



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

3 **R. Dupas¹, J. Salmon-Monviola¹, K. Beven², P. Durand¹, P.M. Haygarth², M.J.**
4 **Hollaway², C. Gascuel-Oudou¹**

5

6 [1] INRA, Agrocampus Ouest, UMR1069 SAS, F-35000 Rennes, France

7 [2] Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom, LA1
8 4YQ

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

10

11 **Abstract**

12 We developed a parsimonious topography-based hydrologic model coupled with a soil
13 biogeochemistry sub-model in order to improve understanding and prediction of Soluble
14 Reactive Phosphorus (SRP) transfer in agricultural headwater catchments. The model
15 structure aims to capture the dominant hydrological and biogeochemical processes identified
16 from multiscale observations in a research catchment (Kervidy-Naizin, 5 km²). Groundwater
17 fluctuations, responsible for the connection of soil SRP production zones to the stream, were
18 simulated with a fully-distributed hydrologic model at 20 m resolution. The spatial variability
19 of the soil phosphorus status and the temporal variability of soil moisture and temperature,
20 which had previously been identified as key controlling factor of SRP solubilisation in soils,
21 were included as part of an empirical soil biogeochemistry sub-model. The modelling
22 approach included an analysis of the information contained in the calibration data and
23 propagation of uncertainty in model predictions using a GLUE “limits of acceptability”
24 framework. Overall, the model appeared to perform well given the uncertainty in the
25 observational data, with a Nash-Sutcliffe efficiency on daily SRP loads between 0.1 and 0.8
26 for acceptable models. The role of hydrological connectivity via groundwater fluctuation, and
27 the role of increased SRP solubilisation following dry/hot periods were captured well. We
28 conclude that in the absence of near continuous monitoring, the amount of information
29 contained in the data is limited hence parsimonious models are more relevant than highly



1 parameterised models. An analysis of uncertainty in the data is recommended for model
2 calibration in order to provide reliable predictions.

3 **1 Introduction**

4 Excessive phosphorus (P) concentrations in freshwater bodies result in increased
5 eutrophication risk worldwide (Carpenter et al., 1998; Schindler et al., 2008). Eutrophication
6 restricts economic use of water and poses a serious health hazard to humans, due to the
7 potential development of harmful cyanobacteria (Bradley et al., 2013; Serrano et al., 2015). In
8 western countries, reduction of point source P emissions in the last two decades has resulted
9 in a proportionally increasing contribution of diffuse sources, mainly from agricultural origin
10 (Alexander et al., 2008; Grizzetti et al., 2012; Dupas et al., 2015a). Of particular concern are
11 dissolved P forms, often measured as Soluble Reactive Phosphorus (SRP), because they are
12 highly bioavailable and therefore a likely contributor to eutrophication.

13 To reduce SRP transfer from agricultural soils it is important to identify the spatial origin of P
14 sources in agricultural landscapes, the biogeochemical mechanisms causing SRP
15 solubilisation in soils and the dominant transfer pathways. Research catchments provide
16 useful data to investigate SRP transport mechanisms: typically, the temporal variations in
17 water quality parameters at the outlet, together with hydroclimatic variables, are investigated
18 to infer spatial origin and dominant transfer pathways of SRP (Haygarth et al., 2012; Outram
19 et al., 2014; Dupas et al., 2015b; Mellander et al., 2015; Perks et al., 2015). Hypotheses
20 drawn from analysis of water quality time series can be further investigated through hillslope
21 monitoring and/or laboratory experiments (Heathwaite and Dils, 2000; Siwek et al., 2013;
22 Dupas et al., 2015c). When dominant processes are considered reasonably known, it is
23 possible to develop computer models, for two main purposes: first, to validate scientific
24 conceptual models, by testing whether model predictions can produce reasonable simulations
25 compared to observations. Of particular interest is the possibility to test the capability of a
26 computer model to upscale P processes observed at fine spatial resolution (soil column,
27 hillslope) to a whole catchment. Second, if the models survive such validation tests, then they
28 can be useful tools to simulate the response of a catchment system to a future perturbation
29 such as changes in agricultural management and climate changes.

30 However, process-based P models generally perform poorly compared to, for example,
31 nitrogen models (Wade et al., 2002; Dean et al., 2009; Jackson-Blake et al., 2015a). This is of
32 major concern because poor model performance suggests poor knowledge of dominant



1 processes at the catchment scale, and poor reliability of the modelling tools used to support
2 management. The origin of poor model performance might be conceptual misrepresentations,
3 structural imperfection, calibration problems, irrelevant model evaluation criteria and
4 difficulties in properly assessing the information content of the available data when it is
5 subject to epistemic error. All five causes of poor model performance are intertwined, e.g.
6 model calibration strategy depends on model performance evaluation criteria, which depend
7 on the way the information contained in the observation data is assessed (Beven and Smith,
8 2015).

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

27 A second key issue, linked to the question of model complexity, concerns model calibration
28 and evaluation. Both calibration and evaluation require assessing the fit of model outputs with
29 observation data. However, observation data are generally not directly comparable with model
30 outputs, because of incommensurability issues and/or because they contain errors (Beven,
31 2006; 2009). Typically, predicted daily concentrations and/or loads are evaluated against data
32 from grab samples collected on a daily or weekly basis. The information content of these data



1 must be carefully evaluated to propagate uncertainty in the data into model predictions.
2 Uncertainty in grab sample data might stem from i) sampling frequency problems and ii)
3 measurement problems (Lloyd et al., 2015). Grab sample data represent a snapshot of the
4 concentration at a given time of the day, which can differ from the flow weighted daily
5 concentration. This difference between observation data and simulation output can be large
6 during storm events in small agricultural catchments, as P concentrations can vary by several
7 orders of magnitudes during the same day (Heathwaite and Dils, 2000; Sharpley et al., 2008).
8 Model evaluation can be severely penalised by this difference, because many popular
9 evaluation criteria such as the Nash-Sutcliffe efficiency (NSE) are sensitive to extreme values
10 and errors in timing (Moriassi et al., 2007). During baseflow periods, it is more likely that grab
11 sample data are comparable to flow-weighted mean daily concentrations, as concentrations
12 vary little during the day and they are usually low in the absence of point sources. However,
13 measurement errors are expected to occur at low concentrations, either due to too long storage
14 times or laboratory imprecision when concentrations come close to detection/quantification
15 limits (Jarvie et al., 2002; Moore and Locke, 2013). Uncertainty in the data can also relate to
16 discharge measurement and input data (e.g. maps of soil P content and rainfall data). In this
17 paper we strive to identify and quantify the different sources of uncertainty in the data when
18 the required quality check tests have been performed. A Generalised Likelihood Uncertainty
19 Estimation (GLUE) “limits of acceptability” approach (Beven, 2006; Beven and Smith, 2015)
20 is used to calibrate/evaluate the model.

21 This paper presents a dominant-process model that couples a topography-based hydrologic
22 model with a soil biogeochemistry sub-model able to simulate daily discharge and SRP loads.
23 The dominant processes included in the hydrologic and soil biogeochemistry sub-models have
24 been identified in previous analyses of multiscale observational data, which have
25 demonstrated on the one hand the control of groundwater fluctuation on connecting soil SRP
26 production zones to the stream (Haygarth et al., 2012; Jordan et al., 2012; Dupas et al., 2015b;
27 2015d; Mellander et al., 2015), and on the other hand the role of antecedent soil moisture and
28 temperature conditions on SRP solubilisation in soils (Turner and Haygarth, 2001; Blackwell
29 et al., 2009; Dupas et al., 2015c). Model development and application was performed in the
30 Kervidy-Naizin catchment in western France with the objectives of: i) testing if the model
31 was capable of capturing daily variation of SRP load, thus confirming hypotheses on
32 dominant processes; ii) develop a methodology to analyse and propagate uncertainty in the



1 data into model prediction using a “limits of acceptability” approach. Model development and
2 analysis of uncertainty in the data are interlinked in this approach.

3 **2 Material and methods**

4 **2.1 Study catchment**

5 **2.1.1 Site description**

6 Kervidy–Naizin is a small (4.94 km²) agricultural catchment located in central Brittany,
7 Western France (48°N, 3°W). It belongs to the AgrHyS environmental research observatory
8 (http://www6.inra.fr/ore_agrhys_eng), which studies the impact of agricultural activities and
9 climate change on water quality (Molenat et al., 2008; Aubert et al., 2013; Salmon-Monviola
10 et al., 2013; Humbert et al., 2014). The catchment (Fig. 1) is drained by a stream of second
11 Strahler order, which generally dries up in August and September. The climate is temperate
12 oceanic, with annual cumulative precipitation and specific discharge averaging 854 ± 179 mm
13 and 290 ± 106 mm, respectively, from 2000 to 2014. Mean annual temperature is $11.2 \pm$
14 0.6°C . Elevation ranges from 93 to 135 m above sea level. Topography is gentle, with
15 maximum slopes not exceeding 5%. The bedrock consists of impervious, locally fractured
16 Brioverian schists and is capped by several metres of unconsolidated weathered material and
17 silty, loamy soils. The hydrological behaviour is dominated by the development of a water
18 table that varies seasonally along the hillslope. In the upland domain, consisting of well
19 drained soils, the water table remains below the soil surface throughout the year, varying in
20 depth from 1 to > 8 m. In the wetland domain, developed near the stream and consisting of
21 hydromorphic soils, the water table is shallower, remaining near the soil surface generally
22 from October to April each year. The land use is mostly agriculture, specifically arable crops
23 and confined animal production (dairy cows and pigs). A farm survey conducted in 2013 led
24 to the following land use subdivisions: 35% cereal crops, 36% maize, 16% grassland and 13%
25 other crops (rape seed, vegetables). Animal density was estimated as high as 13 livestock
26 units ha⁻¹ in 2010. Estimated soil P surplus is $13.1 \text{ kg P ha}^{-1} \text{ yr}^{-1}$ (Dupas et al., 2015b) and soil
27 extractable P in 2013 (Olsen et al., 1954) is $59 \pm 31 \text{ mg P kg}^{-1}$ ($n = 89$ samples). A survey
28 targeting riparian areas highlighted the legacy of high soil P content in these currently
29 unfertilized areas (Dupas et al., 2015c). No point source emissions are recorded but scattered
30 dwellings with septic tanks are present in the catchment.



1 **2.1.2 Hydroclimatic and chemical monitoring**

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

7 To record both seasonal and within storm dynamics in P concentration, two monitoring
8 strategies complemented each other from October 2013 to August 2015: a daily manual grab
9 sampling at approximately the same time (between 16:00 – 18:00 local time) and automatic
10 high frequency sampling during 14 storm events (autosampler ISCO 6712 Full-Size Portable
11 Sampler, 24 one litre bottles filled every 30 min). The water samples were filtered on-site,
12 immediately after grab sampling and after 1-2 days in the case of autosampling. They were
13 analysed for SRP (ISO 15681) within a fortnight. To assess uncertainty in daily SRP
14 concentration related to sampling time, storage and measurement errors, a second grab sample
15 was taken at a different time of the day (between 11:00 – 15:00 local time) in 36 instances
16 during the study period. The second sample was analysed within 24h with the same method;
17 this second dataset is referred to as verification dataset, as opposed to the reference dataset.
18 Among the 36 pairs of comparable daily samples, 12 were taken during storm events and 24
19 during baseflow periods. To assess uncertainty in high frequency SRP concentration during
20 storm events due to delayed filtration of autosampler bottles, 5 grab samples were taken
21 during the course of 4 distinct storms and were filtered immediately. The same lab procedure
22 was used to analyse SRP.

23 **2.1.3 Identification of dominant processes from multiscale observations**

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



1 In-situ monitoring of soil pore water at 4 sites (15 cm and 50 cm depths) in the Kervidy-
2 Naizin catchment has shown that mean SRP concentration in soils was a linear function of
3 Olsen P (Olsen et al., 1954). This reflects current knowledge that a soil P test, or alternatively
4 estimation of a degree of P saturation, can be used to assess solubilisation in soils
5 (Beauchemin and Simard, 1999; McDowell et al., 2002; Schoumans et al., 2015). This linear
6 relationship derived from the data contrasts however with other studies, where threshold
7 values above which SRP solubilisation increases greatly have been identified (Heckrath et al.,
8 1995; Maguire et al., 2002).

9 Soluble Reactive Phosphorus solubilisation in soil varies seasonally according to antecedent
10 conditions of temperature and soil moisture. Dry and/or hot conditions are favourable to
11 accumulation of mobile P forms in soils, while water saturated conditions lead to their
12 flushing (Turner et al., 2001; Blackwell et al., 2009; Dupas et al., 2015c).

13 **2.2 Description of the Topography-based Nutrient Transfer and** 14 **Transformation – Phosphorus (TNT2-P)**

15 TNT2 was originally developed as a process-based and spatially explicit model simulating
16 water and nitrogen fluxes at a daily time step (Beaujouan et al., 2002) in meso-scale
17 catchments (< 50 km²). TNT2-N has been widely used for operational objectives, to test the
18 effect of mitigation options proposed by local stakeholders or public policy-makers (Moreau
19 et al., 2012; Durand et al., 2015), on nitrate fluxes and concentrations in rivers.

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

22 **2.2.1 Hydrological sub-model**

23 The assumptions in the hydrological sub-model are derived from TOPMODEL which has
24 previously been applied to the Naizin catchment (Bruneau et al., 1995; Franks et al., 1998): 1)
25 the effective hydraulic gradient of the saturated zone is approximated by the local topographic
26 surface gradient ($\tan \beta$). It is calculated in each cell of a Digital Elevation Model (DEM) at the
27 beginning of the simulation; 2) the effective downslope transmissivity (parameter T) of the
28 soil profile in each cell of the DEM is a function of the soil moisture deficit (Sd). Hydraulic
29 conductivity decreases exponentially with depth (parameter m, Fig. 2). Hence water fluxes (q)
30 are computed as:



$$1 \quad q = T * \tan\beta * \exp\left(-\frac{sd}{m}\right) \quad (1)$$

2 Based on these assumptions, TNT2 computes an explicit cell-to-cell routing of fluxes, using a
3 D8 algorithm. This explicit cell-to-cell routing of fluxes increases computation times
4 compared to TOPMODEL, for which calculations are grouped according to a distribution of
5 hydrologically similar points, but it allows taking account of spatial interactions between soil
6 and groundwater, which has been shown to improve representation of nutrients fluxes and
7 transformations (Beaujouan et al., 2002).

8 To simulate SRP fluxes, the only modification to the hydrological sub-model aimed to
9 compute water fluxes from each soil layer by integrating [1] between the maximum depth of
10 the soil layer considered and:

11 - estimated groundwater level, if the groundwater table is within the soil layer
12 considered

13 or

14 - the minimum depth of the soil layer considered, if the groundwater table above the
15 soil layer considered

16 In this application of the TNT2-P model, 5 soil layers with a thickness of 10 cm are
17 considered. Hence, 7 flow components are computed in the model:

- 18 - overland flow on saturated surface
- 19 - 5 sub-surface flow components, for each soil layer
- 20 - deep flow, i.e. flow below the 5 soil layers

21 **2.2.2 Soil-P sub-model**

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

- 24 - background SRP concentration in the soil pore water of a given layer is proportional to
25 soil Olsen P;
- 26 - seasonal increases in P availability compared to background conditions are determined
27 by biogeochemical processes, controlled by antecedent temperature and soil moisture.
28 Data show that SRP availability in the soil pore water increases following periods of
29 dry and hot conditions (Dupas et al., 2015c).



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

$$4 \quad F_{SRP \text{ overland}} = Coef_{SRP \text{ overland}} * P_{Olsen} * q_{overland} \quad (2)$$

$$5 \quad F_{SRP \text{ sub-surface}} = Coef_{SRP \text{ sub-surface}} * P_{Olsen} * q_{sub-surface} \quad (3)$$

6 Where $F_{SRP \text{ overland}}$ and $F_{SRP \text{ sub-surface}}$ are SRP transfer via overland flow and sub-surface
7 flow for a given soil layer respectively, $q_{overland}$ and $q_{sub-surface}$ are water flows from the
8 same pathways. $Coef_{SRP \text{ overland}}$ and $Coef_{SRP \text{ sub-surface}}$ are coefficients which vary
9 according to antecedent temperature and soil moisture conditions, such as:

$$10 \quad Coef_{SRP} = Coef_{background} * (1 + F_T * F_S) \quad (4)$$

11 Where $Coef_{SRP}$ is either $Coef_{SRP \text{ overland}}$ or $Coef_{SRP \text{ sub-surface}}$, and F_T and F_S are
12 temperature and soil moisture factors, respectively. F_T and F_S are expressed as:

$$13 \quad F_T = \exp\left(\frac{\text{mean}(\text{temperature}, i \text{ days}) - T1}{T2}\right) \quad (5)$$

$$14 \quad F_S = 1 - \left(\frac{\text{mean}(\text{water content}, i \text{ days})}{\text{maximum water content}}\right)^{S1} \quad (6)$$

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

25 **2.2.3 Input data and parameters**

26 Spatial input data include:

- 27 - A DEM in raster format. Here, a 20 m resolution DEM was used, hence model
28 calculations were made in 12348 grid cells covering a 4.94 km² catchment.



- 1 - A map of soils with homogeneous hydrological parameter value, in raster format.
2 Here, two soil classes were considered by differentiating well-drained (86%) and
3 poorly drained soils (14%) according to Curmi et al. (1998) (Fig. 1).
- 4 - A map of surface Olsen P in raster format and description of decrease in P Olsen with
5 depth for five soil layers between 0-50 cm. Here, the map of Olsen P in the 0-15 cm
6 soil layer was obtained from statistical modelling with the rule-based regression
7 algorithm CUBIST (Quinlan, 1992) using data from 198 soil samples (2013) in an
8 area of 12 km² encompassing the 4.94 km² catchment (Matos-Moreira et al., 2015).
9 To describe how P Olsen decreases with depth, land use information was used. In
10 tilled fields, i.e. all crop rotations including arable crops, Olsen P was assumed to be
11 constant between 0-30 cm and to decrease linearly with depth between 30-50 cm. In
12 no-till fields, i.e. permanent pasture and woodland, Olsen P was assumed to decrease
13 linearly with depth between 0-50 cm. An exponential decrease with depth is more
14 commonly adopted in untilled land (e.g. Haygarth et al., 1998; Page et al., 2005), but a
15 specific sampling in currently untilled areas in the Kervidy-Naizin catchment (Dupas
16 et al., 2015c) has shown that a linear function is more appropriate, probably because
17 of these areas having been ploughed in the past.
- 18 Climate input data include minimum and maximum air temperature, precipitation, potential
19 evapotranspiration, global radiation on a daily basis. The TNT2 model allows for several
20 climate zones to be considered, in which case a raster map of climate zone must be provided
21 to the model. Here, only one climate zone is considered.
- 22 In total, the TNT2-P model includes 15 parameters for each soil type, i.e. 30 parameters in
23 total if two soil drainage classes are considered. To reduce the number of model runs
24 necessary to explore the parameter space using Monte Carlo simulations, several parameters
25 were given fixed values, or a constant ratio between the two soil types was set (Table 1). In
26 the hydrological sub-model, the parameters to vary were identified in a previous sensitivity
27 analysis (Moreau et al., 2013). In the soil sub-model, all the parameters were varied.
- 28 Finally, only 12 parameters were varied independently. Initial parameter ranges for the
29 hydrological sub-model were based on literature-derived values (Moreau et al., 2013) and
30 those for the soil sub-model were based on a preliminary manual trial and error procedure.
31 The SRP concentration for deep flow water was based on actual measurement of SRP in the



1 weathered schist (Dupas et al., 2015c). A constant flux value for domestic sources was set at
2 the 1% percentile of the daily flux between 2007 and 2013 (Dupas et al., 2015b).

3 **2.3 Deriving limits of acceptability from data uncertainty assessment**

4 The Monte Carlo based Generalized Likelihood Uncertainty Estimation (GLUE)
5 methodology has been widely used in hydrology and is described elsewhere (Beven and
6 Freer, 2001; Beven, 2006, 2009). Briefly, the rationale of GLUE is that many model
7 structures and parameter sets can give “acceptable” results, according to one or several
8 performance measures, due to equifinality. Hence, GLUE considers that all models that give
9 acceptable results should be used for prediction. A key issue in GLUE is to decide on a
10 performance threshold to define acceptable models; typically, modellers set a threshold value
11 of a measure such as the Nash-Sutcliffe Efficiency based on their subjective appreciation of
12 data uncertainty or on previously used values. To allow for a more explicit justification of the
13 performance threshold values used, the limits of acceptability approach outlined by Beven
14 (2006) relies on an assessment of uncertainty in the calibration/evaluation data. According to
15 this approach, all model realisations that fall within the limits of acceptability are used for
16 prediction, weighted by a score calculated based on overall performance.

17 Details on how the limits of acceptability for daily discharge and daily SRP load were derived
18 from uncertainty assessment of the observational data are presented below. Input data, such as
19 weather and soil data, also contained uncertainty which were not accounted for in the limits of
20 acceptability due to a lack of data to quantifying them.

21 **2.3.1 Discharge**

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

$$25 \quad Q = a * (h - h_0)^b \quad (7)$$

26 Where Q is discharge, h is stage reading, h_0 is stage reading at zero discharge, a and b are
27 calibrated coefficients. Limits of acceptability were defined as the 90% prediction interval of
28 log-log linear regression (Fig. 3). Estimated acceptability range was $\pm 39\%$ on average. For
29 daily discharge values below 2 mm d^{-1} , fixed acceptability limits were set at the 90%
30 prediction interval for a stage measurement corresponding to 2 mm d^{-1} .



1 **2.3.2 SRP load**

2 Uncertainty in “observed” daily load includes uncertainty in discharge (see 2.3.1.) and
3 uncertainty in SRP concentration. Uncertainty in daily load was estimated summing up
4 relative uncertainty assessed for discharge and SRP concentration. Uncertainty in SRP
5 concentration stems from sampling frequency problems as one grab sample collected on a
6 specific day is incommensurable with the mean daily concentration or load simulated by the
7 model. Further, measurement errors exist that include the effect of storage time (Haygarth et
8 al., 1995). During baseflow periods, measurement error was expected to be the main source of
9 uncertainty because relative measurement error is large for low concentrations, especially
10 when sample storage time exceeds 48h (Jarvie et al., 2002), while concentrations vary little.
11 During storm events, sampling frequency was expected to be the main source of uncertainty
12 because SRP concentration can vary by one order of magnitude within a few hours.
13 Therefore, different acceptability limits were set for both flow conditions. We considered
14 storms as events with $> 20 \text{ l s}^{-1}$ increase in discharge and the following 24h.

15 During baseflow periods, the acceptability limits were derived from the 90% prediction
16 interval of a linear regression model linking pairs of data points sampled on the same day
17 (reference sample between 16:00-18:00, verification sample between 11:00-15:00) and
18 analysed independently (within a fortnight for the reference sample and within 1-2 days for
19 the verification sample). It was assumed that there was no systematic bias between the two
20 datasets due to different sampling time. The reference SRP concentrations were on average
21 13% lower than the verification value but this difference was not statistically significant
22 (Mann-Whitney Rank Sum Test, $p > 0.05$). Hence, the expected underestimation of SRP
23 concentration due to long sample storage appears to be overshadowed by other sources of
24 uncertainty such as variability in SRP concentration during the day of sampling or analytical
25 imprecision at low concentrations. This method encompasses all various sources of
26 uncertainty, which results in prediction intervals much wider than what would result from a
27 mere repeatability test: at the median concentration (0.02 mg l^{-1}), estimated prediction interval
28 was 166% with this method versus 57% with a repeatability test (Fig. 4).

29 During storm events, acceptability limits were derived from the 90% prediction interval of
30 concentration discharge empirical models $C = a \cdot Q^b$ using high frequency autosampler data.
31 A distinct empirical model was used to fit to each storm event monitored and a delay term
32 was introduced manually in the empirical model when a time lag existed between



1 concentration and discharge peaks. The empirical models were then applied to extrapolate
2 concentration estimation during two days at 10 min resolution, for each of the 14 storm events
3 monitored. Finally the 2-day mean “observed” load was estimated as the mean of 10 min
4 loads and uncertainty limits were derived from the 90% prediction interval. In model
5 evaluation, the mean of simulated loads during 2 consecutive days was evaluated against the
6 2-day mean “observed” load for which prediction intervals have been calculated. A 2-day
7 acceptability limit enables to cover the whole of storm events (Fig. 5 and Supplement).

8 When comparing autosampler data with data from immediately filtered samples, the ratio
9 obtained ranged 1-1.6 (mean = 1.3), hence autosampler data were underestimated arguably
10 through adsorption or biological consumption. We used the mean ratio to correct all storm
11 uncertainty intervals by 30% and the range values to extend the upper limit by 60%. During
12 days with a storm event not monitored at high frequency with an autosampler, we considered
13 that the grab sample data did not contain enough information to derive an acceptability
14 interval for daily SRP load.

15 **2.3.3 Model runs and selection of acceptable models**

16 To explore the parameter space, 15,000 Monte Carlo realisations were performed to simulate
17 daily discharge and SRP load during the water years 2013-2014 and 2014-2015. A 7-month
18 initialisation period was run to reduce the impact of initial conditions on simulated results
19 during the study period, from 1 October 2013 to 31 July 2015.

20 To be considered acceptable, model runs must fall within the acceptability limits defined in
21 2.3.1 and 2.3.2. More specifically, 100% of simulated daily discharge, 100% of simulated
22 baseflow SRP load and 100% of simulated storm SRP load had to fall within the acceptability
23 limits. Thus, 572 acceptability tests were performed for discharge, 378 for baseflow SRP load
24 and 14 for storm SRP loads, i.e. 964 evaluation criteria.

25 To evaluate the model performance in more detail, normalized scores were calculated during
26 6 periods (Table 2). To calculate the scores, a difference was calculated between each of the
27 daily simulated discharge, baseflow SRP load and 2-day storm SRP loads and the
28 corresponding observation. This difference was then normalized by the width of the
29 acceptability limit defined for that day, so the score has a value of 0 in the case of a perfect
30 match with observation, -1 at the lower limit and +1 at the upper limit (Fig. 6a). Finally, the
31 median of this ratio was calculated for each of the 6 periods to investigate whether the model



1 tended to underestimate or overestimate discharge and loads at different moments of the year
2 and between the two years.

3 Model runs were successively evaluated for discharge, baseflow SRP load and storm SRP
4 load. To use the models for prediction, each accepted model was given a likelihood weight
5 according to how well it has performed for each of the 964 evaluation criteria. Here a
6 triangular weight was calculated for each evaluation criteria (Fig. 5 b), with the base of the
7 triangle corresponding to the acceptability limit. Calculated weights were then averaged for
8 discharge, baseflow SRP load and storm SRP load respectively and the final likelihood was
9 calculated as the sum of all three averages.

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

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

26 **3 Results**

27 **3.1 Presentation of observation data and calculation of acceptability limits**

28 The two water years studied were highly contrasted in terms of hydrology and SRP loads.
29 Water year 2013-2014 was the wettest in the last 10 years, with cumulative rainfall 1289 mm
30 and cumulative runoff 716 mm. Water year 2014-2015 was an average year (5th wettest in the
31 last 10 years), with cumulative rainfall 677 mm and cumulative runoff 383 mm. Annual SRP



1 load was $0.35 \text{ kg P ha}^{-1} \text{ yr}^{-1}$ in 2013-2014 and $0.17 \text{ kg P ha}^{-1} \text{ yr}^{-1}$ in 2014-2015, i.e. a
2 difference 10% higher than that of discharge. Observed mean SRP concentration during the
3 study period was 0.024 mg l^{-1} .

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

9 **3.2 Model evaluation**

10 First, model runs were evaluated against acceptability limits defined for discharge (Fig. 8).
11 4,120/15,000 models fulfilled the selection criterion for discharge, i.e. they had 100% of
12 simulated daily discharge within the acceptability limits. The NSE estimated for these models
13 ranged from 0.78 to 0.92. The normalized scores calculated seasonally (Fig. 9a) show that
14 simulated discharge is often overestimated in autumn and spring, and underestimated in
15 winter.

16 Then, model runs were evaluated against acceptability limits defined for SRP loads (Fig. 8b).
17 During baseflow periods, 3,730/15,000 models fulfilled the selection criterion for SRP loads,
18 i.e. they had 100% of simulated daily SRP load within the acceptability limits. Among them,
19 1,210 also fulfilled the previous selection criterion for discharge. Normalized scores for
20 baseflow SRP load showed the same trend as for discharge (Fig. 9b), i.e. overestimation in
21 autumn and spring, and underestimation in winter. During storm events, only 5 models
22 fulfilled the selection criterion for SRP loads, i.e. they had 14/14 of simulated 2-day storm
23 SRP loads within the acceptability limits, but none of them also fulfilled the selection criteria
24 for discharge and baseflow SRP loads. Two storm events were particularly difficult to
25 simulate (number 2 and number 9, Fig. 9c), probably because their acceptability interval was
26 very narrow as a result of only small changes in discharge and concentration. To obtain a
27 reasonable number of acceptable models, we relaxed the selection criterion so that the
28 acceptable models had to simulate 12/14 of storm loads within the acceptability limits, in
29 addition to the selection criteria defined for discharge and baseflow SRP load: 418 models
30 were then accepted. Estimated NSE of these 418 models ranged from 0.09 to 0.80 for daily



1 load and from negative values to 0.53 for daily concentrations (this includes all data from the
2 regular sampling).

3 **3.3 Sensitivity analysis and prediction results**

4 According to the HSA generalised sensitivity analysis, simulated discharge was critically
5 sensitive to 10 out of the 12 hydrological parameters varied. Simulated SRP load was
6 critically sensitive to the sub-surface and overland flow parameters during baseflow periods
7 and to the overland flow parameter during storm events. During baseflow periods, SRP load
8 was insignificantly sensitive to the parameter associated with deep flow load. Both baseflow
9 and storm SRP loads were critically sensitive to the parameter related to soil moisture and soil
10 temperature dependent SRP solubilisation (S1, T1 and T2), in addition to respectively 11 and
11 8 hydrological parameters.

12 Fig. 10 shows the daily discharge, SRP load and concentration as simulated by the acceptable
13 models. Simulated SRP load during the water year 2013-2014 ranged 0.77 – 3.28 kg P ha⁻¹ yr⁻¹
14 (median = 1.62 kg P ha⁻¹ yr⁻¹); simulated SRP load during the water year 2014-2015 ranged
15 0.14 – 0.73 kg P ha⁻¹ yr⁻¹ (median = 0.32 kg P ha⁻¹ yr⁻¹). Best estimate of SRP load according
16 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.
17 According to the model, 56 – 61% (median = 58%) of water discharge and 71 – 75% (median
18 = 62%) of SRP load occurred during storm events. Mean SRP concentrations during the two
19 water years ranged 0.013 – 0.043 mg l⁻¹ (median = 0.028 mg l⁻¹), while mean observed SRP
20 concentration was 0.024 mg l⁻¹.

21 **4 Discussion**

22 **4.1 Role of hydrology and biogeochemistry in determining SRP transfer**

23 The fairly good performance of TNT2-P at simulating SRP loads confirms that the
24 hydrological and biogeochemical processes included into the model are dominant controlling
25 factors in the Kervidy-Naizin catchment. The primary control of hydrology in controlling
26 connectivity between soils and streams has been highlighted by many studies analysing water
27 quality time series at the outlet of agricultural catchments (Haygarth et al., 2012; Jordan et al.,
28 2012; Dupas et al., 2015c; Mellander et al., 2015). This modelling exercise also confirmed
29 that SRP solubility was determined by the soil P Olsen content and could vary according to
30 temperature and moisture conditions. The underlying processes have not been identified



1 precisely in the Kervidy-Naizin catchment: independent laboratory experiments have shown
2 that microbial cell lysis resulting from alternating dry and water saturated periods in the soil
3 could be the cause of increased SRP mobility (Turner and Haygarth, 2001; Blackwell et al.,
4 2009). This could explain the moisture dependence of SRP solubility in the model.
5 Furthermore, net mineralisation of soil organic phosphorus could explain the temperature
6 dependence of SRP solubility in the model. These two hypotheses may explain increased SRP
7 solubility in soils in periods of dry and hot conditions and will be further explored by
8 incubation experiment with soils from the Kervidy-Naizin catchments.

9 **4.2 Potential improvements to the model structure according to modelling** 10 **purpose**

11 The TNT2-P model was designed to test hypotheses about dominant processes and for this
12 purpose, a parsimonious model structure was chosen to include only the processes which were
13 to be tested. This parsimonious model structure might contain some conceptual
14 misrepresentations due to oversimplification, and it might not include all the processes
15 necessary for the purpose of evaluating management scenarios. This section discusses
16 whether the simplifications made are acceptable in the context of different catchment types,
17 and to which conditions the model could be made more complex by including additional
18 routines for the purpose of evaluating management scenarios.

19 From a conceptual point of view, the lack of cell-to-cell routing of SRP fluxes might result in
20 erroneous results in some contexts. The fact that all the SRP emitted from each cell through
21 overland flow and sub-surface flow reaches the stream on the same day is acceptable for the
22 catchment studied because groundwater interception of shallow soil layers occurs in the
23 riparian zone only, hence the signal of SRP mobilisation in these soils is generally transmitted
24 to the stream (Dupas et al., 2015c). This simplification would not be acceptable in catchments
25 where soil-groundwater interactions are taking place throughout the landscape, e.g. due to
26 topographic depressions or poorly drained soils. In the latter type of catchment, transmission
27 of the SRP mobilisation signal to the stream is more complex to comprehend (Haygarth et al.,
28 2012), hence a more complex model structure would be required.

29 The reason for this simplification was that we lacked knowledge of SRP re-adsorption in
30 downslope cells and on the long-term fate of re-adsorbed SRP. For a more physically realistic
31 representation of processes, it is likely that an explicit representation of flow velocities and



1 pathways would be necessary, along with an explicit representation of several soil P pools.
2 However, such an explicit representation of processes contradicts the idea of a parsimonious
3 model, which was adopted here for the purpose of identifying dominant processes. In this
4 respect, TNT2-P is an aggregative model rather than a fully distributed model although it is
5 based on a fully distributed hydrological model (Beaujouan et al., 2002). The current spatial
6 distribution allows finer representation of soil-groundwater interactions than semi-distributed
7 models such as SWAT (Arnold et al., 1998), INCA-P (Wade et al., 2002) and HYPE
8 (Lindstrom et al., 2010) but at higher computation cost. It would be interesting to test to
9 which extent moving from an aggregative model with fully distributed information to a semi-
10 distributed model would degrade the model performance and in the same time reduce
11 computation cost. This could be achieved by grouping cells according to a hydrological
12 similarity criterion like in the original TOPMODEL and Dynamic Topmodel (Beven and
13 Freer, 2001; Metcalfe et al., 2015) and do the same for similarity in soil P content.

14 If reducing the number of calculation units proved to reduce computation cost without
15 degrading quality of prediction, it would be possible to include more parameters in the model,
16 for example to simulate SRP re-absorption in downslope cells or include routines to simulate
17 the evolution of soil P content under different management scenarios (Vadas et al., 2011;
18 2012), and still perform a Monte-Carlo based analysis of uncertainty. The question of
19 coupling or not such a soil P routine with the current TNT2-P model will depend on available
20 data and on the length of available time series: studying the evolution of the soil P content
21 requires at least a decade of soil observation data (Ringeval et al., 2014) and probably a
22 longer period of stream data to account for the time delay for a perturbation in the catchment
23 to become visible in the stream (Wall et al., 2013). Thus, the two years of daily stream SRP in
24 the Kervidy-Naizin catchment are not enough to build a coupled soil-hydrology model with
25 an elaborate soil P routine. Therefore, as things stand, it is more reasonable to generate new
26 soil P Olsen maps with a separate model such as the APLE model (Vadas et al., 2012;
27 Benskin et al., 2014) or the 'soil P decline' model used by Wall et al. (2013), and use these
28 maps as input to TNT2-P.

29 Because the current model can simulate response to rainfall, soil moisture and temperature, it
30 could be used to test the effect of climate scenarios on SRP transfer. In Western France, and
31 more generally in Western Europe, the climate for the next few decades is expected to consist
32 of hotter, drier summers and warmer, wetter winter (Jacob et al., 2007; Macleod et al., 2012;



1 Salmon-Monviola et al., 2013) with increased frequency of high intensity rainfall events
2 (Dequé 2007). In these conditions, SRP concentrations and load will seemingly increase
3 compared to today's climate as a result of both an increase in SRP solubility in soil due to
4 higher temperature and more severe drought and an increase in transfer due to wetter winter
5 and more frequent high intensity rainfall events. TNT2-P could be used to confirm and
6 quantify the expected increase in SRP transfer from diffuse sources in future climate
7 conditions.

8 **4.3 Improving information content in the data**

9 Despite relatively large uncertainty in the data used in this study, it was possible to build a
10 parsimonious catchment model of SRP transfer for the purpose of testing hypotheses about
11 dominant processes, namely the role of hydrology in controlling connectivity between soils
12 and streams and the role of temperature and moisture conditions in controlling soil SRP
13 solubilisation. However, the large uncertainties in the calibration data lead to large prediction
14 uncertainty. For example, the SRP load estimated by the behavioural models from 2013 to
15 2015 ranged from 0.45 to 2.0 kg P ha⁻¹ yr⁻¹; hence the width of the credibility interval was
16 160% of the median (0.97 kg P ha⁻¹ yr⁻¹). Similarly, the mean SRP concentration estimated by
17 the behavioural models from 2013 to 2015 ranged from 0.013 to 0.045 mg l⁻¹; hence the width
18 of the credibility interval was 110% of the median (0.028 mg l⁻¹). The large uncertainty in the
19 calibration data, along with a lack of long-term information, also prevents including more
20 detailed processes in the soil routine.

21 To reduce uncertainty in prediction and to build more complex models, several options exist
22 to improve information content in the data. As stated by Jackson-Blake et al. (2015b), “the
23 key to obtaining a realistic model simulation is ensuring that the natural variability in water
24 chemistry is well represented by the monitoring data”. The monitoring strategy adopted in the
25 Kervidy-Naizin catchment should theoretically enable to capture the natural variability in
26 stream SRP concentration, because sampling took place during two contrasting water years,
27 during different seasons and at a high frequency during 14 storm events. The analysis of
28 uncertainty in the data shows that a large part of uncertainty in “observed” SRP concentration
29 originates from sample storage, both unfiltered between the time of autosampling and manual
30 filtration and between filtration and analysis. This is due to SRP being non-conservative.
31 Thus, there is room for improvement in reducing storage time, without increasing further the
32 monitoring frequency. In this respect, the primary interest of investing in high frequency



1 bankside analysers would lie in their ability to analyse water samples immediately in addition
2 to providing near continuous data. Because bankside analysers perform measurements in
3 relatively homogeneous conditions, unlike the manual and autosampler data for which storage
4 time of filtered and unfiltered samples vary, a finer quantification of uncertainty in the
5 measurement data would be possible (e.g. Lloyd et al., 2015).

6 **5 Conclusion**

7 The TNT2-P model was capable of capturing daily variation of SRP loads, thus confirming
8 the dominant processes identified in previous analyses of observation data in the Kervidy-
9 Naizin catchment. The role of hydrology in controlling connectivity between soils and
10 streams, and the role of soil Olsen P, soil moisture and temperature in controlling SRP
11 solubility have been confirmed. The lack of any representation of the short-term effect of
12 management practices did not seem to penalize the model's performance. Their long-term
13 effect on the soil Olsen P could be simulated with an independent model or through an
14 additional sub-model if a longer period of data was available to calibrate it. The modelling
15 approach presented in this paper included an assessment of the information content in the
16 data, and propagation of uncertainty in the model's prediction. The information content of the
17 data was sufficient to explore dominant processes, but the relatively large uncertainty in SRP
18 concentrations would seemingly limit the possibility for including more detailed processes
19 into the model. Data from near continuous bankside analyser will probably allow calibrating
20 more detailed models in the near future.

21 **References**

- 22 Alexander RB, Smith RA, Schwarz GE, Boyer EW, Nolan JV, Brakebill JW. Differences in
23 phosphorus and nitrogen delivery to the gulf of Mexico from the Mississippi river basin.
24 *Environmental Science & Technology* 2008; 42: 822-830.
- 25 Arnold JG, Srinivasan R, Muttiah RS, Williams JR. Large area hydrologic modeling and
26 assessment - Part 1: Model development. *Journal of the American Water Resources*
27 *Association* 1998; 34: 73-89.
- 28 Aubert AH, Gascuel-Odoux C, Gruau G, Akkal N, Fauchoux M, Fauvel Y, et al. Solute
29 transport dynamics in small, shallow groundwater-dominated agricultural catchments:
30 insights from a high-frequency, multisolute 10 yr-long monitoring study. *Hydrology and*
31 *Earth System Sciences* 2013; 17: 1379-1391.



- 1 Beauchemin S, Simard RR. Soil phosphorus saturation degree: Review of some indices and
2 their suitability for P management in Quebec, Canada. Canadian Journal of Soil Science
3 1999; 79: 615-625.
- 4 Beaujouan V, Durand P, Ruiz L, Arousseau P, Cotteret G. A hydrological model dedicated
5 to topography-based simulation of nitrogen transfer and transformation: rationale and
6 application to the geomorphology-denitrification relationship. Hydrological Processes 2002;
7 16: 493-507.
- 8 Benskin CMH, Roberts W. M, Wang Y, Haygarth PM. Review of the Annual Phosphorus
9 Loss Estimator tool – a new model for estimating phosphorus losses at the field scale. Soil
10 Use and Management 2014; 30: 337-341.
- 11 Beven K. A manifesto for the equifinality thesis. Journal of Hydrology 2006; 320: 18-36.
- 12 Beven K. Environmental Modelling – An Uncertain Future? Routledge: London 2009.
- 13 Beven K, Freer J. Equifinality, data assimilation, and uncertainty estimation in mechanistic
14 modelling of complex environmental systems using the GLUE methodology. Journal of
15 Hydrology 2001; 249: 11-29.
- 16 Beven K, Smith P. Concepts of Information Content and Likelihood in Parameter Calibration
17 for Hydrological Simulation Models. Journal of Hydrologic Engineering 2015; 20.
- 18 Beven KJ. Distributed hydrological modelling: applications of the TOPMODEL concept,
19 1997.
- 20 Blackwell MSA, Brookes PC, de la Fuente-Martinez N, Murray PJ, Snars KE, Williams JK,
21 et al. Effects of soil drying and rate of re-wetting on concentrations and forms of phosphorus
22 in leachate. Biology and Fertility of Soils 2009; 45: 635-643.
- 23 Blazkova S, Beven K. A limits of acceptability approach to model evaluation and uncertainty
24 estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech
25 Republic. Water Resources Research 2009; 45.
- 26 Bradley WG, Borenstein AR, Nelson LM, Codd GA, Rosen BH, Stommel EW, et al. Is
27 exposure to cyanobacteria an environmental risk factor for amyotrophic lateral sclerosis and
28 other neurodegenerative diseases? Amyotrophic Lateral Sclerosis and Frontotemporal
29 Degeneration 2013; 14: 325-333.



- 1 Bruneau P, Gascuel-Oudou C, Robin P, Merot P, Beven KJ. Sensitivity to space and time
2 resolution of a hydrological model using digital elevation data. *Hydrological Processes* 1995;
3 9: 69-82.
- 4 Carluer N. Vers une modélisation hydrologique adaptée à l'évaluation des pollutions diffuses:
5 prise en compte du réseau anthropique. Application au bassin versant de Naizin (Morbihan).
6 PhD thesis Université Pierre et Marie Curie 1998.
- 7 Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH. Nonpoint
8 pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* 1998; 8:
9 559-568.
- 10 Curmi P, Durand P, Gascuel-Oudou C, Merot P, Walter C, Taha A. Hydromorphic soils,
11 hydrology and water quality: spatial distribution and functional modelling at different scales.
12 *Nutrient Cycling in Agroecosystems* 1998; 50: 127-142.
- 13 Dean S, Freer J, Beven K, Wade AJ, Butterfield D. Uncertainty assessment of a process-based
14 integrated catchment model of phosphorus. *Stochastic Environmental Research and Risk*
15 *Assessment* 2009; 23: 991-1010.
- 16 Deque M. Frequency of precipitation and temperature extremes over France in an
17 anthropogenic scenario: Model results and statistical correction according to observed values.
18 *Global and Planetary Change* 2007; 57: 16-26.
- 19 Dupas R, Delmas M, Dorioz JM, Garnier J, Moatar F, Gascuel-Oudou C. Assessing the
20 impact of agricultural pressures on N and P loads and eutrophication risk. *Ecological*
21 *Indicators* 2015a; 48: 396-407.
- 22 Dupas R, Gascuel-Oudou C, Gilliet N, Grimaldi C, Gruau G. Distinct export dynamics for
23 dissolved and particulate phosphorus reveal independent transport mechanisms in an arable
24 headwater catchment. *Hydrological Processes* 2015b.
- 25 Dupas R, Gruau G, Gu S, Humbert G, Jaffrezic A, Gascuel-Oudou C. Groundwater control of
26 biogeochemical processes causing phosphorus release from riparian wetlands. *Water*
27 *Research* 2015c.
- 28 Dupas R, Tavenard R, Fovet O, Gilliet N, Grimaldi C, Gascuel-Oudou C. Identifying seasonal
29 patterns of phosphorus storm dynamics with Dynamic Time Warping. *Water Resources*
30 *Research* 2015d.



- 1 Durand P, Moreau P, Salmon-Monviola J, Ruiz L, Vertes F, Gascuel-Oudoux C. Modelling the
2 interplay between nitrogen cycling processes and mitigation options in farming catchments.
3 Journal of Agricultural Science 2015; 153: 959-974.
- 4 Franks SW, Gineste P, Beven KJ, Merot P. On constraining the predictions of a distributed
5 model: the incorporation of fuzzy estimates of saturated areas into the calibration process,
6 Water Resources Research 1998; 34: 787-797.
- 7 Grizzetti B, Bouraoui F, Aloe A. Changes of nitrogen and phosphorus loads to European seas.
8 Global Change Biology 2012; 18: 769-782.
- 9 Hahn C, Prasuhn V, Stamm C, Lazzarotto P, Evangelou MWH, Schulin R. Prediction of
10 dissolved reactive phosphorus losses from small agricultural catchments: calibration and
11 validation of a parsimonious model. Hydrology and Earth System Sciences 2013; 17: 3679-
12 3693.
- 13 Hahn C, Prasuhn V, Stamm C, Schulin R. Phosphorus losses in runoff from manured
14 grassland of different soil P status at two rainfall intensities. Agriculture Ecosystems &
15 Environment 2012; 153: 65-74.
- 16 Haygarth PM, Ashby CD, Jarvis SC. Short-term changes in the molybdate reactive
17 phosphorus of stored soil waters. Journal of Environmental Quality 1995; 24: 1133-1140.
- 18 Haygarth PM, Hepworth L, Jarvis SC. Forms of phosphorus transfer in hydrological pathways
19 from soil under grazed grassland. European Journal of Soil Science 1998; 49: 65-72.
- 20 Haygarth PM, Page TJC, Beven KJ, Freer J, Joynes A, Butler P, et al. Scaling up the
21 phosphorus signal from soil hillslopes to headwater catchments. Freshwater Biology 2012;
22 57: 7-25.
- 23 Heathwaite AL, Dils RM. Characterising phosphorus loss in surface and subsurface
24 hydrological pathways. Science of the Total Environment 2000; 251: 523-538.
- 25 Heckrath G, Brookes PC, Poulton PR, Goulding KWT. Phosphorus leaching from soils
26 containing different phosphorus concentrations in the broadbalk experiment. Journal of
27 Environmental Quality 1995; 24: 904-910.
- 28 Hornberger GM, Spear RC. An approach to the preliminary analysis of environmental
29 systems. J. Environmental Management 1981; 12: 7-18.



- 1 Jackson-Blake LA, Dunn SM, Helliwell RC, Skeffington RA, Stutter MI, Wade AJ. How well
2 can we model stream phosphorus concentrations in agricultural catchments? *Environmental*
3 *Modelling & Software* 2015a; 64: 31-46.
- 4 Jackson-Blake LA, Starrfelt J. Do higher data frequency and Bayesian auto-calibration lead to
5 better model calibration? Insights from an application of INCA-P, a process-based river
6 phosphorus model. *Journal of Hydrology* 2015b; 527: 641-655.
- 7 Jacob D, Barring L, Christensen OB, Christensen JH, de Castro M, Deque M, et al. An inter-
8 comparison of regional climate models for Europe: model performance in present-day
9 climate. *Climatic Change* 2007; 81: 31-52.
- 10 Jarvie HP, Withers PJA, Neal C. Review of robust measurement of phosphorus in river water:
11 sampling, storage, fractionation and sensitivity. *Hydrology and Earth System Sciences* 2002;
12 6: 113-131.
- 13 Jordan P, Melland AR, Mellander PE, Shortle G, Wall D. The seasonality of phosphorus
14 transfers from land to water: implications for trophic impacts and policy evaluation. *Sci Total*
15 *Environ* 2012; 434: 101-9.
- 16 Kirchner JW. Getting the right answers for the right reasons: Linking measurements,
17 analyses, and models to advance the science of hydrology. *Water Resources Research* 2006;
18 42.
- 19 Humbert G, Jaffrezic A, Fovet O, Gruau G, Durand P. Dry-season length and runoff control
20 annual variability in stream DOC dynamics in a small, shallow groundwater-dominated
21 agricultural watershed. *Water Resources Research* 2015.
- 22 Lazzarotto P, Stamm C, Prasuhn V, Flüßler H. A parsimonious soil-type based rainfall-runoff
23 model simultaneously tested in four small agricultural catchments. *Journal of Hydrology*
24 2006; 321: 21-38.
- 25 Lindstrom G, Pers C, Rosberg J, Stromqvist J, Arheimer B. Development and testing of the
26 HYPE (Hydrological Predictions for the Environment) water quality model for different
27 spatial scales. *Hydrology Research* 2010; 41: 295-319.
- 28 Lloyd CEM, Freer JE, Johnes PJ, Coxon G, Collins AL. Discharge and nutrient uncertainty:
29 implications for nutrient flux estimation in small streams. *Hydrological processes* 2015.



- 1 Macleod CJA, Falloon PD, Evans R, Haygarth PM. The effects of climate change on the
2 mobilization of diffuse substances from agricultural systems. In: Sparks DL, editor. *Advances*
3 in *Agronomy*, Vol 115. 115, 2012, pp. 41-77.
- 4 Maguire RO, Sims JT. Soil testing to predict phosphorus leaching. *Journal of Environmental*
5 *Quality* 2002; 31: 1601-1609.
- 6 Matos-Moreira M, Lemerrier B, Michot D, Dupas R, Gascuel-Oudou C. Using agricultural
7 practices information for multiscale environmental assessment of phosphorus risk.
8 *Geophysical Research Abstracts* 2015; 17.
- 9 McDowell R, Sharpley A, Withers P. Indicator to predict the movement of phosphorus from
10 soil to subsurface flow. *Environmental Science & Technology* 2002; 36: 1505-1509.
- 11 Mellander PE, Jordan P, Shore M, Melland AR, Shortle G. Flow paths and phosphorus
12 transfer pathways in two agricultural streams with contrasting flow controls. *Hydrological*
13 *Processes* 2015.
- 14 Metcalfe P, Beven BJ, and Freer J. Dynamic Topmodel: a new implementation in R and its
15 sensitivity to time and space steps. *Environmental Modelling and Software* 2015; 72: 155-
16 172.
- 17 Molenat J, Gascuel-Oudou C, Ruiz L, Gruau G. Role of water table dynamics on stream
18 nitrate export and concentration. in agricultural headwater catchment (France). *Journal of*
19 *Hydrology* 2008; 348: 363-378.
- 20 Moore MT, Locke MA. Effect of Storage Method and Associated Holding Time on Nitrogen
21 and Phosphorus Concentrations in Surface Water Samples. *Bulletin of Environmental*
22 *Contamination and Toxicology* 2013; 91: 493-498.
- 23 Moreau P, Ruiz L, Mabon F, Raimbault T, Durand P, Delaby L, et al. Reconciling technical,
24 economic and environmental efficiency of farming systems in vulnerable areas. *Agriculture*
25 *Ecosystems & Environment* 2012; 147: 89-99.
- 26 Moreau P, Viaud V, Parnaudeau V, Salmon-Monviola J, Durand P. An approach for global
27 sensitivity analysis of a complex environmental model to spatial inputs and parameters: A
28 case study of an agro-hydrological model. *Environmental Modelling & Software* 2013; 47:
29 74-87.



- 1 Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. Model
2 evaluation guidelines for systematic quantification of accuracy in watershed simulations.
3 Transactions of the Asabe 2007; 50: 885-900.
- 4 Olsen SR, Cole CV, Watanbe FS, Dean LA. Estimation of available phosphorus in soils by
5 extraction with sodium bicarbonate 1954.. Circ. 939. USDA, Washington, DC.
- 6 Outram FN, Lloyd CEM, Jonczyk J, Benskin CMH, Grant F, Perks MT, et al. High-frequency
7 monitoring of nitrogen and phosphorus response in three rural catchments to the end of the
8 2011-2012 drought in England. Hydrology and Earth System Sciences 2014; 18: 3429-3448.
- 9 Page T, Haygarth PM, Beven KJ, Joynes A, Butler T, Keeler C, et al. Spatial variability of
10 soil phosphorus in relation to the topographic index and critical source areas: Sampling for
11 assessing risk to water quality. Journal of Environmental Quality 2005; 34: 2263-2277.
- 12 Perks MT, Owen GJ, Benskin CMH, Jonczyk J, Deasy C, Burke S, et al. Dominant
13 mechanisms for the delivery of fine sediment and phosphorus to fluvial networks draining
14 grassland dominated headwater catchments. Science of the Total Environment 2015; 523:
15 178-190.
- 16 Quinlan, J.R. Learning with continuous classes. Proceedings of the 5th Australian Joint
17 Conference On Artificial Intelligence 1992, 343-348.
- 18 Ringeval B, Nowak B, Nesme T, Delmas M, Pellerin S. Contribution of anthropogenic
19 phosphorus to agricultural soil fertility and food production. Global Biogeochemical Cycles
20 2014; 28: 743-756.
- 21 Salmon-Monviola J, Moreau P, Benhamou C, Durand P, Merot P, Oehler F, et al. Effect of
22 climate change and increased atmospheric CO₂ on hydrological and nitrogen cycling in an
23 intensive agricultural headwater catchment in western France. Climatic Change 2013; 120:
24 433-447.
- 25 Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, Paterson MJ, et al.
26 Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year
27 whole-ecosystem experiment. Proceedings of the National Academy of Sciences of the United
28 States of America 2008; 105: 11254-11258.
- 29 Schoumans OF, Chardon WJ. Phosphate saturation degree and accumulation of phosphate in
30 various soil types in The Netherlands. Geoderma 2015; 237: 325-335.



- 1 Serrano T, Dupas R, Upegui E, Buscail C, Grimaldi C, Viel J-F. Geographical modeling of
2 exposure risk to cyanobacteria for epidemiological purposes. *Environment International* 2015;
3 81: 18-25.
- 4 Sharpley AN, Kleinman PJ, Heathwaite AL, Gburek WJ, Folmar GJ, Schmidt JP. Phosphorus
5 loss from an agricultural watershed as a function of storm size. *J Environ Qual* 2008; 37: 362-
6 8.
- 7 Siwek J, Siwek JP, Zelazny M. Environmental and land use factors affecting phosphate
8 hysteresis patterns of stream water during flood events (Carpathian Foothills, Poland).
9 *Hydrological Processes* 2013; 27: 3674-3684.
- 10 Turner BL, Haygarth PM. Biogeochemistry - Phosphorus solubilization in rewetted soils.
11 *Nature* 2001; 411: 258-258.
- 12 Vadas PA, Joern BC, Moore PA. Simulating soil phosphorus dynamics for a phosphorus loss
13 quantification tool. *J Environ Qual* 2012; 41: 1750-7.
- 14 Vadas PA, Jokela WE, Franklin DH, Endale DM. The Effect of Rain and Runoff When
15 Assessing Timing of Manure Application and Dissolved Phosphorus Loss in Runoff1.
16 *JAWRA Journal of the American Water Resources Association* 2011; 47: 877-886.
- 17 Wade AJ, Whitehead PG, Butterfield D. The Integrated Catchments model of Phosphorus
18 dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river
19 systems: model structure and equations. *Hydrology and Earth System Sciences* 2002; 6: 583-
20 606.
- 21 Wall DP, Jordan P, Melland AR, Mellander PE, Mechan S, Shortle G. Forecasting the decline
22 of excess soil phosphorus in agricultural catchments. *Soil Use and Management* 2013; 29:
23 147-154.
- 24 Whitehead P, Young P. Water-quality in river systems – Monte-Carlo analysis. *Water*
25 *Resources Research* 1979; 15: 451-459.

26 **Acknowledgements**

27 This work was funded by the “Agence de l’Eau Loire Bretagne” via the “Trans-P project”.
28 Long-term monitoring in the Kervidy-Naizin catchment is supported by “ORE AgrHyS”.
29 Data of “ORE AgrHyS” can be downloaded from http://www6.inra.fr/ore_agrhys/Donnees.

30



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

	Abbreviation	Unit	Hydrological (H), Phosphorus model (P)	Range poorly drained soils (min-max)	Range well drained soils (min-max)
Lateral transmissivity at saturation	T	m ² d ⁻¹	H	4-8	-> x1.5
Exponential decay rate of hydraulic conductivity with depth	m	m ² d ⁻¹	H	0.02-0.2	0.02-0.2
Soil depth	ho	m	H	0.3-0.8	-> x1
Drainage porosity of soil	po	cm ³ cm ⁻³	H	0.1-0.4	-> x1
Regolith layer thickness	h1	m	H	5-10	-> x4
Exponent for evaporation limit	α	-	H	8 (fixed)	-> x1
kRC parameter for capillary rise	kRC	-	H	0.001 (fixed)	-> x1
n parameter for capillarity rise	n	-	H	2.5 (fixed)	-> x1
Drainage porosity of regolith layer	p1	cm ³ cm ⁻³	H	0.01-0.05	-> x1
Background P release coefficient for subsurface flow	Coef _{SRP} overland	-	P	0-0.015	-> x1
Background P release coefficient for overland flow	Coef _{SRP} sub-surface	-	P	0-0.25	-> x1
Temperature coefficient 1	T1	-	P	5-10	-> x1
Temperature coefficient 2	T2	-	P	2-10	-> x1



Soil moisture coefficient	S1	-	P	0-2	-> x1
SRP concentration in deep flow	SRP_deep	mg l ⁻¹	P	0-0.007	-> x1

1

2 Table 2: Starting and ending dates of periods studied

Name	Starting date	Ending date
Autumn 2013	01 October 2013	31 December 2013
Winter 2014	01 January 2014	31 March 2014
Spring 2014	01 April 2014	31 July 2014
Autumn 2014	01 October 2014	31 December 2014
Winter 2015	01 January 2015	31 March 2015
Spring 2015	01 April 2015	31 July 2015

3

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

6

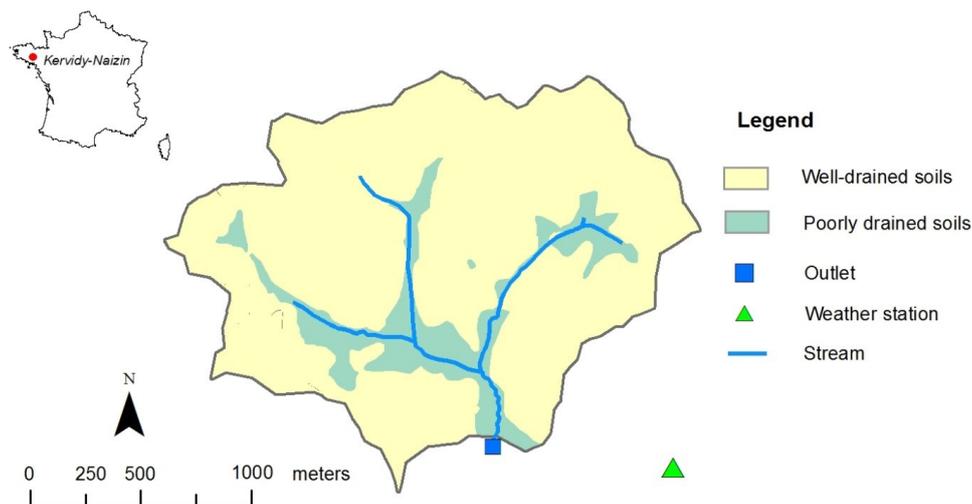
	discharge	baseflow SRP load	storm SRP load
T (poorly drained soils)	.	***	***
m (poorly drained soils)	***	***	***
ho (poorly drained soils)	***	***	.
po (poorly drained soils)	***	***	***
h1 (poorly drained soils)	***	.	.
p1 (poorly drained soils)	***	***	*



soils)				
T (well drained soils)	.		***	***
m (well drained soils)	***		***	***
soils)				
ho (well drained soils)	***		***	.
soils)				
po (well drained soils)	***		***	***
soils)				
h1 (well drained soils)	***		***	.
soils)				
p1 (well drained soils)	***		***	*
soils)				
Coef_sub-surface	.		***	.
Coef_overland	.		***	***
SRP_deep	.	.		*
S1	.		***	***
T1	.		***	***
T2	.		***	***

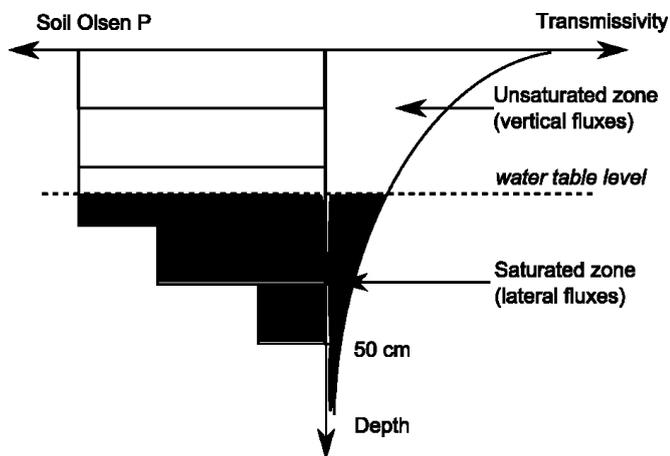
1

2



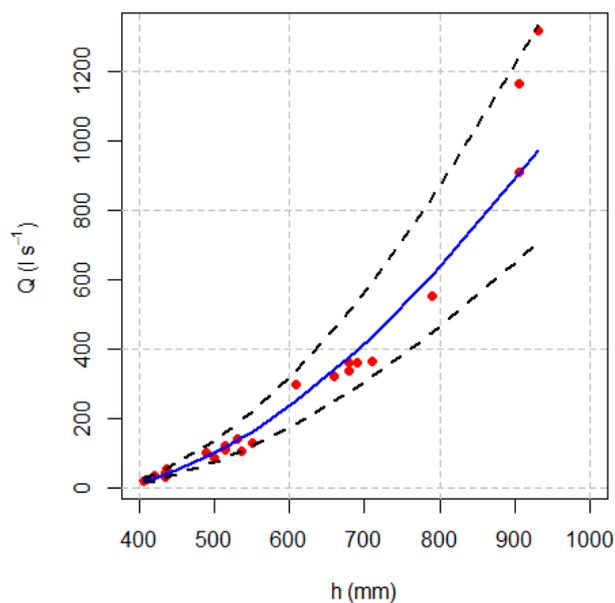
1

2 Fig. 1. Soil drainage classes in the Kervidy-Naizin catchment, Curmi et al. (1998)



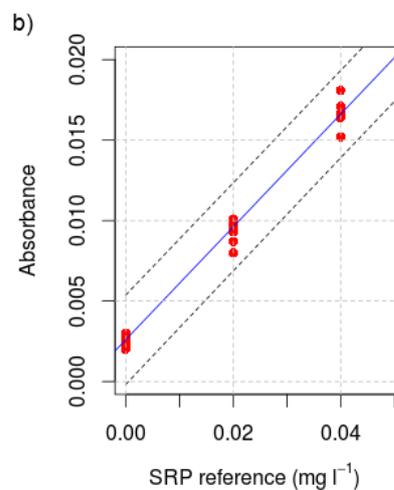
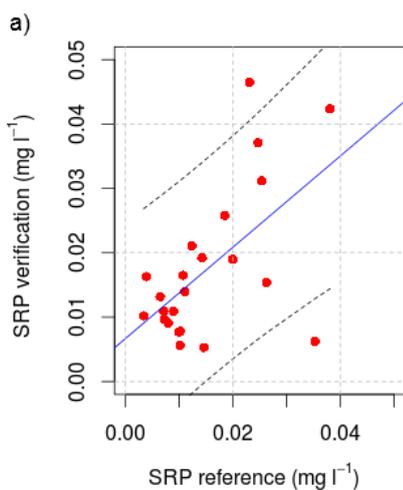
3

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



1

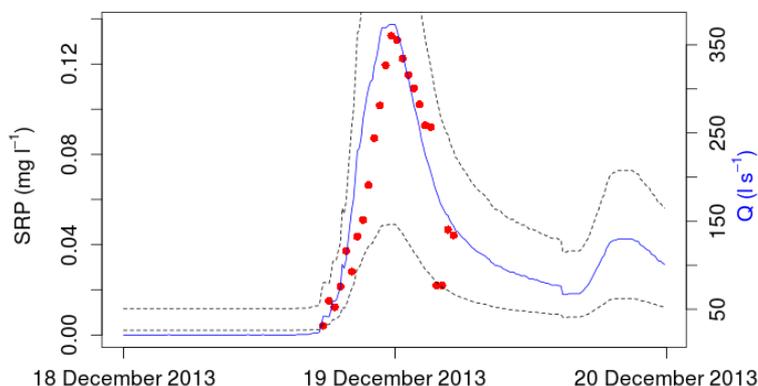
2 Fig. 3 : Rating curve in Kervidy-Naizin; acceptability bounds derived from 90% prediction
 3 interval (blue line: fitting regression; black dots: 90% prediction interval). Red dots represent
 4 the original discharge measurements used to calibrate the stage-discharge rating curve
 5 (Carlier, 1998).



6

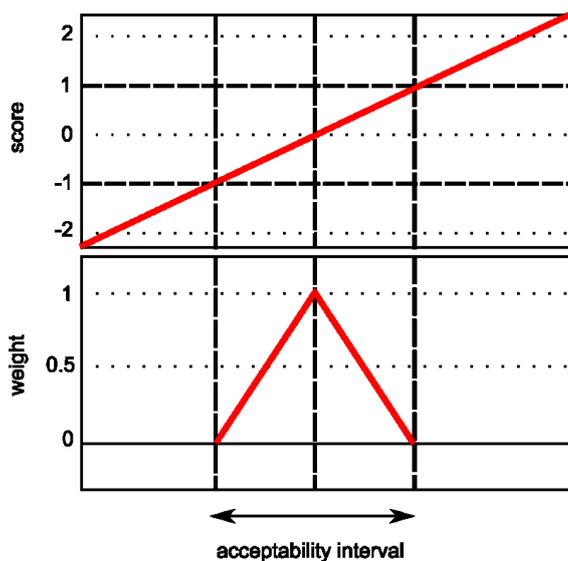


1 Fig. 4: a) linear regression model linking the reference data and a verification dataset; b)
 2 measurement error as estimated from a repeatability test performed by the lab in charge of
 3 producing reference data (blue line: fitting regression; black dots: 90% prediction interval).
 4



5 18 December 2013 19 December 2013 20 December 2013
 6 Fig. 5: Example of an empirical concentration – discharge model; acceptability bounds
 7 derived from 90% prediction interval. Red circles represent the SRP measurements.

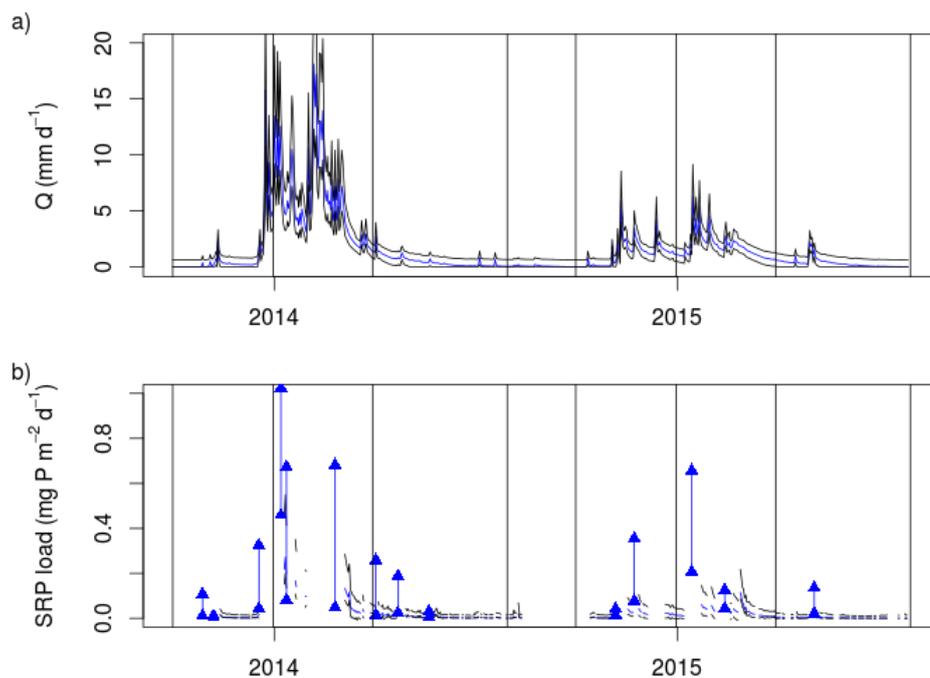
8



9

10 Fig. 6 : a) normalized scores; b) triangular weighting function

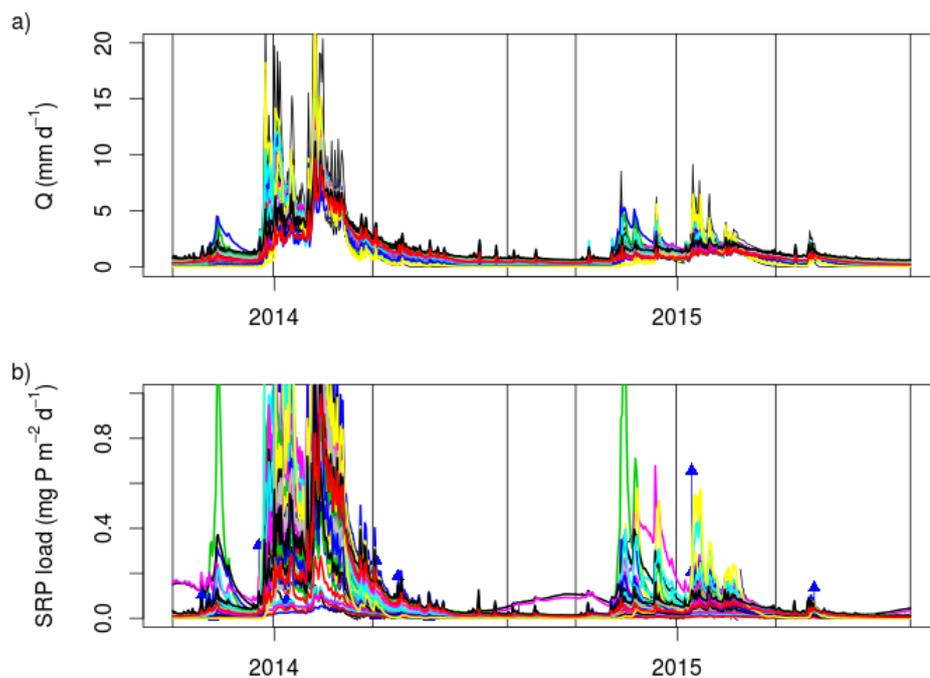
11



1

2 Fig. 7: Acceptability limits for daily discharge (a) and SRP load (b). Blue lines represent best
3 estimates; black lines represent the acceptability limits. Storm loads acceptability limits are
4 represented by vertical blue lines. Black vertical lines represent the starting and ending dates
5 for each season (table 2).

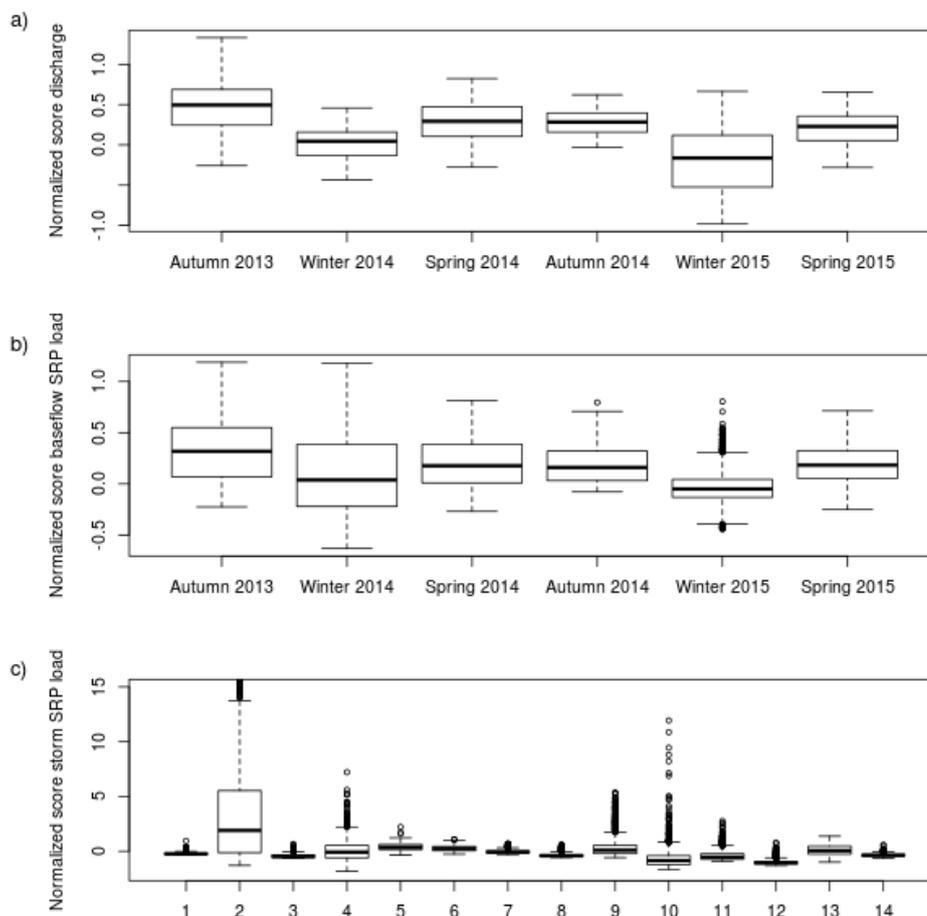
6



1

2 Fig. 8: Example of 50 model runs simulating discharge (a) and daily load (b). Vertical lines
3 represent the starting and ending dates for each season (table 2).

4

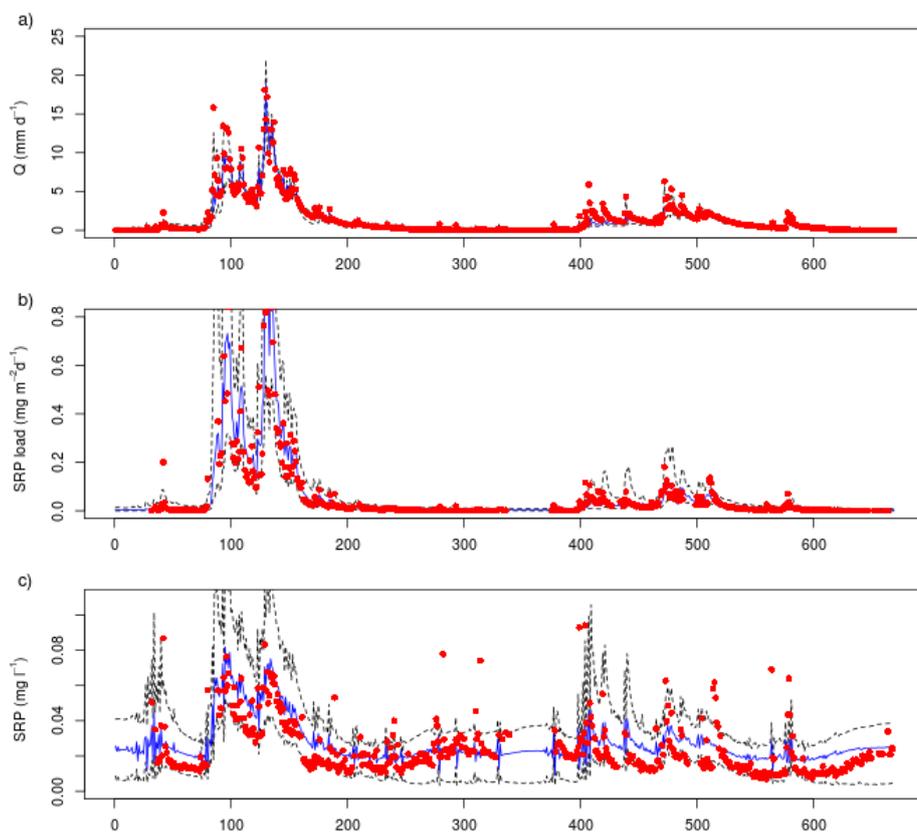


1

2 Fig. 9: Normalized score for daily discharge (a), baseflow SRP load

3 (c).

4



1

2 Fig. 10: Median and 95% credibility interval for daily discharge (a), SRP load (b) and SRP
3 concentration (c). Red circles represent observational data.

4