UNIVERSITY OF CALIFORNIA
RIVERSIDE

Development of Physically-Based and Data-Driven Models to Predict Contaminant Loads in Runoff Water From Agricultural Fields

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Sciences

by

Jing Liang

September 2018

Dissertation Committee:
Dr. Jiří Šimůnek, Chairperson
Dr. Scott A. Bradford
Dr. Hoori Ajami
Dr. Amir Haghverdi

The Dissertation of Jing Liang is approved:

_____

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to start by thanking my two supervisors, Dr. Jiří Šimůnek and Dr. Scott Bradford. This dissertation work would not have been possible without their guidance, support and encouragement. Dr. Jiří Šimůnek and Dr. Scott Bradford provided me with the opportunity to deepen my knowledge bases and responded to my questions and queries promptly. I appreciate their willingness to work on a tight schedule and help me graduate in a timely manner while pursuing a meaningful topic of my choosing. Thank you for giving me the perseverance to never give up on my studies. I would also like to thank all of my dissertation committee and qualifying exam committee members, Dr. Hoori Ajami, Dr. Amir Haghverdi, Dr. Laosheng Wu, Dr. David Crowley, and Dr. Brett Sanders for their valuable comments and suggestions.

Looking back over my own path in the graduate school, I have been lucky to work with motivated collaborators and colleagues. Special thanks to Anne Hartmann, Sarah Helalia, Shirley Yang, Dr. Salini Sasidharan, Dr. Fábio Mallmann, Dr. Giuseppe Brunetti, Dr. Stathis Diamantopoulos, Dr. Hirotaka Saito, Dr. Hongxing Li, Dr. Miaoyue Zhang, Dr. Yusong Wang, Dr. Xiaojun Shen, Dr. Everton Pinheiro, Jinsong Ti , Ye Xu, and all other friends and mentors in the Department of Environmental Sciences at UCR and in the USDA Salinity Laboratory.

I must express my gratitude to my family getting me through and always standing by my side. My mother, my brother, sister in law, mother in law, father in law, grandmother in law, uncle Alan, auntie Grace, cousin Issac, I can't thank you enough for everything you

have done for me. I'd like to present my sincere thankfulness to my dear father, who died three years ago, for being great role in my life. I wish to thank my loving and supportive husband, Wenzhe Li, who has been constant source of support and encouragement. Finally, to my beloved daughter Zhien Li, you are the best thing that ever happen to me!

This research was partially supported by the USDA, ARS, National Program 211 and the Western Riverside County Agricultural Coalition (WRCAC).

ABSTRACT OF THE DISSERTATION

Development of Physically-Based and Data-Driven Models to Predict Contaminant Loads in Runoff Water From Agricultural Fields

by

Jing Liang

Doctor of Philosophy, Graduate Program in Environmental Sciences
University of California, Riverside, September 2018
Dr. Jiří Šimůnek, Chairperson

Contaminants in water at the soil surface can be rapidly transported to streams or locations of surface water storage by runoff. This has been reported to be the primary transport route for contaminants on sloping fields and hillslopes. Generally, contaminants are transported faster and in larger amounts with runoff, and much slower and in smaller amounts with infiltration. Runoff transports are strongly affected by surface topography, soil properties, vegetation, and weather. The risk of transport is also highly dependent on the contaminant load at the soil surface, interactions with soil particles, and transformations. An understanding of and the ability to predict processes that influence the transport and fate of contaminants in runoff water are therefore needed to assess and mitigate risks of contamination of surface water supplies on human health.

Physically-based, spatially-distributed models have the potential to be an efficient tool to examine and optimize the removal of contaminants from agricultural runoff through land-use changes and best management practices. In this research, the existing subsurface version of HYDRUS-1D was adapted to simulate uniform or physical

nonequilibrium flow and reactive solute transport processes during runoff at the soil surface. The numerical results obtained by the new model produced an excellent agreement with an analytical solution for the kinematic wave equation. Additional model tests further demonstrated the applicability of the adapted model to simulate the transport and fate of many different solutes (non-adsorbing tracers, nutrients, pesticides, microbes, and sediments) that undergo equilibrium and/or kinetic sorption and desorption, and first- or zero-order reactions.

Along with PBMs, data-driven models are becoming increasingly popular for describing the behavior of hydrological and water resources systems since these models can be used to complement or even replace physically based-models when there is a lack of required data. Here we propose a new data-driven model as an alternative to a physically-based overland flow and transport model. Several machine learning techniques including Linear Regression (LR), k-Nearest Neighbor regression (kNN), Support Vector Machine with linear (SVM-L) and non-linear (SVM-NL) kernels, and Deep Neural Networks (DNN) (Neural Networks with multiple hidden layers) were explored to find input - output functional relations. The results indicated that the Deep Neural Network (DNN) model with two hidden layers performed the best among selected data-driven models. This DNN model accurately predicted runoff water quantity over a wide range in parameters. It also predicted well runoff water quality for near-equilibrium solute transport over a wide range in parameters.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Background

California produces abundant food supplies for domestic markets and has been recognized as the nation's top agricultural state. However, agricultural activities can discharge a wide range of contaminants into subsurface and surface water resources when improperly managed. Farming and ranching operations produce a variety of agricultural pollutants (e.g., sediment, nutrients, pathogens, pesticides, metals, and salts) that can lead to impairments of local and far-field water quality [Wilcock, 1986; Wilcock et al., 1999]. These diffuse nonpoint sources can directly harm ecosystem and watershed water quality, and adversely affect drinking water supplies. An estimated global use of agricultural pesticides is 3 billion kg each year [Pimental et al., 2004]. Leaching and runoff of these agricultural chemicals can contaminate surface and groundwater resources that serve as the primary source of drinking water throughout North America and Canada [Larson et al., 1997; Nowell et al., 1999; Allen et al., 1993; Gustafson, 1993]. In addition to utilizing pesticides and chemical fertilizers, farmers frequently use the Concentrated Animal Feeding Operations (CAFOs) biomass as a surface applied fertilizer to agricultural fields. These wastes contain high concentrations of salts, organics, nutrients, heavy metals, hormones and antibiotics, and pathogenic microorganisms [Bradford et al., 2008; California Regional Water Quality Control Board, 2013]. Accumulated contaminants at the land surface can be transported to water resources faster and in larger amounts with surface runoff and much slower and in smaller amounts with subsurface flow. Surface runoff from agricultural fields, which may substantially contribute to pollution discharges, has therefore been identified as one of the major causes for water impairment

in agricultural settings [Carpenter et al., 1998; USGS, 1999; Tyrrel and Quinton, 2003]. Understanding of the processes that influence the transport and fate of contaminants in runoff water and the ability to predict them are therefore required to assess and mitigate risks of contamination of surface water supplies.

Surface runoff is typically initiated when the inflow rate (precipitation, snow melt, or runoff) exceeds the soil infiltration capacity and fills surface depressions (Hortonian flow) [Horton, 1933] or when inflow from upland occurs on saturated soil [Dunne and Black, 1970]. A certain fraction of water and contaminants infiltrates into the ground and is stored in the soil profile and underground aquifers. These losses of water delay the initiation and travel time of surface runoff. Both runoff and subsurface transport occur simultaneously and are strongly affected by surface topography, soil properties, vegetation, soil management, and weather.

However, overland and subsurface water flow and contaminant transport are rarely uniform. At the local scale, spatially varying roughness, vegetation, and microtopography influence the distribution of shear stress and create hydrologically active and passive, or relatively immobile flow regions. Surface active regions for overland flow can be formed in the field by rills or connected networks of microdepressions, which route overland water and pollutant fluxes on the soil surface, and where the velocity can be much higher than the average flow [Dunkerley, 2003; Chen et al., 2013]. These hydraulically active regions pose great risks of contamination of water resources because of the presence of large water and contaminant fluxes and reduced time for infiltration and contaminant decay, transformation, and/or interphase

mass transfer (e.g., interactions with soil particles). Surface passive or relatively immobile regions are formed in depressions or in areas with obstructed flow. These regions retain water and pollutants, inhibit overland flow, or cause shallow, slow-moving flow and exchange water and pollutants with surface active zones. These spatially varying surface characteristics generate nonequilibrium overland flow and transport processes.

Obtaining monitoring data using field investigations is very time consuming and expensive, and associated with many experimental difficulties. Furthermore, simple field observations may be difficult to interpret to obtain a complete picture of potential contaminant transport routes and mechanisms, to extrapolate findings to other environmental conditions and climates, and to repeat. Alternatively, mathematical models have the potential to be an efficient tool to examine and optimize the removal of contaminants from overland flow through testing various implementation scenarios of land use change and best management practices [Kirkby and Beven, 1979; Park et al., 1994; Borah and Bera, 2003; Roz, 2011]. Many publicly available overland pollutant transport models have been developed [Beven and Kirkby, 1979; Haith and Shoemaker, 1987; Bathurst and Connell, 1992; Flanagan et al., 1995; Arnold et al., 1998]. These modes can be grouped from simple to complex into three categories based on the model structure: empirical black-box models, lumped conceptual models, and distributed physically-based models. Empirical models, sometimes called data-driven models, are entirely lacking the explicit description of internal physical processes which are involved in the overland flow and transport. Some examples of empirical models are the SCS-

Curve Number [USDA, 1986], regression equations, and machine and deep learning models. These models quantify input–output functional relations for complex rainfall-runoff systems without any reference to the involved physical processes [Hsu et al., 1995; Minns and Hall, 1996; Dawson and Wilby,1998; Dibike et al.,1999; Abrahart and See, 2000]. Lumped conceptual models are designed to approximate physical hydrological mechanisms by connecting subprocesses in the overall hydrological process. These models are represented by reservoir storage models and simplified equations of physical hydrological processes [Devi et al., 2015; Vaze, 2012]. These models have been commonly used for contaminant transport in overland flow but may be inaccurate because they ignore processes acting on water and chemicals at the soil surface (e.g., Soil and Water Assessment Tool [SWAT; Sadeghi and Arnold, 2002], Hydrologic Simulation Program Fortran [HSPF; Donigian et al., 1995], Integrated Nitrogen CAtchment model [INCA; Whitehead et al., 1998; Wade et al., 2002], and COLI [Walker et al., 1990]).

Physically-based, distributed models explicitly account for main hydrologic and contaminant transport processes using mathematical descriptions. Spatial and temporal variations can be incorporated into physically-based models, which makes them more realistic. As mentioned above, the spatial variation of surface roughness, microtopography, vegetation, and soil hydraulic properties strongly influence overland flow and transport processes. It is necessary to incorporate spatial variations of land surface characteristics to accurately predict the hydraulic and hydrologic behavior of overland flow, soil erosion, and contaminant movement. There are many existing two-dimensional and three-dimensional mechanistic models that can consider local scale

parameters (e.g., Parallel Flow [ParFlow; Kollet and Maxwell, 2006], CATchment Hydrology [CATHY; Camporese et al., 2010], HydroGeoSphere [HGS; Therrien et al., 2012], and OpenGeoSys [OGS; Kolditz et al., 2012]). These distributed models use the actual varying microtopography based on Digital Elevation Maps (DEM), land use imagery from satellites, and gridded precipitation as an input. However, they are computationally demanding, and extra effort is needed for collecting the required data. Difficulties in collecting distributed input data and large computational time are possible reasons why distributed models are not more widely used compared to lumped models [Rinsema, 2014]. Over the past several decades, the one-dimensional overland flow models usually use the 1D kinematic wave or 1D diffusion wave equations, which describe surface runoff as sheet flow with a uniform depth and velocity across the homogeneous plane surface (e.g., the KINematic runoff and EROSion model [KINEROS2; Miller et al., 2007; Goodrich et al., 2012; Kennedy et al., 2013] or the Precipitation–Runoff Modeling System [PRMS; Leavesley et al., 1983]). These models neglect local-scale variations in parameters by considering simplified smooth surfaces with a global roughness coefficient and a constant average slope. As a result, these models may have failed to accurately describe the distribution of water depths, fluxes, and contaminant loads [Dunne and Dietrich, 1980; Smith et al., 2011; Cea et al., 2014; Bradford et al., 2015]. However, these models may provide an excellent approximation of overland flow when calibrated to real data but may not be adequate to describe overland contaminant transport that is affected by local-scale parameters. Hence, accurate, physically-based modeling of nonequilibrium water flow and solute transport still

remains a challenge in the field of surface hydrology, and there is currently no widely accepted physically-based model available to simulate overland transport processes and all kinds of reactive chemicals and sediments.

The HYDRUS-1D software [Šimůnek et al., 2016] is a popular numerical computer code for solving water flow and reactive transport in variably-saturated porous media. This code solves the Richards and convection–dispersion equations, which simulate water flow and contaminant transport in the subsurface, respectively. Recent studies indicate that the governing overland flow equation can be written in a similar mathematical formulation as the Richards equation [Weill et al., 2009; Hromadka and Lai, 1985; Panday and Huyakorn, 2004]. The central hypothesis of this dissertation work is that the existing numerical schemes of HYDRUS-1D can also be used to solve the 1D-diffusion wave equation, which usually describes overland flow. Then the existing HYDRUS-1D solute transport models coupled with water flow can be used for simulating advective and dispersive transport of many different kinds of contaminants (e.g., salts, nutrients, pesticides, and microbes) in runoff water by properly selecting the sorption (linear and nonlinear equilibrium sorption, kinetic sorption and desorption on multiple sites), and zero- and first-order production and decay terms [Ahuja, 1986; Wallach et al. 1988; Wallach and Shatai, 1992]. In this case, the new overland flow and transport model can be build based on the HYDRUS-1D codes, which will allow many existing HYDRUS users to easily operate the new models while having access to various features of the HYDRUS software, such as a well-designed GUI, user-friendly graphical input, parameter optimization, and the generation of random scale factors. In addition, the

existing HYDRUS-1D code includes a variety of physical and chemical equilibrium and nonequilibrium models to describe water flow and solute transport processes in the subsurface [Šimůnek and van Genuchten, 2008]. These models form a hierarchical system of models of increasing complexity accounting for both physical (and chemical) equilibrium and nonequilibrium flow, including (i) a uniform flow model, (ii) a dual-porosity (mobile–immobile water) model, (iii) a dual-permeability model, and (iv) a dual-permeability model with immobile water. These conceptual models for subsurface can be adapted to account for both equilibrium and nonequilibrium overland flow and reactive solute transport.

In addition, a physically-based overland flow model coupled with the mass conservation equation for sediments can generate a physically-based soil erosion model. This indicates that the HYDRUS-1D model has a potential to be modified to simulate soil erosion and transport. Many existing soil erosion and sediment yield and transport models have their own capabilities and limitations [e.g., Williams and Berndt, 1977; Wischmeier and Smith, 1978; Knisel, 1980; Singh et al., 1982; Abbott et al., 1986; Storm et al., 1987; Renard et al., 1994; Hanley et al., 1998; Nearing et al., 1999; Woodward, 1999; De Jong et al., 1999; Johnson et al., 2000; Van Oost et al., 2000; Tucker et al., 2001; Ziegler et al., 2001; Torri et al., 2002; Katlin et al., 2003; Goodrich et al, 2012]. The most widely used physically-based soil erosion models (e.g., the Kinematic Runoff and Erosion Model [KINEROS: Woolhiser et al., 1990], the European Soil Erosion Model [EUROSEM: Morgan et al., 1998], etc.) commonly divide soil erosion into rill and interrill erosion. Interrill erosion is sheet erosion that is caused by sheet flow

removing soil in uniform thin layers by forces of raindrops and overland flow whereas the rill erosion removing soil by concentrated water. However, natural systems are more complex and the ignorance of variable spatial distribution of water fluxes may cause significant errors in predicted sediment yield, transport, and erosion. The sheet erosion models can be extended to non-equilibrium sediment transport models by adapting physical equilibrium and nonequilibrium flow concepts.

The use of physically-based models, such as HYDRUS, may require significant computational time as the spatial and temporal scale of the considered problem increases. For example, to simulate runoff and reactive contaminant transport at larger watershed scales can take a much longer time. In contrast to physically-based models, data-driven models are based on functional relationships between input and output variables. Artificial intelligence approaches, such as Machine Learning (ML) and Deep Learning (DL) techniques, are increasingly being used in data-driven models to quantify input-output functional relations for complex systems in hydrology [Hsu et al., 1995; Minns and Hall, 1996; Dawson and Wilby,1998; Dibike et al.,1999; Abrahart and See, 2000; Mjolsness and DeCoste, 2001; Govindaraju and Rao, 2001; Solomatine and Ostfeld, 2008; Roz, 2011; Remesan and Mathew, 2014; Bai et al. 2016; Karandish and Šimůnek, 2016; Fang et al., 2017]. Several ML algorithms for modeling rainfall-runoff processes include Linear Regression (LR), K-Nearest Neighbor regression (kNN), Feed-forward Artificial Neural Networks (ANN), and Support Vector Machine (SVM) models [Loague and Freeze, 1985; Karlasson and Yakowitz, 1987; ASCE, 2000ab; Lin et al., 2006; Nourani et al., 2009]. However, there are no existing studies to compare the performance

of these different machine learning techniques on developing relationships between water flow/contaminant transport in runoff and environmental variables

1.2 Research Objectives

The objective of this dissertation is to develop a comprehensive set of modeling tools for the investigation of many important research problems involving overland flow, reactive transport, and sediment transport processes. To achieve the overall objective of this dissertation, the following specific objectives are:

1. To adapt the existing version of HYDRUS-1D for subsurface processes to simulate similar uniform or physical nonequilibrium flow and reactive solute transport processes during runoff at the soil surface. The developed models will improve our ability to quantify natural flow and transport processes at the soil surface and to enhance our understanding of factors that cause deviations from sheet flow.
2. To test the ability of data-driven models to mimic physically-based models to predict surface runoff water quantity and quality in agricultural settings. Linear Regression (LR), k-Nearest Neighbor regression (kNN), Support Vector Machine with linear (SVM-L) and non-linear (SVM-NL) kernels, and Deep Neural Networks (DNN) (Neural Networks with multiple hidden layers) models will be evaluated on their ability to relate model inputs with outputs.
3. To develop a series of physically-based soil erosion models, which are able to simulate sheet flow, interrill erosion, and interrill erosion during derivations from sheet flow in agricultural settings.

1.3 References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O' Connell, P.E. and Rasmussen, J., 1986. An introduction to the European Hydrological System—Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system. *Journal of hydrology*, 87(1-2), 45-59.

Abrahart, R.J. and See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological processes*, 14, 2157-2172.

Ahuja, L.R., 1986. Characterization and modeling of chemical transfer to runoff. In *Advances in soil science*, 149-188, Springer.

Allen, H.E., Perdue, E.M. and Brown, D.S., 1993. *Metals in groundwater*. CRC Press.

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development 1. *JAWRA Journal of the American Water Resources Association*, 34(1), 73-89.

ASCE Task Committee on Application of Artificial Neural Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology–I: preliminary concepts. *Journal of Hydrologic Engineering*, *5*(2), 115-123.

ASCE Task Committee on Application of Artificial Neural Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology–II: hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124–137.

Bai, Y., Chen, Z., Xie, J. and Li, C., 2016. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology*, 532, 193–206.

Bathurst, J.C. and O'connell, P.E., 1992. Future of distributed modelling: the Systeme Hydrologique Europeen. *Hydrological Processes*, 6(3), 265-277.

Borah, D.K. and Bera, M., 2003. Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *Transactions of the ASAE*, 46(6), 1553.

Bradford, S.A., Headd, B., Arye, G. and Šimůnek, J., 2015. Transport of E. coli D21g with runoff water under different solution chemistry conditions and surface slopes. *Journal of Hydrology*, 525, 760-768.

Bradford, S.A., Segal, E., Zheng, W., Wang, Q., and Hutchins, S. R., 2008. Reuse of concentrated animal feeding operation wastewater on agricultural lands. *Journal of Environmental Quality*, 37(5), 97.

California Regional Water Quality Control Board Santa Ana Region, 2013. General waste discharge requirements for concentrated animal feeding operations (dairies and related facilities) within the Santa Ana Region. Order No. R8-2013-00001.

Cea, L., Legout, C., Darboux, F., Esteves, M. and Nord, G., 2014. Experimental validation of a 2D overland flow model using high resolution water depth and velocity data. *Journal of hydrology*, 513, 142-153.

Camporese, M., Paniconi, C., Putti, M. and Orlandini, S., 2010. Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research*, 46(2).

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N. and Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological applications*, 8(3), 559-568.

Dawson, C.W. and Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal*, 43(1), 47-66.

De Jong, S.M., Paracchini, M.L., Bertolo, F., Folving, S., Megier, J. and De Roo, A.P.J., 1999. Regional assessment of soil erosion using the distributed model SEMMED and remotely sensed data. *Catena*, 37(3-4), 291-308.

Devi, K.G., Ganasri, B.P. and Dwarakish, G.S., 2015. A review on hydrological models, *Aquatic Procedia,* 4, 1001–1007.

Dibike, Y.B., Solomatine, D. and Abbott, M.B., 1999. On the encapsulation of numerical-hydraulic models in artificial neural network. *Journal of Hydraulic research*, 37(2), 147-161.

Donigian Jr, A.S. and Beyerlein, D.C., 1993. Watershed modeling of agrichemicals in Walnut Creek, IA: a plan for preliminary assessment of contributions and impacts. In *Water quality modeling: Proceedings of the International Symposium*, American Society of Agricultural and Biological Engineers.

Dunne, T. and Black, R.D., 1970. An experimental investigation of runoff production in permeable soils. *Water Resources Research*, 6(2), 478-490.

Dunne, T. and Dietrich, W.E., 1980. Experimental study of Horton overland flow on tropical hillslopes. *Journal of Geomorphology*, 35, 40-59.

Fang, K., Shen, C., Kifer, D. and Yang, X., 2017. Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. *Geophysical Research Letters*, 44(21), 11-030.

Flanagan, D.C., Ascough, J.C., Nicks, A.D., Nearing, M.A., Laflen, J.M., Flanagan, D. and Nearing, M., 1995. Overview of the WEPP erosion prediction model. *Technical Documentation: USDA-Water Erosion Prediction Project (WEPP).*

Furman, A., 2008. Modeling coupled surface–subsurface flow processes: A review. *Vadose Zone Journal*, 7(2), 741-756.

Goodrich, D.C., Burns, I.S., Unkrich, C.L., Semmens, D.J., Guertin, D.P., Hernandez, M., Yatheendradas, S., Kennedy, J.R. and Levick, L.R., 2012. KINEROS2/AGWA: model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1561-1574.

Govindaraju, R.S. and Rao, A.R. eds., 2013. Artificial neural networks in hydrology (Vol. 36). *Springer Science & Business Media.*

Gustafson, D.I., 1993. Pesticides in drinking water. *Van Nostrand Reinhold.*

Haith, D.A. and Shoenaker, L.L., 1987. Generalized watershed loading functions for stream flow nutrients 1. *JAWRA Journal of the American Water Resources Association*, 23(3), 471-478.

Hanley, N., Faichney, R., Munro, A. and Shortle, J.S., 1998. Economic and environmental modelling for pollution control in an estuary. *Journal of Environmental management*, 52(3), 211-225.

Horton, R.E., 1933. The role of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, 14(1), 446-460.

Hromadka II, T.V. and Lai, C., 1985. SOLVING THE TWO-DIMENSIONAL DIFFUSION FLOW MODEL. In *Hydraulics and Hydrology in the Small Computer Age, Proceedings of the Specialty Conference*, 555-562.

Hsu, K.L., Gupta, H.V. and Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10), 2517-2530.

Johnson, B.E., Julien, P.Y., Molnar, D.K. and Watson, C.C., 2000. The two-dimensional upland erosion model CASC2D-SED 1. *JAWRA Journal of the American Water Resources Association*, 36(1), 31-42.

Kalin, L. and Hantush, M.M., 2003, Assessment of two physically-based watershed models based on their performances of simulating water and sediment movement". In *Conference Proceedings, 1st Interagency Conference on Research in the Watersheds (ICRW), Benson, AZ*.

Karandish, F., and Šimůnek J., 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of Hydrology,* 543, 892–909.

Karlsson, M. and Yakowitz, S., 1987. Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. *Water Resources Research*, 23(7), 1300-1308.

Kennedy, J.R., Goodrich, D.C. and Unkrich, C.L., 2012. Using the KINEROS2 modeling framework to evaluate the increase in storm runoff from residential development in a semiarid environment. *Journal of Hydrologic Engineering*, 18(6), 698-706.

Kirkby, M.J. and Beven, K.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, 24(1), 43-69.

Knisel, W.G., 1980. CREAMS: a field scale model for Chemicals, Runoff, and Erosion from Agricultural Management Systems [USA]. *United States. Dept. of Agriculture. Conservation research report (USA)*.

Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J.O., Fischer, T., Görke, U.J., Kalbacher, T., Kosakowski, G., McDermott, C.I. and Park, C.H., 2012. OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media. *Environmental Earth Sciences*, 67(2), 589-599.

Kollet, S.J. and Maxwell, R.M., 2006. Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7), 945-958.

Larson, S.J., Capel, P.D. and Majewski, M., 1997. Pesticides in surface waters: Distribution, trends, and governing factors. *CRC Press*.

Lin, J.Y., Cheng, C.T. and Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, *51*(4), 599-612.

Loague, K.M. and Freeze, R.A., 1985. A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research*, 21(2), 229-248.

Miller, S.N., Semmens, D.J., Goodrich, D.C., Hernandez, M., Miller, R.C., Kepner, W.G. and Guertin, D.P., 2007. The automated geospatial watershed assessment tool. *Environmental Modelling & Software*, 22(3), 365-377.

Minns, A.W. and Hall, M.J., 1996. Artificial neural networks as rainfall-runoff models. *Hydrological sciences journal*, 41(3), 399-417.

Mjolsness, E. and DeCoste, D., 2001. Machine learning for science: state of the art and future prospects. *Science*, 293(5537), 2051-2055.

Moore, J.A., Smyth, J.D., Baker, E.S., Miner, J.R. and Moffitt, D.C., 1989. Modeling bacteria movement in livestock manure systems. *Transactions of the ASAE*, 32(3), 1049-1053.

Morgan, R.P.C., Quinton, J.N., Smith, R.E., Govers, G., Poesen, J.W.A., Auerswald, K., Chisci, G., Torri, D. and Styczen, M.E., 1998. The European Soil Erosion Model (EUROSEM): a dynamic approach for predicting sediment transport from fields and small catchments. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group*, 23(6), 527-544.

Nearing, M.A., Simanton, J.R., Norton, L.D., Bulygin, S.J. and Stone, J., 1999. Soil erosion by surface water flow on a stony, semiarid hillslope. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 24(8), 677-686.

Nourani, V., Komasi, M. and Mano, A., 2009. A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water resources management*, 23(14), 2877.

Nowell, L.H., Capel, P.D. and Dileanis, P.D., 1999. Pesticides in stream sediment and aquatic biota: distribution, trends, and governing factors. *CRC Press.*

Park, S.W., Mostaghimi, S., Cooke, R.A. and McClellan, P.W., 1994. BMP impacts on watershed runoff, sediment, and nutrient yields.1. *JAWRA Journal of the American Water Resources Association*, 30(6), 1011-1023.

Pimentel, D., Berger, B., David, F., Newton, M., Wolfe, B., Karabinakis, E., Clark, S., Poon, E., Abbett, E. and Nandagopal, S., 2004. *Water resources, agriculture and the environment (Technical Report).*

Renard, K.G., Laflen, J.M., Foster, G.R. and McCool, D.K., 1994. The revised universal soil loss equation. *Soil erosion research methods*, *2*, 105-124.

Remesan, R. and Mathew, J., 2014. Hydrological data driven modelling: a case study approach, Vol. 1. *Springer.*

Rinsema, J.G., 2014. Comparison of rainfall runoff models for the Florentine catchment (Bachelor thesis), University *of Twente, Enschede, Netherlands*

Roz, E. P., 2011. Water quality modeling and rainfall estimation: a data driven approach (Master thesis). *University of Iowa, Iowa, US.*

Sadeghi, A.M. and Arnold, J.G., 2002. A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales. In *Total Maximum Daily Load (TMDL): Environmental Regulations, Proceedings of 2002 Conference*. American Society of Agricultural and Biological Engineers.

Šimůnek, J., Van Genuchten, M.T. and Šejna, M., 2016. Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone Journal*, *15*(7).

Singh, V.P. and Prasad, S.N., 1982. Explicit solutions to kinematic equations for erosion on an infiltrating plane. *Modelling Components of Hydrologic Cycle*, 515-538.

Smith, M.W., Cox, N.J. and Bracken, L.J., 2011. Modeling depth distributions of overland flows. *Geomorphology*, 125(3), 402-413.

Solomatine, D.P. and Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. *Journal of hydroinformatics*, 10(1), 3-22.

Storm, B. and Jorgensen, G.H., 1987. Simulation of water flow and soil erosion processes with a distributed physically-based modelling system. *Forest hydrology and watershed management/edited by RH Swanson, PY Bernier & PD Woodard*.

Therrien, R., McLaren, R.G., Sudicky, E.A. and Panday, S.M., 2010. HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. *Groundwater Simulations Group, University of Waterloo, Waterloo, ON*.

Torri, D., Borselli, L., Calzolari, C., Yañez, M. and Salvador-Sanchis, M.P., 2002. Soil erosion, land use, soil quality and soil functions: effects of erosion. *Man and soil at the third millennium,* 131-148.

Tucker, G., Lancaster, S., Gasparini, N. and Bras, R., 2001. The channel-hillslope integrated landscape development model (CHILD). In *Landscape erosion and evolution modeling. Springer,* 349-388.

Tyrrel, S.F. and Quinton, J.N., 2003. Overland flow transport of pathogens from agricultural land receiving faecal wastes. *Journal of Applied Microbiology*, 94, 87-93.

USGS, 1999. The Quality of Our Nation's Waters- Nutrients and Pesticides. US Geological Survey Circular. 1225, 82.

Van Oost, K., Govers, G. and Desmet, P., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landscape ecology*, 15(6), 577-589.

Vaze, J., Jordan, P., Beecham, R., Frost, A. and Summerell, G., 2012. Guidelines for rainfall-runoff modelling: towards best practice model application. *Water Cooperative Research Centre, San Francisco*, 47.

Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M.H., Amelung, W., Aitkenhead, M. and Allison, S.D., 2016. Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone Journal*, *15*(5).

Walker, S.E., Mostaghimi, S., Dillaha, T.A. and Woeste, R.E., 1990. Modeling animal waste management practices: impacts on bacteria levels in runoff from agricultural lands. *Transactions of the ASAE*, 33(3), 807-0817.

Wallach, R., Jury, W.A. and Spencer, W.F., 1988. Transfer of chemicals from soil solution to surface runoff: A diffusion-based soil model. *Soil Science Society of America Journal*, 52(3), 612-618.

Wallach, R. and Shabtai, R., 1992. Surface runoff contamination by soil chemicals: Simulations for equilibrium and first-order kinetics. *Water resources research*, 28(1), 167-173.

Weill, S., Mouche, E. and Patin, J., 2009. A generalized Richards equation for surface/subsurface flow modelling. *Journal of Hydrology*, 366(1-4), 9-20.

Whitehead, P.G., Wilson, E.J. and Butterfield, D., 1998. A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I—model structure and process equations. *Science of the Total Environment*, 210, 547-558.

Wilcock, R.J., 1986. Agricultural run-off: a source of water pollution in New Zealand? *New Zealand agricultural science*.

Wilcock, R.J., Nagels, J.W., Rodda, H.J., O'Connor, M.B., Thorrold, B.S. and Barnett, J.W., 1999. Water quality of a lowland stream in a New Zealand dairy farming catchment. *New Zealand Journal of Marine and Freshwater Research*, 33(4), 683-696.

Williams, J.R. and Berndt, H.D., 1977. Sediment yield prediction based on watershed hydrology. *Transactions of the ASAE*, 20(6), 1100-1104.

Wischmeier, W.H. and Smith, D.D., 1978. Predicting rainfall erosion losses-a guide to conservation planning. *Predicting rainfall erosion losses-a guide to conservation planning.*

Woodward, D.E., 1999. Method to predict cropland ephemeral gully erosion. *Catena*, 37(3-4), 393-399.

Woolhiser, D.A., Smith, R.E. and Goodrich, D.C., 1990. A kinematic runoff and erosion model: documentation and user manual, ARS 77. *US Department of Agriculture*.

Ziegler, A.D., Giambelluca, T.W. and Sutherland, R.A., 2001. Erosion prediction on unpaved mountain roads in northern Thailand: validation of dynamic erodibility modelling using KINEROS2. *Hydrological Processes*, 15(3), 337-358.

Chapter 2

Adapting HYDRUS-1D to Simulate Overland Flow and Reactive Transport During Sheet Flow Deviations

ABSTRACT

Surface runoff is commonly described in numerical models using either the diffusion wave or kinematic wave equations which assume that surface runoff occurs as sheet flow with a uniform depth and velocity across the slope. In reality, overland water flow and transport processes are rarely uniform. Local soil topography, vegetation, and spatial soil heterogeneity control directions and magnitudes of water fluxes, and strongly influence runoff characteristics. There is increasing evidence that variations in soil surface characteristics influence the distribution of overland flow and transport of pollutants. These spatially varying surface characteristics can generate deviations from sheet flow such as physical non-equilibrium flow and transport processes that occur only on a limited fraction of the soil surface. Such processes are rarely considered in numerical models. In this study, we first adapted the HYDRUS-1D model, a popular numerical model for solving the Richards equation for variably-saturated water flow and solute transport in porous media, to solve the diffusion wave equation for overland flow at the soil surface. The numerical results obtained by the new model produced an excellent agreement with an analytical solution for the kinematic wave equation. Additional model tests further demonstrated the applicability of the adapted model to simulate the transport and fate of many different solutes (non-adsorbing tracers, nutrients, pesticides, and microbes) that undergo equilibrium and/or kinetic sorption and desorption, and first- or zero-order reactions. HYDRUS-1D includes a hierarchical series of models of increasing complexity to account for both uniform and physical non-equilibrium flow and transport; e.g., dual-porosity and dual-permeability models, up to a dual-permeability model with

immobile water. This same conceptualization was adapted to simulate physical non-equilibrium overland flow and transport at the soil surface. The developed model improves our ability to describe non-equilibrium overland flow and transport processes, and our understanding of factors that cause this behavior.

Note: This research work has been published in the Vadose Zone Journal.

Liang, J., Bradford, S.A., Šimůnek, J. and Hartmann, A., 2017. Adapting HYDRUS-1D to Simulate Overland Flow and Reactive Transport during Sheet Flow Deviations. Vadose Zone Journal, 16(6), doi: 10.2136/vzj2016.11.0113".

2.1 Introduction

Contaminants at the soil surface (e.g., pesticides, heavy metals, and pathogenic microbes) can be rapidly transported to streams or locations of surface water storage by overland flow. This has been reported to be the primary transport route for contaminant dissemination in agricultural settings [Carpenter et al.,1998; USGS, 1999; Tyrrel and Quinton, 2003]. An understanding of and ability to predict processes that influence the transport and fate of contaminants in runoff water is therefore needed to assess and mitigate risks of contamination of surface water supplies on human health [Furman, 2008; Vereecken et al. 2016].

The diffusion wave equation describes surface runoff as sheet flow with a uniform depth and velocity across the slope. In reality, overland water flow and pollutant transport are rarely uniform. Local soil micro-topography, surface roughness, vegetation, and spatial soil heterogeneity vary over distances of centimeters to meters, and they control directions and magnitudes of water fluxes and concentrations of pollutants [Zhang and Cundy, 1989]. There is increasing evidence that overland flow and transport processes often cannot be described using classical overland flow and transport models that assume uniform flow and transport [e.g., Smith et al., 2011; Cea et al., 2014; Bradford et al., 2015]. At the local scale, spatially varying roughness, vegetation, and micro-topography influence the distribution of shear stress and create hydrologically active and passive or relatively immobile flow regions. Surface active regions for overland flow can be formed in the field by rills or connected networks of micro-depressions which route overland

water and pollutant fluxes on the soil surface where velocity can be 2–7 times higher than the average flow [Dunkerley, 2003; Chen et al., 2013]. Great risk of contamination to water resources occurs from hydraulically active regions because of the presence of large water and contaminant fluxes and reduced time for infiltration and contaminant decay, transformation, and/or interphase mass transfer (e.g., interactions with soil particles). Surface passive or relatively immobile regions are formed by regions of depression and obstruction storage that retain water and pollutants, and inhibit overland flow or cause shallow, slow-moving flow and exchange with surface active zones. These spatially varying surface characteristics generate non-equilibrium overland flow and transport processes.

Physically based, spatially distributed models have the potential to be an efficient tool to examine and optimize the removal of contaminants from overland flow through land-use changes and best management practices. Many publicly available overland pollutant transport models have been developed [Beven and Kirkby, 1979; Haith and Shoemaker, 1987; Bathurst and Connell, 1992; Flanagan et al. 1995; Arnold et al. 1997]. Lumped parameter conceptual models have commonly been used for contaminant transport in overland flow but may be inaccurate because they ignore processes acting on water and chemicals at the soil surface {e.g., Soil and Water Assessment Tool (SWAT) [Sadeghi and Arnold, 2002], Hydrologic Simulation Program Fortran (HSPF) [Donigian et al., 1995], Integrated Nitrogen Catchment model (INCA) [Whitehead et al., 1998; Wade et al., 2002], and COLI [Walker et al., 1990]}. Other mechanistic models consider local scale parameters {e.g., Parallel Flow (ParFlow) [Kollet and Maxwell, 2006],

CATchment Hydrology (CATHY) [Camporese et al., 2010], HydroGeoSphere (HGS) [Therrien et al., 2012], OpenGeoSys (OGS) [Kolditz et al., 2012]}. For example, models that use the actual varying micro-topography based on digital elevation maps have become popular [Gómez and Nearing, 2005; Chen et al., 2013; Zhao and Wu, 2015], but they are computationally demanding and extra effort is needed for collecting required data. In an attempt to overcome these limitations, some mechanistic models neglect local scale variations in parameters by considering simplified smooth surfaces with a global roughness coefficient and a constant average slope; e.g., KINematic runoff and EROSion (KINEROS2) [Miller et al., 2007; Goodrich et al., 2012; Kennedy et al., 2013] and Precipitation-Runoff Modeling System (PRMS) [Leavesley et al., 1983]. These models may provide an excellent approximation of overland flow when calibrated to real data but may not be adequate to describe overland transport that is affected by local scale parameters. Hence, accurate physically-based modeling of non-equilibrium water flow and solute transport still remains a challenge in the field of surface hydrology, and there is currently no widely accepted physically-based model available to simulate overland transport processes of all kinds of reactive chemicals.

Recent studies have demonstrated that equations for overland water flow (diffusion wave equation) and transport (advective-dispersion equation, ADE) are equivalent to those for the subsurface (Richards and ADE equations) when using different functional forms for water content, water capacity, and hydraulic conductivity [Weill et al., 2009; Bittelli et al., 2010]. The HYDRUS-1D software [Šimůnek et al., 2016], a popular numerical computer code for solving water flow and reactive transport

in variably-saturated porous media, can thus be modified to simulate overland flow rather than subsurface flow. Existing subsurface flow and transport models in HYDRUS-1D include a hierarchical series of physical non-equilibrium models of increasing complexity, from dual-porosity and dual-permeability models, up to a dual-permeability model with immobile water [Šimůnek and van Genuchten, 2008]. All these non-equilibrium models were derived from Richards and convection-dispersion equations.

The objective of this study is to adapt the existing subsurface version of HYDRUS-1D to simulate similar uniform or physical non-equilibrium flow and reactive solute transport processes during runoff at the soil surface. A detailed description of the implemented equations for overland flow and transport will be presented first. Next, illustrative examples of model results will be shown, with special focus on modeling sorption/desorption processes and physical non-equilibrium flow and transport. A limited comparison of model results and experimental and numerical data that exhibit deviations from sheet flow will also be provided. The developed models improve our ability to quantify natural flow and transport processes at the soil surface, and simulation results enhance our understanding of factors that cause deviations from sheet flow.

2.2 Numerical Models

A variety of physical and chemical equilibrium and non-equilibrium models is available in HYDRUS-1D to describe water flow and solute transport processes in the subsurface [Šimůnek and van Genuchten, 2008]. These models form a hierarchical system of models of increasing complexity to account for both physical equilibrium and non-equilibrium flow, including (a) a uniform flow model, (b) a dual-porosity (mobile-immobile water) model, (c) a dual-permeability model, and (d) a dual-permeability model with immobile water. These conceptual models for subsurface flow were adapted in this study to account for both equilibrium and non-equilibrium overland flow and reactive solute transport, resulting in (a) a uniform flow and transport (UFT) model, (b) a horizontal mobile-immobile (HMIM) model, (c) a vertical mobile-immobile (VMIM) model, (d) an active-passive region (APR) model, (e) a combined APR and HMIM (APR-H) model, and (f) an combined APR and VMIM (APR-V) model. While the mobile-immobile models for overland flow correspond to the dual-porosity models for subsurface flow, the active-passive region models correspond to the dual-permeability subsurface models. Both mobile-immobile and active-passive regions models have two additional subsets with either horizontal or vertical subregions. While the focus in this manuscript is on physical non-equilibrium models that describe conditions when flow and transport occur only on a limited fraction of the soil surface that bypasses regions with little or no flow, the implemented models can also account for chemical non-equilibrium when sorbed concentrations are not in equilibrium with liquid concentrations. The schematics of newly developed overland flow models that will be discussed in the

next section are shown in Figure 2.1. Below we will also describe the governing equations for each of these models. Table 2.1 provides a comparison of different definitions of variables for subsurface and surface models.

Table 2.1. Subsurface and overland flow and transport parameters in the various models

Subsurface Models			**Surface Models**		
Uniform Flow and Transport	Water content	θ	**Uniform Flow and Transport (UFT)**	Water depth	h
	Hydraulic conductivity	$K(h)$		Horizontal Conductivity	$\frac{kh^{2/3}}{n\sqrt{S}}$
	Hydraulic capacity	$C(h)$		Constant	1
	Darcy's velocity	q		Runoff flow rate	Q
Dual-Porosity Model	Mobile water content	θ_{mo}	**Mobile-Immobile Regions Model (HMIM and VMIM)**	Mobile water depth	h_m
	Immobile water content	θ_{im}		Immobile water depth	h_{im}
Dual-Permeability Model	Fracture water content	θ_f	**Active-Passive Regions Model (APR)**	Active region water depth	h_1
	Matrix water content	θ_m		Passive region water depth	h_2
	Fracture hydraulic conductivity	$K_f(h_f)$		Active horizontal conductivity	$\frac{k_1 h_1^{5/3}}{n_1\sqrt{S}}$
	Matrix hydraulic conductivity	$K_m(h_m)$		Passive horizontal conductivity	$\frac{k_2 h_2^{5/3}}{n_2\sqrt{S}}$
	Matrix concentration	c_m		Concentration in the active region	c_1
	Fracture concentration	c_f		Concentration in the passive region	c_2
Dual-Permeability Model with MIM	Water content / concentration in the mobile zone of the matrix region	$\theta_{m,m} / c_{m,m}$	**Active-Passive Regions Model with Immobile Domains (APR-H, APR-V)**	Water depth / concentration in the mobile zone of the active region	h_{1m} / c_{1m}
	Water content / concentration in the immobile zone of the matrix region	$\theta_{im,m} / c_{im,m}$		Water depth / concentration in the mobile zone of the passive region	h_{2m} / c_{2m}
	Fracture water content / concentration	θ_f / c_f		Water depth / concentration in the immobile zone of the active region	h_{1im} / c_{1im}
				Water depth / concentration in the immobile zone of the passive region	h_{2im} / c_{2im}

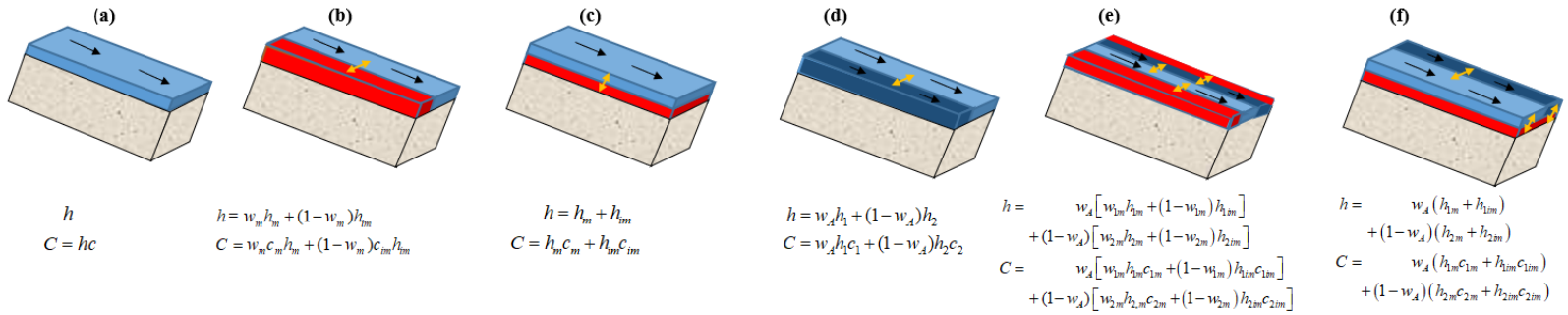


Figure 2.1. Conceptual equilibrium and physical non-equilibrium models for overland flow and solute transport: **(a) Uniform Flow and Transport Model (UFT), (b) Horizontal Mobile-Immobile Regions (HMIM) Model, (c) Vertical Mobile-Immobile (VMIM) Regions Model, (d) Active-Passive Regions (APR) Model, (e) Combined APR and HMIM (APR-H) model (f) Combined APR and VMIM (APR-V) model.** In the plots, h is the surface water depth, h_m and h_{im} in (b) and (c) are surface water depths in the mobile (blue) and immobile (red) flow regions, respectively; h_1 and h_2 in (d) are surface water depths in the passive (dark blue) and active (blue) regions, respectively; h_{1im}, h_{2im}, h_{1m} and h_{2m} in (e) and (f) are water depths in immobile (subscript *im*) (red) and mobile (subscript *m*) zones of the passive (subscript *2*) (dark blue) and active (subscript *1*) (blue) regions, respectively; c are concentrations in corresponding regions, with subscripts having the same meaning as for surface water depths, while C is the total solute content of the liquid phase; black and yellow arrows show directions of flow and exchange, respectively. w_m in (b) is the fraction of the mobile region; w_A in (d, e, and f) is the fraction of the active region; w_{1m} and w_{2m} in (e) are the fractions of the mobile region in the active and passive regions, respectively.

2.2.1 UFT Model

Overland flow and solute transport are commonly described using diffusion wave and advection-dispersion equations, respectively. This approach lumps irregular land surface characteristics into effective parameters that are used to simulate uniform sheet flow and transport (Fig. 2.1a). The diffusion wave equation for overland flow may be written in a similar form to the Richards equation as:

$$\frac{\partial h}{\partial t} = \frac{\partial}{\partial x}\left(\frac{k h^{5/3}}{n\sqrt{S}} \frac{\partial (z+h)}{\partial x} \right) + q \qquad [1]$$

where x is a space coordinate in direction of flow [L; where L denotes units of length], t is time [T; where T denotes units of time], h is the surface water depth [L], q is the source/sink term [LT^{-1}] accounting for precipitation, evaporation, and infiltration, z is the land surface elevation [L], S is the mean local slope [-], n is a Manning's roughness coefficient for overland flow [-], and k is a unit conversion factor [$L^{1/3}T^{-1}$]. Detailed derivation of this equation can be found, for example, in Weill et al. [2009], Hromadka and Lai [1985], Panday and Huyakorn [2004], and Šimůnek [2015]. The n parameter is either dimensionless when the units conversion factor k is used or has units of [$TL^{-1/3}$] when it is not. The parameter k can also be used to convert equation [1] between SI and English units and can be left out when consistent units are used. However, it is a standard practice to use k=1 ($m^{1/3}s^{-1}$) for SI units and k=1.49 ($ft^{1/3}s^{-1}$) for English units.

When rainfall (R, [LT^{-1}]) (or irrigation), evaporation (E, [LT^{-1}]), and infiltration (I, [LT^{-1}]) rates are the only source/sink terms then $q=R-E-I$. Infiltration is usually determined using various empirical, semi-empirical, or physical models. In this work the

infiltration rate I is described using the Horton's equation [Horton, 1939]:

$$I = f_c + \left(f_0 - f_c \right) e^{-k_i t} \qquad [2]$$

where f_0 is the initial infiltration rate [LT^{-1}], f_c is the final equilibrium infiltration rate [LT^{-1}], and k_i is a constant representing the rate of decrease in infiltration [T^{-1}]. According to Eq. [2] infiltration starts at a rate f_0, and then decreases exponentially with time until it reaches an equilibrium infiltration rate f_c. The use of Horton's equation is only the first attempt in our model to consider the infiltration process, since the focus of the manuscript is on overland flow and transport processes. This infiltration model will be replaced in the future by other empirical and/or process-based infiltration models.

Solute transport in overland flow is usually described using the advection-dispersion equation (ADE) of the form:

$$\frac{\partial hc}{\partial t} + \frac{\partial s}{\partial t} = \frac{\partial}{\partial x}\left(hD \frac{\partial c}{\partial x} \right) - \frac{\partial Qc}{\partial x} + h\phi + Rc_r - Ic \qquad [3]$$

where c is the solute concentration in the aqueous phase [ML^{-3}; where M denotes units of mass], c_r is the concentration in rainfall water [ML^{-3}], s is the sorbed solute concentration at the soil surface area [ML^{-2}], D is the effective dispersion coefficient accounting for both molecular diffusion and hydrodynamic dispersion [L^2T^{-1}], ϕ is a sink/source term that accounts for various zero- and first-order or other reactions [$ML^{-3}T^{-1}$], and Q is the runoff flow rate [L^2T^{-1}]. The parameter Q is given as:

$$Q = hU = -\frac{kh^{5/3}}{n\sqrt{S}} \frac{\partial \left(h + z \right)}{\partial x} \qquad [4]$$

where U is a depth-averaged velocity [LT^{-1}] calculated using the Manning-Strickler uniform flow formula [Hromadka and Lai, 1985]. It should be mentioned that the subsurface pore-water velocity and U have the same units, but the subsurface Darcy's velocity and Q do not have the same units. The effect of diffusion on the dispersion coefficient can often be ignored and in this case, D can be defined as the product of the dispersivity (λ, [L]) and U.

A variety of sorption processes can be considered in HYDRUS-1D, including chemical equilibrium described using both linear and nonlinear adsorption isotherms and chemical non-equilibrium described using the first-order kinetic models. These same sorption models may also be employed with overland transport. However, the units for s are different for subsurface [MM^{-1}] and overland [ML^{-2}] transport. Furthermore, the soil bulk density is not required in the overland transport equation, contrary to the subsurface transport equation. In the examples given below, the value of s in the equilibrium model is given as:

$$\frac{\partial s}{\partial t} = \frac{\partial K_D c}{\partial t} \qquad [5a]$$

where K_D is the distribution coefficient [L]. The 1 site kinetic sorption model is written as:

$$\frac{\partial s}{\partial t} = \omega\left(K_D c - s\right) \qquad [5b]$$

where ω is the first-order rate coefficient [T^{-1}] representing kinetic sorption. The combined equilibrium and kinetic sorption model is written as:

$$\frac{\partial s}{\partial t} = \frac{\partial f_{eq} s_{eq}}{\partial t} + \frac{\partial (1 - f_{eq}) s_k}{\partial t} \qquad [5c]$$

where f_{eq} [-] is the fraction of equilibrium sorption sites, and the subscripts *eq* and *k* on *s* denote equilibrium and kinetic sorption sites given by expression similar to Eqs. [5a] and [5b], respectively. The two-site kinetic sorption model is given as:

$$\begin{aligned} \frac{\partial s}{\partial t} &= \frac{\partial s_{k1}}{\partial t} + \frac{\partial s_{k2}}{\partial t} \\ \frac{\partial s_{k1}}{\partial t} &= k_{att1} hc - k_{\det 1} s_{k1} \\ \frac{\partial s_{k2}}{\partial t} &= k_{att2} hc - k_{\det 2} s_{k2} \end{aligned} \qquad [5d]$$

where k_{att1} [T^{-1}] and k_{att2} [T^{-1}] are kinetic sorption rate coefficients, k_{det1} [T^{-1}] and k_{det2} [T^{-1}] are kinetic desorption rate coefficients, and the subscripts 1 and 2 on parameters denote the kinetic sorption sites 1 and 2, respectively. Note that the latter model is commonly also used to describe attachment and detachment processes when modeling transport of particular substances, such as viruses, colloids or pathogens. It should be mentioned that Eq. [5b] can be recast in terms in $k_{att1}=\omega K_D/h$ and $k_{det1}=\omega$.

The upper boundary condition (BC) for water can be either a prescribed water head or flux as follows:

$$\begin{aligned} h(x,t) &= h_0(t) && at \quad x = 0 \\ \frac{k h^{5/3}}{n\sqrt{S}} \frac{\partial (z+h)}{\partial x} &= q_0(t) && at \quad x = 0 \end{aligned} \qquad [6]$$

where h_0 [L] and q_0 [L^2T^{-1}] are the water head and flux at the upper boundary, respectively. The water depth gradient at the lower BC is assumed to be zero:

$$\frac{\partial h}{\partial x} = 0 \qquad at \quad x = L \qquad [7]$$

Two types of BCs can be applied for transport at the upper or lower boundaries. The first-type boundary conditions prescribe the concentration at the boundary:

$$c(x,t) = c_0(x,t) \qquad at \quad x = 0 \ or \ x = L \qquad [8]$$

whereas third-type (Cauchy type) boundary conditions may be used to prescribe the concentration flux at the boundary:

$$\left(hD \frac{\partial c}{\partial x} \right) - Qc = q_0 c_0 \qquad at \ x = 0 \ or \ x = L \qquad [9]$$

in which c_0 is the concentration of the incoming fluid [ML^{-3}].

2.2.2 HMIM Model

A schematic of the HMIM model is shown in Figure 2.1b. This model assumes that the soil surface is horizontally divided, parallel to the direction of water flow, into regions with mobile and immobile water. Water and solutes in the immobile region may be stored, retained, and exchanged with the mobile domain. Similar to Equation [1], the movement of water in the mobile region and moisture dynamics in the immobile region are given as:

$$\begin{aligned} \frac{\partial h_m}{\partial t} &= \frac{\partial}{\partial x}\left(\frac{k h_m^{\,5/3}}{n\sqrt{S}} \frac{\partial (z + h_m)}{\partial x} \right) - \frac{E_w^{MIM}}{w_m} + q_m \\ \frac{\partial h_{im}}{\partial t} &= + \frac{E_w^{MIM}}{1 - w_m} + q_{im} \end{aligned} \qquad [10]$$

where E_w^{MIM} is the water transfer rate between mobile and immobile regions [LT^{-1}], w_m is the ratio of the width of the mobile region and the total width of the soil surface, and

subscripts *m* and *im* denote parameters associated with the mobile and immobile regions, respectively. The parameter E_w^{MIM} is assumed to be proportional to the difference in water depths between the two regions:

$$E_w^{MIM} = \alpha\left(h_m - h_{im}\right) \quad [11]$$

where α [T^{-1}] is the first-order mass transfer coefficient [Simunek et al., 2003]. The average surface water depth of the entire domain is:

$$h = w_m h_m + \left(1 - w_m\right) h_{im} \quad [12]$$

The governing solute transport equations for the mobile and immobile regions are given as follows:

$$\begin{aligned} \frac{\partial h_m c_m}{\partial t} + \frac{\partial s_m}{\partial t} &= \frac{\partial}{\partial x}\left(h_m D_m \frac{\partial c_m}{\partial x}\right) - \frac{\partial Q_m c_m}{\partial x} - \frac{E_s^{MIM}}{w_m} + h_m \phi_m + R_m c_r - I_m c_m \\ \frac{\partial h_{im} c_{im}}{\partial t} + \frac{\partial s_{im}}{\partial t} &= \frac{E_s^{MIM}}{1 - w_m} + h_{im} \phi_{im} + R_{im} c_r - I_{im} c_{im} \end{aligned} \quad [13]$$

where E_s^{MIM} is the solute transfer rate between mobile and immobile regions [$ML^{-2}T^{-1}$]. The parameter E_s^{MIM} is given as:

$$E_s^{MIM} = \omega_m (1 - w_m) h_{im} (c_m - c_{im}) + E_w^{MIM} c^* \quad [14]$$

where ω_m is the solute mass transfer coefficient [T^{-1}], and c^* is a concentration that is equal to c_m for $E_w^{MIM} > 0$ and c_{im} for $E_w^{MIM} < 0$. Note that solute exchange between the two liquid regions is modeled in Eq. [14] as the sum of an apparent first-order diffusion process and advective transport.

It should be mentioned that Eqs. [10] and [13] are written in terms of local-scale mass balances in mobile and immobile regions. To formulate them in terms of the total

region the mass balance equations for mobile and immobile regions need to be multiplied by w_m and (1- w_m), respectively.

2.2.3 VMIM Model

A schematic of the VMIM model is presented in Figure 2.1c. This model assumes that the surface water depth is vertically divided, parallel to the mean surface slope, into an immobile region adjacent to the soil surface and an overlying mobile domain. This conceptual picture is consistent with an immobile depression and/or obstruction storage zone that needs to fill before the initiation of overland flow [Panday and Huyakorn, 2004; Guber et al., 2009]. An immobile region is supposed to account for the effects of small surface roughness and/or vegetation that prevents an immediate surface runoff once ponding is reached. It represents a small water layer that needs to be formed before surface runoff can occur. Water flow is described as:

$$\begin{aligned} \frac{\partial h_m}{\partial t} &= \frac{\partial}{\partial x}\left(\frac{k h_m^{\,5/3}}{n\sqrt{S}} \frac{\partial (z + h_m)}{\partial x} \right) - E_w^{MIM} + R - E \\ \frac{\partial h_{im}}{\partial t} &= + E_w^{MIM} - I \end{aligned} \qquad [15]$$

In contrast to the HMIM model, the parameter E_w^{MIM} is now given as:

$$E_w^{MIM} = \left(R - E + Q_{up} / \Delta L \right) H_o \left(h_{im}^{\max} - h_T \right) \qquad [16]$$

where $h_{im}^{\max}$ is the maximum water depth for the immobile region [L], h_T is the total water depth [L] that is equal to h_m+h_{im}, Q_{up} is flow from upstream [L^2T^{-1}], and ΔL is the length of the surface [L] over which inflow from upstream flows into the immobile region. The Heaviside function (H_o) in Eq. [16] is equal to one when $h_{im}^{\max} > h_T$ and is zero when $h_{im}^{\max}$

$\leq h_T$. Consequently, water flow in the mobile domain only occurs when the immobile region is filled ($h_{im}^{\max} \leq h_T$) at a particular location.

Solute transport equations for the mobile and immobile regions are given as follows:

$$\begin{aligned} \frac{\partial h_m c_m}{\partial t} &= \frac{\partial}{\partial x}\left(h_m D_m \frac{\partial c_m}{\partial x} \right) - \frac{\partial Q_m c_m}{\partial x} - E_s^{MIM} + h_m \phi_m + R c_r \\ \frac{\partial h_{im} c_{im}}{\partial t} + \frac{\partial s_{im}}{\partial t} &= + E_s^{MIM} + h_{im} \phi_{im} - I c_{im} \end{aligned} \qquad [17]$$

Note that the transport equation for the mobile region does not include the sorption term since this region is not in direct contact with soil. The parameter E_s^{MIM} is given as:

$$E_s^{MIM} = \omega_m h_{im} (c_m - c_{im}) H_o\left(h_T - h_{im}^{\max}\right) + \left(R c_r + Q_{up} c_m / \Delta L\right) H_o\left(h_{im}^{\max} - h_T\right) \qquad [18]$$

2.2.4 APR Model

A schematic of the APR model is given in Figure 2.1d. This model assumes that the soil surface is divided, parallel to the direction of water flow, into hydraulically active (fast flow) and passive (slow flow) domains. This model allows for water and solute transport in both active and passive regions, and exchange between these regions. Surface water flow in each region is, in analogy of the dual-permeability subsurface flow model of Gerke and van Genuchten [1993], described using separate diffusion wave equations:

$$\begin{aligned} \frac{\partial h_1}{\partial t} &= \frac{\partial}{\partial x}\left(\frac{k h_1^{5/3}}{n_1 \sqrt{S}} \frac{\partial (z + h_1)}{\partial x} \right) - \frac{E_w^{APR}}{w_A} + q_1 \\ \frac{\partial h_2}{\partial t} &= \frac{\partial}{\partial x}\left(\frac{k h_2^{5/3}}{n_2 \sqrt{S}} \frac{\partial (z + h_2)}{\partial x} \right) + \frac{E_w^{APR}}{1 - w_A} + q_2 \end{aligned} \qquad [19]$$

where E_w^{APR} is the water transfer rate between active and passive regions [LT^{-1}], w_A is the

ratio of the width of the surface active region and the total surface width [-], and the subscripts 1 and 2 refer to hydrologically active and passive regions, respectively. Values of E_w^{APR} and *h* are quantified in a similar manner to Eqs. [11] and [12] as follows:

$$E_w^{APR} = \alpha_{12} (h_1 - h_2) \quad [20]$$

$$h = w_A h_1 + (1 - w_A) h_2 \quad [21]$$

Solute transport is described using separate advection-dispersion equations for each region:

$$\begin{aligned} \frac{\partial h_1 c_1}{\partial t} + \frac{\partial s_1}{\partial t} &= \frac{\partial}{\partial x}\left(h_1 D_1 \frac{\partial c_1}{\partial x} \right) - \frac{\partial Q_1 c_1}{\partial x} - \frac{E_s^{APR}}{w_A} + h_1 \phi_1 + R c_r - I_1 c_1 \\ \frac{\partial h_2 c_2}{\partial t} + \frac{\partial s_2}{\partial t} &= \frac{\partial}{\partial x}\left(h_2 D_2 \frac{\partial c_2}{\partial x} \right) - \frac{\partial Q_2 c_2}{\partial x} + \frac{E_s^{APR}}{1 - w_A} + h_2 \phi_2 + R c_r - I_2 c_2 \end{aligned} \quad [22]$$

where E_w^{APR} is the solute transfer rate between active and passive regions [$ML^{-2}T^{-1}$]. The E_s^{APR} term is quantified in a similar manner to Eq. [14] as:

$$E_s^{APR} = \omega_{12} (1 - w_1) h_2 (c_1 - c_2) + E_w^{APR} c^* \quad [23]$$

where ω_{12} is the solute mass transfer coefficient for transfer between active and passive regions [T^{-1}].

The total flux is obtained as:

$$Q = Q_1 w_A + Q_2 (1 - w_A) \quad [24]$$

and the average flux concentration is:

$$c_T = \frac{w_A c_1 Q_1 + (1 - w_A) c_2 Q_2}{w_A Q_1 + (1 - w_A) Q_2} \quad [25]$$

It should be mentioned that Eqs. [19] and [22] are written in terms of local-scale

mass balances in regions 1 and 2. To formulate them in terms of the entire region the mass balance equations for regions 1 and 2 need to be multiplied by w_A and (1-w_A), respectively.

2.2.5 APR-H Model

A schematic for the combined APR and HMIM (APR-H) model is shown in Figure 2.1e. This model assumes that the soil surface is horizontally divided, in the direction parallel to water flow, into four regions: (i) an active (fast flow) region; (ii) a passive (slow flow) region; (iii) an immobile region that is in contact and exchanges mass (water and/or solute) horizontally with the active region; and (iv) and an immobile region that is in contact and exchanges mass horizontally with the passive region.

Several different surface water depths are considered in the combined APR-H model. Values of h_1 and h_2 are divided into mobile (h_{1m} and h_{2m}) and immobile (h_{1im} and h_{2im}) regions using Eq. [12] such that the average water depth for the entire domain (h) is given as:

$$h = w_A\left[w_{1m}h_{1m} + \left(1-w_{1m}\right)h_{1im}\right] + (1-w_A)\left[w_{2m}h_{2m} + \left(1-w_{2m}\right)h_{2im}\right] \quad [26]$$

The movement of water in active, passive, and immobile zones is given as:

$$\frac{\partial h_{1m}}{\partial t} = \frac{\partial}{\partial x}\left(\frac{k h_{1m}^{5/3}}{n_1\sqrt{S}}\frac{\partial h_{1m}}{\partial x}\right) - \frac{E_w^{APR}}{w_A} - \frac{E_{1w}^{MIM}}{w_A w_{1m}} + q_{1m} \quad [27a]$$

$$\frac{\partial h_{1im}}{\partial t} = + \frac{E_{1w}^{MIM}}{w_A(1-w_{1m})} + q_{1im} \quad [27b]$$

$$\frac{\partial h_{2m}}{\partial t} = \frac{\partial}{\partial x}\left(\frac{k h_{2m}^{5/3}}{n_2\sqrt{S}}\frac{\partial h_{2m}}{\partial x}\right) + \frac{E_w^{APR}}{1-w_A} - \frac{E_{2w}^{MIM}}{(1-w_A)w_{2m}} + q_{2m} \quad [27c]$$

$$\frac{\partial h_{2im}}{\partial t} = \qquad\qquad + \frac{E_{2w}^{MIM}}{(1-w_A)(1-w_{2m})} + q_{2im} \qquad [27d]$$

where E_{1w}^{MIM} and E_{2w}^{MIM} denote the water transfer rate between mobile and immobile domains in active and passive regions [LT^{-1}], respectively. The parameters E_{1w}^{MIM} and E_{2w}^{MIM} are obtained in an analogous fashion to Eq. [11] using separate values of α, h_m, and h_{im} for active (α_1, h_{1m}, and h_{1im}) and passive (α_2, h_{2m}, and h_{2im}) regions, whereas E_w^{APR} is determined using Eq. [20].

The following governing equations describe solute transport in the surface active, passive, and immobile regions:

$$\frac{\partial h_{1m}c_{1m}}{\partial t} + \frac{\partial s_{1m}}{\partial t} = \frac{\partial}{\partial x}\left(h_{1m}D_{1m}\frac{\partial c_{1m}}{\partial x}\right) + \frac{\partial Q_{1m}c_{1m}}{\partial x} - \frac{E_s^{APR}}{w_A} - \frac{E_{s1}^{MIM}}{w_A w_{1m}} + h_{1m}\phi_{1m} + R_m c_r - I_{1m}c_{1m} \qquad [28a]$$

$$\frac{\partial h_{1im}c_{1im}}{\partial t} + \frac{\partial s_{1im}}{\partial t} = \frac{E_{s1}^{MIM}}{w_A(1-w_{1m})} + h_{1im}\phi_{1im} + R_{im}c_r - I_{1im}c_{1im} \qquad [28b]$$

$$\frac{\partial h_{2m}c_{2m}}{\partial t} + \frac{\partial s_{2m}}{\partial t} = \frac{\partial}{\partial x}\left(h_{2m}D_{2m}\frac{\partial c_{2m}}{\partial x}\right) + \frac{\partial Q_{2m}c_{2m}}{\partial x} - \frac{E_s^{APR}}{1-w_A} - \frac{E_{s2}^{MIM}}{(1-w_A)w_{2m}} + h_{2m}\phi_{2m} + R_m c_r - I_{2m}c_{2m} \qquad [28c]$$

$$\frac{\partial h_{2im}c_{2im}}{\partial t} + \frac{\partial s_{2im}}{\partial t} = \frac{E_{s2}^{MIM}}{(1-w_A)(1-w_{2m})} + h_{2im}\phi_{2im} + R_{im}c_r - I_{2im}c_{2im} \qquad [28d]$$

where E_{1s}^{MIM} and E_{2s}^{MIM} denote the solute transfer rate between mobile and immobile domains in active and passive regions [$ML^{-2}T^{-1}$], respectively. The parameters E_{1s}^{MIM} and E_{2s}^{MIM} are obtained in an analogous fashion to Eq. [14] using separate values of ω_m, w_m, h_{im}, c_m, c_{im}, and E_w^{MIM} for active (ω_1, w_{1m}, h_{1im}, c_{1m}, c_{1im}, and E_{1w}^{MIM}) and passive (ω_2, w_{2m},

h_{2im}, c_{2m}, c_{2im}, and E_{2w}^{MIM}) regions, whereas E_s^{APR} is determined using Eq. [23].

2.2.6 APR-V Model

A schematic for the combined APR and VMIM (APR-V) model is shown in Figure 2.1f. Similar to the APR model, the combined model assumes that the soil surface is horizontally divided, in the direction parallel to water flow, into active (fast flow) and passive (slow flow) regions. Each of these regions is further divided vertically into an immobile region adjacent to the soil surface and an overlying mobile domain.

The values of h_1 and h_2 are again divided into mobile (h_{1m} and h_{2m}) and immobile (h_{1im} and h_{2im}) regions for the combined APR-V model. In this case, the average water depth for the entire domain (h) is given as:

$$h = w_A\left[h_{1m} + h_{1im}\right] + (1 - w_A)\left[h_{2m} + h_{2im}\right] \qquad [29]$$

The movement of water in active, passive, and immobile zones is given as:

$$\frac{\partial h_{1m}}{\partial t} = \frac{\partial}{\partial x}\left(\frac{k h_{1m}^{\,5/3}}{n_1\sqrt{S}}\frac{\partial h_{1m}}{\partial x}\right) - \frac{E_w^{APR}}{w_A} - \frac{E_{1w}^{MIM}}{w_A} + R - E \qquad [30a]$$

$$\frac{\partial h_{1im}}{\partial t} = \qquad + \frac{E_{1w}^{MIM}}{w_A} - I_1 \qquad [30b]$$

$$\frac{\partial h_{2m}}{\partial t} = \frac{\partial}{\partial x}\left(\frac{k h_{2m}^{\,5/3}}{n_2\sqrt{S}}\frac{\partial h_{2m}}{\partial x}\right) + \frac{E_w^{APR}}{1 - w_A} - \frac{E_{2w}^{MIM}}{1 - w_A} + R - E \qquad [30c]$$

$$\frac{\partial h_{2im}}{\partial t} = \qquad + \frac{E_{2w}^{MIM}}{1 - w_A} - I_2 \qquad [30d]$$

Values of E_{1w}^{MIM} and E_{2w}^{MIM} are given in this model similar to Eq. [16] using separate values of R, E, h_T, and $h_{im}^{\max}$ for active (h_{1T} and $h_{1im}^{\max}$) and passive (h_{2T} and $h_{2im}^{\max}$) regions.

Solute transport in the surface active, passive, and immobile regions are described using the following model:

$$\frac{\partial h_{1m} c_{1m}}{\partial t} = \frac{\partial}{\partial x}\left(h_{1m} D_{1m} \frac{\partial c_{1m}}{\partial x} \right) + \frac{\partial Q_{1m} c_{1m}}{\partial x} - \frac{E_s^{APR}}{w_A} - \frac{E_{s1}^{MIM}}{w_A} + h_{1m}\phi_{1m} + Rc_r \qquad [31a]$$

$$\frac{\partial h_{1im} c_{1im}}{\partial t} + \frac{\partial s_{1im}}{\partial t} = \qquad + \frac{E_{s1}^{MIM}}{w_A} + h_{1im}\phi_{1im} - I_1 c_{1im} \qquad [31b]$$

$$\frac{\partial h_{2m} c_{2m}}{\partial t} = \frac{\partial}{\partial x}\left(h_{2m} D_{2m} \frac{\partial c_{2m}}{\partial x} \right) + \frac{\partial Q_{2m} c_{2m}}{\partial x} - \frac{E_s^{APR}}{1-w_A} - \frac{E_{s2}^{MIM}}{1-w_A} + h_{2m}\phi_{2m} + Rc_r \qquad [31c]$$

$$\frac{\partial h_{2im} c_{2im}}{\partial t} + \frac{\partial s_{2im}}{\partial t} = \qquad + \frac{E_{s2}^{MIM}}{1-w_A} + h_{2im}\phi_{2im} - I_2 c_{2im} \qquad [31d]$$

Parameters E_{1s}^{MIM}, E_{2s}^{MIM}, and E_s^{APR} are determined in a similar manner to Eqs. [18], [18], and [23], respectively. Note that the APR-V model assumes that there are no interactions between the two immobile regions under the active and passive flow regions. The additional assumption in the current implementation of the APR-V model is that the maximum water depths $h_{im}^{\max}$ in the two immobile regions are the same.

2.3. Numerical Implantation

The HYDRUS software uses the Galerkin-type linear finite element method (FEM) for spatial discretization of the governing partial differential equations and the finite difference method to approximate temporal derivatives. The FEM is commonly used in hydrological models, such as in HydroGeoSphere and OGS. An important advantage of the FEM over the finite difference method to solve the 1D diffusion wave equation is the ease of increasing mesh resolution, so that shock fronts can be simulated

without any numerical oscillations. Since the governing overland flow equations were formulated using the subsurface flow equations with different definitions of the water content, hydraulic capacity, and conductivity coefficients (see Table 2.1), the original fully implicit finite difference scheme with Picard linearization was used to solve the overland flow equation as well, while a Crank-Nicholson finite difference scheme was used to solve the advection-dispersion equations. Details about the numerical solutions of corresponding subsurface flow and transport models are provided in the HYDRUS-1D manual [Simunek et al., 2016]. The same graphical user interface (GUI) in HYDRUS-1D is used to select and execute subsurface and overland flow and transport models.

The numerical implementation of the overland flow equation was verified by comparing numerical results with the analytical solution of the kinematic wave equation [Woolhiser and Liggett, 1967; Singh, 1996]. There was an excellent agreement between the numerical and analytical solutions at different times and steady-state conditions. A mass balance was also calculated for all of the overland flow and transport models to verify the correctness of the implementation of more complex models, for which analytical solutions either do not exist or cannot be derived. Results demonstrated that the numerical implementation conserved mass.

2.4 Applications

This section presents a large number of numerical examples of overland flow and solute transport using the various model formulations. The flow domain is 100 m long and discretized into 101 finite elements. The slope is 1%, and unless otherwise noted the surface roughness n is 0.01. The dispersivity (λ) is set equal to 10 m.

Table 2.2. Simulation conditions of test examples

Models	Figure Numbers	Simulation Conditions
UFT	**Fig. 2**	Total simulation time: 20 mins; constant head inflow (1 cm) for 6 mins; n=0.01. a. no infiltration. b. infiltration parameters: f_0 = 0.00199 cm s^{-1}, f_c = 0.00009433 cm s^{-1}, and k_i = 0.00404 s^{-1}.
	Fig. 3	The same conditions as for Fig. 2a, but additionally with solute inflow. a. solute inflow and no retardation: K_D=0 cm. b. solute inflow and retardation: K_D=1 cm. c. solute inflow and attachment/detachment: k_a=0.01 s^{-1} and k_d=0.001 s^{-1}.
HMIM	**Fig. 4**	The same conditions as for Fig. 2a**.** a. constant w_m=0.5 with α = 0, 0.001, 0.005, and 0.01s^{-1}. b. constant α =0.01 s^{-1} with w_m=0.3, 0.5, and 0.7.
	Fig. 5	Steady state flow conditions; solute inflow for 30 mins; total simulation time 60 min; w_m=0.5, α =0.01 s^{-1}, with ω_m = 0, 0.001, 0.005, and 0.01s^{-1}.
VMIM	**Fig. 6a**	Total simulation time: 60 mins; constant rainfall R=0.00666 cm s^{-1}.s for 20 mins; f_0 = 0.00399 cms^{-1}, f_c = 0.00009433 cms^{-1}, and k_i = 0.000404 s^{-1}; $h_{im}^{\max}$ = 0, 0.3, and 0.5 cm.
APR	**Fig. 7a**	Total simulation time: 60 mins; constant head inflow (1 cm) for 20 mins; n_1 = 0.01, n_2 = 0.05; w_A=0.5; α_{12} =0, 0.001 and 0.01 s^{-1}.
	Fig. 8a	The same conditions as the Fig.5; n_1 = 0.01, n_2 = 0.05; w_A=0.5; ω_{12} =0, 0.001 and 0.01 s^{-1}.
APR-H	**Fig. 7b**	The same conditions as the Fig.7a; $\alpha = 0.01s^{-1}$; w_A=0.5; w_{1m}=0.5, w_{2m}=0.5; α_{12} =0, 0.001 and 0.01 s^{-1}.
	Fig. 8b	The same conditions as the Fig.8a; ω_m = 0.01s^{-1}; w_A=0.5; w_{1m}=0.5, w_{2m}=0.5; ω_{12} =0, 0.001 and 0.01 s^{-1}.
APR-V	**Fig. 6b**	The same conditions as the Fig. 6a**,** but with two different Manning's roughness coefficients. n_1 = 0.01; n_2 = 0.05; α=0.1s^{-1}; $h_{1im}^{\max} = h_{2im}^{\max} = 0.5$cm.

2.4.1 UFT Model

Figure 2.2 presents an example of uniform overland flow over an impervious (Fig. 2.2a) and permeable (Fig. 2.2b) soil surface. In this example, inflow was induced by specifying the pressure head equal to 1 cm at the top boundary for 6 min, and then simulating overland flow for a total of 20 min. Water depths are shown as a function of time at different locations (0, 20, 40, 60, and 80 m) from the top of the slope. The inlet water pulse moves along the impervious soil surface in Figure 2.2a and reaches the bottom of the profile after about 200 s. A maximum water depth of 1 cm that is consistent with the inlet boundary condition is reached and then gradually reduced when the inflow is stopped. In contrast to Figure 2.2a, the water depths are not uniform across the permeable soil surface in Figure 2.2b due to infiltration. Infiltration reduces the water mass that reaches the bottom boundary and delays its arrival. Mass balance calculations indicated that the sum of cumulative discharge, infiltration, and water at the soil surface equaled the cumulative inflow. This indicates that mass was conserved during infiltration.

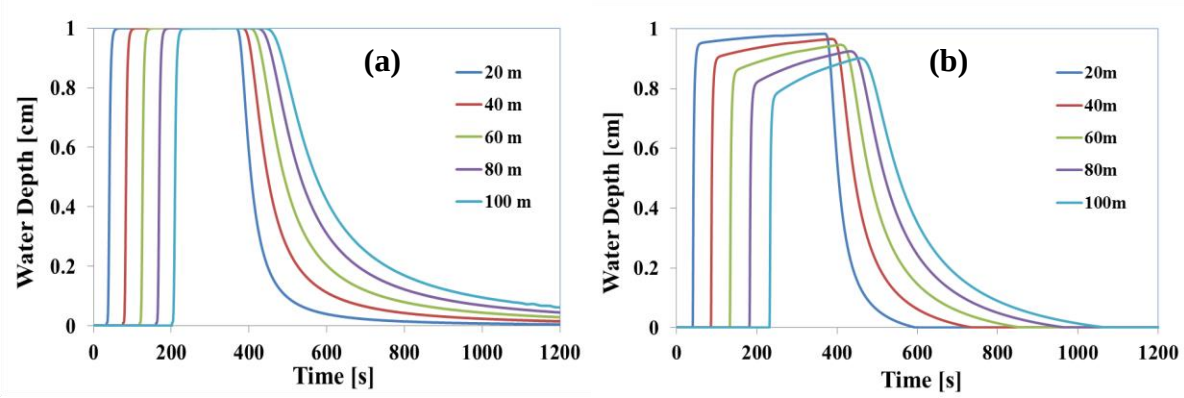


Figure 2.2. Water depths at different locations for the uniform flow model: **(a)** no infiltration; **(b)** infiltration. Parameter values are given in Table 2.2.

The transport of reactive contaminants requires considering the advective-dispersive transport with a range of biogeochemical processes such as adsorption, desorption, volatilization (i.e., mass transfer between different phases), degradation, precipitation, dissolution, etc. To demonstrate the applicability of our reactive transport code, the numerical examples of solute transport with different sorption processes are presented in Fig. 2.3 for the same water flow conditions as in Figure 2.2a. Solute transport was induced using a third-type boundary condition at the inlet with a constant unit concentration for 6 mins. Solute concentrations are shown as a function of time at different locations along the slope. Figure 2.3a shows the simulated transport of a conservative tracer. Similar to water flow (Fig. 2.2a), the tracer reaches the bottom outlet after around 200s. The tracer was never eluted with solute-free water. Consequently, the solute concentration remains constant after the water and tracer inflow stopped due to a decreased volume of runoff water and solute input. Figure 2.3b shows the transport of solute that undergoes linear equilibrium sorption. In this case, the retardation coefficient equals $1+K_D/h$; where K_D was set equal to 1 cm. In comparison to the conservative tracer (Fig. 2.3a), the transport of the solute undergoing linear equilibrium sorption is delayed. The retardation factor is equal to 2 at the maximum water depth of 1 cm, but drastically increases as h goes to zero. Consequently, the delay in solute transport (Fig. 2.3b) becomes more pronounced during the receding limb of the hydrograph (Fig. 2.2a). Figure 2.3c shows the transport of a solute that undergoes kinetic sorption (sorption rate of 0.01 s^{-1}) and desorption (desorption rate of 0.001 s^{-1}). The solute concentration continuously decreases as is it transported over the soil surface during inflow due to kinetic sorption.

The solute concentration rapidly increases after inflow ceases due to a decrease in volume of surface water and continued solute desorption from the land surface. Eventually, the solute concentration reaches an equilibrium concentration level after inflow ceases. The equilibrium aqueous phase concentration is $k_d/h_{min}k_a$ where h_{min} [L] is a minimum water depth.

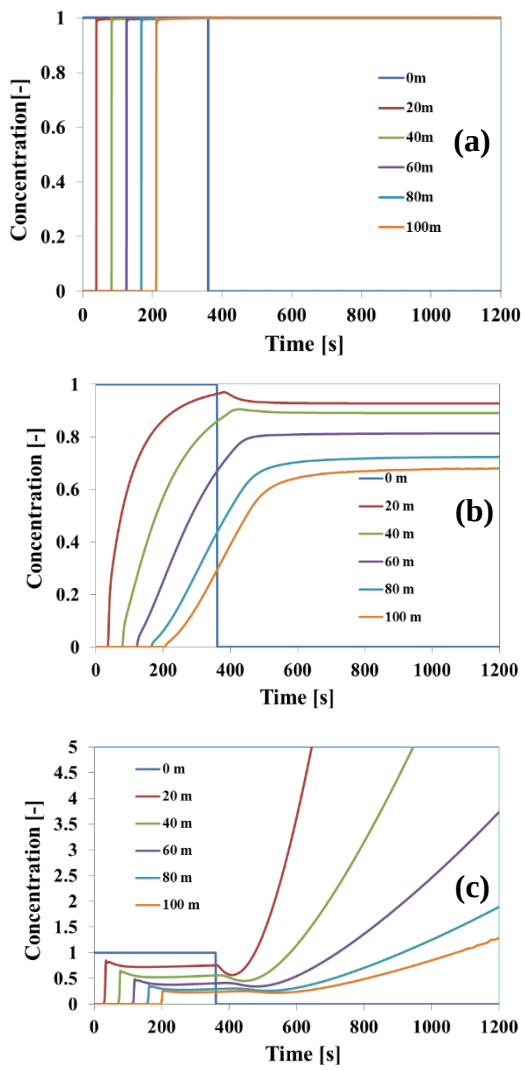


Figure 2.3. Concentrations at different locations for the UFT model: **(a)** conservative tracer, **(b)** solute with linear equilibrium partitioning, and **(c)** solute with kinetic sorption and desorption. Parameter values are given in Table 2.2.

2.4.2 HMIM Model

The HMIM model has two additional input parameters for overland flow compared with the UFT model: (i) the water mass transfer coefficient, α; and (ii) the ratio of the width of the mobile domain to the total width of the surface domain, w_m. The influence of these two parameters on the outflow rate at the bottom boundary as a function of time is demonstrated in Figure 2.4. Similar to Figure 2.2a, the inlet water depth was equal to 1 cm for 6 min and the soil surface was impervious. A constant value of w_m=0.5 and different values of α=0, 0.001, 0.005, and 0.01 s^{-1} are considered in Figure 2.4a. No exchange of water occurs between the mobile and immobile domains when α=0, and the resulting outflow rate is equal to the uniform flow model. Conversely, a fraction of inflow water is transferred to and from mobile and immobile regions when α>0. This process slows down overland flow and delays the arrival of the water front relative to the uniform model. It also produces prolonged tailing in the outflow rate compared with the uniform model. Increasing α produces a faster equilibration of water depths between the mobile and immobile zones and causes a greater delay in outflow. It should be mentioned that the relative importance of α on outflow will increase with smaller values of w_m. A constant value of α=0.001 s^{-1} and different values of w_m=0.3, 0.5, and 0.7 were considered in Figure 2.4b. Increasing w_m produces a slightly earlier outflow arrival time, a higher outflow rate, and less outflow tailing because of the larger mobile domain area. The relative importance of w_m increases with smaller values of α.

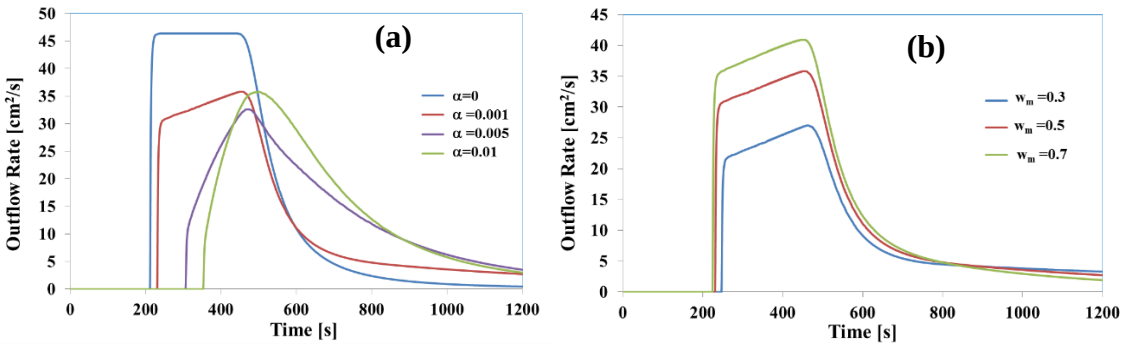


Figure 2.4. Outflow rates at the bottom boundary simulated using the **HMIM Model** with different values of α (s^{-1}) (w_m=0.5) (a) and w_m (α =0.01s^{-1}) (b). Parameter values are given in Table 2.2.

In contrast to the uniform transport model, the HMIM model also depends on the solute mass transfer coefficient (ω_m). Figure 2.5 presents simulated breakthrough curves (BTCs) at the bottom boundary for a 30 min conservative tracer pulse when using the HMIM model with w_m=0.5 and ω_m =0, 0.001, 0.005, and 0.01 s^{-1}. Steady-state overland flow conditions were considered in these simulations to isolate the influence of ω_m from water exchange (e.g., α). The BTC is equivalent to the uniform transport model when ω_m=0. Increasing ω_m produces greater amounts of diffusive solute exchange and a faster equilibration between mobile and immobile domains. Consequently, increasing ω_m produces a greater delay in the initial tracer breakthrough, but also less long-term concentration tailing for the considered parameters. It should be mentioned that the relative importance of ω_m on BTCs will increase with smaller values of w_m.

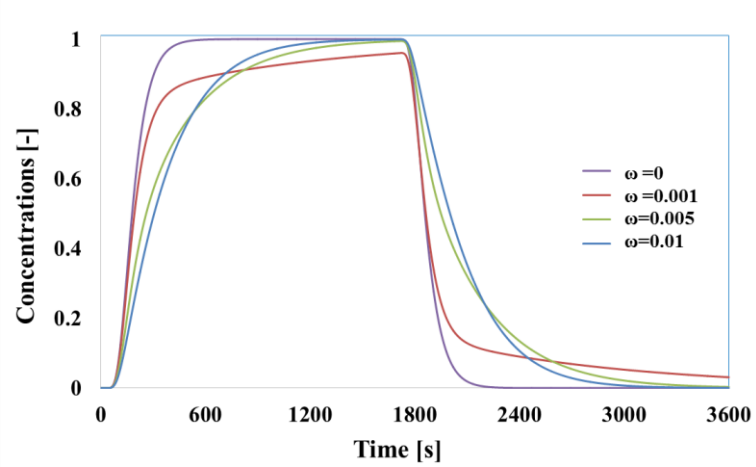


Figure 2.5. Breakthrough curves calculated using the **HMIM Model** with different solute exchange rates ω_m (s^{-1}). Parameter values are given in Table 2.2.

2.4.3 VMIM Model

The VMIM model allows the consideration of depression storage by assigning different values of the maximum water depth for the immobile region (h_{im}^{max}). Figure 2.6a shows plots of the water depth as a function of time at the bottom boundary when using the VMIM model with values of h_{im}^{max} =0, 0.3, and 0.5 cm. In this case, rainfall was uniformly applied over the soil surface for 20 mins at a rate of 0.4 cm min^{-1} and infiltration was also considered (Table 2.2). Figure 2.6a demonstrates that the water level rises from the land surface up to h_{im}^{max} , and that overland flow is initiated when $h_T > h_{im}^{max}$. Overland flow occurs immediately when h_{im}^{max} =0 cm, otherwise overland flow is delayed until the immobile, depression storage, zone is filled. As we expected, higher values of h_{im}^{max} resulted in a greater delay in overland flow. The model also simulated the dynamics

of overland flow and infiltration for 40 min after rainfall ceased. In particular, overland flow continued when $h_T > h_{im}^{max}$, whereas infiltration occurred as long as $h_T > 0$.

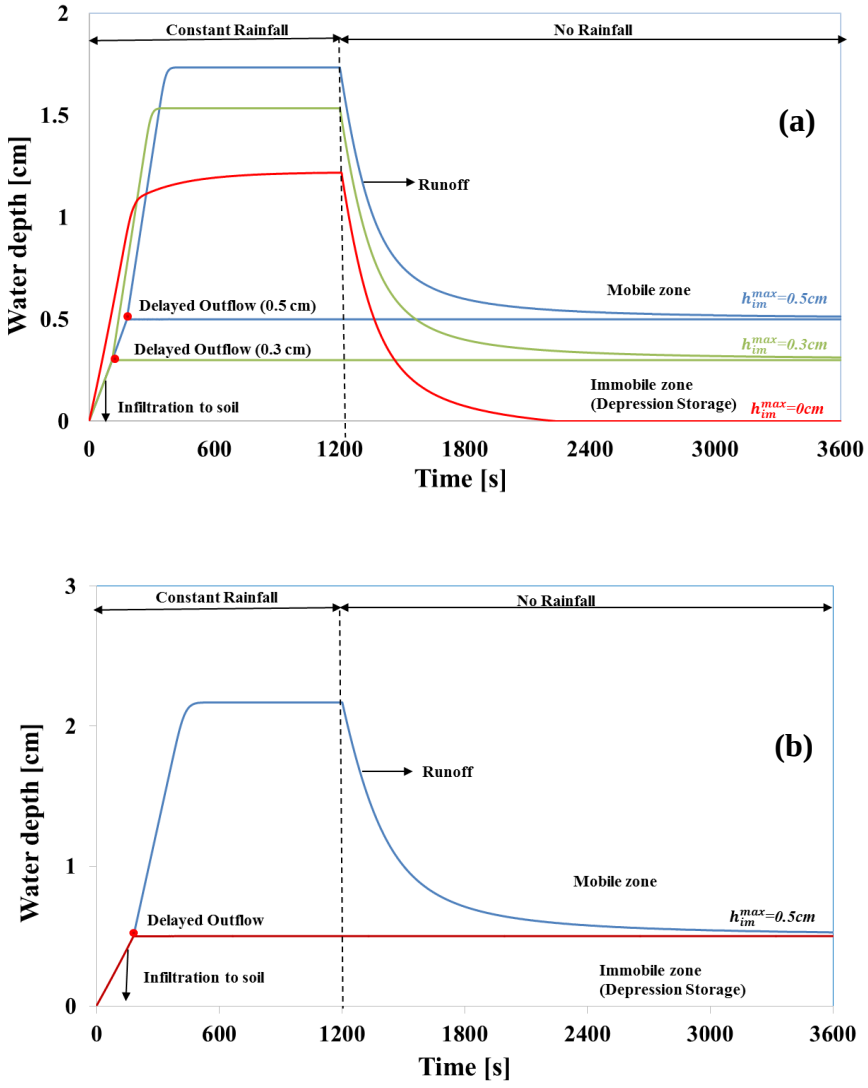


Figure 2.6. Water depths at the bottom boundary simulated using the **VMIM model** (a) and **the APR-V model** (b) with different maximum water depths for the immobile region. Parameter values are given in Table 2.2.

2.4.4 APR Model

Overland flow and transport occurs in two parallel surface regions in the APR model. Water exchange between the active and passive regions is determined by the mass transfer coefficient α_{12}. Separate Manning's roughness coefficients are employed in the surface active (n_1) and passive (n_2) regions to obtain different velocities. Figure 2.7a presents simulated active, passive, and total outflow rates as a function of time when using the APR model with w_A=0.5, n_1=0.01, n_2=0.05, and α_{12}=0, 0.001 and 0.01 s^{-1}. Overland flow over an impervious soil surface was initiated by setting the inlet boundary to a water depth of 1 cm for 20 min. Water moves independently in active and passive domains when α_{12}=0 (red lines). In this case, two outflow peaks are observed with outflow starting after about 200 s and 1000 s in the active and passive domains, respectively, due to differences in the roughness coefficient. An increase in α_{12} produces faster water exchange and equilibration that cause the outflow peaks for the active and passive regions to decrease and increase, respectively. The APR model approaches the behavior of the uniform flow model at the average velocity when α_{12} is very high. It should be mentioned that an increase in n_2 decreases the water velocity in the passive region and the APR model, therefore, approaches the HMIM model when n_2 is very larger. Figure 2.7a illustrates that classic non-equilibrium flow behavior (an early arrival, multiple peaks, and long-term tailing in the outflow rate) can be obtained with the APR model.

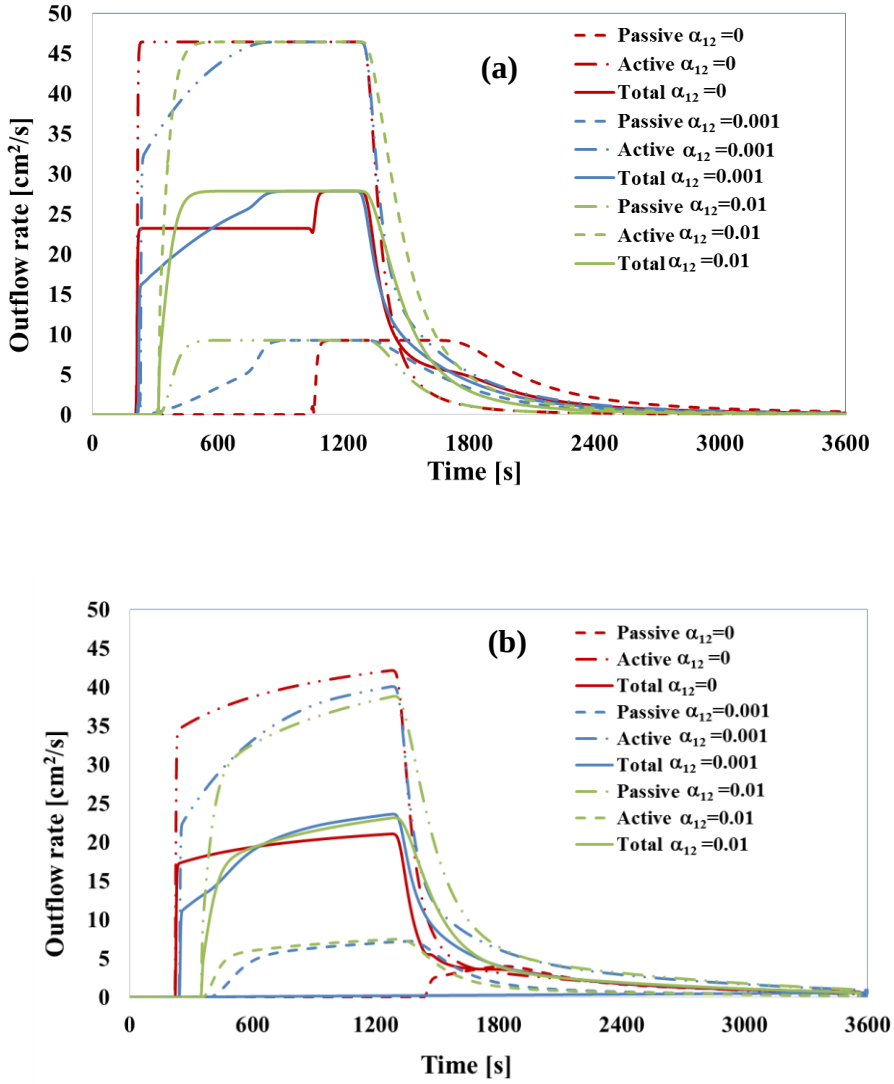


Figure 2.7. Outflow rates at the bottom boundary calculated using the **APR model** (a) and the **APR-H model** (b) with different water exchange coefficients α_{12} (s^{-1}). Parameter values are given in Table 2.2.

Figure 2.8a presents simulated active (dash with dot), passive (dashed line), and total (solid) BTCs at the bottom boundary for a 40 mins conservative tracer pulse when using the APR model with w_A=0.5, n_1=0.01, n_2=0.05, and ω_{12}=0, 0.001, and 0.01 s^{-1}. Steady-state overland flow conditions over an impervious soil surface (an inlet water

depth of 1 cm) were considered in these simulations to isolate the influence of ω_{12} from water exchange (e.g., α_{12}). When the exchange coefficient is equal to zero (red lines), solute moves independently through each of the two surface domains. In this case, the total BTC is the weighed superposition of BTCs from the active and passive regions. Increasing ω_{12} causes greater amounts of diffusive solute exchange and a faster equilibration between the active and passive domains. This produces a decrease and an increase in effluent concentrations in the active and passive regions, respectively. When ω_{12} is very high the BTCs from the two domains converge and resemble the results of the uniform transport model at the average water velocity (Purple line in Fig. 2.4a).

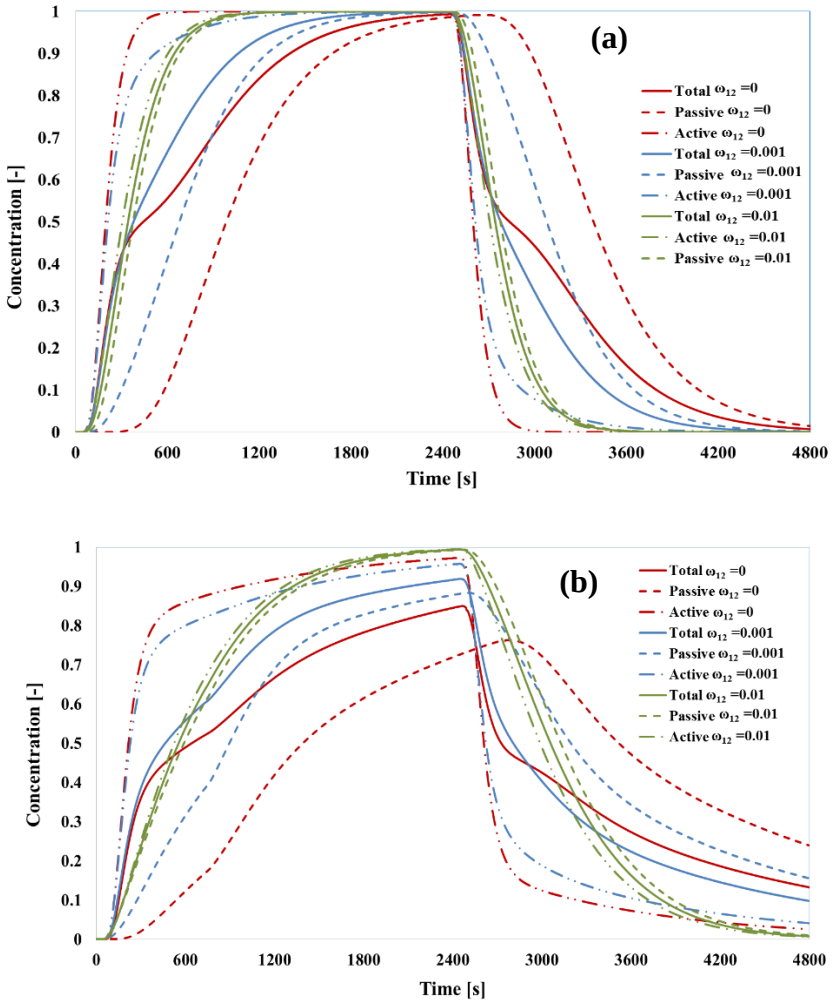


Figure 2.8. Breakthrough curves calculated using the **APR model** (a) and the **APR –H model** (b) with different solute exchange coefficients ω_{12} (s^{-1}). Parameter values are given in Table 2.

2.4.5 APR-V Model

A simulation associated with Figure 2.6a (the VMIM model) was repeated for the combined APR and VMIM (APR-V) model in Figure 2.6b. Additional APR parameters for this simulation included: w_A=0.5, α_{12}=0.1 s^{-1}, n_1=0.01, n_2=0.05, and $h_{1im}^{\max} = h_{2im}^{\max} = 0.5$ cm. Similar to the VMIM model, overland flow in the APR-V model does not occur until

the depression storage zone is filled. Overland flow may then occur at faster and slower rates in the active and passive regions, respectively. The water depth will, therefore, be higher in the passive than the active domain for the same rainfall and infiltration rates in both regions. Consequently, the steady-state water depth at the outlet is higher in the APR-V model than in the VMIM model when α_{12}=0. However, increases in α_{12} cause exchange of water between the passive and active regions, such that the steady-state water depths approach each other when α_{12} is very high.

2.4.6 Solute Wash-off

Figure 2.9a and 2.9b shows plots of a runoff hydrograph and solute wash-off from the impervious soil surface, respectively, when using the several model formulations (UFT, HMIM, APR, and APR-H). In this case, the same inflow (1 cm constant head for 6 min), a runoff flow rate (45 cm^2s^{-1}), the initial soil solute concentration (unit concentration), and solute desorption rate (0.01 s^{-1}) were considered. Other model parameters included: (i) the UFT model with n=0.01; (ii) the HMIM model with w_m=0.5, α =0.01, and ω_m=0.001 s^{-1}; (iii) the APR model with w_A=0.5, n_1=0.005, n_2=0.02, α_{12}=0.001 s^{-1}, and ω_{12}=0.001 s^{-1}; and (iv) the APR-H model with w_A=0.5, w_{1m}=0.5, w_{2m}=0.5, n_1=0.005, n_2=0.02, α_1=0.01 s^{-1}, α_2=0.01 s^{-1}, ω_1=0.01 s^{-1}, ω_2=0.01 s^{-1}, α_{12} =0.001 s^{-1}, and ω_{12} 0.001 s^{-1}. Drastically different hydrographs and solute wash-off behavior are observed when using the various physical non-equilibrium model formulations in comparison to the uniform model for the same initial and runoff flow rate. Clearly, the physical non-equilibrium models can produce hydrographs and/or solute wash-off with

earlier or delayed arrivals, multiple peaks, and prolonged tailing.

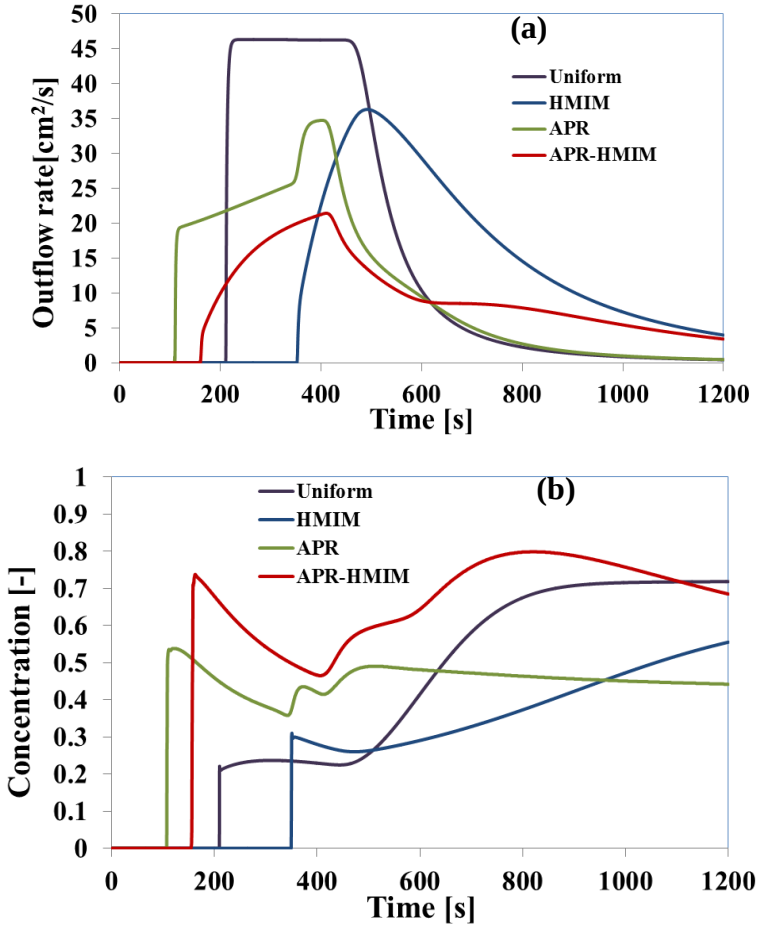


Figure 2.9. Simulated outflow rates (a) and concentrations (b) at the bottom boundary using equilibrium (UFT) and non-equilibrium flow and transport models (HMIM, APR, and APR-H).

2.5 Validation of Physical Non-Equilibrium Models

This section first presents visual evidence for physical non-equilibrium phenomenon during overland flow and transport. The UFT and HMIM models are then used to describe published tracer transport data from this system [Bradford et al., 2015].

Finally, a comparison is made between simulation results from several 1D models (UFT, HMIM, and APR) and a 2D model domain having spatial variations in the Manning roughness coefficient.

A dye tracer experiment was performed in a laboratory runoff chamber. The chamber was 2.25 m long, 0.15 m wide and 0.16 m high. Autoclaved ultrapure quartz sand was uniformly packed into the chamber to a depth of 0.1 m. The chamber slope was then set to 11.8%. Steady-state water flow at a rate of 124 ml min^{-1} was achieved in the chamber before initiating a dye tracer experiment using a peristaltic pump connected to a rain simulator at the upslope portion of the chamber. A step pulse (30 minutes) of 10 ppm yellow-colored fluorescent dye tracer was subsequently pumped to the rain simulator at the same flow rate. Figure 2.10 shows several pictures of water and dye movement at the soil surface during this experiment. The water flow and dye was clearly not uniformly distributed (sheet flow) over the soil surface. The concentrated flow mostly occurred at the edge of the chamber, whereas other areas had little or no overland flow. These observations provide visual evidence that physical non-equilibrium processes were significantly contributing to overland water flow and tracer transport in this experiment.

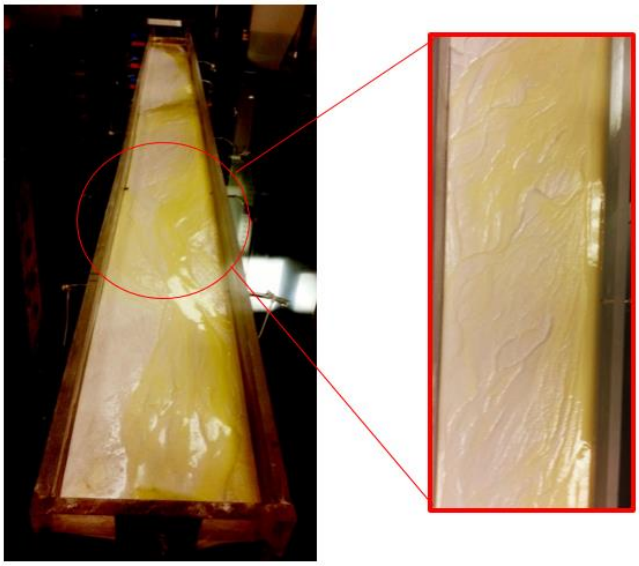


Figure 2.10. Illustrative pictures of the distribution of fluorescent dye tracer and surface water in a 2.25 m long, 0.15 m wide, and 0.16 m high runoff chamber with a 11.8% slope.

Bradford et al. [2015] presented salt tracer (100 mM NaCl solution) data from this same runoff chamber setup when the slope was 5.6, 8.6, and 11.8%. In this case, the eluted solution was collected at the toe of the slope using a fraction collector every 3 mins for 30 mins, then the chamber was eluted with deionized water for another 90 mins. The UFT and the HMIM models were used to simulate the tracer breakthrough curves (BTCs) when the Manning's coefficient was 0.02, the dispersivity was taken as 0.1 of the chamber length [Gelhar et al., 1985, 1992], and other model parameters (w_m and ω_m) were determined by inverse optimization. The fitted parmeters were w_m =0.39E-02, 0.22E-02, and 0.12E-02 and ω_m=0.37E-03, 0.44E-03, and 0.59E-03 s^{-1} for the slopes of 5.6%, 8.6%, and 11.8%, respectively. Figure 2.11a presents observed and simulated BTCs. The values of the Pearson's correlation coefficient (R^2) for the HMIM models at

the slopes of 5.6%, 8.6%, and 11.8% were 0.94, 0.92, and 0.90, respectively. The simulation results indicated that the HMIM model provided a reasonable fit to the experimental data, whereas the UFT model was unable to accurately describe the initial pulse and the prolonged tailing behavior. The value of w_m (the ratio of the width of the mobile region and the total width of the soil surface) were generally very small and tended to decrease with increasing chamber slope. The higher slope generates more dynamic water flow and apparently creates a wider range of immobile regions to store the surface water. A greater rate of diffusive mass transfer (ω_m) also occurred with increasing slope. Figure 2.11b shows an enlarged snapshot of the simulated effluent tracer concentrations during the first 60 s. The HMIM model generates earlier tracer arrival compared to the UFT model. It should be mentioned that these differences in the arrival time between HMIM and UFT models will increase with the domain scale.

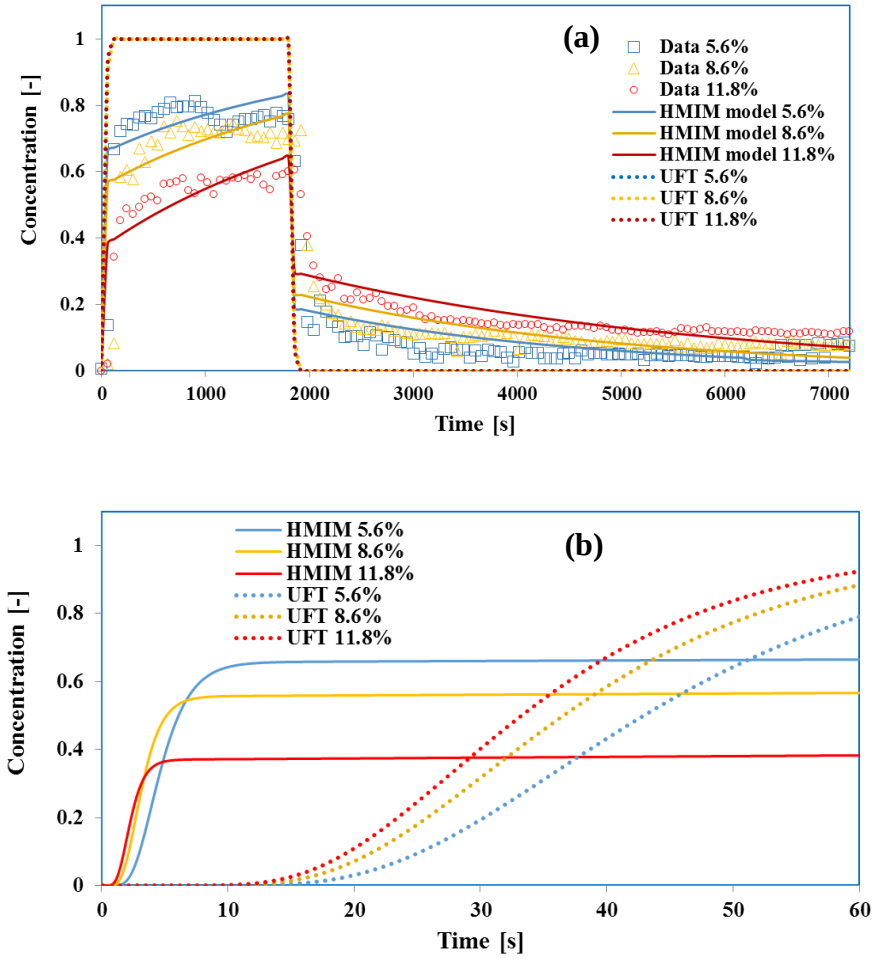


Figure 2.11. Observed and simulated runoff breakthrough curves for a 100mM NaCl tracer when the runoff chamber slope was 5.6%, 8.6%, and 11.8% (a). Simulations for the HMIM (solid lines) and UFT (dashed lines) models are shown. An enlarged snapshot of simulated BTCs during the first 60s is also provided (b).

Most process-based overland flow and transport models neglect local scale parameter heterogeneity by employing average values of the slope and Manning roughness coefficient over the entire surface [Wallach and Van Genuchten, 1990; Deng et al., 2005]. This approach significantly reduces the input parameter requirements and computational time, but may sometimes result in a poor description of experimental data.

The developed HMIM and APR models allow us to simply represent spatial variability in flow and transport parameters in a simple 1D domain with a limited number of parameters. As an illustration, simulation results from several 1D models (UFT, HMIM, and APR) and a 2D model domain having spatial variations in the Manning roughness coefficient were compared. In this case, HYDRUS-2D was similarly adapted as HYDRUS-1D to simulate uniform (sheet) overland flow and transport over a 2D domian. A hypothetical spatially varying roughness coefficient was generated using the HYDRUS-2D GUI in a 10 m X 100 m simulation domain. The roughness coefficient at the edge of the domain was set to 0.005 and rest of the domain was set to 0.1 (Fig. 2.12a) to achieve spatial variations in the overland flow field. As expected, solute transport was much faster in the domain with a smaller roughness coefficient (Fig. 2.12b). The integrated BTCs at the bottom boundary of the 2D domain were subsequently calculated and analyzed using various 1D (UFT, HMIM, and APR) models.

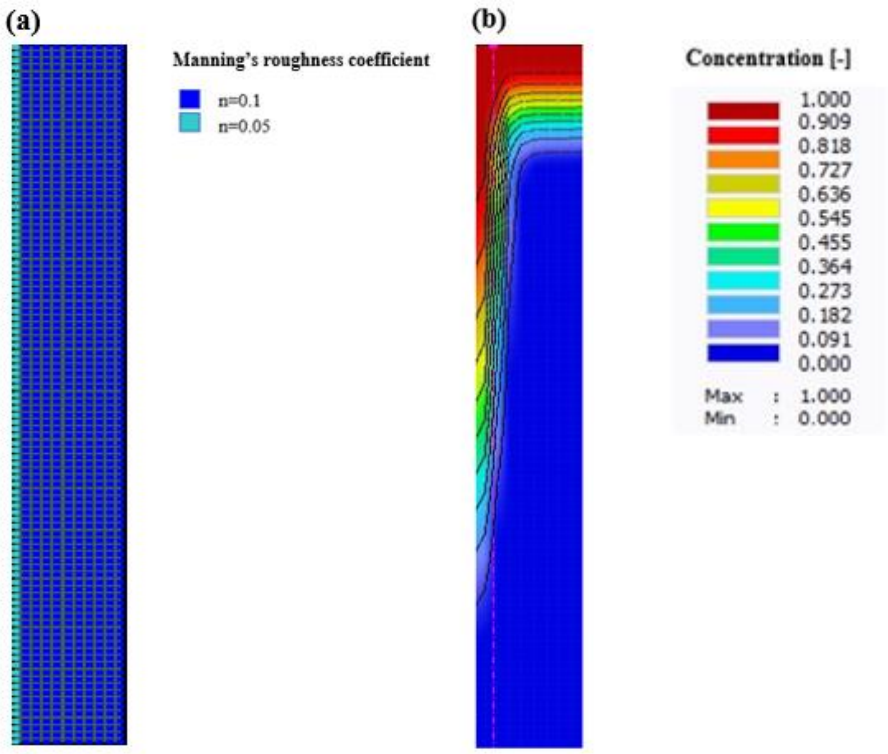


Figure 2.12. The 2D simulation domain showing the distribution of Manning's roughness coefficient (a) and the graphical output of normalized concentration value at time equals to 4060 s (b) when using a version of HYDRUS-2D that was adapted to simulate overland flow and transport in a similar manner to the UFT model.

Figure 2.13 shows the observed (average effluent concentration from the 2D simulation) and simulated BTCs from the 1D models. The 1D UFT model with an averaged Manning's roughness coefficient (n=0.1) provided a poor description of the average 2D data. In contrast, the HMIM and especially the APR models provided a much better description of the average 2D data. The Manning's coefficient set in the APR models are n_1=0.005 and n_2=0.1 for active domain and passive domain, respectively. The values of w_A and ω_{12} were determined by inverse optimization, they were 0.05 and 0.53E-

03 s^{-1}, respectively. For the HMIM model, the Manning's coefficient in mobile domain set equal to 0.005, the inversed w_m and ω_m were equal to 0.09 and 3.50E-03 s^{-1}, respectively. In contrast, the HMIM and especially the APR models provided a much better description of the average 2D data. It should be mentioned that distribution of Manning roughness coefficients in the 2D simulation was selected to achieve physical non-equilibrium flow behavior. Additional research is needed to systematically study and assess the ability of 1D physical non-equilibrium models to describe overland flow and transport on 2D and/or 3D domains with heterogeneous soil surface properties.

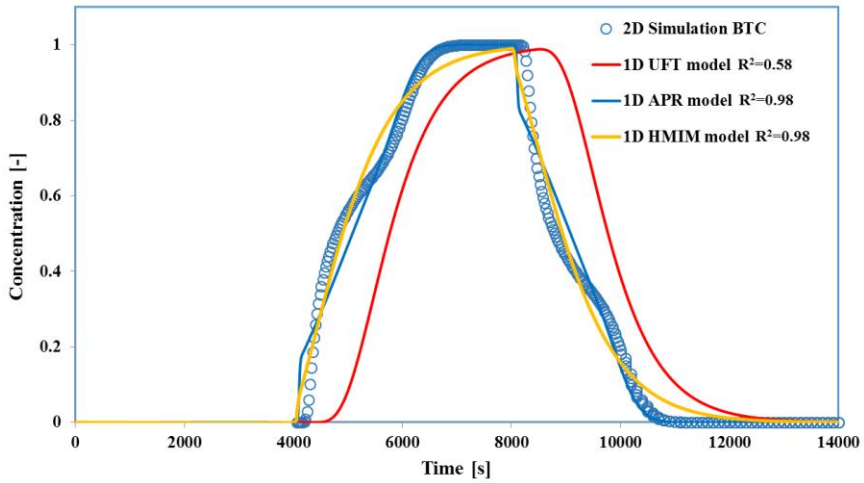


Figure 2.13. Observed (average effluent concentration from the 2D simulation shown in Fig. 12) and simulated BTCs from the 1D UFT, HMIM, and APR models.

2.6 Summary and Conclusion

The popular HYDRUS-1D code was extended to simulate uniform and physically non-equilibrium overland flow and reactive solute transport (such as salts, nutrients, pesticides, and microbes). This code provides information on the temporal and

spatial distribution of water depths and solute concentrations in different phases (e.g., liquid and solid) and regions (e.g., mobile and immobile zones, active and passive regions), as well as on mass balance on the soil surface, inflow, rainfall, infiltration, and evapotranspiration. These models provide a comprehensive set of tools to numerically investigate many important research problems involving overland flow and reactive transport processes. Physical non-equilibrium models may be better suited for studying hydrological processes at the plot and field scale than equilibrium models when spatial patterns of land surface characteristics are poorly characterized. However, the physical non-equilibrium models also involve a relatively large number of parameters which may need to be obtained by calibrating against laboratory or field measurements. The HYDRUS-1D model includes provisions to estimate these parameters by inverse optimization as demonstrated above in the "Validation" section, and a variety of objective functions can be considered based on different measurements (e.g., water fluxes, water depths, resident concentrations, flux concentrations, etc.). Ideally, the model selection should be based on calibration with experimental data. If one model does not provide a satisfactory description of the data, then another model with an additional level of complexity may be considered. The simplest model (with the fewest number of fitting parameters) that accurately describes experimental data is generally preferred.

Additional modifications to the code are needed to consider other infiltration equations, and the full coupling between runoff water and the subsurface.

2.7 References

Ahuja, L.R., Timlin, D.J. and Heathman, G.C., 1991, August. Modelling soluble chemical transfer from soil to overland flow and its transport through macropores to groundwater. In *Hydrological Basis of Ecologically Sound Management of Soil and Groundwater, Proc. Vienna Symp*, 163-172.

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development 1. *JAWRA Journal of the American Water Resources Association*, 34(1), 73-89.

Bathurst, J.C. and O'connell, P.E., 1992. Future of distributed modelling: the Systeme Hydrologique Europeen. *Hydrological processes*, 6(3), 265-277.

Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, 24(1), pp.43-69.

Bittelli, M., Tomei, F., Pistocchi, A., Flury, M., Boll, J., Brooks, E.S. and Antolini, G., 2010. Development and testing of a physically based, three-dimensional model of surface and subsurface hydrology. *Advances in Water Resources*, 33(1), 106-122.

Bradford, S.A., Headd, B., Arye, G. and Šimůnek, J., 2015. Transport of E. coli D21g with runoff water under different solution chemistry conditions and surface slopes. *Journal of Hydrology*, 525, 760-768.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N. and Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological applications*, 8(3), 559-568.

Cea, L., Legout, C., Darboux, F., Esteves, M. and Nord, G., 2014. Experimental validation of a 2D overland flow model using high resolution water depth and velocity data. *Journal of hydrology*, 513, 142-153.

Chen, L., Sela, S., Svoray, T. and Assouline, S., 2013. The role of soil-surface sealing, microtopography, and vegetation patches in rainfall-runoff processes in semiarid areas. *Water Resources Research*, 49(9), 5585-5599.

Camporese, M., Paniconi, C., Putti, M. and Orlandini, S., 2010. Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research*, 46(2).

Deng, Z.Q., De Lima, J.L. and Singh, V.P., 2005. Transport rate-based model for overland flow and solute transport: Parameter estimation and process simulation. *Journal of Hydrology*, *315*(1-4), 220-235.

Donigian Jr, A.S. and Beyerlein, D.C., 1993. Watershed modeling of agrichemicals in Walnut Creek, IA: a plan for preliminary assessment of contributions and impacts. *In: The International Symposium on Water Quality Modeling, Proceedings of the International Symposium,* 192-201.

Dunkerley, D.L., 2003. Determining friction coefficients for interrill flows: the significance of flow filaments and backwater effects. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 28(5), 475-491.

Flanagan, D.C., Ascough J.C., Nicks, A., Nearing, M. and Laflen, J.M., 1995. *Overview of the WEPP erosion prediction model. Technical documentation WEPP* (No. 10). Report.

Furman, A., 2008. Modeling coupled surface–subsurface flow processes: A review. *Vadose Zone Journal*, 7(2), 741-756.

Gelhar, L.W., Mantaglou, A., Welty, C., and Rehfeldt, K.R., 1985. A review of field scale physical solute transport processes in unsaturated and saturated porous media. *In: EPRI Topical Report EA-4190, Electric Power Research Institute, Palo Alto, CA.*

Gerke, H.H. and Van Genuchten, M.T., 1993. A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. *Water resources research*, 29(2), 305-319.

Gómez, J.A. and Nearing, M.A., 2005. Runoff and sediment losses from rough and smooth soil surfaces in a laboratory experiment. *Catena*, 59(3), 253-266.

Goodrich, D.C., Burns, I.S., Unkrich, C.L., Semmens, D.J., Guertin, D.P., Hernandez, M., Yatheendradas, S., Kennedy, J.R. and Levick, L.R., 2012. KINEROS2/AGWA: model use, calibration, and validation. *Transactions of the ASABE,* 55(4), 1561-1574.

Guber, A.K., Yakirevich, A.M., Sadeghi, A.M., Pachepsky, Y.A. and Shelton, D.R., 2009. Uncertainty evaluation of coliform bacteria removal from vegetated filter strip under overland flow condition. *Journal of environmental quality,* 38(4),1636-1644.

Haith, D.A. and Shoenaker, L.L., 1987. Generalized Watershed Loading Functions for Stream Flow Nutrients 1. *JAWRA Journal of the American Water Resources Association*, 23(3), 471-478.

Horton, R.E., 1939. Analysis of runoff-plat experiments with varying infiltration-capacity. *Eos, Transactions American Geophysical Union*, 20(4), 693-711.

Hromadka II, T.V., and Lai, C., 1985. Solving the two-dimensional diffusion flow model. *In Hydraulics and Hydrology in the Small Computer Age, Proceedings of the*

Specialty Conference, 555–562.

Kennedy, J.R., Goodrich, D.C. and Unkrich, C.L., 2012. Using the KINEROS2 modeling framework to evaluate the increase in storm runoff from residential development in a semiarid environment. *Journal of Hydrologic Engineering*, 18(6), 698-706.

Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J.O., Fischer, T., Görke, U.J., Kalbacher, T., Kosakowski, G., McDermott, C.I. and Park, C.H., 2012. OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media. *Environmental Earth Sciences*, 67(2), 589-599.

Kollet, S.J. and Maxwell, R.M., 2006. Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7), 945-958.

Leavesley, G.H., R.W. Lichty, B.M. Troutman, and L.G. Saindon. 1983. Precipitation-runoff modeling system-User's manual: *U.S. Geological Survey Water-Resources Investigations Report*, 83-4238, 207

Miller, S.N., Semmens, D.J., Goodrich, D.C., Hernandez, M., Miller, R.C., Kepner, W.G. and Guertin, D.P., 2007. The automated geospatial watershed assessment tool. *Environmental Modelling & Software*, 22(3), 365-377.

Panday, S. and Huyakorn, P.S., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Advances in water Resources*, 27(4), 361-382.

Sadeghi, A.M. and Arnold, J.G., 2002. A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales. In *Total Maximum Daily Load (TMDL): Environmental Regulations, Proceedings of 2002 Conference*, 56.

Singh, V.P., 1996. Kinematic wave modeling in water resources, surface-water hydrology. *John Wiley & Sons.*

Šimůnek, J., 2015. Implementation of Overland Flow into HYDRUS-1D, HYDRUS Software Series 6a, *Department of Environmental Sciences, University of California Riverside, Riverside, CA*

Šimůnek, J. and van Genuchten, M.T., 2008. Modeling nonequilibrium flow and transport processes using HYDRUS. *Vadose Zone Journal*, 7(2), 782-797.

Šimůnek, J., Van Genuchten, M.T. and Šejna, M., 2016. Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone*

Journal, 15(7).

Šimůnek, J., Jarvis, N.J., Van Genuchten, M.T. and Gärdenäs, A., 2003. Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. *Journal of hydrology*, 272(1-4), 14-35.

Smith, M.W., Cox, N.J. and Bracken, L.J., 2011. Modeling depth distributions of overland flows. *Geomorphology*, 125(3), 402-413.

Therrien, R., McLaren, R.G., Sudicky, E.A. and Panday, S.M., 2010. HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. *Groundwater Simulations Group, University of Waterloo, Waterloo, ON*.

Tyrrel, S.F. and Quinton, J.N., 2003. Overland flow transport of pathogens from agricultural land receiving faecal wastes. *Journal of Applied Microbiology*, 94,87-93.

USGS. 1999. The Quality of Our Nation's Waters-Nutrients and Pesticides. *US Geological Survey Circular*. 1225, 82.

Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., Vanderborght J., Young, M.H., Amelung, W., Aitkenhead, M., Allisson, S.D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Geiser T., Govers, G., Ghezzehei, T., Hallett, P., Hendricks Franssen, H.J., Heppel, J., Horn, R., Huisman, J.A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E.C., Schwen, A., Šimůnek, J., Titak, A., van Dam, J., S.E.A.T.M., van der Zee, Vogel H.J., Vrugt, J.A., Wöhling, T. and Young, I.M., 2016. Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone Journal*, 15(5).

Wallach, R. and Genuchten, M.T., 1990. A physically based model for predicting solute transfer from soil solution to rainfall-induced runoff water. *Water Resources Research*, 26(9), 2119-2126.

Wade, A.J., Durand, P., Beaujouan, V., Wessel, W.W., Raat, K.J., Whitehead, P.G., Butterfield, D., Rankinen, K. and Lepisto, A., 2002. A nitrogen model for European catchments: INCA, new model structure and equations. *Hydrology and Earth System Sciences Discussions*, 6(3), 559-582.

Walker, S.E., Mostaghimi, S., Dillaha, T.A. and Woeste, R.E., 1990. Modeling animal waste management practices: impacts on bacteria levels in runoff from agricultural lands. *Transactions of the ASAE*, 33(3), 807-0817.

Weill, S., Mouche, E. and Patin, J., 2009. A generalized Richards equation for

surface/subsurface flow modelling. *Journal of Hydrology*, 366(1-4), 9-20.

Whitehead, P.G., Wilson, E.J. and Butterfield, D., 1998. A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I—model structure and process equations. *Science of the Total Environment*, 210, 547-558.

Woolhiser, D.A. and Liggett, J.A., 1967. Unsteady, one-dimensional flow over a plane—The rising hydrograph. *Water Resources Research*, 3(3), 753-771.

Zhang, W. and Cundy, T.W., 1989. Modeling of two-dimensional overland flow. *Water Resources Research*, 25(9), 2019-2035.

Zhao, L. and Wu, F., 2015. Simulation of runoff hydrograph on soil surfaces with different microtopography using a travel time method at the plot scale. *PloS one*, 10(6).

Chapter 3

Physics-Informed Data-Driven Models to Predict Surface Runoff Water Quantity and Quality in Agricultural Fields

ABSTRACT

Contaminants can be rapidly transported at the soil surface by runoff to surface water bodies. Physically-based models (PBMs), which are based on the mathematical description of main hydrological processes, are key tools for predicting surface water impairment. Along with PBMs, data-driven models are becoming increasingly popular for describing the behavior of hydrological and water resources systems since these models can be used to complement or even replace physically based-models when there is a lack of required data. Here we propose a new data-driven model as an alternative to a physically-based overland flow and transport model. First, we have developed a physically-based numerical model to simulate overland flow and contaminant transport (the HYDRUS-1D overland flow module). Then, a large number of numerical simulations was carried out to develop a database containing information about the impact of various relevant factors on surface runoff quantity and quality. Numerical simulations were conducted to evaluate the influence of different weather patterns, surface topography, vegetation, soil conditions, contaminants, and best management practices on runoff water quality. Finally, the resulting database involving various input/output surface runoff interactions was used to train data-driven models. Traditional Machine Learning and Deep Learning techniques were explored to find input - output functional relations. The results indicated that the Deep Neural Network (DNN) model with two hidden layers performed the best among selected data-driven models. This DNN model accurately predicted runoff water quantity over a wide range in parameters. It also

predicted well runoff water quality for near-equilibrium solute transport over a wide range in parameters.

3.1. Introduction

A variety of agricultural pollutants result from farming and ranching operations (e.g., sediment, nutrients, pathogens, pesticides, metals, and salts) can lead to impairments of local and far-field water quality. These diffuse nonpoint sources can directly harm ecosystem and watershed water quality, and adversely affect our drinking water supply. Agricultural runoff, generated by rainfall or irrigation events, can transfer pollutants accumulated at the land surface into receiving water bodies. This process has been identified as one of the major causes of water impairment in agricultural settings [Carpenter et al., 1998; USGS, 1999; Tyrrel and Quinton, 2003]. The amount of pollutants transported to surface runoff water is dependent on local soil and crop management, climatic conditions, contaminant properties and environmental factors. The discharge of pollutants from agricultural activities to surface waters can be minimized by locally implementing Best Management Practices (BMPs). However, obtaining monitoring data using field investigations is very time consuming and expensive, and associated with many experimental difficulties. Furthermore, simple field observations may be difficult to interpret to obtain a complete picture of potential contaminant transport routes and mechanisms, and to extrapolate findings to other environmental conditions and climates.

Alternatively, physically-based and data-driven models are potential tools for examining and optimizing the effects of land-use changes and BMPs on surface water

quality [Kirkby and Beven, 1979; Donigian et al., 2002; Borah and Bera, 2003; Roz, 2011]. Indeed, numerical experiments can be cheaply conducted using a physically-based model (PBM) to predict runoff water quality over a wide range of agricultural fields, weather patterns, and BMPs. PBMs explicitly account for main hydrologic and contaminant transport processes using mathematical descriptions. Complex numerical or analytical techniques are commonly used to solve these mathematical descriptions, which may require significant computational time as the spatial and temporal scale of the considered problem increases. In contrast to PBMs, data-driven models are based on functional relationships between input and output variables. Artificial intelligence approaches, such as Machine Learning (ML) and Deep Learning (DL) techniques, are increasingly being used in data-driven models to quantify input-output functional relations for complex systems in hydrology [Hsu et al., 1995; Minns and Hall, 1996; Dawson and Wilby,1998; Dibike et al.,1999; Abrahart and See, 2000; Mjolsness and DeCoste, 2001; Govindaraju and Rao, 2001; Solomatine and Ostfeld, 2008; Roz, 2011; Remesan and Mathew, 2014; Bai et al. 2016; Karandish and Šimůnek, 2016; Fang et al., 2017]. Several ML algorithms for modeling rainfall-runoff processes include Linear Regression (LR), K-Nearest Neighbor regression (kNN), Feed-forward Artificial Neural Networks (ANN), and Support Vector Machine (SVM) models [Loague and Freeze, 1985; Karlasson and Yakowitz, 1987; ASCE, 2000ab; Lin et al., 2006; Nourani et al., 2009]. The ANN technique consists of a neural network with a single hidden layer. In contrast, DL data-driven techniques like Deep Neural Networks (DNN) consist of multiple hidden layers to describe more complex, non-linear structures in data. DL has

been successfully applied in many commercial products that are already available to the public, such as computer vision and speech recognition, language translation, and self-driving cars[Hinton et al., 2012; Lecun et al., 2015; Goodfellow et al., 2016]. In contrast to ML techniques, only a few studies have attempted to use DL techniques to model hydrological processes [Marçais and de Dreuzy, 2017; Fang et al., 2017]. However, no existing DL studies have related water flow and contaminant transport in runoff to environmental variables.

In contrast to PBMs, existing data-driven models do not explicitly account for the underlying physics of hydrological processes. The performance and accuracy of data-driven model algorithms therefore vary depending on the training data set, which are frequently restricted to a small range of site specific conditions. To overcome these limitations, we synthetically generated an extensive database using a PBM that uses standard descriptions of overland flow and transport processes that have been extensively verified to properly represent real field processes [Ahuja et al., 1982; Sharda et al., 1994; Wallach et al., 2001; Goodrich et al., 2002]. This numerically generated database contains information about the impact of a wide range of physical factors on surface runoff quantity and quality in agricultural fields. This database was then used in conjunction with ML and DL techniques to develop correlation relationships between model inputs and outputs from a PBM.

The overall objective of this research was to develop data-driven models to predict contaminant loads in runoff water from agricultural fields. Inputs for these models included information about the agricultural field (slope, length, and area), the soils (type,

roughness, initial water content, and initial contaminant concentration), and precipitation events (rainfall intensity and duration). Outputs from these models included the cumulative water volume and the contaminant mass in runoff water. BMPs to minimize pollution loads can be developed by exploring the relationship between model input and output parameters. Multiple data-driven models were tested on their ability to relate model inputs with outputs.

3.2 Database Preparation

3.2.1 Physically-Based Model

The one-dimensional diffusion-wave and advection-dispersion equations, which are commonly used for describing overland flow and solute transport, respectively, have recently been implemented into the popular HYDRUS-1D code [Liang et al., 2017]. This code has previously been used to numerically solve the Richards equation, which simulates water flow, and convection-dispersion equation, which simulates contaminant transport in the subsurface [Šimůnek et al., 2016]. Recent studies indicated that the governing overland flow equation can be written in a similar mathematical formulation as the Richards equation [Hromadka and Lai, 1985; Panday and Huyakorn, 2004; Weill et al., 2009; Liang et al., 2017] so that existing numerical schemes of HYDRUS-1D can be used to solve the following 1D-diffusion wave equation:

$$\frac{\partial h}{\partial t} - \frac{\partial}{\partial x}\left(\frac{k h^{5/3}}{n_{man}\sqrt{S}} \frac{\partial (z+h)}{\partial x} \right) = R - I \qquad [1]$$

where x is a space coordinate in direction of flow [L; where L denotes units of length], t

is time [T; where T denotes units of time], h is the surface water depth [L], R is rainfall or the irrigation rate [LT^{-1}], and I is the infiltration rate [LT^{-1}], S is the mean local slope [-], and k is a unit conversion factor [$L^{1/3}T^{-1}$]. The parameter n_{man} is a Manning's roughness coefficient for overland flow, and it is dimensionless when k=1 $m^{1/3}s^{-1}$ or has units of $TL^{-1/3}$ when k is not equal to unity.

In this work, the infiltration rate I is described using the Green-Ampt infiltration model [Green and Ampt,1911]. This model is a simplified physically-based approach that is based on fundamental physics to describe the infiltration process as a function of the soil suction head, water content, and soil hydraulic conductivity:

$$I = K\left(\frac{\psi\Delta\theta}{F}+1\right) \quad [2]$$

where ψ is the wetting front soil suction head [L], $\Delta\theta$ is the water content difference [-], K is the saturated hydraulic conductivity [LT^{-1}], and F is the cumulative depth of infiltration [L]. The value of $\Delta\theta$ is defined by using the following equations:

$$\Delta\theta = \theta_e(1-s_e) \quad [3]$$

$$s_e = \frac{\theta_i-(n-\theta_e)}{\theta_e} \quad [4]$$

$$\theta_e = n-\theta_r \quad [5]$$

where s_e is the effective saturation [-], θ_e is the effective water content [L^3L^{-3}], n is the porosity [L^3L^{-3}], θ_i is the initial water content [L^3L^{-3}], and θ_r is the residual water content [L^3L^{-3}].

Solute transport in overland flow is usually described using the advection-dispersion equation (ADE) of the form:

$$\frac{\partial hc}{\partial t}+\frac{\partial s}{\partial t}=\frac{\partial}{\partial x}\left(hD\frac{\partial c}{\partial x}\right)-\frac{\partial Qc}{\partial x}+h\phi+Rc_r-Ic \qquad [6]$$

where c is the solute concentration in the aqueous phase [ML^{-3}; where M denotes units of mass], c_r is the concentration in rainfall water [ML^{-3}], s is the sorbed solute concentration at the soil surface area [ML^{-2}], D is the effective dispersion coefficient accounting for both molecular diffusion and hydrodynamic dispersion [L^2T^{-1}], ϕ is a sink/source term that accounts for various zero- and first-order or other reactions [$ML^{-3}T^{-1}$], and Q is the runoff flow rate [L^2T^{-1}]. The parameter Q is given as:

$$Q=hU=-\frac{kh^{5/3}}{n_{man}\sqrt{S}}\frac{\partial\left(h+z\right)}{\partial x} \qquad [7]$$

where U is a depth-averaged velocity [LT^{-1}] calculated using the Manning-Strickler uniform flow formula [Hromadka and Lai, 1985]. The effect of diffusion on the dispersion coefficient can often be ignored and in such case D can be defined as the product of the dispersivity (λ, [L]) and U. Kinetic sorption/desorption between the solid and aqueous phases can be described using the following equation:

$$\frac{\partial s}{\partial t}=\omega\left(K_D c-s\right) \qquad [8]$$

where K_D is the linear equilibrium partition coefficient [L] and ω is the first-order desorption rate coefficient [T^{-1}]. The product of K_D and ω is proportional to the sorption rate coefficient. Details about the numerical solutions of flow and transport models are provided in the HYDRUS-1D manual [Šimůnek et al., 2008].

3.2.2 Numerical Simulations

A large number of numerical simulations was carried out to develop a database containing information about the impact of various relevant factors on surface runoff quantity and quality. In agricultural settings, land surface contaminants can be easily picked up and transported by surface runoff or can seep into the soil by infiltration. Therefore, simulation scenarios considered in this work mainly focused on the transport of various contaminants in overland flow over different soil and land surfaces under different rainfall rates.

At first, numerous water flow simulations were run with various input parameters shown in Table 3.1. In order to build a realistic training database, a wide range of input parameter values was selected. In these simulations, precipitation was distributed uniformly over the land surface and lasted one hour. Historical records of precipitation frequency for the San Jacinto, CA watershed (http://hdsc.nws.noaa.gov/hdsc/pfds) were employed to generate precipitation-frequency estimates that were employed in simulations. Seven different percent probabilities of exceedance were selected: 1%, 2%, 4%, 10%, 20%, 50%, and 100%. Various agricultural field conditions were obtained by assigning different values of the Manning's roughness coefficient, slope, and length. As a result, 150 different field conditions (6 Manning's roughness coefficients × 5 Slopes × 5 Lengths) were considered. Additionally, four different initial water contents were considered: $\max(\theta_r, 0.1)$, 0.2, 0.3, and the saturated water content, for each soil type. Selected Green-Ampt infiltration parameters (soil porosity, residual water content, suction head, and hydraulic conductivity) were taken from Rawls et al. [1983] (Table

3.2). In all water flow simulations, cumulative water volumes at the bottom outlet at 25 evenly spaced times were outputs. The resulting water flow database thus contained 2,310,000 entries (22 soils × 7 rainfall rates × 150 field conditions × 4 initial water contents × 25 print times).

Solute transport simulations were subsequently conducted for the previously described water flow simulations. In these simulations, it was assumed that a unit concentration (c_i = 1 g/cm^2) of solute was initially distributed along the soil profile. Since the solute transport equation is a linear equation, the results obtained for a unit initial solute concentrations can be simply multiplied by other initial concentrations to get corresponding results. Runoff water generated by rainfall events "washed-off" solutes from the land surface and transported them to the bottom outlet of the field. The "wash-off" process was described using different sorption and desorption rates. The exchange of solute concentrations between solid and liquid phases is controlled by the overland water flow rate and kinetic sorption/desorption parameters (ω and K_D) in Eq. [8]. The kinetic sorption model approaches equilibrium conditions when ω is large, whereas non-equilibrium conditions prevail and slow solute release occurs when ω is small. In this study, the solute transport database considered transport processes and solute desorption for near-equilibrium conditions (ω = 8640 day^{-1}) in order to mitigate the complexity of transport processes. Hence, two additional input parameters, variable K_D shown in Table 3.1 and constant ω = 8640 day^{-1}, were accounted for in these simulations. Solute transport simulations generated the output that included the cumulative solute mass at the bottom outlet at 25 evenly spaced times. The inclusion of two additional parameters to

describe solute wash-off (K_D and ω) resulted in a database that was 5 times bigger than the water flow database. Furthermore, the variance in the solute transport database was higher than the water flow database.

Table 3.1. Input parameters used in the numerical simulations.

Input	Values
Rainfall intensity (cm day^{-1})	21.6, 28.8, 40.8, 52.8, 69.6, 86.4, 105.6
Manning's roughness coefficient	0.01, 0.02, 0.04, 0.07, 0.15, 0.24
Slope (-)	0.01, 0.02, 0.04, 0.08, 0.16
Length (m)	30, 60, 120, 240, 480
K_D (cm)	0.1, 0.5, 1, 1.5, 2
Initial water content, θ_i	θ_r, 0.2, 0.3, θ_s

Table 3.2. Green-Ampt infiltration parameters for various soil classes.

Texture	θ_s	θ_r	θ_e	ψ (cm)	K (cm s^{-1})
Sand	0.437	0.020	0.417	4.950	3.272E-03
Loamy Sand	0.437	0.036	0.401	6.130	8.306E-04
Sandy Loam	0.453	0.041	0.412	11.010	3.028E-04
Loam	0.463	0.029	0.434	8.890	9.444E-05
Silt Loam	0.501	0.015	0.486	16.680	1.806E-04
Sandy Clay Loam	0.398	0.068	0.330	21.850	4.167E-05
Clay Loam	0.464	0.155	0.309	20.880	2.778E-05
Silty Clay Loam	0.471	0.039	0.432	27.300	2.778E-05
Sandy Clay	0.430	0.109	0.321	23.900	1.667E-05
Silty Clay	0.470	0.047	0.423	29.220	1.389E-05
Clay	0.475	0.090	0.385	31.630	8.333E-06
Tilled Sand*	0.524	0.024	0.500	5.940	3.927E-03
Tiled Loamy Sand*	0.524	0.043	0.481	7.356	9.967E-04
Tilled Sandy Loam*	0.544	0.049	0.494	13.212	3.633E-04
Tilled Loam*	0.556	0.035	0.521	10.668	1.133E-04
Tilled Silt Loam*	0.601	0.018	0.583	20.016	2.167E-04
Tilled Sandy Clay Loam*	0.478	0.082	0.396	26.220	5.000E-05
Tilled Clay Loam*	0.557	0.186	0.371	25.056	3.333E-05
Tilled Silty Clay Loam*	0.565	0.047	0.518	32.760	3.333E-05
Tilled Sandy Clay*	0.516	0.131	0.385	28.680	2.000E-05
Tilled Silty Clay*	0.564	0.056	0.508	35.064	1.667E-05
Tilled Clay*	0.570	0.108	0.462	37.956	1.000E-05

* Tillage management, multiply the values for the textural class by 1.2 (20% increase)

3.3 Data-driven Models

In this section, we briefly describe the data-driven models used in this paper. They are divided into two broad classes, linear and non-linear models. The selected linear data-driven models include Linear Regression (LR) and Linear Support Vector Machine (SVM-L). Non-linear models include K-Nearest Neighbor Regression (kNN), Support Vector Machines with non-linear kernel (SVM-NL), and Deep Neural Networks (DNN) with different numbers of hidden layers. We used the previously generated synthetic

databases from the PBM to train the selected data-driven models to predict surface runoff water quantity and quality. All these models were implemented using the open-source Scikit-learn and Keras Libraries in Python 3.5 [Pedregosa et al., 2011; Chollet, 2015].

3.3.1 Linear Regression

LR is the simplest ML model, which is often treated as the baseline method. It applies linear mapping to minimize the sum of least square errors between input data (either experimental or (PBM) model-derived) and output data (of the regression model). Suppose we have a set of training dataset $\left(y_i, x_{ij}\right)$, where i =1, 2, 3…n and j =1, 2, 3…k. In here, i is the index for training samples in the dataset, j is the index for the input parameters. A general linear regression problem can be developed by assuming that output variables y_i are influenced by input parameters x_{ij} as:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i \qquad [9]$$

where β_j are regression coefficients, β_0 is intercept, and e_i represent a deviation between actual and predicted values. The ordinary least squares procedure seeks to minimize the total sum of the residuals. However, since this approach treats data as a matrix, the process will be very computational intensive when we have large data sets. Therefore, gradient-based optimization approaches, which optimize the values of the coefficients by iteratively minimizing the sum of the squared errors for each pair of input and output values are used here [Ticlavilca et al., 2011].

3.3.2 Support Vector Machine

The SVM is another widely used model for both classification and regression tasks. The SVM model was originally developed for classification problems to find a hyperplane, which can separate a data set into one class from those of another class [Vapnik, 2013; Vapnik & Vapnik, 1998]. The SVM models have been successfully extended to solve regression problems by the introduction of the so-called ε-insensitive loss function. This loss function is used to penalize errors that are greater than the threshold ε (an insensitivity zone). A version of the SVM for regression problems is called support vector regression (SVR) and it depends on a subset of the training data. The ε-SVR aims to find the flattest regression function $f(x_{ij})$ that has at most a deviation ε from actual y_i for all training points x_{ij}:

$$f(x_{ij}) = \sum_{j=1}^{k} w_j x_{ij} + b_i \qquad [10]$$

where w_j are weights and b_i is the bias term that should be estimated from the training data. One great benefit of using ε-SVR is the use of kernels, which inherently map the data into a non-linear space depending on the chosen kernel function. Therefore, the above regression can be formulated as [Van Looy et al., 2017]:

$$f(x_{ij}) = \sum_{j=1}^{k} w_j \Phi(x_{ij}) + b_i \qquad [11]$$

where $\Phi(x_{ij})$ denotes nonlinear transformation (kernel functions). Several non-linear kernel functions are available, such as polynomial, radial basis, and sigmoid functions. In this study, we tested both SVR with linear (SVM-L) and non-linear (SVM-NL) sigmoid

kernel functions.

The optimization of ε-SVR is more complicated than the LR model because there is a linear constraint. In this case, the penalty factor C and the slack variables ξ_i and ξ_i^* are introduced to cope with infeasible constraints [Vapnik, 2013]. Hence, w_j and b_j are estimated by minimizing the following objective function:

$$\min \frac{1}{2}|w|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad [12]$$

subject to

$$\begin{cases} y_i - f(x_{ij}) - b_i \le \varepsilon + \xi_i^* \\ f(x_{ij}) + b_i - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad [13]$$

where $|w|^2 = \sum_{j=1}^{k} w_j{}^2$. The penalty factor C determines the trade-off between the flatness of the function $f(x_{ij})$ and the amount up to which deviations larger than ε are tolerated. The slack variables ξ_i and ξ_i^* determine the degree to which data points will be penalized if the error is larger than ε.

3.3.3 K-Nearest Neighbor Regression

K-Nearest Neighbor (kNN) regression is another machine learning method for regression problems. The kNN is a non-parametric regression type of analysis. Predictions are made by computing the distance between all the data points in the database and searching for the number of neighbors (K_{nn}) which are most similar to a new instance. Once the nearest-neighbor list is obtained, the prediction of output y is

achieved by assigning weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, the Euclidean distance d between lth training sample x_{lj} and ith test sample x_{ij} is defined as [Friedman et al., 2013]:

$$d(x_{ij,}, x_{lj}) = \sqrt{\sum_{j=1}^{k} (x_{ij} - x_{lj})^2} \quad [14]$$

The weight w_d can be defined by a selected K_{nn} value and a distance between lth and ith samples:

$$w_d(x_{ij}, x_{lj}) = \frac{\exp(-d(x_{ij}, x_{lj}))}{\sum_{l=1}^{K_{nn}} \exp(-d(x_{ij}, x_{lj}))} \quad [15]$$

Thus, the output y_i can be predicted by

$$y_i = \sum_{l=1}^{K_{nn}} w_d(x_{ij}, x_{lj}) y_l \quad [16]$$

3.3.4 Deep Feed-Forward Network

Neural networks (NN) are the most popular ML models used nowadays. The classical feed-forward network is a neural network that has been widely used in the hydrologic and water quality problems [ASCE, 2000ab; Maier and Dandy, 2000; Sharma et al., 2003; Khalil et al., 2011]. The fundamental components of these algorithms include the input layer, intermediate layers, output layers, and neurons. Adjustable parameters characterize the neurons that are connected to the input and transmit the information from one neuron to another in one direction: from an input

layer, through one or more hidden layers, to an output layer. Figure 1 illustrates the adjustable parameters for an individual neuron. In a network, the input variables x_{ij} are weighted and summed up to produce the hidden neurons S_p (p =1, 2, …P):

$$S_p = \sum_{j=1}^{k} w_{jp} x_{ij} \qquad [17]$$

where p is the number of hidden neurons. A neuron computes an output, based on the weighted sum of all its inputs (S_p), according to an activation function ($f(S_p)$). Finally, an activation function which is applied to S_p provides the final output from this logistic sigmoid [Dawson and Wilby, 1998] (see Fig. 3.1). The sigmoid activation function is given as:

$$f(S_p) = \frac{1}{1+e^{-S_p}} \qquad [18]$$

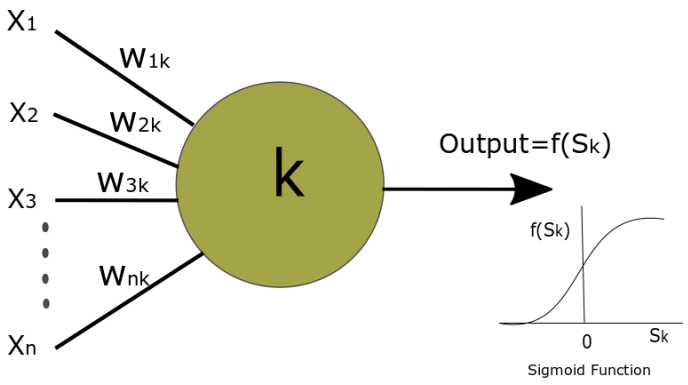


Figure 3.1. Activation of a single neuron.

The so-called Deep Neural Networks (DNN) model is a direct extension of the feed-forward network that contains multiple hidden layers that are stacked as shown in Figure 3.2 [Schmidhuber, 2015]. The DNN model becomes "deeper" as the number of hidden layers increases. Deeper DNN models allow for simulating more complicated, non-linear mapping functions between input and output data. However, DNN models often suffer from overfitting when using too many hidden layers. This problem can be mitigated by dropping layers. It should be noted that the DNN model with one hidden layer is equivalent to the traditional ANN model. We have tested the DNN model with one, two, and three hidden layers to determine the optimum number of hidden layers and to minimize overfitting.

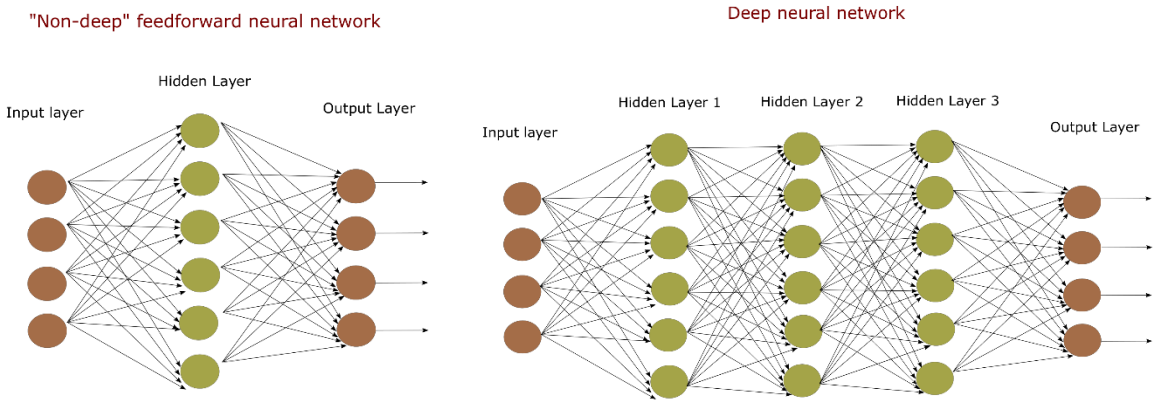


Figure 3.2. A feed-forward (non-deep) neural network (left) and a deep feed-forward neural network (right).

3.4 Model Training and Evaluation

The runoff water database generated using the PBM was divided into training (about 80% of the entire database) and validation (about 20% of the entire database) data sets for the data-driven models. Variable input parameters and output values were used to train the data-driven models by randomly selected 18 out of 22 soil types. After the training was completed, data-driven models were tested on the remaining 4 soil types over many different scenarios. The accuracy and predictive capability were first assessed by making scatter plots of cumulative runoff water volumes (Q_c) for data-driven models and the PBM. The calculated R^2 and the best-fitted regression line were shown in these scatter plots to quantify the goodness of the data-driven model prediction.

In addition, two examples were chosen to show the accuracy between observed and predicted cumulative runoff hydrographs between the PBM and data-driven models, respectively, over a range of initial water contents and soil types. The first example considered four different initial water contents and the following conditions: the rainfall rate of 69.6 cm/day for 1 hr, the surface length of 240 m, a Silt Loam soil, the Manning roughness coefficient of 0.24 (dense grass), and a steep slope of 16%. The total simulation time was 1 hr and 10 minutes so that the water recession process could be fully captured in these simulations. The second example considered surface runoff from Silt Loam, Clay, tilled Silt Clay, and tilled Clay soils under the following conditions: an initially saturated soil, the rainfall rate of 105.4 cm/day for 1 hr, a surface length of 48 m, a slope of 16%, and the Manning roughness coefficient of 0.24.

Statistical measures were used to quantify the ability of various ML and DL models to predict an output from the PBMs. Calculated statistical parameters included the root mean square error (RMSE), the mean absolute error (MAE), the mean bias error (MBE), the model efficiency (EF), and the *Relative Error* Δ_i. These statistical parameters were determined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \quad [19]$$

$$MAE = \frac{\sum_{i=1}^{n}|O_i - P_i|}{n} \quad [20]$$

$$MBE = \frac{\sum_{i=1}^{n}(O_i - P_i)}{n} \quad [21]$$

$$EF = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O_i})^2} \quad [22]$$

$$Relative\ Error\ \Delta_i = \frac{(P_i - O_i)}{\frac{(P_i + O_i)}{2}} \quad [23]$$

where P_i and O_i are the output values predicted by data-driven models and observed data obtained from the PBM, respectively, $\overline{O_i}$ is the average of observed data, and n is the number of observations. It should be noted that data with zero values of surface runoff were removed from this analysis for the purpose of model inter-comparison.

A database for solute wash-off with runoff water (e.g., the cumulative solute mass at the outlet) was developed using the PBM that encompassed a wide range in

rainfall rates, soil physical properties, initial water contents, slopes, field size, Manning roughness coefficients, and kinetic sorption/desorption parameters. As we mentioned before, the solute transport database in this study considered transport processes and solute desorption for near-equilibrium conditions with a constant desorption rate coefficient (ω = 8640 day^{-1}) and different K_D. The solute transport databases had a similar structure as the water flow database, but was 5 times bigger in size.

The contaminant transport is highly related to water flow. The data-driven model with the best performance in predicting the runoff water volume was therefore selected for training on the solute transport databases. Data-driven models were trained against about 80% of these solute transport databases using Q_c, K_D, and other input parameters associated with the water flow database. The remaining 20% of the solute transport databases were employed for validation. Two simulation examples were chosen to further investigate the predictive ability of the trained data-driven models. These simulation scenarios considered the same input variables as for the water flow model (e.g., differences in initial water contents or soil texture) with K_D=2 cm and ω=8640 day^{-1}.

3.5 Results and Discussion

3.5.1 Surface Runoff Quantity

3.5.1.1 Linear Regression

A scatter plot of estimated Q_c using the LR model versus observed Q_c for the PBM validation data set are shown in Figure 3.3a. We can see that R^2 is equal to 0.483,

which indicates that the LR model failed to accurately describe the PBM data. The LR algorithm usually works well with high-bias and low variance data. However, the PBM describes real-world scenarios with complex, non-linear relationships between input and output variables that generate high variance data sets.

The observed (PBM) runoff water volume in Figure 3.3b decreased and was increasingly delayed when the initial soil water content was lower. The observed (PBM) simulations in Figure 3.3c demonstrate that the Clay and tilled Clay textured soils produced the most runoff water because of their lower saturated hydraulic conductivity and infiltration. After rainfall stops, the cumulative runoff water flux continues to increase because water flows slowly over the rough surface before reaching the bottom boundary. The LR model failed to accurately describe runoff water dynamic for these two selected examples, especially when the initial water content was higher (Fig. 3.3b) and the hydraulic conductivity was lower (Fig. 3.3c). In particular, the LR model tended to overestimate Q_c for early times and underestimate Q_c for later times.

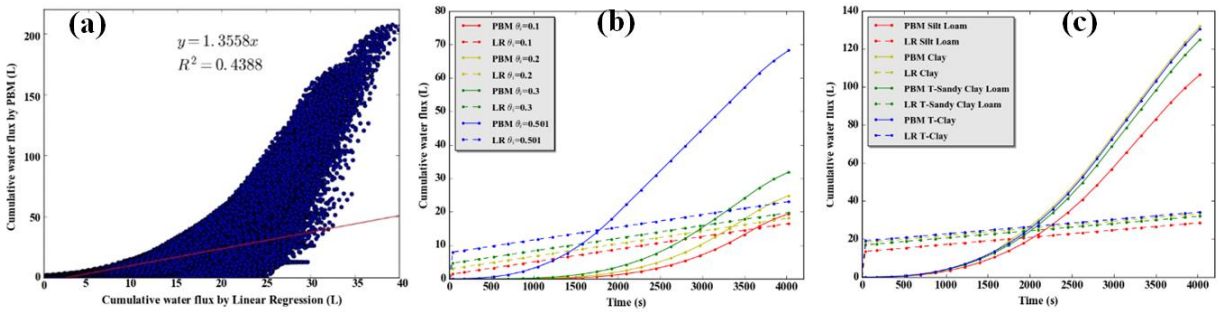


Figure 3.3. Comparison of cumulative water fluxes (Q_c) calculated by the PBM and the Linear Regression (LR) model (a - overall, b - different initial water contents, c - different textures).

3.5.1.2 Support Vector Machine

Similar to the LR model, the Support Vector Machine model with linear kernel functions (SVM-L) produced a linear input-output relationship. Two user-defined parameters for the SVM-L model were the penalty parameter *C* on the error term, which was set to 1, and the maximum number of iterations, which was 2,000 [Joachims, 2002]. Figure 3.4a indicates that this SVM-L model had an even lower R^2 (R^2=0.386) than the LR model (R^2=0.483). Similarly, simulation examples shown in Figure 3.4b (different initial water contents) and 3.4c (different textures) indicate that the SVM-L model completely failed to capture the large values of Q_c at later times. Linear models (LR and SVM-L) are therefore not well suited to describe complex non-linear runoff problems.

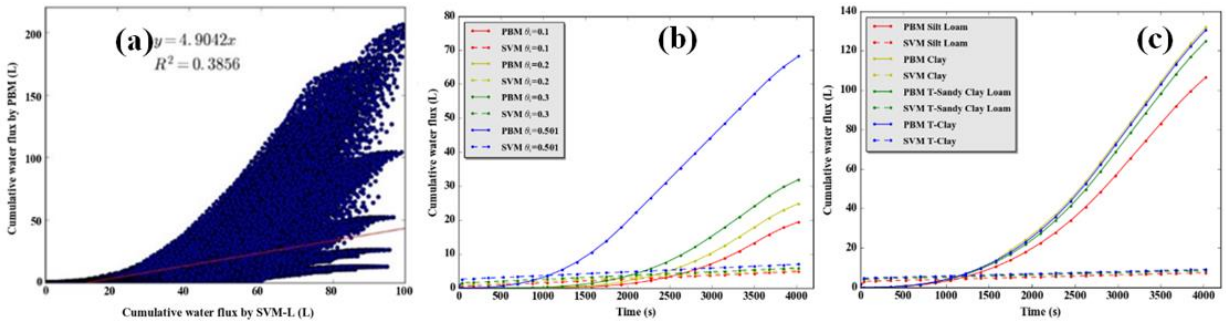


Figure 3.4. Comparison of cumulative water fluxes (Q_c) calculated by the PBM and the Linear Support Vector Machine (SVM-L) model (a - overall, b - different initial water contents, c - different textures).

Many researchers have indicated that the SVM model with non-linear kernels (SVM-NL) can successfully describe surface/subsurface hydrology problems [Dibike et al.,

2001; Bray and Han, 2004; Yu et al., 2006; Karandish and Šimůnek, 2016]. However, previous SVM-NL applications have considered only a relatively small training and validation dataset. Training the SVM-NL model takes a tremendous amount of computational time because of the complexity of this model. For example, we estimated that training the SVM-NL model to our complete runoff water database would take somewhere between a few days and a month. Consequently, only about 20% of the original runoff water training data set were randomly selected to train the SVM-NL model. The trained model was then used for testing against the validation data set with C=1 and ε=0.2. Figure 3.5a shows the scatter plot of estimated Q_c using the SVM-NL model versus observed Q_c from the validation set. The R^2 (R^2=0.931) for the SVM-NL model was much higher than for the LR and SVM-L models. The SVM-NL model also provided a much-improved description of the PBM results for the two examples shown in Figures 3.5b and 3.5c. However, there were still considerable deviations between the SVM-NL model and the PBM, especially for larger print times, higher initial water contents (Fig. 3.5b), and for the Silt Loam soil (Fig. 3.5c).

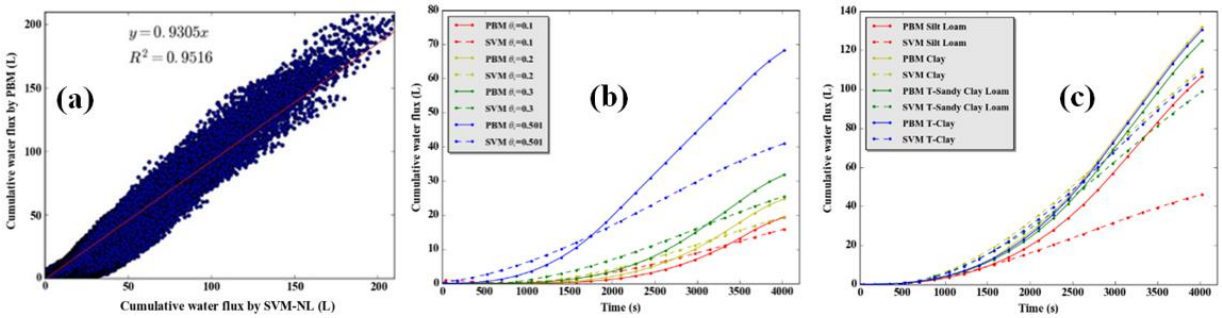


Figure 3.5. Comparison of cumulative water fluxes (Q_c) calculated by the PBM and the Non-Linear Support Vector Machine (SVM-NL) model (a - overall, b - different initial water contents, c - different textures).

3.5.1.3 K-Nearest Neighbor Regression

The kNN models with different numbers of neighbors (the *K* value) were tested while all points in each neighborhood were weighted equally. We observed that the optimal *K* value is equal to 5. The results of calculations with the kNN model are summarized in Figure 3.6. The R^2 (R^2=0.959) for the kNN-5 model was slightly higher than that for the SVM-NL model (R^2=0.9305). However, there were more predicted outliers when the observed Q_c was relatively small. Although the overall runoff trends were correctly captured in Figure 3.6b, the kNN-5 model underestimated the cumulative water flux at every print time and incorrectly predicted that Q_c stopped increasing when rainfall ended. Furthermore, this deviation tended to increase for smaller initial water contents. However, Figure 3.6c shows that the kNN-5 model did a much better job of

predicting large values of Q_c than the SVM-NL model.

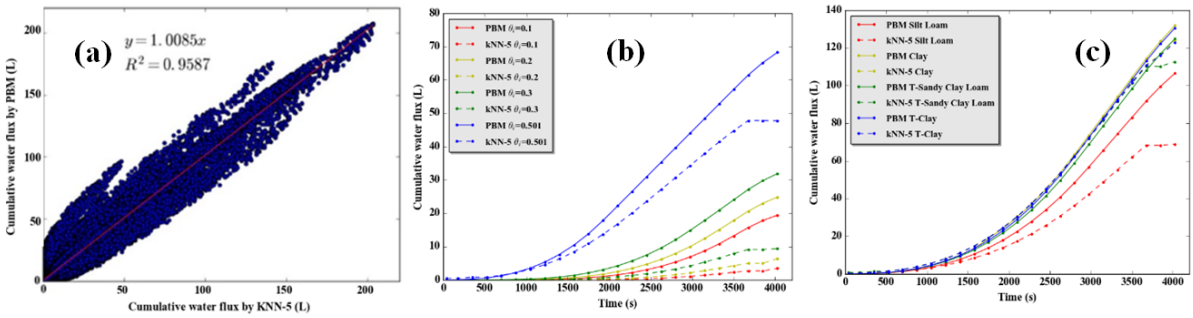


Figure 3.6. Comparison of cumulative water fluxes (Q_c) calculated by the PBM and the k-Nearest Neighbor Regression (kNN-5) model (a - overall, b - different initial water contents, c - different textures).

3.5.1.4 Deep Feed-Forward Neural Network

The DNN models with one, two, and three hidden layers were tested to determine the optimum number of hidden layers and to minimize overfitting. In this study, the optimal number of neurons in each hidden layer was determined by trial and error. We tested possible numbers of neuron from 3 to 20 to see if the model performance was improved. The R^2 and the Root Mean Squared Error (RMSE) were not improved when neuron numbers were increased beyond 16. Therefore, the number of neurons in each hidden layer was set to 16.

Figure 3.7a shows the scatter plot of estimated Q_c using DNN models with three different layers versus observed Q_c for the PBM validation data set. Values of R^2 were equal to 0.975, 0.981, and 0.976 for the DNN-1, DNN-2, and DNN-3 models, respectively. The DNN models provided fairly accurately runoff and obtained higher R^2

values than the other ML models. However, the top and bottom rows of Figure 3.7 indicate that the DNN-1 and DNN-3 models tended to underestimate and overestimate observed values of Q_c, respectively, whereas the DNN-2 model provided the best prediction. This result suggests that a one hidden layer model is not "deep" enough, whereas a three-layer model probably is too complicated and results in overfitting. A more quantitative comparison of various ML models is given below.

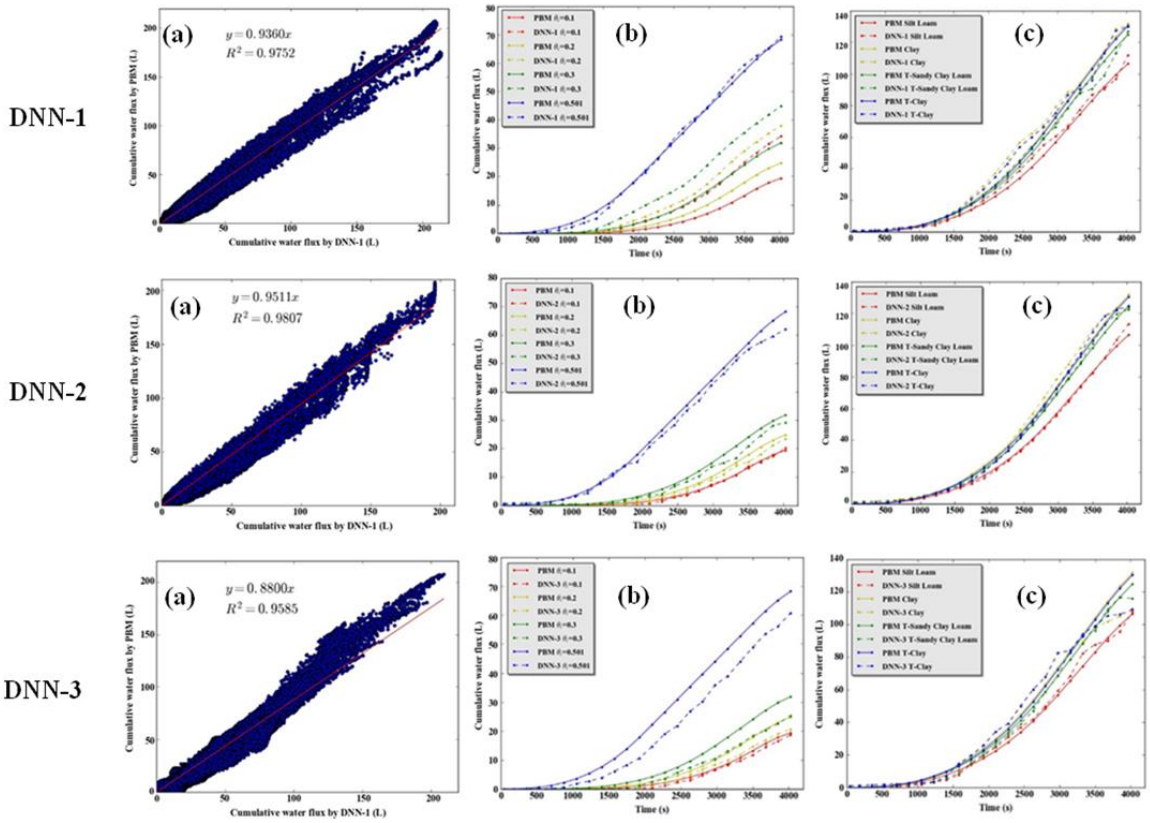
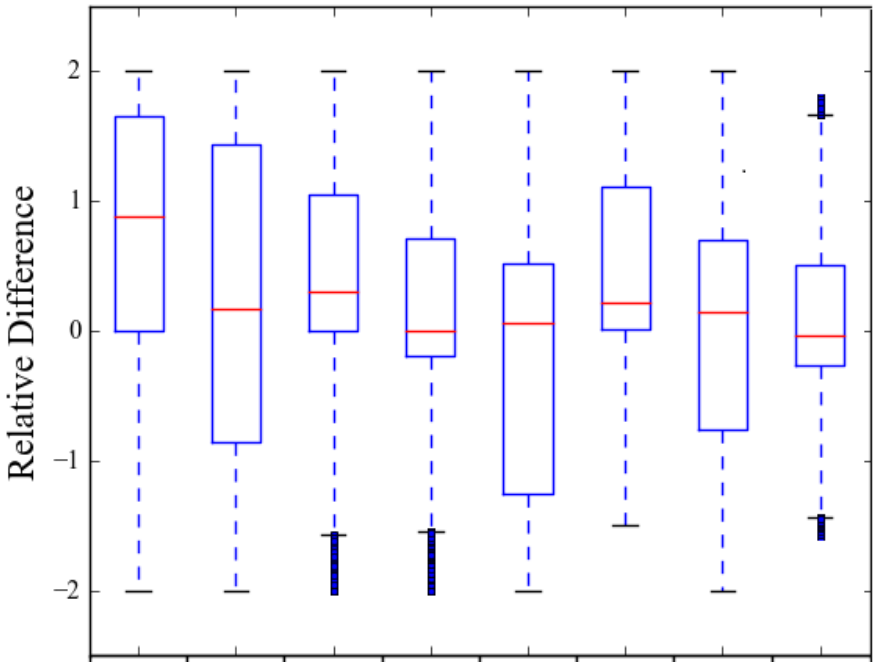


Figure 3.7. Comparison of cumulative water fluxes (Q_c) calculated by the PBM and the Deep Feed-Forward Neural Network (DNN-1 (top row), DNN-2 (middle row), and DNN-3 (bottom row)) models (a - overall, b - different initial water contents, c - different textures).

3.5.1.5 Comparison of Data-driven models

Figure 8 presents a box plot of the relative error distribution for various data-driven models, as well as other statistical parameters for their agreement with the PBM. The LR and SVM-L models show the big error distributions, which indicates that these models were not well suited to predict runoff. The SVM-NL model performed much better than the SVM-L and LR models, but not as well as kNN-5 and DNN models. It is possible that the performance of the SVM-NL model could be further improved if the training data size was increased, but this would also dramatically increase the computational time. The boxplot of relative errors for the kNN-5 model was narrow, but it included a lot of outliers which indicated that the model performance may be different under other conditions. According to Figures 3.7a and 3.8, outliers of the relative error for DNN-1 appeared when Q_c was over 100 L, whereas the DNN-3 model slightly underestimated large Q_c values. The DNN-2 model shows the narrowest box plot of relative errors and the best performance statistics for the considered ML models. This indicates that the DNN-2 model provided the best prediction of runoff water volumes.



	Linear	SVM-L	SVM-NL	kNN-5	DNN-1	DNN-2	DNN-3	DNN-2_S1
RMSE	1.46	1.84	0.47	0.59	0.49	0.30	0.91	4.77
MAE	0.83	0.75	0.25	0.23	0.27	0.14	0.42	1.57
MBE	0.18	-0.57	0.21	0.13	0.19	0.10	0.35	0.18
EF	0.39	0.03	0.94	0.90	0.93	0.97	0.76	0.96

Figure 3.8. Statistical parameters associated with the trained data-driven models of runoff quantity and quality. The root mean square error (RMSE), mean bias error (MBE), mean absolute error (MAE), model efficiency (EF), and a box plot of the relative error distribution for data-driven models are given. The units of RMSE, MBE, and MAE of the runoff quantity and quality are *L* and *g*, respectively. The upper and lower boundaries of the boxes show the 75th and 25th percentile, the whiskers of the box plot show the maximum and minimum values, and the red line within the box is the median value. Blue dot symbols indicated the outliers.

3.5.2 Surface Runoff Quality

We discovered that the DNN-2 model structure with fine adjusted model parameters performs best among all selected data-driven models for predicting the

runoff water volume. We therefore trained the DNN-2 model with the near equilibrium solute transport database. The DNN-2 model for water flow could be directly adjusted to produce a data-driven model that relates physical factors and the cumulative solute mass.

Figure 9a presents a scatter plot of the cumulative solute mass predicted by the DNN-2 model and observed by the PBM under near equilibrium conditions. The constructed DNN-2 relationship between input and output variables for near equilibrium conditions was very accurate. The goodness of the DNN-2 predictions for solute transport near equilibrium conditions is reflected by the statistical parameters (RMSE= 4.77 g, MAE=1.57 g, MBE=0.18 g, and EF=0.96 g) in Figure 8, the value R^2 =0.992 in Figure 3.9a, and the agreement between observed (PBM) and predicted results for the example simulations for various water contents (Fig. 3.9b) and soil types (Fig. 3.9c).

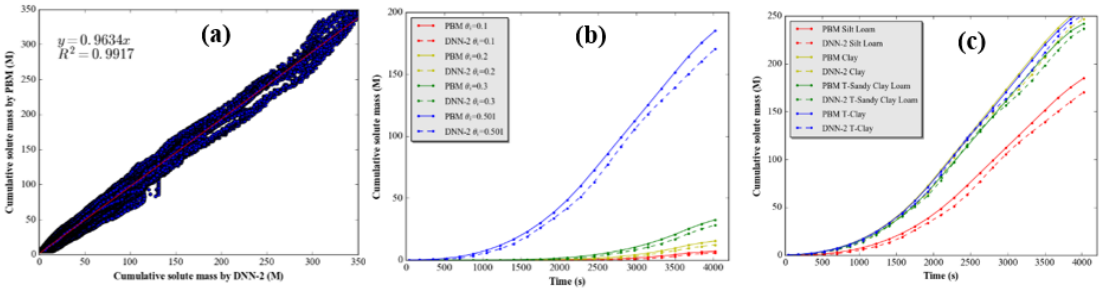


Figure 3.9. Comparison of the cumulative solute mass calculated by the PBM and the DNN-2 model for the near equilibrium training dataset (a - overall, b - different initial water contents, c - different textures).

3.6 Conclusion and Outlook

This study tested the ability of data-driven models to mimic PBMs to predict surface runoff water quantity and quality in agricultural settings. A physically-based overland flow and transport model was used to develop a large database containing information about the impact of various factors on surface runoff quantity and quality. In order to build a realistic training database, a wide range of input parameter values was selected. The input factors included: a) rainfall intensities (for seven different percent probabilities of exceedance of 1%, 2%, 4%, 10%, 20%, 50%, and 100% for the San Jacinto Watershed, Southern California), b) Manning's roughness coefficients (reflecting management practices; 6 values), c) field slopes (5 values), d) field lengths (5 values), e) soil properties (affecting infiltration; 11 soil types with and without tillage), e) initial water contents (4 values), and f) solute sorption parameters (5 conditions). The resulting water flow database thus contained 2,310,000 entries, whereas the solute transport database was 5 times larger.

Multiple data-driven models were tested on their ability to relate/correlate model

inputs with outputs. The following data-driven models were used in the analysis: Linear Regression (LR), k-Nearest Neighbor regression (kNN), Support Vector Machine with linear (SVM-L) and non-linear (SVM-NL) kernels, and Deep Neural Networks (DNN) (Neural Networks with multiple hidden layers). The LR and SWM-L models failed to accurately describe runoff water dynamics, having the regression coefficient (R^2) less than 0.5. In particular, both linear models tended to overestimate the runoff water volume for early times and underestimate it for later times. Both SVM-NL and kNN models performed much better than linear models, having R^2 of 0.93 and 0.96, respectively. However, there were still considerable deviations between the SVM-NL and kNN models and PBM predictions, especially for larger times, higher initial water contents, and for some textures (such as Silt Loam). Finally, the DNN models with one, two, and three hidden layers (DNN-1, DNN-2, and DNN-3) were tested to determine the optimum number of hidden layers and to minimize overfitting of output variables. The best performance was obtained by the DNN model with two hidden layers (DNN-2) (R^2=0.98). The DNN-1 and DNN-3 models tended to underestimate, and overestimate observed values of runoff water volumes, respectively.

In conclusion, DNN techniques exhibited a better capability to reproduce the results of the PBM compare with traditional ML techniques. In particular, results indicate that the simple linear models are not well suited to develop correlations for runoff from agricultural lands because runoff is a complex non-linear hydrological process. Although non-linear models like kNN and SVM-NL can fairly accurately capture the surface water flow dynamic, the DNN models are more promising for dealing with complex hydrologic

problems. Applications of the DNN models inherit benefits of traditional ANNs, but additional hidden layers provide a possibility to further improve the model performance. Similarly, the trained DNN-2 model provided an excellent prediction of near equilibrium solute transport.

It should be noted that the training databases from the PBM could also be augmented to include real monitoring data and/or more complex model formulations (e.g., multidimensional simulations, stochastic simulations, other reactive transport processes, and physical non-equilibrium flow and transport). This approach would preserve the benefits of data-driven models over PBMs (e.g., faster execution times and ability to incorporate real data), while constraining data-driven models to the underlying physics of hydrological processes.

3.7 References

Abrahart, R.J. and See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological processes*, 14(11-12), 2157-2172.

Ahuja, L.R., Sharpley, A.N. and Lehman, O.R., 1982. Effect of Soil Slope and Rainfall Characteristics on Phosphorus in Runoff 1. *Journal of Environmental Quality*, 11(1), 9-13.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, 5(2), 115-123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124-137.

Bai, Y., Chen, Z., Xie, J. and Li, C., 2016. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of hydrology*, 532, 193-206.

Borah, D.K. and Bera, M., 2003. Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *Transactions of the ASAE*, 46(6), 1553.

Bray, M. and Han, D., 2004. Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics*, 6(4), 265-280.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N. and Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological applications*, 8(3), 559-568.

Chollet, F., 2015. Keras: Deep learning library for theano and tensorflow. *URL: https://keras. io/k*, 7(8).

Donigian, A.S. and Love, J.T., 2002. The Connecticut Watershed Model–A tool for BMP impact assessment. *Proceedings of the Water Environment Federation*, *2002*(2), pp.605-624.

Dawson, C.W. and Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal*, 43(1), 47-66.

Dibike, Y.B., Solomatine, D. and Abbott, M.B., 1999. On the encapsulation of numerical-hydraulic models in artificial neural network. *Journal of Hydraulic research*, 37(2), 147-161.

Dibike, Y.B., Velickov, S., Solomatine, D. and Abbott, M.B., 2001. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15(3), 208-216.

Fang, K., Shen, C., Kifer, D. and Yang, X., 2017. Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. *Geophysical Research Letters*, 44(21), 11-030.

Friedman, J., Hastie, T. and Tibshirani, R., 2001. The elements of statistical learning: Springer series in statistics. *Springer.*

Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y., 2016. Deep learning (Vol. 1). *Cambridge: MIT press.*.

Goodrich, D.C., Burns, I.S., Unkrich, C.L., Semmens, D.J., Guertin, D.P., Hernandez, M., Yatheendradas, S., Kennedy, J.R. and Levick, L.R., 2012. KINEROS2/AGWA: model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1561-1574.

Govindaraju, R.S. and Rao, A.R. eds., 2013. Artificial neural networks in hydrology (Vol. 36). *Springer Science & Business Media.*

Green, W.H. and Ampt, G.A., 1911. Studies on Soil Phyics. *The Journal of Agricultural Science*, 4(1), 1-24.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6), 82-97.

Hromadka II, T.V. and Lai, C., 1985. SOLVING THE TWO-DIMENSIONAL DIFFUSION FLOW MODEL. In *Hydraulics and Hydrology in the Small Computer Age, Proceedings of the Specialty Conference*, 555-562.

Hsu, K.L., Gupta, H.V. and Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10), 2517-2530.

Joachims, T., 2002. Learning to classify text using support vector machines: Methods, theory and algorithms (Vol. 186). *Norwell: Kluwer Academic Publishers*.

Karandish, F. and Šimůnek, J., 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of hydrology*, *543*, 892-909.

Karlsson, M. and Yakowitz, S., 1987. Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. *Water Resources Research*, 23(7),1300-1308.

Liang, J., Bradford, S.A., Šimůnek, J. and Hartmann, A., 2017. Adapting HYDRUS-1D to Simulate Overland Flow and Reactive Transport during Sheet Flow Deviations. *Vadose Zone Journal*, 16(6).

Lin, J.Y., Cheng, C.T. and Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), 599-612.

Loague, K.M. and Freeze, R.A., 1985. A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research*, 21(2), 229-248.

Marçais, J. and de Dreuzy, J.R., 2017. Prospective interest of deep learning for hydrological inference. *Groundwater*, 55(5), 688-692.

Maier, H.R. and Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental modelling & software*, 15(1), 101-124.

Minns, A.W. and Hall, M.J., 1996. Artificial neural networks as rainfall-runoff models. *Hydrological sciences journal*, 41(3), 399-417.

Mjolsness, E. and DeCoste, D., 2001. Machine learning for science: state of the art and future prospects. *science*, 293(5537), 2051-2055.

Nourani, V., Komasi, M. and Mano, A., 2009. A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water resources management*, 23(14), 2877.

Panday, S. and Huyakorn, P.S., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Advances in water Resources*, 27(4), 361-382.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*, 2825-2830.

Rawls, W.J., Brakensiek, D.L. and Miller, N., 1983. Green-Ampt infiltration parameters from soils data. *Journal of hydraulic engineering*, 109(1), 62-70.

Remesan, R. and Mathew, J., 2016. Hydrological data driven modelling. *Springer*.

Roz, E.P., 2011. Water quality modeling and rainfall estimation: A data driven approach. *University of Iowa, IA,US.*

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.

Sharda, V.N. and Singh, S.R., 1994. A finite element model for simulating runoff and soil erosion from mechanically treated agricultural lands: 1. Governing equations and solutions. *Water resources research*, 30(7), 2287-2298.

Sharma, V., Negi, S.C., Rudra, R.P. and Yang, S., 2003. Neural networks for predicting nitrate-nitrogen in drainage water. *Agricultural Water Management*, 63(3),169-183.

Šimůnek, J., Šejna, M., Saito, H., Sakai, M. and Van Genuchten, M.T., 2008. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, version 4.0: HYDRUS Software Series 3. *Department of Environmental Sciences, University of California Riverside, Riverside, California, USA, 315*.

Šimůnek, J., Van Genuchten, M.T. and Šejna, M., 2016. Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone Journal*, 15(7).

Solomatine, D.P. and Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. *Journal of hydroinformatics*, 10(1), 3-22.

Ticlavilca, A.M. and McKee, M., 2011. Multivariate Bayesian regression approach to forecast releases from a system of multiple reservoirs. *Water resources management*, 25(2), 523-543.

Tyrrel, S.F. and Quinton, J.N., 2003. Overland flow transport of pathogens from agricultural land receiving faecal wastes. *Journal of Applied Microbiology*, 94, 87-93.

USGS, 1999. The quality of our nation's waters-nutrients and pesticides. *US Geological Survey Circular 1225*, Washington DC, 80.

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J. and Schaap, M.G., 2017. Pedotransfer functions in Earth system science: Challenges and perspectives. *Reviews of Geophysics*, 55(4), 1199-1256.

Vapnik, V., 2013. The nature of statistical learning theory. *Springer science & business media.*

Vapnik, V., 1998. Statistical learning theory.Vol. 3, *Wiley*,

Wallach, R., Grigorin, G. and Byk, J.R., 2001. A comprehensive mathematical model for transport of soil-dissolved chemicals by overland flow. *Journal of Hydrology*, 247(1-2), 85-99.

Weill, S., Mouche, E. and Patin, J., 2009. A generalized Richards equation for surface/subsurface flow modelling. *Journal of Hydrology*, 366(1-4), 9-20.

Yu, P.S., Chen, S.T. and Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328(3-4), 704-716

Chapter 4

Extending the HYDRUS-1D Overland Flow Model to Simulate Soil Erosion During Sheet Flow Deviations

ABSTRACT

The popular HYDRUS-1D code was modified to simulate erosion, sediment transport, and deposition over the land surface. The resulting numerical model has an excellent agreement with the KINEROS model, which confirms that the numerical solution was correctly implemented. Interrill erosion has been assumed to be an equilibrium transport process in many publicly available models. Conversely, sediment transport in interrill flow exhibits characteristics of non-equilibrium transport because of irregularity of hillslope surfaces, soil heterogeneity, and instability of flow. In this study, we propose a series of process-based models to simulate sediment transport during uniform and non-uniform flow. Numerical experiments were conducted to demonstrate the model's applicability and to provide insight erosion processes during sheet-flow derivations. The newly developed erosion models provide a comprehensive set of tools to numerically investigate many important research problems involving sediment erosion, transport, and deposition processes.

4.1. Introduction

Soil erosion is recognized as one of the major environmental threats to sustainable agriculture because the erosion process removes organic matter and nutrients from agricultural fields, which results in a reduction of cultivable soil and a decline in soil fertility. In addition, soil erosion produces sediment downstream, which reduces the capacity of rivers or streams and increases pollution levels of nitrogen and phosphorus, which are often adsorbed to sediments [Morgan, 2009]. Agricultural activities increase the vulnerability of soil to erosion because the removal of the vegetation cover and repeated preparations by tillage leave the soil exposed to energy from raindrops or wind. An estimated 10 million ha of cropland are annually abandoned worldwide due to the lack of productivity caused by erosion [Feath, 1994]. Therefore, knowledge of what happens during individual rainstorms on agricultural land is required to effectively manage soil erosion.

Soil erosion is closely related to rainfall and runoff and is comprised of three major components: erosion due to raindrop splash, interrill erosion by sheet flow, and rill erosion by concentrated flow in rills [Liu et al., 2006]. On well-maintained agricultural land, sheet erosion has been reported to be the predominant source of both soil loss and fluvial sediments [American Society of Civil Engineers, 1970]. Sheet erosion caused by sheet flow removes soil uniformly in thin layers by forces from raindrops and overland flow whereas rill erosion caused by runoff water forms small channels as it concentrates down a slope. Quantitative assessment of sheet erosion is more difficult than rill erosion because sheet erosion is less visible than rill erosion, which is more easily identified on

agricultural fields. In addition, sheet erosion processes are rarely uniform due to spatial variabilities of local soil topography, vegetation, and soil heterogeneity.

Point measurements of erosion are labor-intensive and expensive, and often not reliable because they are affected by a wide spatial and temporal variability of influencing factors. Consequently, mathematical models are commonly used to predict spatial and temporal variations in erosion and to improve erosion management. Many different models and relations have been proposed during the last few decades to describe and predict soil erosion and associated sediment yield [Williams and Berndt, 1977; Wischmeier and Smith, 1978; Knisel, 1980; Singh et al., 1982; Abbott et al., 1986; Storm et al., 1987; Renard et al., 1994; Hanley et al., 1998; Nearing et al., 1999; Woodward, 1999; De Jong et al., 1999; Johnson et al., 2000; Van Oost et al., 2000; Tucker et al., 2001; Ziegler et al., 2001; Torri et al., 2002; Goodrich et al, 2002; Katlin et al., Borah et al., 2017;]. Existing soil erosion models can be divided into two categories: empirical models and physically-based models. Empirical models, such as the universal soil loss equation [USLE model; Wischmeier and Smith, 1978] and the revised USLE [RUSLE; Renard et al., 1991] have been extensively used to predict the total mass of soil erosion at specific location. However, empirical models are not able to describe sediment deposition, pathways taken by eroded material, or the spatial variability of soil erosion. On the other hand, physically-based erosion models, which solve the sediment mass conservation equation, provide much more detailed information than the USLE or its derivatives. Physically-based erosion models such as the Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS) model [Knisel, 1980], the Areal

Nonpoint Source Watershed Environment Response Simulation (ANSWERS) model [Beasley et al., 1980], the Kinematic Runoff and Erosion (KINEROS) model [Woolhiser et al., 1990], the European Soil Erosion Model [EUROSEM; Morgan et al., 1998] and the Water Erosion Prediction Project (WEPP) model [Nearing et al., 1989], are widely used models nowadays.

The WEPP, CREAMS, and EUROSEM models calculate rill and interrill processes separately whereas the KINEROS and ANSWERS models do not explicitly separate rill and interrill erosion. However, all these models considered overland flow and interrill erosion to be uniform processes. In reality, overland flow is rarely uniformly distributed over the land surface and so is soil erosion. An alternative modeling approach is needed to consider non-equilibrium, non-uniform soil erosion and sediment transport under conditions of deviations from sheet flow. Therefore, the overall objective of this study is to develop a series of physically-based soil erosion models capable of simulating not only interrill erosion during sheet flow but also interrill erosion during deviations from sheet flow. This chapter describes the development and applications of such models, which are different from existing physically-based models discussed above, and which are analogous to the non-equilibrium overland flow and solute transport models of Liang et al. [2017].

4.2 Numerical Models

4.2.1 Sheet Erosion Model

The HYDRUS-1D overland flow module [Liang et al., 2017] was extended to simulate soil erosion in overland flow. The surface water depth and runoff flow rates are obtained by solving the diffusive wave equation. The detailed development and numerical implementation of the overland flow module in the HYDRUS-1D software are described in Chapter 2. Sediment transport during uniform overland flow is described using the following equation:

$$\frac{\partial hc_e}{\partial t} = \frac{\partial}{\partial x}\left(hD\frac{\partial c_e}{\partial x} \right) - \frac{\partial Qc_e}{\partial x} - \mu hc_e - \gamma h \qquad [1]$$

where x is a space coordinate in the direction of flow [L], t is time [T], h is the surface water depth [L], c_e is the sediment concentration [ML^{-3}], D is the effective dispersion coefficient accounting for both molecular diffusion and hydrodynamic dispersion [L^2T^{-1}], Q is the runoff flow rate [L^2T^{-1}], μ [T^{-1}] is the first-order decay coefficient, and γ [$ML^{-3}T^{-1}$] is a zero-order decay coefficient. The sediment carrying capacity approach [Miller et al., 2007; Goodrich et al., 2012; Kennedy et al., 2013] was implemented into HYDRUS-1D to model erosion due to the raindrop impact and surface runoff, and sediment deposition. In this case, values of γ and μ in Eq. [1] were defined to be consistent with the sediment carrying capacity approach that is implemented in Kineros2 [Goodrich et al., 2012]. In particular, the value of γ is given as:

$$\gamma = -\frac{\rho_s}{hW} c_f f_{cohesion} \exp\left(-c_h h\right) R^2 - \mu C_m \qquad [2]$$

where $f_{cohesion}$ [-] is the particle cohesion fraction, W [L] is the width of the field, c_f [T] is the splash erosion coefficient, c_h [L^{-1}] is the dampening coefficient, ρ_s [ML^{-3}] is the specific density of the particle, and C_m [ML^{-3}] is the maximum particle carrying capacity. The first-term on the right-hand side of Eq. [2] accounts for erosion due to the raindrop impact, whereas the second term is used to describe hydraulic erosion. Experimental work by Govers [1990] and others using shallow flows over soil have demonstrated that C_m was well described using the Engelund and Hansen [1967] relation:

$$C_m = \frac{0.5 U U_s^3 \rho_s}{g^2 d_p h \left(\gamma_s - 1\right)^2} \quad [3]$$

where g [LT^{-2}] is the acceleration due to gravity, U_s [LT^{-1}] is the shear velocity (e.g., $U_s = \sqrt{ghS}$), d_p [L] is the particle diameter, and γ_s [-] is the suspended specific gravity of the particles.

The sediment carrying capacity approach considers that erosion from surface runoff occurs when $C_m \geq c_e$ and particle deposition occurs when $C_m < c_e$. The value of μ is theoretically equal to the particle settling velocity divided by the hydraulic depth, h. To account for this effect, the value of μ in Eq. [1] is equivalent to the runoff erosion rate coefficient, which takes on different values during runoff erosion and deposition as follows:

$$\mu = f_{cohesion} \frac{v_s}{h} \quad (C_m \geq c_e) \quad [4a]$$

$$\mu = \frac{v_s}{h} \quad (C_m < c_e) \qquad [4b]$$

where v_s [L T^{-1}] is the particle settling velocity. The value of v_s is calculated from the particle size and density, assuming that particles have drag characteristics and terminal fall velocities similar to those of spheres [Fair and Geyer, 1954]. This relation is as follows:

$$v_s^2 = \frac{4}{3}\frac{g(\rho_s - 1)d_p}{C_D} \qquad [5]$$

in which C_D [-] is the particle drag coefficient. The drag coefficient is a function of the particle Reynolds number R_n:

$$C_D = \frac{24}{R_n} + \frac{3}{\sqrt{R_n}} + 0.34 \qquad [6]$$

which is defined as:

$$R_n = \frac{v_s d_p}{\nu} \qquad [7]$$

where ν is the kinematic viscosity of water [L^{2}T^{-1}]. The settling velocity of a particle is found by solving Eqs. [5]-[7] for v_s.

4.2.2 Horizontal Mobile-Immobile (HMIM) Erosion Model

Similar to the HMIM solute transport model [Liang et al., 2017], the HMIM erosion model assumes that the soil surface is horizontally divided, parallel to the direction of water flow, into regions with mobile and immobile water. Sediments in the immobile region may be stored, retained, and exchanged with the mobile domain.

$$\frac{\partial h_m c_{e,m}}{\partial t} = \frac{\partial}{\partial x}\left(h_m D_m \frac{\partial c_{e,m}}{\partial x} \right) - \frac{\partial Q_m c_{e,m}}{\partial x} - \frac{E_e^{MIM}}{w_m} - h_m \mu c_{e,m} - \gamma h_m \qquad [8a]$$

$$\frac{\partial h_{im} c_{e,im}}{\partial t} = \frac{E_e^{MIM}}{(1-w_m)} - h_{im} \mu c_{e,im} - \gamma h_{im} \qquad [8b]$$

where w_m [-] is the ratio of the width of the mobile region to the total width of the soil surface, and the subscripts *m* and *im* denote parameters associated with the mobile and immobile regions, respectively. The parameter E_e^{MIM} [$ML^{-2}T^{-1}$] represents the sediment transfer rate between mobile and immobile regions and is given as:

$$E_e^{MIM} = \omega_{e,m}(1-w_m) h_{im}(c_{e,m} - c_{e,im}) + \alpha_{e,m}(h_m - h_{im}) c_e^* \qquad [9]$$

where $\omega_{e,m}$ [T^{-1}] is the sediment transfer coefficient, $\alpha_{e,m}$ [T^{-1}] is the first-order mass transfer coefficient, and c_e^* is a concentration that is equal to $c_{e,m}$ for $h_m > h_{im}$ and $c_{e,m}$ for $h_m < h_{im}$. It should be mentioned that Eqs. [8a] and [8b] are written in terms of the local-scale mass balance in the mobile and immobile regions, respectively. To formulate them in terms of the total region, the mass balance equations for mobile and immobile regions need to be multiplied by w_m and (1-w_m), respectively.

2.3 Active-Passive Regions (APR) Erosion Model

In analogy of the APR solute transport model [Liang et al., 2017], the APR erosion model assumes that the soil surface is divided, parallel to the direction of water flow, into hydraulically active (fast flow) and passive (slow flow) domains. This model allows sediment transport in both active and passive regions and exchange between these two regions. Soil erosion in each region is described using separate diffusion wave equations:

$$\frac{\partial h_1 c_{e,1}}{\partial t} = \frac{\partial}{\partial x}\left(h_1 D_1 \frac{\partial c_{e,1}}{\partial x} \right) - \frac{\partial Q_1 c_{e,1}}{\partial x} - \frac{E_e^{APR}}{w_A} - \mu_1 h_1 c_{e,1} - \gamma_1 h_1 \qquad [10a]$$

$$\frac{\partial h_2 c_{e,2}}{\partial t} = \frac{\partial}{\partial x}\left(h_2 D_2 \frac{\partial c_{e,2}}{\partial x} \right) - \frac{\partial Q_2 c_{e,2}}{\partial x} + \frac{E_e^{APR}}{1-w_A} - \mu_2 h_2 c_{e,2} - \gamma_2 h_2 \qquad [10b]$$

where E_e^{APR} [$ML^{-2}L^{-1}$] is the sediment transfer rate between active and passive regions, w_A [-] is the ratio of the width of the surface active region and the total surface width, and the subscripts 1 and 2 refer to hydrologically active and passive regions, respectively. Value of E_e^{APR} is quantified in a similar manner to Eq. [9]:

$$E_e^{APR} = \omega_{e,12}(1-w_A)h_2(c_{e,1}-c_{e,2}) + \alpha_{e,12}(h_1-h_2)c_e^* \qquad [11]$$

where $\omega_{e,12}$ [T^{-1}] is the sediment transfer coefficient for transfer between active and passive regions, $\alpha_{e,12}$ [T^{-1}] is the first-order mass transfer coefficient, and c_e^* is a concentration that is equal to $c_{e,1}$ for $h_1 > h_2$ and $c_{e,2}$ for $h_1 < h_2$. The average sediment flux is:

$$c_T^e = \frac{w_A c_{e,1} Q_1 + (1 - w_A) c_{e,2} Q_2}{w_A Q_1 + (1 - w_A) Q_2} \quad [12]$$

It should be mentioned that Eqs. [10a] and [10b] are written in terms of local-scale mass balances in Regions 1 and 2. To formulate them in terms of the entire region, the mass balance equations for Regions1 and 2 need to be multiplied by w_A and (1- w_A), respectively.

4.3 Numerical Examples

Several numerical examples were conducted using the newly developed soil erosion models in order to show the models' applicability. All simulation domains are considered to be 100 m long and 1 m wide, consisting of single size particle grains. The default erosion parameters used in all simulations (unless mentioned otherwise) are given in Table 4.1.

Table 4.1. Default erosion parameters used in simulations.

Input parameters	Values
Particle density, ρ_s (kg/m^3)	2.65×10^3
Splash erosion coefficient, c_f (s)	50
Dampening coefficient, c_h (m^{-1})	656
Temperature, T (°C)	33
Particle cohesion fraction, $f_{cohesion}$ [-]	1

4.3.1 Sheet Erosion Model

Figure 4.1 shows the comparison of simulation results obtained by the KINEROS and HYDRUS sheet erosion models for three different particle sizes with diameters of 0.1, 0.05, and 0.01 mm, respectively. Two different rainfall rates (0.2 and 0.9 mm/s) were

applied to the land surface (brown dashed line) with the Manning's coefficient of 0.04 and the slope of 5%. The simulated results show that the sediment transport rate increased as the rainfall intensity increased because higher rainfall rates produce higher discharge, which in turn leads to the increased eroding capability. Smaller soil particles have higher erodibility and smaller settling velocities than larger soil particles which cause them to travel longer distances along the slope. There is generally an excellent agreement between the KINEROS and HYDRUS sheet erosion models. However, the HYDRUS erosion model predicted lower peak sediment concentration for the 0.01 mm soil particles compared with the KINEROS model. This deviation is likely due to differences in the governing equations and numerical solution of these models; e.g., the HYDRUS overland flow model solves the diffusion wave equation using the finite element method and considers the dispersive sediment flux, whereas the KINEROS model solves the kinematic wave equation using the finite difference approach and neglects the dispersive flux. As a result, flow shock fronts, as well as the corresponding sediment transport rates, may be slightly different.

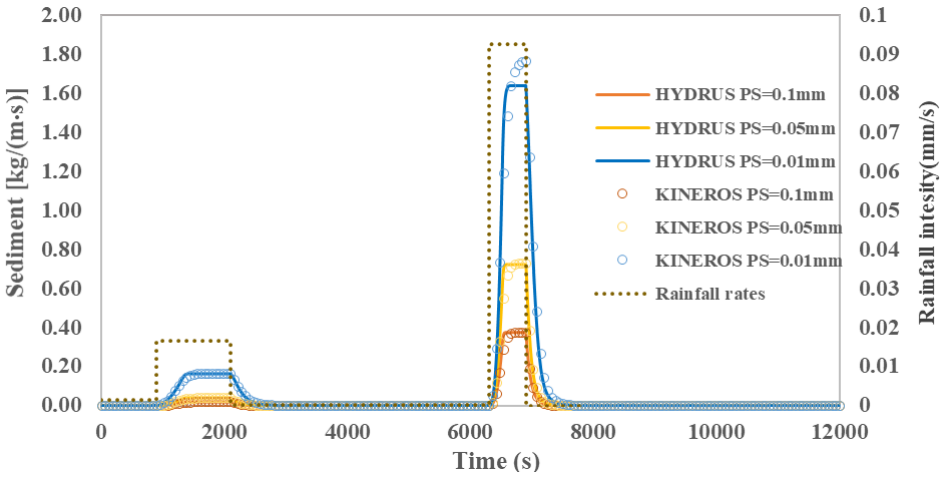


Figure 4.1. Comparison of simulation results obtained by the KINEROS and HYDRUS erosion models. Different colors represent the sediments transport rates (kg/m/s) at the outlet for different soil particle sizes (PS) for the soil profile with the Manning's coefficient of 0.04 and the slope of 5%.

Another numerical example was conducted to show the effect of the slope gradient on soil erosion since the slope gradient is one of the most important factors affecting surface flow erosion. A uniform rainfall of 9.25×10^{-3} mm/s was applied for 1200 s on a land surface with a Manning's roughness coefficient of 0.01 and soil particle size of 0.05 mm. The slope in the top half of the profile was 1% and in the bottom half was 5%. The HYDRUS erosion model can provide temporal and spatial distributions of sediment concentrations. Figure 4.2 shows the sediment concentration fluxes in kg/m/s at different times along the land surface. Sediment concentration fluxes are always higher in the bottom half of the profile with the higher slope. The higher slope produces lower water depths, higher flow velocities, and lower settling velocities of soil particles, and as a result higher sediment concentrations.

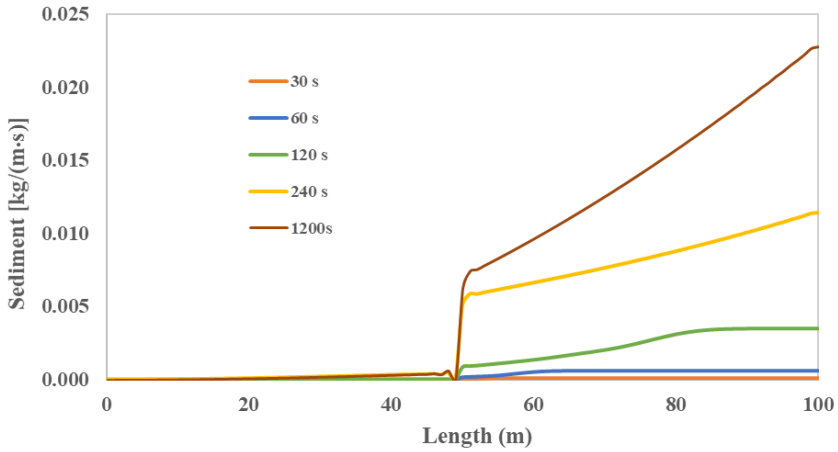


Figure 4.2. Sediment concentration fluxes at different times for an example with uniform rainfall (9.25×10^{-3} mm/s) on the soil surface with two different slopes (1 and 5%), a particle size of 0.05 mm, and a Manning's roughness coefficients of 0.01.

To investigate differences between mechanisms of sediment transport and reactive solute transport, Fig. 4.3 shows the simulation results for the same flow conditions induced by inflow at the upslope inlet and involving either solute wash-off or sediment transport. Surface runoff is induced by a 1 cm constant head at the top of the slope for 360 s. The Manning's coefficient of the land surface was 0.01, the slope was 1%, and the total simulation time was 1200 s. The most sensitive parameters affecting solute and sediment concentrations are sorption/desorption rates and transport capacities of soil particles, respectively. Solute wash-off simulations were conducted using the HYDRUS overland flow one-site kinetic sorption model (eq. [3] in Chapter 2). A uniform unit initial soil solute concentration, the distribution coefficient K_D of 1 m, and different kinetic sorption rates ω (s^{-1}) were considered in the solute wash-off scenarios (Fig. 4.3a). A higher sorption/desorption rate produces a higher initial concentration pulse

because solute detaches faster from the land surface and is transported in surface runoff water rapidly during the inflow period. A lower kinetic sorption rate results in lower initial outflow solute concentrations during inflow. However, once inflow ceases, solute concentrations increase due to continued solute desorption from the land surface into a decreasing surface water layer.

The soil erosion simulations (Fig. 4.3b) were conducted using different values of particle sizes (PS=0.01, 0.025, 0.05, and 1 mm). In these simulations, the transport of particles by runoff water is the main soil erosion factor since there is no impact of raindrops in the inflow simulations. We can observe increased sediment concentrations at the outlet during the inflow period and a gradual decrease in sediment concentrations (due to sedimentation) after inflow ceased. The sediment concentrations are highest for soil particles of 0.01 mm than for larger particle sizes (0.1, 0.05, and 0.025 mm) during the entire simulation time. The sediment settling velocity of the soil particle of 0.01 mm is smaller than for larger particles, producing higher concentrations of smaller eroded soil particles and longer travel distances than for larger soil particles due to the higher particle carrying capacity. Fig. 4.3 illustrates that solute transport and sediment erosion produce quite different concentration profiles under the same flow conditions. The tailing and peak solute concentrations are mainly controlled by sorption/desorption rates whereas the peak sediment concentrations are mainly controlled by the transport carrying capacity and settling velocities when flow conditions are the same.

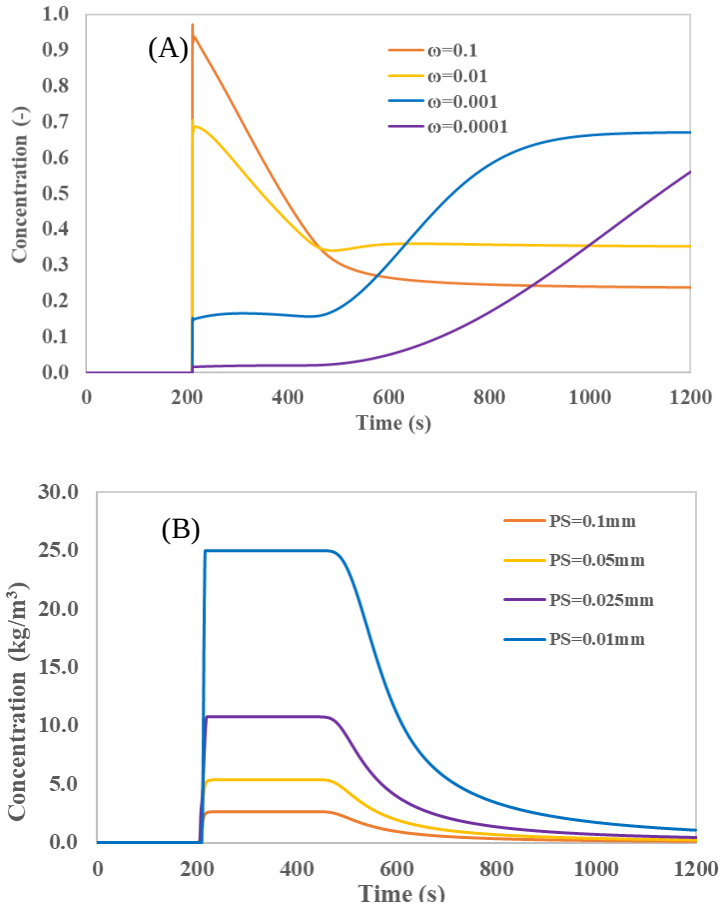


Figure 4.3. Simulated (a) solute concentrations and (b) sediment concentrations at the bottom boundary under the same runoff velocities. Different sorption/desorption rates (ω=0.0001, 0.001, 0.01, and 0.1 s^{-1}) are considered in solute transport simulations and different particle sizes (PS=0.01, 0.025, 0.05, and 1 mm) are considered in sediment transport simulations. The surface runoff is induced by a 1 cm constant head at the top of the slope for 360 s. The Manning's coefficient of the land surface is 0.01 and the slope is 1%.

4.3.2 Horizontal Mobile-Immobile Erosion Model

The example presented in Fig. 4.1 (slope=5%; the Manning coefficient=0.04) for the particle size of 0.1 mm was rerun using the HMIM model. Three additional input parameters were needed compared with the sheet erosion model: (i) the ratio of the width

of the mobile domain to the total width of the surface domain, w_m, (ii) the water mass transfer coefficient, $\alpha_{e,m}$, and (iii) the sediment transfer coefficient $\omega_{e,m}$. The influence of the sediment transfer coefficient on the sediment transport rate at the bottom boundary as a function of time is demonstrated in Fig. 4.4. Constant values of w_m =0.5 and $\alpha_{e,m}$ =0 and different values of $\omega_{e,m}$ = 0, 0.0001, 0.001, and 0.01 s^{-1} are considered in Fig. 4.4. When no exchange of sediment occurs between the mobile and immobile domains ($\omega_{e,m}$ =0), the sediment transport rate is half of that for the sheet erosion model since only half of the total region (the mobile region) contributes to soil erosion. In general, the surface water depth in the immobile region is higher than in the mobile region under the same rainfall intensity because rainfall water accumulates in this region without runoff outflow. Similarly, sediment concentrations (due to raindrop erosion and no outflow) are higher in the immobile region than in the mobile region. Therefore, a fraction of the sediment in the immobile region is transferred to the mobile region when $\omega_{e,m}$ >0, resulting in higher sediment transport rates compared to when $\omega_{e,m}$ =0. An increase in $\omega_{e,m}$ produces a faster equilibration of sediment concentrations between the mobile and immobile zones and produces higher peaks in the sediment transport rate whereas a decrease in $\omega_{e,m}$ produces prolonged tailing.

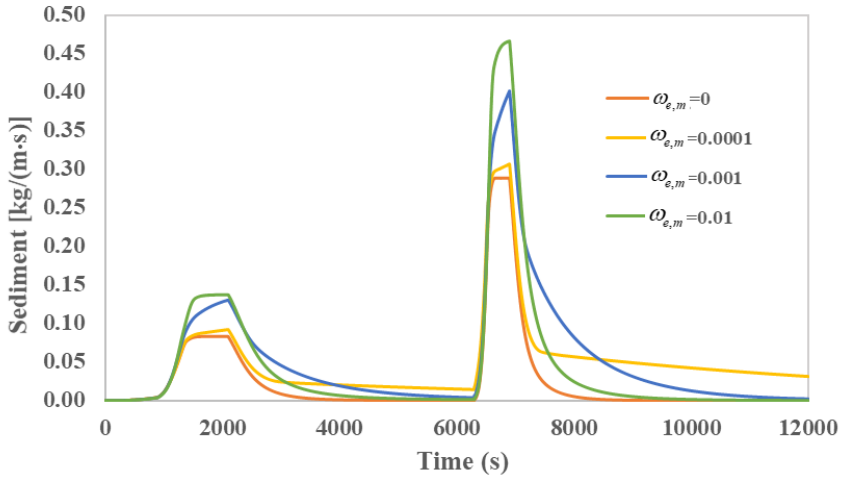


Figure 4.4. Sediment transport rates at the outflow boundary simulated using the horizontal mobile–immobile (HMIM) erosion model with different values of solute exchange rates ($\omega_{e,m}$=0, 0.0001, 0.001, and 0.01 s^{-1}) when rainfall (0.2 and 0.9 mm/s) is applied on the entire soil profile (the Manning coefficient=0.04, slope=5%, w_m =0.5).

Soil erosion due to inflow at the upslope boundary instead of rainfall is simulated in Fig. 4.5. An inlet water depth is equal to 1 cm for 6 min and soil surface is considered to be impervious. The Manning's coefficient is equal to 0.01 and the slope is 1%. A constant value of w_m =0.5, $\omega_{e,m}$ =0, and different values of $\alpha_{e,m}$ =0, 0.001, 0.005, and 0.01 s^{-1} are considered in Fig. 4.5. No exchange of water occurs between the mobile and immobile domains when $\alpha_{e,m}$ = 0 and the resulting sediment transport rates are equal to those predicted by the sheet erosion transport model. Conversely, a fraction of inflow water is transferred to and from mobile and immobile regions when $\alpha_{e,m}$ >0. This process slows down the sediment transport and delays the arrival of sediment to the outlet compared to the sheet erosion model. It also produces prolonged tailing in the sediment

transport rate compared with the sheet erosion model. Increasing $\alpha_{e,m}$ produces a faster runoff water movement from mobile zone to immobile zone, water depth gradient between the moible and immobile zones causes a greater delay in the arrival of sediment to the outlet.

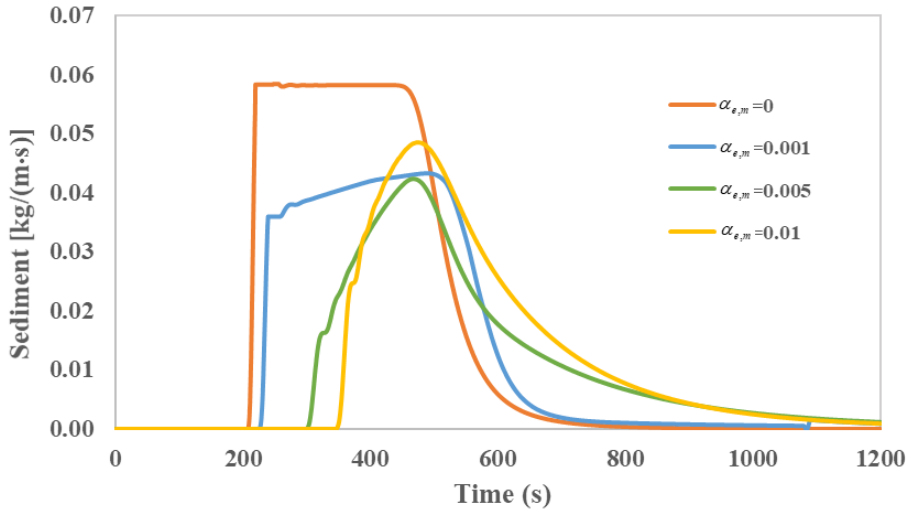


Figure 4.5. Sediment transport rates at the outflow boundary simulated using the horizontal mobile–immobile (HMIM) erosion model with different values of the first-order mass transfer rate ($\alpha_{e,m}$=0, 0.0001, 0.001, and 0.01 s^{-1}) when inflow (h_0= 1 cm) is initiated from the upslope boundary. The Manning coefficient=0.01, slope=1%, and w_m =0.5.

4.3.3 Active-Passive Regions Erosion Model

The sediment transport occurs in two parallel surface regions in the APR model. Water and sediment exchange between the active and passive regions is determined by the water mass transfer coefficient α_{12}^{e} and sediment transfer rate ω_{12}^{e}, respectively. Three different simulation scenarios were considered in this section using the APR

erosion model. Separate Manning's roughness coefficients are used in the surface active (n_1) and passive (n_2) regions to obtain different velocities. Figure 4.6 presents simulated active, passive, and total sediment transport rates as a function of time when using the APR model with $w_A = 0.5$, $n_1 = 0.01$, $n_2 = 0.04$, slope=5% and $\alpha_{e,12} = 0\ s^{-1}$ and $\omega_{e,12} = 0.01\ s^{-1}$. The same rainfall rates as in Fig. 4.1 were applied on the impervious land surface. We can observe that the sediment transport rate at the outflow boundary in the active region is higher than in the passive region. Therefore, the composite sediment transport rate is greater than the sediment transport rate calculated by the sheet erosion model. The active transport region can be interpreted as a rill region with rill erosion, in which the occurrence of concentrated flow produces increased erosion and increased total sediment transport rates.

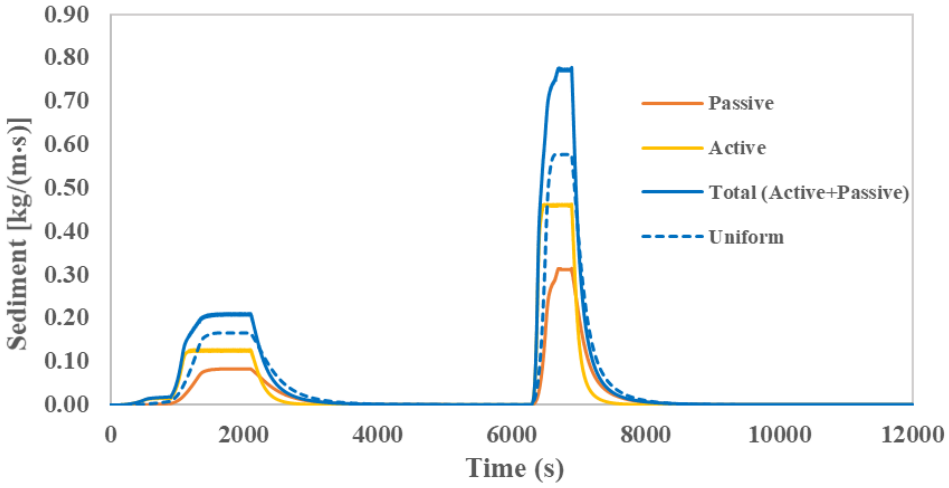


Figure 4.6. Sediment transport rates at the outflow boundary simulated using the Active-Passive Regions (APR) erosion model when rainfall (0.2 and 0.9 mm/s) is applied on the entire soil profile (the Manning coefficients n_1=0.01 and n_2=0.04, slope=5%, w_A =0.5).

Figure 4.7 presents simulated active, passive, and total sediment transport rates as a function of time when using the APR model with w_A = 0.5, n_1 = 0.01, n_2 = 0.05, slope=1%, $\omega_{e,12}$ = 0.01 s^{-1}, and $\alpha_{e,12}$ = 0, 0.01, and 0.001 s^{-1}. Overland across an impervious soil surface was initiated by setting the inlet boundary to a water depth of 1 cm for 20 mins during 60 mins simulation time. The sediment transport occurs independently in the active and passive domains when $\alpha_{e,12}$ =0 (blue lines). In this case, two sediment peaks are observed, with outflow starting after about 200 and 1000 s in the active and passive domains, respectively, due to different values in the roughness coefficient. An increase in $\alpha_{e,12}$ produces faster exchange of water and sediments between the two regions and causes a decrease in the peak of the sediment transport rate in the active region and a corresponding increase in the passive region. Figure 4.7

illustrates that the nonequilibrium sediment transport behavior, such as an early arrival, multiple peaks, and long-term tailing, can be obtained with the APR erosion model. It should be noted that this new modeling approach can represent an alternative approach to the existing interrill-rill erosion models.

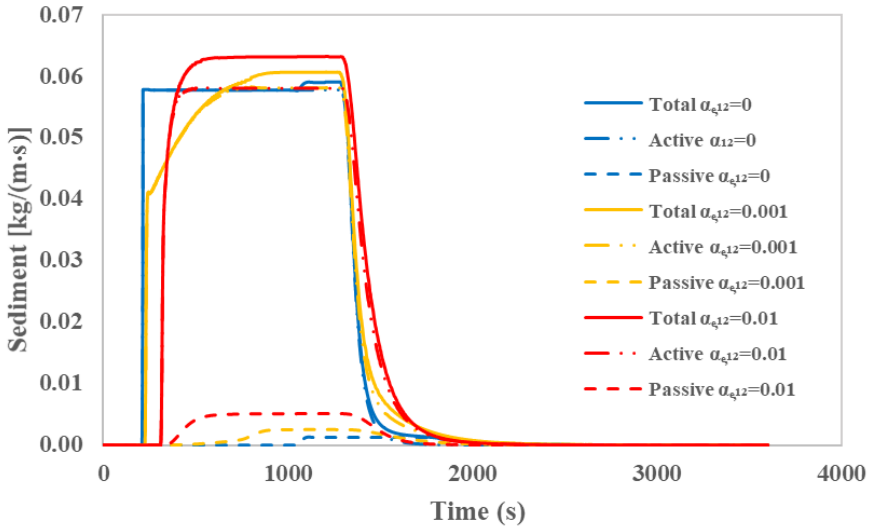


Figure 4.7. Sediment transport rates at the outflow boundary calculated using the active–passive regions (APR) erosion model with different values of the water exchange coefficient ($\alpha_{e,12}$=0, 0.001, and 0.01 s^{-1}) when inflow (h_0= 1 cm for 20 mins) is initiated from the upland. The Manning coefficients n_1=0.01 and n_2=0.05, slope=1%, w_A =0.5, and $\omega_{e,12}$ = 0.01 s^{-1}.

Finally, the effect of particle size on the sediment transport rates in the active and passive regions are shown in Fig. 4.8. In this case, the previous example from Fig. 4.7 is rerun with $\alpha_{e,12}$ =0.001 s^{-1} and particle sizes = 0.1 mm, 0.05 mm, and 0.01 mm. Smaller soil particles produce higher sediment transport rates due to the slower settling velocity in both active and passive regions (Fig. 4.8a and 4.8b, respectively). The velocity in the

passive region is lower than in the active region, resulting in higher sediment transport rates in the active region. However, the sediment transport in the passive region has a significant effect on the tailing and the peak of the composite sediment transport rate.

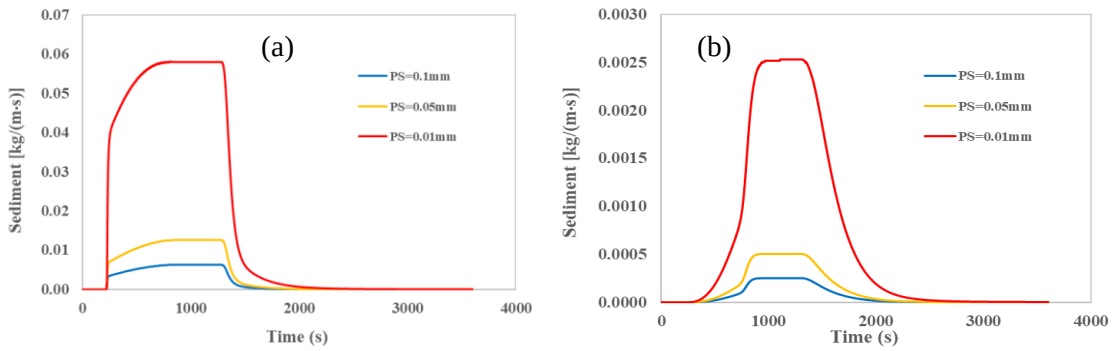


Figure 4.8. Sediment transport rates at the outflow boundary of the active (a) and passive (b) regions calculated using the active–passive regions (APR) erosion model with different particle sizes (0.1, 0.05, and 0.01mm). when inflow (h_0= 1 cm for 20 mins) is initiated from the upland. The Manning coefficients n_1=0.01 and n_2=0.05, slope=1%, w_A =0.5, $\alpha_{e,12}$ = 0.01 s^{-1}, and $\omega_{e,12}$ = 0.01 s^{-1}.

4.4 Summary and Conclusions

The HYDRUS-1D overland flow and transport code was extended to simulate sheet erosion and non-equilibrium erosion during sheet flow deviations. The code provides information on the temporal and spatial distributions of sediment concentrations in different flow regions (e.g., mobile and immobile zones, active and passive regions). The erosion models implemented into HYDRUS-1D provide a comprehensive set of tools to numerically investigate many important research problems involving sediment erosion, transport, and deposition processes. The HMIM (horizontal mobile-immobile) model can be a useful tool to simulate soil erosion at land surfaces containing depressions

and flow obstructions that retain surface water and sediments. The APR model can be an alternative modeling tool to investigate interactions between rill and interrill erosion when spatial patterns of the land surface are poorly characterized. However, these newly developed non-equilibrium sediment transport models may require a relatively large number of additional parameters that needs to be obtained by calibrating against laboratory or field measurements. Similar to HYDRUS-1D overland flow and transport models, the HYDRUS erosion model also includes provisions to estimate these parameters by inverse parameter optimization, and a variety of objective functions can be considered based on available measurements (e.g., water fluxes, water depths, resident concentrations, and flux concentrations). However, the current limitation of the HYDRUS erosion model is that it only considers a single mean soil particle size sediment transport and thus additional modifications are needed to consider entire particle size distributions.

4.5 References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O' Connell, P.E. and Rasmussen, J., 1986. An introduction to the European Hydrological System—Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system. *Journal of hydrology*, 87(1-2), 45-59.

Beasley, D.B., Huggins, L.F. and Monke, A., 1980. ANSWERS: A model for watershed planning. *Transactions of the ASAE*, 23(4), 938-0944.

De Jong, S.M., Paracchini, M.L., Bertolo, F., Folving, S., Megier, J. and De Roo, A.P.J., 1999. Regional assessment of soil erosion using the distributed model SEMMED and remotely sensed data. *Catena*, 37(3-4), 291-308.

Engelund, F. and Hansen, E., 1967. A monograph on sediment transport in alluvial streams. *Technical University of Denmark 0stervoldgade 10, Copenhagen K.*

Faeth, P., 1994. Building the Case for Sustainable Agriculture: Policy Lessons from India, Chile, and Chile, and the Philippines. *Environment: Science and Policy for Sustainable Development*, 36(1), 16-39.

Goodrich, D.C., Burns, I.S., Unkrich, C.L., Semmens, D.J., Guertin, D.P., Hernandez, M., Yatheendradas, S., Kennedy, J.R. and Levick, L.R., 2012. KINEROS2/AGWA: model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1561-1574.

Govers, G., 1990. Empirical relationships for the transport capacity of overland flow. *IAHS publication*, *189*, 45-63.

Hanley, N., Faichney, R., Munro, A. and Shortle, J.S., 1998. Economic and environmental modelling for pollution control in an estuary. *Journal of Environmental management*, 52(3), 211-225.

Johnson, B.E., Julien, P.Y., Molnar, D.K. and Watson, C.C., 2000. The two-dimensional upland erosion model CASC2D-SED 1. *JAWRA Journal of the American Water Resources Association*, 36(1), 31-42.

Kalin, L. and Hantush, M.M., 2003, Assessment of two physically-based watershed models based on their performances of simulating water and sediment movement". In *Conference Proceedings, 1st Interagency Conference on Research in the Watersheds (ICRW), Benson, AZ*.

Kennedy, J.R., Goodrich, D.C. and Unkrich, C.L., 2012. Using the KINEROS2 modeling framework to evaluate the increase in storm runoff from residential development in a semiarid environment. *Journal of Hydrologic Engineering*, 18(6), 698-706.

Knisel, W.G., 1980. CREAMS: a field scale model for Chemicals, Runoff, and Erosion from Agricultural Management Systems. *United States Department of Agriculture. Conservation research report (USA)*.

Liang, J., Bradford, S.A., Šimůnek, J. and Hartmann, A., 2017. Adapting HYDRUS-1D to Simulate Overland Flow and Reactive Transport during Sheet Flow Deviations. *Vadose Zone Journal*, 16(6).

Liu, Q.Q., Chen, L., Li, J.C. and Singh, V.P., 2007. A non‐equilibrium sediment transport model for rill erosion. *Hydrological Processes: An International Journal*, 21(8), 1074-1084.

Miller, S.N., Semmens, D.J., Goodrich, D.C., Hernandez, M., Miller, R.C., Kepner, W.G. and Guertin, D.P., 2007. The automated geospatial watershed assessment tool. *Environmental Modelling & Software*, 22(3), 365-377.

Morgan, R.P.C., 2009. Soil erosion and conservation. *John Wiley & Sons.*

Morgan, R.P.C., Quinton, J.N., Smith, R.E., Govers, G., Poesen, J.W.A., Auerswald, K., Chisci, G., Torri, D. and Styczen, M.E., 1998. The European Soil Erosion Model (EUROSEM): a dynamic approach for predicting sediment transport from fields and small catchments. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group*, 23(6), 527-544.

Nearing, M.A., Foster, G.R., Lane, L.J. and Finkner, S.C., 1989. A process-based soil erosion model for USDA-Water Erosion Prediction Project technology. *Transactions of the ASAE*, 32(5), 1587-1593.

Nearing, M.A., Simanton, J.R., Norton, L.D., Bulygin, S.J. and Stone, J., 1999. Soil erosion by surface water flow on a stony, semiarid hillslope. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 24(8), 677-686.

Renard, K.G., Laflen, J.M., Foster, G.R. and McCool, D.K., 1994. The revised universal soil loss equation. *Soil erosion research methods*, *2*, 105-124.

Singh, V.P. and Prasad, S.N., 1982. Explicit solutions to kinematic equations for erosion on an infiltrating plane. *Modelling Components of Hydrologic Cycle*, 515-538.

Storm, B. and Jorgensen, G.H., 1987. Simulation of water flow and soil erosion processes with a distributed physically-based modelling system. *Forest hydrology and watershed management/edited by RH Swanson, PY Bernier & PD Woodard*.

Torri, D., Borselli, L., Calzolari, C., Yañez, M. and Salvador-Sanchis, M.P., 2002. Soil erosion, land use, soil quality and soil functions: effects of erosion. *Man and soil at the third millennium*, 131-148.

Tucker, G., Lancaster, S., Gasparini, N. and Bras, R., 2001. The channel-hillslope integrated landscape development model (CHILD). In *Landscape erosion and evolution modeling. Springer*, 349-388.

Van Oost, K., Govers, G. and Desmet, P., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landscape ecology*, 15(6), 577-589.

Williams, J.R. and Berndt, H.D., 1977. Sediment yield prediction based on watershed hydrology. *Transactions of the ASAE*, 20(6), 1100-1104.

Wischmeier, W.H. and Smith, D.D., 1978. Predicting rainfall erosion losses-a guide to conservation planning. *Predicting rainfall erosion losses-a guide to conservation planning.*

Woodward, D.E., 1999. Method to predict cropland ephemeral gully erosion. *Catena*, 37(3-4), 393-399.

Woolhiser, D.A., 1989. KINEROS, a kinematic runoff and erosion model. *Documentation and Use Manual*.

Ziegler, A.D., Giambelluca, T.W. and Sutherland, R.A., 2001. Erosion prediction on unpaved mountain roads in northern Thailand: validation of dynamic erodibility modelling using KINEROS2. *Hydrological Processes*, 15(3), 337-358.

Chapter 5

Summary and Conclusions

The overall goal of this doctoral thesis was to develop a comprehensive set of modeling tools for the investigation of many important research problems involving overland flow, reactive transport, and sediment transport processes. Physically-based, spatially-distributed models, or machine learning algorithms trained on large experimental or model-generated datasets have the potential to be an efficient tool to examine and optimize the removal of contaminants from agricultural runoff through land-use changes and best management practices.

In Chapter 2, the popular HYDRUS-1D code was extended to simulate uniform and physical nonequilibrium overland flow and reactive solute transport (such as salts, nutrients, pesticides, and microbes). To demonstrate the applicability of the developed models, a large number of numerical examples was presented. The newly developed non-equilibrium overland flow and transport models included Horizontal Mobile-Immobile (HMIM), Vertical Mobile-Immobile (VMIM), Active-Passive Regions Model (APR), Combined Active-Passive regions and Horizontal Mobile-Immobile (APR-HMIM), and Combined Active-Passive regions and Vertical Mobile-Immobile (APR-VMIM). Using numerical experiments, we observed that the physical nonequilibrium models can produce hydrographs and/or solute BTCs with earlier or delayed arrivals, multiple peaks, and prolonged tailing. Model validation was achieved using published overland transport data in conjunction with the non-equilibrium overland transport model. Physical nonequilibrium models may be better suited than equilibrium models for studying hydrological processes at the plot and field scales when spatial patterns of land surfaces are poorly characterized.

In Chapter 3, a new data-driven model was proposed as an alternative to a physically-based overland flow and transport model. The HYDRUS-1D overland flow and transport model which was developed in Chapter 2 was used to develop a large database containing information about the impact of various relevant factors on the surface runoff quantity and quality. The database containing surface runoff water quantity and quality information was then used to develop correlations between model inputs (various relevant input factors discussed above such as rainfall rate, soil hydraulic conductivity, field slope, initial water content, and the Manning roughness coefficient) and outputs (surface runoff quantity and quality). Multiple data-driven models were tested on their ability to relate/correlate model inputs with outputs. The following Machine Learning (ML) models were used in the analysis: Linear Regression (LR), K-nearest neighbor regression (kNN), Artificial Neural Networks (ANN) (with one hidden layer), Support Vector Machine with linear (SVM-L) and non-linear (SVM-NL) kernels, and Deep Neural Networks (DNN) (Neural Networks with multiple hidden layers). The best performance was obtained by the DNN model with two hidden layers (DNN-2) (R^2=0.98).

In Chapter 4, the HYDRUS-1D overland flow and transport code was further extended to simulate sheet erosion and non-equilibrium erosion during sheet flow deviations. The developed models can be an efficient tool for the investigation of erosion processes under non-equilibrium water flow conditions. The Horizontal Mobile-immobile model and Active Passive Regions erosion models have the ability to predict earlier or

delayed arrivals, multiple peaks, and prolonged tailing in the sediment transport rate compared to traditional erosion modes.

The modeling approaches developed and described in this dissertation improve our ability to describe overland flow and transport processes in complex natural systems. The proposed models have the following features that make then different from other existing models:

1. First, the overland flow, solute transport, and sediment transport models are built based on the existing, widely-used HYDRUS code, which will allow many current HYDRUS users to easily operate the new models while having access to many useful features of the HYDRUS software. For example, the same graphical user interface (GUI) of HYDRUS-1D may be used to select and execute subsurface and overland flow and transport models.
2. Second, the developed models include a hierarchical system of physical equilibrium and nonequilibrium models of increasing complexity, from Uniform Flow and Transport, Horizontal Mobile-Immobile (HMIM), Vertical Mobile-Immobile (VMIM), Active-Passive Regions Model (APR), Combined Active-Passive regions and Horizontal Mobile-Immobile (APR-HMIM), to Combined Active-Passive regions and Vertical Mobile-Immobile (APR-VMIM). If one model does not provide a satisfactory description of the data, then another model with an additional level of complexity may be considered.

3. Third, the framework of data-driven models can be used for more complex scenarios with increased spatial and temporal scales to produce a fast data-driven model, which can mimic the computationally intensive physically-based model.

However, it should be emphasize that future work is requested to better define the physical meaning of additional parameters of the non-equilibrium overland flow and transport models.