Water and nutrient simulations using the HYPE model for Sweden *vs.* the Baltic Sea basin – influence of input-data quality and scale

Berit Arheimer, Joel Dahné, Chantal Donnelly, Göran Lindström and Johan Strömqvist

ABSTRACT

Water resource management is often based on numerical models, and large-scale models are sometimes used for international strategic agreements. Sometimes the modelled area entails several political entities and river basins. To avoid methodological bias in results, methods and databases should be homogenous across political and geophysical boundaries, but this may involve fewer details and more assumptions. This paper quantifies the uncertainty when the same model code is applied using two different input datasets; a more detailed one for the country of Sweden (S-HYPE) and a more general one for the entire Baltic Sea basin (Balt-HYPE). Results from the two model applications were compared for the Swedish landmass and for two specific Swedish river basins. The results show that both model applications may be useful in providing spatial information of water and nutrients at various scales. For water discharge, most relative errors are <10% for S-HYPE and <25% for Balt-HYPE. Both applications reproduced the most mean concentration for nitrogen within 25% of the observed mean values, but phosphorus showed a larger scatter. Differences in model set-up were reflected in the simulation of both spatial and temporal dynamics. The most sensitive data were precipitation/temperature, agriculture and model parameter values.

Key words | evaluation, large-scale, multi-basins, modelling, nutrients, water

Berit Arheimer (corresponding author)
Joel Dahné
Chantal Donnelly
Göran Lindström
Johan Strömqvist
Swedish Meteorological and Hydrological Institute (SMHI),
Research and Development,
60176 Norrköping,
Sweden
E-mail: *Berit.Arheimer@smhi.se*

INTRODUCTION

Several European directives today prescribe characterisation and reporting of water status, e.g. the Water Framework Directive, WFD (2000/60/EC) and the Marine Strategy Framework Directive, MSFD (2008/56/EC). The reasons for this are political in order to harmonise environmental ambitions in the union and provoke water protection measures. Implementation of the directives includes legal aspects, water management procedures, knowledge increase of impact-response processes and communication to stimulate changed behaviours and acceptance of water protection among citizens. The characterisation and reporting of water bodies are often based on monitoring programmes; however, numerical models are also frequently used for interpolation and extrapolation between observations in time and space.

Numerical models are structured ways to integrate knowledge of complex systems. For instance, hydrological water quality models are used for the integration of large data materials and process understanding to achieve new information, e.g. in order to quantify catchment sources and sinks in flow-paths, detect critical areas, analyse scenarios and evaluate impacts from protection measures. When applied on the large scale, numerical models in general have been shown to be very useful for international strategic agreements. Examples are the convention on long-range transboundary air pollution using the EMEP model (Simpson *et al.* 2012), the Baltic Sea action plan using the NEST model (Wulff *et al.* 2009), and the Kyoto protocol using the various climate

doi: 10.2166/nh.2012.010

models in the ensemble of IPCC (e.g. 2007). Models are thus useful to complement monitoring programmes for status characterisation and also for reaching understanding and strategic agreements between several actors and nations. On the other hand, models always include assumptions and it is not evident that the same assumptions are made for various scales; data access and knowledge availability may differ considerably. To avoid methodological bias in results, methods and databases should be homogenous across political and geophysical boundaries, but this may involve fewer details and more assumptions. This paper quantifies the uncertainty when the same model code is applied using two different input datasets; one more detailed for the country of Sweden and one more general for the entire Baltic Sea basin.

Sweden is a country rich in surface water, with more than 100,000 lakes >0.01 km^2 and 118 rivers discharging to the surrounding sea. So far, the water authorities have defined 7,232 lakes and 15,563 river reaches for status reporting. However, the national monitoring programme is rather sparse, e.g. there are only 300 hydrological gauges and some 900 sites where grab samples of nutrient concentrations are taken every month. This situation provides little information for each water body and therefore the Swedish Meteorological and Hydrological Institute (SMHI) was requested by the Swedish government to deliver high-resolution model data to water authorities, to support their WFD work. SMHI then developed the Hydrological Predictions for the Environment (HYPE) model (Lindström *et al.* 2010). The new model was applied nationally for 17,313 sub-basins (on average 28 km^2) covering Sweden, although most of these are ungauged, using a step-wise, multi-basin calibration technique (Donnelly *et al.* 2009; Strömqvist *et al.* 2012).

The major environmental problem linked to water status in Sweden is eutrophication, and therefore, nutrient simulations were focused in the development of the HYPE model. This area is also the major concern for the enclosed Baltic Sea, where HELCOM is the coordination platform for the implementation of the MSFD. The Baltic Sea Action Plan (BSAP from Helcom Secretariat 2007) was agreed upon by ministers from all nine countries around the Baltic Sea in 2007, and it implies severe reductions in nutrient emissions. This strategic agreement has been criticised because it was only based on results from one single model. Nowadays the practice is rather to use several model concepts in ensemble runs to estimate the uncertainty in the overall conclusions to decision makers (e.g. IPCC 2007). In order to increase the model ensemble for the Baltic Sea region, the HYPE model was also applied to the entire Baltic Sea basin.

Assumptions about processes are made during each model set-up procedure. Much effort in the hydrological research community lately has been on developing methods to quantify the uncertainties (starting with e.g. Beven & Binley 1992), but also on developing good modelling practices (e.g. Klemes 1986; Lane & Richards 2001; Refsgaard *et al.* 2010). Differences in input data or model parameter values may give different results and conclusions to decision-makers. To avoid misinterpretations this must be highlighted in sensitivity studies or uncertainty estimates, and be communicated to end-users. The present study gives one example of such a sensitivity study that identifies which part of the model is most sensitive to assumptions in the set-up procedure. This part of the model should be the focus for future data collection and model development.

The HYPE model introduces the ability to model detailed hydrological processes at high resolution simultaneously and homogenously across many river basins. In this paper the model set-up of the country of Sweden (S-HYPE; Arheimer *et al.* 2011a; Strömqvist *et al.* 2012) and the model set-up of the entire Baltic Sea basin (Balt-HYPE; Donnelly *et al.* 2010, 2011) were compared. When using a modelling tool to assess water resources and their quality for a region entailing several political entities or even several river basins, it is an advantage that both the methods and data used are homogenous across such political and geophysical boundaries. Hence, the two model applications described in this study have been implemented using as many homogenous data sets and methods as possible, including uniform calibration, over each of the model application's domains. Nevertheless, each application is based on different input data and different resolution (Figure 1) and is calibrated against different observed datasets. Later on, each application can be part of a larger model ensemble to account for

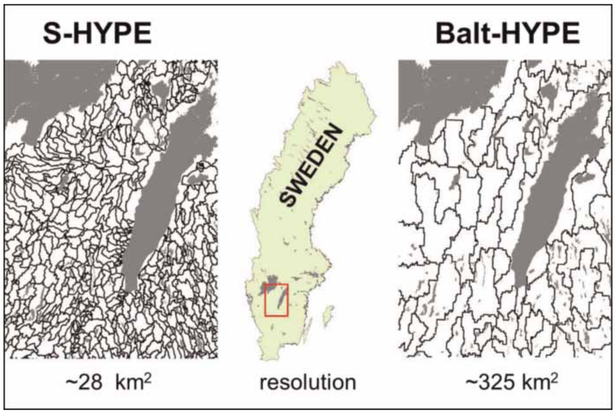

Figure 1 | Different resolution of the Hype model when applied for Sweden, S-HYPE, and for the Baltic basin, Balt-HYPE, respectively. Delineation of sub-basins, for which water and nutrients are calculated, are shown for a region in south-central Sweden.

uncertainties in model-based decision support. The scientific questions put in the paper are:

1. Will homogenous multi-basin approaches give useful model results?
2. How will differences in various set-ups using the same model be reflected in model results?
3. Which parts of the model set-up are most sensitive for the model results?

MATERIAL AND METHODS

The HYPE model (Lindström *et al.* 2010) is a semi-distributed dynamic model, which integrates landscape elements and hydrological compartments along the flow paths with nutrient turnover and transport. Calculations are made on a daily time step in coupled subbasins. Each subbasin is divided into hydrological response units, HRU, according to soil type, vegetation and altitude. The soil profile may be divided into three layers with user specified depths (Figure 2).

Model parameters are either general or related to soil type or land cover depending on the process parameterised. The model simulates processes including snowmelt, surface runoff, surface erosion, macropore flow, tile drainage, groundwater outflow from the individual soil layers, nutrient turnover in soil, and transport/transformation in groundwater, rivers and lakes. The model also accommodates the

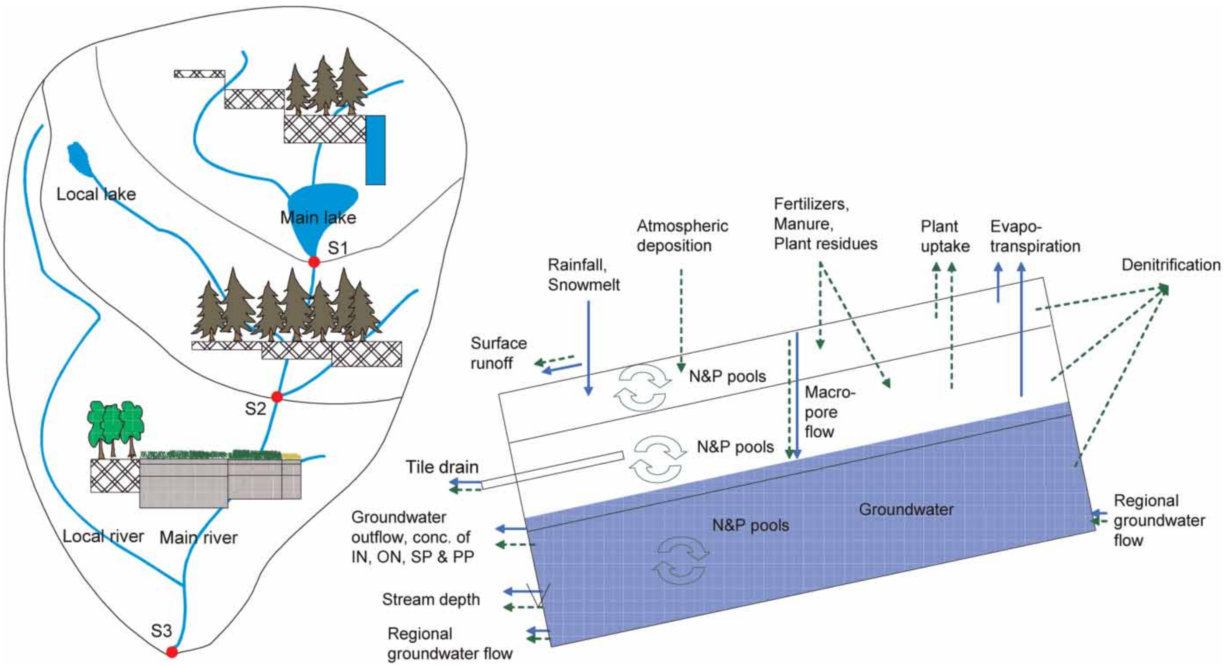

Figure 2 | Schematic figure of the Hydrological Predictions for the Environment (HYPE) model (from Lindström *et al.* 2010). Left: schematic division into sub-basins and land classes according to elevation, soil type, vegetation and lake classes. Result points for each sub-basin S1–S3. Right: schematic model structure within a land class (i.e. a combination of a soil type and a crop), simulated using three soil layers. Solid and dashed arrows show fluxes of water and elements, respectively.

river network and a river routing routine, which enables the calculation of accumulated water discharge and nutrient transport through the mouth of each sub-basin. This routine introduces storage in lakes as well as water flowing in local streams and in the main rivers with continuous mixing of water and nutrients from various sources, which thus implicitly accounts for residence times in each hydrological compartment. Model results may be compared with measurements wherever such are available.

In this paper the recent model set-ups for Sweden and the Baltic Sea basin are evaluated and compared, where different model input data, spatial resolution and observations for calibration were used (Table 1).

S-HYPE

The S-HYPE application (Arheimer *et al.* 2011a, Strömqvist, *et al.* 2012) covers the Swedish landmass including transboundary river basins. About 65% is covered by forest, but there are significant agricultural areas in the south of the country. Sweden has approximately nine million inhabitants. The country is bordered by mountains to the west and a long coastline to the east, meaning that the country is drained by a large number of rivers which start in the west and run eastwards to the Baltic Sea. The S-HYPE application was set up specifically to assist the Swedish implementation of the water WFD by providing nutrient and water information at the local water-body scale defined by the Swedish water authorities. The application is calibrated against 30 daily river discharge stations and 90 water quality stations (with weekly or monthly grab samples), and validated against a further 318 and 582 stations, respectively, for water discharge and quality.

Balt-HYPE

The Balt-HYPE application (Donnelly *et al.* 2011) covers all land areas with discharge to the Baltic Sea, including the Swedish western coast and Glomma River in Norway. Fourteen different countries housing over 85 million inhabitants and over 20 large river basins make up the Baltic Sea basin. Biophysical characteristics range from boreal forests in the north to widespread agriculture in the south, the Scandinavian mountains in the northwest and the Carpathian Mountains in the south. Population and agricultural density increases from north to south. Several large lakes exist in the

Table 1 | Data sources for S-HYPE and Balt-HYPE, respectively

Data type	S-HYPE	Balt-HYPE
Areal extent	476,000 km^2	1.8 million km^2
Median sub-basin resolution	18 km^2	325 km^2
No. sub-basins	17,313	5,128
Topography/routing	Swedish Water Archive (SMHI)[a]	Hydro1K (USGS 2000)
Forcing data	PTHBV 1960–2008[b], resolution = 4 km	ERAMESAN 1980–2004[l], resolution = 11 km
Land cover	CORINE (EIONET 2003)	Globcover 2000[g]
Soil types	Soils database[c] (Swedish Geological Survey, SGU)	European Soils Database (JRC 2006)
Point sources (urban/rural)	Swedish Emissions Database[d]	HYDE Population Database[h], EEA treatment level[i]
Point sources (industrial)	Swedish Emissions Database[d]	EEA Industrial Emissions database[i]
Agriculture data	Statistics Sweden[e], for 22 regions	EUROSTAT and CAPRIS for Nuts regions
Atmospheric deposition	MATCH model[a]	MATCH model[a]
Discharge measurements	SMHI observation[a]	GRDC[j], BALTEX[k]
Nutrient concentration measurements	Swedish Agricultural University (SLU)[f]	European Environmental Agency WISE database[h]

[a]www.smhi.se; [b]Johansson 2002; [c]www.squ.se; [d]www.smed.se; [e]www.scb.se; [f]www.slu.se; [g]www.ionia1.esrin.esa.int; [h]Goldewijk *et al.* 2011; [i]www.eea.europa.eu; [j]http://www.bafg.de/GRDC/EN/Home/homepage_node.html; [k]www.baltex-researc.eu; [l]Jansson *et al.* 2007.

basin including Europe's two largest, Lakes Ladoga and Onega in Russia. The Balt-HYPE application was set up to forecast nutrient and freshwater influxes to the Baltic Sea; however, it can also be used to evaluate discharge and nutrient variables for sub-basins in the basin. The application is calibrated against 35 daily river discharge stations and 20 water quality stations (with seasonal or annual weighted values) at river outlets to the sea. The model is validated against a further 121 daily discharge stations. Water quality can be considered as not validated in this first Baltic Sea basin application, although another 35 water quality stations were included in the evaluation.

Model set-up procedure for each application

Generally, when the detail of the input data used is higher, the smaller the model scale. For example, the Swedish Water Archive, which is used to define sub-basins in the S-HYPE application, uses a topographic model adjusted by hand to truly represent the direction of water flow as observed on the ground. Clearly, this level of detail is impossible at a continental scale using digital elevation models for water divide alienation. At the larger scale, flow routing can only be adjusted to some extent due to lack of local knowledge. This adjustment is done manually where gauged rivers appear to accumulate incorrect upstream areas. The resolution of S-HYPE is 18 times higher than in Balt-HYPE.

There are also some significant differences in input databases for the water quality applications of S-HYPE and Balt-HYPE. The S-HYPE application uses measurements of nutrient loads from point sources as well as national statistical agricultural data for regions which detail crop types, planting dates and fertilisation rates (Strömqvist *et al.* 2012). For Balt-HYPE, agricultural data, including the types, intensities and fertilisation rates of crops grown, are taken from EUROSTAT in the format used for the CAPRIS agrieconomic model (Britz *et al.* 2007). Although more detailed data regarding agricultural statistics are readily available for some of the countries in the model domain, they are not available for all countries, and homogenous inputs for all countries were chosen. Similarly, point source emissions are modelled as a function of the population in each sub-basin, and the dietary protein intake and washing powder/detergent use per country after the methodology proposed by Bouwman *et al.* (2005). Dietary protein intake data were taken from FAOSTAT (2006), washing powder/detergent use was also taken from Bouwman *et al.* (2005), and the gridded population density, both rural and urban was taken from the HYDE database (Klein Goldewijk *et al.* 2011). Nutrient loads from each point source were finally calculated by assuming that the rate of each country's use of primary, secondary and tertiary sewage treatment was applied to the population in each sub-basin. A further advantage of this method is the ability to easily change inputs for modelling of nutrient scenarios.

Daily river discharge data from Swedish (SMHI) and pan-Baltic (BALEX; EWA and GRDC 2009a, b) databases were used to calibrate and validate the parameters describing discharge processes. For water quality modelling, daily nutrient fraction concentrations for nitrogen and phosphorous from Swedish monitoring programmes were used to calibrate the S-HYPE application. Monthly and accumulated seasonal data for total nitrogen (TN) and total phosphorus (TP) from the European Environment Agency's WISE database were used to calibrate the Balt-HYPE application.

Calibration of parameters was made using a step-wise 'representative gauged basin' method (Donnelly *et al.* 2009; Strömqvist *et al.* 2012) in order to link parameters to HRUs (i.e. combined land-uses and soil types). First, all gauged basins with lakes and regulation upstream were excluded. Then, gauged basins with dominant areas of each land-use or soil-type, e.g. forest, agriculture, open land or fine soil, coarse soil, moraine, etc., were chosen as 'representative gauged basins' for each combined land-use and soil-type. Both water and nutrient parameters for each HRU were then determined by optimising a group of stations containing dominant areas of the relevant class. The same unique parameter sets were then applied to all HRUs of a specific type for the whole model domain (i.e. Sweden or Baltic Sea basin). Second, lake parameters were calibrated for a group of gauged basins with unregulated lakes, and parameter values were applied for the whole domain. The model was then run with fixed parameter values to validate the results against all other available monitoring stations, which were then independent.

This calibration methodology can be described as simultaneous in space but stepwise in model routines; hence, a uniform calibration of a multi-basin model. Simultaneous refers to the fact that many river basins are calibrated simultaneously. A uniform calibration approach is defined as a single set of land-use and soil-type linked parameters, simultaneously optimised to the available flow data in the model domain, regardless of whether or not an individual river within the domain is gauged. Each application was run on a daily time-step from 1961 to 2008.

Model comparison

Results from the two model applications were compared with focus on their capability to reconstruct observed time-series of water and nutrients. The variations in both space and time were analysed for the period 1996–2005. Each model application was evaluated separately, and then results for the two applications were compared for the country of Sweden and for two Swedish Rivers. In the study, two types of statistical criteria were used to study model performance; the relative error, RE Equation (1) and, in hydrology, commonly used Nash and Sutcliffe efficiency, NSE (Nash & Sutcliffe 1970; Equation (2)). The RE measures the ability of the model to reproduce a long-term mean value and is expressed as a percentage of the observed mean value (with zero as a perfect fit). The NSE measures the daily error of the model compared with the variance in the observations. A perfect fit corresponds to a value of one and a value larger than zero indicates that the model captures the variance of the observations better than a mean value of the observations. For concentrations, visual inspections were also used due to lack of continuous measured data.

$$\mathrm{RE} = \frac{\bar{M} - \bar{O}}{\bar{O}} \cdot 100[\%] \qquad (1)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} (O_i - M_i)^2}{\sum_{i=1}^{n} \left(O_i - \bar{O}\right)^2} [-] \qquad (2)$$

where M = modelled value (on days with observations) and O = observed value.

Finally, the impact of input data on model results was explored in a sensitivity analysis for the simulations of TN and TP load, and water discharge. This was performed by substituting one dataset at a time between the two model applications. S-HYPE was used as the reference model, to which data was were substituted with input data from the Balt-HYPE model. The following procedures were undertaken as data-sets were exchanged:

1. Agricultural data – information on crop management, e.g. fertiliser applications, substituted with data from the Balt-HYPE model application (Table 1).
2. Lake data – lake depth information substituted with a default value of 10 m depth for all lakes.
3. Point sources – values of emissions from point sources taken from the Balt-HYPE application that are based on population data, consumption and general treatment.
4. Forcing data (P and T) – precipitation and temperature data from the Balt-HYPE application (i.e. the ERAME-SAN grid; Jansson et al. 2007).
5. Balt-HYPE parameters – parameters from the Balt-HYPE application used.
6. Division into high-resolution sub-basins removed – Sweden modelled as 240 main basins instead of 17,313 coupled subbasins.

RESULTS AND DISCUSSION

Comparing model performance

For the whole Baltic Sea basin, the result of the Balt-HYPE model for the period 1995–2005 (Table 2) are of the same magnitude as has been reported in previous studies (e.g. Graham 1999; Pettersson et al. 2000; Lääne 2002; Mörth et al. 2007). Some discrepancies are expected due to the climate during the chosen time period, slightly different model domains and different calculation methods/models. The overall S-HYPE results are also in line with previous reports (e.g. Brandt & Ejhed, 2003; Brandt et al. 2008), although none of the previous studies mentioned are based on homogenous methods, input data and calibration of the whole domain as one single unit.

Table 2 | Annual average water and nutrient load for the period 1995–2005, using the two different HYPE model applications

	Balt-HYPE	Balt-HYPE	S-HYPE
Modelled variable	Baltic Sea basin (1,800,000 km^2)	Swedish surface (476,000 km^2)	Swedish surface (476,000 km^2)
Water discharge (m^3/s)	15,600	6,400	6,280
Nitrogen load (tonnes/yr)	648,000	122,000	106,000
Phosphorus load (tonnes/yr)	29,100	3,370	2,730

Table 2 shows that Balt-HYPE estimated a higher discharge of both water and nutrients than S-HYPE for the Swedish domain. When comparing seasonal patterns it was found that the two applications mainly differed during the month of May (Figure 3). This difference is during the spring flood, where Balt-HYPE gives a much higher estimate in water discharge as well as in TP concentration compared with S-HYPE. This combined effect gives an estimated load of TP more than twice the load from S-HYPE in May. Probably the overestimated spring-flood is an effect of lack of hydropower dams and regulation routines in Balt-HYPE, which then simulates more natural conditions. In the hydropower dams water is stored and released when electricity is needed, which significantly influences the river flow, especially in northern Sweden. During the rest of the year the Balt-HYPE application only gives slightly higher estimates in concentration of both TN and TP than S-HYPE. Nevertheless, on an annual basis Balt-HYPE simulates 16% higher loads of TN and 23% higher loads of TP from Sweden than S-HYPE.

When comparing modelled and observed water and nutrient levels, both model applications show about the same spread in scatter of long-term mean values (Figure 4). The model clearly captures the difference in mean values between large and small rivers, dry and wet regions, and high and low nutrient levels. There is a particularly good correspondence between observed annual data and modelled data for water discharge in S-HYPE. For nutrient concentrations, TN is better captured than TP by the model. It should be remembered that the model was

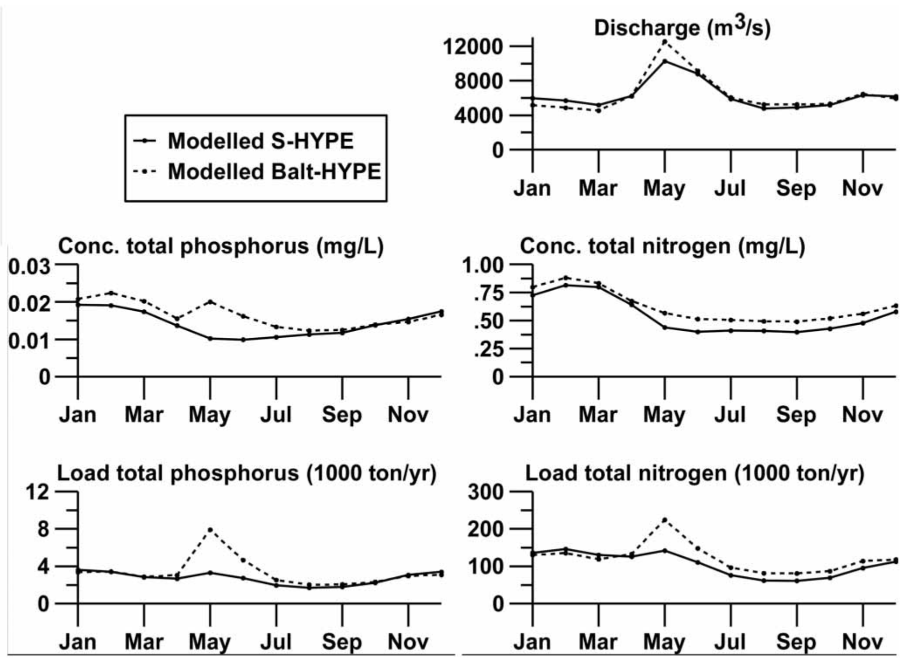

Figure 3 | Modelled water discharge and nutrient load from the Swedish landmass.

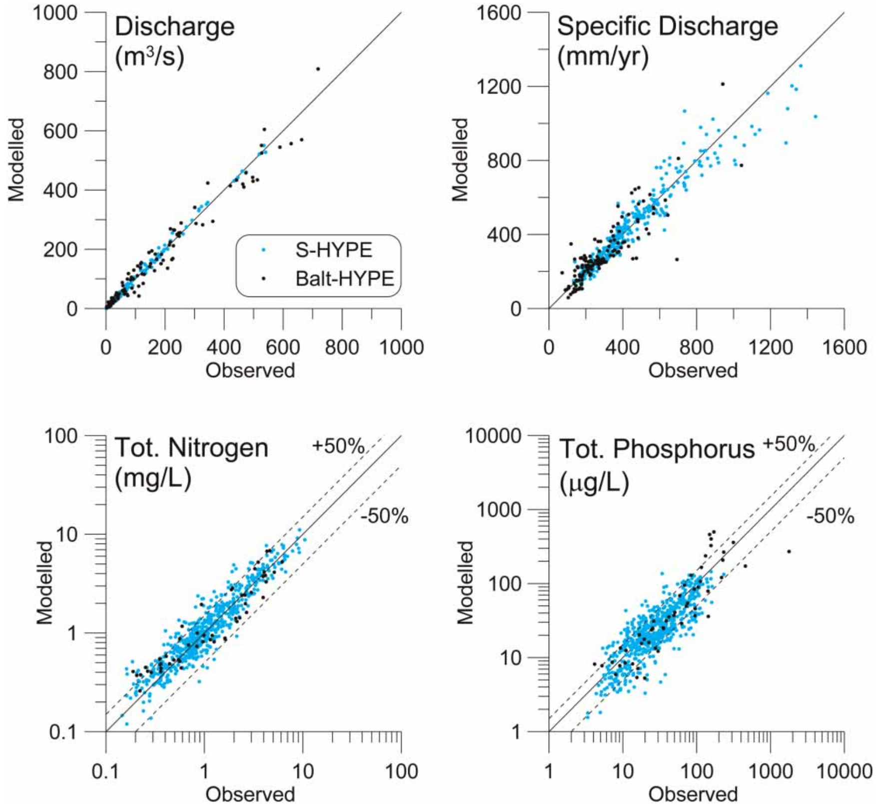

Figure 4 | Mean values of modelled and observed data for all monitored catchments of the S-HYPE and the Balt-HYPE applications, respectively.

calibrated manually using relatively few monitoring sites to receive reasonable results for the whole domain. The present S-HYPE model produces predictions in 17,313 sub-basins, but observations of water discharge and nutrient concentrations are only available in a few hundred sites. Figure 4 can thus be considered to represent ungauged conditions for S-HYPE, as only about 10% of the observations were used for model calibration and the rest for independent model validation. For Balt-HYPE, on the other hand, the few monitoring stations available were all used for calibration of nutrients, while about 75% of the water discharge stations can be considered as independent. Hence, modelled data for ungauged sub-basins, particularly nutrient simulations, are more uncertain using the Balt-HYPE application.

Analyses of REs show that both model applications simulate the long-term volume of river discharge within 25% of the measured volume at nearly all gauging stations compared. The S-HYPE application simulates volumes significantly better, with many gauging stations within 10%. The spatial pattern of model performance indicates that the model captured processes in some regions better than others (Figure 5). For instance, in the Balt-HYPE application, the volumes along the eastern and northern border of the Baltic basin are less well captured by the model. This may be due to the quality of the precipitation grid used as forcing data, but also to the lack of regulation in water power dams. It could also be worthwhile to further evaluate the evapotranspiration algorithm, which in the HYPE model at present relates to temperature, soil moisture and date (Lindström *et al*. 2010).

For the S-HYPE application, the NSE value at most stations was greater than 0.5, with a median value of 0.74 for unregulated basins and 0.55 for regulated basins. The Balt-HYPE application had a median NSE value of 0.51 for unregulated basins, but a median of less than zero for regulated basins. Again, this shows the importance of regulation curves for waterpower dams.

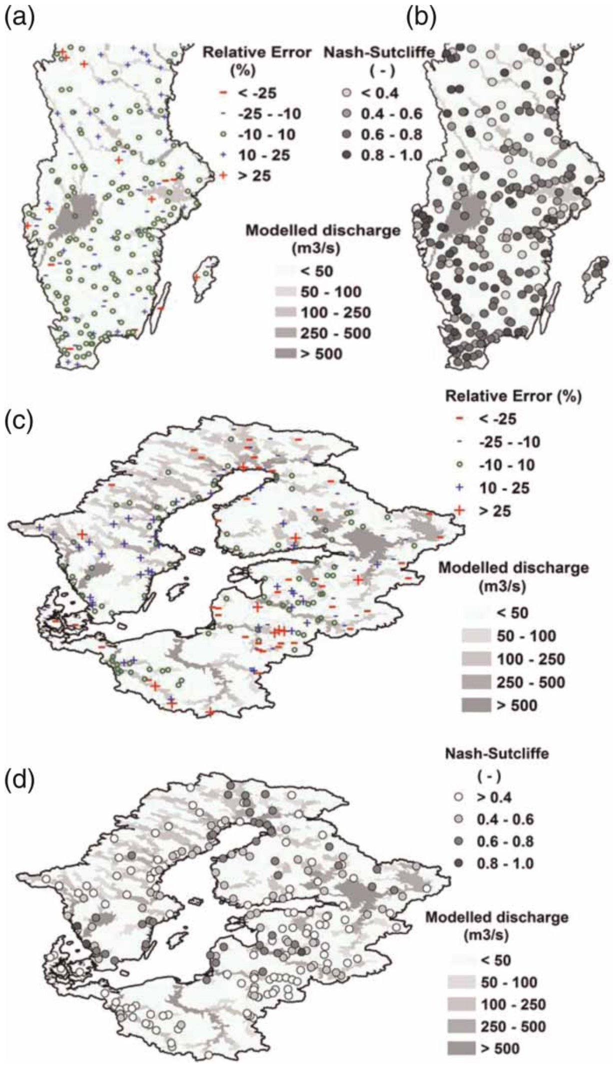

Figure 5 | Modelled performance of water discharge in 348 sites with observations for S-HYPE: (a) Relative error (%), (b) NSE, and in total 256 sites for Balt-HYPE: (c) Relative error (%), and (d) NSE. For S-HYPE only southern Sweden shown for legibility.

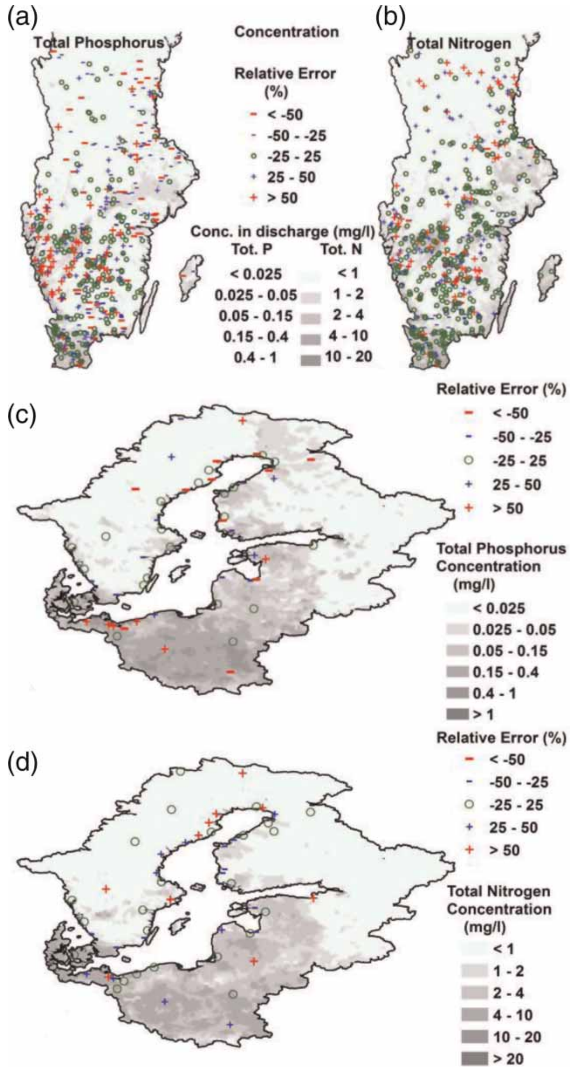

Figure 6 | Relative errors of water quality at monitoring sites, which are in total 672 for S-HYPE: (a) total phosphorus concentrations, (b) total nitrogen concentrations, and 50 sites for Balt-HYPE: (c) total phosphorus concentrations, and (d) total nitrogen concentrations.

Both applications reproduce most mean concentration in TN within 25% of the observed mean values, but TP showed a larger scatter. Three out of four stations showed an RE within 50% for both TN and TP. This is a reasonable result as nutrient concentrations are measured from irregular grab samples, which may not be representative for the general feature of a time-series or even a whole day, whereas the model simulates daily means continuously. For S-HYPE a correlation coefficient was calculated to compare measured and modelled nutrient concentrations and this value was generally above zero for both TN and TP, indicating that the model had some skill in describing the timing of events.

The spatial analysis of model performance of water quality (Figure 6) shows that TP concentrations were less well captured on the Swedish west coast using the S-HYPE. TP is sensitive to rainfall intensity, high runoff and clay soils, which are the conditions in this part of Sweden. Obviously, the influences of these processes are exaggerated in the model as it gives too high levels compared with observations. In contrast to S-HYPE, TP concentrations of the Balt-HYPE were generally underestimated, particularly for rivers in Sweden and Finland. This may be caused by the TP leaching from peat, which has been reported to be high in boreal forests, especially in Finland. TN on the other hand, was often overestimated along the northern Swedish coast. This is probably due to the lack of lake information in Balt-HYPE application, which used another land cover database (Corine vs. Globecover, Table 1) and more assumptions were made for lake depths. This influences the turnover times in lakes and thereby removal by denitrification processes in the model.

When comparing model performance criteria from both applications for the same monitoring sites (Figure 7) it was obvious that S-HYPE shows much higher NSE and lower RE for water discharge than Balt-HYPE. The differences in RE for mean concentrations, however, were less obvious.

Comparing results for two specific rivers

At the river basin scale it is not possible to find enough measurements to capture the spatial variation. Nevertheless, modelled results for the two applications were compared. It is clear from Figure 8 that the Balt-HYPE application gives a much lower concentration of TN in the local runoff compared with S-HYPE, both in the Lagan River and the Emån River. Some similarities can be seen, for example in the north-western part of river Emån catchment where both applications give higher concentrations. Similar results are seen for the area close to the outlet of river Lagan. Even though the concentration of TN in the local runoff differs significantly between the applications, the eventual concentrations in the outlet to the sea are more alike. This situation is probably due to the simulation of denitrification processes in streams and lakes, as mentioned above. Landscape features, particularly lakes, are better described due to the finer resolution; hence, the S-HYPE application simulates more removal of nitrogen by denitrification in the lakes during the transport through the landscape. The impact of model resolution on overall results was further evaluated in the sensitivity study below.

The results on river-basin scale indicate that the coarser resolution Balt-HYPE model can still reproduce reasonable estimates of nutrient concentrations further downstream in river networks. The model performance clearly reflects the calibration method applied. The Balt-HYPE model was only calibrated for river outlets and the results show that observations from sub-basins must be used in the calibration process for more reliable results on the sub-basin scale. Other sources of error might be the data on land cover, agricultural practices and forcing data (e.g. precipitation), the latter influencing mainly phosphorus concentrations. The forcing data are taken from a rather coarse grid. Such a grid delivers mean values for a large region and may thus not reflect events in a sub-basin. This significantly impacts the modelled concentrations as they are dependent on erosion and surface processes, which are event driven.

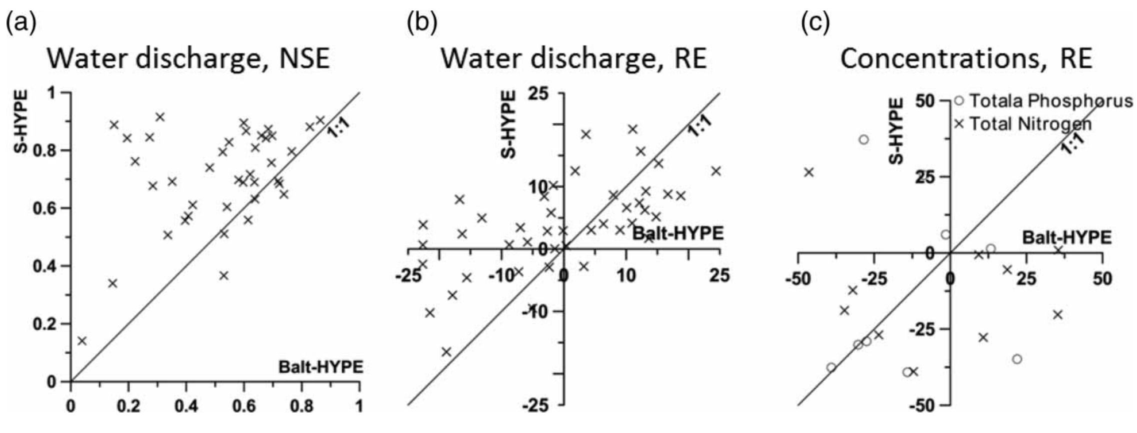

Figure 7 | Model performance of Balt-HYPE vs. S-HYPE in monitoring sites where both models make calculations; (a) Nash–Sutcliffe efficiency (NSE) of water discharge, (b) relative error (%) of water discharge, (c) relative errors (%) for concentrations of nitrogen and phosphorus, respectively.

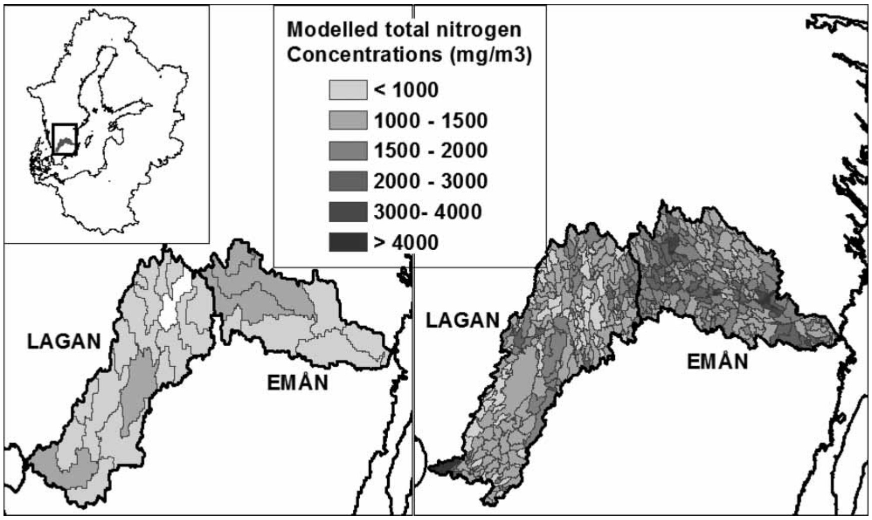

Figure 8 | Comparison of modelled nitrogen concentrations in the (a) Balt-HYPE and (b) S-HYPE applications for Lagan and Emån Rivers in southern Sweden.

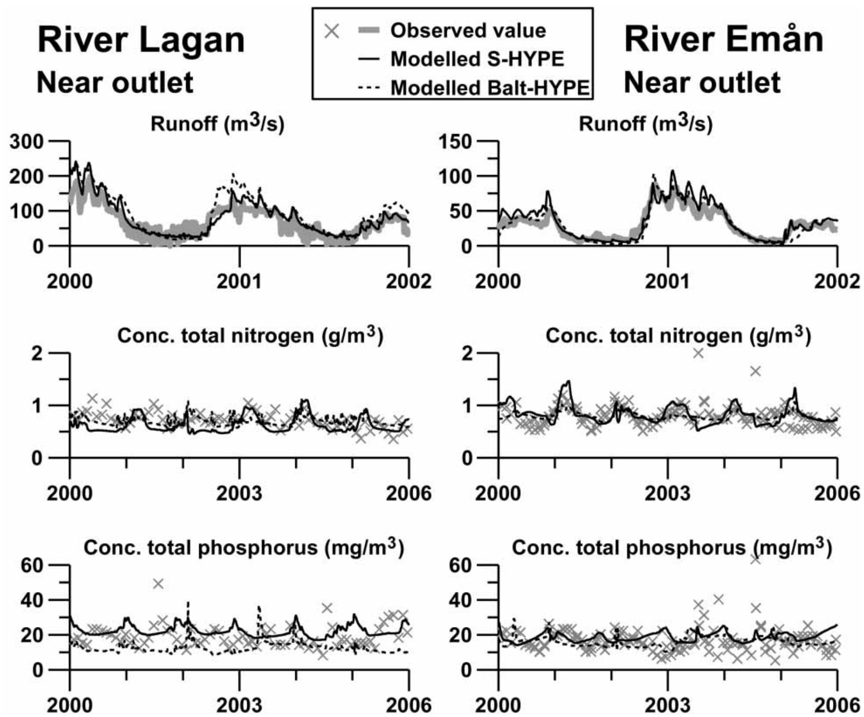

Figure 9 | Observed and modelled results for water discharge, nitrogen and phosphorus concentrations from the S-HYPE and Balt-HYPE applications of the two river basins Lagan River and Emån River, respectively.

Modelled dynamics of water discharge and nutrients were also compared between S-HYPE and Balt-HYPE for the two specific river basins. There was very little difference in the performance of the water discharge simulation for these two rivers, but large differences in how well TN and TP were simulated (Figure 9). The S-HYPE application captured the magnitude of the seasonal variations in TN and TP concentrations as well as the average concentration. The Balt-HYPE application does not manage to capture this seasonal variation, particularly not for TN. This may be expected as the Balt-HYPE application was only calibrated against seasonal and annual averages of TN and TP at river mouths. Nevertheless, the Balt-HYPE model manages to capture the approximate magnitude of the TN and TP concentrations.

Sensitivity study

The differences in model performance between the national (S-HYPE) and the Baltic Sea basin (Balt-HYPE) scale applications can be attributed to either differences in model inputs or differences in calibration (Figure 10). In the Swedish application, more detailed input data on physiography, emissions and meteorology have been used for the higher resolution, while generally available databases and generic methods have been used when modelling the entire Baltic basin (e.g. see data sources in Table 1). The most sensitive factor in the model set-up procedure was found to be parameter values (Figure 10). Using parameter values from Balt-HYPE in the S-HYPE model resulted in 94% higher TP load and 60% higher TN load. This also indicates that parameters are not transferrable between model applications with large differences in input data. TP simulations were also very sensitive to forcing data (i.e. precipitation and temperature), where the dataset with less resolution and thereby fewer extremes resulted in 87% less transport. This is probably an effect of less erosion and less leaching from superficial soil layers when precipitation is not so intense in the model. It highlights how difficult it is to model phosphorous. It also shows that it is important to use precipitation data with adequate resolution of rainfall intensity, and that model parameters have to be revised if the input data are changed. Finally, agricultural data also had a large impact on the simulated TP load, followed by resolution and lake depths, which influence the magnitude of modelled lake residence times and sedimentation processes, which may occur in rivers and lakes in each sub-basin.

The outcome of changing the model resolution is difficult to interpret. Many factors are involved here with compensating or strengthened results in the experiment, and it is hard to judge their internal influence without a more careful analysis of each separately. For instance, (i) average precipitation and temperature were used for entire rivers basins, (ii) each land-use type was aggregated within the sub-basin independently of landscape location, and (iii) lakes, streams and point-sources were aggregated for a whole river. It is interesting to note, however, that only the nutrient simulations were sensitive to the resolution factor and not water discharge simulations. Hence, only water quality issues demand such detailed landscape analysis as in the S-HYPE application, where specific positions of

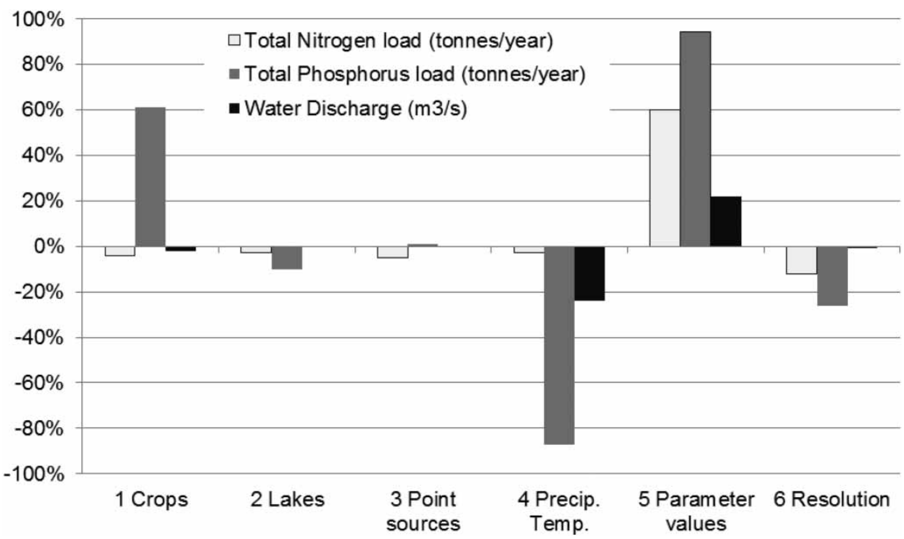

Figure 10 | Sensitivity analysis with one data set substituted at the time, showing long-term annual results for nitrogen load, phosphorous load, and water discharge, from the whole of Sweden to the Baltic Sea. S-HYPE is the reference model, to which data have been substituted with data sets from Balt-HYPE. Bars represent changes of model results in percentages.

landscape features and emissions are combined with forcing data.

Simulation of water discharge was, on the other hand, sensitive to a shift in forcing data, using the same calibrated parameter values. The coarser precipitation and temperature grid of Balt-HYPE gave 24% less water using the S-HYPE parameters. Accordingly, the parameter values from Balt-HYPE gave 22% more water. Most probably the model calibration thus compensated for any bias in the forcing data. It is well known that there is a strong interaction between input data and calibrated parameter values. To sum up, the sensitivity study indicates that, for improving model performance, efforts should be focused on better data for precipitation/temperature, agriculture and model parameter values.

Usefulness and limitations of the two model applications

The results show that it is possible to obtain useful information about water resources using homogenous methods and input data across political and geophysical boundaries. Data from the whole model domain are applied to also make predictions for ungauged basins within the domain. Such data are not yet available, e.g. from national environmental protection agencies or the European Environmental Agency. The new information is useful for mapping large areas with high-resolution (e.g. for the WFD), for regions where no other hydrological model is available, or to consider uncertainties by using a model ensemble. When validated, the models may be used for regional assessments, such as characterisation of status, forecasts, or 'what if' scenarios with changed conditions. This information should be very valuable for both water authorities and large-scale strategic agreements of water-related issues (e.g. for the MSFD).

The exercise of applying the same model using different datasets clearly shows how model performance is affected by the application itself. Both applications managed to simulate water discharge with rather good agreement between model data and observations; however, for water quality the dynamics were not captured with the same precision on the sub-basin level using the large scale application. This implies that the internal processes in the large-scale model application are not described correctly. This has to be checked and corrected before using the model for scenario estimates. Thus, model results so far can be used for characterisation of leaching to the Baltic Sea and probably also for near-time forecasts, but caution should be taken when using the results for high-resolution prediction in time and space or for simulation of scenarios, as these are dependent on trustworthy process description in the model application.

One way to handle the problems of proper model validation is to use a model ensemble to illustrate the range of uncertainty in results. This is applied for climate models (e.g. IPCC 2007) and meteorological model input to hydrological forecasts (e.g. Nobert *et al.* 2010; Arheimer *et al.* 2011b). Ensemble means from several hydrological models have actually been found to give higher performance criteria than each single model used (Viney *et al.* 2005), and recently this has also been applied to water quality models (Exbrayat *et al.* 2010). Hopefully, the HYPE applications will be useful model members in larger model ensembles for strategic decision-making both in Sweden and in the Baltic Sea region.

The applications presented here are still undergoing development and testing. The Balt-HYPE application will be updated with more detailed agricultural and point source input data as such data become available, and it will be calibrated and validated using the Baltic Nest Institute data and the EWA (GRDC 2009b) water discharge database. The model code is also continuously updated and improved. The ambition is also to launch the HYPE model as an open source code, to make better use of the excellence of several research groups and initiate a collaborative modelling community among European hydrologists (Arheimer *et al.* 2011c). The different HYPE applications could be useful as hydrological laboratories, also for researchers to test hypothesis and specific algorithms on the large scale.

CONCLUSIONS

Semi-distributed and process-based hydrological model applications are now available for Sweden (28 km^2 resolution) and the Baltic Sea basin (400 km^2 resolution), using homogenous methods and input data across political and geophysical boundaries. The comparison of model performance shows that:

- Both model applications can serve as useful tools for providing spatial information of water and nutrients; they capture regional variations in mean values between large and small rivers, dry and wet regions, and high and low nutrient levels. However, caution should be exercised when using the nutrient concentrations (especially from the Balt-HYPE model) for ungauged basins, upstream catchments and scenarios.
- Differences in model set-up were reflected in the simulation of both spatial and temporal dynamics. Model performance is clearly better for S-HYPE regarding discharge, while differences in relative errors for nutrient concentrations are less significant. Dynamics in nutrient concentrations were less distinct for specific rivers using the lower resolution of the Balt-HYPE model. For the whole country of Sweden, the spring flood was more pronounced using the coarse resolution and global databases without hydropower regulation, which resulted in about 20% higher simulated nutrient loads.
- A higher spatial resolution is needed for water quality simulations than for water discharge simulations. In particular, the phosphorous simulations were very sensitive to the choice and resolution of the input data.

Site-specific knowledge and data are needed for further development of the models. In particular precipitation/temperature, agriculture and model parameter values should be considered more carefully in the set-up procedure. Future international collaboration is desired to provide an ensemble of models for strategic international agreements in the Baltic Sea region.

ACKNOWLEDGEMENTS

The study was performed at the SMHI Hydrological Research unit, where extensive similar work is carried out, taking advantage of previous work and several projects running in parallel in the group. Hence, input from more people other than the authors was essential for the basis of this study, and we would especially like to acknowledge contributions from Kristina Isberg, Johanna Nilsson, Charlotta Pers, Jörgen Rosberg and Wei Yang. The funding was mainly provided by the Water Management program at SMHI, the Bonus project Ecosupport, and the EU FP7/GMES project Geoland 2, to which we all wish to express our sincere thanks. Finally, we would like to acknowledge the GRDC, BALTEX and SLU for providing observed data. Balt-HYPE results are now available for free downloading at www.smhi.se\balt-hype and S-HYPE data at www.smhi.se/tema/Vattenforvaltning.

REFERENCES

Arheimer, B., Dahné, J., Lindström, G., Marklund, L. & Strömqvist, J. 2011a Multi-variable evaluation of an integrated model system covering Sweden (S-HYPE). *IAHS Publ.* **345**, 145–150.

Arheimer, B., Lindström, G. & Olsson, J. 2011b A systematic review of sensitivities in the Swedish flood-forecasting system. *Atmos. Res.* **100**, 275–284.

Arheimer, B., Wallman, P., Donnelly, C., Nyström, K. & Pers, C. 2011c E-Hypeweb: service for water and climate information – and future hydrological collaboration across Europe? In: *International Symposium on Environmental Software Systems (ISESS), IFIP Advances in Information and Communication Technology* (J. Hřebíček, G. Schimak & R. Denzer, eds). Springer, Heidelberg, 359, pp. 657–666.

Beven, K. & Binley, A. M. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Proc.* **6**, 279–298.

Brandt, M. & Ejhed, H. 2003 TRK-Transport, Retention, Källfördelning. Belastning på havet. Swedish Environmental Protection Agency, Report No. 5247.

Brandt, M., Ejhed, H. & Rapp, L. 2008 Näringsbelastning på Östersjön och Västerhavet 2006: underlag till Sveriges PLC5-redovisning till HELCOM. Swedish Environmental Protection Agency, Report 5815.

Bouwman, A. F., van Drecht, G. & van der Hoek, K. W. 2005 Global and regional surface nitrogen balances in intensive agricultural production systems for the period 1970–2030. *Pedosphere* **15** (2), 137–155.

Britz, W., Pérez, I., Zimmermann, A. & Heckelei, T. 2007 Definition of the CAPRI Core Modelling System and Interfaces with other Components of SEAMLESS-IF. SEAMLESS report no. 26, January 2007.

Donnelly, C., Strömqvist, J. & Arheimer, B. 2011 Modelling climate change effects on nutrient discharges from the Baltic Sea catchment. *IAHS Publ.* **348**, 1–6.

Donnelly, C., Dahné, J., Rosberg, J., Strömqvist, J., Yang, W. & Arheimer, B. 2010 High-resolution, large-scale hydrological modelling tools for Europe. *IAHS Publ.* **340**, 553–561.

Donnelly, C., Dahné, J., Lindström, G., Rosberg, J., Strömqvist, J., Pers, C., Yang, W. & Arheimer, B. 2009 An evaluation of multi-basin hydrological modelling for predictions in ungauged basins. *IAHS Publ.* **333**, 112–120.

EIONET 2003 Central Data Repository. Available from http://cdr.eionet.eu.int. (accessed 14 February 2009)

Exbrayat, J.-F., Viney, N. R., Seibert, J., Wrede, S., Frede, H.-G. & Breuer, L. 2010 Ensemble modelling of nitrogen fluxes: data fusion for a Swedish meso-scale catchment. *Hydrol. Earth Syst. Sci.* **14**, 2383–2397.

FAOSTAT 2006 Food and Agriculture Organisation (FAO) of the UN, Rome. Available from www.apps.fao.org. (accessed 14 March 2009)

Graham, L. P. 1999 Modelling runoff to the Baltic Sea. *Ambio.* **28**, 328–334.

GRDC – Global Runoff Data Centre 2009a *European Water Archiver (EWA)/Global Runoff Data Centre*. Federal Institute of Hydrology (BfG), Koblenz, 2009.

GRDC – Global Runoff Data Centre 2009b *River Discharge Data*, Federal Institute of Hydrology (BfG), Koblenz, 2009.

Goldewijk, K., Beusen, A., de Vos, M. & van Drecht, G. 2011 Holocene=Anthropocene. The HYDE database for integrated global change research over the past 12,000 years. Global Ecology and Biogeography. *Hydrolog. Sci. J.* (in press Feb 2012).

HELCOM Secretariat 2007 Towards a Baltic Sea unaffected by Eutrophication. Helsinki. Available from www.helcom.fi/. (accessed 13 January 2010).

IPCC 2007 *IPCC Fourth Assessment Report: Climate Change 2007 (AR4)*. IPCC, Geneva, Switzerland.

Jansson, A., Persson, C. & Strandberg, G. 2007 2D meso-scale re-analysis of precipitation, temperature and wind over Europe – ERAMESAN Time period 1980–2004. SMHI Reports: Meteorology and climatology no. 112.

Johansson, B. 2002 *Estimation of areal precipitation for hydrological modelling in Sweden*. PhD Thesis, Earth Sciences Centre, Dept Phys. Geog., Göteborg University. Sweden.

JRC 2006 European Soil Database. Available from http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDB/index.htm (accessed 17 February 2009)

Klemes, V. 1986 Operational testing of hydrological simulation models. *Hydrolog. Sci. J.* **31**, 13–24.

Lane, S. A. & Richards, K. S. 2001 The 'validation' of hydrodynamic models: some critical perspectives.. In: M. G. Anderson & P. D. Bates (eds) *Model Validation Perspectives in Hydrological Science*. John Wiley & Sons, Ltd, pp. 413–438.

Lääne, A., Pitkänen, H., Arheimer, B., Behrendt, H., Jarosinski, W., Lucane, S., Pachel, K., Räike, A., Shekhovtsov, A., Svendsen, L. M. & Valatka, S. 2002 *Evaluation of the implementation of the 1988 Ministerial Declaration regarding nutrient load reductions in the Baltic Sea catchment area*. The Finnish Environment Report No. 524. The Finnish Environment Institute (FEI), Helsinki, pp. 175–189.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. & Arheimer, B. 2010 Development and test of the HYPE (Hydrological Predictions for the Environment) model – a water quality model for different spatial scales. *Hydrol. Res.* **41** (3–4), 295–319.

Mörth, C. M., Humborg, C., Eriksson, H., Danielsson, Å., Rodriguez Medina, M., Löfgren, S., Swaney, D. P. & Rahm, L. 2007 Modeling riverine nutrient transport to the Baltic Sea: a large-scale approach. *Ambio* **36** (2–3), 124–133.

Nash, J. E. & Sutcliffe, J. V. 1970. River flow forecasting through conceptual models. Part I. A discussion of principles. *J. Hydrol.* **10**, 282–290.

Nobert, S., Demeritt, D. & Cloke, H. 2010 Informing operational flood management with ensemble predictions: lessons from Sweden. *J. Flood Risk Manage.* **3**, 72–79.

Pettersson, A., Brandt, M. & Lindström, G. 2000. Application of the HBV-N model to the Baltic Sea Drainage Basin. *Vatten* **56**, 7–13.

Refsgaard, J. C., Storm, B. & Clausen, T. 2010 Système Hydrologique Européen (SHE): review and perspectives after 30 years development in distributed physically-based hydrological modeling. *Hydrol. Res.* **41**, 355–377.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á. & Wind, P. 2012 The EMEP MSC-W chemical transport model – Part 1: Model description. *Atmos. Chem. Phys. Discuss.* **12**, 3781–3874.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C. & Lindström, G. (accepted) Water and nutrient predictions in ungauged basins – set-up and evaluation of a model at the national scale. *Hydrolog. Sci. J.* (February 2012, in press)

USGS (US Geological Survey) 2000 Hydro1k Elevation Derivative Database. Available from http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html (accessed 16 February 2009).

Viney, N. R., Croke, B. F. W., Breuer, L., Bormann, H., Bronstert, A., Frede, H., Gräff, T., Hubrechts, L., Huisman, J. A., Jakeman, A. J., Kite, G. W., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M. & Willems, P. 2005 *Ensemble Modelling of the Hydrological Impacts of Land Use Change*. International Congress on Modelling and Simulation 2005, Melbourne, Vic., Aust., pp. 2967–2973.

Wulff, F., Humborg, C., Rodriguez Medina, M., Mörth, M., Savchuk, O. & Sokolov, A. 2009 *Revision of the country allocation of nutrient reductions in the Baltic Sea Action Plan, Section A: Hydrological adjusted riverine loads and atmospheric loads from different countries averaged for 2000–2006*. Technical Report Series No.1, Compilation June 5, 2009. Stockholm, Baltic Nest Institute.

First received 12 January 2011; accepted in revised form 25 June 2011. Available online 31 January 2012