NUREG/CR-6884

Model Abstraction Techniques for Soil-Water Flow and Transport

U.S. Department of Agriculture
Agricultural Research Service

U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Washington, DC 20555-0001

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at NRC's Public Electronic Reading Room at http://www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and *Title 10, Energy*, in the Code of *Federal Regulations* may also be purchased from one of these two sources.

1. The Superintendent of Documents
   U.S. Government Printing Office
   Mail Stop SSOP
   Washington, DC 20402–0001
   Internet: bookstore.gpo.gov
   Telephone: 202-512-1800
   Fax: 202-512-2250
2. The National Technical Information Service
   Springfield, VA 22161–0002
   www.ntis.gov
   1–800–553–6847 or, locally, 703–605–6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:
Address: Office of Administration,
Reproduction and Distribution
Services Section,
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
E-mail: DISTRIBUTION@nrc.gov
Facsimile: 301–415–2289

Some publications in the NUREG series that are posted at NRC's Web site address http://www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, and transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—
The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852–2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—
American National Standards Institute
11 West 42nd Street
New York, NY 10036–8002
www.ansi.org
212–642–4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX), (4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG–0750).

DISCLAIMER: This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any employee, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this publication, or represents that its use by such third party would not infringe privately owned rights.

NUREG/CR-6884

Model Abstraction Techniques for Soil-Water Flow and Transport

Manuscript Completed: September 2005
Date Published: December 2006

Prepared by
Y.A. Pachepsky[1], A.K. Guber[3], M.T. Van Genuchten[2],
T.J. Nicholson[4], R.E. Cady[4], J. Simunek[3], M.G. Schaap[3]

[1] U.S. Department of Agriculture
Agricultural Research Service
Environmental Microbial Safety Laboratory
Beltsville, MD 20705

[2] U.S. Department of Agriculture
Agricultural Research Service
George E. Brown Jr. Salinity Laboratory
Riverside, CA 92705

[3] University of California Riverside
Department of Environmental Sciences
Riverside, CA 92521

[4] U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Washington, DC 20555-0001

T.J. Nicholson, NRC Project Manager

Prepared for
Division of Fuel, Engineering & Radiological Research
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code Y6724

ABSTRACT

This report describes the methodology of model abstraction in subsurface hydrology. Model abstraction is defined as the methodology for reducing the complexity of a simulation model while maintaining the validity of the simulation results with respect to the question that the simulation is being used to address. The need in model abstraction may stem from the need to improve the reliability and reduce uncertainty of simulations, to make the modeling and its results more explicable and transparent to the end users, and to enable more efficient use of available resources in data collection and computations. A comprehensive review of model simplification techniques developed in subsurface hydrology is included. Abstractions of both model structure and model parameter determination are described. A systematic and objective approach to model abstraction is outlined. A case study is presented that is designed to illustrate how model abstraction can affect performance assessment of contaminant migration at a relatively humid site. Although the model abstraction methodology is generic, it is designed to be of practical use to NRC licensing staff in their review of the performance assessment of decommissioning sites and waste disposal facilities. The model abstraction process provides a systematic approach to understand the adequacy of model simplification, and facilitates communication and transparency of the model to regulators, stakeholders, and the general public.

FOREWORD

An important step in the U.S. Nuclear Regulatory Commission (NRC) staff evaluation of the proposed decommissioning of nuclear sites and facilities is the review of licensees' analyses of the performance of the proposed actions during extended periods of future time. In complex cases, licensees often must develop their own conceptual models of their sites and facilities, and use a variety of computer codes to model the potential radionuclide transport pathways. The model development process is often not systematic and does not identify simplifications made to reduce the complexity of the site's features, events, and processes (FEPs) or explain that the simulation results honor the significant FEPs affecting subsurface flow and radionuclide transport assessments. To better understand the model development process and its implications on modeling of complex cases, the NRC partnered with the U.S. Department of Agriculture's (USDA) Agricultural Research Service (ARS) under Interagency Agreement (RES-02-008), "Model Abstraction Techniques for Soil Water Flow and Transport." The objective was to identify, test, and confirm the practicality of various model abstraction techniques for establishing the appropriate combination of site specific models and supporting data collection programs to simulate water flow and contaminant transport through soils.

This report documents the conduct and findings of this research. The research findings provide the technical basis for developing guidance to NRC licensees and NRC licensing staff on reviewing models of water flow and solute transport in soils, sediments, and porous rocks. The report presents a systematic approach for identifying the need for model abstractions, identifies and reviews model abstraction techniques, discusses the process of selection of appropriate model abstraction techniques, and demonstrates the testing and confirmation of the model abstraction approach through an application to field data. The case study demonstrates the effectiveness of model abstraction in selecting an appropriate level of simplicity for simulating soil water fluxes at a relatively humid site where transport may be affected by the presence of soil macropores and related preferential flow phenomena. The case study shows the capabilities, limitations, and usefulness of selected model abstraction methods and confirms the utility and correctness of the model abstraction method through a series of comparisons. The model abstraction process addresses the uncertainty in model structure and in model parameter determination, and facilitates transparency and realism in the site-specific modeling of essential FEPs. The results of this research supports regulatory decisions that will help make modeling activities and decisions on which they are based more efficient and effective.

This report is a technology transfer document. The NRC staff can use it in reviewing subsurface hydrologic models. The implementation of the model abstraction methodology provides a context for more efficient use of modeling in the regulatory process. Earlier versions of this report were used in training workshops at NRC headquarters on June 22–24, 2004, and May 31, 2005. The report was also circulated to cooperating Federal agencies working under the memorandum of understanding on research in multimedia environmental models. The report was peer-reviewed by the researchers who earlier developed the general strategy of hydrologic modeling and uncertainty analysis for nuclear sites and facilities.

Carl J. Paperiello, Director
Office of Nuclear Regulatory Research

CONTENTS

ABSTRACT..... iii
FOREWORD.....…..... v
EXECUTIVE SUMMARY xiii
ACKNOWLEDGEMENTS.....…... xvi
ABBREVIATIONS.....…….. xx
SYMBOLS….....…....….....…… xx
1 INTRODUCTION…… 1
1.1 Background…….. 1
1.2 Model abstraction…….…… 2
1.3 Objectives…….…… 3
2 MODEL ABSTRACTION TECHNIQUES…… 5
2.1 Background……. 5
2.2 Model abstraction techniques for flow and transport models……... 11
2.2.1 Hierarchies of models……. 11
2.2.2 Delimited input domain…….. 14
2.2.3 Scale change…… 16
2.2.3.1 Scale change – upscaling…… 16
2.2.3.2 Scale change – aggregation……. 19
2.2.4 Metamodeling…… 19
2.2.5 Discretization…….. 22
2.2.6 Scaling……. 22
2.2.7 Parameter estimation with pedotransfer functions…….. 22
2.2.7.1. Estimating water flow parameters…….. 26
2.2.7.2. Performance of pedotransfer functions…….. 33
2.2.7.3 Emerging pedotransfer techniques…….. 36
2.3. Model abstraction implementation……... 38
2.3.1 Status of model abstraction in flow and transport modeling……... 38
2.3.2. Model abstraction process……... 39
2.3.2.1 Justify the need for model abstraction…… 39
2.3.2.1.1 Need in abstraction because of the limited observation ability……. 40
2.3.2.1.2 Need in abstraction because of uncertainty in key model outputs…… 43
2.3.2.2 Review the context of the modeling problem…….. 44
2.3.2.3 Define applicable MA techniques…… 45
2.3.2.4 Define and execute model abstractions…… 47
2.4 Concluding remarks……... 48
3 MODEL ABSTRACTION CASE STUDY……. 51
3.1 Introduction……. 51
3.2 Description of experiments……. 51
3.2.1 Field setup…… 51
3.2.2 Dataset overview……. 59
3.2.2.1 Precipitation……. 59
3.2.2.2 Soil water contents, capillary pressures and soil temperatures…… 59

3.2.2.3 Soil water fluxes……... 64
3.2.2.4 Laboratory data on soil water retention…….. 64
3.3 The base model and the need in model abstraction…… 64
3.4 Review of the model abstraction context…….. 73
3.4.1.The key output……. 73
3.4.2 Statistics to evaluate model abstractions……. 73
3.4.3 Available data and their uncertainty……..…….. 75
3.4.3.1 Averaging water contents …..……. 75
3.4.3.2 Temporal persistence in soil water contents...……. 75
3.4.3.3 Surface evaporation……. 83
3.4.4 Additional information to ensure sound abstraction……... 83
3.4.4.1 Evidence of the macropore flow……. 83
3.4.5 Recalibration of the base model……. 88
3.5 Selection and application of model abstraction techniques…….. 92
3.5.1 Model abstraction design…….. 92
3.5.2 Abstraction with a hierarchy of models: using water budget…… 92
3.5.3 Abstraction with the scale change: profile aggregation…… 93
3.5.4 Abstraction of parameter determination: using an ensemble of pedotransfer functions……... 99
3.5.5 Sequential model abstractions…….. 99
3.5.6 Abstraction with metamodeling……... 110
3.5.7 Pedotransfer-estimated vs. laboratory-measured water retention…… 116
3.6 Summary and conclusion…… 118
4 REFERENCES……...……... 121

APPENDICES
A. PEDOTRANSFER FUNCTIONS TO ESTIMATE SOIL WATER RETENTION…… A-1
B. DIURNAL OSCILLATIONS IN WATER CONENT AND CAPILLARY PRESSURE DATA…….. B-1
C. USING TEMPORAL PERSISTENSE TO ESTIMATE DEPTH-AVERAGED WATER CONTENTS WITH MISSING DATA……... C-1
D. ESTIMATING MATRIC UNSATURATED HYDRAULIC CONDUCTIVITY IN DEEP SOIL LAYERS……... D-1
E. MWBUS: ONE-DIMENSIONAL SOIL WATER FLOW MODEL…….. E-1

Figures

2-1 General taxonomy of model abstraction techniques ……...…..… 7
2-2 Categories of model abstraction techniques relevant to flow and transport modeling in subsurface hydrology…… 12
2-3 Hierarchy of models to simulate water flow and solute transport in structured soils or in unsaturated fractured rock…… 13
2-4 Maximum tracer velocity in 34 documented field experiments with variably saturated soils…… 15
2-5 Spatial and temporal operational scales in hydrology…… 17
2-6 Soil field capacity and soil water retention at fixed matric potentials…… 20
2-7 Observed dependencies of the solute dispersivity on distance in unsaturated soils…….. 23
2-8 Examples of scale-dependence in soil saturated hydraulic conductivity …… 24
2-9 Relationship between field and laboratory water contents at the same soil-water matric potential…… 25
2-10 Saturated hydraulic conductivity K_{sat} grouped by textural classes and bulk density classes; (a) median values $K_{sat(50\%)}$ for the groups, (b) difference between third $K_{sat(75\%)}$ and first $K_{sat(25\%)}$ quartiles of the distributions within each group; data for high and low bulk density samples are shown by filled and hollow symbols, respectively…….. 30
2-11 Estimating water contents at 33 kPa (○) and at 1500 kPa (●) with 21 PTFs developed in various region of the world…… 35
2-12 Using the dense auxiliary data to infer soil hydraulic properties…… 37
3-1 Soil profile at the Bekkevoort experimental site……...…… 52
3-2 View of the instrumented trench…… 54
3-3 Locations and numbering of TDR-probes (●), tensiometers (○) and temperature probes (▲)……. 55
3-4 Three passive capillary samplers on the PVC plate…… 57
3-5 Installation of passive capillary samplers…… 58
3-6 Cumulative rainfall and daily precipitation during the experiment…… 61
3-7 Time series of average water contents at each of the measurement depths… 62
3-8 Time series of average capillary pressures at each of the measurement depths 63
3-9 Consecutive images of two-dimensional distributions of soil moisture contents; days from the beginning of the experiment are shown…… 66
3-10 Time series of soil temperature……. 67
3-11 Cumulative rainfall and soil water fluxes measured with passive capillary lysimeters (PCAPS) at two depths…… 68
3-12 Van Genuchten parameters derived from water retention of samples taken from three soil horizons along the 30-m trench……. 69
3-13 Measured cumulative rainfall and simulated cumulative actual infiltration for the different objective functions…… 72
3-14 Wetting-drying periods selected to compute the cumulative water flux as the key output of the base and abstracted models……... 74
3-15 Time series of TDR-measured water contents at the depth of 15 cm (a-d) and precipitation (e) during the period when all probes worked at this depth……. 77
3-16 Cumulative change in soil water storage in several locations along the

trench. Location numbering is given in the Fig. 3-3…… 78
3-17 Capillary pressure gradient during a drying period in the location 10…….. 79
3-18 Daily average water content and pressure head time series at the location 10.. 80
3-19 Water contents in soil profiles for several days of a drying period……... 81
3-20 Estimated unsaturated hydraulic conductivity of soil matrix below 60 cm….. 82
3-21 Estimated daily evaporation for dry periods during the experiment……. 84
3-22 Probability distribution function of estimated daily evaporation…….. 84
3-23 Relationships between pressure head and water content at the 15-cm depth between 140 and 190 day from the beginning of the experiment……. 85
3-24 Drying water retention curve of a soil with macroporosity……..……. 86
3-25 Cumulative water fluxes computed with constant evaporation of 0.86 mm day…… 87
3-26 Time series of the depth-averaged water contents simulated with the calibrated HYDRUS-1D model ……...…… 89
3-27 Soil water fluxes simulated with calibrated Richards medium model…….. 91
3-28 Design of the model abstraction application in this work. The Richards media was simulated with the HYDRUS-1D software; the water budget model was MWBUS…….. 94
3-29 Time series of the depth-averaged water contents simulated with the calibrated MWBUS model…… 95
3-30 Soil water fluxes simulated with calibrated HYDRUS-1D and the MWBUS models……... 96
3-31 F-statistics of soil water flux simulations with different abstraction techniques for depths of 15 and 55 cm…….. 97
3-32 Time series of transect-average water contents: symbol – measured, red lines – calibrated the HYDRUS-1D model, blue lines - calibrated the WMBUS model for homogeneous soil profile…….. 100
3-33 Cumulative water fluxes simulated with calibrated HYDRUS-1D and MWBUS models for homogeneous soil profile ……... 101
3-34 Comparison of field water retention (symbols) at the depth of 15 cm with water retention estimated using 21 pedotransfer functions listed in the Appendix A and the Rosetta software……... 102
3-35 Time series of transect-average water contents: symbol – measured, red lines - simulated with the HYDRUS-1D, blue lines - simulated with the WMBUS model; water retention data and median hydraulic conductivity estimated with pedotransfer functions……... 104
3-36 Soil water fluxes simulated with HYDRUS-1D and MWBUS models; water retention data and median hydraulic conductivity estimated with pedotransfer functions for layered soil profile…….. 105
3-37 Time series of transect-average water contents: symbol – measured, red lines - simulated with the HYDRUS-1D, blue lines - simulated with the WMBUS model for homogeneous soil profile; water retention data and median hydraulic conductivity estimated with pedotransfer functions for topsoil…… 107
3-38 Soil water fluxes simulated with HYDRUS-1D and the MWBUS model; water retention data and median hydraulic conductivity estimated with pedotransfer functions for topsoil profile; error bars show one standard

deviation…… 108
3-39 Soil water fluxes simulated with calibrated HYDRUS-1D and the MWBUS model for the homogeneous soil profile…… 109
3-40 Schematics of the neural network application to abstract simulation models... 111
3-41 Estimated with the artificial neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – MWBUS, artificial neural network input – daily average flux, ANN output – daily average flux, the total number of hidden neurons – 20……. 112
3-42 Estimated with the artificial neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – MWBUS, artificial neural network input – daily average flux, ANN output – daily average flux, the total number of hidden neurons – 12……... 113
3-43 Estimated with the artificial neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – Hydrus-1D, artificial neural network input – daily average flux, ANN output – daily average flux, the total number of hidden neurons – 20……. 114
3-44 Estimated with the neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – HYDRUS-1D, ANN input – weekly average flux, ANN output – weekly average flux, number of hidden neurons – 12……...……. 115
3-45 Probability distribution of the root-mean square error (RMSE) of water content from Monte Carlo simulations for the HYDRUS-1D and the MWBUS models with two sources of soil hydraulic properties and two representations of soil profile…… 117

Tables

2-1 A set of model simplification techniques suggested by Innis and Rextad (1983) for four steps of model development……..…… 6
2-2 Meaning and theoretical basis of model abstraction techniques……. 8
2-3 Regression equation to predict soil water content at specific matrix potential (after Rawls et al., 1982)……. 27
2-4 Regression equation to predict soil water content at specific matrix potential (after Rawls et al., 1983)……. 28
2-5 Saturated hydraulic conductivity (cm day^{-1}) at three probability levels for USDA textural classes (from Rawls et al., 1998)……... 29
2-6 Parameters of the Ahuja equation (2-3) to estimate saturated soil hydraulic conductivity……. 32
2-7 Potential need for model abstraction from the parameter uncertainty standpoint based on Hill (1998) analysis…… 41
3-1 Average soil texture (%) along the trench at the Bekkevoort site…….. 53
3-2 Types and frequencies of the monitoring measurements…… 60
3-3 Coefficients of variation (%) of soil water contents in time……... 65
3-4 Coefficients of variation (%) of soil water content along the trench…….. 65
3-5 Statistics of van Genuchten parameters of water retention in samples along the 30-m rench…… 70
3-6 Calibrated van Genuchten (Eq. 2-2) and van Genuchten-Mualem (Eq.2-5)

parameters of the Richards medium model (HYDRUS-1D)…….. 90
3-7 Calibrated parameters of the water budget model (MWBUS)…… 98
3-8 PTF used to run HYDRUS 1D with Brooks-Corey (BC) and Van Genuchten (VG) water retention equations……... 103

EXECUTIVE SUMMARY

This report describes and demonstrates the methodology of model abstraction in subsurface contaminant hydrology. Model abstraction is defined as *methodology for reducing the complexity of a simulation model while maintaining the validity of the simulation results with respect to the question that the simulation is being used to address.*). Thus, model abstraction reduces the complexity of the system to be simulated to its essential components and processes through a series of conceptualizations, selection of significant processes and appropriate scales, and identification of the associated parameters.

The need in model abstraction addresses the concerns of the NRC licensing staff that the use of overly complex simulation models in decommissioning plans and performance assessment may cause an excessive burden of data collection and computations as well as difficulties in interpreting simulation results and conveying the simulation approach to both technical and lay audiences. The presumed risk of leaving some important process or feature out often leads the model users to employing fairly complex flow and transport models that simulate subsurface systems through detailed numerical grids and associated data inputs, thereby introducing large computational and intensive data-collection requirements. However, many of the detailed features, events and processes represented in these complex models may have limited influence on the performance of a specific site.

The feasibility of model abstraction has been demonstrated in many research and engineering fields that give ample examples of models having strikingly different complexity and yet the same accuracy. In contaminant hydrology, model abstraction is possible because the complexity of flow and transport pathways at a specific site is easy to perceive but difficult to represent in mathematical equations of the model without making strong simplifying assumptions. Different sets of plausible assumptions lead to different models that are consistent with the available observations. The multiplicity of models reflects the multiplicity of possible conceptual approaches to representation of complex subsurface processes in mathematical form tractable within limitations of existing computer and measurement technologies.

The methodology of model abstraction has been developing for more than 30 years in various research and engineering fields. Most of the model abstraction techniques are specific to the type of mathematical models used in a specific field. This report contains an overview and annotated examples of model abstraction in modern contaminant hydrology. Two main targets of abstraction are (1) the model structure, i.e. the formal description of specific processes and their interactions that affect flow and transport variables, and (2) the parameter determination, i.e. the estimation of constant and functions serving as coefficients in model equations.

Structure of hydrologic models is changed with model abstraction via (a) using pre-defined hierarchies of models, (b) delimiting input domain, (c) scale change done by either upscaling or aggregation, and (d) metamodeling. A predefined hierarchy contains a

series of progressively more simple conceptual and corresponding mathematical representations of porous media. The class of model abstraction techniques based on the delimiting input domain utilises the fact that some features, events, or processes may be not relevant for a given set of scenarios or for a given set of model outputs. Scale change provides transitions between four operational scales – core, profile/"pedon", field, and watershed scales that are of interest in contaminant hydrology. Model abstraction with scale change alters model equations, variables and parameters with two classes of methods: upscaling and aggregation. Upscaling model abstraction methods use the fine-scale model and the fine scale media properties to derive the coarse-scale model equations and to relate the coarse-scale and fine-scale transport parameters. On the contrary, model abstraction with aggregation does not assume any relationship between model parameters and equations at the fine and at the coarse scales. Parameters of the coarse-scale model are lumped, and are subject to calibration with field data. Aggregation can be also done without the change in the model equations by combining several soil horizons or geologic strata. Metamodeling seeks to simulate the input-output relationships of the complex model with a simple statistical relationship.

Model abstraction applied to parameterization of hydrologic models does not provide a substitution for the model calibration, but rather seeks to obtain reasonable estimates of parameter values and their variability. Such estimates are useful for setting initial values of parameters for the model calibration, using non-calibrated model in pilot studies and field campaign designs, assigning values of parameters that are shown to be not sensitive, etc. Abstraction of parameter determination affects parameter estimation via (a) discretization, (b) scaling, and (c) pedotransfer functions. The discretization abstraction techniques replace the continuously varying spatial fields of parameters with piece-wise constant spatial distributions. The scaling abstraction techniques are useful when models are coarsened, i.e. the grid cell size is substantially increased but no changes in the model structure is made. The scale-dependence in model parameters is usually encountered in such cases, and scaling relationships are needed to convert the original parameter values and measurement data to the parameters values of the coarsened model. The pedotransfer techniques convert the readily available data to the hydraulic and transport parameters of the unsaturated flow and transport. Those parameters are notoriously difficult to measure, and s substantial effort has been made to estimate these parameters from the data available from soil survey or borehole logs. The empirical functions used for such estimating are often called pedotransfer functions. Large databases have been assembled to encompass variety of soil properties in developing pedotransfer functions in different parts of the world. New powerful heuristic tools, such as artificial neural networks and regression trees, appeared to be useful in the development of pedotransfer functions. Still, the accuracy of pedotransfer functions outside of their development dataset remains essentially unknown, and the use of ensembles of pedotransfer functions is strongly recommended.

The model abstraction implementation has to conform requirements of objectiveness, systematic implementation, comprehensiveness, and efficiency. The MA process starts with an existing *base* model that can be calibrated and used in simulations. The *key output* of the model id defined that provides the necessary and sufficient information to

decide on issues of interest. The base model may need abstraction for one or more of the following reasons (a) the base model includes a complex description of processes that cannot be observed well and yet need to be calibrated; the calibrated values of parameters of those processes are very uncertain, (b) the base model propagates uncertainty in the initial distributions, parameters, and forcing in a manner that creates an unacceptable uncertainty of the key output, (c) the base model produces inexplicable results in terms of the key output, (d) the base model requires an unacceptable amount of resources for computations, data preprocessing, or data post-processing, e.g. the base model is not suitable to be used as a part of a real-time modeling system that requires short computer runtimes, (e) the base model lacks transparency to be explicable and believable to the users of the key output.

The model abstraction process, as described in this report, is a transparent step-by-step formalized procedure of justification of the use of a simplified model. It includes the following steps (1) justify the need for the model abstraction, (2) review the context of the modeling problem, (3) select applicable model techniques, (4) determine the model abstraction directions that may give substantial gain, (5) simplify the base model in each direction. Statistical criteria are provided to justify of the need in model abstraction in case it is related to the uncertainty in calibrated parameter values or in the model key output. The context of the modeling problem is described that has to be reviewed to realize what details and features of the problem are omitted or de-emphasized when the abstraction is performed, and thus to warrant the comprehensiveness and objectiveness of the model abstraction process. Model abstraction can lead to simplifications via (a) the number of processes being considered explicitly, (b) process descriptions, (c) coarsening spatial and temporal support, (d) the number of measurements for the reliable parameter estimation, (e) reduced computational burden, (f) data pre-processing and post-processing. To guide the selection of applicable model abstraction techniques, classes of model abstraction techniques are specified that lead to each type of the simplification. Each abstracted model has to be parameterized and confirmed in the uncertainty context.

The model abstraction case study was design to illustrate the model abstraction process and techniques. The main objective of the test case was to demonstrate how the model abstraction can be applied to understanding and prediction of soil water fluxes at a relatively humid site where transport may be affected by the presence of soil macropores and related preferential flow phenomena. The test site was located in Bekkevoort, Belgium. The 1.5-m deep test trench was dug in a loamy soil and was instrumented to measure soil water content and soil matric potential in 60 locations at five depths hourly, soil temperature in 6 locations at five depths hourly, and soil water fluxes in three locations at two depths once in one to four days. Vegetation was removed and a thin layer of fine gravel covered the soil. Soil water monitoring continued for 384 days. Soil was sampled at 90 locations at three depths to measure soil water retention in laboratory. The base model was the Richards equation of water flow in variably saturated porous media. The model was calibrated using almost 200,000 measurements of soil water content and soil water potential. The key output of the model was total soil water flux at two measurement depths and at the bottom of the soil profile over three wetting-drying periods.

The need in model abstraction was justified by the fact that the calibrated base model predicted substantial runoff whereas no runoff was observed at the site during the monitoring period. The base model produced inexplicable results in terms of the key output.

The review of the modeling context revealed that (a) soil water contents across the trench demonstrated strong temporal persistence, and (b) a substantial preferential flow should occur in soil. The persistence was used to estimate soil bottom flux from water content measurements, and those flux flow estimates were used to assess evaporation from soil surface. Thus, the review of modeling context helped to improve knowledge about boundary fluxes that have not been measured.

Four classes of applicable model abstraction techniques were selected. Model structure abstraction was achieved using the hierarchy of model, aggregation, and metamodeling. The hierarchy of models included the Richards equation and a simple soil water budget model. Aggregation was done by replacing the layered soil profile with the homogeneous profile. Metamodeling was done with the artificial neural network that used daily and weekly precipitation as input to estimate monthly cumulative soil water fluxes. An ensemble of 22 pedotransfer functions was used in model abstraction of parameter determination. HYDRUS1-D, MWBUS, and MATLAB software packages were used to calibrate the abstracted models and to run Monte Carlo simulations to evaluate the uncertainty in calibrated model outputs.

The abstraction with the hierarchy of models was useful. The simple soil water budget model was less accurate in predictions of soil water content compared with the Richards model. However, unlike the more complex Richards model, this simple soil water budget model correctly predicted the absence of runoff and measured cumulative soil water fluxes. The prediction of runoff was an artifact of the Richards model calibration in absence of measured boundary fluxes. This abstracted model appeared to be instrumental in both explaining behavior of the complex model and in predicting the key output – soil water fluxes.

The abstraction with aggregation was not useful in this case. The Richards model was less accurate with respect to soil water contents and continued to generate large simulated runoff when a homogeneous soil layer was introduced.

The abstraction of parameter determination with pedotransfer functions was useful. It showed that the Richards model with parameters in correct ranges is able to correctly simulate soil water fluxes. The ensemble of pedotransfer functions represented field water retention better than data from laboratory soil water retention measurements. Soil water content predictions with the Richards model were more accurate with the ensemble of pedotransfer functions than with laboratory data.

The neural network metamodel was extremely accurate in estimating cumulative soil water fluxes. It was also five orders of magnitude faster than the numerical algorithm coded in HYDRUS-1D.

In general, the model abstraction process was successful in this case study. Model abstraction explained the strange behavior of the complex model, and provided the correct description of the system behavior and plausible parameter ranges. The case has also demonstrated the need in model calibration and uncertainty analysis toolboxes and appropriate user interfaces to perform the model abstraction in the uncertainty context.

Overall, the intensive use of models in subsurface contaminant hydrology has resulted in development of many model abstraction techniques that to-date have been used mostly in research. Potential benefits of model abstraction include improvement of understanding and communication of modeling results, more robust predictions, and better understanding of essential factors and their representation in models. This makes model abstraction an attractive methodology for engineering modeling applications. The model abstraction process can be set as a transparent step-by-step formalized procedure of justification of the use of a simplified model. An important feature of models abstraction is the explicit treatment of model structure uncertainty. The model structure, along with the data uncertainty, and scenario uncertainty, is known to introduce the uncertainty in modeling results. Unlike the uncertainty in input data, in model parameters, and in scenarios, the effect of the model structure uncertainty on the uncertainty in simulation results is usually impossible to quantify in statistical terms. Using model abstraction, a series of models with feasible structures can be built and evaluated in a systematic manner. Each of the models is evaluated from results of an ensemble of simulations by its accuracy to measurement data and by its predictions with respect to scenarios that have not been observed.

It is expected that as the model abstraction methodology will develop, knowing about model abstraction and applying it to the regulated sites will be helpful for the NRC licensing staff. Reviewers will have a means to determine whether the models submitted to support licensing actions adequately represent the site, and whether the investigations adequately represent important features, events, and processes for the site. Managers will have an advisory tool to decide whether the requests for additional information are targeted tat sensitive site parameters and processes. On the other hand, the NRC licensees will have a device to determine whether a simple model that is easy to understand and communicate to regulators, stakeholders, and the general public can adequately represent their site.

ACKNOWLEDGMENTS

This work was supported by the Interagency Agreement RES-02-008 "Model Abstraction Techniques for Soil Water Flow and Transport." The Nuclear Regulatory Commission staff conceived and championed this research.

The authors acknowledge with thanks contributions by the following individuals: Dr. Diederik Jacques of the SCK-CEN, Belgium for having provided experimental data from the Bekkevoort experimental site; Dr. Brent Thomas of RAND Corporation for having brought to our attention the recent extensive work on model abstraction; Professor Shlomo Neuman of the University of Arizona and Dr. Glendon Gee from the Pacific Northwest National Laboratory for their insightful review of a draft version of this manuscript; members of the Working Group II on Uncertainty and Parameter Estimation of the Interagency Steering Committee on Multimedia Environmental Models for their helpful comments on the presentations of this work being in progress.

Abbreviations

1D, 2D	One-dimensional, Two-dimensional
ANN	Artificial neural networks
ANOVA	ANalysis Of VAriance between groups
AQUA	One-dimensional water budget based model in unsaturated soil
ASCII	American Standard Code for Information Interchange.
CEC	Cation Exchange Capacity
CV	Coefficient of variation
FAO	Food and Agriculture Organization of the United Nations
GA	Genetic algorithms
HYDRUS	Software package for simulating the movement of water, heat and multiple solutes in variably saturated media.
MA	Model Abstraction
MACRO	A model of water flow and solute transport in Macroporous soil.
MEPAS	Multimedia Environmental Pollutant Assessment System
MIN3P	Multicomponent reactive transport modeling in variably saturated porous media
MODFLOW	Three-dimensional finite-difference ground-water flow model
MWBUS	Model of Water Budget of Unsaturated Soil
Neuropack	Neural Networks Package for fitting Pedotransfer Functions
NRC	U.S. Nuclear Regulatory Commission
PCAPS	Passive Capillary Samplers
PEST	A nonlinear Parameter Estimation and optimization package
PTF	Pedotransfer function
PVC	Polyvinyl Chloride
Rosetta	Windows-based program that implements artificial neural network to soil hydraulic parameters
SUFI	Parameter estimation methodology
TDR	Time Domain Reflectometer
UCODE	A Computer Code For Universal Inverse Modeling
USDA	United States Department of Agriculture
WGEN	A weather generator

Symbols

Roman symbols

a	a parameter of PTF
A	a parameter of PTF
b	a parameter of PTF
B	a parameter of PTF
C	clay
C	coarse
CA	clay activity

cL	clay loam
clay	percent clay
D	mean depth of sample in centimeters
d_s	geometric mean diameter of soil particle size distribution
E	simulated daily evaporation rate
F	fine
h	capillary pressure (the absolute value of the matric potential)
h_b	bubbling pressure in the Brooks-Corey water retention model
h_e	air entry potential
h_{es}	air entry potential at a standard bulk density
K	number of the working probes at one depth
K_{sat}	saturated hydraulic conductivity
Kw	unsaturated hydraulic conductivity
l	a parameter of unsaturated hydraulic conductivity model
L	loam
lS	loamy sand
m	a parameter of van Genuchten's water retention model
M	medium
MF	medium fine
n	a parameter of van Genuchten's water retention model
N	number of probes
OC	percent organic carbon
OM	percent organic matter
Q	interlayer water flux
R	rainfall rate
R^2	Coefficient of determination
RMSE	Root Mean Square Error
S	sand
S	soil water saturation
sand	percent sand
SAR	sodium adsorption ratio
sC	sandy clay
scL	sandy clay loam
Si	silt
siC	silty clay
sicL	silty clay loam
siL	silt loam
silt	percent silt
sL	sandy loam
$s_{\beta i}$	standard derivation of the relative water content
t	time
topsoil	a parameter of PTF
VF	very fine
w	gravimetric water contents
w_{330}	gravimetric water contents at capillary pressures of 330 cm
w_{15000}	gravimetric water contents at capillary pressures of 15000 cm

z	depth

Greek symbols

α	a parameter of van Genuchten's water retention model
β_{ij}	relative water content
$\overline{\beta}$	average relative water content
γ	a parameter of unsaturated hydraulic conductivity model
$\Delta\theta$	water content change
Δt	time interval
Δz	thickness of the soil layer
δ	a parameter of unsaturated hydraulic conductivity model
θ	volumetric water content
θ_e	effective porosity
θ_{FC}	water field capacity
$\overline{\theta}_j$	average water content
$\hat{\theta}_j$	average water content over working probes at this depth for the sampling time j
θ_r	residual water content
θ_s	saturated water content
λ	pore size distribution index in Brooks-Corey water retention model
ρ_b	soil bulk density
σ_g	geometric standard deviation of soil particle size distribution
τ	resistance parameter
τ_0	the value of resistance parameter at depth z=0
ϕ	porosity in Brooks-Corey water retention model

1 INTRODUCTION

1.1 Background

NRC staff review of performance assessments of nuclear facilities, e.g., decommissioning of the facilities, management of low-level and high-level radioactive waste disposal sites, frequently involves the review of models of water flow and solute transport in soils, sediments, and porous rocks in the vicinity of the nuclear facility (U.S. NRC, 1993, 1994). These models seek to represent complex and highly transient subsurface systems. Representation of those complex systems in existing models ranges from a very simple to a extremely sophisticated (Davis et al., 2004; Meyer et al., 1997; Neuman and Wierenga, 2003; Reilly and Harbaugh, 2004; Hill et al., 2004).

It is generally agreed that models of the appropriate complexity may be required to capture certain flow and transport phenomena (Hunt and Zheng, 1999). How to reach this appropriate complexity remains an ongoing research issue. Suggested systematic approaches are based on the scale of interest. A "downward" approach in hydrologic modeling typically starts with a very simple model to capture the signature of, say, the annual runoff response of a catchment (Milly, 1994). Complexity is added as the spatial and temporal scales become finer. (Jothityangkoon et al., 2001). Neuman and Wierenga (2003) addressed the issue of the suitable level of model complexity in the site-specific hydrogeologic modeling. They indicated that site-specific hydrogeologic heterogeneities and driving mechanisms may be controlled by the features, events and processes belonging to the larger scale(s), and that the models have to account for these controls. At the same time, site-specific hydrogeologic heterogeneities and driving mechanisms may be also significantly influenced by the smaller scale features, events and processes, and this also has to be accounted for in the model. Usually, the complexity of flow and transport pathways at the specific site may be easily perceived, but it is often difficult to represent it in mathematical equations of the model without making strong simplifying assumptions (Beven, 2002). This implies that several different structures of the model could be consistent with the available observations. Such multiplicity of models reflects the multiplicity of possible conceptual approaches to representation of complex subsurface processes in mathematical form tractable within limitations of existing computer and measurement technologies.

It has been amply demonstrated that the increase in complexity of the model does not necessarily mean the increase in its accuracy. Examples of equally accurate models of strikingly different complexity can be found in various fields of modeling applications, such as regional ground water assessments (Kelson et al., 2002), chemical engineering (Diwekar, 1994), marine ecology (Stillman et al., 2000), runoff generation (Jakeman and Hornberger, 1993), population dynamics (Stephens et al., 2002), demographic projections (Smith, 1997), state-wide water supply systems (Palmer and Cohan, 1986), battlefield simulations (Sisti and Farr, 1997), etc.

If the complexity is not inexorably linked with the accuracy, there may exist an opportunity to simplify models. Simplifying models may stem from several rationales. As

motivations for model simplification, Neuman and Wierenga (2003) list narrowly defined contextual or regulatory criteria, limited data or resources, and a quest for transparency. The narrowly defined criteria usually mean that the accuracy of the model is evaluated with respect to some (but not all) variables that this model predicts. Limitations of resources may refer to runtime and also to data pre-processing and post-processing. The transparency may be imperative in communicating model requirements and justifying the use of the specific model.

1.2 Model abstraction

Model abstraction is the methodology of model simplification performed systematically and objectively. Frantz (2002) defined the model abstraction as "*a methodology for reducing the complexity of a simulation model while maintaining the validity of the simulation results with respect to the question that the simulation is being used to address"* (Frantz, 2002). Thus, the model abstraction reduces the complexity of the natural system to be simulated to its essential components and processes through a series of conceptualizations, selection of significant processes, and identification of the associated parameters.

The presumed risk of leaving some important process or feature out often leads the model users to employing fairly complex flow and transport models that simulate subsurface systems through detailed numerical grids and associated data inputs, thereby introducing large computational and intensive data-collection requirements. Many of the detailed features, events and processes represented in these complex models may have limited influence on the performance of a specific site. This will cause an excessive burden of data collection and computations as well as difficulties in interpreting simulation results and conveying the simulation approach to both technical and lay audiences. This has become a concern on the part of both NRC staff and the staff of other regulatory and research federal agencies involved in multimedia modeling. Applying model abstraction in a systematic manner will address this concern.

Successful model abstraction results in a simpler model that renders more easily understandable description of the problem and allows discussions to be focused on the most important aspects. It is often imperative to explicitly acknowledge the abstraction strategy to make a model justifiable and defensible. Finally, model abstraction explicitly deals with uncertainties in modeling. Therefore, model abstraction techniques that can simplify and expedite the assessment of complex systems without significant loss of accuracy would greatly benefit the synthesis and review of performance assessments.

Recently, the issue of model simplification has been addressed for the subsurface hydrologic modeling in the comprehensive strategy developed by Neuman and Wierenga (2003) who advocated the need in systematic approach. This report develops and illustrates such approach.
We note that the term "model abstraction" is sometimes applied to describe the model development, i.e. representation of the complex reality with a specific conceptual and/or

mathematical model, or to describe a development of a set of progressively simplified models. Such meaning of the term is not used in this report.

1.3 Objectives

The general goal of this work was to provide the NRC with an improved analysis capability that would enhance the credibility and defensibility of site performance reviews. Toward this goal, the specific objectives of this study were
(1) to identify model abstraction techniques that may be appropriate for characterizing and simulating water flow and contaminant transport in vadose zone, and
(2) to develop a case study that will demonstrate the efficiency of some of model abstraction techniques for a specific site.
Our presumption was that model abstraction techniques that reduce either the conceptual complexity, or the parameter determination complexity, may also enhance communications with stakeholders and help clarify the scientific bases for decision-making related to the first-order importance issues.

2 MODEL ABSTRACTION TECHNIQUES

2.1 Background

Two definitions of the term 'simple' are pertinent to modeling: (1) easily understood or done, and (2) plain and uncomplicated in form, nature, or design (Weiner and Simpson, 1991). The definition 1 applies to the user perception and to the model use. The definition 2 pertains to the model structure and to the details of modeling. Simplification of models has been analyzed in both directions for more than 30 years. The first comprehensive discussion of model simplification was presented by Meisel and Collins (1973). These authors listed four advantages of using the simplified models:

- they are less expensive to use; such savings will permit more thorough analysis for a given analysis budget;
- they have fewer input requirements;
- they are easier to transfer and/or combine with other models;
- they are easier to interpret; it is easier to understand the properties of and results from a model with a small number of state variables and parameters than the properties and results of one with more.

The first summary of model simplification techniques was compiled by Ziegler (1976) who introduced four categories of the model abstraction techniques:

- dropping unimportant parts of the model,
- replacing some part of the model by a random variable,
- coarsening the range of values taken by a variable,
- and grouping parts of the model together.

Innis and Rextad (1983) noted that model can be simplified at various stages of a modeling project. They listed 17 categories of simplification techniques that were used in ecological modeling (Table 2-1). The authors emphasized that the model simplification has to be conditioned on the purpose of modeling and proposed the first guidelines of using the model abstraction in modeling natural systems.

Frantz (1995) has viewed the model abstraction as a simplification of the model that assures validity of the model for the specific purpose. He introduced a comprehensive taxonomy of the model abstraction techniques presented in Fig. 2-1. Table 2-2 describes the meaning of the classes of techniques proposed by Frantz (1995). The model boundary abstraction changes the real-world factors accounted for in the model by changing the exogenous variables of the model. The model behavior abstraction eliminates the transitions through the system states that are irrelevant to the simulation requirements. Finally, the model form abstraction uses inputs and outputs from the multiple simulations with the model to define the relationship between input and output with a simpler mathematical model. Frantz (1995) noted that different model abstraction techniques tend to be applied simultaneously. Later, other classifications of model abstraction techniques have been proposed (Caughlin and Sisti, 1997; Ziegler, 1998; Yilmaz and Ören, 2004) that have less detail but generally cover the same set of techniques as the Frantz' (1995) classification.

Table 2-1. A set of model simplification techniques suggested by Innis and Rextad (1983) for four steps of model development

No.	Technique	Examples
Step 1. Hypotheses		
1	System organization	In a conservative system, one of compartment can be represented by subtraction; self organization can impose relationships between parameters that effectively eliminate one or more parameters,
2	Filtering	Fluctuations in a state or driving variable at very high frequencies are not sensible to the remainder of the system. For instance, rapid changes in solar insolation have little effect on the growth of a tree.
3	Stochastic features	Stochastic systems may be better represented directly rather than as stochastically disturbed deterministic systems
4	Graph theory	Analysis of diagrams of the interrelations between variables may help to isolate highly interactive subsystems for low resolution treatment
Step 2. Formulation.		
5	Sensitivity-based results	Factors of little supposed importance may be omitted; susbsystems with little effect on system dynamics can be aggregated.
6	Structure and logic analysis	Models often contain numerous model-building artifacts whose removal may simplify the result.
7	Dimensional analysis	Using dimensionless variables can make results more broadly applicable.
8	Repro-Meta modeling	A model of a model may offer such advantages over the original model as (a) finding interrelations not obvious in the data or the model, (b) simulated data can be selected to meet restrictive requirements (e. g., being equally spaced in time), (c) regression of input/output parts applied to all or part of simulation may speed and reduce the cost of projection.
9	Time constants of sub-systems	Widely disparate rates in a model may be eliminated using suitable approximations, e.g. slowly changing variables may be replaced by constants, rapidly changing variables that equilibrate with other variables may be represented with algebraic expressions in place of dynamic equation, thus eliminating a need in large number of parameters.
10	Analytical solutions	Using eigenvalues and eigenvectors instead of solving evolutionary matrix equations
11	Interdependencies	Interdependencies among parameters can be utilized that stem from data or generally accepted theories.
12	Perturbation	Correct linearization of the initial model.
13	Calculation of output moments	Computing statistical moments of the output directly from input data and model structure
Step 3. Coding		
14	Languages	Using the simulation language tailored to the needs of the study
15	Coding improvement	Coding improvements can ease reading and debugging as well as increase the execution speed.
Step 4. Experiments		
16	Variance reduction	Reducing the number of simulation runs needed to estimate an output value within a specified accuracy; accounting for correlations between variables
17	Linear systems	Retaining the dominant eigenvalues and eigenvectors

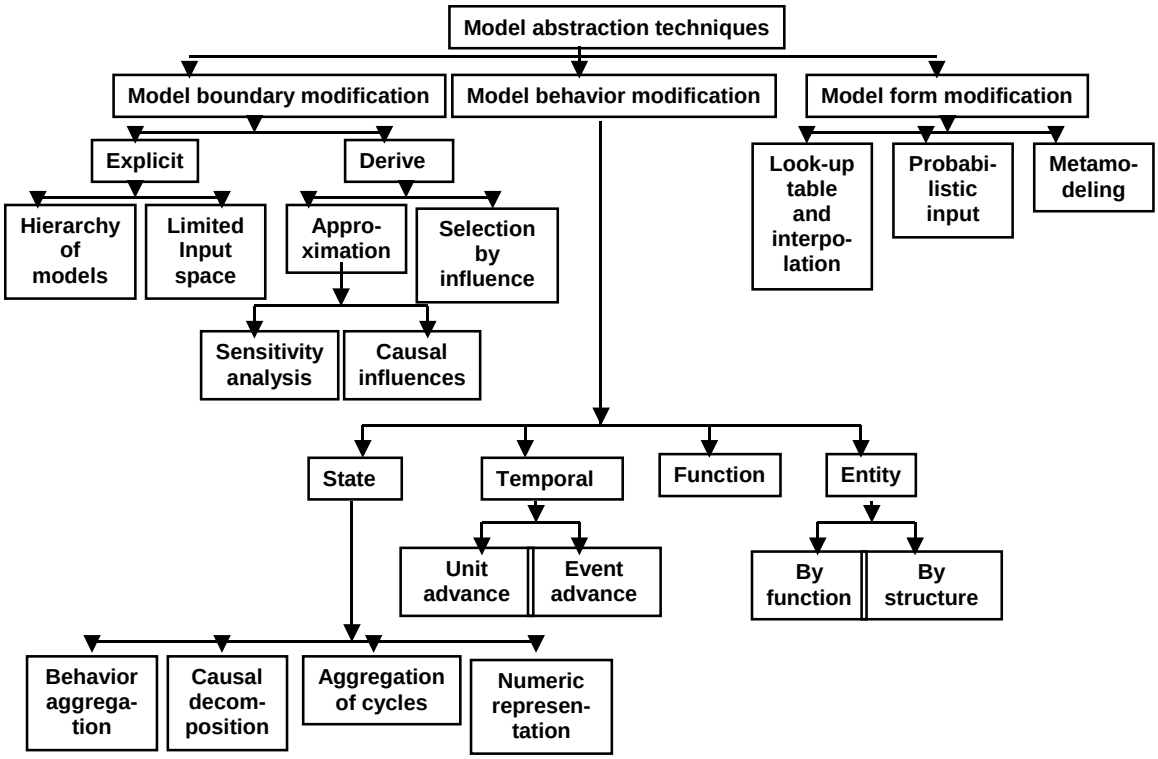

Figure 2-1. General taxonomy of model abstraction techniques after Frantz (1995). With permission.

Table 2-2. Meaning and theoretical basis of model abstraction techniques (after Frantz, 2002)

Class of model abstraction techniques	Broad meaning.	Theoretical background
Hierarchy of models	A pre-defined hierarchy of models can be created to include simplified models and more complex models.	Graph of models (Addanki et al., 1991); compositional modeling (Falkenhauer and Forbus, 1991).
Limited input space	Limit the domain of inputs provided. Some process descriptions are not important or can be simplified because input parameters are in a limited range such that the additional processes have little or no effect on the results.	Pre-defined experimental frame of reference (Zeigler, 1976, Zeigler et al. 1998).
Approximation	Parameter elimination based on requirements of the simulation to be executed rather than being defined externally. Analysis determines which parameters in a model can be eliminated, yielding a new and computationally simpler model that provides approximate results that are within a certain tolerance necessary to meet the requirements of the simulation	Causal relationships between parameters (Nayak, 1992), aggregation of variables (Simon and Ando, 1996).
Selection by influences	Simplifying models by using relationships between variables observed in the simulations. This approach focuses on determining which variables in the model must be exogenous, and requires that the effects of variables on other variables in the model be defined.	Automated modeling (Rickel and Porter, 1994).
Behavior aggregation	Combining system states whose distinctions are irrelevant to the simulation output. Temporal ordering of events may be irrelevant to the final simulation result.	Automated modeling (Rickel and Porter, 1994).
Causal decomposition	Dividing a model into loosely connected components, executing each component separately, and searching for constraints that execution of one component can impose on other components. Where no causal relationship exists, the components may be executed in parallel.	Tractable simulations (Clancy and Kuipers, 1994), causality in model abstraction (Iwasaki and Simon, 1994).

Aggregation of cycles	Combining states that reflect similar sequences and distinctions among the individual sequences are irrelevant to the final outcome.	Aggregation in simulations (Weld, 1986).
Numerical representation	Replacing continuous variables by categorical or nominal variables. For example, combining states characterized by floating point numbers that round to the same integer value.	Developments in genetic algorithms (Mitchell, 1996) and in regression trees (Clark and Pregibon, 1992).
Temporal aggregation	Decreasing temporal resolution. This type of abstraction reduces computational complexity by reducing the granularity (coarseness) of the time advance. This abstraction can be applied in either unit advance or event advance in discrete event simulations	Temporal grain size aggregation (Iwasaki, 1990)
Entity aggregation	Refers to the representation of a collection of "lower level" entities by a "higher level" entity. Entity aggregation may be performed by structure or by function. Aggregating units that perform common or similar functions creates an entity that performs a function that is the aggregate of the individual functions. Aggregation by structure is accomplished by replacing lower level entities by higher level entities. Behavior in such cases is substantially different as a result of abstraction, even if the output is the same. Aggregates functions of entities or subunits but leaves entities in place. Functions are redefined to provide a coarser list of states or output information from existing entities.	Qualitative simulations (Fishwick and Luker, 1991), variable resolution modeling (Hillestad et al, 1992, 1993).
Look-up tables	Simplification of the input-output transformation within a model or model component by means of a decrease in computational effort.	
Probabilistic input	Using randomly generated inputs to develop lumped models.	Fishwick, 1988.
Metamodeling	Approximating the output response surface of the model with a statistical regression or using heuristic data mining tools.	Neural network-based metamodeling Burton (1994)

Recent developments in the model abstraction theory and applications include using the multiresolution modeling as proposed by Fishwick (1995). This type of modeling presumes existence and use of two or more models of different complexity for the same system. Davis and Bigelow (2003) indicated that models of different resolution, or detail, have different strengths and weaknesses. Detailed models are particularly valuable for representing explicitly the underlying phenomena. Successful models of this type are the embodiments in mathematics and/or in computer programs of the fundamental knowledge about the subject in question existing at the moment of the model development. However, the fundamental knowledge about the system under considerations not the only important knowledge (or even the *most* important knowledge) for the decision-making A significant part of our knowledge of the world is low-resolution. Bigelow and Davis (2003) noted that both the insightful strategy-level analysis and the decision support typically require relatively simple models. The most fundamental reason is cognitive: decision makers need to reason about their issues and inject their own judgments and perspectives. They need to construct coherent stories that are convincing to themselves and to others. This implies abstracting, that is reducing of what may be a very complex problem to a relatively small number of variables or cognitive chunks (e.g., 3–5 rather than 10s or 100s). This also implies focusing on the appropriate variables and on the cause-effect relationships. Another fundamental reason for using simpler models is that strategy-level problems may be characterized by massive uncertainty in many dimensions. The appropriate way to address such problems is often the exploratory analysis in which one examines issues across the entire domain of plausible initial states. The exploratory analysis, however, is most effective when the model comprehensively covers the space in which the problem is considered with only 3–12 variables. In such cases, the exploration can be comprehensive and comprehensible. In contrast, if one has a large model and explores by holding hundreds of variables constant while varying only a few (because in reality it is not possible to vary all of them), the results cannot be assessed confidently because the effects of varying the others are unknown. Model abstraction also seems to be very attractive because simple models require much smaller amounts of the experimental data, much less time and efforts for data preparation and post-processing. Analysts and end users can quickly comprehend simple models and their inputs and outputs.

Model simplification has also been recently advocated as a powerful tool to improve the usefulness of complex ecological models for uncovering mechanisms of ecosystem functioning (Van Ness and Scheffer, 2005). Thes authors have proposed a strategy that includes joint exploration of the behavior of a complex model and of its simplified counterpart. The purpose of the simplification in this case is to gain understanding about causes of the patters in the complex model output.

Types of models applied in various research and engineering fields are quite different. Therefore, not all model abstraction techniques are relevant to the particular field of knowledge. For example, abstracting models of the discrete events (i.e. the event advance abstraction in Frantz's classification) is not relevant to subsurface hydrology in its current state where the representation of flow and transport is predominantly continuous. The arsenal of model abstraction techniques developed in linguistics or in

computer science has a very limited use in subsurface hydrology because of fundamental differences in types of models used.

The need for the model simplification in contaminant hydrology was outlined by Neuman and Wierenga (2003) in their comprehensive strategy of hydrologic modeling and uncertainty for nuclear facilities and sites. Neuman and Wierenga emphasized the necessity to carry out the model simplification a systematic and comprehensive manner. In this work, simplifications of model structure and of model parameter determination are considered. Simplifications related to the numerical solution of model equations are outside the scope of this report.

2.2 Model abstraction techniques for flow and transport models

This section present a systematic description of the model abstraction techniques applied in subsurface hydrology. The categories of model abstraction techniques relevant to flow and transport modeling are presented in Fig. 2-2. This classification is based on our personal experience and the extensive literature review that is given below. Although an effort has been made to compile a comprehensive list, the classes of techniques outlined in the next section are neither mutually exclusive nor mutually exhaustive. Two large targets of abstraction are (1) the model structure, i.e. the formal description of specific processes and their interactions that affect flow and transport variables, and (2) the parameter determination, i.e. the estimation of constant and functions serving as coefficients in model equations. Although model calibration is considered to be a mandatory procedure in flow and transport modeling applications, a preliminary estimation of model parameters and their variability parameters is useful in both setting initial values of parameters for the model calibration and using non-calibrated model in pilot studies and field campaign designs. Each category of the classification in Fig. 2-2 represents a variety of specific procedures and techniques that are identified in following sections.

2.2.1 Hierarchies of models

A pre-defined hierarchy of models has been suggested for flow and transport in structured media (Altman et al., 1996). Figure 2-3 shows a schematic representation of increasingly complex models that may be used to simulate preferential flow and transport in macroporous soils or unsaturated fractured rock. The simplest bucket-type model describes the accumulation of water in soil matrix and its discharge when the water content exceeds the field water capacity or during evaporation periods (Fig. 2-3a). The classical approach for simulating flow/ transport processes in vadose zones devoid of macropores or fractures is to use the Richards equations for variably-saturated water flow and the advection-dispersion equation for solute transport (Fig. 2-3b). The simplest situation (Fig. 2-3c) for a fractured medium arises when the Richards and advection-dispersion equations are still used in an equivalent continuum approach, but now with composite (double-hump type) hydraulic conductivity (permeability) curves of the type shown in Mohanty et al. (1997), rather than the classical curve for the relative permeability shown in Fig. 2-3b. More involved dual-porosity type models (Fig. 2-3d)

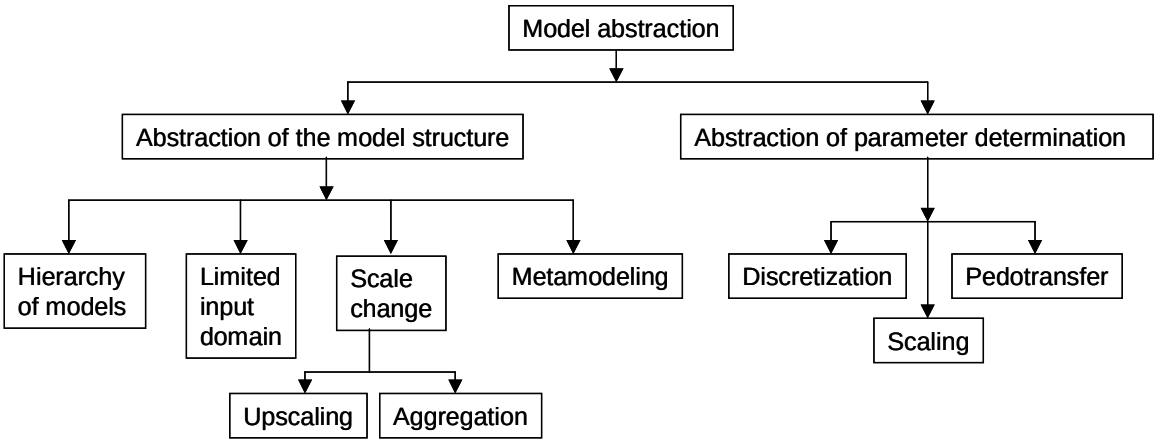

Figure 2-2. Categories of model abstraction techniques relevant to flow and transport modeling in subsurface hydrology.

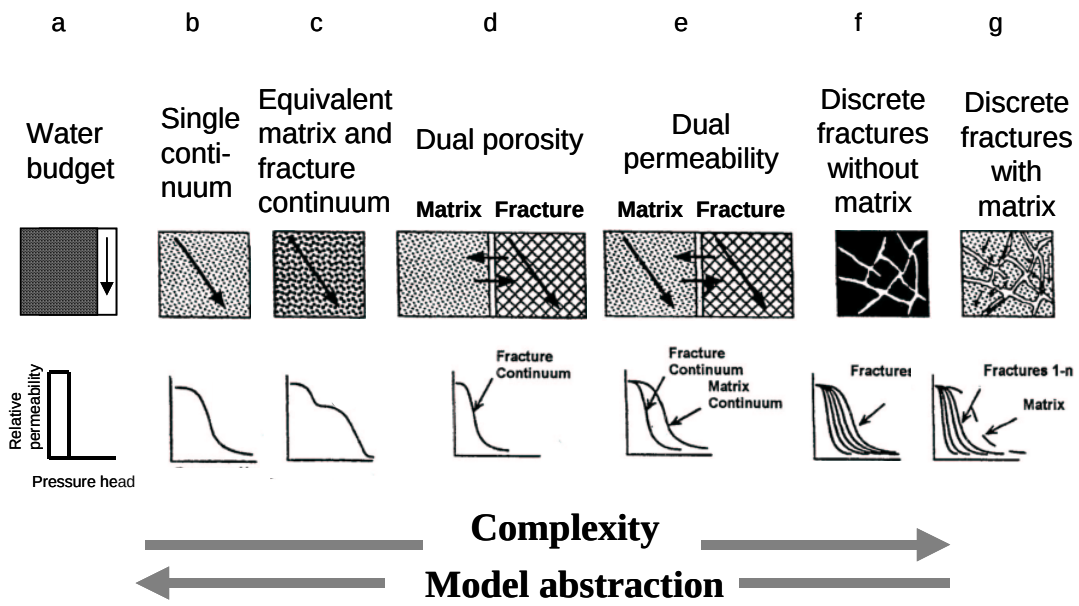

Figure 2-3. Hierarchy of models to simulate water flow and solute transport in structured soils or in unsaturated fractured rock. Modified from (Altman et al., 1996).

result when the medium is partitioned into fracture and matrix pore regions, with water and/or solutes allowed to exchange between the two liquid regions (e.g., Ventrella et al., 2000). Different formulations of this type are possible. For example, one could permit transient variably saturated flow in the fractures only, while allowing water to exchange between the fracture and matrix domains. The latter situation leads to both advective and diffusive exchange of solutes between the fractures and the matrix but still without vertical flow in the matrix (e.g., Zurmühl and Durner, 1996; Zurmühl, 1998). Dual-permeability models (Fig. 2-3e) arise when water flow occurs in both the fracture and the matrix domains. Examples are models by Pruess (1991), Gerke and van Genuchten (1993) and Jarvis (1999). These models use different formulations for the exchange of water between the fracture and matrix regions. In some models (e.g., Wilson et al., 1998) more than two domains are considered, each one having its own hydraulic properties. The modeling approach can be refined further by considering transient flow and/or transport in discrete fractures without (Fig. 2-3f) or with (Fig. 2-3g) interactions between the fractures and matrix. The latter approach is based on the assumption that the flow and transport equations of the fracture network can be solved in a fully coupled fashion with the corresponding equations for the matrix (e.g., Therrien and Sudicky, 1996).

Subsets of this hierarchy that have been used most often include using two region and one-region models of pore space (Nielsen et al., 1986). Other examples of hierarchies use non-equilibrium and equilibrium models for contaminant sorption and transport,as illustrated by Valocchi (1985).

2.2.2 Delimited input domain

This class of model abstraction techniques relies on the observation that some features or processes may be not relevant for a given class of scenarios or for a given set of model outputs. A reduction of the spatial dimensionality is one application of this technique. Two-dimensional representations can be redundant and 1D representation may suffice as shown by Wang et al. (2003). In another example, Guswa and Freuberg (2002) explored a possibility to use the 1D model to characterize solute spreading in the environment with low permeability lenses; they found that 1D macroscopic advective-dispersive equation well matched the results from the 2D model when the equivalent conductivity of a domain was less than the geometric mean conductivity, This example shows that one should expect a change in parameter values if the dimensionality is reduced.

Delimiting input domain may allow rejection of the models in the hierarchy of Figure 2-2 based upon the type of flow being encountered. This is schematically illustrated in Fig. 2-4 (Nimmo, 2002) for the maximum rate of preferential flow over various distances for a large number of field tracer experiments. The absence of saturated or near-saturated flow may make it possible to use the equivalent continuum models, whereas the presence of saturated flow may warrant the use of the dual-porosity or dual-permeability flow and transport models. The presence of substantial periods of unsaturated flow and isolated events of saturated conditions may motivate the use of different models for each of the two types of flow. This use of two different models may be simpler than invoking one integrated model for continuous simulation of both types of flow. Similarly, some types

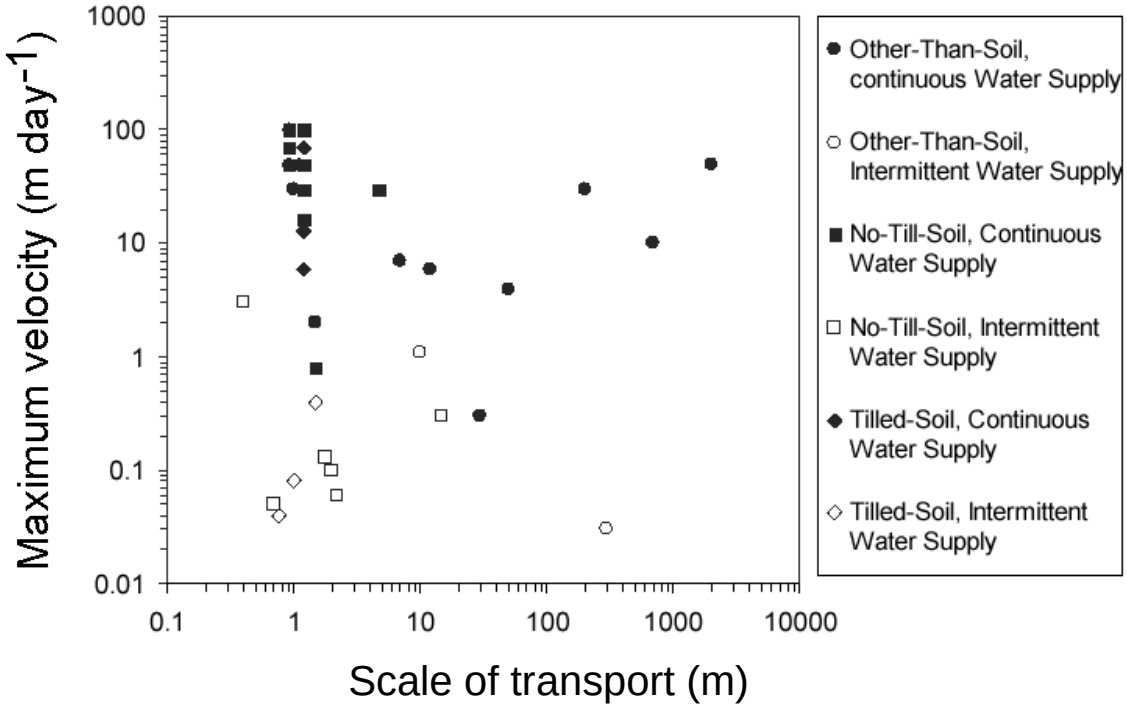

Figure 2-4. Maximum tracer velocity in 34 documented field experiments with variably saturated soils (after Nimmo, 2002, with permission).

of soil water regimes make the steady state flow assumption applicable to the solute transport simulations at filed scale (Destouni, 1991).

The behavioral pattern of the biotic component may also allow abstraction of this type. Guswa et al. (2002) have studied the soil water regime in presence of plants and have found that the mechanistic soil water flow model can be abstracted to the simple bucket-filling model if the plant can extract water from locally wet regions to make up for roots in dry portions of soil column.

2.2.3 Scale change

Scale is a complex concept having multiple connotations. A notion of *support* is important to characterize and relate different scales in subsurface hydrology. Support is the length, area, or volume for which a single value of porous media property is defined and no variations in this and other properties are taken into account. Size of an individual sample and size of a discrete spatial element in the flow and transport model present typical examples of the support. The term "resolution" is used for supports defined in terms of length, and the term "representative elementary volume" is applied for supports defined as volumes. Terms "pixel size" and "grid size" are also used to define resolution. An area or a volume that are sampled with given support determine the *extent* of measurements. Distances between sampling locations define *spacing.* Blöschl and Sivapalan (1995) suggested using the triplet *support-spacing-extent* to characterize the scale of the hydrologic study. In practice of the vadose zone research, supports of core scale, soil profile, or “pedon,” scale, field scale, and watershed scale are defined operationally, and the increase of the linear size by about two orders of magnitude corresponds to the transition from one of these scales to the next, coarser one (Fig. 2-5). Change in spatial scale is usually reflected in the change of temporal scale as shown in this figure.

The transition from a fine to the next, coarser scale in this sequence is referred below as the scale change.

2.2.3.1 Scale change – upscaling

This category of model abstraction recognizes the need in altering model structure when the spatial scale changes. Model equations, variables and parameters change.

Specific techniques of such abstraction employ statistical properties of flow and transport at the fine scale to derive the coarse-scale models. Development of upscaling techniques is very actively researched in hydrology, and a multitude of techniques is developed (Govindaraju, 2002; Rubin, 2003). Most of the upscaling techniques use the some assumptions about the spatial variability of flow, transport and porous media properties. However, no standard technique exists to date. The differences among the techniques are also related to the need of introducing additional, so-called constitutive relationships to compensate for the loss of information that occurs after upscaling (Cushman, 1990).

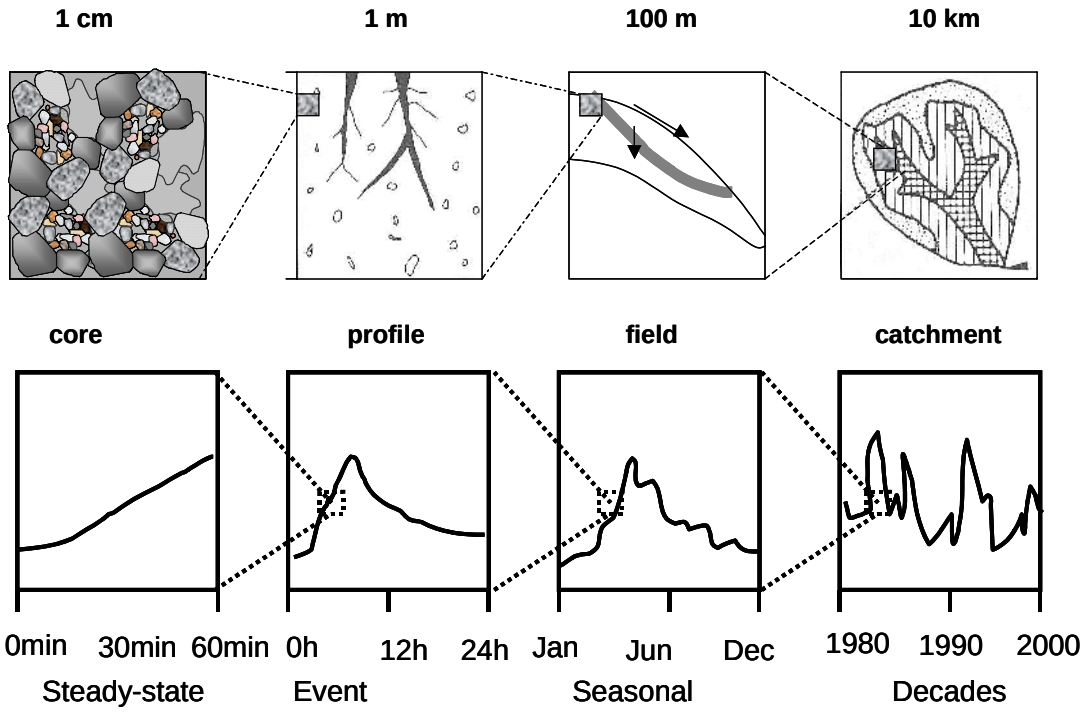

Figure 2-5. Spatial and temporal operational scales in hydrology (after Blöschl and Sivapalan, 1995). With permission, John Wiley & Sons, Inc.

Different constitutive relationships lead to different techniques. Recently it was suggested to carry out upscaling in the probability space (Guadagnini and Neuman, 1999). With this

method, the effective parameters of the flow model depend on the available information about the porous media properties at the fine scale.

Numerical stochastic simulations are also employed in upscaling. In this case the coarse-scale flow and transport is estimated not from upscaled equations but directly from Monte Carlo simulations at fine scale. Recently, using Monte-Carlo simulations became instrumental in upscaling multiphase flow and transport models for unsaturated zone simulations (i.e., Ewing et al., 2000; Li et al., 2003). Khaleel et al. (2002) used Monte Carlo simulations of flow and transport to represent a heterogeneous unsaturated medium with its homogeneous equivalent. Simulations showed that such upscaling also requires introducing new constitutive relationships, such as dependences of hydraulic conductivity anisotropy and dispersivity on moisture content or tension.

Upscaling may involve also the reduction of dimensionality, as demonstrated by Kumar (2004) for upscaling two-dimensional Richards equation to its one-dimensional coarse-scale analog.

To be efficient, all mentioned upscaling techniques have to use the correct statistical representation of rare features at the fine scale, because these features often govern the media behavior at the coarse scale. For example, if hydraulic conductivity is distributed lognormally, then five percent of fine scale representative elementary volumes (REVs) will conduct 95% of all flow, and relatively rare REVs will control the flow at the coarse scale. Macropores are rare in soils and are easy to miss during sampling, but they control water flow and solute transport in soil profile.

The scale finer than the core/core scale is usually referred to as the pore scale. This scale is not of an immediate interest in engineering applications of the flow and transport modeling Nevertheless, an active upscaling research at this scale is being done towards better understanding and modeling of flow and transport at core scale with the possibility to extend upscaling techniques to coarser scales. One example is to combine individual stream tubes into a stochastic transport model (Jury, 1990). Recent developments include modeling solute transport under assumption that individual solute particles undergo non-Brownian motion in a complex pore space. In this case, the classical convective-dispersive equation becomes inapplicable, and other models need to be used. Solute particles in complex and/or reactive pore space may experience long waiting times before making a “leap” in the general direction of transport. Such individual behavior can be integrated using models of fractional kinetics (Benson et al., 2001) or more general continuous-time random walk models (Berkowitz et al, 2001). A similar approach to modeling of the flow in unsaturated soil is also plausible (Pachepsky et al, 2003). Although these methods have been mostly researched to understand transport at the core scale, they are increasingly applied at the coarser scales. For example, the diffusion-limited aggregation model provided a good description of solute transport in soils with

the distance-dependent dispersivity (Flury and Flühler, 1995). Fractional kinetics models have been successfully used to simulate the MADE tracer experiment (Benson et al., 2001).

2.2.3.2 Scale change – aggregation

Change of scale also leads to a change of the model governing equations and parameters with these techniques. However, unlike in upscaling, no relationship is assumed between model parameters at the fine and at the coarse scales. Parameters of the coarse-scale model are deemed to be lumped, and field data are needed to calibrate such coarse scale models. An example is using a bucket-type soil water flow model (Fig. 2-3a) at the field scale whereas Richards equation is used at the soil profile scale (Fig. 2-3b). Soil water retention is parameterized with the field capacity at coarse scale and with soil water retention curves at the fine scale. No relationship between the coarse scale and fine scale parameters exists, because there is no reliable relationship between the water content at field capacity and soil water content at any specific values of soil water tension (Fig. 2-6).

Aggregation can be also done without the change in the model equations by combining several soil horizons or geologic strata. In flow and transport in vadose zone, one common application is to replace a heterogeneous soil profile with an equivalent homogenous profile while retaining the Richards equation as a flow model (Zhu and Mohanty, 2002). In such case, flow and transport parameters are still lumped. Attempts to determine the effective hydraulic properties of the equivalent soil profile from the layer properties have shown that the effective properties depend not only on layer properties but also on the type of the predominant water regime (infiltration or evaporation). Many examples of distributed watershed modeling with various degree of sub-watershed aggregation were recently developed in surface hydrology. Boyle et al. (2001) showed that aggregation from eight to three sub-watersheds does not worsen the model performance whereas aggregating all sub-watersheds in one does.

Aggregation should be distinguished from the model coarsening, when the support is increased without explicit recognition of the increasing non-homogeneity within it. In the latter case, no model abstraction is actually performed.

2.2.4 Metamodeling

This is a group of abstraction methods that use results of multiple simulation runs to extract the information helpful to simplify a complex model.

Metamodeling literally means modeling a model. Also known as the repromodeling (Meisel and Collins, 1973) or the response surface modeling, the metamodeling creates a relatively small and simple empirical model intended to mimic the behavior of a large complex model, that is, to reproduce the object model's input-output relationships. (Davis and Bigelow, 2003). A common way to develop a metamodel is to generate "data" from a

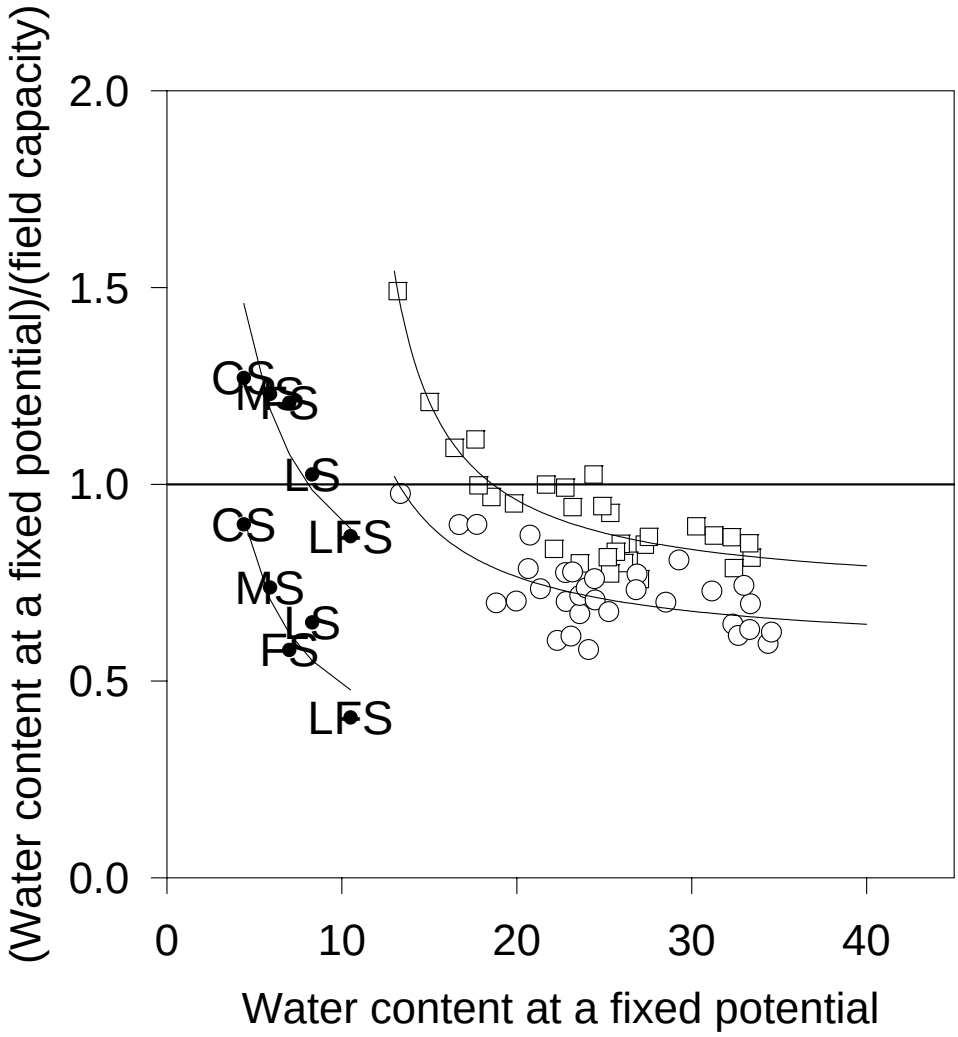

Figure 2-6. Soil field capacity and soil water retention at fixed matric potentials; CS - coarse sand, MS - medium sand, FS - fine sand, LFS - loamy fine sand at -5 and -10 kPa sol matric potentials (data from Rivers and Shipp, 1978); ○ and □ - topsoil and subsoil of varying texture at -33kPa (data from Haise et al., 1955).

number of large-model runs and then to use the statistical methods to relate the model input to the model output without attempting to understand the model's internal working.

Artificial neural networks (ANN) became a popular means to build metamodels because they are powerful approximators and, as such, are used to relate multiple input variables to outputs from the complex model (Hecht-Nielsen, 1990). Recent examples of application of the artificial neural networks to mimic MODFLOW output in a range of scenarios for particular remediation sites have been recently published (Kron and Rosberg, 1998; Masket et al., 2000).

The use of ANN requires development of a large number of training datasets covering the range of possible scenarios of forcing variables. Generating such datasets requires an extensive computing effort, but the computing with a trained ANN is several orders of magnitude faster than the simulation with the original model. Wang and Jamieson (2002) reported an example of regional wastewater planning in which, for reasons of computational efficiency, an artificial neural network was employed to replicate the process-based model in multiple evaluations of the model output during optimization aiming to determine both the best sites and individual discharge standards.

ANN can be used to simulate not only the model output of interest, but also the results of any part of computations performed during a model run. Hassan and Hamed (2001) demonstrated the use of ANN to predict particle trajectories in a particle-tracking algorithm in simulation the plume migration in heterogeneous media.

Using classification and regression trees (sometimes called decision trees) is another heuristic methodology for creating metamodels. It gains popularity because of the transparency of results, ability to retain only significant inputs-predictors, and possibility to work with categorical, or ordinal inputs. No applications of regression trees in subsurface flow and transport modeling have been found, although they were successfully used in hydraulic parameter estimation from readily available data (Rawls and Pachepsky, 2002).

Based on pure statistics, metamodels undoubtedly have problems that may be either minor or major, depending on the context and on the statistical methods used. Davis and Bigelow (2003) listed (a) the failure to tell a story, that is inability of metamodels to provide a sensible explanation of results to a decision-maker, (b) the failure to represent multiple critical components; if a system failure is modeled and the system fails if any of several components fail, then metamodels may not simulate that, instead predicting that a weakness in one component can be compensated by greater strength of another, (c) non-reliable sensitivity of metamodels, and (d) shortcomings in representing the whole input space; statistical metamodel may emulate base-model results reasonably well on average, but fails badly in what appears to be the obscure corners of the input space. Those authors suggest to use statistical metamodels based on the physical phenomena represented in the object model, knowledge of the structure of the object model, or a combination.

2.2.5 Discretization

In some cases, using a set of discrete values of parameters instead of their continuous values leads to substantial simplifications. This simple but very important abstraction is commonly used in flow and transport modeling by employing soil and geological textural maps to define soil or sediment "units" that have the same hydraulic properties - even though the real-world hydraulic properties may gradually change from one unit to another. Zonation of hydraulic conductivity was found efficient when geophysical and hydrogeological data were combined (Hubbard et al., 2003). Automated zonation is now feasible (Tsai, 2003) that gives an advantage of no need in assuming a specific equation to simulate the dependence of parameters on the spatial coordinates.

2.2.6 Scaling

When the model coarsening is performed, i.e. the support is increased but no changes in the model structure is made, the scale-dependence in model parameters will be most probably encountered because the larger units of the hierarchical structures in soils or sediments are enclosed in the extended support volumes. This kind of scale-dependence has been documented for the groundwater flow (Neuman and Di Frederico, 2003). A similar scale-dependence can be clearly observed in variably saturated soils as shown in Fig. 2-7, 2-8, and 2-9. Whereas in groundwater studies scaling of dispersivity was shown to obey the universal scaling law (Neuman, 1990, 1995), the scaling exponents of dispersivity in unsaturated soils appear to be quite variable (Fig. 2-7). The scale-related increase in saturated hydraulic conductivity depends on whether an aggregation of soil horizons has actually occurred in experiments (Fig. 2-8). Substantial differences in core-measured and field-measured water retention (Fig. 2-9) show that the applications of the core scale unsaturated flow parameters at field scale require cautious consideration.

Scaling relationships for the parameters of the unsaturated flow have been documented with the support range of about 1 to 1.5 orders of magnitude (Fig. 2-7, 2-8, and 2-9). They can provide useful relationships for modeling within the core-pedon range of scales if a model is coarsened, i.e. discretization in computations is changed without change in model equations. In such case, scaling provides means of using core scale measurements to infer parameters for increased support size

2.2.7 Parameter estimation with pedotransfer functions

Parameters of the water flow models in variably saturated porous media are nonlinear functions of matric potential or water content. These parameters are notoriously difficult to measure. Measurement of water and solute fluxes in unsaturated soils remains a research issue; no routine methods have been implemented to-date. Therefore, defining parameters using the inverse methods is often problematic. A substantial effort has been made to estimate these parameters from the data available from soil survey or borehole logs. The empirical functions used for such estimating are often called pedotransfer functions. An extensive work has been done to develop them. Proceedings of two International conferences (van Genuchten et al., 1992, van Genuchten et al, 1997) and the

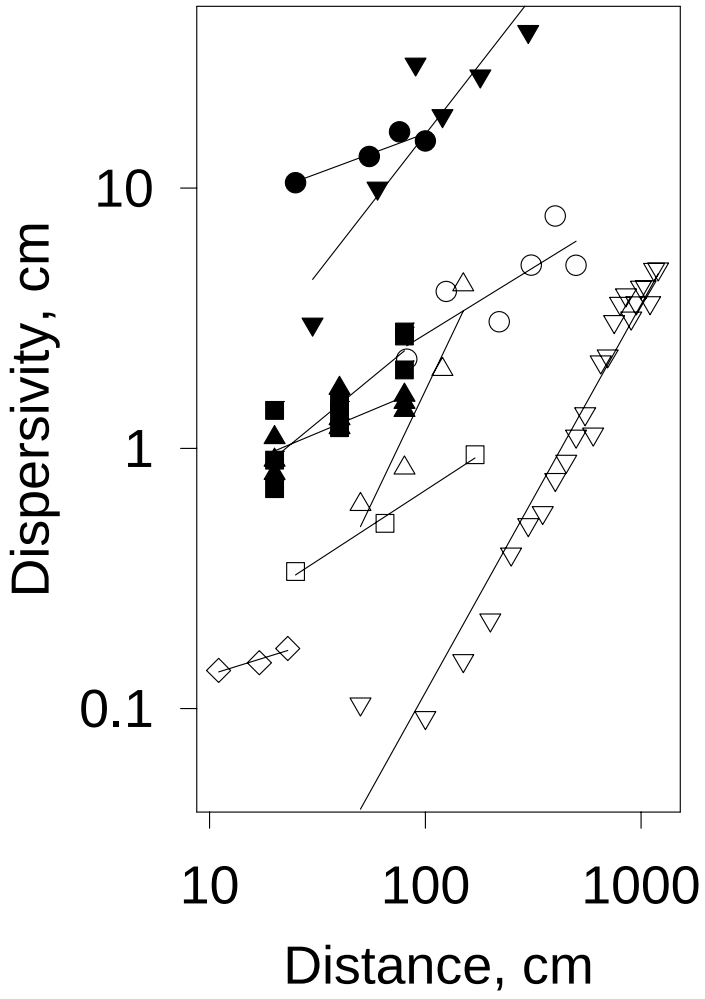

Figure 2-7. Observed dependencies of the solute dispersivity on distance in unsaturated soils (after Pachepsky et al., 2000). With permission, SSSA. Different symbols show data of different authors.

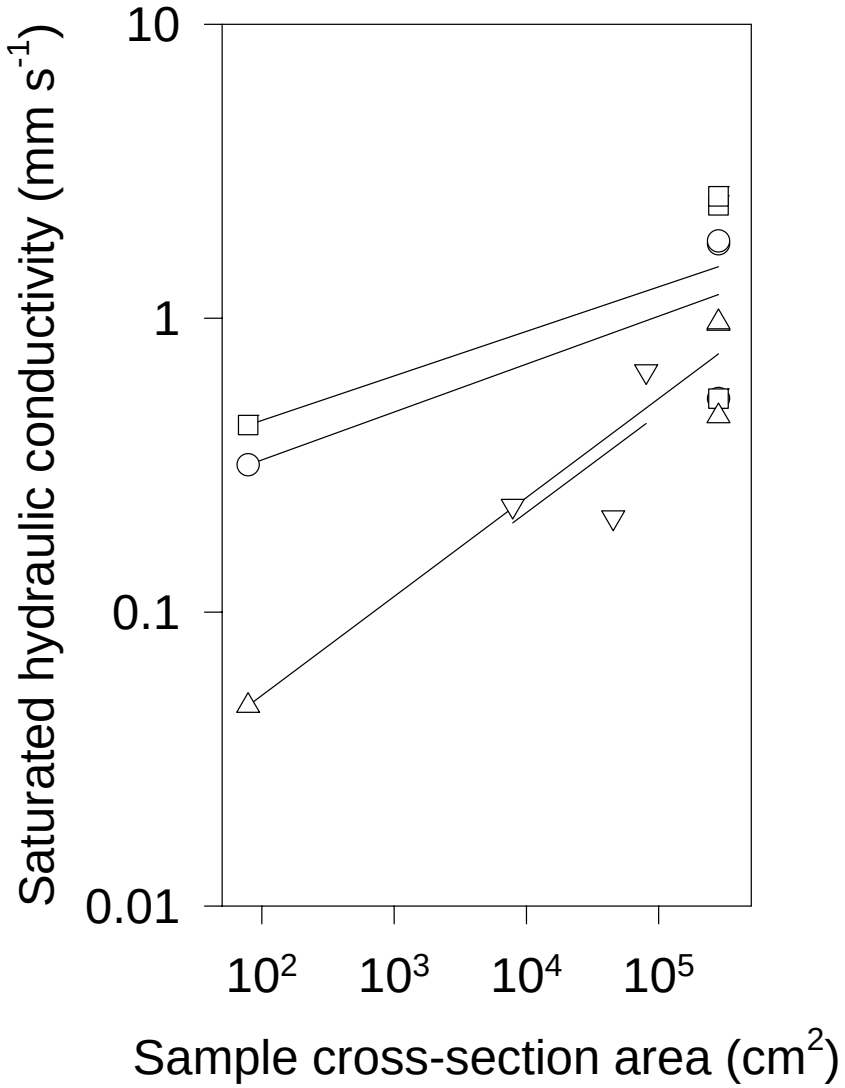

Figure 2-8. Examples of scale-dependence in soil saturated hydraulic conductivity; ○- loamy sand, Ap horizon, □ - loamy sand, A_2 horizon, Δ - sandy clay loam, B_1t horizon, ∇ - clay loam, B_1t horizon (after Pachepsky and Rawls, 2003). With permission, Blackwell Publishing.

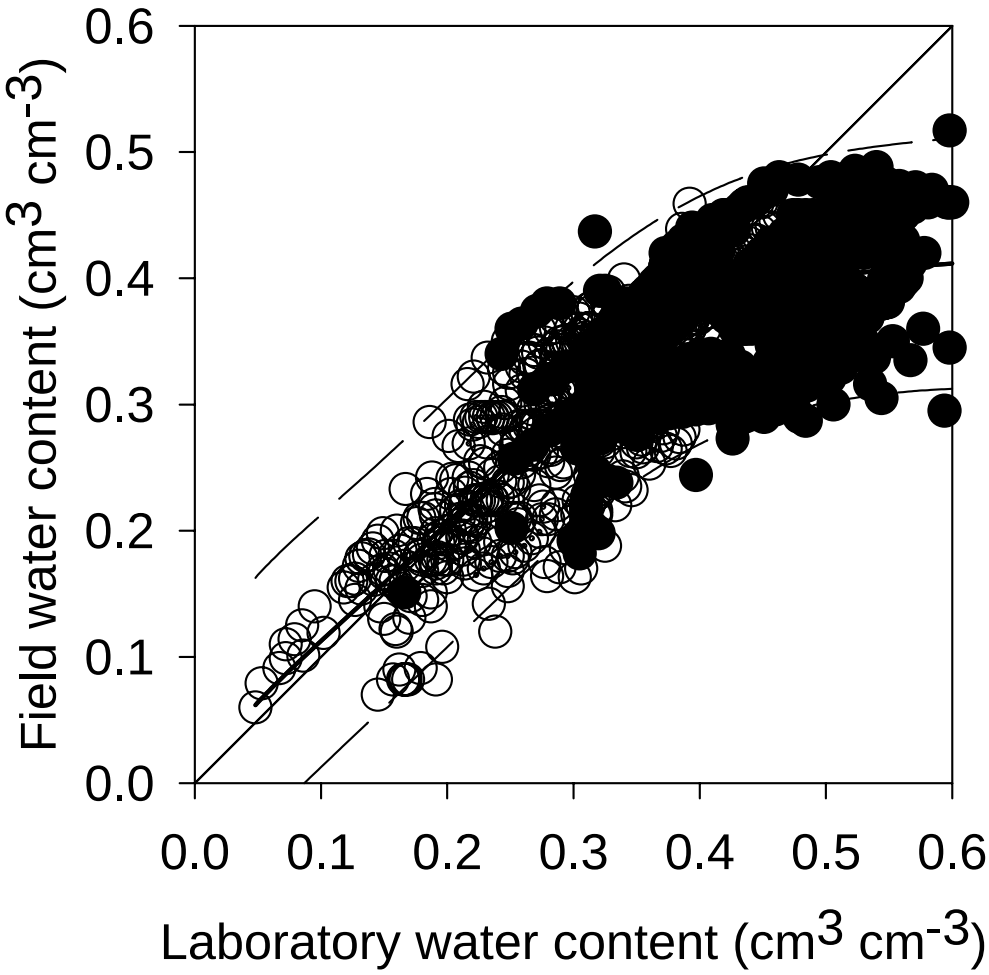

Figure 2-9. Relationship between field and laboratory water contents at the same soil-water matric potential; ○ - samples with sand content more than 50% , ● - samples with the sand content less than 50%; lines show the quadratic regression (_____) with 95% prediction interval (_ _ _). Data from the UNSODA database (Leij et al., 1996). With permission, SSSA.

first book on this topic (Pachepsky and Rawls, 2004) provide a panorama of these fast-developing studies.

2.2.7.1 Estimating water flow parameters

For water retention, one approach consists in estimating soil water contents at several soil water potentials with a separate regression equation for each potential. Rawls et al. (1982) used the US Cooperative Soil Survey Database to develop 12 regression equations to relate soil water contents at 12 matric potentials to sand, clay and organic matter contents (Table 2-3). Later, a similar set of equations has been developed to utilize the values of bulk density along with sand, clay and organic matter contents (Table 2-4). Other predictive equations developed from large databases are presented in the Appendix A.

Another approach to estimate soil water retention involves estimating coefficients in equations that express dependence of soil water contents on soil water potential. The equations that are most often used are the Brooks-Corey equation (Brooks and Corey, 1964):

$$\frac{\theta - \theta_r}{\phi - \theta_r} = \left(\frac{h_b}{h}\right)^{\lambda} \qquad (2\text{-}1)$$

and the van Genuchten equation (van Genuchten, 1980):

$$\frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{1}{\left[1 + (\alpha h)^n\right]^m} \qquad (2\text{-}2)$$

In Eq. (2-1) and (2-2), θ is the volumetric water content, h is the capillary pressure (the absolute value of the matric potential), ϕ is the porosity, θ_s is the saturated water content, θ_r is the residual water content, h_b is bubbling pressure, λ is pore size distribution index, α, m, and n are empirical shape-defining parameters. Various equations to estimate parameters in Eq. (2-1) and (2-2) are presented in the Appendix A. Recently, several new water retention equations were introduced to describe θ-on-ψ dependencies (Kosugi, 1994; Fayer and Simmons, 1995; Perfect et al., 1996, Bird and Perrier, 2000) but they are lacking the pedotransfer functions.

For the saturated hydraulic conductivity, K_{sat} , one approach is to estimate it from soil basic parameters without employing data on soil water retention (i.e., Bloemen, 1980; Loague, 1992). Rawls et al. (1998) assembled a database of more than 1000 experiments in the USA and showed that both median value of K_{sat} and its values on 25% and 75% probability levels can be defined for each USDA textural class if the samples are preliminary separated into high porosity and low porosity groups for each of the classes. These results are shown in Table 2-5. Referring the median Ksat values and the difference between K_{sat} values at 25% and at 75 % probability levels to the average clay content of the class results in a fairly well defined relationships (Fig. 2-10).

Another approach to the K_{sat} estimation benefits from soil pore space models as a source of the K_{sat} equation in which coefficients need to be estimated from soil basic properties. Ahuja and co-workers (1984) introduced a generalized form of the Kozeni-Karman

Table 2-3. Regression equation to predict soil water content at specific matrix potential: $\theta(\psi)$ (cm^3cm^{-3})= a + b∗sand(%) + c∗silt(%) + d∗clay(%) + e∗OM(%) (after Rawls et al., 1982)

Parameters of the linear regression					
ψ, kPa	a	b	c	d	e
-10	0.4118	0-0.0030	0	0.0023	0.0317
-20	0.3121	-0.0024	0	0.0032	0.0314
-33	0.2576	-0.0020	0	0.0036	0.0299
-60	0.2065	-0.0016	0	0.0040	0.0275
-100	0.0349	0	0.0014	0.0055	0.0251
-200	0.0281	0	0.0011	0.0054	0.0200
-400	0.0238	0	0.0008	0.0052	0.0190
-700	0.0216	0	0.0006	0.0050	0.0167
-1000	0.0205	0	0.0005	0.0049	0.0154
-1500	0.0260	0	0	0.0050	0.0158

Table 2-4. Regression equation to predict soil water content at specific matrix pressure: $\theta(\psi)$ ($cm^3 cm^{-3}$)= a + b$*$sand(%) + c$*$clay(%) + d$*$OM(%) + e$*\rho_b$(g cm^{-3}) (after Rawls et al., 1983)

Parameters of the linear regression					
ψ, kPa	a	b	c	d	e
-20	0.4180	-0.0021	0.0035	0.0232	-0.0859
-33	0.3486	-0.0018	0.0039	0.0228	-0.0738
-60	0.2819	-0.0014	0.0042	0.0216	-0.0612
-100	0.2352	-0.0012	0.0043	0.0202	-0.0517
-200	0.1837	-0.0009	0.0044	0.0181	-0.0407
-400	0.1426	-0.0007	0.0045	0.0160	-0.0315
-700	0.1155	-0.0005	0.0045	0.0143	-0.0253
-1000	0.1005	-0.0004	0.0045	0.0133	-0.0218
-1500	0.0854	-0.0004	0.0044	0.0122	-0.0182

Table 2-5. Saturated hydraulic conductivity (cm day^{-1}) at three probability levels for USDA textural classes (from Rawls et al., 1998)

USDA textural class	Soils with high porosity				Soils with low porosity			
	25%	50%	75%	Porosity cm^3cm^{-3}	25%	50%	75%	Porosity cm^3cm^{-3}
Sand	638.4	436.8	230.4	0.44	523.2	218.4	153.6	0.39
Fine sand	566.4	338.4	283.2	0.49	528.0	240.0	163.2	0.39
Loamy sand	468.0	295.2	201.6	0.45	184.8	98.4	74.4	0.37
Loamy fine sand	292.8	148.8	86.4	0.46	278.4	28.8	16.8	0.37
Sandy loam	312.0	134.4	72.0	0.47	74.4	31.2	0.37	0.37
Fine sandy loam	86.4	52.8	24.0	0.45	40.8	19.7	8.2	0.36
Loam	67.2	9.4	3.8	0.47	39.6	14.9	6.7	0.39
Silt loam	89.0	34.6	18.2	0.49	23.8	8.2	2.4	0.39
Sandy clay loam	121.2	18.5	4.8	0.44	26.2	6.7	2.4	0.37
Clay loam	31.4	10.1	5.3	0.48	9.1	1.7	0.5	0.40
Silty clay loam	25.0	8.9	5.5	0.50	33.6	11.8	5.5	0.43
Clay	14.4	4.8	2.2	0.48	16.6	4.3	0.7	0.40

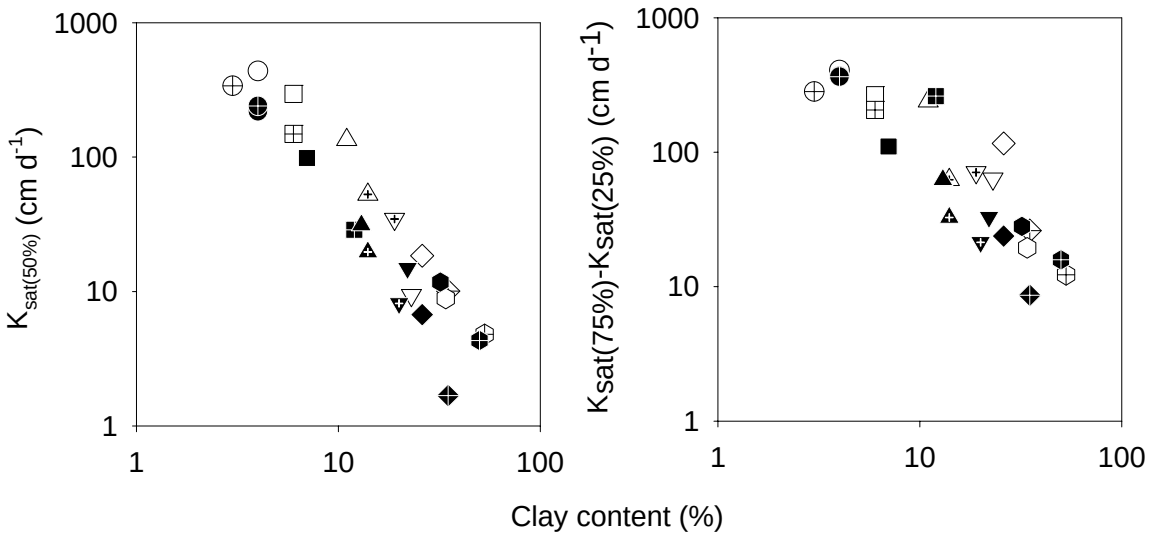

Figure 2-10. Saturated hydraulic conductivity K_{sat} grouped by textural classes and bulk density classes; (a) median values $K_{sat(50\%)}$ for the groups, (b) difference between third $K_{sat(75\%)}$ and first $K_{sat(25\%)}$ quartiles of the distributions within each group; data for high and low bulk density samples are shown by filled and hollow symbols, respectively; ○, ●-sand, ⊕, ⊕ - fine sand, □, ■ - loamy sand, ⊞, ⊞- loamy fine sand, △, ▲- sandy loam, ⊿, ▲ - fine sandy loam, ▽, ▼ - loam, ▽, ▼ - silt loam, ◇, ◆ - sandy clay loam, ◇, ◆ - clay loam, ⬡, ⬢ -silty clay loam, ⊕, ⊕ - clay.

model to relate the saturated hydraulic conductivity to the effective porosity:

$$K_{sat} = B\theta_e^{\ n} \qquad (2\text{-}3)$$

where θ_e is a difference between the total porosity and soil water content at -33 kPa, B and n are parameters which vary among different data sets (Table 2-6).

Estimates of the unsaturated hydraulic conductivity are predominantly based on pore space models that relate unsaturated hydraulic conductivity to the saturated hydraulic conductivity and to geometric parameters of the pore space inferred from water retention data.

A general form of such model has been proposed by Alexander and Skaggs (1984):

$$K(h) = K_{sat} S(h)^{l} \left\{ \frac{\int_0^{S(h)} h(\theta)^{-\gamma}\, d\theta}{\int_0^{1} h(\theta)^{-\gamma}\, d\theta} \right\}^{\delta} \qquad (2\text{-}4)$$

where S is the soil saturation, $S = (\theta-\theta_r)/(\theta_s-\theta_r)$, l, γ, and δ- parameters. When water retention is simulated with the van Genuchten equation (2-2), this equation with γ=1 and α= 2 can be transformed into the van Genuchten-Mualem equation:

$$K \ = K_{sat} (S)^{l} \left[1 - (1 - S^{1/m})^{m} \right]^{2} \qquad (2\text{-}5)$$

Pedotransfer functions need to be used to estimate both K_{sat} and $h(\theta)$ in Eq. (2-5) to evaluate the integral and calculate the unsaturated hydraulic conductivity.

Artificial neural networks (ANN) recently are useful for PTF development. The feed-forward multiplayer ANN, the most popular type for PTF development, has built-in equations (Pachepsky et al., 1996) and is no different in applications than any other regression equation. However, coefficients of this equation are never reported, and one typically has to use software with hardwired ANN equations in it. Rosetta is a Windows-based program that implements artificial neural network results published by Schaap et al. (1998), Schaap and Leij (1998), and Schaap and Leij (2000) and is available from http://www.ussl.ars.usda.gov/models/rosetta/rosetta.HTM. The program implements five PTFs in so-called hierarchical approach (H1, see also Schaap et al, 2001). This approach was chosen to maximize the accuracy of the PTF estimates given the particular data availability. All models in Rosetta estimate saturated hydraulic conductivity and van Genuchten (1980) retention and unsaturated hydraulic conductivity parameters. Through use of the Bootstrap Method, Rosetta is also able to estimate the uncertainty of the estimated hydraulic parameters. Minasny and McBratney (2002) developed the Neuropack software package (http://www.usyd.edu.au/su/agric/acpa/neuropack/neuropack.htm). This package differs from the previously discussed software since it is primarily intended to develop PTFs using neural network-based techniques with the user-supplied. No previously calibrated PTF is shipped with Neuropack except some sample data.

Table 2-6. Parameters of the Ahuja equation (2-3) to estimate saturated soil hydraulic conductivity

Source	B, 10^{-3} m s^{-1}	n
Ahuja et al., 1989	2.9	3.21
Franzmeier, 1991	5.0	3.25
Messing, 1989	1.5 – 9.5	1.8 – 3.0
Timlin et al. 1996	2.1	3.29
Rawls et al., 1992	2.8	4.0

In general, ANNs provide an excellent flexibility in mapping complex 'input-output' dependencies (Hecht-Nielsen, 1990). The use of ANNs has, however, some disadvantages compared with regressions. In particular, the equations built during ANN training are “opaque,” and ANNs do not distinguish inputs by their significance leaving the responsibility to select significant inputs to the user.

Estimating solute transport parameters remains a weak point of the PTF development. Although a progress has been made in estimating solute dispersivity for small soil cores (Griffioen et al., 1998; Oliver and Smettem, 2003; Minasny and Perfect, 2004), no methodology exists for such estimates at the pedon scale and at coarser scales.

2.2.7.2. Performance of pedotransfer functions

Performance of pedotransfer functions (PTF) can be evaluated in terms of their accuracy, reliability, and utility. In broad terms, the accuracy of a PTF can be defined as a correspondence between measured and estimated data for the data set *from which a PTF has been developed*. The reliability of PTF can be viewed as a correspondence between measured and estimated data for the data set(s) *other than the one used to develop a PTF*. Finally, the utility of PTF in modeling can be assessed in terms of the correspondence between *measured and simulated environmental variables*.

PTF accuracy has been characterized using various quantitative measures, such as the mean error, the standard deviation of the mean error, the mean squared error, the determination coefficient R^2, etc. With the same measure, accuracy of existing PTFs varies appreciably. Pachepsky et al. (1999) randomly sampled literature on PTF and found that soil basic parameters typically account for 80 - 90 % of the moisture content variance at capillary pressures of 330 and 15000 cm. The RMSE ranges from 0.02 to 0.07 $m^3\ m^{-3}$ at both levels of capillary pressure. The water content at 330 cm ranges from 0.04 to 0.50 $m^3\ m^{-3}$ and the water content at 15000 cm ranges from 0.02 to 0.35 $m^3\ m^{-3}$ for the most of soils. With such ranges, the PTF RMSE between 0.02 to 0.07 $m^3\ m^{-3}$ is sufficient in some cases, while in other cases it is unsatisfactory, and improvements in PTF accuracy are desirable. The hydraulic conductivity PTFs usually have the accuracy about half order of magnitude.

The reliability of PTFs can be estimated by re-sampling with the data set used in the PTF development, or by using an independent data set. The re-sampling methods generate numerous artificial data from the original experimental data and evaluate statistical properties of interest from its observed variability over all the generated artificial datasets (Good, 1999). Two methods, the bootstrap (Schaap, 2004) and jackknife (Pachepsky and Rawls, 1999) have been used so far. The re-sampling methods do not render any information about the reliability of PTF if it is applied in the region other than that the development data set has been collected at.

PTFs developed from regional databases produce quite adequate results in the regions with similar soil and landscape history. Water retention PTFs developed in Belgium (Vereecken et al., 1989) were the most accurate as compared with 13 others for the data

base of the Northern Germany (Tietje and Tapkenhinrichs, 1996). Water retention PTFs developed for the Hungarian Plain (Pachepsky et al., 1982) were applicable for the Caucasian Piedmont Plain (Nikolaeva et al., 1986). PTFs developed in Australia were more accurate for the Mississippi Delta as compared with other regional PTFs (Timlin et al., 1996a). It remains to be explored whether this observation holds for other cases, and which soil and landscape features have to be similar in two regions to assure the mutual reliability of the PTFs.

PTFs developed from the USA database by Rawls et al (1982) appear to be more robust than PTFs developed from the regional databases. When the reliability of the all-USA PTFs has been compared with the accuracy of several other regional databases, and the PTFs were ranked by their reliability, the all-USA PTFs usually had one of the highest rankings (Tietje and Tapkenhinrich, 1993; Kern, 1995; Timlin et al., 1996b). In general, it is not possible to predict whether a specific PTF will be applicable at a given site or not. An example of an interregional comparison of PTFs is given in Fig. 2-11 where 21 PTF is applied to estimate water contents at 33 kPa and 1500 kPa. (Wösten et al., 2001). This figure illustrates various situations that are encountered in PTF reliability estimations. The PTF #1 has the CEC of clay fraction as an essential input. An average value of the CEC/clay ratio=0.5 was applied to all data to compute the data in Fig. 2-11. PTF # 3 is inaccurate with Oklahoma data, but it is also not particularly accurate with the data used for the development, R^2 being only 0.58. The PTF # 5 is derived from data on 6 soils, and it fails when tested with the large database. Similarly The PTF # 7 is derived from data on only 43 soils. The PTF # 6 for the field capacity was developed with actual field data, and an attempt to equate the field capacity to the water content at 3 kPa was unsuccessful. PTFs # 2 and # 8 were derived for soils with the clay minerals that are distinctly different from those in Oklahoma. They tend to overestimate water retention at 15 bars; for the same reason PTF # 9, PTF # 13, and PTF # 19 tend to underestimate it. PTFs # 18 and 3 developed for tropical soils do not give satisfactory predictions of field capacity, but seem to perform well at 1500 kPa. Using the texture only leads to unsatisfactory results at low water contents (PTF # 17). Overall, PTFs # 11 and # 16 are the best.

The reliability of PTFs may be limited by differences in methods used to measure soil hydraulic properties that may have profound effect on results (Marion et al., 1994). One of the method-related factors affecting the PTF development is the support size as shown in Figures 2-7 and 2-8. Large data bases assembled from different sources to develop PTF often lose in their quality because of the diversity in measurement methods. There also may exist a limit in accuracy and reliability of PTFs caused by temporal variation of soil properties related to the changes in vegetation and soil management. Pachepsky et al. (1992) have reported 20% changes in water retention at -1 kPa and 5% changes in water retention at -30 kPa in soils under wheat during growing season in three different climatic zones. The amount of published data on temporal variations of soils hydraulic properties remains insufficient. The temporal component assessment may be required in PTFs to improve their reliability.

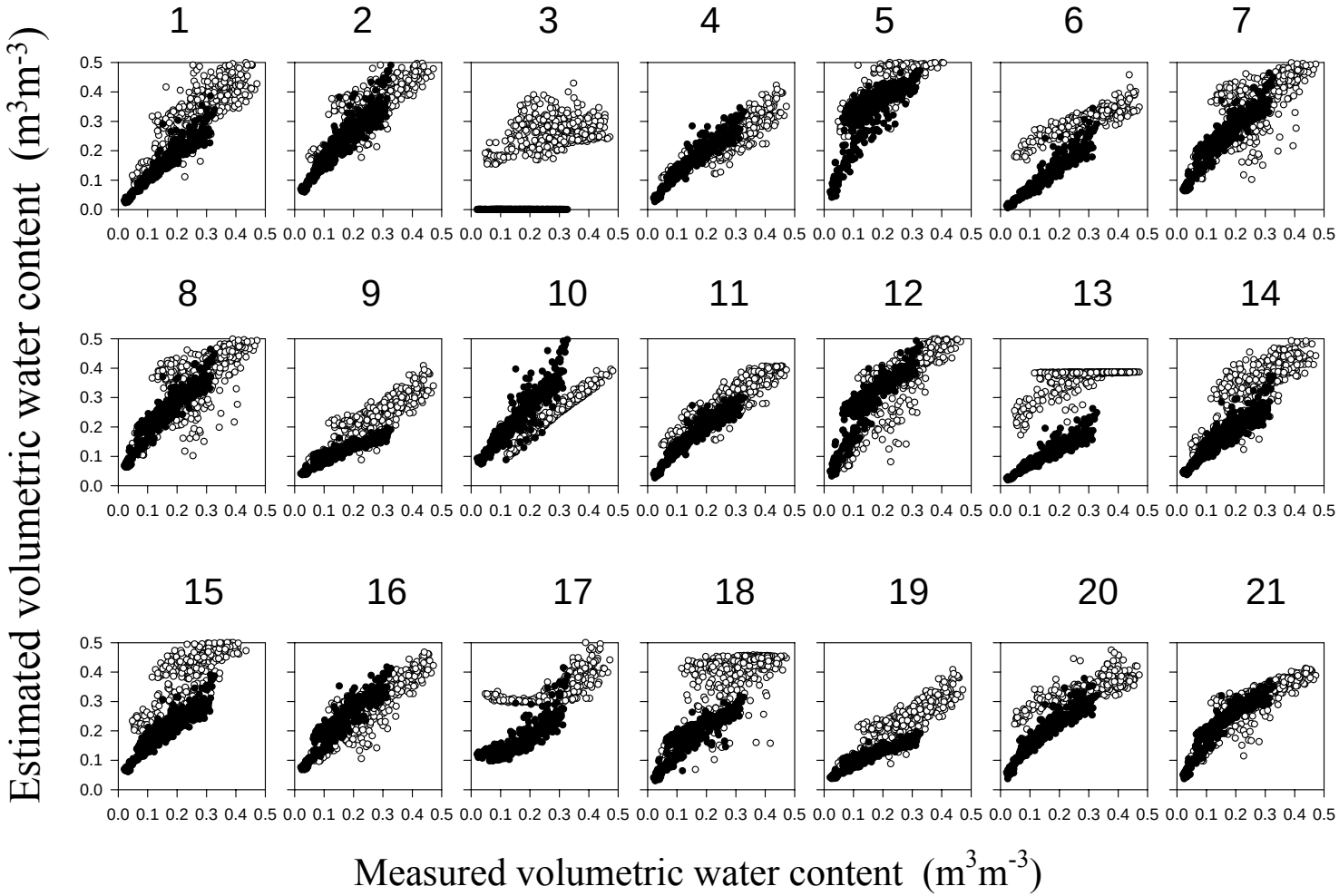

Figure 2-11. Estimating water contents at 33 kPa (○) and at 1500 kPa (●) with 21 PTFs developed in various region of the world Experimental data on soil water retention and soil basic properties have been extracted from the US Cooperative Soil Survey database for the state of Oklahoma. From (Wösten et al., 2001). With permission, Elsevier.

The reliability of a PTF is not directly related to its utility (Pachepsky et al., 1999). The latter is affected by sensitivity of the model to PTF predictions, and also by the uncertainty in other model inputs (Leenhardt et al., 1995). When the PTF uncertainty is factored in the modeling effort, variation in predictions of different PTFs has to be considered along with uncertainty of individual PTF predictions. As the experience of calibrating vadose zone models accumulates, more opportunities appear to compare calibrated and PTF-estimated hydraulic properties (Wang et al., 2003) as well as to use PTF predictions as initial estimates for model calibration (Jacques et al., 2002). Using several different PTFs instead of predictions of a single PTF appears to be a viable option to explore.

2.2.7.3 Emerging pedotransfer techniques

Developments in pedotransfer technologies are mostly associated with using spatially dense physical information related to the soil cover. Schematic of this research is shown in Figure 2-12. The use of topographic information is based on the hypothesis that there may exist some relationship between soil hydraulic properties and topographic variables (Pachepsky et al., 2001) because (a) the basic properties of soil are known to be related to landscape position, and (b) soil hydraulic properties are related to soil basic properties, Romano and Palladino (2002) used the terrain attributes to recalibrate PTF, and soil water retention exhibited strong dependence on terrain attributes in the work by Rawls and Pachepsky (2002) with the US National Cooperative Soil Survey data.

The use of hydrogeophysical, remote sensing, and crop yields data is based on a similar premise. The basic data on soils affect both sensor readings and soil hydraulic properties. Therefore, some relationship between these dense data and hydraulic properties can be expected. Smettem et al. (2004) presented an example of using airborne gamma radiometric sensing to estimate the clay content of surface soils and using a simple pedotransfer function (Smettem et al. 1999) to convert this information into a spatial representation for the slope of the Brooks and Corey water retention function. Timlin et al. (2003) related the yield maps to the soil field capacity.

Using dense auxiliary data in PTFs is an attempt to trade the data quality for data quantity. Since a dense coverage can be treated as an image, the image analysis techniques can be used for segmentation and classification, as well as for evaluation of spatial structure in soil properties. Also, the data assimilation techniques can be suited to combine the soil survey and sensor information (McLaughlin, 1995). PTFs developed with auxiliary data are probably highly site-specific and, therefore, useful for the sites they have been developed for. Nevertheless, the availability of sensor data can make the development of such PTFs a viable component of supplying parameters of the models of water flow in vadose zone.

Since the accuracy of any PTF outside of its development dataset is essentially unknown, the use of an ensemble of PTFs may be considered. Existence of several models that are developed and tested in one region but perform relatively poorly in other regions is fairly common in meteorology. Justifying a selection of a single model has become an

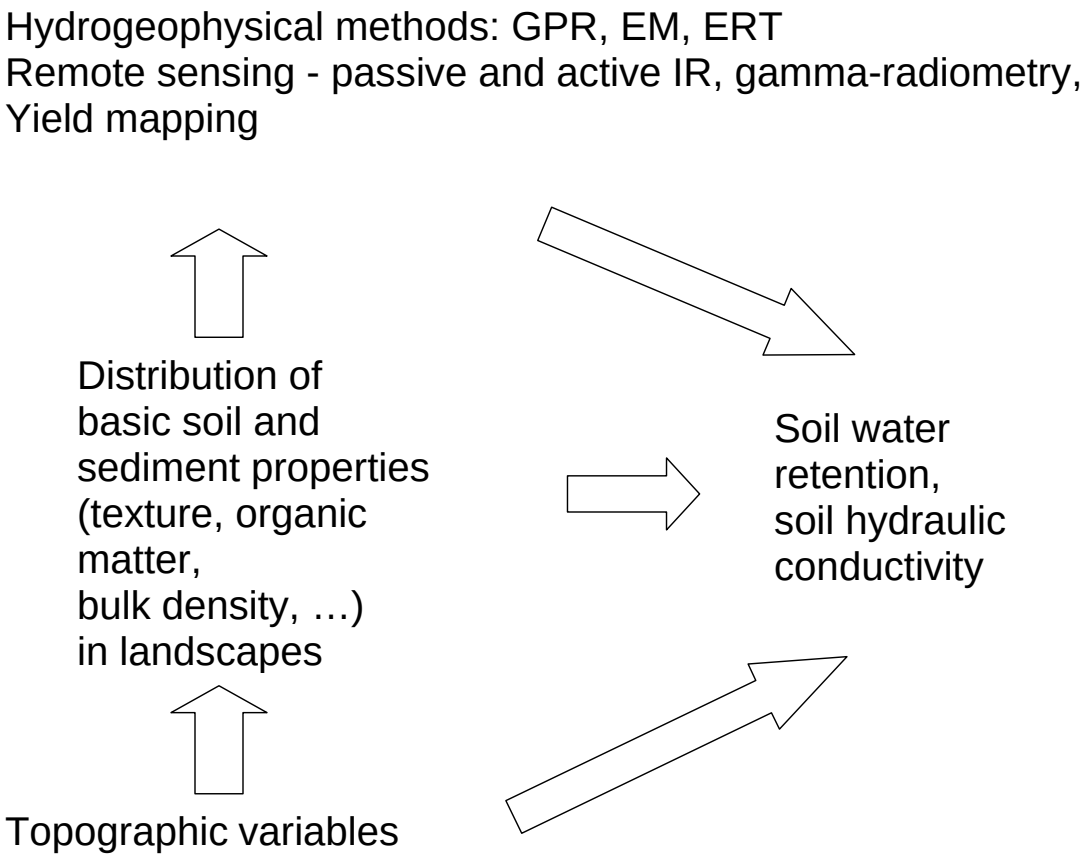

Figure 2-12. Using the dense auxiliary data to infer soil hydraulic properties.

unsolvable problem. The multimodel ensemble prediction method was developed during the last decade to address this dilemma (Molteni et al, 1996). The basic idea is to use a weighted average of prediction obtained from all the models in hand. Since there is no underlying theoretical formalism from which a probability distribution of model uncertainty can be estimated, some pragmatic approach must be sought. One such an approach relies on the fact that climate models have been developed somewhat independently by different research groups. An ensemble comprising such quasi-independent models is referred to as a multi-model ensemble (Palmer et al., 2004). Ensemble forecasts offer a way of filtering the predictable from the unpredictable through averaging – the features that are consistent among ensemble members are preserved, while those that are inconsistent are reduced in amplitude. What is even more important, the ensemble itself, as a sample from possible forecast outcomes, can be used to estimate the forecast uncertainty and the likely structure of forecast errors (Hamill et al. 2004).

Recently, the multimodel ensemble methods have begun to be used in subsurface hydrology. Ye et al. (2004) suggested averaging of spatial variability models in unsaturated fractured tuff in a situation when standard information criteria provide an ambiguous ranking of the models that does not justify selecting one of them and discarding all others. A limited testing outlined in Chapter 3 of this report suggests that the multimodel ensemble prediction of soil hydraulic properties is a promising technique.

2.3. Model abstraction implementation

2.3.1 Status of model abstraction in flow and transport modeling

The status of model abstraction in flow and transport modeling has been recently reviewed by Neuman and Wierenga (2003) They indicate that "…a widely practiced approach has been to simplify complex hydrogeologic systems for modeling purposes in an ad hoc and subjective manner. It is generally not clear that this approach captures adequately all aspects of the complexity that have a significant impact on the problem under consideration. It is also not clear that such models have the space-time resolution required to render them comparable and compatible with site data. Without this, the models remain interesting logical constructs, however, may not be not valid representations of actual site conditions and may yield unreliable results."

Currently, solid justification for a model selection is rarely provided. This refers to both the selection of the conceptual flow and transport models, i.e. process descriptions, flow and transport domain discretization, forcing and initial conditions, and the selection of the method to estimate the hydraulic properties. Comparison of performance of two or more models is rarely made. Most model abstraction methods are not comprehensively tested for modeling of flow and transport in variably saturated soils and sediments. The following describes a systematic approach to the model abstraction.

2.3.2. Model abstraction process

General requirements to the model abstraction process are objectiveness, systematic implementation, comprehensiveness, and efficiency (Neuman and Wierenga, 2003). Model abstraction (MA), as defined in this report, is not about developing models, but rather about modifying or replacing them. The MA process starts with an existing *base* model that can be calibrated and used in simulations. The *key output* of the model id defined that provides the necessary and sufficient information to decide on issues of interest.

The base model may need abstraction for one or more of the following reasons:

- The base model includes a complex description of processes that cannot be observed well and yet need to be calibrated; the calibrated values of parameters of those processes are very uncertain.
- The base model propagates uncertainty in the initial distributions, parameters, and forcing in a manner that creates an unacceptable uncertainty of the key output.
- The base model produces inexplicable results in terms of the key output.
- The base model requires an unacceptable amount of resources for computations, data preprocessing, or data post-processing, e.g. the base model is not suitable be used as a part of an operational modeling system that requires real-time data processing.
- The base model lacks transparency to be explicable and believable to the users of the key output.

The model abstraction process includes the following steps.

1. Justify the need for the model abstraction
2. Review the context of the modeling problem
3. Select applicable MA techniques.
4. Determine MA directions and simplify model in each direction.

The following provides additional information on each of the steps.

2.3.2.1 Justify the need for model abstraction

Any model simplification requires calibration of the abstracted simpler model and its confirmation with multiple model runs. Therefore, it is a separate modeling project that also demands resources, and the need in model abstraction has to be justified.

The lack of transparency of the base model is the reason for abstraction that arises from the perception of potential users or critics of simulation results. The justification for abstraction in such case is outside of the modeling project per se. The incomprehensible results of simulations with the base model or its unacceptable resource demand of base model are the reasons that essentially preclude the completion of the modeling project. The justification for model abstraction is simply the need to finish the project.

The need in abstraction may stem also from the uncertainty in calibrated parameter values or in simulation results. The decision to carry out the model abstraction is based on statistics of parameter estimates and the key output, as discussed below.

2.3.2.1.1 Need in abstraction because of the limited observation ability

A case study of the effect of observability on the model performance has been recently presented by Kekkonen and Jakeman (2001). They compared two catchment hydrologic models. Each model had six parameters. Both models were calibrated with streamflow data. One model explicitly simulated evapotranspiration and was parameterized in this part, whereas another model did not contain this process as described directly. The model without the evapotranspiration submodel provided, in general, more accurate reproductions of streamflow, even on the independent dataset. This difference was more pronounced at drier catchments. This result was attributed to the calibration with the streamflow data that contained relatively little information about evapotranspiration. Authors concluded that the more process complexity one wanted to include in the model structure, the more types of data and larger information content were required to estimate the process parameters and to test the model performance. When only rainfall-runoff data were available, it was difficult to justify a substantial conceptualization of evapotranspiration.

The model calibration gives the necessary quantitative information to decide on the reliability of the parameter estimates with respect to their importance for the key output. A parameter estimate is acceptable if either the uncertainty of this parameter is adequate as compared with estimates of other parameters, or it is unimportant for predictions of interest. A diagnostic table (Table 2-7) has been developed based upon the classification proposed by Hill (1998). In Table 2-7, diagnosis is based on (1) statistical characteristics of the parameter, and (2) the model prediction uncertainty. In this analysis, the basic diagnostic measures are: (1) the composite scaled sensitivity of the parameter; (2) coefficient of variation of the parameter; (3) the scale sensitivity of prediction; (4) correlation coefficients between estimated parameters, and (5) the updated correlations coefficients of the parameter. These statistics are directly available or can be computed according to Hill (1998) from the output of calculations with the available from the most widely used universal parameter estimation codes, such as UCODE (www.mines.edu/igwmc/freeware/ucode/) and PEST (www.parameter-estimation.com/html/pest_overview.com), as well as from the inverse modeling with HYDRUS (Simunek et al, 1999). The diagnostic table 2-7 suggests using the model abstraction where the reliability of calibrated parameters is low. Model structure abstraction or model parameter abstraction is suggested depending on whether the parameters are influential or not.

In variably saturated porous media, the complexity of process description related to the specific parameter or to the group of parameters can preclude a reliable estimation of these parameters with the additional data from available sources. For example, the reliable estimation of parameters of the mobile-immobile zone solute transport model has been found impossible in some field studies (Koch and Flühler, 1993; Field and Pinsky,

2000). However, if some additional experiments can potentially provide the information to improve the parameter estimates, the decision to carry out the model abstraction should be postponed until the data from such experiments are available. An (exaggerated) example is the soil water flow model that includes preferential flow but has been

Table 2-7. Potential need for model abstraction from the parameter uncertainty standpoint based on Hill (1998) analysis.

Importance of the parameter to the key output	Reliability of the parameter estimate	
	Precision of the parameter estimate	
	Poor precision: large C-scaled sensitivity[¶], large coefficient of variation, large confidence interval	Good precision: small C-scaled sensitivity, small coefficient of variation, small confidence interval
Not important: small P-scaled sensitivity[⌂]	Abstraction of parameter determination may be useful	No indications to use model abstraction
Important: large P- scaled sensitivity	Abstraction of the model structure may be useful	No indications to use model abstraction
	Uniqueness of the parameter estimate	
	Poor: the absolute values of C-correlation coefficients[§] of this parameter with some other parameters are close to one	Good: all of the C-correlation coefficients for this parameter have absolute values less than 0.95
Not important: the absolute values of P-correlation coefficients[♪] of this parameter with the same parameters as here remain close to one.	Abstraction of parameter determination may be useful	No indications to use model abstraction
Important: the absolute values of P-correlation coefficients of this parameter with the same parameters as here are low	Abstraction of the model structure may be useful	No indications to use model abstraction

[¶]The C-scaled sensitivity (composite scaled sensitivity in Hill, 1998) shows the effect of the small increment in the parameter value on the simulation of observed values used in calibration. The large composite scaled sensitivity means that that the simulation of observed values is very sensitive to the parameter value.

[⌂]The P-scaled sensitivity (predicted scaled sensitivity) shows the effect of the small increment in the parameter value on the simulated of key output. The large predicted scaled sensitivity means that the key output is very sensitive to the parameter value.

[§]The C-correlation coefficient of two parameters is computed as the parameter correlation coefficient (Hill, 1998) when only simulated observations are used. Large C-correlation coefficient of two parameters means that they cannot be estimated independently from available observations, i.e. cannot be uniquely defined.

[♪]The P-correlation coefficient of two parameters is computed as the parameter correlation coefficient (Hill, 1998) when both simulated observations and the key output are used.

Large P-correlation coefficient means that the correlation in estimated values of parameters, albeit high, does not affect the key output. Small P-correlation coefficient means non-uniqueness of parameter determination affects the key output.
Equations to compute the C-scaled sensitivity, the P-scaled sensitivity, the C-correlation coefficients, the P-correlation coefficients, the coefficients of variation of parameters, and the confidence intervals of parameters are provided by M. Hill (1998).

calibrated with the data on soil water regime observed during the period without substantial rainfall events. No reliable parameters for the preferential flow submodel can be determined in this case. Similarly, an attempt to find a reliable value of the molecular diffusion coefficient from a breakthrough experiment fails if the convective transport and the hydrodynamic dispersion dominate the transport. For the above example with the preferential flow, soil water regime and concentrations of a conservative tracer can be monitored during large rainfall events, then the data for parameterization the preferential flow submodel become available. The decision on abstracting the base model depends on whether these data will be sufficient or not to determine the reliable parameter values.

A detailed set of guidelines for effective model calibration has been developed for groundwater modeling (Hill, 2004). But in flow and transport modeling in variably saturated soils and sediments, the well-tested guidelines for calibration have not been developed so far. However, the general concepts of the groundwater modeling guidelines appear to be applicable to the models with nonlinearities encountered in unsaturated zone modeling. These concepts are:
using the prior information; this means either using the prior probability distributions of the sought parameter values, or using the measurements of auxiliary values that are not directly simulated but can be computed from the simulation results. In the former case, the maximum likelihood method renders for the minimization the composite sum of the squared model errors and squared differences between prior and posterior parameter estimates (Neuman and Wierenga, 2003). In the latter case, the composite sum to minimize includes the squared model errors and squared differences between measured and computed from simulations auxiliary values (Hill, 1998).
using the measurement errors; this often means assigning weights to the model errors that are inversely proportional to the error in the data;

Since the calibration problem is mathematically ill-posed, there exists a probability of finding a set of parameters that do not provide the global minimization of the minimized sum of squared differences. This probability generally increases when the data accuracy is lower. To mitigate this source of error, it is advised to use a preliminary scanning of the plausible parameter value domain as it is done in the SUFI algorithm (Abbaspour et al., 1997; Schmied et al., 2000). Such scanning allows one to define sub-domains in the parameter space where the correct set of parameters is located with the highest probability.

The prior information about flow and transport parameters has to be obtained either from the measurements at similar sites or from the pedotransfer functions. It is strongly advised to use an ensemble of pedotransfer functions developed with different, sufficiently large databases (Appendix A).

2.3.2.1.2 Need in abstraction because of uncertainty in key model outputs

The system behavior has to be predicted under realistic forcing different from the forcing used in the model calibration. Monte Carlo simulations have to be set with correlations between parameters utilized in scenario generation. If the Gaussian distribution can be

used for original or transformed parameters, the method suggested by Clifton and Neuman (1982) should be used to produce random realizations. If the correlations between inputs are known, the uncertainty is substantially reduced (Meyer et al., 1997). Pohlman et al. (2002) modeled density-driven flow and radionuclide transport at the underground nuclear test site and showed that the incorporation of correlation between hydraulic conductivity and recharge significantly reduced the uncertainty in both travel time and transverse plume location. The climate data generator CLIMGEN (http://www.bsyse.wsu.edu/climgen) can be used to generate the weather scenarios.

The uncertainty of the key output has to be quantified using probability distribution function to see the expected range of predictions given the uncertainty of inputs.
If this range is unacceptable for the purposes of performance assessment, i.e. the median predicted value cannot be statistically significantly discerned from the critical value of the key output, then the model abstraction can be considered. Parameters and inputs of the abstracted model may be estimated with lower uncertainty, and the output may become more reliable.

2.3.2.2 Review the context of the modeling problem

The context of the modeling problem has to be reviewed to assure the objectiveness and the comprehensiveness of the model abstraction. It needs to be realized what details and features of the problem are omitted or de-emphasized when the abstraction is performed.

The context of the modeling problem has been defined by Neuman and Wierenga (2003). They list the following issues that constitute the context:

1. What is (are) the question(s) that the base and abstracted models are to address? The answer should consider (a) the potential or existing problem in which modeling is one of the solution instruments, (b) the potential or existing causes of the problem, (c) the issues needing resolution, and (d) the criteria to decide on efficiency of the resolution. The key output has to be provided with the spatial and temporal scale at which it is evaluated. Acceptable accuracy and uncertainty of the model output have to be established from the end users. In some cases there are mandatory regulations on performance measures that articulate the statistics to use in the particular case. If there are no such regulations, then the statistics have to be selected and defined; it should describe simple and clear ideas about the correspondence between the data and simulations, e.g., how the variability of model errors compares with the variability in the data; do the model residuals have trends; is there a systematic relative or absolute error in predictions, etc.

2. What kind of data is available to calibrate the base and the abstracted models and to test them with respect to the key output? The essential condition is to have the database as broad as possible. It has to include the data from public and private sources, cover both quantitative and qualitative (expert) information, and encompass both site-specific and generic information. The list of the base model inputs and outputs provides a convenient template for the necessary part of the

database. It is imperative to have statistics of all available model inputs and measurable model outputs, including (a) the initial distributions of water contents and/or matric potentials and concentrations of solutes of interest, (b) the pore space and surface properties including horizon or layer thickness, porosity, bulk density, adsorption parameters, and some cases redox conditions (c) the forcing that provides the boundary conditions and the source/sink terms, and (d) the model parameters typical for the site. The latter can be inferred from the ensemble of pedotransfer functions provided the basic soil survey and/or borehole log data are available to serve as the pedotransfer inputs. Statistics includes the type of statistical distribution, the median values and variability measures, the information about observed outliers, and the correlations between parameter values.

3. To make sure that the abstracted models are sound, the additional information has to be collected insuring that the abstracted models include descriptions of all essential processes of flow and transport for given site. This information may be of lower quality compared with the necessary part of the database. However, one has to be sure that some small-scale internal heterogeneities will not have a dominant effect on flow and/or transport at the scale of interest. It is easy to become right for a wrong reason without the information on possible effects of processes that are not included in the model. To become wrong for a right reason is much more responsible way in this case.

2.3.2.3 Define applicable MA techniques

MA can lead to simplifications via

- the number of processes being considered explicitly,
- process descriptions,
- coarsening spatial and temporal support
- the number of measurements for the reliable parameter estimation,
- reduced computational burden,
- data pre-processing and post-processing.

The number of processes can be decreased using abstractions with hierarchies of models, and delimiting input space. Using the hierarchy of models entails a smaller number of processes accounted for in the model. Delimiting input space can be done by changing from two-dimensional to one-dimensional representation of flow and transport in variably saturated soils and sediments. Not only this procedure decreases a pre-processing time, demand on the initial distribution data, and a runtime, but it also excludes the need in defining components of transport coefficient tensors that require elaborate measurements to be found by calibrations. If the flow is predominantly vertical, eliminating the second dimension in many cases causes a relatively small change in flow and transport description in variably saturated soils

Process descriptions can be simplified using abstractions with a hierarchy of models, limiting input domain, and scale change. As the complexity of the model porous media in

Fig. 2-2 decreases, both flow and transport can be described by simpler models. Upscaling can generate models for coarser scales. These models can still be applicable for computing the measure of the performance. At the same time, they have a parameter set that is much smaller and simpler. The same is true in case of applying aggregation. However, the coarser scale parameters are 'lumped,' i.e. they represent a joint effect of several processes and of their variation at the subgrid scale. It is usually impossible to relate the values of these parameters to the basic soil and landscape properties; therefore, the prior estimates of their values are not possible.

If the scale is coarsened but the process description is not changed, one should expect the scale-dependence of parameters. There typically exists a mismatch between the observation scale and computational grid scale; therefore an extreme caution has to be exercised when observations are used to render parameter values for simulations. In unsaturated zone modeling, the reliance on laboratory measured hydraulic properties is much higher than in saturated zone projects. Coarsening of the scale typically creates computational grid units that are larger than the laboratory samples. In such case, scaling abstraction techniques have to be applied and the empirical scaling laws have to be established to convert the laboratory measurements to the grid-scale parameters.

Model abstraction does not decrease the number of measurements needed to calibrate the base model. Ultimately, the performance assessment can be conducted with the aid of highly simplified, abstracted model, and yet the base model has to be calibrated and tested to assure that it is able to represent the natural complexity of the hydrological content, and that it can work as a starting point for abstraction. The number and type of measurements affect both parameter estimation and model performance evaluation. The measurements have to provide (a) the reliable values of parameters to which the key model output is sensitive, and (b) the reliable key model output.

The performance of models can be evaluated with a variety of statistics, most of which normalize the variability of errors by the variability in data (Neuman and Wierenga 2003). High variability of the data may cause the statistic value to be relatively low not because the model is good but because the data are poor.

It has been often assumed that spatially and temporally averaged values are less prone to the predictive uncertainty. This is generally true for statistically homogeneous and stationary systems. Spatial-temporal fields of water contents, concentrations, and soil water potentials usually do not conform these statistical requirements. Nevertheless, the change of scale may actually reduce the variability of the data because of the temporal persistence in differences between measurements in different locations (Grayson and Western, 1998). If the persistence is documented, the coarse-scale average water contents, for example, can be estimated with measurements of the lower density.

Reducing computational burden can be achieved using any of the abstractions. Large runtime can be caused by either the numerical stability requirements or by the temporal detail in the data used to provide input or to evaluate the performance measures. Strong nonlinearity of the models for variably saturated water flow imposes substantial

limitations on the integration time step, especially for the coarse-textured soils and sediments. It is always worthwhile to see whether reformulation of boundary conditions on the soil surface (similar to the one done by van Dam and Feddes (2000) or reformulations of water transport equations **(**similar to the one done by Pan and Wierenga (1995)) may improve the computation speed. Similarly, it is advisable to use look-up tables to avoid multiple computations of nonlinear constitutive relationships. Coding the independent submodels that can be executed in parallel, decreases the number of grid variables that participate in each iteration process present in the model. Wu et al. (2002) used parallel computing in simulations of unsaturated flow and transport for the Yukka mountain project; Englert et al. (2004) demonstrated using parallel computing to solve the saturated-unsaturated flow and transport problem for a large area of pesticide application. However, the limitations of numerical methods and the need to run multiple simulations may still call for model abstraction. Scale change via upscaling or aggregation typically results in much smaller number of grid units and usually in much faster simulations. The metamodeling requires fairly large number of preliminary model runs to generate the development and testing datasets, and may be efficient if the number for future runs of a neural network metamodel is substantially larger than the number of runs needed to develop this metamodel.

Data pre- and post-processing may consume a substantial part of the resources available for the modeling project. Most of model abstraction techniques require changes in the input data used for modeling. The input data are expected to be less voluminous for simpler models. However, the data may also be less certain, and a larger number of simulation runs may be needed to account for this uncertainty. In case of abstractions done by scale change or changing dimensionality, a computational unit, i.e. a grid cell, needs to be characterized by the porous media properties and by forcing at the boundaries. Because of the non-linear effects of subgrid heterogeneities on flow and transport, such averaging is never certain, and such uncertainty has to be simulated on top levels of the uncertainty of initial water contents and solute concentrations, pore media properties, and forcing variables at the boundaries at the fine scale.

In case of extremely large numbers of computational units, even the storage of simulated water content, matric potentials, and concentration fields may become an issue. Using wavelets for data compression may be needed to overcome this difficulty (Mallat, 1998, Neupauer and Powell, 2005).

2.3.2.4 Define and execute model abstractions

Model abstraction, in general, results in smaller number of independent parameters to measure/estimate and/or the lesser amount of computations. The particular procedure of MA depends on the purpose of abstraction and on the available resources. Each abstraction is a separate sub-project that requires justification, planning, and milestone-metered execution. The feasibility of a particular model abstraction can be judged on the basis of resources spent cumulatively on five stages of the abstracted model application, namely (1) pre-processing, (2) calibration, (3) simulation, (4) post-processing, and (5) reporting. For the specific model abstraction procedure, it is necessary to make sure that

the software for using the abstracted model is available, that the output of the abstracted model will be compatible with the available data for calibration and testing purposes, that the pre- and post processing tools are in place. The necessary steps are also to estimate the time and computational resources gain from the abstraction, as well as the personnel available to perform and explain the abstraction. The abstracted models need to be calibrated as described in 2.3.2.1.1. Plausible scenarios described in 2.3.2.1.2 allows one to confirm that the abstraction has alleviated or eliminated the problem encountered with the base model, that it did not create any similar or new problems.

A model that is the result of abstraction can be abstracted further if necessary. Such sequential abstraction has to be done in a systematic manner to assure the understanding of the role of each abstraction step in the performance of the final model.

2.4 Concluding remarks

Very complex models were developed as the results of advancements in understanding and quantifying environmental and social phenomena, as well as industrial processes. The complexity of models is caused, in particular, by the large number of simulated processes, large number of interacting entities, differences in scales at which the interactions occur, large number of feedbacks to simulate. Model abstraction, understood as model simplification, is the field of active research in a multitude of scientific and engineering fields. The more actively models are used in a field the more research on model abstraction can be found. Each research or engineering field has specific types of models in use, and the model simplification techniques are field-dependent. The intensive use of models in subsurface contaminant hydrology has resulted in development of many model abstraction techniques. They have been briefly summarized in previous sections.

Model abstraction in contaminant hydrology has been used mostly in research. Potential benefits of model abstraction include improvement of understanding and communication of modeling results, more robust predictions, and better understanding of essential factors and their representation in models. This makes model abstraction an attractive methodology for engineering modeling applications. The model abstraction process, as described in previous section, is a transparent step-by-step formalized procedure of justification of the use of a simplified model. It is demonstrated in the Chapter 3 of this report.

Engineering applications of models abstraction require model calibration and uncertainty analysis toolboxes and appropriate user interfaces. Such toolboxes and interfaces are currently being developed by several agencies (Castleton et al, 2002; Fine et al., 2002). Availability of the software tools will greatly simplify the model abstraction use.

An important feature of models abstraction is the explicit treatment of model structure uncertainty. The model structure, along with the data uncertainty, and scenario uncertainty, is known to introduce the uncertainty in modeling results. Unlike the uncertainty in input data, in model parameters, and in scenarios, the effect of the model structure uncertainty on the uncertainty in simulation results is usually impossible to

quantify in statistical terms without making very strong assumptions. With model abstraction, a series of models with feasible structures can be built and evaluated in a systematic manner. Each of the models is evaluated from results of an ensemble of simulations by its accuracy to measurement data and by its predictions with respect to scenarios that have not been observed. Model abstraction is always performed in the uncertainty context.

It is observed in diverse modeling projects that several models may have different structures but the same accuracy with respect to data of measurements. As it is not possible to decide which model is “better” or “more correct, ” the idea of using an ensemble of models becomes more and more accepted. Model abstraction can generate models of different structures in a systematic way, thus creating a model ensemble.

It is expected that as the model abstraction methodology will develop, knowing about model abstraction and applying it to the regulated sites will be helpful for the NRC licensing staff. Reviewers will have a means to determine whether the models submitted to support licensing actions adequately represent the site, and whether the investigations adequately represent important features, events, and processes for the site. Managers will have an advisory tool to decide whether the requests for additional information are targeted tat sensitive site parameters and processes. On the other hand, the NRC licensees will have a device to determine whether a simple model that is easy to understand and communicate to regulators, stakeholders, and the general public can adequately represent their site.

3 MODEL ABSTRACTION CASE STUDY

3.1 Introduction

Selection of abstraction techniques and their combinations must take into account the uniqueness of the environmental system being modeled. This relates to the concept of "uniqueness of place in environmental modeling" (Beven, 2002; Neuman and Wierenga, 2003). After reviewing numerous available experimental data sets to develop a model abstraction test case study, we selected a humid site that is relevant to NRC decommissioning site reviews.

The first and main objective of the test case was to understand how model abstraction can affect performance assessment of contaminant migration at a relatively humid site where transport may be affected by the presence of soil macropores and related preferential flow phenomena. Specifically, the test case is designed to assess the capabilities, limitations, and usefulness of selected model abstraction methods.

3.2 Description of experiments

3.2.1 Field setup

The experimental field is located at Bekkevoort, Belgium. It is situated at the bottom of a gentle slope and is covered with a meadow. The soil is classified as Eutric Regosol (FAO, 1975). Typically, the top 1 meter shows three soil horizons: an Ap horizon between 0 and 25 cm, a Cl-horizon between 25 and 55 cm and a C2 horizon between 55 and 100 cm (Fig. 3-1).

A trench, 1.2 m deep and 8 m long, was dug at the field site. The site-specific soil description was performed along one side of the trench. The Ap horizon thickness varied along the trench between 8 and 15 cm depth. From a horizontal distance of 1 m from the side in the trench, a 30 cm thick, sandier layer was observed beneath the Ap horizon. The underlying Cl and C2 horizons reached depths between 75 -90 cm. Beneath the C2, a Bt horizon was found. Between 1.5 and 2.2 m from the side of the trench, a 10 cm thick sandy inclusion at a depth of 75 cm was observed. Texture analyses were done at several locations in the trench (Table 3-1). A study of water retention along a 30-m trench in the same soil at the adjacent site (Mallants et al., 1996) was performed. Values of *van Genuchten parameters* were found from water retention measurements done at 5-cm long and 5.1-cm diameter cores with sand-box apparatus for capillary pressures of 1, 5, 10, 50, and 100 cm, and with pressure cell for capillary pressures of 200, 630, 2500, and 15,000 cm.

The grass cover was removed from the experimental area. A plastic sheet covered one side of the trench to isolate the disturbed trench zone. The trench has been instrumented as shown in Fig. 3-2. Volumetric water content was measured with TDR. Sixty two-rod TDR probes (25 cm long, 0.5 cm rod diameter, 2.5 cm rod spacing) were installed along the trench of 5.5 meter at 12 locations with the 50-cm horizontal spacing at five depths of 15,

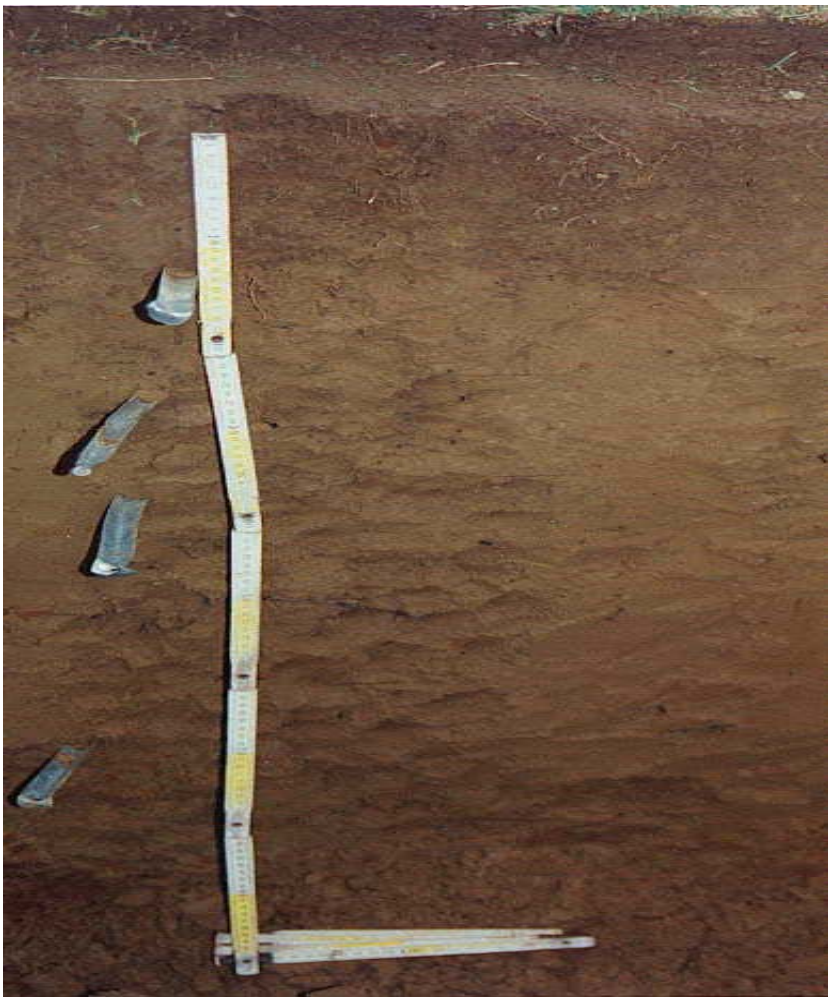

Figure 3-1. Soil profile at the Bekkevoort experimental site (after Jacques, 2000). With permission.

Table 3-1. Average soil texture (%) along the trench at the Bekkevoort site.

Depth, cm	No. samples	>50 μm	50-20 μm	20-10 μm	10-2 μm	<2 μm
15	7	58.6	19.3	6.4	4.5	11.1
35	8	56.7	18.9	7.8	3.2	13.3
55	5	57.3	17.6	6.6	3.7	14.8
75	3	49.6	21.2	7.9	4.4	17.4
95	4	43.8	30.03	7.4	4.5	14

Figure 3-2. View of the instrumented trench (after Jacques, 2000). With permission.

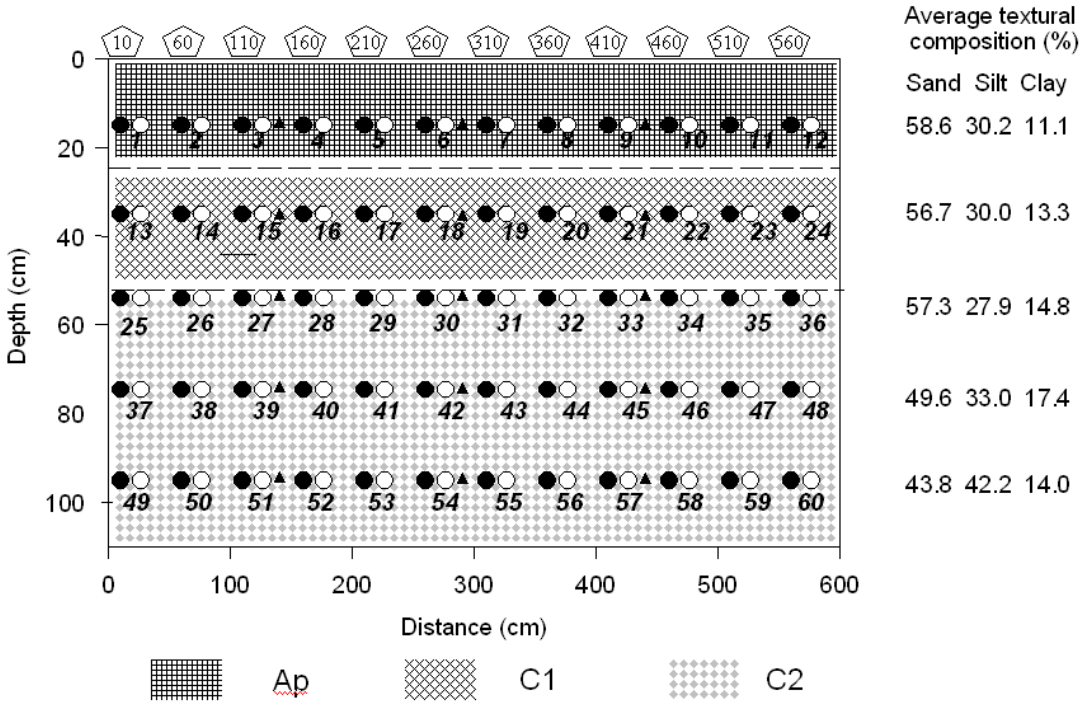

Figure 3-3. Locations and numbering of TDR-probes (●), tensiometers (○) and temperature probes (▲). Patterned rectangles show Ap, C1 and C2 horizons. Dashed lines show the average position of the horizon boundaries. Average values of clay, silt, and sand content are given for the probe installation depths. The numbers in pentagons show the numbering of locations monitored by the five-probe profiling. From (Pachepsky et al., 2005). With permission, SSSA.

35, 55, 75, and 95 cm (Fig. 3-3). Pressure head was measured with tensiometers. Tensiometric porous cups (6 mm diameter, 25 mm long) were installed at a horizontal distance of 10 cm from each of the 60 TDR-probes.
Temperature probes (6 cm long) were installed at 5 depths on distances of 1.3, 2.7 and 4.2 m from the trench.

The TDR-measurements were done with a Tectronix 1502B cable tester. The automated system of Heimovaara and Bouten (1990) was used to control, retrieve, store and analyze the measurements of the travel time of an electromagnetic wave along the TDR - and the soil impedance. Apparent dielectric constants were related to the water content via the site-specific calibration curve (Jacques et al., 1999). One measurement cycle for all TDR-probes took approximately 35 min. The porous cups were connected with water-filled capillary tubes to pressure transducers. These transducers were incorporated in an electronic circuit (SDEC France, Reignac/Indre). Automated measurements of the pressure transducers were controlled and stored by a Campbell CR10X datalogger and AM416 multiplexers (Campbell Scientific Ltd, Leicestershire, UK). The datalogger and multiplexing system also controlled the soil temperature measurements.

Passive capillary samplers (PCAPS) were installed at two 15 and 55 cm depths at the distance of about 5 m from the trench. As discussed by Holder et al. (1991), PCAPS are able to measure water and solute fluxes under both saturated and unsaturated flow conditions by imposing a hanging water column to the soil. After combustion at 400 °C during 8 h (Knutson et al., 1993), the unraveled wicks (Amatex 3/8-inch high density, no.1 0-864KR-02) were mounted in a symmetric way on a 10 mm thick PVC plate. Each wick covered an area of 17 x 17 cm^2. Three wick samplers were put together on one PVC plate. In the middle of each unit (i.e. the 17 x 17 cm^2-area), a hole was drilled from where the wick was connected to a collection chamber through a flexible PVC tube. A nylon mesh was put on the top of the wick filaments to prevent the penetration of soil particles into the PCAPS (Fig. 3-4). For installation in the soil, an excavation of 60 cm width, 10 cm height and 60 cm depth was dug into the trench wall (Fig. 3-5). The top surface of the excavation was carefully levelled. The base plate containing three PCAPS-units was put into the excavation and pushed against the top by hammering wooden blocks under the PCAPS-unit. The flexible tube and collection chambers were then placed at the bottom of the trench in such a way that the distance between the base plate and the bottom of the wick was 90 cm A sampling tube connected the collection chamber with the soil surface. To collect water from the collection chamber, a capillary tube was put into the sampling tube, and suction was applied at the soil surface.

Rainfall was measured and recorded near the trench at a plot with a catch area of 200 cm^2. Cumulative rainfall was written continuously with a floated pen system on a paper (0.1 mm precipitation intervals) fixed on a drum, rotating with a speed of 1 cm per hour.

After all devices were installed, the trench was filled. A thin layer of gravel (1 to 2 cm) was evenly distributed on the study area: (i) to decrease the erosive effect of the rain impact on the bare soil surface, (ii) to minimize the evaporation from the soil surface, and (iii) to decrease the growth of weed on the experimental plot.

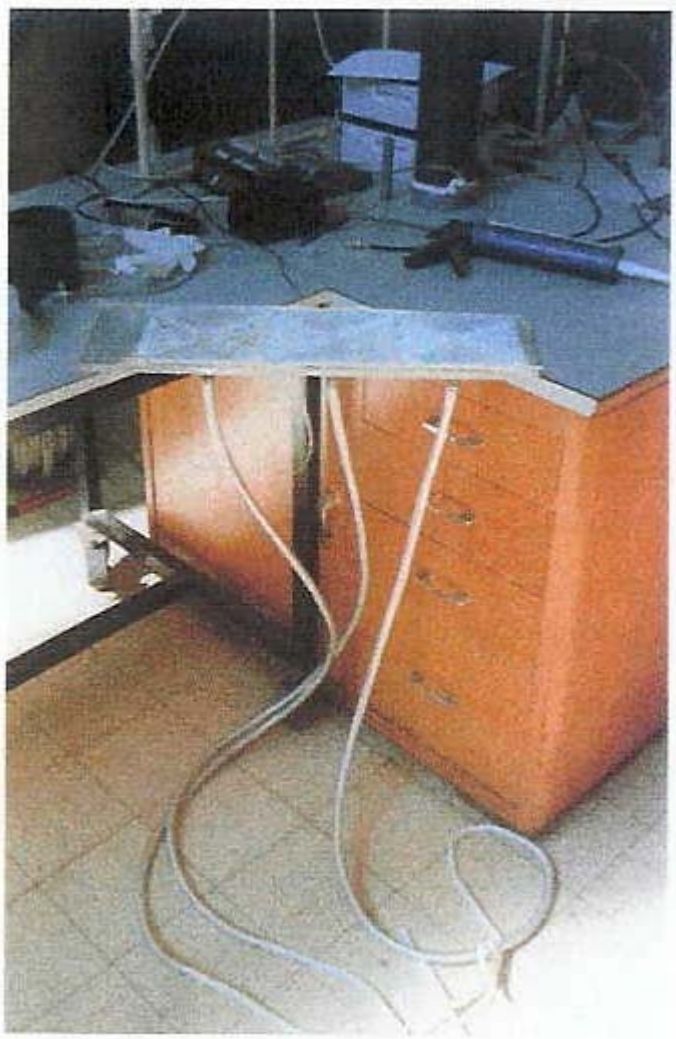

Figure 3-4. Three passive capillary samplers on the PVC plate (after Jacques, 2000). With permission.

Figure 3-5. Installation of passive capillary samplers (after Jacques, 2000). With permission.

.

Field measurements started on March 11, 1998 (day 0) and finished on March 31,1999 (day 384). Weeds were regularly removed from the site during the summer. The tracer ($CaCl_2 \cdot H_2O$)
was manually applied on August 28, 1998 (day 168). On an area of 8 x 2 m^2, a total of 6000 g was applied using a water depth of 0.5 cm resulting in a concentration of 75 $g \cdot l^{-1}$. The experimental area was subdivided in 16 squares of 1 m^2 and solute was applied in three steps (2 times 1.5 liter on each square and then 1 liter on 0.5 m^2). The resident solute concentration was calculated from TDR measurements using calibration curves for each soil horizon.

Frequencies and types of the monitoring measurements are summarized in the Table 3-2. Total of 400,000 measurements were accumulated during the monitoring period.

3.2.2 Dataset overview

3.2.2.1 Precipitation

Fig. 3-6 shows the hourly rainfall rates and the cumulative precipitation during the experiment. A total rainfall amount of 109 cm was recorded during the period of the experiment. The average rainfall in this region is around 80 cm. The amount of rainfall was rather exceptional since it was the largest amount of rainfall during a 65-year period recorded at an official nearby weather station (Ukkel, Brusssels). In the time period of the experiment, the pluviograph in Ukkel had recorded 106 cm.

On days 185 and 186 (13 and 14 September 1998), an extreme rainfall event occurred which resulted in large scale flooding in the region of the experiment. At the experimental site itself, no ponding was observed during the storm. However, flooding occurred at the foot of the hill.
After the storm, there was one more dry period, and an almost linear increase in cumulative precipitation was observed beginning from the day 237.

3.2.2.2 Soil water contents, capillary pressures and soil temperatures

Time series of average at each depth water contents and capillary pressures are shown in Fig. 3-7 and Fig. 3-8. Measured average water content varied in time in a limited range between 0.3 and 0.4 due to the absence of water uptake by plants, limited evaporation and the relatively high continual rainfall during the experiment. As could be expected, the temporal water content variations were more pronounced in the topsoil than at larder depths.

Averaged along the trench capillary pressures varied in time in a relatively small range (Fig. 3-8). The largest temporal variation was observed during the first 200 days of the experiment where periods of heavy rainfall alternated with long dry periods. During the last 180 days of the experiment, the pressure heads showed a fast oscillatory behavior between e.g. 0 and -50 cm at the 15 cm depth and 0 and -20 cm at the 55 cm depth due to the typical rain pattern during the last 100 days.

Table 3-2. Types and frequencies of the monitoring measurements.

Quantity	Device	Frequency	Positions
Water content	TDR-probes	Every 2 hr	5 d x 12p
Resident solute concentration	TDR-probes	Every 2 hr	5 d x 12 p
Pressure head	Tensiometers	Every hr	5 d x 12 p
Temperature	Temperature probes	Every hr	5 d x 3 p
Water fluxes	Passive Cap. Lys	Every 2-3 days	2 d x 3 p
Solute fluxes	Passive Cap. Lys	Every 2-3 days	2 d x 3 p
Rainfall	Pluviograph	Continuously	1

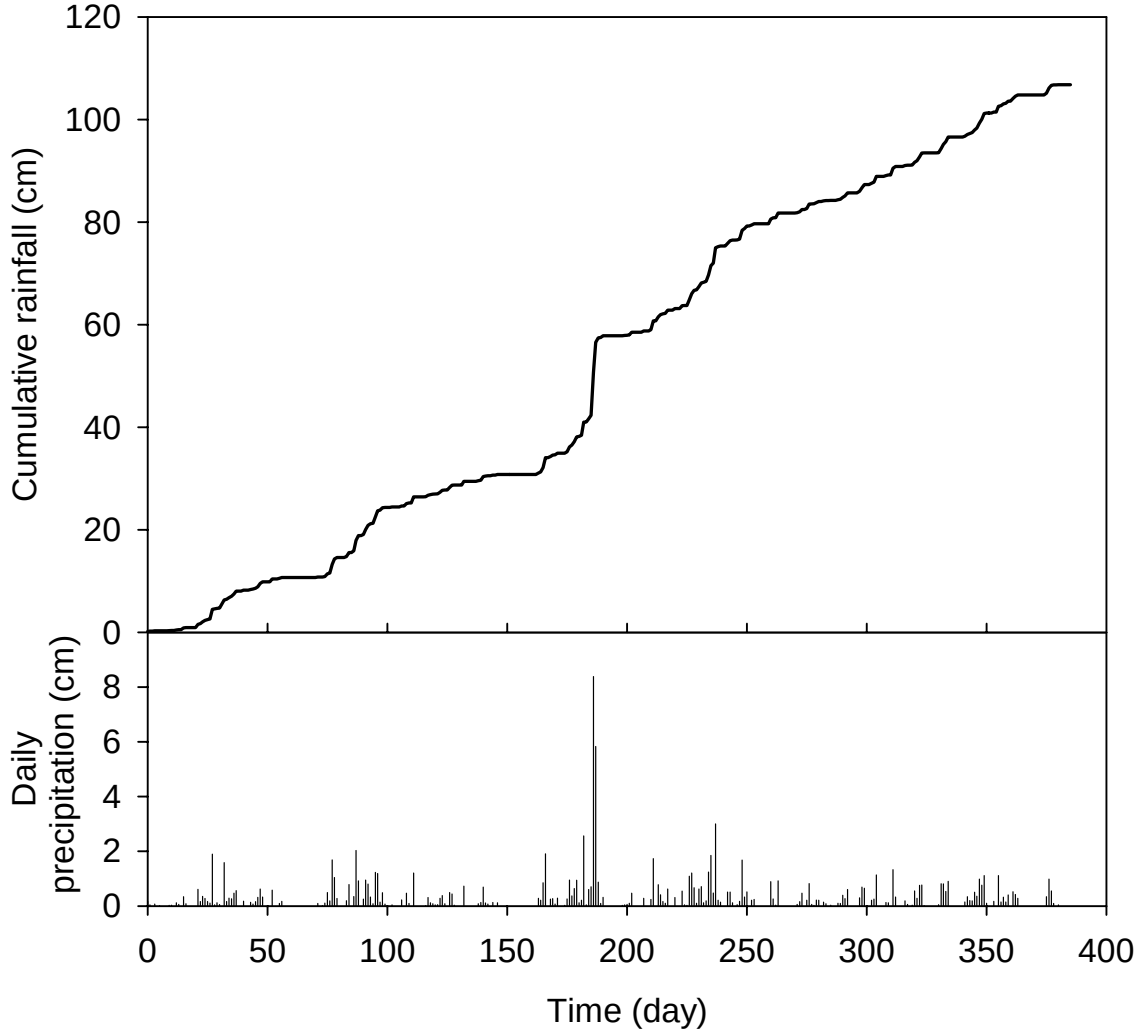

Figure 3-6. Cumulative rainfall and daily precipitation during the experiment.

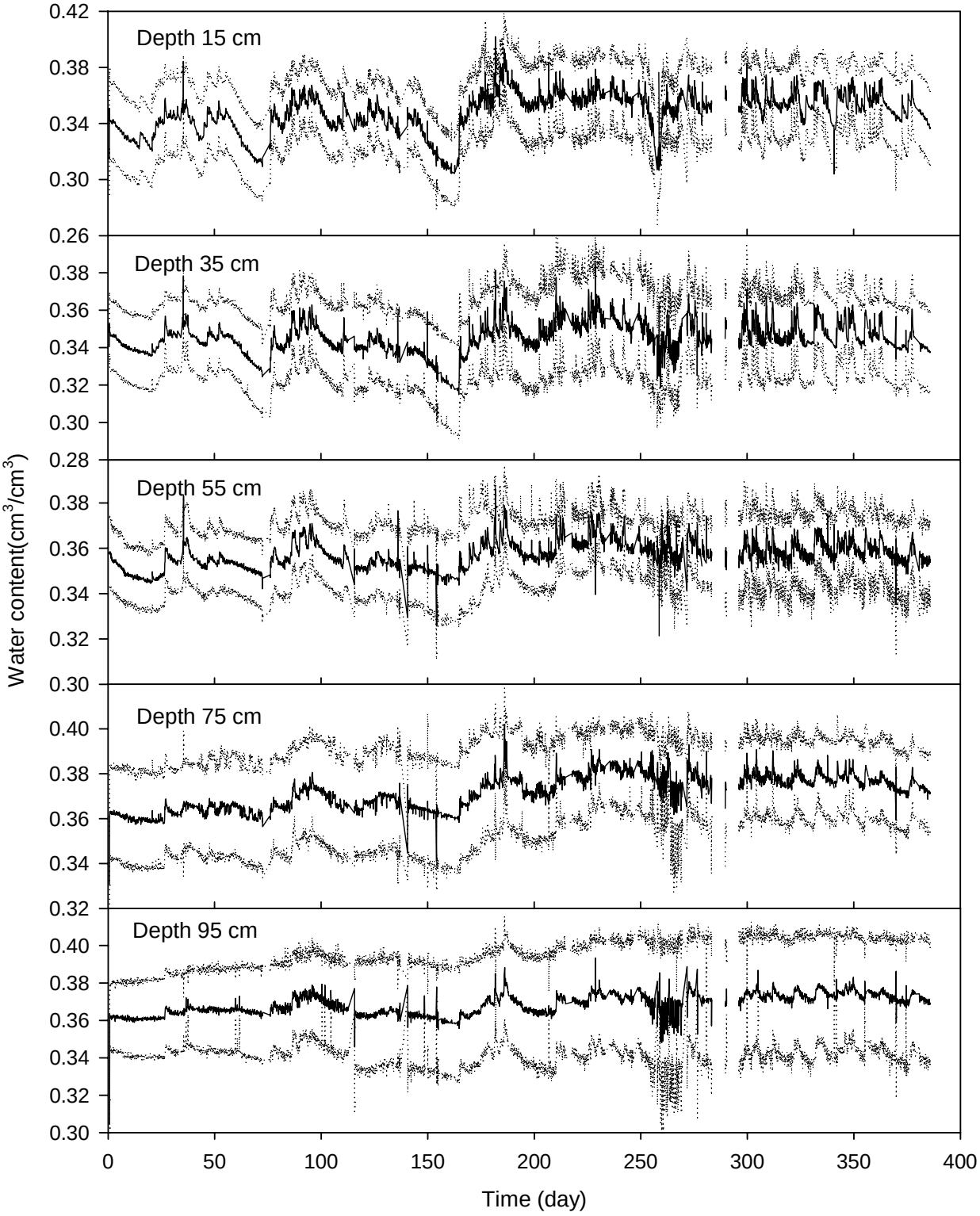

Figure 3-7. Time series of average water contents at each of the measurement depths. Solid lines correspond to the measured values at 50%, dotted lines at 25% and 75% probability levels.

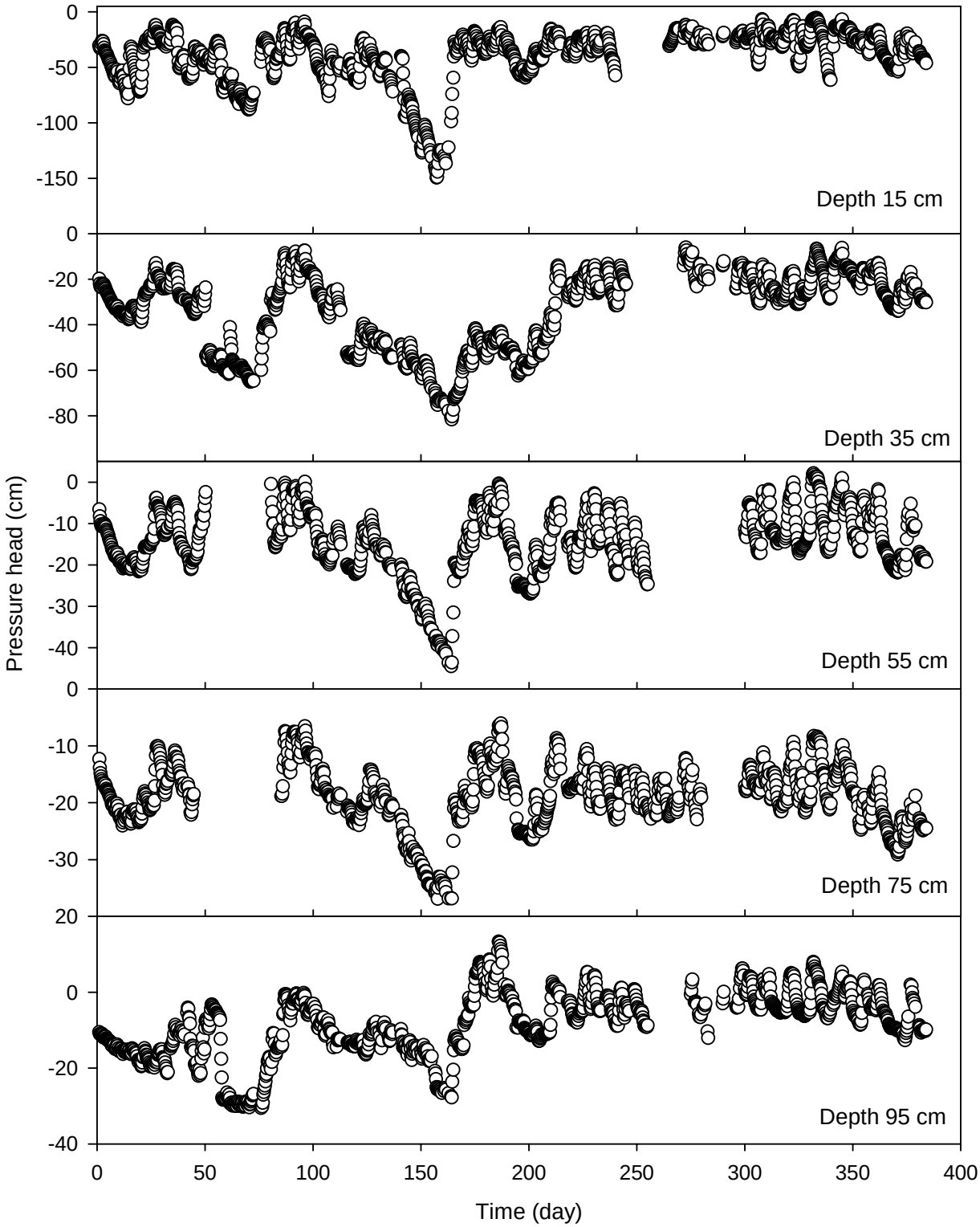

Figure 3-8. Time series of average capillary pressures at each of the measurement depths.

The lateral and temporal variability of soil water contents are characterized in Tables 3-3 and 3-4. Coefficient of variation (CV) of water content in time was in the range from 0.6% to 8.1% (Table 3-4). Mean value of CV decreased with depth from 4.7% to 2.0%. Spatial variations in water content along the trench were more pronounced as compared with temporal variations. The CV values were in the range from 0.4% to 15.5%, which was wider than the range of temporal variability. Mean value of CV for space water content variability did not have depth-related trend. Rainfall was distributed evenly at all locations, and variability in soil water content has been likely a result of soil heterogeneity and hydraulic properties variability.

Two-dimensional spatial distributions of water content for several observation days are shown in Fig. 3-9. There is a definite structure in the data, so that some parts of the profile remain consistently wet, and other parts stay relatively dry.

Time series of averaged along the trench soil temperature are shown in Fig. 3-10. Values of temperature had a typical annual sine wave trend and were in the range between 0.4 and 25.6C° on 25cm depth and in the range between 4.0 and 18.3C° on 95cm depth. The average annual temperature decreased slightly from 11.0 to 10.8C° with depth. The spatial variability of temperature was very small at all depths. The coefficient of variations of temperature had no depth-related trend and was in the range from 1% and 0 to 6%.

3.2.2.3 Soil water fluxes

Cumulative soil water fluxes measured with passive capillary lysimeters are shown in Fig. 3-11. The lysimeters responded to the large precipitation events. The variability in values of the fluxes was substantial and grew as the experiment progressed.

3.2.2.4 Laboratory data on soil water retention

Parameters of the van Genuchten equation (2-2) for the samples taken along 30-m trench are shown in Fig. 3-12. Statistics of van Genuchten parameters are in the Table 3-5. The variability of the van Genuchten parameters α and n was relatively high in all three horizons. There was a weak correlation between saturated water contents θ_s and lg α. In all three horizons, and a moderate correlations between lg α and lg n. The same van Genuchten parameters determined in different horizons did not correlate.

3.3 The base model and the need in model abstraction

The base model was the single continuum model of variably saturated porous media simulated with the Richards equation. The software to solve the Richards equation and to estimate its parameters was the HYDRUS-1D (Simunek et al., 1998). HYDRUS-1D is one

Table 3-3. Coefficients of variation (%) of soil water contents in time.

Coefficients of variation	Depth				
	15 cm	35 cm	55 cm	75 cm	95 cm
Min over all probes at a given depth	2.7	1.4	0.7	0.6	1.2
Max over all probes at a given depth	8.1	5.5	3.1	4.6	2.8
Mean over all probes at a given depth	4.7	3	2	2.2	2

Table 3-4. Coefficients of variation (%) of soil water content along the trench

Coefficients of variation	Depth				
	15 cm	35 cm	55 cm	75 cm	95 cm
Min over all observation times	3.6	3.6	0.4	1.1	2.2
Max over all observation times	11.6	11.2	7.4	9.5	15.5
Mean over all observation times	7.2	6.2	4.3	5.4	7.5

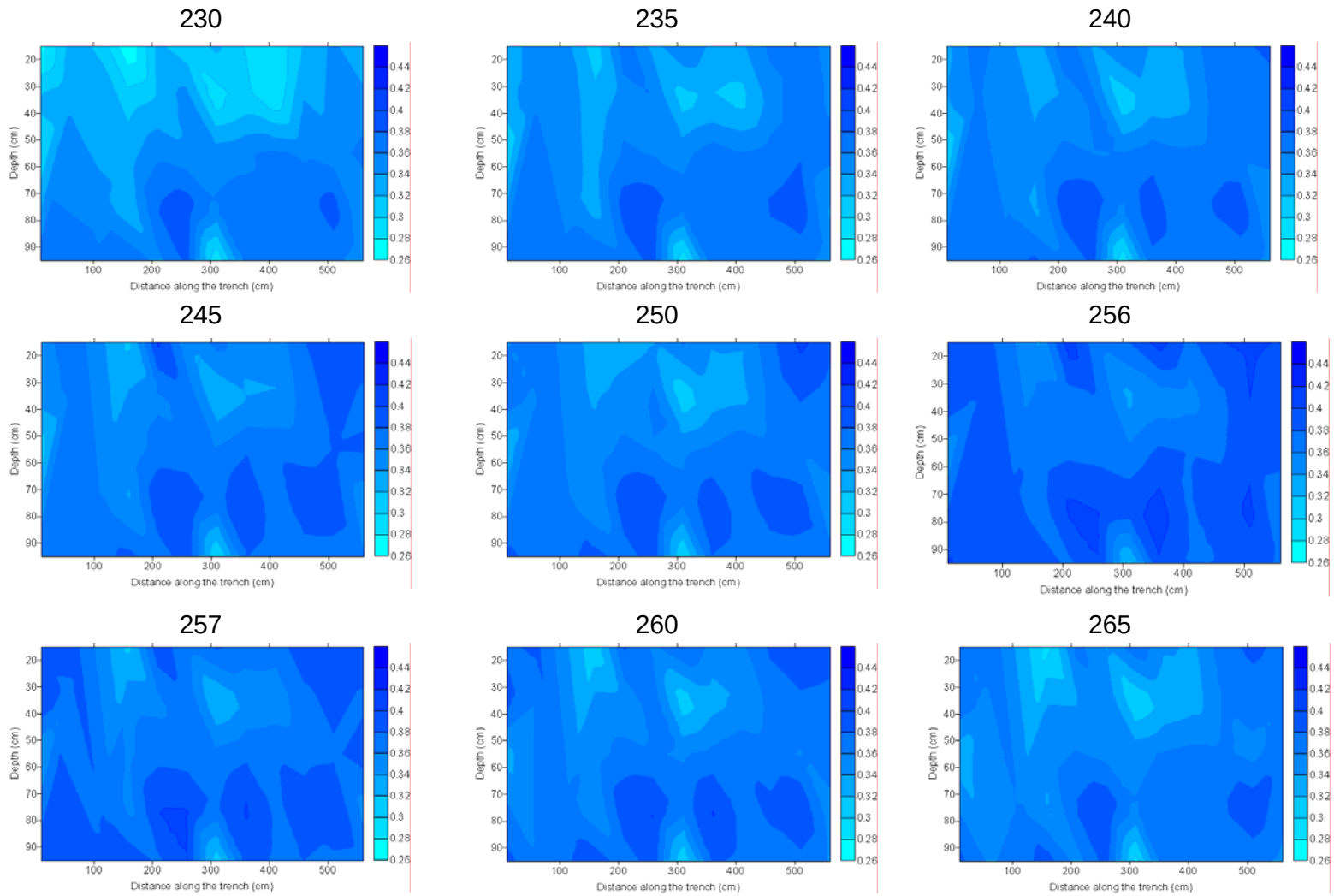

Figure 3-9. Consecutive images of two-dimensional distributions of soil moisture contents; days from the beginning of the experiment are shown.

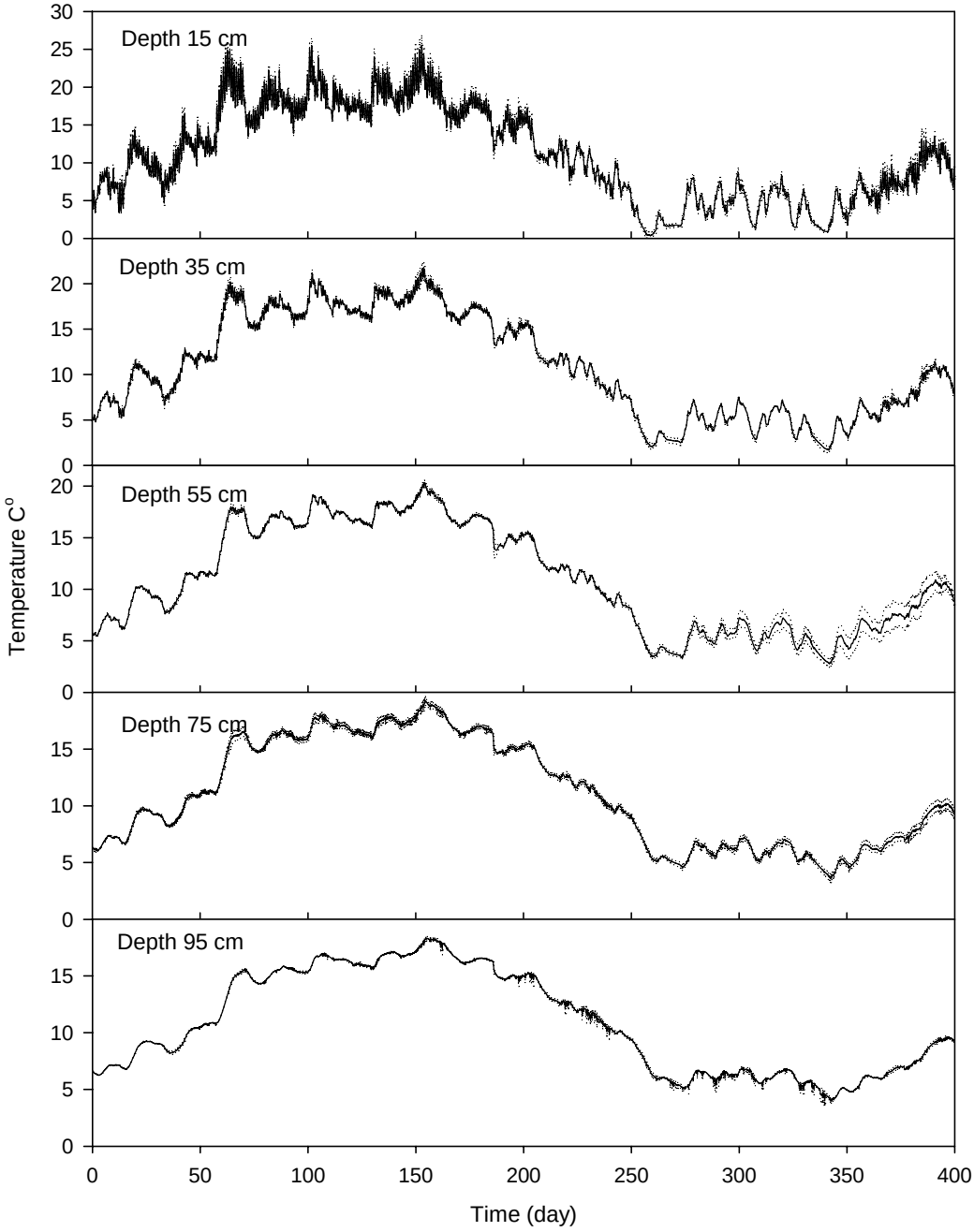

Figure 3-10. Time series of soil temperature. Solid lines correspond to the average measured values, dotted lines correspond to minimum and maximum values at each depth.

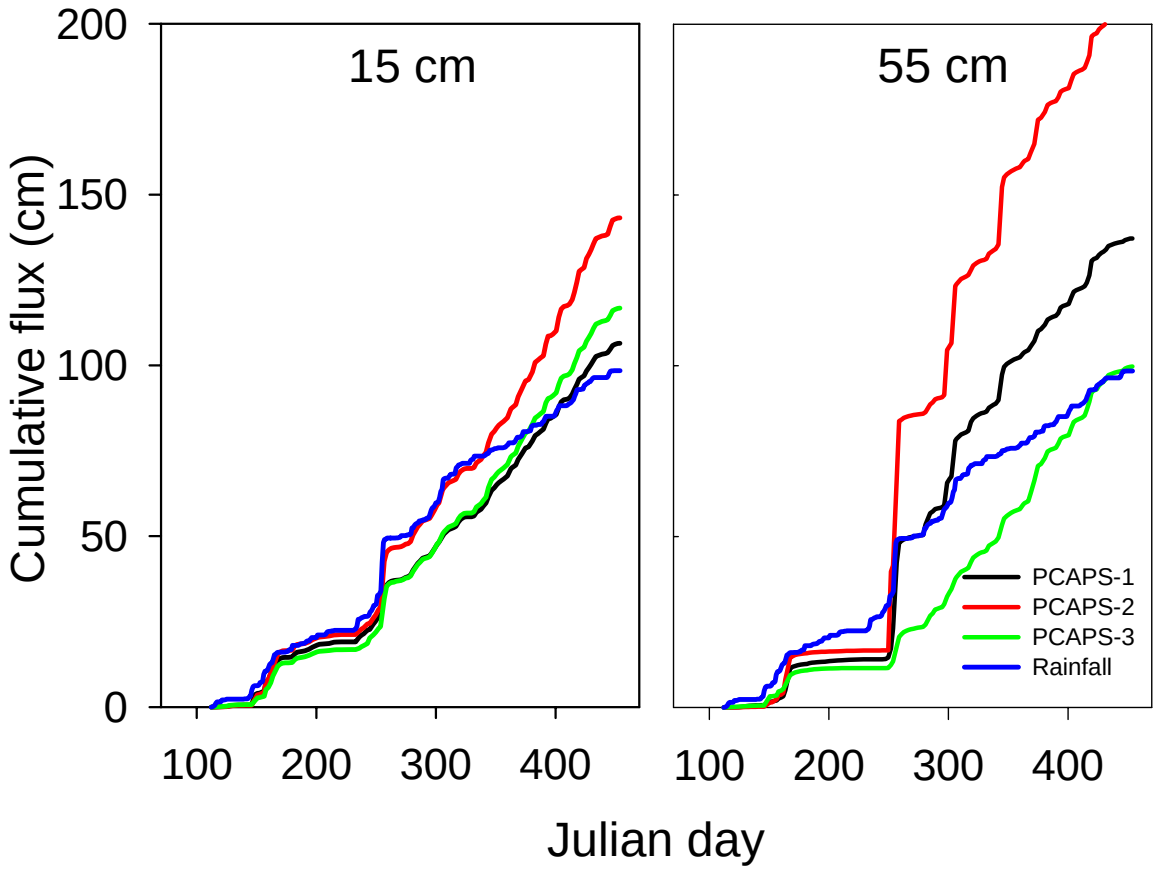

Figure 3-11. Cumulative rainfall and soil water fluxes measured with passive capillary lysimeters (PCAPS) at two depths.

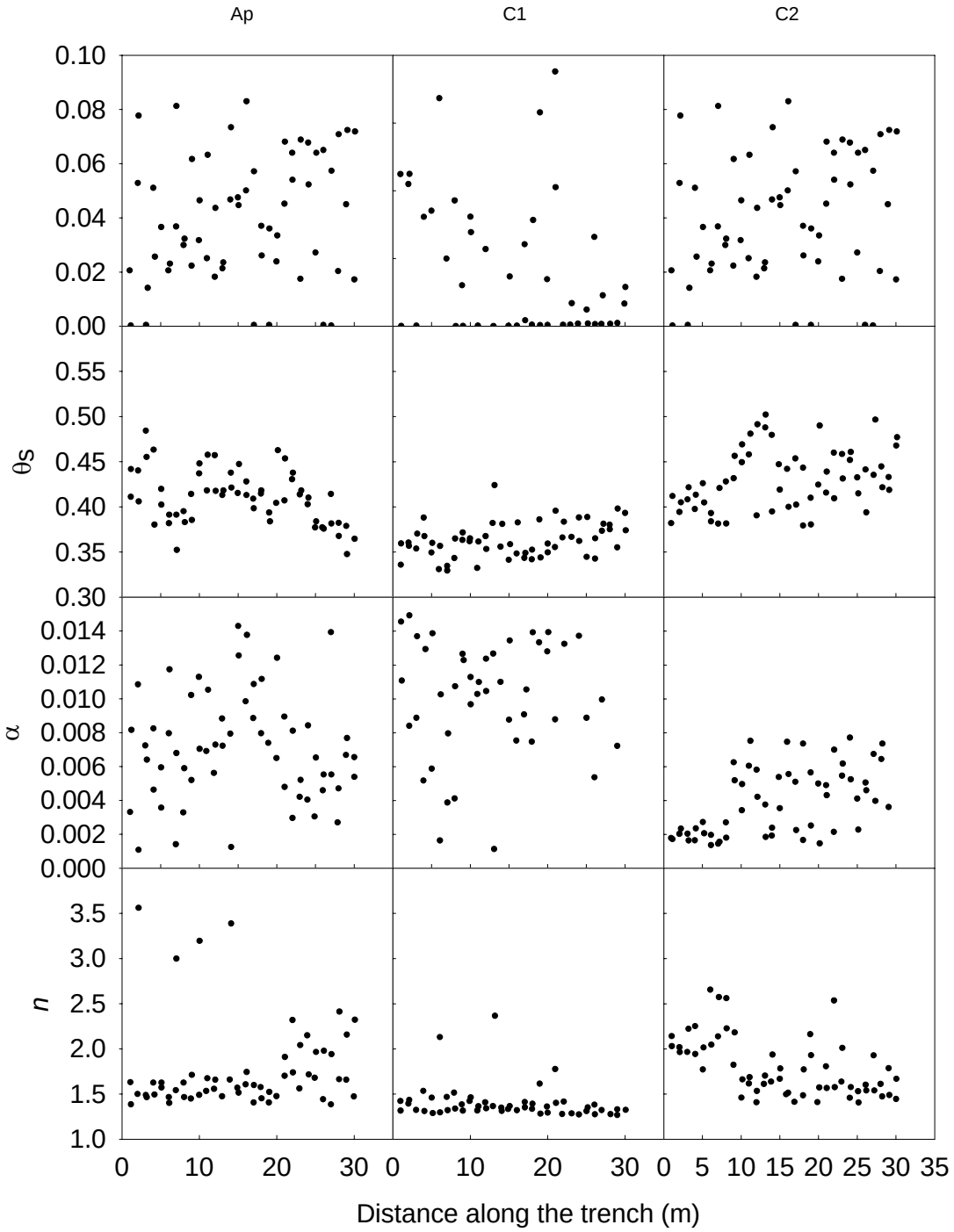

Figure 3-12. Van Genuchten parameters derived from water retention of samples taken from three soil horizons along the 30-m trench.

Table 3-5. Statistics of van Genuchten parameters of water retention in samples along the 30-m trench

	$\theta_{r,Ap}$	$\theta_{r,C1}$	$\theta_{r,C2}$	$\theta_{s,Ap}$	$\theta_{s,C1}$	$\theta_{s,C2}$	lg α_{Ap}	lg α_{C1}	lg α_{C2}	lg n_{Ap}	lg n_{C1}	lg n_{C2}
Min	0.000	0.000	0.000	0.348	0.330	0.379	-2.966	-2.947	-2.865	0.141	0.103	0.148
Average	0.040	0.015	0.044	0.411	0.363	0.431	-2.201	-1.951	-2.466	0.236	0.143	0.249
Max	0.083	0.075	0.101	0.484	0.424	0.502	-1.845	-1.536	-2.113	0.552	0.374	0.424
Coefficient of variation (%)	57.7	135.1	53.8	7.2	5.2	7.7	11.0	13.2	9.7	41.0	34.8	29.9
Correlation coefficients significant at 0.05 significance level												
$\theta_{r,Ap}$	1.000	-	-	-	0.268	-	-0.332	0.349	-	0.678	-	-
$\theta_{r,C1}$		1.000	-	-	-	-0.383	-	-0.338	-	-	0.851	0.254
$\theta_{r,C2}$			1.000	-	-	-	-0.381	-	-0.684	-	-	0.834
$\theta_{s,Ap}$				1.000	-	-	0.430	-	-	-	-	-
$\theta_{s,C1}$					1.000	-	-	0.558	-	-	-	-
$\theta_{s,C2}$						1.000	-	-	0.444	0.376	-0.285	-0.329
lg α_{Ap}							1.000	-	-	-0.453	-	-0.313
lg α_{C1}								1.000	-	0.282	-0.630	-0.333
lg α_{C2}									1.000	-	-	-0.899
lg n_{Ap}										1.000	-	-
lg n_{C1}											1.000	0.340
lg n_{C2}												1.000

of the most widely used codes for unsaturated flow and solute transport modeling that uses the finite element method to solve the Richards equation. It has options for nonisothermal liquid and vapor flow and heat transport. Constitutive relationships include van Genuchten and Brooks and Corey water retention functions. Information on soil texture can be used along with pedotransfer functions to determine water retention and hydraulic conductivity parameters. The code includes inverse modeling capabilities for estimating hydraulic and transport parameters. The upper boundary condition includes standard constant pressure and constant flux conditions in addition to meteorological forcing. Options for the lower boundary condition include unit gradient and seepage face (Scanlon, 2004).

The HYDRUS-1D code was applied to simulate the dataset presented in the sections 3.2.2.1 and 3.2.2.2 (Jacques et al., 2002). These authors calibrated the model using three different measurement sets in the inverse optimization: the water content data averaged by depth for each depth $\theta(t,z)$; the pressure head data averaged by depth for each depth $\psi(t,z)$; the hourly rainfall rates that are smaller than 0.5 cm h^{-1} $q(t,0)$. The upper boundary condition was defined by hourly rainfall rates. The evaporation was assumed to be negligible during the experiment due to the presence of a thin layer of gravel of soil surface. The lower boundary condition was defined as a zero pressure head gradient, i.e. free drainage. The initial condition was determined by linear interpolation of the measured pressure heads between the five depths at day 0.
Two approaches were used for the estimation of the soil hydraulic parameters. First, parameters were estimated using time series of observations for a specific depth assuming a homogeneous soil profile. Parameters were interpreted as effective parameters representative for an equivalent homogeneous porous medium between the soil surface and a particular depth. The objective function included three data sets: $q(t,0)$, $\theta(t,z)$ and $\psi(t,z)$ for the particular depth. Three parameters (θ_s, α and n) of the van Genuchten equation (2-2) and two parameters (K_{sat} and l) of Genuchten-Mualem equation (2-5) were estimated for each of five observation depths. Second, a layered soil profile with different parameters for each layer was assumed. The objective function included eleven data sets: $q(t,0)$ and data sets $\theta(t,z)$ and $\psi(t,z)$ for five depths. Twenty soil hydraulic parameters were estimated within single optimization. Data from passive capillary lysimeters were used neither in calibration nor in the model performance assessment.

The calibrated model reproduced the time series of soil water contents very well. Simulated pressure heads decreased less compared to observations at all depths during the dry period in the summer (day 140-160) and were significantly overpredicted at depth of 35 cm between days 150 and 200 during the wet period (see Jacques et al., 2002 for details). The total simulated cumulative infiltration was only from 50.2 to 83.8% of the total amount of rain (Fig. 3-13). The largest discrepancy was obtained during the heavy rainfall on day 185, and smaller differences during moderate rainfall conditions (e.g. days 100 and 150, and 250 and 350). The model predicted substantial runoff. However, no runoff or ponding were observed in the field.
This inexplicable result indicated the need for the model abstraction.

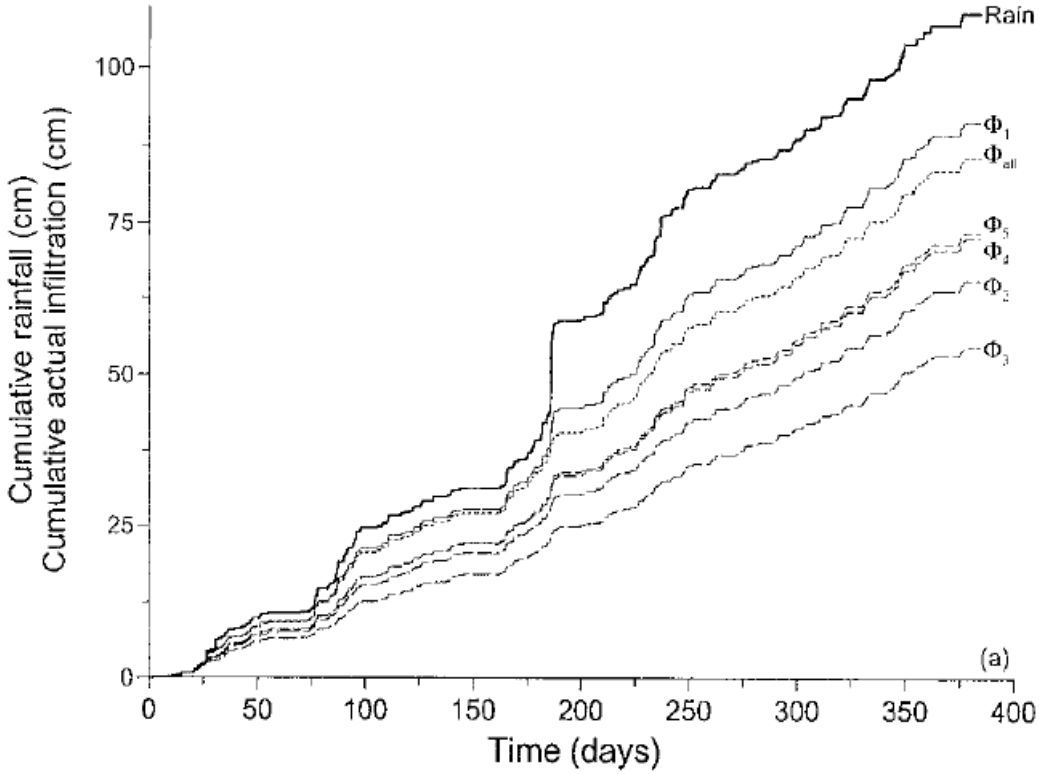

Figure 3-13. Measured cumulative rainfall and simulated cumulative actual infiltration for the different objective functions (Jacques et al., 2002). Φ_1, Φ_2, Φ_3, Φ_4 and Φ_5 denote to the parameter optimization at depths of 15, 35, 55, 75 and 95 of homogeneous soil profile; Φ_{all} denotes to the multi-layered soil profile optimization.

3.4 Review of the model abstraction context

Model abstraction starts with defining purpose and resources. As discussed in the Chapter 2.3.2.2, the main directions of the context review are:

(1) What is (are) the question(s) that the base and abstracted models are to address? What is the key output of the model that quantifies those questions? How the results of abstractions will be evaluated?

(2) What kind of data is available to calibrate the base and the abstracted models and to test them with respect to the key output?

(3) What is the additional information that has to be collected to make sure that the abstracted models include descriptions of essential processes of flow and transport for given site.

3.4.1 The key output

The NRC staff review of performance assessments of nuclear facilities, e.g., decommissioning of the facilities, management of low-level and high-level radioactive waste disposal sites, frequently involves the review of models of water flow and solute transport in soils. Because water flow is the essential driver of the contaminant transport in subsurface, an accurate prediction of water fluxes is the precondition of correct estimating transport of contaminants. Therefore, the soil water fluxes were selected as the key output of the base and abstracted model.

Inspection of the soil water fluxes measured at depths of 15 and 55 cm with PCAPS revealed gaps in measurements and equipment malfunctioning. Soil water fluxes measurements started 40 days after the beginning of water content and pressure head measurements. Time interval between measurements varied from 1 to 8 days. The measured cumulative fluxes after rainfall event on the day 185 were greater then the cumulative rainfall and the slopes of cumulative fluxes curves were steeper then that for rainfall that indicated malfunctioning of PCAPS. For those reasons not all data were used, but three wetting-drying periods were selected as shown in Fig. 3-14. Cumulative soil water fluxes for those periods were used as three values of the key output.

3.4.2 Statistics to evaluate model abstractions

The *F*-test was used to estimate accuracy of flux simulation for each MA technique. The test compares model error with the variation in the experimental data (Whitmore, 1991). The sum of squares due to lack of fit was used to characterize the model error:

$$LOFIT = \sum_{j=1}^{K} n_j (\bar{y}_j - x_j)^2 \qquad (3.1)$$

where n_j is the number of replicates at each depth and time, x_j is the simulated flux for each depth and time, K is the total number of measurement points (K = [number of depth] x [number of time periods]), $\bar{y}_j$ is the mean of the measurement at each depth and time:

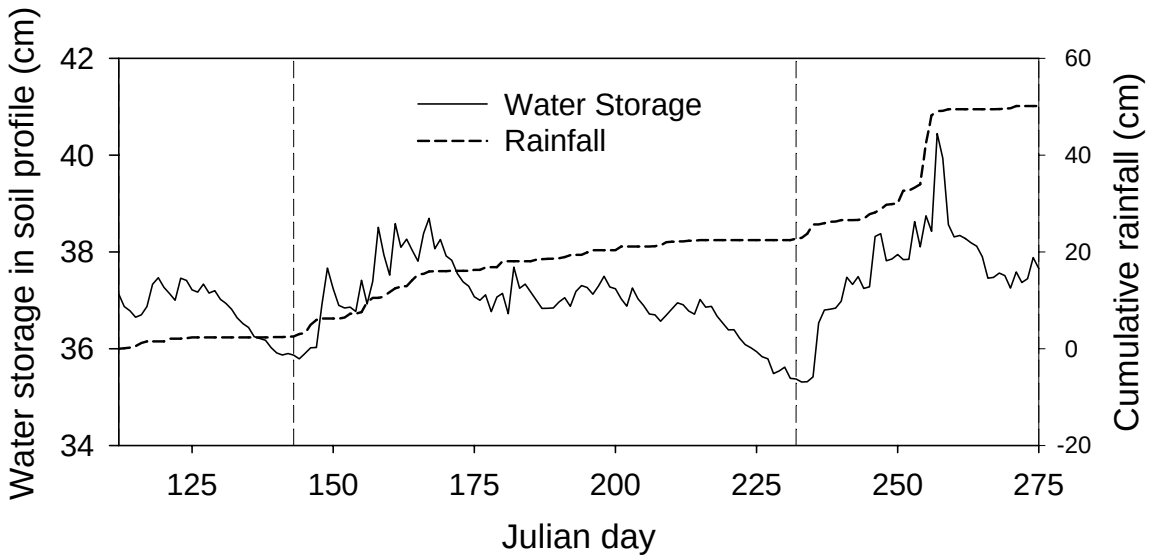

Figure 3-14. Wetting-drying periods selected to compute the cumulative water flux as the key output of the base and abstracted models.

$$\bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij} \qquad (3.2)$$

where y_{ij} is the measured flux for each depth and time.

The sum of squares of the error was used to characterize variability of the measured data:

$$SSE = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2 \qquad (3.3)$$

The F-statistics is computed as:

$$F = \frac{LOFIT}{SSE} \qquad (3.4)$$

The number of degrees of freedom in the *LOFIT* is K, and in the *SSE* is N-K, where N is the total number of measurements including all replications.

Each abstracted model was tested to assess the propagation of the variability in parameters through the model and the resultant variability in the key output and in simulated water contents. The variability in parameters was inferred either form the inverse modeling results or from the ensemble of pedotransfer functions as described below.. Probability distribution functions of the root mean-square error in water contents and the probability distribution functions of the key output were used to document and compare the error propagation.

3.4.3 Available data and their uncertainty

The purpose of the data review is to establish model calibration dataset having reliable statistics of all model inputs and outputs.

3.4.3.1 Averaging water contents

Inspection of time series of water content and pressure head revealed oscillatory behavior that apparently had diurnal periodicity. The data analysis presented in the Appendix B showed that that the oscillations of recorded capillary pressures and soil water contents were related to the temperature on/near soil surface. That could be attributed to the unknown response of the electronic equipment on the surface to the diurnal daily oscillations. To eliminate temperature effect of measurement devices, daily time series of water content and pressure head were averaged for each day of the observation period. The temporal time scale of observations became equal to one day.

3.4.3.2 Temporal persistence in soil water contents

The water content data were incomplete due to malfunctioning of the measurement devices. When removed from soil, all probes were still intact, and were subsequently used in other laboratory or field experiments. The malfunctioning was apparently caused by the connections and switch boxes in the automated measurement system that would stop

working properly resulting in missing data for the rest of the experiment for a given probe. Averaging data from the remaining probes at a given depth could be an option if no persistent difference had existed between TDR-measured water content in different locations. However, the data inspection showed that, when all probes worked, some probes at a given depth consistently showed water contents below average whereas others show water contents above the average. Fig. 3-15 shows measurements made when all probes at the 15-cm depth have worked. The graphs for the individual probes are clearly shifted relative to each other along the water content axis. This demonstrates the temporal persistence in water contents at different locations. Similar graphs were obtained for other depths (data not shown).

Because of the temporal persistence, the malfunctioning of probes with consistently lower than average values would lead to averaging only data from the remaining probes that had water contents consistently higher than the true average. This averaging would lead to overestimation of the average water content at the depth of interest. In contrary, the malfunctioning of probes with consistently higher than average values would lead to the underestimating the average.

The correction for the temporal persistence (Pachepsky et al., 2005) was applied to the data as described in the Appendix C to estimate layer-averaged water contents. This correction also led to the five-fold decrease in the uncertainty of the estimates of layer-averaged water contents because, in case of persistence, the individual probe readings are not independent. The depth-averaged water contents with the standard errors based on the persistence were used in calibrations.

3.4.3.3 Surface evaporation

The calibration of the base model in (Jacques et al., 2002) was based on the assumption that the surface evaporation was absent during the experiment. To evaluate the uncertainty related to this assumption, we used the data from locations where the monitoring was continuous at all observation depths. The cumulative change in water storage in soil profile was computed for those locations. Results in Fig. 3-16 show that distinct periods of water loss ("dry periods") and water accumulation ("wet periods") can be defined for all locations. Inspection of capillary pressure and water content data during dry periods showed that:

(a) capillary pressure gradient at depths 65 cm and 85 cm was between zero and one (Fig. 3-17); positive value of pressure gradient indicated downward water flux from those layers with infiltration,
(b) water content decreased and pressure head increased continuously at all depths (Fig. 3-18)
(c) water content increased linearly with depth in deep parts of soil profile (Fig. 3-19).

We used these data to estimate the matrix unsaturated hydraulic conductivity in the deep part of the profile in the C2 soil horizon. We assumed that (a) no preferential flow occurred

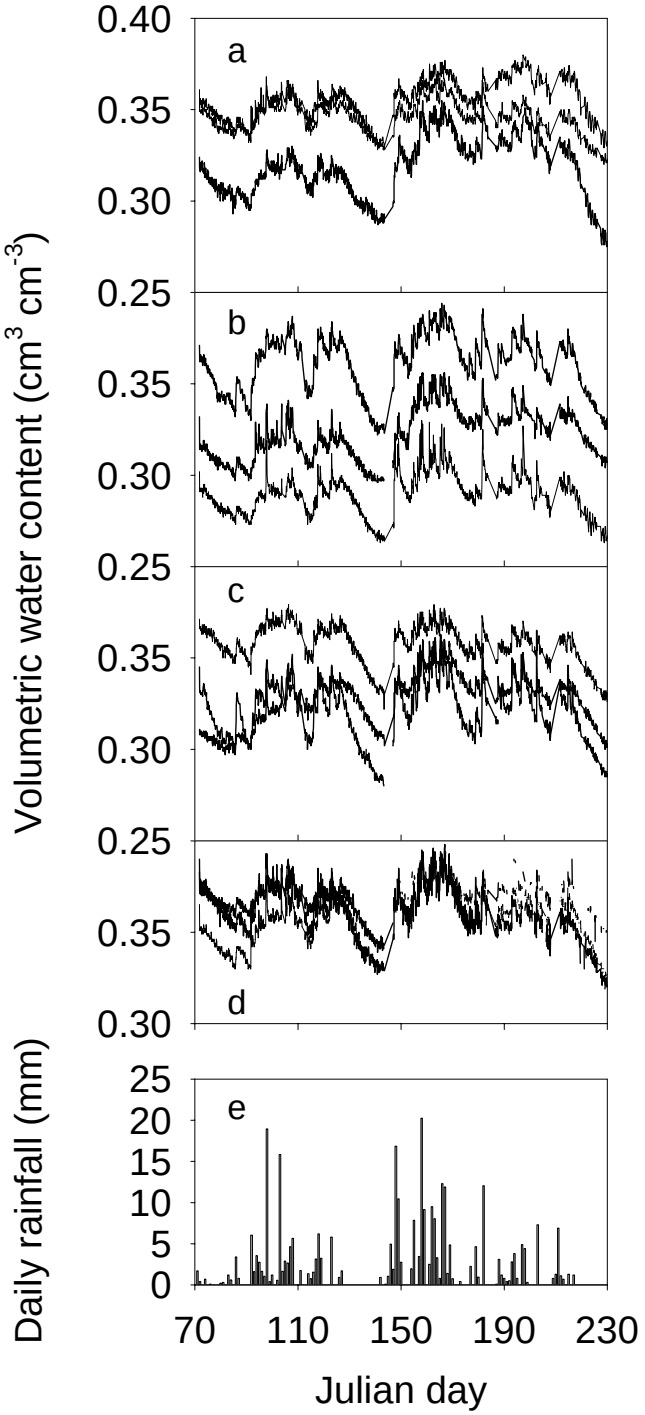

Figure 3-15. Time series of TDR-measured water contents at the depth of 15 cm (a-d) and precipitation (e) during the period when all probes worked at this depth. Probe numbers top to bottom (a) 2, 3, 1, (b) 5, 6, 4, (c) 7, 8, 9, (d) 11, 12, 10; probe numbering is shown in Figure 3-3.

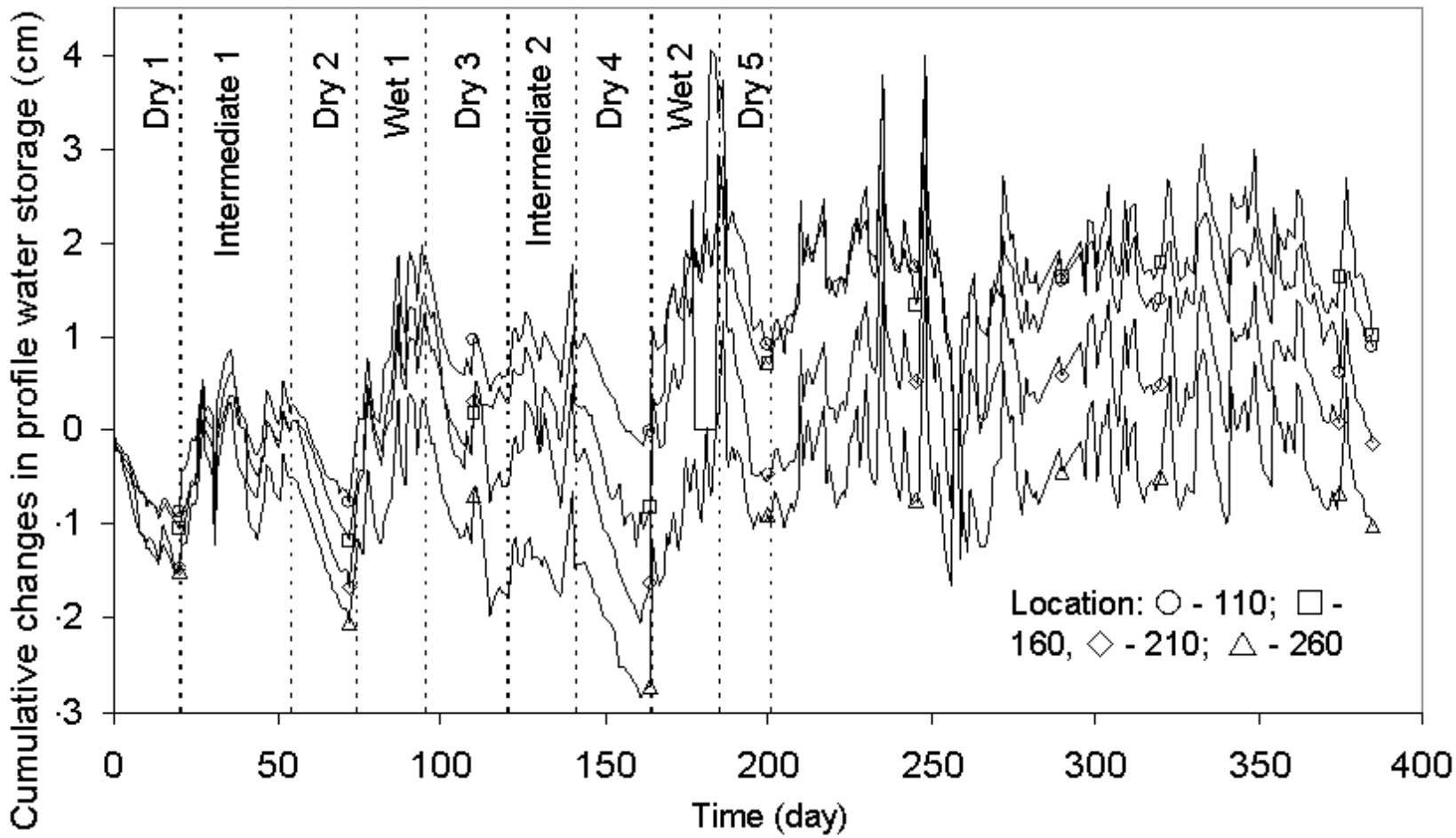

Figure 3-16. Cumulative change in soil water storage in several locations along the trench. Location numbering is given in the Fig. 3-3.

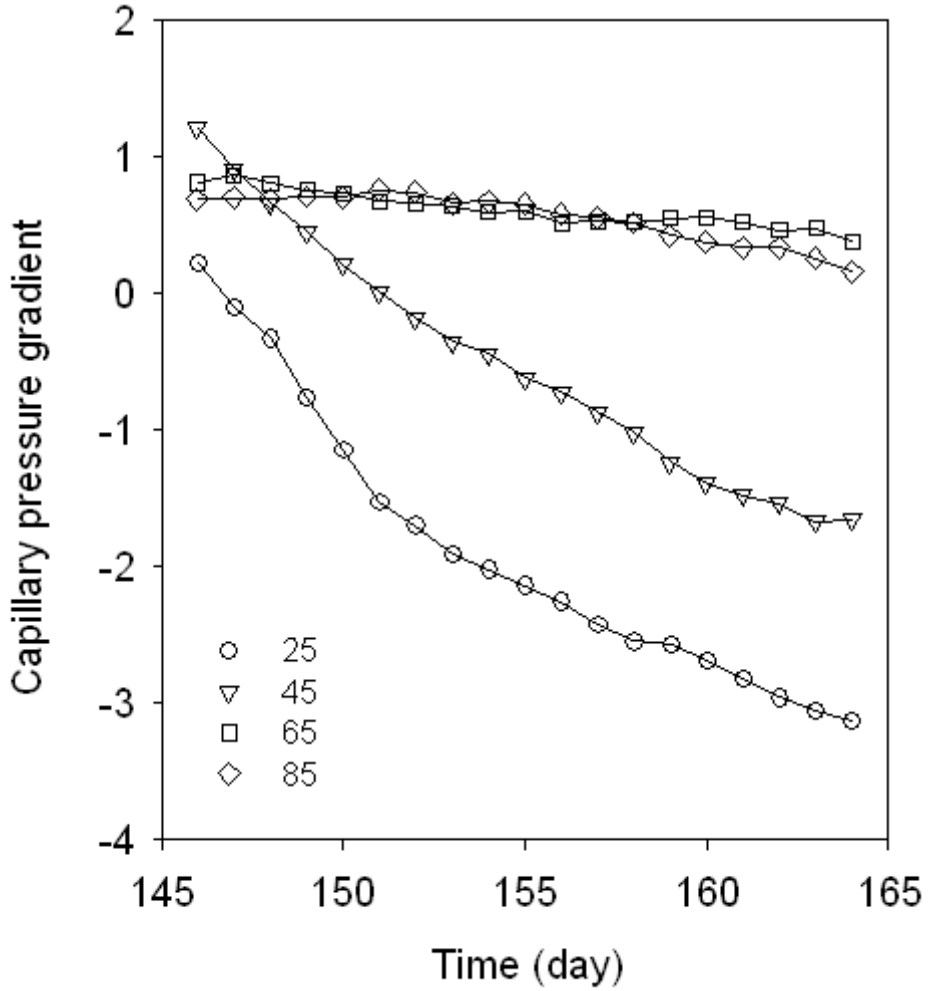

Figure 3-17. Capillary pressure gradient during a drying period in the location 10

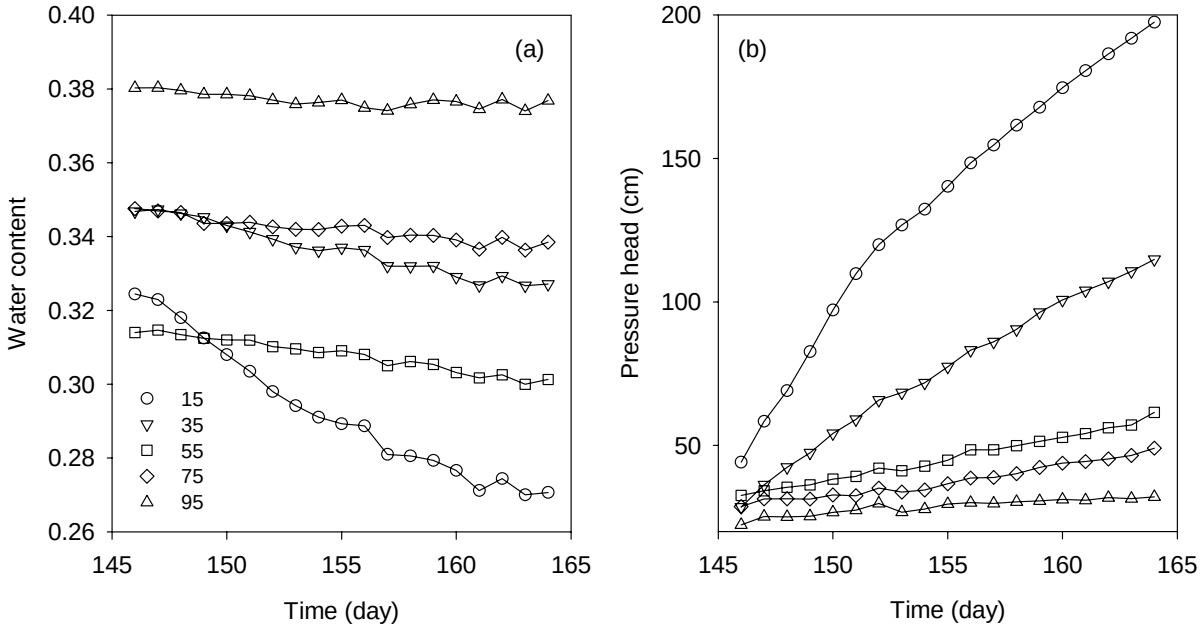

Figure 3-18. Daily average water content and pressure head time series at the location 10.

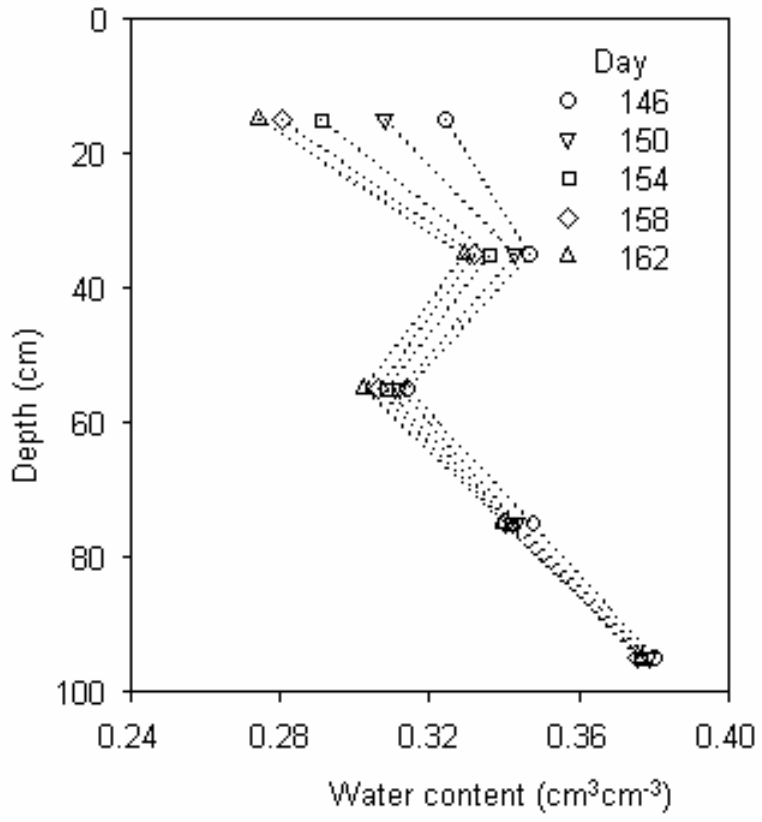

Figure 3-19. Water contents in soil profiles for several days of a drying period.

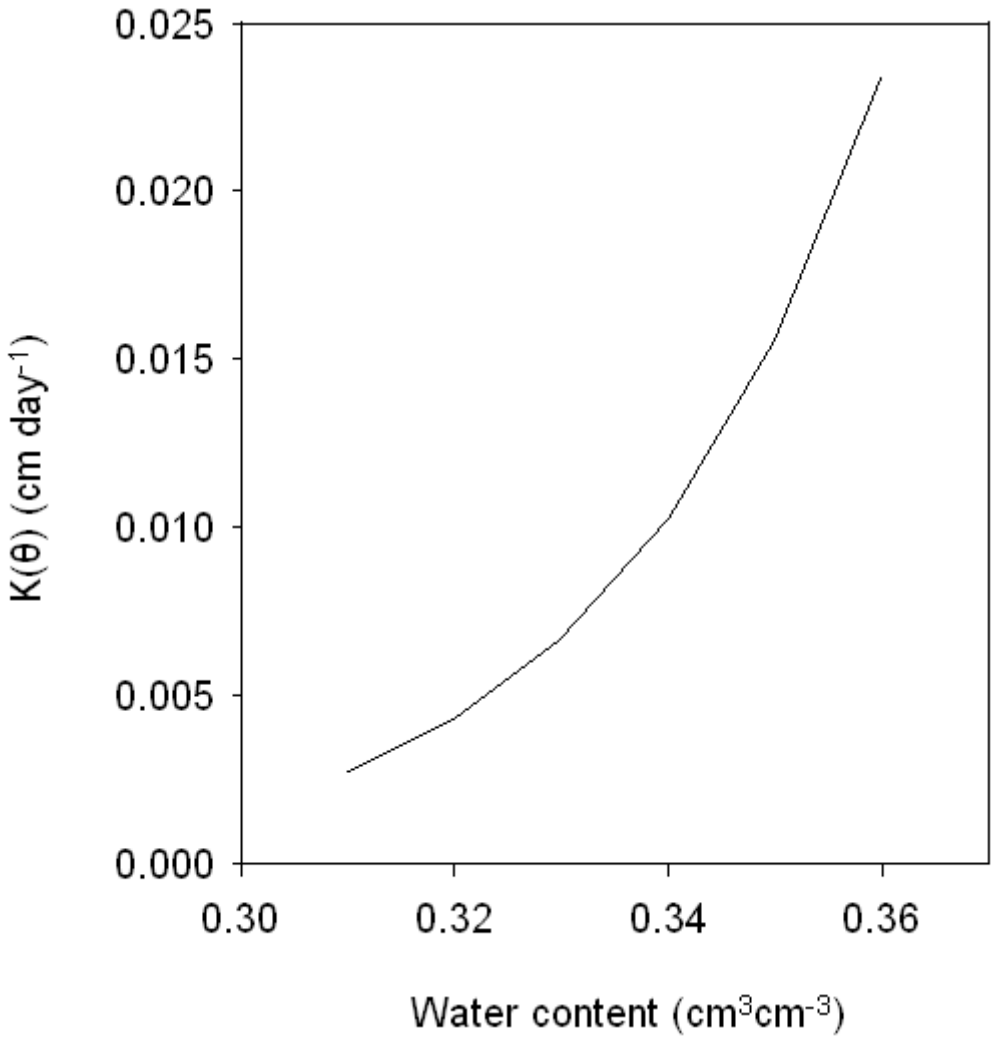

Figure 3-20. Estimated unsaturated hydraulic conductivity of soil matrix below 60 cm.

during the drying period, (b) the dependence of the unsaturated hydraulic conductivity on water content could be assumed the same within the depth range from 65 to 85 cm, (c) the linear interpolation was applicable to water contents. The procedure to estimate unsaturated hydraulic conductivity is given in the Appendix D. It is based on the combining the mass conservation equation and Darcy-Buckingham law and using the one-day time step. The estimated dependence of the unsaturated hydraulic conductivity on volumetric water content is shown in Fig. 3-20. This dependence was then used to compute the daily water flux from the soil profile at the depth of 105 cm as shown in the Appendix D.

Being able to compute the matrix flux at the bottom of the soil profile and changes in water storage of the soil profile, we could estimate evaporative losses during dry periods from the water mass balance. Estimated daily evaporation for five dry periods is shown in Fig. 3-21. The variability between dry periods is less than the variability within dry periods. The Kruskal-Wallis rank sum test did not indicate (p-value = 0.924) difference in mean daily evaporation values between dry periods, and ANOVA did not select the period as a influential factor at 0.05 probability level. The probability distribution function of estimated evaporation is given in Fig. 3-22. The values of daily evaporation were 0.29 mm, 0.86 mm and 1.51 mm at probability levels 25%, 50% and 75%, respectively. About 10% of the estimates are negative; this should be attributed to the error propagation, measurement noise, and a large time step.

3.4.4 Additional information to ensure sound abstraction

The purpose of collecting the additional information is to ensure that the abstracted models includes descriptions of all essential processes of flow and transport for given site. Although this additional information may be of lower quality compared with the necessary part of the database, it may show that some small-scale internal heterogeneities may have a dominant effect on the flow.

3.4.4.1 Evidence of the macropore flow

The presence of the macropore flow was one of the hypothetical reasons why the base model in HYDRUS-1D failed to simulate infiltration after intensive rainfall (Jacques et al. 2002). We analyzed the field water retention data and used the estimates of the surface evaporation from the section 3.3.2.3 to see whether the macropore flow might be present.

Fig. 3-23 shows field water retention at the 15-cm depth for the drying period from day 140 to day 190. The drying retention curves except curves at locations 110 and 310 have a horizontal segments at high saturations. At those segments, water content decreases and capillary pressure remains approximately constant as soil dries. The capillary pressure measured with the tensiometer reflects soil water potential in soil matrix. The decrease in soil water contents reflects the loss of soil water in macropores outside soil matrix. This loss occurs because both evaporation and redistribution. When the macropore water is lost, evaporation begins to draw water from soil matrix. This affects the tensiometer readings, and capillary pressure begins to rise simultaneously with the decrease of soil water contents. The schematic of such soil drying is shown in Fig. 3-24. This type of drying

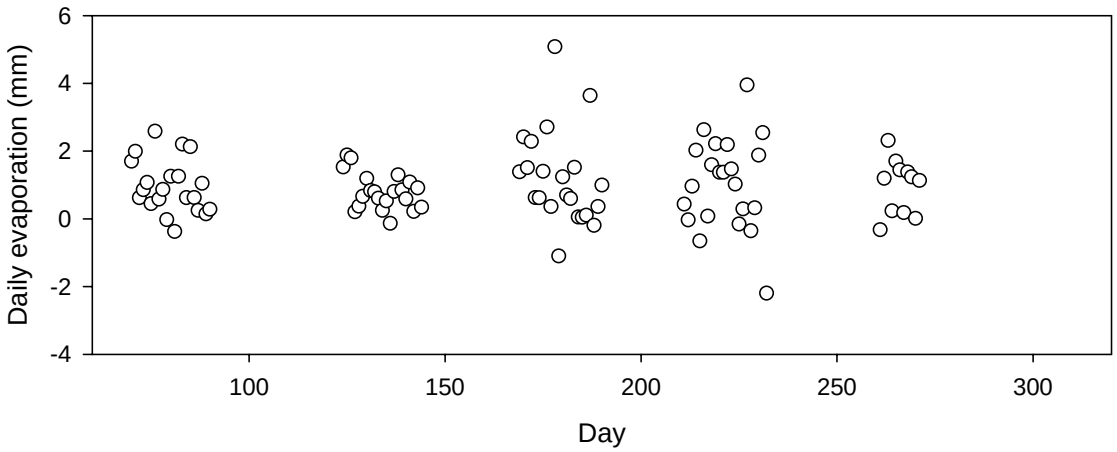

Figure 3-21. Estimated daily evaporation for dry periods during the experiment.

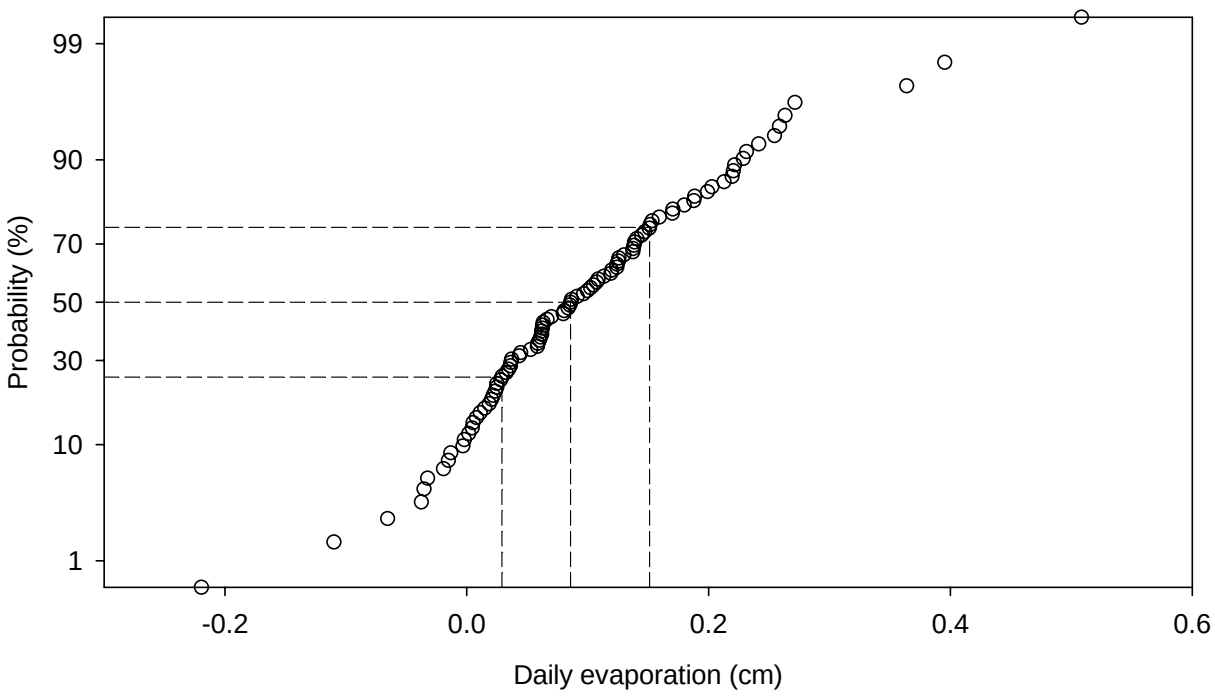

Figure 3-22. Probability distribution function of estimated daily evaporation.

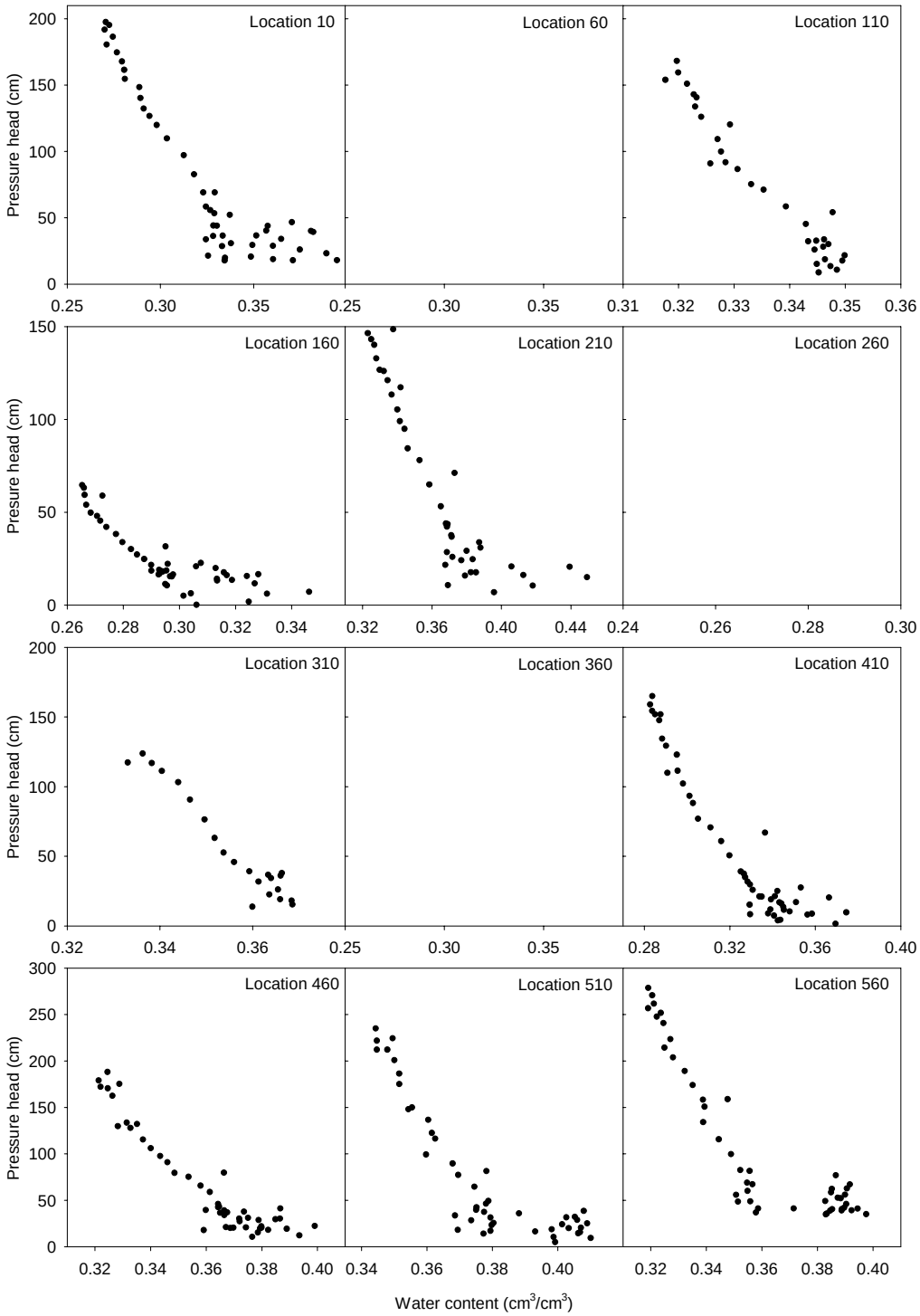

Figure 3-23. Relationships between pressure head and water content at the 15-cm depth between 140 and 190 day from the beginning of the experiment. Location numbering is given in the Fig. 3-3.

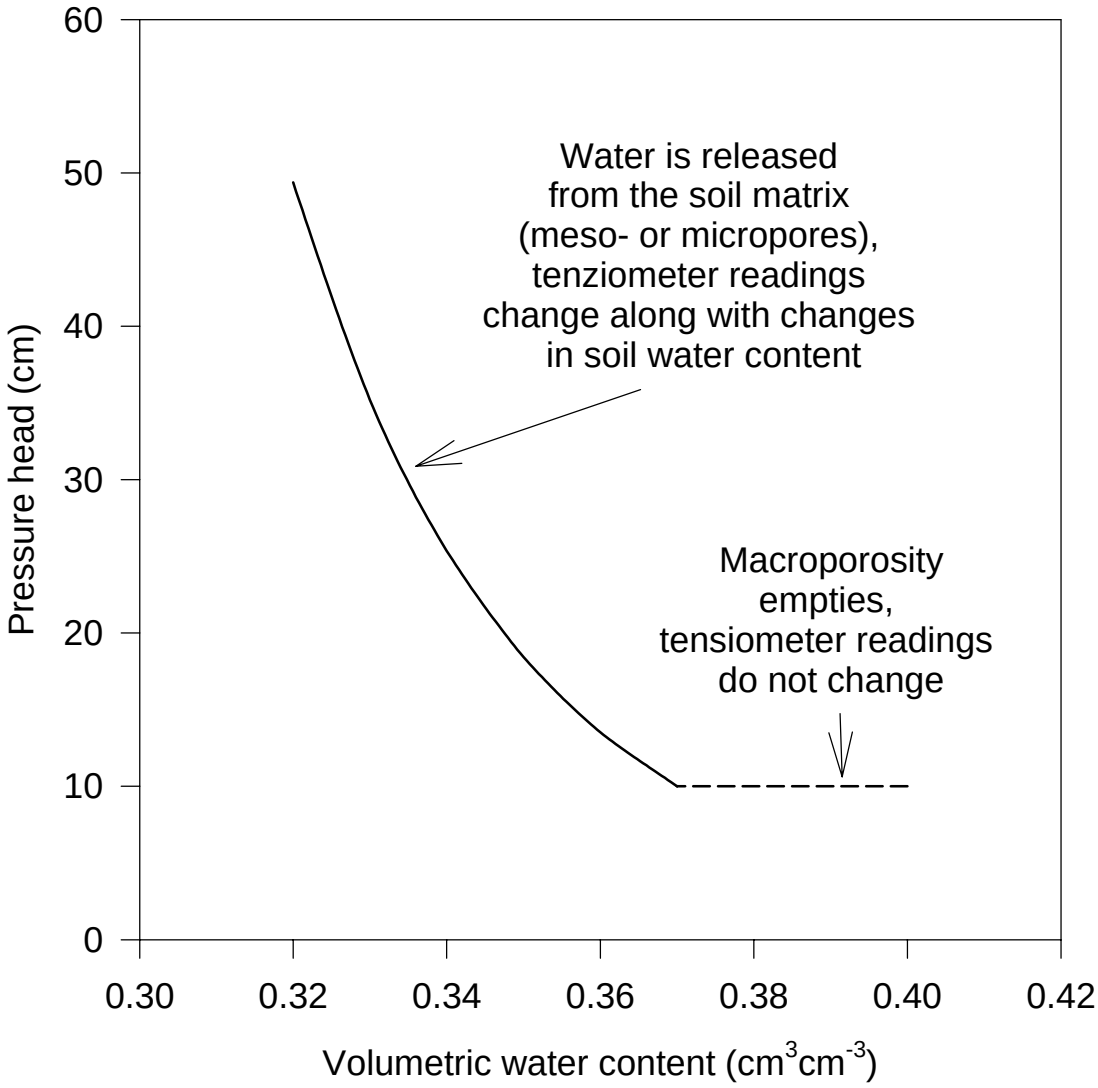

Figure 3-24. Drying water retention curve of a soil with macroporosity.

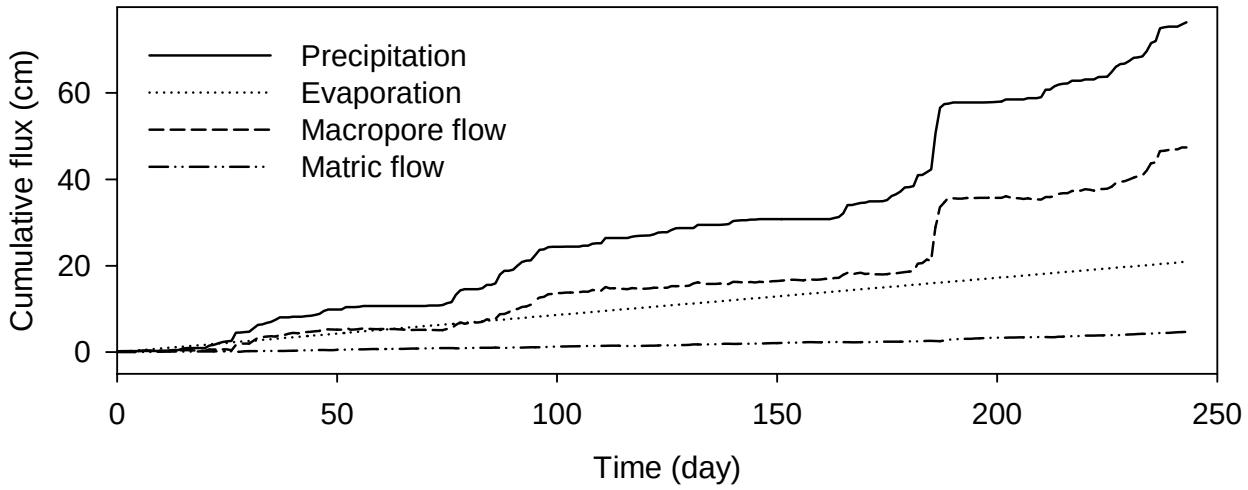

Figure 3-25. Cumulative water fluxes computed with constant evaporation of 0.86 mm day.

indicates the well-developed macroporosity. Therefore, soil at the site has necessary pore structure to exhibit the fast water flow in macropores in close-to-saturation conditions if case macropores have a good connectivity. Pore structure of the soil under study may serve as a conduit for the preferential flow which is, by definition, the non-uniform downward flow is substantially larger than the estimated matrix flow. It develops predominantly during the intermediate and wet periods and generally follows the pattern of precipitation.

The estimates of water balance, hydraulic conductivity, and surface evaporation are not necessarily accurate. Several strong, albeit plausible, assumptions have been made to obtain them. However, those estimates show that the substantial macropore flow in the soil could develop during intensive rainfall events. Macropores present small scale-heterogeneities in pore space that can substantially affect flow and transport in variably saturated subsurface. The abstracted models should include the ability to simulate macropore flow.

3.3.5 Recalibration of the base model

The Richards single continuum model was recalibrated using (a) water contents corrected for daily oscillations and temporal persistence (Sections 3.4.3.1 and 3.4.3.2) and (b) the estimated evaporation rate as boundary condition to see whether the inexplicable behavior of the model will disappear. Soil profile was subdivided into five layers 0-25, 25-45, 45-65, 65-85, and 85-105 cm. Four parameters θ_r, θ_s, α, n of the van Genuchten equation (2-2) and two parameters K_s, l of the Genuchten-Mualem equation (2-4) were estimated for each layer fitting the HYDRUS-1D to the corrected averaged water content. The initial estimates of the parameters were found using data on soil texture (Table 3-1) and the Rosetta software included in the HYDRUS-1D model. The total 30 parameters were estimated during the model calibration.

Calibration resulted in the relatively high accuracy (Fig. 3-26) of water contents. The RMSE (root-mean-square error) was 0.0058 cm^3cm^{-3}, the determination coefficient of the linear regression 'observed vs. simulated' water contents was $R^2 = 0.834$. Calibrated parameters along with their standard errors and linear tolerance intervals are shown in Table 3-6. Parameter θ_r is found with the lowest accuracy because it affects the water retention at high capillary pressures whereas the observed capillary pressures were very low (Fig. 3-8). In contrary, values of θ_s are very reliable. Parameters α and n are defined reliably only in the top layer. In all other layers, the tolerance intervals are very wide but the estimates themselves are acceptable. As indicated by Hill (1998), this suggests that the data in deep layers are insufficient for conclusive evaluation. And indeed, soil in top layer experiences some drying (Fig. 3-26), and therefore simulations are sensitive to the retention curve shape parameters α and n. In deeper layers, no substantial drying occurs, and the retention curve shape parameters α and n cannot be found reliably because all experimental points are located very close to the wet end. The same is true for the parameters of the hydraulic conductivity curve K_{sat} and l.

Although the calibration results were satisfactory (Fig. 3-26), simulated soil water fluxes were still substantially different from measured (Fig. 3-27). The F-statistic of 5.92 computed according Eq. (3-4) was above the critical value of 2.5. The Richards single

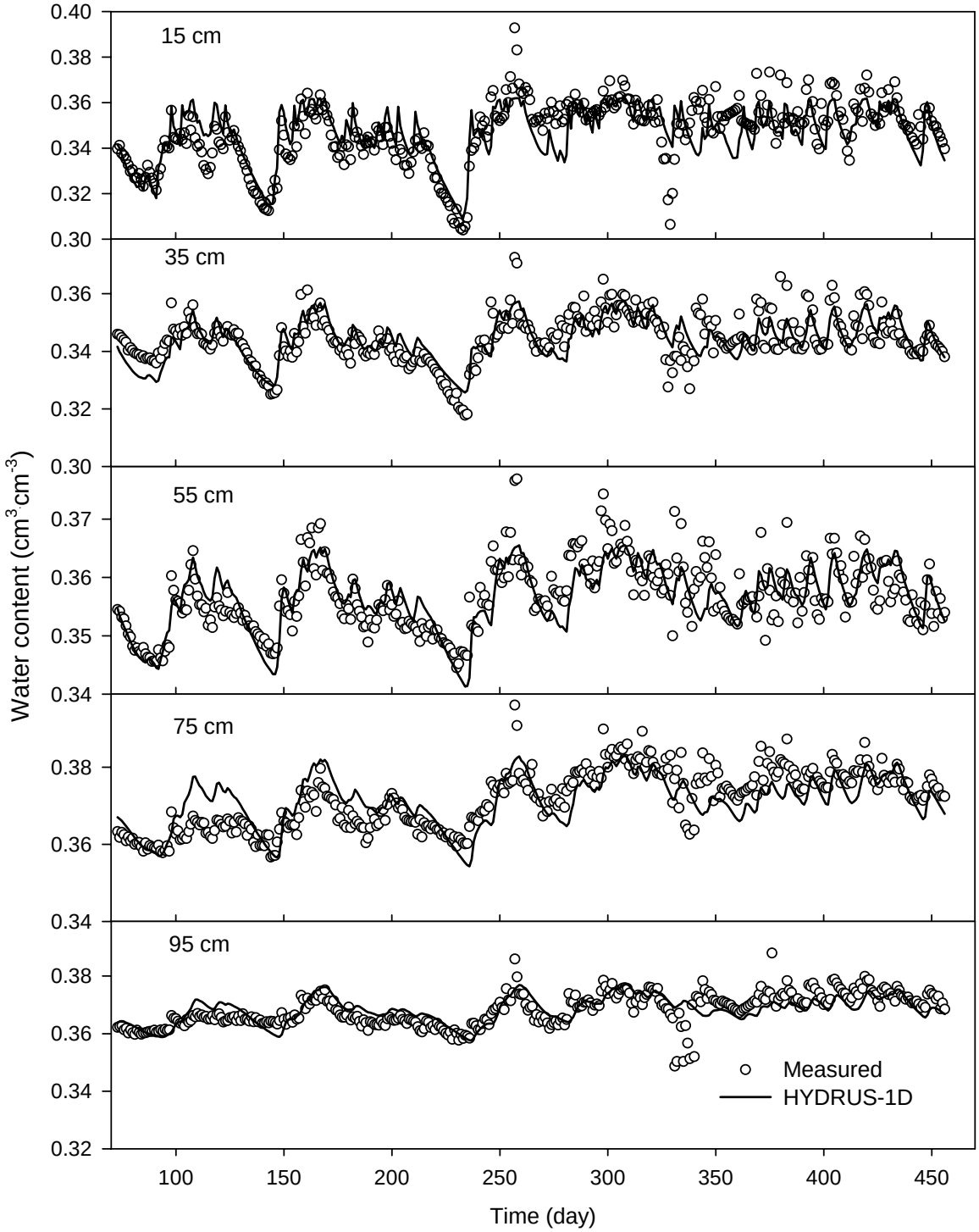

Figure 3-26. Time series of the depth-averaged water contents; ○ – measured, — simulated with the calibrated Richards medium model.

Table 3-6. Calibrated van Genuchten (Eq. 2-2) and van Genuchten-Mualem (Eq.2-5) parameters of the Richards medium model (HYDRUS-1D)

Variable	θ_r	θs	α	n	K_{sat}	l
	cm^3cm^{-3}	cm^3cm^{-3}	cm^{-1}		cm day^{-1}	
			0-25 cm			
Estimate	0.005	0.363	0.0011	1.8	0.066	0.01
Std. error	0.22	0.001	0.0004	0.189	0.006	0.83
Lower 95%	-0.427	0.361	0.0003	1.43	0.054	-1.63
Upper 95%	0.437	0.365	0.0019	2.17	0.078	1.64
			25-45 cm			
Estimate	0.086	0.362	0.00184	1.267	50	0.5
Std. error	1.409	0.008	0.004	1.136	409.61	36.4
Lower 95%	-2.678	0.347	-0.0061	-0.961	-753.34	-70.88
Upper 95%	2.85	0.377	0.0098	3.495	853.34	71.88
			45-65 cm			
Estimate	0.154	0.367	0.00123	1.527	0.207	5
Std. error	1.15	0.004	0.00438	1.35	0.3	54.4
Lower 95%	-2.102	0.359	-0.0074	-1.12	-0.381	-101.7
Upper 95%	2.41	0.375	0.00982	4.174	0.795	111.7
			65-85 cm			
Estimate	0.108	0.388	0.00133	1.492	0.167	0.5
Std. error	1.443	0.008	0.00426	1.441	0.329	24.87
Lower 95%	-2.722	0.372	-0.007	-1.334	-0.478	-48.28
Upper 95%	2.938	0.404	0.00969	4.318	0.812	49.28
			85-105 cm			
Estimate	0.01	0.386	0.00155	1.199	4.159	18.4
Std. error	0.453	0.004	0.00161	0.197	7.646	31.5
Lower 95%	-0.878	0.378	-0.0016	0.812	-10.836	-43.5
Upper 95%	0.898	0.394	0.0047	1.586	19.154	80.3

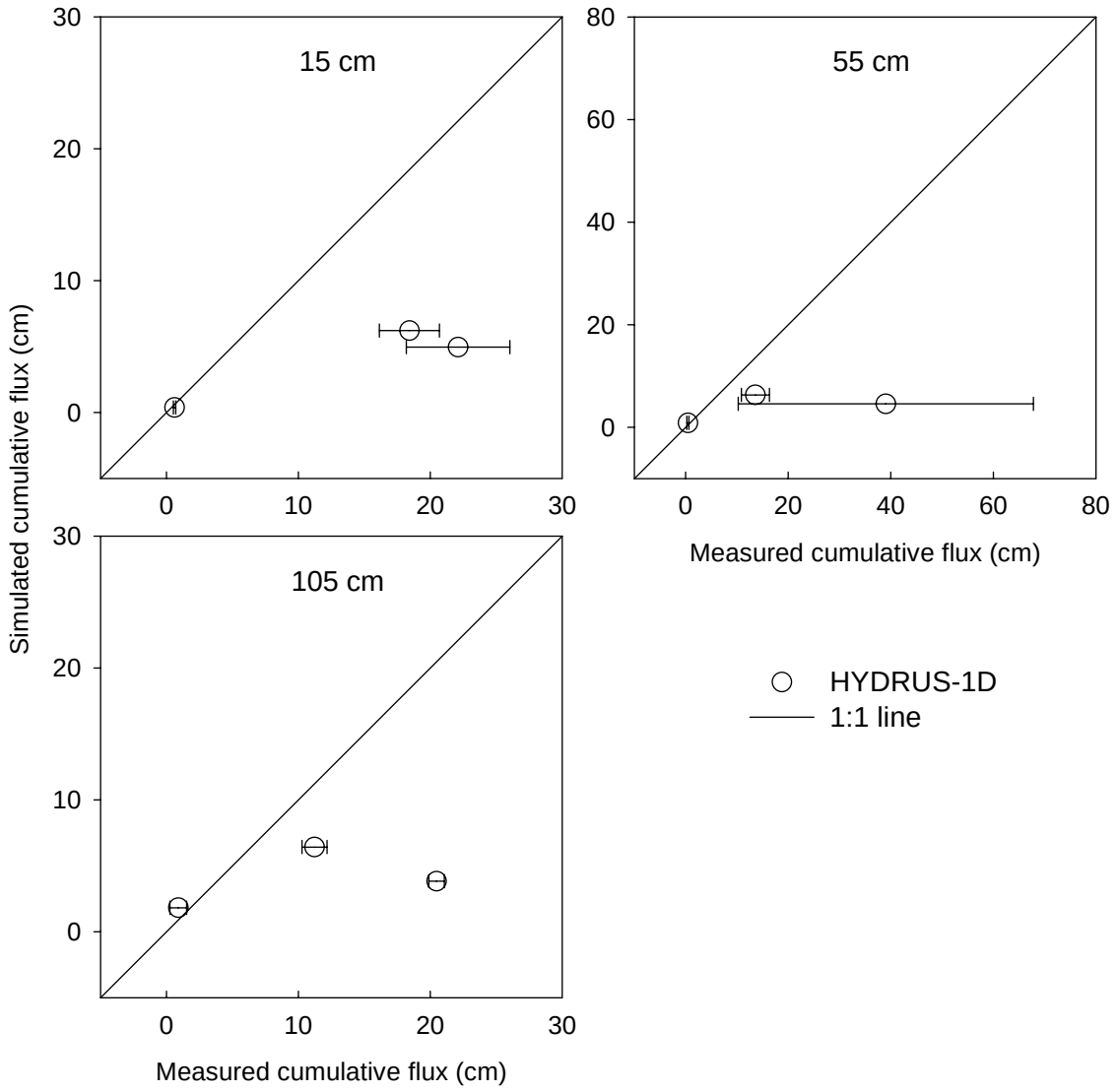

Figure 3-27. Soil water fluxes simulated with calibrated Richards medium model; error bars show one standard deviation

continuum model predicted either water accumulation at the soil surface or simulated substantial surface runoff during large storm events. As a result, simulated water fluxes through soil were much less than measured with passive capillary lysimeters or estimated from water balance. points are located very close to the wet end. The same is true for the parameters of the hydraulic conductivity curve K_{sat} and l.

Although the calibration results were satisfactory (Fig. 3-26), simulated soil water fluxes were still substantially different from measured (Fig. 3-27). The F-statistic of 5.92 computed according Eq. (3-4) was above the critical value of 2.5. The Richards single continuum model predicted either water accumulation at the soil surface or simulated substantial surface runoff during large storm events. As a result, simulated water fluxes through soil were much less than measured with passive capillary lysimeters or estimated from water balance.

Overall, the base model failed to simulate the key output when fitted to spatially and temporally corrected data using corrected boundary conditions. Therefore, the need in model abstraction persisted.

3.5 Selection and application of model abstraction techniques

3.5.1 Model abstraction design

The review of model abstraction techniques listed in Chapter 2 showed that the following four categories of the techniques are applicable to this case study: (a) the abstraction with a hierarchy of models, (b) the abstraction with scale change (aggregation); (b) the abstraction of parameter determination with pedotransfer functions, and (d) the abstraction with metamodeling. The operational scale was not changed, and no data were available to perform discretization or scaling in parameter determination. Therefore, other categories of abstraction were not used.

The design of the model abstraction process is shown in Fig. 3-28. Eight abstractions were performed. The base model is shown as the model 0 in Fig. 3-28. The model 1 is the result of model hierarchy abstraction, the model 2 is the result of the abstraction with scale change (aggregation); the model 3 is the result of the abstraction of parameter determination with pedotransfer functions, the model 4 is the result of the abstraction with metamodeling. Another four models are built as a result of the sequential abstraction as discussed below.

Model abstraction experiments were performed in the uncertainty context. Fifty simulations were done for each abstracted model. The random input was generated according to the statistical distributions of parameter values. Details of the Monte Carlo simulations are given in following subsections.

3.5.2 Abstraction with a hierarchy of models: using water budget

The hierarchy in Fig. 2-2 suggests using water budget model as a simplification of the single continuum, or Richards medium, model. The water budget model MWBUS (Model of Water Budget of Unsaturated Soil) has been used in this work. Detailed description of the model is in the Appendix E. The model is based on the model AQUA (Guber et al., 1995). The model is conceptually similar to the model MEPAS (Whelan et al, 1996). The difference is that MWBUS does not assume the steady-state flow and includes the depth-dependent loss of water for evaporation and/or transport to upper layers as proposed by Suleiman and Ritchie (2003). The model has the ability to simulate macropore flow as required in Section 3.3.4.1. The MWBUS is used in models 1, 5, 7, and 8 shown in Fig. 3-28.

The MWBUS was calibrated on corrected averaged water content using rainfall and estimated evaporation rate as upper boundary condition. The estimated parameters were saturated water content θ_s and field capacity θ_{fc} for five layers, and parameters of the evaporation equation (E4) a, b, and τ_0. The values of saturated hydraulic conductivity were assigned at 50% probability according to soil texture and porosity (Table 2-5) and were not changed during optimization. Thus, the total number of estimated parameters was equal to 13. The version of Marquardt algorithm (van Genuchten, 1981) was used in the calibration.

Results of the MWBUS calibration are shown in Fig. 3-29 and 3-30. As compared with the Richards medium model, the accuracy of the water content simulation with the calibrated MWBUS was worse (compare Fig. 3-29 and Fig. 3-30), the RMSE was 0.00719 cm^3cm^{-3} and the determination coefficient of the linear regression 'observed vs. simulated' water contents was R^2=0.740. However, the correspondence between MWBUS-simulated and measured soil water fluxes was satisfactory (Fig. 3-28). The F-statistic value was below the critical value (Fig. 3-31). Calibrated parameters of the MWBUS model and their statistics are given in Table 3-7. The MWBUS parameters are much more reliable compared with the parameters of the Richards model.

The abstraction from the Richards model to the water budget model not only provided a reasonable estimation of the soil water fluxes as the key output, but also served as a "sanity check" giving a clear indication that the Richards medium model was optimized in the wrong domain of its parameter space.

3.5.3 Abstraction with the scale change: profile aggregation

The layered soil profile was replaced with homogeneous one in this abstraction (model 2 in Fig. 3-28). The HYDRUS-1D was calibrated with averaged water content corrected as described in Sections 3.3.2.1 and 3.3.2.2. Four parameters θ_r, θ_s, α, n of the van Genuchten equation (2-2) and two parameters K_s, l of the Genuchten-Mualem equation (2-4) were estimated for the whole soil profile fitting the HYDRUS-1D model. The total number of estimated parameters was equal to 6. Results of the calibration are shown in Fig. 3-32. Replacing the layered profile with homogeneous profile has more effect on errors in water contents than on errors in flux computations (compare Fig 3-29 with 3-32, and Fig. 3-30 with Fig. 3-33). Water fluxes were not simulated correctly after this abstraction. The F-statistic computed according Eq. 3-4 was above the critical value (Fig. 3-31).

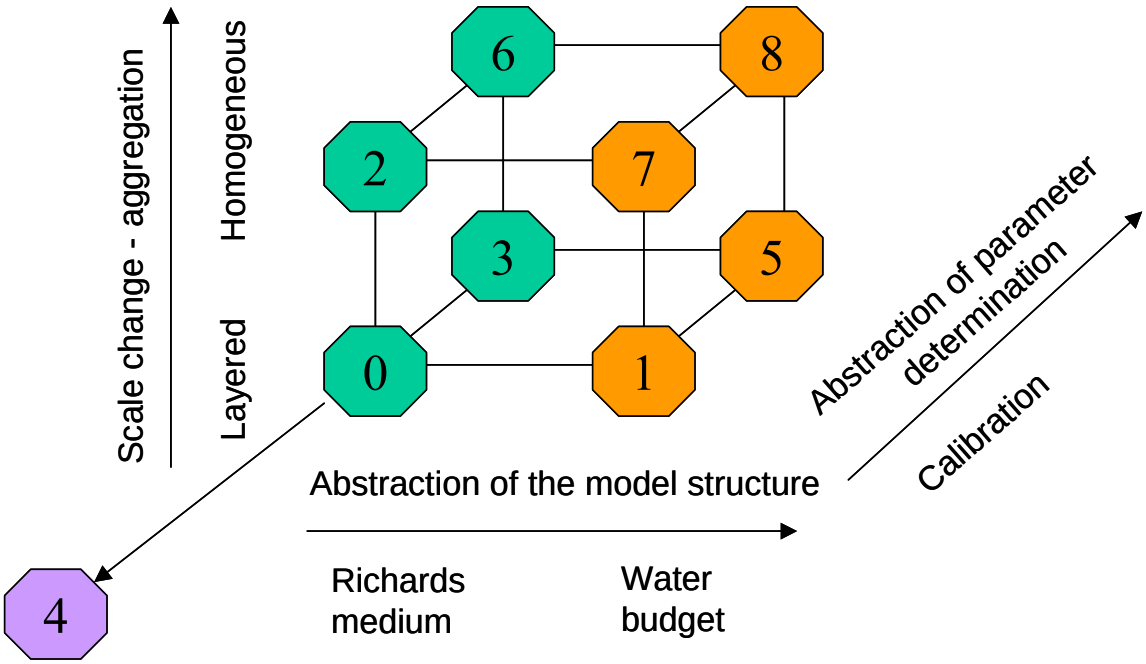

Figure 3-28. Design of the model abstraction application in this work. The Richards media was simulated with the HYDRUS-1D software; the water budget model was MWBUS.

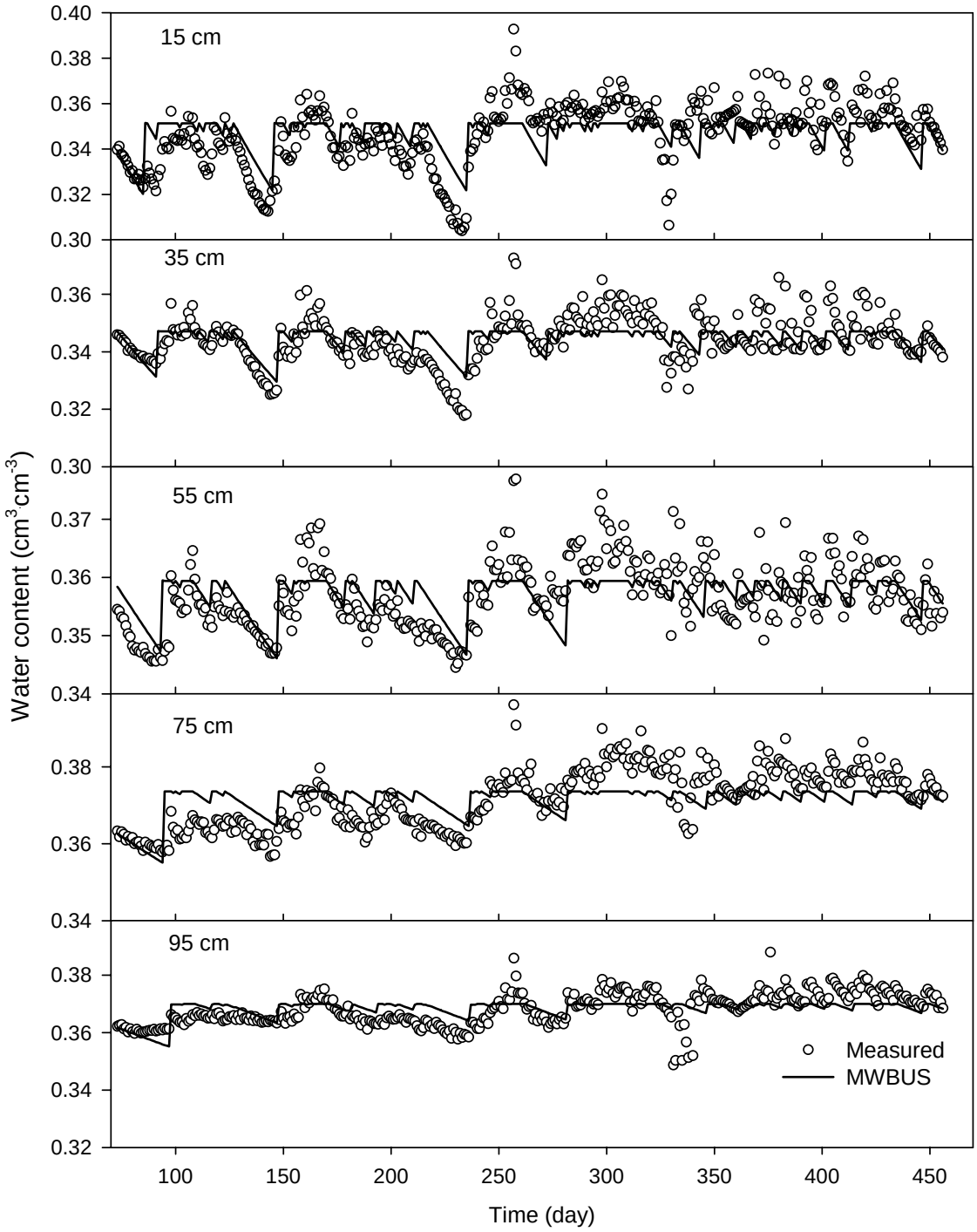

Figure 3-29. Time series of the depth-averaged water contents; ○ – measured, — simulated with the calibrated MWBUS model.

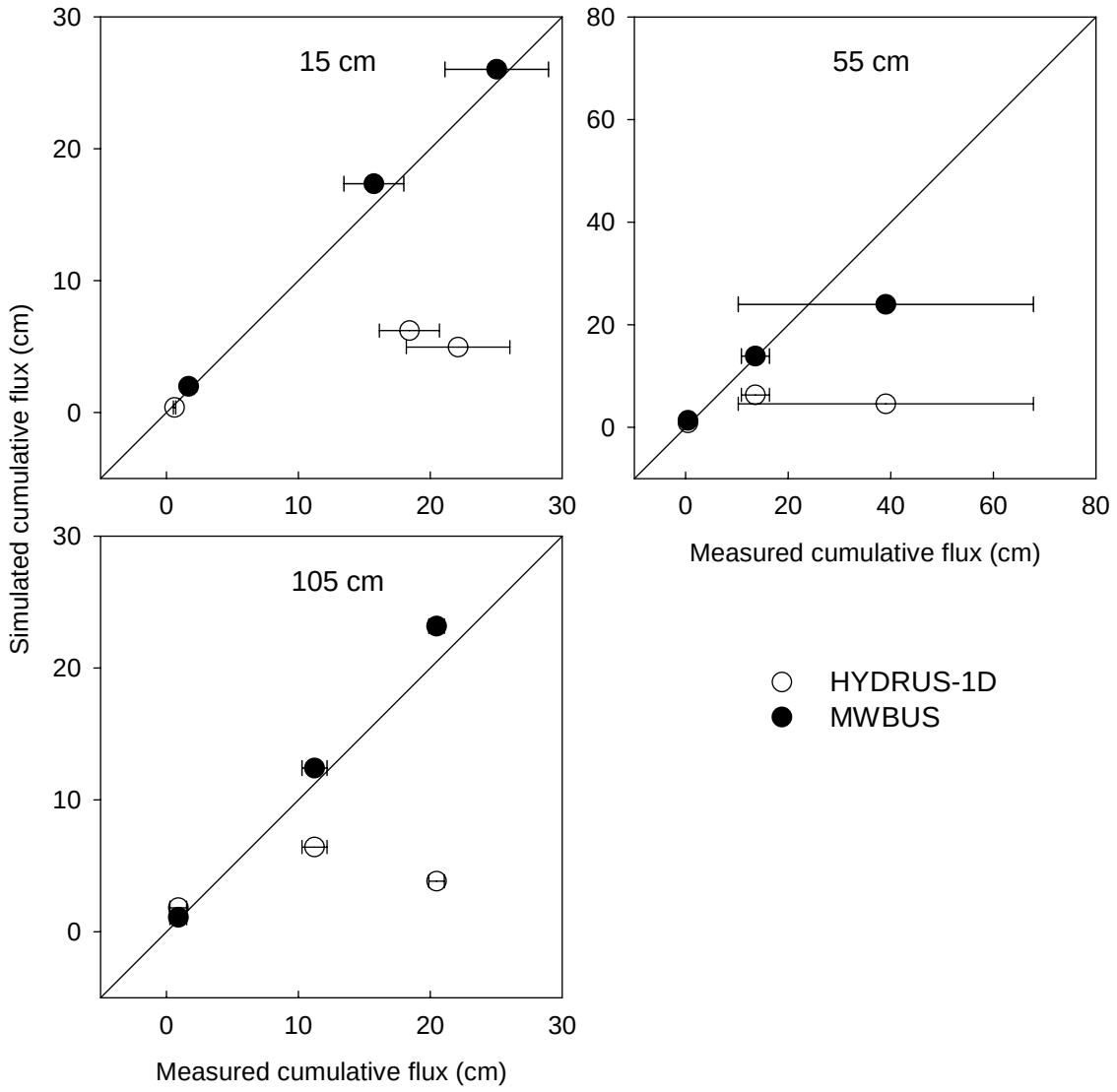

Figure 3-30. Soil water fluxes simulated with calibrated HYDRUS-1D and the MWBUS models; error bars show one standard deviation

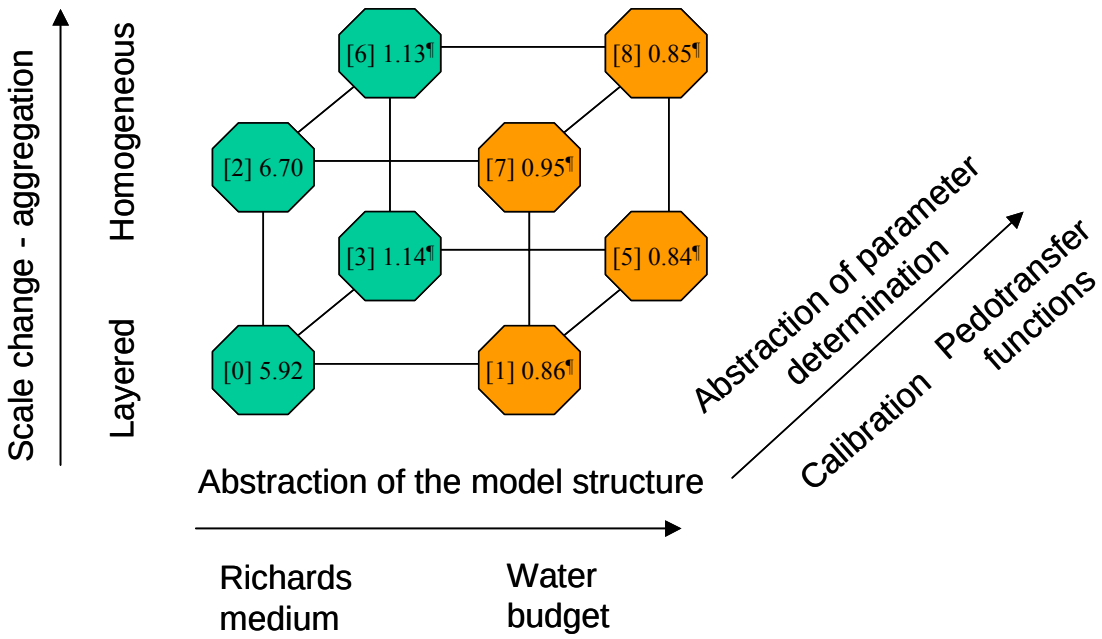

Figure 3-31. F-statistics of soil water flux simulations with different abstraction techniques for depths of 15 and 55 cm.; ¶ - the evaluated F-value is less than the critical F-value at 95% probability; model numbering in brackets is the same as in Fig. 3-28. The critical value of F-statistics is 2.5.

Table 3-7. Calibrated parameters of the water budget model (MWBUS)

Variable	θs	θ_{fc}	K_{sat}	θr[¶]
	cm^3cm^{-3}	cm^3cm^{-3}	cm day^{-1}	cm^3cm^{-3}
		Layered soil profile		
		0-25 cm		
Estimate	0.432	0.353	133.9	0.09
Std. error	0	0.000	0	
Lower 95%	0.432	0.353	133.9	
Upper 95%	0.432	0.354	133.9	
		25-45 cm		
Estimate	0.384	0.348	133.9	0.101
Std. error	0	0.000	0	
Lower 95%	0.384	0.347	133.9	
Upper 95%	0.384	0.349	133.9	
		45-65 cm		
Estimate	0.388	0.360	133.9	0.108
Std. error	0	0.000	0	
Lower 95%	0.388	0.359	133.9	
Upper 95%	0.388	0.361	133.9	
		65-85 cm		
Value	0.388	0.374	14.9	0.123
Std. error	0	0.000	0	
Lower 95%	0.388	0.374	14.9	
Upper 95%	0.388	0.375	14.9	
		85-105 cm		
Estimate	0.388	0.370	14.9	0.109
Std. error	0	0.001	0	
Lower 95%	0.388	0.368	14.9	
Upper 95%	0.388	0.371	14.9	
		Homogeneous soil profile		
Variable	k_0	a	b	
	day	day cm^{-1}		
Estimate	109.7	0.602	1.579	
Std. error	3.2	0.023	0.012	
Lower 95%	103.3	0.557	1.555	
Upper 95%	116.0	0.646	1.603	

[¶]Was estimated from pedotransfer function Rosetta in HYDRUS-1D

3.5.4 Abstraction of parameter determination: using an ensemble of pedotransfer functions

This abstraction corresponds to the model 3 in Fig. 3-28. Calibration was replaced with using pedotransfer functions for both water retention and hydraulic conductivity. Results of application of 21 pedotransfer functions given in the Appendix A and the Rosetta software are shown in Fig. 3-34. None of the PTFs provide a good approximation of the field water retention, although the field variability makes questionable any attempt of PTF-based predictions, just because no data on PTF inputs collocated with the measurement points have been available. Nevertheless, the ensemble of pedotransfer functions creates an envelope that encompasses the range of field water retention data. Similar results were obtained at other depths (data not shown).

The Monte Carlo simulations were performed with HYDRUS-1D model using each pedotransfer function from the Appendix 1 and the Rosetta pedotransfer function to estimate water retention. Table 3-8 indicates whether Brooks-Corey equation (2-1) or van Genuchten equation (2-2) was used in simulations with a specific pedotransfer function. The median values of lognormal distributions of K_{sat} shown in Table 2-5 and in Fig. 2-10 were used from loam and silt loam texture classes at depths ranges 0 to 65 cm and 65 to 105 cm, respectively. Results of water content simulations are shown in Fig. 3-35. The accuracy of those simulations in terms of water contents was low. There has obviously been a mismatch between the simulated initial distributions of pressure heads and the actual ones, and the first period of simulations was characterized by the substantial loss of water through the bottom of the profile. However, the accuracy of soil water fluxes estimates in simulations with pedotransfer functions was satisfactory (Fig. 3-36). The F-statistic value was below the critical value (Fig. 3-31). These results show that the Richards media model is, in principle, capable of correct prediction of soil water fluxes in this case study provided realistic parameter estimates are used.

3.5.5 Sequential model abstractions

Models 5, 6, and 7 in Fig. 3-28 represent application of two model abstraction techniques. The purpose of such sequential application is to find out whether additional gains can be obtained without the loss in accuracy.

Model 5 in Fig. 3-28 represents using pedotransfer functions with soil water budget model. The Monte Carlo simulations were performed with the MWBUS model using each pedotransfer function from the Appendix 1 and the Rosetta pedotransfer function to estimate water retention. Water retention was estimated using average soil texture for depth 15 and 35 cm for topsoil, and 55, 75, 95 cm for subsoil. The Monte Carlo simulations required a decision on how to estimate field capacity from such data. Although there is no

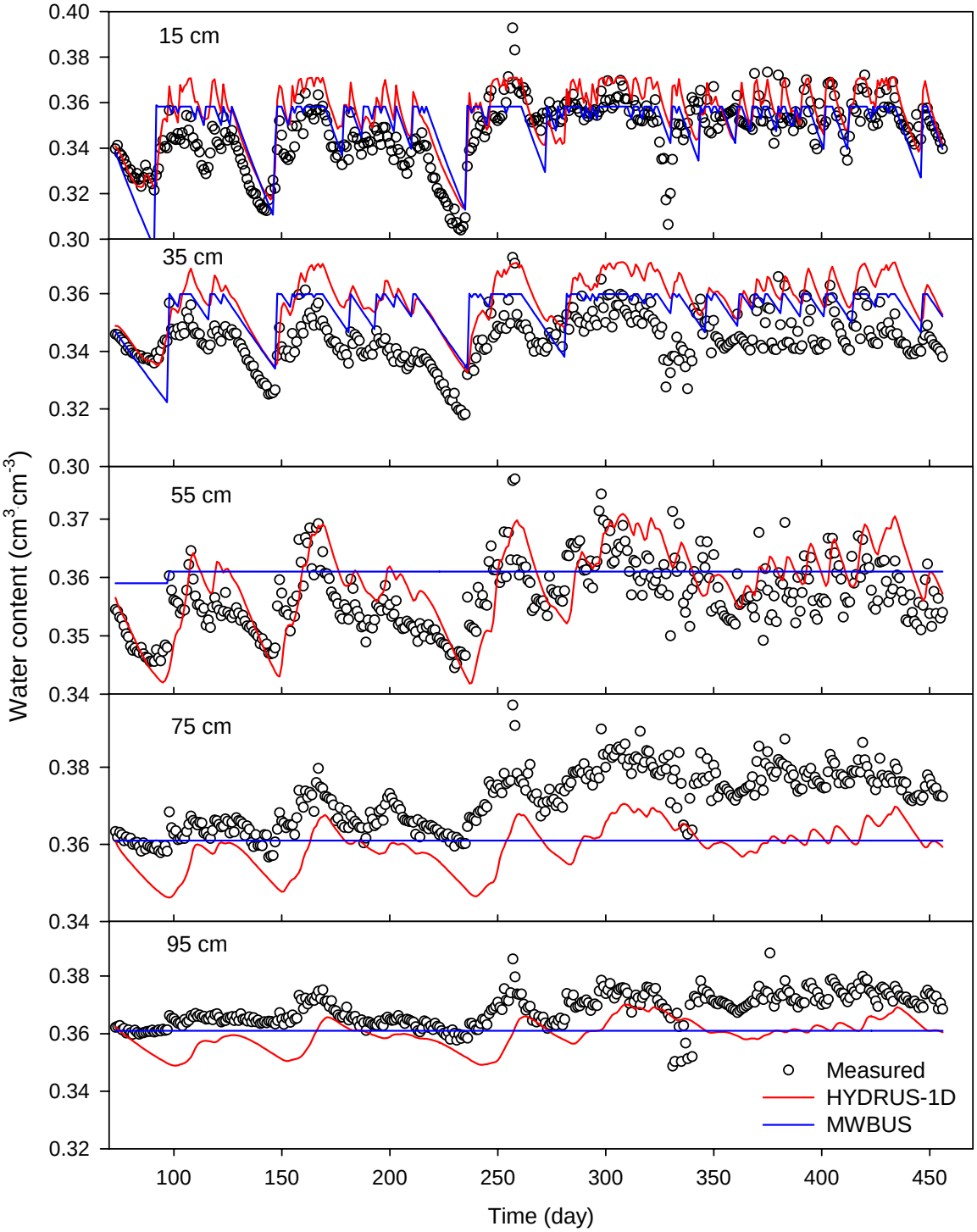

Figure 3-32. Time series of transect-average water contents: symbol – measured, red lines – calibrated the HYDRUS-1D model, blue lines - calibrated the WMBUS model for homogeneous soil profile.

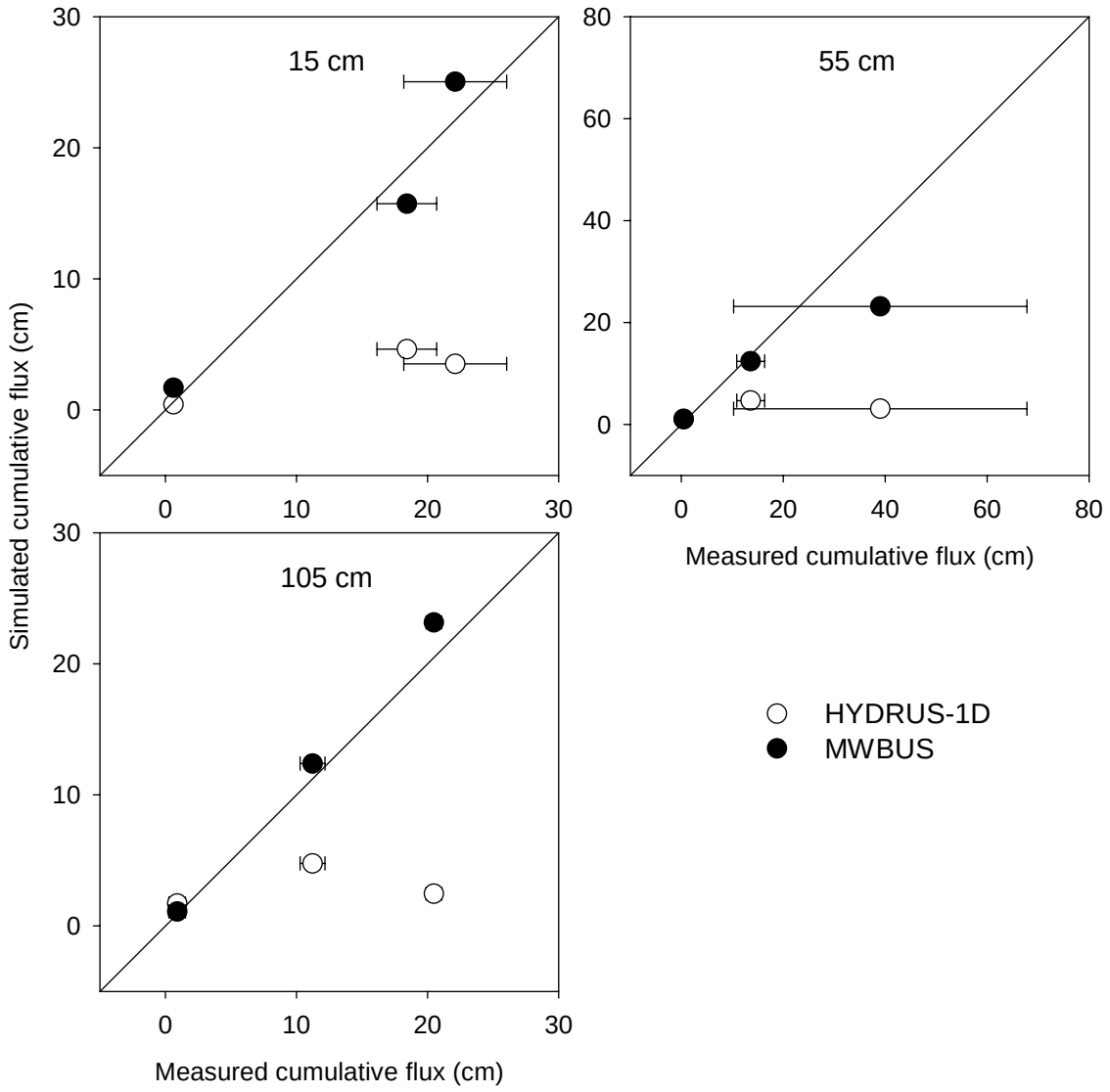

Figure 3-33. Cumulative water fluxes simulated with calibrated HYDRUS-1D and MWBUS models for homogeneous soil profile; error bars show one standard deviation.

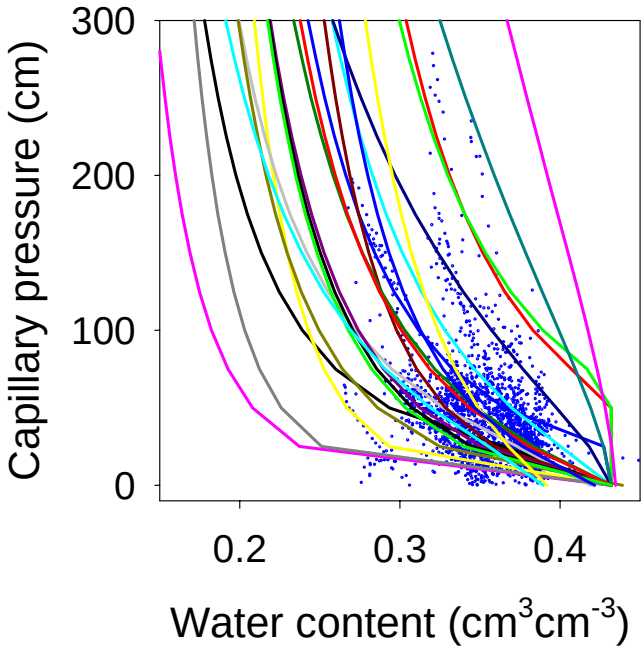

Figure 3-34. Comparison of field water retention (symbols) at the depth of 15 cm with water retention estimated using 21 pedotransfer functions listed in the Appendix A and the Rosetta software.

Table 3-8. PTF used to run HYDRUS 1D with Brooks-Corey (BC) and Van Genuchten (VG) water retention equations

#	Authors		Clay	Silt	Sand	BD[¶]	OM[§]	topsoil
1	Campbell & Shiozawa (1992)	BC	+	-	+	+	-	-
2	Rawls & Brakensiek (1989)	BC	+	-	+	+	-	-
3	Saxton et al. (1986)	BC	+	-	+	-	-	-
4	Oosterveldt & Chang (1980)	BC	+	-	+	-	-	-
5	Williams et al. (1992)	BC	+	-	+	-	+	-
6	Williams et al. (1992)	BC	+	-	+	+	-	-
7	Baumer (1992)	VG	+	-	+	+	+	-
8	Bruand et al. (1994)	VG	+	-	-	-	-	-
9	Canarache (1993)	VG	+	-	-	+	-	-
10	Gupta & Larson (1979)	VG	+	+	+	+	+	-
11	Hall et al. (1977)	VG	+	+	+	+	-	-
12	Mayr & Jarvis (1999)	VG	+	+	+	+	+	-
13	Pachepsky et al. (1982)	VG	+	-	+	+	-	-
14	Petersen et al. (1968)	VG	+	-	-	-	-	-
15	Rajkai & Varallyay (1992)	VG	+	-	+	+	+	-
16	Rawls et al. (1982)	VG	+	+	+	+	-	-
17	Rawls et al. (1992)	VG	+	+	+	+	+	-
18	Verekeen et al. (1989)	VG	+	-	+	+	+	-
19	Tomasella & Hodnett (1998)	VG	+	+	-	-	-	-
20	Wosten et al. (1999)	VG	+	+	+	-	-	+
21	Wosten et al. (1999)	VG	+	+	+	+	+	+
22	Rosetta	VG	+	+	+	+	-	-

[¶]Average bulk density values were taken from Mallants et al. (1996) as 1.426, 1.544, and 1.526 g cm^{-3} in the depth ranges 0 to 25 cm, 25 to 55 cm and 55 to 105 cm, respectively

[§]Organic matter content was set at 2% in the 0-25 cm topsoil, and 0.5% in subsoil.

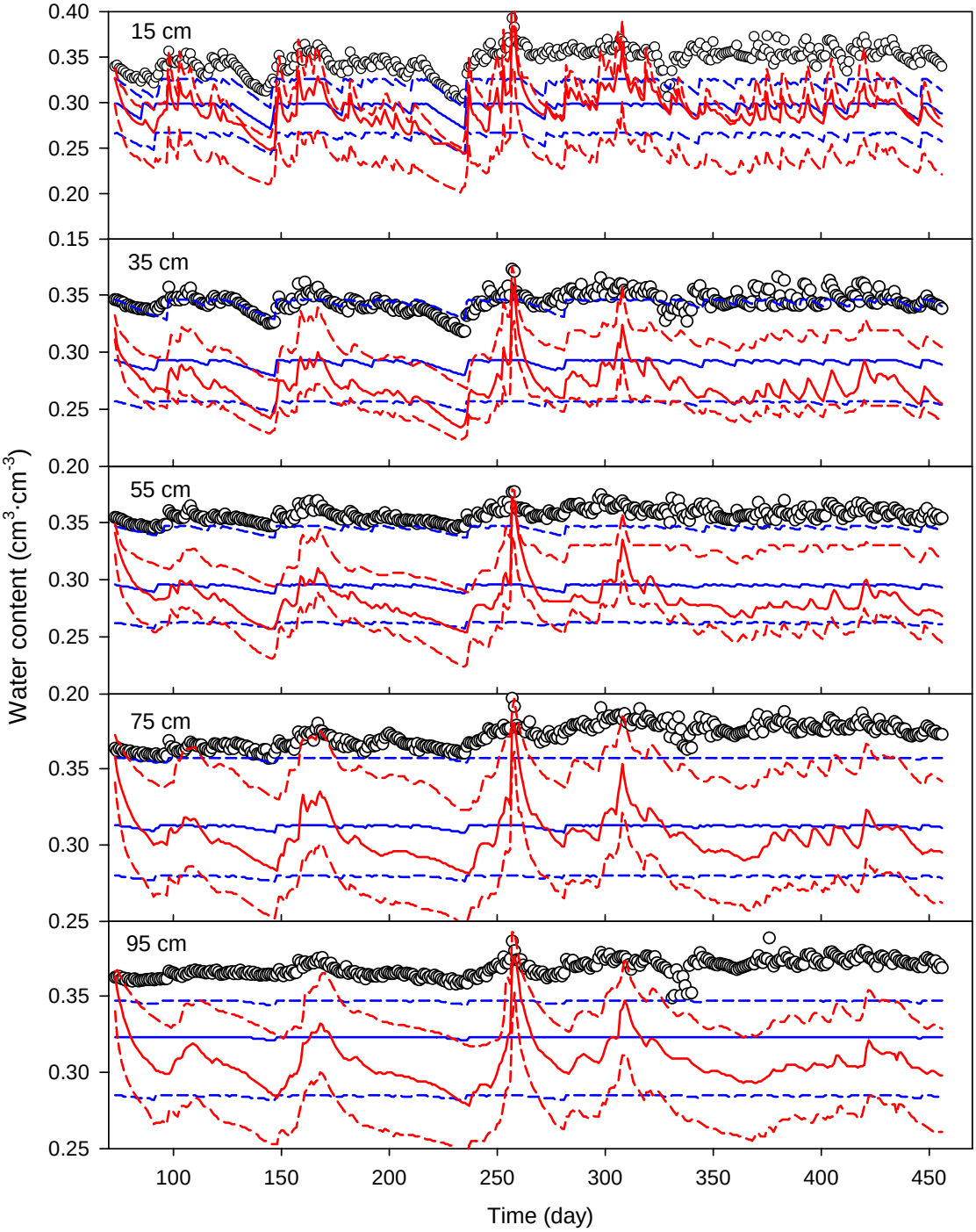

Figure 3-35. Time series of transect-average water contents: symbol – measured, red lines - simulated with the HYDRUS-1D, blue lines - simulated with the WMBUS model; water retention data and median hydraulic conductivity estimated with pedotransfer functions; the solid line corresponds to the median value of simulated water contents for each day, dashed lines correspond to the simulated daily values at 25% and 75% probability levels.

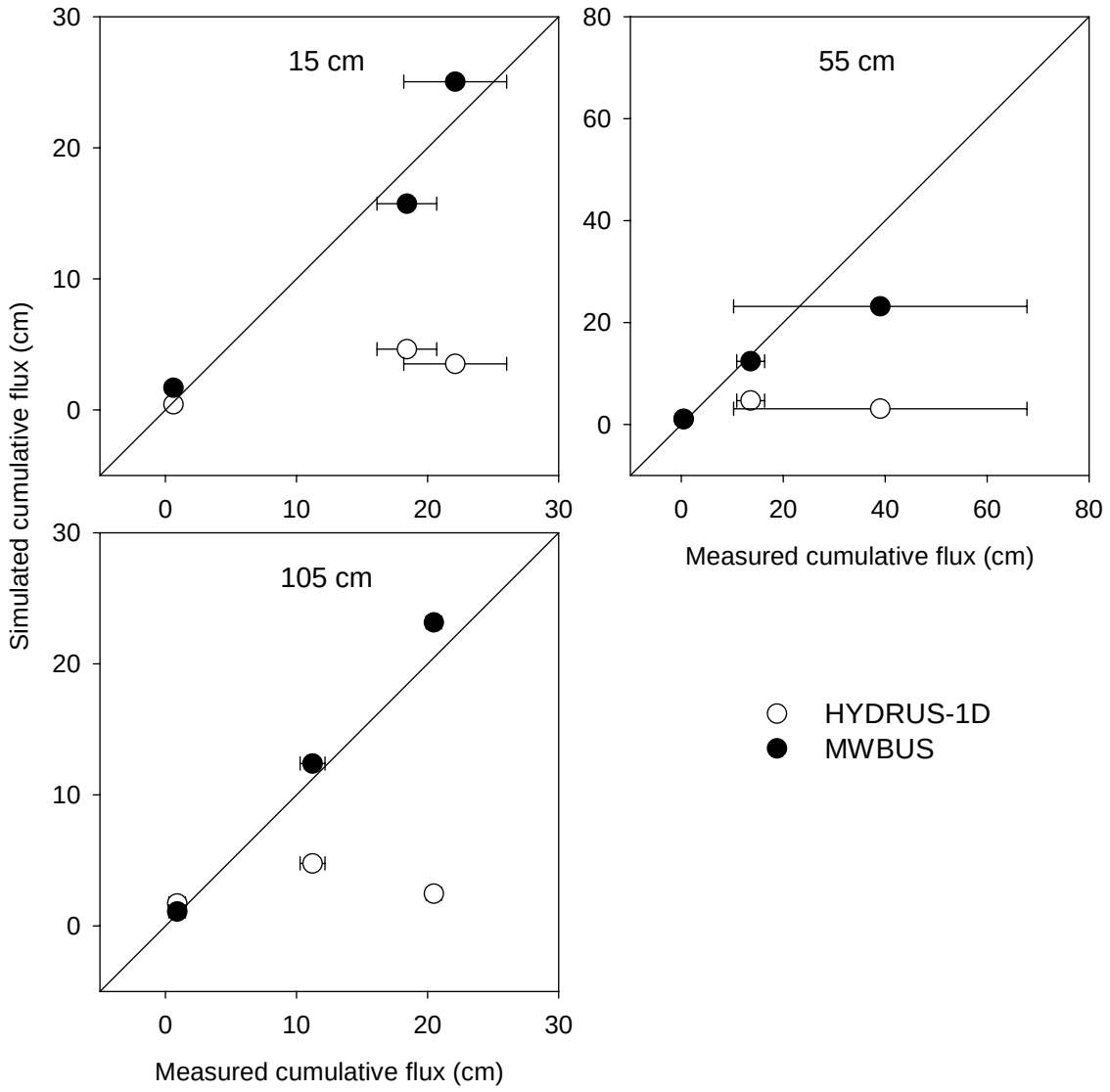

Figure 3-36. Soil water fluxes simulated with HYDRUS-1D and MWBUS models; water retention data and median hydraulic conductivity estimated with pedotransfer functions for layered soil profile; error bars show one standard deviation.

unique capillary pressure value to estimate field capacity from 'water content-capillary pressure' dependencies (see section 2.2 and Fig. 2-6), two often-suggested values of soil water potential for such estimate are 100 cm and 330 cm. We used the laboratory water retention data to select the capillary suction corresponding to the field capacity. The distributions of water content at both capillary pressures were constructed by random sampling the distributions of van Genuchten parameters shown in Fig. 3-12, and the soil water was simulated with both distributions. The root-mean-square errors with water contents at capillary pressure of 100 cm as field capacity values were about two times less than the errors with water contents at capillary pressure of 330 cm (data not shown). Therefore, to perform the MWBUS model with parameters estimated using PTF (Table 3-8), the values of θ_{fc} were evaluated at pressure head of 100 cm. Parameters of the evaporation model for the MWBUS were taken from the calibration as described in 3.5.2 Simulations were run for each function with the median values of lognormal distributions of K_{sat} shown in Table 2-5 and in Fig. 2-10 that were used from loam and silt loam texture classes at depths ranges 0 to 65 cm and 65 to 105 cm, respectively. Results of water content simulations are shown in Fig. 3-35. The accuracy of those simulations was low. There has obviously been a mismatch between water field capacity estimated with PTF and the actual ones. The first period of simulations was characterized by the substantial loss of water through the bottom of the profile. However, the accuracy of soil water fluxes estimates in simulations with pedotransfer functions was satisfactory (Fig. 3-36).

Model 6 in Fig. 3-28 represents using pedotransfer functions after the aggregation of the soil profile is performed. Soil texture was different for topsoil and subsoil in the soil profile (Fig. 3-3). Pedotransfer functions require the soil textural composition as the input. Therefore two cases were considered separately: homogeneous soil profile with topsoil hydraulic properties, homogeneous soil profile with subsoil hydraulic properties.

Results of simulations are shown in Fig. 3-37 and 3-38. Simulated water content and water fluxes were similar to those, obtained for layered soil profile (Fig. 3-35 and 3-36). The Richards media model underestimated water content at all depths. The accuracy of flux predictions was acceptable. The F-statistic computed for water fluxes was below the critical value (Fig. 3-31). Replacing topsoil hydraulic parameters with subsoil parameters did not result in noticeable changes in water content and flux values for both models (data not shown). That might be due to variability in water retention caused by different PTF is greater than that caused by changes in soil texture for particular soil profile.

Overall, the model abstraction with scale change (aggregation) resulted in deterioration of the accuracy of soil water content simulations with both HYDRUS-1D and MWBUS models. Soil water fluxes, however, were simulated well on both abstraction steps.

Model 7 in Fig. 3-28 represents the sequential use of hierarchy of models and aggregation. The MWBUS has the single homogeneous layer and is calibrated for this case. Calibrated parameter values and their statistics are shown in Table 3-7. The tolerance intervals are narrow, parameter values are robust. Two parameters - saturated water content θ_s and field capacity θ_{fc}, - and parameters of the evaporation equation (E4) a, b, and τ_0 were estimated fitting the MWBUS model. The values of saturated hydraulic conductivity were assigned at

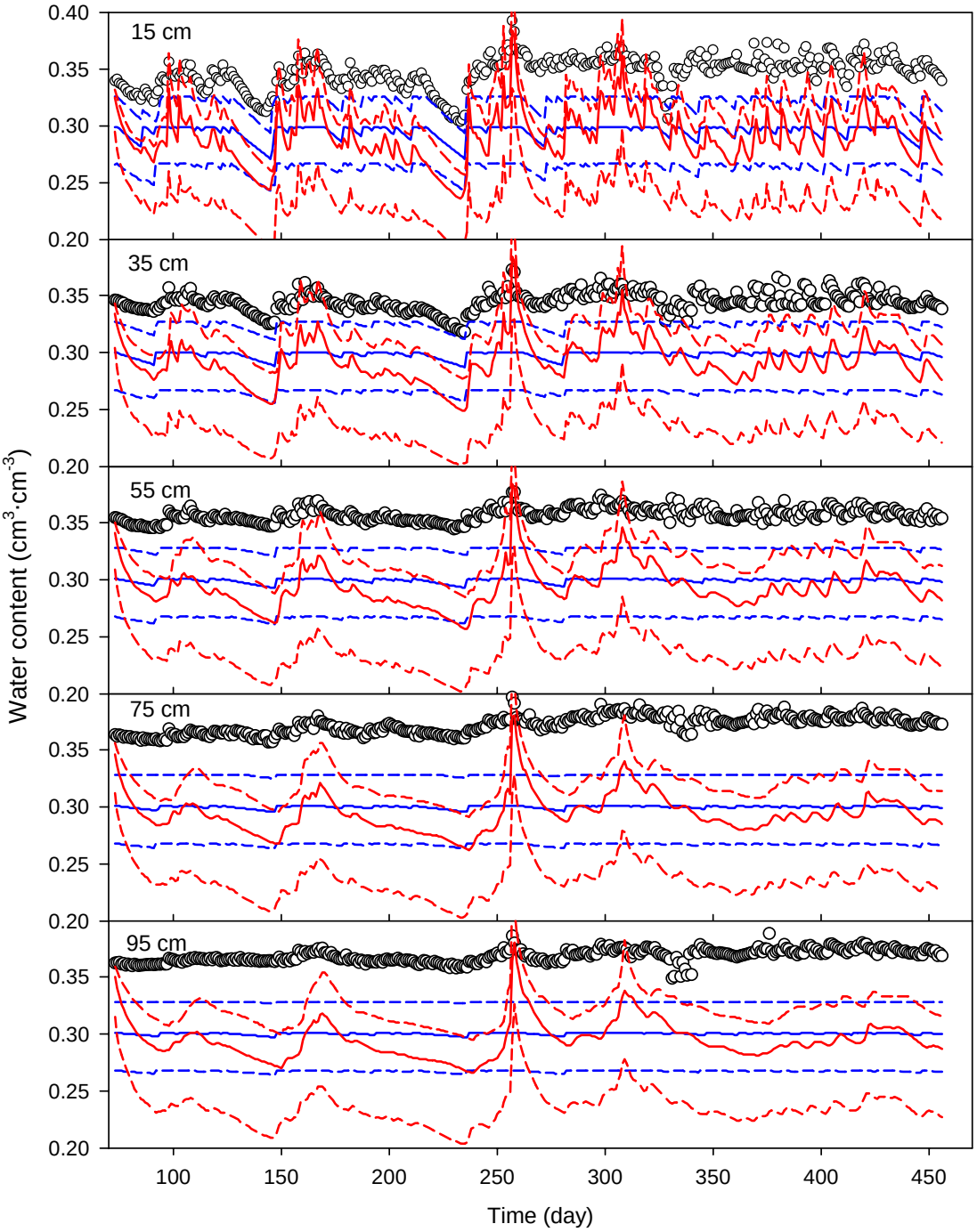

Figure 3-37. Time series of transect-average water contents: symbol – measured, red lines - simulated with the HYDRUS-1D, blue lines - simulated with the WMBUS model for homogeneous soil profile; water retention data and median hydraulic conductivity estimated with pedotransfer functions for topsoil; the solid line corresponds to the median value of simulated water contents for each day, dashed lines correspond to the simulated daily values at 25% and 75% probability levels.

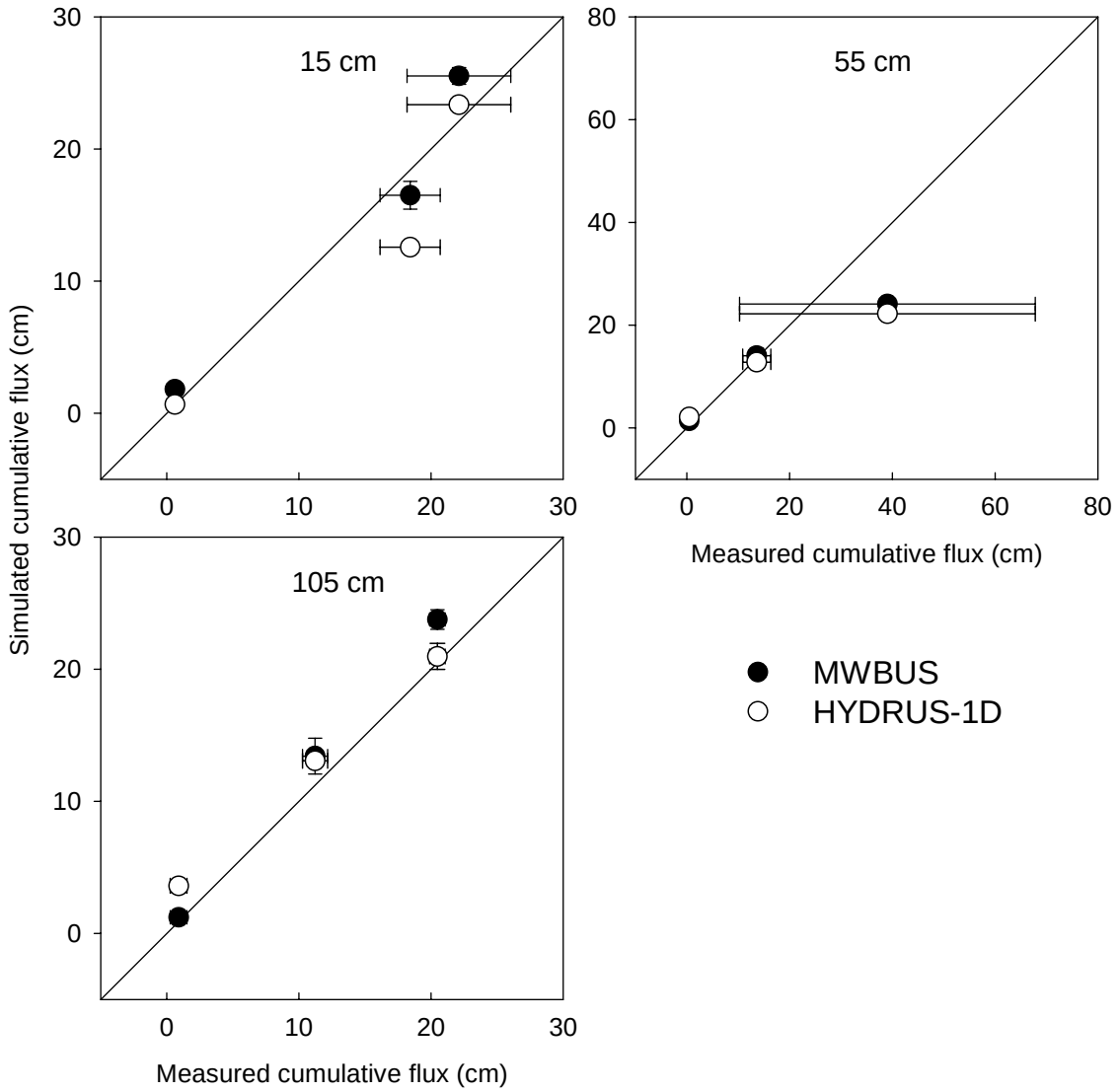

Figure 3-38. Soil water fluxes simulated with HYDRUS-1D and the MWBUS model; water retention data and median hydraulic conductivity estimated with pedotransfer functions for topsoil profile; error bars show one standard deviation.

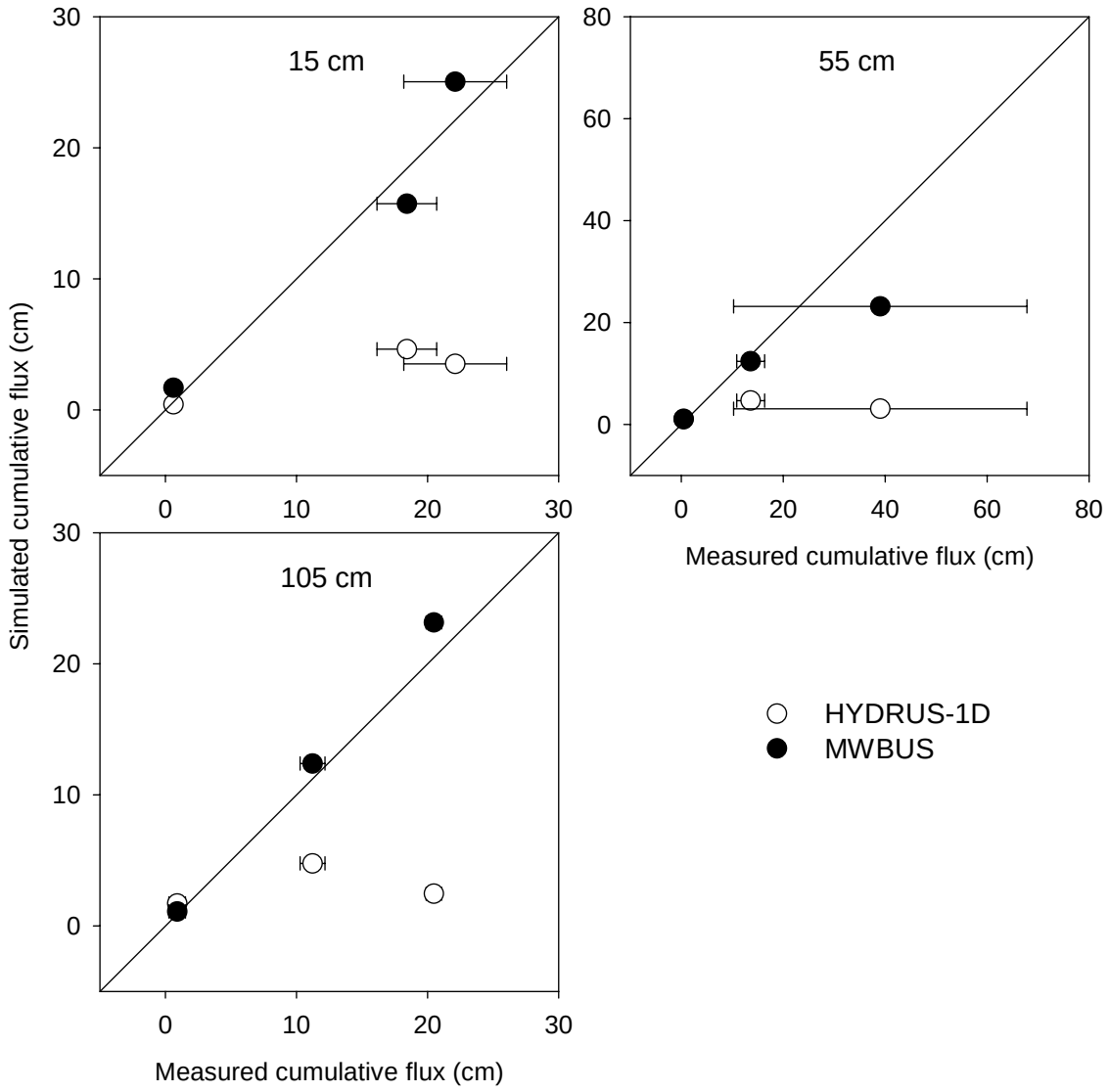

Figure 3-39. Soil water fluxes simulated with calibrated HYDRUS-1D and the MWBUS model for the homogeneous soil profile; error bars show one standard deviation.

50% probability according to soil texture and porosity (Table 2-4) and were not changed during the MWBUS optimization. Thus, the total number of estimated parameters was equal to 5.

Results of the calibration are shown in Fig. 3-32 and 3-33. Replacing the layered profile with homogeneous profile has more effect on errors in water contents than on errors in flux computations (see Fig. 3-29 and 3-30). The MWBUS model simulated water content qualitatively less adequately compared to the MWBUS. The values of water content simulated with the MWBUS did not changed in time at depths 55, 75 and 95 cm and were equal to field capacity. However, the MWBUS model simulated water fluxes more accurately compared to the HYDRUS-1D model when calibrated for homogeneous soil profile (Fig. 3-39).

Model 8 in Fig. 3-28 represents the sequential application of three abstraction steps: hierarchy of models, aggregation and pedotransfer functions. This ultimate simplification leads to further deterioration of the accuracy in water content estimates (Fig. 3-37). However, soil water fluxes are estimated reasonable well (Fig. 3-38).

3.5.6 Abstraction with metamodeling

This type of MA should be used when a large number of model runs may be required for the performance assessment (see Section 2.2.4). We developed an example to show the feasibility of applying an artificial neural network (ANN) to mimic the model behavior in predicting soil water fluxes with meteorological inputs defined at daily, weekly and ten-day time scales. The design of this work is shown in Fig. 3-40. It includes generating probable weather patterns, running the models in the Monte Carlo simulation mode, dividing the data into development and testing subsets, training the neural net to mimic the model output on the development data subset, and testing the net on the testing data subset.
We used the feed-forward ANN as described by Pachepsky et al. (1996). Climatic parameters of Albany, NY were used to generate 300 daily precipitation and evaporation sequences for the April-September period. The sequences were randomly divided into 150 sequences for training and 150 sequences for testing. Feedforward networks were developed using the MATLAB Neural Network toolbox. The minimum number of the hidden neurons was found experimentally to provide the best accuracy with testing dataset. ANN inputs were daily average precipitation and evaporation, or weekly average precipitation and evaporation. Accordingly, ANN outputs were daily soil water fluxes or weekly soil water fluxes.

Results are shown in Fig. 3-41 – 3-44. ANN related inputs and outputs of the models well. The determinations coefficients of regressions 'ANN monthly vs. simulated monthly' ranged from 0.899 to 0.944 when daily inputs and outputs were used with the MWBUS model (Fig. 3-41). There was a marked increase in accuracy of monthly fluxes predictions when weekly input and output were used; the determination coefficients were between 0.959 and 0.971 (Fig. 3-42). There was an opposite trend with the HYDRUS-1D output. The accuracy was higher with the daily input and outputs were used; values of R^2 were in the range between 0.990 and 0.994 (Fig. 3-43). With weekly inputs and outputs, the

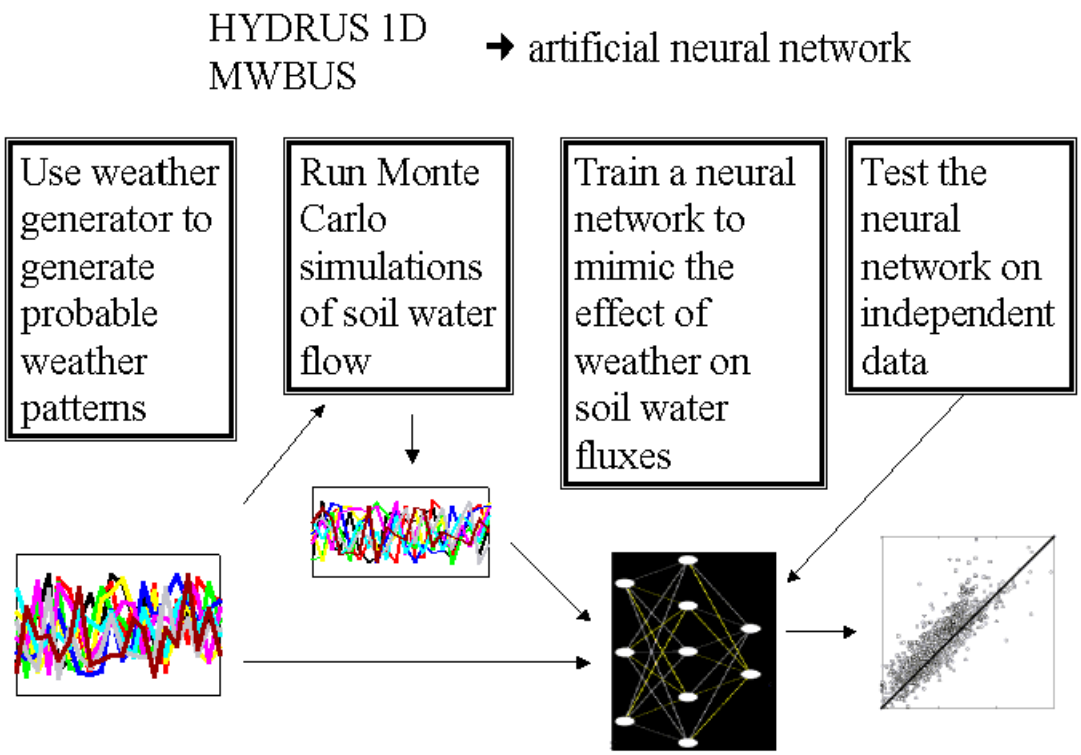

Figure 3-40. Schematics of the neural network application to abstract simulation models.

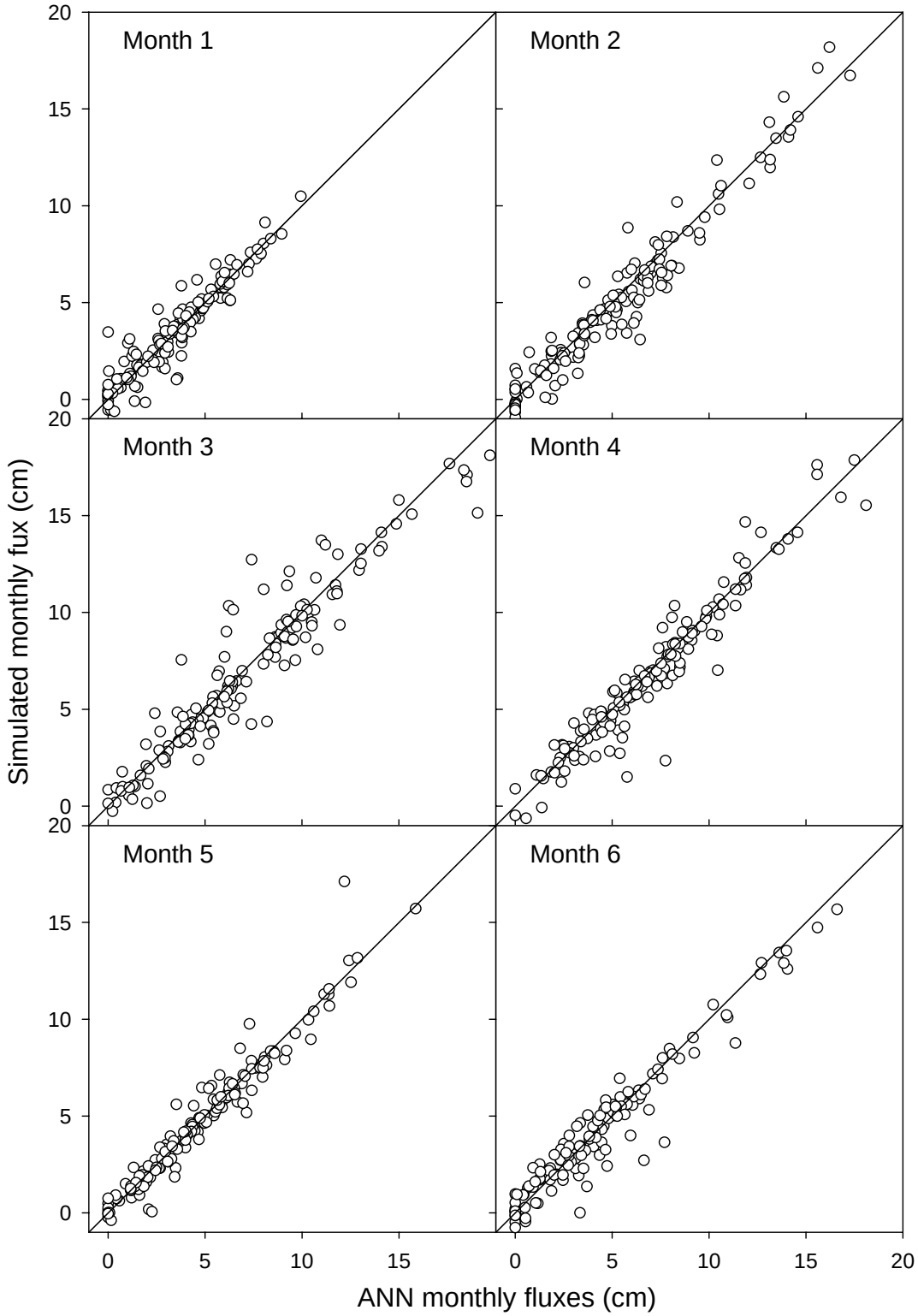

Figure 3-41.Estimated with the artificial neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – MWBUS, artificial neural network input – daily average flux, ANN output – daily average flux, the total number of hidden neurons – 20.

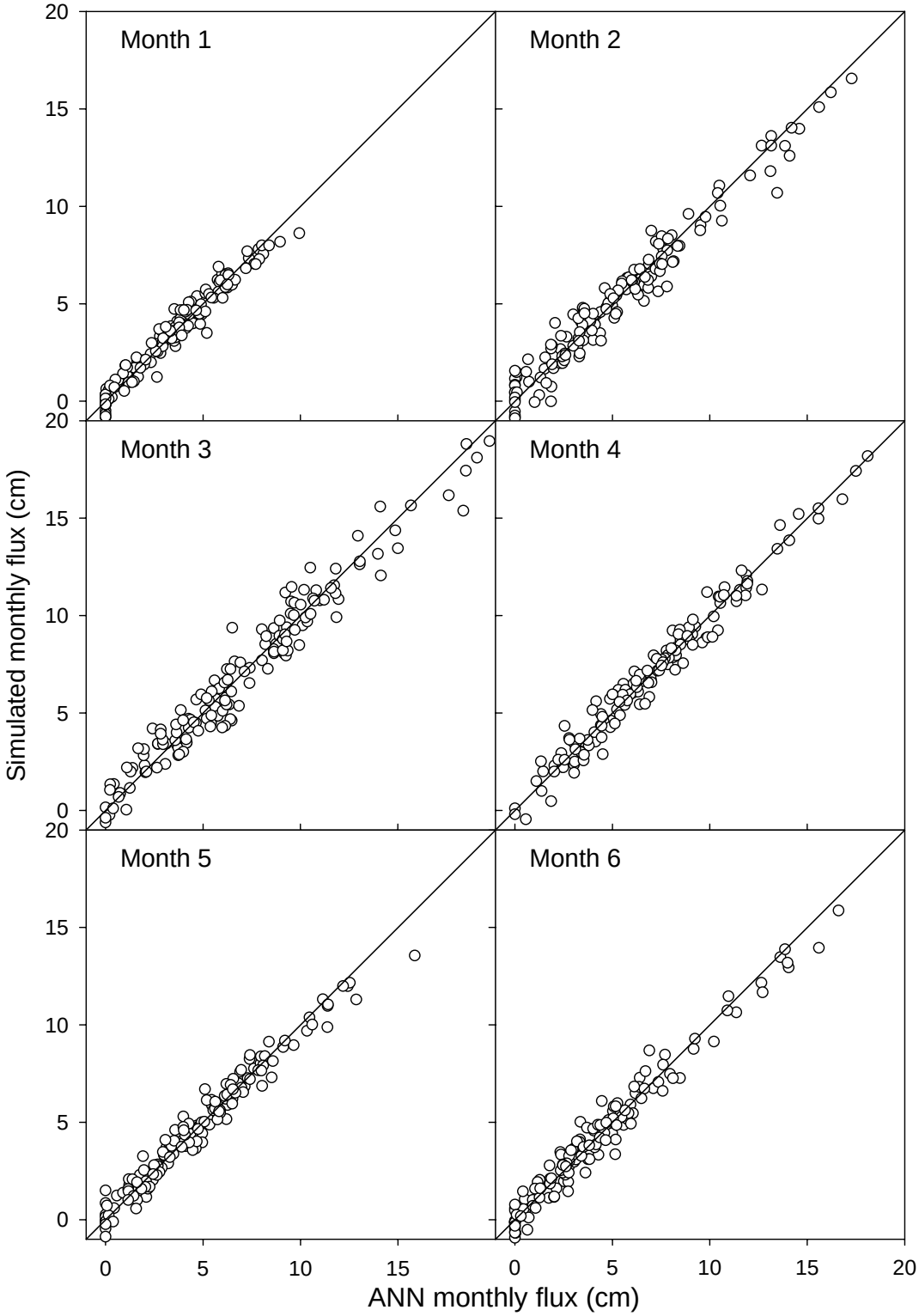

Figure 3-42.Estimated with the artificial neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – MWBUS, artificial neural network input – daily average flux, ANN output – daily average flux, the total number of hidden neurons – 12.

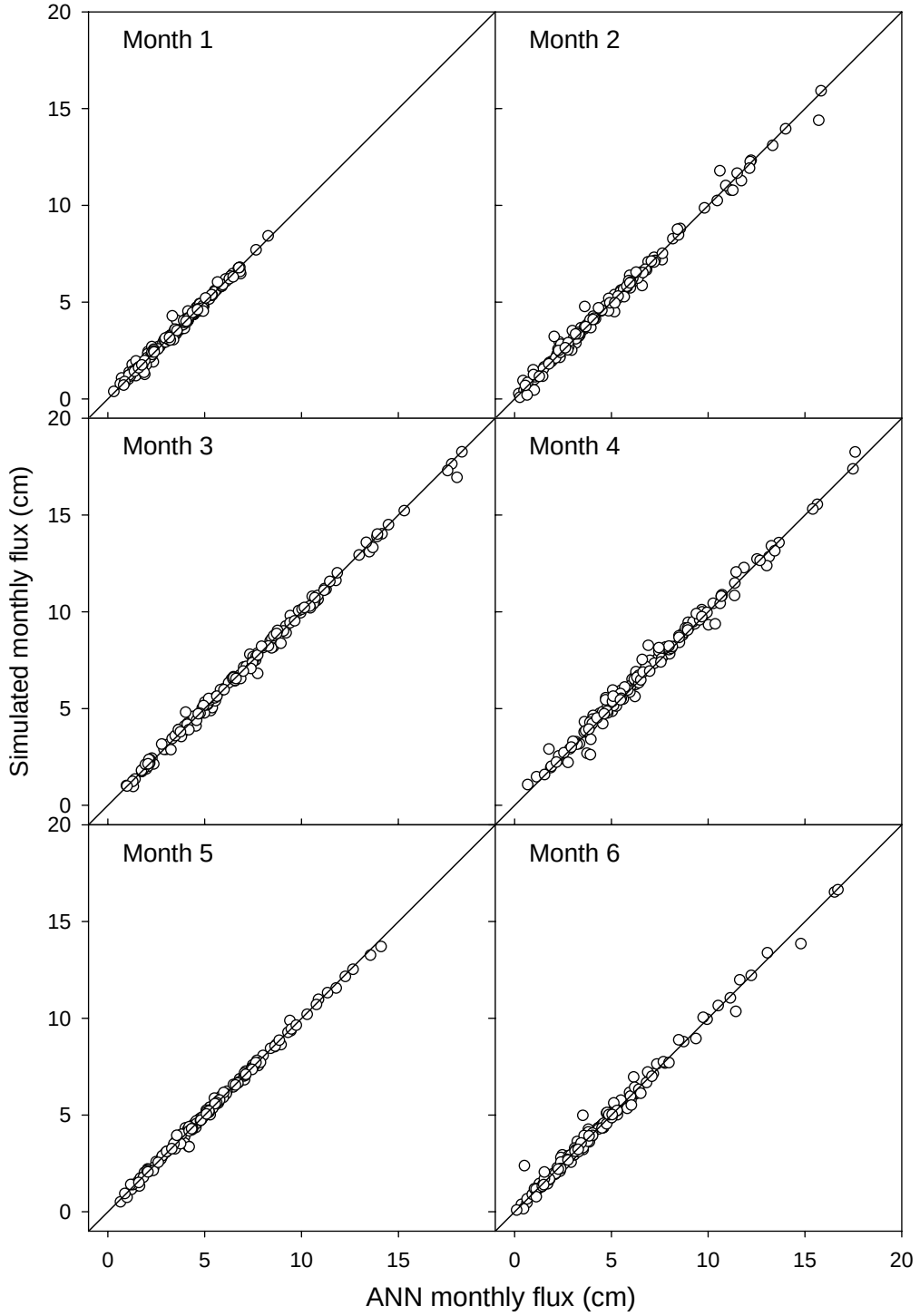

Figure 3-43. Estimated with the artificial neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – HYDRUS-1D, artificial neural network input – daily average flux, artificial neural network output – daily average flux, the total number of hidden neurons – 20.

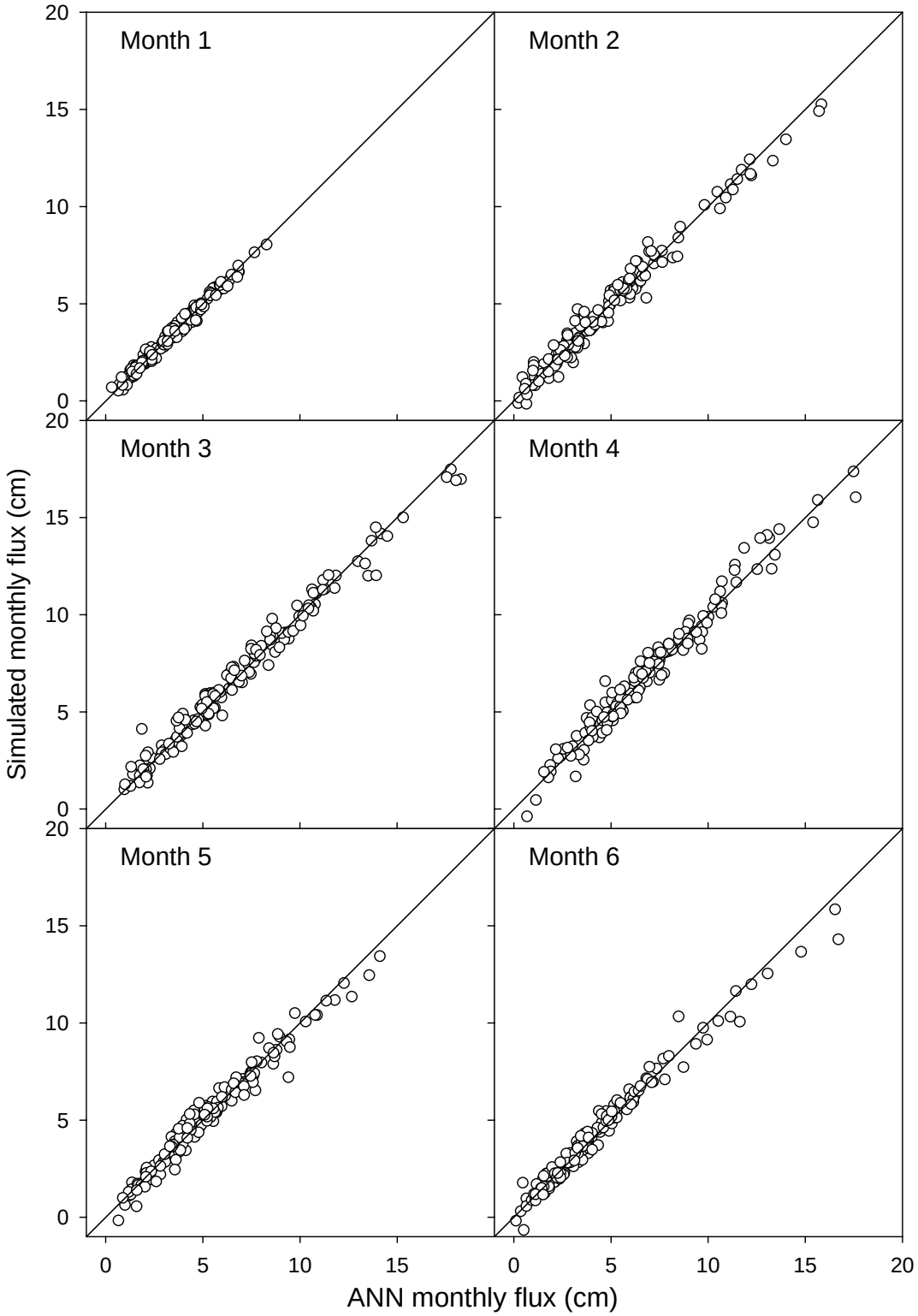

Figure 3-44. Estimated with the neural network and simulated monthly soil water fluxes at the depth of 105 cm. Model – HYDRUS-1D, ANN input – weekly average flux, ANN output – weekly average flux, number of hidden neurons – 12.

accuracy slightly decreased, and more scatter could be seen in the 1:1 diagrams of simulated versus ANN fluxes (Fig. 3-44). Values of values of R^2 were in the range between 0.976 and 0.985 in this case.

Overall, the feedforward neural networks demonstrated an excellent ability to mimic the simulations of soil water fluxes in bare soil under humid climatic conditions.

3.5.7 Pedotransfer-estimated vs. laboratory-measured water retention

Application of pedotransfer functions is often suggested only when measurements of hydraulic properties are not available or are not feasible. Field measurements of soil water retention were available in this case study. Additional simulations were made to estimate the loss of accuracy caused by supplanting PTF-estimated water retention instead of the measured one.

The Monte Carlo simulation was performed for the HYDRUS-1D model with van Genuchten parameters shown in Fig. 3-12 and parameters estimated with PTF (Table 3-8). No correlations between parameters shown in the Table 3-5 were used at this time. The empirical distribution functions were sampled randomly 50 times. The median values of lognormal distributions of K_{sat} (Table 2-5) were used from loam and silt loam texture classes at depths ranges 0 to 65 cm and 65 to 105 cm, respectively.

The comparison of the statistical distributions of the root-mean-square errors obtained in simulations with laboratory data and with pedotransfer functions showed that the accuracy of simulations with PTF was better than with laboratory data (Fig. 3-45). In terms of uncertainty no difference between simulation with laboratory and PTF were obtained, and slight less uncertainty or more narrow range of RMSE for water content simulation was obtained in case of layered soil profile.

The Monte-Carlo simulations with the MWBUS model were performed using water contents at capillary pressure of 100 cm as the field capacity values. The probability distributions of the MWBUS model root-mean-square errors from simulations with laboratory data and with PTF functions showed that the accuracy of simulations with laboratory data was better than with PTF (Fig. 3-45). Water content simulation results were also less uncertain with the laboratory measured water retention than with use PTF.

The probability distributions were similar for layered and homogeneous soil profile with PTF estimated water retention. For laboratory measured water retention more accurate but less uncertain results were obtained for the homogeneous than for layered soil profile.

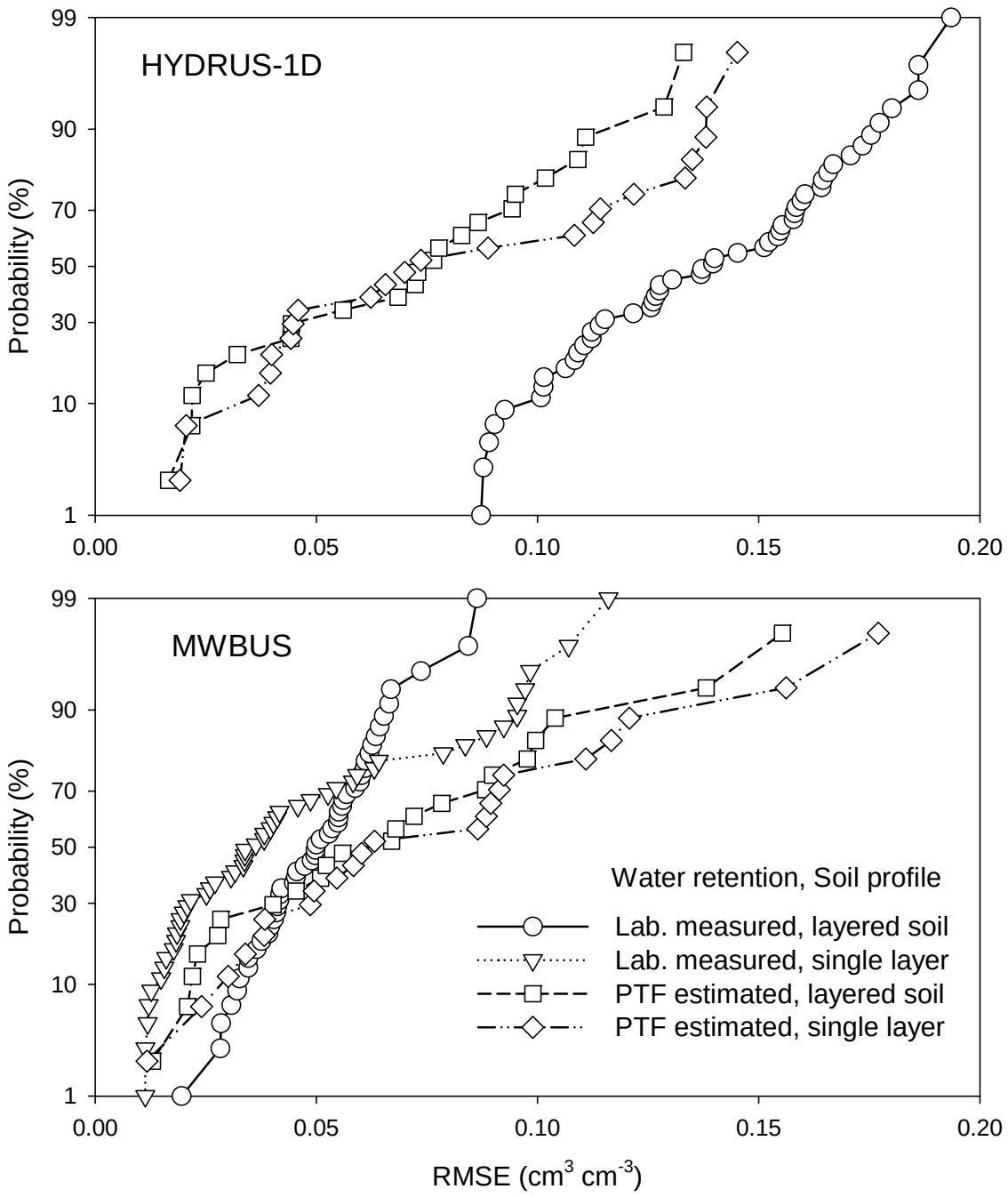

Figure 3-45. Probability distribution of the root-mean square error (RMSE) of water content from Monte Carlo simulations for the HYDRUS-1D and the MWBUS models with two sources of soil hydraulic properties and two representations of soil profile.

3.6 Summary and conclusion

The case study elucidated the model abstraction efficiency in modeling water flow in variably saturated subsurface. The intensive monitoring per se did not warrant the adequate performance of the base model which was the Richards single continuum model. The need in model abstraction was justified by the fact that the calibrated base model predicted substantial runoff whereas no runoff was observed at the site during the monitoring period. The base model produced inexplicable results in terms of the key output.

The review of modeling context helped to improve knowledge about boundary fluxes that have not been measured. In particular, the review of the modeling context revealed that (a) soil water contents across the trench demonstrated strong temporal persistence, and (b) a substantial preferential flow should occur in soil. The persistence was used to estimate soil bottom flux from water content measurements, and the preferential flow estimates were used to assess evaporation from soil surface.

Four classes of applicable model abstraction techniques were selected. Model structure abstraction was achieved using the hierarchy of model, aggregation, and metamodeling. The hierarchy of models included the Richards single continuum medium and a simple soil water budget model. Aggregation was done by replacing the layered soil profile with the homogeneous profile. Metamodeling was done with the artificial neural network that used daily and weekly precipitation as input to estimate monthly cumulative soil water fluxes. An ensemble of 22 pedotransfer functions was used in model abstraction of parameter determination. HYDRUS1-D, MWBUS, and MATLAB software packages were used to calibrate the abstracted models and to run Monte Carlo simulations to evaluate the uncertainty in calibrated model outputs.

The abstraction with the hierarchy of models was useful. The simple soil water budget model was less accurate in predictions of soil water content compared with the Richards model. However, unlike the more complex Richards model, this simple soil water budget model correctly predicted the absence of runoff and measured cumulative soil water fluxes. The prediction of runoff was an artifact of the Richards model calibration in absence of measured boundary fluxes. This abstracted model appeared to be instrumental in both explaining behavior of the complex model and in predicting the key output, i.e. soil water fluxes.

The abstraction with aggregation was not useful in this case. The Richards model was less accurate with respect to soil water contents and continued to generate large simulated runoff when a homogeneous soil layer was introduced.

The abstraction of parameter determination with pedotransfer functions was useful. It showed that the Richards model with parameters in correct ranges is able to correctly simulate soil water fluxes. The ensemble of pedotransfer functions represented field water retention better than data from laboratory soil water retention measurements. Soil water content predictions with the Richards model were more accurate with the ensemble of pedotransfer functions than with laboratory data.

The neural network metamodel was extremely accurate in estimating cumulative soil water fluxes. It was also five orders of magnitude faster than the numerical algorithm coded in HYDRUS-1D. However, the HYDRUS-1D runtime was adequate for the Monte Carlo simulations to evaluate model abstraction in the uncertainty context.

In general, the model abstraction process was successful in this case study. Model abstraction explained the strange behavior of the complex model, and provided the correct description of the system behavior and plausible parameter ranges.

4 REFERENCES

Abbaspour, K.C., van Genuchten, M.T., Schulin, R., and E. Schläppi, "A Sequential Uncertainty Domain Inverse Procedure for Estimating Subsurface Flow and Transport Parameters," *Water Resources Research*, 33(8): 1879–1892, 1997.

Addanki, S., Cremonini, R. and J. Penberthy, "Graphs of Models," *Artificial Intelligence*, 51: 145-177, 1991.

Ahuja, L. R., J. W. Naney, P. E. Green, and D.R. Nielsen, "Macroporosity to Characterize Spatial Variability of Hydraulic Conductivity and Effects of Land Management," *Soil Science Society of America Journal*, 48: 404-411, 1984.

Ahuja, L.R., D.K. Cassel, R.R. Bruce, and B.B. Barnes, "Evaluation of Spatial Distribution of Hydraulic Properties Using Effective Porosity Data," *Soil Science,* 148: 404-411, 1989.

Alexander, L., and R. W. Skaggs, "Predicting Unsaturated Hydraulic Conductivity From Soil Texture," *Journal of Irrigation and Drainage Engineering*, 113: 184-197, 1987.

Altman, S. J., Arnold, B. W., Barnard, R. W., Barr, G. E., Ho, C. K., McKenna, S. A., and R. R. Eaton, "Flow Calculation for Yucca Mountain Groundwater Travel Time," (*GWTT-95*). *SAND96-0819*, Sandia National Laboratories, Albuquerque, NM, 212 pp., 1996.

Amemiya, M., "The Influence of Aggregate Size on Soil Moisture Content Capillary Conductivity Relations," *Soil Science Society of America Proceedings*, 29: 744-748, 1965.

Baumer, O. M. "Predicting Unsaturated Hydraulic Parameters," pp.341-354. in M.Th. van Genuchten, F. J. Leij, and L J. Lund (ed*.) Proc. Int. Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils*. University of California, Riverside, CA, 1992.

Benson, D.A., Wheatcraft S.W., and M.M. Meerschaert, "Application of a Fractional Advection-Dispersion Equation," *Water Resources Research*, 36: 1403-1412, 2000.

Berkowitz, B., Kosakowski, G., Margolin, G., and H. Scher, "Application of Continuous Time Random Walk Theory to Tracer Test Measurements in Fractured And Heterogeneous Porous Media," *Ground Water*, 39: 593-604, 2001.

Beven, K. "Towards a Coherent Philosophy for Modeling the Environment," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 458(2026): 2465-2484, 2002

Bigelow, J. H,. and Davis, P. K.,"*Implications for model validation of multiresolution, multiperspective modeling (MRMPM) and exploratory analysis.* MR-1750," RAND, 2138, Santa Monica, CA, 2003

Bird, N. R. A, and E. M. A. Perrier, "The Water Retention Curve for a Model of Soil Structure With Pore And Solid Fractal Distribution," *European Journal of Soil Science*," 55: 55-67, 2000.

Bloemen, G. W., "Calculation of Hydraulic Conductivities of Soils From Texture and Organic Matter Content," *Z. Pflanzenernahr. Bodenkd.*, 143: 581-605, 1980.

Blöschl, G., and Sivapalan, M., "Scale issues in hydrological modeling", Hydrological processes, 9: 251-290, 1995.

Boyle, D. P., Gupta, H. V., Soroshian, J., Koren, V., Zhang, Z., M. Smith, "Toward Improved Streamflow Forecasts: Value of Semidistributed Modeling," *Water Resources Research*, 37: 2749-2759, 2001.

Brooks, R.H., and A.T.Corey, "Hydraulic Properties of Porous Media," *Hydrology Pap. 3*, Colorado State University, Ft. Collins, 27 pp.,1964.

Bruand, A., Baize, D., and Hardy, M. 1994. Predicting of water retention properties of clayey soil using a single soil characteristic. Soil Use and Management, 10: 99-103.

Burton, R., "Metamodeling: A State of The Art Review," pp. 237-244 in: *Proceedings of the 1994 Winter Simulation Conference* (J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila - Eds). Institute of Electrical and Electronics Engineering, San Francisco, California. 1994.

Campbell, G.S., and S. Shiozawa, "Prediction of Hydraulic Properties of Soils Using Particle Size Distribution and Bulk Density Data," pp.317-328 in M. Th. van Genuchten, F. J. Leij, and L J. Lund (ed.) *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. University of California*, Riverside, CA, 1992.

Canarache, A., "Physical-Technological Maps - A Possible Product of Soil Survey for Direct Use in Agriculture," *Soil Technology*, 6: 3-16, 1993.

Cassel, D. K., Wendroth, O., and D.R. Nielsen, "Assessing Spatial Variability in an Agricultural Experiment Station Field: Opportunities Arising From Spatial Dependence," *Agronomy Journal*, 92: 706-714, 2000.

Castleton, K., Fine, S., Banta, N., Hill, M., Markstrom, S., Leavesley, G., and Babendreier, J., "An Overview of the Uncertainty Analysis, Sensitivity Analysis, and Parameter Estimation (UA/SA/PE) API and How To Implement It", Pp. 171-172. In: Nicholson, T. J., Babendreier, J. E., Meyer, P. D., Mohanty, S., Hicks, B. B., Leavesley, G. H. (Eds.) Proceedings of the International Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling August 19–21, 2003. NUREG/CP-0187, U.S. Nuclear Regulatory Commission, Washington, DC, 2002.

Caughlin, D. and A. F. Sisti, "Summary of Model Abstraction Techniques," *Proc. SPIE Int. Soc. Opt. Eng*. 3083: 2-13, 1997.

Christakos, G. *Modern Spatiotemporal Geostatistics,* Oxford University Press, Oxford, New York, 212 pp., 2000.

Clancy, D., and B. Kuipers, "Model Decomposition and Simulation," *Proceedings of the Eighth International Workshop on Qualitative Reasoning about Physical Systems*, Nara, Japan, pp. 45-54, 1994.

Clark, L. A. and D. Pregibon, "Tree-based Models," pp. 377-419 in S.J.M. Chambers and T. J. Hastie (eds.) *Statistical Models*, Wadsworth, 1991.

Clifton P. M., and S. P. Neuman, "Effects of Kriging and Inverse Modeling on Conditional Simulation of The Avra Valley Aquifer in Southern Arizona," *Water Resource Research*, 18(4): 1215-1234, 1982.

Comegna, V. and A. Basile, "Temporal Persistence of Spatial Patterns of Soil Water Storage in a Cultivated Vesubian Soil," *Geoderma* 62, 299-310, 1994.

Cushman, J.H., "An Introduction to Fluids in Hierarchical Porous Media," pp. 1-6 in J.H. Cushman, ed., *Dynamics of Fluids in Hierarchical Porous Media*, Academic, NY, 1990

Davis, J. A., Yabusaki, S. B., Steefel, C. I., Zachara, J. M., Curtis, G. P., Redden, G. D., Criscenti, L. J. and B. D.Honeyman, "Assessing Conceptual Models for Subsurface Reactive Transport of Inorganic Contaminants," *EOS*, 85(44): 449-455, 2004.

Davis, P. K., and Bigelow, J. H,."*Motivated metamodels: synthesis of cause-effect reasoning and statistical metamodeling*. MR-1570," RAND, 2138, Santa Monica, CA, 2003.

Destouni, G. "Applicability of The Steady State Flow Assumption for Solute Advection in Field Soils," *Water Resources Research*, 27: 2129-2140, 1991.

Diwekar, U. M. "How Simple Can It Be? - A Look at The Models for Batch Distillation," *Computers & Chemical Engineering* 18: S451—S457, 1994.

Englert, A., Hardelauf, H., Vanderborght, J., and H. Vereecken, "Numerical Modelling of Flow and Transport on Massively Parallel Computers," *NIC Symposium 2004, Proceedings*, D. Wolf, G. M¨unster, and M. Kremer (Editors), John von Neumann Institute for Computing, Jülich, *NIC Series*, Vol. 20, ISBN 3-00-012372-5, pp. 409-418, 2004.

European Soil Bureau, 1998. Georeferenced soil database for Europe. European Soil Bureau and Joint Research Centre, EC. Ispra, Italy.

Ewing, R.E., Helmig, R., Finsterle, S., and R. Hinkelmann, "Modeling of Flow and Transport Processes in the Subsurface," pp. 477-480 in *Groundwater Updates*, K. Soto and Y. Iwasa, eds., Tokyo, Springer-Verlag, 2000.

Falkenhainer, B. and K. Forbus, K., "Compositional modeling: finding the right model for the job, " *Artificial Intelligence*, 51 :95-143, 1991

FAO, 1975. *Soil Map of the World at 1:5,000,000*. Volume I. Europe. *UNESCO*, Paris, France, 62 p.

Fayer, M.J. and C.S. Simmons, "Modified Soil Water Retention Functions for all Matric Suctions," *Water Resources Research*, 31: 1233-1238, 1995.

Ferreyra, R. A., Apeztegua, H. P., Sereno, R., and J. W. Jones, "Reduction of Soil Water Spatial Sampling Density Using Scaled Semivariograms and Simulated Annealing," *Geoderma*, 110: 265-289, 2002.

Field, M. S., and P. F. Pinsky. "A two-region nonequilibrium model for solute transport in solution conduits in karstic aquifers", Journal of Contaminant Hydrology, 44: 329-351, 2000

Fine, S., Catleton, K, Banta, N., Markstrom, S., Leavesley, G., Babendreier, J., "Calibration, Optimization, and Sensitivity and Uncertainty Algorithms Application Programming Interface (COSU-API), Pp. A1-A54. In: Nicholson, T. J., Babendreier, J. E., Meyer, P. D., Mohanty, S., Hicks, B. B., Leavesley, G. H. (Eds.) Proceedings of the International Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling August 19–21, 2003. NUREG/CP-0187, U.S. Nuclear Regulatory Commission, Washington, DC, 2002.

Fishwick, P. A. "The Role of Process Abstraction in Simulation*," IEEE Transactions on Systems, Man and Cybernetics*, 18: 18–39, 1988.

Fishwick, P. A. and P. A. Luker, (Eds.), "Qualitative Simulation Modeling and Analysis," Springer Verlag, 1991, 330 pp., 1991.

Fishwick, P. A., "*Simulation Model Design and Execution"*. Englewood Cliffs: Prentice-Hall, 1995

Flury, M. and H. Flühler, "Modeling Solute Leaching in Soils by Diffusion-Limited Aggregation: Basic Concepts and Application to Conservative Solutes," *Water Resources Research*, 31: 2443-2452, 1995.

Frantz, F. K. "A Taxonomy of Model Abstraction Techniques," 2002. http://www.rl.af.mil/tech/ papers/ModSim/ModAb.html

Frantz, K. F., "A taxonomy of model abstraction techniques," In: Alexopoulos, C., Kang, K., Lilegdon, W. R., and Goldsman, D. (Eds.) Proceedings of the 1995 Winter Simulation Conference, pp.1413-142, 1995.

Franzmeier, D.P. "Estimation of Hydraulic Conductivity from Effective Porosity Data for Some Indiana Soils," *Soil Science Society of America Journal,* 55: 1801-1803, 1991.

Gerke, H.H., and M.Th. van Genuchten, "A Dual-Porosity Model for Simulating the Preferential Movement of Water and Solutes in Structured Porous Media," *Water Resources Research*, 29: 305-319, 1993.

Good, P. I., 1999. Resampling methods: a practical guide to data analysis. Birkhäuser, Boston.

Goovaerts, P. and C.N., Chiang, "Temporal Persistence of Spatial Patterns for Mineralizable Nitrogen and Selected Soil Properties," *Soil Science Society of America Journal,* 57, 372-381, 1993.

Govindaraju, R.S. (ed.) " *Stochastic methods in subsurface contaminant*

Grayson, R. B., and A.W. Western, "Towards Areal Estimation of Soil Water Content from Point Measurements: Time and Space Persistence of Mean Response," *Journal of Hydrology,* 207: 68-82, 1998.

Griffioen, J.W., Barry, D.A., and J.Y. Parlange, "Interpretation of Two-Region Model Parameters,"*Water Resources Reserach,* 34, 373-384, 1998.

Guadagnini, A., and Neuman, S.P., "Nonlocal and localized analyses of conditional mean steady state flow in bounded randomly nonuniform domains 1. Theory and computational approach,"Water Resour. Res. 35:2999–3018, 1999.

Guber, A. K., Rawls, W. J, Shein, E. V., and Y. A. Pachepsky, "Effect of Soil Aggregate Size Distribution on Water Retention," *Soil Science,* 168: 223 – 233, 2003.

Guber, A. K., Shein, E.V., Van Itsyuan and A.B.Umarova, "Experimental Information for Mathematical Models of Water Transport in Soil: Assessment of Model Adequacy and Reliability of Prediction," *Eurasian Soil Science*, 31(9): 1020-1029, 1998.

Gupta, S. C., and W. E. Larson, "Estimating Soil Water Retention Characteristics from Particle-Size Distribution, Organic Matter Percent, and Bulk Density," *Water Resources Research*, 15(6): 1633-1635, 1979.

Guswa, A. J. and D. L. Freuberg, "On Using the Equivalent Conductivity to Characterize Solute Spreading in Environments with Low-Permeability Lenses," W*ater Resources Research*, 38(8): 1132 (DOI 10.1029/2001WR000528), 2002.

Guswa, A. J., Celia, M .A., and I. Ridriguez-Iturbe, "Models of Soil Moisture In Ecohydrology: A Comparative Study," *Water Resources Research*, 38(9): 1166 (DOI 10.1029/2001WR000826), 2002.

Haise, H.R., Haas, H.J., and L.R. Jensen, "Soil Moisture of Some Great Plains Soils. II. Field Capacity as Related To 1/3-Atmosphere Percentage and "Minimum Point" as Related to 15- and 26-Atmosphere Percentages," *Soil Science Society of America Pro*ceedings, 19: 20-25, 1955.

Hall, D. G. M., Reeve, M. J., Thomasson, A. I., and V. F. Wright, "Water Retention, Porosity And Density of Field Soils," *Soil Surv. Tech. Monogr*. 9, 74 pp., 1977.

Hamill, T. M., Hansen, J. A., Mullen, S. L., Snyder, C. Workshop on Ensemble Forecasting in the Short to Medium Range. Meeting Summary. *Bulletin of the American Meteorological Society,* 86, (in press), 2005.

Hassan, A. and K. Hamed, "Prediction of Plume Migration in Heterogeneous Media Using Artificial Neural Networks," *Water Resources Research*, 37(3): 605-624

Hecht-Nielsen, R. "Neurocomputing," Addison-Wesley Publishing Company, Reading, MA, pp. 433 pp., 1990.

Heimovaara T.J. and W. Bouten, "A Computer Controlled 36-Channel Time Domain Reflectometry System for Monitoring Soil Water Contents," Water Resources Research, 26: 2311-2316, 1990.

Hill, C. M. "Methods and Guidelines for Effective Model Calibration," U. S. Geological Survey Water Resources Investigations Report 98-4005, Denver, CO, 90 pp., 1998.

Hill, M. C., Middlemis, H., Hulme, P., Poeter, E., Riegger, J., Neuman, S. P., WIlliams, H., and M. Anderson, "Brief Overview of Selected Groundwater Modelling Guidelines," pp. 105-120, in: K. Kovar, J. Bruthans and Z. Hrkal (eds.), *Finite-Element Models, MODFLOW, and More, Solving groundwater problems*, Proceedings, Carlsbad, Czech Republic, September 13-16, 2004.

Hillestad, R. J. and M. L. Juncosa, "Cutting Some Trees to Save the Forest: On Aggregation and Disaggregation in Combat Models", MR-189, RAND, Santa Monica, Calif., 1993.

Hillestad, R., Owen, J. G. and Blumenthal D. Experiments in Variable Resolution Modeling, N-3631-DARPA, RAND, Santa Monica, Calif., 1992.

Holder, M., Brown, K. W., Thomas, J. C., Zabcik, D. and H. E. Murray, "Capillary-Wick Unsaturated Zone Soil Pore Water Samplers," *Soil Science Society of America Journal*, 55: 1195-1202, 1991.

Hubbard, S., J. Chen, J. Peterson, K. Williams, D. Watson, P. Jardine, and M. Fienen. "Characterization & Monitoring at FRC Area 3 using Geophysical Data Conditioned to Hydrogeological Data" at public.ornl.gov/nabirfrc/update_aug03.pdf , 2003

Hunt, R. and C. Zheng, "Debating Complexity in Modeling," *EOS, Transactions*, American Geophysical Union, 80(3): 29, January 19, 1999

Hupet, F. and M. Vanclooster, "Intraseasonal Dynamics of Soil Moisture Variability Within a Small Agricultural Maize Cropped Field," *Journal of Hydrology,* 261, 86–101, 2002.

Innis, G., and E. Rextad, "Simulation model simplification techniques," *Simulation*, 41: 7-15, 1983.

Iwasaki, Y., "Two Model Abstraction Techniques Based on Temporal Grain Size: Aggregation and Mixed Models," *Knowledge Systems Laboratory Report*, Stanford University, 12 pp., 1990.

Iwasaki, Y., and H. Simon, "Causality and Model Abstraction," *Artificial Intelligence,* 67: 143–194, 1992

Jacques D. "Analysis of Water Flow and Solute Transport at The Field Scale," PhD, no 454, Faculteit Land bouwkundige en Toegepaste Biologische Wetenschappen, K.U. Leuven, Belgium, 255 pp., 2000.

Jacques, D., Simunek, J., Timmerman, A., and J. Feyen, "Calibration of Richards and Convection–Dispersion Equations To Field-Scale Water Flow And Solute Transport Under Rainfall Conditions," *Journal of Hydrology*, 259: 15-31, 2002.

Jacques, D., Timmerman, A. and J. Feyen, "Experimental Study of Water Flow and Solute Transport in a Macroporous Soil Under Natural Boundary Conditions: Bekkevoort-Site Description, Experimental Design and Calibration Procedures," Internal Report, no. 53, Institute for Land and Water Management, K.U.Leuven, Belgium, 99 p., 1999.

Jakeman, A.J. and G.M. Hornberger, "How Much Complexity is Warranted in a Rainfall-Runoff Model?" *Water Resources Research*, 29(8): 2637-2649, 1993.

Jarvis, N.J. "Modeling the Impact of Preferential Flow on Nonpoint Source Pollution," pp. 195-221 in: H.M. Selim and L. Ma (eds*.), Physical Nonequilibrium in Soils: Modeling and Applications*, Ann Arbor Press, Chelsea, MI. 1999.

Jothityangkoon, C., Sivapalan, M., and D.L. Farmer, "Process Controls of Water Balance Variability in a Large Semi-Arid Catchment: Downward Approach To Hydrological Modeling," *Journal of Hydrology* 254: 174-198, 2001.

Jury, W.A. "Transfer Function Model of Field-Scale Solute Transport Under Transient Water Flow," S*oil Science Society of America Journal*, 54: 327-332, 1990.

Kachanoski, R.G. and E. De Jong, "Scale Dependence and the Temporal Persistence of Spatial Patterns Of Soil Water Storage," *Water Resources Research,* 24: 85-91, 1988.

Karpouzos, D. K., Delay, F., Katsifarakis, K. L., and G. de Marsily, "A Multipopulation Genetic Algorithm to Solve the Inverse Problem in Hydrogeology. *Water Resource Research*, 37: 2291-2302: 2001.

Kekkonen, T. S., and A. J. Jakeman, "A Comparison of Metric and Conceptual Approaches in Rainfall Runoff Modeling and Its Implications," *Water Resources Research*, 34: 2345-2352, 2001.

Kelson, V.A., Hunt, R.J., and H.M. Haitjema, "Improving a Regional Model Using Reduced Complexity and Parameter Estimation," *Ground Water*, 40(2): 132-143, 2002.

Kern, J. S. "Evaluation of Soil Water Retention Models Based on Basic Soil Physical Properties*," Soil Science Society of America Journal*, 59: 1134-1141, 1995.

Khaleel., R., Yeh, T. C. and Z. Lu, "Upscaled Flow and Transport Properties for Heterogeneous Unsaturated Media," *Water Resources Research*, 38(5): 1053 (DOI 10.1029/2000WR000072), 2002.

Knutson, J. H., Lee, S. B., Zhang, W. Q. and J. S Selker, "Fiberglass Wick Preparation for Use on Passive Capillary Wick Soil Pore-Water Samplers," *Soil Science Society of America Journal*, 57: 1474-1476, 1993.

Koch, S., and H. Flühler. "Solute transport in aggregated porous media: Comparing model independent and dependent parameter estimation", Water, Air, & Soil Pollution, 68: 275-289, 1993.

Kosugi, K.,"Three-Parameter Lognormal Distribution Model for Soil Water Retention," Water Resource Research, 30: 891-901, 1994.

Kron T. D. and D. Rosberg. "An Artificial Neural Network Based Groundwater Flow and Transport Simulator," http://www.isva.dtu.dk/staff/tdk/nn.2.pdf ,1998.

Kumar, P., "Layer Averaged Richard's Equation with Lateral Flow", *Advances in Water Resources* 27: 521 –531, 2004

Leenhardt, D., Voltz, M. and S. Rambal, "A Survey of Several Agroclimatic Soil Water Balance Models with Reference to Their Spatial Application," *Eur. J. Agron*., 4(1): 1-14, 1995.

Leij, F., W. J. Alves, M. Th. van Genuchten, and J. R. Williams, "The UNSODA Unsaturated Soil Hydraulic Database. User's Manual Version 1.0" *EPA/600/R-96/095*, National Risk Management Laboratory, Office of Research and Development, Cincinnati, OH, 103 pp., 1996.

Li, S. G., McLaughlin D., and H. S. Liao, H. S., "A computationally practical method for stochastic groundwater modelIng", *Adv. Water Resour. 26*(11), 1137-1148, 2003.

Loague, K., "Using Soil Texture to Estimate Saturated Hydraulic Conductivity and the Impact on Raifall-Runoff Simulations," *Water Resources Bulletin*, 28: 687-693, 1992.

Mallants, D., Mohanty, B. Jacques, D. and J. Feyen, "Spatial Variability of Hydraulic Properties in a Multi-Layered Soil," *Soil Science*, 161:167-181, 1996.

Mallat, S. G. "A Wavelet Tour of Signal Processing," Academic Press, San Diego, 673 pp., 1998.

Marion, J. M., Or, D. Rolston, D. E., Kavvas, M. I and J. W. Biggar, "Evaluation of Methods for Determining Soil-Water Retentivity and Unsaturated Hydraulic Conductivity," Soil Science, 158: 1-13, 1994.

Martinez-Fernandez, J. and A. Ceballos, "Temporal Stability of Soil Moisture in a Large-Field Experiment in Spain," *Soil Science Society of America Journal*, 67: 1647 – 1656, 2003.

Masket, S, Dibike, Y.B, Jonoski, A and D.P. Solomatine, "Groundwater Model Approximation With ANN for Selecting Optimum Pumping Strategy for Plume Removal," pp. 67-80 in O. Schleider and A. Zijderveld, (eds.), *AI Methods in Civil Engineering Applications2nd Joint Workshop on AI Methods in Civil Engineering Applications*, Cottobus, 2000.

Mayr, T. and N. J. Jarvis, "Pedotransfer Functions to Estimate Soil Water Retention Parameters for a Modified Brooks-Corey Type Model," *Geoderma*, 91: 1-9. 1999.

McLaughlin D. "Recent Developments in Hydrologic Data Assimilation" *Reviews of Geophysics*. *Supplement*: 977-984, US National report to International Union of Geodesy and Geophysics, Paper 95RG00740, American Geophysical Union. Washington, D. C., 1995.

Meisel, W. S. and D. C. Collins, " Repromodeling: An approach to efficient model utilization and interpretation'" IEEE Transactions on Systems, Man, and Cybernetics SMC-3: 349-358, 1973.

Messing, I., "Estimation of Saturated Hydraulic Conductivity in Clay Soils from Moisture Retention Data," *Soil Science Society of America Journal*, 53: 665-668, 1989.

Meyer, P. D., Rockhold, M. L., Gee, G. W., and T. J. Nicholson, "Uncertainty Analyses of Infiltration and Subsurface Flow and Transport for SDMP Sites," *NUREG/CR-6565, PNNL-11705*. U. S. Nuclear Regulatory Commission. Washington, DC 20555-0001, 1997.

Meyer, P.D., Taira, R. Y., and T. J. Nicholson, Hydrologic Uncertainty Assessment for Decommissioning Sites: Hypothetical Test Case Applications. Pacific Northwest

National Laboratory, Richland, WA 99352. NUREG/CR-6695. Prepared for Division of Risk Analysis and Applications Office of Nuclear Regulatory Research U. S. Nuclear Regulatory Commission Washington, DC 20555-0001 NRC JCN W6933, 34 pp., 2001.

Milly, P.C.D., "Climate, Soil Water Storage, And The Annual Average Water Balance," *Water Resources Research,* 30: 2143-2156, 1994.

Minasny, B. and Perfect, E., "Solute adsorption and transport parameters", Pp. 195-224, In: In: Pachepsky, Y. A., and Rawls, W J. (ed.). Development of pedotransfer functions in soil hydrology. Elsevier, 512 pp., 2004.

Minasny, B., and. McBratney, A.B, "The Neuro-M Method for Fitting Neural Network Parametric Pedotransfer Functions," *Soil Science Society of America Journal*, 66: 352-362, 2002.

Mitchell, M., "An Introduction to Genetic Algorithms," MIT Press, Cambridge, MA, 205 p., 1996.

Mohanty, B.P., Bowman R.S., Hendrickx J.M.H., and M.Th.van Genuchten, "New Piecewise-Continuous Hydraulic Functions for Modeling Preferential Flow in an Intermittent Flood-Irrigated Field," *Water Resources Research*, 33: 2049-2063, 1997.

Molteni, F., Buizza, R., Palmer, T.N., and Petroliagis, T. 1996: The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Met. Soc.*, 122, 73-119.

Nayak, P. "Causal Approximations," *Artificial Intelligence*, 70: 277-334, 1994.

Neuman, S. and V. Di Federico, "Multifaceted Nature of Hydrogeologic Scaling and its Interpretation," *Reviews of Geophysics*, 41(3): 1014 (10.1029/2003RG000130), 2003

Neuman, S. P. "Universal Scaling of Hydraulic Conductivities and Dispersivities in Geologic Media," *Water Resources Research*, 26(8): 1749-1758, 1990.

Neuman, S. P., Wierenga, P. J., and T. J. Nicholson, "A Comprehensive Strategy of Hydrogeologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites," *NUREG/CR 6805*. U. S. Nuclear Regulatory Commission. Washington, D.C. 20555-0001, 2003.

Neuman, S.P. "On Advective Dispersion Fractal Velocity in Permeability Fields," *Water Resources Research*, 31(6): 1455-1460. 1995.

Neupauer, R.M.and Powell, K.L., "A fully-anisotropic Morlet wavelet to identify dominant orientations in a porous medium," *Computers & Geosciences* (in press), 2005.

Nielsen, D.R., M. Th. Van Genuchten, and J. W. Biggar. "Water Flow and Solute Transport Processes in the Unsaturated Zone," *Water Resources Research,* 22 (9,suppl.): 89S-108S, 1986.

Nikolaeva, S. A., Pachepsky, Ya. Shcherbakov, R. A. and A. I. Shcheglov. "Modeling Moisture Regime of Ordinary Chernozems," *Pochvovedenie*, 6: 52-59, 1986 (in Russian).

Nimmo, J.R. "What Measurable Properties Can Predict Preferential Flow?" Paper No. 324. *Transactions of the 17th World Congress of Soil Science*, August 14-21, Bangkok, Thailand, 2002.

Oliver, Y.M., and K.R.J. Smettem, "Parameterization of Physically Based Solute Transport Models In Sandy Soils," *Aust. J. Soil Res.,* 41: 771-788, 2003.

Oosterveld, M., and C. Chang, "Empirical Relations Between Laboratory Determinations of Soil Texture and Moisture Characteristic," *Can. Agric. Eng.* 22:149-151, 1980.

Pachepsky, Y. A., Guber, A. K., and D. Jacques, "Temporal Persistence in Vertical Distributions of Soil Moisture Contents," *Soil Science Society of America Journal*, 69(2) (in press), 2005.

Pachepsky, Y. A., Timlin, D. J., and W. J. Rawls, Generalized Richards' Equation to Simulate Water Transport in Unsaturated Soils," *Journal of Hydrology*, 272: 3–13, 2003.

Pachepsky, Y., Benson, D. and W.J. Rawls, "Simulating Scale-Dependent Solute Transport in Soils With the Fractional Advective-Dispersive Equation," *Soil Science Society of America Journal,* 64:1234-1243, 2000.

Pachepsky, Y.A., and Rawls, W. J. (eds.), "Development of Pedotransfer Functions in Soil Hydrology," *Developments in Soil Science*, 30, Amsterdam, Elsevier, 2004.

Pachepsky, Y.A., and W.J. Rawls, "Soil Structure and Pedotransfer Functions," *European Journal of Soil Science*, 54: 443-452, 2003.

Pachepsky, Y.A., Mironenko, E.V., and R.A. Scherbakov, "Prediction and Use of Soil Hydraulic Properties," pp. 203-212 in: M.Th. van Genuchten, F.J. Leij, and L.J. Lund (eds.), *Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils*, Riverside: University of California, 1992.

Pachepsky, Y.A., Timlin, D.J., and W.J. Rawls, "Soil Water Retention as Related to Topographic Variables," *Soil Science Society of America* Journal, 65: 1787-1795, 2001.

Pachepsky, Ya. , Benson D. and W. Rawls, "Simulating Scale-Dependent Solute Transport in Soils With the Fractional Advective-Dispersive Equation," *Soil Science Society of America Journal*, 64: 1234-1243, 2000.

Pachepsky, Ya. A., and W. J. Rawls, "Accuracy And Reliability of Pedotransfer Functions as Affected by Grouping Soils," *Soil Science Society of America Journal,* 63: 1748-1757, 1999.

Pachepsky, Ya. A., Shcherbakov, R. A., Varallyay, Gy. and K.Rajkai, "Statistical Analysis of the Correlation Between the Water Retention and other Physical Soil Properties," *Pochvovedenie* 2: 56-66 (in Russian), 1982.

Pachepsky, Ya. A., W. J. Rawls, and D. J. Timlin, "The Current Status of Pedotransfer Functions: Their Accuracy, Reliability, and Utility In Field- and Regional-Scale Modeling," pp. 223-234 in D. L.Corwin, , K. Loague, and T. R. Ellsworth (Eds.), *Assessment of Non-point Source Pollution in the Vadose Zone*, *Geophysical monograph* 108. American Geophysical Union, Washington, D. C., 1999.

Pachepsky, Ya., D. Timlin, and L. Ahuja. "Estimating Saturated Soil Hydraulic Conductivity Using Water Retention Data and Neural Networks," *Soil Science,* 164:552-560, 1999.

Palmer, R.N. and J.L. Cohan, "Complexity in Columbia River Systems Modeling." *Journal of Water Resources Planning and Management-ASC*E, 112(4): 453-468, 1986.

Palmer, T.N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F.J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson, Development of a European multi-model ensemble system for seasonal to

inter-annual prediction (DEMETER). Bulletin of the American Meteorological Society, 85, 853-872, 2004.

Pan, L. and P.J. Wierenga, "A Transformed Pressure Head Based Approach to Solve Richards' Equation for Variably Saturated Soils," *Water Resources Research*, 31: 925-931, 1995.

Perfect, E., McLaughlin, N. B., Kay, B. D. and G. C. Topp, "An Improved Fractal Equation for the Soil Water Retention Curve," *Water Resources Research*, 32: 281-287, 1996.

Petersen, G. W., Cunningham, R. L., and R. P. Matelski, "Moisture Characteristics of Pennsylvania Soils: I. Moisture Retention as Related To Texture," *Soil Science Society of America Proceedings*, 32: 271-275, 1968.

Pohlman, K. F., Hassan, A. E., Chapman, J. B. "Modeling Density-Driven Flows and Radiounuclide Transport at an Underground Nuclear Test: Uncertainty Analysis and Effect of Parameter Correlation," *Water Resource Research*, 385): 1059 (DOI 10.29/2001WR001047), 2002.

Pruess, K. "A General-Purpose Numerical Simulator for Multiphase Fluid and Heat Flow," *Report LBL-29400*, Lawrence Berkeley Laboratory, Berkeley, CA, 103 p., 1991.

Rajkai, K. and Gy. Várallyay, "Estimating Soil Water Retention from Simpler Properties by Regression Techniques," pp.417-426 in M. Th. van Genuchten, F. J. Leij, and L. 3. Lund (eds.), *Proc. Int. Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils,* University of California, Riverside, CA, 1992.

Rawls, W. J., and D. L. Brakensiek, "Prediction of Soil Water Properties for Hydrologic Modeling," pp.293-299 in E. B. Jones and T. J. Ward (eds.), *Proc. Symp. Watershed Management in the Eighties*, April 30-May 1, 1985, Denver, CO. Am. Soc. Civil Eng., New York, NY. 1985.

Rawls, W. J., and D. L. Brakensiek. 1989. Estimation of Soil Water Retention and Hydraulic P roperties. p.275-300. *In* Morel-Seytoux (ed.) *Unsaturated Flow* in *Hydrologic Modeling; Theory and Practice. Proc.* NATO Adv. Res. Workshop *Hydrology,*" ASCE Press, New York, 2002.

ASI Series, Ser. C, Vol.275. KIuwer Academic Press, Dordrecht, The Netherlands.

Rawls, W. J., and Ya. A. Pachepsky, "Using Field Topographic Descriptors to Estimate Soil Water Retention," *Soil Science*, 167: 423-235, 2002

Rawls, W. J., Brakensiek, D. L., and B. Soni, "Agricultural Management Effects on Soil Water Processes, Part I. Soil Water Retention and Green-Ampt Parameters," *Trans. ASAE* 26: 1747-1752, 1983.

Rawls, W. J., Brakensiek, D. L., and K. E. Saxton, "Estimation of Soil Water Properties," *Trans. ASAE* 25: 1316-1320, 1982.

Rawls, W. J., L R. Ahuja, and D. L Brakensiek. "Estimating Soil Hydraulic Properties from Soils Data," p.329-340 in M. Th. van Genuchten, F. J. Leij, and L. Lund (eds.) *Proc. Int. Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils,* University of California, Riverside, CA. 1992.

Rawls, W.J., Giménez, D., and Grossman, R., "Use of soil texture, bulk density and slope of the water retention curve to predict saturated hydraulic conductivity. Trans. ASAE 41(4): 983-988, 1998.

Reichardt, K., Bacchi, O.O.S., Villagra, M.M., Turatti, A.L. and Z.O., Pedrosa, "Hydraulic Variability in Space and Time in a Dark Red Latosol of The Tropics," *Geoderma,* 60: 159-168, 1993.

Reilly, T. E. and A. W. Harbaugh, "Guidelines for Evaluating Ground-Water Flow Models," *U. S. Geological Survey Scientific Investigation Report 2004-5038*, 30 p., 2004.

Reynolds, S.G., "The Gravimetric Method of Soil Moisture Determination, Part III: An Examination of Factors Influencing Soil Moisture Variability," *Journal of Hydrology,* 11: 288-300, 1970.

Rickel, J., and P. Porter, ``Automated Modeling of Complex Systems to Answer Prediction Questions", Artificial Intelligence Journal, 93(1-2), pp. 201--260, 1997.

Rivers, E.D. and R.F. Shipp, "Available Water Capacity of Sandy and Gravely North Dakota Soils," *Soil Science*, 113: 74-80,1972.

Romano, N., and M. Palladino, "Prediction of Soil Water Retention Using Soil Physical Data and Terrain Attributes," *Journal of Hydrology.*, 265: 56-75, 2002.

Rubin, Y., "*Applied stochastic hydrogeology,*"Oxford University Press, Oxford, 2003

Saxton, K. E., W. J. Rawls, J. S. Romberger, and R. I. Papendick, "Estimating Generalized Soil-Water Characteristics from Texture," *Soil Science Society of America Journal*, 50:1031-1036, 1986.

Scanlon, B. R., Tyler, S. W., P. J. Wierenga, "Hydrologic Issues in Arid, Unsaturated Systems And Implications for Contaminant Transport," *Rev. Geophys.*, 33: 461-490, 1997.

Scanlon, B. Software Review, 2004. http://typhoon.mines.edu/software/igwmcsoft/hydrus1d_review.htm

Schaap, M.G. "Accuracy and uncertainty of PTF predictions". Pp. 33-42. In: Pachepsky, Y. A., and Rawls, W J. (ed.). Development of pedotransfer functions in soil hydrology. Elsevier, 512 pp., 2004.

Schaap, M.G. and F. J. Leij. "Database-related Accuracy and Uncertainty of Pedotransfer Functions," *Soil Science*, 163: 765-779, 1998.

Schaap, M.G. and F.J. Leij, "Improved Prediction of Unsaturated Hydraulic Conductivity with the Mualem-Van Genuchten Model," *Soil Science Society of America Journal,* 64:843-851, 2000.

Schaap, M.G. and F.J. Leij, "Improved Prediction of Unsaturated Hydraulic Conductivity with the Mualem-Van Genuchten Model," *Soil Science Society of America Journal,* 64:843-851, 2000.

Schaap, M.G., Leij, F.G., van Genuchten, M.T., "Neural network analysis for hierarchical prediction of soil hydraulic properties", Soil Sci. Soc. Am. J, 62:847-855, 1998.

Schaap, M.G., Leij, F.J., and M.Th. van Genuchten, "Rosetta: A Computer Program for Estimating Soil Hydraulic Parameters with Hierarchical Pedotransfer Functions," *Journal of Hydrology,* 251:163-176, 2001.

Schmied, B., Abbaspour, K., and R. Schulin, "Inverse Estimation of Parameters in a Nitrogen Model Using Field Data," *Soil Science Society of America Journal*, 64: 533–542, 2000.

Simon, H. and A. Ando, "Aggregation of Variables in Dynamic Systems," *Econometrica*, 29: 111–138, 1996.

Simunek, J., M. Sejna, and M. Th. van Genuchten, "The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version 2.0", IGWMC - TPS - 70, *International Ground Water Modeling Center, Colorado School of Mines, Golden, Colorado,* 186pp., 1998

Sisty, A.F. and S.D. Farr, "Model Abstraction Techniques: An Intuitive Overview," *Air Force Research Laboratory/IFSB*, Rome, New York, 2004. http://www.rl.af.mil/tech/papers/ ModSim/ModAb_Intuitive.html

Smettem, K.R.J., Oliver, Y.M., Pracilio, G., and R.J. Harper, "Data Availability and Scale in Hydrologic Applications," pp.253-272, in*:* Pachepsky, Y. and Rawls, W.J. (eds.), *Development of Pedotransfer Functions in Soil Hydrology.* Amsterdam, Elsevier, 2004.

Smith, S.K., "Further Thoughts on Simplicity and Complexity in Population Projection Models," *International Journal of Forecasting*, 13(4): 557 – 565, 1997.

Soil Survey Staff, 1951. Soil Survey Manual. USDA Handbook No. 18. US Government Printing Office. Washington, D.C.

Statsoft, "STATISTICA for Windows. Computer Program Manual", StatSoft, Inc., Tulsa, CD-ROM, 1999.

Stephens, P. A., Frey-Roos, F. Arnold, W., and W. J. Sutherland, "Model Complexity and Population Predictions. The Alpine Marmot As A Case Study," *Journal of Animal Ecology*, 71: 343–361, 2002.

Stillman R. A. , McGrorty, S., Goss-Custard, J. D. and A. D. West, "Predicting Mussel Population Density and Age Structure: The Relationship Between Model Complexity and Predictive Power," *Marine Ecology Progress Series* 208: 131–145, 2000

Suleiman, A.A. and J. T. Ritchie, "Modeling Soil Water Redistribution Under Second Stage Evaporation," *Soil Science Society of America Journal*, 67(2): 377-386, 2003.

Tamboli P.M., Larson, W.E. and M. Amemya, "Influence of Aggregate Size on Moisture Retention," *Iowa Acad. Sci*. 71: 103-108, 1964.

Therrien, R., and E.A. Sudicky, "Three-Dimensional Analysis of Variably-Saturated Flow and Solute Transport in Discretely-Fractured Porous Media," *Journal of Contaminant Hydrology*, 23: 1-44, 1996.

Tietje, O., and M. Tapkenhinrichs. "Evaluation of Pedo-transfer Functions," *Soil Science Society of America Journal*, 57: 1088-1095, 1993.

Timlin, D. J., L. R. Ahuja, and R. D. Williams, "Soil Hydraulic Parameters for Regional-Scale Applications of Mechanistic Models," pp. 185- 203 in D. L. Corwin and K. Loague (eds.), *Application of GIS to the modeling of non-point source pollutants in the vadose zone. SSSA Special publication 48,* American Society of Agronomy, Madison, WI, 1996b.

Timlin, D. J., Y. A. Pachepsky, and C. L. Walthall. "A Mix of Scales: Topography, Point scales, and Yield Maps," pp. 227-242, in Y. A. Pachepsky, D. E. Radcliffe, and H. M. Selim (eds.), *Scaling Methods in Soil Physics,* CRC Press, Boca Raton, FL, 2003.

Timlin, D. J., Ya. Pachepsky, B. Acock, and F. Whisler. "Indirect Estimation of Soil Hydraulic Properties to Predict Soybean Yield Using GLYCIM," *Agricultural Systems*, 52: 331-353, 1996a.

Tomasella, J. and M.G. Hodnett, "Estimating Soil Water Retention Characteristics from Limited Data in Brazilian Amazonia," *Soil Science,* 163: 190-202, 1998.

Tsai, F. T-C., N-Z. Sun, and W. W-G. Yeh, "Global-Local Optimization Methods for the Identification of Three-Dimensional Parameter Structure in Groundwater Modeling," Water Resources Research, 39, 1043, doi:10.1029/2001WR001135, 2003

U.S. NRC, "Branch Technical Position on Site Characterization for Decommissioning," U. S. Nuclear Regulatory Commission, Washington, DC, 1994.

U.S. NRC, "Site Decommissioning Management Plan," *NUREG-1444,* U. S. Nuclear Regulatory Commission, Washington, DC, 1993.

Unsaturated Flow in Hydrologic Modelling. ArIes, France, June 13-17 1988. NATO

Vachaud, G., Passerat de Silans, A., Balabanis, P., and M. Vauclin, "Temporal Persistence of Spatially Measured Soil Water Probability Density Function," *Soil Science Society of America Journal,* 49: 822- 828, 1985.

Valocchi, A.J. "Validity of The Local Equilibrium Assumption For Modeling Sorbing Solute Transport Through Homogeneous Soils," *Water Resources Research*, 21: 808-820, 1985.

Van Dam, J.C. and R.A. Feddes, "Numerical Simulation of Infiltration, Evaporation and Shallow Groundwater Levels with the Richards Equation," *Journal of Hydrology*, 233: 72-85, 2000.

Van Genuchten, M. T., Leij, F. J., and Lund, L. J. (eds.), "Proceedings of the International Workshop on Indirect methods for Estimating the Hydraulic Properties of Unsaturated Soils", University of California, Riverside, CA, 1992.

Van Genuchten, M. Th, F. J. Leij, and L. Wu (Eds.). "Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media," *Proceedings of the International Workshop*, Riverside, California, October, 22-24, 1997, University of California, Riverside, 1997.

Van Genuchten, M. Th. "A Closed Form Equation for Predicting the Hydraulic Conductivity Of Unsaturated Soils," *Soil Science Society of America Journal*, 44: 892-898, 1980.

Van Genuchten, M.Th., "Non-equilibrium transport parameters from miscible displacement experiments. Research Rep. No. 119," U.S. Salinity Laboratory, USDA-SEA-ARS, Riverside, CA, 1981.

Van Ness, E. H., Scheffer, M, "A strategy to improve the contribution of complex simulation models to ecological theory", Ecological Modelling, 185: 153-164, 2005.

Ventrella, D., Mohanty, B. P., Simunek, J. Losavio, N., and M. Th. van Genuchten, "Water and Chloride Transport in a Fine-Textured Soil: Field Experiments and Modeling," *Soil Science*, 165(8): 624-631, 2000.

Vereckeen, H., Maes, J., Feyen, J., and P. Darius, "Estimating the Soil Moisture Retention Characteristics from Texture, Bulk Density and Carbon Content," *Soil Science,* 148: 389-403, 1989.

Wang, C. G., and D. G. Jamieson, "An Objective Approach to Regional Wastewater Treatment Planning," *Water Resource Research*, 38(3): 1022, (DOI 10.1029/2000WR000062), 2002.

Wang, W, Neuman, S. P., Yao, T., and P. J. Wierenga, "Simulation of Large-Scale Field Infiltration Experiments Using a Hierarchy of Models Based on Public, Generic, and Site Data," *Vadose Zone Journal,* 2: 297–312, 2003.

Weiner, E. S. C. and Simpson, J. A. (Eds.). "The Compact Oxford English Dictionary," Oxford University Press, 1991.

Weld, D. "The Use of Aggregation in Causal Simulation," *Artificial Intelligence*, 30: 47-55, 1986.

Whelan, G., J. P. McDonald, and C. Sato, "Multimedia Environmental Pollutant Assessment System (MEPAS): Groundwater Pathway Formulations," *PNNL-10907,* Pacific Northwest National Laboratory, Richland, WA, 1996.

Whitmore, A.P., "A method for assessing the goodness of computer simulation of soil processes," *J. Soil Sci.*, 42:289-299, 1991.

Williams, J., Ross, P., and K. Bristow, "Prediction of the Campbell Water Retention Function From Texture, Structure, And Organic Matter," pp. 427-442 in M. Th. van Genuchten, F. J. Leij, and L J. Lund (eds.), *Proceedings of the International Workshop on Indirect methods for Estimating the Hydraulic Properties of Unsaturated Soils*, University of California, Riverside, CA, 1992.

Wilson, G.V., Gwo, J.P., Jardine, P.M., and R.J. Luxmoore, "Hydraulic and Nonequilibrium Effects on Multiregion Flow," pp. 37-61 in: H.M. Selim and L. Ma (eds.), *Physical Nonequilibrium in Soils; Modeling and Application*, Ann Arbor Press, Chelsea, MI, 1998.

Wittmuss, H.D., and A.P. Mazurak, "Physical and Chemical Properties of Aggregates in a Brunizem Soil," *Soil Science Society of America Proceedings,* 22: 1-5, 1958.

Wösten J.H.M., A. Lilly, A. Nemes, and C. Le Bas, "Development and use of a database of hydraulic properties of European soils," *Geoderma*, 90, 169-185, 1999.

Wösten, J.H.M., Y. A. Pachepsky, and W. J. Rawls, "Pedotransfer Functions: Bridging the Gap Between Available Basic Soil Data and Missing Soil Hydraulic Characteristics," *Journal of Hydrology,* 251: 123-150, 2001.

Wu, Y.-S., Zhang, K., and K. Pruess, "Massively Parallel Simulation of Flow and Transport in Variably Saturated Porous and Fractured Media," Lawrence Berkeley National Laboratory. Paper LBNL-49407, 8 pp., 2002.

Ye, M., Neuman, S. P., and Meyer, P. D., "Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff,". Water Resour. Res., *40*, W05113, doi:10.1029/2003WR002557, 2004

Yi-Bing Lin and P. A. Fishwick, "Asynchronous Parallel Discrete Event Simulation," *IEEE Transactions on Systems, Man and Cybernetics*, 26: 397-412, 1996.

Yilmaz, L. and T.I. Ören, Dynamic "Model Updating in Simulation with Multimodels: A Taxonomy and a Generic Agent-Based Architecture," *Proceedings of SCSC 2004 - Summer Computer Simulation Conference*, July 25-29, 2004, San Jose, CA., pp. 3-8. 2004.

Zeigler, B. "Theory of Modeling and Simulation," Academic Press, NY, 510 pp., 2000.

Zeigler, B. P., Kim, T. G., and Praehofer, H, "*Theory of Modeling and Simulation*. 2 ed., New York, NY, Academic Press, 1998.

Zeigler, B., "Theory of modeling and simulation," New York, New York: Wiley and Sons, 1976.

Zhang, T., and R. Berndtsson, "Temporal Patterns and Spatial Scale of Soil Water Variability In a Small Humid Catchment," *Journal of Hydrology,* 104: 111-128, 1988.

Zhu, J. and B. P. Mohanty, "Spatial Averaging of Van Genuchten Hydraulic Parameters for Steady State Flow in Heterogeneous Soils: A Numerical Study," *Vadose Zone Journal*, 1: 261-272, 2002.

Zurmühl, T. "Capability Of Convection-Dispersion Transport Models To Predict Transient Water and Solute Movement in Undisturbed Soil Columns," *Journal of Contaminant Hydrology*, 30: 101-110, 1998.

Zurmühl, T., and W. Durner, "Modeling Transient Water and Solute Transport in a Biporous Soil," *Water Resources Research*, 32: 819-829, 1996.

Appendix A. Pedotransfer functions to estimate soil water retention

Most of equations presented below were developed from large (>200) samples databases. Symbols w and θ denote volumetric ($cm^3 cm^{-3}$) and gravimetric ($cm^3\ g^{-1}$) water contents, respectively, subscripts indicate the capillary pressure in cm, clay, and sand denote percentages of textural fractions according the USDA textural classification, OM is the organic matter content, %, OC is the organic carbon content, ρ_b is the bulk density g cm^{-3}, other symbols are defined as they appear.

A.1. *Equations developed to estimate soil water content at fixed soil water potentials*

Baumer (1992) used the US National Soil Survey database to relate gravimetric water contents at capillary pressures of 15000 cm and 330 cm with the equations:

$$w_{15000}=0.01*\rho_b*(0.71+0.45*OM+0.336*clay+0.117*clay*(CA^{3/2})+0.004*clay*SAR) \quad (A1)$$

$$w_{330}=0.01*\rho_b*(15.84+0.746*OM+2.2025*CA^2-0.137*sand+0.743*w_{15000}) \quad \text{if clay} > 10\% \quad (A2)$$

$$w_{330}=0.01*\rho_b*(15.84+0.746*OM+0.02*CA^2*clay^2-0.137*sand+0.743*w_{15000}) \quad \text{if clay} \leq 10\% \quad (A3)$$

where CA is the clay activity, SAR is the sodium adsorption ratio.

Bruand et al. (1994) estimated volumetric water contents at capillary pressures of 15000 cm and 330 cm as

$$\theta_{15000} = (0.008+0.00367*clay)/(0.471+0.00411*clay) \quad (A4)$$

$$\theta_{330} = (0.043+0.004*clay)/(0.471+0.00411*clay) \quad (A5)$$

Canarache (1993) applied regression analysis to the Romanian national database and obtained predictive equations:

$$\theta_{15000}=0.01*\rho_b*(0.2805*clay+0.0009615*clay^2) \quad (A6)$$

$$\theta_{330}=0.01*\rho_b*(2.65+1.105*clay-0.01896*clay^2+0.0001678*clay^3+15.12*\rho_b-6.745*\rho_b^2-0.1975*clay*\rho_b) \quad (A7)$$

Gupta and Larson (1979) used a subset of the US National Cooperative Survey database to derive equations for estimating volumetric water contents at capillary pressures of 15000 cm and 330 cm as

$$\theta_{330}=0.003075*sand+0.005886*silt+0.008039*clay+0.002208*OM-0.1434*\rho_b \quad (A8)$$

$$\theta_{15000}=-0.000059*sand+0.001142*silt+0.005766*clay+0.00228*OM+0.02671*\rho_b \quad (A9)$$

Hall et al. (1977) analyzed a subset of British Soil Survey data and arrived to the equations

$$\theta_{330}=0.01*(20.81+0.45*clay+0.13*silt-5.95*\rho_b) \quad (A10)$$

$$\theta_{15000}=0.01*(1.48+0.84*clay-0.0055*clay^2) \quad (A11)$$

Petersen et al. (1968) worked with the Pennsylvania soil database. Their equations are:

$$\theta_{330}=0.01*(11.83+0.96*clay-0.008*(clay)^2) \quad (A12)$$

$$\theta_{15000}=0.01*(1.74+0.76*clay -0.005(clay)^2) \quad (A13)$$

Rajkai and Varallyay (1992) analyzed the Hungarian national database:

θ_{330}=0.01*(38.62-0.00479*sand-0.0019*(sand/silt)^2) (A14)

θ_{15000}=0.01*(1.39+0.36*clay +0.22*OM^2) (A15)

We used clay content instead of unavailable fine fraction content in Eq. (A14).

Tomasella and Hodnett (1998) have been working with Brazilian soils and came up with the equations:

θ_{330}=0.01*(4.046+0.426*silt+0.404*clay) (A16)

θ_{15000}=0.01*(0.91+0.150*silt+0.396*clay) (A17)

A.1.2 Equations developed to estimate Brooks-Corey parameters

Rawls and Brakensiek (1989) developed the following equations to estimate the Brooks-Corey parameters in Eq. (1)

h_b =exp[5.340+0.185*clay-2.484*ϕ -0.002*clay^2-0.044*sandϕ +0.001*sand^2*ϕ^2-0.009*clay^2*ϕ^2 -0.00001*sand^2*clay+0.009*clay^2*sand-0.0007*sand^2*ϕ 0.000005*clay^2*sand-0.500*ϕ^2*clay] (A18)

λ=exp[-0.784+0.018*sand-1.062*ϕ -0.00005*sand^2-0.003*clay^2+1.111*ϕ^2-0.031*sand*ϕ +0.0003*sand^2*ϕ^2-0.006*clay^2*ϕ^2-0.000002*sand^2*clay+0.008*clay^2*ϕ -0.007*ϕ^2*clay] (A19)

θ_r=-0.018+0.0009*sand+0.005*clay+0.029*ϕ -0.0002*clay^2-0.001*sand*ϕ -0.0002*clay^2*ϕ^2+0.0003*clay^2*ϕ -0.002*ϕ^2*clay (A20)

Campbell and Shiozava (1992) set the value of the residual water content in Eq. (1) equal to zero and transformed this equation to

$$h = h_e\left(\theta / \theta_s\right)^{-b} \quad \text{(A21)}$$

where h_e is "air entry potential". The parameters in (1), estimated from two data sets for British soils, were found to be

$$h_{es} = -0.05 d_g^{-1/2}$$
$$b = -20 h_{es} + 0.2\sigma_g \quad \text{(A22)}$$

where the value of h_{es} corresponds the air entry potential at a standard bulk density, ρ_b of 1.3 g cm^{-3}. The proposed adjustment for bulk density is

$$h_e = h_{es}\left(\rho_b / 1.3\right)^{0.67b} \quad \text{(A23)}$$

The geometric mean diameter d_s and geometric standard deviation are given by

d_g=exp(-0.025- 0.0363silt – 0.0688clay) (A24)

σ_g=exp(0.133silt +0.477clay - ln$^2 d_g$)$^{1/2}$

Saxton et al. (1986) also set the value of the residual water content in Eq. (1) equal to zero and transformed this equation to

$$h = \mathrm{A}\theta^{\mathrm{B}} \quad \text{(A25)}$$

where

A=100*exp(-4.396-0.0715*clay-0.000488*sand^2-0.00004285*sand^2*clay) (A26)

B=-3.140-0.00222*clay^2-0.00003484*sand^2*clay (A27)

Oosterveldt & Chang (1980) transformed Eq. (1) with θ_r=0 to the form

$$\theta = 0.01*\rho_b*(35.367+0.644*\text{clay}-0.251*\text{sand}-0.045*D)*h^{-0.190} \quad \text{(A28)}$$

where D is the mean depth of sample in centimeters.

Williams et al. (1992) transformed Eq. (1) with θ_r=0 to the logarithmic form

$$\ln\theta = A+B\ln h \quad \text{(A29)}$$

and applied it to the Australian database. Pedotransfer equations depended on the availability of data on basic properties. Equations

$$A=2.57+0.238*\ln(\text{clay})-0.000192*\text{sand}^2-0.0137*\text{sand}-0.0926*\ln(\text{OM})+0.0412*\text{OM} \quad \text{(A30)}$$

$$B=-0.403+0.0871*\ln(\text{clay})-0.00077*\text{sand} \quad \text{(A31)}$$

were suggested for cases when the data on organic matter are available, and equations

$$A=1.839+0.257*\ln(\text{clay})+0.381*2-0.0001*\text{sand}^2 \quad \text{(A32)}$$

$$B=-0.303+0.093*\ln(\rho_b)+0.0565*\ln(\text{clay})-0.00003*\text{sand}^2 \quad \text{(A33)}$$

were developed or cases when no information about the organic matter content was available.

Mayr and Jarvis (1999) set θr=0 and porosity ϕ equal to the saturated water θ_s in the Brooks-Corey equation, and combined this equation in the dry range of soil water retention curve

$$\theta = \theta_s (h/a)^{-1/b} \qquad \theta < \theta_i \quad \text{(A34)}$$

with a parabolic equation in the wet range:

$$\theta = \theta_s - \frac{\theta_s h^2 (1-\theta_i/\theta_s)}{a^2 (\theta_i/\theta_s)^{-2b}} \qquad \theta \geq \theta_i \quad \text{(A35)}$$

The water content θ_i and the equivalent pressure head h_i at the matching point are given by:

$$\theta_i = \frac{2b\theta_s}{1+2b} \quad \text{(A36)}$$

$$h_i = a\left(\frac{2b}{1+2b}\right)^{-b} \quad \text{(A37)}$$

Pedotransfer functions developed from a Scandinavian dataset were:

$$\log(a)=-4.9840297533+0.0509226283*\text{sand}+0.1575152771*\text{silt}+0.1240901644*\rho_b -0.1640033143*\text{OC}-0.0021767278*\text{silt}^2+0.0000143822*\text{silt}^3+0.0008040715*\text{clay}^2 +0.0044067117*\text{OC}^2) \quad \text{(A38)}$$

$$\log(1/b)=-0.8466880654-0.0046806123*\text{sand}+0.0092463819*\text{silt}-0.4542769707*\rho_b -0.0497915563*\text{OC}+0.0003294687*\text{sand}^2 0.000001689056*\text{sand}^3+0.0011225373*\text{OC}^2 \quad \text{(A39)}$$

$$\theta_s = 0.2345971971+0.0046614221*\text{sand}+0.0088163314*\text{silt}+0.0064338641*\text{clay}-0.3028160229*\rho_b +1.79762*10^2*\text{sand}^2-3.134631*10^2*\text{silt}^2 \quad \text{(A40)}$$

A.1.3 Equations developed to estimate Brooks-Corey parameters

Wösten et al. (1999) analyzed the all-Europe database and derived pedotransfer functions to estimate van Genuchten parameters in the Eq. (2):

θ_s = 0.7919 + 0.001691*clay - 0.29619*ρ_b - 0.000001491*silt^2 + 0.0000821*OM^2 + 0.02427/clay + 0.01113/silt + 0.01472*ln(silt) - 0.0000733*OM*clay - 0.000619*ρ_b*clay - 0.001183*ρ_b*OM - 0.0001664*topsoil*silt (A41)

α =exp[-14.96 + 0.03135*clay + 0.0351*silt + 0.646*OM +15.29*ρ_b - 0.192*topsoil - 4.671*ρ_b^2 - 0.000781*clay^2 - 0.00687*OM^2 + 0.0449/OM + 0.0663*ln(silt) + 0.1482*ln(OM) - 0.04546*ρ_b *silt - 0.4852*ρ_b*OM + 0.00673*topsoil*clay] (A42)

n =1.+exp[-25.23 - 0.02195*clay + 0.0074*silt - 0.1940*OM + 45.5*ρ_b - 7.24*ρ_b^2 + 0.0003658*clay^2 + 0.002885*OM^2 -12.81/*ρ_b - 0.1524/silt - 0.01958/OM - 0.2876*ln(silt) - 0.0709*ln(OM) -44.6*ln(ρ_b) - 0.02264*ρ_b*clay + 0.0896*ρ_b*OM + 0.00718*topsoil*clay (A43)

where topsoil is an ordinal variable having the value of 1 or of 0. Parameter *m* in Eq. (1) was computed as 1-1/*n*.

These authors have also estimated average values of van Genuchten parameters by the FAO textural classes. Those values are shown in Table A1 and the FAO textural classes are delimited in Fig. A1.

Vereecken et al. (1989) used the Belgian dataset to develop the following pedotransfer functions for van Genuchten equation with *m*=1:

θ_s = 0.81 – 0.283*ρ_b + 0.001*clay (A44)

θ_r = 0.015 + 0.005*clay + 0.014*OC (A45)

α= exp(-2.486 + 0.025*sand – 0.351*clay) (A46)

n=0.81 – 0.283*ρ_b + 0.001*clay (A47)

Pachepsky et al. (1982) applied van Genuchten equation with m=1 and θ_r = 0 to the Hungarian national database and found regression equations:

θ_s = 0.01*(-56.4*ρ_b + 0.00205*clay^2 + 123.79) (A48)

n = 0.336*ρ_b - 0.053 (A49)

α= $10^{-0.04701*\rho b*\text{clay} + 1.513*\rho b - 0.417}$ (A50)

Table A1. Tabulated parameters of van Genuchten equation by FAO textural classes (Wösten et al., 1999)

	FAO textural class	θ_r	θ_s	α	n
Topsoils	Coarse	0.025	0.403	0.0383	1.3774
	Medium	0.010	0.439	0.0314	1.1804
	Medium fine	0.010	0.430	0.0083	1.2539
	Fine	0.010	0.520	0.0367	1.1012
	Very fine	0.010	0.614	0.0265	1.1033
Subsoils	Coarse	0.025	0.366	0.0430	1.5206
	Medium	0.010	0.392	0.0249	1.1689
	Medium fine	0.010	0.412	0.0082	1.2179
	Fine	0.010	0.481	0.0198	1.0861
	Very fine	0.010	0.538	0.0168	1.0730

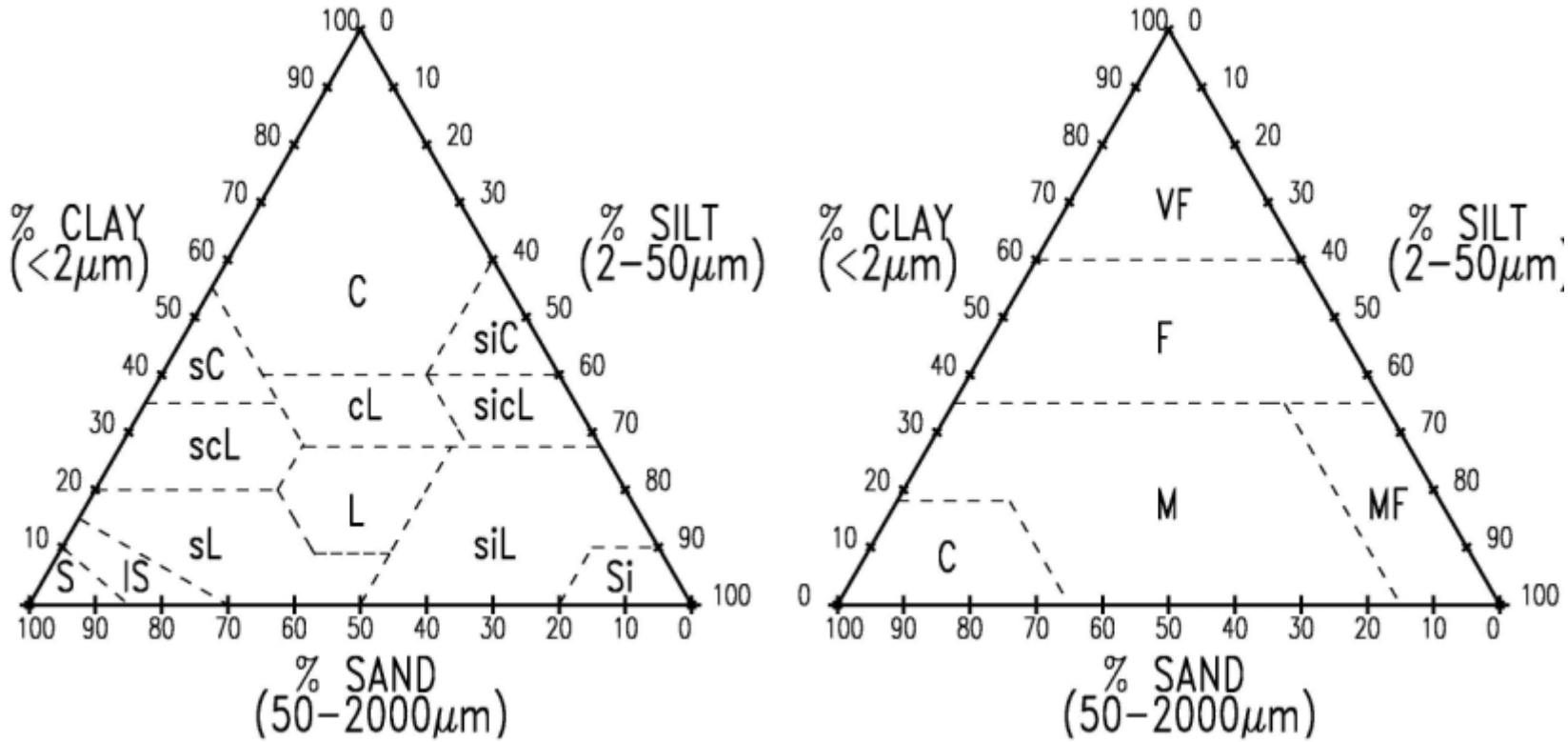

Figure A1. Textural classification systems according to (a) USDA (Soil Survey Staff, 1951) and (b) as defined by the FAO Soil Map of Europe (European Soil Bureau, 1998). USDA classes: S, sand, lS, loamy sand, sL, sandy loam, scL, sandy clay loam, sC, sandy clay, L, loam, cL, clay loam, C, clay, siL, silt loam, sicL, silty clay loam, siC, silty clay, Si, silt; FAO classes, C, coarse, M, medium, MF, medium fine, F, fine, VF, very fine.

Appendix B. Diurnal oscillations in water content and capillary pressure data

Inspection of time series of water content and pressure head revealed oscillatory behavior that apparently had diurnal periodicity. An example of such behavior is shown in Fig. B1 for the 10-day period between day 80 and day 90 from the beginning of the experiment at location "10". Both pressure head and water content (Fig. B1.a and Fig. B1.b) oscillated synchronously at all depths. Temperature time series at different depths had sine wave shape. The amplitude decreased and a phase had been shifted with depth (Fig. B1.c). Our first hypothesis was that the oscillations of water contents and capillary pressures at each depth could be related to temperature at this depth. To test this hypothesis, capillary pressure and temperature time series were subjected to spectral Fourier analysis with the *STATISTICA* software (StatSoft, 1999). Values of periods were equal to 24 hour for all pressure head time series (Fig. B2.a), and for soil temperature time series at depth 15 cm and 35 cm (Fig. B2.b). However, temperature time series at depths below 35 cm did not have the 24-hour period. Therefore, local oscillations of temperature could not be useful in explaining oscillations of capillary pressure and water contents at depths larger than 35 cm. The original hypothesis should be rejected.

The second hypothesis was that the surface temperature was the factor causing oscillations. To test this hypothesis, the time of the daily temperature maximum at each depth was plotted against time (Figure B3). The time when the linear regression line 'Phase shift – depth" crossed the depth axis, i.e. the regression intercept, was not significantly different from the time at which maximum capillary pressure was reached at all depths. A similar result was obtained for water contents (data not shown). This means that the oscillations of recorded capillary pressures and soil water contents were related to the temperature on/near soil surface. That could be attributed to the unknown response of the electronic equipment on the surface to the diurnal daily oscillations. To eliminate temperature effect of measurement devices, daily time series of water content and pressure head were averaged for whole observation period. The temporal time scale of observations became equal to one day.

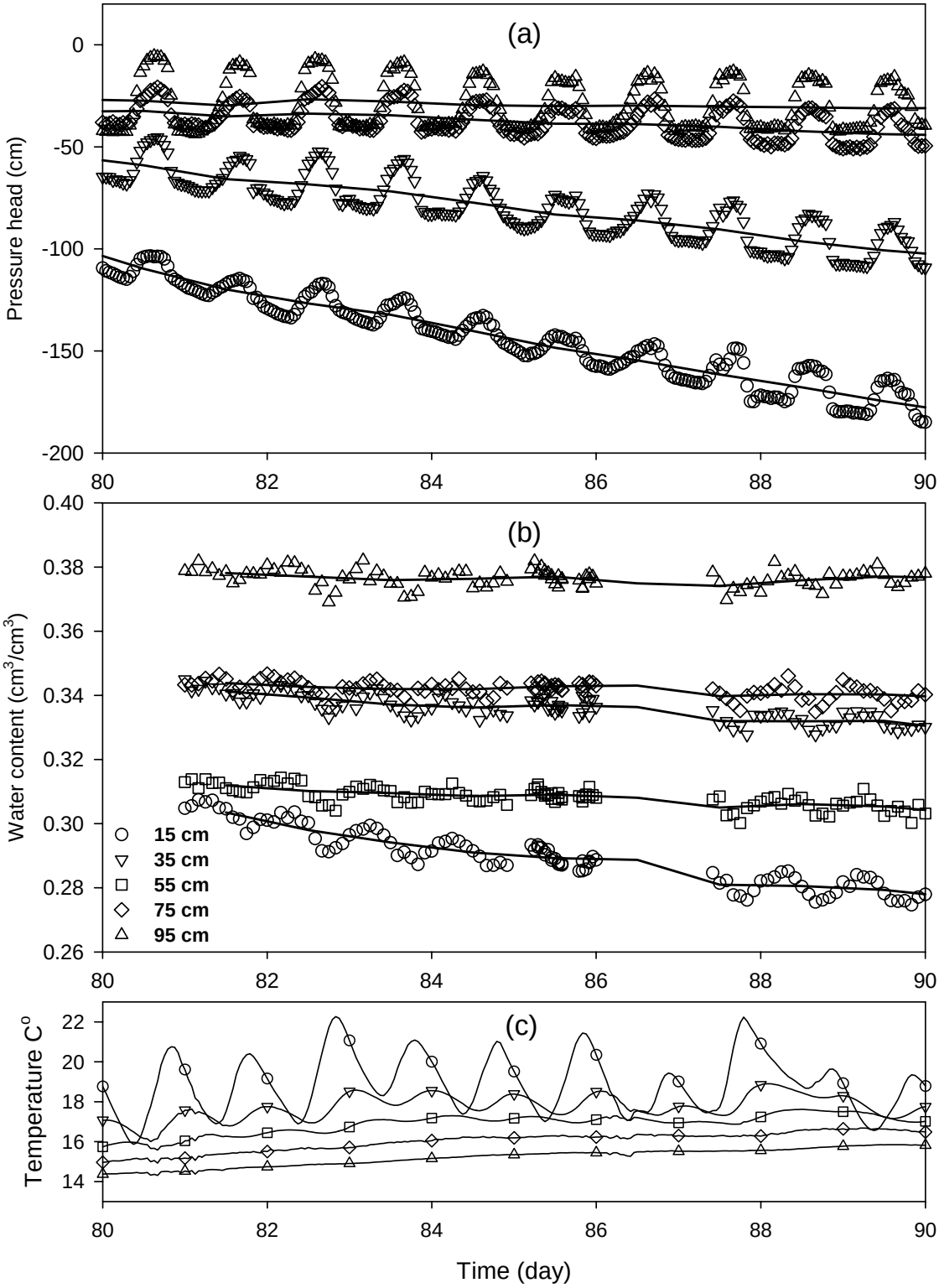

Figure B1. Time series of capillary pressures (a), water content (b) and soil temperature at different depths before (symbols) and after (lines) daily averaging at the location 10 cm.

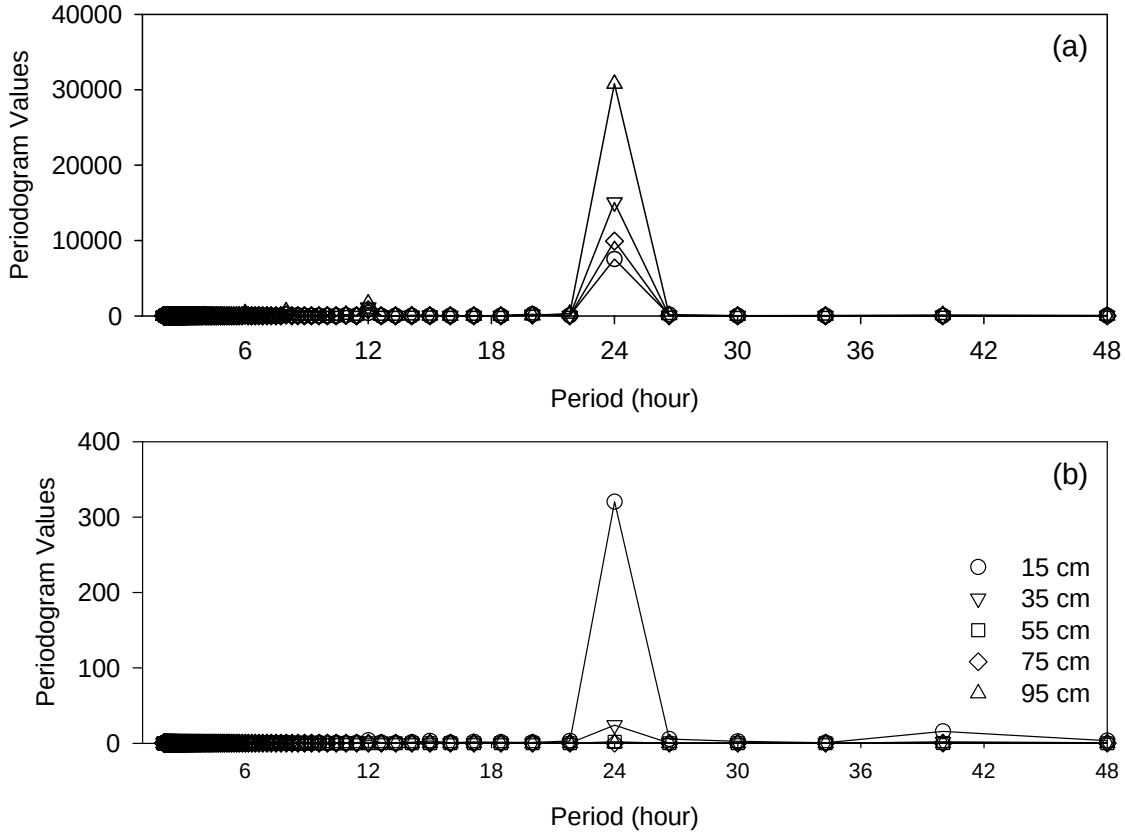

Figure B2. Periodograms for capillary pressure (a) and temperature (b) time series at different depths.

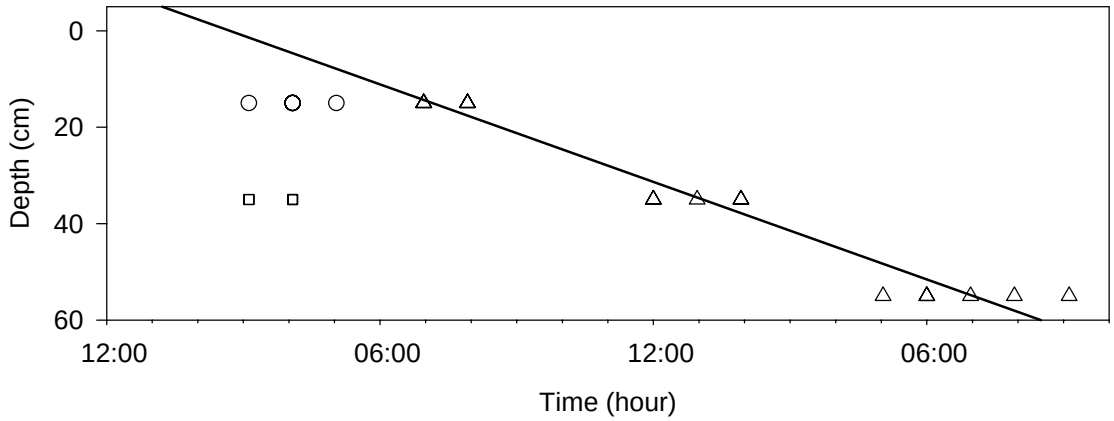

Figure B3. Time of the maximum temperature (Δ) and maximum pressure head at 15 cm (o) and at 35 cm (□) depth at the location "10 cm".

Appendix C. Using temporal persistence to estimate depth-averaged water contents with missing data

C.1 Introduction

When a field or a small watershed is repeatedly surveyed for soil water content, sites often can be spotted where soil is consistently wetter or consistently dryer than average across the study area. Existence of such sites is important for soil management. It is also important for selection of sites to infer the area-average soil water content to use at coarser scale characterization and simulation, i.e. to compare with remote sensing data or establishing field- or catchment-wide antecedent moisture conditions for runoff simulations (Grayson and Western, 1998). The phenomenon has been called time stability, temporal stability, or temporal persistence in spatial patterns of soil water content or in soil water contents. Temporal persistence of water contents at the same depth was documented by Vachaud et al. (1985), Kachanoski and de Jong (1988), Zhang and Berndtsson (1988), Goovaerts and Chiang (1993), Reichardt et al. (1993), and Ferreyra et al. (2002) across areas of various extents.

Relatively less is known about temporal persistence of water content at various depths. Comegna and Basile (1994) found a time-stable spatial structure for the water content in the top 90 cm of the soil profile. Cassel et al. (2000) observed greater temporal persistence of water content in deep soil layers than in shallow layers under a wheat crop. This effect could be attributed to the impact of crop root water uptake. Hupet and Vanclooster (2002) showed a substantial effect of vegetation on spatial and temporal structure in profile soil water distributions. Martinez-Fernandez and Ceballos (2003) did not observe any specific pattern of stability with respect to depth.

Our interest to the temporal persistence in soil water content in soil profile arose from the fact that some of TDR probes positioned along soil transect at five depths with 50-cm spacing eventually began to malfunction. Averaging data from the remaining probes at a given depth could be an option if no persistent difference had existed between TDR-measured water content in different locations. However such persistence clearly manifested itself in our dataset (see, e.g., Fig. 24). Because of that, the malfunctioning of probes with consistently lower than average values would lead to averaging only data from the remaining probes that had water contents consistently higher than the true average. This averaging would lead to overestimation of the average water content at the depth of interest. In contrary, the malfunctioning of probes with consistently higher than average values would lead to averaging only data from the remaining probes that had water contents consistently lower than the true average, and the average water contents would be underestimated.

To use the collected data in model calibration and evaluation, we had to propose a technique to utilize this persistence to remedy the effect of probe malfunctioning on the estimates of the average water content in the layer.

C.2 Temporal persistence in soil water contents

To quantify the persistence, we used an approach similar to the one proposed by Vachaud et al. (1985). The relative water contents β_{ij} for each sampling location *i* at the same depth for the measurement time *j* were computed as:

$$\beta_{ij} = \frac{\theta_{ij}}{\overline{\theta}_j} \tag{C1}$$

where θ_{ij} is the water content measured in location *i* at the *j*th measurement time, and $\overline{\theta}_j$ is the average water content at the *j*th measurement time at the depth of interest:

$$\overline{\theta}_j = \frac{1}{N}\sum_{i=1}^{N}\theta_{ij} \tag{C2}$$

where N is the total number of probes at a given depth (N=12 in our case).

Statistics of values β for all probes are shown in Table C1. Of the total 60 probes, 49 probes had the values of β_{ij} that were larger or less than 1 in more than 95% of cases. This demonstrates that temporal persistence exists in soil profile, and some probes consistently show water contents below average whereas others show water contents above the average. Substantial temporal linear trends in the $\overline{\beta}$ values with coefficients of determination R^2 larger than 0.3 were detected for four probes at the 15-cm depth, and were absent at all other depths.

To test the dependence of the temporal stability on depth, we used values of the standard deviations $s_{\beta i}$ of the β_{ij} values. The larger this standard deviation the less temporal stability is observed for the probe *i*. Inspection of the dependence of $s_{\beta i}$ on depth showed that there was a weak inverse relationship (data not shown). The R^2 value of the linear regression of depth vs. of $s_{\beta i}$ was 0.202 and differed statistically significantly from zero (p=0.001).

C.3 Correcting the average water content for the persistence

Assume that *K* probes of total *N* are working at one depth. The average water content over working probes at this depth for the sampling time *j* is

$$\hat{\theta}_j = \frac{1}{K}\sum_{i=1}^{K}\theta_{ij} \tag{C3}$$

The true average water content is

$$\overline{\theta}_j = \frac{1}{N}\left(\sum_{i=1}^{K}\theta_{ij} + \sum_{i=K+1}^{N}\theta_{ij}\right) \tag{C4}$$

where the first term in parentheses includes data from measurements, and the second includes unknown values that would come from malfunctioning probes should they work. The idea of the technique is to replace the unknown values with their estimates. The estimation consists in replacing the actual value of β_{ij} with the average value of β_{ij} over the period when all probes had worked, $\overline{\beta}_i$

$$\theta_{ij} \approx \overline{\theta}_j \overline{\beta}_i \tag{C5}$$

where i denotes a number of malfunctioning probe ($i = K+1, K+2, .., N$). Values of $\overline{\beta}_i$ are given in Table C1.
Substituting θ_{ij} from Eq. (C5) into Eq. (C4) and using Eq. (C3), one has

$$\overline{\theta}_j \approx \frac{1}{N}\left(K\hat{\theta}_j + \overline{\theta}_j \sum_{i=K+1}^{N} \overline{\beta}_i \right) \tag{C6}$$

Rearranging Eq. (C6) leads to

$$\overline{\theta}_j \approx \frac{K\hat{\theta}_j}{N - \sum_{i=K+1}^{N} \overline{\beta}_i} \tag{C7}$$

The sum of $\overline{\beta}_i$ values is always equal to N because $\sum_{i=1}^{N} \beta_{ij} = \frac{\sum_{i=1}^{N} \theta_{ij}}{\overline{\theta}_j} = \frac{N\overline{\theta}_j}{\overline{\theta}_j} = N$ for any j.

Therefore the denominator in the equation (C7) is equal to $\sum_{i=1}^{K} \overline{\beta}_i$. This leads to the following equation to estimate $\overline{\theta}_j$ by correcting $\hat{\theta}_j$:

$$\overline{\theta}_j \approx \frac{K\hat{\theta}_j}{\sum_{i=1}^{K} \overline{\beta}_i} \tag{C8}$$

Only measured values are included in this equation.

To test the correction technique based on Eq. (C8), we performed simulation experiments. We used all hourly measurements over the period when all 12 probes worked at the 15-cm depth. First, we computed values of $\overline{\beta}_i$ and values $\overline{\theta}_j$. The latter were dubbed 'true average water contents.' Then we randomly removed several probes for each sampling time and estimated the average water content using two methods: (a) the average water content over remaining probes assumed to be working in simulations (i.e., $\hat{\theta}_j$ according Eq. (C3)) , and (b) the corrected value of the average water content according to Eq. (C8). We removed up to eight probes in the experiment I (Fig. C1a and C1b), up to four probes in the experiment II (Fig. C1c and C1d), and one or two probes in the experiment III. As expected, as less probes are removed, the average water content over working probes is a better estimate of the true average (Fig. C1a, C1c, and C1e). Correction according to Eq. (C8) consistently gave much better estimate of the true average $\overline{\theta}_j$ than the average over working probes $\hat{\theta}_j$. The improvement with Eq. (C8) depended on the total number of working probes. The smaller the number of working probes the more efficient was the correction.

When the correction with Eq. (C8) was applied to actual measured water contents during the whole observation period of 360 days, it made substantial changes in some of estimated layer-average water contents. Maximum values for the corrections of the average layer-average water contents were 0.016, 0.013, 0.008, 0.014 and 0.014 v/v at the depths of 15 cm, 35 cm, 55 cm, 75 cm, and 95 cm, respectively. Ranges of observed

layer-average water contents $\hat{\theta}_j$ at the same depths were 0.089, 0.062,0.037, 0.046, and 0.040 v/v, respectively. The corrections were relatively more important at larger depths because of narrower ranges of observed water contents.

C.4 Discussion

The temporal persistence in water regimes was well-expressed at all five studied depths (Fig. 24 and Table C1). Texture, organic matter contents, soil structure, and the number of channels formed by roots, worms and other organisms were mentioned as leading static factors that affect spatial variations of soil water contents (Reynolds, 1970). Vachaud et al. (1985) related the variability of soil water contents and persistence in soil water content distributions to the variations in soil texture. These considerations could be pertinent to this work, although we do not have direct data to pinpoint the specific source of the persistence. In our research area, a study of water retention along a 30-m trench in the same soil at the adjacent site (Mallants et al., 1996) showed substantial small-scale variability. Data on saturated volumetric water content θ_s from this work are shown in Fig. 22. Values of θ_s were measured at 5-cm long and 5.1-cm diameter cores, i.e. at the scale comparable with the scale of TDR measurements. The variability of saturated water contents is comparable with variability in measured water content, indicating that the variations in water contents measured with TDR are realistic. We hypothesized that differences in soil structure can be responsible for those variations. The dependence of water retention on size of soil aggregates has been demonstrated in several studies, i.e. Wittmuss and Mazurak (1958), Tamboli et al. (1964) Amemiya (1965), Chang (1968), and Guber et al. (2002). Such dependence can cause differences in water contents at the same depth if soil matric potentials do not vary much across locations at this depth. This might contribute to the temporal persistence in water contents that we observed.

The proposed approach relies on the temporal persistence. Alternatively, one could try to estimate data for the malfunctioning probes using an interpolation technique, for example using kriging as suggested in spatiotemporal geostatistics (Christakos, 2000). We made an attempt to apply it, and found unsatisfactory results for probes positioned close to the corners of the spatiotemporal domain.

The proposed correction techniques worked reasonably well, mostly because the probability distribution functions of relative water contents were narrow, and the average relative water content was a good approximation for the whole distribution. The wider the distribution the less persistence is observed, and the less useful the proposed correction can be. Temporal trends in values of relative water contents make their distributions wider. Significant temporal trends were observed only for four probes at the 15-cm depth. This corresponds to the conclusion of Cassel et al. (2000) about higher persistence in deeper horizons. These authors attributed such dependence on depth to the root activity. Plant roots were not active in this work. The weakest time persistence at the shallowest observation depth in this work seemed to be more in line with results of Zhang and Berndtsson (1988) and Hupet and Vanclooster (2002) who had documented weaker temporal persistence during dry periods compared with wet ones.

Temporal persistence provides an opportunity to decrease the uncertainty of estimates of spatially averaged water contents. Indeed, if probe readings at a given depth are thought to be independent random values at any measurement time, the differences between those readings will be incorporated in the standard deviation of the average water content at this depth. However, in case of persistence, the individual probe readings are not independent. Therefore, only the variability of relative water contents has to be included in the estimate of the variability of the average soil water content. The variability of relative water contents is much smaller than variability of individual probe readings. Consequently, much smaller values of the standard deviation of the average water content should be expected. One effect of that is an improvement in accuracy of soil water balance computed using average soil water contents at several depths, because the average water contents will be less uncertain (will have smaller variability estimates) if temporal persistence in soil water content distributions is present and taken into account. Another advantage of establishing the temporal persistence in soil water content distributions is allowing removal some of the probes and having their measurements reconstructed from remaining probes. The temporal persistence means a temporal stability of spatial variability in soil water contents that may be useful in estimating uncertainty of measurements with small number of probes. Correct treatments of those topics are outside of the scope of this note, but they seem to be important to be explored.

In summary, the temporal persistence of water content patterns in soil profiles has been observed at a small scale. A correction using temporal persistence has been suggested that can be important in estimating layer-averaged water contents and their uncertainty in case of temporary malfunctioning equipment.

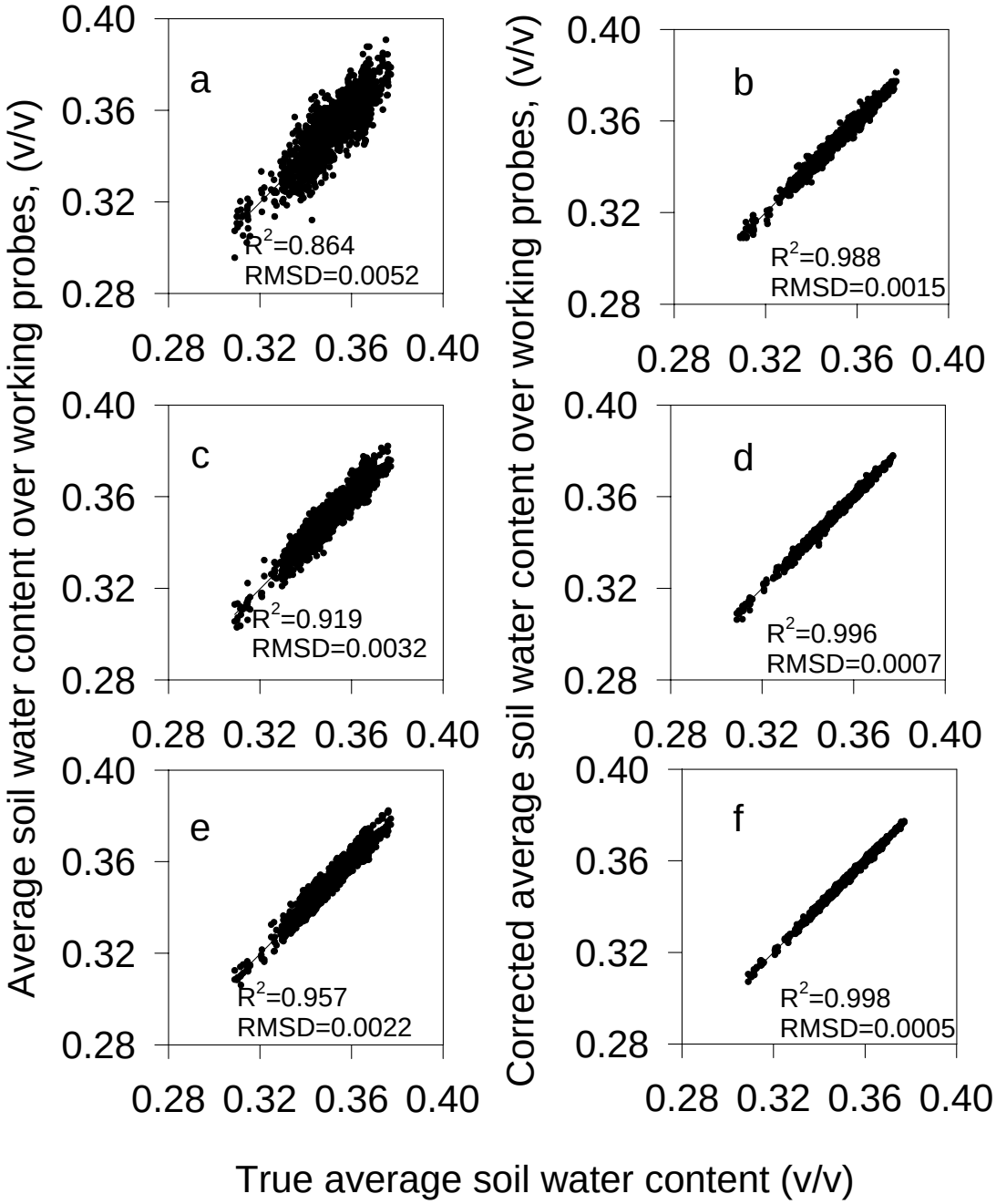

Figure C1. Simulated effect of the correction for temporal persistence on the relationship between actual and estimated layer-average water contents at the depth 15 cm; a, b – from 1 to 8 probes removed, c, d – from 1 to 4 probes removed, e, f – from one to two probes removed; a, b, c – average water content over remaining probes before the correction application, d, e, f - average water content over remaining probes after the correction application. Values of R^2 are the determination coefficients, and values of the RSMD are the root-mean square differences of the linear regression actual vs. estimated average water contents.

Table C1. Statistics of individual TDR probes

i	1*	2*	3	4*	5*	6*	7*	8	9*	10*	11*	12*
dnd	131	0	198	2	8	0	0	0	0	90	132	61
$\overline{\beta}i$	0.93	1.04	1.02	0.85	1.06	0.94	1.06	0.95	0.95	1.04	1.09	1.07
CV_i, %	2.0	1.2	1.2	1.1	1.2	1.9	1.6	2.7	1.9	1.1	1.4	2.1
i	13	14	15*	16*	17	18*	19*	20*	21*	22*	23*	24*
dnd	114	0	273	0	0	0	0	0	0	93	212	61
$\overline{\beta}i$	1.01	1.00	1.03	0.96	0.99	1.03	0.89	0.95	0.95	1.04	1.07	1.07
CV_i, %	1.1	0.9	1.3	0.8	0.7	0.7	1.1	0.8	1.1	0.6	0.6	0.8
i	25*	26	27*	28*	29*	30	31	32*	33	34*	35	36*
dnd	0	301	212	0	0	0	0	116	2	2	127	129
$\overline{\beta}i$	0.93	0.99	0.96	0.95	1.02	1.00	1.02	1.02	1.01	1.02	1.01	1.07
CV_i, %	0.9	1.0	0.5	0.7	0.7	0.7	1.5	0.8	0.9	0.8	2.0	0.8
i	37*	38	39*	40*	41*	42*	43	44*	45	46*	47*	48*
dnd	0	284	0	0	0	0	0	166	109	0	60	178
$\overline{\beta}i$	0.93	1.01	0.97	0.90	1.05	1.02	1.01	1.07	1.01	1.03	1.07	0.93
CV_i, %	1.3	0.6	0.6	0.6	0.6	0.6	0.8	2.4	0.6	0.6	0.8	1.2
i	49*	50*	51*	52*	53	54*	55*	56*	57	58	59*	60*
dnd	0	254	0	0	0	0	0	42	41	1	62	130
$\overline{\beta}i$	1.03	1.05	1.02	1.02	1.00	1.05	0.84	1.02	1.01	1.00	1.02	0.96
CV_i, %	0.8	1.3	0.7	0.6	0.7	0.9	1.3	0.8	0.8	0.7	0.6	0.9

Probe number *i* according Fig. 1, *dnd* = the total number of days with no data, $\overline{\beta}i$ = mean ratio of the probe "*i*" water contents to the depth-average water contents, CV_i =coefficient of variation of the ratios of the probe "*i*" water content to the depth-average water contents. Values $\overline{\beta}i$ CV_i were computed from hourly data. Probes marked with the asterisk showed water contents that were different from the depth-average water contents in more than 95% of cases.

Appendix D. Estimating matric unsaturated hydraulic conductivity in deep soil layers

We assumed that (a) no preferential flow occurred during the drying period, (b) the dependence of the unsaturated hydraulic conductivity on water content could be assumed the same within the depth range from 65 to 85 cm, (c) the linear interpolation was applicable to water contents. A simple mass balance computational scheme shown in Fig. D1 was used.

The Darcy-Buckingham law and the mass conservation equations were used to calculate interlayer water fluxes. Darcy equations were written for interlayer flux at depths 65 cm and 85 cm as

$$q_{65} = k_{65}\left(\frac{h_{75} - h_{55}}{75 - 55} + 1\right); \qquad q_{85} = k_{85}\left(\frac{h_{95} - h_{75}}{95 - 75} + 1\right), \tag{D1}$$

where q is interlayer water flux, cm day^{-1}; Kw is unsaturated hydraulic conductivity, cm day^{-1}; h is the capillary pressure, cm, subscripts indicate depths.

The mass conservation in the soil layer 55-75 cm is:

$$\Delta\theta_{75}\cdot 20 = \Delta t(q_{65}-q_{85}), \tag{D2}$$

where $\Delta\theta_{75}\cdot 20$ is daily water storage change at 20 cm layer; Δt is time interval equal to 1 day.

Combining equation (1) and (2), one obtains:

$$\Delta\theta_{75} \cdot 20 / \Delta t = k_{65}\left(\frac{h_{75} - h_{55}}{20} + 1\right) - k_{85}\left(\frac{h_{95} - h_{75}}{20} + 1\right) \tag{D3}$$

Unsaturated hydraulic conductivity values k_{65} and k_{85} are unknown variables in the equation (D3).

We assumed the Averianov's relationship between Kw and water content θ for depth of 65 cm and 85 cm:

$$k = K_{sat}\left(\frac{\theta - \theta_r}{\theta_s - \theta_r}\right)^n, \tag{D4}$$

where K_{sat} is the saturated hydraulic conductivity; θ_s, θ_r are the saturated water content and the and residual water content, respectively; n is exponent. The residual water content was set at to 0.001 cm^{3}cm^{-3}.

Coefficients k_{65} and k_{85} in equation (3) refer to interlayer water content at 65cm and 85 cm depths, and water contents were linearly interpolated to those depths.

Equation (D3) was fitted to the water storage changes at layer 65-85 cm shown in Fig. 28. Fitted curve described observed data adequately. The root mean square error of fitting was equal 0.0346 cm, whereas standard deviation of measured water storage changes was equal 0.0355 cm. Estimated values of parameters were K_{sat} = 0.024 cm·day^{-1}, θ_s = 0.36 cm^{3}cm^{-3}, and n = 14.3.

This function was used for calculation of water flux at the depth of 105 cm as:

$$q_{105} = k_{85}\left(\frac{h_{95} - h_{75}}{20} + 1\right) - \Delta\theta_{95} \cdot 20 / \Delta t \qquad \text{(D5)}$$

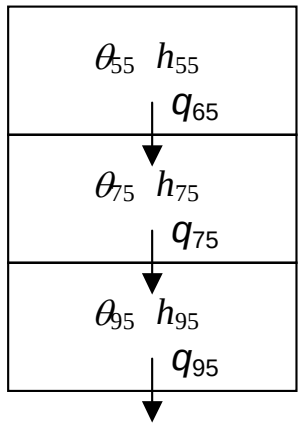

Figure D1. Water flow scheme to estimate the matric unsaturated hydraulic conductivity.

Appendix E. MWBUS: one-dimensional soil water flow model

The MWBUS (model of water balance of unsaturated soils) model has been developed to provide a simple and robust description of water flow in layered soils with the daily time step**.** The model functioning is pictured in Fig. E1.

Soil is divided in layers of thickness Δz_i, i=1,2,…, M. A downward water flow from a layer "i"can occur if water content θ_i exceeds the value of water field capacity $\theta_{FC,I}$ of this layer. The value of flux q_i from the ith soil layer is calculated as:

$q_i = min(q_1,q_2)$

$$q_1 = q_{i-1} + (\theta_i - \theta_{FC,i})\Delta z_i/\Delta t \qquad \text{(E1)}$$

$q_2 = min(K_{sat,i}, K_{sat,i+1})$

where q_{i-1} is the incoming water flux to ith layer, $K_{sat,i}$ is the saturated hydraulic conductivity of ith layer, Δt is the time interval set to one day.

The value of flux q_0 is calculated from the rainfall rate R as:

$$q_0 = min(R, K_{sat,1}) \qquad \text{(E2)}$$

Fluxes q_i, I=1,2,…,M are calculated recursively top-to-bottom. Fluxes may be corrected beginning from any internal layer to account for complete saturation of the layer.

The layer water contents θ_i decrease due to evaporation and/or loss to the upward flow. The exponential equation describes the decrease in θ with time (Suleiman & Ritchie, 2003):

$$\theta_i = \theta_{r,i} + (\theta_{i,initial} - \theta_{r,i})\exp(-t/\tau) \qquad \text{(E3)}$$

where τ is the resistance parameter, $\theta_{r,i}$ is the residual water content. The daily change in water content is

$$\Delta\theta_i = -\frac{\theta_i - \theta_{r,i}}{\tau}$$

The value of τ increases with the depth z. Suleiman and Ritchie (2003) have found that a power function gives a good fit to relationships between τ and z for several soils and evaporation scenarios:

$$\tau = \tau_0 + az^b \qquad \text{(E4)}$$

where τ_0 is the value of resitance parameter at depth z=0, a and b are parameters of the evaporation model.

The simulated daily evaporation rate E is:

$$E = \sum_{i=1}^{n} \Delta\theta_i \Delta z_i \qquad \text{(E6)}$$

The number of soil layers n subject to evaporation is adjusted to make the value of E equal to the evaporation in the model input. An adjustment over those layers is made to provide the exact equality.

The set of the MWBUS model parameters includes the saturation water contents $\theta_{s,i}$, the field capacity values $\theta_{FC,i}$, the residual water content values $\theta_{r,i}$, and the saturated hydraulic

conductivity values $K_{sat,I}$ for each layer, and the parameter of evaporation/upward loss submodel τ_0, a, and b.

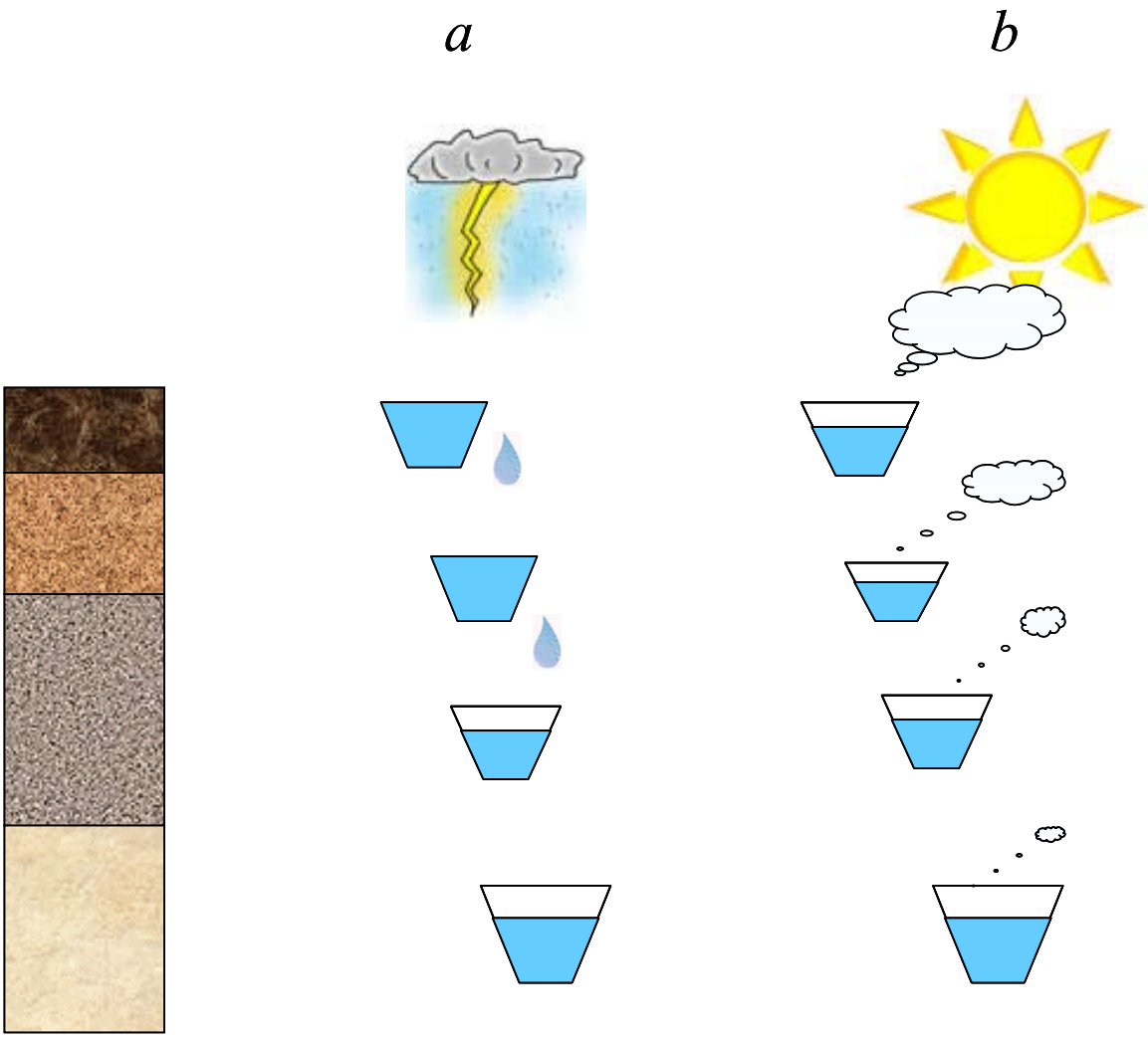

Figure E1. Schematic of the modeling soil water flow with the MWBUS model; a- infiltration regime, a downward water flow from a layer to the layer below occurs if the layer water content exceeds the field capacity of this layer; b – evaporation regime; layers lose water exponentially in time; the characteristic time of the water loss increases with the depth.

NRC FORM 335
(9-2004)
NRCMD 3.7

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER
(Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/CR-6884

2. TITLE AND SUBTITLE

Model Abstraction Techniques for Soil-Water Flow and Transport

3. DATE REPORT PUBLISHED

MONTH	YEAR
December	2006

4. FIN OR GRANT NUMBER

Y6724

5. AUTHOR(S)

Yakov A. Pachepsky 1, Andrey Guber 2, M.T. van Genuchten 1, T.J. Nicholson 3, R.E. Cady 3, J. Simunek 2 and M.G. Schaap 2

1 U.S. Department of Agriculture/Agricultural Research Service; 2 University of California at Riverside; 3 U.S. Nuclear Regulatory Commission/Office of Nuclear Regulatory Research

6. TYPE OF REPORT

Technical

7. PERIOD COVERED *(Inclusive Dates)*

September 2002 - August 2005

8. PERFORMING ORGANIZATION - NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Environmental Microbial Safety Laboratory
Agricultural Reseach Service
U.S. Department of Agriculture
Beltsville, MD 20705

9. SPONSORING ORGANIZATION - NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

Division of Fuel, Engineering & Radiological Research
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555

10. SUPPLEMENTARY NOTES

T.J. Nicholson, NRC Project Manager

11. ABSTRACT *(200 words or less)*

This report describes the methodology of model abstraction in subsurface hydrology. Model abstraction is defined as the methodology for reducing the complexity of a simulation model while maintaining the validity of the simulation results with respect to the question that the simulation is being used to address. The need in model abstraction may stem from the need to improve the reliability and reduce uncertainty of simulations, to make the modeling and its results more explicable and transparent to the end users, and to enable more efficient use of available resources in data collection and computations. A comprehensive review of model simplification techniques developed in subsurface hydrology is included. Abstractions of both model structure and model parameter determination are described. A systematic and objective approach to model abstraction is outlined. A case study is presented that is designed to illustrate how model abstraction can affect performance assessment of contaminant migration at a relatively humid site. Although the model abstraction methodology is generic, it is designed to be of practical use to NRC licensing staff in their review of the performance assessment of decommissioning sites and waste disposal facilities. The model abstraction process provides a systematic approach to understand the adequacy of model simplification, and facilitates communication and transparency of the model to regulators, stakeholders, and the general public.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

flow and transport modeling
metamodeling
model abstraction
model hierarchy
model simplification
model structure
parameter estimation
pedotransfer function
soil water
subsurface hydrology
uncertainty

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page)

unclassified

(This Report)

unclassified

15. NUMBER OF PAGES

16. PRICE

PRINTED ON RECYCLED PAPER

Printed
on recycled
paper

Federal Recycling Program