.25 kg P ha^{-1} yr^{-1} (median = 1.68 kg P ha^{-1} yr^{-1}); simulated SRP load during the water year 2014–2015 ranged from 0.14 to 0.73 kg P ha^{-1} yr^{-1} (me-
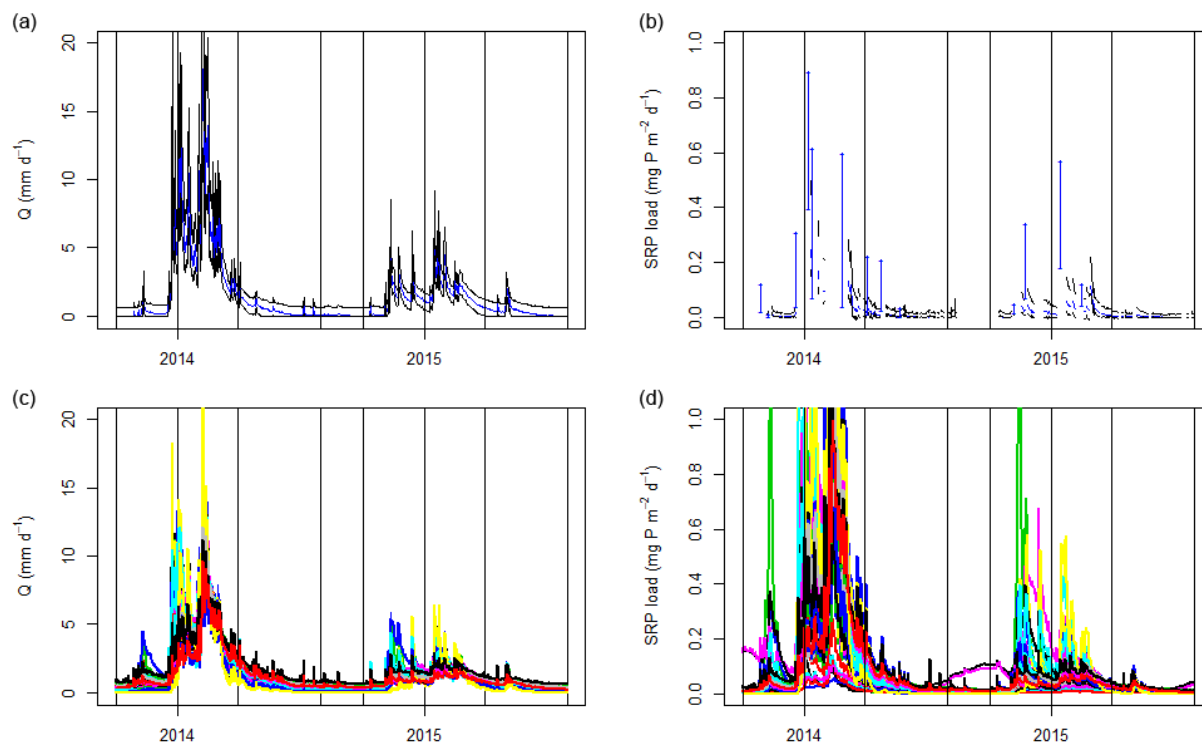
Figure 7. Acceptability limits for daily discharge **(a)** and SRP load **(b)**. Blue lines represent best estimates; black lines represent the acceptability limits. Storm loads acceptability limits are represented by vertical blue lines. An example of 50 model runs simulating discharge **(c)** and daily load **(d)**. Black vertical lines represent the starting and ending dates for each season (Table 2).

dian $= 0.34$ kg P ha^{-1} yr^{-1}). Best estimate of SRP load according to observation data was 0.35 kg P ha^{-1} yr^{-1} in 2013–2014 and 0.17 kg P ha^{-1} yr^{-1} in 2014–2015. According to the model, 49–55 % (median $= 52$ %) of water discharge and 66–70 % (median $= 67$ %) of SRP load occurred during storm events. Mean SRP concentrations during the two water years ranged from 0.014 to 0.044 mg L^{-1} (median $= 0.029$ mg L^{-1}), whereas mean observed SRP concentration was 0.024 mg L^{-1}.

4 Discussion

4.1 Role of hydrology and biogeochemistry in determining SRP transfer

The fairly good performance of TNT2-P at simulating SRP loads provides further support that the hydrological and biogeochemical processes included into the model are dominant controlling factors in the Kervidy–Naizin catchment (i.e. the modelling hypotheses could not be rejected based on these results, except for two storm events). The primary control of hydrology in controlling connectivity between soils and streams has been highlighted by many studies analysing water quality time series at the outlet of agricultural catchments (Haygarth et al., 2012; Jordan et al., 2012; Dupas et al., 2015c; Mellander et al., 2015). This modelling exercise also

provides further support that SRP solubility can be satisfactorily represented by the soil Olsen P content and could vary according to temperature and moisture conditions. The underlying processes have not been identified precisely in the Kervidy–Naizin catchment; independent laboratory experiments have shown that microbial cell lysis resulting from alternating dry and water-saturated periods in the soil could be the cause of increased SRP mobility (Turner and Haygarth, 2001; Blackwell et al., 2009). This could explain the moisture dependence of SRP solubility in the model. Furthermore, net mineralization of soil organic phosphorus could explain the temperature dependence of SRP solubility in the model. These two hypotheses may explain increased SRP solubility in soils in periods of dry and hot conditions and will be further explored by incubation experiment with soils from the Kervidy–Naizin catchments.

4.2 Potential improvements to the model structure according to modelling purpose

The TNT2-P model was designed to test hypotheses about dominant processes and for this purpose, a parsimonious model structure was chosen to include only the processes that were to be tested. This parsimonious model structure might contain some conceptual misrepresentations due to oversimplification, and it might not include all the processes nec-
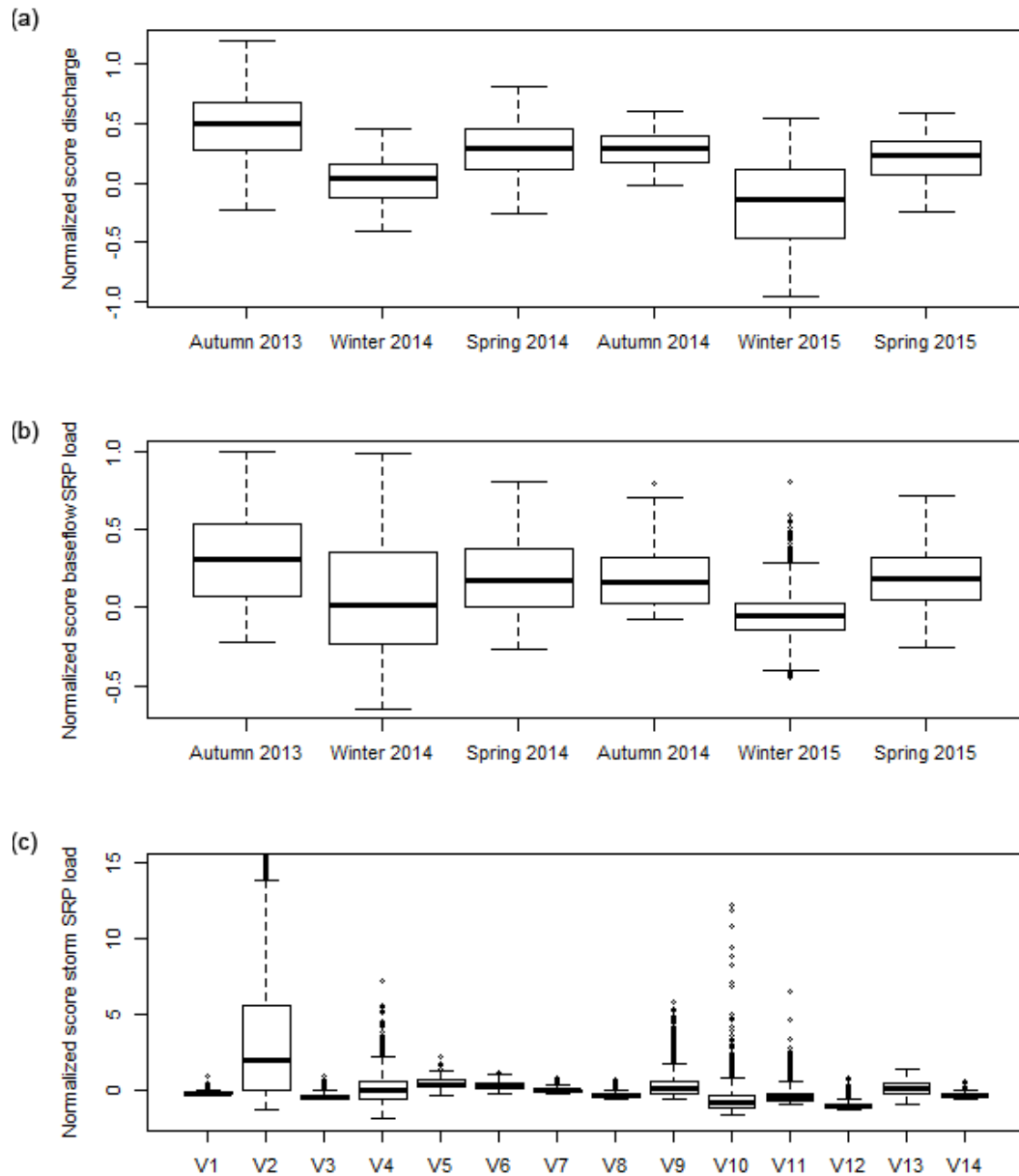
Figure 8. Normalized score for daily discharge **(a)**, baseflow SRP load **(b)** and storm SRP load **(c)**.

essary for the purpose of evaluating management scenarios. This section discusses whether the simplifications made are acceptable in the context of different catchment types, and to which conditions the model could be made more complex by including additional routines for the purpose of evaluating management scenarios.

From a conceptual point of view, the lack of cell-to-cell routing of SRP fluxes might result in erroneous results in some contexts. The fact that all the SRP emitted from each cell through overland flow and sub-surface flow reaches the stream on the same day is generally acceptable for the catchment studied, because groundwater interception of shallow soil layers occurs in the riparian zone only; hence, the signal of SRP mobilization in these soils is generally transmitted to the stream (Dupas et al., 2015c). This simplification, however, does not seem to be acceptable for all the storm events in the study catchment, as the SRP load evaluation criteria had to be relaxed to obtain acceptable model results. It would also not be acceptable in catchments where soil–groundwater interactions are taking place throughout the landscape, e.g. due to topographic depressions or poorly drained soils. In the latter type of catchment, transmission of the SRP mobilization signal to the stream is more complex (Haygarth et
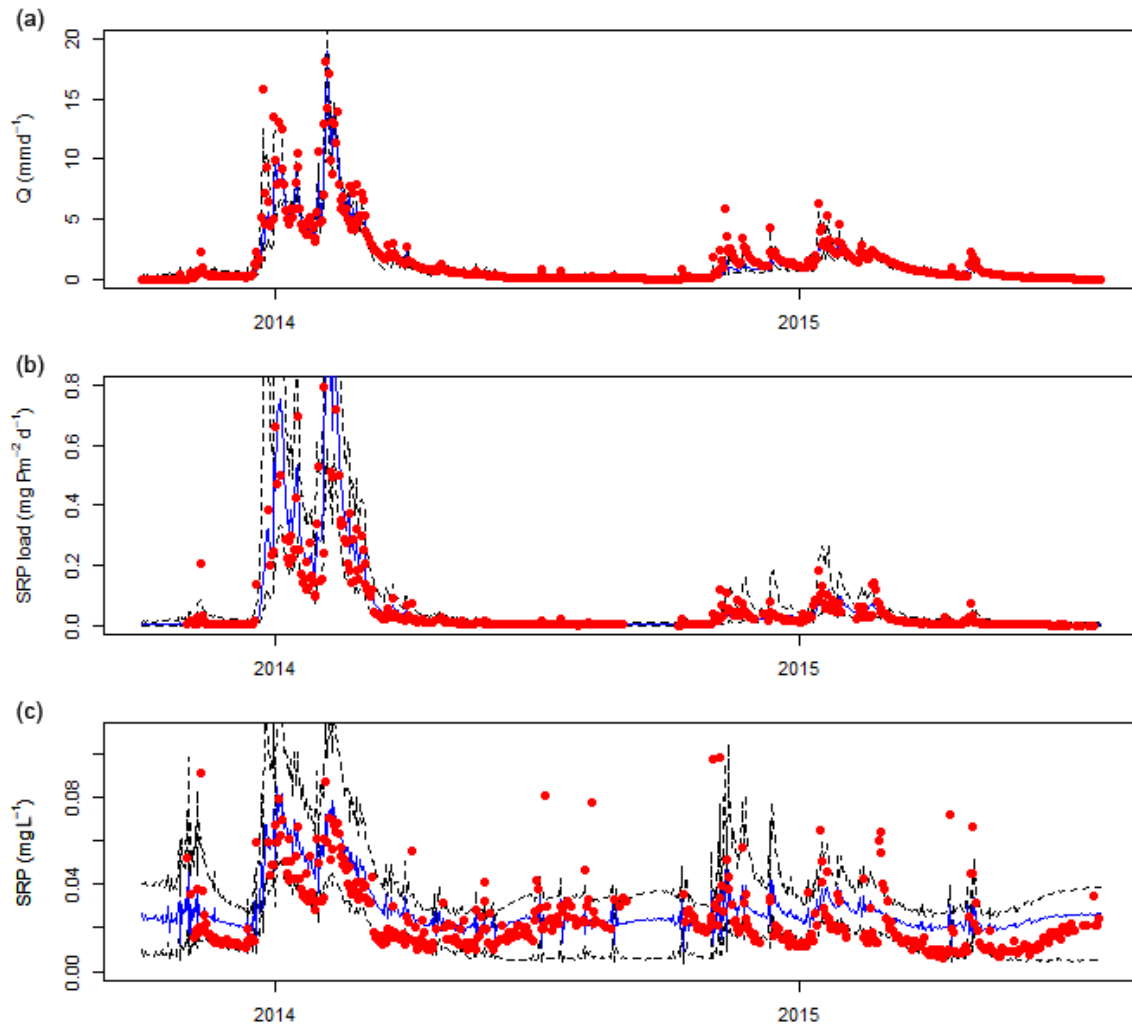
Figure 9. Median and 95 % credibility interval for daily discharge **(a)**, SRP load **(b)** and SRP concentration **(c)**. Red circles represent observational data.

al., 2012); hence, a more complex model structure would be required.

The reason for this simplification was that we lacked knowledge of SRP re-adsorption in downslope cells (or on suspended sediments in the stream network) and on the long-term fate of re-adsorbed SRP. For a more physically realistic representation of processes, it is likely that an explicit representation of flow velocities and pathways would be necessary, along with an explicit representation of several soil P pools. However, such an explicit representation of processes contradicts the idea of a parsimonious model, which was adopted here for the purpose of identifying dominant processes. In this respect, TNT2-P is an aggregative model rather than a fully distributed model although it is based on a fully distributed hydrological model (Beaujouan et al., 2002). The current spatial distribution allows for finer representation of soil–groundwater interactions (i.e. the time-varying extent of the riparian wetland area) than semi-distributed mod-

els such as SWAT (Arnold et al., 1998), INCA-P (Wade et al., 2002a) and HYPE (Lindstrom et al., 2010) but at higher computational cost. It would be interesting to test to what extent moving from an aggregative model with fully distributed information to a semi-distributed model would degrade the model performance while reducing computational cost. This could be achieved by grouping cells according to a hydrological similarity criterion like in the Dynamic TOPMODEL (Beven and Freer, 2001a; Metcalfe et al., 2015) and do the same for similarity in soil P content. Reducing computation time is critical in the context of a GLUE analysis because this method requires the parameter space to be sampled adequately to identify those models to be considered acceptable. This is debatable here because 12 parameters were varied and only 20 000 model runs were performed. It is therefore possible that some regions of the parameter space with acceptable models might not have been sampled.

If reducing the number of calculation units proved to reduce computational cost without degrading quality of prediction, it would be possible to include more parameters in the model, for example to simulate SRP re-absorption in downslope cells or include routines to simulate the evolution of soil P content under different management scenarios (Vadas et al., 2012), and still perform a Monte Carlo-based analysis of uncertainty. The question of coupling or not coupling such a soil P routine with the current TNT2-P model will depend on available data and on the length of available time series; studying the evolution of the soil P content requires at least a decade of soil observation data (Ringeval et al., 2014) and probably a longer period of stream data to account for the time delay for a perturbation in the catchment to become visible in the stream (Wall et al., 2013). Thus, the 2 years of daily stream SRP in the Kervidy–Naizin catchment are not enough to build a coupled soil–hydrology model with an elaborate soil P routine. Therefore, as things stand, it is more reasonable to generate new soil Olsen P maps with a separate model such as the APLE model (Vadas et al., 2012; Benskin et al., 2014) or the "soil P decline" model used by Wall et al. (2013), and use these maps as input to TNT2-P.

Because the current model can simulate response to rainfall, soil moisture and temperature, it could be used to test the effect of climate scenarios on SRP transfer. In western France, and more generally in western Europe, the climate for the next few decades is expected to consist of hotter, drier summers and warmer, wetter winters (Jacob et al., 2007; Macleod et al., 2012; Salmon-Monviola et al., 2013) with increased frequency of high-intensity rainfall events (Déqué, 2007). In these conditions, SRP concentrations and load will seemingly increase compared to today's climate as a result of both an increase in SRP solubility in soil due to higher temperatures and more severe drought, and an increase in transfer due to wetter winters and more frequent high-intensity rainfall events. TNT2-P could be used to confirm and quantify the expected increase in SRP transfer from diffuse sources in future climate scenarios, and to determine whether those predicted changes are significant relative to the uncertainty in predictions under current climate variability.

4.3 Improving information content in the data

Despite relatively large uncertainty in the data used in this study, it was possible to build a parsimonious catchment model of SRP transfer for the purpose of testing hypotheses about dominant processes, namely the role of hydrology in controlling connectivity between soils and streams and the role of temperature and moisture conditions in controlling soil SRP solubilization. However, the large uncertainties in the calibration data lead to large prediction uncertainty. For example, the SRP load estimated by the behavioural models from 2013 to 2015 ranged from 0.48 to 1.99 kg P ha^{-1} yr^{-1}; hence, the width of the credibility interval was 150 % of the median (1.0 kg P ha^{-1} yr^{-1}). Similarly, the mean SRP con-

centration estimated by the behavioural models from 2013 to 2015 ranged from 0.014 to 0.044 mg L^{-1}; hence, the width of the credibility interval was 102 % of the median (0.029 mg L^{-1}). The large uncertainty in the calibration data, along with a lack of long-term information, also prevents including more detailed processes in the soil routine.

To reduce uncertainty in prediction and to build more complex models, several options exist to improve information content in the data. As stated by Jackson-Blake and Starrfelt (2015), "the key to obtaining a realistic model simulation is ensuring that the natural variability in water chemistry is well represented by the monitoring data". The monitoring strategy adopted in the Kervidy–Naizin catchment should theoretically enable one to capture the natural variability in stream SRP concentration, because sampling took place during two contrasting water years, during different seasons and at a high frequency during 14 storm events. The analysis of uncertainty in the data shows that a large part of uncertainty in "observed" SRP concentration originates from sample storage, both unfiltered between the time of autosampling and manual filtration and between filtration and analysis. This is due to SRP being non-conservative. Thus, there is room for improvement in reducing storage time, without further increasing the monitoring frequency. In this respect, the primary interest of investing in high-frequency bankside analysers would lie in their ability to analyse water samples immediately in addition to providing near-continuous data. Because bankside analysers perform measurements in relatively homogeneous conditions, unlike the manual and autosampler data for which storage time of filtered and unfiltered samples vary, a finer quantification of uncertainty in the measurement data would be possible (e.g. Lloyd et al., 2016).

Finally, alternative methods to statistical models could be used to derive acceptability limits (in this study three statistical models are used: the rating curve, the SRP concentration uncertainty during baseflow periods and the storm event interpolation model) because statistical models have at least three shortcomings: (i) they lump the uncertainty linked to the timing of sampling, the immediate or delayed filtration of the samples, the storage time and the analytical error; (ii) the formula chosen adds error to the already existing measurement errors because empirical models are not perfect representations of the system dynamics; (iii) they assume a parametric distribution and temporally independent errors, which are not always verified in practice. As an alternative, nonparametric methods could be used, but these methods generally require a large number of data points and they are not suitable for extrapolation to extreme values.

5 Conclusion

The TNT2-P model was capable of capturing daily variation of SRP loads, thus confirming the dominant processes identi-

fied in previous analyses of observation data in the Kervidy–Naizin catchment. The role of hydrology in controlling connectivity between soils and streams, and the role of soil Olsen P, soil moisture and temperature in controlling SRP solubility have been confirmed. The lack of any representation of the short-term effect of management practices did not seem to reduce the model's performance. Their long-term effect on the soil Olsen P could be simulated with an independent model or through an additional sub-model if a longer period of data was available to calibrate it. The modelling approach presented in this paper included an assessment of the information content in the data, and propagation of uncertainty in the model's prediction. The information content of the data was sufficient to explore dominant processes, but the relatively large uncertainty in SRP concentrations would seemingly limit the possibility for including more detailed processes into the model. Data from the near-continuous bankside analyser will probably allow for calibrating more detailed models in the near future.

6 Data availability

Data of "ORE AgrHyS" can be downloaded from http://www6.inra.fr/ore_agrhys/Donnees (ORE AgrHyS, 2015).

The Supplement related to this article is available online at doi:10.5194/hess-20-4819-2016-supplement.

Edited by: J. Freer
Reviewed by: T. Krueger and P. G. Whitehead

References

Alexander, R. B., Smith, R. A., Schwarz, G. E., Boyer, E. W., Nolan, J. V., and Brakebill, J. W.: Differences in phosphorus and nitrogen delivery to the gulf of Mexico from the Mississippi river basin, Environ. Sci. Technol., 42, 822–830, 2008.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment - Part 1: Model development, J. Am. Water Resour. As., 34, 73–89, 1998.

Aubert, A. H., Gascuel-Odoux, C., Gruau, G., Akkal, N., Faucheux, M., Fauvel, Y., Grimaldi, C., Hamon, Y., Jaffrézic, A., Lecoz-Boutnik, M., Molénat, J., Petitjean, P., Ruiz, L., and Merot, P.: Solute transport dynamics in small, shallow groundwater-dominated agricultural catchments: insights from a high-frequency, multisolute 10 yr-long monitoring study, Hydrol. Earth Syst. Sci., 17, 1379–1391, doi:10.5194/hess-17-1379-2013, 2013.

Beauchemin, S. and Simard, R. R.: Soil phosphorus saturation degree: Review of some indices and their suitability for P management in Quebec, Canada, Can. J. Soil Sci., 79, 615–625, 1999.

Beaujouan, V., Durand, P., Ruiz, L., Aurousseau, P., and Cotteret, G.: A hydrological model dedicated to topography-based simulation of nitrogen transfer and transformation: rationale and application to the geomorphology-denitrification relationship, Hydrol. Process., 16, 493–507, 2002.

Benskin, C. M. H., Roberts, W. M., Wang, Y., and Haygharth, P. M.: Review of the Annual Phosphorus Loss Estimator tool – a new model for estimating phosphorus losses at the field scale, Soil Use and Manage., 30, 337–341, 2014.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, 2006.

Beven, K.: Environmental Modelling – An Uncertain Future?, Routledge, London, UK, 2009.

Beven, K. and Freer, J.: A dynamic TOPMODEL, Hydrol. Process., 15, 1993–2011, 2001a.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001b.

Beven, K. and Smith, P.: Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models, J. Hydrol. Eng., 20, A4014010, 2015.

Blackwell, M. S. A., Brookes, P. C., de la Fuente-Martinez, N., Murray, P. J., Snars, K. E., Williams, J. K., and Haygarth, P. M.: Effects of soil drying and rate of re-wetting on concentrations and forms of phosphorus in leachate, Biol. Fert. Soils, 45, 635–643, 2009.

Bruneau, P., Gascuelodoux, C., Robin, P., Merot, P., and Beven, K.: Sensitivity to space and time resolution of a hydrological model using digital elevation data, Hydrol. Process., 9, 69–81, 1995.

Carluer, N.: Vers une modélisation hydrologique adaptée à l'évaluation des pollutions diffuses: prise en compte du réseau anthropique. Application au bassin versant de Naizin (Morbihan), PhD thesis Université Pierre et Marie Curie, Paris, France, 1998.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., and Smith, V. H.: Nonpoint pollution of surface waters with phosphorus and nitrogen, Ecol. Appl., 8, 559–568, 1998.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., and Smith, P. J.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resour. Res., 51, 5531–5546, 2015.

Curmi, P., Durand, P., Gascuel-Odoux, C., Merot, P., Walter, C., and Taha, A.: Hydromorphic soils, hydrology and water quality: spatial distribution and functional modelling at different scales, Nutr. Cycl. Agroecosys., 50, 127–142, 1998.

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

Dequé, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global Planet. Change, 57, 16–26, 2007.

Dupas, R., Delmas, M., Dorioz, J. M., Garnier, J., Moatar, F., and Gascuel-Odoux, C.: Assessing the impact of agricultural pressures on N and P loads and eutrophication risk, Ecol. Indic., 48, 396–407, 2015a.

Dupas, R., Gascuel-Odoux, C., Gilliet, N., Grimaldi, C., and Gruau, G.: Distinct export dynamics for dissolved and particulate phosphorus reveal independent transport mechanisms in an arable headwater catchment, Hydrol. Process., 29, 3162–3178, 2015b.

Dupas, R., Gruau, G., Gu, S., Humbert, G., Jaffrezic, A., and Gascuel-Odoux, C.: Groundwater control of biogeochemical processes causing phosphorus release from riparian wetlands, Water Res., 84, 307—314, 2015c.

Dupas, R., Tavenard, R., Fovet, O., Gilliet, N., Grimaldi, C., and Gascuel-Odoux, C.: Identifying seasonal patterns of phosphorus storm dynamics with dynamic time warping, Water Resour. Res., 51, 8868–8882, 2015d.

Durand, P., Moreau, P., Salmon-Monviola, J., Ruiz, L., Vertes, F., and Gascuel-Odoux, C.: Modelling the interplay between nitrogen cycling processes and mitigation options in farming catchments, J. Agr. Sci., 153, 959–974, 2015.

Franks, S. W., Gineste, P., Beven, K. J., and Merot, P.: On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res., 34, 787–797, 1998.

Grizzetti, B., Bouraoui, F., and Aloe, A.: Changes of nitrogen and phosphorus loads to European seas, Glob. Change Biol., 18, 769–782, 2012.

Hahn, C., Prasuhn, V., Stamm, C., and Schulin, R.: Phosphorus losses in runoff from manured grassland of different soil P status at two rainfall intensities, Agr. Ecosyst. Environ., 153, 65–74, 2012.

Hahn, C., Prasuhn, V., Stamm, C., Lazzarotto, P., Evangelou, M. W. H., and Schulin, R.: Prediction of dissolved reactive phosphorus losses from small agricultural catchments: calibration and validation of a parsimonious model, Hydrol. Earth Syst. Sci., 17, 3679–3693, doi:10.5194/hess-17-3679-2013, 2013.

Haygarth, P. M., Ashby, C. D., and Jarvis, S. C.: Short-term changes in the molybdate reactive phosphorus of stored soil waters, J. Environ. Qual., 24, 1133–1140, 1995.

Haygarth, P. M., Hepworth, L., and Jarvis, S. C.: Forms of phosphorus transfer in hydrological pathways from soil under grazed grassland, Eur. J. Soil Sci., 49, 65–72, 1998.

Haygarth, P. M., Page, T. J. C., Beven, K. J., Freer, J., Joynes, A., Butler, P., Wood, G. A., and Owens, P. N.: Scaling up the phosphorus signal from soil hillslopes to headwater catchments, Freshwater Biol., 57, 7–25, 2012.

Heathwaite, A. L. and Dils, R. M.: Characterising phosphorus loss in surface and subsurface hydrological pathways, Sci. Total Environ., 251, 523–538, 2000.

Heckrath, G., Brookes, P. C., Poulton, P. R., and Goulding, K. W. T.: Phosphorus leaching from soils containing different phosphorus concentrations in the Broadbalk experiment, J. Environ. Qual., 24, 904–910, 1995.

Hornberger, G. M. and Spear, R. C.: An approach to the preliminary analysis of environmental systems, J. Environ. Manage., 12, 7–18, 1981.

Humbert, G., Jaffrezic, A., Fovet, O., Gruau, G., and Durand, P.: Dry-season length and runoff control annual variability in stream DOC dynamics in a small, shallow groundwater-dominated agricultural watershed, Water Resour. Res., 51, 7860–7877, 2015.

Jackson-Blake, L. A., Dunn, S. M., Helliwell, R. C., Skeffington, R. A., Stutter, M. I., and Wade, A. J.: How well can we model stream phosphorus concentrations in agricultural catchments?, Environ. Modell. Softw., 64, 31–46, 2015.

Jackson-Blake, L. A. and Starrfelt, J.: Do higher data frequency and Bayesian auto-calibration lead to better model calibration? Insights from an application of INCA-P, a process-based river phosphorus model, J. Hydrol., 527, 641–655, 2015.

Jacob, D., Barring, L., Christensen, O. B., Christensen, J. H., de Castro, M., Dequé, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellstrom, E., Lenderink, G., Rockel, B., Sanchez, E., Schar, C., Seneviratne, S. I., Somot, S., van Ulden, A., and van den Hurk, B.: An inter-comparison of regional climate models for Europe: model performance in present-day climate, Climatic Change, 81, 31–52, 2007.

Jarvie, H. P., Withers, J. A., and Neal, C.: Review of robust measurement of phosphorus in river water: sampling, storage, fractionation and sensitivity, Hydrol. Earth Syst. Sci., 6, 113–131, doi:10.5194/hess-6-113-2002, 2002.

Jordan, P., Melland, A. R., Mellander, P. E., Shortle, G., and Wall, D.: The seasonality of phosphorus transfers from land to water: Implications for trophic impacts and policy evaluation, Sci. Total Environ., 434, 101–109, 2012.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, W03S04, doi:10.1029/2005WR004362, 2006.

Krueger, T., Quinton, J. N., Freer, J., Macleod, C. J. A., Bilotta, G. S., Brazier, R. E., Hawkins, J. M. B., and Haygarth, P. M.: Comparing empirical models for sediment and phosphorus transfer from soils to water at field and catchment scale under data uncertainty, Eur. J. Soil Sci., 63, 211–223, 2012.

Lazzarotto, P., Stamm, C., Prasuhn, V., and Flühler, H.: A parsimonious soil-type based rainfall-runoff model simultaneously tested in four small agricultural catchments, J. Hydrol., 321, 21–38, 2006.

Lindstrom, G., Pers, C., Rosberg, J., Stromqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrol. Res., 41, 295–319, 2010.

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., Coxon, G., and Collins, A. L.: Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams, Hydrol. Process., 30, 165–152, 2016.

Macleod, C. J. A., Falloon, P. D., Evans, R., and Haygarth, P. M.: The effect of climate change on the mobilization of diffuse substances from agricultural systems, in: Advances in Agronomy, edited by: Sparks, D. L., Advances in Agronomy, 115, 41–77, 2012.

Maguire, R. O. and Sims, J. T.: Soil testing to predict phosphorus leaching, J. Environ. Qual., 31, 1601–1609, 2002.

Matos-Moreira, M., Lemercier, B., Michot, D., Dupas, R., and Gascuel-Odoux, C.: Using agricultural practices information for multiscale environmental assessment of phosphorus risk, Geophysical Research Abstracts, 17, 2015.

McDowell, R., Sharpley, A., and Withers, P.: Indicator to predict the movement of phosphorus from soil to subsurface flow, Environ. Sci. Technol., 36, 1505–1509, 2002.

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26, 4078–4111, 2012.

Mellander, P. E., Jordan, P., Shore, M., Melland, A. R., and Shortle, G.: Flow paths and phosphorus transfer pathways in two agricultural streams with contrasting flow controls, Hydrol. Process., 3504–3518, doi:10.1002/hyp.10415, 2015.

Metcalfe, P., Beven, K., and Freer, J.: Dynamic TOPMODEL: A new implementation in R and its sensitivity to time and space steps, Environ. Modell. Softw., 72, 155–172, 2015.

Molenat, J., Gascuel-Odoux, C., Ruiz, L., and Gruau, G.: Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (France), J. Hydrol., 348, 363–378, 2008.

Moore, M. T. and Locke, M. A.: Effect of Storage Method and Associated Holding Time on Nitrogen and Phosphorus Concentrations in Surface Water Samples, B. Environ. Contam. Tox., 91, 493–498, 2013.

Moreau, P., Ruiz, L., Mabon, F., Raimbault, T., Durand, P., Delaby, L., Devienne, S., and Vertes, F.: Reconciling technical, economic and environmental efficiency of farming systems in vulnerable areas, Agr. Ecosyst. Environ., 147, 89–99, 2012.

Moreau, P., Viaud, V., Parnaudeau, V., Salmon-Monviola, J., and Durand, P.: An approach for global sensitivity analysis of a complex environmental model to spatial inputs and parameters: A case study of an agro-hydrological model, Environ. Modell. Softw., 47, 74–87, 2013.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, T. Asabe, 50, 885–900, 2007.

Olsen, S. R., Cole, C. V., Watanbe, F. S., and Dean, L. A.: Estimation of available phosphorus in soils by extraction with sodium bicarbonate Circ. 939, USDA, Washington, D.C., USA, 1954.

ORE AgrHyS: AgrHyS database, available at: http://www6.inra.fr/ore_agrhys/Donnees, 2015.

Outram, F. N., Lloyd, C. E. M., Jonczyk, J., Benskin, C. McW. H., Grant, F., Perks, M. T., Deasy, C., Burke, S. P., Collins, A. L., Freer, J., Haygarth, P. M., Hiscock, K. M., Johnes, P. J., and Lovett, A. L.: 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, doi:10.5194/hess-18-3429-2014, 2014.

Page, T., Haygarth, P. M., Beven, K. J., Joynes, A., Butler, T., Keeler, C., Freer, J., Owens, P. N., and Wood, G. A.: Spatial variability of soil phosphorus in relation to the topographic index and critical source areas: Sampling for assessing risk to water quality, J. Environ. Qual., 34, 2263–2277, 2005.

Perks, M. T., Owen, G. J., Benskin, C. M. H., Jonczyk, J., Deasy, C., Burke, S., Reaney, S. M., and Haygarth, P. M.: Dominant mechanisms for the delivery of fine sediment and phosphorus to fluvial networks draining grassland dominated headwater catchments, Sci. Total Environ., 523, 178–190, 2015.

Quinlan, J. R.: Learning with continuous classes, Proceedings of the 5th Australian Joint Conference On Artificial Intelligence, 343–348, 1992.

Ringeval, B., Nowak, B., Nesme, T., Delmas, M., and Pellerin, S.: Contribution of anthropogenic phosphorus to agricultural soil fertility and food production, Global Biogeochem. Cy., 28, 743–756, 2014.

Rode, M. and Suhr, U.: Uncertainties in selected river water quality data, Hydrol. Earth Syst. Sci., 11, 863–874, doi:10.5194/hess-11-863-2007, 2007.

Salmon-Monviola, J., Moreau, P., Benhamou, C., Durand, P., Merot, P., Oehler, F., and Gascuel-Odoux, C.: Effect of climate change and increased atmospheric CO_2 on hydrological and nitrogen cycling in an intensive agricultural headwater catchment in western France, Climatic Change, 120, 433–447, 2013.

Schindler, D. W., Hecky, R. E., Findlay, D. L., Stainton, M. P., Parker, B. R., Paterson, M. J., Beaty, K. G., Lyng, M., and Kasian, S. E. M.: Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment, P. Natl. Acad. Sci. USA, 105, 11254–11258, 2008.

Schoumans, O. F. and Chardon, W. J.: Phosphate saturation degree and accumulation of phosphate in various soil types in The Netherlands, Geoderma, 237, 325–335, 2015.

Serrano, T., Dupas, R., Upegui, E., Buscail, C., Grimaldi, C., and Viel, J.-F.: Geographical modeling of exposure risk to cyanobacteria for epidemiological purposes, Environ. Int., 81, 18–25, 2015.

Sharpley, A. N., Kleinman, P. J., 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. Environ. Qual., 37, 362–368, 2008.

Siwek, J., Siwek, J. P., and Zelazny, M.: Environmental and land use factors affecting phosphate hysteresis patterns of stream water during flood events (Carpathian Foothills, Poland), Hydrol. Process., 27, 3674–3684, 2013.

Turner, B. L. and Haygarth, P. M.: Biogeochemistry – Phosphorus solubilization in rewetted soils, Nature, 411, 258–258, 2001.

Vadas, P. A., Joern, B. C., and Moore, P. A.: Simulating soil phosphorus dynamics for a phosphorus loss quantification tool, J. Environ. Qual., 41, 1750–1757, 2012.

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

Wade, A. J., Whitehead, P. G., Hornberger, G. E., and Snook, D.: On Modelling the flow controls on macrophytes and epiphyte dynamics in a lowland permeable catchment: the River Kennet, southern England, Sci. Total Environ., 282–283, 395–417, 2002b.

Wall, D. P., Jordan, P., Melland, A. R., Mellander, P. E., Mechan, S., and Shortle, G.: Forecasting the decline of excess soil phosphorus in agricultural catchments, Soil Use Manage., 29, 147–154, 2013.

Whitehead, P. and Young, P.: Water-quality in river systems – Monte-Carlo analysis, Water Resour. Res., 15, 451–459, 1979.

Whitehead, P. G. and Hornberger, G. M.: Modeling algal behavior in the river thames, Water Res., 18, 945–953, 1984.