Algorithmic classification and statistical modelling of coastal settlement patterns in Mesolithic south-eastern Norway

Abstract

This paper presents and contrasts procedures and conceptual underpinnings associated with statistical modelling and algorithmic classification in the study of past locational patterns. This was done by applying the methods of logistic regression and random forest to a case study of coastal Mesolithic settlement patterns in southern Norway—a context that has not been subject to formal locational pattern analysis in the past. While the predictive accuracy of the the two methods was comparable, the different strengths and weaknesses associated with the methods offered a firmer foundation on which to both draw and moderate substantive conclusions. The main findings were that among the considered variables, the exposure of sites was the most important driver of Mesolithic site location in the region, and while some diachronic variation is evident, the differences detected appear to both be of a different and far more modest nature compared to that which has previously been proposed. All employed data, the Python script used to run the analyses in GRASS GIS, as well as the R script used for subsequent statistical analysis is freely available in an online repository, allowing for a complete scrutiny of the steps taken.

Keywords: Site locational patterns, Statistical modelling, Machine learning, Mesolithic, South-eastern Norway

1   Introduction

Logistic regression has been the most commonly applied statistical method in the study of settlement locational patterns in archaeology (e.g. Kvamme 1988; Warren and Asch 2000; Woodman and Woodward 2002; Verhagen and

1

Whitley 2020). However, in recent years it has been suggested that machine learning methods such as random forest might represent a valuable alternative (e.g. Verhagen and Whitley 2012; Harris, Kingsley and Sewell 2015). This paper contributes to this discourse by providing an applied case study that employs both methods. The paper thus contrasts the disparate modelling procedures associated with statistical modelling and algorithmic classification, while also using techniques within these frameworks that have seen little previous application in archaeology. While the paper compares these approaches, the ultimate aim of the study was to elucidate coastal settlement patterns in Mesolithic Norway. As such, the aim is not a balanced comparison of the potential performance of the methods by applying them to idealised data sets, but rather to demonstrate and compare their use as applied to a real world archaeological material with its own range of idiosyncrasies.

The Mesolithic in Norway has typically been understood in terms of a continuous increase in societal variability, coupled with a decrease in mobility (e.g. Bergsvik 2001; Bjerck 2008; Glørstad 2010:97). This is to be clearly reflected in the location of coastal sites relative to a range of environmental variables. However, while frequent reference to overarching settlement patterns and general developments in locational preferences throughout the period can be found in the literature, there is an overall lack of formal treatment of these patterns (Åstveit 2014). The substantive goals of the study were consequently to i) formalise and quantitatively evaluate previously suggested environmental determinants for coastal site location in Mesolithic Norway, and ii) attempt to determine if the relative importance of these variables change across major chronological transitions. This was done by means of a case study, represented by excavated and surveyed coastal Mesolithic sites from the municipalities of Larvik, Porsgrunn and Bamble, situated in south-eastern Norway.

2   Archaeological context and case study

Human presence in Norway has not been recorded earlier than around 9400–9300 BCE, signifying the beginning of the Early Mesolithic (EM). The classic perception of the EM is of a phase characterised by coastal exploitation by highly mobile groups (e.g. Bjerck 2008; Fuglestvedt 2012). The subsequent Middle Mesolithic (MM), set to last from around 8200–6300 BCE, has been somewhat contested in the literature, where discussions have concerned whether the phase has more in common with the highly mobile EM (Mansrud 2014), or if it signifies the emergence of what are deemed classic Late Mesolithic (LM) traits of semi-sedentary and more varied locational pat-

terns in south-eastern Norway (Solheim and Persson 2016). The LM, lasting from around 6300–3900 BCE, is to mark a more definite overall decrease in mobility (Bjerck 2008; Glørstad 2010). While the LM is only defined to last until 3900 BCE, settlement patterns in the first part of the Early Neolithic is largely seen as a continuation of the LM in the region (Glørstad 2009). Sites dating up to around 3600 BCE were therefore included in this analysis as part of the LM.

An important characteristic of the region as whole is that due to isostatic rebound following the retreat of Fennoscandian Ice Sheet, the land has continuously risen faster than the sea in south-eastern Norway. In addition, Mesolithic sites located beneath the marine limit are almost exclusively believed to have been utilised when they were situated on or a few meters from the shoreline. This means that the sites are distributed chronologically in the landscape, where older sites are situated at higher elevations than younger sites (Breivik, Fossum and Solheim 2018; Solheim 2020). Due to the rate of rebound they were also shore bound within relatively limited time-spans, meaning the sites are usually constituted of chronologically distinct artefact inventories. Here it is assumed that all sites under study were located on the shoreline. While not unproblematic, I believe that aggregating the results across as many cases as is done here will suppress noise caused by any sites potentially misclassified as shore-bound.

The area of concern is comprised of the municipalities Porsgrunn, Bamble and Larvik in Vestfold and Telemark county (Figure 1). The region has been subject to extensive survey and excavation activity in the recent decade or so, leading to a dramatic increase in research activity and available data (e.g. Solheim and Damlien 2013; Jaksland and Persson 2014; Melvold and Persson 2014; Reitan and Persson 2014; Solheim 2017).

3 Site data

The most readily available, comprehensive and standardised source of information on archaeological investigations in Norway is held in the national heritage and monuments database *Askeladden*, which is managed by the Directorate for Cultural Heritage. However, as the database is mainly aimed at cultural resource management, the information it holds is predominantly of a bureaucratic nature. The only consistently available information of interest here is the extent of the sites, given by spatial geometries. All site and stray find records from Bamble, Porsgrunn and Larvik were retrieved from the database and the records manually gone over to exclude any undefinable and non-Stone Age records. Parallel to this, each record was given a quality score

Quality	Definition	Count
1	Site delineated by use of a GNSS-device, or a securely georeferenced record. Extensive database entry.	160
2	Secure spatial data. Slight damage to site or somewhat lacking database record.	247
3	Secure spatial data. Damaged site, such as outskirts of quarry, and/or very limited database entry.	71
4	Surveyed by archaeologists. However, the database entry is extremely limited/unclear, the site geometry is only given as a point or small automatically generated circle, and/or finds are from the topsoil of a field, making the actual site location and extent uncertain.	107
5	Likely site but uncertain spatial information. Typical example is recurring stray finds in a field or other larger area.	92
6	Single stray find or unverified claims/suggestions of possible site.	105

Table 1: Quality scoring system for Stone Age site records in *Askeladden* based on available information and the certainty with which it was originally included in the database. Records of quality 4 or below were excluded from further analysis.

based on the criteria given in Table 1. The process resulted in 477 records being deemed certain enough to allow for analysis.

The sites were then ascribed to each of the three chronological phases based on the lowest elevation value of the polygon representing the site, using the mean of the corresponding regional shoreline displacement curve given in Figure 2. All sites situated at elevations corresponding to a shoreline date lower than 3600 BCE were then removed, resulting in a final 462 records being retained for analysis.

4   Sampling scheme

To understand where sites are located in the landscape ultimately also depends on our knowledge of where they are verifiably *not* located, as this allows us to get at the degree of selectivity exhibited by prehistoric inhabitants (Kvamme and Jochim 1989:2). As Stone Age sites are typically discovered by means of test-pitting in southern Norway, surveys do not provide a continuous field of known presence or absence of archaeological material, and no

verified negative data is readily available from sources such as *Askeladden*. Random samples of control points representing assumed non-sites were therefore generated across the landscape (e.g. Kvamme 1988; Fisher et al. 1997). These were first constrained to not fall in areas holding higher slope values than any of the sites. This was done both because surveys are commonly directed by degree of slope (e.g. Nielsen et al. 2016:369), and because this excludes extremely steep areas that it is reasonable to assume have not been in competition to be chosen for site locations. Similarly, the samples were also constrained to fall within the elevation range of the sites, as areas outside of this would not have represented alternative coastal locations in the Mesolithic. Finally, samples were restricted to not fall in present day lakes, both because several of these are likely to have been present or part of the sea in the Mesolithic, and because they are typically only partially surveyed when water levels are low.

As the most substantial survey projects have followed from the expansion of infrastructure, this has led to severe spatial structuring in the retrieval of archaeological material. Two sampling frames were employed in an attempt to work around this issue. The first involved constraining the random sample by a 500 m buffer around the sites, in addition to the environmental constraints. This is somewhat arbitrary, but follows from the survey strategy that has been employed for highway projects, where a 400 m tract along the proposed highway is surveyed, with additional extensions for areas such as planned intersections and lay-bys (Eskeland 2017:7). The second sampling frame involved creating a convex hull around these buffers, in addition to the environmental constraints. This larger hull constraint is likely to include more variability in the background populations for the variables considered, but it is less certain that any observed pattern is not simply an effect of including areas in the analysis where no archaeological investigation has taken place (cf. Kvamme 2020:217). While the smaller buffer sampling frame is not a safeguard against this effect, and the considerable reduction in sample space is likely to exclude variability that was relevant for locational choices, any locational pattern that can be observed using both sampling schemes would result in a far more convincing result in terms of whether actual locational preferences have been discerned. The sample sizes were set to the arbitrarily large 1000 for each of the two sampling strategies for each of the three phases, totalling at 6000 random samples.

5 Locational variables

Below follows a presentation of the considered environmental covariates. These are mainly derived from previous informal studies of Mesolithic settlement patterns in Norway. Of these, exposure are among the most central as this has been emphasised in studies undertaken in most areas of the country (Bjerck 2008; Berg-Hansen 2009; Åstveit 2014). Exposure is also the locational variable that has been most explicitly linked to mobility, where a common sentiment is that higher mobility is reflected in higher degrees of site exposure. As societies in earlier periods are typically understood to have been more mobile, the general trend is to be an overall decrease in site exposure through the Mesolithic (e.g. Lindblom 1984; Jaksland 2001; Bjerck 2008; Breivik, Fossum and Solheim 2018). Svendsen (2014) has, however, pointed out that the exposure term as used in Norwegian Mesolithic research has referred to both lack of shelter from wind and sea, but also degree of visibility from sites. In an attempt to get at this distinction, exposure is treated by two variables.

5.1 Wind fetch

Wind fetch is a fundamental concept in the study of ocean systems and environments that is commonly employed outside archaeology to estimate the exposure of locations to wind-generated wave action (Laing 1998, for archaeological applications see Nitter, Elvestad and Selsing 2013; Nitter and Coolen 2018). Wind-generated wave action is a complex phenomenon that can depend on a wide range of parameters. However, the most commonly applied models for wave exposure are founded in so-called fetch-based indices. Fetch is defined as the unobstructed distance of sea across which the wind travels towards a given location. The longer the uninterrupted distance the wind can travel the surface of the ocean, the higher waves will tend to be. Wave prediction models based on fetch involve various degrees of complexity, starting from relatively simplistic models that only take account of the length of the fetch, as is done here. While more complex models allow for more accurate predictions that also move beyond estimation of mere relative exposure (Malhotra and Fonseca 2007; Bekkby et al. 2008; Sundblad et al. 2014), investigations into the predictive power of fetch-based models have found that those based solely on fetch length still perform reasonably well in predicting the presence and absence of biological indicators of wave exposure (Burrows, Harvey and Robb 2008; Hill et al. 2010).

Here, estimation of average fetch length was based on the procedures presented by Ekebom, Laihonen and Suominen (2003) and Tolvanen and

Suominen (2005). With the sea-level raised just above each random point and the centre point for the site polygons, this involved first drawing a line every 7.5 degrees in circumference around each point. The lines were then first clipped by the raised coastline, and then by the present day coastline in the larger region (Figure 3). The average length of the lines were then returned. The coastline has also changed considerably outside of the study region throughout the Mesolithic, which means that the the average fetch of the more exposed locations has likely been underestimated. However, as the average fetch lengths will largely be determined by more immediate surroundings, this seemed like less of a problem for the purposes of mapping general locational patterns.

5.2   Visibility

Viewshed analysis was employed to investigate visibility patterns (Conolly and Lake 2006:225–233; Gillings and Wheatley 2020). As what is of interest here are general visibility patterns, unsophisticated measures of total visible area from the analysed locations was deemed sufficient. The outer radii of the viewsheds was set to 5000 and 10,000 meters to represent views at a long and medium distance. These distances are somewhat arbitrary, but relatively common when mapping the size of views where a target has not been defined (e.g. Lake and Woodman 2000; Lopez-Romero 2008; Garcia 2013). The elevation of the observer point was set to 1.65 m above ground, and r.viewshed, the employed GRASS module, was set to take the curvature of the earth into account.

   The sea level was raised to the lowest elevation of each site polygon, and the viewsheds computed from the centre points. For raising the sea level to the random point locations, a circular buffer with a radius of 11.9 meters was generated around each point. This equates to the median site size. Results were returned as the number of cells visible from each point.

5.3   Shoreline displacement

Wren, Costopoulos and Hawley (2018) recently investigated the relationship between Wemindji settlement location and rapid isostatic rebound in James Bay, Canada. They found that the sites tend to be situated at locations where the topography will have resulted in a higher than expected degree of coastal stability within 2 km radius of the sites, and a higher than expected variability in shoreline emergence within 20 km of the sites. This is taken to reflect a settlement strategy favouring site locations with stable local surroundings that at the same time offer ecological variability within

wider resource catchments. Variation in when areas emerged from the sea means that these will be at various stages of ecological succession, following the transition from oceanic, to littoral, to finally a terrestrial environment.

A similar approach is taken here, with a few adjustments to the methodology. First of all the shoreline displacement curves for the study region are only reasonably reliable up to around 100 masl. The 32 EM sites situated above 100 masl were consequently excluded from this part of the analysis. Furthermore, the shoreline displacement in south-eastern Norway cannot be approximated by a low-order polynomial. The maximum radius explored around sites and non-sites was therefore set to 1 km, simply due to uncertainty towards the outer edges of the distribution of sites.

The emergence raster in Figure 4 was created by reclassifying the elevation raster based on the regional displacement curves. As the assumption here is that all sites have been situated on the shoreline, all areas lower than the lowest elevation of the site polygons and non-site buffers were excluded from analysis for each location. Standard deviation in year of emergence was then retrieved using buffers with radii of 50, 500 and 1000 meters around the point locations.

5.4   Island location

Several authors have pointed to the fact that a substantial amount of sites seem to have been situated on small islands, and especially so in the earlier parts of the Mesolithic (e.g. Nyland 2012; Bjerck and Zangrando 2013; Bjerck 2017). Behavioural explanations of this assertion has followed a general emphasis on boating and resource exploitation in archipelago environments, but has also been explicitly related to extensive seal hunting in the earliest part of the Mesolithic (Bjerck 2017), as well as concentrations of marine productivity in the lee of smaller islands (Breivik 2014; Schmitt 2015). For this analysis a context dependent algorithm available in the provided Python script was devised to identify if point locations would have been situated on islands, assuming they were shore-bound. If yes then the size of the island was retrieved.

5.5   Sediments and infiltration ability

Sites have also been suggested to be located relative to different soils and sediments (Bergsvik 1995; Berg-Hansen 2009). The relevance of this is typically suggested to follow from a desire to situate sites on well-drained locations. The sediment data used here was retrieved from the Geological Survey of Norway and has a resolution of 1:20,0000. The sediment data comes provided

with an infiltration score, relating to how well the sediments drain and clean wastewater. While these are in part related to the potential for biological breakdown of toxins, they are also determined by degree of permeability (Geological Survey of Norway 2015). Thus, while the original motivation for the classification does not perfectly align with my intended use, it does appear to capture the relevant properties of the sediments. The infiltration level for each site was found by taking the mode on each site polygon and each of the non-site buffers.

5.6 Aspect

Aspect has also been suggested to be important to locational choices in Mesolithic Norway. This is typically seen in relation to shelter (Berg-Hansen 2009:47), a desire to face either land or the open sea (Darmark, Viken and Johannessen 2018), or to situate sites at locations that receive more sunlight throughout the day (Mikkelsen 1989:58; Berg-Hansen 2009:113). Here it was decided to focus on the possible relevance of aspect as related to sunlight, as the other proposed reasons for why aspect should be relevant should mostly be captured by other variables. The aspect of each location, returned as degrees from north, was therefore transformed to deviance from south by finding the absolute distance of each value from 180 degrees (Spencer and Bevan 2018).

6 Quantitative framework

6.1 Model building strategies

The first statistical method applied is logistic regression, which forms the backbone of site locational analysis in archaeology. The most common approach to model building with regression methods in archaeology could be said to fall in the category of stepwise minimisation, sometimes also called data-driven model tuning, as this often starts with analysis of the relationship between predictors and the response, independently of the multivariate model (Conolly and Lake 2006:182–183). This becomes very problematic when variables are included or excluded from further analysis based on the statistical significance of this relationship, as has sometimes been the case (e.g. Stančič and Veljanovski 2000:152–153; Duncan and Beckman 2000:36). This is in effect a form of p-value driven, forward stepwise variable selection, and is hampered by severe issues (Harrell 2015:71–72). First of all this approach would exclude a variable without considering whether it might

be an important predictor once other variables are controlled for, or if it might alter the effect of other predictors in the model. If such effects are present, the variable should be included in the model irrespective of whether or not it is a good predictor of the independent variable (Kohler and Parker 1986:416; James et al. 2013:89. Secondly, this method has a tendency to inflate model performance. This follows from the fact that the final model specification is typically evaluated using standard methods that do not take the model tuning process into account. Not accounting for the total number of variables considered, including those that were discarded, breaks the assumptions underlying most standard statistical hypothesis testing and estimation methods, leading to a deflation of p-values and confidence intervals, and an inflation of R^2 values and coefficients (Harrell 2015).

However, the term stepwise selection is typically taken to refer to a form of automated model selection that has also seen application in archaeology (e.g. Bevan et al. 2013; Visentin and Carrer 2017; Spencer and Bevan 2018; Wachtel et al. 2018). Here, independent variables are excluded or included in the model at sequential steps that are stopped once a statistic that determines the relative success of the models does not improve. The Akaike and Bayesian information criteria (AIC and BIC, respectively) are the most frequently applied statistics for stepwise model selection in archaeology. Stepwise selection based on AIC and BIC are seen as an improvement over simple minimisation by use of p-values as a stopping criteria, as they take the total number of variables considered into account in the evaluation of the final model. The fundamental strategy of automated stepwise model selection has, however, received considerable criticism (e.g. Henderson and Denison 1989; Derksen and Keselman 1992; Chatfield 1995; Malek, Berger and Coburn 2007; Burnham, Anderson and Huyvaert 2011; Harrell 2015), and the implications of multiple testing and model tuning are more profound and difficult to take account of in a satisfactory way, leading Harrell (2015:69) to conclude that 'AIC is just a restatement of the p-value, and as such, doesn't solve the severe problems with stepwise variable selection other than forcing us to use slightly more sensible α values.'

The most basic model building strategy associated with regression techniques is the forced entry, direct or simultaneous technique, where all independent variables are included from the start, with no consideration of the relative importance, or the order by which the variables should be included (e.g. Stoltzfus 2011). This is typically best suited when there is little reason to hold any a priori assumptions about the nature of these aspects, and is the approach taken here. The main drawback of this approach is the danger of overfitting and including irrelevant variables that only contribute an increase in standard errors.

Given the dramatic increases in computational power over the last decades, iteratively fitting hundreds or even thousands of models has now become tractable. Faced with this, automated model specification through stepwise selection could appear to be one way to handle this flood of information. Furthermore, regression models can offer insight into the complex net of relative and absolute variable importance, the relationship between these, and an estimate of the confidence with which these assertions can be made. But, as shown above, several authors have argued that this potential is lost with the implementation of stepwise model selection. To fully utilise their capacity arguably requires decisive and rigorous implementation of a well-thought-out, thorough and at times time-consuming modelling strategy (e.g. Harrell 2015:94–99). It could perhaps be argued that in the face of a fragmented and sparse material such as that frequently encountered in archaeology, a looser and more exploratory modelling procedure is warranted. Woodman and Woodward (2002) have argued the precise opposite. They contend that this uncertainty is precisely where careful modelling is most necessary and beneficial, as this can help both elucidate the nature of the fragmentation and guide an analysis that could otherwise be led astray.

Relevant to this is the delineation that Breiman (2001b) makes between what he terms a data modelling culture and an algorithmic modelling culture in statistics. The data modelling culture is based on an assumption that the data under study represents independent draws from a data-generating process where the value on a response variable follows a stochastic model of the form f(parameters, predictor variables, random noise). The algorithmic modelling culture, on the other hand, only makes the assumption that the data represents independent results of a complex and mostly unknowable underlying multivariate process. The goal is to identify an algorithm that best predicts the outcome of this underlying process. Good prediction indicates a better emulation of the underlying process than a solution that predicts poorly. A normal criticism of this focus is that prediction is not explanation, and that despite predictive success, the algorithms in use are often too impenetrable to contribute much in the way of explanation. Breiman (2001b:210) states that this dichotomy of interpretability and explanation as opposed to prediction is contrived: 'The goal is not interpretability, but accurate information'. While regression models can offer a highly interpretable output in the form of coefficients and confidence intervals, they are argued to be easily distorted by resting on assumptions concerning the nature of the underlying distributions that are virtually never met in the analysis of real world phenomena. Additionally, they appear to often be outperformed when it comes to prediction. While the output of machine learning techniques is less interpretable, it is certainly less opaque than the original process under

11

study. Hence, it is argued, they can provide far more reliable and accurate insight, even if this insight holds less detail.

6.2   Random forests

Relatively high predictive power, ease of implementation and its non-parametric nature are among the elements that have led to the popularity of random forests (Hastie, Tibshirani and Friedman 2017:587–604). Random forest is an ensemble method that leverages the increased predictive power following from aggregating the results of a large number of de-correlated, high-variance, low-bias decision trees (Breiman 2001a). Decision trees for classification are based on identifying what variables best partition the data into classificatory groups at sequential binary splits (Baxter 2003:116–118). At each split the variable to use is decided based on how well the candidate variables partitions the data into the correct groups. The degree of discriminatory success is also termed purity, and is often given by the Gini impurity score, which represents how often a random observation would be labelled incorrectly on the given split. While classification trees are highly interpretable and often perform well in classifying the original training data, they have a tendency towards high variance (Hastie, Tibshirani and Friedman 2017:312).

To avoid overfitting while reducing the amount of variance, random forest works by combining bootstrap aggregation of the training data with a random selection of predictor variables to be used in the construction of each individual tree (Box 1). Thus, for each tree the training data is first randomly redrawn with replacement, which results in around 2/3 of the original data being evaluated in the training of the entire forest—a process also known as bagging. The remaining 1/3 is known as the out-of-bag (OOB) sample, and can be used for subsequent model validation and variable importance estimates. Following each resampling, a classification tree is grown. For each node in the tree, a random subset of predictors is drawn, and the predictor returning the lowest Gini impurity score is selected at that node split. The number of randomly selected predictors evaluated at each node is recommended to start at $\sqrt{p}$ for classification trees, where p is the number of independent variables. The number of variables to be evaluated at each split, the parameter *mtry*, is then explored during model tuning. As there is no danger of overfitting associated with generating too many trees, the arbitrary large number of 5000 trees were generated for each forest here. While this came at a computational cost, it meant only the *mtry* parameter required tuning.

Box 1: Random Forest for classification (modified from Hastie, Tibshirani and Friedman 2017:588).

1. For B bootstrap samples $b = 1, ..., B$:
   a) Draw sample Z* of size N from training data X with replacement.
   b) Grow tree T_b by iterating over nodes:
      i. Select m random variables from p predictors.
      ii. Select variable giving the best split.
      iii. if split reduces impurity:
         Split the node into two new nodes.
      iiii. $else$:
         Define as terminal node.
2. Output classification tree ensemble $\{T_b\}_1^B$.

Random forest prediction at point x is then given as $\hat{C}_{rf}^B(x) = majority\ vote$ $\{\hat{C}_b(x)\}_1^B$, where $\hat{C}_b(x)$ is the prediction of the bth classification tree.

An important output of random forests is the variable-importance estimates that rank the variables based on their ability to correctly classify the input data. Following Parr et al. (2018), variable importance was here found by means of the permutation importance method. This is based on first running the OOB-sample through the final forest to establish the baseline accuracy. Then the column of values for a independent variable is randomly shuffled, and the OOB-sample is run through again. The difference in accuracy between the two runs then gives the variable importance measure. It is crucial to note that these can only be used to evaluate the relative importance of variables, and do not provide estimates of effect size.

6.3 Model validation

Internal model validation was done by means of bootstrap resampling for the logistic regression models (Kvamme 1988; Verhagen 2009). The purpose of applying this method is to replace distributional assumptions and asymptotic results with computationally derived confidence intervals (Fox 2008:605), and to avoid both fitting and evaluating model performance using the exact same data (Hastie, Tibshirani and Friedman 2017:249–254). This entailed iteratively resampling the original data and fitting each model 9999 times. For each iteration the coefficients were retrieved, in addition to the Brier score and Area under the receiver operating curve (AUC). The Brier score is what's knowns as a proper accuracy score as it summarises calibration, the difference between the predicted probability and the observed

probability of an outcome, as well as model resolution, the spread of the predictive distribution (Rufibach 2010). The measure ranges from 0 to 1, where 0 represents perfect accuracy. The score is not especially meaningful on its own, however, and is best suited for comparative purposes. The AUC score has a more intuitive interpretation than the Brier score, but only considers the discriminatory abilities of the model (Hosmer, Lemeshow and Sturdivant 2013:173–182). The AUC ranges from 0 to 1, where 1 represents perfect classification, and 0.5 represents no improvement over random classification. The accuracy scores are reported as the mean result across bootstrap samples. The coefficients were evaluated by means of 95% bootstrap percentile intervals, constituted by the range between the $2.5th$ and $97.5th$ percentiles of coefficient values generated across the bootstrap samples (Fox 2008:595).

Variable importance estimates and model performance measures for the random forests were found using nested k-fold cross-validation instead of the bootstrap (Box 2, and below for imputation). The reason for this is that while the logistic regression models were fit using the simultaneous technique, a model tuning step was included for the random forests to identify the optimal number of random variables to be evaluated at each split in the classification trees (the parameter $mtry$). Single bootstrap resampling results in some overlap between the original data and the resampled observations, meaning tuning and validating with this technique would likely lead to overfitting and overestimation of model performance (Hastie, Tibshirani and Friedman 2017:250). Basic cross-validation involves randomly splitting the data into k folds, training the model using all but one fold, and then validating the result against the retained fold (Verhagen 2009:92; Hastie, Tibshirani and Friedman 2017:241–249). This is then repeated until each fold has been used as the validation set. The nested version of cross-validation entails that for each of these validation steps the folds not retained are split into an additional k folds. Here a complete sweep of possible $mtry$ settings was done for each inner cycle, and the $mtry$ achieving the best performance by means of AUC was returned and used for the current outer fold. The final AUC and Brier scores, as well as variable importance measures are reported as averaged across the outer validation cycle.

```
┌─────────────────────────────────────────────────────────────────┐
│              Box 2: Nested 5-fold cross-validation of random forests.              │
│                                                                                        │
│ For each of 5 imputed datasets:                                                        │
│      1. Randomly split into 5 outer folds.                                             │
│      2. For each outer fold:                                                           │
│            a. Split the other 4 outer folds into 5 inner folds.                        │
│            b. For each inner fold:                                                     │
│                      i. For $P$ predictors $mtry = 1, ..., P$:                         │
│                            Create random forest using $mtry$.                          │
│                      ii. Return $mtry$ achieving highest AUC score.                    │
│            c. Return $mtry$ achieving the highest AUC across inner folds.              │
│            d. Run random forest using optimal $mtry$ on outer fold.                    │
│            e. Report Brier score, AUC score and variable importance.                   │
│      3. Return mean Brier, AUC and variable importance from outer folds.               │
│ Return mean Brier, AUC and variable importance from imputed datasets.                  │
└─────────────────────────────────────────────────────────────────┘
```

6.4   Model parameters and preprocessing

Below follows a presentation of predefined model parameters and the pre-
processing steps that were taken. With the exception of for imputation, the
response variable was kept out of this process to maintain inferential validity.

   With the exception of aspect deviating from south, continuous variables
were transformed for the logistic regression models by taking the natural
logarithm due to extreme distributions. Continuous variables were then nor-
malised to take on values between 0 and 1 to better facilitate a comparison
of the impact of each variable, although at the cost of distorting the inter-
pretability of effect sizes (cf. Spencer and Bevan 2018). In the random forests,
the variables were included using raw values. For a second set of models dir-
ectly comparing the the location of sites between the different phases, as
opposed to sites and non-sites, all variables with the exception of the binary
island/mainland variable were transformed by finding the percentile rank of
each site value as compared to the values in the corresponding hull sample.
In an attempt to achieve more sensible diachronic comparisons, this adjusted
the values by variation in the surrounding landscape as captured by the hull
samples for each phase.

   There are a host of different methods for identifying and dealing with col-
linearity (e.g. Dormann et al. 2012; Tomaschek, Hendrix and Baayen 2018).
The approach taken here was a relatively straightforward one, where Pear-
son's r and Spearman's ρ was found for each predictor variable pair for each
chronological phase. In the case of problematic levels of correlation between
predictor variables across all phases ($|r \ or \ \rho| > 0.8$), all but one of the correl-

ating variables were removed from analysis. The choice of what variable to retain was based on what variable appeared to best summarise the collinear group in substantive terms (Table 2).

Following the lack of previous quantitative studies on which to base this analysis, linearity with the logit could not reasonably be assumed for the predictors in the logistic regression models. To allow for non-linearity, pre-specified natural splines with default quantile knots could have been included (Harrell 2015:26–28). However, with 104 sites for the EM representing the smallest effective sample size, the inclusion of more than 10 model parameters would likely lead to overfitting, following the most lenient $m/10$ rule provided by Harrell (2015:72–73), where m is effective sample size. Consequently, as there were no grounds on which to prioritise among what variables to include with splines, none were used. Similarly, due to the large number of possible two-way interaction effects, no interaction terms were included in the models as there were no grounds on which to prioritise among these (cf. Harrell 2015:36–38).

In total, shoreline emergence was missing for 32 sites and 725 non-sites, while infiltration level was given as unclassified for one site and 138 non-sites. Imputation of these missing values was deemed most appropriate. Unless the number of cases that have missing data is very low, or these miss data completely at random, which is rarely the case, the alternative of deleting observations can lead to dramatic loss of information and bias in all model estimates (Nakagawa and Freckleton 2008). Imputation for the logistic regression models was done using predictive mean matching (van Buuren and Groothuis-Oudshoorn 2011). This is based on performing OLS-regression on the column with missing values, using the other variables in the data as predictors. From the resulting coefficients a random draw is then made, and these coefficients are used to predict the entire column containing missing values, including those values that are not missing. Each missing value is then given the originally observed value of one of its predicted closest neighbours by a random draw, typically among the five closest cases. This process is then ideally repeated multiple times, and subsequent analysis is based on a pooling of the results from running each imputed dataset through the full analysis. As the imputations done for the logistic regression were to be nested in a bootstrap, a single imputation was performed per bootstrap sample (Brand et al. 2018).

Random forest provides its own method for imputing missing values based on its proximity and classification capabilities (Breiman 2003). This works by first setting the missing values to the average of the column and then growing and running the dataset through a random forest. For each classific-ation tree in the forest, each observation that ends up in the same terminal

Variable	Abbreviation	Measure	Collinearity	New name
Location (island/mainland)	loc	0 = island, 1 = mainland		
Island size	isl_si	Continous		
Infiltration ability	infil	Ordinal 1–4		
Deviation from south	dev_south	Continous		
Average fetch	fetch	Continous		
Viewshed 5 km	view_5	Continous	A: +	view
Viewshed 10 km	view_10	Continous	A: -	
Emergence 50 m	emerg_50	Continous	B: +	emerg_shdist
Emergence 500 m	emerg_500	Continous	B: -	
Emergence 1 km	emerg_1k	Continous	B: +	emerg_lgdist

Table 2: Variable names and handling of collinear variables. Letter indicate collinearity group and + indicates that the variable was retained and used to represent the group in subsequent analysis.

node as the observation with the missing value is counted as similar. The value to be imputed is then given as the average across the proximal observations, weighted by the number of times they ended up in the same terminal node. This entire process is then repeated using the new values instead of the column average. Here five sequential random forests were grown, with each forest growing 5000 trees. This process was then repeated five times, creating five imputed datasets that each were run through the estimation and validation process described above.

7 Results

The results from comparing sites and non-sites are given in Figures 5–10 and the results from comparing the different phases in Figures 13–14. It is striking that the two different sampling frames for non-sites largely return coinciding results for each phase, and that apart from some minor fluctuations, the random forests and logistic regression models perform about equally well. It is worth repeating that this has not been a controlled and balanced test of their potential performance, not least because the methods were subjected to different validation procedures. Confirming the findings by use of methods with different strength and weaknesses is nonetheless a benefit as this makes the final results more reliable, and in the case of discrepancies, more nuanced. This is especially true as the sample sizes did not allow for much complexity in the logistic regression models.

Moving on to the substantive results, Hosmer, Lemeshow and Sturdivant

(2013:177) state that AUC scores from 0.6–0.8 could generally be taken to indicate poor to acceptable discrimination, which is achieved when comparing sites to non-sites across all results given in Figures 5–10. This indicates that while the included variables are capturing some amount of patterning associated with the sites, a lot remains unexplained. The two exposure variables is driving most of the achieved result. This is irrespective of phase and whether the samples are from the hull or buffer constraints, and is evident in both variable importance estimates for the random forests and the coefficients of the logistic regression models. This is with the exception of the viewsize coefficient for the LM, which, while also tending towards positive values, is not significantly different from zero. The otherwise consistent nature of the relationship between view size and fetch length is somewhat surprising, however, where the coefficients indicate that larger view sizes and smaller average fetch lengths characterises the location of the sites. This pattern could consequently speak to a settlement strategy that favours open immediate surroundings that at the same time are sheltered from larger expanses of open sea. When sites are compared across phases using the percentile rank adjusted values, the exposure variables have a steady impact around zero, indicating that this is a fairly stable pattern throughout the Mesolithic.

While the only consistently significant variables for the EM is related to the exposure of sites, both the MM and LM sites have a tendency to be situated in locations with a high degree of variation in shoreline emergence in close proximity around the sites, while the LM sites also display a tendency to have stable surroundings within larger catchments. The most immediate behavioural implication of this result could be that the sites in both phases represent locations from where resources within a short distance were exploited, as this points towards ecological variation in close proximity of the sites. One perspective might also follow from the fact that the shoreline emergence variables are based on a transformation from elevation to year of emergence, which means that they are not entirely divorced from the relief of the local landscape, or possibly, depending on if the sites were strictly shore-bound or not, bathymetry. An implication of this could consequently be that the variation in the immediate proximity of the MM and LM sites is related to good fishing grounds (Darmark, Viken and Johannessen 2018; Mjærum and Mansrud 2020). The relevance of the significantly stable wider catchments in the LM could then be that this reflects a more predictable pattern for the exploitation of flora and fauna. As it is not treated directly here, a possible increase in the relevance of bathymetric variation in the MM and LM has to be left as a suggestion, however, and the nature of the apparent relevance of shoreline emergence as a whole could clearly benefit from being evaluated while also considering bathymetric and topographic variation

directly.

For the second set of models that compares the location of sites across phases, the island variables were driving most of the result simply due to the fact that the number of islands has varied dramatically throughout the Mesolithic (Figure 11). The results were therefore plotted leaving out the island variables to make the comparisons between the other variables discernable in the plots while still taking account of their effect. In addition, the accuracy scores are given as the accuracy achieved by the two island variables subtracted from the accuracy achieved by the full model. The minute effect sizes of the few significant coefficients in these models, combined with the fact that the variables apart from the island variables only contribute minute changes in the accuracy scores, indicates that the settlement patterns, excluding the island variables, largely coincide across all phases. Furthermore, following the logistic regression models comparing sites to non-sites, only sites from the MM appear to have a consistent significant propensity towards being situated relative to islands, where these have been slightly favoured, and sites have a higher tendency to be situated on smaller islands than could be expected by chance (Figure 7–8). While the corresponding random forests do not place as much emphasis on these variables, this might be caused by some collinearity between the variables, causing their importance to be split between the two. Their impact is, however, relatively small even when combined, which could indicate that even though significant in the logistic regression models, the island variables might not have been of a huge importance in the MM either. In summary, there are certainly not grounds on which to claim that the settlement patterns across the Mesolithic were the same with respect to the considered variables, but the overall findings do indicate that these differences have not been substantial.

8   Conclusion

This study was aimed at elucidating general settlement patterns among coastal sites from the Mesolithic of south-eastern Norway through statistical modelling and algorithmic classification. Although many of the employed procedures followed from idiosyncrasies in the data under analysis, presenting the modelling rationale and execution in some detail allowed for the demonstration of some concrete techniques with wider applicability, and hopefully contributes to an increased awareness of the necessity of employing thorough and conscious modelling strategies in archaeological research.

Given the lack of previous quantitative investigations on which to base this study, few assumptions concerning the nature of the inter-relationship

between independent variables and their relationship with the dependent variable could reasonably be made. This put a substantial pressure on the modelling procedures associated with the logistic regression models. While these aspects can have a large effect on the final models, their elucidation have to be balanced with a concern for the potentially adverse effects of model tuning and so-called data dredging that might undermine any findings. While an attempt at navigating this as best as possible was made, it was therefore also decided to apply the non-parametric machine learning method of random forest. This provided a robust, although less informative output with which to compare the more nuanced, but sensitive results of logistic regression. Given the similar performance of the two methods, as applied in this context, logistic regression should be given preference in the choice between the two, due to it's probabilistic and more informative output.

The main substantive findings were that the Mesolithic settlement patterns, as captured by the considered covariates, appear to largely coincide across the period in the region, and that as a whole these are mainly driven by the exposure variables. While there are some indications of diachronic variation, this is not substantial enough to warrant any decisive conclusion at this stage, and this variation does certainly not appear to be as marked as has commonly been suggested. Taking a cue from Åstveit (2014), this suggests that instead of seeing these societies as situated somewhere on a linear trend-line moving towards ever increasing degrees of sedentism and societal complexity, a more sensible point of departure in their study might be to consider them hunter-fisher-gatherer societies and to assume a null-hypothesis of no temporal difference. It is, however, worth underscoring the overall limited performance of the models, as improvement of these might alter the picture. Furthermore, the chronological phases have here been treated with no consideration of the stability or homogeneity of patterns within the individual phases, nor the nature of the transition between these. While it was possible to identify some patterns employing this framework, it is not given that these represent the best chronological categories for the aggregation of the sites (Reitan 2016). Similarly, no consideration of potential spatial variation has been considered, nor any discrepancies associated with different site types. These dimensions are likely to hold variation not discerned here, and as such represent clear candidates for further analysis. Nonetheless, as long as the scale and assumptions underlying the results are kept in mind when they are being interpreted, the coarser perspective and adherence to an established chronological framework arguably constitutes a reasonable starting point in the analysis of a material that has not been subject to formal quantitative treatment in the past.

Data accessibility statement

Data, code, figures, text files and associated licences for this paper is available as an online repository on Zenodo at 10.5281/zenodo.3948120 [blinded DOI] as well as on GitHub at [omitted for blinding].

Acknowledgements

References

Åstveit, LI, 2014. Noen synspunkt på den tidligmesolittiske bosetningen i Sør-Norge. *Primitive tider*, 16:87–104.

Baxter, MJ, 2003. *Statistics in Archaeology*. Chichester: John Wiley & Sons.

Bekkby, T, Isachsen, PE, Isæus, M and Bakkestuen, V, 2008. GIS modeling of wave exposure at the seabed: a depth-attenuated wave exposure model. *Marine Geodesy*, 31(2):117–127.

Berg-Hansen, IM, 2009. *Steinalderregistrering - Metodologi og forskningshistorie i Norge 1900–2000 med en feltstudie fra Lista i Vest-Agder*. Oslo: Museum of Cultural History, University of Oslo.

Bergsvik, KA, 1995. Bosetningsmønstre på kysten av Nordhordland i steinalder. En geografisk analyse. *Arkeologiske skrifter*, 8:111–130.

Bergsvik, KA, 2001. Sedentary and mobile hunterfishers in Stone Age western Norway. *Arctic Anthropology*, 38(1):2–26.

Bevan, A, Crema, E, Li, X and Palmisano, A, 2013. Intensities, interactions, and uncertainties: some new approaches to archaeological distributions. In: Bevan, A and Lake, M (eds.) *Computational approaches to archaeological spaces*, pp. 27–52. Walnut Creek: Left Coast Press.

Bjerck, HB, 2008. Norwegian Mesolithic trends: A review. In: Bailey, G and Spikins, P (eds.) *Mesolithic Europe*, pp. 60–106. Cambridge: Cambridge University Press.

Bjerck, HB, 2017. Settlements and seafaring: Reflections on the integration of boats and settlements among marine foragers in Early Mesolithic Norway and the Yámana of Tierra del Fuego. *The Journal of Island and Coastal Archaeology*, 12(2):276–299.

Bjerck, HB and Zangrando, AF, 2013. Marine ventures: Comparative perspectives on the dynamics of early human approaches to the seascapes of Tierra del Fuego and Norway. *The Journal of Island and Coastal Archaeology*, 8(1):79–90.

Brand, J, van Buuren, S, le Cessie, S and van den Hout, W, 2018. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. *Statistics in Medicine*, 38(2):210–220.

Breiman, L, 2001a. Random forests. *Machine learning*, 45(1):5–32.

Breiman, L, 2001b. Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231.

Breiman, L, 2003. *Manual - setting up, using, and understanding random forests V4. 0. 2003*. Available from: `https://www.stat.berkeley.edu/~breiman/papers.html`.

Breivik, HM, 2014. Palaeo-oceanographic development and human adaptive strategies in the Pleistocene–Holocene transition: A study from the Norwegian coast. *The Holocene*, 24(11):1478–1490.

Breivik, HM, Fossum, G and Solheim, S, 2018. Exploring human responses to climatic fluctuations and environmental diversity: Two stories from Mesolithic Norway. *Quaternary International*, 465:258–275.

Burnham, KP, Anderson, DR and Huyvaert, KP, 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1):23–35.

Burrows, MT, Harvey, R and Robb, L, 2008. Wave exposure indices from digital coastlines and the prediction of rocky shore community structure. *Marine Ecology Progress Series*, 353:1–12.

Chatfield, C, 1995. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 158(3):419–444.

Conolly, J and Lake, M, 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.

Darmark, K, Viken, S and Johannessen, LS, 2018. A good place. Stone Age locations in southern Norway: A diachronic GIS approach. In: Reitan, G and Sundström, L (eds.) *Kystens steinalder i Aust-Agder: Arkeologiske undersøkelser i forbindelse med ny E18 Tvedestrand-Arendal*, pp. 5–16. Oslo: Cappelen Damm Akademisk.

Derksen, S and Keselman, HJ, 1992. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2):265–282.

Dormann, CF, Elith, J, Bacher, S, Buchmann, C, Carl, G, Carré, G, Marquéz, JRG, Gruber, B, Lafourcade, B, Leitão, PJ, Münkemüller, T, McClean, C, Osborne, PE, Reineking, B, Schröder, B, Skidmore, AK, Zurell, D and Lautenbach, S, 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46.

Duncan, RB and Beckman, KA, 2000. The application of GIS predictive site location models within Pennsylvania and West Virginia. In: Wescott, KL and Brandon, RJ (eds.) *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit*, pp. 33–58. London: Taylor & Francis.

Ekebom, J, Laihonen, P and Suominen, T, 2003. A GIS-based step-wise procedure for assessing physical exposure in fragmented archipelagos. *Estuarine, Coastal and Shelf Science*, 57(5-6):887–898.

Eskeland, KF, 2017. E18 Langangen Rugtvedt, 16/06999, Porsgrunn og Bamble kommune. Tech. rep., Telemark County Municipality. Report, Archaeological survey.

Fisher, P, Farrelly, C, Maddocks, A and Ruggles, C, 1997. Spatial analysis of visible areas from the Bronze Age cairns of Mull. *Journal of Archaeological Science*, 24(7):581–592.

Fox, J, 2008. *Applied regression analysis and generalized linear models*. Throusand Oaks: Sage Publications, 2nd ed.

Fuglestvedt, I, 2012. The pioneer condition on the Scandinavian Peninsula: the last frontier of a 'Palaeolithic way' in Europe. *Norwegian Archaeological Review*, 45(1):1–29.

Garcia, A, 2013. GIS-based methodology for Palaeolithic site location preferences analysis. A case study from Late Palaeolithic Cantabria (Northern Iberian Peninsula). *Journal of Archaeological Science*, 40(1):217–226.

Geological Survey of Norway, 2015. *Produktspesifikasjon: ND_løsmasser, versjon 3.0*. Only available in Norwegian: https://www.ngu.no/en/topic/datasets.

Gillings, M and Wheatley, D, 2020. Gis-based visibility analysis. In: Gillings, M, Hacigüzeller, P and Lock, G (eds.) *Archaeological Spatial Analysis: A Methodological Guide*, pp. 313–332. London & New York: Routledge.

Glørstad, H, 2009. The northern province? The neolithisation of southern Norway. In: Glørstad, H and Prescott, C (eds.) *Neolithisation as if history mattered. Processes of Neolithisation in North-Western Europe*, pp. 135–168. Gothenburg: Bricoleur Press.

Glørstad, H, 2010. *The structure and history of the Late Mesolithic societies in the Oslo fjord area 6300–3800 BC*. Gothenburg: Bricoleur Press.

Harrell, FE, 2015. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. New York: Springer, 2nd ed.

Harris, MD, Kingsley, RG and Sewell, AR, 2015. *Archaeological Predictive Model Set*. Harrisburg: Commonwealth of Pennsylvania, Department of Transportation. Final report. Volume 7.

Hastie, T, Tibshirani, R and Friedman, J, 2017. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2nd ed.

Henderson, DA and Denison, DR, 1989. Stepwise regression in social and psychological research. *Psychological Reports*, 64(1):251–257.

Hill, NA, Pepper, AR, Puotinen, ML, Hughes, MG, Edgar, GJ, Barrett, NS, Stuart-Smith, RD and Leaper, R, 2010. Quantifying wave exposure in shallow temperate reef systems: applicability of fetch models for predicting algal biodiversity. *Marine Ecology Progress Series*, 417:83–95.
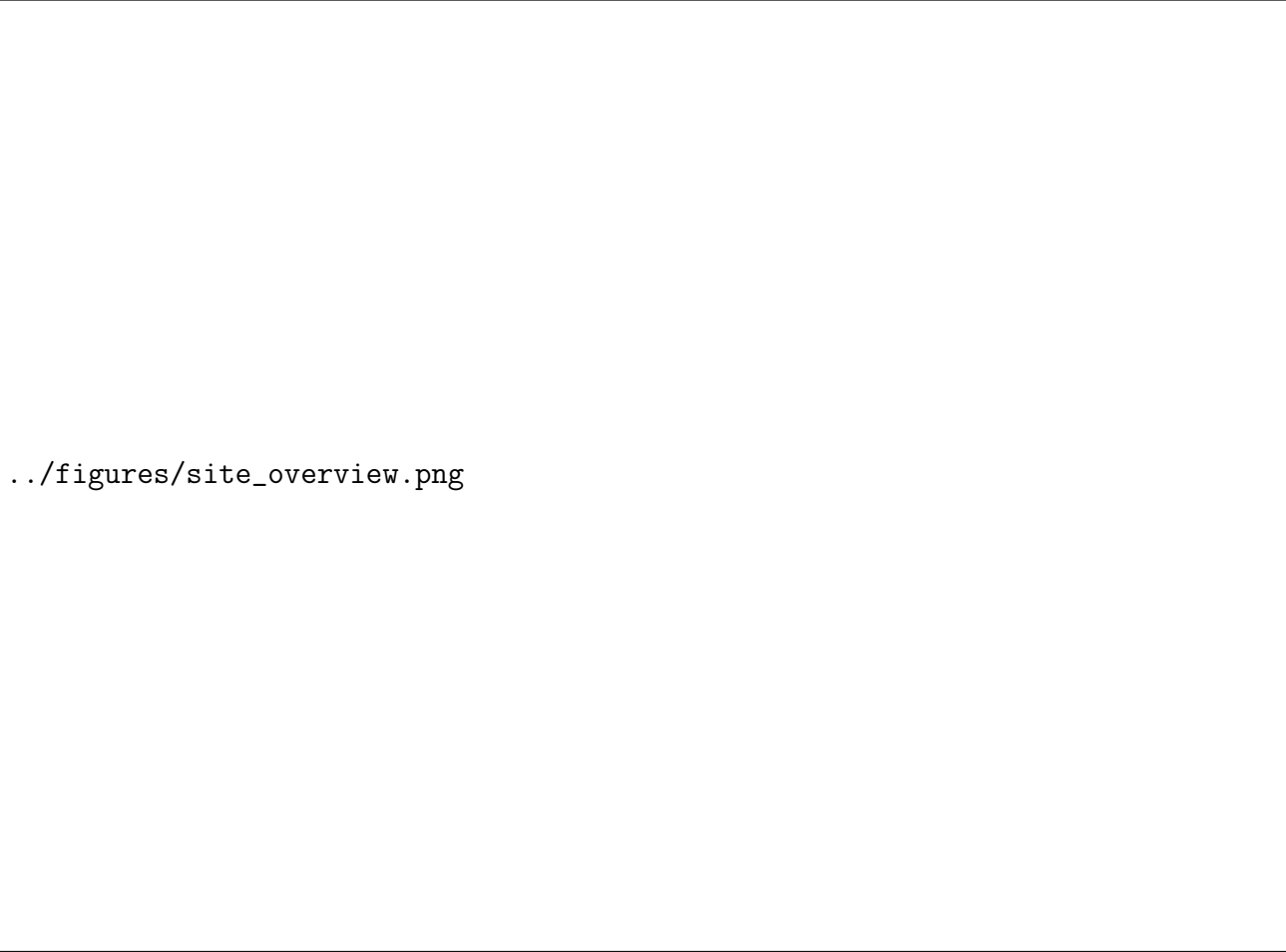
Hosmer, DW, Lemeshow, S and Sturdivant, RX, 2013. *Applied logistic regression*. John Wiley & Sons, 3rd ed.

Jaksland, L, 2001. *Vinterbrolokalitetene: en kronologisk sekvens fra mellom- og senmesolitikum i Ås, Akershus*. Oslo: Museum of Cultural History, University of Oslo.

Jaksland, L and Persson, P (eds.), 2014. *E18 Brunlanesprosjektet. Bind 1. Forutsetninger og kulturhistorisk sammenstilling*. Oslo: Museum of Cultural History, University of Oslo.

James, G, Witten, D, Hastie, T and Tibshirani, R, 2013. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

Kohler, TA and Parker, SC, 1986. Predictive models for archaeological resource location. In: Schiffer, MB (ed.) *Advances in archaeological method and theory*, vol. 9, pp. 397–452. New York: Academic Press.

Kvamme, KL, 1988. Development and testing of quantitative models. In: Judge, JW and Sebastian, L (eds.) *Quantifying the present and predicting the past: theory, method, and application of archaeological predictive modeling*, pp. 325–428. Denver: US Department of the Interior, Bureau of Land Management.

Kvamme, KL, 2020. Analysing regional environmental relationships. In: Gillings, M, Hacigüzeller, P and Lock, G (eds.) *Archaeological Spatial Analysis: A Methodological Guide*, pp. 212–230. London & New York: Routledge.

Kvamme, KL and Jochim, MA, 1989. The environmental basis of Mesolithic settlement. In: Bonsall, C (ed.) *The Mesolithic in Europe. Papers presented at the third international symposium, Edinburgh 1985*, pp. 1–12. Edinburgh: John Donald.

Laing, AK, 1998. An introduction to ocean waves. In: *Guide to Wave Analysis and Forecasting*, pp. 1–14. Geneva: World Meteorological Organization, 2nd ed.

Lake, M and Woodman, PE, 2000. Viewshed analysis of site location on Islay. In: Mithen, S (ed.) *Hunter–Gatherer Landscape Archaeology: The Southern Hebrides Mesolithic Project 1988–98. Vol. 2: Archaeological fieldwork on Colonsay, computer modelling, experimental archaeology, and final interpretations*, pp. 497–503. Cambridge: McDonald Institute for Archaeological Research.

Lindblom, I, 1984. Former for økologisk tilpasning i Mesolitikum, Østfold. *Universitetets Oldsaksamling Årbok 1982/1983*, pp. 43–86.

Lopez-Romero, EGdlA, 2008. Characterizing the evolution of visual landscapes in the late prehistory of south-west Morbihan (Brittany, France). *Oxford Journal of Archaeology*, 27(3):217–239.

Malek, MH, Berger, DE and Coburn, JW, 2007. On the inappropriateness of stepwise regression analysis for model building and testing. *European Journal of Applied Physiology*, 101(2):263–264.

Malhotra, A and Fonseca, MS, 2007. WEMo (Wave Exposure Model): formulation, procedures and validation. Tech. rep., National Ocean Service of the National Oceanic and Atmospheric Administration. NOAA Technical Memorandum NOS NCCOS #65.

Mansrud, A, 2014. Mobil eller bofast? Erverv, landskap og mobilitet i mellommesolittiske kystsamfunn i Øst-Norge (8300–6300 f. Kr.). *Norsk Maritimt Museums årbok*, 2013:67–108.

Melvold, S and Persson, P (eds.), 2014. *Vestfoldbaneprosjektet – Arkeologiske undersøkelser i forbindelse med ny jernbane mellom Larvik og Porsgrunn. Bind 1*. Oslo: Museum of Cultural History, University of Oslo.

Mikkelsen, E, 1989. *Fra jeger til bonde: utvklingen av jordbrukssamfunn i Telemark i steinalder og bronsealder*. Oslo: Universitetets Oldsaksamling, University of Oslo.

Mjærum, A and Mansrud, A, 2020. Resource management in Late Mesolithic Eastern Norway? Fishing in the coastal, interior and mountain areas and its socio-economic implications. In: Schülke, A (ed.) *Coastal Landscapes of the Mesolithic, Human engagement with the coast from the Atlantic to the Baltic sea*, pp. 264–299. London & New York: Routledge.

Nakagawa, S and Freckleton, RP, 2008. Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*, 23(11):592–596.

Nielsen, SV, Åkerstrøm, J, Stokke, JSF and Eskeland, KF, 2016. Quartz utilization along the coast of southern Norway: Results from a Stone Age survey in Aust-Agder. In: Bjerck, HB, Breivik, HM, Fretheim, E, Piana, B, Skar, A, Tivoli, A and Zangrando, AFJ (eds.) *Marine Ventures: Archaeological Perspectives on Human-Sea Relations*, pp. 367–381. Sheffield: Equinox Publishing.

Nitter, M and Coolen, J, 2018. Any way the wind blows... Wind fetch as a determinant factor of the quality of landing sites. In: von Carnap-Bornheim, C, Daim, F, Ettel, P and Warnke, U (eds.) *Harbours as objects of interdisciplinary research – archaeology + history + geosciences*, pp. 45–56. Mainz: Verlag des Römisch-Germanischen Zentralmuseums.

Nitter, M, Elvestad, E and Selsing, L, 2013. Maritime site protection and the fetch method: an example from Rogaland, Norway. *International Journal of Nautical Archaeology*, 42(1):87–102.

Nyland, AJ, 2012. Lokaliseringsanalyse av tidligmesolittiske pionerboplasser. In: Glørstad, H and Kvalø, F (eds.) *Havvind: Paleogeografi og arkeologi*, pp. 70–96. Oslo: Norwegian Maritime Museum.

Parr, T, Turgutlu, K, Csiszar, C and Howard, J, 2018. Beware default Random forest importances. Explained.ai: https://explained.ai/rf-importance/index.html.

Reitan, G, 2016. Mesolittisk kronologi i Sørøst-Norge – et forslag til justering. *Viking*, 79:23–51.

Reitan, G and Persson, P (eds.), 2014. *Vestfoldbaneprosjektet–Arkeologiske undersøkelser i forbindelse med ny jernbane mellom Larvik og Porsgrunn. Bind 2.* Oslo: Museum of Cultural History, University of Oslo.

Rufibach, K, 2010. Use of Brier score to assess binary predictions. *Journal of Clinical Epidemiology*, 63(8):938–939.

Schmitt, L, 2015. Early colonization, glacial melt water, and island mass effect in the archipelago of Western Sweden: A case history. *Oxford Journal of Archaeology*, 34(2):109–117.

Solheim, S (ed.), 2017. *E18 Rugtvedt-Dørdal. Arkeologiske undersøkelser av lokaliteter fra steinalder og jernalder i Bamble kommune, Telemark fylke.* Oslo: Museum of Cultural History, University of Oslo.

Solheim, S, 2020. Mesolithic coastal landscapes. Demography, settlement patterns subsistence economy in southeastern norway. In: Schülke, A (ed.) *Coastal Landscapes of the Mesolithic, Human engagement with the coast from the Atlantic to the Baltic sea*, pp. 44–72. London & New York: Routledge.

Solheim, S and Damlien, H (eds.), 2013. *E18 Bommestad-Sky. Undersøkelser av lokaliteter fra mellommesolitikum, Larvik kommune, Vestfold fylke.* Oslo: Museum of Cultural History, University of Oslo.

Solheim, S and Persson, P, 2016. Marine adaptation in the Middle Mesolithic of south-eastern Norway. In: Bjerck, HB, Breivik, HM, Fretheim, E, Piana, B, Skar, A, Tivoli, A and Zangrando, AFJ (eds.) *Marine Ventures: Archaeological Perspectives on Human-Sea Relations*, pp. 255–270. Sheffield: Equinox Publishing.

Sørensen, R, Henningsmoen, KE, Høeg, H and Gälman, V, 2014. Holocene landhevningsstudier i søndre Vestfold og sørøstre Telemark – revidert kurve. In: *Vestfoldbaneprosjektet. Arkeologiske undersøkelser i forbindelse med ny jernbane mellom Larvik og Porsgrunn kommune. Bind 1*, pp. 236–247. Oslo: Museum of Cultural History, University of Oslo.

Sørensen, R, Høeg, H and Gälman, V, 2015. Revidert strandlinjeforskyvningskurve for bamble. Tech. rep. Report for the Project E18 Rugtvedt-Dørdal. Museum of Cultural History, Unversity of Oslo.

Spencer, C and Bevan, A, 2018. Settlement location models, archaeological survey data and social change in Bronze Age Crete. *Journal of Anthropological Archaeology*, 52:71–86.

Stančič, Z and Veljanovski, T, 2000. Understanding Roman settlement patterns through multivariate statistics and predictive modelling. In: Lock, G (ed.) *Beyond the Map: Archaeology and Spatial Technologies*, pp. 147–156. Amsterdam: IOS Press.

Stoltzfus, JC, 2011. Logistic regression: a brief primer. *Academic Emergency Medicine*, 18(10):1099–1104.

Sundblad, G, Bekkby, T, Isæus, M, Nikolopoulos, A, Norderhaug, K and Rinde, E, 2014. Comparing the ecological relevance of four wave exposure models. *Estuarine, Coastal and Shelf Science*, 140:7–13.

Svendsen, F, 2014. Kommentar til Leif Inge Åstveit:≪Noen synspunkt på den tidlig-mesolittiske bosetningen i Sør-Norge≫. *Primitve Tider*, 16:121–127.

Tolvanen, H and Suominen, T, 2005. Quantification of openness and wave activity in archipelago environments. *Estuarine, Coastal and Shelf Science*, 64(2-3):436–446.

Tomaschek, F, Hendrix, P and Baayen, RH, 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71:249–267.

van Buuren, S and Groothuis-Oudshoorn, K, 2011. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1–67.

Verhagen, P, 2009. Predictive models put to the test. In: Kamermans, H, van Leusen, M and Verhagen, P (eds.) *Archaeological Prediction and Risk Management. Alternatives to Current Practice*, pp. 71–122. Leiden: Leiden University Press.

Verhagen, P and Whitley, TG, 2012. Integrating archaeological theory and predictive modelling: A live report from the scene. *Journal of Archaeological Method and Theory*, 19(1):49–100.

Verhagen, P and Whitley, TG, 2020. Predictive spatial modelling. In: Gillings, M, Hacigüzeller, P and Lock, G (eds.) *Archaeological Spatial Analysis: A Methodological Guide*, pp. 231–246. London & New York: Routledge.

Visentin, D and Carrer, F, 2017. Evaluating Mesolithic settlement patterns in mountain environments (Dolomites, Eastern Italian Alps): the role of research biases and locational strategies. *Archeologia e Calcolatori*, 28(1):129–154.

Wachtel, I, Zidon, R, Garti, S and Shelach-Lavi, G, 2018. Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north Israel and north-east China. *Journal of Archaeological Science*, 92:28–36.

Warren, RE and Asch, DL, 2000. A predictive model of archaeological site location in the Eastern Prairie Peninsula. In: Wescott, KL and Brandon, JR (eds.) *Practical applications of GIS for archaeologists: a predictive modeling kit*, pp. 5–32. London: Taylor & Francis.

Woodman, PE and Woodward, M, 2002. The use and abuse of statistical methods in archaeological site location modelling. In: Wheatley, D, Earl, G and Poppy, S (eds.) *Contemporary themes in archaeological computing*, pp. 39–43. Oxford: Oxbow.

Wren, CD, Costopoulos, A and Hawley, M, 2018. Settlement choice under conditions of rapid shoreline displacement in Wemindji Cree Territory, subarctic Quebec. *Quaternary International*. In press.

Figure 1: Location of the study area in Norway, and overview of the sites. The green points represent the 478 sites that were initially considered for further analysis. These were subsequently reduced to 462 to only include sites given a shoreline date to the Mesolithic.
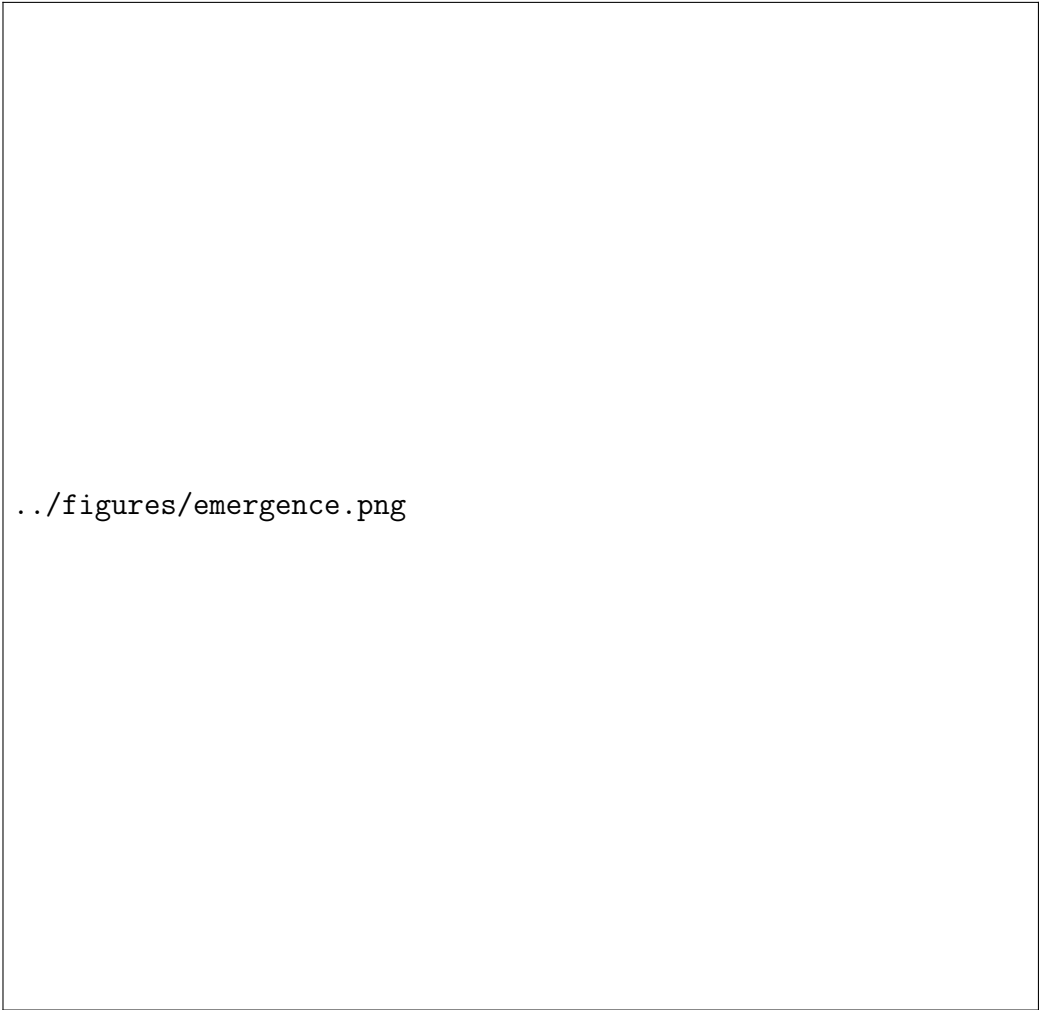
Figure 2: Shoreline displacement curves for the study region (after Sørensen et al. 2014; Sørensen, Høeg and Gälman 2015). The transitions between the chronological phases are given by the vertical lines. The Bamble curve covers Bamble municipality, and the Gunnarsrød curve Porsgrunn and Larvik (Figure 2).

Figure 3: Average fetch as estimated from site with Askeladden ID 22346-1. A) Sea-level has been raised to 34 meters above present. The lines are drawn from the point that is the average coordinate of the underlying site polygon. These are then clipped as they intersect the polygons representing land from the vectorised DTM with 10 m resolution. B) The five fetch lines that escape the inlet in which the site is located continue. C) Outside of the area covered by the regional DTM the present day shoreline is used for clipping the lines. This is based on a 25 m resolution DTM.

../figures/emergence.png

Figure 4: Reclassified elevation raster by year of emergence given in calendar years. The area west of the red line has been reclassified using the mean of the Bamble displacement curve, and the area to the east using the mean of the Gunnarsrød displacement curve (Figure 2).

Figure 5: Early Mesolithic - Hull sample. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

Figure 6: Early Mesolithic - Buffer sample. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

../figures/mm_hull_fig.png

Figure 7: Middle Mesolithic - Hull sample. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

../figures/mm_buff_fig.png

Figure 8: Middle Mesolithic - Buffer sample. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

Figure 9: Late Mesolithic - Hull sample. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

Figure 10: Late Mesolithic - Buffer sample. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.
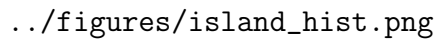
../figures/island_hist.png

Figure 11: Number of random samples on islands and mainland across the different phases. It is evident that this proportion has varied dramatically throughout the Mesolithic.

Figure 12: Early Mesolithic - Middle Mesolithic. The two island variables have been excluded from the plots, and the accuracy scores are given as the difference between that achieved by the full model and only by the island variables. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

Figure 13: Early Mesolithic - Late Mesolithic. The two island variables have been excluded from the plots, and the accuracy scores are given as the difference between that achieved by the full model and only by the island variables. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.

Figure 14: Middle Mesolithic - Late Mesolithic. The two island variables have been excluded from the plots, and the accuracy scores are given as the difference between that achieved by the full model and only by the island variables. A) Map overview. B) Logistic regression, percentile bootstrap. C) Random forest, permutation variable importance.