



Type

Research brief

Area

Kapiti Research Station, Kenya

Partners

The International Livestock Research Institute (ILRI), Lancaster University, Technical University of Munich, Scotland's Rural College, CGIAR Initiative on Livestock and Climate.

Machine Learning-Based Gridded Soil Mapping for Kapiti Research Station and Wildlife Conservancy, Kenya

Research objective

To develop a high-resolution digital soil map for Kapiti Wildlife Conservancy, using advanced geospatial technologies and machine learning techniques. This study aims to provide detailed, spatially explicit information on soil properties, enabling improved pasture management, forage production, climate resilience and sustainable land use practices in the conservancy.

Authors

Ambica Paliwal¹, Fredah Cherotich¹, Sonja Leitner¹, Fiona Pearce², Mariana Rufino³, John Quinton², Ram Dhulipala¹, Mazdak Salavati⁴, Ilona Gluecks¹ and Anthony Whitbread¹

¹ILRI, ²Lancaster University, ³Technical University of Munich, ⁴Scotland's Rural College



INITIATIVE ON
Livestock and Climate

ILRI



Background

Soil information is fundamental for the effective management of pastures and the maintenance of high forage quality, as it directly influences the health and productivity of plant species that livestock depend on for nourishment (Reid et al., 2004; Kumar and Jhariya 2015; Karaca et al 2021). Soil properties, such as texture, nutrient content, moisture retention, and organic matter, determine the capacity of pastures to sustain forage growth and support diverse, nutrient-rich vegetation. Healthy soils ensure that the forage available to livestock is not only abundant but also high in the nutrients necessary for livestock growth, reproduction, and milk production. Without a clear understanding of soil conditions, pasture degradation through overgrazing, soil erosion, and nutrient depletion can occur, leading to reduced forage quality and diminished livestock productivity. Additionally, soil information is critical for strategic decisions on grazing patterns, pasture rotation, and the application of fertilizers or soil amendments, all of which are vital for improving forage resilience in the face of climate variability and other environmental stressors.

In this context, digital soil mapping (DSM) has emerged as an invaluable tool for providing detailed, spatially explicit information on soil properties across large and often heterogeneous landscapes (McBratney et al 2003). Unlike traditional methods of soil surveying, which are often labor-intensive, costly, and time-consuming, DSM leverages modern technologies such as satellite remote sensing,

machine learning algorithms, and geostatistical models to create high-resolution soil maps. These maps allow land managers, farmers, and policymakers to identify key soil characteristics, such as areas of nutrient deficiency or susceptibility to erosion, that require specific management practices. By integrating DSM into pasture and rangeland management, stakeholders can implement targeted interventions that improve soil health, enhance forage quality, and optimize livestock production. Furthermore, DSM facilitates the monitoring of soil changes over time, ensuring that land degradation is addressed promptly and that sustainable practices can be implemented to restore and preserve soil ecosystems. In the long run, digital soil mapping plays a crucial role in supporting resilient rangeland ecosystems, promoting sustainable livestock production, and mitigating the impacts of climate change on pasture-based systems.

The first version of Global SoilGrids, introduced in 2014 with a 1 km spatial resolution (Hengl et al 2014), was initially a 'proof of concept.' It demonstrated that global soil profile compilations could be employed in an automated system to generate complete and consistent spatial predictions of soil properties and classes. However, since its release, various colleagues have recognized and pointed out some of the limitations associated with this early version. This study attempts to provide soil grids on 30 m spatial resolution using soil data.

Methodology

Study Area

The study was conducted at the Kapiti Research Station and Wildlife Conservancy of the International Livestock Research Institute (ILRI) located in the semi-arid savanna region (1°35.8'W-1°40.9'W, 36°6.4'E-37°10.3'E) of the Athi-Kapiti ecosystem in Kenya. The Kapiti ranch is in Machakos County, southern Kenya, covering an area of 13,000 ha (Figure 1a). The climate is typical for semi-arid savannas, with precipitation below potential evapotranspiration. The mean annual precipitation is 550 mm, with rainfall being distributed in a bimodal precipitation regime (March-May and November-December). Approx. 80% of the annual precipitation occurs during these two periods. The mean annual temperature is 20.2° C, with 4° C of annual variation and substantial day and night variability. The soils in the study area are Salid Sodic Pellic Vertisols (Magnesic), locally known as “black cotton soils” (Leitner et al., 2024).

Field Data Collection

The evaluation soil survey was carried out using high-res Google Earth images and in situ reconnaissance sampling the soil with a soil auger was conducted to delineate the soil mapping units. Such type of soil survey is carried out for preliminary investigations of soil conditions of a given area. The data on soil parameters was collected from 14 geo-coordinates at 4 different depth levels - 0-18 cm, 18-41cm, 41-74cm and 74-150cm in the month of January and May 2023 according to the FAO ‘Guidelines for soil description’ (FAO, 2006, 87pg). The soils were analysed for texture, total carbon (C%), total nitrogen (N%), pH, electrical conductivity (EC), exchangeable potassium (K), calcium (Ca), magnesium (Mg), and sodium (Na). From each soil profile we further generated 40 more soil data points for our analysis (Figure 1b). We did the exploratory analysis of each soil parameter by creating separate histograms for each parameter.

Satellite data/ covariates ML Approach

The soil co-variates/predictors used in the study is primarily based on satellite remote sensing data. Land use land cover (ESA 2021), monthly maximum and mean temperatures of past three years (2020-2022) (Abatzoglou et al., 2018), mean monthly rainfall of past three years (Abatzoglou et al., 2018), average of monthly evapotranspiration (ET) of past three years (Abatzoglou et al., 2018), mean monthly Normalized Difference Vegetation Index (NDVI) of past three years generated using Sentinel 2 with 10m spatial resolution (ESA 2021), slope and elevation (SRTM 2013).

All the processing and analysis was conducted using R packages (raster, sp, gdal, ggplot, terra, caret and randomForest). All the covariates were brought to 30m spatial resolution using cubic convolution method. We divided the soil points into testing and training datasets (70:30). We used training data to train the model and remaining test data to predict the accuracy. Specifically, we overlaid the training points over the covariate stack prepared the regression matrix and fitted the spatial prediction model. We identified the highly correlated variables and removed it from our analysis and conducted variable inflation factor analysis to measure the amount of collinearity (VIF>5) for remaining variables. We then predicted values for soil parameters using random forest models and created variable importance plots for all the soil parameters. Furthermore, though random forest is robust to overfitting, we ensured that our models did not overfit the data by running a fivefold cross validation analysis where we used 70% of the data for training and 30% of the data for testing.

We validated our satellite estimates by comparing estimated values after the two-step approach with the observed soil values at the field scale for each soil parameter. Accuracy was evaluated using R2 which is a common approach in satellite data estimation.

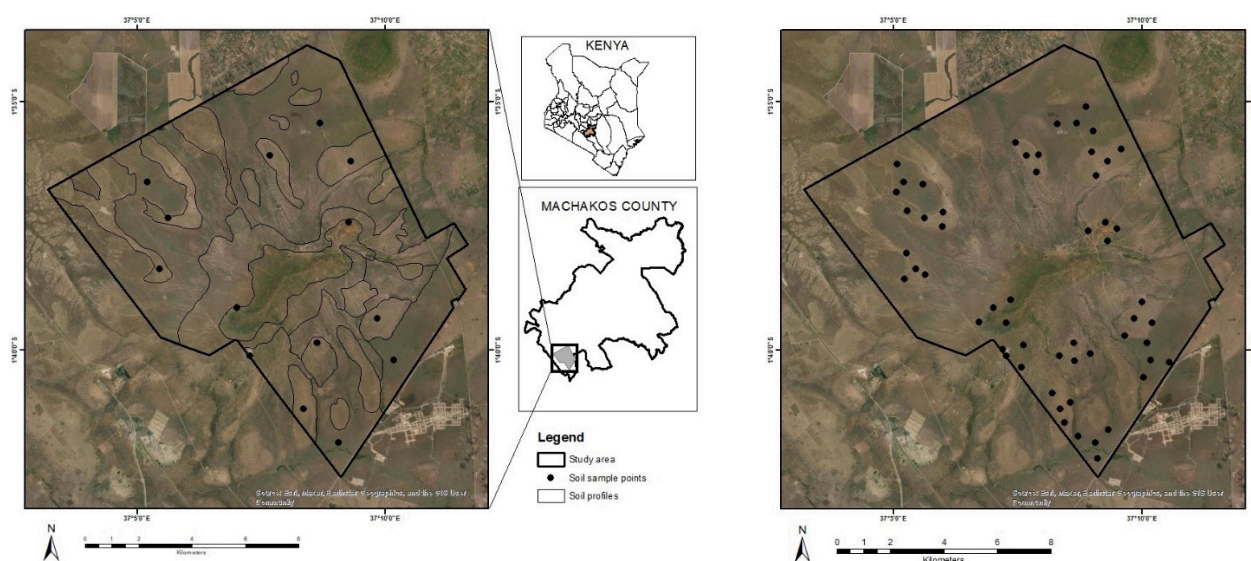


Figure 1. (a) Study Area Map of Kapiti Ranch along with data points from which samples were collected; **(b)** Data points that were generated from different profiles.

Key findings

- Exploratory analysis for each soil variable provided the range of values that exists for each soil parameter in the area (Figure 2). The histograms display the distribution of key soil properties such as carbon content (C%), nitrogen content (N%), pH, electrical conductivity (EC), and exchangeable cations (K, Ca, Mg, Na) across the sampled locations.

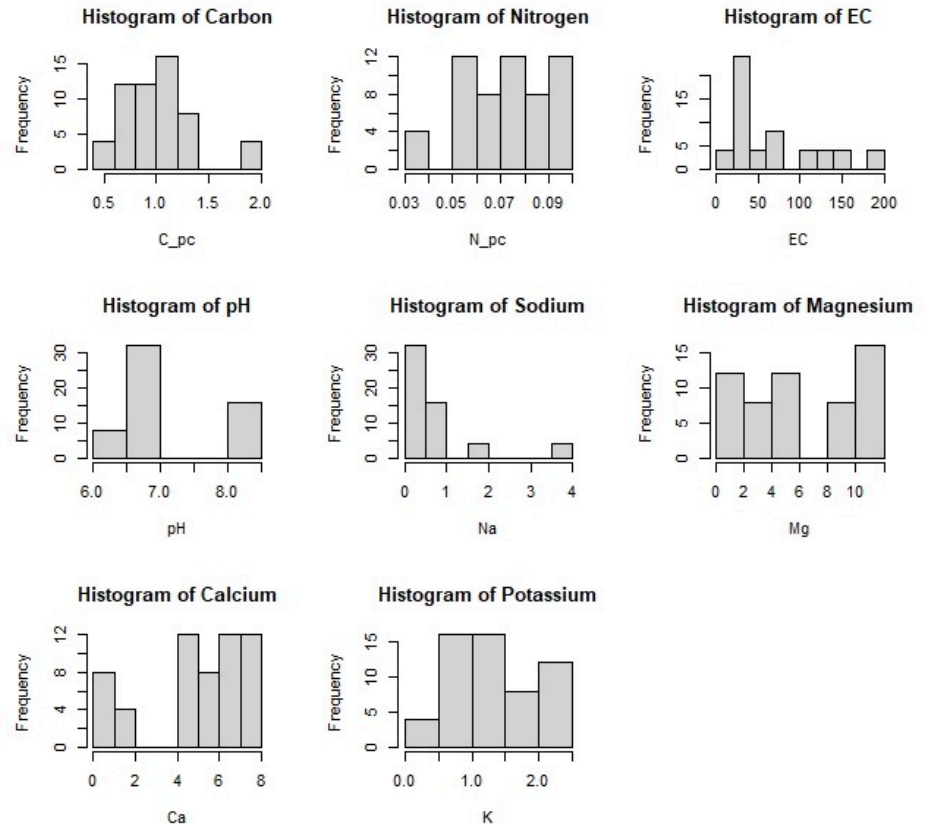


Figure 2. Histograms of soil parameters.

- We validated our predicted values for all the soil parameters by plotting it against the observed data (testing data that we did not use in the analysis). Scatter plots comparing observed soil parameters with values predicted by the machine learning model (Figure 3). The plots illustrate the model's accuracy in predicting soil properties from the satellite-derived covariates.

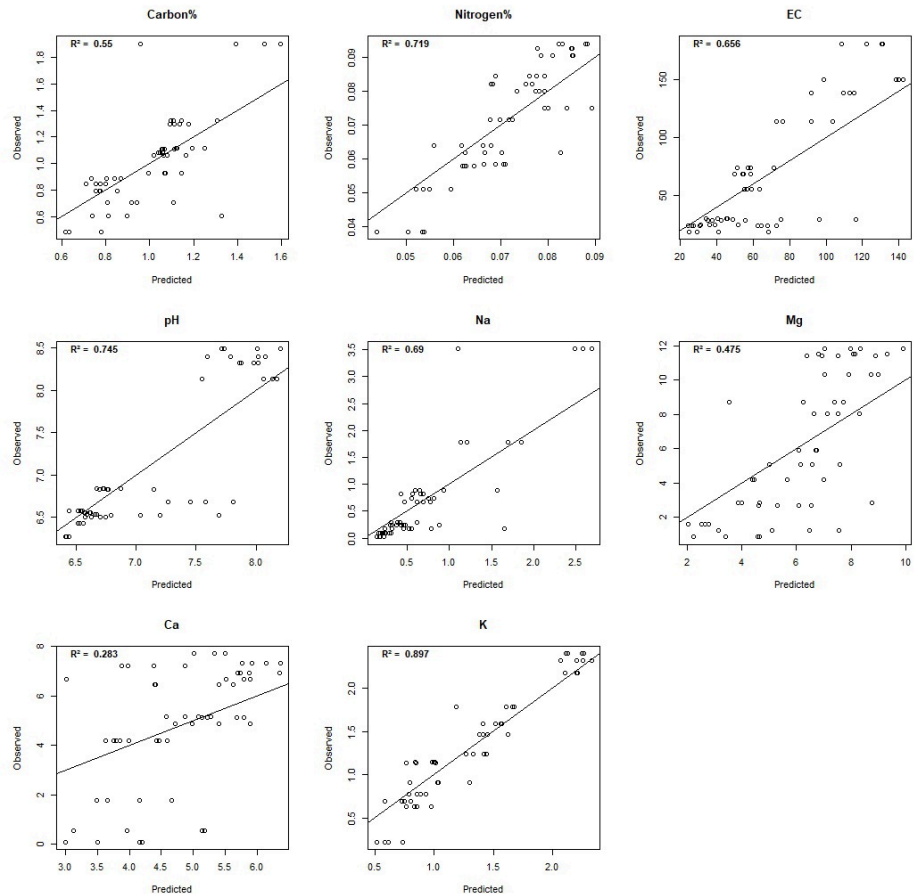


Figure 3. Observed vs. predicted soil parameter values.

Key findings (cont).

3. Figure 4 shows variable importance plots derived from random forest models showing the top five predictor variables for each soil parameter. The plots highlight the key satellite-based covariates, such as land cover, NDVI, and climate data, that influence soil property predictions.

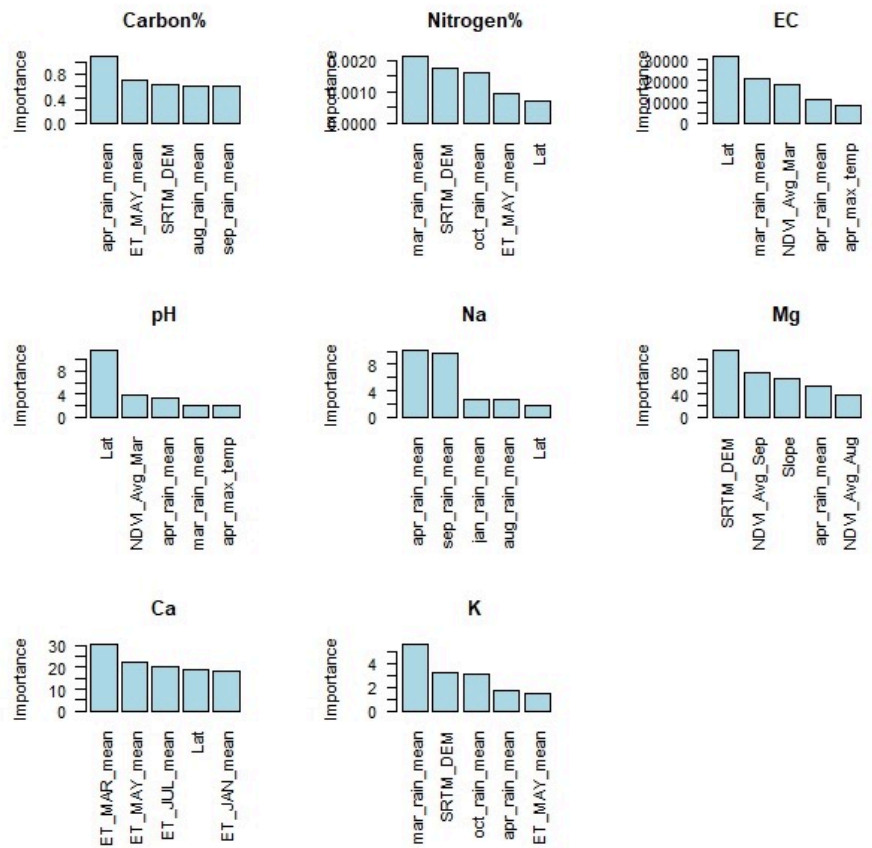


Figure 4. Variable Importance for Soil Parameter Prediction, showing top 5 predictor variables.

4. As a result, we generated predicted maps for soil parameters across the Kapiti Wildlife Conservancy, generated using random forest models (Figure 5).

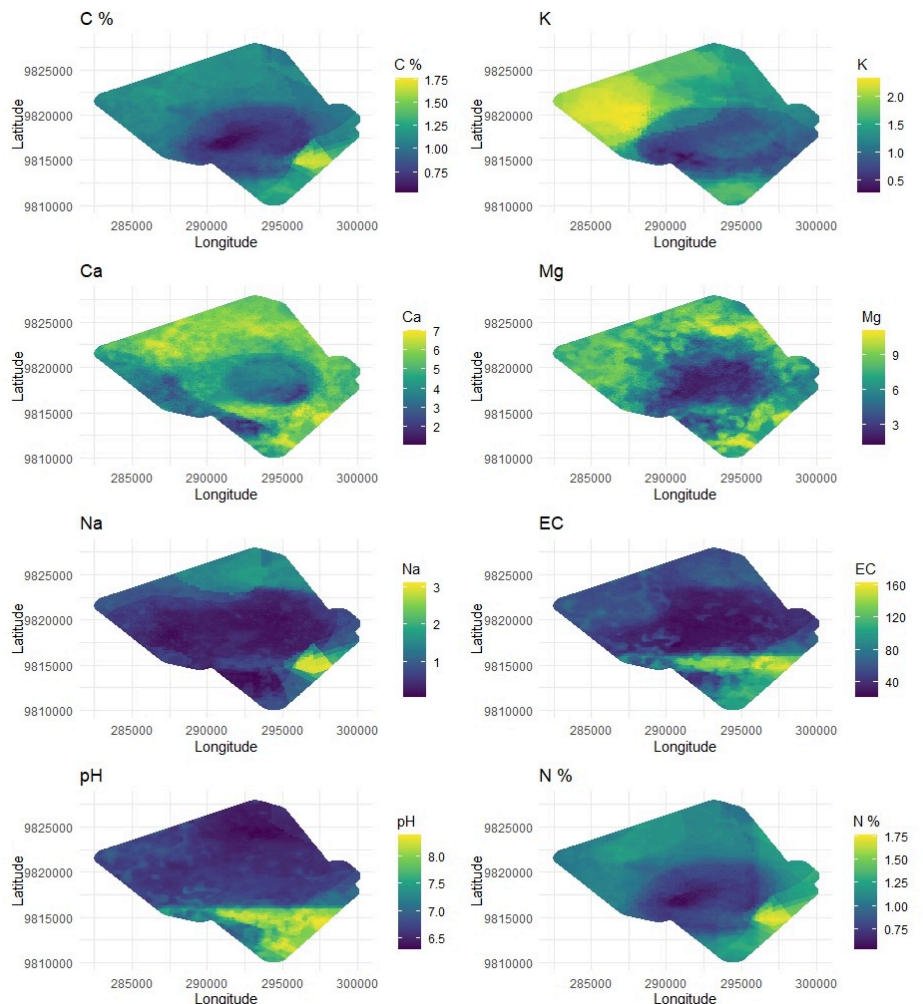


Figure 5. Predicted maps for soil parameters across the Kapiti Ranch, generated using random forest models.

Conclusion

The application of machine learning for digital soil mapping at Kapiti Research Station and Wildlife Conservancy demonstrates the potential for enhanced precision and efficiency in soil analysis. By integrating satellite remote sensing data with advanced machine learning algorithms, this approach enables the generation of high-resolution maps that provide spatially explicit information on critical soil properties. This method significantly improves upon traditional soil surveying techniques, which are often labor-intensive and costly, by offering a faster, scalable solution suitable for large and heterogeneous landscapes. The insights gained from these digital soil maps allow for more informed decisions regarding pasture management and grazing patterns, which are essential for maintaining healthy rangeland ecosystems.

Furthermore, the predictive power of machine learning models enhances the ability to monitor soil conditions over time, identify degradation risks, and implement timely interventions to prevent further land degradation. This capability is crucial for promoting climate-resilient land use practices and optimizing livestock productivity in the face of environmental variability. Overall, the adoption of digital soil mapping through machine learning offers a sustainable and innovative approach to soil and pasture management, supporting long-term conservation and productivity goals at Kapiti.

References

- Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a High-Resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* 5, 170191 (2018). <https://doi.org/10.1038/sdata.2017.191>
- European Space Agency (ESA). ESA WorldCover 10m 2021 v200. 2021. Available online: <https://worldcover2021.esa.int> (accessed on July 24, 2024).
- Karaca, S.; Dengiz, O.; Demirag, L.; Ozkan B.; Dedeoglu, M.; Gulser, F.; Sargin, B.; Demirkaya, S.; Ay, A. An assessment of pasture soils quality based on multiindicator weighting approaches in semi-arid ecosystem, *Ecol. Ind.*, 121 (2021), Article 107001, Volume 121, 2021, 107001, <https://doi.org/10.1016/j.ecolind.2020.107001>.
- Kumar, T and Jhariya, D.C. Land quality index assessment for agricultural purpose using multicriteria decision analysis (MCDA) *Geocarto Int.*, 30 (7) (2015), pp. 822–41, [10.1080/10106049.2014.997304](https://doi.org/10.1080/10106049.2014.997304).
- Hengl, T. ; de Jesus J.M. ; MacMillan, R.A. ; Batjes, N.H.; Heuvelink, G.B.M.; Ribeiro, E.; et al. (2014) SoilGrids1km – Global Soil Information Based on Automated Mapping. *PLoS ONE* 9(8): e105992. <https://doi.org/10.1371/journal.pone.0105992>
- Leitner, S.M. ; Carbonell, V. ; Mhindu, R.L. ; Zhu, Y.; Mutuo, P.; Butterbach-Bahl, K.; Merbold, L. Greenhouse gas emissions from cattle enclosures in semi-arid sub-Saharan Africa: the case of a rangeland in South-Central Kenya. *Agriculture, Ecosystems & Environment*. 367 (2024), <https://doi.org/10.1016/j.agee.2024.108980>.
- McBratney A.B.; Santos M.L.M.; Minasny B. On digital soil mapping. *Geoderma*, 117 (1–2) (2003), pp. 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
- Reid, R.S.; Thornton, P.K.; McCrabb, G.J.; Kruska, R.L. ; Atieno, F. ; Jones, P.G. Is it possible to mitigate greenhouse gas emissions in pastoral ecosystems of the tropics? *Environ. Dev. Sustain.*, 6 (2004), pp. 91–109. <https://doi.org/10.1023/B:ENVI.0000003631.43271.6b>.
- SRTM, NASA JPL Shuttle Radar Topography Mission (SRTM) Version 3.0. 2013. Available online: <https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003> (accessed on July 24, 2024).



INITIATIVE ON
Digital Innovation

Research-based evidence and solutions for digital innovations to accelerate transformation of agrifood systems, with an emphasis on inclusivity and sustainability.

More information: on.cgiar.org/digital

Acknowledgements

The authors would like to acknowledge the funding received for this activity from the CGIAR Initiative on Digital Innovation. The research brief activity describing digital soil mapping is part of the Digital Innovation project to develop a digital twin for the Kapiti research facility. The authors would also like to acknowledge the support of the staff and management of Kapiti, and the support from the Livestock and Climate Initiative.

This publication has been prepared as an output of the CGIAR Initiative on Digital Innovation, which researches pathways to accelerate the transformation towards sustainable and inclusive agrifood systems by generating research-based evidence and innovative digital solutions. This publication has not been independently peer reviewed. Responsibility for editing, proofreading, and layout, opinions expressed, and any possible errors lies with the authors and not the institutions involved. The boundaries and names shown and the designations used on maps do not imply official endorsement or acceptance by the International Livestock Research Institute (ILRI), CGIAR, our partner institutions, or donors. In line with principles defined in the CGIAR Open and FAIR Data Assets Policy, this publication is available under a CC BY 4.0 license. © The copyright of this publication is held by ILRI. We thank all funders who supported this research through their contributions to the CGIAR Trust Fund.