Estimating Glaciers Ice Thickness with Machine Learning

Master's Thesis

in Atmospheric Sciences

Submitted to the
Faculty of Geo- and Atmospheric Sciences
of the
University of Innsbruck

in Partial Fulfillment of the Requirements for the Degree of
Master of Science

by
Matteo Castellani

Advisor
Assistant Professor Dr. Fabien Maussion

Innsbruck, December 2020

To be decided

Preface

Abstract

Contents

Chapter 1

Introduction

1.1 Motivation

(1) Hype in machine learning (both in academic and business world).

(2) Hype in estimating glacier ice thickness

(3) Most models use physical based approaches. Use a statistical one.

1.2 State of Research

Background of the literature: GlaThiDa, ITMIX (Farinotti et al. (2017)), etc

1.3 Goals and Outline

(1) How well can Machine Learning Algorithm predict glaciers ice thickness.

(2) How do 3 different machine learning algorithm compare with each other in estimating glaciers ice thickness.

(3) If we use those algorithm to estimate the total volume of glaciers in the alps, how do these model compare to some of the physical based ones for this region.

Chapter 2

Methodology

This seems like it will contain a lot of stuff.... Maybe i can avoid some?

2.1 Glathida Dataset

2.1.1 GlaThiDa

GlaThiDa

2.1.2 RGI

RGI

2.1.3 linking them

linking them using OGGM

2.1.4 Some stats about GlaThiDa

GlaThiDa Statistics about glaciers distributions and so on

2.2 Choosing Features

2.2.1 Putting together the features

How did we get the features for training: OGGM.

2.2.2 Which features did we choose

The features we chose and why.

2.3 Machine Learning Models

What they are and how they work on the high level

2.3.1 Tuning parameters

How did we decide which parameters to use (maybe we can just leave it for the appendix)

2.3.2 SVM

Explain Support Vector Machine (do i need to write down the specific mathematics?)

2.3.3 Random Forest

Random Forest

2.3.4 Linear Regression

Linear Regression

2.4 Training method

use sklearn train_test_split 20 times

2.5 Scoring method

metrics to compare the goodness of models

2.6 Features Importance

what is feature importance.

2.6.1 Shuffle

Shuffle

2.6.2 partial dependence plot

partial dependence plot

Chapter 3

Results

3.1 Linear Regression Results

Results from the linear regression model

(1) Score plots

(2) Volume spread plots

(3) Glacier maps

(4) Features importance

(5) Volume for the whole alps

3.2 SVM

Results from the SVM

(1) Score plots

(2) Volume spread plots

(3) Glacier maps

(4) Features importance

(5) Volume for the whole alps

3.3 Random forest

Results from the Random Forest

(1) Score plots

(2) Volume spread plots

(3) Glacier maps

(4) Features importance

(5) Volume for the whole alps

Chapter 4

Discussion

4.1 Machine Learning models comparison

4.1.1 Score comparison

How well do the models predict values in the test dataset and how do the models compare with each other.

4.1.2 Spread Comparison

How much spread can we expect across different slpits between train and test data for all the model.

4.1.3 Feature importance comparison

Do different models give more importance to different features when making predictions?

4.2 Alps volume comparison

How do the predicted volumes for the whole alps from the machine learning models compare to the volumes predicted by some physical based models.

Chapter 5

Conclusions

To be decided

Appendix A

Tuning models parameters?

If needed could be used to explain for example choosing parameters etc.

Bibliography

Farinotti, D., et al., 2017: How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment. *Cryosphere*, **11 (2)**, 949–970, doi:10.5194/tc-11-949-2017.

Acknowledgments