AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2016WR020132

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

Supporting Information:
- Supporting Information S1

Correspondence to:
L. Jackson-Blake,
ljb@niva.no

Citation:
Jackson-Blake, L. A., J. E. Sample, A. J. Wade, R. C. Helliwell, and R. A. Skeffington (2017), Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P, *Water Resour. Res.*, *53*, doi:10.1002/2016WR020132.

Received 16 NOV 2016
Accepted 22 MAY 2017
Accepted article online 26 MAY 2017

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

L. A. Jackson-Blake[1,2], J. E. Sample[1,2], A. J. Wade[3], R. C. Helliwell[2], and R. A. Skeffington[3]

[1]Norwegian Institute for Water Research, Oslo, Norway, [2]The James Hutton Institute, Aberdeen, UK, [3]Department of Geography and Environmental Science, University of Reading, Reading, UK

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, while 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 among the water quality modeling 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.

Plain Language Summary Catchment models may be useful tools for managing water quality, for example for exploring how water quality may change in the future under different land management, land use or climate. However, models are only useful if they capture the right processes, otherwise there is a risk of management decisions being based on unreliable information. There is now a growing awareness that many catchment water quality models used today are too complex. This makes it difficult, time-consuming and expensive to set models up, and reduces the reliability of their predictions. We have therefore developed a new, simple model to predict phosphorus concentrations in rivers, one of the biggest causes of troublesome algal blooms in fresh waters. The simple model was compared with one of the standard models in common use and was found to perform as well, despite being substantially simpler to set up and use. This supports the idea that current water quality models are too complex, and that modelers need to put more effort into assessing whether they are using appropriate tools.

1. Introduction

Dynamic, process-based 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 modeling frameworks [*Jakeman and 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 modeling studies used just five models: SWAT [*Arnold and Fohrer*, 2005; *Arnold et al*., 1998], INCA [*Wade et al*., 2002a, 2002b; *Whitehead et al*., 1998], AnnAGNPS [*Binger and Theurer*, 2005; *Young et al*., 1989], HSPF [*Bicknell et al*., 2001; *Donigian et al*., 1995], and HBV, now superseded by HYPE [*Lindström et al*., 2010]. These models are semidistributed, 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 and 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 and 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 and 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 and 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 modeling approaches.

There are sound theoretical reasons for why simpler models should be chosen over more complex ones (see for example *MacKay* [2003], chapter 28). With over-parameterized models, different parameter sets will give almost identical fits to the calibration data [e.g., *Beven and 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. First, 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]. Second, 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 autocalibration, 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 prerequisite 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 and 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 data sets, 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 modeling 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 for 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 subcatchment 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 seminatural 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 μg L^{-1}, respectively.

3. Modeling Approach

3.1. INCA-P

INCA-P is a process-based, semidistributed 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*., 2002b] and has since undergone several phases of redevelopment. 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 version 1.4.4.

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), while 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, while plant uptake is also dependent on soil moisture and season. Sediment delivery to the stream uses equations from INCA-sed [*Jarritt and 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

10.1002/2016WR020132

therefore affected by the same processes. In-stream, the model includes effluent inputs, water abstractions, sediment settling and resuspension, bank erosion, P sorption reactions in the water column and in the streambed and biological uptake of P from the water column and the streambed. 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 fourth-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 supporting information Text S1, and only a brief description is provided below.

A full description of model aims and scope is given in supporting information 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 and Starrfelt*, 2015; *Jackson-Blake et al*., 2016]. These provided recommendations for model improvements, some of which were incorporated into INCA-P (version 1.4 onward). However, one of the main recommendations was for model simplification. While 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, while 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 semidistributed, so the catchment may be split into subcatchments and associated reaches. The model is run for each subcatchment in turn, and outputs from each reach are fed into the main stem sequentially down-stream. Land may be further subdivided 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 subdivided to account for differences in erodibility (for example into 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 supporting information section 3.2).

A full description of SimplyP model processes is given in supporting information section 4. In brief, the following sets of processes are included: snow accumulation and melt (supporting information section 4.1.1); rainfall-runoff (supporting information section 4.1.2); in-stream hydrology (supporting information section 4.1.3); sediment delivery to the watercourse and in-stream transport (supporting information section 4.2); and terrestrial and in-stream P processes (supporting information 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 base

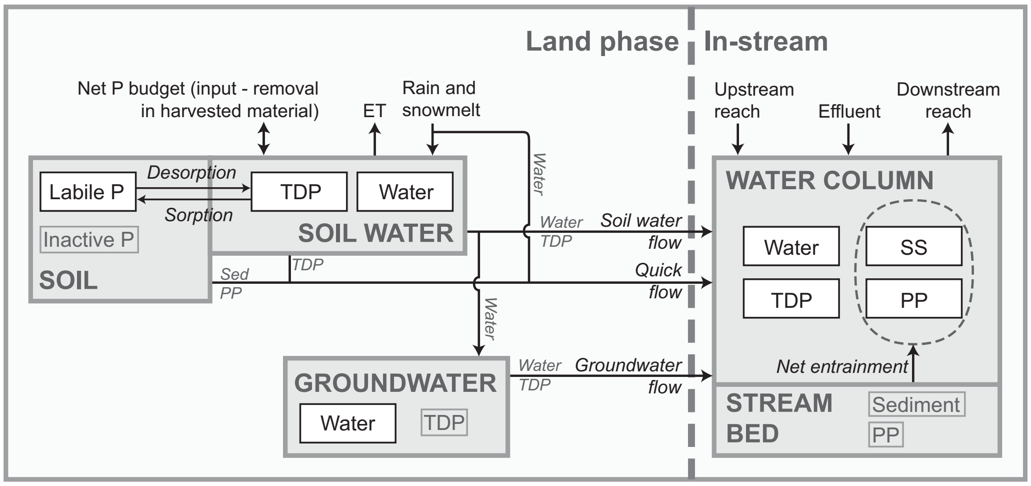

Figure 1. Schematic of the main stores, processes, and pathways included in the model. White boxes show the state variables whose volume (water) or mass (sediment, P species) is tracked. Variables within small gray boxes are implicitly included in the model, but are not tracked. Arrows show fluxes within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

flow 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 supporting information 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:

$$SS = E_{sus} {Q_r}^k \tag{1}$$

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 subcatchment 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 and 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, while 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; for example 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 supporting information 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 supporting information 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, while 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, such that in-stream sinks and sources are balanced. While 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 supporting information 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 supporting information 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 min 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 subcatchments (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 seminatural, although when calculating fertilizer and manure inputs to agricultural land, area-weighted inputs from a finer-representation of the land use were used. Seminatural land incorporated rough grazing, heather moorland, and deciduous and coniferous woodland. For more details, including percent land use per subcatchment, see *Jackson-Blake et al*. [2016].

For SimplyP, the catchment was considered as a single unit, rather than being separated into subcatchments 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 seminatural land (50%).

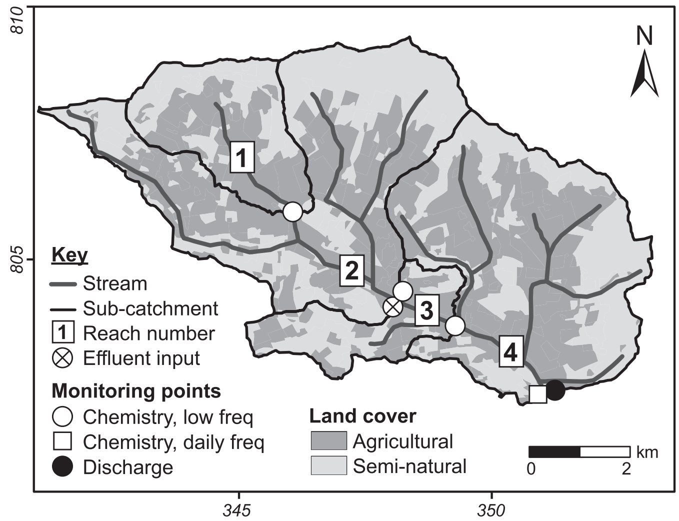

Figure 2. The Tarland catchment, with simplified land use, the location of sewage treatment effluent inputs and monitoring points, and the subcatchments used in the INCA-P application. Eastings and northings (km) are relative to the British National Grid.

Input data to drive both models included air temperature and precipitation, derived from the UK Met Office 5 km gridded data set, 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 eight parameters, the majority based on soil properties.

3.5. Model Calibration and Validation and Scenario Analysis

Ideally, a single autocalibration 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 autocalibration of all uncertain parameters is not possible—the parameter space is too large to be explored within reasonable time scales [e.g., *Jackson-Blake and Starrfelt*, 2015], and the choice of which subset of parameters to include in the analysis is subjective. Autocalibration often does not lead to improved model performance compared to manual calibration [*Boyle et al*., 2000], particularly for dynamic variables such as phosphorus—previous autocalibration of INCA-P in the study catchment resulted in less realistic simulated TDP dynamics than manual calibration [*Jackson-Blake and 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 autocalibration using a sophisticated MCMC algorithm [*Jackson-Blake and 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 stepwise 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 streambed sediment and PP during the model run.

Table 1. SimplyP Model Parameters, Including Default Values, Recommended Ranges and Possible Data Sources[a]

Type	Parameters[b]	Units	Description	Spatial[c]	Tarland	Default	Min	Max	Data Sources
Snow	$D_{snow,0}$	mm	Initial snow depth		0	0	0	10,000	Meteorological records
	f_{DDSM}	mm dd°C^{-1}	Degree-day factor for snow melt		2.74	2.74	1.6	6	Literature, e.g., U.S. Department of Agriculture (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., Supporting Information Appendix, Figure 1)
	*beta	None	Base flow index		0.70	0.60	0	1	Local or global databases [e.g., *Beck et al.*, 2013]
	*T_g	days	Base flow 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, top reach		1.0	1.0	> 0	N/A	Q observations
Sediment	C_{cover}	None	Vegetation cover factor (ratio of erosion rates under the land class versus 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 nonlinear 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 (see 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 areal mass, important in determining the initial soil labile P mass		95	100	>0	800	Soils data (bulk density and depth of P-rich soil)
	*TDP_{eff}	kg d^{-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]

[a]Q is discharge.
[b]Parameters likely to be key in most settings are marked with an asterisk. Many of those without an asterisk are optional.
[c]Describes whether the parameter varies spatially by land use (LU), and in which case by which LU type (A: agricultural, S: seminatural, Ar: arable, IG: improved grassland), or by sub-catchment/reach (SC/R).

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 seminatural values with a half-life of 7–9 years [*McCollum*, 1991; *Syers et al.*, 2008], so, 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; Table 1. Soil depths in the range 7–10 cm were obtained, 14–20 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 streambed.

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 subcatchment/reach, SimplyP has 13 ODEs (supporting information Table 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 supporting information Table 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 eight of these are optional (before taking

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 subcatchment and/or reach	64	1
Not measurable (purely calibrated)	43	4 or 5
Additional parameters held constant over land use/reach/subcatchment	12	4
Parameters used in the study catchment	48 (additional 45 to simplify the setup by removing processes)	22

spatial variability into account), depending on the level of process-representation desired. By comparison, INCA-P has 146 parameters (assuming one reach and subcatchment, varying three parameters by two land classes for dissolved P processes and by three classes for soil erodibility for comparability with SimplyP). The version of INCA-P described in the tests performed here (four reaches with three land classes) involves around 48 calibrating parameters, which we believe is simple compared to previous applications [e.g., *Crossman et al.*, 2013; *Jin et al.*, 2015]. A further ~45 parameters were carefully assigned values to help simplify the model structure and turn certain processes off, leaving around 53 unused parameters.

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

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

4.2. Model Performance in the Tarland Burn

4.2.1. Model Calibration and Validation Results

During the calibration period, discharge performance statistics were slightly better for SimplyP than for INCA-P (Table 3). This is in part because a snow accumulation and melt routine was included in SimplyP and not in the simple model used to generate HER input for INCA-P, although this only affects a few winter flow peaks (Figure 4). Small discharge peaks were also often slightly better simulated by SimplyP, particularly during base flow, which may not have been the case had a more complex hydrology model been used to generate input for INCA-P. The models performed similarly for SS, which is noteworthy given the dramatically simpler process-representation in SimplyP. Slightly improved SS performance statistics for SimplyP are likely to be due to the slightly better discharge simulation. The TP and PP simulations are also similar: output from SimplyP is less biased than output from INCA-P and the distribution of the simulated PP data in particular is closer to that of the observations (Figure 3). Both variables however have a lower correlation coefficient than INCA-P output. The story is clearer for the TDP simulation, with SimplyP scoring higher in all model performance statistics during the calibration period (Table 3), and also producing distributions of

Table 3. Model Performance Statistics in the Calibration and Validation Period

		Calibration					Validation				
Variable	Model	*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

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

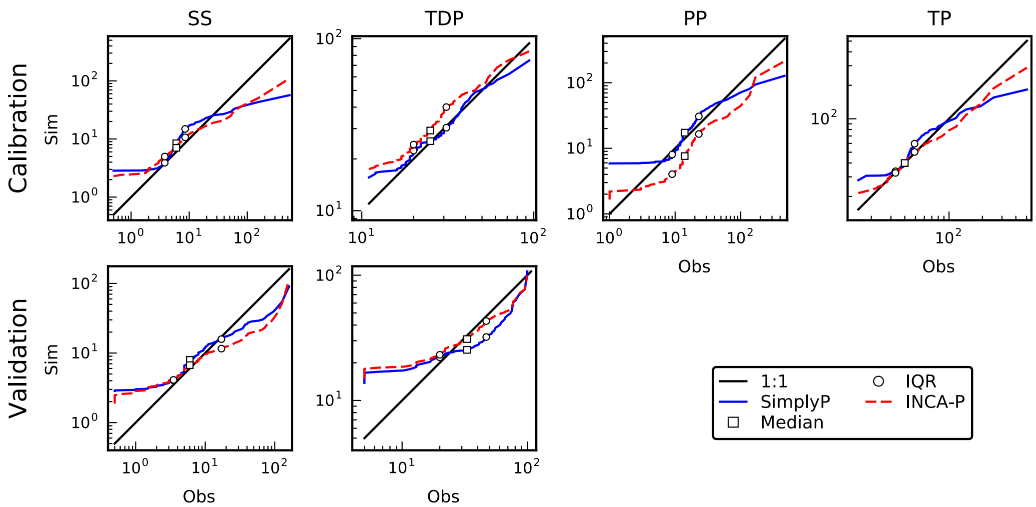

Figure 3. Q-Q plots for the calibration and validation periods. Quantiles of the simulated data are plotted against corresponding quantiles of the observed data; if observed and simulated data are from similar distributions, points will lie close to the 1:1 line. Median and interquartile ranges (IQR)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

simulated data that were more comparable to the observations (Figure 3). The slightly improved TDP dynamics may be related to the better discharge simulation. However, an improvement in simulated TDP is 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 way the ODEs are formulated or solved in INCA-P (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), while SS had lower bias but a smaller correlation coefficient (Table 3). There has been a shift in base flow 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 base flow, 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 base flow, and even small differences in simulated base flow 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 and Sudheer*, 2008; *Schaefli and 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].

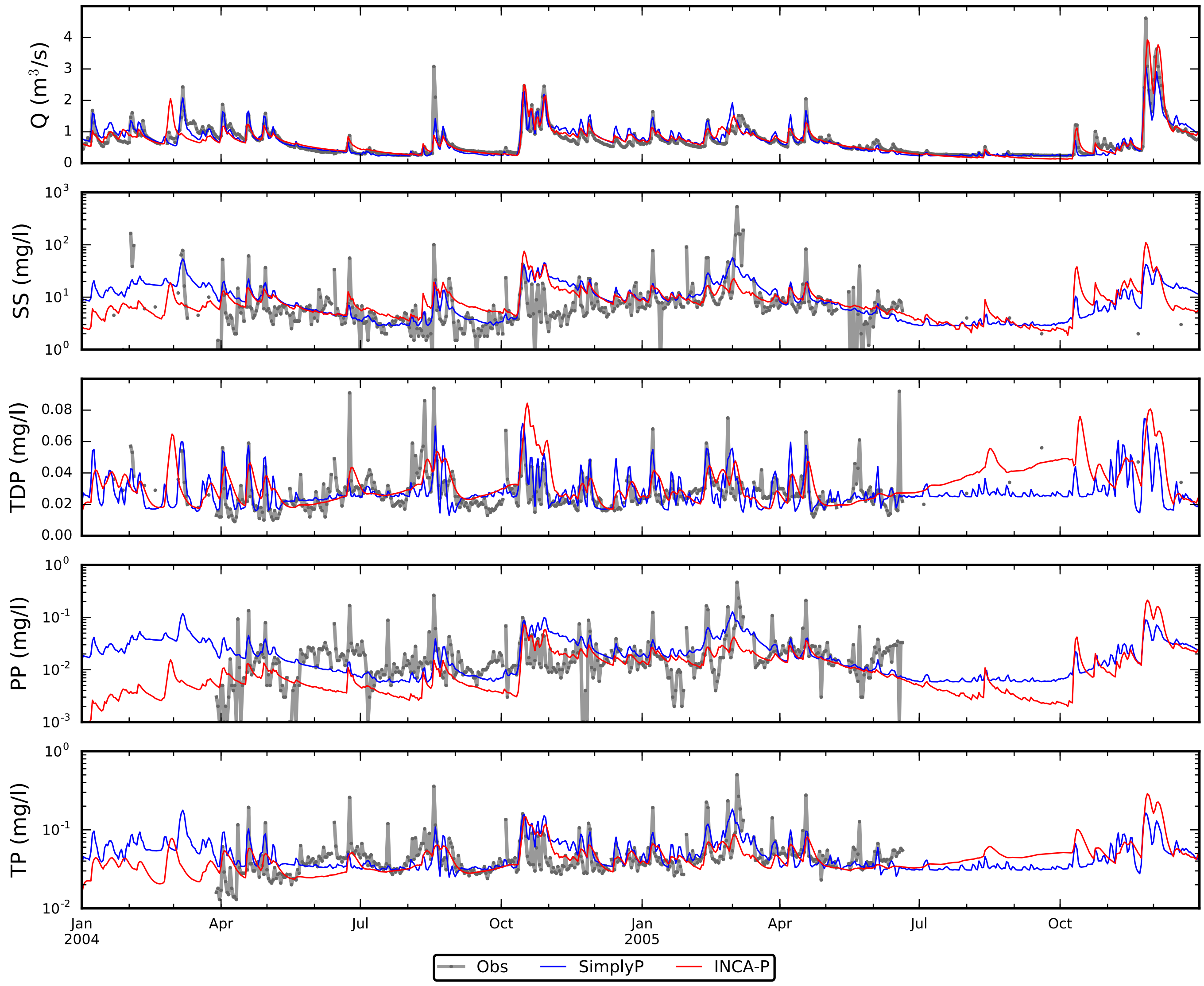

Figure 4. Time series of observed and simulated discharge (Q) and water quality during the calibration period. Note the log scales for SS, PP, and TP.

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.

While 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

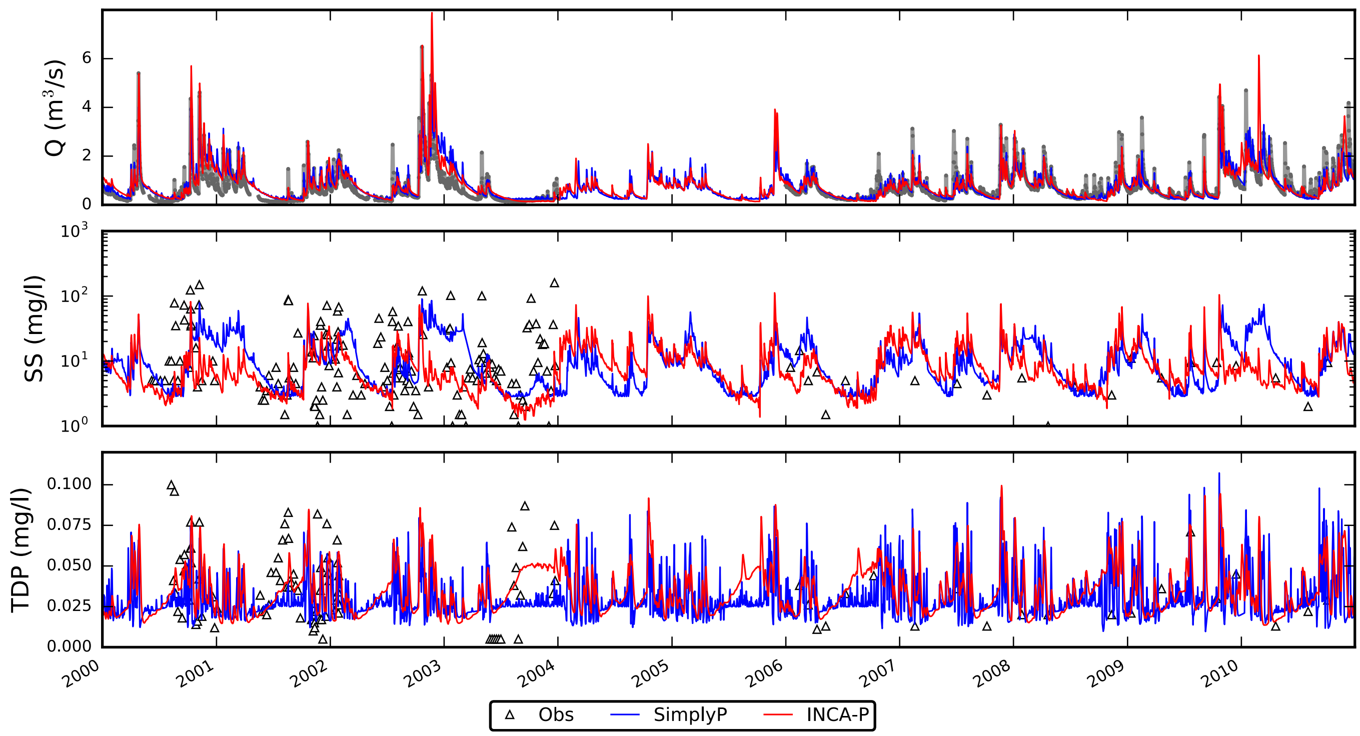

Figure 5. Time series of observed and simulated discharge (Q) and water quality during the validation period. Note the log scales for SS, PP, and TP.

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 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, while INCA-P has around 148. Only 4 or 5 SimplyP parameters are "free," in that they 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, while 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

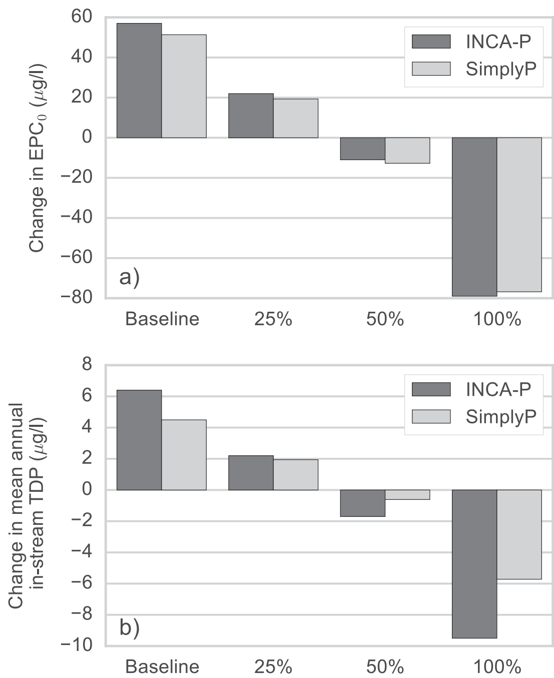

Figure 6. Simulated changes over a 30 year period in (a) agricultural soil water EPC_0 (the equilibrium phosphorus concentration, closely linked to soil water TDP concentration), and (b) in-stream mean TDP concentration. Results are shown for baseline "business as usual" inputs and for three fertilizer and manure reduction scenarios.

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. First, there has been a desire to include process-understanding and data derived from plot-scale studies. However, nonlinearities between small-scale and catchment-scale processes mean up-scaling is often inappropriate [e.g., *Kirchner*, 2006; *Oreskes and 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 land management and climate. Second, there has been a desire to make models versatile and widely applicable, aiming for "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 semidistributed 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 the supporting information (section 5 and Table 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 among the water quality modeling community as to whether today's models are appropriate and fit for purpose.

Supporting Information References

The description of SimplyP in the supporting 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 and Kavetski*, 2010; *Croke et al.*, 2006; *Dari et al.*, 2015; *Domagalski and Johnson*, 2011; *Fenicia et al.*, 2011; *Gan and Luo*, 2013; *Hindmarsh*, 1983; *Holman et al.*, 2008; *House*, 2003; *Jarvie et al.*, 2013a; *Jarvie et al.*, 2013b; *Jordan-Meille et al.*, 2012; *Kavetski and Clark*, 2011; *Kavetski et al.*, 2006a; *Kavetski et al.*, 2006b; *Kleinman et al.*, 2011; *Lefrançois et al.*, 2007; *Leopold and Maddock* 1953; *Luo et al.*, 2012; *Menzel*, 1980; *Merritt et al.*, 2003; *Neal and Jarvie*, 2005; *Oeurng et al.*, 2010; *Radcliffe and 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 and Smith*, 1965; *Wischmeier and Smith*, 1978, *Wittenberg*, 1999; *Wolman et al.*, 1964].

Acknowledgments

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 modeling. 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 supporting information section 2 for more details. To obtain an INCA-P executable, email rmc@niva.no.

References

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

Allen, R., L. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration—Guidelines for computing crop water requirements, *FAO Irrigation and drainage paper 56. FAO, Rome*, *300*(9), D05109.

Arnold, J. G., and N. Fohrer (2005), SWAT2000: Current capabilities and research opportunities in applied watershed modelling, *Hydrol. Processes*, *19*, 563–572.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998), Large area hydrologic modeling and assessment. Part I: Model development, *J. Am. Water Resour. Assoc.*, *34*, 73–89.

Asselman, N. E. M. (2000), Fitting and interpretation of sediment rating curves, *J. Hydrol.*, *234*, 228–248.

Bagnold, R. (1966), An approach to the sediment transport problem from general physics, *U.S.Geol. Surv. Prof. Pap. 422-I*.

Beck, H. E., A. I. Dijk, D. G. Miralles, R. A. Jeu, T. R. Mcvicar, and J. Schellekens (2013), Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, *Water Resour. Res.*, *49*, 7843–7863, doi:10.1002/2013WR013918.

Bergfur, J., B. O. L. Demars, M. I. Stutter, S. J. Langan, and N. Friberg (2012), The tarland catchment initiative and its effect on stream water quality and macroinvertebrate indices, *J. Environ. Qual.*, *41*, 314–321.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298.

Bicknell, B. R., J. C. Imhoff, J. L. Kittle Jr., T. Jobes, A. Donigian Jr., and R. Johanson (2001) *Hydrological Simulation Program-Fortran: HSPF Version 12 User's Manual*, AQUA TERRA Consult., Mountain View, Calif.

Binger, R., and F. Theurer (2005), *AnnAGNPS Technical Processes Documentation, Version 3.2*, USDA-ARS, Natl. Sediment. Lab., Oxford.

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

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*, 3663–3674.

Chapra, S. C. (2008), *Surface Water-Quality Modeling*, Waveland Press, Long Grove, Ill.

Clark, M. P., and D. Kavetski (2010), Ancient numerical daemons of conceptual hydrological modeling: 1: Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, *46*, W10510, doi:10.1029/2009WR008894.

Colby, B. (1956), *Relationship of Sediment Discharge to Streamflow*, *Open File Rep.*, *56–27*, 170 pp.

Couture, R.-M., K. Tominaga, J. Starrfelt, S. J. Moe, Ø. Kaste, and R. F. Wright (2014), Modelling phosphorus loading and algal blooms in a Nordic agricultural catchment-lake system under changing land-use and climate, *Environ. Sci. Processes Impacts*, *16*, 1588–1599.

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

Crossman, J., M. N. Futter, S. K. Oni, P. G. Whitehead, L. Jin, D. Butterfield, H. M. Baulch, and P.J. Dillon (2013), Impacts of climate change on hydrology and water quality: Future proofing management strategies in the Lake Simcoe watershed, Canada, *J. Great Lakes Res.*, *39*, 19–32.

Dari, B., V. Nair, J. Colee, W. Harris, and R. Mylavarapu (2015), Estimation of isotherm parameters: A simple and cost-effective procedure, *Front. Environ. Sci.*, *3*, 70.

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

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

Donigian, A., Jr., B. Bicknell, J. Imhoff, and V. Singh (1995), Hydrological Simulation Program-Fortran (HSPF), in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 395–442, Water Resour. Publ., Highlands Ranch, Colo.

Eurostat (2013), Agri-environmental indicator fact sheet—Risk of pollution by phosphorus, in *European Union (EU) Agri-Environmental Indicator Fact Sheets*. [Available at http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_risk_of_pollution_by_phosphorus, accessed 17 Feb. 2016.]

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

Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resour. Res.*, *47*, W11511, doi:10.1029/2011WR010748.

Gan, R., and Y. Luo (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.

Hahn, C., V. Prasuhn, C. Stamm, P. Lazzarotto, M. Evangelou, and R. Schulin (2013), Prediction of dissolved reactive phosphorus losses from small agricultural catchments: Calibration and validation of a parsimonious model, *Hydrol. Earth Syst. Sci.*, *17*, 3679–3693.

Hindmarsh, A. C. (1983), ODEPACK, A systematized collection of ODE solvers, in *IMACS Transactions on Scientific Computation*, vol. 1, edited by R. S. Stepleman et al., pp. 55–64, North-Holland, Amsterdam.

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

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

Jackson-Blake, L. A., and J. Starrfelt (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, *J. Hydrol.*, *527*, 641–655.

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

Jackson-Blake, L. A., A. J. Wade, and M. N. Futter (2016) The INtegrated CAtchment model of Phosphorus dynamics (INCA-P): Description and demonstration of new model structure and equations, *Environ. Modell. Software*, *83*, 356–386.

Jain, S. K., and K. Sudheer (2008), Fitting of hydrologic models: A close look at the Nash–Sutcliffe index, *J. Hydrol. Eng.*, *13*, 981–986.

Jakeman, A. J., and G. M. Hornberger (1993) How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, *29*, 2637–2649.

Jakeman, A. J., and R. A. Letcher (2003), Integrated assessment and modelling: Features, principles and examples for catchment management, *Environ. Modell. Software*, *18*, 491–501.

Jakeman, A. J., R. A. Letcher, and J. P. Norton (2006), Ten iterative steps in development and evaluation of environmental models, *Environ. Modell. Software*, *21*, 602–614.

Jarritt, N. P., and D. S. L. Lawrence (2007), Fine sediment delivery and transfer in lowland catchments: Modelling suspended sediment concentrations in response to hydrological forcing, *Hydrol. Processes*, *21*, 2729–2744.

Jarvie, H. P., A. N. Sharpley, B. Spears, A. R. Buda, L. May, and P. J. A. Kleinman (2013a), Water quality remediation faces unprecedented challenges from "Legacy Phosphorus," *Environ. Sci. Technol.*, *47*, 8997–8998.

Jarvie, H. P., A. N. Sharpley, P. J. A. Withers, J. T. Scott, B. E. Haggard, and C. Neal (2013b), Phosphorus mitigation to control river eutrophication: Murky waters, inconvenient truths, and "Postnormal" science, *J. Environ. Qual.*, *42*, 295–304.

Jin, L., P. G. Whitehead, S. Sarkar, R. Sinha, M. N. Futter, D. Butterfield, J. Caesar, and J. Crossman (2015), Assessing the impacts of climate change and socio-economic changes on flow and phosphorus flux in the Ganga river system, *Environ. Sci. Processes Impacts*, *17*, 1098–1110.

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

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

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

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

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

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

Kleinman, P., A. Sharpley, A. Buda, R. Mcdowell, and A. Allen (2011), Soil controls of phosphorus in runoff: Management barriers and opportunities, *Can. J. Soil Sci.*, *91*, 329–338.

Krueger, T., J. Freer, J. N. Quinton, and C. J. A. Macleod (2007), Processes affecting transfer of sediment and colloids, with associated phosphorus, from intensively farmed grasslands: A critical note on modelling of phosphorus transfers, *Hydrol. Processes*, *21*, 557–562.

Lazar, A. N., D. Butterfield, M. N. Futter, K. Rankinen, M. Thouvenot-Korppoo, N. Jarritt, D. S. Lawrence, A. J. Wade, and P. G. Whitehead (2010), An assessment of the fine sediment dynamics in an upland river system: INCA-Sed modifications and implications for fisheries, *Sci. Total Environ.*, *408*, 2555–2566.

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

Leopold, L. B., and T. Maddock Jr. (1953), The hydraulic geometry of stream channels and some physiographic implications, *U.S. Geol. Surv. Prof. Pap.*, *252*, 57 pp.

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

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

Mackay, D. J. (2003) *Information Theory, Inference and Learning Algorithms*, Cambridge Univ. Press, New York.

Martin-Ortega, J., et al. (2015), A transdisciplinary approach to the economic analysis of the European Water Framework Directive, *Ecol. Econ.*, *116*, 34–45.

Mccollum, R. E. (1991), Buildup and decline in soil phosphorus: 30-Year trends on a typic umprabuult, *Agron. J.*, *83*, 77–85.

Mccray, J. E., S. L. Kirkland, R. L. Siegrist, and G. D. Thyne (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, in *CREAMS: A Field-Scale Model for Chemicals, Runoff, and Erosion from Agricultural Management Systems*, *Conserv. Res. Rep. 26*, edited by W. G. Knisel, pp. 486–492, U.S. Dep. Agric., Washington, D. C.

Merritt, W. S., R. A. Letcher, and A. J. Jakeman (2003), A review of erosion and sediment transport models, *Environ. Modell. Software*, *18*, 761–799.

Messiga, A. J., N. Ziadi, D. Plénet, L. E. Parent, and C. Morel (2010), Long-term changes in soil phosphorus status related to P budgets under maize monoculture and mineral P fertilization, *Soil Use Manage.*, *26*, 354–364.

Mooij, W. M., et al. (2010), Challenges and opportunities for integrating lake ecosystem modelling approaches, *Aquat. Ecol.*, *44*, 633–667.

Neal, C., and H. P. Jarvie (2005), Agriculture, community, river eutrophication and the Water Framework Directive, *Hydrol. Processes*, *19*, 1895–1901.

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

Oreskes, N., and K. Belitz (2001), Philosophical issues in model assessment, in *Model Validation: Perspectives in Hydrological Science*, edited by M. G. Anderson and P. D. Bates, pp. 23–41, John Wiley, London.

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

Pappenberger, F., K. Beven, K. Frodsham, R. Romanowicz, and P. Matgen (2007), Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach, *J. Hydrol.*, *333*, 275–287.

Perrin, C., C. Michel, and V. Andréassian (2001), Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *J. Hydrol.*, *242*, 275–301.

Radcliffe, D. E., and M. L. Cabrera (2006), *Modeling Phosphorus in the Environment*, CRC Press, Boca Raton, Fla.

Radcliffe, D. E., J. Freer, and O. Schoumans (2009), Diffuse phosphorus models in the United States and Europe: Their usages, scales, and uncertainties, *J. Environ. Qual.*, *38*, 1956–1967.

Ratliff, L. F., J. T. Ritchie, and D. K. Cassel (1983), Field-measured limits of soil water availability as related to laboratory-measured properties, *Soil Sci. Soc. Am. J.*, *47*, 770–775.

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer (2009), Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, *Bioinformatics*, *25*, 1923–1929.

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

Robson, B. J. (2014), When do aquatic systems models provide useful predictions, what is changing, and what is next?, *Environ. Modell. Software*, *61*, 287–296.

Sample, J. (2015), *Statistics Notes for Environmental Modelling*, GitHub. [Available at https://github.com/JamesSample/enviro_mod_notes, accessed 19 Jun. 2017.]

Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrol. Processes*, *21*, 2075–2080.

Seibert, J. (2003), Reliability of model predictions outside calibration conditions, paper presented at the Nordic Hydrological Conference (Røros, Norway 4–7 August 2002), *Nord. Hydrol.*, *34*, 477–492.

Sharpley, A. N. (1980), The enrichment of soil phosphorus in runoff sediments, *J. Environ. Qual.*, *9*, 521–526.

Sharpley, A., H. P. Jarvie, A. Buda, L. May, B. Spears, and P. Kleinman (2013), Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment, *J. Environ. Qual.*, *42*, 1308–1326.

Sivapalan, M. (2005), Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale, in *Encyclopedia of Hydrological Sciences*, vol. 1, edited by M. G. Anderson, pp. 193–220, John Wiley, Australia.

Sivapalan, M. (2006), *Predictions in Ungauged Basins: Promise and Progress*, Int. Assoc. of Hydrol. Sci.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002), Bayesian measures of model complexity and fit, *J. R. Stat. Soc., Ser. B*, *64*, 583–639.

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

Stutter, M. I., S. J. Langan, and R. J. Cooper (2008), Spatial contributions of diffuse inputs and within-channel processes to the form of stream water phosphorus over storm events, *J. Hydrol.*, *350*, 203–214.

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

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

Syers, J. K., A. E. Johnston, and D. Curtin (2008), Efficiency of soil and fertilizer phosphorus use. Reconciling changing concepts of soil phosphorus behaviour with agronomic information, in *Fertilizer and Plant Nutrition Bulletins*, Food and Agric. Organ. of the U.N., Rome.

Trimble, S. W. (2010), Streams, valleys and floodplains in the sediment cascade, in *Sediment Cascades: An Integrated Approach*, edited by T. Burt and R. Allison, pp. 307–343, Wiley-Blackwell, U. K.

Twarakavi, N. K. C., M. Sakai, and J. Šimůnek (2009), An objective analysis of the dynamic nature of field capacity, *Water Resour. Res.*, *45*, W10410, doi:10.1029/2009WR007944.

U.S. Department of Agriculture (USDA) (2004), *National Enginnering Handbook, Part 630—Hydrology*, chap. 11 (Snowmelt), U.S. Dep. of Agric., Nat. Resour. Conserv. Serv.

Van Dijk, A. (2010), Climate and terrain factors explaining streamflow response and recession in Australian catchments, *Hydrol. Earth Syst. Sci.*, *14*, 159–169.

Van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*, 10–23.

Van Meter, K. J., and N. B. Basu (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, A. J., P. Durand, V. Beaujouan, W. W. Wessel, K. J. Raat, P. G. Whitehead, D. Butterfield, K. Rankinen, and A. Lepisto (2002a), A nitrogen model for European catchments: INCA, new model structure and equations, *Hydrol. Earth Syst. Sci.*, *6*, 559–582.

Wade, A. J., P. G. Whitehead, and D. Butterfield (2002b), 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.

Wade, A. J., D. Butterfield, T. Griffiths, and P. G. Whitehead (2007), Eutrophication control in river-systems: An application of INCA-P to the River Lugg, *Hydrol. Earth Syst. Sci.*, *11*, 584–600.

Watson, A., and R. Evans (2007), Water erosion of arable fields in North-East Scotland, 1985–2007, *Scot. Geogr. J.*, *123*, 107–121.

Wellen, C., A.-R. Kamran-Disfani, and G. B. Arhonditsis (2015), Evaluation of the current state of distributed watershed nutrient water quality modeling, *Environ. Sci. Technol.*, *49*, 3278–3290.

Whitehead, P. G., E. J. Wilson, and D. Butterfield (1998), A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I — model structure and process equations, *Sci. Tot. Environ.*, *210–211*, 547–558.

Whitehead, P. G., J. Crossman, B. B. Balana, M. N. Futter, S. Comber, L. Jin, D. Skuras, A. J. Wade, M. J. Bowes, and D. S. Read (2013) A cost-effectiveness analysis of water security and water quality: Impacts of climate and land-use change on the River Thames system, *Philos. Trans. R. Soc. A*, *371*, 20120413.

Wischmeier, W. C., and D. D. Smith (1965), *Predicting Rainfall-Erosion Losses From Cropland East of the Rocky Mountains*, U.S. Dep. of Agric., Washington, D. C.

Wischmeier, W. C., and D. D. Smith (1978), *Predicting Rainfall Erosion Losses—A Guide to Conservation Planning*, U.S. Dep. of Agric., Washington, D. C.

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

Wolman, M., J. Miller, and L. Leopold (1964), *Fluvial Processes in Geomorphology*, W. H. Freeman, San Francisco.

Young, R. A., C. Onstad, D. Bosch, and W. Anderson (1989), AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds, *J. Soil Water Conserv.*, *44*, 168–173.