Appendix B

Data Assimilation Approach

This approach is based on Bayes' theorem to construct posterior distribution. The *a priori* information is defined as $B(x)$ to be used for an input parameter vector x. The *a posteriori* information $p(x)$ of the vector x can be defined as:

$$p(x) = r \cdot B(x) \cdot L(x)$$

where r is an appropriate normalization constant, and $B(x)$ can be combined with the information obtained from comparisons between model outputs and observed data. $L(x)$ is the likelihood function, which can be used to test the fit between observed and modelled value, and be expressed as following by assuming model errors are independent and following Gaussian distribution:

$$L(x) = \prod_{i=1}^{n} [(2\pi\sigma)^{-0.5} e^{-(u-u')^2 (2\sigma^2)^{-1}}]$$

$$u = \{u_i, i = 1,\ 2,\ \ldots, m\}$$

$$u' = \{u_i', i = 1, 2,\ \ldots\ldots, m\}$$

where u and u' denote the measured and simulated NEP (in $gCm^{-2}d^{-1}$), respectively, σ represents estimated standard deviation of each data point, which related to the given model parameters and, thus, represents a combination of measurement error and process representation error.

The Metropolice-Hastings (MH) algorithm was used to obtain sequences of random samples from the distribution of model parameters. After initialization, the acceptance distribution is judged to pick up a new parameter sample from a candidate parameter set. The acceptance probability can be defined as follows,

$$p = min[1, p(\psi_k \mid x) / p(\psi_{k-1} \mid x)]$$

where ψ is a parameter set that includes four parameters in this study ψ=(V_{max}, J_{max}, m, R_{10}). p is ranging between 1 and 0 at each iterative process. This is used to chose the candidate parameters (the vector x) until reaching the last k iteration.

This study conducted MCMC for the model simulations in each month of 2006 at seven forest flux towers. Limited by the space, here we present an example of posterior distribution of the four constrained parameters for ENB situation in March 2006 (Fig. B1). Our experiment indicates that the simulations demonstrated stabilized convergences after 3500 loops in the iterations (e.g. see Fig. B2).

Another experiment demonstrated similar convergences in optimizing both R_{10} and Q_{10} (Fig. B3), and did not affect the convergences of other parameters (Vmax, Jmax, and m) and the NEP simulates. The sensitivity of the five key parameters was tested at the beginning of this study. With value increase of each parameter, all NEP simulates increased as well. Nonetheless, comparing the increases of parameter values, modelled NEP had different rates of relative increases, such as 42.6%, 8.6%, 15.0%, and 17.8% faster than increases of Vmax, m, R_{10}, and Q_{10} respectively, and 18.6% slower than the increase of Jmax. The comparison suggests that Vmax, Jmax and Q_{10} are more sensitive to simulate NEP than other parameters.

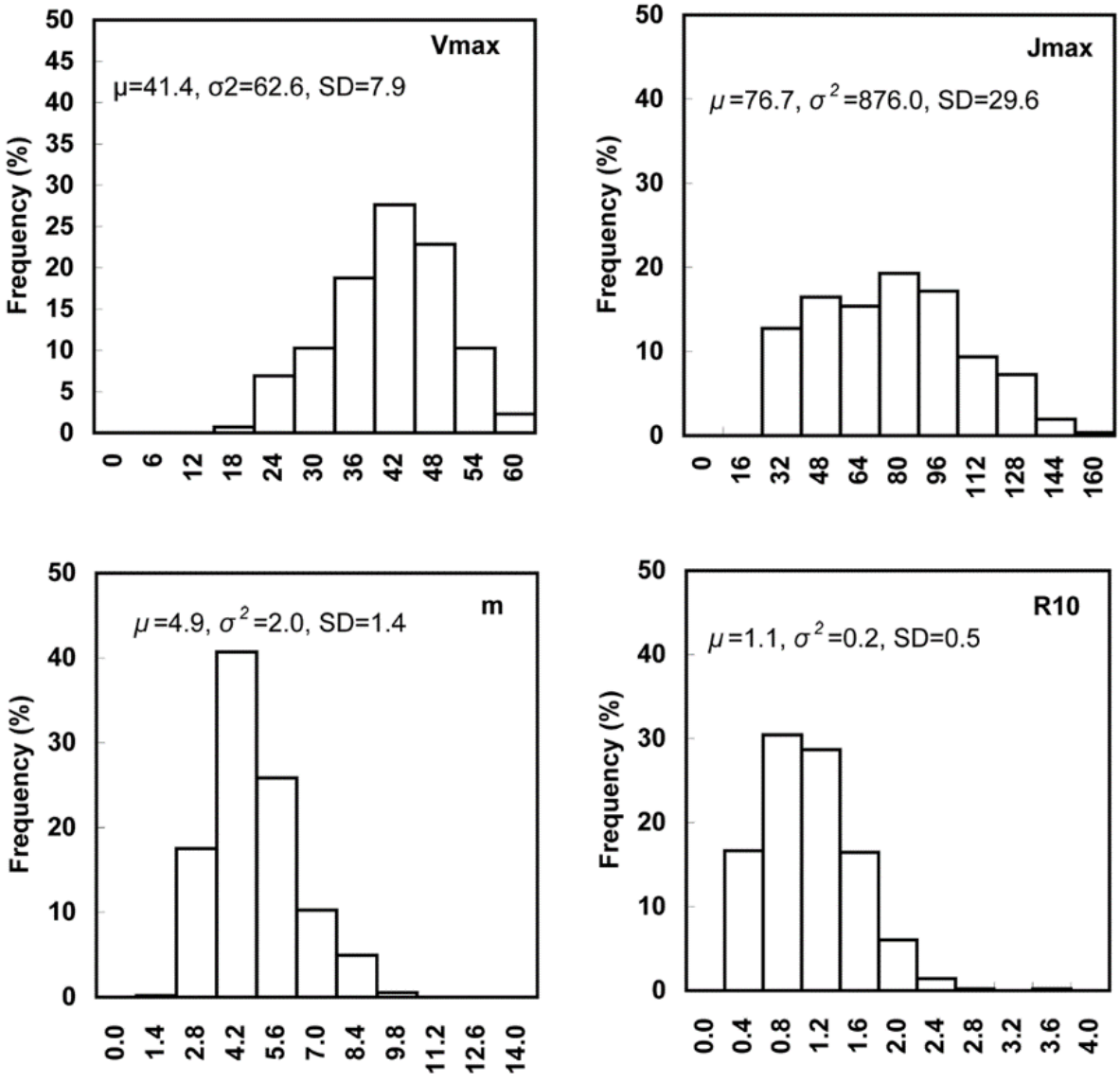

FIG. B1. An example of posterior distribution of the four constrained parameters. These simulations were performed to optimize parameters for ENB situation in March of 2006. Each simulation uses 5000 samples to construct the distribution.

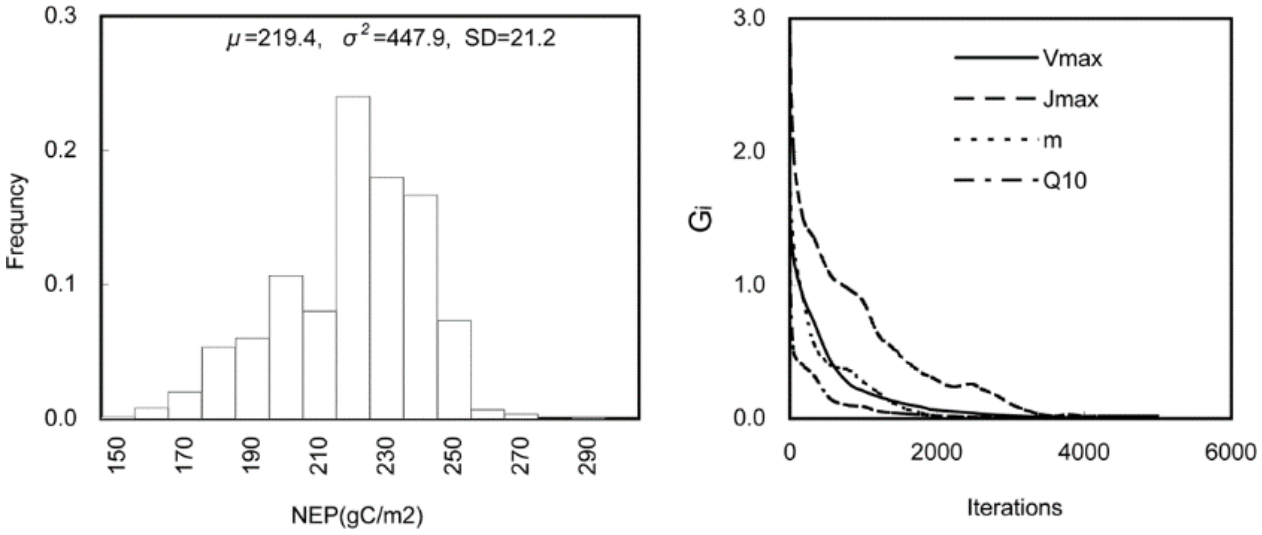

FIG. B2. An example of the convergence estimate for the posterior parameters illustrated in Fig. B1. The convergence stabilized after 3500 loops in the iterations based on the convergence diagnostic index ($G_i=D_i(k-1)/D_i(k)$).

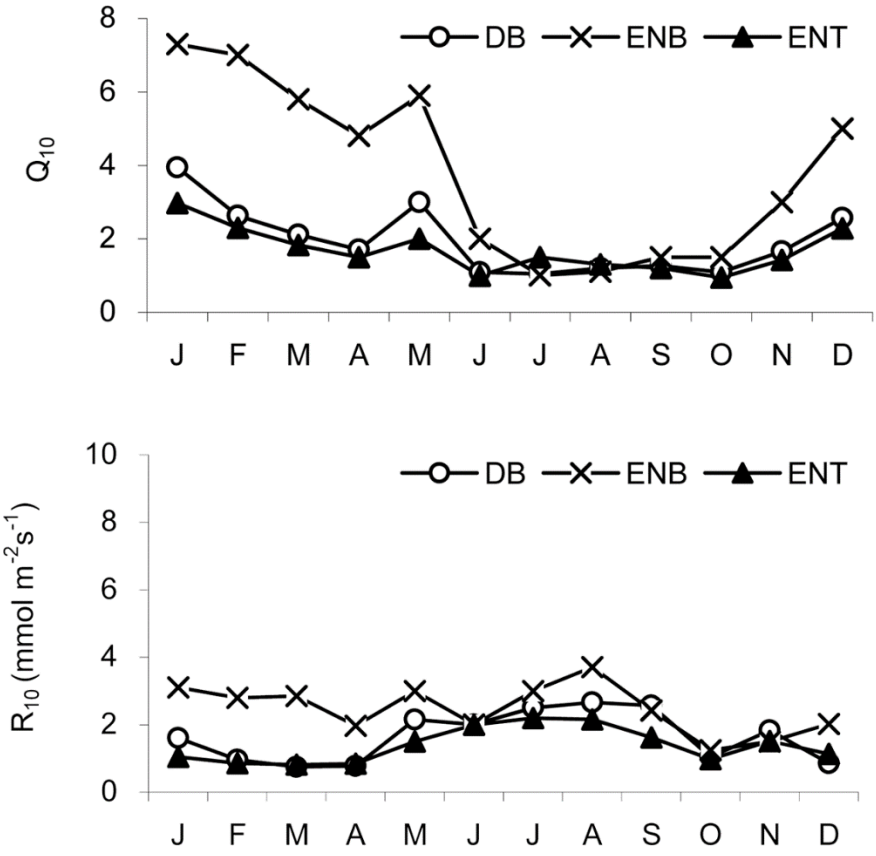

FIG. B3. An example of seasonal variation of parameters R_{10} and Q_{10} for the different forest ecosystems under investigation (2006). In this experiment, the parameter R_{10} and Q_{10} were tested concurrently with Vmax, Jmax, and m. The forest types are defined as same as in the Fig. 2.