

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Use of new generation geospatial data and technology for low cost
drought monitoring and SDG reporting solution

A thesis presented in partial fulfillment of the requirement for the
degree of

Master of Science
in
Computer Science

at Massey University, Manawatū, New Zealand.

Mohammad Hossain Dehghan-Shoar

2018

Abstract

Food security is dependent on ecosystems including forests, lakes and wetlands, which in turn depend on water availability and quality. The importance of water availability and monitoring drought has been highlighted in the Sustainable Development Goals (SDGs) within the 2030 agenda under indicator 15.3. In this context the UN member countries, which agreed to the SDGs, have an obligation to report their information to the UN. The objective of this research is to develop a methodology to monitor drought and help countries to report their findings to UN in a cost-effective manner.

The Standard Precipitation Index (SPI) is a drought indicator which requires long-term precipitation data collected from weather stations as per World Meteorological Organization recommendation. However, weather stations cannot monitor large areas and many developing countries currently struggling with drought do not have access to a large number of weather-stations due to lack of funds and expertise. Therefore, alternative methodologies should be adopted to monitor SPI.

In this research SPI values were calculated from available weather stations in Iran and New Zealand. By using Google Earth Engine (GEE), Sentinel-1 and Sentinel-2 imagery and other complementary data to estimate SPI values. Two genetic algorithms were created, one which constructed additional features using indices calculated from Sentinel-2 imagery and the other data which was used for feature selection of the Sentinel-2 indices including the constructed features. Followed by the feature selection process two datasets were created which contained the Sentinel-1 and Sentinel-2 data and other complementary information such as seasonal data and Shuttle Radar Topography Mission (SRTM) derived information.

The Automated Machine Learning tool known as TPOT was used to create optimized machine learning pipelines using genetic programming. The resulting models

yielded an average of 90 percent accuracy in 10-fold cross validation for the Sentinel-1 dataset and an average of approximately 70 percent for the Sentinel-2 dataset. The final model achieved a test accuracy of 80 percent in classifying short-term SPI (SPI-1 and SPI-3) and an accuracy of 65 percent of SPI-6 by using the Sentinel-1 test dataset. However, the results generated by using Sentinel-2 dataset was lower than Sentinel-1 (45 percent for SPI-1 and 65 percent for SPI-6) with the exception of SPI-3 which had an accuracy of 85 percent.

The research shows that it is possible to monitor short-term SPI adequately using cost free satellite imagery in particular Sentinel-1 imagery and machine learning. In addition, this methodology reduces the workload on statistical offices of countries in reporting information to the SDG framework for SDG indicator 15.3. It emerged that Sentinel-1 imagery alone cannot be used to monitor SPI and therefore complementary data are required for the monitoring process.

In addition the use of Sentinel-2 imagery did not result in accurate results for SPI-1 and SPI-6 but adequate results for SPI-3. Further research is required to investigate how the use of Sentinel-2 imagery with Sentinel-1 imagery impact the accuracy of the models.

Acknowledgements

I would like to thank my kind supervisors Prof. Hans Guesgen, Dr. Sunil Lal and Dr. Lorenzo De Simone for their guidance and support during my study. In addition I would like to thank my family members for helping me in my time of need in particular my father and mother for their kind support.

Contents

List of Tables	4
List of Figures	6
1 Introduction	8
1.1 Impacts of drought	8
1.2 Types of drought	9
1.3 Monitoring drought	11
1.4 Objectives	11
2 Literature Review	13
2.1 Remote sensing	13
2.2 Synthetic Aperture Radar (SAR)	14
2.2.1 Microwave imagery and soil moisture	16
2.2.2 Sentinel-1 imagery pre-processing	17
2.3 Multispectral satellites (cameras)	18
2.3.1 Pre-processing of multispectral imagery	20
2.4 Cloud computing for remote sensing data	21
2.5 Drought indicators	22
2.5.1 Evapotranspiration	22
2.5.2 The Standard Precipitation Index	23
2.5.3 SPI strengths and weaknesses	24
2.5.4 Applications of SPI	24

2.5.5	SPI formula	25
2.6	Machine learning	26
2.6.1	The Artificial Neuron	27
2.6.2	Artificial Neural Networks	28
2.6.3	Ensemble Machine Learning Algorithms	30
2.6.4	The issue of overfitting	32
2.6.5	Curse of Dimensionality	33
2.6.6	Genetic algorithms for feature selection and model optimization	35
3	Materials and Methods	38
3.1	Sentinel-2 methodology	42
3.2	Sentinel-1 methodology	49
4	Sentinel-1 results and discussion	53
4.1	Sentinel-1 results	53
4.1.1	Correlation heatmap between Sentinel-1 input data and SPI values	53
4.2	Sentinel-1 discussion	54
5	Sentinel-2 results and discussion	57
5.1	Sentinel-2 results	57
5.1.1	New constructed features for SPI-1	57
5.1.2	New constructed features for SPI-3	58
5.1.3	New constructed features for SPI-6	58
5.1.4	Feature selection results	58
5.1.5	The selected features for each of the SPI values	60
5.1.6	Correlation heatmaps for selected Sentinel-2 features and SPI values	61
5.2	Sentinel-2 discussion	63
5.2.1	Selected features for SPI-1	64
5.2.2	Selected features for SPI-3	65
5.2.3	Selected features for SPI-6	67
5.2.4	Feature selection outcomes	69
5.2.5	Final results of the Sentinel-2 methodology	69

6 Conclusion	71
Bibliography	73

List of Tables

2.1	Specification of different microwave bands	14
2.2	Different types of polarization	16
2.3	Bands specification for the Sentinel-2 satellite	19
2.4	Different versions of SPI	26
2.5	Different SPI values and their labels	26
2.6	Possible locations of the object in one, two and three dimensions . . .	34
3.1	Different SPI values and their classifications scheme based on Mckee (1993) [1]	39
3.2	List of weather stations used for creating the models	41
3.3	List of weather stations used for testing the models	42
3.4	Atmospherically corrected bands retrieved from Sentinel-2	44
3.5	Remote sensing indices calculated by using the Sentinel-2 spectral bands	45
3.6	Seasonal data used for the Iranian data	47
3.7	Seasonal data used for the New Zealand data	47
4.1	Cross validation and testing scores achieved by using the Sentinel-1 methodology and the XGboost classifier	53
5.1	Best constructed features for SPI-1	57
5.2	Best constructed features for SPI-3	58
5.3	Best constructed features for SPI-6	58
5.4	Results generated following the feature selection process	59

5.5	Selected features of the Sentinel-2 datasets	60
5.6	Accuracy of the Sentinel-2 methodology by using the Iran and New Zealand data	69

List of Figures

1.1	Chart displaying different types of droughts and their impacts on the environment	10
1.2	Variables commonly used to monitor different types of drought	10
2.1	Illustration of a passive satellite	13
2.2	Illustration of an active satellite	14
2.3	Illustration of differently polarized signals interacting with the environment	15
2.4	Reflection of microwave signals by soils containing different water content	16
2.5	Change in reflectance at different wavelengths for various objects . . .	20
2.6	Effects of the atmosphere on incoming shortwave radiation	21
2.7	A simplified illustration of variables contributing to Evapotranspiration	23
2.8	Drought occurrence and its intensity derived from VHI	25
2.9	Illustration of an Artificial Neuron	28
2.10	Illustration of the Sigmoid and ReLU activation functions	28
2.11	Illustration of a backpropagational neural network	29
2.12	Illustration of Gradient Descent	30
2.13	Combining the results from different learners by using majority voting to achieve a final solution	31
2.14	Demonstration of the effect of max depth for a classification problem	32

2.15	Illustration of an object present in a search space with different dimensions	34
2.16	Illustration of a search space and the global maximum	35
2.17	Illustration of Population, Chromosomes and Genes in genetic algorithms	36
2.18	Flowchart of a genetic algorithm	37
3.1	The locations of the weather stations used from Iran	40
3.2	The locations of the weather stations used from New Zealand	40
3.3	Circular buffer created and used to extract eight pixels surrounding the weather station. The black pixel represents the weather station location	43
3.4	Depiction of earth at perihelion and aphelion	44
3.5	Sentinel-2 methodology flowchart	49
3.6	Sentinel-1 methodology flowchart	52
4.1	Correlation heatmap for Sentinel-1	54
4.2	Processes contributing to change in soil moisture	55
5.1	Chart generated by using the genetic algorithm for feature selection (SPI-1)	59
5.2	Correlation heatmap for selected Sentinel-2 features and SPI-1	61
5.3	Correlation heatmap for selected Sentinel-2 features and SPI-3	62
5.4	Correlation heatmap for selected Sentinel-2 features and SPI-6	62
5.5	The impact of slope on water flow and soil water absorption	64
5.6	Change in reflectance in Red and NIR bands for healthy and unhealthy vegetation	65
5.7	The variation of Multispectral Scanner (MSS) pixel positions corresponding to growing vegetation, as related to the Tasseled Cap transformation (Kauth and Thomas, 1976)	66
5.8	The planes and axes of the Tasseled Cap transformation	67
5.9	Demonstration of different stages of the vegetation life cycle	68