Supplementary Information "Hybrid data-model driven prediction of soil thickness in a mountainous watershed"

Qina Yan[1], Haruko Wainwright[1], Baptiste Dafflon[1], Sebastian Uhlemann[1], Carl I. Steefel[1], Nicola Falco[1], Jeffrey Kwang[2], Susan S. Hubbard[1]

[1]Earth and Environmental Science Area, Lawrence Berkeley National Laboratory, Berkeley, CA., USA.
[2] Department of Geosciences, University of Massachusetts Amherst, Amherst, MA, USA.

Correspondence to: Qina Yan (qinayan@lbl.gov); Haruko Wainwright (HMWainwright@lbl.gov)

Posterior distribution of the parameters

We used the Bayesian method to calculate the posterior distribution of the parameters in this hybrid-model (Eqn. 6-8) and estimated the maximum a posteriori (MAP).

The parameter set is defined as a vector θ, including six parameters (except for K_s in Eqn. 3b). We define the prior distribution for θ, $p(\boldsymbol{\theta})$, which is the products of independent uniform distributions for the parameters. We define the real soil thickness data vector $\mathbf{z}$ (which is a m-vector with m field sampling points: $\{z_1, z_2, \ldots, z_m\}$) as the sum of predicted soil thickness and error vector $\boldsymbol{\varepsilon}$:

$$\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{\varepsilon} \tag{S1}$$

where $\boldsymbol{y}$ is the predicted soil thickness based on the parameter sets:

$$\boldsymbol{y} = f(\boldsymbol{\theta}). \tag{S2}$$

We assume that each element in the error vector follows a normal distribution with a standard deviation σ. We estimate the standard deviation of the measurement error based on the discrepancy between the auger and CPT measurements.

Using the Bayes' rule, we define the posterior distribution as:

$$p(\boldsymbol{\theta}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{S3}$$

Since the error vector is normal, the likelihood $p(\boldsymbol{z}|\boldsymbol{\theta})$ is a normal distribution.

To compute the posterior distribution, we use the sampling-resampling scheme (Smith and Gelfand, 1992). We sample the parameter set at each grid following the grid search $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots., \boldsymbol{\theta}^{(N)}\}$, where $N = 6^6 = 46656$ and predict the soil thickness at the measurement locations $\{\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots., \boldsymbol{y}^{(N)}\}$ where $\boldsymbol{y}^{(i)} = f(\boldsymbol{\theta}^{(i)})$

$$p(\boldsymbol{\theta} = \boldsymbol{\theta}^{(i)}|\boldsymbol{z}) = \frac{l^{(i)}}{\sum_{j=1}^{N} l^{(i)}} \tag{S4}$$

$p(\boldsymbol{\theta}|\boldsymbol{z})$= distribution

$$\boldsymbol{y}^{(1)} = \{y_1^{(1)}, \ldots, y_m^{(1)}\}$$

Where the likelihood $l^{(i)}$ is defined as the normal distribution with the mean $\boldsymbol{y}^{(i)}$ and the standard deviation σ as:

$$l^{(i)} = N\left(\boldsymbol{z} - \boldsymbol{y}^{(i)}, \sigma\right)$$

$$l^{(i)} = N\left(z_{\boldsymbol{1}} - y_1^{(i)}, \sigma\right) \ldots\ldots N\left(z_{\boldsymbol{m}} - y_m^{(i)}, \sigma\right) \quad \text{(S5)}$$

Since we assume that each data point ($j = 1 \ldots . m$) is independent, we can represent the likelihood as the product of normal distributions of individual measurements.

$$l^{(i)} = C \prod_{j=1}^{m} \exp\left[-\frac{\left(z_{\boldsymbol{j}} - y_j^{(i)}\right)^2}{2\sigma^2}\right] = C \exp\left[\sum_{j=1}^{m} -\frac{\left(z_{\boldsymbol{j}} - y_j^{(i)}\right)^2}{2\sigma^2}\right]$$

$$\hat{l}^{(i)} = argmax\left\{C \exp\left[\sum_{j=1}^{m} -\frac{\left(z_{\boldsymbol{j}} - y_j^{(i)}\right)^2}{2\sigma^2}\right]\right\} \quad \text{(S6)}$$

Note that C is the normalization factor. The standard deviation of the error is equal to the standard deviation of the difference. We then obtained the MAP estimate as the parameter set that provides the maximum posterior probability. The marginal distribution of each parameter can be calculated by the summation of the posterior distributions given each fixed parameter value (Fig. S8).

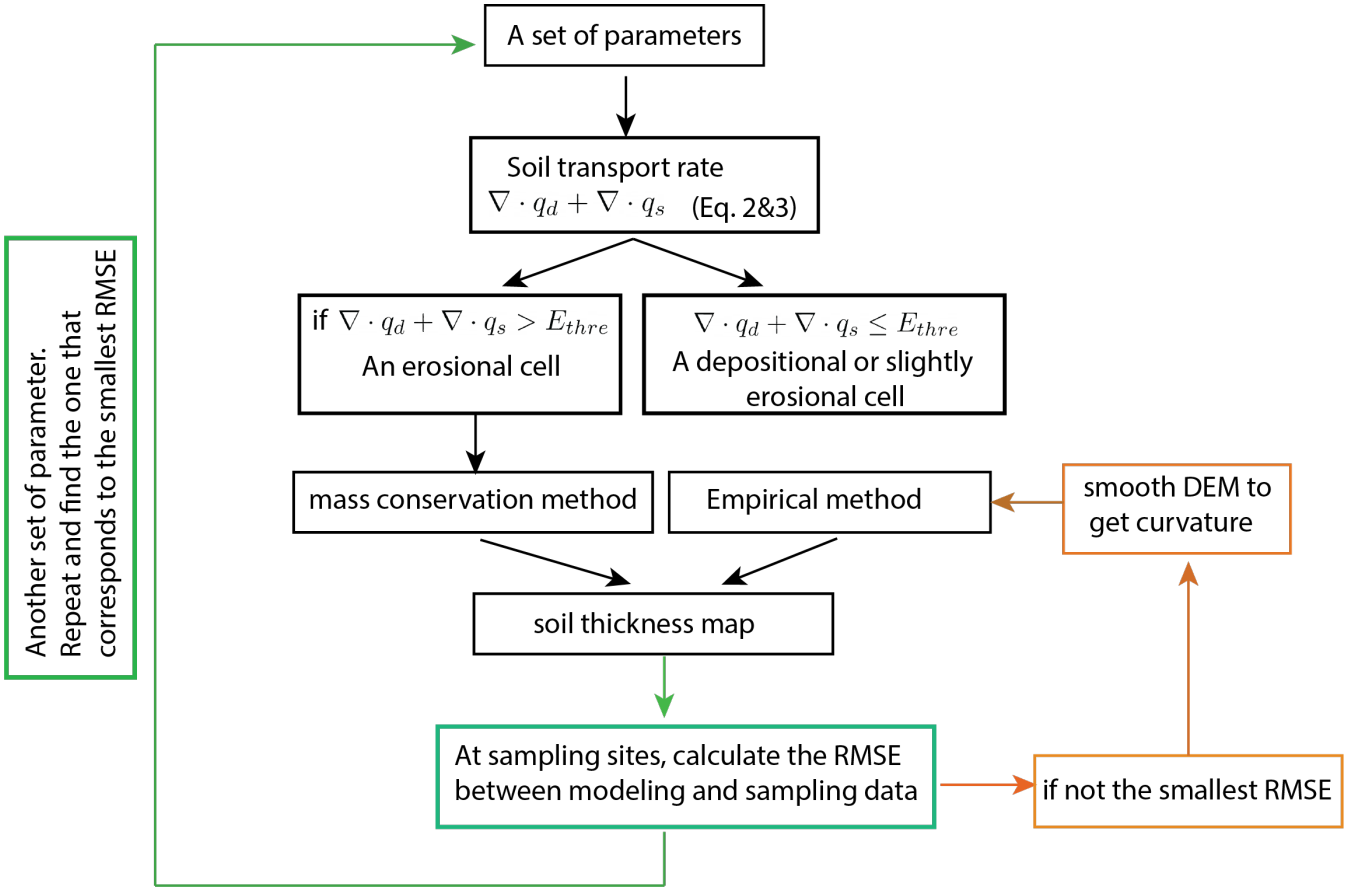

Figure S1: The workflow of the hybrid method.

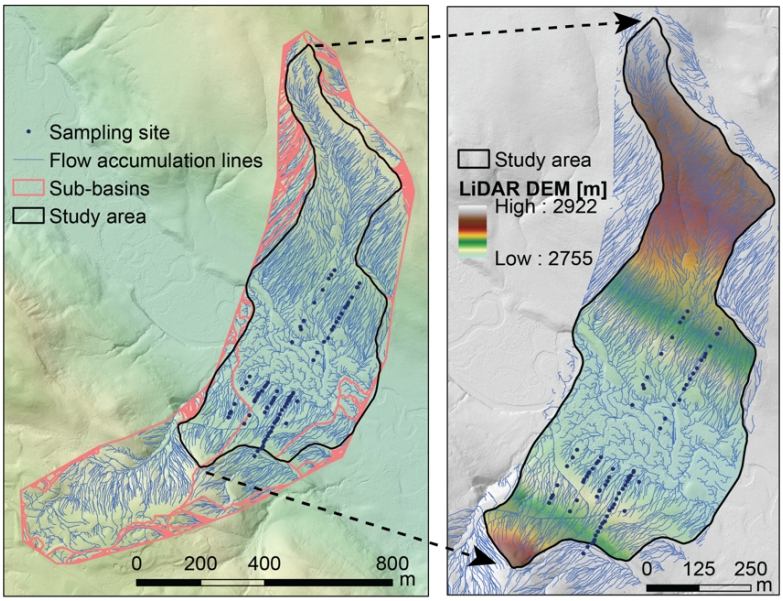

Figure S2: Drainage area delineation. The drainage areas are created by locating the pour points at the edges of the analysis window (where water would pour out of the raster), as well as sinks, then identifying the contributing area above each pour point.

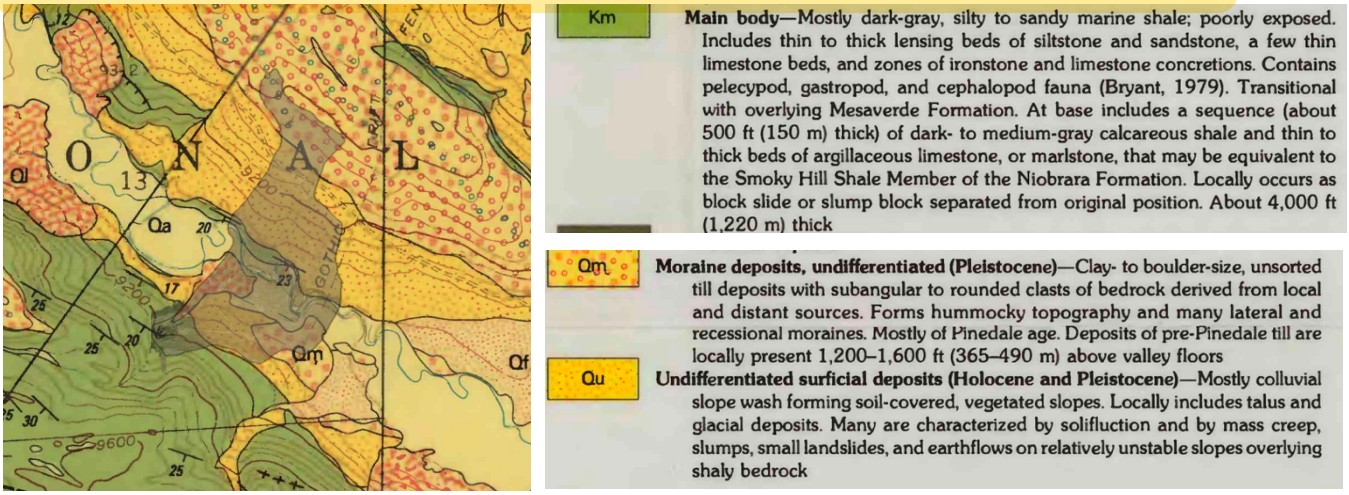

Figure S3: A geological map of parent materials and deposits.

Figure S4: Field images of saprolite layers from auger sampling.

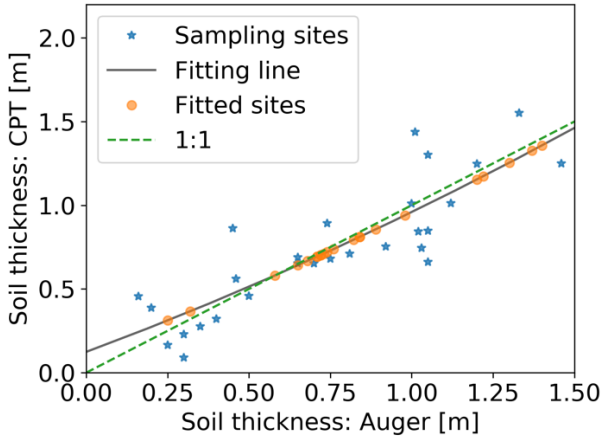

Figure S5: Fitting the soil thickness of the CPT data using the auger data. Correlation = 0.86, Root-minimum-square-error = 0.20 m.

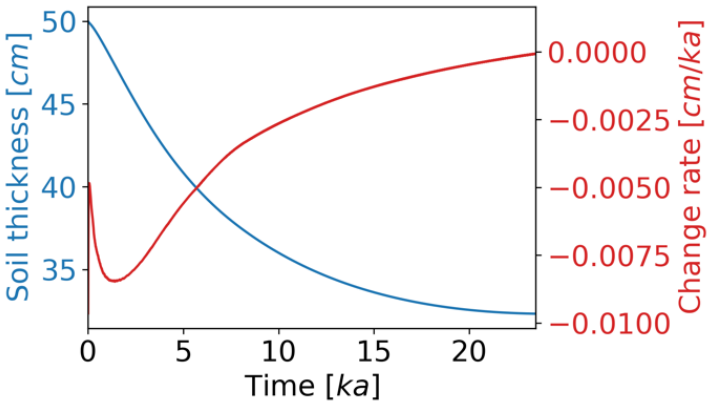

Figure S6: Spatial mean values of erosional sites of soil thickness evolution over time. The initial soil thickness is 0.5 m, time step is 1 year, and the initial elevation is the current DEM data. The boundary condition is Neumann boundary condition, the surface transport fluxes around the edge is zero. The time step is 1 yr, and the diffusion coefficient is 1.1×10^{-3} m^2/yr for the north-facing hillslope and 1.8×10^{-3} m^2/yr for the south-facing hillslope.

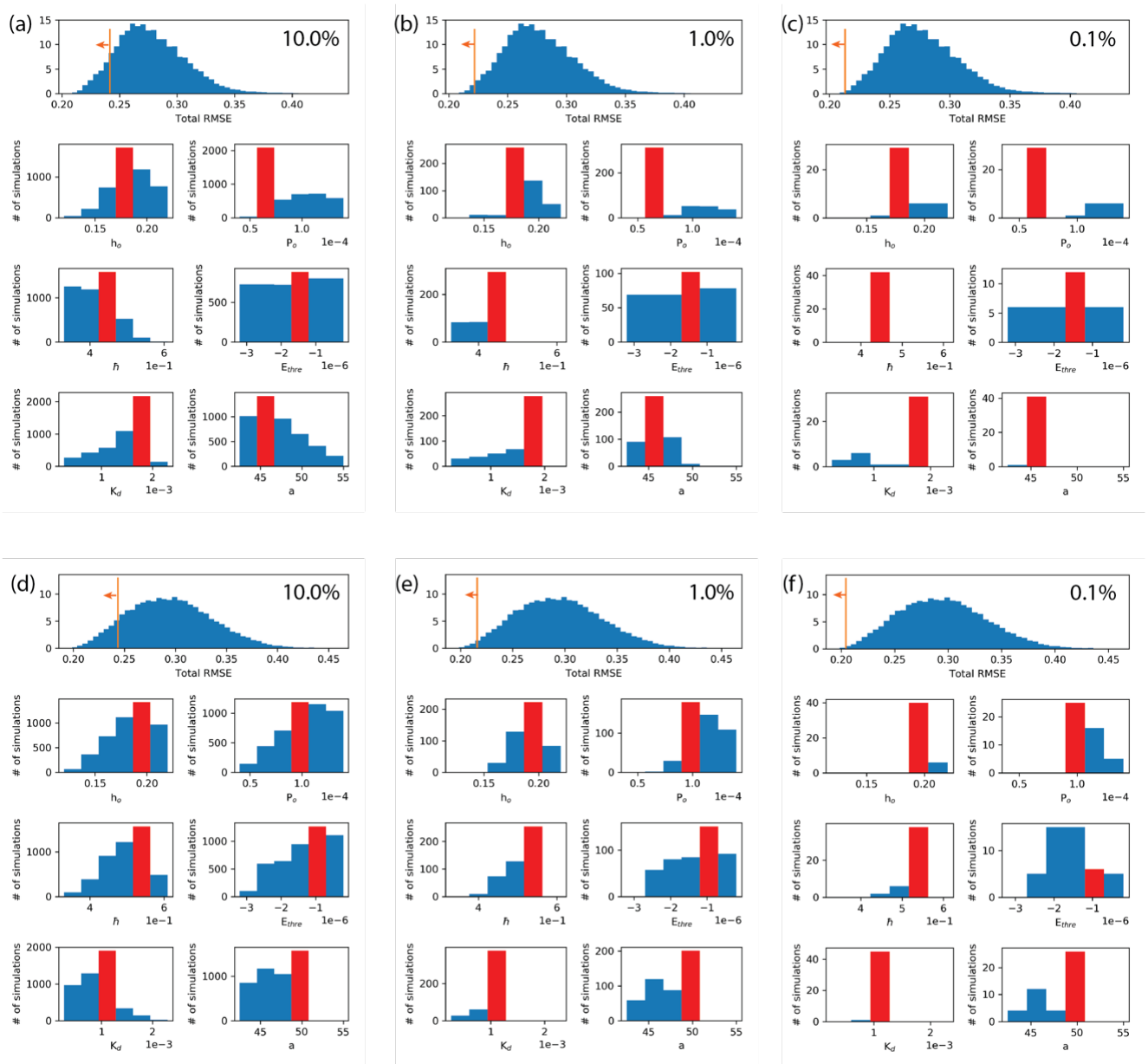

Figure S7: Probability density function and histogram plots from a series of simulations of a grid search. The root-mean-square error (RMSE) between the simulated and measured soil thickness. (a-c) The RMSE for the south-facing hillslope and the corresponding histogram plots which show the distribution of each parameter that corresponds to 10.0%, 1.0%, and 0.1% of the smallest RMSE values, respectively; (d-f) The RMSE of north-facing hillslope and the corresponding histogram plots which show the distribution of each parameter that corresponds to 10.0%, 1.0%, and 0.1% of the smallest RMSE values, respectively. The red color bar represents the parameter that provides the global minimum between simulation and field measurement.

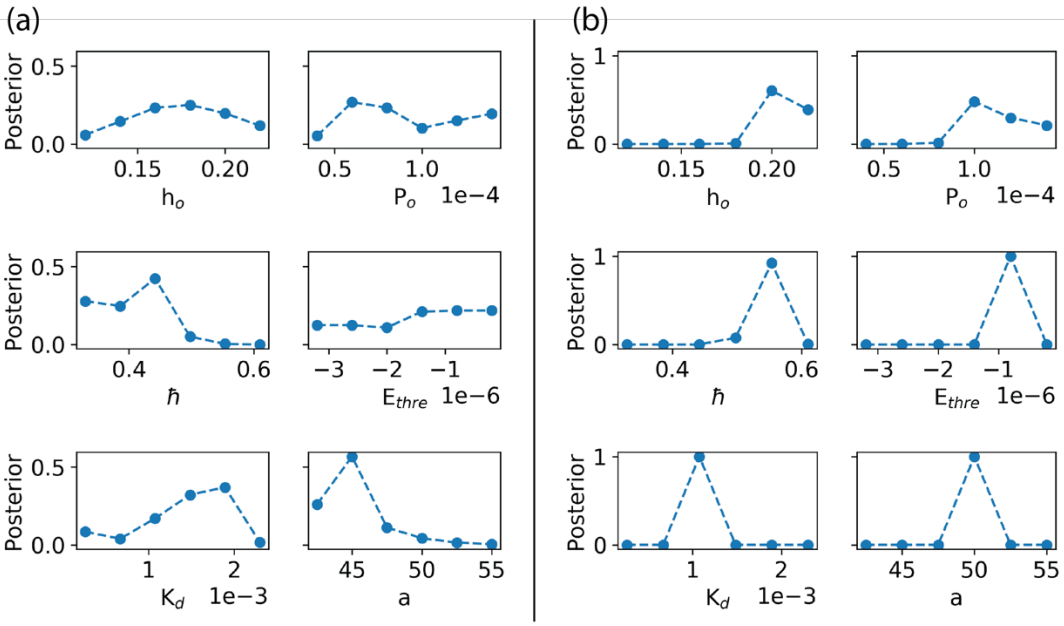

Figure S8: Marginal posterior probabilities of six parameters. (a) The posterior probabilities of the six parameters for the south-facing hillslope; (b) The posterior probabilities of the six parameters for the north-facing hillslope.

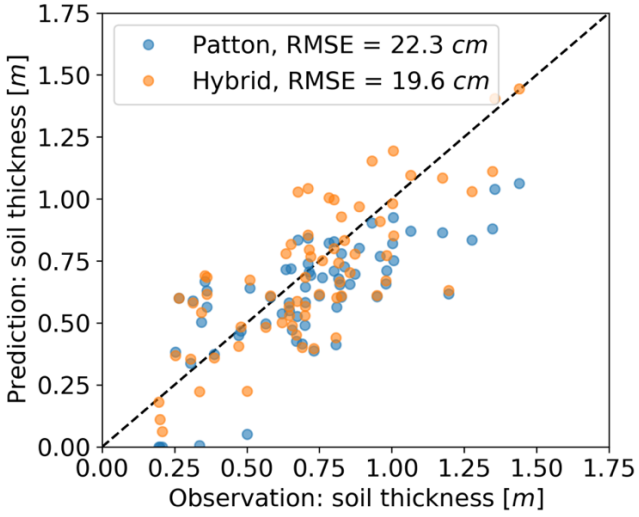

Figure S9: Parameter calibrations for the hybrid model. The root-mean-square error (RMSE) and Pearson's correlation between sampling and simulation results are calculated for the south-facing (a) and north-facing (b), respectively.

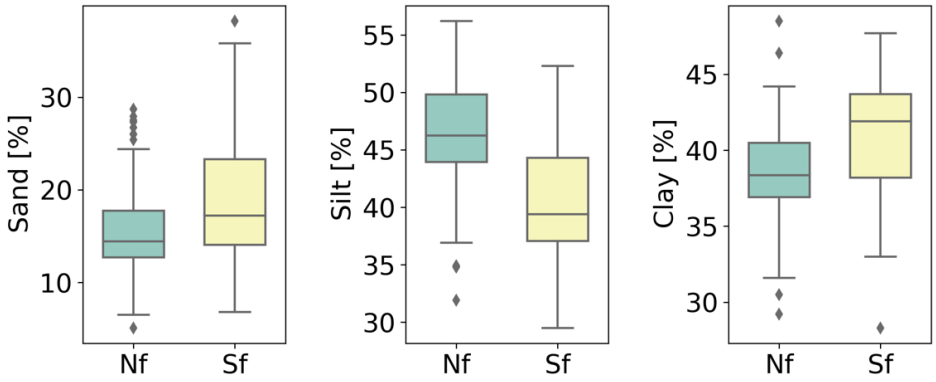

Figure S10: Soil texture at two hillslopes. Nf = North-facing hillslope, Sf = South-facing hillslope.

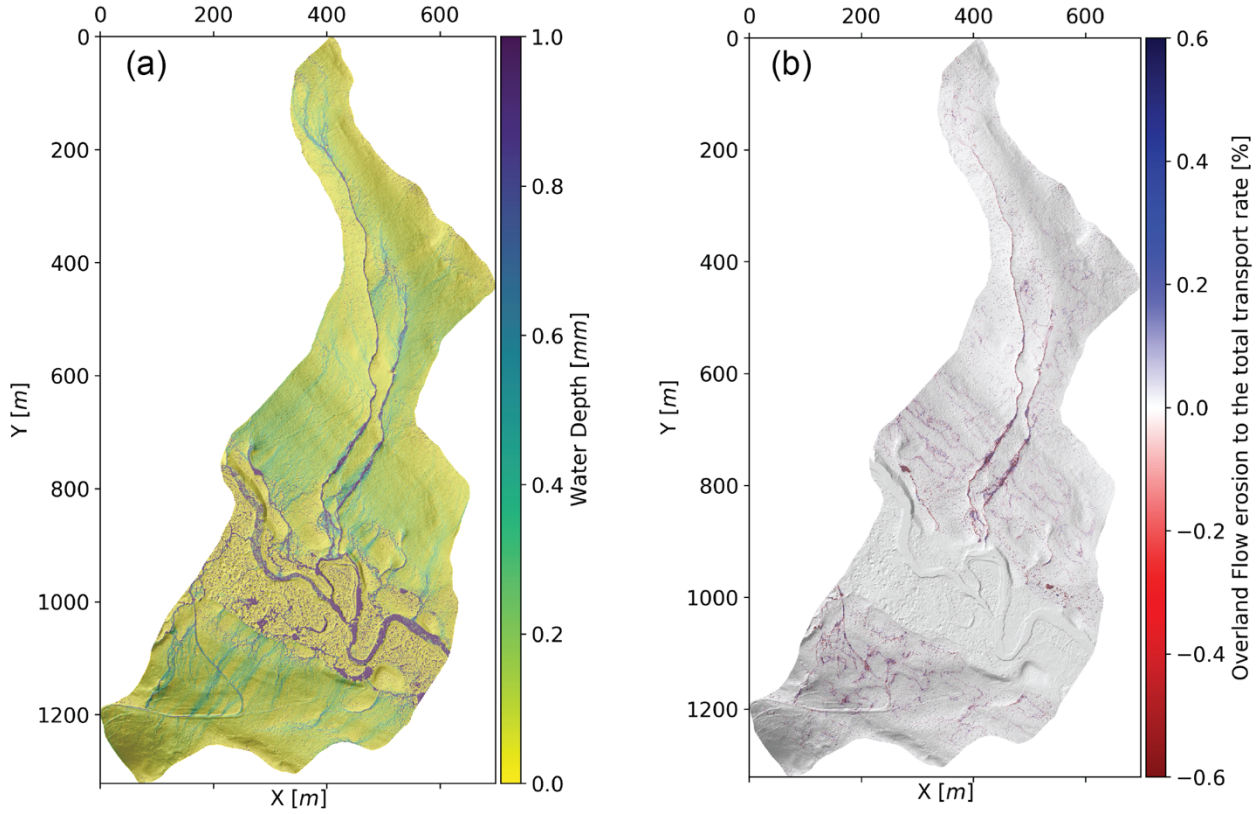

Figure S11: (a) water depth of overland flow at a steady-state, which occurs after 6 days with constant rainfall = 363 mm/yr, and the time step $\Delta t \approx 1sec$; (b) The rate of the overland flow erosion rate to the total soil transport rate. The overland flow erosion mechanism is from Equation 3.

Table S1. A list of topographic variables for the correlation with soil thickness:

Variable names	Explanation
Lidar_2018_Aspect_NEON	Topographical aspect computed at the original pixel resolution of 1m
Aspect_NEON_10m	Topographical aspect computed at a pixel resolution of 10m
Aspect_NEON_10m_x3	Topographical aspect computed at a pixel resolution of 10m and smoothed considering a 3x3-pixel moving window
Aspect_NEON_10m_x5	Topographical aspect computed at a pixel resolution of 10m and smoothed considering a 5x5-pixel moving window
Aspect_NEON_10m_x9	Topographical aspect computed at a pixel resolution of 10m and smoothed considering a 9x9-pixel moving window
C_unc	Uncertainty of the estimated leaf carbon content derived by airborne hyperspectral data (pixel resolution of 1 m)
C	Leaf carbon content derived by airborne hyperspectral data (pixel resolution of 1 m)
N_unc	Uncertainty of the estimated leaf nitrogen content derived by airborne hyperspectral data (pixel resolution of 1 m)
N	Leaf nitrogen content derived by airborne hyperspectral data (pixel resolution of 1 m)
CN_unc	Uncertainty of the estimated leaf carbon/nitrogen ratio derived by airborne hyperspectral data (pixel resolution of 1 m)
CN	Leaf carbon/nitrogen derived by airborne hyperspectral data (pixel resolution of 1 m)
chm_mosaic	Canopy height model (i.e., plant height) at 1 m pixel resolution
LMA	Leaf mass area derived by airborne hyperspectral data (pixel resolution of 1 m)
LMA_unc	Uncertainty of the estimated leaf mass area derived by airborne hyperspectral data (pixel resolution of 1 m)
LWC	Leaf water content derived by airborne hyperspectral data (pixel resolution of 1 m)
Curv_NEON_10m	Topographical curvature computed at a pixel resolution or 10m
Curv_NEON_10m_x3	Topographical curvature computed at a pixel resolution or 10m and smoothed considering a 3x3-pixel moving window
Curv_NEON_10m_x5	Topographical curvature computed at a pixel resolution or 10m and smoothed considering a 5x5-pixel moving window
Curv_NEON_10m_x9	Topographical curvature computed at a pixel resolution or 10m and smoothed considering a 9x9-pixel moving window
dsm_mosaic	Digital surface model at 1 m resolution
dsm_mosaic_10m	Digital surface model at 10 m resolution
dtm_mosaic	Digital terrain model at 1 m resolution

dtm_mosaic_10m	Digital terrain model at 10 m resolution
FlowAcc_NEON	Topographical flow accumulation computed at the original pixel resolution of 1m
FlowAcc_NEON_10m	Topographical flow accumulation computed at a pixel resolution of 10m
FlowAcc_NEON_10m_x3	Topographical flow accumulation computed at a pixel resolution of 10m and smoothed considering a 3x3-pixel moving window
FlowAcc_NEON_10m_x5	Topographical flow accumulation computed at a pixel resolution of 10m and smoothed considering a 5x5-pixel moving window
FlowAcc_NEON_10m_x9	Topographical flow accumulation computed at a pixel resolution of 10m and smoothed considering a 9x9-pixel moving window
NEON_mosaic_NDNI	Normalized difference nitrogen index computed from the airborne hyperspectral data
NEON_mosaic_NDVI	Normalized difference vegetation index computed from the airborne hyperspectral data
NEON_mosaic_NDWI	Normalized difference water index computed from the airborne hyperspectral data
NEON_mosaic_liq_water	Canopy water content estimated from the airborne hyperspectral data
SlopeDeg_NEON	Topographical aspect computed at the original pixel resolution of 1m
SlopeDeg_NEON_x3	Topographical slope in degree computed at a pixel resolution of 1m and smoothed considering a 3x3-pixel moving window
SlopeDeg_NEON_x9	Topographical slope in degree computed at a pixel resolution of 1m and smoothed considering a 9x9-pixel moving window
SlopeDeg_NEON_10m	Topographical slope in degree computed at a pixel resolution of 10m
SlopeDeg_NEON_10m_x3	Topographical slope in degree computed at a pixel resolution of 10m and smoothed considering a 3x3-pixel moving window
SlopeDeg_NEON_10m_x5	Topographical slope in degree computed at a pixel resolution of 10m and smoothed considering a 5x5-pixel moving window
SlopeDeg_NEON_10m_x9	Topographical slope in degree computed at a pixel resolution of 10m and smoothed considering a 9x9-pixel moving window
Srad_10m	Solar radiation computed at a pixel resolution of 10m.
TPI_NEON_10m_x3	Topographical position index computed at a pixel resolution of 10m and smoothed considering a 3x3-pixel moving window
TPI_NEON_10m_x5	Topographical position index computed at a pixel resolution of 10m and smoothed considering a 5x5-pixel moving window
TPI_NEON_10m_x9	Topographical position index computed at a pixel resolution of 10m and smoothed considering a 9x9-pixel moving window
TWI_NEON_10m	Topographical wetness index computed at a pixel resolution of 10m
TWI_NEON_10m_x3	Topographical wetness index computed at a pixel resolution of 10m and smoothed considering a 3x3-pixel moving window

TWI_NEON_10m_x5	Topographical wetness index computed at a pixel resolution of 10m and smoothed considering a 5x5-pixel moving window
TWI_NEON_10m_x9	Topographical wetness index computed at a pixel resolution of 10m and smoothed considering a 9x9-pixel moving window
UPslope_NEON_10m	local upslope drainage area computed at a pixel resolution of 10m
UPslope_NEON_10m_x3	local upslope drainage area computed at a pixel resolution of 10m and smoothed considering a 3x3-pixel moving window
UPslope_NEON_10m_x5	local upslope drainage area computed at a pixel resolution of 10m and smoothed considering a 5x5-pixel moving window
UPslope_NEON_10m_x9	local upslope drainage area computed at a pixel resolution of 10m and smoothed considering a 9x9-pixel moving window

References

A. F. M. Smith and A. E. Gelfand, Bayesian statistics without tears: a sampling-resampling perspective, Am. Statistician, 46 (1992), pp. 84–88.