Process-based eco-hydrological modelling of nutrient loads to the Baltics from three Estonian watersheds

Working Group Final Report

Project NORRA: Development of data-modelling system and the decision support tool for the integrated marine and inland water management

Csilla Farkas[a], Rain Elken[b], Juan Manuel Garcia Diaz[b], Anatoli Vassiljev[c], Attila Nemes[a], Johannes Deelstra[a], Andreas Porman[b], Tiia Pedusaar[b],Per Stålnacke[a], Peeter Ennet[b],

[a] Norwegian Institute of Bioeconomy Research

[b] Estonian Environment Agency

[c] Tallinn University of Technology

Estonian Environmental Agency

Ås, Tallinn, 2017

Contents

General overview, the aim of this Working Group report

It is now evident that fresh- and marine-water ecosystems have long been overloaded by nutrients that originate, among others, from land-based agriculture, and the manifold point-type anthropogenic sources, such as waste-water treatment plants. The situation in this sense is rather acute in the Baltic countries and their waters. This situation will not improve until we identify the true sources and pathways of pollutants and quantify their contribution, advance our process-understanding to describe the mechanisms through which the pollution occurs, and by what steps we may be able to influence that, and quantify what, if any such steps – today popularly termed 'mitigation measures' – may produce certain levels of benefit sustainably.

Various advancements have already been made in a number of the above steps, but the applied mitigation measures are often local and scarce, their effects are often influenced by other unknown and uncontrolled factors, and their installation is too recent to yield noticeable changes due to e.g. the nutrient retention characteristics of the ecosystem. It is therefore of paramount importance that we monitor and use environmental data towards long-term planning in order to sustain or improve the state of our environment. Environmental modeling is a growingly important tool for future planning. One main branch of such modeling – the use of process-based models – is a generally data intensive, but in exchange rather detailed way of quantifying natural processes.

As part of the greater context of the NORRA project, this Working Group reports on its work towards (a) cataloguing existing environmental data for selected pilot-watersheds of Estonia that are needed to run process-based simulation models for those watersheds; (b) identifying suitable, and feasibly usable models, and calibrate, validate and test them under Estonian conditions; (c) evaluating the models' performance and capability to become decision support tools; and (d) and advising on data and knowledge gaps towards further future progress.

Introduction to eco-hydrological catchment modeling

Background to simulation modeling in this project

Decision makers actively seek assistance from environmental information systems to be informed of the status of environmental resources and variables, and to help assist policy-making. Such information systems rely on long-term field monitoring data, as well as data from computer-based models of various complexity, among them numerical simulations models.

With recent advancements in computing facilities and in the development of a range of models, numerical simulation modeling is increasingly becoming the tool of choice when it comes to assessing the anticipated impact of certain natural or human-induced changes to/in our environment (Kværnø et al., 2013). As the understanding of natural processes by the scientific community keeps improving, this improvement continually translates into a better ability to quantify those processes by such models (Deelstra, 2014). Changes that can be addresses by simulation studies include a wide range of scenario and impact assessment studies driven by e.g. planned land-use or land management changes (Farkas et al., 2013), industrial, urban infrastructure and other facilities planning, etc.

This report summarizes the work of a working group within the NORRA project, aiming to assess the suitability of different dynamical models to describe Estonian eco-hydrological conditions. It has also been an expressed goal that the participating groups attempt to improve understanding of surface, subsurface and in-stream processes that are most relevant in Estonia, and collaboratively try to identify the main constrains and future tasks of applying dynamical eco-hydrological models in Estonia.

In this project, we used 5 different models in parallel to achieve the established goals, and simulated the water and nutrient-transport and loads in 3 selected watersheds of Estonia. We first summarize some important aspects and limitations of eco-hydrological modeling in general, and some specifics of the multi-model approach that we have taken.

Overview of eco-hydrological catchment models – benefits and constrains

The continuous dynamic models that consist of mathematical descriptions of physical, biogeochemical and hydrochemical processes, and combine significant elements of both physical and conceptual semi-empirical nature can be called process-based eco-hydrological models. An eco-hydrological process-based model for a river catchment necessarily contains a hydrological module as a basic feature. Another necessary part is a vegetation and soil sub-model. Also, such a model usually includes the sub-models for biogeochemical cycles (mainly nitrogen and phosphorus) with a certain level of complexity. The hydrological, vegetation and biogeochemical sub-models are usually coupled in order to include important interactions and feedbacks between the processes, like water and nutrient drivers for plant growth, water transpiration by plants, nutrient transport with water, etc. Usually, vertical and lateral fluxes of water and nutrients in catchments are modelled separately, whereas meteorological parameters are used as external drivers.

It is an important dilemma, how detailed an eco-hydrological model should be at the catchment scale. Model complexity by itself should not be seen as a binding necessity. Often, a complex phenomenon or process can be described mathematically in a simplified form and parametrized using more easily available information. When that is the case, using a simple model is preferable compared to using a complex model and a great degree of detail in describing and parameterizing the process, which is often problematic, lacks proper data, and the control of the overall model behavior is difficult.

Different models operate at intrinsic spatial-temporal resolutions that they were designed for. The spatial resolution, scale of application, and objective of the study are interrelated. Processes that are evidenced at some smaller scales may behave differently at larger scales. Therefore information obtained from experiments and observations at a small temporal or spatial scale cannot directly and automatically be transferred to larger scales. Similarly, large-scale observations cannot be used directly for small-scale simulations. As a rule of thumb, the deeper one intends to examine a phenomenon - and thus intends to go to smaller modeling scales - the more parameters will be required to describe processes, which may be an overwhelming task and may result in massive and uncontrollable uncertainties.

It is possible to classify eco-hydrological models by their spatial and temporal resolution, which will help the user in identifying the desirable model in this aspect (
Figure 1).

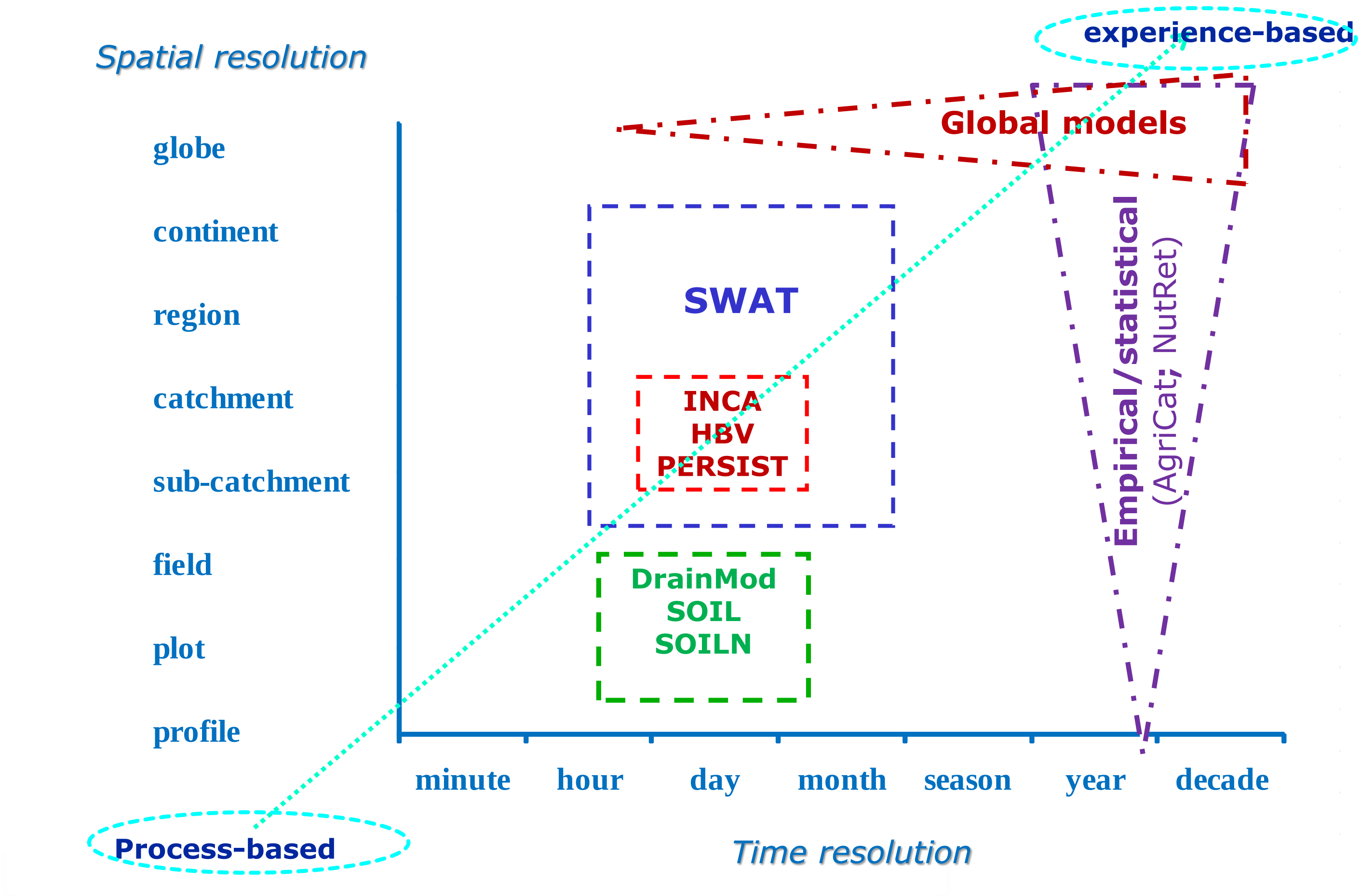


Figure 1. Spatial-temporal resolution of a select list of eco-hydrological simulation models

Benefits and constraints of using multiple models in parallel

Ensemble modeling – i.e. the use of a number of models in combination – is a natural part of weather and climate modeling today. This practice, however, has not set foot yet in environmental modelling, although the research area of estimating soil hydraulic properties as input to numerical simulation models now growingly applied one or another version of such techniques (Baker and Ellison, 2008). Ensemble modeling carries a number of benefits and potential over the use of a single model.

Models can differ in their theory and structure, but also in the information that they require. As a result, their sensitivity, scale of support and scale of command may also differ. Their use is easy to justify if it is difficult to determine which, if any, single model may be superior to others. In ensemble modeling, the main aim is not to make the single model perfect, but to capture the trend that multiple models agree on. The ensemble will amplify trends that are common among models, while by-chance predictions will be softened. The outputs, therefore, can be interpreted – qualitatively or quantitatively - as a measure of uncertainty.

In soils related predictions, two different types of ensemble models have been explored. Guber et al. (2009) used a bag of 19 published pedotransfer functions in an ensemble prediction scheme to parameterize an eco-hydrological model that solves the Richards' Equation to calculate water flow. In their scheme, they used different models that required different sources and levels of input and that had different structure. A more popular approach – and one that is more simple to implement – is the use one of several schemes to resample data of the main data pool, and use those to develop a given number of predictive models of the same structure, which are then statistically pooled to give a prediction – optionally with a measure of uncertainty. Such schemes include e.g. bagging and bootstrapping, and have been used by several authors in the field of helping to parameterize simulation models (e.g. Schaap et al., 2004; Nemes et al., 2010). Although it technically belongs to the latter type of ensemble modeling, Monte-Carlo simulations, or similar techniques may also serve the purpose, since the technique involves parameterizing multiple model runs slightly differently as selected model parameters are recurrently sampled from a pre-determined frequency distribution. Such techniques are typically applied in parameter sensitivity and uncertainty analysis.

The first type of ensemble modeling is rare, since it typically involves an excessive amount of work that cannot be automated like the latter ensemble types (data or parameter resampling). Their specific value is in that multiple model types – and the trends that they produce - can be compared, and the findings are not conditional on having to pre-selecting and accepting a model concept or structure. Very often, however – and this was the case in this project – different models often evaluate environmental metrics of interest differently, and the results may not be easy to match. In such cases, conversions, expert-interpretations, or additional scaling may be necessary. Apart from the benefit of the ability to potentially discard (an) outlying model(s), each approach has the inevitable extra benefit of being able to produce some metric of uncertainty to the output, which can then be expert-interpreted, and potentially propagated further to additional studies or models, if desirable.

User-bias

While objective metrics are used when calibrating and validating simulation models, it is inevitable that users make choices in e.g. model parameterization that are not standard choices, but are somewhat influenced by their personal preferences. Such can be the case with e.g. various resembling parameter sets that may yield very similar model results. Different modelers may reason differently why one parameter set or the other should be preferred to reach the same goal. The background of the modeler may also have some influence on the model parameterization, in that a crop scientist will likely be more knowledgeable and comfortable with adjusting crop related parameters, while a soil scientist may do the same with soil parameters. This aspect may have significance if different groups work in parallel, and if some or all models are not-auto-calibrated using the same initial parameterization.

Unless some obvious discrepancy is found, we see no reason to deem any alternate solutions insufficient if otherwise their simulation metrics are up to standard, and the calibration/validation that was performed covers all aspects of the expected use of the model.

Nevertheless, we discussed any potential aspects in this study that may be related to user bias. In this project we first addressed this matter via awareness, i.e. the recognition that this factor exists and may have some impact on the outcome of the modeling results. Second, we performed the relevant modeling exercises involving more than one researcher in the work-flow. Most model simulations were either performed jointly, or were followed up by consulting with (an)other modeling expert(s) in the project. This has been achieved through recurrent meetings between researchers of the different institutions.

On data-driven limitations

Simulation modeling studies, especially when multiple models have to be parameterized, inevitably face the situation of missing information, whether those are model parameters (e.g. site- or area-specific constants or soil parameters), or driver variables (weather data, etc.). Such missing information can originate from multiple sources, that include but are not limited to e.g.:

- Routine or targeted data collection normally taking place at a different scale (e.g. water outflow from fields)
- Differences in methodology or standards between the model and the area/country of application (e.g. different soil particle-size distribution standards)
- The information not being readily available from the information source

It can also happen that some of such inputs have simply not been determined, when e.g. particular data collection has not been in the focus before, or lack of resources did not allow proper data collection or monitoring.

Lack of data can pose different degrees of difficulty to the modeler to overcome, and will require different strategies to overcome. A user can opt to consult model documentation for reference values – which are somewhat generic – or can look up earlier case studies for values used under similar circumstances. In certain cases, interpretation or up/down scaling of information from studies at other scales can provide useful information towards parameterizing a model. This step involves expert judgement as a resource that modeling studies often have to rely on. Expert knowledge and understanding of underlying physical processes and/or of the equations behind the modeled processes (that are typically simplifications of reality) can very often yield sufficient information to set certain model parameters satisfactorily. Of course, it is important to emphasize that this should be done in consultation with experts familiar with the local conditions. Simulation modeling based studies typically encounter several of the listed problems, and the solution is eventually found using a combination of tools. In this report, examples of such data inferences – and how they were addressed – are provided at the appropriate sections.

The consequence of having to use such solutions to fill data gaps is the increased risk of introducing uncertainties into the study by propagating both random and non-random (systematic) errors. The user is advised to experiment with the model to learn about the model's sensitivity to the setting of different parameters, in order to be informed about greater or lesser risks of error due to uncertainty in the parameterization that can be interpreted later. Some model developers assist the user by providing a tool that is suitable for this task, while in other cases the modeler needs to run trial-and-error simulations with alternative parameterizations.

Advantages and disadvantages of multi-model simulations have been introduced earlier. Even if it is a costly and labor intensive approach, using multiple models to address the same problem can help the scientist – and eventually the policy maker – to identify likely trends and the unlikely outliers, which may be the result of imperfections in parameterizing a model. This approach is now actively used to the benefit of society in the form of forecasting daily weather, or the chances of dangerous weather extremes. Our study adapted the multi-model simulation approach for the same reason, i.e. to help reduce the potential risk of relying on a single model and its parameterization.

General limitations in eco-hydrological modeling

The regionalization of models has been a recurring theme in the atmospheric and hydrological sciences over the last few decades. Such studies are an inevitable part of building a pro-active rather than re-active approach to responding to projected climatic changes, or other changes such as changes in land use or land cover. Models and modelling results in general are continually improving, but still remain limited in many ways, due to a combination of many factors, such as e.g. inadequate or unreliable environmental or other support data and limited funding available to collect more data; or limited (quantitative) knowledge of natural processes and their limited representation in simulation models.

It has long been a dilemma whether models or their support data are the limiting factor in a study. Models are being developed and fine-tuned constantly by their developers. It is generally seen that model performance will greatly depend on how well it is parameterized, how detailed support data are available for it to be calibrated on, and how efficient its calibration was. In the vast majority of large-scale modeling applications there are gaps in data availability – and it was also the case in this study. A number of modules in the simulation models had to be generically parameterized, lacking more detailed and/or local data. An example is the hydraulic characterization of the soils used. It usually yields a more efficient investment of resources to collect additional support data in such cases, than to invest in a more complex model, since the data are likely the limiting factor in the quantitative characterization of our knowledge.

Model calibration and validation is another extremely important step of the modeling process, which, unfortunately requires significant resources and effort in that detailed field data should be collected. Examples of such data collection are given by Iital (2005) or Bechmann and Deelstra (2013), presenting methodologies and results of environmental monitoring programs of small agricultural catchments in Estonia and Norway, respectively. In Estonia, runoff and nutrient loss data have been collected in small agricultural catchments. This, however, recently has been stopped, even though the size of these catchments is suitable for model calibration/validation since processes like nutrient retention are less dominant compared to larger catchments. Once calibrated for the smaller catchments, the same parameter settings could be used to model surface and subsurface runoff and nutrient loads for larger catchments. Another important aspect, especially when it comes to the simulation of nutrient losses from agriculture-dominated catchments, is to take into consideration the dominant flow paths. For example if subsurface drainage systems exists for creating optimal cropping conditions, those are undoubtedly important pathways for runoff and nutrients. Therefore, simulation models for such cases have to be able to properly simulate these processes. In addition, natural drainage may also be an important flowpath for water and nutrients.

Data collected in monitoring programs is extensive and at the same time costly; but of good enough quality to be used for modelling purposes. The availability of validation data is often in connection with the general economical situation, and the availability of resources for the subject in question. Lack of validation data is more often a problem in developing countries than in Western countries, although this is not a hard rule. Our study was somewhat limited in model validation terms – as it was detailed in each of the relevant modeling reports. While some validation did take place, such efforts should be extended in the future. For the current application, the approach of interpreting the simulation results in relative terms and drawing conservative conclusions should help reduce the risks posed by a lack of extensive model validation.

Data used as input to each of the simulation models have their own intrinsic uncertainties, originating from e.g. (a) their natural variability, b) the timing and sensitivity of our measurements and (c) the used techniques of data collection. These are all sources of error and uncertainty. In studies where future scenarios are generated, the uncertainty of extrapolations using imperfect models also have to be recognized. Additionally studies like this build on successive layers of simulation modeling, where the uncertainty of the output of one layer is passed on to the next layer of modeling (e.g. hydrology models and nutrient-transport models). It takes a large effort to address, quantify and reduce such uncertainties, which is a problem that is rarely addressed sufficiently. One way of quantifying certain sources of uncertainties is to work with distributions of stochastic data, rather than to use any chosen (mean) value. This can be done in e.g. a Monte-Carlo simulation scheme, but using such a technique was beyond the means of this project.

There is also a limitation introduced to simulation based scenario studies by not being able to assume a number of potential future changes. This study, for example, incorporates our current knowledge on the type and distribution of land-use, and the use of current agro-techniques. Changes to any of those factors may yield significant changes when their effects are up-scaled to the regional or national level. Therefore, it is desirable, when establishing new policies, subsidy-systems, etc. to (re-)evaluate the effect of any such factors, when improved information becomes available.

Apart from some general limitations seen by the simulation based studies, a number of specific limitations are recognized – and partially listed in the relevant section later. We briefly cite two examples here. The limited availability of measured discharge data as well as ground-water data limited the calibration and validation of the SWAT model for the water availability simulations. While we attempted to eliminate any potential biases to the best of the modelers' knowledge during model calibration, the resulting modeling uncertainty still has to be factored in. In terms of the applied modeling techniques, certain limitations are present in the moisture regime module of some of the applied models that can to be noted. While the field-capacity approach to approximate soil water transport (aka 'bucket'-type model) is accepted and frequently used by model developers and users, this approach is not able to account for the potential benefit of upward capillary rise from the ground-water.

Simulated metrics

There were four metrics that were of concern in this project, and hence were variables of interest. First and foremost, in order to be able to simulate the transport of nutrients and sediments, the hydrological balance of the studied catchments had to be simulated reasonably well. The primary metric that is typically field-observed and can also be simulated is the outflow from a river catchment. If outflow is not successfully simulated, typically the other metrics in question will not be simulated successfully either. Suspended sediment (SS), NO_3-nitrogen (NO_3-N) and total phosphorus (TP) concentrations were then subsequently simulated and assessed. These metrics are of main concerns when it comes to water quality issues related to nutrient/pollutant loads from land-based agriculture. Water discharge data was available on a daily basis in units of m^3/sec, while information on SS, NO_3-N, TP was available on a monthly or bi-monthly basis in units of mg/L.

Materials and Methods

Study areas

Estonia's general geographical and climatic characteristics

Estonia consists mainly of lowlands bordered by the Baltic Sea, Latvia, and Russia. It has numerous lakes and many rivers where water bodies in total comprise approximately 5% of the area of the country. Estonian rivers are typically short with a small catchment areas and therefore relatively scarce in water. The river system, however, is dense. In terms of drainage, different Estonian rivers are divided into four natural river basin districts: Narva-Peipsi river basin district, the Gulf of Finland river basin district, the Gulf of Riga river basin district, and the river basin district of islands.

There are 10 rivers longer than 100 km. The longest is the Võhandu River – 162 km, then the Pärnu River – 144 km. These are followed by Põltsamaa, Pedja, Kasari, Emajõgi and Keila River. 15 rivers with catchment areas greater than 1000 square kilometers exist, whereas the entire catchment area of the Narva River is greater than the territory of the Republic of Estonia. The catchment area of the River Emajõgi, located almost in its entirety in Estonia, forms 22% of the country's territory. The river with highest fall is Piusa; the elevation difference of its source and mouth is 208 m. The highest stream gradient, 3.5 m/km is on the River Mustoja, which flows into the Gulf of Finland, while the lowest is on River Emajõgi with only 0.04 m/km.
The specificity of Estonian nature lies in the occurrence of the karstic feature (subsurface streams, swallow holes, etc) in Northern Estonia and the islands. Due to karst, some rivers flow partly underground.

While Estonia is a flat country, much of its area is forested or marshy. Approximately 25 percent of the land is considered arable. Permanent pastures comprise 11 percent of land use. The climate in Estonia is similar to Nordic climate, having a mixture of coastal and inland influences. Estonia's marine location keeps the climate moderate along the coast. Inland, temperatures are typically more extreme. Summers in Estonia are generally cool, with temperatures rarely exceeding 20°C. Winters are cold, with temperatures usually remaining below freezing from mid-December to late February. July and August are the wettest months. The annual average precipitation at the river basins is about 600–750 mm with the average potential evapotranspiration rate of 300-500 mm annually. Rain and melting snow cause some flooding of rivers in the spring.

Characterization and justification of the selected study areas

Three river watersheds have been selected to test and evaluate the performance of chosen simulation models to describe watershed hydrology and water-quality measures, and to assess if the combination of these models and available data is suitable to describe and represent similar metrics for entire Estonia. The three watersheds possess somewhat different characteristics.

All the watersheds in this project, Vihterpalu, Keila and Leivajõgi are located in northern part of Estonia. They are typical Estonian lowland rivers that drain into the Baltic Sea. Vihterpalu River's watershed has a total area of 480 km2, Keila River's water yield is collected from an area of 631.79 km2 and Leivajõgi's watershed is sized 84.85 km2. The river basins were considered upward of their respective gauge stations (Vihterpalu, Keila and Pajupea). All the basins are covered mostly by forests and agricultural landscapes.

Different types of models require and allow different types and levels of representation of distinct areas. For example, the point models only allow using a soil profile as basic unit, and it is up to the user to allocate what area that represents. Box-type catchment models, like INCA, allow the user to delineate a limited number of land-use areas, combined with a given soil type. GIS-based catchment models think in layers of information, and hence will internally delineate a finite, but large number of units based on an overlaid combination of a land-use map and a soil map – allowing various combinations of both types of information. This latter approach is the most detailed and advanced approach, and hence we use that as the example that we present.

Using the Corine Land Cover map and performing an analysis using GIS, information about the land use distribution in each basin and sub basin was extracted (Table 1). For example, we can observe that 72% of the Vihterpalu basin is covered by forest and semi natural areas and the major concentration of this land use is at subbasin Vihterpalu 2 with 36 % coverage of the total basin area. Keila River is covered at about the same proportion by forest and agricultural areas; only at subbasin Maidla the agricultural areas dominate (51% of the area). Out of the three studied basins, the Keila basin has the largest proportion of point-source type pollutants. Leivajõgi has a bit more heterogeneous land use distribution with 56% of forest areas and 42% of agricultural areas at the entire basin level. Forest areas are mostly located at subbasin Leivajõgi 3 (34% of the basin area), followed by subbasin Leivajõgi 1 with 12%. The other two subbasins are mostly covered by agricultural areas. In general, the artificial surfaces, wetlands and water bodies are not influential land covers in any of the three basins.

	Basin Vihterpalu		Subbasin Vihterpalu1			Subbasin Piirsalu 1			Subbasin Vihterpalu2			Subbasin Vihterpalu		
	Total Area km2	480.18		Total Area km2	72.48		Total Area km2	55.08		Total Area km2	247.14		Total Area km2	105.48
	Area km2	% Basin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area
Artificial surfaces	2.15	0	1.13	0	2	0.10	0	0	0.92	0	0	0.00	0	0
Agricultural areas	77.70	16	11.84	2	16	6.40	1	12	34.50	7	14	24.96	5	24
Forest and semi natural areas	344.46	72	59.36	12	82	46.64	10	85	170.96	36	69	67.51	14	64
Wetlands	54.41	11	0.15	0	0	1.94	0	4	40.76	8	16	12	2	11
Water bodies	1.45	0.3	0.00	0	0	0.00	0	0	0.00	0	0	1	0.3	1
Total	480.18	100	72.48	15	100	55.08	11	100.00	247.14	51	100	105.48	22	100

	Basin Keila		Subbasin Keila1			Subbasin Keila2			Subbasin Maidla			Subbasin Keila		
	Total Area km2	631.79		Total Area km2	127.20		Total Area km2	240.53		Total Area km2	51.40		Total Area km2	212.66
	Area km2	% Basin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area
Artificial surfaces	15.04	2	1.74	0	1	5.54	1	2	0.37	0	1	7.39	1	3
Agricultural areas	290.09	46	55.15	9	43	109.49	17	46	26.08	4	51	99.37	16	47
Forest and semi natural areas	292.11	46	57.32	9	45	113.48	18	47	22.61	4	44	98.71	16	46
Wetlands	34.54	5	12.99	2	10	12.01	2	5	2.34	0	5	7	1	3
Water bodies	0.00	0	0.00	0	0	0.00	0	0	0.00	0	0	0	0	0
Total	631.79	100	127.20	20	100	240.53	38	100.00	51.40	8	100	212.66	34	100

	Basin Leivajõgi		Subbasin Leivajõgi 1			Subbasin Leivajõgi 2			Subbasin Leivajõgi 3			Subbasin Pajupea		
	Total Area km2	84.85		Total Area km2	12.20		Total Area km2	18.45		Total Area km2	44.09		Total Area km2	10.11
	Area km2	% Basin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area	Area km2	% Basin Area	% Subbasin Area
Artificial surfaces	0.44	1	0.00	0	0	0.00	0	0	0.26	0	1	0.19	0	2
Agricultural areas	35.98	42	2.35	3	19	11.91	14	65	14.43	17	33	7.28	9	72
Forest and semi natural areas	47.54	56	9.84	12	81	6.53	8	35	28.52	34	65	2.64	3	26
Wetlands	0.89	1	0.00	0	0	0.00	0	0	0.89	1	2	0	0	0
Water bodies	0.00	0	0.00	0	0	0.00	0	0	0.00	0	0	0	0	0
Total	84.85	100	12.20	14	100	18.45	22	100.00	44.09	52	100	10.11	12	100

Table 1: Vihterpalu, Keila, and Leivajõgi landuse distribution at Basin and Subbasin levels.

Altogether, the three watersheds represent various types of landscapes, water and pollution sources and river types. We note, however, that these choices were partly driven by necessity, as these three are some of the limited number of watersheds that are (or have been) monitored in Estonia.

Data monitoring at the study areas

There are four meteorological stations relevant for this study that belong to the KAUR (Estonian Environment Agency) network, namely Pakri MJ, Lääne-Nigula MJ, Tallinn-Harku AJ, Kuusiku MJ and Kehra HJ. These stations also collect data on precipitation, but they are outside the examined basins. Each basin has a gauging discharge station that measures the discharge at the downstream end of each river.
The water quality measurement, including the determination of TN, TP, nitrate, ammonia and suspended sediment concentrations were based on grab water samples, collected at the gauging stations twice in a month.

The NORRA database contains all the monitored data used in this Report.

Models' description

Guidelines to selecting models (best practices, etc.)

There are numerous considerations to be accounted for, when simulation modeling based studies are designed, and the actual simulation models or model packages are chosen (Waveren et al., 1999; Farkas and Hagyo, 2010; Deelstra et al., 2010). The outcome of a particular simulation based study is heavily dependent on – besides the model itself - the quality, resolution and amount of the input data available and used, the quality and extent of the expert knowledge about locally prevailing conditions, as well as the validity of any assumptions that are inevitably made while parameterizing the model (Waveren et al., 1999; Deelstra et al., 2010). For this reason, we have found it important that a balance is found between e.g. model quality and relevance to the given area, the model's resolution both spatially and temporally vs. the resolution and availability of the base data, model simplicity and ease of use and the experts' familiarity with the given simulation model(s). The potential for linkages to/with pre-existing studies as well as the capability to address issues of stakeholders' interests are also examples for considerations that point beyond the idea of choosing the 'best model' in terms of strictly its scientific complexity and acceptance. This project served as an excellent basis for working with various models and evaluating the general experience with them in terms of the cost-benefit balance for the user.

Saloranta et al. (2003) established a set of operational and functional selection criteria for (computer) models whose application is intended to support decision making related to a particular water management issue. However, these criteria, the so-called "benchmark-criteria" can also guide potential model users in selecting the appropriate model for use in other areas as well. The benchmark criteria are presented in the form of 14 questions – with a 3-tier response system – through which each model can be evaluated.

Based on the benchmark criteria, a preliminary model evaluation has been performed to select simulation models for the NORRA project process-based simulation modelling tasks.

Models available for the team have been evaluated and the list of criteria that was deemed most important is as follows:

Q1.1. How well does the model's output relate to the management task?
Q1.2. How well does the model's spatio-temporal resolution match the requirements of the task?
Q1.3. How well has the model been tested?
Q1.4. How complicated is the model in relation to the task?
Q1.5. How is the balance between the model's input data and data availability?
Q1.8. How is the peer acceptance for the model with scientific theory?
Q3.5. How is the model's flexibility for adaptation and improvements

Our project, however, had a somewhat different set of goals than most classic modelling studies do, in terms of model selection and modelling work. It was our expressed approach that we use multiple simulation models to perform the same task, instead of picking one model. We have summarized reasons for this approach earlier. When this is the case, the above approach for model selection is useful information for documentation and discussion purposes, and we have followed a systematic evaluation of models. It was, however, decided that the working group will use several models that are feasible to use: (a) some team members are already familiar with it; (b) its capabilities match with the task; and (c) there is no obvious data gap that prohibits its use entirely. The agreed model selection yielded a list of models that carry different characteristics, and that are presented in the following chapters.

It is expected that any systematic change to climatic features will also have an influence on the longer term water balance of a given watershed. The choice of a simulation model to simulate surface and subsurface water balance and water availability in the study areas was somewhat more complex, given the overlapping expertise among (but also within) the institutions of the project partners.

Within the frames of the NORRA project, we further developed the model selection procedure by introducing Excel sheets containing not just the "benchmark" questions, but also scoring and explanation for model's evaluation. We believe, that this gives an opportunity to get more coherent evaluations from the different experts.

Tables 3a-3c show some selected parts of the question sheets for model evaluation. The scores are automatically summed up for each of the evaluated model in a separate sheet (Table 2).

		MODEL NAME					
		SWAT	HYPES	INCA	Mod_4	Mod_5	Mod_6
Total score		23	20	23	21		
Number of "0" scores		0	0	0	0		
	RELEVANCE	12	11	15	11		
Total score	SensAnal	2	2	1	1		
	Ease of Use	9	7	7	9		
	RELEVANCE	0	0	0	0		
Number of "0" scores	SensAnal	0	0	0	0		
	Ease of Use	0	0	0	0		
Expertise of the expert, completing the table in various models		4	3	5	4		
(please, give a score from 1 to 5)							
5	experienced in using the model						
4	has applied the model for 1 or 2 sites						
3	has limited experience with the model						
2	has seen papers/presentations about the model application						
1	heard about the model for the first time						

Table 2: Final outcome of the model's evaluation procedure, following the criteria of Saloranta et al. (2003)

	Please, give your scores (0, 1 or 2) within the red frame only. We appriciate any additional comments in columns M-P.								
		SWAT	HYPES	INCA	Soiln	Good (2)	Adequate (1)	Inadequate (0)	Items for taking decision
1. MODEL APPLICABILITY AND RELEVANCE FOR THE MANAGEMENT TASK									
1.1	How well does the model's output relate to the management task?	2	2	2	1	at least 2 items			the model's output can be directly related to the "core" of the management task
									the model's output (relevant to the management task) consists of variables that are commonly applied and easy to measure and
									the model allows the simulation of a variety of relevant management operations
							if valid		the model's output can be relate to the management task via clear, well-known, and well-established links
								at least 1 item	the model's output is peripheral in relations to the management task
									the links between the model's output and management task are not clear or
1.2	How well does the model's span and resolution in time and space compare with the requirements of the management task?	2	2	1	1	all items			the model can be run with any desired spatial and temporal resolution
									the model can be run over the desired spatial and temporal span (e.g. It allows simulations to be run over many years)
							if valid		there are restrictions on the model's spatial or temporal resolution or span, but the model is still expected to produce useful and meaningful results for the management task.
								if valid	the model's spatial and temporal resolution or span cannot be chosen to be appropriate for the management task
1.3	How well has the model been tested?	2	1	2	2	at least 3 items			there are at least 10 documented previous model applications
									at least five model applications are published in peer-reviewed journals
									the model has been evaluated against independent data sets
									the model has been evaluated in various conditions or geographical regions
									some previous model use and evaluation is closely related to the management task in question
							at least 2 items		there are at least three reported model applications
									the model has been evaluated in different conditions or geographical regions
									the model is specific to the site of the management task
								at least 1 item	the model is site-specific to other type of site than that of the management task and it has not been evaluated in different conditions or geographical regions
									there are less than three documented previous model applications
1.4	How complicated is the model in relation to the management task?	1	1	2	1	if valid			the model has an optimally simple structure, i.e., it includes mostly only those processes and parameters that are known to be relevant for the management task.
							on of the items		the model has a somewhat too complicated structure, i.e., most of the model's processes and parameters are relevant, but the model seemingly includes also some irrelevant processes and parameters
									its alternatively, the model is somewhat too simple, i.e. relevance to the management task could be enhanced somewhat (but not radically) by introducing some additional processes
								on of the items	the model is too complex, and most of the model's features could clearly be omitted or simplified (or a more simple model could be chosen) without loss in model relevance for the management task
									alternatively, model is too simple, and many key processes relevant to the management task are not included
1.5	How is the balance between the model's input data requirements and data availability?	1	1	2	1	if valid			the required model input data are available from monitoring and field observations, either from the management site or from other applicable site close to it
							if valid		most of the required model input data are available from monitoring and field observations, either from the management site or from other applicable site close to it; however, some surrogate input data (e.g. results from other models, or data from other sites) must be used
								if valid	majority of the required model input data are not available from monitoring and field observations from the management site (or from other applicable site close to it)
1.6	How is the identifiability of the model parameters?	1	1	2	1	at least 1 item			all relevant model parameter values are well documented in scientific literature or can be estimated directly based on available
									available data (corresponding to model output variables) will allow establishment of all relevant model parameter values via model calibration
							if valid		there seems to be enough data or documentation available to allow an adequate estimate of most of the relevant model parameter values (either directly or via model calibration)
								if valid	there are clearly not enough calibration data or other parameter documentation available to allow for an adequate establishment of many of the relevant model parameter values
1.7	How easily are the model results understood and interpreted?	1	2	2	2	if valid			nonspecialist users are generally capable of understanding and interpreting the model output results
							if valid		assistance from research staff or modelling specialists is necessary to clarify ant interpret the model's output results
								at least 1 item	expert skills, long experience, and deep insight (e.g., those of a model developer) are needed to understand and interpret the model results
									much "tacit" (i.e., difficult-to-express) knowledge or intuition is involved in the interpretation of the model results.
1.8	How is the peer acceptance for the model and the model's consistency with scientific theory?					at least 2 items			the model has gained wide and international acceptance among the scientific community
									the model is widely used in many countries
									the whole model is based on well-established scientific theory
							at least 1 item		the model is used and gained peer-acceptance mostly locally/nationally
									most of the model components are based on well-established science

ReadMe | Relevance | SensAnal | EaseOfUse | SCORE

Table 3a: Model evaluation using the benchmark criteria of Saloranta et al. (2003) – Relevance

	Please, give your scores (0, 1 or 2) within the red frame only. We appriciate any additional comments in columns M-P.								
		SWAT	HYPES	INCA	Mod_4	Good (2)	Adequate (1)	Inadequate (0)	Items for taking decision
2. HOW WELL IS THE MODEL SUITED FOR SENSITIVITY AND UNCERTAINTY ANALYSES AND HOW WELL HAVE THESE ANALYSES BEEN PERFORMED AND DOCUMENTED?									
2	How well is the model suited for sensitivity and uncertainty analyses and how well have these analyses been performed and documented?	2	2	1	1	at least 3 items			Thorough analysis of model sensitivity has been performed and reported
									model sensitivity analyses is published in peer-reviewed journal(s)
									a variety of sensitivity/uncertainty analysis techniques or software can easily and with reasonable effort be applied to the model
									the model software contains tools for sensitivity/uncertainty analysis
									uncertainty ranges, associated with the model parameter values, can be adequately established
							at least 1 item		screening of the most sensitive model parameters has been done and published in technical report(s)
									sensitivity/uncertainty analysis techniques or software can be applied to the model, but this will be a rather laborious task
								if valid	no model sensitivity analyses has been performed because the model is generally not suitable for adequate analysis of sensitivity/uncertainty
	Total score	2	2	1	1				
	Number of "0" scores	0	0	0	0				

Table 3b: Model evaluation using the benchmark criteria of Saloranta et al. (2003) – Sensitivity

Please, give your scores (0, 1 or 2) within the red frame only.
We appriciate any additional comments in columns M-P.

		SWAT	HYPES	INCA	Mod_4	Good (2)	Adequate (1)	Inadequate (0)	Items for taking decision
3. MODEL TRANSPARENCY, EASE OF UNDERSTANDING AND EASE OF USE									
3.1	How is the model's version control?	2	2	2	2	all items			different model versions are numbered and description of version development exists
									it is easy to check the version of the executable model
									user manual and other model documentation matches with the particular model version
							one of the items		model versions are numbered
									user manual and other model documentation is known to be sufficiently consistent with the particular model versionű
									alternatively, only one model version exists
								if valid	no consistent numbering between different model versions exists
3.2	How are the model's user manual and tutorial?	2	1	1	2	at least 2 items			instructions for use are comprehensive and detailed, yet operative and clear
									the scope of the model, its application domain, input file structure, and parameter estimation methods are explained
									there are application examples, or a well-structured tutorial section
							if valid		user manual is less comprehensive, but includes clear operating instructions
								if valid	adequate user manual is not available
3.3	How are the model's technical documentation?	2	1	1	2	at least 2 items			model documentation gives comprehensive and detailed description of the processes, algorithms, and numerical methods
									the science behind the model is reviewed in the documentation
									documentation is published in peer-reviewed scientific journal(s)
							if valid		technical document of the model processes and equations are available
								if valid	no adequate technical document of the model and its structure is available
3.4	How are the model's interactiveness, user-friendliness, and suitability for end-user participation?	1	1	2	1	at least 3 items			the model is well-structured, transparent, and has informative user interface with easy visualisation of the model output
									input data format is user-friendly and model parameters are easily modified (or the model is connected to parameter databases)
									active user support is available, wither from model developers or from a user-group
									non-specialist users are generally capable of running the model
									the model can contribute to the process of negotiation among relevant stakeholders
							at least 1 item		the model is less transparent and the facilitation of a model specialist is required to guide the model user
									the model has well-functioning user interface offering the user some insight and control on model parameters and functioning
								at least 1 item	the model is an "opaque box", and allows the user no interaction with the model and its parameters
									only a specialist (e.g. model developer) can use the model
3.5	How is the model's flexibility for adaptations and improvements?	2	2	1	2	at least 2 items			the model's source code is available to the model user and is well structured and documented
									the model is flexible, i.e., different processes can easily be added to (or removed from) the model in the form of e.g. , add-in modules
									the model is easily adaptable for inclusion in integrated model systems
							one of the items		the model's source code is available to the model user
									alternatively, the model's source code is not generally available, but model developers may give support for adaptation and improvements
								if valid	the model's source code is not available and no active model development exists
	Total score	9	7	7	9				
	Number of "0" scores	0	0	0	0				

Table 3c: Model evaluation using the benchmark criteria of Saloranta et al. (2003) – "Easiness-of-use"

Table 4. shows an example of a somewhat similar assessment of numerous hydrological models, but from a different angle. While their assessment was not based on the questions by Saloranta et al. (2003), but rather focusing on the models' capabilities, there is still a significant overlap between the two lists of evaluation criteria.

From both evaluations, the SWAT and INCA models emerged as the most potent models to use for the hydrological simulations at the watershed scale. From the scientific point of view, both the models present widely used and well accepted, conceptually well-established models that have the ability to be flexibly used in various environments and levels of data availability. From the practical point of view, the partners have voted for using SWAT and INCA because of i) their availability, ii) having experience at both institutes in applying these models under various geographical and climatic conditions; iii) the ease at which support is available from the developers and – in case of SWAT - from the huge SWAT user group World-wide if needed; and iv) the already proven ability of the models to describe hydrological processes under conditions similar to those in focus.
The HYPE model got lower scores because it was less known within the Consortium due to its novelty.

Model layer	Processes	DrainMod	Coup	HBV	INCA	SWAT
Above ground vegetation zone	Precipitation	Driving	Driving	Driving	Driving	Driving
	Snow dynamics/snowmelt	Calculated	Calculated	Calculated	Calculated	Calculated
	Interception	Indirectly	Calculated	Calculated	Indirectly	Calculated
	Transpiration	Indirectly	Calculated	Calculated	Indirectly	Calculated
Soil surface	Evaporation	Indirectly	Calculated	Calculated	Indirectly	Calculated
	Surface runoff	Calculated	Calculated	Calculated	Calculated	Indirectly
	Infiltration	Calculated	Calculated	Indirectly	Indirectly	Indirectly
Unsaturated zone	Bypass/ macropore flow	NO	Calculated	Indirectly	NO	Calculated
	Plant water uptake	Indirectly	Calculated	Indirectly	Indirectly	Calculated
	Soil water redistribution	NO	Calculated	Calculated	NO	Uniform
	Capillary rise	Calculated	Calculated	NO	NO	NO
	Water flow in frozen soil	Indirectly	Calculated	Calculated	NO	at saturation
	Lateral flow to stream	NO	NO	Calculated	Calculated	Calculated
	Subsurface drainage flow	Indirectly	Calculated	NO	Indirectly	Indirectly
	Percolation to sat. zone	Calculated	Calculated	Calculated	Calculated	Calculated
Saturated zone	Lateral inflow	Parameter	Parameter	NO	NO	NO
	Capillary rise to unsat. zone	NO	Calculated	Calculated	NO	Indirectly
	Recharge to deep aquifer	NO	NO	NO	NO	Calculated
	Base flow	Calculated	NO	Calculated	Calculated	Calculated
CONFINING LAYER						
DEEP AQUIFER						

Table 4: Model evaluation using comparative matrix's

Selected models in brief

As we introduced earlier, we have opted to use a set of diverse criteria to choose which models we intended to work with; ranging from the model being internationally accepted and documented, and capable to do the desired task, via their complexity and data requirement to project partners' familiarity with each of the models. The models we selected for this study, followed by the corresponding expert(s) within Project are:

- HBVlight and PERSiST – hydrological models that provide hydrology input for the INCA model family – *study conducted by Csilla Farkas, Rain Elken*
- INCA – consisting of INCA-N and INCA-P water quality models - *Csilla Farkas, Rain Elken*
- SWAT – hydrology and water quality model - *Juan Manuel Garcia Diaz, Andreas Porman and Tiia Pedusaar for hydrology; Rain Elken and Csilla Farkas for water quality*
- SOIL and SOIL-N - hydrology and water quality model - *Anatoli Vassiljev*
- HYPE - *Rain Elken*

Description of the selected models (with tabular information for transparency)

PERSiST/HBVlight & INCA

PERSiST (Precipitation, Evapotranspiration and Runoff Simulator for Solute Transport) is a simple daily-based rainfall-runoff hydrology model developed in cooperation with the Swedish University of Agricultural Sciences and the University of Reading (UK). At its core, PERSiST is a conceptual, semi-distributed, so called bucket-type model which simulates water fluxes from precipitation through the terrestrial part of a catchment into rivers and streams. The model requires daily input of air temperature and precipitation from one or more sites as driving data. For calibration PERSiST requires measured stream flow at one or more stations in a river.

The main reason for the usage of PERSiST within the Project is its capability to generate hydrologic input to water quality INCA models. INCA models rely on external time series of hydrologically effective rainfall (HER - the fraction of precipitation that directly contributes to runoff) and soil moisture deficits (SMD - the difference between the current depth of water and the water holding capacity) which can be directly produced by PERSiST.

For the user the model in one single executable file (*Figure 2* with initial screen) where catchment parameters and driving meteorological data need to be input as separate text files in specific formats.

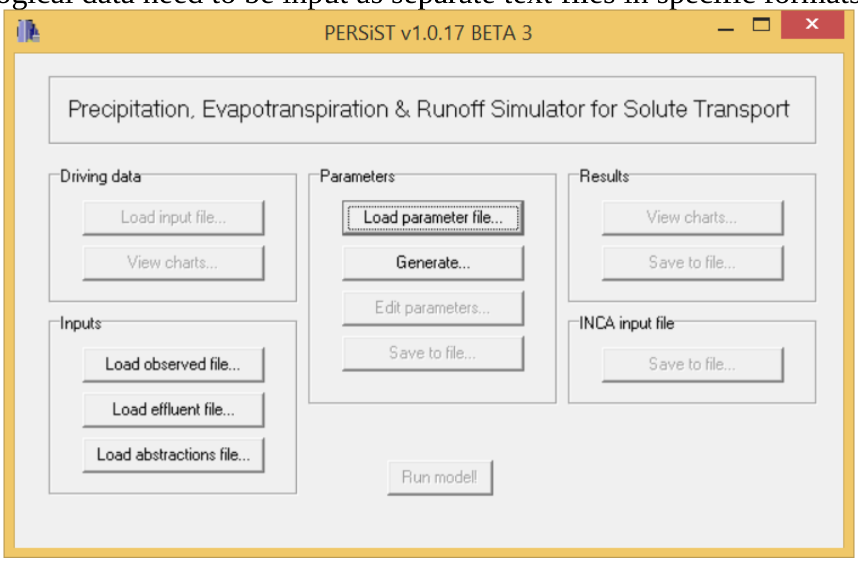


Figure 2. PERSiST model initial screen

HBVlight is a semi-distributed, conceptual hydrological model that describes the essential characteristics of the precipitation-runoff process. It is a simplified version of the HBV eco-hydrological model developed by the Swedish Meteorological and Hydrological Institute **(**SMHI). HBVlight simulates the volumes of water stored as snow, subsurface water and streamflow. The model does calculations for 10 elevation bands within a catchment in order to take into account the elevation variation of the driving precipitation and temperature data. Each elevation band may be divided into a maximum of four computational elements - three land use zones with different vegetation and soil types and one lake area.

Similar to PERSiST, HBVlight has a simple user-interface (*Figure 3*) with parameters and driving data loaded as text files.

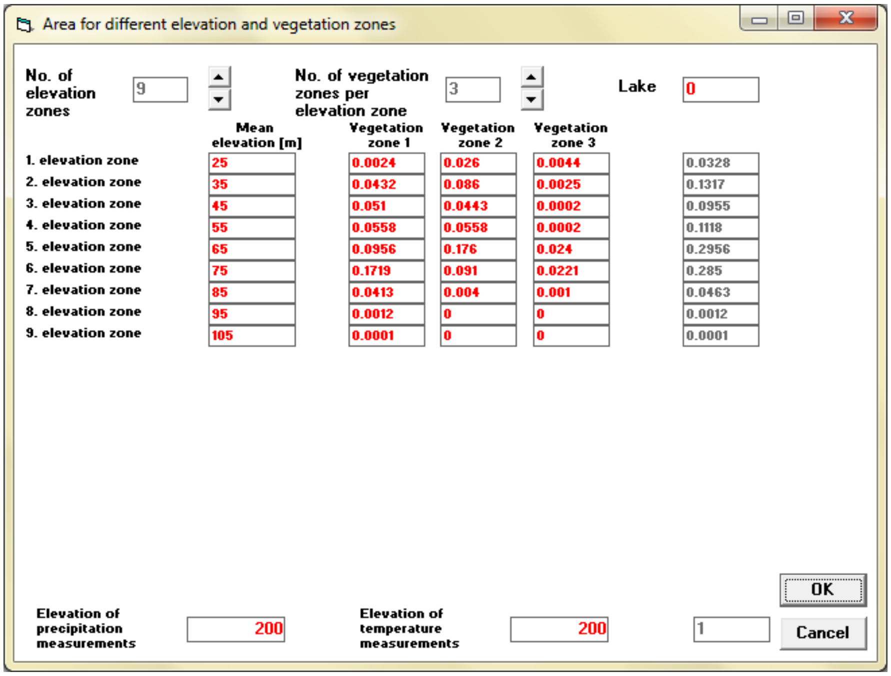


Figure 3. HBVlight model screen for parameterisation

The **INCA** (Integrated Nitrogen Model for European Catchments) is a semi-distributed stochastic catchment-scale model to assess the impact of point and diffuse pollution sources on in-stream chemistry in an integrated form. It is developed by the University of Reading (UK) and provides process-based representation of the factors and processes controlling nutrient dynamics in both the land and the in-stream proportion of catchments. One of the key drivers while designing the model was to minimize input data requirements. For simulating nitrogen and phosphorus processes the model has two separate executable files – INCA-N and INCA-P for nitrogen and phosphorus respectively. In the INCA model, hydrologically effective rainfall (HER – modelled with PERSiST or HBVlight) is input to the soil water storage module, driving water flow through the catchment. Hydrology within a catchment is modelled using a simple two-box approach, with key “tanks” of water in the reactive soil zone and deeper groundwater zone. Flows from the soil and groundwater zones are controlled by residence times in the “tanks”. To account for spatial distribution, the user has to define up to six landuse/soil classes, which should remain uniform for the whole catchment, but can be differentiated within sub-catchments. The modeling structure is based on stream definition of a discrete set of linked reaches where, for the purpose of modelling chemical processes, each reach is assumed to be fully mixed. Each reach may have input from diffuse sources (from land) or point sources such as wastewater effluent discharges.
Figure 4 illustrates the classification of landuse/soil classes and reach structure on the example of the Keila catchment.

INCA Code	INCA_name	Area %	LAND USE	SOIL
AGR_L	Agric_Loam	**30.76**	Agricultural areas Agricultural areas Urban	loam peat all
AGR_SL	Agric_Sandy Loam	**18.55**	Agricultural areas	loam, sandy loam sand
For_L	Forest_Loam	**20.20**	Forest and semi natural areas	loam
For_P	Forest_Peat	**10.18**	Forest and semi natural areas	peat
For_SL	Forest_Sandy Loam	**14.98**	Forest and semi natural areas	loam, sandy loam sand
Wet	Wetland	**5.34**	Wetland	all (peat)

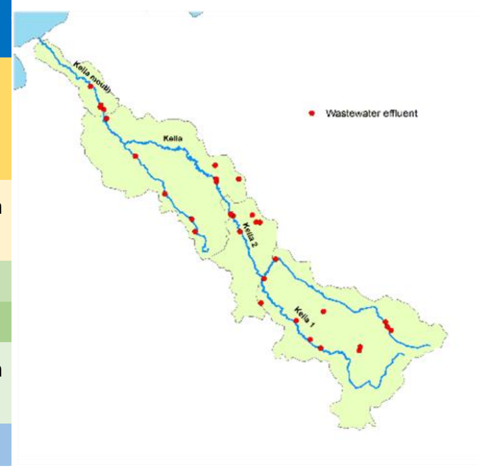


Figure 4. Land use combined with soils in the INCA model and reach structure (example for Keila)

SWAT

The **S**oil and **W**ater **A**ssessment **T**ool (***SWAT***) (Neitsch et al. 2009) is a continuous time, semi-distributed watershed-scale model that operates on a daily time step. SWAT is physically based and developed to quantify the impact of land management practices in large, complex watersheds. SWAT requires information on weather, soil properties, topography, vegetation, and land management practices in the watershed. The physical processes associated with water movement, sediment movement, crop growth, nutrient cycling, etc. are directly modeled by SWAT using these input data. For modeling purposes, a watershed may be partitioned into a number of sub-watersheds or sub-basins, which are spatially connected. Input information for each sub-basin is grouped into hydrologic response units or HRUs (*Figure 5*). HRUs are lumped land areas comprised of unique land cover, soil, slope, and management combinations. Runoff is predicted separately for each HRU and routed to obtain the total runoff for the watershed. SWAT calculates canopy storage (water intercepted by vegetative surfaces), infiltration, redistribution (movement of water through a soil profile after input of water), evapotranspiration (ET and PET), lateral subsurface flow, base flow and surface runoff. Surface runoff is computed using a modification of the SCS curve number method. The curve number method varies non-linearly with the moisture content of the soil. The curve number drops as the soil approaches the wilting point and increases to near 100% as the soil approaches saturation. The model increases runoff for frozen soils but still allows significant infiltration when the frozen soils are dry.

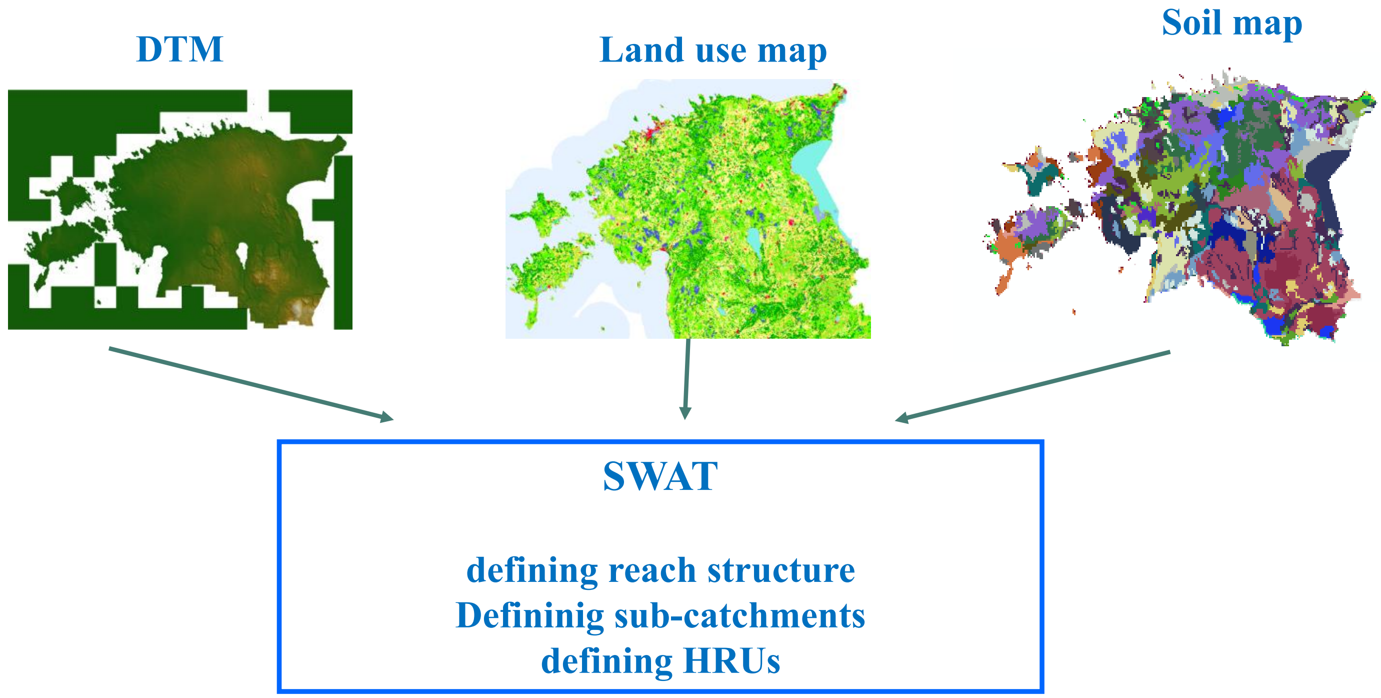


Figure 5. Identification of sub-basins and HRU's by the SWAT model

SOIL and SOILN

The models SOIL and MACRO were developed to provide data needed for nitrogen-leaching models but these models can only be used to model small fields. These models are one-dimensional, developed for use in small homogeneous areas at the field or plot scale. In this study, both models (SOIL and MACRO) were used in succession with different versions of the SOILN model such that the results from SOIL and MACRO (e.g. soil moisture content, water flows between layers, and soil temperature) are used as input to the SOILN model.

The SOIL model simulates water and heat processes in soil taking into account the plant cover. The basic structure in the model is the depth profile of the soil. Two coupled differential equations for the water and heat flow represent the central part of the model. These equations are solved with an explicit numerical method (Jansson 1991). Meteorological data, most importantly, precipitation, air temperature, air humidity, wind speed, and cloudiness, are the driving variables to the model.

The MACRO model considers the division of the soil profile into the micro- and macro pore. Soil macro pores (e.g. root and worm holes, structural shrinkage cracks) allow rapid non-equilibrium fluxes of water in soil (Beven and Germann, 1982), and consequently influence the leaching of nitrogen. Larsson & Jarvis (1999) showed that such influence might be quite significant. Hydrological models developed for watersheds usually ignore the non-equilibrium water movement. In this study, in addition to the SOIL, the MACRO model was used because some authors (Litaor et al., 2008) have indicated the presence of macro pores in peat soils.

The SOILN model simulates major C and N-flows in soils and plants. The model has a daily time step and simulates flow and state variables at a field level. Input variables are daily data on the air temperature and solar radiation and data on soil heat and water conditions simulated by the SOIL or MACRO model. The soil is divided into layers. In each layer, mineral N is represented by one pool for ammonium N and one for nitrate N. Ammonium N is usually regarded as immobile whereas nitrate form is transported with the water fluxes (a special option can also make ammonium mobile). The ammonium pool is increased by the nitrogen supplied from manure application, mineralization of organic material and by atmospheric deposition, and it is decreased by immobilization to an organic material, nitrification to the nitrate pool and plant uptake. The nitrate store increases through the nitrification of the ammonium pool, fertilization and atmospheric deposition. The leaching, denitrification and plant uptake reduce the amount of nitrate N in the soils. Water flows that transport nitrate N between the layers are responsible for nitrogen leaching. The rate of the decomposition of organic matter depends on soil moisture and temperature conditions. Nitrogen dynamics of the organic matter is governed by C flows and mineralization or immobilization depend on the C/N ratio of the decomposed material and availability of mineral N (Johnsson et al., 1987). The models were adapted according to the scheme described in (Vassiljev et al., 2004). The scheme includes calculations for the different soil profiles and simulation of water movement in the river system.

HYPE

Within the project, the HYPE model was selected as the model to be incorporated into the integrated web-based Airviro modelling system of EERC (the Estonian Environment Research Centre). This hydrological catchment model has been developed by the Swedish Meteorological and Hydrological Institute (SMHI). The model simulates the flow of water and the transport of substances through the soil, river and lakes to the river outlet (Arheimer et al., 2008; Lindström et al., 2009). The HYPE model is a semi-distributed hydrological model for water and water quality. It simulates water and nutrient concentrations in the landscape over time, most often HYPE is run at daily time steps. Its spatial division is related to sub-catchments (in the model called sub-basins) and classes (non-located fractions of the sub-basin area separated by land use/vegetation, soil type, or elevation). *Figure 6* shows an example of a catchment divided into two sub-basins, and 4 SLC (soil-landuse)-classes represented with different colour. SLC-s compare to hydrological response units (HRUs) in other models, with its parameters controlling their operation. Within a sub-basin, HYPE simulates different hydrological compartments; snow pack, soil (three layers) including shallow groundwater, rivers and lakes. In addition, it simulates the coupling between sub-basins through routing of river flow. It's also possible to take into account anthropogenic processes like reservoirs, flow regulation, irrigation, and abstractions. The schematic overview of processes involved in HYPE is shown in *Figure 7*.

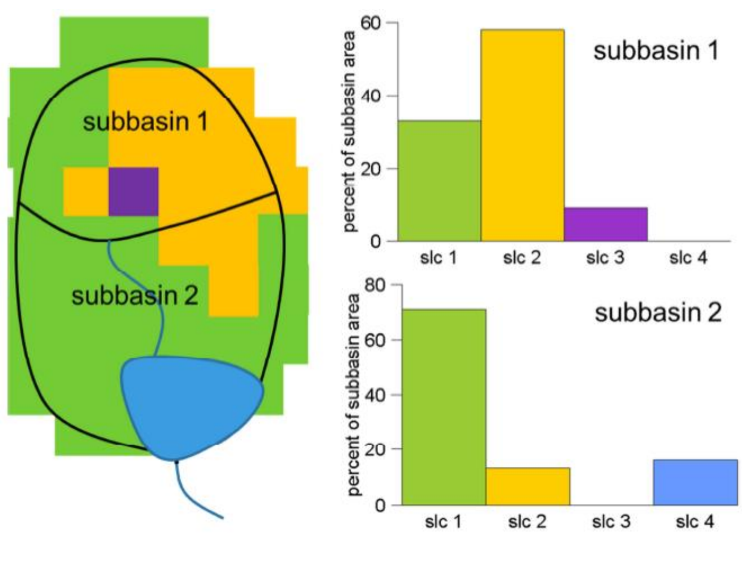


Figure 6. Example of the spatial division of a catchment into sub-basins and soil-landuse classes (SLC).

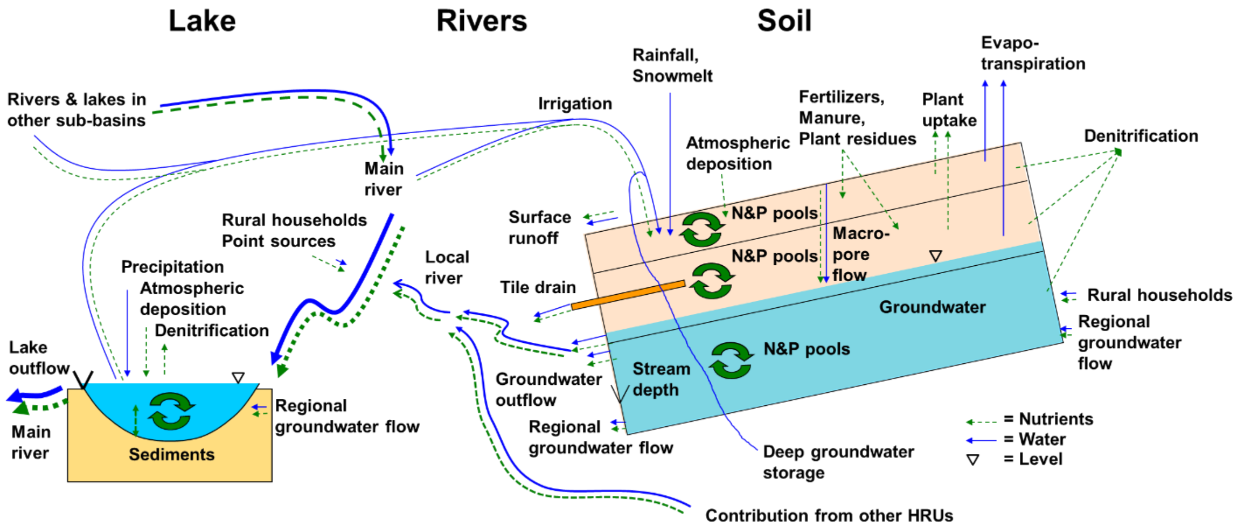


Figure 7. Schematic description of processes in the HYPE model.

HYPE is an open source model. By default it doesn't provide any graphical user interface; the model operates with input-output text files where the user has to set the model up in a time consuming process. Airviro is a system primarily designed for air quality modelling and management. It has also been developed by SMHI, and the Apertum IT AB internet technology company and has been continuously developed since the early nineties. EERC is one of some 80 places world-wide where this system is being installed and used. The Airviro system provides a user-friendly interface and is capable of handling large amounts of data and lots of users who can run different models simultaneously, and assists the user with a prepared model set-up platform. The main target group of users are environmental specialists from different organizations and authorities who deal with water related issues. The Airviro system provides the user a lot of options to visually represent modeling results and they can be easily shared between users.

The general plan in the project was to set up the model for whole Estonia, where end-users can then rely on the system for their specific tasks, after manipulating some task-specific parameters. While Apertum IT AB did the integration of HYPE into Airviro, Keskkonnaagentuur (the Estonian Environment Agency) carried out the model set-up and preparation. As part of this step, basic calibration and validation was done on the three pilot catchments in this project, while a more in-depth calibration is ongoing work.

Model setup and parameterizing the selected models

Prior to setting up a model, it is necessary to collect all the information related to the physical basin's representation, weather data and hydrological data. This step is usually time consuming since the available datasets are rarely in the condition to be readily usable, they often need a range of manipulations, such as gap filling, re-formatting, re-classifying or estimating these input data. *Figure 8* shows the example of the procedure to setup the SWAT model. The physical inputs described later, land use map, soil map and slope bands from the DEM were overlaid, to define the Hydrological Response Units HRUs, which is a unique combination of land use, soil and slope for each sub-basin that is thereby defined. Subdividing the basin into areas having unique land use and soil combinations enables the model to reflect differences in evapotranspiration and other hydrologic conditions for various coverages and soils.

In this project, ArcSWAT, an ArcGIS extension and graphical user interface for the SWAT model, has been used to process all the information that goes into the model. Model inputs are discussed in a separate section of this report. Once the inputs have been entered, the models' embedded processes need to be parameterized prior to first run. It is, however, important that simulation models undergo a manual or automated calibration process of parameters in order to optimize how observations are reproduced and validation on independent data to quantify the model performance.

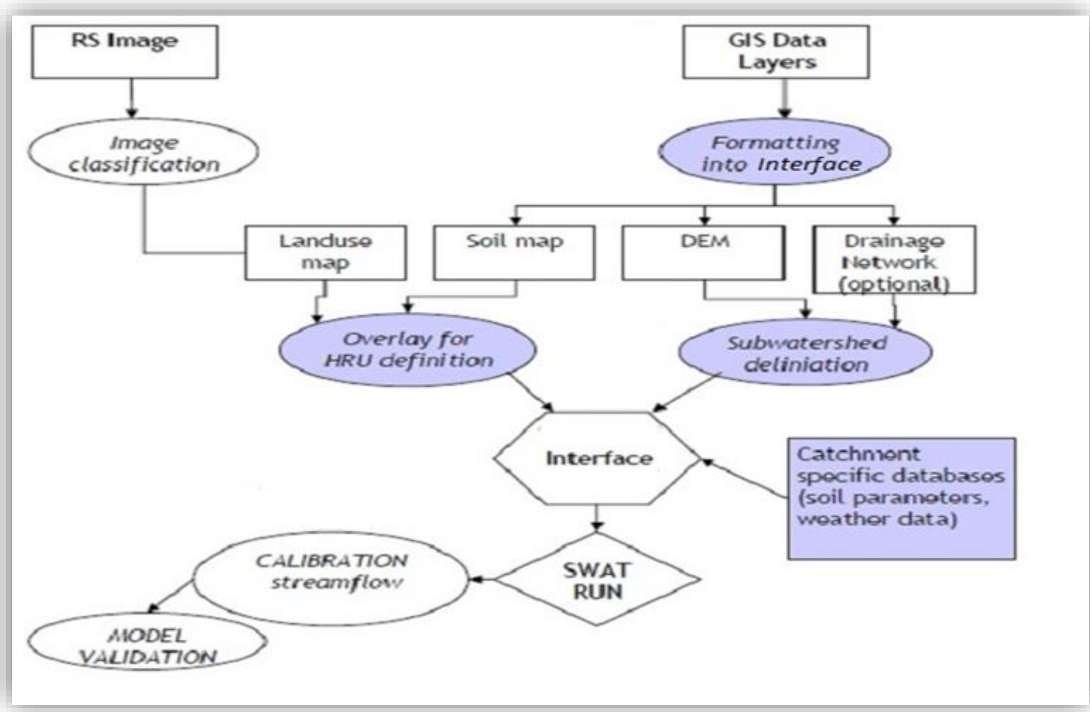
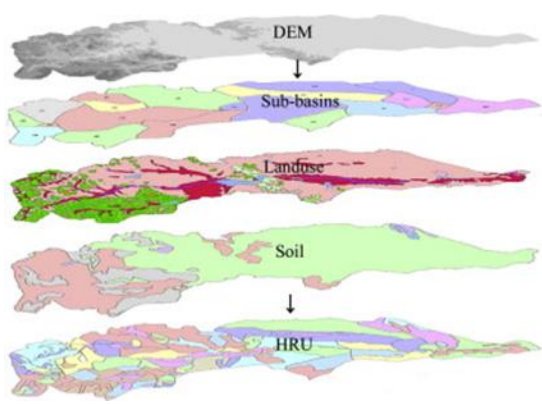


Figure 8: Schematic representation of setting up the SWAT model



Depending on data availability and modelling accuracy, one sub-basin may have one or several HRUs defined. In this study, dominant HRUs were used to match the complexity of the model with other models involved in this project like Inca and HBV-light. This option chooses the land use, soil and slope combination of the largest potential HRU in the sub-basin.

	Vihterpalu	Keila	Leivajõgi
Number of HRUs	19	11	11
Number of Subbasins	19	11	11
Number of meteo station used during calibration and validation	2	2	1
Number of radar grid points (discussed later)**	121	157	24
Number of virtual meteo stations, areal precipitation from (interpolation + radar) (discussed later)**	19	11	11

Table 5: Summary of sub-basins, HRUs and meteorological observation locations for the 3 catchments.

The Working Group made a very strong effort to be able to standardize the study cases among the various models - to the extent it was possible – in order to limit duplication of work, and reduce the number of factors that differ between model runs. Latter is important because we would like to be able to conclude about results between models, and the more the methodology or data differs, the less we are able to cite that the models being the cause of any differences.

Our efforts included e.g.:

1. Using common data sources and estimation methods (*Figure 9*)
2. Using common initial and boundary conditions (*Tables 6 and 7*)
3. Using harmonising model parameters, where possible (*Table 8*)
4. Using common calibration and validation periods
5. Using common model evaluation techniques

The common data source was the NORRA data platform (*Figure 9*), available for all the modellers and containing information on both, model input data (meteorological data, soil information etc.) and reference data for model's calibration (time series of measured discharge and water quality data).

Examples for setting up common initial and boundary conditions, and harmonising model parameters are given in *Tables 6-8*.

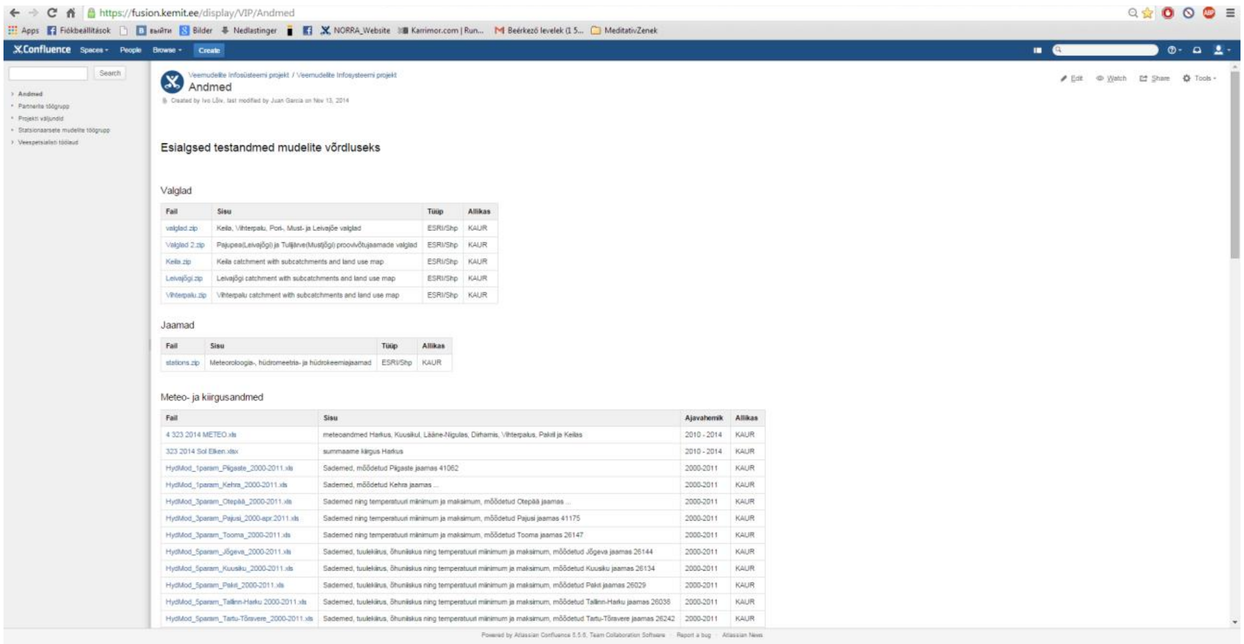


Figure 9: The common data platform in the NORRA project

	A	DrainMod	INCA	SOIL available	SOIL used	SWAT available	SWAT used	Synchronised value for the bulk run	SOIL parameter name	
4	**soil heat initial conditions**	+	-	+	+	-	-			**F - forest**
5	uniform temperature profile	+	-	+	+	-	-		SoilInitTempConst (AP)	**A- agricultural land**
6	distribution of soil temperature in the soil profile temp(z)-measured	+	-	+	+	-	-	**According to Table 1.**		
7	distribution of soil temperature in the soil profile temp(z)-estimated	-	-	+	-	-	-			
8	heat(z)	-	-	+	-	-	-			
9	**soil water initial conditions**	+	+	+	+	+	+			
10	distribution of water content in the soil profile water content(z)-measured	+	-	+	-	-	-			
11	uniform soil water content	+	-	+	+	-	-	**= 36.1 (RK8); = 15.1 (JE3)**	InitialWaterContent (AP)	
12	uniform vertical flow	-	-	+	-	-	-			
13	soil water initial flow (lateral)	-	+	-	-	-	-			
14	distribution of water potential in the soil profile pressure head(z) - measured	-	-	+	-	-	-			
15	uniform pressure head	-	-	+	-	-	-		InitialPressureHead (AP)	
16	Initial water storage as a fraction of field capacity	-	-	-	-	+	+	**= 1 (field capacity)**		
17	**Snow pack initial conditions**	+	+	+	-	+	-			
18	initial snow water content	-	-	-	-	+	-	**= 0**		
19	snow depth	+	-	+	+	-	-	**= 0**	SnowDepthInitial (AP)	
20	snow depth as water equivalent	-	+	-	-	-	-	**= 0**		
21	snow mass	-	-	+	-	-	-		SnowMassInitial (AP)	
22	**groundwater initial conditions**	+	+	+	+	+	-			
23	groundwater initial flow	-	+	-	-	-	-	**= 0.001 m^3/s**		
24	groundwater level	+	-	+	+	+	+	**= -0.5 m**	InitialGroundWater (AP)	
25	Initial depth of water in shallow aquifer	-	-	-	-	+	+			
26	Initial depth of water in deep aquifer	-	-	-	-	+	+			
27	**land phase initial conditions**	-	+	-	-	+	+			
28	inital curve number for soil moisture conditions	-	-	-	-	+	+			
29	curve number for management	-	-	-	-	+	-			
30	direct runoff initial flow	-	+	-	-	-	-	**= 1.1E-6**		
31	**instream initial conditions**	-	+	-	-	-	-			
32	instream flow	-	+	-	-	-	-	**= 0.0073 m^3/s**		
33	Crop	+	-	+	+	+	+		All the crop parameters will be set accorging to the crop fiel, that Sigrun has.	
34	intial leaf area index	-	-	+	+	+	-			
35	intial dry weight biomass	-	-		-	+	-			
36	initial root depth	+	-		-	-	-	**= 0.1 cm (A); = 200 cm (F)**		
37	initial temperature sum to maturity	-	-		-	+	-			
38	forest age	-	-	-	-	+	+			

Table 6: Example for setting up common initial conditions for different models

	Drainmod	INCA	Coup		SWAT	
			available	used	available	used
UPPER BOUNDARY CONDITIONS						
weather data						
daily average temperature	+	+	+	+	+	+
daily precipitation	+	+	+	+	+	+
global radiation	-	-	-	+	+	+
net radiation	-	-	-	+	-	-
air humidity	-	-	-	+	+	+
wind speed	-	-	-	+	+	+
hydraulically effective rainfall (HER)	-	+	-	-	-	-
potential evapotranspiration (PET)	+	-	+	+	+	+
measured	+	-	+	+	+	-
Penman method (modelled)	-	-	+	+	+	+
Penman method (direct input)	+	-	-	-	+	+
Thornthwaite method (modelled)	+	-	-	-	-	-
LOWER BOUNDARY CONDITIONS						
soil heat flow lower boundary:	+	-	+	+	-	-
base temperature as the lower boundary	+ (3.5 C°)	-	-	-	-	-
constant heat flow in time	-	-	+	+	-	-
soil temperature cycle in time	-	-	+	-	-	-
soil water lower boundary:	+	inbuilt	+	+	+	+
no flow (impermeable layer)	+	+	+	+	+	+
reacharge to deep aquifer	-	-	-	-	+	-
ground water level in time		-	+	-		-
pressure head in time	+	-	+	-	-	-
free drainage	-	-	+	-		-
seepage flow	+	-	+	-		-
unit grad flow	-	-	+	-	-	-

Table 7: Example for setting up common boundary conditions for different models

	A	B	C	D	E	F	G	H	I	J
1		PERSIST	INCA	SOIL	SWAT	PERSIST	INCA	SOIL	SWAT	
2						Values used in the BULK run				
3	**Soil parameters**									
4	clay content	-	-	+	+	Table 3				A - agricultural land
5	sand	-	-	+	+					F - forest
6	silt	-	-	(-)	+					
7	rock fragment content	-	-	(-)	+					
8	soil moist albedo	-	-	-	+					
9	organic matter	-	-	+	+					
10	bulk density	-	-	-	+					
11	**Soil hydraulic properties**									
12	saturated water content, Θs	+	-	+	-	Table 4				
13	field capacity, Θ_{FC}	+	-	+	-					
14	residual water content, Θ_{RES}	+	-	+	-					
15	Measured values of soil water retention curve	+	-	+	-					
16	van Genuchten alpha	-	-	+	-					
17	van Genuchten n	-	-	+	-					
18	van Genuchten m=1-1/n	-	-	+	-					
19	wilting point	+	-	+	-					
20	available water capacity, $\Theta_{AV}=\Theta_{FC}-\Theta_{RES}$	-	-	-	+					
21	ratio of total to available water in soil = $\Theta s/\Theta_{AV}$	-	+	-	-					
22	saturated hydraulic cond., vertical	+	+	+	+					
23	saturated hydraulic cond. Lateral	+	-	-	-					
24	saturated hydraulic cond., matrix	-	-	+	-					
25	maximum infiltration rate (mm/day)	-	+	-	-		5			
26	Crack/byplass flow	-	-	+ (inactive)	+ (inactive)					
27	**Winter: snow**									
28	MeltCoefAirTemp	-	-	+	-			2 (F); 4(A)		
29	Temperature, above which all precipitation is rain	-	-	+	-			2		
30	Specific heat capacity due to freeze /thaw	-	+	-	-		7.1			
31	Rain/snow dividing tepmerature (deg C)	+	-	+	+	0		0	0	
32	Soil thermal conductivity coefficient	-	+	+	-		0.61	0.61		
33	Thermal conductivity coefficient b (W/m deg C)	+	-	-	-	0.39				
34	Thermal conductivity coefficient a (W/m deg C)	+	-	-	-	1.33				
35	Water retention capacity of snow	-	-	+	-			0.07		
36	Snow pack temperature lag factor	+	-	-	+	1			1	
37	Snowfall temperature	-	-	-	+				1	
38	Snow melt base temperature	+	-	-	+	1.5			1.5	
39	Maximum melt rate for snow	-	-	-	+				4.5	
40	Minimum melt rate for snow	-	-	-	+				4.5	
41	Snow melt degree-day coefficient (mm/dd-deg C)	+	+	-	-	2.24(F); 3(A)	2.24(F); 3(A)			
42	Water equivalent of snow	-	+	calculated	-		0.3	calculated		
43	Diurnal phase lag of air temperature (deg)	+	-	-	-	8				
44	Density of (new) snow (kg / m3)	+	-	calculated	+	100		calculated	100	
45	Critical ice content above which infiltration stops (cm3/cm3)	+	-	-	-	0.2				
46	**Winter: frost**									
47	FreezepointF0	-	-	+	-					
48	FreezepointF1	-	-	+	-					
49	FreezepointFW1	-	-	+	-					
50	Unfrozen water content as a function of temperature	+	-	-	-	Table 2.				
51	**Surface runoff/storage or surface water management**									
52	The maximum surface pool cover - SPMaxCover	-	-	+	-					
53	The potential surface cover - SPCovPot	-	-	+	-					
54	Amount of water on the surface at complete soil cover - SPCoverTotal	-	-	+	-					
55	Surface runoff (from surface pool) coefficient - SurfCoef	-	-	+	-			0.8		
56	Max amount of water stored in the surface wihtout runoff(cm)	+	-	+	-	0.5		0.5		
57	Kirkham's depth for flow to drains (cm)	+	-	-	-	0.5				
58	Surface runoff lag time (day)	-	+	-	+		0.5(F); 0.2(A)		0.5(F); 0.2(A)	
59	Curve number	-	-	-	+					
60	Manning's n value for overland flow	-	-	-	+					
61	Curve number calculation methods	-	-	-	+					
62	**Interception**									
63	WaterCapacityBase	-	-	+	-			0		
64	WaterCapacityPerLAI	-	-	+	-			0.2		
65	Maximum canopy storage	-	-	-	+				0	
66	**Transpiration**									
67	CanDensMax	-	-	+	-			0.7		
68	CondMax	-	-	+	-			0.02		
69	CondRis	-	-	+	-			5.0 E6		
70	CondVPD	-	-	+	-			100		
71	RoughLMin	-	-	+	-			0.01		
72	WindLessExchangeCanopy	-	-	+	-			0.001		
73	Lower limit of water conternt in the root zone (cm3 / cm3)	+	-	-	-	0.17				
74	Limiting water table (cm)	+	-	-	-	30				
75	Plant uptake compensation factor	-	-	-	+				0	
76	**Evaporation**									
77	Penman surface resistance parameter - psiRs_1p	-	-	+	-			200		
78	LAI cointribution to aerodynamic resistance - RaIncreaseWithLAI	-	-	+	-			50		
79	Surface roughness length for bare soil - RoughLBareSoilMom	-	-	+	-			0.001		
80	Calculate sublimation Yes/NO	-	-	-	+				yes	
81	Soil evaporation compensation factor - ESCO	-	-	-	+				0	
82	**Crop parameters**									
83	growth season starting date	-	time series	-	-					
84	growth period	-	time series	-	-					
85	growth curve offset	-	+	-	-		0.1			
86	growth curve amplitude	-	+	-	-		0.8			
87	maximum rooting depth	+	-	+	+	1(A); 2(F)		driving variable	1(A); 2(F)	
88	root distribution	time series	-	time series	-	Table 5		Table 5		
89	canopy height	-	-	time series	-					

Table 8: Example for harmonizing the parameters uses in different models

Solutions to upscale model results to the catchments

As presented earlier, upon model selection, one makes choices in terms of model theory and their capabilities. In terms of representing a catchment, the models selected for this project utilize three different approaches for water and nutrient routing. The chosen catchment models are all semi-distributed models. Fully distributed models are capable of capturing the true spatial distribution of input variables, but are extremely data- and calculation-intensive. Their use may not pay off in data-scarce situations at all. On the contrary, semi-distributed models are less data and calculation intensive. They consider smaller homogeneous units, sub-basins or hydrological response units (HRUs), for which the meteorological and physical parameters are lumped, and the model returns the 'average behavior' of these smaller units, which are then aggregated ('routed') for pre-defined locations downstream – and eventually for the catchment outlet. Among semi-distributed models, there are two dominant approaches to solve such routing. Some models rely on a true, geography-based routing using true GIS background. Such models in this study are SWAT and HYPE. Other models - like HBV, PERSiST and INCA in this study – only connect the output of one sub-basin as input to the next using retention parameters that are inputs. These solutions are integrated parts of the models, and are implemented via model input and parameterization during model run. Each model's core documentation describes the process.

Point models, such as the SOIL model, work differently. Those models can only model finite small areas by assuming that a soil profile represents that area homogeneously, and unlike in nature, the resulting surface and subsurface water flow cannot directly be routed to an adjacent area. In such cases, some solution is needed to channel water and solutes through the landscape and the river system, which solution requires calculations outside the model. The approach taken in this project is described herein.

Differences between the small homogeneous fields and the heterogeneous watersheds can be quite significant. Various types of land cover and soils can be included by dividing the watershed into subareas with similar characteristics. The zone of aeration extends from the surface to the water table and is usually thin in areas located close to permanent streams and quite thick in areas located far from streams and especially on hills. The soil profile with a thin zone of aeration will be saturated very quickly, forming surface runoff. The soil profile with a thick zone of aeration needs much more water for saturation and very rarely forms surface runoff. The contribution of the aeration zones with different thicknesses in the formation of a water flow and nitrogen transport depends on their areal fractions of the watershed. The quantity and quality of water fluxes from the fractions were calculated with the SOIL/MACRO and SOILN models and the results were aggregated to represent the entire catchment. Daily inflow to the river system from a watershed represented by N profiles was calculated as:

$$I_t = \sum_{i=1}^{N} I_{i,t} * k_i \qquad \text{(Eq. 1)}$$

where t is time, $I_{i,t}$ represents water discharge from the area represented by each field (profile) i at the time t, and k_i - the fraction of the watershed area occupied by the i-th soil profile. The discharge measured from the watershed can be used to calibrate k_i. This approach is commonly used in hydrological modelling and gives good results.

In this study, the measured water flow was used to calibrate k_i. It was mentioned above that different soil profiles produce different waves of water flow. For instance, soil profiles with a thin zone of aeration produce a water flow after any storm, including summer periods. On the contrary, soil profiles with a thick zone of aeration produce a water flow only during wet periods (autumn, winter and spring). Every year, there are usually dozens of waves formed under different conditions. Different impact of different soil profiles gives an opportunity to use the shape of the hydrographs to find the fraction of the whole watershed represented by the soil profile *i*.

Inflow at any point on the watershed travels a certain distance to reach the outlet of the watershed. During this travel it undergoes changes caused by channel storage. The transformation undergone by the inflow is due to (a) the translation effect and (b) the storage effect, consisting of a time lag and shape modification (Singh, 1988). Comparison of the results from a linear model with those from the St Venant equations applied for Estonian rivers (Vassiljev et al. 2004) showed that in common simple cases (without backwater effect or other complex phenomena), both of the approaches showed the same precision. Therefore, the linear river routing model was used in this investigation. The linear model for water movement is:

$$Q_t = \sum_{\tau=1}^{\max}[I_{t-\tau+1}h_\tau] + Q_{gr,t} \qquad \text{(Eq. 2)}$$

where Q_t is the water discharge at the outlet of the watershed at time *t*, *I* represents the inflow to the river system, h_τ is the ordinate of the response function, where τ represents the consecutive numbers of these ordinates (from 1 to max).

The response function may be approximated by a flexible function with a low number of parameters. Some standard optimization procedures may be used to find the parameters of the response function. In this study, the representation of the response function suggested by Kalinin and Miljukov (1957) was used. This can be described as:

$$h_\tau = \frac{1}{k\Gamma(n)}\left(\frac{\tau}{k}\right)^{n-1}\exp\left(-\frac{\tau}{k}\right) \qquad \text{(Eq. 3)}$$

where $\Gamma(n)$ is the gamma function, and *n* and *k* are fitting parameters.

Capabilities and limitations of the selected simulation models

Models differ in their capacity to be able to simulate the potential impact of mitigation measures, given the differences in their theory, architecture and included processes. A separate section is allotted for this subject in a later part of the report.

Models' input data

General data needs

The models used in this study generally need i) driving input data, consisting of time series of various meteorological data; ii) information on soil, land use, slope etc. and iii) reference data for model calibration (time series of measured discharge and water quality data).

Data availability for the pilot areas (sources, references to data providing agencies, services)

The chosen process-based models rely on a range of inputs, as dictated by their data needs. The user often faces a multitude of trade-offs while choosing a model and its options to handle inputs. It is because one may intuitively choose a more complex solution to describe a natural process, but the availability (or lack of) and quality-issues with environmental data may limit the benefits of using such complex processes. A realistic evaluation of the balance in available data and models and model options remains a challenging task for every modeler. We summarize the core data that were used in this project using the example of the SWAT model's case, and note that each model's data needs are somewhat different. We refer the interested reader to the models' core documentation for further details. The SWAT model requires the following core types of data:

- Physical inputs: DEM (Digital Elevation Model) elevation maps, land cover map and soil map including soil hydraulic properties – all of which are used to define the Hydrological Response Units (HRUs) and their core properties.
- Weather inputs: precipitation, air temperature (daily min and max), solar radiation, wind speed and relative humidity.
- Hydrological data: stream flow data.
- Point-source information: location, amount, timing.

Physical inputs

DEM (Digital Elevation Model)

A 10 meters resolution map of Estonia provided by Maa-amet (Estonian Land Board) http://www.maaamet.ee, was used for delineation and slope calculations (*Figure 10*).

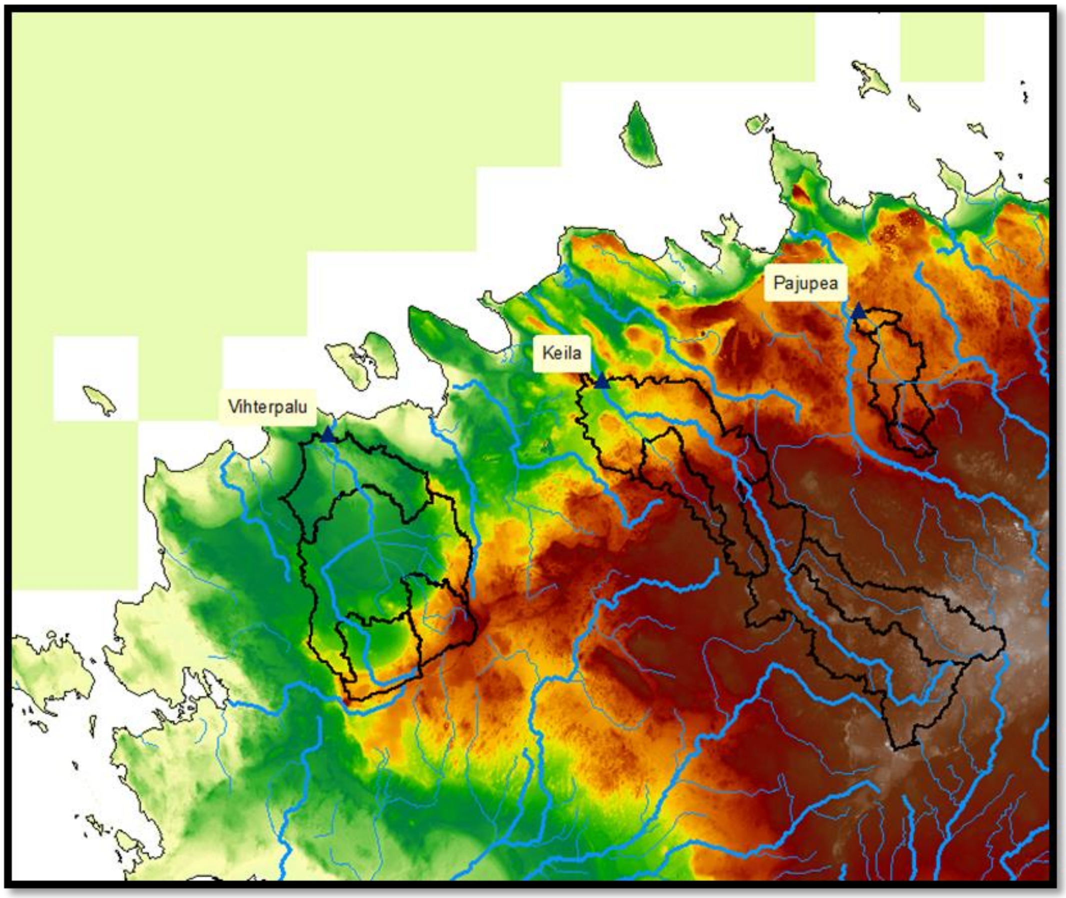


Figure 10: Digital Elevation Model of NW Estonia

The basin and sub-basins definition consist of terrain analysis using a digital elevation Model (DEM), to define the contributing area to any stream or river, which is more easily defined in mountain areas than in flat areas. All the basins and sub-basins were calculated using both permanent stations and temporary stations. We obtaining the following areas after the delineation (*Table 10*):

Basin Vihterpalu		Subbasin Vihterpalu1		Subbasin Piirsalu 1		Subbasin Vihterpalu2		Subbasin Vihterpalu	
Total Area km2	480.18	Total Area km2	72.48	Total Area km2	55.08	Total Area km2	247.14	Total Area km2	105.48
Basin Keila		Subbasin Keila1		Subbasin Keila2		Subbasin Maidla		Subbasin Keila	
Total Area km2	631.79	Total Area km2	127.20	Total Area km2	240.53	Total Area km2	51.40	Total Area km2	212.66
Basin Leivajõgi		Subbasin Leivajõgi 1		Subbasin Leivajõgi 2		Subbasin Leivajõgi 3		Subbasin Pajupea	
Total Area km2	84.85	Total Area km2	12.20	Total Area km2	18.45	Total Area km2	44.09	Total Area km2	10.11

Table 10: Basin and sub-basin areas of the 3 catchment areas delineated using DEM.

Land Cover map (CLC)

The Corine Land Cover (CLC) 1:100.000 raster map has been used in the project. The CLC consists of a geographical database describing vegetation and land use in 44 classes, grouped in three nomenclature levels. It covers entire Europe and gives information on the status and the changes of the environment. We were using the CLC 2006 raster map 100m, version 12/2009 for Estonia but it was necessary to transform the CLC classification codes into the land cover/plant classification system (4-letter codes) that SWAT recognizes. We present the applicable map in this section (*Figure 11*), but presented and discussed the resulting classifications in the chapter that describes the study areas.

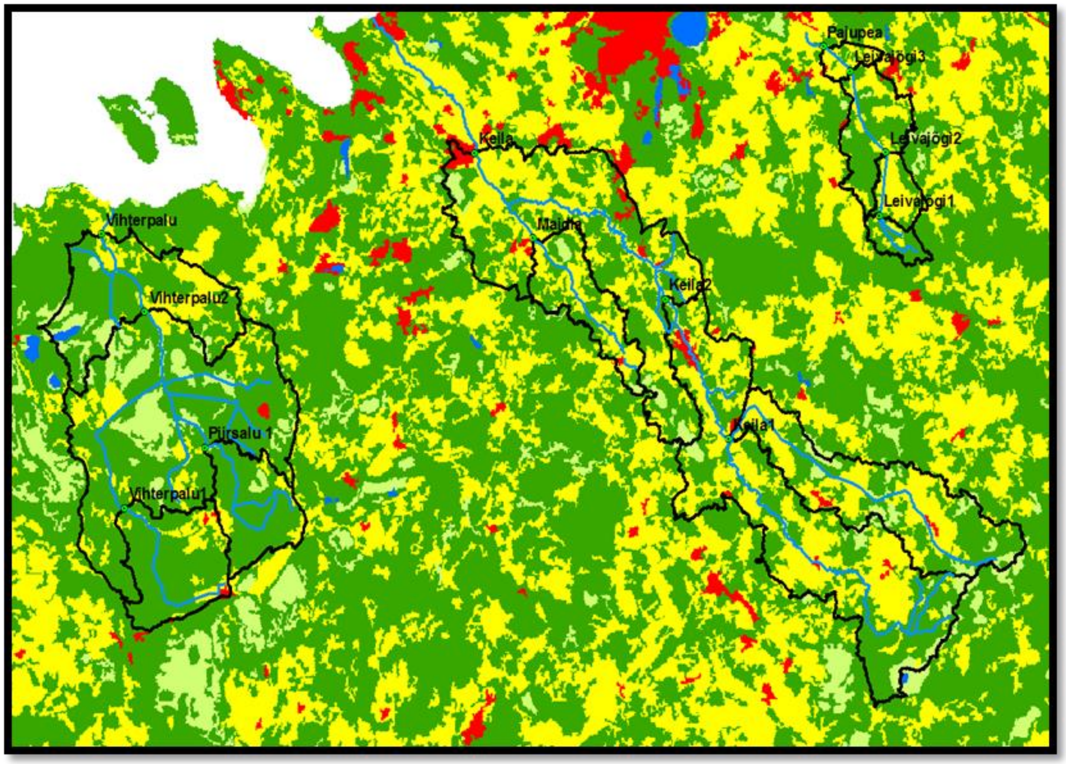


Figure 11: Corine Land Cover (CLC) LABEL 1 Classification

Soil map

We used the ESDB (European Soil Database) map of dominant soils for Estonia as source of basic soil information (*Figure 12*).

The core soil database available for the project corresponds to the Harmonized World Soil Database (HWSD) v 1.2, and represents the dominant soils for Estonia. It is a 30 arc-second raster database in which the soils' spatial distribution is presented, along with aggregated information on soil profiles and their associated properties. Soil properties – namely soil hydraulic properties - needed for process-based modeling are usually not readily available in such databases, and thus need to be estimated from other soil information or using expert judgment at the worst case. A number of soil (physical) properties were available in the database, which were relevant towards estimating a number of the necessary model inputs. These properties – presented for top- and sub-soils separately - are:

- Soil Classification according to the FAO 1985 system
- Clay content (% of particles of diameter equivalent <0.002 mm in the <2mm (fine earth) fraction)
- Silt content (% of particles of diameter equivalent between 0.002 to 0.05 mm)
- Sand content (% of particles of diameter equivalent between 0.05 to 2 mm)
- Organic carbon content (%)

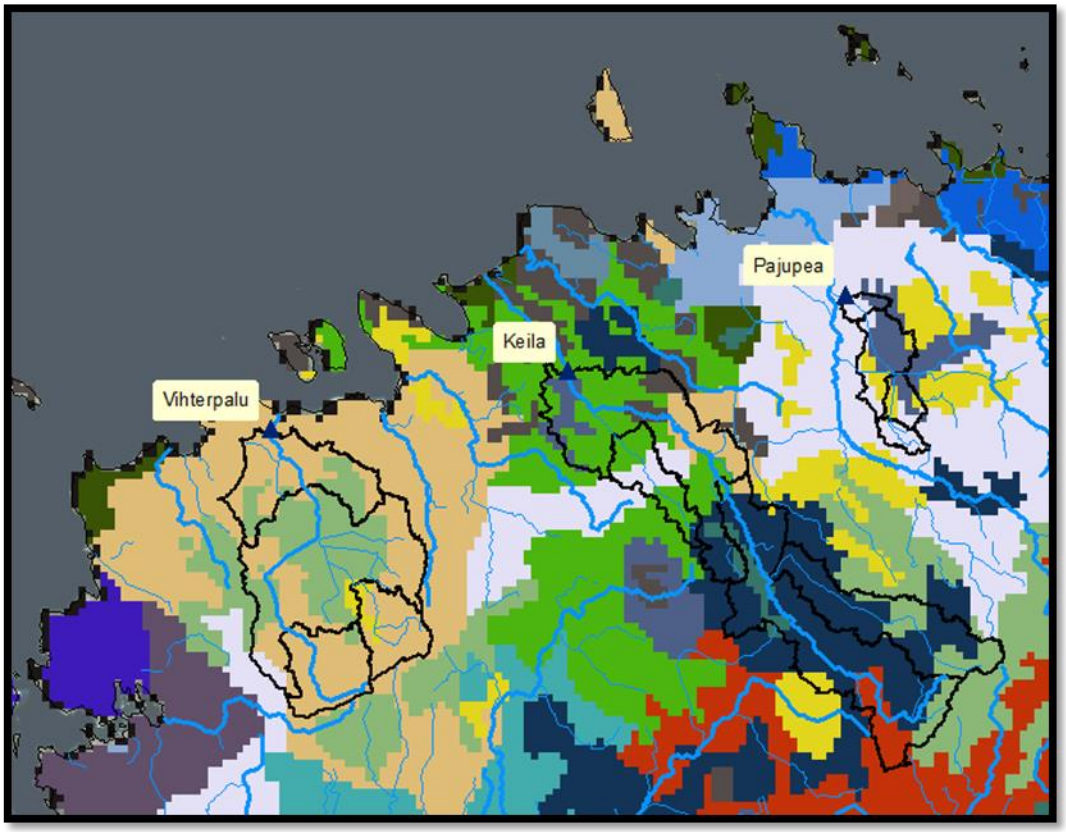


Figure 12: Map of dominant soils for Estonia (ESDB (European Soil Database))

The SWAT model, however, requires information on a different list of soil and soil related properties, which are seldom part of such generic databases. These are as follows:

- Hydrological Group (A B C D)
- Maximum rooting depth of soil profile (SOL_ZMX)
- Moist bulk density (g/cc) (SOL_BD1)
- Moist soil Albedo (SOL_ALB1)
- Soil erodibility (K) (USLE_K1).
- Available water capacity (cc/cc) (SOL_AWC1)
- Saturated hydraulic conductivity (mm/hr) (SOL_K1)

In order to successfully complete the simulation runs, this project used a range of tools in different estimation schemes to obtain the above listed soil properties. Those schemes are presented in the next chapter on inference to missing data.

Weather inputs

Weather data from the Meteorological stations network were used for the core simulations from the following stations: Pakri MJ, Lääne-Nigula MJ, Tallinn-Harku AJ, Kuusiku MJ. Among the required weather related data, precipitation is typically that shows the largest spatial heterogeneity, and so increasing the density of precipitation data is always desirable. Therefore, we also included the Kehra location that has additional precipitation records. All these stations belongs to Keskkonnaagentuur (Estonian Environment Agency) and all data were collected at a daily time step for the period 2000-2011. We used the following information for the simulations to establish the 'atmospheric demand' in the model, representing potential evapotranspiration:

- Precipitation (mm/day).
- Air temperature (Min and Max) (Degrees Celsius).
- Solar radiation. (Mj/m2/day).
- Wind speed (m/s).
- Relative Humidity (%).

Figure 13 in the next sub-section presents the geographical distribution of meteorological stations. One can observe that all the rain gauges in this network are located outside the river basins – none of them are inside any of the catchments. This challenge triggered the idea of experimenting with an alternate weather-data source, the application of radar-based data, which is detailed in a separate chapter below.

Hydrological data

Stream flow data measurements are not necessary as input to run the model, but rather they are essential for model calibration and validation. The following stations were of use in this project:

- Three stations from the hydro network (Vihterpalu, Keila and Pajupea) for calibration and validation purposes,
- Six temporary stations (locations of interest to get outputs from the model) (*Figure 13*):

    - **Vihterpalu basin**: Vihterpalu 1, Pirsalu 1, Vihterpalu 2.
    - **Keila basin**: Keila 1, Keila 2, Maidla.

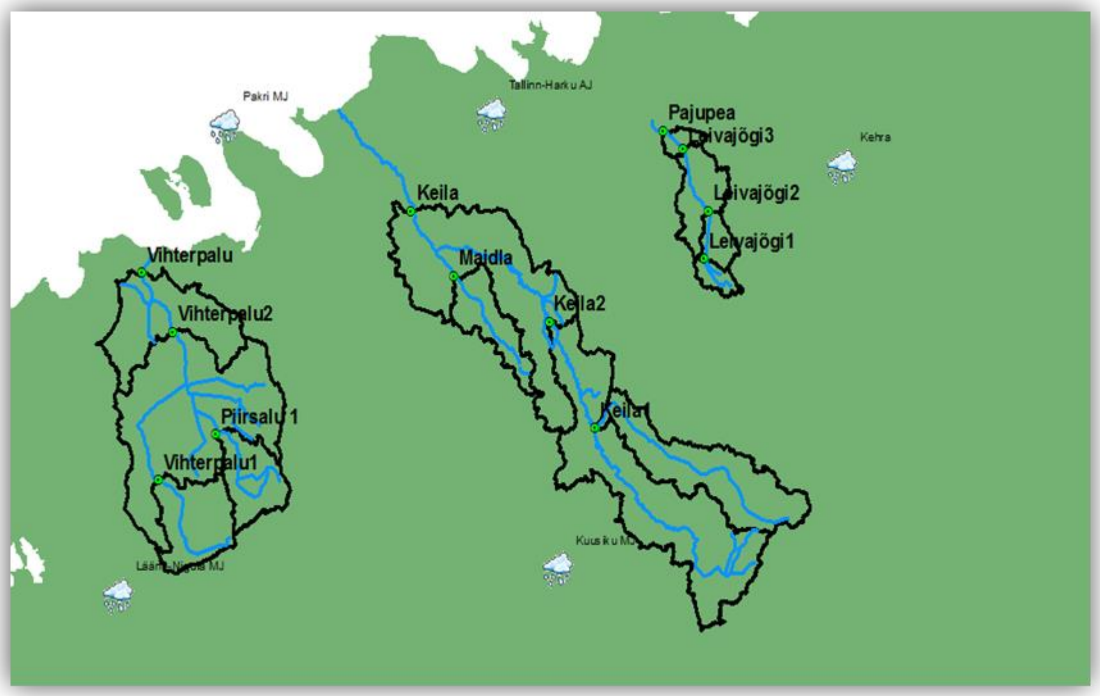


Figure 13: Meteorological stations, hydro network stations and locations of interest

We were installing a Keller water level sensor in each location of interest from May 2014 to November 2015, to collect water level readings in an hourly time step; while a field work campaign began simultaneously to measure river discharge in order to collect enough measurement data to establish a rating curve.

Point source pollutants

Wastewater effluents were used as direct point source input for the INCA and SWAT models. Data was obtained from the relevant database maintained by the Estonian Environment Agency. Up to year 2009, point sources loads data was available only annually, while since 2010 loads are measured four times a year. The variables that are input to the simulation models are: wastewater discharge, total nitrogen, total phosphorus, total suspended sediments. For the two models, the following subdivision of total nitrogen and phosphorus loads was made based on expert opinion:

Nitrate (NO_3)	50% of total nitrogen
Ammonium (NH_4)	40% of total nitrogen
Organic nitrogen	10% of total nitrogen
Phosphate ($PO4^{3-}$)	80% of total phosphorus
Organic phosphorus	20% of total phosphorus

Land management data

Data on fertiliser application in Estonia is available only at the national level by National Statistics, but it did not provide sufficient certainty on local application rates and timings, and therefore we didn't use it. Fertilization and tillage information was input based on expert knowledge and after coonsulting the calibration data on water quality.

Specific data needs of models (only the significant ones need to be listed for general understanding)

The HYPE model within the Airviro system used a time period and weather input that was different from the other models. The task that was performed with HYPE was different than that with the other models: while the three catchments were evaluated, HYPE was used to simulate the hydrology of the entire country of Estonia. This, obviously, required weather input for the entire country, which could not be satisfied using the data collections that we listed elsewhere in this report – i.e. 5 meteorological stations or the Tallinn-Harku radar. In Airviro, a module has been developed that extracts precipitation and temperature input directly from the output of the High Resolution Limited Area Model (HIRLAM). The HIRLAM model is maintained by Keskkonnaagentuur in Estonia, and has a spatial resolution of 11 km. Naturally HIRLAM is a model for weather forecasting and the forecasts are renewed at regular time steps, as more and more recent actual observations become available. When a real observation becomes available, the forecast product is substituted with the observed data and the Airviro module collects the real data. By using HIRLAM model data, good spatial resolution has been achieved, and as an extra benefit, data for transboundary catchments is also provided. The limitation of using these data is that HIRLAM is only available since 2009, and we did not find a way to use data from observation stations for preceding years, which situation is likely to improve in the future.

Inference to missing data

Users more often than not are in the situation that there are either gaps in their data series or required variables are simply not available. In such cases, various techniques are applied as tools of inference for the required data. Here we cite two such cases that required action in this project:

(a) the estimation of soil hydraulic and related properties for the HRUs,
(b) the interpolation of weather data from locations outside the watersheds.

Estimation of soil (hydraulic) properties

Estimation or derivation of these properties required a range of techniques and considerations, which are described below. After deriving each of those properties for each unique soil layer, they were coupled to the applicable map units.

Hydrological Group

The U.S. Natural Resources Conservation Service (USDA-NRCS) classifies soils into four main hydrological groups, based on their infiltration characteristics, i.e. presenting different potential for surface runoff. Soil properties that influence runoff potential are those that the minimum infiltration rate for a bare, un-frozen soil, such as the depth of seasonally variable water-table, saturated hydraulic conductivity, and/or the depth to a restrictive/non-permeable layer. The grouping took place using expert judgement after consulting the detailed criteria and the available soil information from the soil map.

Maximum rooting depth of soil profile

Generic soil maps cannot provide locally relevant and crop specific rooting depths. We followed common practice for cases when there is no knowledge of substantially restrictive layers, and the modeled soil profile is not very deep, and established rooting depth to be equivalent to the depth of the modeled soil profile, which was 1m in this study.

Moist bulk density

The proper estimation of moist bulk density is hampered by the fact that international databases either lack information on the definition of how stored bulk density data were obtained, or those are typically dry bulk density values.

Documentation on the Harmonized World Soil Database v 1.2 suggests one of two processes to estimate soil bulk density. Upon reviewing the results of both processes, we have opted to retain the values derived based on the analysis of data in various regional SOTER databases (i.e. the SOTWIS Database). This process allowed the estimation of bulk density by soil type and depth, based on available data on soil texture, organic matter content and porosity. Similarly to what has been documented for HWSD, we observed much more realistic estimates for specialty cases (e.g. extreme high OC Histosols) using this method than using the alternative one.

Moist bulk density was then subsequently used to calculate saturated water content, in order to avoid potentially substantial mismatches between the two variables, in case another estimation or calculation technique is used.

Moist soil Albedo

There have been multiple ways suggested to approximate moist soil albedo, and we adapted the approach proposed by the developers of the SWAT model, in order to reduce the uncertainty introduced by differences in understanding, as well as due to the availability of the necessary input. The procedure calculates albedo as a function of soil organic matter/carbon content.

Soil erodibility (K)

A number of approaches are recognized worldwide to approximate the soil erodibility factor, which expresses the inherent erodibility of a particular soil. Upon choosing a method to approximate the applicable value, one needs to consider the great uncertainty in each of those values – which becomes evident when methods are compared to each other – and the availability of the necessary variables to derive the estimate. For example, soil structure information is more rarely available than soil texture information.

We have chosen to adapt the look-up table derived by Stewart et al. (1975) for US-EPA, which indicates the general magnitude of the K-factor as a function of organic matter content and soil textural class. Every value was individually judged, and expert judgement was used for soils that were out of range for the table (i.e. Histosols with very high OC).

Available water capacity and saturated hydraulic conductivity

Methodology to derive these variables is described together, given that the same advanced estimation method and the same data source was used.

Available water capacity has been defined as the difference between the amounts of water held at 'field capacity' minus the amount held at 'wilting point'. Latter is commonly approximated by the laboratory measured water content at -1500 kPa pressure, while the earlier has a definition that is still among the most debated definitions in soil physics. Nevertheless, since the SWAT model is to be used, the definitions understood in its documentation were adhered to: the amount of water held at -33kPa is considered as 'field capacity', with the option to consider a different value (-10kPa) for sands, where sands are loosely defined. We note that this approach is primarily adhered to in the U.S.

Estimation of the above properties, and of saturated hydraulic conductivity, requires an external source-database. Since such database was not available for the project from Estonia, we looked into the potential use of internationally available soil hydraulic databases, such as that of USDA-NRCS, the data compilation behind the Rosetta estimation software, the European HYPRES database, or the also European EU-HYDI database. Given that EU-HYDI is a database that was assembled from many European countries, it is recent (2014) and has the greatest depth of relevant information we anticipated that this database has the greatest relevance for Estonian conditions and for the given task.

A subset of soils were used from the EU-HYDI database that had original measured data (i.e. not interpolated) on the above water retention points, saturated hydraulic conductivity, as well as soil particle-size fractions according to the FAO system, as well as organic carbon content and bulk density. The selection was subsequently error checked to discard any obviously erroneous data/samples, which yielded 2269 samples. The method introduced by Nemes et al. (2006) has been used to estimate the given properties. This encompasses using a k-nearest neighbor algorithm to provide local estimates of the given soil hydraulic properties, based on the source-database. While details of the method are not provided here, we mention that by 'local estimate' we refer to the specialty of this technique (vs. nearly all other known estimation techniques in soil hydrology) that it is not based on a generic set of equations, but an estimate is formulated for each sample real-time using only samples that resemble the properties of the actual sample (i.e. that are its nearest neighbors). The algorithm we used also embedded a sub-sampling algorithm, which allowed for multiple estimations from the same source-database, and thereby the calculation of an uncertainty metric to each of the estimated values. For each of the represented top and subsoils in the Estonian soil map, we therefore generated estimates of -10kPa, -33kPa and -1500kPa water retention, and saturated hydraulic conductivity, and each value had an uncertainty metric associated to them. Calculation of the available water capacity was then calculated from the relevant values, using -33kPa water contents 'field capacity' for all soil textures.

Meteorological data – a special case with radar data for future applications

Here we refer back to a statement made in the section on collecting weather data for modeling. The available meteorological stations were scarce and fell outside the catchments (*Figure 13*). Therefore, while those were the only 'hard-data' available on the required climatic variables, we understand that (1) there is a degree of uncertainty associated with those, and that (2) this situation is not expected to change soon, which will affect similar future studies as well. In response, in order to assist potential future applications, we experimented with an alternate weather-data source, radar-based precipitation data.

Harku radar was installed in 2009 but became fully operational in 2010. It is a Vaisala WMR200 dual polarization Doppler weather radar that scans in 15 minute sets. The radar data was collected from the radar files archives (IRIS raw file) for January 2010 until 31 of December 2011. We note that this period is much shorter than the available time-series using the meteorological stations, and therefore at this stage it is more sufficient for testing the methodology than to derive overarching conclusions.

We were interested in precipitation accumulation in (mm/d). The daily precipitation is based on the horizontal reflectivity (dBZ) at 0.5 km elevation and the pCAPPI product from Harku radar. The Z-R relationship was used to convert radar reflectivity to rainfall intensity (mm/h) as Z=200R^1.6, where Z is the corrected reflexivity dBZ, and R is the rainfall rate in mm/h. Rainfall values were accumulated at 18UTC as daily precipitation in order to match the accumulation period of the radar data with that of the meteorological stations data (PR24He at 18UTC). Finally, for each location, the daily precipitation for each location (meteorological stations and grid points), in total 308 files altogether, were extracted from the radar archives in a csv format.

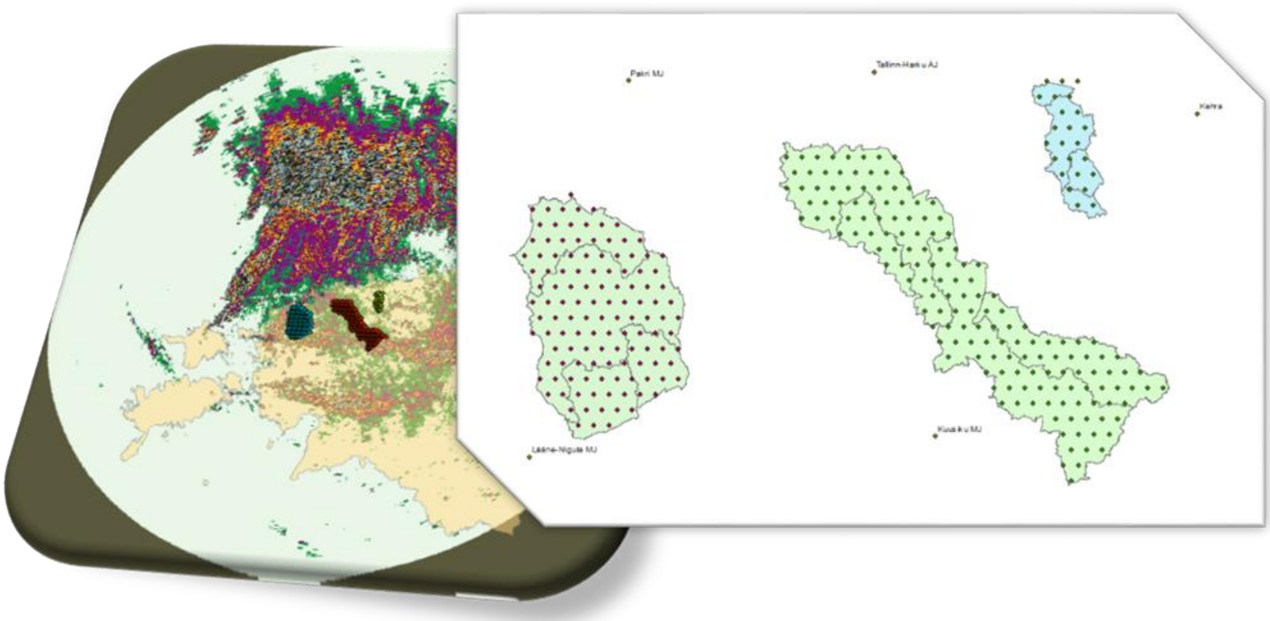
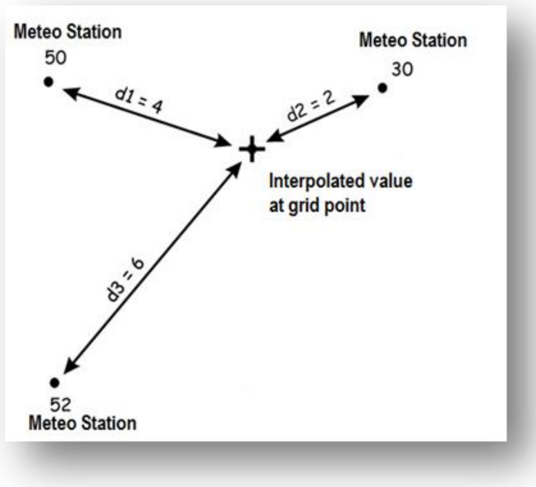


Figure 14: Example image of Tallinn-Harku radar precipitation (left) and the derived 2km grid used in this project.

There was a need to synchronize precipitation from weather stations and the Harku radar to facilitate working with the SWAT model. SWAT uses the approach to distribute the precipitation in the basin, that each sub-basin takes the precipitation data from the precipitation station that is closest to the centroid of that sub-basin. This approach was used during the calibration and validation period of 2000-2011, by using precipitation data only from the meteorological stations.

To use the radar data we had to bridge a large temporal and spatial gap between data from the data sources. In order to resolve this mismatch, we were interpolating the rain gauge data at the grid points for the period 2000 to 2009 using the Inverse Distance Weighted (IDW) method, and then added the radar data extracted at same grid points for 2010 to 2011. We decided to create a grid of 2x2km to extract the daily precipitation from the radar at each point. This process has given us 121 points in and adjacent to the Vihterpalu catchment, 157 for Keila and 24 for Leivajõgi. We also extracted daily radar data at the same locations where the meteostations (Pakri MJ, Lääne-Nigula MJ, Tallinn-Harku AJ, Kuusiku MJ and Kehra HJ) are.



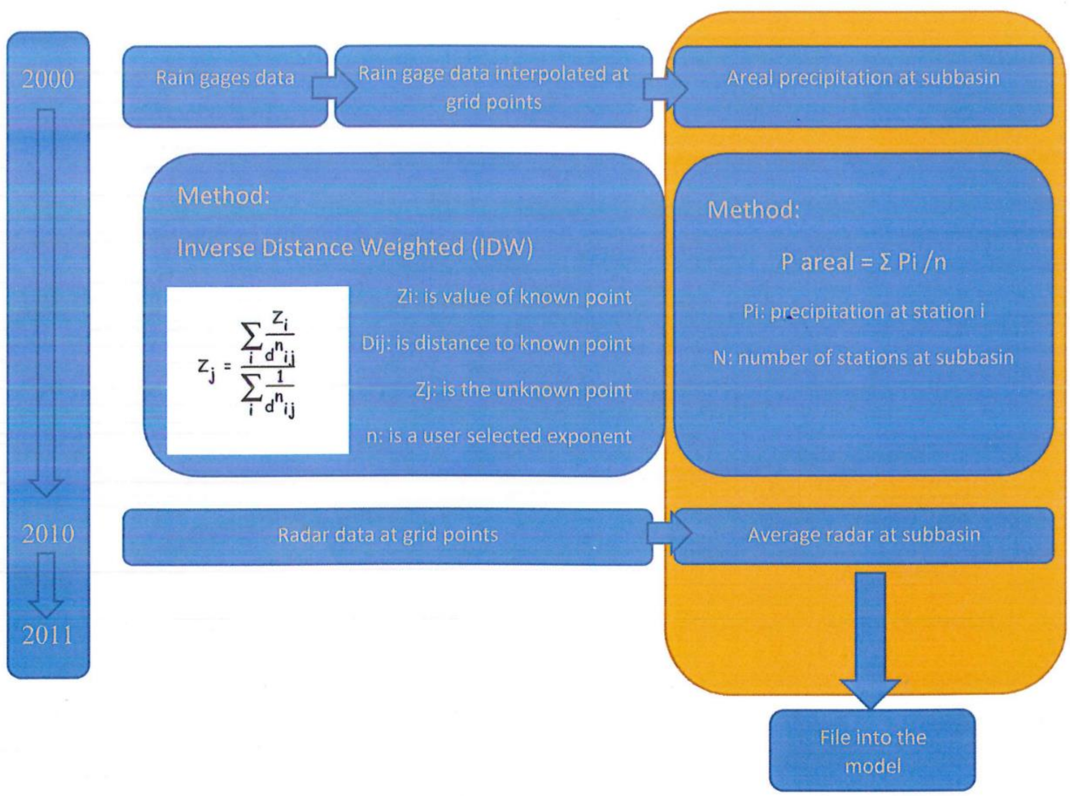


Figure 15: Schematic representation of the base idea of Inverse Distance Weighting and supporting the SWAT model with weather data from the radar and from meteorological stations.

Given the approach the SWAT model takes in distributing the precipitation, if one sub-basin contains more than one grid point, only the closest grid data point to the centroid of the sub-basin will be taken and the rest will be dismissed. In order to avoid losing the spatial precipitation pattern, the areal precipitation was aggregated to get the average precipitation over the sub-basin area, obtaining a final daily value of precipitation per sub-basin. It is still a single value for the sub-basin, but it relies on multiple grid points.

The precipitation input used during the calibration and validation period 2000-2011 originates from the meteo stations themselves, without any modifications, allowing the model to distribute the precipitation in each sub-basin. One of the aims when implementing radar data was that it has then become possible to combine the interpolation at grid points, adding the radar data at these grid points and generating an indication of areal precipitation in each sub-basin. This resulted in having a 'virtual station' per each sub-basin. This extra precipitation dataset was added into the model, replacing the original dataset to verify if there is any improvement in the efficiency of the model.

Model calibration and validation

Why is it necessary

While some challenge the notion, it is dominantly seen and accepted in the modeling community that it is necessary and beneficial to calibrate and validate a simulation model against local benchmark data. If such data do not exist or are insufficient or unreliable, it may be argued that calibrating the model does not make much sense. Nevertheless, the modeler – and especially any stakeholder that relies on the modeling results – needs to understand the risks that this poses: using a model that has been developed elsewhere and applying any recommended, generic, parameter sets have been developed elsewhere carries enormous risks and uncertainties – the size of which are unknown. To mention a classic example, the rainfall-runoff factor in the otherwise simple Universal Soil Loss Equation (USLE) that was established in the US could not be used in the UK, given that the daily amount of rainfall fell in a completely different intensity and pattern in the two countries (short, intensive storms vs. prolonged drizzle).

It is therefore understood by this modeling group that calibration and validation on independent benchmark data is really necessary. Model calibration entails setting up a base run for the model, and comparing the simulated and measured benchmark data using the selected statistical metrics. In the likely scenario that this first run does not yield the best possible results, selected (recommended) model parameters, coefficients are then iteratively modified by the user - in either a trial-and-error manner, or by using an auto-calibration tool that is provided with the model (e.g. SWAT-CUP for SWAT) – and the output patterns and statistical metrics are checked again. The effective, systematic adjustment of such calibration parameters requires expert understanding of their function and inter-relations, and what type of change is expected because of their modification.

Calibration, for instance, targets to help avoid e.g. systematic biases in the simulations, helps avoid a pattern-like behavior in estimation errors, or helps the model in reflecting the occurrence of any extreme events that are often missed. In catchment hydrology, a multitude of factors affect each and every case that suggest that the same parameter sets are not applicable in different locations. The model user often faces hard choices when it comes to performing the model calibration, as adjustments often work against each other, and also because improvements at one end often cause a decrease in simulation accuracy at the other end. What criteria to use, what order of preference to keep in adjusting parameters, and in general the question of when should calibration be deemed ready are all decisions that are hard to give (or get) advice on. The targets to reach are also often dependent on the overall goal of the study; while one study may seek good annual averages for the benchmark variable, others may be more interested in matching extreme events, even if the annual averages returned by the model are sub-optimal.

Model calibration can be a daunting task, and it often becomes the most laborious part of setting up and running an environmental simulation model. Nevertheless, this is the only tool that is in the hand of a modeler to ensure that systematic biases – in space and time – and scale issues (e.g. extreme events) are addressed as best as possible.

Model calibration and validation in the NORRA project

As introduced, prior to model calibration that user typically faces a number of hard choices, as it is rarely the case that all the necessary information is available for the modeler. In this project the following tasks had to be addressed.

- Decision about the reasonably attainable temporal and spatial scale
- Collection of data, and decisions about methods and/or alternate sources to fill data gaps.
- Decisions about preferred and attainable model processes for individual models as well as for reasons of later model comparability

Upon inquiring for and collecting the necessary and available data, the group consulted again and agreed on compromises that had to be made due to both model limitations and data availability. The elaborate process of model calibration was made by the individual modelers named earlier, or in some cases by joint work of two modelers. Apart from the SWAT model, which offers the option of using an auto-calibration tool called SWAT-CUP, all other models were manually calibrated. This process is most often very elaborate, and requires a great depth of expert knowledge and experience.

Out of the available data, it was decided in the consortium that in general the period of 2000-2005 was going to be used for model calibration, and the successive period of 2006-2011 was going to be used as a period of model validation. Model validation, in this context means that the parameterization optimized and established for the calibration period is used to simulate the validation period as independent data, without any attempt to match the measured field data via any parameter or other adjustment.

In certain sub-tasks, such as e.g. the experimentation with the use of radar-based precipitation data as input, a different calibration and validation period was used due to differences in data availability. These are noted at the appropriate sections.

Model calibration for catchment hydrology and for sediment and nutrient losses were done in separate steps. Hydrological processes are primary drivers of sediment and nutrient transport processes. Some models – or their combinations – require an established hydrology background to be able to handle particulate and nutrient transport processes. It is also desirable to separate these processes in order to simplify the model calibration, which still involves the adjustment of dozens (if not more) of parameters for each such module.

Metrics of model evaluation

Model performance was evaluated using three statistical types of metrics, each capable of evaluating somewhat different aspects of the simulations: the Nash-Sutcliffe model efficiency coefficient (NS), the regression coefficient (R^2) and the bias percentage (PBIAS).

The Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970) is a dimensionless, normalized statistic that determines the magnitude of the residual variance relative to the variance in the measured data. The Nash-Sutcliffe efficiency ranges from $-\infty$ to 1; improved model performance is indicated as the NS approaches 1, while a value of zero or negatives indicate that simulated values are no better than the mean of observed values. NS is calculated as:

$$NS = 1 - \left(\frac{\sum_{i=1}^{n} (Q_i - Q_i')^2}{\sum_{i=1}^{n} (Q_i - \bar{Q})^2} \right)$$

where Q_i is the measured value (discharge), Q_i' is the simulated value, $\bar{Q}$ is the average measured value, and n is the number of data points.

The regression coefficient (R^2) is a standard regression type metric that has been widely used for model evaluation in literally every natural science. The R^2 value describes the degree of collinearity in the measured and simulated data. R^2 ranges from 0 to 1, with higher values indicating less error variance, and typically values greater than 0.5 are considered acceptable. R^2 is calculated as:

$$R^2 = \frac{\left[\sum_i \left(Q_{m,i} - \overline{Q}_m\right)\left(Q_{s,i} - \overline{Q}_s\right) \right]^2}{\sum_i \left(Q_{m,i} - \overline{Q}_m\right)^2 \sum_i \left(Q_{s,i} - \overline{Q}_s\right)^2}$$

where *m* stands for measured, *s* indicates simulated, and all other notations are as for NS (above).

Percent bias (PBIAS, %) is an error index, that indicates the average tendency of the simulated data to be greater or smaller than the corresponding observed data. The optimal value of PBIAS is 0, small absolute values are indicating accurate model simulation. Positive values indicate model underestimation, and negative values indicate model overestimation (Gupta et al., 1999).
PBIAS is calculated as:

$$\text{PBIAS} = \left[\frac{\sum_{i=1}^{n} \left(Y_i^{obs} - Y_i^{sim}\right) * (100)}{\sum_{i=1}^{n} \left(Y_i^{obs}\right)} \right]$$

Using a combination of performance indicator types helps in obtaining a robust idea on the performance of a simulation model. Ideally, one obtains high NS and R^2 and low PBIAS, but it is rarely this simple in practice. While there is no consensus on specific coefficient values for the daily time step, we present the performance ratings for monthly simulations (*Table 11*), as suggested in e.g. Moriasi et al. (2007).

It has to be noted that it is much more difficult to achieve good modelling statistics at the daily time step than at monthly or annual steps. This is because at the monthly time scale, a lot of smoothing is taking place, given the time-scale of the most dominant processes in catchment hydrology and sediment transport, and so at the monthly step there is a significant degree of smoothing involved, which is easier for the model to capture. In other words, it is much easier to predict a mean value (or similar) than to predict the fluctuations.

Performance Rating	RSR	NSE	PBIAS (%)		
			Streamflow	Sediment	N, P
Very good	0.00 ≤ RSR ≤ 0.50	0.75 < NSE ≤ 1.00	PBIAS < ±10	PBIAS < ±15	PBIAS< ±25
Good	0.50 < RSR ≤ 0.60	0.65 < NSE ≤ 0.75	±10 ≤ PBIAS < ±15	±15 ≤ PBIAS < ±30	±25 ≤ PBIAS < ±40
Satisfactory	0.60 < RSR ≤ 0.70	0.50 < NSE ≤ 0.65	±15 ≤ PBIAS < ±25	±30 ≤ PBIAS < ±55	±40 ≤ PBIAS < ±70
Unsatisfactory	RSR > 0.70	NSE ≤ 0.50	PBIAS ≥ ±25	PBIAS ≥ ±55	PBIAS ≥ ±70

Table 11: General performance ratings for simulations at the monthly time step (Moriasi et al., 2007)

Results

Modeling catchment hydrology

Model calibration and validation results

The evaluation statistics of both model calibration and validation are provided in *Table 12*. Figures of calibration/validation model runs vs. the available flow data are provided in Annex for the interested reader. First, a few notes need to be provided.

1. It was decided early in the project that the MACRO model will only be run for one selected catchment in order to be able to evaluate the effect of its capabilities without much burden on the modelers.
2. Some models could not be successfully run for some of the catchments. This was due to some unknown model stability issues, which would have required an excessive effort – and joint work with the model developers – to overcome. It was therefore decided by the working group that pursuing the goal of completing the table below without gaps is beyond the means of this project and should be left as a future task.

DISCHARGE	STAT	CATCHMENT BASED MODELS								PROFILE BASED MODELS			
		HBV		PERSIST		SWAT		INCA		SOIL		MACRO	
		CALIB	VALID	CALIB	VALID	CALIB	VALID	CALIB	VALID	CALIB	VALID	CALIB	VALID
Keila	N-S	0.52	0.49			0.71	0.64	0.65	0.42	0.78	0.69		
	PBIAS	2%	2%			3%	-5%	9%	16%	-1%	-3%		
Vihterpalu	N-S	0.14		0.63	0.63	0.66	0.59	0.55	0.41	0.71	0.62		
	PBIAS	31%		-17%	-22%	2%	17%	14%	24%	-1%	7%		
Leivajogi	N-S			0.63	0.70	0.62	0.62	0.33	0.35	0.55	0.5	0.71	0.55
	PBIAS (%)			16%	23%	9%	-1%	40%	45%	14%	11%	-7%	6%

Table 12: Results for catchment hydrology using catchment- or point-models. Color coding reflects the rate of success of the model calibration and validation according to the recommendations by Moriasi et al., (2007). (Table 11)

The results reflect generally successful simulations of water flow patterns in the three catchments, using the various introduced models. In general, modeling the hydrology of the Vihterpalu catchment – the catchment with the least amount of anthropogenic influence was least challenging, although the simulation of Keila's hydrology was rather successful too. The modeling of Leivajogi presented both weaker model results, as well as model instability problems. We have encountered a number of challenges in both calibrating and validating the models. We highlight a few specific cases that will help present some concerns and findings of the hydrology modeling task.

Figure 16 presents an example of what often happens – and happened in this project – when simulating soil hydrological response. The presented data shows how the SOIL model's ability to simulate peak flows varied. The fact that the model often missed the observed peaks – primarily their volume, not their timing – is most likely due to not being able to simulate the quick pathway of macropore flow, by which the retention time of water in/on land was prolonged. Water is then simulated to either leave the area with a time-delay and potentially different pattern of outflow, or leave the simulated system via another process (e.g. evaporation, plant-uptake) or as a loss (e.g. groundwater recharge). In either case, this is a water balance concern for the modeler.

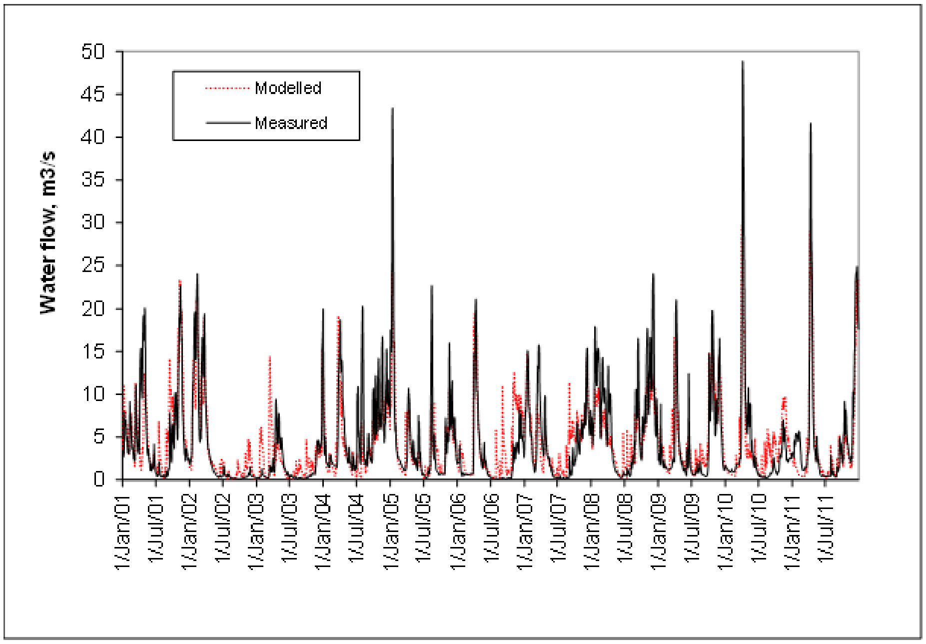
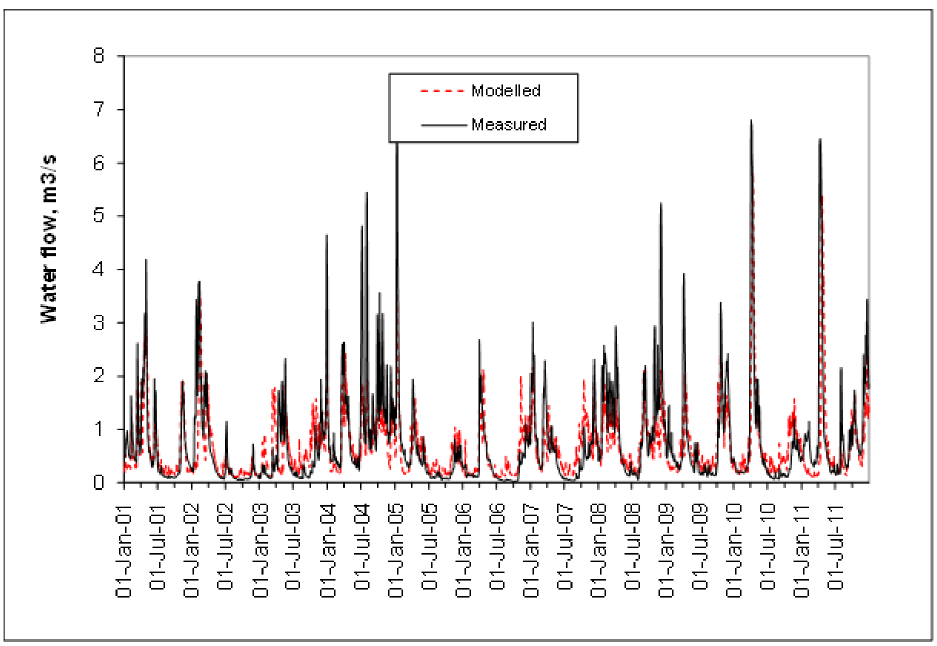


Figure 16: Simulation of the hydrology of the Vihterpalu catchment for the entire 2001-2011 period using the SOIL(N) model.

To indicate how differently a macropore-capable model can handle this aspect of water redistribution, we present the comparison between the SOIL and MACRO models for the Leivajogi catchment in *Figure 17*. The MACRO model was substantially more capable of simulating the flow peaks than the SOIL model; on occasion even overshooting the peak. The improved calibration and validation metrics with the MACRO model are also apparent from *Table 12*.



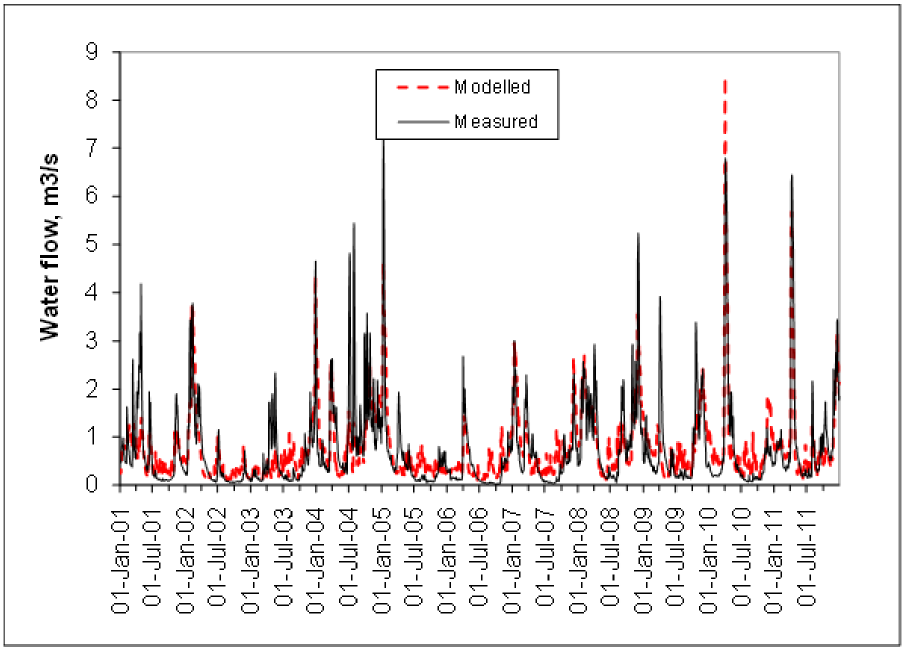


Figure 17: Simulation of the hydrology of the Leivajogi catchment for the entire 2001-2011 period using the SOIL(N) model (top) and the MACRO model (bottom).

Related to this point is that models – especially macropore-capable models - should be enhanced to account for artificial (subsurface) drainage processes that are important pathways to the transport of large volumes of water in a short period of time. Drainage itself, but especially when coupled with macropore flow, is a fast-flow process that carries great importance in describing the hydrological balance of an area; failure to capture this process will result in misjudging the amounts and rates of water and nutrient transport in the environment.

Utility of alternate sources for meteorological data

An attempt was made to perform and present a case in which the workgroup used a novel type of data source to obtain an aerial coverage of precipitation data. As presented, radar-based precipitation data was used as input to the SWAT model. A separate report by some authors (Manuel Garcia et al.) in the same team and project present the case in more detail, we only present and discuss a single, representative case.

Figure 18 presents a comparison of weather-station and radar-based precipitation data. It is obvious, and the pattern was similar for the other locations as well, that the currently available radar-based data severely underestimates the precipitation amounts signaled by weather stations; where the differences often exceed 50%. Consequently, it is expected that when such weather data are used, simulated water flows will be much lower than when station data are used.

	1	2	3	4	5	6	7	8	9	10	11	12	total	% difference
Keskmine ilmateenistus Sademete hulk (mm) 1981-2010	**50**	**35**	**37**	**31**	**42**	**69**	**72**	**83**	**64**	**74**	**63**	**53**	**672**	
pcp Kehra 2010	**25.7**	**60.7**	**61.6**	**32.9**	**66.3**	**60.7**	**109.3**	**49.5**	**94.7**	**66.8**	**113.5**	**88.9**	831	
pcp Kehra RADAR 2010	**2.6**	**21.4**	**37.7**	**26**	**34.9**	**20.6**	**26.1**	**31.5**	**33**	**39.9**	**63.4**	**36.6**	374	55.01
pcp Kehra 2011	**58.5**	**31.9**	**32.2**	**13.3**	**33.2**	**76.9**	**118.4**	**52.5**	**113.1**	**69.5**	**49.1**	**87.1**	736	
pcp Kehra RADAR 2011	**23.2**	**11.3**	**14.4**	**12.3**	**21.2**	**47.5**	**55.6**	**26.9**	**74.3**	**54.6**	**18.1**	**63.9**	423	42.46

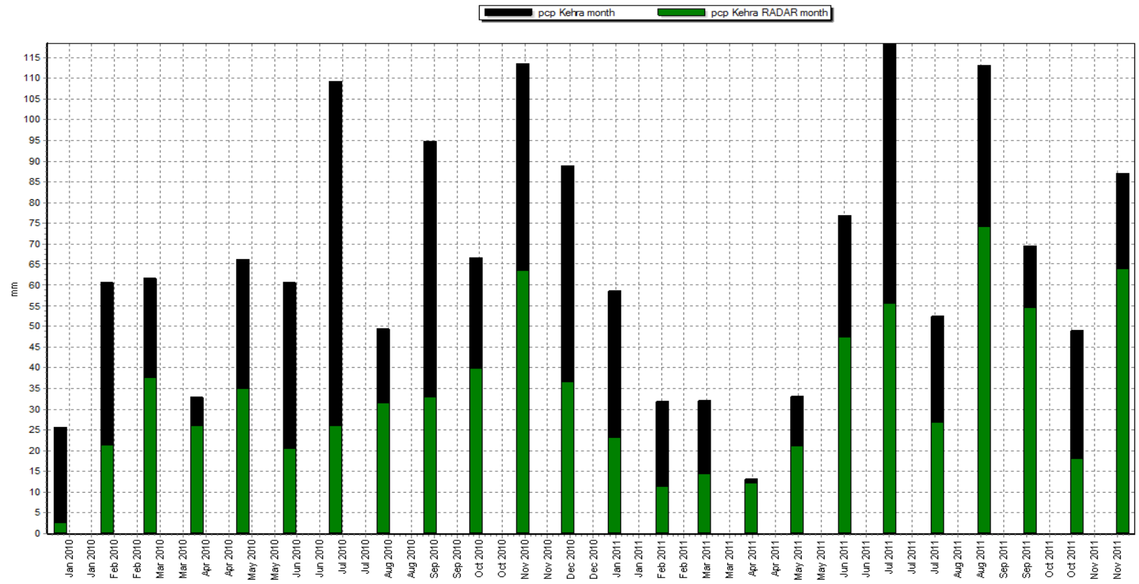


Figure 18: Station vs. radar precipitation data for the Kehra radar. Source: http://www.ilmateenistus.ee/kliima/kliimanormid/sademed/

Figure 19 presents the case of simulating water flow at the Pajupea permanent station using the SWAT model, and a combination of interpolated weather data (2001-2009) and radar-based precipitation (2010-2011) presented above. It is clearly visible how the simulated hydrograph signals low flow conditions, and is unable to return the patterns of measured flow when the water input by precipitation is significantly reduced.

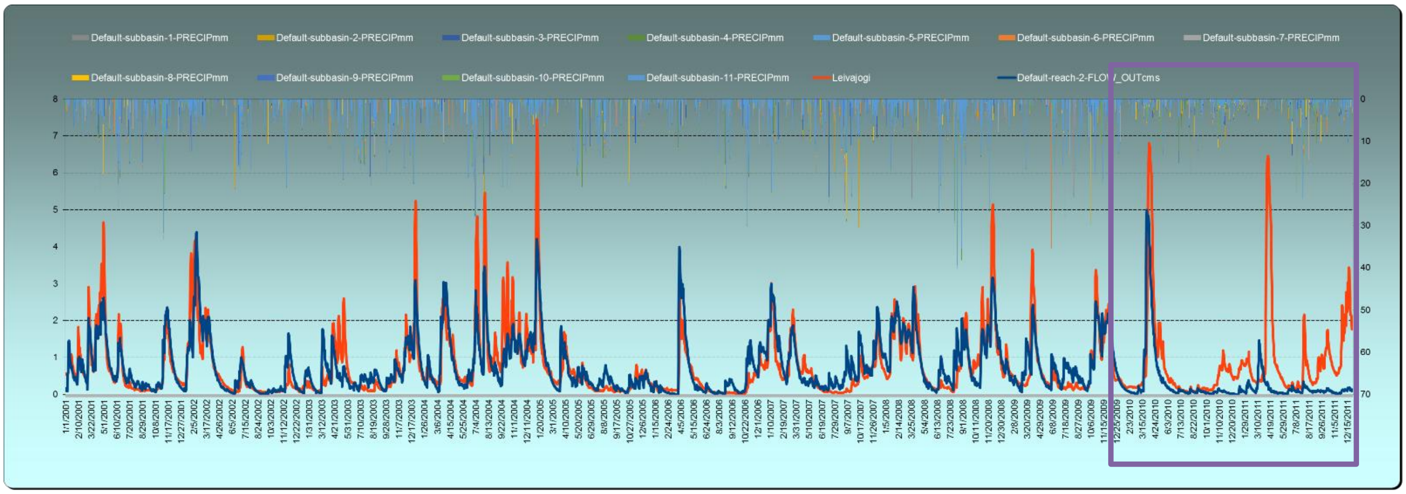


Figure 19: Simulation of water flow at the Pajupea permanent station using the SWAT model, and a combination of interpolated weather data (2001-2009) and radar-based precipitation (2010-2011). Red is observed flow, blue is modelled.

Discussion, assessment

In interpretation of the results, we first have to remind the reader that our model evaluation statistics were obtained on simulation catchment hydrology on a fine temporal resolution – i.e. on daily basis – which is much more difficult to fine-tune than doing the same on monthly data which the model-evaluation metrics by Moriasi et al., (2007) were meant for. Daily data reflect hydrological processes (fluctuations) that are much more difficult to capture because those are a product of short-response events (often extreme events) for which data may not be readily available, and for which the models may not contain sufficient (or any) processes to handle. This means that given the obtained results, we have to be optimistic about the catchment hydrology modeling part of our project. We therefore grade our model results to be successful and suitable for further work, i.e. sediment and nutrient transport simulation, as well as further advisory on the given catchments' hydrology.

Modeling catchment hydrology on a daily basis is challenging but it is necessary to keep refining the temporal scale of simulations, since a number of influential phenomena – and driving forces – occur at a very fine temporal scale. Intensive storms, for example occur at the minute or hourly scale. For many areas, the frequency and intensity of such storms is projected to increase in the future, and thus understanding and proper modeling of their impact is of paramount importance. It is for this reason that we elected to use the finest possible temporal scale to be simulated in this study, even though the data support to that was often not optimal. The applied simulation models presented a compromise. Given the complexity of parameterizing processes that take place at different scales, it is typical that detailed models do not simulate processes at larger scales, and vice versa. The chosen catchment models were directly relevant for the areas of application in this study. The point models require further interpretation and aggregation to be made relevant at larger scales. Apart from lack of detailed data, latter is a challenge that users face when attempting to apply event-based numerical models. Such models are able to emulate processes of fine temporal scale, but are typically limited in spatial terms – at least now in the 2010s. It will be desirable in the future to advance those models, or to couple them with catchment scale models, in order to be able to account for events at fine temporal scales, such as intensive/extreme rainfall events. Till then, daily time-step should be considered to be the best-available, but also the most desirable solution to use.

The availability of meteorological data was a source of uncertainty in our project. Meteorological stations from which data were available were scarce, and they fell outside the catchments. This meant that we had to rely on interpolations. There are multiple ways to interpolate such data, each carrying its inherent uncertainty. However, neither of them carry the capability to represent sub-grid spatial variability that is not representable by a distance-based gradient, especially if the landscape between the available stations is also heterogeneous. The most affected important weather parameter is (the amount and temporal distribution of) precipitation, which often varies along short distances. This will likely make the modeling study miss existing – potentially intensive – local rainfall events, and cause unexplained peaks in the observed flow.

In order to make an attempt to overcome this problem in the future, this project ran an experiment, by using an alternate source for precipitation. Radar stations can provide information on spatially distributed rainfall, which can then be matched with the individual modeling units of the modeled catchments. The effort appeared to be feasible. However, in our study, unfortunately the mismatch between total indicated rainfall by the two data sources was large enough, and was systematic, to yield an impossible water balance and to overcome the purpose of the study. We believe that the Doppler radar will eventually be a valuable data source for the likes of study that this project also conducted. The technique, however, is an indirect technique, which will first need to be fine tuned (calibrated?) to indicate realistic precipitation estimates. Unless and until the alternative, novel source is able to represent the ground-measured data with good statistical certainty, simulations of water flow will be strongly biased and will not represent true conditions. An overarching statistical evaluation of such data should precede their practical use for such modeling purposes.

As briefly mentioned earlier, model instability affected our work. Instability in this context refers to a numerical instability, where the internal calculations in some model process do not converge to a solution, which then halts the model, or yields unreasonable results. It is often obvious that the model output exists but it is unreasonable, while in other cases it requires expert judgement to understand and conclude that the output is not in the expectable range, or water-balance components do not add-up. In either case, the model is not ready to assist in drawing conclusions. Our study was somewhat affected by this, and the reasons could not be eliminated by reasonable effort. When we could not avoid such instability, we did/do not communicate the modeling results – as noted earlier.

Barring above mentioned instability/convergence problems, the listed simulation models were mostly calibrated manually by the respective experts by systematically changing calibration parameters in a trial-and-error scheme. The SWAT model, however, has also been calibrated using SWAT-CUP, the auto-calibration tool provided with the model. Modelers in this project experienced that the auto-calibration tool may not provide better calibration statistics than manual calibration did – which was the case for the Leivajogi catchment, for example. Auto-calibration may take away some burden from the modeler, but may also need expert supervision to provide expectably good calibration results. We think that this is due to the nearly impossible task to automatically fit a model to rather specific situations – especially if input data may be scarce or difficult to trust. The calibration tool may yield a set of calibrated parameters that present a 'local minimum' on the model error domain, while expert knowledge may help modify some parameters to achieve yet better model fit metrics. We advise the modeler to critically evaluate any auto-calibrated parameter sets for reasonability.

The areas of application presented some very specific soil conditions – in certain areas – that were borderline or outside the scope of the applied models, or the data sources available to provide estimates of soil properties. Peat soils present very specific and rather extreme properties. These 'soils' may be formed from a very small percentage of mineral material (clay, silt, sand) and a high to extremely high (up to 70-80%) amount of organic material, with very high porosities, and un-soil-like soil hydraulic properties that are difficult to characterize and represent. Their properties are also rather heterogeneous within the same group of organic soils. Correspondingly, internationally available data sources lack proper data to provide estimates of their hydraulic properties with good certainty, and model developers also handled the representation and internal parameterization for such soils with low priority. In this study this became evident when applying the 1D point models and their parameterization, and when attempting to parameterize the other models for areas represented by such soils. This raises concerns about the applicability of the applied models – or in fact all models – for such soil conditions. This problem is not country specific, however, but also affects Estonia and our case studies, and calls for more data collection on organic soils.

We have simulated catchment hydrology using the different simulation models separately, while we have introduced 'ensemble modeling' and its advantages earlier. To re-capture, ensemble modeling means using individual models that can perform the same task, and later relying on their (weighted) average as the outcome of modeling. This is commonly done today in e.g. weather forecasting. In this study we were short of resources to develop and perform true ensemble modeling, but we established the knowledge base for it in that we evaluated the data needs and availability, catalogued the models' capabilities and limitations – as well as the modeling expertise within this workgroup. We recommend, and prepare to be able to implement, ensemble modeling as one of the first steps in any follow-up to this project.

Modeling nutrient transport

Model calibration and validation results

The evaluation statistics of both model calibration and validation for sediment and nutrient transport are provided in *Tables 13a-c*. Figures of model-wise calibration/validation results are provided in Annex for the interested reader. The results for modeling sediment and nutrient transport were substantially less successful than the simulation of catchment hydrology – with some promising exceptions. The success of simulations was more model dependent in this case than in case of modeling catchment hydrology. There are, however, frequently occurring low values of the Nash-Sutcliffe efficiency metric, signaling general problems with the simulations. Similarly to hydrology modeling, we have encountered a number of challenges in both calibrating and validating the models. We highlight a few specific cases that will help present some concerns and findings of the hydrology modeling task.

		NITRATE		SUSPENDED SEDIMENT		TP	
		SWAT	INCA_N	SWAT	INCA_P	SWAT	INCA_P
		CALIB	CALIB	CALIB	CALIB	CALIB	CALIB
Keila	R^2	0.22	0.72	0.22	0.59		0.61
	N-S	-4.50	0.54	0.10	0.25		0.13
Vihterpalu	R^2	0.21	0.61	0.23	0.33		0.11
	N-S	-0.64	0.47	-2.54	-0.09		*-1.48/0.5*

Table 13a. Results and statistics with nutrients – catchment based models

		SOILN		*SOILN for MACRO*	
		Calib	*Valid*	*Calib*	*Valid*
Leivajõgi	*N-S*	*0.41*	*0.32*	*0.43*	*0.32*
	PBIAS	*-2.18*	*-3.12*	*15*	*17*
Vihterpalu	*N-S*	*-0.12*	*-0.2*		
	PBIAS	*-0.8*	*-18.2*		
Keila	*N-S*	*-0.24*	*-0.71*		
	PBIAS	*8.92*	*7.18*		

Table 13b. Results and statistics for nitrate concentration – profile based models

		NITRATE				SS		TP	
		SWAT	INCA_N	SOILN	SOILN for Macro	SWAT	INCA_P	SWAT	INCA_P
		CALIB	CALIB	CALIB	CALIB	CALIB	CALIB	CALIB	CALIB
Keila	R^2	0.22	0.72			0.22	0.59		0.61
	N-S	-4.50	0.54	-0.24		0.10	0.25		0.13
Vihterpalu	R^2	0.21	0.61			0.23	0.33		0.11
	N-S	-0.64	0.47	-0.12		-2.54	-0.09		*-1.48/0.5*
Leiva	R^2								
	N-S	-0.64	-0.19	0.41	0.43		-0.23		0.23

Table 13c. Overview of calibration results for the different models used in the Project

The example shown in *Figure 20* presents the modeling of nitrate N concentrations in the Leivajogi catchment as a result of leaching and runoff, modeled using the MACRO model. Despite the less-than-perfect model efficiency metrics, the model does manage to indicate the generally observed concentration, although with much variation in the success of catching the amplitude of peaks. Peaks in N concentrations are substantially influenced by release from point sources and the application of manure and fertilizers (and the weather pattern that follows), which was an area of significant uncertainty in this study. Information on point sources, such as input from waste-water treatment plants was not available. This made the modelers to add some point source input using estimated timing and amounts to attempt to explain previously unexplained peaks (e.g. as in *Figure 20*). Another potential source of N input is fertilization and manure application, of which we also did not have sufficient, locally specific information.

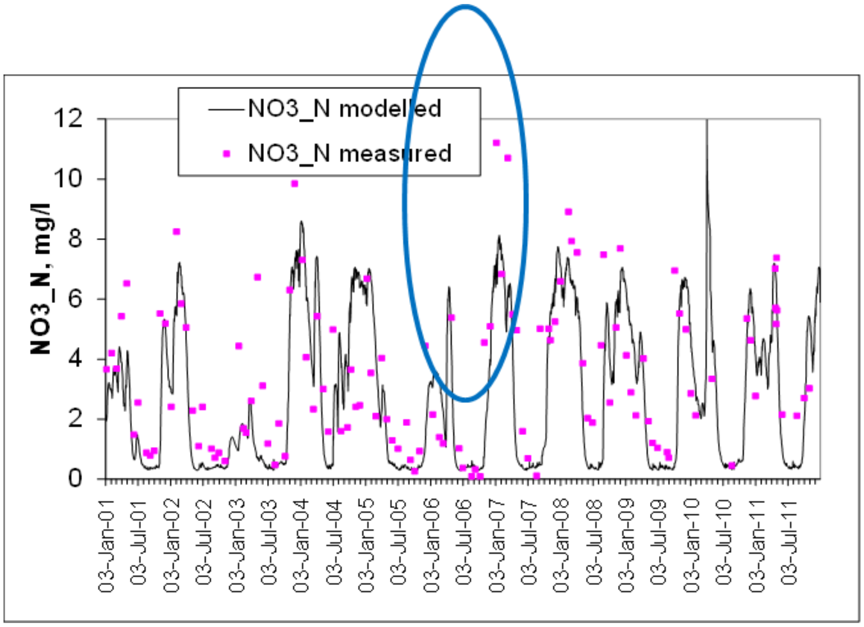


Figure 20: Simulation of nitrate N concentrations in the Leivajogi catchment for the entire 2001-2011 period using the MACRO model.

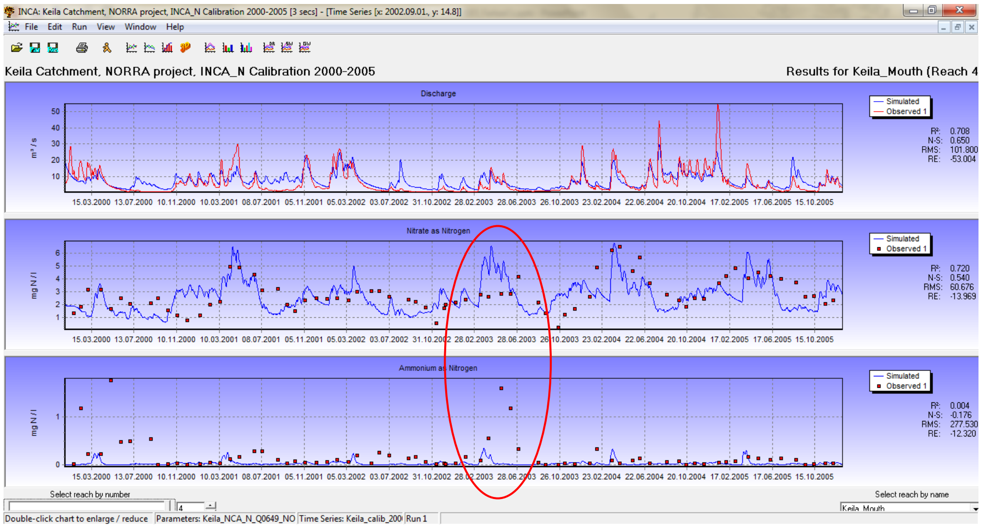


Figure 21: Simulation of ammonia and nitrate N concentrations in the Keila catchment (calibration period) using the INCA_N model.

Some simulation results also gave indications that there are uncertainties around the forms of nitrogen in the river, which roots back to uncertainties regarding the source of N. Nitrates and ammonia in rivers typically dominantly originate from different sources, i.e. nitrates from fertilization, and ammonia from animal waste as well as waste water treatment plants. *Figure 21* shows an example where it appears that the total amount of nitrogen was probably better simulated (perceived as the sum of the lower two panels) than the individual components, i.e. nitrate (middle-panel) and ammonia (lower-panel) separately. This is further indication of the importance of having sufficient reference data available for model calibration and validation.

Figures 22 and 23 demonstrate the INCA-P and the SWAT results with respect to suspended sediment concentration.

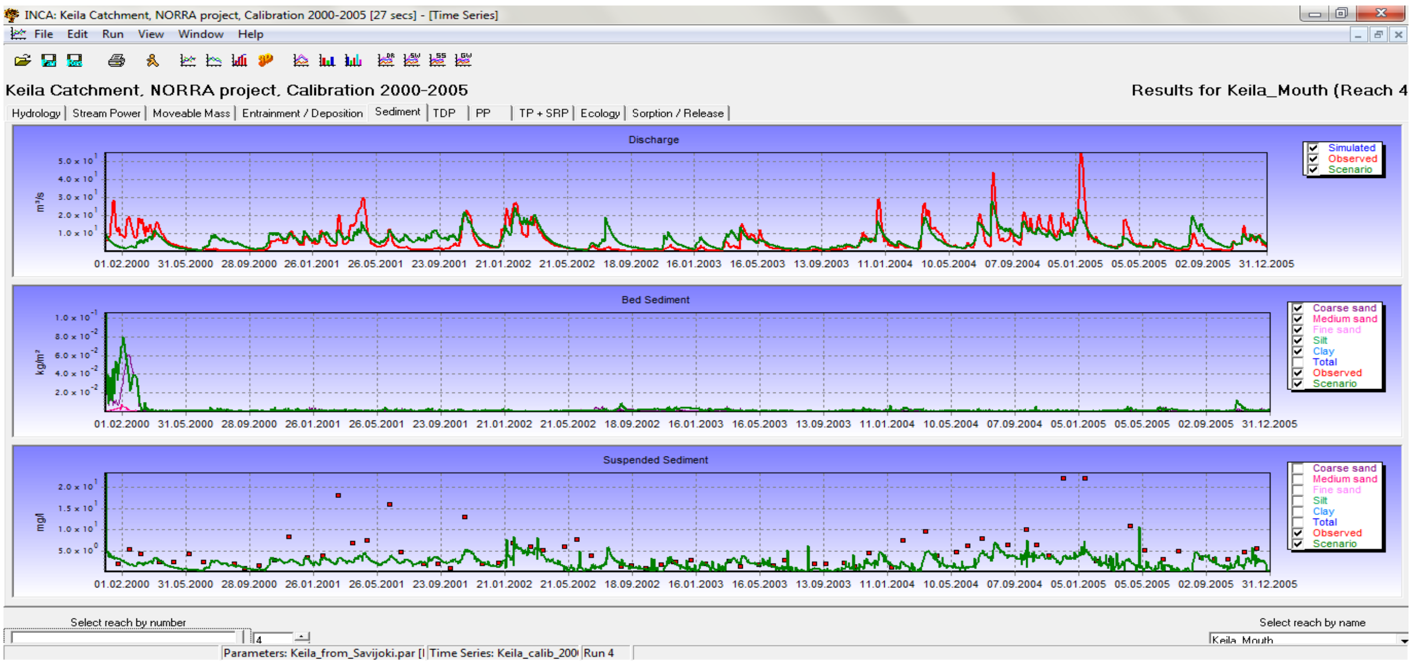


Figure 22: Simulation of suspended sediment concentrations in the Keila catchment using the INCA_P model.

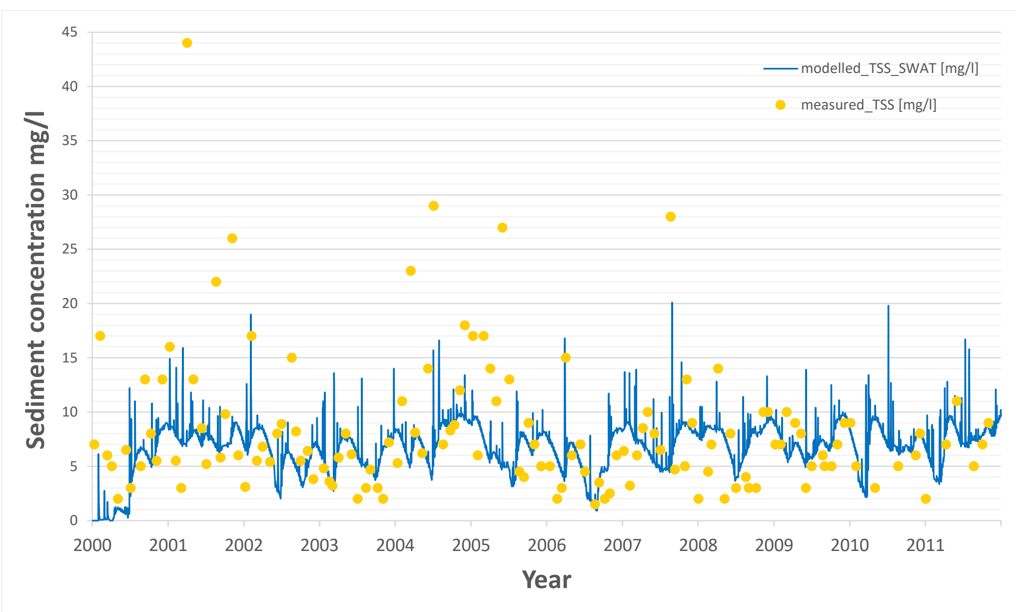


Figure 23: Simulation of suspended sediment concentrations in the Leivajögi catchment using the SWAT model.

Both Figures 22 and 23 indicate, that the model's can describe, in general, the sediment concentrations in the river, but do not capture the extremely high concentration values. The reason for that, most probably is, that these models are run on a daily time step, while extremely high soil losses are mostly generated by short-term extreme precipitation events. Thus, event-based models, operating on an hourly or even finer time step are needed to describe the physics, standing behind flash flood and extreme erosion.

Discussion, assessment

Simulating the hydrological and nutrient balance of catchments with less anthropogenic influence with more success seems logical: the simulation of anthropogenic point-type effects is difficult, especially when the available data is scarce. In our cases, two of the most significant such sources were point-source polluters and agricultural fertilization – both of which are a lot more significant when simulating nutrient concentrations and loads in surface waters. Point source polluters comprise mostly of waste-water treatment plants that can release large amounts of nutrients downstream. For this project, no or very scarce information was available about point source polluters, and thus the model calibration/validation could not account for their contribution to the nutrient loads, hampering model performance. Lacking such data, it could only be speculated by the modelers when and how much effluent may have been released into the river upstream.

Various land-uses typically yield different amounts of sediment and nutrient losses when exposed to the same amount and intensity of rain. Models that can differentiate between land-uses – typically all catchment models - require input on land-use specific losses. Lacking such data, in this project, this information was assumed by expert judgement. The only available information source was national statistics on fertilizer application, which was consulted prior to assuming certain levels and timing of fertilizer application in the models.

It needs to be noted that apart from the uncertainties that arise from unknown, scarce or uncertain source information on nutrient loads, the effect of modeling errors in hydrology modeling also propagates through to nutrient and sediment modeling. Even if total loads are known, a mis-representation of river flow will result in mis-judged concentration values, and vice versa. Even if the catchment hydrology is modeled perfectly, unknowns in nutrient applications and loads will make their modeling carry great uncertainty. And we finally note, without much discussion, that model instabilities did affect our nutrient modeling as much as it did the hydrological modeling, and it was unfortunately beyond the means of this project to be able to mitigate some of those cases.

General discussion and outlook

Numerical/dynamic simulation models are continually being improved by their developers – and via feedback by independent users – and hence they are growingly seen as tool of choice when it comes to quantifying environmental phenomena, or predicting the effect of any future scenarios. However, the user needs to understand and accept that the results of numerical models – just as the models themselves - are bound by limitations.

Models work well to the extent and quality as the included processes and the underlying data allow them to. Models are developed via concept-understanding as well as empirical observations. They are also tested against other, independent observations. Model performance is weakened if those observations are spatially or temporally scarce, low quality or entirely missing. In the context of this project, we were able to identify a number of such problematic areas. For example, data on point-source type pollution (e.g. waste-water treatment) were not, or rather scarcely available (only annual or quarterly data). There was no data available to the project on the diffuse (aerial) use of fertilizers – as major N and P sources – in agricultural areas. Neither was there sufficient information on nutrient retention times in surface waters – both streams and lakes. When such important data are lacking, the modeler can only rely on expert opinion, or generic estimates based on other locations, which always increases the degree of uncertainty to the results. At the same time, original data also present an unknown – sometimes sizeable – degree of uncertainty as well as error, which should be reasonably accounted for.

Models themselves also present imperfections and limitations. Models are often experimented with in a context that is beyond its original means, and thus the user may find that certain known processes are not included in a model. Moreover, different developers may have applied concepts of different types and complexities in their respective models. In the first case, the user will either have to ignore the process entirely, or find an expert-based solution (often a tweak in the data) to mimic the process or phenomenon. In the latter case, advantages and disadvantages of certain mathematical solutions are often seen on a case-by-case basis, and their performance usually interacts with the availability and quality of input data. Such is the case, for instance with the soil hydraulic properties or the calculation of evapotranspiration. These often pose known and quantifiable, or unknown degrees of uncertainties.

Calibration of models is a sensitive part of any simulation-based study. Models can, and should, be recalibrated when any new knowledge or data becomes available. Hence, the necessity to re-calibrate a model to perform the same task may re-appear for future studies on the same areas. Calibration tools and techniques may also evolve with time, so the complexity and burden of this task may change in the future.

Overall, we advise the user of our results to consider its limitations resulting from imperfections to the available models, and in the context of the scarcity – or lack of – desirable data.

Simulation results are often taken by their face value, whereas there can be large uncertainties to the estimated values. Such uncertainties can originate from many sources – several were listed previously – and their importance is often not communicated. We want to emphasize awareness of the uncertainty problem, and how to handle it. Studies are often not funded to the degree that researchers are able to support their studies with uncertainty estimates, although attempts are being made continually. We recommend that (a) research results are interpreted with uncertainties in mind; (b) trends and patterns, rather than individual values are focused at, and (c) that future studies are initiated and funded such that uncertainty analysis can become an explicit part of the study.

When studies at larger scales are performed, data availability concerns often derive the necessity to upscale or extrapolate information – input or output – to larger areas. This has also affected our study. It requires great awareness and care towards deciding to what extent certain information can be extrapolated, and any decision on extrapolations should be made by joint discussions and understanding. Some types of data/information can present greater variability over short distances than others (e.g. precipitation over air temperature). It also has to be ensured that the information to be extrapolated does not introduce noticeable bias to the representation of a larger area.

In order to be able to link the proper causes and effects, one also need to consider that every step of upscaling usually results in loss of capability to establish such cause-effect connections. For example, while being able to simulate the annual loading of N and P to the Baltic Sea is of major significance, a national scale study of that kind usually lacks the level of detail that would help establish whether e.g. waste water treatment plants as point sources, or agriculture as a diffuse polluter – and within that area, what type of activity – is the main source, and is the most beneficial environmentally and economically to mitigate. This is a major source of uncertainty for the decision-maker, since it becomes very difficult to judge effectively if a new policy can be expected to yield the expected benefits. It would be very important to collect data on sources and the apportionment of pollution in order to help effectively mitigate the arising pollution.

When new information-collection campaigns are designed and initiated, it may yield great dividends if environmental modelers are invited to discuss their information needs for effective work. They may be able to recommend ways of 'what' information and 'how' it should be collected that may provide very good cost-benefit ratio towards eliminating data gaps and certain sources of uncertainty in future studies. Long-term monitoring stations, when already established, could collect much useful information at low added cost, for example, but it is equally important to try to eliminate gaps in the collected monitoring data. The proposed monitoring strategy will have a regional perspective, in addition to the specific national needs. We need to emphasize the "sustainability" of any monitoring strategy, i.e. that the proposed activities shall not be extravagant with regard to future capacities of the individual countries. One way to enhance the sustainability is to design monitoring programs that are both suitable and attractive for research and educational purposes. For these reasons, it is also important that the applied measurement methods and procedures are sufficiently advanced to comply with international scientific standards.

The main limitation to the work of this working group was the lack of data as inputs to the simulation models, and often the spatial representativeness of data that were available. This reports cites a number of such issues, and points at a number of solutions that were used to try to overcome such lack of data. It is the expressed opinion of this working group that while existing simulation models are not perfect, it is rather the data quality and quantity available for use that was limiting this work, and hence improving that situation would be the first course of action if future improvements are expected in modeling the hydrology and nutrient losses from catchments. This is true for both input information (e.g. fertilizer use) as well as simulated variables (e.g. effluent amounts and concentrations).

At the same time, the modeler needs to appropriately judge the value in using a more data demanding, but otherwise better model when the necessary data are unavailable or rather uncertain. Some models, given their internal set of equations that drive the simulation of processes, may not be applicable across the range of conditions that needs to be simulated. This project ran into complications in simulating the hydrological balance of peat soils; a specific set of soils with extremely high organic matter content and very low bulk density. In some cases, overcoming such situations takes specific expert knowledge to alter the model's input while continually monitor the output to remain meaningful and reasonable.

We envisage these dynamic models to be suitable and appropriate to use to analyze and predict the impact of predicted climatic changes, as well as changes in land-use on agricultural hydrology and nutrient and pollutant transport and losses. Dynamic models are the best-available tools, and combined with the already accumulated experience, they will be suitable to inform the policy maker as well as the public about the direction and expected magnitude of changes. More than one of the applied process-based models are suitable to be used at a larger – national – scale, and they can even be combined in a model ensemble. The benefits of ensemble modeling have been discussed earlier in this report, but its implementation was beyond the means of this project, and remains a recommended direction of future work. This being said, we repeat that it is necessary to expand the network of environmental data collection to areas other than the already known research-watersheds, and to ensure that existing monitoring stations stay on-line as well. Data quality and availability are both crucial to be able to inform and validate our models, and hence form the basis for any modeling study.

Looking forward, we have gained knowledge on capabilities and limitations of the various models that we used in this study. This has special significance towards planning to use these models to simulate mitigation measures. Some of the models cannot simulate the effect of certain mitigation measures, or if they can, that is only possible via an expert-based alteration of inputs, i.e. such application is not self-evident. The simulation of effects of mitigation measures also assumes that proper data are available about the current situation. For example, this project was not able to secure wide-spread information on fertilizer application timings and rates, or crop yield data, that constitute crucial reference information for scenario modeling.

In terms of advice as to what model to recommend for future use, there is an inherent tradeoff between models within and much beyond this project or its constituting tasks. Some models are rather complex and require a vast amount of input information – e.g. the SWAT model in our study. The INCA model family requires less input, but its capabilities are more limited e.g. in terms of simulating the expected impact of implemented or future mitigation measures. Yet another step in the same direction would be to elect to use simple statistical models – that require even fewer input - in place of process based models. Such statistical models, however, most often ignore processes and phenomena that all specialized scientists would agree are important, even if their impact may not be evident from existing (poor?) data collections.

When selecting model(s), the goals of a modeling study and the availability of support and input information need to be harmonized, which is partly accounted for in the model selection process that we recommended earlier. However, we need to re-emphasize that the lack of existing data and its poor spatial representativeness should not be an excuse to eliminate complex models, but should rather be a warning that our data collection is behind the capabilities of our ever-developing predictive tools, and hence field monitoring and the collection of other relevant data should continue and expand.

References

Abbaspour, K.C., Johnson, A. and M.Th. van Genuchten. 2004. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. Vadose Zone Journal 3(4), 1340-1352.

Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. and R. Srinivasan. 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. Journal of Hydrology, 333:413-430.

Arheimer, B., Lindström, G., Pers, C., Rosberg, J. and J. Strömqvist. 2008. Development and test of a new Swedish water quality model for small-scale and large-scale applications. XXV Nordic Hydrological Conference, Reykjavik, August 11-13, 2008. NHP Report No. 50, pp. 483-492.

Baker, L. and D. Ellison. 2008. The wisdom of crowds — ensembles and modules in environmental modelling. Geoderma, 147, 1-7.

Bechmann M. and Deelstra J. 2013. Agriculture and Environment – Long-term monitoring in Norway. Akademika forlag, 392 p.

Beven, K. and P.F. Germann. 1982. Macropores and water flow in soils. Water Resour. Res., 18, 1311–1325.

Deelstra J., Farkas Cs., Engebretsen A., Kværnø S., Beldring S., Olszewska, A. and Nesheim L. 2010. Can we simulate runoff from agriculture-dominated watershed? Comparison of the DrainMod, SWAT, HBV, COUP and INCA models applied for the Skuterud catchment. In: Grzybek A. (ed.), "Modelling of biomass utilisation for energy purpose". Bioforsk FOKUS 5(6): 119-128.

Deelstraa J., Ital A., Povilaitis A., Kyllmard K., Greipsland I., Blicher-Mathiesen G., Jansons V., Koskiaho J., Lagzdins A., 2014. Hydrological pathways and nitrogen runoff in agricultural dominated catchments in Nordic and Baltic countries. Agriculture, Ecosystems & Environment, Vol. 195:211-219.

Farkas Cs, and Hagyó A. 2010. Applicability of profile and catchment scale simulation models for estimating the effects of climate change on water regime. „Klíma-21" Füzetek, 62:59-74. (in Hungarian)

Farkas Cs., Hernádi H., Makó A. and Máté F. 2011. Estimating climate change effects on soil water balance elements of Hungarian Calcic Chernozem soils. Agrokémia és Talajtan, Volume 60:4-56.

Farkas, C, Beldring, S, Bechmann, M and J. Deelstra. 2013. Soil erosion and phosphorus losses under variable land use as simulated by the INCA-P model. Soil Use and Management, 29(1), 124–137. doi: 10.1111/j.1475-2743.2012.00430.x

Guber, A.K., Pachepsky, Ya.A., van Genuchten, M.Th., Simunek, J., Jacques, D., Nemes, A., Nicholson, T.J. and R.E. Cady. 2008. Multimodel simulation of water flow in a field soil using pedotransfer functions. Vadose Zone Journal 8(1), 1-10.

Gupta, H.V., Sorooshian, S. and P.O. Yapo. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. J. Hydrologic Eng., 4(2), 135-143.

Iital, A. 2005. Monitoring of Surface Water Quality in Small Agricultural Watersheds: Methodology and optimation of monitoring network. Doctoral Thesis. Tallinn University of Technology Faculty of Civil Engineering, Department of Environmental Engineering.

Jansson, P.-E. 1991. SOIL water and heat model, technical description. Internal paper from Soil Science Department, Swed. Univ. of Agricultural Sciences, pp. 46.

Johnsson, H., Bergström, L., Jansson, P.-E. and K. Paustian. 1987. Simulated nitrogen dynamics and losses in a layered agricultural soil. Agric. Ecosyst. Environ., 18, 333–356.

Kalinin, G.P. and P.I. Miljukov. 1957. On the computation of unsteady flow in open channels. Met. Gidrol. 10, 10–18. (in Russian).

Kværnø S.H., Farkas Cs., Stenrød M., Eklo O.M., Nemes A., Stolte J., Deelstra J., Engebretsen A. 2013. Model simulations for scenario analyses and risk assessment at the catchment scale. In: Bechmann, M. and Deelstra, J. (eds.): Agriculture and Environment – Long-term monitoring in Norway Akademika Publishing, Trondheim, ISBN: 978-82-321-0014-9.

Larsson, M.H. and N.J. Jarvis. 1999. A dual-porosity model to quantify macropore flow effects on nitrate leaching. J. Environ. Qual., 28, 1298–1307.

Lindström, G., Pers, C.P., Rosberg, R., Strömqvist, J. and Arheimer, B. 2010. Development and test of the HYPE (Hydrological Predictions for the Environment) model – A water quality model for different spatial scales. Hydrology Research 41.3-4:295-319.

Litaor, M.I., Eshel, G.,Sade, R., Rimmer, A. and M. Shenker. 2008. Hydrogeological characterization of an altered wetland. J. Hydrol., 349, 333-349.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and T.L. Veith. 2007. Model Evaluation Guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE, Vol 50(3), 885-900.

Nash, J.E. and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models: Part 1. A discussion of principles. J. Hydrology 10(3): 282-290.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R. and J.R. Williams. 2009. Soil and Water Assessment Tool. Theoretical Documentation. Version 2009. Grassland, Soil and Water Research Laboratory. Agricultural Research Service.

Nemes, A., Rawls, W.J. and Ya.A. Pachepsky. 2006. Use of a non-parametric nearest-neighbor technique to estimate soil water retention. Soil Sci. Soc. Am. J. 70(2): 327-336. DOI: 10.2136/sssaj2005.0128.

Nemes, A., Timlin, D.J. and B. Quebedeaux. 2010. Ensemble approach to provide uncertainty estimates of soil bulk density in support of simulation-based environmental risk assessment studies. Soil Sci. Soc. Am. J. 74(6):1938-1945. doi:10.2136/sssaj2009.0370.

Saloranta T.M., Kamari J., Rekkolainen S. and Malve O. 2003. Benchmark Criteria: A Tool for Selecting Appropriate Models in the Field of Water Management. Environmental Management. 23(3):322-333.

Schaap, M.G., Nemes, A. and M.Th. van Genuchten. 2004. Comparison of Models for Indirect Estimation of Water Retention and Available Water in Surface Soils, Vadose Zone J., 3(4), 1455–1463. doi:10.2136/vzj2004.1455.

Singh, V.P. 1988. Hydrologic Systems, vol. 1. Rainfall-Runoff Modeling. Prentice Hall, Englewood Cliffs, New Jersey, USA.

Stewart, B. A., Woolhiser, D.A., Caro, J.H. and M.H. Frere. 1975. Estimating Potential Erosion. In: Control of Water Pollution From Cropland. Vol 1, 7-25. U. S. Environmental Protection Agency Report No EPA-600/2-75-026a or USDA Report No ARS-H-5-1. Washington, DC.

Vassiljev, A., Grimvall, A. and M. Larsson. 2004. A dual-porosity model for nitrogen leaching from a watershed. Hydrological Sciences Journal, 49(2).

Van Waveren R.H., Groot S., Scholten H., van Geer F.C., Wösten J.H.M., Koeze R.D. and Noort J.J. 1999. Good modelling practice handbook. STOWA Report 1999-05. Available at: http://harmoniqua.wur.nl/public/Reports/Existing%20Guidelines/GMP111.pdf

Wilby, R.L. 1997. Contemporary hydrology. John and Sons, England.