Journal of Archaeological Science: Reports 41 (2022) 103330

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Comparing contemporaneous hunter-gatherer and early agrarian settlement systems with spatial point process models: Case study of the Estonian Stone Age

Kaarel Sikk [a,*], Geoffrey Caruso [b], Alar Rosentau [c], Aivar Kriiska [d]

[a] Luxembourg Centre for Contemporary and Digital History (C^{2}DH), University of Luxembourg, Maison des Sciences Humaines, Campus Belval 11, porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg
[b] Department of Geography and Spatial Planning, University of Luxembourg, and Luxembourg Institute of Socio-Economic Research, Maison des Sciences Humaines, Campus Belval 11, porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg
[c] Department of Geology, Institute of Ecology and Earth Sciences University of Tartu, CHEMICUM – Ravila 14A, 50411 Tartu, Estonia
[d] Department of Archaeology, Institute of History and Archaeology, University of Tartu, 18 Ülikooli St., 50090 Tartu, Estonia

ARTICLE INFO

Keywords:
Settlement patterns
Point-process model
Hunter-gatherers
Corded Ware Culture
Neolithic transition
Maxent
Landscape archaeology

ABSTRACT

Inductive locational models have been used for decades to map the probability of past settlements and identify the preferred environmental conditions for habitation. In this study we apply inductive modelling to compare the spatial structure of the settlement systems of hunter-fisher-gatherer groups (Narva and Combed Ware Culture) and early agrarian communities (Corded Ware Culture) in Stone Age Estonia.

We conceptualise settlement system formation as a point process and develop a first order point process model representing the environmental suitability for habitation based on geomorphological, soil and proximity to water. We use MaxEnt and the SDMTune machine learning framework for building the model, variable selection and estimation. The model is applied to the two communities and the effects of the variables and the resulting spatial patterns compared.

The statistical analysis indicated higher predictive power for hunter-fisher-gatherer sites, which might result from higher variety of agrarian activities, different socio-economic organization or effects of spatial structure of the landscape.

The spatial comparison indicates significant differences between the suitable environments for habitation between the two groups. While the hunter-fisher-gatherer population had an entirely shoreline connected settlement system the Corded Ware people inhabited the areas further away from water bodies. This resulted in significantly expanded potential space with differing spatial configuration for the incoming agrarian groups, possibly allowing tolerated immigration. The results also indicate there was a certain overlap of areas considered suitable habitation by both cultural groups, which might have caused a competition for land.

1. Introduction

The tradition of creating inductive locational models of archaeological sites spans several decades. These are mostly regression models based on environmental covariates which have been used to search for areas with a high probability of site occurrence (e.g. Mehrer et al., 2005; Kvamme et al., 2005; Verhagen and Whitley, 2012). The explanatory interpretation of model results tends to be limited to the exploration of individual variables and descriptions of the geographical distribution of model outputs.

The goal of this paper is to present a spatial inductive model exploration, quantitative comparison and comparative interpretation of modelling results. The case study is based on Stone Age Narva culture (NW), Comb Ceramic culture (CWC) and Corded Ware culture (COCW) in the Estonian area. These cultures represent a timespan of c. 1000 years, with early agrarian society migrating and inhabiting a region with existing hunter-gatherer cultural groups. The cultures (e.g. Jaanits, 1955; Jaanits et al., 1982; Kriiska, 2000; Lõugas et al., 2007; Kriiska et al., 2017; Kriiska et al., 2020) and their different principles of settlement choice (Sikk et al., 2020a) have been thoroughly studied in the

* Corresponding author.
E-mail address: kaarel.sikk@uni.lu (K. Sikk).

https://doi.org/10.1016/j.jasrep.2021.103330
Received 11 June 2021; Received in revised form 15 November 2021; Accepted 20 December 2021
Available online 3 January 2022

region. So far there has been no comparative spatial research of the two populations.

We used environmental data from combined Narva culture and Comb Ware culture (NW-CWC) sites and COCW sites. For a region in northern Estonia we created spatial environmental suitability rasters for residence and compared them with each other (Whitford, 2019; Vidal-Cordasco and Nuevo-López, 2021). To provide environmental data for modelling we used available Holocene relative sea-level (Rosentau et al., 2013; Muru et al., 2017), geomorphological and soil data and built palaeo reconstructions of the region for two different time periods.

Landscape surveys for archaeological sites have led to the observation that locations of sites from certain periods are easier to find than others. In Estonia, NW-CWC settlement sites have been methodically found using landscape-based survey tactics while COCW sites are mostly found "by accident" (e.g. Kriiska, 2000). We can state that archaeologists' implicit mind models can predict the locations of NW-CWC sites but not COCW sites. This provides a strong basis to assume very different settlement systems that require explanations beyond just the influence of individual environmental variables.

The models were therefore explored statistically for their variable contribution and performance. We compared their spatial outputs and interpreted the differences using existing knowledge of the settlement patterns for the cultures concerned.

2. Method

Settlement patterns are formed by subsequent habitation events in different locations. In this paper we generalise these locations as points, as is typical practice for archaeological locational models. This allows us to describe settlement patterns as point patterns and their formation as point processes, mechanisms that produce point patterns. Mathematically a point process is a stochastic generation of points in space; in the case of a homogeneous Poisson point process it is determined by intensity λ and results in complete spatial randomness. It is obvious that settlement locations are chosen not as random locations in space but on the basis of a multitude of conditions perceived by settlers. We can call this a non-homogeneous point process and in this case the constant intensity λ is replaced by a deterministic function $\lambda(s)$, where s is any spatial variable (typically environmental) or a set of variables.

To describe the underlying laws, a point process model (PPM) can be constructed and explored. PPMs can be used to characterise point processes and also to distinguish between the effects of first- and second-order properties. The former refers to externally induced global effects (e.g. environment) on the average intensity of points in certain locations and the second to systemic effects emerging from the relationship between the points, or settlements in our case (Bailey and Gatrell, 1995; Crema et al., 2010). In this study we are interested in how the environment influences settlement choice so we can describe it as a heterogeneous Poisson process that includes only first-order effects. The occurrence of the choice event thus remains independent from other events, taking into account only local environmental conditions.

The empirical data allows us to apply a PPM that describes the hypothetical underlying generating mechanism and makes it possible to give theoretical descriptions of phenomena. In the current case we explore how the environment influenced settlement choice and settlement pattern formation. We define the developed PPM as an archaeological environmental effect model (EEM).

To train our EEM models and output them as habitation suitability rasters we use MaxEnt, a machine-learning tool for model training (Phillips et al., 2017b). MaxEnt has been shown (Renner et al., 2015) to be effective in training PPM models in general and for archaeological site prediction in particular, with only minor loss of explanatory power (Yaworsky et al., 2020; Vernon et al., 2020). MaxEnt applies the expectancies comparison principle and maximum entropy principle to find distributions of variable values that are predicted as suitable, in the current case, for habitation (for a detailed explanation see Elith et al., 2011; Wachtel et al., 2020). It estimates the role of environmental covariates in settlement choice, constraining the geographic probability distribution to be as close as possible to absolute entropy, i.e. spatial homogeneity in the case of point patterns.

MaxEnt allows the creation of an empirical density estimate which is converted through a link function (the complementary log–log transform being recommended) into a point process intensity, which is then interpreted as suitability for habitation in a given location (Vernon et al., 2020). Probability rasters are then created indicating suitability for settlement choice in the research area and can be used to explore the spatial configuration of the effects of environmental suitability.

The approach for exploring environmental effects using an EEM is almost identical to the practice of archaeological predictive modelling in terms of environmental covariates and methodology. Because of having only contemporary environmental data and very sparse information about links between sites, the most feasible predictive model for prehistoric archaeology is the first-order PPM.

EEM explicitly describes the influence of the environment on the settlement pattern formation process. It also enables the study of spatial characteristics of specific landscapes and environmental determinism as a distinct analytical category for settlement choice. Some methodology (e.g. niche overlap and niche breadth measures) for the study of spatial characteristics has been developed in the field of ecological species distribution modelling (SDM) and imported to the archaeological domain through eco-cultural niche modelling (ECNM; Banks et al., 2006; Banks, 2017; Whitford, 2019). The modelling practices also use similar data and a similar methodology. For example the MaxEnt toolkit was originally developed for SDM. SDM is based on species with lesser complexity in their socio-economic behaviour and technology use, so their outputs are considered to be almost entirely determined by the environment.

The purpose of ECNM is to explore the concept of ecological niches; it enables the identification and analysis of complex relationships between cultural systems and the niches used. The method includes analysis of the characteristics of niches as spatial units, including the use of comparative measures (Whitford, 2019; Vidal-Cordasco and Nuevo-López, 2021). It is mostly applied on a continental scale (for some recent regional studies, see Whitford, 2019; Vidal-Cordasco and Nuevo-López, 2021) including non-local variables like climate and brings with it the more general concept of niche. Niche acts as an ecological container for species or societies activities. It is not intended to explain experiential decisions of individual choices which are influenced by local environmental features like elevated sandy patches but also socio-cultural interactions. Those can lead to patterns within a particular niche. Therefore on a smaller scale, varied socio-economic behaviour may have a more complex impact on the environmental determinism to choices that could not be explained through the niche framework.

We create EEMs of hunter-fisher-gatherer (NW-CWC) and agrarian (COCW) settlement patterns and compare these models (see Section 4 for description). This comparison describes deterministic environmental features that also lead to spatial differences between the settlement systems. We explore statistical measures and predictive power of the models, the effect of individual environmental covariates and the spatial configuration of the spatially explicit output of the models (Section 5).

3. Archaeological and environmental data

The locational data of this study is based on settlement locations of three archaeological phenomena: Narva culture (5200–3900 BCE), Comb Ceramic Culture (3900–1750 BCE) and Corded Ware culture (2800–2000 BCE) (Kriiska et al., 2020): Fig. 1. The first two groups are hunter-fisher-gatherers with similar subsistence systems. As it has been shown that the environmental characteristics of their settlement sites are virtually indistinguishable (Sikk et al., 2020a) they are grouped together as NW-CWC. We study and compare the location choice of two settlement groups: hunter-fisher-gatherer groups (NW-CWC) and

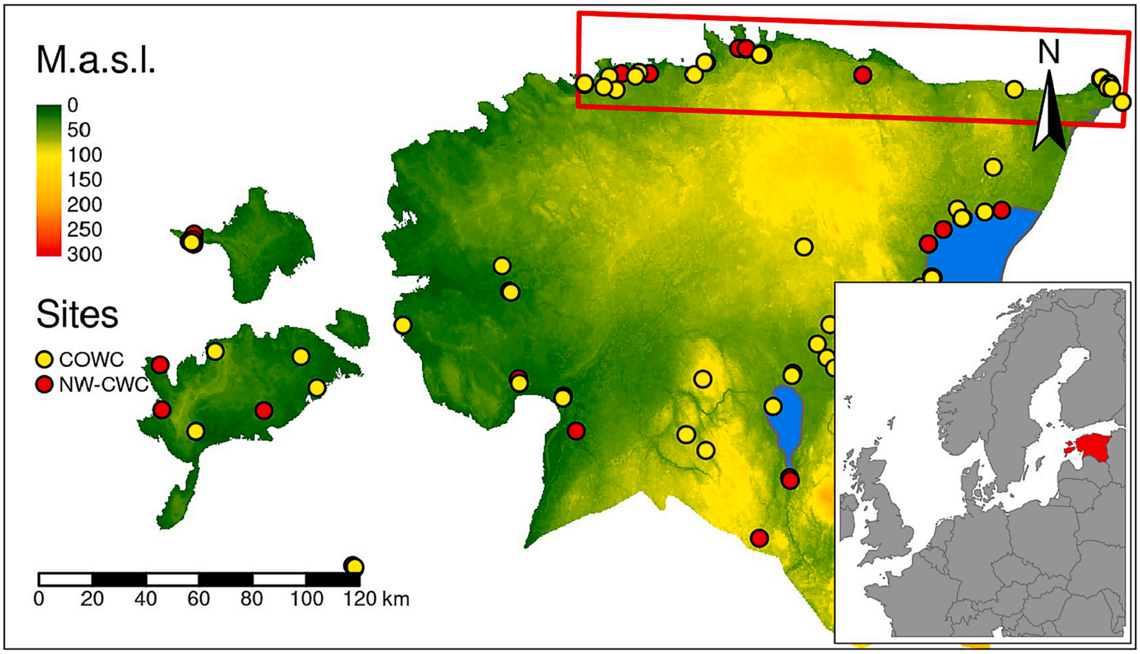

Fig. 1. Overview of the Stone Age archaeological sites used in this research (for detailed site distribution maps see Sikk et al 2020). The spatially explicit model was created based on the area in northern Estonia (surrounded by red boundaries). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

agrarian groups (COCW).

The sites of the phenomena can be distinguished by their use of pottery which also helps to filter out temporary work sites that might not indicate habitation. NW, the first in the region to use pottery, was a hunter-fisher-gatherer subsistence culture with a likely dominant aquatic economy (e.g. Kriiska et al., 2017; Oras et al., 2017). There is believed to have been cultural and genetic continuity from the first inhabitants arriving in Estonia about 9000 cal. BC (e.g. Jaanits et al., 1970; Mittnik et al., 2018; Kriiska et al., 2020). The settlements are situated close to water, on islands, lake shores, river deltas and the beaches of marine lagoons (Sikk et al., 2020a and the literature cited therein).

A similar pattern was used by the hunter-fisher-gatherer subsistence CWC peoples who arrived with a migration from the east (e.g. Jaanits et al., 1982; Saag et al., 2017). Settlement location choice was very similar but the larger area contained several specialised and central sites (Kriiska et al., 2020).

COCW developed in Central Europe as a result of immigration from eastern steppe regions of Eastern Europe and also dispersed to Estonia as a result of immigration (Allentoft et al., 2015; Haak et al., 2015; Saag et al., 2017; Saag et al., 2021). The immigration took place at a time when CWC was also present in Estonia, resulting in simultaneous habitation (Kriiska et al., 2020). In considered territory COCW was an agrarian or semi-agrarian society (Kriiska, 2000; Lõugas et al., 2007) and despite offering no indication of a new level of social organisation and complexity, it had a very different settlement pattern and settlement choice principles (Sikk et al., 2020a). Its inhabitants built their settlements on the north Estonian klint edge, river floodplains, ancient coastal formations in the Baltic Sea region and small drumlins on the lowlands near Võrtsjärv. While hunter-fisher-gatherer sites have been only found in close proximity to water bodies (Sikk et al., 2020a), COCW sites are also situated further away. There are several cases of COCW people reusing previous hunter-fisher-gatherer sites if the shorelines had receded from the region.

The empirical data informing us about past settlement choices is derived from environmental conditions in the known settlement sites of the two groups. The previously published dataset of Stone Age settlement sites in Estonia (Sikk et al., 2020b) is used as a source of site locations and two groups are selected from the data. There are 99 NW-CWC and 67 COCW settlement locations. The settlements and background environmental locations (n = 10,000) used for model training were chosen from the whole area of Estonia. The background locations were chosen using spatially structured sampling representing subregions with known archaeological sites to avoid model overfitting (Anderson and Raza, 2010).

Each location (sites and background) was associated with environmental data representing environmental influence on settlement choice. The variables used generally belong to three interrelated categories: distance to water, geomorphology and soil. Distance to water was divided into distances to sea, lakes and rivers. The second group consists of various variables describing location's position in relation to landforms. Examples include slope, topographic wetness index, which quantifies locations control of hydrological processes; and topographic position index (TPI_100) which quantifies locations relative elevation in comparison to its surroundings. Third group includes various characteristics of soil including its standard classification, wetness, texture and sandiness (for more details see Sikk et al., 2020a, (Kmoch et al., 2021). Individual variables were chosen based on previous statistical analysis (Sikk et al., 2020a) and during the model building process (see 4.3).

Because the goal of the model is to explain the influence of the natural environment on settlement choice, only variables representing abiotic, non-constructed and non-cultural characteristics were used, omitting any possible human-made environmental features like roads or fields. It should be said that soil data could be influenced by agricultural practices, but for this reason we experimented with a large set of soil variables published by (Kmoch et al., 2021). Digital elevation rasters (5 m resolution) provided by the (Estonian Land Board, 2019) and soil data were acquired for regions with settlement locations and the area of the spatial model output. The digital elevation models (DEMs) were used to create palaeo reconstructions of two distinct moments in time: 3900 BCE, used for NW-CWC settlement reconstruction, and 2500 BCE, used for COCW settlement reconstruction.

Based on the current DEMs, palaeo reconstructions of the past topography and shoreline were created and distance to the sea calculated as a separate raster layer. The palaeogeographical reconstructions for the time slices 3900 BCE and 2500 BCE are based on the GIS approach (Rosentau et al., 2011), with the palaeo-sea-level surfaces being subtracted from the 5x5m LiDAR-DTM (Estonian Land Board, 2019). Initial palaeo-sea-level surfaces were interpolated using a point kriging approach from the database of coastal formations (Saarse et al., 2003) for the Litorina Sea (5500 BCE) and for the modern Baltic Sea (1892–1991 CE) based on sea-level measurements and geodetic data (Ekman, 1996). These were used as reference surfaces, and two new time slices for 3900 BCE and 2500 BCE were interpolated between the reference surfaces considering a linear decay in shoreline tilting and average relative sea levels taken from the shore displacement curves for the Tallinn (Muru et al., 2017) and Narva-Luga areas (Rosentau et al., 2013). For the Tallinn area, the cultural layer thicknesses were also subtracted from the DTM based on data by (Arbeiter, 1993).

The geomorphological layers of the past were derived from DEMs and known water level changes using SAGA GIS 7 (Conrad et al., 2015). For this two spatially explicit residential potential rasters were created for the coastal region of north-eastern Estonia (Fig. 1). The region was chosen because of its general suitability for habitation during both periods as it has several known subregions with existing sites. Most of the study area is unexplored for archaeological sites, potentially raising the possibility of subsequent model validation using landscape surveys.

4. Modelling methods

4.1. Toolkit for training inductive point process model

We used the MaxEnt program with the “SDMTune” model tuning library (Vignali et al., 2020) written in R. SDMTune was used as it provides all the tools needed to run a complete modelling workflow, which in this case included data management for training, testing and validation, data-driven variable selection, model parameter tuning, validation and generation of raster maps of models (Fig. 2).

4.2. Variable selection

One of the significant statistical problems in combined archaeological and environmental data is the collinearity of the predictor variables which can introduce bias into models (Peterson et al., 2011). There is high potential for correlation between environmental variables, especially as geomorphological data is derived from digital elevation models. It is also known that geomorphological variables correlate with soils which are formed as processes influenced by local topographies (Moore et al., 1993). We therefore needed to reduce the statistical collinearity (ridged correlation) between the environmental variables (Fig. 3).

The initial model was created using a set of existing predictor variables (see Table 1); redundant variables were removed while building the model. The data-driven variable selection process was executed using the varSel function implemented in the “SDMTune” library. The function iterates all the variables in the order of maxent percent contribution. If the variable is found to have high Pearson correlation coefficient (R2) with any others a jackknife test is run and one of the variables is removed based on the model performance. The algorithm is iterated until highly correlated variables have been removed (R2 > 7).

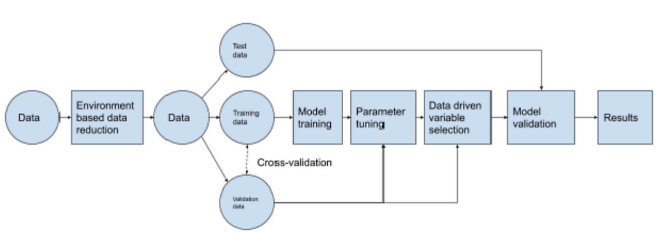

Fig. 2. Workflow diagram used for the modelling process with SDMTune library, based on (Vignali et al., 2020).

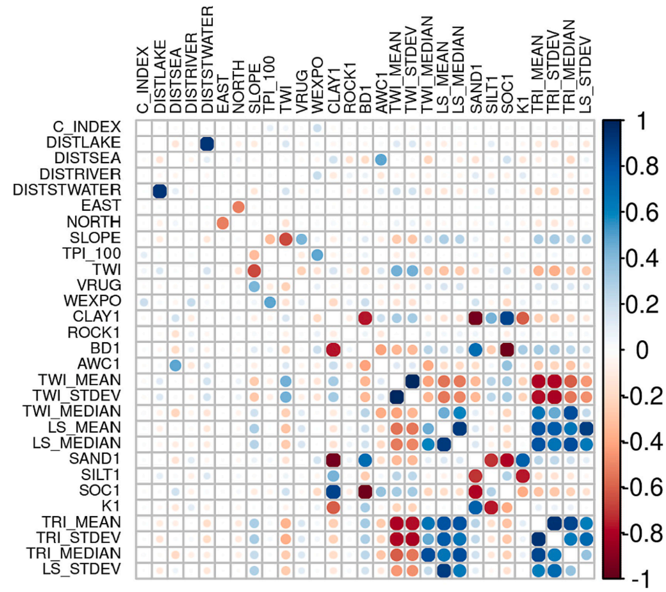

Fig. 3. Correlation matrix of the predictor variables of the background points used for initial data evaluation.

To account for model simplicity, variables not contributing to model performance were also removed using the SDMTune importance-based variable reduction function reduceVar (with threshold value of 1). Some variables which were highly correlated with others and led to lower model performance were also removed manually.

Several candidate models were then evaluated and the model with the best performance was selected. We preferred a model with more general soil categories (WRB main classification) as variables with high numbers of categories led to model overfitting. The resulting set of variables for the models were different for the NW-CWC and COCW groups but both models still included significant variables from all three main variable classes: distance to water, geomorphology and soil (Table 1).

The separation of distances to water by water body types increased model performance. All distances were also included as logarithms of the values, which better reflected intuitive understanding of proximity to water and also significantly increased model performance.

We also experimented with variables based on aggregates of larger regions (soil patches) provided by (Kmoch et al., 2021), but only LS_FACTOR provided significant input to the model. We also tested topographic position index with different ranges (10 m, 50 m, 100 m, 300 m, 1000 m) and based on this chose decided to include only the variable with 100 m range.

4.3. Model hyperparameters and performance and validation

MaxEnt models can be tweaked using several model “hyperparameters” which influence model output. These are the regularisation multiplier and number of iterations while training the model and five feature classes (linear, quadratic, product, hinge and threshold) that can be used in combination (for functional overview see (Morales et al., 2017). The default setup is to use all the feature classes and a regularisation multiplier of 1. To find the hyperparameter set with the best performance while avoiding overfitting we ran the SDMTune function optimizeModel (Vignali et al., 2020). The function uses an evolutionary algorithm to search for the hyperparameter configuration that returns the best metric for prediction. The model configurations for NW-CWC and COCW are displayed in Table 2.

The model data was divided into three groups: training, validation and test datasets each having more than 3000 items. All the modelling steps were evaluated with a cross-validation procedure using training

Table 1
Variables used for modelling with permutation importance in the final models. These values are indicated for the variables that were considered important for describing settlement location choice and included in the final models.

Variable	Name	Permutation importance (NW-CWC)	Permutation importance (COCW)
DISTRIVER	Distance to river (log)	38.40	17.2
DISTSEA	Distance to sea (log)	26.60	9.9
DISTLAKE	Distance to lake (log)	7.40	19.5
WRB_MAIN	WRB soil class	7.40	19.5
WEXPO	Wind exposition	7.10	8.3
VRUG	Vector ruggedness index	6.20	11
LS_MEAN	LS factor (mean of the patch)	2.70	
TWI	Topographic wetness index	2.20	
AWC1	Available water capacity	1.40	4.4
C_INDEX	Convergence index	0.60	0.5
LS_MEDIAN	LS factor (median of the patch)		8.7
SLOPE	Slope		0.6
EAST	Eastness (aspect)		0.3
ROCK1	Rockiness (%)		
SOC1	Soil organic matter		
BD1	Bulk density		
CLAY1	Clay % in soil		
DISTSTWATER	Distance to sea or lake (log)		
K1	saturated hydraulic conductivity		
LS_STDEV	LS factor (stdev of the patch)		
NORTH	Northness (aspect)		
SAND1	Sand % in soil		
SILT1	Silt % in soil		
TPI_100	Topographic position index (100 m range)		
TRI_MEAN	Terrain ruggedness index (mean of the patch)		
TRI_MEDIAN	Terrain ruggedness index (median of the patch)		
TRI_STDEV	Terrain ruggedness index (stdev of the patch)		
TWI_MEAN	Topographic wetness index (mean of the patch)		
TWI_MEDIAN	Topographic wetness index (median of the patch)		
TWI_STDEV	Topographic wetness index (stdev of the patch)		

Table 2
Model configuration hyperparameters used and resulting binary classification threshold values.

Model	Regularisation parameter	Feature classes	Iterations	Thresholds
NW-CWC	1.8	linear, hinge	500	0.09
COCW	2.6	linear, hinge	500	0.21

and validation subsets with 500 iterations. The final model was then evaluated using receiver-operator curves (ROC) and summarised with the AUC (area under the ROC curve) (Merow et al., 2013), which is standard practice for both archaeological predictive models and ecological niche models (Fig. 6).

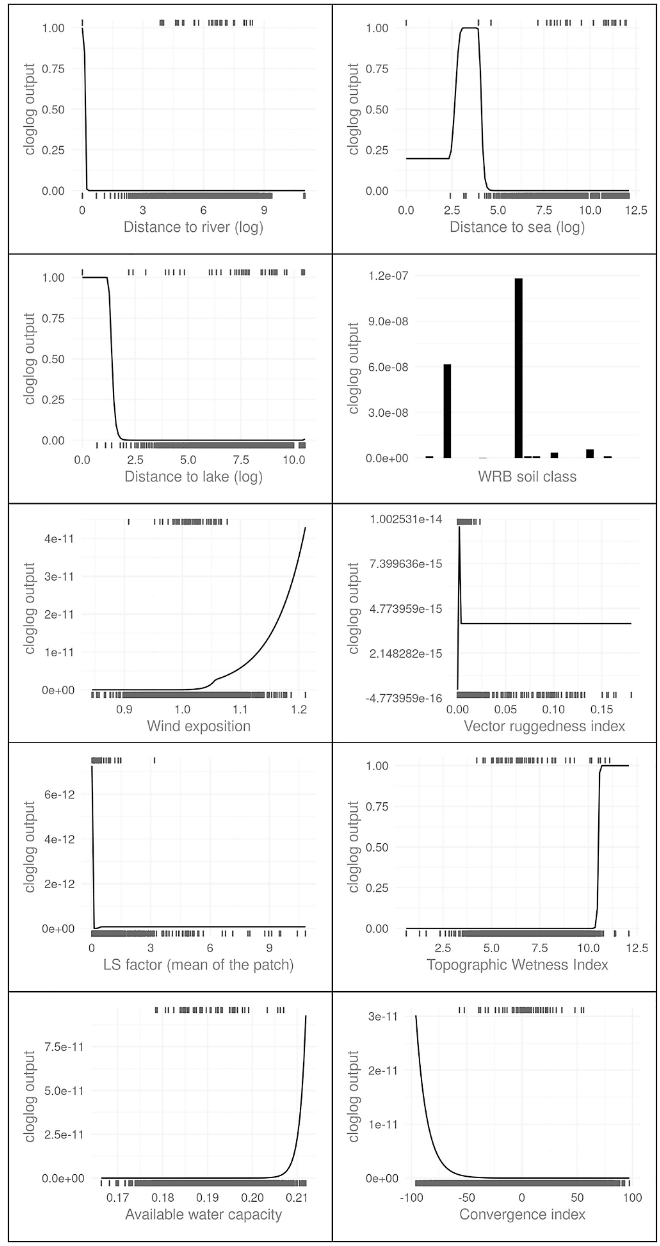

Fig. 4. Marginal response curves for NW-CWC variables.

4.4. Spatial output

We generate a high-resolution raster map of the suitability of the environment for settling. The spatial output is in the 0 to 1 range as a relative probability raster, with 1 indicating highest suitability. The spatially explicit output is generated for a selected region in northern Estonia only (Fig. 1, red rectangle) to benefit the high-resolution study. Raster models of the area were created based on predictor variables for corresponding site groups using complementary log–log link (cloglog; Phillips et al., 2017a) transformation for model prediction. The process resulted in spatially explicit probability rasters with different spatial configurations (Fig. 7). For this study the general suitable area and niche overlap values were computed (see 5.2).

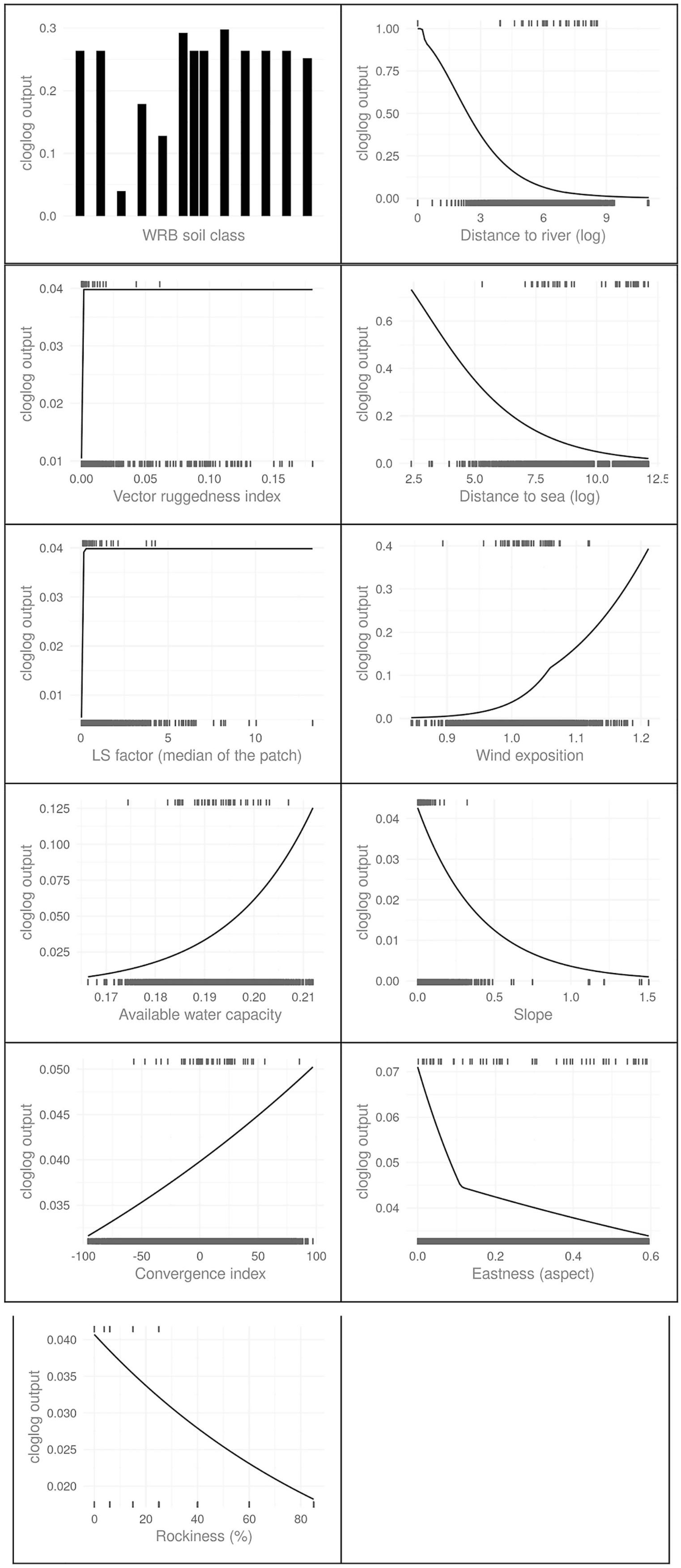

Fig. 5. Marginal response of continuous variables used in the final COCW model.

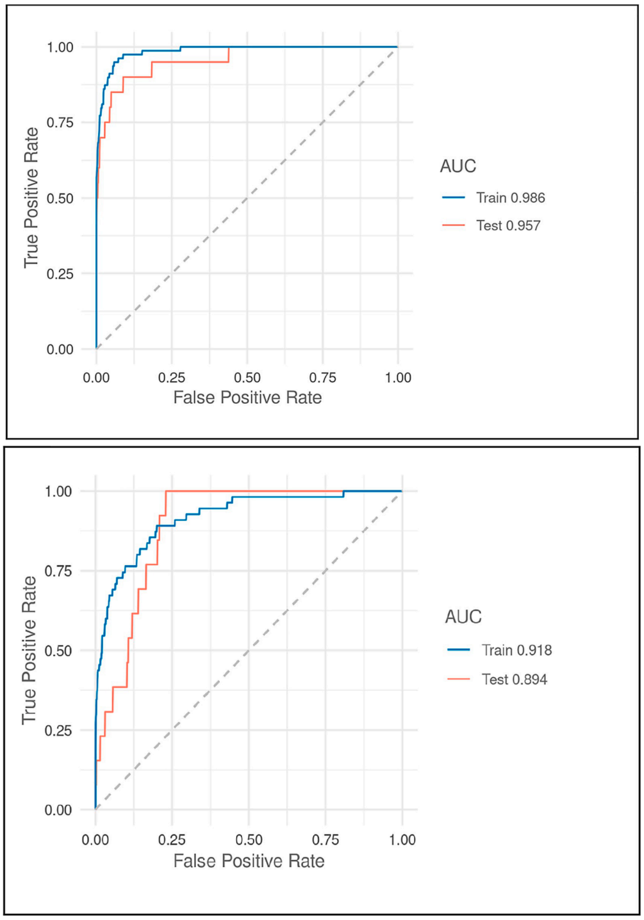

Fig. 6. AUC curves indicating performance of NW-CWC (upper) and COCW (lower) models.

5. Model results

5.1. Variables and their contribution

The contribution of individual environmental variables was measured using the permutation importance metric (Table 1). This assesses the importance of variables by shuffling values and calculating resulting model performance loss. We also explored marginal response curves (Figs. 4, 5) for variables, created by holding all other covariates at their zero-centred means and predicting the density with a MaxEnt PPM using all the data. The results were converted into a relative probability (Vernon et al., 2020) using the complementary log–log transform (cloglog) (Baddeley et al., 2010; Phillips et al. 2017a). Marginal response indicates how potential suitability responds to changes in variables. Variable importance values together with variable response curves informed us about the variable values that determine suitability for habitation (Phillips et al., 2006; Phillips and Dudík, 2008).

It should be emphasised that the impact of variables does not imply direct causation as the same effect could potentially be expressed through a combination of other variables. The relationship between environmental variables and human choices is complex. For example combined slope and (modified) topographic and soil attributes are known to correlate with vegetation (Moore et al., 1993). So it is not always clear whether the landscape form itself is more important for humans or the vegetation or abiotic resources available in specific places. Although a causal link cannot be directly inferred from these measures, they still give insights into the first-order effects behind the point process and also model performance.

Metrics confirmed previous knowledge (e.g. Moora et al., 1935; Jussila and Kriiska, 2004; Sikk et al., 2020a) that hunter-gatherer settlement choice (NW-CWC group) was largely determined by proximity to water; closeness to rivers was the most important factor, making river deltas particularly sought-after places. Soil type at the location and various geomorphological variables had a slight effect. These include wind exposition, which might be an indirect factor, as it is likely that the windiness of the location was not a priority. It might describe the effect of higher, dry locations on the shoreline, which tend to be north- and east-facing and more exposed to wind.

For the COCW settlement this effect of proximity to water was significantly smaller, although proximity to rivers was still important. Geomorphological (wind exposition, slope, median of slope length and steepness factor) and soil (soil type, available water capacity) variables were more important. These variables describe higher, flat, less rugged locations with slopes nearby. The soils indicated as suitable by the model are podzols, regosols and retisols. The model therefore includes klint edges in northern Estonia and river floodplains, both of which have been associated with early agriculture. Previous coastal formations known from empirical data can also be included in the model. The difference in variable importance clearly illustrates the expected shift from shoreline-connected hunter-fisher-gatherer subsistence to an agrarian culture.

The exclusion of topographic position index and the minor influence of topographic wetness index contradicts causal thinking because it is known that dry locations were preferable as habitation sites. Previous research using the same data also showed that this variable on its own had a significant effect on settlement choice (Sikk et al., 2020a). In the modelling process the choice was better described using other correlated variables.

5.2. Predictability

The MaxEnt toolkit provides tools to evaluate models' predictive performance with receiver-operator curves (ROC), which we summarised here with the AUC illustrating the relationship between true and false positive prediction rates (Phillips et al., 2006; Merow et al., 2013). The AUC values were calculated using both training and test datasets.

The statistic quantitatively indicated that model performance could be one of the reasons behind the existing observation that COCW site locations are harder to predict based on current knowledge than NW-CWC sites. During the model development all the model pairs had NW-CWC models with significantly higher AUC values than COCW models. Although the AUC score for the final COCW model was reasonably good it still indicated lower predictive performance (Fig. 6).

5.3. Statistical metrics

To compare the suitability models we calculated two measures: the total suitable area and niche overlap using the MaxEnt toolkit. The general area indicates the region that could have been considered suitable during the periods' eco-cultural systems. It was calculated by converting the general raster model outputs into binary form, indicating suitable and unsuitable pixels using thresholds (see Liu et al., 2013) provided by the MaxEnt system. The suitable pixels were then added together, resulting in a total suitable total area of 100.16 km^2 for NW-CWC and 1201.06 km^2 for COCW. This indicates an area about ten times larger that our model considered suitable to agrarian COCW groups in comparison to hunter-fisher-gatherer NW-CWC groups. These areas can not be taken as absolutes as the area of predicted presence is strongly affected by grain size, but they are sufficient for comparing models of identical environments.

Niche overlap helps to measure the degree of geographical similarity of the two model predictions. We used the I statistic (Warren et al., 2008) provided by the MaxEnt toolkit which reports values between 0 and 1, with 0 indicating no niche overlap and 1 indicating total niche overlap. The I statistic value for the NW-CWC and COCW models was 0.331, which indicates minor to moderate overlap between suitable

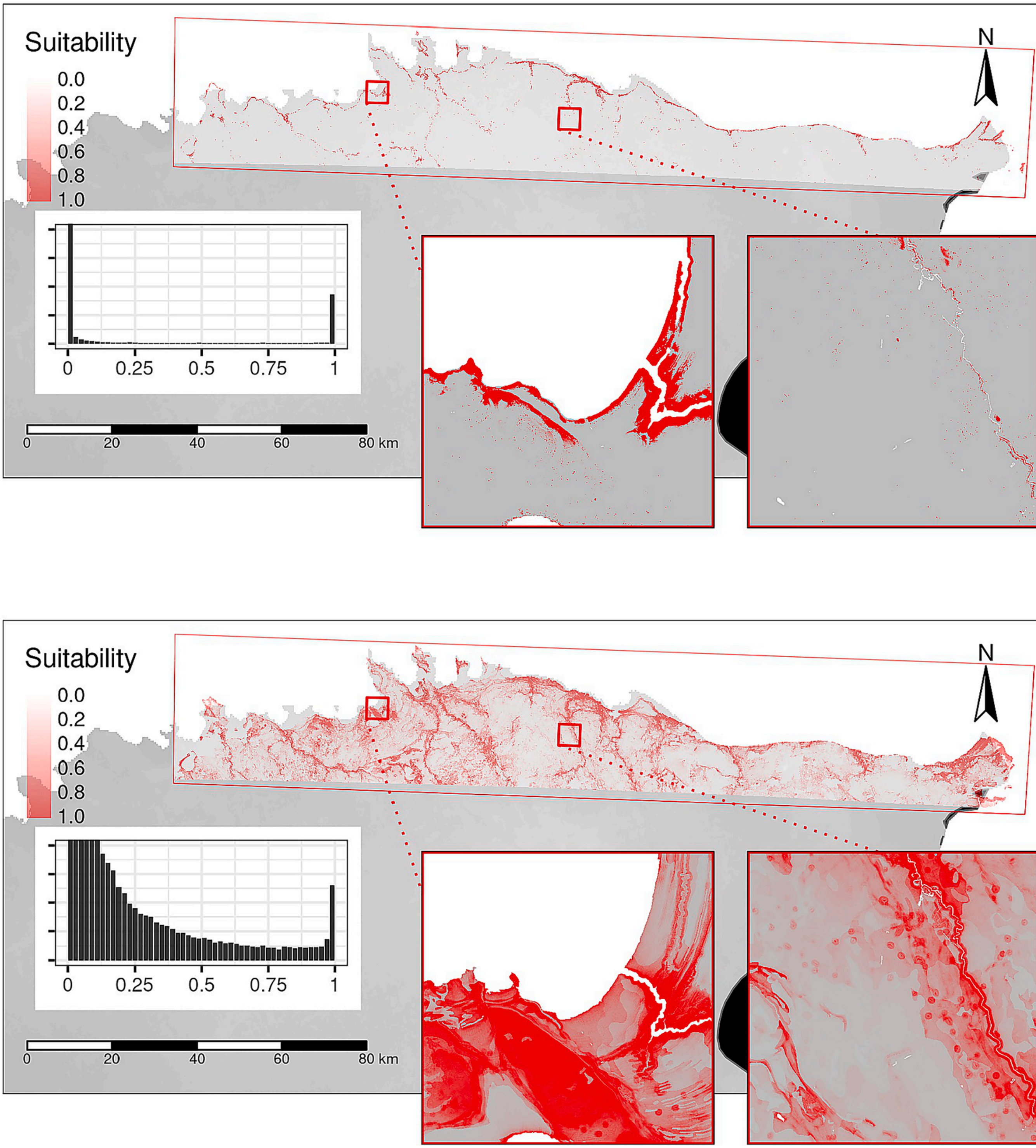

Fig. 7. Point process model output for the research area indicating suitability for NW-CWC (upper) and COCW (lower) habitation. The two zoomed areas show river bay and inland riverbanks; the bar chart shows general raster value distribution.

areas.

5.4. Spatial structure

Spatially explicit environmental suitability rasters indicate a significant difference between general suitable environmental areas for hunter-fisher-gatherer NW-CWC and agrarian COCW (see Fig. 7). While NW-CWC settlement choice is restricted to areas on the seashore and close to riverbanks, COCW settlement choice is free of this restriction, giving a tenfold increase in potential residential areas.

The model indicates that riverbanks and areas near sea and lakes were considered suitable by both settlement groups. For COCW, the suitability of regions close to water bodies was more variable and significant areas inland were also considered as suitable environments. As a result the spatial configuration of the suitable residential area was more homogeneous and the total potential residential area was much larger.

It can also be observed that the probability distributions for site presences are different. The NW-CWC model results in a more clear-cut distinction between higher and lower probability areas. For the NW-CWC model the areas can be separated into low and very high probability habitation, whereas the COCW model has a significant number of in-between areas.

6. Discussion and conclusion

6.1. Changing landscape

Our results confirm knowledge retrieved by fieldwork and univariate statistics (Sikk et al., 2020a) about the suitable environmental conditions for habitation. The results were extended by spatially explicit suitability models which led to an overview of potential habitation zones in general. It indicated a significant expansion of usable areas for COCW compared with NW-CWC. The larger area available for early agrarian societies could have made tolerated immigration (as previously proposed by Kriiska, 2019) possible as the newcomers could inhabit previously actively uninhabited areas. The new type of settlement system resulting from the migration of agrarian societies probably created new patterns of trade and social connections.

Despite new potentially habitable areas there was also a significant overlap of highly suitable areas close to rivers. The overlap of used residential regions is not only theoretical but can also be observed in archaeological material, for example in the Narva region, where both NW-CWC and COCW settlements were close to the Narva river mouth, although not contemporaneously (Kriiska et al., 2016). The overlap implies that there could have been competition for certain areas between contemporaneous hunter-gatherers and early agrarian societies. In most cases COCW settlements followed previous hunter-gatherer habitation (e.g. Kriiska and Nordqvist, 2012, Fig. 10), indicating expansion.

It is also possible that the overlap was only potential and did not result in competition. COCW sites in river flood meadows made riverbanks attractive to farmers, who only used them in certain regions, which is universally reflected in models. There could have also been an exchange resulting in COCW pottery being brought to CWC sites which ultimately led to possible misclassification of some of the hunter-fisher-gatherer sites as COCW sites.

6.2. Model predictability

Similarly to archaeologists' observations, the study revealed that COCW models do not predict settlement site locations as well as NW-CWC models. From the perspective of fieldwork, lower predictability could be an indication of larger suitable areas, as also shown by the model comparison. But the lower model performance suggested that another explanation is needed.

Various systemic effects on the predictive performance of inductive models have long been discussed. For example (Ebert and Kohler, 1988) have argued in the context of forager studies that spatial heterogeneity, temporal expectedness and economic intensification influence settlement location predictability. Below we discuss three possible reasons for differing modelling performance.

6.3. Mixing settlement choice modes in one model

It is known that early agrarian sites used locations with varied environmental conditions, for example on the klint in northern Estonia and on river flood meadows in southern Estonia. The COCW model unified these different modes into one model. Furthermore, COCW habitations encompassed areas usable for agricultural land but also locations suitable for hunter-fisher-gatherers, sometimes even reusing their sites on sandy patches on the lake shoreline and riverbanks. This might be an indication of a more complex environmental effect including different site functions and available environment.

It is possible that the COCW settlement pattern contains two or more settlement choice modes occupying different spaces. For example one could be similar to hunter-fisher-gatherer mode and another to agrarian mode. Two distinct modes addressed through a single PPM could significantly lower model performance. Similar situations could arise if NW-CWC sites are misidentified as COCW sites because of pottery exchange, as proposed above (for effects of misidentification see e.g. Costa et al., 2015).

The related possible effect of temporal predictability of the model is currently not observable in archaeological data. COCW use of both agricultural lands and riverbanks and shorelines might indicate a different seasonal mode for COCW settlement which needs to be studied further. Different mobility modes also influence settlement patterns but a simulation study of foragers (Sikk and Caruso, 2020) has shown that mobility driven by the same resources does not significantly influence settlement location choice.

6.4. Effects of spatial configuration

It is also known that both niche breadth and the spatial structure of the environment influence the predictive power of models. In a dedicated study evaluating the predictive power of species distribution models (Connor et al., 2018) it was shown that model accuracy significantly deteriorated for homogeneous environments with a grain-size increase in observations, which is natural to archaeology. The results indicate that habitat specialists were more accurately modelled than generalist species. We can presume a similar situation for human cultures as a diversified economy provides generalist spatial features.

(Ebert and Kohler, 1988) suggest that spatial heterogeneity or patchiness of the attractions required for settlement determines the effects of site location. The more evenly distributed such attractions are on the landscape, the more determined and thus predictable settlement locations are. This effect can potentially explain the difference between NW-CWC and COCW patterns, as NW-CWC settlements are restricted by water bodies and thus have very specific regions which are suitable for habitation, while COCW settlements are not restricted in a similar way. The EEM then indicates that COCW inhabitants had a very different perception of the landscape which made it possible to extend the potential residential area inland. Agrarian perspective transformed the perception of the landscape, giving it a more homogeneous spatial configuration with large areas of almost equally suitable locations for living. This can at least partly explain the lesser predictive power of the model.

6.5. Socio-economic organisation

In addition to eco-spatial principles, social factors, which can be considered second-order properties of a point process, can introduce spatial autocorrelation to the system and influence the predictability of site locations. The general question regarding this issue is: what proportion of human behaviour is immediate and based on perception (can be explained by proximity arguments) and what proportion is systematically organised? (Ebert and Kohler, 1988). Intensification of hunter-gatherer economy and movement towards agricultural economic mode has been associated with increased population, increased packing, decreased residential mobility, increased storage and production of storable food resources. This kind of intensification has been considered to improve the predictability of settlement site locations (Ebert and Kohler, 1988).

The increased predictability of NW-CWC settlement may suggest that they were undergoing economic intensification but that COCW were not, despite their agrarian subsistence. It might be possible that COCW brought with them agrarian culture developed elsewhere but immigrating groups did not have social properties associated with intensification.

It has also been hypothesised that as social complexity grows, settlement choice is determined less by environmental factors and more by social factors (Altschul, 1988; Kvamme et al., 2005), which are less observable in archaeological records. The reason for this is twofold: firstly horizontal – more socially complex societies have diverse economic activities which can result in a multitude of different residential sites which follow different location choice principles and might be hard to isolate (see Section 6.3). Secondly, social complexity leads to a vertical, hierarchical structure, which also has an impact on the spatial structure of settlement patterns. (Altschul, 1988) argues that increased social complexities also led to development on different, central or "magnetic" sites which stood out from the general picture.

The known pattern of COCW settlement in the region contradicts the social complexity hypothesis, as central sites are known for the period; such sites are only known for the subsequent Bronze Age period. Archaeological material also contradicts social complexity, as monumental burial sites and most outstanding prestige items were introduced later, during the Bronze Age.

6.6. Conclusion

This study mostly confirmed previous knowledge of settlement choice principles. It explicitly indicated the geographical difference between suitable residential spaces available for hunter-gatherer NW-CWC and early agrarian COCW cultures, which might explain the conflictless immigration process. Potential residential regions have a certain overlap that can be further studied to explore the dynamic relations, including possible spatial competition, between hunter-gatherers and COCW during the general conversion to agrarian mode.

The created models provide robustly different performance measures which can be explained by mutually non-exclusive spatial and eco-social effects. These include hypothetical effects influencing systemic environmental determinism, such as spatial configuration of suitable spaces and economic intensification and diversity, which can be addressed by further quantitative modelling and simulation studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research has been supported by the Luxembourg National Research Fund (FNR) (10929115) and Estonian Research Council (PRG243).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jasrep.2021.103330.

References

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., Duffy, P.R., Ebel, A.V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Price, T.D., Saag, L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V.I., Szeverényi, V., Tóth, G., Trifanova, S.V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Population genomics of Bronze Age Eurasia. Nature 522 (7555), 167–172. https://doi.org/10.1038/nature14507.

Altschul, J., 1988. Models and the modelling process. In: Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling. U.S. Department of the Interior, Bureau of Land Management, Denver, pp. 61–89.

Anderson, R.P., Raza, A., 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela: Effect of study region on models of distributions. J. Biogeogr. 37, 1378–1393. https://doi.org/10.1111/j.1365-2699.2010.02290.x.

Arbeiter, R., 1993. Engineering Geology Maps for the Tallinn area in scale 1:5000. Sheets 97–99 and 126–128.

Baddeley, A., Berman, M., Fisher, N.I., Hardegen, A., Milne, R.K., Schuhmacher, D., Shah, R., Turner, R., 2010. Spatial logistic regression and change-of-support in Poisson point processes. Electron. J. Statist. 4, 1151–1201. https://doi.org/10.1214/10-EJS581.

Bailey, T.C., Gatrell, A.C., 1995. Interactive Spatial Data Analysis. Longman Scientific & Technical Essex.

Banks, W.E., 2017. The application of ecological niche modeling methods to archaeological data in order to examine culture-environment relationships and cultural trajectories. Quaternaire. Revue de l'Association française pour l'étude du Quaternaire 271–276. 10.4000/quaternaire.7966.

Banks, W.E., d'Errico, F., Dibble, H.L., Krishtalka, L., West, D., Olszewski, D.I., Peterson, A.T., Anderson, D.G., Gillam, J.C., Montet-White, A., 2006. Eco-cultural niche modeling: new tools for reconstructing the geography and ecology of past human populations. PaleoAnthropology 4, 68–83.

Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J., Wang, F., Liu, J., 2018. Effects of grain size and niche breadth on species distribution modeling. Ecography 41 (8), 1270–1282. https://doi.org/10.1111/ecog.03416.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8 (7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Costa, H., Foody, G., Jiménez, S., Silva, L., 2015. Impacts of species misidentification on species distribution modeling with presence-only data. ISPRS Int. J. Geo-Inf. 4, 2496–2518. https://doi.org/10.3390/ijgi4042496.

Crema, E.R., Bevan, A., Lake, M.W., 2010. A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record. J. Archaeol. Sci. 37 (5), 1118–1130. https://doi.org/10.1016/j.jas.2009.12.012.

Ebert, J.I., Kohler, T.A., 1988. The theoretical basis of archaeological predictive modeling and a consideration of appropriate data-collection methods. In: Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling. U.S. Department of the Interior, Bureau of Land Management, Denver, pp. 97–172.

Ekman, M., 1996. A consistent map of the postglacial uplift of Fennoscandia. Terra Nova 8 (2), 158–165. https://doi.org/10.1111/j.1365-3121.1996.tb00739.x.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists: statistical explanation of MaxEnt. Divers. Distrib. 17, 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x.

Estonian Land Board, 2019. Elevation Data. Estonian Land Board, Tallinn.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R.G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S.L., Risch, R., Rojo Guerra, M.A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K.W., Reich, D., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522 (7555), 207–211. https://doi.org/10.1038/nature14317.

Jaanits, L., 1970. Kultuuri arengupidevusest Eestis üleminekul keskmiselt nooremale kiviajale. In: Schmiedehelm, M., Jaanits, L., Selirand, J. (Eds.), Studia Archaeologica in Memoriam Harri Moora. Valgus, Tallinn, pp. 81–87.

Jaanits, L., 1955. Neoliitilised asulad Eesti NSV territooriumil, in: Muistsed Asulad Ja Linnused. Arheoloogiline kogumik I. Eesti Riiklik Kirjastus, pp. 176–201.

Jaanits, L., Laul, S., Lõugas, V., Tõnisson, E., 1982. Eesti esiajalugu. Eesti Raamat, Tallinn.

Jussila, T., Kriiska, A., 2004. Shore displacement chronology of the Estonian Stone Age. Estonian J. Archaeol. 8 (1), 3. https://doi.org/10.3176/arch.2004.1.01.

Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., Pärtel, M., Ostonen, I., Uuemaa, E., 2021. EstSoil-EH: a high-resolution eco-hydrological modelling parameters dataset for Estonia. Earth Syst. Sci. Data 13, 83–97. https://doi.org/10.5194/essd-13-83-2021.

Kriiska, A., 2019. Lühivaade III aastatuhandesse eKr ehk mida me teame voi vähemalt arvame end teadvat Eesti alal elanud nöörkeraamika kultuuri inimestest. Humanitaarteadused ja kunstid 100aastases rahvusülikoolis. Ed. R. Altnurme. Tartu Ülikooli Kirjastus, Tartu 13–24.

Kriiska, A., 2000. Corded Ware culture sites in north-eastern Estonia, in: De Temporibus Antiquissimis Ad Honorem Lembit Jaanits. Muinasaja Teadus. pp. 59–79.

Kriiska, A., Gerasimov, D.V., Nordqvist, K., Lisitsyn, S.N., Sandell, S., Kholkina, M.A., 2016. Stone Age Research in the Narva–Luga Klint Bay Area in 2005–2014. Iskos 21.

Kriiska, A., Lang, V., Mäesalu, A., Tvauri, A., Valk, H., 2020. Eesti ajalugu I. Eesti esiajalugu [Estonian History I. Estonian Prehistory]. Tartu Ülikooli ajaloo ja arheoloogia instituut, Tartu.

Kriiska, A., Nordqvist, K., 2012. Arheoloogilised väljakaevamised Narva-Jõesuu IIa neoliitilisel asulakohal 2010. aastal, in: Märgilised Mälestised. Uurimusi Narva Piirkonna Ajaloost., Narva Muuseumi Toimetised. Narva Muuseum, Narva, pp. 14–37.

Kriiska, A., Oras, E., Lõugas, L., Meadows, J., Lucquin, A., Craig, O.E., 2017. Late Mesolithic Narva stage in Estonia: pottery, settlement types and chronology. Estonian J. Archaeol. 21 (1), 52. https://doi.org/10.3176/arch.2017.1.03.

Kvamme, K.L., 2005. There and back again: Revisiting archaeological locational modeling, in: GIS and Archaeological Site Location Modeling. CRC Press, pp. 23–55.

Liu, C., White, M., Newell, G., Pearson, R., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. J. Biogeogr. 40 (4), 778–789. https://doi.org/10.1111/jbi.12058.

Lõugas, L., Kriiska, A., Maldre, L., 2007. New dates for the Late Neolithic Corded Ware Culture burials and early husbandry in the East Baltic region. Archaeofauna 16, 21–31.

Mehrer, M.W., Wescott, K.L. (Eds.), 2005. GIS and Archaeological Site Location Modeling, 0 ed. CRC Press, Boca Raton. 10.1201/9780203563359.

Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36 (10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., Furtwängler, A., Andrades Valtueña, A., Feldman, M., Economou, C., Oinonen, M., Vasks, A., Balanovska, E., Reich, D., Jankauskas, R., Haak, W., Schiffels, S., Krause, J., 2018. The genetic prehistory of the Baltic Sea region. Nat. Commun. 9, 442. https://doi.org/10.1038/s41467-018-02825-9.

Moora, H., Laid, E., Mägiste, J., Kruus, H., 1935. Eesti Ajalugu I. Esiajalugu ja muistne vabadusvõitlus. Eesti kirjanduse Selts, Tartu.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57 (2), 443–452. https://doi.org/10.2136/sssaj1993.03615995005700020026x.

Morales, N.S., Fernández, I.C., Baca-González, V., 2017. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. PeerJ 5, e3093. 10.7717/peerj.3093.

Muru, M., Rosentau, A., Kriiska, A., Lõugas, L., Kadakas, U., Vassiljev, Jüri, Saarse, L., Aunap, R., Küttim, L., Puusepp, L., Kihno, K., 2017. Sea level changes and Neolithic hunter-fisher-gatherers in the centre of Tallinn, southern coast of the Gulf of Finland, Baltic Sea. The Holocene 27 (7), 917–928. https://doi.org/10.1177/0959683616678462.

Oras, E., Lucquin, A., Lõugas, L., Tõrv, M., Kriiska, A., Craig, O.E., 2017. The adoption of pottery by north-east European hunter-gatherers: Evidence from lipid residue analysis. J. Archaeol. Sci. 78, 112–119. https://doi.org/10.1016/j.jas.2016.11.010.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B. (Eds.), 2011. Ecological Niches and Geographic Distributions (MPB-49)Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press.

Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E., 2017a. Opening the black box: an open-source release of Maxent. Ecography 40 (7), 887–893. https://doi.org/10.1111/ecog.2017.v40.i710.1111/ecog.03049.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3-4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31 (2), 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x.

Phillips, S.J., Dudík, M., Schapire, R.E., 2017b. Maxent software for modeling species niches and distributions (Version 3.4. 1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/.

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I., 2015. Point process models for presence-only analysis. Methods Ecol. Evol. 6, 366–379. https://doi.org/10.1111/2041-210X.12352.

Rosentau, A., Muru, M., Kriiska, A., Subetto, D.A., Vassiljev, J., Hang, T., Gerasimov, D., Nordqvist, K., Ludikova, A., Lõugas, L., Raig, H., Kihno, K., Aunap, R., Letyka, N., 2013. Stone Age settlement and Holocene shore displacement in the Narva-Luga Klint Bay area, eastern Gulf of Finland: Stone Age settlement and Holocene shore displacement, Gulf of Finland. Boreas n/a-n/a. 10.1111/bor.12004.

Rosentau, A., Veski, S., Kriiska, A., Aunap, R., Vassiljev, J., Saarse, L., Hang, T., Heinsalu, A., Oja, T., 2011. Palaeogeographic model for the SW Estonian coastal zone of the Baltic Sea. In: The Baltic Sea Basin. Springer-Verlag, Berlin, pp. 165–188.

Saag, L., Varul, L., Scheib, C.L., Stenderup, J., Allentoft, M.E., Saag, L., Pagani, L., Reidla, M., Tambets, K., Metspalu, E., Kriiska, A., Willerslev, E., Kivisild, T., Metspalu, M., 2017. Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. Curr. Biol. 27, 2185–2193.e6. https://doi.org/10.1016/j.cub.2017.06.022.

Saag, L., Vasilyev, S.V., Varul, L., Kosorukova, N.V., Gerasimov, D.V., Oshibkina, S.V., Griffith, S.J., Solnik, A., Saag, L., D'Atanasio, E., Metspalu, E., Reidla, M., Rootsi, S., Kivisild, T., Scheib, C.L., Tambets, K., Kriiska, A., Metspalu, M., 2021. Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. Science. Advances 7, eabd6535. https://doi.org/10.1126/sciadv.abd6535.

Saarse, L., Vassiljev, J., Miidel, A., 2003. Simulation of the Baltic Sea Shorelines in Estonia and Neighbouring Areas. J. Coastal Res. 19, 261–268.

Sikk, K., Kriiska, A., Johanson, K., Sander, K., Vindi, A., 2020a. Environment and settlement location choice in Stone Age Estonia. EJA.

Sikk, K., Caruso, G., 2020. A spatially explicit agent-based model of central place foraging theory and its explanatory power for hunter-gatherers settlement patterns formation processes. Adaptive Behavior (Evolution of Cultural Complexity) 377–397. https://doi.org/10.1177/1059712320922915.

Sikk, K., Kriiska, A., Johanson, K., Sander, K., Vindi, A., 2020b. Estonian Stone Age settlement sites: a dataset. https://doi.org/10.5281/zenodo.3775415.

Verhagen, P., Whitley, T.G., 2012. Integrating Archaeological Theory and Predictive Modeling: a Live Report from the Scene. J Archaeol Method Theory 19, 49–100. https://doi.org/10.1007/s10816-011-9102-7.

Vernon, K.B., Yaworsky, P.M., Spangler, J., Brewer, S., Codding, B.F., 2020. Decomposing Habitat Suitability Across the Forager to Farmer Transition. Environ. Archaeol. 1–14 https://doi.org/10.1080/14614103.2020.1746880.

Vidal-Cordasco, M., Nuevo-López, A., 2021. Difference in ecological niche breadth between Mesolithic and Early Neolithic groups in Iberia. J. Archaeolog. Sci.: Rep. 35, 102728 https://doi.org/10.1016/j.jasrep.2020.102728.

Vignali, S., Barras, A.G., Arlettaz, R., Braunisch, V., 2020. SDMtune : An R package to tune and evaluate species distribution models. Ecol. Evol. 10, 11488–11506. https://doi.org/10.1002/ece3.6786.

Wachtel, I., Zidon, R., Shelach-Lavi, G., 2020. Using the Maximal Entropy Modeling Approach to Analyze the Evolution of Sedentary Agricultural Societies in Northeast China. Entropy 22, 307. https://doi.org/10.3390/e22030307.

Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution: International Journal of Organic. Evolution 62, 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x.

Whitford, B.R., 2019. Characterizing the cultural evolutionary process from eco-cultural niche models: niche construction during the Neolithic of the Struma River Valley (c. 6200–4900 BC). Archaeol Anthropol Sci 11, 2181–2200. https://doi.org/10.1007/s12520-018-0667-x.

Yaworsky, P.M., Vernon, K.B., Spangler, J.D., Brewer, S.C., Codding, B.F., 2020. Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument. PLoS ONE 15, e0239424. https://doi.org/10.1371/journal.pone.0239424.