Received: 25 May 2021 | Accepted: 21 January 2022

DOI: 10.1002/saj2.20385

Soil Science Society of America Journal

SOIL PHYSICS & HYDROLOGY

Heuristics-enhanced geospatial machine learning (SaaS) of an ancient Mediterranean environment

Karl B. W. Svatos[1] | **Zubair Ahmed**[2] | **Angela Recalde-Salas**[2] | **Tom Young**[2]

[1]Western Australia State Agricultural Biotechnology Center, Murdoch Univ., 90 South St., Murdoch, Western Australia, Australia

[2]Faculty of Science and Engineering, Curtin Univ. of Technology, Bentley Drive, Perth, Western Australia

Correspondence
Karl B. W. Svatos, Western Australia State Agricultural Biotechnology Center, Murdoch Univ., 90 South St., Murdoch, Western Australia, Australia.
Email: k.svatos@murdoch.edu.au

Assigned to Associate Editor Richard Whalley.

Funding information
iPREP WA Industry and PhD Research Engagement Program Administered formerly through Edith Cowan University.

Abstract

Raw soil core physical data used in machine learning algorithms with corresponding spatial remotely sensed data is an emerging science. Using data derived from soil core samples previously collected in Universal Transverse Mercator zone 50 (Western Australia) and remotely sensed data, a model that predicted ground movement (GM) was developed specific to Australian Standards manual AS 1726–2017. This is the first approach for Australian soils and first in the world for soils older than 200 million yr. The model developed reliably predicted GM with 91.1% accuracy. The error obtained from the prediction is within acceptable limits currently used by engineers in calculations concerning soil classification for engineering purposes. Concerning the remotely sensed data analyzed, accuracy of the Atterberg limits method might be improved if additional information about soil structure (layering and horizon) or other variables (seasonal data) are built into this model. This model can be used to save on construction material costs, reduce the potential for human error associated with data collection and sample manipulation, but also fast-track (by up to 6 wk based on current wait times) building approvals while ensuring compliance to the relevant legislation. This platform also reduces the environmental effects of invasive drilling techniques. A requirement within principles of sustainable building practices, and associated with current standards commonly used by structural engineers who may seek better understanding of soil properties in Australia as a software service (with application potential in North America).

1 | INTRODUCTION

The surface geology and geomorphology of soils habiting the Swan Coastal Plain in Western Australia are ancient, highly weathered, and diverse. The parent material, which includes an igneous formation located west of the Darling Range dating back at least 2,600 million yr, informs the soil physical and geological landscape of that locality; broadly, framework silicate–based, iron-rich sands that gradually increase to more calcareous mixtures near the wave-affected coastline (McArthur, 2004). Modeling of geological age within this geographical location constitutes a broad range in terms of soil era and thus has had limited uptake concerning the building of new AI-based models, unlike other industries where computational advancements may be rapidly outstripping concise data collection strategies but conversely are having significant positive statistical gains in the order of 5–10% (improvement in prediction accuracy) (Arama et al., 2020; Guo et al., 2021; Yu et al., 2021) (Figure 1).

Abbreviations: GM, ground movement; ISRIC, International Soil Reference and Information Centre; SaaS, software as a service; UTM, Universal Transverse Mercator.

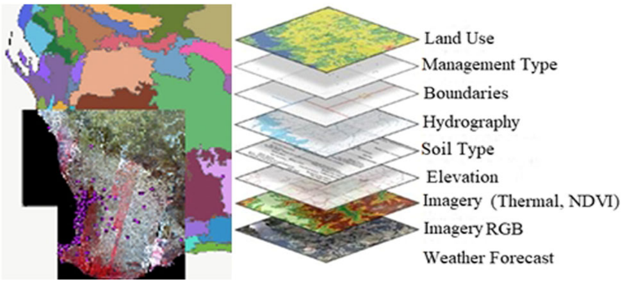

FIGURE 1 The soil lithology map of Western Australia (McArthur, 2004) and overlaying RGB Landsat satellite Universal Transverse Mercator zone 50 images (black background, soil core logged, and geo-referenced *.tiff; location data coordinates) for laboratory samples used to create the polygon of the representative dataset GPS points for predicting ground movement based on Atterberg values (McBride, 2002) determined for the laboratory core logged datasets (left). Constructed images of the various inputs and data sources that can be used to program an algorithmic model based on laboratory datasets from any number of potential sources (real-time API's, satellite data, SQL databases, etc.) (right). The layer cake nature shows how multiple indices, including imagery and weather forecasting, can be used to create statistical models to assess effects of multiple co-factor environmental parameters (e.g., on physical or chemical association with ground movement), which can then be used to determine the influence on any constructed algorithm(s) or for future prediction based on prevailing conditions at any given time. The two main approaches include machine learning and statistical modeling. In each case, to enable real-time decision making the data must be converted to digital formats and then instances created to process the data

Core Ideas

- High-resolution spatial analysis gives deep soil insights about an ancient Mediterranean soil.
- We discuss a new soil machine learning engineering prediction method and corresponding software platform.
- Eliminating misconceptions about machine learning through sound statistics consolidates Atterberg values used to determine "ground movement" in the context of Australian (and international) standards.

The accepted legislated standards used to classify soils to determine site suitability in the building and construction industry in Australia are listed in Appendix A of the Australian Standards manual AS 1726–2017 (Australian Standards, 2017), which focuses on consistency limits and soil plasticity that and is broadly related to physical properties of some fine grained soils and provides information about their engineering properties (e.g., shear strength and compressibility). These parameters are the standard input for many soil-based investigation programs. Their correct definition and usage in such programs are essential. (Osman, 2007).

Concerning calculating how much any soil might move over any given period or number of seasons throughout the life of a building structure, a general term currently in use is ground movement (GM) (Cameron, 1989). Ground movement is based on traditional geological classification of a type of "problem soil" that may exhibit expansion or shrinkage that damages foundations and building structures. This phenomenon occurs in many parts of the world, including in Australia and North America, although, technically there is no such thing as a "problem soil type" per se. There are several publications related to sampling techniques for the purpose of classifying and defining the properties of such problem soils. One such example is "heave" as it relates to plasticity index and percentage clay fraction (Jones, 2017; Van der Merwe, 1964).

Simulations of GM concerning unsaturated clay in North America (Rees & Thomas (1993, 7) and is analogous to Terzaghi's classical theory of one-dimensional consolidation for saturated soils. They conclude that the approach they adopted "was capable of producing realistic predictions of seasonal ground movement." This summary of soil and structural engineering related representations (GTM-7, 2015; McBride, 2002) coupled with review (Johnson et al., 2020; Vorwerk et al, 2015) related to the physical calculation of Atterberg limits generally represent GM hereinafter for the purpose of the software as a service (SaaS) platform (Table 1).

Concerning the calculation of GM used in the building and construction industry (for calculations and legislation pur-

TABLE 1 Typical ground movement (GM) values between 0 and >75 cm used to help determine soil classification

Ground movement	Soil classification
cm	
0	A
1–20	S
20–40	M
40–60	H1
60–75	H2
>75	E
Problem site	P

Note. A, mostly sand and rock sites with little or no GM expected; E, extremely reactive sites that can experience extreme GM from moisture changes; H, highly reactive clay sites that can experience high GM from moisture changes; M, moderately reactive clay or silt sites that can experience moderate GM from moisture changes; P, problem sites, which can include soft soils, such as soft clay or silt, varying depths of fill, loose sands, landslips, mine subsidence, collapsing soils, soils subject to erosion, reactive sites subject to abnormal moisture conditions, or sites that cannot be classified otherwise; S, slightly reactive clay sites with only slight GM from moisture changes expected.

poses), several mixed-use models have been presented but with little or no clear focus on the industry requirements concerning noninvasive drilling approaches, or standardization. As companies and scientists look to move more data to the cloud SaaS to benefit from opportunities related to building sustainability and data sharing, herein lies both a knowledge gap and an opportunity (Svatos, 2021). More recent research has provided insights and potential database metadata clusters (and metrics) for correlations (Armstrong et al., 2007 Sutherland et al., 2014; Qu et al., 2018; Tan et al., 2021), but these may also fall short concerning industry standards on data collection.

The International Soil Reference and Information Centre (ISRIC) SoilGrids (a java-based cloud *.html console API) provides predictions for standard numeric soil properties (organic C, bulk density, cation exchange capacity, pH, soil texture fractions, and coarse fragments) at seven standard depths (0, 5, 15, 30, 60, 100, and 200 cm). Based on the American soil classification system (taking into account that the majority of American soils are <200 million yr old), averages for depth intervals are derived by taking a weighted average of the predictions using numerical integration of the trapezoidal rule (Hengl et al., 2017). This algorithm then provides a generalized surface and subsurface prediction of the aforementioned soil properties based on current geological era correlation data (Hengl et al., 2017). However, where the classification of ancient soils skews correlation outside allowable model tolerances, many or all data points have to be omitted to populate the algorithm map. This omission may also be partly due to areas incorrectly reflecting significant differences in era and age between the Australian and American classification systems that does not account for occurrences of soils >200 million yr in geological time (McArthur, 2004). An ISRIC spokesperson recently said that the ISRIC is not working on the prediction of soil classes at the global scale. The number of profiles with assessed classification in each of the three systems is not very large; at this moment, it is not yet possible to substantially improve on the class prediction (Luís Moreira de Sousa, personal communication, 16 July 2020). And yet, the entire global reference dataset (GPS grid points) is kept on the map grid for use. This nonstandardized method (for prediction at the global scale) is not limited to soil prediction models. Concerning vector-based algorithms for statistical applications (including for soil) or other quantitative modeling methods, including those utilizing machine learning, biases (considering predictive significance) appear to be underreported in the scientific literature, yet are increasingly used to explain the significance of the methods. (Louca et al., 2018).

Data-driven machine learning techniques have increasingly been used for modeling and prediction of the composition and rates of sedimentation (e.g., in sedimentary basins). As such, the existing mathematical and machine learning models for forecasting river sediment deposition is oft driven by non-remotely sensed data without the required complementary ground-truthing for calculating tolerances or absolute significance (Qu et al., 2018). This soil moisture investigation uses a machine learning approach based on experimental data and Landsat5-thematic mapper images, specific for the Mega-City Beijing. Qu et al. (2018) aims to demonstrate that remote sensing and unsupervised machine learning techniques, coupled with an appropriate validation metric, can be used to quickly forecast regions that are subject to future river sediment deposition environments. They claim that their support vector classifier (SVC) method trained with remote sensing and grayscale data achieved an accuracy of 76.69%. However, without the support vector classifier, it is a fair assumption that the predictive accuracy of such a model would be significantly less biased concerning predictive significance, but also considerably less accurate (Awan et al., 2021; Huang et al., 2019; Sheffield et al., 2012).

But, to what extent do computer libraries and available statistical visualization and computational hardware setups allow for predictions (with or without ground-truthing of algorithmically extrapolated data) independent of “accuracy” concerning applied AI used in soil analyses? The Python3 geospatial data abstraction library may be used for converting satellite image *.tiff files to *.csv files, and R (computer language) is used for unsupervised learning algorithms including those used by AI currently. Increasingly the R libraries can optimize numbers of clusters for pairwise plotting of feature bands and cluster validation (Karmakar et al., 2019; R Core Team, 2022). However, where vast distances need classification for soil physical and chemical properties via index correction or vector algorithms, these big-data clusters create significant processing bottlenecks (compared with genetic data clustering, which also use this type of approach). These types of models use substantial graphical processing unit resources, and thus are inefficient or not Turing complete. (Fong et al., 2018; Keight et al., 2018; Reynolds et al., 2019).

The aims of this study are: (a) to characterize the major landforms, surface features, and physical specifications of a standardized set of core sample logs (based on a laboratory database used to calculate GM) specific to Western Australia’s ancient and unique soils, including those along Swan Coastal Plain; (b) to validate the formulated equations and indices using complementary statistical modeling and machine learning methods; (c) to validate GM soil equation(s) using existing remote sensing databases and software, to correlate or extrapolate information concerning GM for areas without laboratory data in the data polygon, to simulate lack of access to the site, and to minimize future site destruction; and (e) to embed the kernel into a cloud-based container for high-speed computation and software as a service (SaaS) API access.

2 | MATERIALS AND METHODS

2.1 | Validation of the laboratory-based geotechnical data

Ground movement values (determined from the accurate conversion of physical laboratory measurements based on Atterberg limits), using the industry standards (Australian Standards, 2017; McBride, 2002) were first completed on 5,000 core samples collected from Western Australia's Swan Coastal Plain and surrounding areas (Table 1). The raw results data were then uploaded to an online database AWS S3 bucket (Amazon Web Services). The entire dataset of laboratory samples stored in the AWS bucket consisted of 3,000 datasets that were formatted from *.pdf format into *.xlsx and then *.csv. These datasets were also cross-referenced by suburb and general location to remove inconsistent GPS locations from their physical addresses and plotted over the Western Australian lithology map (Figure 1).

The laboratory database GM*.csv files were then deconstructed to look for relationships between the sample observations (based on the theoretical algorithm vs. the calculated values) to verify the Atterberg method. Relationships between approximately 30 soil characteristics (from the laboratory *.csv files) and an additional 20 characteristics from the available remote sensed geo-referenced *.tiff and GPS logged satellite or online databases (Landsat, Sentinel I and II, and ISRIC) were then determined for the Universal Transverse Mercator (UTM) zone 50.

2.2 | Remote geospatial ground movement prediction based on statistical relationships

The methods used for machine learning and statistical modeling (Tan et al., 2021; Yu et al., 2021) implemented here have previously been described (Ahmed et al., 2019; Recaldes et al., 2020; Svatos, 2021). The theory behind the installation and configuration of such implementations for environmental simulation has also previously been described (Svatos, 2018). Briefly, the two complementary methods were set up and preconfigured to be remotely accessible from a remote secure shell terminal. The methods included a statistical modeling approach using the QGIS graphical user interface and command-line interface through the QGIS software suite (QGIS Development Team), and a machine learning approach using a *.py (python file) with preconfigured commands implemented through the command-line of a pythonic kernel and console API, running on the long-term support version of the Ubuntu 18.04 operating system. To enable near real-time automation of these statistical methods, a pipeline for the baseline implementation (workflow in the cloud) was created. Using Amazon EC2, the instance was set up and installed with access to an S3 bucket containing the laboratory database and remotely sensed raw datasets (Amazon).

From the cloud shell, the statistical models were used to determine GM based on the relationships between soil characteristics and GM (using 80% of the data for training and 20% for testing); the two analytical approaches compiled algorithms for downstream supervised learning assessments (within the GPS polygon). Finally, to determine the effectiveness of the statistical modeling and machine learning pipeline, the workflow algorithms were used to see how accurate the correlations were with open-source geospatial data for determining GM (Table 1).

3 | RESULTS AND DISCUSSION

The scientific models developed through this research (based on laboratory datasets of the Atterberg limit industry standards) are the first to incorporate machine learning into GM calculations in Western Australia, and the first for soils older than 200 million yr due to the age of soils studied (Figure 1). Based on the statistical modeling approach, the best determining correlation variables for GM were liquid limit, plasticity, plasticity index, and linear shrinkage, due to the nature of the relationships between GM and soil physical properties that support its measurement (as expected). The generalized additive model output from the QGIS terminal, which was set for spatial and temporal autocorrelation, with additional smoothing, fitting, weight, and flexibility parameters, explained 91.1% of total deviation of the dataset (Figure 2). Similarly, for the machine learning approach, the overall prediction potentials for R^2 = .05 and R^2 = .10 were 65 and 83%, respectively. The prediction potential of both statistical methods supports their use (and the dataset) for making estimates of GM in the determination of foundation thickness (in the building and construction industry) within the GPS polygon, and possibly remotely, without the need for core samples and additional laboratory analyses.

Using the 30 remotely sensed spatially available datasets to determine GM showed inconsistent results (Figure 3). The remote sensing data models that were identified via both the machine learning and statistical modeling approaches do not support the models developed (Dastbaz et al., 2018), even though the same Landsat datasets were used in our models. Our models were less than 50% as accurate, which leads to questions concerning the significant deviation concerning similar models used by others (Arama et al., 2020; Armstrong et al., 2007; Louca et al., 2018; Qu et al., 2018). In our spatial model, the data were not significant enough to warrant the use of the model in its current form with a

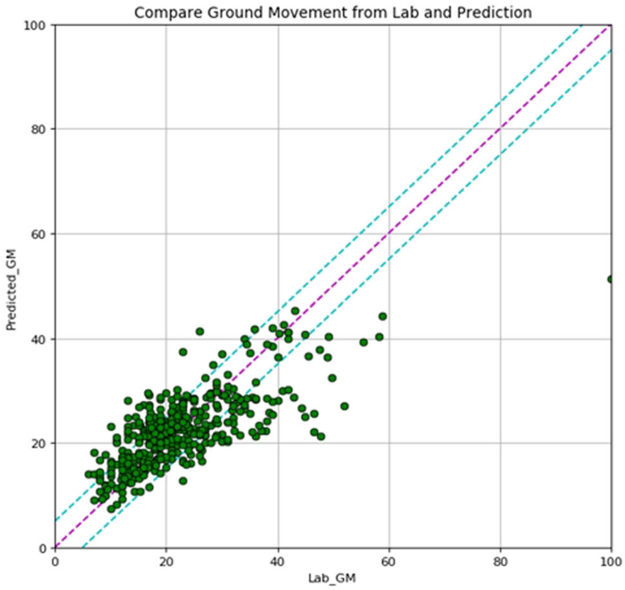

FIGURE 2 Potential for ground movement (GM) in cm from the generalized additive model based on the laboratory dataset (within the GPS polygon Figure 1.) was best described by the "moisture equation" log{E(Yij)} = β0 + β1Ll + β2Pl + β3Year + β4Season + β5GeologicalZone + β6Friabilty + β7USC + β8Colour + β9Moistue + β10P0.425 + β11Latitude + β12Longitude, which explained 91.1% of deviation from expected values ($P < .05$)

high enough level of certainty concerning industry standards. Although, we were able to predict whether the ground would move or not, more data are needed to predict the extent of the movement confidently (without introducing qualitative bias).

Detailed information about built-up areas is valuable for mapping complex urban environments. Although many classification algorithms for surface areas have been developed, they are rarely tested from the perspective of feature engineering and feature learning for predictions of subsurface soil parameters (Sheffield et al., 2012). Better accuracy of the spatial model might be obtained if soil structure (layering) and other variables analyzed in future data collections of Atterberg limits (McBride, 2002) are factored in concerning the geological effects of time. The ground moves as the seasons change; therefore, there is a correlation between this amount of movement and soil type, including the subsurface saturation zone.

However, for a significant correlation to be made, the model data need additional stratigraphic observations about the soil horizon so that two-dimensional soil lithology maps can then be used to piece together a four-dimensional geological time stratigraphic soil map for the creation of larger four-dimensional time-stamped polygons (outside of UTM zone 50). For example, the model gave a better result in classifying movement vs. nonmovement, which might indicate smaller scale influences. Traditionally, seasonality is used in agricultural field modeling where small-scale calculations are improved with field measurements to validate aspects of crop morphology with remotely sensed data (Alexakis et al., 2017; Cai et al., 2019). Significant improvements of the accuracy may thus be obtained when categorical outputs corresponding to site classification (surface water mapping or groundwater bore logs) are built into the Atterberg limit laboratory data collection and geotechnical procedures, irrespective of physical time deviation between measurements. Although, this has yet to be demonstrated as effective concerning standardized data collection protocols (Guo et al., 2021; Tan et al., 2021; Yu et al., 2021).

Remotely sensed big-data sets, in terms of the structural, spectral, and textual features, correspond to the satellites generating the data and is a sophisticated science (Dey et al., 2018; Fong et al., 2018; Peng et al., 2020; Svatos, 2021; Tan et al., 2021). Investigating the character of remotely sensed big-data becomes an essential need to help verify (if available) laboratory datasets for improved model accuracy, especially where data are scarce. These are the first models to predict ground movement in soils older than 200 million yr, a first for Mediterranean climates, a first for Australia, and possibly a first for the Southern Hemisphere (Arama et al., 2020). The model developed from laboratory data is reliable for industry use concerning ground movement within the GPS coordinates of the UTM zone 50 polygon (Figure 1). However, even better accuracy of the spatial model might be obtained if additional information on soil structure (layering) and other variables (that may be sourced) are built into the continually improving automated machine learning pipeline with software improvements (ARM, 2019; Hagendorff, 2020; Oracle, 2020). The future of building and construction is now. The success of this type of model can potentially save tens of thousands of dollars in construction materials, reduce the potential human error associated with data collection and samples manipulation, and save up to 6 wk of building time while ensuring the compliance of the relevant local, state, and federal legislation. In summary, this approach can: (a) allow designers and engineers to prepare better quality and more accurate quotes by providing more information prior to site visits; (b) save time through less maintenance of on-site equipment; (c) provide a potential for less construction wastage, less materials required, and less site works–preparation through cost savings on engineering contingency fees (potential saving of up to 34% of construction materials required); (d) provide less subjective, more consistent estimates; create the possibility for further full data capture and analytics improvements; and (e) reduce down-time and bottlenecks within the overall design building and approval process, which has been amplified by COVID-19.

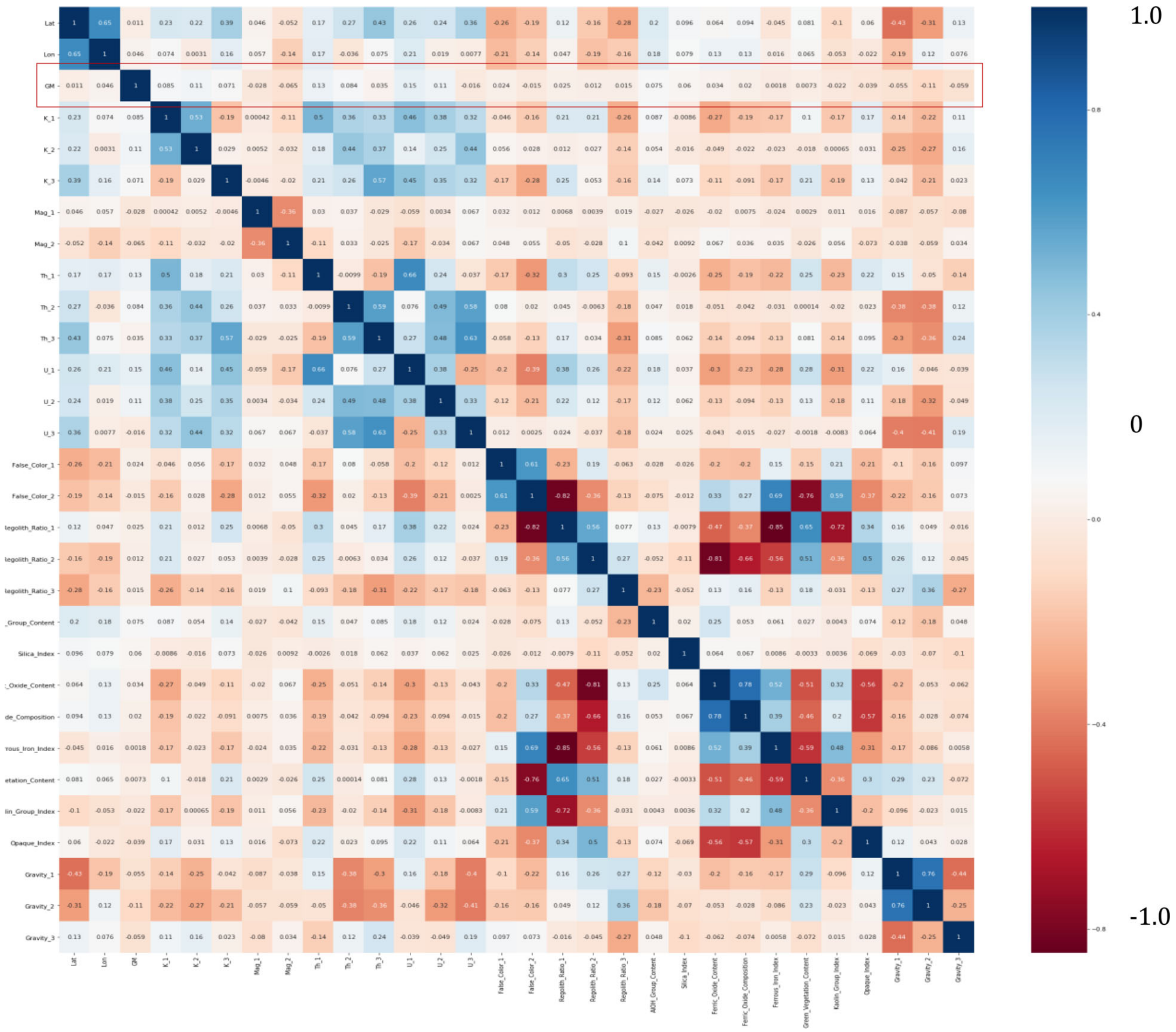

FIGURE 3 The combined heat map for a total of 30 soil physical and chemical properties concerning remotely sensed datasets used (including additional derivates): AlOH_Group_Content; False_Color; Ferric_Oxide_Composition; Ferrous_Iron_Index; GM, ground movement); Green_Vegetation_Content; gravity; K, potassium; Kaolin_Group_Index; Lat, latitude; Lon, longitude; Mag, magnesium; Opaque_Index; Regolith_Ratio; Silica_Index; Th, thorium; U, uranium. Ground movement in cm (Table 1) with red border with its corresponding prediction potential for the Universal Transverse Mercator zone 50 based on the 29 other properties metrics (including outside the GPS polygon from Figure 1)

ACKNOWLEDGMENTS

The author wishes to acknowledge uDrew for their help in supporting the student authors through an internship administered by iPREP Western Australia. Tom Young collected the original datasets and wrote the algorithms concerning Atterberg limits used to formulate the hypothesis. Tom Young and Karl Svatos conceived, from concept to creation, the idea for an automated pipeline for statistical prediction of soil properties (laboratory geophysical soil dataset calculations). Karl Svatos architected the .NET batch assembly for automated machine learning through Amazon, AWS, SaaS, and API. Karl Svatos wrote the manuscript, and handled versioning and editing on request of the journal. Thanks to Dr. Angela Recalde-Salas and Dr. Zubair Ahmed, who conducted critical mathematical modeling based on the requirements of the project during their internship, and to James Adams, and Kris and Kate Zima who contributed by helping link the command-line interface to the Jupyter console, and helped manage the internship team.

AUTHOR CONTRIBUTIONS

Karl B. W. Svatos: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Supervision; Validation; Writing – original draft; Writing – review & editing. Zubair Ahmed: Data curation; Software. Angela Recaldes: Data curation; Software. Tom Young:

Conceptualization; Data curation; Funding acquisition; Software; Supervision; Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Karl B. W. Svatos https://orcid.org/0000-0003-1468-0601
Angela Recalde-Salas https://orcid.org/0000-0003-3383-8902

REFERENCES

Ahmed, N., Mahmud, S., Elahi, M. M. L., Ahmeda, S., & Sujauddin, M. (2019). Forecasting river sediment deposition through satellite image driven unsupervised machine learning techniques. *Remote Sensing Applications: Society and Environment*, *13*, 435–444. https://doi.org/10.1016/j.rsase.2018.12.011

Alexakis, D. D., Mexis, F. D. K., Vozinaki, A. E. K., Daliakopoulos, I. N., & Tsanis, I. K. (2017). Soil moisture content estimation based on Sentinel-1 and auxiliary earth observation products. A hydrological approach. *Sensors*, *17*,(6), 1455. https://doi.org/10.3390/s17061455

Arama, Z. A., Gençdal, H. B., Nuray, S. E., & Yücel, M. (2020). The applicability of regression analysis and artificial neural networks to the prediction process of consistency and compaction properties of high plastic clays. *International Conference on Harmony Search Algorithm*, pp. 295–305. Springer. https://doi.org/10.1007/978-981-15-8603-3_26

ARM. (2019). Arm AI trust manifesto. https://www.arm.com/blogs/blueprint/wp-content/uploads/2019/11/Arm-AI-Trust-Manifesto-2019.pdf

Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F., & Dwivedi, G. (2021). Imputation of missing data with class imbalance using conditional generative adversarial networks. *Neurocomputing*, *453*, 164–171. https://doi.org/10.1016/j.neucom.2021.04.010

Armstrong, L., Diepeveen, D., & Maddern, R. (2007). The application of data mining techniques to characterize agricultural soil profiles. In P. Christen, P. Kennedy, J., Li, I. Kolyshkina, & G. Williams (Eds.), *Proceedings of Australasian Data Mining Conference* (pp. 85–99). (AusDM'07). Australian Computer Society. https://dl.acm.org/doi/proceedings/10.5555/1378245

Cai, Y., Guan, K., Lobell, D., Potieter, A. B., Wang, S., Peng, J., Asseng, S., Zhang, Y., You, L., & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology*, *274*, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010

Cameron, D. A. (1989). Test for reactivity and prediction of ground movement. *Australian Civil Engineering Transactions*, *CE31*, 121–132. http://hdl.handle.net/102.100.100/261334?index=1

Dastbaz, M., Arabnia, H., & Akhgar, B. (2018). *Technology for smart futures*. Springer. https://doi.org/10.1007/978-3-319-60137-3

Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A. S., & Satapathy, S. C. (2018). *Internet of things and big data analytics toward next-generation intelligence*. Springer. https://doi.org/10.1007/978-3-319-60435-0

Fong, S., Song, W., Wong, R., Bhatt, C., & Korzun, D. (2018). Framework of temporal data stream mining by using incrementally optimized Very Fast Decision Forest. In N. Dey, A. Hassanien, C. Bhatt, A. Ashour, & S. Satapathy (Eds.), *Internet of things and big data analytics toward next-generation intelligence*. Springer. https://doi.org/10.1007/978-3-319-60435-0_19

Guo, Z., Yu, K., Jolfaei, A., Bashir, A. K., Almagrabi, A. O., & Kumar, N. (2021). Fuzzy detection system for rumors through explainable adaptive learning. *IEEE Transactions on Fuzzy Systems*, *29*, 3650–3664. https://doi.org/10.1109/TFUZZ.2021.3052109

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds & Machines*, *30*, 99–120. https://doi.org/10.1007/s11023-020-09517-8

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez Ruiperez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*, e0169748. https://doi.org/10.1371/journal.pone.0169748

Huang, F., Yu, Y., & Feng, T. (2019). Automatic extraction of impervious surfaces from high resolution remote sensing images based on deep learning. *Journal of Visual Communication and Image Representation*, *58*, 45361. https://doi.org/10.1016/j.jvcir.2018.11.041

Johnson, T., Cameron, D., Moore, G., & Brien, C. (2020). Ground movement in a moderately expansive soil subject to rainfall infiltration through pervious paving. *Ecological Engineering*, *158*, 106022. https://doi.org/10.1016/j.ecoleng.2020.106022

Jones, G. A. (2017). An empirical preliminary prediction of heave (technical note). *Journal of the South African Institution of Civil Engineering*, *59*, 4. https://doi.org/10.17159/2309-8775/2017/v59n4a7

Karmakar, B., Das, S., Bhattacharya, S., Sakar, R., & Mukhopadhyay, I. (2019). Tight clustering for large datasets with an application to gene expression data. *Scientific Reports*, *9*, 3053. https://doi.org/10.1038/s41598-019-39459-w

Keight, R., Al-Jumeily, D., Hussain, A., Fergus, P., & Mustafina, J. (2018). Big data and data science applications for independent and healthy living. In M. Dastbaz, H. Arabnia, & B. Akhgar (Eds.), *Technology for smart futures* (pp. 77–111). Springer. https://doi.org/10.1007/978-3-319-60137-3_5

Louca, S., Doebeli, M., & Parfrey, L. W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome*, *6*, 41. https://doi.org/10.1186/s40168-018-0420-9

McArthur, W. M. (2004). *Reference soils of south-western Australia*. Department of Agriculture and Food. http://www.researchlibrary.agric.wa.gov.au/books/4/

McBride, R. A. (2002). 2.9 Atterberg limits. In J. H. Dane & G. C. Topp (Eds.), *Methods of Soil Analysis: Part 4. Physical Methods 5.4* (pp. 389–398). SSSA. https://doi.org/10.2136/sssabookser5.4

New York Department of Transportation. (2015). *Test method for liquid limit, plastic limit, and plasticity index, geotechnical test method GTM-7*. Geotechnical Engineering Bureau, New York State Department of Transportation. https://www.dot.ny.gov/divisions/engineering/technical-services/technical-services-repository/GTM-7b.pdf

Osman, N. Y. (2007). *The development of a predictive damage condition model of light structures on expansive soils using hybrid artificial intelligence techniques*. [Doctoral dissertation, Swinburne University of Technology]. http://hdl.handle.net/1959.3/27968

Oracle. (2020). *Oracle® advanced support gateway security guide*. Oracle Corporation. http://docs.oracle.com

Peng, B., Guan, K., Zhou, W., Jiang, W. C., Frankenberg, C., He, L., & Koeler, P. (2020). Assessing the benefit of satellite-based solar-induced chlorophyll fluorescence in crop yield prediction. *International Journal of Applied Earth Observation and Geoinformation*, *90*, 102126. https://doi.org/10.1016/j.jag.2020.102126

Qu, Y., Qian, X., Song, H., Xing, Y., Li, Z., & Tan, J. (2018). Soil moisture investigation utilizing machine learning approach based experimental data and Landsat5-TM images: A case study in the mega city Beijing. *Water*, *10*, 423. https://doi.org/10.3390/w10040423

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://R-project.org/

Recaldes, A., Erbe, C., Salgado-Kent, C., & Parsons, M. (2020). Non-song vocalizations of humpback whales in Western Australia. *Frontiers in Marine Science*, *7*, 141. https://doi.org/10.3389/fmars.2020.00141

Rees, S. W., & Thomas, H. R. (1993). Simulating seasonal ground movement in unsaturated clay. *Journal of Geotechnical Engineering*, *119*, 7. https://doi.org/10.1061/(ASCE)0733-9410(1993)119:7(1127)

Reynolds, D., Ball, J., Bauer, A., Davey, R., Griffiths, S., & Zhou, J. (2019). CropSight: A scalable and open-source information management system for distributed plant phenotyping and IoT-based crop management. *Gigascience*, 8. https://doi.org/10.1093/gigascience/giz009

Sheffield, J., Wood, E. F., & Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature*, *491*, 435–438. https://doi.org/10.1038/nature11575

Standards Australia. (2017). *Geotechnical site investigations 2017 update*. ISBN: 0726278785

Sutherland, G., Chasmer, L. E., Petrone, R. M., Kljun, N., & Devito, K. J. (2014). Evaluating the use of spatially varying versus bulk average 3D vegetation structural inputs to modelled evapotranspiration within heterogeneous land cover types. *Ecohydrology*, *7*, 1545–1549. https://doi.org/10.1002/eco.1477

Svatos, K. B. W. (2018). Commercial silicate phosphate sequestration and desorption leads to a gradual decline of aquatic systems. *Environmental Science and Pollution Research*, *25*, 5386–5392. https://doi.org/10.1007/s11356-017-0846-9

Svatos, K. B. W. (2021). *Data-driven, Mediterranean, A.I. simulations and the ethical implications of embedded qualitative bias in digital twin deepfake games*. [Doctoral thesis, Murdoch University]. https://researchrepository.murdoch.edu.au/id/eprint/63264/

Tan, L., Yu, K., Bashir, A. K., Cheng, X., Ming, F., Zhao, L., & Zhou, X. (2021). Toward real-time and efficient cardiovascular monitoring for COVID-19 patients by 5G-enabled wearable medical devices: A deep learning approach. *Neural Computing and Applications*, 1–14. https://doi.org/10.1007/s00521-021-06219-9

Van der Merwe, D. H. (1964). The prediction of heave from the plasticity index and percentage clay fraction of soils. *South Africa Institute Civil Engineers*, *6*, 103–116. https://doi.org/10.10520/AJA10212019_21190

Vorwerk, S., Cameron, D., & Keppel, G. (2015) Clay soil in suburban environments: Movement and stabilization through vegetation. In B. Indraratna, J. Chu, & C. Rujikiatkamjorn (Eds.), *Ground improvement case histories* (pp. 655–682). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-08-100191-2.00022-8

Yu, K., Guo, Z., Shen, Y., Wang, W., Lin, J. C., & Sato, T. (2021). Secure artificial intelligence of things for implicit group recommendations. *IEEE Internet of Things Journal*, https://doi.org/10.1109/JIOT.2021.3079574

How to cite this article: Svatos, K. B. W., Ahmed, Z., Recalde-Salas, A., & Young, T. (2022). Heuristics-enhanced geospatial machine learning (SaaS) of an ancient Mediterranean environment. *Soil Science Society of America Journal*, 1–8. https://doi.org/10.1002/saj2.20385