**Supplemental Information**

Table S1 – Additional information on data layers used as input to random forests analysis.

| Data Type | Units | Original Resolution | Source |
|---|---|---|---|
| Slope percent | percentage | 30m | http://www.opentopography.org/ |
| Aspect | orientation | 30m | http://www.opentopography.org/ |
| Solar radiation | watt hours per $m^2$ | 30m | http://www.opentopography.org/ |
| Distance to runoff | meters | 30m | http://www.opentopography.org/ |
| Soil organic carbon, 0 cm | 5 g / kg | 250m | https://opengeohub.org/about-openlandmap/ |
| Soil moisture, 10cm | 33kPa field capacity | 250m | https://opengeohub.org/about-openlandmap/ |
| Soil pH, 10 cm | pH in $H_2O$ | 250m | https://opengeohub.org/about-openlandmap/ |
| Soil class | USDA classes | 250m | https://opengeohub.org/about-openlandmap/ |
| Soil nitrogen, 0-5 cm | cg/kg | 250m | www.isric.org/explore/soilgrids |

Using relict farming plots from 200-300 years ago to model farming conditions of 2000 or more years ago requires some discussion. Two issues to address are the nature of farming carried out in the Egiin Gol valley that left our relict field evidence and the dynamism of the forest-steppe ecotone as related to the soil variables selected for the classification analysis. In the first case, techniques for dry farming with runoff irrigation systems and soil preparation technologies were likely quite conservative over centuries of farming practice. One of the points raised by reviewers is the nature of plowing technologies available during these different periods since advances in plow design could have made available farming areas previously unusable. Cast iron plowshares discovered at the Xiongnu period settlement of Ivolga (Davydova 1995) are similar to those used by Han Dynasty farmers to the south and by contemporaneous farming communities in the region of Manchuria. These plowing technologies comprise horse or oxen drawn wooden plows capped with cast iron shares in a configuration that Molnár (1982) refers to as the rectangular plow. Based on historical and linguistic evidence, Molnár traces the use of this plow technology in Mongolia up to the period in question and even as late as the 1950s in some regions (Molnár 1982). Given this research, the toolset available for working the soil at Egiin Gol 200 years ago was probably not dramatically different from that in use during the Xiongnu period.

The relict fields at Egiin Gol may be a good proxy for pre-modern farming; however, the geographical and pedological variables chosen to characterize these farm plots also require consideration. The Egiin Gol valley is situated within the ecotone of the forest-steppe of northern Mongolia which is sensitive to climatic modulation and, therefore, presents a dynamic landscape. While the

geographic variables of slope, solar exposure, aspect, and hydrological network have a degree of longevity, soil variables are subject to greater impacts from climatic shifts within the valley. For an informative overview of current efforts and best practices related to soil process modeling see Vereecken and colleagues (2016). We are encouraged to employ soil variables in a proxy model based on the long-term climatic baseline established by Barna's (2023) lake core sediment analysis from Lake Khargal, 75 km from the Egiin Gol valley. According to the palynological reconstruction, maximum humidity occurs in the Lake Khargal drainage at approximately 2000 cal yr B.P., followed by gradually increasing aridity and temperatures. These changes are subtle, and there is no dramatic shift in vegetation communities over the period observed. Furthermore, average climatic and environmental regimes during the Xiongnu period are not radically different from the conditions known from Egiin Gol over the past few decades, most notably during the 1990s when our survey project was carried out.

Soil formation or pedogenesis is a complex process involving local geology, climate, relief, hydrology, landforms, biotic process, and human activities. Although some degree of climatic consistency is apparent from limnological, geomorphological, and archaeobotanical analyses (Barna 2023; Harrison 2000; Jackson et al. 2002; Trigg 2003; Prouse 2004), we are under no illusions about the significant differences between contemporary soils and those of 2000 B.P. It is important to re-emphasize, however, that our classification is merely a 'best available' estimate of areas with potential for pre-modern styles of cultivation. This analysis appeals to the logic of recent re-conceptions of pedogenesis in which the standard conception of 'soil as mirror of the landscape' has been replaced by an understanding of 'soil as landscape memory' (Richter and Yaalon 2011). In other words, our model assumes that the best approximation available for understanding soils of the past, will be the soils of today.

Barna, J. An 11,600-Year Reconstruction of Vegetation Communities, Moisture Availability, and Land Use Changes at Lake Khargal, Northern Mongolia. Master's Thesis, Univiversity of Alaska, Department of Geology: Fairbanks, AK, USA, 2023.

Davydova, A.V. *Ivolginskii Arkheologicheskii Kompleks: Ivolginskoe Gorodishche*; AziatIKA: Saint Petersburg, Russia, 1995.

Harrison, B. Geomorphology of the Egiin Gol Valley. Unpublished Geological Field Report; Earth and Environmental Sciences Department, New Mexico Tech: Socorro, NM, USA, 2000.

Jackson, M.; Harrison, B.; Honeychurch, W. Geomorphic controls on the distribution of archaeological sites at Egiin Gol, northern Mongolia. In Proceedings of the 2002 Annual Meeting of the Geological Society of America, Denver, CO, USA, 13–20 April 2002.

Molnár, Á. The plough and ploughing among the Altaic peoples. *Central Asiatic J.* 1982, 26, 3/4, 215-224.

Prouse, T. Phytoliths and Their Meaning for Steppe Archaeology: Ecological Adaptations of Nomadic Pastoralists of Central Asia Based on the Phytolith Analysis from Two Central Asian Regions. Master's Dissertation, University College London, Department of Archaeology, London, UK, 2004.

Richter, D.; Yaalon, D. "The changing model of soil" revisited. *Soil Sci. Soc. Am. J.* 2011, 76, 766–778. doi:10.2136/sssaj2011.0407.

Trigg, H. Report on Flotation and Wood Samples from Mongolia. Unpublished Botanical Analysis Report; Department of Anthropology, University of Massachusetts: Boston, MA, USA, 2003.

Vereecken, H.; Schnepf, A.; Hopmans, J.; Javaux, M.; Or, D.; Roose, T.; Vanderborght, J.; Young, M.; Amelung, W.; itkenhead, M.; Allison, S.; Assouline, S.; Baveye, P.; Berli, M.; Brüggemann, N.; Finke, P.; Flury, M.; Gaiser, T.; Govers, G.; Ghezzehei, T.; Hallett, H.; Hendricks Franssen, H.; Heppell, J.; Horn, R.; Huisman, J.; Jacques, D.; Jonard, F.; Kollet, S.; Lafolie,; Lamorski, K.; Leitner, D.; McBratney, A.; Minasny, B.; Montzka, C.; Nowak, W.; Pachepsky, Y.; Padarian, J.; Romano, N.; Roth, K.; Rothfuss, Y.; Rowe, E.; Schwen, A.; Šimůnek, J.; Tiktak, A.; Van Dam, J.; van der Zee, S.; Vogel, H.; Vrugt, J.; Wöhling, T.; Young, I. Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone J.* 2016, 15, 5, 2-57. doi:10.2136/vzj2015.09.0131.

Table S2 – Archaeological sites used in analysis.

The Egiin Gol habitation sites are primarily dated based on diagnostic ceramics and two decades of subsequent survey in other parts of Mongolia has contributed greatly to our knowledge of pottery and periodization. For that reason, the classification of sites by period has changed slightly since their original discovery and the list presented below differs from those initially presented in the dissertations of the three co-directors (Honeychurch 2004; Amartuvshin 2004; Wright 2006). In addition, radiocarbon dating of the more questionable Egiin Gol habitation sites is ongoing. The sites selected for this study are those upon which all co-directors agree and which have 25 sherds or more and, in the case of Xiongnu sites, are 0.25 ha in extent or larger. These selection criteria exclude several sites that are likely pot-drops or are very small or have assemblages with very few diagnostics. EGS 278 is included here as a Xiongnu habitation despite having a C14 date that is slightly later than the classic Xiongnu period. This decision was based on the presence of diagnostic Xiongnu sherds and some doubt about the quality of the original radiocarbon analysis which is being redated. Likewise in a few cases where diagnostics ceramics are entirely clear as to period, sites were included despite having comparatively small size.

Bronze and Early Iron Age (BEIA) habitation sites included for this analysis are given below. Three sites listed in Wright et al. 2023 as BEIA have been left out because they are small, have few diagnostics or are under review for periodization. (EGS 28, 34, 454).

One site (EGS 194) has been added because it very likely has a Late Bronze Age phase although it also has an earlier occupation during the Eneolithic period as explained by Wright and colleagues (2023).

BEIA habitations included for this analysis:

| **Site** | **Size (square meters)** |
|---|---|
| EGS 14 | 4900 |
| EGS 15 | overlap w/ EGS 14 |
| EGS 132 | 3600 |
| EGS 146 | 1500 |
| EGS 153 | 200 |
| EGS 194 | 4000 |
| EGS 254 | 400 |
| EGS 279 | 625 |
| EGS 280 | 400 |
| EGS 296 | 1050 |
| EGS 332 | 56 |

Xiongnu period habitations included for this analysis:

| **Site** | **Size (square meters)** |
|---|---|
| EGS 14/17 | 7650 |
| EGS 36/34 | 40600 |
| EGS 110 | 11000 |
| EGS 131 | 3000 |
| EGS 278 | 3600 |
| EGS 301 | 26000 |
| EGS 455 | 26730 |
| EGS 463 | 2500 |
| EGS 486 | 7800 |
| EGS 247 | 2800 |
| EGS 297 | 27600 combined |
| EGS 298 | 4550 combined |

| | |
|---|---|
| EGS 299 | 10000 combined |

Table S3 – Specifications and diagnostics for the random forests model.

**Model Characteristics**

| | |
|---|---|
| Number of Trees | 275 |
| Leaf Size | 1 |
| Tree Depth Range | 57-105 |
| Mean Tree Depth | 81 |
| % of Training Available per Tree | 100 |
| Number of Randomly Sampled Variables | 3 |
| % of Training Data Excluded for Validation | 20 |

**Model Out of Bag Errors**

| | |
|---|---|
| Number of Trees | 275 |
| MSE | 15.454 |
| 0 | 9.649 |
| 1 | 32.019 |

**Top Variable Importance**

| Variable | | Importance | Percent |
|---|---|---|---|
| pH_10CM | (soil pH) | 5.73 | 14 |
| W_10CM | (soil moisture) | 5.54 | 14 |
| C_0CM | (soil organic carbon | 5.14 | 13 |
| ASPECT | (aspect) | 4.81 | 12 |
| SOLAR | (solar radiation) | 4.60 | 12 |
| SLOPE | (percent slope) | 4.33 | 11 |
| RUNOFF | (distance to runoff) | 4.20 | 11 |
| N_0-5CM | (soil nitrogen) | 4.09 | 10 |
| CLASSES | (soil class) | 1.34 | 3 |

## Training Data: Classification Diagnostics

| Category | F1-Score | MCC | Sensitivity | Accuracy |
|---|---|---|---|---|
| 0 | 0.99 | 0.96 | 1.00 | 0.99 |
| 1 | 0.97 | 0.96 | 0.95 | 0.99 |

Predictions for the data used to train the model compared to the observed categories for those features.

## Validation Data: Classification Diagnostics

| Category | F1-Score | MCC | Sensitivity | Accuracy |
|---|---|---|---|---|
| 0 | 0.89 | 0.62 | 0.91 | 0.84 |
| 1 | 0.72 | 0.62 | 0.70 | 0.84 |

Predictions for the test data (excluded from model training) compared to the observed values for those test features.