Open Access Journal

Journal of Power Technologies 94 (3) (2014) 165–175

journal homepage:papers.itc.pw.edu.pl

Determination of Pumped Storage Capacity Combining the Entropy Weighting Method and Principal Component Analysis

Zhan'an Zhang[*,a], Xingguo Cai[a], Destung Parkun[b]

[a] *School of Electrical Engineering and Automation*
Harbin Institute of Technology, Harbin 150001, China
[b] *Pohang University of Science and Technology*
Pohang 790-784, Republic of Korea

Abstract

The aim of the study is to evaluate methods for determining appropriate pumped storage capacities. This is especially relevant, since pumped storage units are, today, viewed as some of the best means of storing large amounts intermittently-produced power in order to meet peak demands on power supply grids. The determination of appropriate pumped storage capacity is a problem of integrated decision-making. The entropy weighting method and principal component analysis are combined to determine the optimum pumped storage capacity, taking into account several representative indices, whilst using the syntropy method to standardize the data indicators. The entropy weighting method is used to determine the weighting of the indicators, while principal component analysis offers reduction of the dimensions. The optimal solution is then determined by ranking the evaluation values for each design. This method can avoid subjectivity in the weighting assignment and simplifies the calculations to an evaluation problem which contains multiple evaluation indices. Using the principle of energy-saving scheduling, the peak-shaving method is applied to the dispatching over a typical daily load in order to verify the rationality of the calculated pumped storage capacity. The example analysis, here, shows that it is reasonable to determine the optimum pumped storage capacity by using this combination of the entropy weighting method and principal component analysis.

Keywords: Pumped storage, Capacity determination, Integrated decision-making, Entropy weighting method, Principal component analysis

1. Introduction

For power grids, especially those based mainly on thermally generated power, it is difficult to meet both the peak load and the changing gap between peak load and valley load. The problem is exacerbated by massive intermittent demands for extra energy from users connected to the grid [1]. A pumped storage (P-S) unit has the benefits of a high ramp-up rate, with quick starting and stopping, whilst it has low running costs, cleaning requirements, frequency and phase modulation and minimal black start power demands. Thus P-S units are considered to represent the most appropriate technology for meeting demands for large amounts of intermittent power and meeting the peak loads in power grids [2–4]. However, determining a reasonable storage capacity is the key problem.

[*] Corresponding author

Email address: zza.163.com@126.com (Zhan'an Zhang[*,])

In the past, the optimal P-S capacity has always been determined by individual factors, such as unit operating cost, loss of load probability, peak-regulation proportion, wind and nuclear power capacity, generator utilization hours, and other indices. Therefore, it was inevitable that mismatch problems would appear. To determine appropriate P-S capacity is an issue of comprehensive decision-making and integration, and at present the main methods are the analytic hierarchy process (AHP), fuzzy comprehensive evaluation [5, 6], the entropy weighting method, principal component analysis (PCA) and combinations thereof. Currently, the AHP and fuzzy methods are rarely used alone for evaluation because the index weightings represent contingencies which are very subjective, even when judged by experts. Hence, various combination algorithms have been widely applied to combination evaluations, such as the AHP and fuzzy methods [7, 8], AHP and other algorithms [9, 10] and fuzzy and other algorithms [11, 12]. In order to avoid subjectivity in the process of determining the weightings, the entropy method [13, 14] and PCA algorithm [15, 16] are used extensively.

There are also some evaluation research approaches which combine the entropy method and the PCA algorithm [17, 18]. However, the applicability of this and their rationality are short of reasonable explanation, and the methods for standardization of the data are not identical, so this has a significant impact on the resulting comprehensive evaluation. There is currently no application for combining the entropy weighting method and PCA in studies for determining optimum P-S capacity.

The key issues are the choice of evaluation indices and determination of the index weighting in the overall evaluation process. In this paper, the data indicators are standardized using the syntropy method, which makes the evaluation results more credible. Using the entropy weighting method to determine the weighting of indicators is more objective than PCA, although the latter can also take the variance contribution rate into account for its weighting. Using PCA to reduce the dimensions of the indicators can simplify the calculations to just an evaluation problem containing multiple evaluation indices.

Finally, we insert the calculated optimal P-S capacity into the operation of an example system, to assess the accuracy of this value derived from the combination of the entropy weighting method and principal component analysis. Since the peak-shaving method is particularly suitable for scheduling grids dominated by thermal power, it is used to represent the dispatching of a typical daily load to demonstrate the correctness and feasibility of the predictions of our method for energy-saving scheduling.

The conclusion validates our method as a means of avoiding the subjectivity of weighting assignment and for simplifying the calculations required, as well as for providing an effective method for multi-index comprehensive evaluations.

2. Methods of Determining P-S capacity

The benefits of P-S stations can be divided into the static and dynamic benefits. Static benefits include the capacity benefit, peak-valley filling efficiency and coal saving benefits; the dynamic efficiency benefits mainly include frequency-modulation, phase-modulation, back-up, load tracking, black start, and improved system reliability. The following section explains some of the commonly used evaluation indices:

2.1. Capacity benefit

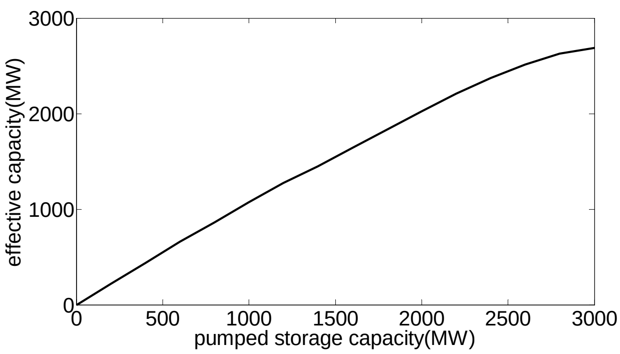

Figure 1: Curve for P-S capacity vs. the effective capacity

A P-S station can meet peak loads just as a conventional hydropower station does, thereby effectively obviating the need for other sources, such as thermal power generation, and thus reducing the investment needed for these other power sources. The resulting economic benefits are called capacity benefits. The

P-S effective capacity curve of a typical system is shown in Fig. 1 [19].

The effective capacity curve of the example system shows that, for P-S capacity it is not simply a case of the bigger the better, as excess amounts of P-S capacity can actually reduce the efficiency.

2.2. *Considering the capacity of wind and nuclear power*

Because of the uncertainty of wind power output, the dispatchers have to restrict wind power usage or to abandon this source altogether when anti-peaking appears. Using wind power capacity to maintain appropriate P-S capacity means that wind energy can be effectively stored and released when needed, therefore reducing the amount of abandoned wind power and increasing wind energy utilization. If appropriate P-S capacity is installed in the power system, the penetration of wind power in the grid can therefore be improved.

Because nuclear power has an essentially non-adjustable output, such units can run in tandem with P-S units of appropriate capacity to enable regulation of the overall power output, and this can greatly improve the efficiency and safety of the associated nuclear power plants.

2.3. *Peak-regulation proportion*

The system peak-regulation proportion is the ratio between the capacity of the adjustable units and the total system capacity. By calculating the deficit of peaking capacity in meeting the peak load in the grid, we can determine the required installed capacity of the necessary P-S units. This assumes that all the P-S capacity is used to its fullest extent and that the P-S unit can generate at its rated output,

$$R_G = \frac{P_{\max} - P_{\min} - 2P_{P-S}}{P_N} \cdot 100\% \tag{1}$$

where R_G is the system peak-regulation proportion; P_{max} is the maximum output of all the units and P_{min} is the minimum; P_N is the combined rated output of all the units; P_{P-S} is the required P-S capacity. Only when the difference of peak load accounting for the largest load ratio is less than or equal to the system peak-regulation proportion is the system's peak contradiction resolved [20].

2.4. *The amount of coal consumption and coal saved*

The coal-saving benefit of P-S refers to the difference in coal consumption for the case under review, with and without the inclusion of a P-S station. Although operation of the P-S units will result in additional cost because of the power used, the overall coal consumption will drop as a result of the improved operating mode of the thermal power station, so the actual coal- saving benefit trade-off depends on both of these factors.

The P-S stations use power from those coal stations with the minimum coal consumption in order to pump water, but are then used to generate electricity in place of those thermal power units with the highest coal consumption, so this is the origin of the benefits of P-S in this alternative.

The annual coal-saving formula can be expressed as

$$B = \Delta b \cdot W \tag{2}$$

$$\Delta b = b_f - \eta b_{ch} \tag{3}$$

where B is the annual coal-saving amount of the P-S plants (t), W is the annual generation of electricity (kWh), Δb is the coal-saving amount per unit of electricity (g/kWh), b_f is the equivalent coal consumption of the units that were replaced by the P-S units (g/kWh), η is the ratio of pumping to generation for the P-S units, b_{ch} is the coal consumption of the unit providing the electricity for the pumping (g/kWh) [21].

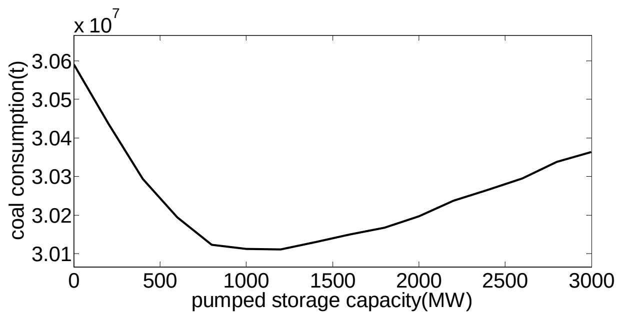

Figure 2: P-S capacity vs. the coal consumption curve

The coal consumption curve and coal-saving curve of a typical system are shown in Figures 2 and 3, respectively [19]. It can be seen from the figures that

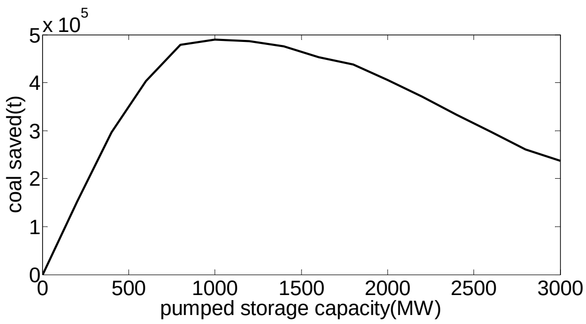

Figure 3: P-S capacity vs. the amount of coal saved

the P-S units should be chosen with an appropriate capacity.

2.5. Environmental benefits

The environmental benefits of P-S power plants are embodied in the transfer of the power production options to those stations with more efficient coal usage, and in the resultant reduction in thermal power generation emissions. In order to calculate the environmental benefits of P-S power plants, the effects of the reduction in pollutants can be unified and their value calculated according to the formula as follows:

$$C_e = \sum_{i=1}^{k} \frac{E_{ri}}{N_i} \tag{4}$$

where C_e is the total equivalent value of pollutant emission reductions, E_{ri} is the i-th pollutant emission reduction, N_i is the equivalent value of the i-th pollutants. Here the pollutant emissions can be CO, NO_X, SO_2 and smoke [21].

2.6. Spare capacity benefits

Spare capacity includes emergency and additional-load capacity, where the reserves provided by P-S power plants are the principal dynamic benefits, since they are able to supply emergency power to the grid if part of the system has failed or is under greater-than-expected load, thereby reducing any outages by increasing the available supply to meet the load. The spare capacity benefits in China are hard to assess because of the current immaturity of its electricity market mechanism.

2.7. Frequency and phase modulation

By starting up quickly and providing an adjustable output across a wide range, P-S units can adapt to sudden load changes and maintain the frequency stability of the grid. Therefore P-S stations, by bearing the load regulation and meeting the changing daily load requirements, can provide load tracking benefits.

A P-S station can supply or absorb reactive power at run-time, which can reduce the need for reactive power compensation equipment, maintain system voltage level stability and therefore result in phase modulation benefits.

2.8. LOLP

LOLP (Loss of Load Probability) is the probability of insufficient power availability. It means that the power generation system margin has a less than zero probability in the system. The *LOLP* can be obtained from the sum of probabilities that the outage capacity is greater than the power generation margin. The smaller the value of *LOLP*, the higher the power system reliability,

$$LOLP = P(X \geq C_s - L) \tag{5}$$

where X refers to system outage capacity, C_s is system installed capacity, L is the daily maximum load.

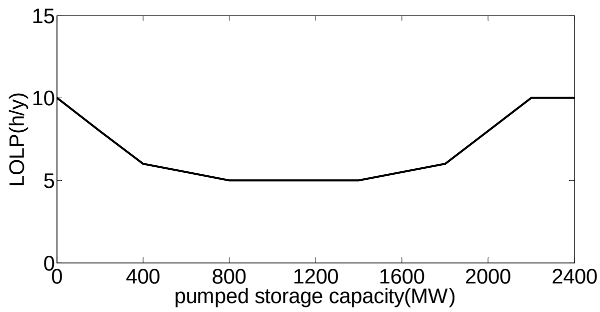

Figure 4: *LOLP* curve

A curve showing the relationship between the installed P-S capacity and the *LOLP* of a typical system is shown as Fig. 4. As the graph indicates, the *LOLP* value first decreases and then rises again with an increase in P-S capacity. The P-S capacity should therefore be chosen appropriately. Overcapacity does not reduce the probability of power outage; on the contrary, it reduces the supply reliability of the power system.

2.9. Black-start benefits

System recovery after blackouts over large areas is known as black start. A P-S station can start without an external power source and can therefore supply power to drive other units, so that the power system can be restored in the shortest time.

P-S units characteristically require few auxiliaries, need only a small amount of self-power and are quick to start. Therefore, they represent ideal and convenient black-start power supplies. Whilst it is obviously beneficial to the reliable operation of the grid to use P-S units as the black-start power source, it is difficult to quantify the benefits.

At present, with the current state of China's electricity market structure and its electricity pricing system, the auxiliary services are not perfect, and therefore the dynamic benefits of P-S stations are hard to calculate accurately.

3. Introduction to entropy and PCA

3.1. Entropy

Derived from thermodynamics, the concept of entropy can be applied to statistical mechanics, information theory and other disciplines. According to information theory, information is a measure of the degree of order, while entropy is a measure of disorder. Entropy is essentially a quantitative measure of uncertainty, which can be used to describe the use of information to assess risk more accurately and easily. Considering a probability test with n results, if we set these results with a discrete probability p_i ($i = 1 \sim n$), then the entropy is defined as

$$H(p_i) = -\sum_{i=1}^{n} p_i \ln p_i \tag{6}$$

where $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$.

The basic properties of entropy are:

1. Symmetry, the entropy value of p_i ($i = 1 \sim n$) has nothing to do with its position in the sequence.
2. Non-negativity, i.e., $H(p_i) \geq 0$ ($i = 1 \sim n$).
3. Additivity, the system's entropy is equal to the sum of all the entropy components.
4. Extremum property, the maximum entropy $H_{max}(p_i) = \ln n$ when $p_i = 1/n$ ($i = 1 \sim n$).
5. Concavity and convexity, $H(p_i)$ is a symmetrical concave function of all the variables.

To make the $\ln p_i$ meaningful, we generally assume that $p_i \ln p_i = 0$ when $p_i = 0$.

The entropy value reflects the extent of disorder, which can be used to the impact of the amount of data on the calculations. The greater the entropy value, the less the contribution of the information. Conversely, the smaller the entropy, the greater the contribution of the information [22].

3.2. Principal component analysis (PCA)

PCA is a commonly-used multiple statistical method for dimension reduction. The main idea of PCA is to use a number of unrelated new variables, instead of the more usual ones, and then to combine them in a linear manner to form a single variable. The new variable is called the component, and it retains the information from the original variables as far as possible. From the angle of statistical analysis, the contained information can be expressed by variance, and the bigger the variance, the more meaningful the information.

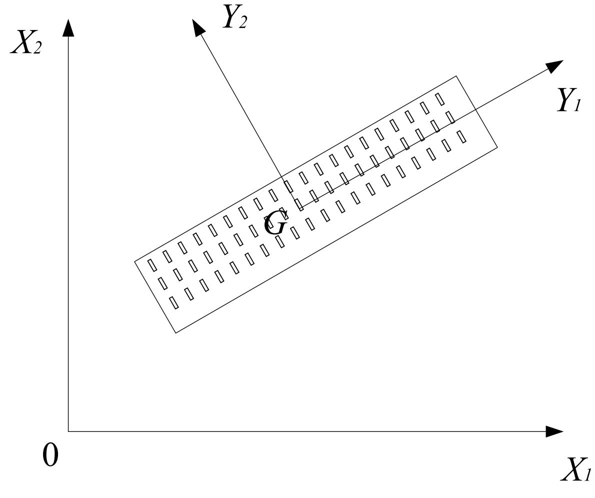

Figure 5: Dimension-reducing principle schematic diagram of PCA

The value of PCA is the translation and rotation of coordinates. Suppose that there is a two-dimensional data table. The data is distributed in a long series, and G is the centre of gravity, as shown in Fig. 5. Let us move the origin of the coordinate system to G, and then rotate the coordinate system for transformation. Taking the direction of maximum data variation as the axis and ignoring small variations, we can get an orthogonal coordinate system Y_1GY_2. Then any two-dimensional space problems can be reduced to one-dimensional analysis. The same thinking promotes the establishment of high-dimensional space.

Suppose the index data are composed by n plans and p evaluation indexes $X=(x_{ij})_{n\times p}$, where x_{ij} is the value of the i-th plan to the j-th index.To standardize the index data with (7) ,

$$z_{ij} = \frac{x_{ij} - \overline{x}_j}{\overline{s}_j} \tag{7}$$

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \overline{s}_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2} \tag{8}$$

where $\bar{x}_j$ and $\bar{s}_j$ are the mean and mean square error of the j-th index respectively. The standardized index data $Z = \left(z_{ij}\right)_{n\times p} = \left(Z_1, Z_2, ..., Z_p\right)$ satisfied with

$$E\left(Z_j\right) = 0,\ D\left(Z_j\right) = 1\ (j = 1, 2, \ldots, p) \tag{9}$$

The means are 0 and mean square errors are 1 for all p variables, which is called the Z-score method. We can establish a standardized data correlation coefficient matrix $R = \left(r_{ij}\right)_{p\times p}$, where r_{ij} reflects the degree of correlation between Z_i and Z_j.

$$r_{ij} = \frac{cov(Z_i, Z_j)}{\sqrt{D(Z_i)}\sqrt{D(Z_j)}} \tag{10}$$

where $cov(Z_i, Z_j)$ is the covariance between Z_i and Z_j. Since as established by (9), the correlation coefficient matrix and the covariance matrix are equal, R, the correlation coefficient, can be obtained as

$$R = \frac{1}{n-1} Z^T Z \tag{11}$$

Solving the eigenvalue of R, if there are q eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q \geq 0$, the corresponding orthogonal eigenvectors to eigenvalues can be expressed as $A = \left(a_1, a_2, \ldots, a_q\right)$, then q principal components are

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{p1} \\ a_{12} & a_{22} & \cdots & a_{p2} \\ \vdots & \vdots & & \vdots \\ a_{1q} & a_{2q} & \cdots & a_{pq} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix} \tag{12}$$

Turn Equation (12) into matrix form as

$$Y = A^T Z \tag{13}$$

The principal component has the following properties:

$$cov(y_i, y_j) = \begin{cases} 0 & i \neq j \\ \lambda_i & i = j \end{cases} \tag{14}$$

Equation (14) shows that the principal components are unrelated to each other, and the eigenvalue of y_i is the variance of the principal component. The contribution rate of principal component variance to the total variance is as

$$\rho_i = \lambda_i / \sum_{j=1}^{n} \lambda_j \tag{15}$$

The contribution rate ρ_i reflects the i-th principal component and includes some of the original variable information. Therefore, the first principal component has the greatest contribution to the variance, while the contributions of the subsequent ones are progressively smaller. Defining the former m principal components for the cumulated variance contribution rate as

$$\rho_{sum} = \sum_{i=1}^{m} \lambda_i / \sum_{j=1}^{n} \lambda_j \tag{16}$$

In order to achieve the objective of dimension reduction, if the cumulated variance contribution rate of the former m principal components is more than 85%, then we can take the former m principal components for the original p indexes, and the m principal components constitute a comprehensive evaluation function as

$$f = \rho_1 y_1 + \rho_2 y_2 + \cdots + \rho_m y_m \tag{17}$$

where the principal component weight is its contribution rate to the variance.

The PCA is based on the evaluation of dimension reduction, meaning that the evaluation calculation can be simplified, thereby also greatly reducing the workload of the decision makers [23].

4. Comprehensive evaluation model based on entropy and PCA

1. Taking m evaluation indicators to analyze n schemes, we can get the evaluation matrix as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \tag{18}$$

2. The standardization method used for the data has a significant impact on the comprehensive evaluation result. The Z-score method is widely used, principally in order to increase the difference between the data, and to facilitate the evaluation, although the principal component has to be extracted from the correlation coefficient matrix. In this study, the syntropy method is used to standardize the data indicators. According to the mechanism analysis and our professional knowledge, we can define positive and negative effect indices. For the positive effect index a_{ij}^+, the bigger the value, the better, while for the negative effect index a_{ij}^-, the smaller, the better, therefore normalizing the two types of indicator. Define the ratio of a_{ij} and a_j^* for the close degree of a_{ij} on a_j^*,

$$D_{ij} = \begin{cases} a_{ij}^+/a_j^* & a_j^* = \max\{a_{ij}^+\} \quad for \quad a_{ij}^+ \\ a_j^*/a_{ij}^- & a_j^* = \min\{a_{ij}^-\} \quad for \quad a_{ij}^- \end{cases} \tag{19}$$

Setting $D = \sum_{i=1}^n \sum_{j=1}^m D_{ij}$, standardizing D_{ij} with D, then $d_{ij} = D_{ij}/D$, $d_j = \sum_{i=1}^n d_{ij}$.

3. Obtain the standardized evaluation matrix of indicators

$$d_{ij} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nm} \end{bmatrix} \tag{20}$$

4. Calculate H_j , the entropy of indicator j,

$$H_j = -\frac{1}{\ln n} \sum_{i=1}^{n} \frac{d_{ij}}{d_j} \ln \frac{d_{ij}}{d_j} \quad j = 1, ..., m \tag{21}$$

5. For the unbiased index weighting formula,

$$w_j = \frac{1 - H_j}{m - \sum_{j=1}^m H_j} \tag{22}$$

where: $\sum_{j=1}^m w_j = 1$, $(j = 1, ..., m)$. The bigger the weighting and the greater the influence of the indicators, the greater their contribution to the scheme evaluation. Although the PCA method can also take the variance contribution rate as the weighting, using the entropy weighting method to determine the weighting of the indicators is more objective than with PCA.

6. If PCA is carried out on d_{ij}, and we calculate the correlation coefficient matrix R, then we can calculate the corresponding eigenvalues λ,

$$\lambda = diag(\lambda_n) \quad n = 1, ..., p \tag{23}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 \tag{24}$$

and calculate the variance contribution rate ρ_i, if $\rho_{sum} > 85\%$, then take the former m principal components for the original p indexes.

7. Calculate the evaluation value of n schemes with the entropy method

$$F_i = \sum_{j=1}^{m} w_j d_{i,j}, \quad i = 1, \ldots, n \tag{25}$$

Sort the values of F_i, and the maximum value of F_i is the optimal solution [17].

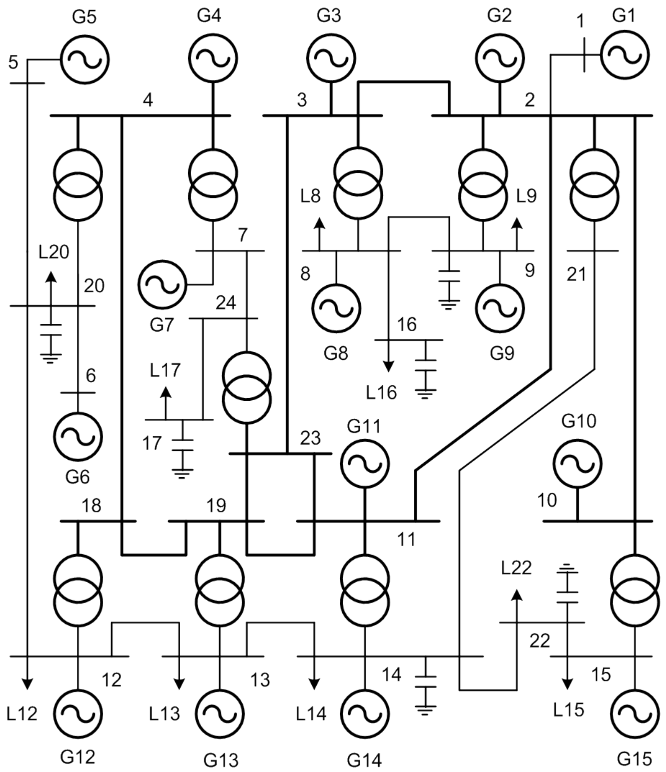

Figure 6: Power grid system diagram

5. Example analysis

5.1. The determination of P-S capacity

A simplified power grid system, as sketched in Fig. 6, can be used to illustrate the proposed method and dispatch model. The thermal, hydro and wind power capacities are 15000 MW, 1000 MW and 3000 MW, respectively. The planned P-S units to put into system are 400 MW, 800 MW, 1200 MW, 1600 MW and 2000 MW. The operation mode of each P-S unit will use its full capacity both for pumping and generation.

In Fig. 6, G5, G8 and G9 are wind farms, of which the maximum outputs of the wind generators are 1.5 MW. G1, G7 and G15 are conventional hydro stations and G10 is P-S station with a capacity of 1200 MW, the parameters of which are shown in Table 1 where, respectively, V_{max} and V_{min} are the maximum and minimum reservoir volumes, Q_{max} and Q_{min} represent the maximum and minimum turbine flows, while Y_{max} and Y_{min} refer to the highest and lowest heads for the stations. P_{max} is the maximum output and η is the efficiency of the unit.

Table 1: Parameters of hydro and P-S stations

Parameters	Hydro1 (G1)	Hydro2 (G15)	Hydro3 (G7)	P-S (G10)
V_{max}, 10^8m^3	17	30	10	0.12
V_{min}, 10^8m^3	7	15	8	0.1
P_{max}, MW	100	600	300	1200
Q_{max}, m^3/s	240	1200	1200	350
Q_{min}, m^3/s	40	50	40	0
Y_{max}, m	60	60	30	445
Y_{min}, m	50	50	25	405
η	0.9	0.95	0.95	0.75

Table 2: Parameters of thermal power plants

P_N, MW	Num.	Ramp rate, MW/min	P_{min}, %	C_C, g/kWh	S_S, t/time
600	8	12	55	310	150
300	16	6	57	330	60
200	20	4	60	350	35
100	14	2	65	390	17

The others are thermal power plants, the main parameters of which are listed in Table 2, where P_N represents the rated output of a unit, P_{min} is the ratio of the minimum output to the rated output of the generator, C_C is the average coal consumption of the unit, while S_S refers to the coal consumption per single start/stop for the unit.

As the dynamic benefits of a P-S station are hard to calculate accurately, we have selected a few representative indices in order to evaluate the P-S capacity, such as the effective coefficient of its substitution for thermal power capacity, the system coal consumption, the amount of coal saved, the peak-regulation proportion and the *LOLP*, which are shown in Table 3. Where C_T is the coefficient of the equivalent alternative thermal capacity, S_C is the coal saved, F_C refers to the system annual coal consumption and R_G

Table 3: The value of each index under various plans

Plan	P-S, MW	C_T	F_C, 10^4t	S_C, 10^4t	R_G, %	$LOLP$, h/y
1	400	1.1	3029.38	29.62	34	6
2	800	1.0775	3012.35	47.9	38	5
3	1200	1.0617	3011.02	48.64	42	5
4	1600	1.0281	3014.98	45.26	46	5.5
5	2000	1.013	3019.65	40.54	50	8

refers to the peak-regulation proportion [22].

The standardized evaluation matrix of indicators is as follows:

$$d_{ij} = \begin{bmatrix} 1.0000 & 0.9939 & 0.6090 & 0.68 & 0.833 \\ 0.9795 & 0.9996 & 0.9848 & 0.76 & 1.000 \\ 0.9652 & 1.0000 & 1.0000 & 0.84 & 1.000 \\ 0.9346 & 0.9987 & 0.9305 & 0.92 & 0.909 \\ 0.9209 & 0.9971 & 0.8335 & 1.00 & 0.625 \end{bmatrix}$$

The corresponding weighting for each index can be obtained (as shown in Table 4).

Table 4: Table of weighting for each index

Weighting	C_T	F_C	S_C	R_G	$LOLP$
H_j	0.9997	0.9999	0.9909	0.9943	0.9916
d_j	4.8	4.9893	4.3578	4.2	4.367
w_j	0.0127	0.0042	0.3856	0.2415	0.3559

The correlation coefficient matrix R is

$$R = \begin{bmatrix} 1.0000 & -0.3275 & -0.3676 & -0.9932 & 0.5133 \\ -0.3275 & 1.0000 & 0.9986 & 0.3509 & 0.5644 \\ -0.3676 & 0.9986 & 1.0000 & 0.3886 & 0.5283 \\ -0.9932 & 0.3509 & 0.3886 & 1.0000 & -0.5157 \\ 0.5133 & 0.5644 & 0.5283 & -0.5157 & 1.0000 \end{bmatrix}$$

The eigenvalues of R can be obtained from $\lambda = diag\,(0, 0.0049, 0.0713, 2.209, 2.7148)$, then $\lambda_1 = 2.7148$, $\lambda_2 = 2.209$. The corresponding weightings are therefore 0.3559 and 0.2415, respectively.

The variance contribution rates are $\rho_1 = 0.543$, $\rho_2 = 0.4418$, $\rho_{sum} = \rho_1 + \rho_2 = 98.48\% > 85\%$, therefore we can take the former 2 principal components for the original 5 indices, calculating the result as in formula (25) and then we obtain F_i as $F_i = (0.4607, 0.5394, 0.5588, 0.5457, 0.4639)$.

The maximum value is 0.5588, as a result, the third plan is the optimal solution, thus a 1200 MW capacity P-S station should be installed in the power grid, and this is consistent with results concluded in the literature [19], because, in the actual system it was indeed decided to install 1200 MW P-S units.

5.2. Scheduling analysis

If we assume that a 1200 MW P-S unit is put into the power system we can carry out a scheduling analysis in order to verify the rationality of the above results. In accordance with the principle of energy-saving dispatching, the generating sequence of the grid is as follows:

1. The renewable energy units which have no adjustment ability should be dispatched first. These units include wind generators lacking output control, nonadjustable hydro and solar;
2. Then the hydro power units which have regulation ability;
3. Combined heat and power units, the electrical output power of which is related to the quantity of heat produced;
4. Coal-fired generators, using a scheduling sequence from low to high in terms of coal consumption per unit of electricity generated;
5. Oil-fired generators.

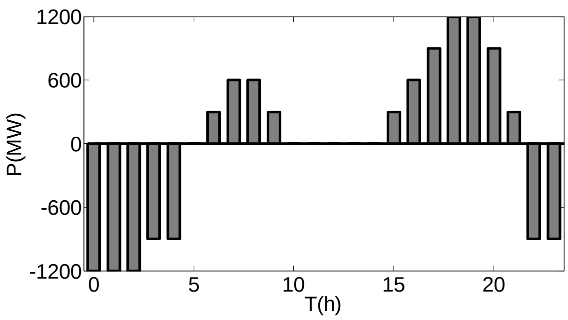

Figure 7: Load difference between actual load and average load

The peak-shaving method is used here to dispatch the load using the principle of energy-saving scheduling, which is particularly suitable for scheduling a system dominated by thermally produced electricity. Therefore, the hydro output is used

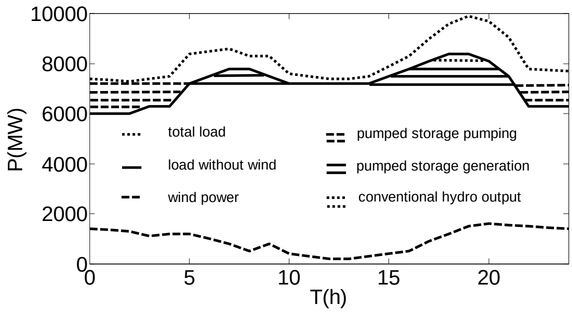

Figure 8: Load dispatch curve

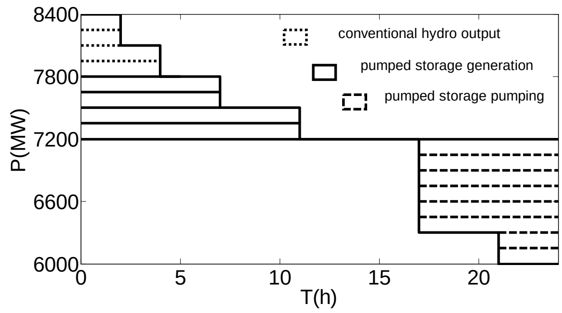

Figure 9: Load duration curve

to cover the peak load in order to reduce the coal consumption of the system. The load difference between actual load and average load, load dispatch and load duration curves are shown in Figs 7 to 9, respectively. Net losses are ignored here to simplify the calculation.

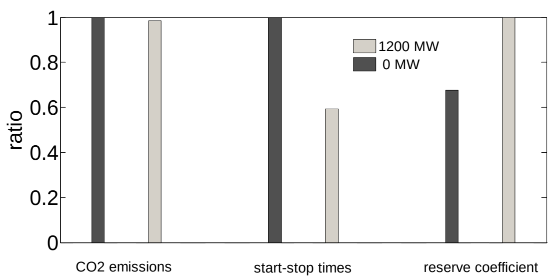

Figure 10: Indicator comparisons with and without 1200 MW P-S

The comparison of the different indicators is shown in Fig. 10 when the 1200 MW P-S units are put into operation in the example grid. It can be seen from the figure that the reductions in CO_2 emissions are not significant. However, the start-stop times of the thermal units are greatly reduced, and the reserve coefficient increases significantly. The P-S station can obtain 3.9×10^6 yuan in a typical day, which is the maximum benefit under the peak-valley time-of-use tariff (as listed in Table 5).

Table 5: Peak-valley time-of-use tariff

Load period	Time, hour	Price, yuan
Peak load	6~9, 17~20	1.5
Flat load	5, 10~16, 21	1.0
Valley load	0~4, 22~24	0.5

The numerical and scheduling results prove that P-S units can have an effect on improving the economy and security of the system, whilst also improving the operating conditions of thermal generation units. The equivalence between these results and our calculated outcomes indicates the value and effectiveness of our proposed method.

6. Conclusion

Our comprehensive evaluation method combines the entropy weighting method and PCA to determine optimum P-S capacity. The weightings of the evaluation indices are determined by the entropy weighting method, and the dimensions are reduced by using PCA. This method can avoid the subjectivity of weighting assignment and therefore simplifies the calculations to an evaluation problem containing multiple evaluation indices.

Under the current state of the electricity market in China, where the auxiliary services are not perfect, some evaluation indices for P-S are difficult to assess, however the effectiveness of P-S stations has been widely recognized by decision makers, although the dynamic benefits of P-S stations are hard to calculate accurately.

As an effective evaluation method, which combines entropy and PCA algorithms, our suggested calculations are widely applicable in multi-index comprehensive evaluations.

Acknowledgments

In the process of writing this paper, I should like to express our thanks for the help and guidance of my colleagues and supervisor.

References

[1] A. Poullikkas, Optimization analysis for pumped energy storage systems in small isolated power systems, Journal of Power Technologies 93 (2) (2013) 78–89.

[2] S. Papaefthimiou, E. Karamanou, S. Papathanassiou, M. Papadopoulos, Operating policies for wind-pumped storage hybrid power stations in island grids, IET Renewable Power Generation 3 (3) (2009) 293–307.

[3] P. D. Brown, J. A. Pecas Lopes, M. A. Matos, Optimization of pumped storage capacity in an isolated power system with large renewable penetration, IEEE Trans on Power Systems 23 (2) (2008) 523–531.

[4] S. V. Papaefthymiou, E. G. Karamanou, S. A. Papathanassiou, M. P. Papadopoulos, A wind-hydro-pumped storage station leading to high RES penetration in the autonomous island system of ikaria, IEEE Transactions on Sustainable Energy 1 (3) (2010) 163–172.

[5] S. M. Chen, S. J. Niou, Fuzzy multiple attributes group decision-making based on fuzzy preference relations, Expert Systems with Applications 38 (4) (2011) 3865–3872.

[6] A. Soroudi, M. Ehsan, R. Caire, N. Hadjsaid, Possibilistic evaluation of distributed generations impacts on distribution networks, IEEE Transactions on Power Systems 26 (4) (2011) 2293–2301.

[7] M. T. Isaai, A. Kanani, M. Tootoonchi, H. R. Afzali, Intelligent timetable evaluation using fuzzy AHP, Expert Systems with Applications 38 (4) (2011) 3718–3723.

[8] H. K. Chan, X. J. Wang, G. R. T. White, N. Yip, An extended fuzzy-AHP approach for the evaluation of green product designs, IEEE Transactions on Engineering Management 60 (2) (2013) 327–339.

[9] Y. H. Chiang, Using a combined AHP and PLS path modelling on blog site evaluation in Taiwan, Computers in Human Behavior 29 (4) (2013) 1325–1333.

[10] W. H. Chen, Quantitative decision-making model for distribution system restoration, IEEE Transactions on Power Systems 25 (1) (2010) 313–321.

[11] Q. Wang, D. M. Kilgour, K. W. Hipel, Fuzzy real options for risky project evaluation using least squares montecarlo simulation, IEEE Systems Journal 5 (3) (2011) 385–395.

[12] S. K. Tyagi, M. Akram, Human reliability evaluation for offshore platform musters using intuitionistic fuzzy sets, IEEE Transactions on Fuzzy Systems 21 (6) (2013) 1115–1122.

[13] D. Liu, M. Liu, Application of sample entropy on measuring precipitation series complexity in jiansanjiang branch bureau of china, Nature Environment and Pollution Technology 12 (2) (2013) 249–254.

[14] B. Hsiao, C. C. Chern, C. R. Chiu, Performance evaluation with the entropy based weighted russell measure in data envelopment analysis, Expert Systems with Applications 38 (8) (2011) 9965–9972.

[15] S. Bouhouche, L. L. Yazid, H. Tarek, B. Jurgen, Inferential sensor-based adaptive principal components analysis for mechanical properties prediction and evaluation, Measurement 46 (9) (2013) 3683–3690.

[16] A. Banimostafa, S. Papadokonstantakis, K. Hungerbuhler, Evaluation of EHS hazard and sustainability metrics during early process design stages using principal component analysis, Process Safety and Environmental Protection 90 (1) (2012) 8–26.

[17] M. F. Qi, Z. G. Fu, Y. Jing, Y. Ma, A comprehensive evaluation method of power plant units based on information entropy and principal component analysis, in: Proceedings of CSEE, Vol. 33, 2013, pp. 58–64.

[18] X. Q. Wen, Y. H. Wang, C. Chen, S. G. Peng, Evaluation on land intensive utilization based on the combination of principal component analysis and entropy value method, Journal of Anhui Agri. Sci 36 (28) (2008) 12372–12373.

[19] D. P. Guo, Research on benefits and economy for pumped storage station, Master's thesis, Dalian University of Technology, China (2001).

[20] F. Cao, L. Z. Zhang, Determination of pumped-storage plant capacity with peak-regulation proportion, Electric Power Automation Equipment 27 (6) (2007) 47–50.

[21] J. C. Cui, D. H. Liu, W. L. Liang, F. Xie, H. Y. Chen, Analysis on economic and environmental benefit of pumped-storage station, Electric Power 40 (1) (2007) 5–10.

[22] Z. A. Zhang, X. G. Cai, Determination of pumped storage capacity based on entropy, Electric Machines and Control 18 (3) (2014) 34–39.

[23] H. Z. Nie, S. Nie, Y. Qiao, P. Lv, Comprehensive decision-making of alternative transmission network planning based on principal component analysis, Power System Technology 34 (6) (2010) 134–137.