Quality control of Sea Level Data

National Marine Data and Information Service, China

Guo Fengyi

Abstract

In order to get higher quality some quality control have to be done and quality control and quality assurance methods have to be designed. In this report some quality control methods of sea level data are described. These methods are widely used in application program of quality control in China. General and reasonable check is available.

1 Preface

Along China coast there are several stations where tide was observed (See figure 1). The sea level data obtained from these stations are sent to CNODC (China National Oceanographic Data Center). Users can get the data from CNODC. Before these data are contributed to users, CNODC has to do some work in order do get higher quality. This is quality control. In this report we only describe some quality control methods which are used to check the sea level data sent to CNODC.

2 Quality Control Procedures

Up to now many quality control methods of sea level data are designed in the world. But the following procedures are widely used and are available.

a. Data field verification
b. Data relationship verification
c. Statistical test

2.1 Data Field Verification

The data field verification procedure performed a variety of check on individual data fields to assure that the characteristics of any given field matched characteristics defined as acceptable for that field. The tests were performed to assure that:

a) Each field contained the expected data type.
b) Data values were written within acceptable ranges.
c) Fields filled with some codes contained acceptable code values
d) Mandatory fields were present

The following table is a sample for data field verification:

latitude	0 <= X <= 90
longitude	0 <= X <= 180
year	start date <= X <= end date

month	1 <= X <= 12
date	1 <= X <= 31
hour	0 <= X <= 24
time	0 <= X <= 60

where X = Observation Value

2.2 Data Relationship verification

The data relationship verification procedure is used to evaluate the appropriateness of data values in relation to other data values. The test is performed to assure that every value is acceptable and reasonable.

2.3 Data Statistical test

The statistical tests are used to determine the acceptability of data values in comparison to existing values. For example, we know that the depth of sea at one point is nearly fixed except the effect of sea level, especially in open sea.

The following formula was used:

$$M-3.5*SD < X < M + 3.5*SD$$

where M = Mean value
SD = Standard Deviation
X = Observation value

3. Quality Assurance

In order to get higher quality for sea level data the State Ocean Administration (SOA) of China published “Specification for Oceanographic Survey” in 1961. The “Specification for Oceanographic Survey” was modified in 1975 and 1990. In the specification the advantageous technology, data processing procedures and many information codes are adopted. According to the specification the observations at every station along China coast must be performed following the specification. General speaking by the means of the specification the obtained data at the above stations is reliable and acceptable. The quality of the data is assured.

4. Data Field Verification

After CNODC receive any sea level data from the above stations, some quality control work must be done. First of all we have to do general check , that is data field verification. It contains the following information:

a. For one station the name must be same at different files and it’s unique.

b. For one station the location must be same at different files and it's unique.
c. The month of sea level data must satisfy the following condition:

1<=M<=12

where M is month

d. The date of sea level data must satisfy the following condition:

1<=d<=D(M)

where M is the maximum date of the month

e. The hour of sea level data must satisfy the following condition:

0<=H<=24

where H is hour

f. Every field must have no illegal codes.

5 Data Relationship verification

We know that the tide livel is changing regularly with time. The tide level of one station at one moment can be calculated by means of Lagrange interposition formula. The following formula are used:

R(k)=2/3(R(k+1)+R(k-1))-1/6(R(k+2)+R(k-2))..... (1)
(k=0,1,.....K)

where R(k) is mean value of sea level at the moment k
K is number of obtained sea levels
R(k-2), R(k-1), R(k+1), R(k+2) respectively represent the value of sea level at the moment k-2, k-1, k+1, k+2
According to the continuity of tide process the calculated value should be nearly same as the observed value of sea level. If the difference between calculated value and observed value at the moment k is too large, then one of the observed values is doubtable.

Now the problem is how to judge the doubtable value. We use the following formula:
Abs (Rk-Rm) >M……..……(2)
where Rk is observed value of sea level at the moment k
Rm is a mean value
M = 4.374SD
4.374 is a constant. It is calculated by normally distribution function
SD is Standard Deviation
If Abs (Rk-Rm) >M, the value Rk is doubtable.

We find that the mean value and standard deviation of tide level at different point of

one tide curve at one station is different, for example, at the station Mawei.

Mean Value and Standard Deviation at Different Point of Tide Curve
(Mawei 1976)

Time	Mean Value(cm)	Standard Deviation(cm)
one hour before low tide	9.6	5.0
at low tide	-20.3	14.0
one hour after low tide	- 1.7	15.7
two hour after low tide	7.1	6.9
three hour after low tide	1.7	3.8
other time	0.5	3.8

This table shows that the mean value and standard deviation at different point of one tide curve is different. So the value of Rk at different point of one tide curve should be considered respectively and Different standard at different point of one tide curve should be adopted .

In order to determine any doubtable data and it's correct value, we also need to compare the observed tide level with astronomical tide level. The astronomical tide level is calculated by harmonic constant of tide.

6 Data Statistical test

We find that some errors of sea level data arise from zero point of rod. Sometimes the zero point of rod at one station may be changed for some reason, but we don't know. In this case the sea level data is not correct. Using data statistical test this variation con be found. From the monthly mean sea level we can find this problem. Using the harmonic constant of main component tide , such as K1 and M2, we can judge the reliability of a data set at one tide station. For example, Table 2 shows 1961-1963 monthly sea level at Lianyungang station. We can find that the zero point of rod from September to November 1962 was changed and the sea level data are not reliable.

Table 2 1961-1963 Monthly Sea Level(cm)
(Lianyungang Station)

Month	1961	1962	1963
1	279.5	272.9	255.2
2	282.9	277.4	263.6
3	288.0	264.5	279.7
4	292.8	280.4	285.5
5	297.8	294.5	292.3
6	310.7	300.7	301.8
7	310.1	329.7	308.6
8	322.5	322.6	312.3
9	315.4	350.1	310.3
10	313.0	351.4	294.0
11	296.2	336.1	288.9
12	284.0	283.5	277.7

7 Error Message and Analysis

After quality control we may find some error messages. We have to design some methods to judge the messages in order to correct them. The following procedures are used

a) If it is a clerical error, we just correct it.
b) If it is a kind of systematic errors, we can correct them also.
c) If it is a doubtable data, we can ask the observer of the data. If the observer can not tell the data is right or not, we just add a quality flag.
d) If we find that some data set is not reliable, we must do some analysis in order to find some reason and correct them. Before we correct them, we just delete them from some files.

Coastal Station Map of China