Supplementary Files

1. Horizontal Predictive Control Model Algorithm

Horizontal predictive control model algorithm is described as follows:

Input: Training samples input item *input_train*, Output item *output_train*;

Test sample input item *input_test*, Output item *output_test*;

Weak predictor number *K*;

Output: Strong predictor *SP* and mean square error *mse*;

BEGIN

1 For i = 1:K

2 If i == K

PSO algorithm optimizes the initialization parameters of the *SVM* and establishes an optimal *SVM* weak predictor;

3 Else

PSO algorithm optimizes the *RNN* parameters and establishes an optimized *RNN* weak predictor;

4 End If;

5 Train the *i* weak *PSO*-prediction model until convergence;

6 Calculate the weighted prediction error of the training samples $error\_yc_i(x)$ and absolute error sum E_i;

7 Adjust the weight $w(x)$ of the training sample according to the prediction error and normalize it;

8 Calculate the weight at_i of the weak prediction model based on E_i.

9 End For;

10 Weak predictor weights are normalized and composed of strong predictors *SP* according to weights;

11 Bring the test sample *input_test* into the *SP* to get the *mse* between the predicted value and the actual value *output_test*.

END

2. Abnormal Data Detected of Water Level Data in Guxiandu Station in 2011

As shown in Figure 1, the first half is a histogram (quarterly) of data quality inspection at Guxiandu station. Suspicious data are mainly concentrated in the second and third quarters, less in the first and fourth quarters, while the fourth quarter has the least amount of abnormal data. The main reason is that the fluctuation of water level in summer is large, the water level in winter is relatively stable, and the variation of time series increases the difficulty of control.

The lower part of Figure 1 is a histogram of detection statistics by suspicious level. There are more types of data such as vacancy detection and suspicious level 1. The data of suspicious level 1 is mostly caused by the rapid rise of the water level in two or three days in the second quarter and the third quarter, and the lag of the statistical control method. The other two predictive control models have better adaptability. Time-varying inspection detects less suspicious data, mainly because the time-varying threshold is not set strictly in the second and third quarters. In addition, if the data does not pass the extreme value detection, it will be marked and directly controlled by the predictive control model to give the prediction value and confidence interval, no longer through the time-varying detection steps.

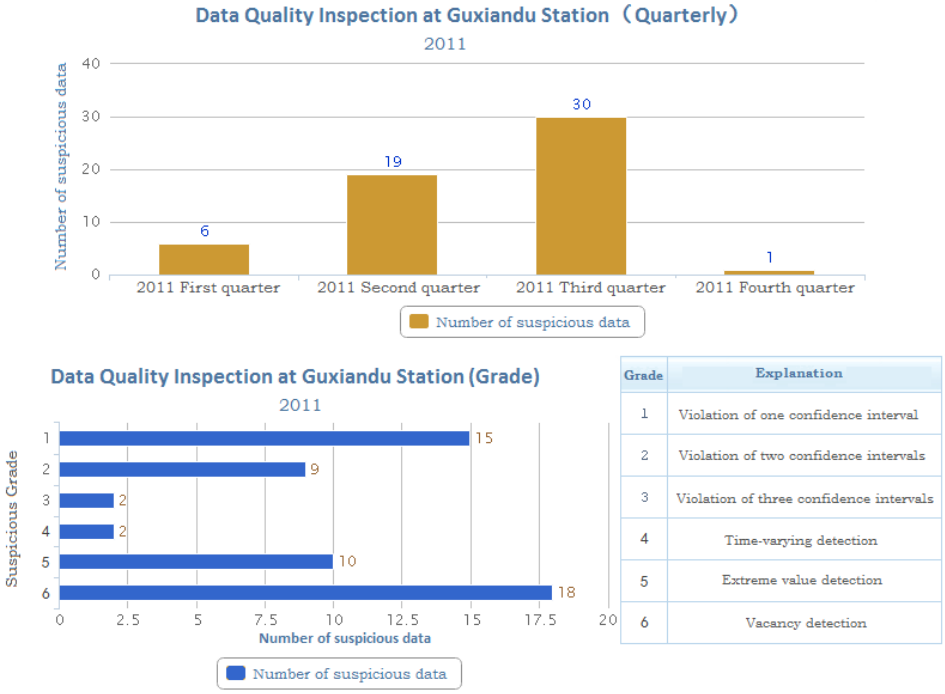

Figure S1. Statistical chart of data quality test results of Guxiandu station.

The following table classifies the suspicious grades and lists several sets of abnormal data of the water level at Guxiandu station in tabular form (as shown in Table 1).

Table S1. Abnormal check list of water level hourly data in Guxiandu station in 2011.

Grade	Data	Detection Value(m)	Suggested Value 1 (m)	Suggested Value 2 (m)
6	2011-06-26 20:00:00	0	17.071	17.073
	2011-06-26 21:00:00	0	16.995	17.063
	2011-06-26 22:00:00	0	16.989	17.063
	2011-06-26 23:00:00	0	16.993	17.053
	2011-06-27 03:00:00	0	16.925	16.927
	2011-06-27 04:00:00	0	16.912	16.915
	2011-06-27 08:00:00	0	16.935	16.932
	2011-12-13 08:00:00	0	13.470	13.468
5	2011-06-25 18:00:00	27.30	17.231	17.232
	2011-06-25 19:00:00	27.30	17.232	17.230
	2011-07-17 04:00:00	24.10	15.641	15.640
	2011-07-17 05:00:00	24.10	15.643	15.645

	2011-07-17 06:00:00	24.10	15.647	15.649
4	2011-04-23 10:00:00	13.40	13.732	13.730
	2011-08-18 19:00:00	15.01	14.532	14.529
3	2011-06-05 20:00:00	14.45	14.241	14.272
	2011-06-05 21:00:00	14.65	13.332	14.391
2	2011-06-05 22:00:00	14.71	14.431	14.522
	2011-06-05 23:00:00	14.83	14.486	14.722
	2011-06-06	14.91	14.636	14.892
1	2011-06-06 01:00:00	14.99	14.841	14.903
	2011-06-06 02:00:00	15.07	14.869	15.025
	2011-06-06 03:00:00	15.09	14.975	15.069

Note: Table S1 lists the typical data time points in the test results of the Guxiandu station.

The vacancy data of Grade 6 are mainly concentrated in 2011-06-26:20:00 to 2011-06-27:08:00, and there are 13 hours without data records. The problem may be caused by machine failure. The proposed value 1 is the predicted value of the horizontal control model, it provides the recommended value in the missing time. The proposed value 2 is the predicted value of the longitudinal control model. Because the missing time is 13 consecutive hours, and the model predicts the next time according to the data of historical adjacent time. In order to ensure the accuracy of the predicted value, the model only provides short-term recommendations.

Grade 5 is the result of extreme testing, mainly concentrated in the second quarter (2011-06-2518:00 and 2011-06-2519:00) and the third quarter (2011-07-17 04:00 to 2011-07-17 06:00). It is found that the extremum detection results deviate from the peripheral value more, and mainly large, these may be due to the machine short-term monitoring instrument failure.

Grade 4 is the time-varying detection result. Since the set time-varying rate is large, there is less suspicious data detected by time-varying.

Grades 1, 2, and 3 are the results of the continuous hydrological data method in this paper. The results are mainly concentrated in 2011-06-05 20:00 to 2011-06-06-06 03:00, the water level rises from 13.89 m in 2011-06-05 19:00 to 15.09 m in 2011-06-06-06 03:00, the water level rises more than 1 m in 8 hours. For the water level variation rule of the Guxiandu station, it belongs to the abnormal time period. Due to the rapid rising speed of water level, the control interval obtained by the control model (prediction and statistics) cannot keep up with the changing speed of water level, which leads to the appearance of suspicious data, more of which is a suspicious sequence.

A total of 56 suspicious data were detected from the hourly water level data of Guxiandu Station in 2011, accounting for 0.64% of the total data. Except for the vacant data, the proportion of suspicious data detected by the proposed method was 0.41%.