Are current dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P

Article

Accepted Version

Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P

L.A. Jackson-Blake[a,b], J.E. Sample[a,b], A.J. Wade[c], R.C. Helliwell[b], R.A. Skeffington[c]

[a] Norwegian Institute for Water Research (NIVA), Gaustadalleen 21, 0349 Oslo, Norway

[b] The James Hutton Institute, Macaulay Drive, Aberdeen, AB15 8QH, UK

[c] Department of Geography and Environmental Science, University of Reading, RG6 6DW, UK

Corresponding author: Leah Jackson-Blake (ljb@niva.no)

Key Points:

- We developed a new parsimonious dynamic catchment phosphorus model, SimplyP
- SimplyP performed as well in calibration, validation and scenarios as a well-established, substantially more complex model
- Results support the hypothesis that water quality models are too complex and suggest wider simplification exercises may be warranted

Abstract

Catchment-scale water quality models are increasingly popular tools for exploring the potential effects of land management, land use change and climate change on water quality. However, the dynamic, catchment-scale nutrient models in common usage are complex, with many uncertain parameters requiring calibration, limiting their usability and robustness. A key question is whether this complexity is justified. To explore this, we developed a parsimonious phosphorus model, *SimplyP*, incorporating a rainfall-runoff model and a biogeochemical model able to simulate daily streamflow, suspended sediment, and particulate and dissolved phosphorus dynamics. The model's complexity was compared to one popular nutrient model, INCA-P, and the performance of the two models was compared in a small rural catchment in northeast Scotland. For three land use classes, less than six SimplyP parameters must be determined through calibration, the rest may be based on measurements, whilst INCA-P has around 40 unmeasurable parameters. Despite substantially simpler process-representation, SimplyP performed comparably to INCA-P in both calibration and validation and produced similar long-term projections in response to changes in land management. Results support the hypothesis that INCA-P is overly complex for the study catchment. We hope our findings will help prompt wider model comparison exercises, as well as debate amongst the water quality modelling community as to whether today's models are fit for purpose. Simpler models such as SimplyP have the potential to be useful management and research tools, building blocks for future model development (prototype code is freely available), or benchmarks against which more complex models could be evaluated.

1. Introduction

Dynamic, process-based integrated catchment models are designed to represent the processes governing catchment hydrology and water quality, and may therefore be useful tools for catchment management. Models can be used, for example, to interpolate sparse monitoring data, to highlight knowledge and data gaps and help design monitoring strategies, as well as to provide evidence to support decision making. Integrated catchment models are increasingly called upon, for example, to explore potential future water quality under scenarios of changing management, land use and climate, often as part of wider integrated modelling frameworks (Jakeman & Letcher, 2003, Martin-Ortega *et al.*, 2015). As a result, many catchment-scale nutrient and sediment models have been developed during the last few decades.

In a recent review, Wellen *et al.* (2015) found that the majority of recent nutrient modelling studies used just five models: SWAT (Arnold & Fohrer, 2005, Arnold *et al.*, 1998), INCA (Wade *et al.*, 1999, Wade *et al.*, 2002, Whitehead *et al.*, 1998), AnnAGNPS (Binger & Theurer, 2005, Young *et al.*, 1989), HSPF (Bicknell *et al.*, 2001, Donigian Jr *et al.*, 1995) and HBV, now superseded by HYPE (Lindström *et al.*, 2010). These models are semi-distributed, mass balance models, which simulate the daily dynamics of nutrient transport in catchments. Although the models differ in their structures, each requires on the order of >50 to 100s of user-supplied parameters for hydrology, sediment and phosphorus (P) simulations to be performed. Many of these cannot be measured directly and are highly uncertain, and their values must therefore be determined by calibrating the model to observations. In most catchments, calibration is carried out using end-of-pipe measurements of discharge and water chemistry, the latter often only sampled infrequently. Previous analyses have suggested that there is only enough information in discharge data to constrain a small number (<6) of hydrology model parameters

during calibration (Jakeman & Hornberger, 1993), and it seems unlikely that in-stream concentration data will provide enough additional information for tens or even hundreds of additional parameters to be constrained (Kirchner, 2006). Overall, there is a growing awareness that, given data and process knowledge limitations, current catchment water quality models are overly-complex and there is a need for more parsimonious models that capture the dominant modes of behavior at the scale of interest (Jackson-Blake & Starrfelt, 2015, Jakeman *et al.*, 2006, Kirchner, 2006, Krueger *et al.*, 2007, Radcliffe *et al.*, 2009). Examples of some of the new simpler dynamic nutrient models being developed include the Rainfall-Runoff Phosphorus model (Hahn *et al.*, 2013, Van Meter & Basu, 2015), which focuses on identifying critical source areas and simulating discharge and soluble reactive phosphorus (SRP) concentration in grassland catchments, and an analytical model for quantifying time lags between implementing measures and seeing reductions in surface water nitrate concentrations (Van Meter & Basu, 2015). These models are relatively limited in their aims and scope, and there is a need to assess to what extent the aims of more complex models like INCA and SWAT, which attempt to simulate a wider range of nutrient species and processes, could be achieved using simpler modelling approaches.

There are sound theoretical reasons for why simpler models should be chosen over more complex ones (see e.g. MacKay, 2003, Chapter 28). With over-parameterized models, different parameter sets will give almost identical fits to the calibration data (e.g. Beven & Binley, 1992), and yet may yield very different predictions of how the system will behave as conditions change, and can therefore perform poorly in validation (Seibert, 2003). Model complexity often therefore turns out to be unjustified in practice (e.g. Perrin *et al.*, 2001). There are related practical reasons for choosing simpler models. Firstly, over-complexity leads to difficulties in using the model to test hypotheses about the dominant processes operating within a catchment, as the real processes may be masked by too much flexibility introduced by unnecessary parameters (Jakeman *et al.*, 2006, Kirchner, 2006). Secondly, large parameter spaces lead to difficulties in model calibration because of parameter non-identifiability, whether due to structural or practical non-identifiability (as defined in Raue *et al.*, 2009). There are then related problems with auto-calibration, sensitivity and uncertainty analyses, as the parameter space increases exponentially with the number of parameters, eventually becoming too large to be searched within realistic time frames. Uncertainty analyses are becoming a pre-requisite for model applications, yet as only subsets of the parameter space of more complex models can be searched, these analyses become somewhat subjective and incomplete (Jackson-Blake & Starrfelt, 2015, Pappenberger *et al.*, 2007), and the meaningfulness of estimated uncertainty intervals is often questionable. Finally, as models become more complex they become more time-consuming to set up and require larger calibration and evaluation datasets, increasing the financial burden associated with model applications and limiting the size of the user group. Ultimately, over-complexity therefore reduces a model's usefulness for supporting research and real world decision-making.

Simplicity should not be a goal in itself – the ultimate test of a model is how well it can simulate the system of interest, usually assessed through validation. The relative performance of different model structures in validation depends to some extent on data availability – more detailed data permit more rigorous testing, potentially allowing identification of more complex model structures. However, in applications where the data available for calibration and validation are "sparse" (usually the case for catchment simulations, even in well-studied catchments), simpler models may be expected to perform just as well as complex ones.

The hypothesis driving this study is therefore that the most popular catchment-scale dynamic water quality models are unnecessarily complex. Testing this hypothesis will require a concerted effort from many modelling groups, with comparative studies of a range of different model structures in a wide variety of catchments. In this study, we begin this model evaluation process with the comparison of one popular and representative water quality model, INCA-P, and a newly-developed parsimonious catchment phosphorus model, SimplyP. The comparison is carried out in a rural Scottish catchment which contains the majority of land uses and P-related processes found in temperate rural regions, so results are likely to be applicable more widely. Our hypothesis is that INCA-P is overly-complex for the catchment, in which case we would expect similar model performance skill metrics from INCA-P and SimplyP.

2. Study area

Model applications were carried out in the Tarland Burn catchment, upstream of the James Hutton Institute monitoring point at Coull (51 km^2). The Tarland is a rural sub-catchment of the River Dee in northeast Scotland. Land use is a mixture of agriculture (primarily spring barley, improved grassland and rough grazing), upland heath and forestry. Humus iron podzols and brown forest soils tend to be associated with agricultural land, with peaty podzols under semi-natural land on the hills fringing the catchment. The main settlement is the village of Tarland, which has a small wastewater treatment works (around 600 people); septic tanks serve the remainder of the catchment (several hundred people). Water quality is of concern, primarily due to inputs of nutrients and sediments from agriculture. The Tarland Burn is the most upstream tributary of the ecologically-sensitive River Dee to have impaired water quality, and works have therefore been underway during the last decade to reduce sediment and nutrient inputs to the water course (Bergfur *et al.*, 2012). During the 2000-2010 period, the catchment had a mean annual rainfall of 966 mm, a mean annual runoff of 451 mm yr^{-1}, and median total dissolved P (TDP) and total P (TP) concentrations of 25 and 40 μgl^{-1}, respectively.

3. Modelling approach

3.1 INCA-P

INCA-P is a process-based, semi-distributed catchment model for simulating the daily transport of sediment, dissolved and particulate P from catchments to streams, as well as subsequent in-stream transport and processing. It is similar to other popular water quality models in terms of its structure and the number of processes and parameters it includes, and is therefore representative of the current dynamic phosphorus models in popular usage. During the last decade, INCA-P has been applied to a variety of catchments throughout Europe (e.g. Couture *et al.*, 2014, Martin-Ortega *et al.*, 2015, Wade *et al.*, 2007, Whitehead *et al.*, 2013), Canada (e.g. Crossman *et al.*, 2013) and India (Jin *et al.*, 2015), to explore how P dynamics may respond to changes in land management and climate. The model was originally developed in the early 2000s (Wade *et al.*, 2002) and has since undergone several phases of re-development. The most recent model description is given in Jackson-Blake *et al.* (2016), and only a brief description is given here. In this study, we used INCA-P version1.4.4.

INCA-P requires input time series of air temperature, precipitation, hydrologically effective rainfall (HER; the water from precipitation and snowmelt which contributes to runoff), and soil moisture deficit (SMD, the difference between the soil moisture content and field capacity). HER and SMD are derived from an external hydrology model.

Within INCA-P, water is delivered to the water course via three flow paths: quick flow, which accounts for overland flow and other rapid flow pathways, soil water flow and groundwater flow. Quick flow is generated through infiltration and saturation excess overland flow, soil water flow is generated by HER, and groundwater flow is derived from soil water flow via percolation. All flows transport dissolved P (as TDP), whilst sediment and particulate P (PP) are only transported via quick flow. Within the terrestrial compartment, P is present in the soil as solid labile or inactive soil P, or as dissolved P in soil water. Within the soil water and groundwater, TDP concentrations are controlled by sorption equilibria. Terrestrial P inputs include solid and liquid fertilizer, manure and atmospheric deposition; the major terrestrial sinks are plant uptake and adsorption. The rates of plant uptake, weathering and immobilization are dependent on temperature, whilst plant uptake is also dependent on soil moisture and season. Sediment delivery to the stream uses equations from INCA-sed (Jarritt & Lawrence, 2007, Lazar *et al.*, 2010), including process representations for splash detachment, flow erosion and the transport capacity of quick flow. PP is associated with sediment, and therefore affected by the same processes. Instream, the model includes effluent inputs, water abstractions, sediment settling and resuspension, bank erosion, P sorption reactions in the water column and in the stream bed and biological uptake of P from the water column and the stream bed. To simulate biological uptake, the model also simulates the dynamics of epiphyte and macrophyte biomass within stream reaches.

The rate of change in volume or mass of model state variables with respect to time is described by a series of ordinary differential equations (ODEs), solved within each time step using an adaptive 4^{th} order Runge-Kutte-Merson method.

3.2 SimplyP

A new simple hydrology, sediment and phosphorus model, SimplyP, was developed for comparison to INCA-P. A full description of SimplyP model structure, equations, numerical methods and priorities for future model development is given in the supplementary information (SI, Text 1), and only a brief description is provided below.

A full description of model aims and scope is given in SI Section 3.1. Briefly, the development of SimplyP was motivated by results of several studies evaluating INCA-P (Jackson-Blake *et al.*, 2015, Jackson-Blake & Starrfelt, 2015, Jackson-Blake *et al.*, 2016). These provided recommendations for model improvements, some of which were incorporated into INCA-P (version 1.4 onwards). However, one of the main recommendations was for model simplification. Whilst a certain amount of simplification can be achieved through parameterization, this requires a high level of familiarity with the model, is time-consuming and prone to errors, and there are limits to the amount of simplification that can be achieved. The aim here was therefore to carry out a more substantial process of simplification, whilst maintaining sufficient complexity for the model to be useful in hypothesis and scenario testing. SimplyP retains a number of similarities with INCA-P (Section 4.1). Other areas of the model were inspired by experiences in applying INCA-P, by assumptions used in other water quality models, or by well-established process understanding. A particular design aim was for the process representation to be simple enough to allow parameter values to be constrained using available data, by both keeping the number of parameters requiring calibration to a minimum, and aiming for as many parameters as possible to be in principle measurable.

Like INCA-P and other popular mechanistic water quality models, SimplyP is dynamic with a daily time step and is spatially semi-distributed, so the catchment may be split into sub-catchments and associated reaches. The model is run for each sub-catchment in turn, and outputs from each reach are fed into the main stem sequentially down-stream. Land may be further sub-divided based on simple 'land classes': for dissolved P processes, two classes are considered, a 'high P' and a 'low P' class. Land within a given class should have a similar gross annual P balance, soil P content and hydrological characteristics; in a typical rural catchment the high P class could include fertilized grassland and arable land, with everything else assigned to the low P class. For sediment and particulate P processes, the high P class may be sub-divided to account for differences in erodibility (e.g. improved grassland versus arable land). Finally, if arable land is present, the proportion of spring versus autumn-sown crops may be taken into account, along with the variation in soil erodibility through time. For convenience, there is also the possibility of a 'newly-converted' land class, to take into account legacy soil P when agricultural land becomes disused, or the lack of legacy soil P in new agricultural land (see SI Section 3.2).

A full description of SimplyP model processes is given in SI Section 4. In brief, the following sets of processes are included: snow accumulation and melt (SI Section 4.1.1); rainfall-runoff (SI Section 4.1.2); in-stream hydrology (SI Section 4.1.3); sediment delivery to the watercourse and in-stream transport (SI Section 4.2); and terrestrial and in-stream P processes (SI Section 4.3). A summary of the main stores and fluxes of water, sediment and P is provided in Figure 1. Three terrestrial flow paths are taken into account: (1) Quick flow, to simulate inputs to the watercourse during larger rainfall events and when soils are dry and little soil water flow occurs. This was simply calculated by assuming quick flow is proportional to incoming precipitation and is routed instantaneously to the stream; (2) Soil water flow, responsible for TDP leaching from soils and groundwater recharge. Inputs to the soil water are from rainfall and snowmelt; outputs are through evapotranspiration and soil water flow, and soil water flow is assumed to only occur once the soil water content is above field capacity; (3) Groundwater flow, important for controlling baseflow TDP concentrations. Groundwater recharge occurs through percolation from the soil water. In-stream reach volume and discharge are then estimated.

Sediment processes are represented in a highly simplified manner (see SI Section 4.2 for details). Briefly, in-stream suspended sediment (SS) concentration has long been known to be well-explained by a simple power law with in-stream discharge (Bagnold, 1966), as:

$$SS = E_{sus}\, Q_r^{k} \quad [1]$$

where Q_r is in-stream discharge and E_{sus} and k are constants (Colby, 1956). This simple power law is taken as the basis for predicting the combined sediment inputs to the stream reach from both the land phase and in-stream entrainment. The E_{sus} parameter is divided into a calibrated scaling factor and a number of factors which are known to affect terrestrial sediment delivery and in-stream entrainment rates, such as sub-catchment and channel slope and land cover. The land cover factor may be varied throughout the year to take into account periods of higher soil erodibility (Watson & Evans, 2007).

P is represented in the soil in three forms: TDP in the soil water, labile soil P and inactive soil P. The masses of dissolved and labile soil P change through time, whilst the mass of inactive soil P is constant. Labile soil P and TDP are assumed to be in equilibrium, related via a linear relationship (McCray *et al.*, 2005). A number of assumptions are made to help parameterize this

relationship, e.g. it is assumed that the inactive soil P content is the same in the high and low P classes, and that the low P class does not contain labile soil P and has soil water TDP concentrations around zero. The difference in total soil P content between the two classes is therefore all potentially labile P in the high P class, built up during fertilizer and manure additions (for a discussion of the issues relating to incorporating agronomic soil test P measurements into the model, see SI Section 4.3.1e). Initial soil water TDP concentration is calibrated within plausible ranges, and used to calculate a gradient for the labile P versus TDP relationship, which can then be used to relate soil labile P and TDP concentration outside the calibration period (see SI Section 4.3.1). Fertilizer, manure and plant uptake fluxes are grouped together into a single gross annual P balance parameter, which is then evenly applied or subtracted over the course of the year. This representation of soil P processes is highly simplified. In reality, soil P is present in a continuum of interlinked states of varying extractability, and hysteresis effects are common in P transfers between states. However, the understanding of how detailed soil chemical processes upscale to the catchment-scale is arguably not yet advanced enough to be usefully incorporated into a catchment-scale model.

Quick flow TDP concentration is assumed to be the same as soil water TDP concentration. Groundwater TDP concentration is set as a constant, given generally high P sorption capacities in and below mineral soil horizons. TDP is then transported to the reach via all three terrestrial flow pathways, whilst PP dynamics are linked to SS dynamics, taking into account enrichment of PP relative to parent material due to the selective transport of finer-grained, more P-rich material (Sharpley, 1980). Incidental P losses, potentially large P fluxes washed into watercourses when rainfall events coincide with fresh fertilizer and manure applications, are not yet included, due to difficulties in capturing the high spatial and temporal variability of these events and the relatively detailed management knowledge required.

In-stream, the model only includes the dilution of diffuse and point source P inputs and down-stream transport. This simple formulation assumes that in-stream processing is in a state of dynamic equilibrium, i.e. in-stream sinks and sources are balanced. Whilst there is insufficient data to suggest otherwise in the study catchment, it would be straightforward to add a simple retention or loss factor to the model for catchments where in-stream retention is thought to be important, informed, for example, by results of large-scale empirical studies (e.g. Alexander *et al.*, 2004).

The rate of change in volume or mass of model state variables with respect to time is described by a series of ODEs. To reduce errors introduced by numerical approximations, model ODEs were formulated as continuous functions, avoiding thresholds, and were solved within each time step using the LSODA solver (Hindmarsh, 1983) (see SI Section 4.5).

SimplyP requires input time series of daily precipitation, air temperature and potential evapotranspiration (PET). Model outputs include time series of daily fluxes and flow-weighted daily mean concentrations of TDP, PP and SS and daily mean flow. The state of the internal stores may also be output (e.g. snow depth, volumes and flows from the two water stores, and P masses in the different stores).

SimplyP v1.0 model code is open source and freely-available for download (https://github.com/LeahJB/SimplyP). See SI Section 2 for details and instructions.

3.3 Data for model calibration and testing

Data from the catchment outflow from 2000 to 2010 were used for model calibration and testing. Weekly chemistry sampling took place between 2000 and the end of 2003, with some daily sampling during rainfall events. Daily samples were collected between February 2004 and June 2005, providing 15 months of daily data. After June 2005, infrequent irregular sampling continued for the rest of the period. Measured parameters included SS, TDP and soluble reactive P (SRP) concentrations; daily 2004-2005 samples were also analyzed for total P (TP), allowing PP to be calculated (as TP – TDP). Daily discharge data is also available (daily means calculated from 15-minute data). For further details of monitoring and analytical methods, see Stutter *et al.* (2008).

A comprehensive set of additional data were compiled to help parameterize both models, including land use data, estimates of sewage effluent inputs from septic tanks and the sewage treatment works, soil solution and groundwater TDP concentrations, and soils data, including soil total P content and bulk density. See Jackson-Blake *et al.* (2015) for a full description of these data.

3.4 Model setups in the study catchment

The INCA-P application used in this study is described in Jackson-Blake *et al.* (2016). Briefly, the catchment was split into four reaches and associated sub-catchments (Figure 2). Four parameters were varied by reach: reach width and slope, initial bed sediment silt mass and effluent inputs. Three land classes were considered: arable, improved grassland and semi-natural, although when calculating fertilizer and manure inputs to agricultural land, area-weighted inputs from a finer-representation of the land use were used. Semi-natural land incorporated rough grazing, heather moorland, deciduous and coniferous woodland. For more details, including percent land use per sub-catchment, see Jackson-Blake *et al.* (2016).

For SimplyP, the catchment was considered as a single unit, rather than being separated into sub-catchments and reaches. Only one partially unconstrained SimplyP parameter may vary by reach, the effluent input, and for this first test of the model sewage effluent inputs from the sewage treatment works and septic tanks were summed and added at the top of the single simulated reach. The catchment area was then grouped similarly to the INCA-P setup: the high P class (50% of the catchment area) included arable land (20%) and improved grassland (30%). Upland heath, forestry and rough grazing were grouped into the low P class as semi-natural land (50%).

Input data to drive both models included air temperature and precipitation, derived from the UK Met Office 5 km gridded dataset, and PET, estimated using the FAO56 Penman Monteith method (Allen *et al.*, 1998). HER and SMD time series for INCA-P were calculated using a water balance model (described in Jackson-Blake *et al.*, 2015), which in turn required inputs of daily precipitation and PET. This water balance model has 8 parameters, the majority based on soil properties.

3.5 Model calibration, validation and scenario analysis

Ideally, a single auto-calibration procedure would be used to calibrate both INCA-P and SimplyP, rather than relying on manual calibration which could introduce modeler bias. However, the complexity of INCA-P means that full auto-calibration of all uncertain parameters

is not possible – the parameter space is too large to be explored within reasonable time scales (e.g. Jackson-Blake & Starrfelt, 2015), and the choice of which subset of parameters to include in the analysis is subjective. Auto-calibration often does not lead to improved model performance compared to manual calibration (Boyle *et al.*, 2000), particularly for dynamic variables such as phosphorus – previous auto-calibration of INCA-P in the study catchment resulted in less realistic simulated TDP dynamics than manual calibration (Jackson-Blake & Starrfelt, 2015). For this study, both models were therefore calibrated manually. The INCA-P calibration does however represent what we feel is the best possible setup for the Tarland catchment: much time has been spent calibrating INCA-P in the study catchment, initially through manual calibration and an investigation of different model structures (Jackson-Blake *et al.*, 2015), then through auto-calibration using a sophisticated MCMC algorithm (Jackson-Blake & Starrfelt, 2015), and finally through manual calibration informed by the results of previous calibrations (Jackson-Blake *et al.*, 2016). By contrast, the SimplyP calibration was done within a few hours, and therefore represents a first test of the model's potential.

The calibration period was 2004-2005, encompassing the 15 months when daily discharge and surface water chemistry data are available. Daily discharge and sparser water chemistry data for the period 2000-2010 (excluding 2004-2005) were then used for model validation.

Manual calibration was done in a step-wise manner: hydrology-related parameters were adjusted until an acceptable discharge calibration was obtained, then sediment-related parameters and finally P-related parameters, with several iterations. Additional data taken into account in calibration included one-off measurements of soil solution TDP concentrations from agricultural soils in the catchment. Model performance was evaluated using the procedure recommended in Jackson-Blake *et al.* (2015), using a combination of: (1) visual assessment of time series, (2) comparison of distributions of observed and simulated data using quantile-quantile (QQ) plots, and (3) model performance statistics, including Spearman's Rank correlation coefficient and model bias. Nash Sutcliffe efficiency (NS) and NS on logged data were used to assess the discharge simulation. NS was not used for water chemistry parameters as previous work has shown it to be poor at discriminating between realistic and unrealistic P simulations (Jackson-Blake *et al.*, 2015). In addition, the INCA-P calibration included checking for plausible changes in masses of soil P and stream bed sediment and PP during the model run.

For both models, the calibration procedure also involved adjusting the initial soil P store so that simulated agricultural soil water TDP concentration changed at an appropriate rate over longer model runs: long-term monitoring experiments suggest that, in the absence of fertilizer inputs, soil P in agricultural land should drop to near semi-natural values with a half-life of 7 to 9 years (McCollum, 1991, Syers *et al.*, 2008), i.e. within around 35 years in the study catchment. In INCA-P, this involved adjusting the soil depth parameter. In SimplyP, the soil areal mass parameter was adjusted (combining soil depth and bulk density; SI Table 7). Soil depths in the range 7 to 10 cm were obtained, 14 to 20 cm taking soil porosity into account. This is plausible, given that soil P decreases with depth and is highest in the top 20 cm in agricultural soils (Syers *et al.*, 2008). Calibrated parameter values for SimplyP are given in Table 1.

Several sensitivity tests were then carried out with SimplyP to test the model's ability to explore future scenarios of change. The model was run for a 30-year period with a number of reductions in net P inputs relative to the 2000 – 2010 baseline, corresponding to reductions in fertilizer and manure applications of 25%, 50% and 100%. Agricultural land in the catchment

has an estimated P balance (inputs from fertilizer and manure minus outputs via harvesting) of around 10 kg ha^{-1} yr^{-1} (Messiga *et al.*, 2010) and fertilizer and manure inputs are estimated at around 24 kg ha^{-1} yr^{-1}. The 25% and 50% reduction scenarios are therefore economically-feasible, corresponding to annual P balances of around 4 and -2 kg ha^{-1} yr^{-1}, which should have little impact on crop yields for a number of years. The 100% reduction scenario is included as a sensitivity test. Results were compared to previous results for the same set of scenarios derived using INCA-P (Jackson-Blake *et al.*, 2016).

4. Results

4.1 Comparison of model process-representation and complexity

SimplyP has a number of features in common with INCA-P, such as three terrestrial flow paths, a simple split of P into TDP and PP, and a split of soil P into labile and inactive stores. Many processes included in INCA-P have however been omitted from SimplyP, the most important P-related ones being: (1) the removal of seasonal variability in soil water TDP concentrations; (2) simplification of quick flow generation, by removing the process-representation of infiltration excess and saturation excess; (3) extensive simplification of the sediment-related equations, including removing the process-representation of splash detachment, flow erosion, the transport capacity of quick flow and in-stream entrainment and deposition; (4) controls on groundwater TDP concentration, which in SimplyP is considered to be constant; (5) simplification of the in-stream P processes, including removal of the macrophyte and epiphyte biomass equations and of the simulation of separate P processes in the water column and the stream bed.

Another difference is the incorporation of a hydrology model into SimplyP. We see this as a great improvement, removing the need for an external hydrology model and simplifying the calibration procedure, as water quality simulations are extremely dependent on the hydrology simulation, so time-consuming iterative calibration of separate hydrology and water quality models is often required for good model performance.

Model complexity is in part reflected by the number of state variables included in the model. For each sub-catchment/reach, SimplyP has 13 ODEs (Table SI-9), plus four for calculating daily in-stream fluxes from instantaneous fluxes (not present in INCA-P to our knowledge, but required to calculate volume-weighted daily mean concentrations). This is substantially fewer than INCA-P, despite the fact that SimplyP includes a hydrology model: INCA-P has 28 ODEs before land use variability is taken into account and 52 in an equivalent setup to SimplyP with 3 land use classes.

Model complexity is also in part reflected by the number of model parameters. SimplyP parameters are described in Table 1 and INCA-P parameters in Jackson-Blake *et al.* (2016). Both models require a number of well-constrained parameters which are generally not included in the calibration procedure (e.g. catchment area, areas of land classes, slopes and reach lengths; described in Table SI-10 for SimplyP). Excluding these, the total number of parameters in both models, split by process or type, is summarized in Table 2. SimplyP has 23 parameters that are to some extent unconstrained, 24 – 27 when spatial variability between land classes is taken into account. At least 8 of these are optional (before taking spatial variability into account), depending on the level of process-representation desired. By comparison, INCA-P has 146 parameters (assuming one reach and sub-catchment, varying three parameters by two land

403 classes for dissolved P processes and by three classes for soil erodibility for comparability with
404 SimplyP). The version of INCA-P described in the tests performed here (4 reaches with 3 land
405 classes) involves around 48 calibrating parameters, which we believe is simple compared to
406 previous applications (e.g. Crossman *et al.*, 2013, Jin *et al.*, 2015). A further ~45 parameters
407 were carefully assigned values to help simplify the model structure and turn certain processes
408 off, leaving around 53 unused parameters.

409 Only one SimplyP parameter varies by sub-catchment or reach, the effluent TDP input,
410 so model complexity will not increase substantially in larger systems. For INCA-P, 64
411 parameters may be varied by sub-catchment or reach, resulting in the potential for highly
412 parameterized model setups.

413 Another important issue is the extent to which model parameters may be informed by
414 data (Table 2). Around 43 INCA-P parameters are not measurable and must be determined
415 purely through calibration, and 14 are not measurable and yet are thought to exert a key
416 influence on model output (Jackson-Blake *et al.*, 2016). This can be problematic for model
417 calibration. Meanwhile, the majority of SimplyP parameters may be based on measured data or
418 data derived from literature reviews, and only 4 or 5 must be determined purely through
419 calibration (Table 1). One of these relates to the sediment simulation, the rest are hydrology
420 parameters. This is promising, as water quality models are particularly sensitive to hydrology
421 parameters (Dean *et al.*, 2009, Jackson-Blake & Starrfelt, 2015, van Griensven *et al.*, 2006), so
422 there is a good chance of these parameters being identifiable. Even in the most complex setup in
423 which all 27 SimplyP parameters are used, it should therefore be feasible to search the entire
424 parameter space as part of an auto-calibration or uncertainty analysis, provided the user has data
425 to inform the parameter values (Table 1) and sufficient discharge and water quality observations
426 (including SS, TDP and PP concentrations under the full range of flow conditions).

427 ## 4.2 Model performance in the Tarland Burn

428 ### *4.2.1 Model calibration and validation results*

429 During the calibration period, discharge performance statistics were slightly better for
430 SimplyP than for INCA-P (Table 3). This is in part because a snow accumulation and melt
431 routine was included in SimplyP and not in the simple model used to generate HER input for
432 INCA-P, although this only affects a few winter flow peaks (Figure 4). Small discharge peaks
433 were also often slightly better simulated by SimplyP, particularly during baseflow, which may
434 not have been the case had a more complex hydrology model been used to generate input for
435 INCA-P. The models performed similarly for SS, which is noteworthy given the dramatically
436 simpler process-representation in SimplyP. Slightly improved SS performance statistics for
437 SimplyP are likely to be due to the slightly better discharge simulation. The TP and PP
438 simulations are also similar: output from SimplyP is less biased than output from INCA-P and
439 the distribution of the simulated PP data in particular is closer to that of the observations (Figure
440 3). Both variables however have a lower correlation coefficient than INCA-P output. The story is
441 clearer for the TDP simulation, with SimplyP scoring higher in all model performance statistics
442 during the calibration period (Table 3), and also producing distributions of simulated data that
443 were more comparable to the observations (Figure 3). The slightly improved TDP dynamics may
444 be related to the better discharge simulation. However, an improvement in simulated TDP is
445 apparent even during flow events when SimplyP and INCA-P produced comparable discharge

simulations (Figure 4): SimplyP TDP peaks tend to be less broad and more responsive to hydrological inputs, something which could not be achieved with INCA-P even when the soil water time constants were reduced to below the values used in SimplyP. The reasons for this are not clear, but could relate to the formulation or solving of the ODEs in INCA-P (although this could not be checked as the code is closed source).

The story is similar during the validation period, when model performance statistics for SimplyP were slightly better than those for INCA-P for discharge and TDP (aside from a slightly larger bias in TDP), whilst SS had lower bias but a smaller correlation coefficient (Table 3). There has been a shift in baseflow discharge over time during the validation period (Figure 5), likely due to a change in channel cross section. Both models therefore over-estimate summer discharge during the first half of the period and under-estimate it later on. Although TDP performance statistics are slightly better for SimplyP, TDP is somewhat under-estimated during baseflow, something which is less of an issue in INCA-P during the validation period (Figure 5), and which leads to the slightly larger difference in distributions of observed and simulated data for SimplyP (Figure 3). This is likely due to differences in the discharge simulation, as SimplyP tends to simulate slightly higher discharge during baseflow, and even small differences in simulated baseflow discharge lead to large differences in simulated concentration.

Note that NS coefficients are only reported for water quality variables in Table 3 for comparability with other studies, and were not used when assessing model performance. Although NS coefficients might suggest simulations from both models are inadequate, various authors have pointed out problems with using NS as a measure of model performance (e.g. Jain & Sudheer, 2008, Schaefli & Gupta, 2007) and we have argued elsewhere that NS is particularly poor for measuring the performance of P models in agricultural catchments, where NS values above 0.2 are rare unless point sources dominate (Jackson-Blake *et al.*, 2015).

The results presented here constitute a small fraction of the behavior that could be compared between the two models. However, the indications are that, at least during calibration and testing within similar conditions to the calibration period, SimplyP appears to perform comparably for PP, SS and TP, and perhaps slightly outperform INCA-P in terms of discharge and TDP. The two models therefore appear to be as capable, at least in this catchment, of simulating daily concentrations and discharge. The broader question of whether either model performs well enough for model output to be useful is a valid one, the answer depending largely on what the output will be used for and how it is presented. This is discussed further in Jackson-Blake *et al.* (2015).

4.2.2 Scenario analysis

Results from the fertilizer and manure reduction scenarios are shown in Figure 6 in terms of the change in agricultural soil water EPC_0 (the equilibrium TDP concentration of zero sorption) and in-stream mean annual TDP concentration over the 30-year period. Results from the two models tell the same story: under the baseline and 25% reduction scenario P is still being added surplus to crop requirements, and so EPC_0 continues to rise, resulting in an increase over time in in-stream TDP concentration during rainfall events and a higher annual mean. Meanwhile, the 50% and 100% reduction scenarios result in net plant uptake of P from the soil, gradually depleting the labile P store and causing reductions in simulated EPC_0, soil water TDP inputs to the stream and therefore lower mean annual in-stream TDP concentration. However,

there is an important lag in the time for improvements to be seen, with only small decreases in in-stream TDP concentrations during the first 5 years of the simulation (less than a 15% reduction compared to the baseline even for the 100% reduction scenario; data not shown). The full benefits of the measures are only realized by the end of the 30-year period, the time taken for near full depletion of the labile soil P store.

Whilst terrestrial compartment results were similar for the two models, in-stream TDP results differed slightly, with a larger effect simulated by INCA-P. This may be because the in-stream TDP peaks simulated by INCA-P are slightly too broad (Section 4.2.1), resulting in over-estimation of the influence of agricultural inputs on mean in-stream TDP concentrations.

The similarity in results does not mean either model is right, but it does show that SimplyP is as capable of predicting the dynamics of legacy soil P as INCA-P. For both models to produce more robust output, more long term soil P data is needed to help constrain the parameterization as well as improved understanding of how soil P extractability changes at the catchment scale as soil P stores become depleted. Furthermore, a potentially important process currently missing from both models is the link between soil P content and crop uptake of P, as it is probably unrealistic to expect uptake to be unchanged as soil P stocks become depleted.

SimplyP has been run with just one kind of future scenario here – the effect of changing terrestrial P balances. However, like INCA, the model can be used to simulate a number of other broad-scale measures. Being a catchment-scale model, it is particularly suited to looking at the potential impacts of changes in land use, climate and effluent inputs. The effectiveness of measures aimed at reducing sediment inputs to the stream may also be simulated through changing the terrestrial erodibility parameters (informed for example by literature on the Universal Soil Loss Equation (Kinnell, 2010)) or through the use of a delivery reduction factor.

5. Discussion and conclusions

We set out to test the hypothesis that INCA-P is overly-complex when applied in a Scottish agricultural catchment. To do this, a new simple catchment phosphorus model was developed, SimplyP, and model structure and performance were compared to INCA-P. SimplyP is substantially more streamlined than INCA-P, with up to 28 parameters, whilst INCA-P has around 148. Only 4 or 5 SimplyP parameters are ‘free’, i.e. cannot be informed by observations, compared to around 45 for INCA-P. This reduction in complexity is despite the fact that SimplyP includes a rainfall-runoff module, whilst INCA-P relies on output from an external hydrology model (not included in this comparison of complexity). In the study catchment, both models performed similarly during calibration and validation. Both models also produced similar results in a scenario assessment, with identical implications for diffuse pollution mitigation and decision support. Results therefore support the hypothesis that INCA-P is overly-complex in the study catchment.

Although limited to just one study site, P dynamics in the study catchment are controlled by similar processes to those operating in the majority of temperate regions, with a mixture of land uses, hydrological flow paths and P inputs from both agriculture and sewage. This result is therefore likely to be transferable to other study areas. In addition, INCA-P is similar in complexity and structure to other popular catchment water quality models, so this conclusion is likely to apply to other models and water quality variables. Results are consistent with long-established theory and recent thinking in catchment science (e.g. Kirchner, 2006, Sivapalan,

2006), and provide further support for the idea that a more parsimonious approach to simulating catchment water quality is warranted.

Overall, there are strong arguments, backed up by the findings presented here, that the current generation of catchment-scale, dynamic water quality models are too complex. This complexity has likely been driven by two factors. Firstly, there has been a desire to include process-understanding and data derived from plot-scale studies. However, non-linearities between small-scale and catchment-scale processes mean up-scaling is often inappropriate (e.g. Kirchner, 2006, Oreskes & Belitz, 2001); catchment-scale responses are often simpler than might be anticipated from detailed process knowledge (Sivapalan, 2005). We therefore need a better understanding of catchment-scale behavior, requiring more comprehensive spatially-distributed data collection across catchments as well as high frequency monitoring of watercourses. Technological improvements in remote sensing and in-stream sensors are beginning to yield exciting new data, but more effort is needed to constrain soil water, groundwater and effluent chemistry, and to determine longer-term trends in soil and groundwater chemistry, especially in response to changes in e.g. land management and climate. Secondly, there has been a desire to make models versatile and widely applicable, i.e. to produce 'one-size-fits-all' models. This has helped us think about the variety of processes that could operate in different areas, but for any given study area is likely to result in overly complex models. Balancing the demands of model realism and parsimony remains a significant challenge, and resolving the tension between the two can only be achieved by assessing the performance of models with different structures (e.g. Fenicia *et al.*, 2006), preferably within statistical model comparison frameworks (e.g. Spiegelhalter *et al.*, 2002). For this, community-based modular model frameworks offer perhaps the best hope for the future (e.g. Mooij *et al.*, 2010, Robson, 2014).

The initial aim in developing SimplyP was a proof-of-concept that simple can be as good as complex. However, we believe that SimplyP also has the potential to fill an important gap, attempting to be both process-based and dynamic, maintaining a spatially semi-distributed setup, differentiating between soluble and particulate P phases, incorporating hydrology and a variety of flow paths, and yet having far fewer parameters than other popular water quality models. SimplyP also retains sufficient complexity to be used to investigate scenarios relevant for research, policy and land management. It was markedly quicker (and therefore cheaper) to set up and calibrate than INCA-P, and it should be feasible to include all parameters in an auto-calibration/uncertainty analysis procedure, and therefore in a formal model comparison framework. The fact that most parameters are physically-meaningful is also likely to help with generalization and transferability to other (perhaps more data-poor) areas (Sivapalan, 2005). At a more fundamental level, the reduction in the number of parameters should make validation exercises more effective for diagnosing structural problems with the model, as model behavior becomes less dependent on parameter tuning and more on model structure. This in turn means that the simpler model should be more useful for testing hypotheses about system behavior (Kirchner, 2006). The hope is therefore that SimplyP could provide a benchmark when choosing between different models, a building block for future model development, or, given its advantages over more complex models, be a useful tool in its own right (prototype code is freely available, see Section 3.2).

For SimplyP to become a robust tool in its own right, a number of further developments are recommended. The first priority is for more testing in a range of contrasting study sites, to

establish whether any of the extra processes available in more complex models are required in certain areas. Additional potential model improvements are summarized in SI Section 5 and Table SI 12. Many of these suggestions involve an increase in complexity, and would need to be justified by demonstrating improved model performance in validation.

Overall, we hope that the model development and simple comparison exercise presented here will help prompt wider model comparison and simplification, and more generally encourage debate amongst the water quality modelling community as to whether today's models are appropriate and fit for purpose.

Supplementary references

The description of SimplyP in the supplementary information cites many additional studies, reproduced here in the main text to ensure they are indexed, included in citation records and given appropriate credit (Bowes *et al.*, 2005, Chapra, 2008, Clark & Kavetski, 2010, Croke *et al.*, 2006, Dari *et al.*, 2015, Domagalski & Johnson, 2011, Fenicia *et al.*, 2011, Gan & Luo, 2013, Hindmarsh, 1983, Holman *et al.*, 2008, House, 2003, Jarvie *et al.*, 2013a, Jarvie *et al.*, 2013b, Jordan-Meille *et al.*, 2012, Kavetski & Clark, 2011, Kavetski *et al.*, 2006a, Kavetski *et al.*, 2006b, Kleinman *et al.*, 2011, Lefrançois *et al.*, 2007, Leopold & Maddock Jr, 1953, Luo *et al.*, 2012, Menzel, 1980, Merritt *et al.*, 2003, Neal & Jarvie, 2005, Oeurng *et al.*, 2010, Radcliffe & Cabrera, 2006, Ratliff *et al.*, 1983, Renard *et al.*, 1991, Sample, 2015, Sharpley *et al.*, 2013, Stollenwerk, 1996, Stutter *et al.*, 2010, Stutter *et al.*, 2009, Trimble, 2010, Twarakavi *et al.*, 2009, Wischmeier & Smith, 1965, Wischmeier & Smith, 1978, Wittenberg, 1999, Wolman *et al.*, 1964).

Acknowledgements and Data

Many thanks to Dmitri Kavetski for advice on formulating ODEs and to Marc Stutter and Andy Vinten for useful discussions which contributed to the design of SimplyP. This work was funded by the Rural and Environment Science and Analytical Services (RESAS) division of the Scottish Government and by the Nordic Research Council-Nordforsk project #74306, e-Infrastructure for river-basin modelling. Data to reproduce the conclusions reported here are available for download, including parameter values, input data and observed data (https://github.com/LeahJB/SimplyP/tree/Hydrology_Model/Tarland_Data_WRR2016). SimplyP parameter values for scenario analysis are described in Section 3.5. SimplyP model code may be downloaded from https://github.com/LeahJB/SimplyP, see SI Section 2 for more details. To obtain an INCA-P executable, email rmc@niva.no.

References

Alexander R, Smith R, Schwarz G (2004) Estimates of diffuse phosphorus sources in surface waters of the United States using a spatially referenced watershed model. *Water Science and Technology,* **49**, 1-10.

Allen R, Pereira L, Raes D, Smith M (1998) *Crop evapotranspiration – guidelines for computing crop water requirements. FAO irrigation and drainage paper 56*.

Arnold JG, Fohrer N (2005) SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes,* **19**, 563-572.

618 Arnold JG, Srinivasan R, Muttiah RS, Williams JR (1998) Large area hydrologic modeling and
619 assessment part I: Model development1. *JAWRA Journal of the American Water*
620 *Resources Association,* **34**, 73-89.
621 Asselman NEM (2000) Fitting and interpretation of sediment rating curves. *Journal of*
622 *Hydrology,* **234**, 228-248.
623 Bagnold R (1966) An approach to the sediment transport problem. *General Physics Geological*
624 *Survey, Prof. paper*.
625 Beck HE, Dijk AI, Miralles DG, Jeu RA, Mcvicar TR, Schellekens J (2013) Global patterns in
626 base flow index and recession based on streamflow observations from 3394 catchments.
627 *Water Resources Research,* **49**, 7843-7863.
628 Bergfur J, Demars BOL, Stutter MI, Langan SJ, Friberg N (2012) The Tarland Catchment
629 Initiative and Its Effect on Stream Water Quality and Macroinvertebrate Indices. *J.*
630 *Environ. Qual.,* **41**, 314-321.
631 Beven K, Binley A (1992) The future of distributed models: Model calibration and uncertainty
632 prediction. *Hydrological Processes,* **6**, 279-298.
633 Bicknell BR, Imhoff JC, Kittle Jr JL, Jobes T, Donigian Jr A, Johanson R (2001) Hydrological
634 simulation program-Fortran: HSPF version 12 user's manual. *AQUA TERRA*
635 *Consultants, Mountain View, California*.
636 Binger R, Theurer F (2005) AnnAGNPS Technical Processes documentation, version 3.2. In:
637 *USDA-ARS, National Sedimentation Laboratory.* pp Page.
638 Bowes MJ, House WA, Hodgkinson RA, Leach DV (2005) Phosphorus–discharge hysteresis
639 during storm events along a river catchment: the River Swale, UK. *Water Research,* **39**,
640 751-762.
641 Boyle DP, Gupta HV, Sorooshian S (2000) Toward improved calibration of hydrologic models:
642 Combining the strengths of manual and automatic methods. *Water Resources Research,*
643 **36**, 3663-3674.
644 Chapra SC (2008) *Surface water-quality modeling*, Waveland press.
645 Clark MP, Kavetski D (2010) Ancient numerical daemons of conceptual hydrological modeling:
646 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research,* **46**,
647 W10510.
648 Colby B (1956) Relationship of sediment discharge to streamflow. pp Page, US Dept. of the
649 Interior, Geological Survey, Water Resources Division.
650 Couture R-M, Tominaga K, Starrfelt J, Moe SJ, Kaste Ø, Wright RF (2014) Modelling
651 phosphorus loading and algal blooms in a Nordic agricultural catchment-lake system
652 under changing land-use and climate. *Environmental Science: Processes & Impacts,* **16**,
653 1588-1599.
654 Croke BF, Andrews F, Jakeman AJ, Cuddy SM, Luddy A (2006) IHACRES Classic Plus: a
655 redesign of the IHACRES rainfall-runoff model. *Environmental Modelling & Software,*
656 **21**, 426-427.
657 Crossman J, Futter MN, Oni SK *et al.* (2013) Impacts of climate change on hydrology and water
658 quality: Future proofing management strategies in the Lake Simcoe watershed, Canada.
659 *Journal of Great Lakes Research,* **39**, 19-32.
660 Dari B, Nair V, Colee J, Harris W, Mylavarapu R (2015) Estimation of Isotherm Parameters: A
661 Simple and Cost-effective Procedure. *Frontiers in Environmental Science,* **3**.

662 Dean S, Freer J, Beven K, Wade AJ, Butterfield D (2009) Uncertainty assessment of a process-
663 based integrated catchment model of phosphorus. *Stochastic environmental research and*
664 *risk assessment,* **23**, 991-1010.
665 Domagalski JL, Johnson HM (2011) Subsurface transport of orthophosphate in five agricultural
666 watersheds, USA. *Journal of Hydrology,* **409**, 157-171.
667 Donigian Jr A, Bicknell B, Imhoff J, Singh V (1995) Hydrological Simulation Program-Fortran
668 (HSPF). *Computer models of watershed hydrology.*, 395-442.
669 Eurostat (2013) Agri-environmental indicator fact sheet - risk of pollution by phosphorus. In:
670 *European Union (EU) agri-environmental indicator fact sheets.* pp Page.
671 Fenicia F, Kavetski D, Savenije HHG (2011) Elements of a flexible approach for conceptual
672 hydrological modeling: 1. Motivation and theoretical development. *Water Resources*
673 *Research,* **47**, n/a-n/a.
674 Fenicia F, Savenije HHG, Matgen P, Pfister L (2006) Is the groundwater reservoir linear?
675 Learning from data in hydrological modelling. *Hydrol. Earth Syst. Sci.,* **10**, 139-150.
676 Gan R, Luo Y (2013) Using the nonlinear aquifer storage–discharge relationship to simulate the
677 base flow of glacier- and snowmelt-dominated basins in northwest China. *Hydrol. Earth*
678 *Syst. Sci.,* **17**, 3577-3586.
679 Hahn C, Prasuhn V, Stamm C, Lazzarotto P, Evangelou M, Schulin R (2013) Prediction of
680 dissolved reactive phosphorus losses from small agricultural catchments: calibration and
681 validation of a parsimonious model. *Hydrology and Earth System Sciences,* **17**, 3679.
682 Hindmarsh AC (1983) ODEPACK, A Systematized Collection of ODE Solvers, RS Stepleman et
683 al.(eds.), North-Holland, Amsterdam,(vol. 1 of), pp. 55-64. *IMACS transactions on*
684 *scientific computation,* **1**, 55-64.
685 Holman IP, Whelan MJ, Howden NJK, Bellamy PH, Willby NJ, Rivas-Casado M, Mcconvey P
686 (2008) Phosphorus in groundwater—an overlooked contributor to eutrophication?
687 *Hydrological Processes,* **22**, 5121-5127.
688 House WA (2003) Geochemical cycling of phosphorus in rivers. *Applied Geochemistry,* **18**, 739-
689 748.
690 Jackson-Blake LA, Dunn SM, Helliwell RC, Skeffington RA, Stutter MI, Wade AJ (2015) How
691 well can we model stream phosphorus concentrations in agricultural catchments?
692 *Environmental Modelling & Software,* **64**, 31-46.
693 Jackson-Blake LA, Starrfelt J (2015) Do higher data frequency and Bayesian auto-calibration
694 lead to better model calibration? Insights from an application of INCA-P, a process-based
695 river phosphorus model. *Journal of Hydrology,* **527**, 641-655.
696 Jackson-Blake LA, Wade AJ, Futter MN *et al.* (2016) The INtegrated CAtchment model of
697 Phosphorus dynamics (INCA-P): description and demonstration of new model structure
698 and equations. *Environmental Modelling & Software,* **83**, 356-386.
699 Jain SK, Sudheer K (2008) Fitting of hydrologic models: a close look at the Nash–Sutcliffe
700 index. *Journal of hydrologic engineering,* **13**, 981-986.
701 Jakeman AJ, Hornberger GM (1993) How much complexity is warranted in a rainfall-runoff
702 model? *Water Resources Research,* **29**, 2637-2649.
703 Jakeman AJ, Letcher RA (2003) Integrated assessment and modelling: features, principles and
704 examples for catchment management. *Environmental Modelling & Software,* **18**, 491-
705 501.
706 Jakeman AJ, Letcher RA, Norton JP (2006) Ten iterative steps in development and evaluation of
707 environmental models. *Environmental Modelling & Software,* **21**, 602-614.

Jarritt NP, Lawrence DSL (2007) Fine sediment delivery and transfer in lowland catchments: modelling suspended sediment concentrations in response to hydrological forcing. *Hydrological Processes,* **21**, 2729-2744.

Jarvie HP, Sharpley AN, Spears B, Buda AR, May L, Kleinman PJA (2013a) Water Quality Remediation Faces Unprecedented Challenges from "Legacy Phosphorus". *Environmental Science & Technology,* **47**, 8997-8998.

Jarvie HP, Sharpley AN, Withers PJA, Scott JT, Haggard BE, Neal C (2013b) Phosphorus Mitigation to Control River Eutrophication: Murky Waters, Inconvenient Truths, and "Postnormal" Science. *J. Environ. Qual.,* **42**, 295-304.

Jin L, Whitehead PG, Sarkar S *et al.* (2015) Assessing the impacts of climate change and socio-economic changes on flow and phosphorus flux in the Ganga river system. *Environmental Science: Processes & Impacts,* **17**, 1098-1110.

Jordan-Meille L, Rubæk GH, Ehlert PaI *et al.* (2012) An overview of fertilizer-P recommendations in Europe: soil testing, calibration and fertilizer recommendations. *Soil Use and Management,* **28**, 419-435.

Kavetski D, Clark MP (2011) Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes,* **25**, 661-670.

Kavetski D, Kuczera G, Franks SW (2006a) Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology,* **320**, 173-186.

Kavetski D, Kuczera G, Franks SW (2006b) Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis. *Journal of Hydrology,* **320**, 187-201.

Kinnell PIA (2010) Event soil loss, runoff and the Universal Soil Loss Equation family of models: A review. *Journal of Hydrology,* **385**, 384-397.

Kirchner JW (2006) Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research,* **42**, W03S04.

Kleinman P, Sharpley A, Buda A, Mcdowell R, Allen A (2011) Soil controls of phosphorus in runoff: Management barriers and opportunities. *Canadian Journal of Soil Science,* **91**, 329-338.

Krueger T, Freer J, Quinton JN, Macleod CJA (2007) Processes affecting transfer of sediment and colloids, with associated phosphorus, from intensively farmed grasslands: a critical note on modelling of phosphorus transfers. *Hydrological Processes,* **21**, 557-562.

Lazar AN, Butterfield D, Futter MN *et al.* (2010) An assessment of the fine sediment dynamics in an upland river system: INCA-Sed modifications and implications for fisheries. *Science of the Total Environment,* **408**, 2555-2566.

Lefrançois J, Grimaldi C, Gascuel-Odoux C, Gilliet N (2007) Suspended sediment and discharge relationships to identify bank degradation as a main sediment source on small agricultural catchments. *Hydrological Processes,* **21**, 2923-2933.

Leopold LB, Maddock Jr T (1953) The hydraulic geometry of stream channels and some physiographic implications. pp Page.

Lindström G, Pers C, Rosberg J, Strömqvist J, Arheimer B (2010) Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology research,* **41**, 295-319.

Luo Y, Arnold J, Allen P, Chen X (2012) Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China. *Hydrol. Earth Syst. Sci.,* **16**, 1259-1267.

754 Mackay DJ (2003) *Information theory, inference and learning algorithms*, Cambridge university
755 press.
756 Martin-Ortega J, Perni A, Jackson-Blake L *et al.* (2015) A transdisciplinary approach to the
757 economic analysis of the European Water Framework Directive. *Ecological Economics,*
758 **116**, 34-45.
759 Mccollum RE (1991) Buildup and Decline in Soil Phosphorus: 30-Year Trends on a Typic
760 Umprabuult. *Agronomy Journal,* **83**, 77-85.
761 Mccray JE, Kirkland SL, Siegrist RL, Thyne GD (2005) Model Parameters for Simulating Fate
762 and Transport of On-Site Wastewater Nutrients. *Ground Water,* **43**, 628-639.
763 Menzel R (1980) Enrichment ratios for water quality modeling. *CREAMS: A Field-Scale Model*
764 *for Chemicals, Runoff, and Erosion from Agricultural Management Systems*
765 *Conservation Research Report Number 26, May, 1980. p 486-492, 1 Fig, 2 Tab, 11 Ref.*
766 Merritt WS, Letcher RA, Jakeman AJ (2003) A review of erosion and sediment transport
767 models. *Environmental Modelling & Software,* **18**, 761-799.
768 Messiga AJ, Ziadi N, Plénet D, Parent LE, Morel C (2010) Long-term changes in soil
769 phosphorus status related to P budgets under maize monoculture and mineral P
770 fertilization. *Soil Use and Management,* **26**, 354-364.
771 Mooij WM, Trolle D, Jeppesen E *et al.* (2010) Challenges and opportunities for integrating lake
772 ecosystem modelling approaches. *Aquatic Ecology,* **44**, 633-667.
773 Neal C, Jarvie HP (2005) Agriculture, community, river eutrophication and the Water
774 Framework Directive. *Hydrological Processes,* **19**, 1895-1901.
775 Oeurng C, Sauvage S, Sánchez-Pérez J-M (2010) Dynamics of suspended sediment transport and
776 yield in a large agricultural catchment, southwest France. *Earth Surface Processes and*
777 *Landforms,* **35**, 1289-1301.
778 Oreskes N, Belitz K (2001) Philosophical issues in model assessment. *Model validation:*
779 *Perspectives in hydrological science,* **23**.
780 Panagos P, Borrelli P, Meusburger K, Alewell C, Lugato E, Montanarella L (2015) Estimating
781 the soil erosion cover-management factor at the European scale. *Land Use Policy,* **48**, 38-
782 50.
783 Pappenberger F, Beven K, Frodsham K, Romanowicz R, Matgen P (2007) Grasping the
784 unavoidable subjectivity in calibration of flood inundation models: A vulnerability
785 weighted approach. *Journal of Hydrology,* **333**, 275-287.
786 Perrin C, Michel C, Andréassian V (2001) Does a large number of parameters enhance model
787 performance? Comparative assessment of common catchment model structures on 429
788 catchments. *Journal of Hydrology,* **242**, 275-301.
789 Radcliffe DE, Cabrera ML (2006) *Modeling phosphorus in the environment*, CRC Press.
790 Radcliffe DE, Freer J, Schoumans O (2009) Diffuse Phosphorus Models in the United States and
791 Europe: Their Usages, Scales, and Uncertainties. *J. Environ. Qual.,* **38**, 1956-1967.
792 Ratliff LF, Ritchie JT, Cassel DK (1983) Field-Measured Limits of Soil Water Availability as
793 Related to Laboratory-Measured Properties1. *Soil Science Society of America Journal,*
794 **47**.
795 Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009)
796 Structural and practical identifiability analysis of partially observed dynamical models by
797 exploiting the profile likelihood. *Bioinformatics,* **25**, 1923-1929.
798 Renard KG, Foster GR, Weesies GA, Porter JP (1991) RUSLE: Revised universal soil loss
799 equation. *Journal of soil and Water Conservation,* **46**, 30-33.

800 Robson BJ (2014) When do aquatic systems models provide useful predictions, what is
801 changing, and what is next? *Environmental Modelling & Software,* **61**, 287-296.
802 Sample J (2015) Statistics notes for environmental modelling. pp Page, GitHub.
803 Schaefli B, Gupta HV (2007) Do Nash values have value? *Hydrological Processes,* **21**, 2075-
804 2080.
805 Seibert J (2003) Reliability of Model Predictions Outside Calibration Conditions. *Paper*
806 *presented at the Nordic Hydrological Conference (Røros, Norway 4-7 August 2002),* **34**,
807 477-492.
808 Sharpley A, Jarvie HP, Buda A, May L, Spears B, Kleinman P (2013) Phosphorus Legacy:
809 Overcoming the Effects of Past Management Practices to Mitigate Future Water Quality
810 Impairment. *J. Environ. Qual.,* **42**, 1308-1326.
811 Sharpley AN (1980) The Enrichment of Soil Phosphorus in Runoff Sediments1. *J. Environ.*
812 *Qual.,* **9**, 521-526.
813 Sivapalan M (2005) Pattern, process and function: elements of a unified theory of hydrology at
814 the catchment scale. *Encyclopedia of hydrological sciences*.
815 Sivapalan M (2006) *Predictions in ungauged basins: promise and progress*, International Assn
816 of Hydrological Sciences.
817 Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model
818 complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical*
819 *Methodology),* **64**, 583-639.
820 Stollenwerk KG (1996) Simulation of phosphate transport in sewage-contaminated groundwater,
821 Cape Cod, Massachusetts. *Applied Geochemistry,* **11**, 317-324.
822 Stutter MI, Demars BOL, Langan SJ (2010) River phosphorus cycling: Separating biotic and
823 abiotic uptake during short-term changes in sewage effluent loading. *Water Research,* **44**,
824 4425-4436.
825 Stutter MI, Langan SJ, Cooper RJ (2008) Spatial contributions of diffuse inputs and within-
826 channel processes to the form of stream water phosphorus over storm events. *Journal of*
827 *Hydrology,* **350**, 203-214.
828 Stutter MI, Langan SJ, Lumsdon DG, Clark LM (2009) Multi-element signatures of stream
829 sediments and sources under moderate to low flow conditions. *Applied Geochemistry,* **24**,
830 800-809.
831 Syers JK, Johnston AE, Curtin D (2008) Efficiency of soil and fertilizer phosphorus use.
832 Reconciling changing concepts of soil phosphorus behaviour with agronomic
833 information. In: *Fertilizer and plant nutrition bulletins.* pp Page, Rome, Food and
834 Agriculture Organisation of the United Nations (FAO).
835 Trimble SW (2010) Streams, valleys and floodplains in the sediment cascade. *Sediment*
836 *Cascades: An Integrated Approach*, 307-343.
837 Twarakavi NKC, Sakai M, Šimůnek J (2009) An objective analysis of the dynamic nature of
838 field capacity. *Water Resources Research,* **45**, n/a-n/a.
839 Usda (2004) National Enginnering Handbook, Part 630 - Hydrology, Chapter 11 (Snowmelt). pp
840 Page, United States Department of Agriculture, Natural Resources Conservation Service.
841 Van Dijk A (2010) Climate and terrain factors explaining streamflow response and recession in
842 Australian catchments. *Hydrology and Earth System Sciences,* **14**, 159-169.
843 Van Griensven A, Meixner T, Grunwald S, Bishop T, Diluzio M, Srinivasan R (2006) A global
844 sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal*
845 *of Hydrology,* **324**, 10-23.

Van Meter KJ, Basu NB (2015) Catchment legacies and time lags: a parsimonious watershed model to predict the effects of legacy storage on nitrogen export. *PloS one,* **10**, e0125971.

Wade AJ, Butterfield D, Griffiths T, Whitehead PG (2007) Eutrophication control in river-systems: an application of INCA-P to the River Lugg. *Hydrol. Earth Syst. Sci.,* **11**, 584-600.

Wade AJ, Durand P, Beaujouan V *et al.* (1999) A nitrogen model for European catchments: INCA, new model structure and equations. *Hydrol. Earth Syst. Sci.,* **6**, 559-582.

Wade AJ, Whitehead PG, Butterfield D (2002) 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.

Watson A, Evans R (2007) Water erosion of arable fields in North-East Scotland, 1985 – 2007. *Scottish Geographical Journal,* **123**, 107-121.

Wellen C, Kamran-Disfani A-R, Arhonditsis GB (2015) Evaluation of the Current State of Distributed Watershed Nutrient Water Quality Modeling. *Environmental Science & Technology,* **49**, 3278-3290.

Whitehead PG, Crossman J, Balana BB *et al.* (2013) A cost-effectiveness analysis of water security and water quality: impacts of climate and land-use change on the River Thames system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* **371**.

Whitehead PG, Wilson EJ, Butterfield D (1998) A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I — model structure and process equations. *Science of the Total Environment,* **210–211**, 547-558.

Wischmeier WC, Smith DD (1965) *Predicting rainfall-erosion losses from cropland east of the Rocky Mountains.,* Washington DC, US Department of Agriculture (USDA).

Wischmeier WC, Smith DD (1978) *Predicting rainfall erosion losses - a guide to conservation planning,* Washington DC, US Department of Agriculture (USDA).

Wittenberg H (1999) Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes,* **13**, 715-726.

Wolman M, Miller J, Leopold L (1964) Fluvial processes in geomorphology. *San Francisco*.

Young RA, Onstad C, Bosch D, Anderson W (1989) AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. *Journal of soil and water conservation,* **44**, 168-173.

Table 1: SimplyP model parameters, including default values, recommended ranges and possible data sources. 'Spatial' column describes whether the parameter varies spatially by land use (LU), in which case by which LU type (A: agricultural, S: semi-natural, Ar: arable, IG: improved grassland), or sub-catchment/reach (SC/R). Parameters likely to be key in most settings are marked with an asterisk. Many of those without an asterisk are optional. Q is discharge.

Type	Param	Units	Description	Spatial	Tarland	Default	Min	Max	Data sources
Snow	$D_{snow,0}$	mm	Initial snow depth	–	0	0	0	10000	Meteorological records
	f_{DDSM}	mm dd°C^{-1}	Degree-day factor for snow melt	–	2.74	2.74	1.6	6	Literature, e.g. USDA (2004)
Hydrology	*T_s	days	Soil water time constant	LU (A, S)	A: 2 S: 10	A: 1 S: 10	> 0	30	*Calibration*
	f_{quick}	none	Proportion of precipitation routed to quick flow	–	0.02	0.02	0	0.2	*Calibration*
	alpha	none	PET reduction factor	–	1	1	0.4	1.2	Literature, e.g. Allen *et al.* (1998)
	*FC	mm	Soil field capacity	–	290	300	100	400	Soils database, or from soil texture using conversion charts (e.g. Appendix, Figure A1)
	*beta	none	Baseflow index	–	0.70	0.60	0	1	Local or global databases (e.g. Beck *et al.*, 2013)
	*T_g	days	Baseflow recession constant	–	65	65	> 0	100	May be estimated from Q data using methods of Van Dijk (2010); see Beck *et al.* (2013) for a global analysis
	$Q_{g,min}$	mm d^{-1}	Minimum groundwater flow	–	0.4	0.0	0	2	*Calibration*
	a	m^{-2}	Gradient of stream velocity-Q relationship	–	0.5	0.5	0.1	0.8	Empirically-derived from paired velocity and Q measurements (e.g. from flow gauging)
	$Q_{r0\ init}$	m^3 s^{-1}	Initial in-stream Q	–	1.0	1.0	> 0	N/A	Q observations
Sediment	C_{cover}	None	Vegetation cover factor (ratio of erosion rates under the land class vs bare soil)	LU (Ar, IG, S)	A: 0.2 S: 0.021 IG: 0.09	A: 0.2 S: 0.021 IG: 0.09	0	1	(R)USLE literature, e.g. Panagos *et al.* (2015)
	*E_M	kg mm^{-1}	Sediment input scaling factor	–	1500	1500	0	5000	*Calibration*
	*k_M	none	Sediment input non-linear coefficient	–	2.0	2.0	1.2	3	Empirical relationship between Q and SS observations or literature (e.g. Asselman, 2000)
	$d_{maxE,spr}$	none	Julian day with max erodibility; spring-sown crops	–	60	60	1	365	Local agronomic practices
	$d_{maxE,aut}$	none	Julian day with max erodibility, autumn-sown crops	–	304	304	1	365	Local agronomic practices
Dissolved P	*$P_{soilConc}$	mg kg^{-1}	Initial total soil P content	LU (A, S)	A: 1458 S: 873	A: 1458 S: 873	0-400	>3000	Soils database. Estimate from soil test P data using an empirical relationship
	*$P_{netInput}$	kg ha^{-1} yr^{-1}	Net annual P input to the soil (negative if uptake > input); S fixed at 0	LU (A)	10	10	-30	30	Fertilizer and manure application surveys, literature for P uptake, national P balance inventories (e.g. eurostat, 2013, for EU countries)
	*$EPC_{0,init}$	mg l^{-1}	Initial soil water TDP concentration on agricultural land	LU (A)	0.1	0.1	0	2	Direct measurements, literature
	*$M_{soil,m2}$	kg m^{-2}	Soil mass per m^2, important in determining the initial soil labile P mass	–	95	100	>0	800	Soils data (bulk density and depth)

Type	Param	Units	Description	Spatial	Tarland	Default	Min	Max	Data sources
	*TDP_{eff}	kg day^{-1}	Reach effluent TDP inputs	SC/R	0.1	0	0	N/A	Water company/environment protection agency data
	*TDP_g	mg l^{-1}	Groundwater TDP concentration	–	0.02	0	0	2	Direct measurements or literature
PP	*E_{PP}	none	PP enrichment factor	–	1.6	1	1	6	Direct measurements or literature (e.g. Sharpley, 1980)

885 Table 2: Comparison of numbers of model parameters required for SimplyP and INCA-P.

Category	INCA-P (excluding hydrology model parameters)	SimplyP
Total, no spatial variability	138	23
Total, with spatial variability (land class)	146	24 – 27
Parameters that vary by sub-catchment and/or reach	64	1
Not measurable (purely calibrated)	43	4 or 5
Additional parameters held constant over land use/reach/sub-catchment	12	4
Parameters used in the study catchment	48 (additional 45 to simplify the setup by removing processes)	22

886

887 Table 3: Model performance statistics in the calibration and validation period.

Variable	Model	Calibration					Validation				
		n[a]	Bias (%)	SR[b]	NS[c]	NS (logs)[c]	n[a]	Bias (%)	SR[b]	NS[c]	NS (logs)[c]
Q	SimplyP	716	0	0.92	0.80	0.81	3213	12	0.87	0.73	0.72
	INCA-P		0	0.91	0.73	0.79		12	0.87	0.55	0.72
SS	SimplyP	448	-6	0.54	0.13	0.10	189	-27	0.23	0.13	-0.10
	INCA-P		-16	0.46	0.02	0.34		-43	0.31	0.05	0.33
TP	SimplyP	428	0	0.25	0.16	-0.01	0				
	INCA-P		-14	0.37	0.07	0.13					
PP	SimplyP	428	-3	0.19	0.10	-0.27	0				
	INCA-P		-43	0.28	0.01	-0.06					
TDP	SimplyP	449	0	0.41	0.12	0.05	105	-11	0.54	0.15	0.22
	INCA-P		9	0.34	-0.24	-0.14		7	0.44	0.10	0.17

888 [a]Number of observations, [b]Spearman's Rank correlation coefficient, [c]Nash Sutcliffe efficiency of
889 untransformed or logged data. NS are only provided for water quality parameters for
890 comparability with other studies.
891

Figure 1: Schematic of the main stores, processes and pathways included in the model. White boxes show the state variables whose volume (water) or mass (sediment, P species) is tracked. Variables within small grey boxes are implicitly included in the model, but are not tracked. Arrows show fluxes within and between compartments. P: phosphorus, SS: suspended sediment, TDP: total dissolved P, PP: particulate P, ET: evapotranspiration.

Figure 2: The Tarland catchment, with simplified land use, the location of effluent inputs and monitoring points, and the sub-catchments used in the INCA-P application. Eastings and northings (km) are relative to the British National Grid.

Figure 3: Q-Q plots for the calibration and validation periods. Quantiles of the simulated data are plotted against corresponding quantiles of the observed data; if observed and simulated data are from similar distributions, points will lie close to the 1:1 line. Median and interquartile ranges (IQR) are shown. Units are mg l^{-1} for suspended sediment (SS) and µg l^{-1} for all P species. Note log scales.

Figure 4: Time series of observed and simulated discharge (Q) and water quality during the calibration period. Note the log scales for SS, PP and TP.

Figure 5: Time series of observed and simulated discharge (Q) and water quality during the validation period. Note the log scales for SS, PP and TP.

Figure 6: Simulated changes over a 30-year period in (a) agricultural soil water EPC_0 (the equilibrium phosphorus concentration, closely linked to soil water TDP concentration), and (b) in-stream mean TDP concentration, comparing the first and last 5 years of the study period. Results are shown for baseline 'business as usual' inputs and for three fertilizer and manure reduction scenarios (reductions of 25%, 50% and 100%).

Figure 1.

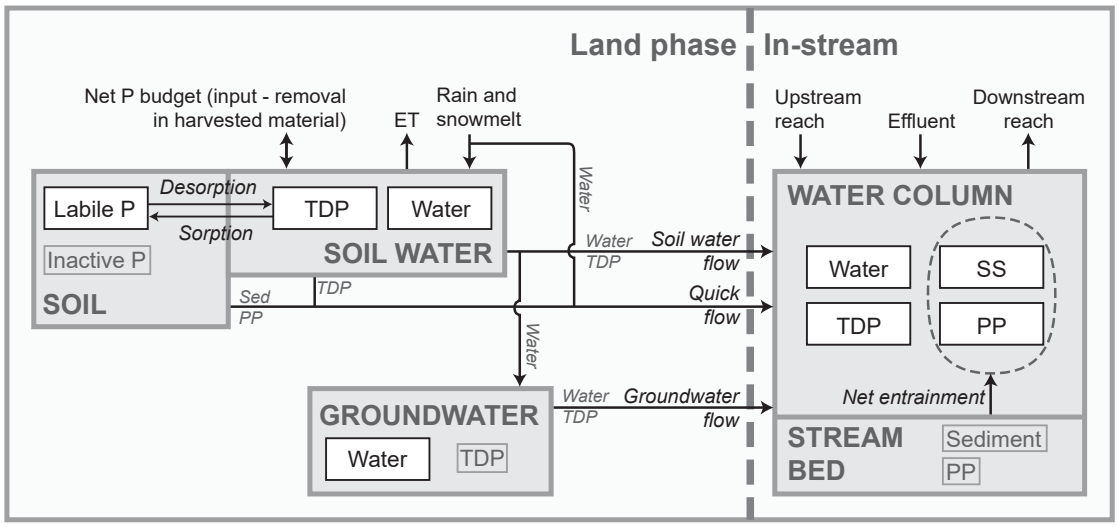

Land phase
In-stream
Net P budget (input - removal in harvested material)
ET
Rain and snowmelt
Upstream reach
Effluent
Downstream reach
Labile P
Desorption
Sorption
TDP
Water
SOIL WATER
Inactive P
SOIL
Water
TDP
Soil water flow
Sed
PP
TDP
Quick flow
Water
GROUNDWATER
Water
TDP
Water
TDP
Groundwater flow
WATER COLUMN
Water
SS
TDP
PP
Net entrainment
STREAM BED
Sediment
PP

Figure 2.

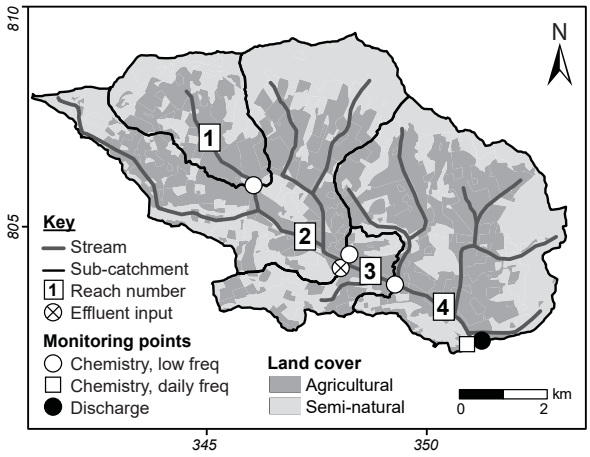

810
805
345
350
N
1
2
3
4
Key
Stream
Sub-catchment
1 Reach number
Effluent input
Monitoring points
Chemistry, low freq
Chemistry, daily freq
Discharge
Land cover
Agricultural
Semi-natural
km
0
2

Figure 3.

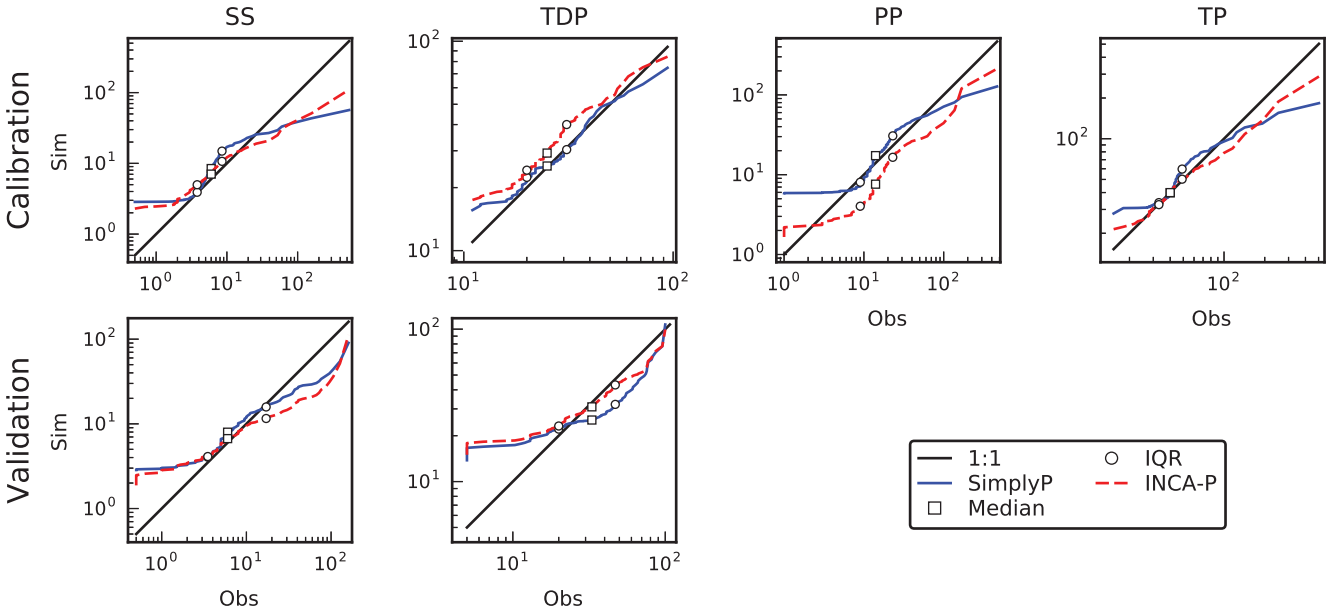

SS
TDP
PP
TP
Calibration
Validation
Sim
Obs
1:1
SimplyP
Median
IQR
INCA-P

Figure 4.

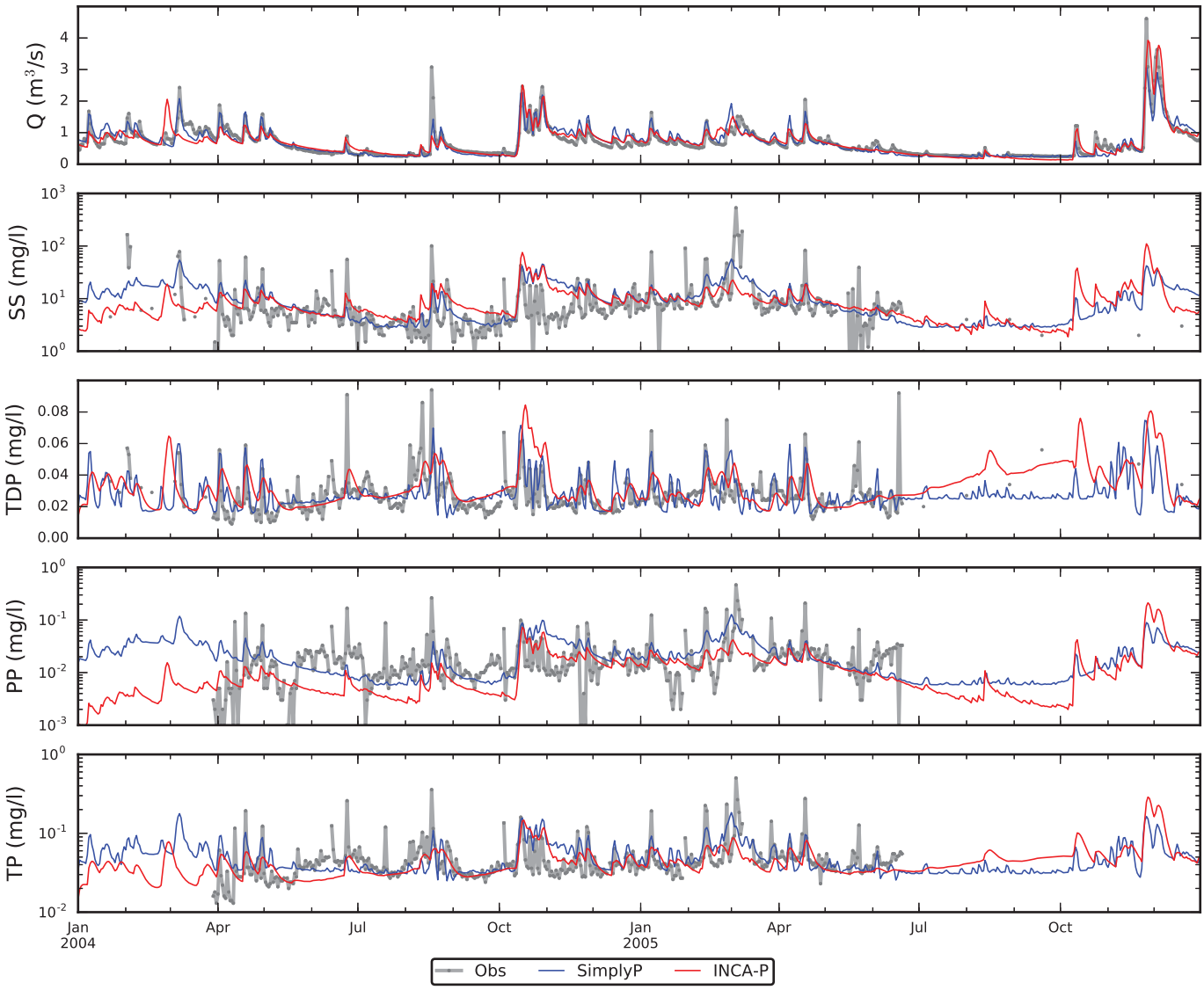

Q (m3/s)
SS (mg/l)
TDP (mg/l)
PP (mg/l)
TP (mg/l)
Jan 2004
Apr
Jul
Oct
Jan 2005
Apr
Jul
Oct
Obs
SimplyP
INCA-P

Figure 5.

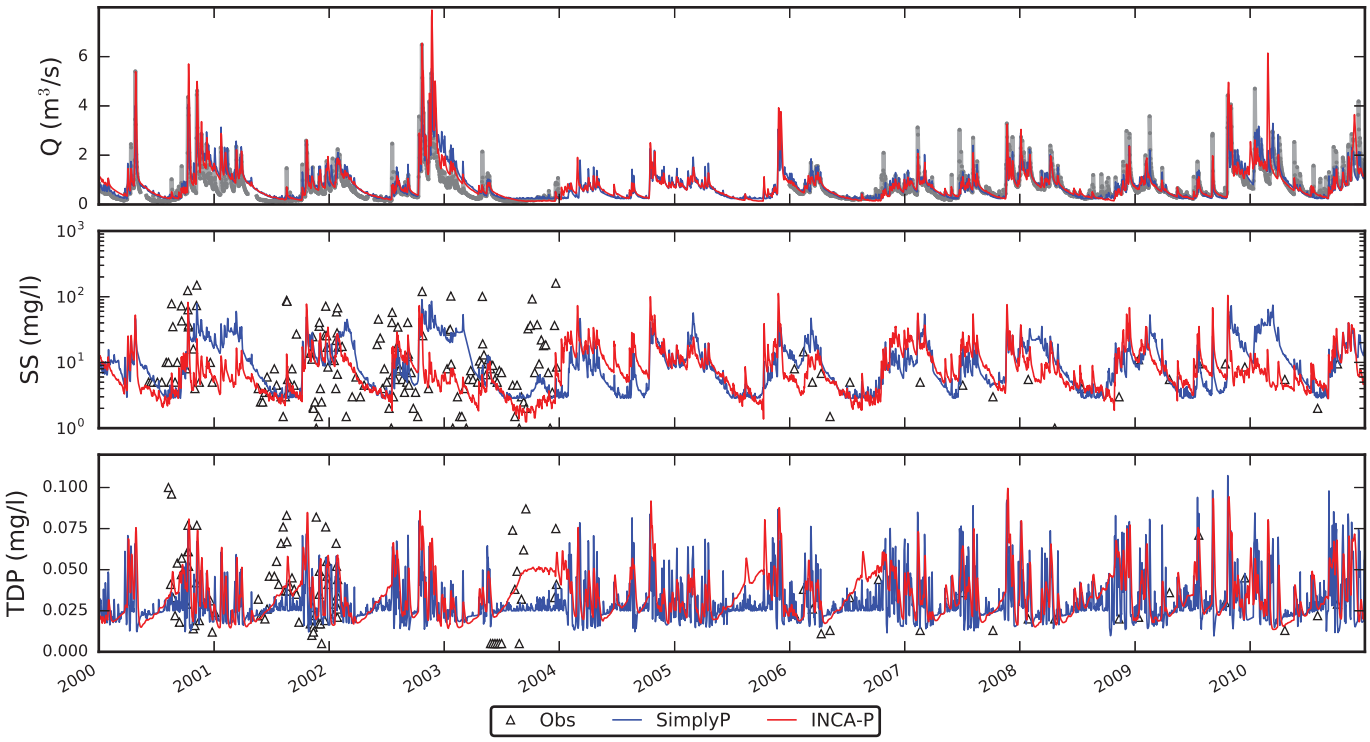
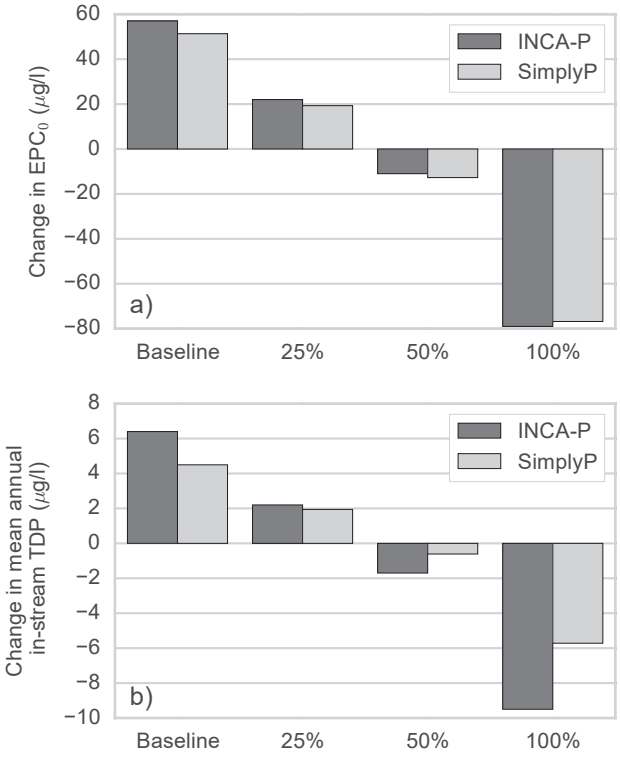

Q (m3/s)
0
2
4
6
SS (mg/l)
10^0
10^1
10^2
10^3
TDP (mg/l)
0.000
0.025
0.050
0.075
0.100
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Obs
SimplyP
INCA-P

Figure 6.


60
40
20
0
−20
−40
−60
−80
Change in EPC$_0$ (μg/l)
INCA-P
SimplyP
a)
Baseline
25%
50%
100%
8
6
4
2
0
−2
−4
−6
−8
−10
Change in mean annual in-stream TDP (μg/l)
INCA-P
SimplyP
b)
Baseline
25%
50%
100%

Water Resources Research

Supporting Information for

Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P

L.A. Jackson-Blake[a,b], J.E. Sample[a,b], A.J. Wade[c], R.C. Helliwell[b], R.A. Skeffington[c]

[a] Norwegian Institute for Water Research (NIVA), Gaustadalleen 21, 0349 Oslo, Norway
[b] The James Hutton Institute, Macaulay Drive, Aberdeen, AB15 8QH, UK
[c] Department of Geography and Environmental Science, University of Reading, RG6 6DW, UK

Contents of this file

Introduction

This supplementary information contains a full description of the SimplyP model, a newly-developed parsimonious catchment-scale dynamic water quality model for simulating hydrology, phosphorus and sediment dynamics in catchments. The model was originally developed in 2015 at the James Hutton Institute (Scotland), as part of L. Jackson-Blake's PhD. In the following document, a full description is provided of model aims, scope and scale, the processes and equations included in SimplyP v1.0 and the underlying scientific rationale, a description of numerical methods used to solve the equations, and a suggestion of model development priorities.

Text S1.

SimplyP v1.0 model description

Contents

1. Introduction

SimplyP is a parsimonious catchment-scale dynamic water quality model, which simulates hydrology, phosphorus (P) and sediment dynamics in catchments. The model was originally developed in 2015 at the James Hutton Institute (Scotland), as part of L. Jackson-Blake's PhD. The model is under development, lead by NIVA (Norway), with input from the James Hutton Institute. Here, we present model aims, scope and scale (Section 3) and give a detailed overview of the processes and equations included in SimplyP v1.0, as well as the underlying scientific rationale. A suggestion of model development priorities is provided in Section 5.

2. Model availability

SimplyP v1.0 model code is open source (https://github.com/LeahJB/SimplyP). Running the model requires a Python installation able to run iPython notebooks (e.g. WinPython; http://winpython.sourceforge.net/); instructions for installing WinPython and running the model using an example dataset are provided in the GitHub repository. Model parameters are input to the model via a simple Excel interface; recommendations for default, minimum and maximum values and potential data sources are provided in Table 11.

3. Model aims, scope and scale

3.1 Model aims

SimplyP aims to be process-based, i.e. model equations reflect hypotheses about the processes governing system behaviour. The aim in developing SimplyP was to minimize the process representation to only those processes that appear to control the catchment response, whilst maintaining the flexibility and functionality required for the model to be useful in hypothesis and scenario testing. The process representation aims to be simple enough to allow parameter values to be constrained using available observational data. What is presented here is a first prototype, the aim being to attempt a proof-of-concept that simple can be as good as complex. The hope is that this simple model could provide a useful benchmark when choosing between different models, or, after further testing, be a useful modelling tool in its own right.

SimplyP was developed with a number of potential uses in mind, including: (1) interpolation of sparse monitoring data, to provide more ecologically-relevant in-stream phosphorus concentrations or more accurate estimates of loads delivered downstream to lakes or estuaries; (2) hypothesis testing and highlighting knowledge and data gaps. This in turn could be used to help design monitoring programmes, highlight experimental needs, and prioritise areas for future model development; (3) exploring the potential response of the system to future change, especially in terms of anticipated storm and low-flow dynamics; and (4) providing evidence to support decision-making, for example to help set water quality and load reduction goals and to advise on the best means of achieving those goals.

A particular requirement at present is for models which can predict time lags in the system due to stores of 'legacy' P, P which has accumulated in the catchment along transport pathways in the land-freshwater continuum. These legacy P stores may become net sources of P if inputs to the store are reduced, potentially confounding management efforts aimed at reducing in-stream P concentrations (Sharpley et al., 2013). Stores of legacy P include: (1) soil P, from historic applications of fertilizer and manure in excess of crop requirements; (2) terrestrial sediment-bound P, stored in areas where local topography promotes sediment deposition; (3) groundwater P, due to percolation of P-rich water from agriculture or sewage; (4) P in up-stream impoundments (e.g. lakes, reservoirs and canals); and (5) P in in-stream bed sediments. The latter may be sourced from the deposition of particulate P in areas of slower flow (e.g. pools, floodplains), or from the adsorption of dissolved P from the water column, particularly downstream of sewage treatment works. The aim here is for the model to reproduce the first of these stores, soil P, as this is often the largest and the most pervasive legacy source of P to the environment (Jarvie et al., 2013a; Kleinman et al., 2011; Sharpley et

al., 2013). The remaining legacy P stores are beyond the scope of this initial version of the model, though recommendations are provided for how more legacy P sources could be incorporated.

In the future, it is likely that well-devised auto-calibration and uncertainty analysis routines will become pre-requisites for model applications. To date, attempts to bring about a shift in modeller behaviour have focused on improving algorithms and making them more available, but have largely over-looked other important barriers such as the subjectivity of the analyses (Jackson-Blake and Starrfelt, 2015) and the often prohibitively large amounts of time required to set up and conduct such analyses. We believe that the use of auto-calibration and uncertainty analysis tools would increase substantially if more attention were paid to: (1) developing models with fewer parameters, so all uncertain parameters can be included in an analysis; (2) reducing the number of interacting parameters, to reduce non-identifiability issues and the need for time-consuming *ad hoc* code to be written; and (3) maximising the number of parameters which can in principle be measured, and therefore given informative priors. SimplyP was developed with these three aims in mind. In practice, the number of parameters that can be calibrated in a given study area depends on data available for model calibration and for constraining model parameters to plausible ranges. However, studies have rarely successfully explored more than 40-dimensional parameter spaces (e.g. Dean et al., 2009; Jackson-Blake and Starrfelt, 2015; Starrfelt and Kaste, 2014), so an upper limit of 40 parameters was decided on.

3.2 Model temporal and spatial scale

The model is dynamic and currently runs at a daily time step, short enough to capture much of the variability in catchment hydrology and the associated delivery of dissolved and particulate matter to the water course, yet not so short that computing run times become prohibitively long when the model is run to explore longer term trends. A daily time step is also compatible with the majority of widely-available meteorological data sets used to drive the model.

The model is spatially semi-distributed, as a compromise between the complexity of fully-distributed methodologies and the lack of spatial process representation in fully lumped models. The catchment is broken down into perceived biophysical regions, thereby allowing a certain amount of the spatial variability in processes and outputs to be simulated, whilst reducing input data requirements and computing run times. The main disadvantage compared to fully-distributed models is that the interconnectedness of different parts of the landscape is not included, but in most areas this is justified by the lack of input data for a finer-scale division of the landscape and the reduced computing run times. The catchment area may be split into:

1. *Sub-catchments and associated reaches*, which should be defined based on the presence of monitoring stations, sewage treatment work inputs, or major changes in terrestrial conditions such as geology, topography or soil type. The model is run for each sub-catchment in turn, and reach outputs are fed sequentially down-stream.

2. *Grouped soil properties and broad land management, termed the land class*. A summary of the land class sub-divisions is given in Table 1. For dissolved P processes, two land classes are considered, a '*high P*' class and a '*low P*' class. Land within a class should have a similar annual P budget, soil P content and hydrological characteristics, or area-weighted properties should be used. A third optional '*newly-converted*' class may also be included, to take into account legacy soil P when land use conversion occurs, or the lack of legacy soil P in new agricultural land. For sediment and particulate P processes, the high P class may be further sub-divided in two to account for differences in erodibility (e.g. agricultural land could be split into improved grassland and arable land). Finally, the proportion of spring versus autumn-sown crops that make up any arable land may be taken into account when calculating the variation in soil erodibility due to plant cover through time.

   In highly agricultural areas where there is no 'low P' land, if the model is to be used to explore impacts of changing fertilizer or manure inputs or land use, the expected inactive P content for semi-

natural land in the area must still be provided, to provide a reference point for the P enrichment of agricultural soils, but the 'low P' class coverage of the catchment would be set to zero.

<table>
<tr><th>Processes</th><th colspan="3">Landscape division</th></tr>
<tr><td>Dissolved P</td><td colspan="2">High P
e.g. arable land, improved grassland</td><td>Low P
e.g. unfertilized forest, moorland, rough grassland</td></tr>
<tr><td rowspan="3">Sediment & particulate P</td><td colspan="2">High P</td><td rowspan="3">Low P</td></tr>
<tr><td colspan="2">High erodibility*
e.g. arable</td><td rowspan="2">Low erodibility*
e.g. improved grassland</td></tr>
<tr><td>Spring-sown*</td><td>Winter-sown*</td></tr>
</table>

*Table 1: Land class sub-divisions available for use in SimplyP. These differ according to the process being simulated (dissolved or particulate phosphorus processes). *Optional*

4. SimplyP model structure and equations

A summary of the main stores and fluxes of water, sediment and P is provided in Figure 1. The model includes the following components:

- A snow accumulation and melt model (Section 4.1.1)
- Rainfall-runoff (Section 4.1.2) and in-stream hydrology (Section 4.1.3)
- Sediment delivery to the watercourse and in-stream transport (Section 4.2)
- Terrestrial and in-stream P processes (Section 0)

Dissolved and particulate P are simulated separately as total dissolved P (TDP) and particulate P (PP). PP is assumed to be sediment-bound and no distinction is made between organic and mineral P.

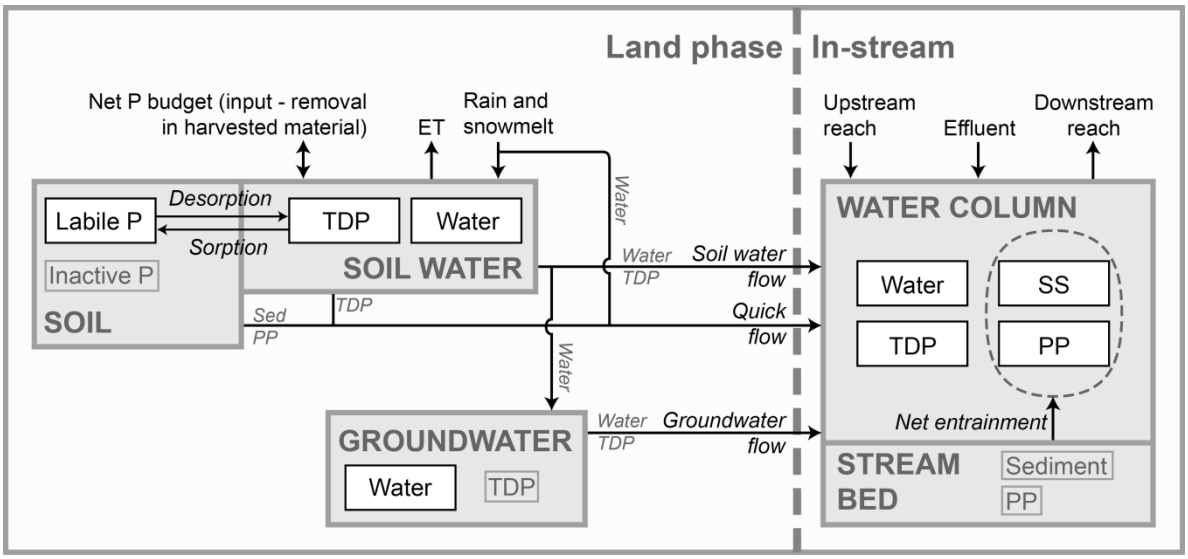


Figure 1: Schematic of the main stores, processes and pathways included in the model. White boxes show the state variables whose volume (water) or mass (sediment, P species) is tracked. Arrows show fluxes within and between compartments. P: phosphorus, SS: suspended sediment, TDP: total dissolved P, PP: particulate P, ET: evapotranspiration.

4.1 Hydrology

4.1.1 Hydrological inputs

Input time series of precipitation and potential evapotranspiration (PET) are required. If the snow module is not used, then these are used to drive hydrological processes in the model. If the snow accumulation and melt

module is used, then air temperature must also be supplied, and a time series of rain plus snowmelt is calculated and used to drive hydrological processes in the model. Parameters and variables used in the hydrological input equations are defined in Table 2.

Variable	Description	Units	Source
D_{snow}	Snowpack depth	mm	Equation 3
$D_{snow,0}$	Initial snowpack depth	mm	Input parameter
f_{DDSM}	Degree-day factor for snowmelt	mm/degree-day °C	Input parameter
P	Hydrological input to the model	mm day^{-1}	Equation 4
$P_{melt,max}$	Potential maximum snowmelt	mm day^{-1}	Equation 2
P_{rain}	Precipitation as rain	mm day^{-1}	Equation 1
P_{snow}	Precipitation as snow	mm day^{-1}	Equation 1
P_{total}	Total precipitation	mm day^{-1}	Input time series
T_{air}	Mean daily air temperature	°C	Input time series

Table 2: Parameters and variables used in calculating hydrological inputs to the model

Within the snow model, precipitation falls as snow when the mean daily air temperature is below 0°C (Equation 1). Potential daily snow melt is calculated using a simple degree-day factor method (USDA, 2004), assuming that melting begins once the air temperature rises above 0°C (Equation 2). Both temperature thresholds are hard-coded into the model at present, but could be readily converted to user-supplied parameters. The degree-day approach to simulating snow melt is one of the simplest formulations. Key limitations are discussed in USDA (2004), in particular: (1) the degree-day factor is assumed to be constant, whilst in reality it varies seasonally and by location; (2) snow melt is only controlled by temperature, and therefore ignores important factors such as snow density; and (3) it is not valid for rain-on-snow events.

Equation 1: Partitioning of total precipitation into rain, P_{rain}, and snow, P_{snow} (mm day^{-1})

When $T_{air} > 0$: $P_{rain} = P_{total}, P_{snow} = 0$
Otherwise: $P_{rain} = 0, P_{snow} = P_{total}$

Equation 2: Potential daily snowmelt, $P_{melt,max}$ (mm day^{-1})

When $T_{air} > 0$: $P_{melt,\,max} = f_{\mathrm{DDSM}}(T_{air} - 0)$

Snow pack depth is then calculated as initial depth plus the change due to snowfall and snowmelt, with the latter limited by the snow pack depth (Equation 3). As snowfall and melt are constant over a day, there is no need for Equation 3 to be formulated as a differential equation. Finally, the sum of rain and snowmelt is used as input to the hydrology model (Equation 4).

Equation 3: Snow pack depth, D_{snow} (mm), where superscript *t* denotes the current time step, and *t-1* the previous time step. This calculation requires a user-supplied initial snowpack depth, $D_{snow,0}$

$$D_{snow}^{t} = D_{snow}^{t-1} + P_{snow}^{t} - \mathrm{minimum}(P_{melt,max}^{t}, D_{snow}^{t-1})$$

Equation 4: Hydrological inputs to the model, P (mm day^{-1})

$$P = P_{rain} + P_{melt}$$

An example of output from the snow module for the Tarland Burn catchment for the period 2004-2005 is shown in Figure 2.

The snow model uses mean daily air temperature as input, which could lead to under- or over-estimation of snowfall and snowmelt. In areas where there is significant winter accumulation of snow, a possible future modification to the model would be to use a more sophisticated representation of temperature variation throughout a day, for example using a triangular or a sinusoidal shape.

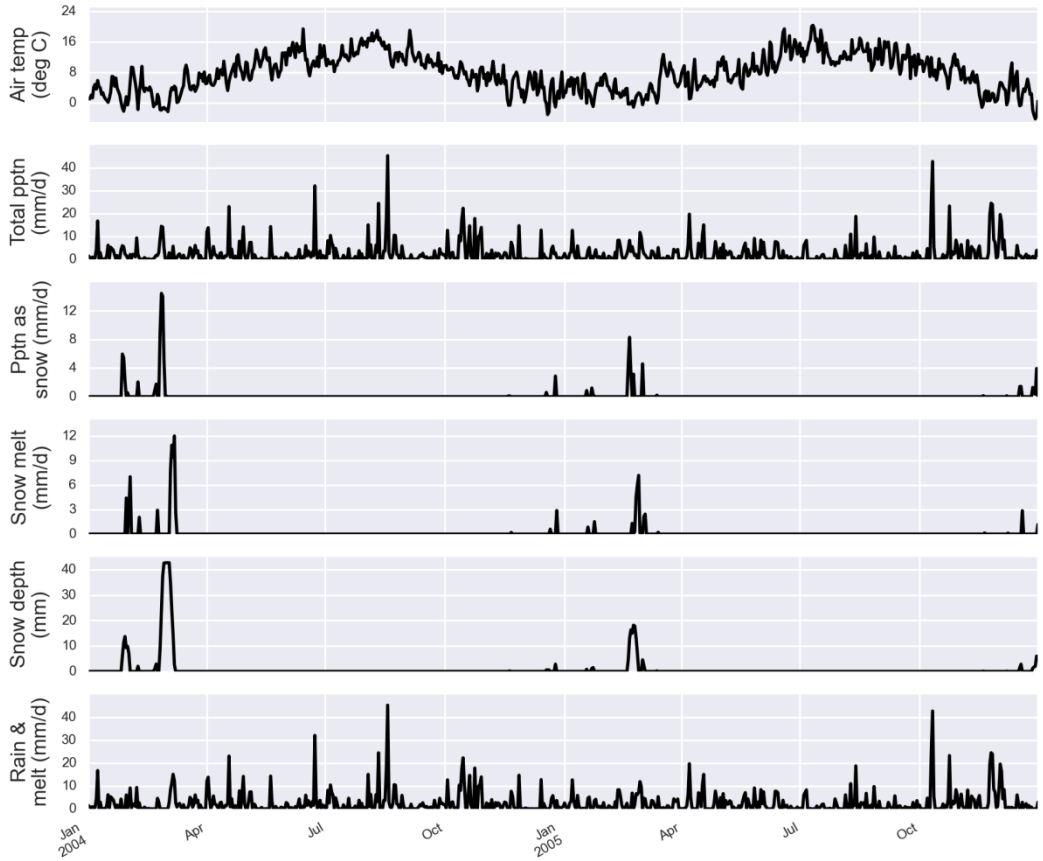


Figure 2: Inputs to and outputs from the snow accumulation and melt module, using input data for the Tarland Burn catchment. 'Pptn' is precipitation. Snow depths are given as mm of water equivalent.

4.1.2 Terrestrial hydrology

A hydrology model provides the foundation for any water quality model, as the flow of water transports matter from the land to the stream. To simulate hydrology alone, simple models may be appropriate, with just a couple of calibration parameters (e.g. IHACRES, Croke et al., 2006). However, to be able to simulate the transport of both dissolved and particulate P under varying flow conditions, a slightly more complex representation of terrestrial hydrology is needed. Here, three terrestrial flow paths are taken into account: (1) quick flow, to simulate water, sediment and P inputs to the watercourse during larger rainfall events and when soils are dry (and therefore little soil water flow occurs); (2) soil water flow, responsible for TDP leaching from soils and groundwater recharge, and (3) groundwater flow, which is important for baseflow P concentrations.

The soil water and groundwater equations used were based on those described by Sample (2015), who provides a full description of their derivation (last accessed March 2016). For convenience, volumes and fluxes are expressed as water depths per unit area in units of mm or mm day^{-1}. Associated real volumes can be derived by multiplying by the catchment area, A_{SC} (km^2), with unit conversions: 1 m^3 = 10^3 A_{SC} mm. Parameters and variables used in the terrestrial hydrology equations are defined in Table 3.

Variable	Description	Units	Source
α	Potential evapotranspiration correction factor	none	Input parameter
β	Base flow index	none	Input parameter
dQ_g/dt	Change in groundwater flow with time	mm day^{-1} day^{-1}	Equation 13
dQ_s/dt	Change in soil water flow with time	mm day^{-1} day^{-1}	Equation 10
dQ_s/dV_s	Change in soil water flow with soil water volume	mm day^{-1} mm^{-1}	Equation 9
dV_g/dt	Change in groundwater volume with time	mm day^{-1}	Equation 11
dV_s/dt	Change in soil water volume with time	mm day^{-1}	Equation 6
E_a	Actual evapotranspiration	mm day^{-1}	Equation 7
E_p	Potential evapotranspiration	mm day^{-1}	Input time series
$f_E(V_s)$	Function to limit evapotranspiration when soil water level drops below field capacity	none	Equation 7
f_{quick}	Proportion of hydrological inputs to the soil that contribute to quick flow	none	Input parameter
$f_{sw}(V_s)$	Function to limit soil water flow once soil water level reaches field capacity	none	Equation 8
P	Hydrological inputs to the soil (rain plus snowmelt)	mm day^{-1}	Equation 4
Q_g	Groundwater flow	mm day^{-1}	Equation 13
$Q_{g,min}$	Minimum groundwater flow	mm day^{-1}	Input parameter
Q_q	Quick flow	mm day^{-1}	Equation 5
Q_s	Soil water flow	mm day^{-1}	Equation 10
T_g	Baseflow recession constant	days	Input parameter
T_s	Soil water time constant	days	Input parameter
V_g	Groundwater volume	mm	Equation 11
V_{FC}	Field capacity	mm m^{-1}	Input parameter
V_s	Soil water volume	mm	Equation 6

Table 3: Parameters and variables used in the terrestrial hydrology equations

a) Quick flow

Quick flow is conceptualised to include a host of rapid flow pathways, including infiltration and saturation excess overland flow, tile drainage, runoff from impervious surfaces and macropore or preferential flow through soils. In practice, it is difficult to differentiate between these various rapid flow paths when calibrating using in-stream discharge data. Therefore as a starting point all were lumped into a single input to the stream, calculated as a function of incoming precipitation. A number of options for characterising quick flow were considered (in order of increasing complexity):

1. Assume quick flow is directly proportional to incoming precipitation and is routed instantaneously to the stream. This involves just one calibration parameter, f_{quick}, the proportion of precipitation that contributes to quick flow:

$$Q_q = f_{quick}P$$

2. Assume quick flow only occurs above some precipitation threshold, I_{max}, with all precipitation routed instantaneously to the stream above this threshold. As with option (1), this involves just one calibration parameter:

$$Q_q = \max\{(P - I_{max}), 0\}$$

3. A third option builds on option 2, including a factor to describe the proportion of the precipitation that contributes to quick flow once the threshold has been exceeded. It therefore has two parameters, f_{quick} and I_{max}:

$$Q_q = f_{quick} \max\{(P - I_{max}),0\}$$

4. Adopt one of the above approaches to describe the hydrological inputs to a quick store of water, and track the change in volume and flow of water in the store. This involves an additional user-input parameter, the time constant of the store, and therefore involves two or three user-supplied parameters. This approach is used by many process-based catchment water quality models, e.g. INCA.

Option 4 was discounted as being unnecessarily complex: the time constant is likely to be set to a small value in most catchments, and so the assumption in options 1 to 3 of quick flow directly entering the stream without a time lag is usually justifiable. As a starting point option 1 was chosen, being the simplest (Equation 5), but future work comparing options within a formal statistical framework would be useful.

Equation 5: Quick flow inputs to the stream, Q_q (mm day^{-1})

$$Q_q = f_{quick}P$$

Despite being simple, this approach allows summer flow events to be simulated, as a proportion of all precipitation enters the stream even when soil water level is below field capacity (often the case during the summer in temperate regions), when no soil water flow is simulated. It is important to be able to simulate these summer flow events, as they are often associated with nutrient peaks. Several models address the problem by adding in an additional quick flow path when soil water drops below field capacity (e.g. PERSiST, Futter et al., 2014). However, it is conceptually more consistent for this process to occur whatever the soil water level, as in Equation 5.

An important limitation of the adopted approach is a lack of seasonality in the generation of quick flow; in reality, quick flow is likely to be higher when soils are saturated, which is often the case during winter in temperate regions. A potential future extension to the model would be to include this saturation excess flow, which would require a re-formulation of the soil water equations.

b) Soil water

The soil water ordinary differential equations (ODEs) are solved separately for semi-natural and agricultural land, so the equations described in this section are present in the model for both land use classes. The change in soil water volume with time is given by Equation 6. Inputs to the soil water are from rainfall and snowmelt (Section 4.1.1), taking into account the proportion that is routed to quick flow; outputs are through evapotranspiration (E_a) and soil water flow.

Equation 6: Rate of change in soil water volume, V_s, with respect to time (mm day^{-1})

$$\frac{dV_s}{dt} = (1 - f_{quick})P - E_a - Q_s$$

E_a is calculated from a user-supplied time series of potential evapotranspiration (E_P), taking into account: (1) differences between land cover and topography in the study catchment compared to reference values used to compute E_p, through the optional use of a correction factor, and (2) moisture limitation once soil water drops below field capacity (Equation 7). To achieve the link between evapotranspiration and soil water content, an additional variable is needed in Equation 7 which tends to 1 as V_s approaches field capacity and to 0 as V_s tends to 0. A convenient function which displays this behaviour is $f_E(V_s) = 1\text{-}e^{\mu V_s}$, where μ is a tuning parameter that determines the shape of the curve (Fenicia et al., 2011). The value of μ is determined within the model as a function of field capacity: for the desired behaviour, E_A should be close to E_P when the soil water is at field capacity, i.e:

$$E_P = k\,E_A \text{ when } V_s = V_{FC}, \text{ where } k \text{ is near 1}$$

Substituting this into Equation 7 and re-arranging gives $\mu = -\ln(1\text{-}k)/V_{FC}$. The $f_E(V_s)$ function is plotted in Figure 3 for a variety of values of k and for the minimum and maximum likely field capacity values (100 to 450 mm/m; Ratliff et al., 1983). Based on this plot, k was fixed at 0.99 so that $f_E(V_s)$ is close to 1 at field capacity and drops away relatively quickly below field capacity.

Equation 7: Actual evapotranspiration, E_a (mm day^{-1}). The function $f_E(V_s)$ limits evapotranspiration when soil water content drops below field capacity

$$E_a = \alpha E_p \, f_E(V_s) = \alpha E_p \, (1 - e^{-\mu V_s}), \text{where } \mu = \frac{-\ln(0.01)}{V_{FC}}$$

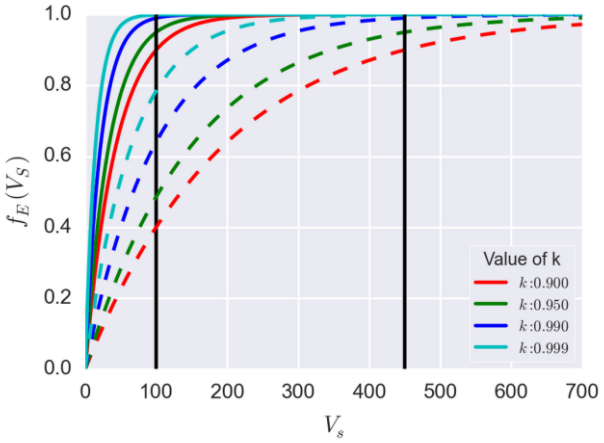


Figure 3: Relationship between $f_E(V_s)$ and soil water volume (V_s/mm) for a variety of k factors, where $f_E(V_s) = 1\text{-}e^{\mu V_s}$ and $\mu = \text{-}ln(k)/V_{FC}$. The vertical lines mark the minimum and maximum likely values for field capacity (FC). Solid lines: FC=100; dashed lines: FC=450 mm m^{-1}.

It is assumed that soil water flow only takes place when the soil water level is above field capacity and that flow is proportional to the volume of water above field capacity (Equation 8). The constant of proportionality is $1/T_s$, where T_s is the soil water time constant. The additional function in Equation 8, $f_{sw}(V_s)$, prevents soil water flow from occurring when soil water content drops below field capacity. One option is for this function to involve a set of logical conditions (e.g. $f_{sw}(V_s) = 0$ when $V_s > V_{FC}$; otherwise $f_{sw}(V_s) = 1$). However, this kind of logic introduces non-differentiable discontinuities into the ODE. Instead, a continuous sigmoid function was used (Fenicia et al., 2011), which yields a curve which switches rapidly from zero to one around field capacity (Figure 4).

Equation 8: Discharge from the soil water store, Q_s (mm day^{-1})

$$Q_s = \frac{1}{T_s}(V_s - V_{FC}) \, f_{sw}(V_s) = \frac{1}{T_s}(V_s - V_{FC}) \left(\frac{1}{1 + e^{(V_{FC} - V_s)}} \right)$$

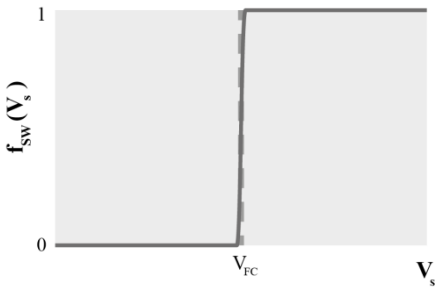


Figure 4: Schematic demonstrating the form $f_{sw}(V_s)$ takes when a sigmoid curve is used

To derive an equation for the rate of change in soil water flow with time, dQ_s/dt, Equation 8 must be differentiated with respect to time. To do this, dQ_s/dV_s is first defined by differentiating Equation 8 with respect to soil water volume (Equation 9), which is then used to derive dQ_s/dt (Equation 10).

Equation 9: Rate of change in soil water flow, Q_s, with respect to soil water volume (day^{-1})

$$\frac{dQ_s}{dV_s} = \frac{(V_s - V_{FC})e^{V_{FC} - V_s}}{T_s(e^{V_{FC} - V_s} + 1)^2} + \frac{1}{T_s(e^{V_{FC} - V_s} + 1)}$$

Equation 10: Rate of change in soil water flow, Q_s, with respect to time (mm day^{-2})

$$\frac{dQ_s}{dt} = \frac{dV_s}{dt}\frac{dQ_s}{dV_s}$$

Finally, a proportion (β) of the discharge from the soil box is assumed to percolate to groundwater, whilst the remainder of the soil water flow ($1 - \beta Q_s$) is assumed to travel to the stream along shallow flow pathways.

c) Groundwater

The rate of change of groundwater volume with time is controlled by the balance of inputs from the soil water and outputs via groundwater flow (Equation 11). Percolation from the soil water is calculated as a fraction (β) of the soil water outflow (where β is the baseflow index). Groundwater flow is assumed to be directly proportional to groundwater volume (Equation 12), and can be differentiated with respect to V_g to give $1/T_g$, where T_g is the baseflow recession constant. As with the soil water flow, the rate of change in groundwater flow with time is then calculated as the product of dV_g/dt and dQ_g/dV_g (Equation 13). The baseflow recession constant is determined through model calibration or from discharge observations using hydrograph separation techniques (e.g. Van Dijk, 2010). Beck et al. (2013) have produced a global dataset of baseflow index and recession constants which could provide useful starting points for model calibration.

Equation 11: Rate of change of groundwater volume, V_g, with time (mm day^{-1})

$$\frac{dV_g}{dt} = \beta Q_s - Q_g$$

Equation 12: Relationship between groundwater flow and volume

$$Q_g = \frac{1}{T_g}V_g; \text{therefore} \frac{dQ_g}{dV_g} = \frac{1}{T_g}$$

Equation 13: Rate of change of groundwater flow, Q_g, with time (mm day^{-1} day^{-1})

$$\frac{dQ_g}{dt} = \frac{dV_g}{dt}\frac{dQ_g}{dV_g} = \frac{\beta Q_s - Q_g}{T_g}, \text{where } Q_g = \text{maximum}(Q_{g,min}, Q_g)$$

The minimum groundwater flow parameter in Equation 13, $Q_{g,min}$, prevents groundwater flow from dropping below a user-specified threshold. Without this, sustained periods when soil water is below field capacity may result in little groundwater recharge and under-simulation of low flows (red line in Figure 5 for the Tarland catchment). This threshold is implemented by testing whether the groundwater flow at the end of the day is below the user-specified threshold. If it is, the initial conditions at the start of the next time step are set to the threshold. It is therefore possible for groundwater flow within a day to drop below the threshold, but generally not by much.

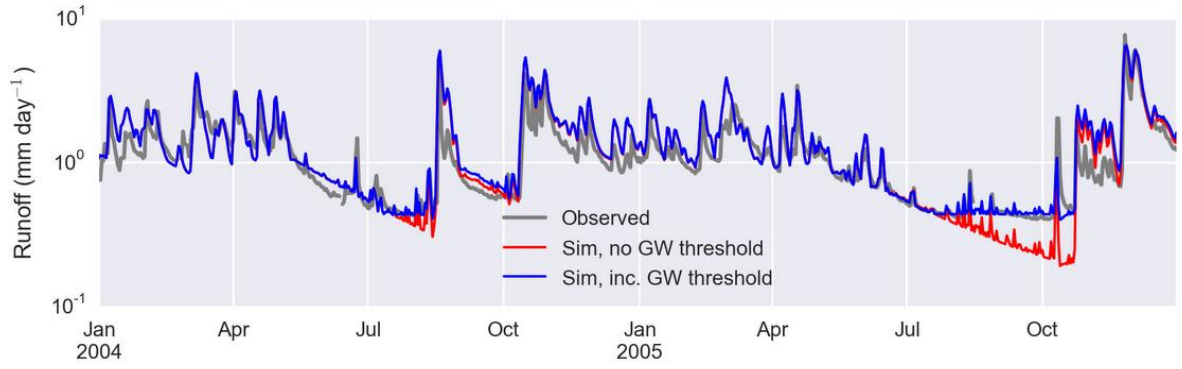


Figure 5: Comparison of simulated runoff with and without inclusion of a minimum groundwater flow in the Tarland catchment. Note log scale.

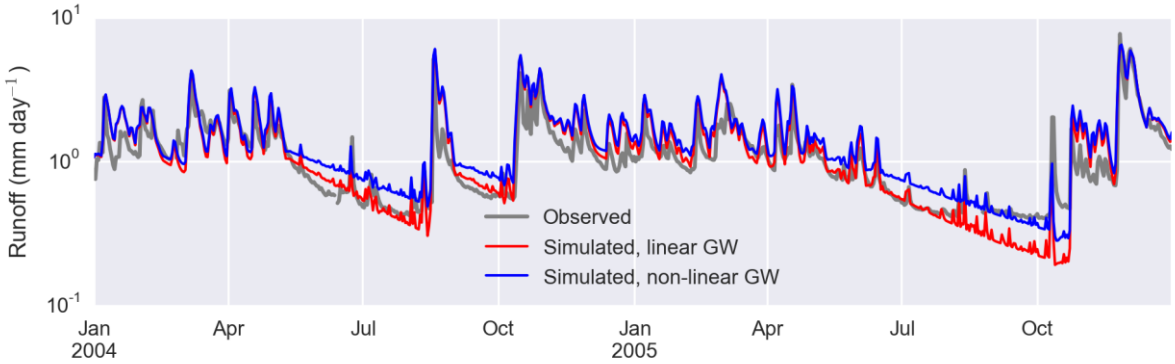


Figure 6: Comparison of simulated runoff generated using a linear and a non-linear groundwater reservoir in the Tarland. Note the log scale.

Poor representation of low flows is a common weakness in hydrological simulations (Fenicia et al., 2006), leading to potentially large errors in simulated solute concentrations during the ecologically-sensitive summer period. Whilst preventing groundwater flow from dropping below a certain threshold improves the simulation, it is not an ideal solution: (1) it circumvents the fact that something is wrong with the model conceptualisation if sustained summer baseflows cannot be simulated, and (2) it does not satisfy the internal catchment water balance, as the excess water required to sustain groundwater at the user-specified threshold is not sourced from within the catchment. This approach therefore works for the purposes of developing a prototype model, and indeed is that adopted by e.g. INCA. However, alternative options were also considered, and should be a priority for future development of the model. These include:

1) Replacing the linear relationship between groundwater volume and flow with a non-linear relationship, so that $V_g = T_g Q^k$, as described by Wittenberg (1999), where k, the non-linear coefficient, reflects the influence of aquifer properties on Q_g. Equations for the rates of change in groundwater volume and flow therefore become:

$$\frac{dQ_g}{dV_g} = \frac{V_g^{\frac{1-k}{k}}}{kT_g^{\frac{1}{k}}} \text{ (which simplifies to } \frac{1}{\mathrm{T_g}} \text{ when k = 1)}$$

$$\frac{dV_g}{dt} = \beta Q_s - \left(\frac{V_g}{T_g}\right)^{\frac{1}{k}}$$

$$\frac{dQ_g}{dt} = \frac{dQ_g}{dV_g}\frac{dV_g}{dt}$$

A comparison of simulations using a linear and a non-linear groundwater store in the Tarland are shown in Figure 6. In this case the non-linear model produces a flatter recession, but simulated low flows are

too high. Discussions as to whether the groundwater storage-discharge relationship should be considered linear or non-linear have been on-going in the literature for some time, and certainly in some cases a non-linear reservoir appears to perform better (e.g. Gan and Luo, 2013). Other researchers have added in additional storage boxes to help improve model performance (e.g. Luo et al., 2012). In cases this may be justified, but Fenicia et al. (2006) make a strong argument for non-linear reservoirs often appearing to perform better only when important fluxes into or out of the groundwater storage zone have been neglected.

2) The second option considered was therefore that the model is missing an additional flux into the groundwater, e.g. slow recharge from the unsaturated zone. Possible alternative formulations of the soil water store, which would allow for some groundwater recharge to occur when the soil water level drops below field capacity, include:

    a. Adding in a small, constant flux from the soil water box to the groundwater, which is independent of soil water content. This would require a user-calibrated parameter to set the flux, and a function as used in Equation 7 to prevent negative soil water volumes.

    b. Changing the shape of the sigmoid function used to limit soil water drainage below field capacity, so that it behaves less like a step function and allows some continued percolation to groundwater below field capacity. This is a mechanistically justifiable option, as soils continue to produce water when water contents drop below field capacity, although the fluxes involved vary with soil type and texture (Twarakavi et al., 2009). Some practicalities of this option would need careful consideration, e.g. choosing an appropriate shape parameter for all possible values of field capacity, to ensure zero flow when water level drops to zero.

        The model only considers average conditions over the catchment, whilst in reality, although catchment-average soil water content may be at or below field capacity, soils in parts of the catchment (e.g. at the foot of slopes or in hollows) are likely to remain above field capacity, and therefore able to feed water into groundwater and surface watercourses. This is a limitation of using a semi-distributed approach, and provides further justification for using a less steep sigmoid function.

In summary, the simple linear store with a user-supplied minimum groundwater flow provides a working solution to the problem of simulating summer low flows, and is suitable for the purposes of the prototype model being developed here. However, improved realism should be achievable without much increase in model complexity, and exploring options to achieve this should be a high priority for any future development efforts.

d) Example model output

An example of terrestrial hydrology output for the Tarland Burn catchment for the period 2004-2005 is shown in Figure 7, using time constants of 1 day for agricultural land and 10 days for semi-natural land. For both land use classes, soil water volume is at or above field capacity for most of the winter and then drops below field capacity during the summer, when quick flow provides the only soil input to the stream. The larger soil water time constant in semi-natural land means that it responds more slowly to changing inputs than agricultural soil water, with lower, broader peaks in flow and volume after rainfall. The groundwater time constant is much longer than the soil water time constants (65 days), so changes in groundwater flow and volume are a highly damped version of the changes in the soil water stores.

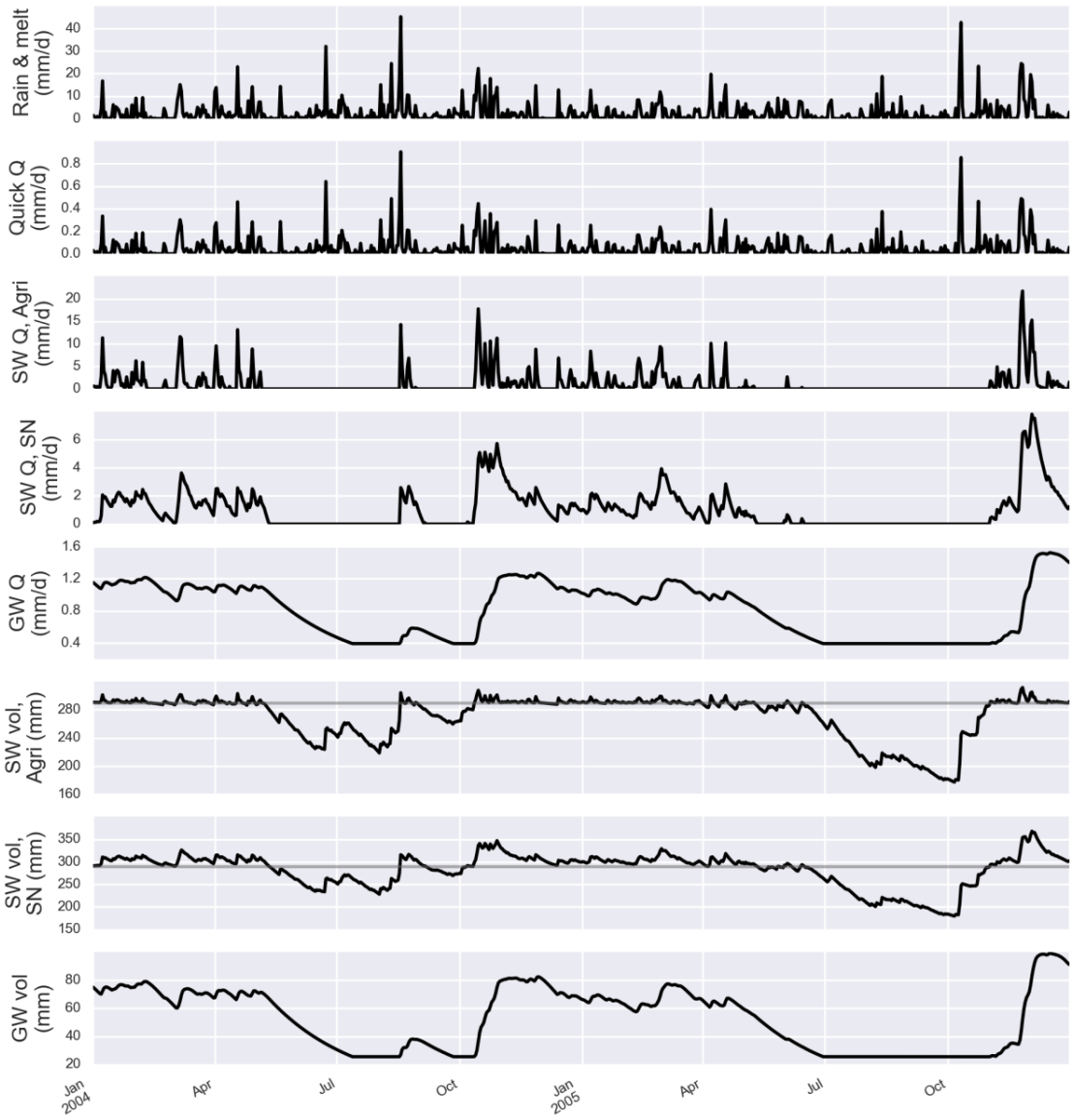


Figure 7: Terrestrial compartment hydrology results in the Tarland catchment. Q: discharge, Vol: volume, SW: soil water, GW: groundwater, Agri: agricultural land, SN: semi-natural land. Grey lines on the soil water volume plots mark the field capacity. N.B. volumes and fluxes are per unit area.

4.1.3 In-stream hydrology

Observed discharge can often be simulated without accounting for within-reach water travel times. However, the water quality model requires estimates of reach volume and discharge, to allow concentrations and fluxes from the reach to be calculated. Parameters and variables used in the instream hydrological equations are defined in Table 4. Note that, as in the terrestrial compartment, all volumes and fluxes are calculated using units of depth per unit area. The simulated daily mean flow is then converted to cumecs for comparison with observations (1 $m^3 s^{-1} = 10^3\ 86400^{-1}\ A_{SC}$ mm day^{-1}, where A_{SC} is sub-catchment area in km^2).

Variable	Description	Units	Source
a	Gradient of in-stream velocity-discharge relationship	m^{-2}	Input parameter
b	Exponent of in-stream velocity-discharge relationship	None	Constant (0.42)
β	Base flow index	None	Input parameter
dQ_r/dt	Rate of change in reach discharge with time	mm day^{-1} day^{-1}	Equation 17
$dQ_{r,av}/dt$	Rate of change in daily mean flow with time	mm day^{-1} day^{-1}	Equation 18
dV_r/dt	Rate of change in reach water volume with time	mm day^{-1} day^{-1}	Equation 14
f_A	Proportion of agricultural land in the sub-catchment	None	Input parameters (f_{Ar} + f_{IG})
f_S	Proportion of semi-natural in the sub-catchment	None	Input parameter
L_{reach}	Reach length	M	Input parameter
Q_g	Groundwater flow	mm day^{-1}	Equation 13
Q_q	Quick flow	mm day^{-1}	Equation 5
Q_s^A	Soil water inputs from agricultural land	mm day^{-1}	Equation 10
Q_s^S	Soil water inputs from semi-natural land	mm day^{-1}	Equation 10
Q_r	Outflow from the reach	mm day^{-1}	Equation 17
$Q_{r,av}$	Daily mean flow from the reach	mm day^{-1}	Equation 18
$Q_{r,US}$	Inputs to the reach from upstream reaches	mm day^{-1}	Model calculates
T_r	Reach time constant	Days	Equation 16
V_r	Reach water volume	Mm	Equation 14

Table 4: Parameters and variables used in the instream hydrology equations

The instream hydrology equations used are similar to those used in other water quality models. The change in water volume in each reach is assumed to be proportional to the difference between input and output fluxes (Equation 14). Input fluxes are from quick flow, soil water flow, groundwater flow and inflow from upstream reaches, and water leaves via the reach outflow. The soil water input to the reach is calculated as the sum of inputs from semi-natural and agricultural land in the catchment.

Equation 14: Rate of change in reach volume, V_r, with time (mm day^{-1})

$$\frac{dV_r}{dt} = Q_q + (1-\beta)(f_A Q_s^A + f_S Q_s^S) + Q_g + Q_{r,US} - Q_r$$

Flow downstream, Q_r, is assumed to be proportional to the reach volume (Equation 15), with a constant of proportionality $1/T_r$ (where T_r is the reach time constant in days). T_r is not however constant as it is in the terrestrial stores, but is inversely proportional to water velocity to account for the shorter residence time of faster flowing water (Equation 16). Water velocity is estimated using an empirical relationship with discharge (Equation 16) (Chapter 14, Chapra, 2008; Leopold and Maddock Jr, 1953), where a is a user-supplied parameter, derived where possible from velocity profiling carried out as part of flow gauging, and b is a non-linear coefficient. The a parameter is the most site-specific, whilst b is usually between 0.3 and 0.5. To reduce model complexity and interactions between these two parameters, the value of b was therefore set as a constant with a value of 0.42, corresponding to the average from over 200 river cross sections in the US and Europe (Wolman et al., 1964). The units of a (m^{-2}) assume that observed velocity and discharge have units of m s^{-1} and m^3 s^{-1}, respectively.

Equation 15: Relationship between reach discharge, Q_r (mm day^{-1}) and volume, V_r (mm)

$$V_r = T_r Q_r = \frac{L_{reach}}{86400\ a} Q_r^{(1-b)}$$

Equation 16: Reach time constant, T_r (days), where U is the water velocity (m s^{-1})

$$T_r = \frac{L_{reach}}{U}; U = aQ_r^b, \text{ so } T_r = \frac{L_{reach}}{86400\ aQ_r^b}$$

The rate of change in discharge with time is given by Equation 17. The time-varying nature of T_r gives rise to the (1-b) factor in the denominator. To get a time series of daily mean flow, the rate of change in discharge with time is integrated with initial conditions set to zero at the start of each day (Equation 18), to provide the

total flux of water leaving the reach over a day. The units of this daily mean flow time series are then converted to cumecs for comparison with observations.

Equation 17: The rate of change in reach discharge, Q_r, with time (mm day^{-1} day^{-1})

$$\frac{dQ_r}{dt} = \frac{1}{T_r(1-b)}\left(Q_q + (1-\beta)(f_A Q_s^A + f_S Q_s^S) + Q_g + Q_{r,US} - Q_r\right)$$

Equation 18: Change in daily mean flow, $Q_{r,av}$ (mm day^{-1}), with time

$$\frac{dQ_{r,av}}{dt} = Q_r$$

The daily mean flow is used to determine the flux from the up-stream reach, $Q_{r,US}$. For the top reach, this is set to zero. For all other reaches, it is set equal to the daily mean flow from the upstream reach for the same time step. In this sense, the reaches are not fully coupled, as simulated within-day changes in flow are not cascaded from reach to reach. However, this assumption simplifies the coding considerably and should be appropriate in all but large, highly flashy systems.

4.2 Sediment processes

4.2.1 In-stream suspended sediment

In-stream SS concentrations are highly variable spatially and temporally, changing in response to terrestrial delivery, stream bank erosion, entrainment of bed sediment material and sediment deposition on floodplains and the stream bed. Terrestrial delivery is in turn controlled by the detachment of soil particles by raindrop impact, flow erosion and transport through the catchment, with deposition in areas of lower water velocity. Important factors affecting in-stream processes include stream power, particle size and type, stream bed morphology, macrophyte cover and antecedent conditions (Merritt et al., 2003). Incorporating these processes into a model would require many parameters to be calibrated, even with a simple formulation. In intensively-studied catchments some terrestrial data may be available to help constrain parameter values, but in most areas calibration is done using only in-stream SS concentration time series, and it is doubtful that these time series contain enough information for such a level of process representation.

Despite the varied and complex processes which represent sediment fluxes to and within streams, in-stream SS concentration has long been known to be well-explained by a simple power law with in-stream discharge: $SS = aQ_r^b$, where SS is suspended sediment concentration, Q_r is in-stream discharge, and a and b are parameters determined by regression (Colby, 1956). There are therefore strong indications that, at the catchment scale, this complexity may be simplified to a remarkably straightforward relationship. Here, this simple power law is therefore taken as the basis for simulating the change in SS mass with time in each stream reach. Parameters and variables used in the sediment equations are defined in Table 5.

Variable	Description	Units	Source
A_{SC}	Sub-catchment area	km^2	Input parameter
C_{cover}	Erosion soil cover factor	none	Equation 26
$C_{measures}$	Erosion management factor	none	Input parameter
dM_{sus}/dt	Rate of change in reach suspended sediment mass with time	kg day^{-1}	Equation 19
$dM_{sus,tot}/dt$	Change in total daily flux of sediment from the stream reach	kg day^{-1}	Equation 23
E_M	Sediment input scaling factor	kg mm^{-1}	Input parameter
E_{sus}	Sediment-discharge rating coefficient	kg mm^{-1}	Equation 21
f_{Ar}	Fraction of arable land in the sub-catchment	none	Input parameter
f_{IG}	Fraction of improved grassland in the sub-catchment	none	Input parameter
f_S	Fraction of semi-natural land in the sub-catchment	none	Input parameter
k_M	Instream erosion and entrainment non-linear coefficient	none	Input parameter
M_{input}	Flux of sediment to the reach from terrestrial and instream sources	kg day^{-1}	Equation 20
M_{sus}	Reach suspended sediment mass	kg	Equation 19
$M_{sus,DS}$	Flux of sediment downstream out of the reach	kg day^{-1}	Equation 22
$M_{sus,US}$	Flux of sediment from upstream reaches	kg day^{-1}	Model calculates
$M_{sus,tot}$	Total daily flux of sediment from the stream reach	kg day^{-1}	Equation 23
Q_r	Reach discharge	mm day^{-1}	Equation 17
$Q_{r,av}$	Daily mean discharge	mm day^{-1}	Equation 18
S_{SC}	Sub-catchment slope	°	Input parameter
S_{reach}	Reach slope	°	Input parameter
SS_r	Mean daily concentration of suspended sediment in the reach	mg l^{-1}	Equation 24
V_r	Volume of water in the reach	mm	Equation 14

Table 5: Parameters and variables used in the in-stream sediment equations

The rate of change in SS mass with time is controlled by the difference between sediment input and output fluxes (Equation 19). Inputs from the terrestrial compartment, in-stream channel erosion and entrainment are grouped into a single input term, M_{input}, assumed to be related to in-stream flow using a power law (Equation 20). It was originally envisaged that Equation 19 would include separate terms for sediment delivery from the land (proportional to quick flow) and instream entrainment and erosion (proportional to instream discharge). However, adding a quick flow-dependent term to Equation 20 made little difference to model output and was therefore discarded, and the entrainment term was re-formulated to represent both terrestrial and instream inputs.

Equation 19: Rate of change in suspended sediment mass in the stream reach, M_{sus}, with time (kg day^{-1}), where superscript *i* denotes the land use class. f^i is one of f_{Ar}, f_{IG} or f_S.

$$\frac{dM_{sus}}{dt} = \left(\sum_i f^i M^i_{input}\right) + M_{sus,US} - M_{sus,DS}$$

Equation 20: Flux of sediment to the stream from terrestrial and in-stream sources, M_{input} (kg day^{-1})

$$M_{input} = E_{sus} Q_r^{k_M}$$

This simple treatment of the sediment inputs to the reach assumes that the majority of in-stream SS is generated from within- or near-channel sources, as is often observed (Bowes et al., 2005). A further assumption is made that these near-channel sediment sources are directly controlled by catchment erodibility and delivery to the watercourse, so that a change in terrestrial erodibility causes an instant reduction in in-stream SS concentration. This assumption is based on the observation that the coefficient in Equation 20, E_{sus}, relates to the erodibility of soils in the catchment (reviewed in Asselman, 2000). To build in a link between terrestrial processes and in-stream SS, the value of E_{sus} therefore incorporates data on the relative differences in expected erosion fluxes from different land units used in the Universal Soil Loss Equation (USLE; Kinnell, 2010; Renard et al., 1991; Wischmeier and Smith, 1965, 1978). E_{sus} is therefore calculated per land class (e.g. arable, improved grassland and semi-natural) and sub-catchment by multiplying a user-calibrated scaling factor with factors representing erodibility and sediment delivery to the stream (Equation 21). These factors include average slope (S_{SC}, which affects the transport capacity of quick flow), a

vegetation cover factor (C_{cover}) and a management factor ($C_{measures}$), all of which may vary with land use, and the cover factor may vary through the year if desired (Section 4.2.2).

Equation 21: Sediment-discharge rating coefficient, E_{sus}, (kg mm^{-1}), where superscript *i* indicates that the variable or parameter varies by land use, and superscript *j* by sub-catchment.

$$E_{sus}^{i,j} = E_M S_{reach}^{j} S_{SC}^{i,j} C_{cover}^{i} C_{measures}^{i}$$

The cover factor, C_{cover}, describes the ratio between the erodibility of a bare soil plot and the land use class; its value therefore ranges from 1 (maximum erodibility) to 0 (no erosion). Cover factors can be sourced from USLE-related literature reviews, selecting an appropriate geo-climatic region and range of vegetation and crop types. For example, Panagos et al. (2015) have collated crop type factors for typical European crops (Appendix, Table A1) and other European land cover (Appendix, Table A2). The user can also incorporate relative differences in inherent soil erodibility of different land classes into the relative differences in C_{cover}, i.e. differences due to soil properties such as texture and organic matter content (generally termed the *K* factor in USLE-related literature). A potential future extension to the model for areas which are dominated by semi-natural land could be to build in a link between the cover factor on semi-natural land and pressures which are known to increase soil erodibility such as grazing, burning and tree felling.

The management factor, $C_{measures}$, allows the user to explore the effects of sediment reduction measures and should be in the range 0 to 1, where 0 implies 100% reduction in sediment yield. Measures that could be taken into account could, for example, relate to the effects of tillage, cover crops or measures to reduce sediment connectivity to the stream (e.g. buffers or fences). The model requires the user to know the effectiveness of the chosen measure for a given land class. Values for effectiveness can be obtained for example from the USLE literature (e.g. Panagos et al., 2015) or from experimental work within the study catchment.

Sediment input to the watercourse should not only vary with terrestrial erodibility and transport capacity, but also with changes in in-stream inputs. A large body of empirical and theoretical work has shown that in-stream sediment transport is controlled by stream power, i.e. the rate of energy expenditure on the stream bed and banks (Bagnold, 1966). In-stream sediment inputs to the watercourse were therefore assumed to be proportional to stream power per unit length, $\omega=\rho g S_{reach} Q_r$. This equation includes terms for the density of water (ρ) and gravitational acceleration (g), both of which are constant and therefore grouped into the user-supplied scaling factor (E_M) in Equation 21. The remaining term, the reach slope (S_{reach}) is then an additional factor in Equation 21.

Additional sediment inputs to the stream reach are via flow from any upstream reaches and sediment is lost via flow from the reach (Equation 22). The integral of the flux out of the reach over each day, starting with an initial condition of 0, then provides a time series of the total sediment flux from the reach per day (Equation 23), which is used to calculate daily mean SS concentration (Equation 24). The latter is output by the model for comparison with observations.

Equation 22: Reach suspended sediment outflow from the reach bottom, $M_{sus,DS}$ (kg day^{-1})

$$M_{sus,DS} = \frac{M_{sus}}{V_r} Q_r$$

Equation 23: Rate of change in the flux of sediment from the reach, $M_{sus,tot}$ (kg day^{-1})

$$\frac{dM_{sus,tot}}{dt} = M_{sus,DS}$$

Equation 24: Mean daily concentration of SS in the stream reach, SS_r (mg l^{-1})

$$SS_r = \frac{M_{sus,tot}}{Q_{r,av}} \frac{1}{A_{SC}}$$

Connectivity between sediment source areas and the water course is an important factor in determining sediment yield to a water course. This is not explicitly accounted for at present, although it is indirectly included in the value assigned to the scaling factor, E_M, in a given area. A potential improvement would be to explicitly include a connectivity factor in Equation 20, for example by: (1) using drainage density as a proxy for connectivity, (2) only considering the characteristics of land within a certain distance of a drainage ditch/watercourse, as land most likely to be a potential sediment source area; (3) factoring in field size and the presence of barriers to flow such as walls and hedges; (4) calibrating the sediment yield scaling factor by sub-catchment, rather than keeping it constant over the whole catchment. In addition, the sub-catchment slope factor could be altered to be more representative of potential sediment source areas, for example it could be the average slope of land within a certain distance of a watercourse.

Assuming that the amount of sediment in near-channel sources is directly proportional to terrestrial erodibility is a big simplification. An implication of this assumption is that a reduction in terrestrial erodibility causes an instant reduction in in-stream SS concentration. In reality, hysteresis is often seen in the sediment-discharge relationship, reflecting the change in sediment source distance or supply during a rainfall event (e.g. Oeurng et al., 2010). As the store of sediment in near-channel sources is not tracked, there is no ability to simulate source-exhaustion over successive storm events (Bowes et al., 2005). Test applications are required to determine whether this is an issue in a given study catchment. Over the longer term, we might expect a time lag between changes in erodibility and in-stream effects as in-stream sources become exhausted. For example, typical lag times of decades or more have been reported for the retention of bulk sediment in river channels (Trimble, 2010). However, here we are primarily concerned with the fine sediment fraction (silt and clay), as the fraction that is of greatest significance for P transport due to its relatively high P content. This finer sediment fraction is substantially more mobile, with residence times of less than a year reported for many rivers, excluding storage on floodplains (Sharpley et al., 2013, and references therein).

Other issues with this simple formulation are that sediment deposition is not explicitly accounted for, and so the model cannot simulate a net flux of sediment from the water column to the stream bed. This could be a problem in some areas, for example in catchments with a big difference in slope and sediment supply between reaches. Future work is needed to determine under which circumstances deposition needs taking into account, and how it should best be done in the simplest way possible. Another potential future improvement could be to split bank erosion and entrainment from terrestrial delivery. Bank erosion and entrainment may be a key sediment input (Lefrançois et al., 2007), and assuming terrestrial sediment reduction measures cause a proportional reduction in in-stream sediment load ignores the fact that bank erosion will be unaffected by these measures. Finally, the instream sediment equations only consider allochthonous particles sourced from the catchment. However, autochthonous particles, generated in-stream by biological processes, may make up an important part of suspended matter (Stutter et al., 2009), and in some areas it may be appropriate to consider both.

4.2.2 Dynamic cover factor

On arable land, the key risk period for soil erosion is between preparation of the seed bed and establishment of the crop, when fine ploughing results in bare soils with low cohesion. To take this temporal change in erodibility into account in the model, there is the option to replace the user-supplied constant cover factor on arable land (hereafter termed $C_{cov,av}$) with a dynamic factor, which changes during the course of the year (hereafter termed C_{cover}). This dynamic cover factor is calculated so that its annual average is the user-specified cover factor, to preserve the user's knowledge of the ratio in long-term erodibility between land use classes. In most temperate areas, high-risk seed beds will be present during both spring and autumn, as many farmers cultivate a mixture of spring- and autumn-sown crops. Arable land may therefore be further sub-

divided into spring-sown (e.g. spring cereals) and autumn sown (e.g. winter cereals) if the user wishes to take dynamic erodibility into account.

Two options were explored for calculating the dynamic cover factor (associated parameters and variables are defined in Table 6). The first approach uses a cosine wave (Equation 25). This is appealing in that it is smooth and only requires one user-supplied parameter, the date of maximum erodibility. In much of northern Europe, for example, this could be around the end of February (day 60) for spring-sown crops and the end of October (day 300) for autumn-sown crops. The amplitude of the curve can be fixed to ensure that the annual mean is equal to $C_{cov,av}$ (Equation 25). The downside of this formulation is that it assumes that points of maximum and minimum erodibility are half a year apart. For autumn-sown cereals in particular this is unlikely.

Variable	Description	Units	Source
C_{cover}	Erosion soil cover factor	None	Equation 26 (if dynamic) or $C_{cover,av}$
$C_{cover,av}$	Average soil cover factor for erodibility	None	Input parameter
$C_{measures}$	Erosion management factor	None	Input parameter
$d_{E,max}$	Date of maximum soil erodibility	None	Input parameter (for spring- and autumn-sown)
$d_{E,start}$	Start of the high erodibility period	None	Model calculates ($D_{E,max}$ -30)
$d_{E,end}$	End of the high erodibility period	None	Model calculates ($D_{E,max}$+30)
d_{year}	Julian day of the year	None	Model calculates
f_{spr}	Fraction of arable land that is spring sown	None	Input parameter

Table 6: Parameters and variables used in the erodibility equations

Equation 25: Dynamic crop cover factor for arable land calculated using a cosine wave, where superscript *i* indicates the crop type (spring-sown or winter-sown)

$$C_{cover}^{i} = \frac{C_{cov,av}^{i}}{2}\left(\cos\left(\frac{2\pi}{365}d_{year} + d_{Emax}^{i}\right) + 1\right)$$

The alternative approach adopted instead was to assume a constant cover factor throughout the year, apart from during a high risk period defined by a user-specified maximum erodibility date. Within this high risk period, the factor follows a triangular shape (Figure 8), which smooths the transition from lower to higher erodibility risk, to take into account differences in ploughing dates across the catchment. The difference between $C_{cov,av}$ and the dynamic cover factor outside the high risk period is calculated so that the annual average of the dynamic cover factor equals $C_{cov,av}$ (Figure 8 and Equation 26). To provide comparable simplicity to the sine curve option, the length of the high erodibility period is fixed. This period should encompass the likely variability in seed bed presence in both space and time in the catchment, whilst not being so wide as to reduce the cover factor during the remainder of the year to below the average values on improved grassland and semi-natural land – we would generally expect arable land, with its higher proportion of bare soil, compaction and tramlines, to have higher erosion risk than the other two land classes. A period of two months (60 days) provided a good compromise, although the sensitivity of the model to this factor could be assessed in the future. Within the high risk period, the value of the dynamic cover factor is then calculated by linear interpolation (Equation 26), and the results for spring- and autumn-sown crops are averaged to give the overall factor to be used in Equation 21.

Equation 26: Dynamic cover factor for arable land, C_{cover}, calculated using a triangular wave during the high risk period, where superscript *i* indicates the crop type (spring-sown or winter-sown)

If the day of the year, d_{year}, is within the period $d_{E,max}{}^{i}$ ± 30 days:

$$\text{If } d_{year} < d_{E,max}\text{: } C_{cover}^{\,i} = C_{cov,av}^{i} + \frac{\left(1 - C_{cov,av}^{i}\right)\left(d_{year} - d_{E,start}^{i}\right)}{d_{E,max}^{i} - d_{E,start}^{i}}$$

$$\text{If } d_{year} > d_{E,max}\text{: } C_{cover}^{\,i} = 1 + \frac{\left(C_{cov,av}^{i} - 1\right)\left(d_{year} - d_{E,max}^{i}\right)}{d_{E,end}^{i} - d_{E,max}^{i}}$$

Otherwise:

$$C_{cover}^{i} = C_{cov,av}^{i} - \left(\frac{60\left(1 - C_{cov,av}^{i}\right)}{2(365 - 60)}\right)$$

Averaging over arable land classes:

$$C_{cover} = f_{spr} C_{cover}^{spring} + \left(1 - f_{spr}\right) C_{cover}^{autumn}$$

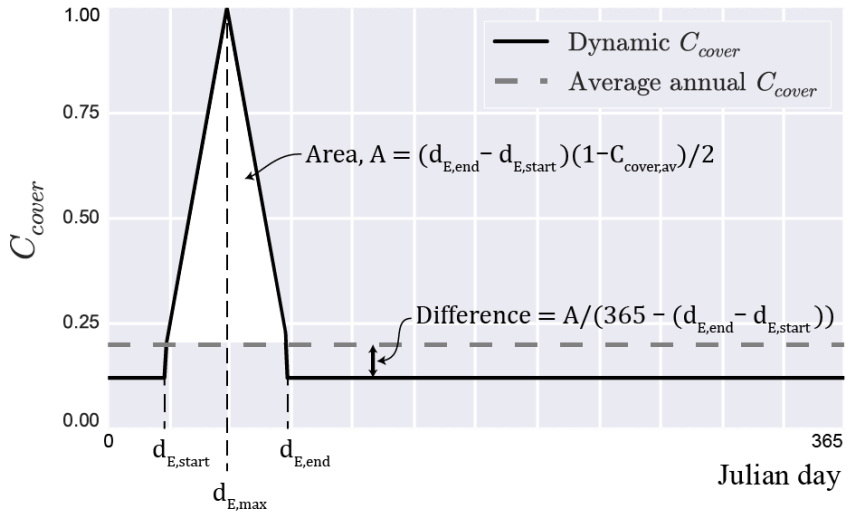


Figure 8: Schematic illustrating the shape of the dynamic crop cover factor (C_{cover}) over a year. The period of high erodibility starts on $d_{E,start}$, is at its maximum at $d_{E,max}$ and finishes at $d_{E,end}$.

Model testing, preferably in a formal statistical framework, is required to determine whether the additional complexity of the dynamic cover factor is warranted. A good place to test this would be in an area where sediment fluxes are believed to change throughout the year in response to changing arable crop cover. It would also be worth exploring how suitable the dynamic cover factor formulation is for autumn-sown crops, which are likely to have relatively bare soils throughout the winter.

4.3 Phosphorus processes

4.3.1 Soil processes

a) Overview

Soil P processes are calculated separately for two land use classes – a 'high P' class and a 'low P' class. Below, the high P class is referred to as agricultural land, and the low P class as semi-natural, but other land uses could be assigned to these classes. For particulate P fluxes, a further distinction may be made between high and low erodibility land uses in the high P class, e.g. between arable and improved grassland, to take the different erodibility (and therefore sediment and PP transport) into account (Section 4.2.1). The model also includes the ability to simulate semi-natural land newly-converted from agricultural land and vice versa (discussed further in Section 4.3.1 c).

In the model, P is present in the soil in three forms: (1) dissolved P (TDP) in the soil water, (2) labile soil P, which can take part in sorption reactions with soil water TDP, and (3) inactive soil P. The masses of dissolved and labile soil P change through time, whilst the mass of inactive soil P is constant. Parameters and variables used in the soil P equations are defined in Table 7.

Variable	Description	Units	Source
A_{SC}	Sub-catchment area	km^2	Input parameter
β	Baseflow index	None	Input parameter
dP_{labile}/dt	Rate of change of labile soil P mass with time	kg day^{-1}	Equation 33
$dTDP_s/dt$	Rate of change of soil water TDP mass with time	kg day^{-1}	Equation 34
EPC_0	Agricultural soil equilibrium P concentration of zero sorption	kg mm^{-1}	Equation 29 or $EPC_{0,user}$
$EPC_{0,user}$	User-supplied initial EPC_0 for agricultural soil	mg l^{-1}	Input parameter
f_i	Fraction of land use in each of the possible land use classes, *i*, including agricultural (A; $f_{Ar} + f_{IG}$), arable (Ar), improved grassland (IG), semi-natural (SN), and newly-converted versions of all 3 (NC_i)	None	Input parameters
K_f	Soil P adsorption coefficient	mm kg $soil^{-1}$	Equation 28
M_{soil}	Soil mass	kg	Equation 32
$M_{soil,m2}$	Soil mass per m^2 (soil depth × bulk density)	kg m^{-2}	Input parameter
$P_{inactive}$	Soil inactive P mass	kg	Equation 30
P_{labile}	Soil labile P mass	kg	Equation 33
$P_{labile,0}$	Initial soil labile P content	kg	Equation 31
$P_{netInput}$	Net annual input (or uptake) of P to the land class	kg ha^{-1} yr^{-1}	Input parameter
$P_{soilConc,A}$	Total soil P content in agricultural land as a mass ratio	mg P (kg soil)$^{-1}$	Input parameter
$P_{soilConc,S}$	Total soil P content in semi-natural land as a mass ratio	mg P (kg soil)$^{-1}$	Input parameter
Q_q	Quick flow	mm day^{-1}	Equation 5
Q_s	Soil water flow	mm day^{-1}	Equation 10
$TDP_{g,conc}$	Groundwater TDP concentration	mg l^{-1}	Input parameter
$TDP_{g,land}$	Groundwater TDP input to the reach	kg day^{-1}	Equation 37
$TDP_{q,land}$	Quick flow TDP input to the reach	kg day^{-1}	Equation 36
TDP_s	TDP mass in the soil water	kg	Equation 34
$TDP_{s,land}$	Soil water TDP input to the reach	kg day^{-1}	Equation 35
V_s	Soil water volume	mm	Equation 6

Table 7: Parameters and variables used in the soil and groundwater P equations

b) Interactions between soil P and dissolved soil water P

The labile soil P store and dissolved P in the soil water are assumed to be in equilibrium, and a simple linear relationship is used to relate soil total P concentration and EPC_0, the equilibrium TDP concentration at which there is no net sorption or desorption of P (Equation 27; Figure 9). This relationship can be conceptualised as accounting for sorption and mineralization/immobilization reactions. This linear relation cannot simulate P saturation, but because of its simplicity it is recommended for use in catchment models, as long as soil water TDP concentration is below 1 to 10 mg l^{-1} (McCray et al., 2005).

Equation 27: Relationship between soil P content ($P_{soil,conc}$; kg P kg $soil^{-1}$) and soil water TDP concentration (expressed as the equilibrium P concentration, EPC_0; kg mm^{-1})

$$P_{soil,conc} = K_f EPC_0 + 10^{-6} P_{soilConc,S}$$

Soil water in semi-natural areas tends to have low TDP concentrations (Q_{75}<5 μg l^{-1}; unpublished James Hutton Institute data), implying tight retention of any P released from the soil matrix. It is therefore assumed that semi-natural land does not contain labile soil P, that semi-natural soil water TDP concentration is zero, and that all soil P in semi-natural land is in the inactive soil P store (Equation 30). The soil P content of semi-natural land is therefore used as the y-axis intercept in Equation 27, and the inactive soil P store on agricultural land is assumed to be the same as in semi-natural land (Figure 9).

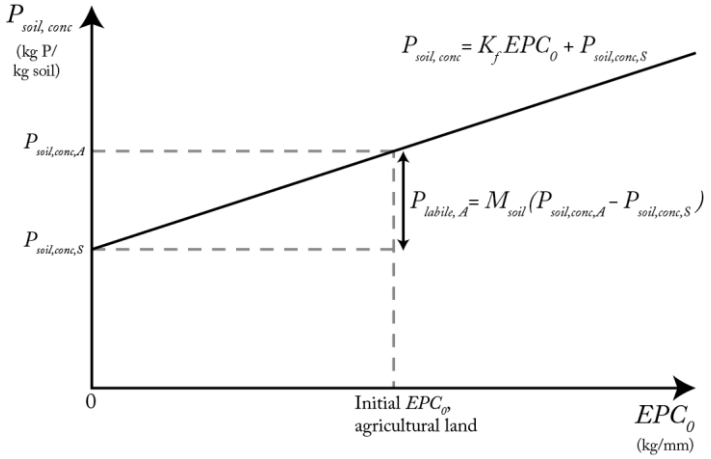


Figure 9: Illustration of the conceptual model used to link soil P content and soil water TDP concentration, expressed as EPC_0, the equilibrium TDP concentration of zero sorption.

Having defined the y-intercept, there are then two unknowns in Equation 27 for a given soil total P content: the adsorption coefficient (K_f) and EPC_0. It is very important for simulated soil water TDP concentrations to be in the right range, as in diffuse pollution-dominated systems near-surface soil water flow often controls in-stream TDP peaks. Initial agricultural soil water TDP concentration, assumed to be equivalent to soil water EPC_0, was therefore chosen to be the calibrated parameter, and K_f is calculated internally by the model using Equation 28. Once agricultural EPC_0 has been determined through calibration, K_f output by the model from the calibration period is then supplied as an input parameter for model runs in simulation mode and EPC_0 is calculated instead within the model (Equation 29). Initial EPC_0 is still an input parameter in any validation and scenario runs, but merely to provide the initial conditions for the soil water P content. This formulation removes much of the need for careful thinking and parameterisation from the modeller, thereby reducing the risk of highly inappropriate sorption equations, and therefore unrealistic simulations of the impacts of changes in fertilizer or manure inputs or land use change.

Equation 28: The adsorption coefficient, K_f (mm kg soil^{-1}), calculated during the calibration period

$$K_f = \frac{10^{-6}\left(P_{soilConc,A} - P_{soilConc,S}\right)}{EPC_0}$$

Equation 29: EPC_0, the equilibrium TDP concentration of zero sorption (kg mm^{-1} in the equation below; supplied by the user in mg l^{-1}). N.B. the user may also opt for EPC_0 to be constant

$$EPC_0 = \frac{P_{labile}}{K_f M_{soil}}$$

Having assumed that the inactive soil P store on agricultural land is equivalent to the total soil P content on semi-natural land, initial labile P content in agricultural land is then calculated as the difference between cultivated and semi-natural total P content (Equation 31). This assumption is only appropriate if soils under the two land classes have similar P sorption capacity (controlled primarily by iron oxide and clay content).

Equation 30: Inactive soil P content in agricultural and semi-natural land, $P_{inactive}$ (kg)

$$P_{inactive} = 10^{-6} P_{soilConc,S}\, M_{soil}$$

Equation 31: Initial labile soil P content in agricultural land, $P_{labile,0}$ (kg)

$$P_{labile,0} = 10^{-6}\left(P_{soilConc,A} - P_{soilConc,S}\right) M_{soil}$$

Converting from soil P concentration to mass of P in the labile and inactive stores requires an estimate of the sub-catchment soil mass (Equation 32). The user-specified areal soil mass parameter ($M_{soil,m2}$) can be

calculated by multiplying soil bulk density with an estimate of soil depth; topsoil depth is recommended, being the depth of soil which contains highest soil P concentrations. These two parameters are lumped together as one user input parameter to reduce parameter non-identifiability issues during auto-calibration.

Equation 32: Sub-catchment topsoil mass, M_{soil} (kg)

$$M_{soil} = M_{soil,m^2} 10^6 A_{SC}$$

The change in mass of P in the labile P store due to a change in TDP concentration in the soil water can then be calculated and is assumed to control the rate of change in soil labile P in agricultural land with time (Equation 33). Inputs from fertilizer and manure are all assumed to occur in liquid form and to bring about an increase in labile P mass through adsorption. Conversely, a change in labile P content may cause a change in soil water TDP concentration due to a shift in EPC_0 (Equation 29). The model also includes the option for EPC_0 to be constant over time, to simplify the model for short model runs where no shift in soil water TDP concentration is expected.

Equation 33: Rate of change in soil labile P mass, P_{labile}, with time (kg day^{-1})

$$\frac{dP_{labile}}{dt} = K_f M_{soil} \left(\frac{\mathrm{TDP}_s}{V_s} - \mathrm{EPC}_0 \right)$$

The rate of change of soil water TDP mass with time in agricultural land is controlled by the balance of input and output fluxes (Equation 34). Potential inputs are from the application of fertilizer and manure and net release of soil P. Outputs are via plant uptake, sequestration into soil P, soil water flow and quick flow. The latter is included as it is assumed that quick flow also inherits soil water TDP concentration. Fertilizer, manure and plant uptake fluxes are grouped together into a single annual P budget parameter ($P_{netInput}$; kg ha^{-1} yr^{-1}), which is then evenly applied (or subtracted, if there is net output) over the course of the year. This grouping greatly reduces the number of parameters required, and the $P_{netInput}$ parameter may be readily informed by budgeting studies or published national P surplus values (e.g. eurostat, 2013). This simplified treatment of terrestrial P inputs and outputs assumes a relatively constant TDP concentration in soil water throughout the year. Whilst this is clearly a simplification, previous modelling work has suggested the additional complexity involved in attempting to simulate the daily changes in soil water TDP concentration in response to variations in fertilizer, manure and plant uptake fluxes is not justified (Jackson-Blake et al., 2015). At present the model takes as input a single constant value. An easy future extension to the model would be to provide the option for this to be replaced by an annual time series of values.

Equation 34: Rate of change of soil water TDP mass, TDP_s, with time (kg day^{-1})

$$\frac{d\ \mathrm{TDP}_s}{dt} = \frac{100\ A_{SC}}{365} P_{netInput} - \frac{dP_{labile}}{dt} - Q_s \frac{\mathrm{TDP}_s}{V_s} - Q_q \frac{\mathrm{TDP}_s}{V_s}$$

This representation of soil P processes greatly simplifies the actual processes involved. In reality soil P is present in a continuum of interlinked states of varying extractability, and hysteresis effects are common in P transfers between the states. However, the understanding of how detailed soil chemical processes upscale to the catchment-scale is arguably not yet advanced enough for fine-scaled geochemical principles to be usefully incorporated into a catchment-scale model, and there is certainly a lack of data to constrain such processes at a catchment scale. A potential future change to the model would however be to use more detailed geochemical models or lab experiments to derive isotherm parameters for a suite of soil types, using soil properties such as Fe and Al oxide content (Dari et al., 2015).

Example soil P results for agricultural soils in the Tarland catchment are shown in Figure 10, assuming an initial labile soil P content of 585 mg kg^{-1}, an initial EPC_0 of 0.1 mg l^{-1}, soil depth of 9.5 cm and annual net P inputs of 10 kg ha^{-1} yr^{-1}. Soil water TDP concentrations are slightly higher than the EPC_0 because of P inputs to the soil water, resulting in net adsorption and a gradual increase in labile P and EPC_0 over time.

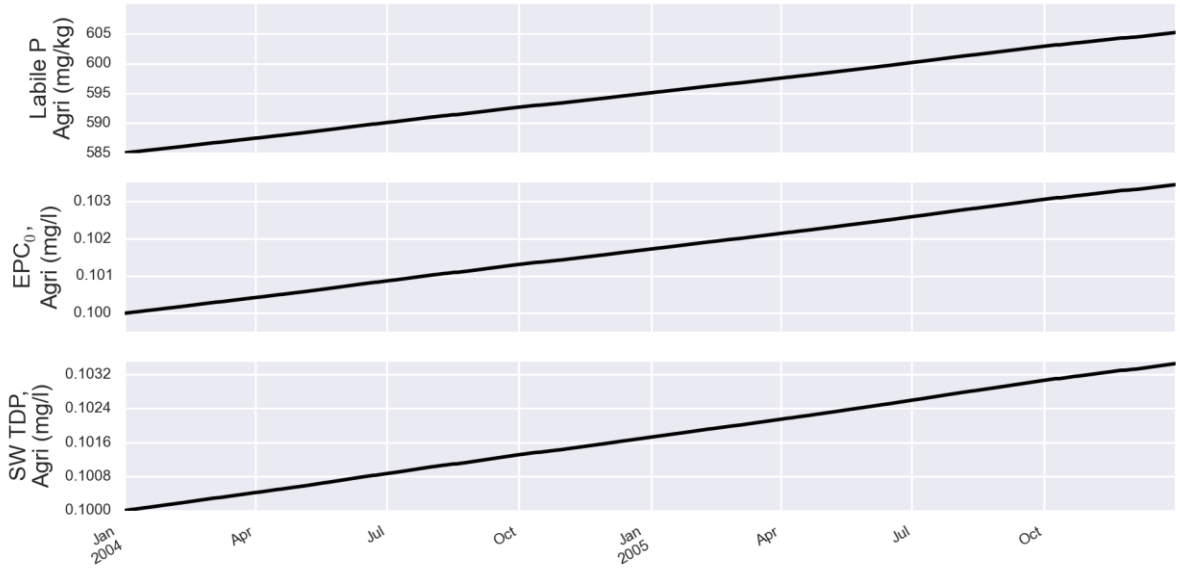


Figure 10: Simulated change in agricultural soil labile P content, soil water (SW) EPC_0 and soil water TDP concentration over a two year period, assuming a soil depth of 9.5 cm and net P inputs of 10 kg ha^{-1} $year^{-1}$.

c) Newly-converted land

A potential use of the model is to explore the impact of land use change on surface water P concentrations and loads. This land use change could be the conversion of agricultural land to semi-natural land, or vice versa. Land that has recently changed use will retain many of its previous characteristics, and cannot therefore be grouped with long-established land of the same use. Of particular importance for P and water quality is legacy soil P, the store of P that builds up in agricultural soil over years of high P application rates. Such legacy P may result in sustained high leaching and PP losses from old agricultural land potentially for several decades after reductions in fertilization (Jarvie et al., 2013b). Likewise the lack of legacy soil P in new agricultural land, recently converted from semi-natural land, may result in low P losses compared to well-established agricultural land.

To allow such effects to be simulated, a third 'newly-converted' land class was introduced into the model. Within each sub-catchment, this newly-converted land can be either semi-natural (initial labile soil P is equivalent to that on agricultural land) or agricultural (no labile soil P at the start of the model run). Two additional ODEs are then introduced: the change in labile soil P on newly-converted land with time and the change in soil water TDP with time. These take the same form as Equation 33 and Equation 34, respectively. EPC_0 is also calculated for the newly-converted class, and there is one additional user-input parameter, $P_{netInput}$. On new semi-natural land this is likely to be a negative, with net uptake of P from the soil. Otherwise, newly converted semi-natural land is grouped with existing semi-natural land, and likewise for agricultural, for other parameterisations and processes.

d) TDP flux to the stream from soil water and quick flow

The TDP input to the stream transported by soil water flow is calculated by summing the inputs from agricultural land and any newly-converted semi-natural or agricultural land (Equation 35).

Equation 35: TDP input to the reach from soil water flow, $TDP_{s,land}$ (kg day^{-1}), where *i* denotes the land use class (*A*: agricultural, *NC*: newly-converted semi-natural or agricultural)

$$\mathrm{TDP}_{s,land} = \sum_{i=A,NC} f_i(1-\beta)Q_s\left(\frac{\mathrm{TDP}_s^i}{V_s}\right)$$

The flux of TDP to the water course from the reach via quick flow is given by Equation 36. The majority of the flow pathways that make up quick flow inputs to the stream interact with the soil surface (e.g. infiltration and saturation excess flow) or are sourced from soil water (e.g. tile drainage). It was therefore assumed that quick flow inherits soil water TDP concentration. This may be conservative in areas where there is substantial runoff from farmyards, which may have TDP (and PP) concentrations which are well above those found in soil water, or if quick flow events occur directly after fertilizer or manure are applied.

Equation 36: Quick flow TDP input to the reach from the land, $TDP_{q,land}$ (kg day^{-1}), where *i* denotes the land use class (*A*: agricultural, *NC*: newly-converted semi-natural or agricultural)

$$\mathrm{TDP}_{q,land} = \sum_{i=A,NC} f_i \, Q_q \left(\frac{\mathrm{TDP}_s^i}{V_s} \right)$$

e) Incorporating soil test P data

Soil P content is usually measured for agronomic purposes, where the aim is to measure plant-available P to help choose appropriate fertilizer application rates. Many procedures exist which aim to determine plant-available soil P; indeed, there are more than ten official methods used in Europe alone (Jordan-Meille et al., 2012). This soil test P data is much more common than total soil P data, so it would be useful for the model to be able to incorporate soil test P data, rather than relying on total soil P data alone. In principle soil test P could replace labile soil P in Equation 27, as long as two additional relationships can be defined:

(1) A relationship between soil test P and total soil P, as total soil P content is still required for the PP simulation. There is always likely to be considerable scatter in this relationship due to varying soil chemistry, composition and texture. An analysis of Scottish soils data, for example, revealed a weak but significant relationship between total soil P and Modified Morgan's P: $P_{total} = 61.1\, P_{soilTest} + 529$ (n=77, R^2=0.62, p<0.001), although as is clear in Figure 11, the uncertainty in total P predictions based on this equation would be substantial.

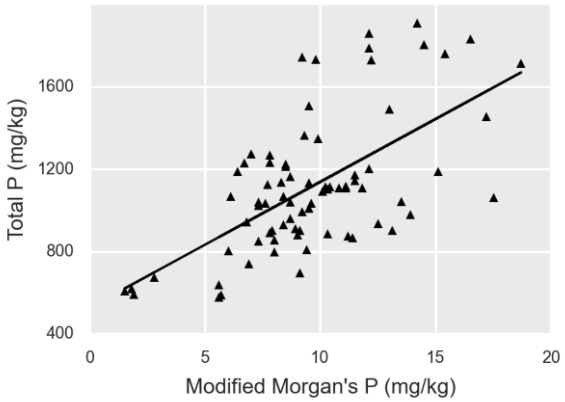


Figure 11: Relationship between soil test P (Modified Morgan's P) and total soil P from a range of Scottish soils (M. Stutter, pers. comm.)

(2) A relationship describing the proportion of net total P inputs (e.g. from fertilizer and manure minus uptake) that enters the labile, soil test P store. Model testing showed that this ratio cannot be the same as that derived from the relationship between soil test P and total soil P. If this ratio were the same, a total P to soil test P ratio of around 60 (from the Scottish soils dataset above) would imply that 1/60th of net total P inputs enter the available P store. However, with a net P input of 10 kg ha^{-1} yr^{-1} (typical for Scottish catchments), model testing showed that a ratio of <25 is required to maintain plant-available P stores at their present-day value, and that the model is extremely sensitive to the value chosen.

The fact that these two additional relationships must be independently specified causes an increase in model complexity that was not felt to be justified by an increase in model realism. In particular, if linear relationships are used to relate total soil P and soil test P, and to relate soil test P and soil water EPC_0, then by definition there will also be a linear relationship between total soil P and soil water EPC_0, meaning the increase in complexity is not justified. It was therefore decided to maintain the model structure described in Section 4.3.1 (b). However, if soil test P data are available instead of total soil P data for a given area, the user could use a linear regression between soil test P and total soil P to derive the latter (e.g. using the regression derived above for Scottish soils data). A potential future extension to the model is for this calculation to be performed internally, using user-specified regression coefficients.

4.3.2 Groundwater phosphorus

The concentration of TDP in groundwater, as in soils, depends on historic P inputs and the P sorption capacity of the aquifer matrix. The latter is often largely controlled by the iron oxide content of the aquifer, with groundwater pH and dissolved oxygen content also playing an important role (Domagalski and Johnson, 2011), whilst co-precipitation with calcium carbonate is important in calcareous aquifers. These processes substantially reduce P mobility within the sub-surface, and traditionally groundwater has therefore only been considered to have negligible TDP concentrations and to dilute surface water P concentrations. However, it is now being increasingly recognised that groundwater TDP concentrations can become elevated by anthropogenic activities (e.g. Holman et al., 2008). Some areas in which groundwater has become enriched in P, for example due to elevated agricultural P inputs or prolonged disposal of sewage effluent, may potentially become net sources of P for decades after the original source of P has been removed (Stollenwerk, 1996).

Ideally, the model would therefore include a link between soil water and groundwater TDP concentrations, taking into account aquifer and groundwater geochemistry. However, the process-understanding needed to formalise this link is arguably not yet developed enough, and certainly the data on groundwater and aquifer geochemistry is hard to come by in many areas. There are therefore good practical reasons for adopting a simpler approach which does not take this process understanding into account. In addition, changes in groundwater TDP concentration over time are likely to be slow, being buffered by potentially large groundwater residence times and sorption reactions, and smaller than associated changes in soil water and effluent fluxes which will drive any changes in groundwater state. For these reasons, a very simplistic approach is taken in the model: dissolved P transport to the stream occurs via groundwater flow, but groundwater TDP concentration is assumed to be constant through time, maintained at a user-specified value ($TDP_{g,conc}$; Equation 37. Variables are defined in Table 7).

Equation 37: Groundwater TDP input to the reach, $TDP_{g,land}$ (kg day^{-1})

$$\mathrm{TDP}_{g,land} = A_{SC} Q_g \mathrm{TDP}_{g,conc}$$

This simplistic approach is a good starting point, but is only valid where the groundwater TDP concentration is indeed unlikely to change over the course of the model run. This is likely to be the case over sub-decadal time periods, in areas with large aquifers, where the groundwater matrix is rich in Fe oxides and/or where the groundwater is oxic. Caution should otherwise be employed. This is a potential area for future model development, the ultimate aim being to derive a simple relationship between soil water and groundwater TDP concentration using readily-measurable groundwater properties. A linear relationship could be a suitable starting point. The gradient of the line could be derived by making some assumptions to help estimate the location of two points on the line: point 1 could be current measured groundwater TDP concentration, corresponding with sub-catchment averaged soil water TDP concentration. The second point would require an estimate of the 'pristine', non-anthropogenic groundwater TDP concentration (e.g. based on space-for-time substitution of measured data), associated with soil water TDP concentrations in semi-natural land, plus knowledge or assumptions of the time taken for semi-natural groundwater to reach its current state.

4.3.3 In-stream phosphorus

The in-stream P process model simulates in-stream dilution of diffuse and point source P inputs and down-stream transport. Parameters and variables used in the in-stream P equations are defined in Table 8.

Variable	Description	Units	Source
A_{SC}	Sub-catchment area	km^2	Input parameter
dPP_r/dt	Change of PP mass in the reach with time	kg day^{-1}	Equation 40
$dPP_{r,df}/dt$	Daily flux of PP from the stream reach	kg day^{-1}	Equation 41
$dTDP_r/dt$	Rate of change of TDP mass in the reach with time	kg day^{-1}	Equation 38
$dTDP_{r,df}/dt$	Daily flux of TDP from the stream reach	kg day^{-1}	Equation 41
E_{PP}	Particulate P enrichment factor	None	Input parameter
f_i	Fraction of land use in each of the possible land use classes, *i*, including agricultural (A; $f_{Ar} + f_{IG}$), arable (Ar), improved grassland (IG), semi-natural (SN), and newly-converted versions of all 3 (NC_i)	None	Input parameters
M_{input}	Sediment mass input to the reach from the land and in-stream entrainment	kg day^{-1}	Equation 20
M_{soil}	Soil mass	kg	Equation 32
$P_{inactive}$	Soil inactive P mass	kg	Equation 30
P_{labile}	Soil labile P mass	kg	Equation 33
PP_{input}	PP input to the reach from the land and in-stream entrainment	kg day^{-1}	Equation 39
PP_r	Mass of PP in the reach	kg	Equation 40
$PP_{r,conc}$	Daily mean concentration of PP in the reach	mg l^{-1}	Equation 42
$PP_{r,df}$	Mean daily flux of PP from the reach	kg day^{-1}	Equation 41
$PP_{r,US}$	PP input to the reach from upstream reaches	kg day^{-1}	Model calculates
Q_q	Quick flow	mm day^{-1}	Equation 5
Q_r	Instantaneous reach discharge	mm day^{-1}	Equation 17
$Q_{r,av}$	Daily mean reach flow	mm day^{-1}	Equation 18
TDP_{eff}	Effluent TDP input to the reach	kg day^{-1}	Input parameter
$TDP_{g,land}$	Groundwater TDP input to the reach	kg day^{-1}	Equation 37
$TDP_{q,land}$	Quick flow TDP input to the reach	kg day^{-1}	Equation 36
TDP_r	Mass of TDP in the reach	kg	Equation 38
$TDP_{r,conc}$	Daily mean concentration of TDP in the reach	mg l^{-1}	Equation 42
$TDP_{r,df}$	Mean daily flux of TDP from the reach	kg day^{-1}	Equation 41
$TDP_{r,US}$	TDP input to the reach from upstream reaches	kg day^{-1}	Model calculates
$TDP_{s,land}$	Soil water TDP input to the reach	kg day^{-1}	Equation 35
V_r	Reach volume	mm	Equation 14

Table 8: Parameters and variables used in the in-stream P equations.

a) TDP

The rate of change in instream TDP mass with time depends on the difference between input and output fluxes. A simplistic representation is used, where inputs are from the land phase (quick, soil water and groundwater flow), sewage and upstream reaches, and the only output is reach outflow (Equation 38).

Equation 38: Rate of change in reach TDP mass, TDP_r, with time (kg day^{-1})

$$\frac{d\,\mathrm{TDP}_r}{dt} = \mathrm{TDP}_{q,land} + \mathrm{TDP}_{s,land} + \mathrm{TDP}_{g,land} + \mathrm{TDP}_{eff} + \mathrm{TDP}_{r,US} - Q_r \frac{\mathrm{TDP}_r}{V_r}$$

This simple formulation assumes that instream TDP is in a state of dynamic equilibrium, i.e. instream sinks of TDP (e.g. adsorption and biological uptake) are balanced by instream sources (e.g. desorption and mineralisation). Studies have indicated that this balance between sources and sinks may in fact change through the year in response to changing concentrations (e.g. Stutter and Lumsdon, 2008). Retention is thought to be particularly important during low flows, due to biological uptake and sorption (House, 2003). Omitting TDP sinks from the model could therefore result in other P sources being over-estimated, particularly groundwater and effluent inputs. To determine whether an in-stream TDP sink is significant in a particular study catchment, data are needed to quantify the sink directly or to provide good constraints on groundwater and effluent inputs. In-stream processes may also become net sources or sinks of TDP if

equilibrium has not yet been reached or is disturbed, e.g. by a change in external loading. For example, reducing effluent P inputs may cause a reduction in water column TDP concentration, which may cause P-enriched bed sediments to release TDP until a new equilibrium state is reached (Stutter et al., 2010). By not incorporating this potential interaction between bed sediments and the water column, the model is not able to represent legacy sewage effluent P. However, as explained in Section 3.1, this legacy store is thought to be much smaller than other potential legacy P stores in the catchment, and to have a more rapid turnover time.

In many countries, soluble reactive P (SRP) is used by regulators to assess compliance with water quality standards, rather than TP or TDP. It is therefore desirable to be able to convert TDP to SRP. Models such as INCA-P do this using a simple linear regression with regression parameters supplied by the user. This approach works well in the majority or rivers, and as a starting point is recommended here. It has not been coded into the model, but the user can easily transform model output using their own regression equation. Different TDP sources may have very different SRP:TDP ratios, and a potential future extension could be to factor in different SRP:TDP ratios of different P sources.

b) PP

All PP is assumed to be sediment-bound, and so PP inputs to the water column are assumed to be proportional to sediment inputs (Equation 39). The mass of PP input to the stream is therefore simply the mass of sediment transported to the stream from each land use class, M_{input}, multiplied by the P content of the soil in that class and assuming unlimited soil P. This approach has the same limitations and caveats as the sediment equations (Section 4.2.1). It also assumes that changes in the P content of soils are instantly reflected in the composition of in-stream sediments, discounting time lags or potential in-stream PP stores. In-stream stores of P are relatively small and short-lived compared to terrestrial stores (Jarvie et al., 2013a), so it was felt that the additional complexity needed to account for these in-stream processes was not justified.

Equation 39: PP input to the reach, PP_{input} (kg day^{-1}), where superscript i denotes the land use class

$$\mathrm{PP}_{input} = E_{PP}\, M^{i}_{input} \sum_{i} f^{i} \frac{\left(P^{i}_{labile} + P_{inactive}\right)}{M_{soil}}$$

An enrichment factor, E_{PP}, represents the selective delivery to the stream of finer particles enriched in P compared to source soils (Sharpley, 1980). The use of a constant enrichment factor is a simplification; in reality there is a well-documented decrease in PP loss with increasing erosion (Radcliffe and Cabrera, 2006). This is because as runoff and erosion increase there is less particle size sorting, so P-enriched finer sediment makes up a smaller proportion of the total sediment mobilised. When assembling enrichment ratio information for the CREAMS model, Menzel (1980) concluded that a log-relationship between enrichment ratio and sediment yield appears to hold in a variety of conditions: $\ln(E_{PP}) = 2.00 - 0.16\ln(Y_{sed})$, where Y_{sed} is the sediment yield (kg ha^{-1}). A potential future extension to the model would be to investigate whether this additional complexity is warranted in some areas.

In Equation 39, PP inputs to the stream are summed over up to six land use classes: arable, improved grassland, semi-natural, and newly-converted versions of each (Section c). The sediment flux for newly-converted land corresponds to the same flux for established land. The rate of change in the mass of PP in the water column with time is then given by Equation 40. As with TDP, a simplistic representation of the associated fluxes is adopted: inputs are from the land phase and entrainment (grouped as one flux) and upstream reaches; the only output is via outflow from the reach. Sewage PP inputs are not included, as the majority of effluent P tends to be in dissolved form (Neal et al., 2005; Withers and Jarvie, 2008). As a starting point, in-stream desorption of TDP from PP is not accounted for. This is likely to be justified in rivers with relatively short residence times, but potentially less so in larger, slower-flowing rivers.

Equation 40: Rate of change in reach PP mass, PP_r, with time (kg day^{-1})

$$\frac{d\,\mathrm{PP}_r}{dt} = \mathrm{PP}_{input} + PP_{r,US} - Q_r \frac{\mathrm{PP}_r}{V_r}$$

c) Daily fluxes and mean daily concentrations of TDP and PP

Time series of total daily fluxes of TDP and PP leaving the reach are obtained by integrating the instantaneous fluxes with respect to time, starting each day with initial conditions of zero (Equation 41). These total daily fluxes are then used in the calculation of daily mean TDP and PP concentrations (Equation 42), converted from units of kg mm^{-1} to mg l^{-1} for comparison with observations.

Equation 41: Rate of change of daily flux of dissolved and particulate P from the stream reach (kg day^{-1})

$$\frac{d\mathrm{TDP}_{r,df}}{dt} = Q_r \frac{\mathrm{TDP}_r}{V_r};\ \frac{d\mathrm{PP}_{r,df}}{dt} = Q_r \frac{\mathrm{PP}_r}{V_r}$$

Equation 42: Daily mean concentrations of TDP and PP in the stream reach, $TDP_{r,conc}$ and $PP_{r,conc}$ (mg l^{-1})

$$\mathrm{TDP}_{r,conc} = \frac{\mathrm{TDP}_{r,df}}{Q_{r,av}} \frac{1}{A_{SC}};\ \mathrm{PP}_{r,conc} = \frac{\mathrm{PP}_{r,df}}{Q_{r,av}} \frac{1}{A_{SC}}$$

4.4 Summary of equations, initial conditions and input parameters

Table 9 summarises the 19 ODEs which are solved simultaneously for each reach in the catchment. For the first time step initial conditions must be supplied for each ODE. To define these, three parameters are specified by the user: in-stream flow in the top reach, total soil P content and soil water TDP concentration in the 'high P' class. If the snow module is run, initial snow depth is also required. All other initial conditions are derived from these parameters or using simple assumptions – details are provided in Table 9. Initial instream masses of SS, PP and TDP are set to 0, so a burn-in period is required (the length of burn-in depends on the residence time in the reach, but should be of the order of days – weeks).

ODE	**Equation**	**Initial conditions (first time step)**
dV_s/dt For agricultural & semi-natural	Equation 6	$V_{s,0} = V_{FC}$
dQ_s/dt For agricultural & semi-natural	Equation 10	$Q^i_{s,0} = \frac{V_s^i - V_{FC}}{T_s^i\left(1+e^{V_{FC}-V_s^i}\right)}$; where superscript i denotes the land class
dV_g/dt	Equation 11	$V_{g,0} = Q_{g,0}\, T_g$
dQ_g/dt	Equation 13	$Q_{g,0} = \beta\, Q_{s,0}$
dV_r/dt	Equation 14	$V_{r,0} = Q_{r,0}\, T_{r,0}$; where $T_{r,0} = \frac{L_{reach}}{8.64\times10^4 a Q_{r0}^{0.42}}$
dQ_r/dt	Equation 17	Reach 1: Input parameter $Q_{r0,init}$ (units converted to mm day^{-1} in model) Downstream reaches: $Q_{r,0} = Q_{r,av}$ from the upstream reach for day 1
dQ_{r_av}/dt	Equation 18	0.0
dM_{sus}/dt	Equation 19	0.0
$dM_{sus,out}/dt$	Equation 23	0.0
dP_{labile}/dt For arable & newly-converted	Equation 33	Equation 31
$dTDP_s/dt$ For agricultural & new semi-natural	Equation 34	$TDP_{s,0} = A_{SC}\, EPC_{0,user}\, V_s$
$dTDP_r/dt$	Equation 38	0.0
$dTDP_{r,out}/dt$	Equation 41	0.0
dPP_r/dt	Equation 40	0.0
$dPP_{r,out}/dt$	Equation 41	0.0

Table 9: Ordinary differential equations (ODEs) solved within the model and information on how the initial conditions are defined for the first time step.

The model requires a number of parameters which should be calculated for example using a GIS (Table 10). These parameter values are likely to be well-constrained, and therefore generally will not form part of any calibration procedure (although if they are uncertain they could be included in an uncertainty analysis).

Param	Units	Description
A_{SC}	km^2	Sub-catchment area
f_{Ar}	none	Proportion of arable or other high soil P, high erodibility land (excluding newly-converted from SN)
f_{IG}	none	Proportion of improved grassland or other high soil P, moderate erodibility land (excluding newly-converted from SN)
f_S	none	Proportion of semi-natural and other low soil P land (excluding newly-converted from agricultural)
f_{NC_Ar}	none	Proportion of newly-converted arable land (from SN)
f_{NC_IG}	none	Proportion of newly-converted improved grassland (from SN)
f_{NC_S}	none	Proportion of newly-converted semi-natural land (from agricultural)
f_{spr}	none	Proportion spring-sown crops make to total arable land area (assume rest is autumn-sown)
S_{Ar}	degrees	Mean slope of arable land in the sub-catchment
S_{IG}	degrees	Mean slope of improved grassland in the sub-catchment
S_{SN}	degrees	Mean slope of semi-natural land in the sub-catchment
L_{reach}	m	Reach length
S_{reach}	degrees	Reach slope (ideally length-weighted)

Table 10: General model parameters derived using a GIS, whose values will usually be kept constant during model calibration.

The remaining model parameters are likely to be less well constrained, and are summarised in Table 11, together with suggested default values, recommended ranges, and potential data sources that could be used to inform the parameter values. An additional model parameter, $C_{measures}$, is not a calibration parameter and is only given a value when the user wishes to investigate the impact of sediment reduction measures on in-stream SS and PP concentrations or loads. There are 23 parameters in Table 11, 24-27 when spatial variability between land use classes is taken into account. Only one of these varies by sub-catchment or reach (effluent inputs), and so model complexity will not increase substantially in larger systems compared to smaller ones. At least 8 of these model parameters are optional (before taking spatial variability into account; Table 11), so in a given setup the actual number of parameters requiring calibration may be much less than 27. Even in the most complex setup in which all 27 parameters are required, an algorithm could potentially search the entire parameter space so that all could be auto-calibrated, provided the user has a full suite of water quality observations for calibration and testing (i.e. observed discharge, suspended sediment, dissolved and particulate P concentrations under the full range of flow conditions). In addition, plausible ranges for the majority of parameters may be based on measured data or data derived from literature reviews. Only four or five parameters must be determined purely through calibration (Table 11). One of these unmeasurable parameters relates to the suspended sediment simulation (E_M); the rest are hydrology parameters.

Table 11: SimplyP model parameters, including default values, recommended ranges and possible data sources. 'Spatial' column describes whether the parameter varies spatially by land use (LU), in which case by which LU type (A: agricultural, S: seminatural, Ar: arable, IG: improved grassland), or sub-catchment/reach (SC/R). Parameters likely to be key in most settings are marked with an asterisk. Many of those without an asterisk are optional. Q is discharge

Type	Param	Units	Description	Spatial	Tarland	Default	Min	Max	Data sources
Snow	$D_{snow,0}$	mm	Initial snow depth	–	0	0	0	10000	Meteorological records
	f_{DDSM}	mm dd°C^{-1}	Degree-day factor for snow melt	–	2.74	2.74	1.6	6	Literature, e.g. USDA (2004)
Hydrology	*T_s	days	Soil water time constant	LU (A, S)	A: 2 S: 10	A: 1 S: 10	> 0	30	*Calibration*
	f_{quick}	none	Proportion of precipitation routed to quick flow	–	0.02	0.02	0	0.2	*Calibration*
	alpha	none	PET reduction factor	–	1	1	0.4	1.2	Literature, e.g. Allen et al. (1998)
	*FC	mm	Soil field capacity	–	290	300	100	400	Soils database, or from soil texture using conversion charts (e.g. Appendix, Figure A1)
	*beta	none	Baseflow index	–	0.70	0.60	0	1	Local or global databases (e.g. Beck et al., 2013)
	*T_g	days	Baseflow recession constant	–	65	65	> 0	100	May be estimated from Q data using methods of Van Dijk (2010); see Beck et al. (2013) for a global analysis
	$Q_{g,min}$	mm d^{-1}	Minimum groundwater flow	–	0.4	0.0	0	2	*Calibration*
	a	m^{-2}	Gradient of stream velocity-Q relationship	–	0.5	0.5	0.1	0.8	Empirically-derived from paired velocity and Q measurements (e.g. from flow gauging)
	Q_{r0_init}	m^3 s^{-1}	Initial in-stream Q	–	1.0	1.0	> 0	N/A	Q observations
Sediment	C_{cover}	None	Vegetation cover factor (ratio of erosion rates under the land class vs bare soil)	LU (Ar, IG, S)	A: 0.2 S: 0.021 IG: 0.09	A: 0.2 S: 0.021 IG: 0.09	0	1	(R)USLE literature, e.g. Panagos et al. (2015)
	*E_M	kg mm^{-1}	Sediment input scaling factor	–	1500	1500	0	5000	*Calibration*
	*k_M	none	Sediment input non-linear coefficient	–	2.0	2.0	1.2	3	Empirical relationship between Q and SS observations or literature (e.g. Asselman, 2000)
	$d_{maxE,spr}$	none	Julian day with max erodibility; spring-sown crops	–	60	60	1	365	Local agronomic practices
	$d_{maxE,aut}$	none	Julian day with max erodibility, autumn-sown crops	–	304	304	1	365	Local agronomic practices
Dissolved P	*$P_{soilConc}$	mg kg^{-1}	Initial total soil P content	LU (A, S)	A: 1458 S: 873	A: 1458 S: 873	0-400	>3000	Soils database. Estimate from soil test P data using an empirical relationship
	*$P_{netInput}$	kg ha^{-1} yr^{-1}	Net annual P input to the soil (negative if uptake > input); S fixed at 0	LU (A)	10	10	-30	30	Fertilizer and manure application surveys, literature for P uptake, national P balance inventories (e.g. eurostat, 2013, for EU countries)
	*$EPC_{0,init}$	mg l^{-1}	Initial soil water TDP concentration on agricultural land	LU (A)	0.1	0.1	0	2	Direct measurements, literature
	*$M_{soil,m2}$	kg m^{-2}	Soil mass per m^2, important in determining the initial soil labile P mass	–	95	100	>0	800	Soils data (bulk density and depth)
	*TDP_{eff}	kg day^{-1}	Reach effluent TDP inputs	SC/R	0.1	0	0	N/A	Water company/environment protection agency data
	*TDP_g	mg l^{-1}	Groundwater TDP concentration	–	0.02	0	0	2	Direct measurements or literature
PP	*E_{PP}	none	PP enrichment factor	–	1.6	1	1	6	Direct measurements or literature (e.g. Sharpley, 1980)

4.5 Numerical methods

The model equations summarised in Section 4.4 must be solved numerically. Simply put, this involves providing initial conditions describing the state of the system, and then using the model equations to project forward in time to predict the new state of the system at the end of the time step. This then becomes the initial condition for the next time step, and the process is repeated. This process of numerical approximation introduces errors, and minimizing these errors by formulating and solving the governing model equations in a robust way is an important part of the model development process. Indeed, Clark and Kavetski (2010) suggest that in some cases numerical errors may be larger than model structural errors. Additional benefits of a robust numerical model include a smoother objective function relating model input parameters and model output (Kavetski et al., 2006a), which may reduce model calibration difficulties by allowing powerful classical parameter analysis techniques for optimisation and uncertainty analysis to be used. Kavetski et al. (2006b) argue that many of the difficulties in hydrological modelling over the last two decades, which have prompted the development of complex parameter estimation tools, are in fact due to (1) discontinuous model structures, where sharp internal thresholds introduce non-smoothness into the objective function, and to (2) poor choice of ODE solver. An attempt was made to avoid the first of these issues by avoiding thresholds in the model equations (e.g. using continuous functions rather than logic checks in equations in Section 4.1.2b). The final part of the model-building process was then to choose an appropriate ODE solver. Three factors were taken into account: (1) whether the solver is appropriate for stiff or non-stiff equations, related to the time-stepping scheme used (see below); (2) popularity, and (3) availability.

Differential equations may be categorised as stiff or non-stiff. Generally speaking, stiff equations include some terms which can lead to rapid variation in the solution, which means that certain numerical methods for solving them are unstable unless the step size taken is excessively small in relation to the smoothness of the exact solution. Time-stepping in this context relates to the sub-steps the model time step is broken down into by the solver. Most of the classical numerical methods for solving ODEs are only suitable for non-stiff ODEs (e.g. the simple Euler method, the 4^{th} order and various adaptive Runge-Kutta methods and the multi-step Adams' method). If applied to stiff systems, these methods are likely to be inaccurate or prohibitively slow. Many ODE systems are stiff in practice, and a wide range of off-the-shelf ODE solvers are available which are able to adapt their time-stepping routine and fluctuate between using stiff or non-stiff solvers. It is therefore surprising that many catchment models continue to use simple, often unsuitable solvers (Kavetski and Clark, 2011). Here, the LSODA solver was chosen, taken from the FORTRAN ODPEPACK library (Hindmarsh, 1983). LSODA starts using a non-stiff method (an Adams predictor-corrector method) and dynamically monitors the data, if necessary switching to a stiff method (the multi-step Backward Differentiation Formula method). LSODA is both widely-used and easy to implement using the SciPy.integrate module's odeint function.

The solver's error tolerance parameters affect the precision of the result, with a smaller error tolerance resulting in more time steps and greater precision but longer run times. Testing showed an approximate log relationship between run times and the relative error tolerance parameter (*rtol*), with a decrease in run times of around 70% for an increase in *rtol* from the default of 1×10^{-8} to 0.05. *rtol* was set to 0.01, to optimise the trade-off between decrease in run time and loss in accuracy. The maximum number of within-timestep function evaluations was set to 5000, to prevent the solver reaching the maximum threshold before finding a solution within the required error tolerance, which could introduce errors.

5. Future model development priorities

A number of potential areas for model improvement are highlighted throughout Section 4 and are summarised in Table 12. Most of these suggestions involve an increase in model complexity, and before being adopted any increase in complexity needs to be justified by demonstrating a substantial increase in model performance, preferably within a statistical model comparison framework. To help prioritise areas for model improvement from within this list (or indeed to highlight other areas), the model needs to be tested in

a range of contrasting catchments, including catchments where internal processes and fluxes have been measured.

Additional general recommendations for model improvement include:

- On a practical level, the model is currently slow to run compared to professionally-coded models such as INCA-P. This may be because of the choice of ODE solver, which is sophisticated compared to the solvers employed by most water quality models, or because the model is coded in Python rather than a faster, lower-level language such as C^{++}. Other solvers should be investigated for speed.
- At present, there is no flexibility in the model in terms of the number of land use classes, which are fixed at two (for dissolved P processes) or three (for sediment and PP). This reduces the versatility of the model, and future development to increase flexibility in this regard could widen the appeal of the model. However, it would also require a re-conceptualisation of the soil P equations.
- As pointed out by Adams et al. (2016), new monitoring techniques mean that high frequency P concentration measurements are now available, e.g. at 30 minute resolution, and yet many popular water quality models are only able to simulate at a daily resolution. A simple change to the model described here would be to allow the user to specify the time step required.
- Improved representation of critical source areas, by taking spatial variability in hydrology, sediment and phosphorus sources, mobilisation processes and transport/delivery pathways into account in a fuller way.

Hydrology and snow:

- Add in a PET calculation, so that the model can be run using just precipitation and temperature as inputs.
- Refinements needed to the simple degree-day approach to simulating snow melt in areas with higher snowfall?
- Include a more detailed representation of temperature variation throughout a day in the snow melt calculation.
- Is model performance improved by adding a parameter to define a lower threshold for precipitation inputs, below which quick flow is zero?
- Should quick flow be varied by land class? If so, is there still a need for different soil water time constants in the land classes?
- Add an upper limit to the soil water volume (at the saturation capacity); water above this is routed to quick flow.
- Replace the minimum groundwater flow threshold parameter with a more process-based representation. E.g. factor in percolation from the soil when soil water level drops below field capacity, but at a reduced rate.

Sediment:

- Should the sediment equations be amended to attempt to track the store of sediment in the near-channel sources, to be able to simulate source-exhaustion?
- Soil erodibility may also be affected by soil wetness, which could be factored in to Equation 21.
- The slope factor in Equation 21 could be altered to be more representative of potential sediment source areas. E.g. the average slope of land within a certain distance of the watercourse.
- An additional factor could be introduced to Equation 21 to represent connectivity between sediment source areas and the watercourse.
- The C_{cover} factor for semi-natural land could be extended to incorporate knowledge on the factors which affect erodibility, such as grazing, burning and felling.
- More testing is required to determine whether the additional complexity of a dynamic C_{cover} factor is warranted, and if so whether the adopted approach is suitable.
- Under what circumstances should sediment deposition to the stream bed be taken into account, and how could this relatively complex process be represented in a simple way?
- Can reach sediment (and PP) inputs be split into inputs from the land versus bank erosion in a simple way? Work on sediment fingerprinting may help.
- Should a distinction be made between allochthonous & autochthonous in-stream sediment and PP?

Phosphorus:

- Add in an option for the net P input parameter to be dynamic, for example as a user-input time series. In addition, a link between soil P content and net P uptake would improve simulations of the longer term soil P dynamics (though is likely to require soil test P to be simulated).
- Geochemical models or lab experiments could inform the adsorption coefficient in Equation 27, e.g. by providing a range of parameters for a suite of soil types with a range of P sorption capacities.
- Add the ability to input soil P as soil test P rather than total soil P. This could be a user-specified linear relationship, but a more advanced geochemical representation may be required for meaningful results to be obtained.
- A dynamic PP enrichment factor could replace the constant, for example linked to discharge.
- To simulate legacy groundwater TDP, a simple link is needed between soil water and groundwater TDP concentration, predictable using readily-measurable groundwater properties.
- Add in an option for effluent inputs to be read in from a time series, rather than being constant.
- When/where is it necessary to explicitly account for in-stream TDP sinks (e.g. adsorption or biological uptake of sewage effluent P)? A simple decay factor may be sufficient.
- When/where should P desorption from the stream bed be simulated? How should this be done?
- For reaches with longer residence times, a link may be needed between in-stream PP and TDP.
- Sewage PP inputs could be added.
- More process-based representation of septic tank inputs.
- Ability to predict SRP concentrations, either using a simple linear regression between TDP and SRP, or by taking into account the TDP:SRP ratio of agricultural versus sewage effluent P inputs.

Table 12: Examples of areas for future model development.

References

Allen, R., Pereira, L., Raes, D., and Smith, M. (1998). "Crop evapotranspiration – guidelines for computing crop water requirements. FAO irrigation and drainage paper 56."

Asselman, N. E. M. (2000). Fitting and interpretation of sediment rating curves. *Journal of Hydrology* **234**, 228-248.

Bagnold, R. (1966). An approach to the sediment transport problem. *General Physics Geological Survey, Prof. paper*.

Beck, H. E., Dijk, A. I., Miralles, D. G., Jeu, R. A., McVicar, T. R., and Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research* **49**, 7843-7863.

Bowes, M. J., House, W. A., Hodgkinson, R. A., and Leach, D. V. (2005). Phosphorus–discharge hysteresis during storm events along a river catchment: the River Swale, UK. *Water Research* **39**, 751-762.

Chapra, S. C. (2008). "Surface water-quality modeling," Waveland press.

Clark, M. P., and Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research* **46**, W10510.

Colby, B. (1956). "Relationship of sediment discharge to streamflow," Rep. No. 2331-1258. US Dept. of the Interior, Geological Survey, Water Resources Division.

Croke, B. F., Andrews, F., Jakeman, A. J., Cuddy, S. M., and Luddy, A. (2006). IHACRES Classic Plus: a redesign of the IHACRES rainfall-runoff model. *Environmental Modelling & Software* **21**, 426-427.

Dari, B., Nair, V., Colee, J., Harris, W., and Mylavarapu, R. (2015). Estimation of Isotherm Parameters: A Simple and Cost-effective Procedure. *Frontiers in Environmental Science* **3**.

Dean, S., Freer, J., Beven, K., Wade, A. J., and Butterfield, D. (2009). Uncertainty assessment of a process-based integrated catchment model of phosphorus. *Stochastic environmental research and risk assessment* **23**, 991-1010.

Domagalski, J. L., and Johnson, H. M. (2011). Subsurface transport of orthophosphate in five agricultural watersheds, USA. *Journal of Hydrology* **409**, 157-171.

eurostat (2013). Agri-environmental indicator fact sheet - risk of pollution by phosphorus. *In* "European Union (EU) agri-environmental indicator fact sheets".

Fenicia, F., Kavetski, D., and Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research* **47**, n/a-n/a.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L. (2006). Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrol. Earth Syst. Sci.* **10**, 139-150.

Futter, M. N., Erlandsson, M. A., Butterfield, D., Whitehead, P. G., Oni, S. K., and Wade, A. J. (2014). PERSiST: a flexible rainfall-runoff modelling toolkit for use with the INCA family of models. *Hydrol. Earth Syst. Sci.* **18**, 855-873.

Gan, R., and Luo, Y. (2013). Using the nonlinear aquifer storage–discharge relationship to simulate the base flow of glacier- and snowmelt-dominated basins in northwest China. *Hydrol. Earth Syst. Sci.* **17**, 3577-3586.

Hindmarsh, A. C. (1983). ODEPACK, A Systematized Collection of ODE Solvers, RS Stepleman et al.(eds.), North-Holland, Amsterdam,(vol. 1 of), pp. 55-64. *IMACS transactions on scientific computation* **1**, 55-64.

Holman, I. P., Whelan, M. J., Howden, N. J. K., Bellamy, P. H., Willby, N. J., Rivas-Casado, M., and McConvey, P. (2008). Phosphorus in groundwater—an overlooked contributor to eutrophication? *Hydrological Processes* **22**, 5121-5127.

House, W. A. (2003). Geochemical cycling of phosphorus in rivers. *Applied Geochemistry* **18**, 739-748.

Jackson-Blake, L. A., Dunn, S. M., Helliwell, R. C., Skeffington, R. A., Stutter, M. I., and Wade, A. J. (2015). How well can we model stream phosphorus concentrations in agricultural catchments? *Environmental Modelling & Software* **64**, 31-46.

Jackson-Blake, L. A., and Starrfelt, J. (2015). 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. *Journal of Hydrology* **527**, 641-655.

Jarvie, H. P., Sharpley, A. N., Spears, B., Buda, A. R., May, L., and Kleinman, P. J. A. (2013a). Water Quality Remediation Faces Unprecedented Challenges from "Legacy Phosphorus". *Environmental Science & Technology* **47**, 8997-8998.

Jarvie, H. P., Sharpley, A. N., Withers, P. J. A., Scott, J. T., Haggard, B. E., and Neal, C. (2013b). Phosphorus Mitigation to Control River Eutrophication: Murky Waters, Inconvenient Truths, and "Postnormal" Science. *J. Environ. Qual.* **42**, 295-304.

Jordan-Meille, L., Rubæk, G. H., Ehlert, P. A. I., Genot, V., Hofman, G., Goulding, K., Recknagel, J., Provolo, G., and Barraclough, P. (2012). An overview of fertilizer-P recommendations in Europe: soil testing, calibration and fertilizer recommendations. *Soil Use and Management* **28**, 419-435.

Kavetski, D., and Clark, M. P. (2011). Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes* **25**, 661-670.

Kavetski, D., Kuczera, G., and Franks, S. W. (2006a). Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology* **320**, 173-186.

Kavetski, D., Kuczera, G., and Franks, S. W. (2006b). Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis. *Journal of Hydrology* **320**, 187-201.

Kinnell, P. I. A. (2010). Event soil loss, runoff and the Universal Soil Loss Equation family of models: A review. *Journal of Hydrology* **385**, 384-397.

Kleinman, P., Sharpley, A., Buda, A., McDowell, R., and Allen, A. (2011). Soil controls of phosphorus in runoff: Management barriers and opportunities. *Canadian Journal of Soil Science* **91**, 329-338.

Lefrançois, J., Grimaldi, C., Gascuel-Odoux, C., and Gilliet, N. (2007). Suspended sediment and discharge relationships to identify bank degradation as a main sediment source on small agricultural catchments. *Hydrological Processes* **21**, 2923-2933.

Leopold, L. B., and Maddock Jr, T. (1953). "The hydraulic geometry of stream channels and some physiographic implications," Rep. No. 2330-7102.

Luo, Y., Arnold, J., Allen, P., and Chen, X. (2012). Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China. *Hydrol. Earth Syst. Sci.* **16**, 1259-1267.

McCray, J. E., Kirkland, S. L., Siegrist, R. L., and Thyne, G. D. (2005). Model Parameters for Simulating Fate and Transport of On-Site Wastewater Nutrients. *Ground Water* **43**, 628-639.

Menzel, R. (1980). Enrichment ratios for water quality modeling. *CREAMS: A Field-Scale Model for Chemicals, Runoff, and Erosion from Agricultural Management Systems Conservation Research Report Number 26, May, 1980. p 486-492, 1 Fig, 2 Tab, 11 Ref.*

Merritt, W. S., Letcher, R. A., and Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software* **18**, 761-799.

Neal, C., Jarvie, H. P., Neal, M., Love, A. J., Hill, L., and Wickham, H. (2005). Water quality of treated sewage effluent in a rural area of the upper Thames Basin, southern England, and the impacts of such effluents on riverine phosphorus concentrations. *Journal of Hydrology* **304**, 103-117.

Oeurng, C., Sauvage, S., and Sánchez-Pérez, J.-M. (2010). Dynamics of suspended sediment transport and yield in a large agricultural catchment, southwest France. *Earth Surface Processes and Landforms* **35**, 1289-1301.

Panagos, P., Borrelli, P., Meusburger, K., Alewell, C., Lugato, E., and Montanarella, L. (2015). Estimating the soil erosion cover-management factor at the European scale. *Land Use Policy* **48**, 38-50.

Radcliffe, D. E., and Cabrera, M. L. (2006). "Modeling phosphorus in the environment," CRC Press.

Ratliff, L. F., Ritchie, J. T., and Cassel, D. K. (1983). Field-Measured Limits of Soil Water Availability as Related to Laboratory-Measured Properties1. *Soil Science Society of America Journal* **47**.

Renard, K. G., Foster, G. R., Weesies, G. A., and Porter, J. P. (1991). RUSLE: Revised universal soil loss equation. *Journal of soil and Water Conservation* **46**, 30-33.

Sample, J. (2015). Statistics notes for environmental modelling. GitHub.

Sharpley, A., Jarvie, H. P., Buda, A., May, L., Spears, B., and Kleinman, P. (2013). Phosphorus Legacy: Overcoming the Effects of Past Management Practices to Mitigate Future Water Quality Impairment. *J. Environ. Qual.* **42**, 1308-1326.

Sharpley, A. N. (1980). The Enrichment of Soil Phosphorus in Runoff Sediments1. *J. Environ. Qual.* **9**, 521-526.

Starrfelt, J., and Kaste, O. (2014). Bayesian uncertainty assessment of a semi-distributed integrated catchment model of phosphorus transport. *Environmental Science: Processes & Impacts* **16**, 1578-1587.

Stollenwerk, K. G. (1996). Simulation of phosphate transport in sewage-contaminated groundwater, Cape Cod, Massachusetts. *Applied Geochemistry* **11**, 317-324.

Stutter, M. I., Demars, B. O. L., and Langan, S. J. (2010). River phosphorus cycling: Separating biotic and abiotic uptake during short-term changes in sewage effluent loading. *Water Research* **44**, 4425-4436.

Stutter, M. I., Langan, S. J., Lumsdon, D. G., and Clark, L. M. (2009). Multi-element signatures of stream sediments and sources under moderate to low flow conditions. *Applied Geochemistry* **24**, 800-809.

Stutter, M. I., and Lumsdon, D. G. (2008). Interactions of land use and dynamic river conditions on sorption equilibria between benthic sediments and river soluble reactive phosphorus concentrations. *Water Research* **42**, 4249-4260.

Trimble, S. W. (2010). Streams, valleys and floodplains in the sediment cascade. *Sediment Cascades: An Integrated Approach*, 307-343.

Twarakavi, N. K. C., Sakai, M., and Šimůnek, J. (2009). An objective analysis of the dynamic nature of field capacity. *Water Resources Research* **45**, n/a-n/a.

USDA (2004). "National Enginnering Handbook, Part 630 - Hydrology, Chapter 11 (Snowmelt)." United States Department of Agriculture, Natural Resources Conservation Service.

Van Dijk, A. (2010). Climate and terrain factors explaining streamflow response and recession in Australian catchments. *Hydrology and Earth System Sciences* **14**, 159-169.

Wischmeier, W. C., and Smith, D. D. (1965). "Predicting rainfall-erosion losses from cropland east of the Rocky Mountains.," US Department of Agriculture (USDA), Washington DC.

Wischmeier, W. C., and Smith, D. D. (1978). "Predicting rainfall erosion losses - a guide to conservation planning," US Department of Agriculture (USDA), Washington DC.

Withers, P. J. A., and Jarvie, H. P. (2008). Delivery and cycling of phosphorus in rivers: A review. *Science of the Total Environment* **400**, 379-395.

Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes* **13**, 715-726.

Wolman, M., Miller, J., and Leopold, L. (1964). Fluvial processes in geomorphology. *San Francisco*.

Appendix A: Data to help with model parameterisation

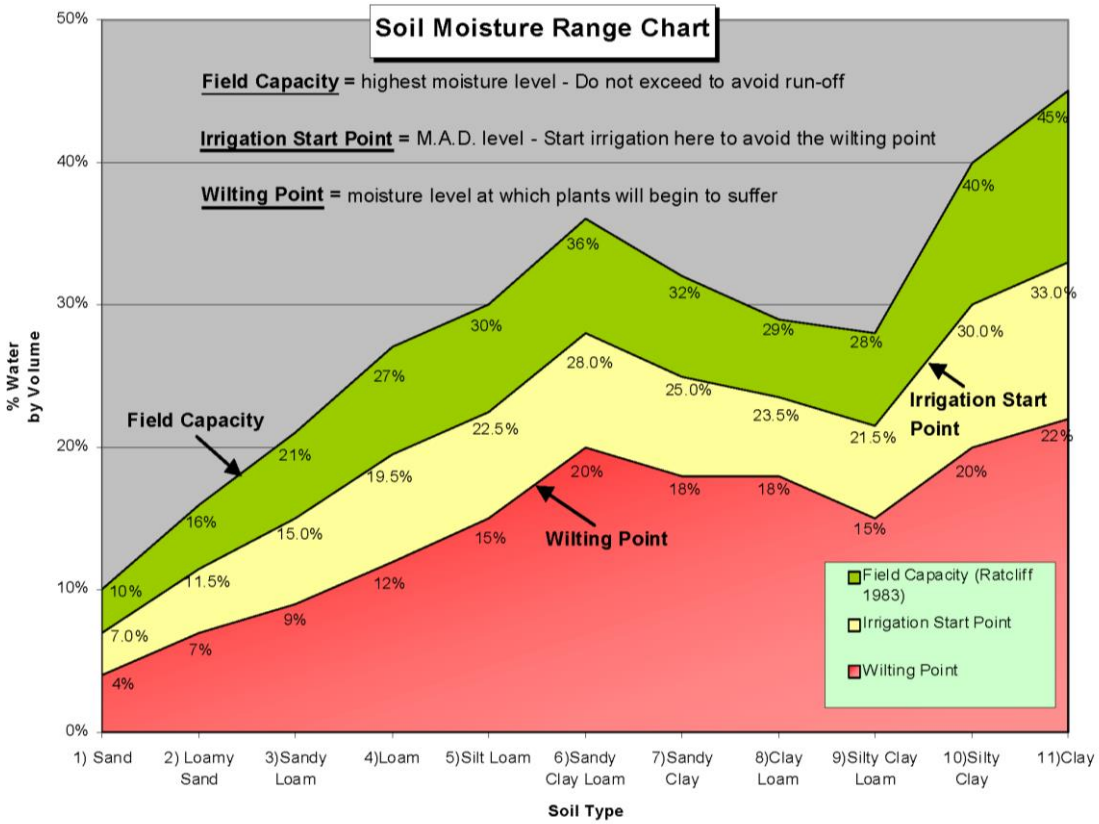


Figure A1: Chart for estimating field capacity based on soil texture, if only the latter is known.
Source: ftp://ftp.dynamax.com/turf_irrigation/Soil%20Moisture%20Range%20Chart.pdf

Crop type	C-factor
Common wheat and spelt	0.2
Durum wheat	0.2
Rye	0.2
Barley	0.21
Grain maize – corn	0.38
Rice	0.15
Dried pulses (legumes) and protein crop	0.32
Potatoes	0.34
Sugar beet	0.34
Oilseeds	0.28
Rape and turnip rape	0.3
Sunflower seed	0.32
Linseed	0.25
Soya	0.28
Cotton seed	0.5
Tobacco	0.49
Fallow land	0.5

Table A1: USLE cover factors for typical European crops (Panagos et al., 2015)

Group	Detailed class	Description	C-factor	UK mean
Permanent crops	Vineyards	Vineyards	0.15–0.45	
	Fruit trees & berry plantations	Fruit trees or shrubs: single/mixed fruit species, fruit trees with permanently grassed surfaces	0.1–0.3	
	Olive groves	Olive trees	0.1–0.3	
Pastures	Pastures	Dense graminoid grass cover of floral composition, not under a rotation system. Mainly used for grazing.	0.05–0.15	0.0867
Heterogeneous agricultural areas	Annual crops associated with permanent crops	Non-permanent crops (arable land or pasture) associated with permanent crops (<25% non-associated crops)	0.07–0.35	
	Complex cultivation patterns	Small parcels of annual crops, pasture and/or permanent crops (each occupy less than 75% of the total area)	0.07–0.2	0.1201
	Principally agriculture, significant areas of natural vegetation	Principally agricultural, with natural areas (agricultural land occupies 25 to 75% of the area)	0.05–0.2	0.1068
	Agro-forestry	Annual crops or grazing land under forested cover	0.03–0.13	
Forests	Broad-leaved, coniferous and mixed forest	Principally trees including shrub and bush understories	0.0001–0.003	0.0011
Scrub and/or herbaceous vegetation	Natural grasslands	Low productivity grassland, often on rough and uneven ground	0.01–0.08	0.0319
	Moors and heathland	Low and closed cover dominated by bushes, shrubs and herbaceous plants	0.01–0.1	
	Sclerophyllous vegetation	Bushy sclerophyllous vegetation, including maquis (dense, shrubby) and garrige	0.01–0.1	
	Transitional woodland-shrub	Bushy or herbaceous vegetation with scattered trees	0.003–0.05	0.0183
Open spaces with little or no vegetation	Beaches, dune, sands	Beaches, dunes and expanses of sand/pebbles. Coastal or continental	0	
	Bare rocks	Scree, cliffs, rocks and outcrops	0	
	Sparsely vegetated areas	Includes steppes, tundra, badlands. Scattered high-altitude vegetation	0.1–0.45	0.1825
	Burnt areas	Areas affected by recent fires, still mainly black	0.1–0.55	
	Glaciers and perpetual snow	Land covered by glaciers or permanent snowfields	0	

Table A2: USLE cover factors collated for European land cover classes (Panagos et al., 2015)