ASSESSING THE FEASIBILITY OF USING PEDOTRANSFER FUNCTIONS AND AUTOMATIC CALIBRATION METHODS FOR ESTIMATING DRAINMOD INPUTS

By

Ahmed Mohamed Abdelbaki Ahmed

B.Sc. Civil Engineering (2000)
Faculty of Engineering Cairo University
El Fayoum Branch, Fayoum, Egypt

Fayoum University, Fayoum, Egypt

A Thesis Submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
In
IRRIGATION AND HYDRAULICS

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
July 2011
ASSESSING THE FEASIBILITY OF USING P EDOTRANSFER
FUNCTIONS AND AUTOMATIC CALIBRATION METHODS FOR
ESTIMATING DRAINMOD INPUTS

By

Ahmed Mohamed Abdelbaki Ahmed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
In
IRRIGATION AND HYDRAULICS

Under the Supervision of

Prof. Dr. Mohamed El Sherbini
Kiwan
Professor of Irrigation and Drainage
Faculty of Engineering, Cairo
University

Prof. Dr. Mohamed A. Youssef
Biological and Agricultural
Engineering Department, NCSU, USA

Dr. Ismail Fathy Naguib
Associate Professor
Faculty of Engineering, Cairo
University

Dr. Emad El-Din Ibrahim El Giddawy
Assistant Professor, Faculty of
Engineering, Fayoum University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
July 2011
ASSESSING THE FEASIBILITY OF USING PEDOTRANSFER
FUNCTIONS AND AUTOMATIC CALIBRATION METHODS FOR
ESTIMATING DRAINMOD INPUTS

By

Ahmed Mohamed Abdelbaki Ahmed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
In
IRRIGATION AND HYDRAULICS

Approved by the
Examining Committee

Prof. Dr. Mohamed El Sherbini Kiwan, Thesis Main Advisor

Prof. Dr. Mohamed A. Youssef, Member

Prof. Dr. Mohamed Mohamed Nour El-Din Member

Prof. Dr. Kamal El-Din Milad Soliman Member

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
July 2011
Engineer: Ahmed Mohamed Abdelbaki Ahmed
Date of Birth: 2/5/1978
Nationality: Egyptian
E-mail: Ahmedabdelbaky2002@yahoo.com
Phone: 002-0100977229
Address: Egypt, Fayoum, Etsa, Kalhana
Registration Date: 13/12/2006
Awarding Date: 
Degree: Ph.D.
Department: Irrigation and Hydraulics

Supervisors: Prof. Dr. Mohamed El Sherbini Kiwan
Prof. Dr. Mohamed A. Youssef (BAE Dept. NCSU, USA)
Dr. Ismail Fathy Naguib
Dr. Emad El-Din Ibrahim El Giddawy (Fayoum Univ.)

Examinars: Prof. Dr. Mohamed El Sherbini Kiwan
Prof. Dr. Mohamed A. Youssef (BAE Dept. NCSU, USA)
Prof. Dr. Mohamed M. Nour El-Din (Ein Shams Univ.)
Prof. Dr. Kamal El-Din Milad Soliman


Key Words: Pedotransfer Functions, Automatic Calibration, DRAINMOD, Hydraulic Conductivity, Soil Water Characteristics Curve.

Summary: DRAINMOD is one of the widely used hydrologic models for drainage water management. Direct measurements of soil parameters are time consuming and costly compared with indirect methods such as Pedotransfer functions (PTFs) and automatic calibration. The goal of this study was to assess the feasibility of running DRAINMOD using input parameters estimated by PTFs or adjusted by automatic calibration process. First, all available PTFs that predict K_{sat} and SWCC have been evaluated. Then, the feasibility of running the model using inputs predicted by the best performing PTFs was tested. Finally, developing an automatic calibration framework for calibrating nine model inputs and compared the model outputs with the case of manual calibration. DRAINMOD shows a good performance in predicting subsurface drainage compared to its predictions using measured/calibrated inputs. Also, automatic calibration was found to be an effective and efficient tool for determining the model inputs compared to the manual calibration.
# CONTENTS

LIST OF TABLES .......................................................................................................................... XI

LIST OF FIGURES ....................................................................................................................... XIV

LIST OF SYMBOLES AND ABBREVIATIONS ........................................................................ XVIII

ACKNOWLEDGMENTS ................................................................................................................ XX

ABSTRACT ................................................................................................................................... XXII

## CHAPTER (1) INTRODUCTION

1.1 Introduction .............................................................................................................................. 1

1.2 Statement of the Problem ........................................................................................................ 2

1.3 Research Subject ..................................................................................................................... 3

1.4 Research Objectives ............................................................................................................... 4

1.5 Research Design and Methods ............................................................................................... 4

1.5.1 Evaluation of Pedotransfer Functions ................................................................................. 5

1.5.2 Feasibility of Running DRAINMOD using Inputs predicted by PTFs ............................... 5

1.5.3 Automatic Calibration of DRAINMOD ............................................................................... 6

1.6 Thesis Outline ......................................................................................................................... 6

## CHAPTER (2) LITERATURE REVIEW

### PART (1) PEDOTRANSFER FUNCTIONS

2.1 Historical Review .................................................................................................................... 10

2.2 Classification of Pedotransfer Functions .............................................................................. 12

2.2.1 Point Pedotransfer Functions .......................................................................................... 12

2.2.2 Continuous Pedotransfer Functions ............................................................................... 13

2.3 Development of Pedotransfer Functions ............................................................................. 15

2.3.1 Regression Analysis ......................................................................................................... 15

2.3.2 Artificial Neural Network Analysis .................................................................................. 16
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.3</td>
<td>Group Method of Data Handling (GMDH)</td>
<td>18</td>
</tr>
<tr>
<td>2.4</td>
<td>Evaluation and Calibration of Pedotransfer Functions</td>
<td>20</td>
</tr>
<tr>
<td>2.5</td>
<td>Uncertainty in Pedotransfer Functions</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>PART (2) AUTOMATIC CALIBRATION</td>
<td>27</td>
</tr>
<tr>
<td>2.7</td>
<td>Introduction</td>
<td>28</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Stages of Auto-Calibration</td>
<td>28</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Parameter Specification</td>
<td>28</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Parameter Estimation</td>
<td>30</td>
</tr>
<tr>
<td>2.8</td>
<td>Local Search Optimization Algorithms</td>
<td>32</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Rosenbrock Method</td>
<td>32</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Downhill Simplex Method</td>
<td>33</td>
</tr>
<tr>
<td>2.9</td>
<td>Global Search Optimization Algorithms</td>
<td>33</td>
</tr>
<tr>
<td>2.9.1</td>
<td>Shuffled Complex Evolution (SCE-UA)</td>
<td>33</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Genetic Algorithm (GA)</td>
<td>35</td>
</tr>
<tr>
<td>2.9.3</td>
<td>Multi Objective Complex Evolution (MOCOM-UA)</td>
<td>36</td>
</tr>
<tr>
<td>2.9.4</td>
<td>Multiobjective Shuffled Complex Evolution Metropolis algorithm (MOSCEM-UA)</td>
<td>37</td>
</tr>
<tr>
<td>2.9.5</td>
<td>Parallel SCE-UA Algorithm</td>
<td>38</td>
</tr>
<tr>
<td>2.9.6</td>
<td>Strength Pareto Evolutionary Algorithm 2 (SPEA2)</td>
<td>40</td>
</tr>
<tr>
<td>2.9.7</td>
<td>Non dominated Sorting Genetic Algorithm II (NSGA-II)</td>
<td>41</td>
</tr>
<tr>
<td>2.9.8</td>
<td>Epsilon Dominance NSGAII (ε-NSGAII)</td>
<td>42</td>
</tr>
<tr>
<td>2.9.9</td>
<td>Non dominated Sorting Particle Swarm Optimization (NSPSO)</td>
<td>43</td>
</tr>
<tr>
<td>2.10</td>
<td>Single Objective Optimization</td>
<td>45</td>
</tr>
<tr>
<td>2.11</td>
<td>Multi Objective Optimization</td>
<td>45</td>
</tr>
<tr>
<td>2.12</td>
<td>Comparison between Optimization Techniques</td>
<td>46</td>
</tr>
<tr>
<td>2.13</td>
<td>Choosing Optimization Technique</td>
<td>50</td>
</tr>
</tbody>
</table>
CHAPTER (3)
DESCRIPTION OF THE HYDROLOGICAL MODEL DRAINMOD

3.1 Introduction ........................................................................................................................................ 52
3.2 Model Development .......................................................................................................................... 53
3.3 Model Components .......................................................................................................................... 55
3.4 Model Input Parameters .................................................................................................................... 55
3.5 Model Outputs .................................................................................................................................... 56
3.6 Model Testing ..................................................................................................................................... 56
3.7 Model Sensitivity to Input Parameters ............................................................................................... 57
3.8 Model Application ............................................................................................................................. 58

CHAPTER (4)
EVALUATION OF PEDOTRANSFER FUNCTIONS OF SATURATED HYDRAULIC CONDUCTIVITY

4.1 Introduction ........................................................................................................................................ 59
4.2 Previous Evaluations of PTFs of Saturated Hydraulic Conductivity .............................................. 60
4.3 Pedotransfer Functions for Predicting Saturated Hydraulic Conductivity ....................................... 63
  4.3.1 Group 1 Pedotransfer functions .................................................................................................. 63
  4.3.2 Group 2 Pedotransfer Functions ................................................................................................ 65
  4.3.3 Group 3 Pedotransfer Functions ................................................................................................. 66
4.4 Soil Databases .................................................................................................................................... 68
4.5 Statistical Analysis ............................................................................................................................ 72
4.6 Results and Discussions .................................................................................................................... 74
  4.6.1 Performance of Group 1 PTFs .................................................................................................... 74
  4.6.2 Performance of Group 2 PTFs .................................................................................................... 76
  4.6.3 Performance of Group 3 PTFs .................................................................................................... 80
  4.6.4 Performance of All Pedotransfer Functions Evaluated using Same Dataset ............................... 84
4.7 Comparison between the Results of This Study and Previous Studies ............................................. 89
4.8 Application Procedure of Pedotransfer Functions ............................................................................ 90
4.9 Summary and Conclusions ................................................................. 92

CHAPTER (5)
EVALUATION OF PEDOTRANSFER FUNCTIONS FOR PREDICTING THE
SOIL WATER CHARACTERISTIC CURVE FOR U.S. SOILS

5.1 Introduction.......................................................................................... 95
5.2 Pedotransfer Functions of Soil Water Content ...................................... 101
  5.2.1 Point Pedotransfer Functions 101
  5.2.2 Continuous Pedotransfer Functions 106
5.3 Soil Database ...................................................................................... 117
5.4 Statistical Evaluation .......................................................................... 119
5.5 Evaluation Methods ........................................................................... 120
5.6 Results and Discussions ...................................................................... 121
  5.6.1 Overall Evaluation of PTFs 121
  5.6.2 Evaluation the Performance of PTFs within the FAO Texture Classes 122
  5.6.3 Evaluation the Performance of PTFs at certain Matric Potential Heads. 127
5.7 Comparison between PTFs Performance in the Current Study with Previous
Studies........................................................................................................ 128
5.8 Potential Application of the Study Results .......................................... 132
5.9 Conclusions.......................................................................................... 134

CHAPTER (6)
ASSESSING THE FEASIBILITY OF DRAINMOD APPLICATION USING SOIL
HYDRAULIC PROPERTIES ESTIMATED BY PEDOTRANSFER FUNCTIONS

6.1 Introduction.......................................................................................... 135
6.2 Materials and Methods........................................................................ 138
  6.2.1 Study Sites and Measured Soil Hydraulic Properties 139
  6.2.2 Pedotransfer Functions Used for Predicting Input Parameters 140
  6.2.3 Model Evaluation 142
6.2.4 Statistical Evaluation Criteria ........................................................................ 143

6.3 Results and Discussion ..................................................................................... 145
  6.3.1 PTF-Estimated Vs. Measured/Calibrated $K_{\text{sat}}$ Values ...................... 145
  6.3.2 DRAINMOD Application with $K_{\text{sat}}$ Predicted by PTFs Group (1) ...... 146
  6.3.3 DRAINMOD Application with $K_{\text{sat}}$ Predicted by PTFs Group (2) ...... 149
  6.3.4 Assessing DRAINMOD Application with SWC Data Predicted by PTFs ... 151
  6.3.5 DRAINMOD Application with $K_{\text{sat}}$ and SWC Data Predicted by PTFs 153
  6.3.6 Comparison between DRAINMOD predictions using $K_{\text{sat}}$ and SWC Data predicted by ROSETTA and PTFs 155
  6.3.7 Comparison between DRAINMOD Predictions Using Calibrated and PTF-Estimated $K_{\text{sat}}$ and SWC Data on Monthly Basis 156
  6.3.8 A Predictive analysis using DRAINMOD with PTF-estimated $K_{\text{sat}}$ and SWC Data 158

6.4 Summary and Conclusion .................................................................................. 160

CHAPTER (7)
AN AUTOMATIC CALIBRATION FRAMEWORK FOR THE DRAINAGE WATER MANAGEMENT MODEL, DRAINMOD

7.1 Introduction ........................................................................................................ 162

7.2 Materials and Methods ..................................................................................... 164
  7.2.1 Optimization Algorithm (SCE-UA) ......................................................... 164
  7.2.2 Description of DRAINMOD ................................................................. 166
  7.2.3 Calibration Parameters ............................................................................. 167
  7.2.4 Objective Functions and Evaluation Criteria ......................................... 168
  7.2.5 Description of Study Sites .................................................................. 169

7.3 Results and Discussions ..................................................................................... 170
  7.3.1 Automatic Calibration vs. Manual Calibration for DRAINMOD ............. 170
  7.3.2 Comparison between Two Automatic calibration Methods ..................... 172
  7.3.3 Sensitivity of Results to Population Size ................................................. 174

7.4 Conclusion ........................................................................................................ 176
CHAPTER (8)

CONCLUSION AND RECOMMENDATIONS

8.1 Main Results .................................................................................................................. 178
  8.1.1 Best Performing PTFs for Predicting Saturated Hydraulic Conductivity 178
  8.1.2 Best Performing PTFs for Predicting Soil Water Characteristic Curve 179
  8.1.3 Application of DRAINMOD Using Soil Properties Estimated by PTFs 180
  8.1.4 Automatic Calibration of DRAINMOD 180
8.2 Recommendations and Future Research ................................................................. 181
  8.2.1 Including of Recently Developed PTFs 181
  8.2.2 Evaluation of PTFs for non U.S. Soils 181
  8.2.3 Application of Other Hydrologic Models Using Soil Properties Estimated by
         PTFs 182

REFERENCES ..................................................................................................................... 183

APPENDIX (1) ..................................................................................................................... 196

COMPUTER PROGRAM FOR SHUFLED COMPLEX EVOLUTION
ALGORITHM FOR AUTOMATIC CALIBRATION OF DRAINMOD .......... 196
LIST OF TABLES

Table (2.1) Some Point Pedotransfer functions and their inputs and outputs (Acutis and Donatelli, 2003) .......................................................... 13
Table (2.2) Regression coefficients of Gupta and Larson (1979) equation for predicting soil water content ......................................................... 16
Table (2.3) PTFs validation: mean difference (MD, m³ m⁻³) and root of the mean squared difference (RMSD, m³ m⁻³) statistics for the whole data set. After Ungaro and Calzolari (2001) ................................................................. 23
Table (2.4) Evaluation of the PTFs for estimating field capacity after Givi et al. (2004) ... 24
Table (2.5) Evaluation of the PTFs for estimating wilting point after Givi et al. (2004) ...... 25
Table (2.6) Parameters of MIKE SHE model indentified for model parameterization by Madsen (2003) ................................................................. 29

Table (3.1) Some of the field tests for DRAINMOD ....................................................... 57

Table (4.1) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of six PTFs evaluated by Tietje and Hennings (1996) ........... 60
Table (4.2) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of eight PTFs evaluated by Wagner et al. (2001) ............... 63
Table (4.3) Group 1 Pedotransfer functions (PTFs require inputs of effective porosity) .... 64
Table (4.4) Group 2 Pedotransfer functions (PTFs that require inputs of %sand, %silt, and %clay, bulk density and porosity) ......................................................... 66
Table (4.5) Group 3 pedotransfer functions (PTFs that require inputs of particle size distribution bulk density and organic matter content) ......................................................... 67
Table (4.6) A statistical summary of dataset 1 ................................................................. 69
Table (4.8) A statistical summary of data set 3 ................................................................. 72
Table (4.9) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) for group 1 PTFs ......................................................... 74
Table (4.10) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of group 2 Pedotransfer functions ........................................ 78
Table (4.11) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of group 3 Pedotransfer functions ........................................ 82
Table (4.12) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of groups 1, 2, and 3 Pedotransfer functions using the same dataset ........................................................................................................... 86
Table (4.13) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of PTFs of saturated hydraulic conductivity in current study compared to results ........................................................................................................... 90
Table (4.14) Best PTFs to estimate $K_{sat}$ according to the available input parameters. .......... 92

Table (5.1) Results of evaluation of thirteen PTFs for predicting soil water content by Tietje and Tapkenhinrichs (1993) ......................................................................................................................... 97
Table (5.2) Results of evaluation of eleven PTFs for predicting the soil water characteristics curve by Schaap et al. (2004) ......................................................................................................................... 100
Table (5.3) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Gupta and Larson (1979) ..................... 102
Table (5.4) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Rawls et al. (1982) ................................. 103
Table (5.5) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Al Majou et al. (2007) ...................... 104
Table (5.6) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Adhikary et al. (2008) .................. 105
Table (5.7) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Homae and Firouzi (2008) ............. 105
Table (5.9) Number of soil water content measurements ($N$) and mean soil characteristics of data sets corresponding to FAO texture classes ........................................ 118
Table (5.10) Goodness of fit of point and continuous PTFs of soil water characteristics curve in different texture classes ........................................................................................................ 124
Table (5.11) Goodness of fit of five point and seventeen continuous PTFs of soil water characteristics curve in different matric potentials ...................................................... 128
Table (5.12) PTFs validation results: root mean square error ($m^3 m^{-3}$) for seven independent datasets ......................................................................................................................... 130

Table (6.1) measured soil properties for the four study sites ............................................ 140
Table (6.2) Pedotransfer functions used for predicting saturated hydraulic conductivity .... 141
Table (6. 3) Pedotransfer functions used for predicting soil water characteristics data ........... 141
Table (6. 4) Description of datasets used for DRAINMOD simulation with different input parameters .............................................................................................................................................. 143
Table (6. 5) Description of statistical measures used for comparing model predictions to observations ........................................................................................................................................ 144
Table (6. 6) Soil properties, PTF-estimated, and measured/calibrated $K_{sat}$ values for the four study sites ........................................................................................................................................... 146
Table (6. 7) Statistical comparison between observed and simulated drainage using $K_{sat}$ values predicted by PTFs group (1) ........................................................................................................................................ 148
Table (6. 8) Statistical comparison between observed and simulated drainage using $K_{sat}$ values predicted by PTFs group (2) ........................................................................................................................................ 151
Table (6. 9) Statistical comparison between observed and simulated drainage using SWC data predicted by PTFs ....................................................................................................................................... 152
Table (6. 10) Statistical comparison between observed and simulated drainage using $K_{sat}$ and SWC data predicted by PTFs ...................................................................................................................................... 155
Table (6. 11) Statistical comparison between observed and simulated drainage using $K_{sat}$ and SWC data predicted by ROSETTA and PTFs ................................................................................................ 156
Table (6. 12) Modeling efficiency of DRAINMOD simulated drainage using measured/calibrated and PTF- estimated $K_{sat}$ and SWC data on monthly basis .............................................. 157
Table (6. 13) DRAINMOD simulated overall drainage outflows with conventional and controlled drainage managements using calibrated and PTFs predicted soil parameters ....... 159
LIST OF FIGURES

Figure (2. 1) the continuous relationship between soil water content and soil water pressure head .................................................................................................................................................. 14

Figure (2. 2) A schematic view of a three-layer cascade forward network. After Merdun et al. (2006) .......................................................................................................................................................... 17

Figure (2. 3) Artificial neural network model: w represents the layer weights, b the bias, f is the transfer function, and θ is the output (water content). After Demuth et al. (2005) ........................................................................................................................................................................ 17

Figure (2. 4) GMDH network for the estimation of the Brooks and Corey saturation water content θs (m$^3$ m$^{-3}$) from clay (%), organic carbon content (%) and bulk density (Mg m$^{-3}$) by Ungaro et al., 2005. .................................................................................................................................................. 20

Figure (2. 5) Flowchart of the shuffled complex evolution algorithm (Duan et al., 1993) ........ 34 et al., 1998) .................................................................................................................................................................................. 37

Figure (2. 7) Flowchart of the MOSCEM-UA algorithm (Vrugt et al., 2003) ...................... 38

Figure (2. 8) Simplified representation of the Parallel SCE-UA algorithm (Sharma et al., 2006) ........................................................................................................................................................................ 39

Figure (2. 9) Observed and simulated Runoff obtained from three different calibration algorithms (Madsen et al., 2002) ........................................................................................................................................... 47

Figure (2. 10) Comparison between five multiple optimization algorithms used for automatic calibration of SWAT2000 (Tolson and Shoemaker 2006) ...................... 48

Figure (3. 1) Schematic of water management system with drainage to ditches or drain tubes After Skaggs (1980) .............................................................................................................................................. 54

Figure (3. 2) Schematic of water management system with subsurface drains that may be used for drainage or sub-irrigation After Skaggs (1980) .................................................................................. 55

Figure (3. 3) Water table depth during one month as an output of DRAINMOD Skaggs (1980) .................................................................................................................................................. 56
Figure (3. 4) Effect of errors in input data on some of the outputs of DRAINMOD after Skaggs (1980) ........................................................................................................................................ 58

Figure (4. 1) geometric mean of error ratio (GMER) of fourteen PTFs evaluated by Minasny and McBratney (2000) ........................................................................................................... 61

Figure (4. 2) geometric standard deviation of error ratio (GSDER) of fourteen PTFs evaluated by Minasny and McBratney (2000) ........................................................................................................... 61

Figure (4. 3) Measured vs. predicted saturated hydraulic conductivity of eight PTFs evaluated by Wagner et al. (2001) ........................................................................................................................................ 62

Figure (4. 4) Textural classification system according to the FAO Soil Map of Europe (1998), the FAO classes are: C (coarse), M (medium), MF (medium fine), F(fine), VF (very fine). ........................................................................................................................................ 70

Figure (4. 5) Distribution of sand and clay contents for the dataset used for evaluating PTFs of Group (2) ........................................................................................................................................ 70

Figure (4. 6) Distribution of sand and clay contents for the soil database used for evaluating Pedotransfer functions of Group (3) PTFs ......................................................................................... 72

Figure (4. 7) Frequency distribution of the three datasets of measured saturated hydraulic conductivities used for the evaluation of PTFs ................................................................................. 73

Figure (4. 8) Geometric mean of error ratio for PTFs of group (1) ............................................................................................................................... 75

Figure (4. 9) Geometric standard deviation of error ratio for PTFs of group (1) ............................................................................................................................... 75

Figure (4. 10) Measured vs. predicted values of K_{sat} of group 1 PTFs ........................................................................................................................................ 76

Figure (4. 11) Geometric mean of error ratio for PTFs of group (2) ............................................................................................................................... 79

Figure (4. 12) Geometric standard deviation of error ratio for PTFs of group (2) ............................................................................................................................... 79

Figure (4. 13) Measured vs. predicted values of K_{sat} of group 2 PTFs ........................................................................................................................................ 80

Figure (4. 14) Geometric mean of error ratio for PTFs of group (3) ........................................................................................................................................ 82

Figure (4. 15) Geometric standard deviation of error ratio for PTFs of group (3) ........................................................................................................................................ 83
Figure (4. 16) Measured vs. predicted values of $K_{sat}$ (group 3 PTFs) ........................................ 84

Figure (4. 17) Geometric mean of error ratio for all PTFs of groups 1, 2, and 3 ......................... 87

Figure (4. 18) Geometric standard deviation of error ratio for all PTFs of groups 1, 2, and 3 ................................................................. 87

Figure (4. 19) Measured vs. predicted values of saturated hydraulic conductivity predicted by all Pedotransfer functions using the same dataset ........................................ 88

Figure (5. 1) Predicted vs. measured soil moisture contents at matric potentials $h$ of -0.3, -3, and -10 kPa for nine PTFs were evaluated by Cornelis et al. (2001) ....................... 98

Figure (5. 2) Predicted vs. measured soil moisture contents at matric potentials $h$ of -31, -98, and -1500 kPa for nine PTFs were evaluated by Cornelis et al. (2001) ............... 99

Figure (5. 3) Distribution of sand and clay contents for the soil database used for evaluating Pedotransfer functions in this study ......................................................... 118

Figure (5. 4) Predicted vs. measured SWC of five point and six continuous PTFs represent the variation of soil water content predictions with different texture classes ................................................................. 125

Figure (5. 5) Predicted vs. measured SWC of eleven continuous PTFs represent the variation of soil water content predictions with different texture classes....................... 126

Figure (5. 6) Root mean square error values of nine Pedotransfer functions from six previous studies compared to the results obtained from the current study ............. 131

Figure (6. 1) Observed and predicted DRAINMOD outflows using $K_{sat}$ values predicted by PTFs group (1) ................................................................. 148

Figure (6. 2) modeling efficiency between Observed and predicted DRAINMOD outflows using $K_{sat}$ values predicted by PTFs group (1) ........................................ 149

Figure (6. 3) Observed and predicted DRAINMOD outflows using $K_{sat}$ values predicted by PTFs group (2) ................................................................. 150
Figure (6. 4) Modeling efficiency between Observed and predicted DRAINMOD outflows using $K_{\text{sat}}$ values predicted by PTFs group (2) ................................................................. 151

Figure (6. 5) Observed and predicted DRAINMOD outflows using SWC data predicted by PTFs ........................................................................................................ 152

Figure (6. 6) Modeling efficiency between Observed and predicted DRAINMOD outflows using SWC data predicted by PTFs ................................................................. 153

Figure (6. 7) Observed and predicted DRAINMOD outflows using $K_{\text{sat}}$ and SWC data predicted by PTFs ........................................................................................................ 154

Figure (6. 8) Modeling efficiency between Observed and predicted DRAINMOD outflows using $K_{\text{sat}}$ and SWC data predicted by PTFs ................................................................. 155

Figure (6. 9) Percent of reduction in simulated outflows between conventional and controlled drainage using calibrated and PTFs predicted input parameters ................. 159
# LIST OF SYMBOLS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>AWC</td>
<td>Available Water Capacity</td>
</tr>
<tr>
<td>BD</td>
<td>Bulk Density Mg m$^{-3}$</td>
</tr>
<tr>
<td>CL</td>
<td>Percent of clay content (%)</td>
</tr>
<tr>
<td>Dc</td>
<td>DRAINMOD simulation using calibrated input parameters</td>
</tr>
<tr>
<td>D-K1</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the first rank PTF that requires inputs of effective porosity only</td>
</tr>
<tr>
<td>D-K2</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the second rank PTF that requires inputs of effective porosity only</td>
</tr>
<tr>
<td>D-K3</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the third rank PTF that requires inputs of effective porosity only</td>
</tr>
<tr>
<td>D-KK1</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the first rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-KK2</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the second rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-KK3</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the third rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-K-θ(1)</td>
<td>DRAINMOD simulation using $K_{sat}$ and $θ(h)$ values predicted by</td>
</tr>
<tr>
<td>D-K-θ(2)</td>
<td>DRAINMOD simulation using $K_{sat}$ and $θ(h)$ values predicted by</td>
</tr>
<tr>
<td>D-K-θ(3)</td>
<td>DRAINMOD simulation using $K_{sat}$ and $θ(h)$ values predicted by</td>
</tr>
<tr>
<td>Do</td>
<td>Observed values</td>
</tr>
<tr>
<td>D-RS1</td>
<td>DRAINMOD simulation using $K_{sat}$ and $θ(h)$ values predicted by ROSETTSA using sand, silt and clay contents</td>
</tr>
<tr>
<td>D-RS2</td>
<td>DRAINMOD simulation using $K_{sat}$ and $θ(h)$ values predicted by ROSETTSA using sand, silt, and clay contents and bulk density</td>
</tr>
<tr>
<td>D-RS3</td>
<td>DRAINMOD simulation using $K_{sat}$ and $θ(h)$ values predicted by ROSETTSA using sand, silt, and clay contents, bulk density, and water content at 33 kPa</td>
</tr>
</tbody>
</table>
D-RS4 DRAINMOD simulation using $K_{\text{sat}}$ and $\theta(h)$ values predicted by ROSETTSA using sand, silt, and clay contents, bulk density, and water contents at 33, 1500 kPa

D-01 DRAINMOD simulation using $\theta(h)$ values predicted by the first rank PTF that requires inputs of particle size distribution and bulk density

D-02 DRAINMOD simulation using $\theta(h)$ values predicted by the second rank PTF that requires inputs of particle size distribution and bulk density

D-03 DRAINMOD simulation using $\theta(h)$ values predicted by the third rank PTF that requires inputs of particle size distribution and bulk density

EF Modelling Efficiency

FC Field Capacity

GA Genetic Algorithm

$K_{\text{sat}}$ Saturated hydraulic conductivity (cm hr$^{-1}$)

ME Mean Error

N a measure of the pore-size distribution

OC Organic Carbon

OM Percent of organic matter content (%)

PSD Particle Size Distribution

PTFs Pedotransfer Functions

PWP Permanent Wilting Point

R Correlation Coefficient

$R^2$ Coefficient of determination

RMSE Root Mean Square Error

SA Percent of sand content (%)

SCE-UA Suffled Complex Evolution- University of Arizona

SI Percent of silt content (%)

SWCC Soil Water Characteristic curve

$A$ related to the inverse of the air entry pressure

$\theta_{10}$ Soil water content at 10 kPa potential head

$\theta_{1500}$ Soil water content at 1500 kPa potential head

$\theta_{33}$ Soil water content at 33 kPa potential head

$\Theta_{r}$ Resedual soil water content

$\Theta_{s}$ Saturated soil water content

$\Phi$ Porosity (%)

$\Phi_{e}$ Effective Porosity
AKNOLEDGMENTS

First of all, thanks to ALLAH.

It is very difficult to say enough to express how thankful I am to Dr. Mohamed A. Youssef, my thesis advisor from North Carolina State University, USA, who has supervised me since April, 2008. He inspired and influenced me in many ways. His sound advice, encouragement and patience helped me build up my research skills and overcome many challenges and frustrations. He treated me not only as a student but also as a friend and a collaborator and was always supportive in a respectful way whenever help was needed. I believe that the things that I learned from him will benefit me over the course of my lifetime.

Great thanks to my advisors from Cairo University; Dr. Mohamed El-Shirbini Kiwan and Dr. Ismail M. F. Naguib for their help and advice during my study in Egypt. Thanks to Dr. Kiwan for his visit to me in USA in which he gave me a good contribution in my work. Also, I would like to thank my advisor from Fayoum University, Dr. Emad I. El-Giddawy for his help, which supported me since my study in the master degree and extended to my Ph. D. degree.

Special thanks to my examining committee Prof. Dr. Mohamed Mohamed Nour El-Din, irrigation and hydraulics department head, faculty of engineering, Ein Shams University. Prof. Dr. Kamal El-Din Milad Soliman, Irrigation and Hydraulics department, faculty of engineering, Cairo University.

I can never forget to thank my beloved wife, Somaya Ibrahim for her support and help not only in my study but in all my life. Without her support, this work would not be success. Special thanks to my kids, Youssef, Mohamed, and Rofaidah.

I would like to thank a lot of people for their help,

Prof. Dr. Abdallah S. Bazar (Cairo University)
Dr. R. W. Skaggs (DRAINMOD author, NCSU, USA)

Dr. Robert O. Evans (BAE, Department head, NCSU, USA)

Dr. Brian Philips (NCSU, USA)

Eng. Ismail M. Ismail
ABSTRACT

DRAINMOD is one of the widely used computer simulation models for drainage water management. The applicability of the model depends on the availability of its input parameters. Direct and indirect methods are available to feed DRAINMOD with the required inputs. Direct measurements of soil parameters are time consuming and costly compared with indirect methods such as the use of Pedotransfer functions (PTFs) and automatic calibration. PTFs is an empirical functions that can predict the model inputs such as saturated hydraulic conductivity ($K_{sat}$) and soil water characteristics curve (SWCC) from some simple and available soil physical properties like particle size distribution, bulk density, porosity, and organic matter content. Automatic calibration is the process of adjusting the model inputs within a specified range to optimize an objective function between model predictions and observations. The goal of this study was to assess the feasibility of running DRAINMOD using input parameters (e.g. $K_{sat}$ and SWCC) estimated by PTFs or adjusted by automatic calibration process. The work in this study was divided into four parts. In the first two parts, we evaluated all available PTFs in the literature until 2010 (e.g. the time of the study) that depend on available inputs in the soil databases for predicting $K_{sat}$ and SWCC respectively. In the third part, we tested the feasibility of running the model using inputs predicted by the best performing PTFs that evaluated earlier. In the fourth part, we developed an automatic calibration framework using the Shuffled Complex Evolution-University of Arizona (SCE-UA) algorithm for calibrating nine of the most sensitive model inputs and compared the results of automatic calibration with other results of manual and other automatic calibration techniques.

The results identified the best performing PTFs for predicting $K_{sat}$ and SWCC according the available soil properties in the soil databases and according to the
soil texture classes. Using the best performing PTFs to estimate the inputs of DRAINMOD, the model shows a good performance in predicting subsurface drainage compared to its predictions using measured/calibrated inputs. Also, automatic calibration was found to be an effective and efficient tool for determining the model inputs compared to the manual calibration.

The results of this study provide alternative methods for determining the input parameters of DRAINMOD or any other hydrologic model instead of field measurements, which are time consuming and costly especially with the large scale applications.
CHAPTER (1)

INTRODUCTION

1.1 Introduction

Many agricultural applications such as irrigation and drainage are very important to be studied. Agricultural drainage is the main factor to enhance crop production. Therefore, agricultural drainage has a great economic impact. Also, irrigation has a great impact on the crop production and management of water resources. In addition, irrigation and drainage are the main keys for solving the salinity problems. Hydrological models are useful tools to study these agricultural applications.

Hydrological models are simplified representations of the natural hydrologic system. In each case, the choice of the model to be applied depends mainly on the objective of the modeling but also on the available inputs. DRAINMOD is one of the widely used hydrological models in United States and all over the world to simulate agricultural drainage. DRAINMOD was evaluated and calibrated in many countries and showed good results. Model inputs are the main restriction for the model applications. These inputs are feasible to be measured for small areas and field scales but it will be time consuming and costly to be measured for large areas and watershed scales.

In this study, we evaluated and developed some alternative methods to predict the soil properties, which are required to facilitate DRAINMOD applications on the large and watershed scales. Pedotransfer functions (PTFs) and model calibration are proposed to feed the hydrologic models with inputs. PTFs are a term used in soil science literature, which can be defined as predictive functions of certain soil properties from other more available, easily, routinely, or
cheaply measured properties. For example, some soil hydraulic properties such as saturated hydraulic conductivity and soil water characteristics curve can be estimated from the soil texture, organic matter and bulk density. Also, model calibration is the process of adjusting the input parameters to the models to satisfy the best fit between the model predictions and observations. In this research study, many Pedotransfer functions will be discussed and evaluated. Using the best performing PTFs, the model predictions using soil inputs predicted by PTFs and measured/calibrated soil properties will be compared. Finally, DRAINMOD will be automatically calibrated using multi-objective calibration algorithm to provide another accurate and efficient method to feed the model with the required input data. The results of this study will provide a fast and accurate method to obtain the inputs of DRAINMOD, which will facilitate the applications of the model over watershed and large scales.

1.2 Statement of the Problem

The application of process based hydrological models requires the estimation of a number of soil hydraulic properties that are difficult, time consuming and costly to be measured. This would have severely limited the use of these models. Along with this increase in complexity is an increase in the difficulty of obtaining input parameters. A model such as DRAINMOD requires numerous spatially referenced moisture retention and hydraulic conductivity data. While these data may be available on a few experimental catchments, on the great majority, especially those in the developing world, they are not. Therefore, the only practical way to get the data necessary to run these modern processed- passed models is by using representative Pedotransfer functions or automatic calibration.

Pedotransfer functions are regression equations that relate readily available, easily measured soil physical and chemical data either directly or indirectly to soil hydraulic properties. PTFs can estimate several soil hydraulic properties such as
soil water characteristics curve and saturated hydraulic conductivity which are the most important soil hydraulic properties used as input parameters in many hydrological models. PTFs can estimate soil hydraulic properties from simple soil properties such as soil texture, organic matter and bulk density. These are either already available from soil survey or can be quickly and easily measured.

Many developments approaches such as regression analysis, artificial neural network and group method of data handling used to develop Pedotransfer functions. A large number of PTFs were developed in recent years to estimate different soil hydraulic properties. The performance of these PTFs has a great variability even under the same conditions. With the large number of developed PTFs in the literature, the performance of these PTFs has to be evaluated to indentify the best performing functions to be used for predicting model inputs. DRAINMOD predictions using measured/calibrated and PTFs predicted soil properties was compared. Another method for predicting the model inputs is the automatic calibration technique. DRAINMOD was automatically calibrated using multiobjective calibration algorithm to adjust the most important soil inputs to facilitate the application of the model on large and watershed scales.

1.3 Research Subject

Knowledge of the soil hydraulic properties is indispensable to solve many soil and water management problems related to agriculture, ecology, and environmental issues. Knowledge of these properties can be by the actual measurements, calibration, or estimation. Actual measurements are costly and time consuming even if a large number of measurements are required in the study area. To facilitate the model application, alternative techniques were discussed. These alternative techniques are Pedotransfer function and automatic calibration. PTFs are empirical equations to get the required soil hydraulic properties from the simple soil properties which are available from the soil survey. Automatic
calibration is the process of adjusting the model inputs to match the best fit between model predictions and observations. By using the PTFs and auto calibration the user of the model can determine the saturated hydraulic conductivity (K_{sat}) and soil water characteristics (SWC) data from simple soil properties such as percentage of sand, silt, clay and organic matter, and bulk density. Results of this study will facilitate the application of the hydrologic model DRAINMOD on large scale, where a lot of measured input data of soil properties are required.

1.4 Research Objectives

The goal of this study is to develop easy and applicable methods to obtain the soil input parameters required in the hydrological model DRAINMOD to facilitate the model applications on a large and watershed scale, where a large number of measurements are required. Two methods were discuss to estimate the model inputs; Pedotransfer functions and automatic calibration. This research study aims to evaluate the performance of a large number of PTFs available in the soil science literature and testing the feasibility of running DRAINMOD using soil properties inputs predicted by the best performing PTFs. Also, this study aims to develop an automatic calibration framework for DRAINMOD to provide another alternative for estimating the soil properties inputs to the model.

1.5 Research Design and Methods

The work in this study was divided into three stages to achieve the research objectives. The first stage is to evaluate the performance of a large number of Pedotransfer functions available in the soil science literature. This stage was divided into two sections. First, evaluation of Pedotransfer functions for predicting saturated hydraulic conductivity (i.e. twenty four PTFs). Second, evaluation of Pedotransfer functions for predicting soil water characteristics data (i.e. twenty two PTFs). The second stage is to test the feasibility of running the hydrologic
model DRAINMOD using input parameters predicted by the best performing PTFs previously evaluated in stage 1. The last stage in this study is to develop an automatic calibration framework for DRAINMOD to be another alternative method for estimation of model input parameters.

1.5.1 Evaluation of Pedotransfer Functions

Twenty four PTFs for predicting saturated hydraulic conductivity and twenty two PTFs for predicting soil water characteristics curve were evaluated in this study. The functions were evaluated using a datasets of US soils collected from different international databases (UNSODA database, SOILVISION database, WISE database, and NRCS database). The PTFs of $K_{sat}$ were divided into three groups according to the required inputs; the first group (6 PTFs) requires only the effective porosity as an inputs, the second group (9 PTFs) requires the particle size distribution, porosity, and bulk density as an inputs, and the third group (9 PTFs) requires particle size distribution, bulk density, and organic matter as an inputs. The PTFs of soil water characteristic curve were divided into two groups; point PTFs (5 PTFs), which predict the soil water content at a certain matric potentials; and continuous PTFs (17 PTFs), which predict the relationship between soil water content and matric potentials as a continuous function. All PTFs were evaluated at different texture classes and potential heads. After these evaluations, the best performing PTFs were identified for using in the next stage.

1.5.2 Testing the Feasibility of Running DRAINMOD using Inputs predicted by PTFs

The performance of DRAINMOD applications using soil input parameters predicted by PTFs was evaluated. Data from four US sites (two sites from Indiana, one site from Iowa, and one site from North Carolina) were used in this study. The model was used to simulate subsurface drainage using measured/calibrated and PTFs estimated soil inputs. The results demonstrate the ability of running
DRAINMOD using soil parameters predicted by PTFs. DRAINMOD is able to simulate the subsurface drainage with saturated hydraulic conductivity and/or soil water characteristics data predicted from PTFs. For all sites, statistical comparisons showed excellent agreement between the simulated and observed outflows when SWC data predicted from PTFs were used. Also, DRAINMOD showed good performance with $K_{\text{sat}}$ values predicted from PTFs. Applying DRAINMOD with $K_{\text{sat}}$ values and SWC data both predicted from PTFs, the model showed good results compared to the observation. The deviation of the simulated values form the observed values is increased when both $K_{\text{sat}}$ and SWC data predicted from PTFs.

1.5.3 Automatic Calibration of DRAINMOD

DRAINMOD was automatically calibrated using the Shuffle Complex Evolution – University of Arizona (SCE-UA) algorithm (Duan et al., 1992). Measured drainage and water table depths from three U.S. sites were used in the analysis. The model has been recently calibrated and validated in these sites; in two sites (Tide Water Experimental Station (TES) from North Carolina and STORY CITY from Iowa), the model was manually calibrated (Youssef et al., 2006; Thorp et al., 2009) and in the third site (SEPAC from Indiana), the model was automatically calibrated (Wang et al., 2006a). The model was automatically calibrated in these three sites using SCE-UA algorithm and the result were compared with the results of (Youssef et al., 2006; Thorp et al., 2009; Wang et al., 2006a).

1.6 Thesis Outline

The work in this thesis was divided into eight chapters. The first three chapters described the procedure of the work and the tools used in the study such as the computer model, DRAINMOD, and pedotransfer functions. Chapters from four to seven represent the significant work of the study. Each chapter of these
four chapters was prepared as a published paper, which was published in international meetings and conferences in USA and Canada. Finally, chapter eight concludes the entire work and presents the main points and results of the study. Brief description of the work is presented below.

In chapter 4, Twenty four Pedotransfer functions for predicting saturated hydraulic conductivity were evaluated. To overcome the problem of lacking the measurements of some soil properties in the soil databases, the functions were divided into three groups according to its input requirements. For each group of the Pedotransfer functions, a soil dataset has been selected to use in its evaluation. To test the performance of all PTFs as one group, a common dataset that has all the input requirements of all functions was used in evaluation. The results showed that the PTFs developed by Suleiman and Ritchie (2001) and Minasny and McBratney (2000) are the best models to estimate $K_{sat}$ when the available measurements is the effective porosity only. If more measurements like particle size distribution, bulk density and porosity are available; the PTF developed by Cosby et al. (1984) is the best model for predicting $K_{sat}$. By dividing the dataset into four texture classes, the Rosetta SSC (Schaap et al., 2001) showed the best performance in very fine - fine class. PTFs by Cosby et al. (1984) and Rosetta SSC are the best models for predicting $K_{sat}$ in fine medium soils. In medium class soil, PTF of Cosby et al. (1984) has the best performance. In coarse soils, PTFs by Puckett at al. (1985), Julia et al. (2004) and Cosby et al. (1984) are the best models for predicting $K_{sat}$. If more soil properties measurements are available like organic matter in addition to particle size distribution and bulk density, the PTF developed by Nemes et al. (2005) is the best model for predicting $K_{sat}$. Finally if all soil properties measurements are available in the soil database, the PTFs developed by Cosby et al. (1984), Nemes et al. (2005), Saxton et al. (1986), Saxton and Rawls (2006), and Julia et al. (2004) are the best models for predicting $K_{sat}$ respectively.
In chapter 5, twenty two PTFs for predicting soil water characteristics curve were evaluated. The evaluation and comparison of the twenty two Pedotransfer functions that were considered in this study enabled us to draw the following conclusions about the PTFs. The function developed by Rawls et al. (1982) showed the best performance among the five point Pedotransfer functions followed by the PTF developed by Adhikary et al. (2008). The Pedotransfer functions developed by Cosby et al. (1984), Brakensiek and Rawls (1985), Mayer and Jarvis (1999), and Schaap et al. (2001) (Rosetta SSC) showed the best performance among the seventeen continuous PTFs considered in this study. Also, the performance of the continuous Pedotransfer functions is almost better than the performance of the point PTFs due to its continuous results. Another conclusion, the performance of Pedotransfer functions changes with the variation of soil texture classes and with the variation of the potential heads. The Pedotransfer functions that showed a good performance when evaluated using the entire dataset are almost the PTFs that showed the good performance in different texture classes and in different matric potentials. The results indicated that the PTFs perform better with the higher potential heads and this could be due to the little error in measurements of water content at these pressures.

In chapter 6, the performance of DRAINMOD applications using soil input parameters predicted by PTFs was evaluated. Data from four US sites (two sites from Indiana, one site from Iowa, and one site from North Carolina) were used in this study. The model was used to simulate subsurface drainage using measured/calibrated and PTFs estimated soil inputs. The results demonstrate the ability of running DRAINMOD using soil parameters predicted by PTFs. DRAINMOD is able to simulate the subsurface drainage with saturated hydraulic conductivity and/or soil water characteristics data predicted from PTFs. for all sites, statistical comparisons showed excellent agreement between the simulated and observed outflows when SWC data predicted from PTFs were used. Also,
DRAINMOD showed good performance with $K_{\text{sat}}$ values predicted from PTFs. Applying DRAINMOD with $K_{\text{sat}}$ values and SWC data both predicted from PTFs, the model showed good results compared to the observation. The deviation of the simulated values form the observed values is increased when both $K_{\text{sat}}$ and SWC data predicted from PTFs.
CHAPTER (2)

LITERATURE REVIEW

PART (1) PEDOTRANSFER FUNCTIONS

2.1 Historical Review

Pedotransfer functions (PTFs) are a term used in soil science literature, which can be defined as predictive functions of certain soil properties from other more available, easily, routinely, or cheaply measured properties. This concept arises in soil science as information on soil survey is now highly demanded.

The term Pedotransfer function was coined by Bouma, 1989 as translating data we have into what we need. The most readily available data come from soil survey, such as field morphology, soil texture, structure and organic carbon content. Pedotransfer functions add value to this basic information by translating them into estimates of other more laborious and expensively determined soil properties. These functions fill the gap between the available soil data and the properties which are more useful or required for a particular model or quality assessment. Pedotransfer functions utilize various regression analysis and data mining techniques to extract rules associating basic soil properties with more difficult to measure properties.

Although not formally recognized and named until 1989, the concept of the Pedotransfer function has long been applied to estimate soil properties that are difficult to determine. Many soil science agencies have their own (unofficial) rule of thumb for estimating difficult-to-measure soil properties. Probably because of the particular difficulty, cost of measurement, and availability of large databases, the most comprehensive research in developing PTFs has been for the estimation of water retention curve and hydraulic conductivity.
As cited by Bouma (1989), the first PTF came from the study of Briggs and McLane (1907) (Eqn. 2.1). They determined the wilting coefficient, which is defined as percentage water content of a soil when the plants growing in that soil are first reduced to a wilted condition from which they cannot recover in an approximately saturated atmosphere without the addition of water to the soil, as a function of particle-size:

\[
\text{Wilting Coefficient} = 0.01 \times SA + 0.12 \times SI + 0.57 \times CL \quad (2.1)
\]

Where SA, SI, CL are the percent of sand, silt and clay contents.

With the introduction of the field capacity (FC) and permanent wilting point (PWP) concepts by Veihmeyer and Hendricksen (1927), research during the period 1950-1980 attempted to correlate particle-size distribution, bulk density and organic matter content with water content at field capacity (FC), permanent wilting point (PWP), and available water capacity (AWC).

In the 1960s various papers dealt with the estimation of FC, PWP, and AWC, notably in a series of papers by Salter and Williams (1965). They explored relationships between texture classes and available water capacity, which are now known as class PTFs. They also developed functions relating the particle-size distribution to AWC, now known as continuous PTFs. They asserted that their functions could predict AWC to a mean accuracy of 16%.

In the 1970s more comprehensive research using large databases was developed. A particularly good example is the study by Hall et al. (1977) from soil in England and Wales; they established field capacity, permanent wilting point, available water content, and air capacity as a function of textural class, and as well as deriving continuous functions estimating these soil-water properties. In the USA, Gupta and Larson (1979) developed 12 functions relating particle-size
distribution and organic matter content to water content at potentials ranging from -4 kPa to -1500 kPa.

With the flourishing development of models describing soil hydraulic properties and computer modeling of soil-water and solute transport, the need for hydraulic properties as inputs to these models became more evident. Clapp and Hornberger (1978) derived average values for the parameters of a power-function water retention curve, sorptivity and saturated hydraulic conductivity for different texture classes. In probably the first research of its kind, Bloemen (1977) derived empirical equations relating parameters of the Brooks and Corey hydraulic model to particle-size distribution.

In Germany, Lamp and Kneib (1981) introduced the term pedofunction, while Bouma and van Lanen (1986) used the term transfer function. To avoid confusion with the term transfer function used in soil physics and in many other disciplines, Bouma (1989) later called it Pedotransfer function. Since then, the development of hydraulic PTFs has become a boom research topic, first in the US and Europe, South America, Australia and all over the world. Although most PTFs have been developed to predict soil hydraulic properties, they are not restricted to hydraulic properties. PTFs for estimating soil physical, mechanical, chemical and biological properties have also been developed.

2.2 Classification of Pedotransfer Functions

2.2.1 Point Pedotransfer Functions

These PTFs are empirical functions that predict the soil hydraulic properties at a certain point as discrete values not as a continuous relationship. As example of the point Pedotransfer functions are the PTFs which predict the soil water content at a pre-defined potential (Gupta and Larson, 1979; Rawls et. al, 1982). The most frequently estimated (Θ) are at -10, -33 kPa corresponding to field capacity and at
-1500 kPa corresponding to permanent wilting point which are needed to determine plant available water content. Table (2.1) represents some examples of point PTFs and their inputs and outputs (Minasny et. al. 1999).

Table (2.1) Some Point Pedotransfer functions and their inputs and outputs (Acutis and Donatelli, 2003)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baumer</td>
<td>PSD, OC</td>
<td>BD, FC, WP</td>
</tr>
<tr>
<td>Brakensiek</td>
<td>PSD, OC, BD</td>
<td>SWC</td>
</tr>
<tr>
<td>British Soil Survey Subsoil</td>
<td>PSD, OC, BD</td>
<td>SWC</td>
</tr>
<tr>
<td>British Soil Survey Topsoil</td>
<td>PSD, OC, BD</td>
<td>SWC</td>
</tr>
<tr>
<td>EPIC</td>
<td>PSD, BD</td>
<td>FC, WP</td>
</tr>
<tr>
<td>Hutson</td>
<td>PSD, BD</td>
<td>SWC</td>
</tr>
<tr>
<td>Manrique</td>
<td>PSD, BD</td>
<td>FC, WP</td>
</tr>
<tr>
<td>Rawls</td>
<td>PSD</td>
<td>BD, FC, WP</td>
</tr>
</tbody>
</table>

PSD, particle size distribution; OC, organic carbon; BD, bulk density; FC, soil water content at field capacity; WP, soil water content at wilting point and SWC, water content at several tensions

Point PTFs for predicting soil water characteristics curve can take the form of equation (2.2). The value of soil water content ($\theta$) at a certain matric potential ($h$) depends on the value of the regression parameters ($a$, $b$, $c$, $d$ and $e$), which has different values at a pre-defined potential heads.

$$\theta_h = a \times SA + b \times SI + c \times CL + d \times OM + e \times BD$$  (2.2)

2.2.2 Continuous Pedotransfer Functions

Continuous or functional PTFs are based on the assumption that the relationship between available soil data and the required soil hydraulic properties can be described adequately by a hydraulic model that is a closed-form equation with a certain number of parameters. The advantage of continuous PTFs is that
the hydraulic characteristics are described as continuous curves, thus allowing the computation of hydraulic values at arbitrary pressures. Continuous PTFs have gained a considerable popularity with examples found in (Rawls and Brakensiek, 1985; Wosten and van Genuchten, 1988; Vereecken et al., 1989; Scheinost et al., 1997; Minasny et al., 1999; Schaap and Leij, 1998; and Minasny and McBratney, 2002). Continuous Pedotransfer functions can take the form of Equation (2.3), which developed by van Genuchten, 1980.

\[ \theta(h) = \theta_r + \frac{(\theta_s - \theta_r)}{\left[1 + (\alpha h)^{\frac{1}{n}}\right]^{\frac{1}{n}}} \]  

(2.3)

where (h) is the pressure head (in kPa); the parameters \( \Theta_r \) and \( \Theta_s \) are residual and saturated water contents (cm\(^3\) cm\(^{-3}\)), respectively; and \( \alpha \) is related to the inverse of the air entry pressure, and (n) is a measure of the pore-size distribution (van Genuchten, 1980). Figure (2.1) represents the continuous relationship between soil water content \( \Theta \) and the pressure head (h).

Figure (2.1) the continuous relationship between soil water content and soil water pressure head
2.3 Development of Pedotransfer Functions

2.3.1 Regression Analysis

Regression analysis is a technique was used for the modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables (explanatory variables). The dependent variable in the regression equation is modeled as a function of the independent variables, corresponding parameters ("constants"), and an error term. The error term is treated as a random variable. It represents unexplained variation in the dependent variable. The parameters are estimated so as to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least squares method, but other criteria have also been used.

Regression analysis is a technique used in development of Pedotransfer function. This technique divided into two approaches, multi linear regression and extended non linear regression (Minasny et al. 1999). The most common method used in point estimation PTF is to employ multiple linear regressions. Multiple linear regression is also used in parametric PTFs. Parameters of the hydraulic models are estimated by fitting the model to water-retention data with nonlinear regression and empirical relationships between basic soil properties and model parameters are then formed. Scheinost et al. 1997 found difficulty in estimating the scaling and shape parameter $\alpha$ and $n$ of the van Genuchten equation using the linear regression approach. Realizing the over parametization too many adjustable parameters relative to number of data points of the van Genuchten equation, they proposed the following approach:

1. set up the expected relationship between the parameters of the hydraulic model and soil properties,
2. insert the relationship into the model and estimate the parameters of the relationship simultaneously by fitting the extended model using nonlinear regression to all data.

As an example for PTFs that have been developed using the multi regression technique, Gupta and Larson 1979 developed point PTFs (Eqn. 2.4) for predicting soil water content at 12 matric potential ranging from 4 kPa to 1500 kPa (Table 2.2)

\[ \theta_P = a \times SA + b \times SI + c \times CL + d \times OM + e \times BD \]

Table (2. 2) Regression coefficients of Gupta and Larson (1979) equation for predicting soil water content.

<table>
<thead>
<tr>
<th>Tension (K Pa)</th>
<th>Regression coefficients</th>
<th>Correlation coefficient, R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a x 10^3</td>
<td>b x 10^3</td>
</tr>
<tr>
<td>4</td>
<td>7.053</td>
<td>10.242</td>
</tr>
<tr>
<td>7</td>
<td>5.678</td>
<td>9.228</td>
</tr>
<tr>
<td>10</td>
<td>5.018</td>
<td>8.548</td>
</tr>
<tr>
<td>20</td>
<td>3.890</td>
<td>7.066</td>
</tr>
<tr>
<td>33</td>
<td>3.075</td>
<td>5.886</td>
</tr>
<tr>
<td>60</td>
<td>2.181</td>
<td>4.557</td>
</tr>
<tr>
<td>100</td>
<td>1.563</td>
<td>3.620</td>
</tr>
<tr>
<td>200</td>
<td>0.932</td>
<td>2.643</td>
</tr>
<tr>
<td>400</td>
<td>0.483</td>
<td>1.943</td>
</tr>
<tr>
<td>700</td>
<td>0.214</td>
<td>1.538</td>
</tr>
<tr>
<td>1000</td>
<td>0.076</td>
<td>1.334</td>
</tr>
<tr>
<td>1500</td>
<td>-0.059</td>
<td>1.142</td>
</tr>
</tbody>
</table>

2.3.2 Artificial Neural Network Analysis

A recent approach for fitting PTFs is to use artificial neural networks ANN. The ANN is a network of many simple processors “units” or “neurons” each possibly having a small amount of local memory. The units are connected by communication channels “connections” which usually carry numeric data, encoded by any of various means, and often organized into subgroups or layers
Minasny et al. (1999). Figure (2.2) and Figure (2.3) represent an example of three layer artificial neural network.

Figure (2.2) A schematic view of a three-layer cascade forward network. After Merdun et al. (2006)

Figure (2.3) Artificial neural network model: w represents the layer weights, b the bias, f is the transfer function, and θ is the output (water content). After Demuth et al. (2005)
The mathematical model of an ANN comprises of a set of simple functions linked together by weights. The network consists of a set of input units, a set of output units and a set of hidden units, which link the inputs to outputs. The hidden units extract useful information from input units and use them to predict the output units. Pachepsky et. al. (1996) used ANNs to estimate water content at eight hydraulic potentials as well as van Genuchten parameters from particle-size and bulk density data for 230 soil samples. They found that for point estimation PTFs, the ANN performed better than the regression method, but for parametric PTFs the performance of both approaches was comparable. Schaap et al. (1998) estimated van Genuchten parameters for 1209 soil samples from the USA using ANNs. They found that ANN performed better than four published PTFs, and that the accuracy of prediction generally increased if more input data were used, but there was always a considerable difference between predicted and measured values. One of the advantages of neural networks compared to traditional regression PTFs is that they do not require a priori regression model, which relates input and output data and in general is difficult because these models are not known (Schaap and Leij, 1998)

2.3.3 Group Method of Data Handling (GMDH)

Group method of data handling (GMDH) is a family of inductive algorithms for computer-based mathematical modeling of multi-parametric datasets that features fully-automatic structural and parametric optimization of models.

As cited by Ungaro et. al. (2005), GMDH was first developed by A.G. Ivakhenenko in the late 1960s. GMDH is a self organizing inductive method that simultaneously finds the structure of the model and the dependence of modeled
system outputs on the values of the most significant inputs to the system (Muller and Lemke, 2000). The flow of information from the inputs to the system output takes place through a statistical network resulting from the combination of a number of functional elements organized as a sequence of linked nodes. The output from each node represents the input to the following node. In the choice of the new variables that contribute effectively to improving the estimation. Nevertheless, differently from neural networks, which require an arbitrary a priori choice of the initial structure of the network, i.e. the number of layers and the number and transfer functions of nodes, GMDH networks are a self-organizing structure, i.e. the network structure evolves during the estimation process learning from the features of the data set. The GMDH networks can be thus considered as universal structure identifiers that organize themselves inductively from data. The number of layers and nodes is estimated in agreement with a minimum external criterion by means of maximum likelihood techniques. In this respect, GMDH resemble stepwise multiple regression techniques, limiting the number of input variables to be retained in the model, but describe the relationships between input variables and system output with a level of accuracy generally superior to that of multiple regressions and often contain a number of regression coefficients generally inferior to that of traditional polynomial regression. According to Muller et al. (1998), neural networks are deductive methods that cannot identify complex objects because of their requirement of a large amount of a priori information. On the other hand, GMDH are inductive algorithms and can be considered as a regression-based method that combines the best of both statistic and neural networks. The most popular base function used in GMDH is the gradually complicated Kolmogorov-Gabor polynomial.\(^{(Eqn. \ 2.5)}\)

\[
Y(x_1, x_2, \ldots, x_n) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \quad (2.5)
\]
where $X(x_1, x_2, \ldots, x_k)$; - input variables vector; $A(a_1, a_2, \ldots, a_k)$; - vector of coefficients or weights. Components of the input vector $X$ can be independent variables, functional forms or finite difference terms. Other non-linear reference functions, such as difference, probabilistic, harmonic, logistic can also be used.

The method allows finding simultaneously the structure of model and the dependence of modeled system output on the values of most significant inputs of the system. Figure (2.4) shows an example of the networks developed by the GMDH method.

Figure (2.4) GMDH network for the estimation of the Brooks and Corey saturation water content $\Theta_s$ (m$^3$ m$^{-3}$) from clay (%), organic carbon content (%) and bulk density (Mg m$^{-3}$) by Ungaro et al., 2005.

2.4 Evaluation and Calibration of Pedotransfer Functions

Since Pedotransfer functions have been developed using certain datasets, the applications of these PTFs outside the development datasets give different results. When a decision has to be taken about the pedotransfer function which can be utilized for the prediction of soil hydraulic parameter, a statistical analysis have to be conducted between these PTFs to select the PTF of the higher performance.
These statistical analyses are called the evaluation of pedotransfer functions. A common method to evaluate pedotransfer functions is to plot the measured values against the predicted values and the correlation between them is used for model evaluation (Givi et al., 2004). Ideally, the relationship between the measured and predicted values should be linear with a slope of unity and intercept of zero. Although this method may be satisfactory for fitting an empirical model to observed data, it is inadequate for evaluating the performance of mechanistic models (Kobayashi and Salam, 2000). Generally, correlation-based statistics in conjunction with two other statistics, root-mean-squared error (RMSE), and mean deviation (MD), also called bias, are used to evaluate the performance of models. However, these statistics are not consistent with each other in their assumptions, therefore, Kobayashi and Salam (2000) have derived the following relationship among mean-squared deviation (MSD), squared bias (SB), mean-squared variation (MSV), squared difference between standard deviations (SDSD), and the lack of positive correlation weighted by the standard deviations (LCS) as follows:

\[ MSD = SB + MSV = SB + SDSD + LCS \]  \hspace{1cm} (2.6)

where

\[ MSD = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \] \hspace{1cm} (2.7)

And \( x_i \) is the simulated value, \( y_i \) is the measured value, and \( n \) is the number of observations

\[ SB = (\bar{x} - \bar{y})^2 \] \hspace{1cm} (2.8)

where \( \bar{x} \) and \( \bar{y} \), are the average values of measured and predicted data.

\[ MSV = \frac{1}{n} \sum_{i=1}^{n} [(x_i - \bar{x}) - (y_i - \bar{y})]^2 \] \hspace{1cm} (2.9)
\[
SDSD = (SD_s - SD_m)^2
\]

Standard deviation of simulated values \((SD_s) = \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{\frac{1}{2}} \quad (2.10)\]

Standard deviation of measured values

\[
(SD_m) = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2\right]^{\frac{1}{2}}
\]

\[
LCS = 2 \times SD_s \times SD_m \times (1 - r) \quad (2.12)
\]

where \(r\) is the correlation coefficient.

The accuracy of model performance is usually judged from the correlation coefficient, \(r\), however, MSD is a more comprehensive evaluator of model performance. It includes LCS, which incorporates the role of \(r\) in the computation of MSD. Moreover, the values of SB and SDSD can also provide greater insight into performance of a model.

Accuracy of the regression equations for derivation of PTFs was evaluated using \(R^2\) and RMSE between the measured or fitted and predicted values of a given hydraulic parameter Merdun et al. (2006).

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \quad (2.13)
\]

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \quad (2.14)
\]

where \(y_i\) denotes the measured value, \(\hat{y}_i\) is the predicted value, \(\bar{y}_i\) is the average of the measured value \(y\), and \(N\) is the total number of observations. In addition to these two criteria, mean error (ME) was used in the evaluation of prediction accuracy of both ANN and regression and calculated as:

\[
ME = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)}{N} \quad (2.15)
\]
Negative and positive values of ME indicate under- and over-estimation of PTFs for a given parameter, respectively. And also visual evaluation was made by regressing measured against predicted values of each parameter. In addition, the analysis of variance (ANOVA) was performed using SAS (SAS, 1999) to determine whether differences between ANN and regression in predicting soil hydraulic parameters are significant or not.

Ungaro and Calzolari (2001) used the data collected on benchmark soils from experimental sites in the Pianura Padano-Veneta, Northern Italy, stored in local soil data bases to test the reliability of nine Pedotransfer functions to derive soil water retention properties, required as input to simulation models of pollutant transport in soils. In order to assess the validity of the selected Pedotransfer functions and to provide operative guidelines for their selection, quantitative error indices mean difference, and square root of the mean squared difference, were calculated and the results is given in Table (2.3)

<table>
<thead>
<tr>
<th>PTFs</th>
<th>No of obse.</th>
<th>MD mean</th>
<th>MD stand. dev.</th>
<th>MD min.</th>
<th>MD max.</th>
<th>RMSD mean</th>
<th>RMSD stand. dev.</th>
<th>RMSD min.</th>
<th>RMSD max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saxton et al.,1986</td>
<td>119</td>
<td>0.0004</td>
<td>0.0923</td>
<td>-0.1599</td>
<td>0.3438</td>
<td>0.0698</td>
<td>0.0498</td>
<td>0.0100</td>
<td>0.4181</td>
</tr>
<tr>
<td>Rawls, Brak., 1985</td>
<td>139</td>
<td>0.0716</td>
<td>0.0651</td>
<td>-0.1641</td>
<td>0.2685</td>
<td>0.0882</td>
<td>0.0482</td>
<td>0.0276</td>
<td>0.2385</td>
</tr>
<tr>
<td>Scheinost et al., 1997</td>
<td>139</td>
<td>0.1194</td>
<td>0.0712</td>
<td>-0.0626</td>
<td>0.316</td>
<td>0.1249</td>
<td>0.0543</td>
<td>0.0051</td>
<td>0.2872</td>
</tr>
<tr>
<td>Scheinost et al., 1997</td>
<td>72</td>
<td>0.0715</td>
<td>0.0592</td>
<td>-0.0628</td>
<td>0.1992</td>
<td>0.0867</td>
<td>0.043</td>
<td>0.0063</td>
<td>0.2064</td>
</tr>
<tr>
<td>Vereecken et al., 1989</td>
<td>139</td>
<td>0.0896</td>
<td>0.0637</td>
<td>-0.0578</td>
<td>0.3221</td>
<td>0.0915</td>
<td>0.0503</td>
<td>0.0137</td>
<td>0.2936</td>
</tr>
<tr>
<td>Jarvis et al., 1997</td>
<td>139</td>
<td>-0.012</td>
<td>0.0604</td>
<td>-0.3512</td>
<td>0.1172</td>
<td>0.0589</td>
<td>0.0389</td>
<td>0.0083</td>
<td>0.3622</td>
</tr>
<tr>
<td>Rawls et al., 1982</td>
<td>139</td>
<td>0.0475</td>
<td>0.0876</td>
<td>-0.1128</td>
<td>0.4873</td>
<td>0.0920</td>
<td>0.0682</td>
<td>0.0116</td>
<td>0.458</td>
</tr>
<tr>
<td>Rawls et al., 1982</td>
<td>139</td>
<td>0.0671</td>
<td>0.0870</td>
<td>-0.0727</td>
<td>0.7116</td>
<td>0.0949</td>
<td>0.0845</td>
<td>0.0232</td>
<td>0.4238</td>
</tr>
<tr>
<td>Rawls et al., 1982</td>
<td>139</td>
<td>0.00001</td>
<td>0.0086</td>
<td>-0.0236</td>
<td>0.205</td>
<td>0.0168</td>
<td>0.0101</td>
<td>0.0028</td>
<td>0.0584</td>
</tr>
</tbody>
</table>
Givi et al. (2004) evaluate Thirteen pedotransfer functions (PTFs), namely Rosetta PTF, Brakensiek, Rawls, British Soil Survey Topsoil, British Soil Survey Subsoil, Mayr-Jarvis, Campbell, EPIC, Manrique, Baumer, Rawls–Brakensiek, Vereecken, and Hutson. These PTFs were evaluated for accuracy in predicting the soil moisture contents at field capacity (FC) and wilting point (WP), of fine-textured soils of the Zagros mountain region of Iran. The results of these evaluations are presented in Tables (2.4, 2.5).

<table>
<thead>
<tr>
<th>PTFs</th>
<th>r</th>
<th>SD_s</th>
<th>SD_m</th>
<th>SB</th>
<th>SDSD</th>
<th>LCS</th>
<th>MSD</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSST</td>
<td>0.668</td>
<td>0.031</td>
<td>0.064</td>
<td>0.003</td>
<td>0.00107</td>
<td>0.0012</td>
<td>0.005</td>
<td>0.2433</td>
<td>0.3369</td>
</tr>
<tr>
<td>Rawls</td>
<td>0.690</td>
<td>0.036</td>
<td>0.064</td>
<td>0.004</td>
<td>0.00073</td>
<td>0.0014</td>
<td>0.006</td>
<td>0.2025</td>
<td>0.3988</td>
</tr>
<tr>
<td>Barkensiek</td>
<td>0.373</td>
<td>0.057</td>
<td>0.064</td>
<td>0.002</td>
<td>0.00004</td>
<td>0.0046</td>
<td>0.007</td>
<td>0.2490</td>
<td>0.3378</td>
</tr>
<tr>
<td>BSSS</td>
<td>0.509</td>
<td>0.040</td>
<td>0.064</td>
<td>0.007</td>
<td>0.00055</td>
<td>0.0025</td>
<td>0.010</td>
<td>0.2202</td>
<td>0.3222</td>
</tr>
<tr>
<td>Baumer</td>
<td>0.466</td>
<td>0.047</td>
<td>0.064</td>
<td>0.007</td>
<td>0.00028</td>
<td>0.0032</td>
<td>0.011</td>
<td>0.2085</td>
<td>0.3452</td>
</tr>
<tr>
<td>Epic</td>
<td>0.096</td>
<td>0.028</td>
<td>0.064</td>
<td>0.008</td>
<td>0.00128</td>
<td>0.0032</td>
<td>0.012</td>
<td>0.3407</td>
<td>0.0427</td>
</tr>
<tr>
<td>Vereecken</td>
<td>-0.326</td>
<td>0.048</td>
<td>0.064</td>
<td>0.007</td>
<td>0.00025</td>
<td>0.0082</td>
<td>0.015</td>
<td>0.4757</td>
<td>-0.246</td>
</tr>
<tr>
<td>Hutson</td>
<td>0.060</td>
<td>0.007</td>
<td>0.064</td>
<td>0.011</td>
<td>0.00323</td>
<td>0.0008</td>
<td>0.016</td>
<td>0.3390</td>
<td>0.0068</td>
</tr>
<tr>
<td>Campbell</td>
<td>0.340</td>
<td>0.023</td>
<td>0.064</td>
<td>0.013</td>
<td>0.00166</td>
<td>0.0019</td>
<td>0.017</td>
<td>0.2783</td>
<td>0.1239</td>
</tr>
<tr>
<td>Ra-Brak.</td>
<td>0.438</td>
<td>0.029</td>
<td>0.064</td>
<td>0.016</td>
<td>0.00127</td>
<td>0.0021</td>
<td>0.020</td>
<td>0.2292</td>
<td>0.1977</td>
</tr>
<tr>
<td>Rosetta</td>
<td>0.736</td>
<td>0.018</td>
<td>0.063</td>
<td>0.038</td>
<td>0.00203</td>
<td>0.0006</td>
<td>0.041</td>
<td>0.1682</td>
<td>0.2154</td>
</tr>
<tr>
<td>Manrique</td>
<td>-0.012</td>
<td>0.038</td>
<td>0.064</td>
<td>0.053</td>
<td>0.00064</td>
<td>0.0050</td>
<td>0.058</td>
<td>0.2242</td>
<td>-0.007</td>
</tr>
<tr>
<td>Mayr-Jarvis</td>
<td>0.415</td>
<td>0.015</td>
<td>0.064</td>
<td>0.075</td>
<td>0.00236</td>
<td>0.0011</td>
<td>0.078</td>
<td>0.1316</td>
<td>0.1002</td>
</tr>
</tbody>
</table>

SDs: standard deviation of simulated values (m³/m³), SDm: standard deviation of measured values (m³/m³), SB: squared bias (m³/m³), SDSD: squared difference between standard deviations (m³/m³), LCS: lack of positive correlation weighted by the standard deviations (m³/m³), and MSD: mean-squared deviation (m³/m³).
Table (2. 5) Evaluation of the PTFs for estimating wilting point after Givi et al. (2004)

<table>
<thead>
<tr>
<th>PTFs</th>
<th>r</th>
<th>SDx</th>
<th>SDm</th>
<th>SB</th>
<th>SDSD</th>
<th>LCS</th>
<th>MSD</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSSS</td>
<td>0.70</td>
<td>0.03</td>
<td>0.05</td>
<td>0.00001</td>
<td>0.00032</td>
<td>0.0012</td>
<td>0.0016</td>
<td>0.1422</td>
<td>0.486</td>
</tr>
<tr>
<td>BSST</td>
<td>0.77</td>
<td>0.03</td>
<td>0.05</td>
<td>0.00057</td>
<td>0.00062</td>
<td>0.0008</td>
<td>0.0020</td>
<td>0.1288</td>
<td>0.434</td>
</tr>
<tr>
<td>Barkensie</td>
<td>0.70</td>
<td>0.03</td>
<td>0.05</td>
<td>0.00039</td>
<td>0.00029</td>
<td>0.0013</td>
<td>0.0020</td>
<td>0.1168</td>
<td>0.493</td>
</tr>
<tr>
<td>EPIC</td>
<td>0.66</td>
<td>0.02</td>
<td>0.05</td>
<td>0.00000</td>
<td>0.00115</td>
<td>0.0008</td>
<td>0.0020</td>
<td>0.1953</td>
<td>0.267</td>
</tr>
<tr>
<td>Hutson</td>
<td>0.58</td>
<td>0.01</td>
<td>0.05</td>
<td>0.00000</td>
<td>0.00174</td>
<td>0.0007</td>
<td>0.0024</td>
<td>0.2295</td>
<td>0.157</td>
</tr>
<tr>
<td>Rawls</td>
<td>0.68</td>
<td>0.04</td>
<td>0.05</td>
<td>0.00113</td>
<td>0.00017</td>
<td>0.0015</td>
<td>0.0028</td>
<td>0.094</td>
<td>0.526</td>
</tr>
<tr>
<td>Campbell</td>
<td>0.75</td>
<td>0.02</td>
<td>0.05</td>
<td>0.00137</td>
<td>0.00087</td>
<td>0.0007</td>
<td>0.0030</td>
<td>0.1346</td>
<td>0.364</td>
</tr>
<tr>
<td>Ra-Brak.</td>
<td>0.75</td>
<td>0.04</td>
<td>0.05</td>
<td>0.00194</td>
<td>0.00014</td>
<td>0.0012</td>
<td>0.0033</td>
<td>0.0651</td>
<td>0.595</td>
</tr>
<tr>
<td>Baumer</td>
<td>0.53</td>
<td>0.05</td>
<td>0.05</td>
<td>0.00162</td>
<td>0.00000</td>
<td>0.0028</td>
<td>0.0045</td>
<td>0.0887</td>
<td>0.522</td>
</tr>
<tr>
<td>Vereecken</td>
<td>0.18</td>
<td>0.04</td>
<td>0.05</td>
<td>0.00637</td>
<td>0.00024</td>
<td>0.0038</td>
<td>0.0104</td>
<td>0.3144</td>
<td>0.131</td>
</tr>
<tr>
<td>Manrique</td>
<td>0.74</td>
<td>0.03</td>
<td>0.05</td>
<td>0.00975</td>
<td>0.00054</td>
<td>0.0009</td>
<td>0.0112</td>
<td>0.0518</td>
<td>0.442</td>
</tr>
<tr>
<td>Rosetta</td>
<td>0.64</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01706</td>
<td>0.00129</td>
<td>0.0009</td>
<td>0.0193</td>
<td>0.0704</td>
<td>0.253</td>
</tr>
<tr>
<td>Mayr-</td>
<td>0.76</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04489</td>
<td>0.00232</td>
<td>0.0002</td>
<td>0.0474</td>
<td>0.0626</td>
<td>0.118</td>
</tr>
</tbody>
</table>

### 2.5 Uncertainty in Pedotransfer Functions

Pedotransfer functions have been developed particularly to translate readily available information into the variables that are needed in simulation models (Bouma 1989). An important factor that needs consideration is the uncertainty in the predictions. Uncertainty in the Pedotransfer functions can result from bias in the model, uncertainty in the parameters, and errors in the measurements of the input variables Minasny and Mcbratney (2002). In many applications, it is no longer acceptable only to model a process and then present the best prediction from the output of the model. The uncertainty in the output must also be considered, and the main contributors to the uncertainty must be identified. Uncertainty analysis is required to gauge the reliability of the predictions.

Vereecken at al. (1992) assessed the error in the hydraulic Pedotransfer functions and evaluated its effect on the output of a soil-water movement model. As cited by Minasny and Mcbratney (2002), Schaap and Leij (1998) considered
the uncertainty in the hydraulic transfer functions based on the accuracy of the database from which the functions were generated. Christiaens and Feyen (2001) analysis the uncertainties of different methods predicting soil hydraulic properties and their consequence when used in hydrological model. Most analysis considers the uncertainty due to error in the model or parameter uncertainty. Minasny et. al. (1999) analyzed the uncertainty in water retention predictions as a result of the error in the Pedotransfer functions and input data. They showed that uncertainty in the parameters is usually small compared with the uncertainty due to error in input variables and not surprisingly found that the error in propagation of error in the predictions. There are many uncertainties inherent in the input data, and consequently the Pedotransfer functions and simulation models can produce varied result.

Rubio et al. (2005) evaluated the use of the Pedotransfer functions (PTFs) included in the ROSETTA model which used to predict the soil water retention curve and compared them with ones obtained specifically for silt-loam soils. They compared The Pedotransfer methods with water retention curves obtained by a simpler sample-averaging method. Also, they evaluated the uncertainty of each method in terms of the mean error (ME) and the root mean square error (RMSE) of the predictions. Also, they calculated effort, in terms of time consumed by field sampling and laboratory determinations for the application to 100 samples. The results indicated that the efficiencies of the methods are similar to those obtained by other authors, but they found with homogeneous soils that the ratios between quality of results and effort favour the simpler approaches.
PART (2) AUTOMATIC CALIBRATION

2.6 Introduction

Calibration is the process of modifying the input parameters to a numerical model until the output from the model matches an observed set of data. The process of model calibration is normally done either manually or by using computer-based automatic procedures. In manual calibration, a trial-and-error parameter adjustment is made. In this case, the goodness-of-fit of the calibrated model is basically based on a visual judgment by comparing the simulated and the observed hydrographs. For an experienced hydrologist it is possible to obtain a very good and hydrologically sound model using manual calibration. However, since there is no generally accepted objective measure of comparison, and because of the subjective judgment involved, it is difficult to assess explicitly the confidence of the model simulations. Furthermore, manual calibration may be a very time consuming task, especially for an inexperienced hydrologist. In automatic calibration, parameters are adjusted automatically according to a specified search scheme and numerical measures of the goodness-of-fit. As compared to manual calibration, automatic calibration is fast, and the confidence of the model simulations can be explicitly stated.

Finally, model calibration does not guarantee reliability of model predictions. The parameter values obtained during calibration and the subsequent predictions made using the calibrated model are only as realistic as the validity of the model assumptions for the study. Therefore, even after calibration, there is potentially a great deal of uncertainty in results that arises simply because it is too unlikely to find error-free observational data.
2.7 Stages of Auto-Calibration

The two most important stages of calibration are parameter specification and parameter estimation. In the first stage, parameter specification, the parameters that need to be adjusted are selected, and in the parameter estimation stage, the optimal or near optimal values for the parameters are found (Sorooshian and Gupta, 1995).

2.7.1 Parameter Specification

Parameter specification or parameter identification is the first step in the automatic calibration. In this step, the calibrated parameters are identified by set of statistical criteria. This step is very significant especially for hydrologic models that have a large number of parameters, which cannot be calibrated at a time. The parameters that have high impact on the model output are the key parameters to the model and are selected to be calibrated.

Model Parameterization: In the model parameterization, the available field data, such as geological descriptions from well-logs, pumping test analysis, maps of soil profiles, soil analysis (texture, density, retention curves), and vegetation maps should be used to define the spatial patterns of the parameter values to describe the most significant variations. This is often done by defining a conceptual model with appropriate parameter classes of geological units, soil types, vegetation types etc. For each class, some parameters are then assessed directly from field data while other parameters may be subject to calibration. The challenge is to formulate a relatively simple model parameterization in order to provide a well-posed calibration problem but at the same time keep it sufficiently distributed in order to capture the spatial variability of key model parameters. Table (2.6) present an example of MIKE SHE model parameterization by Madsen (2003).
It should be noted that model parameterization and model calibration is an iterative process. If the calibration results in poorly defined parameter values, one should reconsider the model parameterization and define a simpler conceptual model that includes fewer calibration parameters. On the other hand, if the model is not able to sufficiently describe the spatial variability reflected in the observations, one should consider distributing key model parameters or including other process descriptions in the calibration.

**Parameter sensitivity analysis:** Sensitivity analysis (SA) is the study of how the variation (uncertainty) in the output of a mathematical model can be apportioned, qualitatively or quantitatively, to different sources of variation in the input of a model. In more general terms uncertainty and sensitivity analyses investigate the robustness of a study when the study includes some form of mathematical modeling. While uncertainty analysis studies the overall uncertainty in the conclusions of the study, sensitivity analysis tries to identify what source of uncertainty weights more on the study's conclusions.
Hill (1998) used sensitivity analysis to measure the change in simulated values with respect to each parameter. Spear and Hornberger, 1980 introduced a generalized sensitivity analysis procedure based on Monte Carlo sampling where a number of randomly generated parameter sets is evaluated and compared. The results of such analyses, however, should be carefully interpreted. The sensitivities in Hill, 1998 depend on the parameter values, and hence sensitivity statistics evaluated at some initial parameter values may be very different from the statistics obtained using other parameter sets. In addition, the sensitivity statistics do not properly account for parameter correlations, implying that parameters that seem to be insensitive may have important correlations with other parameters that are essential for the model behavior.

2.7.2 Parameter Estimation

Parameter estimation is most important stage in automatic calibration. In this stage, the best set of parameters is obtained by optimizing the objective functions. This stage includes the selection of the objective function that represent the difference between the measurement and the simulation, the selection of optimization algorithm, the selection of historical data, and the selection of termination criteria at which the optimization process is terminated. These components of parameter estimation process are discussed as follow:

**Selection of Objective Functions:** The objective function is a function associated with an *optimization problem* which determines how good a solution is. The objective function chosen by the modeler represents the computation of a numerical measure of the difference between the simulated output and the observed (measured) values. The goal of automatic calibration is to optimize (maximize or minimize, as appropriate) this objective function. Traditional measures used in single objective optimization include least squares [or daily root-mean-square (DRMS)] and, more recently, maximum likelihood functions.
It is important to note that, in general, trade-offs exist between the different objective functions. In these occurrences, the use of a single objective function may be questionable and it would be advisable to take into account various objective functions by considering the calibration problem in a multi-objective framework. In hydrology, most of the studies related to multi-objective calibration have investigated the use of two-objective functions and few ones have looked into the use of three or more functions (Madsen et al., 2002; Schoops et al., 2005; Parajka et al., 2007). The selected objective function implicitly gives rise to a response surface, and an iterative optimization algorithm or search procedure is used to search this surface for an “optimum” point, within the user-defined constraints (usually parameter ranges).

**Selection of Optimization Algorithms:** An optimization algorithm is a numerical method or algorithm for finding a value \( x \) such that \( f(x) \) is as small (or as large) as possible, for a given function \( f \), possibly with some constraints on \( x \). Here, \( x \) can be a scalar or vector of continuous or discrete values.

**Selection of Historical data:** As part of the automatic calibration process, a historical period of data against which to calibrate the model must also be selected. In selecting a calibration period, the goal is to choose a dataset representative of the various hydrologic phenomena experienced by the watershed. Research by Yapo et al. (1996) indicates that approximately 8 years of data, specifically including some of the wettest years, are adequate to ensure a quality calibration.

**Selection of Termination Criteria:** Termination criteria are needed to determine when to stop the iterative search. Methods that have been used include parameter convergence, function convergence, or a maximum number of iterations. When an algorithm is unable to appreciably change parameter values,
and improve the objective function value, parameter convergence is achieved. Function convergence occurs when the algorithm is unable to improve the objective function beyond a predefined increment in one or more iterations. A calibrator also may set a maximum number of iterations to stop the search procedure, ensuring that the algorithm does not enter an endless loop.

2.8 Local Search Optimization Algorithms

Local search algorithms move from solution to solution in the space of candidate solutions (the search space) until a solution deemed optimal is found or a time bound is elapsed. Local search algorithm starts from a candidate solution and then iteratively moves to a neighbor solution. This is only possible if a neighborhood relation is defined on the search space. Typically, every candidate solution has more than one neighbor solution; the choice of which one to move to is taken using only information about the solutions in the neighborhood of the current one, hence the name local search. When the choice of the neighbor solution is done by taking the one locally maximizing the criterion, the metaheuristic takes the name, hill climbing.

Termination of local search can be based on a time bound. Another common choice is to terminate when the best solution found by the algorithm has not been improved in a given number of steps. Local search algorithms are typically incomplete algorithms, as the search may stop even if the best solution found by the algorithm is not optimal. This can happen even if termination is due to the impossibility of improving the solution, as the optimal solution can lie far from the neighborhood of the solutions crossed by the algorithms.

2.8.1 Rosenbrock Method

The Rosenbrock method (Rosenbrock, 1960) is a 0th order search algorithm (it means it does not require any derivatives of the target function. Only
simple evaluations of the objective function are used). Yet, it approximates a
gradient search thus combining advantages of 0th order and 1st order strategies. This method is particularly well suited when the objective function does not require a great deal of computing power. In such a case, it's useless to use very complicated optimization algorithms. We will lose our time in the optimization calculations instead of making a little bit more evaluations of the objective function which will lead, at the end, to a shorter calculation time. If the objective function takes lots of time to evaluate (more than a few seconds), you should use a more complex algorithm.

2.8.2 Downhill Simplex Method

The Downhill Simplex Method or Nelder-Mead technique was proposed by John Nelder & R. Mead (1965) and is a technique for minimizing an objective function in a many-dimensional space. The method works with a number of rules. The starting point is used to construct a simplest, a shape with m+1 points, where m is the number of parameters. Thus for a two parameter problem there are three points, a triangle. The program calculates the objective function at each point of the simplex on the objective function surface. The simplex method is relatively robust and numerically less complicated but it can be inefficient (slow) for simple problems.

2.9 Global Search Optimization Algorithms

2.9.1 Shuffled Complex Evolution (SCE-UA)

SCE-UA is a global search algorithm for the minimization of a single function for up to 16 parameters (Duan et al., 1992). It combines the direct search method of the simplex procedure with the concept of a controlled random search of (Nelder and Mead [1965), a systematic evolution of points in the direction of global improvement, competitive evolution (Holland, 1995) and the concept of
complex shuffling. Figure (2.5) presents the flowdiagram of the SCE-UA algorithm, which can be explained as follow. In a first step (zero-loop), SCE-UA selects an initial ‘population’ by random sampling throughout the feasible parameters space for \( p \) parameters to be optimized (delineated by given parameter ranges). The population is portioned in to several “complexes” that consist of \( 2p+1 \) points. Each complex evolves independently using the simplex algorithm. The complexes are periodically shuffled to form new complexes in order to share the gained information. It searches over the whole parameter space and finds the global optimum with a success rate of 100% (Sorooshian et al. 1993).

![Flowchart of the shuffled complex evolution algorithm (Duan et al., 1993)](image)

Figure (2. 5) Flowchart of the shuffled complex evolution algorithm (Duan et al., 1993)
SCE-UA has been widely used in watershed model calibration and other areas of hydrology such as soil erosion, subsurface hydrology, remote sensing and land surface modeling (Duan, 2003). It was generally found to be robust, effective and efficient (Duan, 2003).

The SCE-UA has also been applied with success on SWAT for the hydrologic parameters (Eckardt and Arnold, 2001) and hydrologic and water quality parameters (van Griensven et al., 2002).

2.9.2 Genetic Algorithm (GA)

A genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms (EA) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination).

Genetic algorithms are implemented in a computer simulation in which a population of abstract representations (called chromosomes or the genotype of the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves toward better solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated
due to a maximum number of generations, a satisfactory solution may or may not have been reached

2.9.3 Multi Objective Complex Evolution (MOCOM-UA)

The MOCOM-UA method is a general purpose multi objective global optimization technique developed by Yapo et al., 1998. MOCOM-UA is effective and efficient for broad class of problems. This method combined the strategy of the controlled random search (Price, 1987) with the competitive evolution (Holland, 1975), Pareto ranking (Goldberg, 1989), and multi objective downhill simplex search. The MOCOM-UA strategy is illustrated in Fig. (2.6). It start with an initial sample of s points distributed randomly throughout the n dimensional feasible parameter space that represent the initial parameter uncertainty. For each point, the multi objective functions are computed. Then, the population is ranked and sorted using the pareto ranking procedure suggested by (Goldberg, 1989). Then, Simplexes of n+1 points are selected from the sample according to a robust rank based selection method (Whitley, 1989). Each simplex is evolved in an improvement direction using a multi objective extension of the downhill simplex. The evolution process produces new point that is better than the original points. The calibration terminates automatically when all points in the sample become mutually non-dominated.
2.9.4 Multiobjective Shuffled Complex Evolution Metropolis algorithm (MOSCEM-UA)

MOSCEM-UA was developed by Vrugt et al. (2003a). The algorithm combines a Markov Chain Monte Carlo sampler with the Shuffle Complex Evolutionary algorithm (SCE-UA) algorithm (Duan et al., 1992), while seeking Pareto optimal solutions using an improved fitness assignment approach. MOSCEM-UA combines the complex shuffling of the SCE-UA (Duan et al., 1992, 1993) with the probabilistic covariance-annealing process of the Shuffle Complex Evolution Metropolis-UA algorithm (Vrugt et al., 2003b). The MOSCEM-UA algorithm is presented in Fig. (2.7) and the sequence of the algorithm is: Firstly, a uniformly distributed initial population is divided into complexes within which parallel sequences are created after sorting the population based on fitness values. Secondly, the sequences are evolved iteratively toward a
multivariate normally distributed set of solutions. The moments (mean and covariance matrix) of the multivariate distribution change dynamically because they are calculated using the information from the current evolution stage of sequences and associated complexes. Finally, the complexes are reshuffled before the next sequence of evolution.

**Figure (2.7) Flowchart of the MOSCEM-UA algorithm (Vrugt et al., 2003)**

### 2.9.5 Parallel SCE-UA Algorithm

In the parallel version of the SCE-UA algorithm developed involves the following steps (Sharma et al., 2006) (Fig. 2.8):
1. Sample points are generated from the feasible space (using upper and lower bounds of the parameters). The criterion function values are computed using these sampled points.

2. The sampled points are sorted and ranked in ascending order based on the criterion function values. This will result in the smallest criterion function value generating parameters at the top of the sampled parameter list.

3. The root node partitions the complexes and distributes the complexes on to multiple processors

4. Complexes evolve on multiple processors instead of a single processor.

5. The evolved complexes from multiple processors are sent back to the root node to be combined into a single population.

6. The root node checks for the loop criteria and continues or stops accordingly.

Figure (2. 8) Simplified representation of the Parallel SCE-UA algorithm (Sharma et al., 2006)
2.9.6 **Strength Pareto Evolutionary Algorithm 2 (SPEA2)**

SPEA2 is a multiobjective evolution algorithm developed by Zitzler et al., 2001. It represents an improvement from the original Strength Pareto Evolutionary Algorithm (Zitzler and Thiele, 1999). SPEA2 overcomes limitations of SPEA by using an improved fitness assignment, bounded archiving, and a comprehensive assessment of diversity using k-means clustering. SPEA2 requires users to specify the upper bound on the number of nondominated solutions that are archived. If the number of non-dominated solutions found by the algorithm is less than the user-specified bound then they are copied to the archive and the best dominated individuals from the previous generation are used to fill up the archive. If the size of Nondominated set is larger than the archive size, a k-means clustering algorithm comprehensively assesses the distances between archive members. A truncation scheme promotes diversity by iteratively removing the individual that has the minimum distance from its neighbouring solutions. The archive update strategy in SPEA2 helps to preserve boundary (outer) solutions and guide the search using solution density information. SPEA2 has 5 primary parameters that control the algorithm’s performance: (1) population size, (2) archive size, (3) the probability of mating, (4) the probability of mutation, and (5) the maximum run time. The overall algorithm is as follows:

**Input:** 
N (population size)

\[ \bar{N} \] (archive size)

T (maximum number of generations)

**Output:** 
A (nondominated set)

**Step 1:** Initialization: Generate an initial population \( P_0 \) and create the empty archive (external set) \( \bar{P}_o = \emptyset \) Set \( t = 0 \).

**Step 2:** Fitness assignment: Calculate fitness values of individuals in \( P_t \) and \( \bar{P}_t \).
**Step 3:** Environmental selection: Copy all nondominated individuals in $P_t$ and $P_t$ to $P_{t+1}$. If size of $P_{t+1}$ exceeds $N$ then reduce $P_{t+1}$ by means of the truncation operator, otherwise if size of $P_{t+1}$ is less than $N$ then fill $P_{t+1}$ with dominated individuals in $P_t$ and $P_t$.

**Step 4:** Termination: If $t \geq T$ or another stopping criterion is satisfied then set $A$ to the set of decision vectors represented by the nondominated individuals in $P_{t+1}$. Stop.

**Step 5:** Mating selection: Perform binary tournament selection with replacement on $P_{t+1}$ in order to fill the mating pool.

**Step 6:** Variation: Apply recombination and mutation operators to the mating pool and set $P_{t+1}$ to the resulting population. Increment generation counter ($t = t + 1$) and go to Step 2.

### 2.9.7 Non dominated Sorting Genetic Algorithm II (NSGA-II)

The non-dominated sorting genetic algorithm II (NSGA-II) is a multi-objective optimization technique which is described in detail by Deb et al. (2002). The NSGA-II algorithm may be stated as follows:

1. Create a random parent population of size $N$.
2. Sort the population based on the non-domination.
3. Assign each solution a fitness (or rank) equal to its non-domination level (minimization of fitness is assumed).
4. Use the usual binary tournament selection, recombination, and mutation operators to create a new offspring population of size $N$.
5. Combine the offspring and parent population to form extended population of size $2N$. 

41
(6) Sort the extended population based on non-domination. (7) Fill new population of size \(N\) with the individuals from the sorting fronts starting from the best.

(8) Invoke the crowding comparison operator to ensure diversity if a front can only partially fill the next generation (this strategy is called ‘‘nicheing’’).

(9) Repeat the steps (2)–(8) until the stopping criterion is met. The stopping criterion may be a specified number of generations.

Confesor and Whittaker, 2007 used the NSGA-II algorithm for automatic calibration of the soil and water assessment tool (SWAT model). The calibrated SWAT model simulated well the daily stream flow of the Calapooia watershed for a 3-year period. The daily Nash-Sutcliffe coefficients were 0.86 at calibration and 0.81 at validation.

2.9.8 Epsilon Dominance NSGAII (\(\varepsilon\)-NSGAII)

The \(\varepsilon\)-NSGAII algorithm is an extension to NSGA-II algorithm (Deb et al., 2002) by adding the concepts of \(\varepsilon\)-dominance (Laumanns et al., 2002), adaptive population sizing, and a self termination scheme to reduce the need for parameter specification. The values of \(\varepsilon\), specified by the users represent the publishable precision or error tolerances for each objective. A high precision approximation of the Pareto optimal set can be captured by specifying very small precision tolerances \(\varepsilon\). The goal of employing \(\varepsilon\)-dominance is to enhance the coverage of nondominated solutions along the full extent of an application’s tradeoffs, or in other words, to maintain the diversity of solutions. \(\varepsilon\)-NSGAII is binary coded and real coded. The \(\varepsilon\)-NSGAII uses a series of “connected runs” where small populations are exploited to pre-condition the search with successively adapted population sizes. Pre-conditioning occurs by injecting current solutions within the epsilon-dominance archive into the initial generations of larger population runs. This scheme bounds the maximum size of the population to four times the number
of solutions that exist at the user specified ε resolution. There are 4 major parameters that need to be specified for ε-NSGAII (1) the probability of mating, (2) the probability of mutation, (3) the maximum run time, and (4) the initial population size. The maximum run time is defined as the upper limit on the time the user is willing to invest in search. Although epsilons must be specified for every objective, these values are defined by the properties of the application not the evolutionary algorithm. In any optimization application, it is recommended that the user specify the publishable precision or error tolerances for their objectives to avoid wasting computational resources on unjustifiably precise results.

2.9.9 Non dominated Sorting Particle Swarm Optimization (NSPSO)

This method was developed by Liu, 2009 and applied for auto calibrating the NAM rainfall runoff model. NSPSO is a multi-objective hybrid algorithm to combine single-objective particle swarm optimization (PSO) with NSGA-II operations without losing performance on establishing the Pareto front. The method begins with initial search points and velocities for each agent are usually generated randomly within the allowable range. The population P is sorted according to nondominance. Thus, minimization of fitness is assumed. The current searching point is set to p best for each agent. A random point from the non-dominated solutions is set to gbest for an agent. And then the velocity and position can be modified for the agent using single objective PSO formula. From the first generation onward, the procedure is different. A combined population R \(\frac{1}{4} P [ Q\) is formed. The population will be size 2N. Then, the population is sorted according to non-domination. The new parent population is formed by adding solutions from the first front till the size exceeds N. Let us say that the set Fl is the last non-dominated set beyond which no other set can be accommodated. In general, the count of solutions in all sets from F1 to Fl would be larger than population size. To choose exactly N population members, we sort the solutions of the last front
using the crowded comparison operator, in the descending order and choose the best solutions needed to fill all population slots [12]. As the NSPSO progresses, the entire population tends to converge to the global Pareto front, provided these evolutionary strategies are given sufficient prediction accuracy.

The NSPSO combines the strengths of these advanced operations (a fast non-dominated sorting approach, crowding distance ranking, elitist strategy, mutation, and selection operations) with single-objective PSO search. The hybrid algorithm is presented below:

Step 1: Generate an initial population $P$ (Population size = $N$) and velocity for each individual (agent or particle) in a feasible space; Set the maximum speed $v_{i_{\text{max}}}$ ($v_{i_{\text{max}}} = \text{its upper bound minus lower bound}$) for a variable.

Step 2: Sort the population based on the non-domination and crowding distance ranking.


Step 4: Assign each individual a fitness (or rank) equal to its nondomination level (minimization of fitness is assumed).

Step 5: Randomly choose one individual as gbest for $N$ times from the non-dominated solutions, and modify each searching point using previous PSO formula and the gbest:

Step 6: Do mutation operator (Coello et al., 2004).

Step 7: Combine the offspring and parent population to form extended population of size $2N$.

Step 8: Sort the extended population based on non-domination and fill the new population of size $N$ with individuals from the sorting fronts starting to the best.
Step 9: Modify the pbesti of each searching point: If current rank of the new individual (offspring) $P_{k+1}^i$ is smaller than or equal to the previous one (parent) in $R$, replace the pbesti with current individual; otherwise keep the previous pbesti:

Step 10: Perform step (2)–(9) until the stopping criterion is met. The main differences of our approach with respect to the other proposals existing in the literature (Coello et al., 2004; Li, 2003; Raquel & Naval, 2005; Fieldsend & Singh, 2002) are:

- We add selection operator to the multi-objective particle swarm algorithm.
- We do not need external repository to save the pbest and gbest.
- Selection regime for the choosing of global best (gbest) and personal best (pbest) for swarm members is based on elitist operation.
- The programming code is much short compared with the other proposals and simple to implement to any optimization problem.

### 2.10 Single Objective Optimization

During the past three decades, considerable research has been performed on the development of automated methods to aid the model calibration process. The classical single-objective automatic approach, based on optimization theory, requires the definition of a mathematical measure (an objective function such as least squares or maximum likelihood) of the differences between the observed and simulated hydrograph. An optimization algorithm is then used to adjust the parameters toward values that minimize (or maximize, if appropriate) this function. Although the method is fast and objective, it has not received widespread acceptance among operational hydrologists. For example, NWS personnel have explored the use of automatic calibration methods (with various single-objective functions) but found that visual inspection of the hydrograph reveals areas of concern, such as poor matching of recessions and unacceptable flow biases. Use of
different objective functions has not helped to resolve this problem. Therefore, poor confidence in the capabilities of automatic methods has inhibited their usage at RFCs for speeding model calibration.

2.11 Multi Objective Optimization

The classical single-objective automatic calibration approach, although fast and objective, has not received widespread acceptance among operational hydrologists.

2.12 Comparison between Optimization Techniques

Moulin et al., 1993 compared five currently used algorithms: Simplex algorithm, Powell algorithm, Marquardt algorithm, Quasi-Newton algorithm, and the one specific for compartmental modeling provided by the software SAAM. They found that the Simplex algorithm can be more efficient in case of a good a priori knowledge of parameters, and less efficient in case of a poor a priori knowledge.

Gupta et al., 1999 compared the capability of the SCE-UA automatic procedure with the interactive MCL multistage semiautomated (manual) method developed by Brazil (1988) for calibration of the SAC-SMA model. The results indicate that the state of the art of automatic calibration methods has been elevated to the point where such methods can be considered seriously as a viable alternative to the manual approach.

Madsen et al. (2002) compared three multi objective algorithms to calibrate the MIKE 11/NAM rainfall-runoff model (Fig 2.9). The optimization algorithms are shuffled complex evolution algorithm, clustering and simulated annealing algorithm, and knowledge based expert system. There results illustrate the
problem of non-uniqueness of model parameters and non of the method are non superior with respect to all performance measures considered.

Figure (2. 9) Observed and simulated Runoff obtained from three different calibration algorithms (Madsen et al., 2002)

Kim et al. 2006 compared automatic calibration method using the parameter estimation software (PEST) with manual calibration for the Hydrological Simulation Program-Fortran (HSPF) model. The results showed that the automatic calibrated using PEST performed better than the manual calibration with respect to $R^2$, $E$, and $RMSE$ for both the calibration and validation periods.

Tolson and Shoemaker 2006 compared four multiple optimization algorithms for the automatic calibration of the SWAT2000 model (Fig. 2.10). The optimization algorithms considered were the Shuffled Complex Evolution (SCE),
real-valued simple Genetic Algorithm (GA), multi-start Simplex and Monte Carlo Sampling (MCS) algorithms. In addition, they developed a new algorithm called the Global Greedy Search (GGS) algorithm. GGS algorithm was compared to the popular SCE algorithm. The results showed that the SCE algorithm is better than Simplex, GA and MCS algorithms. However, the new GGS algorithm was demonstrated to find comparable or better final solutions than SCE and GGS found good calibration solutions substantially faster than SCE.

![Comparison between five multiple optimization algorithms used for automatic calibration of SWAT2000 (Tolson and Shoemaker 2006)](image)

Figure (2.10) Comparison between five multiple optimization algorithms used for automatic calibration of SWAT2000 (Tolson and Shoemaker 2006)

Tang et al. 2006 compared three multiobjective evolution algorithms (Epsilon Dominance Nondominated Sorted Genetic Algorithm- II (ε-NSGAIi), the multiobjective Shuffled Complex Evolution Metropolis algorithm (MOSCEM-UA), and the Strength Pareto Evolutionary Algorithm 2 (SPEA2). They used three test cases to compare the algorithms. For the first case, the ε-NSGAII attained the best overall performance for the test function suite followed by SPEA2 while MOSCEM-UA was not able to solve the test function suite reliably. In the second case, SPEA2 attained statistically superior performance in all metrics at the 99%
confidence level. MOSCEM-UA and ε-NSGAII attained results that were competitive with one another. In the third case, SPEA2 and ε-NSGAII had statistically equivalent performance metrics, although SPEA2 was slightly more reliable. They concluded that, Overall, SPEA2 is an excellent benchmark algorithm for multiobjective hydrologic model calibration.

Goswami and Oconnorsix, 2007 used six automated strategies of parameter optimization for calibration of the conceptual soil moisture accounting and routing (SMAR) model for rainfall-runoff simulation in two catchments, one small and the other large. The methods used were: the genetic algorithm, particle swarm optimization, Rosenbrock's technique, shuffled complex evolution of the University of Arizona, simplex search, and simulated annealing. A comparative assessment is made using the Nash-Sutcliffe model efficiency index and the mean relative error (MRE) to evaluate the performance of each optimization method. It was found that the probabilistic global population-based optimization method of simulated annealing is considered best in terms of having the least variability of parameter values in successive tests, thereby alleviating the phenomenon of equifinality in parameter optimization, and also in producing the lowest MRE in verification.

Ndiritu, 2009 compared the automatic calibration of the Pitman model using the shuffled complex evolution algorithm (SCE-UA) with the manual calibration for Nine sub catchments from three southern African basins. Automatic calibration was found to obtain considerably better performance for 2 of the 9 sub catchments modeled suggesting, not unexpectedly, that manual calibration is more prone to achieving suboptimal parameter sets than automatic calibration. For one basin, automatic calibration obtained slightly better performance than manual calibration but gave notably poorer validation performance suggesting inadequacy of model structure.
Liu, 2009 compared the performance of the multiobjective algorithms NSPSO and NSGA-II for calibrating the MIKE SHE model. The simulation results suggest that the proposed optimization framework of NSPSO algorithm is able to achieve good solutions as well diversity compared to the NSGA-II optimization framework.

2.13 Choosing Optimization Technique

A number of studies have been conducted to compare the SCE-UA and other global and local search procedures for model calibration (Duan et al., 1992; Gan and Biftu, 1996; Kuczera, 1997; Franchini et al., 1998). Duan et al. (1993) compared the SCE-UA method with the controlled random search (CRS2) method and a multi-start algorithm (MSX) on seven well established test functions from the literature and demonstrated the superiority of the SCE-UA method. Sorooshian et al. (1993) used a ‘true’ parameter set for the Leaf River watershed to generate ‘synthetic’ stream flows to calibrate the Sacramento model with several optimization algorithms, including the SCE-UA and the multi-start simplex (MSX) methods. They found that the SCE-UA located the ‘true’ parameter values at a 100% success rate while the MSX failed to locate the global optimum in all trials. Cooper et al. (1997) investigated the performance of three probabilistic optimization techniques for calibrating the Tank model. These methods were the SCE-UA, genetic algorithms (GA) and simulated annealing (SA) methods. They found that out of the three global optimization methods, SCE-UA provided better estimates of the optimal solution than GA and SA methods. SCE-UA was also the best in terms of efficiency as expressed by the number of iterations for convergence. They concluded that the superior performance of SCE-UA was perhaps due to its strategy of concurrently exploring several different promising regions of attraction. Kuczera (1997) compared four algorithms, namely the SCE-UA method, the GA (with traditional crossover) and multiple random starts (using
either simplex or quasi-Newton local searches) for parameter identification of the modified Surface inFiltration Baseflow (SFB) model (Boughton, 1984). In his case study, the SCE-UA algorithm was found to be most robust and efficient. Thyer et al. (1999) compared the performance of two probabilistic global optimization methods: SCEUA and the three-phase simulated annealing algorithm (SA-SX). Both algorithms were used to calibrate two parameter sets (a reduced, well-identified parameter set and the full parameter set) of a modified version of the SFB model using data from two Australian catchments that have low and high runoff yields. For the reduced, well-identified parameter set, the algorithms have a similar efficiency for the low-yielding catchment, but SCEUA is almost twice as robust. Although the robustness of the algorithms is similar for the high-yielding catchment, SCE-UA is six times more efficient than SA-SX. Ndiritu and Daniell (2001) observed that the simple GA is inappropriate for global optimum location and hence they modified the simple GA, by using three strategies for its improvement. They further compared their improved GA with the SCE-UA and noted that SCE-UA outperformed the improved GA on two of the three optimization problems that they investigated. Thus, they concluded that SCE-UA would be the first preference for the optimization of unfamiliar continuous variable problems including rainfall-runoff model calibration. These studies by various researchers have demonstrated that the SCE-UA method is a robust and efficient search algorithm.
CHAPTER (3)

DESCRIPTION OF THE HYDROLOGICAL MODEL DRAINMOD

3.1 Introduction

DRAINMOD is a computer simulation model developed by Dr. Wayne Skaggs at the Department of Biological & Agricultural Engineering, North Carolina State University, Raleigh, NC in 1980. The model simulates the hydrology of poorly drained, high water table soils on an hour-by-hour, day-by-day basis for long periods of climatological record (e.g. 50 years). The model predicts the effects of drainage and associated water management practices on water table depths, the soil water regime and crop yields. It has been used to analyze the hydrology of certain types of wetlands and to determine whether the wetland hydrologic criterion is satisfied for drained or partially drained sites. The model is also used to determine the hydraulic capacity of systems for land treatment of wastewater. The model has been successfully tested and applied in wide variety of geographical and soils conditions. In the last 20 years, the model's capability has been extended to predict the effects of drainage and water management practices on the hydrology and water quality of agricultural and forested lands both on field and watershed scale.

The latest version, DRAINMOD 6.0, combines the original DRAINMOD hydrology model with DRAINMOD-N (nitrogen sub-model developed by Youssef et al., 2005) and DRAINMOD-S (salinity sub-model) into a Windows based program. The new version includes a graphical user interface that allows easy preparation of input data sets, running simulations as well as displaying model outputs. In addition to organizing the hydrology, nitrogen, and salinity components of DRAINMOD, the interface facilitates analyses of the effect of
drainage system design on subsurface drainage, surface runoff, SEW30, crop yield, and nitrogen loss in surface and subsurface drainage by automatically editing drainage design parameters (e.g. drain spacing & drain depth) over a specified range, simulating the different designs and graphically displaying the results. The interface also calculates the runoff volume from surrounding areas that drain to a site and adds that runoff volume to a DRAINMOD water balance of the site. Version 6.0 also includes routines for soil temperature modeling and considers freezing and thawing effects on drainage processes.

3.2 Model Development

DRAINMOD is based on one-dimensional water balance in the soil profile and uses long duration weather data to simulate the performance of drainage systems. The model was developed specifically for shallow water table, poorly drained soils. The model avoids the complex numerical methods by using approximate methods to quantify the hydrologic components (Figure 3.1), including subsurface drainage, sub irrigation, infiltration, evapotranspiration (ET), and surface runoff. For example, subsurface drainage is computed using the Houghoudt equation, which assumes an elliptical water table shape. The change in water table depth is based on the assumption that the soil water profile above the water table is drained to equilibrium with the water table. Hourly rainfall is used to compute infiltration using a modified Green-Ampt method. Excess rainfall fills surface storage and any remaining rainfall is considered runoff. Evapotranspiration is computed from potential evapotranspiration as limited by soil water availability. Actual evapotranspiration is the amount that can be supplied from the water table plus the amount available from the unsaturated zone.
The major process of the model is a water balance for the soil profile. The water balance for a time increment of $\Delta t$ is expressed in Eqn. (3.1).

$$\Delta V_a = D + ET + DS - F$$  \hspace{1cm} (3.1)  

where $\Delta V_a$ is the change in the air volume (cm), $D$ is lateral drainage (cm) from the section, $ET$ is evapotranspiration (cm), $DS$ is deep seepage (cm), and $F$ is infiltration (cm) entering the section during time interval $\Delta t$. All of the right-hand side terms are computed in terms of the water table elevation, soil water content, soil properties, site and drainage system parameters, and atmospheric conditions. The amount of runoff and storage on the surface is computed from a water balance at the soil surface for each time increment (Eqn. 3.2).

$$P = F + \Delta S + RO$$  \hspace{1cm} (3.2)  

where $P$ is the precipitation (cm), $F$ is infiltration (cm), $\Delta S$ is the change in volume of water stored on the surface (cm), and $RO$ is runoff (cm) during time interval $\Delta t$. 

Figure (3.1) Schematic of water management system with drainage to ditches or drain tubes After Skaggs (1980)
3.3 Model Components

Model components are the precipitation, infiltration, surface drainage, subsurface drainage, sub irrigation, evapotranspiration, soil water distribution, and rooting depth (Figure 3.2).

![Diagram of water management system with subsurface drains](image)

Figure (3. 2) Schematic of water management system with subsurface drains that may be used for drainage or sub-irrigation After Skaggs (1980)

3.4 Model Input Parameters

The input data requirements for the DRAINMOD are soil properties, crop inputs, water management system parameters and climatological input data. In many cases, all of the input data needed in the model will not be available from conventional data sources. Furthermore, it may not be possible to measure, or otherwise directly determine, the data, and the needed inputs will have to be approximated using Pedotransfer function or adjusting input parameters through model calibration. The most important parameters in the model is soil Properties Inputs, saturated hydraulic conductivity, soil water characteristics Θ(h), Drainage volume – water table depth relationship, Upward flux, Green-Ampt equation parameters, Crop input data, Drainage system parameter, and Effective drain radius. A number of input data sets are distributed with the model. These depict the model setup for various water table management systems. Sample input data
sets are included for designing subsurface drainage, controlled drainage and subirrigation systems. The author also has a number of weather data sets for areas in the US.

### 3.5 Model Outputs

There are a number of output options. Daily, monthly and annual summaries include infiltration, evapotranspiration, depth to the water table (Figure 3.3), runoff, drainage volume, number of work days based on soil air volume, drought and wet stresses. The model also produces an annual ranking for computing recurrence intervals for many of the outputs.

![Image of water table depth during one month as an output of DRAINMOD Skaggs (1980)](image)

**Figure (3. 3)** Water table depth during one month as an output of DRAINMOD Skaggs (1980)

### 3.6 Model Testing

The model was field tested in North Carolina (Skaggs, 1982). Through both cooperative and independent efforts, the model has been subsequently tested in many parts of the world. Some of these efforts are listed in Table (3.1). In addition to these tests of DRAINMOD, the nitrogen and salinity portions of the model have been tested by Breve et al. (1997), Kandil et al. (1995) and Merz and Skaggs (1998). Other field testing is currently on going for these versions of the model.
Many of the field tests have conducted sensitivity analyses. In addition to these, the developer conducted extensive sensitivity analyses (Skaggs, 1980a and b; Skaggs, 1982). Other sensitivity analyses have targeted the soil hydraulic properties. These include Anderson et al. (1987) and Workman and Skaggs (1994).

<table>
<thead>
<tr>
<th>Area</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohio</td>
<td>Skaggs et al. (1981); Nolte et al. (1983); Desmond et al. (1996)</td>
</tr>
<tr>
<td>Georgia</td>
<td>Shirmohammadi et al. (1991)</td>
</tr>
<tr>
<td>Illinois</td>
<td>Mostaghimi et al. (1985); Kurien et al. (1997)</td>
</tr>
<tr>
<td>Louisiana</td>
<td>Fouss et al. (1987); Fouss et al. (1989)</td>
</tr>
<tr>
<td>Iowa</td>
<td>Sanoja et al. (1990)</td>
</tr>
<tr>
<td>Virginia</td>
<td>McMahon et al. (1988)</td>
</tr>
<tr>
<td>Australia</td>
<td>Cox et al. (1994)</td>
</tr>
<tr>
<td>Canada</td>
<td>Madramootoo (1990); Shukla et al. (1994)</td>
</tr>
<tr>
<td>Italy</td>
<td>Bixio and Bortolini (1997)</td>
</tr>
<tr>
<td>Mexico</td>
<td>Quej and Palacios-Valez (1996)</td>
</tr>
</tbody>
</table>

### 3.7 Model Sensitivity to Input Parameters

Expected goodness of fit depends on the amount of field-specific input data. The model is sensitive to saturated hydraulic conductivity and soil properties related to evapotranspiration and water balances in the unsaturated zone (Figure 3.4). Field specific input data sets can often produce simulated water table depths within a few cm of the observed water table data Skaggs (1982). In particular, lateral saturated hydraulic conductivity and a number of the relationships derived from the soil water characteristic including water table depth versus volume drained. The nitrogen component is sensitive to many of the rate factors including denitrification rates. Field measurements of many of these inputs can minimize the
amount of calibration and make the model results more representatives of the field observations

![Figure (3. 4) Effect of errors in input data on some of the outputs of DRAINMOD after Skaggs (1980)](image)

3.8 Model Application

The model was developed for application to drainage and water management problems in areas with poorly drained, high water table soils. Simulations of proposed system designs are conducted over a long period of weather record to evaluate the effectiveness of the design with varying weather conditions. Application of the model to other areas is generally discouraged since the model's procedures were developed based on soil water relationships in high water Table soils.
CHAPTER (4)

EVALUATION OF PEDOTRANSFER FUNCTIONS OF SATURATED HYDRAULIC CONDUCTIVITY

4.1 Introduction

Determination of the saturated hydraulic conductivity $K_{\text{sat}}$ is needed for many studies and applications related to irrigation, drainage, water movement and solute transport in the soil. Due to spatial variability of soil hydraulic properties, large numbers of measurements are often required to properly characterize such properties even at the field scale. $K_{\text{sat}}$ can be obtained from direct laboratory or field measurements, which become too costly and time consuming when large numbers of $K_{\text{sat}}$ values are required for large scale applications. Alternatively, $K_{\text{sat}}$ can be indirectly estimated in terms of the more widely available soil properties such as particle size distribution, bulk density, porosity, and organic matter content. This alternative involves using what is called pedotransfer functions (PTFs) (Bouma et al., 1989). The term PTF means transferring the data that we have to the data that we need. Many approaches have been used to develop PTFs including multi linear regression (MLR) (e.g. Brakensiek et al., 1984; Vereecken et al., 1989; Wösten et al., 1999), artificial neural network (ANN) (e.g. Schaap et al., 1998; Merdun et al., 2005; Parasaruman et al., 2006; Agyare et al., 2007), and group method of data handling (GMDH) (e.g. Ungaro et al., 2005; Nemes et al., 2005). In the last two decades, several PTFs have been developed to estimate saturated hydraulic conductivity (e.g. Saxton et al., 1986; Jabro, 1992; Rawls et al. 1998; Wosten et al., 2001; Julia et al., 2004; Nemes et al., 2005; Saxton and Rawls, 2006; Spychalski et al., 2007; Guarracino et al., 2007). These PTFs have a wide range of variability in regard to the accuracy of prediction and/or the input
requirements. Due to the lack of input data, some PTFs have limited applications (e.g. Rawls et al., 1993; Mbonimpa et al., 2002).

### 4.2 Previous Evaluations of Pedotransfer Functions of Saturated Hydraulic Conductivity

Relatively few studies have been conducted to evaluate the performance of the PTFs for predicting $K_{sat}$. Tietje and Hennings (1996) evaluated the performance of six PTFs (Cosby et al., 1984; Saxton et al., 1986; Brakensiek et al., 1984; Vereecken et al., 1990; and Campbell, 1985) using a data set of 1067 measured $K_{sat}$ for German soils (Table 4.1). Their results showed that three PTFs (Cosby et al., 1984; Saxton et al., 1986; Vereecken et al., 1989) overestimated $K_{sat}$ while the other three functions underestimated $K_{sat}$.

Table (4.1) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of six PTFs evaluated by Tietje and Hennings (1996)

<table>
<thead>
<tr>
<th>PTF author</th>
<th>GMER</th>
<th>GSDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brakensiek et al., 1984</td>
<td>0.17</td>
<td>11.52</td>
</tr>
<tr>
<td>Saxton et al., 1986</td>
<td>1.19</td>
<td>7.72</td>
</tr>
<tr>
<td>Vereecken et al., 1990</td>
<td>1.65</td>
<td>12.91</td>
</tr>
<tr>
<td>Cosby et al., 1984</td>
<td>1.42</td>
<td>7.86</td>
</tr>
<tr>
<td>Campbell, 1985</td>
<td>0.21</td>
<td>10.28</td>
</tr>
<tr>
<td>Bloemen, 1980</td>
<td>0.33</td>
<td>21.64</td>
</tr>
</tbody>
</table>

Minasny and McBratney (2000) used 462 samples of Australian soils to evaluate eight PTFs (K01: K08) of $K_{sat}$ (Campbell, 1985; Campbell and Shiozawa, 1994; Cosby et al., 1984; Puckett et al., 1984; Dane and Puckett, 1994; Brakensiek et al., 1984; Saxton et al., 1986; Schaap et al., 1998). Also, they used the same dataset to develop six PTFs for predicting $K_{sat}$ (K09:K14). They divided the database into three texture classes: sandy, loamy and clayey. They found that the function of Dane and Puckett (1994) (K04) has the best performance for sandy soils, the function of Cosby et al. (1984) (K03) has the best performance for loamy
soils and the Rosetta SSC-BD (an ANN model that requires % sand, %silt, %clay, and bulk density) (Schaap et al., 1998) (K07) has the best performance for clayey soils. The results of the evaluation of the fourteen PTFs are shown in Figures (4.1) and (4.2).

Figure (4.1) geometric mean of error ratio (GMER) of fourteen PTFs evaluated by Minasny and McBratney (2000)

Figure (4.2) geometric standard deviation of error ratio (GSDER) of fourteen PTFs evaluated by Minasny and McBratney (2000)
Sobieraj et al. (2001) evaluated the performance of nine PTFs of $K_{sat}$ (Brakensiek et al., 1984; Campbell and Shiozawa, 1994; Cosby et al., 1984; Jabro, 1992; Puckett et al., 1984; Dane and Puckett, 1994; Saxton et al., 1986; two functions by Schaap et al., 1998) (Fig. 4.3). They found that the functions of Rosetta SSC (an ANN function that requires % sand, %silt, %clay) (Schaap et al., 1998), Rosetta SSC-BD (Schaap et al., 1998) and Jabro, 1992 have the best performance in predicting $K_{sat}$. Lastly, Wagner et al. (2001) evaluated eight PTFs of $K_{sat}$ (Campbell, 1985; Vereecken et al., 1990; Cosby et al., 1984; Brakensiek et al., 1984; Wösten, 1997; Wösten et al., 1999; Saxton et al., 1986; Rawls et al., 1998) using 63 soil samples from Germany. The statistical analysis of the eight functions (Table 4.2) showed that The PTFs of Brakensiek et al. (1984), Campbell (1985), Saxton et al. (1986) and Wösten et al. (1999) tend to underestimate $K_{sat}$, while the other models generally overestimate $K_{sat}$.

![Figure (4.3) Measured vs. predicted saturated hydraulic conductivity of eight PTFs evaluated by Wagner et al. (2001)](image-url)
Table (4. 2) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of eight PTFs evaluated by Wagner et al. (2001)

<table>
<thead>
<tr>
<th></th>
<th>GMER</th>
<th>GSDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campbell, 1985</td>
<td>0.25</td>
<td>13.66</td>
</tr>
<tr>
<td>Vereecken et al., 1990</td>
<td>2.21</td>
<td>9.45</td>
</tr>
<tr>
<td>Cosby et al. (1984)</td>
<td>1.37</td>
<td>12.73</td>
</tr>
<tr>
<td>Brakensiek et al. (1984)</td>
<td>0.23</td>
<td>19.89</td>
</tr>
<tr>
<td>Wosten et al. (1997)</td>
<td>2.00</td>
<td>9.95</td>
</tr>
<tr>
<td>Wosten et al. (1999)</td>
<td>0.73</td>
<td>12.51</td>
</tr>
<tr>
<td>Saxton et al. (1986)</td>
<td>0.67</td>
<td>15.76</td>
</tr>
<tr>
<td>Rawls et al. (1998)</td>
<td>2.76</td>
<td>6.65</td>
</tr>
</tbody>
</table>

Many PTFs (e.g. Julia et al., 2004; Spychalski et al., 2007; Nemes et al., 2005; Saxton and Rawls, 2006; Guarracino et al., 2007; Li et al., 2007; Weynants et al., 2009) have been developed since the last evaluations by Wagner et al. (2001) and Sobieraj et al. (2001). The previous evaluations have also used relatively small soil data bases. The goal of this part of the study is to evaluate the performance of twenty four PTFs for predicting $K_{sat}$, including newly developed functions, using a relatively large data base of US soils that includes approximately 2000 data records.

### 4.3 Pedotransfer Functions for Predicting Saturated Hydraulic Conductivity

Twenty four Pedotransfer functions for estimating $K_{sat}$ were evaluated in this study. The PTFs were classified into three groups according to their input requirements.

### 4.3.1 Group1 Pedotransfer functions

This type of PTFs, first developed by Ahuja et al. (1984), is an empirical relationship between saturated hydraulic conductivity and effective porosity (Eq. 4.1).
\[ K_{\text{sat}} = c \cdot \varphi_e^m \]  \hspace{1cm} (4.1)

where \( \varphi_e \) is the effective porosity, defined as the difference between the total porosity and field capacity (water content at 33 kPa matric potential), c and m are empirical coefficients. The input requirements of these PTFs are available in most soil data bases. Six functions of this type were evaluated in the study (Table 4.3). Forrest et al. (1985) measured \( K_{\text{sat}} \) in the laboratory from 118 undisturbed soil cores from Australia and used these measurements to obtain the c and m coefficients (F1). Minasny and McBratney (2000) used 462 samples from Australian soils to develop a PTF for predicting \( K_{\text{sat}} \) with an effective porosity based on a field capacity determined at 10 kPa (F2). Suleiman and Ritchie (2000) used data from 11 homogeneous textural–class soils and several international and American soils to obtain the c and m coefficients of Ahuja et al.’s model (F3). Spychalski et al. (2007) developed three PTFs for predicting \( K_{\text{sat}} \) in terms of the effective porosity. They used a dataset of 35 measured soil samples and considered the field capacity as the water content at 10 kPa matric potential. One of these functions (F4) is in the form of Ahuja et al.’s model and the other two PTFs (F5, F6) are multi-linear regressions relating the hydraulic conductivity to the effective porosity (Table 4.3).

Table (4.3) Group 1 Pedotransfer functions (PTFs require inputs of effective porosity)

<table>
<thead>
<tr>
<th>PTF ID</th>
<th>Formula (cm/hr) ( \dagger )</th>
<th>Development Dataset</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>( K_{\text{sat}} = 0.1 \times \exp(10.8731 + 3.914 \ln(\varphi_e)) )</td>
<td>118 Australia</td>
<td>Forrest et al. (1985)</td>
</tr>
<tr>
<td>F2</td>
<td>( K_{\text{sat}} = 2319.055 \varphi_e^{3.66} )</td>
<td>462 Australia</td>
<td>Minasny &amp; McBr. (2000)</td>
</tr>
<tr>
<td>F3</td>
<td>( K_{\text{sat}} = 467.5 \varphi_e^{3.15} )</td>
<td>60 International</td>
<td>Suleiman and Ritchie (2000)</td>
</tr>
<tr>
<td>F4</td>
<td>( K_{\text{sat}} = 4031.57 \varphi_e^{3.295} )</td>
<td>35 USA</td>
<td>Spychalski et al. (2007)</td>
</tr>
<tr>
<td>F5</td>
<td>( K_{\text{sat}} = -2.52 + 581.598 \varphi_e^{1.5} - 6966.14 \varphi_e^{2.5} + 11693.78 \varphi_e^3 )</td>
<td>35 USA</td>
<td>Spychalski et al. (2007)</td>
</tr>
<tr>
<td>F6</td>
<td>( K_{\text{sat}} = -3.51 \times 18154.6 \varphi_e^{1.5} - 12213.8 \ln \varphi_e - 6925.78 \varphi_e / \ln \varphi_e )</td>
<td>35 USA</td>
<td>Spychalski et al. (2007)</td>
</tr>
</tbody>
</table>

\( \dagger K_{\text{sat}} \), saturated hydraulic conductivity (cm/hr); \( \varphi_e \), effective porosity
4.3.2 Group 2 Pedotransfer Functions

This group of PTFs requires more inputs than group 1. The PTFs of this group predict $K_{sat}$ in terms of particle size distribution data (%sand, %silt, and %clay), bulk density, and total porosity. Nine PTFs were identified under this category (Table 4.4). Puckett et al. (1984) used 42 US soil samples to develop an exponential function for predicting $K_{sat}$ in terms of clay content (F7). Using a larger dataset of 577 US samples, Dane and Puckett (1994) modified the PTF developed by Puckett et al. (1984) (F8). Julia et al. (2004) used 2178 measured samples from Spain to develop a PTF for predicting $K_{sat}$ in terms of sand content (F9). Cosby et al. (1984) used multi-linear regression to develop a PTF using measured soil properties of 1448 US soil samples with inputs of sand and clay contents (F10). Saxton et al. (1986) used 230 US data points to derive a PTF for predicting $K_{sat}$ in terms of sand and clay contents (F11). With the same database, Brakensiek et al. (1984) developed a PTF for predicting $K_{sat}$ in terms of clay and sand contents and soil porosity (F12). Jabro (1992) used 350 measured samples from international soils to develop PTF for predicting $K_{sat}$ with inputs of silt and clay contents and bulk density (F13). Schaap et al. (2001) developed an artificial neural network model called Rosetta for predicting $K_{sat}$ and soil water content using five levels of input requirements. Two of these five models (Rosetta SSC, Rosetta SSC-BD) were included in this group of PTFs (F14, F15).
Table (4.4) Group 2 Pedotransfer functions (PTFs that require inputs of %sand, %silt, and %clay, bulk density and porosity)

<table>
<thead>
<tr>
<th>PTF ID</th>
<th>Formula (cm/hr) †</th>
<th>Development Dataset</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>F7</td>
<td>(K_{sat} = 15.696 \exp(-0.1975CL))</td>
<td>42 USA</td>
<td>Puckett et al. (1984)</td>
</tr>
<tr>
<td>F8</td>
<td>(K_{sat} = 30.384 \exp(-0.144CL))</td>
<td>577 USA</td>
<td>Dane &amp; Puckett (1994)</td>
</tr>
<tr>
<td>F9</td>
<td>(K_{sat} = 0.0920e^{0.0492SA})</td>
<td>2178 Spain</td>
<td>Julia et al. (2004)</td>
</tr>
<tr>
<td>F10</td>
<td>(K_{sat} = 2.54 \times 10^{(-0.6+0.022SA-0.0064CL)})</td>
<td>1448 USA</td>
<td>Cosby et al. (1984)</td>
</tr>
<tr>
<td>F11</td>
<td>(K_{sat} = \exp[12.01 - 0.0755SA + (-3.895 + 0.03671SA + 0.1103CL + 0.00087546CL^{2})/\theta_{s}])</td>
<td>230 USA</td>
<td>Saxton et al. (1986)</td>
</tr>
<tr>
<td>F12</td>
<td>(K_{sat} = \exp[19.52388 + 8.96847 - 0.028212CL + 0.00018107SA2 - 0.0094125CL2 - 0.019492CL2\phi^2 + 0.0000173SA2CL + 0.02733CL2\phi^2 + 0.001434SA2\phi^{2} - 0.0000035CL2SA])</td>
<td>230 USA</td>
<td>Brakensiek et al. (1984)</td>
</tr>
<tr>
<td>F13</td>
<td>(\log K_{sat} = 9.56 - 0.81\log SI - 1.09\log CL - 4.64BD)</td>
<td>350 International</td>
<td>Jabro (1992)</td>
</tr>
<tr>
<td>F14</td>
<td>Artificial neural network needs inputs of (sand, silt and clay contents)</td>
<td>1306 USA &amp; Europe</td>
<td>&amp; Schaap et al. (2001)</td>
</tr>
<tr>
<td>F15</td>
<td>Artificial neural network needs inputs of (sand, silt and clay contents and bulk density)</td>
<td>1306 USA &amp; Europe</td>
<td>&amp; Schaap et al. (2001)</td>
</tr>
</tbody>
</table>

†\(K_{sat}\), saturated hydraulic conductivity (cm/hr); SA, sand content (%); SI, silt content (%); CL, clay content (%); BD, bulk density (gm/cm\(^3\)); \(\phi\), total porosity (cm\(^3\)/cm\(^3\))

4.3.3 Group 3 Pedotransfer Functions

This group of PTFs requires inputs of particle size distribution, bulk density, and organic matter content. The lack of organic matter measurements in soil databases limits the applicability of the PTFs of this group. Nine PTFs were identified under this category (Table 4.5). Julia et al. (2004) used 2178 measured samples from Spanish soils to develop a PTF for predicting \(K_{sat}\) in terms of sand, clay and organic matter contents (F16). Using the same data set, Saxton and Rawls (2006) developed a PTF that requires the same input parameters (F17). Wösten et al. (1999) used 1136 soil samples from the HYdraulic PRoperties of European Soils database (HYPRES) (Wösten et al., 1999) to develop a PTF for predicting \(K_{sat}\) in terms of clay, silt and organic matter contents and bulk density (F23). Then,
Wöstien et al. (2001) developed two PTFs for Dutch soils, one function for sandy soils that predicts $K_{sat}$ in terms of silt and organic matter contents and bulk density (F18), and the other function for loamy and clayey soils with inputs of clay and organic matter contents and bulk density (F19).

Vereecken et al. (1990) used 182 measured soil samples from Belgium to develop a PTF for predicting $K_{sat}$ in terms of sand, clay, and organic matter contents and bulk density (F20). Recently, Weynants et al. (2009) used the same dataset to develop a PTF for predicting $K_{sat}$ in terms of sand and organic matter contents and bulk density (F21). Nemes et al. (2005) used group method of data

![Table (4. 5) Group 3 pedotransfer functions (PTFs that require inputs of particle size distribution bulk density and organic matter content.](image)

<table>
<thead>
<tr>
<th>PTF ID</th>
<th>Formula (cm/hr) †</th>
<th>Development Dataset</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>F16</td>
<td>$K_{sat}=0.1*(-4.994+0.56728SA-0.131CL-0.0127OM)$</td>
<td>2178 Spain</td>
<td>Julia et al. (2004)</td>
</tr>
<tr>
<td>F17</td>
<td>$K_{sat}=\theta_s-\theta_{33}$ where $\theta_s$, $\theta_{33}$, and $\lambda = f(SA, CL, OM)$</td>
<td>2178 Spain</td>
<td>Saxton &amp; Rawls (2006)</td>
</tr>
<tr>
<td>F18</td>
<td>$K_{sat}=0.04167*\exp(45.8-14.34BD+0.001481SI^2-27.5BD^{-1}-0.891\ln(SI)-0.34\ln(OM))$</td>
<td>832 Netherlands</td>
<td>Wosten et al. (2001)</td>
</tr>
<tr>
<td>F19</td>
<td>$K_{sat}=0.04167*\exp(-42.6+8.71OM+61.9BD-20.79BD^{-2}-0.2107OM^2-0.0162CL<em>OM-5.382BD</em>OM)$</td>
<td>832 Netherlands</td>
<td>Wosten et al. (2001)</td>
</tr>
<tr>
<td>F20</td>
<td>$K_{sat}=0.04167*\exp[20.62-0.96lnCL-0.66lnSA-0.46lnOM-8.43BD]$</td>
<td>182 Belgium</td>
<td>Vereecken et al. (1990)</td>
</tr>
<tr>
<td>F21</td>
<td>$K_{sat}=\exp(1.9582+0.0308SA-0.6142BD-0.1566OM)/24$</td>
<td>182 Belgium</td>
<td>Weynants et al. (2009)</td>
</tr>
<tr>
<td>F22</td>
<td>$K_{sat}=0.04167*(10^{0.571+0.956Z4})$ , $Z4= f(SA,CL,BD,OM)$</td>
<td>886 USA</td>
<td>Nemes et al. (2005)</td>
</tr>
<tr>
<td>F23</td>
<td>$K_{sat}=0.04167*\exp[7.755+0.0352SI+0.93-0.976BD-0.00048CL^2-0.000322SI^2+0.001SI^{-1}-0.0748OM^{-1}-0.643\lnSI-0.0139(BD.CL)-0.167(BD,OM)+0.0298CL-0.03305SI]$</td>
<td>1136 Europe</td>
<td>Wosten et al. (1999)</td>
</tr>
<tr>
<td>F24</td>
<td>$K_{sat}=0.04167*\exp(13.262-1.914lnSA-0.974lnSI-0.058CL\cdot 36+1.709lnOM+2.885OM-8.026\lnBD)$</td>
<td>36 China</td>
<td>Li et al. (2007)</td>
</tr>
</tbody>
</table>

† $K_{sat}$, saturated hydraulic conductivity (cm/hr); SA, sand content (%); SI, silt content (%); CL, clay content (%); OM, organic matter content (%); BD, bulk density (g/cm$^3$); $\theta_s$, total porosity (cm$^3$/cm$^3$); $\theta_{33}$, water content at 33 kPa (cm$^3$/cm$^3$).
handling to develop a PTF for predicting $K_{sat}$ using 886 samples from American soils (F22). Finally, Li et al. (2007) made 36 measurements of $K_{sat}$ for seven soil profiles collected from northern China and developed a PTF for predicting $K_{sat}$ in terms of sand, silt, clay and organic matter contents and bulk density (F24).

4.4 Soil Databases

The U.S. soils data used in evaluating the PTFs were obtained from different sources. Most of the data were obtained from SOILVISION, an international soil database that contains more than 6000 records of measured soil properties including soil water retention data, saturated hydraulic conductivity, particle size distribution, bulk density, porosity, and organic matter content (Fredlund et al., 1996; 1998). This database encompasses more than twelve databases including the UNSODA database (Leij et al., 1996). SOILVISION is available online at http://www.soilvision.com/soilvision_downloads.shtml. We obtained 160 records of saturated hydraulic conductivity, particle size distribution, bulk density, porosity, and organic matter content, measured for Oklahoma state soils (Mohanty et al., 2002). This small data set is available online at http://disc.sci.gsfc.nasa.gov/fieldexp/SGP97/arssl.html. We also used a database for New Jersey soils, obtained from the Natural Resources Conservation Service’s (NRCS) soil database (http://soils.usda.gov/survey/research/investigation.html). This database includes measurements of $K_{sat}$, particle size distribution, porosity, bulk density, organic matter content and soil water retention data. Additionally, we used small data sets found in the literature (e.g. Dane and Puckett, 1994; Rawls et al., 1998). Many soil records of the databases did not have all the input requirements for all PTFs. Therefore, the soil data obtained from all databases (Figure 4.5) were grouped into three categories according to the input requirements of the three groups of PTFs.
**Data set 1**, used for evaluating the first group of PTFs, includes 1911 records of saturated hydraulic conductivity, soil porosity and soil water contents at 10 and 33 kPa, measured for U.S. soils. These measurements are generally available in most soil databases. A statistical summary of data set 1 is given in Table (4.6).

<table>
<thead>
<tr>
<th></th>
<th>( \varphi ) (cm(^3)/cm(^3))</th>
<th>( \theta_{10} ) (cm(^3)/cm(^3))</th>
<th>( \theta_{33} ) (cm(^3)/cm(^3))</th>
<th>( K_{sat} ) (cm/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0.24</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>max</td>
<td>0.92</td>
<td>0.79</td>
<td>0.70</td>
<td>450.00</td>
</tr>
<tr>
<td>avg</td>
<td>0.44</td>
<td>0.19</td>
<td>0.16</td>
<td>19.29</td>
</tr>
<tr>
<td>Stdv</td>
<td>0.07</td>
<td>0.11</td>
<td>0.11</td>
<td>29.32</td>
</tr>
<tr>
<td>med</td>
<td>0.44</td>
<td>0.19</td>
<td>0.16</td>
<td>8.33</td>
</tr>
</tbody>
</table>

Min, minimum value; max, maximum value; avg, average; stdv, standard deviation; med, median; \( \varphi \), total porosity; \( \theta_{10} \), soil water content at 10 kPa matric potential; \( \theta_{33} \), soil water content at 33 kPa matric potential

**Data set 2**, used for evaluating the second group of PTFs, includes measurements of saturated hydraulic conductivity, particle size distribution, bulk density. The total porosity is required as an input for this group of PTFs, so it was estimated based on the measured bulk density and estimated particle density of 2.65 gm/cm\(^3\). The number of records for U.S. soils in this data set was 956, about half the number of records for data set 1. The smaller size of data set 2, compared to data set 1, was primarily due to the lack of particle size distribution measurements in the soil databases. A statistical summary of data set 2 is given in Table 4.7. The data set was divided into five subsets according to soil texture (USDA classification) and the functions were evaluated using each data subset. Due to the small number of records available for the very fine and fine texture classes, the two classes were combined together into one class.
Figure (4. 4) Textural classification system according to the FAO Soil Map of Europe (1998), the FAO classes are: C (coarse), M (medium), MF (medium fine), F(fine), VF (very fine).

Figure (4. 5) Distribution of sand and clay contents for the dataset used for evaluating PTFs of Group (2)
Table 4.7. A statistical summary of data set 2.

<table>
<thead>
<tr>
<th></th>
<th>sand</th>
<th>silt</th>
<th>clay</th>
<th>BD</th>
<th>K&lt;sub&gt;sat&lt;/sub&gt;&lt;sup&gt;†&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0.90</td>
<td>0.04</td>
<td>0.02</td>
<td>0.41</td>
<td>0.02</td>
</tr>
<tr>
<td>max</td>
<td>99.96</td>
<td>85.83</td>
<td>64.00</td>
<td>2.13</td>
<td>530.64</td>
</tr>
<tr>
<td>avg</td>
<td>57.44</td>
<td>27.06</td>
<td>15.49</td>
<td>1.48</td>
<td>14.91</td>
</tr>
<tr>
<td>stdv</td>
<td>27.52</td>
<td>20.33</td>
<td>12.38</td>
<td>0.19</td>
<td>86.69</td>
</tr>
<tr>
<td>med</td>
<td>63.00</td>
<td>21.75</td>
<td>13.29</td>
<td>1.49</td>
<td>2.53</td>
</tr>
</tbody>
</table>

Min, minimum value; max, maximum value; avg, average; stdv, standard deviation; med, median; BD, bulk density<sup>†</sup> avg, stdev, and med values for K<sub>sat</sub> were calculated according to the lognormal distribution equations.

**Data set 3** is the soil data set that was used for evaluating the third group of PTFs. This data set includes measurements of saturated hydraulic conductivity, particle size distribution, bulk density and organic matter content. The number of records of U.S soils in this data set is 678, which is considerably less than the other two data sets, primarily because of the lack of organic matter content measurements in the soil data bases. A statistical summary of data set 3 is given in Table 4.8 and Figure 4.6. This data set was also divided into five subsets similar to data set 2 and the functions were evaluated using each subset. Due to the small number of records available for the very fine and fine texture classes, the two classes were combined together into one class.
4.5 Statistical Analysis

Since $K_{\text{sat}}$ exhibits long-tailed distribution (Figure 4.7), it is best described by lognormal distributions and log transformations are usually used for developing the PTFs (Minasny and McBratney, 2000). The statistical evaluation of the PTFs was conducted using the geometric mean error ratio (GMER, Eq. 4.3) and the geometric standard deviation of the error ratio (GSDER, Eq. 4.4), which are
calculated from the error ratio ($\varepsilon$, Eq. 4.2) of the measured vs. predicted values (Tietje and Hennings, 1996; Minasny and McBratney, 2000; Wagner et al., 2001).

![Graph showing frequency distribution of saturated hydraulic conductivities](image)

**Figure (4.7)** Frequency distribution of the three datasets of measured saturated hydraulic conductivities used for the evaluation of PTFs.

$$\varepsilon = \frac{K_p}{K_m}$$  \hspace{1cm} (4.2)

$$\text{GMEM} = \exp \left( \frac{1}{n} \sum_{i=1}^{n} \ln(\varepsilon_i) \right)$$  \hspace{1cm} (4.3)

$$\text{GSDER} = \exp \left[ \left( \frac{1}{n-1} \sum_{i=1}^{n} [\ln(\varepsilon_i) - \ln(\text{GMEM})]^2 \right)^{1/2} \right]$$  \hspace{1cm} (4.4)

where $K_p$ and $K_m$ are predicted and measured hydraulic conductivities, respectively. A GMER of one corresponds to an exact match between measured and predicted values; GMER less than one indicates underestimation and GMER greater than one indicates overestimation by the predictive model. A GSDER of 1 corresponds to a perfect match and it grows with the deviation from the measured values.
4.6 Results and Discussions

4.6.1 Performance of Group 1 PTFs

The results of evaluating the six PTFs of group 1 are summarized in Table (4.9) and Fig. (4.10). All functions of this group overestimated $K_{sat}$ as indicated by the greater than one GMER for all functions. The model developed by Suleiman and Ritchie (2000), F3, gave the closest predictions to the measured values (GMER=1.28) followed by Minasny and McBratney’s (2000) model, F2, with a GMER of 1.89 (Table 4.9, Figures 4.8, 4.9). The predictions of the other four models were about five to six times higher than the measured values. The three models developed by Spychalski et al. (2007) (F4, F5, and F6) gave very close predictions even though they have different formulae. The poor performance of these three models could be attributed to the relatively small size (n=35 records) of the data set that was used in developing the models. The GSDER values of the six PTFs varied over a small range from 5.26 to 7.00. The function with the least GMER, F3, has a GSDER value of 5.88. However, the function with the second least GMER, F2, has the highest GSDER. Thus the results of this evaluation indicate that the function developed by Suleiman and Ritchie (2000) is the most accurate function among group 1 PTFs. A similar conclusion can be reached by visual inspection of the scatter diagrams of Fig. (4.10).

<table>
<thead>
<tr>
<th>PTFs</th>
<th>GMER</th>
<th>GSDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>5.18</td>
<td>6.22</td>
</tr>
<tr>
<td>F2</td>
<td>1.89</td>
<td>7.00</td>
</tr>
<tr>
<td>F3</td>
<td>1.28</td>
<td>5.88</td>
</tr>
<tr>
<td>F4</td>
<td>5.66</td>
<td>6.55</td>
</tr>
<tr>
<td>F5</td>
<td>6.06</td>
<td>5.26</td>
</tr>
<tr>
<td>F6</td>
<td>6.10</td>
<td>5.55</td>
</tr>
</tbody>
</table>
Figure (4.8) Geometric mean of error ratio for PTFs of group (1)

Figure (4.9) Geometric standard deviation of error ratio for PTFs of group (1)
4.6.2 Performance of Group 2 PTFs

The results of evaluating group 2 PTFs are summarized in Table (4.10) and Figures (4.11, 4.12, 4.13). The results of evaluating the PTFs for different soil textural classes showed that group 2 functions tend to over-estimate $K_{sat}$ for the coarse textural class and under-estimate $K_{sat}$ for the other textural classes. The function F10 (Cosby et al., 1984) showed the best performance for all textural classes but the coarse textural class. The GMER values of F10 were close to one for all textural classes with a small range of 0.84 for the very fine-fine textural class to 1.17 for the coarse textural class. The F14 (Rosetta SSC model, Schaap et al., 2001) showed a performance very close to F10 for the very fine-fine (GMER=0.84 for F10 and 0.74 for F14) and fine-medium (GMER=1.02 for F10
and 1.08 for F14) textural classes but it was outperformed by F10 for the medium (GMER=0.88 for F10 and 0.41 for F14) and the coarse (GMER=1.17 for F10 and 1.55 for F14) textural classes. Both F8 (Dane and Puckett, 1994) and F13 (Jabro, 1992) showed performance comparable to F10 for the medium textural class (GMER=1.19 for F8, 0.86 for F13, and 0.88 for F10), but performed poorly for the other textural classes. The function F7 (Puckett et al., 1984) had the best performance for the coarse textural class, closely followed by F9 (Julia et al., 2004) and F10 (GMER=1.07 for F7, 1.14 for F9, and 1.17 for F10). However, F7 showed poor performance for the other textural classes especially for the very fine-fine textural class. The results also show that the general performance of the PTFs of group 2 is significantly higher for the coarse textural class, compared with the very fine-fine textural class. For the coarse textural class, the GMER and GSDER values for the nine functions of group 2 were in the range of 1.07 to 2.72 and 4.5 to 9.47, respectively. The corresponding ranges of GMER and GSDER for the very fine-fine textural class were 0.00 to 1.96 and 8.74 to 110.00, respectively.

The results of evaluating the PTFs using the entire dataset indicated that F10 had also the highest overall performance with a very close to one GMER of 0.98 and the least GSDER of 5.98 among the nine functions of this group. The functions F14, F15, F11, and F8 followed F10 with GMER values ranging from 0.75 to 1.29. The GMER values were less than one for seven out of the nine evaluated functions, which indicates that group 2 PTFs tend to underestimate $K_{sat}$.
Table (4. 9) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of group 2 Pedotransfer functions

<table>
<thead>
<tr>
<th></th>
<th>Fine &amp; Very Fine</th>
<th>Medium Fine</th>
<th>Medium</th>
<th>Coarse</th>
<th>Total Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GMER</td>
<td>GSDER</td>
<td>GMER</td>
<td>GSDER</td>
<td>GMER</td>
</tr>
<tr>
<td>F7</td>
<td>0.00</td>
<td>17.53</td>
<td>0.46</td>
<td>8.79</td>
<td>0.24</td>
</tr>
<tr>
<td>F8</td>
<td>0.07</td>
<td>13.72</td>
<td>2.62</td>
<td>7.63</td>
<td>1.19</td>
</tr>
<tr>
<td>F9</td>
<td>0.37</td>
<td>12.13</td>
<td>0.25</td>
<td>6.93</td>
<td>0.46</td>
</tr>
<tr>
<td>F10</td>
<td>0.84</td>
<td>10.59</td>
<td>1.02</td>
<td>6.92</td>
<td>0.88</td>
</tr>
<tr>
<td>F11</td>
<td>0.33</td>
<td>8.74</td>
<td>2.13</td>
<td>7.12</td>
<td>0.55</td>
</tr>
<tr>
<td>F12</td>
<td>0.05</td>
<td>110.00</td>
<td>0.06</td>
<td>20.25</td>
<td>0.51</td>
</tr>
<tr>
<td>F13</td>
<td>1.96</td>
<td>18.55</td>
<td>0.59</td>
<td>8.01</td>
<td>0.86</td>
</tr>
<tr>
<td>F14</td>
<td>0.74</td>
<td>8.89</td>
<td>1.08</td>
<td>6.72</td>
<td>0.41</td>
</tr>
<tr>
<td>F15</td>
<td>0.68</td>
<td>10.20</td>
<td>0.64</td>
<td>6.30</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Figure (4. 11) Geometric mean of error ratio for PTFs of group (2)

Figure (4. 12) Geometric standard deviation of error ratio for PTFs of group (2)
The results of evaluating group 3 PTFs are summarized in Table (4.11) and Figures (4.14, 4.15, and 4.15). The results showed that F22 (Nemes et al., 2005) consistently had good performance for all texture classes while other functions, including F16, F17, F19, and F23, showed better performance for only one or two textural classes. The function F19 had the most accurate predictions of $K_{\text{sat}}$ for the
very fine-fine and fine-medium textural classes with GMER values of 0.99 and 0.6, respectively. F17, closely followed by F22 and F16, are the most accurate functions for predicting $K_{sat}$ for the medium textural class (GMER=0.91 for F17, 1.11 for F22, 1.16 for F16). F22 closely followed by F23 are the highest performing functions for the coarse textural class with GMER values of 1.11 and 1.15, respectively. None of the functions F18, F20, F21, and F24 was among the highest performing functions for any of the textural classes. The high performing functions for the medium and coarse textural classes have closer to one GMER and smaller GSDER than the high performing functions for the very fine-fine and fine-medium textural classes, which indicates that the PTFs of this group are more effective in predicting $K_{sat}$ for the medium and coarse textural classes.

The results of evaluating the nine functions using the entire data set showed that F22 has the highest overall performance with the lowest GMER of 1.11 and second to lowest GSDER of 5.80. Functions F16 and F17 followed F22 with GMER values of 0.80 and 0.88, respectively, and GSDER values of 6.03 and 6.35, respectively. Despite this comparable overall performance to F22, both F16 and F17 had significantly less accurate predictions of $K_{sat}$ for the very fine-fine and fine-medium textural classes. Overall, group 3 PTFs tend to under-predict $K_{sat}$. The results also indicated that the models developed using data for U.S. soils, such as F22 and F17, performed better than the other models.
Table (4.10) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of group 3 Pedotransfer functions

<table>
<thead>
<tr>
<th></th>
<th>Fine &amp; Very Fine</th>
<th>Medium Fine</th>
<th>Medium</th>
<th>Coarse</th>
<th>Total Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GMER</td>
<td>GSDER</td>
<td>GMER</td>
<td>GSDER</td>
<td>GMER</td>
</tr>
<tr>
<td>F16</td>
<td>0.64</td>
<td>8.73</td>
<td>0.10</td>
<td>10.17</td>
<td>1.16</td>
</tr>
<tr>
<td>F17</td>
<td>0.12</td>
<td>14.21</td>
<td>0.14</td>
<td>10.48</td>
<td>0.91</td>
</tr>
<tr>
<td>F18</td>
<td>6.09</td>
<td>18.02</td>
<td>9.56</td>
<td>10.03</td>
<td>3.50</td>
</tr>
<tr>
<td>F19</td>
<td>0.99</td>
<td>13.94</td>
<td>0.60</td>
<td>9.42</td>
<td>0.47</td>
</tr>
<tr>
<td>F20</td>
<td>1.55</td>
<td>12.08</td>
<td>6.13</td>
<td>8.31</td>
<td>0.83</td>
</tr>
<tr>
<td>F21</td>
<td>0.24</td>
<td>9.85</td>
<td>0.08</td>
<td>9.63</td>
<td>0.32</td>
</tr>
<tr>
<td>F22</td>
<td>2.10</td>
<td>9.91</td>
<td>0.45</td>
<td>8.15</td>
<td>1.11</td>
</tr>
<tr>
<td>F23</td>
<td>1.85</td>
<td>10.37</td>
<td>0.29</td>
<td>10.02</td>
<td>1.86</td>
</tr>
<tr>
<td>F24</td>
<td>0.74</td>
<td>20.99</td>
<td>3.05</td>
<td>8.25</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Figure (4.14) Geometric mean of error ratio for PTFs of group (3)
Figure (4.15) Geometric standard deviation of error ratio for PTFs of group (3)
4.6.4 Performance of All Pedotransfer Functions Evaluated using Same Dataset

The twenty four PTFs were evaluated using data set 3, which includes all input requirements of all functions. The purpose of this evaluation was to identify the best performing functions when the availability of inputs is not a limiting factor in selecting the PTF. The results of this evaluation are summarized in Table (4.12)
and Figures (4.17, 4.18, and 4.19). The results showed that the performance of PTFs was not affected by the size of the data set. For group 1 PTFs, F3 and F2 showed the highest overall performance with GMER of 1.34 and 1.39, respectively. The function F10 had the highest overall performance among group 2 PTFs with a GMER of 1.02, followed by F11, F14, and F15, with GMER of 0.79, 0.74, and 0.72, respectively. The function F10 was the highest performing function among the functions of the three groups, followed by the function F22.

The performance of the twenty four PTFs was examined for every textural class. For the very fine-fine textural class, the functions F1 and F4 of group 1 and F13 of group 2 had a very close to one GMER (from 1.01 to 1.20) but had high GSDER values ranging from 18.99 to 20.77. The large scatter around the GMER, indicated by the high GSDER values for these functions, can also be observed by examining the scatter diagrams for these functions (Fig. 4.19). The functions F14 of group 2 and F16 of group 3 had reasonable values of GMER (0.76 and 0.64, respectively) and GSDER (9.27 and 8.73, respectively). The function F10, the highest performing function for this textural class according to the previous evaluation with the larger data set 2, followed F14 and F16 with a GMER of 0.59 and a GSDER of 10.72. For the fine-medium textural classes, the functions F8, F13 and F11 of group 2 and F1 of group 1 were the highest performing functions with GMER of 1.06, 1.06, 0.90, and 1.20, respectively, and GSDER values varying over a small range of 7.10 to 10.82. Both F10 and F14, the highest performing functions according to the previous evaluation with the larger data set 2, showed relatively poor performance when evaluated using the smaller data set 3 (GMER=0.38 for F10, 0.41 for F14; GSDER=8.12 for F10, 8.54 for F14). These results indicated that, the performance of the PTFs for this textural class is clearly influenced by the size of the data set used in the evaluation. For the medium textural class, the group 3 PTFs F17, F22, and F16 had the highest performance followed by the group 2 PTFs F10 and F11. Many functions, including F3 of
group 1, F9, F10, F11, F14 and F15 of group 2, and F22 and F23 of group 3, performed well in predicting $K_{\text{sat}}$ for the coarse textural class.

Table (4.11) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of groups 1, 2, and 3 Pedotransfer functions using the same dataset

<table>
<thead>
<tr>
<th></th>
<th>Fine &amp; Very Fine</th>
<th>Medium Fine</th>
<th>Medium</th>
<th>Coarse</th>
<th>Total Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GMER</td>
<td>GSDER</td>
<td>GMER</td>
<td>GSDER</td>
<td>GMER</td>
</tr>
<tr>
<td><strong>Group (1) PTFs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>1.17</td>
<td>20.77</td>
<td>1.20</td>
<td>7.10</td>
<td>6.12</td>
</tr>
<tr>
<td>F2</td>
<td>0.27</td>
<td>21.23</td>
<td>0.14</td>
<td>15.62</td>
<td>1.84</td>
</tr>
<tr>
<td>F3</td>
<td>0.56</td>
<td>17.04</td>
<td>0.50</td>
<td>6.64</td>
<td>1.87</td>
</tr>
<tr>
<td>F4</td>
<td>1.20</td>
<td>18.99</td>
<td>0.63</td>
<td>13.77</td>
<td>6.26</td>
</tr>
<tr>
<td>F5</td>
<td>4.19</td>
<td>14.31</td>
<td>2.37</td>
<td>9.93</td>
<td>8.40</td>
</tr>
<tr>
<td>F6</td>
<td>2.54</td>
<td>17.10</td>
<td>1.35</td>
<td>12.03</td>
<td>7.84</td>
</tr>
<tr>
<td><strong>Group (2) PTFs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F7</td>
<td>0.00</td>
<td>58.13</td>
<td>0.19</td>
<td>13.62</td>
<td>0.33</td>
</tr>
<tr>
<td>F8</td>
<td>0.02</td>
<td>32.32</td>
<td>1.06</td>
<td>10.82</td>
<td>1.64</td>
</tr>
<tr>
<td>F9</td>
<td>0.27</td>
<td>10.12</td>
<td>0.09</td>
<td>8.26</td>
<td>0.66</td>
</tr>
<tr>
<td>F10</td>
<td>0.59</td>
<td>10.72</td>
<td>0.38</td>
<td>8.12</td>
<td>1.25</td>
</tr>
<tr>
<td>F11</td>
<td>0.31</td>
<td>11.16</td>
<td>0.90</td>
<td>8.78</td>
<td>0.76</td>
</tr>
<tr>
<td>F12</td>
<td>0.01</td>
<td>77.09</td>
<td>0.07</td>
<td>8.75</td>
<td>0.55</td>
</tr>
<tr>
<td>F13</td>
<td>1.01</td>
<td>20.35</td>
<td>1.06</td>
<td>8.97</td>
<td>0.89</td>
</tr>
<tr>
<td>F14</td>
<td>0.76</td>
<td>9.27</td>
<td>0.41</td>
<td>8.54</td>
<td>0.56</td>
</tr>
<tr>
<td>F15</td>
<td>0.48</td>
<td>11.37</td>
<td>0.47</td>
<td>8.48</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>Group (3) PTFs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F16</td>
<td>0.64</td>
<td>8.73</td>
<td>0.10</td>
<td>10.17</td>
<td>1.16</td>
</tr>
<tr>
<td>F17</td>
<td>0.12</td>
<td>14.21</td>
<td>0.14</td>
<td>10.48</td>
<td>0.91</td>
</tr>
<tr>
<td>F18</td>
<td>6.09</td>
<td>18.02</td>
<td>9.56</td>
<td>10.03</td>
<td>3.50</td>
</tr>
<tr>
<td>F19</td>
<td>0.99</td>
<td>13.94</td>
<td>0.60</td>
<td>9.42</td>
<td>0.47</td>
</tr>
<tr>
<td>F20</td>
<td>1.55</td>
<td>12.08</td>
<td>6.13</td>
<td>8.31</td>
<td>0.83</td>
</tr>
<tr>
<td>F21</td>
<td>0.24</td>
<td>9.85</td>
<td>0.08</td>
<td>9.63</td>
<td>0.32</td>
</tr>
<tr>
<td>F22</td>
<td>2.10</td>
<td>9.91</td>
<td>0.45</td>
<td>8.15</td>
<td>1.11</td>
</tr>
<tr>
<td>F23</td>
<td>1.85</td>
<td>10.37</td>
<td>0.29</td>
<td>10.02</td>
<td>1.86</td>
</tr>
<tr>
<td>F24</td>
<td>0.74</td>
<td>20.99</td>
<td>3.05</td>
<td>8.25</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Figure (4.17) Geometric mean of error ratio for all PTFs of groups 1, 2, and 3

Figure (4.18) Geometric standard deviation of error ratio for all PTFs
Figure (4.19) Predicted vs. measured values of saturated hydraulic conductivity predicted by all Pedotransfer functions using the same dataset.
4.7 Comparison between the Results of This Study and Previous Studies

Results of this study were compared with results of four previous studies, which evaluated PTFs for predicting $K_{\text{sat}}$ (Table 4.13). Three of the four studies (Tietje and Hennings, 1996; Minasny and McBratney, 2000; Wagner et al., 2001) used GMER and GSDER as performance measures while the fourth study (Julia et al., 2004) used the mean deviation (MD) and the root mean square deviation (RMSD) as performance measures. Eleven functions, out of the 24 functions evaluated in this study, have been previously evaluated at least once. The functions F10 and F11 have been evaluated in the four previous studies. The function F12 has been evaluated three times and the functions F8 and F20 have been evaluated twice. The rest of the eleven functions have been evaluated only once.

The best performing function in this study, F10, had an overall good performance in the previous studies but it was not always the highest performing function. For example, the function F11 outperformed F10 in the evaluations conducted by Tietje and Hennings (1996) and Julia et al. (2004). The current study did not show a bias in F10’s predictions. However, F10 over-estimated $K_{\text{sat}}$ in three of the previous studies (Tietje and Hennings, 1996; Minasny and McBratney, 2000; Wagner et al., 2001) with GMER values ranging from 1.37 to 1.46 and under-estimated $K_{\text{sat}}$ in the more recent evaluation by Julia et al. (2004). The relatively high performance of F11 in our study is consistent with the previous studies with the exception of the evaluation conducted by Minasny and McBratney (2000), which showed a very poor performance of F11 as indicated by the close to zero GMER. Both the current study and the study by Julia et al. (2004) showed a good performance of functions F15 and F16 with tendency to under-estimate $K_{\text{sat}}$, as indicated by the less than one GMER and the negative value of the MD. The function F12 performed better in this study than the previous studies, which consistently showed poor performance of F12 with GMER ranging from 0.17 to 0.23. The performance of the functions F2, F7, F8, F9, F20 and F23 obtained in
this study was not consistent with previous studies. Our study showed that F2, F8, and F23 reasonably predicted $K_{sat}$ with tendency to over-estimate its value (GMER in the range of 1.29-1.89). These functions however, under-estimated $K_{sat}$ in previous evaluations. The other three functions F7, F9, and F20 under-estimated $K_{sat}$ in our study but over-estimated $K_{sat}$ in the previous studies.

Table (4.12) Geometric mean of error ratio (GMER) and geometric standard deviation of error ratio (GSDER) of PTFs of saturated hydraulic conductivity in current study compared to results

<table>
<thead>
<tr>
<th>Authors</th>
<th>N</th>
<th>F2</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>F15</th>
<th>F16</th>
<th>F20</th>
<th>F23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tietje and Hennings (1996)</td>
<td>1067</td>
<td>GMER</td>
<td>1.42</td>
<td>1.19</td>
<td>0.17</td>
<td>1.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minasny and McBratney</td>
<td>462</td>
<td>GMER 0.682</td>
<td>0.429</td>
<td>1.463</td>
<td>0.098</td>
<td>0.211</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wagner et al. 2001</td>
<td>63</td>
<td>GMER 2.656</td>
<td>1.723</td>
<td>1.716</td>
<td>1.757</td>
<td>1.865</td>
<td>1.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Julia et al. 2004</td>
<td>2178</td>
<td>MD 2.152</td>
<td>0.16</td>
<td>0.039</td>
<td>0.508</td>
<td>0.202</td>
<td>-0.256</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current study</td>
<td>956</td>
<td>GMER 1.89</td>
<td>1.6</td>
<td>0.7</td>
<td>0.91</td>
<td>1.11</td>
<td></td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>678</td>
<td>GSDER 9.46</td>
<td>12.91</td>
<td>8.40</td>
<td>5.95</td>
<td>5.47</td>
<td>5.60</td>
<td></td>
<td>5.78</td>
<td>6.03</td>
<td>9.26</td>
<td>6.09</td>
</tr>
</tbody>
</table>

4.8 Application Procedure of Pedotransfer Functions

The results of this study can facilitate the large scale (watershed or basin scale) application of hydrologic and water quality models. The $K_{sat}$ values required by these models can be predicted using the Pedotransfer functions. First, the available soil’s data are identified for each soil series in the watershed. According to the data availability, the best performing PTF for the textural class of this soil
series can be selected from Table (4.14). The expected (mean) value of $K_{\text{sat}}$ is then calculated using the selected PTF. The GSDER values listed in Tables (4.9) to (4.12) can be used to establish the minimum and maximum ranges of $K_{\text{sat}}$, which will be required for model auto-calibration, sensitivity and uncertainty analyses. For large scale applications, the outlined procedure is expected to be automated and integrated into the hydrologic models.

Abdelbaki and Youssef (2010) assessed the feasibility of applying the widely used drainage water management model, DRAINMOD (Skaggs, 1980), using soil hydraulic properties estimated by PTFs. They used in the analysis data from four U.S. agricultural drained sites, having different soils, crops, drainage systems, and climatological conditions, for which the model has been previously calibrated and validated (Wang et al., 2006; Youssef et al., 2006; Ale et al., 2009; Thorp et al., 2009). Following the procedure outlined in this article, $K_{\text{sat}}$ values were estimated for each of the four sites using the best performing PTFs for the textural class of the soil on each site. Measured annual drainage was compared to predicted drainage using estimated and calibrated $K_{\text{sat}}$ values. For the four sites, the normalized root mean square errors (NRMSEs) in yearly drainage predicted by the calibrated model were in the range of 9 to 24% with corresponding Nash-Sutcliffe modeling efficiencies (EF) ranging from 0.62 to 0.91. The use of PTF estimated-$K_{\text{sat}}$ moderately increased the NRMSE to 19-30% and decreased the EF to 0.22-0.85. The relatively small errors in DRAINMOD predictions induced by using PTFs-estimated $K_{\text{sat}}$ support the use of this method to estimate $K_{\text{sat}}$ for large scale applications of hydrologic models such as DRAINMOD. Despite the results of the study by Abdelbaki and Youssef (2010), it should be emphasized that using PTFs to predict $K_{\text{sat}}$ has limitations and should cautiously be used only when obtaining such an important soil hydraulic property is practically infeasible. Despite their limitations, PTFs can be very useful for predicting ranges of $K_{\text{sat}}$ for stochastic modeling.
Table (4.13) Best PTFs to estimate $K_{sat}$ according to the available input parameters.

<table>
<thead>
<tr>
<th>Texture class</th>
<th>Available input parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effective porosity</td>
</tr>
<tr>
<td>Very Fine &amp; Fine</td>
<td>F14-F10-F15</td>
</tr>
<tr>
<td>Medium Fine</td>
<td>F10-F14</td>
</tr>
<tr>
<td>Medium</td>
<td>F3-F2</td>
</tr>
<tr>
<td>Coarse</td>
<td>F7-F9-F10</td>
</tr>
<tr>
<td>Total Dataset</td>
<td>F10</td>
</tr>
</tbody>
</table>

4.9 Summary and Conclusions

The goal of this part of the study was to evaluate twenty four Pedotransfer functions for predicting saturated hydraulic conductivity for U.S. soils. The functions were divided into three groups according to their input requirements. Six functions of group 1 predict $K_{sat}$ in terms of the effective porosity. Nine functions of group 2 require particle size distribution, bulk density, and total porosity as inputs. The input requirements of the nine functions of group 3 include particle size distribution, bulk density, and organic matter content. The U.S. soils data used in evaluating the PTFs were mostly obtained from the SOILVISION soil data base. Three data sets were formed to meet the input requirements of the three groups of PTFs. The PTFs of group 1, which have the least input requirements, were evaluated using the largest data set (1911 records). The PTFs of group 2 were evaluated using a data set that includes 956 records. The PTFs of group three were evaluated using the smallest data set (678 records) because of the lack of organic matter content data in most soil data bases. For testing PTFs of groups 2 and 3, the soil data sets were further divided according to the USDA texture classification into four classes and PTFs were evaluated and ranked according to
their performance in predicting $K_{sat}$ for the entire soil’s data set and for each soil textural class.

The results showed that the PTFs developed by Suleiman and Ritchie (2000) and Minasny and McBratney (2000) are the best models to estimate $K_{sat}$ when the effective porosity is the only available soil’s data (group 1). Group 2 functions tend to over-estimate $K_{sat}$ for the coarse textural class and under-estimate $K_{sat}$ for the other textural classes. The PTF developed by Cosby et al. (1984) has the highest overall performance among the PTFs of group 2. The Rosetta SSC (Schaap et al., 2001) model showed the best performance for the very fine-fine textural class. The PTFs of Cosby et al. (1984) and Rosetta SSC are the best models for predicting $K_{sat}$ for the fine-medium textural class. For medium textural class, the PTF of Cosby et al. (1984) has the best performance. The PTFs of Puckett at al. (1984), Julia et al. (2004) and Cosby et al. (1984) are the best models for predicting $K_{sat}$ for the coarse texture soils.

Overall, group 3 PTFs tend to under-predict $K_{sat}$. The PTF developed by Nemes et al. (2005) is the best model for predicting $K_{sat}$ among the PTFs of group 3. The PTF developed by Wösten et al. (2001) had the most accurate predictions for the very fine-fine and fine-medium textural classes. The PTF of Saxton and Rawls (2006), closely followed by the PTFs of Nemes et al. (2005) and Julia et al. (2004), were the most accurate functions for predicting $K_{sat}$ for the medium textural class. The PTFs of Nemes et al. (2005) and Wosten et al. (1999) were the highest performing functions for the coarse textural class. If the availability of soil inputs are not a limiting factor, the five PTFs that have the highest overall performance, ranked from the most to least accurate, are those developed by Cosby et al. (1984), Nemes et al. (2005), Saxton et al. (1986), Saxton and Rawls (2006), and Julia et al. (2004), respectively. Three of these five functions have recently been developed
The results of this part of the study indicate that PTFs can predict $K_{sat}$ with acceptable levels of accuracy. The general performance of the PTFs is considerably higher for the coarse textural class, compared with the very fine-fine textural class. The models developed using data for U.S. soils performed better than the other models.
CHAPTER (5)

EVALUATION OF PEDOTRANSFER FUNCTIONS FOR PREDICTING THE SOIL WATER CHARACTERISTIC CURVE FOR U.S. SOILS

5.1 Introduction

Soil hydraulic properties are key factors that regulate the movement of groundwater and transport of solutes. These properties are important inputs to hydrologic and water quality models. One of the most important hydraulic properties of soils is the relationship between soil water content and soil matric potential, commonly referred to as soil water characteristic curve (SWCC) or soil water retention curve (van Genuchten, 1980). However, SWCC is not a readily available soil property primarily because of the cost and time of measurement especially with large scale (watershed and basin scale) applications. Instead of the direct measurement of SWCC, which is practically impossible, pedotransfer functions (PTFs) (Bouma, 1989) have been developed to indirectly predict the SWCC from more readily available soil properties such as, particle size distribution, organic matter, and bulk density. These soil physical properties are widely available in several soil databases (Leij et al., 1996; Fredlund et al., 1996, 1998; Batjes, 2008). Several approaches have been used to develop PTFs including multi-linear regression (MLR) (e.g. Rawls et al., 1982; Vereecken et al., 1989; Wösten et al., 1999; Saxton and Rawls, 2006), artificial neural network (ANN) (e.g. Schaap and Leij, 1998; Merdun et al., 2005; Baker and Ellison, 2008) and group method of data handling (GMDH) (e.g. Ungaro et al., 2005; Nemes et al., 2005). Two main groups of PTFs for predicting SWCC have been developed: discrete or point PTFs and continuous PTFs. The point PTFs predict the soil water content at specific matric potentials (e.g. Gupta and Larson, 1979; Rawls et al.,
Continuous PTFs are closed-form equations that predict a continuous relationship between soil water content and matric potential (e.g. Brooks and Corey, 1964; Campbell, 1974; Van Genuchten, 1980; Saxton et al., 1986; Hutson and Cass, 1987; Saxton and Rawls, 2006). Regardless of the approach used in developing the PTFs, the majority of them have been developed using a relatively small size datasets and are expected to perform differently, especially when applied to soil conditions that are significantly different from those under which the functions were developed.

Several studies have been conducted in the past two decades to compare the performance of PTFs. Tietje and Tapkenhinrichs (1993) tested thirteen PTFs using a data set of 1116 soil water content-matric potential data points measured for German soils. The results of their study are presented in Table (5.1). They concluded that the PTFs had poor performance in predicting the parameters of the closed form equation developed by Van Genuchten (1980), the PTFs developed by Rawls and Brakensiek (1985) showed comparable results to those developed by Vereecken et al. (1989), and the PTFs developed by Saxton et al. (1986) and Cosby et al. (1984) showed poor performances, especially for soils with large organic matter content. Another study conducted by Cornelis et al. (2001), they compared nine PTFs to estimate SWCC (Figures 5.1, 5.2) using 69 soil water content-matric potential data points covering a wide range of textures within Flanders (Belgium). They concluded that the PTF developed by Vereecken et al. (1989), which developed using a dataset from the same region (Belgium) as the most accurate, whereas the PTF developed by Rawls and Brakensiek (1985) showed poorer results.
Table (5.1) Results of evaluation of thirteen PTFs for predicting soil water content by Tietje and Tapkenhinrichs (1993)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Applicable to</th>
<th>MD (cm$^3$/cm$^3$)</th>
<th>RMSD (cm$^3$/cm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Husz (1967)</td>
<td>68</td>
<td>-0.0601</td>
<td>0.071</td>
</tr>
<tr>
<td>Renger (1971)</td>
<td>100</td>
<td>0.0129</td>
<td>0.0575</td>
</tr>
<tr>
<td>Gupta and Larson (1979)</td>
<td>88</td>
<td>0.0244</td>
<td>0.0591</td>
</tr>
<tr>
<td>Arya and Paris (1981)</td>
<td>100</td>
<td>0.00095</td>
<td>0.0611</td>
</tr>
<tr>
<td>Rawls et al. (1982) (A)</td>
<td>68</td>
<td>-0.006</td>
<td>0.0631</td>
</tr>
<tr>
<td>Rawls et al. (1982) (B)</td>
<td>79</td>
<td>0.0164</td>
<td>0.0551</td>
</tr>
<tr>
<td>Rawls et al. (1982) (C)</td>
<td>74</td>
<td>0.0039</td>
<td>0.0366</td>
</tr>
<tr>
<td>Cosby et al. (1984)</td>
<td>100</td>
<td>-0.0019</td>
<td>0.0668</td>
</tr>
<tr>
<td>Puckett et al. (1985)</td>
<td>48</td>
<td>-0.0334</td>
<td>0.0748</td>
</tr>
<tr>
<td>Rawls and Brakensiek</td>
<td>100</td>
<td>-0.0527</td>
<td>0.0751</td>
</tr>
<tr>
<td>Saxton et al. (1986)</td>
<td>98</td>
<td>-0.0149</td>
<td>0.0796</td>
</tr>
<tr>
<td>Vereecken et al. (1989)</td>
<td>100</td>
<td>-0.0145</td>
<td>0.0531</td>
</tr>
<tr>
<td>Tyler and Wheatcraft</td>
<td>98</td>
<td>-0.0236</td>
<td>0.0768</td>
</tr>
</tbody>
</table>

MD, mean difference; RMSD, root mean square difference
Figure 5.1 Predicted vs. measured soil moisture contents at matric potentials $\psi$ of -0.3, -3, and -10 kPa for nine PTFs were evaluated by Cornelis et al. (2001)
Figure (5.2) Predicted vs. measured soil moisture contents at matric potentials $h$ of $-31$, $-98$, and $-1500$ kPa for nine PTFs were evaluated by Cornelis et al. (2001)
Recently, Schaap et al. (2004) used a wide international dataset of 47,435 records containing 113,970 observed water contents to test eleven PTFs. They found that all PTFs underestimate the soil water content and the PTF developed by Schaap et al. (1998) that requires inputs of sand, silt and clay contents, bulk density, and water contents at 33, 1500 kPa was the most accurate model with the minimum root mean square error (RMSE) value.

<table>
<thead>
<tr>
<th>Pressure head (cm)</th>
<th>60</th>
<th>100</th>
<th>330</th>
<th>1000</th>
<th>2000</th>
<th>15000</th>
<th>Weighted Error</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of points</td>
<td>1288</td>
<td>7309</td>
<td>47162</td>
<td>352</td>
<td>10429</td>
<td>47435</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosetta model (1)</td>
<td>0.071</td>
<td>0.081</td>
<td>0.071</td>
<td>0.056</td>
<td>0.061</td>
<td>0.049</td>
<td>0.0625</td>
<td>0.069</td>
</tr>
<tr>
<td>Rosetta model (2)</td>
<td>0.069</td>
<td>0.078</td>
<td>0.067</td>
<td>0.05</td>
<td>0.056</td>
<td>0.045</td>
<td>0.0582</td>
<td>0.064</td>
</tr>
<tr>
<td>Rosetta model (3)</td>
<td>0.07</td>
<td>0.079</td>
<td>0.061</td>
<td>0.052</td>
<td>0.062</td>
<td>0.048</td>
<td>0.0574</td>
<td>0.064</td>
</tr>
<tr>
<td>Rosetta model (4)</td>
<td>0.05</td>
<td>0.049</td>
<td>0.022</td>
<td>0.056</td>
<td>0.06</td>
<td>0.045</td>
<td>0.0396</td>
<td>0.047</td>
</tr>
<tr>
<td>Rosetta model (5)</td>
<td>0.052</td>
<td>0.052</td>
<td>0.021</td>
<td>0.05</td>
<td>0.047</td>
<td>0.025</td>
<td>0.0292</td>
<td>0.032</td>
</tr>
<tr>
<td>Rawls and Brakensiek (1985) (1)</td>
<td>0.074</td>
<td>0.082</td>
<td>0.067</td>
<td>0.038</td>
<td>0.049</td>
<td>0.04</td>
<td>0.0571</td>
<td>0.063</td>
</tr>
<tr>
<td>Rawls and Brakensiek (1985) (2)</td>
<td>0.069</td>
<td>0.08</td>
<td>0.066</td>
<td>0.038</td>
<td>0.05</td>
<td>0.04</td>
<td>0.0562</td>
<td>0.064</td>
</tr>
<tr>
<td>Cosby et al. (1984) model (1)</td>
<td>0.079</td>
<td>0.083</td>
<td>0.071</td>
<td>0.034</td>
<td>0.049</td>
<td>0.043</td>
<td>0.0599</td>
<td>0.062</td>
</tr>
<tr>
<td>Cosby et al. (1984) model (2)</td>
<td>0.075</td>
<td>0.083</td>
<td>0.072</td>
<td>0.038</td>
<td>0.052</td>
<td>0.046</td>
<td>0.0616</td>
<td>0.063</td>
</tr>
<tr>
<td>Vereecken et al. (1989)</td>
<td>0.07</td>
<td>0.087</td>
<td>0.073</td>
<td>0.033</td>
<td>0.058</td>
<td>0.049</td>
<td>0.0636</td>
<td>0.066</td>
</tr>
<tr>
<td>Wosten et al. (1999)</td>
<td>0.07</td>
<td>0.08</td>
<td>0.061</td>
<td>0.037</td>
<td>0.052</td>
<td>0.044</td>
<td>0.0551</td>
<td>0.058</td>
</tr>
</tbody>
</table>

URMSE, unbiased root mean square error; RMSE, root mean square error

Over the last two decades, a considerable number of PTFs have been proposed for predicting SWCC to be a useful tool to provide the hydrological models with the required input parameters (e.g. Cosby et al., 1984; Vereecken et al., 1989; Schaap et al., 1998; Wösten et al., 1999; Rajkai et al., 2004; Saxton and Rawls, 2006; Li et al., 2007; Rubio, 2008; Adhikary et al., 2008). These functions have been developed under different conditions and using many different datasets either in volume or in its source (Table 5.2).
Studies showed that the accuracy of predictions of PTFs depends on the soil texture class at which the water content is predicted (Tietje and Tapkenhinrichs, 1993; Cornelis et al., 2001; Thakur et al., 2007). Other studies showed that the accuracy of PTFs depends on the pressure head at which the water content is predicted (Kern, 1995; Ungaro and Calzolari, 2001; Givi et al., 2004; Schaap et al., 2004; Nemes and Rawls, 2006).

The aim of this study is to evaluate the accuracy of 22 PTFs for predicting SWCC for United States soils, where climatic and geological conditions are different from those used for developing the PTFs. Also, the study aims to identify the best performing PTFs corresponding to different textural classes and matric potential heads to be a good reference for further model applications.

5.2 Pedotransfer Functions of Soil Water Content

5.2.1 Point Pedotransfer Functions

This type of PTFs predicts the soil water content at specific matric potentials as discrete points. Five point PTFs are evaluated in this study. Gupta and Larson (1979) used 43 different soil materials originating from 10 locations in eastern and central USA to develop 12 regression equations that estimate soil water content at matric potentials ranging from 4 to 1500 kPa. The PTFs developed by Gupta and Larson, 1979 (Eq. 5.1) require particle size distribution, organic matter content, and bulk density data as inputs.
Table (5.3) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Gupta and Larson (1979)

<table>
<thead>
<tr>
<th>Tension (kPa)</th>
<th>Regression coefficients</th>
<th>Correlation coefficient (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a x 10^3</td>
<td>b x 10^3</td>
</tr>
<tr>
<td>4</td>
<td>7.053</td>
<td>10.242</td>
</tr>
<tr>
<td>7</td>
<td>5.678</td>
<td>9.228</td>
</tr>
<tr>
<td>10</td>
<td>5.018</td>
<td>8.548</td>
</tr>
<tr>
<td>20</td>
<td>3.890</td>
<td>7.066</td>
</tr>
<tr>
<td>33</td>
<td>3.075</td>
<td>5.886</td>
</tr>
<tr>
<td>60</td>
<td>2.181</td>
<td>4.557</td>
</tr>
<tr>
<td>100</td>
<td>1.563</td>
<td>3.620</td>
</tr>
<tr>
<td>200</td>
<td>0.932</td>
<td>2.643</td>
</tr>
<tr>
<td>400</td>
<td>0.483</td>
<td>1.943</td>
</tr>
<tr>
<td>700</td>
<td>0.214</td>
<td>1.538</td>
</tr>
<tr>
<td>1000</td>
<td>0.076</td>
<td>1.334</td>
</tr>
<tr>
<td>1500</td>
<td>-0.059</td>
<td>1.142</td>
</tr>
</tbody>
</table>

\[ \theta = a \cdot \text{SA} + b \cdot \text{SI} + c \cdot \text{CL} + d \cdot \text{OM} + e \cdot \text{BD} \quad (5.1) \]

where \( \theta \) is the volumetric soil water content (m^3 m^-3), a, b, c, d and e are regression coefficients, SA, SI, CL are the sand, silt, and clay contents(%) respectively, OM is the organic matter content(%), and BD is the bulk density (Mg m^-3). The regression coefficients of PTFs in Eqn. (5.1) are presented in Table (5.3) for 12 pressure heads ranging from 4 kPa to 1500 kPa.

Rawls et al. (1982) developed 10 regression equations using data originating from 2543 horizons from across the USA to estimate volumetric water content at matric potentials ranging from 10 to 1500 kPa. The PTFs of Rawls et al., 1982 (Eq. 5.2) require only particle size distribution and organic matter content.
The regression coefficients of Rawls et al. (1982) PTFs are presented in Table (5.4) for 10 pressure heads ranging from 10 kPa to 1500 kPa.

Table (5.4) regression coefficients of point pedotransfer functions for predicting soil water characteristic curve developed by Rawls et al. (1982)

<table>
<thead>
<tr>
<th>Tension (K Pa)</th>
<th>Regression coefficients</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>10</td>
<td>0.4118</td>
<td>-0.0030</td>
</tr>
<tr>
<td>20</td>
<td>0.3121</td>
<td>-0.0024</td>
</tr>
<tr>
<td>33</td>
<td>0.2576</td>
<td>-0.0020</td>
</tr>
<tr>
<td>60</td>
<td>0.2065</td>
<td>-0.0016</td>
</tr>
<tr>
<td>100</td>
<td>0.0349</td>
<td>--</td>
</tr>
<tr>
<td>200</td>
<td>0.0281</td>
<td>--</td>
</tr>
<tr>
<td>400</td>
<td>0.0238</td>
<td>--</td>
</tr>
<tr>
<td>700</td>
<td>0.0216</td>
<td>--</td>
</tr>
<tr>
<td>1000</td>
<td>0.0205</td>
<td>--</td>
</tr>
<tr>
<td>1500</td>
<td>0.0260</td>
<td>--</td>
</tr>
</tbody>
</table>

Al Majou et al. (2007) developed point PTFs using a dataset of 320 soil horizons from France. Their dataset comprising 90 top soils and 230 subsoil horizons. They established the PTFs according to the texture (textural class PTFs) in the Commission of the European Communities (CEC) triangle and then according to both the texture and bulk density (texturo-structural class PTFs). The point PTFs developed by Al Majou et al., 2007 (Eqn. 5.3) predict the soil water content at seven matric potential ranging from 1 to 1500 kPa using inputs of particle size distribution, bulk density, and organic matter content.

$$\theta = a + b \cdot \text{SA} + c \cdot \text{SI} + d \cdot \text{CL} + e \cdot \text{OM} \quad (5.2)$$

$$\theta = a + b \cdot \text{CL} + c \cdot \text{SI} + d \cdot \text{OM} + e \cdot \text{BD} \quad (5.3)$$
The regression coefficients of Al Majou et al. (2007) PTFs are presented in Table (5.5) for 7 pressure heads ranging from 1 kPa to 1500 kPa.

<table>
<thead>
<tr>
<th>Tension (kPa)</th>
<th>Regression coefficients</th>
<th>Correlation coefficient (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>1</td>
<td>0.4701</td>
<td>0.0026</td>
</tr>
<tr>
<td>3.3</td>
<td>0.3556</td>
<td>0.0029</td>
</tr>
<tr>
<td>10</td>
<td>0.2620</td>
<td>0.0034</td>
</tr>
<tr>
<td>33</td>
<td>0.1301</td>
<td>0.0038</td>
</tr>
<tr>
<td>100</td>
<td>0.0184</td>
<td>0.0045</td>
</tr>
<tr>
<td>330</td>
<td>-0.0504</td>
<td>0.0047</td>
</tr>
<tr>
<td>1500</td>
<td>-0.0786</td>
<td>0.0045</td>
</tr>
</tbody>
</table>

Adhikary et al. (2008) developed point PTFs for predicting SWCC using 800 measured soil water content-matric potential data points of Indian soils. They developed seven regression equations to predict soil water content at seven matric potentials ranging from 10 to 1500 kPa. The PTFs developed by Adhikary et al., 2008 (Eqn. 5.4) require only particle size distribution as an input.

\[ \theta = (a + b \cdot SA + c \cdot SI + d \cdot CL)/100 \quad (5.4) \]

The regression coefficients of Adhikary et al. (2008) PTFs are presented in Table (5.6) for 7 pressure heads ranging from 10 kPa to 1500 kPa.
Using 185 measured SWCC of Iranian soils, Homae and Firouzi (2008) developed two types of continuous and point PTFs. Some of the developed PTFs require the gypsum content as an input while the other PTFs require only the particle size distribution. Only, the PTFs that require available inputs (Eqn. 5.5) were evaluated in this study. The Point PTFs developed by Homae and Firouzi (2008) predict the soil water content at six matric potentials ranging from 10 to 1500 kPa.

\[
\theta = a + b \cdot SA + c \cdot BD + d \cdot \log(CL) \quad (5.5)
\]

The regression coefficients of Homae and Firouzi (2008) PTFs are presented in Table (5.7) for 6 pressure heads ranging from 10 kPa to 1500 kPa.
5.2.2 Continuous Pedotransfer Functions

Continuous PTFs for predicting soil water content are closed form equations to simulate the relationship between SWC and matric potential. Continuous PTFs are more reliable to be applied in hydrological models than point PTFs due to its continuous results. Several equations have been suggested to simulate the soil water characteristic curve. Brooks and Corey (1964) suggested Eqn. (5.6) to simulate the relationship between soil water content ($\theta$) vs. matric potential head ($h$).

$$\frac{\theta - \theta_r}{\theta_s - \theta_r} = \left[\frac{H_b}{h}\right]^\lambda \quad (5.6)$$

Where $\theta$ is the volumetric soil water content ($m^3 m^{-3}$), $\theta_s$ is the saturated soil water content ($m^3 m^{-3}$) (water content at zero matric potential), $\theta_r$ is the residual soil water content ($m^3 m^{-3}$), $H_b$ is the bubbling pressure kPa, $\lambda$ is the pore size distribution index (dimensionless). The parameters of Brooks-Corey equation were determined by Rawls and Brakensiek (1985). Their regression equations were formulated using 2543 measured soil water content-matric potential data points for American and international soils. The PTFs developed by Rawls and Brakensiek (1985) require the soil porosity, clay and sand contents as inputs.

$$\ln(H_b) = 5.34 + 0.185 \times CL - 2.484 \times \varphi - 0.002 \times CL^2 - 0.044 \times SA \times \varphi$$
$$- 0.617 \times CL \times \varphi + 0.001 \times SA^2 \times \varphi^2 - 0.009 \times CL^2 \times \varphi^2$$
$$- 0.00001 \times SA^2 \times CL + 0.009 \times CL^2 \times \varphi - 0.0007 \times SA^2 \times \varphi$$
$$+ 0.000001 \times CL^2 \times SA + 0.5 \times \varphi^2 \times CL$$
\[
\ln(\lambda) = -0.784 + 0.018 \times SA - 1.062 \times \varphi - 0.00005 \times SA^2 - 0.003 \times CL^2 \\
+ 1.111 \times \varphi^2 - 0.031 \times SA \times \varphi + 0.0003 \times SA^2 \times \varphi^2 \\
- 0.006 \times CL^2 \times \varphi^2 - 0.000002 \times SA^2 \times CL + 0.008 \times CL^2 \times \varphi \\
- 0.007 \times \varphi^2 \times CL
\]

\[
\theta_r = -0.018 + 0.0009 \times SA + 0.005 \times CL + 0.029 \times \varphi - 0.0002 \times CL^2 \\
- 0.001 \times SA \times \varphi - 0.0002 \times CL^2 \times \varphi^2 + 0.0003 \times CL^2 \times \varphi \\
- 0.002 \times \varphi^2 \times CL
\]

Recently, Campbell (1974) modified Brooks and Corey (1964) equation by neglecting the term of residual water content to be in the form of Eqn. (5.7).

\[
\frac{\theta}{\theta_s} = \left[ \frac{H_b}{h} \right]^\lambda \quad (5.7)
\]

Using 1448 measured soil water content- matric potential data points for American soils, Cosby et al. (1984) developed PTFs for predicting the parameters of Campbell’s equation in terms of sand and clay contents.

\[
\theta_s = -0.142 \times SA + 50.5
\]

\[
H_b = 10^{-0.0095 \times SA + 1.54}
\]

\[
\lambda = -\frac{1}{0.157 \times CL + 3.1}
\]

One of the most used equations for predicting soil water characteristic curve is the equation developed by Van Genuchten, 1980 (Eqn. 5.8). This equation has better performance to predict soil water content especially at matric potentials near saturation.

\[
\frac{\theta - \theta_r}{\theta_s - \theta_r} = \left[ \frac{1}{1 + ((\alpha h)^n)} \right]^{1 - \frac{1}{n}} \quad (5.8)
\]
Vereecken et al. (1989) used 182 horizons of 40 Belgium soil series to solve for the parameters of van Genuchten equation. They developed their PTFs using the method of multiple linear regressions. The PTFs developed by Vereecken et al. (1989) require sand, clay, and organic matter contents, and bulk density as inputs.

\[
\theta_f = 0.015 + 0.005(\text{CL}) + 0.014(\text{OM})
\]

\[
\theta_s = 0.81 - 0.283(\text{BD}) + 0.001(\text{CL})
\]

\[
\alpha = \exp\left(-2.486 + 0.025(\text{SA}) - 0.351(\text{OM}) - 0.2671(\text{BD}) - 0.023(\text{CL})\right)
\]

\[
n = 2.781^{[0.053-0.009(\text{SA})-0.013(\text{CL})+0.00015(\text{SA})^2]}
\]

Wösten et al. (1999) used multi linear regression to develop class and continuous PTFs to predict the parameters of Van Genuchten equation. They used dataset from 4030 horizons from all over Europe. The PTFs developed by Wösten et al. (1999) require sand, silt, clay and organic matter contents and bulk density as inputs.

\[
\theta_s = 0.792 + 0.0017 \times \text{CL} - 0.2962 \times \text{BD} - 0.0000015 \times \text{SI}^2 + 0.000082 \\
\times \text{OM}^2 + 0.0243 \times \text{CL}^{-1} + 0.01113 \times \text{SI}^{-1} + 0.0147 \times \ln(\text{SI}) \\
- 0.00007 \times \text{OM} \times \text{CL} - 0.00062 \times \text{BD} \times \text{CL} - 0.00118 \times \text{BD} \\
\times \text{OM} - 0.000166 \times \text{SI}
\]

\[
\ln(\alpha) = -14.9 + 0.0313 \times \text{CL} + 0.035 \times \text{SI} + 0.646 \times \text{OM} + 15.29 \times \text{BD} - \\
0.192 - 4.671 \times \text{BD}^2 - 0.00078 \times \text{CL}^2 - 0.007 \times \text{OM}^2 + 0.045 \times \text{OM}^{-1} + \\
0.066 \times \ln(\text{SI}) + 0.148 \times \ln(\text{OM}) - 0.0454 \times \text{BD} \times \text{SI} - 0.485 \times \text{BD} \times \text{OM} + \\
0.0067 \times \text{CL}
\]
Rubio (2008) developed two site-specific PTFs to predict the parameters of van Genuchten equation for two vegetation covers. The first function for forest land and requires inputs of sand content and bulk density. The second function for grass land and requires inputs of organic matter content and bulk density.

Rubio (2008) (1)

\[ \theta_r = 0.148 - 0.112 \times BD \]
\[ \theta_s = 0.857 - 0.247 \times BD \]
\[ \alpha = 0.062 + 0.018 \times SA - 0.009 \times SA^2 + 0.00021 \times SA^3 \]
\[ n = 1.229 - 0.081 \times BD \]

Rubio (2008) (2)

\[ \theta_r = 0.158 - 0.035 \times OM + 0.002 \times OM^2 \]
\[ \theta_s = 1.102 - 0.432 \times BD \]
\[ \alpha = 0.723 - 0.444 \times BD \]
\[ n = 1.304 - 0.113 \times BD \]

Schaap et al. (2001) developed artificial neural network model with graphical user interface called (Rosetta) for predicting the parameters of Van Genuchten equation. Rosetta can be used for predicting saturated hydraulic conductivity (\(K_{sat}\)) and Van Genuchten equation’s parameters (\(\theta_r, \theta_s, \alpha, n\)). The model has five different levels of input requirements. In this study, the models Rosetta-SSC (needs inputs of %sand, %silt, and %clay) and Rosetta SSC-
BD (needs inputs of %sand, %silt, %clay, and bulk density) were evaluated due to availability of their inputs.

The Eqn. proposed by Van-Genuchten (1980) can be presented in another form by removing the term of residual soil water content ($\theta_r$) to produce the following equation (Eqn. 5.9).

\[
\theta = \frac{\theta_s}{[1+(\alpha h)^n]^{(1-\frac{1}{n})}} \quad (5.9)
\]

Many studies have been conducted to develop PTFs for predicting the parameters of Eqn. 9. Rajkai et al. (2004) used 305 SWCC from the Hungarian soil database to develop four sets of PTFs explaining the three parameters of Van Genuchten equation. Due to the Lacking of input requirements, two of these sets were evaluated in this study. The PTFs developed by Rajkai et al. (2004) require sand, silt, clay and organic matter contents and bulk density as inputs.

\[
\theta_s = 123.76 - 65.37 \times BD - 0.28 \times OM - 0.000048 \times CL^2 - 1.99 \times \ln(CL)
\]
\[
+ 12.46 \times BD^2 - 0.054 \times BD \times SA + 0.14 \times \frac{SA}{SI} + 0.00049 \times BD^2 \times CL^2
\]

\[
\ln(\alpha) = 16.97 + 0.12 \times BD \times SI + 0.22 \times CL - 9.34 \times \ln(SI) - 0.039 \times BD^2 \times CL + 0.21 \times SI - 0.0029 \times CL^2 - 0.435 \times \frac{SA}{SI} - 0.00093 \times BD^2 \times SI^2
\]
Using a dataset of 320 horizons (90 top soils and 230 sub soils) from France, Al Majou et al. (2007) developed continuous PTFs to predict the three parameters of Van Genuchten equation. Their PTFs require inputs of sand, clay, and organic matter contents, and bulk density.

\[
\begin{align*}
\ln(n) &= -0.069 + 0.32 \times BD - 0.007 \times OM - 0.000009 \times CL^2 + 0.00147 \times \\
\ln(CL) &= 0.00011 \times BD \times SA - 0.0064 \times BD^2 \times CL + 0.0015 \times \frac{SA}{SI} + \\
&\quad 0.000081 \times BD^2 \times CL^2
\end{align*}
\]

Using a dataset of 320 horizons (90 top soils and 230 sub soils) from France, Al Majou et al. (2007) developed continuous PTFs to predict the three parameters of Van Genuchten equation. Their PTFs require inputs of sand, clay, and organic matter contents, and bulk density.

\[
\theta_s = 1.1658 - 0.0032 \times CL - 0.4737 \times BD + 2 \times 10^{-7} \times SI^2 - 0.0001 \times OM^2 \\
+ 0.0373 \times CL^{-1} + 0.0131 \times SI^{-1} - 0.0072 \times \ln SI + 0.00003 \\
\times OM \times CL + 0.0022 \times BD \times CL - 0.0002 \times BD \times OM - 0.0001 \\
\times SI
\]

\[
\ln(a) = 25.61 + 0.0439 \times CL + 0.1129 \times SI + 1.1914 \times OM + 32.21 \times BD \\
- 10.48 \times BD^2 - 0.0009 \times CL^2 - 0.0146 \times OM^2 - 0.3781 \\
\times OM^{-1} - 0.0178 \times \ln SI - 0.1032 \times \ln OM - 0.1 \times BD \times SI \\
- 0.6001 \times BD \times OM
\]

\[
\ln(n) = -15.29 - 0.0659 \times CL + 0.0115 \times SI - 0.2115 \times OM + 12.33 \times BD \\
- 1.3578 \times BD^2 + 0.0006 \times CL^2 + 0.0031 \times OM^2 + 4.0005 \\
\times BD^{-1} + 2.2003 \times SI^{-1} + 0.1643 \times OM^{-1} - 0.1205 \times \ln SI \\
+ 0.2693 \times \ln OM - 9.9367 \times \ln BD + 0.003 \times BD \times CL + 0.0694 \\
\times BD \times OM
\]

111
Zacharias and Wessolek (2007) used dataset of 676 measured SWCC (353 top soils, 323 sub soils) to develop continuous PTFs for predicting the three parameters of Van Genuchten Eqn. Their dataset was selected from two large international databases: the IGBP-DIS soil database for pedotransfer function development and the Unsaturated Soil Hydraulic Database (UNSODA). Zacharias and Wessolek (2007) divided the dataset into two ranges according to the sand content and developed two sets of PTFs; one for sand content less than 66.5% and the other set for sand content greater than or equal 66.5%. These PTFs require inputs of sand and clay contents, and bulk density.

Zacharias and Wessolek (2007) for sand content < 66.5%

\[
\theta_s = 0.890 - 0.001 \times CL - 0.322 \times BD \\
\ln(\alpha) = -0.648 + 0.023 \times SA + 0.044 \times CL - 3.168 \times BD \\
n = 1.392 - 0.418 \times SA^{-0.024} + 1.212 \times CL^{-0.704}
\]

For sand content \(\geq\) 66.5%

\[
\theta_s = 0.788 + 0.001 \times CL - 0.263 \times BD \\
\ln(\alpha) = -4.197 + 0.013 \times SA + 0.076 \times CL - 0.276 \times BD \\
n = -2.562 + 7 \times 10^{-9} \times SA^{4.004} + 3.750 \times CL^{-0.016}
\]

Using 63 measured SWCC from China, Li et al. (2007) used multiple regressions to develop PTFs for predicting the three parameters of Van-Genuchten equation. They used these functions in the regional simulation of a wheat and maize cropping agro ecosystem for the 1998–1999 rotation year. The PTFs developed by Li et al. (2007) require inputs of sand, silt, clay, and organic matter contents, and bulk density.

\[
\ln(\theta_s) = -1.531 + 0.212 \times \ln(SA) + 0.006 \times SI - 0.051 \times OM - 0.566 \times \ln(BD)
\]
\[
\ln(\alpha) = -67.408 - 0.04 \times SI - 0.670 \times \ln(SI) - 2.189 \times OM + 1.41 \times \ln(OM) \\
+ 78.4 \times BD - 121.33 \times \ln(BD)
\]

\[
n = 1.488 + 0.002 \times \ln(SI) + 0.013 \times CL - 0.248 \times \ln(CL) + 0.048 \times \ln(OM) \\
+ 0.451 \times \ln(BD)
\]

Saxton et al. (1986) simplify the Brooks-Corey (1964) model by dividing the SWCC into three ranges (Eqn. 5.10): from 0 to \(h_b\), from \(h_b\) to 10 kPa, and from 10 to 1500 kPa. They developed PTFs to predict the coefficients A and B using 44 measured SWCC. Their equations require inputs of sand and clay contents.

\[
\theta = \begin{cases} \\
\theta_s & \text{for } 0 < h < h_b \\
\theta_{10} + \frac{(10-h)(\theta_s-\theta_{10})}{(10-h_b)} & \text{for } h_b \text{ to } 10 \text{ kpa} \\
\left(\frac{h}{A}\right)^{-\frac{1}{B}} & \text{for } 10 < h < 1500 \text{ kpa}
\end{cases} \quad (5.10)
\]

\[
A = \text{Exp}[-4.396 - 0.0715 \times CL - 4.88 \times 10^{-4} \times SA^2 - 4.285 \times 10^{-5} \times SA^2 \\
\times CL] \times 100
\]

\[
B = -3.140 - 2.22 \times 10^{-3} \times CL^2 - 3.484 \times 10^{-5} \times SA^2 \times CL
\]

Adhikary et al. (2008) used 600 measured SWCC from India to develop PTFs for predicting the coefficients A and B in the Eqn. of Saxton et al. (1986) (Eqn. 10).

\[
\log(A) = 4.013 - 0.018 \times SA + 0.151 \times CL + 0.001 \times SA \times CL
\]

\[
\log(-B) = 0.122 \times CL + 0.026 \times SA
\]

Recently, Saxton and Rawls (2006) modified the model of Saxton et al. (1986) by dividing the SWCC into three different ranges: from 0 to \(h_b\), from \(h_b\) to
33 kPa, and from 33 to 1500 kPa. They developed computer program called SOIL WATER CHARACTERISTICS, which predict the SWCC based on Eqn. (5.11) in terms of sand, clay, and organic matter content. This program is available online at http://hydrolab.arsusda.gov/soilwater/Index.htm.

\[
\theta = \begin{cases} 
\theta_s & \text{for } 0 < h < h_b \\
\theta_{33} + \frac{(33-h)(\theta_s-\theta_{33})}{(33-h_b)} & \text{for } h_b \text{ to } 33 \text{ kpa} \\
\left(\frac{h}{h_b}\right)^{-1} & \text{for } 33 < h < 1500 \text{ kpa}
\end{cases}
\]  

Finally, Hutson and Cass (1987) introduced Eqn. (5.12) as a modified Brooks–Corey type model (Brooks and Corey, 1964), in which water characteristics curve is described as;

\[
\theta = \begin{cases} 
\theta_s \cdot \left(\frac{h}{a}\right)^{-1} & \text{For } \theta < \theta_i \\
\theta_s + h^2 \left(1 - \frac{\theta_i}{\theta_s}\right) & \text{For } \theta \geq \theta_i
\end{cases}
\]

Where \( \theta_i = \frac{2b\theta_s}{1 + 2b} \) and \( h_i = a \left(\frac{2b}{1 + 2b}\right)^{-b} \)

\[ \log(a) = -4.9840297533 + 0.0509226283 \times SA + 0.1575152771 \times SI \\
+ 0.1240901644 \times BD - 0.1640033143 \times OM - 0.0021767278 \times SI^2 + 1.438224 \times 10^{-5} \times SI^3 + 8.040715 \times 10^{-4} \times CL^2 \\
+ 0.0044067117 \times OM^2 \]

\[ \log\left(\frac{1}{b}\right) = -0.8466880654 - 0.0046806123 \times SA + 0.0092463819 \times SI \\
- 0.4542769707 \times BD - 0.0497915563 \times OM + 3.29468 \times 10^{-4} \times SA^2 - 1.689056 \times 10^{-6} \times SA^3 + 0.0011225373 \times OM^2 \]

\[ \theta_s = 0.2345971971 + 0.0046614221 \times SA + 0.0088163314 \times SI + \\
0.0064338641 \times CL - 0.3028160229 \times BD + 1.79762 \times 10^{-5} \times SA^2 - \\
3.134631 \times 10^{-5} \times SI^2 \]

Tables (5.8a), (5.8b) summarized the basic characteristics of all PTFs evaluated in this study including its input requirements, development datasets (size and source), and its output parameters. The PTFs were selected according to its inputs availability, where any other function requires input parameters other than the available measures in the soil database (particle size distribution, bulk density, porosity, and organic matter content) were excluded.
Table (5.8) Basic characteristics of point PTFs evaluated in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>PTF</th>
<th>Input Parameters</th>
<th>Development Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Clay</td>
<td>Silt</td>
</tr>
<tr>
<td>1</td>
<td>Gupta &amp; Larson (1979)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>Rawls et al. (1982)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>Al Majou et al. (2007)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>Adhikary et al. (2008)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>Homae &amp; Firouzi (2008)</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table (5.8b). Basic characteristics of continuous PTFs evaluated in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>PTF</th>
<th>Input Requirements</th>
<th>Development Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Clay</td>
<td>Silt</td>
</tr>
<tr>
<td>7</td>
<td>Cosby et al. (1984)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>Rawls &amp; Brakensiek</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>19</td>
<td>Saxton et al. (1986)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>22</td>
<td>Mayer &amp; Jarvis (1999)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>20</td>
<td>Saxton &amp; Rawls (2006)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>21</td>
<td>Adhikary et al. (2008)</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Continuous PTFs for predicting the parameters of Brooks & Corey (1964) Eqn. or its modified forms:

<table>
<thead>
<tr>
<th>No.</th>
<th>PTF</th>
<th>Input Requirements</th>
<th>Development Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Vereecken et al.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>9</td>
<td>Wösten et al. (1999)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10</td>
<td>Rosetta SSC</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>11</td>
<td>Rosetta SSCBD</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>12</td>
<td>Rubio (2008)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>13</td>
<td>Rubio (2008)2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>14</td>
<td>Rajkai et al. (2004)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>15</td>
<td>Rajkai et al. (2004)2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>16</td>
<td>Al Majou et al. (2007)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>17</td>
<td>Zacharias &amp; Wess.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>18</td>
<td>Li et al. (2007)</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

BD, bulk density; OM, organic matter content; Rosetta SSC, artificial neural network needs inputs of % sand, %silt, %clay; Rosetta SSC-BD, artificial neural network needs inputs of % sand, %silt, %clay, bulk density; BC, Brooks & Corey (1964) parameters; CAM, Campbell (1974) parameters; VG4, four Van Genuchten (1980) parameters; VG3, three Van Genuchten (1980) parameters; SAX, Saxton (1986) parameters; HC, Hutson and Cass (1987) parameters.
5.3 Soil Database

Two large international soil databases were used for evaluation of PTFs; SOILVISION database, ISRIC-WISE database. The soil database (SOILVISION) developed by Fredlund et al. (1996), (1998), is an international database includes more than 6000 records. SOILVISION database has measurements of SWCC, $K_{sat}$ and other soil physical properties like particle size distribution, bulk density, soil porosity and organic matter content. This international database encompasses more than 12 soil databases including UNSODA, Walter Rawls database,…etc. SOILVISION database is available online at http://www.soilvision.com/soilvision_downloads.shtml. In SOILVISION database, soil water content is measured for 7 to 15 pressure heads for each record. A subset of 303 SWCC for United States soils contains 3246 measured soil water contents vs. pressure head were selected from the SOILVISION database. The second database used in this study is the International Soil Reference and Information Center-World Inventory of Soil Emission Potentials (ISRIC-WISE) harmonized global soil profile dataset (Version 3.1) (Batjes, 2008). This database contains 47,833 soil records collected from 149 countries all over the world. In ISRIC-WISE database, soil samples have measurements of particle size distribution, organic matter contents, bulk density and measured soil water content for at least one matric potential head of 10, 33, and 1500 kPa. A dataset of 1743 soil samples of United States soils were selected from the ISRIC-WISE database. Most of the US samples contain measured soil water content at one pressure head only. Therefore, the 1743 soil samples have a total of 1754 measured soil water contents. ISRIC-WISE database is available online at http://www.isric.org/UK/About+Soils/Soil+data/Thematic+data/Soil+Profile+Data/default.htm. The total number of samples was selected from both datasets was 2046 samples having 5000 measured soil water content vs. pressure head. Description of the databases used in this study shown in Figure (5.3) and table
The two datasets were classified according to the Food and Agriculture Organization (FAO) texture classes (very fine, fine, medium fine, medium, coarse). According to FAO classification, more than 88% of SOILVISION database and 57% of the ISRIC-WISE database are medium and coarse soils. The number of very fine class is limited (2.75% of SOILVISION database and 6.8% of the ISRIC-WISE database).

Table (5.9). Number of soil water content measurements (N) and mean soil characteristics of data sets corresponding to FAO texture classes.

<table>
<thead>
<tr>
<th>Texture class</th>
<th>SOILVISION Dataset</th>
<th>ISRIC -WISE Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Sand</td>
</tr>
<tr>
<td>Very Fine</td>
<td>89</td>
<td>19.0</td>
</tr>
<tr>
<td>Fine</td>
<td>140</td>
<td>30.2</td>
</tr>
<tr>
<td>Medium Fine</td>
<td>151</td>
<td>7.0</td>
</tr>
<tr>
<td>Medium</td>
<td>1324</td>
<td>50.4</td>
</tr>
<tr>
<td>Coarse</td>
<td>1542</td>
<td>87.1</td>
</tr>
<tr>
<td>Total</td>
<td>3246</td>
<td>64.1</td>
</tr>
</tbody>
</table>

N, number of observed samples; OM, organic matter content (%); BD, bulk density (gm/cm^3)

Figure (5.3) Distribution of sand and clay contents for the soil database used for evaluating Pedotransfer functions in this study.
5.4 Statistical Evaluation

Two approaches can be considered to compare the measured and predicted SWCCs. In the first approach, the chosen validation indices are numerically calculated based on predicted and measured water contents at certain matric potentials (Schaap et al., 2004; Givi et al., 2004; Merdun et al., 2005; Al Majou et al., 2007). In the second approach, measured and predicted functions are integrated between top and bottom boundaries for the measured and PTF-estimated SWCCs (Tietje and Tapkenhinrichs, 1993; Cornelis et al., 2001). However, the second approach requires continuous functions to be fitted to the measured and predicted θ(h) data pairs. In this study, only the numerical approach was applied for the following reasons. First, the measured soil water contents in the soil database are discrete and it is preferable to compare it with discrete predicted water content at the same matric potentials. Second, considerable number of records in the soil database contains measured soil water content at one pressure head only (10, 33, or 1500 kPa), therefore, the fitted functions between measured values will be inaccurate. Third, the continuous PTFs can predict the water content as continuous and discrete points as well. Therefore, continuous PTFs can predict the water content at all matric potential as in the measurements.

Accuracy of PTFs predictions was evaluated by calculating the coefficient of determination ($R^2$) and the root mean square error (RMSE) between the measured and predicted soil water contents (Merdun et al., 2005) according to the following equations (Eqns. 5.13, 5.14).

\[
R^2 = 1 - \frac{\sum_i^n(\theta_{m,i} - \theta_{p,i})^2}{\sum_i^n(\bar{\theta}_m - \bar{\theta}_m)^2} \quad (5.13)
\]

\[
RMSE = \sqrt{\frac{\sum_i^n(\theta_{m,i} - \theta_{p,i})^2}{N}} \quad (5.14)
\]

Where $\theta_{m,i}$ is the measured soil water content ($m^3 m^{-3}$), $\theta_{p,i}$ is the predicted soil water content ($m^3 m^{-3}$), $\bar{\theta}_m$ is the average of the measured values, and $N$ is the
total number of observations. In addition to these two criteria, mean error (ME) was used in the evaluation of prediction accuracy of PTFs (Eqn. 5.15).

\[ ME = \frac{\sum_{i=1}^{N}(\theta_{m,i} - \theta_{p,i})}{N} \] (5.15)

Negative and positive values of ME indicate under- and over-estimation of the predictive model respectively. Also, visual evaluation was made by regressing predicted against measured values of each function. The functions were ranked according to the criteria that the PTF has the best performance is the function that has minimum value of root mean square error (RMSE) and maximum value of coefficient of determination ($R^2$).

\section*{5.5 Evaluation Methods}

To compare the performance of PTFs under different conditions, the predictions of the 22 PTFs were evaluated in three different ways. First, predictions of PTFs were compared to actual measurements for the entire dataset (2046 samples that makes 5000 data pairs of soil water content vs. potential head). Second, to test the performance of PTFs with the variability of soil texture classes, the soil dataset was divided into five texture classes according to FAO classification (i.e. very fine, fine, medium fine, medium, coarse classes) then PTFs were evaluated using the soil samples of each texture class. Finally, to test the performance of PTFs with the variability of matric potential heads, PTFs were evaluated at three certain pressure heads (10, 33, 1500 kPa). These pressure heads were selected because all PTFs (point and continuous) can predict the water content at these points. Also, predicted water content at these points are needed in many hydrological models that require inputs of permanent wilting point ($\theta_{1500}$), field capacity ($\theta_{10}$, $\theta_{33}$) and available water capacity (AWC).
5.6 Results and Discussions

The predicted soil water content was compared with the measurements. For point PTFs, the functions predict the water content directly. Therefore, the good performance of PTFs refers to the performance of the PTF in predicting water content. In case of continuous PTFs, the functions predict the parameters of closed form equation. Therefore, the good performance refers to the performance of the closed form equation in predicting the water content.

5.6.1 Overall Evaluation of PTFs

All functions (point and continuous PTFs) were evaluated using the entire dataset. The results are summarized in Table (5.10), Figures (5.4, 5.5). Many PTFs showed similar results with a little difference in RMSE up to 0.01$\text{m}^3\text{m}^{-3}$ specially for continuous PTFs. Due to little difference in PTFs predictions, the highest two results of point PTFs and the highest four results of continuous PTFs were taken into consideration (see table 3). The results showed that the PTFs developed by Rawls et al.(1982) and Adhikary et al. (2008) showed the best performances in the point PTFs. Also, the models developed by Cosby et al. (1984), Rawls and Brakensiek (1985), Schaap et al. (1998) (Rosetta SSC) and Mayer and Jarvis (1999) showed the highest performance in continuous PTFs. For the rest of the functions, some PTFs showed good performance (e.g. Al Majou et al., 2007; Rosetta SSC-BD; Zacharias and Wessolek, 2007; Rubio, 2008) while the other functions showed poorer performance. Comparing the results of the continuous and point PTFs, almost the two groups showed similar results. The similarity between point and continuous PTFs results return to the structure of the dataset, which has more than 70% of the measured water content that can be predicted by the point PTFs. Also, the results showed that the PTFs that developed from US soils, performed better than others (e.g. Rawls et al., 1982; Cosby et al., 1984; Rawls and Brakensiek, 1985). Some PTFs developed from non US soils
(e.g. Adhikary et al., 2008; Al Majou et al., 2007; Wösten et al., 1999; Mayer and Jarvis, 1999) showed comparable results. In addition, the volume of development dataset effects the PTFs performance; PTFs that developed from large dataset showed better performance (e.g. Rawls et al., 1982; Wösten et al., 1999; Schaap et al. 2001 (Rosetta SSC, Rosetta SSC-BD); Rawls and Brakensiek, 1985) While the poorer results showed by PTFs that developed from small dataset (e.g. Li et al., 2007, Rubio, 2008).

5.6.2 Evaluation the Performance of PTFs within the FAO Texture Classes

Since the PTFs are based mainly on particle size distribution in its predictions, the variability in texture class of the soil leads to change in PTFs performance. Also, the performance of some PTFs depends on the characteristics of the dataset from which it was developed and its performance changed if it was applied for soil conditions different from its development database. To assess the effect of variability in soil texture on PTFs performance, the entire dataset was classified according to the FAO classification system into five soil subsets and PTFs were evaluated using each soil subset. Table (5.10) represents the results of evaluation of the 22 PTFs within the FAO texture classes. The results showed that many PTFs have similar performance with a little difference in RMSE up to 0.01 m$^3$ m$^{-3}$ specially in continuous PTFs. The best two results of point PTFs and the best four results of continuous PTFs were selected. The results showed that the functions that had good performance in the overall evaluation also had good performance in different texture classes.

For point Pedotransfer functions, the PTFs developed by Rawls et al. (1982), which had the highest performance with the entire dataset, had the best performance in medium fine, medium and coarse soils. The PTF developed by Adhikary et al. (2008) have the highest performance in very fine, fine, medium fine soils. The PTF developed by Gupta and Larson (1979) perform well in medium and coarse soils. The PTF developed by Al Majou et al. (2007) has good
performance in fine soils and the PTF developed by Homaee and Firouzi (2008) has good performance in very fine soils.

For continuous PTFs, the functions that had the best performance with the entire dataset (Cosby et al. (1984), Rawls and Brakensiek (1985), Schaap et al. (1998) (Rosetta SSC) and Mayer and Jarvis (1999) almost had the best performance in all texture classes. The PTFs developed by Cosby et al. (1984) had the best performance in all texture classes. The PTFs developed by Rawls and Brakensiek (1985) had a good performance in fine, medium fine and coarse soils. The PTFs developed by Zacharias and Wessolek (2007) and by Schaap et al. (1998) (Rosetta SSC-BD) had good performance in medium and coarse soils. The PTFs developed by Saxton et al. (1986) and Saxton and Rawls (2006) had good performance in very fine soils. Finally, the PTFs developed by Mayer and Jarvis (1999) had good performance in very fine and fine soils. As mentioned before, the PTFs developed from US soils (Cosby et al., 1984; Rawls and Brakensiek, 1985) performed better than other PTFs developed from non US soils. Also the PTFs developed from larger dataset (Rawls et al., 1982; Wösten et al., 1999; Saxton and Rawls, 2006) perform better than other PTFs developed using small datasets.
Table (5. 10). Goodness of fit of point and continuous PTFs of soil water characteristics curve in different texture classes.

<table>
<thead>
<tr>
<th>Point Pedotransfer Functions</th>
<th>Very Fine Soils</th>
<th>Fine Soils</th>
<th>Medium Fine Soils</th>
<th>Medium Soils</th>
<th>Coarse Soils</th>
<th>Total dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ME</td>
<td>RMSE</td>
<td>EF</td>
<td>ME</td>
<td>RMSE</td>
<td>EF</td>
</tr>
<tr>
<td>Gupta &amp; Larson (1979)</td>
<td>0.093</td>
<td>0.134</td>
<td>-0.453</td>
<td>0.061</td>
<td>0.098</td>
<td>-0.113</td>
</tr>
<tr>
<td>Rawls et al. (1982)</td>
<td>0.063</td>
<td>0.097</td>
<td>0.154</td>
<td>0.028</td>
<td>0.066</td>
<td>0.493</td>
</tr>
<tr>
<td>Al Majou et al. (2007)</td>
<td>-0.003</td>
<td>0.081</td>
<td>0.426</td>
<td>-0.021</td>
<td>0.066</td>
<td>0.500</td>
</tr>
<tr>
<td>Adhikary et al. (2008)</td>
<td>-0.001</td>
<td>0.067</td>
<td>0.557</td>
<td>-0.018</td>
<td>0.061</td>
<td>0.577</td>
</tr>
<tr>
<td>Homaei and Firooz (2008)</td>
<td>-0.007</td>
<td>0.076</td>
<td>0.437</td>
<td>0.014</td>
<td>0.070</td>
<td>0.445</td>
</tr>
<tr>
<td>Continuous PTFs for predicting the parameters of Brooks and Corey (1964) Eqn. or its modified forms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cosby et al. (1984)</td>
<td>0.004</td>
<td>0.082</td>
<td>0.539</td>
<td>-0.015</td>
<td>0.066</td>
<td>0.472</td>
</tr>
<tr>
<td>Rawls and Brak. (1985)</td>
<td>-0.013</td>
<td>0.106</td>
<td>0.234</td>
<td>-0.008</td>
<td>0.073</td>
<td>0.345</td>
</tr>
<tr>
<td>Saxton et al. (1986)</td>
<td>-0.003</td>
<td>0.082</td>
<td>0.548</td>
<td>0.055</td>
<td>0.083</td>
<td>0.157</td>
</tr>
<tr>
<td>Mayer and Jarvis (1999)</td>
<td>0.008</td>
<td>0.088</td>
<td>0.472</td>
<td>0.003</td>
<td>0.075</td>
<td>0.316</td>
</tr>
<tr>
<td>Saxton and Rawls (2006)</td>
<td>-0.019</td>
<td>0.088</td>
<td>0.472</td>
<td>0.020</td>
<td>0.081</td>
<td>0.202</td>
</tr>
<tr>
<td>Adhikary et al. (2008)</td>
<td>0.033</td>
<td>0.105</td>
<td>0.258</td>
<td>-0.016</td>
<td>0.103</td>
<td>-0.291</td>
</tr>
<tr>
<td>Continuous PTFs for predicting the parameters of Van Genuchten, 1980 Eqn. or its modified form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vereecoen et al. (1989)</td>
<td>0.003</td>
<td>0.088</td>
<td>0.471</td>
<td>0.050</td>
<td>0.092</td>
<td>-0.038</td>
</tr>
<tr>
<td>Wosten et al. (1999)</td>
<td>0.027</td>
<td>0.098</td>
<td>0.348</td>
<td>0.026</td>
<td>0.076</td>
<td>0.300</td>
</tr>
<tr>
<td>Rosetta SSC</td>
<td>0.057</td>
<td>0.100</td>
<td>0.319</td>
<td>0.046</td>
<td>0.084</td>
<td>0.127</td>
</tr>
<tr>
<td>Rosetta SSC-BD</td>
<td>0.075</td>
<td>0.118</td>
<td>0.055</td>
<td>0.065</td>
<td>0.096</td>
<td>-0.119</td>
</tr>
<tr>
<td>Rubio (2008)</td>
<td>0.069</td>
<td>0.105</td>
<td>0.253</td>
<td>0.054</td>
<td>0.088</td>
<td>0.060</td>
</tr>
<tr>
<td>Rubio (2008)</td>
<td>0.067</td>
<td>0.124</td>
<td>-0.049</td>
<td>-0.001</td>
<td>0.078</td>
<td>0.254</td>
</tr>
<tr>
<td>Rajkai et al. (2004)</td>
<td>0.023</td>
<td>0.111</td>
<td>0.164</td>
<td>0.003</td>
<td>0.066</td>
<td>0.471</td>
</tr>
<tr>
<td>Rajkai et al. (2004)</td>
<td>0.112</td>
<td>0.143</td>
<td>-0.379</td>
<td>0.112</td>
<td>0.132</td>
<td>-1.132</td>
</tr>
<tr>
<td>Al-Majou et al. (2007)</td>
<td>0.054</td>
<td>0.113</td>
<td>0.135</td>
<td>0.016</td>
<td>0.081</td>
<td>0.188</td>
</tr>
<tr>
<td>Zacharias and Wessolek (2007)</td>
<td>0.122</td>
<td>0.151</td>
<td>-0.557</td>
<td>0.068</td>
<td>0.100</td>
<td>-0.222</td>
</tr>
<tr>
<td>Li et al. (2007)</td>
<td>0.226</td>
<td>0.278</td>
<td>-4.263</td>
<td>0.085</td>
<td>0.153</td>
<td>-1.883</td>
</tr>
</tbody>
</table>

Note: the marked cells show the first and second rank in the point PTFs, first four ranks in the continuous PTFs. ME, mean error; RMSE, root mean square error; $R^2$, coefficient of determination
Figure (5.4) Predicted vs. measured SWC of five point and six continuous PTFs represent the variation of soil water content predictions with different texture classes.
Figure (5.5) Predicted vs. measured SWC of eleven continuous PTFs represent the variation of soil water content predictions with different texture classes.
5.6.3 Evaluation the Performance of PTFs at certain Matric Potential Heads.

Table (5.11) represents the statistical comparison between the measured and predicted soil water contents from the 22 PTFs at three matric potential heads; 10, 33, and 1500 kPa. Three subsets were created from the soil database. These three subsets have water contents measured at 10, 33, and 1500 kPa matric potentials respectively. These potential heads represent the near saturated water content (θ\textsubscript{10}), the field capacity (θ\textsubscript{33}), and the wilting point (θ\textsubscript{1500}) were selected due to the importance of these variables, which are needed as inputs for many hydrological models. As mentioned above, the best two point PTFs and the best four continuous PTFs were selected (see Table 5.11).

The water content at 10 kPa matric potential is best predicted by PTFs developed by Adhikary et al. (2008) and Rawls et al. (1982) respectively as point PTFs. Better results can be obtained from Rosetta SSC, Cosby et al. (1984), Rosetta SSC-BD, Rawls and Brakensiek (1985) as continuous PTFs. However, the continuous PTFs show better performance than point PTFs especially in the wet part of the soil water characteristic curve.

The water content at field capacity (θ\textsubscript{33}) can be predicted by PTFs developed by Gupta and Larson (1979) and Rawls et al. (1982) respectively as point PTFs. Better performance showed by PTFs of Cosby et al. (1984), Mayer and Jarvis (1999), Adhikary et al. (2008) and Wösten et al. (1999) as continuous PTFs. The water content at wilting point (θ\textsubscript{1500}) can be predicted by PTFs of Rawls et al. (1982) and Homaee and Firouzi (2008) respectively as point PTFs. Better results showed by PTFs of Al Majou et al. (2007), Vereecken et al. (1989), Rawls and Brakensiek (1985) and Cosby et al. (1984) as continuous PTFs. The results indicated that the PTFs perform better with the higher potential heads and this could be due to the little error in measurements of water content at these pressures.
Table (5.11) Goodness of fit of five point and seventeen continuous PTFs of soil water characteristics curve in different matric potentials.

<table>
<thead>
<tr>
<th>PTFs</th>
<th>4 KPa</th>
<th>10 KPa</th>
<th>33 KPa</th>
<th>1500 KPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ME</td>
<td>RMSE</td>
<td>EF</td>
<td>ME</td>
</tr>
<tr>
<td></td>
<td>ME</td>
<td>RMSE</td>
<td>EF</td>
<td>ME</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Point Pedotransfer Functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gupta and Larson (1979)</td>
<td>0.052</td>
<td>0.099</td>
<td>0.459</td>
<td>0.032</td>
</tr>
<tr>
<td>Rawls et al. (1982)</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>0.025</td>
</tr>
<tr>
<td>Al Majou et al. (2007)</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-0.016</td>
</tr>
<tr>
<td>Adhikari et al. (2008)</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-0.023</td>
</tr>
<tr>
<td>Homaee and Firozzi (2008)</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-0.013</td>
</tr>
<tr>
<td>Continuous PTFs for predicting the parameters of Brooks and Corey (1964) Eqn. or its modified forms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cosby et al. (1984)</td>
<td>-0.015</td>
<td>0.092</td>
<td>0.531</td>
<td>-0.009</td>
</tr>
<tr>
<td>Rawls and Brakensiek (1985)</td>
<td>0.005</td>
<td>0.092</td>
<td>0.528</td>
<td>0.017</td>
</tr>
<tr>
<td>Saxton et al. (1986)</td>
<td>0.074</td>
<td>0.121</td>
<td>0.188</td>
<td>0.074</td>
</tr>
<tr>
<td>Mayer and Jarvis (1999)</td>
<td>-0.006</td>
<td>0.084</td>
<td>0.607</td>
<td>-0.003</td>
</tr>
<tr>
<td>Saxtom and Rawls (2006)</td>
<td>-0.111</td>
<td>0.168</td>
<td>-0.579</td>
<td>-0.077</td>
</tr>
<tr>
<td>Adhikari et al. (2008)</td>
<td>-0.084</td>
<td>0.194</td>
<td>-1.095</td>
<td>-0.066</td>
</tr>
<tr>
<td>Continuous PTFs for predicting the parameters of Van Gemenchten, 1980 Eqn. or its modified form</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vereeenen et al. (1989)</td>
<td>0.146</td>
<td>0.174</td>
<td>-0.696</td>
<td>0.133</td>
</tr>
<tr>
<td>Wosten et al. (1999)</td>
<td>-0.073</td>
<td>0.190</td>
<td>-1.011</td>
<td>-0.012</td>
</tr>
<tr>
<td>Rosetta SCC</td>
<td>-0.011</td>
<td>0.091</td>
<td>0.539</td>
<td>0.005</td>
</tr>
<tr>
<td>Rosetta SCC-BD</td>
<td>-0.017</td>
<td>0.085</td>
<td>0.598</td>
<td>0.015</td>
</tr>
<tr>
<td>Rubio (2008) (1)</td>
<td>0.009</td>
<td>0.101</td>
<td>0.435</td>
<td>0.018</td>
</tr>
<tr>
<td>Rubio (2008) (2)</td>
<td>-0.133</td>
<td>0.179</td>
<td>-0.794</td>
<td>-0.103</td>
</tr>
<tr>
<td>Rajkai et al. (2004) (1)</td>
<td>-0.094</td>
<td>0.154</td>
<td>-0.323</td>
<td>-0.099</td>
</tr>
<tr>
<td>Rajkai et al. (2004) (2)</td>
<td>0.040</td>
<td>0.139</td>
<td>-0.077</td>
<td>0.051</td>
</tr>
<tr>
<td>H. Al-Majou et al. (2007)</td>
<td>-0.088</td>
<td>0.164</td>
<td>-0.508</td>
<td>-0.050</td>
</tr>
<tr>
<td>Zacharias &amp; Wessolek (2007)</td>
<td>-0.020</td>
<td>0.089</td>
<td>0.557</td>
<td>0.010</td>
</tr>
<tr>
<td>Li et al. (2007)</td>
<td>-0.146</td>
<td>0.192</td>
<td>-1.049</td>
<td>-0.131</td>
</tr>
</tbody>
</table>

Note: the marked cells show the first and second rank in the point PTFs, first four ranks in the continuous PTFs; ME, mean error; RMSE, root mean square error; R², coefficient of determination.

### 5.7 Comparison between PTFs Performance in the Current Study with Previous Studies

Results of this study were compared with results of five previous studies, which evaluated PTFs for predicting SWCC (Table 5.12 and Fig. 5.6). All the five studies used the RMSE as statistical measures. Nine functions, out of the 22 functions evaluated in this study, have been previously evaluated at least twice. The PTFs of Rawls and Brakensiek (1985) and Vereeenen et al. (1989) have been
evaluated in the five previous studies. The PTFs of Rawls et al. (1982) and Cosby et al. (1984) have been evaluated three times. The rest of the nine PTFs have been evaluated two times.

The results showed good agreement with the previous studies. The best performing PTF in this study (Rawls et al., 1982) showed better performance in the studies of Tietje and Tapkenhinrichs (1993) and Cornelies et al. (2001), while showed poorer performance in the study of Ungaro and Calzolari (2001). The PTF of Vereecken et al. (1989), which has the poorer performance in this study, showed the best performance in the study of Cornelies et al., 2001, which used an evaluation dataset from the same region (Belgium) of its development dataset. The two models, Rosetta-SSC and Rosetta-SSCBD were evaluated in two previous studies. The performance of the two models was better in the study of Schaap et al. (2004), while it was poorer than the results of this study as shown in the study of Schaap et al. (1998). The change in performance of the two models in these two studies is related to the size of the evaluation database. The three models of Gupta and Larson (1979), Saxton et al. (1986), and Wösten et al. (1999) were previously evaluated twice and showed better results. The variability in PTFs performance with different studies refer to the uncertainty in PTFs predictions, which caused by many reasons such as uncertainty in measurements, characteristics of soil database. Despite the uncertainty in PTFs predictions, its results are acceptable and it is considered to be an alternative of the field and lab measurements.
Table (5.12) PTFs validation results: root mean square error ($m^3 \cdot m^{-3}$) for seven independent datasets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of dataset</td>
<td>1116 (Germany)</td>
<td>1177 (USA)</td>
<td>1209 (USA)</td>
<td>69 Belgium</td>
<td>139 Italy</td>
<td>47435 NRCS</td>
<td>5000 USA</td>
</tr>
<tr>
<td>No. of samples</td>
<td>13</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>Gupta &amp; Larson 1979</td>
<td>0.0591</td>
<td>--------</td>
<td>0.0602</td>
<td>--------</td>
<td>0.0602</td>
<td>--------</td>
<td>0.091</td>
</tr>
<tr>
<td>Rawls et al. 1982</td>
<td>0.0631</td>
<td>0.091</td>
<td>0.0653</td>
<td>0.092</td>
<td>0.092</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>Braken. &amp; Rawls 1985</td>
<td>0.0751</td>
<td>0.080</td>
<td>0.101</td>
<td>0.0774</td>
<td>0.0882</td>
<td>0.0634</td>
<td>0.082</td>
</tr>
<tr>
<td>Cosby et al. 1984</td>
<td>0.0668</td>
<td>--------</td>
<td>0.111</td>
<td>--------</td>
<td>0.0616</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>Vereecken et al. 1989</td>
<td>0.0531</td>
<td>0.036</td>
<td>0.098</td>
<td>0.0412</td>
<td>0.0915</td>
<td>0.0657</td>
<td>0.144</td>
</tr>
<tr>
<td>Wosten et al. 1999</td>
<td>--------</td>
<td>--------</td>
<td>0.0518</td>
<td>--------</td>
<td>0.0575</td>
<td>0.098</td>
<td></td>
</tr>
<tr>
<td>Rosetta SSC</td>
<td>0.0796</td>
<td>0.078</td>
<td>0.104</td>
<td>0.0698</td>
<td>0.0635</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>Rosetta SSC-BD</td>
<td>0.0796</td>
<td>0.078</td>
<td>0.087</td>
<td>0.0698</td>
<td>0.0641</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>Saxton et al. 1986</td>
<td>0.0796</td>
<td>0.078</td>
<td>0.104</td>
<td>0.0698</td>
<td>0.0635</td>
<td>0.083</td>
<td>0.110</td>
</tr>
</tbody>
</table>
Figure (5. 6) Root mean square error values of nine Pedotransfer functions from six previous studies compared to the results obtained from the current study.
5.8 Potential Application of the Study Results

Direct measurement is the best and accurate way to obtain soil parameters that are needed for hydrologic models applications. However, direct measurements are feasible for field scale and small scale applications that need limited number of measurements. With large scale (watershed or basin scale) applications, direct measurements are time consuming and expensive, which make it infeasible. The authors recommend using the results of this study to facilitate the applications of hydrologic models by predicting the SWCC date required for these models using the best performing PTFs. Table 6 listing the best performing PTFs (best two point PTFs and best four continuous PTFs) corresponding to soil texture class and matric potential head. The user is recommended to use the PTFs for predicting SWCC according to the ranks listed in Table 6. Using the best performing PTF, the user can estimate an average value for SWCC, then, using the RMSE value, upper and lower limits can be identified.
Table 5.13. Best performing point and continuous PTFs corresponding to different texture classes and matric potentials

<table>
<thead>
<tr>
<th>Texture Class or Matric Potential</th>
<th>Point PTFs</th>
<th>Continuous PTFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Fine</td>
<td>(1) Adhikary et al. (2008)</td>
<td>(1) Saxton et al. (1986)</td>
</tr>
<tr>
<td></td>
<td>(2) H Al Majou et al. (2007)</td>
<td>(2) Rajkai et al.</td>
</tr>
<tr>
<td></td>
<td>(2) Rawls et al. (1982)</td>
<td>(2) Rajkai et al.</td>
</tr>
<tr>
<td></td>
<td>(3) Cosby et al. (1984)</td>
<td>(4) Rosetta SSC</td>
</tr>
<tr>
<td></td>
<td>(2) Gupta &amp; Larson (1979)</td>
<td>(2) Cosby et al. (1984)</td>
</tr>
<tr>
<td></td>
<td>(2) Gupta &amp; Larson (1979)</td>
<td>(2) Cosby et al. (1984)</td>
</tr>
<tr>
<td>33 KPa</td>
<td>(1) Gupta &amp; Larson (1979)</td>
<td>(1) Cosby et al. (1984)</td>
</tr>
<tr>
<td></td>
<td>(2) Homae and Firouzi (2008)</td>
<td>(2) Vereecken et al.</td>
</tr>
<tr>
<td></td>
<td>(2) Adhikary et al. (2008)</td>
<td>(2) Brake &amp; Rawls</td>
</tr>
<tr>
<td></td>
<td>(3) Mayer &amp; Jarvis (1999)</td>
<td>(4) Rosetta SSC</td>
</tr>
</tbody>
</table>

Following the procedure outlined in this article, Abdelbaki and Youssef (2010) assessed the feasibility of applying the widely used drainage water management model, DRAINMOD (Skaggs, 1980), using SWCC estimated by the three best performing PTFs. They used in the analysis data from four U.S. agricultural drained sites, having different soils, crops, drainage systems, and climatological conditions, for which the model has been previously calibrated and validated (Wang et al., 2006; Youssef et al., 2006; Ale et al., 2009; Thorp et al., 2009). Measured annual drainage was compared to predicted drainage using estimated and calibrated SWC values. For the four sites, the normalized root mean square errors (NRMSEs) in yearly drainage predicted by the calibrated model
were in the range of 9 to 24% with corresponding Nash-Sutcliffe modeling efficiencies (EF) ranging from 0.62 to 0.91. The use of PTF estimated- SWCC moderately increased the NRMSE to 16-32% and decreased the EF to 0.35-0.89. The relatively small errors in DRAINMOD predictions induced by using PTFs-estimated SWCC support the use of this method to estimate SWCC for large scale applications of hydrologic models such as DRAINMOD.

5.9 Conclusions

The evaluation and comparison of the 22 PTFs that considered in this study enabled us to draw the following conclusions about the PTFs. The function developed by Rawls et al. (1982) showed the best performance among the five point PTFs followed by the PTF developed by Adhikary et al. (2008). The PTFs developed by Cosby et al. (1984), Rawls and Brakensiek (1985), Mayer and Jarvis (1999), and Schaap et al. (2001) (Rosetta SSC) showed the best performances among the 17 continuous PTFs considered in this study. Also, the performance of the continuous PTFs is almost better than the performance of the point PTFs due to its continuous results. Another conclusion, the performance of PTFs changes with the variation of soil texture classes and with the variation of the potential heads. The PTFs that showed a good performance when evaluated using the entire dataset are almost the PTFs that showed the good performance in different texture classes and in different matric potentials. The results indicated that the PTFs perform better with the higher potential heads and this could be due to the little error in measurements of water content at these pressures. Also, the PTFs that developed using US soils showed better results than other functions.
CHAPTER (6)

ASSESSING THE FEASIBILITY OF DRAINMOD APPLICATION USING SOIL HYDRAULIC PROPERTIES ESTIMATED BY PEDOTRANSFER FUNCTIONS

6.1 Introduction

DRAINMOD (Skaggs, 1980) is a field-scale hydrologic model that was developed in early 1980s to simulate the hydrology of poorly drained high water table soils. DRAINMOD and the companion carbon and nitrogen model, DRAINMOD-N II (Youssef et al., 2005), are used to design sustainable crop production systems on artificially drained lands. Compared with many hydrologic models, DRAINMOD is easy to use, requires relatively few inputs, and yet provides quite accurate predictions. In the last two decades, many researchers have extensively tested the model for different climatic conditions, soil types, and farming practices (e.g. Skaggs et al., 1981; Fouss et al., 1987; Sabbagh et al., 1993). In these evaluations, the model was calibrated and validated against field measured water table and subsurface drain flow data. During the calibration process, model inputs of soil hydraulic properties, which are usually based on field/lab measurements, are often adjusted within “acceptable range” to minimize the errors between measured and predicted water tables and drain flows. The calibrated model consistently showed high performance in predicting water table fluctuation and drain flow during the validation periods of these model evaluations. However, any large scale application of the model will be without calibration and will also be based on estimated rather than measured soil hydraulic properties.

DRAINMOD predictions are most sensitive to lateral saturated hydraulic conductivity ($K_{sat}$) of the soil profile followed by the unsaturated hydraulic
properties estimated from the soil water characteristic (SWC) relationship (volumetric water content vs. pressure head) (Skaggs, 1980; Skaggs, 1982; Anderson et al., 1987; Workman and Skaggs, 1994; Haan and Skaggs, 2003; Wang et al., 2006b). For large scale applications of DRAINMOD, the direct measurement of \( K_{\text{sat}} \) and SWC data will be costly and time consuming. Alternatively these inputs can be estimated from readily available soil data (e.g. particle size distribution, bulk density, and soil organic matter content) using pedotransfer functions (PTFs) (Bouma, 1989; Schaap et al., 2001).

A few studies have been conducted to evaluate the feasibility of applying DRAINMOD with \( K_{\text{sat}} \) and SWC data predicted using PTF models (Borin et al. 2000; Qi et al., 2006; Salazar et al., 2008). In Italy, Borin et al. (2000) compared measured water table depths and drain flows to values predicted by DRAINMOD using three levels of details of soil inputs: 1) soil texture and porosity of top layer are the only available data; 2) soil texture and porosity are known for the entire soil profile; 3) \( K_{\text{sat}} \) and SWC data are measured for the entire soil profile. For scenarios 1 and 2, \( K_{\text{sat}} \) and SWC data were estimated using PTFs developed by Rawls et al. (1993) and Rawls et al. (1982) respectively. Borin et al. concluded that DRAINMOD was able to describe water table fluctuation and drain discharge under limited input of soil hydraulic properties. Two similar studies have recently been conducted in Iowa, USA (Qi et al., 2006) and Sweden (Salazar et al., 2008) to investigate the impact of using soil hydraulic properties predicted by the PTF model ROSETTA (Schaap et al., 2001) on the accuracy of DRAINMOD predictions of drainage volumes. Qi et al. (2006) conducted a study to determine the sufficient level of information to be used as inputs to the PTF model, ROSETTA, for obtaining DRAINMOD inputs (i.e. \( K_{\text{sat}} \) and SWC data). They used three levels of information obtained from soil survey. Level 1, they used soil texture and bulk density as inputs to ROSETTA. Level 2, they used soil texture, bulk density, and two water contents at 33, 1500 kPa as inputs to ROSETTA.
Level 3, they used calibrated $K_{sat}$ and SWC data based on initial values determined according to level 1. DRAINMOD predictions of subsurface drainage using these three levels of inputs were compared to observations. The results showed that ROSETTA with soil survey data provides a quick and easy way to feed DRAINMOD with the required input parameters. Likewise, Salazar et al. (2008) used the PTF-model, ROSETTA, with four different levels of inputs; texture class; soil texture; soil texture and bulk density; and soil texture, bulk density, and two water contents at 33, 1500 kPa to determine the $K_{sat}$ values required for DRAINMOD. The model was applied using the four groups of ROSETTA-estimated and laboratory-measured $K_{sat}$ values. The model predictions of drainage outflows using the five $K_{sat}$ groups were compared to observations. They concluded that ROSETTA-estimated $K_{sat}$ values can be used in DRAINMOD applications as accurate as the laboratory measured values.

Over the past two decades, considerable number of PTFs have been suggested for predicting $K_{sat}$ (e.g. Cosby et al., 1984; Jabro, 1992; Minasny and McBratney, 2000) and SWCC (e.g. Wösten et al., 1999; Zacharias and Wessolek, 2007; Homae and Firouzi, 2008, Adhikary et al., 2008). Several studies have been conducted to evaluate the performance of different PTFs for predicting $K_{sat}$ and SWC data (Tietje and Tapkenhinrichs, 1993; Tietje and Hennings, 1996; Minasny and McBratney, 2000; Wagner et al., 2001; Cornelis et al., 2001; Julia et al., 2004). Abdelbaki et al. (2009a, b) have identified the best performing PTFs for predicting $K_{sat}$ and SWC data for U.S. soils. They evaluated 24 PTFs for predicting $K_{sat}$ and 22 PTFs for predicting SWCC using soil databases covering wide range of the U.S. soils. They divided the databases into five texture classes (very fine, fine, medium fine, medium, and coarse soils) and identified the best performing PTFs for predicting $K_{sat}$ and SWCC for each soil texture class. Abdelbaki et al. (2009a, b) showed that ROSETTA, which has been previously used with DRAINMOD (Qi et al., 2006; Salazar et al., 2008) was not the best
performing PTF for predicting $K_{\text{sat}}$ and SWCC for all soil texture classes. The goal of this study was to extensively assess the feasibility of running DRAINMOD using soil hydraulic parameters ($K_{\text{sat}}$ and/or SWCC) predicted by the best performing PTFs.

6.2 Materials and Methods

DRAINMOD has recently been tested against field measured data collected from four U.S. agricultural drained sites; three sites are located in the U.S. Midwest (Wang et al., 2006a; Ale et al., 2009; Thorp et al., 2009) and one site is located in south eastern U.S. (Youssef et al., 2006). These sites have different soils, crops, drainage systems, and climatological conditions. In these evaluations, the model was calibrated and validated against field measured subsurface drain flow (and water table) data. During the calibration process, measured soil hydraulic properties have been adjusted to minimize the errors between measured and predicted drain flow (and water table) data. The results of these four evaluations have indicated that the calibrated model accurately predicted drain flows (and water table fluctuation). In order to assess the feasibility of using DRAINMOD with PTFs-predicted soil properties, drain flows predicted by the calibrated model for the four sites were compared to drain flows predicted by DRAINMOD with soil hydraulic properties estimated by PTFs. In this study, the soil hydraulic properties for the four sites have been estimated using the best performing PTFs for each site. Then, DRAINMOD model with PTFs-estimated soil parameters was used to simulate the hydrology of the four sites and predict subsurface drainage volumes. These predicted drainage rates have been compared to measured values and previously predicted values by the calibrated model for each site.
6.2.1 Study Sites and Measured Soil Hydraulic Properties

Four sites from the United States were used in this study. Two sites from Indiana; The experimental drainage field at Southeast Purdue Agricultural Center (SEPAC); and Purdue University’s water quality field station (WQFS). The third field is near Story City, IA (42.2° N, 93.6° W), the fourth site is the tidewater Experiment station (TES) from north Carolina is a nearly flat, 13.8-ha site. The SEPAC site was under conventional drainage management and planted with continuous corn crop and divided into east and west blocks each block has three drain spacing; 5,10, 20 m. Six years of measured subsurface drainage are available from 1985 to 1990. In the WQFS site, Forty-eight treatment plots (12 cropping system treatments replicated four times) were established in a randomized complete block design, with 12 treatment plots per block. Eleven years of measured subsurface drainage from 1995-2005 are available and were used in this study. The study site near Story City, Iowa is a 22 ha field has Kossuth silty clay loam and Ottosen clay loam soil types. The site under controlled drainage and has available measured subsurface drainage from 1996-2005. Finally, the fourth site located at the Tidewater Experiment Station (TES) near Plymouth, in the North Carolina lower coastal plain. TES site is divided into eight experimental field plots, four of which were used in this study. Two sets of parallel subsurface drains were installed on the site in 1985 and 1991. The drains in the first set were 0.8 to 1.0 m below ground surface and spaced 23 m apart. Drains in the second set were located midway between those in the first set, but at deeper depths of 1.1 to 1.3 m. Measured subsurface drainage values for six years (1992-1997) are available and used in this study. The available measured soil parameters for the four sites are presented in Table (6.1).
Table (6.1) measured soil properties for the four study sites

<table>
<thead>
<tr>
<th>Site</th>
<th>Layer</th>
<th>Sand</th>
<th>Silt</th>
<th>Clay</th>
<th>BD</th>
<th>φ</th>
<th>θ₁₀</th>
<th>θ₃₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEPAC</td>
<td>1</td>
<td>22.00</td>
<td>65.80</td>
<td>12.20</td>
<td>1.42</td>
<td>0.395</td>
<td>0.316</td>
<td>0.275</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>20.00</td>
<td>64.40</td>
<td>15.60</td>
<td>1.65</td>
<td>0.420</td>
<td>0.351</td>
<td>0.331</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>17.70</td>
<td>63.30</td>
<td>19.00</td>
<td>1.66</td>
<td>0.405</td>
<td>0.365</td>
<td>0.343</td>
</tr>
<tr>
<td>WQFS</td>
<td>1</td>
<td>22.00</td>
<td>65.80</td>
<td>12.20</td>
<td>1.25</td>
<td>0.559</td>
<td>0.365</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>20.00</td>
<td>64.40</td>
<td>15.60</td>
<td>1.34</td>
<td>0.456</td>
<td>0.385</td>
<td>0.345</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>17.70</td>
<td>63.30</td>
<td>19.00</td>
<td>1.66</td>
<td>0.455</td>
<td>0.423</td>
<td>0.398</td>
</tr>
<tr>
<td>Story</td>
<td>1</td>
<td>22.00</td>
<td>33.00</td>
<td>45.00</td>
<td>1.60</td>
<td>0.562</td>
<td>0.472</td>
<td>0.425</td>
</tr>
<tr>
<td>City</td>
<td>2</td>
<td>21.00</td>
<td>33.00</td>
<td>46.00</td>
<td>1.65</td>
<td>0.540</td>
<td>0.454</td>
<td>0.409</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>47.00</td>
<td>29.00</td>
<td>24.00</td>
<td>1.90</td>
<td>0.442</td>
<td>0.373</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>35.00</td>
<td>40.00</td>
<td>25.00</td>
<td>1.90</td>
<td>0.340</td>
<td>0.288</td>
<td>0.261</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>35.00</td>
<td>40.00</td>
<td>25.00</td>
<td>1.90</td>
<td>0.321</td>
<td>0.273</td>
<td>0.247</td>
</tr>
<tr>
<td>TES</td>
<td>1</td>
<td>65.70</td>
<td>21.40</td>
<td>12.90</td>
<td>1.37</td>
<td>0.415</td>
<td>0.351</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>66.60</td>
<td>17.40</td>
<td>16.00</td>
<td>1.69</td>
<td>0.334</td>
<td>0.293</td>
<td>0.270</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>77.30</td>
<td>10.90</td>
<td>11.80</td>
<td>1.69</td>
<td>0.359</td>
<td>0.218</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>79.30</td>
<td>12.90</td>
<td>8.10</td>
<td>1.48</td>
<td>0.398</td>
<td>0.244</td>
<td>0.223</td>
</tr>
</tbody>
</table>

φ, the soil porosity (cm³/cm³); θ₁₀, the measured soil water content at 10 kPa potential head; θ₃₃, the measured soil water content at 33 kPa potential head.

6.2.2 Pedotransfer Functions Used for Predicting Input Parameters

Different from the previous studies (Qi et al., 2006; Salazar et al., 2008), which used the PTFs models (ROSETTA) (Schaap et al., 2001) for predicting the inputs of DRAINMOD, we identified the best performing in predicting $K_{sat}$ and SWCC for US soils (Abdelbaki et al., 2009 a, b). For selecting the PTF of interest, the user of the model has to define the available measured soil properties and the soil texture class. In this study, according to the available measured soil parameters, the $K_{sat}$ can be predicted by two groups of PTFs; group1, which require only the effective porosity to predict $K_{sat}$; group 2, which require the particle size distribution and bulk density as inputs to predict $K_{sat}$. The SWC data can be predicted from PTFs that require the particle size distribution and bulk density as input parameters. For each PTFs group the first three ranks PTFs were used. For the four study sites, $K_{sat}$ was predicted using Group1 PTFs from the PTFs developed by (Suleiman and Ritchie, 2001; Minasny and McBratney, 2000;
and Forrest et al., 1985) (Table 6.2). For SEPAC, WQFS, and STORY CITY (layers 3, 4, 5) the texture class is medium and PTFs group2 are (Cosby et al., 1984; Dane and Puckett, 1994 and Jabro et al., 1992) and the PTFs for predicting SWCC are (Zacharias et al., 2007; Cosby et al., 1984; Rosetta SSC-BD) (Table 6.3). For STORY CITY (layers 1,2) the texture class is fine and PTFs group2 are (Rosetta SSC, Cosby et al. 1984; Rosetta SSC-BD) and the PTFs for predicting SWCC are (Adhikary et al., 2008; Cosby et al., 1984; Homae and Firouzi, 2008). For TES site the texture class is coarse and PTFs group 2 are (Puckett et al., 1985; Julia et al. 2004 and Cosby et al., 1984) and the PTFs for predicting SWCC are (Zacharias et al., 2007; Rosetta SSC-BD; Cosby et al., 1984).

Table (6. 2) Pedotransfer functions used for predicting saturated hydraulic conductivity

<table>
<thead>
<tr>
<th>Layers</th>
<th>STORY CITY</th>
<th>SEPAC</th>
<th>WQFS</th>
<th>STORY CITY</th>
<th>TES (NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture class</td>
<td>Fine</td>
<td>1,2</td>
<td>1,2,3</td>
<td>Medium</td>
<td>Coarse</td>
</tr>
<tr>
<td>PTFs group (1)K1</td>
<td>…………..Suleiman and Ritchie (2001).……………………………………..</td>
<td>K2</td>
<td>………..Minasny and McBratney (2000).……………………………………..</td>
<td>K3</td>
<td>……………Forrest et al. (1985).……………………………………..</td>
</tr>
<tr>
<td>KK1</td>
<td>Rosetta SSC</td>
<td>Cosby et al. (1984)</td>
<td>Puckett et al. (1985)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

K1, K2, K3 are the three best performing PTFs of group 1 for predicting $K_{sat}$; KK1, KK2, KK3 are the three best performing PTFs of group 2 for predicting $K_{sat}$.

Table (6. 3) Pedotransfer functions used for predicting soil water characteristics data

<table>
<thead>
<tr>
<th>Layers</th>
<th>STORY CITY</th>
<th>SEPAC</th>
<th>WQFS</th>
<th>STORY</th>
<th>TES (NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture class</td>
<td>Fine</td>
<td>1,2</td>
<td>1,2,3</td>
<td>Medium</td>
<td>Coarse</td>
</tr>
<tr>
<td>03</td>
<td>Homae and Firouzi</td>
<td>Rosetta SSC-BD</td>
<td>Cosby et al. (1984)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\theta_1$, $\theta_2$, $\theta_3$ are the three best performing PTFs for predicting the SWCC.
6.2.3 Model Evaluation

To evaluate the feasibility of applying DRAINMOD for practical purposes using indirect estimated soil hydraulic data, it was proposed that the model be run with 17 different sets of input soil hydraulic properties (Table 6.4). DRAINMOD simulated subsurface outflows using calibrated $K_{\text{sat}}$ values and SWC data (Dc dataset). Also, DRAINMOD simulate subsurface outflows using calibrated SWC data in addition to $K_{\text{sat}}$ values predicted from the first three ranks PTFs that require only the effective porosity as inputs (D-K1, D-K2, and D-K3). Using more inputs for predicting $K_{\text{sat}}$, DRAINMOD simulated the subsurface outflows using $K_{\text{sat}}$ values predicted from the first three ranks PTFs that require the particle size distribution and bulk density as inputs (D-KK1, D-KK2, and D-KK3). With the calibrated values of $K_{\text{sat}}$ with SWC data predicted from the first three ranks PTFs, DRAINMOD simulates the subsurface outflows (D-\(\theta\)1, D-\(\theta\)2, and D-\(\theta\)3). Using predicted $K_{\text{sat}}$ values and SWC data, DRAINMOD simulates subsurface outflows with the first three ranks PTFs for predicting $K_{\text{sat}}$ and SWC data (D-K-\(\theta\) (1), D-K-\(\theta\) (2), and D-K-\(\theta\) (3)). Lastly, four simulations were conducted using $K_{\text{sat}}$ and SWC data that are both estimated by the PTF-model, ROSETTA using four different levels of inputs. DRAINMOD outputs obtained from the 17 dataset were compared to the measured subsurface drainage (Do) for each study site.
### Table 6.4 Description of datasets used for DRAINMOD

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do</td>
<td>Observed values</td>
</tr>
<tr>
<td>Dc</td>
<td>DRAINMOD simulation using calibrated input parameters</td>
</tr>
<tr>
<td>D-K1</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the first rank PTF that requires inputs of effective porosity only</td>
</tr>
<tr>
<td>D-K2</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the second rank PTF that requires inputs of effective porosity only</td>
</tr>
<tr>
<td>D-K3</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the third rank PTF that requires inputs of effective porosity only</td>
</tr>
<tr>
<td>D-KK1</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the first rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-KK2</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the second rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-KK3</td>
<td>DRAINMOD simulation using $K_{sat}$ values predicted by the third rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-01</td>
<td>DRAINMOD simulation using $\theta(h)$ values predicted by the first rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-02</td>
<td>DRAINMOD simulation using $\theta(h)$ values predicted by the second rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-03</td>
<td>DRAINMOD simulation using $\theta(h)$ values predicted by the third rank PTF that requires inputs of particle size distribution and bulk density</td>
</tr>
<tr>
<td>D-K-0(1)</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by the first ranks PTFs.</td>
</tr>
<tr>
<td>D-K-0(2)</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by the second ranks PTFs.</td>
</tr>
<tr>
<td>D-K-0(3)</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by the third ranks PTFs.</td>
</tr>
<tr>
<td>D-RS1</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by ROSETTSA using sand, silt and clay contents</td>
</tr>
<tr>
<td>D-RS2</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by ROSETTSA using sand, silt, and clay contents and bulk density</td>
</tr>
<tr>
<td>D-RS3</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by ROSETTSA using sand, silt, and clay contents, bulk density, and water content at 33 kPa</td>
</tr>
<tr>
<td>D-RS4</td>
<td>DRAINMOD simulation using $K_{sat}$ and $\theta(h)$ values predicted by ROSETTSA using sand, silt, and clay contents, bulk density, and water contents at 33, 1500 kPa</td>
</tr>
</tbody>
</table>

#### 6.2.4 Statistical Evaluation Criteria

Drainage rates predicted by the model using calibrated/measured and PTF-estimated $K_{sat}$ and SWC data were compared to measured values for each site. Statistical performance measures used for comparing predicted and measured subsurface drainage are the Normalized Root Mean Square Error (NRMSE) (Eqn. 6.1),
and Nash-Sutcliffe modeling efficiency (EF) (Eqn. 6.2),

\[ EF = \frac{\sum_{i=1}^{N}(O_i - \bar{O})^2 - \sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \]  \hspace{1cm} (6.2)

where \( N \) is the number of data points, \( P_i \) and \( O_i \) are predicted and observed drainage in year \( i \), respectively, and \( \bar{O} \) is observed yearly drainage. The descriptions of these statistical measures are shown in (Table 6.5).

<table>
<thead>
<tr>
<th>Statistical Measure</th>
<th>Performance Measures</th>
<th>Range</th>
<th>Best Value</th>
<th>Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Root Mean Square Error (NRMSE)</td>
<td>Estimating of the normalized standard deviation of the model predictions</td>
<td>( 0, \infty )</td>
<td>0</td>
<td>Smaller value, better performance</td>
</tr>
<tr>
<td>Nash-Sutcliffe Modeling Efficiency, EF</td>
<td>Compares the predicted values to the mean of the observed values</td>
<td>(-\infty, \infty )</td>
<td>1</td>
<td>( 0–1 ): the model is a better predictor than the mean of the observed values, negative: the mean of observed values is a better predictor than the model; perfect model</td>
</tr>
</tbody>
</table>
6.3 Results and Discussion

6.3.1 PTF-Estimated Vs. Measured/Calibrated $K_{sat}$ Values

Table (6.6) presents measured/calibrated and PTF-estimated $K_{sat}$ values by group 1 PTFs (K1, K2, and K3) and by group 2 PTFs (KK1, KK2, and KK3) for the entire soil profile of each study site. The equivalent $K_{sat}$ values of the soil profiles and modeling efficiency of DRAINMOD simulated annual subsurface drainage using these sets of $K_{sat}$ values as inputs are also presented. The best two values of equivalent $K_{sat}$ compared to the measured/calibrated values were selected. The results indicated that the accuracy of DRAINMOD predictions depends mainly on the performance of PTFs in predicting $K_{sat}$. The best two results of simulated subsurface drainage compared to the observations obtained by using $K_{sat}$ values estimated by the two best performing PTFs. Also, the deviation of DRAINMOD predictions increases with the increase of PTFs deviation in predicting $K_{sat}$ values. Group 2 PTFs, which requires more detailed soil information as inputs for estimating $K_{sat}$, showed the best performance for all sites except for the TES site, which has coarser soil profile.
**Table (6.6) Soil properties, PTF-estimated, and measured/calibrated $K_{\text{sat}}$ values for the four study sites**

<table>
<thead>
<tr>
<th>Site</th>
<th>Depth cm</th>
<th>Sand %</th>
<th>Silt %</th>
<th>Clay %</th>
<th>BD m$^{-3}$</th>
<th>$\theta_{10}$ m$^{-3}$</th>
<th>$\theta_{33}$ m$^{-3}$</th>
<th>PTF-estimated $K_{\text{sat}}$</th>
<th>Measured/Calibrated $K_{\text{sat}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEPA</td>
<td>1 25</td>
<td>22.0</td>
<td>65.8</td>
<td>12.2</td>
<td>1.42</td>
<td>0.395 0.316 0.275</td>
<td>0.59 0.21 1.31 0.98</td>
<td>5.24 2.06</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>2 30</td>
<td>20.0</td>
<td>64.4</td>
<td>15.6</td>
<td>1.65</td>
<td>0.420 0.351 0.331</td>
<td>0.23 0.13 0.41 0.88</td>
<td>3.21 0.14</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>3 120</td>
<td>17.7</td>
<td>63.3</td>
<td>19.0</td>
<td>1.66</td>
<td>0.405 0.365 0.343</td>
<td>0.07 0.02 0.10 0.79</td>
<td>1.97 0.10</td>
<td>0.26</td>
</tr>
</tbody>
</table>

**WQFS**

<table>
<thead>
<tr>
<th>Site</th>
<th>Depth cm</th>
<th>Sand %</th>
<th>Silt %</th>
<th>Clay %</th>
<th>BD m$^{-3}$</th>
<th>$\theta_{10}$ m$^{-3}$</th>
<th>$\theta_{33}$ m$^{-3}$</th>
<th>PTF-estimated $K_{\text{sat}}$</th>
<th>Measured/Calibrated $K_{\text{sat}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 5</td>
<td>22.0</td>
<td>65.8</td>
<td>12.2</td>
<td>1.25</td>
<td>0.559 0.365 0.337</td>
<td>4.08 5.74 14.58 0.98</td>
<td>5.24 12.68</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>2 30</td>
<td>20.0</td>
<td>64.4</td>
<td>15.6</td>
<td>1.34</td>
<td>0.456 0.385 0.345</td>
<td>0.46 0.14 0.97 0.88</td>
<td>3.21 3.77</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>3 147</td>
<td>17.7</td>
<td>63.3</td>
<td>19.0</td>
<td>1.66</td>
<td>0.455 0.423 0.398</td>
<td>0.06 0.01 0.07 0.79</td>
<td>1.97 0.10</td>
<td>2.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>Depth cm</th>
<th>Sand %</th>
<th>Silt %</th>
<th>Clay %</th>
<th>BD m$^{-3}$</th>
<th>$\theta_{10}$ m$^{-3}$</th>
<th>$\theta_{33}$ m$^{-3}$</th>
<th>PTF-estimated $K_{\text{sat}}$</th>
<th>Measured/Calibrated $K_{\text{sat}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 15</td>
<td>22.0</td>
<td>33.0</td>
<td>45.0</td>
<td>1.60</td>
<td>0.562 0.472 0.425</td>
<td>0.89 0.35 2.20 0.45</td>
<td>0.60 0.15</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>2 60</td>
<td>21.0</td>
<td>33.0</td>
<td>46.0</td>
<td>1.65</td>
<td>0.540 0.454 0.409</td>
<td>0.77 0.29 1.85 0.48</td>
<td>0.58 0.12</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>3 90</td>
<td>47.0</td>
<td>29.0</td>
<td>24.0</td>
<td>1.90</td>
<td>0.442 0.373 0.337</td>
<td>0.39 0.13 0.78 1.64</td>
<td>0.96 0.01</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>4 150</td>
<td>35.0</td>
<td>40.0</td>
<td>25.0</td>
<td>1.90</td>
<td>0.340 0.288 0.261</td>
<td>0.16 0.05 0.26 1.16</td>
<td>0.83 0.01</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>5 299</td>
<td>35.0</td>
<td>40.0</td>
<td>25.0</td>
<td>1.90</td>
<td>0.321 0.273 0.247</td>
<td>0.13 0.03 0.20 1.16</td>
<td>0.83 0.01</td>
<td>3.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>Depth cm</th>
<th>Sand %</th>
<th>Silt %</th>
<th>Clay %</th>
<th>BD m$^{-3}$</th>
<th>$\theta_{10}$ m$^{-3}$</th>
<th>$\theta_{33}$ m$^{-3}$</th>
<th>PTF-estimated $K_{\text{sat}}$</th>
<th>Measured/Calibrated $K_{\text{sat}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 28</td>
<td>65.7</td>
<td>21.4</td>
<td>12.9</td>
<td>1.37</td>
<td>0.415 0.351 0.325</td>
<td>0.23 0.10 0.42 1.23</td>
<td>2.32 3.24</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>2 44</td>
<td>66.6</td>
<td>17.4</td>
<td>16.0</td>
<td>1.69</td>
<td>0.334 0.293 0.270</td>
<td>0.08 0.02 0.11 0.67</td>
<td>2.42 3.17</td>
<td>2.02</td>
</tr>
<tr>
<td></td>
<td>3 28</td>
<td>77.3</td>
<td>10.9</td>
<td>11.8</td>
<td>1.69</td>
<td>0.359 0.218 0.196</td>
<td>1.54 1.77 4.35 1.53</td>
<td>4.09 4.54</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>4 140</td>
<td>79.3</td>
<td>12.9</td>
<td>8.1</td>
<td>1.48</td>
<td>0.398 0.244 0.223</td>
<td>1.93 2.42 5.75 3.17</td>
<td>4.52 5.07</td>
<td>0.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>Depth cm</th>
<th>Sand %</th>
<th>Silt %</th>
<th>Clay %</th>
<th>BD m$^{-3}$</th>
<th>$\theta_{10}$ m$^{-3}$</th>
<th>$\theta_{33}$ m$^{-3}$</th>
<th>PTF-estimated $K_{\text{sat}}$</th>
<th>Measured/Calibrated $K_{\text{sat}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 35</td>
<td>22.0</td>
<td>65.8</td>
<td>12.2</td>
<td>1.42</td>
<td>0.395 0.316 0.275</td>
<td>0.59 0.21 1.31 0.98</td>
<td>5.24 2.06</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>2 30</td>
<td>20.0</td>
<td>64.4</td>
<td>15.6</td>
<td>1.65</td>
<td>0.420 0.351 0.331</td>
<td>0.23 0.13 0.41 0.88</td>
<td>3.21 0.14</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>3 120</td>
<td>17.7</td>
<td>63.3</td>
<td>19.0</td>
<td>1.66</td>
<td>0.405 0.365 0.343</td>
<td>0.07 0.02 0.10 0.79</td>
<td>1.97 0.10</td>
<td>0.26</td>
</tr>
</tbody>
</table>

- BD, Bulk Density; $\theta_1$, total porosity; $\theta_{33}$, soil water content at 10 & 33 kPa matric potentials; K1, K2, K3 are the $K_{\text{sat}}$ values estimated by the three best performing PTFs of group 1; KK1, KK2, KK3 are the $K_{\text{sat}}$ values estimated by the three best performing PTFs of group 2; †, equivalent value of $K_{\text{sat}}$ for soil profile; ††, modeling efficiency of DRAINMOD predictions; *, depth from surface; Marked values are the best two predictions.

### 6.3.2 Assessing DRAINMOD Application with $K_{\text{sat}}$ Predicted by PTFs Group (1)

The results of DRAINMOD simulations using $K_{\text{sat}}$ values predicted by group 1 PTFs, which require only the effective porosity as input, are presented in Figures (6.1, 6.2) and Table (6.7). The model poorly predicted yearly subsurface drainage for all sites except for the North Carolina site (TES). DRAINMOD
predictions of drainage volumes were closely correlated with the transmisivity of the soil profile (defined as the sum of the product of $K_{sat}$ and thickness of each layer in the soil profile above the impervious layer), which is a direct function of estimated $K_{sat}$. Group 1 PTFs considerably under-estimated $K_{sat}$ values for the three U.S. Midwest sites (SEPAC, WQFS, and Story City) and as a result, DRAINMOD drastically under-predicted drainage volumes for these sites (Figure 6.1). According to the statistical performance measures listed in Table 6, all DRAINMOD predictions of yearly drainage volumes for the Midwest sites, which are based on $K_{sat}$ values estimated by group 1 PTFs, are considered unsatisfactory. Using $K_{sat}$ estimated by third ranked function of group 1 gave relatively better predictions of yearly drainage for the three sites, compared with DRAINMOD predictions based on $K_{sat}$ estimated by first and second ranked functions. The NRMSE in predicting yearly drainage by the calibrated model was in the range of 15-24% for the three U.S. Midwest sites. The corresponding modeling efficiencies ranged from 0.62 to 0.91. The use of $K_{sat}$ values estimated by the best performing function of group 1 PTFs resulted in a substantial increase in the NRMSEs in DRAINMOD predictions of yearly drainage (28 to 63%) and a substantial decrease in the corresponding EF values (-1.60 to 0.47). For the TES site, the use of $K_{sat}$ estimated by the first ranked function gave the closest predictions to measured drainage (NRMSE=15% and EF=0.63) and DRAINMOD performance was relatively comparable to the calibrated model for the site (NRMSE=9% and EF=0.88). The relatively better predictions of yearly drainage for the TES site are attributed to the relatively smaller difference between the profile transmissivities determined based on calibrated and group 1 PTFs-predicted $K_{sat}$ values.
Figure (6.1) Observed and predicted DRAINMOD outflows using $K_{\text{sat}}$ values predicted by PTFs group (1)

Table (6.7) Statistical comparison between observed and simulated drainage using $K_{\text{sat}}$ values predicted by PTFs group (1)

<table>
<thead>
<tr>
<th></th>
<th>SEPAC</th>
<th>WQFS</th>
<th>STORY CITY</th>
<th>TES (NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NRMSE</td>
<td>EF</td>
<td>NRMSE</td>
<td>EF</td>
</tr>
<tr>
<td>$D_c$</td>
<td>20.9</td>
<td>0.70</td>
<td>24.1</td>
<td>0.62</td>
</tr>
<tr>
<td>$D_{K1}$</td>
<td>40.3</td>
<td>-0.11</td>
<td>72.2</td>
<td>-2.37</td>
</tr>
<tr>
<td>$D_{K2}$</td>
<td>67.3</td>
<td>-2.10</td>
<td>83.4</td>
<td>-3.51</td>
</tr>
<tr>
<td>$D_{K3}$</td>
<td>27.9</td>
<td>0.47</td>
<td>63.4</td>
<td>-1.60</td>
</tr>
</tbody>
</table>
The soil textural class appears to be an important factor affecting the performance of group 1 PTFs and consequently DRAINMOD predictions which are based on K_{sat} estimated using these PTFs. Group 1 PTFs performed poorly for the three Midwest sites, which have fine to medium texture soils. These PTFs performed relatively better for the TES site, which has coarse texture soil.

6.3.3 Assessing DRAINMOD Application with K_{sat} Predicted by PTFs Group (2)

Figures (6.3, 6.4) and Table (6.8) summarize the results of the DRAINMOD simulations using K_{sat} values estimated by group 2 PTFs, which require the particle size distribution and bulk density as inputs. Generally, the use of K_{sat} values estimated by group 2 PTFs has greatly improved DRAINMOD predictions of yearly drainage volumes. The accuracy of DRAINMOD
predictions using $K_{\text{sat}}$ estimated by the first ranked function of group 2 PTFs were comparable to the accuracy of the calibrated model for the three US Midwest sites. For these three sites, the NRMSEs in yearly drainage predicted by the calibrated model were in the range of 15 to 24% with corresponding modeling efficiencies ranging from 0.62 to 0.91. The use of $K_{\text{sat}}$ estimated by the first ranked PTF of group 2 slightly increased the NRMSE (19-30%) and slightly decreased the EF (0.43-0.85). The second ranked PTF out performed the first ranked function for the WQFS site. For this site, the yearly drainage volumes predicted using the $K_{\text{sat}}$ estimated by the second ranked PTF were considerably closer to measured and predicted by the calibrated model than the drainage volumes predicted using the first ranked PTF. However, the use of $K_{\text{sat}}$ estimated by group 2 PTFs resulted in poor predictions of drain flow for the TES site. For this site, the NRMSE and EF for the calibrated model were 9% and 0.88, respectively. The corresponding NRMSE and EF for yearly drainage predicted by DRAINMOD using the best performing PTF among group 2 were 22% and 0.22. For this site, yearly drainage predicted using group 1 PTFs were more accurate than those based on group 2 PTFs.

Figure (6. 3) Observed and predicted DRAINMOD outflows using $K_{\text{sat}}$ values predicted by PTFs group (2)
6.3.4 Assessing DRAINMOD Application with SWC Data Predicted by PTFs

Results of DRAINMOD simulations using PTFs-predicted SWC data and measured/calibrated $K_{sat}$ values are presented in Figures (6.5, 6.6) and Table (6.9). The results showed that the SWC data has a relatively smaller effect on DRAINMOD predictions of yearly drainage than $K_{sat}$. All DRAINMOD simulations using SWC data predicted by PTFs underestimated the annual drainage rates for all sites except for the WQFS site when SWC data predicted by the first ranked PTF were used in the simulation (Figure 6.5). The use of the SWC data predicted by the first ranked PTF resulted in good agreement between the
simulated and observed yearly drainage for the four study sites. The statistical performance measures showed that the performance of DRAINMOD was comparable to the calibrated model, especially for the SEPAC and Story City sites (Table 8). DRAINMOD performance in predicting yearly drainage was directly related to the performance of different PTFs in predicting SWC. The most accurate predictions of yearly drainage for all sites are those based on SWC data estimated by the first ranked PTF. The use of SWC data estimated by the second ranked PTF resulted in the second accurate predicted drainage. The least accurate predictions of drainage volumes resulted from using SWC data estimated by the third ranked PTF.

Figure (6.5) Observed and predicted DRAINMOD outflows using SWC data predicted by PTFs

Table (6.9) Statistical comparison between observed and simulated drainage using SWC data predicted by PTFs

<table>
<thead>
<tr>
<th></th>
<th>SEPAC</th>
<th>WQFS</th>
<th>STORY CITY</th>
<th>TES (NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NRMSE</td>
<td>EF</td>
<td>NRMSE</td>
<td>NRMS</td>
</tr>
<tr>
<td>Do</td>
<td>20.9</td>
<td>0.7</td>
<td>24.1</td>
<td>0.6</td>
</tr>
<tr>
<td>Dθ1</td>
<td>24.0</td>
<td>0.6</td>
<td>31.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Dθ2</td>
<td>29.8</td>
<td>0.3</td>
<td>31.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Dθ3</td>
<td>36.0</td>
<td>0.1</td>
<td>34.5</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure (6.6) Modeling efficiency between Observed and predicted DRAINMOD outflows using SWC data predicted by PTFs

6.3.5 **Assessing DRAINMOD Application with $K_{sat}$ and SWC Data Predicted by PTFs**

These simulations were conducted to assess the overall effect of using both $K_{sat}$ and SWC data estimated by PTFs on the accuracy of DRAINMOD predictions of yearly drainage. This scenario represents the most likely case when DRAINMOD is to be applied at large scale. Three simulations were conducted using the best three performing PTFs for predicting $K_{sat}$ and SWC data for each of the four sites. The functions $\theta_1$, $\theta_2$, and $\theta_3$ (Table 6.3) are consistently ranked from the best to the least performing functions for predicting SWC data for all sites. The three best performing PTFs for predicting $K_{sat}$ varied depending on the site. For example, K1, K2, and KK1 are the best performing functions for the TES site. Likewise, KK1 is the best performing function for the SEPAC site, followed by KK3 and K3. The results of these simulations are presented in Figures (6.7, 6.8) and Table (6.10).

As expected, the use of both PTF-estimated $K_{sat}$ and SWC data as inputs to DRAINMOD resulted in less accurate predictions of yearly drainage, compared
with predicted drainage based on using only PTF-estimated \( K_{\text{sat}} \) or PTF-estimated SWC. For example, the NRMSE in predicting yearly drainage for the Story City was 19% (Table 6.8) when only \( K_{\text{sat}} \) estimated by the best performing function (KK1) was used, 16% (Table 6.9) when only SWC data estimated by the best performing function (D01) was used, and 26% (Table 6.10) when both same PTF-estimated \( K_{\text{sat}} \) and SWC data were used. The corresponding EF values for the three simulations were 0.85, 0.89, and 0.7, respectively. The same trend can be observed for the WQFS and TES sites. For the SEPAC site, the use of both \( K_{\text{sat}} \) and SWC data estimated by PTFs resulted in slightly better predictions of yearly drainage, compared with predicted drainage based on PTF-estimated SWC data only. Another important observation is that best performing PTF for estimating \( K_{\text{sat}} \) and the best performing PTF for estimating SWC are the only combination that resulted in reasonable DRAINMOD predictions of yearly drainage.

Figure (6.7) Observed and predicted DRAINMOD outflows using \( K_{\text{sat}} \) and SWC data predicted by PTFs
Table (6.10) Statistical comparison between observed and simulated drainage using $K_{\text{sat}}$ and SWC data predicted by PTFs

<table>
<thead>
<tr>
<th></th>
<th>SEPAC NRMS</th>
<th>SEPAC EF</th>
<th>WQFS NRMS</th>
<th>WQFS EF</th>
<th>STORY CITY NRMS</th>
<th>STORY CITY EF</th>
<th>TES (NC) NRMS</th>
<th>TES (NC) EF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_c$</td>
<td>20.9</td>
<td>0.70</td>
<td>24.1</td>
<td>0.62</td>
<td>14.8</td>
<td>0.91</td>
<td>8.6</td>
<td>0.88</td>
</tr>
<tr>
<td>$D_{K1}$</td>
<td>23.2</td>
<td>0.63</td>
<td>33.1</td>
<td>0.29</td>
<td>26.1</td>
<td>0.70</td>
<td>20.5</td>
<td>0.32</td>
</tr>
<tr>
<td>$D_{K2}$</td>
<td>45.8</td>
<td>-</td>
<td>49.8</td>
<td>-</td>
<td>64.2</td>
<td>-</td>
<td>22.8</td>
<td>0.16</td>
</tr>
<tr>
<td>$D_{K3}$</td>
<td>38.6</td>
<td>-</td>
<td>45.3</td>
<td>-</td>
<td>60.5</td>
<td>-</td>
<td>27.8</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure (6.8) Modeling efficiency between Observed and predicted DRAINMOD outflows using $K_{\text{sat}}$ and SWC data predicted by PTFs

6.3.6 Comparison between DRAINMOD predictions using $K_{\text{sat}}$ and SWC Data predicted by ROSETTA and PTFs

Table (6.11) shows the statistical measures of DRAINMOD simulations using $K_{\text{sat}}$ and SWC data predicted by ROSETTA (Schaap et al., 2001) and the PTFs of this study. ROSETTA is an artificial neural network model for predicting the soil properties using different levels of inputs (i.e. particle size distribution, bulk density, and water content) and was used by many researchers to feed DRAINMOD with the soil properties inputs (Qi et al., 2006; Salazar et al., 2008). ROSETTA was used to predict the $K_{\text{sat}}$ and SWC data based on four levels of
inputs; particle size distribution (ROSETTA-SSC); particle size distribution and
bulk density (ROSETTA-SSCBD); PSD, BD and water content at 33 kPa; PSD,
BD and two water contents at 33, 1500 kPa. DRAINMOD results using calibrated
parameters and K_{sat} and SWC data predicted by the most accurate PTFs were
compared to the results obtained using the four different sets of data predicted by
ROSETTA. The results indicated that the model performance when running using
input parameters predicted by PTFs is better than its performance when running
using input parameters predicted by ROSETTA. Except the SEPAC site,
DRAINMOD showed bad performance when running using input parameters
predicted by ROSETTA (EF range from -0.03 to -2.99). These results indicated
the importance of determining the best performing PTFs before running the
hydrologic model using predicted soil properties.

<table>
<thead>
<tr>
<th></th>
<th>SEPAC</th>
<th>WQFS</th>
<th>STORY CITY</th>
<th>TES (NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NRMSE</td>
<td>EF</td>
<td>NRMSE</td>
<td>EF</td>
</tr>
<tr>
<td>Dc</td>
<td>20.89</td>
<td>0.70</td>
<td>24.14</td>
<td>0.62</td>
</tr>
<tr>
<td>DKθ1</td>
<td>23.24</td>
<td>0.63</td>
<td>33.12</td>
<td>0.29</td>
</tr>
<tr>
<td>DRS1</td>
<td>27.69</td>
<td>0.47</td>
<td>68.57</td>
<td>-2.04</td>
</tr>
<tr>
<td>DRS2</td>
<td>33.49</td>
<td>0.23</td>
<td>70.43</td>
<td>-2.21</td>
</tr>
<tr>
<td>DRS3</td>
<td>28.27</td>
<td>0.45</td>
<td>62.64</td>
<td>-1.54</td>
</tr>
<tr>
<td>DRS4</td>
<td>26.73</td>
<td>0.51</td>
<td>71.69</td>
<td>-2.33</td>
</tr>
</tbody>
</table>

6.3.7 Comparison between DRAINMOD Predictions Using Calibrated and
PTF-Estimated K_{sat} and SWC Data on Monthly Basis

Table (6.12) presents the modeling efficiency of DRAINMOD simulated
monthly drainage for the four study sites using measured/calibrated and PTF-
estimated K_{sat} and SWC data. The simulated periods were six years for SEPAC
site (i.e. 1985-1990), six years for TES site (i.e. 1992-1997), seven years for
WQFS site (i.e. 1999-2005), and ten years for STORY CITY site (i.e. 1996-2005). The results showed a good agreement between model results for both scenarios. The overall efficiency considering the total simulated period showed a good performance for DRAINMOD when using PTF-estimated parameters for all sites. The poorer performance was at TES site (EF=0.49) while the best performance was at SEPAC (Spacing =5m), Story city sites (EF= 0.73, 0.72). In some simulated years, the model performance using PTF-estimated parameters was better than its performance using measured calibrated parameters (e.g. years 1, 2 for SEPAC site).

Table (6.12) Modeling efficiency of DRAINMOD simulated drainage using measured/calibrated and PTF-estimated $K_{sat}$ and SWC data on monthly basis

<table>
<thead>
<tr>
<th>Year</th>
<th>SEPAC (S=20)</th>
<th>SEPAC (S=10)</th>
<th>SEPAC (S=5)</th>
<th>WQFS</th>
<th>STORY CITY</th>
<th>TES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DC</td>
<td>DK01</td>
<td>DC</td>
<td>DK01</td>
<td>DC</td>
<td>DK01</td>
</tr>
<tr>
<td>Year 1</td>
<td>0.71</td>
<td>0.77</td>
<td>0.73</td>
<td>0.75</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td>Year 2</td>
<td>0.77</td>
<td>0.88</td>
<td>0.84</td>
<td>0.91</td>
<td>0.76</td>
<td>0.84</td>
</tr>
<tr>
<td>Year 3</td>
<td>0.70</td>
<td>0.67</td>
<td>0.67</td>
<td>0.64</td>
<td>0.65</td>
<td>0.52</td>
</tr>
<tr>
<td>Year 4</td>
<td>0.85</td>
<td>0.63</td>
<td>0.83</td>
<td>0.76</td>
<td>0.84</td>
<td>0.87</td>
</tr>
<tr>
<td>Year 5</td>
<td>0.94</td>
<td>0.53</td>
<td>0.87</td>
<td>0.60</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>Year 6</td>
<td>0.67</td>
<td>0.44</td>
<td>0.41</td>
<td>-0.19</td>
<td>0.44</td>
<td>0.36</td>
</tr>
<tr>
<td>Year 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Year 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.77</td>
<td>0.65</td>
<td>0.73</td>
<td>0.58</td>
<td>0.73</td>
<td>0.73</td>
</tr>
</tbody>
</table>

† Year 1 is 1985 for SEPAC site, 1999 for WQFS site, 1996 for STORY CITY site, 1992 for TES site
6.3.8 A Predictive analysis using DRAINMOD with PTF-estimated $K_{\text{sat}}$ and SWC Data

In the previous analysis, both measured drainage and drainage predicted by the calibrated model for the four sites were compared to drainage predicted by the model using PTF-estimated $K_{\text{sat}}$ and SWC data. The results of the analysis quantified the errors in DRAINMOD predictions of yearly drainage induced by using PTF-estimated $K_{\text{sat}}$ and SWC rather than the calibrated/measured values. The use of PTF-estimated $K_{\text{sat}}$ and SWC data resulted in a substantial increase in the NRMSEs by 11% for SEPAC, 37% for WQFS, 76% for Story City, and 138% for TES sites. This demonstrates the importance of the calibration and validation of DRAINMOD using field measured data. The previous analysis, however, does not assess the use of DRAINMOD with PTF-estimated $K_{\text{sat}}$ and SWC in a predictive mode, in which the model simulates different scenarios of land use and water management practices.

In this analysis, the effect of controlled drainage on reduction of yearly drainage volumes was assessed for the four sites using both the calibrated model and the model that uses PTF-estimated $K_{\text{sat}}$ and SWC data. The results of the simulations (Table 6.13, Figure 6.9) reveal an important finding. The use of PTF-estimated $K_{\text{sat}}$ and SWC resulted in an over-prediction/under-prediction of yearly average drainage by 15 to 21% for conventional drainage and by 7 to 23% for controlled drainage. Despite this relatively large difference in the predicted yearly average drainage for each drainage management scenario, the predicted annual reduction in drain flow, on a percentage basis, because of implementing controlled drainage at the four sites was very similar (0.8 to 6.2% difference in the percent reduction). Thus the errors in DRAINMOD predictions induced by the use of PTF-estimated soil hydraulic properties does not have a large impact on the accuracy of the model prediction of the system’s response (e.g. annual drainage rate) to different management scenarios (e.g. conventional and controlled drainage). These
results have an important implications regarding the large scale application of DRAINMOD model using the PTF-estimated soil hydraulic properties.

Table (6.13) DRAINMOD simulated overall drainage outflows with conventional and controlled drainage managements using calibrated and PTFs predicted soil parameters

<table>
<thead>
<tr>
<th></th>
<th>Conv.</th>
<th>Ctrl.</th>
<th>(Δ (%))</th>
<th>Conv.</th>
<th>Ctrl.</th>
<th>(Δ (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEPAC</td>
<td>15.6</td>
<td>12.0</td>
<td>23.1</td>
<td>18.1</td>
<td>12.8</td>
<td>33.2</td>
</tr>
<tr>
<td>WQFS</td>
<td>15.9</td>
<td>10.5</td>
<td>34.1</td>
<td>18.2</td>
<td>11.8</td>
<td>35.8</td>
</tr>
<tr>
<td>Story City</td>
<td>20.0</td>
<td>18.5</td>
<td>6.3</td>
<td>15.9</td>
<td>14.3</td>
<td>6.1</td>
</tr>
<tr>
<td>TES (NC)</td>
<td>43.9</td>
<td>38.5</td>
<td>12.2</td>
<td>52.0</td>
<td>46.0</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Figure (6.9) Percent of reduction in simulated outflows between conventional and controlled drainage managements using calibrated and PTFs predicted input parameters
6.4 Summary and Conclusion

The objective of this study was to assess the feasibility of running the drainage water management model, DRAINMOD, using soil hydraulic properties estimated by PTFs. In two previous studies, Abdelbaki et al., 2009a, b have identified the best performing PTFs for predicting $K_{\text{sat}}$ and SWCC by evaluating 24 PTFs for estimating $K_{\text{sat}}$ and 22 PTFs for predicting SWCC using U.S. soil database. DRAINMOD was used to simulate annual and monthly drainage rates for four US sites in which the model was previously calibrated and validated. We have conducted 17 model simulations for each study site using measured/calibrated and PTF-estimated soil hydraulic parameters. The first simulation represents the calibrated model for the site in which, measured/calibrated $K_{\text{sat}}$ and SWC data were used in the analysis. Then, six simulations have been conducted using measured/calibrated SWC data with $K_{\text{sat}}$ values estimated by two groups of PTFs; group1 was the best three PTFs for predicting $K_{\text{sat}}$ using the effective porosity as input, group2 was the best three PTFs for predicting $K_{\text{sat}}$ using particle size distribution and bulk density as inputs. Another six simulations have been conducted; three simulations using measured/calibrated $K_{\text{sat}}$ values with PTF-estimated SWC data estimated by the three best performing PTFs, three simulations using $K_{\text{sat}}$ and SWC data both estimated by the three best performing PTFs for each site. Lastly, four model simulations have been conducted using $K_{\text{sat}}$ and SWC data predicted by the PTF model, ROSETTA, using four different levels of inputs.

The results showed that the hydrological model, DRAINMOD, can be run using soil hydraulic parameters estimated by PTFs. However, the use of PTF-estimated parameters with DRAINMOD increase the deviation from the observed values compared to the model predictions using measured/calibrated inputs, especially, when the number of PTF-estimated parameters increased. Also, the results showed that DRAINMOD was more sensitive to $K_{\text{sat}}$ values than SWC data.
when one of them was estimated by PTFs. The use of $K_{\text{sat}}$ and SWC data both estimated by PTFs increase the error from the case that only one of the two parameters was estimated by PTFs. Compared with model predictions using soil hydraulic parameters estimated by ROSETTA, DRAINMOD predictions using PTF-estimated parameters were more accurate and close to the observations.
CHAPTER (7)

AN AUTOMATIC CALIBRATION FRAMEWORK FOR THE DRAINAGE WATER MANAGEMENT MODEL, DRAINMOD

7.1 Introduction

Model calibration is the process of adjusting model input parameters to minimize the error between model predictions and the observed data (Liu et al., 2009). Parameters adjustment can be done either manually or automatically by using computer optimization algorithms. In manual calibration, model parameters are adjusted by a trial-and-error matching between model predictions and observations (Madsen, 2000). In this case, the performance of the calibrated model and the time required for the calibration process are based on the experience of the hydrologist with the model. Manual calibration is time consuming and expensive especially with complex models, which have a large number of sensitive parameters. On the other hand, automatic calibration is an efficient and easy way to calibrate complex hydrologic models. Using automatic calibration, the goodness of fit and performance of the calibrated model will be optimized during a considerable short time compared to manual calibration process. The automatic calibration is done using optimization algorithms, which run the model thousands of times with different sets of inputs and assess the goodness of fit between model predictions and measurements according to an objective function and select the optimum solution. A considerable number of optimization algorithms are available (e.g. shuffled complex evolution (Duan et al., 1992); simulated annealing (Kirkpatrick et al., 1983; Monte Carlo Simulation, (MCS) (Bobba et al., 1995));

During the last decade, several studies have been conducted to use the optimization algorithms for automatic calibration of hydrological models such as
SWAT model (Eckhardt and Arnold, 2001; Muleta and Nicklow, 2005; Van Liew et al., 2005; Eckhardt et al., 2005; Lin and Radcliffe, 2005; Tolson and Shoemaker, 2005; Lakshmi et al., 2006; Confesor and Whittaker, 2007; Green and Griensven, 2008), NAM model (Madsen, 2000; Madsen and Kristensen, 2002, Liu, 2009), MIKE SHE model (Madsen, 2003), SAC-SMA and SNOW-17 Models (Gupta et al., 1999; Hogue et al., 2000; Aghnami and Hogue, 2005; Hogue et al., 2006), PIT model (Liu et al., 2007), and HSPF model (Doherty and Johnston, 2003; Kim et al., 2006).

DRAINMOD (Skaggs, 1980) is one of the widely used hydrological models for drained lands with high water table soils. In the last two decades, many researchers have extensively tested the model for different climatic conditions, soil types, and farming practices (e.g. Skaggs et al., 1981; Fouss et al., 1987; Sabbagh et al., 1993; Youssef et al., 2006; Thorp et al., 2009.). In these studies, the model was manually calibrated and validated against field measured water table depths and drainage outflows data. Wang et al., 2006a calibrated DRAINMOD for a study site in Indiana using automatic calibration by Monte Carlo simulation method. The study of Wang et al. was the only study to calibrate DRAINMOD using automatic calibration approach. The development of automatic calibration methods for hydrologic models like DRAINMOD has been an important advancement in environmental simulation modeling and studies have shown the benefits of automatic calibration in terms of saving human time (Ajami et al., 2004).

The goal of this study was to develop and test an automatic calibration framework for the hydrological model DRAINMOD using the Shuffled Complex Evolution-University of Arizona (SCE-UA) algorithm (Duan et al., 1992; Duan et al., 2003).
7.2 Materials and Methods

7.2.1 Optimization Algorithm (SCE-UA)

The optimization algorithm used in this study is the Shuffled Complex Evolution-University of Arizona (SCE-UA) is a global search algorithm for the optimization of a single objective function for up to 16 parameters. Detailed description of this algorithm can be found in (Duan et al., 1992). In brief, the procedure of this method can be summarized in the following steps. First, SCE-UA selects an initial population, which is number of parameters sets, by random sampling throughout the feasible parameters space for p parameters to be optimized. The population size is defined by estimating the number of complexes (group of parameter sets) and the number of parameter sets in each complex to be (2P+1). In this study, the number of calibrated parameter is nine and the number of complexes is eight and each complex has 19 parameter sets, then, the population size is 152 parameter sets. The second step is to run the calibrated model using these parameter sets and calculate the objective function between the model predictions and the observation for each simulation. Third, the population size is sorted according to the value of the objective function (i.e. the population is ranked from the minimum to the maximum value of the objective function). Fourth, The population is divided into several groups (complexes) in such a way that the ith complex contains every NGS*(K-1)+i ranked points, where NGS is the number of complexes, K=1,2,......NPG, NPG is the number of points in each complex Fifth, Each complex evolves independently using the simplex algorithm (Nelder and Mead, 1965) by developing new points that has less objective functions. Sixth, the complexes are periodically shuffled to form new complexes in order to share the gained information. Lastly, the algorithm test for the termination criteria, which is the condition to reach the global optimum either it is the value of the objective function or exceed the maximum number of simulations. If the termination criteria was not met, the previous steps will be repeated. this
algorithm searches over the whole parameter space and finds the global optimum with a success rate of 100% (Sorooshian et al. 1993).

SCE-UA has been widely used in watershed model calibration and other areas of hydrology such as soil erosion, subsurface hydrology, remote sensing and land surface modeling (Duan, 2003). It was generally found to be robust, effective and efficient (Duan, 2003). The SCE-UA has also been applied with success to SWAT model for the hydrologic parameters (Eckardt and Arnold, 2001) and hydrologic and water quality parameters (van Griensven et al., 2002). Also, SCE-UA has been tested and compared with other optimization algorithms (Table 1) and was found to be the best effective and efficient algorithm to get the global optimum in all studies.

Table (7.1) Some studies compared the performance of SCE-UA with other optimization algorithms for automatic calibration of different hydrologic models

<table>
<thead>
<tr>
<th>Calibrated model</th>
<th>Tested optimization algorithms</th>
<th>Best performance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seven well established test</td>
<td>SCE-UA, Controlled random search (CRS2), and Multi-start algorithm (MSX)</td>
<td>SCE-UA</td>
<td>Duan et al. (1992)</td>
</tr>
<tr>
<td>Sacramento model</td>
<td>SCE-UA, and Multi-start simplex (MSX)</td>
<td>SCE-UA</td>
<td>Sorooshian et al. (1993)</td>
</tr>
<tr>
<td>Tank model</td>
<td>SCE-UA, Genetic Algorithms (GA), and Simulated Annealing (SA)</td>
<td>SCE-UA</td>
<td>Cooper et al. (1997)</td>
</tr>
<tr>
<td>Surface inFiltration Baseflow (SFB) model</td>
<td>SCE-UA, Genetic Algorithms (GA), Multiple Random Starts, and Newton local searches</td>
<td>SCE-UA</td>
<td>Kuczera (1997)</td>
</tr>
<tr>
<td>Modified version of the SFB model</td>
<td>SCE-UA, Three-phase Simulated Annealing Algorithm (SA-SX)</td>
<td></td>
<td>Thyer et al. (1999)</td>
</tr>
<tr>
<td>Rainfall-Runoff model</td>
<td>SCE-UA, and Simple Genetic Algorithm</td>
<td></td>
<td>Ndiritu and Daniell (2001)</td>
</tr>
<tr>
<td>Rainfall-Runoff model</td>
<td>SCE-UA, Clustering and Simulated Annealing (SA), Knowledge-based expert system</td>
<td></td>
<td>Madsen and Kristensen (2002)</td>
</tr>
</tbody>
</table>
7.2.2 Description of DRAINMOD

DRAINMOD (Skaggs, 1980) is a field scale water management model developed to simulate the performance of drainage and water table management systems for shallow water table soils, and it has been widely used in the United States and worldwide over the last two decades. It conducts a water balance for soil column midway between two adjacent drains or ditches on a day-by-day, hour-by-hour basis and calculates infiltration, evapotranspiration (ET), subsurface drainage, surface runoff, subirrigation, deep seepage, water table depth on daily, monthly and yearly basis (Skaggs, 1999). DRAINMOD simulates different drainage management systems including conventional drainage, controlled drainage, subirrigation, and combined controlled drainage/subirrigation systems. The model has different types of inputs including soil input parameters (saturated hydraulic conductivity (Ksat), Soil Water Characteristic Curve (SWCC), climatic input parameters (e.g. rainfall, temperature, and evapotranspiration), and cropping system parameters (e.g. planting and harvesting dates, root depths). In the last three decades, the model has extensively been tested for different climatic conditions, soil types, and farming practices (e.g. Skaggs et al., 1981; Fouss et al., 1987; Sabbagh et al., 1993; Youssef et al., 2006). In these studies, the model was manually calibrated and validated against field measured water table and subsurface drain flow data.

DRAINMOD has recently been calibrated against field measured data collected from three U.S. agricultural drained sites; two sites are located in the U.S. Midwest (Wang et al., 2006a; Thorp et al., 2009) and one site is located in south eastern U.S. (Youssef et al., 2006). The model was calibrated automatically for the first site and manually for the other two sites. In this study, the model was recalibrated automatically using the global optimization algorithm (SCE-UA) and the model performance was compared with the previous calibrations.
7.2.3 Calibration Parameters

Calibration parameters are the input parameters of a calibrated model that are adjusted within a specific range in order to minimize the difference between the model predictions and the measurements through an objective function. These parameters are chosen according to its impact on the model outputs. Also, the availability and feasibility of measurement of these parameters are important factors.

Several studies have been conducted to identify the most sensitive parameters in DRAINMOD (Anderson et al., 1987; Workman and Skaggs, 1994; Haan and Skaggs, 2003; Wang et al., 2006b). In addition to these, the developer conducted extensive sensitivity analyses (Skaggs, 1980; Skaggs, 1982). Skaggs (1980) identified the most sensitive parameters to be the saturated hydraulic conductivity (Ksat), the soil-moisture characteristic (h-θ) curve, drained volume-water table depth (V-D) relationship, steady-state upward flux-water table depth (U-D) relationship, and the Green-Ampt infiltration parameters (A and B). Sabbagh and Fox (1999) evaluated the effects of the most sensitive parameters identified by Skaggs (1980) on predictions of DRAINMOD. Haan and Skaggs (2003) identified nine parameters to have the greatest impact on the model predictions. These parameters were the saturated hydraulic conductivity of the four layers (Ksat1, Ksat2, Ksat3, and Ksat4), saturated water content (θs), residual water content (θr), drainage coefficient (Dc), maximum surface storage (STMAX), and minimum air volume (MAV). Wang et al., 2006b chose eight parameters to be included in uncertainty analysis. These parameters were (Ksat1, Ksat2, Ksat3, θs1, θs2, θs3, STMAX), and vertical hydraulic conductivity of the restrictive layer (Kvert). In the current study we selected nine parameters to be included in the automatic calibration process. These parameters are (Ksat1, Ksat2, Ksat3, Ksat4, Ksat5, Kvert, Dc, STMAX, Hv) where Hv is the pizometric head of the restrictive layer. These parameters were selected for calibration due its impact on the model
predictions and their cost of measurements, especially with large scale applications.

7.2.4 Objective Functions and Evaluation Criteria

The automatic calibration process has been conducted for DRAINMOD by comparing the predicted drainage and/or water table depths with field measurements for three U.S. sites. The auto-calibration strategy was to minimize a multi-objective function based on the Absolute Percent Error (APE) Eqn. (1) between the predictions (Pi) and the observations (Oi) and the Nash–Sutcliffe efficiency (EF) (Nash and Sutcliffe, 1970) Eqn (2). The multi-objective function is an aggregated function (Fagg1, 2) (Eqn. 4,5) for the drainage and water table depth respectively. The aggregated functions combined the two statistical measures according to the methods of Madsen (2000) and Madsen and Kristensen (2002). The transformation constants 1 and 0.5 were chosen according to Wang et al., 2006a for differences in the magnitudes of the different measures so that both transformation functions have about the same influence on the aggregated objective function. The objective function was chosen according to Eqn. (6) according to the available measured data.

\[
APE = 100 \times \left[ \frac{\sum_{i=1}^{N} P_i - \sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} O_i} \right] \quad (7.1)
\]

\[
EF = \frac{\sum_{i=1}^{N} (O_i - \overline{O})^2 - \sum_{i=1}^{N} (P_i - O_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2} \quad (7.2)
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2} \quad (7.3)
\]

\[
F_{agg1} = [(1 - EF)^2 + (APE + 0.5)^2]^{1/2}_{DRN} \quad (7.4)
\]

\[
F_{agg2} = [(1 - EF)^2 + (APE + 0.5)^2]^{1/2}_{WT} \quad (7.5)
\]
The predicted subsurface drainage and water table depths using the values of the calibrated parameters in this study and the same predictions in the studies of Wang et al., 2006a; Youssef et al., 2006; Thorp et al., 2009 were compared with field measurements. The Root Mean Square Error (RMSE) Eqn. (3) and the Nash–Sutcliffe efficiency (EF) were used to compare the model predictions and the measurements. The comparison of the model performance in this study with other studies under the same conditions is to test the effectiveness and efficient of auto-calibration using SCE with other calibration methods.

### 7.2.5 Description of Study Sites

Measured subsurface drainage and/or water table depths data from three U.S. sites have been used in this study for calibrating DRAINMOD. The model was previously calibrated and validated in these sites. Table 2 summarizes the main features of the three sites (Tidewater Experiment Station (TES) site, eastern North Carolina; Story City site, central Iowa; and Southeast Purdue Agricultural Center (SEPAC) site, south eastern Indiana) including soil types, drainage systems, and crop rotations. For more detailed description of the sites, readers are referred to Youssef et al., 2006 for TES site, Thorp et al., 2009 for Story City site, and Wang et al., 2006a for SEPAC site.

<table>
<thead>
<tr>
<th>Site</th>
<th>No of years</th>
<th>Soil textural class</th>
<th>Drainage system Depth (cm)</th>
<th>Spacing (m)</th>
<th>Conv./Ctrl.</th>
<th>Crop rotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TES</td>
<td>6</td>
<td>Sandy loam</td>
<td>116</td>
<td>22.72</td>
<td>both</td>
<td>Corn-Wheat-Wheat-Soybean</td>
</tr>
<tr>
<td>STORY CITY</td>
<td>10</td>
<td>Silty clay loam</td>
<td>145</td>
<td>27.4</td>
<td>Conv.</td>
<td>Corn-Soybean</td>
</tr>
<tr>
<td>SEPAC</td>
<td>6</td>
<td>Silt loam</td>
<td>75</td>
<td>5,10,20</td>
<td>Conv.</td>
<td>Continuous Corn</td>
</tr>
</tbody>
</table>

Table (7. 2) Main features of study sites
7.3 Results and Discussions

7.3.1 Automatic Calibration vs. Manual Calibration for DRAINMOD

DRAINMOD has been automatically calibrated using SCE-UA algorithm for the TES and STORY CITY sites in which the model was recently calibrated and validated manually (Youssef et al., 2006; Thorp et al., 2009). The results of the calibrated parameters and model predicted subsurface drainage for the two calibration strategies are shown in Tables 3, and 4. Four parameters (Ksat1, Ksat2, Ksat3, and Ksat4) were calibrated for each plot in the TES site and seven parameters (Ksat1, Ksat2, Ksat3, Ksat4, Ksat5, Kvert, and STMAX) were calibrated for the STORY CITY site. The range of the calibrated parameters was identified by lower and upper limits according to its values in manual calibration. The value of the lower limit was 25% and the upper limit was 250% of the parameter value in manual calibration. For the two study sites, the model was automatically calibrated under the same conditions of the previous manual calibrations using the same measurements. Also, the same calibrated parameters were justified. The global optimum was reached after (10000-15000) model runs.

DRAINMOD predicted subsurface drainage outflows using the automatically and manually calibrated inputs were compared with field measurements using two statistical measures; modeling efficiency (EF) and root mean square error (RMSE). The results showed that the model performance is better in the case of automatic calibration (Table 4 and Figure 1). The modeling efficiency was increased for all sites. EF was increased from (0.74, 0.73, 0.76, 0.87, and 0.84) to (0.80, 0.78, 0.79, 0.87, and 0.89) respectively. Also, the RMSE was reduced by 2% to 17% due to using automatic calibration. The results indicated that the SCE-UA algorithm is more effective and efficient in locating the global optimum.
Table (7.3) Values of calibrated parameters for TRS and Story City sites under automatic and manual calibrations

<table>
<thead>
<tr>
<th>Site</th>
<th>Parameters</th>
<th>Range</th>
<th>Automatic Calibration by SCE-</th>
<th>Manual Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>From</td>
<td>To</td>
<td></td>
</tr>
<tr>
<td>TES-Plot2</td>
<td>$K_{sat1}$</td>
<td>0.89</td>
<td>8.90</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>$K_{sat2}$</td>
<td>0.51</td>
<td>5.10</td>
<td>3.35</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.34</td>
<td>3.35</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>$K_{sat4}$</td>
<td>0.12</td>
<td>1.18</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>$K_{sat1}$</td>
<td>0.85</td>
<td>8.45</td>
<td>2.93</td>
</tr>
<tr>
<td>TES-Plot3</td>
<td>$K_{sat2}$</td>
<td>0.39</td>
<td>3.90</td>
<td>3.84</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.24</td>
<td>2.38</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>$K_{sat4}$</td>
<td>0.11</td>
<td>1.10</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>$K_{sat1}$</td>
<td>0.99</td>
<td>9.90</td>
<td>3.65</td>
</tr>
<tr>
<td>TES-Plot4</td>
<td>$K_{sat2}$</td>
<td>0.65</td>
<td>6.50</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.43</td>
<td>4.25</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>$K_{sat4}$</td>
<td>0.11</td>
<td>1.13</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$K_{sat1}$</td>
<td>0.96</td>
<td>9.63</td>
<td>2.31</td>
</tr>
<tr>
<td>TES-Plot5</td>
<td>$K_{sat2}$</td>
<td>0.47</td>
<td>4.70</td>
<td>3.18</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.26</td>
<td>2.55</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>$K_{sat4}$</td>
<td>0.11</td>
<td>1.05</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>$K_{sat1}$</td>
<td>0.88</td>
<td>8.75</td>
<td>4.32</td>
</tr>
<tr>
<td></td>
<td>$K_{sat2}$</td>
<td>0.88</td>
<td>8.75</td>
<td>5.92</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.88</td>
<td>8.75</td>
<td>6.58</td>
</tr>
<tr>
<td>Story City</td>
<td>$K_{sat4}$</td>
<td>0.88</td>
<td>8.75</td>
<td>4.01</td>
</tr>
<tr>
<td></td>
<td>$K_{sat5}$</td>
<td>0.88</td>
<td>8.75</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>$K_{vert}$</td>
<td>0.00015</td>
<td>0.0015</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>STMAX</td>
<td>0.13</td>
<td>1.25</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Table (7.4) Comparison between DRAINMOD performance under automatic and manual calibrations

<table>
<thead>
<tr>
<th>Site</th>
<th>Automatic Calibration</th>
<th>Manual Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EF</td>
<td>RMSE</td>
</tr>
<tr>
<td>TES-Plot2</td>
<td>0.80</td>
<td>1.19</td>
</tr>
<tr>
<td>TES-Plot3</td>
<td>0.78</td>
<td>1.32</td>
</tr>
<tr>
<td>TES-Plot4</td>
<td>0.79</td>
<td>1.44</td>
</tr>
<tr>
<td>TES-Plot5</td>
<td>0.87</td>
<td>1.17</td>
</tr>
<tr>
<td>Story City</td>
<td>0.89</td>
<td>0.96</td>
</tr>
</tbody>
</table>
7.3.2 Comparison between Two Automatic calibration Methods for DRAINMOD

Wang et al. (2006a) calibrated DRAINMOD for SEPAC site using the Monte Carlo Simulation (MCS) method. They calibrated eight hydrologic parameters ($\theta_{s1}$, $\theta_{s2}$, $\theta_{s3}$, Ksat1, Ksat2, Ksat3, Kvert, and STMAX) using measured subsurface drainage for the two years 1988, and 1989. In this study, DRAINMOD has been automatically calibrated using SCE-UA algorithm for the same. Table 5 shows the results of the calibrated parameters and their ranges for the two calibration processes. Under the same conditions, five hydrologic parameters (Ksat1, Ksat2, Ksat3, Kvert, and STMAX) were calibrated for the west site (W20) and east site (E20) for 20 m drain spacing. The two sites were calibrated using drainage outflows measured for the period from January, 1, 1988 to December, 31, 1989.
Table (7.5) Calibration parameters values for SEPAC site under two approaches of auto-calibration

<table>
<thead>
<tr>
<th>Site</th>
<th>Parameters</th>
<th>Range</th>
<th>Automatic Calibration by SCE-UA</th>
<th>Automatic Calibration by MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>From</td>
<td>To</td>
<td></td>
</tr>
<tr>
<td>W20</td>
<td>$K_{sat1}$</td>
<td>0.03</td>
<td>2</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>$K_{sat2}$</td>
<td>0.03</td>
<td>0.65</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.05</td>
<td>0.67</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>$K_{vert}$</td>
<td>0.0005</td>
<td>0.003</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>STMAX</td>
<td>0.3</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>E20</td>
<td>$K_{sat1}$</td>
<td>0.03</td>
<td>2</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>$K_{sat2}$</td>
<td>0.03</td>
<td>0.65</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$K_{sat3}$</td>
<td>0.05</td>
<td>0.67</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>$K_{vert}$</td>
<td>0.0005</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>STMAX</td>
<td>0.3</td>
<td>1</td>
<td>0.434</td>
</tr>
</tbody>
</table>

DRAINMOD predicted subsurface drainage using the calibrated parameters obtained from the two methods were compared to field measurements from the west and east sites for three different drain spacing (20m, 10m, and 5m) for the period from 1985-1990 and the results are presented in Table (7.6) and Figure (7.2). Two statistical measures (EF, RMSE) were used in the analysis. The results indicated that DRAINMOD performance is better in the case of automatic calibration using SCE-UA algorithm. Excluding the E10 simulation, the modeling efficiency was modified from (0.86, 0.79, 0.78, 0.75, 0.57, and 0.66) to (0.89, 0.82, 0.82, 0.80, 0.45, and 0.79). Also, the RMSE was reduced by 10% to 20% with automatic calibration using the SCE-UA algorithm. The slight enhancement in the model performance because the two calibrations were conducted using an automatic techniques. The results indicated that the SCE-UA algorithm is more effective and efficient in locating the global optimum for DRAINMOD.
Table (7. 6) DRAINMOD performances under two automatic calibration methods

<table>
<thead>
<tr>
<th>Site</th>
<th>Automatic Calibration by SCE-UA</th>
<th>Automatic Calibration by MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EF</td>
<td>RMSE</td>
</tr>
<tr>
<td>W20</td>
<td>0.89</td>
<td>0.59</td>
</tr>
<tr>
<td>W10</td>
<td>0.82</td>
<td>1.05</td>
</tr>
<tr>
<td>W5</td>
<td>0.82</td>
<td>1.21</td>
</tr>
<tr>
<td>E20</td>
<td>0.80</td>
<td>0.52</td>
</tr>
<tr>
<td>E10</td>
<td>0.45</td>
<td>0.93</td>
</tr>
<tr>
<td>E5</td>
<td>0.79</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Figure (7. 2) Predicted vs. observed annual drainage for DRAINMOD automatically calibrated using MCS and SCE-UA algorithms for the SEPAC site

7.3.3 Sensitivity of Results to Population Size

Population size is the number of model simulations that made by the optimization algorithm during the automatic calibration process to reach the termination criteria. This number is defined by the model user according to the desired model performance and the time availability. The optimization algorithm can reach the global optimum (optimum set of calibrated parameters) even before reaching the population size if the desired objective function was achieved. To study the sensitivity of the calibration results to population size, we conducted a
series of optimization runs in which all the conditions were identical, except that the population size was varied as 500, 1000, 2000, 5000, 10000, and 20000 runs. Table (7.7) shows the results of DRAINMOD performance (EF, RMSE) for the six optimization runs. The results showed that the performance of the calibrated model was increased with the increase of the population size. Compared to the manual calibration in TES and STORY CITY sites and previous automatic calibration in SEPAC sites, the current automatic calibration using SCE-UA algorithm shows better results for all simulated years (with the exception of the SEPAC (E10) site). The optimization algorithm reaches the global optimum after the 10000 runs for the TES and SEPAC (W20) sites, after 5000 runs for the STORY CITY site, and after 2000 runs for the SEPAC (E20) site. The minimum population size (500 simulations) shows the poorer model performance, however, it shows better model performance than the previous studies (Youssef et al., 2006; Thorp et al., 2009; Wang et al., 2006a). The average percent reduction in RMSE was larger for the TES site (9%) and STORY SITY site (13%) in which the model was previously manually calibrated than the SEPAC site (7%) in which the model was previously automatically calibrated. The authors recommending the model users to set the population size at 10000 model simulations to achieve the global optimum within acceptable period of time. The time consuming in the calibration process depends on the population size and the time of each model simulation, which depends on the number of the simulated years.
Table (7.7) Change in DRAINMOD performance with the increase of population size of SCE-UA

<table>
<thead>
<tr>
<th>Site</th>
<th>Manual Cali.</th>
<th>Automatic calibration by SCE-UA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Years EF RMSEE</td>
<td>EF 500 1000 2000 5000 10000 20000</td>
</tr>
<tr>
<td>TES</td>
<td>1992 0.79 1.43</td>
<td>0.82 0.82 0.82 0.82 0.83 0.83</td>
</tr>
<tr>
<td></td>
<td>1993 0.93 1.14</td>
<td>0.93 0.93 0.93 0.93 0.93 0.93</td>
</tr>
<tr>
<td></td>
<td>1994 0.26 1.21</td>
<td>0.38 0.38 0.40 0.40 0.41 0.41</td>
</tr>
<tr>
<td></td>
<td>1996 0.84 1.24</td>
<td>0.84 0.84 0.84 0.84 0.84 0.84</td>
</tr>
<tr>
<td>STORY CITY</td>
<td>1997 0.22 1.23</td>
<td>0.23 0.24 0.25 0.36 0.36 0.36</td>
</tr>
<tr>
<td></td>
<td>1998 0.83 1.49</td>
<td>0.88 0.88 0.88 0.89 0.89 0.89</td>
</tr>
<tr>
<td></td>
<td>1999 0.97 0.72</td>
<td>0.96 0.96 0.97 0.97 0.97 0.97</td>
</tr>
<tr>
<td></td>
<td>2000 -125 0.26</td>
<td>-204 -191 -343 -106 -106 -106</td>
</tr>
<tr>
<td></td>
<td>2001 0.53 1.50</td>
<td>0.74 0.74 0.58 0.84 0.84 0.84</td>
</tr>
<tr>
<td>SEPEC</td>
<td>W20 0.86 0.69</td>
<td>0.90 0.90 0.90 0.90 0.90 0.90</td>
</tr>
<tr>
<td></td>
<td>W10 0.79 1.14</td>
<td>0.81 0.81 0.81 0.81 0.82 0.82</td>
</tr>
<tr>
<td></td>
<td>W5 0.78 1.35</td>
<td>0.81 0.81 0.81 0.81 0.82 0.82</td>
</tr>
<tr>
<td></td>
<td>E20 0.75 0.59</td>
<td>0.79 0.79 0.80 0.80 0.80 0.80</td>
</tr>
<tr>
<td></td>
<td>E10 0.58 0.81</td>
<td>0.44 0.44 0.45 0.45 0.45 0.45</td>
</tr>
<tr>
<td></td>
<td>E5 0.66 1.52</td>
<td>0.78 0.78 0.79 0.79 0.79 0.79</td>
</tr>
</tbody>
</table>

7.4 Conclusion

The water management model DRAINMOD (Skaggs, 1980) was automatically calibrated using the Shuffle Complex Evolution – University of Arizona (SCE-UA) algorithm (Duan et al., 1992). Measured drainage and water table depths from three U.S. sites were used in the analysis. The model has been recently calibrated and validated in these sites. In two sites (Tide Water Experimental Station (TES) from North Carolina and STORY CITY from Iowa), the model was manually calibrated (Youssef et al., 2006; Thorp et al., 2009) and in the third site (SEPAC from Indiana), the model was automatically calibrated (Wang et al., 2006a). The model was automatically calibrated in these three sites using SCE-UA algorithm and the results were compared with the results of (Youssef et al., 2006; Thorp et al., 2009; Wang et al., 2006a). The results showed that the SCE algorithm is more effective and efficient in locating the global optimum. Also, the enhancing in the model performance when shifting from manual calibration to automatic calibration is
greater than the enhancing when shifting from automatic calibration using MCS to automatic calibration using SCE. For the NC and IA sites, the EF was modified from (0.74, 0.73, 0.76, 0.87, and 0.84) to (0.80, 0.78, 0.79, 0.87, and 0.89) and for the IN site the EF was modified from (0.86, 0.79, 0.78, 0.75, 0.57, and 0.66) to (0.89, 0.82, 0.82, 0.80, 0.45, and 0.79). Results of this study will reduce the effort and time consuming in model calibration and make it easy to get hydrologically sound model.
CHAPTER (8)

CONCLUSION AND RECOMMENDATIONS

8.1 Main Results

The work in this study was to assess the feasibility of using Pedotransfer functions (PTFs) and automatic calibration methods for estimating the soil hydraulic properties like saturated hydraulic conductivity and soil water characteristics curve, which are considered to be the most important inputs to the DRAINMOD model. A big number of PTFs was found in the literature. To identify the best performing PTFs, two evaluation studies for PTFs of $K_{sat}$ and SWCC for U.S. soil databases. Using the best performing PTFs, DRAINMOD was used to predict the subsurface outflows using soil properties estimated by PTFs and using measured/calibrated soil parameters. Finally, the automatic calibration method was used for calibrating DRAINMOD and the results were compared to the results obtained from manual calibration. The main results of the study as well as the recommendation for future research are presented below.

8.1.1 Best Performing PTFs for Predicting Saturated Hydraulic Conductivity

Twenty four Pedotransfer function for predicting saturated hydraulic conductivity were evaluated in this study. The results showed that the PTFs developed by Suleiman and Ritchie (2001) and Minasny and McBratney (2000) are the best models to estimate $K_{sat}$ when the available measurements is the effective porosity only. If more measurements like particle size distribution, bulk density and porosity are available; the PTF developed by Cosby et al. (1984) is the best model for predicting $K_{sat}$. By dividing the dataset into four texture classes, the Rosetta SSC (Schaap et al., 2001) showed the best performance in very fine - fine class.
PTFs by Cosby et al. (1984) and Rosetta SSC are the best models for predicting $K_{\text{sat}}$ in fine medium soils. In medium class soil, PTF of Cosby et al. (1984) has the best performance. In coarse soils, PTFs by Puckett at al. (1985), Julia et al. (2004) and Cosby et al. (1984) are the best models for predicting $K_{\text{sat}}$. If more soil properties measurements are available like organic matter in addition to particle size distribution and bulk density, the PTF developed by Nemes et al. (2005) is the best model for predicting $K_{\text{sat}}$. Finally if all soil properties measurements are available in the soil database, the PTFs developed by Cosby et al. (1984), Nemes et al. (2005), Saxton et al. (1986), Saxton and Rawls (2006), and Julia et al. (2004) are the best models for predicting $K_{\text{sat}}$ respectively.

8.1.2 Best Performing PTFs for Predicting Soil Water Characteristic Curve

Twenty two Pedotransfer functions for predicting soil water characteristics curve were evaluated in this study. The results showed that the model developed by Rawls et al. (1982) showed the best performance among the five point PTFs followed by the PTF developed by Adhikary et al. (2008). The PTFs developed by Cosby et al. (1984), Brakensiek and Rawls (1985), Mayer and Jarvis (1999), and Schaap et al. (2001) (Rosetta SSC) showed the best performance among the seventeen continuous PTFs considered in this study. Also, the performance of the continuous PTFs is almost better than the performance of the point PTFs due to its continuous results. Another conclusion, the performance of PTFs changes with the variation of soil texture classes and with the variation of the potential heads. The PTFs that showed a good performance when evaluated using the entire dataset are almost the PTFs that showed the good performance in different texture classes and in different matric potentials. The results indicated that the PTFs perform better with the higher potential heads and this could be due to the little error in measurements of water content at these pressures.
8.1.3 Application of DRAINMOD Using Soil Properties Estimated by PTFs

DRAINMOD was applied to predict the subsurface drainage using measured/calibrated and PTF-estimated soil hydraulic properties. The results showed that DRAINMOD is able to simulate the subsurface drainage with saturated hydraulic conductivity and/or soil water characteristics data predicted from PTFs. For all sites, statistical comparisons showed excellent agreement between the simulated and observed outflows when SWC data predicted from PTFs were used. Also, DRAINMOD showed good performance with $K_{sat}$ values predicted from PTFs. Applying DRAINMOD with $K_{sat}$ values and SWC data both predicted from PTFs, the model showed good results compared to the observation. The deviation of the simulated values from the observed values is increased when both $K_{sat}$ and SWC data predicted from PTFs. The conclusion is DRAINMOD can successfully running using PTFs predicted $K_{sat}$ and SWC data instead of measured soil parameters.

8.1.4 Automatic Calibration of DRAINMOD

The water management model DRAINMOD (Skaggs, 1980) was automatically calibrated using the Shuffle Complex Evolution – University of Arizona (SCE-UA) algorithm (Duan et al., 1992). Measured drainage and water table depths from three U.S. sites were used in the analysis. The model has been recently calibrated and validated in these sites; in two sites (Tide Water Experimental Station (TES) from North Carolina and STORY CITY from Iowa) the model was manually calibrated (Youssef et al., 2006; Thorp et al., 2009) and in the third site (SEPAC from Indiana) the model was automatically calibrated (Wang et al., 2006a). The model was automatically calibrated in these three sites using SCE-UA algorithm and the result were compared with the results of (Youssef et al., 2006; Thorp et al., 2009; Wang et al., 2006a). The results showed that the SCE algorithm is more effective and efficient in locating the global optimum. Also, the
enhancing in the model performance when shifting from manual calibration to automatic calibration is greater than the enhancing when shifting from automatic calibration using MCS to automatic calibration using SCE. For the NC and IA sites, the EF was modified from (0.74, 0.73, 0.76, 0.87, and 0.84) to (0.80, 0.78, 0.79, 0.87, and 0.89) and for the IN site the EF was modified from (0.86, 0.79, 0.78, 0.75, 0.57, and 0.66) to (0.89, 0.82, 0.82, 0.80, 0.45, and 0.79). Results of this study will reduce the effort and time consuming in model calibration and make it easy to get hydrologically sound model.

8.2 Recommendations and Future Research

8.2.1 Including of Recently Developed PTFs

In this study, we evaluated twenty four PTFs for predicting saturated hydraulic conductivity and twenty two PTFs for predicting soil water characteristics curve. This number of PTFs does not represent the total number of available PTFs, but it represents the number of PTFs, which require input parameters (i.e. particle size distribution, bulk density, porosity, and organic matter content) that are simple and available in the soil databases. There are other types of PTFs, which require more difficult and not always available inputs such as PH number, gypsum and phosphorous contents are not considered in this study. In addition, Pedotransfer functions are a hot topic in modern science and development of new PTFs is continuous. Our recommendation is to include these newly developed PTFs into the evaluated PTFs of this study and update the list of best performing models. Also, the PTFs that require more difficult or unavailable inputs can be included whenever their input requirements are available in the soil databases.

8.2.2 Evaluation of PTFs for non U.S. Soils

The evaluated PTFs in this study were tested using U.S. soil databases due to its availability. As we mentioned before, the performance of PTFs is changing if it
were applied on different database from which used in its development. The application of the evaluated PTFs for non U.S. soil is limited due to the difference in the soil properties, which is the main input to the PTFs. Our recommendation to apply the result of this study for non U.S. soils is reevaluation of PTFs using the new soil database and identifying the best performing functions. For other countries (e.g. Egypt), which have lack of measured soil databases, we recommend to construct databases for these countries by more soil survey and measurements and collect small data sets in one soil database and make it available for researchers.

8.2.3 Application of Other Hydrologic Models Using Soil Properties Estimated by PTFs

We tested the feasibility of running DRAINMOD using soil hydraulic properties estimated by PTFs and the model showed good performance compared to its predictions using measured/calibrated inputs. The concept of using PTF-estimated inputs with DRAINMOD can be repeated with other hydrologic models (SWAT, RZWQ models), which require the same inputs of DRAINMOD.
REFERENCES


187


APPENDIX (1)

COMPUTER PROGRAM FOR SHUFFLED COMPLEX EVOLUTION ALGORITHM FOR AUTOMATIC CALIBRATION OF DRAINMOD

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                  AUTOMATIC CALIBRATION OF DRAINMOD
% METHOD: SHUFFLED COMPLEX EVOLUTION ALGORITHM (Duan et al., 2004)
% BY
%                          AHMED MOHAMED ABDELBAKI
%                         FAYOUm UNIVERSITY, EGYPT
% VISITING SCHOLAR IN BAE DEPARTMENT
% (NCSU, USA) AND PHD STUDENT (CAIRO UNIVERSITY, EGYPT)
% AND MOHAMED A. YOUSSEF
%ASSISTANT PROFESOUR IN BAE DEPARTMENT (NCSU, RALEIGH, NC, USA)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [bestx,bestf] =
sceua(x0,bl,bu,maxn,kstop,pcento,peps,ngs,iseed,iniflg)

% Definition:
% x0 = the initial parameter array at the start;
% = the optimized parameter array at the end;
% f0 = objective function corresponding to initial parameters.
% = objective function corresponding to optimized paramet.
% bl = the lower bound of the parameters
% bu = the upper bound of the parameters
% iseed = the random seed number (for repetitive testing purpose)
% iniflg = flag for initial param. array (=1, included it in
%initial population; otherwise, not included)
% ngs = number of complexes (sub-populations)
% npg = number of members in a complex
% nps = number of members in a simplex
% nspl = number of evolution steps before shuffling
% mings = minimum number of complexes required during the
optimization process
% maxn = maximum number of function evaluations allowed during
optimization
% kstop = maximum number of evolution loops before convergence
% percento = % change allowed in kstop loops before convergence

% LIST OF LOCAL VARIABLES
% x(.,.) = coordinates of points in the population
% xf(.) = function values of x(.,.)
% xx(.) = coordinates of a single point in x
% cx(.,.) = coordinates of points in a complex
Global BESTX BESTF ICALL PX PF
% THE CALIBRATED PARAMETERS ARE THE SATURATED HYDRAULIC
CONDUCTIVITY %FOR THE FIVE LAYERS, VERTICAL CONDUCTIVITY OF THE
RESTRICTIVE LAYER, %DRAINAGE COEFFICIENT, MAXIMUM SURFACE
STORAGE, AND THE PIZOMETRIC %HEAD OF RESTRICTIVE LAYER i.e.
% (K_{sat}1--K_{sat}2--K_{sat}3--K_{sat}4--K_{sat}5--Kv--Dc--STmax--Hv)
% Initialize SCE parameters:
bl=[0.03 0.03 0.05 0 0 0.0005 2 0.3 0];
bu=[2 0.65 0.67 0 0 0.003 2 1 0];
Xo=0.5*(bl+bu);
Bound = bu-bl;
nopt=length(Xo);
npg=2*nopt+1;
nps=nopt+1;
nspl=npg;
ngs=8;
mings=ngs;
npt=npg*ngs;
iseed=5;
maxn= 20000;
Peps=0.05;
kstop=50;
pcento=1e+3;
% Create an initial population to fill array x(npt,nopt):

rand ('seed',iseed);
x=zeros (npt,nopt);
For i=1:npt;
    x(i,:)=bl+rand(1,nopt).*bound;
End;

nloop=0;
icall=0;
For i=1:npt;

xf(i) = functn(nopt,x(i,:));
icall = icall + 1;
End;
f0=xf(1);

% Sort the population in order of increasing function values;
[xf,idx]=sort (xf);
x=x (idx,:);

% Record the best and worst points;
bestx=x(1,:); bestf=xf(1);
worstx=x(npt,:); worstf=xf(npt);
BESTF=bestf; BESTX=bestx; ICALL=icall;

% Compute the standard deviation for each parameter
Xnstd=STD(x);

% computes the normalized geometric range of the parameters
For n=1:9;
    If bound (1, n) ==0;
       Gnrng=0.7;
    Else
       Gnrng=exp (mean (log ((max(x)-min(x)). /bound)));
    End
End

Disp ('the Initial Loop: 0');
Disp ('BESTF: ' num2str (bestf));
Disp ('BESTX: ' num2str (bestx));
Disp ('WORSTF: ' num2str (worstf));
Disp ('WORSTX: ' num2str (worstx));
Disp (' ');

% Check for convergence;
If icall >= maxn;
    disp('*** OPTIMIZATION SEARCH TERMINATED BECAUSE THE LIMIT');
    disp('ON THE MAXIMUM NUMBER OF TRIALS ');
    disp(maxn);
    disp('HAS BEEN EXCEEDED. SEARCH WAS STOPPED AT TRIAL
NUMBER: ');
    disp(icall);
    disp('OF THE INITIAL LOOP!');
End;

If gnrng < peps;
    disp('THE POPULATION HAS CONVERGED TO A PRESPECIFIED SMALL
PARAMETER SPACE');
End;

% Begin evolution loops:
nloop = 0;

criter=[];
criter_change=1e+5;

While icall<maxn & gnrng>peps & criter_change>pcento;
   Nloop=nloop+1;

% Loop on complexes (sub-populations);
   For igs = 1: ngs;

      % Partition the population into complexes (sub-populations);
         k1=1: npg;
         k2= (k1-1)*ngs+igs;
         Cx (k1, :) = x (k2, :);
         cf (k1) = xf(k2);

% evolves sub-population igs for nspl steps:
   For loop=1: nspl;

% Select simplex by sampling the complex according to a linear
% probability distribution
   lcs(1) = 1;
   For k3=2: nps;
       For iter=1:1000;
           lpos = 1 + floor(npg+0.5-sqrt((npg+0.5)^2 - npg*(npg+1)*rand));
           idx=find(lcs(1:k3-1)==lpos, 1); if isempty(idx); break; end;
           lcs(k3) = lpos;
       End;
       lcs=sort(lcs);
   End;

% Construct the simplex:
   s = zeros (nps,nopt);
   s=cx(lcs,:); sf = cf(lcs);

   [snew,fnew,icall]=cceua(s,sf,bl,bu,icall,maxn);

% Replace the worst point in Simplex with the new point:
   s(nps,:) = snew; sf(nps) = fnew;

% Replace the simplex into the complex;
   cx(lcs,:) = s;
   cf(lcs) = sf;

% Sort the complex;
   [cf,idx] = sort (cf); cx=cx(idx,:);

% End of Inner Loop for Competitive Evolution of Simplexes
End;
% Replace the complex back into the population;
x (k2,:) = cx(k1,:); xf(k2) = cf(k1);

% End of Loop on Complex Evolution;
End;

% shuffled the complexes;
[xf,idx] = sort(xf); x=x(idx,:);
PX=x; PF=xf;

% Record the best and worst points;
bestx=x(1,:); bestf=xf(1);
worstx=x(npt,:); worstf=xf(npt);
BESTX=[BESTX;bestx]; BESTF=[BESTF;bestf]; ICALL=[ICALL;icall];

% Compute the standard deviation for each parameter
xnstd=std(x);

% computes the normalized geometric range of the parameters
For n=1:9;
    If bound(1,n)==0;
        gnrng=0.7;
    Else
        gnrng=exp(mean(log((max(x)-min(x))./bound)));
    End
End
disp(['Evolution Loop: ' num2str(nloop) ' - Trial - ' num2str(icall)]);
disp(['BESTF  : ' num2str(bestf)]);
disp(['BESTX  : ' num2str(bestx)]);
disp(['WORSTF : ' num2str(worstf)]);
disp(['WORSTX : ' num2str(worstx)]);
disp(' ');

% Check for convergence;
If icall >= maxn;
    disp('*** OPTIMIZATION SEARCH TERMINATED BECAUSE THE LIMIT');
disp(['ON THE MAXIMUM NUMBER OF TRIALS ' num2str(maxn) ' HAS BEEN EXCEEDED!']);
End;

If gnrng < peps;
    disp('THE POPULATION HAS CONVERGED TO A PRESPECIFIED SMALL PARAMETER SPACE');
End;
criter=[criter;bestf];
If (nloop >= kstop);
criter_change=abs (criter(nloop)-criter(nloop-kstop+1))*100;

criter_change=criter_change/mean(abs(criter(nloop-kstop+1:nloop)));

if criter_change < pcento;
    disp(['THE BEST POINT HAS IMPROVED IN LAST ' num2str(kstop)'
        LOOPS BY ', '...LESS THAN THE THRESHOLD ' num2str(pcento) '
    '])
    disp('CONVERGENCY HAS ACHIEVED BASED ON OBJ. FUNCTION
    CRITERIA!!!')
end;

% End of the Outer Loops
end;
end;
end;

disp(['SEARCH WAS STOPPED AT TRIAL NUMBER: ' num2str (icall)])
end;

disp(['NORMALIZED GEOMETRIC RANGE = ' num2str (gnrng)])
end;

disp(['THE BEST POINT HAS IMPROVED IN LAST ' num2str (kstop)'
        LOOPS BY ', num2str (criter_change) '
    ']);
end;

% END of Subroutine sceua
return;

%***************************************************************************

% SUBROUTINE TO CALCULATE THE OBJECTIVE FUNCTION BETWEEN %MODEL
PREDICTIONS (WT DEPTH AND SUBSURFACE DRAINAGE) AND OBSERVATIONS
Function f=functn(nopt,x)
%--- for the input parameters in a file
Fid=fopen ('inputparam.txt','w');
for k=1: nopt;
    fprintf (fid,'%5.3f   ', x (k));
end;

Fclose (fid)
String = 'DMHYDRO.exe C:\DrainMod\inputs\hyd85_93.prj';
Dos (string)
% THIS IS THE IMPORTANT PREDICTIONS OF DRAINMOD USED FOR
CALIBRATION
[dvlp dtwtp]=GetPredicted('predicted.txt');
% THIS PART CONTAIN THE DAILY MEASURED VALUES OF DRAINAGE AND WT
DEPTH
[dvolo dtwto]=GetPredicted('observed.txt');
NDAYS=length (dvlp);
%--- call the objective function
f=objective (dvlp,dtwtp,dvolo,dtwto,NDAYS);
return
}

Function [dvlp dtwtp]=GetPredicted(strr)
Fid=fopen(strr,'r');
A = fscanf(fid,'%g %g %g %g %g %g\n',[6,inf]);
Dvlp=A (4, :);
Dtwp=A (5, :);
fclose(fid)
return

%***************************************************************************
Function [objValue]=objective(dvolp,dtwtp,dvolo,dtwto,n)
[dvolp_aver dvolp_sum]=averg(dvolp);
[dtwtp_aver dtwtp_sum]=averg(dtwtp);
[dvolo_aver dvolo_sum]=averg(dvolo);
[dtwto_aver dtwto_sum]=averg(dtwto);
%!---- ABSOLUTE PERCENT ERROR-------(APE)-----------------------------!
%! FOR DRAINAGE VOLUMES
APEDvol= (dvolp_sum-dvolo_sum)/dvolo_sum;
APEDvol=abs (APEDvol);
%! FOR WATER TABLE DEPTH
APEDtwt= (dtwtp_sum-dtwto_sum)/dtwto_sum;
APEDtwt=abs (APEDtwt);
%! ----- MODELLING EFFICIENCY---------- (EF) -------------------------!
%! FOR DRAINAGE VOLUMES
sum1=norm (dvolp-dvolo);
sum1=sum1*sum1;
sum2=norm (dvolo-dvolo_aver);
sum2=sum2*sum2;
EfDvol=1-sum1/sum2
%! FOR WATER TABLE DEPTH
sum1=norm (dtwtp-dtwto);
sum1=sum1*sum1;
sum2=norm (dtwto-dtwto_aver);
sum2=sum2*sum2;
EfDtwt=1-sum1/sum2;
%!----------------------------------------------------------------------------!
%! IF WATER TABLE DEPTHERS AND SUBSURFACE DRAINAGE WAS CALIBRATED
% THE
% OBJECTIVE FUNCTION WILL BE:-
% objValue=0.6*sqrt ((1-EfDvol)*(1-EfDvol)+(APEDvol+0.5)*(APEDvol+0.5)) ...
% +0.4*sqrt((1-EfDtwt)*(1-EfDtwt)+(APEDtwt+0.5)*(APEDtwt+0.5));
%! IF THE SUBSURFACE DRAINAGE ONLY WAS CALIBRATED THE OBJECTIVE
% FUNCTION
%! WILL BE:-
% ObjValue=sqrt ((1-EfDvol)*(1-EfDvol)+(APEDvol+0.5)*(APEDvol+0.5));
Return
Function [aver summ]=averg(vect)
    summ=sum(vect);
    Aver=mean (vect);
Return
%! THIS IS THE SUBROUTINE FOR GENERATING A NEW POINT IN A SIMPLEX
Function [snew,fnew,icall]=ccceu(a,s,sf,bl,bu,icall,maxn)
%! s(.) = the sorted simplex in order of increasing function
%! values
%! s(.) = function values in increasing order
%! LIST OF LOCAL VARIABLES
%! sb(.) = the best point of the simplex
%! sw(.) = the worst point of the simplex
w2(.) = the second worst point of the simplex
fw = function value of the worst point
ce(.) = the centroid of the simplex excluding wo
snew(.) = new point generated from the simplex
iviol = flag indicating if constraints are violated
   = 1, yes
   = 0, no

[nps, nopt] = size(s);
n = nps;
m = nopt;
Alpha = 1.0;
Beta = 0.5;

% Assign the best and worst points:
Sb = s(1, :); fb = sf(1);
Sw = s(n, :); fw = sf(n);

% Compute the centroid of the simplex excluding the worst point:
Ce = mean(s(1: n - 1, :));

% Attempt a reflection point
snew = ce + alpha*(ce - sw);

% Check if is outside the bounds:
ibound = 0;
s1 = snew - bl; idx = find(s1 < 0); if ~isempty(idx); ibound = 1; end;
s1 = bu - snew; idx = find(s1 < 0); if ~isempty(idx); ibound = 2; end;

If ibound >= 1;
   snew = bl + rand(1, nopt).*(bu - bl);
End;
fnew = functn(nopt, snew);
icall = icall + 1;

% Reflection failed; now attempt a contraction point:
If fnew > fw;
   snew = sw + beta*(ce - sw);
   fnew = functn(nopt, snew);
   icall = icall + 1;
% Both reflection and contraction have failed, attempt a random point:
   If fnew > fw;
      snew = bl + rand(1, nopt).*(bu - bl);
   End;
End;

% END OF CCE
Return;