Supplementary Information

An Indirect Simulation-Optimization Model for Determining Optimal TMDL Allocation under Uncertainty

Text S1. Methodology of BRRT v2

S1.1. General Principles

We begin with an introduction of the general principles of BRRT v2. The BRRT v2 is an updated version of Bayesian TREED model developed by [1]. Taking N_2O EF for example, combining the conditional distribution of Y with prior $p(T)$ for model structure, the posterior probability $p(T|X,Y)$ is calculated as

$$p(T|X,Y) \propto p(T)\,p(Y|X,T) \tag{S1}$$

up to a norming constant, where *EF* and *X* are emission factor and environmental factors, respectively; $X = \{x_k\}$; *i.e.*, *N* and x_k in Equation (1). *T* represents a binary tree. Its terminal nodes and splitting rules correspond to the sub-functions, *i.e.*, Equation (1) and the sub-domains, respectively. Thus, specification of $p(T|X,Y)$ consists of two basic components.

First is the tree prior $p(T)$, which could be specified independently according to Bayes' theorem. $p(T)$ is determined by specifying two functions for an binary tree *T*: the probability $p(l,T)$ that terminal node *l* is to be split and the probability $p(\rho|l,T)$ of assigning splitting rule $\rho = \{x_k \le S_k\}$ to node *l* if it is split, which are defined as

$$p(T) = p(l,T)\,p(\rho|l,T) \tag{S2a}$$

where

$$p(l,T) = \tau(1+d)^{-\upsilon} \tag{S2b}$$

$$p(\rho|l,T) = \left(K(M_l - 1)\right)^{-1} \tag{S2c}$$

and x_k is environmental factors applied for sub-domain division; τ (<1) and υ (>0) are hyperparameters; *d* is the depth of node *l*; *K* and M_l are the number of x_k and observations in terminal node *l*, respectively.

Second is the marginal likelihood $p(Y|X,T)$, which can be obtained as integral form of regression coefficients $\theta_l = \{\lambda_l, \sigma_l\}$:

$$p(Y|X,T) = \prod_{l=1}^{L} \int \prod_{m=1}^{M_l} p\left(Y_{lm}|x_{klm},\theta_l\right) p\left(\theta_l|T\right) p\left(x_{kl}|T\right) d\theta_l \tag{S3a}$$

where

$$Y_l|\theta_l \ \text{iid} = N\left(x_l^T\lambda_l,\sigma_l^2\right) \tag{S3b}$$

$$\lambda_l|\sigma_l = N\left(\bar{\lambda}_l,\sigma_l^2/a\right) \tag{S3c}$$

$$\sigma_l^2 = v\omega / \chi_v^2 \tag{S3d}$$

$$p\left(x_{kl}\middle|T\right) = \frac{C_K^n}{C_K^0 + C_K^1 + \ldots + C_K^K}, \quad \forall k = 1,\ldots,n \tag{S3e}$$

and $p(Y_{lm}|x_{klm}, \theta_l)$, $p(\theta_l|T)$, and $p(x_{kl}|T)$ are probabilities of Y, regression coefficients, and selection of n of regression variables x_k in terminal l, respectively, in which n is the number of selected x_k. EF_{lm} and x_{klm} the EF and regression variable k of the mth observation in the lth terminal node, respectively, where $x_{klm} \in X$, $m = 1,2,\ldots, M_l$. All of Y_l are assumed to be independent and identically distributed, in which $x_l^T \lambda_l$ corresponds to Equation (1). θ_l represent a set of regression coefficients λ_l and the variance σ_l of EF in terminal l. λ_l is Normal-distributed, defined as $\lambda_l = \{ \Delta EF_l(x_k), b_{kl}, c_l \}$. σ_l is inverse Gaussian-distributed with mean v and shape parameter ω, in which v and ω are the other two hyperparameters of the BRRT v2. The choice of hyperparameters τ (<1), υ (>0), a (= 1 or 3), v (= 3), and ω (= 0.404 or 0.1173) is determined with minimum cost function (Freeman *et al.*, 2009). The ranges of $\bar{\lambda}_l$ are determined based on the marginal responses of EF to x_k.

It should be noted that the abovementioned selection of regression variables in Equation (S3e), as the **first** improvement of the BRRT v2, is achieved through both random and deterministic processes. First, regression variables are randomly chosen based on prior $p(x_{kl}|T)$ for each terminal node; Second, the randomly-chosen x_k are further screened by stepwise forward regression [2]. The forms of Y_l and $p(Y|X,T)$ are updated accordingly. Contrary to the BTREED [1], variable selection helps reduce the number of redundant variables used in regression equations for individual terminal nodes.

Consequently, progressive multi-restart stochastic search (hereafter PMRS) algorithm is designed for determining the posterior probability of T with minimum cost function. Based on the above prior $p(T)$ and $p(\theta_l|T)$, Metropolis-Hastings algorithm could be used to search for a Markov chain sequence of trees. Starting with an initial tree T^0, it simulates iteratively the transitions from T^i to T^{i+1} through two steps: (1) generate a candidate value $T*$ by five-mode stochastic search with probability $q(T^i, T^*)$; (2) set $T^{i+1} = T*$ with acceptance probability:

$$\delta = \text{Min}\left\{ \Phi = \frac{q(T^*,T^i)}{q(T^i,T^*)} \frac{p(EF|X,T^*)p(T^*)}{p(EF|X,T^i)p(T^i)}, 1 \right\} \tag{S4}$$

Else, set $T^{i+1} = T^i$ and return to step 1. When $i+1 > I$ where I is maximum number of iterations, then stop. Moreover, the transition from T^i to T^{i+1} is processed by stochastically choosing among five modes (S1.2): GROW, PRUNE, CHANGE, SWAP, and GREEDY.

Finally, from Markov chain sequence, the optimal tree (T^{final}) is selected based on the criterion of minimum cost function. In our model, cost function that is minimized is Bayesian Information Criterion (BIC, [3]):

$$BIC = \log(MSE) \cdot n + \log(n) \cdot (K+1) \cdot L \tag{S5}$$

where *MSE* is the mean squared error based on 5-fold cross-validations; n the number of samples (= $\sum_l^L M_l$); and $K + 1$ the number of coefficients and intercepts in regression equation l. Certainly, other commonly used tree criteria would be also useful, such as maximization of marginal likelihood $p(EF|X,T)$. The associated piecewise functions for all terminal nodes can be expressed as:

$$Y_l = X_l^T \lambda_l \pm \varepsilon_l, \text{ where } X_l \in \Omega_l, \forall l \tag{S6}$$

And Ω_l represents the sub-domain of X_l. It should be noted that regression coefficients λ_l are not determined by multiple linear regression after stochastic search (e.g., [3]), but PMRS process directly. In BRRT v2, the cost function that is minimized is BIC. Other commonly used cost function would be also acceptable, such as maximization of $p(EF|X,T)$.

S1.2. Five Modes for Stochastic Search

The details of PMRS algorithm and the associated five modes are introduced in the following two sub-sections. In contrast to conventional Bayesian tree-based models [1,3], a PMRS algorithm is designed in BRRT v2. This algorithm accelerates traversing the entire optima of trees to identify the optimal tree via a minimum number of model executions.

The **second** improvement of BRRT v2 compared with BTREED [1] is the addition of a new operation—GREEDY—into five-mode stochastic search, which can fast identify best split or prune when approaching the optimal tree. This means that it can avoid the situation stabilizing in local sub-domains [3]. Here we simply introduce the details of each mode orderly but take pains to elaborate the GREEDY mode.

(1) **GROW**. A terminal node l is randomly selected based on prior $p(l,T)$ and then split into two new nodes by randomly assigning it a splitting rule $\rho = \{x_k \leq S_k\}$ according to $p(\rho|l,T)$. The corresponding transition kernel q and $p(T^*)$ are deduced as follows:

$$q(T^i, T^*) = p(GROW, T^i) \cdot \frac{1}{a_1} \cdot p(\rho|\eta, T) \tag{S7a}$$

$$q(T^*, T^i) = \left(1 - p(GROW, T^*)\right) \cdot \frac{1}{b_2'} \tag{S7b}$$

$$p(T^*) = \frac{p(T^i)p(\rho|i,T)}{1 - p(\eta, T)} \cdot p(\eta) \cdot \left(1 - p(\eta', T)\right)^2 \tag{S7c}$$

where $p(GROW, T^i)$ or $p(GROW, T^*)$ is the probability of choosing GROW mode; a_1 is the number of terminal nodes that can be split; b_2' is the number of internal nodes in T^* with terminal children nodes; η is the terminal node that is split for GROW mode and the internal nodes before CHANGE and SWAP; η' represents the two children nodes of η for GROW mode.

(2) **PRUNE**. A parent of two terminal nodes is randomly selected and turned into a terminal node by collapsing the nodes below it. The corresponding transition kernel q and $p(T^*)$ are calculated by:

$$q(T^i, T^*) = \left(1 - p(GROW, T^i)\right) \cdot \frac{1}{b_2} \tag{S8a}$$

$$q(T^*, T^i) = p(GROW, T^*) \cdot \frac{1}{b_1 - 1} \cdot p(\rho|\eta', T^*) \tag{S8b}$$

$$p(T^*) = \frac{p(T^i)\left(1 - p(\eta', T^i)\right)}{\left(1 - p(\eta_1, T^i)\right)^2 p(\eta', T^i)p(\rho|\eta', T^i)} \tag{S8c}$$

where η′ represents the internal node to be pruned of its two children nodes η1.

(3) **CHANGE**. An internal node is randomly selected and reassigned a new splitting rule ρ′ according to $p(\rho|l,T)$;

$$q(T^i, T^*) = q(T^*, T^i) \tag{S9a}$$

$$p(T^*) = p(T^i) \cdot \prod_{\eta'=1}^{n'} p(\rho'|\eta', T) \prod_{\eta=1}^{n'} p(\rho|\eta, T)^{-1} \tag{S9b}$$

where η′ represents the internal nodes after CHANGE and SWAP; n' is the number of both the chosen node and all nodes below it.

(4) **SWAP**. A parent-child pair that are both internal nodes is randomly selected, and the splitting rules are swapped, unless the other child has the same rule, in which case the splitting rule of the parent is swapped with that of both children. Chipman *et al*. [1] pointed that this search converged toward regions of higher posterior probability $p(T|X, EF)$ or marginal likelihood $p(EF|X, T)$. The corresponding transition kernel q and $p(T^*)$ are identical to that of CHANGE mode.

(5) **GREEDY**. Previous studies, such as Jung *et al*. [3], grow deterministic tree by GREEDY algorithm as starting point (e.g., TRIAL), and launch stochastic search for subsequent Markov chain sequence of trees. However, our strategy is to insert GREEDY directly into five-mode stochastic research as one choice of modes. The methodology of GREEDY is same as CART, but only one terminal nodes is split or sub-tree of an intermediate node is pruned for each transition from T^i to T^{i+1}. When implementing GREEDY mode, a terminal node is randomly selected based on prior $p(l, T)$ and only the terminal node $\hat{l}$ in which the corresponding residual sums of squares (RSS) is minimum among all terminal nodes in T^* has chance to be split into two children nodes. If the fact that RSS within terminal node $\hat{l}$ is larger than that of its two child nodes, that is

$$RSS\left(T^i\right) > RSS\left(T^*\right) \tag{S10a}$$

where

$$RSS\left(T^i\right) = \sum_{m=1}^{M_{\hat{l}}} \left[O\_EF_{\hat{l}m} - S\_EF_{\hat{l}m} \right]^2 \tag{S10b}$$

$$RSS\left(T^*\right) = \sum_{l'=1}^{2} \sum_{m=1}^{M_{l'}} \left[O\_EF_{l'm} - S\_EF_{l'm} \right]^2 \tag{S10c}$$

is satisfied, we set $T^{i+1} = T^*$. Where l' represents two child nodes in T^* of terminal node l. Else, an intermediate node t (as parent T_t of any sub-tree T_l) is randomly selected and pruned upward into a terminal node when the complexity cost reaches a minimum. We then also set $T^{i+1} = T^*$. Otherwise, $T^* = T^i$ and the search returns to generate another candidate value T^* with probability $q(T^i, T^*)$ again. The corresponding transition kernel q and $p(T^*)$ are same as that of GROW or PRUNE modes.

S1.3. PMRS Algorithm

In contrast to conventional Bayesian tree-based models [1,3], a progressive multi-restart stochastic search algorithm in BRRT v2, as the **third** improvement, is designed, which accelerates traversing the entire optima of trees to identify the optimal tree via a minimum number of model executions.

The basic strategy of the progressive multi-restart stochastic search is to synchronously run a number ($j = 1,...,J_k$) of restart scenarios of Markov chain sequence $T_j^0...T_j^i$ and progressively eliminate underperforming trees before executing subsequent running stages, until the optimal tree is identified in final running stages with minimum cost function.

Specifically, J_k denotes the total number of restart scenarios in stage k, and I denotes the user-specified number of iterations in the stochastic search. For restart stage 1, each restart scenario j (where $j \leq J_k$, $k = 1$) begins with a single node initial tree T^0 and ends in a multi-node tree with a structure obtained through the stochastic search process, where the Metropolis-Hastings algorithm is used to simulate a Markov chain sequence of trees by randomly choosing among the five stochastic operations. Specifically, if the root node is chosen, it is split where probability $p(\text{GROW})$ is 1; if an internal node is chosen, it is randomly reassigned a splitting rule, its splitting rule is swapped, or it is deterministically split or pruned, where $p(\text{CHANGE})$, $p(\text{SWAP})$, or $p(\text{GREEDY})$ is 1/3; if a terminal node is selected, it is randomly split, or pruned, or deterministically split or pruned, where $p(\text{GROW})$, $p(\text{PRUNE})$, or $p(\text{GREEDY})$ is 1/3. Subsequently, a candidate tree T^* is generated with the acceptance probability δ If $\delta \geq U(0,1)$, T^* can be accepted as the next value $T^{i+1} = T^*$; otherwise, the current value of T^i is retained, such that $T^{i+1} = T^i$.

Restart stage 1 is considered complete when all the J_1 initial trees have been run to generate J_1 different resultant trees. Upon completion of stage 1, the BIC of all the J_1 trees are ranked from highest to lowest, and the first $\text{Int}(J_k/2)$ trees with higher $\underset{i}{Min}\,BIC_j(T^i), \forall j$ are used as the initial trees for restart stage 2. The process continues until the optimal tree structure is obtained in restart stage K, where $\text{Int}(J_K/2) = 0$.

S1.4. Solution Procedure

To facilitate the real-world application of the BRRT v2, the source code was written in the VC++ language and comprises the following five steps (ask authors for executable file BRRT v2.exe if interested):

Step #1: *Sensitivity analysis of hyperparameters.* Explore a range of hyperparameters $\tau(<1)$, $\upsilon(>0)$, a (= 1 or 3), and ω (= 0.404 or 0.1173) and determine their optimal values when *BIC* is minimum;

Step #2: *Setup for stochastic search.* Set the initial number of restart J_1 and iterations I per comparison stage;

Step #3: *Model implementation.* Calibrate BRRT v2.exe by *X-EF* dataset to generate Markov chain sequence and calculate the corresponding *BIC*s;

Step #4: *Selection of the T^{final}.* Select optimal T^{final} with minimum *BIC*;

Step #5: *Output.* Export a set of piecewise functions and the associated sub-domains of X_l.

Text S2. Methodology of EILP

To clearly describe the theory of the EILP algorithm, the abovementioned minimum load reduction model [Equation (1)] was generalized as

$$Min \ Z^{\pm} = \sum_{j}^{n} c_j^{\pm} x_j^{\pm} \tag{S11a}$$

$$s.t \ \sum_{i}^{n} a_{ij}^{\pm} x_j^{\pm} \le b_i^{\pm}, \forall i \tag{S11b}$$

$$x_j^{\pm} \ge 0 \tag{S11c}$$

Equation (2) could be decomposed into two submodels corresponding to the lower and upper bounds of the objective function. In this paper, we present the computational procedure of the EILP algorithm, including the decomposition of the upper and lower bounds of $Z^{\pm}$ and their corresponding constraints, a feasibility analysis of the optimal solution space, and a tradeoff analysis for decision making under EI-type uncertainty, which was explained in detail by [4,5].

(i) Determination of the upper and lower bounds and expected values of $Z^{\pm}$

Suppose that k of $c_j^{\pm}$ ($j = 1, 2, \ldots, n$) is positive ($c_j^{\pm} \ge 0, j = 1, s, \ldots, k$) and the others are negative ($c_j^{\pm} \le 0, j = k + 1, k + 2, \ldots, n$). When considering the uncertainties of $c_j^{\pm}$, the lower and upper bounds of the objective function could be determined by the MILP model as $Z^{\pm} = [Z^-, Z^+]$; otherwise, they were determined by the EILP model as the appropriate interval ($AI^{\pm} = [AI^-, AI^+]$). Moreover, the optimal expected values $E[Z^{\pm}]$ of the objective function could then be derived according to $AI^{\pm}$. Thus, the lower and upper bounds of $Z^{\pm}$, $AI^{\pm}$, and $E[Z^{\pm}]$ were defined as follows:

$$Z^- = \sum_{j=1}^{k} c_j^- x_j^- + \sum_{j=k+1}^{n} c_j^- x_j^+ \tag{S12a}$$

$$Z^+ = \sum_{j=1}^{k} c_j^+ x_j^+ + \sum_{j=k+1}^{n} c_j^+ x_j^- \tag{S12b}$$

$$AI^{\pm} = [AI^-, AI^+] = [\frac{Z^{\leftarrow} + Z^-}{2}, \frac{Z^{\rightarrow} + Z^+}{2}] \tag{S12c}$$

$$E[Z^{\pm}] = \frac{1}{2}\left(AI^- + AI^+\right) \tag{S12d}$$

where c_j^+ and c_j^- were the lower and upper bounds of $c_j^{\pm}$, respectively, and the upper and lower bounds of of $Z^{\pm}$, Z^f, and $AI^{\pm}$ are defined as

$$Z^+ = \sum_{j=1}^{k} c_j^+ x_j^+ + \sum_{j=k+1}^{n} c_j^+ x_j^- \tag{S12e}$$

$$Z^- = \sum_{j=1}^{k} c_j^- x_j^- + \sum_{j=k+1}^{n} c_j^- x_j^+ \tag{S12f}$$

$$Z^{\rightarrow} = \sum_{j=1}^{k} c_j^- x_j^+ + \sum_{j=k+1}^{n} c_j^- x_j^- \tag{S12g}$$

$$Z^{\leftarrow} = \sum_{j=1}^{k} c_j^+ x_j^- + \sum_{j=k+1}^{n} c_j^+ x_j^+ \tag{S12h}$$

$$AI^+ = \sum_{j=1}^{k} 0.5(c_j^+ + c_j^-)x_j^+ + \sum_{j=k+1}^{n} 0.5(c_j^+ + c_j^-)x_j^- \tag{S12i}$$

$$AI^- = \sum_{j=1}^{k} 0.5(c_j^+ + c_j^-)x_j^- + \sum_{j=k+1}^{n} 0.5(c_j^+ + c_j^-)x_j^+ \tag{S12j}$$

(ii) Determination of the corresponding constraints

To minimize the objective function, the relationships between the decision variables $\mathbf{X}^{\pm}$ and the left-hand side coefficients $\mathrm{A}^{\pm}$ should be $x_j^- \rightarrow |a_{ij}^{\pm}|^+ Sign(a_{ij}^{\pm})$ ($\forall j$ = 1, 2,..., k) and $x_j^+ \rightarrow |a_{ij}^{\pm}|^- Sign(a_{ij}^{\pm})$ ($\forall j = k + 1, k + 2,..., n$) for the first submodel Z^-. To ensure that $x_j^+ \geq x_{jopt}^-$ for j = 1, 2,..., k and $x_j^- \leq x_{jopt}^+$ for $j = k + 1, k + 2,..., n$ in the second submodel Z^+, the corresponding relationships should be $x_j^+ \rightarrow |a_{ij}^{\pm}|^- Sign(a_{ij}^{\pm})$ ($\forall j$ = 1, 2,..., k) and $x_j^- \rightarrow |a_{ij}^{\pm}|^+ Sign(a_{ij}^{\pm})$ ($\forall j = k + 1, k + 2,..., n$), where x_{jopt}^- ($\forall j = 1, 2,..., k$) and x_{jopt}^+ ($\forall j = k + 1, k + 2,..., n$) are the optimal solutions of the first submodel. This leads to EI-type solutions for the lower and upper bounds of the objective functions (Equation (3a)–(3d)) with $Z_{jopt}^- < Z_{jopt}^+$.

Therefore, the corresponding constraints for Z^- are written as

$$\sum_{j=1}^{k} |a_{ij}^{\pm}|^+ Sign(a_{ij}^{\pm})x_j^- + \sum_{j=k+1}^{n} |a_{ij}^{\pm}|^- Sign(a_{ij}^{\pm})x_j^+ \leq b_i^+, \quad \forall i \tag{S13a}$$

$$x_j^{\pm} \geq 0, \forall j = 1, 2,..., n \tag{S13b}$$

and Z^+ in the second submodel is then solved subject to the following constraints:

$$\sum_{j=1}^{k} |a_{ij}^{\pm}|^+ Sign(a_{ij}^{\pm})x_j^- + \sum_{j=k+1}^{n} |a_{ij}^{\pm}|^- Sign(a_{ij}^{\pm})x_j^+ \leq b_i^-, \quad \forall i \tag{S14a}$$

$$x_j^+ \geq x_{jopt}^- \quad \text{for } j = 1, 2,..., k \tag{S14b}$$

$$x_j^- \leq x_{jopt}^+ \quad \text{for } j = k + 1, k + 2,..., n \tag{S14c}$$

$$x_j^{\pm} \geq 0, \forall j = 1, 2,..., n \tag{S14d}$$

(iii) Feasibility analysis of the optimal solution space

To ensure that the optimal solution $\mathbf{X}_{opt}^{\pm} = \left\{ x_{opt}^{\pm} = [x_{jopt}^-, x_{jopt}^+] | \forall j = 1, 2,..., n \right\}$ is absolutely feasible, extra constraints for minimizing $Z^{\pm}$ should be added to Equation (5) as follows:

$$\sum_{j=1}^{k-p} (|a_{\delta j}^{\pm}|^- x_j^+ - |a_{\delta j}^{\pm}|^+ x_{jopt}^-) - \sum_{j=k+1}^{n-q} (|a_{\delta j}^{\pm}|^+ x_j^- - |a_{\delta j}^{\pm}|^- x_{jopt}^+) \leq 0, \forall \delta \tag{S15}$$

where δ is also the number of constraints in Equation (4a) that meet $\sum_{j=1}^{k} \left|a_{\delta j}^{\pm}\right|^{+} Sign(a_{\delta j}^{\pm})x_{jopt}^{-} + \sum_{j=k+1}^{n} \left|a_{\delta j}^{\pm}\right|^{-} Sign(a_{\delta j}^{\pm})x_{jopt}^{+} = b_{\delta}^{+}$ as well as its $a_{\delta j}^{\pm} \le 0$ for $j = k - p + 1,..., k, k + 1,..., k - q$.

In the minimum load reduction model, $\mathbf{Y}^{i\pm}$ in Equation (S11) was derived from the process-oriented simulation model. Through integration with the BRRT model, Equation (S11) could be solved using the abovementioned EILP algorithm. Z^{-} [Equation (S12a)] with the corresponding constraints [Equation (S13)] should be formulated and solved prior to solving Z^{+} [Equation (S12b)] with the relevant constraints Equations (S14) and (S15). The corresponding optimal solutions of the BRRT-EILP model were $Z_{opt}^{\pm}$ and $E[Z_{opt}^{\pm}]$ with $\mathbf{X}_{opt}^{\pm} = \left\{x_{opt}^{\pm} = [x_{jopt}^{-}, x_{jopt}^{+}] \| \forall j = 1, 2,..., n\right\}$, which was the approximate solution of the traditional simulation-optimization model. Hence, $Z_{opt}^{\pm}$ and $E[Z^{\pm}]_{opt}$ could provide useful results in the following extreme and non-extreme tradeoff analyses of risk-based TMDL allocation.

(iv) Tradeoff analysis for decision making under EI-type uncertainty

According to the optimal solutions for $\mathbf{X}_{opt}^{\pm}$, the solution space could be separated to reflect the relationship between the benefits of TMDL allocation and the risk levels of violating constraints. When solution space was absolutely feasible, Z_{opt}^{-} or Z_{opt}^{+} were the largest or smallest load reductions with the highest or lowest risk levels, respectively, of optimization system violations due to the three types of interval coefficients ($a_{ij}^{\pm}$, $b_{i}^{\pm}$, and $c_{j}^{\pm}$). Conversely, $E[Z^{\pm}]_{opt}$ was a non-extreme or expected value of load reduction corresponding to a moderate risk level due to $a_{ij}^{\pm}$, $b_{i}^{\pm}$, and $c_{j}^{\pm}$. A general tradeoff analysis under uncertainty was developed by Guo *et al.* [6] for minimizing load reductions as follows:

(i) The first decision alternative as the appropriate tradeoff analysis was $\mathbf{X}_{1} = \left\{x_{j} \middle| x_{jopt}^{-} \text{for } j = 1, 2,..., k; x_{jopt}^{+} \text{for } j = k+1,..., n\right\}$ with the lowest total load reduction AI_{opt}^{+}, corresponding to the highest risk level of violating the constraints.

(ii) The second decision alternative as the appropriate tradeoff analysis was $\mathbf{X}_{2} = \left\{x_{j} \middle| x_{jopt}^{+} \text{for } j = 1, 2,..., k; x_{jopt}^{+} \text{for } j = k+1,..., n\right\}$ with the highest total load reduction AI_{opt}^{-}, corresponding to the lowest risk level of violating the constraints.

(iii) The three decision alternative as the non-extreme tradeoff analysis was $\mathbf{X}_{3} = \left\{x_{j} \middle| \mathbf{C}^{+}\mathbf{X}^{\pm} = \mathbf{C}^{-}\mathbf{X}^{\pm} = E[Z^{\pm}]_{opt}\right\}$ with an expected value of system benefit $E[Z^{\pm}]_{opt}$, corresponding to a moderate risk level of violating the constraints.

Text S3. Tradeoff Analysis of Scenario 1 and 2

1. Minimum Total Load Reductions of Inorganic Nitrogen for Scenario 1

(i) The first decision alternative in extreme tradeoff analysis was $\mathbf{X}_{1} = \left\{x_{jl}^{-} \middle| x_{11opt}^{-} = 0.42; x_{j1opt}^{-} = 0 \text{ for other } j\right\}$ with the lowest total load reduction Z_{opt}^{-} of 1870 kg of inorganic nitrogen. It corresponded to the highest risk level of violating constraint;

(ii) The second decision alternative in extreme tradeoff analysis was $\mathbf{X}_{2} = \left\{x_{jl}^{+} \middle| x_{11opt}^{+} = 0.42; x_{51opt}^{+} = 0.23; x_{j1opt}^{+} = 0 \text{ for other } j\right\}$ with the highest total load reduction Z_{opt}^{+} of 1989 kg of inorganic nitrogen, which corresponded to the lowest risk level of violating constraint;

(iii) The third decision alternative in non-extreme tradeoff analysis was
$\mathbf{X}_3 = \left\{ x_{jl}^{\pm} \middle| \sum_j^8 c_{jl}^{\pm} x_{jl}^{\pm} = 1929.5 \text{ where } x_{11opt}^{\pm} = 0.42; x_{51opt}^{\pm} = [0, 0.23] \text{ and } x_{j1opt}^{\pm} = 0 \text{ for other } j \right\}$ with the expected value of total load
reduction $E[Z^{\pm}]_{opt}$ of 1929.5 kg of inorganic nitrogen, which corresponded to the moderate risk
level of violating constraint;

2. Minimum Total Load Reductions of Inorganic Phosphorus for Scenario 1

(i) The first decision alternative in extreme tradeoff analysis was
$\mathbf{X}_1 = \left\{ x_{jl}^{-} \middle| x_{12opt}^{-} = 0.62; x_{22opt}^{-} = 0.79; x_{j2opt}^{-} = 0 \text{ for other } j \right\}$ with the lowest total load reduction Z_{opt}^{-} of 406.2 kg
of inorganic phosphorus. It corresponded to the highest risk level of violating constraint;

(ii) The second decision alternative in extreme tradeoff analysis was
$\mathbf{X}_2 = \left\{ x_{jl}^{+} \middle| x_{12opt}^{+} = 0.74; x_{22opt}^{+} = 0.79; x_{j2opt}^{+} = 0 \text{ for other } j \right\}$ with the highest total load reduction Z_{opt}^{+} of 424.2 kg
of inorganic phosphorus, which corresponded to the lowest risk level of violating constraint;

(iii) The third decision alternative in non-extreme tradeoff analysis was
$\mathbf{X}_3 = \left\{ x_{jl}^{\pm} \middle| \sum_j^8 c_{jl}^{\pm} x_{jl}^{\pm} = 415.2 \text{ where } x_{12opt}^{\pm} = [0.62, 0.74]; x_{22opt}^{\pm} = 0.79 \text{ and } x_{j2opt}^{\pm} = 0 \text{ for other } j \right\}$ with the expected value of total
load reduction $E[Z^{\pm}]_{opt}$ of 415.2 kg of inorganic phosphorus, which corresponded to the moderate
risk level of violating constraint;

3. Minimum Total Load Reductions of Inorganic Nitrogen for Scenario 2

(i) The first decision alternative in extreme tradeoff analysis was $\mathbf{X}_1 = \left\{ x_{jl}^{-} \middle| x_{11opt}^{-} = 0.42; x_{j1opt}^{-} = 0 \text{ for other } j \right\}$
with the lowest total load reduction Z_{opt}^{-} of 1870 kg of inorganic nitrogen. It corresponded to the
highest risk level of violating constraint;

(ii) The second decision alternative in extreme tradeoff analysis was
$\mathbf{X}_2 = \left\{ x_{jl}^{+} \middle| x_{11opt}^{+} = 0.42; x_{51opt}^{+} = 0.71; x_{j1opt}^{+} = 0 \text{ for other } j \right\}$ with the highest total load reduction Z_{opt}^{+} of 2235 kg
of inorganic nitrogen, which corresponded to the lowest risk level of violating constraint;

(iii) The third decision alternative in non-extreme tradeoff analysis was
$\mathbf{X}_3 = \left\{ x_{jl}^{\pm} \middle| \sum_j^8 c_{jl}^{\pm} x_{jl}^{\pm} = 2052.5 \text{ where } x_{11opt}^{\pm} = 0.42; x_{51opt}^{\pm} = [0, 0.71] \text{ and } x_{j1opt}^{\pm} = 0 \text{ for other } j \right\}$ with the expected value of total load
reduction $E[Z^{\pm}]_{opt}$ of 2052.5 kg of inorganic nitrogen, which corresponded to the moderate risk level
of violating constraint;

4. Minimum total load reductions of inorganic phosphorus for Scenario 2

(i) The first decision alternative in extreme tradeoff analysis was
$\mathbf{X}_1 = \left\{ x_{jl}^{-} \middle| x_{12opt}^{-} = 0.62; x_{22opt}^{-} = 0.79; x_{j2opt}^{-} = 0 \text{ for other } j \right\}$ with the lowest total load reduction Z_{opt}^{-} of 406.2 kg of
inorganic phosphorus. It corresponded to the highest risk level of violating constraint;

(ii) The second decision alternative in extreme tradeoff analysis was
$\mathbf{X}_2 = \left\{ x_{jl}^{+} \middle| x_{12opt}^{+} = 0.8; x_{22opt}^{+} = 0.79; x_{j2opt}^{+} = 0 \text{ for other } j \right\}$ with the highest total load reduction Z_{opt}^{+} of 433.6 kg of
inorganic phosphorus, which corresponded to the lowest risk level of violating constraint;

(iii) The third decision alternative in non-extreme tradeoff analysis was $\mathbf{X}_3 = \left\{ x_{jl}^{\pm} \middle| \sum_j^8 c_{jl}^{\pm} x_{jl}^{\pm} = 419.9 \text{ where } x_{12opt}^{\pm} = [0.62, 0.8]; x_{22opt}^{\pm} = 0.79 \text{ and } x_{j2opt}^{\pm} = 0 \text{ for other } j \right\}$ with the expected value of total load reduction $E[Z^{\pm}]_{opt}$ of 419.9 kg of inorganic phosphorus, which corresponded to the moderate risk level of violating constraint.
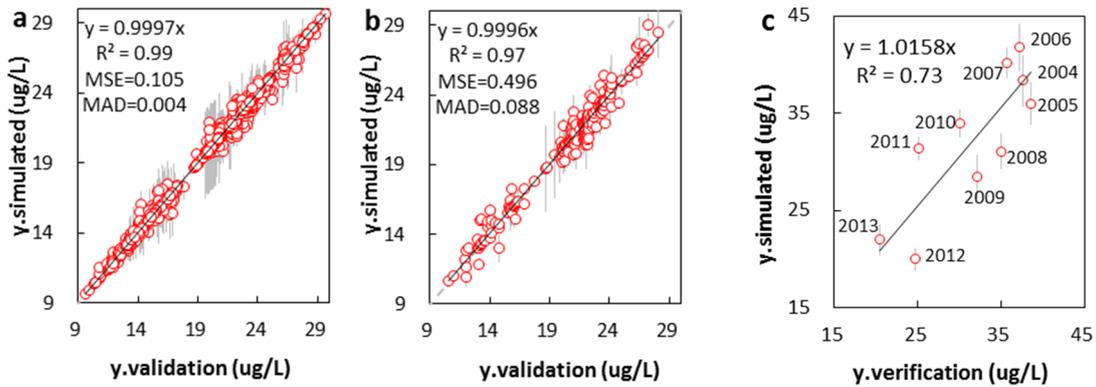
Figure S1. Comparisons on accuracy of BRRT v2 with previous algorithms. (**a**) and (**b**) represent model calibration ($n = 1800$) and validation ($n = 200$), respectively. (**c**) indicates the model verification by 10-year observations from Swift Creek Reservoir Reports.
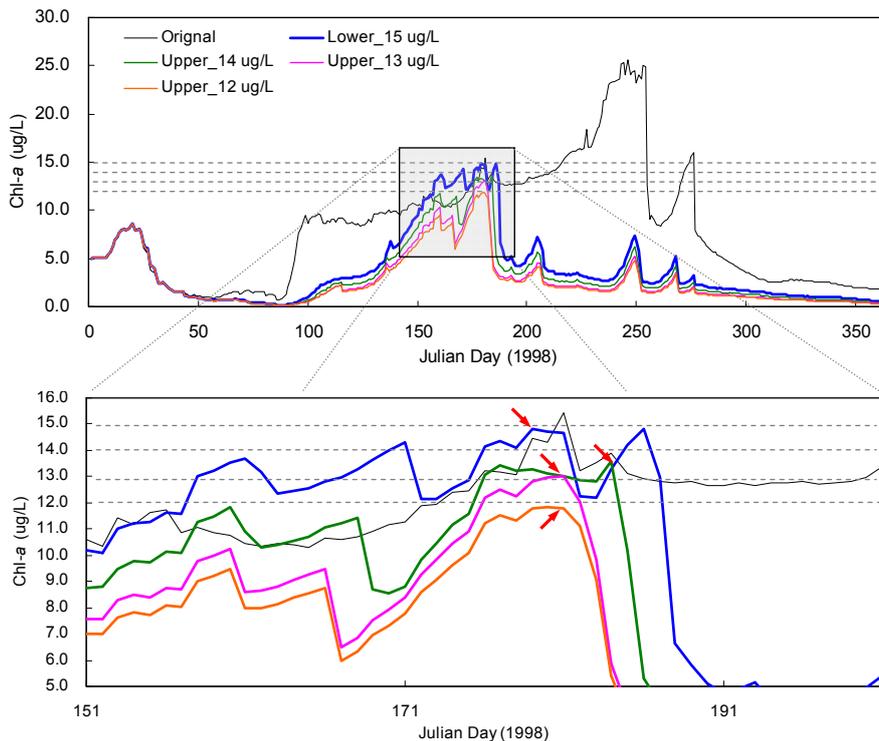
Figure S2. Simulated Chl-*a* concentrations in SCR outlet of optimal load reductions using the CE-QUAL-W2 v3.1. Original line means the simulated Chl-*a* concentrations in SCR outlet without the consideration of nutrient reduction, whereas the other lines indicate the simulated Chl-*a* concentrations under different nutrient reduction scenarios. Arrows indicate the maximum values of Chl-*a* concentrations under different scenarios. The part within gray box is zoomed in below to demonstrate the maximum values of Chl-*a* concentrations.

References

1. Chipman. H.A.; George, E.I.; Mcculloch, R.E. Bayesian treed models. *Mach. Learn* **2002**, *48*, 299–320.
2. Zhou, F.; Huang, G.H.; Guo, H.C.; Zhang, W.; Hao, Z.J.; Spatio-temporal patterns and source apportionment of coastal water pollution in Eastern Hong Kong. *Water Res.* **2007**, *41*, 3439–3449.
3. Jung, M.; Reichstein, M.; Bondeau, A. Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model, *Biogeosciences* **2009**, *6*, 2001–2013.
4. Zhou, F.; Guo, H.C.; Chen, G.X.; Huang, G.H. Interval linear programming: A revisit. *J. Environ. Inform.* **2008**, *11*, 1–10.
5. Zhou, F.; Huang, G.H.; Chen, G.X.; Guo, H.C. Enhanced-interval linear programming. *Eur. J. Oper. Res.* **2009**, *199*, 323–333.
6. Guo, H.C.; Gao, W.; Zhou, F. Three-level trade-off analysis for decision making in environmental engineering under interval uncertainty. *Eng. Optim.* **2014**, *46*, 377–392.