SOC Calibration Framework

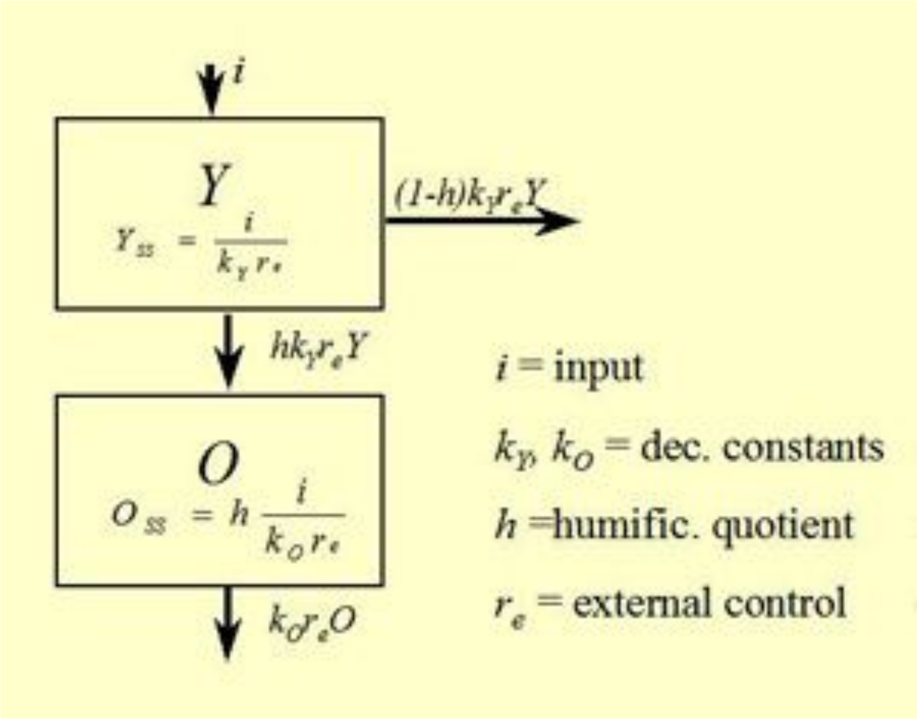
Francis Durnin-Vermette,
Arumugam Thiagarajan

Environment and
Climate Change Canada

Outline

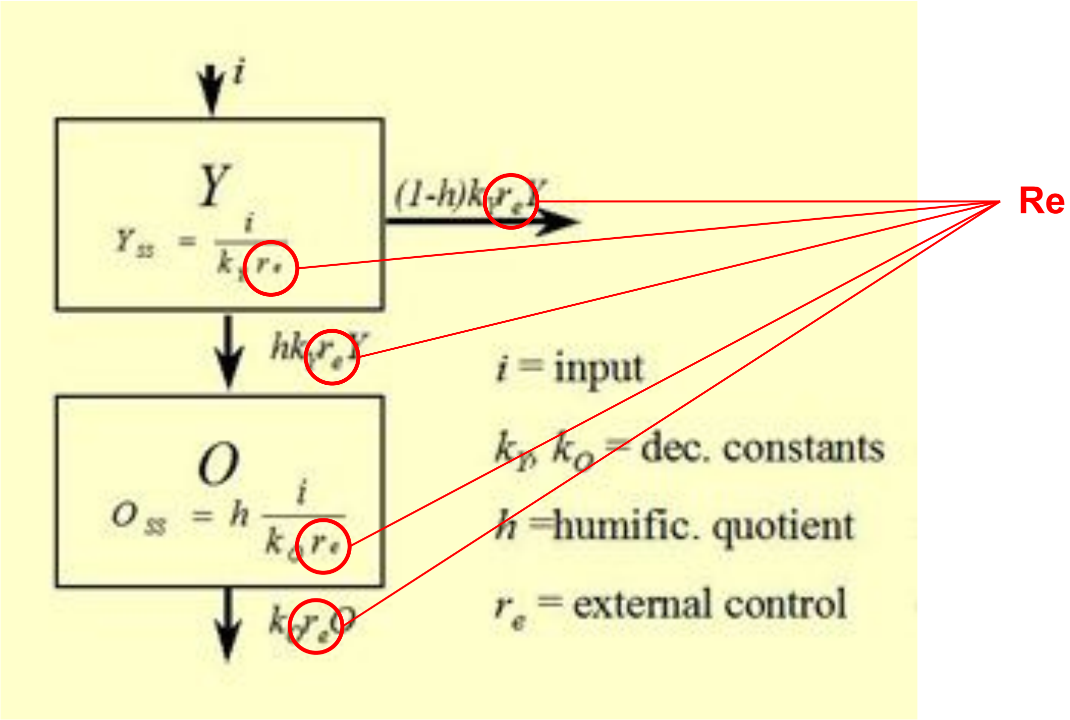
1. Migrating ICBM Re calculator from C# to R
2. Development of the Calibration Framework in R

ICBM



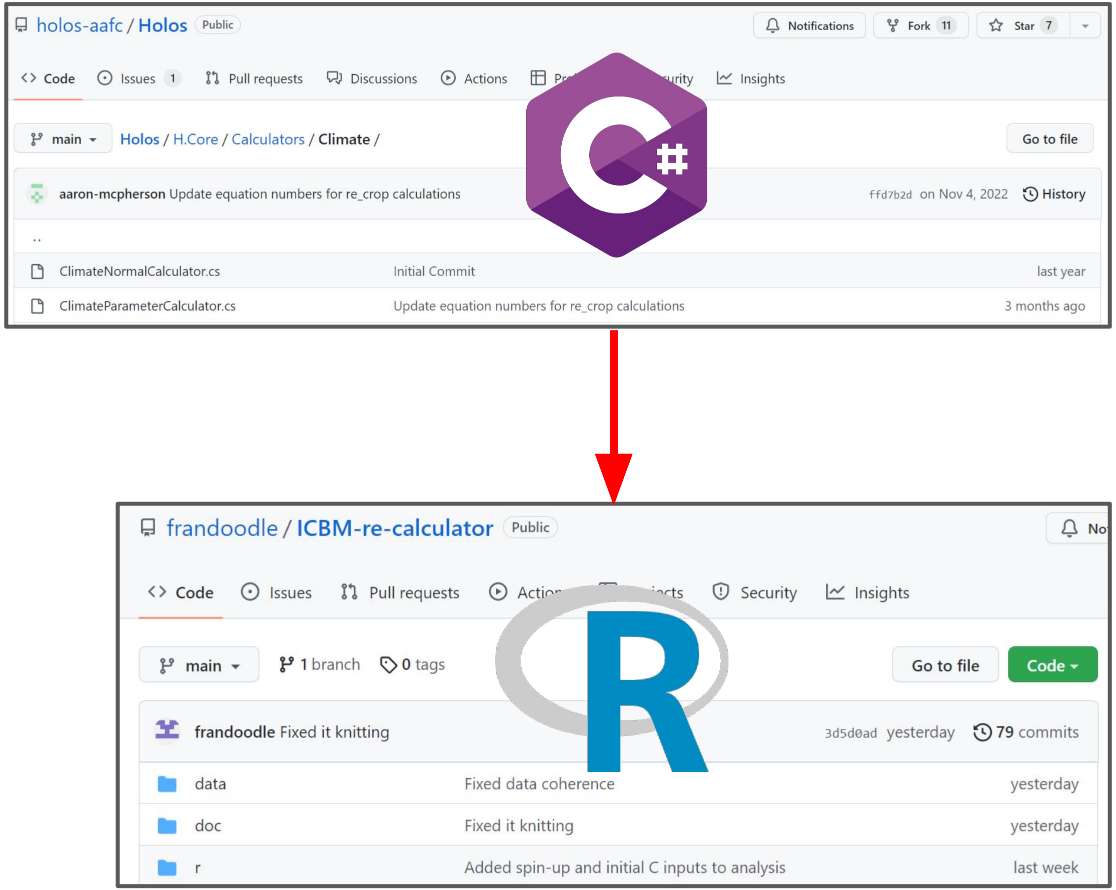
i
Y
$Y_{ss} = \frac{i}{k_Y r_e}$
$(1-h)k_Y r_e Y$
$hk_Y r_e Y$
O
$O_{ss} = h \frac{i}{k_O r_e}$
$k_O r_e O$
i = input
k_Y, k_O = dec. constants
h =humific. quotient
r_e = external control

ICBM


i
Y
$Y_{ss} = \frac{i}{k_Y r_e}$
$(1-h)k_Y r_e Y$
$hk_Y r_e Y$
O
$O_{ss} = h\frac{i}{k_O r_e}$
$k_O r_e O$
i = input
k_Y, k_O = dec. constants
h =humific. quotient
r_e = external control
Re

ICBM Re Calculator

- Migrated the Holos ICBM Re calculator (C#) into R code
- The current public repository for this code is ICBM-re-calculator (https://github.com/frandoodle/ICBM-re-calculator)



ICBM Re Calculator

- Standalone Re calculation also implemented

```
calculate_re <- function(YearInputTable,
                         yield,
                         perennial,
                         SoilOrganicC_Percent,
                         ClayContent,
                         SandContent,
                         alfa = 0.7,
                         SoilTopThickness = 250,
                         Temp_min = -3.78,
                         Temp_max = 30,
                         r_s = 0.42,
                         r_wp = 0.18,
                         ReferenceAdjustment = 0.10516,
                         r_c = NA,
                         tillage_soil = "Brown",
                         tillage_type = "Intensive Tillage",
                         irrigation_region = "Canada",
                         irrigation_use_estimate = FALSE,
                         irrigation = 0,
                         ...)
```

ICBM Re Calculator

- Standalone Re calculation also implemented
- Full code walkthrough, documentation, and QC testing found in doc/walkthrough_re.html

Running ICBM r_e Calculator

Francis Durnin-Vermette

2022/11/22

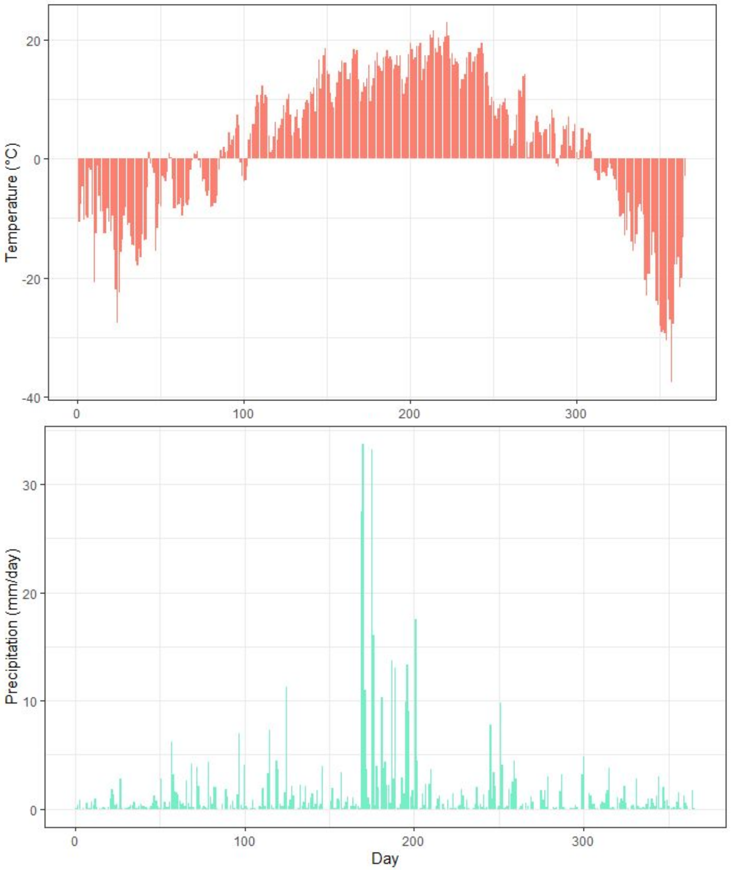
- 1 Quick Start - r_e Calculator
- 2 Walkthrough
  - 2.1 Input data
  - 2.2 Green area index (GAI) dynamics
    - 2.2.1 Eq. 2.2.1-1 through Eq. 2.2.1-3
  - 2.3 Water content at wilting point and field capacity
    - 2.3.1 Eq. 2.2.1-4 through Eq. 2.2.1-10
  - 2.4 Soil temperature
    - 2.4.1 Eq. 2.2.1-11 & Eq. 2.2.1-12
  - 2.5 Surface temperature
    - 2.5.1 Eq. 2.2.1-13 & Eq. 2.2.1-14
  - 2.6 Soil Temperature
    - 2.6.1 Sidenote: handling recursive equations
    - 2.6.2 Eq. 2.2.1-15 & Eq. 2.2.1-16
  - 2.7 Irrigation
    - 2.7.1 Monthly distribution of irrigation
    - 2.7.2 Eq. 2.2.1-17 & Eq. 2.2.1-18
  - 2.8 Crop Evapotranspiration
    - 2.8.1 Eq. 2.2.1-19 & Eq. 2.2.1-20
  - 2.9 Soil Available Water
    - 2.9.1 Eq. 2.2.1-21 through Eq. 2.2.1-24
  - 2.10 Water Balance
    - 2.10.1 Eq. 2.2.1-25 through Eq. 2.2.1-35
  - 2.11 Decomposition rate - effect of soil temperature
    - 2.11.1 Eq. 2.2.1-36 & Eq. 2.2.1-37
  - 2.12 Decomposition rate - effect of soil moisture
    - 2.12.1 Eq. 2.2.1-38 through Eq. 2.2.1-43
  - 2.13 Climate Factor (re_{crop})
    - 2.13.1 Eq. 2.2.1-44 through Eq. 2.2.1-46
  - 2.14 Tillage Factor (r_c)
  - 2.15 Climate/management Factor (r_e)
    - 2.15.1 Eq. 2.2.1-47

ICBM Re Calculator - QC Testing

Using Ellerslie 1983 site + climate data…

Yield = 2181
Clay = 0.39
Sand = 0.17

- Holos implementation: $\mathbf{r_e = 1.036275}$
- R implementation: $\mathbf{r_e = 1.036275}$



Development of a Generic Calibration framework

1. Spinup
   a. Reach steady state for typical site inputs
   b. Calculate C pool proportions
2. Sensitivity
   a. Global sensitivity analysis
   b. Determine the most important parameters for model calibration
3. Bayesian calibration
   a. Assume uniform prior distribution of model parameters
   b. Re-weight model parameters to favour those that **maximize accuracy** against measured SOC values
   c. Create posterior distribution by sampling from re-weighted parameters
   d. Use posterior to inform parameter value selection, uncertainty analysis

Spinup algorithm

Spinup procedure follows FAO (2020) recommendations:

For most models:

1. Run model for 10,000 years with 1 t/ha C input
2. Run model for 1,000 more years with C inputs large enough to reach the initial C stocks for the site

If model has a steady state solution (e.g. ICBM, IPCC T2):

1. Reach SS for the average of the first 10 years of C inputs to calculate C pool sizes

```
spinup <- function(site_data,
                   climate_data,
                   initial_c,
                   model,
                   ...)
```

Spinup algorithm - IPCC

1. Take site and climate data tables
2. Get average for all **numeric** site columns
3. Get first row for all **non-numeric** site columns
4. Get climate data from first year of input data

	location_name	sand_px	clay_px	tillage	crop	grain_yield_kgha
	<chr>	<dbl>	<dbl>	<chr>	<chr>	<dbl>
1	Ellerslie	17	39	CT	BAR	2181.
2	Ellerslie	17	39	CT	BAR	3370.
3	Ellerslie	17	39	CT	BAR	2971.

Spinup algorithm - IPCC

1. Take site and climate data tables
2. Get average for all **numeric** site columns
3. Get first row for all **non-numeric** site columns
4. Get climate data from first year of input data
5. Get steady-state pool sizes using the average values
6. Calculate **pool size proportions**
7. Convert into C stocks for use in model

	location_name	sand_px	clay_px	tillage	crop	grain_yield_kgha
	<chr>	<dbl>	<dbl>	<chr>	<chr>	<dbl>
1	Ellerslie	17	39	CT	BAR	2181.
2	Ellerslie	17	39	CT	BAR	3370.
3	Ellerslie	17	39	CT	BAR	2971.

```r
spinup_results <- sitedata %>%
    slice(1:10) %>%
    spinup(climate_data = climate_data,
           initial_c = 1000, #g C / m^2
           model = "ipcct2")
```

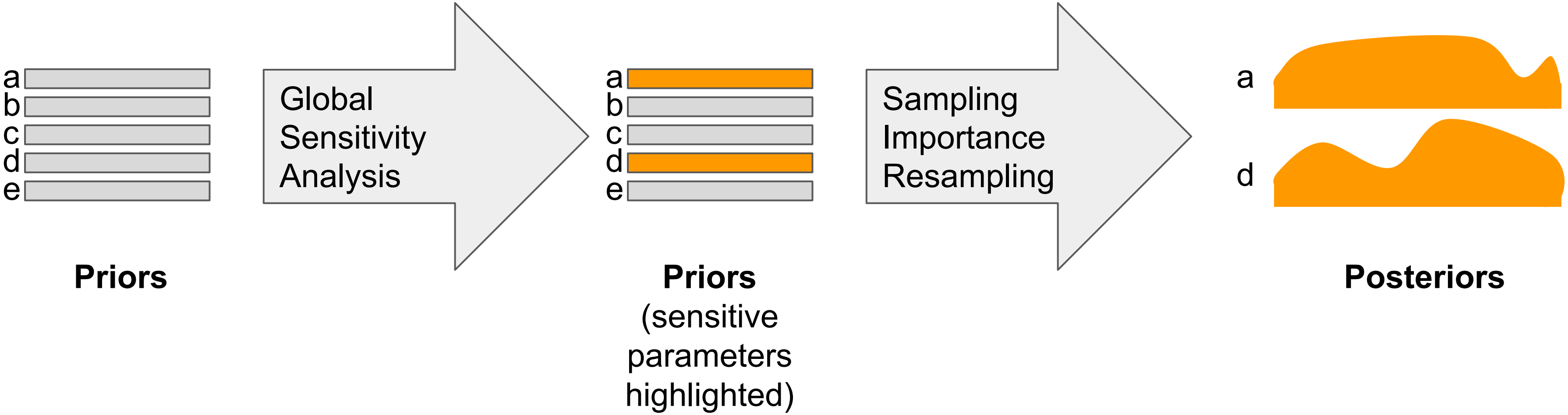
```
$init_active
[1] 6.242881

$init_slow
[1] 59.68814

$init_passive
[1] 934.069
```

Sensitivity & Bayesian calibration

From Gurung et al. (2020): “Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty”



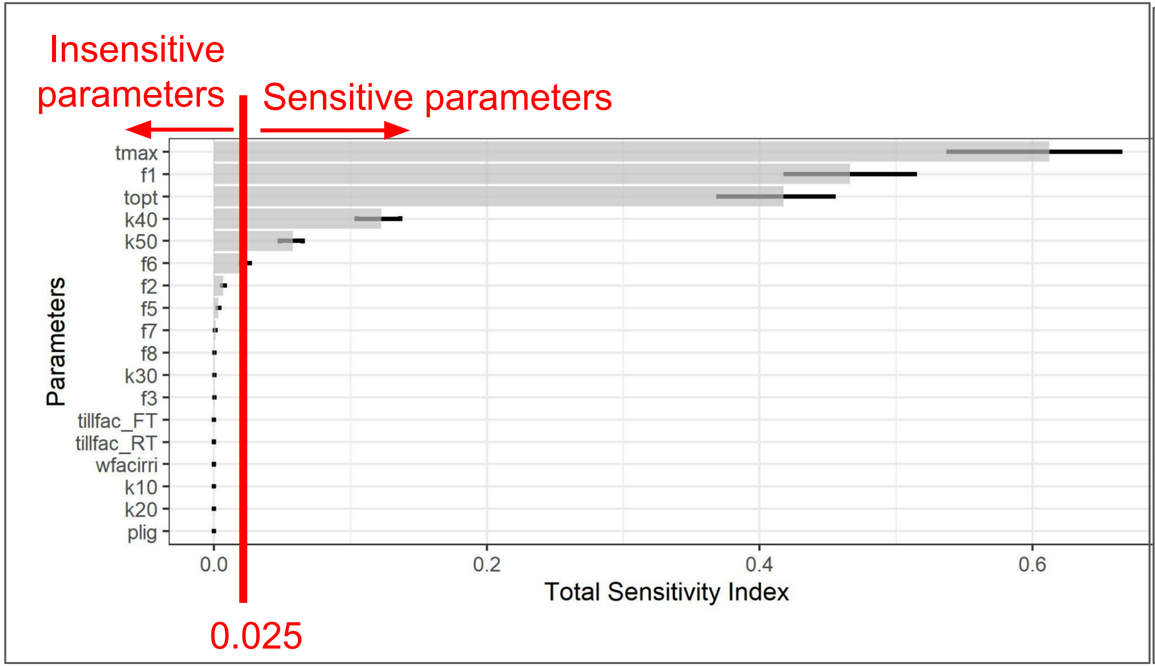
Sensitivity

- Global sens. analysis using either Sobol-Jansen or Fast99 from the `sensitivity` package
    - S-J and Fast99 are similar GSA methods
    - Both are able to compute total parameter effects (main + interactions)
    - Fast99 is faster, but both methods converge at large # of parameters (e.x. ~100)
    - (Saltelli et al., 1999)
- Parallelized using the `parallel` package

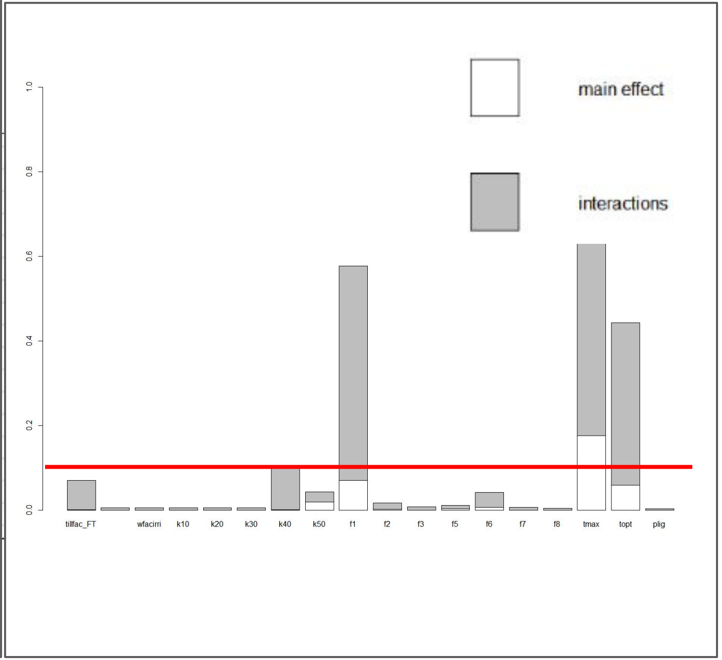
```
gsa <- function(site_data,
                climate_data,
                initial_c,
                parameter_bounds,
                sample_size = 10,
                method = "soboljansen")
```

Sensitivity

Sobol-Jansen

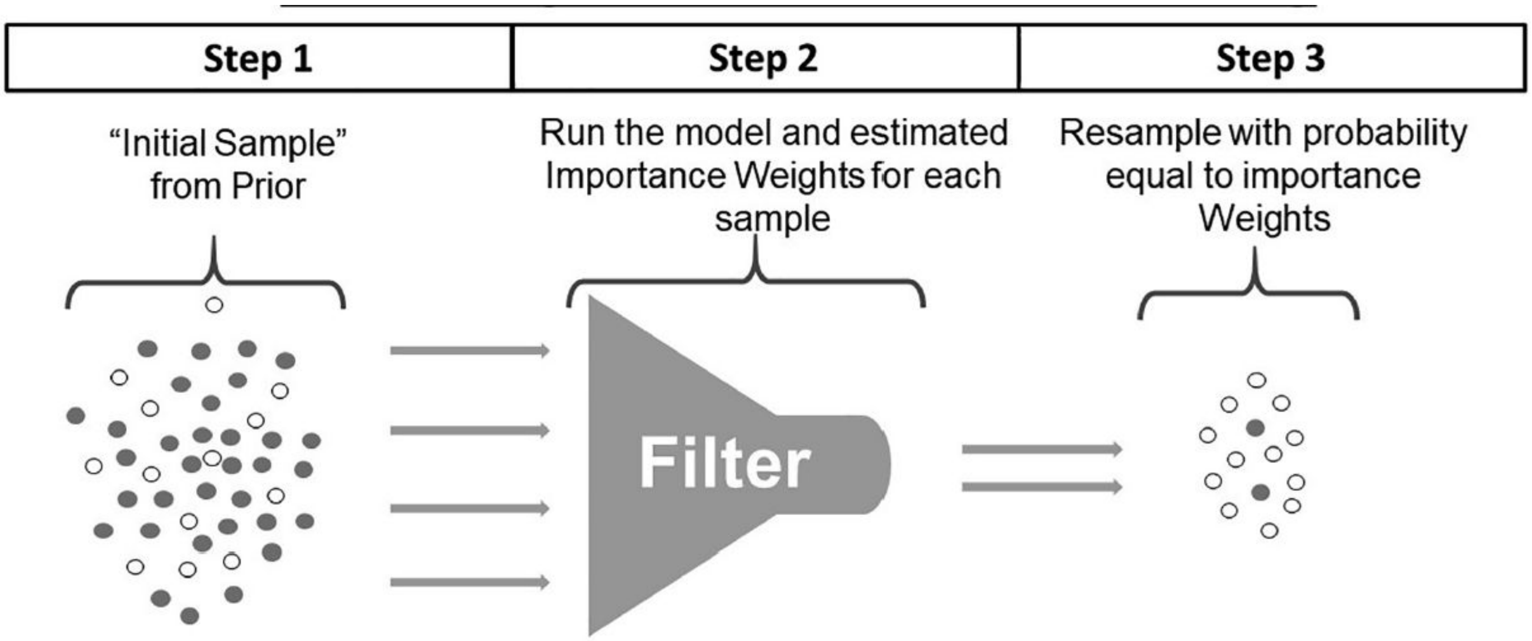


Fast99



Sampling Importance Resampling

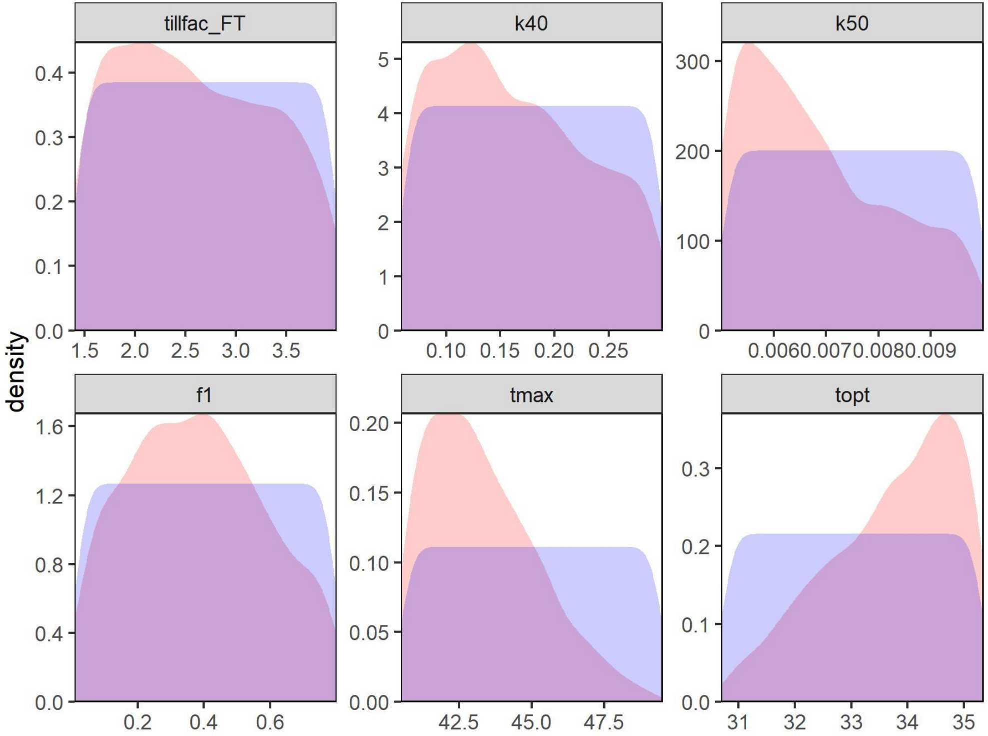
- Latin hypercube sampling
- Parallelized



Prior vs. posterior

1. Sample from **prior**
2. Assign weights using log-likelihood (compare against measured values)
3. Resample from **weighted prior** to get **posterior**

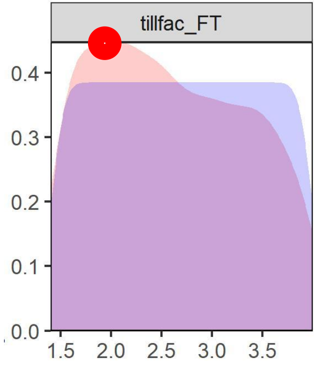
```
sir <- function(site_data,
                climate_data,
                initial_c,
                parameter_bounds,
                sample_size = 100,
                resample_size = 10)
```

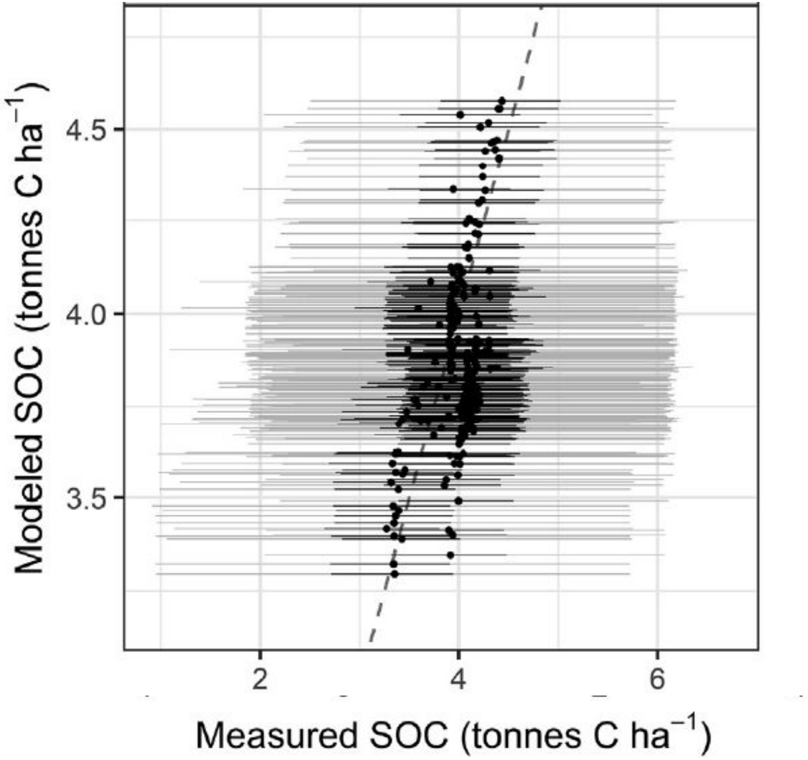


Uncertainty analysis

1. Input distributions of sensitive parameter(s)
   a. Take random samples
   b. Take maximum / median value
2. Run model and compare

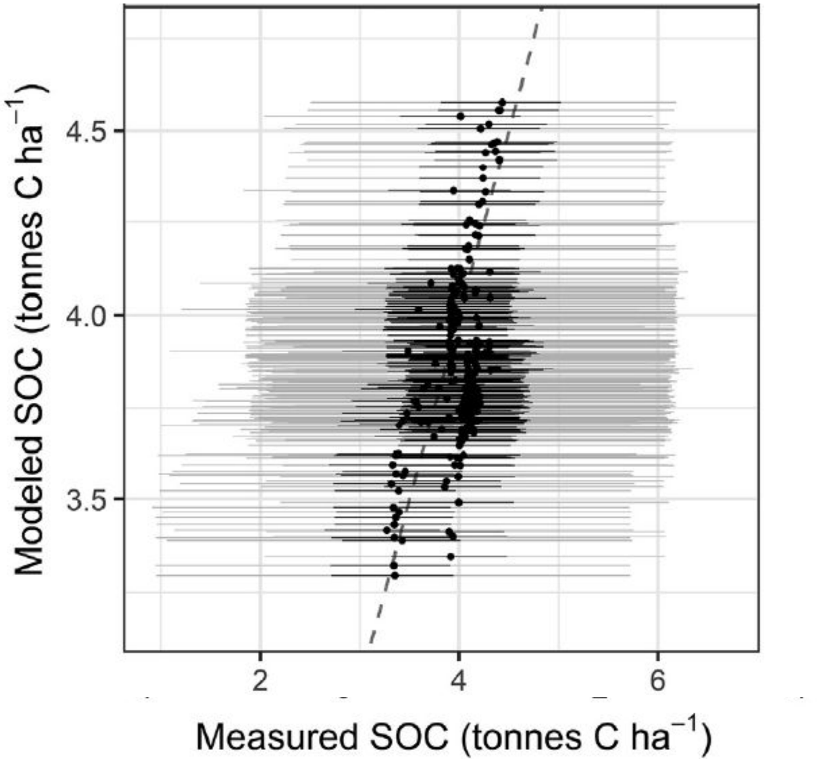
```
montecarlo <- function(site_data,
                       climate_data,
                       initial_c,
                       distribution,
                       sample_size = 10)
```





Uncertainty analysis

1. Input distributions of sensitive parameter(s)
   a. Take random samples
   b. Take maximum / median value
2. Run model and compare



Sources of uncertainty:

- Model parameterization
- Misrepresentation of physical processes in model

Accounted for

- SOC/weather/C input measurement

Not accounted for

Performing analyses and comparing results

1. Spinup
2. Sensitivity
3. Bayesian calibration
4. Validation + uncertainty analysis

- Experiments separated by model
- Standard output / figure format
- Everything required to re-run analysis automatically saved

```
results
└───ipcct2
    └───theme_time_samples
        ├───code
        │       functions.r
        │       main.rmd
        │
        ├───figures
        │       sensitivity.png
        │       sir.png
        │       uncertainty.png
        │       validation.png
        │
        ├───inputs
        │       climate_data.csv
        │       data_sources.txt
        │       parameters.csv
        │       site_data.csv
        │
        └───outputs
                parameters_calibrated.csv
                posterior.csv
                prior.csv
                validation.csv
```

Next steps: generic setup

- Make every step in the experimental setup model-agnostic
- Easier inter-model comparison
- Possible multi-model ensemble approach

```
Lkhood=foreach(i=1:nrow(X),
               .combine = rbind,
               .packages = c("parallel",
                             "doParallel",
                             "tidyverse"),
               .export = c("run_ipcct2",
                           "IPCCTier2SOMmodel",
                           "loglik",
                           "run_ipcct2_calculate_loglik")) %dopar%

             run_ipcct2_calculate_loglik(site_data = site_data[[site_n]
                                         climate_data = climate_data,
                                         i)
```

```
Lkhood=foreach(i=1:nrow(X),
               .combine = rbind,
               .packages = c("parallel",
                             "doParallel",
                             "tidyverse"),
               .export = c("run_chosen_model",
                           "loglik",
                           "run_model_calculate_loglik")) %dopar%

             run_model_calculate_loglik(site_data = site_data[[site_n]],
                                        climate_data = climate_data,
                                        i)
```

References

FAO, 2020. Technical Specifications and Country Guidelines for Global Soil Organic Carbon Sequestration Potential Map.

Gurung, R., Ogle, S. M., Breidt, F. J., Williams, S. B., & Parton, W. J. (2020). Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. Geoderma, 376, 114529. https://doi.org/10.1016/j.geoderma.2020.114529

Saltelli, A., Tarantola, S., & P.-S. Chan, K. (1999) A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output, Technometrics, 41:1, 39-56, DOI: 10.1080/00401706.1999.10485594