Real-Time Quality Control of Wave Observations in the North Sea

MARÍA PAULA ETALA* AND GERRIT BURGERS

Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

(Manuscript received 12 July 1993, in final form 22 March 1994)

ABSTRACT

The use of ocean wave data in new data assimilation techniques prompted the development of a real-time quality control system for wave height and wave period observations. Over the North Sea, a relatively large number of wave observations, as well as a reliable wave model, are available. Therefore, a variety of tests can be made, allowing for a quite sophisticated wave-data quality control system.

The system uses some of the ideas of Gandin's comprehensive quality control. In particular, an ensemble of single-variate tests is used instead of one large multivariate test, and data are not rejected and thrown away in successive tests but can be rehabilitated until the very end.

Three months of data have been used to test the system. The decisions of the system are found to agree with manual analyses. The statistical properties of the accepted and rejected data are checked. After quality control, the residuals of the quality control tests have almost Gaussian distributions.

1. Introduction

Ocean wave forecast models are driven by surface wind or wind stress fields. Until recently, they did not make use of wave observations to fix an initial model state. Because the dynamics of surface waves is such that initial errors decay rather than amplify with time, it is possible to obtain good initial wave states by spinning up the wave model with wind fields from wind-model analyses. Currently, however, techniques for the assimilation of wave height measurements from satellites have been tested successfully (Janssen et al. 1989; Stratton 1990).

Over the North Sea, the density of conventional surface observations of waves is much higher than the density of *ERS-1* satellite observations. So for a regional wave model of the North Sea it is sensible to consider conventional wave observations. In this paper only wave height and wave period observations are considered.

Although some observations are obtained automatically by instruments, the majority consists of visually obtained human estimates. The often low reliability of visually estimated wave data is a well-known problem and wave-data quality control is common practice in climatological and model verification applications. In these cases the quality control is done a posteriori and usually requires some form of human intervention. But for an operational model that assimilates wave data one needs an automatic system that works in real time. Such a system will be described in this paper.

Because of the relative abundance of wave observations on the North Sea, over 250 observations a day, it is meaningful to make interpolation checks on the wave observations. And, as the application to data assimilation implies, one can compare observations to the first guess of a wave model. So, in the North Sea, a variety of tests can be made, allowing for a relatively sophisticated wave-data quality control system.

In data assimilation, the purpose of quality control is to exclude those measurements that would have had a negative impact on the performance of the model, and to accept the ones with a positive impact. Often, one thinks of quality control as rejecting or correcting the "wrong" measurements. This can be a useful concept, but one should remember that in many situations there is no clear-cut distinction between "right" and "wrong" measurements, and that from a theoretical point of view the problem of quality control is rather subtle. In practice, most applications assume that the errors of the useful observations have a Gaussian distribution. Then the quality control should try to remove the non-Gaussian tails.

Our approach to quality control follows some of the ideas of Gandin's (1988) comprehensive quality control. Comprehensive quality control as contrasted to more conventional methods of quality control is discussed in section 2.

Basically, a comprehensive quality control system consists of two parts: a set of checks for each observation and a decision-making algorithm (DMA) that

---

[*] Permanent affiliation: Servicio Meteorológico de la Armada, Buenos Aires, Argentina.

Corresponding author address: G. Burgers, Oceanographic Research, KNMI, P.O. Box 201, 3730 AE De Bilt, the Netherlands.

rejects or accepts each observation. How these parts are implemented in the quality control system for wave heights and wave periods can be found in sections 3, 4, and 5.

A first version of the system and its test on an extreme storm in December 1990 can be found in Etala (1991). Further tests and developments have also used data of the fall of 1991. The performance of the wave-data quality control system is discussed in section 6.

2. Comprehensive versus conventional quality control

In most operational systems, the quality control of meteorological data is done at several stages that are distributed over various modules of the system [see, e.g., Hollingsworth et al. (1985)]. The checks are ordered in increasing severity. Rejected data are not passed on to the next check. Except that in some cases quality flags are set to mark suspected data, no information flows from one stage to the next. At early stages, relatively simple tests are employed, like a check against climatology. This is done to eliminate the most serious and obvious errors as soon as possible. Usually, the final stage is done together with the interpolation step of the data assimilation procedure [the idea can be traced back as far as, e.g., Gandin (1962)]. Data are required not to be too many standard deviations away from the analysis, otherwise they are rejected. Iterative methods can help to be more cautious in the rejection of data.

Another approach has been proposed and developed by Gandin and has been applied successfully to the quality control of rawinsonde data on geopotential height and temperature (Gandin 1969, 1988, 1989; Collins and Gandin 1990). This is the approach called complex or comprehensive quality control (CQC). All CQC is done in one separate module of the operational system. A CQC system consists of two ingredients. The first ingredient is a set of checks. In CQC, all checks are applied to all data. Usually, the checks are relatively simple ones, like univariate interpolation in one direction. In principle, the checks can be of very different types. The second ingredient is a DMA to reject or accept data on the basis of the results of the checks. Also in CQC there is some kind of iteration implied, since some of the checks that involve comparing different data will have to be redone if some of the data are suspicious.

One of the main motivations of Gandin for CQC was that in conventional quality control corrupted rawinsonde data are usually thrown away at an early stage although the redundancy in the data would have permitted the reconstruction of some of them. In CQC such a reconstruction is often feasible, as all data are considered together.

This was not our motivation, since for heights and periods of sea surface waves it is in general not possible to reconstruct corrupted data. We have chosen CQC mainly because it can accomodate checks of a very different nature, and one can be very flexible in the design of the DMA. On the one hand, this means one is not automatically led to use the Gaussian distributions of optimal interpolation, which may not be always adequate to discriminate between acceptable and non-acceptable data. The classical example is that of a tide gauge in a coastal basin: there *no* deviation from the previous measurement usually is an indication that something is wrong. On the other hand, in CQC it is relatively straightforward to transfer the practice of manual quality control methods into an automated algorithm. In principle one could design CQC as an expert system with a DMA that reflects the way decisions are taken by human analysts.

Note that the distinction between CQC and other quality control schemes is not an absolute one. Also in CQC, some data are rejected immediately to speed up the convergence of the method, notably those "data" that simply indicate missing data and data that have values that are unphysical and beyond hope for correction. Because of the employment of error flags and the like, also in other quality control schemes, most decisions are taken at the last step. And in principle, one could, by exploiting all the information contained in the multivariate covariance matrix, make a conventional quality control algorithm that is as sophisticated as a CQC algorithm.

3. The data

The data consist of reports on the Global Telecommunication System (GTS) of the World Meteorological Organization and of reports from the Meetnet Noordzee (MNZ), a network of measuring instruments at twelve buoys and platforms in the North Sea set up by the Dutch authorities. The MNZ data have been monitored for years at the Royal Netherlands Meteorological Institute (KNMI). We know them to be reliable. The quality of the GTS data, most of which are visual estimates, is much less uniform.

The reports are not evenly distributed in space and time. The majority of the reports originate either from the oil and gas fields between Scotland and Norway or from the south of the North Sea, between Holland and England. Each main synoptical time, at 0000, 0600, 1200, and 1800 UTC, about 50 reports are received at KNMI, some more during the day than during the night. In between, at 0300, 0900, 1500, and 2100 UTC, about 20 reports are received. The few reports that come in at other times have not been considered in this study.

The data used are the mean period T and the significant wave height H, or, in case of some of the GTS reports, the visually measured wind sea and swell period and height. From the latter, the characteristic height and a period are obtained as follows:

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

$$T = \frac{H_w^2 P_w + H_{w1}^2 P_{w1} + H_{w2}^2 P_{w2}}{H_w^2 + H_{w1}^2 + H_{w2}^2}, \tag{2}$$

where H_w, H_{w1}, H_{w2} denote the significant height of the wind sea and the first and second swell groups, respectively, and where P_w, P_{w1}, P_{w2} denote the period of the wind sea and the first and second swell groups, respectively. If the period information of a group is missing, then this has no consequence for the calculation of the characteristic wave height. But if a missing wind sea or swell height corresponds to a reported period, also the mean period cannot be calculated, since the mean period depends both on the reported periods and the reported heights. Often stations do not separate wind sea and swell groups and put everything into H_w and P_w. No conversion formulas are applied between characteristic and significant heights, neither between various observed and model mean periods. To what extent this causes difficulties we will discuss at the end of section 6.

In addition, the data often contain the surface wind speed and direction. They are used in one of the internal consistency checks (see section 4).

It is a common error that observers report the sea state twice, both as wind sea and as swell. Swell and wind directions as well as the directions of the first and second swell group must differ by more than 20°. If they do not, then the heights must differ by more than 1 m. If this condition is not met, then the corresponding swell is removed and the remaining swell or wind-sea height is set to the maximum of the two compared heights and the period to the mean of both periods, weighted in H^2.

In the quality control procedure, the data are checked against model first guesses from the Nederlands Wave Model (NEDWAM). NEDWAM (Burgers 1990), the operational North Sea wave model of KNMI, is a regional implementation of the third-generation wave model WAM (WAMDI 1988). The grid points of NEDWAM are 75 km apart. The model calculates the full 2D wave spectrum at every grid point. The quality control system uses only the model significant wave heights and mean periods.

Before entering the actual quality control, heights and periods that fail to pass some very generous limits are rejected: heights with $H < 0$ m or $H > 25$ m and periods with $T < 2$ s or $T > 30$ s. Usually this only eliminates values that correspond to missing data indicators. For saturated wind sea with no swell, $T = 2$ s corresponds to a wind speed of about 3 m s^{-1} and a wave height of about 0.25 m. The particular limit $T = 2$ s was chosen because our wave model cannot handle shorter periods. However, as wave periods in the North Sea are virtually never shorter than 2 s (see, e.g., van Moerkerken 1991), this is hardly a limitation.

TABLE 1. Local checks, interpolation checks, and their correspondances.

	IC	LC
Observation	1	4
Observation vs first guess	2	5
Observation vs previous observation	3	6

TABLE 2. Period limits as a function of height. For stations that do not report swell separately, given H in a range in the first column, T should lie within the corresponding limits in the second column. For stations that do report swell separately, given H_w in a range in the first column, P_w should lie within the corresponding limits in the third column, and given H_{w1} (H_{w2}) in a range in the first column, P_{w1} (P_{w2}) should lie within the corresponding limits in the fourth column.

H, H_w, H_{w1}, H_{w2} (m)	T (s)	P_w (s)	P_{w1}, P_{w2} (s)
0.0–0.25	0–5	0–3	0–5
0.25–0.75	0–11	0–6	0–11
0.75–1.25	0–13	0–8	0–13
1.25–1.75	0–15	0–10	0–15
1.75–2.25	0–17	0–12	0–17
2.25–3.0	0–19	0–13	0–19
3.0–6.5	3–21	3–15	0–21
6.5–10.5	6–23	6–17	0–23
10.5–15.0	8–30	8–20	0–30
15.0–20.5	8–30	10–24	0–30
20.5–25.0	8–30	12–30	0–30

4. The checks of the wave-data quality control system

The checks can be divided into two types: local checks, which only involve the observation that is being checked, and interpolation checks, which compare the observation that is being checked to other yet unchecked observations in the same area. Both types are applied to three kinds of values: the observation itself, the difference between the observation and the model first guess, and the difference between the observation and a previous observation. So there are three interpolation checks (1, 2, and 3) and three local checks (4, 5, and 6). Each local check has a corresponding interpolation check (CIC), as shown in Table 1.

Check 4 examines the internal consistency of the observations. This is the only check where periods and heights are cross-checked and where the heights and periods of the swell groups are taken into consideration. Also, it is the only check that does not result in a number, but only in a yes/no decision: suspected or not suspected. Given the period, the height should be between certain limits. Similarly, given the height, the observed wind speed should be between certain limits. For reports that do contain swell, all this is done separately for wind sea and swell. The suspicion limits are

TABLE 3. For stations that do not report swell separately, given the wind speed FF in a range in the first column, H should lie within the corresponding limits in the second column. For stations that do report swell separately, given FF in a range in the first column, H_w should lie within the corresponding limits in the third column.

FF (kt)	H (m)	H_w (m)
0–15	0.0–25	0.0–2
15–25	0.5–25	0.5–5
25–35	0.5–25	0.5–8
35–45	1.0–25	1.0–12
45–60	1.5–25	1.5–15
60–99	2.0–25	2.0–20

listed in Table 2 and Table 3. Moreover, if swell is reported, and the reported wind speed is zero, then there should be no wind sea. If the wave height does not pass this check, then this is reason to suspect the wave period too. Check 4 can be applied always.

Check 5 considers the difference between the observation and the nearest grid point of the model first-guess field. Also check 5 can be applied always.

Because the MNZ observations are more reliable than the GTS ones, we use a reliability weight s_i in the comparison of different observations. Here, $s_i = 3$ for MNZ observations, and $s_i = 1$ for GTS observations.

Check 6 tries to compare the observation to a previous one at the last main synoptical time before the current time. The previous observation should be at least as reliable as the current one, and its position should be within 75 km from the current one. So it is not always possible to apply this check. If there happen to be more candidates, the closest one is selected.

If around a given point there are enough other observations, interpolation checks can be made. Check 1 compares the observed value at a given point to the value that results from the interpolation of observed values around the point. Check 2 does the same for observed minus first-guess differences. Check 3, which does the same for observed minus previous observation differences, can only be made if both the observation in question can be checked and enough observations around it can be checked against previous ones.

The correlation length in the North Sea is of the order of a few hundred kilometers. For the unnormalized weight w_{ik} of the interpolation to position i of the value at position k ($i \neq k$), we have chosen

$$w_{ik} = \frac{s_k/s_i}{1 + d_{ik}^2}, \tag{3}$$

where d_{ik} is the distance between position i and k in grid units of 75 km. We make interpolations only when there are enough observations to be interpolated. This can be achieved in many ways, we require that the following criteria should be met. First, only points within 300 km are considered, and the sum of the unnormalized weights of those points should be larger than 0.3. Next, the circle around i is divided into eight octants, and in each octant the sum of the w_{ik} is calculated. Also the minimum distance $d_{\min} = \min(d_{ik})$ is calculated. If $d_{\min} > 150$ km, then there should be no more than three consecutive octants with total weight less than 0.1, and if 75 km $< d_{\min} < 150$ km, then there should be no more than five consecutive octants with total weight less than 0.1. If $d_{\min} < 75$ km, then there is no further restriction.

If a check is made, it yields a yes/no result that indicates whether the observation has passed the check or not. Except for check 4, this is done by comparing the observation to a reference value, which produces a *residual* for each check. The observation is considered *suspected* if the magnitude of the residual exceeds the *allowable limit* for the check. The allowable limits for the absolute values of the residuals are given by (for heights in meters and for periods in seconds):

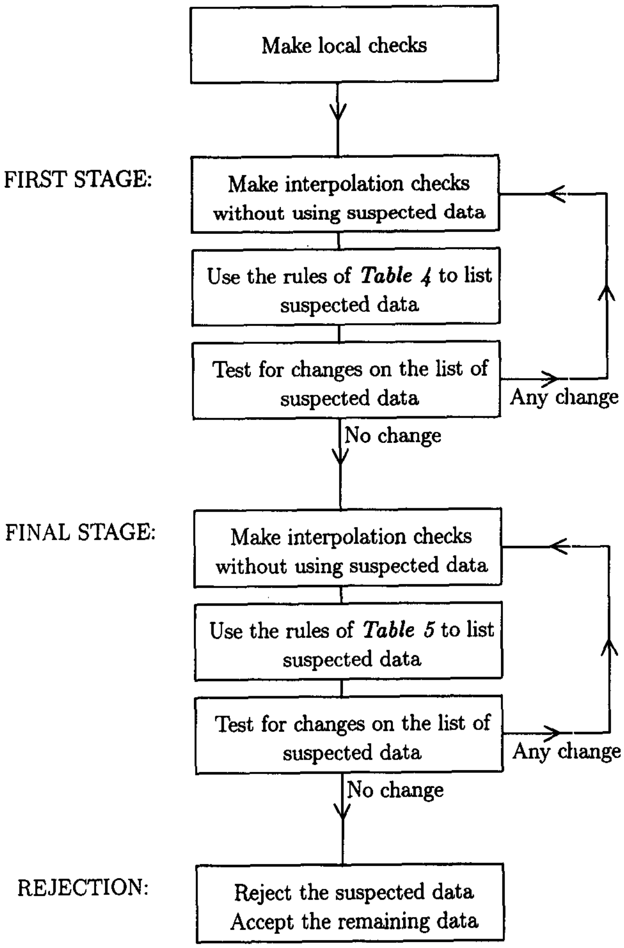

FIG. 1. Flowchart of the DMA.

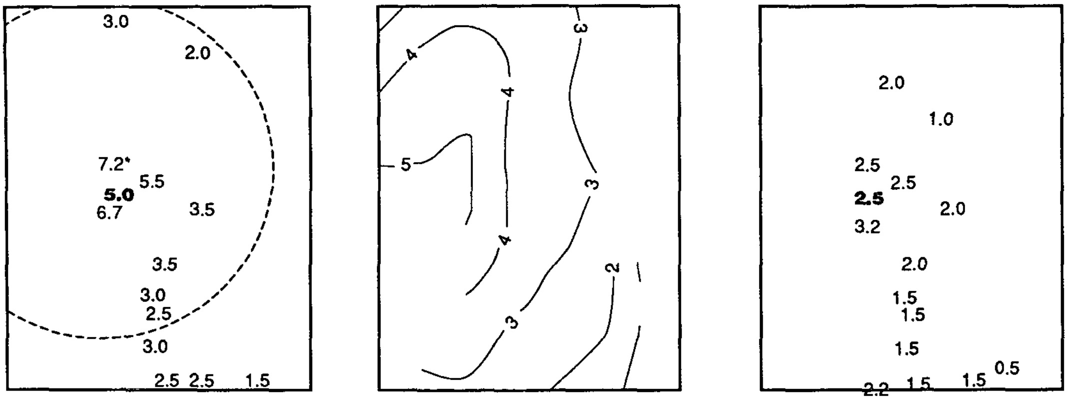

FIG. 2. (a) Wave heights (m) after the first stage of the DMA. MNZ observations are printed in boldface. The observation of 7.2 m marked by an asterisk is suspected. The dashed line is the limit for the interpolation checks of this observation. The observation passed the interpolation with the surrounding observations (check 1) and the internal consistency check (check 4), but failed all the other checks. The large 4.7-m difference with the previous observation exceeded the limit of 1.6 m for test 6 by such a large factor that the observation did meet the conditions of the last line in Table 4 and was suspected. (b) Model first guess for (a). (c) Accepted wave heights from 6 h before (a).

$$
\begin{aligned}
(\Delta H)^1_{\text{lim}} &= s_i^{1/2}(1.25 + 0.15H_{\text{obs}}) \\
(\Delta H)^2_{\text{lim}} = (\Delta H)^5_{\text{lim}} &= s_i^{1/2}(1.25 + 0.15H_{\text{fgs}}) \\
(\Delta H)^3_{\text{lim}} = (\Delta H)^6_{\text{lim}} &= s_i^{1/2}(1.25 + 0.15r_{\text{prv}}H_{\text{prv}}) \\
(\Delta T)^1_{\text{lim}} &= s_i^{1/2}[\max(1.8, 0.3T_{\text{obs}})] \\
(\Delta T)^2_{\text{lim}} = (\Delta T)^5_{\text{lim}} & \\
&= s_i^{1/2}\{\max[1.8, 0.25\min(T_{\text{obs}}, T_{\text{fgs}})]\} \\
(\Delta T)^3_{\text{lim}} = (\Delta T)^6_{\text{lim}} & \\
&= s_i^{1/2}\{\max[1.8, 0.25r_{\text{prv}}\min(T_{\text{obs}}, T_{\text{prv}})]\}. \qquad (4)
\end{aligned}
$$

The indices obs, fgs, and prv stand for observation, first guess, and previous observation, respectively. The factor r_{prv} is 0.66 (1) if the previous observation is from 3 (6) h before. As already mentioned above, the reliability weights are $s_i = 3$ for MNZ reports and $s_i = 1$ for GTS reports. The allowable limits are larger than the standard deviations of the quantities. However, while the standard deviation is *smaller* for more reliable observations, the allowable limits are *larger:* the more reliable the observation, the more probable it is that a large residual is due to an exceptional situation at sea rather than due to some exceptional circumstances in the measuring process. Clearly, the determination of the suspicion limits involves some tuning.

5. The decision-making algorithm of the wave-data quality control system

The DMA is the central part of the system. It evaluates the information given by the different checks and eventually either accepts or rejects the data. For the construction of the DMA it is particularly true that it is necessary to try it out on real data. In the present implementation, the DMA is the same for wave heights as for wave periods.

It is natural to assume that the first-guess field was wrong, if in a given area all observations fail the test

TABLE 4. Conditions for suspicion in the first stage: NO—check failed, NC—check could not be made. Blanks mean any possibility: either NO, NC, or check passed.

Check 1	Check 2	Check 3	Check 4	Check 5	Check 6
NC	NC	NC	NO	NO	
NC	NC	NC	NO		NO
NC	NC	NC		NO > $1.5(\Delta)^5_{\text{lim}}$	NO > $1.5(\Delta)^6_{\text{lim}}$
NC	NC	NC		NO > $2(\Delta)^5_{\text{lim}}$	NC
NO	NO		NO		
NO		NO	NO		
NO	NO			NO > $1.25(\Delta)^5_{\text{lim}}$	
	NO	NO		NO > $1.25(\Delta)^5_{\text{lim}}$	
NO		NO			NO > $1.25(\Delta)^6_{\text{lim}}$
	NO	NO			NO > $1.25(\Delta)^6_{\text{lim}}$

against the first guess. So, one of the guidelines is that a local check (LC) that fails can be compensated by a CIC that does not fail. Similarly, the failing of an interpolation check is taken less seriously if the corresponding LC is passed.

A problem with the interpolations is that wrong observations spoil the interpolation checks for right observations. Therefore, the DMA is iterated several times. In each iteration, the checks are applied to all observations. The DMA then singles out a number of suspected data. The suspected data are candidates for rejection. In the next iteration, the suspected data do not contribute to the interpolated values in the interpolation tests. The new set of suspected data is compared to the old one. There may be new suspected data, and some of the old suspected data may be rehabilitated. Usually, a stationary set of suspected data is reached after a few iterations.

Actually, the DMA has two stages. Both stages are iterated a number of times. Data with strong indications of being wrong are spotted at the first stage. The second stage is more critical. The second stage starts when the first stage has reached a stationary state. Finally, after the second stage has reached a stationary state, the suspected data are rejected and the remaining ones are accepted. A flowchart of the DMA is given in Fig. 1.

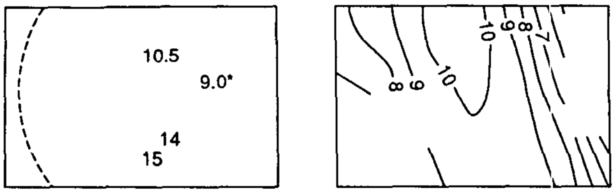

FIG. 3. (a) Wave heights (m) that passed the first stage. The observation of 9 m marked by an asterisk is suspected in the final stage. The dashed line is the limit for the interpolation checks of this observation. The difference with the interpolation from surrounding observations of −3.7 m failed check 1, and the difference with the first guess of 3.2 m failed check 5. Checks with the past (3 and 6) could not be made. Suspicion then follows from line 8 in Table 5. Probably, a more refined DMA would have noticed the sign difference in the residuals and would not have suspected this observation. (b) Model first guess for (a).

A complication for the DMA is that in general it is not possible to apply all the checks to all the data. So the conditions to suspect an observation depend on whether or not it was possible to compare it to surrounding and previous data.

The precise formulation of the DMA may be a little tortuous but we shall give a few examples to illustrate it, at least to show that intuitively it makes sense.

Most of the data suspected in the first stage are those whose mismatch is obvious, as the one marked by an

TABLE 5. Conditions for suspicion in the second stage: NO—check failed, NC—check could not be made. Blanks mean any possibility: either NO, NC, or check passed.

Check 1	Check 2	Check 3	Check 4	Check 5	Check 6
NC	NC	NC	NO	NO	
NC	NC	NC	NO		NO
NC	NC	NC		NO	$\lvert(\Delta)^5 + (\Delta)^6\rvert > 1.5(\Delta)^5_{\text{lim}}$
NC	NC	NC		$(\Delta)^5 > 0.5(\Delta)^5_{\text{lim}}$	NO, $\lvert(\Delta)^5 + (\Delta)^6\rvert > 1.5(\Delta)^5_{\text{lim}}$
NC	NC	NC		NO $> 1.5(\Delta)^5_{\text{lim}}$	NC
NO	NO	NC			
NO		NC	NO		NC
NO		NC		NO	NC
	NO	NC		NO	NC
		NC	NO	NO	NO
NO	NO	NO			
NO			NO	NO	
NO			NO		NO
NO				NO	NO
	NO		NO	NO	
	NO		NO		NO
	NO			NO	NO
		NO	NO	NO	
		NO	NO		NO
		NO		NO	NO
NO	NO		NO		
NO	NO			NO	
NO	NO				NO
NO		NO	NO		
NO		NO		NO	
NO		NO			NO
	NO	NO	NO		
	NO	NO		NO	
	NO	NO			NO

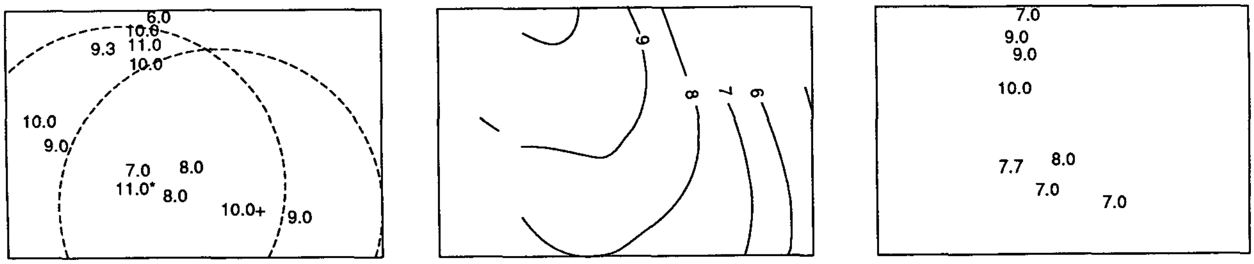

FIG. 4. (a) Wave periods (s) that passed the first stage. The data of 11 s marked by an asterisk (*) and 10 s by a cross (+) are suspected in the final stage. The dashed lines are the limits for the interpolation checks of these observations. For 11-s data, checks with the past (3 and 6) could not be made. The difference with the interpolated value (check 1) of 2.9 s was just accepted (limit 3.3 s). The mismatch of 3.2 s with the interpolation of the observation minus first-guess differences (check 2, limit 2.2 s) was not accepted and also the residual of check 5 against the first guess of 2.3 s (limit 2.2 s) was just a little too large to be acceptable. Line 9 in Table 5 marks such an observation as suspect. Observation (+) only passed the internal consistency check 4 and the interpolation with surrounding observations (check 1, the difference of 1.7 s was accepted). Check 3, interpolation of past minus present differences, could not be made. The observation failed the remaining checks 2, 5, and 6 (e.g., the difference of 3 s with the previous observation is too large) and is suspected according to line 17 of Table 5. (b) Model first guess for (a). (c) Accepted wave periods from 6 h before (a).

asterisk in Fig. 2. The conditions for suspicion in the first stage are summarized in Table 4. If interpolation checks have been made, a necessary condition for suspicion is that at least one LC–CIC pair fails. Observations that pass two of the LC–CIC pairs are never suspected. If only the three local checks are made, a necessary condition for suspicion is that two of the three fail.

The conditions for suspicion in the second stage are summarized in Table 5. They are more strict than in the first stage. Data for which no checks with the past could be made are suspected if they do not pass the first-guess checks 2 and 5, or if they do not pass another pair of checks that includes check 1, the interpolation of surrounding observations. In Fig. 3 an example is given of a suspected observation that failed the interpolation to surrounding observations. The observations in Fig. 4 failed both check 2 and 5, while the one marked by a cross also failed check 6. Figure 5 shows a suspected observation for which all tests could be made. Such data are suspected if at least three of the checks fail, including an interpolation check. Whenever interpolation checks are passed, data can be accepted no matter the differences with the first guess and/or previous observations. This is especially useful for situations with deficient first guesses and peaks of storms, where limits may not be large enough. For an example, see again Fig. 3, where data that are 4 and 5 m higher than the first guess are accepted. The most difficult situation to handle for the DMA is when only tests 4 and 5 could be made. Usually test 4 is passed. Then the observation is accepted as long as the residual of test 5 is not more than 1.5 times the allowable limit. We took a rather generous limit because these isolated observations tend to occur in the northern part of the grid where the quality of the model first guess is less high than in the south.

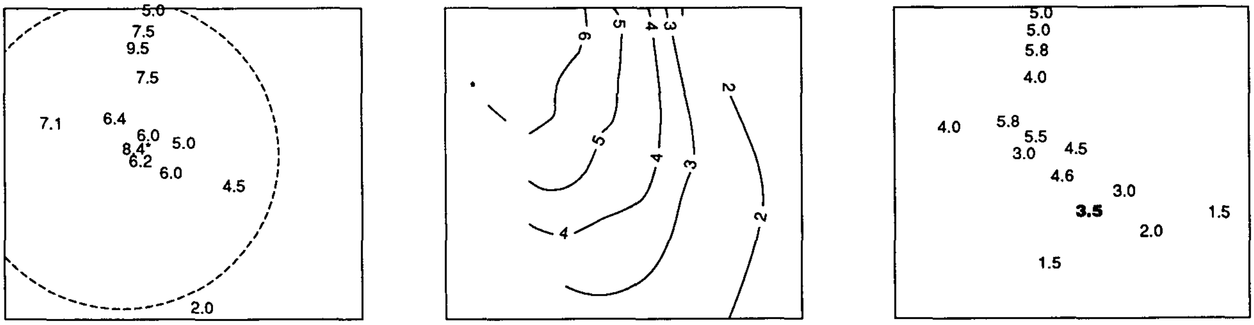

FIG. 5. (a) Wave heights (m) that passed the first stage. MNZ observations are printed in boldface. The observation marked by an asterisk is suspected in the final stage. The dashed line is the limit for the interpolation checks of this observation. All checks could be made. The internal consistency check 4 was passed, as was check 1 with surrounding observations and check 2 with surrounding observation minus first-guess differences. But the observation failed check 5 against the first guess, and also both the local check and the interpolation check against the past (checks 3 and 6). Having failed three checks including an interpolation check, this observation is suspected. (b) Model first guess for (a). (c) Accepted wave heights from 6 h before (a).

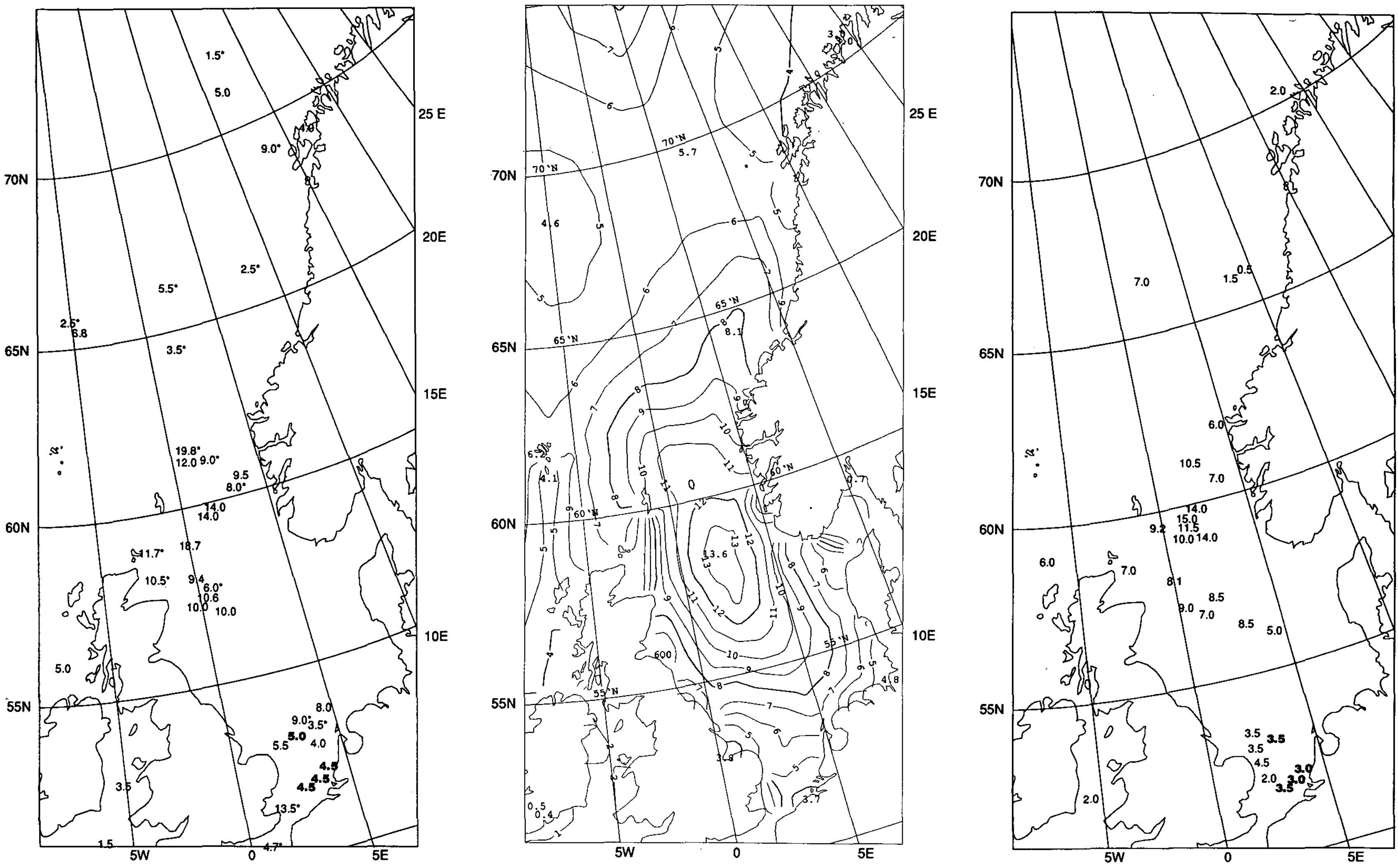

FIG. 6. (a) Wave height observations for 1200 UTC 12 December 1990. Rejected observations are marked by asterisks. MNZ observations are printed in boldface. (b) First-guess wave height field for 1200 UTC 12 December 1990. (c) Quality-controlled wave height observations for 0600 UTC 12 December 1990.

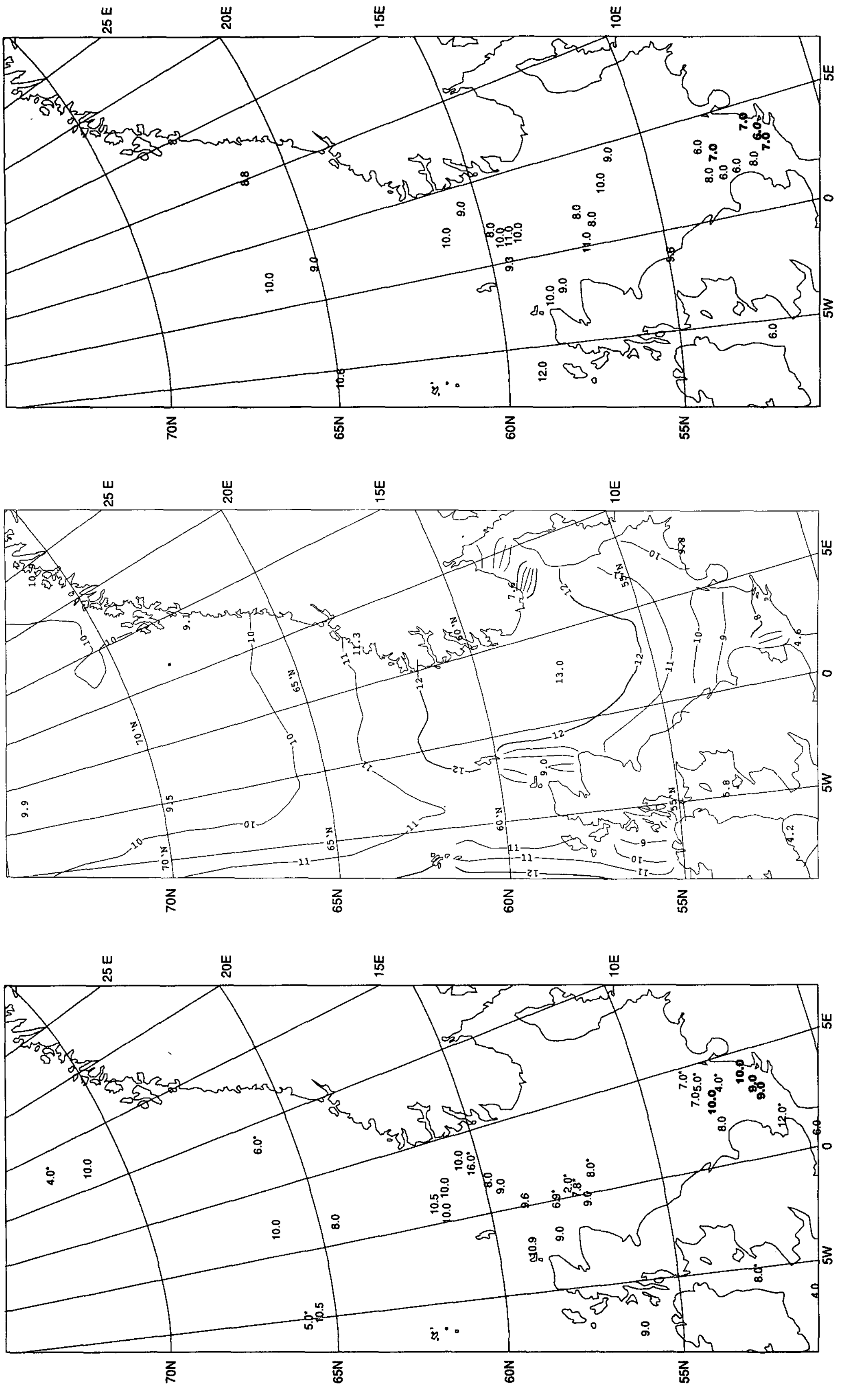

FIG. 7. (a) Wave period observations for 1200 UTC 12 December 1990. Rejected observations are marked by asterisks. MNZ observations are printed in boldface. (b) First-guess wave period field for 1200 UTC 12 December 1990. (c) Quality-controlled wave period observations for 0600 UTC 12 December 1990.

TABLE 6. Final results of the QC and the results of the individual checks.

	GTS + MNZ			MNZ		
	Total	Accepted	Rejected	Total	Accepted	Rejected
Heights						
Total	20 654	18 826	1828	3914	3914	0
1 OK	13 990	13 834	165	3375	3375	0
1 NO	1456	204	1252	0	0	0
1 NC	5208	4788	420	539	539	0
2 OK	13 932	13 869	63	3375	3375	0
2 NO	1514	169	1345	0	0	0
2 NC	5208	4788	420	539	539	0
3 OK	9405	9302	103	2864	2864	0
3 NO	521	129	392	0	0	0
3 NC	10 728	9395	1333	1050	1050	0
4 OK	20 267	18 710	1557	3911	3911	0
4 NO	387	116	271	3	3	0
5 OK	18 032	17 817	215	3914	3914	0
5 NO	2627	1014	1613	0	0	0
6 OK	12 732	12 526	206	3503	3503	0
6 NO	966	438	528	5	5	0
6 NC	6956	5862	1094	406	406	0
Periods						
Total	19 263	15 538	3725	3911	3094	7
1 OK	11 229	10 978	251	3279	3276	3
1 NO	2331	213	2118	0	0	0
1 NC	5703	4347	1356	632	628	4
2 OK	10 852	10 784	68	3270	3270	0
2 NO	2708	407	2301	9	6	3
2 NC	5703	4347	1356	632	628	4
3 OK	7450	7252	198	2731	2731	0
3 NO	566	140	426	1	0	1
3 NC	11 247	8146	3101	1179	1173	6
4 OK	19 024	15 430	3594	3897	3890	7
4 NO	279	108	131	14	14	0
5 OK	14 011	13 771	240	3888	3887	1
5 NO	5252	1767	3485	23	17	6
6 OK	10 750	10 252	498	3484	3481	3
6 NO	1037	347	690	5	2	3
6 NC	7476	4939	2537	22	21	1

6. Performance of the quality control system

We have tested the system on archived raw observations from three months, December 1990, October 1991, and December 1991. As stated before, the first guesses came from the operational KNMI regional wave model NEDWAM. The model is run four times a day, so for the main synoptical times (0000, 0600, 1200, and 1800 UTC) model analyses are used, and for the intermediate times (0300, 0900, 1500, and 2100 UTC) +3-h forecasts. The quality-controlled observations were not assimilated into the wave model.

An example of how the quality control scheme performs in extreme conditions is shown in Fig. 6. The situation chosen is from 1200 UTC 12 December 1990. The first panel shows the reported wave heights at that time, rejected ones are being marked by an asterisk. The second and third panel show the other information used by the quality control scheme: the model first-guess field for 1200 UTC, and the observations of 0600 UTC accepted by the quality control, respectively. The same information, but for periods, is given in Fig. 7. Going into the details of this situation will help to get an insight of the behavior of the quality control scheme. Hereafter we shall refer to Figs. 6 and 7, already mentioned above.

The ships reporting 9-m wave height near the northern coast of Norway and 3.5-m height at 64°30′N, 1°54′E do not fulfill the conditions required to make interpolation checks and are rejected only for great disagreement with the first guess. The 2.5-m height and 5-s period ship observation at 65°48′N, 5°W is in complete disagreement with the close observation of a 6.8-m wave height and 10.5-s period. Looking into the reports, wind seas are similar, while the latter reports a 6.5-m height, 11-s period swell. This one is an example of the very usual case of a ship failing to report the swell. The ship height observations of 19.8 m at

TABLE 7. Bias and standard deviations of the checks after quality control. Units are meters for heights and seconds for periods.

	GTS			MNZ		
	Number	Bias	σ	Number	Bias	σ
Heights						
Check 1	10 663	0.03	0.74	3375	−0.03	0.38
Check 2	10 663	0.03	0.72	3375	−0.03	0.35
Check 3	6 567	0.00	0.67	2864	0.00	0.35
Check 5	14 912	0.18	0.90	3914	0.14	0.47
Check 6	9 456	0.02	0.80	3508	0.00	0.53
Periods						
Check 1	7 915	−0.24	0.98	3276	0.24	0.81
Check 2	7 915	−0.24	0.98	3276	0.23	0.76
Check 3	4 661	0.00	0.82	2731	0.00	0.66
Check 5	11 634	−0.42	1.27	3904	0.51	0.87
Check 6	7 116	0.03	0.84	3483	0.00	0.84

61°36′N, 1°18′E and 18.7 m at 59°N, 0°24′E fail, by a large margin, the first guess and the interpolation checks. The platform located at 57°42′N, 0°54′E reports 6-m height and 2-s period, which fail the height-period cross-check, and it does not pass any other check. From earlier and later reports it seems to have malfunctioned from 12 h earlier until the end of the storm, probably damaged. The ship 8- and 9-m wave heights near the southwestern coast of Norway are rejected because they match neither the closest observations nor their differences with the first guess, although differences with previous observations are not too large. First, the 10-s period MNZ observation at 53°12′N, 3°12′E as well as the 4-s ship observation that is close to it were suspected. At a further step, the 5-s ship period nearby was also suspected and, in this way, the MNZ period was rehabilitated, although it matches neither the first guess nor previous observation. The 6.9-s wave period observed by platform at 58°6′N, 0°6′E is low when compared to the first guess and surrounding periods are in better accordance with the model results. The 7.8-s wave period observation at 57°30′N, 0°30′E passes the interpolation check, but fails both the first guess and the previous observation check. As no other checks could be made, the 8-s period near 57°N, 1°E is rejected only for its disagreement with the first guess. Such isolated data rejections only by near-the-limit discrepancy with the first guess or the past are always a little bit doubtful.

In extreme conditions a higher than average proportion of the data is rejected, although the above situation is exceptional in this respect. This is both because it is harder to make accurate observations in such situations and because in extreme cases it is more likely that the allowable limits of the tests will be exceeded. In total, some 20 000 pairs of wave height and wave period observations have been processed, more than 100 per day. About 10% of the heights and 20% of the periods were rejected, as one can see in Table 6. Observations from the MNZ stations constitute some 20% of all observations. They are virtually never rejected.

Table 6 also shows the results of the individual tests. Not always all the checks can be applied, in particular check 3 can be applied in less than half of the cases. Check 4, "the internal consistency check," is not very discriminating because almost all observations pass this check. However, for observations that cannot be subjected to interpolation checks, every piece of information helps. Our interpolation procedure seems to work all right in the sense that almost all the accepted data pass the interpolation test, but almost no rejected ones. The QC procedure does not filter exclusively observations that are in agreement with the model: a substantial number of non-MNZ observations that fail test 5, comparison with the model, are accepted as well.

Although the suspicion limits for MNZ observations are larger than for non-MNZ observations—so they are accepted more easily by the QC—the distribution of their residuals is much narrower than the distributions of the accepted non-MNZ observations, as shown in Table 7. This justifies our assumption that MNZ observations are more reliable than the average.

After quality control, the distributions of the residuals of the checks are much more Gaussian than before quality control. Both for periods and for heights, we have checked this for data grouped in classes according to the first-guess size. For each class and test we calculated the mean and standard deviation of the residuals, and we also binned the results in 1.5-m- or 1.5-s-wide bins. After quality control in all cases, the con-

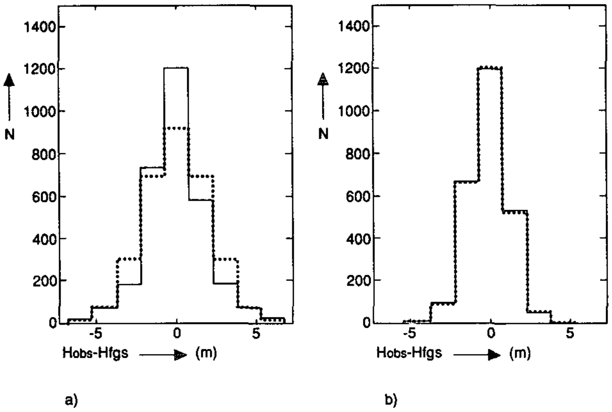

FIG. 8. Histograms of residuals of the observed minus the model first-guess height values (a) before and (b) after quality control for heights between 3.75 and 6.25 m (solid lines). The quality control accepted 2556 of the 3079 observations; 15 of the rejected observations did not even fit in the bounds (±7.25 m) of the left histogram. The bias and standard deviation before quality control are 0.0 and 2.0 m, respectively: after quality control they are −0.1 and 1.2 m. The dotted lines in the figures show how the histograms would have looked if the residuals were distributed according to Gaussian distributions with the same bias and standard deviation as the actual distributions.

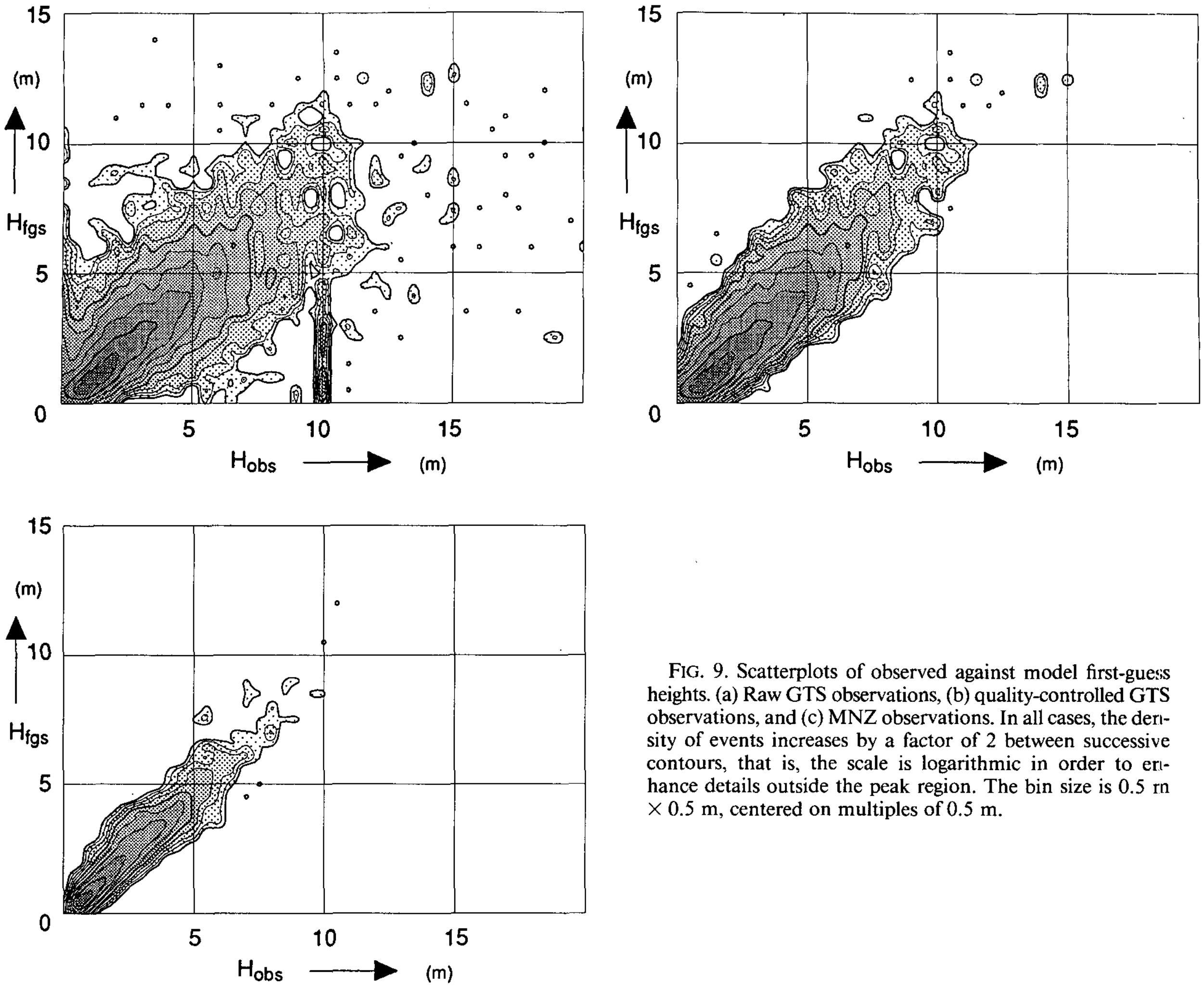

FIG. 9. Scatterplots of observed against model first-guess heights. (a) Raw GTS observations, (b) quality-controlled GTS observations, and (c) MNZ observations. In all cases, the density of events increases by a factor of 2 between successive contours, that is, the scale is logarithmic in order to enhance details outside the peak region. The bin size is 0.5 m × 0.5 m, centered on multiples of 0.5 m.

tents of the central bin never deviated more than a few percent from the contents the central bin would have had if the same number of data were distributed according to a Gaussian distribution with the same mean and standard deviation. This is illustrated in Fig. 8 for the residuals of the local check for first-guess heights between 3.75 and 6.25 m.

The next series of figures shows scatterplots of observed versus first-guess data. Instead of marking the individual data, the data have been binned. The bin size was 1 s for periods and 0.5 m for heights. The resulting field has been contoured. The scale is logarithmic, between two contour lines the density increases with a factor of 2, the outer contour being the boundary of the area with no data.

Both for heights, Fig. 9, and for periods, Fig. 10, a comparison is made between raw GTS, quality-controlled GTS, and MNZ observations. The latter observations correlate best with the model first guesses. Heights correlate better than periods. GTS heights and periods cover a much wider range than MNZ ones. The impact from the quality control on the GTS observations is clear. The well-known preference in GTS data for erroneous reports of wave heights of 10 m and wave periods of 20 s results in the conspicuous maxima in the raw data plots around these values.

After quality control, for periods there are still substantial biases for periods. MNZ periods are generally higher than first-guess periods, while GTS periods tend to be shorter. One of the reasons is that the wave model has a tendency to underestimate long wave periods. Another reason is that there are several ways to assign a wave period and a wave height to a wave spectrum. For heights the differences are small, but for periods the differences are considerable. The MNZ stations use probably the $T_{1/3}$, the mean period of the highest third of the waves. This period is larger than the mean zero-upcrossing period used by many of the GTS stations, the difference can be up to 3 s for periods around 10 s. The period used by the first guesses is the $T_{0,-1}$, which is in general slightly lower than the $T_{1/3}$. During most of the test period, the MNZ data were received through

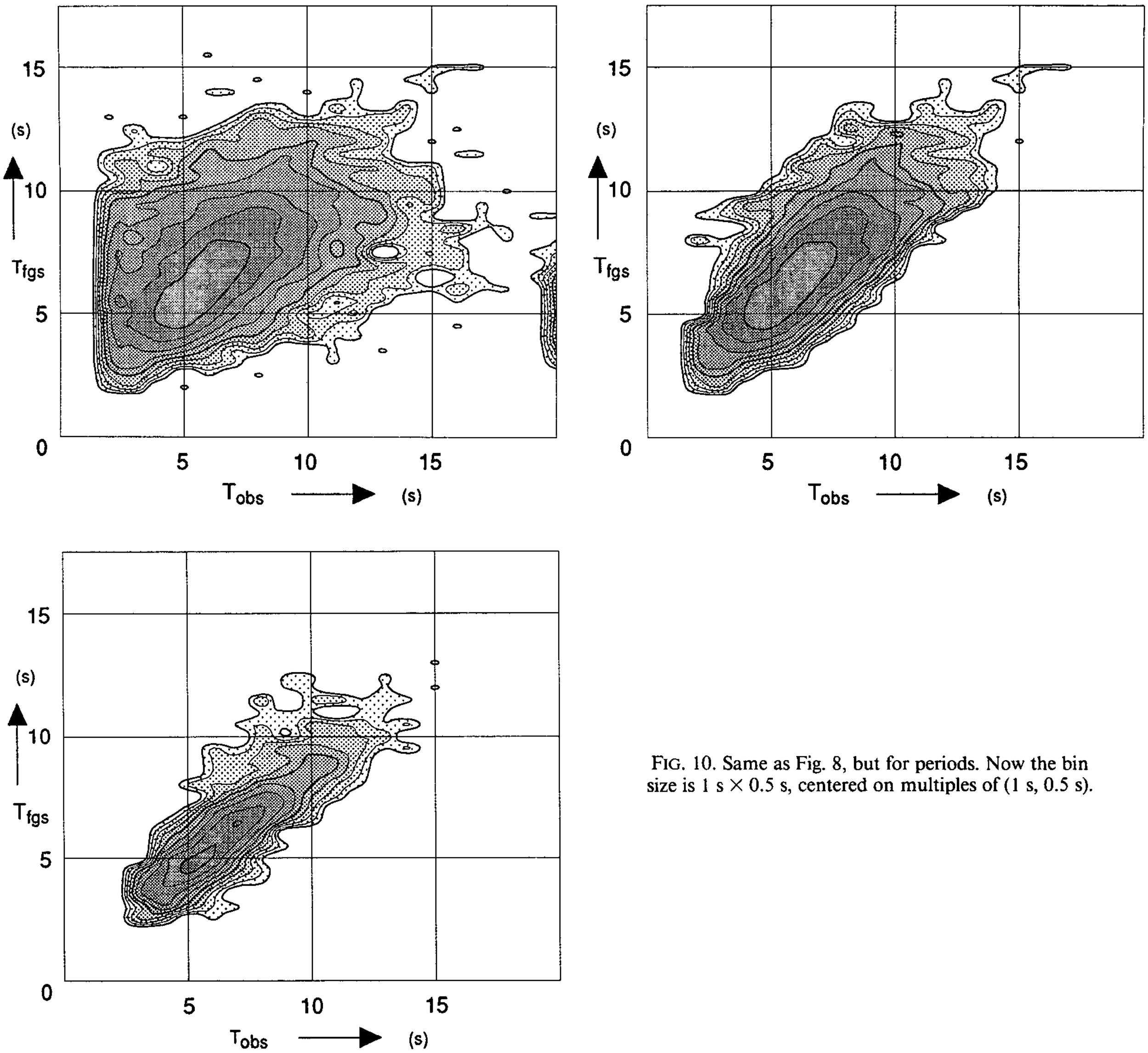

FIG. 10. Same as Fig. 8, but for periods. Now the bin size is 1 s × 0.5 s, centered on multiples of (1 s, 0.5 s).

the GTS system, but after the last week of the test period, the MNZ data, including a period that we know to be $T_{0,-1}$, are sent directly to the quality control system, thus circumventing the GTS. However, for the GTS observations the problem remains, especially if one needs quality-controlled data for, for example, data assimilation. We have not yet attempted to establish which kind of period the regular reporting GTS stations are supplying.

7. Conclusions

The low reliability of visually observed wave data is a well-known problem. This makes a good quality control of wave data necessary. We have presented a real-time wave-data quality control system that uses a combination of tests to decide whether an observation is accepted or rejected. Two types of observations are treated: significant wave heights and mean wave periods. The method needs a first-guess wave field and performs best if the observations are close enough to be compared, in which case the decision is not so dependent on the first-guess quality.

The system has been tested extensively in the North Sea, where observations are abundant. The performance has been judged in two ways. First, in a number of test cases it was checked whether the decisions of the system were not counterintuitive or clearly wrong. Second, it was verified that after quality control the errors had an almost Gaussian distribution. The data of the MNZ stations were shown to be quite reliable. They could be used for data assimilation in a wave model. In addition to observations of MNZ stations, many other observations were accepted, extending the coverage provided by the MNZ stations.

Acknowledgments. We would like to thank Prof. Lev Gandin for his comments on an earlier version of this paper, and Drs. Evert Bouws and Roel van Moerkerken for helping with retrieving the data. One of us (MPE) wishes to thank Dr. Gerbrand Komen for his invitation to visit KNMI and for his continuous encouragement during and after her stay. She is also thankful to KNMI for financial support.

REFERENCES

Burgers, G., 1990: A guide to the NEDWAM wave model. KNMI Scientific Report WR-90-04, 81 pp.

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

Etala, M. P., 1991: A real-time wave data quality control algorithm. KNMI Tech. Rep. TR-139, 29 pp.

Gandin, L. S., 1962: On operational techniques for computational analysis of meteorological fields. Main Geophysical Observatory (in Russian). *Trudy,* **124,** 18–29.

——, 1969: On automatic quality control of current meteorological information (in Russian). *Meteor. Gidrol.,* **3,** 3–13.

——, 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.,* **116,** 1137–1156.

——, 1989: Comprehensive quality control. *Proc. ECMWF Seminar on Data Assimilation and the Use of Satellite Data, Vol. 1,* Reading, UK, ECMWF, 121–149.

Hollingsworth, A., A. C. Lorenc, M. S. Tracton, K. Arpe, G. Cats, S. Uppala, and P. Kallberg, 1985: The response of numerical weather prediction systems to FGGE II-b data. Part I: Analyses. *Quart. J. Roy. Meteor. Soc.,* **111,** 1–66.

Janssen, P. A. E. M., P. Lionello, M. Reistad, and A. Hollingsworth, 1989: Hindcast and data assimilation studies with the WAM model during the Seasat period. *J. Geophys. Res.,* **94,** 973–993.

Moerkerken, R. van, 1991: LAM and NEDWAM statistics over the period October 1990–April 1991. KNMI Tech. Report TR-137, 74 pp.

Stratton, R. A., 1990: Remotely sensed data for wave forecasting. *Meteor. Mag.,* **119,** 9–17.

WAMDI Group (S. Hasselmann, K. Hasselmann, E. Bauer, P. A. E. M. Janssen, G. J. Komen, L. Bertotti, P. Lionello, A. Guillaume, V. J. Cardone, J. A. Greenwood, M. Reistad, L. Zambresky, and J. A. Ewing), 1988: The WAM model—A third generation ocean wave prediction model. *J. Phys. Oceanogr.,* **18,** 1775–1810.