EPSC Abstracts
Vol. 14, EPSC2020-474, 2020
https://doi.org/10.5194/epsc2020-474
Europlanet Science Congress 2020

Using supervised machine learning methods to improve the selection of analogue sites for studying habitability of the sub-surface ocean of Europa

Alvaro del Moral, Victoria Pearson, Mark Fox-Powell, and Karen Olsson-Francis
AstrobiologyOU, School of Environment, Earth and Ecosystem Sciences, STEM Faculty, The Open University, Milton Keynes, United Kingdom. (alvaro.delmoral@open.ac.uk)

Introduction

Owing to the limited amount of data from other planetary bodies in our Solar System, terrestrial analogues are frequently used to further the understanding of extraterrestrial environments. One of the moons of Jupiter, Europa has been selected for future exploration in our Solar System. Measurements carried out by the Galileo spacecraft suggest the existence of a tidally heated, sub-surface ocean that could harbour life [1]; yet the composition of the ocean is unknown as there have been no direct measurements [2, 3]. The information about Europa is thus limited but has been complemented by modelling surface and interior processes [4] (fig. 1). Surface features and data from Galileo initially suggested an ocean dominated by magnesium sulfate; however, recent observations suggest it is dominated by sodium chloride with a high concentration of magnesium [5, 6, 7]. Considering the limited data available about the interior of this moon, analogue studies can be used to extend our understanding of the crucial constraints on microbial habitability below the ice shell.

To date, terrestrial analogue studies for Europa have focused on subglacial lakes, cold springs, and sulfate lakes. These sites, however, have been selected based on a single or few parameters. Some are of interest due to their pH level, temperature, or main ionic composition. This approach often ignores or compromises other important variables that may influence habitability, such as pressure, dissolved gases or minor ions. The aim of this work is to develop a quantitative method to evaluate the similarity between terrestrial analogue sites and models of the Europan ocean. This will improve the selection of appropriate terrestrial analogues, the consequent validity of the results from microbiological and geochemical studies, and thus, their correlation to studies of the potential habitability of Europa.

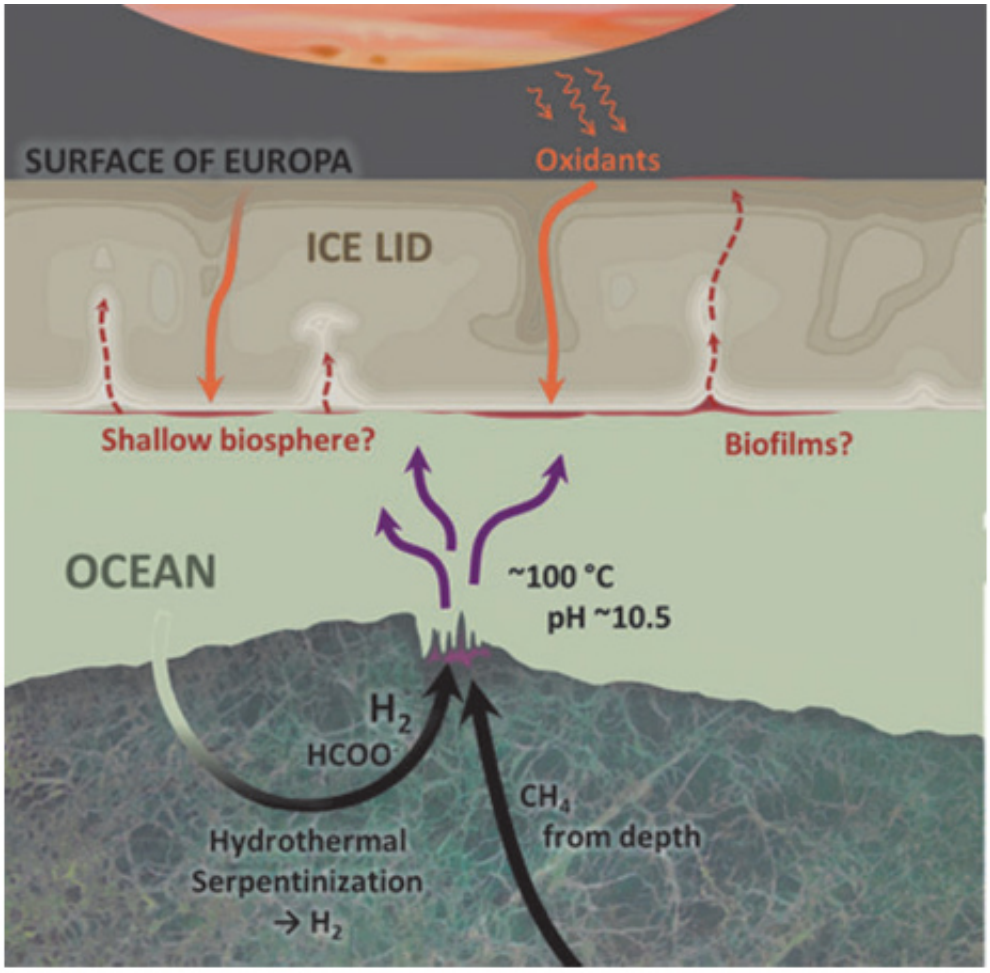

Figure 1: Cross-section of the potential structure of Europa's ocean. Reproduced from Russell *et al.* 2017

Objective analogue site selection requires a quantitative method with which to evaluate the similarity between terrestrial analogue sites and models of the Europan ocean and that takes into account all major variables that could determine habitability. Given the number of possible variables, classifying analogue environments is an inherently multivariate problem. Current classification systems for aqueous environments tend to fall short as they rely on human input, use inconsistent parameters or do not account for all relevant physicochemical conditions. To address this issue, machine learning models can be used, which are self-improving algorithms able to learn what a categorised dataset looks like and then make predictions on what category new data will fit into [8].

Methods

A terrestrial analogue database was constructed with published chemical (ionic composition), temperature, pressure, and pH data from potential analogue sites, including subglacial lakes, saline lakes and hydrothermal environments. Using a range of multivariate classification methods and clustering solutions, categories were defined by the full range of known physicochemical characteristics. Machine learning algorithms were trained with such categories and the highest accuracy model was selected.

New samples (such as new potential analogue sites, data from a Europan model or even direct measurements) can be fed into the trained model and it will output a so-called similarity score. This score is based on how similar the model calculates this new sample is to each of the categories of

the training dataset. The category with the highest similarity score contains the sites that should be considered analogous to the new sample; in our case this new sample can be any Europan composition model found in the literature.

Cross validation methods and genetic algorithms were used to improve the models and thus the accuracy of the similarity scores. Cross-validation creates subsets of the database and assigns one to be used to evaluate the model and the rest to train it. The process then repeats by selecting a different subset to be used to evaluate the model, leaving the rest as part of the training set, until every subset has been used to both train the model and test it. Genetic algorithms operate on the principle of biological evolution, by generating multiple versions of the model, changing certain parameters stochastically (mutation) and selecting those with the desired metrics to populate the subsequent generation of versions of the model. These processes were applied to several machine learning algorithms to select the ones with the highest accuracy.

Results

We applied the model of selecting possible analogues to existing estimates of Europa's ocean chemistry. Initial classification (accuracy 90.6%) attempts suggest the use of low temperature sites: either sulphate-saline dominated lakes like Basque Lake (Canada) or Last Chance (Greenland) (similarity 96.0%); or calcium-rich lakes like Don Juan Pond (Antarctica) or El Chichón (Mexico) (similarity 99.9%). The similarity between the physicochemical data of the modelled Europan ocean and the terrestrial analogue sites can be calculated with high accuracy, justifying the use of microbial communities or geochemical samples from these environments for studying the habitability of extra-terrestrial environments. Ongoing work is improving the predictive accuracy of our machine learning model. The results from this work demonstrate that machine learning algorithms can improve the selection process of field sites for the study of microbial ecology and astrobiology.

References

[1] Russell et al. (2017). Astrobiology, 17(12), 1265–1273. **[2]** Kivelson et al. (2000). Science, 289(5483), 1340–1343. **[3]** Pappalardo et al. (1999). Journal of Geophysical Research E: Planets, 104(E10), 24015–24055. **[4]** Jin & Ji (2012). Science China: Physics, Mechanics and Astronomy, 55(1), 156–161. **[5]** Trumbo et al. (2019). Science Advances, 5(6), 2–6. **[6]** Zolotov & Shock (2001). Journal of Geophysical Research E: Planets, 106(E12), 32815–32827. **[7]** Ligier et al. (2016). The Astronomical Journal, 151(6), 163. **[8]** Mander et al. (2013). Proceedings of the Royal Society B: Biological Sciences, 280(1770).