water

MDPI

Article

A New National Water Quality Model to Evaluate the Effectiveness of Catchment Management Measures in England

Barry Hankin [1,*], Johan Strömqvist [2], Chris Burgess [3], Charlotta Pers [2], Sally Bielby [1], Beatriz Revilla-Romero [1] and Linda Pope [3]

[1] JBA Consulting, Mersey Bank House, Second Floor, Barbauld Street, Warrington WA1 1WA, UK
[2] Swedish Meteorological and Hydrological Institute (SMHI), SE-601 76 Norrköping, Sweden
[3] Environment Agency, Deanery Road, Bristol BS1 5AH, UK
* Correspondence: barry.hankin@jbaconsulting.com; Tel.: +44-(0)-192-557-0876

Received: 10 July 2019; Accepted: 1 August 2019; Published: 3 August 2019

check for updates

Abstract: This investigation reports on a new national model to evaluate the effectiveness of catchment sensitive farming in England, and how pollution mitigation measures have improved water quality between 2006 and 2016. An adapted HYPE (HYdrological Predictions for the Environment) model was written to use accurate farm emissions data so that the pathway impact could be accounted for in the land phase of transport. Farm emissions were apportioned into different runoff fractions simulated in surface and soil layers, and travel time and losses were taken into account. These were derived from the regulator's 'catchment change matrix' and converted to monthly load time series, combined with extensive point source load datasets. Very large flow and water quality monitoring datasets were used to calibrate the model nationally for flow, nitrogen, phosphorus, suspended sediments and faecal indicator organisms. The model was simulated with and without estimated changes to farm emissions resulting from catchment measures, and spatial and temporal changes to water quality concentrations were then assessed.

Keywords: water quality model; Catchment Sensitive Farming; HYPE model; diffuse pollution

1. Introduction

The objective of this investigation was to develop a new national, unsteady, process-based water quality model to assess the effectiveness of the U.K. government's catchment sensitive farming (CSF) initiative. CSF is an advice-led initiative delivering targeted support that enables farmers to take action to reduce pollution. It is delivered through a network of CSF officers, bringing together land management and capital works to address water quality, alongside other environmental issues [1]. The Environment Agency are required to carry out a formal evaluation of the improvements in environmental quality resulting from programs of changes in agricultural practice, which includes measures to reduce losses of phosphorus, nitrogen, sediment and faecal indicator organisms (FIOs) from farmyards and fields in defined priority areas across England. This formal evaluation is carried out every four years, and in other years measures to reduce losses from agriculture are planned and implemented. The planning and evaluation requires both land use and water quality modelling, including losses from farms and the impact of these losses on in-stream water quality, at both a local and a national scale.

Detecting catchment scale change due to agriculture remains a challenge when combined with changes in point source discharges, weather patterns and a reduction in national scale in-stream monitoring. At a limited number of river sites, however, high frequency in-stream monitoring has been implemented as part of the CSF programme, which allows us to look at long-term and seasonal changes in water quality. The final measure of the impact of changes in agricultural practice is in the water quality at a network of points across England, where water quality is regularly monitored. The data from these sites was used to validate the model.

Recent developments in modelling losses from farms at a field and catchment scale, such as Farmscoper, SEPARATE, Catchment Change Matrix (CCM), NEAP-N [2] and PSYCHIC-P [3] have given us an improved understanding of the loads of substances applied to farmland, and how these loads change with farming practices. However, the national scale models that the Agency generally uses were focused on point source discharges directly to watercourse, and required an external model to 'decay' pollutant loads from farms so they could be entered directly into the watercourse. The national models were also steady-state statistical models, which do not provide long-term daily output but annual statistics.

To evaluate CSF, the Environment Agency (EA) required a more process-based model that could integrate the land phase of the environment with the water phase at a daily timestep, and that could also include a time series of point source releases to look at the impact of changes in discharge limits. The model needed to include the whole of England and provide output on a small catchment scale. This would allow the CSF project to look at the impact not only of changes in the past, but also model potential changes in farming practice to predict their outcome.

The influence of CSF is more complex than simply the measure directly recommended by the operational staff, and the programme also facilitates other agri-environment agreements, helping to maximize the benefits of these agreements for improving water quality. For this reason, the current evaluation looks at interventions, such as buffer strips and woodland, from CSF and other agri-environment schemes.

This project attempts to quantify the benefits to water quality from CSF alone and in combination with other agri-environment schemes, and the reductions predicted represent the combined reduction from all modelled interventions (The results presented here differ from those presented in the final CSF evaluation because they stop at 1 January 2017, one year before the end of the full CSF evaluation period, and they include reductions from all modelled agricultural interventions). For the final CSF evaluation results, please refer the latest CSF publication [4].

A key requirement of the new model was to be able to represent diffuse pollution as the composition of many discrete farm-scale 'point sources' of pollution and evaluate the myriad of pathways so that these can be aggregated to a small catchment scale and evaluated alongside point source loads at the national scale. In this paper, we investigate a modified version of the HYPE model that accommodates variation in hydrological pathway impact travel times between farm-scale emissions and the watercourse network. This is achieved through incorporating travel time in the definition of a hydrological response unit, along with land use and soil characteristics, such that emissions into the different runoff fractions at different proximities to watercourses can experience different levels of assimilation.

2. Methodology

2.1. Model Framework

HYPE is an open-source semi-distributed Swedish hydrological water quality nutrient package, simulating water flow and pollution transport and transformation from precipitation, through soil, river and lakes to the river outlet [5–7]. In a typical model application, the modelled domain is divided into connected sub-basins. The landscape is further divided into soil-land classes with similar hydrological responses, which can be grouped and modelled efficiently at the national scale (typically between 50 and 100 classes). A key modification to HYPE developed for this study has been to include an additional hydrological 'travel-time' class, based on hydrological-pathway analysis of national remotely sensed topographic data. Although this does not influence hydrological response, it enables banding of the diffuse pollution sources, combined with a modified half-life of pollutants to represent more realistic losses across different pathways. In this paper, the new model and driving datasets are introduced before describing the new functionality.

2.2. Model Setup

The model encompasses the European Water Framework Directive (WFD) river network and catchment waterbody reporting units for England. These 4091 waterbodies were further sub-divided into 5117 sub-basins in order to split catchments at reliable flow gauges to allow for more accurate flow calibration. Each sub-basin was further divided into hydrological response units (HRUs) that share common runoff characteristics based on soil type, land cover and travel time to the nearest watercourse. There are 234 HRUs nationally, based on 5 soil-texture classifications, 14 land cover classes based on [8], and 5 travel time classes. The different HRUs each have three soil layers with variable depths (and depths-to-drainage), and these respond differently to rainfall and antecedent conditions.

2.2.1. River Network and Lakes

Lakes can have a large influence on runoff and water quality characteristics of a catchment, so it is important to include their impact explicitly. Generally, lakes are represented with a depth, area and crest level above which spills commence based on a weir relationship. The Environment Agency has data for 5500 lakes, and these lakes were filtered and classified depending on geometry in relation to the sub-basin they were in, and whether they were in-line or off-line to the river network. The lake data were then quality assessed—resulting in 4989 lakes—to be classified into '*olakes*' and '*ilakes*'. *Olakes* are the more significant lakes and control the catchment outlet, but also receive and mix water from higher up the network, whereas *ilakes* are offline lakes draining parts of the local catchment only. The attributes, such as area and watershed draining area, for the different parameters were then set for the categorised lakes as part of the hydrological calibration.

2.2.2. Precipitation and Temperature Data

HYPE is driven by daily timestep rainfall and temperature data, which were derived from the Met Office UKCP09 land surface climate observations daily temperature and precipitation at 5 km resolution [9]. The data were obtained as netCDF files, one file per year, per variable (rainfall and mean temperature), containing daily data stored on a fixed 5 km grid (see Appendix B). The gridded 5 km data were intersected with the new HYPE Sub-Basin spatial data. An intersected area field was added and the fractional area of each cell was used to apportion rainfall totals to each of the 5117 sub-basins.

2.2.3. Point Source Pollution Loads

Point sources were based on three core types of effluent discharge that were all added as monthly point source load time series, using new functionality in HYPE written for this project:

- Effluent with flow and water quality time series—exported from the Environment Agency Water Information Systems KISTERS (WISKI) and Water Information Management System (WIMS) archives (http://environment.data.gov.uk/water-quality/view/landing), respectively.
- Time series based on the use of defaults for qualitative consented discharges calculated from default 'average' values for different types of effluent.
- Time series that are purely flow using consented abstraction data, incorporated based on the data used in the national SIMulated CATchment (SIMCAT) model [10].

The point source pollution loads were included for a range of inputs—in total 40,156 time series were generated across the different determinands and included descriptive consented discharges, alongside effluent from sewage treatment works.

2.2.4. Water Quality Monitoring Data and Assumptions

Scripts were written in R to process national monitoring programme data for the following chemical determinands: phosphorus (soluble phosphorus, SP, and total phosphorus, TP); nitrogen (inorganic nitrogen, IN); faecal indicator organisms (FIO) and suspended solids (SS). The Environment Agency sometimes models different determinand types to the original HYPE model, and in these cases surrogate determinands have been used. The following assumptions were made for the national scale model where measurements were not available:

- SP is approximated by monitored soluble reactive phosphate (SRP).
- The ratio IN:TN (total nitrogen) ~ 0.9.
    - The ratio IN:ON (Organic Nitrogen) ~ 1:9 (this simplified relationship is compensated by the model calibration).
- The ratio of SP:TP (total phosphorus) ~ 0.9.
    - The ratio SP:PP = 1:9.

In total, 14,800 time series datasets were compiled into a monthly time series over the modelling period with which to compare model predictions.

2.2.5. Farm Pollution Loads

The diffuse loads were also prepared as monthly time series data representing the excess emissions of farming practices and provided from the Environment Agency's Catchment Change Matrix (CCM) [1]. The CCM database combines data on farm-level land use over time with coefficients derived from multiple runs of the ADAS Farmscoper model [11–13] to create annual excess loads per farm. These annual loads are converted to monthly values using a literature values that distribute the total annual load to monthly values based on relevant farming practices [3,14,15]. The format for these diffuse load time series is new for HYPE, and comprises a single diffuse monthly load time series file for each of the 234 HRUs for each chemical, apportioned to each of the 5117 sub-basins requiring some large geo-spatial intersections. These diffuse load files were prepared by the Environment Agency for baseline (no interventions) and post-intervention scenarios. The post-intervention scenarios reflect a reduction that has been 'put into practice', and was therefore used in calibration, with the difference in loads examined to understand change.

Inputs from other diffuse sources, such as woodland and diffuse urban sources, were added to the input file as a constant rate based on calibration of catchments with, for example, a high proportion of woodland and low agricultural coverage, ensuring a mass balance at monitoring locations.

2.3. Simulation

The model was simulated for seven years to 'spin-up' the water levels in the soil stores, so it has been driven by the UKCIP09 data since 1993, and the assessment period extends from 2000 to December 2016 (later data was not available at the time of writing). Having pre-processed different datasets to setup the sub-basins, the HRUs, the river network and lakes, the predicted daily flows were calibrated against 777 reliable long-term records spanning the period 2000–2016 using two core global performance measures of regional Nash–Sutcliffe efficiency (NSE) and median Kling–Gupta efficiency (KGE) [16], along with numerous time series and spatial plots assessed by eye.

2.4. New HYPE Functionality

Figure 1 shows a simplified schematic of a catchment split into three HRUs with different runoff characteristics. Farms 1 and 2 are both within HRU2, and for simplicity have the same emissions of nutrients. However, Farm 2 poses a greater risk of contamination of the watercourse (solid blue line) because it is much closer to the watercourse and there is less time for decay and assimilation in the soils.

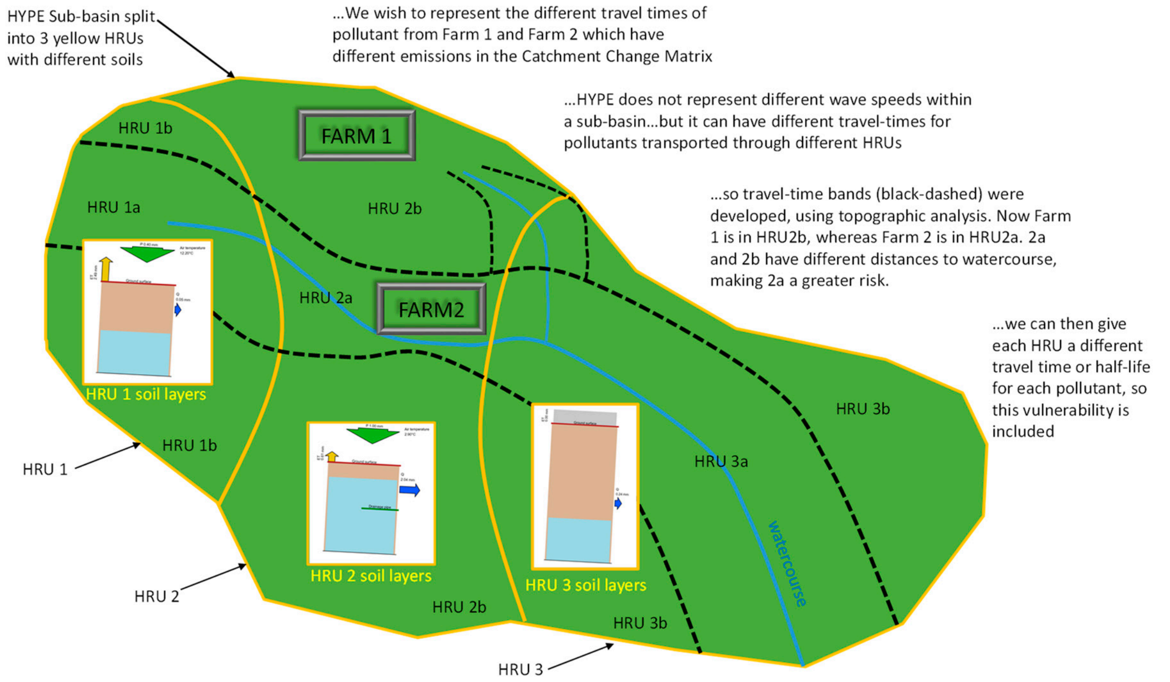

Figure 1. Schemetisation of new HYPE functionality.

In the modified version of HYPE, the HRUs are split by travel-time to the closest watercourse based on topographic analysis. Farm 1 is now differentiated from Farm 2 by the fact it is more remote from the watercourse than Farm 2. Pollution emitted by Farm 1 is given a reduced half-life compared with that emitted by Farm 2 to reflect the increase travel-time (more time for losses). The change in concentration due to decay is calculated according to Equation (1):

$$C_{t+1} = C_t \times e^{-\ln(2) \times \frac{\Delta t}{expdec}}, \tag{1}$$

where Δt is the time step (days, the time step here is 1 day) and *expdec* is the half-life parameter (days). The loads are added to the soil on a daily basis, and it is assumed they are gradually varying for the approximations to hold. The loads may be passed into the soil at different depths, firstly pooled at the soil surface and then moving into the first, second or third soil layers. Typically the first soil layer represents the soil down to the ploughing depth, the second layer extends down to the rooting depth and the third layer represents the ground water zone. In this application, all loads were added to the pool at the soil surface. The whole pool is included in the determination of runoff concentration, i.e., all will be assumed dissolved except for the pool on top of the soil, for which substances are 'washed-out' during precipitation events.

It was assumed that the decay parameter *expdec* is valid for a load introduced on a land parcel with a certain travel-time to the stream (ToT_{REF}, a parameter), and we artificially sped up the decay for HRUs with longer travel times than the reference time of travel and slowed it down for HRUs with shorter travel times:

$$C_{t+1} = C_t \times e^{-\ln(2) \times \frac{\Delta t}{expdec \times f(ToT)}}, \tag{2}$$

$$f(ToT) = \left(\frac{ToT_{REF}}{ToT_{SLC}}\right)^{ToTexp}, \tag{3}$$

where ToT_{SLC} is the time-of-travel for the HRU, and *ToTexp* is the parameter (1 in the simple case).

2.5. Travel Time Classification of Landscape

Travel time classes were developed using an 'impedance' weighted hydrological pathway length calculation. Conceptually, this is an average travel time of a parcel of water being transported across the faster responding overland and near-surface soil layers, including transport via tile drainage. A national 10 m re-sampled digital terrain model (DTM) was first analysed using the ArcGIS 'hydrology' toolset (http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/an-overview-of-the-hydrology-tools.htm) to produce national slope, flow direction and flow accumulation rasters. These were used in combination with a pathway-length tool that determined the hydrological pathway length to the nearest watercourse for every 10m land parcel. The hydrological pathway length was weighted with the reciprocal of an estimated near-surface speed (impedance) to produce an estimated plume travel-time.

The approach accounts for variable slopes experienced along each part of the hydrological pathway, although there are large uncertainties in estimating average velocities. To do this, the local overland flow velocity was first estimated based on Manning's equation, using the slope analysis of the DTM and an approximate roughness. This was then reduced by a factor, f_v, to estimate plume-average transport rates across agricultural hillslopes where field drains will also form significant preferential flow pathways. Transportation rates of the order of 0.03 m/s have been measured from hillslope-scale tracer experiments, where significant macro-pores are present [17]. It is argued that the majority of pollution will be transported in the fast-responding runoff, so f_v was adjusted to give compatible average rates (giving $f_v \sim 30$).

Ultimately this scaling influences the pollution decay rate (half-life) set during water quality calibration, so it is the combination of f_v and *expdec* that is important. Despite the approximations involved, the approach enables, at a national scale, differential travel times for a plume of pollutant to be used, reflecting the local slope conditions experienced by a plume along the hydrological pathway to the closest watercourse.

2.6. Calibration Data

The National River Flow Archive (https://nrfa.ceh.ac.uk/) daily flow data was assessed for over 800 gauges across England, assessing information on the data quality on the website, and through generation of flow duration curves and time series using a set of automated R scripts over the calibration period between 2000 and 2016. Only very obvious outliers were removed (>3 orders of magnitude), and datasets with gaps in the record were discarded, which resulted in 777 calibration gauges. The spatial data for the HYPE sub-basins were adjusted until these gauges were all at outlets, allowing for a more accurate calibration.

For water quality data, over 14,400 time series were generated across the different chemical determinands, and a highly automated R Script was written to check for outliers, trends (Mann–Kendal), and step-changes (Pettitt test). Very few datasets were rejected, and trends were not removed as with steady-state modelling, since in the new HYPE models the through-time changes. A step change in the observation data should be reflected in the model since it incorporates through-time changes in the inputs, such as sudden reductions in phosphorus concentrations due to chemical stripping at a point source.

3. Results and Discussion

3.1. Flow Calibration

A series of advanced calibration tools were used and developed to assess performance, including the HYPETools R scripts (see Appendix A), an on-line HYPE model performance tool, where results can be instantly visualised (Figure 2), and new R Scripts comparing the predictions with and without catchment measures and the predicted differences were written.

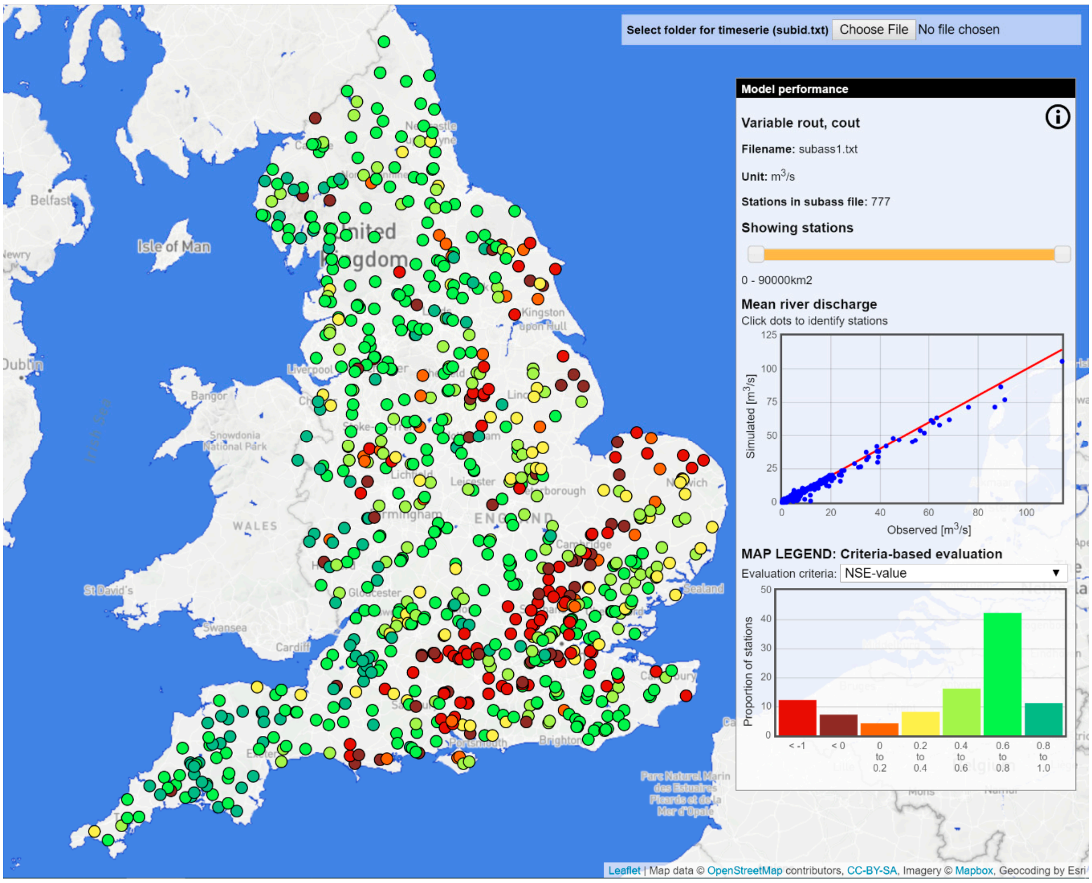

Figure 2. National performance map (based on relative error in the mean flow). The figure is a screenshot from the online calibration tool.

Having attained a water balance through adjustments mainly to the evaporation rates, the flows were calibrated in a multi-step process. The calibration strategy included first isolating small catchments comprising a high proportion of a key soil/land use class. This might result in 20 small catchments with 60% medium-grain size soils. Key parameters (Table 1) for this group were then adjusted within acceptable bounds—either manually or by using auto-calibration. With an improved performance at the 20 gauges, the model was then run nationally to check that the adjustment did not make overall performance worse.

Table 1. Sensitive flow calibration parameters and ranges.

Parameter	Description	Min	Max
wcfc	field capacity	0.05	0.4
wcep	effective porosity in soil	0.01	0.4
rrcs1	soil recession coefficient	0.05	0.5
rrcs2	soil recession coefficient	0.005	0.1
srrate	fraction for surface runoff	0.01	0.4
trrcs	recession coefficient for tile drains	0.1	0.5
mperc2	maximum percolation capacity from soil layer 2 to soil layer 3	2	40

This was repeated until improvements could no longer be made easily, ultimately resulting in a regional NSE of 0.89 across 777 flow gauges, and a median KGE of 0.67. The regional NSE is the NSE over all 777 time series (i.e., 777 time series were stacked on top of each other for the assessment).

Such figures hide the variability in performance, but the areas of worst performance can be identified using the HYPE web tool, and it was evident from the time series that the model under-performs in chalk geologies (the red-dots in Figure 2). In order to refine the flows further in these locations, a groundwater module is considered necessary, linking surface catchments to the underlying larger basins that govern the baseflows across multiple catchments. Areas with an extensive network of canals and large areas where the river flow is controlled by sluice gates and pumping stations also caused poor model performance at some monitoring sites.

3.2. Water Quality Calibration

The water quality calibration was very dependent on attaining a strong flow balance, as without this the fluxes of pollution and loss rates will be biased. The relatively strong national flow calibration provided some confidence that the key physically-based parameters controlling uptake, losses and half-life of different pollutants could be modified realistically. It was important to first ensure a mass balance, and then, as before, iterate the calibration to generate improved performance compared with observational data. Sensitive water quality parameters are shown in Table 2.

Table 2. Sensitive water quality calibration parameters and ranges. SP = soluble phosphorus, IN = inorganic nitrogen, FIO = faecal indicator organisms.

Parameter	Description	Min	Max
SPdecay, INdecay etc.	Half life parameters for new loss rate model (d)	300	3000
T1expdec	Half life parameter for FIO in surface water (d)	1	100
denitwl, denitwrm	User defined parameter for denitrification in lakes and rivers (-)	10^{-5}	10^{-3}
macuptspr	rate controlling riverine uptake of soluble P by macrophytes (-)	10^{-5}	10^{-3}

The calibration for the model development phase paused once an acceptable level was achieved, and further work will be carried out to improve this in phase 2 of the model development. Overall, the correlation was strong for catchments with higher flows and larger catchments (Figure 3), but also improves when using catchments with a higher proportion of arable land cover. For phosphorus the correlation was not as strong, but the overall performance for larger catchments was acceptable (Figure 4).

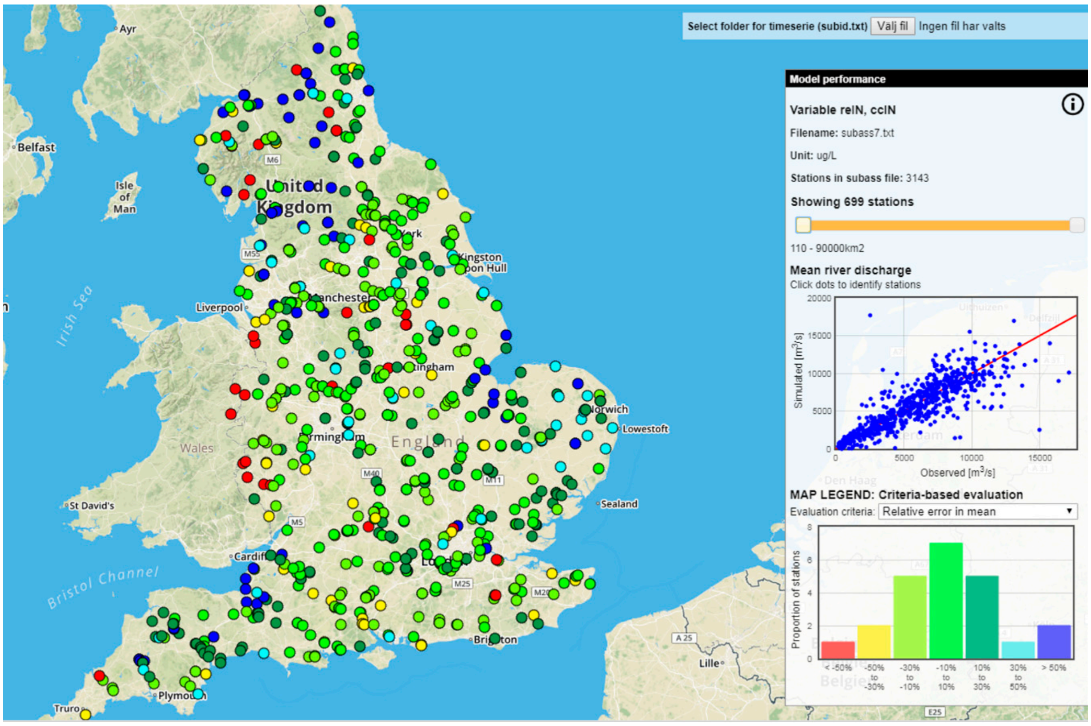

Figure 3. National performance for inorganic nitrogen (IN) using a flow weighted performance measure for catchments > 110 km^2.

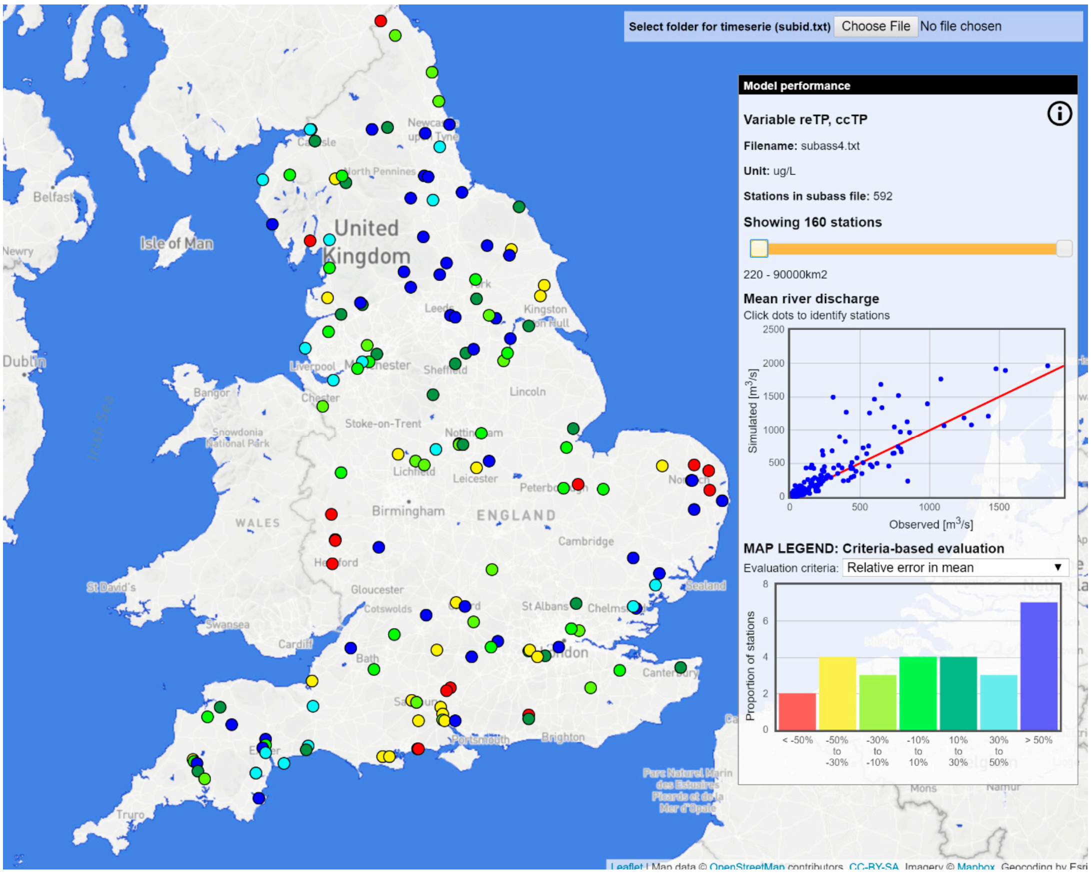

Figure 4. National performance for total P (TP) for catchments > 210 km^2.

4. Discussion

Quantifying the Effectiveness of Catchment Measures

The calibrated model was next used with farm emissions representing a baseline, assuming there had been no agricultural intervention measures, and then with these emissions was adjusted to reflect reality. Figure 5 shows a heat map for the predicted annual average percentage reduction of TP and TN across the country averaged over 10 years of the CSF programme (2006–2016), with hotspots around the North West and East Anglia, showing a distribution that is qualitatively similar to the distribution of effort (Figure 1 in [18]). The 10% reduction over baseline in Figure 5 is also commensurate with that predicted for 2014 across the key pollutants for catchments with the most significant reductions, which was expected to range between 4–7% [18].

Figure 6 shows a range of larger reductions through time for the 100 catchments with the greatest load reduction for suspended solids (based on the concentration and flows at the catchment outlet), showing reductions that are commensurate with the findings of the 2014 predictions (inset box plots), where it was considered that CSF was most effective in terms of delivering sediment reductions, with predicted reductions varying greatly across catchments up to 36 per cent, with an average agricultural load reduction of 12 per cent [18]. This is similar to the HYPE model predictions in Figure 6.

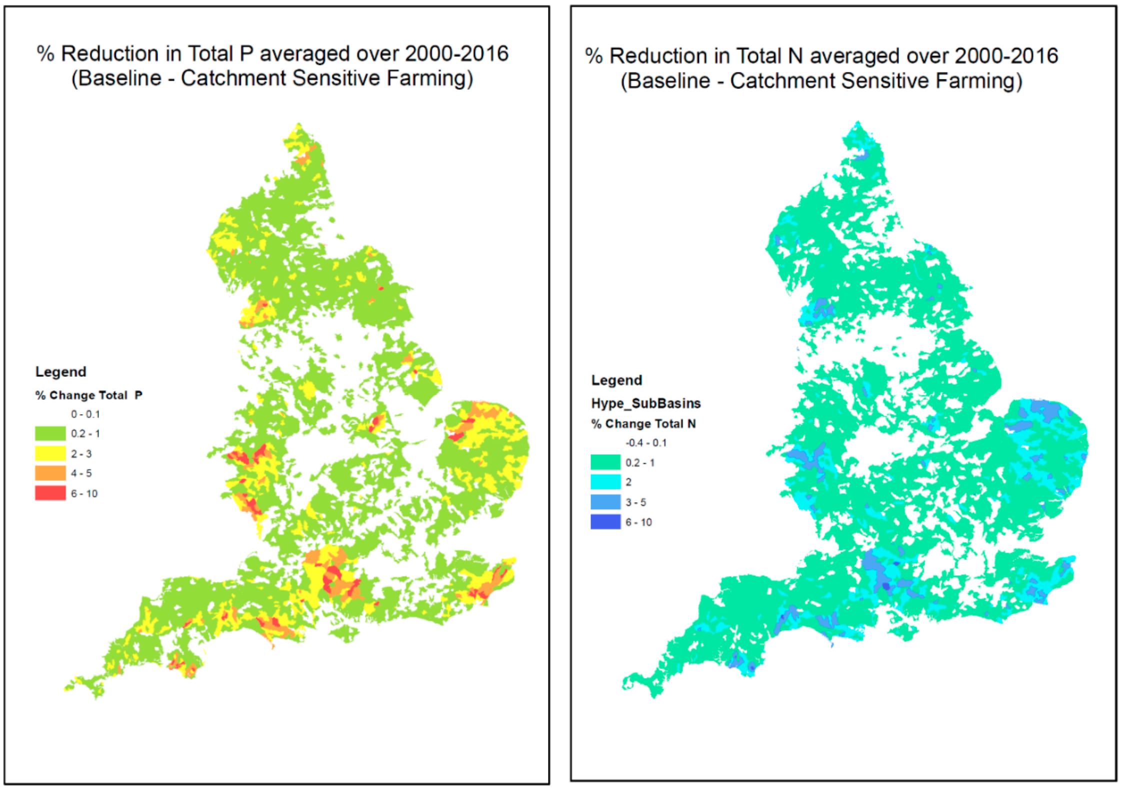

Figure 5. Predictions of percentage reduction in total phosphorus and total nitrogen loads based on the Catchment Sensitive Farming Programme and agri-environment schemes.

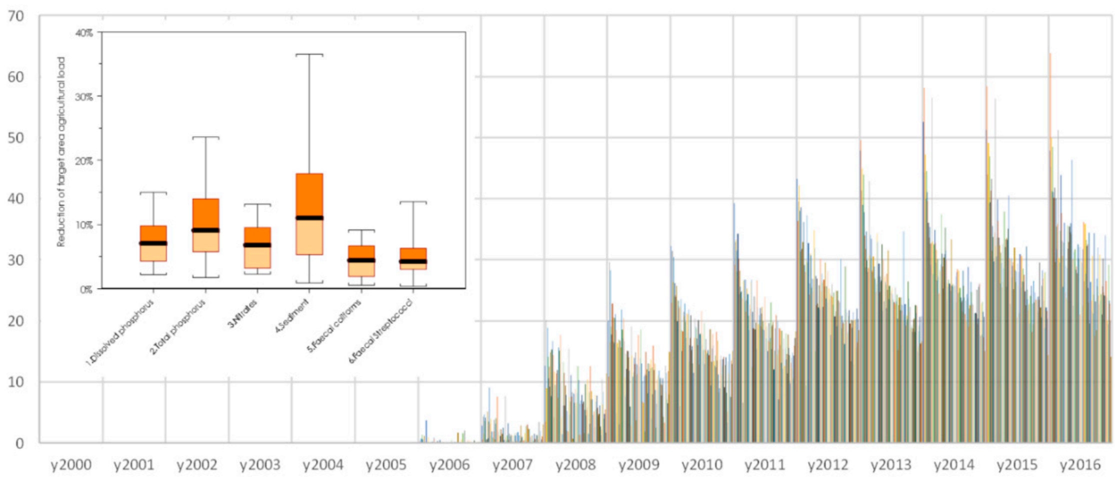

Figure 6. Year-on-year predictions of percentage reduction in suspended solids for targeted catchments compared with 2014 catchment sensitive farming (CSF) predictions (inset, after EA, 2014).

There is insufficient space to show all the results at all sites individually, so Figures 5–8 demonstrate some of the predicted changes compared with observations. The difference between post-intervention and baseline is not always discernible, so the peak differences are shown in the captions. Figure 7 highlights how the model predicted the general reduction in the in-river concentrations, more as a result of reduction in point sources than diffuse, and for this catchment only a very small change has been made as a result of changes to agricultural practices.

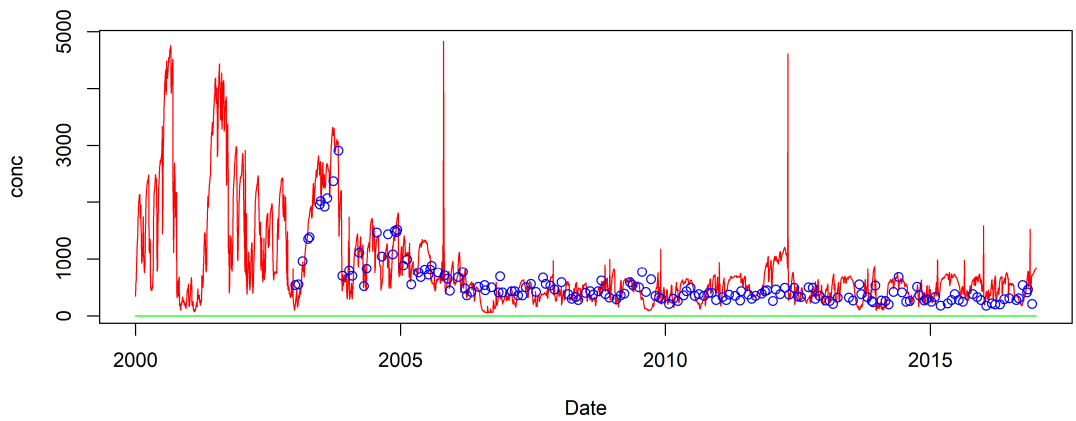

Figure 7. Time series plot of baseline concentration of TP (ug/L) from 2000 to 2016 with observations in blue highlighting changing because of point source improvements (sewage treatment work upgrade). Diffuse improvement was at most 3.96 ug/L.

The improvements in relative error in mean going from baseline (no measures) to CSF (actual measures calibrated against) were 2.0% for IN, 2.7% for SS and 0.4% for TP, but worse for FIO at around −6%. This last anomaly arose because there was much less data for FIO, and the data that did exist were very noisy. Assessing the same measure for the key site with higher frequency FIO data (see Figure 7) shows that the relative error in the mean improved by 2.0%.

Figure 8 demonstrates a more significant change between predicted SS with and without the interventions, rising to a 10% change, with some of the detected high concentration pollution events captured periodically. Of course, the sampling error in the monthly samples is such that the full variation in SS cannot be captured, but evidence of capturing the order of magnitude of the event-based pollution episodes provides some confidence.

Figure 9 shows the noisier response of FIOs (modelled as a tracer, T1, with calibrated half-life). The observations suggest that the water quality is actually more sensitive than modelled, although the general pattern is discernible. This site is on the Wyre catchment, targeted for CSF and monitored in particular with higher frequency sampling than is typically the case. The reductions in FIOs can be up to 10%.

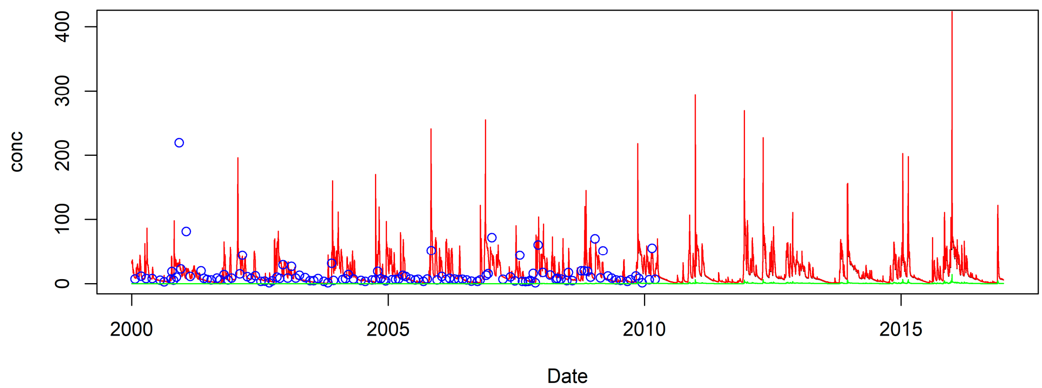

Figure 8. Time series plot of baseline (red) and the difference between baseline and post-intervention (green) in prediction of suspended solids (ug/L) from 2000 to 2016. Maximum diffuse concentration improvement was 14.11 ug/L. Monitoring data is represented by blue circles.

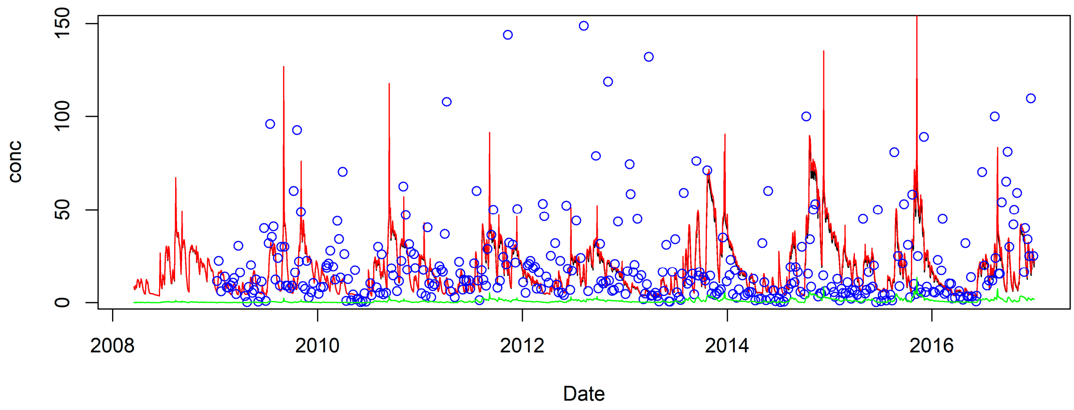

Figure 9. Time series plot of baseline (red), post-intervention (black) and differences (green) in prediction of faecal indicator organisms (FIOs) (counts per billion) from 2000 to 2016. Maximum diffuse concentration improvement was 13.28 ug/L.

5. Conclusions

A new rainfall driven, process-based, unsteady, semi-distributed rainfall runoff and water quality model at the WFD waterbody scale was developed using an innovative adaptation of the HYPE software to allow better use of the increasingly rich spatial data for farming emissions, whilst retaining core hydrological concepts, such as flow pathway partitioning and turn-over times in soils. This new functionality of HYPE has made the incorporation of 'load time-series' of diffuse and point source pollutants possible, and has made it more compatible with the available data and methods used in the U.K., and potentially other countries. The framework allows diffuse sources to be treated more like many small point sources, which is ultimately correct. The flexibility also allows incorporation of any number of event-driven pollution episodes to be added realistically—for instance, it would now be possible to add the very episodic loads from combined sewer overflows or emergency outfalls. With the rise in event-monitoring at sewage treatment works, this becomes a real possibility, and it may help capture more of the variance in the observations. Variability in the boundary conditions will always make modelling time series of concentrations difficult, and thus it is so important to include measurements in small upstream catchments so the variability in diffuse inputs can be characterised.

The modelling shows that the effectiveness of a national programme of distributed farming measures can be assessed, and it produces outputs similar to other independent estimates, although the uncertainties should now be more fully appraised. This will be possible given the relatively short run-time of approximately 1 h 40 min for this national scale model with all the details and thousands of diffuse and point sources of pollution over the 17 year daily simulation. With an estimate of prediction uncertainty, the model can also be used in forward-mode to understand future scenarios for policymaking [19].

For evaluating the CSF programme, the modelling uncovers the expected order of magnitude changes and patterns in the CSF programme, with the following observations:

- For total P, the maximum percentage reduction in average load between CSF and baseline scenarios over the full period (2000–2016) was 10% in targeted catchments, with some catchments changing up to 14%.
- For total N, the maximum percentage reduction in average load between CSF and baseline scenarios over the full period (2000–2016) was 11% in targeted catchments, with some catchments changing up to 16% in one year.
- For suspended solids (at 105 °C), the maximum percentage reduction in average load between CSF and baseline scenarios over the full period (2000–2016) was 20.6% in targeted catchments, with some catchments changing up to 70% in one year.

- For FIOs, the maximum percentage reduction in average load between CSF and baseline scenarios over the full period (2000–2016) was 12.5% in targeted catchments, with some catchments changing up to 65% in one year.

These conclusions are based on the process-based, unsteady national modelling of 5117 catchments, just smaller than the WFD waterbody catchments. The model combines 40,156 point discharge and abstractions, and diffuse loads from the Catchment Change Matrix that are spatially distributed across 234 hydrological response units intersected with the 5117 catchments.

- A range of physical processes including nutrient cycling, sediment erosion and event driven pollution episodes can begin to be taken into account at the WFD waterbody scale, nationally.
- Event-driven pollution episodes and step changes in pollution levels can be modelled with the new model so change through time can be evaluated. With further calibration/investment, the model could be used to project different farm-programme management measures and for example climate change scenarios—since the model is driven by temperature and precipitation, different scenarios can be easily modelled.
- New functionality has been added to the HYPE model to attempt to account for pathway-impact effects of farm pollution at different distances from the watercourse. This requires further evaluation, but the results so far are promising
- The model is well calibrated for flows, apart from on chalk catchments, where it is recommended a groundwater model is incorporated in the longer term
- The total loads for changes in N and P through time and the percentage improvement through time is of a similar order of magnitude to the expected change based on previous studies (Environment Agency, 2014).

Author Contributions: Conceptualization, B.H., J.S., C.B.; methodology, B.H., J.S., C.B., S.B., B.R.-R., L.P.; software, C.P., J.S., B.H.; validation, B.H., J.S., B.R.-R., S.B., L.P.; formal analysis, B.H., J.S., C.B., S.B.; investigation, all; resources, B.H., C.B.; data curation, C.B., L.P.; writing—original draft preparation, B.H., J.S., C.B., L.P.; writing—review and editing, B.H., L.P., C.B.; visualization, B.H., J.S.; supervision, B.H., J.S.; project administration, B.H., C.B.; funding acquisition, C.B., L.P., B.H.

Funding: This research was funded by the U.K. Department for Environment, Farming and Rural Affairs (Defra).

Acknowledgments: This investigation, model conceptualisation, software development and model-build-calibration has been a concerted, collaborative effort by a small team spanning the English Environment Agency, SMHI-Sweden and JBA Consulting.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A Software and Data

- HYPE is developed by SMHI (http://hypecode.smhi.se/). The code is open source.
- Hype tools for analysis of results: https://github.com/rcapell/HYPEtools/releases/tag/v0.4-0.
- The on-line HYPE model performance tool http://hypeweb.smhi.se/model-water/hype-tools/modelview/.
- The modified version of the HYPE code was written by Charlotta Pers and was 'branched' off from HYPE version 5.5.0, and will be made available in some form at http://hypecode.smhi.se/ in the future.

Appendix B Open Datasets

- Rainfall and temperature gridded data: http://catalogue.ceda.ac.uk/uuid/319b3f878c7d4cbfbdb356e19d8061d6.
- Flow data: https://nrfa.ceh.ac.uk/.
- Water quality data: http://environment.data.gov.uk/water-quality/view/landing.

References

1. Burgess, C.; Pope, L. *Modelling the effectiveness of Catchment Sensitive Farming (2006–2018) at Reducing Diffuse Water Pollution from Agriculture in England—Technical Report to Natural England*; The Parliamentary Office of Science and Technology: York, UK, 2019.
2. Anthony, S.; Quinn, P.; Lord, E. Catchment scale modelling of nitrate leaching. *Appl. Biol.* **1996**, *46*, 23–32.
3. Davison, P.; Withers, P.J.A.; Lord, E.; Betson, M.J.; Stromqvist, J. Psychic—A process-based model of phosphorus and sediment mobilisation and delivery within agricultural catchments. Part 1: Model description and parameterisation. *J. Hydrol.* **2008**, *350*, 290–302. [CrossRef]
4. *Environment Agency Catchment Sensitive Farming Evaluation Report—Water Quality*; Phases 1 to 4 (2006–2018); Natural England Publication: York, UK, 2019.
5. Arheimer, B.; Lindström, G.; Pers, C.; Riosberg, J.; Strömqvist, J. Development and test of a new Swedish water quality model for small-scale and large-scale applications. In Proceedings of the XXV Nordic Hydrological Conference, NHP Report No. 50, Reykjavik, Iceland, 11–13 August 2008; pp. 483–492.
6. Lindström, G.; Pers, C.P.; Rosberg, R.; Strömqvist, J.; Arheimer, B. Development and test of the HYPE (Hydrological Predictions for the Environment) model—A water quality model for different spatial scales. *Hydrol. Res.* **2010**, *41*, 295–319. [CrossRef]
7. Strömqvist, J.; Arheimer, B.; Dahné, J.; Donnelly, C.; Lindström, G. Water and nutrient predictions in ungauged basins—Set-up and evaluation of a model at the national scale. *Hydrol. Sci. J.* **2012**, *57*, 229–247. [CrossRef]
8. Hollaway, M.J.; Beven, K.J.; Benskin, C.M.H.; Collins, A.; Evans, R.; Falloon, P.; Forber, K.J.; Hiscock, K.M.; Kahana, R.; Macleod, C.J.; et al. The challenges of modelling phosphorus in a headwater catchment: Applying a 'limits of acceptability' uncertainty framework to a water quality model. *J. Hydrol.* **2018**, *558*, 607–624. [CrossRef]
9. Met Office. UKCP09: Met Office Gridded Land Surface Climate Observations—Daily Temperature and Precipitation at 5 km Resolution. Centre for Environmental Data Analysis. 2018. Available online: http://catalogue.ceda.ac.uk/uuid/319b3f878c7d4cbfbdb356e19d8061d6 (accessed on 2 August 2019).
10. Warn, T. *SIMCAT 11.4 A Guide to Reference for Users*; Environment Agency: Bristol, UK, 2010.
11. Gooday, R.D.; Anthony, S.G. *Mitigation Method-Centric Framework for Evaluating Cost-Effectiveness. Defra Project WQ0106(3)*; Defra: London, UK, 2010.
12. Gooday, R.D.; Anthony, S.G.; Chadwick, D.R.; Newell-Price, P.; Harris, D.; Duethmann, D. Modelling the cost-effectiveness of mitigation methods for multiple pollutants at farm scale. *Sci. Total Environ.* **2014**, *468*, 1198–1209. [CrossRef] [PubMed]
13. Gooday, R.D.; Anthony, S.G.; Durrant, C.; Harris, D.; Lee, D.; Metcalfe, P.; Newell-Price, P.; Turner, A. *Farmscoper Extension*; Final Report for Defra Project SCF0104; Defra: London, UK, 2015.
14. Misselbrook, T.H.; Chadwick, D.R.; Chambers, B.J.; Smith, K.A.; Webb, J.; Demmers, T.; Sneath, R.W. *Inventory of Ammonia Emissions from UK Agriculture—2003*; Inventory Submission Report; Defra Contract AM0127; Defra: London, UK, 2004.
15. Hellsten, S.; Dragosits, U.; Place, C.J.; Misselbrook, T.H.; Tang, Y.S.; Sutton, M.A. Modelling seasonal dynamics from temporal variation in agricultural practices in the UK ammonia emission inventory. In *Acid Rain-Deposition to Recovery*; Springer: Dordrecht, The Netherlands, 2007; pp. 3–13.
16. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
17. Anderson, A.; Anderson, E.; Weiler, M.; Alila, Y.; Hudson, R.O. Dye staining and excavation of a lateral preferential flow network. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 935–944. [CrossRef]

18. Natural England. Catchment Sensitive Farming Phase 3 Delivery Report. 2014. Available online: http://publications.naturalengland.org.uk/publication/6312755155959808?category=6919090 (accessed on 2 August 2019).
19. Hankin, B.G.; Bielby, S.; Pope, L.; Douglass, J. Catchment scale sensitivity and uncertainty in water quality modelling: Catchment scale sensitivity and uncertainty in water quality modelling. *Hydrol. Process.* **2016**, *30*, 4004–4018. [CrossRef]