



Can soil information add value for the archaeological predictive model?

Mathias Bellat^{1,2}, Benjamin Glissmann^{1,3}, Ruhollah Taghizadeh-Mehrjardi^{1,2,4},
Tobias Rentschler^{1,2,4,5}, Paola Sconzo⁶, Peter Pfälzner^{1,3}, Thomas Scholten^{1,2,4}

¹ SFB 1070 ResourcesCultures, ² Department of Geosciences, Chair of Soil Science and Geomorphology, ³ Institute for Ancient Near Eastern Studies (IANES)
⁴ Cluster of Excellence Machine Learning: New Perspectives for Science, ⁵ Digital Humanities Center, ⁶ Department of Culture and Society, University of Palermo

International Congress on the Archaeology of the Ancient Near East Copenhagen, 22-26 May 2023
Session: 1. Near Eastern Archaeology in the field and in the lab, b. Methods and theory

RESOURCE CULTURES

Resource
Complexe

- Network of knowledges, technics, practices, peoples, places and objects/medias

Resource
Assemblage

- State and process at the same time, movements
- Historical contingence of heterogenous practices

Resource
Cultures

- Grasp the overlaps, recurrent phenomena and inherent dependencies
- Diversity of resource cultures in space and time

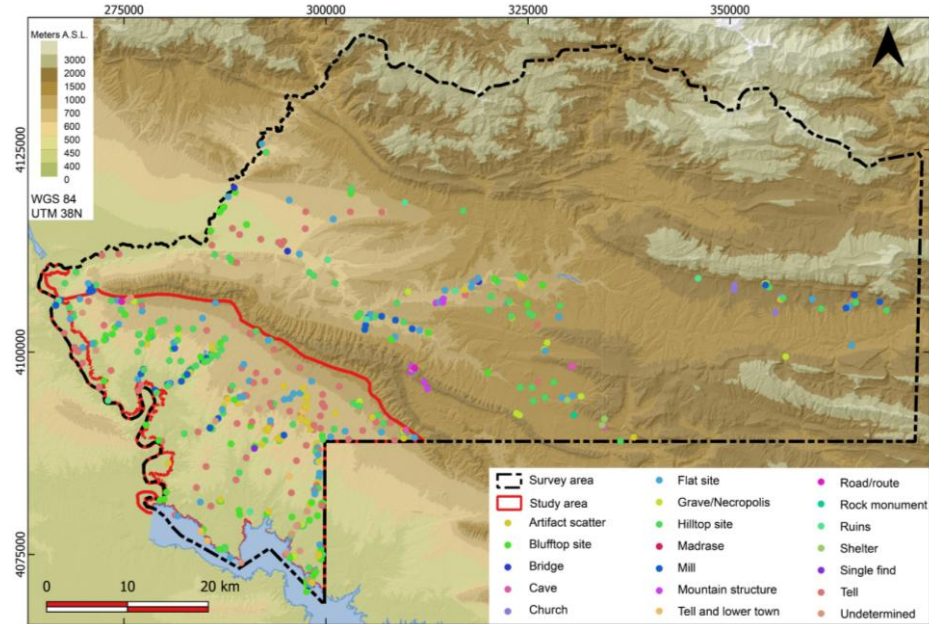
SETTLEMENTS PATTERNS

“Resource hunting” strategy from Mesopotamian empires (monopole, acculturation...)

≠

Local indigenous uses of resources (close environment)

Are soil data relevant in revealing settlements in Iraqi Kurdistan?



Study area with archaeological settlements location

MOVEMENT AND POWER INFLUENCE IN THE SELEVANI PLAIN (4TH AND 3RD MIL B.C.)

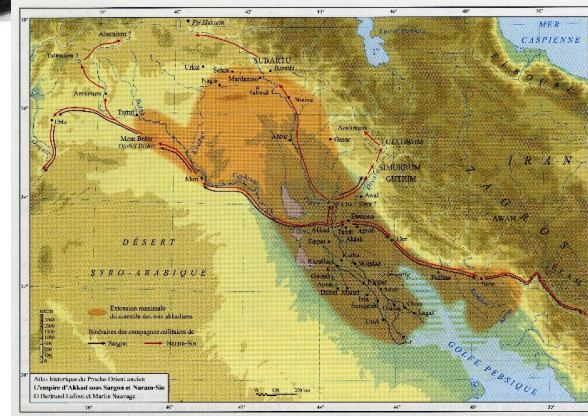
- Military Campaigns and Expansion of the Akkadian Empire
- Statue from Bassetki with Inscription of *Naram Sin*
- Region is part of a inter-regional route system (3. millenium BC) with peak in old assyrian trade network
- Maradaman (=Bassetki) is mentioned as one of the stops along the trade routes



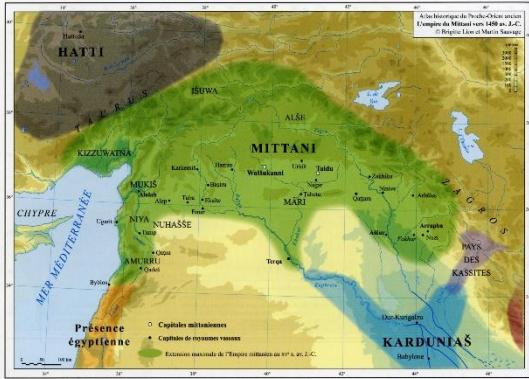
Aruz 2003,
Art of the First Cities



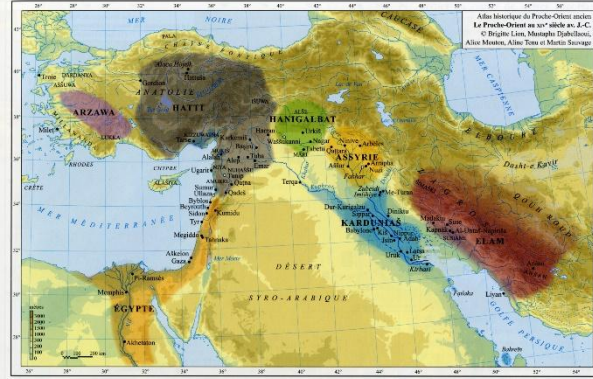
From Sauvage 2021,
Atlas Historique du Proche-Orient Ancien



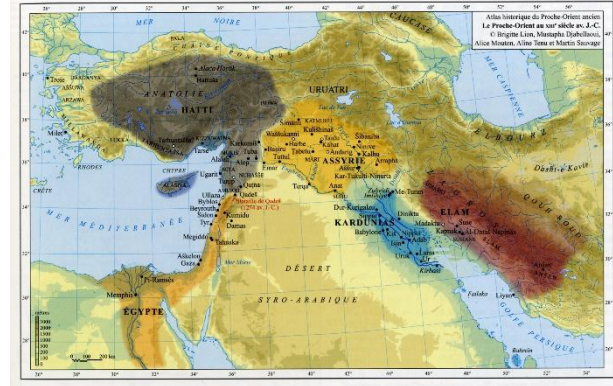
MOVEMENT AND POWER INFLUENCE IN THE SELEVANI PLAIN (2TH MIL B.C.)



15th century BC



14th century BC



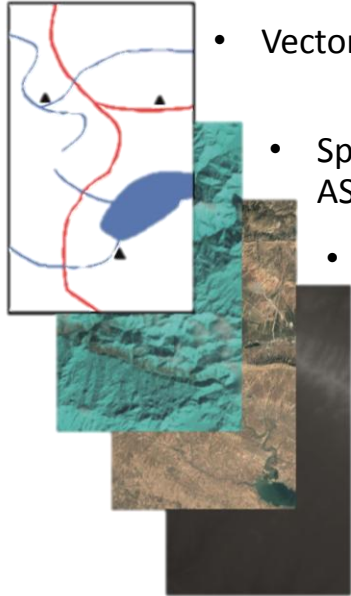
13th century BC

- 15th century: part of Mitannian Empire
- 14th – 13th centuries: Middle Assyrian Empire
- Part of Neo Assyrian Empire with Relief of Mila Mergi



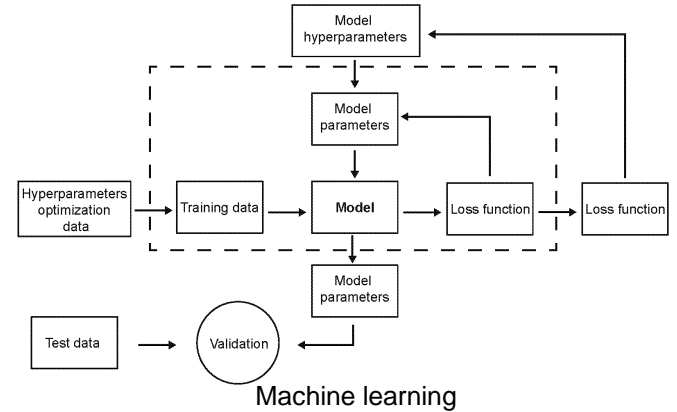
Relief of Mila Mergi with
Inscription of *Tiglathpilesar III*

From Sauvage 2021,
*Atlas Historique du
Proche-Orient Ancien*



- Vector: Sites, roads, rivers, hollow ways...
- Spectral data: Landsat 8/9, Sentinel 2, ASTER (NIR, color, thermal bands...)
- Ortho-images : Copernicus, CORONA, HEXAGON
- Others raster: DEM, Soil data, Models (precipitation, temperature...)

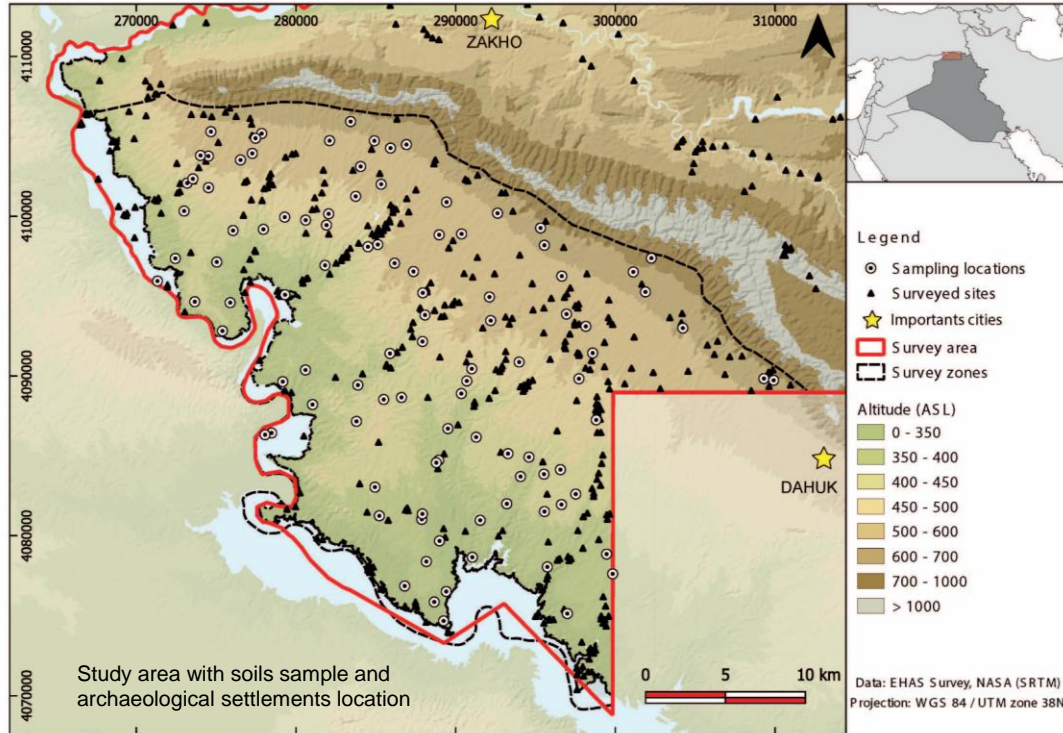
ARCHAEOLOGICAL PREDICTIVE MODELS (APM)



Collect data on the field
Prepare data (ICP, FTIR, ML)
Compute soil maps

Probability map of sites
Analyse important factors

Analyse the patterns for each periods and sites type (clusters, attraction areas)

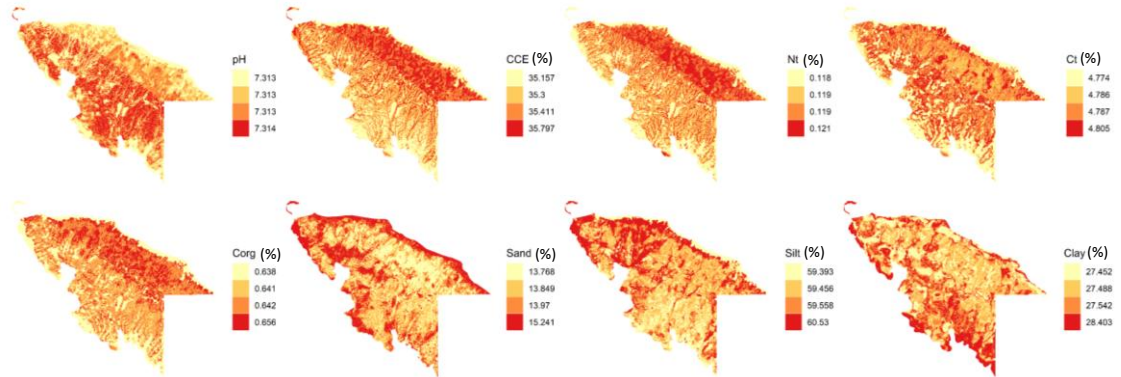


2022 SOIL SAMPLING CAMPAIGN

- 800 km²
- 336 surveyed sites
- 101 soils samplings locations in 2022
- 435 soils samples token
- Cluster Latin Hypercube Sampling used for the sampling design
- 8 different soils properties measured
- 50 soils samples measured and 385 predicted *via* FTIR (MIR)

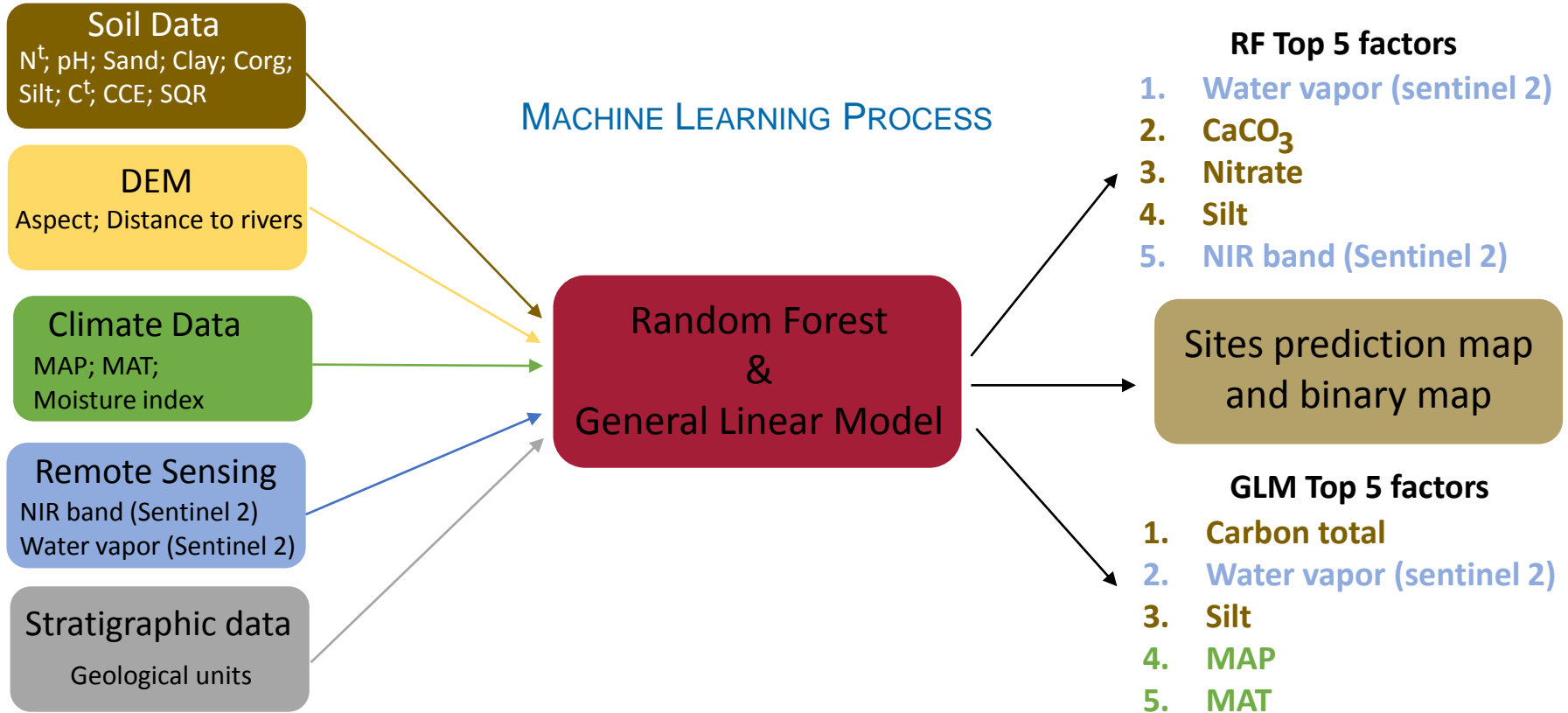
Data	Source
Aspect	DEM
Distance to rivers	DEM
Mean annual precipitation (MAP)	Climate data
Mean annual temperature (MAT)	Climate data
Moisture index	Climate data
NIR band (Sentinel 2)	Remote sensing
Water vapor (Sentinel 2)	Remote sensing
Geological unit	Stratigraphic data

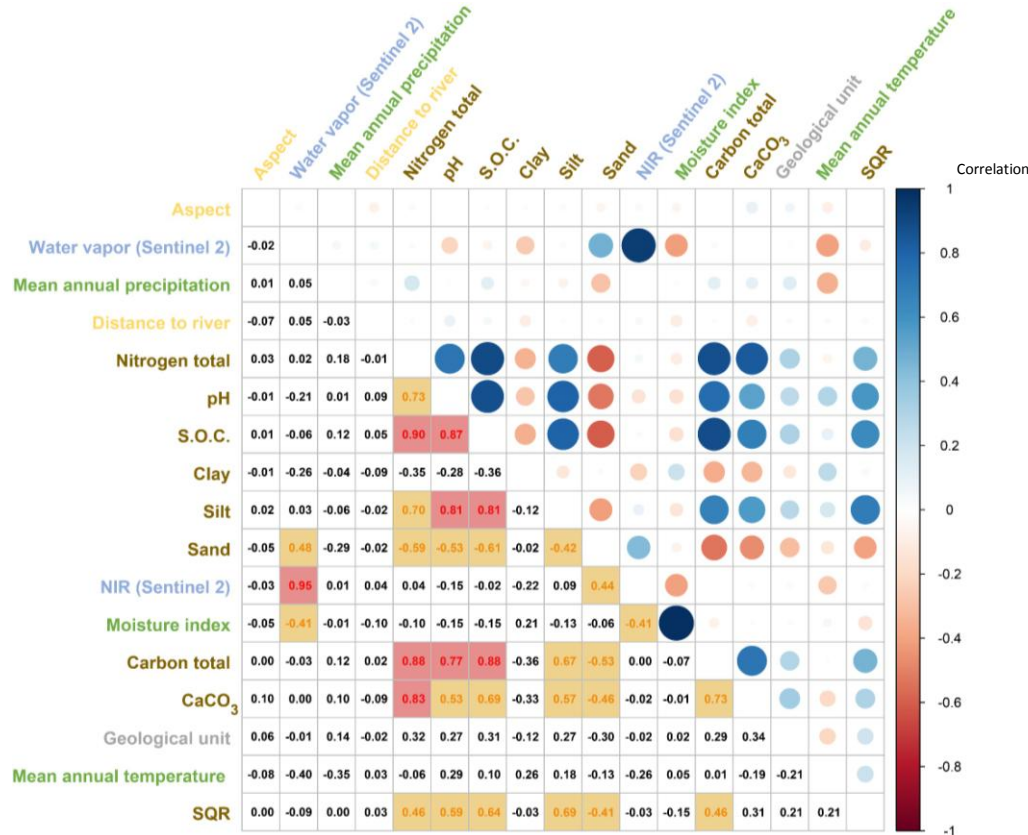
DATA



Map of the soil properties predicted from soil samples of 2022 sampling campaign with random forest

- 51 different predictors were used at the beginning
- The 8 predictors with the lowest auto-correlation were used for the final model

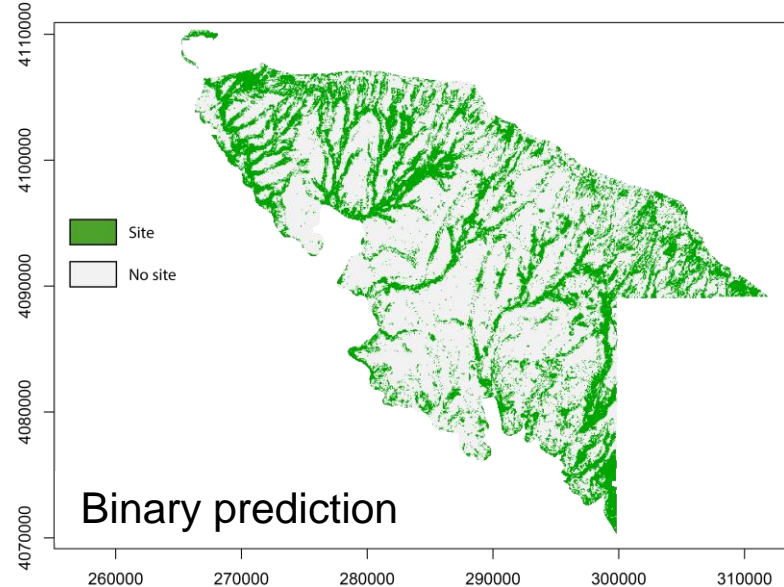
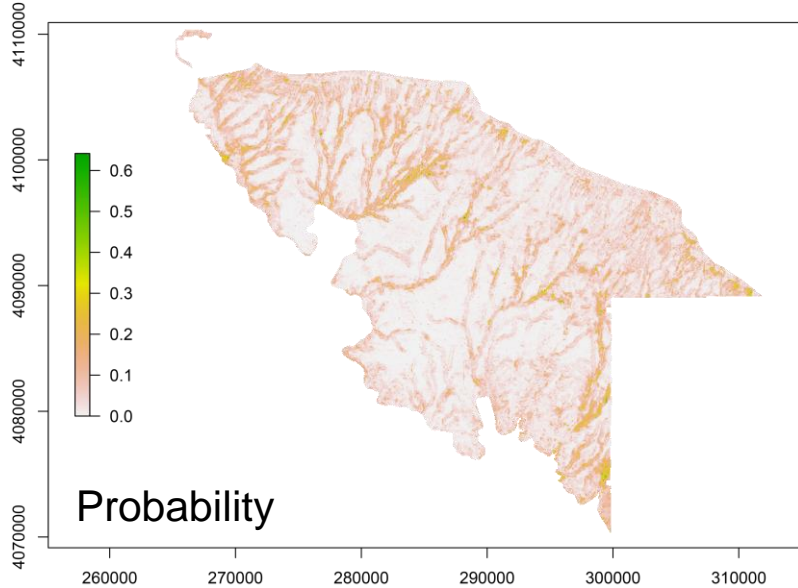




SOIL DATA CORRELATION

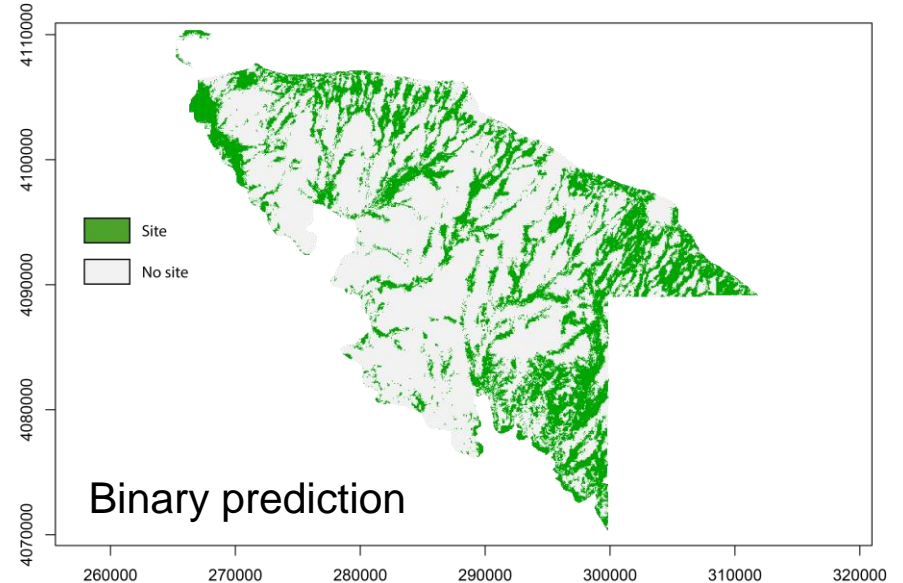
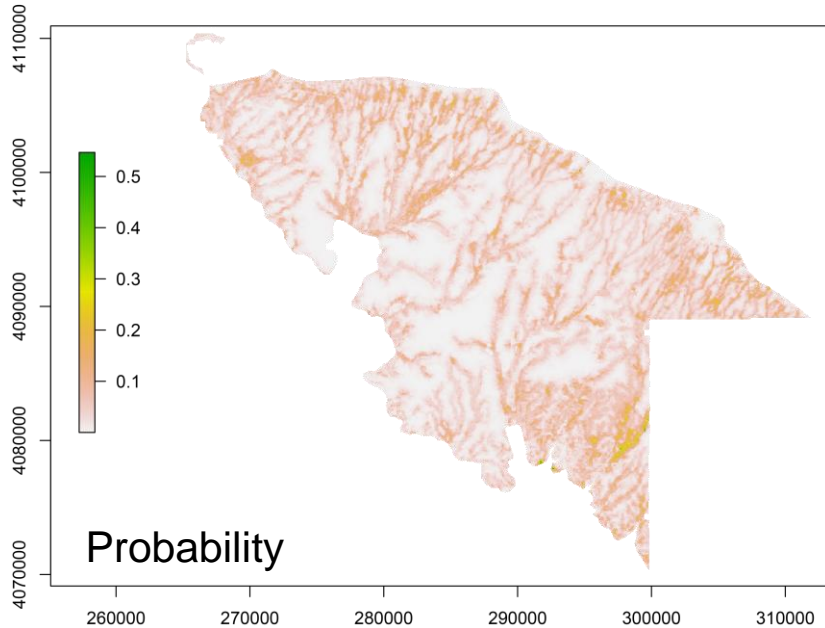
- Some factors have a low variation (soil type, soil quality ranking)
- A high auto-correlation between soil factors
- $0.4 > 0.75$ or $-0.4 < -0.75$: medium correlation
- $0.75 >$ or $-0.75 <$: high correlation
- A need to find new soil parameters for more homogenous APM models

RANDOM FOREST PREDICTION MAP FOR ARCHAEOLOGICAL SETTLEMENTS



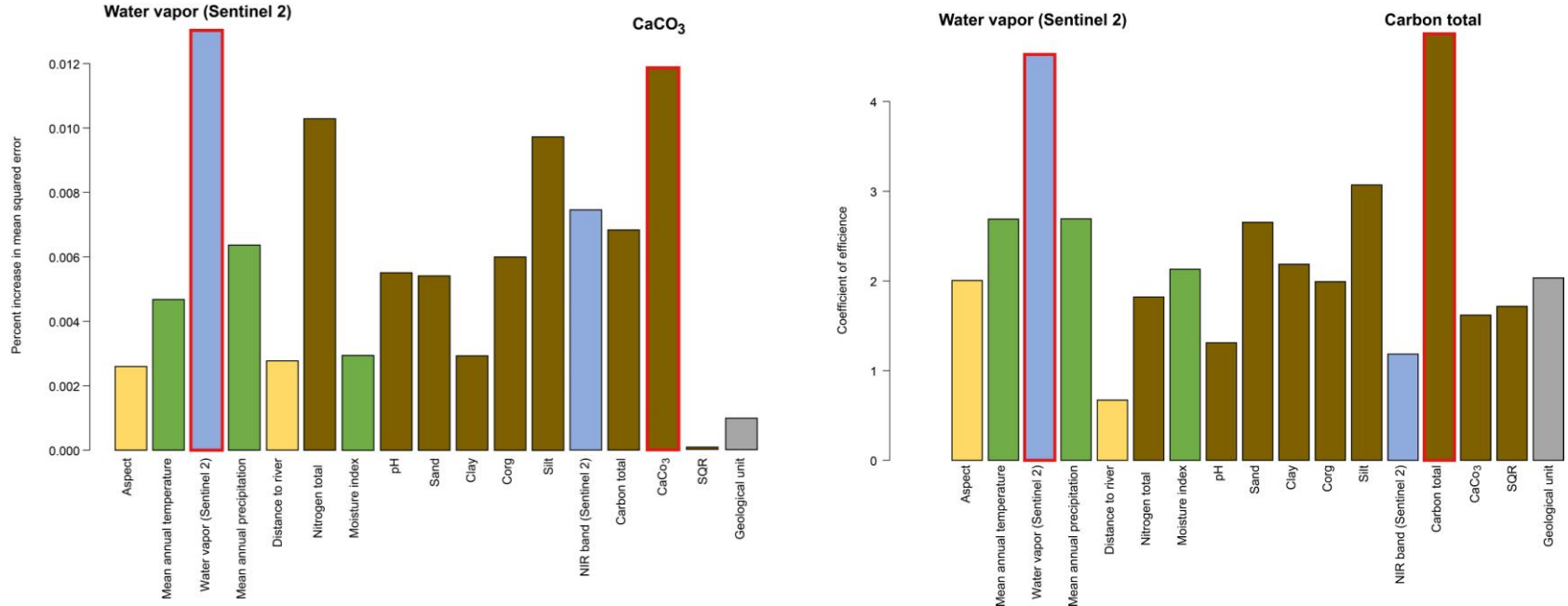
- Random forest (RF) was used for the prediction as a machine learning method
- The precision of RF model is 73% which can still be improved

GENERAL LINEAR MODEL PREDICTION MAP FOR ARCHAEOLOGICAL SETTLEMENTS



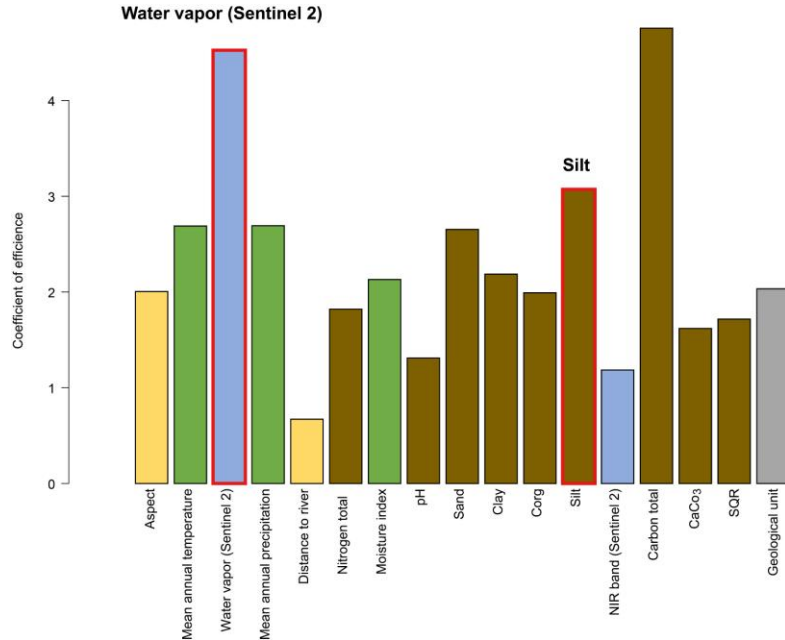
- General Linear Model (GLM) was used for the prediction as a statistical method
- The precision is lower than RF but much more easier to run

IMPORTANT FACTORS (FEATURE SELECTION)



Most important factors of RF model (left) and GLM model (right)

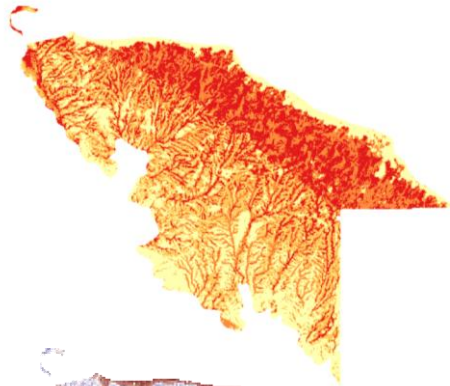
IMPORTANT FACTORS (FEATURE SELECTION)



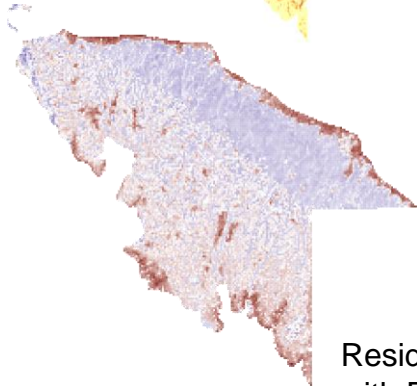
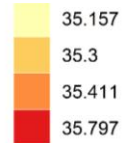
Most important factors of GLM model with auto-correlation taken into account

- The **Water vapor** from the Sentinel 2 is the most important factor in both RF and GLM
- For the RF model **CaCO₃** is the most important factor
- For the GLM model **Carbon total** is the most important factor
- If we take the auto-correlation into account **Silt** is then the most important factor with a low correlation

INTERPRETATIONS



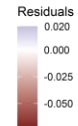
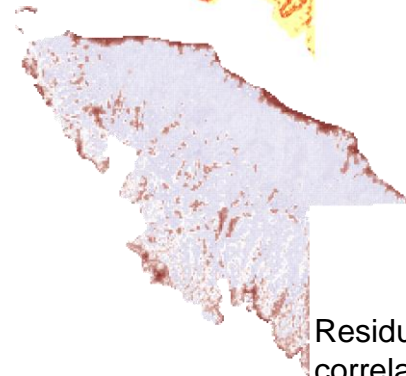
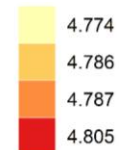
CaCO₃ (%)



Residuals of CaCO₃ correlation
with RF prediction map

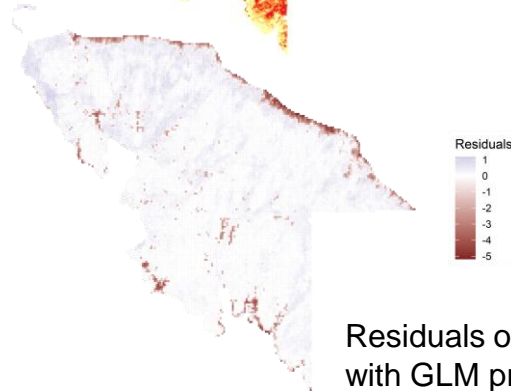
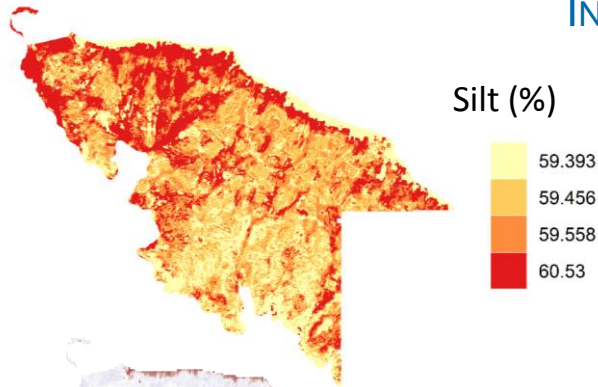


Carbon total (%)



Residuals of Carbon total
correlation with GLM prediction map

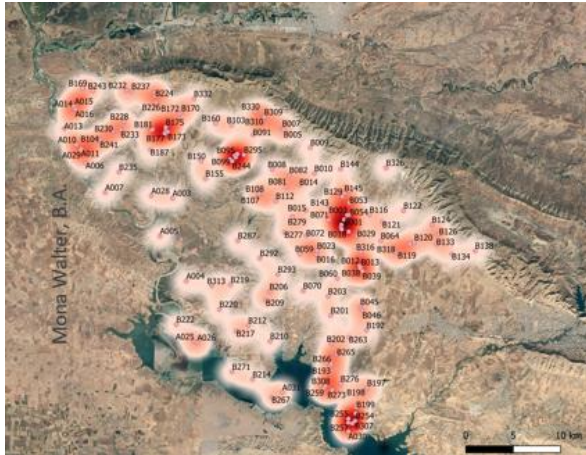
INTERPRETATIONS



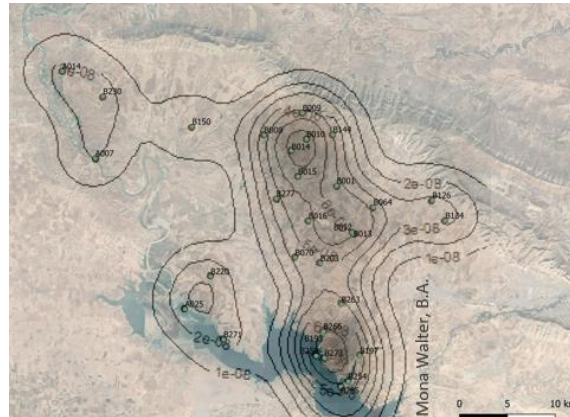
Residuals of Silt correlation
with GLM prediction map

- High calcium carbonate concentration could be a **reason** or a **result** for settlements installation
- High carbon concentration is related to land use and can be a **result** of settlements installation
- Silt concentration is related to the type of soil and is a **reason** for settlements installation

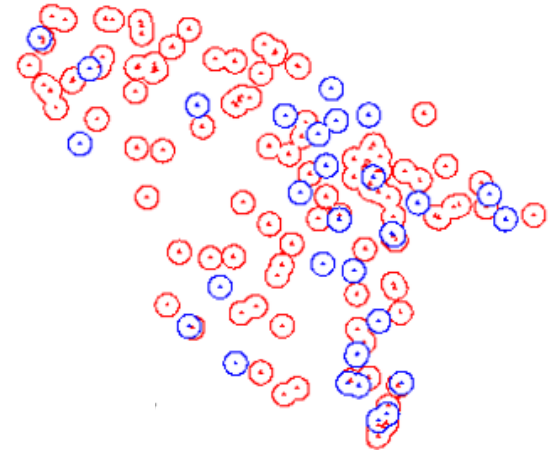
SETTLEMENTS CLUSTERS AT THE BRONZE AGE (1600 – 1350 B.C.)



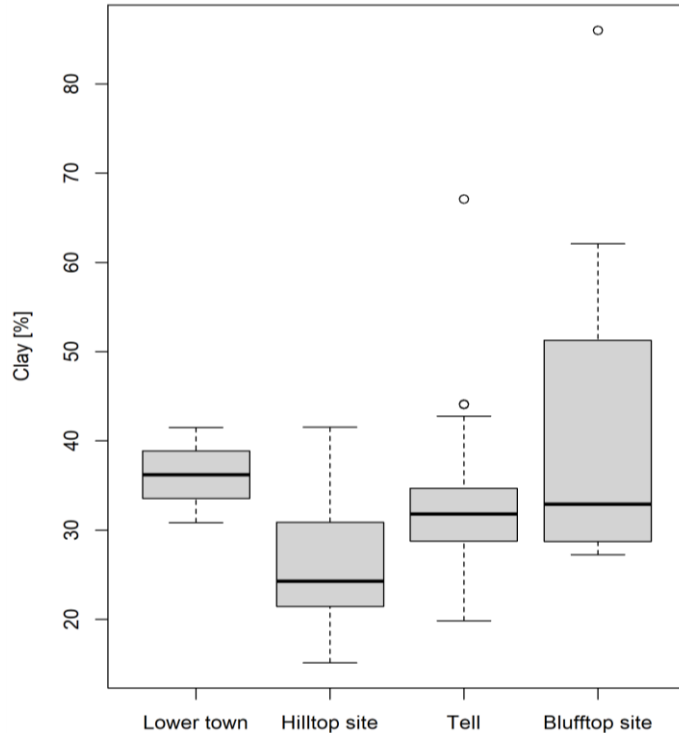
Hot spot analysis (M. Walter)



Core density estimation (M. Walter)



Buffer zone analysis (M. Walter)



Clay content by site type (n = 66)

SETTLEMENTS CLASSIFICATION WITH SOIL PARAMETERS

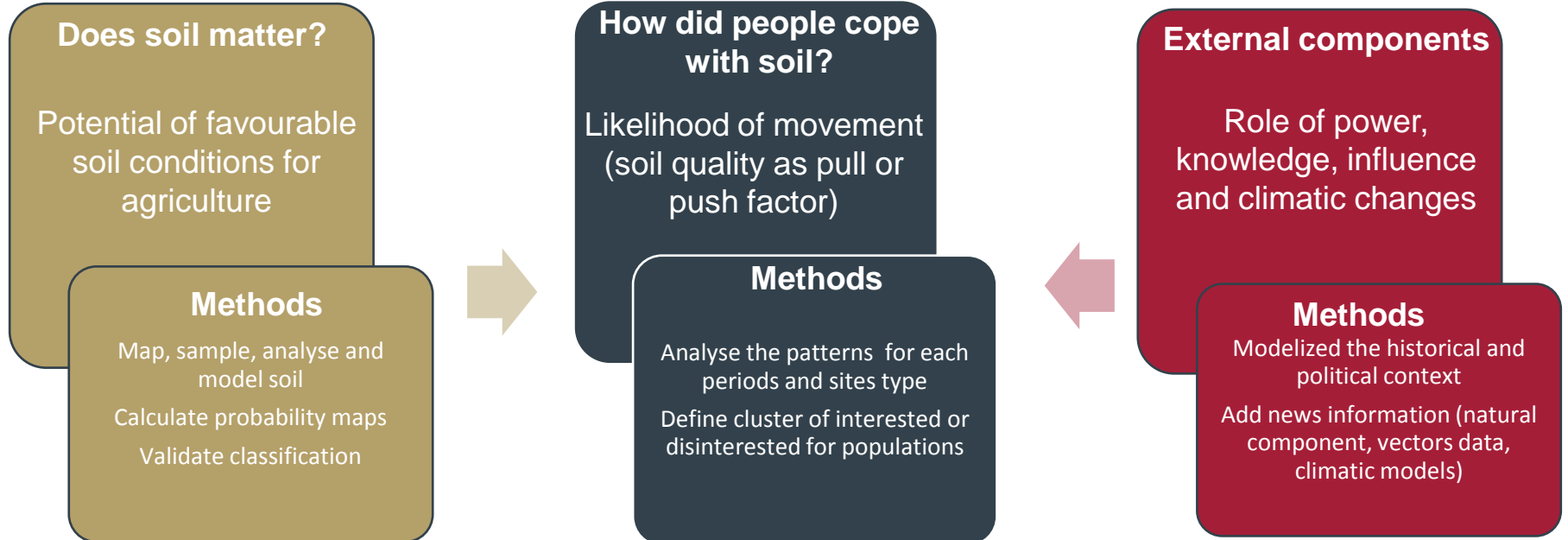
Anova on clay content per site type $\Pr(>F) = 0.002$

Tukey HSD table points out significant differences between

- Hilltop/Blufftop sites (p.value = **0.0012**)
- Hilltop/Tell sites (p.value = **0.0266**)

Clay content **seam to have a significant influence** on site type

INFLUENCE OF SOILS IN SETTLEMENTS PREDICTION



- ❑ Machine learning combined with statistics can underline unseen relation between sites and their environments
- ❑ Soil information can be used for spotting the likelihood of settlements toward different chronological periods
- ❑ Soil information can predict but also differentiate the types of sites and maybe chronological distinction in the future

Soil information in particular Carbon, CaCO₃ and Silt are relevant for predicting archaeological settlements at a meso/macro-scale

Thank you for your attention



mathias.bellat@uni-tuebingen.de

