skb.se

SKB P-23-04

ISSN 1651-4416
ID 2002186
April 2023

WIM Forsmark 1.1

A GIS-based machine-learning tool for predicting wetland extent

Jean-Marc Mayotte
Svensk Kärnbränslehantering AB

Mårten Strömgren
Umeå Universitet

Keywords: wetlands, landscape development, machine-learning, GIS

Data in SKB's database can be changed for different reasons. Minor changes in SKB's database will not necessarily result in a revised report. Data revisions may also be presented as supplements, available at www.skb.se.

This report is published on www.skb.se

Abstract

This report describes the methodology used to develop a machine-learning tool that can be used to predict the extent of wetlands at Forsmark. The tool uses the “wetland identification modelling” toolbox in ArcGIS® Pro. 10 individual wetland areas in Sweden are used to train 10 individual machine-learning algorithms which are used to predict wetland locations and extents at Forsmark. Wetland predictions for Forsmark are given in raster format. The suite of algorithms used in this study, as well as the specific methods used to train, evaluate and apply the algorithms when predicting wetland extents at Forsmark, is given the name “WIM Forsmark” and a version number of “1.1”. Preliminary results suggest that WIM Forsmark 1.1 is capable of predicting the extent of wetlands at Forsmark and produces results similar to an alternative, more complicated machine-learning tool produced by the Swedish University of Agricultural Sciences (SLU). It is therefore suggested that WIM Forsmark 1.1 could be used to predict the position and extent of future wetlands at Forsmark and should be considered in studies of landscape development. Furthermore, the study postulates that WIM Forsmark 1.1 may be able to give information surrounding the uncertainty of the wetland predictions.

Sammanfattning

Den här rapporten beskriver metodiken för att utveckla ett maskininlärningsverktyg som kan användas för att prediktera våtmarker i Forsmark. Verktyget använder verktygslådan ”wetland identification modelling” i ArcGIS® Pro. 10 olika våtmarksområden i Sverige används för att träna 10 olika maskininlärningsalgoritmer som används för att prediktera läge och utbredning av våtmarker i Forsmark. Våtmarksprediktion för Forsmark erhålls i rasterformat. Uppsättningen av algoritmerna som används i denna studie, liksom de specifika metoderna använda för att träna, evaluera och applicera algoritmerna vid prediktion av våtmarksutbredning i Forsmark har namngivits till ”WIM Forsmark” med versionsnummer ”1.1”. Preliminära resultat antyder att WIM Forsmark 1.1 är kapabel att prediktera läge och utbredning av våtmarker i Forsmark och producerar resultat som är jämförbara med ett alternativt, mer komplicerat maskininlärningsverktyg utvecklat av Sveriges Lantbruksuniversitet (SLU). Det föreslås därför att WIM Forsmark 1.1 skulle kunna användas för att prediktera läge och utbredning av framtida våtmarker i Forsmark och bör övervägas i övriga studier om landskapsutveckling. Studien postulerar dessutom att WIM Forsmark 1.1 kan ge information kring osäkerheten i våtmarksprediktionerna.

Content

1 Introduction

It is the responsibility of the Swedish Nuclear Fuel and Waste Management Company (SKB) to manage the management of both radioactive waste and spend nuclear fuel in Sweden. The existing Repository for Short-Lived Radioactive Waste (SFR) and the site for the Spent Fuel Repository is in Forsmark (Figure 2 1). An extension to the SFR facility is planned to extend the capabilities of the repository. As a part of the license applications for the extension of SFR and the construction of the final repository for spent fuel, SKB is continually assessing the long-term radiological safety each future repository.

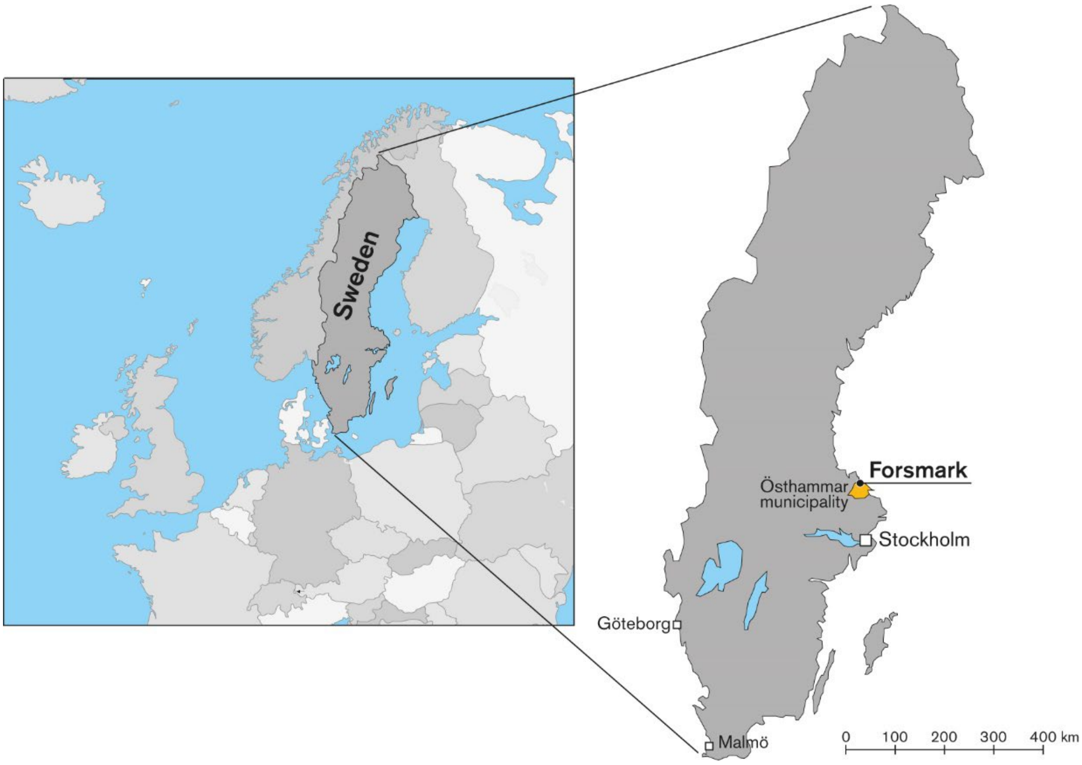

Figure 1-1. Location of the Forsmark site in Sweden (right) and in context with the countries in Europe (left). The site is situated in the Östhammar municipality, which belongs to the County of Uppsala.

Hydrological and hydrogeological modelling is an important part of the safety assessments for the existing and planned repositories at Forsmark. Comprehensive analyses of the future hydrology and future near-surface hydrogeology of the Forsmark area were performed for the license application for the SFR extension (Werner et al. 2013) and site-selection for the Spent Fuel Repository (Bosson et al. 2010). A key point of these modelling studies was that they were required to examine the hydrology as affected by a postulated future climate and landscape over very large time-spans (10^5 years for SFR and 10^6 years for the Spent Fuel Repository). The time-spans under consideration are so large that geological events such as large-scale bulk transports of sediment via erosion processes, shore-line retreat due to isostatic rebound and future ice ages must be accounted for. Geological events like these can cause significant changes to quaternary geology and topography of an area, and the nature of these changes are only generally understood at regional scales.

For the hydrological and near-surface hydrogeological investigations for Forsmark, models are parameterized using topographical and geological information with <10 – 80 m^2 of spatial resolution. There are significant uncertainties regarding the topographical and geological characteristics of a landscape in the very far future at these smaller spatial scales which can affect the predictive capacity of the hydrological models. A review of the safety assessment for the Spent Fuel Repository (SKB 2011) conducted by the Swedish Radiation Safety Authority (SSM) stated that the deterministic nature by which the future landscape at Forsmark was considered in the safety analyses was insufficient and that uncertainties in the future landscape need to be accounted for in the dose modelling (Section 7.1.2.2, SSM 2018). This implies that landscape development modelling needs to be conducted within an uncertainty framework. This further implies that several realizations of the future landscape may need to produced thus strengthening the case for using relatively simple models which require minimal computational and personnel resources.

The primary purpose of this study is focused on developing a model that is capable of predicting the extent of future wetlands at Forsmark while providing information on the likelihood of the predictions. The model should also be easy to set up and run using minimal personnel and computational resources in order to reduce the time required to produce wetland predictions in case several different realizations are needed. The model should also be relatively easy to explain in order to increase transparency in order to ease the burden on external reviewers.

This study proposes that machine learning is a viable alternative to process-based modelling of wetland development.

2 Methodology

2.1 Overview

This study uses the Wetland Identification Model (WIM) developed by O'Neil et al. (2018, 2019) to generate individual predictions of wetland extent at Forsmark. The WIM tool is available as a part of the Arc Hydro toolset in ArcGIS® Pro (ESRI 2021). The WIM tool uses a pre-processed[1] digital elevation model (DEM), raster data on the spatial location of surface water bodies (lakes and streams), and observed data on wetland extent to predict the extent of wetlands in areas missing observations. A DEM is used to calculate the topographic wetness index (TWI) (Beven and Kirby 1979), depth to water index (DTW) and the curvature of the landscape (Ågren et al. 2014).[2] The TWI, DTW and curvature are used as predictor variables in a "random trees" machine learning algorithm which is trained and tested using observed wetland data (Breiman 2001).

Once deemed satisfactory, the WIM algorithm can be used to predict the occurrence of wetlands in areas lacking data. Applying the algorithm to new DEMs is a relatively quick process that is not computationally intense. However, a bias in the predictions needs to be assumed given that the algorithm is trained using measured data from a location where environmental factors (defined as climate, topography and wetland development rate within the context of this study) may differ from those of the location where new predictions are being made. For this reason, it is essential that algorithms are only used to predict wetland extent in areas with similar environmental conditions.

It is hypothesized that WIM could be used to help predict wetland extent at Forsmark within the context of the safety analyses for both SFR and the Spent Fuel Repository. To accomplish this, any potential WIM algorithm would have to be trained using a landscape with environmental factors characteristic of the Forsmark landscape in the far future (i.e. $+10^3 – +10^6$ years).

Past and planned safety analyses assume substantial uncertainty surrounding the climate, topography and hydrological conditions of the future Forsmark landscape. It is therefore not reasonable to assume that a single landscape proxy can be chosen which could adequately represent the future landscape of Forsmark. This study proposes that a suite of WIM algorithms, trained using a variety of landscapes, could be used simultaneously when predicting wetland extent for the future Forsmark landscape. This would not only ensure that a large variation in climactic, topographic and wetland ages are considered in the algorithm training, but could also provide some insight into the uncertainty surrounding the predictions by considering prediction frequency across the suite of algorithms.

2.1.1 Uncertainty management using WIM

As previously mentioned in Section 2.1, climate, topography and wetland development rate are considered to be the primary factors which are capable of being accounted for using the WIM tool. While topography is considered directly by the WIM tool (via the DEM inputs), climate and wetland age are not implicitly considered in the training of the WIM algorithms. Instead, it is up to the user to train the algorithm using areas where both the climate and approximate wetland age are similar to that of the area where wetland extent is to be predicted. In the context of the safety analyses pursued for the SFR and Spent Fuel Repository, prediction of wetland extent needs to be performed for landscapes in the far future where the climate and conditions conducive for wetland development are highly uncertain.

As briefly mentioned in Section 1, it is essential that any tool that is used to predict the future state of the Forsmark landscape is capable of providing information on the uncertainty surrounding predictions. The current distribution of the WIM tool in ArcGIS provides Boolean outputs of wetland predictions; raster cells either do or do not contain a wetland. However, raw outputs from the random trees algorithms are reported as probabilities which the WIM tool in ArcGIS then converts to

[1] Preprocessing includes smoothing (removal of data "noise") and producing a hydroconditioned DEM (filling local sinks and generating a flow-direction raster). The DEM may also be used to generate a surface water raster (i.e. locations of lakes and streams).

[2] TheTWI, DTW and curvature are calculated at the same resolution as the input DEM and used within the WIM tool as raster inputs.

Boolean predictions.[3] This suggests that it is possible to instead use the probabilistic outputs from WIM, instead of the default Boolean outputs, to try and quantify the uncertainty of the prediction as quantified by the random tress algorithm.

While this step may help in quantifying the uncertainty surrounding wetland predictions for an area based on a single trained algorithm, the user must be confident that the wetland area used to train the algorithm is characteristically similar (i.e. climatologically and topographically) to the area where the predictions are being made. The tool in development will be used to predict wetland extent for the landscape at Forsmark in the far future which includes a substantial area of land which currently resides at the bottom of the Baltic Sea. This means that it is not realistic to assume that a user would be able to select any one single area that will be characteristically similar for this future landscape as the climactic conditions of this future landscape are highly uncertain.

In order to overcome this issue this study proposes that a suite of WIM algorithms can be considered simultaneously when predicting wetland extent. The variation of the climactic, topographic, and wetland development conditions across the selected landscapes used to train the individual algorithms is intended to represent the uncertainty surrounding these factors in the far future. By considering each output simultaneously (i.e. overlay raster outputs), the user could then ascertain some degree of the uncertainty in the predictions by examining the frequency of predictions for a given location (i.e. the number of simultaneous predictions for a single raster cell).

In this study, 10 different Swedish landscapes are used to train 10 different WIM algorithms. Each individual algorithm is then simultaneously applied to the area where wetland extent and location is to be predicted. Results are then presented in two different forms: one using the default "Boolean" outputs of the WIM tool in ArcGIS for the suite of algorithms, and results which attempt to present the uncertainty of the prediction by examining the probabilistic outputs from the suite of algorithms.

When considering the Boolean outputs of WIM, results are presented in the form of "hit maps" wherein each predicted raster cell has a value range of zero to 10: a value of zero indicates that none of the algorithms have predicted a wetland in that cell, and a value of 10 indicates that every algorithm has predicted a wetland in that cell (Figure 2-1). When considering the probabilistic outputs, a mean probability of wetland prediction, or "p-means", is calculated for each cell using the outputs of the 10 different WIM algorithms. Results are then presented as a raster where each cell containing a probability of prediction.

This study proposes that, by presenting results in this fashion, the user will be able to gather some insight into the uncertainty of the wetland predictions that may be applied to future analyses. The hit maps may be applied within a less "formal" uncertainty framework which does not require the user to motivate the use of probabilities which, as calculated via the WIM tool in ArcGIS, are largely black-box in nature and may not be meant to be further applied within a larger uncertainty framework. The user could instead use qualitative reasoning to include/exclude predictions (e.g. only consider cells with five or more hits) instead of purporting to have a rigorous, statistical understanding of the predictions. On the other hand, the p-means presentation of the results may be better suited for an application of the results as their use is relatively intuitive (e.g. only consider cells with a probability greater than 0.50) and would likely not require a substantial amount of exploratory text before results could be further explored. However, it is not entirely certain how the "p-means" method of presenting results would compare to the results of a single WIM algorithm (via an examination of the probabilistic outputs) that was trained using all 10 of the Swedish landscapes simultaneously. However, questions surrounding the use of the WIM results are considered outside the scope of this study. This study will instead present both types of results with the assumption that any further application of the tool discussed herein will assess the assumptions and caveats associated with the tool's results prior to the application of said results.

[3] For all cells with a probability of predicted wetland greater than 0.50, the cell is reported as containing a wetland. For all cells with a probability of predicted wetland less than 0.50, the cell is reported as containing no wetland.

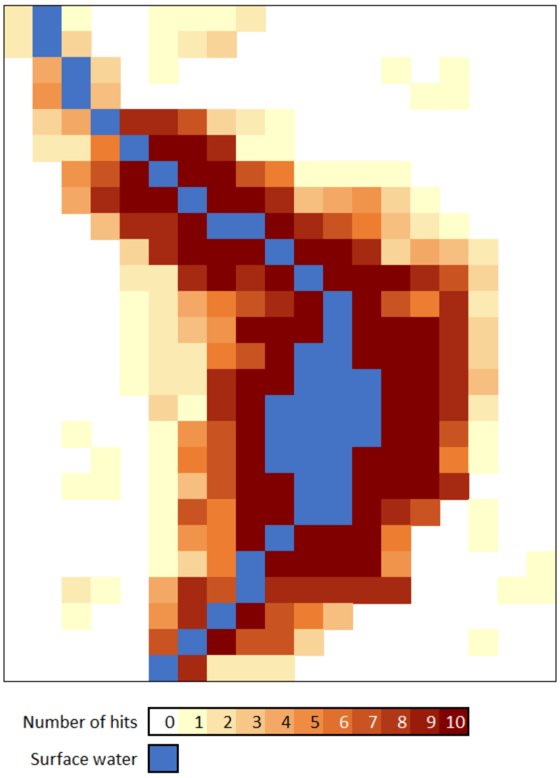

Figure 2-1*. Example of a "hit map" used to display results from 10 different WIM algorithms. Note: this figure is produced using synthesized data.*

2.1.2 Alternative tool: SLU's land "wetness" map

The Swedish University of Agricultural Sciences (SLU) has developed a tool based on the work of Lidberg et al. (2020) which is used to predict the "wetness" of a forested landscape in areas where observations may not be available. The tool uses machine learning and other publicly available data to generate maps of potentially "wet" soils (Lidberg et al. 2020). Outputs of the tool are given as raster data where the soil is classified as one of four different "wetness" classes ranging from "dry" to "wet"[4]. The "soil wetness maps" are publicly available for download via the Swedish Forestry Agency[5].

Both SLU's tool for generating the "wetness maps" and the WIM tool in ArcGIS® Pro use machine learning to aid in their predictions. Both tools require a DEM as input and both use the DTW and TWI (both calculated based on a hydro-conditioned DEM[6]) as predictor variables. However, SLU's tool uses additional predictor variables (generated using the DEM) and requires additional geographic inputs which WIM does not. SLU's tool requires five separate raster inputs and uses 24 predictor variables to train the algorithm (Ågren and Lidberg 2020) while WIM requires only one raster input and uses three predictor variables to train the algorithm (O'Neil et al. 2018)[7].

At the beginning of this study, it was decided that SLU's wetland prediction tool was not well suited for the intended application of the tool for two primary reasons: a) the SLU tool is not publicly available and any new estimates of wetland extent, especially for areas that do not currently exist,

[4] The four "wetness classes" (*fuktighetsklasser* in Swedish) are (*SV*/EN): *torr*/dry, *frisk-fuktig/*mesic-moist", *fuktig-blöt*/moist-wet, and *blöt*/wet. The wet class is used for mapping of surface waterbodies.
[5] SLU's" soil wetness map" (*Markfuktighetskarta* in Swedish) for Sweden - https://www.skogsstyrelsen.se/sjalvservice/karttjanster/geodatatjanster/rest/
[6] A hydo-conitioning is a process where minor adjustments are performed on the DEM in order to ensure that routing of runoff is calculated correctly. This involves "smoothing" of the DEM in order to remove local outliers in the elevation data and filling of hydrological sinks.
[7] It should be noted that, while not required by either tool, additional input data showing the observed locations of surface water bodies (i.e. lakes and streams) can be used to modify/verify the processed data layers which are further used to calculate the TWI.

would limit SKB's ability to quickly adapt the tool to the needs of the safety analyses. And b) because the SLU tool requires more input data and uses more predictor variables, any analyses of prediction uncertainty would become more complicated due to the larger amount of input data needed for the SLU tool. The WIM tool is available to all with an ArcGIS® Pro which increases SKB's ownership of the results and modelling methodology, and uncertainty analyses are relatively simple (compared to the SLU tool) given that its only data input is the DEM and the location of surface waterbodies. For these reasons, this study has chosen to examine WIM as a tool to aid in the prediction of future wetlands instead of the SLU tool.

2.1.3 Model application

This study plans to produce a suite of WIM models which will be used to help predict the extent of wetlands for the future landscape at Forsmark. The model will be applied to the portions of the Forsmark DEM[8] relevant for the safety analyses of both SFR and the Spent Fuel Repository. The model may also be applied to several alternate versions of the Forsmark DEM which have been manipulated in order to account for erosion and sedimentation processes.[9]

The two primary purposes of modelling the extent of future wetlands at Forsmark are: a) to aid in the delineation of future biosphere objects, i.e. the areal demarcation of the landscape wherein special focus is given in the modelling of radionuclide transport through the regolith and at the land surface, and b) to aid in the prediction of peat development by helping parameterize the extent of future wetlands.

The suite of WIM models produced in this study is called "WIM Forsmark 1.1". The workflow used in the production of WIM Forsmark 1.1 is presented in Figure 2-2. A version number is given in order to denote potential additions/subtractions of individual WIM algorithms to/from the suite and/or to denote changes in the individual WIM algorithms which make up the suite. Any changes in the number of algorithms used and/or the training of the algorithms would warrant an update of the version number as would any updates or deviations from the workflow presented in Figure 2-2. It is assumed that any updates to the version number will not warrant a comprehensive re-iteration of the modelling methods described herein and that study will suffice as the primary documentation of the modelling methodology.

[8] The current version of the Forsmark DEM (at the time of this report was published) is presented in Petrone and Strömgren (2020)

[9] Future safety analyses will most likely consider results from the landscape development model UNTAMO (Gunia and Gunia 2021) which may include several updated DEMs.

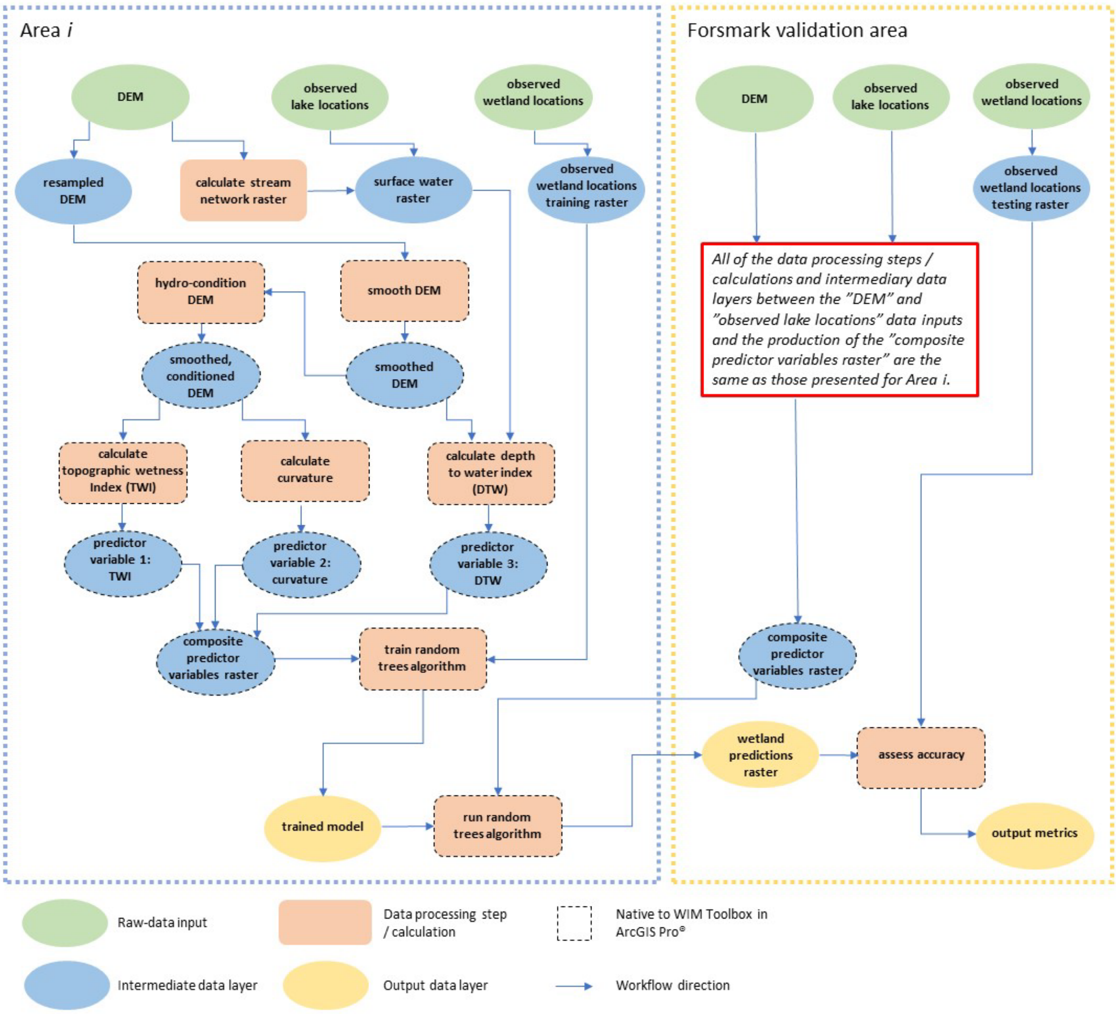

Figure 2-2. WIM Forsmark 1.1 workflow. Figure is largely informed by Figure 2 in ESRI (2021).

2.1.4 Presentation of results

The workflow presented in Figure 2-2 details the flow of data through WIM Forsmark 1.1. The primary output of the workflow being the trained random trees models, the wetland prediction raster and the performance metrics of the wetland predictions for the specific random-trees algorithm under investigation. The workflow in Figure 2-2 does not, however, specify how results of the 10 individual algorithms is presented. As mentioned in Section 2.1.1, this study presents two different methods to examine the wetland predictions produced by WIM Forsmark 1.1:

- The first method examines predictions from an amalgamation of all 10 individual algorithms; an examination of the "hit-maps" (see Section 2.1.1) is presented along-side this presentation of results.
- The second method examines the probability of wetland prediction for each cell averaged across the 10 individual algorithms used in WIM Forsmark 1.1 (see Section 2.1.1). This method is termed "p-means" in this study.

2.1.5 Definition of "wetland" in the context of this study

Training of the individual WIM algorithms that make up WIM Forsmark 1.1 requires data for observed wetlands (see Section 2.1 and Figure 2-2). All data for observed wetlands was taken from the online services provided by The Swedish Land Survey (Lantmäteriet) (see Section 2.2.3). Therefore, within the context of this study, the definition of a wetland (denoted as "marshlands by Lantmäteriet) is taken from the description of the data layers used as input data. The definition of a wetland, according to Lantmäteriet, is divided into two parts where each respective definition corresponds to the description of a single data-layer; the observed wetland data used in this study is an amalgamation of the two data layers defined in Table 2-1.

Table 2-1. Data layer names and definitions for marshlands (or wetlands in the context of this study) according to Table 81 in Lantmäteriet 2020.

Data layer name	Description
"Marshland (wetland)*"*	*"Commonly peat forming fen with shrubs and grass sedge. The area is usually accessible for walking. Can be covered by trees or completely free from trees as well as just a few trees"*
"Marsland (wetland)*, almost impassable"*	*"The area is usually hard to access and can be waterlogged. Peat forming watery fens and soft bed without vegetation. Overgrown lakes with reed. Can be covered by trees or completely free from trees as well as just a few trees."*

2.2 Input data used for training of WIM algorithms

2.2.1 Areas used for training of individual WIM algorithms

As mentioned above in Section 2.1.1, WIM Forsmark 1.1 consists of a suite of 10 separate WIM algorithms that are examined simultaneously in the prediction of wetland extent. The methods used in the production of the WIM algorithm specific to each area i ($i = 1\text{-}10$) are presented in Figure 2-2. The location of each area i is presented in Figure 2-3.

All of the areas were chosen due to their similar topographical conditions (i.e. mean percent slope of the area) and land use characteristics relative to Forsmark (Table A-1 and A-2 and Figures A-1 – A-9). One exception to this is the Krycklan area where the topographic gradient is significantly larger than that for Forsmark (the mean percent slope of Krycklan and Forsmark is 9.9% and 3.1% respectively, Table A-1). Krycklan was included anyway as it is often considered as a hydrological proxy of the Forsmark landscape in the far future. The island of Gräsö, directly east of Forsmark, is considered separate from the terrestrial portions of the Forsmark investigation area. Due to much higher prevalence of exposed bedrock on Gräsö relative Forsmark (Figure B-1), it is assumed that the conditions conducive for wetland formation on Gräsö will be significantly different than those at Forsmark and it would therefore be beneficial to consider the areas separately when training the individual WIM algorithms within the context of this study.

The areas selected for the algorithm training span the entire length and breadth of Sweden. This was done in order to try and capture a wide range of climatological conditions that would affect wetland development. Temperature and precipitation statistics from the Swedish Meteorological and Hydrological Institute (SMHI) for the 10 areas used for training of the WIM algorithms are presented in Appendix C.

Delineation of the area boundaries was done using watershed delineations from SMHI[10]. The SMHI watershed IDs used to delineate the 10 areas in Figure 2-3 are presented in Table A3.

[10] Data publicly available via SMHI's *Svenskt vattenarkiv* (SVAR 2012): https://www.smhi.se/data/utforskaren-oppna-data/vattendrag-svar-2012

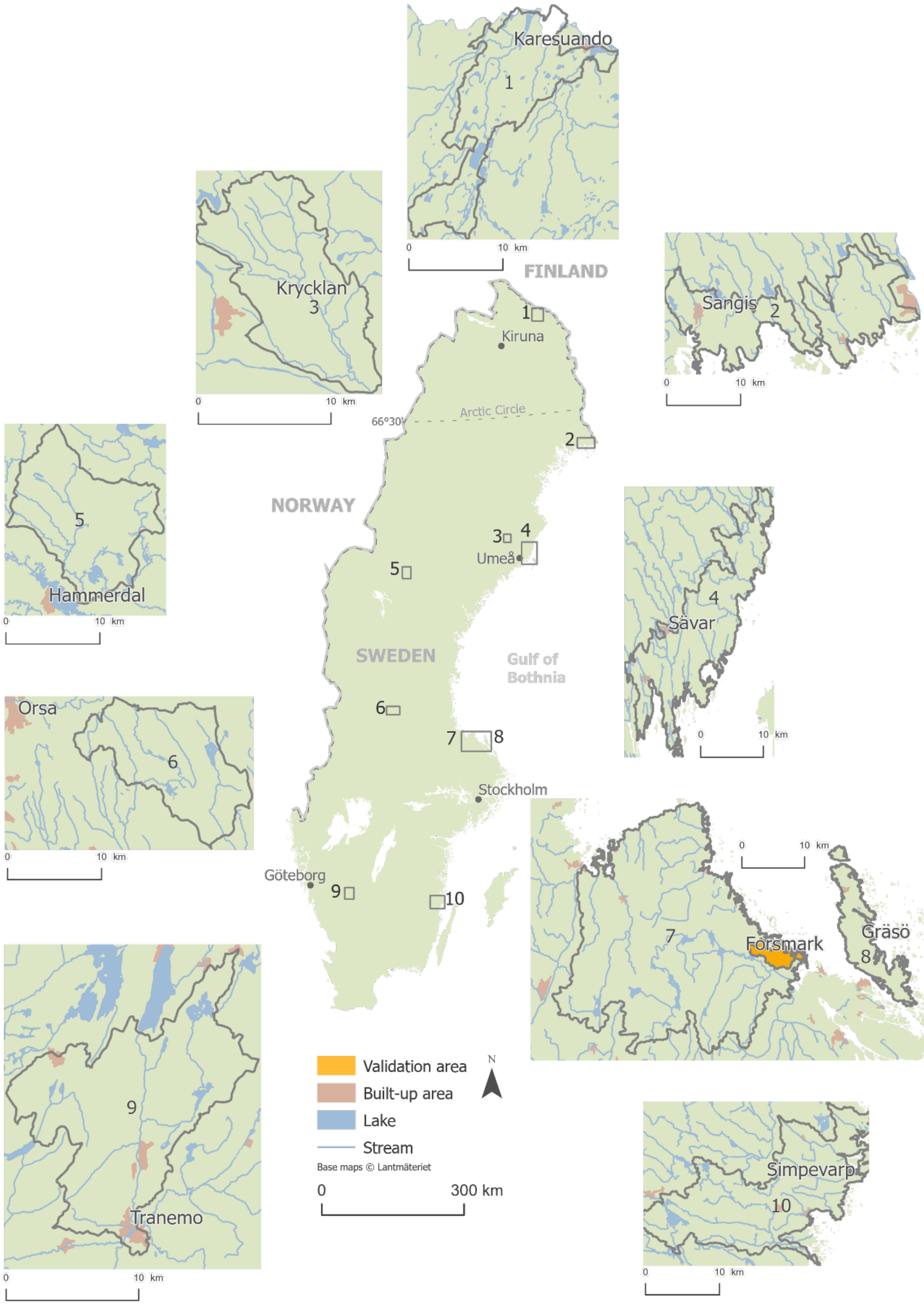

Figure 2-3. The 10 individual areas used for the training of the 10 WIM algorithms that make up WIM Forsmark 1.1. Note that at the eastern edge area 7: Forsmark a smaller area is delineated; this smaller area corresponds to the Forsmark validation area and is excluded from "area 7: Forsmark".

2.2.2 Elevation data

Elevation data in the form of a digital elevation model (DEM) is a primary data input in the WIM tool in ArcGIS Pro® (ESRI 2021) and, therefore, in the production of WIM Forsmark 1.1 (Figure 2-2). The DEM is used to calculate the TWI, DTW and curvature (Section 2.1 and Figure 2-2). The DEM is also used to generate surface water input rasters which are used in the calculation of the DTW (Section 2.2.4 and Figure 2-2).

DEMs for each of the 10 areas was downloaded via the online services provided by Lantmäteriet.[11] DEMs for each area were resampled from a 2×2 m to a 10×10 m grid using a "nearest neighbour" algorithm in order to be compatible with the WIM tool in ArcGIS Pro®. The DEM is then "clipped" using the boundaries for each area (Figure 2-3). The resulting DEMs for each area are presented in Appendix A.

Within the WIM tool in ArcGIS Pro®, a "smoothing algorithm" is applied to the DEM (Figure 2-2) in order to remove variations in the data deemed too small to be indicative of an actual topographic feature (ESRI 2021). For this study, the "median" smoothing method is used with a smoothing width of 50 m (see Section 7.2 in ESRI 2021).

2.2.3 Observed wetlands

Land-use data which maps the location of observed wetlands is a primary data input in the WIM tool in ArcGIS Pro® (ESRI 2021). The observed locations are input as 10×10 m raster data using the same grid as the DEM (see Section 2.2.2). Observed wetland data from each training area *i* is used only for the training of the individual algorithms (Figure 2-2). Observed wetland data used to test the performance of the individual algorithms is discussed later in Section 2.3.

Observed wetland data was downloaded via the online services provided by Lantmäteriet.[12] At the time of this study, data wetland data from Lantmäteriet was provided as in vector format (i.e. a "shapefile") which was then converted to a raster with the same grid as the DEM. Wetland data for each area is presented in Appendix A.

Lantmäteriet limits wetland observations to wetlands with a cohesive area of at least 2500 m^2 (Table 81, Lantmäteriet 2020); any observed wetlands with areas smaller than this are not used in the training of the individual WIM algorithms. The accuracy of the location data for wetlands is ±20 m (Table 68, Lantmäteriet 2020). Data on observed wetlands does not account for dried or artificially drained wetlands that are currently used or have previously been used as productive forest land (Table 81, Lantmäteriet 2020). It should also be noted that the data on observed wetlands also does not account for "old" wetlands that have been drained and are currently being used as arable land.

2.2.4 Surface water

The locations of surface waterbodies such as lakes, streams and rivers are a primary data input in the WIM Forsmark 1.1 workflow (Figure 2-1). Stream and lake locations are input as 10×10 m raster data using the same grid as the DEM (see Section 2.2.2). Surface water data is used in the calculation of the DTW (see Section 2.1 and ESRI 2021).

Data for lakes uses the same dataset used for the observed wetlands[11], however, the accuracy of the location data for shoreline location is ±5-10 m (instead of ±20 m for wetlands, see Section 2.2.3) depending on whether the shoreline is or is not diffuse in nature. Much like the wetland data (see Section 2.2.3), original data was proved as a shapefile which was then converted to a raster with the same grid as the DEM (Lantmäteriet 2020).[11] Due to the fact that many wetlands are located surrounding lakes, raster conversion of the lakes and wetland shapefiles was conducted simultaneously in order to avoid overlapping raster cells.

In this study, data for the stream network in each area was estimated using the "D8" flow-accumulation algorithm available in the "Hydrology toolset" in ArcGIS PRO® as applied to the

[11]©Lantmäteriet: https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/markhojdmodell-nedladdning/

[12]©Lantmäteriet: https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/hydrografi-visning-inspire/#qry=HY.PhysicalWaters.Wetland

DEM for each respective area. An accumulation threshold of 4 ha was used to determine which cells would be included in the stream network. The stream network raster was then combined with the lake raster.[13]

When using the WIM tool in ArcGIS Pro®, observed stream-network data can be used as a direct input or combined with the calculated stream network results from a flow-accumulation algorithm via tertiary data-processing steps. For this study, observed data for stream location was only used to help verify the calculated stream network.

2.3 Assessing performance of WIM Forsmark 1.1

In this study, the performance of the individual WIM algorithms in WIM Forsmark 1.1 is tested against their ability to predict wetlands within the “Forsmark validation area” (Figure 2-4). When using the WIM tool in ArcGIS Pro®, the user is instructed to delineate the area used for algorithm training into a “train” area and a “test” area; the former is used in the algorithm training and the latter is used to measure algorithm performance (ESRI 2021). For this study, algorithm performance is assessed using the Forsmark validation area; this means that training of the 10 individual algorithms is conducted using the entirety of the areas shown in Figure 2-3.

Delineation of the Forsmark validation area is delineated according to the eight hydrological catchments presented in Brundberg et al. 2004. Data for the observed wetlands and lakes shown in Figure 2-4 is taken from the same data sources discussed in Sections 2.2.3 and 2.2.4. The DEM for the area (Figure A-10) corresponds to the terrestrial portions of the 10×10 m Forsmark DEM as presented in Petrone and Strömgren (2020). The drainage network used to generate wetland predictions within the Forsmark validation area (not pictured in Figure 2-4) is produced using the same methods discussed in Section 2.2.4 together with the DEM for the Forsmark validation area (see Figure 3-2and Figure 3-1).

In this study, the performance of WIM Forsmark within the Forsmark validation is assessed in three different ways:

- The primary method of assessing performance examines the amalgamated WIM Forsmark 1.1 predictions against the observed wetland data.
- A secondary assessment of model performance is presented which examines each of the algorithms included in WIM Forsmark 1.1 individually. This is done in order help provide a point of comparison at which the potential added value of using the amalgamated predictions to predict wetland extent may be assessed.
- Finally, a preliminary assessment of model performance for the p-means method is presented wherein the prediction accuracy is examined by only accounting for cells with a mean probability of prediction greater than 0.50.

The hit-maps produced in this study are not used explicitly when investigating model performance, i.e. no investigation of model performance as a function of the number of hits is examined in this study. However, it is the opinion of the authors that this should be investigated if future work with WIM Forsmark 1.1 is pursued.

[13] The accumulation threshold used to generate the stream network should be consistent with the threshold used in the calculation of the TWI predictor variable (see Figure 2-2).

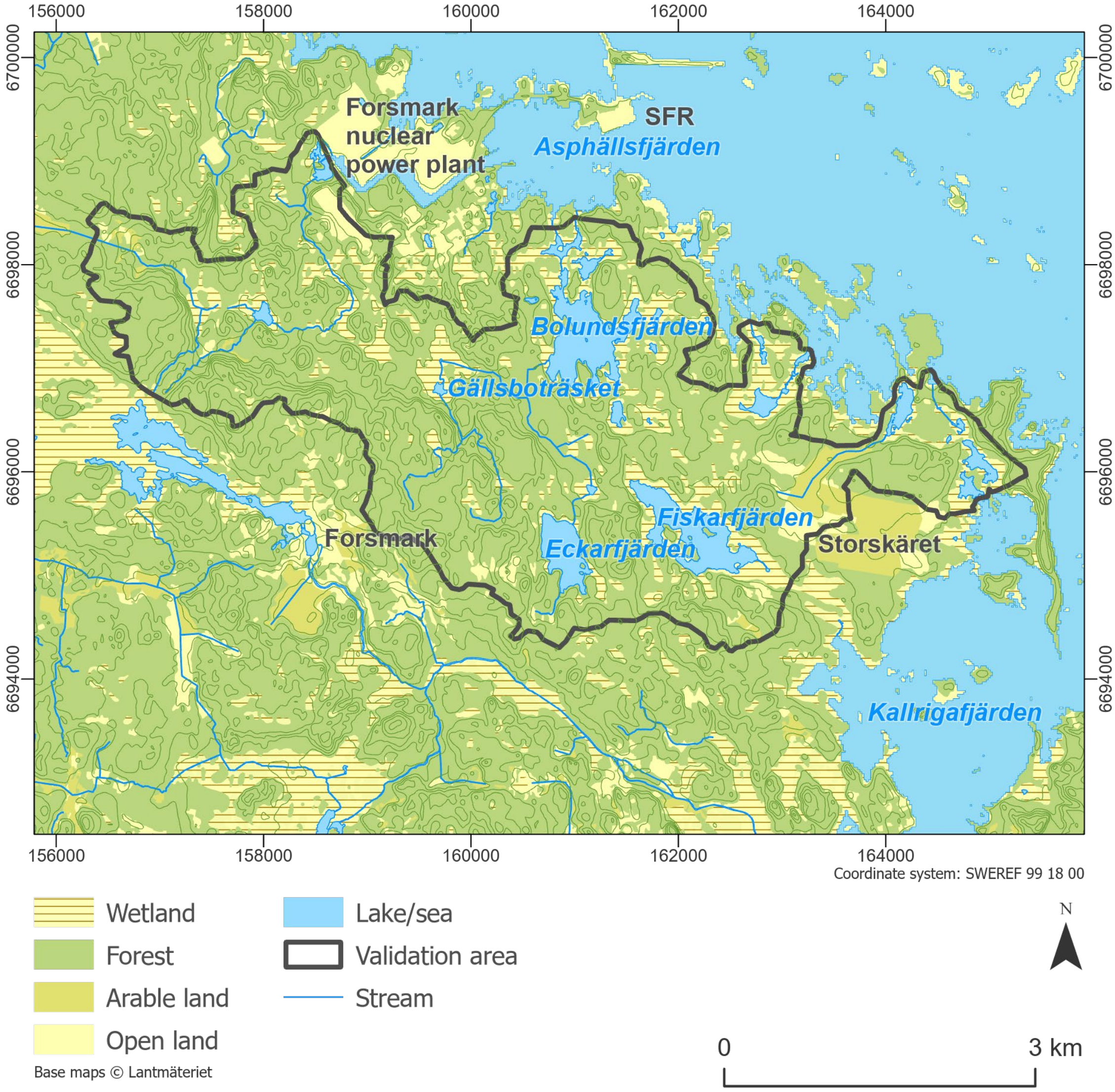

***Figure 2-4.** Delineation and land-use classifications in the Forsmark validation area. The stream network shown in the figure is represents the observed data for the area and is not directly used in wetland prediction.*

Following the training of the individual algorithms, individual algorithm performance, as well as the performance of the amalgamated predictions, is assessed against the observed wetland data within the Forsmark validation area. Algorithm performance is reported using three metrics:

Precision: The ratio of true positive predictions (i.e. positions where the algorithm correctly predicted the existence of a wetland) with the total number of positive positions (i.e. all positions where the algorithm predicted a wetland both correctly and incorrectly) as shown in Equation 2-1. Precision is used to help quantify the level with which an algorithm may overpredict the wetlands; values close to zero indicate substantial overprediction and values close to one indicate near-perfect prediction.

$$Precision = \frac{true\ positive\ predictions}{all\ positive\ predictions} \qquad 2\text{-}1$$

Recall: The ratio of true positive predictions to all of the true wetlands (i.e. positions where observed data indicates a wetland) as shown in Equation 2-2. Recall helps quantify the "detection rate" of the algorithm or the extent to which an algorithm underpredicts the existence of wetlands; values close to zero indicate substantial underprediction and values close to one indicate near-perfect prediction.

$$Recall = \frac{true\ positive\ predictions}{all\ true\ wetlands} \tag{2-2}$$

F_1-score: The harmonic mean of the precision and the recall according to equation 2-3. The F_1-score can be used examine the over- and underprediction rates simultaneously; values close to zero indicate poor overall performance and values close to one indicate near-perfect performance.

$$F_1 = 2\frac{precision \cdot recall}{precision + recall} \tag{2-3}$$

Furthermore, the amalgamated model predictions are also compared to wetlands predicted by SLU's land "wetness" maps (see Section 2.1.2) in order assess the performance of WIM Forsmark 1.1 compared to an alternate, peer-reviewed wetland prediction tool.

3 Results

3.1 Wetland prediction within the Forsmark validation area

Wetland predictions for the Forsmark validation area presented as an amalgamation of wetland predictions from the 10 individual WIM algorithms included in WIM Forsmark 1.1, along with a comparison of how the predictions correspond with the observed wetland data for the same area (see Section 2.3) are presented in Figure 3-1 and Figure G-1. Wetland predictions for the Forsmark validation area from each of the 10 individual algorithms included in WIM Forsmark 1.1 are presented in Appendix E. Wetland predictions for the Forsmark validation area presented as a "hitmap" (see Section 2.1.1) are shown in Figure 3-2 and Figure F-1.

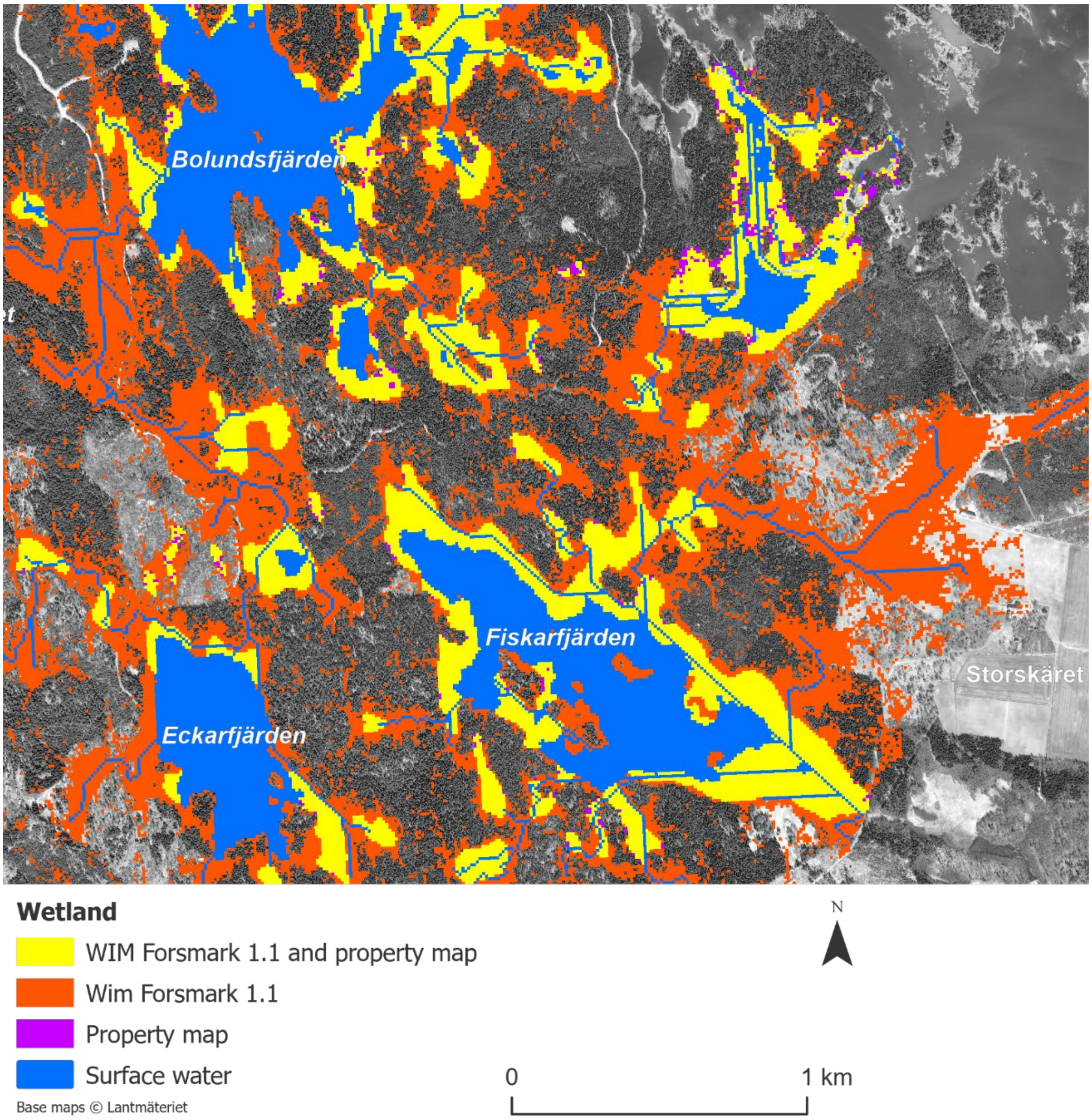

Figure 3-1. Extraction of Figure G-1. Wetland predictions, presented as the amalgamated results from the 10 individual algorithms included in WIM Forsmark 1.1, and observed wetlands (Lantmäteriet 2020) within a ~3.5×3 km portion of the Forsmark validation area surrounding the lakes Bolundsfjärden, Eckarfjärden and Fiskarfjärden. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

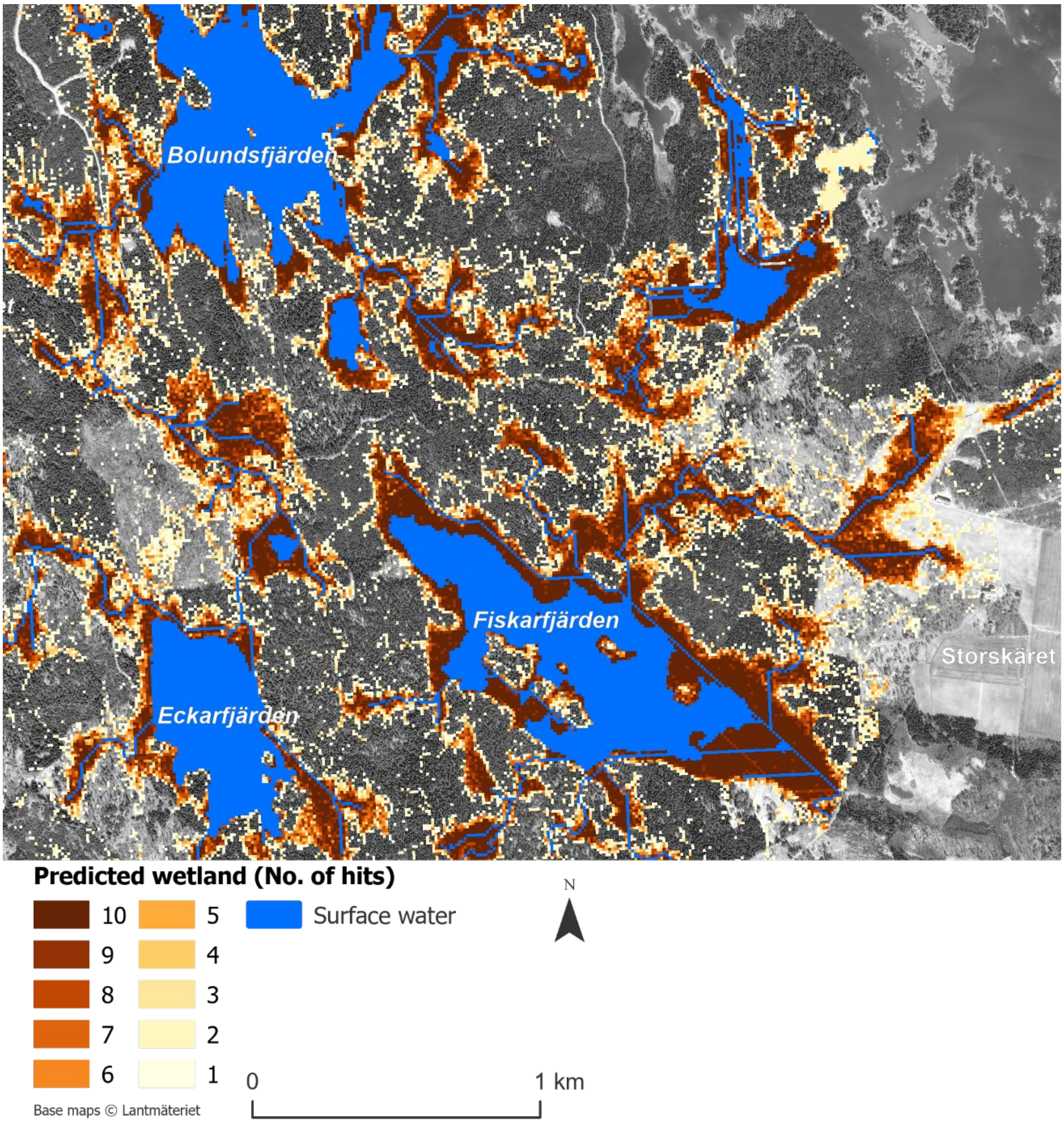

Figure 3-2. Extraction of Figure F-1. Wetland predictions within a ~3.5×3 km portion of the Forsmark validation area surrounding the lakes Bolundsfjärden, Eckarfjärden and Fiskarfjärden; results are presented as a hitmap with each "hit" representing a cell predicted to contain a wetland according to one of the 10 individual algorithms that make up WIM Forsmark 1.1. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

Wetland predictions for the Forsmark validation area presented using the p-means methodology with a mean probability of greater than 0.50 (see Section 2.1.1 and 2.1.4), along with a comparison of how the predictions correspond with the observed wetland data for the same area (see Section 2.3) are presented in Figure 3-3 and Figure I-1. Wetland predictions for the Forsmark validation area using the p-means methodology with a probability of 0-1 are shown in Figure 3-4 and Figure I-2.

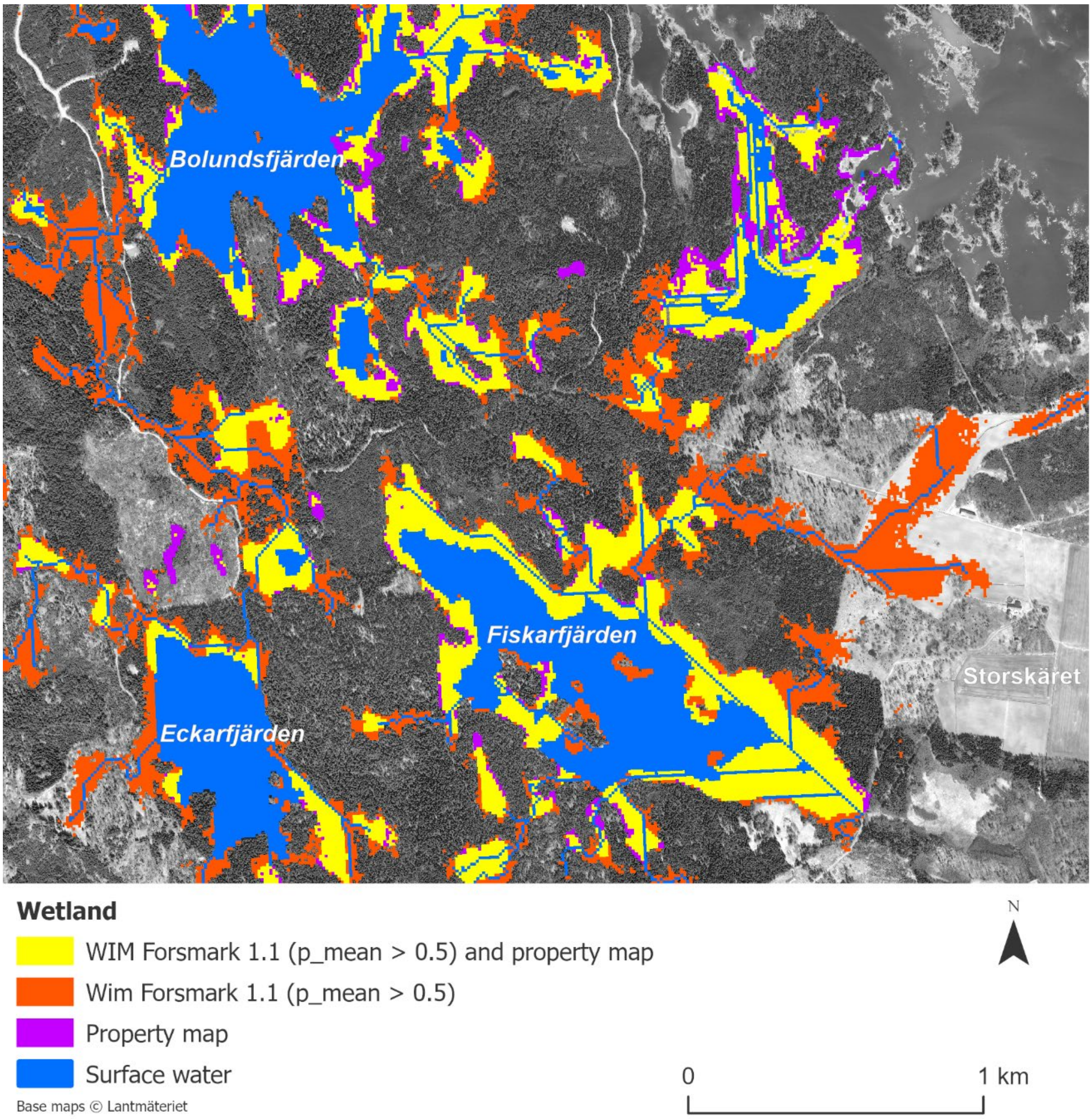

***Figure 3-3.** Extraction of Figure I-1. Wetland predictions presented as all cells with a mean probability of a positive wetland prediction greater than 0.50 according to the p-means methodology, and observed wetlands (Lantmäteriet 2020) within a ~3.5×3 km portion of the Forsmark validation area surrounding the lakes Bolundsfjärden, Eckarfjärden and Fiskarfjärden. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.*

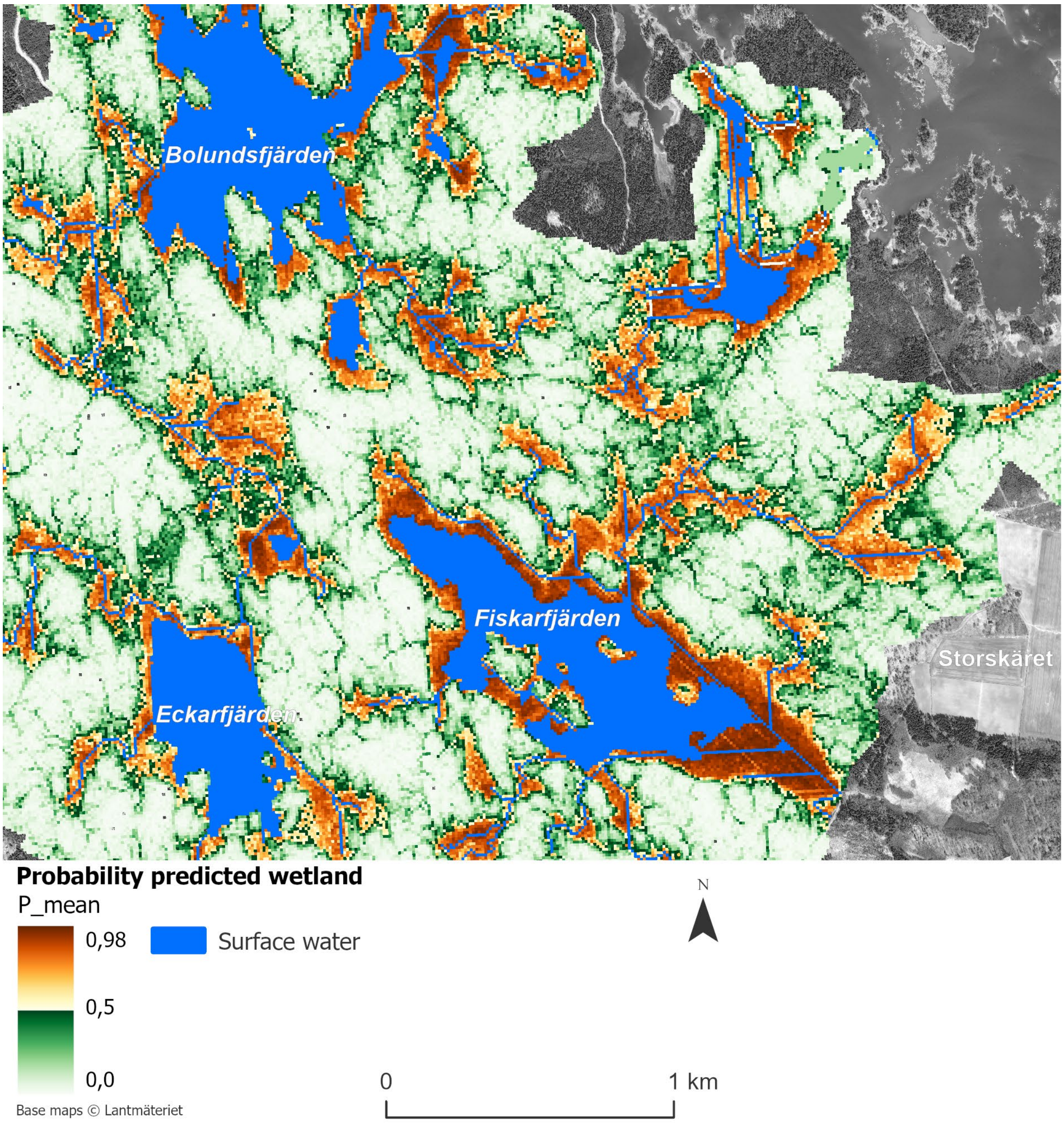

Figure 3-4. Extraction of Figure I-2. Wetland predictions within a ~3.5×3 km portion of the Forsmark validation area surrounding the lakes Bolundsfjärden, Eckarfjärden and Fiskarfjärden; results are presented as the mean probability of a positive wetland prediction for each cell according to the p-means methodology. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

3.2 Algorithm performance

Performance metrics (see Section 2.3) for the 10 individual algorithms included in WIM Forsmark 1.1, the amalgamation of all of the algorithms and the p-means method ($p > 0.50$) are presented in Table 3-1. The average *Precision*, *Recall* and F_1*-score* for the methods considered were 0.476, 0.799 and 0.588 respectively (Table 3-1).

For the 10 individual algorithms (Maps showing predictions from the individual algorithms are presented in Appendix E), results indicated that the algorithm trained using "area 3: Krycklan" was the worst performing (i.e. lowest F_1*-score*) of all of the individual algorithms considered as this algorithm resulted in the largest overprediction of wetlands in the Forsmark validation area (*Precision* = 0.347), see Figure E-3. The best performing algorithm (i.e. highest F_1*-score*) was trained using "area 1: Karesuando"; this is due to above average values for both the *Precision* and *Recall*, see Figure E-1.

For the amalgamation of all of the algorithms (i.e. predictions from all 10 algorithms considered simultaneously) results indicated that this method produced the lowest F_1*-score*; this indicates that the amalgamation greatly overpredicted the existence of wetlands within the Forsmark validation area (*Precision = 0.315*) according to the observed data. However, this method had the highest recall score of all of the methods used to assess model performance meaning that the amalgamated results was most capable of predicting cells with observed wetland data. Further work with WIM Forsmark 1.1 should include an assessment of model performance using the hit-map in order to further quantify the predictive capacity of the amalgamated results.

The performance metrics for the p-means methodology (see Sections 2.1.1 and 2.1.4), which examines all cells with an average probability of a positive wetland prediction greater than 0.50, indicated that this method produced the highest F_1*-score* for all of the methods used to assess algorithm performance. This was due to the values of both *Precision* and *Recall* being well above average.

Table 3-1. Performance metrics for the 10 individual algorithms. Algorithm performance is assessed using the Forsmark validation area (Figure 2-4). Algorithm names can be found in Appendix D.

Area	Reference location	Precision	Recall	F_1-score
1	Karesuando	0.509	0.807	0.624
2	Sangis	0.541	0.694	0.608
3	Krycklan	0.347	0.903	0.501
4	Norum	0.447	0.842	0.584
5	Hammerdal	0.556	0.724	0.629
6	Skattungbyn	0.533	0.738	0.619
7	Forsmark	0.541	0.694	0.608
8	Gräsö	0.463	0.825	0.593
9	Tranemo	0.448	0.796	0.573
10	Simpevarp	0.479	0.792	0.597
	Amalgamation	0.315	0.956	0.474
	P-means (p>0.50)	0.536	0.815	0.647

3.2.1 Predicted wetlands in drained agricultural areas

Observed wetland data from Lantmäteriet does not account for previous wetland areas that may have been drained to be used as agricultural land. This implies that at least some of the overprediction of wetlands seen for all of the individual algorithms included in WIM Forsmark 1.1 (Table 3-1) may be due to the prediction of wetlands in drained areas that do not appear in the observed data. Data showing a limited extent of observed drainage ditches was obtained from SLU[14] in order to investigate whether WIM Forsmark 1.1 predicted the existence of wetlands in areas that may have been drained for agricultural use. Amalgamated wetland predictions in a portion of the Forsmark area is plotted together with data on observed locations of drainage ditches in Figure 3-5. Results show that WIM Forsmark 1.1 does predict the existence of wetlands in drained areas that are currently used as arable land. It is assumed that these areas likely contained wetlands prior to being drained.

[14] Data for drainage ditches (obtained via email from SLU on August 26th, 2021) is "working material" used by SLU in the studies with the wetness maps and is not yet publicly available.

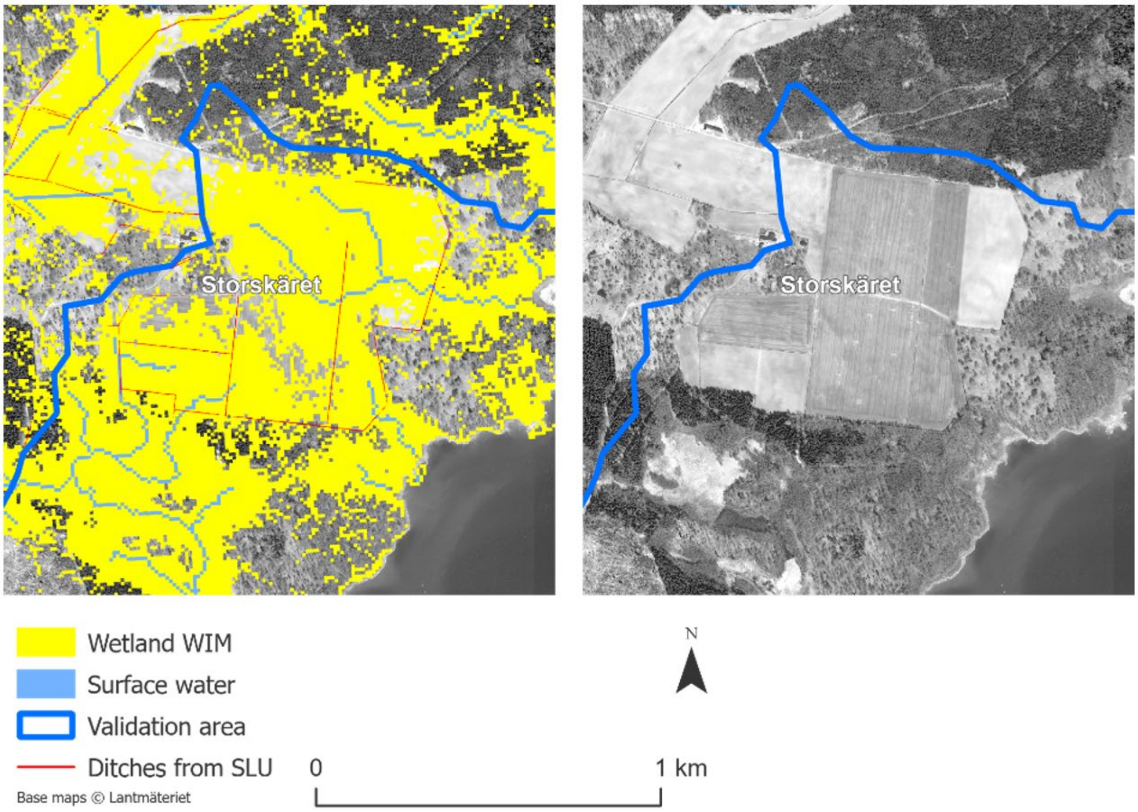

Figure 3-5. Amalgamated wetland predictions within and outside the eastern extents of the Forsmark validation area, observed drainage ditches an aerial photo of the area (left) and the same aerial photo without added data showing the arable land in the area (right). The waterways shown (left) represent the surface water data inputs which were calculated using the flow-accumulation algorithm in ArcGIS® Pro in Section 2.2.4.

3.3 Model predictions compared to SLU wetness map

Wetland predictions for the Forsmark validation area presented as an amalgamation of wetland predictions from the 10 individual WIM algorithms included in WIM Forsmark 1.1, along with a comparison of how the predictions correspond with both the observed wetland data and SLU's wetness map (see Section 2.1.2) is presented in Figure 3-6. and Figures H-1, H-2 and H-3 . Results indicate that wetland predictions from WIM Forsmark 1.1 agree remarkably well with SLU's wetness map's predictions when examining the amalgamation of the "moist-mesic" and "wet-moist" classifications (Figure 3-6. and Figure H-1); performance of the wetland predictions from WIM Forsmark 1.1 was markedly worse when compared to the SLU maps for "moist-mesic" and "wet-moist" individually (Figures H-2 and H-3 respectively). It should however be noted that the definition of a "wetland" differs between that used in this report and that used for the SLU wetness maps: the definition of the observed wetlands used in this study is based primarily on the classification of vegetation (see Section 2.1.5) while SLU's definition centers around a qualitative definition of the soil-moisture content (see Section 2.1.2).

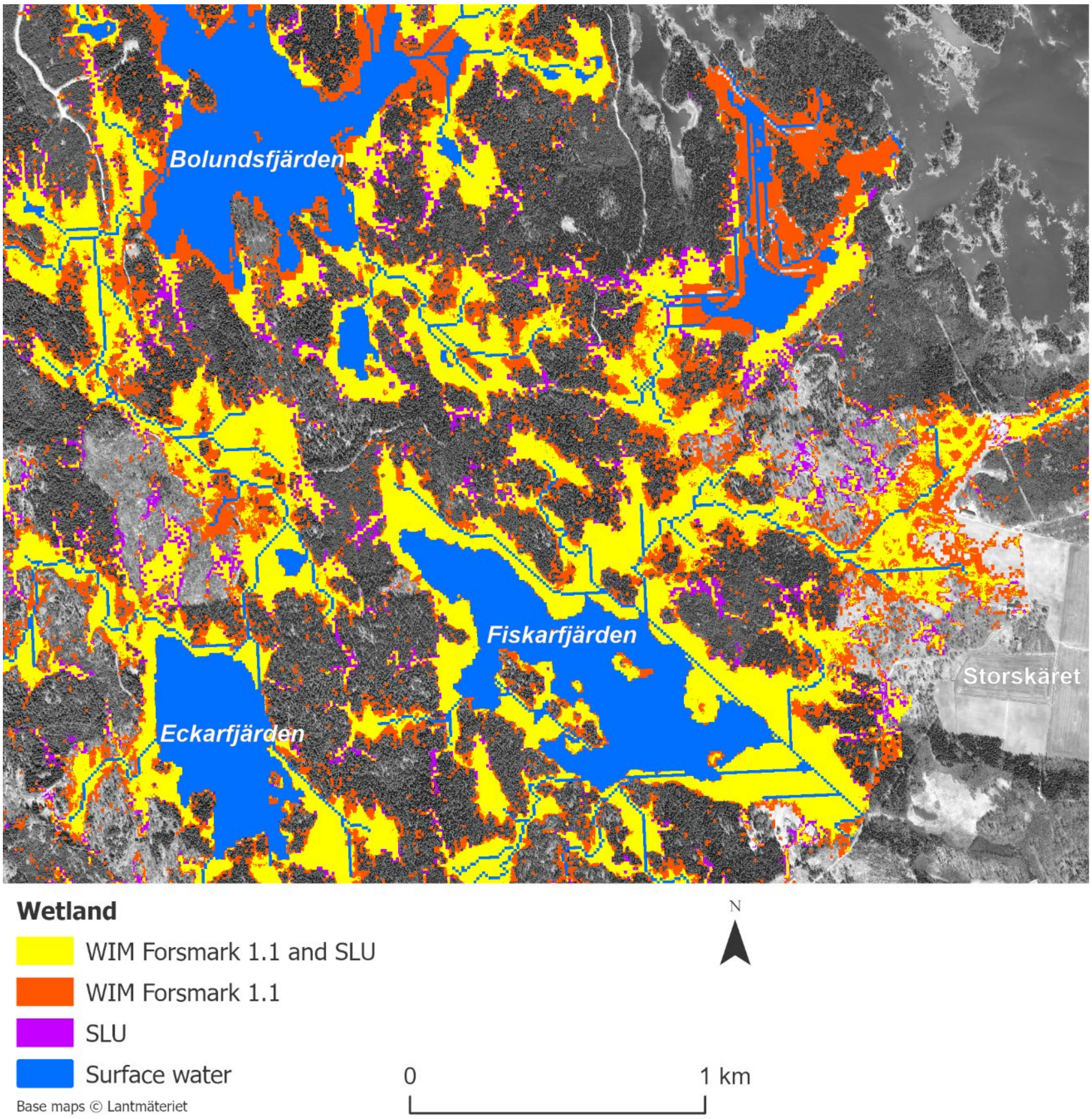

Figure 3-6. Extraction of Figure H-1. Wetland predictions, presented as the amalgamated results from the 10 individual algorithms included in WIM Forsmark 1.1, observed wetlands and wetland predictions from SLU's wetness maps within a ~3.5×3 km portion of the Forsmark validation area surrounding the lakes Bolundsfjärden, Eckarfjärden and Fiskarfjärden. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4. Wetness classes "mesic-moist" (frisk-fuktig) and "moist-wet" (fuktig-blöt) from the SLU wetness map are used in the comparison (see Section 2.1.2).

…

4 Suggested future use of WIM Forsmark 1.1

As stated in Section 1, this study proposes that machine learning is a viable alternative to process-based modelling of wetland development. Results indicate that the WIM methodology and the amalgamation of algorithms which comprise WIM Forsmark 1.1 is capable of predicting observed wetlands within the Forsmark area. Results also indicate that WIM Forsmark 1.1 is capable of largely reproducing the results of the SLU wetness map in spite of the relative simplicity of WIM Forsmark 1.1 relative to the machine learning algorithms used to produce the SLU wetness maps. Furthermore, the machine learning algorithms used to produce the SLU wetness maps are not yet publicly available which therefore limits the extent to which these algorithms can be used to produce predictions of future wetlands within postulated landscapes. The authors therefore suggest that WIM Forsmark 1.1 (or an updated version) can be used as a tool for predicting the extents of future wetlands and should be considered in landscape development studies which focus on the Forsmark area.

This study briefly examined the performance of WIM Forsmark 1.1 in relation to agricultural areas. The training of the algorithms incorporated in WIM Forsmark 1.1 did not consider the effects that drained agricultural areas may or may not have on wetland predictions. Future work with WIM Forsmark 1.1 should examine the validity of the predictions when pertaining to current or future agricultural areas.

References

SKB's (Svensk Kärnbränslehantering AB) publications can be found at www.skb.com/publications.

Beven K J, Kirkby M J, 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrological Sciences Bulletin, 24:1, 43–69.

Bosson E, Sassner M, Sabel U, Gustafsson L-G, 2010. Modelling of present and future hydrology and solute transport at Forsmark. SR-Site Biosphere. Updated 2013-01. SKB R-10-02, Svensk Kärnbränslehantering AB.

Breiman L, 2001. Random Forests. Machine Learning 45, 5–32.

Brundberg A-K, Carlsson T, Blomqvist P, Brydsten L, Strömgren M, 2004. Forsmark site investigation – Identification of catchments, lake-related drainage parameters and lake habitats. SKB P-04-25, Svensk Kärnbränslehantering AB.

ESRI, 2021. Arc Hydro: Wetland Identification Model. Redlands, CA: Environmental Systems. Research Institute.

Gunia M, Gunia K, 2021. UNTAMO model description – Model version 3.5. SKB P-21-26, Svensk Kärnbränslehantering AB.

Lantmäteriet, 2020. Product Description: GSD-Property map, vector. Document version 6.2.7. Lantmäteriet, Sweden.

Lidberg W, Nilsson M, Ågren A, 2020. Using machine learning to generate high-resolution wet area maps for planning forest management: A study in a boreal forest landscape. Ambio 49, 475–486.

O'Neil G L, **Goodall J L**, **Watson, L T, 2018**. Evaluating the potential for site-specific modification of LiDAR DEM derivatives to improve environmental planning-scale wetland identification using random forest classification. Journal of Hydrology, 559, 192–208.

O'Neil G L, Saby L, Band L E, Goodall J L, 2019. Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model. Water Resources Research, 55.

Petrone J, Strömgren M, 2020. Baseline Forsmark – Digital elevation model. SKB R-17-06, Svensk Kärnbränslehantering AB.

SKB, 2011. Long-term safety for the final repository for spent nuclear fuel at Forsmark. Main report of the SR-Site project. Updated 2015-05. SKB TR-11-01, Svensk Kärnbränslehantering AB.

SSM, 2018. Strålsäkerhet efter slutförvarets förslutning. SSM 2018:07, Strålsäkerhetsmyndigheten, Sweden.

Werner K, Sassner M, Johansson E, 2013. Hydrology and near-surface hydrogeology at Forsmark - synthesis for the SR-PSU project. SR-PSU Biosphere. SKB R-13-19, Svensk Kärnbränslehantering AB.

Ågren A M, Lidberg W, Strömgren M, Ogilvie, J, Arp P A, 2014. Evaluating digital terrain indices for soil wetness mapping–a Swedish case study. Hydrology and Earth System Sciences, 18(9), 3623–3634.

Ågren A, Lidberg W, 2020. Dokumentation nya hydrografiska kartor – vattendrag och SLU Markfuktighetskartor. Sveriges lantbruksuniversitet, Sweden.

Appendix A Topography and land-use the areas used for training of algorithms

Table A-1. Elevation and slope statistics for each of the 10 areas used to train the WIM algorithms that make up WIM Forsmark 1.1. All statistics were calculated using the DEMs respective for each area. DEMs for each area were downloaded using Lantmäteriet's online services.[a] Note: the same statistics are reported for the Forsmark validation area which was not used in algorithm training.

ID	Area name	Elevation			Slope (%)			
		Max	*Min*	*Range*	*Mean*	*STDV*	*Max*	*Min*
1	Karesuando	453.9	320.4	134.0	3.0	4.2	71.7	0.0
2	Sangis	60.2	-1.7	61.9	3.2	5.0	98.5	0.0
3	Krycklan	372.6	107.0	265.6	9.9	9.3	185.5	0.0
4	Norum	62.1	-0.1	62.2	3.0	3.1	65.2	0.0
5	Hammerdal	385.6	302.9	82.7	3.5	4.4	60.7	0.0
6	Skattungbyn	392.4	246.0	146.4	3.5	3.3	58.3	0.0
7	Forsmark	55.2	-10.0	65.2	3.1	3.0	82.7	0.0
8	Gräsö	27.1	-0.1	27.1	5.0	4.1	62.7	0.0
9	Tranemo	276.9	150.7	126.2	5.1	5.4	101.6	0.0
10	Laxemar	110.8	-3.6	114.3	7.4	6.1	79.0	0.0
	Forsmark Validation area	27.4	-1.6	29.0	3.8	3.3	32.8	0.0

a: ©Lantmäteriet: https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/markhojdmodell-nedladdning/

Table A-2. Total area, proportion of wetlands and proportion of land use for each of the 10 areas used to train the WIM algorithms that make up WIM Forsmark 1.1. Note: the same statistics are reported for the Forsmark validation area which was not used in algorithm training. Land use data for each area was downloaded using Lantmäteriet's online services.[a]

ID	Area name	Total Area (m^2)	Wetland (%)[b]			Land use (%)[c]				
			Total	*In forest*	*In open land*	*Lake*	*Forest*	*Open land*	*Arable land*	*Developed*
1	Karesuando	1.639×10^8	34.9	0.5	33.4	11.5	50.1	37.6	0.0	0.0
2	Sangis	2.468×10^8	18.3	9.6	8.8	1.9	82.5	12.8	2.1	0.2
3	Krycklan	1.215×10^8	8.1	5.4	2.7	0.6	93.8	3.7	1.8	0.0
4	Norum	3.192×10^8	17.9	9.6	8.3	3.0	81.3	12.2	3.1	0.5
5	Hammerdal	1.619×10^8	34.3	15.7	18.6	4.0	74.0	20.5	1.5	0.0
6	Skattungbyn	1.509×10^8	25.5	13.4	12.1	3.0	84.3	12.5	0.2	0.0
7	Forsmark	8.938×10^8	10.8	5.9	4.9	2.6	80.7	9.3	7.4	0.0
8	Gräsö	1.002×10^8	3.8	1.9	1.9	0.4	84.2	11.4	4.0	0.0
9	Tranemo	1.804×10^8	14.2	9.6	4.6	3.1	69.8	12.9	12.7	1.5
10	Laxemar	3.044×10^8	3.3	1.5	1.8	4.7	86.0	5.9	3.4	0.0
	Forsmark Validation area	1.948×10^7	13.1	3.6	9.5	7.5	78.1	12.7	1.7	0.0

a: ©Lantmäteriet: https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/hydrografi-visning-inspire/#qry=HY.PhysicalWaters.Wetland
b: Percentages are calculated relative the total area
c: Percentages are calculated relative the total area. Wetlands are assumed to exist in either a forested or an open landscape, i.e. the statistics reported under 'Land use – Forest' and 'Land use – Open land' may or may not contain wetlands. The statistics reported under the 'Land use' columns sum to 100% for each respective area.

Table A-3. The *Svenskt vattenarkiv* (SVAR 2012) catchment IDs (listed in comma-deliminated form) for the main- and sub-catchments used to delineate the 10 areas used to train the algorithms used in WIM Forsmark 1.1. 12-digit IDs containing a "-" indicate that the catchment is a "sub-basin". Five-digit IDs that do not contain a "-" indicate that the catchment is a "main catchment".

Area name	Area ID	SMHI catchment ID
Karesuando	1	758977-175857, 760776-177497, 760804-177366, 760741-176313, 760684-177152, 759398-176020, 759481-176034, 759172-175457, 759505-176123, 761033-176142, 759150-175943, 759625-176072, 761277-176507, 759720-176027, 759279-175746, 760999-176342, 759856-176168, 760806-176499, 761255-176708, 761267-176650, 761105-176953
Sangis	2	732555-184877, 732544-187258, 732979-184722, 732761-184831, 733232-184793, 732011-186799, 733048-184870, 732731-185797, 732414-185285, 732697-185974, 732615-185421, 732774-187357, 733312-187676, 732659-185466, 732696-185485, 733590-187486, 732945-186658, 732649-187029, 732717-185384, 732727-184952, 732670-186175, 732792-186267, 732716-185088, 732391-186445, 732857-186066, 732409-185687, 732842-185741, 732857-185513, 733120-187291, 733374-187244, 733125-187027, 732723-187905, 733086-185217, 733093-185275, 732925-187482, 733106-187550, 732991-185241, 732206-186976
Krycklan	3	712659-169980, 712982-169611, 712323-169978, 713141-169627, 712229-170068, 713548-169451, 712821-169332, 712482-169517, 713350-169154, 713172-169136, 713355-169235, 712742-169722, 713087-169457, 713374-169328
Norum	4	711559-174978, 711436-174783, 711237-174844, 708715-173940, 710770-174663, 708851-174073, 709580-174269, 711486-174894, 711785-174889, 711003-174802, 710592-174441, 711000-174361, 710560-174136, 710535-174536, 710374-174552, 710217-174390, 711417-174646, 710109-174403, 710310-174312, 711149-174519, 711242-174626, 708758-173529, 708757-173790, 709207-173589, 709279-173691, 709958-174448, 709811-173637, 709380-174080, 709201-173856, 709640-173923, 708103-173369, 707855-173482, 709240-172796, 709143-172731, 709385-173293, 709598-173329, 709771-173281, 709445-173333, 709369-173142, 709312-173342, 709544-173195, 708183-172902, 708172-173439, 708796-173251, 708540-172972, 708562-173418, 709108-173100, 708565-173342, 708164-172846, 708402-172842, 708818-173029, 708898-172918, 708064-172847, 708984-173426, 709173-172907, 708908-173007
Hammerdal	5	705390-148350, 705381-148128, 706256-148244, 706550-148088, 705541-148627, 705568-148239, 706843-147805, 705819-148033, 706255-147748, 705590-148190, 705846-148193, 705559-148064, 705801-147965, 705534-148148, 706039-147703, 706255-148711, 706586-147673, 706642-147579, 706514-147905
Skattungbyn	6	676905-145172, 677563-145665, 677362-144530, 677042-145019, 676791-145449, 676470-145713, 676727-145982, 677293-145224, 677489-145077, 677133-145392, 677217-145917, 677269-145796, 677237-145546
Forsmark	7	54055, 55000, 55056
Gräsö	8	26059
Tranemo	9	637874-135277, 637555-135206, 638416-135438, 638666-135357, 638963-135817, 638695-135875, 639366-135836, 638608-134746
Laxemar	10	636578-155016, 637033-154600, 635774-154026, 635692-154142, 636262-154861, 636090-152653, 636524-153063, 636336-153466, 635916-153969, 636210-153946, 636951-155398, 636862-155291, 636430-154920, 637086-155019, 637364-155203, 637309-155066, 635959-153712, 635692-153888, 636175-154352, 635864-152650, 636760-152993, 636755-153128, 636691-155175, 636014-153136, 636687-154890, 635573-154461

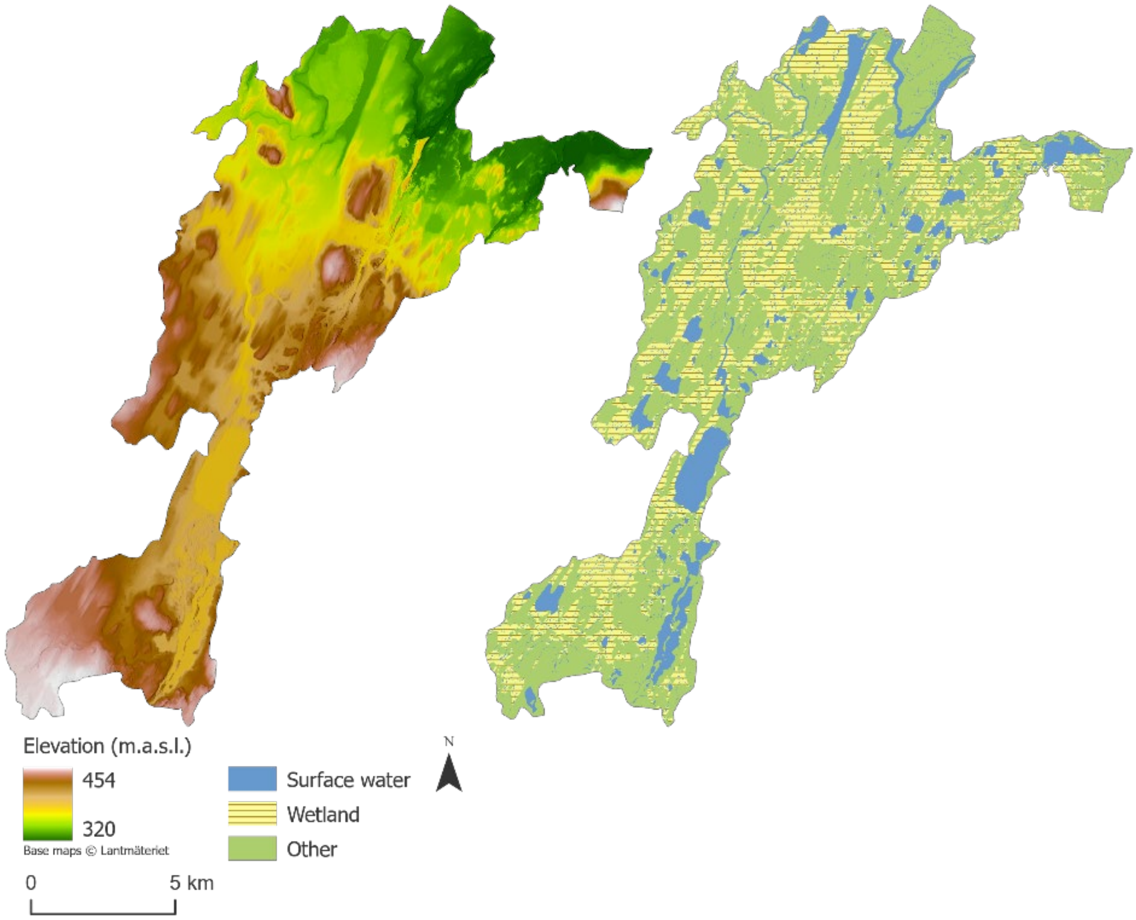

Figure A-1. Maps showing the DEM (left) and the surface water, wetland cover, and "other" land use (right) for Area 1: Karesuando. All land use data that is neither classified as surface water or wetland is classified as other.

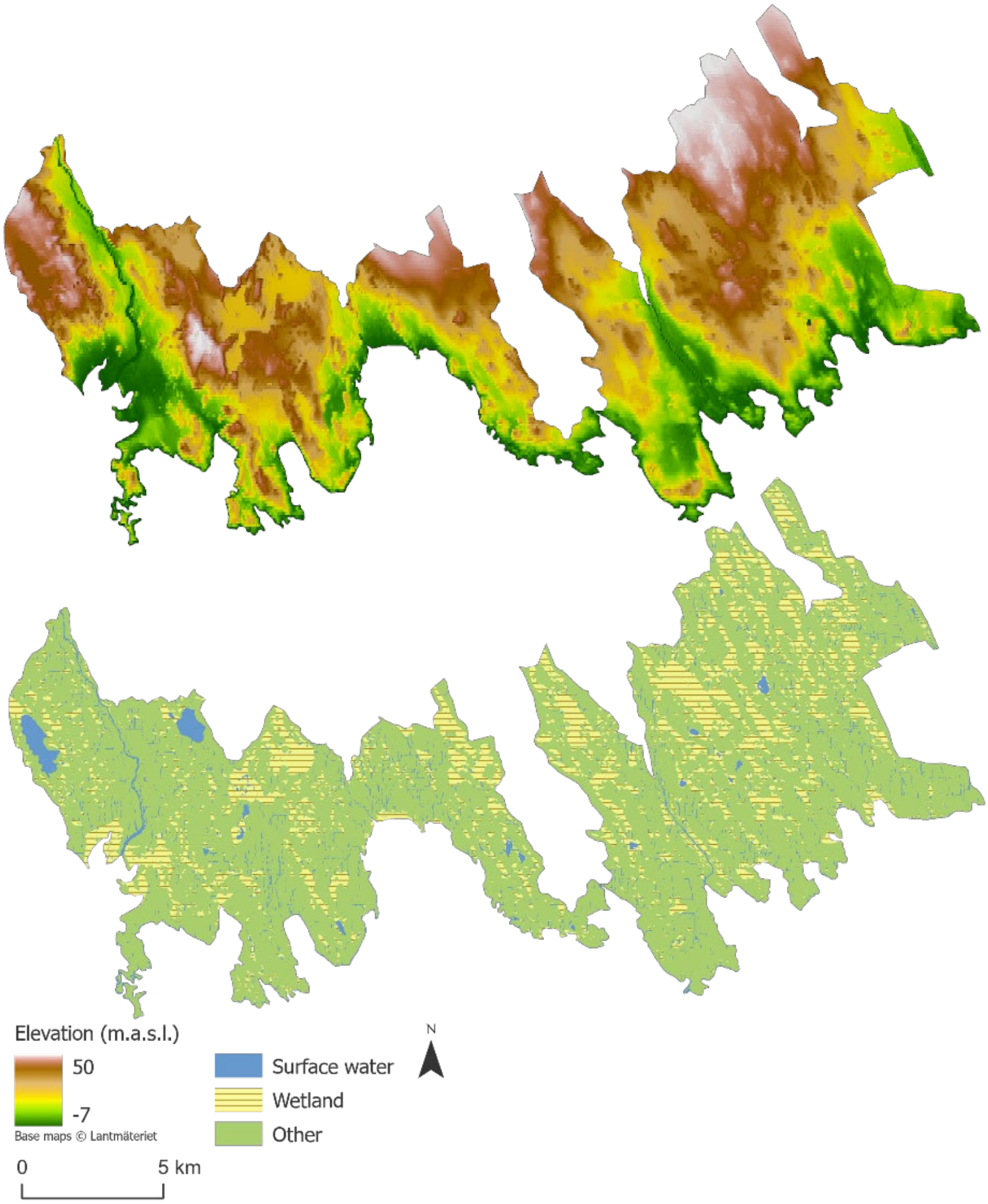

Figure A-2. Maps showing the DEM (top) and the surface water, wetland cover, and "other" land use (bottom) for Area 2: Sangis. All land use data that is neither classified as surface water or wetland is classified as other.

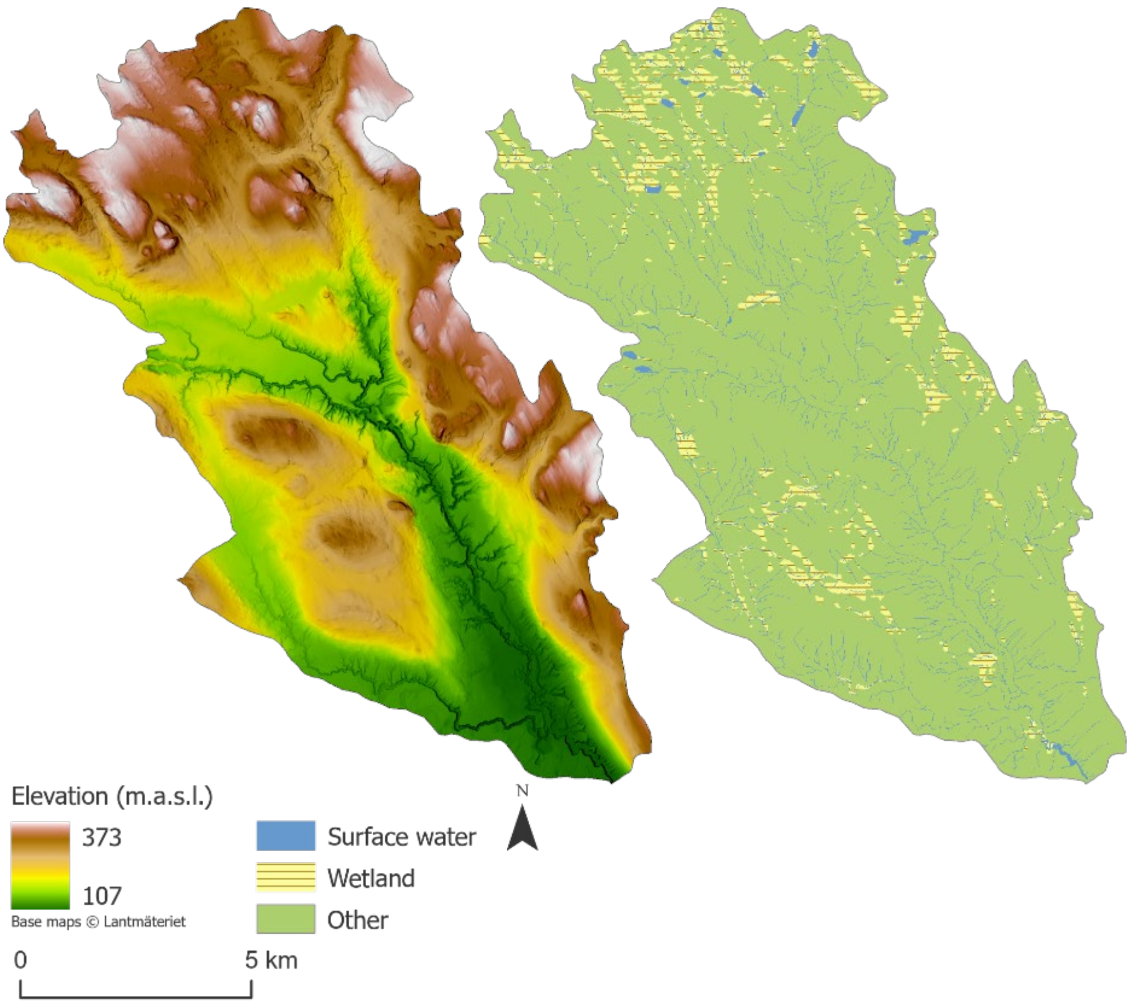

Figure A-3. Maps showing the DEM (left) and the surface water, wetland cover, and "other" land use (right) for Area 3: Krycklan. All land use data that is neither classified as surface water or wetland is classified as other.

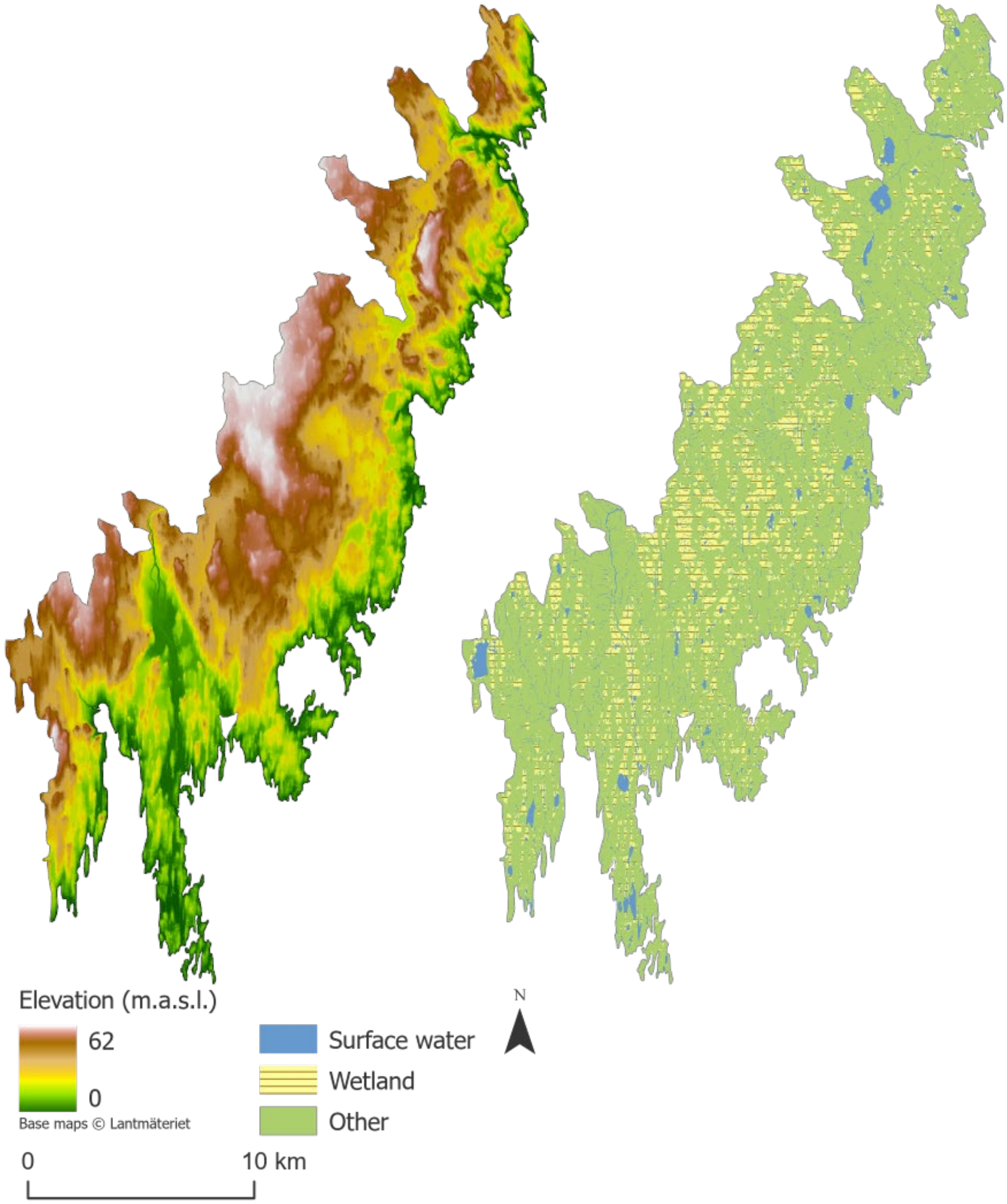

Figure A-4. Maps showing the DEM (left) and the surface water, wetland cover, and "other" land use (right) for Area 4: Savar. All land use data that is neither classified as surface water or wetland is classified as other.

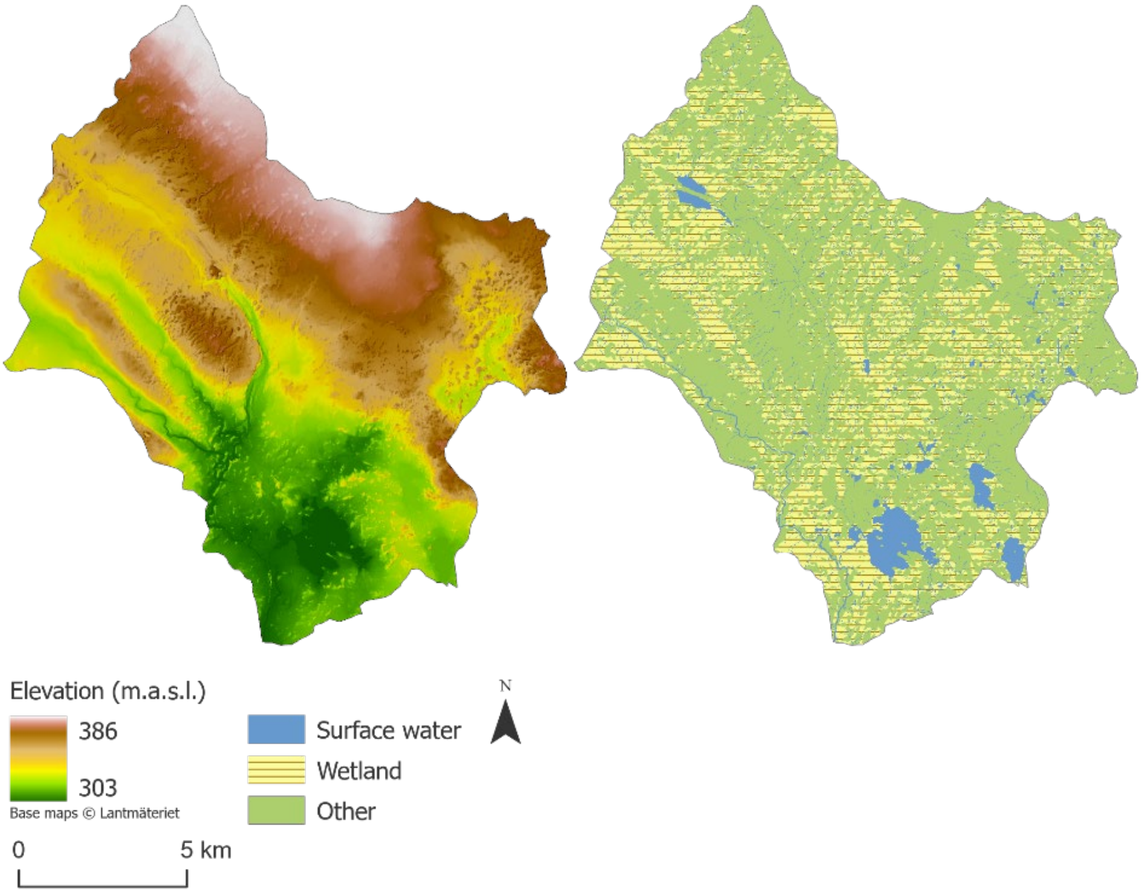

Figure A-5. Maps showing the DEM (left) and the surface water, wetland cover, and "other" land use (right) for Area 5: Hammerdal. All land use data that is neither classified as surface water or wetland is classified as other.

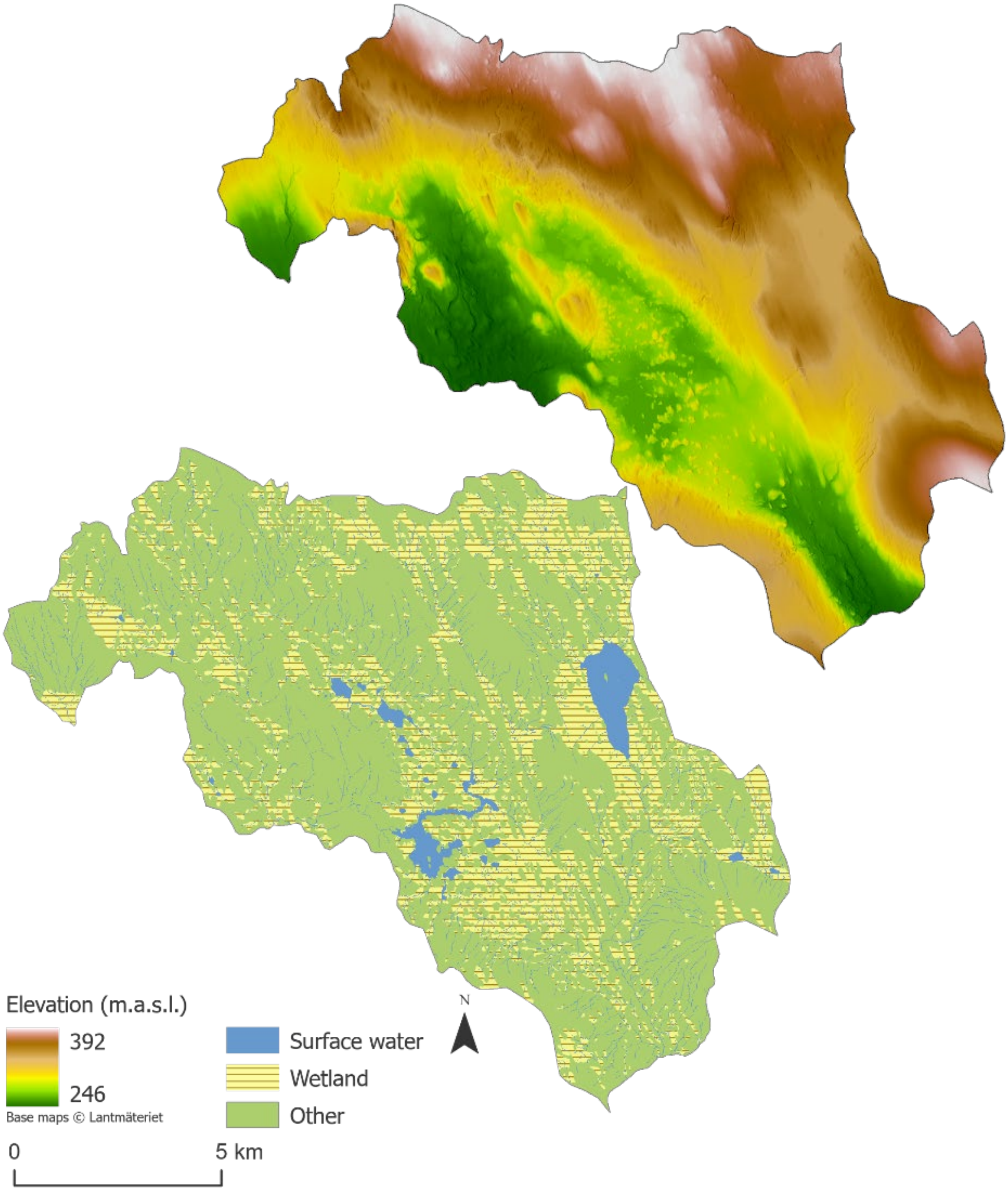

Figure A-6. Maps showing the DEM (top) and the surface water, wetland cover, and "other" land use (bottom) for Area 6: Orsa. All land use data that is neither classified as surface water or wetland is classified as other.

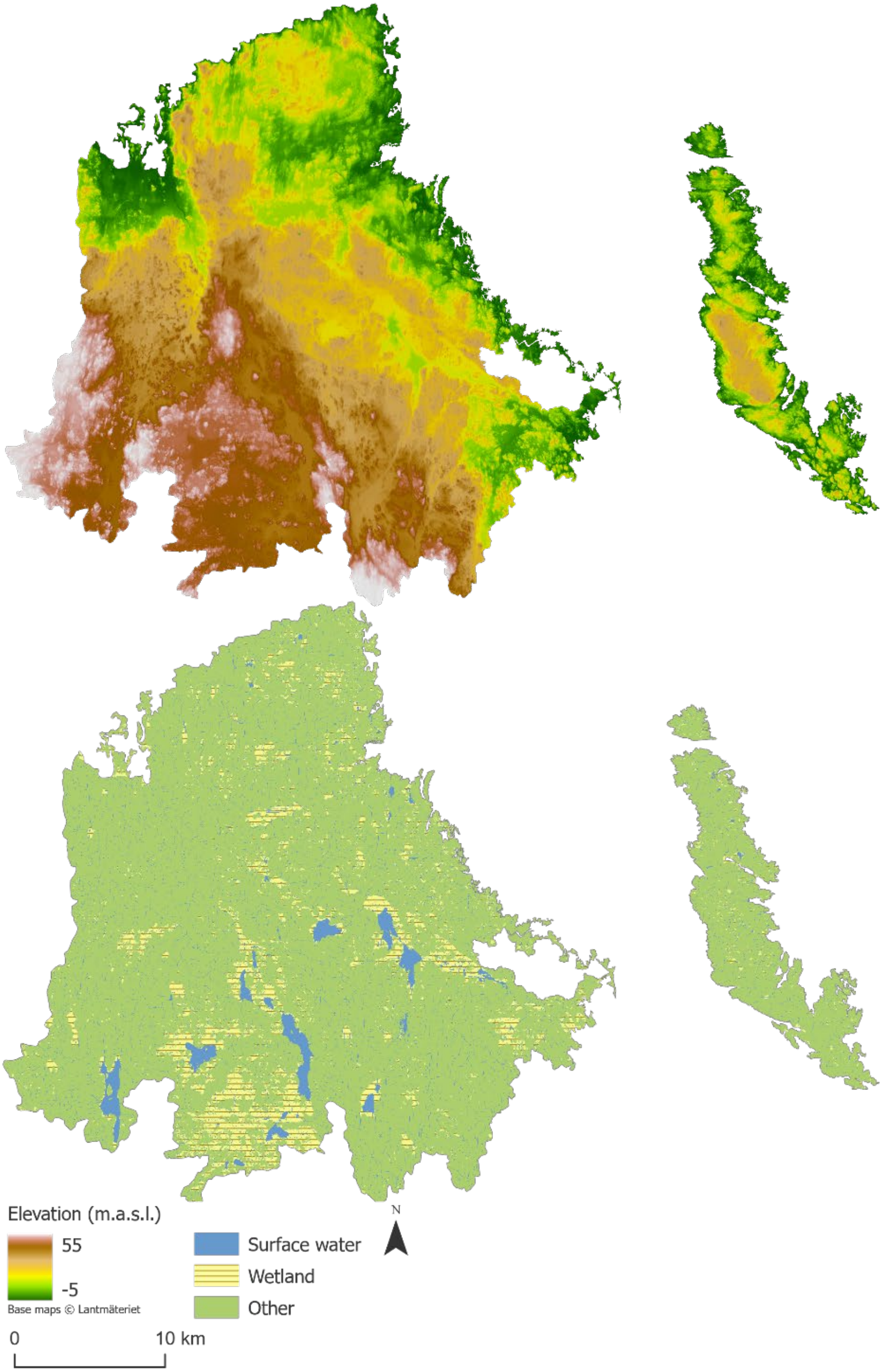

Figure A-7. Maps showing the DEM (top) and the surface water, wetland cover, and "other" land use (bottom) for Areas 7 and 8: Forsmark and Gräsö. All land use data that is neither classified as surface water or wetland is classified as other. Note that the Forsmark validation area is excluded from the maps.

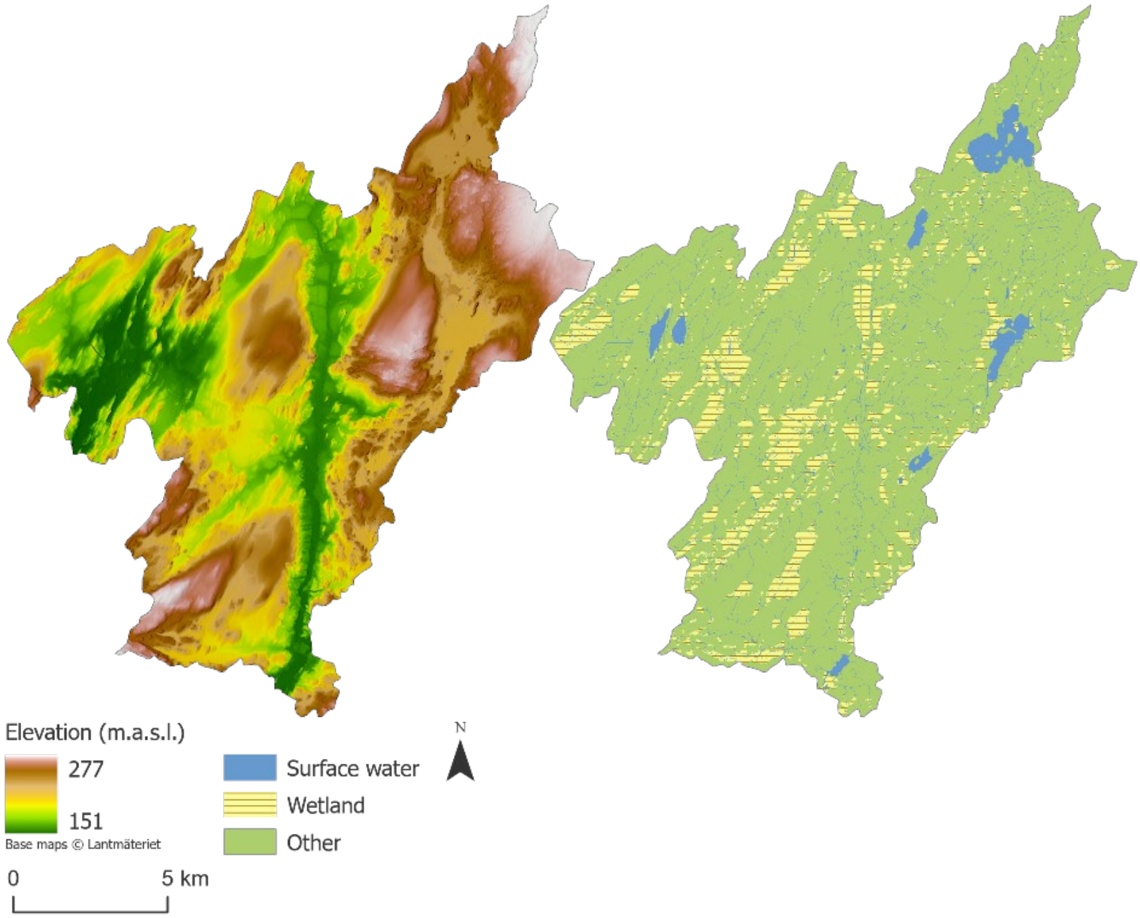

Figure A-8. Maps showing the DEM (left) and the surface water, wetland cover, and "other" land use (right) for Area 9: Tranemo. All land use data that is neither classified as surface water or wetland is classified as other.

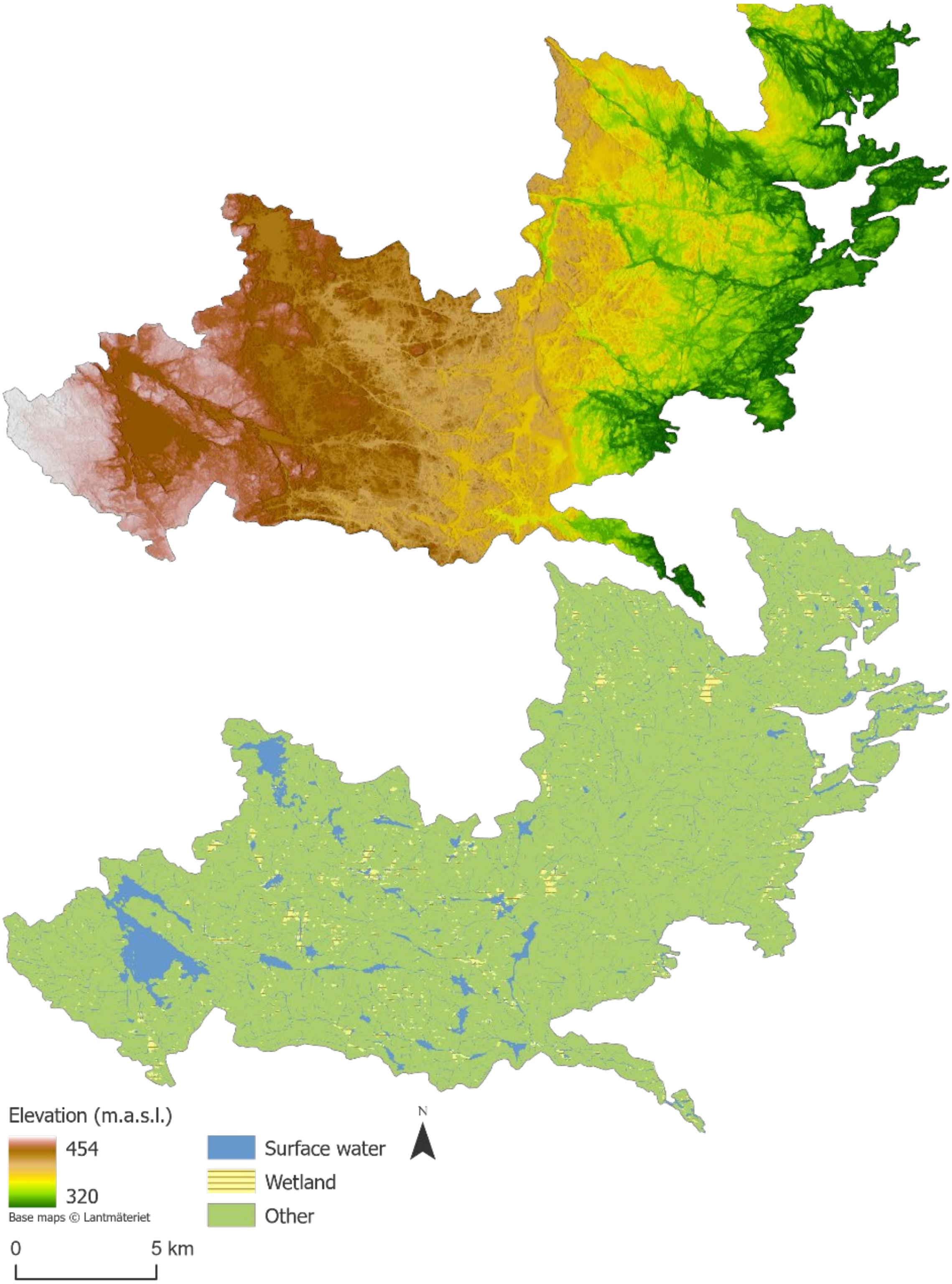

Figure A-9. Maps showing the DEM (top) and the surface water, wetland cover, and "other" land use (bottom) for Area 10: Simpevarp. All land use data that is neither classified as surface water or wetland is classified as other.

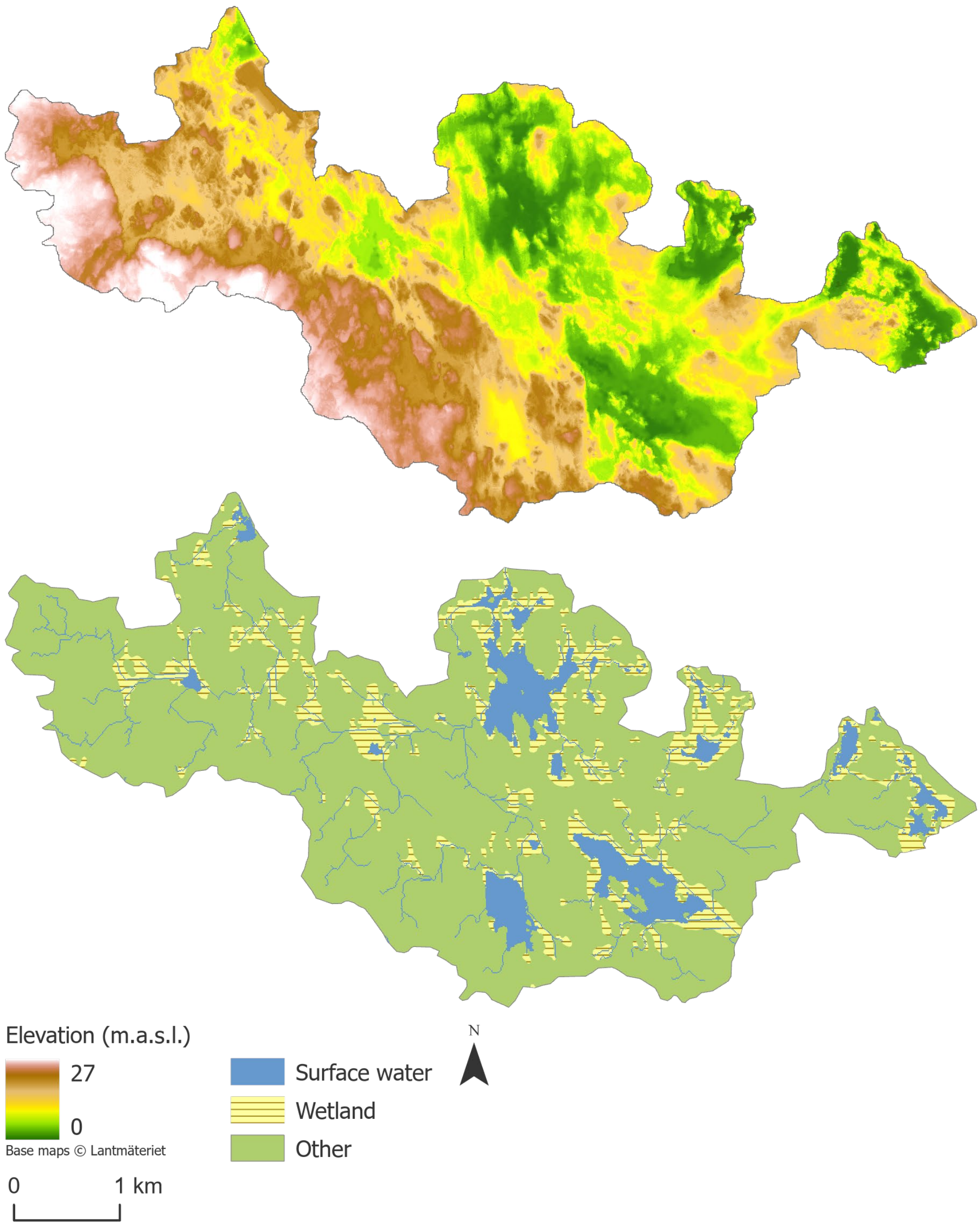

Figure A-10. Maps showing the DEM (top) and the surface water, wetland cover, and "other" land use (bottom) for the Forsmark validation area. All land use data that is neither classified as surface water or wetland is classified as other.

Appendix B SGU soil map of Forsmark and Gräsö

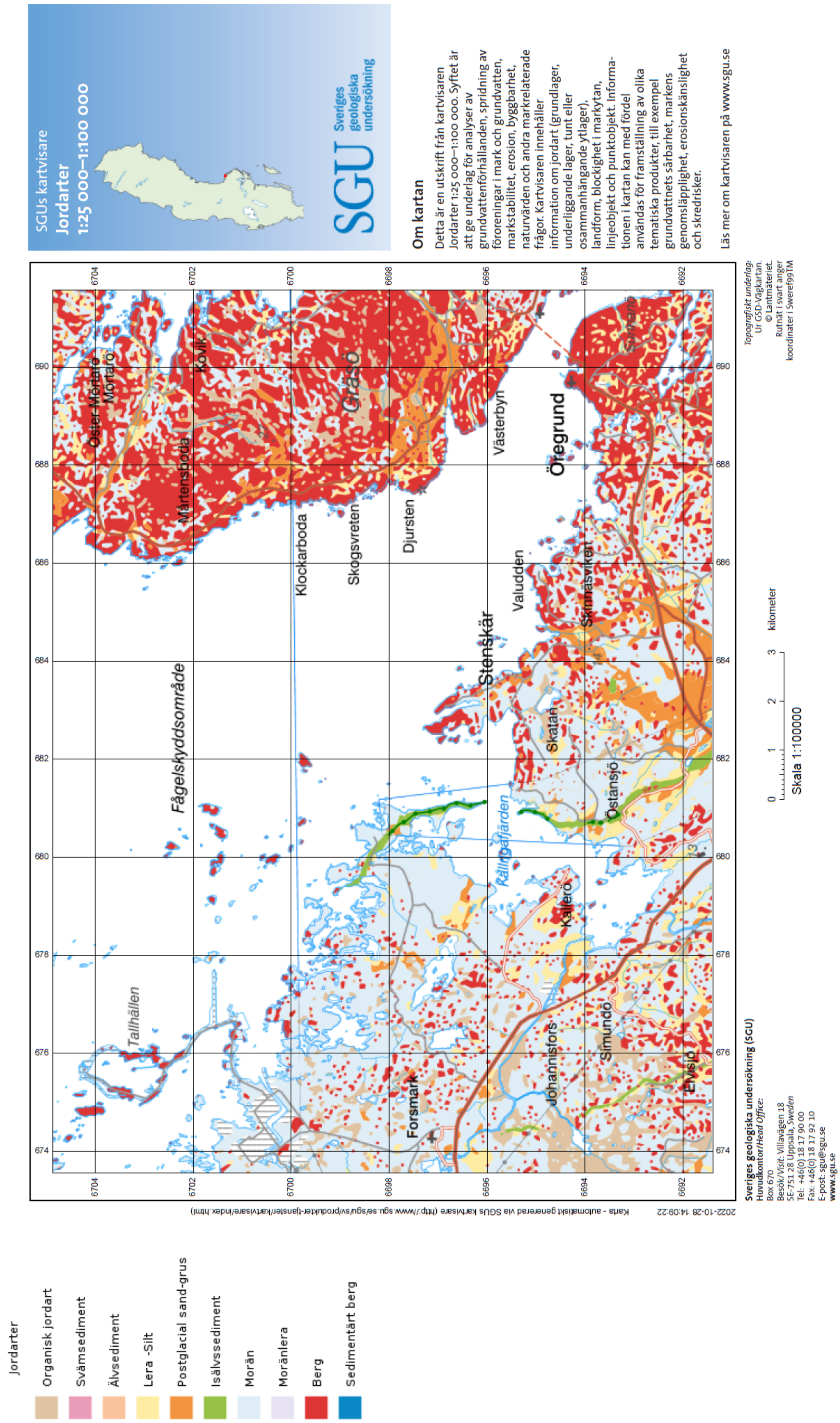

Figure B-1. SGU Surface soil map of portions of Forsmark and Gräsö.

Appendix C SMHI temperature and precipitation statistics for Sweden

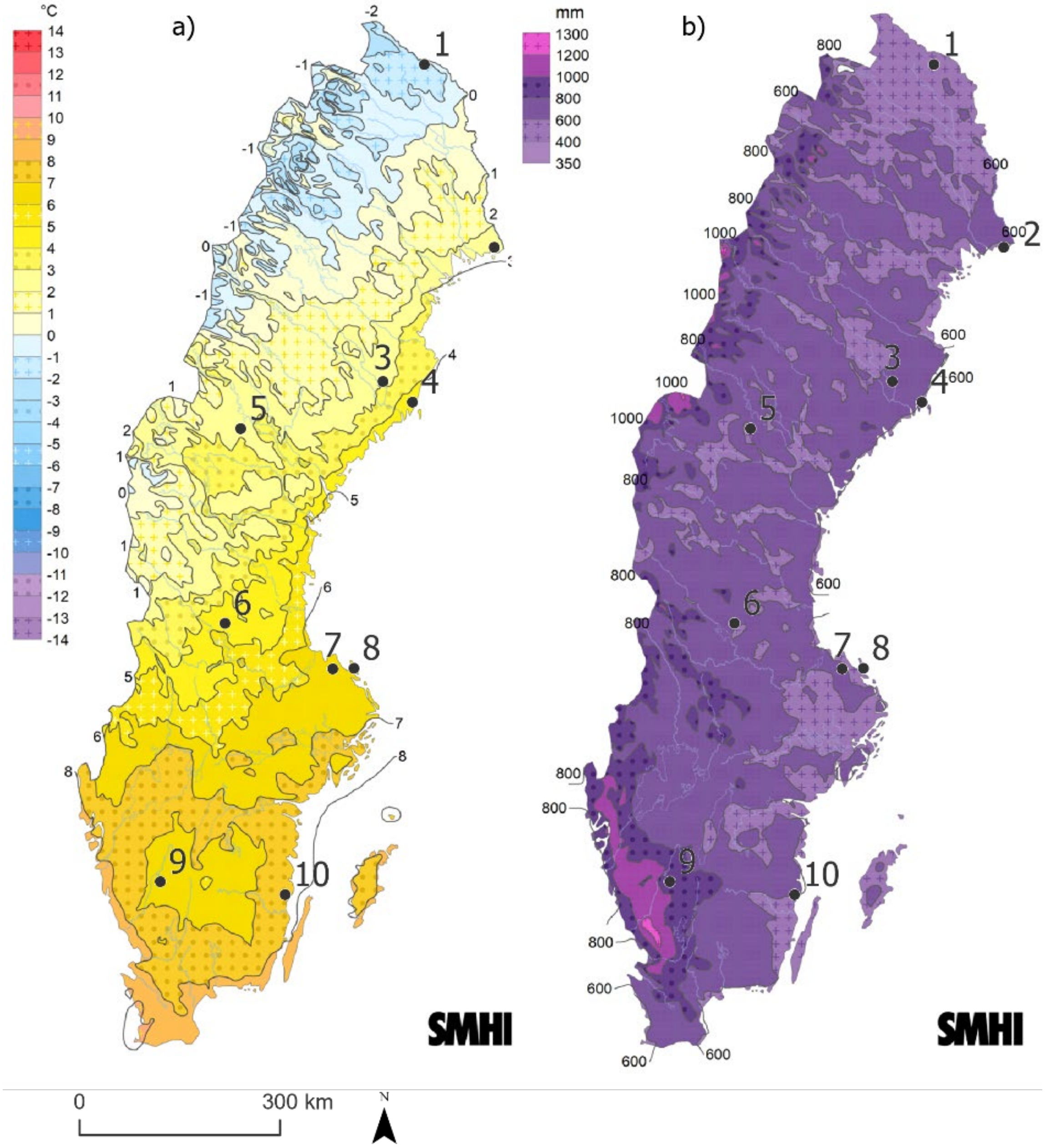

Figure C-1. Average temperature and yearly precipitation in for Sweden for the period 1991 – 2020

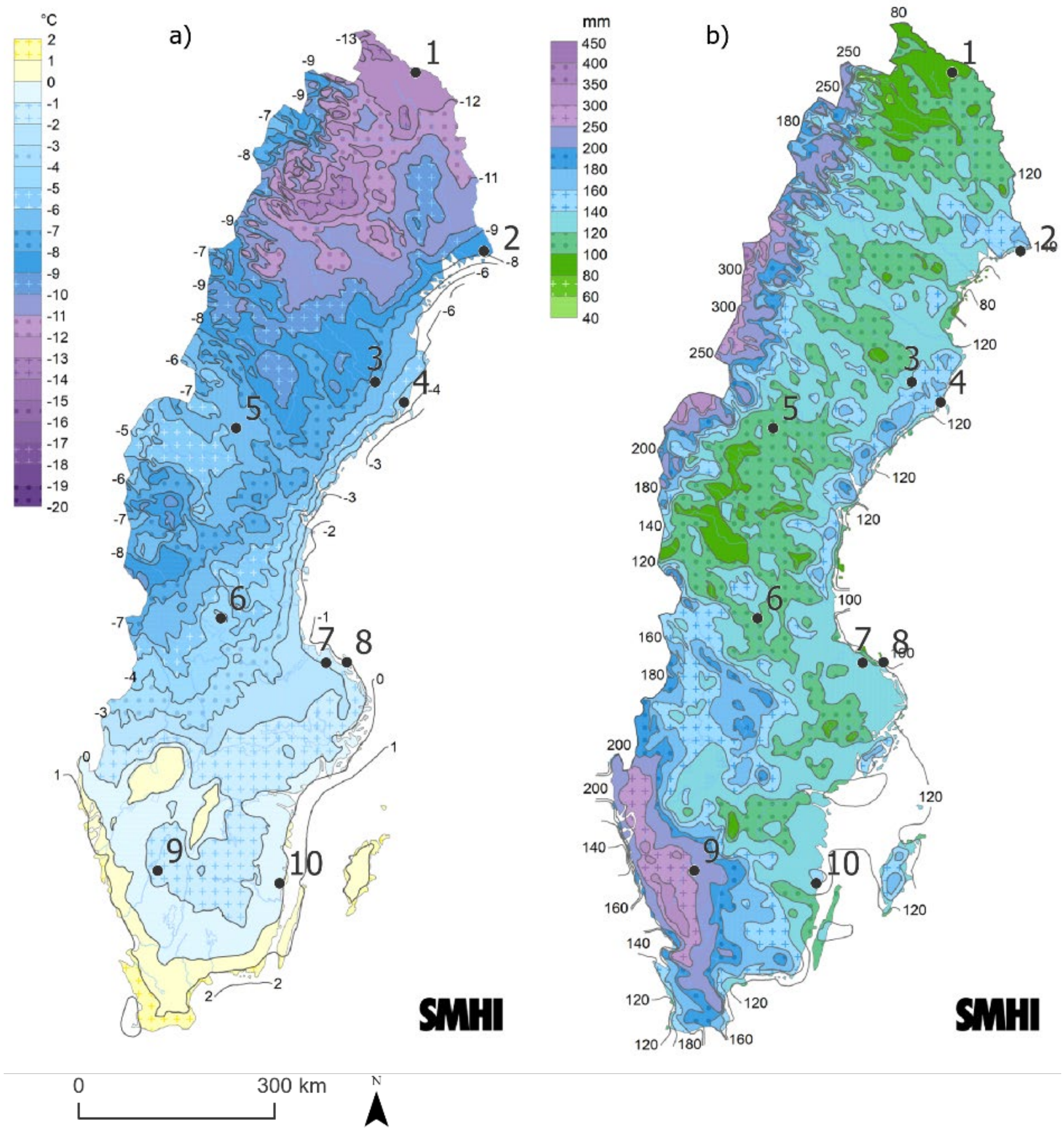

Figure C-2. Average winter temperature and winter precipitation for Sweden for the period 1991 – 2020

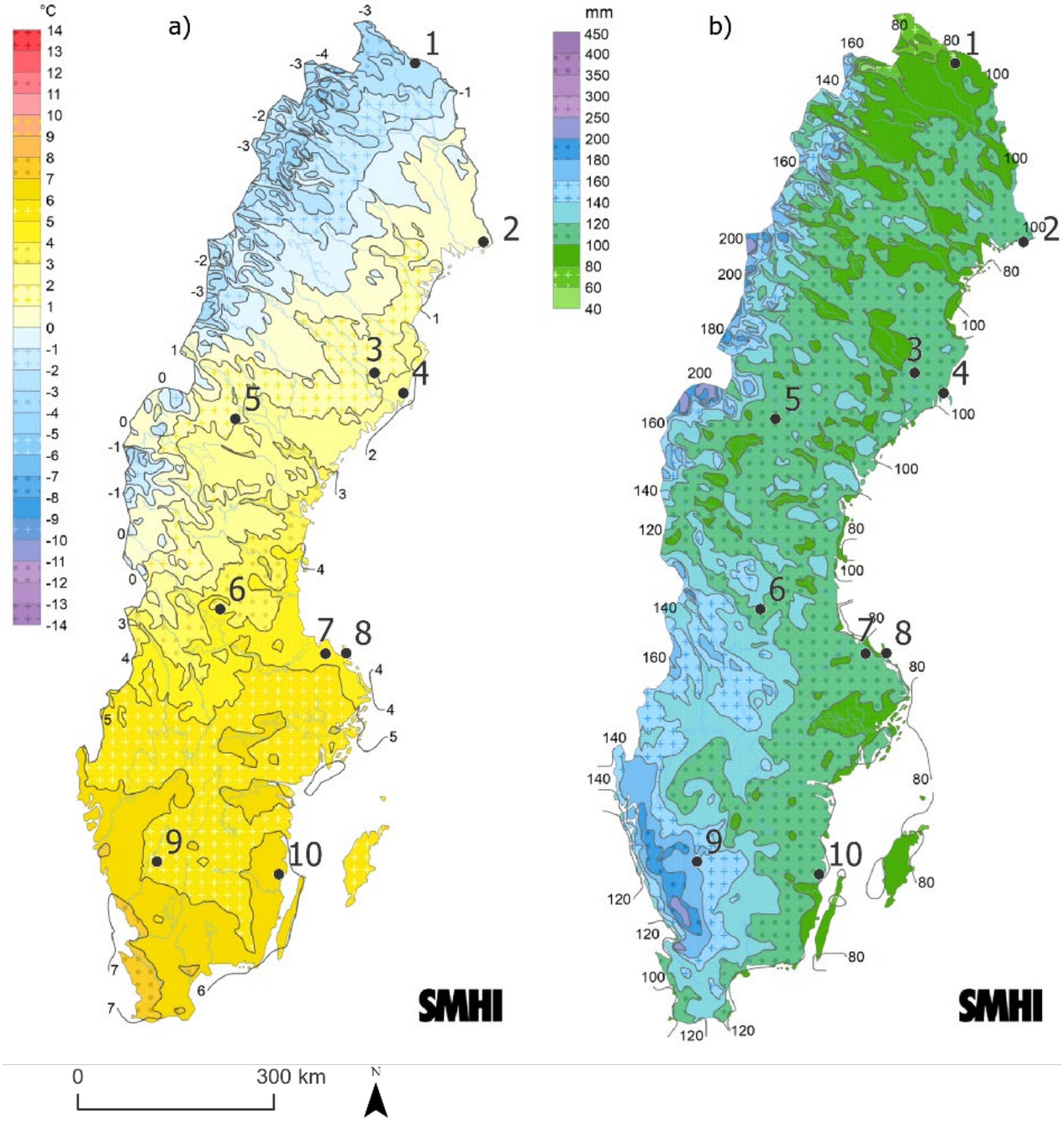

Figure C-3. Average spring temperature and spring precipitation for Sweden for the period 1991 – 2020

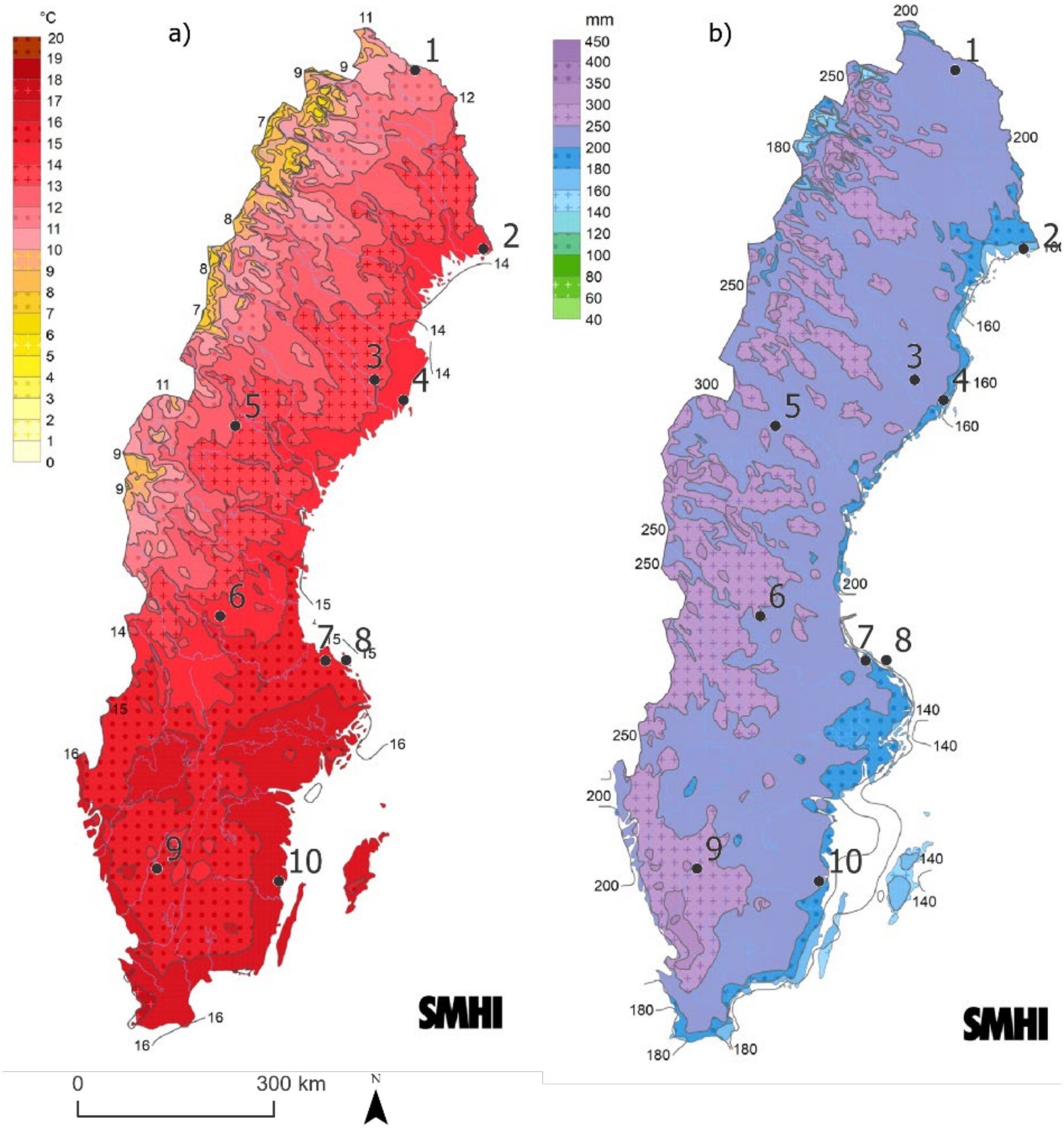

Figure C-4. Average summer temperature and summer precipitation for Sweden for the period 1991 – 2020

Appendix D WIM Forsmark 1.1 algorithm names and details

Table D-1. Areas and their respective algorithm names and file-sizes. All algorithms have the file-type "JOBLIB"

Area	Reference location	Algorithm name	Size (KB)
1	Karesuando	Train_model_1_1	2544698
2	Sangis	Train_model_2_1	1978310
3	Krycklan	Train_model_3_1	497844
4	Norum	Train_model_4_1	2568630
5	Hammerdal	Train_model_5_1	2637563
6	Skattungbyn	Train_model_6_1	1942300
7	Forsmark	Train_model_7_1	4288116
8	Gräsö	Train_model_8_1	91274
9	Tranemo	Train_model_9_1	1289532
10	Simpevarp	Train_model_10_1	355962

All algorithms for WIM Forsmark 1.1 are stored on SVN at the following address:
svn://svn.skb.se/projekt/Otherprojects/Landscape/WIM Forsmark/WIM Forsmark 1.1

Appendix E Wetland prediction for individual algorithms in WIM Forsmark 1.1

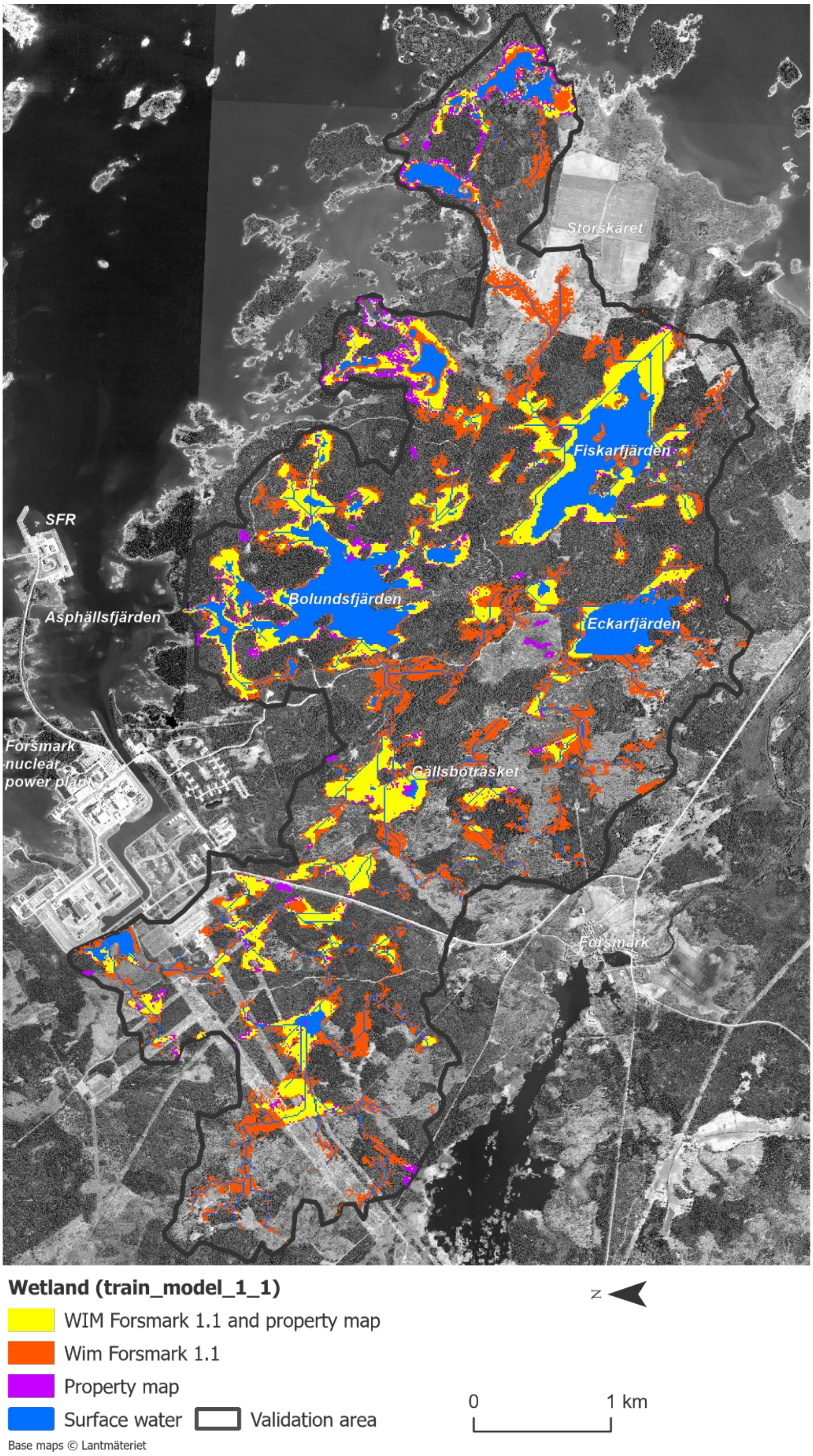

Figure E-1. Wetland predictions for algorithm trained using data from Area 1: Karesuando (algorithm name "Train_model_1_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

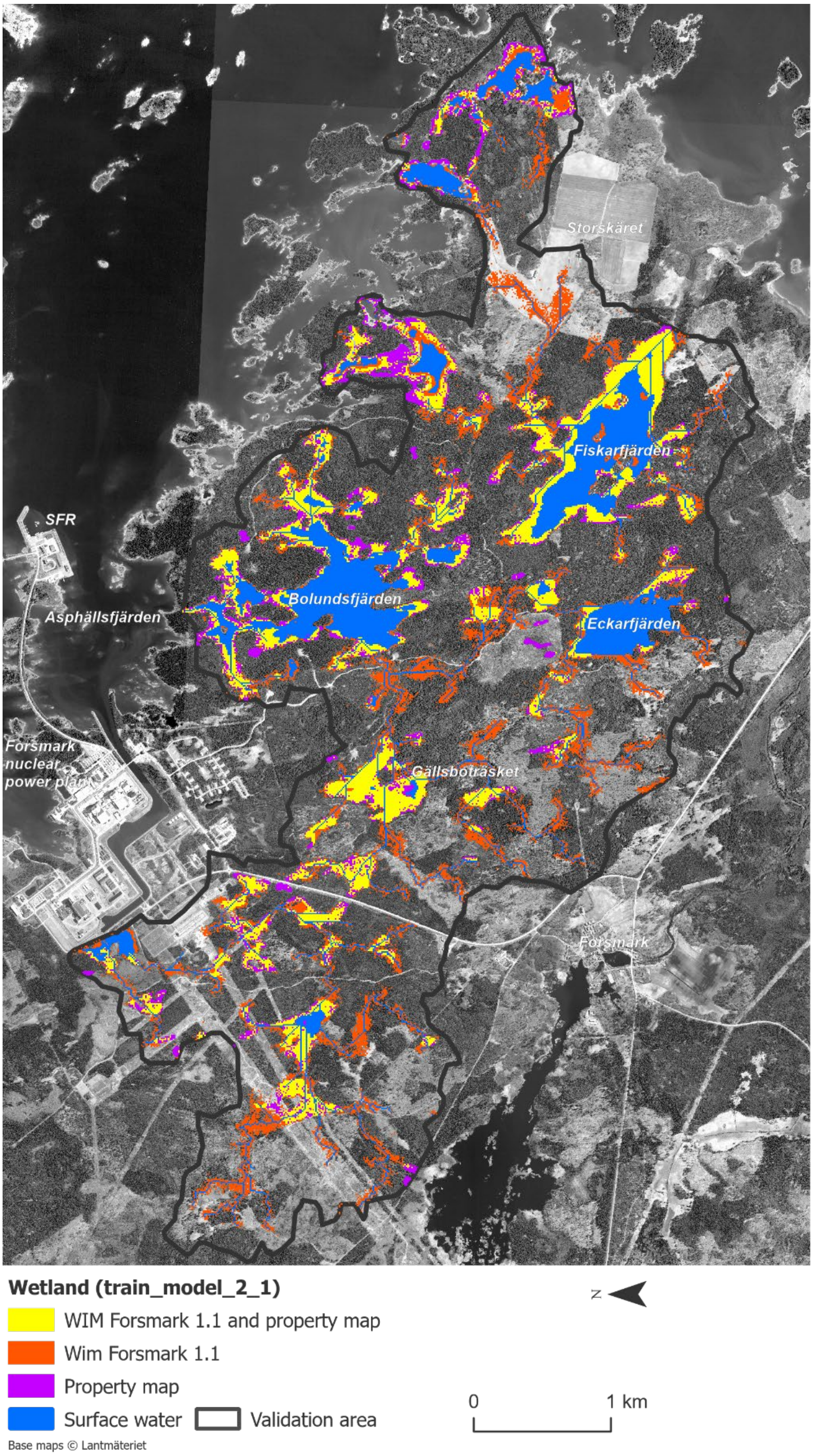

Figure E-2. Wetland predictions for algorithm trained using data from Area 2: Sangis (algorithm name "Train_model_2_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

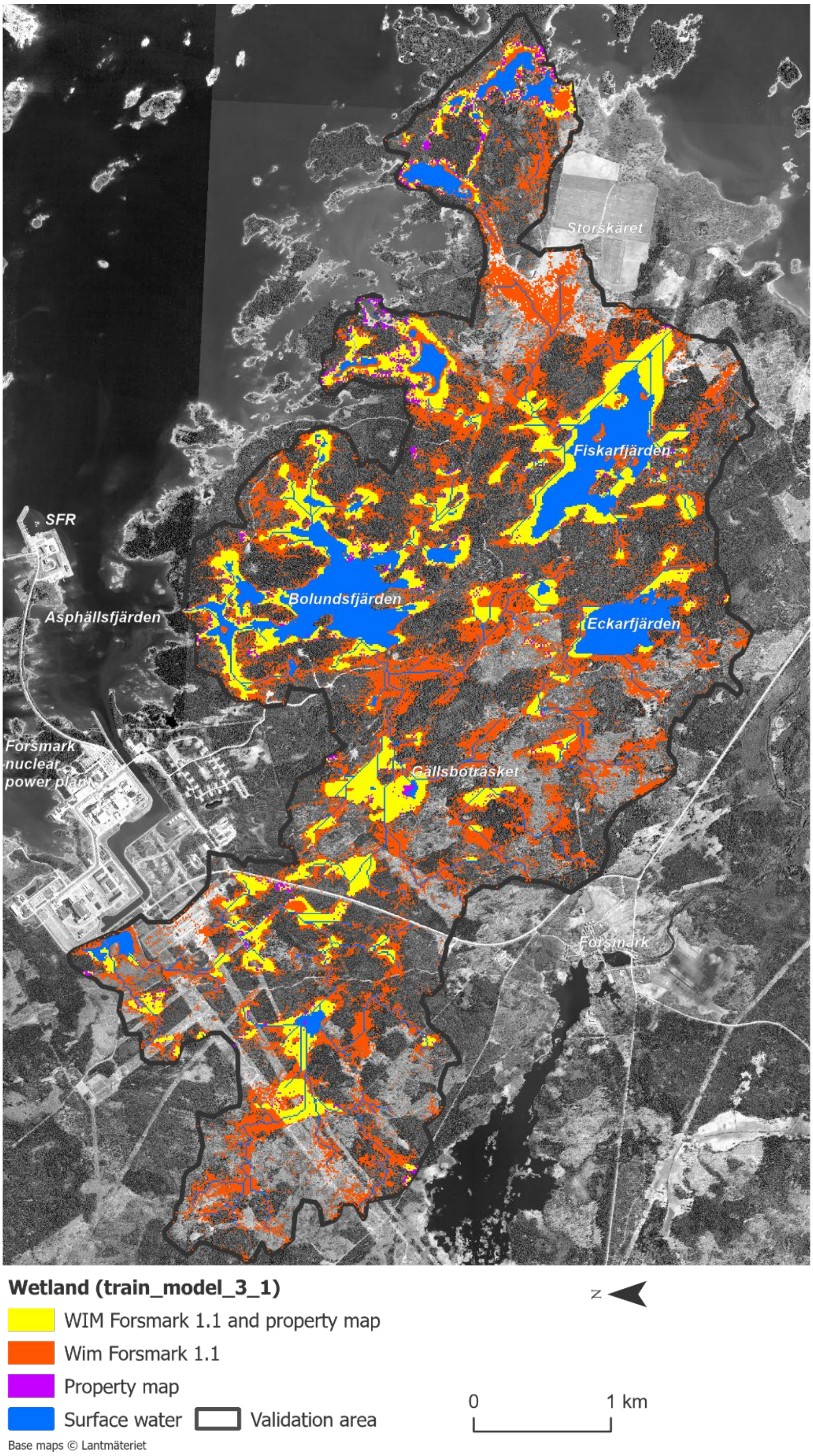

Figure E-3. Wetland predictions for algorithm trained using data from Area 3: Krycklan (algorithm name "Train_model_3_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

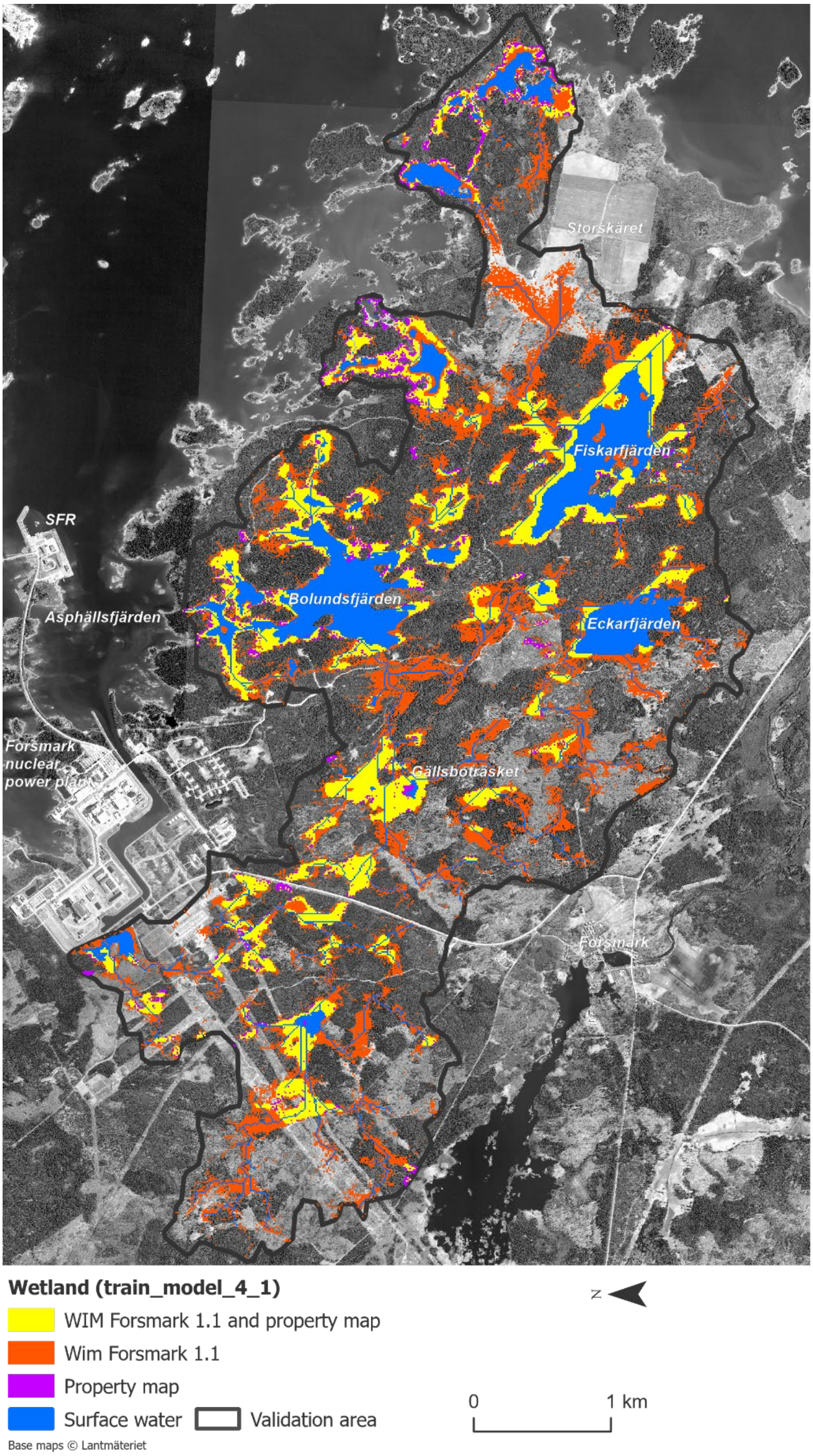

Figure E-4. Wetland predictions for algorithm trained using data from Area 4: Norum (algorithm name "Train_model_4_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

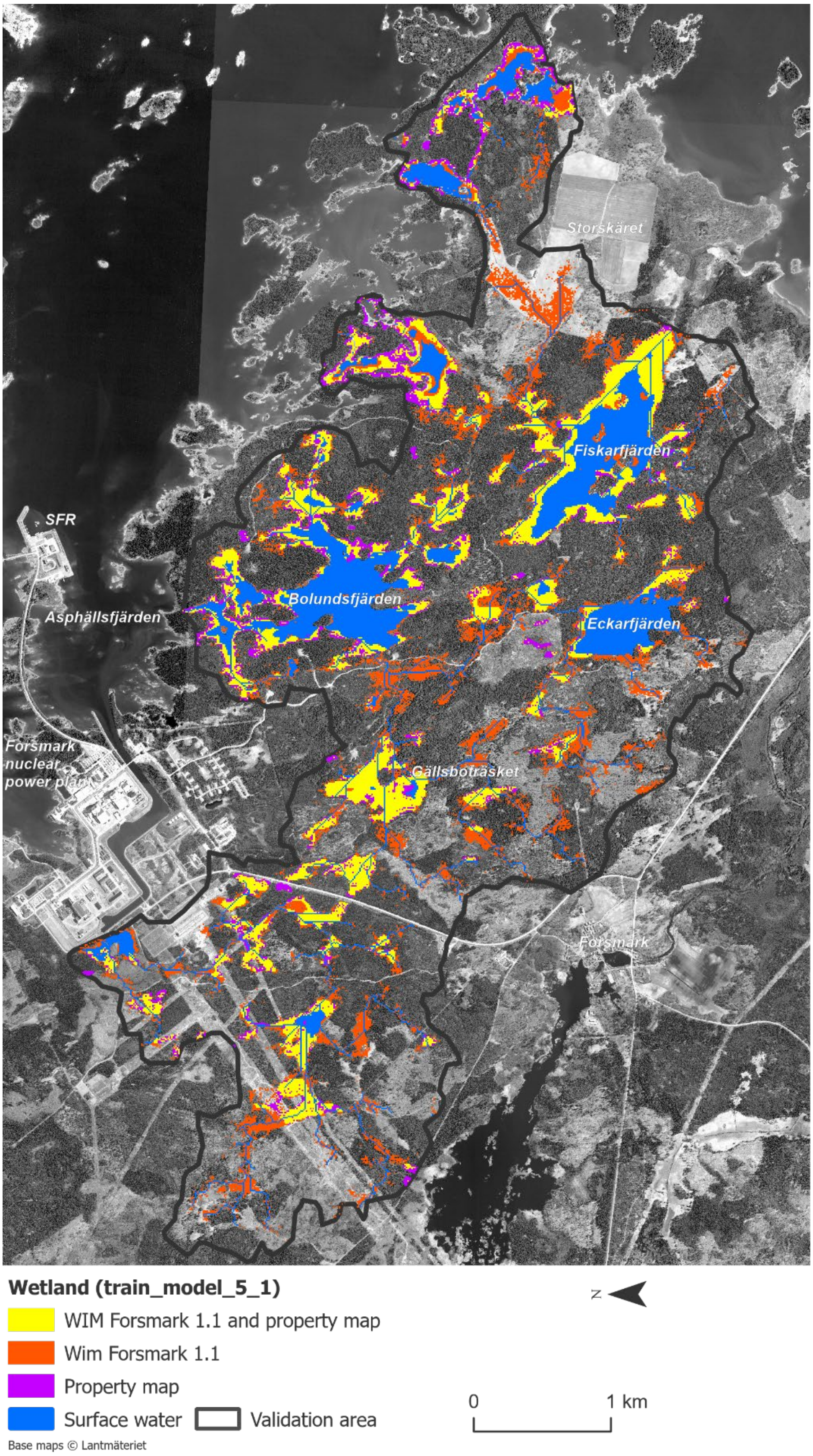

Figure E-5. Wetland predictions for algorithm trained using data from Area 5: Hammerdal (algorithm name "Train_model_5_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

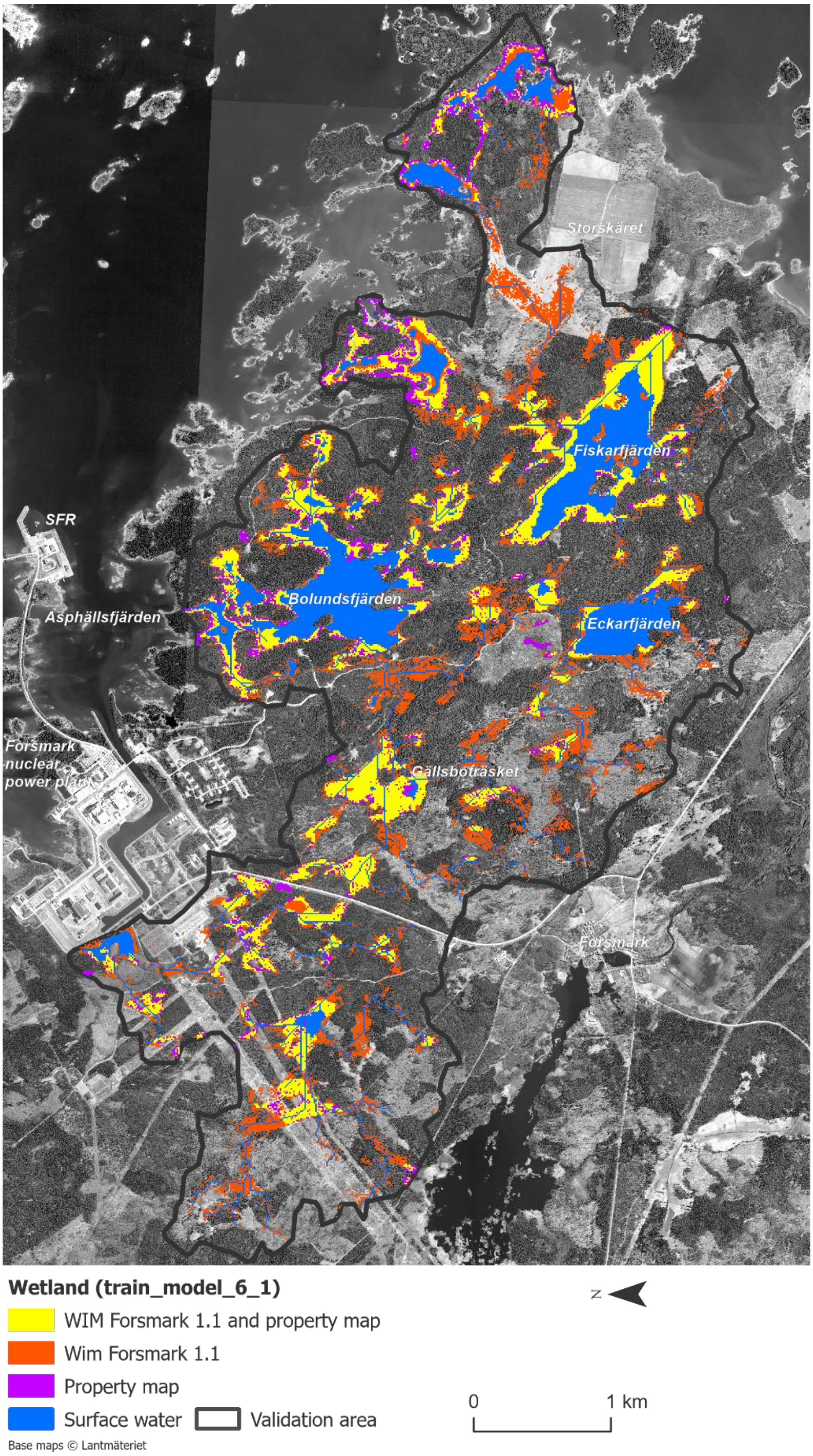

Figure E-6. Wetland predictions for algorithm trained using data from Area 6: Skattungbyn (algorithm name "Train_model_6_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4

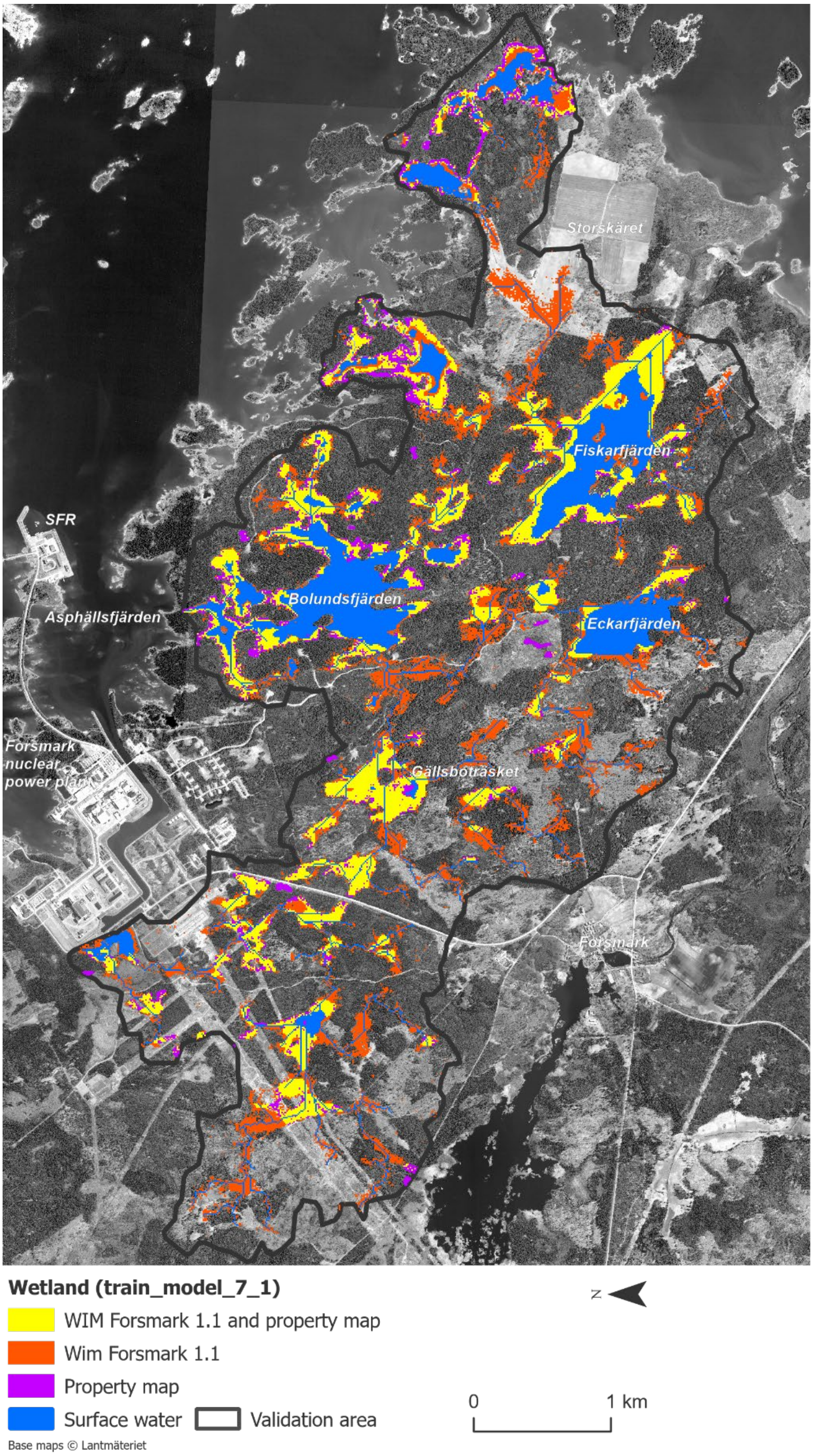

Figure E-7. Wetland predictions for algorithm trained using data from Area 7: Forsmark (algorithm name "Train_model_7_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

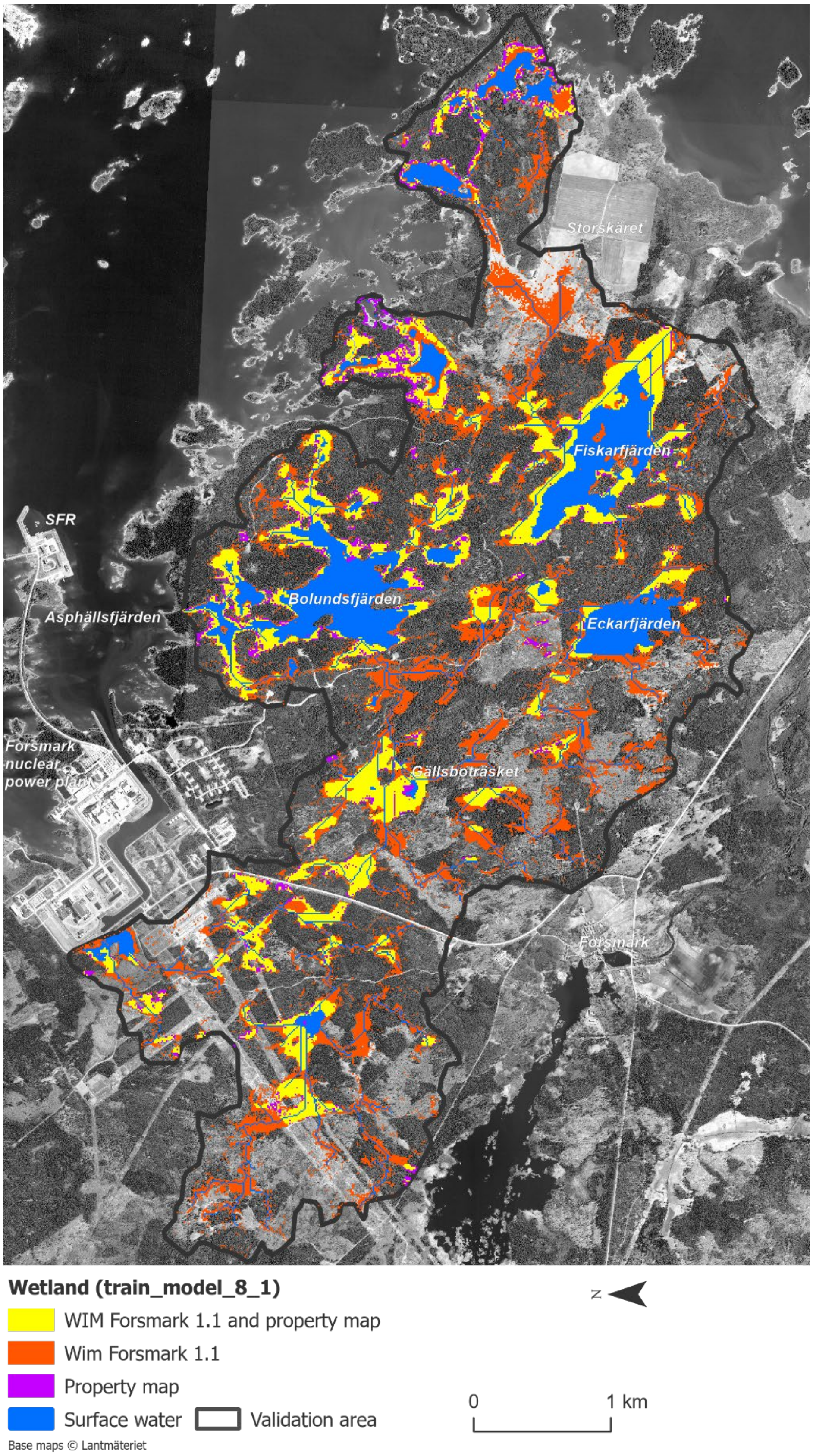

Figure E-8. Wetland predictions for algorithm trained using data from Area 8: Gräsö (algorithm name "Train_model_8_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

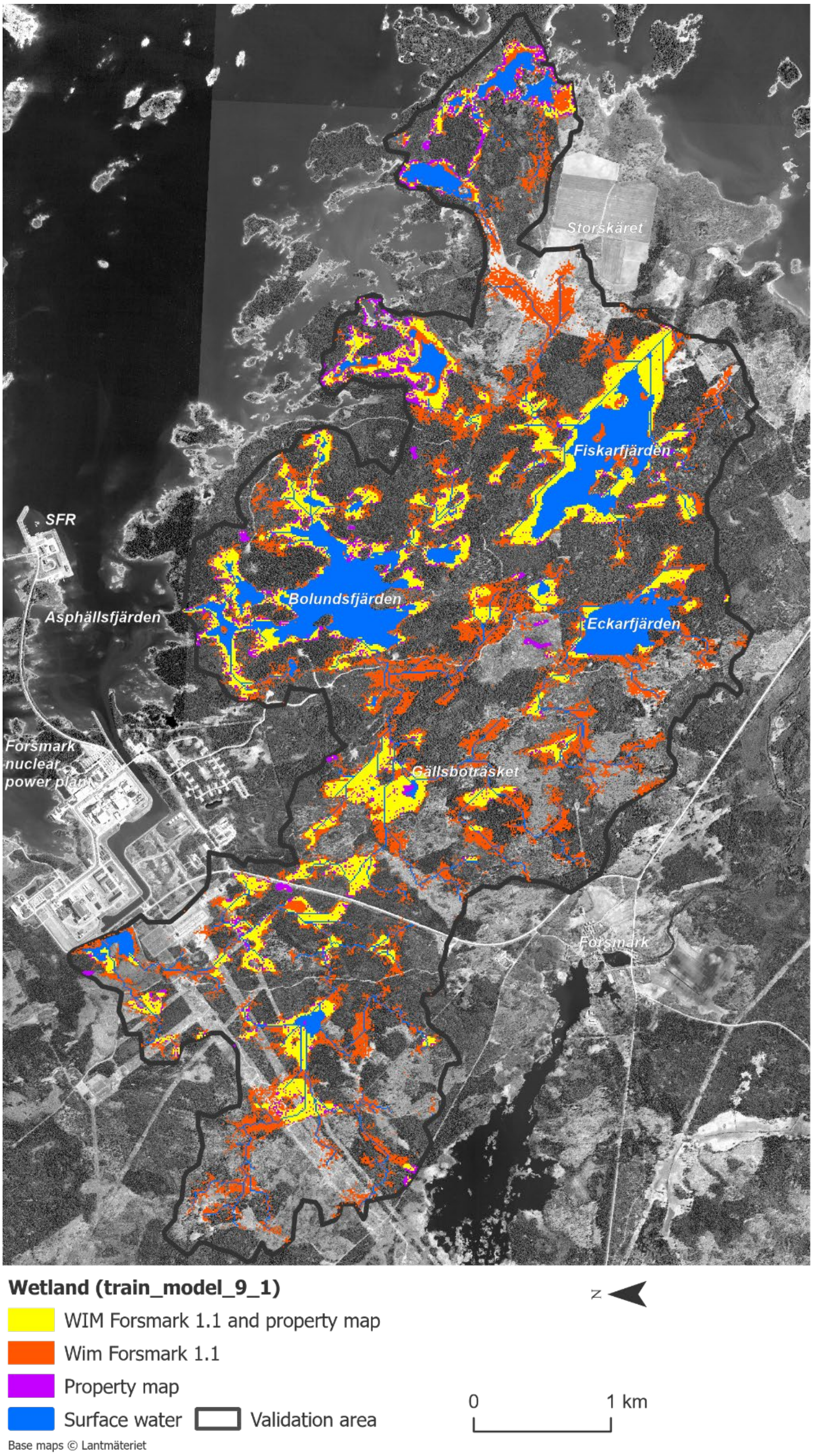

***Figure E-9.** Wetland predictions for algorithm trained using data from Area 9: Tranemo (algorithm name "Train_model_9_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.*

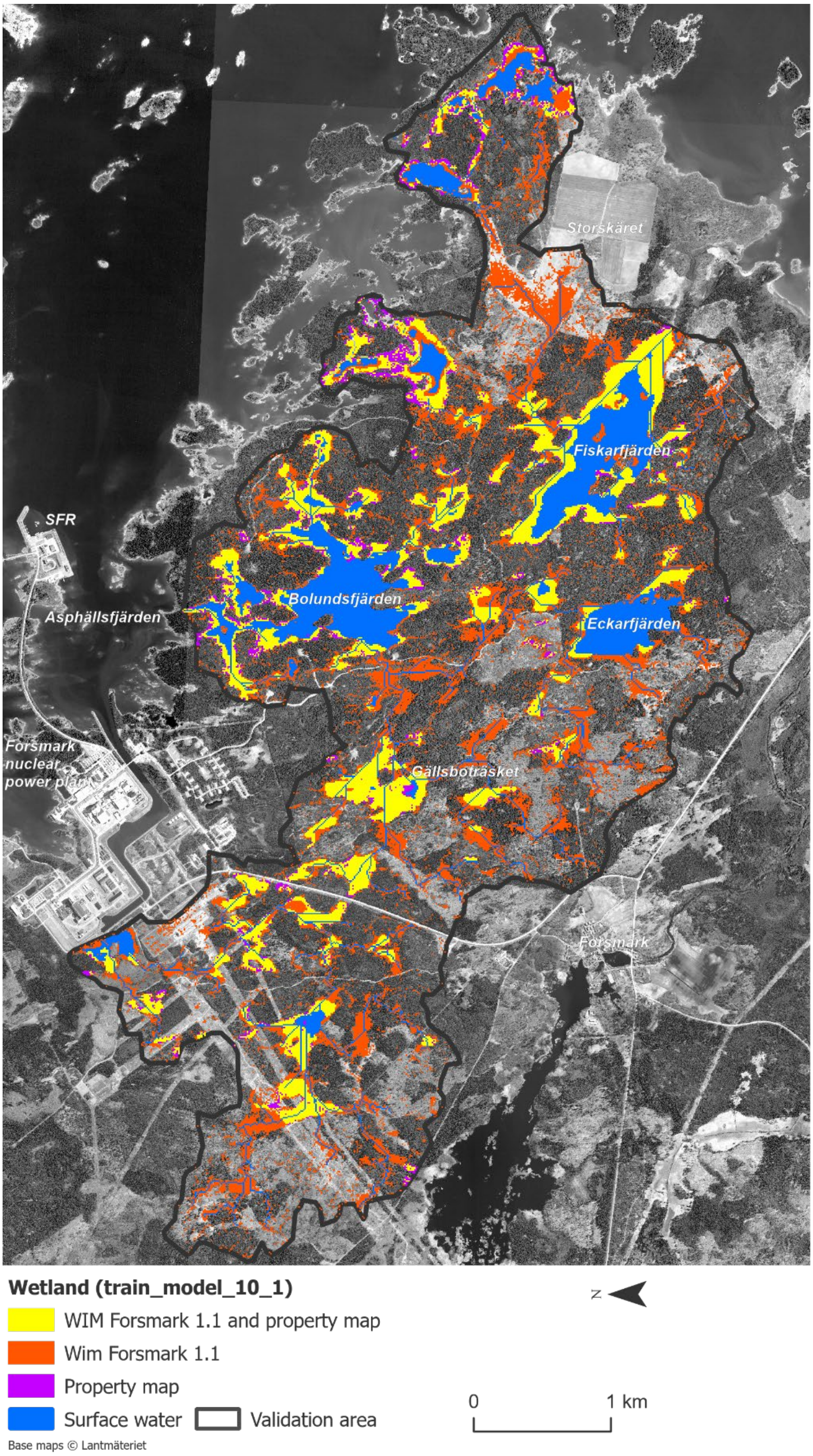

Figure E-10. Wetland predictions for algorithm trained using data from Area 10: Simpevarp (algorithm name "Train_model_10_1") and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

Appendix F Wetland prediction “hitmap” for the Forsmark validation area

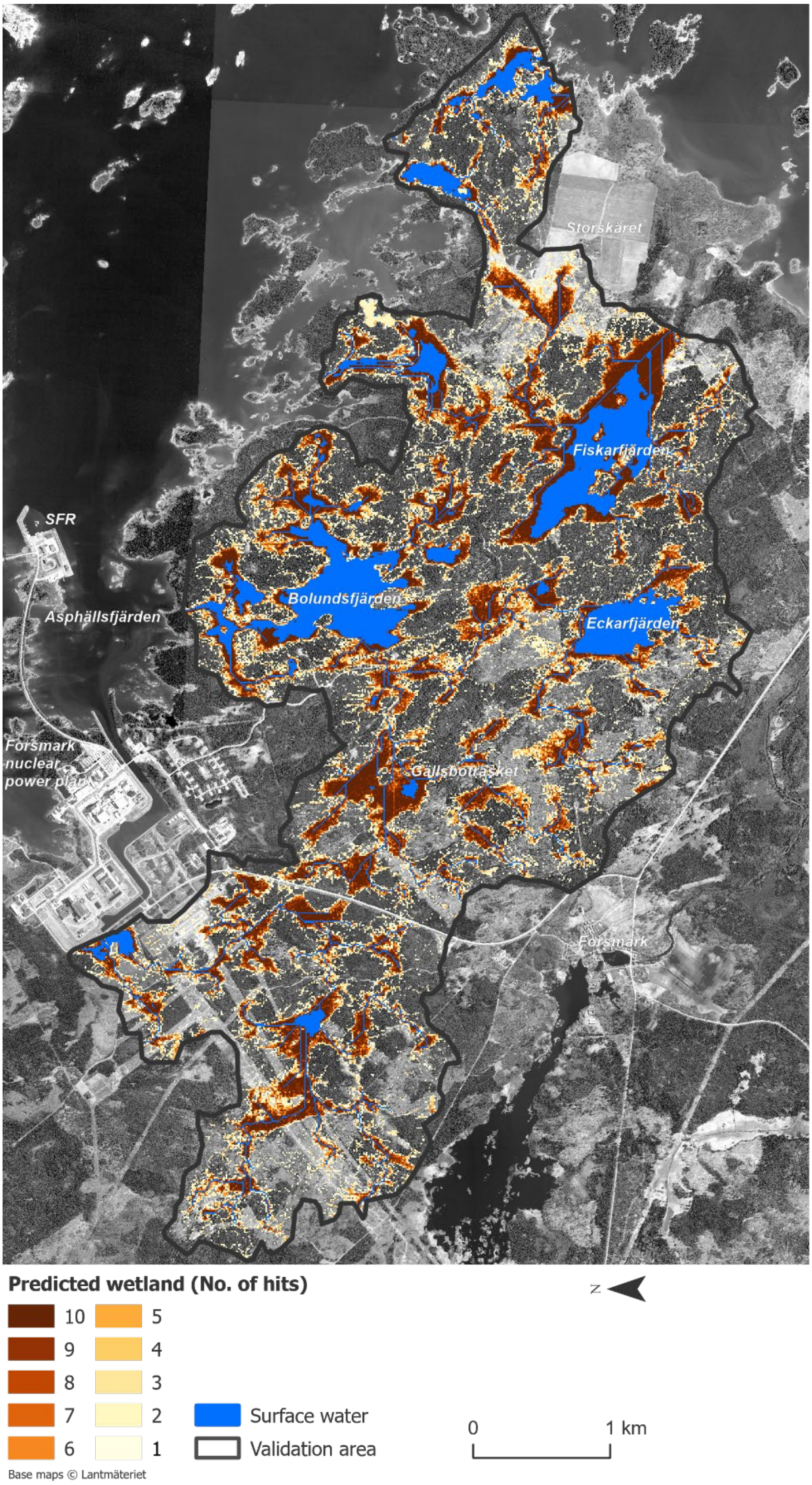

Figure F-1. Wetland predictions for the Forsmark validation area; results are presented as a hitmap with each “hit” representing a cell predicted to contain a wetland according to one of the 10 individual algorithms that make up WIM Forsmark 1.1. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.

Appendix G Amalgamated predictions for Forsmark validation area

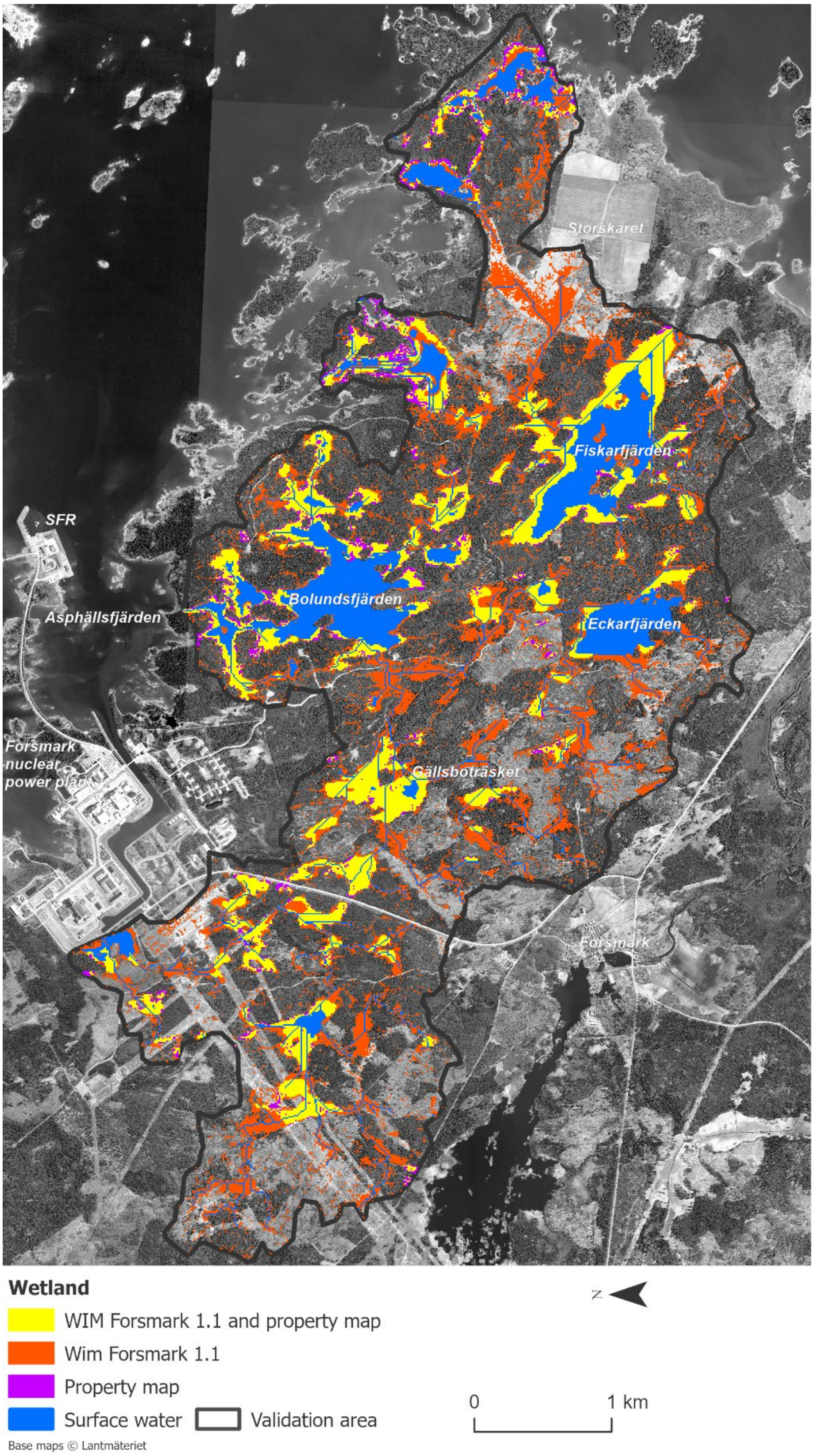

Figure G-1. Wetland predictions, presented as the amalgamated results from the 10 individual algorithms included in WIM Forsmark 1.1, and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4. Relative the total wetland area (i.e. sum of the predicted wetland area and the observed wetlands), overlap between the predicted and observed wetlands accounts for 66.9% of the total wetland area, predicted wetlands that do not overlap with an observed wetland accounts for 31.6% of the total wetland area and observed wetlands that were not predicted accounts for 1.6% of the total area.

Appendix H Amalgamated predictions for Forsmark validation area and wetland predictions from SLU wetness map

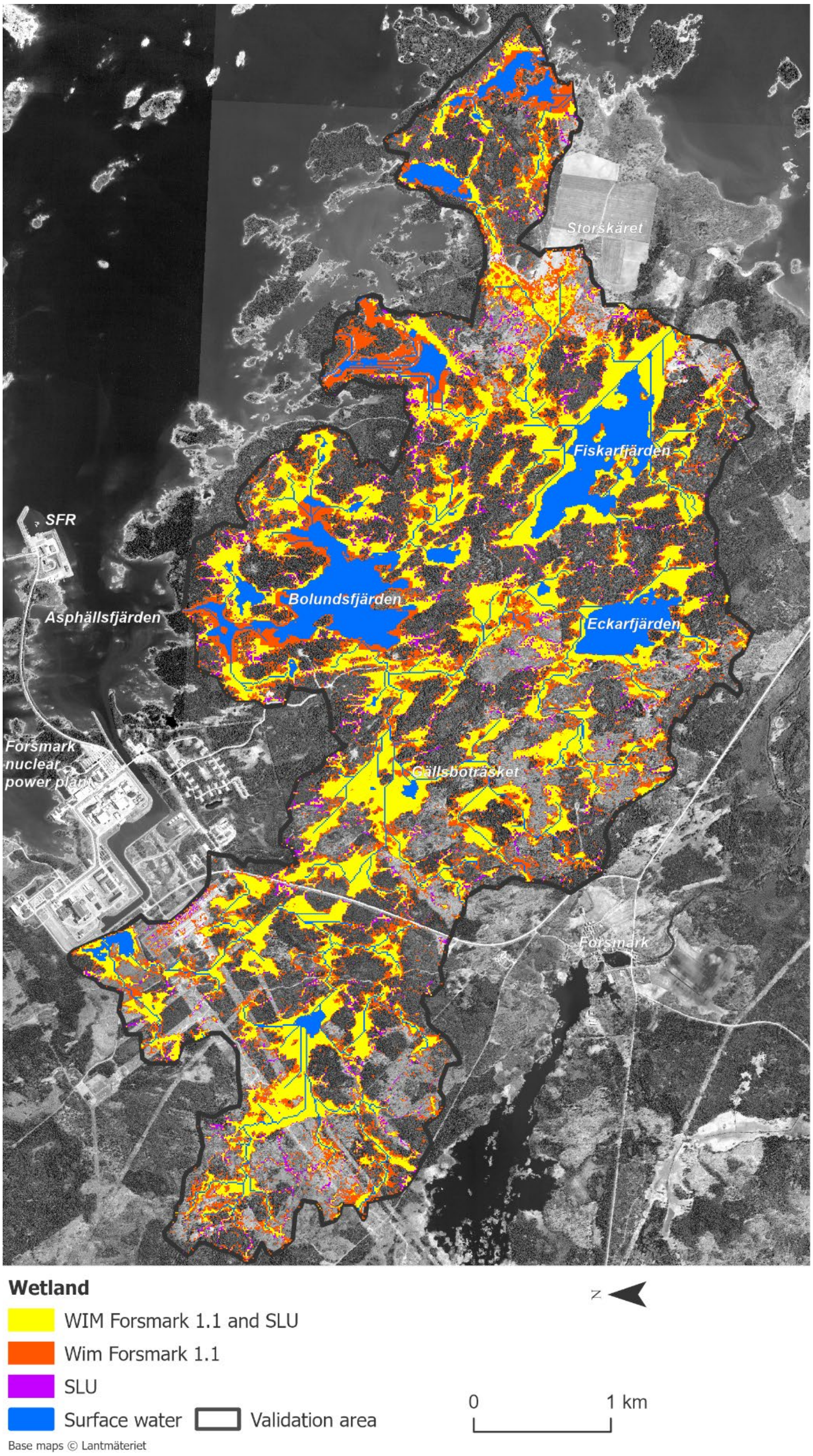

Figure H-1. Wetland predictions, presented as the amalgamated results from the 10 individual algorithms included in WIM Forsmark 1.1 and wetland predictions from SLU's wetness maps (combined "mesic-moist" (frisk-fuktig) and "moist-wet" (fuktig-blöt)) within the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4. Relative the total wetland area (i.e. sum of the WIM Forsmark 1.1. predicted wetland area and the SLU wetness map), overlap between the WIM Forsmark 1.1 and predicted wetlands and the SLU wetness map accounts for 59.3% of the total wetland area, WIM Forsmark 1.1 predicted wetlands that do not overlap with the SLU wetness maps accounts for 31.2% of the total wetland area and areas on the SLU wetness map that WIM Forsmark 1.1 did not predict as wetlands accounts for 9.4% of the total area.

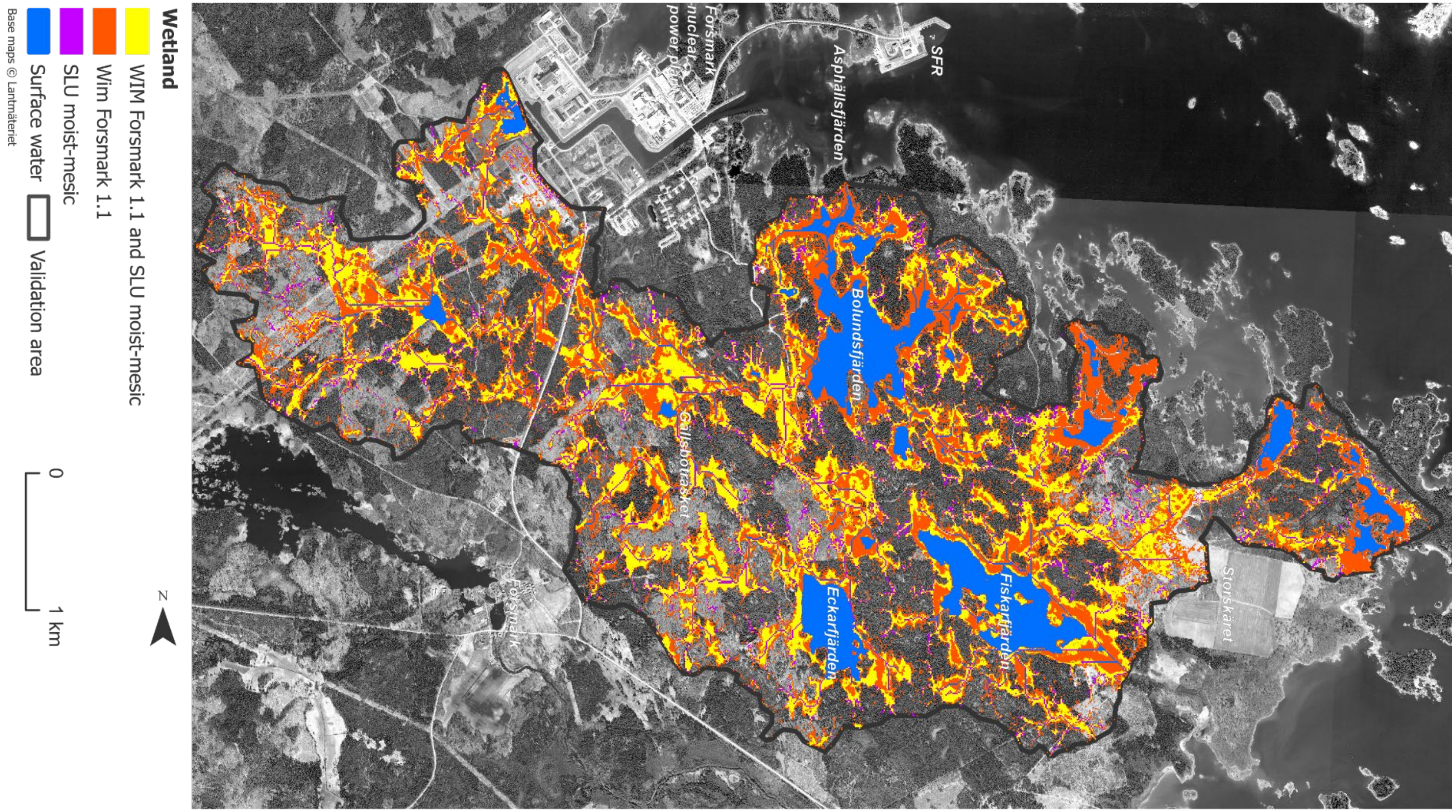

Figure H-2. Wetland predictions, presented as the amalgamated results from the 10 individual algorithms included in WIM Forsmark 1.1 and wetland predictions from SLU's wetness maps ("moist-mesic (frisk-fuktig)) within the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4. Relative the total wetland area (i.e. sum of the WIM Forsmark 1.1. predicted wetland area and the SLU wetness map for "moist-mesic"), overlap between the WIM Forsmark 1.1 and predicted wetlands and the SLU wetness map accounts for 43.7% of the total wetland area, WIM Forsmark 1.1 predicted wetlands that do not overlap with the SLU wetness maps accounts for 47.6% of the total wetland area and areas on the SLU wetness map that WIM Forsmark 1.1 did not predict as wetlands accounts for 8.8% of the total area.

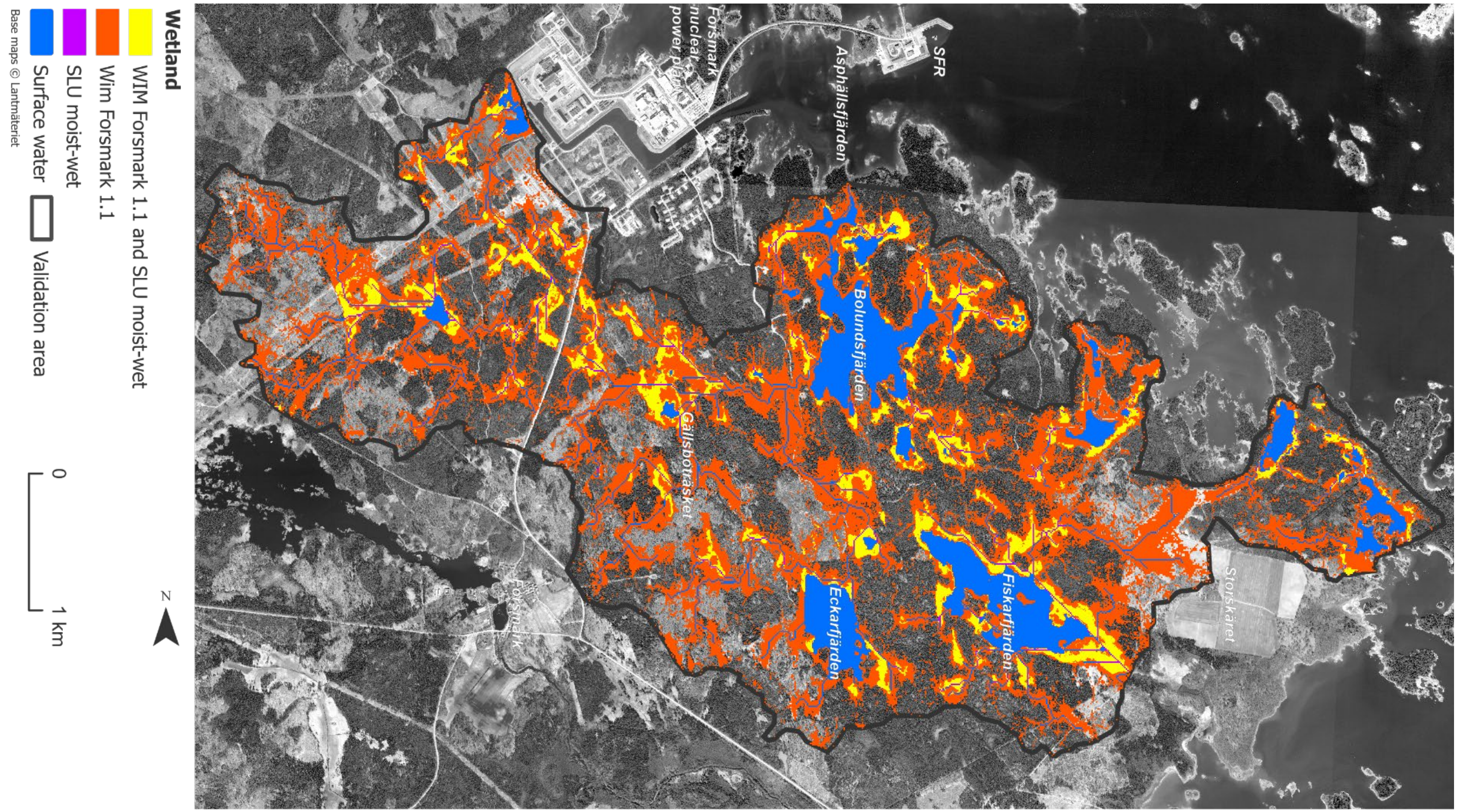

Figure H-3. Wetland predictions, presented as the amalgamated results from the 10 individual algorithms included in WIM Forsmark 1.1 and wetland predictions from SLU's wetness maps ("moist-wet (fuktig-blött)) within the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4. Relative the total wetland area (i.e. sum of the WIM Forsmark 1.1. predicted wetland area and the SLU wetness map for "moist-mesic"), overlap between the WIM Forsmark 1.1 and predicted wetlands and the SLU wetness map accounts for 17.6% of the total wetland area, WIM Forsmark 1.1 predicted wetlands that do not overlap with the SLU wetness maps accounts for 82.2% of the total wetland area and areas on the SLU wetness map that WIM Forsmark 1.1 did not predict as wetlands accounts for 0.2% of the total area.

Appendix I Wetland predictions using "p-means" for Forsmark validation area

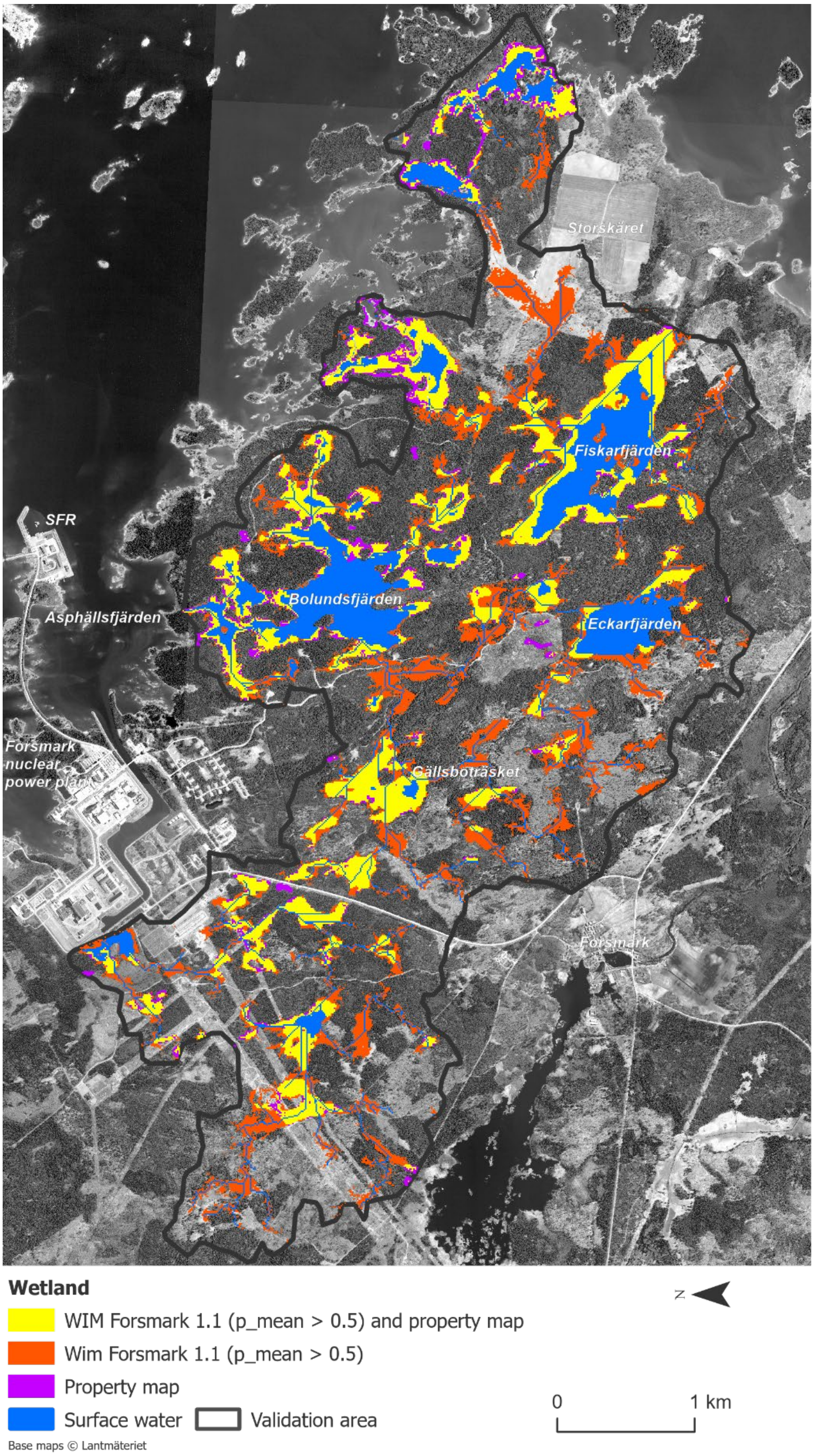

***Figure I-1.** Wetland predictions, presented using the p-means method (p > 0.50), and observed wetlands for the Forsmark validation area. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.*

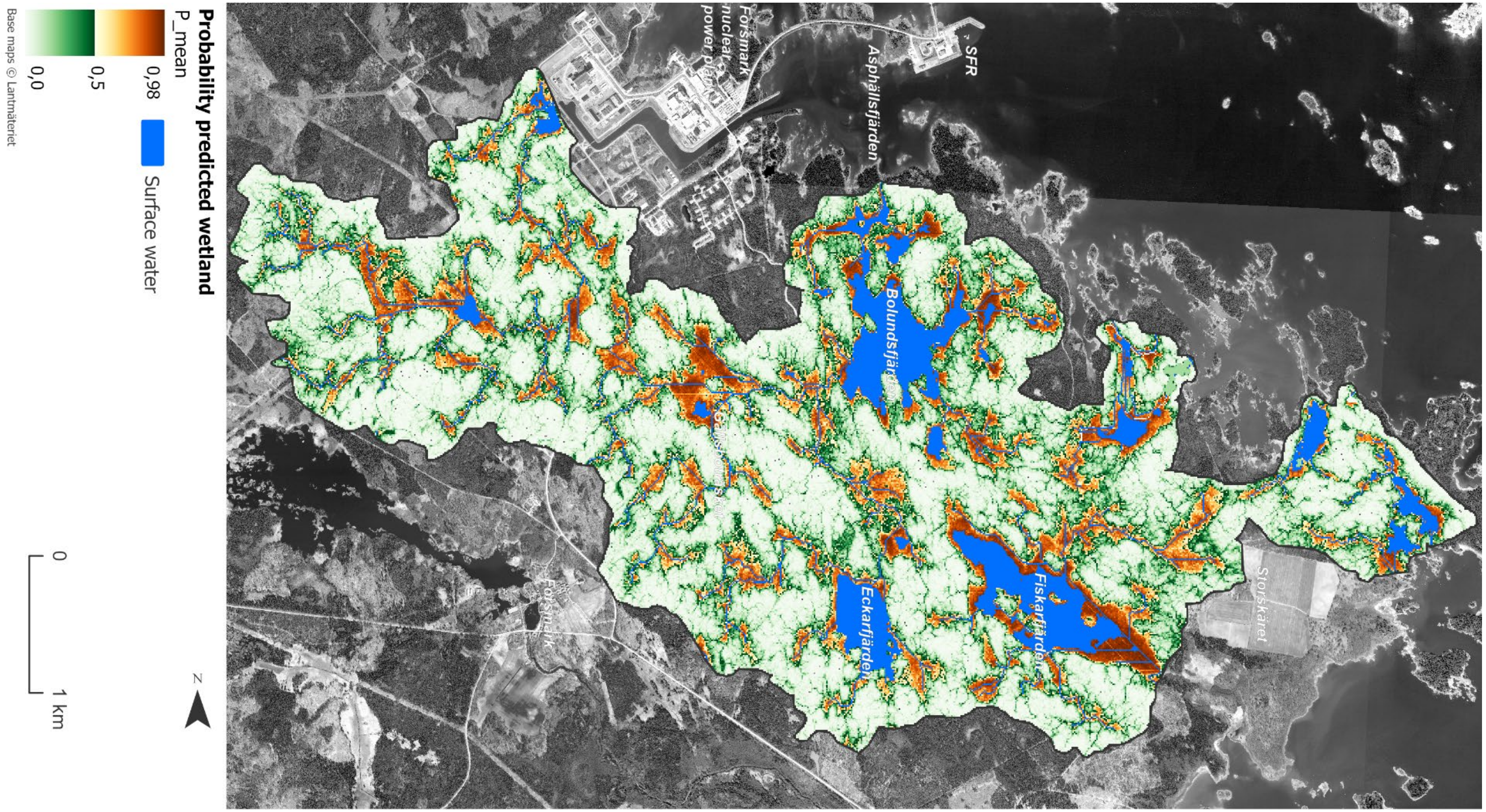

Figure I-2. Wetland predictions for the Forsmark validation area presented using the p-means. The surface water shown in the figure (lakes and waterways) represent the surface water data inputs discussed in Section 2.2.4.