ministerie van verkeer en waterstaat

koninklijk nederlands meteorologisch instituut

A real-time wave data quality control algorithm

María Paula Etala

Technical reports; TR-139
Technische rapporten; TR-139

De Bilt, 1991

de bilt 1991

publicatienummer: Technical reports=
technische rapporten; TR-139 (OO)
Division of Oceanographic Research

postbus 201
3730 AE de bilt
wilhelminalaan 10
tel. (030) 206911
telex 47096
fax (030) 210407

On leave from: Servicio Meteorológico de La Armada,
Buenos Aires, Argentina

auteur: María Paula Etala

U.D.C.: 551.46.062.5
551.466.3

ISSN: 0169-1708

ISBN: 90-369-2009-4

A real-time wave data quality control algorithm

María Paula Etala

Technical reports; TR-139
Technische rapporten; TR-139

De Bilt, 1991

A REAL-TIME WAVE DATA QUALITY CONTROL ALGORITHM

María Paula Etala [1]

Department of Oceanographic Research

KNMI, De Bilt, The Netherlands

1 - INTRODUCTION

Wave data quality control is a common practice in climatological or model verification applications, and it is based on both objective criteria and human experience. Traditional wave models do not use wave observations. Initial fields are retrieved only from wind fields. Currently, new data assimilation techniques in wave models are being tested in order to improve the forecast by assimilating wave observations.

This paper gives a first approach to real-time quality control of wave height and period observations from ships, buoys and platforms in the North Sea and Norwegian Sea. The method follows, in a simple manner, the idea of Comprehensive Quality Control (CQC) (Gandin,1988), first applied at the Hydrometeorological Center in Moscow and already in practice at the National Meteorological Center in Washington (W.G. Collins and L.S. Gandin, 1990) for height and temperature rawinsonde data.

The principle of the method is not to reject data by the application of sucessive checks, but to apply at least two more or less independent checks and then evaluate the results by a so-called Decision Making Algorithm (DMA). The DMA can either reject, accept, suspect, correct or rehabilitate the data.

---

[1] Permanent address: Servicio Meteorológico de la Armada, Buenos Aires, Argentina.

It would be very difficult, if not impossible, to obtain for wave data all the benefits from the method when applied to rawinsonde data. Reported values are not computed but observed, so most errors are observational and those are the most difficult to treat. In addition, the relationships between different variables are less direct for waves, making cross-checks highly complicated.

Some freedom has been taken in the choice of the checks of the present approach, in particular with respect to the hierarchial level required by the CQC.

2 - THE QUALITY CONTROL APPROACH

2.1 - THE DATA

The data used are the mean period and the significant wave height from buoys and platforms and the windsea and swell period and height measured visually from ships. Characteristic height and mean period are obtained from the latter in the following manner:

$$(1) \quad H_c = \sqrt{H_w^2 + H_{w1}^2 + H_{w2}^2}$$

$$(2) \quad T = \frac{H_w^2 P_w + H_{w1}^2 P_{w1} + H_{w2}^2 P_{w2}}{(H_w^2 + H_{w1}^2 + H_{w2}^2)}$$

where: H_w is the significative height of wind waves;
P_w is the observed period of wind waves;
H_{w1}, H_{w2} are the significative height of swell 1 and 2, respectively;
P_{w1}, P_{w2} are the observed periods of swell 1 and swell 2, respectively.

No conversion is made between characteristic and significative wave height, neither between buoy, ship and modelled mean periods. Equations (1) and (2) are evaluated after check n°4, where windsea and swell control takes place, as will be seen in the following section.

Observational errors for height are asumed to be proportional to the height value, as it was done by the data assimilation algorithm. These errors are :

$E_h = 0.3 + 0.20\ H_c$ for visual observations, in meters.

$E_h = 0.3 + 0.05\ H_c$ for measured observations, in meters.

Observational errors for period are proportional to the mean period for measured observations and are constant for visual observations. They are :

$E_t = 2$ for visual observations, in seconds.

$E_t = 0.5 + 0.10\ T$ for measured observations, in seconds.

Constants have been adjusted to follow aproximately the values in WMO (1988).

2.2 - THE CHECKS

Two general types of checks, interpolation checks (IC) and checks against limits (LC), are applied to three different types of values: observation itself, difference between observation and first-guess and difference between current observation and past observation. So, there are three interpolation checks (nº 1, 2, 3) and three limit checks (nº 4, 5, 6). Each limit check has its corresponding interpolation check (CIC), as shown in Table 1.

	IC	LC
Obs.	1	4
Obs.vs.FG	2	5
Obs.vs.Past obs.	3	6

Table 1. Limit checks, interpolation checks and their correspondences.

2.2.1 - LIMIT CHECKS

The first check which is applied is **check n°4**. This is the only check where data can be simply rejected. If the data are provided by buoys or platforms then it just compares height and period against maximum and minimum limits and period versus height. This check is summarized in Tables 2 and 3. The lower limit of 2 sec. for the period is only due to a limitation in the Nedwam wave model discretization of the spectrum (Burgers, 1990). Data assimilation in Nedwam (Burgers et al., 1990) is the current application of this approach, as will be seen in section 3. As shorter periods are possible and may be correct, the rejection because of this lower limit is made after the complete check has been performed.

If the data come from a ship then several checks are applied with respect to windsea, swell 1 and swell 2, as follows. Each reported height in these three groups is taken into account in formula (1) although the period in the same group may be missing. But if in one group a height is missing and period is reported, then the characteristic height is not calculated from (1) and it is considered to be missing. For example, if

$$H_w = 6$$
$$P_w = 8$$
$$H_{w1} = 8$$
$$P_{w1} = 99$$
$$H_{w2} = 99$$
$$P_{w2} = 99$$

then, from (1), $H_c = 5$m but mean period is not calculated because P_{w1} is missing and swell 1 actually exists.

The same criterium is applied for periods but, in addition to that, as the mean period in (2) depends on both parameters in each group, whenever period is reported but height is missing in one group, mean period is also considered to be missing.

H<0 or H>50	....	H rejected
H>35	....	H suspected
T≤2 or T>30	....	T rejected
T>20	....	T suspected

Table 2. Check nº4. Gross limit checks for buoys and platforms. H in 0.5m, T in sec.

H (0.5m)	T (sec)
0	0 - 5
1	0 - 11
2	0 - 13
3	0 - 15
4	0 - 17
5	0 - 19
6 - 12	3 - 21
13 - 20	6 - 23
21 - 50	8 - 30

Table 3. Check nº4. Height vs period for buoys and platforms. Data that do not fit the criteria are both suspected.

Any data beyond the limits shown in Table 4 are rejected and the corresponding total value is not calculated.

H_w, H_{w1}, H_{w2} < 0 or > 50 ..	H_c rejected
P_w, P_{w1}, P_{w2} < 0 or > 30 ..	T rejected

Table 4. Check nº4. Gross limit checks for windsea and swell. H in 0.5m, P in sec.

Windsea height (H_w) is checked against wind speed (FF), which is considered to be correct. H_w data that do not fit limits shown in Table 5 are suspected.

Windsea period is also checked against wind speed, following Table 6 and 7. Data that fail in fitting the criteria are suspected.

Windsea height is checked against windsea period as shown in Table 8. Data that do not fit the criteria are both suspected.

Swell height is checked against swell period following Table 9. Data that do not fit the criteria are both suspected.

FF (kts)	H_w (0.5m)
0	0
1 - 15	0 - 4
16 - 25	1 - 10
26 - 35	1 - 16
36 - 45	2 - 24
46 - 60	3 - 30
61 - 99	4 - 40

Table 5. Check nº4. Windsea height versus wind speed.

P_w (sec)	FF (kts)
0	< 10

Table 6. Check nº4. Wind waves period versus wind speed.

FF (kts)	P_w (sec)
0	0

Table 7. Check nº4. Wind speed versus windsea period.

H_w(0.5m)	P_w(sec)
0	0 - 3
1	0 - 6
2	0 - 8
3	0 - 10
4	0 - 12
5	0 - 13
6 - 12	3 - 15
13 - 20	6 - 17
21 - 29	8 - 20
30 - 40	10 - 24
41 - 50	12 - 30

Table 8. Check nº4. Height vs period for windsea.

H_{w1}, H_{w2}(0.5m)	P_{w1}, P_{w2}(sec)
0	0 - 5
1	0 - 11
2	0 - 13
3	0 - 15
4	0 - 17
5	0 - 19
6 - 12	0 - 21
13 - 20	0 - 23
21 - 50	0 - 30

Table 9. Check nº4. Height vs period for swell.

It is an usual error that observers report the highest wind waves as swell. Swell and wind directions as well as swell 1 and swell 2 directions must differ in more than 20°. If they do not, then heights must differ in more than 1m. If this condition does not follow, the corresponding swell is removed but the data is not suspected. The remaining swell or windsea height is the maximum of the two compared heights and the period is the mean of both periods, weighted in H^2 (E. Bouws, personal communication).

Data that have not been suspected are subjected to the last limit check, which is shown in Table 10.

$H_w > 40$	.....	H_w suspected
H_{w1}, $H_{w2} > 35$	.....	H_{w1}, H_{w2} suspected
$P_w > 20$	.....	P_w suspected
$P_{w1}, P_{w2} > 25$	.....	P_{w1}, P_{w2} suspected

Table 10. Check nº4. Gross limit checks for windsea and swell. H in 0.5m, P in sec.

Finally, H_c and T are calculated from (1) and (2). If the mean period is less or equal than 2 sec., then it is rejected.

Limits and criteria in this check have been taken from Stam (1989) and from R. van Moerkerken (personal communications).

The next check, **check n°5**, considers the difference between the observation and the nearest point of the first-guess field. The limit differences are shown in Table 11. Every observation beyond these limits that has not been rejected by the previous check is flagged as suspected.

Note that in this and in the further checks the quantities in consideration are the total significant height and the mean period.

$\|H - Hfg\|$	2
$\|T - T_{fg}\|$	2

Table 11. Check n°5. Limits for the difference between observations and first-guess for height and period in meters and seconds, respectively.

Check n°6 takes into account past observations that have been considered correct by the previous control from 3 or 6 hours earlier. The closest past observation for each current observation is selected. If there is not any past observation within a maximum distance of 90 km from the current observation point or the current observation has been rejected by check n°4, the check is not applied.

The limits allowed to the difference between current and past observations are shown in Table 12.

	t - 3 (hs)	t - 6 (hs)
$\|H - H_p\|$	2	3
$\|T - T_p\|$	2	3

Table 12. Check n°6. Limits for difference between current and past observations for height and period in meters and seconds, respectively.

Data that are beyond these limits are flagged as suspected. It is obvious that greater differences can occur, but time series of the North Sea indicate that the most usual cases can be within these limits (R. van Moerkerken, 1990, 1991). Anyway, the data are just suspected by this check, as well as by check n°5, and would only be rejected if other checks react. Being too permissive in this stage would go against the performance of the total approach, as will be seen in section 2.3. It is clear that experience with this method is the best way to tune these kind of details and can be used as a feed-back for further improvement.

2.2.2. - INTERPOLATION CHECKS

These checks apply to the same parameters as the limit checks: observed height and mean period **(check n°1)**, observed height and period minus first-guess **(check n° 2)** and observed height and period minus past observation **(check n°3)**. The test structure is exactly the same for all of them.

Some conditions are required concerning distance and distribution of the points in order to make an interpolation. Observations within a maximum distance D_{max} = 300 Km around the interpolation point are considered. This circle of D_{max} radius is divided into octants and it is required that the observations are distributed in at least two non-consecutive octants. So, there should be at least two points, distributed for example, as shown in Figure 1-a, but not as in Figure 1-b. This condition is overridden if the minimum distance of the points to the interpolation point is less than 75 Km. This distance corresponds to one grid unit of Nedwam.

Interpolated values are obtained for each non-rejected observation point from surrounding non-rejected observations.
The interpolation is just a weighted mean of observations. We define the interpolation function as

$$f(X,P_0) = \sum_{i=1}^{n} w(P_i,P_0)\, x_i \tag{3}$$

where $X = (x_1,x_2,\ldots,x_n) = [x(P_1),x(P_2),\ldots,x(P_n)]$ is the array of the values of the parameter that is being interpolated to P_0, in the interpolating points P_i. The weight function is

$$w(P_i,P_0) = \frac{1 / [1+D^2(P_i,P_0)]}{\sum_{j=1}^{n} 1 / [1+D^2(P_j,P_0)]} \tag{4}$$

where $D(P_i,P_0)$ is the distance between each point and the interpolation point. This function was preffered to Cressman's

$$w(P_i,P_0) = \frac{[D_{max}^2-D^2(P_i,P_0)] / [D_{max}^2+D^2(P_i,P_0)]}{\sum_{j=1}^{n} \{[D_{max}^2-D^2(P_j,P_0)] / [D_{max}^2+D^2(P_j,P_0)]\}}$$

for its rapid decrease with distance.

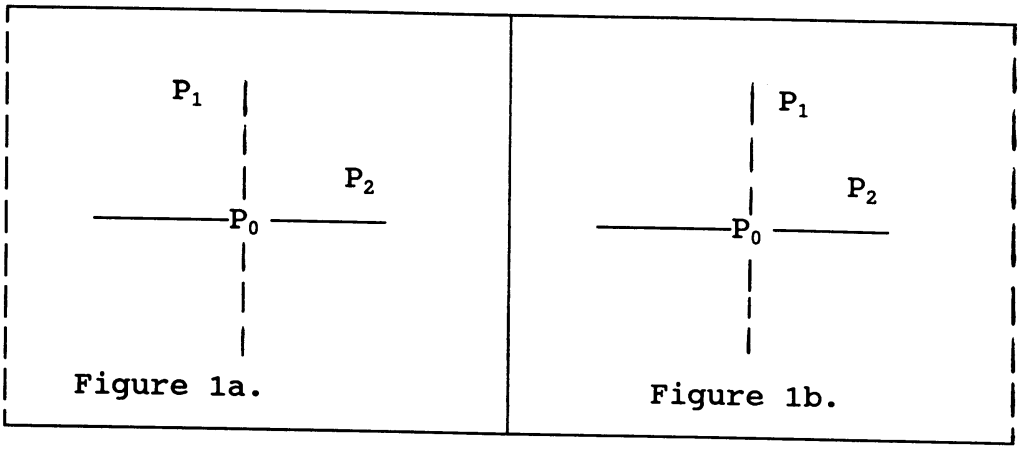

Figure 1.a - An example of the minimum condition distribution for interpolation.
Figure 1.b - An example of a on no-interpolation condition.
P_0 = interpolation point; P_1,P_2 = interpolating points.

Let's introduce now the "RMS comparison error" (E_c) as

$$(5) \qquad E_c^2 (P_0) = \overline{[\, f(X,P_0) - x(P_0)\,]^2}$$

and if we assume that the errors of the parameter in different points are uncorrelated and that the mean error is null, i.e.

$$\overline{r_i \; r_j} = \delta_{ij} \; \epsilon_i^2$$

and

$$\overline{r_i} = 0$$

where r_i is the error in P_i and δ_{ij} is the Dirac delta function, then E_c becomes

$$(6) \qquad E_c(P_0) = \sqrt{E^2_{int}(P_0) + \epsilon^2 (P_0)}$$

where E_{int} is the interpolation error and $\epsilon(P_0) = \epsilon_0$ is the error in P_0 inherent to the parameter that is being interpolated.

ϵ is the random observational error (ROE) if we are interpolationg observations, it is the sum of ROE and estimated model error if we are interpolating differences between observation and first-guess or, finally, it is the sum of current and past observation ROE if we are interpolating the difference of these two values.

The expectation value of the error due to interpolation is

(7) $$E^2_{int} = [\ f(X_t,P_0) - x_t(P_0)\ ]^2 + \sum_{i=1}^{n} w^2(P_i,P_0)\ \epsilon^2(P_0)$$

where $X_t = [x_t(P_1),x_t(P_2),\ldots,x_t(P_n)]$ is the array of true values in the interpolating points P_i. The first term in (6) is not known but the second term can be calculated and it is the lower limit of E^2_{int}. So, the interpolation error is estimated as follows :

(8) $$E^2_{int} = \max \{\ [\ f(X,P_c) - x(P_c)\ ]\ ,\ \sum_{i=1}^{n} w^2(P_i,P_0)\ \epsilon^2(P_0)\ \}$$

where P_c is the closest point to P_0 (G. Burgers, personal communication).

It is possible to obtain the quantity [$f(X,P_i) - x(P_i)$] for all interpolating points P_i but it would not be a realistic measure of $E_c(P_0)$. The interpolation problem at the border of the circle is not representative of the actual problem. That is why $E_c(P_0)$ is calculated by replacing (8) into (6).

Now that we have $E_c(P_0)$ we are able to make the final evaluation of the check. In general, the modulus of the difference between the true value in point P_0 and the interpolated value should be less than K times E_c to be accepted.

(9) $$|\ x(P_0) - f(X,P_0)\ | \leq K\ E_c$$

It is usual in meteorological applications that K=4. In this application we take K=1.5 for the interpolation of observations and K=1 for the interpolation of differences, because observational errors are large enough to do this.

If (9) does not follow in an interpolation check then the observation is flagged as suspected by that check.

2.2.3 - THE TABLE OF RESULTS

Results of the different checks for each observation are compiled in a table. The flag "OK" means that the data has been accepted by the check, the flag "NO" means that it has been suspected and the flag "NC" means that the check has not been applied to that observation, either because it was not possible or because the data had been rejected previously.

Together with the flag appears the difference between each parameter and the limit of the test. Flags "OK" are in correspondence with negative values and "NO"s with positive values.

2.3 - THE DECISION

This is the most important part of the approach and it is particularly the one that requires further adjustments according to experience with the method. It evaluates the information given by the different checks and eventually rejects the data depending on this evaluation. This is done in several steps. After each step, interpolation checks are repeated using only the remaining observations and the table of results is updated. Usually, some data that have been flagged because of nearby wrong data are now rehabilitated. Occasionally, an accepted observation can now be suspected by an IC.

These steps are grouped in two stages. The first stage applies to all sources: ships, buoys and platforms. Data with strong indications of being definitely wrong are rejected at this stage. The second one applies only to ships and buoys and looks at finer details, mostly on the basis of interpolation checks. In general, measured data from platforms are trusted and are never rejected just for nearby visual or buoy data. That is why they do not pass through the final stage.

First Stage

Step 1. Those data for which IC's could not be carried out must fit more strictly to LC's in order to be accepted. If two LC's react, the data are rejected, unless it is a platform which is not suspected by check n°4. So, the only condition for rejection of a platform is that check n°4 and another LC react. If check n°4 accepts the observation, it is not rejected even if it does not match first-guess or past observation. If it is a ship or a buoy and check n°6 (past) could not be made and check n°5 (first-guess) reacts, it must do so for more than 2 m or sec for height and period, respectively, to be rejected.

For those data for which IC's have been made the condition for rejection is the following: Any LC fails for more than 0.5, also its CIC fails, as well as another LC and another IC. In other words, two LC's and two IC's must react and at least two of them should correspond. The reason is that a correct CIC rehabilitates the LC because it means the condition is more or less the same for all observations in the area and limits could be too restrictive.

Table n°13 illustrates some examples of conditions for rejection in this step.

Step 2. The conditions for cases of no interpolation checks are the same as in Step 1. The check for observations with no-interpolation conditions is made in every step because conditions for interpolation may change as we update the table of results after the rejection of data.

If an observation does not match the general limit test (check n°4) and its CIC (check n°1) then it is rejected. If another limit check is wrong, a difference of 0.5 above the limit, reaction of its CIC and of another IC is requested to reject the data. Examples are shown in table n°14.

After repeating the IC's this step is iterated until there are no more data satisfying these conditions, up to a maximum limit of three times. If this limit has to be applied then it means that rejected data are making the IC's worse instead of improving them. In that case a warning appears in the final output.

At the end of this stage remain suspected only:

a) data only suspected by IC's, that means that any nearby point can be wrong;

b) data with wrong limits but not suspected by IC's, which means that limits can be too restrictive;

c) data only suspected by a LC different from check n°4 and its CIC, which means that either past observation or first-guess can be wrong;

d) data only suspected by a LC and an IC different from its CIC, which means that the LC perhaps is too restrictive because nearby observations assess such a difference with past observation or first-guess. If the condition follows it can fail for two LC's or two IC's (see Table 19), but never two IC's and two LC's can be wrong.

Check / Source	1	2	3	4	5	6
PLAT/SHIP/BUOY	NC	NC	NC	NO	NO	
PLAT/SHIP/BUOY	NC	NC	NC	NO		NO
SHIP/BUOY	NC	NC	NC		NO	NO
SHIP/BUOY	NC	NC	NC		NO >2	NC
PLAT/SHIP/BUOY	NO	NO		NO	NO>.5	
PLAT/SHIP/BUOY	NO	NO			NO>.5	NO
PLAT/SHIP/BUOY		NO	NO	NO	NO>.5	
PLAT/SHIP/BUOY		NO	NO		NO>.5	NO

Table n°13. Stage 1 - Step 1. Examples of conditions for rejection. Blanks mean either NO, OK or NC.

Check / Source	1	2	3	4	5	6
PLAT/SHIP/BUOY	NO			NO		
PLAT/SHIP/BUOY	NO	NO			NO>.5	
PLAT/SHIP/BUOY		NO	NO		NO>.5	
PLAT/SHIP/BUOY		NO	NO			NO>.5
PLAT/SHIP/BUOY	NO		NO			NO>.5

Table n°14. Idem 13 for Stage 1 - Step 2.

Final Stage

This stage does not apply to platforms. If data to which interpolation checks have not been applied do not match the first-guess and any other LC they are rejected. This is the same for every step.

Step 1. In this step, observations that fail for a LC and an IC different for its CIC, but CIC result is very close to the limit are rejected. That is, suspected data of type a), b) and c) are the same as in the first stage, but now the remaining type d) suspected data must pass with some margin the CIC. Table 15 shows an example of rejection in this step.

Check / Source	1	2	3	4	5	6
SHIP/BUOY	NO	OK<.1			NO	

Table n°15. Idem 13 for Stage 2 - Step 1.

Step 2. Now those data that have failed only for a LC and its CIC are tested. If another IC is very close to the limit of suspection, then it is rejected. So, the remaining suspected data of type c), failing for CIC, must pass the rest of the IC's with some margin. Table 16 gives an example of rejection.

Check / Source	1	2	3	4	5	6
SHIP/BUOY	OK<.1		NO			NO

Table n°16. Idem 13 for Stage 2 - Step 2.

Step 3. Remaining suspected data of type d) are now completely rejected. Table 17 shows an example. So, only type a), b) and c) should be still suspected. It can happen that conditions for rejection in previous steps now occur because of the updating of the table of results. These conditions are not always checked again but the whole decision procedure is iterated after step 4 has been completed, to be sure that all data satisfying rejection conditions are removed.

Check / Source	1	2	3	4	5	6
SHIP/BUOY			NO		NO	

Table n°17. Idem 13 for Stage 2 - Step 3.

Step 4. Data that have been suspected by at least two IC's (type a)) are rejected in this last step. It is illustrated in Table 18.

The two stages are repeated until there are no more rejected data, up to a maximum of five times. If any data were still rejected in the fifth iteration, a warning message would appear in the standard output. This would mean that the approach is not working properly and that it leads to the degradation of the quality of accepted data.

All data that still remain suspected after these four steps and iterations, if they have been neccessary, are now accepted. They are summarized in Table 19 and they are :

a) data only suspected by one IC;

b) data with wrong limits but accepted by IC's;

c) data only suspected by a LC different from check n°4 and its CIC, but passing other IC's with some margin.

It means that ship and buoy data are never rejected just by one IC, neither by only wrong limits. If IC's are performed, the possibility of wrong first-guess or past observation is also considered (but not both at the same time), if the checks that do not refer to them are correct.

Check / Source	1	2	3	4	5	6
SHIP/BUOY	NO	NO				

Table n°18. Idem 13 for Stage 2 - Step 4.

3 - IMPLEMENTATION

The data quality control is a system of routines called by the wave data entry routine (OBPROC) in the main data assimilation routine of NEDWAM (ASWAVE).

A diagram of the system is shown in figure 2, where the routines in black letters are the ones that have been added for this quality control approach.

The root routine of the system is QUACON, which performs the LC's, calls the routine for IC's (INTERP), makes the first table of results and calls the decision making routine (DMA).

INTERP first calls CONDIT to check whether conditions for interpolation are fulfilled in each point. Sometimes CONDIT has to call MINCON to see if the nearest point is closer than 1 grid unit. PAINTE is the routine called by INTERP to calculate the value of the interpolation function in each point.

There are some auxiliar routines, like PLOTIE, which plots the observation points in an x-y diagram. MINCON is also used as an auxiliar routine just to look for the closest point to the observation point. WFILE11 opens and closes input and output files and INCDATE is used to change the date to 3 or 6 hours earlier to look for past observations.

Source \ Check	1	2	3	4	5	6	Type
PLAT	NC	NC	NC	OK	NO	NO	b
PLAT	NO	NO	NO/NC	OK	OK	OK/NC	a
PLAT	OK	OK	OK/NC	NO	NO	NO/NC	b
PLAT	OK	NO	OK/NC	OK	NO	OK/NC	c
PLAT	OK	OK	NO	OK	OK	NO	c
PLAT	OK	NO	OK/NC	NO	OK	NO	d
PLAT	OK	NO	NO	NO	OK	OK	d
PLAT	OK	OK	NO/NC	NO	NO	OK/NC	d
PLAT	NO	OK	OK/NC	OK	NO	NO	d
PLAT	NO	OK	NO	OK	NO	OK	d
PLAT	NO	NO	OK/NC	OK	OK	NO	d
SHIP/BUOY	NC	NC	NC	OK	OK	NO	b
SHIP/BUOY	NC	NC	NC	OK	NO	OK	b
SHIP/BUOY	NC	NC	NC	OK	NO <2	NC	b
SHIP/BUOY	NO	OK	OK/NC	OK	OK	OK/NC	a
SHIP/BUOY	OK	NO	OK/NC	OK	OK	OK/NC	a
SHIP/BUOY	OK	OK	NO	OK	OK	OK	a
SHIP/BUOY	OK	OK	OK/NC	NO	NO		b
SHIP/BUOY	OK>.1	NO	NC	OK	NO	OK/NC	c
SHIP/BUOY	OK>.1	OK>.1	NO	OK	OK	NO	c

Table n°19. Summary of minimum conditions to accept the data.

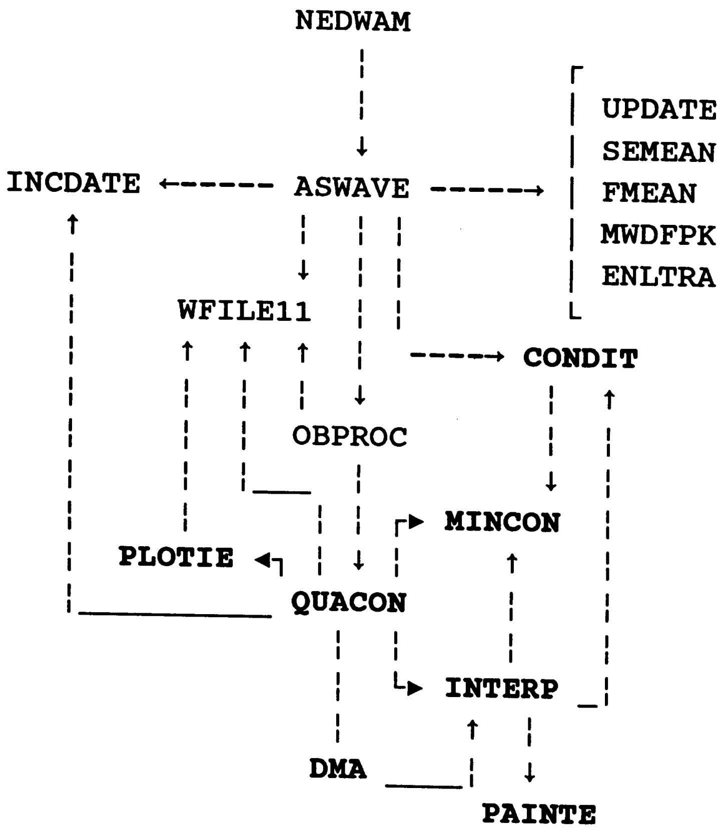

Figure 2. The data assimilation and quality control system.

Two output files are generated by the quality control for each data assimilation time and one for the whole run. They are:

NEWA_QCT_$dtg, where $dtg is the data assimilation time. This is the input file of past observations for the next data assimilation time.It contains two lists of accepted observations, one of periods and another of heights, and two lists of rejec ted observations, also one for periods and one for heights. The last table of results for each rejected observation is saved in this file. Also the two types of possible warning messages are saved here if necessary, i.e. if step 2 of stage 1 needed more than three iterations and if the whole DMA needed more than five iterations.

NEWA_PLT_$dtg contains the plot of the positions of the observations.

OUT is just to obtain more information about the decision. The table of results and the rejected observations in each step are saved here.

Some examples of these outputs can be found in the following section.

4 - APPLICATION IN EXTREME CONDITIONS

A hindcast was made for the severe storm that occured in December 1990 in the North Sea. Wave heights over 10 m were observed with quick variation in time. An example has been chosen of what would be the most extreme conditions for the quality control approach, i.e. the explosive increase in height that led to the peak of the storm. This situation corresponds to 12 December 12 GMT. Here, some large errors were found in the first guess and also it was the time when the largest temporal variations occured (van Moerkerken, in preparation).

The situation ocurred with periods on 10 December at 00 GMT is also interesting. There is a great mix of medium and short periods reported in the central area of the grid. It may be difficult for the decision when two well defined different groups of values are present, as in this case.

The plot of the period observations for 10 December 00 GMT is shown in **figure 3**. A letter is assigned to each observation for identification. Observed values are typed manually. Rejected periods are marked with asteriks.

The part of the main output file, NEWA_QCT_90121000, where rejected data are listed is shown in figure 4. Some BUOY's appear as PLAT's in this runs and so are treated by the algorithm.

The approach begins by rejecting observation "n" with T=3 sec. because of mismatch with first-guess and also with surrounding differences (check n°2) and observations (check

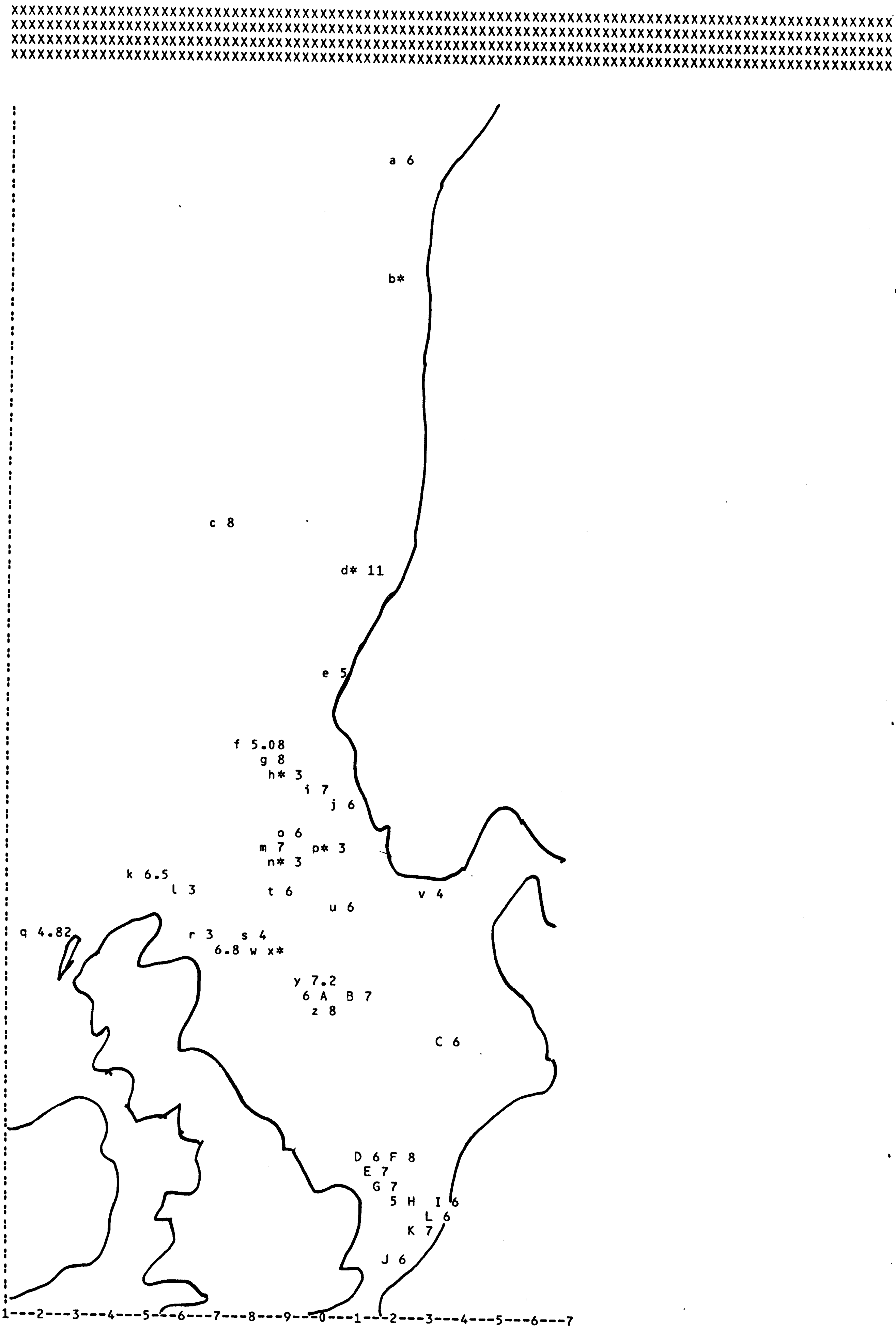

Figure 3. Period observations for 90121000

n°1). It is 4.46 sec lower than first-guess and 3 sec lower than past observation. The differences shown in **figure 4** for IC's correspond to those over the limit K E_c. In the following step "h" is rejected because now it does not fit IC's. In another step, "p" is also rejected. As all of them fail for first-guess LC, all these rejections are made in successive iterations of step 2 of the 1^{st} stage. The warning of degradation of IC's appears because it was done more than three times, but the approach is correct and it is reasonable that IC's are degraded in this case.

The situation of 12 December 12 GMT can be visualized from figures 5 and 6. **Figure 5** is the first-guess field from Nedwam and in **figure 6** the height observations are plotted. The scale is the same in both pictures.

Rejected observations are listed in **figure 7** and they can also be found as the marked points in figure 6.

Several observations are rejected in the 1^{st} step of the 1^{st} stage: "f" point, with H_c= 2.5 m is isolated, but it is 4.42 m lower than the first-guess, as it can be seen in figure 7, so it is rejected; "k", with H_c= 19.8 m is 9.52 m higher than the first-guess, 9.30 m higher than past observation and it does not fit IC's; "q", with H_c= 18.68 m has a reported swell of 18 m that exceeds the suspicion limit of check n°4, but it would be rejected anyway because of first-guess and IC's; "x" is a buoy that appears as a platform and, appart from being 6 m below the first-guess, it is suspected by check n°4 for having a reported period of 2 sec.; "C" does not pass anyone of the IC's and the first-guess is in agreement with the other observations, but perhaps the problem is that it is in an area of rapid depth decrease.

```
REJECTED PERIODS
b Y5KW999.00   NC  0.00  NC  0.00  NC  0.00  NC  0.00  NC  0.00  NC  0.00
x GBXW999.00   NC  0.00  NC  0.00  NC  0.00  NC  0.00  NC  0.00  NC  0.00
d JXCX 11.00   NO  1.10  NO  1.85  NC  0.00  OK -9.00  NO  4.18  NC  0.00
n PLAT  3.00   NO  0.51  NO  0.24  OK -0.37  OK-17.00  NO -4.46  OK -3.00
h GBRH  3.00   OK -0.12  NO  0.01  NO  0.36  OK-17.00  NO -4.65  NO -5.00
p LHDO  3.00   NO  0.07  NO  0.04  OK -3.76  OK-17.00  NO -4.83  OK -1.00
*WARNING* Data rejected in stage 1 result in degrading IC
```

Figure 4. List of rejected period observations for 10 December 00 GMT as can be found in the QCT file.

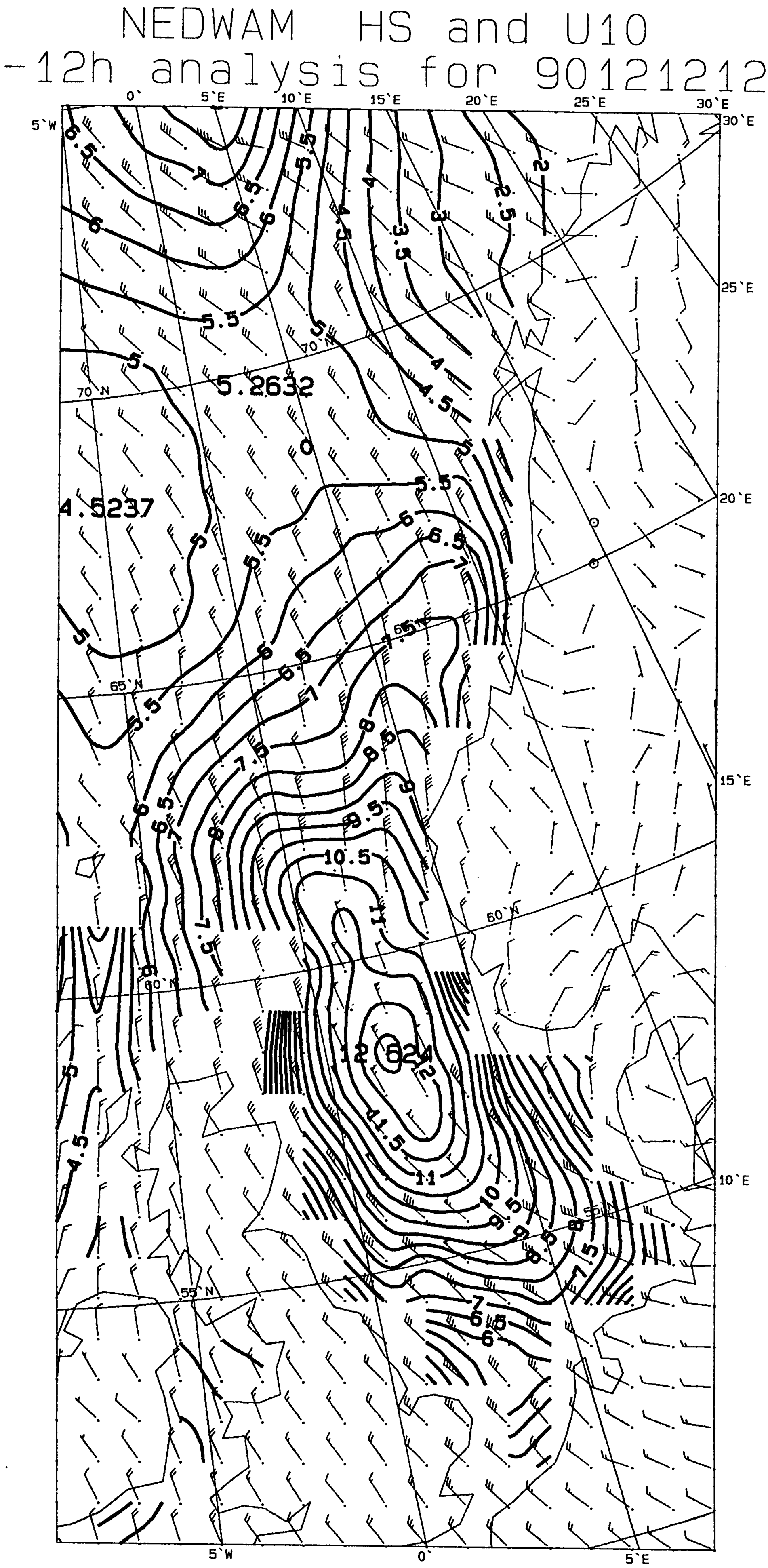

Figure 5. Nedwam first-guess field for 90121212

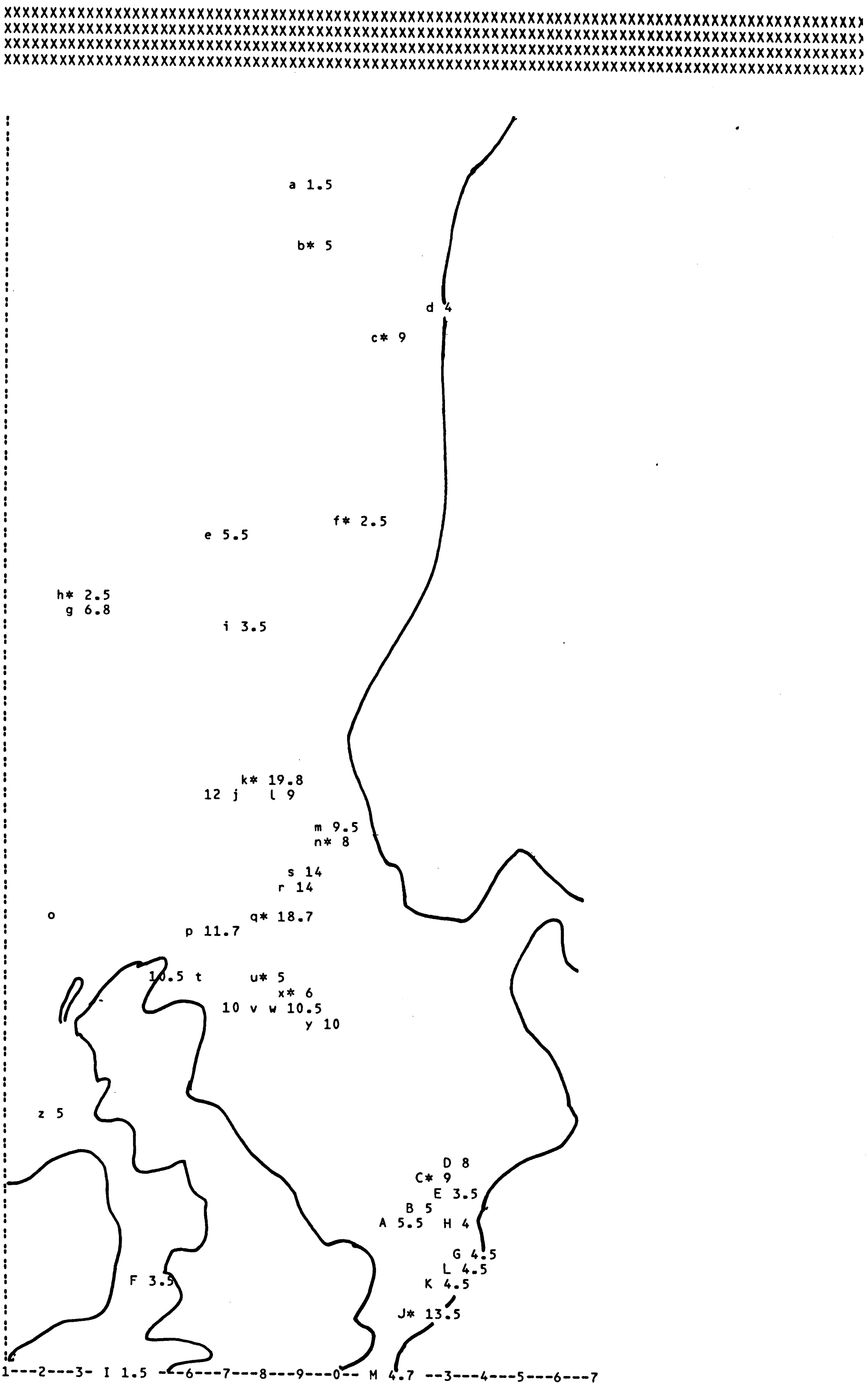

Figure 6. Wave-height observations for 90121212

Observation "c" is 5.47 m over the first-guess and the latter is in accordance with surrounding observations. Nevertheless, "b" is rejected in the last step only because of IC's.

It can be thought that "u" has been rejected because "x" had been rejected before, but even then "u" did not pass IC's. It is rejected afterwards because it has only one wrong LC.

"g" is preferred to "h" because it fits first-guess and "h" does not.

"n" does not match the first-guess, neither interpolation of differences and the interpolation of observations is very close to the limit of rejection.

In both situations we can see that observations have been rejected in general with strong arguments and that the remaining data are all consistent for the analysis.

REJECTED HEIGHTS

o	UWGB	999.00	NC	0.00	NC	0.00	NC	0.00	NC	0.00	NC	0.00	NC	0.00
f	URJP	2.50	NC	0.00	NC	0.00	NC	0.00	OK	-17.50	NO	-4.42	NC	0.00
k	GCDF	19.80	NO	2.20	NO	4.67	NO	2.24	OK	-3.50	NO	9.52	NO	9.30
q	GBRH	18.68	NO	0.34	NO	0.77	NC	0.00	NO	0.50	NO	7.23	NC	0.00
x	PLAT	6.00	NO	2.49	NO	4.47	NC	0.00	NO	99.00	NO	-6.07	NC	0.00
C	PLAT	9.00	NO	1.16	NO	1.42	NO	2.82	OK	-8.50	NO	3.03	NO	3.50
c	EVKA	9.00	NO	1.33	NO	1.03	NC	0.00	OK	-11.00	NO	5.47	NC	0.00
u	PLAT	5.00	NO	4.24	NO	6.18	NC	0.00	OK	-12.50	NO	-6.04	NC	0.00
J	PLAT	13.50	NO	7.38	NO	7.49	NC	0.00	OK	-4.00	NO	9.27	NC	0.00
h	UUWI	2.50	NO	1.54	NO	1.14	NC	0.00	OK	-17.50	NO	-2.42	NC	0.00
n	ULYS	8.00	OK	-0.03	NO	0.64	OK	-3.73	OK	-12.00	NO	-2.89	OK	-1.00
b	ESIE	5.00	NO	0.92	NO	0.65	NC	0.00	OK	-15.00	OK	1.34	NC	0.00

Figure 7. List of rejected height observations for 12 December 12 GMT as in the QCT file.

5 - CONCLUSIONS

The low reliability of wave data, particularly of the visually observed ones, is a well known problem. The necessity of a good quality control for any intent of wave data assimilation is not a subject for discussion. It is evident as well in the examples shown in the previous section.

The quality control presented here is appropiate for real-time use and does not need any human intervention. Outputs are presented in a clear manner to be easily checked by forecasters and provide wide information about the basis on which decisions have been taken for each rejected observation.

Up to now, experience with the method indicates that it is good for application in the North Sea where observations are abundant. It performs well, even in extreme conditions, although it can be improved in several aspects. For example, statistics in reference to model errors, observational errors, temporal differences and spatial correlations can be included; control of position for ships and buoys; depth considerations; no-interpolation through land points; rehabilitation of rejected data; repetition of IC's whenever an observation is rejected; inclusion of more possibilities and more knowledge about data in the decision, for instance: if a ship reports a short period where there are longer periods or swell, it is possible that the ship has not reported swell and so, the short period may be wrong. Anyway, experience with the method is probably the best source of knowledge to improve it.

Hierarchial level of the checks is not strict, as there are IC's together with LC's, but it is very difficult to apply any kind of cross-checks between wave data, as it can be done in the atmosphere, for example with hydrostatic balance.

Extension of the method to regions of more sparse data is doubtful, but some application to remote sensed data can be tried, for example, in South Atlantic.

6 - REFERENCES

Burgers G., 1990. "A Guide to the Nedwam Wave Model", KNMI Scientific Report WR-90-04, De Bilt.

Burgers G., Quanduo G. and de las Heras M.M.,1990. "Towards a wave-data assimilation system for a regional ocean-wave model", KNMI memorandum 00-90-20. [unpubl. manuscript]

Collins, W.G. and L.S. Gandin, 1990. "Comprehensive quality control at the National Meteorological Center", Mon. Wea. Rev. 118, 2752-2767.

Gandin, L.S., 1988. "Complex quality control of meteorological observations", Mon. Wea. Rev., 116, 1137-1156.

Gandin L.S., 1989. "Comprehensive Quality Control", ECMWF Seminar Proceedings, Data Assimilation and the Use of Satellite Data, 5-9 September 1988, Vol. I, Reading.

Stam M., 1989. "Revised Quality Control Procedures of Marine Climatological Data Collected from Dutch Voluntary Observing Ships and Bureau Marine Affairs", memorandum KNMI Afdeling Klimatologische Dienst KD-89-01. [unpubl. manuscript]

van Moerkerken R., 1991. "LAM and NEDWAM Statistics over the Period October 1990 - April 1991", KNMI, De Bilt. (Technical reports; TR-137).

van Moerkerken R., 1990. "LAM and NEDWAM Statistics over the Period October 1989 - December 1989", KNMI memorandum 00-90-11. [unpubl. manuscript]

Wilkerson J.C. and Earle M.D., 1990. "A Study of Differences between Environmental Reports by Ships in the Voluntary Observing Program and Measurements from NOAA Buoys", J. Geophys. Res. Vol.95, N° C3, 3373-3385.

WMO, 1988. "Guide to Wave Analysis and Forecasting", WMO N° 702, Geneva.

ACKNOWLEDGEMENTS

The author wants to thank Gerrit Burgers for his guidance, the revision of the programs and the text and his useful suggestions and Evert Bouws for retrieving the data.