SOCRATES-Retrieval

User's Manual

Seho Kim (kim3002@purdue.edu)

School of Aeronautics and Astronautics
Purdue University

Version 1.0

Contents

Chapter 1

Introduction

1.1 Overview

SOCRATES-Retrieval is a retrieval module of SOCRATES to simulate the retrieval of subsurface soil moisture (RZSM) and vegetation water content (VWC) using synthetic observations of multi-frequency/polarimetric signals of opportunity (SoOp). This simulation tool integrates the SCoBi model [1] and the Principle of Maximum Entropy (POME) model to provide a bistatic SoOp environment and resulting measurements in a comprehensive manner. The current release of SOCRATES-Retrieval boasts the following capabilities:

- Soil moisutre profile and VWC retrieval using multi-frequency/polarimetric SoOp.
- Sensitivity analysis of retrieval accuracy to SoOp system parameters.
- Antenna property realizations including gain pattern.
- Geophysical data realizations including soil moisture and vegetation.

Use SOCRATES for more functionality such as generic SoOp coverage analysis and trade study of SoOp constellation design.

1.2 Compilation

The GNU Compiler Collection with C11 standard on a Unix or Linux platform is recommended to compile the code. Additional software packages OpenMPI and netCDF are required for the compilation. Latest versions of the packages are recommended. OpenMPI is for multi-processing the retrieval algorithm. This is not used in the synthetic observation generation. NetCDF is for creating, reading, and writing the NetCDF files.

SOCRATES-Retrieval can be accessed from the following **github** repository: `https://github.itap.purdue.edu/RadioNavigationLab/SOCRATES-Retrieval`. To install from the Git repository:

1. Clone the code from Git:

   `$git clone git@github.itap.purdue.edu:RadioNavigationLab/SOCRATES-Retrieval.git`

2. Open a terminal on the top-level directory (i.e., **.../SOCRATES-Retrieval**) and run `$ make` to compile the simulator. Note that `$ make clean` deletes the existing object files.

3. An executable file `SOCRATES-Retrieval` should be created on the top-level directory.

1.3 Getting Started

SOCRATES provides a MATLAB script to guide the workflow of end-to-end retrieval simulation. Open **main.m** in the top-level directory and follow the instructions in Chapter 2.

1.4 Citation

SOCRATES-Retrieval is open-source software under GNU General Publich Licence (GPL) and freely available with its documentation. However, the developers of the simulator would appreciate those who cite the corresponding study below in the case they are used:

Seho Kim, James Garrison, and Mehmet Kurum, "Retrieval of subsurface soil moisture and vegetation water content from multi-frequency SoOp-Reflectometry: Sensitivity Analysis," *IEEE Transactions of Geoscience and Remote Sensing*, 2023.

For any questions, please contact Seho Kim (kim3002@purdue.edu).

1.5 Disclosure of Third-party Code

This software includes material copied from or derived from the following sources:

- Some math kits copied from 42:
  https://github.com/ericstoneking/42, Copyright (C) 2010 United States Government as represented by the Administrator of the National Aeronautics and Space Administration.
- The Principle of Maximum Entropy (POME) model translated from Python code:
  https://github.com/Vikalp86/soil-moisture-profile, Copyright (C) 2019 Vikalp86.
- The SCoBi model translated from MATLAB code:
  https://github.com/impresslab/SCoBi, Copyright (C) 2017-2018 Mehmet Kurum, Orhan Eroglu, Dylan R. Boyd.

Chapter 2

Retrieval Simulation

The m-script **main.m** provides a step-by-step guideline to complete the retrieval simulation including the processing of USCRN dataset, synthetic observation generation, and retrieval. In this chapter, we introduce each step listed in **main.m**. It is recommend to run one section at a time to avoid any confusion in the simulation. Note that some sections may take several hours. Read instructions carefully and understand the functionality before running each section. Open **main.m** in MATLAB and click the *Run Section* icon in the *Editor* section of the toolbar or click the *Run Advance* icon to proceed to the next section after running one section.

2.1 Step 1: Processing Source Data

First, we collect all necessary ground data and combine them into a single NetCDF file for each USCRN station. For the demonstration, we process data for *CO_Cortez_8_SE* station for 2016. See the description below for user parameters, input, and output. Set the user parameters and run the section.

- User parameters:
  - Stations of interest: Select USCRN stations of interest from **./inputs/station_list.xlsx** and put them into **./inputs/station_in.txt**. Note that the first line of the text file must be the total number of the stations of interest listed in the file. The name of each station must be provided from the second line separated by a line. The default station is *CO_Cortez_8_SE*.
  - `years`: Set a vector of the year of interest between 2016 and 2020. The default value is [2016].
  - `months`: Set a vector of the start month and the end month. The start month must be less than the end month. The default value is [1, 12].
  - `orbits`: Set a cell array of the orbit. Select 'ISS' and/or 'SSO'. The default value is {'ISS'}.
- Input:
  - Soil moisture: Hourly in-situ soil moisture data at 5 depths (5, 10, 20, 50, 100 cm) for each selected USCRN station in the selected years are required. Once stations and years are selected, **processUSCRNdata.m** will automatically download the data for you. The data will be located in **./data/USCRN/**.
  - NDVI: Daily NDVI (Normalized Difference Vegetation Index) data for each selected USCRN station is required. The 5-year NDVI for all (7) stations listed are provided in **./data/MODIS-NDVI/**. They are linearly interpolated from MOD13Q1 of MODIS, 16-day NDVI at 250-m resolution, for each station.
  - Incidence angle set: Hourly incidence angle data of three transmitting systems (Orbcomm, MUOS, and GPS) formed on the same area is required. A set of incidence angles was randomly sampled as an input for each station. This data is provided in **./data/geometry/**.

- Output:
  - Ground data file: A combined data for soil moisture and vegetation for each selected USCRN station. This is used in Step 2-1 as an input file for generating synthetic observation. This data will be created in **./data/processed/** in a NetCDF file format. The naming convention is 'USCRN_hourly_`[year]`M`[start month]`M`[end month]`_`[orbit]`_`[station name]`.nc'.
  - Ground data plot: A plot of a ground data file. This plot presents the time series of the following data: volumetric soil moisture (VSM), precipitation, soil temperature, the daily minimum air temperature, vegetation water content (VWC), and incidence angles of Orbcomm, MUOS, and GPS. This plot will be created in **./data/processed/**. The naming convention is 'USCRN_hourly_`[year]`M`[start month]`M`[end month]`_`[orbit]`_`[station name]`.png'. Figure 2.1 is the example for the default setting.

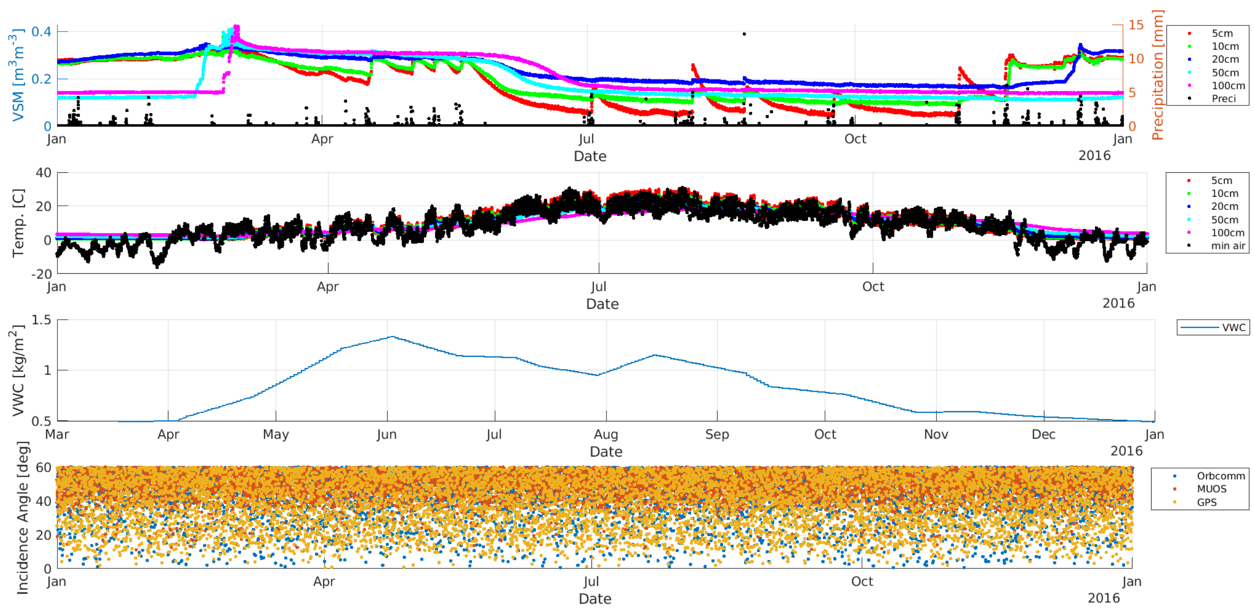

Figure 2.1: (Step 1) Ground data of *CO_Cortez_8_SE* station after processing.

2.2 Step 2-1: Generating Synthetic Observations

Second, we generate synthetic reflectivity observation at multi-frequency and multi-polarization. The ground data file output from Step 1 is required for this step. **runForwardModel.m** updates input files for SOCRATES-Retrieval with the user parameters defined in Step 1 and runs SOCRATES-Retrieval as a forward mode. SOCRATES-Retrieval filters invalid ground samples (frozen soil/air and missing data) and simulates reflectivities at four frequencies (137, 255, 370, and 1575.42 MHz) and all polarizations (RHCP, LHCP, vertical, and horizontal) for each sample. Note that this overwrites any existing files and may take several hours to complete. See the description below for user parameters, input, and output. Set the user parameters and run the section.

- User parameters: Same as Step 1.
- Input:
  - Ground data file: A NetCDF file created from Step 1 that contains the combined data for soil moisture and vegetation data for each USCRN station.

- SOCRATES-Retrieval input files: See Chapter 3 for details on input files for SOCRATES-Retrieval.

- Output:
  - Synthetic observation file: A simulated multi-frequency/polarimetric reflectivity for each station for the valid soil moisture and vegetation data. The file will be located in **./results/forward/**. The naming convention is 'USCRN_hourly_[`year`]M[`start month`]M[`end month`]_[`orbit`]_[`station name`]_forward.nc'.

2.3 Step 2-2: Plotting Synthetic Observations

This step plots the time series of simulated reflectivities at all frequencies and polarizations along with VSM, VWC, and incidence angles for each station and year. See the description below for user parameters, input, and output. Set the user parameters and run the section.

- User parameters: Same as Step 1.
- Input:
  - Ground data file from Step 1.
  - Synthetic observation file from Step 2.
- Output:
  - Synthetic observation plot: A plot of a synthetic observation data file. This plot presents the time series of the following data: VSM, reflectivity at RHCP, LHCP, vertical (X), and horizontal (Y) polarizations, VWC, and incidence angles of Orbcomm, MUOS, and GPS. This file will be located in **./figures/forward/**. The naming convention is 'USCRN_hourly_[`year`]M[`start month`]M[`end month`]_[`orbit`]_[`station name`]_TS.png'. Figure 2.2 is the example for the default setting.

2.4 Step 3: Estimating *b* Parameters

Third, we estimate the b parameters for each frequency and polarization to relate the vegetation optical depth (VOD) to the VWC. The b parameter is required in Step 4 for retrieval. We use 10% of the samples (hourly data) randomly selected as a training data for the linear regression of the b parameter. See the description below for user parameters, input, and output. Set the user parameters and run the section.

- User parameters: Same as Step 1.
- Input:
  - Ground data file from Step 1.
  - Synthetic observation file from Step 2.
- Output:
  - b parameter text file: A text file listing the estimates of the b parameter and the RMSE of linear regression for all frequencies and polarizations. The file path is **./inputs/b_parameters.txt**. This file name must be provided in **Inp_Retrieval.txt**. See Section 3.4 for details.
  - b parameter MAT-file: A MAT-file containing the b parameters, VWC, and VOD. The file path is **./results/inverse/b_parameters.mat**. This file is only for storing and plotting the b parameters and related data.
  - b parameter plot: A plot of b parameters of all frequencies and polarizations estimated by linear regression relating VOD to VWC. This file is located in **./figures/**. The file name is 'b_linearRegression.png'. Figure 2.3 is the example for the default setting.

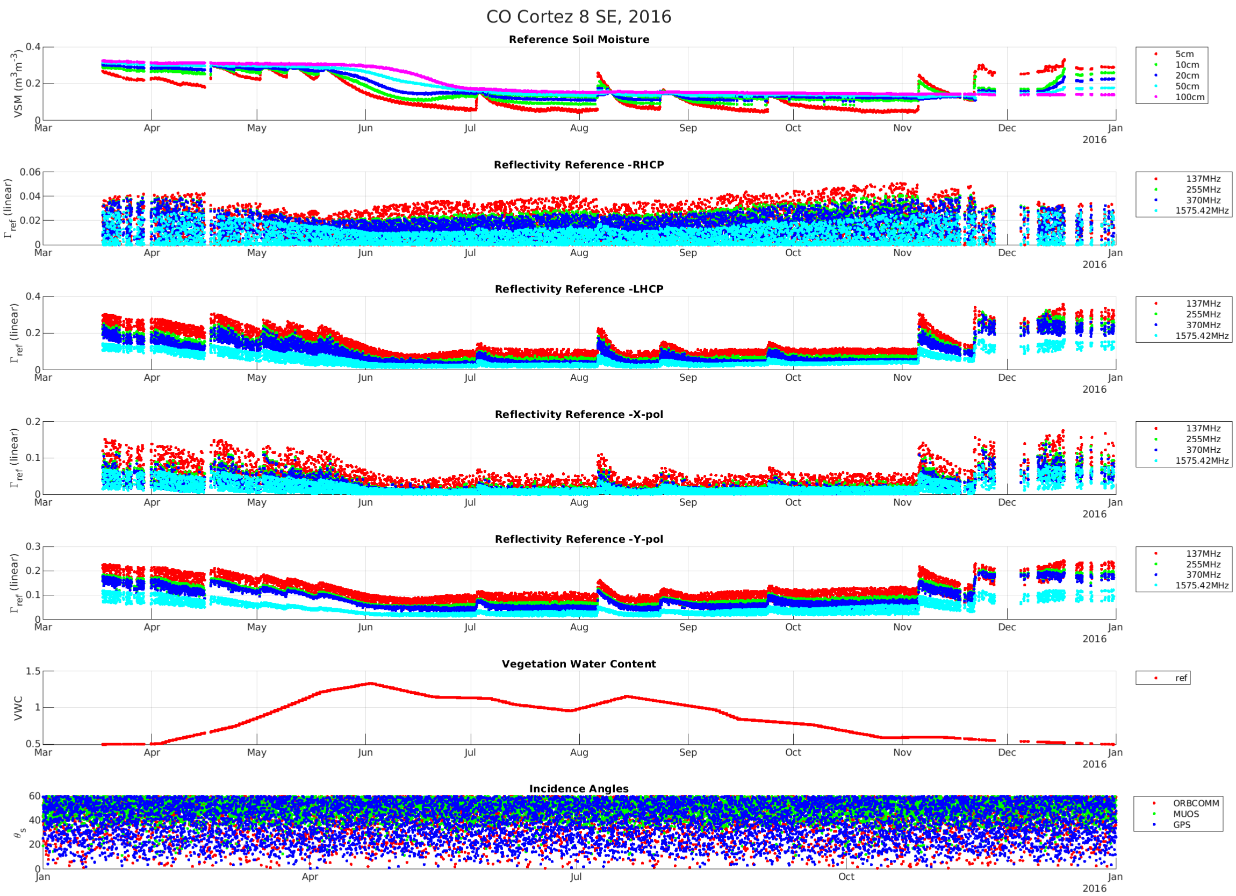

Figure 2.2: (Step 2-2) Synthetic observation of *CO_Cortez_8_SE* station.

2.5 Step 4-1: Running Retrieval Simulations

Fourth, we define the SoOp-R system parameters for retrieval and run the retrieval simulation. The synthetic observation file generated from Step 2-1 is required for this step. **runInverseModel.m** updates the input files for SOCRATES-Retrieval and runs SOCRATES-Retrieval as a inverse mode. See the description below for user parameters, input, and output. Set the user parameters and run the section.

- User parameters:
  - `freq`: Set a cell array of numeric arrays representing frequency combinations. The notation of the frequency is as follows; 1: 137 MHz, 2: 255 MHz, 3: 370 MHz, 4: 1575.42 MHz. A user can put multiple combinations in the cell array. For example, to simulate all combinations of 137 MHz and 255 MHz, this parameter should be set as {[1,2], [1], [2]}. The default value is {[1,2,3,4]}, which indicates the use of one combination including all four frequencies.
  - `polRx`: Set a cell array of strings representing the receiver polarization. The notation of the receiver polarization is as follows: 'R': RHCP, 'L': LHCP, 'X': Vertical, 'Y': Horizontal. Any combination of these polarizations are possible. As the transmitter's polarizations are RHCP, we recommend to use LHCP ('L') or a dual linear polarization ('XY'). A user can put multiple polarizations in the cell array. The default value is {'XY'}.
  - `stddev`: Set a numeric value for the reflectivity error. This value defines the fraction of the reflectivity value to be used as the standard deviation of zero-mean Gaussian random error added to the reflectivity. The default value is 0.04, or 4%. For details, see the description on the parameter η in [2].

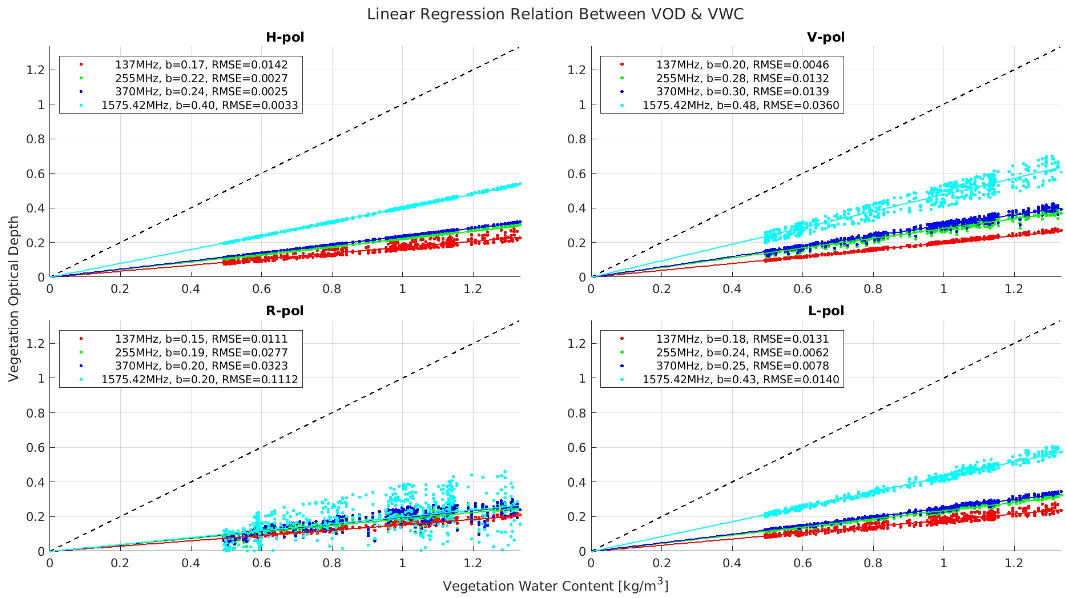

Figure 2.3: (Step 3) The b parameter estimated from *CO_Cortez_8_SE* station.

- `period`: Set a numeric value for the interval of retrievals in hours. The default value is 168 hours, or 7 days. For details, see the description on the parameter P in [2].
- `window`: Set a numeric value for the observation time window in hours. All observations will be randomly distributed within the given time window. The time window will be centered at the retrieval time defined above. The default value is 12 hours. For details, see the description on the parameter T in [2].
- `np`: Set a numeric value for the number of processors used in retrieval. SOCRATES-Retrieval uses the MPI-based multi-processing to speed up the series of retrievals. One processor simulates one retrieval at a time. The default value is 24.

- Input:
  - Synthetic observation file from Step 2.
  - SOCRATES-Retrieval input files: See Chapter 3 for details on input files for SOCRATES-Retrieval.

- Output:
  - Retrieval file: A simulated retrival for soil moisture and vegetation data of each station. The file will be located in **./results/inverse/**`[System parameters]`**/**. The naming convention of [`System parameter`] is 'p[`period`]w[`window`]f[`freq`][`polRx`]std[`stddev`][`orbit`]'. The naming convention of the retrieval file is 'USCRN_p[`period`]w[`window`]std[`stddev`]_[`orbit`]_[`station name`]_ret.nc'.

2.6 Step 4-2: Analyzing Retrieval Results

Lastly, we compute and plot the error statistics of the retrieval results. See the description below for user parameters, input, and output. Set the user parameters and run the section.

- User parameters: Same as Step 4-1 except `np`.
- Input:

– Ground data file from Step 1.

– Synthetic observation file from Step 2.

– Retrieval file from Step 4-1.

- Output:

  – RMSE files: Text files listing the root mean square error (RMSE) of retrievals. All stations listed in **./inputs/station_in.txt** are included in the error anaylsis. The files will be located in **./results/inverse/**[`System parameters`]**/**. The naming convention of [`System parameter`] is 'p[`period`]w[`window`]f[`freq`][`polRx`]std[`stddev`][`orbit`]'. There are three files for each system parameter: RMSE of soil moisture profiles at the five depths (5, 10, 20, 50, 100 cm), RMSE of soil moisture profiles at all depths, RMSE of VWC. The naming conventions are

    * USCRN_p[`period`]w[`window`]std[`stddev`]_[`orbit`]_RMSE_SMP.txt,
    * USCRN_p[`period`]w[`window`]std[`stddev`]_[`orbit`]_RMSE_SMP_alldepth.txt,
    * USCRN_p[`period`]w[`window`]std[`stddev`]_[`orbit`]_RMSE_VWC.txt,

    respectively.

  – Sum total file: A text file listing the total number of retrieved soil moisture profiles for each station and the total sum of profiles of all stations. All stations listed in **./inputs/station_in.txt** are included in the file. The file will be located in **./results/inverse/**[`System parameters`]**/**. The naming convention of [`System parameter`] is 'p[`period`]w[`window`]f[`freq`][`polRx`]std[`stddev`][`orbit`]'. The naming convention of the file is 'USCRN_p[`period`]w[`window`]std[`stddev`]_[`orbit`]_nProfiles.txt'

  – Analysis plots: Plots of error analyis of retrieval results. The plots will be located in **./figures/inverse/**[`System parameters`]**/**. The naming convention of [`System parameter`] is 'p[`period`]w[`window`]f[`freq`][`polRx`]std[`stddev`][`orbit`]'. There are five types of files:

    1. USCRN_hourly_[`year`]M[`start month`]M[`end month`]_ [`orbit`]_SM_matchups.png: One-to-one matchups of reference and retrieved soil moisture including all selected stations. Figure 2.4 is the example of the default setting.
    2. USCRN_hourly_[`year`]M[`start month`]M[`end month`]_[`orbit`]_VWC _matchups.png: One-to-one matchups of reference and retrieved VWC including all selected stations. Figure 2.5 is the example of the default setting.
    3. USCRN_hourly_[`year`]M[`start month`] M[`end month`]_[`orbit`]_SMP_RMSE.png: RMSE of soil moisture profiles as a function of depth for each station. This plot includes all selected stations. Figure 2.6 is the example of the default setting.
    4. USCRN_hourly_[`year`]M[`start month`]M[`end month`]_[`orbit`] _[`station name`]_TS_ret.png: The time series of reference and retrieved soil moisture and VWC for each station. Figure 2.7 is the example of the default setting.
    5. USCRN_hourly_[`year`] M[`start month`]M[`end month`]_[`orbit`]_[`station name`]_SA.png: The time series of parameters for simulated annealing for each station, such as the number of iterations, the number of jumps (from the local minima), the CPU time taken in retrieval, and the cost value. Figure 2.8 is the example of the default setting.

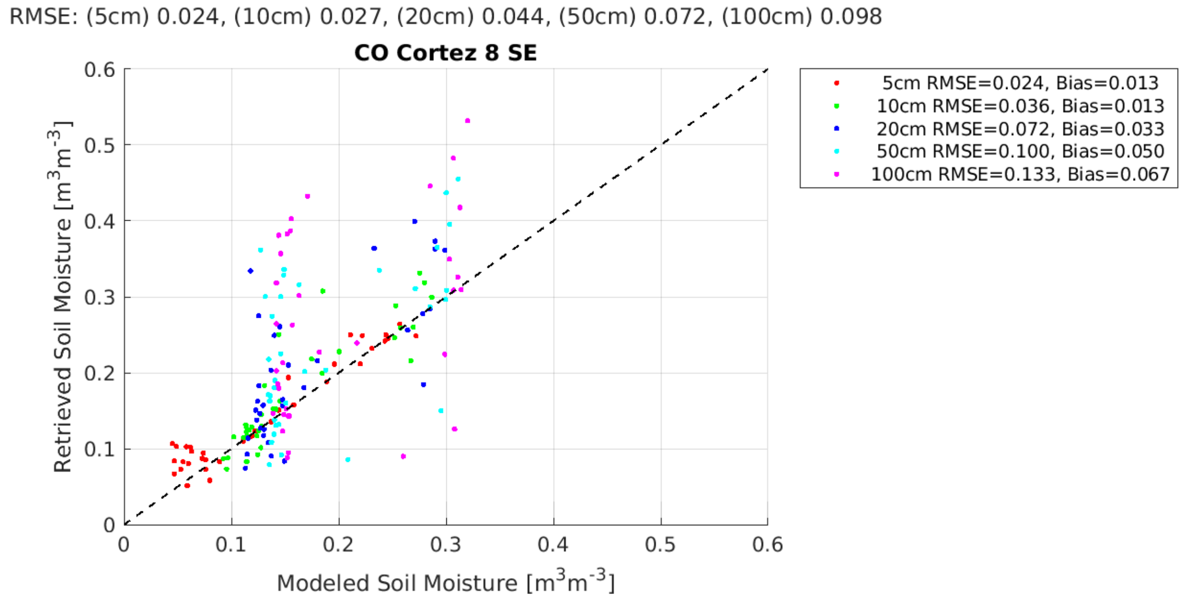

Figure 2.4: (Step 4-2) One-to-one matchups of reference and retrieved soil moisture at *CO_Cortez_8_SE* station.

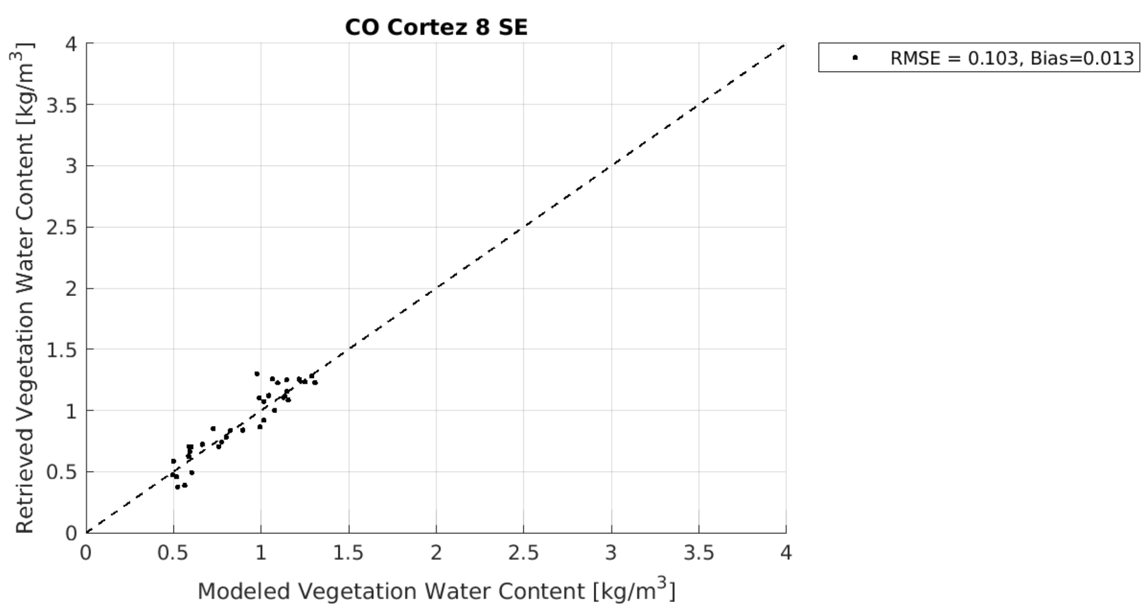

Figure 2.5: (Step 4-2) One-to-one matchups of reference and retrieved VWC at *CO_Cortez_8_SE* station.

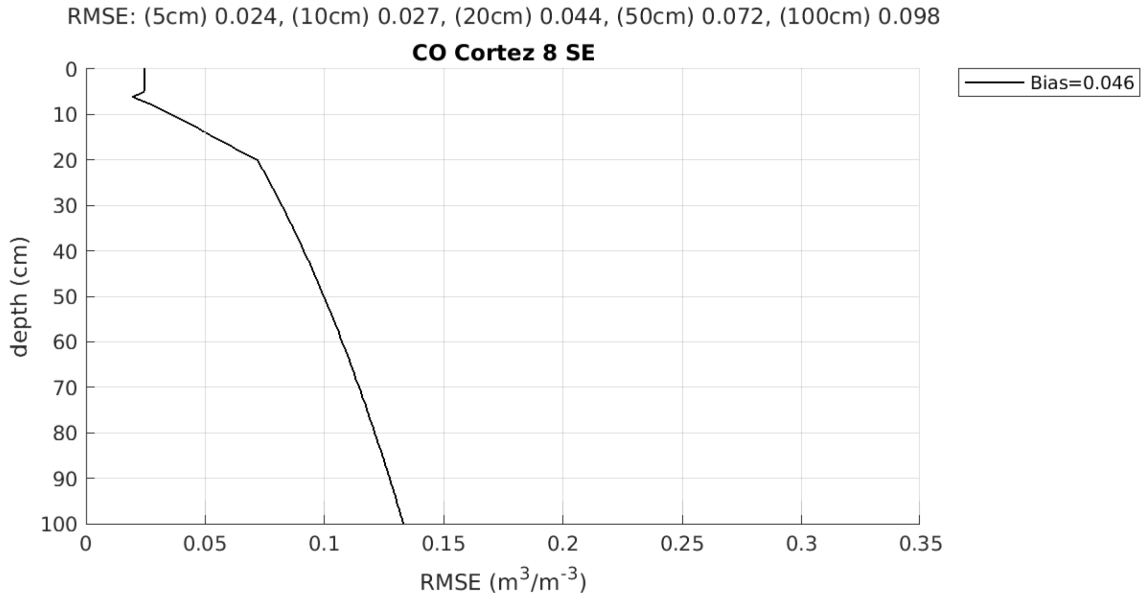

Figure 2.6: (Step 4-2) RMSE of soil moisture profiles as a function of depth at *CO_Cortez_8_SE* station.

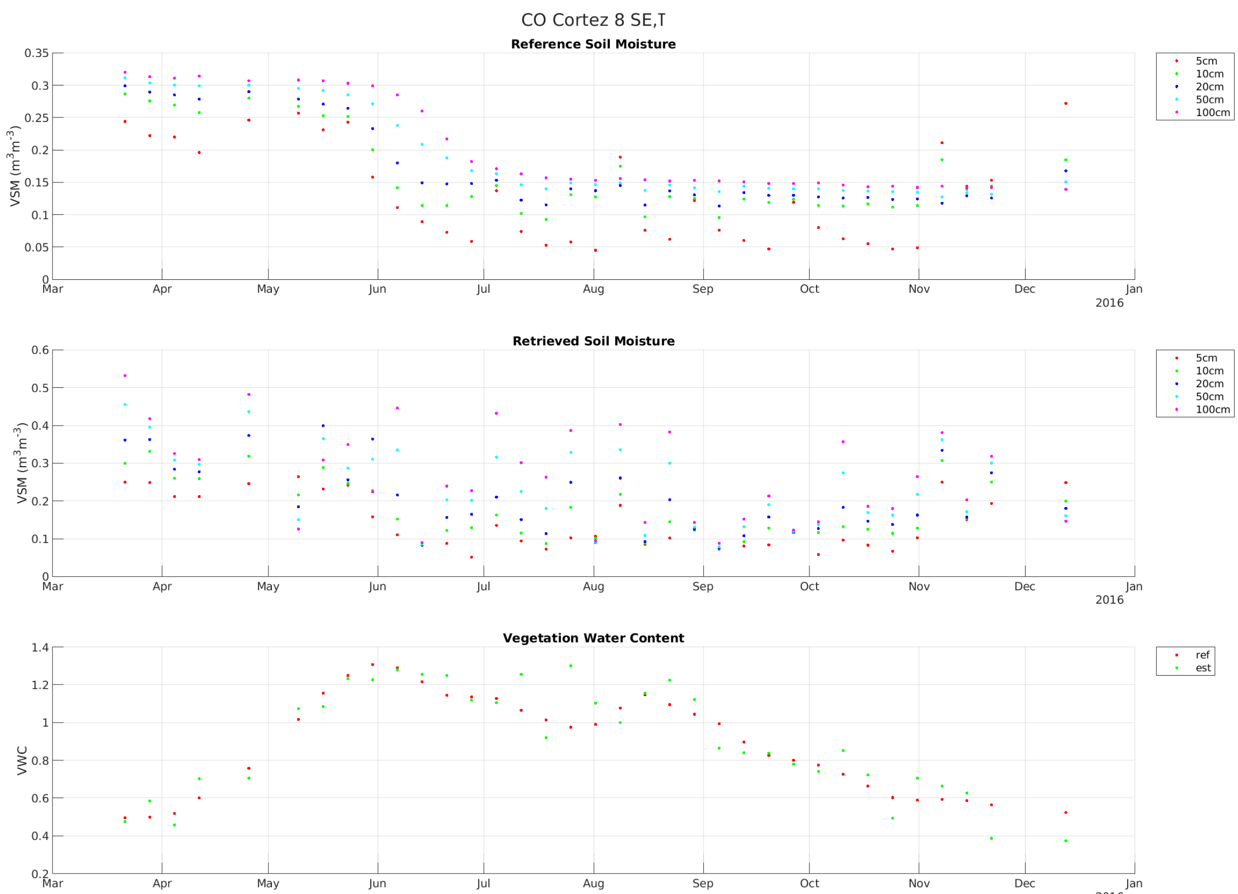

Figure 2.7: (Step 4-2) The time series of reference and retrieved soil moisture and VWC at *CO_Cortez_8_SE* station.

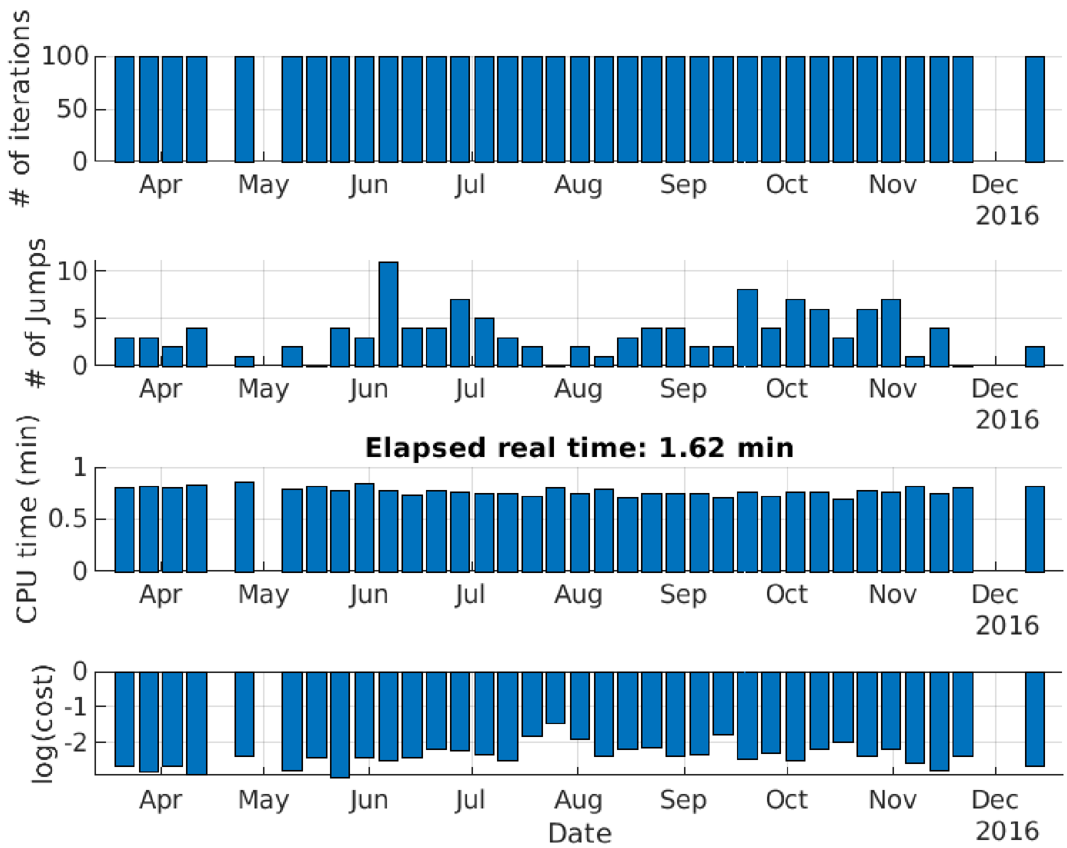

Figure 2.8: (Step 4-2) The time series of parameters for simulated annealing at *CO_Cortez_8_SE* station.

Chapter 3

Input Files

3.1 Inp_Main.txt

This file is the top-level input file which defines the ground and vegetation structure to be observed:

- `Dynamic Data`:
  - `POME Model`:
    - `Total Depth`: The total depth of the ground structure in centimeters.
    - `Layer Discretization`: The thickness of a single soil layer in millimeters. The total number of layers is determined by the total depth divided by this layer thickness.
    - `Initial Inflection Depth`: The initial guess of the inflection point in centimeters.
  - `USCRN`:
    - `USCRN Station List File Name`: The name of a text file providing the list of the USCRN stations of interest that SOCRATES-Retrieval will look into. The first line of the file must be the total number of the stations of interest listed in the file. The name of each USCRN station must be provided from the second line of the file separated by a line. This file should be located in **./inputs** folder.
    - `Start Year, End Year`: The starting and ending years. The years in between will be included.
    - `Start Month, End Month`: The starting and ending months. The months in between will be included. The start month must be less than the end month.
- `Static Data`:
  - `Vegetation`: `TRUE/FALSE` turns on/off the vegetation layer on the ground structure.
  - `RMSH Roughness of Soil Surface`: Root-Mean-Square-Height roughness of soil surface in centimeters.

3.2 Inp_Fixed.txt

This file describes input parameters for fixed observation:

- `Receiver`:
  - `Receiver Altitude`: The altitude of a receiver in kilometers. The altitude makes only difference on the received power calculations.
  - `Receive Antenna Gain, Polarization`: The antenna gain of the receiver should be given in decibels. The polarization of the receiver can be chosen to be: `R`: RHCP, `X`: linear vertical polarization, or `RX`: dual polarization - RHCP and linear vertical. SOCRATES-Retrieval simulates co- and cross-polarization simultaneously.

- `Antenna Pattern Input`: `IDEAL/GAUSSIAN/USER_DEFINED` selects the receiver antenna pattern. SOCRATES-Retrieval currently supports three different receiver antenna pattern generation methods.
- `(GAUSSIAN) HPBW, Sidelobe, X-pol, Res`: The properties of a simple generalized Gaussian antenna pattern: the half-power bandwidth in degrees, first side-lobes level in dB, cross-polarization level in dB, and minimum sensitivity in degrees. Skip this if `IDEAL` or `USER_DEFINED` is selected.
- `(USER_DEFINED) Antenna Pattern File Name`: If the receiver antenna pattern is selected to be `USER_DEFINED`, four Excel spreadsheets are required, which holds the normalized voltage values for co- and cross-polarizations for X and Y ports. Antenna pattern should be provided such that in each sheet of the Excel file, columns represent the theta-angle look-up and rows represent the phi-angle look-up. Theta (θ) angles have a 180° scan, while phi (ϕ) angles have a 360° scan. Note that the antenna pattern files do not have headers within the Excel spreadsheet. Each sheet consists of numbers only. This file should be located in **./ant_pat** folder. Skip this if `IDEAL` or `GAUSSIAN` is selected.

- `Transmitter`:
  - `Number of Transmitters`: The number of transmitting systems (or constellations). The following input sets must be provided for each transmitting system.
  - `Transmitter Frequency`: The transmitter antenna's operating frequency in MHz units.
  - `Transmitter Altitude`: The transmitter's altitude should be given in kilometers.
  - `Transmitter Polarization`: The polarization of the transmitter can be chosen to be: `R`: RHCP, `L`: LHCP, `X`: linear vertical polarization, or `Y`: linear horizontal polarization.

3.3 Inp_Veg.txt

This file describes input parameters for vegetation structure based on the discrete scatterer approach used in the SCoBi model [1].

- `Number of Vegetation Types`: The number of vegetation types. The following input sets must be provided for each vegetation.

- `Vegetation`:
  - `Label`: The label or name of the vegetation.
  - `Number of Layers`: The number of the vegetation's sublayers. The following input sets must be provided for each sublayer.
  - `Thickness`: The thickness of the sublayer in millimeters.
  - `Number of Included Kinds, Particle ID of Kinds`: The number of scatterer Kinds included in the sublayer. The particle IDs of included Kinds must be provided separated by spaces. The particle ID of each Kind must be defined under `Vegetation Scatterer Definition` section.

- `Vegetation Scatterer Definition`:
  - `Number of Scatterer Kinds`: The total number of scatterer Kinds to define. The following input sets must be provided for each Kind.
  - `Particle ID`: The particle ID can be assumed the Kind name, which must be defined by two characters, where the first character must be one of the following letters: `L`: Leaf, `B`: Branch, `T`: Trunk, `N`: Needle, or `W`: Water, and the second charater should be integer to represent a specific Kind (e.g. "`L1`", "`L2`", "`B1`", "`T3`", etc.). This ID is used in `Vegetation` section to link the vegetatoin sublayers to the scatterer Kinds.
  - `Density`: The number of scattering particles within one cubic meter.

- Dimensions: Three dimension parameters of either a cylinder (Branch, Trunk, or Needle) or a elliptical disk (Leaf of Water) separated by spaces. The first dimension is either the radius of the start of a cylinder or major axis of an elliptical disk in meters. The second dimension is either the radius of the end of a cylinder or minor axis of an elliptical disk in meters. The third dimension is either the length of a cylinder or thickness of an elliptical disk in meters.
- Volumetric Water Content: The volumetric water content of a particle.
- Interval Angle of Orientation: The minimum and maximum interval value of the scatterer orientation separated by spaces that is assumed to be uniformly distributed between two angular boundaries.

3.4 Inp_Retrieval.txt

This file describes input parameters for retrieval.

- Receive Polarization: Receiver antenna polarization can be a single or a dual polarziation defined by a combination of the following: R: RHCP, L: LHCP, X: linear vertical polarization, or Y: linear horizontal polarization.
- Standard Deviation of Reflectivity Error Ratio: The reflectivity error parameter defining the standard deviation of additive zero-mean Gaussian random error, which is set to be a fraction of the reflectivity.
- Measurement Period, Time window: The observation period P and the observation time window T in hours.
- The b parameter input file name: The name of a text file listing the b parameters created by **estimate_b_parameter.m**. This file should be located in **./inputs** folder.
- POME Model:
  - Total Depth: The total depth of the ground structure to retrieve in centimeters.
  - Layer Discretization: The thickness of a single soil layer in millimeters. The total number of layers is determined by the total depth divided by this layer thickness.
  - Initial Inflection Depth: The initial guess of the inflection point in centimeters.

References

[1] Mehmet Kurum, Manohar Deshpande, Alicia T Joseph, Peggy E O'Neill, Roger H Lang, and Orhan Eroglu. Scobi-veg: A generalized bistatic scattering model of reflectometry from vegetation for signals of opportunity applications. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):1049–1068, 2018.

[2] Seho Kim, James Garrison, and Mehmet Kurum. Retrieval of subsurface soil moisture and vegetation water content from multi-frequency soop-reflectometry: Sensitivity analysis. 2023.