Application of Cluster Analysis for Finding Operational Patterns of Multireservoir System during Transition Period

Duan Chen[1]; Arturo S. Leon, M.ASCE[2]; Parnian Hosseini[3]; Nathan L. Gibson[4]; and Claudio Fuentes[5]

Abstract: Operational objectives and/or constraints of a reservoir system may need to be shifted at certain periods (i.e., transition periods) due to seasonal considerations of human interest and ecological benefits. Despite the fact that operational schemes in the transition periods are critical and of great interest to reservoir operation practice, the problem has received little attention in the literature. This paper presents a study on cluster analysis for identifying patterns of operational schemes during transition periods. The test case corresponds to ten major reservoirs of the Federal Columbia River Power System (FCRPS) in the United States. The operation horizon consists of two weeks during which the objectives of the reservoir system are shifted based on seasonal consideration for fish migration and survival. An optimization model based on an evolutionary algorithm is used to derive the optimal operational schemes under various inflow scenarios. A K-Spectral Centroid algorithm (K-SC) is applied to the resulting operational schemes to find clusters of the schemes based on similarities of their temporal shapes. By investigating the relations between the clusters and the inflow scenarios, general patterns of operational schemes are identified. The analyses offer insights into the operational schemes during the transition period and broaden the understanding of short-term reservoir operation with shifting operational objectives. **DOI:** 10.1061/(ASCE)WR.1943-5452.0000772.

Author keywords: Cluster analysis; Multireservoir system; Operational patterns; Shifting objectives.

Introduction

Reservoir operation normally provides multiple benefits to human interests, including flood control, hydropower generation, and irrigation. Recently, restoration of river ecosystems is being considered in reservoir operation to address growing concerns of ecological and environmental protection. Flow requirements for the biota in the river—i.e., fish communities (Cardwell et al. 1996; Chen et al. 2013), riparian vegetation (Morrison and Stone 2015; Richter and Richter 2000), and macroinvertebrate communities (Maynard and Lane 2012)—are considered for adapting reservoir operation. However, some of the requirements regarding the river ecosystem are seasonal, e.g., fish migration, and they are normally emphasized only during specific periods. As a result, the operational considerations (either the objectives or constraints or both) are shifted at specific times (i.e., transition periods). Reservoir operation schemes during a transition period are expected to achieve an optimal trade-off between the operational objectives both before and after the transition.

[1]Senior Engineer, Changjiang River Scientific Research Institute, Wuhan 430010, China (corresponding author). ORCID: http://orcid.org/0000-0002-1996-370X. E-mail: chenduan@mail.crsri.cn

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Houston, Houston, TX 77204. E-mail: aleon3@central.uh.edu

[3]Ph.D. Candidate, School of Civil and Construction Engineering, Oregon State Univ., Corvallis, OR 97331. E-mail: hosseinp@oregonstate.edu

[4]Associate Professor, Dept. of Mathematics, Oregon State Univ., Corvallis, OR 97331. E-mail: gibsonn@oregonstate.edu

[5]Assistant Professor, Dept. of Statistics, Oregon State Univ., Corvallis, OR 97331. E-mail: fuentesc@science.oregonstate.edu

Note. This manuscript was submitted on March 16, 2016; approved on January 6, 2017; published online on April 12, 2017. Discussion period open until September 12, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496.

Shifting operational objectives have been frequently discussed in the context of long-term planning studies (Lund 1996; Wurbs 1991). The shifts occur mostly because the original objectives and/or constraints are replaced with others that can better serve the new requirements for the reservoir system. These changes of the operational considerations are due to regional economic development or climate impacts (Jager and Smith 2008; Li et al. 2010; Loucks 1992; Raje and Mujumdar 2010), which typically happen during a relatively long time frame such as decades. The shift of objectives and/or constraints in this long time frame context may have an influence on the short-term reservoir operation due to the connection between long-term water control plans and the prescribed rules for short-term operation. However, the influence is mostly significant for a long time scale such as years. For a short-term operation, the shift of operational objectives and/or constraints within the long-term planning are not considered.

In the context of short-term reservoir operation (i.e., within one year), many studies considering ecological interests have been made. However, most of these studies highlight the implementation of ecological interests in reservoir operations (Chen et al. 2015; Homa et al. 2005) and focus on achieving an optimal trade-off between the original human interests, e.g., power generation, and the added ecological interests, e.g., ecological flow (Olivares 2008; Suen and Wang 2010). Very few studies have been conducted on reservoir operation in a transition period during which the objectives and/or constraints are shifted from one set to another due to seasonal requirements of the river ecosystem. Eschenbach et al. (2001) emphasized the need for reservoir managers to adapt quickly to changing objectives. Smith et al. (2007) argued that shifting operational objectives and constraints on ecological interests is a future challenge of reservoir operation for meeting dynamic and changing requirements. These discussions show the need for investigating optimal schemes for reservoir operations during a transition period.

Optimal schemes for reservoir operation are typically obtained by intensive simulation or, alternatively, by optimization algorithms. In

J. Water Resour. Plann. Manage., 2017, 143(8): 04017028

addition to traditional optimization approaches such as Newton's method, evolutionary algorithms—e.g., genetic algorithms—have been receiving increasing attention in reservoir operation (Atiquzzaman et al. 2006; Prasad and Park 2004; Reed et al. 2013; Yandamuri et al. 2006; Yin and Yang 2011) due to their ability to find global (and not just local) optima. Data-mining techniques are also applied frequently for identifying operational schemes of reservoir operation (Bessler et al. 2003; Wei and Hsu 2008). Among them, cluster analysis (CA) has been found to have many applications in reservoir operation due to its advantage for identifying patterns from massive data (Ponnambalam et al. 2002; Suen 2011).

The main purpose of this study is to use a CA approach to identify operational scheme patterns for reservoir operation during a transition period. A case study of ten reservoirs in the Columbia River, United States is considered. Fifty-one different inflow hydrograph scenarios based on historical records from 1965 to 2015 are used. For each inflow scenario, the optimal operational scheme is derived using a genetic algorithm and then a clustering method is used to group and identify patterns of operational schemes.

The remainder of the paper is organized as follows. In the section on Optimization Model Setup, the study case—the Big-Ten reservoir system of the Federal Columbia River Power System (FCRPS)—is briefly introduced. The objective and the constraints of the optimization model during a transition period, as well as modelling of the reservoir system, are described. The inflow scenarios used for the optimization model are introduced and their statistics are briefly discussed. The CA Method section introduces the K-Spectral Centroid algorithm (K-SC), which is an efficient clustering technique recently developed (Yang and Leskovec 2011). By comparing it with the k-means method, which is widely used for CA, advantages of applying the K-SC to reservoir operational schemes are discussed. The index for determining the number of clusters in the K-SC is also described. In the Results and Discussions section, the optimal operational schemes and the identified patterns are presented. The practical benefits of the identified operational patterns are also discussed. Finally, the main results are summarized in the "Conclusions" section.

Optimization Model Setup

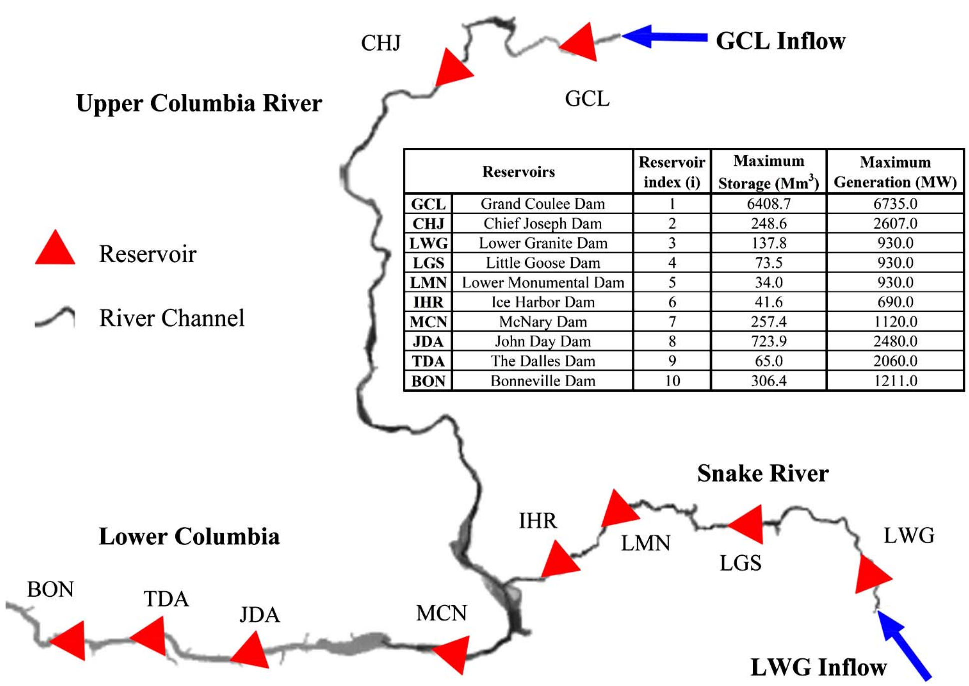
Study Case

The Big-Ten reservoir system—the ten large reservoirs of the FCRPS in the United States—is considered as a study case. Grand Coulee reservoir (GCL), located in the upper Columbia River, is of the storage type and dominates the system by accounting for nearly 80% of the storage. Other reservoirs are mostly run-of-river type, characterized by having relatively small storage. The river–reservoir network and some of the reservoir characteristics are presented in Fig. 1.

The Big-Ten reservoir system provides multiple benefits, e.g., power generation, flood control, and fish migration. However, some of the reservoirs have seasonal requirements and the operational objectives are only required during specific periods (Chen et al. 2016; Schwanenberg et al. 2014). From April to August, the reservoir system is operated to help migration of juvenile anadromous fish by maintaining specific operation pool levels (SOPs) and spilling a certain amount of flow (called fish flow). The reservoir system no longer has the fish flow nor the SOP requirements during September. Therefore the objectives of reservoir operation are shifted after August 31st (called the shift date).

Objectives

An hourly optimization model was used for finding the optimal operational schemes during the transition period. The time horizon for operating the reservoir system as short-term is normally two weeks (Chen et al. 2016). In order to investigate the overall performance of the reservoir system during the transition period, the optimization period in the study was set to two weeks with one week before and after the shift date. The decision variables in the model were the total outflows at each reservoir and at each time interval (i.e., hour). The nondominated sorting genetic algorithm [NSGA-II, (Deb et al. 2002)], one of the most widely used evolutionary algorithms, was selected as the optimization method.



Reservoirs		Reservoir index (i)	Maximum Storage (Mm^3)	Maximum Generation (MW)
GCL	Grand Coulee Dam	1	6408.7	6735.0
CHJ	Chief Joseph Dam	2	248.6	2607.0
LWG	Lower Granite Dam	3	137.8	930.0
LGS	Little Goose Dam	4	73.5	930.0
LMN	Lower Monumental Dam	5	34.0	930.0
IHR	Ice Harbor Dam	6	41.6	690.0
MCN	McNary Dam	7	257.4	1120.0
JDA	John Day Dam	8	723.9	2480.0
TDA	The Dalles Dam	9	65.0	2060.0
BON	Bonneville Dam	10	306.4	1211.0

Fig. 1. Sketch of the ten-reservoir system in the Columbia River

The population (i.e., candidate solutions) of the NSGA-II was set to 50 and the generation (i.e., iteration times) was set to a relatively large number (10,000) to ensure convergence.

An important objective of the reservoir system is to meet power load in the region, as well as to gain maximum revenue from power generation. Power generated that exceeds the load can be sold in the power market. On the other hand, energy needs to be purchased if a load deficit occurs. Net electricity is defined as hydropower generated minus the load. The revenue is quantified by multiplying the net electricity by real-time prices from the power market. The revenue objective is expressed as

$$\max \sum_{t=1}^{T}\left\{\left[\left(\sum_{i=1}^{N_r} \mathrm{PG}_t^i\right) - \mathrm{PL}_t\right] \times \mathrm{PR}_t\right\} \tag{1}$$

where PG = hydropower generated in the system (MWh); PL = power (MWh) that is needed for meeting the load (MW) in the region; PR = market price for hydropower (dollars/MWh); t = time, e.g., in hours; T = optimization period, i.e., 3,360 h (14 days); the index i indicates individual reservoirs in the system; and N_r = total number of reservoirs. The price of hydropower for the two-week period was predetermined by an economic model (Chen et al. 2014) and was treated as a deterministic parameter in this study. It should be noted that the formulation of the objective was mainly for demonstrating the effect of objective shifting on the reservoir operation. The operating agency—the Bonneville Power Administration—primarily aims to reduce the total operational cost rather than to make a profit, as is true of other nonprofit federal agencies. An alternative objective can be formulated for reducing the operational cost.

Other constraints of reservoir operation, such as maintaining the SOP and the fish flow, are described below.

Constraints

In order to assist juvenile salmon and steelhead species in surface passage past the dams, most of the reservoirs in the system are required to spill a certain amount of flow through nonturbine structures such as sluices or gates (Schwanenberg et al. 2014). These flow requirements are expressed as either a fixed flow rate or a percentage of the total outflow of a reservoir (NOAA Fisheries 2014), as follows:

$$Q_{s,i}^t = Q_{sr,i} \quad (\text{for } i = 5, 7, 8, 9) \tag{2}$$

$$Q_{s,i}^t = \frac{q_{sr,i}}{100} Q_{\mathrm{out},i}^t \quad (\text{for } i = 3, 4, 6, 10) \tag{3}$$

where Q_s = spill flow; Q_{sr} = fixed fish flow requirement; q_s = flow rate; and Q_{out} = total outflow from reservoir. According to the Biological Opinion issued by NOAA Fisheries (2014), the Grand Coulee ($i = 1$) and Chief Joseph ($i = 2$) reservoirs are not required to satisfy any fish flow requirement.

In concert with the purpose of assisting fish migration, the forebay elevations of reservoirs in the system are required to be kept within specific ranges, i.e., the SOP. The SOP requirements are expressed as follows:

$$\mathrm{SOP}_{\mathrm{lower},i} \le H_{r,i}^t \le \mathrm{SOP}_{\mathrm{upper},i} \tag{4}$$

where H_r = forebay elevation; and $\mathrm{SOP}_{\mathrm{lower}}$ and $\mathrm{SOP}_{\mathrm{upper}}$ = lower and upper boundaries for the SOP requirement, respectively.

Other operational constraints considered in the model include lower and upper limits on forebay elevations, turbine flows, and power outputs and ramping limits on reservoir outflows, forebay elevations, and tail water elevations. These constraints are considered to be common practice for reservoir operation and therefore are not listed for brevity.

The short-term operation of reservoirs is known to be greatly dependent on initial and ending conditions (Lund 1996) such as reservoir forebay elevations. Different initial and ending forebay elevation conditions often lead to various operational schemes that are too different to compare. To exclude the effects of initial and ending conditions, a fixed initial forebay elevation and a restriction on ending forebay elevation are considered. In the study, the historical forebay elevation of a normal year (1986) at the end of August 24th (the day before beginning date of optimization) was used as initial condition. On the other hand, the reservoir forebay elevations at the end of optimization period are expected to stay within a target range in order to fulfill their future obligations. These target ranges are commonly decided by middle-term or long-term optimization models (Lund 1996), which are not included in this study. Instead, the historical forebay elevation of 1986 at the end of September 7th (end date of optimization) was used as a reference ending condition. In order to avoid equality constraints, a small deviation was allowed for the forebay elevation at the end of the period to approximate the reference ending condition

$$H_{\mathrm{tar},i} - \Delta.D_{w,i} \le H_{r,i}^t \le H_{\mathrm{tar},i} + \Delta.D_{w,i} \tag{5}$$

where H_{tar} = reference forebay elevation at the end of the period; Δ = deviation percentage; and D_w = maximum water depth at reservoir i. The deviation percentage for Grand Coulee reservoir was set to 0.25% due to its large storage, corresponding to only 0.04 m in water depth. For the other reservoirs the deviation percentage was set to 10%.

Reservoir System Modelling

The reservoir storages at each time step were modelled through the following equation (i.e., continuity equation) in order to conserve the mass:

$$V_i^{t+1} - V_i^t = [(Q_{\mathrm{in},i}^t + Q_{\mathrm{in},i}^{t+1})/2 - (Q_{\mathrm{out},i}^t + Q_{\mathrm{out},i}^{t+1})/2]\Delta t \tag{6}$$

where V = reservoir storage; Q_{in} and Q_{out} = inflow to and outflow from reservoirs, respectively; and Δt = time step. The inflows are inputs to the model and the outflows are the decision variables. Water losses due to evaporation were not considered in the model due to the short time frame under consideration.

The forebay elevations were obtained from the established forebay-storage curves. The tail waters were obtained using a regression equation involving the reservoir outflow and the forebay elevation of the downstream reservoir. The turbine flow was modelled by relating the outflow with the fish flow requirement through the following procedures:

$$Q_{tb}^t = \begin{cases} Q_{tb\_\min}^t & \text{if } Q_{tb\_\min} \le Q_{\mathrm{out},i}^t < Q_{sr,i} + Q_{tb\_\min} \\ Q_{\mathrm{out},i}^t - Q_{sr,i} & \text{if } Q_{sr,i} + Q_{tb\_\min} \le Q_{\mathrm{out},i}^t < Q_{sr,i} + Q_{tb\_\max} \\ Q_{tb\_\max} & \text{if } Q_{sr,i} + Q_{tb\_\max} \le Q_{\mathrm{out},i}^t \\ Q_{\mathrm{out},i}^t & \text{else} \end{cases} \tag{7}$$

where Q_{tb} = turbine flow; and $Q_{tb\_\min}$ and $Q_{tb\_\max}$ = allowed minimum and maximum turbine flows, respectively.

The power generation was computed based on the turbine flow and the water head (a function of forebay elevation and tail water elevation) with project-aggregated coefficients

$$N_{d,i}^t = K_i(H_{r,i}^t - \mathrm{TW}_i^t) \times Q_{tb}^t \quad (8)$$

where N_d = power output; TW = tail water elevation; and K = coefficient to express the overall efficiency of turbine, which was aggregated as one value for each project (reservoir).

The flow propagation within the reservoir–river network was modelled using the Muskingum-Cunge routing method with calibrated coefficients. Most of the propagation times in the river between two reservoirs were 1–3 h except the river reach between CHJ reservoir and MCN reservoir, which had an average propagation time of 21 h.

Inflow Scenarios

There are two inflows to the reservoir system—the inflow from upstream of the GCL reservoir (GCL inflow), and the inflow from upstream of the LWG reservoir (LWG inflow). Other inflows, mostly side inflows from small tributaries, provide relatively small water volumes and hence are omitted in this study. Historical records of the two inflows with six-hour intervals from 1965 to 2015 (Fig. 2) were used as the multiple inflow scenarios. The period considered in the study is two weeks ranging from August 25th to September 7th. Because the optimization model was an hourly time step, the inflow data were linearly interpolated.

To better characterize the inflow scenarios, two indexes are proposed in the study. The total inflow volume of the two weeks period is certainly important for fulfilling objectives/constraints of reservoir operation. Inflow volumes of Week One and of Week Two are also important because shifting of reservoir operation involves temporal water usage competition. The first index is the total volume ratio (TVR), which is defined as the ratio of total inflow volume in a given year for the two-week period to that of a benchmark year for the same period. The year 1986, the normal water year, was used as the benchmark year. If the TVR value of one inflow is larger than one, the total inflow volume for the two-week period is larger than the benchmark year, implying relative water abundance, and vice versa. Another index, the weekly volume ratio (WVR), is defined as the ratio of inflow volume of Week One to that of Week Two. The index aims to represent temporal distribution of the inflow between Week One and Week Two.

The histograms of the indexes from the 51 inflow scenarios are shown in Fig. 3. For the GCL inflow, the observed mean and standard deviation of the TVR index were 1.00 and 0.18, respectively. It follows that, on average, the total inflow volume of the two weeks was equal to that of the benchmark year, although large variability was observed among the different scenarios. On the other hand, the observed mean and standard deviations of the WVR index for GCL inflow were 1.10 and 0.14, respectively. This suggests that, on average, the inflow volume of Week One was significantly larger than that of Week Two, showing an important variability of inflows during the two-week period. For LWG inflow, the observed means of the TVR and the WVR indexes were 0.99 and 1.0, respectively. These values suggest that the total inflow volume of the two weeks for the LWG inflow was (on average) a little less than that of the benchmark year, but the inflow volumes of the first and second week were (on average) nearly the same.

The study considers multiple inflow scenarios in order to identify general patterns of reservoir operation during the transition period. Each optimization for a given inflow scenario is called an experiment. Each experiment resulted in a set of outflows and associated forebay elevation trajectories. The trajectory of the forebay elevation is one of the primary means to represent operation of reservoirs and each of these trajectories is an operational scheme for reservoir operation. For a given reservoir system, the forebay elevation trajectory is influenced by the initial and ending conditions of the forebay elevation, as well as the inflow. Because the initial and ending forebay elevation of the optimization model were almost invariant in each experiment (described in the Constraints section), the study focuses solely on the relationships between forebay elevation trajectories and reservoir inflows.

CA Method

Cluster Analysis refers to the group of techniques that are designed to separate a set of objects or observations into different groups or *clusters* according to their similarities or proximities. Due to its generality, the problem has been extensively studied and a number of solutions and methodologies have been proposed in the literature, going back to Hartigan's Rule (e.g., Fuentes and Casella 2009; Hartigan 1975; Sugar and James 2003; Tibshirani et al. 2001). Among different techniques, the k-means clustering

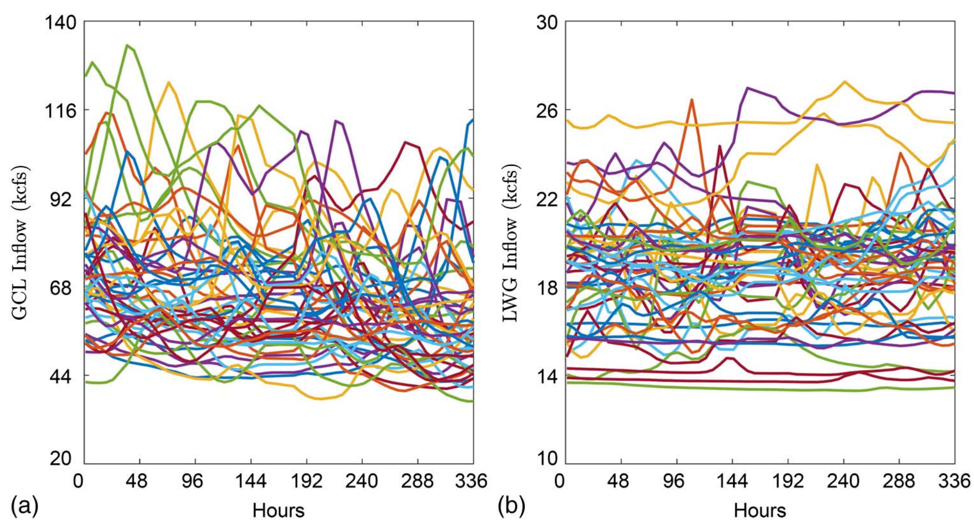


Fig. 2. Historical inflows for the period between August 25th and September 7th between 1965 and 2015 at (a) Grand Coulee (GCL); (b) Lower Granite (LWG)

J. Water Resour. Plann. Manage., 2017, 143(8): 04017028

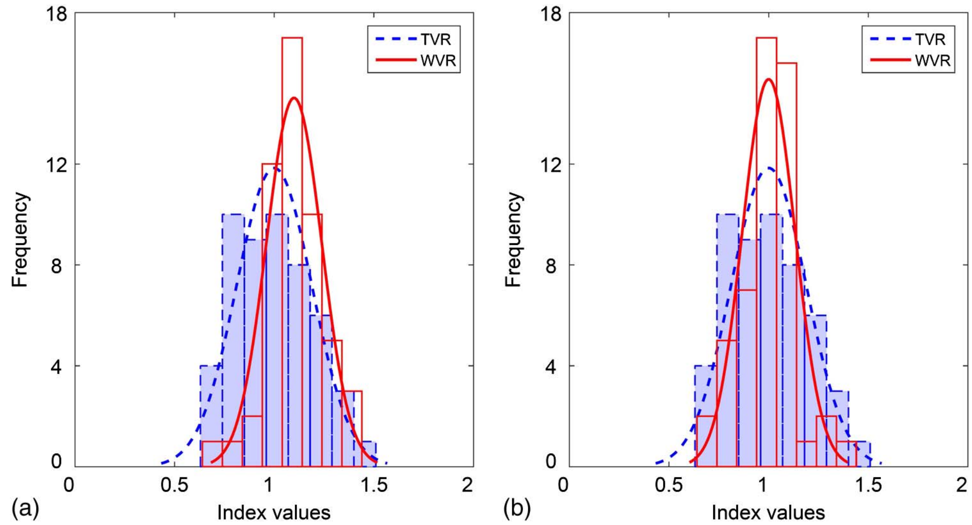


Fig. 3. Histograms of the TVR and the WVR indexes for (a) Grand Coulee (GCL) inflow; (b) Lower Granite (LWG) inflow

algorithm (Dhillon and Modha 2001; Hartigan and Wong 1979) has been a widely used method for CA. More recently, with the advances in genetics, image processing, and machine learning, new variations of the problem have become increasingly popular, including clustering and classification of curves, with obvious implications in pattern recognition, as discussed in (Zhang et al. 2015). The operational schemes of reservoir operation are time-series data (i.e., curves) which may have similar patterns even under different inflow conditions. Identifying patterns of operational schemes helps to gain a generalized understanding on reservoir operation during the transition period.

K-SC Algorithm

The *K*-SC algorithm (Yang and Leskovec 2011) is a recently developed method for finding distinct temporal patterns of time-series data. For a given N set of time series and the number of clusters K, the goal of the *K*-SC algorithm is to find an assignment of each time series and the centroid of each cluster so that a function of a distance metric is minimized. In a similar way to the *K*-means clustering algorithm, the *K*-SC algorithm iterates a two-step procedure: an assignment step and a refinement step. The *K*-SC algorithm starts with a random initialization of the cluster centers. In the assignment step, each data time series is assigned to the closest cluster, and in the refinement step the cluster centroids are updated. By alternating the two steps, the sum of the distances between the members of the same cluster is minimized, and the assignment of N sets of time series into K clusters is completed. The *MATLAB* code of the *K*-SC algorithm can be found at the Stanford Large Network Dataset Collection (SNAP) provided by Leskovec et al. (2014).

K-SC Algorithm versus *K*-Means Method

The two clustering methods were compared based on their applications to reservoir operational schemes. In this study, the operational schemes are time series, each representing specific actions or decisions over time. Similar shapes of these operational schemes suggest similar decisions on reservoir operation that can be grouped on the same cluster. Therefore it is essential to have a metric that can appropriately measure the shape similarity of two time series. For *k*-means, a simple distance metric, Euclidean, is adopted. The Euclidean metric measures the overall distance between two curves and tends to focus on only the global peaks of the curves. Under this metric, two time series may have a large distance due to a scale (in volume) or shifting (in position) effect even if their temporal shapes are similar. By contrast, the *K*-SC algorithm uses a distance metric $D(x_j, x_k)$ that is invariant to scaling and shifting (Yang and Leskovec 2011), defined as

$$D(x_j, x_k) = \min_{\lambda,q} \frac{\|x_j - \lambda.x_{k(q)}\|}{\|x_j\|} \tag{9}$$

where $\|\|$ = l_2 norm; λ = scaling coefficient; and q = shifting coefficient measured by q time units that are used to shift x_k. The metric works by finding the optimal value of the alignment q and the scaling coefficient λ for matching the shapes of the two time series.

To compare the *k*-means and *K*-SC methods, six artificial operational schemes were designed, some of them with similar shapes (Fig. 4). However, the volumes (i.e., scale) within those with similar shapes were different. For example, Schemes 1 and 2 had a difference of 20% in terms of scale. Some shapes were very close in terms of scale, such as Schemes 5 and 6 with only 3% difference. For reservoir operation, two operational schemes are defined as similar if their temporal shapes are similar despite their scales and shift. The rationale for the definition is discussed in relation to the clusters that are found by *k*-means and *K*-SC which are shown in Fig. 4.

The six artificial operational schemes should be easily classified into four clusters by direct observation. The members in each cluster are {①, ②}, {③, ④}, {⑤}, {⑥}. Scheme ① and scheme ② are different in scale but are very similar in terms of temporal shape. From the reservoir operator's perspective, these two share a similar operational pattern which decreases (either water level or outflow) along with time, hits a valley point, and then increases after that. In the same manner scheme ③ and scheme ④ are similar. Scheme ⑤ and scheme ⑥ should be considered different operational patterns even though they are very close in magnitude. As shown in Fig. 4, *k*-means clusters the six operational schemes as {①, ②}, {③}, {④, ⑥}, {⑤}. It turns out that *k*-means fails to recognize the relation between schemes 3 and 4, producing incorrect clusters. On the other hand, the *K*-SC method is able to find the desirable clusters.

Another advantage of *K*-SC is the robustness in presence of outliers. *K*-means is more sensitive to outliers because it considers the average of time series for a cluster center. *K*-SC instead scales

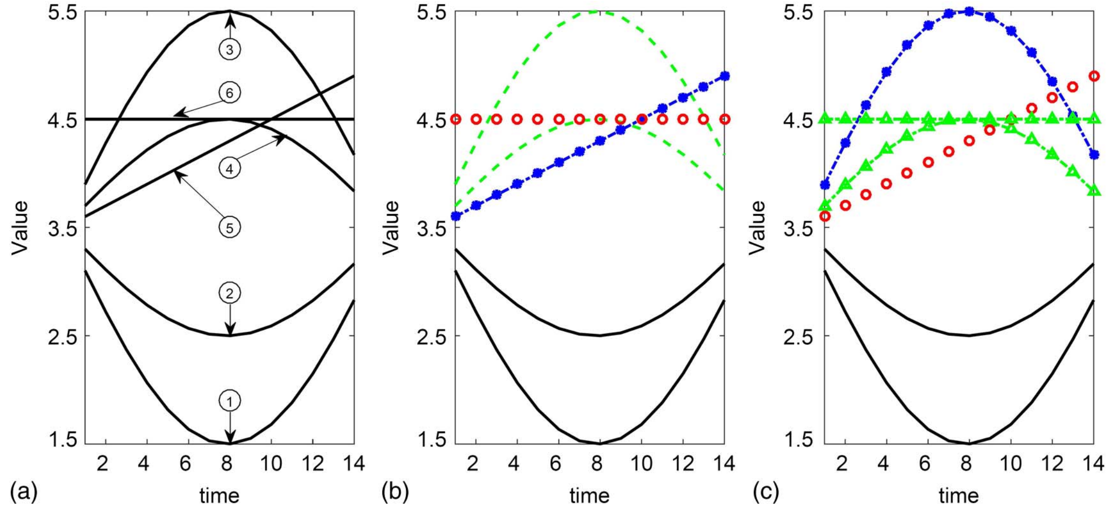


Fig. 4. Artificial operational schemes clustered by the k-means and the K-SC methods

each time series differently to find a cluster center, and therefore the influence of outliers is largely decreased.

Number of Clusters

Similar to most clustering methods, K-SC also needs to be specify the number of clusters in advance. The Silhouette (Kaufman and Rousseeuw 2009), an index to measure how well each object lies within its cluster, is used for determining the number of clusters. The Silhouette index for the ith point, S_i, is defined as $S_i = (x_i - y_i)/\max(x_i, y_i)$, where x_i is the average distance from the ith point to the other points in the same cluster, and y_i is the minimum average distance from the ith point to points in a different cluster, minimized over clusters. The index is within the range of $[-1,1]$, and the higher the value, the better the clustering. The case study measured how the Silhouette index (on average) varied with the number of clusters for the operational schemes of each reservoir and determined the number of the clusters with the highest Silhouette index. Fig. 5 shows the relations between the Silhouette index and different number of clusters for three reservoirs in the Big-Ten system as an example. From these relations, the optimal

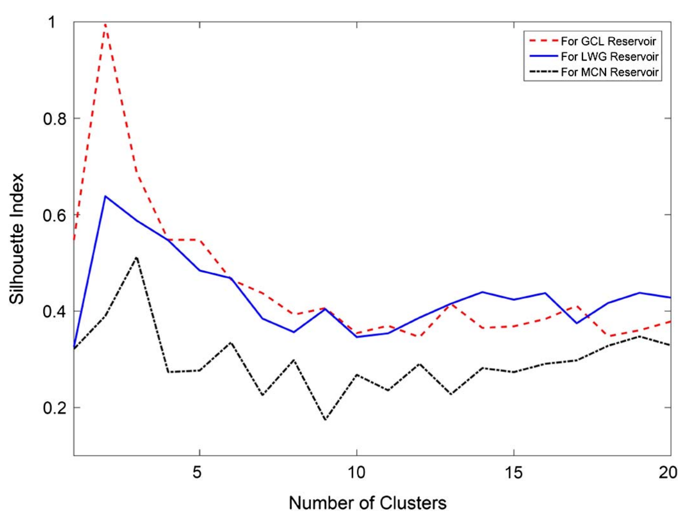


Fig. 5. Average Silhouette index for the GCL reservoir, the LWG reservoir, and the MCN reservoir

number of clusters for the GCL reservoir, the LWG reservoir, and the MCN reservoir can be determined as 2, 2, and 3, respectively.

Results and Discussion

Optimal Operational Schemes and Clusters

Among the ten reservoirs, the GCL and the LWG are the two most upstream reservoirs and their operation certainly influences the downstream reservoirs. The MCN, which is located immediately downstream of the confluence of the Snake River and the upper Columbia River (Fig. 1), also plays an important role in the system. Therefore these three reservoirs were selected to demonstrate the operation of the ten reservoirs. Most of the other reservoirs are run-of-river reservoirs, which pass inflow from the upstream reservoir. For simplicity, the operation of these reservoirs is not discussed herein, although all ten reservoirs were considered in the modeling. The optimal forebay elevation of the selected reservoirs under multiple inflow scenarios was obtained from the optimization model and is shown in Fig. 6.

The groups of the forebay elevation that are clustered by the K-SC algorithm are also shown in Fig. 6. The centroid of each group, which demonstrates a mean result of the corresponding cluster, is illustrated as well.

Two distinctive clusters or groups [Figs. 6(a and d)] were found in the collection of forebay elevations of the GCL reservoir. For Group 1, the forebay elevation gradually decreased (with oscillation) in Week One and increased in Week Two. In contrast, the forebay scenarios in Group 2 show that the forebay elevation increased (with oscillation) in Week One, achieving a maximum elevation at the end of Week One or at the beginning of Week Two. After the maximum forebay elevation was attained, the forebay elevation decreased until the end of Week Two.

The forebay elevations of the LWG reservoir were also clustered into two groups [Figs. 6(b and e)]. Even though the forebay elevations in Week One were all restrained in a certain range because of the SOP requirement, the trajectories have clear patterns. For Group 1, the forebay elevations initially decreased and then increased in Week One. The forebay elevations were maintained at a high level in Week Two. For Group 2, during the first week the forebay elevation was initially increased and then decreased until the end of Week Two, resulting in an opposite operational strategy to that of Group 1.

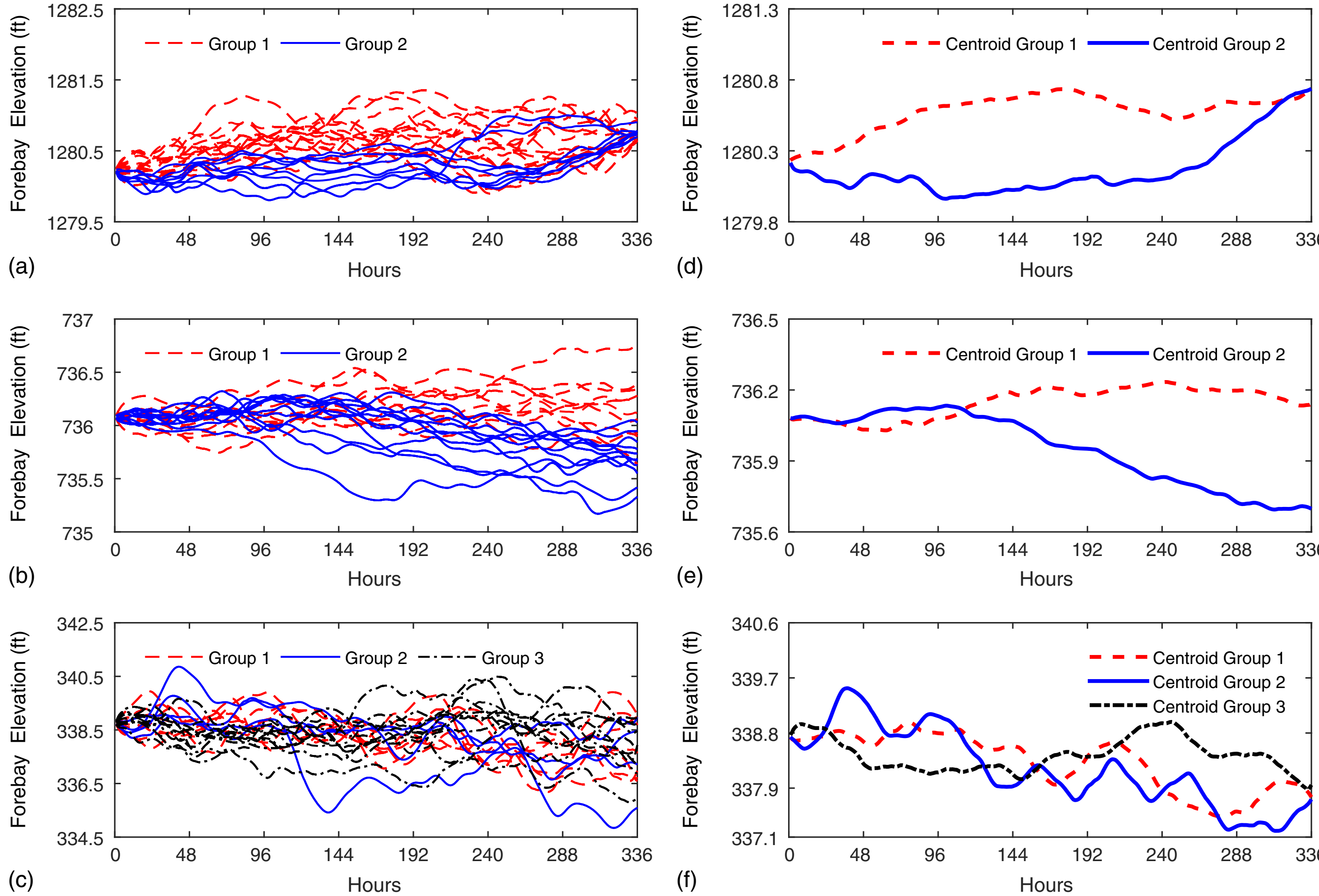


Fig. 6. Forebay elevation trajectories and corresponding clusters of the GCL reservoir, the LWG reservoir, and the MCN reservoir

Three clusters were identified for the forebay elevation of the MCN reservoir [Figs. 6(c and f)]. For Group 1, the forebay elevations mainly decreased in Week One and then increased in the first half of Week Two. After that, the forebay elevations decreased until the end of Week Two. Groups 2 and 3 were similar in terms of temporal shape for Week Two, during which the forebay elevations were mainly decreased (with oscillation). However, these two groups adopted different operational schemes for Week One. The forebay elevations of Group 2 rapidly increased and then decreased, whereas for Group 3 the forebay elevations maintained a constant level in the first half-week and then decreased.

Relations between Inflow Scenarios and Clusters

Based on the forebay elevation clusters of each reservoir, the TVR and WVR indexes of the inflows, specifically the GCL inflow and the LWG inflow, can be grouped accordingly. Note that each inflow has these two indexes. For instance, two groups were identified in the GCL forebay elevation [Figs. 6(a and d)], with 38 solutions in Groups 1 and 13 solutions in Group 2. Because each forebay elevation curve (one member in a group) is associated with one inflow scenario, the TVR indexes of all 38 inflow scenarios that are associated with Group 1 can be put in one group. The other 13 inflow scenarios that are associated with Group 2 were classified as another group, shown in Fig. 7(a). Similarly, the WVR index was classified into two groups as shown in Fig. 7(b). Correspondingly, the TVR and WVR indexes of the two inflows can also be classified based on the forebay elevation groups of the LWG reservoir and the MCN reservoir [shown in Figs. 7(c and d) and Figs. 7(e and f)].

The groups of the TVR index showed no interesting results. However, clearly separated clusters (or regions) were found for the WVR index. As can be seen in Fig. 7(b), the WVR index of the GCL inflow in Group 1 mostly adopted values higher than 1.0 and in Group 2 these values were mostly lower than 1.0. Interesting results were also found for the WVR index of the LWG inflow [Fig. 7(d)]. In Group 1 the WVR index of the LWG inflow adopted values lower than 1.0, and in Group 2 these values were mostly higher than 1.0. Three groups were found for the WVR index based on the three groups of the forebay elevations for the MCN reservoir [Fig. 7(f)]. Members of Group 1 are all in the upper-right region in which the WVR index of the GCL inflow and that of the LWG inflow were both higher than 1.0. Most of the scenarios of Group 2 are located in the lower-left region in which the WVR index of the GCL inflow was lower than 1.0 and the WVR index of the LWG inflow was mostly lower than 1.05. Most of the members in Group 3 are in the lower-right region in which the WVR index of the GCL inflow was higher than 1.0 and the WVR index of the LWG inflow was lower than 1.0.

Patterns of Reservoir Operation

By linking the definition of the WVR index with the two groups of the forebay elevation for the GCL reservoir, it is clear that different operational schemes need to be adopted when the volume of GCL

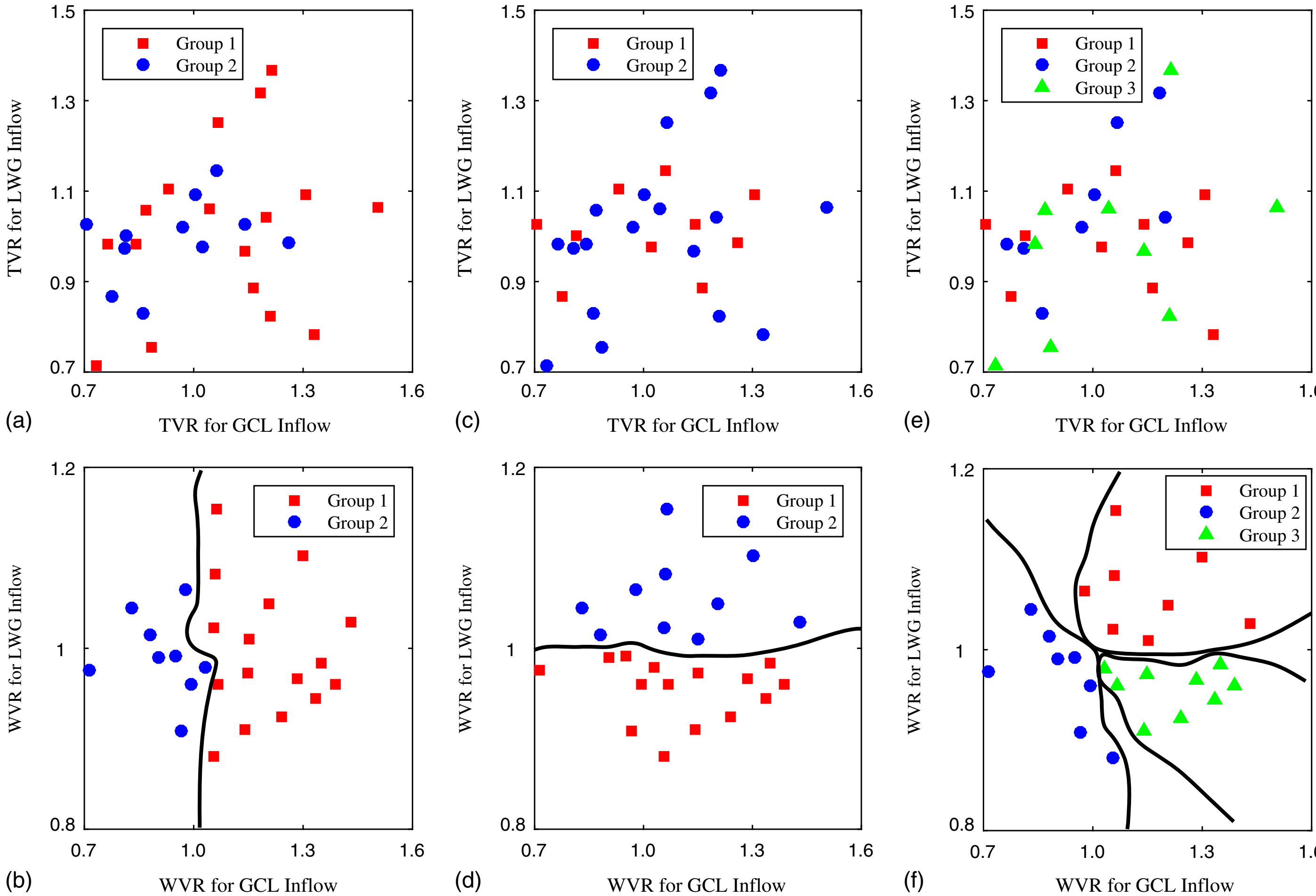


Fig. 7. Grouping of TVR and WVR indexes of inflows based on the forebay elevation groups of the GCL reservoir, the LWG reservoir and the MCN reservoir

inflow in Week One is smaller or greater than that of Week Two. When the volume of GCL inflow in Week One is smaller than in Week Two (i.e., WVR lower than 1.0), the operation should adopt the scheme of Group 1, which should use the storage of this reservoir (forebay elevation is decreased) during Week One to increase its outflow. This would decrease the power generation (and power revenue) in this reservoir because flow is released when water level is relatively low. However, the outflow increases in the GCL reservoir during Week One for meeting the fish flow requirement. This operational scheme tries to obtain a balanced solution between human interests and ecological benefits. In contrast, when the volume of GCL inflow in Week One is larger than in Week Two (i.e., WVR greater than one), the system should adopt the scheme of Group 2, which should store water in Week One (forebay elevation is increased) when inflow is relatively high during this week. The high inflow from upstream of GCL ensures that fish flow requirements for the four reservoirs on the lower Columbia River (MCN, JDA, TDA, and BON) are satisfied in Week One. Continuing to release water from the GCL reservoir would no longer be needed for fish flow because the other four reservoirs with fish flow requirement (LWG, LGS, LMN, and IHA) are on the Snake River. Therefore the optimal operation of the system under this situation is to store the excess water (after satisfying fish flow requirements) during Week One to produce more power during Week Two.

Another pattern was identified for the operation of the LWG reservoir. The association between forebay elevation groups with the WVR index shows that the LWG reservoir should adopt a different operational scheme when the volume of LWG inflow in Week One is smaller or greater than that of Week Two. When the volume of LWG inflow in Week One is smaller than in Week Two (i.e., WVR index lower than one), the LWG reservoir should release more water in Week One to fulfill the fish flow requirement. Thus its forebay elevation is decreased, as shown in Figs. 6(b and e) (Group 1). During Week Two, the forebay elevation maintains a high level for generating more power with the same outflow, which helps to compensate the power loss in Week One. On the other hand, the LWG reservoir would store some water in Week One when the volume of LWG inflow in Week One is larger than in Week Two (i.e., WVR index is higher than one), after the fish flow requirement is met. Higher forebay elevations can be obtained in this way, as shown in Figs. 6(b and e) (Group 2). This resulting high forebay elevation and the increased outflow in Week Two help to produce more power.

The operation of the MCN reservoir is influenced by the operation of reservoirs on the upper Columbia River (GCL and CHJ) and the operation of reservoirs on the Snake River (LWG and the other three reservoirs). Patterns for the MCN reservoir result from three combinations of the GCL reservoir operation and the LWG reservoir operation. For instance, the scheme of Group 3 needs to be adopted for the MCN reservoir when the GCL reservoir operation adopts its Group 2 scheme (when the WVR index of GCL inflow is higher than one) and the LWG reservoir operation adopts its Group 1 scheme (when WVR index of LWG inflow is lower than one). In this case, because the fish flow requirement for downstream reservoirs can be fulfilled by the operation of upstream reservoirs, the scheme would only pass the inflow from GCL and LWG in Week One. A relatively high forebay elevation can also be maintained in this way. More water would be released in Week Two for generating more power, thus maximizing revenue. The rationale of this scheme is to improve the power objective when fish flow requirements are met.

The identified patterns offer an insight into the various operational schemes during the transition period. Reservoir operators can benefit from these patterns because they could choose an operational scheme depending on a forecasted hydrological regime. Note that the accuracy of the forecast influences the selection of the patterns. These patterns can also be used as prior information for online optimization, which will diminish the effort for finding optimal solutions. The identified patterns provide a good initial starting point for the optimization model. Therefore the efficiency of optimization models can be improved with the assistance of the patterns.

Conclusions

Different patterns of operational schemes were identified for the ten-reservoir system of the FCRPS during a transition period. These patterns are found to be highly correlated with the index of WVR, which represents a ratio between volume of water in Week One and that of Week Two. In contrast, the patterns showed nearly no correlation with the index of TVR, which represents a ratio between total water volume for the two-week period of a specific year and that of a benchmark year. The comparison indicates that reservoir operation during objective shifting is more sensitive to temporal distribution of the inflow (i.e., the WVR index) than to the total volume of the inflow (i.e., the TVR index). Therefore, the WVR index is the main driver for adopting different operational schemes.

The identified patterns help to provide a general understanding for the operational schemes during the transition period, which can be used as prior knowledge for better online optimization performance. The method of the *K*-SC was found to outperform the widely used *k*-means for clustering reservoir operational schemes with more informative patterns.

Acknowledgments

This work was supported by the Bonneville Power Administration through projects TIP258 and TIP342. The first author would like to thank support from the National Natural Science Foundation of China (51425902, 51479188) and the Fundamental Research Funds for the Central Universities (CKSF2016009/SL). The authors also thank the anonymous reviewers for their insightful comments and suggestions.

References

Atiquzzaman, M., Liong, S.-Y., and Yu, X. (2006). "Alternative decision making in water distribution network with NSGA-II." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2006)132:2(122), 122–126.

Bessler, F. T., Savic, D. A., and Walters, G. A. (2003). "Water reservoir control with data mining." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2003)129:1(26), 26–34.

Cardwell, H., Jager, H. I., and Sale, M. J. (1996). "Designing instream flows to satisfy fish and human water needs." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1996)122:5(356), 356–363.

Chen, D., Leon, A. S., Gibson, N. L., and Hosseini, P. (2016). "Dimension reduction of decision variables for multireservoir operation: A spectral optimization model." *Water Resour. Res.*, 52(1), 36–51.

Chen, D., Leon, A. S., and Hosseini, P. (2014). "Optimizing short-term operation of a multireservoir system during transition of objectives and constraints." *World Environmental and Water Resources Congress 2014*, ASCE, Reston, VA, 1093–1105.

Chen, D., Li, R., Chen, Q., and Cai, D. (2015). "Deriving optimal daily reservoir operation scheme with consideration of downstream ecological hydrograph through a time-nested approach." *Water Resour. Manag.*, 29(9), 3371–3386.

Chen, Q., Chen, D., Li, R., Ma, J., and Blanckaert, K. (2013). "Adapting the operation of two cascaded reservoirs for ecological flow requirement of a de-watered river channel due to diversion-type hydropower stations." *Ecol. Modell.*, 252, 266–272.

Deb, K., Pratap, A., and Agarwal, S. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE Trans. Evolut. Comput.*, 6(2), 182–197.

Dhillon, I. S., and Modha, D. S. (2001). "Concept decompositions for large sparse text data using clustering." *Mach. Learn.*, 42(1–2), 143–175.

Eschenbach, E. A., Magee, T., Zagona, E., Goranflo, M., and Shane, R. (2001). "Goal Programming decision support system for multiobjective operation of reservoir systems." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2001)127:2(108), 108–120.

Fuentes, C., and Casella, G. (2009). "Testing for the existence of clusters." *Stat. Oper. Res. Trans.*, 33(2), 115–146.

Hartigan, J. A. (1975). *Clustering algorithms*, Wiley, New York.

Hartigan, J. A., and Wong, M. A. (1979). "Algorithm AS 136: A K-means clustering algorithm." *J. R. Stat. Soc. Ser. C*, 28(1), 100–108.

Homa, E. S., Vogel, R. M., Smith, M. P., Apse, C. D., Huber-Lee, A., and Sieber, J. (2005). "An optimization approach for balancing human and ecological flow needs." *Proc., EWRI World Water and Environmental Resources Congress*, ASCE, Anchorage, AK, 1–12.

Jager, H. I., and Smith, B. T. (2008). "Sustainable reservoir operation: Can we generate hydropower and preserve ecosystem values?" *River Res. Appl.*, 24(3), 340–352.

Kaufman, L., and Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*, Wiley, New York.

Leskovec, J., Krevl, A., and Datasets, S. N. A. P. (2014). "Stanford large network dataset collection." Stanford Univ., Stanford, CA.

Li, L., Xu, H., Chen, X., and Simonovic, S. P. (2010). "Streamflow forecast and reservoir operation performance assessment under climate change." *Water Resour. Manage.*, 24(1), 83–104.

Loucks, D. P. (1992). "Water resource systems models: Their role in planning." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1992)118:3(214), 214–223.

Lund, J. R. (1996). "Developing seasonal and long-term reservoir system operation plans using HEC-PRM." *HEC-RD-40*, Hydrologic Engineering Center, Davis, CA.

Maynard, C. M., and Lane, S. N. (2012). "Reservoir compensation releases: Impact on the macroinvertebrate community of the Derwent River, Northumberland, UK—A longitudinal study." *River Res. Appl.*, 28(6), 692–702.

Morrison, R. R., and Stone, M. C. (2015). "Investigating environmental flows for riparian vegetation recruitment using system dynamics modelling." *River Res. Appl.*, 31(4), 485–496.

NOAA Fisheries. (2014). "Endangered Species Act section 7(a)(2). Supplemental biological opinion. Consultation on remand for operation of the Columbia River power system." ⟨http://www.westcoast.fisheries.noaa.gov/publications/hydropower/fcrps/2014_supplemental_fcrps_biop_final.pdf⟩ (Jan. 3, 2016).

Olivares, M. A. (2008). "Optimal hydropower reservoir operation with environmental requirements." Univ. of California Davis, Davis, CA.

Ponnambalam, K., Karray, F., and Mousavi, S. (2002). "Optimization approaches for reservoir systems operation using computational intelligence tools." *Syst. Anal. Modell. Simul.*, 42(9), 1347–1360.

Prasad, T. D., and Park, N.-S. (2004). "Multiobjective genetic algorithms for design of water distribution networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2004)130:1(73), 73–82.

Raje, D., and Mujumdar, P. P. (2010). "Reservoir performance under uncertainty in hydrologic impacts of climate change." *Adv. Water Resour.*, 33(3), 312–326.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., and Kollat, J. B. (2013). "Evolutionary multiobjective optimization in water resources: The past, present, and future." *Adv. Water Resour.*, 51, 438–456.

Richter, B. D., and Richter, H. E. (2000). "Prescribing flood regimes to sustain riparian ecosystems along meandering rivers." *Conserv. Biol.*, 14(5), 1467–1478.

Schwanenberg, D., Xu, M., Ochterbeck, T., Allen, C., and Karimanzira, D. (2014). "Short-term management of hydropower assets of the Federal Columbia River Power System." *J. Appl. Water Eng. Res.*, 2(1), 25–32.

Smith, B. T., Jager, H. I., and Sale, M. J. (2007). "Optimization of hydropower resources: Current practices and opportunities for improvement." *ORNL/TM-2006/91*, Oak Ridge National Laboratory, Oak Ridge, TN.

Suen, J.-P. (2011). "Determining the ecological flow regime for existing reservoir operation." *Water Resour. Manage.*, 25(3), 817–835.

Suen, J.-P., and Wang, R.-H. (2010). "Optimal reservoir operation considering downstream water quality and environmental flow needs." *World Environmental and Water Resources Congress 2010: Challenges of Change*, ASCE, Reston, VA, 2592–2601.

Sugar, C. A., and James, G. M. (2003). "Finding the number of clusters in a dataset: An information-theoretic approach." *J. Am. Stat. Assoc.*, 98(463), 750–763.

Tibshirani, R., Walther, G., and Hastie, T. (2001). "Estimating the number of clusters in a data set via the gap statistic." *J. R. Stat. Soc. Ser. C*, 63(2), 411–423.

Wei, C.-C., and Hsu, N.-S. (2008). "Derived operating rules for a reservoir operation system: Comparison of decision trees, neural decision trees and fuzzy decision trees." *Water Resour. Res.*, 44(2), 1–11.

Wurbs, R. A. (1991). "Optimization of multiple-purpose reservoir systems operations: A review of modeling and analysis approaches." *Research Documents*, U.S. Army Corps of Engineers, Hydrologic Engineering Center, Washington, DC.

Yandamuri, S. R., Srinivasan, K., and Murty Bhallamudi, S. (2006). "Multiobjective optimal waste load allocation models for rivers using nondominated sorting genetic algorithm-II." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2006)132:3(133), 133–143.

Yang, J., and Leskovec, J. (2011). "Patterns of temporal variation in online media." *Proc., 4th ACM Int. Conf. on Web Search and Data Mining*, ACM, New York, 177–186.

Yin, X. A., and Yang, Z. F. (2011). "Development of a coupled reservoir operation and water diversion model: Balancing human and environmental flow requirements." *Ecol. Modell.*, 222(2), 224–231.

Zhang, Z., Pati, D., and Srivastava, A. (2015). "Bayesian clustering of shapes of curves." *J. Stat. Plann. Infer.*, 166, 171–186.