

Combinatorial Optimization for
Parameter Identification of a Rainfall-Runoff Model
Applied to 120 Dam River Basins in Japan
(日本全国 120 ダム流域を対象にした
組合せ最適化による流出モデルのパラメータ同定法)

February 16, 2020

International Course Program of Civil Engineering,
Undergraduate School of Global Engineering, Faculty of Engineering,
Kyoto University

Sosuke ICHIHASHI

ABSTRACT

Basic dataset including basin topography and surface flow direction in 1-sec (30 m) resolution have been developed for the entire Japan, and it becomes possible to apply a distributed hydrologic model to the whole Japan. However, it is not feasible to optimize hydrologic model parameters to numerous river basins individually. Moreover, model parameters cannot be optimized well for river basins without sufficient amount of data in good quality. In recent years, however, much research has focused on parameter optimization methods with a lot of observation data. In this study, we examined combinatorial optimization to identify a limited number of optimal parameter sets that improve the performance of a distributed hydrologic model in dam reservoir basins whole around Japan and explored the relation between basin characteristics and the optimal parameter sets.

We identified an appropriate objective function for the optimization and parameter sets reflecting various runoff mechanisms and performed combinatorial optimization to identify optimal parameter sets among them. Then, the improvements of model performances due to the optimization were measured with different evaluation indices. Lastly, the relation between the optimal parameter sets, discharge characteristics and geographical features were discussed.

The combination of the relative error in peak discharge and Pearson correlation coefficient was identified as the appropriate evaluation index and was used in the objective function. The improvements of model performances with a limited number of optimal parameter sets were confirmed in different evaluation indices for a number of basins, and the results show that optimal parameter sets generally correspond to different discharge characteristics and that there is a correlation between the discharge characteristics and the sum of ratios of granite rock, quaternary volcanic rock and tertiary volcanic rock in bedrock.

CONTENTS

1.	INTRODUCTION	1
2.	METHODS	5
2.1	Rainfall-Runoff-Inundation Model (RRI Model).....	5
2.1.1	Basic Equation and Diffusive Wave Approximation	5
2.1.2	Runoff on The Slope.....	6
2.1.3	Numerical Algorithm.....	7
2.2	Global Optimization for Parameter Identification of a Rainfall-Runoff Model	7
2.3	Combinatorial Optimization for Parameter Identification of a Rainfall-Runoff Model.....	8
2.4	Combinatorial Optimization for Parameter Identification of the RRI Model.....	9
2.4.1	Optimization Settings	9
2.4.2	Selection of Representative Parameter Sets	11
2.4.3	Combinatorial Optimization with the Representative Parameter Sets	13
2.5	Clustering of Dam Reservoir Basins by Default Simulation Performances	16
3.	RESULTS AND DISCUSSIONS.....	17
3.1	Evaluation of the RRI Simulation with the Default Parameter Set.....	17
3.2	Differences in Hydrographs Selected for Different Objective Functions	18
3.3	Best Parameter Set among the 40 Representatives for Each Basin	18
3.4	Combinatorial Optimization Results.....	19

3.5	Runoff Mechanisms and Parameter Sets	20
3.6	Relation between Basin Geology, Optimal Parameter Sets and Discharge Characteristics Categories	23
4.	CONCLUSION.....	27
	APPENDIX.....	28
	Figures and Tables.....	28
	References.....	77
	ACKNOWLEDGEMENT	80

1. INTRODUCTION

Recently, climate extreme events have significant impacts on the human society and are the key issues in dealing with climate change. Increasing trends in precipitation extremes due to the impact of climate change on the hydro-climatology have been found by Solomon et al.¹. For example, Fujibe et al.² found that four-hourly, daily and hourly heavy precipitation increased in Japan in the 20th century. Furthermore, the frequency of precipitation extremes is projected to increase in the future³.

As a result, the frequency and intensity of hydrological hazards have been increased. For example, Typhoon Hagibis hit Japan in October 2019, and more than 21 rivers were flooded. Therefore, flood prediction plays a vital role in maintaining the human society under climate change.

Flood prediction is composed of river discharge predictions based on precipitation data and inundation predictions based on the predicted river discharge data. On the one hand, lumped hydrological models, which deal with each watershed as a single aggregate unit and calculate discharges at a river mouth, have been conventionally used for predictions of river discharges. The lumped models require relatively less computation due to their simple model structures. On the other hand, distributed hydrologic models, which predict hydrological state variables at arbitrary points by incorporating spatially varying land characteristics and precipitation data, require relatively more computation. With the improvement of computational capabilities and spatially distributed data availability as well as model developments, recently, distributed models have been widely used in practice. The lumped and distributed models are called process-based models because they reflect or conceptualize physical processes. In contrast, model based solely on the input data is called data-driven model. For example, regression models such as the artificial neural network model (ANN model) have been utilized to predict river runoff. However, data-

driven models require huge amount of data in good quality for accurate predictions, therefore, they are suitable for river basins with numerous observation data.

Rainfall-Runoff Inundation model (RRI model) was developed by Sayama et al.⁴ and was applied to the Chao Phraya River basin with 1-2 km grid cells. The RRI model is a distributed hydrological model which simulates rainfall-runoff and flood inundation simultaneously. In recent years, basic dataset, which contains basin topography and surface flow direction, has been developed by Yamazaki et al.⁵. The resolution of the data is 1-sec (30 m) for the entire Japan. As a result, it becomes possible to apply the RRI model to the whole Japan in a consistent manner.

Hydrologic model parameters are usually optimized with observation data in order to improve the accuracy. A number of studies have been conducted on optimizations of hydrologic model parameters. For example, Duan et al.⁶ developed the shuffled complex evolution (SCE-UA) method, which is able to effectively optimize hydrologic model parameters. However, it still takes time to optimize hydrologic model parameters, and it requires sufficient amount of observation data such as cross-sections, water levels and discharges of the rivers.

There are two major obstacles in optimizing rainfall-runoff models including the RRI model for many river basins with different sizes. First, enormous computation resources are required to optimize parameters for each basin. In the case of Japan, there are more than 35,000 rivers, so it is computationally too expensive to optimize model parameters for each basin. Second, parameters cannot be optimized well for river basins with few observation data with good quality. In this sense, flood discharge data in dam reservoir basins are suitable for parameter optimizations. However, even if parameter optimizations are successfully conducted for dam reservoir basins, parameters for the other basins cannot be identified. In principle, parameters of distributed models, which model physical processes, should reflect

geographical characteristics in hydrologic phenomena. Thus, if some representative optimal parameter values are found at dam reservoir basins around Japan and are regionalized, parameter values for all the river basins in Japan can be identified.

To deal with this problem, combinatorial optimization⁷ is introduced. Combinatorial optimization uses combinatorial techniques to solve discrete optimization problems, which seek to obtain the best solution from a finite set of options. In other words, an optimal combination is searched effectively from all the given options. By employing combinatorial optimization to identify parameter values of a rainfall-runoff model, a limited number (set to five in this study) of optimal combinations of parameter values (parameter sets) are selected from a number of parameter sets for many river basins based on an evaluation index, and optimal parameter set for each basin selected from the optimal parameter sets is shown. Though the obtained parameter sets are not globally optimal, i.e. different from the results obtained by conventional global optimization conducted in each basin, it can be an effective way to get suitable parameter sets for the dam reservoir basins around Japan while limiting the number of optimal parameter sets. Moreover, since optimal parameter set for each basin is selected from a limited number of optimal parameter sets, each basin can be classified based on its optimal parameter set. Therefore, if the optimal parameter sets and river basin features such as geological, geographical data and the resulting runoff mechanisms are related, it would be possible to assign the selected optimal parameter sets for basins lacking observation data including many parts of the downstream of dam reservoir basins according to geological and geographical conditions. Hence the introduced approach is important not only to develop the whole Japanese flood modeling with parameter optimizations but also to regionalize hydrologic characteristics. Even though they are the most fundamental hydrologic studies, identifying the relation between hydrological characteristics and the geographic features have not been well studied.

Singh⁸ pointed out the effect of basin characteristics, such as surficial soil permeability, geology and topography, on the shapes of flow duration curves in basins in Illinois, while Mimikou and Kaemaki⁹ used multiple regression and regionalized flow duration curves with morphoclimatic characteristics in basins in Greece. Though they estimate discharge characteristics based on basin features, they do not give sufficient consideration on whether the relation is applicable to basins other than the targeted ones. To deal with this issue, the effect of individual factors such as climate, geology, surface soil layer and vegetation on discharge characteristics should be studied for a sufficiently large number of river basins. Mushiake et al.¹⁰ studied the effects of climate and basin geology on flow regime of Japanese mountainous river basins. Soulsby et al.¹¹ studied the effect of surface soil layer on runoff in a basin in Scotland. Burt and Swank¹² examined the effect of vegetation on discharge characteristics in the western North Carolina. Though these studies provide important insights into the relation between basin features and discharge characteristics, most of the studies reviewed so far suffer from the fact that their findings are based on either a limited number of river basins or average daily discharge data, and few have studied runoff mechanisms and its factors in a number of river basins with event-based time scale.

Based on the above background, this study applies a combinatorial optimization technique to the RRI model parameterization focusing on 121 dam reservoir basins positioned whole around Japan. The main objectives of this study are (1) to obtain optimal parameter sets reflecting different runoff mechanisms for 121 dam basins in Japan to enhance the accuracy of river discharge predictions of the RRI model; (2) to evaluate the combinatorial optimization results with different evaluation indices; and (3) to show the relation between runoff mechanisms and geological and geographical features in high flow events.

2. METHODS

2.1 Rainfall-Runoff-Inundation Model (RRI Model)

The RRI model is a two-dimensional grid cell based hydrodynamic model capable of simulating rainfall runoff and flood inundation processes. The overview of the RRI model is shown in Figure 1.

2.1.1 Basic Equation and Diffusive Wave Approximation

There are two model equations used in the RRI model: the continuity equation (1) and two-dimensional shallow water flow equation (2) and (3).

$$\frac{\partial h}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} = r - f \quad (1)$$

$$\frac{\partial q_x}{\partial t} + \frac{\partial uq_x}{\partial x} + \frac{\partial vq_x}{\partial y} = -gh \frac{\partial H}{\partial x} - \frac{\tau_x}{\rho_w} \quad (2)$$

$$\frac{\partial q_y}{\partial t} + \frac{\partial uq_y}{\partial x} + \frac{\partial vq_y}{\partial y} = -gh \frac{\partial H}{\partial y} - \frac{\tau_y}{\rho_w} \quad (3)$$

where h : water depth, q_x : unit width discharge in x direction, q_y : unit width discharge in y direction, r : rainfall intensity, f : infiltration rate, u : flow velocity in x direction, v : flow velocity in y direction, H : water level, τ_x : shear stress in x direction, τ_y : shear stress in y direction, ρ_w : water density.

The second terms of the right side of Equation (2) and (3) are calculated with the Manning's equation.

$$\frac{\tau_x}{\rho_w} = \frac{gn^2 u \sqrt{u^2 + v^2}}{h^{1/3}} \quad (4)$$

$$\frac{\tau_y}{\rho_w} = \frac{gn^2 v \sqrt{u^2 + v^2}}{h^{1/3}} \quad (5)$$

where n : Manning's roughness coefficients for land surface (n_{slope}) and river channels (n_{river}).

Applying the diffusive wave approximation, inertial terms (the left side of Equation (2) and (3)) are neglected.

Furthermore, considering x and y directions separately, the following equations are derived:

$$q_x = -\frac{1}{n} h^{5/3} \sqrt{\left| \frac{\partial H}{\partial x} \right|} \operatorname{sgn} \left(\frac{\partial H}{\partial x} \right) \quad (6)$$

$$q_y = -\frac{1}{n} h^{5/3} \sqrt{\left| \frac{\partial H}{\partial y} \right|} \operatorname{sgn} \left(\frac{\partial H}{\partial y} \right) \quad (7)$$

where sgn : signum function.

The RRI model spatially discretizes the continuity equation (1) as follows:

$$\frac{dh^{i,j}}{dt} + \frac{q_x^{i,j-1} - q_x^{i,j}}{\Delta x} + \frac{q_y^{i-1,j} - q_y^{i,j}}{\Delta y} = r^{i,j} - f^{i,j} \quad (8)$$

where i : row index of grid cells, j : column index of grid cells, $q_x^{i,j}$: discharge in x direction from a grid cell at (i, j) , $q_y^{i,j}$: discharge in y direction from a grid cell at (i, j) .

Combining the equations (6), (7) and (8), water depth and discharge are calculated at each grid cell for each time step.

2.1.2 Runoff on The Slope

The RRI model simulates rainfall runoff with vertical infiltration in flat areas and with saturated subsurface flow in mountainous areas for a better representation. A key aspect of the RRI model is that it can simulate both surface and subsurface flows with the same algorithm. Three settings are available to adjust the model for different land uses: overland flow, vertical infiltration and infiltration excess overland flow, and saturated subsurface and saturation excess overland flow. Replacing Equation (6)

with Equation (9), the model reflects the effect of subsurface flow¹³. The basic equation of saturated subsurface flow in x direction is as follows:

$$q_x = \begin{cases} -k_m d_m \left(\frac{h}{d_m}\right)^\beta \frac{\partial H}{\partial x} & (h \leq d_m) \\ -k_a(h - d_m) \frac{\partial H}{\partial x} - k_m d_m \frac{\partial H}{\partial x} & (d_m < h \leq d_a) \\ -\frac{1}{n}(h - d_m)^{\frac{5}{3}} \sqrt{\left|\frac{\partial H}{\partial x}\right|} \operatorname{sgn}\left(\frac{\partial H}{\partial x}\right) - k_a(h - d_m) \frac{\partial H}{\partial x} - k_m d_m \frac{\partial H}{\partial x} & (d_a < h) \end{cases} \quad (9)$$

where k_a : lateral saturated hydraulic conductivity, β : exponent parameter related to unsaturated hydraulic conductivity, k_m : lateral unsaturated hydraulic conductivity ($k_m = k_a/\beta$), d_m : total thickness of soil layer (D) times unsaturated soil porosity (γ_m), d_a : total thickness of soil layer (D) times effective soil porosity (γ_a). The same equation is applied in the y direction.

2.1.3 Numerical Algorithm

Fifth order Runge-Kutta method with adaptive time step control is applied to solve the continuity equation and two-dimensional shallow water flow equation. This method solves an ordinary differential equation by the general fifth order Runge-Kutta formula and estimates its error by an embedded forth order formula to control the time step.

2.2 Global Optimization for Parameter Identification of a Rainfall-Runoff Model

Performance of a rainfall-runoff model depends on how well it is calibrated, i.e. how well model parameters are adjusted. There have been many attempts to get proper parameter values. For example, Masri et al.¹⁴ proposed “adaptive random search (ARS)” method, which was later modified by Pronzato et al.¹⁵. Another approach called the multistart procedure based on the simplex algorithm (MSX) was suggested by Johnson and Pilgrim to deal with multiple optimization problems. Furthermore, Duan et

al.⁶ presented the shuffled complex evolution (SCE-UA) method and showed that the SCE-UA method outperforms the ARS method and the MSX method in terms of effectiveness in parameter identification of a rainfall-runoff model.

However, it is not feasible to use the SCE-UA method for calibration of a rainfall-runoff model when it is applied to a large number of river basins. What is worse, parameters cannot be optimized well for river basins with limited observation data with good quality. In this sense, flood inflow data into dam reservoirs are suitable for parameter optimizations. Nevertheless, even if parameter optimizations are successfully conducted for dam reservoir basins, parameters for the other basins, for example downstream of the dam reservoir basins, cannot be identified. In principle, parameters of distributed models, which represent physical processes, should reflect geographical characteristics in hydrologic phenomena. Thus, if some representative optimal parameter values are found at dam reservoir basins in the whole Japan and are well regionalized, parameter values for all the river basins in Japan may be identified. To achieve this goal, we propose a combinatorial optimization for parameter identification of a rainfall-runoff model.

2.3 Combinatorial Optimization for Parameter Identification of a Rainfall-Runoff Model

Combinatorial optimization combines techniques from combinatorics and linear programming to solve discrete optimization problems, which seek to obtain the best solution among a finite set of options⁷. In other words, an optimal combination is searched effectively from all the given options. By employing combinatorial optimization to identify parameter values of a rainfall-runoff model, a limited number (set to five in this study) of optimal parameter sets are selected from a number of candidates for all the river basins based on an evaluation index together with an optimal parameter set for each basin. In other words, the method searches a suitable parameter set for each dam reservoir basin while limiting the number of

optimal parameter sets. In order to apply this method, first discrete values are prepared for each model parameter. Then, simulations with combinations of these discrete parameter values (parameter sets) are evaluated based on an objective function. Finally, optimal parameter sets are selected with combinatorial optimization. In this study, we utilize combinatorial optimization for parameter identification of the RRI model in order to obtain five parameter sets for 121 dam river basins in Japan. Based on the identified parameter sets, the relation between the parameters and basin geography are explored.

2.4 Combinatorial Optimization for Parameter Identification of the RRI Model

2.4.1 Optimization Settings

A combinatorial optimization is conducted to identify parameter values of the RRI model for the selected 121 dam river basins in Japan (the selected basins are shown in Table 1). The country is divided into 14 regions by adjusting a typical regional division so that there is not much difference in their areas and that the boundaries do not cross the targeted basins.

The parameters considered are Manning's roughness coefficient for land surface n_{slope} , effective soil porosity γ_a , unsaturated soil porosity γ_m , subsurface and lateral permeability k_a and exponent parameter related to unsaturated hydraulic conductivity β . For each parameter, six discrete values are prepared (values are shown in Table 2). The prepared values include the default parameter values, which represent quick runoff responses by ignoring unsaturated subsurface flow process, of the RRI model for forest area because the most abundant land use in dam river basins in Japan is forest.

The data collection period of rainfall and inflow discharges to the dams is from June 2002 to November 2018. The rainfall data used in this study are a composite product of radar and ground gauged data provided obtained by Japan Meteorological Agency (JMA). The reason why the data collection period

starts in 2002 is that the spatial resolution of the rainfall data product is 5 km before April 2001. The same rainfall periods are selected for all the basins in each region to run the model simultaneously for all the basins in each region. For different regions, different flood periods are selected because flood periods vary from region to region. The targeted flood periods are selected in the following steps.

Step 1: For each dam, sort the dates in the data collection period in the descending order of maximum daily inflow discharges and select candidate flood periods in which inflow discharges are increased. The flood period is set to about five to ten days long to cover increases and decreases in inflow discharges at the dams.

Step 2: For each region, select ten flood periods by giving priority to the one with the largest maximum daily inflow discharges in more dams because there are several dams in each region (the selected periods are shown in Table 3).

As a result, when focusing on specific dams, an increase in inflow discharge does not always occur in all the ten flood periods, and rainfall intensities vary from dam to dam. Checking the data, it was found that there was no actual increase in inflow discharge in about three out of the ten flood periods at each dam. Thus, seven flood periods with the largest total amount of rainfall among them are selected as targeted events for each dam reservoir basin, and they are used for the optimization. Twenty dam reservoir basins are excluded from the scope of the optimization as the number of observed rainfalls is less than seven in those basins.

The objective of the optimization is to obtain five parameter sets from which we can select a single parameter set that improve the performance of discharge prediction for all the precipitation events in each basin.

Unfortunately, it is not feasible to identify five optimal parameter sets from the given parameter sets with combinatorial optimization. Since there are five parameters with six discrete values each, the number of given parameter sets is 6^5 . To evaluate parameter sets in combinatorial optimization, the RRI model has to be run with every parameter set in every basin, so the model is run 6^5 times for each event in each basin. Thus, it is computationally too expensive to use combinatorial optimization with all the given parameter sets. To deal with this issue, we first reduce the number of parameter sets by selecting representative parameter sets while conserving parameter sets reflecting a variety of runoff mechanisms. To identify five optimal parameter sets with combinatorial optimization in a reasonably short time, 40 representative parameter sets are selected out of 6^5 .

2.4.2 Selection of Representative Parameter Sets

In order to reduce the number of parameter sets while keeping model responses related to different runoff mechanisms, parameter sets leading to similar simulation results are clustered accordingly.

First of all, two dam river basins, Hiyoshi and Shimouke, are selected to classify the parameter sets by their simulation results. The reason why they are selected is that they have different geologic features, which can lead to different runoff mechanisms. Hiyoshi's bedrock is mainly composed of accretionary wedge and sedimentary rock (Figure 2), while Shimouke's bedrock is mostly igneous rock (Figure 3). One single-peak precipitation event and one double-peak precipitation event are selected for each basin because they can also result in different runoff mechanisms.

Then, the RRI model is run for each event with all the given parameter sets. The resulting four hydrographs of each parameter set are discretized as shown in Figure 4. In other words, we take a discharge value every hour and make a vector consisting of those values for each parameter set. In this

manner, simulation results of each parameter set for the four events turn into a point in a high dimensional space.

Finally, all the parameter sets are clustered in that high dimensional space into 40 clusters with the k-means algorithm with the k-means++ initialization algorithm in Euclidean distance¹⁶. The k-means algorithm begins with a random set of cluster centers, and it iteratively assigns points to clusters and recomputes cluster centers until the centers no longer change. However, there are many cases in which the k-means algorithm produces arbitrarily bad clustering results due to its initialization process. The k-means++ initialization algorithm chooses the centers one by one while current chosen centers stochastically affects the choice of the next center. In this manner, the k-means++ improves the accuracy of the k-means clustering. Arthur and Vassilvitskii¹⁶ showed that the k-means++ improves the accuracy of the k-means by at least 10%, and it often performs much better. As a result of the clustering, parameter sets with similar hydrographs are assigned to the same cluster. The parameter set closest to each cluster center is selected as the representative of that cluster.

The obtained 40 representative parameter sets are shown in Table 4. The representative parameter sets are sorted in the descending order of the sum of the simulated peak discharges for the four events. Therefore, parameter sets with smaller index numbers tend to result in larger simulated peak discharges. The hydrographs of the discharges simulated with the representative and all the given parameter sets and observed discharges are shown in Figure 5. The selected parameter sets seem to represent different runoff mechanisms and mostly cover the observation data and simulations of the given parameter sets.

2.4.3 Combinatorial Optimization with the Representative Parameter Sets

Combinatorial optimization is conducted on the 40 parameter sets selected in the previous step. The objective of the optimization is to obtain five parameter sets from which we can select a single parameter set that improve the performance of the RRI model for all the seven targeted events in each basin.

First of all, the RRI model is run with the 40 parameter sets for the seven targeted precipitation events in each basin. And the simulation results are evaluated with the observation discharge data and an evaluation index. The evaluation indices examined are the Pearson correlation coefficient (CC), the Kling-Gupta efficiency¹⁷ (KGE), the Nash-Sutcliffe efficiency (NSE) and the relative error in peak discharge (EPD). The indices are calculated with the following equations.

$$CC = \frac{\sum_{t=1}^T (Q_o^t - \bar{Q}_o)(Q_s^t - \bar{Q}_s)}{\sqrt{(\sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2)(\sum_{t=1}^T (Q_s^t - \bar{Q}_s)^2)}} \quad (10)$$

$$KGE = 1 - \sqrt{(CC - 1)^2 + (\beta - 1)^2 + (\alpha - 1)^2} \quad (11)$$

$$\beta = \frac{\bar{Q}_s}{\bar{Q}_o} \quad (12)$$

$$\alpha = \frac{\bar{Q}_s / \sqrt{\frac{1}{T} \sum_{t=1}^T (Q_s^t - \bar{Q}_s)^2}}{\bar{Q}_o / \sqrt{\frac{1}{T} \sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2}} \quad (13)$$

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_s^t - Q_o^t)^2}{\sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2} \quad (14)$$

$$EPD = \frac{Q_{p,s} - Q_{p,o}}{Q_{p,o}} \quad (15)$$

where Q_o^t : observed discharge at time t , Q_s^t : simulated discharge at time t , \overline{Q}_o : mean observed discharge in an event, \overline{Q}_s : mean simulated discharge in an event, β : bias ratio, α : variability ratio, $Q_{p,s}$: simulated peak discharge, $Q_{p,o}$: observed peak discharge.

And they are standardized for better performances of the optimization. For the CC , KGE and NSE , values less than 0 are replaced by 0 (the standardized indices are denoted as CC_{std} , KGE_{std} , NSE_{std} , respectively). For the EPD , values more than 1 is replaced by 1, then, the standardized EPD (EPD_{std}) is obtained with the following equation because the indices are maximized in this combinatorial optimization problem.

$$EPD_{std} = 1 - |EPD| \quad (16)$$

The optimizations with the CC_{std} , KGE_{std} , NSE_{std} , EPD_{std} , and their combinations are examined.

Now, the optimal parameter sets are selected by solving the combinatorial optimization problems formulated below.

$$\max \sum_{i=1}^{N_{basin}} \sum_{m=1}^{N_{param}} y_{i,m} \times score_{i,m} \quad (17)$$

s.t.

$$\sum_{m=1}^{N_{param}} y_{i,m} = 1 \quad (i = 1, 2, \dots, N_{basin}) \quad (18)$$

$$y_{i,m} \leq x_m \quad (i = 1, 2, \dots, N_{\text{basin}}; m = 1, 2, \dots, N_{\text{param}}) \quad (19)$$

$$\sum_{m=1}^{N_{\text{param}}} x_m \leq N_{\text{opt}} \quad (20)$$

where,

$$y_{i,m} = \begin{cases} 0 & (\text{do not select parameter set } m \text{ for dam } i) \\ 1 & (\text{select parameter set } m \text{ for dam } i) \end{cases} \quad (21)$$

$$x_m = \begin{cases} 0 & (\text{parameter set } m \text{ is not in selected parameter sets}) \\ 1 & (\text{parameter set } m \text{ is in selected parameter sets}) \end{cases} \quad (22)$$

$$\text{score}_{i,j,m} = (\text{evaluation index value for dam } i \text{ event } j \text{ paraemter set } m) \quad (23)$$

$$\text{score}_{i,m} = \sum_{j=1}^{N_{\text{event}}} \text{score}_{i,j,m} \quad (i = 1, 2, \dots, N_{\text{basin}}; m = 1, 2, \dots, N_{\text{param}}) \quad (24)$$

$$N_{\text{basin}} = 101 \quad (\text{number of dam river basins}) \quad (25)$$

$$N_{\text{event}} = 7 \quad (\text{number of events}) \quad (26)$$

$$N_{\text{param}} = 40 \quad (\text{number of parameter sets before optimization}) \quad (27)$$

$$N_{\text{opt}} = 5 \quad (\text{number of parameter sets to be selected in optimization}) \quad (28)$$

The constraint (18): One parameter set is selected for each basin.

The constraint (19): Parameter sets are selected only from the selected parameter sets.

The constraint (20): The number of selected parameter sets is less than N_{opt} .

Performances of two linear problem solvers are evaluated: Gurobi Optimizer and MATLAB. Their calculation times are compared on combinatorial optimization problem with different initial numbers of parameter sets. All the experimental results reported are conducted on Windows 10 Pro with two Intel Xeon Gold 6134 CPUs with the clock speed of 3.20 [GHz] and 192 [GB] of RAM. The median calculation time of ten trials for each initial number of parameter sets is shown in Table 5 and Figure 6. Since Gurobi solves problems in a shorter time than MATLAB, Gurobi is used to conduct the combinatorial optimization. As shown in Figure 6, it is expected that the problem gets computationally expensive if the initial number of parameter sets is fifty or larger. And that is why forty representative parameter sets are selected in the previous section.

2.5 Clustering of Dam Reservoir Basins by Default Simulation Performances

By classifying dam reservoir basins based on the performance of the RRI model with the default parameter set, the performance of the combinatorial optimization and the relation between the model parameter values and dam reservoir basin features can be understood deeply. Therefore, the basins are classified with the EPD_{std} and CC_{std} of the RRI model with the default parameter set for forest (the term ‘default parameter set’ will be used solely when referring to the default parameter set for forest).

Firstly, the RRI model is run with the default parameter set for forest for the seven targeted events in each dam reservoir basin. Secondly, the EPD_{std} and CC_{std} are calculated for each event in each basin. Thirdly, each basin is represented by a point in a 14-dimensional space where the axes are EPD_{std} for each event in the descending order of the total amount of rainfall and CC_{std} for each event in the descending order of the total amount of rainfall, respectively. Finally, the basins are clustered into four clusters by the k-means algorithm with the k-means++ initialization algorithm in Euclidian distance¹⁶.

3. RESULTS AND DISCUSSIONS

3.1 Evaluation of the RRI Simulation with the Default Parameter Set

The RRI model with the default parameter set is run for the seven targeted events for each dam reservoir basin. And the average of the scores for the events is calculated with CC_{std} , KGE_{std} , NSE_{std} and EPD for each basin. The values for all the basins are shown in Table 6. Moreover, their spatial distributions are shown in Figure 7, Figure 8, Figure 9 and Figure 10. Note that values for basins with less than seven events are excluded and EPD values more than 1 are replaced by 1.

The EPD is the relative error between the simulated peak discharge and observed peak discharge, therefore, it gets close to 0 if the simulated peak discharges match the observed ones well. As shown in Table 6 and Figure 10, the EPD gets far from 0, which means the model overestimates the peak discharges, in most of the basins in Hokkaido and Kanto regions. The CC evaluates how well the simulated fluctuations match the observed ones. As shown in Table 6 and Figure 7, the CC is more than 0.80 in most of the basins, therefore, the RRI model with the default parameter set simulate the fluctuations in the discharges well. However, the CC is lower in Hokkaido and Kanto regions, so there is still some room to improve. The KGE and NSE have both aspects of the indices mentioned above, i.e. they evaluate both fluctuations and discharge errors. Therefore, these scores for the basins in Hokkaido and Kanto regions are also low as shown in Table 6, Figure 8 and Figure 9. To see the performances of these evaluation indices, the combinatorial optimizations with KGE_{std} , NSE_{std} and the sum of the CC_{std} and EPD_{std} ($EPDCC$) are conducted below to determine the suitable evaluation index for the optimization.

3.2 Differences in Hydrographs Selected for Different Objective Functions

The optimizations with the KGE_{std} , NSE_{std} , and $EPDCC$ are conducted, and different optimal parameter sets are selected for different evaluation indices used. The obtained simulations with these optimal parameter sets in Hiyoshi (event 1 and 5) and Shimouke (event 1 and 7) are shown in Figure 11, Figure 12, Figure 13 and Figure 14. The simulations with the optimal parameter sets of the combinatorial optimizations with the NSE_{std} and $EPDCC$ match the observation better than the one with the KGE_{std} for the event 1 in Hiyoshi, while the one with the KGE_{std} is better than the others for the event 5 in Hiyoshi. Therefore, from these two cases, it is not clear which evaluation index is suitable for being used in the objective function of the combinatorial optimization. However, the simulation with the optimal parameter sets of the combinatorial optimization with the KGE_{std} and NSE_{std} largely underestimate the peak discharges for the events in Shimouke. It may be because these simulations match the observations better than the ones with $EPDCC$ at the baseflow recession parts. Though the KGE and NSE are widely used evaluation indices for rainfall-runoff models, they do not put sufficiently large weight on the peak discharge error, and they tend to give good scores even if the peak discharge error is not negligibly small. Thus, the $EPDCC$ is selected to be used as the evaluation index of this study as we find it matches our goal most among the tested indices. Though the KGE_{std} and NSE_{std} are not used in the combinatorial optimization, they are used to evaluate the combinatorial optimization results.

3.3 Best Parameter Set among the 40 Representatives for Each Basin

To understand how well the 40 representative parameter sets reflect various runoff mechanisms and improve the simulations, parameter sets that achieve the largest sum of the $EPDCC$ for the seven targeted among the representative parameter sets for each basin are obtained as shown in Table 7 (the term ‘opt parameter sets’ will be used solely when referring to these parameter sets). And the spatial distribution

is shown in Figure 15. Moreover, the CC_{std} , KGE_{std} , NSE_{std} and EPD for their simulations are shown in Table 8. And their spatial distributions are shown in Figure 16, Figure 17, Figure 18 and Figure 19. As previously mentioned, the representative parameter sets are sorted in the descending order of the sum of the simulated peak discharges in the four targeted events. Thus, parameter sets with smaller ID numbers tend to result in simulations with larger peak discharges. Comparing the index scores of the simulations of the default and opt parameter sets, the simulation results are improved with the opt parameter sets in most of the basins. However, the evaluation index scores are not much improved from the default parameter set in some basins. It may be because some of them have upstream dam reservoirs that can affect the inflow discharges. For example, Amagase dam in Kinki region has Seta River Araizeki weir in the upstream. Unfortunately, the selected opt parameter sets vary from basin to basin, and it is hard to regionalize the opt parameter sets. Therefore, the number of parameter sets should be limited with the combinatorial optimization. The opt parameter set and index score for each basin is used to evaluate the performance of the parameter set selected in the combinatorial optimization (the term ‘c_opt parameter set’ will be used solely when referring to the optimal parameter sets selected in the combinatorial optimization).

3.4 Combinatorial Optimization Results

The optimal parameter sets obtained by the combinatorial optimization are shown in Table 9. The selected c_opt parameter set for each basin is shown in Table 10, and their spatial distributions are shown in Figure 20. Also, the average CC_{std} , KGE_{std} , NSE_{std} and EPD of their simulations for the seven targeted events are shown in Table 11, and their spatial distributions are shown in Figure 21, Figure 22, Figure 23 and Figure 24. The EPD has been improved in most of the basins even in Haokkaido and Kanto region, where EPD is large with the default parameter set. The KGE_{std} and NSE_{std} are also

improved in many basins. Though some basins still have relatively low scores, those are the ones even the opt parameter sets have low scores. To examine the overall improvements of each evaluation index and to compare it with the ones of the default and opt parameter sets, the cumulative relative frequency plots of evaluation index scores for all the targeted events for the default, opt and c_opt parameter sets are shown in Figure 25. To see how the number of parameter sets selected in the combinatorial optimization affect the model performances, the combinatorial optimization is also conducted to obtain ten c_opt parameter sets, and its performance is plotted in the same figure (to avoid disambiguation, the terms ‘c_opt(5) parameter set’ and ‘c_opt(10) parameter set’ will be used when referring to the optimal parameter sets selected in the combinatorial optimization aiming to select five and ten optimal parameter sets, respectively; when the term ‘c_opt parameter set’ is used, it refers ‘c_opt(5) parameter set’).

The plots indicate that the c_opt(5) parameter sets improve the evaluation index scores compared to the default parameter set, and the scores are close to the ones of the c_opt(10) and opt parameter sets, which are selected without any limitations on the total number. Therefore, it seems that the c_opt(5) parameter sets achieve almost as high model performance as the c_opt(10) and opt parameter sets. Thus, the performance will not be largely enhanced even if the number of parameter sets selected is increased from five. Also, it seems that the CC_{std} is not well improved as the KGE_{std} , NSE_{std} and EPD_{std} through the combinatorial optimizations. It may be because there is not much room left for the improvements as the CC_{std} is already large for many basins before the optimizations.

3.5 Runoff Mechanisms and Parameter Sets

To deeply understand the model performance with c_opt parameter sets and the relation between the model parameter values and dam reservoir basin features, the basins are classified with the EPD_{std} and CC_{std} of the RRI model with the default parameter set for the seven targeted events. The dam reservoir

basins and the clusters to which they are clustered are shown in Table 12. The centroids of the clusters are shown in Table 13. The centroid of the cluster 1 has relatively small EPD and large CC_{std} for all the events. Since cluster centroids show the typical features of the cluster members, it implies that the RRI model with the default parameter set can simulate precipitation events well regardless of the total amount of rainfall for the dam reservoir basins in the cluster 1. In contrast, for the centroid of the cluster 2, the EPD gets larger, and the CC_{std} gets smaller as the total amount of rainfall of events decreases. Thus, the RRI model with the default parameter set can simulate precipitation events with the large total amount of rainfall well, but it starts overestimating the river discharges as the total amount of rainfall decreases for the basins in the cluster 2. Additionally, the centroid of the cluster 3 has large EPD and small CC_{std} for all the targeted events, while the centroid of the cluster 4 has negative EPD and small CC_{std} for all the targeted events. It indicates that the RRI model with the default parameter set overestimates the river discharges regardless of the total amount of rainfall of precipitation events for the basins in cluster 3, while it underestimates the river discharges regardless of the total amount of rainfall for the basins in cluster 4. Therefore, all the basins can be categorized into four categories corresponding to the four clusters (Category 1, 2, 3 and 4) based on the clusters they belong to. And the basins in the same categories may have similar runoff mechanisms. The spatial distributions of the category members are shown in Figure 26. Since most of the dam reservoir basins in the Category 4 is in Okinawa region, there may be some regional factor for the underestimation by the RRI model with the default parameter set such as inaccurate rainfall observations in that region. Though our goal is to select as less c_{opt} parameter sets as possible, it is reasonable to obtain at least about five c_{opt} parameter sets because the clustering result implies that there are at least four different discharge characteristics in Japan.

To see the optimization performance in each category, the cumulative relative frequency plots of the evaluation index scores in all the targeted events for the default, opt and c_opt parameter sets are shown by categories in Figure 27, Figure 28, Figure 29 and Figure 30. The plots show that the combinatorial optimizations increase the evaluation index scores in all the categories. Comparing the plots of the Category 1 and 2, the c_opt(5), c_opt(10) and opt parameter sets increase the evaluation scores to some extent for basins in the Category 1, but they increase the scores much more for basins in the Category 2. The possible reason is the following. Since the evaluation scores are already high for most of the targeted events under the default settings for the basins in the Category 1, there is not much room to improve for the basins in the Category 1. For the basins in the Category 2, the scores are lower for the events with the smaller total amount of rainfall. Therefore, there is larger room to improve the score for the Category 2 than the Category 1.

The ratio of the c_opt parameter sets selected for the basins in each category is shown in Figure 31. As mentioned above, parameter sets with lower index numbers tend to result in larger peak discharges. Thus, it makes sense that most of the c_opt parameters in the Category 4 is the parameter set 2, which tends to result in larger peak discharges than the default, as the RRI model with the default parameter set usually underestimates the river discharges in the basins in the Category 4. Moreover, the majority of the c_opt parameter sets in the Category 1 are the parameter set 2 and 17, and it may be because parameter sets resulting in the similar simulations to the default parameter set are selected as the RRI model with the default parameter set already simulates the river discharges well for the Category 1. Comparing the ratios of the c_opt parameter sets in the Category 1, 2 and 3, the c_opt parameter sets in the Category 1 are mainly the ones with relatively small index numbers, while the c_opt parameter sets in the Category 3 are mainly the ones with relatively large index numbers, and the ratios in the Category 2 is somewhere

between them. The possible reason is that the basins in the Category 1 have quicker runoff responses, while the ones in the Category 3 have slower runoff responses than the others.

3.6 Relation between Basin Geology, Optimal Parameter Sets and Discharge Characteristics Categories

When comparing basin geology and discharge characteristics, geology classification reflecting different runoff mechanisms should be used. Yokoo and Oki¹⁸ discuss which surface geology classification best explain the shape of flow duration curves in Japanese mountainous basins among five different geology classifications and argue that Mushiake's classification and classification based on age are relatively effective. Therefore, we utilize Mushiake's classification, where rocks are categorized into quaternary volcanic rock (Vq), tertiary volcanic rock (Vt), granite rock (G), tertiary deposit (T), Mesozoic deposit (M) and Paleozoic deposit (P). There have been several attempts to see the relation between basin geology and discharge characteristics by using Mushiake's classification. Mushiake et al.¹⁰ investigated the effect of basin geology on flow regime in several river basins in Kanto region and argue that low flow discharge is largest in basins mainly composed of quaternary volcanic rock, followed in order by those of granite rock and tertiary volcanic rock and those of Mesozoic deposit and Paleozoic deposit. They also investigated the effect at 124 gauging stations in Japan and argue that annual maximum discharge is smallest in basins mainly composed of quaternary volcanic rock, followed in order by those of granite rock, those of tertiary volcanic rock, those of Mesozoic deposit and those of Paleozoic deposit. Moreover, Jitousono and Takeshita¹⁹ added metamorphic rock to the classification, analyzed the effect of basin geology on flow regime based on average daily discharge data in 163 dam reservoir basins in Japan and argue that function of head water recharge, which corresponds to water-holding capacity, is highest in basins mainly composed of quaternary volcanic rock, followed in order by those of tertiary

volcanic rock, granite rock and tertiary deposit, those of Paleozoic deposit and metamorphic rock and those of Mesozoic deposit. What is more, Kato et al.²⁰ monitored rainfall and runoff in basins upstream of Ibi river in Japan mainly composed of granite rock, Mesozoic deposit and Paleozoic deposit and found that lag time (runoff peak response time) is relatively short, recession limb is steep, and discharge is constant in granite rock basins, while in Mesozoic deposit and Paleozoic deposit basins, there are two different discharge characteristics: one has relatively long lag time and inconstant discharges, while the other has relatively short lag time and constant discharges. Though these studies provide important insights into the relation between basin and discharge characteristics, the studies reviewed so far suffer from the fact that their findings come from either a limited number of river basins or daily discharge data. Therefore, the present research may generalize or provide deeper insights into the effect of basin geology on discharge characteristics by exploring it with hourly discharge data in a number of dam reservoir basins positioned whole around Japan.

Since both selected optimal parameter set and discharge characteristics category explain the runoff mechanism of each basin, the relation between basin geology and discharge characteristics may be shown by looking at the basin geology in each category or for each optimal parameter set. Thus, the ratio of V_q, V_t, G, T, M and P is calculated for each category in the following steps. Firstly, the bedrock geology data are obtained for all the dam reservoir basins by clipping the seamless digital geological map of Japan 1:200,000²¹ for the basin. Secondly, in each basin, the area of each rock type is calculated, and their ratios are calculated based on the areas. Lastly, for each category, the ratio of each rock type is obtained by taking average of its ratios in all the basins in that category. The result is shown in Figure 32. It seems that the average ratio of Mesozoic deposit is larger in Category 1 than the other categories, while the sum of the average ratios of quaternary volcanic rock, tertiary volcanic rock and granite rock

is larger in Category 2 and 3. Since the RRI model with the default parameter set tend to perform well for basins in Category 1 and overestimate river discharges for basins in Category 2 and 3, the result may imply that the RRI model with the default parameter set tend to perform well for basins largely composed of Mesozoic deposit, while it overestimates river discharges as the sum of the ratios of quaternary volcanic rock, tertiary volcanic rock and granite rock gets larger.

Further, the ratio of Vq, Vt, G, T, M and P for each c_opt parameter set is shown in Figure 33. Though the parameter set 2 seems to be an exception, it seems that basins with c_opt parameter sets resulting in smaller peak discharges tend to have larger sum of the ratios of quaternary volcanic rock, tertiary volcanic rock and granite rock, while basins with c_opt parameter sets reflecting quicker runoff processes tend to have relatively large ratio of Mesozoic deposit. Therefore, it corresponds to the result shown in Figure 32.

The results broadly support the findings of Mushiake et al.¹⁰ that annual maximum discharge is smallest in basins mainly composed of quaternary volcanic rock, followed in order by those of granite rock, those of tertiary volcanic rock, those of Mesozoic deposit and those of Paleozoic deposit. Also, the results may imply that basins mainly composed of Mesozoic deposit have relatively low water-holding capacity and lead to relatively quick runoff, while those of quaternary volcanic rock, tertiary volcanic rock and granite rock have relatively high water-holding capacity and lead to relatively slow runoff. The possible reasons are the following. Quaternary volcanic rock, tertiary volcanic rock and granite rock tend to have relatively high permeability or more voids and fractures, and some part of water permeates through their fractures and is kept for relatively long time. In addition, basin slope may be relatively gentle in the basins with quaternary volcanic rock, tertiary volcanic rock and granite rock as argued by Jitousono and Takeshita²², therefore, it leads to relatively high water-holding capacity in those basins. If this is the case,

this finding is consistent with the result obtained by Jitousono and Takeshita¹⁹ that water-holding capacity is highest in basins mainly composed of quaternary volcanic rock, followed in order by those of tertiary volcanic rock, granite rock and tertiary deposit, those of Paleozoic deposit and metamorphic rock and those of Mesozoic deposit. This also accords with the finding of Mushiake et al.¹⁰ that low flow discharge is largest and recession limb is most gentle in basins mainly composed of quaternary volcanic rock, followed in order by those of granite rock and tertiary volcanic rock and those of Mesozoic deposit and Paleozoic deposit because low flow discharge and recession limb of hydrograph largely depend on base flow that can have positive correlations with water-holding capacity of bedrocks. Last but not least, this hypothesis is consistent with the fact that the RRI model with the default parameter set overestimates river discharges as the total amount of rainfall decreases for the basins in Category 2, which bedrock is largely composed of quaternary volcanic rock, tertiary volcanic rock and granite rock, for the same reason. Overall, this study confirms the relation between basin geology and discharge characteristics found in previous studies, which focus on a limited number of river basins or daily discharge data, are generally applicable to most of the river basins around Japan.

4. CONCLUSION

We conducted combinatorial optimization for parameter identification of a distributed rainfall-runoff model targeting 121 basins positioned whole around Japan to enable the model to be applied to all the river basins in Japan with suitable parameter values. This study has identified that the combination of the Pearson correlation coefficient and relative peak discharge error is a suitable evaluation index for simulated river discharges when peak discharge should be a key component besides fluctuations of discharge in the evaluation. This study also identified that five parameter sets improve the model performance in the 121 basins which may have different runoff mechanisms. In general, therefore, a limited number of parameter sets may improve performances of a rainfall-runoff model in various basins. Further, the results show that the parameter sets largely improve the model performances in the basins where the model with the default parameter values overestimate the river discharges as the total amount of rainfall gets small. An implication of this is possibly that a rainfall-runoff model performs well with suitable parameter sets reflecting runoff mechanisms of targeted basins regardless of the total amount of rainfall. Further, the results show that parameter set resulting in relatively small discharges is selected for the basins with larger sum of granite rock, quaternary volcanic rock and tertiary volcanic rock in the bedrocks. It strengthens the idea that basin geology is an important factor in runoff mechanisms.

Whilst this study did not confirm that suitable parameter values of a rainfall-runoff model can be determined based on the basin characteristics, it did partially substantiate the relation between suitable parameter set and basin geology, and it shows the effectiveness of combinatorial optimization in parameter identification of a rainfall-runoff model. Yet, greater efforts are needed to clarify the relation between basin features and discharge characteristics. A reasonable approach to tackle this issue could be exploring the effect of surface soil, bedrock and basin topographies on discharge characteristics.

APPENDIX

Figures and Tables

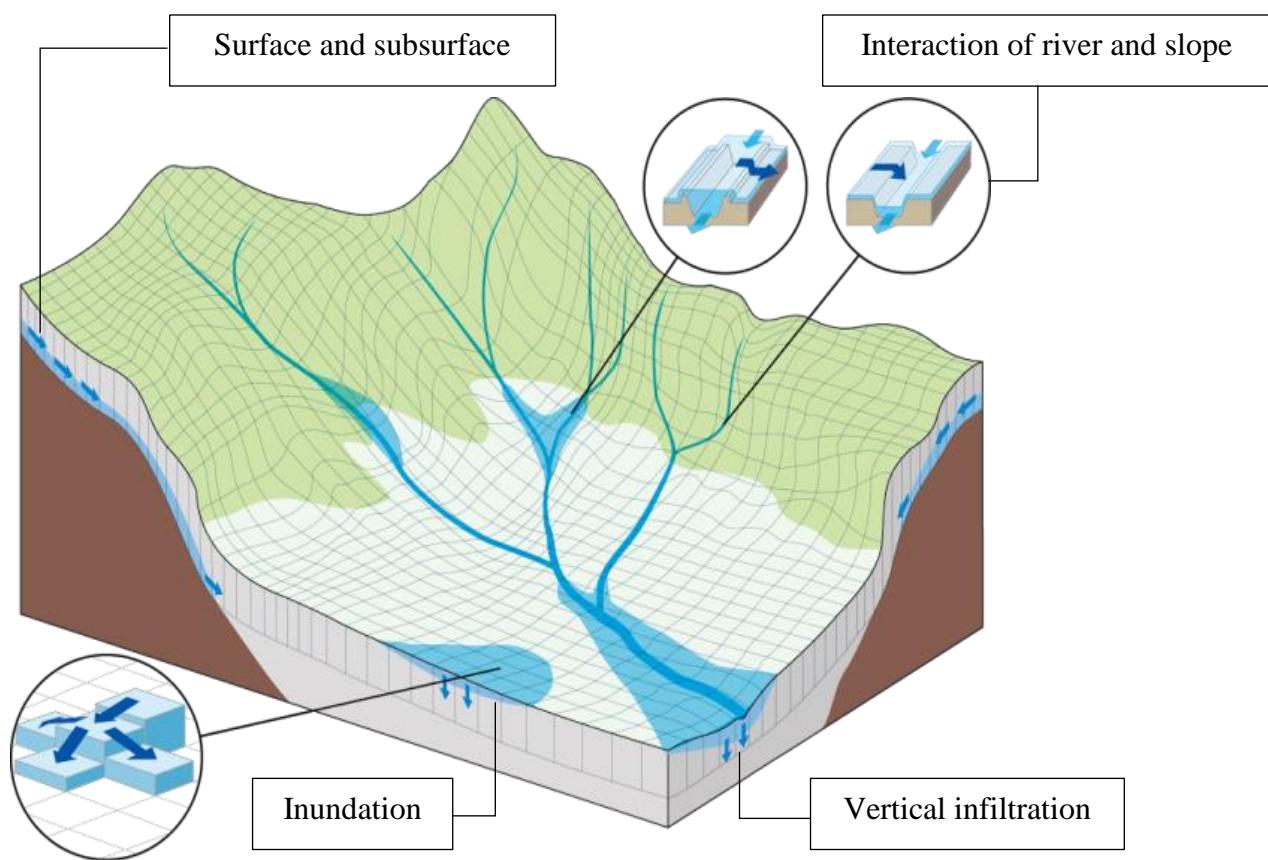


Figure 1: Overview of Rainfall-Runoff-Inundation model (RRI model)

Table 1: Selected 121 dam reservoir basins in Japan

Region		Dam Name			
Hokkaido_1	Kanoko	Satsunaigawa	Tokachi		
Hokkaido_2	Chubetsu Kanayama	Hoheikyo Katsurazawa	Iwaonai Rumoi	Izarigawa Taisetsu	Jozankei Takisato
Hokkaido_3	Nibutani	Pirika			
Tohoku		Aseishigawa Kamafusa Shichikashuku Yuda	Gassan Miharu Shijushida	Gosho Nagai Shirakawa	Isawa Naruko Tamagawa
Kanto		Aimata Kawamata Sonohara Yunishigawa	Fujiwara Kusaki Takizawa	Futase Miyagase Urayama	Ikari Naramata Wataraseyusuichi
Hokuriku_1	Oishi	Okawa	Omachi	Sagurigawa	
Hokuriku_2	Kuzuryu	Managawa	Tedorigawa	Unaduki	
Chubu	Agigawa	Hachisu	Iwaya	Koshibu	Makio
	Maruyama	Misogawa	Miwa	Nagashima	Origawa
	Shintoyone	Tokuyama	Ure	Yahagi	Yokoyama
Kinki	Amagase	Hinachi	Hitokura	Hiyoshi	Muro
	Nunome	Otaki	Sarutani	Shorenji	Takayama
Chugoku	Haiduka	Haji	Hattabara	Nukui	Obara
	Shimadigawa	Shitsumi	Sugesawa	Tomata	Tono
	Yasaka				
Shikoku	Ikeda	Ishitegawa	Kanogawa	Nagayasuguchi	Nakasujigawa
	Nomura	Odo	Sameura	Shingu	Tomisato
	Yanase				
Kyushu	Egawa	Kasegawa	Kyuragi	Matsubara	Midorikawa
	Ryumon	Shimouke	Terauchi	Tsuruda	Yabakei
Okinawa	Aha	Arakawa	Benoki	Fukuji	Fungawa
	Haneji	Kanna	Kurashiki	Taiho	

Table 2: Discrete parameter values (the default values for forest are underlined)

Notation	Unit	Parameter values					
n_{slope}	$m^{-1/3}s$	0.20	0.30	0.40	0.50	<u>0.60</u>	0.70
γ_a	-	0.30	0.35	0.40	<u>0.45</u>	0.50	0.55
γ_m	-	<u>0.00</u>	0.05	0.10	0.15	0.20	0.25
k_a	m/s	0.030	0.050	0.075	<u>0.100</u>	0.150	0.200
β	-	<u>4</u>	5	6	7	8	9

Table 3: Start and end time of precipitation events

All the observation data is from 1 am of the start date to 11:30 pm of the end date.

Region	ID	Start Date	End Date	Region	ID	Start Date	End Date
Hokkaido_1	1	8/16/2016	8/25/2016	Chubu	1	9/1/2018	9/7/2018
	2	7/2/2018	7/8/2018		2	9/27/2018	10/3/2018
	3	9/15/2017	9/21/2017		3	9/17/2011	9/23/2011
	4	8/28/2016	9/3/2016		4	8/31/2011	9/6/2011
	5	8/8/2014	8/14/2014		5	7/11/2010	7/18/2010
	6	8/31/2011	9/9/2011		6	10/19/2017	10/26/2017
	7	8/16/2006	8/22/2006		7	10/27/2017	11/2/2017
	8	1/10/2012	1/16/2012		8	7/3/2018	7/10/2018
	9	5/23/2010	5/29/2010		9	4/22/2018	4/28/2018
	10	12/24/2006	12/30/2006		10	9/16/2016	9/25/2016
Hokkaido_2	1	8/31/2011	9/9/2011	Kinki	1	9/13/2013	9/19/2013
	2	8/19/2016	8/28/2016		2	8/31/2011	9/6/2011
	3	8/21/2010	8/27/2010		3	9/18/2011	9/24/2011
	4	9/8/2012	9/14/2012		4	10/19/2017	10/25/2017
	5	6/30/2018	7/9/2018		5	8/6/2014	8/12/2014
	6	5/8/2006	5/14/2006		6	8/13/2014	8/19/2014
	7	5/1/2012	5/7/2012		7	9/26/2018	10/5/2018
	8	8/13/2011	8/19/2011		8	8/21/2018	8/27/2018
	9	8/3/2014	8/9/2014		9	10/5/2009	10/11/2009
	10	9/8/2014	9/14/2014		10	9/27/2012	10/4/2012
Hokkaido_3	1	11/8/2009	11/14/2009	Chugoku	1	7/2/2018	7/11/2018
	2	8/16/2006	8/22/2006		2	10/19/2017	10/26/2017
	3	9/16/2017	9/23/2017		3	8/31/2011	9/6/2011
	4	8/19/2007	8/25/2007		4	8/1/2014	8/8/2014
	5	5/1/2012	5/7/2012		5	9/15/2017	9/20/2017
	6	8/20/2016	8/26/2016		6	8/29/2013	9/7/2013
	7	8/8/2014	8/14/2014		7	8/22/2015	8/28/2015
	8	8/10/2010	8/16/2010		8	9/27/2018	10/3/2018
	9	7/26/2010	8/4/2010		9	9/6/2018	9/12/2018
	10	8/29/2016	9/4/2016		10	7/16/2006	7/23/2006
Tohoku	1	9/16/2011	9/25/2011	Shikoku	1	6/28/2018	7/10/2018
	2	12/24/2006	12/30/2006		2	8/7/2014	8/13/2014
	3	12/24/2006	12/30/2006		3	9/27/2018	10/3/2018
	4	6/21/2011	6/29/2011		4	9/6/2018	9/12/2018
	5	4/30/2012	5/9/2012		5	7/12/2007	7/18/2007
	6	9/14/2013	9/20/2013		6	9/14/2011	9/23/2011
	7	8/14/2016	8/25/2016		7	8/30/2011	9/5/2011

	8	8/25/2018	9/3/2018		8	10/10/2014	10/16/2014
	9	7/20/2017	7/26/2017		9	8/31/2013	9/7/2013
	10	8/22/2017	8/28/2017		10	7/12/2015	7/19/2015
Kanto	1	10/20/2017	10/26/2017	Kyushu	1	7/2/2018	7/11/2018
	2	9/18/2011	9/24/2011		2	6/17/2016	6/26/2016
	3	9/13/2013	9/19/2013		3	6/30/2006	7/9/2006
	4	9/28/2018	10/4/2018		4	7/18/2006	7/26/2006
	5	9/3/2007	9/9/2007		5	7/10/2012	7/17/2012
	6	9/6/2015	9/12/2015		6	7/24/2009	7/29/2009
	7	6/16/2012	6/22/2012		7	6/26/2009	7/3/2009
	8	8/31/2011	9/8/2011		8	6/14/2011	6/23/2011
	9	7/26/2011	8/2/2011		9	7/3/2017	7/12/2017
	10	8/19/2016	8/25/2016		10	7/8/2010	7/17/2010
Hokuriku_1	1	6/30/2017	7/8/2017	Okinawa	1	9/13/2012	9/19/2012
	2	7/16/2006	7/22/2006		2	8/24/2012	8/30/2012
	3	7/15/2017	7/21/2017		3	6/13/2018	6/19/2018
	4	10/20/2017	10/26/2017		4	7/17/2015	7/23/2015
	5	7/3/2018	7/11/2018		5	10/8/2014	10/14/2014
	6	9/7/2015	9/13/2015		6	7/6/2014	7/12/2014
	7	9/19/2011	9/25/2011		7	4/29/2012	5/5/2012
	8	7/23/2006	7/29/2006		8	8/2/2011	8/9/2011
	9	7/9/2006	7/15/2006		9	10/1/2011	10/7/2011
	10	10/27/2017	11/2/2017		10	5/26/2010	6/1/2010
Hokuriku_2	1	7/15/2006	7/22/2006				
	2	7/3/2018	7/9/2018				
	3	7/3/2011	7/9/2011				
	4	8/7/2014	8/13/2014				
	5	7/3/2017	7/9/2017				
	6	10/20/2017	10/26/2017				
	7	9/2/2018	9/8/2018				
	8	6/17/2013	6/23/2013				
	9	7/27/2013	8/2/2013				
	10	6/26/2018	7/2/2018				

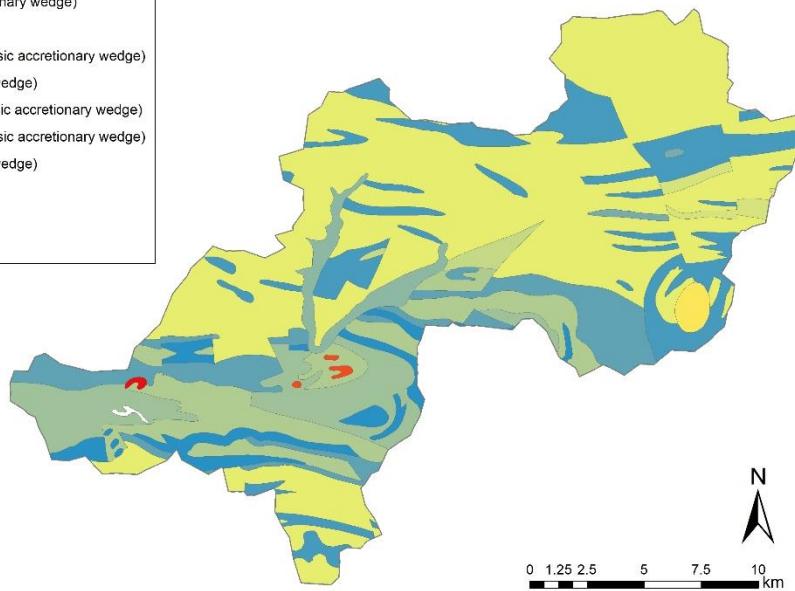
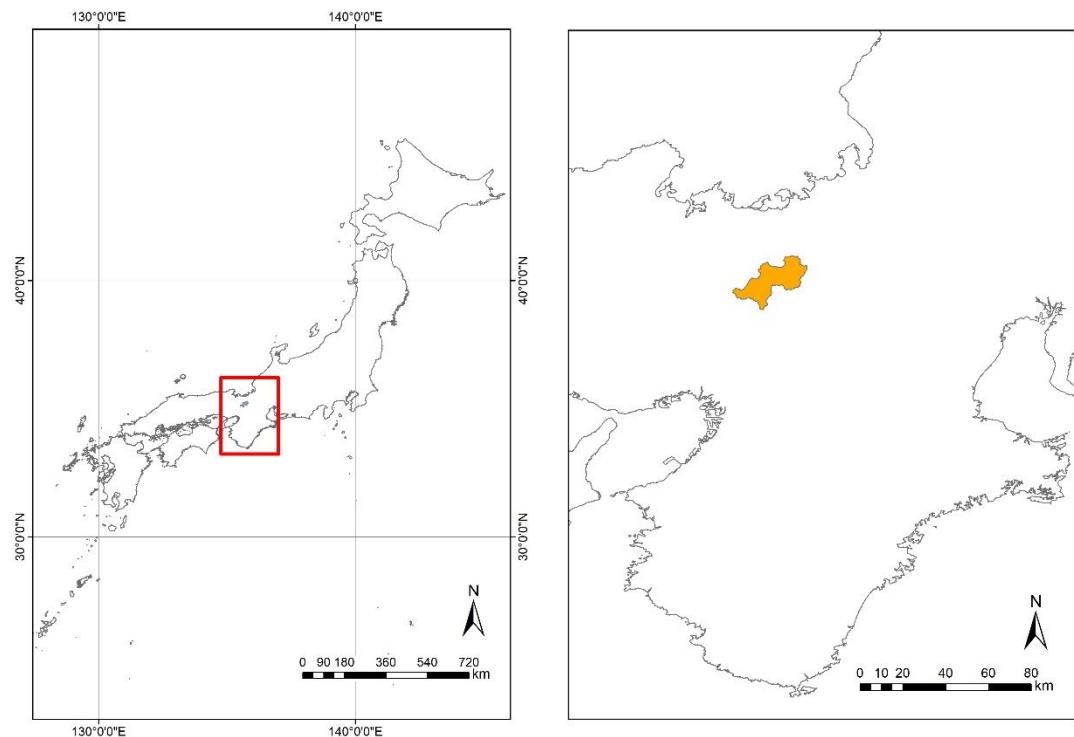


Figure 2: Geology in Hiyoshi dam reservoir basin

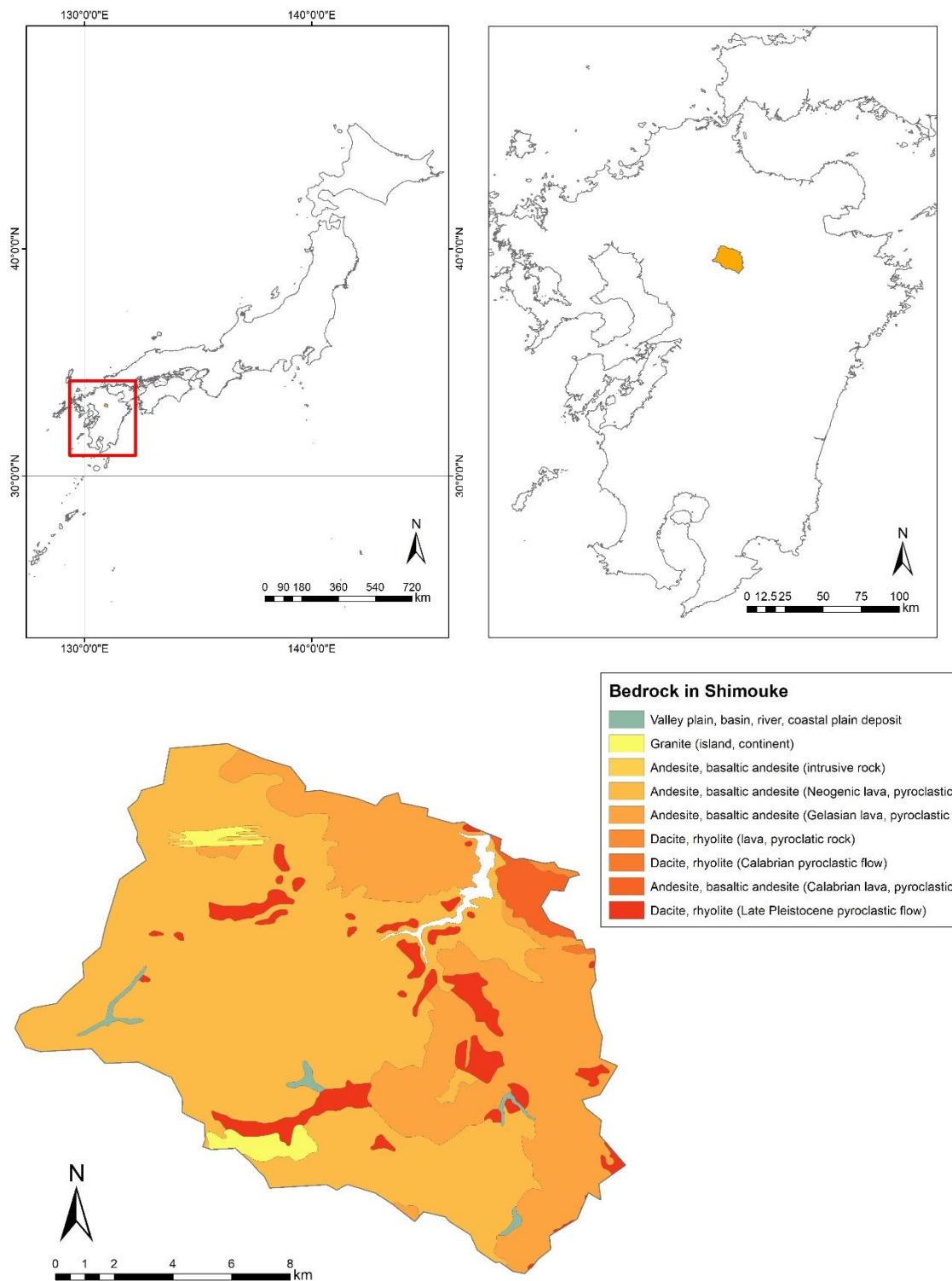


Figure 3: Geology in Shimouke dam reservoir basin

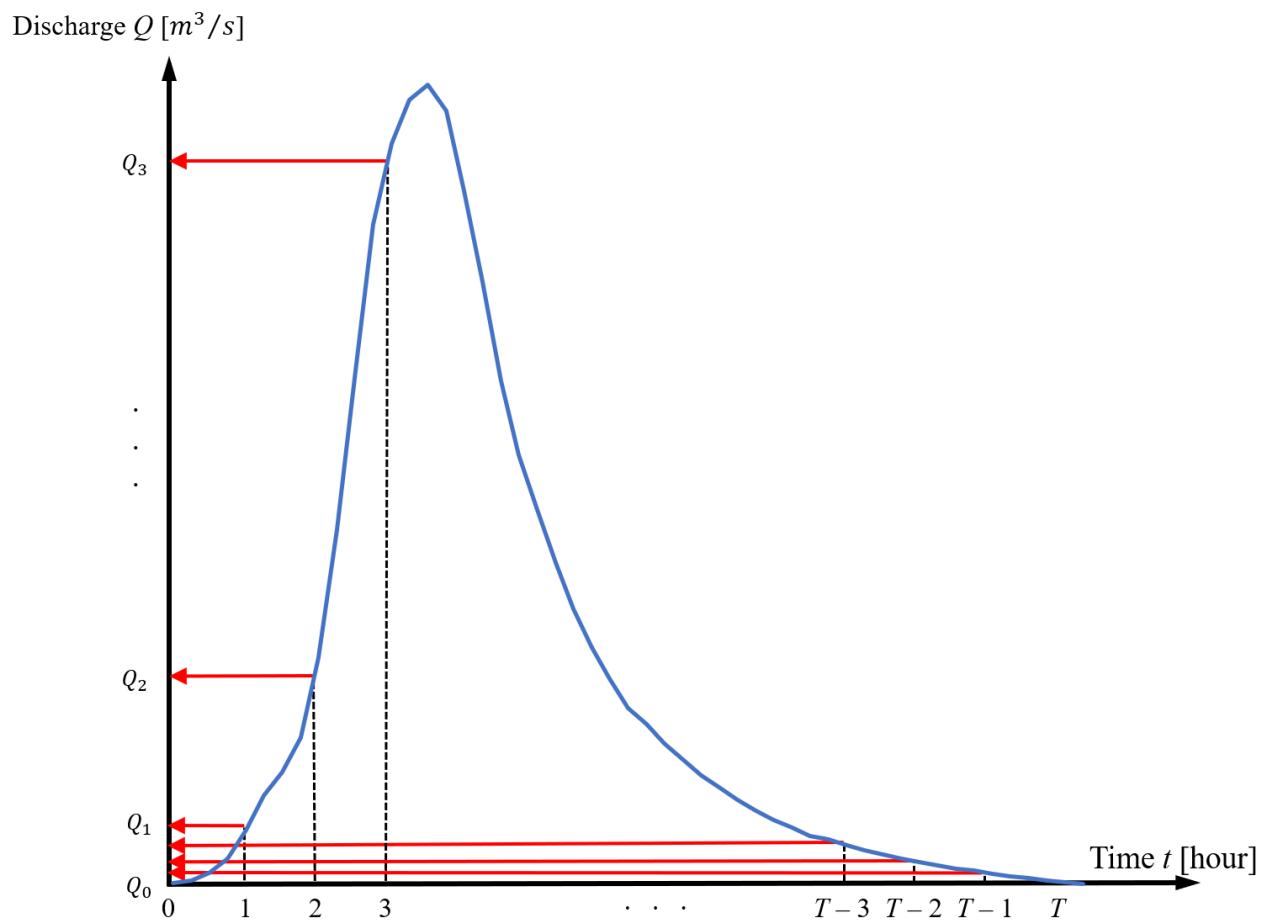


Figure 4: Discretization of hydrographs for clustering

Table 4: Representative parameter sets

ID	n_{slope} [m ^{-1/3} s]	γ_a [-]	γ_m [-]	k_a [m/s]	β [-]	$Q_{peak-sum}$ [m ³]
1	0.40	0.35	0.00	0.20	4	5542.22
2	0.30	0.40	0.05	0.20	6	5483.43
3	0.40	0.35	0.10	0.20	7	5450.76
4	0.40	0.30	0.15	0.15	6	5285.43
5	0.40	0.50	0.00	0.20	4	5282.19
6	0.40	0.45	0.05	0.15	6	5145.45
7	0.40	0.40	0.10	0.15	6	5138.47
8	0.40	0.35	0.00	0.10	4	5118.42
9	0.50	0.35	0.05	0.10	7	5086.71
10	0.40	0.30	0.20	0.15	6	5044.15
11	0.40	0.40	0.15	0.15	6	4906.03
12	0.70	0.35	0.10	0.10	6	4873.71
13	0.40	0.30	0.20	0.10	7	4856.94
14	0.30	0.40	0.00	0.08	4	4846.68
15	0.40	0.40	0.05	0.08	6	4782.40
16	0.40	0.40	0.15	0.10	6	4622.88
17	0.50	0.40	0.10	0.08	6	4581.51
18	0.50	0.35	0.15	0.08	7	4573.69
19	0.40	0.50	0.00	0.08	4	4533.80
20	0.40	0.30	0.20	0.05	9	4425.67
21	0.20	0.40	0.20	0.10	7	4421.77
22	0.50	0.50	0.05	0.08	7	4419.04
23	0.20	0.55	0.15	0.10	4	4260.96
24	0.40	0.45	0.15	0.08	6	4172.17
25	0.40	0.50	0.00	0.05	4	4139.23
26	0.30	0.40	0.20	0.08	6	4097.45
27	0.40	0.50	0.05	0.05	9	4058.76
28	0.60	0.30	0.25	0.05	7	3869.84

29	0.50	0.50	0.10	0.05	5	3807.40
30	0.20	0.50	0.25	0.15	7	3797.74
31	0.40	0.45	0.15	0.05	5	3781.11
32	0.30	0.50	0.00	0.03	4	3702.12
33	0.50	0.50	0.05	0.03	7	3468.25
34	0.20	0.55	0.20	0.08	6	3461.95
35	0.20	0.50	0.25	0.10	7	3376.31
36	0.40	0.50	0.10	0.03	7	3287.05
37	0.50	0.40	0.25	0.05	7	3050.02
38	0.40	0.50	0.15	0.03	8	2964.92
39	0.20	0.55	0.25	0.05	6	2692.61
40	0.40	0.50	0.25	0.03	6	2407.88

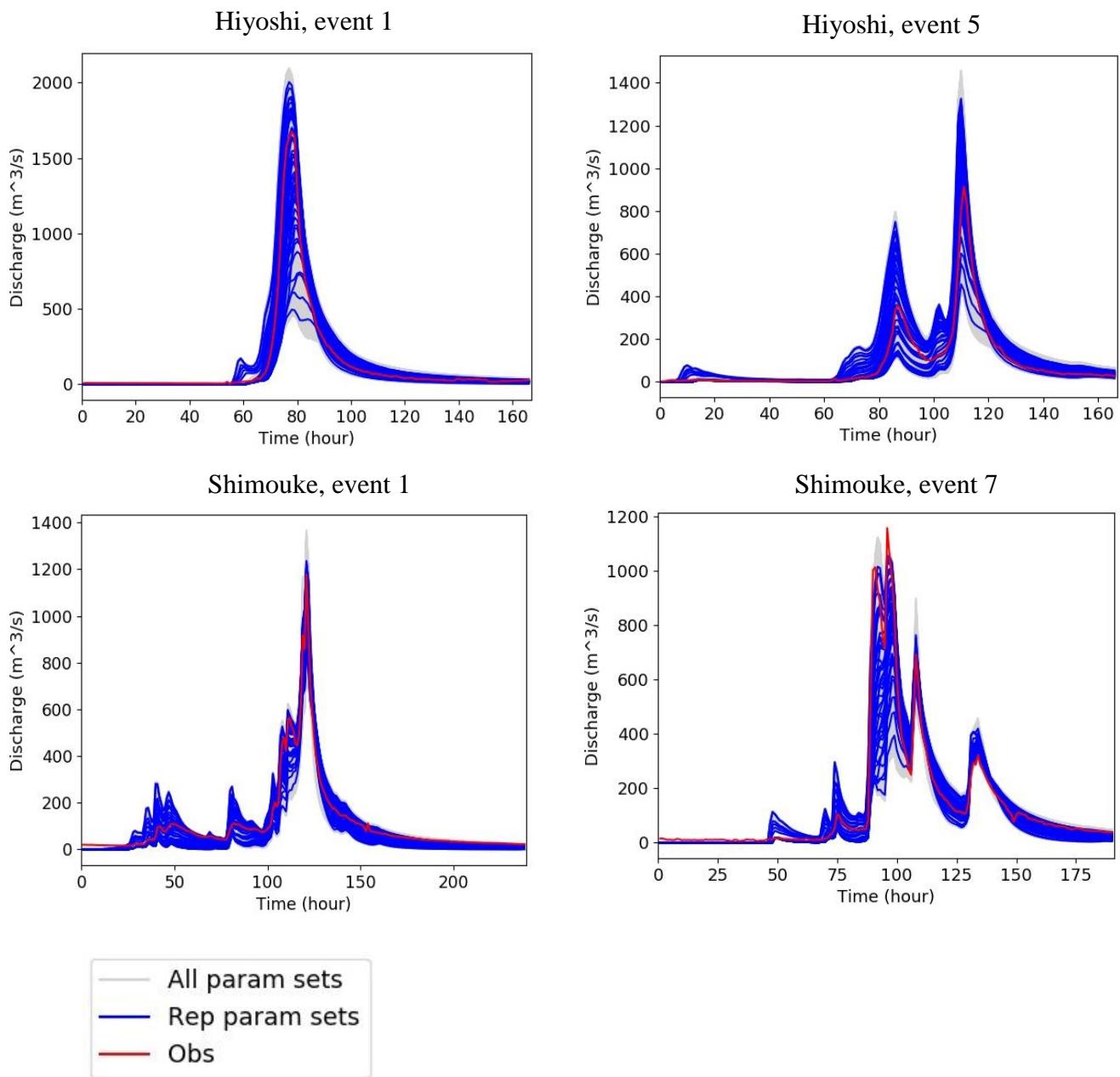


Figure 5: Hydrographs of representative parameter sets

Table 5: Comparison of calculation time of Gurobi and MATLAB Optimizer (median of ten trials)

Num of param set	Time (Gurobi) [s]	Time (MATLAB) [s]
10	4.38	144.06
12	5.74	198.67
14	7.37	430.18
16	14.52	852.54
18	21.24	1630.08
20	25.89	2752.89
22	33.62	-
24	49.30	-
26	62.77	-
28	111.70	-
30	152.29	-
32	166.53	-
34	272.93	-
36	408.10	-
38	501.88	-
40	666.45	-

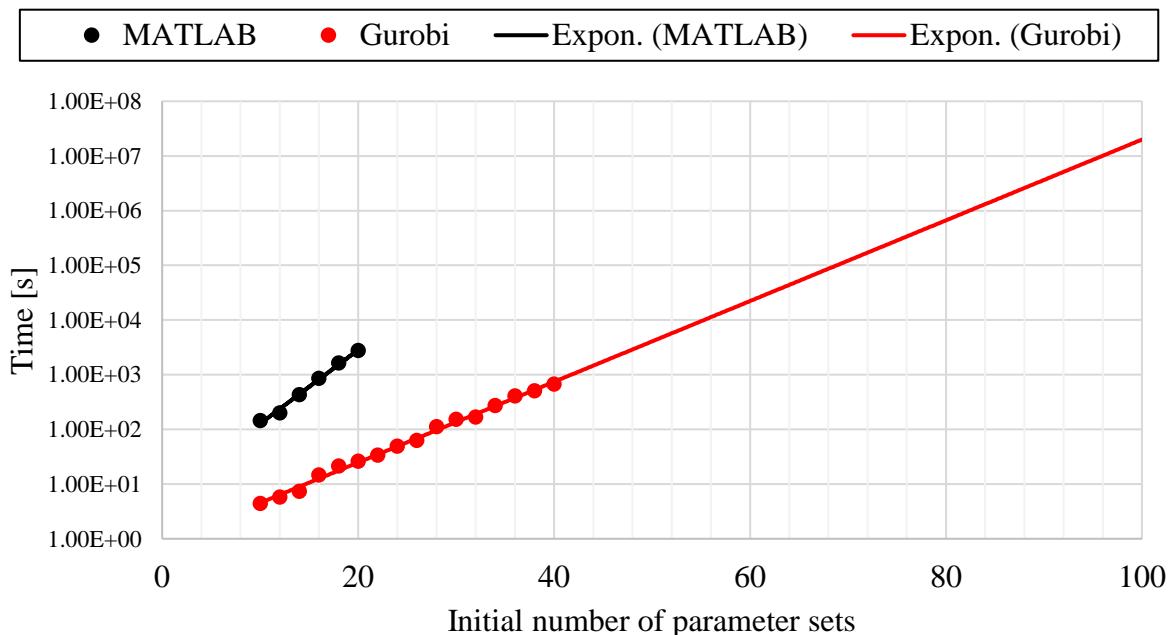


Figure 6: Comparison of calculation time of Gurobi and MATLAB Optimizer
(vertical axis is in logarithmic scale; lines are drawn by regression based on the measured points)

Table 6: Average scores for the RRI model with the default parameter set in each basin

Scores for the basins without enough observation data are set to -999.

Dam	CC_{std}	KGE_{std}	NSE_{std}	EPD	Dam	CC_{std}	KGE_{std}	NSE_{std}	EPD
Agigawa	0.95	0.61	0.50	0.49	Naramata	0.86	0.61	0.60	-0.05
Aha	0.81	0.52	0.64	-0.44	Naruko	0.95	0.57	0.39	0.15
Aimata	0.89	0.50	0.14	0.50	Nibutani	0.85	0.47	0.28	0.55
Amagase	0.47	0.05	0.00	1.00	Nomura	0.97	0.61	0.77	0.00
Arakawa	0.60	0.03	0.19	-0.40	Nukui	0.82	0.08	0.13	0.59
Aseishigawa	0.86	0.20	0.01	0.58	Nunome	0.95	0.58	0.74	-0.01
Benoki	0.72	0.26	0.38	-0.34	Obara	-999	-999	-999	-999
Chubetsu	0.81	0.21	0.00	0.93	Odo	0.96	0.81	0.86	0.02
Egawa	0.94	0.80	0.84	-0.13	Oishi	0.85	0.58	0.31	0.14
Fujiwara	0.74	0.47	0.06	0.65	Okawa	0.79	0.26	0.17	0.71
Fukuji	0.74	0.40	0.51	-0.50	Omachi	0.69	0.19	0.00	0.83
Fungawa	0.72	0.32	0.50	-0.57	Origawa	-999	-999	-999	-999
Futase	0.88	0.52	0.27	0.61	Otaki	0.94	0.87	0.84	-0.09
Gassan	0.77	0.19	0.14	0.49	Pirika	0.80	0.62	0.66	-0.21
Gosho	0.94	0.82	0.75	-0.02	Rumoi	-999	-999	-999	-999
Hachisu	0.91	0.64	0.58	0.11	Ryumon	0.93	0.78	0.74	-0.10
Haiduka	-999	-999	-999	-999	Sagae	0.91	0.70	0.61	-0.13
Haji	0.94	0.57	0.46	0.38	Sagurigawa	-999	-999	-999	-999
Haneji	0.81	0.44	0.54	-0.36	Sameura	0.94	0.82	0.84	-0.13
Hattabara	0.97	0.64	0.70	0.28	Sarutani	0.97	0.86	0.91	-0.11
Hinachi	0.94	0.67	0.65	0.08	Satsunaigawa	0.92	0.64	0.61	0.05
Hitokura	0.96	0.57	0.54	0.36	Shichikashuku	0.83	0.39	0.36	0.55
Hiyoshi	0.94	0.56	0.66	0.18	Shijushida	0.91	0.34	0.08	0.91
Hoheikyo	0.88	0.32	0.22	0.17	Shimadigawa	0.85	0.62	0.37	0.36
Ikari	0.90	0.53	0.23	0.66	Shimokubo	0.86	0.43	0.21	0.62
Ikeda	0.93	0.66	0.66	0.15	Shimouke	0.95	0.79	0.75	-0.07
Isawa	-999	-999	-999	-999	Shingu	0.87	0.51	0.43	0.15
Ishibuchi	-999	-999	-999	-999	Shintoyone	0.93	0.66	0.52	0.23
Ishitegawa	0.93	0.21	0.22	0.69	Shirakawa	0.89	0.50	0.39	0.28
Iwaonai	0.73	0.29	0.02	0.83	Shitsumi	-999	-999	-999	-999
Iwaya	0.90	0.61	0.28	0.30	Shorenji	0.96	0.80	0.88	-0.14
Izarigawa	0.89	0.10	0.00	0.79	Sonohara	0.88	0.50	0.22	0.47
Jozankei	0.89	0.23	0.01	0.71	Sugesawa	0.95	0.50	0.68	0.16
Kamafusa	0.90	0.58	0.50	0.35	Taiho	-999	-999	-999	-999
Kanayama	0.81	0.27	0.07	0.89	Taisetsu	0.84	0.42	0.24	0.20
Kanna	0.56	0.06	0.30	-0.63	Takayama	0.96	0.65	0.60	0.32
Kanogawa	0.97	0.65	0.77	0.04	Takisato	0.90	0.31	0.14	0.79

Kanoko	0.78	0.10	0.01	0.95	Takizawa	-999	-999	-999	-999
Kasegawa	-999	-999	-999	-999	Tamagawa	0.93	0.68	0.70	-0.08
Katsurazawa	0.90	0.58	0.27	0.49	Tase	0.91	0.13	0.00	1.00
Kawaji	0.61	0.25	0.28	-0.70	Tedorigawa	0.93	0.70	0.56	0.12
Kawamata	0.91	0.56	0.33	0.40	Terauchi	0.94	0.81	0.80	-0.04
Koshibu	0.83	0.28	0.00	0.90	Tokachi	0.79	0.25	0.07	0.71
Kurashiki	-999	-999	-999	-999	Tokuyama	-999	-999	-999	-999
Kusaki	0.89	0.63	0.55	0.10	Tomata	0.94	0.63	0.42	0.22
Kuzuryu	0.93	0.68	0.53	0.09	Tomisato	0.93	0.80	0.78	-0.06
Kyuragi	0.90	0.08	0.01	0.36	Tono	-999	-999	-999	-999
Makio	0.93	0.73	0.66	0.09	Tsuruda	0.97	0.73	0.76	0.37
Managawa	0.86	0.19	0.00	0.78	Unaduki	0.87	0.20	0.00	0.80
Maruyama	-999	-999	-999	-999	Urayama	0.91	0.51	0.24	0.62
Matsubara	0.92	0.59	0.17	0.62	Ure	0.87	0.57	0.68	-0.30
Midorikawa	0.94	0.61	0.34	0.29	Wataraseyusuichi	-999	-999	-999	-999
Miharu	0.89	0.63	0.42	0.42	Yabakei	0.94	0.69	0.85	-0.33
Misogawa	0.71	0.22	0.00	1.00	Yagisawa	0.87	0.58	0.35	0.23
Miwa	0.74	0.29	0.00	0.92	Yahagi	0.94	0.49	0.11	0.40
Miyagase	0.91	0.72	0.59	-0.08	Yanase	0.91	0.71	0.59	0.06
Muro	0.96	0.52	0.58	0.21	Yasaka	0.91	0.23	0.10	0.73
Nagai	-999	-999	-999	-999	Yokoyama	0.67	0.06	0.00	1.00
Nagashima	-999	-999	-999	-999	Yuda	0.97	0.71	0.73	0.07
Nagayasuguchi	-999	-999	-999	-999	Yunishigawa	-999	-999	-999	-999
Nakasujigawa	0.94	0.66	0.77	-0.09					

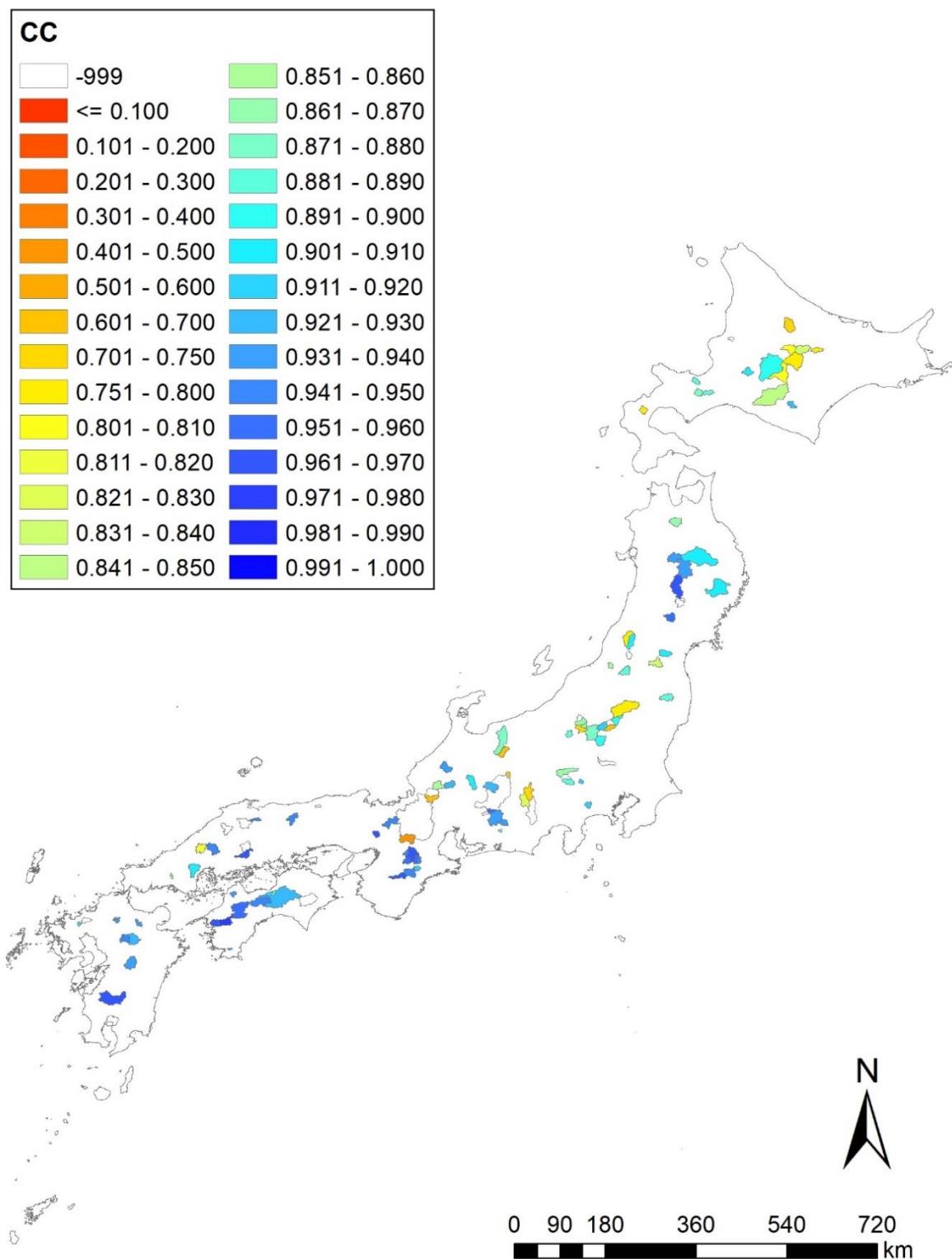


Figure 7: Spatial distribution of the average CC_{std} for the default parameter set

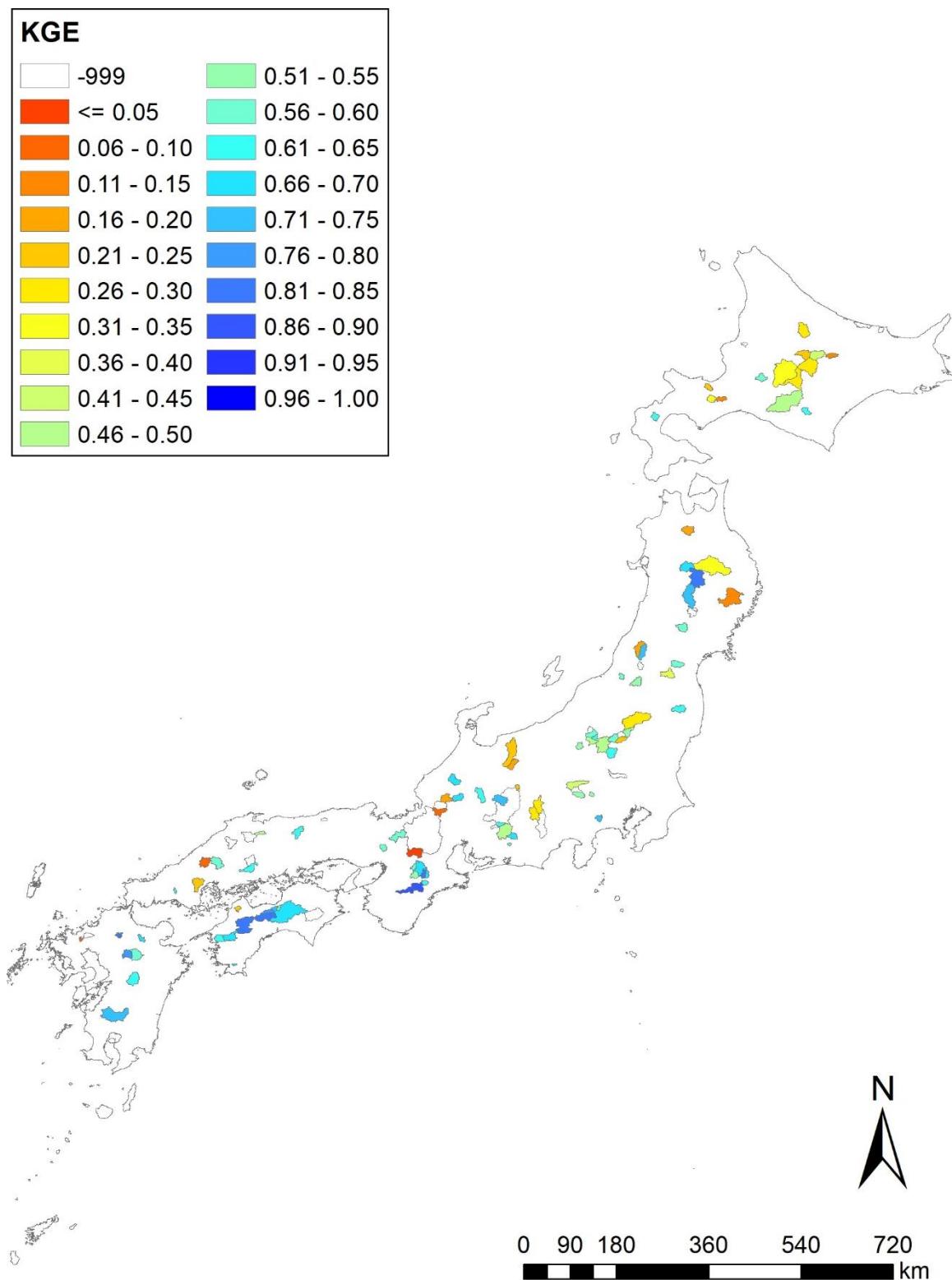


Figure 8: Spatial distribution of the average KGE_{std} for the default parameter set

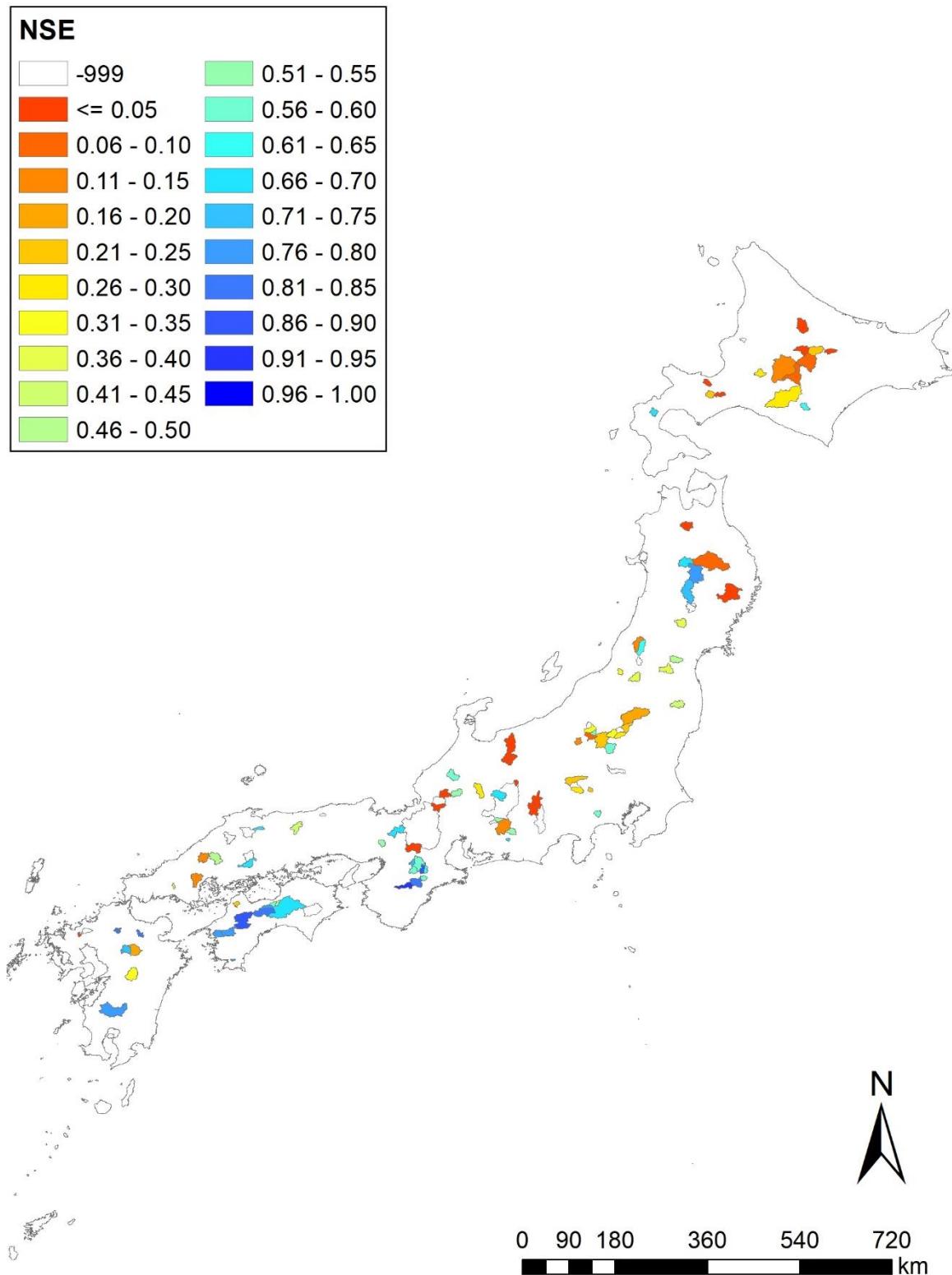


Figure 9: Spatial distribution of the average NSE_{std} for the default parameter set

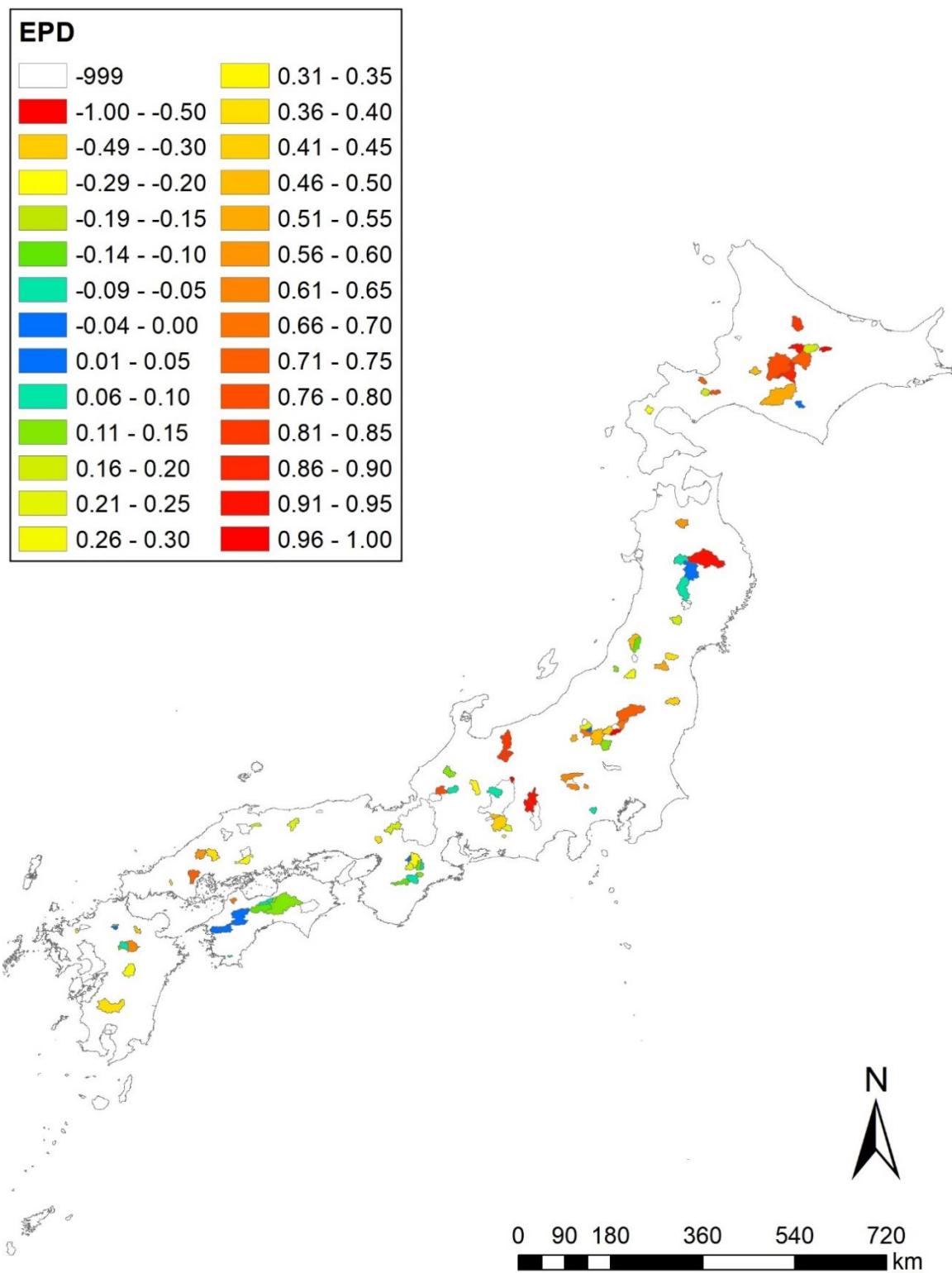


Figure 10: Spatial distribution of the average *EPD* for the default parameter set

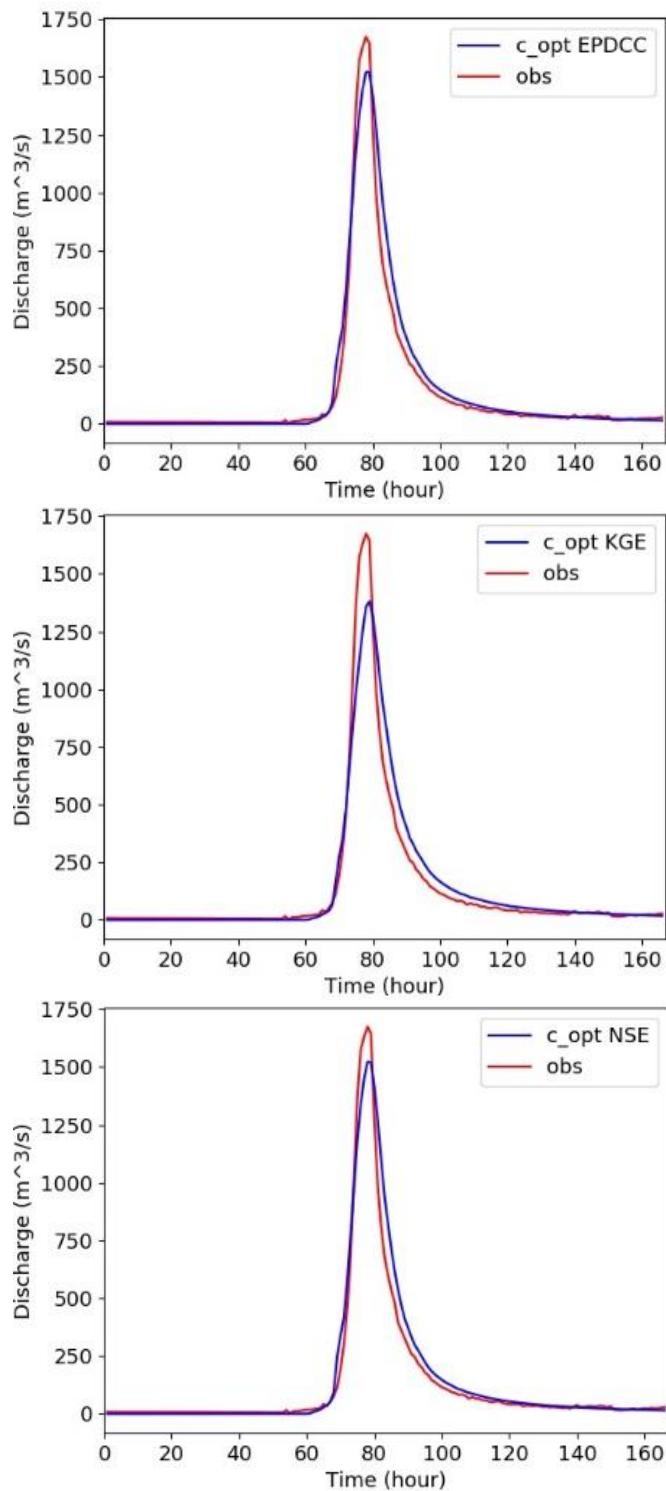


Figure 11: Resulting simulations with different evaluation indices in Hiyoshi (event 1)

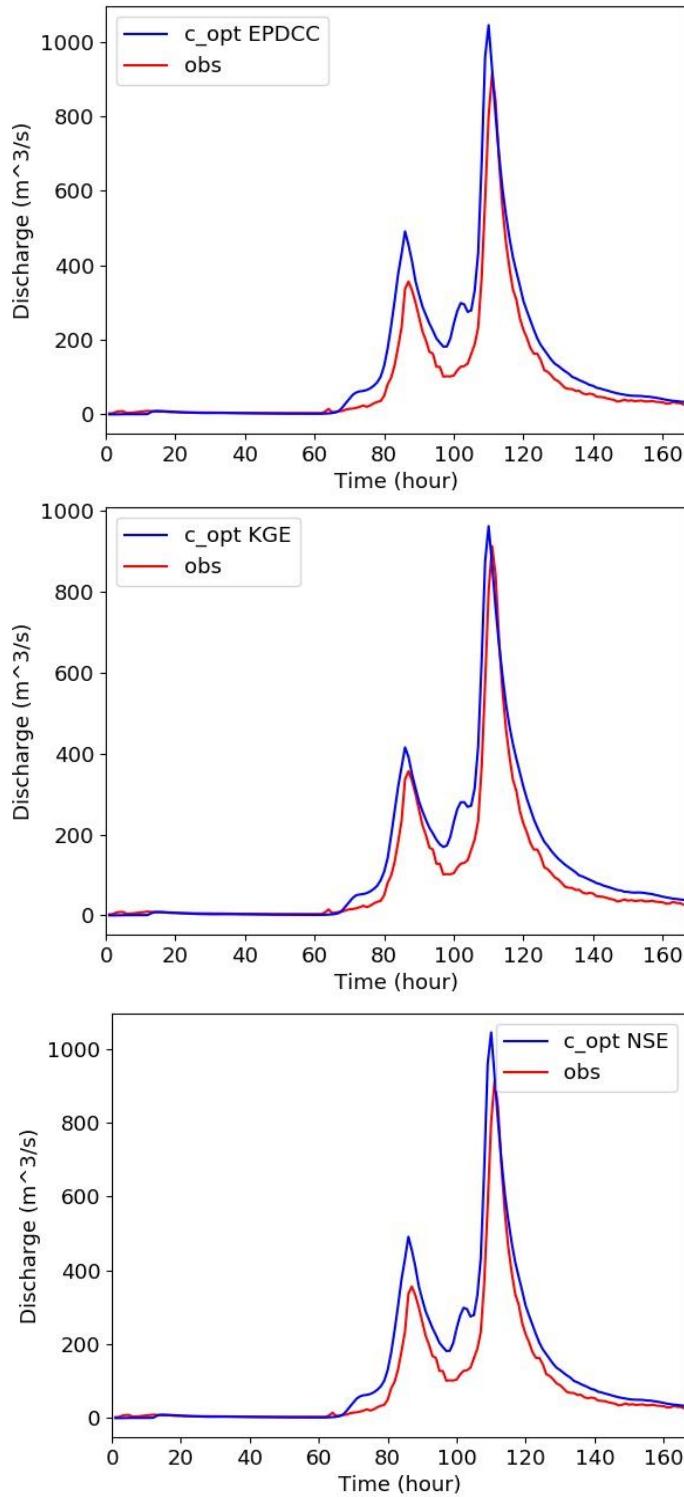


Figure 12: Resulting simulations with different evaluation indices in Hiyoshi (event 5)

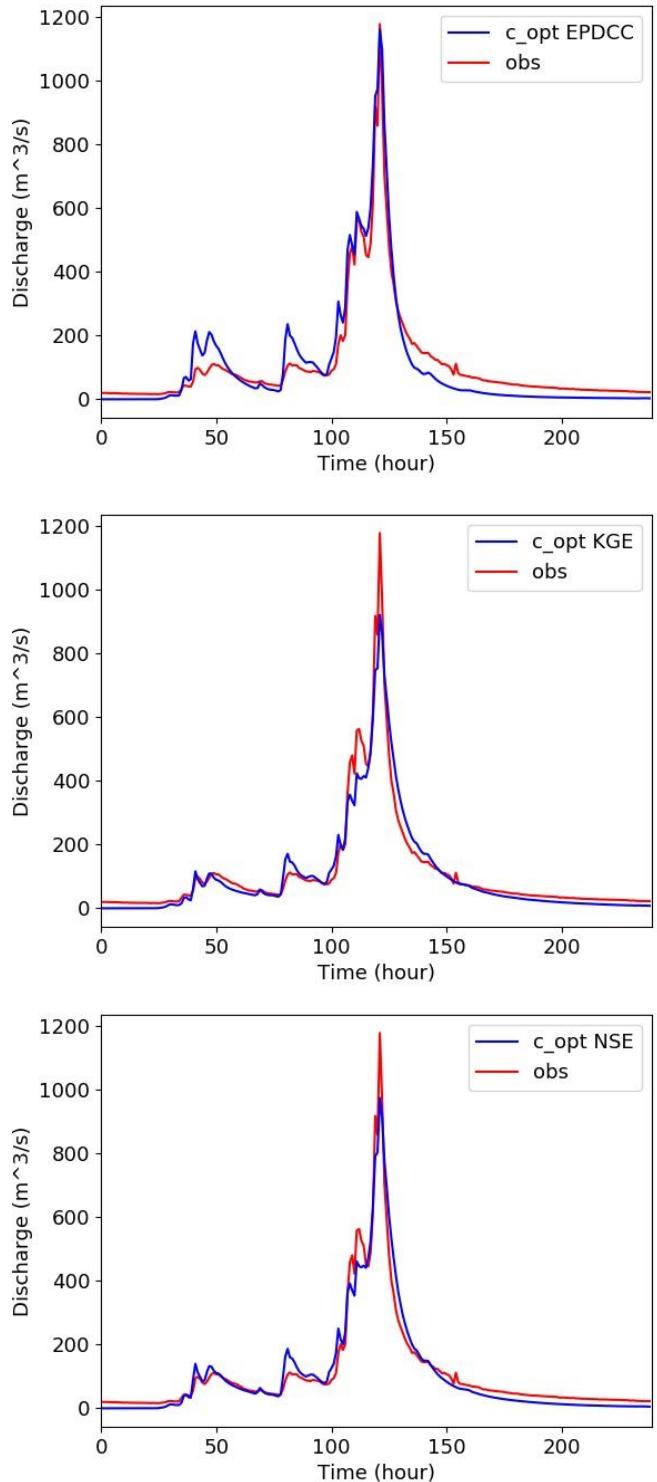


Figure 13: Resulting simulations with different evaluation indices in Shimouke (event 1)

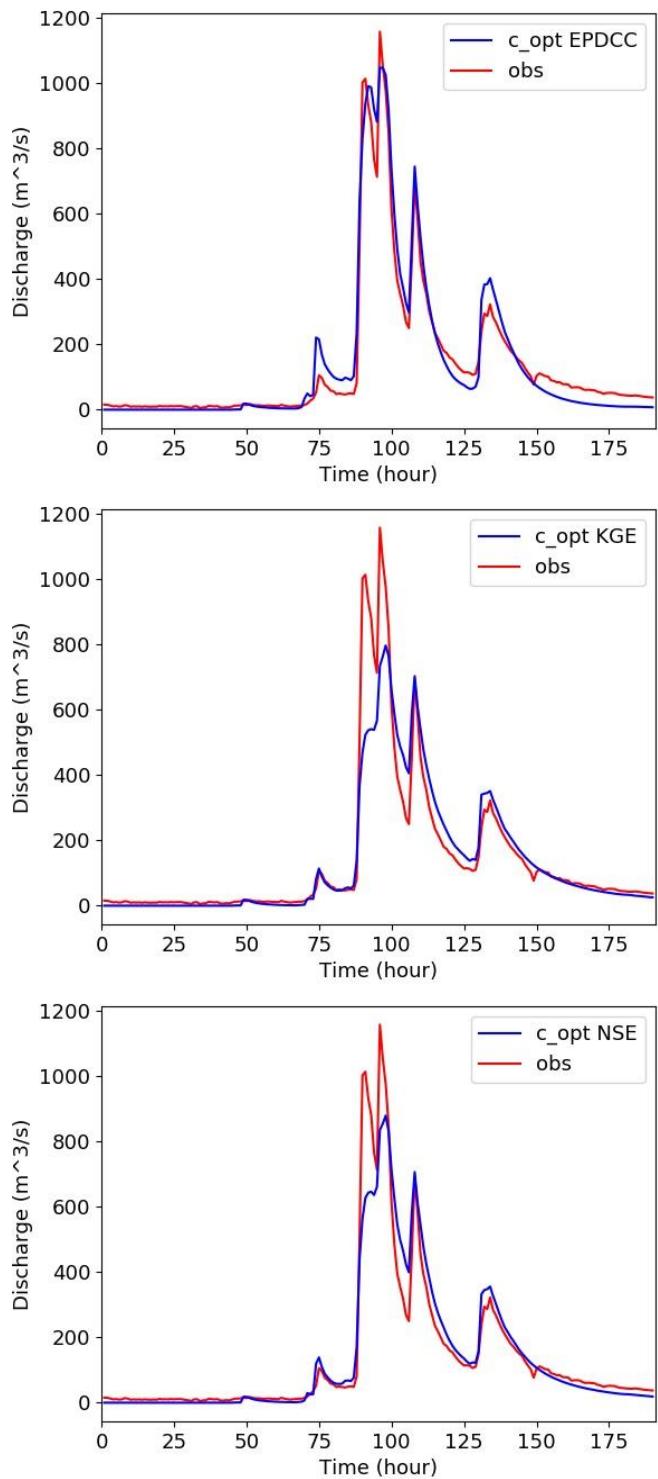


Figure 14: Resulting simulations with different evaluation indices in Shimouke (event 7)

Table 7: Best parameter set among the 40 representatives (opt parameter sets) for each basin

Optimal parameter sets for the basins without enough observation data are set to -1.

Dam	Opt Param Set	Dam	Opt Param Set
Agigawa	32	Naramata	2
Aha	1	Naruko	15
Aimata	27	Nibutani	17
Amagase	38	Nomura	14
Arakawa	1	Nukui	17
Aseishigawa	17	Nunome	19
Benoki	5	Obara	-1
Chubetsu	20	Odo	15
Egawa	3	Oishi	6
Fujiwara	32	Okawa	23
Fukuchi	1	Omachi	36
Fungawa	1	Origawa	-1
Futase	33	Otaki	11
Gassan	25	Pirika	8
Gosho	6	Rumoi	-1
Hachisu	11	Ryumon	9
Haiduka	-1	Sagae	5
Haji	27	Sagurigawa	-1
Haneji	3	Sameura	3
Hattabara	22	Sarutani	3
Hinachi	22	Satsunaigawa	19
Hitokura	16	Shichikashuku	7
Hiyoshi	22	Shijushida	33
Hoheikyo	9	Shimadigawa	27
Ikari	36	Shimokubo	17
Ikeda	32	Shimouke	3
Isawa	-1	Shingu	19
Ishibuchi	-1	Shintoyone	29
Ishitegawa	33	Shirakawa	22
Iwaonai	29	Shitsumi	-1
Iwaya	12	Shorenji	3
Izarigawa	21	Sonohara	32
Jozankei	18	Sugesawa	4
Kamafusa	12	Taiho	-1
Kanayama	33	Taisetsu	2
Kanna	1	Takayama	32
Kanogawa	9	Takisato	33

Kanoko	36	Takizawa	-1
Kasegawa	-1	Tamagawa	2
Katsurazawa	27	Tase	36
Kawaji	40	Tedorigawa	22
Kawamata	16	Terauchi	15
Koshibu	38	Tokachi	17
Kurashiki	-1	Tokuyama	-1
Kusaki	15	Tomata	22
Kuzuryu	15	Tomisato	10
Kyuragi	27	Tono	-1
Makio	3	Tsuruda	32
Managawa	36	Unaduki	36
Maruyama	-1	Urayama	29
Matsubara	40	Ure	1
Midorikawa	32	Wataraseyusuichi	-1
Miharu	32	Yabakei	1
Misogawa	31	Yagisawa	14
Miwa	36	Yahagi	27
Miyagase	7	Yanase	10
Muro	19	Yasaka	36
Nagai	-1	Yokoyama	40
Nagashima	-1	Yuda	19
Nagayasuguchi	-1	Yunishigawa	-1
Nakasujigawa	7		

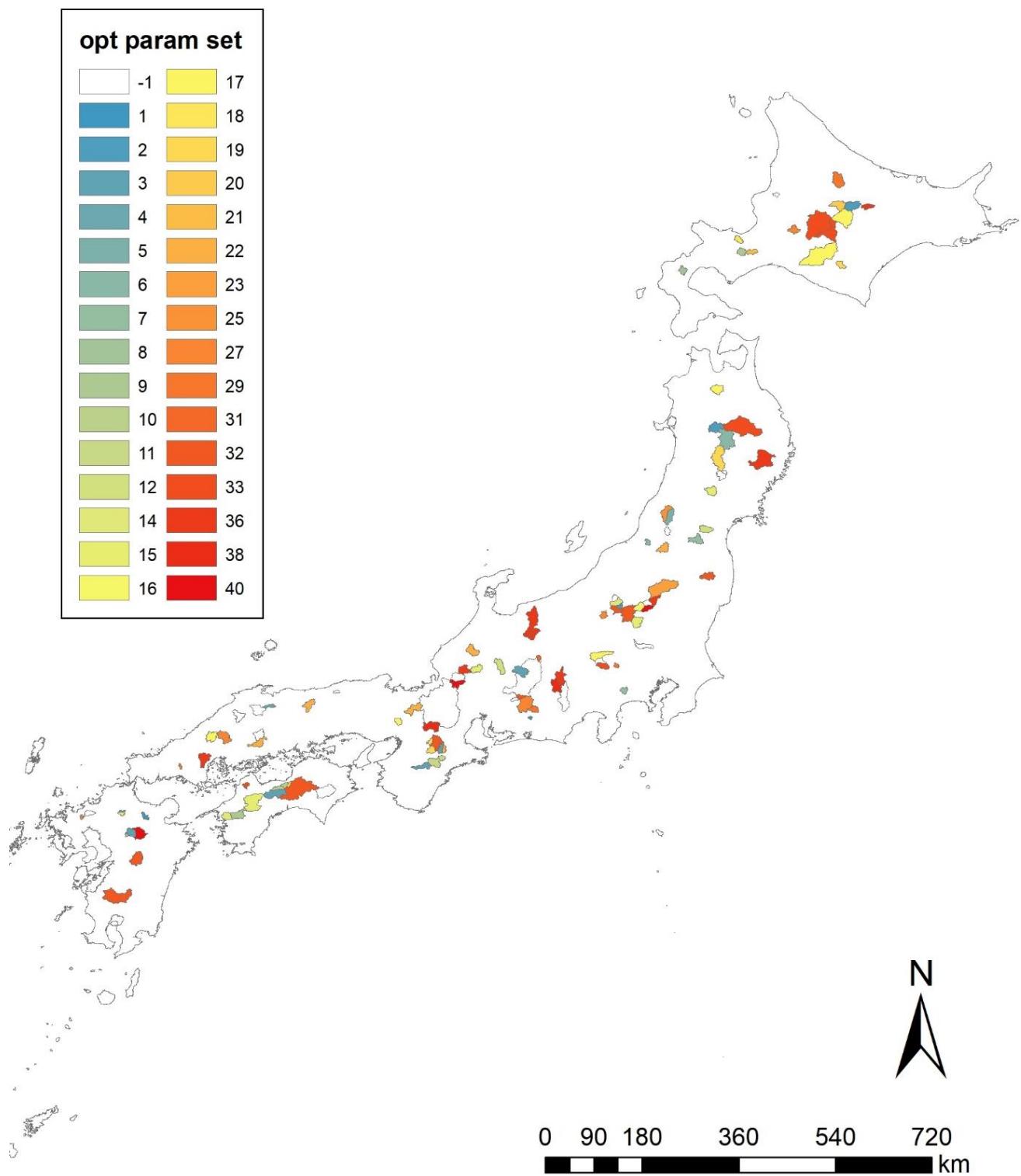


Figure 15: Spatial distribution of best parameter sets among the 40 representatives (opt parameter sets)

Table 8: Evaluation index scores for the opt parameter sets

Scores for the basins without enough observation data are set to -999.

Dam	CC_{std}	KGE_{std}	NSE_{std}	EPD	Dam	CC_{std}	KGE_{std}	NSE_{std}	EPD
Agigawa	0.94	0.72	0.81	0.17	Naramata	0.89	0.60	0.54	-0.04
Aha	0.86	0.63	0.73	-0.31	Naruko	0.94	0.64	0.63	-0.04
Aimata	0.93	0.66	0.64	0.13	Nibutani	0.92	0.57	0.62	0.01
Amagase	0.57	0.08	0.00	0.58	Nomura	0.97	0.62	0.80	-0.05
Arakawa	0.68	0.07	0.22	-0.19	Nukui	0.81	0.21	0.38	0.09
Aseishigawa	0.84	0.59	0.44	-0.11	Nunome	0.94	0.59	0.77	-0.08
Benoki	0.75	0.27	0.35	-0.26	Obara	-999	-999	-999	-999
Chubetsu	0.77	0.30	0.15	-0.16	Odo	0.98	0.81	0.94	-0.01
Egawa	0.95	0.64	0.85	-0.01	Oishi	0.87	0.54	0.33	0.16
Fujiwara	0.75	0.58	0.36	0.02	Okawa	0.84	0.49	0.41	0.09
Fukuji	0.76	0.50	0.54	-0.41	Omachi	0.83	0.25	0.13	0.07
Fungawa	0.79	0.50	0.61	-0.44	Origawa	-999	-999	-999	-999
Futase	0.97	0.73	0.86	0.05	Otaki	0.97	0.79	0.89	-0.04
Gassan	0.76	0.19	0.15	0.36	Pirika	0.80	0.63	0.65	-0.17
Gosho	0.97	0.75	0.84	-0.01	Rumoi	-999	-999	-999	-999
Hachisu	0.91	0.65	0.68	0.05	Ryumon	0.95	0.74	0.75	-0.04
Haiduka	-999	-999	-999	-999	Sagae	0.91	0.64	0.42	0.09
Haji	0.96	0.80	0.84	-0.04	Sagurigawa	-999	-999	-999	-999
Haneji	0.80	0.61	0.55	-0.30	Sameura	0.97	0.79	0.89	-0.05
Hattabara	0.98	0.72	0.84	0.02	Sarutani	0.99	0.83	0.94	0.00
Hinachi	0.95	0.71	0.76	-0.03	Satsunaigawa	0.94	0.67	0.69	-0.04
Hitokura	0.97	0.73	0.87	-0.14	Shichikashuku	0.93	0.52	0.47	0.18
Hiyoshi	0.96	0.73	0.86	0.03	Shijushida	0.89	0.64	0.57	0.16
Hoheikyo	0.86	0.45	0.27	-0.17	Shimadigawa	0.92	0.73	0.69	0.08
Ikari	0.95	0.79	0.83	-0.07	Shimokubo	0.94	0.71	0.81	0.02
Ikeda	0.97	0.74	0.87	-0.11	Shimouke	0.96	0.73	0.76	0.11
Isawa	-999	-999	-999	-999	Shingu	0.87	0.51	0.48	0.07
Ishibuchi	-999	-999	-999	-999	Shintoyone	0.96	0.80	0.85	-0.04
Ishitegawa	0.94	0.45	0.52	0.18	Shirakawa	0.87	0.57	0.56	-0.09
Iwaonai	0.80	0.55	0.43	0.29	Shitsumi	-999	-999	-999	-999
Iwaya	0.95	0.60	0.66	0.05	Shorenji	0.98	0.74	0.89	0.02
Izarigawa	0.81	0.34	0.29	0.08	Sonohara	0.93	0.57	0.76	-0.02
Jozankei	0.85	0.53	0.49	-0.19	Sugesawa	0.94	0.70	0.71	-0.09
Kamafusa	0.92	0.71	0.72	0.02	Taiho	-999	-999	-999	-999
Kanayama	0.88	0.56	0.36	0.18	Taisetsu	0.87	0.51	0.31	0.26
Kanna	0.63	0.13	0.38	-0.51	Takayama	0.95	0.75	0.84	-0.02
Kanogawa	0.98	0.73	0.82	0.05	Takisato	0.89	0.55	0.39	0.14

Kanoko	0.86	0.63	0.44	0.19	Takizawa	-999	-999	-999	-999
Kasegawa	-999	-999	-999	-999	Tamagawa	0.95	0.67	0.69	0.01
Katsurazawa	0.93	0.76	0.76	0.05	Tase	0.90	0.55	0.48	-0.14
Kawaji	0.84	0.23	0.32	-0.73	Tedorigawa	0.93	0.75	0.76	-0.07
Kawamata	0.96	0.76	0.79	0.06	Terauchi	0.96	0.80	0.86	-0.07
Koshibu	0.92	0.64	0.69	-0.01	Tokachi	0.85	0.53	0.47	0.10
Kurashiki	-999	-999	-999	-999	Tokuyama	-999	-999	-999	-999
Kusaki	0.93	0.72	0.68	0.03	Tomata	0.96	0.75	0.71	0.06
Kuzuryu	0.92	0.72	0.64	-0.03	Tomisato	0.95	0.71	0.87	-0.03
Kyuragi	0.89	0.18	0.12	0.13	Tono	-999	-999	-999	-999
Makio	0.94	0.72	0.76	0.10	Tsuruda	0.95	0.82	0.85	0.11
Managawa	0.91	0.59	0.61	-0.04	Unaduki	0.78	0.42	0.15	0.07
Maruyama	-999	-999	-999	-999	Urayama	0.97	0.77	0.83	0.09
Matsubara	0.89	0.67	0.66	0.17	Ure	0.88	0.63	0.68	-0.17
Midorikawa	0.95	0.74	0.80	-0.04	Wataraseyusuichi	-999	-999	-999	-999
Miharu	0.88	0.63	0.60	-0.05	Yabakei	0.94	0.78	0.86	-0.15
Misogawa	0.91	0.58	0.37	0.23	Yagisawa	0.86	0.60	0.39	0.12
Miwa	0.94	0.65	0.62	0.18	Yahagi	0.95	0.68	0.62	0.08
Miyagase	0.94	0.71	0.69	-0.03	Yanase	0.92	0.67	0.67	0.05
Muro	0.95	0.55	0.66	0.12	Yasaka	0.93	0.68	0.67	0.03
Nagai	-999	-999	-999	-999	Yokoyama	0.83	0.53	0.34	0.09
Nagashima	-999	-999	-999	-999	Yuda	0.97	0.70	0.76	-0.02
Nagayasuguchi	-999	-999	-999	-999	Yunishigawa	-999	-999	-999	-999
Nakasujigawa	0.96	0.73	0.81	0.02					

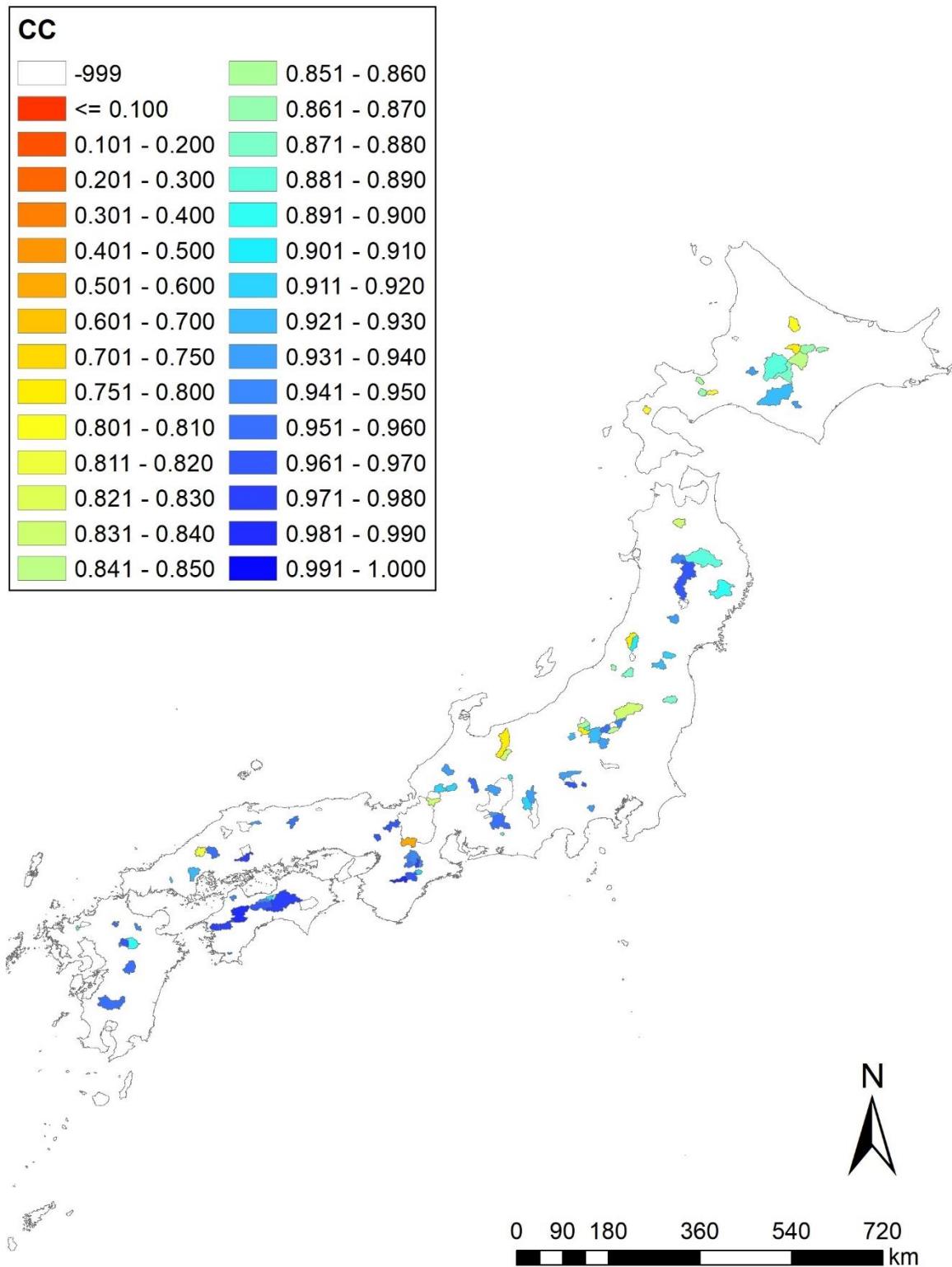


Figure 16: Spatial distribution of CC_{std} for the opt parameter sets

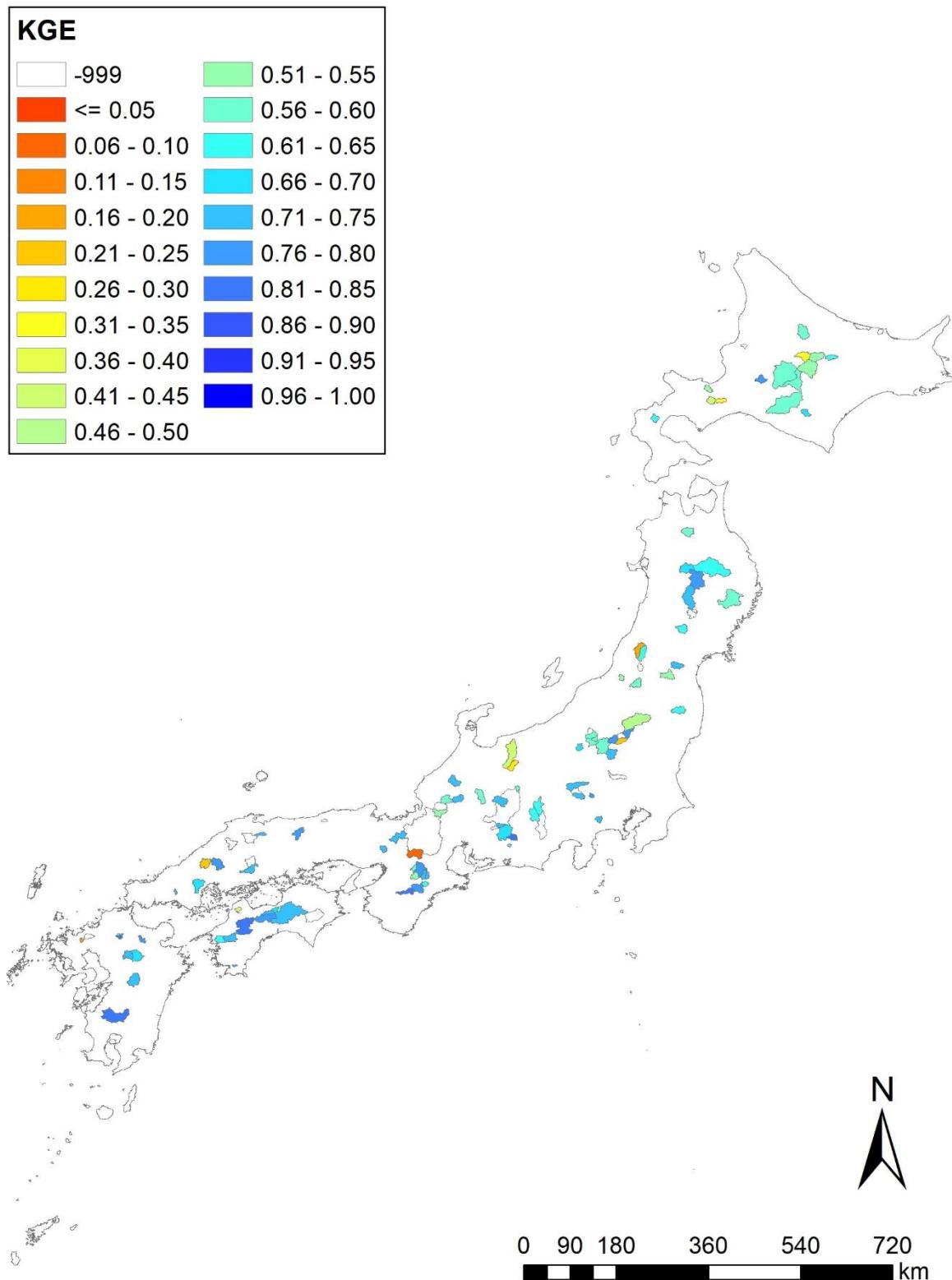


Figure 17: Spatial distribution of KGE_{std} for the opt parameter sets

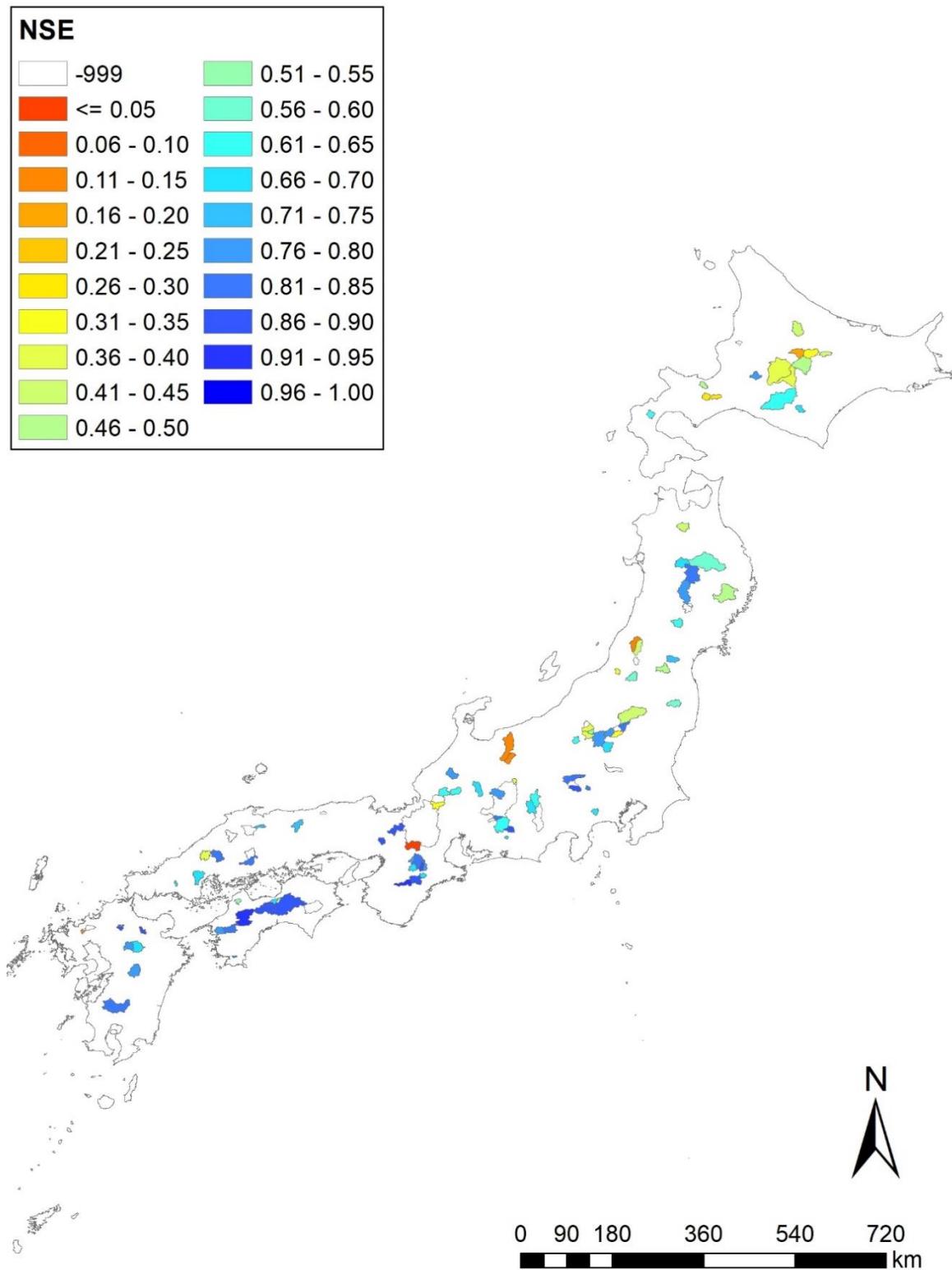


Figure 18: Spatial distribution of NSE_{std} for the opt parameter sets

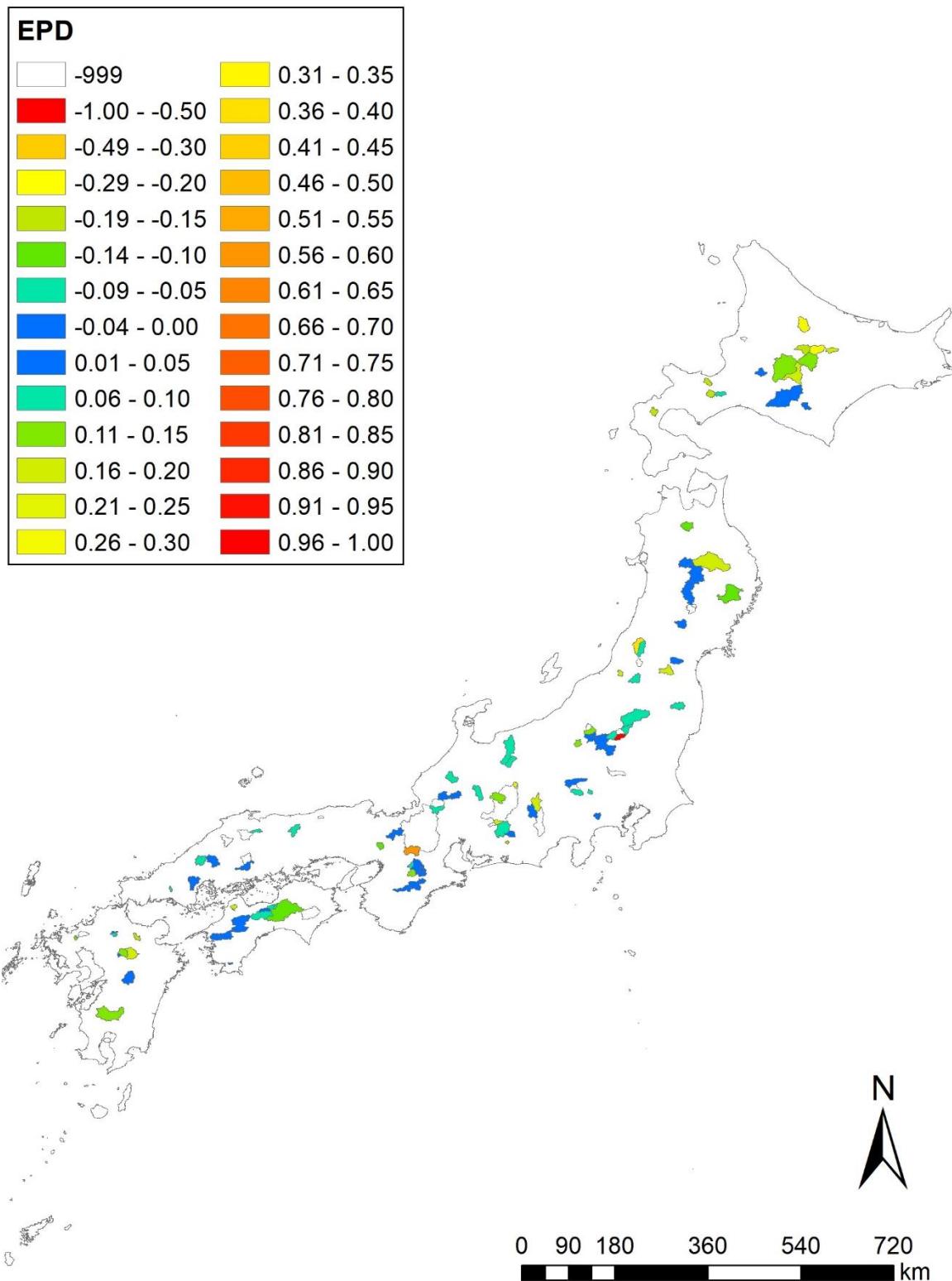


Figure 19: Spatial distribution of *EPD* for the opt parameter sets

Table 9: Optimal parameter sets obtained by the combinatorial optimization (c_opt parameter sets)

ID	n_{slope} [m $^{-1/3}s$]	γ_a [-]	γ_m [-]	k_a [m/s]	β [-]
2	0.3	0.40	0.05	0.200	6
17	0.5	0.40	0.10	0.075	6
22	0.5	0.50	0.05	0.075	7
33	0.5	0.50	0.05	0.030	7
38	0.4	0.50	0.15	0.030	8

Table 10: The c_opt parameter set for each basin

The c_opt parameter set IDs for the basins without enough observation data are set to -1.

Dam	c_opt ID	Dam	c_opt ID	Dam	c_opt ID
Agigawa	33	Kawamata	17	Shichikashuku	33
Aha	2	Koshibu	38	Shijushida	33
Aimata	22	Kurashiki	-1	Shimadigawa	22
Amagase	38	Kusaki	22	Shimokubo	17
Arakawa	2	Kuzuryu	22	Shimouke	2
Aseishigawa	17	Kyuragi	22	Shingu	22
Benoki	2	Makio	22	Shintoyone	33
Chubetsu	38	Managawa	33	Shirakawa	22
Egawa	2	Maruyama	-1	Shitsumi	-1
Fujiwara	33	Matsubara	38	Shorenji	2
Fukuchi	2	Midorikawa	33	Sonohara	17
Fungawa	2	Miharu	2	Sugesawa	17
Futase	33	Misogawa	38	Taiho	-1
Gassan	22	Miwa	33	Taisetsu	2
Gosho	2	Miyagase	17	Takayama	33
Hachisu	17	Muro	22	Takisato	33
Haiduka	-1	Nagai	-1	Takizawa	-1
Haji	22	Nagashima	-1	Tamagawa	2
Haneji	2	Nagayasuguchi	-1	Tase	33
Hattabara	22	Nakasujigawa	22	Tedorigawa	22
Hinachi	22	Naramata	2	Terauchi	17
Hitokura	17	Naruko	22	Tokachi	17
Hiyoshi	22	Nibutani	17	Tokuyama	-1
Hoheikyo	2	Nomura	22	Tomata	22
Ikari	33	Nukui	17	Tomisato	17
Ikeda	17	Nunome	22	Tono	-1

Isawa	-1	Obara	-1	Tsuruda	33
Ishibuchi	-1	Odo	17	Unaduki	38
Ishitegawa	33	Oishi	22	Urayama	33
Iwaonai	33	Okawa	17	Ure	2
Iwaya	22	Omachi	33	Wataraseyusuichi	-1
Izariigawa	38	Origawa	-1	Yabakei	2
Jozankei	17	Otaki	2	Yagisawa	22
Kamafusa	22	Pirika	2	Yahagi	33
Kanayama	33	Rumoi	-1	Yanase	17
Kanna	2	Ryumon	2	Yasaka	33
Kanogawa	22	Sagae	2	Yokoyama	38
Kanoko	38	Sagurigawa	-1	Yuda	22
Kasegawa	-1	Sameura	2	Yunishigawa	-1
Katsurazawa	33	Sarutani	2		
Kawaji	38	Satsunaigawa	22		

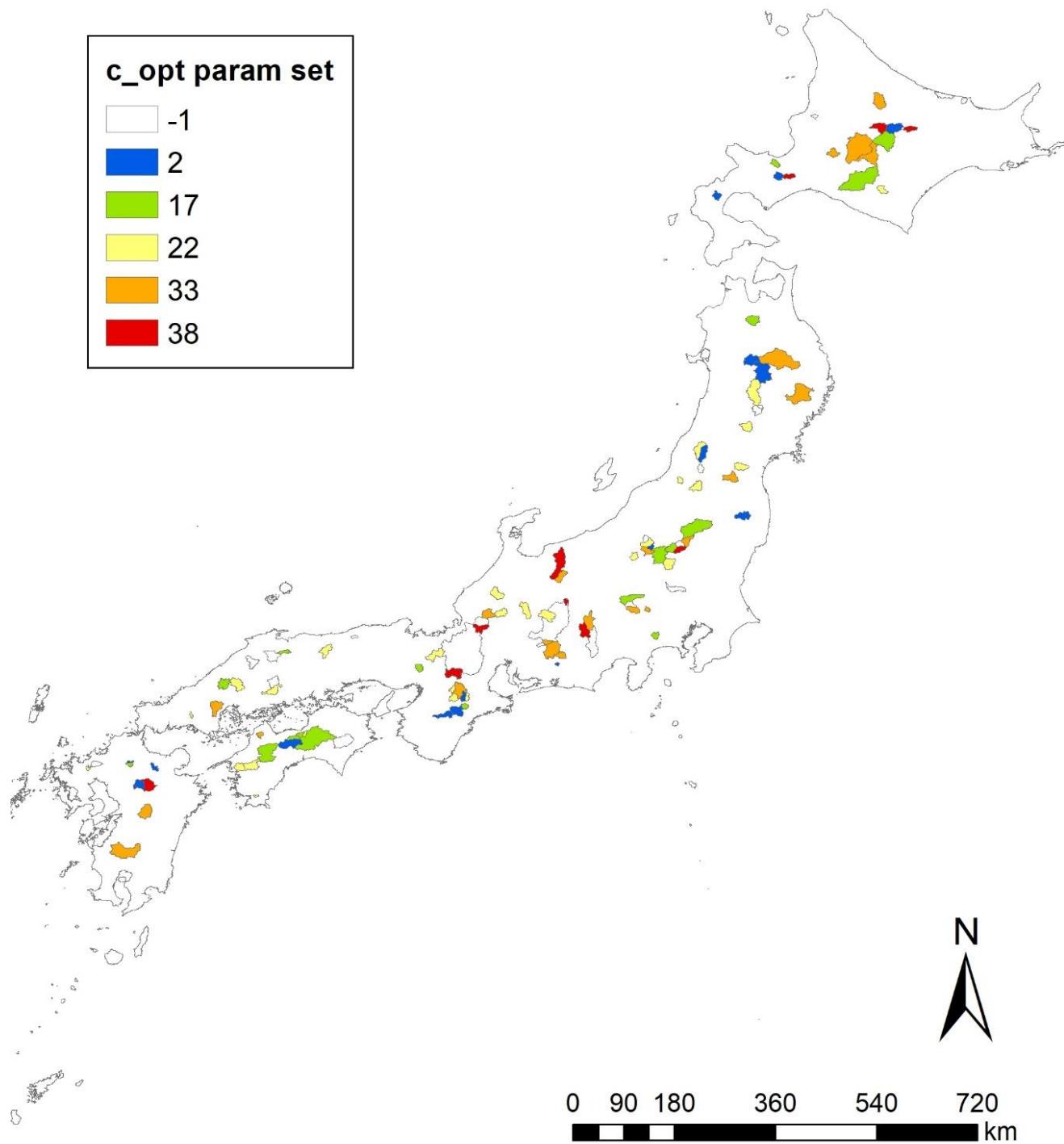


Figure 20: Spatial distributions of the c_{opt} parameter sets

Table 11: Average evaluation index scores of the simulation with the c_opt parameter sets
 Scores for the basins without enough observation data are set to -999.

Dam	CC_{std}	KGE_{std}	NSE_{std}	EPD	Dam	CC_{std}	KGE_{std}	NSE_{std}	EPD
Agigawa	0.92	0.70	0.76	0.09	Naramata	0.89	0.60	0.54	-0.04
Aha	0.85	0.68	0.71	-0.32	Naruko	0.94	0.66	0.65	-0.08
Aimata	0.93	0.61	0.54	0.28	Nibutani	0.92	0.57	0.62	0.01
Amagase	0.57	0.08	0.00	0.58	Nomura	0.96	0.69	0.83	-0.19
Arakawa	0.66	0.11	0.22	-0.21	Nukui	0.81	0.21	0.38	0.09
Aseishigawa	0.84	0.59	0.44	-0.11	Nunome	0.93	0.65	0.77	-0.14
Benoki	0.74	0.39	0.32	-0.22	Obara	-999	-999	-999	-999
Chubetsu	0.78	0.45	0.27	-0.10	Odo	0.99	0.78	0.95	-0.06
Egawa	0.94	0.69	0.81	0.03	Oishi	0.84	0.58	0.38	-0.06
Fujiwara	0.74	0.60	0.44	-0.14	Okawa	0.84	0.50	0.40	0.16
Fukuchi	0.75	0.53	0.52	-0.43	Omachi	0.86	0.28	0.07	0.41
Fungawa	0.78	0.54	0.59	-0.46	Origawa	-999	-999	-999	-999
Futase	0.97	0.73	0.86	0.05	Otaki	0.95	0.83	0.83	0.01
Gassan	0.75	0.28	0.12	0.37	Pirika	0.79	0.55	0.63	-0.12
Gosho	0.96	0.71	0.77	0.09	Rumoi	-999	-999	-999	-999
Hachisu	0.91	0.69	0.68	0.00	Ryumon	0.93	0.68	0.60	0.11
Haiduka	-999	-999	-999	-999	Sagae	0.91	0.67	0.58	-0.04
Haji	0.97	0.78	0.82	0.11	Sagurigawa	-999	-999	-999	-999
Haneiji	0.83	0.55	0.54	-0.25	Sameura	0.96	0.79	0.86	-0.03
Hattabara	0.98	0.72	0.84	0.02	Sarutani	0.97	0.86	0.89	0.01
Hinachi	0.95	0.71	0.76	-0.03	Satsunaigawa	0.92	0.59	0.73	-0.13
Hitokura	0.98	0.74	0.90	-0.02	Shichikashuku	0.90	0.49	0.45	-0.01
Hiyoshi	0.96	0.73	0.86	0.03	Shijushida	0.89	0.64	0.57	0.16
Hoheikyo	0.89	0.39	0.21	0.11	Shimadigawa	0.92	0.72	0.67	0.18
Ikari	0.96	0.75	0.76	0.08	Shimokubo	0.94	0.71	0.81	0.02
Ikeda	0.97	0.66	0.81	0.07	Shimouke	0.95	0.75	0.73	0.10
Isawa	-999	-999	-999	-999	Shingu	0.88	0.56	0.53	0.04
Ishibuchi	-999	-999	-999	-999	Shintoyone	0.95	0.81	0.84	-0.10
Ishitegawa	0.94	0.45	0.52	0.18	Shirakawa	0.87	0.57	0.56	-0.09
Iwaonai	0.82	0.52	0.44	0.35	Shitsumi	-999	-999	-999	-999
Iwaya	0.95	0.65	0.64	0.09	Shorenji	0.98	0.76	0.86	0.04
Izarigawa	0.81	0.33	0.36	-0.01	Sonohara	0.96	0.84	0.87	0.04
Jozankei	0.84	0.47	0.44	0.08	Sugesawa	0.94	0.64	0.72	-0.15
Kamafusa	0.93	0.70	0.69	0.10	Taiho	-999	-999	-999	-999
Kanayama	0.88	0.56	0.36	0.18	Taisetsu	0.87	0.51	0.31	0.26
Kanna	0.60	0.15	0.36	-0.52	Takayama	0.94	0.76	0.83	-0.09
Kanogawa	0.96	0.75	0.85	-0.10	Takisato	0.89	0.55	0.39	0.14

Kanoko	0.77	0.43	0.27	0.01	Takizawa	-999	-999	-999	-999
Kasegawa	-999	-999	-999	-999	Tamagawa	0.95	0.67	0.69	0.01
Katsurazawa	0.90	0.78	0.73	-0.12	Tase	0.91	0.56	0.41	0.25
Kawaji	0.76	0.28	0.34	-0.71	Tedorigawa	0.93	0.75	0.76	-0.07
Kawamata	0.97	0.73	0.71	0.16	Terauchi	0.96	0.78	0.87	-0.10
Koshibu	0.92	0.64	0.69	-0.01	Tokachi	0.85	0.53	0.47	0.10
Kurashiki	-999	-999	-999	-999	Tokuyama	-999	-999	-999	-999
Kusaki	0.93	0.73	0.70	-0.03	Tomata	0.96	0.75	0.71	0.06
Kuzuryu	0.92	0.73	0.67	-0.08	Tomisato	0.96	0.80	0.90	-0.11
Kyuragi	0.90	0.16	0.06	0.22	Tono	-999	-999	-999	-999
Makio	0.95	0.78	0.82	-0.05	Tsuruda	0.93	0.77	0.78	0.05
Managawa	0.93	0.42	0.37	0.15	Unaduki	0.74	0.42	0.24	-0.17
Maruyama	-999	-999	-999	-999	Urayama	0.96	0.77	0.81	0.03
Matsubara	0.93	0.72	0.69	0.27	Ure	0.89	0.66	0.69	-0.20
Midorikawa	0.95	0.71	0.80	-0.08	Wataraseyusuichi	-999	-999	-999	-999
Miharu	0.85	0.53	0.55	0.02	Yabakei	0.96	0.82	0.90	-0.23
Misogawa	0.84	0.52	0.38	-0.10	Yagisawa	0.82	0.64	0.42	-0.11
Miwa	0.94	0.54	0.42	0.39	Yahagi	0.93	0.75	0.75	-0.11
Miyagase	0.93	0.74	0.76	-0.14	Yanase	0.93	0.73	0.70	-0.03
Muro	0.94	0.67	0.73	0.06	Yasaka	0.94	0.51	0.61	0.26
Nagai	-999	-999	-999	-999	Yokoyama	0.82	0.52	0.16	0.50
Nagashima	-999	-999	-999	-999	Yuda	0.95	0.77	0.80	-0.15
Nagayasuguchi	-999	-999	-999	-999	Yunishigawa	-999	-999	-999	-999
Nakasujigawa	0.95	0.70	0.82	-0.15					

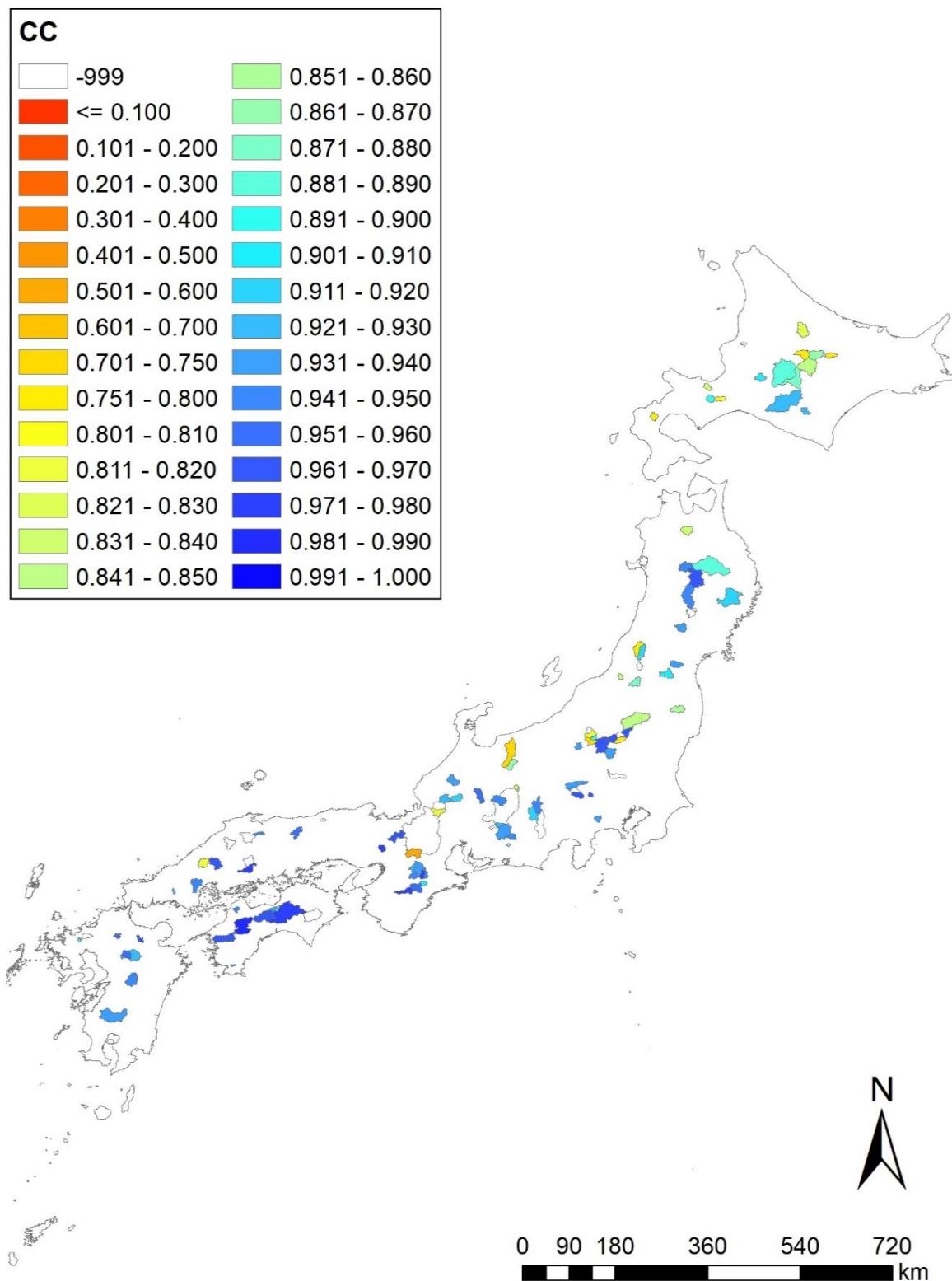


Figure 21: Spatial distributions of average CC_{std} for the c_{opt} parameter sets

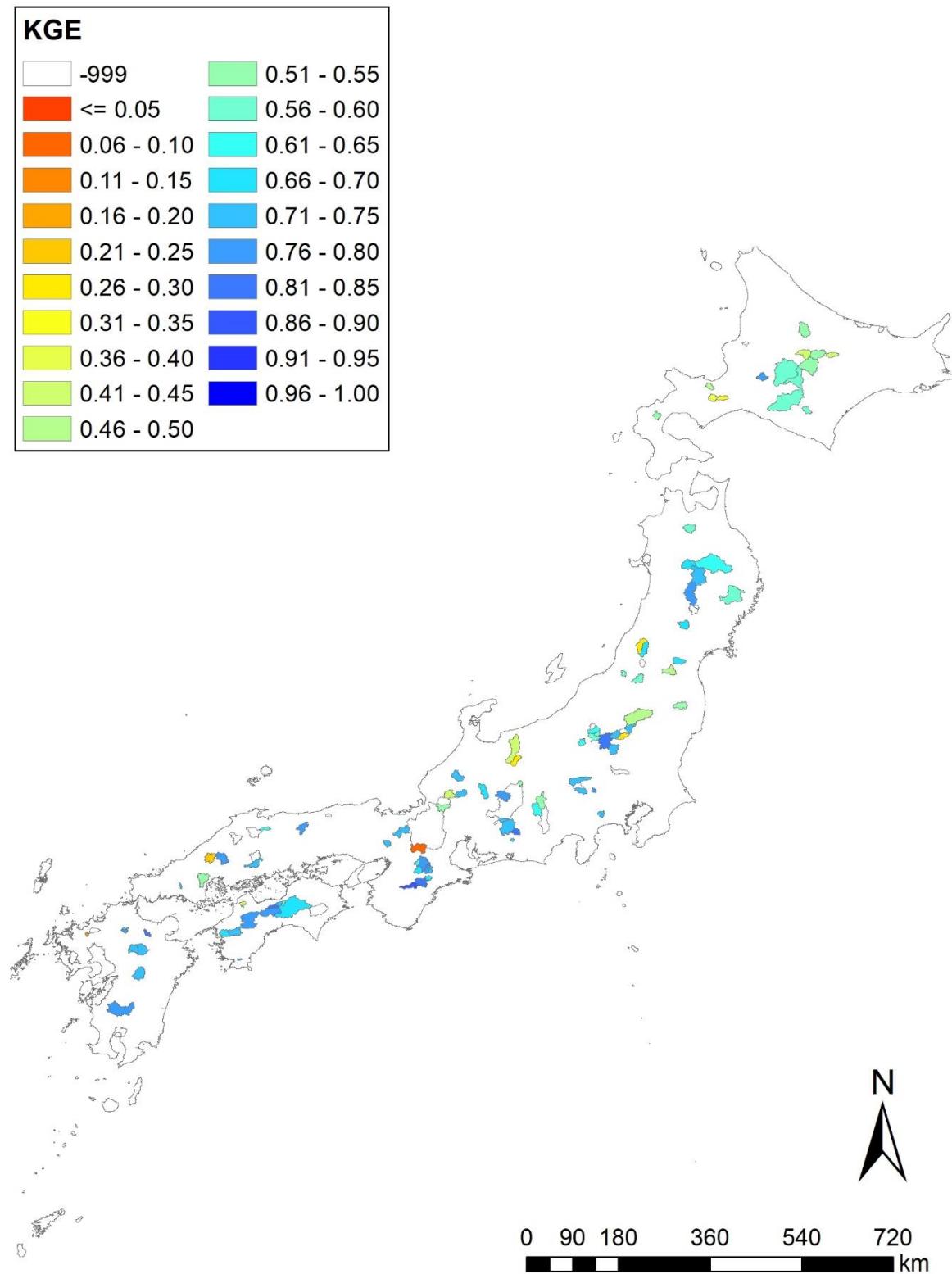


Figure 22: Spatial distributions of average KGE_{std} for the c_{opt} parameter sets

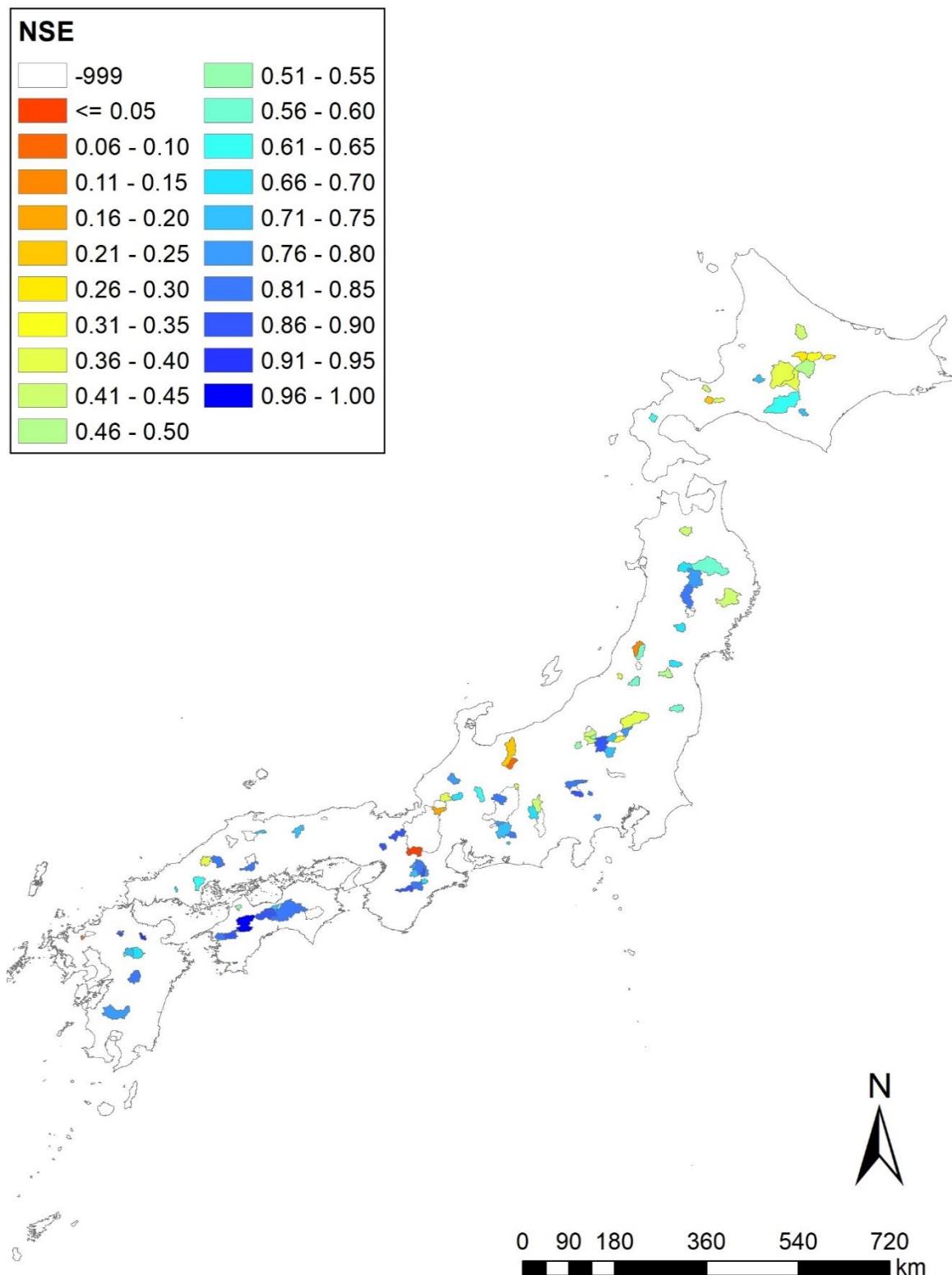


Figure 23: Spatial distributions of average NSE_{std} for the c_{opt} parameter sets

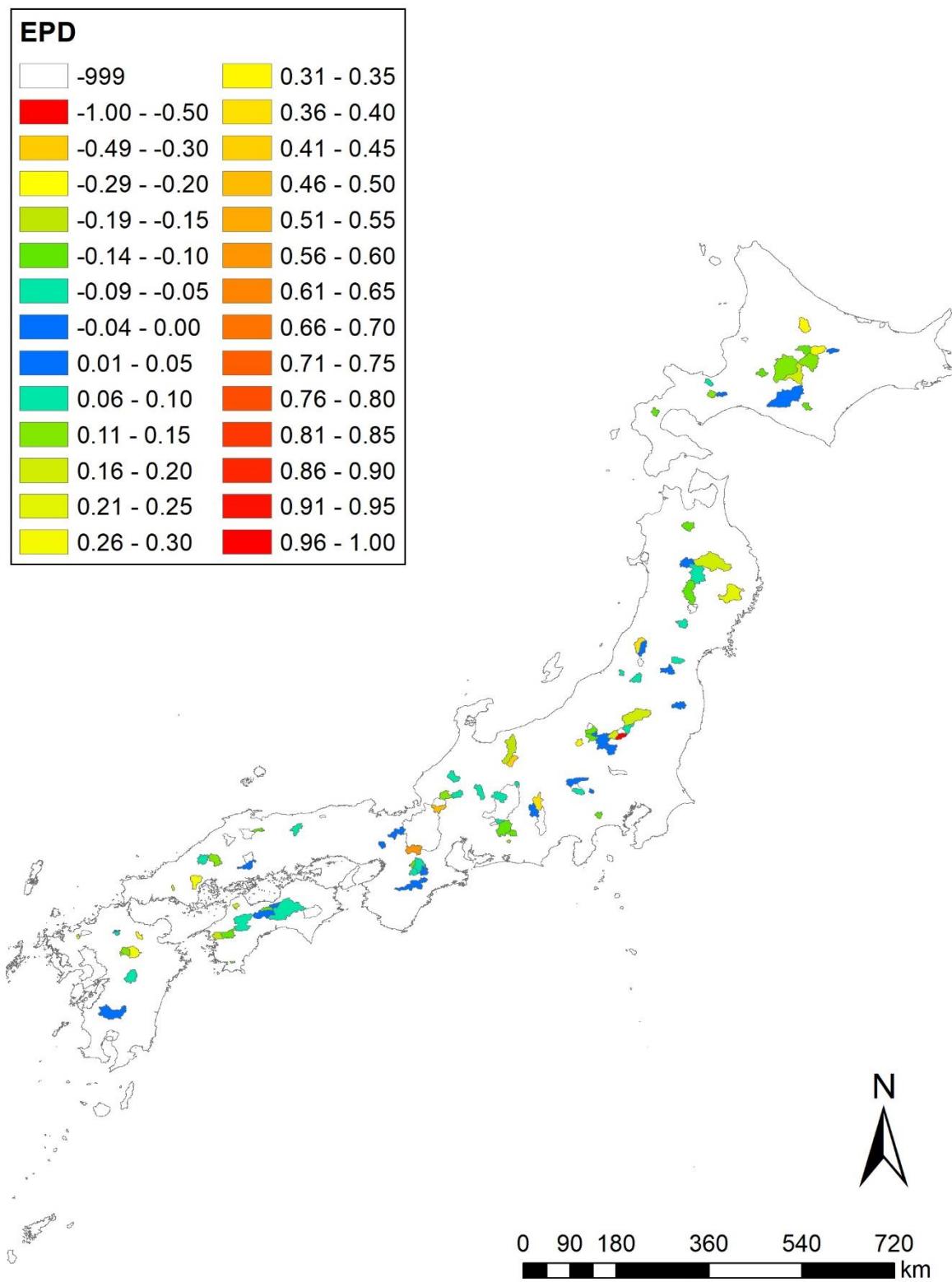


Figure 24: Spatial distributions of average *EPD* for the *c_opt* parameter sets

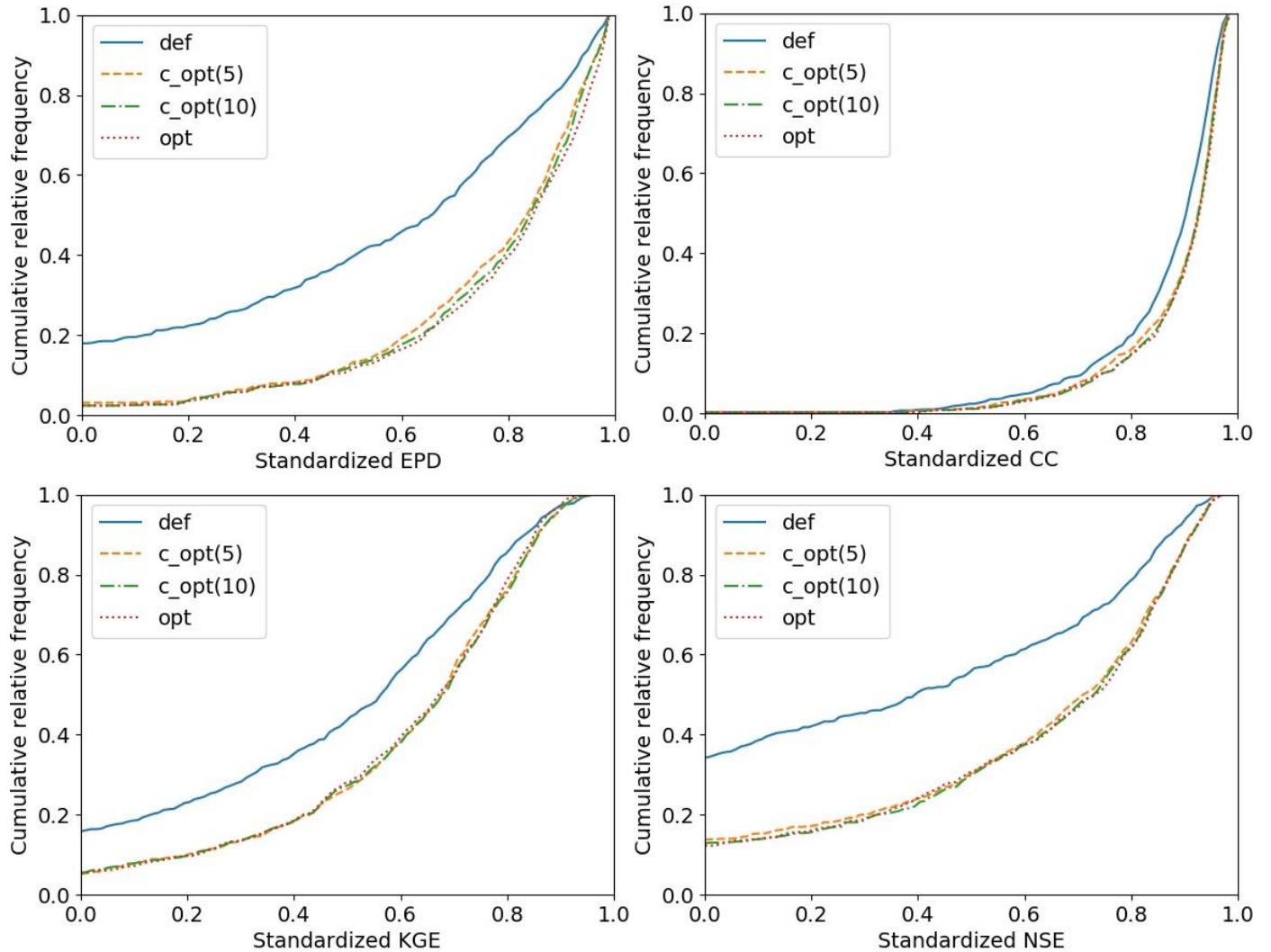


Figure 25: Cumulative relative frequency of evaluation index scores for the default, opt and c_opt parameter sets

Table 12: Dam reservoir basins and the clusters they belong to

Cluster IDs for the basins without enough observation data are set to -1.

Dam	Cluster	Dam	Cluster	Dam	Cluster
Agigawa	2	Kawamata	2	Shichikashuku	2
Aha	4	Koshibu	3	Shijushida	3
Aimata	2	Kurashiki	-1	Shimadigawa	2
Amagase	3	Kusaki	1	Shimokubo	2
Arakawa	4	Kuzuryu	1	Shimouke	1
Aseishigawa	2	Kyuragi	2	Shingu	1
Benoki	4	Makio	1	Shintoyone	1
Chubetsu	3	Managawa	3	Shirakawa	2
Egawa	1	Maruyama	-1	Shitsumi	-1
Fujiwara	3	Matsubara	2	Shorenji	1
Fukuchi	4	Midorikawa	2	Sonohara	2
Fungawa	4	Miharu	2	Sugesawa	1
Futase	2	Misogawa	3	Taiho	-1
Gassan	2	Miwa	3	Taisetsu	1
Gosho	1	Miyagase	1	Takayama	2
Hachisu	1	Muro	1	Takisato	3
Haiduka	-1	Nagai	-1	Takizawa	-1
Haji	2	Nagashima	-1	Tamagawa	1
Haneji	4	Nagayasuguchi	-1	Tase	3
Hattabara	2	Nakasujigawa	1	Tedorigawa	1
Hinachi	1	Naramata	1	Terauchi	1
Hitokura	2	Naruko	1	Tokachi	3
Hiyoshi	1	Nibutani	2	Tokuyama	-1
Hoheikyo	2	Nomura	1	Tomata	1
Ikari	2	Nukui	2	Tomisato	1
Ikeda	1	Nunome	1	Tono	-1
Isawa	-1	Obara	-1	Tsuruda	2
Ishibuchi	-1	Odo	1	Unaduki	3
Ishitegawa	2	Oishi	1	Urayama	2
Iwaonai	3	Okawa	3	Ure	4
Iwaya	2	Omachi	3	Wataraseyusuichi	-1
Izarigawa	3	Origawa	-1	Yabakei	1
Jozankei	3	Otaki	1	Yagisawa	1
Kamafusa	2	Pirika	4	Yahagi	2
Kanayama	3	Rumoi	-1	Yanase	1
Kanna	4	Ryumon	1	Yasaka	3
Kanogawa	1	Sagae	1	Yokoyama	3

Kanoko	3	Sagurigawa	-1	Yuda	1
Kasegawa	-1	Sameura	1	Yunishigawa	-1
Katsurazawa	2	Sarutani	1		
Kawaji	4	Satsunaigawa	1		

Table 13: Centroids of the clusters

Rain is in descending order of the total amount of rainfall.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
<i>EPD</i> (rain rank 1)	-0.10	0.28	0.66	-0.46
<i>EPD</i> (rain rank 2)	0.00	0.30	0.80	-0.43
<i>EPD</i> (rain rank 3)	0.06	0.42	0.81	-0.35
<i>EPD</i> (rain rank 4)	0.05	0.46	0.93	-0.36
<i>EPD</i> (rain rank 5)	0.07	0.52	0.82	-0.40
<i>EPD</i> (rain rank 6)	0.04	0.66	0.98	-0.47
<i>EPD</i> (rain rank 7)	0.09	0.53	0.94	-0.66
<i>CC_{std}</i> (rain rank 1)	0.88	0.86	0.67	0.53
<i>CC_{std}</i> (rain rank 2)	0.85	0.80	0.60	0.39
<i>CC_{std}</i> (rain rank 3)	0.87	0.80	0.57	0.59
<i>CC_{std}</i> (rain rank 4)	0.87	0.85	0.63	0.55
<i>CC_{std}</i> (rain rank 5)	0.86	0.79	0.66	0.59
<i>CC_{std}</i> (rain rank 6)	0.87	0.76	0.56	0.35
<i>CC_{std}</i> (rain rank 7)	0.86	0.80	0.60	0.37

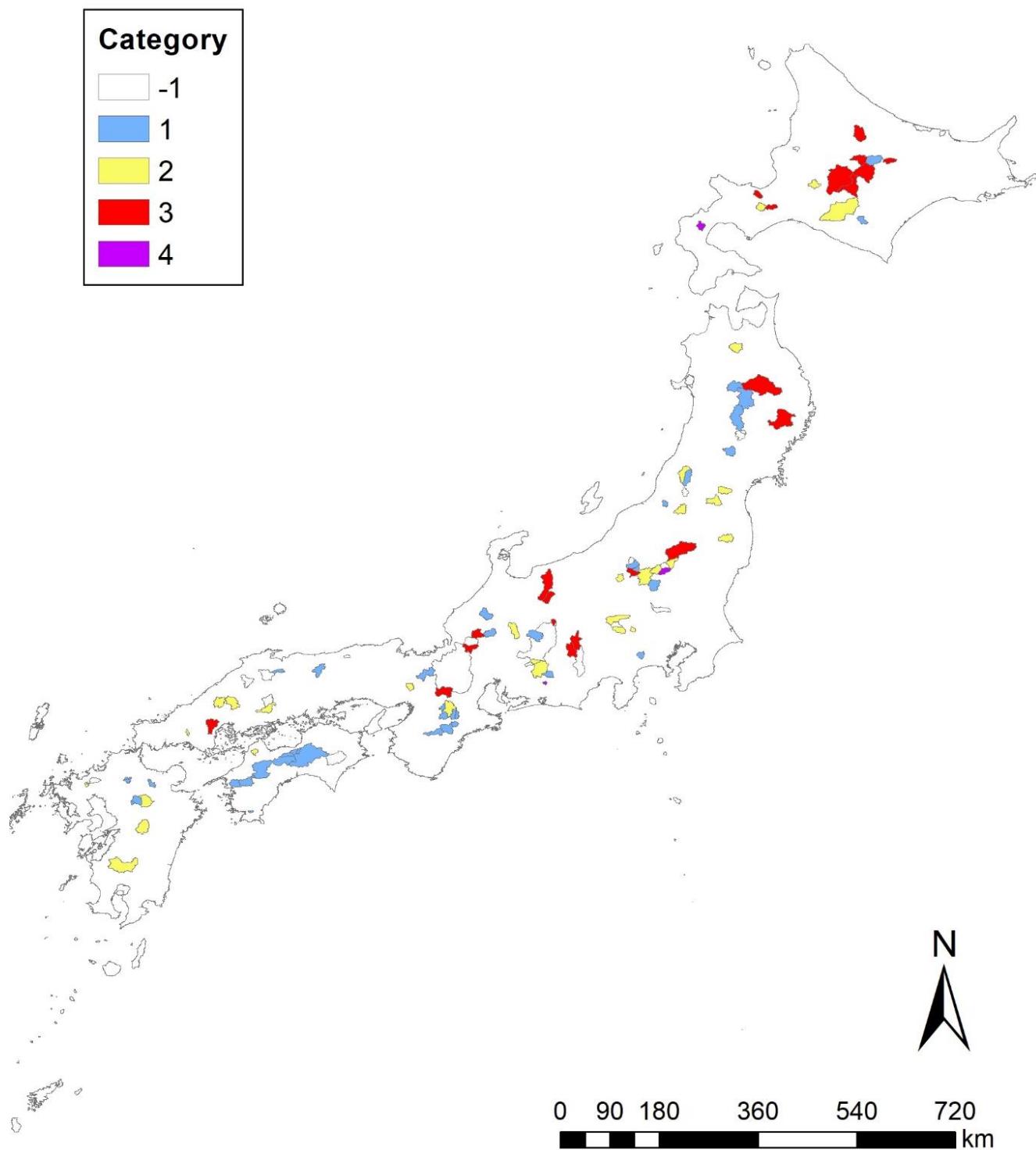


Figure 26: Spatial distributions of category members

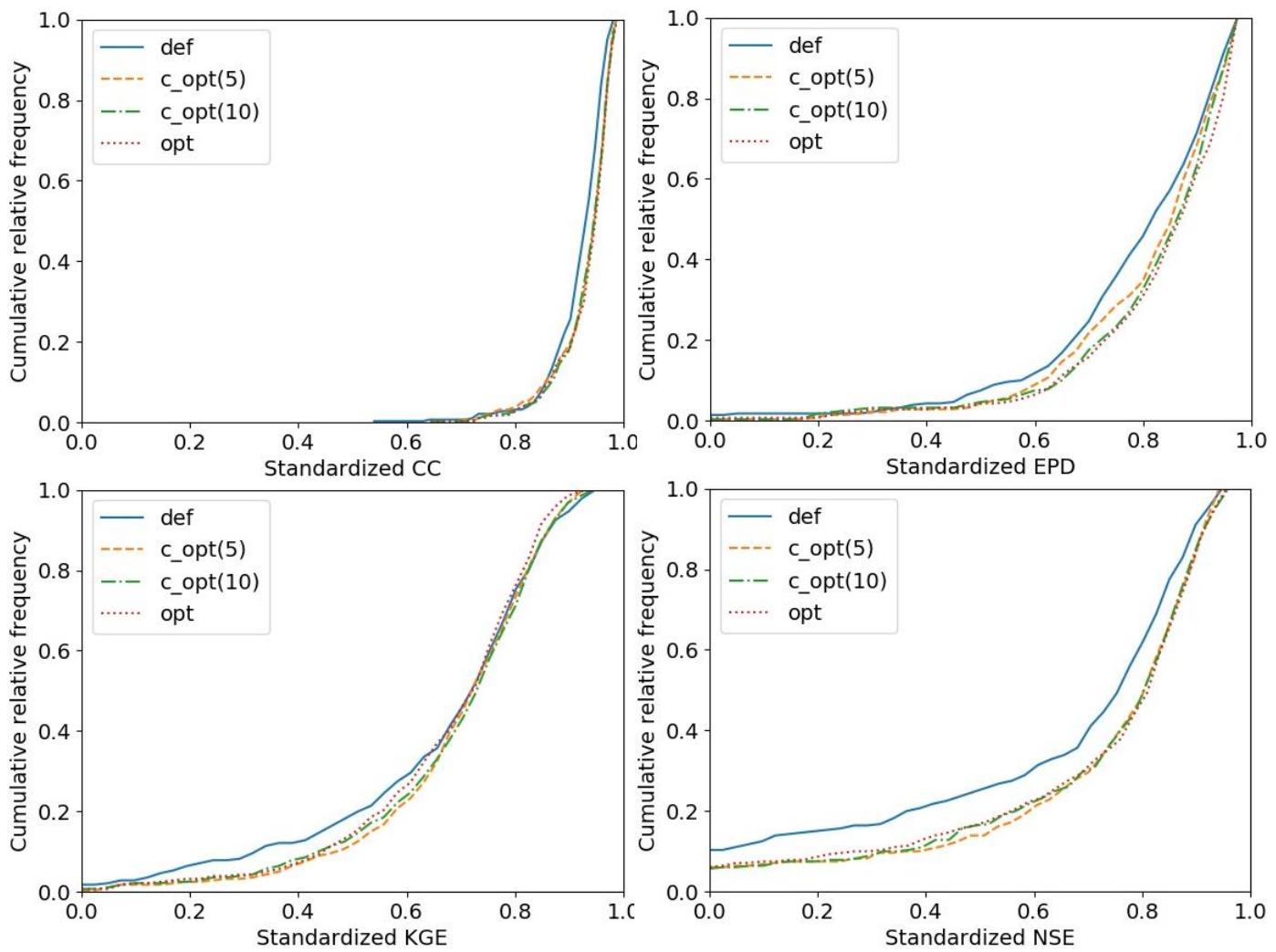


Figure 27: Cumulative relative frequency plots of the evaluation index scores for the basins in the Category 1

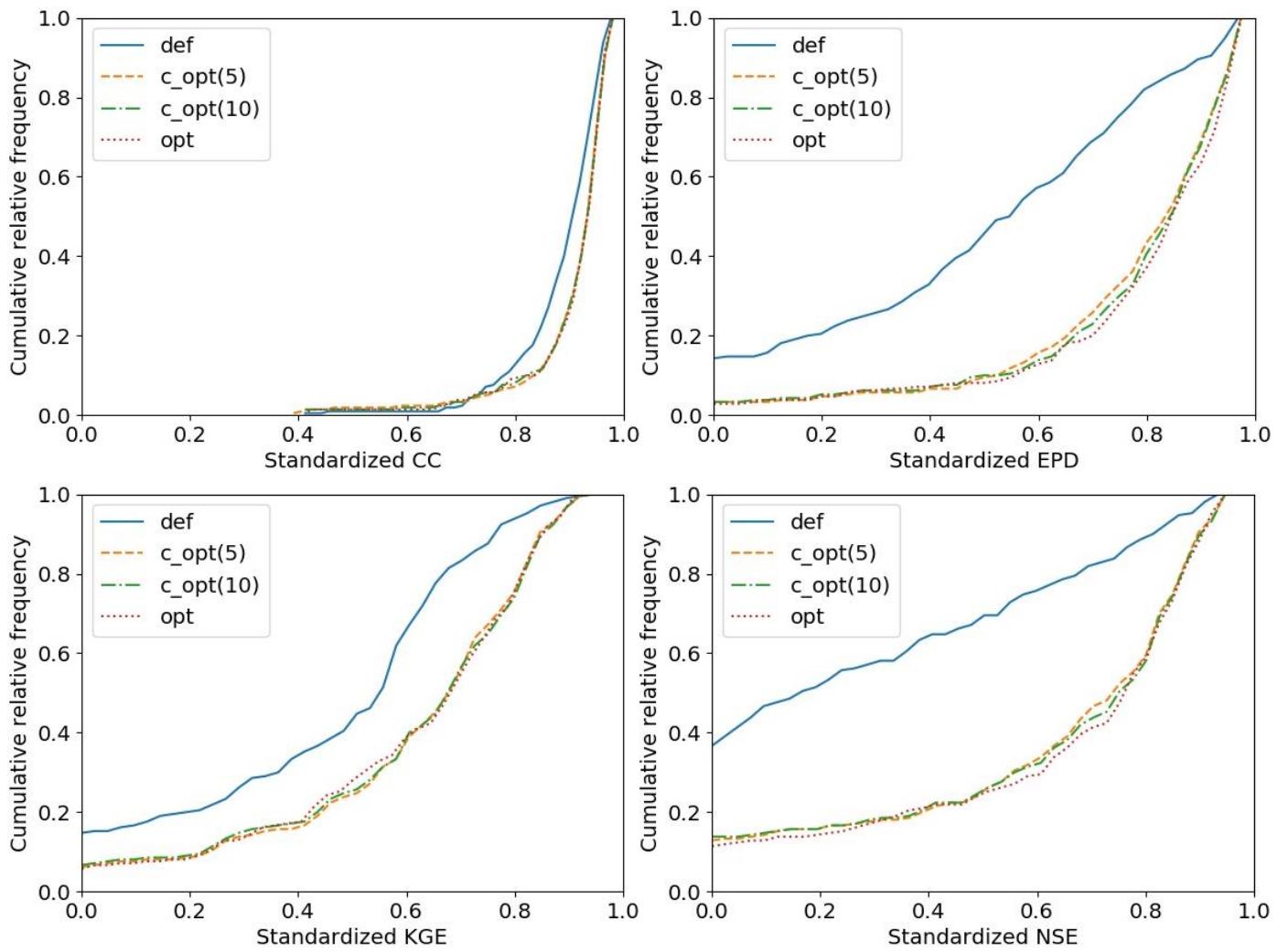


Figure 28: Cumulative relative frequency plots of the evaluation index scores for the basins in the Category 2

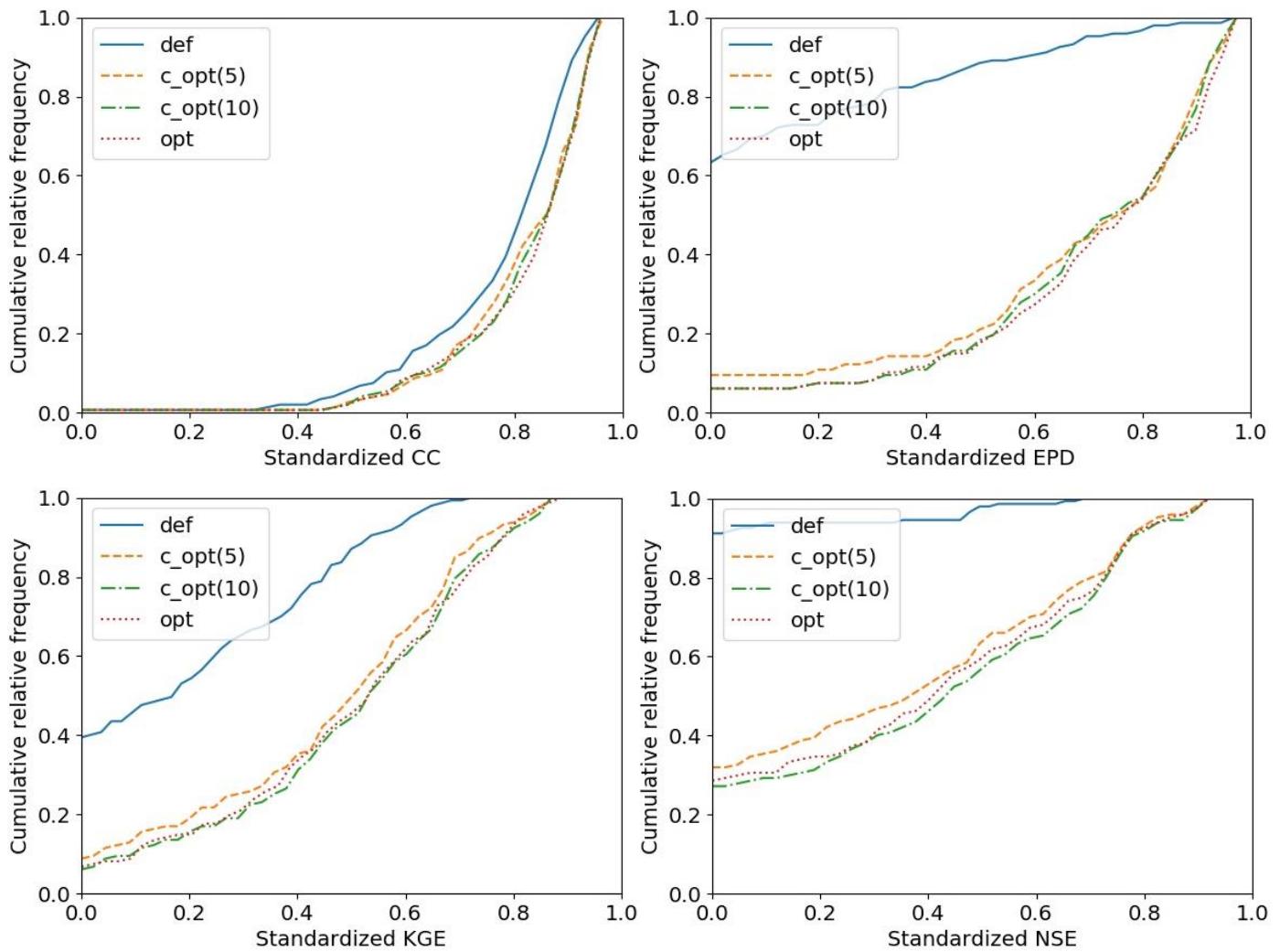


Figure 29: Cumulative relative frequency plots of the evaluation index scores for the basins in the Category 3

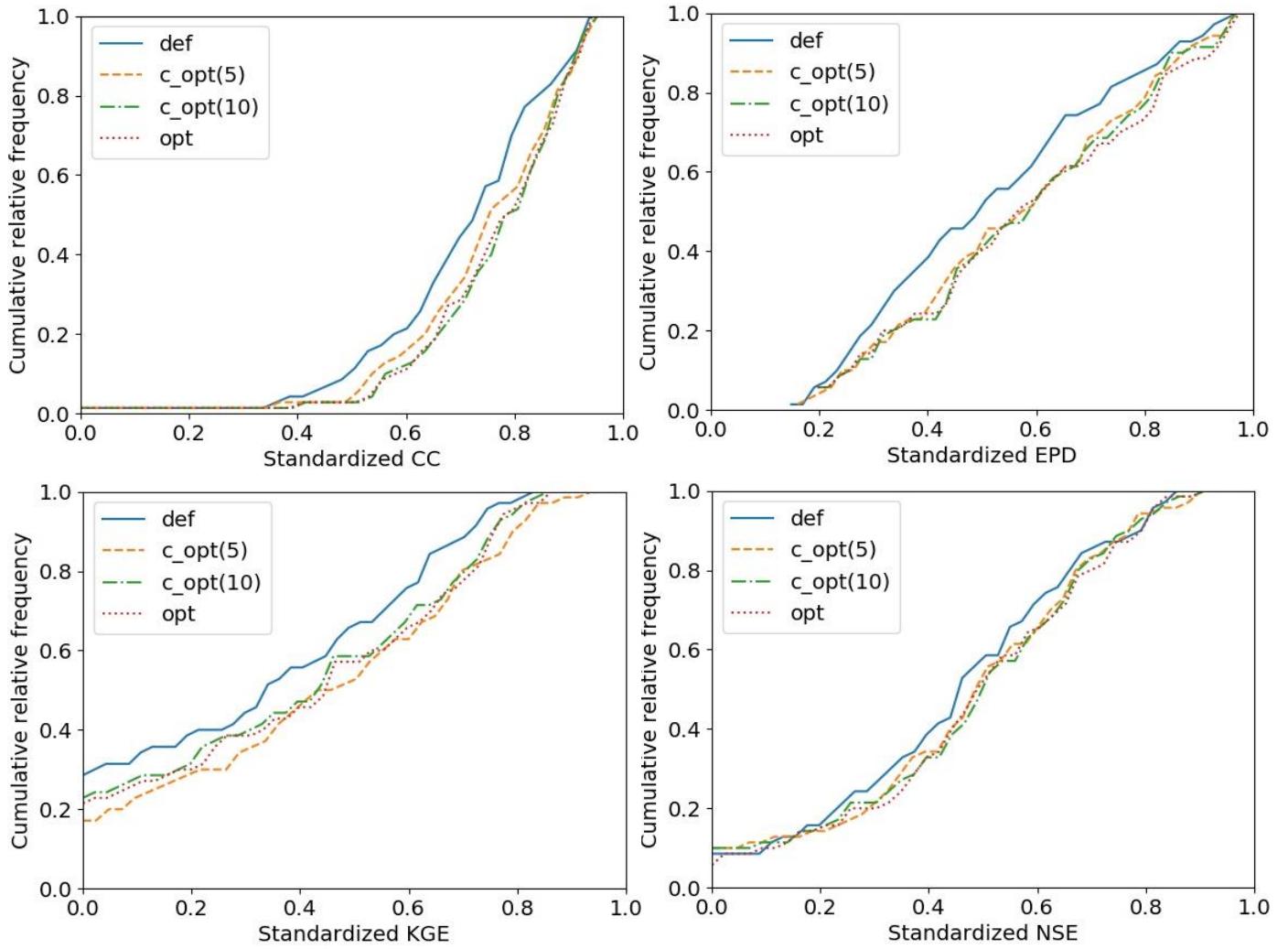


Figure 30: Cumulative relative frequency plots of the evaluation index scores for the basins in the Category 4

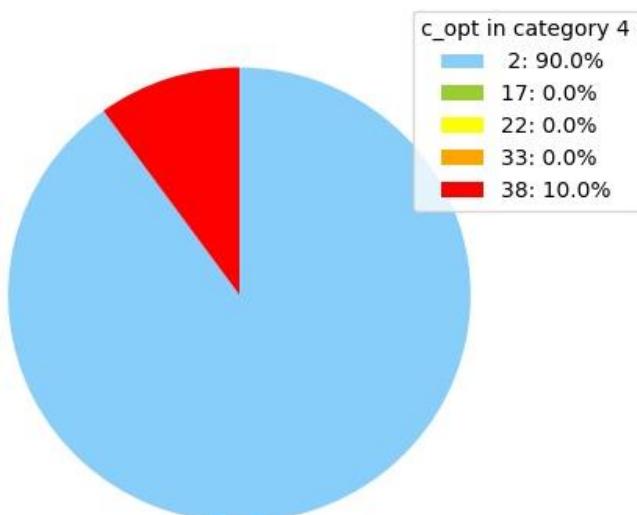
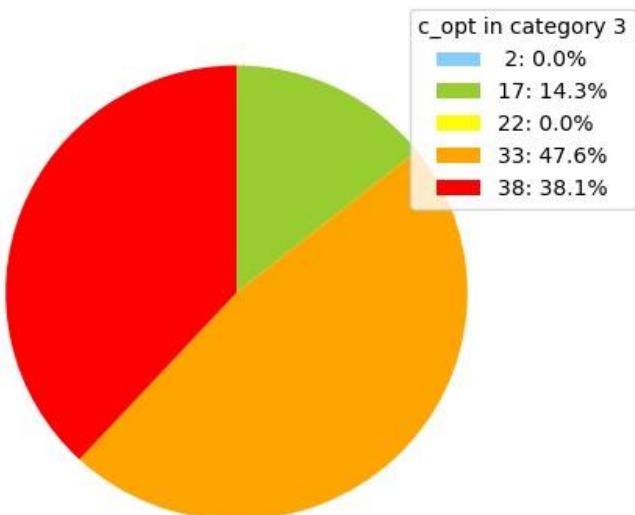
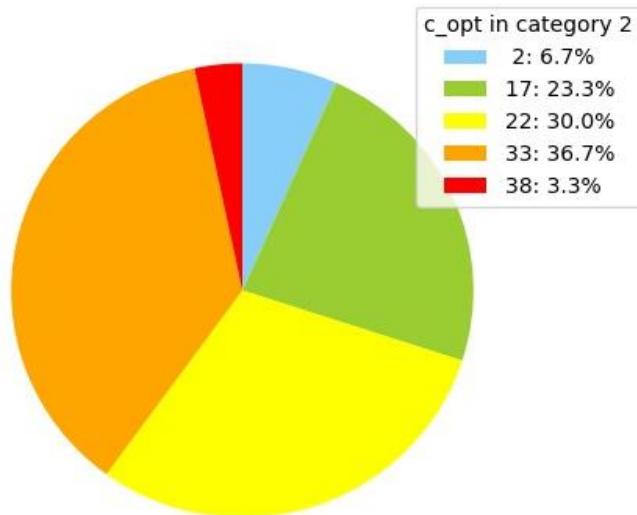
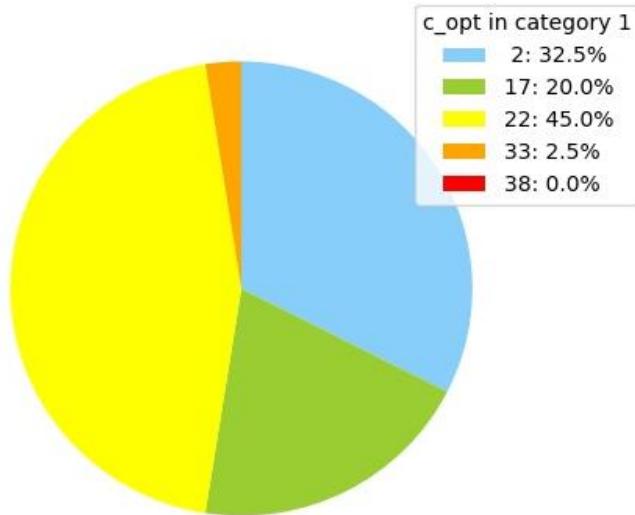


Figure 31: Ratio of c_opt parameter sets in each category

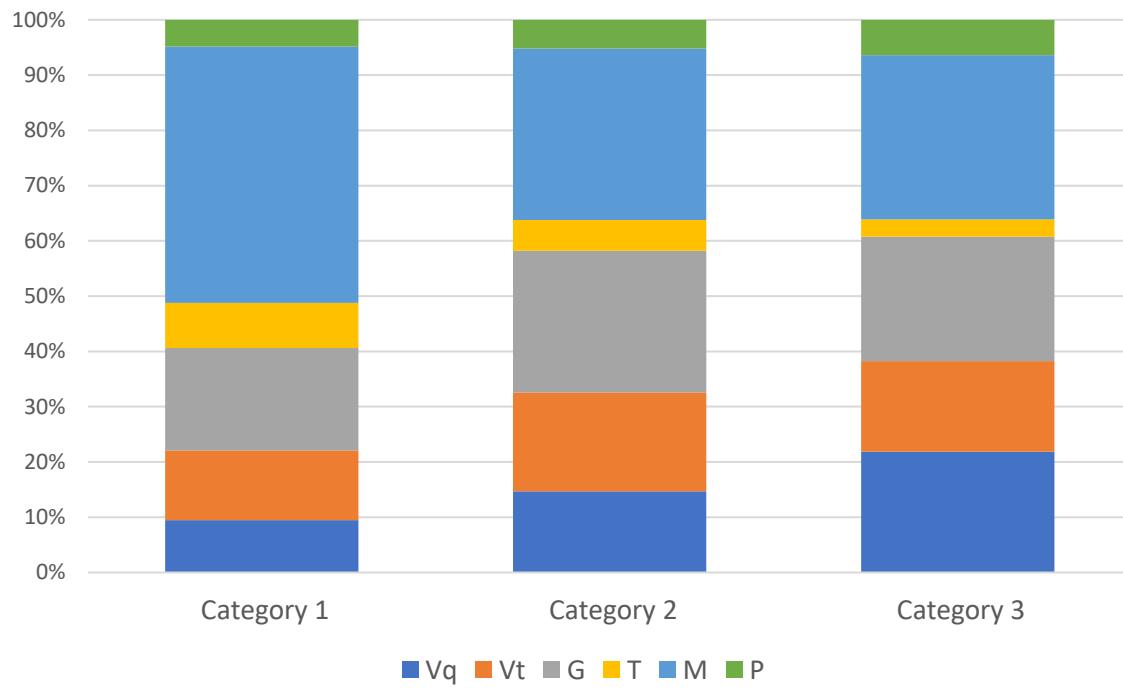


Figure 32: Ratios of Bedrock Types for Each Category

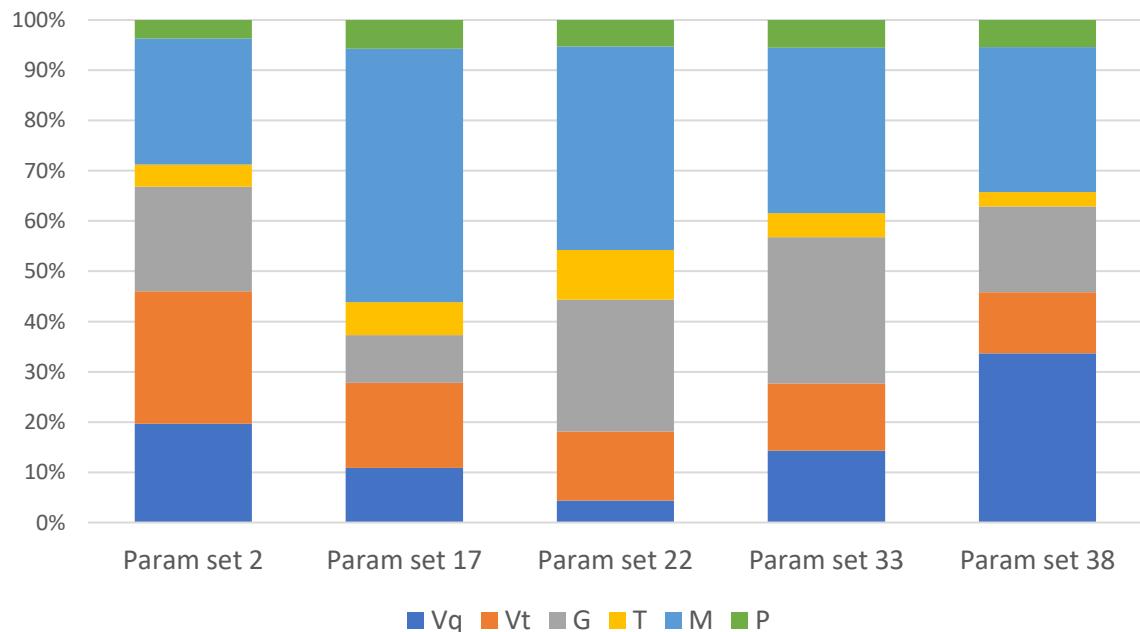


Figure 33: Ratio of bedrock types for each c_opt parameter set

References

- [1] S. Solomon, Dawei Qin, Martin Manning, ZeNa Chen, Melinda Marquis, K. Avery, M. Tignor and H. Miller. 2007. Climate Change 2007: The Physical Science Basis. Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press.
- [2] Fumiaki Fujibe, Nobuo Yamazaki, Mitsugi Katsuyama and Kenji Kobayashi. 2005. The increasing trend of intense precipitation in Japan based on four-hourly data for a hundred years. SOLA, 1, 41-44
- [3] Climate Change Adaptation Research Group, NILM, MLIT. 気候変動適応策に関する研究（中間報告）. 2013. URL: <http://www.nilim.go.jp/lab/bcg/siryou/tnn/tnn0749.htm>. Accessed on 2019/12/10.
- [4] Takahiro Sayama, Yuya Tatebe, Shigenobu Tanaka. 2015. Hydrologic sensitivity of flood runoff and inundation: 2011 Thailand floods in the Chao Phraya River basin. Nat. Hazards Earth Syst. Sci., 15, 1617-1630.
- [5] Dai Yamazaki, Saeka Togashi, Akira Takeshima, Takahiro Sayama. 2018. High-resolution flow direction map of Japan. Journal of Japan Society of Civil Engineers, Ser B1 (Hydraulic Engineering), 74, 163-168.
- [6] Qingyun Duan, Soroosh Sorooshian, Vuai Gupta. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resource Research, 28, 4, 1015-1031.
- [7] William J. Cook, William H. Cunningham, William R. Pulleyblank, Alexander Schrijver. 1997. Combinatorial Optimization. 1st ed. New York, Wiley.
- [8] Krishan P. Singh. 1971. Model flow duration and streamflow variability. Water Resource Research, 7, 1031-1036.

- [9] Maria Mimikou, Stella Kaemaki. 1985. Regionalization of flow duration characteristics. *Journal of Hydrology*, 82, 77-91.
- [10] Katumi Mushiake, Yutaka Takahashi and Yoshihisa Ando. 1981. Effects of basin geology on river-flow regime in mountainous areas of Japan. *Journal of JSCE*, 309, 51-62.
- [11] Christopher Soulsby, Doerthe Tetzlaff, P. Rodgers, S. Dunn, S. Waldron. 2006. Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: an initial evaluation. *Journal of Hydrology*, 325, 197–221.
- [12] T. P. Burt, W. T. Swank. 1992. Flow frequency responses to hardwood-to-grass conversion and subsequent succession. *Hydrological Processes*, 6, 179-188.
- [13] Yasuto Tachikawa, Gen Nagatani, Kaoru Takara. 2004. Development of stage discharge relationship equation incorporating saturated-unsaturated flow mechanism (in Japanese), *JSCE Annual Journal of Hydraulic Engineering*, 48, 7-12. (in Japanese with English Abstract).
- [14] George A. Bekey, Sami F. Masri. 1983. Random search techniques for optimization of nonlinear systems with many parameters. *Mathematics and Computers in Simulation*, 25, 3, 210-213,
- [15] Luc Pronzato, Eric Walter, Alain Venot, Jean-Francois Lebruchec. 1984. A general-purpose global optimizer: Implementation and applications. *Mathematics and Computers in Simulation*, 26, 5, 412-422.
- [16] David Arthur and Sergei Vassilvitskii. 2007. k-means++: the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1027-1035.
- [17] Hoshin V. Gupta, Harald Kling, Koray K. Yilmaz, Guillermo F. Martinez. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80-91.

- [18] Yoshiyuki Yokoo, Taikan Oki. 2010. Effects of climate, topography, soil, geology and land use on flow regimes in Japanese mountainous watersheds. JSCE Annual Journal of Hydraulic Engineering, 54, 469-474. (in Japanese with English Abstract).
- [19] Takashi Jitousono, Keiji Takeshita. 1988. Relations between flow regime and basin conditions in mountainous areas: II. Effects of basin geology on flow regime (in Japanese). Bulletin of the Kagoshima University Forest, 15, 15-38.
- [20] Yuko Kato, Yuichi Onda, Takahisa Mizuyama, Ken-ichiro Kosugi, Ai Yoshikawa, Maki Tsujimura, Koji Hata, Masao Okamoto. 2000. The difference of runoff peak response time in upstream of Ibi River underlain by different geology. Journal of the Japan Society of Erosion Control Engineering, 53, 4, 38-43. (in Japanese with English Abstract).
- [21] Geological Survey of Japan, AIST (ed.). 2015. Seamless digital geological map of Japan 1: 200,000. May 29, 2015 version. Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology.
- [22] Takashi Jitousono, Keiji Takeshita. 1987. Relations between flow regime and basin conditions in mountainous areas: III. Effects of basin topography and geology on flow regime (in Japanese). Bulletin of the Kagoshima University Forest, 16, 61-81.

ACKNOWLEDGEMENT

It would be impossible to complete this graduation thesis without the support from individuals and parties who in a way or another had contributed their valuable times and ideas. In this acknowledgement opportunities, I would like to express my sincere gratitude to those individuals.

My deepest appreciation goes to Prof. Takahiro Sayama whose enormous support and insightful comments were invaluable during the course of my study.

I also owe a very important debt to Ms. Karin Yamamoto, Dr. Masafumi Yamada and Mr. Yoshito Sugawara whose sharp indications were an enormous help to me.

Further, Mr. Koji Matsumoto helped me relax and refresh myself in the laboratory by chatting or playing table tennis with me.

Furthermore, I would also like to express my gratitude to my family for building a house near the laboratory. I could come to the laboratory whenever I would like to work.

Last but not least, I am grateful to all the people who helped me and gave me useful advices during the research.