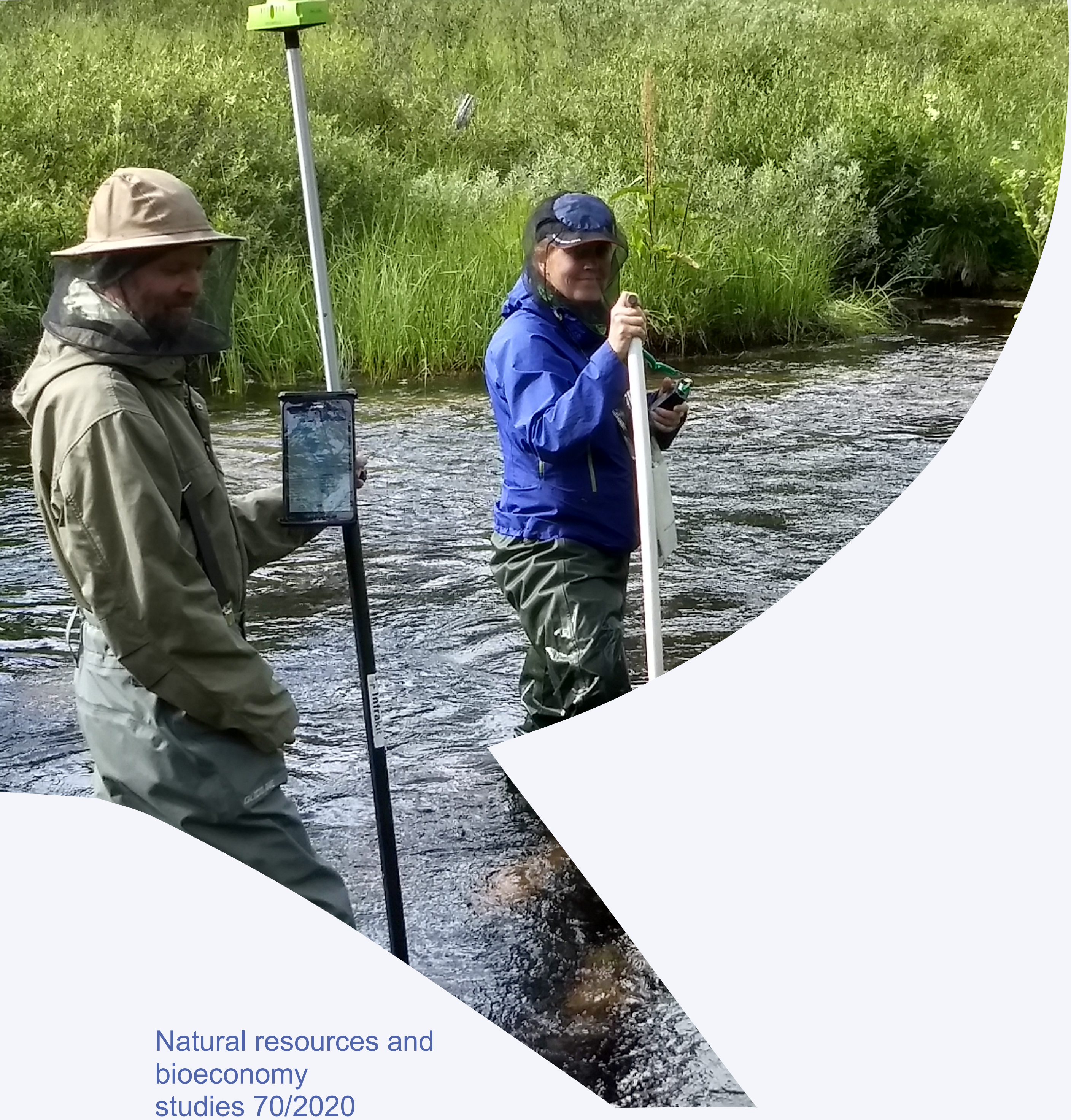

Natural resources and
bioeconomy
studies 70/2020

The impact of groundwater upwelling on the Tornionjoki trout

Project report

Atso Romakkaniemi, Pekka Jounela and Olli van der Meer

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Natural resources and bioeconomy studies 70/2020

The impact of groundwater upwelling on the Tornionjoki trout

Project report

Atso Romakkaniemi, Pekka Jounela and Olli van der Meer

Natural Resources Institute Finland, Helsinki 2020

How to refer this report:
Romakkaniemi, A., Jounela, P. & Olli van der Meer. 2020. The impact of groundwater upwelling on the Tornionjoki trout : Project report. Natural resources and bioeconomy studies 70/2020. Natural Resources Institute Finland. Helsinki. 29 s.

Atso Romakkaniemi, ORCID ID, https://orcid.org/0000-0001-9696-117X

ISBN 978-952-380-057-1 (Print)
ISBN 978-952-380-058-8 (Online)
ISSN 2342-7647 (Print)
ISSN 2342-7639 (Online)
URN http://urn.fi/URN:ISBN:978-952-380-058-8

Authors: Atso Romakkaniemi, Pekka Jounela and Olli van der Meer
Publisher: Natural Resources Institute Finland (Luke), Helsinki 2020
Year of publication: 2020
Cover photo: Atso Romakkaniemi
Printing house and publishing sales: PunaMusta Oy, http://luke.juvenesprint.fi

Summary

Atso Romakkaniemi[1], Pekka Jounela[2], Olli van der Meer[3]

[1]Natural Resources Institute, Paavo Havaksen tie 3, 90014 Oulun yliopisto

[2]Natural Resources Institute, Itäinen Pitkäkatu 4 a, 20520 Turku

[3]Tmi Olli van der Meer (consultancy), Hiomontie 14, 90850 Martinniemi

The effects of groundwater upwelling on the occurrence of fish species trout and sculpins were studied in the tributary streams Valkeajoki and Kuerjoki of the River Äkäsjoki (part of the Tornionjoki river system), northern Finland. Four main study areas with altogether six study sites were selected for data collection. Within these sites, altogether 81 sampling lines across stream channel were studied, comprising of 558 sampling points (1 sqm each). Habitat characteristics known to affect trout occurrence were mapped, and the presence of groundwater was identified/mapped from the existing TIR remote sensing data and by water and sediment temperature measurements. The chosen rivers are generally strongly fed by groundwater, which together with cool weather kept river water temperature low. The presence/absence data of fish species was collected by point sampling with a standard electrofishing equipment. Altogether 348 fish were caught, out of which 285 were trout parr and 63 were sculpins.

Trout and sculpins occurrence was modeled using self-organizing map (SOM), and random forests (RF) models. SOM was used to visualize interactions between predictors of the present study including trout, sculpin and no fish occurrences. Trout and sculpin occurrence patterns were recognized using RF model.

The analyses indicated that trout parr prefer sampling points with low water temperatures, low sediment temperatures, low river flow, high oxygen levels, small distance from the riverbank, low water depth and plain bottom. Thus, our analyses indicate similar general habitat choices for trout as found out in various other studies, but also a preference to the points of groundwater upwelling which is a new finding. The data was not suffient to indicate, why trout seem to favour strong presence of groundwater. However, the possible effects of groundwater upwelling on the distribution and abundance of submerged macrophytes as well as on the species composition of benthic macroinvertebrate communities is discussed in the light of trout habitat choices.

The study was carried out during 2017–2018 in the project "Groundwater and trout" (Luke's project ID 41007-00107300). The project was co-financed by the incomes from selling fishing licenses in Tornionjoki, Koneen säätiö and Luke.

Keywords: Groundwater, trout, habitat, upwelling, rivers, water temperature

Tiivistelmä

Atso Romakkaniemi[1], Pekka Jounela[2], Olli van der Meer[3]

[1]Luonnonvarakeskus, Paavo Havaksen tie 3, 90014 Oulun yliopisto

[2]Luonnonvarakeskus, Itäinen Pitkäkatu 4 a, 20520 Turku

[3]Tmi Olli van der Meer, Hiomontie 14, 90850 Martinniemi

Pohjavesipurkaumien vaikutusta kalojen (taimen, simput) esiintymiseen tutkittiin Valkeajoen ja Kuerjoen vesistöissä, jotka ovat Pohjois-Suomessa sijaitsevan Tornionjoen vesistöön kuuluvan Äkäsjoen sivuvesistöjä. Aineistojen keruuseen valittiin näistä vesistöistä neljä aluetta, joihin perustettiin yhteensä kuusi tutkimuskohdetta. Näissä tutkimuskohteissa aineistoa kerättiin yhteensä 558:sta yhden neliömetrin suuruisesta pisteestä, jotka sijaitsivat yhteensä 81:lla joen poikki perustetuilla linjoilla. Tutkimuspisteistä kerättiin habitaattidata muuttujista, joiden ennakolta tiedetään vaikuttavan taimenen esiintymiseen. Lisäksi pohjaveden esiintymistä/purkautumista tutkimuspisteissä arvioitiin pohjautuen sekä lämpökameralla tehtyihin ilmakuvauksiin että mittaamalla jokiveden ja pohjasedimentin lämpötilaa. Tutkimukseen valituilla jokialueilla esiintyy yleisesti ottaen runsaasti pohjavettä, mikä yhdessä aineistojen keruun ajankohdan viileän sään kanssa piti jokiveden lämpötilan alhaisena. Kalojen esiintymistä tutkimuspisteissä kartoitettiin tyypillisesti virtavesien kalastotutkimuksissa käytetyllä sähkökalastuslaitteella. Koekalastuksissa saatiin saaliiksi yhteensä 348 kalaa, joista 285 yksilöä oli taimenenpoikasia ja 63 yksilöä oli simppuja.

Taimenten ja simppujen esiintymistä suhteessa kartoitettuihin muuttujiin analysoitiin ns. itseorganisoituvien karttojen (self-organizing map, SOM) ja ns. Random Forest (RF) -menetelmällä. SOM-menetelmä visualisoi selitettävien muuttujien (kalojen esiintyminen) ja selittäjien välisiä yhteyksiä. RF-menetelmällä määritettiin ko. yhteyksien muotoja.

Analyysitulosten mukaan taimenenpoikaset suosivat tutkimuspisteitä, joissa oli alhainen veden ja sedimentin lämpötila, suhteellisen hidas virtausnopeus, korkea happipitoisuus, lyhyt etäisyys joen rantaviivaan, matala veden syvyys ja vesikasveista paljas pohja. Tulokset siis tukevat kirjallisuudessa esitettyjä yleisiä taimenen habitaattipreferenssejä, mutta uutena tuloksena havaittiin taimenen suosivan myös pohjaveden purkauspaikkoja. Kerätyt aineistot eivät ole riittäviä selittämään, miksi taimen suosii runsasta pohjaveden läsnäoloa. Mahdollisina syinä pohditaan pohjaveden vaikutusta pohjakasvillisuuden ja/tai pohjaeliöstön esiintymisen ja lajiston muokkaajana taimenelle suotuisaksi.

Tutkimus toteutettiin vuosina 2017–2018 projektissa “Groundwater and trout” (Luken projektinumero 41007-00107300). Projektia yhteisrahoittivat Tornionjoen kalastuslupatulot, Koneen säätiö ja Luke.

Asiasanat: Pohjavesi, taimen, habitaatti, kumpuaminen, joet, vedenlämpö

Contents

1. Introduction

Salmo trutta or sea trout is an anadromous form of brown trout. Anadromous and local forms of brown trout coexist in same rivers and local trout may spawn with anadromous ones (Jonsson 1985). Existence of healthy, productive sea trout populations requires habitats, which are well suited to the specific life stages occupied by the species along its migration routes. That is, spawning rivers must contain spawning grounds and juvenile rearing habitats, which provide shelter, food and high survival. Feeding areas at sea must provide enough good-quality food and low enough predation pressure. Smolt and spawning migration route must allow for easy, safe and fast migration up- and downstream. Finally, fishing pressure must be maintained at sustainable levels.

In the Tornionjoki river system, sea trout is found to spawn mostly in tributary streams and rivers located 200-300 km upstream from the sea (Ikonen *et al.* 1985; Bergelin & Karlström 1985). Most of the tributaries have been affected by human activities, including timber floating, catchment drainage, forestry, road constructions, agriculture and mining. The River Äkäsjoki which together with its tributaries forms the most important spawning area of Tornionjoki sea trout (Ikonen *et al.* 1985, Palm *et al.* 2019).

Many human activities are directly or indirectly affecting the quantity and quality of groundwater and surface water supplies. Trout are often found in groundwater dominated river catchments (Zorn *et al.* 2002; Gosselin *et al.* 2012), which indicates trout's preference to rivers with high groundwater supply. Places of groundwater inflow may provide shelter against adverse environmental conditions and thus may increase survival and production of trout (Bowlby & Roff 1986). The role of groundwater for trout, however, is not much studied and the existing studies are conducted in conditions clearly different from those prevailing in the R. Tornionjoki catchment. It is therefore important to increase our knowledge about groundwater-surface water interactions and their role for trout in our conditions.

Thermal infrared (TIR) remote sensing enables location and quantification of groundwater inputs (Dugdale 2016). Drs. Anne Rautio and Kirsti Korkka-Niemi from the Department of Geosciences and Geography, University of Helsinki, conducted large-scale aerial infrared surveys to map river water temperatures in the Äkäsjoki main stem and in some of the tributaries to R. Äkäsjoki. The preliminary analyses indicate good success for exact identification of groundwater discharge in the mapped reaches (e.g., large small-scale spatial variation in surface water temperatures, up to several degrees of Celsius). Direct measurements of water temperature and its spatial variation within a channel can be used parallel to the TIR remote sensing, to indicate both the general presence of groundwater in a river section and to map e.g. the dimensions of confluence plumes within a channel. Single in-channel points of groundwater upwelling can be identified and located by sediment temperature measurements. However, temperature measurements are useful for these purposes only during the seasons when the surface water temperature differs from the groundwater temperature (4 degrees Celcius).

A joint analysis of spatial distribution of groundwater upwelling, together with detailed survey data on spatial variation in fluvial fish fauna, would allow for new insights into the role of groundwater for trout habitat choices. For this purpose, a collaborative project between Luke and the University of Helsinki was run in 2017–2018. The project was co-financed by the incomes from selling fishing licenses in Tornionjoki, Koneen säätiö and Luke. This report describes the main activities and outcomes of the project.

2. The study area

The Äkäsjoki water course is a tributary of the Tornionjoki river system and it flows into the River Muonionjoki in the western Finnish Lapland. The Äkäsjoki catchment covers 649 km^2 of sparsely populated northern boreal terrain dominated by coniferous forests. The main river (Äkäsjoki) is 46 km long and its average discharge is 7.2 m^3s^{-1}. Several smaller tributaries flow into the main river, among largest of which are Kuerjoki and Valkeajoki (Figure 1). Trout inhabit most rivers of the water course and the highest densities of trout juveniles (parr) are frequently found in the Valkeajoki and Kuerjoki (e.g. Vähä *et al.* 2007). These rivers are also spawning areas of sea trout. The fish fauna in fluvial habitats of the Äkäsjoki system is typical for the region and consist of several salmonid species (salmon, grayling, whitefish and trout). Salmon used to occur only on the lowermost main river, but the species has recently colonized the whole lower main river and also the lowermost short sections of the Kuerjoki and Valkeajoki. Interestingly, Cottus species (European and Siberian sculpins) which are found elsewhere in the river system are not found in Kuerjoki upstream from the waterfall Kuerlinkka (which is located 1 km upstream from Kuerjoki's confluence to Äkäsjoki main river).

Four main study areas (stretches of river) were selected for data collection, two from Valkeajoki (hereafter called Lower and Upper Valkeajoki) and two from Kuerjoki catchment (hereafter called Kuerjoki and Vitsajoki; Figure 1). Moreover, at the Lower Valkeajoki and Kuerjoki areas the groundwater-fed brooks (one brook in both areas) creating a confluence plume were also studied from their lowermost tens of meters. Thus, altogether six sampling sites were identified for the analyses. The criteria for choosing these sampling sites were as follows:

- the sites have points of groundwater upwelling, either on the side of the river channel or within the river channel
- High trout parr densities are expected to occur
- The sites represent somewhat different habitat characteristics

Groundwater upwelling within the main study areas was identified/mapped from the existing TIR remote sensing data (all areas) and by sediment temperature measurements (all other areas except Lower Valkeajoki). The Lower and Upper Valkeajoki are located rather close to each other and the habitat characteristics seem rather similar. In the Upper Valkeajoki numerous in-channel (hyporheic) upwelling points of groundwater and a cold alcove were observed, while on the Lower Valkeajoki a groundwater-fed brook (sampling site called A-oja) joined the main river with a confluence plume. Other sources of groundwater may also occur in the Lower Valkeajoki, but they were not mapped. The Kuerjoki has also a confluence plume from a groundwater-fed brook (called Puukko-oja site) as well as a few hyporheic upwelling points and lateral seeps. The river at Kuerjoki site is larger and trout parr densities were lower than elsewhere. The Vitsajoki (tributary to Kuerjoki) was selected as the last study site due to its seemingly homogenous, fast-flowing habitat and several groundwater sources including cold alcoves, lateral seeps and hyporheic upwelling points.

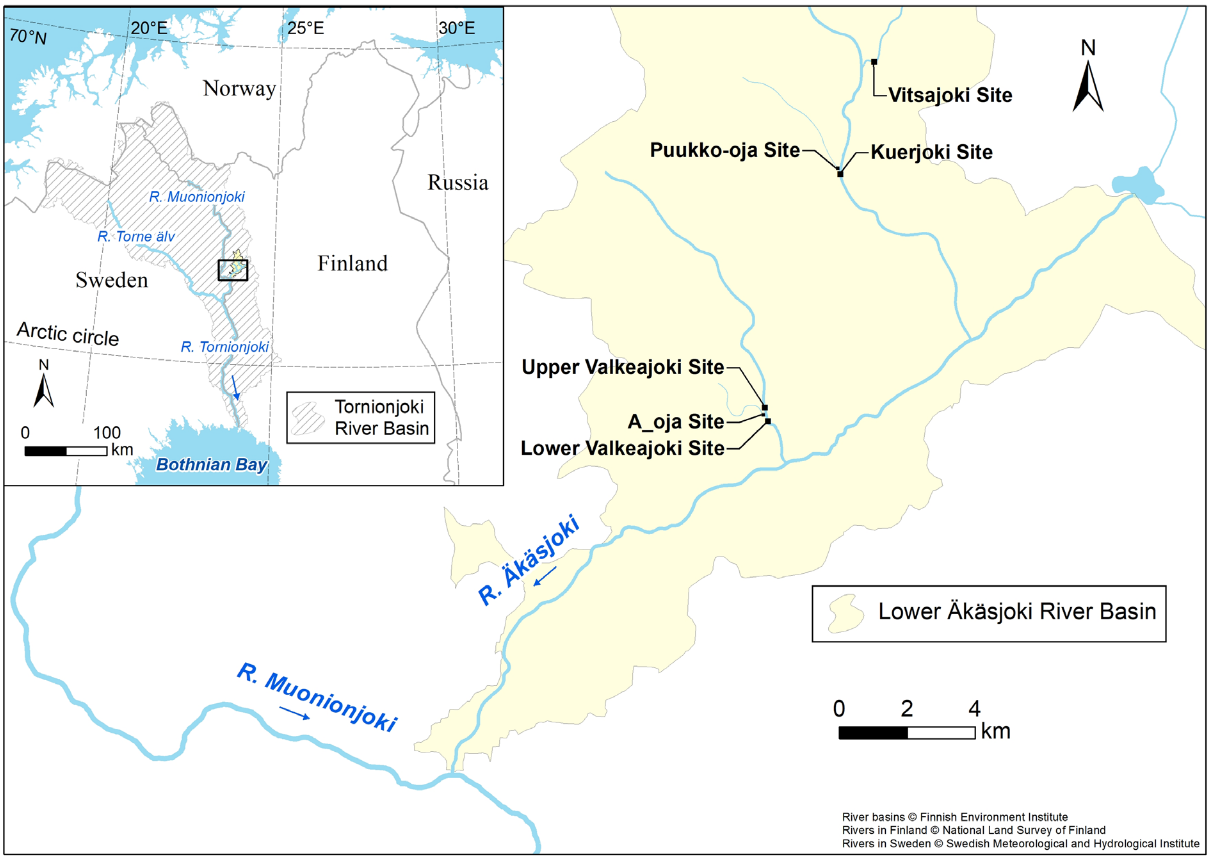

Figure 1. Four main study areas (stretches of river) were selected for data collection, two from Valkeajoki (Lower and Upper Valkeajoki sites) and two from Kuerjoki catchment (Kuerjoki and Vitsajoki sites). Moreover, at the Lower Valkeajoki and Kuerjoki areas the groundwater-fed brooks (A-oja and Puukko-oja, respectively) were also studied.

3. Materials and methods

3.1. Spatial design of data collection

Within each study site, data collection was based on a number of horizontal sampling lines crossing perpendicular the river channel. These lines were close to each other in the vicinity of the confluence of the groundwater-fed brooks at Lower Valkeajoki and Kuerjoki and located further away from each other the longer the distance to the confluence. However, distances between the lines were predefined in order to avoid subjective selection of the exact locations. At the Upper Valkeajoki and Vitsajoki no clear main source of groundwater was indentified, but instead numerous in-channel upwelling points together with cold alcoves; there location of the sampling lines did not follow any specific pattern.

Along each line a rope was attached to cross the river just above the water surface. There were marks in the rope 1 meter apart from each other. The rope and the marks were used to define a series of about 1 sqm large sampling points along the rope across the river (Figures 2–5).

In total, 81 sampling lines comprising of 558 sampling points were studied:

- 21 sampling lines and 131 sampling points in Lower Valkeajoki, including A-oja
- 12 sampling lines and 68 sampling points in Upper Valkeajoki
- 20 sampling lines and 192 sampling points in Kuerjoki, including Puukko-oja
- 28 sampling lines and 191 sampling points in Vitsajoki

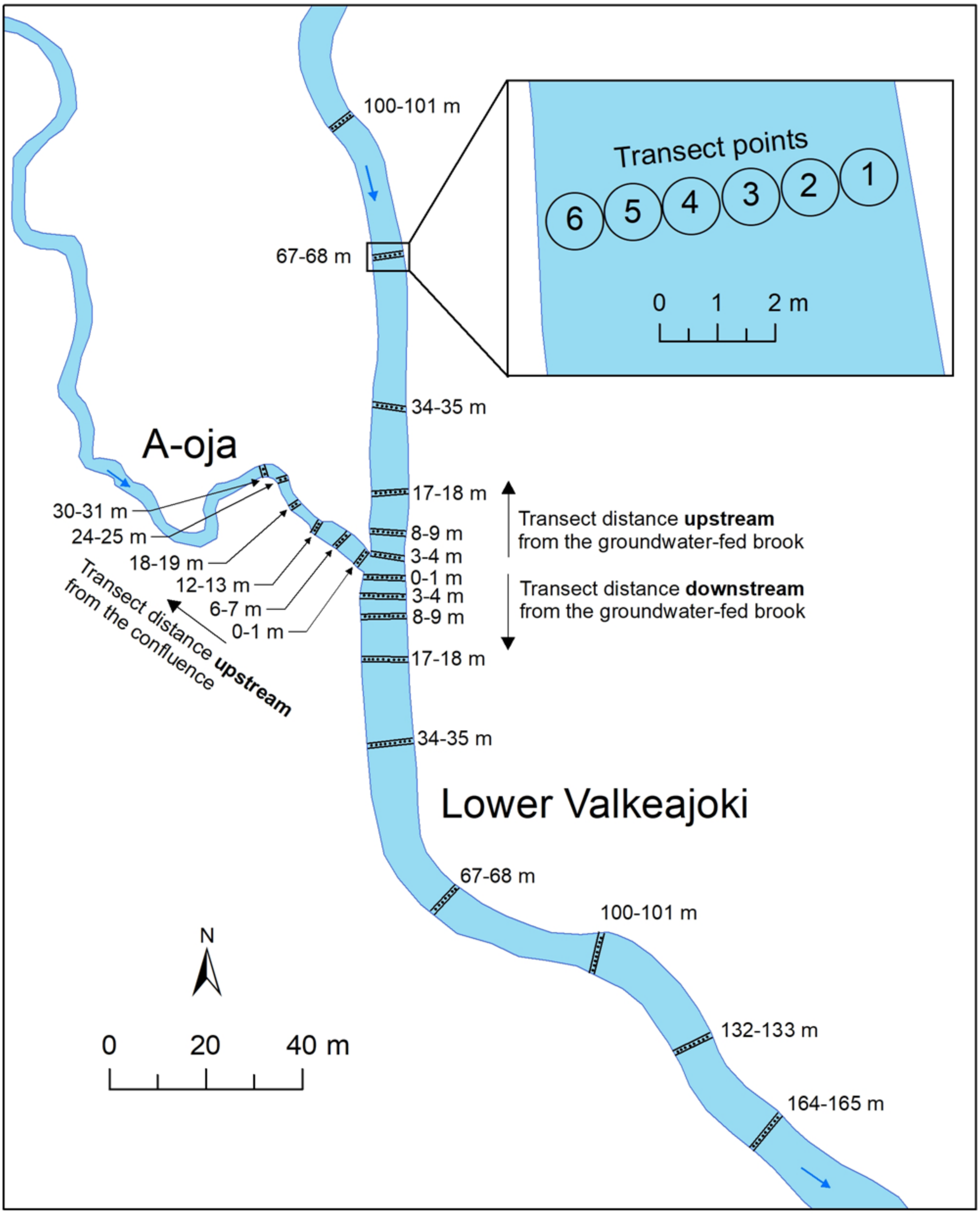

Figure 2. Locations of the sampling lines at the Lower Valkeajoki and A-oja sites and one enlargened sampling line as an example showing each sampling point.

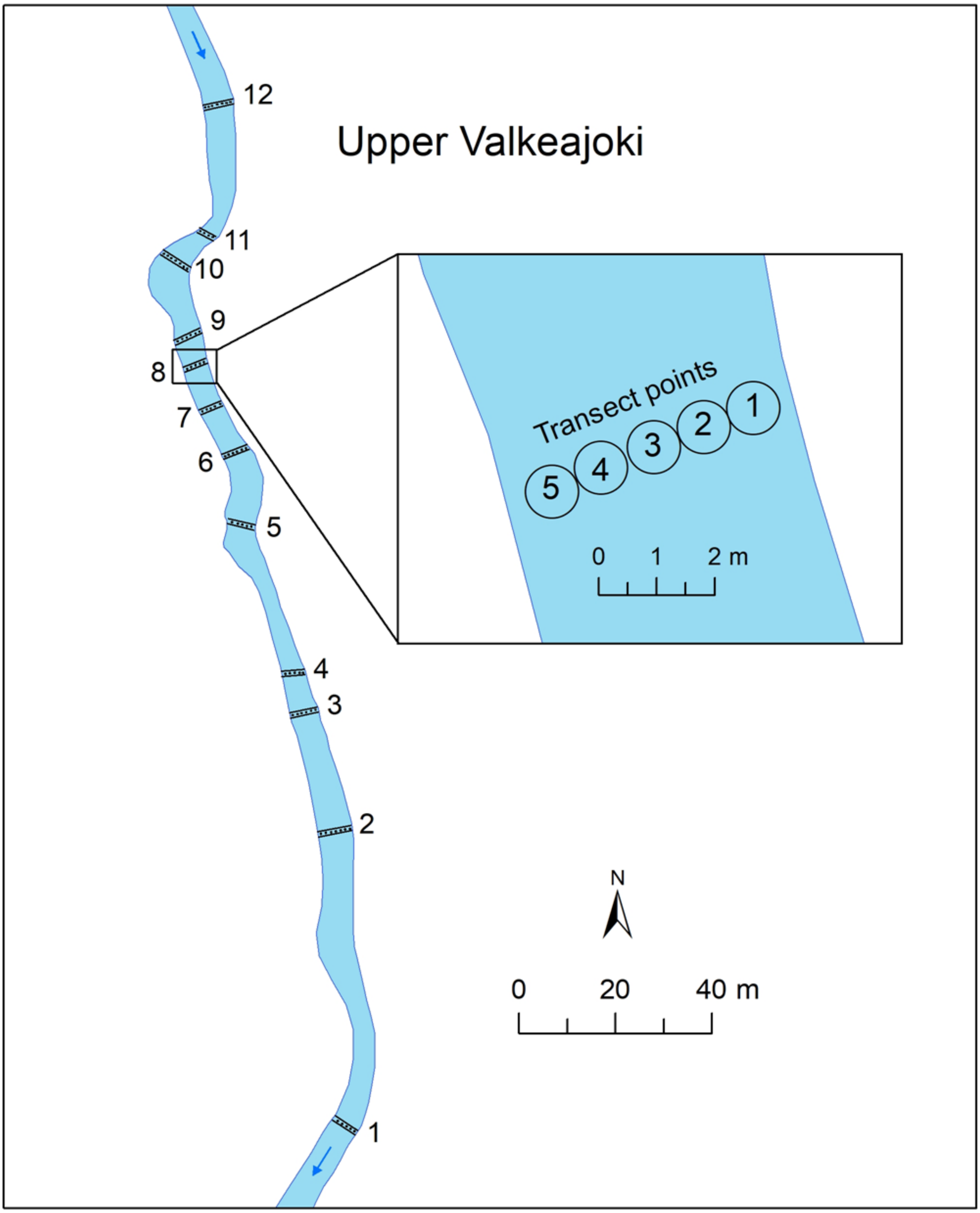

Figure 3. Locations of the sampling lines at the Upper Valkeajoki site and one enlargened sampling line as an example showing each sampling point.

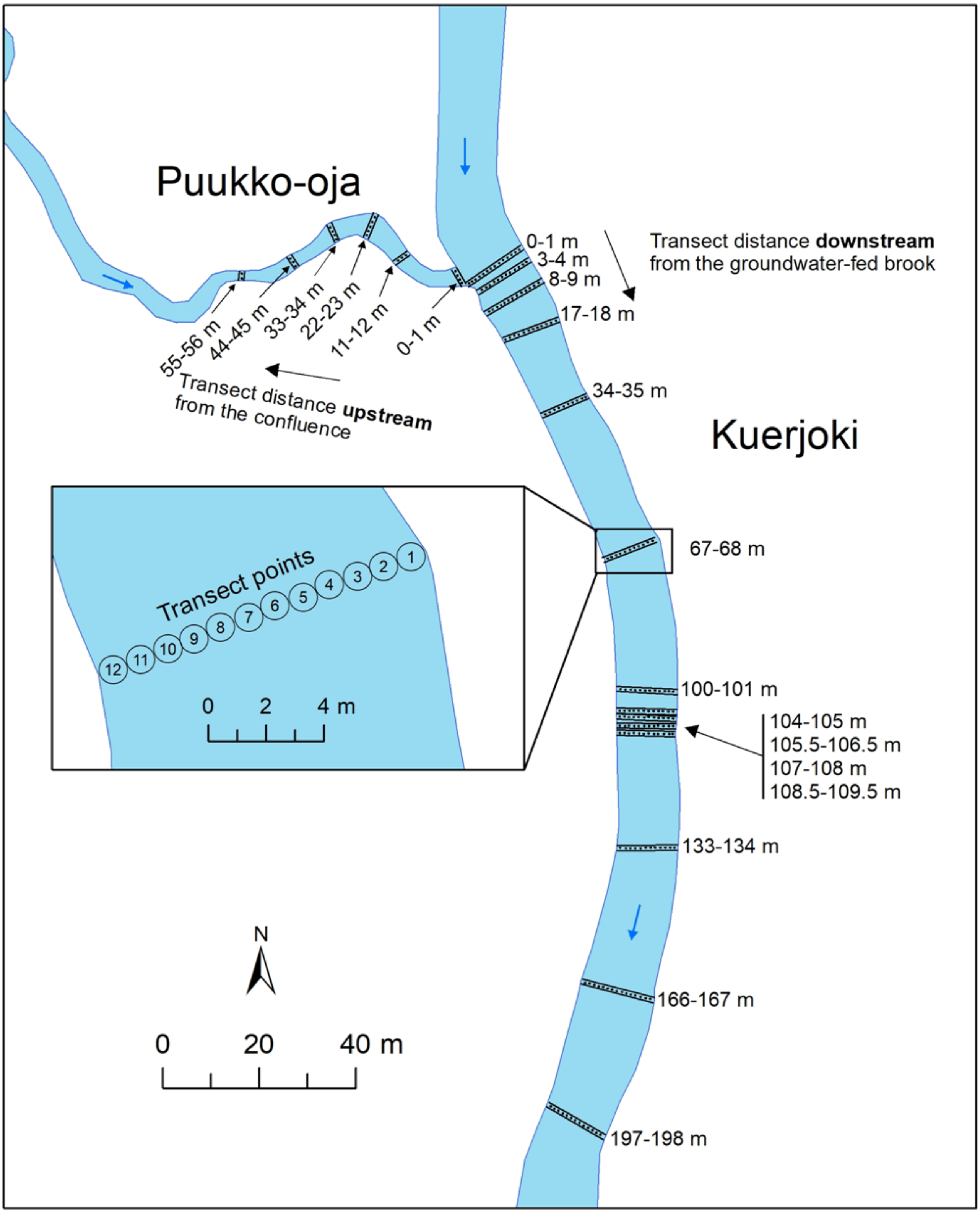

Figure 4. Locations of the sampling lines at the Kuerjoki and Puukko-oja sites and one enlargened sampling line as an example showing each sampling point. The sampling lines from 104-105 m to 108.5-109.5 m are located around one strong in-channel groundwater upwelling point.

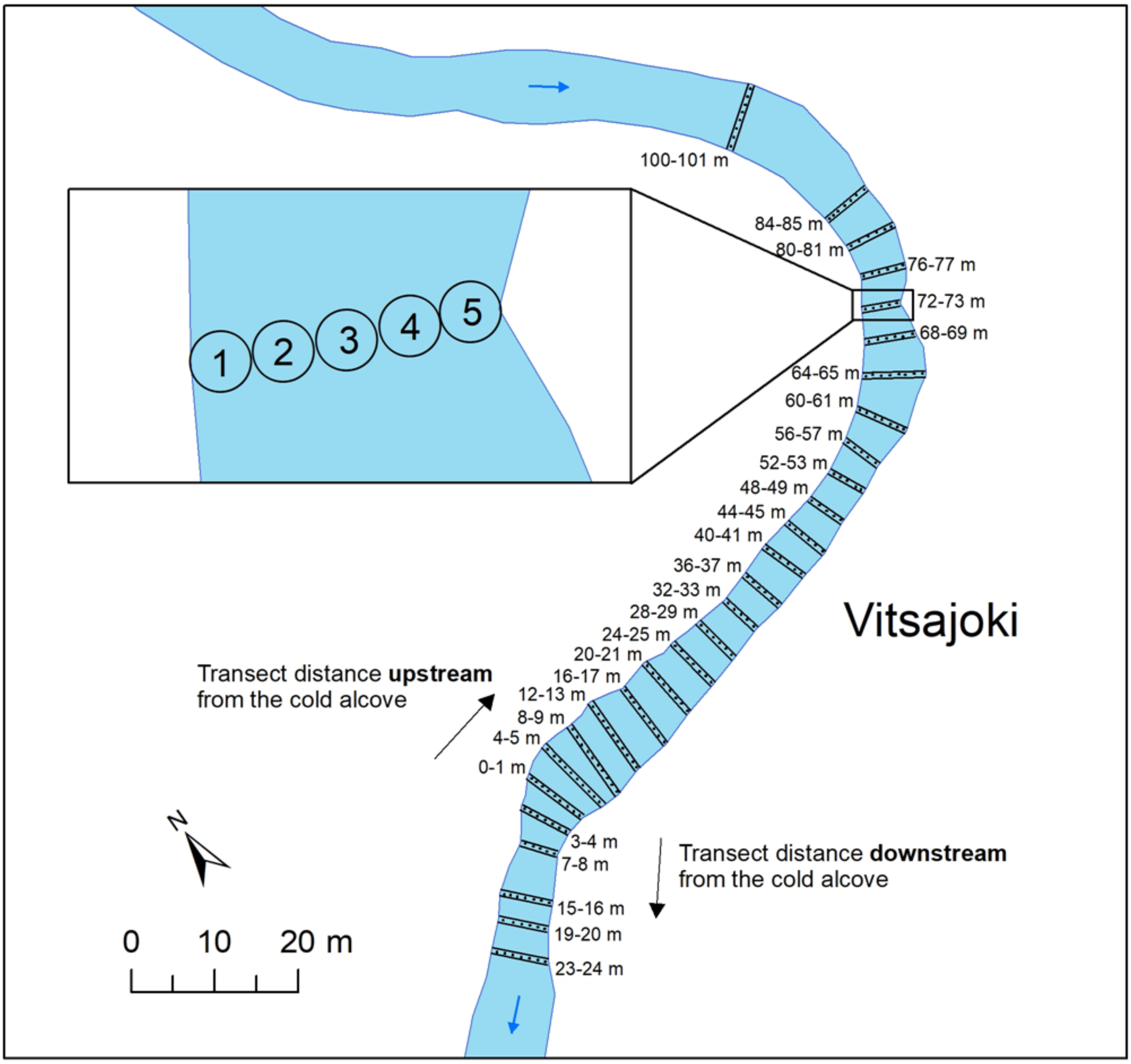

Figure 5. Locations of the sampling lines at the Vitsajoki site and one enlargened sampling line as an example showing each sampling point.

3.2. Fish data

A standard bankside electrofishing equipment including a generator, a separate control box and electrodes was used to collect fish data from the sampling points. Pulsed direct current with a frequency of 50/second and 600 V was used in fishing. The anode was 24 cm in its diameter and based on previous test fishing, an effective fishing area of this anode with the above settings covers a circle with about 50 cm radius.

The fishing team consisted of a person using the anode and a person catching fish with a dipnet. Dipnet was placed on the river bottom one meter downstream the rope. Then the anode was dipped half meter downstream the rope directly above the dipnet. Electricity was turned on and the anode was moved into the dipnet along the water flow. Sampling points were fished this way across the river, starting from one bank and finishing fishing to the opposite bank. Catch of each sampling point was placed in separate small buckets and the catch was treated and recorded after fishing of all the sampling points of a sampling line.

From the catch, the species of each individual was identified (except the very smallest sculpins the species of which could not be identified by eye) and the total length of each specimen was measured. In case age of trout would be needed for any later analyses, ages of trout were identified: young-of-the-year (YOY) trout were identified based on their length, while scale samples for ageing were taken from larger specimens. Individual weight of fish longer than 50 mm was also measured.

Altogether 348 fish were caught, out of which 285 were trout and 63 were sculpins (Cottus Gobio and Cottus poecilopus):

- 190 specimens (out of which 140 trout) in the Lower Valkejoki, including A-oja
- 75 specimens (62 trout) in the Upper Valkeajoki
- 25 specimens (all trout) in the Kuerjoki, including Puukko-oja
- 58 specimens (all trout) in the Vitsajoki

3.3. Habitat data

3.3.1. Physical, hydraulic and macrovegetation data

From each sampling line and point, the following habitat characteristics were measured: channel width (to the nearest 10 cm), water depth (to the nearest 1 cm), water velocity (meter/second), bottom substrate size (see details below) and instream vegetation (see details below).

Water velocity was measured at the 0.6 depth. From part of the sampling points measurements were made by a “Flowtracker” acoustic doppler velocimeter, and from the rest of the sampling points by a “Schiltknecht MiniAir20” flow meter. A fraction of the sampling points was measured using both devices. Flowtracker measurements were observed rather unstable and unexpectedly small, possibly due to the instream vegetation creating small-scale eddies in the near-bottom and mid-water water flow. Flowtracker also gave clearly lower velocity values compared to the flow meter measurements from the same sampling points. The flow meter measurements were regarded as more trustworthy, therefore a conversion factor was estimated and applied to the Flowtracker measurements in the sampling points from which velocity was measured only by Flowtracker. As a result, comparable estimates reflecting the prevailing water velocity were obtained for each sampling point.

The proportions of various bottom substrates were visually assessed to the nearest 10% using a following classification:

Name	Substrate diameter (mm)
Organic/fine	< 0.2
Sand	0.2–2
Gravel	2–20
Stone 1	20–100
Stone 2	100–200
Boulder	> 200

Similarly, the coverage of bottom vegetation was assessed to the nearest 10% using the following classification:

Name (species family)
Plain bottom (no vegetation)
Water moss (Fontinalis spp.)
Water-milfoil (Myriophyllum spp.)
Sedge (Carex spp.)

The above classifications were assessed from aerial photos, except in the Vitsajoki site where the classification was done in the field.

3.3.2. Identification of the extent of groundwater presence

Water and sediment temperature measurements were used as the principal indicators of the groundwater supply. In each sampling point the water temperature was measured from the middle of the water column. Each measurement was compared against the average of water temperature across all the sampling points in the same a horizontal sampling line. This way one could identify major groundwater supplies located on either side of the river (groundwater rich confluence plumes, cold alcoves and lateral seeps), which cool the near-bank water before fully mixing to the rest of the river water.

Sediment temperatures were measured approximately 5 cm below the bottom surface. These measurements covered all sampling points except those of Lower Valkeajoki and A-oja. Sediment temperature, when compared against the mid-water temperature in the same sampling point, was considered to indicate points of in-channel groundwater supply.

The larger the difference between temperatures in the above-mentioned comparisons was the stronger the groundwater influence was considered in a sampling point. In Upper Valkeajoki sediment temperatures were mapped throughout the studied river stretch with a grid of 1-1.5 m distance between measurements. Figure 6 shows the points where sediment temperature was found to be at least 3 degrees colder than the river water, apparently indicating substantial groundwater upwelling. Groundwater upwelling points were not as abundant in the other studied sites as in the Upper Valkeajoki.

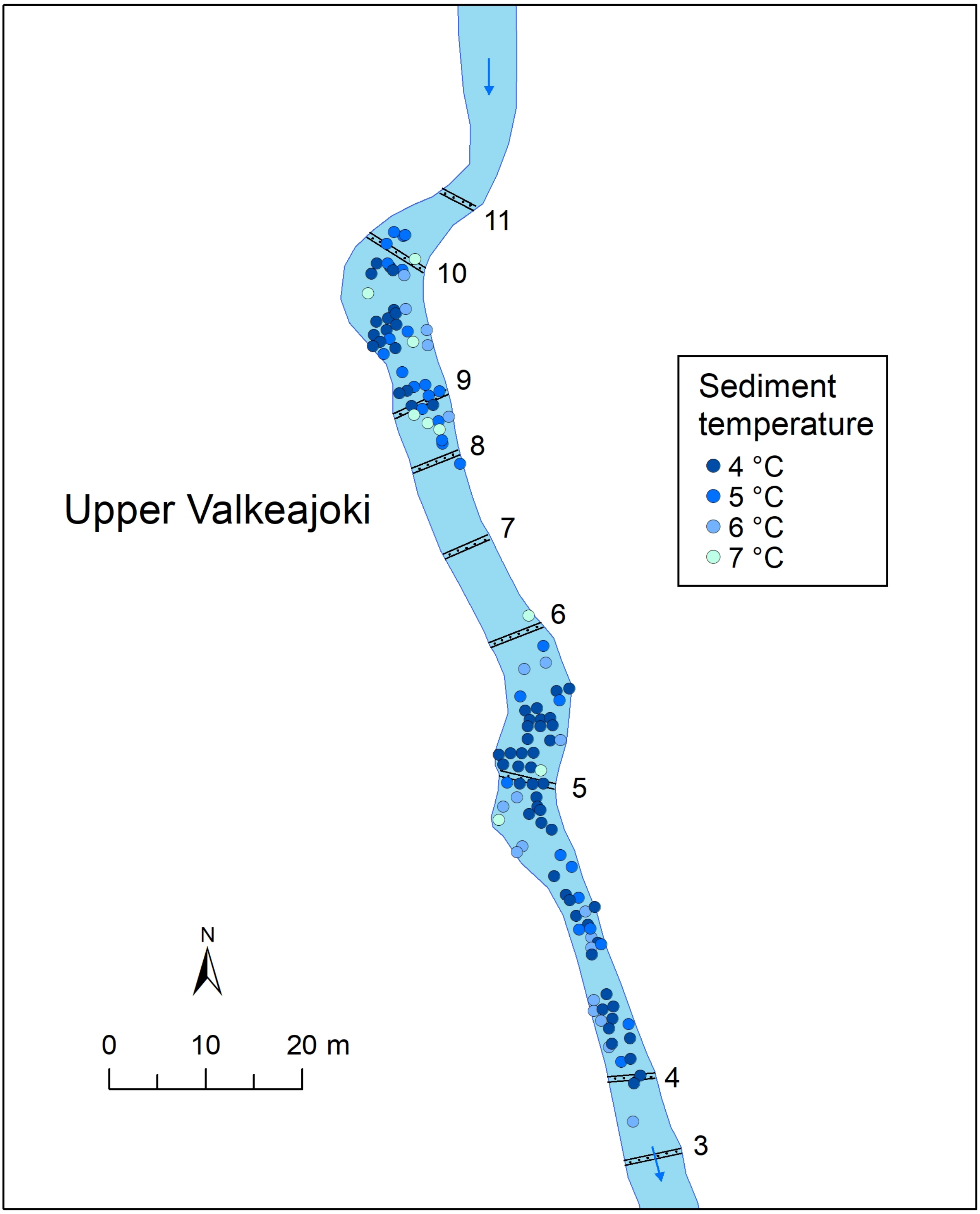

Figure 6. Cold anomalies of the sediment temperature (difference between the water and the sediment temperatures being 3 degrees or more) in the Upper Valkeajoki site. The prevailing water and sediment temperatures were about 10 degrees (Figure 11).

3.4. Data analyses

The predictors of trout and sculpin occurrence were: sampling site (6 sites), substrate diameter (5 size categories), vegetation species (3 categories), size of plain bottom, pH, oxygen level, conductivity, sediment temperature, water flow, water depth, water temperature, water temperature in i^{th} sampling point minus average water temperature in the same horizontal sampling line across the river, and rescaled distance from the river bank of a sampling point (range = 0, 0.1,…, 0.5) where 0 refers to river bank and 0.5 refers to the horizontally midmost sampling point across the river.

Sediment temperature had not been measured in two sampling sites (Lower Valkeajoki, including the site A-oja) and hence, additional model runs were done without information on sediment temperature with an aim to take into account consistent information from all sampling sites.

Statistical methods

Trout and sculpin occurrences were modeled using self-organizing map (SOM, Kohonen 1982, 2001), and random forests (Breiman 2001) models.

At first, SOM was used to visualize interactions between predictors of the present study including trout, sculpin and no fish occurrences. In general, SOM is an unsupervised dimensionality reduction method that visualizes high-dimensional data (here: 26 predictors, and 567 non-missing valued samples) in low-dimensional (typically two-dimensional) map. The two dimensions of the SOM map was then grouped, i.e., clustered. This two-stage procedure, first using SOM to produce the prototypes that are then clustered in the second stage, has been found to perform well when compared with direct clustering of the data (Vesanto and Alhoniemi 2000). The two dimensions of SOM were clustered using k-means algorithm (Kohonen 2014) and using Davies Boulding validity index (Davies and Boulding 1979) as a performance criterion. In the parameter optimization, SOM net sizes (x and y dimensions) and the number of clusters parameter k of the k-means algorithm were altered using grid search until the minimum of the Davies Boulding index was found. Each trial SOM consisted of 1000 training rounds and the learning rate function was inverse-of-time, which ensures that all input samples have approximately equal influence on the results. In parameter optimization, the SOM net size (x and y dimensions) was not allowed to exceed the map size rule (of thumb) of Vesanto and Alhonniemi (2000; N(nodes) = 5 x sqrt(N_rows)).

Trout and sculpin occurrence patterns were recognized using random forests (RF) model (Breiman 2001). RF is an ensemble classifier of decision trees produced from bagging (bootstrap aggregating; see Breiman 1996) and a randomized variant of the tree induction algorithm. In ecological studies, RF has been applied for species distribution models (SDMs), e.g. for species conservation and biodiversity management purposes, and in the context of climate change (Guo *et al.* 2015). One general reason for using ensemble SDMs has been to reduce uncertainty and stability in predictions, especially when compared with a single SDM, such as a generalized linear model or a generalized additive model (Marmion *et al.* 2009a,b, Grenouillet *et al.* 2011, Guo *et al.* 2015). The parameters (number of trees, maximum depth) of the RF model were estimated using 10-fold cross-validation (Kohavi 1995) applied to the grid-search. After finding the best parameters based on 10 non-overlapping test sets, the best model was applied with all data once. SMOTE up-sampling (Chawla et al. 2002) was used to balance class-imbalance of three response classes (trout, sculpin, no fish). The sample size in RF modeling was 1344 rows with 10.4% missing data. Missing data imputation was not used.

All statistical analyses were performed using RapidMiner software (version Studio Large 9.3.000., https://rapidminer.com /, Mierswa *et al.* 2006).

4. Results

Studied habitats

Altogether, field work focused on 558 sampling points located in 81 sampling lines. However, sediment temperature was not measured in the Lower Valkeajoki and A-oja sites and the analyses which included sediment temperature could therefore be carried out only from 411 sampling points covering 67 sampling lines.

Channel width is clearly the smallest in the groundwater fed brooks flowing into the main rivers, while the widest river channel is in the Kuerjoki site:

Site	Channel with, average (m)	Measured range (min-max) of the channel width (m)
Lower Valkeajoki	7.7	4.5–9.1
A-oja	2.9	1.6–4.0
Upper Valkeajoki	5.6	4.0–7.7
Kuerjoki	12.4	10–14.5
Puukko-oja	3.5	1.7–4.8
Vitsajoki	7.0	4.5–10.0

Water depth is most stable across the sampling points of Lower Valkeajoki, while the largest depth variation is found among the Upper Valkeajoki and Vitsajoki sampling points (Figure 7). The shallowest site is Puukko-oja and the deepest site is Kuerjoki.

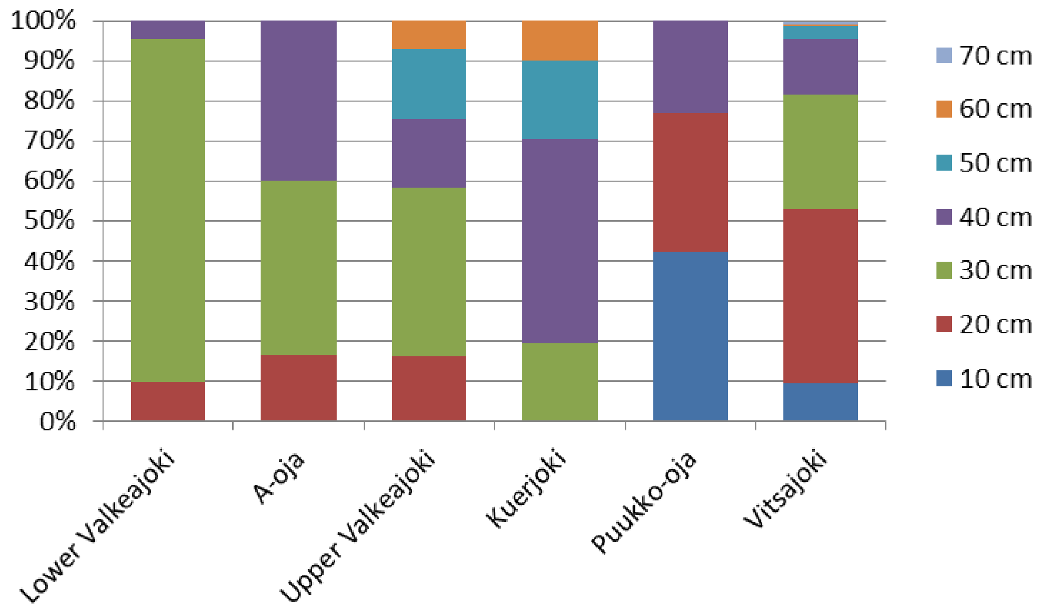

Figure 7. Proportion of sampling points in each depth class (rounded to the nearest 10 cm) by sampling site.

Bottom substrate varies the most in the Upper Valkeajoki and Kuerjoki sites (Figure 8). Lower Valkeajoki has the overall smallest and Vitsajoki the largest substrate diameters. A lot of variation in the water velocity is found within every sampling site (Figure 9).

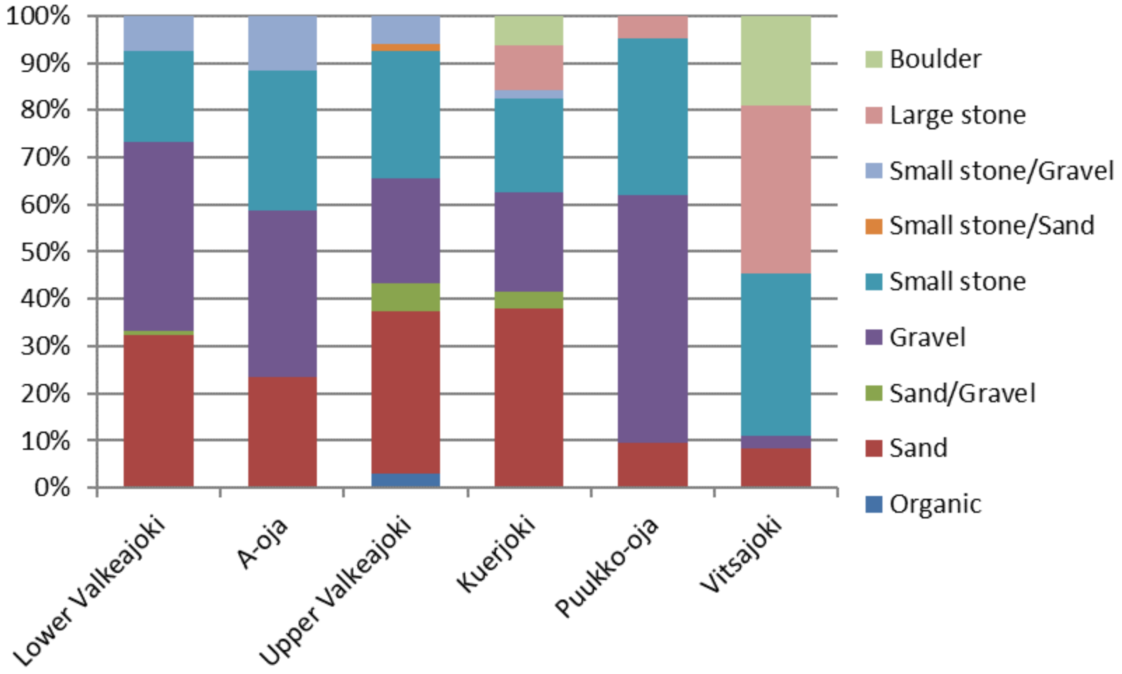

Figure 8. Proportion of sampling points by modal bottom substrate and sampling site.

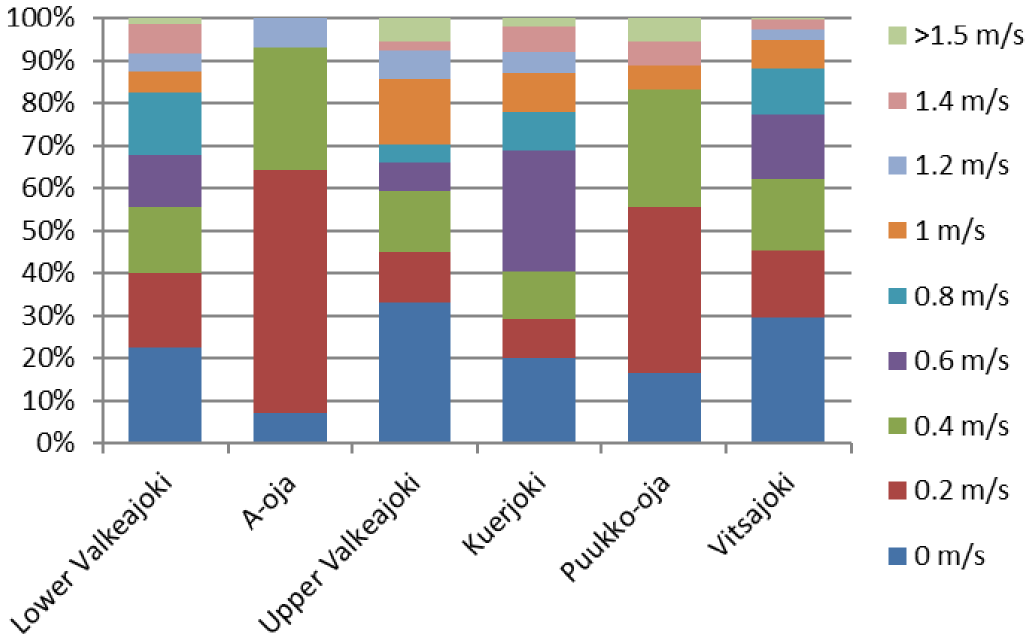

Figure 9. Proportion of sampling points in each water velocity class (rounded to the nearest 0.2 m/s) by sampling site.

The largest areas with plain bottom are found in the Lower and Upper Valkeajoki sites, while Kuerjoki and Vitsajoki sites have the highest coverage of macrophytes (Figure 10). Water moss is by far the most abundant macrophyte, while sedges and water-milfoils are found only in minority of sampling points.

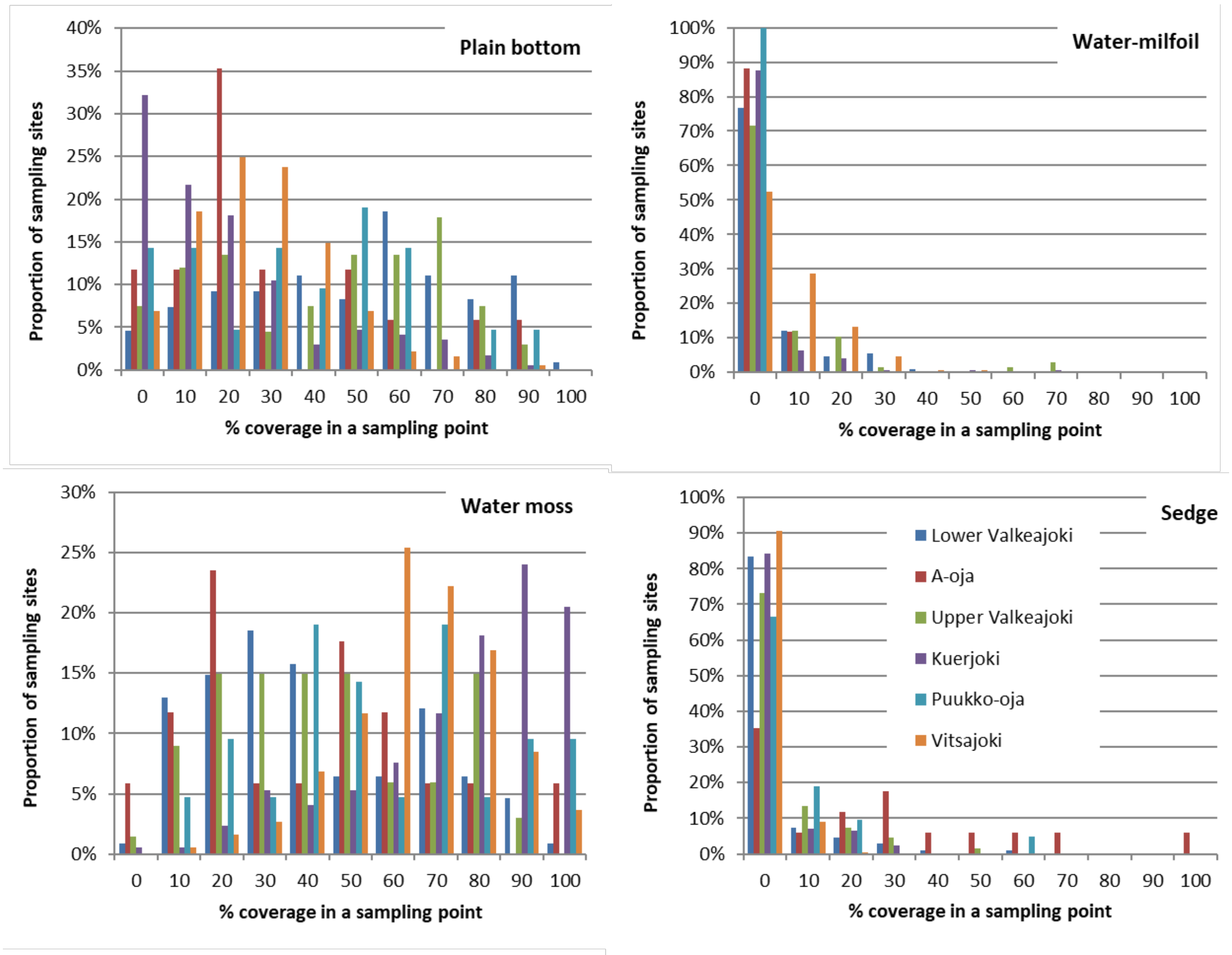

Figure 10. Proportion of sampling points with % coverage by bottom vegetation type (and plain bottom) and by sampling site.

Water temperatues varied clearly less than sediment temperarures between the sampling points (Figure 11). In the Valkeajoki Upper and Vitsajoki sites temperature measurements indicate numerous points of in-channel groundwater upwelling, while in Kuerjoki and especially Puukko-oja sediment temperatures are closer to the ambient water temperature. The fairly large variation in water temperatures measured at Kuerjoki site indicates existence of the confluence plume from Puukko-oja, but it also indicates the fact that field data at Kuerjoki was gathered over two long days with varying air temperature. In the Lower Valkeajoki and A-oja sites water temperatures varied very little (between 9.2–10.2 degrees and 9.6–9.7 degrees, respectively). Although the confluence plume created by A-oja brook flowing into Valkeajoki is relatively well groundwater fed, the main river Valkeajoki at the same reach is also much fed by groundwater sources located above the confluence. Therefore, at Lower Valkeajoki site the small-scale spatial differences appear small in terms of the groundwater influence.

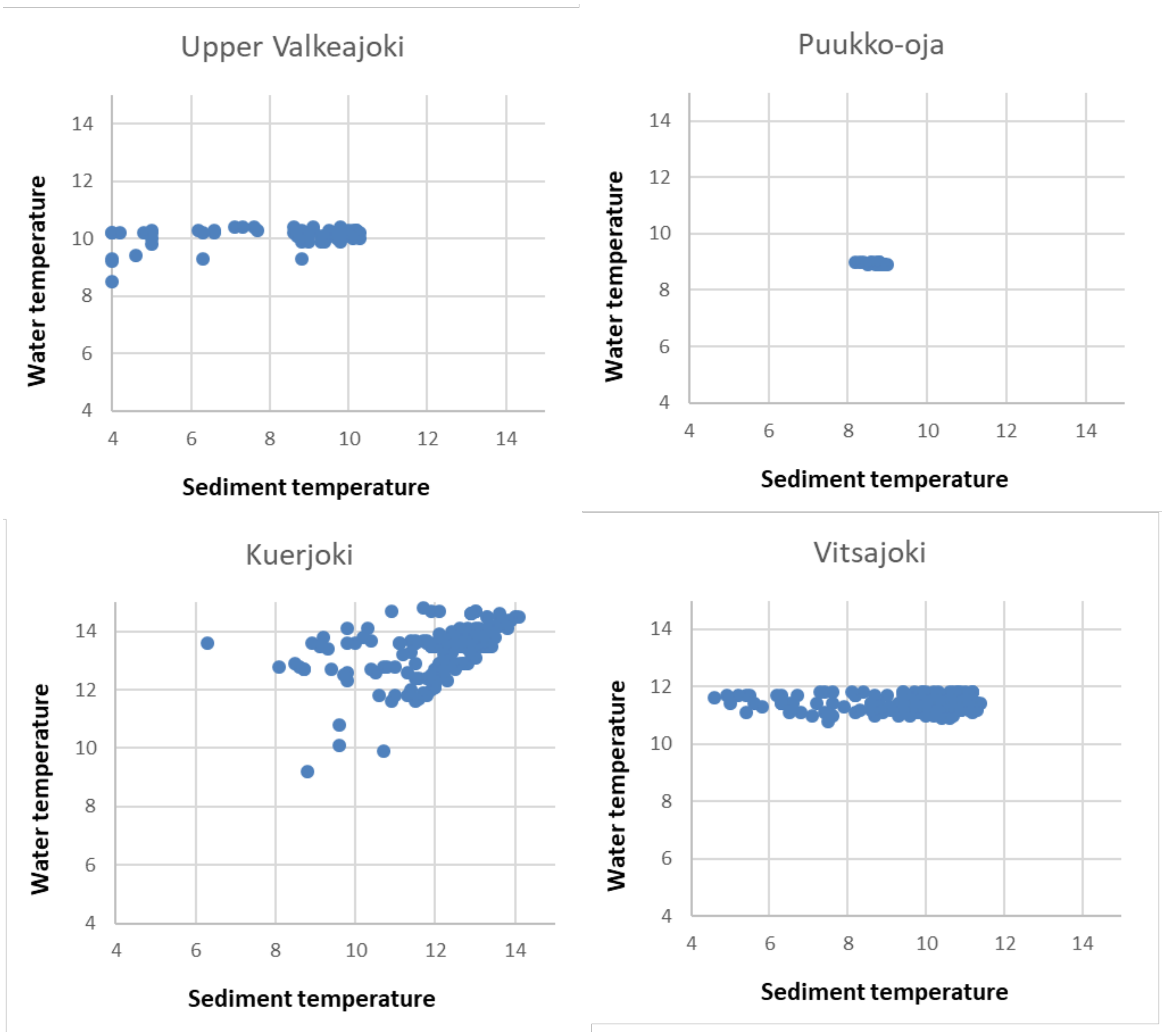

Figure 11. Sediment temperature vs. corresponding water temperature in each sampling point shown by sampling site. No sediment temperature data exists from the Valkeajoki Lower and A-oja sites.

Trout and sculpin occurrence with information on sediment temperature

The self-organing map (explained variance 42.01%; four clusters, Davies Boulding index 0.50) shows that trout generally prefers sampling points with low water temperatures, low sediment temperatures, low river flow, high oxygen levels, small distance from the river bank, low water depth, and sampling points where both sediment temperature and water temperature are colder than the average water temperature of the same horizontal sampling line and plain bottom (Figure 12). Sculpin generally prefers sampling points with mediocre trout densities and low water temperatures.

The random forests model (cv-accuracy: 86.55%) suggests that trout occurrence is higher in lower sediment temperatures, lower river flow, nearer to the river bank and in sampling points where sediment temperature is colder than the average temperature of the same horizontal sampling line (Figure 13). That is, the key findings of the random forests model underline the key findings of the self-organizing map model.

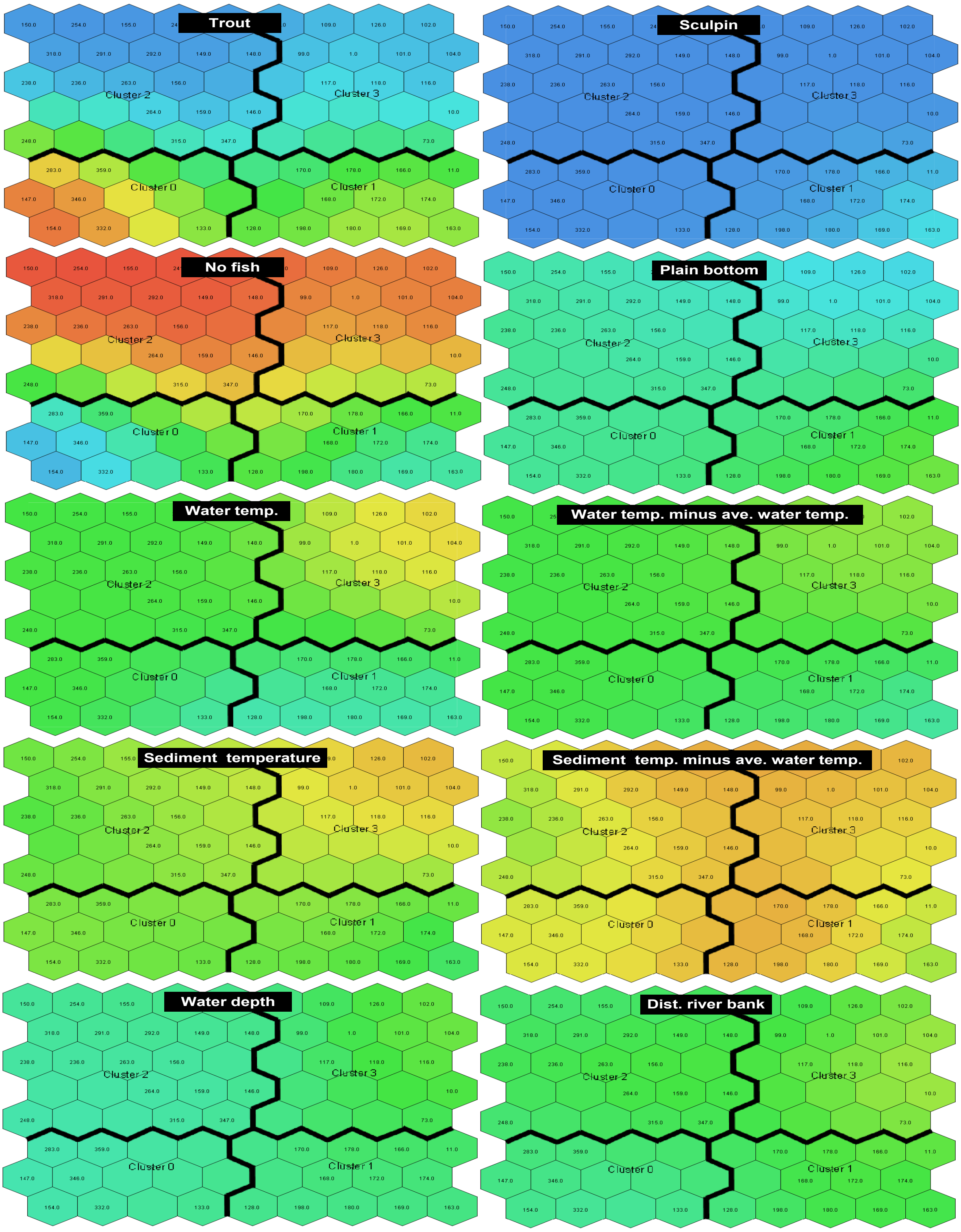

Figure 12. Self-organizing maps of most important variables with four clusters separated by thick black lines in each figure. More reddish color refers to a higher value of a variable and more bluish color refers to a smaller value of a variable (range from min to max:). For example in the top-leftmost figure, greenish and reddish clusters 0 and 1, respectively, are higher trout occurrence sampling point clusters. Each sample (id, row) remains (stays) in the same SOM node (cell) in each figure. The sample size of sculpin was low (63) and hence the general coloring of the top-rightmost sculpin figure is bluish.

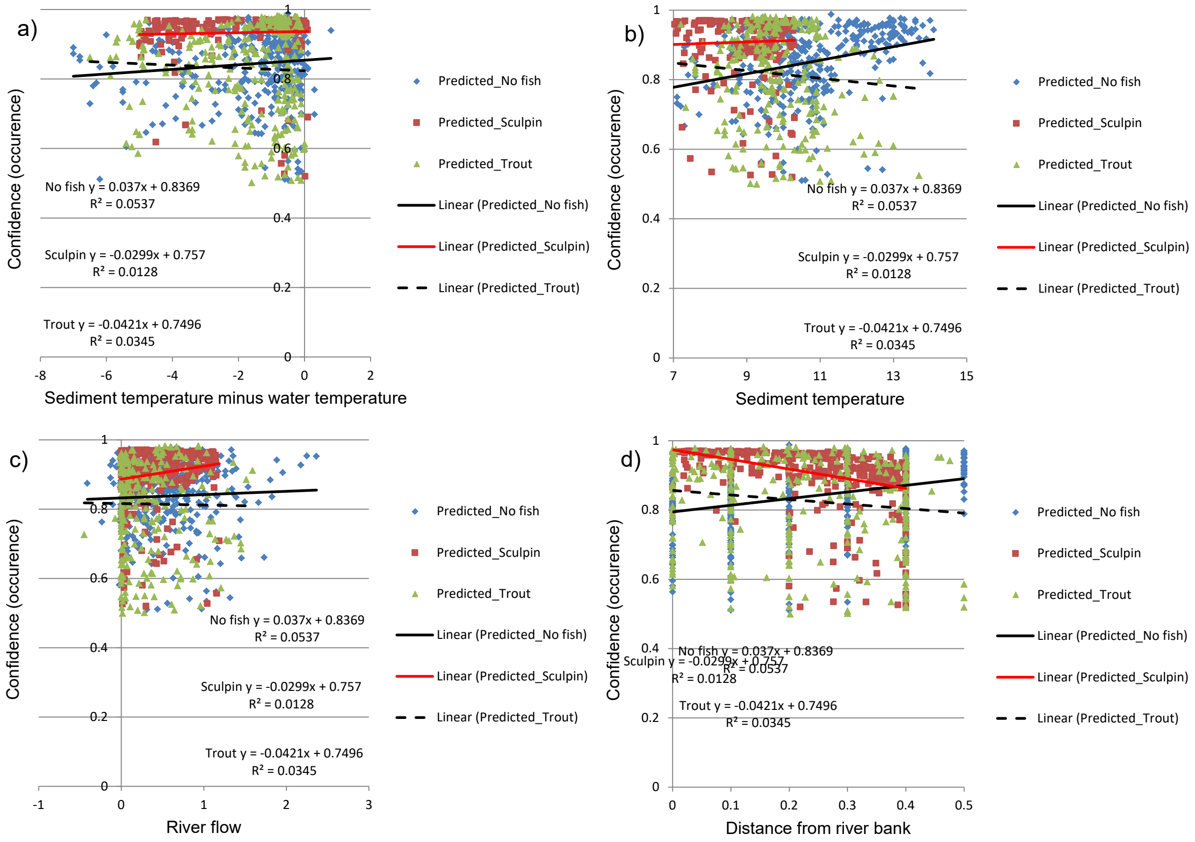

Figure 13. Random forests model based confidences of occurrence for trout, sculpin and no fish in relation to a) sediment temperature minus average water temperature of the same horizontal sampling line, b) sediment temperature, c) river flow, and d) distance from the river bank. In general, the confidence of trout occurrence is higher in sampling points where sediment temperature is colder than the average temperature of the same horizontal sampling line, in colder sediment temperatures, lower river flow and lower distance from the riverbank.

Trout and sculpin without information on sediment temperature

The self-organing map (explained variance 42.72%; four clusters, Davies Boulding index 0.50) shows that trout generally prefers sampling points with low water temperatures, low river flow, high oxygen levels, small distance from the river bank, low water depth, and sampling points where water temperature is colder than the average temperature of the same horizontal sampling line and plain bottom (Figure 14).

The random forests model (cv-accuracy: 78.57%) suggests that trout occurrence is higher in colder water temperatures, nearer to the riverbank, lower river flow and in sampling points where water temperature is colder than the average temperature of the same horizontal sampling line (Figure 15). That is, the key findings of the random forests model again underline the key findings of the self-organizing map model.

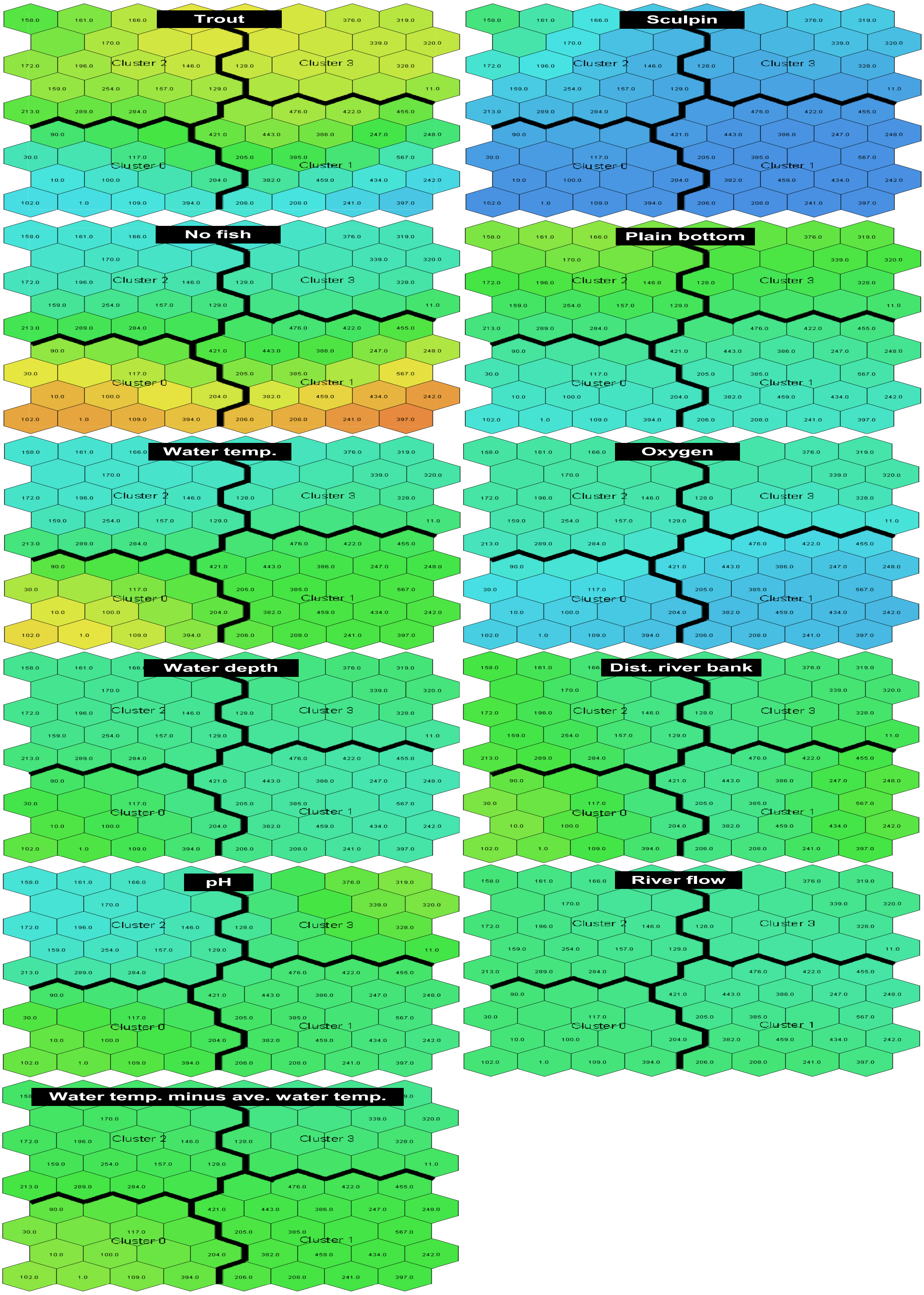

Figure 14. Self-organizing maps of most important variables with four clusters separated by thick black lines in each figure. More reddish color refers to a higher value of a variable and more bluish color refers to a smaller value of a variable (range from min to max:). For example in the top-leftmost figure, more greenish clusters 2 and 3, respectively, are higher trout occurrence sampling point clusters. Each sample (id, row) remains (stays) in the same SOM node (cell) in each figure.

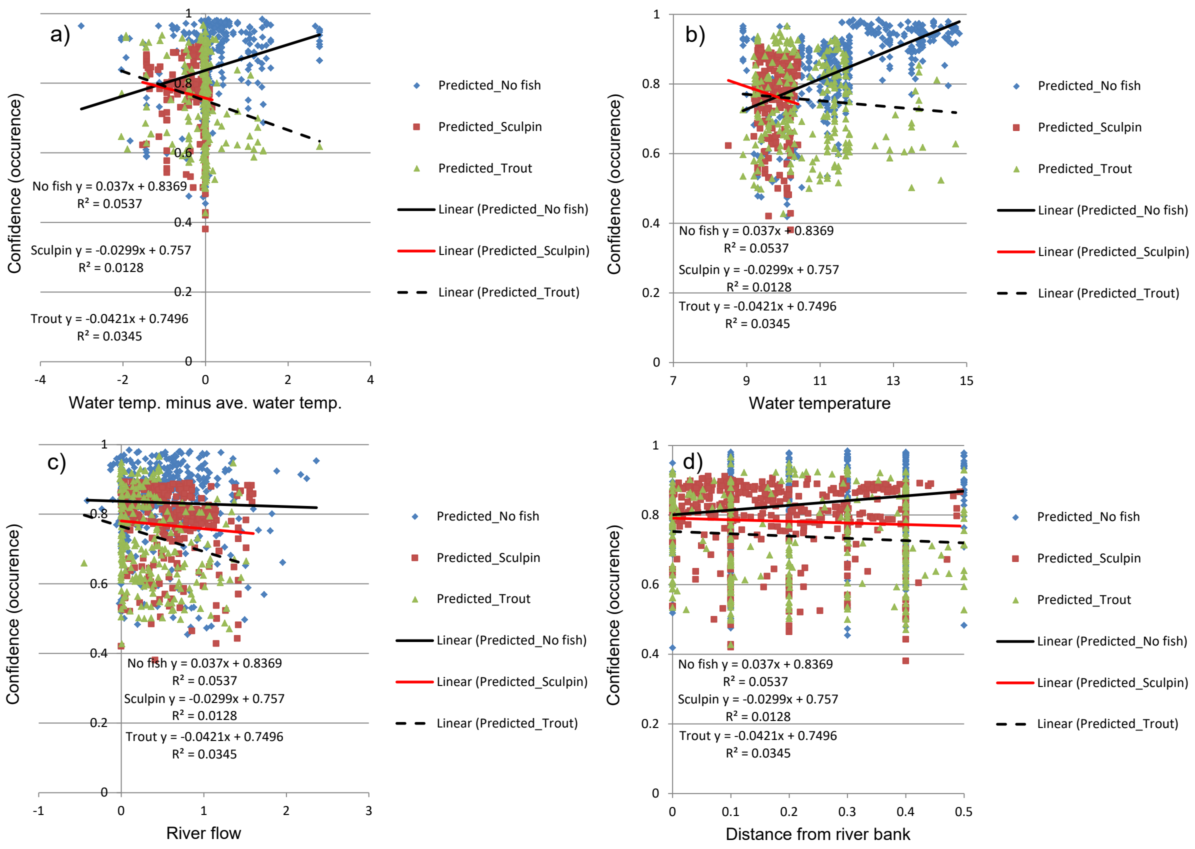

Figure 15. Random forests model based confidences of occurrence for trout, sculpin and no fish in relation to a) water temperature minus average water temperature of the same horizontal sampling line, b) water temperature, c) river flow and d) distance from the river bank. In general, the confidence of trout occurrence is higher in sampling points where water temperature is colder than the average temperature of the same horizontal sampling line, in colder water temperatures, lower river flow and lower distance from the riverbank.

5. Discussion

Our study area is located in the Finnish Lapland, where summertime temperatures are typically low. Air temperatures were low also during the data collection, although the field period was purposely timed to August, when summer temperatures still prevail, but YOY trout have grown large enough to be caught by electrofishing. The chosen rivers are generally strongly fed by groundwater, which together with cool weather kept river water temperature low. Our finding that trout prefer the coldest spots (below 10 degrees) even in these temperatures is especially interesting, given that the temperature optimum for trout growth is found at temperature above 10 degrees (Elliot 1975, Jensen 1990, Elliot *et al.* 1995).

The key findings obtained using two model types and four model runs are highly in concordance in that trout occurrence is highest in cold sediment temperatures, cold water temperatures, near to the river bank and in sampling sites where both sediment temperature and water temperature are colder than the average temperature of the same horizontal sampling line. The results seem robust also to the used subsets of data (with/without sediment temperature data).

Giraudel and Lek (2001) compared SOM and some conventional statistical methods (e.g. PCA, NMDS) for ecological community ordination and found specific benefits of using SOM. For instance SOM allows the visualisation of interspecific association even if it differs in different parts of the data space (as in our study), and the sample units and the species abundance can be seen in the same figure and hence the analysis is becoming easier. In addition, SOM averages data and thus removes noise (Vesanto *et al.*, 1998). If outliers exist in the dataset, each of them affects only one map unit and its neighbourhood. The other areas of the map are not affected by these data (Kaski, 1997).

Cutler *et al.* (2007) listed advantages of RF compared to other statistical classifiers used in ecology that include (1) very high classification accuracy; (2) a novel method of determining variable importance; (3) ability to model complex interactions among predictor variables; (4) flexibility to perform several types of statistical data analysis, including regression, classification, survival analysis, and unsupervised learning; and (5) an algorithm for imputing missing values. Very high classification accuracy was also highlighted by Fernandez-Delgado *et al.* (2014) who found that among 179 classifiers arising from 17 methodological families, the best classifiers were different versions of RF. In the present study the statistical performance of both RF models (its 10-fold cross-validation accuracy) were also good (RF: 86.55% and 78.57%).

We were not able to measure exact microhabitat characteristics of trout, but habitat variables and presence/absence of trout were recorded with a resolution of 1 sqm spatial units. Neither did we measure habitat availability, which together with the test fishing data would allow for construction of habitat preference curves. Nevertheless, our analyses indicate similar general habitat choices for trout as found out in various other studies: trout inhabit mostly shallow near-bank (or narrow channel) habitat with low to moderate water flow, and substrates dominated by small particle sizes (reviewed by e.g. ICES 2011).

Temperature differences in the study area can only arise from different amounts of groundwater presence. Our analysis is able to remove the effect of measured correlated habitat parameters, thus occurrence of groundwater alone, directly or indirectly, seems to attract trout. We have no hypothesis, why trout seem to favour strong presence of groundwater. However, some studies suggest that variation in groundwater upwelling can affect the distribution and abundance of submerged macrophytes (e.g. Loeb and Hackley 1988, Lillie and Barko 1990). Frandsen *et al.* (2012) demonstrated that groundwater seepage stimulates the growth of aquatic macrophytes. If places of groundwater upwelling have more macrophytes, they offer more hiding places which are important for trout parr. In our study we collected information about macrophytes, but their impacts on occurrence of trout seem to be ambiguous, also somewhat varying by the macrophyte species in question. The resolution of the data we collected may in this case not be good enough, e.g. because measurements were largely based on interpretation of aerial photos. Apart from the possible effects of groundwater on macrophytes, the species composition of benthic macroinvertebrate communities seems to be affected by the presence of groundwater (e.g. Sun *et al.* 2019). Thus, it is possible that groundwater upwelling points may locally change the bottom fauna and affect trout distribution via prey species preference. Consequently, it is possible that the preference of trout to groundwater presence is actually caused by some factor(s) correlated with groundwater presence rather than groundwater *per se*.

Presence of groundwater seemed to attract trout also on the mesohabitat level, which is indicated by the highest trout occurrence found in the study sites with the coldest overall water temperatures (Valkeajoki and A-oja sites). The mesohabitat effects on trout may arise from, e.g., the fact that groundwater provides baseflows in summer and winter and at times may provide moderated habitat conditions and possibly refugia for salmonids (Heggenes *et al.* 2011).

Groundwater influx may have negative effects to egg incubation if the groundwater has low oxygen content. Hypoxic groundwater may also have high concentration of harmful dissolved components, which may damage eggs. Nika (2011) found that upwelling of hypoxic groundwater increased mortality among trout eggs and caused premature fry emergence after hatching. He also found sea trout to prefer spawning sites with downwelling of surface water into the streambed, instead of groundwater upwelling sites. In this study we were not able to study selection of trout spawning sites against groundwater upwelling, because too few spawning nests were observed in the study sites in order to carry out a proper data analysis. Most of the trout we caught were YOY parr. By August these parr could have moved far away from their hatching places and therefore their locations do not tell much about the locations of spawning nests. However, our results do not indicate any avoidance of groundwater even among the youngest trout. This may be due to different characteristics of the groundwater in our study area, compared to that of Nika's study. Also, trout parr are typically dwelling on the water column (on bottom surface) rather than inside the bottom substrate, which may make the difference to what kind of groundwater (hypoxic or oxidized) they are exposed to. Indeed, as our results indicate preference of trout to high oxygen contents and at the same time preference to high groundwater presence, it is likely that the groundwater of our study sites is well oxidized when found in water column.

The results indicate partial seggregation of sculpins from trout against the measured habitat characteristics. Sculpins seem to favour, among other things, lower Ph, higher oxygen contents and deeper water than trout. Also, sculpins did not favour the coldest spots with the highest groundwater upwelling. However, sculpins are not present at all in the Kuerjoki, Puukko-oja and Vitsajoki sites due to the impassable water fall downstream from these study sites (Figure 1). This may affect the results concerning habitat choice of sculpins.

Acknowledgements

Anne Rautio and Kirsti Korkka-Niemi from the Department of Geosciences and Geography, University of Helsinki, have participated in planning of the study and in the data collection. Tommi Karesvuori has interpreted aerial photographs, to identify some of the habitat characteristics. Miska Haapsalo, University of Jyväskylä, mapped trout spawning nests in the study sites.

References

Bergelin, U. & Karlström, Ö. 1985. Havsöringen i sidovattendrag till Torne älvs vattensystem. Fiskeriintendenten, övre norra distriktet. Meddelande nr 5. 36 s.
Bowlby, J.N. & Roff, J.C. 1986. Trout biomass and habitat relationships in Southern Ontario streams. Transactions of the American Fisheries Society 115: 503–514.
Breiman, L. 1996. Bagging predictors. Machine Learning 24: 123−140.
Breiman, L. 2001. Random Forests. Machine Learning. 45 (1): 5–32.
Chang, Y.W. & Lin, C.J. 2008. Feature ranking using linear SVM. In Journal of Machine Learning Research (JMLR) W&CP, volume 3, pages 53-64, WCCI2008 workshop on causality, Hong Kong, June 3–4 2008.
Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. 2002. SMOTE: Synthetic Minority Over-Sampling Technique. J. Artificial Intelligence Research, vol. 16, pp. 321–357.
Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A. & Hess, K.T. 2007. Random forests for classification in ecology. Ecology, 88, pp. 2783–2792.
Dugdale, S.J. 2016. A practitioner's guide to thermal infrared remote sensing of rivers and streams: recent advances, precautions and considerations. WIREs Water 2016. doi: 10.1002/wat2.1135.
Elliot, J.M. 1975. The growth rate of brown trout (*Salmo trutta* L.) fed on maximum rations. J Anim Ecol 44, 805-821.
Elliott, J.M., Hurley, M.A. & Fryer, R.J. 1995. A new, improved growth model for brown trout, S*almo trutta.* Functional Ecology 9, 290–298.
Fernandez-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. 2014. Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine Learning Research, 15(1), pp. 3133–3181.
Frandsen, M., Nilsson, B. Engesgaard, P. & Pedersen, O. 2012. Groundwater seepage stimulates growth of aquatic macrophytes. Freshwater Biology 57(5): 907–921.
Grenouillet, G., Buisson, L., Casajus, N. & Lek, S. 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography 34: 9−17.
Guo, C., Lek, S., Ye, S., Li, W., Liu, J. & Li, Z. 2015. Uncertainty in ensemble modelling of large-scale species distribution: effects from species characteristics and model techniques. Ecol Model 306: 67−75.
Davies, D.L. & Bouldin, D.W. 1979. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2): 224–227.
Gosselin, M.-P., Maddock, I. & Petts, G. Mesohabitat use by brown trout (Salmo trutta) in a small groundwater-dominated stream. River Res. Applic. 28: 390–401.
Heggenes, J., Bremset, G. & Brabrand, Å. 2011. Groundwater, critical habitats, and behaviour of Atlantic salmon, brown trout and Arctic char in streams. – NINA Report 654, 28 pp.
ICES. 2011. Study Group on data requirements and assessment needs for Baltic Sea trout (SGBALANST), 23 March 2010 St. Petersburg, Russia, By correspondence in 2011. ICES CM 2011/SSGEF:18. 54 pp.
Ikonen, E., Jutila, E., Koljonen, M-L., Pruuki, V. & Romakkaniemi, A. 1985. Tornionjoen vesistön meritaimenkantojen tila, geneettiset erot ja viljelytarpeet. Riista- ja kalatalouden tutkimuslaitos, Monistettuja julkaisuja nro 57. 103 s.
Jensen, A.J. 1990. Growth of young migratory brown trout *Salmo trutta* correlated with water temperature in Norwegian rivers. J Anim Ecol 59, 603–614.
Jonsson, B. 1985. Life history patterns of freshwater resident and sea-run migrant brown trout in Norway. Transactions of the American Fisheries Society 114: 182–194.
Kaski, S. 1997. Data exploration using self-organizing maps. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82.
Kohonen, T. 1982. Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics 43 (1): 59–69. doi:10.1007/bf00337288.

Kohonen T. Self-Organizing Maps, Springer Verlag, 2001.
Kohonen, T. 2014. MATLAB Implementations and Applications of the Self-Organizing Map, Unigrafia Oy, Helsinki, Finland.
Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the fourteenth international joint conference on artificial intelligence. pp. 1137–1143. Morgan Kaufmann. San Mateo.
Lillie R.A. & Barko J. 1990. Influence of sediment and groundwater on the distribution and biomass of Myriophyllum spicatum in Devils Lake, Wisconsin. Journal of Freshwater Ecology, 5, 417–426.
Loeb S.L. & Hackley S.H. 1988. The distribution of submerged macrophytes in lake Tahoe, California and Nevada, and the possible influence of groundwater seepage. Verhandlungen der Internationale Vereinigung für Theoretische und Angewante Limnologie, 23, 1927–1933.
Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. 2009a. Evaluation of consensus methods in predictive species distribution modelling. Divers Distrib 15: 59−69.
Marmion, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. 2009b. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol Model 220: 3512−3520.
Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M. & Euler, T. 2006. Yale: Rapid prototyping for complex data mining tasks. Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining (KDD-06).
Nika, N. 2011. Reproductive ecology and success of sea trout Salmo trutta L. In a small lowland stream of western Lithuania. Dissertation, Klaipeda University, Lithuania, 52 p.
Palm, S., Romakkaniemi, A., Dannewitz, J., Jokikokko, E., Pakarinen, T., Huusko, R., Broman, A. & Sutela, T. 2019. Tornionjoen lohi-, meritaimen- ja vaellussiikakannat – yhteinen ruotsalais-suomalainen biologinen selvitys sopivien kalastussääntöjen arvioimiseksi vuodelle 2019 (Torneälvens bestånd av lax, havsöring och vandringssik – gemensamt svensk-finskt biologiskt underlag för bedömning av lämpliga fiskeregler under 2019). Sveriges lantbruksuniversitet, Luonnonvarakeskus. 55 p.
Sun, Y., Takemon, Y. & Yamashiki, Y. 2019. Freshwater spring indicator taxa of benthic invertebrates. Ecohydrology & Hydrobiology, https://doi.org/10.1016/j.ecohyd.2019.02.003.
Vesanto, J., Himberg, J., Siponen, M. & Simula, O. 1998. Enhancing SOM based data visualization. In Proceedings of the International Conference on Soft Computing and Information/Intelligent Systems (IIZUKA'98), pages 64-67, Iizuka, Japan.
Vesanto, J. & Alhoniemi, E. 2000. Clustering of the self-organizing map. IEEE Trans. Neural Netw., vol. 11, no. 3, pp. 586–600, May 2000.
Vähä, V., Romakkaniemi, A., Ankkuriniemi, M., Keinänen, M., Pulkkinen, K. & Mäntyniemi, S. 2007. Lohi- ja meritaimenkantojen seuranta Tornionjoessa vuonna 2006 (Monitoring of the salmon and trout stocks in the River Tornionjoki in 2006). Kala- ja riistaraportteja 405: 1–51 + liitteet. (in Finnish with extended English summary).
Zorn, T.G., Seelbach, P.W. & Wiley, M.J. 2002. Distributions of Stream Fishes and their Relationship to Stream Size and Hydrology in Michigan's Lower Peninsula. Transactions of the American Fisheries Society 131:70–85.

luke.fi

Natural Resources Institute Finland
Latokartanonkaari 9
FI-00790 Helsinki, Finland
tel. +358 29 532 6000