Data in Brief 33 (2020) 106342

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

The first version of nation-wide open 3D soil database for Sri Lanka

Eranga M. Wimalasiri[a,*], Ebrahim Jahanshiri[a,*], T.A.S.T.M. Suhairi[a], Ranjith B. Mapa[b], Asha S. Karunaratne[c], Lal P. Vidhanarachchi[c], Hasika Udayangani[c], N.M.M. Nizar[a], Sayed N. Azam-Ali[a]

[a] *Crops for the Future UK, Chelmsford, Essex CM2 7PJ, England, UK*
[b] *Department of Soil Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka*
[c] *Department of Export Agriculture, Faculty of Agricultural Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka*

A R T I C L E I N F O

Article history:
Received 27 August 2020
Revised 9 September 2020
Accepted 18 September 2020
Available online 24 September 2020

Keywords:
Digital soil mapping
Process-based crop models
SRICANSOL
Tropical soils, legacy soil data

A B S T R A C T

Soil data for Sri Lanka are available through semi-detailed series maps that were developed based on limited soil profile data combined with expert knowledge. This data plays a vital role in decisions at national and regional levels. However, the present format of this database does not allow for their wider use in crop simulation modelling and other related agricultural research that require finer scale data. This is due to the fact that cross-country profile data are not harmonised based on standard depths. Several attempts were made to produce digital soil data for Sri Lanka at different geographic scales, however, a completely harmonised data that covers variability across depths and properties is yet to be made available. In this article, we describe the first version of the open digital soil database that was developed using a database of 122 locations across the country. Soil properties were harmonised for standard depths using equal-area quadratic smoothing splines. Out of several interpolation methods that were evaluated for univariate interpolation, maps which were produced with the least overall error (RMSE) in the process of cross-validation were selected. The newly developed digital soil database contains 9 soil

* Corresponding authors.
E-mail addresses: eranga.wimalasiri@cropsforthefutureuk.org (E.M. Wimalasiri), e.jahan@cropsforthefutureuk.org (E. Jahanshiri).

https://doi.org/10.1016/j.dib.2020.106342

properties; pH, bulk density, cation exchange capacity, organic carbon, volumetric moisture content at 0.33 and 15 bars levels, sand silt and clay content. Moreover, the data are available for five standard depth layers as 0–5, 5–15, 15–30, 30–60 and 60–100 cm in raster format at 1 km spatial resolution. Both interpolated property maps and their error maps were stored in an open repository and made available for public use. The first version of all maps is also showcased online through open web mapping services. The repository will be gradually updated with higher resolution and more accurate maps as more samples become available and better interpolation method are used. This data could provide complementary information for insight generation at finer scales where limited local informaiton about soils hinders agricultural development.

Specifications Table

Subject	Environmental Science (General)
Specific subject area	Digital Soil Mapping, Soil data harmonisation
Type of data	Image Digital maps (quantitative), Metadata (Attributes)
How data were acquired	Soil depths were harmonised using thin equal-area quadratic smoothing splines. The SplineTool v2 (available at: http:// www.asris.csiro.au) was used. The data were analysed using ESRI ArcGIS 10.7.1 Geostatistical Analyst to create interpolation of the soil observed data; GIS Spatial Analyst extension was used to create the raster according to the area of interest. Instruments: Desktop computer with Intel ® Core (TM)i7-4600U CPU@2.10 GHz with 16GB RAM
Data format	Raw Raster images (GeoTIFF format)
Parameters for data collection	Observed raw soil data of Sri Lanka published by Mapa et al. [1–4] were used. Both soil physical and chemical properties (pH, bulk density, cation exchange capacity, organic carbon, volumetric moisture content (VMC) at 0.33 bars level (drainage upper limit), VMC at 15 bars level (wilting point), clay, sand and silt) were collected for all available horizons.
Description of data collection	The observed dataset contains 9 properties for 122 locations across Sri Lanka. The data covered different horizons, therefore, properties were harmonised for standard depths (0–5, 5–15, 15–30, 30–60 and 60–100 cm) before interpolations. Soil analysis procedures were previously described [3]. The observed data were interpolated using various techniques to create a continuous map of all properties. Different interpolation methods were used to create predictive maps and error map for soil properties. The maps were created using high resolution raster images with 600 dpi and projected using WGS 1984 datum to make them more suitable.
Data source location	Institution: Crops For The Future Research Centre City/Town/Region: Semenyih, Selangor Country: Malaysia Primary data sources: Observed soil data by Mapa et al. [1–4] were used as primary data.

(*continued on next page*)

Data accessibility	Only mapping figures in a template for 5 depths are available with the paper. The digital data are available at Mendeley Data. DOI: 10.17632/5sc7njfcyn.1 URL: https://data.mendeley.com/datasets/5sc7njfcyn/1 Instructions for accessing these data: The data can be visualised using any GIS software online and offline. Web portal Soil property data are visualized using QGIS Cloud and ArcGIS online. Link to the online web maps: URL: https://cropbase.co.uk/src/srism/
Related research article	E.M. Wimalasiri, E. Jahanshiri, T.A.S.T.M. Suhairi, H. Udayangani, R.B. Mapa, A.S. Karunaratne, L.P. Vidhanarachchi, S.N. Azam-Ali, Basic Soil Data Requirements for Reliable Crop Yield Simulations in Process-Based Crop Models as a basis for Crop Diversification, Sustainability, 12 (2020) 7781. https://doi.org/10.3390/su12187781

Value of the Data

- Current Soil data of Sri Lanka are primarily comprised of the distribution of units. This format does not allow for its wider use in process-based crops models. Since the national level digital soil database is not available, this dataset provides the first digital soil data at the national level at 1 km spatial resolution. The maps are created based on soil properties rather than soil series which makes it easier to extract exact information at any desired location.
- As an agricultural country with around 50% of agricultural lands, soil knowledge plays a vital role in agricultural and related research in Sri Lanka. Any decision at local, regional and national scale regarding agricultural development requires accurate information about soils. The data for 9 soil parameters is made available for different depths which allows them to be used in process-based crop models. The data enables understanding the crop performance in areas where soil profile data are not available.
- The data is provided at all locations within Sri Lanka. This allows virtually any yield simulation to be performed at any desired location using calibrated models. For example, Wimalasiri et al. [5] has used the same dataset to create a continuous yield map of rice across the whole country. The data can be used to fill gaps in observed data which are normally scarce.
- With the growing concern on climate change and the prevalence of climate adaptation studies, the data can be used in process-based crop models to create regional level insights. The dataset contains basic soil parameters that are required by the Agricultural Production System Simulator (APSIM) and potential other crop models.
- This dataset will also provide a primer for yield simulation of neglected and underutilised crops that can play a major role in food security and nutrition of a growing population.
- As this data is available publicly, many other possibilities with this data could emerge in the future. Potential uses for this data could be for modelling soil processes, urban development and hydrology.

1. Data Description

This dataset contains interpolated maps for 9 soil properties; pH, bulk density (gcm^{-3}), cation exchange capacity (cmol + kg^{-1}), organic carbon content (%), volumetric moisture content (VMC) at 0.33 bars level (drainage upper limit – mm mm^{-1}), VMC at 15 bars level (wilting point – mm mm^{-1}), clay (%), sand (%) and silt content (%). The data are available at 5 different depths as

0–5 cm, 5 15 cm, 15–30 cm, 30 60 cm and 60–100 cm. The maps were created using various interpolation methods that closely mimic the spatial variability. The predicted maps were created at 5 different depths for each soil attribute at 1 km spatial resolution.

The dataset contains 45 individual soil property maps (9 properties x 5 depths) and 30 standard error maps. It should be noted that the figure templates for all the depths per soil property shown in this paper (Figs. 1–9) are for illustration purposes only. Individual maps are available at the open repository (https://data.mendeley.com/datasets/5sc7njfcyn/1). The database is mainly divided into two components as soil property maps and standard error maps. Under each category, individual soil properties are available in different folders which were renamed according to the respective soil property. Individual soil properties were further divided into sub-folders based on the depth and renamed accordingly. Therefore, a single folder contains image files in GEOTIFF format along with geographic information systems supported files (TFW, XML and OVR) of one soil property for the respective depth.

In this article, Figs. 1–4 shows pH, organic carbon content, bulk density and cation exchange capacity along with their error maps for all the standard depths. The (VMC) at 0.33 bars level which is also known as drainage upper limit and VMC at 15 bars level (wilting point) are shown in Figs. 5–6. Soil textural properties are shown in Fig. 7 (clay content), Fig. 8 (sand content) and Fig 9 (silt content).

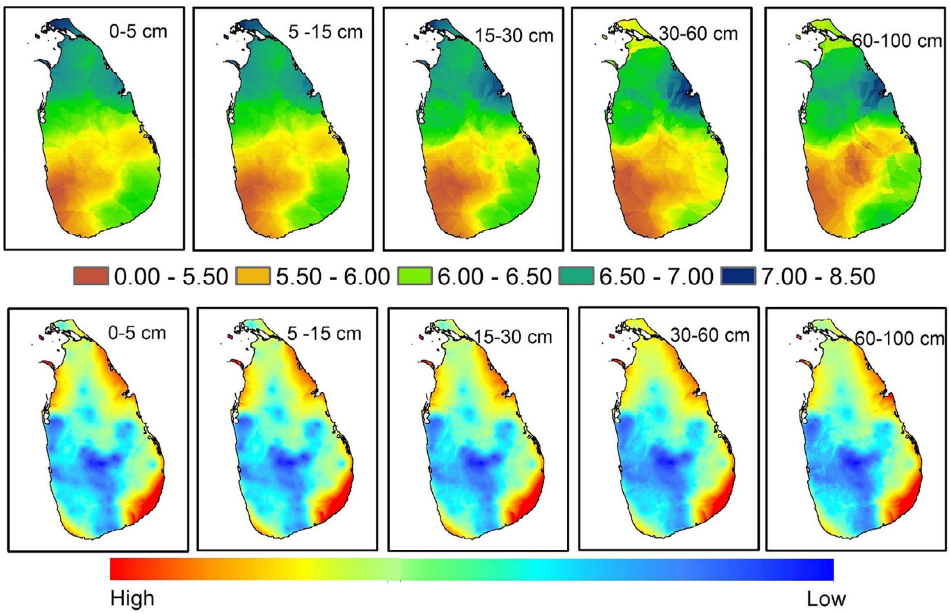

Fig. 1. Interpolated soil pH maps throughout 5 standard soil depths (above) and their error maps (below).

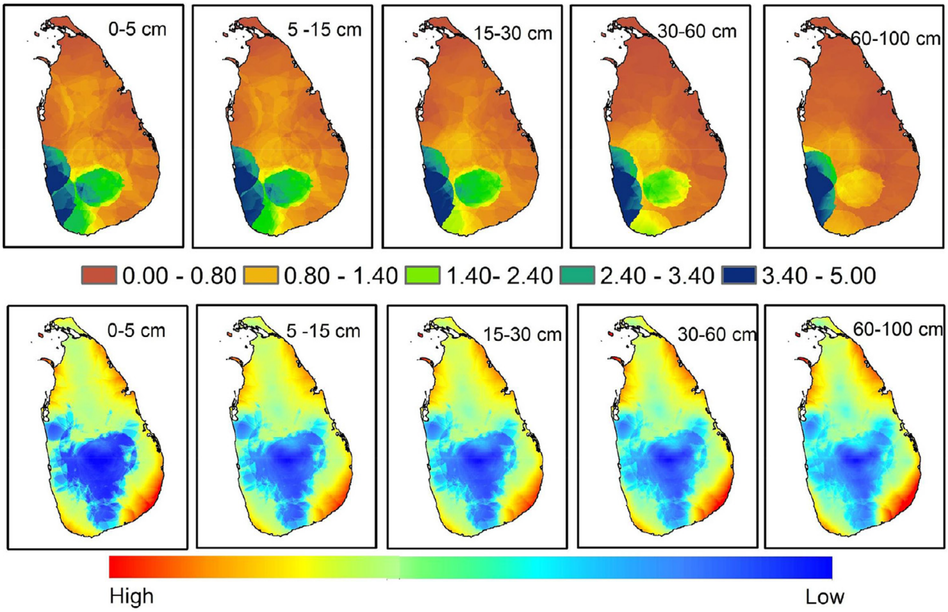

Fig. 2. Interpolated soil organic carbon content (%) maps throughout 5 standard soil depths (above) and their error maps (below).

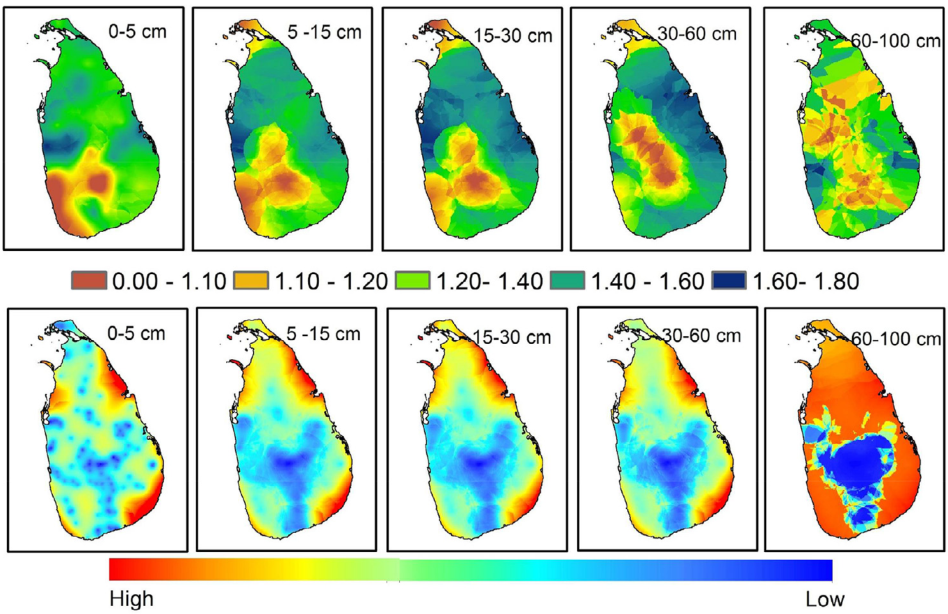

Fig. 3. Interpolated soil bulk density (gcm^{-3}) maps throughout 5 standard soil depths (above) and their error maps (below).

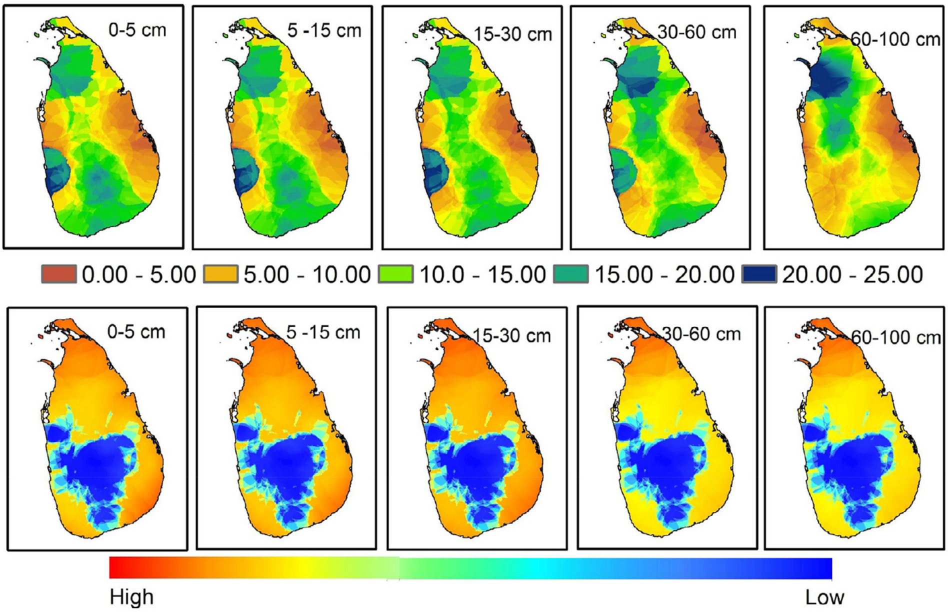

Fig. 4. Interpolated soil cation exchange capacity (cmol+ kg^{-1}) maps throughout 5 standard soil depths (above) and their error maps (below).

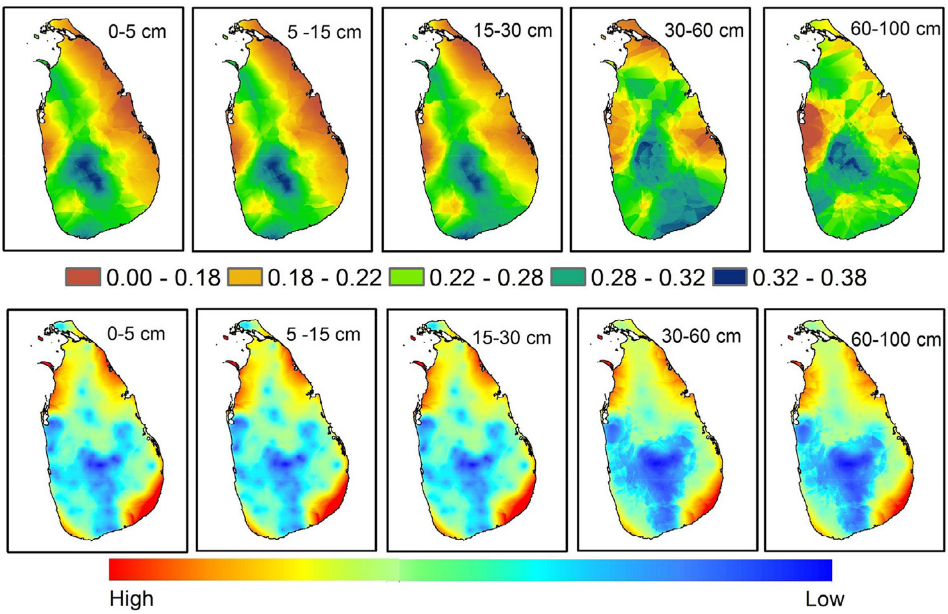

Fig. 5. Interpolated soil volumetric moisture content at 0.33 bars level (drainage upper limit – mm mm^{-1}) maps throughout 5 standard soil depths (above) and their error maps (below).

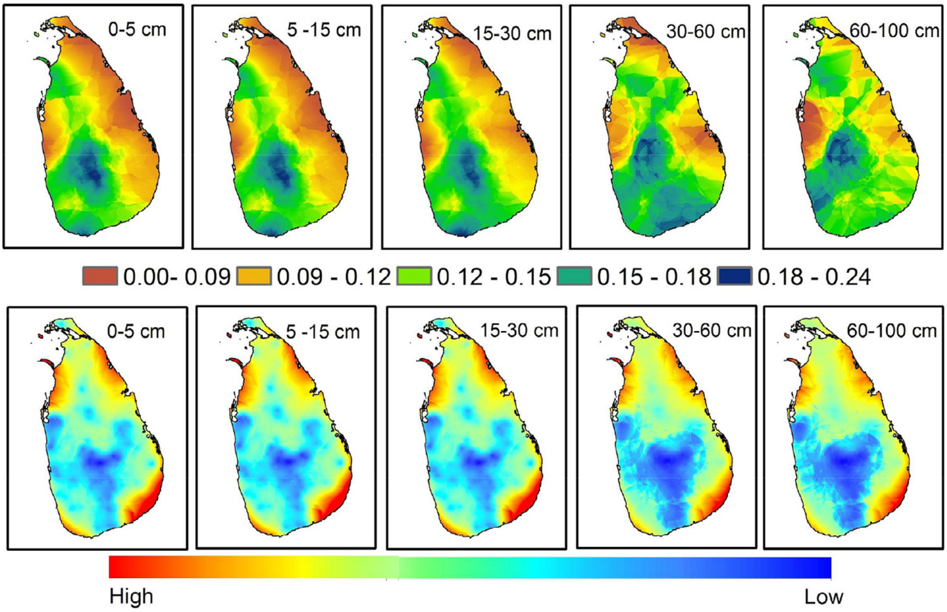

Fig. 6. Interpolated soil volumetric moisture content at 15 bars level (wilting point – mm mm^{-1}) maps throughout 5 standard soil depths (above) and their error maps (below).

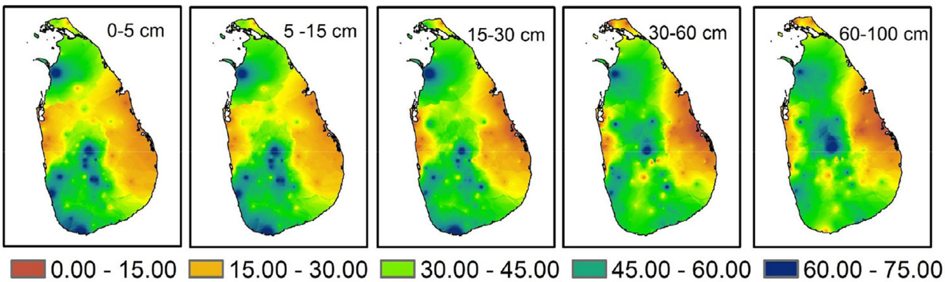

Fig. 7. Interpolated clay content (%) maps throughout 5 standard soil depths.

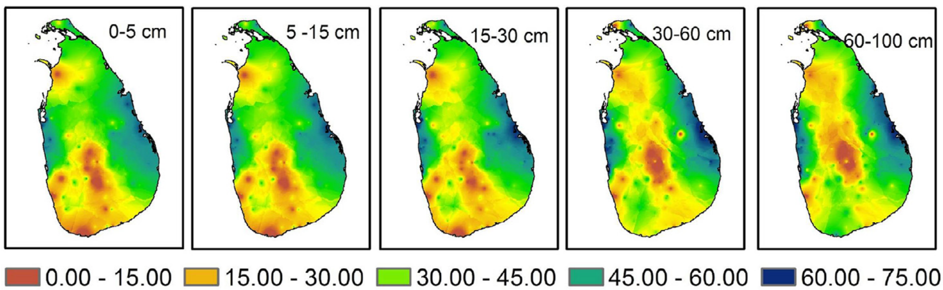

Fig. 8. Interpolated sand content (%) maps throughout 5 standard soil depths.

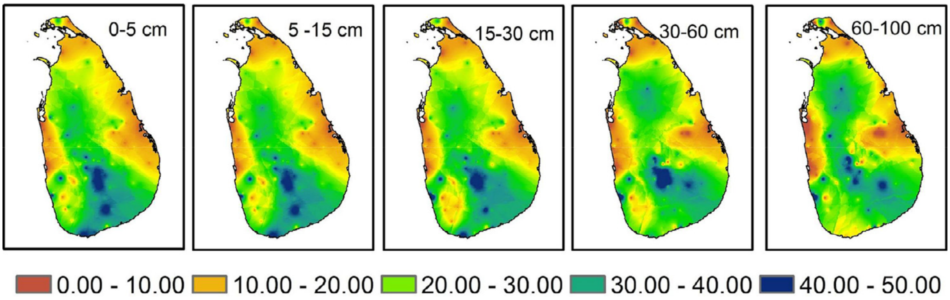

Fig. 9. Interpolated silt content (%) maps throughout 5 standard soil depths.

2. Experimental Design, Materials and Methods

Observed soil data from two databases were used to develop interpolated maps. The SRICANSOL is the most comprehensive and up-to-date soil database in Sri Lanka that describes 110 benchmark soil series across the country [1–3]. Twelve soil series in the northern region that were not included in the SRICANSOL were also added [4]. These 122 locations covered all the land use types in Sri Lanka. Soil sampling locations used for interpolations were previously described by Wimalasiri et al. [5].

The observed data were available for different horizons, showing an enormous variation across the country. Therefore, the depths were harmonised into 5 standard depths as 0–5, 5–15, 15–30, 30–60 and 60–100 cm using equal-area quadratic smoothing splines [6]. Detailed soil analysis methods were previously described [3]. In the observed soil, pH was determined using a 1:2.5 soil-water ratio. Parallel to the international standards for soil pH determination (ISO 10390: 2005), pH was converted to the ratio of 1:5 water as follows [7].

$$\mathrm{pH1}:5\mathrm{W} = 0.14 + 0.99 \times \mathrm{pH1}:2.5\mathrm{W},$$

where W is the water.

In the next step, maps of all soil properties were created for the entire landmass of Sri Lanka. The observed soil property data were interpolated using different interpolation methods to predict the spatial distribution of the observed data. Summary statistics of soil properties used for interpolation mapping are provided as supplementary materials (Table S1). Statistical interpolation methods have the ability to utilise the spatial correlation among the observations to predict values at unsampled locations [8]. In this study, univariate Kriging, Cokriging, Empirical Bayesian Kriging and deterministic methods such as inverse distance weighting, radial basis function and other interpolation methods were examined for all properties across all depths. The maps were created for both soil chemical and physical properties. The cross-validation results demonstrated that radial basis function (RBF) was the best interpolation method for mapping soil physical (clay, sand and silt) properties. This is in agreement with the findings of Rodrigues et al. [9] that RBF shows the best result for mapping physical properties with relatively small errors. However, for other soil properties (pH, bulk density, cation exchange capacity, organic carbon content, volumetric moisture content (VMC) at 0.33 and 15 bars level) Empirical Bayesian Kriging showed lowest RMSE. Similar studies also have shown that Kriging method is superior for univariate soil chemical property mapping [10,11]. Due to sparse nature of the dataset smooth circular neighbourhood selection method was selected for pH. The neighbourhood type of all other parameters was standard circular. The semivariogram type was completely regularized spline in soil texture while linear method was used for all other parameters. Interpolation methods and their prediction accuracy for soil attributes are summarised in the Table 1.

Table 1
Interpolation methods and their prediction accuracy for soil attributes at different layers.

		RMSE*				
Soil Parameter	Geostatistics Method	0–5 cm	5–15 cm	15–30 cm	30–60 cm	60–100 cm
pH	Empirical Bayesian Kriging	0.770	0.756	0.848	1.273	1.500
OC	Empirical Bayesian Kriging	2.649	2.620	2.616	2.185	1.651
BD	Empirical Bayesian Kriging	0.184	0.274	0.281	0.390	0.668
CEC	Empirical Bayesian Kriging	14.949	14.765	14.301	12.761	8.177
DUL	Empirical Bayesian Kriging	0.090	0.086	0.077	0.088	0.115
LL15	Empirical Bayesian Kriging	0.082	0.078	0.070	0.074	0.091
Clay	Radial Basis Function	10.838	10.694	11.678	13.377	14.735
Sand	Radial Basis Function	15.970	15.861	16.482	20.565	21.368
Silt	Radial Basis Function	8.159	8.151	8.688	8.717	8.685

Where OC = Organic carbon, BD = Bulk density, CEC = Cation exchange capacity, DUL = Drainage upper limit 606 (VMC33) and LL15 = Wilting point (VMC1500).
* RMSE values were obtained from Wimalasiri et al. [5].

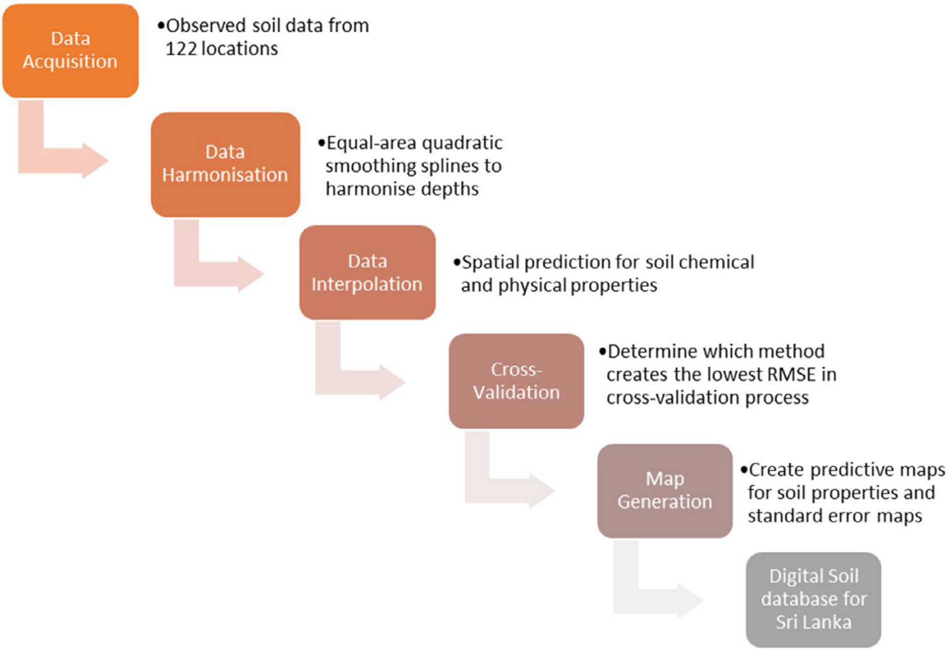

Fig. 10. Flow chart for the generation of Sri Lanka digital soil database.

The complete flow chart of the process is shown in Fig. 10. As more soil data are becoming available, more accurate results can be achieved by combining the primary data with auxiliary information. The online repository will be gradually updated with higher resolution maps in the future.

Ethics Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2020.106342.

CRediT authorship contribution statement

Eranga M. Wimalasiri: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft. **Ebrahim Jahanshiri:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft. **T.A.S.T.M. Suhairi:** Methodology, Software, Validation, Formal analysis, Writing - original draft, Visualization. **Ranjith B. Mapa:** Writing - review & editing. **Asha S. Karunaratne:** Writing - review & editing. **Lal P. Vidhanarachchi:** Writing - review & editing. **Hasika Udayangani:** Methodology, Data curation. **N.M.M. Nizar:** Writing - original draft. **Sayed N. Azam-Ali:** Writing - review & editing.

References

[1] R.B. Mapa, S. Somasiri, S. Magarajah, Soils of the Wet Zone of Sri Lanka: Morphology, Characterization and Classification: Special Publication No. 1., Soil Science Society of Sri Lanka, Sarvodaya Wishva Lekha, Sri Lanka, 1999.

[2] R.B. Mapa, A.R. Dassanayake, H.B. Nayakekorale, Soils of the Intermediate Zone of Sri Lanka: Morphology, Characterization and Classification. Special Publication No. 4, Soil Science Society of Sri Lanka, Sarvodaya Wishva Lekha, Sri Lanka, 2005.

[3] R.B. Mapa, S. Somasiri, A.R. Dassanayake, Soils of the Dry Zone of Sri Lanka: Morphology, Characterization and Classification. Special Publication No. 7, Soil Science Society of Sri Lanka, Sarvodaya Wishva Lekha, Sri Lanka, 2010.

[4] R.B. Mapa, Characterization of Soils in the Northern Region of Sri Lanka to Develop a Soil Data Base for Land Use Planning and Environmental Applications, National Research Council of Sri Lanka, Sri Lanka, 2016.

[5] E.M. Wimalasiri, E. Jahanshiri, T.A.S.T.M. Suhairi, H. Udayangani, R.B. Mapa, A.S. Karunaratne, L.P. Vidhanarachchi, S.N. Azam-Ali, Basic soil data requirements for reliable crop yield simulations in process-based crop models as a basis for crop diversification, Sustainability 12 (2020) 7781 https://doi.org/10.3390/su12187781.

[6] T.F.A. Bishop, A.B. McBratney, G.M. Laslett, Modelling soil attribute depth functions with equal-area quadratic smoothing splines, Geoderma 91 (1999) 27–45 https://doi.org/10.1590/S1413-7054201500020000110.1016/S0016-7061(99)00003-8.

[7] C. Kabała, E. Musztyfaga, B. Gałka, D. Łabuńska, P. Mańczyńska, Conversion of soil pH 1:2.5 KCl and 1:2.5 H_2O to 1:5 H_2O: conclusions for soil management, environmental monitoring, and international soil databases, Pol. J. Environ. Stud. 25 (2016) 647–653 https://doi.org/10.1590/S1413-7054201500020000110.15244/pjoes/61549.

[8] F.W.A. Júnior, E.M. Silveira, J.M. de Mello, C.R. de Mello, J.R.S. Scolforo, Change detection in Brazilian savannas using semivariograms derived from NDVI images, Ciênc. Agrotec. 39 (2015) 103–109 https://doi.org/10.1590/S1413-70542015000200001.

[9] M.S. Rodrigues, D.C. Alves, V.C. de Souza, A.C. de Melo, A.M.N. Lima, J.C. Cunha, Spatial interpolation techniques for site-specific irrigation management in a mango orchard, Com. Sci. 9 (2018) 93–100 https://doi.org/10.14295/cs.v9i1.2645.

[10] A. Bostani, M.M. Rahman, D. Khojasteh, Spatial mapping of soil properties using geostatistical methods in the Ghazvin plains of Iran, Mod. Appl. Sci. 11 (2017) 23–37.

[11] M. Nikpey, M. Sedighkia, M.B. Nateghi, J. Robatjazi, Comparison of spatial interpolation methods for mapping the qualitative properties of soil, Adv. Agric. 5 (2017) 1–15.