Maximum Entropy Modelling Result Paper for First Phase Development of the New Territories North – San Tin / Lok Ma Chau Development Node – Investigation

TABLE OF CONTENTS

List of Tables

List of Plates

1 INTRODUCTION

1.1 Project Background

1.1.1.1 San Tin / Lok Ma Chau Development Node (STLMC DN, or the Project) is located to the west of Kwu Tung North and Fanling North New Development Areas (NDAs) and Fanling and Sheung Shui New Towns, and to the northeast of Yuen Long and Tin Shui Wai New Towns. STLMC DN is bisected by San Tin Highway into northern and southern parts, and bounded by the Shenzhen River and Sam Po Shue wetland to the north; LMC Loop to the northeast; some village settlements and Ki Lun Shan to the east; San Tin Barracks and Ngau Tam Shan to the south; and Tam Mei Barracks and some residential developments to the southwest. Existing Shek Wu Wai village and the village clusters bounded by San Tin Highway, San Tin Tsuen Road and Tung Wing On Road are excluded from the boundary of STLMC DN. The Project boundary of STLMC DN is about 611 ha.

1.1.1.2 Under the EIA for this Project, a cultural heritage impact assessment (CHIA) is required to be conducted in accordance with the criteria and guidelines as stated in Section 2 of Annexes 10 and 19 of the *Technical Memorandum on Environmental Impact Assessment Process* (EIAO-TM) as well as the requirements given in Clause 3.4.13 and Appendix L of the Study Brief (No. ESB-340/2021).

1.1.1.3 As part of the archaeological impact assessment, archaeological fieldwork has been conducted between November 2022 and December 2022. While the archaeological fieldwork under the EIA has conducted in as much accessible area as possible within the Project area, it is still not possible to establish conclusive result based purely on the fieldwork.

1.1.1.4 In the light of this, archaeological predictive modelling is utilised to help establish the archaeological potential within the Project boundary. The archaeological potential makes up of the archaeological findings from the archaeological fieldwork and the result from the archaeological prediction model in the following sections.

1.2 The Use of GIS Prediction Models in Archaeology

1.2.1.1 Geographical Information System (GIS) is a computer system for collecting, storing, integrating, visualising, and analysing of spatial data. The use of GIS and related spatial technologies have played a significant role in archaeology, especially to understand human and their relationship with the natural environment[1]. One of earliest and widespread uses of GIS in archaeology is the development of predictive model.

1.2.1.2 Predictive modelling is a quantitative method aiming to predict the possibility of archaeological presence at a given location in different environment and period. This is done based on the premise that a certain type of archaeological site tends to occur at the same kind of "place"[2], where "place" can be quantified into different kind of environmental spatial data, including elevation, aspect and gradient, etc. This model is useful when studying a widespread extent, where not every location can be studied comprehensively.

1.2.1.3 In predictive modelling for archaeological potential, there are usually two classes of training data needed: presence and absence data. The training data are used to train the prediction model to classify presence data and absence data, in case of archaeology, classifying location of having and not having archaeological site[3]. Presence/Absence data refers respectively to the existence/lack of archaeological data at a particular location. In archaeology, however, while presence data refers to the location of where archaeological

[1] Conolly, J. 2008. Geographical Information Systems and Landscape Archaeology. In David, B. and Thomas, J. Eds. *Handbook of Landscape Archaeology*. 583–595.

[2] Refrew, C. & Bahn, P. (2016). *Archaeology: Theories, Methods and Practice (6th ed.)*. United Kingdom: Thames & Hudson.

[3] Bickler, S. (2021). Machine Learning Arrives in Archaeology. *Advances in Archaeological Practice*. 9(2). 186-191. https://doi.org/10.1017/aap.2021.6.

sites may exist, the absence data does not usually mean a "no". In fact, there is no true absence data since the absence of site is not equivalent to the location being not suitable for past human settlement, it is just simply no data from archaeological survey. Therefore, a model that does not rely upon having absence data is required.

1.2.1.4 One such approach is the usage of *Maximum Entropy Modelling* (MaxEnt). MaxEnt is a machine learning model which has been widely used as ecological niche or species distribution modelling. The model compares environmental variables and the location of the presence of the species, and calculates the probability based on the above premise on predictive modelling[4]. For applying MaxEnt modelling in archaeology, one of the advantages of is that it requires the sampling of presence archaeological sites data only[5]. The ability to use presence-only data avoided the problems of having to create a set of site absence data. The presence data is used to calculate against the environmental variables that generate a model for the interpretation by archaeologists with quantitative results. Therefore, MaxEnt modelling is adopted to develop archaeological prediction model for this Project.

1.3 Real-life Applications of Archaeological Prediction Models

1.3.1.1 There are examples of MaxEnt being used in predictive modelling, such as the prediction of hunter-gatherer sites in prehistoric archaeological site in Fuxin (China) and Upper Galilee (Israel)[6], cave sites in Indonesia[7], Southern Pampa (Argentina)[8].

1.3.1.2 The study of prehistoric sites in Fuxin, China and Upper Galilee, Israel makes use of presence data of the known locations of prehistoric settlements. The study also aims to compare two methods of prediction, logistic regression and MaxEnt. Both methods are able to generate useful predictions in the two areas. In this study, MaxEnt is shown to be more efficient since the model can be constructed with a small dataset and it only requires archaeological site presence data, while logistic regression requires both archaeological site presence and absence data.

1.3.1.3 The study of prehistoric cave site in Indonesia makes use of presence data of known prehistoric cave sites. The study has three objectives, including:

1. to extract environmental data from various sources related to the cave sites;
2. producing a spatial model from the environmental data for input in MaxEnt model; and
3. to produce a prediction from the MaxEnt model and test the accuracy of how well it predicts the location of prehistoric cave sites.

1.3.1.4 The environmental variables used in the study are valley-hill classification map, distance to valley base, elevation, slope, aspect, distance to watercourse, lineament density and distance to lineament. The model performs reasonably well at identifying environments that has archaeological potential based on the environmental variables considered. This suggests the model is useful for highlighting areas that warrant further archaeological investigation.

1.3.1.5 The study in Argentina makes use of presence data of hunter-gatherer sites to generate a prediction using the MaxEnt model. The study utilised the controlling and adjustment of bias and model parameter of the maxent model to produce an effective prediction. The

[4] Kvamme, K.L. (2006). There and Back Again: Revisiting Archaeological Locational Modeling. In Mehrer, M.W. and Wescot, K.L. Eds. *GIS and Archaeological Site Location Modeling*. United Kingdom: Routledge. 3-38.

[5] Yaworsky, P.M. et al. (2020). Advancing Predictive Modeling in Archaeology: An Evaluation of Regression and Machine Learning Methods on the Grand Stair-case Learning Methods on the Grand Staircase-Escalante National Monument. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0239424.

[6] Wachtel, Ido & Zidon, Royi & Garti, Shimon & Shelach-Lavi, Gideon. (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. 22-36. https://doi.org/10.1016/j.jas.2018.02.001.

[7] Luthfi, A. M., Sigit, H. M. and Bowo, S. (2019). MaxEnt (Maximum Entropy) model for predicting prehistoric cave sites in Karst area of Gunung Sewu, Gunung Kidul, Yogyakarta. *Proceedings of the SPIE*. 113110B. https://doi.org/10.1117/12.2543522

[8] Rafuse, D.J., 2021. A Maxent Predictive Model for Hunter-Gatherer Sites in the Southern Pampas, Argentina. *Open Quaternary*, 7(1), p.6. https://doi.org/10.5334/oq.97.

study also aimed to identify the key environmental factors that affect the distribution of hunter-gatherer sites in the southern Pampas region. Environment variables used in the study are elevation, slope, aspect, Topographic Wetness Index, distance to watercourse, distance to waterbody and distance to toolstone source. The study demonstrated that MaxEnt is good at capturing real patterns in how the archaeological potential prediction responds to the environmental variables used in a statistically significant way.

2 METHODOLOGY

2.1.1.1 This section discusses the methodology in developing the archaeological prediction models for STLMC DN.

2.2 Data Collection and Preparation

2.2.1 ArcMap

2.2.1.1 ArcMap is a geospatial processing program primarily used to view, edit, create and analyse geospatial data. In this study, ArcMap version 10.8.1 is used to process and visualize the input archaeological sites and environmental variable data and the output of the prediction model.

2.2.2 Consideration over the Extent of The Training Data for MaxEnt Modelling

2.2.2.1 Sea level affects the local landscapes in Hong Kong and result into different scenario of the landscape in the Project area. The sea level changes became more or less stable at about 2,000BP, which this period overlapped with the historical period in Hong Kong[9]. Hence, for a better use of the GIS predictive modelling using MaxEnt, we conduct two models, one for prehistoric (before Bronze Age) and the other for historical period (from *Tang* to *Qing* dynasties) respectively. Currently, there lacks very solid understanding of human settlement (or the lack of) after the Bronze Age and before the Tang dynasty, and there is no substantial presence data available from any archaeological sites in Hong Kong that could provide useful understanding for this period of time. Hence it is assumed, based on current academic understanding, that there is no human settlement in Hong Kong between after Bronze Age and before the Tang dynasty.

2.2.2.2 The extent of the historical prediction model covers the Project area and the 500m assessment area that projected from the Project area boundary. The extent would allow the MaxEnt model to consider all traditional villages in San Tin and those at Mai Po in the west, the Mai Po Site of Archaeological Interest, as well as the landscape information within. The *Man* clan has settled in this area since the historical period, and thus the presence data would be adequate.

2.2.2.3 For the model to generate useful prediction result, presence data is required to represent possible archaeological sites as training data for the MaxEnt model. However, there is no known prehistoric presence data within the 500m assessment area for this Project. Lacking such data would make the MaxEnt model generate no useful result because there is nothing for the model to identity workable variables for prehistoric sites. In the light of this, we include presence data of the prehistoric period from Tuen Mun, Yuen Long, Sheung Shui and Fanling by including sites of archaeological interest with prehistoric discoveries as presence data. This allows MaxEnt to calculate the environmental variables over the prehistoric presence data.

2.2.3 Presence Dataset: Archaeological Sites Data

Sites of Archaeological Interest (SAIs)

2.2.3.1 Known SAIs in Hong Kong have been defined and recorded by AMO. The areas covered by boundaries of SAIs could be translated into spatial data. Hence, the extents of the SAIs are utilized so that these areas represent an area of higher archaeological potential for both historical and prehistoric sites.

Traditional Villages

2.2.3.2 The extents of traditional villages are also utilized to reflect settlement patterns and human decision making when selecting habitable landscape in Hong Kong. It is noted that some

[9] Fyfe, J. A et al. (2000). The Quaternary Geology of Hong Kong. Hong Kong: Civil Engineering Department.

traditional villages overlapped or are in vicinity of SAIs with archaeological potential in the historical period (such as Siu Hang Tsuen in Tuen Mun and Ping Che in Fanling). This suggests that traditional villages are useful in reflecting the environmental preference of settlements by past humans during historical period. The extent of each traditional village is also used as an indicator of presence of archaeological potential in the historical period.

Existing Archaeological Field Survey Data

2.2.3.3 Existing location with discovery of archaeological materials from field survey are also utilised as presence data. These findings include archaeological findings from previous archaeological investigation as well as archaeological survey conducted under this Project (**Section 12.6.2** in EIA report refers), where findings from the historical period can be identified in various field scanning location at A5, A10 and A12. The locations of findings are translated into spatial data as presence data for the model.

2.2.3.4 It is noted that previous archaeological survey did not identify archaeological materials (**Section 12.4.3** in EIA report refers). Thus, past survey in this area has no contribution to the predictive modelling.

2.2.4 Environmental Variables

2.2.4.1 Environmental variables refer to factors such as aspect, elevation, slope and geology that may or may not have influenced where past humans chose to settle and form sites. These variables can be represented as *raster* maps or *vector* layers in a GIS. When compared to the locations of known archaeological potential used to train the model, they can reveal the environmental conditions that are most commonly associated with where potential sites tend to occur. The MaxEnt model examines the variables and conditions that are most closely associated with the location of the training sites. It then predicts that other areas within the study area possessing those same suitable environmental characteristics likely have potential to contain archaeological remains, even if no sites have yet been discovered there. This allows the model to identify locations of higher archaeological potential to help focus surveys and investigations efforts.

2.2.4.2 Both raster and vector datasets are processed into *ASCII* format to be utilised in the MaxEnt Model.

2.2.4.3 The environmental variables used in the model were chosen based on variables that had been shown to be relevant in similar past studies applying MaxEnt modeling to archaeological research (**Section 1.3** refers). The variables likely to influence site locations in this area include:

- ***Aspect***: certain aspects may be preferable, especially in some climates.
- ***Elevation***: many sites tend to form on level ground at certain elevation, avoiding extremes.
- ***Slope***: gentler slopes are usually more suitable for human settlement and occupation.
- ***Geology***: different rock types and soil compositions influence the habitability and resource potential of a location.
- ***Distance to coast***: coastal zones often concentrate sites, though this may vary by time period and culture.
- ***Distance to alluvial deposits***: floodplains and river valleys historically attracted human settlement, though original river channels have since changed course over time.
- ***Distance to Hill***: sites may cluster closer to hills for reasons of defense, aspect, or resource access.
- ***Direction to Hill***: the aspect in relation to hills may have been important for some past groups.

2.2.4.4 Together, these environmental variables aim to capture major landscape factors that could have influenced where and why past humans chose to settle within the study area. By comparing the variable conditions present at known archaeological sites to those across the entire region, the MaxEnt model seeks to identify which combinations of variables and variable ranges define the most suitable environments for site formation.

2.2.4.5 A number of other environmental datasets were taken into account as input for the prediction model, such as roughness of land, greenness of the location and distance to modern watercourse, etc. However, these other datasets were excluded due to them being random, irrelevant or disturbed by modern landscaping. Furthermore, test runs with different variables during model development help identifying environmental factors that are influencing the predictions. Higher influence indicates the variable should be used in the model.

2.2.4.6 The environmental variables used in the study are presented in Table 1.

Table 1 Environmental Variables for MaxEnt Modelling

<table>
<tr><th>Presence Data</th><th colspan="2">Source</th></tr>
<tr><td>Sites of Archaeological Interest (SAIs)[10]</td><td colspan="2" rowspan="3">Prehistoric:
- Sample points of SAIs with archaeological interest of prehistoric period; and
- Existing Archaeological Field Survey Data
Historical:
- Sample points from area of SAIs with archaeological interests of historical period;
- Traditional villages; and
- Existing Archaeological Field Survey Data.</td></tr>
<tr><td>Traditional Villages</td></tr>
<tr><td>Existing Archaeological Fieldwork Data</td></tr>
<tr><th>Environmental Variable</th><th>Source</th><th>Unit</th></tr>
<tr><td>Aspect</td><td rowspan="3">Hong Kong climate, vegetation, and topography raster data[11]</td><td>Degree (°)</td></tr>
<tr><td>Elevation</td><td>Meter (m)</td></tr>
<tr><td>Slope</td><td>Degree (°)</td></tr>
<tr><td>Geology</td><td>CEDD Geological Map[12]</td><td>Geological Deposit</td></tr>
<tr><td>Distance to Coast*</td><td>Ancient coastline extracted based on marine deposits. Distant calculated with Euclidean Distance.</td><td>Meter (m)</td></tr>
<tr><td>Distance to Alluvial Deposits**</td><td>Watercourse extracted based on geology. Distant calculated with Euclidean Distance.</td><td>Meter (m)</td></tr>
<tr><td>Distance to Hill***</td><td>Hilly landscape extracted based on solid geology. Distant calculated with Euclidean Distance.</td><td>Meter (m)</td></tr>
<tr><td>Direction to Hill****</td><td>Hilly landscape extracted based on solid geology. Direction calculated with Euclidean Direction.</td><td>Degree (°)</td></tr>
</table>

* Value of closest distance to marine deposits
** Value of closest distance to alluvial deposit (Qa)
*** Value of closest distance to solid geology
**** Value of bearing to closest solid geology

2.3 Model Development

2.3.1.1 The MaxEnt algorithm is ran with the software developed by Steven J. Philips, Miroslav Dudík and Robert E. Schapire[13] using the presence archaeological site data and environmental variables as input. Two separate models are developed for the prediction of archaeological potential for the prehistoric and historical periods.

2.3.1.2 The MaxEnt model allows different parameters settings that can maximise the effectiveness of the model, thus improving the accuracy. Multiple tests runs were

[10] Antiquities and Monuments Office. *List of Sites of Archaeological Interest in Hong Kong (as at Nov 2012)*. Retrieve from https://www.amo.gov.hk/filemanager/amo/common/form/list_archaeolog_site_eng.pdf.

[11] Morgan, B. and Guénard, B. 2019. New 30m Resolution Hong Kong Climate, Vegetation and Topography Rasters Indicate Greater Spatial Variation Than Global Grids within Urban Mosaic. *Earth System Science Data*. 11(3). 1083-1098. https://doi.org/10.5194/essd-11-1083-2019.

[12] Civil Engineering and Development Department. *Geological Maps in 1:20,000 & 1:100,000*. Retrieved from https://www.geomap.cedd.gov.hk/GEOOpenData/eng/GeologicalMap.aspx.

[13] Philips, S.J., Dudík, M. and Schapire, R.E. (2023). Maxent Software for Modeling species Niches and Distribution (Version 3.4.1). Available from url: https://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2023-4-28.

conducted using different parameters, environmental variables, presence data and processing extent. In this study, the model is set with the default values of the software.

2.3.1.3 The model generates results of a continuous prediction map in *ASCII* format. The model also produces:

1. a table showing the 'percent contribution' and 'permutation importance', indicating how each variable contributes to the model and the significance of each variable on the model.
2. response curves of each environmental variable, it displays the predicted probability against the values of each environmental variable, which is useful in identifying the range of each environmental variable with higher predicted probability.
3. an Area Under Curve (AUC) output which is used to evaluate the effectiveness of the model.

2.3.1.4 The prediction map is then overlayed with a No Potential layer. This included areas with 1) modern development (i.e. road and pond) and 2) marine superficial deposits, where the archaeological deposits have already been destroyed. Also, marine deposits (Qm on geological map) that indicated the extent of ancient sea is also considered No Potential layer.

2.4 Model Evaluation

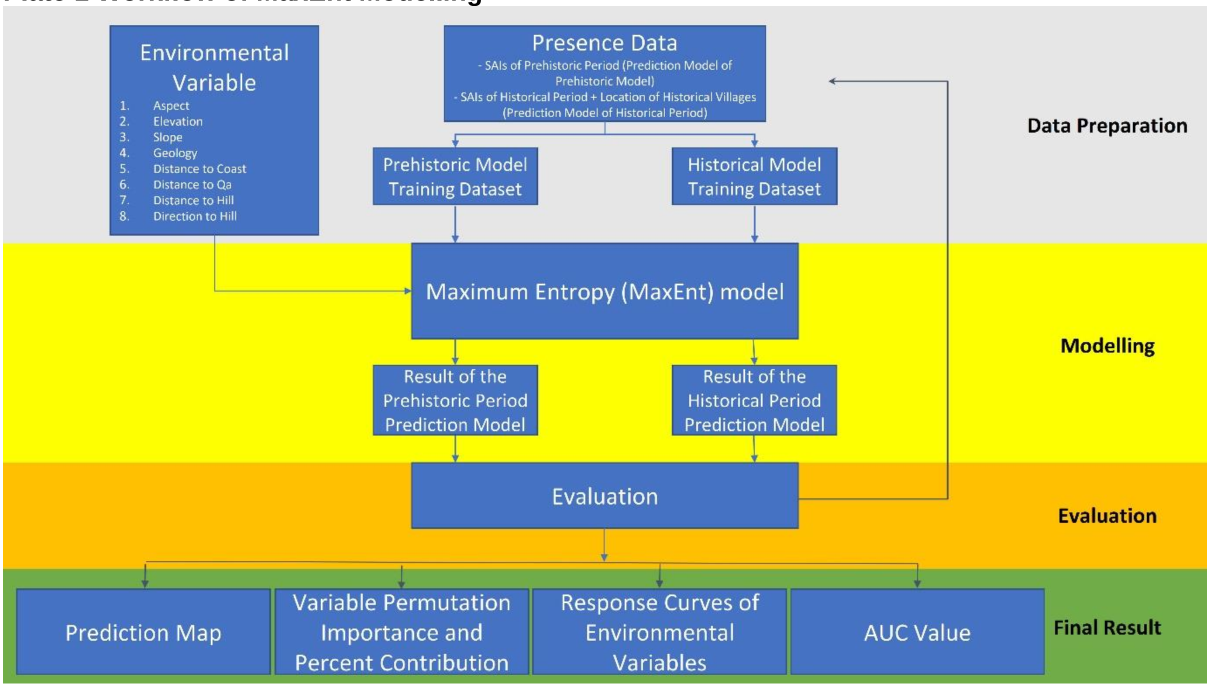
2.4.1 AUC value

2.4.1.1 A common approach in evaluating MaxEnt is by calculating the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) curve[14]. The AUC-Roc shows how capable the model is at distinguishing between positive locations (where archaeological potential areas are actually present, and negative locations (where there is no archaeological potential). The AUC values range from 0 to 1. While higher values indicating better model performance, a value at 0.5 suggest that the environmental variables have no correlations with the archaeological potential predictions by MaxEnt. The AUC values of the prediction models for historical and prehistoric period in the following sections are calculated by the by the MaxEnt model.

2.4.1.2 Evaluation was done for each run for the model and the parameters. Dataset has been modified for another run. The modelling stages are repeated for each run of the model until the AUC-ROC value is acceptable and above 0.7. A graph of the MaxEnt modelling workflow is presented below in **Plate 1**.

[14] Fourcade, Y. Engler, J.O., Rödder, D. and Secondi, J. 2014. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0097122.

Plate 1 Workflow of MaxEnt Modelling



3 MODEL RESULT

3.1 Introduction

3.1.1.1 This section displays the results generated by the historical and prehistoric prediction modelas by MaxEnt. Three crucial results for each environmental variable are generated, namely percent contribution, permutation importance and response curves of each environmental variables.

3.1.1.2 Percent contribution and permutation importance are two metrics the help to evaluate the relative importance of the environmental variables for the prediction of archaeological potential by MaxEnt model. Percent contribution measures the weight of particular variable contributed to the model results based on how MaxEnt algorithm utilise it. The calculation was made by the MaxEnt algorithm by first making predictions based on only one variable at a time. Then another variable is added to the algorithm to check if the predictions has improved and assigns a contribution score based on that improvement. The variables that improve the predictions the most get higher contribution scores.

3.1.1.3 On the other hand, permutation importance measures the relevancy of each variable in predicting the selection of past human settlement location. It works by randomly mixing up the values of one variable at a time, then seeing how much the prediction model of archaeological potential changes as a result. This change would trigger decrease to the AUC-ROC value. The more this change cause decrease in the AUC-ROC value, the more important is this variable to the prediction model.

3.1.1.4 Furthermore, response curves portray how the predicted archaeological potential varies across the range of values for respective environmental variable. The y-axis of the graph is the predicted probability, and the x-axis is the value of variables. Curves shapes reveals conditions of each variable that favours higher or lower potential per statistical patterns. It should be noted that these curves only consider single variable in isolation and overlooks the correlations between variables.

3.2 Prediction Model for Prehistoric Period

3.2.1 Variable Contributions and Importance

3.2.1.1 **Table 2** shows the results of the environmental variables used in the prediction model for the prehistoric period.

Table 2 Percentage Contribution and Permutation Importance of each Environmental Variable in Prehistoric Prediction Model

Variable	Percent Contribution	Permutation Importance
Aspect	3.0	2.1
Elevation	31.1	38.1
Slope	5.3	4.2
Geology	12.4	4.2
Distance to Coast	17.1	22.7
Distance to Alluvial Deposits	26.2	16.3
Distance to Hill	3.3	10.1
Direction to Hill	1.6	2.3
Total	100	100

3.2.1.2 As shown in **Table 2,** elevation has the highest percent contribution at 31.1%, indicating the algorithm utilized elevation heavily when generating the predictions. Distance to coast also has a relatively high percent contribution of 17.1%.

3.2.1.3 However, both elevation and distance to coast have even higher permutation importance values of 38.1% and 22.7%. This means that they contained the most useful information for distinguishing areas of higher archaeological potential.

3.2.1.4 In contrast, variables like aspect, slope and direction to hill have lower permutation importance despite their percent contributions. This indicates the MaxEnt algorithm overestimated their significance, when in reality they likely had less relevance for predicting where archaeological sites actually formed.

3.2.1.5 Overall, the variables with the highest permutation importance (elevation, distance to coast and distance to alluvial deposits) have the most meaningful insights into past human behaviours and site distributions based on environmental influences. These factors probably had the greatest impact on where archaeological sites actually occurred.

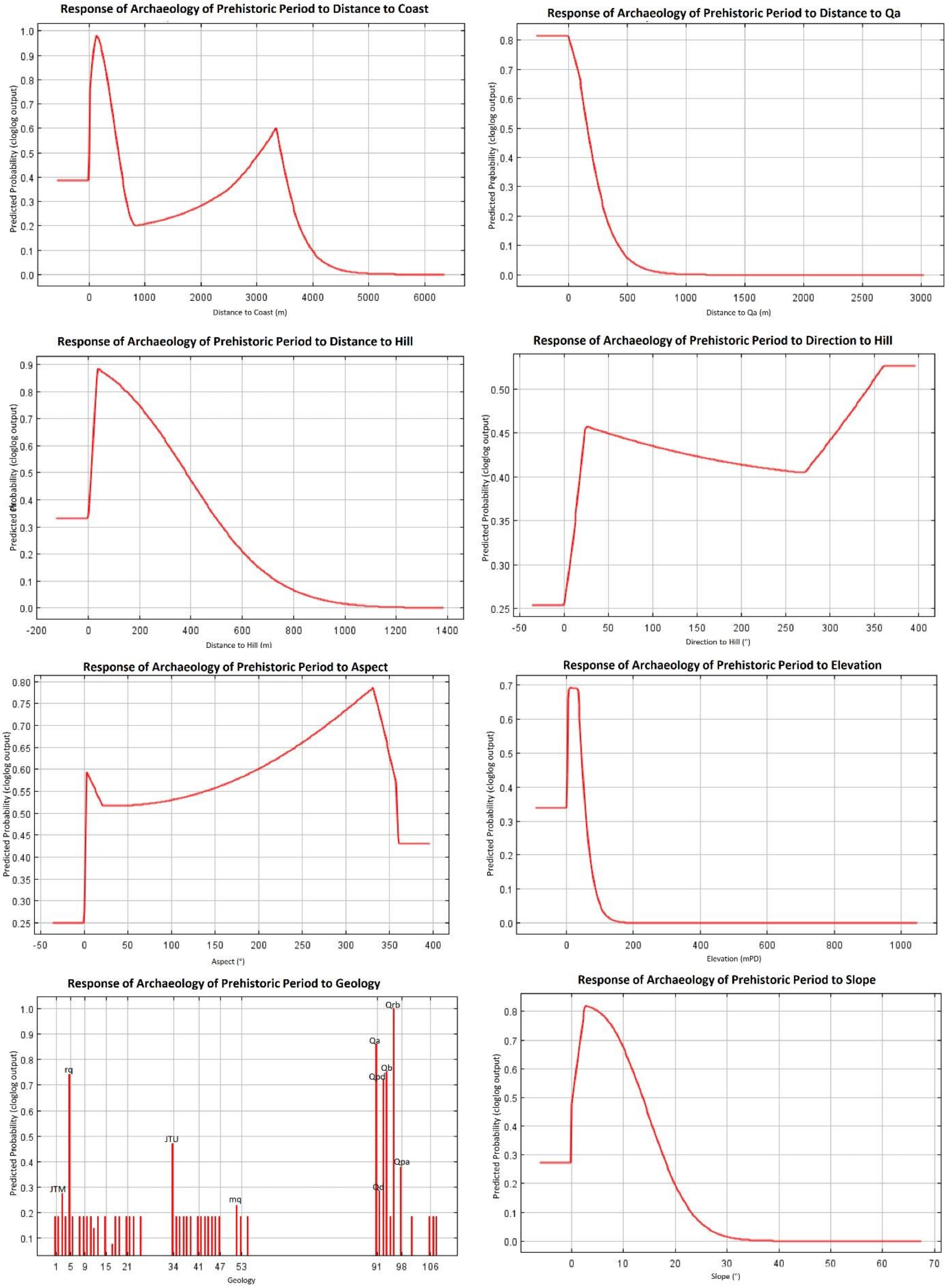
3.2.2 Response Curves of Variables

3.2.2.1 The response curves illustrate how environmental variables influence the MaxEnt model results. It reveals the variables that has the most impact on predicting archaeological potential. Curves that change sharply and differ across the entire range of a variable likely indicate that the variable significantly affected where ancient human sites were formed.

3.2.2.2 On the other hand, curves that remain fairly similar and give an even probability across all values for a variable tend to represent aspects with minimal relevance for determining which landscape features were more favourable for ancient human settlement and activities. These variables typically contribute very little to the prediction, correctly reflecting their weak predictive power considered individually.

3.2.2.3 In simple terms, the steeper and more variable the curve for an environmental factor, the more important that factor likely was in influencing where ancient humans chose to settle and the activities they undertook. However, factors with flatter, more uniform curves probably did not significantly impact ancient human site locations and behaviours. The response curves of environmental variables for the MaxEnt model in prehistoric period is displayed in Plate 2.

Plate 2 Response Curve of Archaeology of Prehistoric Period to Environmental Variables



3.2.2.4 The ***aspect*** response curve illustrates its influence on predicted archaeological potential of the prehistoric period. Aspect ranges from 0° to 360°, with 0° denoting north and 180° south.

- 0° to 90° (NE): Low potential (0.5-0.55 probability)

- 90° to 180° (SE): Lower potential (0.55-0.6 probability)
- 180° to 270° (SW): Higher potential (0.6-0.7 probability, rising)
- 270° to 360°/0°(NW): Highest potential (0.7-0.8 probability, peaking at 340° then declining)

3.2.2.5 The curve suggests NW-facing aspects had relatively greater potential. However, minor contribution (3%) of aspect means the model placed little weight on it.

3.2.2.6 The ***elevation*** response curve reveals its influence on predicted archaeological potential of the prehistoric period. Elevation ranged from 0mPD to +1000mPD, categorized as:

- 0mPD to +50mPD: Higher potential (higher probability, >0.35)
- +50mPD to +1000mPD: Lower potential (declining probability as elevation increases)

3.2.2.7 The curve suggests lower elevations had relatively greater potential. This aligns with the high contribution (31.1%) of elevation, indicating the model weighted it heavily.

3.2.2.8 The ***geology*** response curve reveals its influence on predicted archaeological potential. Geological types ranged from solid to superficial deposits. The graph depicts these types as bars, with longer bars indicating higher probabilities.

3.2.2.9 Geological types with relatively higher probabilities (>0.2) include Qa (alluvium), Qpa (Pleistocene alluvium), Qd (colluvium) Qpd (Pleistocene colluvium), Qb (beach deposits) and Qrb (raised deposits). Some solid geology also generated higher probabilities where presence data occurred within them.

3.2.2.10 This indicates superficial deposits, especially alluvial and colluvial, yielded relatively greater potential. This aligns with moderate contribution (12.4%) of geology, indicating the model moderately weighted it.

3.2.2.11 The ***slope*** response curve reveals its influence on predicted archaeological potential. Slope ranged from 0° to 90°, categorized as:

- 0° to 15°: Higher potential (higher probability >0.5)
- 15° to 90°: Lower potential (declining probability as slope increases)

3.2.2.12 The curve suggests gentler slopes had relatively greater potential. However, this contrasts with minor contribution (5.3%) of slope, indicating the model placed little weight on it.

3.2.2.13 The ***distance to coast*** response curve reveals its influence on predicted archaeological potential. Distance ranged continuously from 0m to over 3000m.

3.2.2.14 The curve has a bimodal shape, indicating two distance ranges with relatively higher probabilities:

- 0m to 500m: Higher potential (peak probability)
- 500m to 3000m: Potential generally decreases as distance to coast increases
- 3000m to 3500m: Moderate potential (secondary peak)

3.2.2.15 This aligns with the high contribution (17.1%) of distance to coast, indicating the model weighted it heavily. Locations close to the coast, especially within 500m, likely exhibited relatively greater potential.

3.2.2.16 The ***distance to alluvial deposits*** response curve reveals its influence on predicted archaeological potential. Distances ranged continuously from 0m to over 1000m. The curve indicates locations within 100m generally exhibited higher probabilities (>0.5), with potential decreasing as distance increases beyond 100m. Potential is highest (peak probability) at 0m, i.e. locations sitting directly on alluvial deposits. This aligns with distance to the high contribution (26.2%) of alluvial deposits, indicating the model weighted it heavily.

Proximity to alluvial deposits, especially within 100m, likely conferred relatively greater potential.

3.2.2.17 The ***distance to hill*** response curve reveals its influence on predicted archaeological potential. Distances ranged continuously from 0m to over 1000m.

3.2.2.18 The curve indicates locations within 500m generally exhibited higher probabilities (>0.5), with potential peaking around 25m and declining as distances approach 0m. Potential is lowest at 0m, i.e. locations sitting directly on hills. This contrasts with minor contribution (3.3%) of distance to hill, as the model placed little weight on it. Proximity to hills, but not direct contact, likely conferred relatively greater potential.

3.2.2.19 The ***direction to hill*** response curve reveals its influence on predicted archaeological potential. Direction to hill ranges from 0° to 360°, with 0° denoting north and 180° south.

3.2.2.20 The curve reveals a nearly flat, featureless response, with probability generally <0.5 across all directions.

3.2.2.21 The insignificant contribution (1.6%) of direction to hill indicates this variable poorly predicts archaeological potential. This suggests direction to hill provides little useful information for distinguishing locations likely to contain archaeological remains. The uniform probability across directions indicates this variable exhibited no meaningful influence over where sites actually formed in the past.

3.2.2.22 Curve shapes revealing little variation across an environmental factor often denote variables with minimal true significance for determining landscape settings that facilitated human settlement and activities in the past. Such variables tend to exhibit extremely low contributions that reflect their poor predictive abilities.

3.2.3 Summary

3.2.3.1 Response curve analysis suggests two general environmental settings exhibited relatively higher archaeological potential during the prehistoric period:

1. Coastal flats:
   - Lower elevations (0m to 50m)
   - Gentler slopes (0° to 10°)
   - Proximity to the coast (0m to 500m and 3000m to 3500m)
   - Proximity to alluvial deposits (0m to 250m from Qa)
2. Hill foots:
   - Foot slopes and gentle slopes (<15°)
   - Aspect facing Deep Bay in the northwest
   - Proximity to the hill (within 400m)

3.2.3.2 Overall, the coastal flats appeared to offer the most favourable conditions for past human settlement and activity. Lower-lying alluvial areas near the coast but set back from active beach zones may have balanced advantages of resources, arable land and transport routes with fewer constraints.

3.2.3.3 Hill foots likely conferred advantages due to prospects, aspect, drainage and river views, though they exhibited less potential than coastal flats. Top of hills showed the lowest potential, indicating humans preferred slopes over peaks during the prehistoric period. Furthermore, direction to hills exhibited no significant influence, suggesting relative orientation to hills mattered little compared to other factors like slope, elevation and coastality.

3.2.3.4 In summary, landscape setting appears to have strongly influenced archaeological potential during the prehistoric period, with coastal flats offering the most favourable conditions according to statistical patterns in the input data.

3.2.3.5 Presence data for locations with known archaeological potential during the prehistoric period exhibit two general patterns:

1. Coastal flats: Sites occurring on the coastal plains tend to cluster in Tuen Mun, including at Lau Fau Shan, Ngau Hom Shek and Mong Tseng. The environmental variable values for these locations - like low elevation, proximity to alluvial deposits and the coast - generally fall within the high probability ranges indicated by the response curves.
2. Hill foots: Sites situated on gently sloping hills tend to concentrate in Fanling and Sheung Shui, particularly around Wah Shan and Hung Leng. The environmental characteristics of these locations - such as slope, aspect and distance to hills - also typically align with the higher probability ranges revealed in the corresponding curves.

3.2.3.6 This correspondence between the environmental conditions at known archaeological locations and the high probability ranges predicted by the model strengthens confidence in the response curves' accuracy. They appear to have successfully identified the key landscape variables influencing site distributions and potential during the prehistoric period. **Table 3** supplement the values of each environmental variables of the areas.

Table 3 Environmental Variables Value of Presence Data Location of Prehistoric Period

Occurrence Area	Elevation (mPD)	Slope (°)	Distance to Coast (m)	Distance to Qa (m)	Geology
Ngau Hom Shek	6 – 86 (mean = 35)	3 - 16 (mean = 11)	0 – 426 (mean = 191)	42 – 400 (mean = 231)	gfm, Qpd, Qb, Qa
Mong Tseng	4 – 18 (mean = 9)	1 – 14 (mean = 5)	30 – 150 (mean = 95)	30 – 175 (mean = 101)	Cts, Qpa, Qpd
Wah Shan	7 – 39 (mean = 19)	2 – 21 (mean = 9)	2612 – 3064 (mean = 2854)	67 -375 (mean = 198)	JTM, Qpd, Qpa
Hung Leng	12 – 49 (mean = 26)	2 – 20 (mean = 10)	3141 – 3595 (mean = 3367)	0 – 458 (mean = 228)	JTM, Qpd

3.2.4 AUC Value

3.2.4.1 To review the confidence level of the model prediction, a train-test split was adopted when developing the model. Among all the presence data inputted into the model, 80% of the data is used to train and develop the model, while 20% of the data is used to test the model's accuracy in predicting the archaeological potential for unknown locations. MaxEnt will generate an AUC value on the testing data, which indicates the predictive power of the model (i.e. how fitting is the model result to the testing data)[15]. An AUC closer to 1 implies a better model performance.

3.2.4.2 The AUC values of the MaxEnt model for the prehistoric period is 0.905 for testing data. This suggests the MaxEnt model can predict the probability of high archaeological potential in prehistoric period with reasonable accuracy.

[15] Phillips, Steven J., AT&T Research. (2017). A Brief Tutorial on Maxent. Retrieved from https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf.

3.3 Prediction Model for Historical Period

3.3.1 Variable Contributions and Importance

3.3.1.1 **Table 4** displays the results of the prediction model for the historical period.

Table 4 Percentage Contribution and Permutation Importance of each Environmental Variable in Historical Prediction Model

Variable	Percent Contribution	Permutation Importance
Aspect	0.6	0.3
Elevation	38.8	14.8
Slope	1.4	13.1
Geology	16	8.4
Distance to Coast	34	54.7
Distance to Qa	7	6.9
Distance to Hill	0	0.7
Direction to Hill	2.2	1.2
Total	100	100

3.3.1.2 As shown in **Table 4,** based on the percent contribution values, the two most important variables for the model are elevation and distance to coast, together accounting for over 70% of the model contribution. Aspect has the lowest percent contribution at only 0.6%.

3.3.1.3 However, the permutation importance values show a somewhat different pattern. Distance to coast has the highest permutation importance at 54.7%, followed by elevation at 14.8% and slope at 13.1%. Aspect again has the lowest importance at 0.3%.

3.3.1.4 This suggests that while elevation and distance to coast increase the model gain the most when included, distance to coast has the greatest impact on model accuracy. This indicates distance to coast may be the most influential variable for actually predicting archaeological potential.

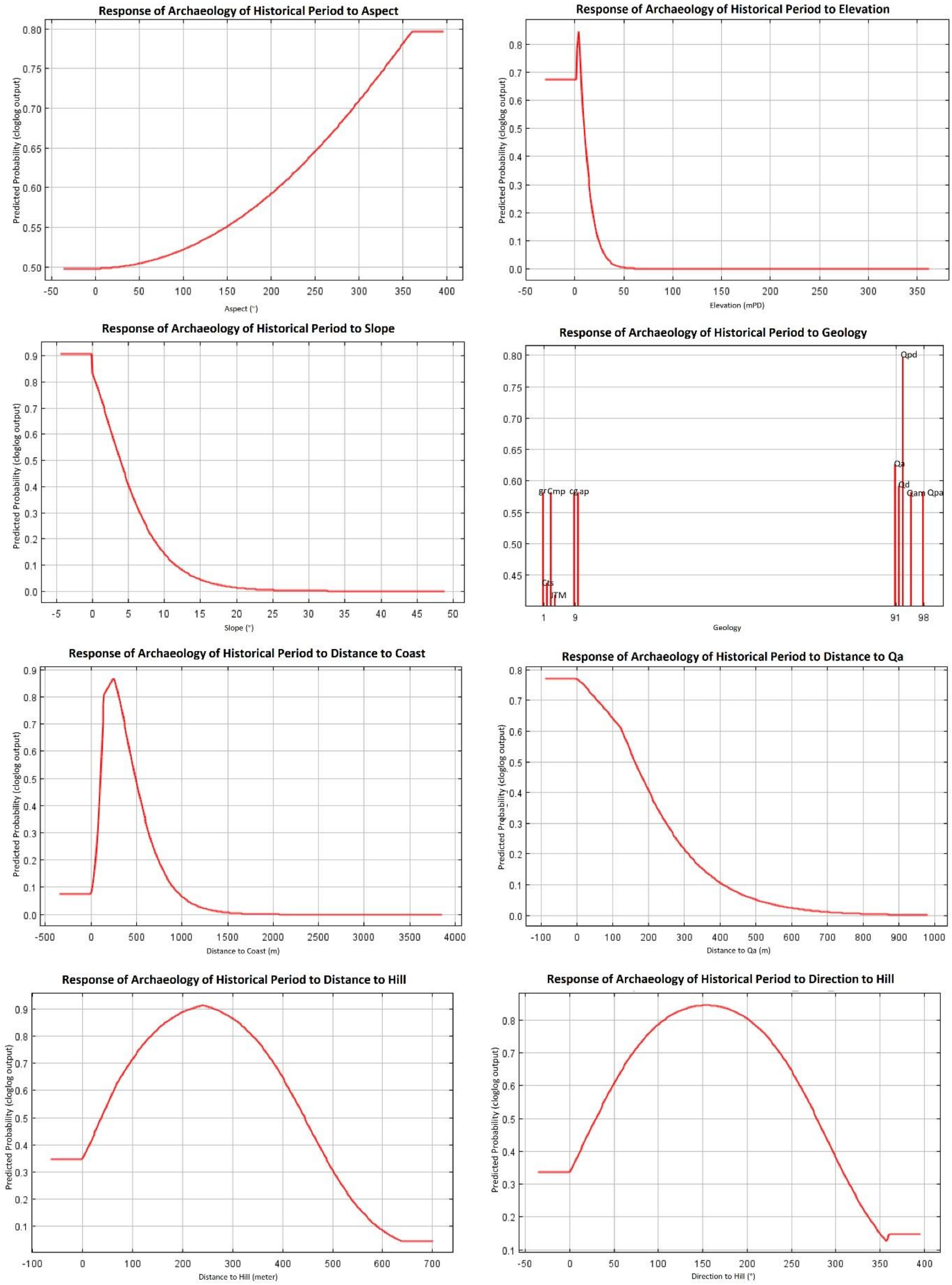
3.3.1.5 Slope and geology have higher permutation importance relative to their percent contribution, suggesting they have more predictive power in the model than implied by their contribution to the overall gain.

3.3.1.6 In summary, both sets of values point to elevation, distance to coast and slope as the most important variables. However, permutation importance indicates distance to coast may be the single strongest predictor of archaeological potential, though elevation contributes the most to model gain. Aspect appears to be the least influential variable in both analyses.

3.3.2 Response Curves of Variables

3.3.2.1 The response curves of environmental variables for the MaxEnt model in historical period is displayed in Plate 3.

Plate 3 Response Curve of Archaeology of Historical Period to Environmental Variables



3.3.2.2 The ***aspect*** response curve illustrates its influence on predicted archaeological potential of the historical period. Aspect ranges from 0° to 360°, with 0° denoting north and 180° south.

- 0° to 90° (NE): Lowest archaeological potential (0.5 probability)

- 90° to 180° (SE): Medium-low archaeological potential (0.5-0.55 probability)
- 180° to 270° (SW): Medium-high archaeological potential (0.55-0.65 probability)
- 270° to 360°/0°(NW): Highest archaeological potential (0.65-0.8 probability)

3.3.2.3 This response suggests that NW and SW-facing slopes were preferred locations for human settlement and activity during the historical period. However, minor contribution (0.6%) of aspect means the model placed little weight on it.

3.3.2.4 The ***elevation*** response curve reveals its influence on predicted archaeological potential of the prehistoric period. Elevation ranged from 0mPD to +350mPD, categorized as:

- 0mPD to +20mPD: Higher potential (higher probability, >0.35)
- +20mPD to +350mPD: Lower potential (probability declining sharply as elevation increases; reaching 0 at +60mPD)

3.3.2.5 The curve suggests lower elevations had relatively greater potential. This aligns with the high contribution (38.8%) of elevation, indicating the model weighted it heavily.

3.3.2.6 The ***slope*** response curve reveals its influence on predicted archaeological potential. Slope ranged from 0° to 90°, categorized as:

- 0° to 5°: Higher potential (higher probability >0.5)
- 5° to 90°: Lower potential (declining probability as slope increases)

3.3.2.7 The curve suggests gentler slopes had relatively greater potential. However, this contrasts with minor contribution (1.4%) of slope, indicating the model placed little weight on it.

3.3.2.8 The ***geology*** response curve reveals its influence on predicted archaeological potential. Geological types ranged from solid to superficial deposits. The graph depicts these types as bars, with longer bars indicating higher probabilities.

- The solid geology types (Cts, Jtm, Cmp) generally have lower probabilities below 0.6, particularly Cts and Jtm below 0.5.
- The superficial geology types such as colluvium (Qd), marine sand deposits (Qam), Pleistocene alluvium (Qpa) have probabilities around 0.57 to 0.59, slightly higher than the solid types.
- The alluvial deposit (Qa) has a probability of around 0.62, slightly higher still.
- The Pleistocene colluvium (Qpd) has the highest probability of around 0.8.

3.3.2.9 This suggests that areas on Pleistocene colluvium have the greatest archaeological potential. These deposits are formed by hill slope sediment accumulation, likely conducive to human settlement and activity. Solid geology types have the lowest potential, possibly due to their association will hilly regions. Alluvial deposit shows intermediate potential.

3.3.2.10 This indicates superficial deposits, especially alluvial and colluvial, yielded relatively greater potential. This aligns with moderate contribution (16%) of geology, indicating the model moderately weighted it.

3.3.2.11 The ***distance to coast*** response curve reveals its influence on predicted archaeological potential. Distance ranged continuously from 0m to over 3000m.

3.3.2.12 The curve has a bimodal shape, indicating two distance ranges with relatively higher probabilities:

- 0m to 500m: Higher potential (peak probability at around 250m)
- Beyond 500m: Potential generally decreases as distance to coast increases

3.3.2.13 This aligns with the high contribution (34%) of distance to coast, indicating the model weighted it heavily. Locations close to the coast, especially within 500m, likely exhibited relatively greater potential.

3.3.2.14 The ***distance to alluvial deposits*** response curve reveals its influence on predicted archaeological potential. Distances ranged continuously from 0m to over 1000m. The curve indicates locations within 150m generally exhibited higher probabilities (>0.5), with potential decreasing as distance increases beyond 150m. Potential is highest (peak probability) at 0m, i.e. locations sitting directly on alluvial deposits. However, it has a low percent contribution of (7%) of alluvial deposits, while its response curve still indicates it predicts archaeological potential to some extent. It may have limited predictive power individually but still improves the model when included.

3.3.2.15 The ***distance to hill*** response curve reveals its influence on predicted archaeological potential. Distances ranged continuously from 0m to over 1000m.

3.3.2.16 The curve indicates locations between 50m and 450m are generally exhibited higher probabilities (>0.5), with potential peaking around 250m and declining as distances approach 0m. Potential is lowest at 0m, i.e. locations sitting directly on hills. This variable has the lowest contribution of all variables at 0% based on percent contribution values. This suggests it provides no gain to the MaxEnt model when included.

3.3.2.17 The ***direction to hill*** response curve reveals its influence on predicted archaeological potential. Direction to hill ranges from 0° to 360°, with 0° denoting north and 180° south.

3.3.2.18 The parabolic shape of the response curve indicates a relationship between direction to hill and archaeological potential. The curve peaks at around 100° to 200°, suggesting the highest probability occurs when hills are located to the southeast of a site. The probability is around 0.8 at the peak, higher than for other directions.

3.3.2.19 The insignificant contribution (2.2%) of direction to hill indicates this variable poorly predicts archaeological potential. This suggests direction to hill provides little useful information for distinguishing locations likely to contain archaeological remains. While this trend is evident, the limited shape and range of the curve indicate direction to hill is not an especially useful predictor of archaeological potential, in line with its low percent contribution.

3.3.3 Summary

3.3.3.1 The response curves reveal various landscape features correlated with higher archaeological potential during the historical period. Coastal flatlands below 500 meters from the shoreline showed the highest potential. Proximity to coast and low elevation were key predictors according to the contribution values.

3.3.3.2 Areas within 150 meters of rivers also indicated higher potential, though distance to alluvial deposits had only a limited contribution to the model. Locations on alluvial and colluvial deposits like Qd, Qpd and Qpa similarly correlated with higher potential, suggesting river terraces and gentle hillslopes favouring human settlement.

3.3.3.3 Northwest facing aspects potentially due to their orientation toward the Sham Chun River and Deep Bay showed higher potential, though aspect had only a minor influence. Locations approximately 100 to 450m from hills, similar to historical villages situated in front of hills, also indicated higher potential. Potential increased when the hills were located to the south, as is the case for many San Tin villages facing north toward the river.

3.3.3.4 However, direction and distance to Hill provided negligible gain to the model. In summary, the response curves demonstrate that the landscape features most predictive of archaeological potential during the historical period included low-lying coastal land, locations near rivers and facing northwest, and areas in front of hills facing the river.

3.3.3.5 While some variables provided little overall explanatory power according to their percent contribution values, their response curves still revealed meaningful correlations with potential. This indicates the variables do not fully capture all factors influencing settlement patterns during this period.

3.3.3.6 Most archaeological sites from the historical period are located in coastal flatlands. Places like San Tin, Mai Po, Chau Tau and Lok Ma Chau. The environmental conditions at these sites generally match the landscape features that showed the highest potential in the model. Features like:

- Being within 500 meters of the coast
- Having elevations under 20 meters
- Having slopes under 5 degrees
- Proximity to rivers and alluvial or colluvial deposits

3.3.3.7 This alignment between the locations of known archaeological sites and the landscape features predictive of potential in the model helps validate the results. It shows that the model is successfully identifying landscape settings conducive to human settlement during this period.

3.3.3.8 The fact that most historical settlements are found in these low-lying coastal and alluvial landscapes suggests these environmental conditions were favourable for human occupation. They likely provided access to resources, arable land, trade routes and other necessities for villages and farmsteads.

3.3.3.9 **Table 5** supplement the values of each environmental variables of the areas.

Table 5 Environmental Variables Value of Presence Data Location of Historical Period

Areas	Elevation (mPD)	Slope (degree)	Distance to Coast (meter)	Distance to Qa (meter)	Geology
Mai Po Area	4 – 32 (mean = 12)	0 – 17 (mean = 4)	90 – 408 (mean = 268)	0 – 192 (mean = 82)	JTM, Qd, Qam, cmp, Qa
San Tin Area	2 – 10 (mean = 5)	0 – 4 (mean = 1)	90 – 446 (mean = 286)	0 – 192 (mean = 62)	JTM, Qpd, Qa, Qam
Chau Tau Area	4 – 8 (mean = 5)	0 – 5 (mean = 2)	551 – 713 (623)	42 – 175 (mean = 110)	Qpa, Qpd
Lok Ma Chau Area	4 – 9 (mean = 5)	1 – 13 (mean = 5)	67 – 127 (mean = 100)	300 – 384 (mean = 343)	Qd

3.3.4 AUC Value

3.3.4.1 To review the confidence level of the model prediction, a train-test split was adopted when developing the model. Among all the presence data inputted into the model, 80% of the data is used to train and develop the model, while 20% of the data is used to test the model's accuracy in predicting the archaeological potential for unknown locations. MaxEnt will generate an AUC value on the testing data, which indicates the predictive power of the model (i.e. how fitting is the model result to the testing data)[16]. An AUC closer to 1 implies a better model performance.

3.3.4.2 The AUC values of the MaxEnt model in historical period is 0.945 for testing data. This suggests the model has excellent discriminatory power in predicting the probability of archaeological sites from the historical period.

[16] Phillips, Steven J., AT&T Research. (2017). A Brief Tutorial on Maxent. Retrieved from https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf.

4 ARCHAEOLOGICAL POTENTIAL

4.1 Introduction

4.1.1.1 The prediction model can provide useful initial indication of archaeological potential, but it should be used cautiously and cannot be taken at face value due to inherent limitations. While the environmentally deterministic[17] approach has been criticised for minimising cultural factors, it also has value in identifying potential environmental influences on past human decisions. Biases in the data and methodology can affect the accuracy of the results. Hence, archaeological investigation under this Project and previous archaeological studies are essential to complement, validate and refine the output of the predictive modelling, establishing archaeological potential areas with high confidence.

4.1.1.2 Previous archaeological findings from prehistoric period have uncovered evidence of past human occupation in Hong Kong. Archaeological sites from the prehistory have been discovered on the relatively flat area along coastal beaches in the western Hong Kong near the Zhujiang River Delta[18]. Prehistoric archaeological remains have been found in locations such as Lau Fau Shan, Ngau Hom Shek and Mong Tseng. The geographical condition in these areas would have made them attractive places for human settlement during the prehistoric period.

4.1.1.3 **Table 6** presents the known archaeological periods previously identified in Hong Kong, with sensitive landscape to each period in general.

Table 6 Known Archaeological Periods in Hong Kong

Period	Years	Representative Sites	Landforms with Archaeological Potential
Middle to Late Neolithic	7000BC – 3500BC	Lung Kwu Chau, Tai Wan (Lamma Island), Yung Long	Coastal plains, low hills, river terraces near estuaries
Bronze Age	3500BC – 2500BC	Tai Wan (Lamma Island), Lo So Shing, Man Kwok Tsui	
Song Dynasty	AD960 – AD1279	Mong Tseng Wai, Tai Hom Tsuen	Plains, valleys, river terraces
Ming and Qing Dynasty	AD1368 – AD1912	Chok Ko Wan, Kowloon Walled City	

4.2 Archaeological Potential Areas Predicted by the MaxEnt Model

4.2.1.1 The prediction model for prehistoric period has identified 8 areas to have archaeological potential in prehistoric period, namely Mai Po Area, Hop Shing Wai Area, San Tin Area, Mai Po Lung Area, Siu Hum Tsuen Area, Pang Loon Tei Area, Hang Tau Area and Ngau Tam Mei (North) Area. They are generally located in coastal flats and hill foots. Their locations are presented in **Figure 12.60**.

4.2.1.2 The prediction model for historical period has identified 4 areas to have archaeological potential in historical period, namely Mai Po Area, Hop Shing Wai Area, San Tin Area and Shek Wu Wai Area. They are generally located in coastal flats. Their locations are presented in **Figure 12.61**.

[17] Conolly, J. 2008. Geographical Information Systems and Landscape Archaeology. In David, B. and Thomas, J. Eds. *Handbook of Landscape Archaeology*. 583–595.
[18] 商志*譚* 、吳偉鴻 (2010) 《香港考古學叙研》。北京：文物出版社。

Table 7 Summary of Areas of Archaeological Potential Predicted by MaxEnt Models

	Model for Prehistoric Period (Figure 12.60)	Model for Historical Period (Figure 12.61)
1	Mai Po Area	Mai Po Area
2	Hop Shing Wai Area	Hop Shing Wai Area
3	San Tin Area	San Tin Area
4	Mai Po Lung Area	Shek Wu Wai
5	Siu Hum Tsuen Area	
6	Pang Loon Tei Area	
7	Hang Tau Area	
8	Ngau Tam Mei (North) Area	

Mai Po Area (Identified in Prehistoric Model and Historical Model)

4.2.1.3 The area has a coastal landscape that would have favoured prehistoric and historical human settlements (see **Figure 12.60** and **Figure 12.61** refer). The land sits at an elevation approximately between +4mPD and +32mPD (mean = +12mPD) with a slope generally lower than 10°. The area is close to the coast, between 85m and 633m approximately. It faces west towards the Sham Chun River and Deep Bay. Such coastal setting have been known to have supported historical settlement such as the village cluster at San Tin and the Mai Po Lo Wai, as well as prehistoric sites on the coast of Deep Bay in Tuen Mun. The area is also near to the Mai Po SAI, which has known archaeological potential.

Hop Shing Wai Area (Identified in Prehistoric Model and Historical Model)

4.2.1.4 Hop Shing Wai Area is similar to Mai Po Area in having a coastal landscape that would have supported prehistoric and historical human settlements (see **Figure 12.60** and **Figure 12.61** refer). The land sits at an elevation approximately between +4mPD and +43mPD (mean = +9mPD) with slopes generally lower than 10°. The area is close to the coast, between 30m and 295m approximately. It sits in front of a small hill facing north towards the Sham Chun River and Deep Bay. Such coastal setting has known historical settlement such as the village cluster at San Tin, and the Mai Po Lo Wai, as well as known prehistoric coastal setting in the southwestern coast of Deep Bay in Tuen Mun. In the light of this, this piece of land is predicted to contain archaeological potential in both historical and prehistoric period.

4.2.1.5 The original landscape of this area was on the coast before modern development kicked in since the 20th century. However, the degree of the disturbance to archaeological potential in this area remains uncertain by the existing land use being temporary storage (which may have involved some site formation works). Thus, the area is assumed to have archaeological potential dating from prehistoric and historical period.

San Tin Area (Identified in Prehistoric Model and Historical Model)

4.2.1.6 The area has known historical potential as the villages in San Tin could be dated to the early Ming dynasty (see **Figure 12.60** and **Figure 12.61** refer). The San Tin Area is a flat coastal plan with an elevation between +2mPD and +10mPD and slopes of 0° to 4° approximately. The historical villages of San Tin is situated in front of a small hill facing northwest. They are close to the coast, between 90m and 446m (mean = 286m) approximately. These landscape characteristics (a low-lying, gently sloping coastal plain close to a source of fresh water) are features that would have favoured human settlements. The presence of Ming dynasty villages indicates the area has high potential to contain archaeological remains dating from this period.

Mai Po Lung Area (Identified in Prehistoric Model)

4.2.1.7 The area is located on a low-lying coastal plain with characteristics that would have favoured human settlements (see **Figure 12.60** refers). The elevation ranges from +4mPD to +31mPD (mean = +11mPD) and slopes range from 2° to 20° (mean = 7°) approximately. It is located close to the coast, between 228m and 560m (mean = 420m) approximately. The model results suggested that these landscapes have a higher probability of containing archaeological site from prehistoric period.

4.2.1.8 However, the archaeological potential had likely been impacted by modern development. The construction of San Tin Highway was commenced in the late 1980s. Large scale site formation and slope-cutting works were performed[20], any archaeological deposits (if any) would have already been destroyed. Nevertheless, the remaining areas might still contain archaeological potential in prehistoric period.

Siu Hum Tsuen Area and Pang Loon Tei Area (Identified in Prehistoric Model)

4.2.1.9 The areas have a river terrace landscape on gentle hill slopes that would have favoured settlement during the prehistoric period (see **Figure 12.60** refers). Similar landscape can be found in SAI of prehistoric period, such as Po Leng SAI and Sheung Shui Wah Shan SAI, which are also located on river terrace at the foot of hills. Both areas are located on a higher elevation between +2mPD and +48mPD (mean = +25mPD) and on steeper slopes between 1° and 18° (mean = 8°) approximately since they are situated on hill slopes. They are also further away from the coast, over 1000m to 2000m away. Meanwhile, the geology of the area is mainly alluvial and colluvial deposits (Qa, Qpa and Qpd), features typical of river terrace. These areas have landscape setting similar to the Po Leng SAI to the east (outside the assessment area), where prehistoric findings have been found. In the light of this, these three areas would have archaeological potential in the prehistoric period.

Hang Tau Area (Identified in Prehistoric Model)

4.2.1.10 Hang Tau area is predicted to have archaeological potential from prehistoric period by the MaxEnt Model (see **Figure 12.60** refers). It is located on the eastern side of Saddle Pass between Ki Lun Shan and Ngau Tam Shan. The area is located on a hillslope within a river valley. Within the 500m Assessment Area, its geology mainly comprises of alluvial and colluvial deposits (Qa, Qpa and Qpd), which are features typical of river terrace. It is located on a higher elevation between +14mPD and +40mPD approximately and on steeper slopes between 1° and 16° approximately. It is noted that the extent of the Hang Tau Area might extend beyond the 500m Assessment Area.

Ngau Tam Mei (North) Area (Identified in Prehistoric Model)

4.2.1.11 Ngau Tam Mei (North) area is predicted to have archaeological potential from prehistoric period by the MaxEnt Model (see **Figure 12.60** refers). The area is located on a southern hillslope of Ngau Tam Shan within the Ngau Tam Mei river valley. Within the 500m Assessment Area, its geology mainly comprises of colluvial deposits (Qpd) with some alluvial deposits (Qa and Qpa) in the south. It is located on a higher elevation between +14mPD and +28mPD approximately and on steeper slopes between 2° and 15°

[20] Lands Department. (1988). *1:8000, 4000 ft., A15545 [aerial photo]*. Lands Department. Retrieved from https://www.hkmapservice.gov.hk/OneStopSystem/map-search.

approximately. It is noted that the extent of the Ngau Tam Mei (North) Area might extend beyond the 500m Assessment Area.

Shek Wu Wai Area (Identified in Historical Model)

4.2.1.12 Shek Wu Wai Area is located to the north of the existing Shek Wu Wai village next to San Tin Highway (see **Figure 12.61** refers). It is located on river terraces that would have favoured human settlement in the historical period. The elevation ranges from +4mPD to +8mPD approximately and slopes range from 1° to 9° approximately. It sits on the colluvial deposits and is in close proximity to alluvial deposits (within 40 meters). The model results suggested that this area have a higher probability of containing archaeological site from historical period.

4.2.1.13 To the south of the Shek Wu Wai Area identified in the MaxEnt model, although the model does not predict it to have higher probability for having archaeological deposits from historical period, field scanning area A10 contains surface finds with blue and white porcelain sherds. Although the artefacts were identified along with many other modern waste deposits on the ground surface, the nature of this waste deposit was possible from the nearby Shek Wu Wai when older buildings were demolished for the newer village houses. The archaeological potential is likely to be secondary deposit from as refuse by the past humans in Shek Wu Wai.

5 BILIBOGRAPHY

Antiquities and Monuments Office. *List of Sites of Archaeological Interest in Hong Kong (as at Nov 2012)*. Retrieve from https://www.amo.gov.hk/filemanager/amo/common/form/list_archaeolog_site_eng.pdf.

Bickler, S. (2021). Machine Learning Arrives in Archaeology. *Advances in Archaeological Practice*. 9(2). 186-191. https://doi.org/10.1017/aap.2021.6.

Civil Engineering and Development Department. (2013). *Planning and Engineering Study on Development of Lok Ma Chau Loop – Investigation - Environmental Impact Assessment. 10 Cultural Heritage Impact.* https://www.epd.gov.hk/eia/register/report/eiareport/eia_2122013/PDF/EIA/S10%20-%20CH%20v16.pdf.

Civil Engineering and Development Department. *Geological Maps in 1:20,000 & 1:100,000*. Retrieved from https://www.geomap.cedd.gov.hk/GEOOpenData/eng/GeologicalMap.aspx.

Conolly, J. 2008. Geographical Information Systems and Landscape Archaeology. In David, B. and Thomas, J. Eds. *Handbook of Landscape Archaeology*. 583–595.

Fourcade, Y. Engler, J.O., Rödder, D. and Secondi, J. 2014. Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0097122.

Fyfe, J. A et al. (2000). The Quaternary Geology of Hong Kong. Hong Kong: Civil Engineering Department.

Kvamme, K.L. (2006). There and Back Again: Revisiting Archaeological Locational Modeling. In Mehrer, M.W. and Wescot, K.L. Eds. *GIS and Archaeological Site Location Modeling*. United Kingdom: Routledge. 3-38.

Luthfi, A. M., Sigit, H. M. and Bowo, S. (2019). MaxEnt (Maximum Entropy) model for predicting prehistoric cave sites in Karst area of Gunung Sewu, Gunung Kidul, Yogyakarta. *Proceedings of the SPIE*. 113110B. https://doi.org/10.1117/12.2543522

Morgan, B. and Guénard, B. 2019. New 30m Resolution Hong Kong Climate, Vegetation and Topography Rasters Indicate Greater Spatial Variation Than Global Grids within Urban Mosaic. *Earth System Science Data*. 11(3). 1083-1098. https://doi.org/10.5194/essd-11-1083-2019.

Philips, S.J., Dudík, M. and Schapire, R.E. (2023). Maxent Software for Modeling species Niches and Distribution (Version 3.4.1). Available from url: https://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2023-4-28.

Phillips, Steven J., AT&T Research. (2017). A Brief Tutorial on Maxent. Retrieved from https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf.

Rafuse, D.J., 2021. A Maxent Predictive Model for Hunter-Gatherer Sites in the Southern Pampas, Argentina. *Open Quaternary*, 7(1), p.6. https://doi.org/10.5334/oq.97.

Refrew, C. & Bahn, P. (2016). *Archaeology: Theories, Methods and Practice (6th ed.)*. United Kingdom: Thames & Hudson.

Wachtel, Ido & Zidon, Royi & Garti, Shimon & Shelach-Lavi, Gideon. (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. 22-36. https://doi.org/10.1016/j.jas.2018.02.001.

Yaworsky, P.M. et al. (2020). Advancing Predictive Modeling in Archaeology: An Evaluation of Regression and Machine Learning Methods on the Grand Stair-case Learning Methods on the Grand Staircase-Escalante National Monument. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0239424.

商志譚 、吳偉鴻 (2010) 《香港考古學叙研》。北京：文物出版社。