On the Time Validity of John Philip's Two-Term Rainfall Infiltration Model

Yifu Gao

May 2020

Submission in partial satisfaction of requirements

for the degree of

MASTER OF SCIENCE

in Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Irvine

Thesis committee:

Associate Professor Jasper A. Vrugt, Chair

Professor Efi Foufoula-Georgiou

Associate Professor Russell Detwiler

On the Time Validity of John Philip's Two-Term Rainfall Infiltration Model

By

Yifu Gao

Contents

List of Figures

List of Tables

Acknowledgement

I would like to express the my deepest thanks to my committee chair, Professor Jasper A. Vrugt, who's a talented and friendly person offering me generous help and guidance on my research. His rigorous and logical attitude towards scientific research has always been promoting me to do better works. It is such an honor for me to meet him during my master program. I am so glad that I can continue my academic life as his doctoral student.

I would like to thank my committee members, Professor Efi Foufoula-Georgiou and Professor Russell Detwiler, for kindly spending their time and offering me valuable evaluation and suggestions to my thesis. Their instructions on stochastic hydrology and ground water hydrology have enormously broadened my knowledge of data analysis and hydrology.

Lastly, I want express my deepest gratitude to my parents, who has been supporting me financially and emotionally through my life. It is you who let me choose the life I want to live and make me stronger. I can never make those accomplishments without you.

Abstract

On the Time Validity of John Philip's

Two-Term Rainfall Infiltration Model

by

Yifu Gao

Master of Science in Civil and Environmental Engineering

University of California, Irvine, 2020

Associate Professor Jasper A. Vrugt, Chair

Rainfall infiltration, the process wherein water enters the soil surface and replenishes moisture in the vadose zone, is an important component of the water balance and hydrologic cycle. Infiltration guarantees a continued availability of moisture to sustain root water uptake, plant growth, groundwater recharge and soil structure. There are several ways to estimate rainfall infiltration rates and volumes. The most rigorous approach would use a partial differential equation (Richards' equation), coupled, if necessary, with a surface water routine and groundwater model (Darcy's law), to describe infiltration into variably-saturated soils. Analytic solutions of Richards' equation and/or Darcy's law and empirical infiltration functions may work well under certain conditions (deep-drained soils with uniform initial moisture content) and/or single rainfall events. Among all research conducted, the Philip's two-term infiltration model, $I = S\sqrt{t} + cK_s t$,

where I [L] is the cumulative infiltration, S [L·T$^{-1/2}$] signifies the sorptivity, K_s [L·T^{-1}] denotes the saturated soil hydraulic conductivity, c is a unitless curve fitting coefficient and t denotes time in units of length, has found widespread use and applicability. This model is particularly easy to use as (1) it only has three unknown parameters, (2) the least squares parameter values are easily determined from experimental data using linear regression, and (3) two of the estimated parameters, S [L·T$^{-1/2}$] and K_s [L·T^{-1}], have a clear physical significance. In favor of this simplicity, Philip's two-term infiltration model eliminates higher-order terms of a polynomial series of time that account for the effect of gravity on infiltration. This effect becomes more important at later times as a larger proportion of the soil reaches saturation and the soil water pressure head gradient becomes negligible. As a result, Philip's two-term infiltration model, $I = S\sqrt{t} + cK_s t$, has a limited time validity, t_{valid} [T]. In his work, Philip provides theoretical guidance on the time validity of his two-term infiltration model. This time validity is of great importance as it determines the time span of experimental infiltration data to use for parameter estimation.

In this research, we explore the time validity of Philip's two-term infiltration model using Bayesian inference and the , Soil Water Infiltration Global (SWIG) database. This database consists of a large ensemble of measured cumulative infiltration curves of a wide variety of soils worldwide. Essentially, we test, benchmark and evaluate the approach of Jaswal et al. (2020) on measured data rather than synthetic infiltration data simulated with HYDRUS-1D. The methodology consists of two parts. First, we determine the values of the parameters S

$[\,\mathrm{L}\cdot\mathrm{T}^{-1/2}\,]$ and K_s $[\,\mathrm{L}\cdot\mathrm{T}^{-1}\,]$ via Bayesian inference of the Haverkamp infiltration equation using the **D**iffe**R**ential **E**volution **A**daptive **M**etropolis (DREAM) algorithm. As semi-implicit solution of Richards' equation, the Haverkamp model is valid for the entire duration of the infiltration experiment. Then, the posterior distribution of the sorptivity and saturated soil hydraulic conductivity of each measured infiltration curve are used in Philip's two-term infiltration model to determine the optimal value of the coefficient c via linear regression. We implement the Bayesian information criterion (BIC) to return, as byproduct of our analysis, the optimal time validity of Philip's two-term infiltration model. The uncertainty of the time validity, t_valid $[\,\mathrm{T}\,]$, can be estimated by evaluating the different posterior samples of S $[\,\mathrm{L}\cdot\mathrm{T}^{-1/2}\,]$ and K_s $[\,\mathrm{L}\cdot\mathrm{T}^{-1}\,]$. We particularly focus on the "best" samples of each soil type in the SWIG database as results confirm that the temporal resolution of the infiltration data plays a critical role. Results demonstrate that coarse textured soils (e.g. sand, loamy sand, sandy loam) have a rather small value of t_valid $[\,\mathrm{T}\,]$ ranging between 0.10 hour to 1.00 hour. Medium textured soils (sandy clay loam, loam, clay loam) exhibit somewhat larger values of the time validity ranging between 1.00 hour to 4.76 hours. Unfortunately, the measured infiltration curves in the SWIG database did not allow us to determine adequate values of the time validity for fine textured soils. The time validity, t_valid $[\,\mathrm{T}\,]$ of clay loam, silty clay loam, silty clay, and clay soils was simply equal to the time of the last infiltration measurement. In other words, the experiments did not last long enough to determine accurately their respective time validity.

All results were compared to those of Jaswal et al. (2020) using synthetic infiltration data. This analysis made evident that (1) the measurement errors of the infiltration data increase the uncertainty of t_{valid} [T]; (2) The much poorer measurement (time) resolution of the infiltration data in the SWIG database makes it difficult to accurately determine the time validity of Philip's two-term infiltration model; (3) For fine textured soils, the infiltration experiments were of insufficient length to reliably estimate the value of t_{valid} [T]. Altogether, we conclude that it is not particularly easy to estimate the time validity of Philip's two-term infiltration model from measured cumulative infiltration data. A large cohort of the infiltration experiments in the SWIG database lack the temporal resolution and necessary length of the experiment to warrant an accurate determination of the time validity. Thus, we recommend using synthetic infiltration data simulated derived from numerical solution of Richards' equation to determine an approximate time validity for each soil type. The resulting estimates of t_{valid} [T] can then serve as guidelines for analysis of real-world infiltration experiments.

1 Introduction

1.1 Definition of infiltration

Infiltration is the process of water entering the soil from the ground surface under the influence of gravitational and capillary forces. This water may originate from rainfall or irrigation and will be stored in the pores of the soil matrix. If the top soil cannot immediately absorb all the water at the soil surface, then water may pond on the soil surface and/or runoff and concentrate in other areas depending on the local topography. Infiltration replenishes the moisture in the unsaturated zone and, hence, is an important component of the soil water balance and hydrologic cycle. At the field scale, we can write the water balance as follows

$$P + W - ET - R = I \tag{1.1}$$

Where P [L] signifies precipitation, W [L] denotes irrigation, ET [L] is the evapotranspiration, R [L] is the runoff and I [L] represents the cumulative infiltration. Evapotranspiration is sum of transpiration and evaporation, $ET = E + T$. The cumulative infiltration can be further written as

$$I = \Delta S + G \tag{1.2}$$

Where ΔS signifies the change in storage, G is the drainage or groundwater flow.

Infiltration plays an important role in hydrology cycle. According to the water balance equation, infiltration constitutes the recharge of the soil water and the

drainage or groundwater flow. Through recharging the soil layer with water, infiltration maintains the groundwater level, soil moisture, the structure of the soil, and the depth of water bodies. Besides, infiltration provides plants with water source, enhancing the structure of the root zone. Due to the infiltration, the proportion of runoff from the precipitation gets smaller (compared with no infiltration), reducing the flood risk to a large extent.

1.2 Mechanisms and factors of infiltration

Due to the significant impact of infiltration on hydrology cycle, the study of infiltration attracts great amounts of scientific research and experiments on that, with a product of many mathematical models. Before the deeper learning of the infiltration model, it is suggested that we have a good understanding of the mechanisms and factors of infiltration.

The mechanism of 1-D infiltration rate or vertical flow can be described as a result of hydraulic gradient in the vadose zone (unsaturated zone):

$$i = K(\theta)\frac{\partial H}{\partial z} = K(\theta)\left(\frac{\partial h}{\partial z} + 1\right) \tag{1.3}$$

Where z [L] is the vertical distance, H [L] is the hydraulic head which can be written as a combination of soil water pressure head and potential head: $h + z$. $K(\theta)$ [$\mathrm{L}\cdot\mathrm{T}^{-1}$] is the hydraulic conductivity, which depends on soil moisture content θ [$\mathrm{L}^3\cdot\mathrm{L}^{-3}$]. This upper partial differential equation effectively explains the mechanisms of infiltration. As there is water present at the ground surface, there is great hydraulic gradient between the surface and subsurface since there is nearly no moisture in the subsurface zone. This head difference or gradient

drives the water down into the soil. The moisture content increases as water infiltrates in, causing the soil water pressure head to increase. Eventually, the soil get saturated and θ $[\mathrm{L}^3{\cdot}\mathrm{L}^{-3}]$ goes to the saturated moisture content θ_s $[\mathrm{L}^3{\cdot}\mathrm{L}^{-3}]$, which is a constant for a given soil. At the same time, $K(\theta)$ $[\mathrm{L}{\cdot}\mathrm{T}^{-1}]$ will turn to K_s $[\mathrm{L}{\cdot}\mathrm{T}^{-1}]$ as the saturated hydraulic conductivity. If a constant head gradient presents at the surface, it is reasonable to conclude that we will get a constant infiltration rate after saturation.

To further interpret infiltration, it is essential to study the controlling factors of infiltration. Generally, there are three significant factors:

(1) Precipitation

The intensity and duration of precipitation are essential controlling factors of infiltration. For a rainfall with small intensity, the hydraulic gradient at the surface is relatively small, contributing to a slow infiltration rate because the water infiltrates into the soil and continues going down under the capacity of the soil, leaving no runoff at the surface. The infiltration rate will be equal to the rainfall intensity. By contrast, an intense precipitation leads to an overload of water on the ground, where the rainfall intensity overwhelms the infiltration capacity the soil can hold and runoff forms. Given such situation, the infiltration rate at the surface will be equal to its infiltration capacity at saturation.

(2) Soil properties

The soil properties associated with the soil types are influential factors of infiltration. Richards (1931) introduced a partial differential equation to study the unsaturated groundwater flow. The Richards equation vesus vertical flow could

be written as [Richards 1931]:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial h}{\partial z}\right)\right] \tag{1.4}$$

Where $\frac{\partial h}{\partial z}$ is the vertical gradient of soil water pressure head, θ [$\mathrm{L^3 \cdot L^{-3}}$] and $K(\theta)$ [$\mathrm{L \cdot T^{-1}}$] are soil properties which represent the moisture content and hydraulic conductivity. Similar to other hydrologic cases, the unsaturated zone flow is a result of hydraulic gradient, but different in the flow media which varies spatially with respect to soil types, soil structures, and so on, leaving it hard to solve the unsaturated flow from a microscopic aspect. This leads to studies on the relationships between the soil properties and numerical solutions to the partial differential equation. The water retention function curve (WRF) is a commonly used measure to study the relationship between soil water pressure head and soil moisture content [Van Genuchten 1980]:

$$S_{\mathrm{e}}(h) = [1 + (\alpha|h|)^n]^{-m} \tag{1.5}$$

Where α is the reciprocal of the air entry value; m is a parameter equals to $1 - 1/n$, and n is the parameter that is proportional to the inverse of the slop of the water retention curve of that soil; S_{e} [-] signifies the effective saturation:

$$S_{\mathrm{e}}(h) = \frac{\theta - \theta_{\mathrm{r}}}{\theta_{\mathrm{s}} - \theta_{\mathrm{r}}} \tag{1.6}$$

Where θ [$\mathrm{L^3 \cdot L^{-3}}$] is the current moisture content; θ_{s} [$\mathrm{L^3 \cdot L^{-3}}$] is the saturated moisture content and θ_{r} [$\mathrm{L^3 \cdot L^{-3}}$] is the residual moisture content. **Fig 1.1** shows the water retention fucntion curves for 12 types of USDA soils, using parameters estimated from Carsel and Parrish (1988) through HYDRUS-1D software [Vogel et al. 1996]. **Table 1.1** provides information on the Van Genuchton

Table 1.1: Water retention function parameters of 12 soils

Texture	θ_r (cm^3/cm^3)	θ_s (cm^3/cm^3)	α -	n -
Sand	0.045	0.430	0.145	0.627
Loamy Sand	0.057	0.410	0.124	0.561
Sandy Loam	0.065	0.410	0.075	0.471
Sandy Clay Loam	0.100	0.390	0.059	0.324
Sandy Clay	0.100	0.380	0.027	0.187
Loam	0.078	0.430	0.036	0.359
Silty Loam	0.067	0.450	0.020	0.291
Silt	0.034	0.460	0.016	0.27
Clay Loam	0.095	0.410	0.019	0.237
Silty Clay Loam	0.089	0.430	0.010	0.187
Silty Clay	0.070	0.360	0.005	0.083
Clay	0.068	0.380	0.008	0.083

parameters of the 12 soils. [Carsel and Parrish 1998]

(3) Soil types

The type of the soil is determined through a soil classification process which primarily takes the particle size as an index for separating the soils. Particle size affects the soil properties to a great extent, where it affects the packing

Table 1.2: Water retention function parameters of 12 soils

Texture	Texture Class (*)	K_s (cm/hour)
Coarse sands	Coarse	50.81
Sands, Loamy sands	Coarse	15.24-50.81
Sandy loam	Mod. coarse	5.08-15.24
Loam, Silty loam, Silt	Medium	1.52-5.08
Clay loam, Sandy clay loam, Silty clay loam	Mod. fine	0.51-1.52
Sandy clay, Silty clay, Clay	Fine and very fine	0.15-0.51

*The texture class is specified according to the particle size. (e.g. coarse: 0.5 – 1.0 mm in diameter)

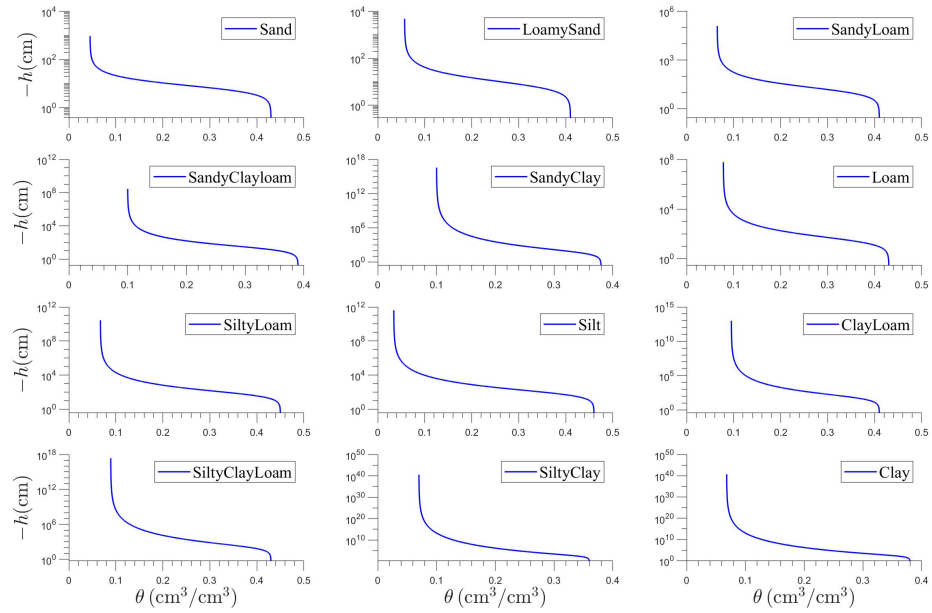

Figure 1.1: Water retention function curves of 12 types of soil

of the particles and the empty space between the particles. Thus, a difference in particle size leads to the variation of properties like porosity, saturated and residual moisture content, and so on. The soil classification system (e.g. USDA) classifies soils into gravel, sand, silt, and clay (in the order of decreasing particle size) where 12 types are further determined (sand, loamy sand, loam, clay loam, etc.). **Fig 1.1** shows how θ_{s} $[\mathrm{L}^3{\cdot}\mathrm{L}^{-3}]$, θ_{r} $[\mathrm{L}^3{\cdot}\mathrm{L}^{-3}]$ varies with respect to the soil types. **Table 1.2** also illustrates the relationship between saturated hydraulic conductivity and soil types.

1.3 Infiltration model and Philip model

Like previously discussed, the mechanism of an infiltration process in the unsaturated soil is based on a partial differential equation, which is difficult to solve straightly. Consequently, the recent years' studies of infiltration have turned to the numerical models. Unlike the vertical flow equation, infiltration models can easily produce simulation of the cumulative infiltration and the optimization of the models will enable the simulations to proximate the observation. The numeric models usually make assumptions to simplify the infiltration process (e.g. assumptions on the time, simplification of the infiltration mechanisms). This leads to the model structural error which is inevitable. Despite this, the residuals (difference between simulations and observations) can be minimized to get an optimal estimation of the infiltration by analyzing the best fitted parameters of the model. Therefore, infiltration model is a more efficient and effective approach to study the infiltration than the partial differential equation, where the focus of model study is to increase the model efficiency by minimizing the residuals.

There are two major methods to study the infiltration model: analytical and empirical methods. The empirical approach focuses on using numeric models and parameters. The process or experiment of fitting empirical models to the data enables researchers to derive the optimal estimated parameters regarding the sample. One typical empirical method is Horton model [Horton 1941], which expresses the infiltration rate as an exponential decay regarding the time with a parameter k.

The analytical method is aimed at solving the vertical flow equation and deriving the mathematical solutions to that partial differential equation by simplifying the infiltration mechanisms. One typical analytical infiltration model is Green-Ampt model [Green and Ampt 1991], which simplifies the wetting front by assuming a constant soil water pressure head at the wetting front:

$$I = (\theta_{\mathrm{s}} - \theta_{\mathrm{i}}) \cdot \sqrt{2D(\theta_{\mathrm{s}})t} \quad \text{(short time)} \tag{1.7}$$

$$I = K_{\mathrm{s}} \cdot t \quad \text{(long time)} \tag{1.8}$$

Where $D(\theta_{\mathrm{s}})$ $[\,\mathrm{L}^2 \cdot \mathrm{T}^{-1}\,]$ is the soil diffusivity, which is defined as: $D(\theta_{\mathrm{s}}) = K_{\mathrm{s}} \cdot \frac{\mathrm{d}h}{\mathrm{d}\theta}$. Besides, there are many analytical approaches to simulate the infiltration, for example: Philip (polynomial series) [Philip 1957a], Mein and Larson (combination of surface ponding and Green-Ampt) [Mein and Larson 1973], Smith (two branched model for ponding time and infiltration rate decay) [Smith 1978], Haverkamp (semi-implicit equation solving the Richard equation) [Haverkamp, Parlange, et al. 1990], Valiantza (linearlized two-parameter infiltration derived from Philip's two-term model) [Valiantza 2010]. These models varies within parameters and dimensions, where each model has their own restriction on time stage and condition of inflow (e.g. steady rainfall intensity).

Among all the numeric models discussed, Philip infiltration model receives the greatest popularity and has found wide applications in soil hydraulic property study. The Philip 1-D infiltration model was first introduced by John Philip (1957), providing a numeric solution to the partial differential equation describing the vertical infiltration. This model describes the cumulative infiltration

through polynomial series:

$$I(t) = A_1 t^{1/2} + A_2 t + A_3 t^{3/2} + A_4 t^2 + ... \tag{1.9}$$

Where $A_1, A_2, A_3...$ are the polynomial parameters of different units. Philip [Philip 1957b] illustrated that A_1 is defined as the 'sorptivity', which is a term that measures the soil's ability to uptake or absorb the liquid through capillarity. The second parameter A_2 is a product of saturated hydraulic conductivity K_{s} [$\mathrm{L{\cdot}T^{-1}}$] and an unknown coefficient c.

The Philip 1-D model shows two advantages: (1) Given measured data, the optimal parameters can be easily derived through linear regression. (2) Polynomial series can help weight each term at different time, showing the different stages of infiltration. Despite this, the dimension of Philip greatly affects the simulation of infiltration, where Runge's phenomenon [Runge 1931] probably exists when we adopt high-dimensional Philip model. This will lead to the fluctuation of simulated infiltration when analyzing high resolution of time. Additionally, unreasonable interpolation of data points might happen due to the Runge's phenomenon. To address this problem, the two-term (d=2) version of Philip model was introduced, which is the exact formula that has been widely used:

$$I(t) = S \cdot t^{1/2} + c \cdot K_{\mathrm{s}} t \tag{1.10}$$

The Philip's two-term receives popularity based on several reasons: (1) it only has three unknown parameters, (2) the least squares parameter values are easily determined from experimental data using linear regression, and (3) two of the estimated parameters, S [$\mathrm{L{\cdot}T^{-1/2}}$] and K_{s} [$\mathrm{L{\cdot}T^{-1}}$], have a clear physi-

cal significance. Regardless of these advantages, the Philip's two-term removes the later terms from the the polynomial series, which function as the effect of gravity (the infiltration is dominated by the potential gradient) as the soil get close to saturation. This means that Philip's two-term might simulate effective infiltration given a limited time length, where the difference of simulated curves using two-term and polynomial series will start to increase after that time. This specific time is called the time validity t_{valid} of Philip's two-term model. The time validity is really important because it tells the feasible time length to analyze the data, which produces the optimal model efficiency of Philip's two-term. Based on the existence of t_{valid}, it is not suggested to directly use the entire length of data while studying the soil hydraulic properties using Philip's two-term.

The objective of this paper is to explore the soil hydraulic parameters: S, K_{s}, c and time validity of Philip's two-term model with respect to 10 types of soils given the infiltration data from experimental data (SWIG database). In this paper, a new approach of finding the time validity was introduced through using Bayesian Information Criterion as a criterion (BIC) [Schwarz 1978] for model efficiency to test the optimal performance of Philip's two-term and its corresponding time as t_{valid}. The general steps of this research are as followed. First, in Section 3.1, we adopted the semi-implicit Haverkamp model (a model which does not have restrictions of time validity) to explore the optimal S $[\,\text{L}\cdot\text{T}^{-1/2}\,]$ and K_{s} $[\,\text{L}\cdot\text{T}^{-1}\,]$ of the samples. In section 3.2, 1-D infiltration Data from the SWIG database were collected and preprocessed by analyzing the measuring

instrument and plotting the infiltration curves, and low-quality data were discarded. Then, in section 3.3, the algorithm **D**iffe**R**ential **E**volution **A**daptive **M**etropolis (DREAM) [Vrugt 2016] was adopted to analyze the data and derive the posterior distribution and optimum of these two parameters. In Section 3.4, the optimal S $[\mathrm{L{\cdot}T^{-1/2}}]$ and K_s $[\mathrm{L{\cdot}T^{-1}}]$ were input to the Philip's two-term model. To explore the t_valid, we use linear regression method to calculate the optimal coefficient c and BIC by analyzing different time lengths. The optimal coefficient c and t_valid were derived with respect to the minimum BIC. In section 4, several results were presented which were derived by analyzing different samples. Section 4.1 covered the results derived from the total 646 samples. To address the issue of low resolution of time and measurement errors, best sample of each soil type were analyzed again in section 4.2. Results of the DREAM output posterior distribution of S $[\mathrm{L{\cdot}T^{-1/2}}]$ and K_s $[\mathrm{L{\cdot}T^{-1}}]$ were presented. Tables covering the optimal parameter values were illustrated. To resolve the problem of low resolution of measured times, interpolation was done and new results were derived in section 4.3, where uncertainty analysis of the t_valid and coefficient c were also covered, along with an output of the soil textural triangle of time validity. Section 5 states the conclusions and limitations of this research.

2 SWIG database

The Soil Water Infiltration Global (SWIG) database [Mehdi Rahmati 2018] covers the infiltration curves (time series based cumulative infiltration) produced by

Soil No	%	%	%			%	mm		`%	g/cm3
Code	Clay	Silt	Sand	Texture Cla	TextureCode	Gravel	dg	Sg	OC	Db
216	40.000	40.000	20.000	SILTY CLA	11		0.015	13.112	1.050	1.280
217										
218										
219										
220	14.000	38.000	48.000	LOAM	6		0.096	12.199	1.010	1.370
221	14.000	38.000	48.000	LOAM	6		0.096	12.199	1.010	1.370
222	14.000	38.000	48.000	LOAM	6		0.096	12.199	1.010	1.370
223	32.000	34.000	34.000	CLAY LOA	9		0.032	16.753	0.990	1.290
224	32.000	34.000	34.000	CLAY LOA	9		0.032	16.753	0.990	1.290

Figure 2.1: Missing information on soil properties in SWIG database

experiments conducted in 54 countries from a global scale (majorly contributed by Iran, China, and USA). According to this database, over 5,000 samples were tested, providing researchers a great support of real-world soil water infiltration data, along with location information, measuring instrument, land use and so on. In addition to data-related information, the database contains information on soil properties like soil texture, bulk density, initial moisture content, saturated moisture content and so on. Therefore, the SWIG database is of enormous value for analyzing real-world infiltration data and studying infiltration models through fitting the models to the data. Unlike computer simulated data (e.g. HYDRUS 1D infiltration data) which is sensitive only to the input soil hydraulic parameters, the SWIG database provides realistic data which are subject to a great many factors such as measuring instrument, measured time, location, precipitation, etc. Through analyzing real-world data, scientist will not only estimate soil hydraulic parameters and optimize the infiltration model, but also explore more information on the model structural error (the error of the model itself).

Since the SWIG database consists of infiltration data from worldwide, there

are great variations of the measured data with respect to the length of time, time interval, measurement instrument, and size of data. This is a common phenomenon to deal with the real-world data where data quality requires to be double-checked and a preprocess of the data is also needed before analyzing. For instance, 5,023 samples were tested for infiltration and other soil properties, where only 3,842 of the total samples cover the information of soil texture (see **Fig 2.1**), which is a critical property of the samples providing information on the soil constituents and particle size. Even though it is possible to analyze the data and other soil properties to infer the soil texture, discarding those samples will always be a good approach for preciseness. Another example is that the measurement instrument plays a significant role in choosing the infiltration model. In fact, the parameters and structure of the infiltration models are affected by the dimension of the flow (1-D or 3-D infiltration). And the choosing of the measuring instrument will seriously affect how the water flow through the unsaturated sample. To illustrate, disc infiltrometer and single-ring infiltrometer are commonly considered as measuring 3-d infiltration [Bouwer 1986] while double-ring infiltrometer usually measures the 1-d infiltration. In this research, the targeted model is based on vertical infiltration (1-d). Thus, filtering out samples measured by double-ring infiltrometer will be suggested.

3 Methods

3.1 Haverkamp 1-D Infiltration Model

The Haverkamp 1-D cumulative infiltration model [Haverkamp, Ross, et al. 1994] was selected to estimate the optimal parameters of soil sorptivity and saturated hydraulic conductivity. As the objective of this research is to determine the time validity for Philip model, a problem has come up that the soil saturated hydraulic conductivity (K_{s} [$\mathrm{L{\cdot}T^{-1}}$]) and sorptivity (S [$\mathrm{L{\cdot}T^{-1/2}}$]) could not be directly derived through analyzing Philip model. Given that the model input (time) is not determined, it will be meaningless to apply statistical methods (e.g. linear regression, MCMC) to this model for parameter optimization. An effective approach to deal with this problem is to use another infiltration model which also contains K_{s} [$\mathrm{L{\cdot}T^{-1}}$] and S [$\mathrm{L{\cdot}T^{-1/2}}$] as model parameters. That is the reason why Haverkamp 1-D infiltration model was introduced to this study. This model is a semi-implicit equation which shares two parameters with Philip model and solves the cumulative infiltration [Haverkamp, Ross, et al. 1994]:

$$\frac{2(K_{\mathrm{s}}-K_{\mathrm{i}})^2}{S_0^2}\cdot t(1-\beta) = 2\cdot\frac{(K_{\mathrm{s}}-K_{\mathrm{i}})(I-K_{\mathrm{s}}t)}{S_0^2} - ln\Big\{\frac{1}{\beta}\cdot exp\left[\frac{2\beta(K_{\mathrm{s}}-K_{\mathrm{i}})(I-K_{\mathrm{s}}t)}{S_0^2}\right]+\frac{\beta-1}{\beta}\Big\} \tag{3.1}$$

Where I and t are the cumulative infiltration and time. β is a shape parameter that is unitless. K_{i} [$\mathrm{L{\cdot}T^{-1}}$] and K_{s} [$\mathrm{L{\cdot}T^{-1}}$] are the initial (at t=0) and saturated hydraulic conductivity, S [$\mathrm{L{\cdot}T^{-1/2}}$] signifies the sorptivity, which was defined by John Philip as an estimation of the soil's ability to capture liquid

due to capillary effects [Philip 1957b]. The Haverkamp implicit formulation is derived through solving Richards equation for cumulative infiltration using analytical methods [Lassabatere et al. 2009; Haverkamp, Parlange, et al. 1990; Richards 1931]. The Richards equation vesus vertical flow could be written as:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial h}{\partial z} + 1\right)\right] \tag{3.2}$$

Where h represents the water pressure head, z is the vertical distance, θ [$\mathrm{L^3 \cdot L^{-3}}$] is the moisture content, $K(\theta)$ [$\mathrm{L \cdot T^{-1}}$] is the hydraulic conductivity which is a function of θ [$\mathrm{L^3 \cdot L^{-3}}$]. Therefore, the advantage of Haverkamp model is clear that it is flexible to any time input without restriction to the time validity, allowing the whole time series input for parameter optimization. Another advantage of applying Haverkamp model is that, the two models share the same parameters K_s [$\mathrm{L \cdot T^{-1}}$] and S [$\mathrm{L \cdot T^{-1/2}}$]. Through importing the optimal K_s [$\mathrm{L \cdot T^{-1}}$] and S [$\mathrm{L \cdot T^{-1/2}}$] as known parameter, we improve the efficiency of solving Philip model by reducing the dimension of the model from four to two. (Initially four unknowns: K_s [$\mathrm{L \cdot T^{-1}}$], S [$\mathrm{L \cdot T^{-1/2}}$], c and t_valid; Now two unknowns: c and t_valid) What's more, solving the Philip model will change from a non-linear problem to a linear problem where linear regression method can be easily conducted to find the optimal coefficient c.

Parameter optimization for the Haverkamp 1-D model is a four-dimension problem where K_i [$\mathrm{L \cdot T^{-1}}$], K_s [$\mathrm{L \cdot T^{-1}}$], S [$\mathrm{L \cdot T^{-1/2}}$], β are required to be estimated. We can reduce the dimension of this problem by writing the initial hydraulic conductivity as a function of the K_s [$\mathrm{L \cdot T^{-1}}$]. Some of the samples in the SWIG database provide information on initial moisture content θ_i [$\mathrm{L^3 \cdot L^{-3}}$], which can

be used to derive the initial hydraulic conductivity through the following steps. According to Van Genuchton (1980), the hydraulic conductivity for unsaturated soils is a function of the effective saturation (dimensionless water content) and saturated hydraulic conductivity:

$$K(S_{\mathrm{e}}) = K_{\mathrm{s}} S_{\mathrm{e}}^{L} \left[1 - \left(1 - S_{\mathrm{e}}^{\frac{1}{m}}\right)\right]^{2} \tag{3.3}$$

Where L is an empirical parameter which is normally assumed to be 0.5; m is a parameter equals to $1 - 1/n$, where n is the parameter that is proportional to the inverse of the slop of the water retention curve of that soil; Information on parameter n is available in Table 1.1. S_{e} [-] signifies the effective saturation:

$$S_{\mathrm{e}} = \frac{\theta - \theta_{\mathrm{r}}}{\theta_{\mathrm{s}} - \theta_{\mathrm{r}}} \tag{3.4}$$

Where θ [$\mathrm{L}^3 \cdot \mathrm{L}^{-3}$] is the current moisture content; θ_{s} [$\mathrm{L}^3 \cdot \mathrm{L}^{-3}$] is the saturated moisture content and θ_{r} [$\mathrm{L}^3 \cdot \mathrm{L}^{-3}$] is the residual moisture content. In refer to **Table 1.1**, the residual and saturated moisture content of different types of soil are also provided. Therefore, through combining Eq 3.3 and Eq 3.4, we can derive the equation for computing the initial hydraulic conductivity, which is a function of the saturated hydraulic conductivity:

$$K_{\mathrm{i}} = A \cdot K_{\mathrm{s}} \tag{3.5}$$

Where:

$$A = \left(\frac{\theta - \theta_{\mathrm{r}}}{\theta_{\mathrm{s}} - \theta_{\mathrm{r}}}\right)^{0.5} \left[1 - \left(1 - \left(\frac{\theta - \theta_{\mathrm{r}}}{\theta_{\mathrm{s}} - \theta_{\mathrm{r}}}\right)^{\frac{1}{m}}\right)^{m}\right] \tag{3.6}$$

Given the input of m, θ_{i} [$\mathrm{L}^3 \cdot \mathrm{L}^{-3}$], θ_{r} [$\mathrm{L}^3 \cdot \mathrm{L}^{-3}$], and θ_{s} [$\mathrm{L}^3 \cdot \mathrm{L}^{-3}$], coefficient A will be a constant for a specific soil sample. Combined with the formulas of unsaturated hydraulic conductivity [Van Genuchten 1980], parameter estimation

of Haverkamp 1-D model turns to the estimation of K_{s} [$\mathrm{L{\cdot}T^{-1}}$], S [$\mathrm{L{\cdot}T^{-1/2}}$], and β.

To derive the optimal K_{s} [$\mathrm{L{\cdot}T^{-1}}$], S [$\mathrm{L{\cdot}T^{-1/2}}$], and β of a sample, the usual method is: (1)Input the parameters and time to the Haverkamp 1-D model; (2)Solve the cumulative infiltration; (3)Compare the simulations with real data and apply a Markov Chain Monte Carlo method (will be discussed in the following section) to derive the posterior distribution and optimal value for each parameter. One deficiency of using the Haverkamp 1-D model is that it is an semi-implicit formula which does not directly provide the solution of the cumulative infiltration. To solve this equation, Haverkamp et al. (1994) introduced the approximation of the solutions to the quasi-exact equation based for very short time, short time, and long time [Haverkamp, Ross, et al. 1994]:

$$I_{1D}(\text{very short}) = S\sqrt{t} \tag{3.7}$$

$$I_{1D}(\text{short}) = S\sqrt{t} + \left[\frac{2}{3-\beta}\cdot(K_{\mathrm{s}} - K_{\mathrm{i}}) + K_{\mathrm{i}}\right] t \tag{3.8}$$

$$I_{1D}(\text{long}) = K_{\mathrm{s}} t \tag{3.9}$$

Based on the current study, the validity time interval for the upper equation is not determined for all the soil types [Lassabatere et al. 2009; Vandervaere, Vauclin, and Elrick 2000b]. Some previous studies indicated that for sands and loams, the validity times for the first approximation is too short to determine the sorptivity [Vandervaere, M. Vauclin, and Elrick 2000a]. So it is relatively tough to choose which upper equation to be used. Instead of using the approximation of the solution, the root-finding methods was applied to solve the equation

in our study. There are some built-in functions in MATLAB (e.g. vpasolve, fzero, etc) that could be adopted. In this research, the application of MCMC method requires huge amounts of computer simualtions to derive the posterior distribution of each parameters, where the built-in root-finding functions are time-consuming. A secant method was introduced in this paper to solve the Haverkamp 1-D equation, which efficiently produces the result.

A secant method is a root-finding method that makes use of the secant lines of two points on the curve. This is analogous to the Newton's method but secant method directly adopts the secant lines instead of the derivative to update the root [Avriel 1976]. The mathematical expression of secant method can be decribed as a recurrence relation:

$$x_n = x_{n-1} - \frac{f_{n-1}}{k} \tag{3.10}$$

Where k is the slope derived from the two starting points x_{n-2}, x_{n-1} and their function outputs f_{n-2}, f_{n-1}:

$$k = \frac{f_{n-1} - f_{n-2}}{x_{n-1} - x_{n-2}} \tag{3.11}$$

Starting with two points x_0, x_1 on the curve, the two 'roots' will continually be updated through iterations based on their secant lines and reach the root while function outputs of these points give zero value or values close to zero (based on the setup of tolerance level). In our research, the target function becomes the left hand side of the equation minus the right hand side, which is a function

of cumulative infiltration:

$$F(I) = \frac{2(K_\mathrm{s} - K_\mathrm{i})^2}{S_0^2} \cdot t(1-\beta) - 2 \cdot \frac{(K_\mathrm{s} - K_\mathrm{i})(I - K_\mathrm{s}t)}{S_0^2} + ln\Big\{\frac{1}{\beta} \cdot exp\left[\frac{2\beta(K_\mathrm{s} - K_\mathrm{i})(I - K_\mathrm{s}t)}{S_0^2}\right] + \frac{\beta - 1}{\beta}\Big\} \tag{3.12}$$

The implementation of the secant method could be written as:

$$I_n = I_{n-1} - F(I_{n-1})\frac{I_{n-1} - I_{n-2}}{F(I_{n-1}) - F(I_{n-2})} = \frac{F(I_{n-1})I_{n-2} - F(I_{n-2})I_{n-1}}{F(I_{n-1}) - F(I_{n-2})} \tag{3.13}$$

Unlike Newton's method requiring an analytical solution to the derivative of the target function: $\frac{\mathrm{d}F(I)}{\mathrm{d}I}$, which also takes more time for calculation, the secant method straightly uses the slope of the two points, contributing to a good computing efficiency. This is the reason why this method was adopted in this research.

3.2 Data preprocess

Data preprocess was done to retrieve reasonable and effective data for future analysis. Regardless of the adequate information provided by the SWIG database, there are still much missing information with respect to soil texture class (about 1,211 samples) and some other parameters like initial moisture content (which is an important input parameter that will affect the simulations). Before drawing information from the data, a careful analysis of the applicable soil samples was conducted to enhance the quality of the data to be estimated. The first thing to process the data is to identify those samples with clear information on texture class. This will make it easier for the later work on summarizing

results for different types of soil. For instance, Sand and loamy sand share the same boundary for hydraulic saturated conductivity according to **Table 1.2**. If MCMC simulations was applied to derive the optimal K_{s} $[\,\mathrm{L}{\cdot}\mathrm{T}^{-1}\,]$ for an unknown sample, which falls into the boundary of 15.24-50.81 [cm/hour], we can hardly determine whether it is a sand soil or loamy sand soil. A total of 646 1-D infiltration samples (see **Fig 3.1**) were filtered out from the 3,842 samples with respect to the measuring instrument. According to the database, a total of 13 instruments (e.g. single ring, double ring, rainfall simulator, etc.) were adopted to measure the cumulative infiltration, where the state of 1-D (vertical) or 3-D infiltration measurement was not determined. To address this issue, soil samples measured through double ring infiltrometer were assumed to have 1-D infiltration. The double ring infiltrometer has inner and outer ring, where a constant head of the water level is often operated in the outer ring to ensure no leakage from the inner ring [Gregory et al. 2005]. Consequently, measurement from the inner ring will yield a vertical infiltration (one-dimensional). Instruments like single ring infiltrometer can overestimate the vertical flow because the infiltration beneath the cylinder will expand to a three-dimensional condition [Bouwer 1986].

The incoherent infiltration data were identified based on the discrepancies of infiltration curves and the differences of the length of data in time and infiltration. **Fig 3.1** provides the 1-D infiltration curves from the SWIG database. There are conspicuous deviations of infiltration curves in some soil samples which share the same soil texture and construction. For instance, given the clay soil samples

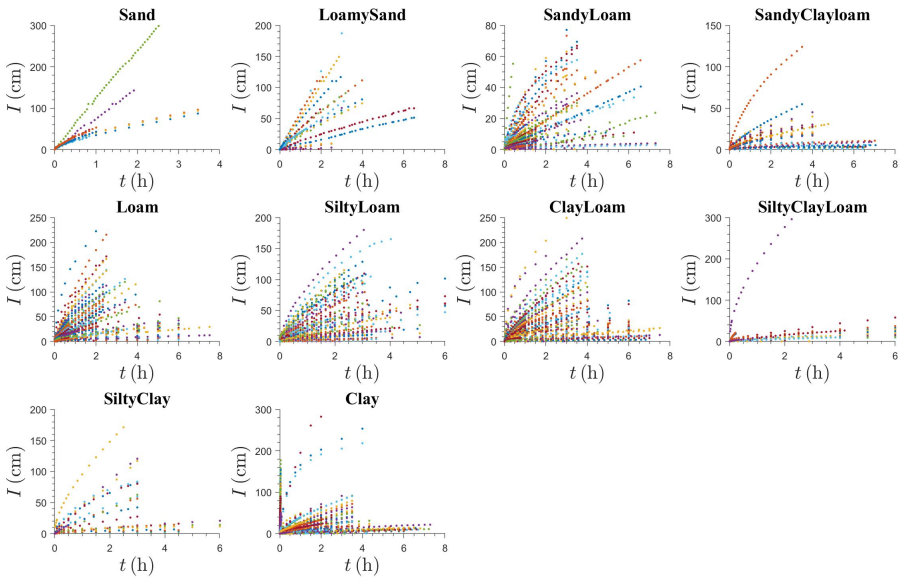

Figure 3.1: 646 samples of 1-D infiltration data from SWIG database

in **Fig 3.1**, it is irrational to have the those infiltration data which approaches 200 cm in a really short time interval. Those data should be viewed as low quality data. One typical reason for having those data is that the actual time data is measured in days rather than hours. To deal with those incoherent data, one method is to plot the infiltration rate versus the time series to check those outliers (unreasonable data will exhibit a clear bias of the infiltration rate compared with others). Another approach is to estimate the parameter K_{s} $[\mathrm{L}{\cdot}\mathrm{T}^{-1}]$ for each sample through MCMC methods. Given the information from **Table 1.1** and the soil texture, the unreasonable samples will be directly identified by checking whether the K_{s} $[\mathrm{L}{\cdot}\mathrm{T}^{-1}]$ is within the boundary for that type of soil. The second approach was valued because the estimated K_{s} $[\mathrm{L}{\cdot}\mathrm{T}^{-1}]$ can

be straightly input into the Philip model, leading to a linear model problem (as K_{s} $[\,\mathrm{L}{\cdot}\mathrm{T}^{-1}\,]$ is known) which can be easily solved through linear regression.

As a consequence of data-preprocessing, we not only conduct experiments on the total 646 one-dimensional infiltration samples, but do further analysis on partial high-quality data of the 646 samples as well.

3.3 DREAM algorithm

The **D**iffe**R**ential **E**volution **A**daptive **M**etropolis (DREAM) [Vrugt 2016] was adopted to derive the posterior distribution of the model parameters and find the optimum. This algorithm is a multi-chain MCMC method which incorporates initial sampling, partial crossover, multi-chain parallel computing, prior distribution, family of likelihood functions, modal jump of the proposals, boundary handling, outlier checking, posterior exploration and sampling, convergence diagnostics, etc [Vrugt, Braak, C.J.F, et al. 2008a]. Substantial advantages has been found using DREAM algorithm such as the flexibility to low and high dimensional models, enforcing detailed balance of Markov Chain, the capability of tuning the orientation of the chains, the efficiency to reach covergence, and so on. The fundamental processes of the DREAM algorithm are generating proposals, computing Metropolis Ratio (accept the proposal or not), searching for the outliers (chains that stuck at a locally optimized region), and adaptively tuning the orientation through adjusting the selection probability of the crossover value and reinitializing the outlier chains [Vrugt, Braak, Gupta, et al.

Table 3.1: Parameter boundary for DREAM analysis

Parameter	S (cm/hour$^{0.5}$)	K_s (cm/hour)	β (-)
Maximum	25.00	100.00	2.00
Minimum	0.00	0.00	0.00

2008b]. The general process of the DREAM algorithm could be shown from the following steps:

(1) Create N chains with initial states (Initial sampling):

$$\mathbf{X}_{(0)} = \{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \mathbf{x}_{(3)}, \dots \mathbf{x}_{(N)}\} \tag{3.14}$$

Where $\mathbf{X}_{(0)}$ is a $N \times p$ matrix, which signifies the initial generation of the multi-chains; x_i is a $1 \times p$ vector, where p is the dimension (number of parameters) of the estimated model. Note that the parameter boundary was set provided by **Table 3.1**.

(2) Starting from the first chain, select $\{r_1, r_2, r_3, \dots, r_6\}$ randomly from $\{1, 2, 3, \dots, N\}$ without replacement as the chosen parent chains;

(3) Generate a total of p values $\{z_1, z_2, z_3, \dots, z_p\}$ from $u \sim (0, 1)$;

(4) Set initial partial crossover value and selection probability:

$$\mathrm{CR} = \{\frac{1}{n_{\mathrm{CR}}}, \frac{2}{n_{\mathrm{CR}}}, \dots, 1\} \tag{3.15}$$

Where default value of n_{CR} is assumed to be 3 which shows good results in low dimensional problems. [Vrugt 2016] This means the maximum crossover value enforces a full crossover. Then initialize the selection probability:

$$p_{\mathrm{CR}} = \{\frac{1}{n_{\mathrm{CR}}}, \frac{1}{n_{\mathrm{CR}}}, \dots, \frac{1}{n_{\mathrm{CR}}}\} \tag{3.16}$$

The equal selection probability means that we have equal initial probability of selecting CR $= 1/n_{\text{CR}}$ or CR $= 2/n_{\text{CR}}$, ... or CR $= 1$. Given the selected crossover value, compare $z_1, z_2, z_3, \ldots, z_p$ with the crossover value, assign those z_i which are smaller than the CR to **set A** and keep those z_j which are greater than CR in **set B**. As a result, **set A** corresponds to elements which will be updated through differential evolution while **set B** corresponds to elements that will remain unmodified.

(5) To start creating proposals, parts of the elements (**set A**) of the parameters undergo crossover. The crossover coefficient is previously assumed: $\gamma(\delta, p^*) = 2.4/\sqrt{2\delta p^*}$ (80% chance, where p^* is the size of **set A** and δ is a presumed value between 1 and 3) or $\gamma(\delta, p^*) = 1$ (20% chance, this helps implement modal jumps). Another scaling part of the parents is: $\boldsymbol{\zeta} = 1 + u_{p^*}(-0.05, 0.05)$. Lastly, an vector of error coefficient $\boldsymbol{\xi}$ is drawn from $N \sim (0, 10^{-6})$ to add to the generated proposal. For the residual elements, keep those elements for the new proposal with no modifications. A general equation of this differential evolution process [Storn and Price 1997] can be written as:

$$\mathrm{d}\mathbf{X}_{\text{A}}^{j} = [\mathbf{1}_{p^*} + \boldsymbol{\zeta}] \cdot \gamma(\delta, p^*) \cdot \left(\sum_{i=1}^{\delta} \mathbf{X}_{(0)}^{r_i} - \sum_{l=\delta+1}^{2\delta} \mathbf{X}_{(0)}^{r_l} \right) + \boldsymbol{\xi} \tag{3.17}$$

$$\mathrm{d}\mathbf{X}_{\text{B}}^{j} = 0 \tag{3.18}$$

Where $\mathrm{d}\mathbf{X}_{\text{A}}^{j}$ is the jump distance vector for the jth chain with a size of $1 \times p^*$, $\mathbf{X}_{(0)}^{j}$ is the jth initial (parent) chain. The proposal now could be written:

$$\mathrm{d}\mathbf{X}_{\text{p}}^{j} = \mathbf{X}_{(0)}^{j} + \mathrm{d}\mathbf{X}_{\text{A}}^{j} \tag{3.19}$$

(6) Enforce the boundary handling. It is possible that that outbound proposals are created. Outbound proposals are not acceptable and may slow the chains' speed to converge. To deal with this problem, there are several boundary handling methods. The 'reflection' method was adopted in this research. To handle the proposals that exceed the parameter's upper boundary:

$$X_{\mathrm{p}} = 2X_{\mathrm{p}}^{\mathrm{up}} - X_{\mathrm{p}} \tag{3.20}$$

For those proposals which are smaller than the parameter's lower boundary:

$$X_{\mathrm{p}} = 2X_{\mathrm{p}}^{\mathrm{low}} - X_{\mathrm{p}} \tag{3.21}$$

(7) Compute the acceptance rate using the log-likelihood function (Gaussian likelihood function where measurement error integrated out):

$$\begin{aligned}\log(L) = -\frac{1}{2}n\log(2\pi) + \frac{1}{2}n\log(n-1) - \\ \frac{1}{2}(n-1) - \frac{1}{2}n\log\left(\sum_{i=1}^{n}(I_i - \tilde{I}_i)^2\right)\end{aligned} \tag{3.22}$$

Where n is the size of data; I_i is the ith data and $\hat{I}_i$ is the ith simulation. For analyzing a time series of data, the size of the data is constant. This leads to the log-likelihood as a function of sum of squared residuals [Eliason 1993]:

$$\log\left(L(x|\tilde{I})\right) = -\frac{1}{2}n\log\left(\sum_{i=1}^{n}(I_i - \tilde{I}_i)^2\right) + C \Rightarrow \log(L) \propto -\frac{1}{2}n\log\left(\sum_{i=1}^{n}(I_i - \tilde{I}_i)^2\right) \tag{3.23}$$

Where C is a normalized constant which is a function of the data size (how many data points). The probability to accept proposal (or Metropolis Ratio) [Hastings 1970] could also be written as:

$$p_{\mathrm{acc}} = \alpha(x_{\mathrm{p}}|x_i) = \frac{P(x_{\mathrm{p}}|\tilde{I})}{P(x_i|\tilde{I})} = \frac{P(x_{\mathrm{p}})L(x_{\mathrm{p}}|\tilde{I})}{P(x_i)L(x_i|\tilde{I})} \tag{3.24}$$

Since the conditional probability density of $P(x_{\mathrm{p}}|\tilde{I})$ is extremely small (due to large size of data), log-density (or log-likelihood) is commonly applied to derive the probability density of the hypothesis. In this case, we can rewrite the Metropolis Ratio:

$$p_{\mathrm{acc}} = \alpha(x_{\mathrm{p}}|x_i) = \frac{\exp\{\log[P(x_{\mathrm{p}})L(x_{\mathrm{p}}|\tilde{Y})]\}}{\exp\{\log[P(x_i)L(x_i|\tilde{Y})]\}} = \frac{\exp[F(x_{\mathrm{p}})]}{\exp[F(x_{\mathrm{p}})]}$$
$$= \exp[F(x_{\mathrm{p}}) - F(x_i)] \tag{3.25}$$

Where $F(x_{\mathrm{p}})$ is the log-Density (log-prior + log-likelihood) of the proposal, $F(x_i)$ is the log-Density (log-prior + log-likelihood) of the current parent chain; Then, we can determine whether accept the proposals ($u \sim (0,1) \leq \alpha(x_{\mathrm{p}}|x_i)$) or reject ($u \sim (0,1) \geq \alpha(x_{\mathrm{p}}|x_i)$).

(8) Given the new generation, update the jumping distance of this partial crossover and update selection probability. The jumping distance is a $1 \times p$ vector where each element corresponds to the jumping distance for that specific crossover value. Taking the first chain $\mathbf{x}_1$ for example (assuming the crossover value is 1/3), we now have the crossover part of $\mathbf{X}_1$ which can be written as $\mathrm{d}\mathbf{X}^1_{(w)}$. w signifies the generation, since we are now creating the first generation, $w = 1$. $\mathrm{d}\mathbf{X}^1_{(w)}$ is a $1 \times p$ vector (assuming the element is $\mathrm{d}x_i$, $i = 1, 2, \ldots, p$). The jumping distance of the selected crossover value for chain $\mathbf{x}_1$ could be updated by:

$$J_{\mathrm{d,CR=1/3}} = J_{\mathrm{d,CR=1/3}} + \sum_{i}^{N^*} \left(\frac{\mathrm{d}x_i}{\sigma_i}\right)^2 \tag{3.26}$$

Where σ_i is the standard deviation of the ith parameter at the $(w-1)$th generation. $\mathrm{d}x_i$ is the ith element of the crossover part of chain $\mathbf{x}_1$. N^* is the number

of times that the crossover value (e.g. $CR = 1/3$) has been used. As the iteration proceeds, the jumping distance of each crossover value will be updated and increase theoretically.

To enforce better efficiency of the algorithm, the update of selection probability is one effective approach. Since the choosing of different crossover value will result in different jumping distance of the proposals, it is essential to weight the probability of selecting each crossover value. Given the updated jump distance, the selection probability for the crossover value will be (take $n_{\mathrm{CR}} = 3$ for example):

$$p_{\mathrm{CR}=1/3} = \frac{Q_1}{Q_1 + Q_2 + Q_3} \tag{3.27}$$

$$p_{\mathrm{CR}=2/3} = \frac{Q_2}{Q_1 + Q_2 + Q_3} \tag{3.28}$$

$$p_{\mathrm{CR}=1} = \frac{Q_3}{Q_1 + Q_2 + Q_3} \tag{3.29}$$

Where:

$$Q_1 = \frac{J_{\mathrm{d,CR}=1/3}}{T_{\mathrm{d,CR}=1/3}},\ Q_2 = \frac{J_{\mathrm{d,CR}=2/3}}{T_{\mathrm{d,CR}=2/3}},\ Q_1 = \frac{J_{\mathrm{d,CR}=1}}{T_{\mathrm{d,CR}=1}} \tag{3.30}$$

$T_{\mathrm{CR=k}}$ represents how many times that the crossover value k has been used to generate proposals. Therefore, after each time a proposal of the chain is generated, the selection probability of the crossover value will be updated.

(8) go to step (2) again and form the loop. End the iteration when N proposal chains has been created. (9) Check for the outlier chains and enforce reinitialization. For models with complex structure and high dimensions, some of the chains may converge to a local optimum (local maximum likelihood) as the iteration proceeds. The jump distance may be not long enough for the chain to

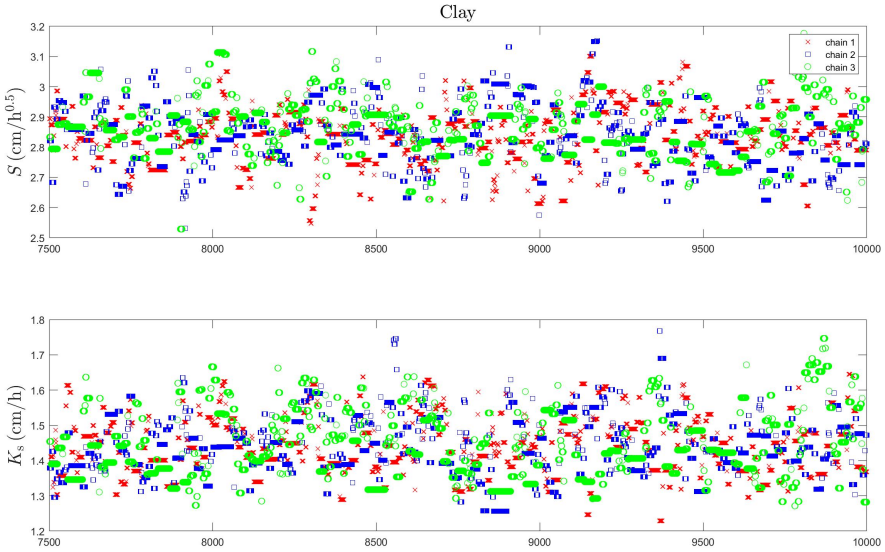

Figure 3.2: Performance of Markov Chain of the last 25% samples

reach a new space with higher likelihood. To find those outlier chains, the mean log-density of the second half of the N chains need to be calculated and compared. Those chains which perform badly will exhibit small mean log-density. Those outliers will be patched through reinitialization.

(10) Keep on the iteration from step (2) to step (9). After thousands of generations are created, the multi-chain samples will reach convergence. The last 25% of the posterior samples will be stored and be used to plot the posterior distribution of each parameter. Optimum parameters will also be explored through the posterior samples.

After running the DREAM algorithm, the posterior distribution and optimum of each parameter can be derived for each sample. The optimum corresponds

the parameter values which give the best model simulation using Gaussian likelihood as the criterion. Taking the posterior, the statistics (e.g. mean value, standard deviation, and 95% interval of the posterior samples) of posterior samples can be analyzed. **Fig 3.2** shows the performance of Markov Chain (taking the last 25% posterior samples) with respect to two hydraulic parameters of clay soil. Those chains are considered as converged according to the diagnostic method of R-statistic. [Gelman and Rubin 1992].

3.4 Linear regression and Bayesian Information Criterion

The parameter estimation using DREAM algorithm provided the optimal S [cm·hour$^{-1/2}$] and K_{s} [cm·hour^{-1}] value for each sample. Through taking the sorptivity and saturated hydraulic conductivity as known values, the Philip's two-term model reduced its dimension from 3 (S, K_{s}, and c unknown) to 1 (c unknown). The model simulation using Philip's two-term could be written as (for a sample having n data):

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix} = S \cdot \begin{bmatrix} \sqrt{t_1} \\ \sqrt{t_2} \\ \vdots \\ \sqrt{t_n} \end{bmatrix} + K_{\mathrm{s}} \cdot c \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \tag{3.31}$$

Where $I_1, \ldots, I_n$ [cm] are the simulated infiltration. Given the S [cm·hour$^{-1/2}$] and K_{s} [cm·hour^{-1}] as known input parameters, the linear regression method was adopted to derive the optimal coefficient c when estimating a sample with n data. By taking the derivative of the sum of squared residuals (SSR) with

respect to coefficient c, we derived the following equation to determine the optimal coefficient c where the derivative of SSR goes to zero:

$$c_{\text{opt}} = (\mathbf{X}' \times \mathbf{X})^{-1} \times \mathbf{X} \times \mathbf{Y} \tag{3.32}$$

Where $\mathbf{X}$ is the Jacobian Matrix and $\mathbf{Y}$ is the a function of the measured data, which are given by:

$$\mathbf{X} = K_{\text{s}} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \tilde{I}_1 \\ \tilde{I}_2 \\ \vdots \\ \tilde{I}_n \end{bmatrix} - S \cdot \begin{bmatrix} \sqrt{t_1} \\ \sqrt{t_2} \\ \vdots \\ \sqrt{t_n} \end{bmatrix} \tag{3.33}$$

Where $\tilde{I}_1, \ldots, \tilde{I}_n$ [cm] signifies the measurement data. The time validity of Philip's two-term model represents the time when best model efficiency presents (closest simulations to the data). In this research, the Bayesian Information Criterion (BIC) was chosen as the criterion for model efficiency:

$$BIC = -2\log\left(L(c|\tilde{I})\right) + k \cdot \log(n) \tag{3.34}$$

Where $L(c|\tilde{I})$ is the likelihood derived from the coefficient c given the data $\tilde{I}$, which is a function of the SSR. k is the dimension of the model which is equal to 1 because we are only estimating the coefficient c. n is the number of data points. To find the time validity of a sample, several measured times were tested by estimating the optimal coefficient c and calculating the Bayesian Information Criterion (BIC) [Schwarz 1978] each time. The general steps to estimate time validity are as follows:

(1) Start from the first sample, input the corresponding optimal K_{s} [cm·hour^{-1}]

and S [cm·hour$^{-1/2}$] to the Philip's two-term model.

(2) Extract the data, start with analyzing i=3 data points.

(3) Apply linear regression to determine the optimal value of coefficient c with respect to the extracted data.

(4) Use the optimal coefficient c to calculate the sum of squared residuals (SSR). Compute the likelihood and BIC. Store the result of BIC.

(5) $j = j + 1$, go back to step (2), form the iteration and end the loop when $j = n + 1$. This means all the data points (time) have been tested.

(6) Take the time which corresponds to the minimum of BIC as the t_{valid} of that sample.

(7) Go back to step (1) and test another sample. Terminate the loop when all the samples have been analyzed.

Unlike the conventional way to use BIC as an approach to balance the model accuracy and model complexity, we used the BIC as a function of the number of data points and determine the optimal number of data points and its corresponding time as t_{valid}. One thing to note is that the determination of t_{valid} is limited to some factors, for example, time resolution, measurement errors. This means many of the low-quality data might not be feasible to be analyzed.

4 Results

4.1 Entire samples

The entire 646 samples were analyzed firstly to find if there were any patterns shown by the whole dataset. This means we derived estimated optimal K_s [cm·hour^{-1}] and S [cm·hour$^{-1/2}$], coefficient c and time validity for each sample. **Table 4.1** and **Table 4.2** give the statistics of optimal parameters based on the 646 samples. Given the statistics, it can be found that many samples give outbound or irrational estimation of the hydraulic parameters along with considerable uncertainties, compared with parameter boundary provided by **Table 1.2**. **Fig 3.1** could probably explain this phenomenon that infiltration curves show conspicuous divergence for each type of soil. Many existent sharp cumulative infiltration curves are contradictory to the corresponding soil properties, which should be considered as low-quality data or outliers.

Looking at the output of t_{valid}, all the soils show a relatively small mean value and a great uncertainty. **Fig 4.1** provide the soil texture triangle of t_{valid} (mean value) derived from each sample, where t_{valid} seems to cluster at the early start of the infiltration for each soil type, which is theoretically not acceptable. Like previously discussed, a large proportion of the 646 samples exhibit low quality with respect to their curves. In addition, the determination of t_{valid} is affected by the resolution of time, where accurate estimation requires short measurement time intervals. Many of the samples were measured around every 0.5 hour, which leads to huge uncertainty of the t_{valid}. The measurement error also has

Table 4.1: Mean and standard deviation of the optimal parameters of the entire samples

Texture	S (cm/hour$^{0.5}$)		K_s (cm/hour)	
	mean	σ	mean	σ
Sand	9.7478	0.4666	49.0052	24.9594
Loamy Sand	6.0253	3.5429	23.0827	18.5525
Sandy Loam	5.3194	3.3629	6.7734	12.3299
Sandy Clay Loam	7.0245	3.4027	3.4787	6.9935
Loam	8.0303	2.8802	18.8525	23.3504
Silty Loam	7.4138	3.4363	14.7104	14.7544
Clay Loam	7.3974	3.0825	11.2082	17.6753
Silty Clay Loam	6.0539	3.0123	13.3343	30.6414
Silty Clay	5.8011	3.4552	14.1768	18.6325
Clay	6.4194	3.9048	31.4140	41.6185

Table 4.2: Mean and standard deviation of the t_{valid} [hour]of the entire samples

Texture	t_{valid} (hour)	
	Mean	σ
Sand	0.1314	0.0745
Loamy Sand	0.7704	1.0386
Sandy Loam	0.9922	1.5033
Sandy Clay Loam	0.7981	1.7652
Loam	0.3427	1.0872
Silty Loam	0.6480	0.9825
Clay Loam	0.8437	1.7074
Silty Clay Loam	0.7130	1.3708
Silty Clay	1.0250	1.7035
Clay	0.8294	1.5607

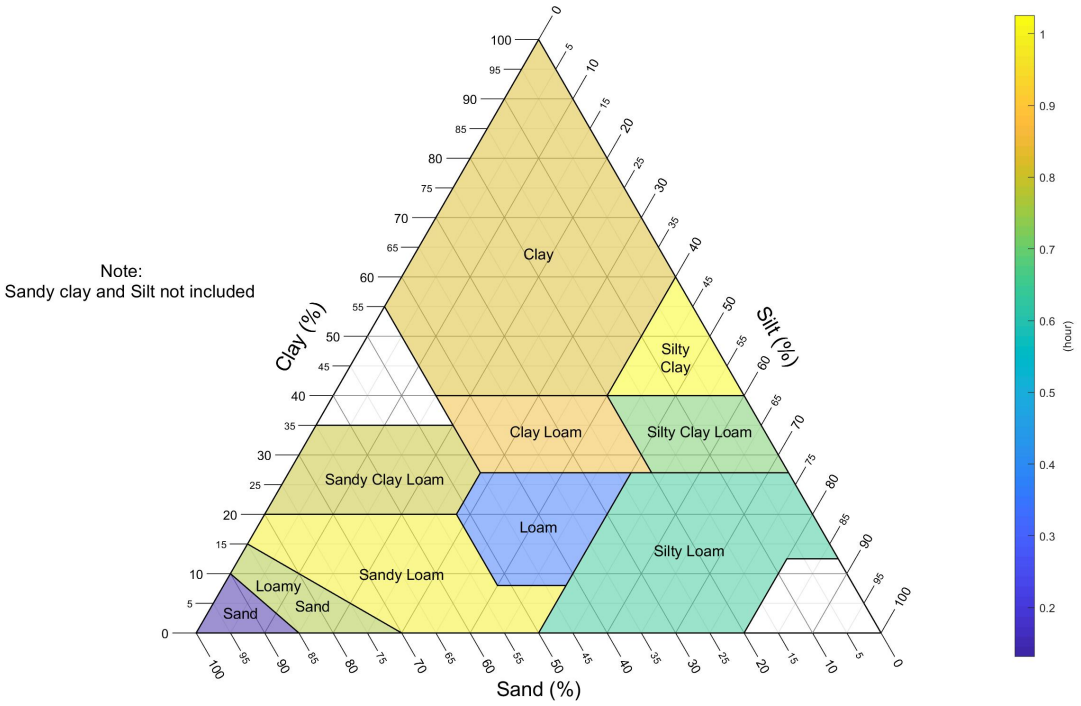

Figure 4.1: Soil texture triangle of t_{valid} derived from 646 samples

impact on the analysis by affecting the results of linear regression. A typical measurement error can be illustrated in **Fig 4.2** where cumulative infiltration tends to decrease in the late time. To balance some disparate data points, the linear regression method would derive some optimal coefficient c which might not be reasonable (e.g. extremely positive or negative value), leading to the effect on calculating the BIC, then coming into the conclusion of the literally small t_{valid} at the start of the infiltration.

Based on the upper findings, we concluded that many samples are not applicable

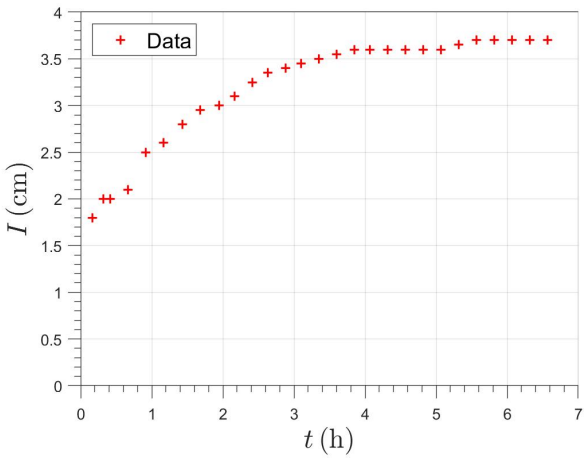

Figure 4.2: Typical measurement error issue

to the time validity and hydraulic parameters analysis due to the low-quality data (e.g. irrationally sharp curves, low resolution of measured time, apparent measurement errors). The great uncertainty of each parameter can hardly enable us to explore any patterns for each soil type. Therefore, as the next steps, it is essential to isolate some good-quality data with respect to the following criteria: (1) The cumulative infiltration is reasonable where estimation of the hydraulic parameters are acceptable. (2) The measurement points should be as many as possible and the time intervals are small enough for more accurate estimation of t_{valid}. (3) The measurement data exhibit relatively small measurement errors where infiltration curves are smoother.

4.2 Best sample from each soil type

10 samples (see **Fig 4.3**) were isolated from the 646 samples as the best sample from each soil type. Those samples were selected based on the criteria discussed above. From the plots of the data, each sample is measured under a relatively long time with feasible measured cumulative infiltration. Although there are still inevitable measurement errors and the measured times are not quite of high resolution, these data show superior quality compared with other samples in the dataset. Through analyzing those best samples, the DREAM algorithm can produce better estimation of the optimum which gives more accurate simulated curves and smaller sum of squared residuals. A higher resolution of measured time enables us to explore more accurate t_{valid} with smaller uncertainty.

The DREAM algorithm estimates the optimal parameters for each soil type, along with the posterior distribution, which can be used to analyze the posterior statistics of each parameter. **Table 4.3** provides the DREAM output of the optimal parameters. **Table 4.4** covers the statistics of DREAM posterior. **Fig 4.4** and **Fig 4.5** illustrate the posterior distribution of S [cm·hour$^{-1/2}$] and K_{s} [cm·hour^{-1}]. From these tables and figures, it can generally concluded that the Markov Chains generated using DREAM algorithm have effectively converged, where the posterior samples converged to the specific range, given the prior distribution of each parameter. Given the optimal parameters, the simulated curves using Haverkamp model were plotted in **Fig 4.6**, which are satisfying to show the 'true' values of the parameters derived from DREAM. This shows the excellent analysis of DREAM algorithm. By taking those optimum as the input

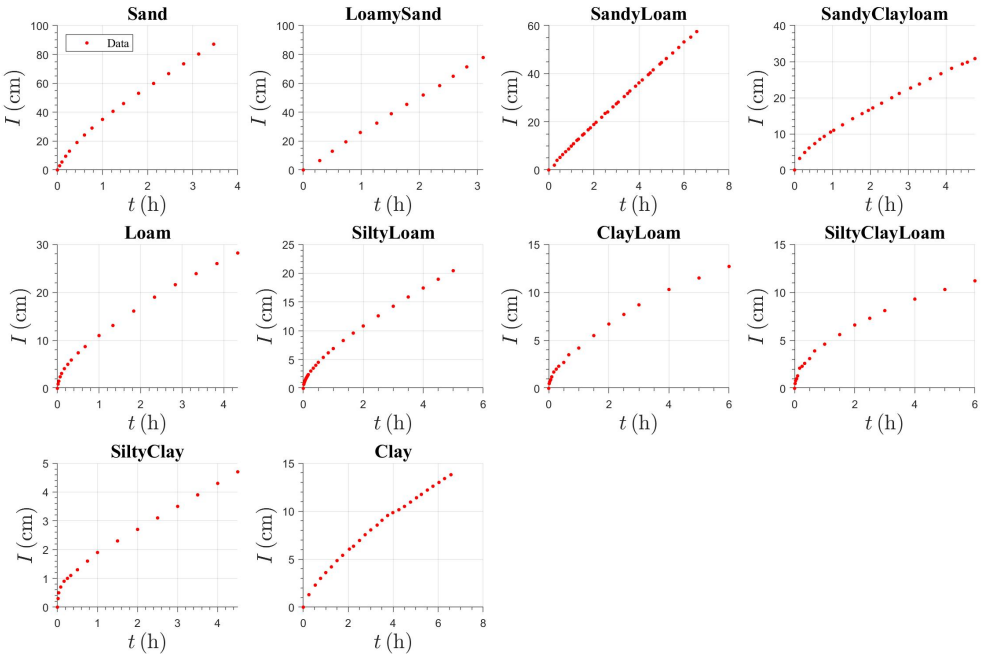

Figure 4.3: Best sample from each soil type

Table 4.3: Optimal parameter estimation from DREAM

Texture	S (cm/hour$^{0.5}$)	K_s (cm/hour)	β (-)
Sand	20.6845	18.2727	0.0350
Loamy Sand	2.7939	24.8784	0.0122
Sandy Loam	2.9685	8.4559	0.0030
Sandy Clay Loam	8.1994	3.7050	0.0326
Loam	8.9526	3.1824	0.0742
Silty Loam	5.2363	2.3422	0.0092
Clay Loam	3.8281	0.8477	0.0156
Silty Clay Loam	4.6060	0.0010	0.9988
Silty Clay	1.7993	0.6610	1.9949
Clay	2.8217	1.3934	0.0794

Table 4.4: DREAM posterior statistics of S [cm·hour$^{-1/2}$] and K_s [cm·hour^{-1}]

Texture	S (cm/hour$^{0.5}$)				K_s (cm/hour)			
	2.5%	Mean	Median	97.5%	2.5%	Mean	Median	97.5%
Sand	18.9966	21.0636	20.9378	23.8725	17.1944	19.3937	19.3736	21.6321
Loamy Sand	1.7941	5.4593	5.3476	9.5585	24.4609	24.9383	24.9635	25.2752
Sandy Loam	2.7461	2.9924	2.9780	3.2789	8.3973	8.4912	8.4907	8.5871
Sandy Clay Loam	8.1254	8.2379	8.2339	8.3775	3.6546	3.9402	3.9160	4.3480
Loam	8.7841	9.0221	9.0230	9.3325	3.0413	3.5337	3.5090	4.1436
Silty Loam	5.2102	5.2428	5.2420	5.2814	2.3283	2.4210	2.4076	2.5813
Clay Loam	3.6133	3.8763	3.8698	4.1792	0.7442	1.0431	1.0383	1.4222
Silty Clay Loam	4.4342	4.5634	4.5704	4.6528	0.0042	0.1403	0.1013	0.4519
Silty Clay	1.6226	1.7705	1.7779	1.8975	0.2310	0.5678	0.6093	0.7376
Clay	2.6580	2.8317	2.8257	3.0164	1.3085	1.4466	1.4355	1.6309

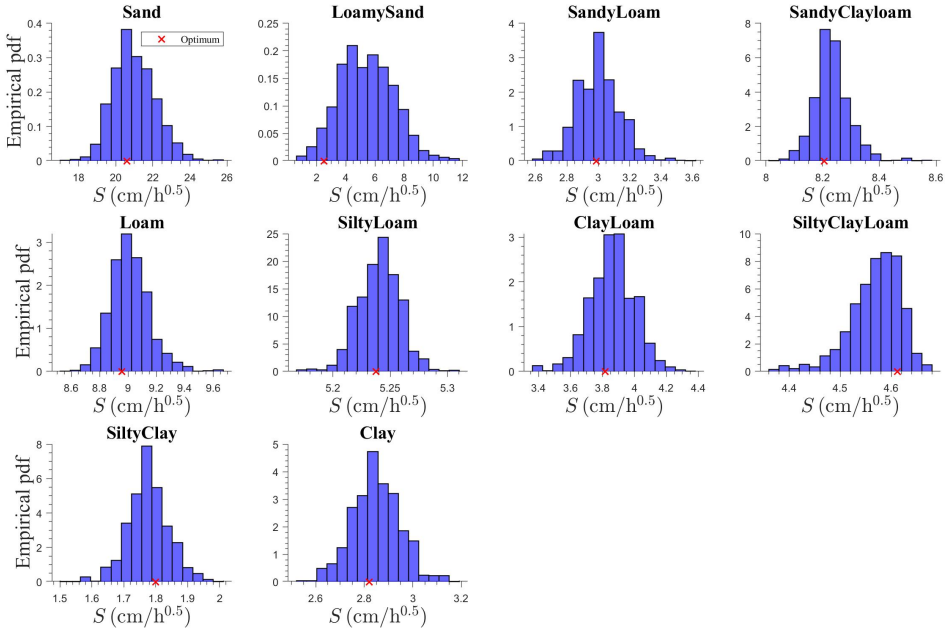

Figure 4.4: Posterior distribution of S [cm·hour$^{-1/2}$] of each sample

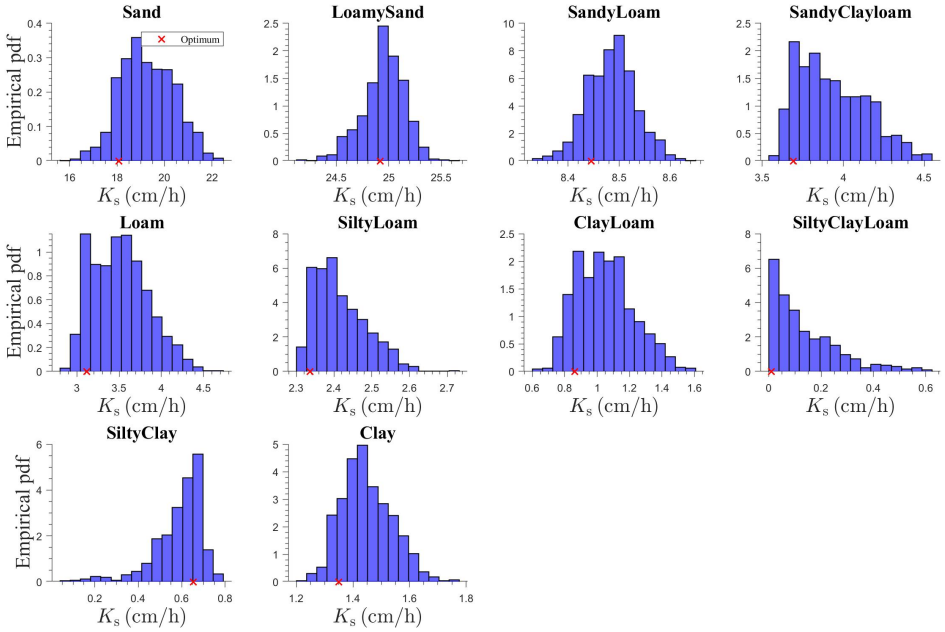

Figure 4.5: Posterior distribution of K_s [cm·hour^{-1}] of each sample

parameters to the Philip's two-term model, the optimal coefficient c and t_{valid} were determined through the methodology described in section 3.4. To explore the uncertainty of c and t_{valid}, 500 DREAM posterior samples were randomly selected from the last 25% of the generations. Those posterior samples of S and K_s were taken as the model input of Philip's two-term. The output of optimal coefficient c and t_{valid} of each soil type is provided in **Table 4.5**. **Fig 4.7** shows the soil texture triangle of the estimated t_{valid}. In addition to the results from SWIG database, The results derived from analyzing HYDRUS-1D simulated data [Jaiswal 2019] are also provided to compare results derived from different datasets. The t_{char} was covered as well, which is the time validity calculated

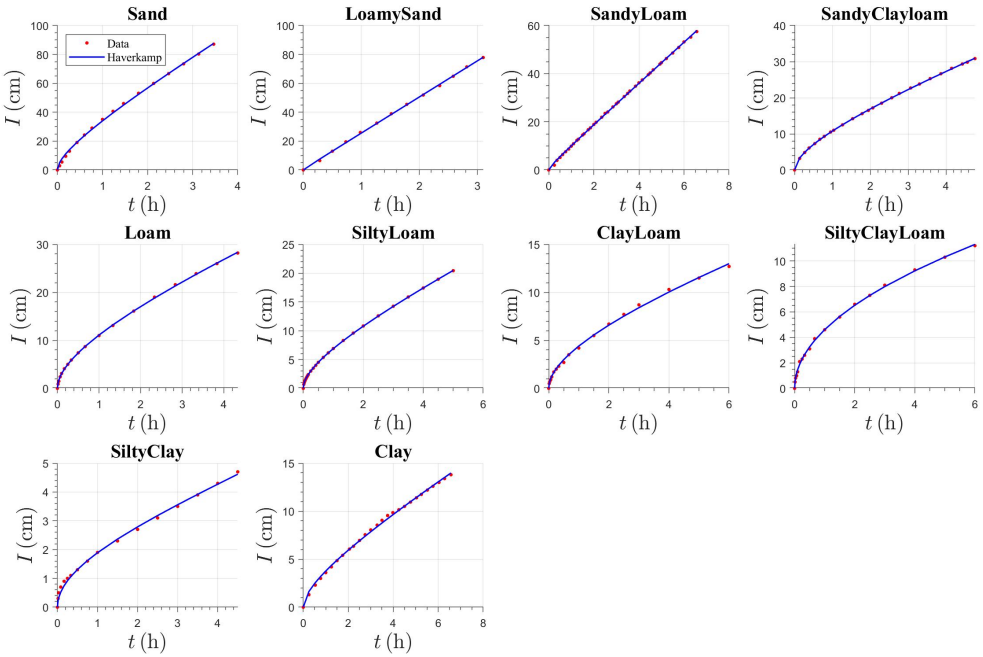

Figure 4.6: Haverkamp Simulated curves using the DREAM optimum

and proposed by Philip. The results from SWIG database and HYDRUS-1D share relatively close estimation of t_{valid} for some coarse soils and fine soils (e.g. sand, sandy loam, silty loam, clay). Comparing the results from SWIG database and Philip, there are close estimation of t_{valid} for coarse and medium soil (e.g. sand, loamy sand, sandy clay loam, loam). **Table 4.7** shows the uncertainty of t_{valid} given by the 95% range. Based on the output DREAM posterior samples, 500 samples of each parameter were randomly selected from each posterior distribution to derive the estimation of t_{valid}. By sorting the 500 outputs of the t_{valid}, we calculated the 95% interval of the t_{valid}. Many of the

Table 4.5: Estimated t_{valid} derived from the best samples

Texture	t_{valid}(SWIG) (hour)	t_{valid}(HYDRUS-1D) (hour)	t_{char}(Philip) (hour)	σ(SWIG)
Sand	0.1833	0.1159	0.0962	0.0000
Loamy Sand	0.7333	0.2453	0.1805	0.0000
Sandy Loam	0.5000	0.6383	0.7505	0.0000
Sandy Clay Loam	2.3000	1.1552	1.4918	0.2697
Loam	4.3333	1.995	4.4343	1.3418
Silty Loam	5.0000	4.1833	13.4444	2.2937
Clay Loam	0.5000	4.7842	31.1021	1.1905
Silty Clay Loam	6.0000	43.0141	51.1837	0.6209
Silty Clay	3.5000	30.6717	306.2500	0.4338
Clay	6.5667	6.2032	26.0100	0.0000

uncertainty output of coarse soil are approaching or equal to zero mainly due to the low resolution of the measured time. For example, for the best sample of loamy sand, there are only four data points before t goes to 1 hour.

Those evident differences between the results for other soils can be attributed to two aspects. Unlike the HYDRUS-1D data which were simulated with 100 data points and high resolution of time, the SWIG database provides real-world measured data with low resolution (even best samples were selected), which affect the estimation of time validity and lead to poor results of the uncertainty. Besides, due to the short period of measured time, the data provided by SWIG could hardly be applied to estimate the true t_{valid} of some fine and very fine soils, where t_{valid} goes to the maximum of the measured time $t_{\max}$. Apparently, the issue that the measured length of time is too short plays a limiting factor on exploring the t_{valid}. To address the issue of low resolution, we conducted our next step analysis by interpolating data points to the original data.

Table 4.6: Uncertainty of t_{valid} from best sample of each soil type

Texture	t_{valid} (hour)			
	2.5%	Mean	Median	97.5%
Sand	0.1833	0.1833	0.1833	0.1833
Loamy Sand	0.7333	0.7333	0.7333	0.7333
Sandy Loam	0.5000	0.5000	0.5000	0.5000
Sandy Clay Loam	2.3000	2.3694	2.3000	4.7667
Loam	1.3333	2.6909	1.8333	4.3333
Silty Loam	0.4167	2.4723	0.4167	5.0000
Clay Loam	0.3333	0.9705	0.5000	4.0000
Silty Clay Loam	4.0000	5.7690	6.0000	6.0000
Silty Clay	3.0000	3.7008	3.5000	4.5000
Clay	6.5667	6.5464	6.5667	6.5667

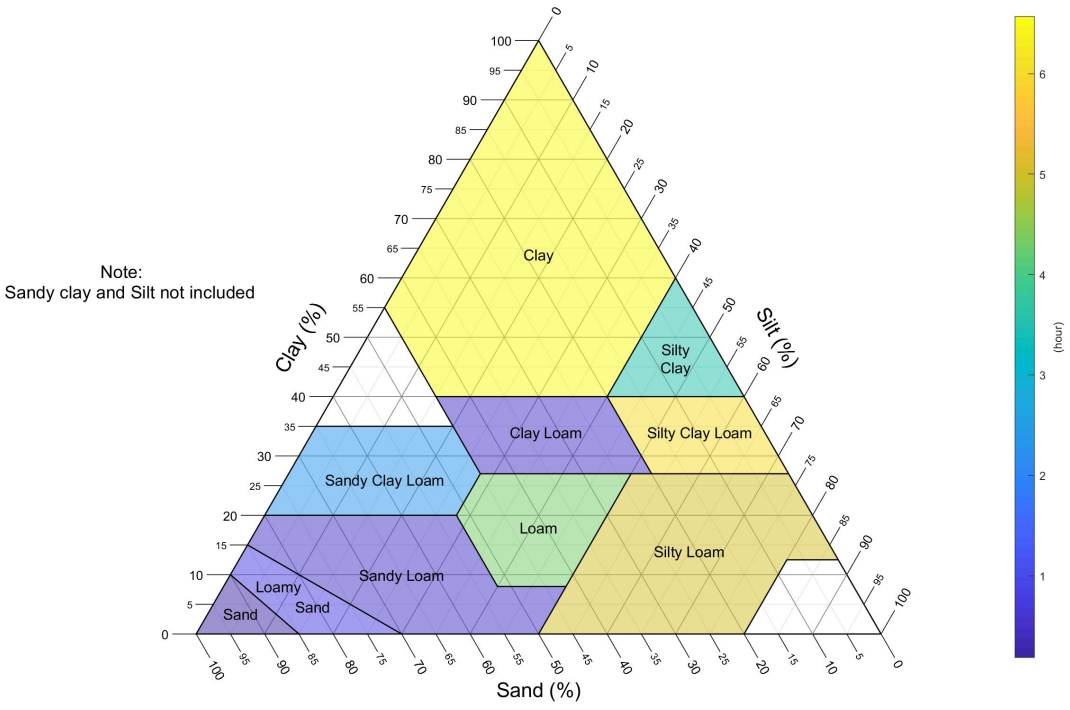

Figure 4.7: Soil texture triangle of t_{valid} derived from the best samples

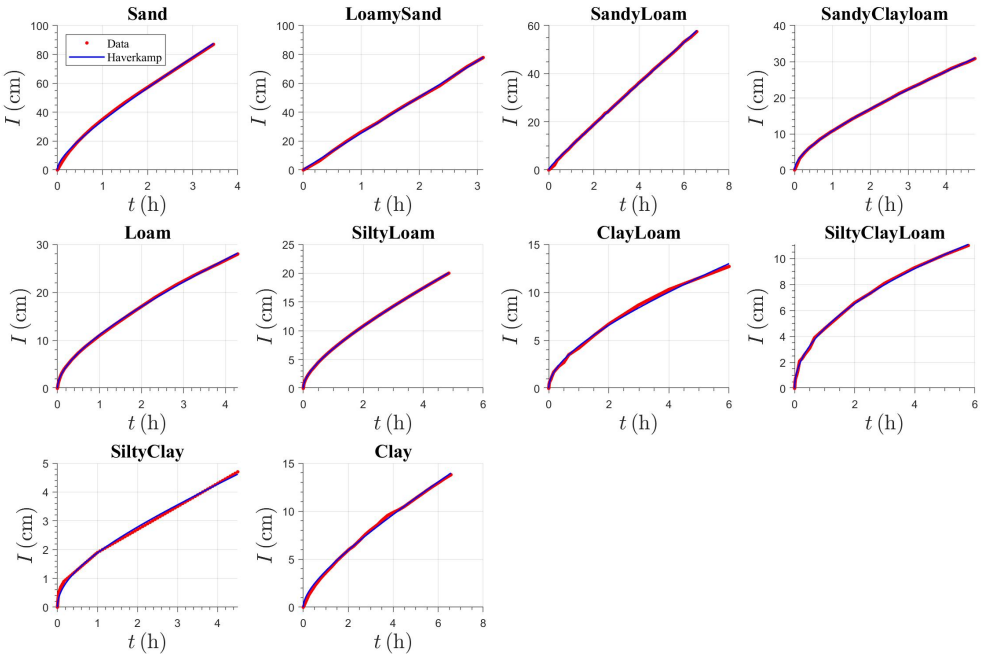

Figure 4.8: Best samples with interpolated data

4.3 Best samples with interpolated times

Given the results derived from last section, low resolution of measured times is a major issue that should be addressed. Thus, we again conducted the analysis discussed before but with interpolated measured times in this section. To manage the interpolation and ensure high resolution of time, the new infiltration time series were created by setting the cumulative infiltration as 0 cm to the maximum $I_{\max}$, where the interval is defined as 0.05 cm. For instance, the $I_{\max}$ of the best sand sample is 87.1 cm. The cumulative infiltration for this sample will be initially set to: 0.00 cm, 0.05 cm, 0.10 cm, ..., 87.1 cm. Based on the

Table 4.7: Estimated t_{valid} derived from best samples with interpolated temporal data

Texture	t_{valid}(SWIG) (hour)	t_{valid}(HYDRUS-1D) (hour)	t_{valid}(Philip) (hour)	σ(SWIG)
Sand	0.0451	0.1159	0.0962	0.0000
Loamy Sand	0.1760	0.2453	0.1805	0.0651
Sandy Loam	0.1125	0.6383	0.7505	0.0114
Sandy Clay Loam	4.7667	1.1552	1.4918	2.0552
Loam	4.3333	1.995	4.4343	1.2384
Silty Loam	4.8529	4.1833	13.4444	0.0000
Clay Loam	4.8358	4.7842	31.1021	1.7094
Silty Clay Loam	6.0000	43.0141	51.1837	0.3141
Silty Clay	3.5013	30.6717	306.2500	0.4238
Clay	6.5667	6.2032	26.0100	1.9902

original data and the new infiltration series, we did the linear interpolation of the new temporal data. Consequently, interpolation were conducted for each best sample, where the new interpolated data are shown in **Fig 4.8**. The new time series of infiltration of each sample is of much higher resolution compared with the original data. The simulated curves are also included in **Fig 4.8** using the optimal parameters derived from Section 4.2 because it is reasonable to assume that the linear interpolation of data will not make great different to the estimated optimal parameters S [cm·hour$^{-1/2}$] and K_{s} [L·T^{-1}]. Based on the new interpolated data, the experiment of time validity analysis was tested again. Given the optimal parameters, the output of t_{valid} [hour] was covered in **Table 4.7**, along with a soil triangle of t_{valid} [hour] in **Fig 4.9**.

Looking at the result derived from interpolated data, the new estimated t_{valid} [hour] exhibits proximation of the output from HYDRUS-1D or Philip with

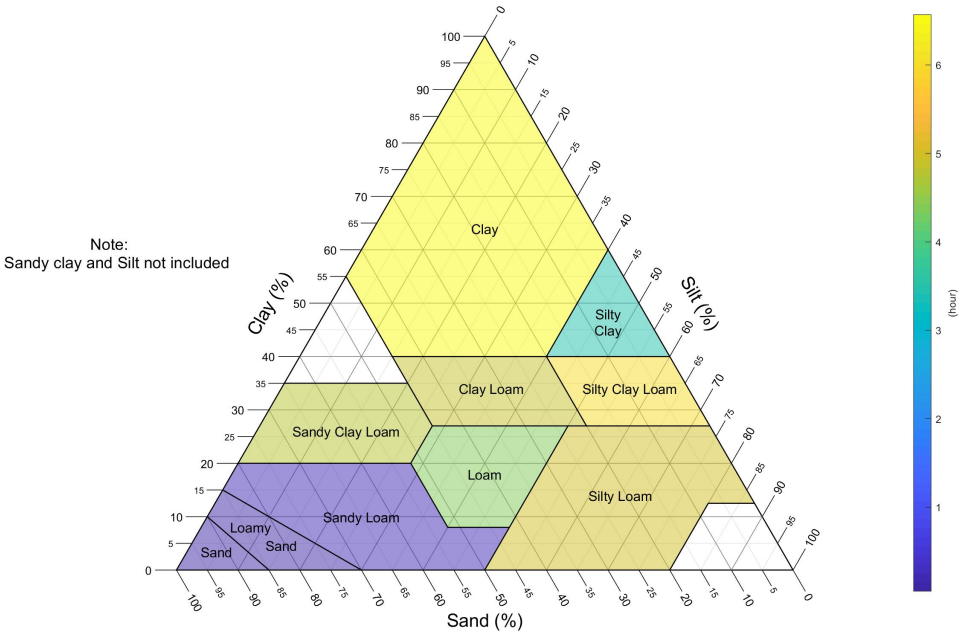

Figure 4.9: Soil texture triangle of t_{valid} [hour] derived from new interpolated data

Table 4.8: Uncertainty of t_{valid} from best samples with interpolated temporal data

Texture	t_{valid} (hour)			
	2.5%	Mean	Median	97.5%
Sand	0.0451	0.0451	0.0451	0.0451
Loamy Sand	0.0066	0.0611	0.0349	0.2854
Sandy Loam	0.0938	0.1139	0.1125	0.1375
Sandy Clay Loam	0.0125	3.4450	4.7667	4.7667
Loam	1.6167	3.0458	3.7917	4.2879
Silty Loam	4.8549	4.8549	4.8549	4.8549
Clay Loam	1.2692	3.0699	4.2083	5.0833
Silty Clay Loam	4.8500	5.6528	5.7778	5.7778
Silty Clay	2.9375	3.7059	3.6250	4.5000
Clay	0.1058	5.8561	6.5667	6.5667

Table 4.9: Estimated coefficient c corresponding to the t_{valid} [hour]

Texture	c (-)
Sand	-0.5260
Loamy Sand	0.6378
Sandy Loam	-0.3052
Sandy Clay Loam	0.7405
Loam	0.7220
Silty Loam	0.7394
Clay Loam	0.6815
Silty Clay Loam	1.7295
Silty Clay	0.1610
Clay	0.6540

respect to sand, loamy sand, loam, silty loam, clay loam, and clay. By randomly selecting 500 samples fron the DREAM posterior samples as input to the Philip's two-term, the uncertainty of t_{valid} [hour] (**Table 4.8**) and coefficient c (**Table 4.10**) were derived. The 95% intervals of t_{valid} [hour] derived from SWIG database include most of the results from HYDRUS-1D and Philip, except for soils like clay loam, silty clay loam and silty clay. The major reason why the uncertainty interval could not cover results of those soils is that the best samples of SWIG do not have measured times of 8 hours or longer. Regardless of the limitation of short measured periods, the uncertainty output indicates that using SWIG could make estimation of the t_{valid} [hour] but with large uncertainty, which is mainly attributed to the measurement error and minorly model structural error of the Philip model. Like previous discussed, HYDRUS-1D simulates synthetic data with no measurement error, where the uncertainty of the estimated t_{valid} [hour] is only subjected to the time intervals. Thus, due to the

Table 4.10: Uncertainty of the estimated coefficient c

Texture	c (-)			
	2.5%	Mean	Median	97.5%
Sand	-1.2818	-0.5272	-0.5242	-0.4459
Loamy Sand	-4.4547	-0.7252	-0.4456	0.7805
Sandy Loam	-0.5237	-0.3003	-0.3005	-0.0799
Sandy Clay Loam	-16.9183	-3.9026	0.6687	0.7416
Loam	0.4390	0.5898	0.5907	0.7294
Silty Loam	0.6718	0.7159	0.7184	0.7435
Clay Loam	0.0365	0.4557	0.4741	0.9528
Silty Clay Loam	-0.3804	4.9716	0.2042	10.1740
Silty Clay	0.0377	0.2702	0.2434	0.6654
Clay	-4.1509	0.1972	0.6950	0.7961

measurement error, the use of real-world data (SWIG) to explore time validity of Philip's two-term is probably deficient compared with using HYDRUS-1D. With large uncertainties of the t_{valid} [hour], it is then generally suggested that we propose a 95 % interval of the t_{valid} [hour] with respect to different soils rather than specific singular value of t_{valid} [hour] for each soil.

5 Conclusion

From the results of best samples with interpolated data, we can come to the conclusion that coarse soils (e.g. sand, loamy sand, sandy loam) have extremely short t_{valid} [hour] from 0.10 hour to 1.00 hour. Medium soils (sandy clay loam, loam, clay loam) have medium lengths of t_{valid} [hour] from 1.00 hour to 4.76 hours. The t_{valid} [hour] of the fine soils (clay loam, silty clay loam, silty clay, clay) could hardly be determined based our data due to the 95% intervals cover the measured t_{max}. Based on the experiments and results, we could only conclude that the 'true' t_{valid} [hour] of those soils will be longer than the results provided in **Table 4.7**. Despite this, the output 95% interval of t_{valid} [hour] of the coarse soils can be viewed as important results, providing the valid time that should be adopted to estimate the soil hydraulic properties using Philip's two-term model. The t_{valid} [hour] of the medium soils exhibits conspicuous and large uncertainties, which could hardly be used as a reference of feasible t_{valid} [hour] for other experimental tests.

Looking at the methodology that has been adopted, we conducted three experiments which differ with respect to the data that were analyzed. Through the steps of using Haverkamp model and DREAM algorithm, it shows that Haverkamp 1-D semi-implicit equation is a effective model to simulate the cumulative equation, which can be efficiently solved using a secant method and is not limited by time validity. By comparing the results derived from the best samples with interpolated data and the t_{char} proposed by Philip, the use of Bayesian Inference Criterion can be a feasible approach to derive the time va-

lidity of Philip's two-term model where two results both follow the pattern that coarse soils have shorter t_{valid} [hour] while fine soils have longer t_{valid} [hour]. The differences of results estimated from different data sources greatly specifies the drawbacks of using real-world data (SWIG) to study the soil hydraulic properties and t_{valid} [hour] of Philip's two-term model. The real-world data has the following issue when doing the analysis: (1) Data exhibit conspicuous measurement error, which will result in great uncertainty of estimation of t_{valid} [hour]; (2) Data are measured at large time interval, which means that the low-resolution measured times will lead to significant errors when estimating the t_{valid} [hour]; (3) Data are not measured under a relatively long period, which makes it hard to derive the 'true' t_{valid} [hour] of fine soils. Even though interpolation could be adopted to address the second problem, the other two issues could hardly be resolved or eliminated when using real-world data. This comes to the conclusion that instead of real-world data, using synthetic data (e.g. HYDRUS-1D) is a suggested approach to explore the t_{valid} [hour] of Philip's two-term model, which numerically computes infiltration from Richards' equation and eliminates the effect of measurement error.

References

Vandervaere, Jp, M. Vauclin, and D.E. Elrick (July 2000a). "Transient flow from tension infiltrometers: I. The two-parameter equation". In: *Soil Science Society of America Journal* 64, pp. 1263–1272. DOI: 10.2136/sssaj2000.6441263x.

— (July 2000b). "Transient flow from tension infiltrometers: II. Four methods to determine sorptivity and conductivity". In: *Soil Science Society of America Journal* 64, pp. 1272–1284.

Vrugt, J.A., Cajo ter Braak, C.J.F, et al. (Nov. 2008a). "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspce Sampling". In: *International Journal of Nonlinear Sciences and Numerical Simulation 10 (2009) 3* 10. DOI: 10.1515/IJNSNS.2009.10.3.273.

Vrugt, J.A., Cajo ter Braak, Hoshin Gupta, et al. (Oct. 2008b). "Equifinality of Formal (DREAM) and Informal (GLUE) Bayesian Approaches in Hydrologic Modeling?" In: *Stochastic Environmental Research and Risk Assessment* 23, pp. 1011–1026. DOI: 10.1007/s00477-008-0274-y.

Richards, L.A. (1931). "Capillary conduction of liquids through porous mediums". In: *Physics* 1.5, pp. 318–333.

Runge, C. (1931). "uber empirische Funktionen und die Interpolation zwischen aquidistanten Ordinaten". In: *Zeitschrift fur Mathematik und Physik* 46, pp. 224–243.

Horton, R.E. (1941). "An approach toward a physical interpretation of infiltration-capacity". In: *Biometrika* 5.C, pp. 399–417.

Philip, J.R. (1957a). "The theory of infiltration: 1. the infiltration equation and its solution". In: *Soil science* 83.5, pp. 345–358.

— (1957b). "The theory of infiltration: 4. Sorptivity and algebraic infiltration equations". In: *Soil science* 84.3, pp. 257–264.

Hastings, W.K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, pp. 97–109.

Mein, R.G. and C.L. Larson (1973). "Modeling infiltration during a steady rain". In: *Water Resource Research* 9.2, pp. 384–394.

Avriel, M (1976). *Nonlinear Programming: Analysis and Methods*. ISBN: 0-13-623603-0.

Schwarz, Gideon E. (1978). "Estimating the dimension of a model". In: *Annals of Statistics* 6.2, pp. 461–464.

Smith, R.E. (1978). "A parameter-efficient hydrologic infiltration model". In: *Water Resource Research* 14.3, pp. 533–538.

Van Genuchten, Martinus (Sept. 1980). "A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1". In: *Soil Science Society of America Journal* 44. DOI: 10.2136/sssaj1980.03615995004400050002x.

Bouwer, H (1986). "Intake rate: Cylinder infiltrometer". In: *Methods of Soil Analysis*, pp. 825–843.

Haverkamp, R., J.Y. Parlange, et al. (1990). "Infiltration under ponded conditions: 3. A predictive equation based on physical parameters". In: *Soil Science* 149, pp. 292–300.

Green, W.H. and G.A. Ampt (1991). "Studies on soil phyics". In: *he Journal of Agricultural Science* 4.1, pp. 1–24.

Gelman, A.G. and D.B. Rubin (1992). "Inference from iterative simulation using multiple sequences". In: *Stat. Sci.* 7, pp. 457–472.

Eliason, S.R. (1993). *Nonlinear Programming: Analysis and Methods*. ISBN: 0-8039-4107-2.

Haverkamp, R., P.J. Ross, et al. (1994). "Three-dimensional analysis of infiltration from the disc infiltrometer: 2. physically based infiltration equation". In: *Water Resources Research* 30.11, pp. 2931–2935.

Vogel, Tomas et al. (Aug. 1996). "The HYDRUS Code for Simulating One-Dimensional Water Flow, Solute Transport, and Heat Movement in Variably-Saturated Media. Version 5.0". In: DOI: 10.13140/RG.2.1.3456.7525.

Storn, R. and K. Price (Jan. 1997). "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces". In: *Journal of Global Optimization* 11, pp. 341–359. DOI: 10.1023/A:1008202821328.

Carsel, R.F. and R.S. Parrish (1998). "Developing joint probability distributions of soil water retention characteristics". In: *Water resources research* 24.5, pp. 755–769.

Gregory, Justin H. et al. (2005). "Analysis of Double-Ring Infiltration Techniques and Development of a Simple Automatic Water Delivery System". In: *Applied Turfgrass Science* 2.1, pp. 1–7.

Lassabatere, L. et al. (2009). "Numerical evaluation of a set of analytical infiltration equations". In: *Water Resources Research* 45.12. DOI: 10.1029/2009WR007941.

Valiantza, J.D. (2010). "New linearized two-parameter infiltration equation for direct determination of conductivity and sorptivity". In: *Journal of Hydrology* 384.1-2, pp. 1–13.

Vrugt, J.A. (2016). "Markov chain Monte Carlo simulation using the DREAM software parckage:Theory, concepts and MATLAB implementation". In: *Environmental Modelling Software* 75, pp. 273–316.

Mehdi Rahmati, et al (2018). "Development and analysis of the Soil Water Infiltration Global database". In: *Earth Systems Science Data* 10, pp. 1237–1263.

Jaiswal, P. (2019). "Bayesian Determination of the Soil Hydraulic Parameters and the Time Validity of Philip's Two-Term Infiltration Equation From Measured Infiltration Data". In: *UC Irvine* ProQuest ID: Jaiswal_uci_0030M_15834. DOI: escholarship.org/uc/item/76n9152m.