Supporting Information for

Single- and Multi-Objective Parameter Optimization of Common Land Model based on the revised sequential Monte Carlo algorithm

Cong Xu[1], Gaofeng Zhu[1], Yang Zhang[1], Kun Zhang[2]

[1]College of Earth and Environmental sciences, Lanzhou University, Lanzhou 730000, China
[2]School of Biological Sciences and Institute for Climate and Carbon Neutrality, the University of Hong Kong, Hong Kong SAR, China

Corresponding author: Gaofeng Zhu (zhugf@lzu.edu.cn)

Contents of this file

S1 Evaluating the effectiveness and efficiency of the revised PEM-SMC algorithm

Here, we couducted two case studies to compare the differences in optimization effectiveness and efficiency between the original PEM-SMC algorithm and its revised version.

S1.1 Case study 1: synthetic 5-dimensional bimodal normal distribution

To emulate the common feature of the parameter surface with multiple modes in Land Surface Models (LSMs), the first case study involves a synthetic multi-dimensional bimodal normal distribution:

$$\pi(\theta) = \frac{1}{3}N_d(\theta; -\mathbf{5}, 1) + \frac{2}{3}N_d(\theta; \mathbf{5}, 1) \quad (1)$$

where θ denotes a five-dimensional parameter vector, with each parameter constrained with a feasible range of -10 to 10. The values -5 and 5 correspond to the true parameter values at the two peaks of the distribution function. The 1 represents the covariance of the tow normal distributions.

Regarding optimization efficiency, for the same number of particles(Np=8000) and evolutions(S=1000), the original PEM-SMC algorithm required 87 seconds to run, whereas the modified PEM-SMC algorithm only took 53 seconds. This 40% reduction in run time is particularly significant for complex models with lengthy single-run times. In terms of optimization effectiveness, the enhanced PEM-SMC algorithm demonstrated no significant differences from the original in aspects such as particle diversity, evolutionary progression, and final particle distribution (Figure S1). When considering the optimized values of the five-dimensional parameters, both the improved and original PEM-SMC algorithms consistently achieved the posterior marginal distributions of the parameters, successfully identifying the true values for each parameter (Figure S2). Thus, the enhanced PEM-SMC algorithm substantially reduces run time while maintaining the original's optimization performance.

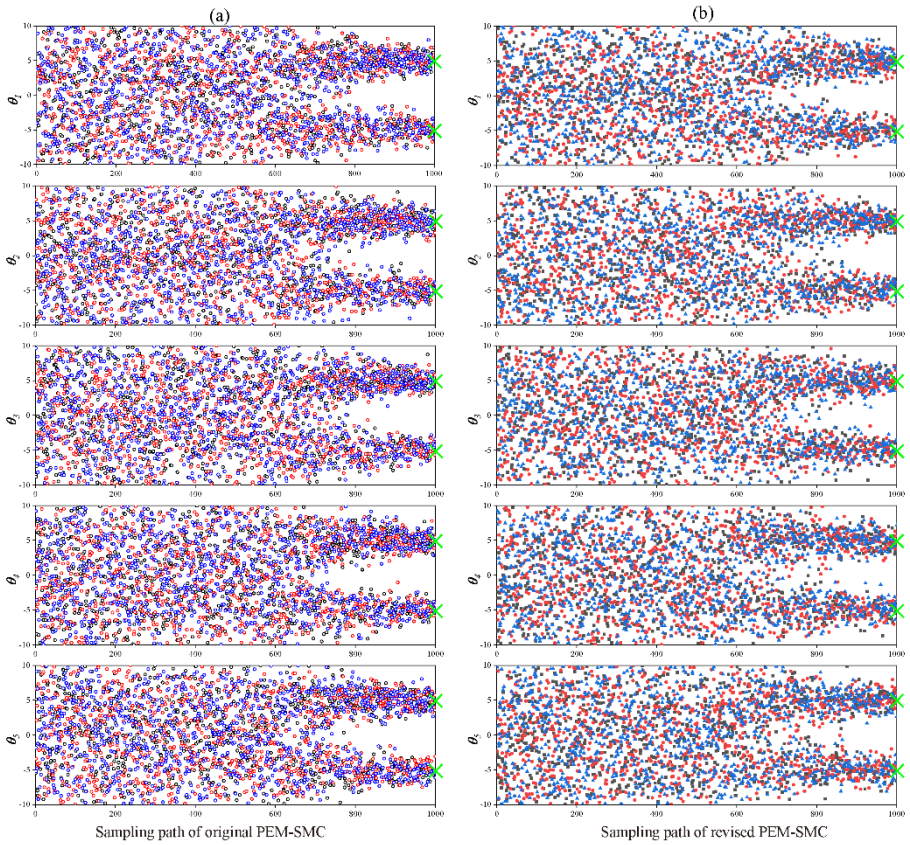

Figure S1. Evolutionary of the parameter in three randomly selected particles along the sampling path toward a five-dimension bimodal posterior distribution using (a) original PEM-SMC and (b) revised PEM-SMC samplers. The true target parameter values, -5 and +5, are denoted by '×' at the right hand side. Each particle is uniquely identified by distinct symbols and colors. Despite the graphical depiction being limited to three particles, the randomness inherent in particle selection suggests that the convergence patterns of additional particles would yield a similar profile to the one presented.

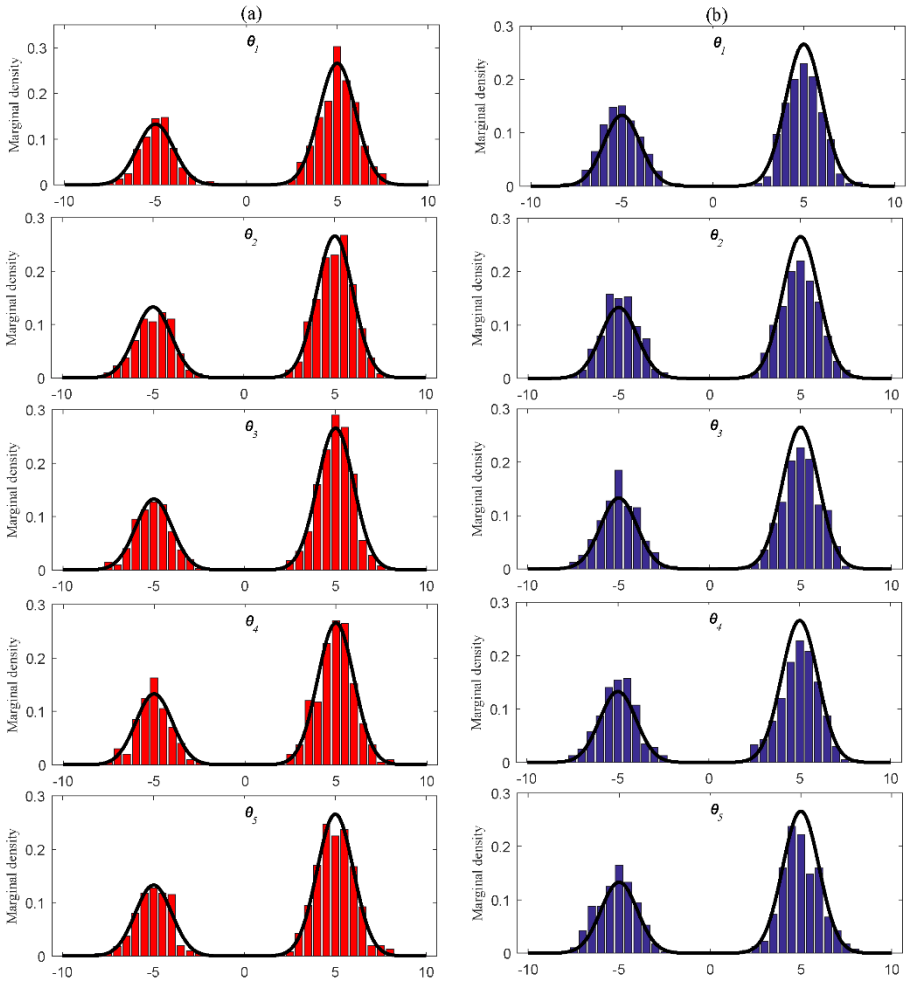

Figure S2. The posterior marginal distributions of the five-dimension parameters ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$) as generated by (a) original PEM-SMC and (b) revised PEM-SMC samplers. The solid black line depcits the target distribution.

S1.2 Case study 2: a benchmark experiment

To ensure the accuracy of the subsequent optimization results, we performed a benchmark experiment on the PEM-SMC sampler algorithm. For each optimization scenario, we generated a simulated LE and NEE output dataset using the default values of the selected parameters and treated the simulated datasets as "observation datasets" to calibrate the selected parameters. By comparing the optimized values with the default values of the parameters, we can then determine whether the selected parameters and the PEM-SMC algorithm are adequate for the parameter calibration of CoLM. The particle sampling evolution results of selected parameters in three optimization scenarios are shown in Figure S3-S5. It is worth noting that although we have graphically displayed the evolution for only five particles, the randomness of particle selection guarantees that the graphical examination of the convergence of the algorithms for other particles would yield a similar picture. We can see that all the particles converge to a limiting distribution of the prior range at the last evolutionary iteration, and the median values of the posterior distributions of all parameters optimized by the PEM-SMC sampler match well with the actual parameter values in both single-objective and multi-

objective optimization scenarios. Therefore, we believe that the PEM-SMC algorithm is highly stable and robust in searching for the most appropriate model parameter combinations under the constraints of the observed datasets.

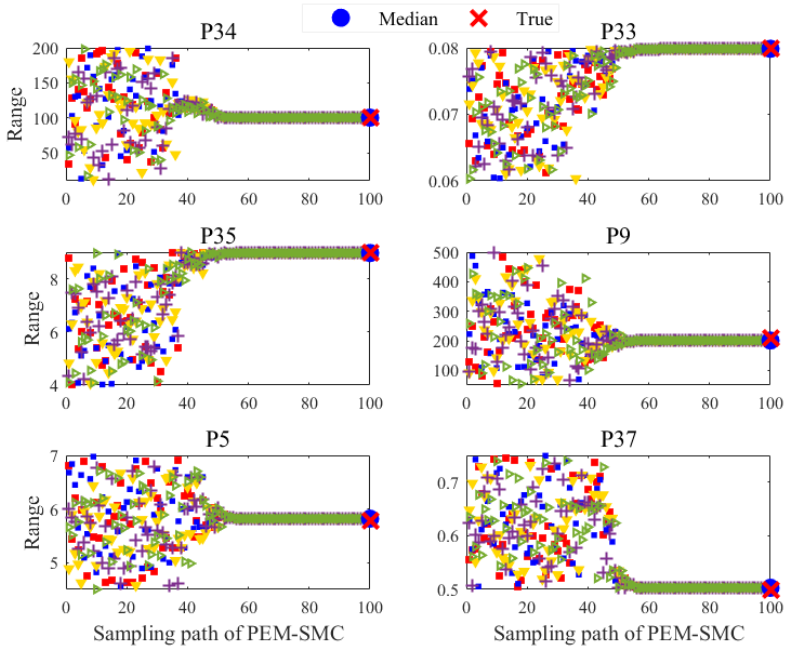

Figure S3. Transitions of the sampled parameter values in five randomly selected particles during the sampling path in the optimization scenario of Opt_ALL. The vertical axis denotes the prior range of the parameter and the horizontal axis denotes the number of evolutions. The cross symbol at the right-hand side of each plot indicates the actual parameter values and the solid circles are the median value of the posterior distribution of the parameters.

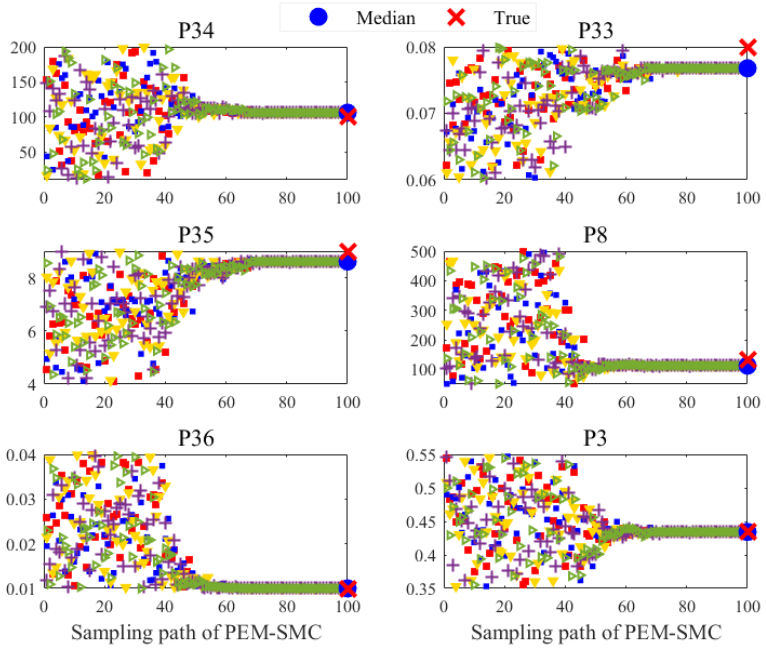

Figure S4. The same as in Figure S3 but for the selected parameters in the optimization scenario of Opt_LE.

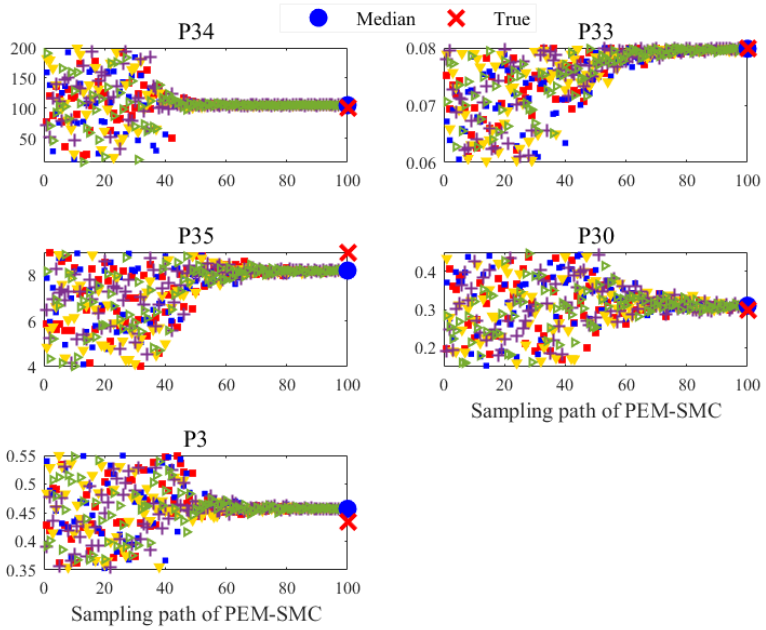

Figure S5. The same as in Figure S3 but for the selected parameters in the optimization scenario of Opt_NEE.

S2 The configuration of the tunable parameters

The forty chosen static parameters of the CoLM are delineated in Table S1, with their prior ranges determined as follows: For the parameters depending on the vegetation and soil type, such as the light reflection and transmission coefficients of the canopy (P20-P27), vegetation biochemical parameters (P19, P29, P31, P35, P36), and some soil physical parameters that empirically produced based on the soil texture characteristics (P2-P9), we defined the range as the lower and upper bounds among all the forest types in the 24-category(USGS) vegetation dataset and all the 17-category soil types, respectively. For some other parameters that cannot be calculated explicitly, we determined their ranges in the evergreen needleleaf forest ecosystem based on their physical meaning and commonly recognized conclusions. For example, the aerodynamic roughness length (Z_{0m}) is empirically estimated as being about 1/10 of the height of the roughness elements (Wallace et al., 2006). Therefore, we define the range of this parameter as 2-3 according to the approximate tree height (20-30m)of the needleleaf forest. In addition, the default value of the quantum efficiency of vegetation photosynthesis("effcon") is 0.08 in all forest ecosystems. To make it variable and have basic biochemical significance, we set its range to 0.06-0.08 according to Landsberg et al, (2011). For some parameters that are defined as constants (P28, P30, P32), we add and subtract half of their values based on the original values as their variable ranges.

Table S1. Adjustable parameters for sensitivity analysis and optimization in CoLM.

Para.	Parameter	Physical meaning	Category	Min	Max

P1	zlnd	roughness length for soil surface	soil	0.005	0.015
P2	θ_{sat-s}	surface soil porosity	soil	0.35	0.55
P3	θ_{sat-d}	deep soil porosity	soil	0.35	0.55
P4	B_s	Clappp and Hornberger “b” parameter	soil	4.5	7
P5	B_d	Clappp and Hornberger “b” parameter	soil	4.5	7
P6	K_{sat-s}	saturated hydraulic conductivity of surface soil	soil	0.001	0.176
P7	K_{sat-d}	saturated hydraulic conductivity of deep soil	soil	0.001	0.176
P8	ψ_s	minimum soil suction of surface soil	soil	50	500
P9	ψ_d	minimum soil suction of deep soil	soil	50	500
P10	wtfact	fraction of shallow groundwater area	soil	0.15	0.45
P11	wimp	factor controlling whether water is impermeable	soil	0.01	0.1
P12	pondmx	maximum ponding depth for soil surface	soil	5	15
P13	csoilc	the drag coefficient for the soil under the canopy	soil	0.002	0.006
P14	zsno	roughness length for snow	snow	0.0012	0.0036
P15	capr	tuning factor of soil surface temperature	soil	0.17	0.51
P16	cnfac	Crank Nicholson factor	canopy	0.25	0.5
P17	dewmx	maximum ponding of leaf area	canopy	0.05	0.15
P18	Z_{0m}	aerodynamic roughness length	canopy	2	3
P19	χ_L	leaf-angle distribution factor	canopy	0.01	0.25
P20	$\alpha_{V,l}$	leaf reflectance, visible, live	canopy	0.07	0.1
P21	$\alpha_{V,d}$	leaf reflectance, visible, dead	canopy	0.16	0.36
P22	$\alpha_{N,l}$	leaf reflectance, near IR, live	canopy	0.35	0.45
P23	$\alpha_{N,d}$	leaf reflectance, near IR, dead	canopy	0.39	0.58
P24	$\delta_{V,l}$	leaf transmittance, visible, live	canopy	0.05	0.07
P25	$\delta_{V,d}$	leaf transmittance, visible, dead	canopy	0.001	0.3
P26	$\delta_{N,l}$	leaf transmittance, near IR, live	canopy	0.1	0.25
P27	$\delta_{N,d}$	leaf transmittance, near IR, dead	canopy	0.001	0.38
P28	slti	slope of low-temperature inhibition function	canopy	0.1	0.3
P29	hlti	1/2 point of low-temperature inhibition function	canopy	278	288
P30	shti	slope of high-temperature inhibition function	canopy	0.15	0.45
P31	hhti	1/2 point of high-temperature inhibition function	canopy	303	313
P32	sqrtdi	the inverse of the square root of the leaf dimension	canopy	2.5	7.5
P33	α	quantum efficiency of vegetation photosynthesis	canopy	0.06	0.08
P34	V_{max25}	maximum carboxylation rate 25°C	canopy	10	200
P35	m	the slope of the conductance-photosynthesis model	canopy	4	9
P36	b	intercept of conductance-photosynthesis model	canopy	0.01	0.04
P37	N	coefficient of leaf nitrogen allocation	canopy	0.5	0.75
P38	smpmax	wilting point potential	canopy	-2e5	-1e5
P39	smpmin	restriction for min of soil potential	soil	-1e8	-9e7
P40	E_{trmax}	maximum transpiration for vegetation	canopy	1e-4	100e-4

S3 The leaf stomatal photosynthesis-conductance module of CoLM

In CoLM, leaf stomatal resistance is coupled to leaf photosynthesis in a manner way:

$$\frac{1}{r_s} = m\frac{A}{c_s}\frac{e_s}{e_i}p_s + b \quad (2)$$

where r_s is leaf stomatal resistance; $m(P35)$ and $b(P36)$ are the slope and intercept of the linear regression equation, respectively; c_s and e_s are the CO2 concentration and vapor pressure at the leaf surface, respectively; e_i is saturation vapor pressure inside the leaf; p_s is atmospheric pressure at the surface; A is the leaf photosynthesis and is calculated by:

$$\mathrm{A} = \min\left(\mathrm{w_c}, \mathrm{w_j}, \mathrm{w_e}\right) \quad (3)$$

$$\mathrm{w_c} = \frac{(\mathrm{c_i}-\Gamma_*)\mathrm{V_{max25}}}{\mathrm{c_i}+\mathrm{K_c}(1+\mathrm{o_i}/\mathrm{K_o})} \quad (4)$$

$$\mathrm{w_j} = \frac{(\mathrm{c_i}-\Gamma_*)4.6\phi\alpha}{\mathrm{c_i}+2\Gamma_*} \quad (5)$$

$$\mathrm{w_e} = 0.5\ \mathrm{V_{max}} \quad (6)$$

where $\mathrm{w_c}$ and $\mathrm{w_j}$ are the Rubisco and light limited rate of carboxylation, respectively; $\mathrm{c_i}$ and $\mathrm{o_i}$ are CO2 and O2 concentration, respectively; $\mathrm{K_c}$ and $\mathrm{K_o}$ are Michaelis-Menten constants for CO2 and O2, respectively; Γ_* is the CO2 compensation point; $\alpha(P33)$ is the quantum efficiency; ϕ is the absorbed photosynthetically active radiation; $\mathrm{V_{max}}$ is the maximum rate of carboxylation varies with temperature, foliage nitrogen, and soil water:

$$\mathrm{V_{max}} = \mathrm{V_{max25}}\mathrm{a_{vmax}}^{\frac{\mathrm{T_v}-25}{10}}\mathrm{f(N)f(T_v)f(w)} \quad (7)$$

where $\mathrm{V_{max25}}(P34)$ is the maximum rate of carboxylation at 25 ℃; $\mathrm{a_{vmax}}$ is a temperature sensitivity paramter; $\mathrm{f(N)}, \mathrm{f(T_v)}, \text{and } \mathrm{f(w)}$ are the limitations of foliage nitrogen, temperature, and soil water, respectively.

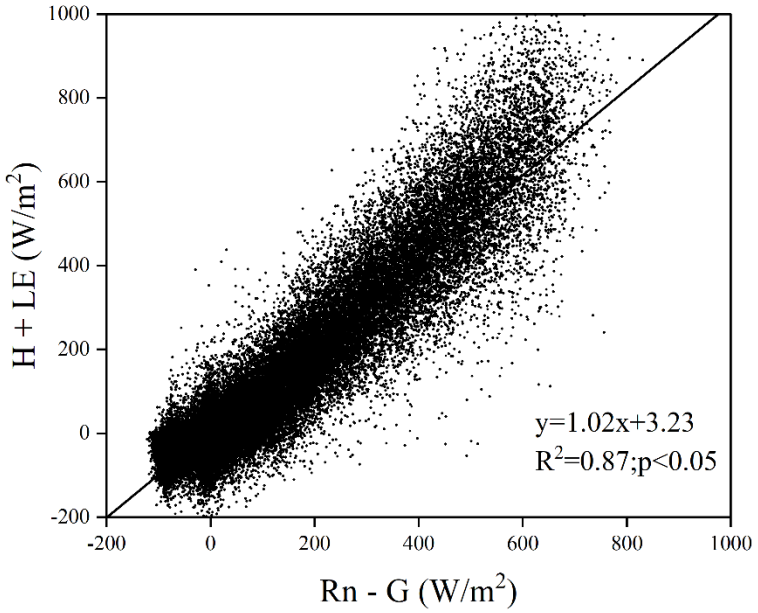

Figure S6. The multi-year(2015-2020) energy balance closure of flux observation in the study site. Note that the results above are obtained after eliminating moments with missing or invalid values for net radiation, sensible heat flux, latent heat flux, and soil heat flux observations.

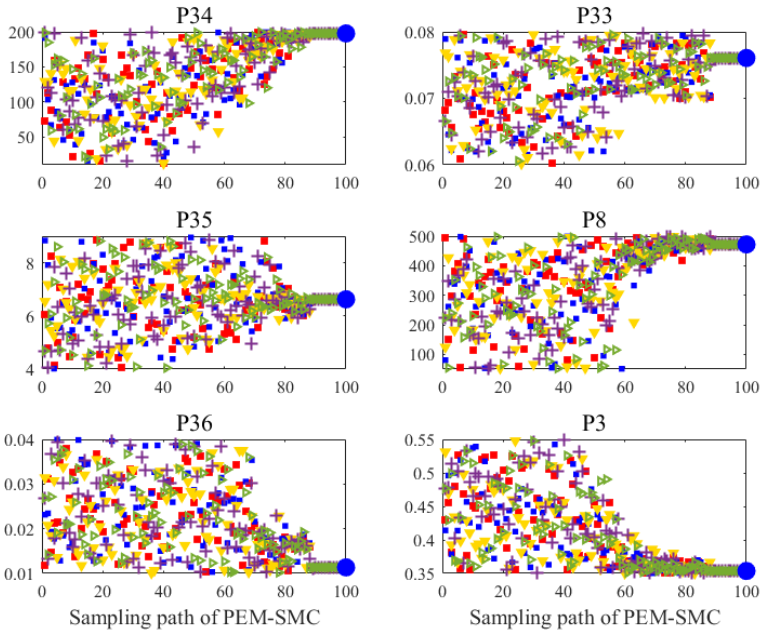

Figure S7. Transitions of the sampled parameter values in five randomly selected particles during the sampling path in the optimization scenario of Opt_LE. The vertical axis denotes the prior range of the parameter and the horizontal axis denotes the number of evolutions. The solid circles are the median value of all particles at the end of evolutions.

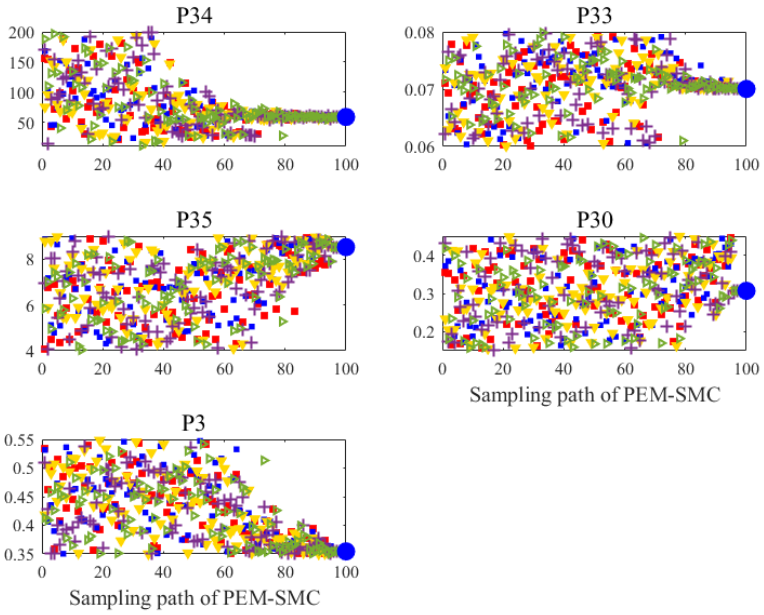

Figure S8. The same as in Fig. S7 but for the optimization scenario of Opt_NEE.

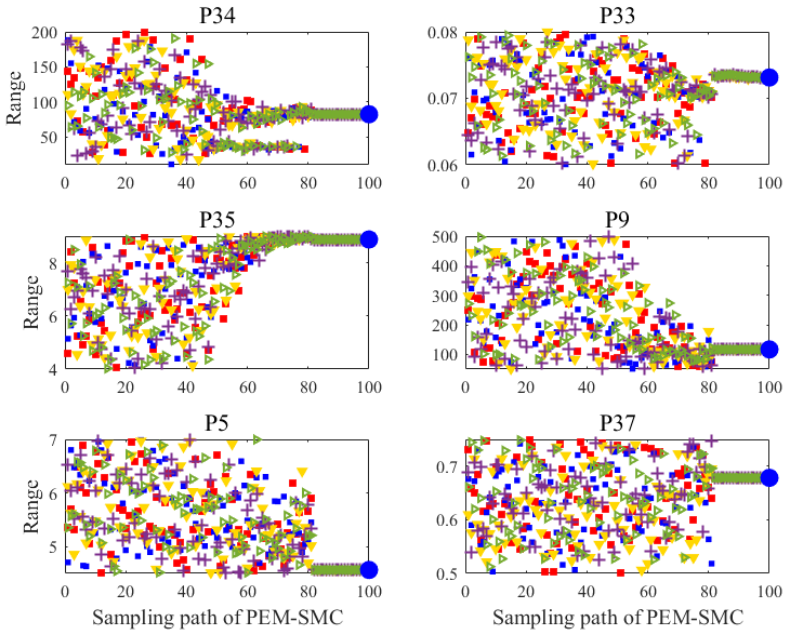

Figure S9 The same as in Fig. S7 but for the optimization scenario of Opt_ALL.