

Studies in Big Data

Volume 72

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series “Studies in Big Data” (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

** Indexing: The books of this series are submitted to ISI Web of Science, DBLP, Ulrichs, MathSciNet, Current Mathematical Publications, Mathematical Reviews, Zentralblatt Math: MetaPress and Springerlink.

More information about this series at <http://www.springer.com/series/11970>

Pradeep Kumar Garg · Rahul Dev Garg ·
Gaurav Shukla · Hari Shanker Srivastava

Digital Mapping of Soil Landscape Parameters

Geospatial Analyses using Machine Learning
and Geomatics

Pradeep Kumar Garg
Geomatics Section, Civil Engineering
Department
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand, India

Rahul Dev Garg
Geomatics Section, Civil Engineering
Department
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand, India

Gaurav Shukla
Surveying and Geomatics Section, Civil
Engineering Department
Maharishi Markandeshwar University
Ambala, Haryana, India

Hari Shanker Srivastava
Indian Institute of Remote Sensing (IIRS)
Dehradun, Uttarakhand, India

ISSN 2197-6503

ISSN 2197-6511 (electronic)

Studies in Big Data

ISBN 978-981-15-3237-5

ISBN 978-981-15-3238-2 (eBook)

<https://doi.org/10.1007/978-981-15-3238-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

About This Book

This book begins with the fundamental concepts behind the mapping of soil-landscape parameters in geospatial domain. As readers go through the chapters of this book, they will learn how machine learning and geomatics can be applied for more efficient mapping, understanding, and management of ‘soil’.

The judicious utilization of the piece of land is the biggest and most important challenging issue. This is especially true in light of the rapid urbanization worldwide, which requires continuous monitoring of resource consumption. This book provides a clear idea of how machine learning can be used to explore a remote sensing data to monitor the key parameters, below the surface, at the surface, and above the surface. Not only will readers gain insight into the approaches, they will also learn challenges and issues associated with the digital mapping of these parameters. This book also provides understanding, into the selection of data set to represent soil-landscape relationships. Readers will also be able to understand the complex and interconnected linkages between soil-landscape parameters under a range of soil and climatic conditions. The readers gain insight into utilization of a network of satellite-based Earth observations to provide a solution towards smart farming and smart land management.

The first two chapters prepare the readers to comprehend the basic concepts, approaches, and issues/challenges associated with digital mapping of soil parameters. All other chapters develop reader’s ability to handle the complexity involved in the digital mapping of three key parameters, i.e. below the surface: soil moisture, at the surface: soil class, and above the surface: crop cover.

Our potential readers will be researchers and geo-scientist, exploring digital mapping techniques for resource management, environmental planning, and sustainable development, as well as the college and university students studying geospatial technologies. It is also hoped that book contents will also attract and inspire the individuals now in the industry who requires framework in the field related to smart land planning, crop management, precision agriculture (satellite farming), and sustainable-city development in geospatial domain.

Contents and Coverage

The book comprises 6 chapters. Chapter 1: “Concept of Digital Mapping”—covers fundamental of digital mapping, key soil-landscape parameters, different approaches, and important challenging issue in digital mapping of soil-landscape parameters. Chapter 2: “Different Approaches on Digital Mapping of Soil-Landscape Parameters”—provides an in-depth explanation of various approaches for the mapping of soil classes, crop cover, and soil moisture using different machine learning algorithms. It also covers the selected method, structure, and assessment systems. Chapter 3: “Selection of Suitable Variables and Their Development”—covers the suitable variables and their development related to soil properties and parent material, climate, organisms/vegetation, relief/topography, relative position, soil moisture, and surface characteristics. Satellite data and its pre-processing are also presented for development of suitable variables. Chapter 4: “Digital Soil Mapping: Implementation and Assessment”—is devoted to approach and assessment of digital soil mapping. It discusses the various approaches that have been, or could be, used for fitting relationships between soil properties or classes and soil-forming factors. It also discusses the data layers that have been, or could be, used to describe the soil-forming factors and their importance in digital soil mapping. Chapter 5: “Prediction Models for Crop Mapping”—describes the framework for implementation of models for crop-type prediction using climate variables in association with texture, vegetation, soil moisture, phenology, topography, parent material, and soil, followed by comparative performance analysis of different ensemble models. It also discusses the potential of different variables for crop-type classification. Chapter 6: “Spatial Soil Moisture Prediction Model Over an Agricultural Land”—discusses the framework for surface soil moisture prediction model. It also discusses the model evaluation based on the goodness-of-fit test and probability of non-exceedance on field and SMAP (Soil Moisture Active and Passive) data.

A Bibliography at the end of each chapter is for the benefits of the students.

Contents

1	Concept of Digital Mapping	1
1.1	Key Soil-Landscape Parameters	1
1.2	Retrieval Approaches	3
1.3	Underlying Concepts and Challenges	6
1.3.1	Challenges Associated with Acquisition Techniques	7
1.3.2	Challenges Associated with Analytical Algorithms	7
	References	9
2	Different Approaches on Digital Mapping of Soil-Landscape Parameters	13
2.1	State of the Art in Digital Soil Mapping (DSM)	13
2.2	Machine Learning in Crop Cover Mapping	23
2.3	State of the Art in Soil Moisture Estimation	26
	References	37
3	Selection of Suitable Variables and Their Development	47
3.1	Introduction	47
3.2	Satellite Data and Pre-Processing	49
3.2.1	Landsat-8	49
3.2.2	Modis	49
3.2.3	Digital Elevation Model (DEM)	50
3.2.4	Synthetic Aperture Radar (SAR)	50
3.3	Predictor Variables and Their Development	50
3.3.1	Soil Properties and Parent Material	51
3.3.2	Climate	53
3.3.3	Organisms/Vegetation	54
3.3.4	Relief/Topography	57
3.3.5	Relative Position	58
3.3.6	Soil Moisture and Surface Characteristics	58
	References	60

4	Digital Soil Mapping: Implementation and Assessment	65
4.1	Introduction	65
4.2	Material and Methods	67
4.2.1	Background to Implementation	67
4.2.2	Soil Spatial Prediction Models	69
4.2.3	Model Construction and Implementation	76
4.2.4	Performance Evaluation	76
4.2.5	Parameters to Access Model Behaviour	81
4.3	Results and Discussion	81
4.3.1	RF Dependency on Parameters (<i>mtry</i> , <i>ntree</i>)	81
4.3.2	Identification of Effective Indicators	82
4.3.3	Accuracy Assessment	83
4.3.4	Comparative Performance Analysis	83
4.3.5	Influence of Training Data Set Size	86
4.3.6	Influence of Noise in the Training Set	88
4.4	Summary	88
	References	90
5	Prediction Models for Crop Mapping	93
5.1	Introduction	93
5.2	Modelling Environment	94
5.2.1	Modelling Methods	94
5.2.2	Construction and Implementation	99
5.2.3	Performance Evaluation	101
5.3	Results and Discussion	102
5.3.1	Identification of Effective Indicators	102
5.3.2	Class-Separability Analysis	105
5.3.3	Comparative Performance Analysis	105
5.4	Summary	114
	References	114
6	Spatial Soil Moisture Prediction Model Over an Agricultural Land	117
6.1	Introduction	117
6.2	Modelling Framework	120
6.2.1	Modelling Methods	120
6.2.2	Model Development and Implementation	123
6.2.3	Identification of Effective Indicators for Soil and Cover	124
6.2.4	Model Evaluation	127

6.3	Result and Discussion	128
6.3.1	Models Performance on Field Samples on Site I	128
6.3.2	Model Performance on the 9 km Modelling Grid in Site II	134
6.4	Summary	135
	References	137

About the Authors

Professor Pradeep Kumar Garg is faculty in Civil Engineering Department of Indian Institute of Technology, Roorkee. He has also served as vice chancellor, Uttarakhand Technical University, Dehradun. He completed his B.Tech. in 1980 and his M.Tech. in 1982, both from the University of Roorkee, India (now IIT Roorkee). He did a Ph.D. from University of Bristol, UK, and postdoctoral research work at the University of Reading, UK. He joined the Department of Civil Engineering at IIT Roorkee in 1982. Dr. Garg has published about 93 research papers, guided 7 Ph.D. thesis, 47 M.Tech. thesis, authored one textbook on Remote Sensing, organized 22 training programmes, and prepared two educational films under UGC programme. He has completed 13 research projects and 25 consultancy projects. He is Fellow member of 5 Technical Societies and life member of 15 Technical Societies. His main areas of research interest are remote sensing and GIS applications.

Professor Rahul Dev Garg is faculty in Civil Engineering Department of Indian Institute of Technology, Roorkee. He graduated with a bachelor's in technology in Civil Engineering in 1989, master's in technology in 1993, and a Ph.D. in 2004 from IIT Roorkee. He has also served as a scientist from 1993 to 2007 in Indian Institute of Remote Sensing (IIRS), Dehradun. Dr. Garg has published about 110 research papers, guided 8 Ph.D. thesis, 47 M.Tech. thesis, authored one textbook on Remote Sensing, organized 22 training programmes, and prepared two educational films under UGC programme. He has completed 11 research projects and 24 consultancy projects. He is Fellow member of 3 Technical Societies and life member of 10 Technical Societies. His main areas of interest are land surveying, remote sensing, GIS, GPS, digital image processing, SAR interferometry, and GPR.

Dr. Gaurav Shukla is currently working as a faculty member in surveying and geomatics section, Civil Engineering Department, Maharishi Markandeshwar (Deemed to be University) University, Mullana, Haryana, India. He is also a nodal coordinator of MHRD initiative, virtual lab programme. He is postgraduated in geomatics from the Indian Institute of Technology (ISM), Dhanbad, in 2011 and

completed his Ph.D. from Indian Institute of Technology, Roorkee, India, in 2018. Dr. Shukla has published 7 journal papers and organized 2 training programmes. His main areas of research interest include nonparametric approaches to retrieval of Earth's parameters, GNSS reflectometry, remote sensing, and GIS applications.

Dr. Hari Shanker Srivastava is a senior scientist, faculty and Group Head, Programme, Planning and Evaluation Group (PPEG) of Indian Institute of Remote Sensing (IIRS), ISRO, Dehradun, India. He is gold medallist in M.Sc. Physics and Ph.D. in Physics on Synthetic Aperture Radar (SAR). He is Ph.D., M.Tech., M.Sc. and B.Tech. guide at IIRS, IIT-Dhanbad, IIT-Roorkee, IIT-BHU, BITS, Pilani etc. His research interest is in the field of Microwave Remote Sensing focusing on multi-parametric SAR, passive microwave radiometers, Interferometry, Polarimetry and Polarimetric Interferometry for soil moisture, agriculture, wetland, forestry, human settlement, radiometric calibration of active sensors and validation of soil moisture derived from passive sensors. With more than 150 research publications to his credit, he is reviewer to a number of reputed journals published by IEEE, Elsevier, Taylor & Francis, Springer, SPIE etc.

Abbreviations

ALOS	Advanced Land Observing Satellite
ANN	Airborne Visible/Infrared Imaging Spectrometer
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CART	Classification and Regression Tree
CDT	CART Decision Trees
CEA	CART Ensembles ARCing
CEB	CART Ensemble Bagger
CKC	Cohen's Kappa coefficient
CTI	Compound Topographic Index
DEM	Digital Elevation Model
DSM	Digital Soil Mapping
DTM	Digital Terrain Model
EM	Electromagnetic
GIS	Geographical Information System
GPS	Global Positioning System
IRS	Indian Remote Sensing
ISLE	Importance Sampled Learning Ensemble
JCC	Jaccard's Coefficient of Community
jNSM	Java Newhall Simulation Model
LISS	Linear Imaging Self Scanning Sensor
LR	Linear Regression
MDA	Mean Decrease in Accuracy
MDG	Mean Decrease in Gini
MLR	Multinomial Logistic Regression
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NPV	Negative Predictive Value
OLI	Operational Land Imager
PPV	Positive Predictive Value

RADAR	Radio Detection and Ranging
RF	Random Forest
RH	Right Circular Transmit and Horizontal Receive
RISAT	Radar Imaging Satellite
RS	Remote Sensing
RV	Right Circular Transmit and Vertical Receive
SAR	Synthetic Aperture Radar
SMAP	Soil Moisture Active and Passive
SMPM	Soil Moisture Prediction Model
SMR	Soil Moisture Regime
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
TNGBM	TreeNet Gradient Boosting Machine
TNR	True Negative Rate
TPR	True Positive Rate
USGS	United States Geological Survey
UTM	Universal Transverse Mercator projection
VH	Vertical Transmit and Horizontal Receive
VV	Vertical Transmit and Vertical Receive

List of Figures

Fig. 2.1	Structure of digital soil mapping and digital soil assessment systems (Carre et al. 2007)	15
Fig. 2.2	T_s-F_v space.	31
Fig. 2.3	LST-NDVI triangle space	32
Fig. 2.4	Red-Near infrared feature space	33
Fig. 3.1	Variables related to soil properties and parent material: a IMI; b CTI; c Filtered CTI; and d True colour composite of Landsat-8 bands 4 (red), 3 (green), 2 (blue) images	52
Fig. 3.2	Soil moisture calendar of two classes of SMR at Aligarh [(a) ustic: typic tempustic] and Bharatpur [(b) aridic: weak aridic]	53
Fig. 3.3	Variables related to climate: a SMR; b TAR; c TAP; d PS; e MAT; and f Isothermality	55
Fig. 3.4	Variables related to the organism: a NDVI; b NDVI divergence measure by the 3×3 window; c NDVI variance measure by the 3×3 window; and d Phenological Variability.	56
Fig. 3.5	Variables related to relief or topography: DEM derived a Slope; b Aspect and Filtered DEM derived; c Slope; d Curvature image	57
Fig. 3.6	SAR derived data. a False colour composite of RISAT-1 polarization bands RH (red), RV (green), 0 (blue); b False colour composite of Sentinel-1A polarization bands VH (red), VV (green), 0 (blue).	59
Fig. 4.1	Structure of DSM and accuracy assessment systems.	72
Fig. 4.2	Error rate plot showing dependency on a number of trees and predictive variables.	73
Fig. 4.3	a Kappa coefficient plot for <i>mtry</i> and <i>nree</i> values; b Importance measure on the basis of SD score.	74
Fig. 4.4	Variable importance contribution of different bands in terms of a MDA and b MDG.	75

Fig. 4.5 RF predicted soil map of study area 77

Fig. 4.6 **a** CDT; **b** RF; and **c** CEB predicted soil map of study area 84

Fig. 4.7 Performance based on class specific measures 84

Fig. 4.8 Performance based on multiclass measures and time 85

Fig. 4.9 Overall performance based on multiclass measures and time 85

Fig. 4.10 Outlier analysis for each sample 86

Fig. 4.11 CKC plot for models trained by subsets of the original training data set. 87

Fig. 4.12 CKC plot against % of noise introduced. 88

Fig. 5.1 OOB error rate convergence for *mtry* values 4, 5, and 6 with 1000 trees 99

Fig. 5.2 Crop and non-crop map of the study area using RF model. 100

Fig. 5.3 Crop and non-crop maps of the study area using different models; **a** RF; **b** TNGB; **c** CEB; **d** CEA; **e** CDT; and **f** ISLE 101

Fig. 5.4 Potentiality of variables based on **a** MDA and **b** MDG score. 103

Fig. 5.5 Local importance plot showing potentiality of variables in each class 104

Fig. 5.6 Performance of methods on the basis of class specific measures 112

Fig. 5.7 Performance of methods on the basis of multiclass measures and total time (dark boundary) 113

Fig. 5.8 Overall performance of methods without time and with time factor (dark boundary) 113

Fig. 6.1 SMPM modelling approach. 120

Fig. 6.2 Flow chart of SMPM 121

Fig. 6.3 **a** MDA score and **b** MDG score for crop and soil class study. Average value of score is considered as cut-off value for selection of variables of major importance. The variables above the cut-off value are presented with light green and reddish brown colour within their textured bar for crop and soil class mapping, respectively 125

Fig. 6.4 Correlation between predicted and actual surface soil moisture content for model **a** RF with $\sigma_{RH}^0, \sigma_{RV}^0$; **b** RF with $\sigma_{RH}^0/\sigma_{RV}^0$; **c** RF with $\sigma_{VH}^0, \sigma_{VV}^0$; **d** RF with $\sigma_{VH}^0/\sigma_{VV}^0$; **e** SVM with $\sigma_{RH}^0, \sigma_{RV}^0$; **f** SVM with $\sigma_{RH}^0/\sigma_{RV}^0$. Dotted line is the 45° bisector line 129

Fig. 6.5 Correlation between predicted and actual surface soil moisture content for model **a** SVM with $\sigma_{VH}^0, \sigma_{VV}^0$; **b** SVM with $\sigma_{VH}^0/\sigma_{VV}^0$; **c** ANN with $\sigma_{RH}^0, \sigma_{RV}^0$; **d** ANN with $\sigma_{RH}^0/\sigma_{RV}^0$; **e** ANN with $\sigma_{VH}^0, \sigma_{VV}^0$; **f** ANN with $\sigma_{VH}^0/\sigma_{VV}^0$. The dotted line is the 45° bisector line 130

Fig. 6.6 Probability of non-exceedance for model **a** RF with $\sigma_{RH}^0, \sigma_{RV}^0$; **b** RF with $\sigma_{RH}^0/\sigma_{RV}^0$; **c** RF with $\sigma_{VH}^0, \sigma_{VV}^0$; **d** RF with $\sigma_{VH}^0/\sigma_{VV}^0$; **e** SVM with $\sigma_{RH}^0, \sigma_{RV}^0$; **f** SVM with $\sigma_{RH}^0/\sigma_{RV}^0$; **g** SVM with $\sigma_{VH}^0, \sigma_{VV}^0$; **h** SVM with $\sigma_{VH}^0/\sigma_{VV}^0$; **i** ANN with $\sigma_{RH}^0, \sigma_{RV}^0$; **j** ANN with $\sigma_{RH}^0, \sigma_{RV}^0$; **k** ANN with $\sigma_{VH}^0, \sigma_{VV}^0$; **l** ANN with $\sigma_{VH}^0/\sigma_{VV}^0$. The horizontal light dotted line shows an error value of 5% and the dark dotted horizontal line shows an error value of 10% 133

Fig. 6.7 Correlation between predicted and SMAP L4_SM, for model **a** RF with $\sigma_{RH}^0, \sigma_{RV}^0$; **b** RF with $\sigma_{RH}^0/\sigma_{RV}^0$; **c** RF with $\sigma_{VH}^0, \sigma_{VV}^0$; and **d** SVM with $\sigma_{RH}^0, \sigma_{RV}^0$. Dotted line is the 45° bisector line 135

Fig. 6.8 Surface soil moisture map at 9 km grid of **a** SMAP L4_SM; and model predicted using **b** RF with $\sigma_{RH}^0, \sigma_{RV}^0$; **c** RF with $\sigma_{RH}^0/\sigma_{RV}^0$; **d** RF with $\sigma_{VH}^0, \sigma_{VV}^0$; **e** RF with $\sigma_{VH}^0/\sigma_{VV}^0$; **f** SVM with $\sigma_{RH}^0, \sigma_{RV}^0$ 136

List of Tables

Table 2.1	Summary of few research works on DSM	20
Table 2.2	Comparison of traditional ground-based techniques	28
Table 2.3	Summary presenting the advantages and limitation of remote sensing data for near-surface soil moisture estimation.	30
Table 3.1	Predictor variables associated with six covariates categories	48
Table 4.1	Class wise local importance of variables in descending order of importance	70
Table 4.2	Performance measures of classifiers RF, CART decision tree, and CART ensembles bagger.	78
Table 4.3	Grading of performance	80
Table 5.1	Separability analysis between different crop and non-crop classes using JM distance method	106
Table 5.2	Performance measures of classifiers RF, CDT, and CEB	107
Table 5.3	Performance measures of classifier CEB/A, ISLE, and TNGBM	110
Table 5.4	Performance scale for model evaluation	112
Table 6.1	Performance measures of RF, SVM, and ANN models on testing samples on Site I. The MSE, RMSE, and MAE values are in percentage (%)	131