ARIEL UNIVERSITY

MASTER THESIS SUBMISSION

A Machine Learning Approach for Predicting Water Retention Curves of Porous Materials

Author:

Or Haim Anidjar

Supervisors:

Dr. Amos Azaria

Dr. Noam Hazon

Dr. Arcady Beriozkin

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in the

Department of Computer Science

February 28, 2019

Declaration of Authorship

I, Or Haim Anidjar, hereby declare that this thesis proposal entitled, "**A Machine Learning Approach for Predicting Water Retantion Curves of Porous Materials**" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Or Haim Anidjar

Date: February 28, 2019

Ariel University

Abstract

Faculty of Natural Sciences

Department of Computer Science

Master of Science

A Machine Learning Approach for Predicting Water Retantion Curves of Porous Materials

by Or Haim Anidjar

In this thesis, we explore an existing linear inter-functional relationship, between two families of porous media characteristic curves: a wetting curve, and a drying curve, in order to develop a model predicting a function of family B, from a known function of A and a dataset consisted from couples of curves from families A and B.

This study concerns with developing a novel machine learning approach to modeling the relationship between two families of pairwise associated functions. This relationship includes both regular and random constituents, and it is represented by a limited set $\mathcal{L}$ of N known pairs of associated functions, a function of the first family and a function of the second family in each pair. The suggested approach is applied to the subject of capillary hysteresis, for predicting the boundary drying function (BDF), pertaining to the second family, from the known boundary wetting function (BWF), pertaining to the first family, resting on the regularity found in training set $\mathcal{L}$ ($\mathcal{L} \subseteq \overline{\mathcal{L}}$) of k ($k \leq N$) pairs. Set $\mathcal{L}$ is defined on the principle of k nearest neighbors in the sense that the k functions of the first family within set $\mathcal{L}$ appear to be the nearest neighbors of the known BWF, associated with the sought BDF. Prediction of the desired BDF from its associated known BWF is obtained as a product of two mappings: (a) a nonlinear mapping of the known BWF (belonging the first family) to its corresponding hypothetical drying function (as defined in the hysteresis theory of Mualem, 1984), and (b) a linear mapping of this latter function to the desired BDF (belonging the second family). The latter mapping is based on (i) optimization of the k functions of the second family within set $\mathcal{L}$, using the leave-one-out cross-validation procedure, in order to minimize the influence of the statistical scattering of the pairs on the predictive reliability, and (ii) obtainment of the sought BDF as a linear combination of the optimized k functions. The predicted boundary drying curves

indicate a generally acceptable agreement with the measured ones. An important advantage of the proposed approach is a possibility of permanent updating the suggested predictive model by incorporating new measured data, what enhances its trustworthiness.

Keywords: Hysteretic Systems, Machine Learning, Supervised Learning Algorithms, Linear Combination, k-Nearest Neighbors.

Contents

List of Figures

Dedicated to my loved ones.

Chapter 1

Introduction

In this thesis, we develop a predictive machine learning model describing a linear inter-functional relationship between two families of porous media characteristic curves: wetting curve, denoted by $S_w(\psi)$, and drying curves, denoted by $S_d(\psi)$. Both of the curves are representing a state of water saturation S, as function of capillary pressure (denoted by ψ). The suggested approach is a deterministic one in sense that different characteristic functions of each individual porous medium (such as $S_w(\psi), S_d(\psi)$ being water permeability, as well as $K(S)$, which defines a grain size distribution curve) are deterministically interrelated to each other, since they are all defined by the same pore space topology.

Our work focuses on the prediction of a $S_d(\psi)$ curve, given a $S_w(\psi)$ one, based on a dataset consisted from N couples of $\langle S_{w_i}(\psi), S_{d_i}(\psi)\rangle_{i=1}^{N}$ curves, by learning the transformation G between them (as presented in Figure 1.1). As will be presented later (in Chapter $\#2$), the relationship between $S_w(\psi)$ and $S_d(\psi)$ curves is not linear, hence we will use in an intermediate curve (denoted by $S_d^{\circ}(\psi)$), which can be easily computed from $S_w(\psi)$, and defined as a hypothetical drying curve, assuming that all of the hysterons are parallely connected, i.e. independent of each other. Since such a scenario is almost surreal, the $S_d^{\circ}(\psi)$ curves are serving as a baseline for prediction of a $S_d(\psi)$ curve, and in Chpater $\#2$ we will prove that there exist a linear relationship between $S_d(\psi)$ and $S_d^\circ(\psi)$ curves, and will explain how to convert the data set into $S_d^\circ(\psi)$ curves, in order to exploit this relationship.

The two curves $S_w(\psi), S_d(\psi)$ are creating a capillary hysteresis loop, described physically by presentation of the porous space as a system of a large number of elementary hysterons, together with the hypothetical drying curve, in Fig.[1.2].

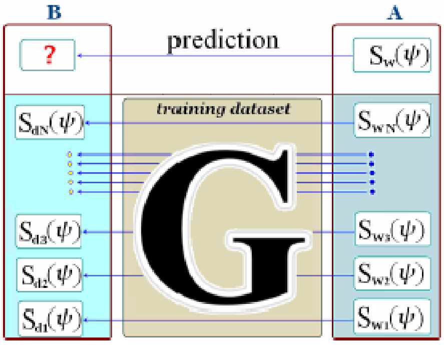

FIGURE 1.1: Illustration of the dataset consisted from n wetting and drying curves, where final goal is to predict a new drying curve, based on its matching wetting curve and known data.

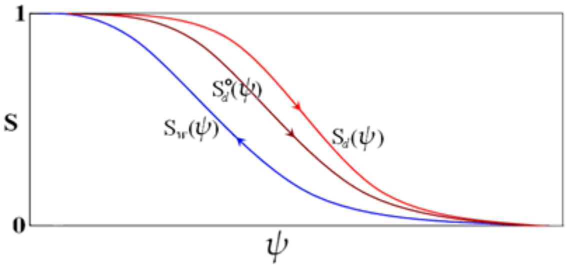

FIGURE 1.2: Schematic representation of the hysteretic loop created by wetting (blue) and drying (red) curves $S_w(\psi)$ and $S_d(\psi)$, respectively, and hypothetical drying curve (brown) $S_d^\circ(\psi)$ according to Mualem (1984). Arrows show the directions of processes.

A hysteron, defined as an element having two parameters: (1) entry and (2) exit, is a fundamental element in hysteretic systems. In capillary hysteresis, the entry parameter is a pressure of water filling ψ_w - wetting, whereas the exit parameter, is a pressure of emptying ψ_d - drying (see Figure 1.3). Of these two parameters, the drying pressure is higher.

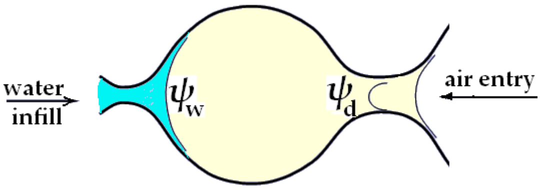

FIGURE 1.3: Illustration of a capillary hysteron. Capillary pressure value ψ_w is associated with the position of the "water-air" interface directly preceding the infill; Capillary pressure value ψ_d is associated with the position of the "water-air" interface directly preceding the air entry.

A system of a large number of hysterons would behave hysteretically even if all of the hysterons will be independent of each other. However, the real hysteresis is manifested much more pronouncedly, because of interdependence of the hysterons. When two hysterons are connected in series, then one of them can be blocked by another. By arranging all of the hysterons on the plane (ψ_d, ψ_w), we obtain a probability density function of the hysteron volume distribution, denoted by $f(\psi_d, \psi_w)$. The domain of definition of $f(\psi_d, \psi_w)$, is a triangle on (ψ_d, ψ_w) plane, where $\psi_d \geq \psi_w$.

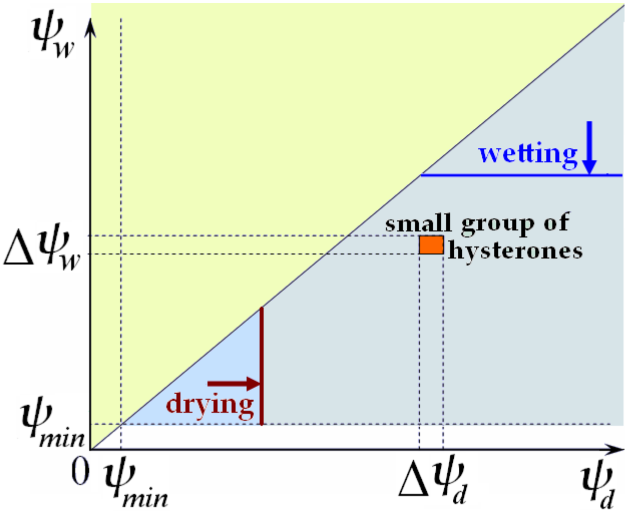

FIGURE 1.4: Domain of definition of $f(\psi_d, \psi_w)$.

Using one of the admissible assumptions regarding the volume distribution of hysterons (e.g. hysteresis model of Mualem, 1984), the function $f(\psi_d, \psi_w)$ can be determined by $S_w(\psi)$ only. Knowledge of function $f(\psi_d, \psi_w)$ is necessary for describing the function $S_d(\psi)$. However, the function $f(\psi_d, \psi_w)$ is not sufficient for this purpose, because of inter-dependence of the hysterons manifested as the blockage phenomenon. If we assume independence of the hysterons (i.e. their parallel connection), the usage of function $f(\psi_d, \psi_w)$ leads to the hypothetical drying curve $S_d^\circ(\psi)$ that underestimates the real extent of the hysteretic loop.

The objective of the suggested work is to develop a machine learning model predicting the main drying curve $S_d(\psi)$ from the known main hypothetical drying curve $S_d^\circ(\psi)$, which is inferred from its appropriate $S_w(\psi)$ curve. This model should take advantage of the ability to convert the $S_w(\psi)$ curves into $S_d^\circ(\psi)$ ones, in order to explore the linear relationship between the two families of functions: $S_d^\circ(\psi)$ and $S_d(\psi)$. The solution is based on finding an optimal linear combination of a number of hypothetical drying curves $S_d^\circ(\psi)$, with coefficients $\{k_i\}_{i=1}^t$ (for some natural number t, which represent the number of vectors considered on linear combination) in order to use the same coefficients for predicting $S_d(\psi)$ curves. To do this, we suggest a supervised machine learning algorithm, namely $k - Nearest - Neighbors$ by computing the t nearest neighbors given a specific $S_d^\circ(\psi)$ curve, and then integrate linear algebra tools into the model, by exploiting the coefficients extracted from the linear combination, and then use them on predicting a $S_d(\psi)$ curve.

Clearly, exploitation of the relationship between $S_w(\psi)$ and $S_d(\psi)$ curves, which is based on the linear relationship between $S_d^\circ(\psi)$ and $S_d(\psi)$ curves, must be mathematically proven, and will be presented in Chapter $\#2$.

Chapter 2

Linear Relationship Between Source and Image Functions

The following section will be dedicated for explaining the linear relationship between the discussed source and image functions, then we will give a direct proof for a conversion of those families of functions into another such a families, that is denoted by U.

The HDF (Figure 1.2) obtained in the hypothetical drying process when assuming mutual independence of the hysterons, is calculated according to the hysteresis theory of Mualem (1984):

$$S_d^{\circ}(\psi) = 2S_w(\psi) - S_w^2(\psi) \tag{2.1}$$

Suppose we have at our disposal a set $\overline{\mathcal{L}}$ of N porous media, each of them with its pair of known functions $S_{w_i}(\psi)$ and $S_{d_i}(\psi)$ $(i = 1, 2, ..., N)$. Let set $\overline{\mathcal{L}}$ be a representative sample of an infinite collection of porous media. Therefore, this collection defines three families of functions: the boundary wetting functions (BWF), the hypothetical drying functions (HDF) and the boundary drying functions (BDF). The functions of these families (as presented in Figure 2.1) are monotonically decreasing from 1 to 0 with zero derivative (from right) at $\psi = 0$ (Figure 1.2) and single inflection point. As follows from

Eq.(2.1), the relationship between the families of BWF and HDF is a non-linear mapping. Regarding the relationship between the families of HDF and BDF, in the following we consider a linear mapping between them, including its physical meaning.

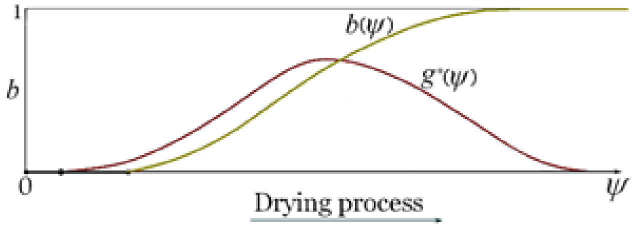

FIGURE 2.1: Schematic representation of differential water capacity function $g^{\circ}(\psi)$ and blockage function $b(\psi)$.

The hysteresis between $S_w(\psi)$ and $S_d^{\circ}(\psi)$ is explained by the cumulative effect of individual hysteretic behaviors of all the hysterons constituting the porous space, when assuming their mutual independence. The observed hysteresis loop, created by $S_w(\psi)$ and $S_d(\psi)$ is significantly wider (Figure 1.2) than that existing between $S_w(\psi)$ and $S_d^{\circ}(\psi)$, because actually in the drying process an interdependence of some part of the hysterons is involved, due to specific morphology of the porous medium. This interdependence has a form of blockage phenomenon occurring when several hysterons prevent their adjacent ones from emptying. The extent of the blockage phenomenon is expressed by the blockage function defined by Mualem (1984) as:

$$b(\psi) = \frac{1 - S_d(\psi)}{1 - S_d^{\circ}(\psi)} \tag{2.2}$$

Define also a differential water capacity function (DWC):

$$g^{\circ}(\psi) = -\frac{dS_d^{\circ}(\psi)}{d\psi} \tag{2.3}$$

that has a meaning of a probability density function. Functions $b(\psi)$ and $g^{\circ}(\psi)$ are illustrated in Fig. [2.1]. Using Eq.(2.3) we rewrite Eq.(2.2) as follows:

$$1 - S_d(\psi) = \int_0^{\psi} b(\psi) g^{\circ}(\varphi) d\varphi \tag{2.4}$$

For function $b(\psi)$ we define its bivariate extension $B(\varphi, \psi)$:

$$B(\varphi, \psi) = \begin{cases} b(\psi) & \varphi \leq \psi \\ 0 & \varphi > \psi \end{cases} \tag{2.5}$$

which is illustrated in the next figure (Figure 2.2):

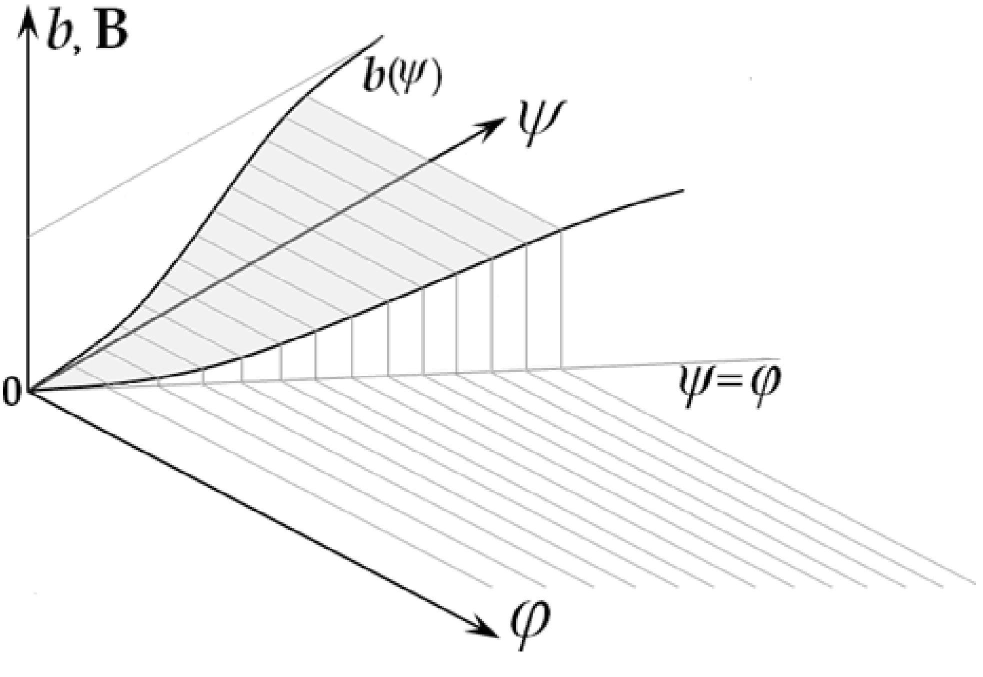

FIGURE 2.2: Schematic illustration of function $B(\varphi, \psi)$.

According to Eq.(2.5) we have $b(\psi) = B(0, \psi)$. Function $B(\varphi, \psi)$ enables to rewrite Eq.(2.4) for each i^{th} medium in the following form:

$$1 - S_{d_i}(\psi) = \int_0^{\overline{\psi}} B_i(\varphi, \psi) g_i^{\circ}(\varphi) d\varphi \quad (i = 1, 2, .., N) \tag{2.6}$$

where $\overline{\psi}$ denotes a minimal value, common for all of the N media, at which every $S_{d_i}(\psi)$ as well as every $g_i^\circ(\psi)$ attains zero.

Taking into account that $\int\limits_0^{\psi} g_i^\circ(\varphi)d\varphi = 1$ ($i = 1, 2, .., N$), we obtain from Eq.(2.6):

$$S_{d_i}(\psi) = \int\limits_0^{\overline{\psi}} [1 - B_i(\varphi, \psi)] g_i^\circ(\varphi) d\varphi \quad (i = 1, 2, .., N) \tag{2.7}$$

Along with the N individual functions $1 - B_i(\varphi, \psi), (i = 1, 2, .., N)$, there exists a set of kernel functions equally satisfying N equations (2.7) at once. Function $D(\varphi, \psi)$ of the form:

$$D(\varphi, \psi) = \sum_{j=1}^{N} \lambda_j(\psi) h_j(\varphi) \tag{2.8}$$

will belong to this set if it would satisfy N equations:

$$S_{d_i}(\psi) = \int\limits_0^{\overline{\psi}} [\sum_{j=1}^{N} \lambda_j(\psi) h_j(\varphi)] g_i^\circ(\varphi) d\varphi \quad (i = 1, 2, .., N) \tag{2.9}$$

where $h_j(\varphi), (j = 1, 2, .., N)$ are are the first N elements of a certain orthonormal basis (e.g. the Fourier trigonometric basis, Legendre polynomials) and $\lambda_j, (j = 1, 2, .., N)$ are the unknown functions to be found from this system of N equations.

Denote α_{ij} by:

$$\alpha_{ij} = \int\limits_0^{\overline{\psi}} g_i^\circ(\varphi) h_j(\varphi) d\varphi \tag{2.10}$$

With this designation Eqs.(2.9) are rewritten into the following system of linear algebraic equations:

$$S_{d_i}(\psi) = \alpha_{i1}\lambda_1(\psi) + .. + \alpha_{ij}\lambda_j(\psi) + .. + \alpha_{iN}\lambda_N(\psi) \tag{2.11}$$

that yields solution $\lambda_j(\psi), (j = 1, 2, .., N)$ for each prescribed value of ψ. Therefore, we have the linear mapping:

$$S_{d_i}(\psi) = \int\limits_0^{\overline{\psi}} D(\varphi, \psi) g_i^\circ(\varphi) d\varphi \quad (i = 1, 2, .., N) \tag{2.12}$$

Function $D(\varphi, \psi)$ can be thought of as a generalized modified blockage function, common for the set of the N measured media. Other functions of the above set of kernel functions can be written in the form $\sum\limits_{j=1}^{M} \lambda_j(\psi) h_j(\varphi)$, where $M > N$ and functions $\lambda_j(\psi), (j = 1, 2, .., M)$ satisfy the following undetermined system of equations:

$$S_{d_i}(\psi) = \alpha_{i1}\lambda_1(\psi) + .. + \alpha_{ij}\lambda_j(\psi) + .. + \alpha_{iM}\lambda_M(\psi) \quad (i = 1, 2, .., N) \tag{2.11*}$$

The graph of the operator in Eq.(2.12) passes via all the N pairs $(g_i^\circ(\varphi), S_{d_i}^\circ(\psi))$, $(i = 1, 2, .., N)$ scattered around the supposed actual interfunctional DWCF-BDF regularity.

Therefore, the operator must be too sensitive with respect to alternate exclusion or replacement of a one or a few pairs, what implies its low predictive reliability.

Based upon the linear operator in Eq.(2.12) acting from the span of DWCF family to the span of BDF family, we can proceed to determining a direct linear mapping between the HDF and BDF families. Suppose, we have to predict BDF for some porous medium based on its DWCF, the latter derived from the known BWF by Eqs.(2.1) and (2.3).

Denote this DWCF by $g_p^\circ(\psi))$. Consider k $(k \leq N)$ functions of $g_i^\circ(\psi))$ $(i = 1, 2, ., k)$, that are the k nearest neighbors of $g_p^\circ(\psi))$ taken from the N known DWCF pertaining to the given dataset of the N measured media. The 'nearness' is defined by the Euclidean norm. Let function $\widetilde{g}_p^\circ(\psi)$ be the best normalized approximation of $g_p^\circ(\psi)$, that can be attained by linear combination of $g_i^\circ(\psi)$, $(i = 1, 2, .., k)$, i.e.

$$\widetilde{g}_p^\circ(\psi) = \omega_1 g_1^\circ(\psi) + .. + \omega_k g_k^\circ(\psi) \tag{2.13}$$

with coefficients $\omega_i(i = 1, 2, .., k)$ such that:

$$\sum_{i=1}^{k} \omega_i = 1 \tag{2.14}$$

The image of $g_p^\circ(\psi)$ by the linear operator in Eq.(2.12) is:

$$\widetilde{S}_{d_p}(\psi) = \omega_1 S_{d_1}(\psi) + .. + \omega_k S_{d_k}(\psi) \tag{2.15}$$

On the other hand, integrating Eq.(2.13):

$$\int\limits_0^\psi \widetilde{g}_p^\circ(\varphi) d\varphi = \omega_1 \int\limits_0^\psi g_1^\circ(\varphi) d\varphi + .. + \omega_k \int\limits_0^\psi g_k^\circ(\varphi) d\varphi \tag{2.16}$$

yields:

$$\widetilde{S}_{d_p}^\circ(\psi) = \omega_1 S_{d_1}^\circ(\psi) + .. + \omega_k S_{d_k}^\circ(\psi) \tag{2.17}$$

where Eq.(2.14) is taken into account. The linear combinations (2.15) and (2.17) have the same coefficients $\omega_i(i = 1, 2, .., k)$, therefore instead of operator (2.12) we can consider the direct linear mapping from the span of HDF family to the span of BDF family. Accordingly, instead of seeking the best normalized approximation of function $g_p^\circ(\psi)$ (Eq. 2.13) by its k nearest neighbors we can seek the best normalized approximation of function $\widetilde{S}_{d_p}(\psi)$ by the k nearest neighbors of the latter. The obtained coefficients $\omega_i(i = 1, 2, .., k)$ appearing in Eq.(2.17) should be used for determining function $\widetilde{S}_{d_p}(\psi)$ by Eq.(2.15), as presented in Fig.[2.3]. Just as the operator in Eq.(2.12) cannot serve as a reliable predictive tool because of its oversensitivity, function $\widetilde{S}_{d_p}(\psi)$ cannot be

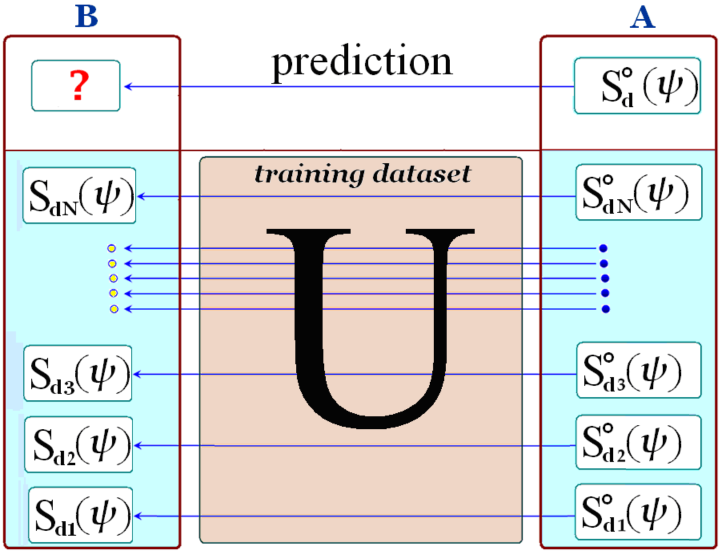

FIGURE 2.3: Illustration of dataset conversion (denoted by the transformation U), from HDF family to BDF family.

considered as a reliable prediction of the sought BDF. Therefore, the following stage of this study is an optimization procedure necessary for reducing of influence of the statistical scattering of HDF-BDF pairs around the actual regularity.

So far, after a brief introduction to our work, and demonstration of how is it possible to convert functions families $\{g_i^\circ(\psi)\}_{i=1}^n$ and $\{1 - S_{d_i}(\psi)\}_{i=1}^n$ into functions families $\{S^\circ_{d_i}(\psi)\}_{i=1}^n$ and $\{S_{d_i}(\psi)\}_{i=1}^n$ (respectively), Chapter #3 would be dedicated for related work, whereas in Chapter #4 will present a description of our soultion method, and Chapter #5 would show a set of experiemental results, comparing different baseline methods to our approach. Finally, we present a discussion in Chpater #6, then summary and conclusions in Chpater #7.

Chapter 3

Related Work

As mentioned earlier, the main idea behind our problem's solution combines supervised machine learning models, and common linear algebra principles. Whereas our work concentrates on the prediction of drying curves based on the wetting curves, Lamorski et al.(2014, 2016) have developed a machine learning based model, namely Support Vector Machine (SVM), for the prediction of the inverse problem, i.e. the prediction of the wetting curves based on the drying curves. They have estimated the main wetting branch of the Soil Water Retention Curve (SWRC) based on the knowledge of the main drying branch and other, optional, basic soil characteristics such as Particle Size Distribution (PSD), Bulk Density (BD), Organic Content (OC), and Soil Specific Surface (SSS). The construction was consisted of different sets of input parameters for each SVM model. All of the models used information related to the drying branches of SWRC's, by fitting the Van Genuchten(2005) model (a method of representation of water retention curves) to measured retention data points. This process resulted in the extraction of 15 features. Some of the models have used additional soil characteristics as input parameters (i.e., PSD, BD, OC, and SSS), as well as soil physical parameters, since they were correlated with the wetting branch of the SWRC. Our suggested approach of using k-NN for inter-functional prediction, is related to the extraction of linear combination coefficients. As far as we know, no usage of k-NN has ever served for inter-functional regression using linear combination technique and data synthesizing. Though, there are some works that are involved with extraction of linear combination coefficients in order to improve the accuracy of predictive models. E.g is the one of Xu et al.(2012), that have shown a method for determining the weights of linear combination coefficients in order to achieve better accuracy of different predictive models, alike monthly mean rainfall, and aviation engine's residual life; Fu et al.(2016) have predicted the coefficients of a linear combination, by changing the distribution of an original FASTQ (a text-based format for storing both a biological sequence and its corresponding quality scores) file, through a linear combination prediction for an improved compression, using existing compression algorithms. The main output of this study is the prediction of the $S_d(\psi)$ curve (BDF), given the $S_w(\psi)$ curve (BWF). This prediction can be seen as a simple multi dimensional regression problem, yet, the regression output is a multi-dimensional vector on which every coordinate is inter-related with its neighbor coordinates, using the optimization algorithm that has been developed in this work. An example for a multi-label classification problem has seen light in Wan et al.(2017), on which they developed an ensemble transductive learning method to tackle the multi-label classification problem, in predicting the multi localization of chloroplast proteins at the sub-subcellular level. More specificly, given a protein in a dataset, its composition-based sequence information and profile-based evolutionary information, has been compared with those of other proteins in the dataset. The comparison led to two similarity vectors which are weighted-combined to constitute an ensemble feature vector. Then, a transductive learning model based on the least squares and nearest neighbor algorithms have been proposed in order to process the ensemble features. To the best of our knowledge, our work is the first to combine multi-dimensional and inter-functional regression with data synthesizing, based on a given small sampled dataset.

Chapter 4

Our Approach

As main approach, the algorithm first transforms all of the wetting curves $\{S_{w_i}(\psi)\}_{i=1}^{n}$ into hypothetical drying curves $\{S^{\circ}_{d_i}(\psi)\}_{i=1}^{n}$. The algorithm then picks the k nearest hypothetical drying curves in order to re-represent each of the hypothetical drying curves $\{S^{\circ}_{d_i}(\psi)\}_{i=1}^{n}$. Following to that, the algorithm computes a coefficients series that minimizes the L_2 norm of the difference between a given hypothetical drying curve, and its representation as a linear combination resulted by the computed coefficients. This step results in the extraction of a coefficeints series $\{k_i\}_{i=1}^{k}$ for the presentation of the drying curve's prediction, with the same coefficient series, applied on the k nearest **drying** curves. First, we will present two baseline methods for linear combination's computation that lays under the construction of the method proposed in this study;

4.1 Baseline Methods

In the simplest usage of k-NN model that mentioned above, we are given an $S^{\circ}_{d}(\psi)$ curve, then the algorithm looks for the k nearest hypothetical drying curves that serving as neighbors, and sets the prediction of $S_d(\psi)$ by the adequate drying curves being equally weighted, i.e. the weight of each drying curve in the representation of $S_d(\psi)$ as a linear combination of drying curves is exactly $\frac{1}{k}$. In the more sophisticated method, the algorithm finds a coefficient series that serve as a linear combination and minimize the L_2 norm of the difference between a given $S_d^\circ(\psi)$ curve, and its representation as a linear combination of its k nearest hypothetical drying curves. Then, the algorithm applies the coefficient series on the drying curves, namely the algorithm tries to minimize the L_2 norm of the difference between the real drying curve $S_d(\psi)$, and its prediction which is represented as a linear combination of its k nearest drying curves.

4.2 Drying Curves Optimization

The optimization is aimed to reduce the influence of statistical scattering of the HDF-BDF pairs around the actual regularity. Accordingly, the objective of the optimization procedure is to create N synthetic boundary drying curves, such that when substituted into Eq. (2.17) instead of $S_{d_i}(\psi)$ $(i = 1, 2, .., N)$, they would yield a reliable prediction $S_{d_p}(\psi)$, for the desired BDF of a porous medium with only measured HBDF, $S_{d_p}^\circ(\psi)$, as follows:

$$S_{d_p}(\psi) = \omega_{p_1}\widehat{S}_{d_1}(\psi) + ... + \omega_{p_k}\widehat{S}_{d_k}(\psi) \tag{4.1}$$

Here, $\widehat{S}_{d_i}(\psi)$ $(i = 1, 2, .., N)$ denote the synthetic boundary drying curves, and the coefficients $\omega_{p_1}, \omega_{p_2}, .., \omega_{p_k}$ are the same as obtained for approximation of the HBDF by Eq.(2.17). To describe the synthetic boundary drying curve of an i^{th} medium we derived the following one-parametric expression:

$$\widehat{S}_{d_i}(\psi) = S_{d_i}(\psi) + \beta(\frac{S_{d_i}(\psi)}{1 - S_{d_i}(\psi) + S_{d_i}^2(\psi)} - S_{d_i}(\psi)) \tag{4.2}$$

where parameter β can vary in the range [-1,1]. For initiating the optimization procedure, we need:

i . To provide the N pairs of known HDF-BDF.

ii . To compose for each i^{th} medium $(i = 1, 2, .., N)$ the list of its k nearest neighboring media, altogether N lists. The number k of the nearest neighbors, found as the optimal number, is assumed to be the same for all N media.

iii . To find for each i^{th} HDF the best approximation by linear combination (Eq. 2.17) of its k nearest neighbors:

$$S^{\circ}_{d_i}(\psi) \sim \omega_{i_1} S^{\circ}_{d_1}(\psi) + ... + \omega_{i_k} S^{\circ}_{d_k}(\psi), \quad (i = 1, 2, .., N) \tag{4.3}$$

iv . To compose for each i^{th} medium $(i = 1, 2, .., N)$ a list of those media for whom this i^{th} medium is a neighbor, altogether N lists. The number of such media, that have an ith medium as a neighbor, depends on i. This number is denoted hereafter by J_i.

v . To define a routine returning for an i^{th} medium $(i = 1, 2, .., N)$ the current form of the synthetic boundary drying curve:

$$\breve{S}_{d_i}(\psi) = S_{d_i}(\psi) + \beta(\tfrac{S_{d_i}(\psi)}{1 - S_{d_i}(\psi) + S^2_{d_i}(\psi)} - S_{d_i}(\psi)), \quad (-1 \leq \beta \leq 1) \tag{4.2*}$$

Calls to this routine should be made in the course of greedy search of the optimal values of parameter β needed for each i^{th} $(i = 1, 2, .., N)$ synthetic boundary drying curve. The final optimized form of the synthetic boundary drying curve is denoted by $\widehat{S}_{d_i}(\psi)$, i.e. $\breve{S}_{d_i}(\psi)$ tends to $\widehat{S}_{d_i}(\psi)$ during the optimization procedure.

vi . To define a routine, returning for each iterative current value of parameter β (Eq. 4.2*) pertaining to an i^{th} medium, the current transient predictions of the BDF of all J_i media for whom this i^{th} medium is a neighbor. In particular, the current prediction of the BDF for an l^{th} medium $(l = 1, 2, .., J_i)$, for which the i^{th} medium as a neighbor, we have:

$$S^p_{d_l}(\psi) = \omega_{l_1} \breve{S}_{d_1}(\psi) + .. + \omega_{l_m} \breve{S}_{d_m}(\psi) + .. + \omega_{l_k} \breve{S}_{d_k}(\psi) \tag{4.1*}$$

for $(m = 1, .., k, l = 1, ..J_i)$, where $S^p_{d_l}(\psi)$ designates the current prediction of the BDF for an l^{th} medium, $m = 1, 2, .., k$ indicate sequential numbers of the k nearest neighbors of the l^{th} medium. Accordingly, one of these sequential numbers is related to the i^{th} medium.

4.2.1 Training

The algorithm performs a greedy search of the values of parameter β (Eq. 4.2*) for each of the N media. The sequence of tasks, to be carried out with respect to every i^{th} medium, can be described by example of the 1^{st} medium. The tasks are performed with respect to each iterative current value of β, and they are as follows:

1. The parameter β is assigned a new iterative value of β.

2. The algorithm calculates the current form of the synthetic boundary drying curve $\breve{S}_{d_1}(\psi)$ (Eq. 4.2*) corresponding to the above current value of β.

3. The current form of $\breve{S}_{d_1}(\psi)$, in its turn, invokes computation of transient predicted BDF from (Eq. 4.1*), $S^p_{d_l}(\psi)$ $(l = 1, .., J_1)$, for the media that have the 1_{st} medium as a neighbor.

4. Each l^{th} computed transient predicted BDF $(l = 1, .., J_1)$ is compared with the corresponding actual BDF, and the algorithm calculates the current prediction error:

$$\rho_l = ||S_{d_l}(\psi) - S^p_{d_l}(\psi)||, (l = 1, .., J_1) \tag{4.4}$$

   where symbol $||\cdot||$ indicates the Euclidean norm.

5. Having the above J_1 values of the current prediction error (Eq. 4.4) the algorithm calculates the current cumulative prediction error for the

current iterative value of β:

$$\overline{\rho} = \sum_{l=1}^{J_1} ||S_{d_l}(\psi) - S^p_{d_l}(\psi)|| \tag{4.5}$$

Quantity $\overline{\rho}$ serves as an objective function of the optimization procedure. The tasks $1-5$ are carried out for each iteration of β until $\overline{\rho}$ is minimized. Once $\overline{\rho}$ is minimized, the corresponding value of β is saved as a temporary optimal value pertaining to the 1^{st} medium, and the algorithm proceeds to the 2^{nd}, 3^{rd} and so forth up to the N^{th} medium. After obtainment of the N temporary optimal values of β, each with respect to its medium, the above tasks $1-5$ have to be repeated (for all N media) several more times, because some of these originally obtained values of β (especially a few first of them) turn out to be no longer optimal due to subsequent changes in the synthetic boundary drying curves. Eventually, after running the algorithm several times along the tasks $1-5$, the process converges and the N obtained values of β as well as the N corresponding synthetic boundary drying curves are saved as the optimal values of β and the final synthetic boundary drying curves $\widehat{S}_{d_i}(\psi)$, respectively.

4.2.2 Test

Suppose we need to predict the BDF, $S_{d_p}(\psi)$, from the known BWF, $S_{w_p}(\psi)$, of some medium, not belonging to the dataset. The prediction of the missing curve is obtained by the following steps:

1. Obtainment of the HBDF $S^\circ_{d_p}(\psi)$, from the known BWC using Eq. (2.1).
2. Finding the k nearest neighbors of function $S^\circ_{d_p}(\psi)$, from among the N known function $S^\circ_{d_i}(\psi)$ $(i = 1, 2, .., N)$.

3. Finding the best approximation of $S^\circ_{d_p}(\psi)$ by a linear combination of the above k nearest neighbors (Eq. 2.17):

$$S^\circ_{d_p}(\psi) \sim \omega_{p_1} S^\circ_{d_1}(\psi) + ... + \omega_{p_k} S^\circ_{d_k}(\psi) \tag{4.6}$$

4. Obtainment of the desired prediction of the HBDF:

$$S_{d_p}(\psi) = \omega_{p_1} \widehat{S}_{d_1}(\psi) + ... + \omega_{p_k} \widehat{S}_{d_k}(\psi) \tag{4.7}$$

Test phase can be described by Algorithm[1] which follows next;

Algorithm 1 Drying Curve Prediction

Input: Set of $n \in \mathbb{N}$ hystertic loops, $k \in \mathbb{N}$ nearest neighbors, test wetting curve $S_{w_p}(\psi)$
Output: Drying curve $S_{d_p}(\psi)$ prediction

```
$S_{d_p^\circ}(\psi) \leftarrow 2S_{w_p}(\psi) - S^2_{w_p}(\psi)$
$S_{d_p^\circ}(\psi) \leftarrow \sum_{i=1}^{k} k_i \cdot S_{d_i^\circ}(\psi)$
for $1 \leq i \leq k$ do:
    $\beta_i \leftarrow Load\_Betha\ (S_{d_i}(\psi))$
    $\widehat{S_{d_i}}(\psi) \leftarrow S^\circ_{d_i}(\psi) + \beta(\frac{S^\circ_{d_i}(\psi)}{1-S^\circ_{d_i}(\psi)+S^\circ_{d_i}(\psi)^2} - S^\circ_{d_i}(\psi))$
end for
$S_{d_p}(\psi) \leftarrow \sum_{i=1}^{k} k_i \cdot \widehat{S_{d_i}}(\psi)$
return $S_{d_p}(\psi)$
```

Chapter 5

Experiements and Results

We developed an experimental platform to validate the hypothesis that using synthetic drying curves for predicting the correspondant drying curve of a given weeting curve, outperforms the traditional methods. We have found intresteting results that confirms the research hypothesis and the mathematical developments, after collecting a dataset of size $N = 21$ media sampled. It is important to pay attention that the variance between the presented hysteretic loops in each figure is reflected in the following properties: (i) the thickness of the loop, i.e. the distance between the wetting and drying curves, and (ii) the values of the capillary pressure ψ on which the wetting and drying curves are saturated. The workflow of the experiements was as follows;

5.1 Data Preparation

The measured hysteresis loops have been collected from the literature worldwide. There are $N = 21$ porous media used in this study: Rubicon sandy loam (Topp, 1969), Caribou silt loam (Topp, 1971), Norfolk sandy loam (Hopmans and Dane, 1986), Bloomfield sand (Bruce and Klute, 1963), Avondale clay loam (Watson et al., 1975), Adelaide dune sand (Talsma, 1970), Del Monte fine sand (Liakopoulos, 1966), Rideau clay loam (Topp, 1971), Fontaine bleau sand (Bruce and Klute, 1956), Mason county fine sand (Bruce and Klute, 1963), Sandy soil (Santini, 1981), Berea sandstone (Raeesi et al., 2014), Hopman (Mualem, 1976), Boise sandstone (Raeesi et al., 2014), Silica fine F-95 sand (Muraleetharan et al., 2009), Borden sand (Demond and Roberts, 1991), Chiba sandy soil (Gallage et al., 2013), Fine sand F-100 (Hong et al., 2016), Plainfield sandy loam (Nimmo and Miller, 1986), Sand 50-500 micron (Jackson et al., 1965), Glass beads 75 Micron (Bruce and Klute, 1963), Bentheimer sandstone (Ruspini et. al., 2017), Uniform fine quartz sand (Ray and Morris, 1995), and Aggregated glass beads 5003 (Mualem, 1976).

The dataset was represented as a set of images alike the one seen in Fig.[1.2]. The only difference between Fig.[1.2] and the dataset is the range of ψ axis on each image, that varied between different ranges. As a preprocess step for the data preparation, we have sampled by using a digitizing program, each of the drying and wetting curves in different locations in order to create a skeleton of each curve. Then, we have used spline curves in order to reconstruct an approximation for the original shape of each curve, so that each curve was represented as a discrete vector of size 451. The reason for this length, is that we have found that whenever $\psi \leq 0$, it holds that $S_w(\psi) = S_d(\psi) = 1$, since when there is no capillary pressure - saturation is full, i.e. equals 1. On the other hand, we have found for all of the 21 media, that whenever $\psi > 451$, it holds that $S_w(\psi) = S_d(\psi) = 0$, means that saturation is empty. As a result, we have created for each medium a couple of vectors of size 451 - $S_w(\psi), S_d(\psi)$ for $0 \leq \psi \leq 450$ in spaces of 1. By Mualem addmissible assumption (Eq. 2.1), which claims that $S_d^{\circ}(\psi) = 2S_w(\psi) - S_w^2(\psi)$, we have created the adequate S_d° vector for each medium, on the same range of ψ.

The data is available at the following link: Hysteretic Loops Data

5.2 Comparative Methods

When having in hand $S_w(\psi), S_d^\circ(\psi), S_d(\psi)$ curves for each medium, we have constructed the following seven different models, such that the difference between them was expressed by (i) the source and image functions (i.e. from wetting or hypothetical drying curves, to drying curves), and (ii) the computation of the linear combination's coefficient series.

The first model represents our solution, and its name is **Hypothetical To Drying With Beta** (HTDWB) - in this main model, the source functions are the hypothetical drying curves, and the image functions are the drying curves. This main model is based on the construction from the Drying Curves Optimization phase (presented in subsection 3.1.2), and outperformed all of the models presented in this work. As can be seen in Fig.[5.1], HTDWB outperforms all the other six baselines models, in that the standard deviation is so low (0.015), implying that our method is very robust to the number of neighbors selected. The standard deviation σ for each model was calculated as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{X})^2} \tag{5.1}$$

Where N is the number of media, k is the neighbors amount considered in the experiements, x_i represents the sum of predictions error (defined later in Eq. 5.2) for a given number of neighbors i, and $\overline{X}$ is the average sum of predictions error.

The rest of the models have served as baseline and comparison models, as follows:

1. **Hypothetical To Drying Linear Combination** (HTDLC) - Same as HTDWB, but differs in the optimization phase which is missing. In this

model, the linear combination (LC) minimizes the L_2 norm of the difference between a given $S_d^\circ(\psi)$ curve, and its representation as a linear combination of its k nearest hypothetical drying curves.

2. **Hypothetical To Drying Equal** (HTDE) - This model differs from HTDWB and HTDLC models, in the coefficient series extraction, and has no optimization process. After finding of k nearest hypothetical drying curves for a given $S_d^\circ(\psi)$ one, the weight of each corresponding drying curve in the representation of $S_d(\psi)$ as a linear combination of drying curves is exactly $\frac{1}{k}$.

3. **Mualem** - As mentioned in the introduction, an hypothetical drying curve is a curve that satisfies the assumption that all of the hysterons are parallely connected, i.e. independent of each other. Since such a scenario is almost surreal in nature, the $S_d^\circ(\psi)$ curves can serve as a baseline for prediction of a drying curve $S_d(\psi)$. Thus, in this model there was no construction at all, and the prediction of each drying curve $S_d(\psi)$ was defined as the computation of hypothetical drying curve $S_d^\circ(\psi)$, which equals to $2S_w(\psi) - S_w^2(\psi)$.

The last three models are **Wetting To Drying With Beta** (WTDWB), **Wetting To Drying Linear Combination** (WTDLC), and **Wetting To Drying Equal** (WTDE). This three models are equivalent to HTDWB, HTDLC and HTDE (respectively), apart from the fact that they are using wetting curves $S_w(\psi)$ directly (instead of hypothetical drying curves $S_d^\circ(\psi)$), for the prediction of drying curves $S_d(\psi)$. Although linear relationship is assumed to being hold in this models between $S_w(\psi)$ curves and $S_d(\psi)$ curves, no such a relationship has been proved yet.

5.3 Expreiemntal Evaluation

The following steps are describing the approach for accomplishing the goal of this study, i.e. how each of the seven models depicted earlier performs for perdicting the drying curve of a given wetting curve.

- **Cross Validation -** In all of the models presented in section 4.2 apart from the one of **Mualem**, the data has been splitted into training and test sets using Leave-One-Out Cross Validation, so that any time that the algorithm has tested a medium, the rest of the media have served as training set. Important to note is that in the main model of this work, which is HTDBW (as well as WTDWB), the cross validation process has been done twice. This is due to that in the first performance, the cross validation has served for data splitting, whereas in the second time it served for the training phase of the optimization process, i.e. for β's value computation for a given medium. Clearly, in Mualem's model there was no need to split data into training and test sets. This is due to that given an $S_w(\psi)$ curve, the prediction of the drying curve $S_d(\psi)$ was defined as the hypothetical drying curve $S_d^\circ(\psi)$.
- **Searching for Neighbors** - For each tested medium, the algorithm computes its $S_d^\circ(\psi)$ curve's euclidian-distance from the rest of the $S_d^\circ(\psi)$ curves of all media, then sorted in increasing order of the neighbors distance. For the experiemental evaluation we have used with any value of $k \in \{1, 2, .., 12\}$ nearest neighbors, which is approximately an half of the dataset size. For each value of k, the algorithm extracted a coefficients series of size k, in order to represent the $S_d^\circ(\psi)$ curve of the tested medium, as a linear combination of its k nearest $S_d^\circ(\psi)$ curves.
- **Prediction Step** - After extracting the coefficients series for a given test medium, the algorithm picked the $S_d(\psi)$ curves that belongs to the

tested medium's neighbors, then computed a dot product between the $S_d(\psi)$ curves and the coefficients vector. This computation resulted finally in a prediction for $S_d(\psi)$ curve for tested medium.

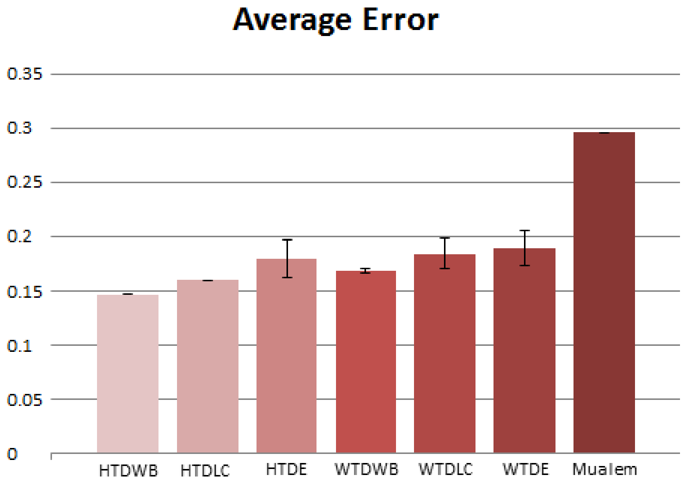

FIGURE 5.1: Bar chart of average error of the main method in this study - HTDWB, compared to the six baselines models described earlier. The standard deviation of each model performance, is attached above to its adequate bar.

- **Performance of Models** - Eventually, after repetition on prediction step for each medium, we have statistically assessed the performance of our approach, and compared it to the baseline methods, as presented in Fig.[5.1]. The comparison has been done by computing difference between vectors norm, i.e. for every number of neighbors k considered between 1 to 12, and for every porous media among the 21 exist, we have accomulated absolute value of L_2 norm of difference between the predicted curve, and the true drying curve (known beforehead from data) *for each loop* from the hysteretic loops dataset, divided by the norm of the true drying curve. This accomulated result was divided by the number of media N. More formally, each column in Fig.[5.1] was computed as:

$$error_{M_k} = \frac{\sum_{i=1}^{N} \frac{||S_{d_i}(\psi) - S_{p_i}(\psi)||}{||S_{d_i}(\psi)||}}{N} \tag{5.2}$$

Where $error_{M_k}$ represnts the error of an arbitrary one of the seven models for a specific number of neighbors $k \in \{1, 2, .., 12\}$, i is an index for each medium among the $N = 21$ exist in the dataset, $S_{d_i}(\psi)$ is the real drying curve of medium i, and $S_{p_i}(\psi)$ is the prediction of model M for $S_{d_i}(\psi)$.

Figures [5.2]-[5.5] are demonstrating examples for drying curves predictions using the main model of this work which is HTDWB, for several hystertic loops of different four media. Each figure contains the title that represents the medium's name including its nearest neighbros names, and four curves: (i) wetting curve $S_w(\psi)$, (ii) drying curve $S_d(\psi)$, (iii) hypothetical drying curve $S_d^{\circ}(\psi)$, and (iv) drying curve's prediction $S_{d_{Pred}}(\psi)$.

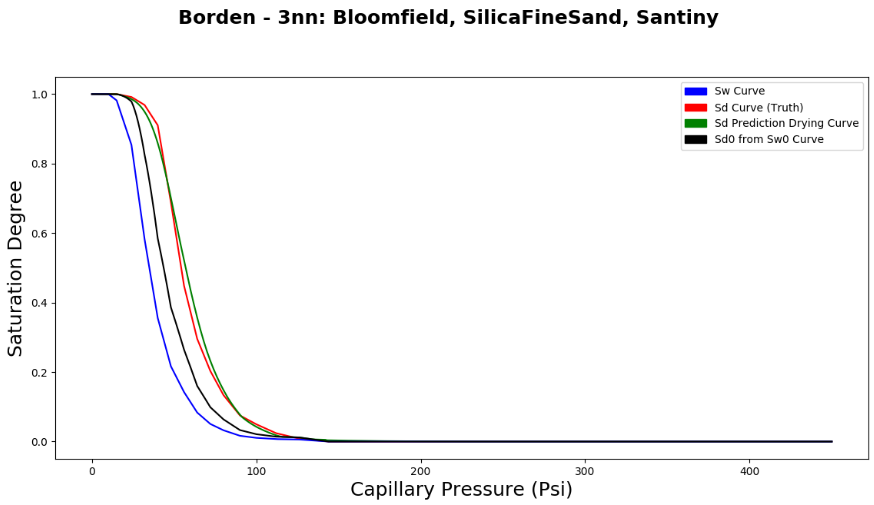

FIGURE 5.2: Prediction for medium named **Borden** with its 3 nearest neighbors, ordered in ascending order by their distance: **Bloomfield**, **Silica Fine Sand** and **Santiny**.

Whereas figures [5.2]-[5.4] are demonstrating cases on which the prediction curve (green curve) has almost conjoined with the red line (the real drying curve), Fig.[5.5] shows a bad exmaple of the prediction process. This bad prediction may be caused as a consequent of a lack in close enough neighbors. Finally, the performance of each among the seven models described,

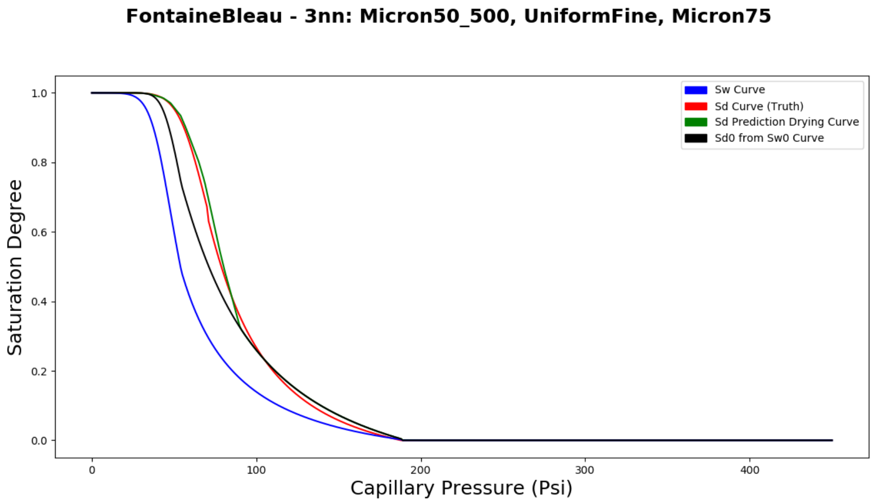

FIGURE 5.3: Prediction for medium named **Fontaine Bleau** with its 3 nearest neighbors, ordered in ascending order by their distance: **Micron 50-500**, **Uniform Fine** and **Micron 75**.

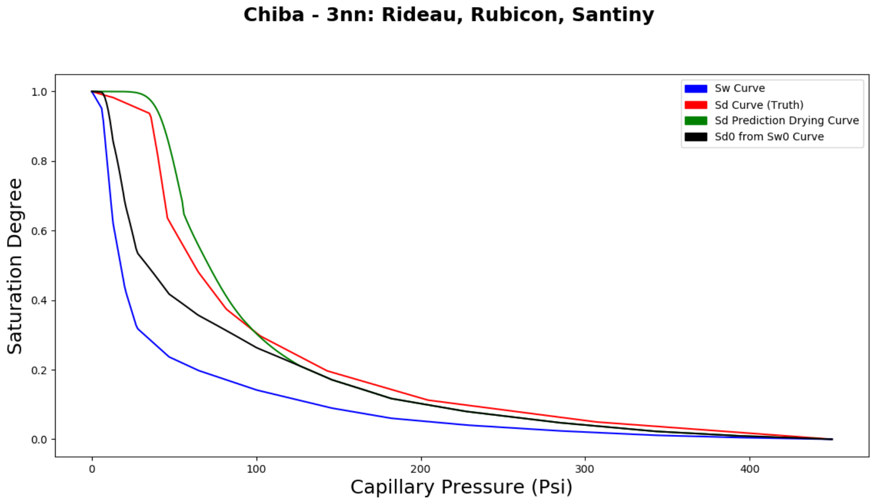

FIGURE 5.4: Prediction for medium named **Chiba** with its 3 nearest neighbors, ordered in ascending order by their distance: **Rideau**, **Rubicon** and **Santiny**.

for each k value is presented in Table [5.1]. Each row represents a different k value (between 1 to 12), and each column represents one of the six models constructed in this work: HTDWB, HTDLC, HTDE, WTDWB, WTDLC, WTDE. The only model that has no column on its own, is the seventh model which refers to **Mualem**'s model, due to the fact that no k value affects its prediction results.

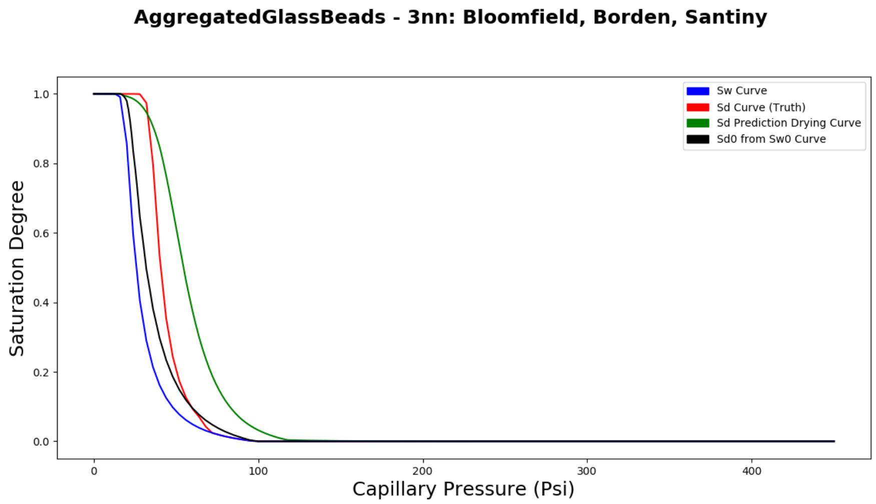

FIGURE 5.5: Prediction for medium named **Aggregated Glass Beads** with its 3 nearest neighbors, ordered in ascending order by their distance: **Bloomfield**, **Borden** and **Santiny**.

Results by neighbors list						
Neighbors / Model	**HTDWB**	**HTDLC**	**HTDE**	**WTDWB**	**WTDLC**	**WTDE**
12	0.147	0.174	0.209	0.165	0.209	0.219
11	0.147	0.170	0.201	0.167	0.203	0.212
10	0.148	0.161	0.191	0.168	0.197	0.204
9	0.147	0.162	0.190	0.168	0.191	0.198
8	0.147	0.163	0.188	0.168	0.186	0.192
7	0.148	0.162	0.184	0.169	0.179	0.184
6	0.148	0.161	0.178	0.169	0.180	0.186
5	0.147	0.156	0.168	0.170	0.183	0.184
4	0.147	0.148	0.160	0.169	0.175	0.177
3	0.145	0.158	0.170	0.171	0.181	0.185
2	0.147	0.161	0.165	0.165	0.159	0.166
1	0.146	0.146	0.146	0.171	0.171	0.171

TABLE 5.1: Results table of the seven models constructed, **HTDWB, HTDLC, HTDE, WTDWB, WTDLC** and **WTDE**. The numerical value on each cell of the internal table represents absolute value's accomulation of L_2 norm, of the difference between the predicted curve, and the true drying curve *for each loop* from the hysteretic loops dataset. The error of **Mualem**'s model was carried on 0.296 as can be seen in Fig. [5.1]. Clearly, this error is independent on the number of neighbors considered, since the prediction process is directly from the wetting to the drying curves.

Chapter 6

Discussion

The predicted BDF have been produced for each medium of the available collection of N measured porous media, and they are presented in Figures. [5.2]-[5.5]. Each model-predicted BDF has been obtained when the rest N-1 media have served as the training dataset for model calibration, such that all N predicted BDF have been obtained by leave-one-out cross-validation. As we can see in Fig.[5.5] the displayed predicted BDF generally indicate a partial success of the predictive model performance. Observed shortcomings of the model performance can be attributed to the following reasons:

i . The restriction of $\omega_{p_i} \geq 0$ is imposed on the linear combinations that approximating the given HDF, i.e. $\omega_{p_1} S^{\circ}_{d_1}(\psi) + .. + \omega_{p_k} S^{\circ}_{d_k}(\psi)$.

ii . Shortage of data; wide scattering of HDF-BDF pairs around the actual regularity.

iii . Statistical nature of Eq.(2.1) underlying the suggested theory.

In addition, there are three other possible reasons to be thoroughly considered in the subsequent study:

i . Use of the non-optimized functions $S^{\circ}_{d_i}(\psi)$ for approximative linear combinations.

ii . Use of the approximative linear combinations of the form $\omega_{p_1} S^{\circ}_{d_1}(\psi) + .. + \omega_{p_k} S^{\circ}_{d_k}(\psi)$ instead of the corresponding linear combinations of the normalized form $\frac{\omega_{p_1} S^{\circ}_{d_1}(\psi)}{||S^{\circ}_{d_1}(\psi)||} + .. + \frac{\omega_{p_k} S^{\circ}_{d_k}(\psi)}{||S^{\circ}_{d_k}(\psi)||}$.

iii . Use of the Euclidean distance between functions $S^{\circ}_{d_p}(\psi)$ and $S^{\circ}_{d_i}(\psi)$ while the relative Euclidean distance $\frac{||S^{\circ}_{d_p}(\psi) - S^{\circ}_{d_i}(\psi)||}{||S^{\circ}_{d_p}(\psi)||}$ may be more suitable.

The restriction $\omega_{p_i} \geq 0$ adopted in the optimization algorithm, was imposed in order to preserve monotonic behavior of the linear combinations $\omega_{i_1} S^{\circ}_{d_1}(\psi)$ $+.. + \omega_{i_k} S^{\circ}_{d_k}(\psi)$ $(i = 1, 2, .., N - 1)$ that should approximate the given HDF $S^{\circ}_{d_p}(\psi)$. This restriction reduces the quality of approximation, measured by the Euclidean distance between $\omega_{i_1} S^{\circ}_{d_1}(\psi) + .. + \omega_{i_k} S^{\circ}_{d_k}(\psi)$ and $S^{\circ}_{d_p}(\psi)$. Enhancing the available database should enable to remove this restriction, because generally more neighbors would fall within a close vicinity of $S^{\circ}_{d_p}(\psi)$, what should weaken non-monotonic oscillations of $\omega_{i_1} S^{\circ}_{d_1}(\psi) + .. + \omega_{i_k} S^{\circ}_{d_k}(\psi)$.

The "statistical scattering" of pairs $(S^{\circ}_{d_i}(\psi), S_{d_i}(\psi))$ with respect to the actual interfunctional regularity is thought to be a consequence of an actual variation of properties of the different porous media rather than of incorrections of the measured data used. Nevertheless, for objective judgment regarding the "statistical scattering" and the model performance one has to take into account that:

i . Determination of functions $S^{\circ}_{d_i}(\psi)$ is based on Eq.(2.1), which itself is not exact and reflects statistical properties inherent to the morphology of porous spaces.

ii . The measurements of the wetting and drying curves have been carried out using different measurement methods and devices. Both factors affect the "statistical scattering".

Chapter 7

Summary and Conclusions

After the publication of the Mualem (1984) dependent domain theory of capillary hysteresis, i.e. more than three decades, the transition from the HDF to the BDF remained a missing link in the hysteresis modelling based on this theory. This is explained probably by significant difficulty of accounting for the spatial arrangement of the hysterons, that causes the blockage phenomenon. The suggested model is aimed to fill up this missing link. Prediction of the desired BDF from its associated known BWF is obtained as a product of two mappings: (i) a nonlinear mapping of the known BWF to its corresponding HDF (Eq. 2.1), and (ii) a linear mapping of this latter function to the desired BDF, by the suggested algorithm HTDWB. We discern two approaches to predictive modeling of the interfunctional bijective relationships between the porous media characteristic functions: (i); the integral operator approach, as presented by Eq. (2.12), and (ii) the direct transformation implemented in the machine learning algorithm HTDWB, suggested in this paper. The developed framework enables formulating the inverse model intended to prediction of the BW functions from the known BDF.

We appraise the predicted results, obtained with the available data as they are, as encouraging, what justifies further efforts for improvement of the machine learning modeling methodology as well as for widening the database.

The suggested modeling approach can be applied in other prediction problems, related to the realms of Geomechanics and Physics of flow through porous media; among them are the relationship between the soil density as a function of the water saturation, S, under certain effective stress and that under another effective stress; the relationship between the water permeability as a function of S and S as a function of the capillary pressure, ψ; the relationship between the specific area of the "air-water" interface (related to the Helmholtz free energy) as a function of S and S as a function of ψ; the relationship between the grain size distribution (in granular media) and the pore size distribution. An important advantage of the suggested machine learning modeling approach is a possibility of permanent improving its predictive ability using newly incoming measured data.

Bibliography

[1] Mualem, Y., A modified dependent-domain theory of hysteresis, *Soil Science. Vol.137, No.5:283-291, 1984.*

[2] Mualem, Y., A catalogue of the hydraulic properties of unsaturated soils, *Research Project 442. Technion, Israel Institute of Technology, Haifa, Israel, 1976. 100 p.*

[3] K. Lamorski, C. SBawiNski, F. Moreno, G. Barna, W.Skierucha, J. L. Arrue, Modelling Soil Water Retention Using Support Vector Machines with Genetic Algorithm Optimisation, *Hindawi Publishing Corporation, The Scientific World Journal, Volume 2014.*

[4] K. Lamorski, J. Simůnek, C. Slawiński, J. Lamorska, An estimation of the main wetting branch of the soil water retention curve based on its main drying branch using the machine learning method, *Water Resour. Res.,53, 1539–1552, doi:10.1002/2016WR019533.*

[5] Marcel G. Schaap, Martinus Th. van Genuchten, A Modified Mualem–van Genuchten Formulation for Improved Description of the Hydraulic Conductivity Near Saturation, *Vadose Zone Journal 5:27–34 (2005).*

[6] Baohua Xu, Dong Han, Chao Xu, Linear Fixed Weight Combination Prediction Model and Model Optimum Seeking Method, *13th IEEE ITHERM Conference, 2012.*

[7] Jiabing Fu, Yacong Ma, Shoubin Dong, A lossless F ASTQ quality scores file compression algorithm based on linear combination prediction, *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).*

[8] Shibiao Wan, Man-Wai Mak, Sun-Yuan Kung, Transductive Learning for Multi-Label Protein Subchloroplast Localization Prediction, *IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 14, No. 1, January/February 2017.*

[9] Topp, G.C., Soil-water hysteresis measured in a sandy loam compared with the hysteretic domain model, *Soil Science Society of America Proceedings, 33(5):645–651.*

[10] Topp, G.C, Soil-water hysteresis: the domain model theory extended to pore interaction conditions, *Soil Science Society of America Proceedings, 35(2):219–225.*

[11] Hopmans, J. W., Dane, J. H, Temperature dependence of soil water retention curves, *Soil Science Society of America Journal, 50(3), 562-567. https://doi.org/10.2136/sssaj1986.03615995005000030004x.*

[12] Bruce R. R., A. Klute, Measurements of Soil Moisture Diffusivity from Tension Plate Outflow Data, *Soil Sci. Soc. Am. J. 1963; 27: 18-21. doi:10.2136/sssaj1963.03615995002700010011x.*

[13] Watson, K.K., R.J. Reginato, R.D. Jackson, Soil water hysteresis in a field soil, *Soil Sci. Soc. Am. Proc. 1975; 39(2): 242-246.*

[14] Talsma T, Hysteresis in Two Sands and the Independent Domain Model, *Water Resour. Res. 1970; 6(3): 964-970. https://doi.org/10.1029/WR006i003p00964.*

[15] Liakopoulos, A.C, Theoretical approach to the solution of the infiltration problem, *International Association of Scientific Hydrology. 1966; 11(1):69–110.*

[16] Bruce, R. R., A. Klute, The Measurement of Soil Moisture Diffusivity, *Soil Sci. Soc. Am. J. 1956; 20:458-462. doi:10.2136/sssaj1956.03615995002000040004x.*

[17] Santini. A, Natural replenishment of aquifers, *National Research Council of Italy (in Italian). 1981; CNR Paper No. 72: 53-89.*

[18] Behrooz Raeesi, Norman R. Morrow, Geoffrey Mason, Capillary Pressure Hysteresis Behavior of Three Sandstones Measured with a Multistep Outflow–Inflow, *Apparatus Vadose Zone Journal (2014) 13 (3): vzj2013.06.0097. RESEARCH ARTICLE | MARCH 01, 2014.*

[19] Muraleetharan K.K., Ch. Liu, Ch. Wei, T.C.G. Kibbey, L. Chen, An elastoplatic framework for coupling hydraulic and mechanical behavior of unsaturated soils, *International Journal of Plasticity. 2009; 25(3): 473–490.*

[20] Demond, A. H. and P. V. Roberts, Effect of interfacial forces on two-phase capillary pressure-saturation relationships, *Water Resources Research. v. 27, no. 3, pp. 423-437.*

[21] Gallage C., J. Kodikara, T. Uchimura, Laboratory measurement of hydraulic conductivity functions of two unsaturated sandy soils during drying and wetting processes, *Soils and Foundations (Japanese Geotechnical Society). 2013; 53(3): 417-430.*

[22] Won-Taek Hong, Young-Seok Jung, Seonghun Kang and Jong-Sub Lee, Estimation of soil-water characteristic curves in multiple-cycles using membrane and TDR system, *Materials 2016, 9(12), 1019.*

[23] Nimmo J. R, Semi-empirical model of soil water hysteresis, *Soil Sci. Soc. Am J. 1992; 56(6): 1723-1730.*

[24] Jackson R. D., Reginato R. J., Van Bavel, C.H.M, Comparison of measured and calculated hydraulic conductivities of unsaturated soils, *Water Resour. Res. 1965; 1(3): 375–380.*

[25] Ruspini L.C., R. Farokhpoor, P.E.Øren, Pore-scale modeling of capillary trapping in water-wet porous media: A new cooperative pore-body filling model, *Advances in Water Resources 2017; 108: 1-14.*

[26] Ray R. P., Morris K. B., Automated laboratory testing for soil/water characteristic curves, *Proc. 1st Int. Conf. Unsaturated Soils, E. Elsevier, 1995; 547-552.*