(Photo: Lars Maltha Rasmussen)

New Ways to Test and Improve Methods –

Quality Assurance Activities in Breeding Bird Monitoring

Jan Blew, Dersau, FRG

Introduction

Bird monitoring has to deal with a large variability of bird numbers in time and space. To cope with this, a strong and accurate harmonization of methods is of utmost importance. The Trilateral Monitoring of Breeding Birds in the Wadden Sea in the frame of the Trilateral Monitoring and Assessment Program (TMAP) is such a large-scale program.

The TMAP attempts to improve the comparability of monitoring methods in the Wadden Sea. Trilateral monitoring guidelines have been prepared for the different parameters which require appropriate quality assurance programs to ensure that the data are suitable for the purpose for which they have been collected and that the level of accuracy is compatible with the objectives of the monitoring program.

Activities to strengthen the data quality of breeding bird numbers in the Wadden Sea has been a major task since the beginning of the trilateral breeding bird monitoring in 1993. But because there were no standardized procedures to test and elaborate field tests on methods at the national and international level innovative steps had to be found. First insights will be published here.

Since 1995, the monitoring of coastal breeding birds in the Danish-German-Dutch Wadden Sea has been carried out using standardized methods (Hälterlein et al. 1995). Within the frame of the TMAP the members of the Joint Monitoring Group of Breeding Birds (JMBB) – between 1993 and 2001 – collected data at a total of seven Quality Assurance Meetings (QAM). Those meetings had – besides expert discussions, workshops, seminars and motivation – the aim to conduct comparative counts on breeding bird census plots. These comparative counts shall help to further improve, calibrate and standardize the counting techniques.

Within the Wadden Sea, quality assurance methods for breeding bird monitoring had not been available so far. Thus, those meetings started out as pioneer work. "Learning by doing" was the motto: the field tests changed from meeting to meeting, different species, habitats and methods were used and data quality were not steady due to changing conditions, participants and objectives. Besides, no scientific method of the tests was available. Consequently, these first analyses should be – in light of their provisional character - viewed with some caution.

Material and Methods

A compilation and evaluation of the available data of seven QAM has been carried out (Blew 2003).

In Hälterlein et al. (1995) six different counting methods are described:

A: aerial counts aided by aerial photographs – for large colonies,

B: ground counts of breeding pairs – for colonies which can be viewed/scanned with a spotting scope,

C: ground counts of individual birds in flight above the colony – for colonies which cannot be viewed/scanned with a spotting scope,

D: ground counts of nests,

E: ground counts of territorial/breeding pairs (observing territorial behaviour etc.),

F: ground counts of all birds present at counting time (excluding non-breeding birds).

During the comparative counts the stochastic and the systematic errors cause variation of results. In other sources the „stochastic error" is called „sampling error", the „systematic error" is called „measurement error".

The systematic error is the difference between the mean of many counts and the actual number of birds present and is the predictable part of the error (Rappoldt et al. 1985). Sources of systematic errors can be – even within one counting method - the counting circumstances (weather, habitat and conspicuousness of the birds, disturbances etc.), time of day, the optic used, the numbers of birds present, the individual experience of the counter, time and effort taken for counting and others (e.g. Hälterlein 1996). All efforts should be taken to keep the systematic error as low as possible; in general this can be achieved by standardization of methods, a thorough description and a meticulous application of the methods.

To assess the systematic error, the scale and accuracy of the results yielded during QAM can be compared with "control numbers". Control numbers are either results of aerial or nest counts or the results of those experienced counters who are most familiar with the counting plots and in most cases carry out those counts on a regular basis.

But even if all efforts are made to minimize the systematic error, a stochastic error still exists, meaning counters will still count with a certain variation. This stochastic error is the "within-situation-variation" of the data and is described in terms of the "ratio of standard deviation over mean" – RSD, that is the proportion, the standard deviation represents from the mean (100*STD/AVG) (Rappoldt et al. 1985). Unfortunately, the stochastic and the systematic error cannot always be separated in the overall variation of a count.

During seven QAM a total of 25 comparative counts have been available for an analysis. If an analysis is carried out per species, this number increases to a total of 92 species counts.

The assessment of both systematic and stochastic error is carried out separately for counts of colony species and counts of non-colony species. „Colony-species with less than 50 individuals in the counting plot" also fall under this category.

Results

Colony species – stochastic counting error

For the 18 colony species counts, the magnitude of the individual counting error ("within-situation-variation") lies within the limits of literature values (e.g. Rappoldt et al. 1985). It turned out that there is no clear relation between the size of a colony and the stochastic error (Fig. 1 for an overview, Fig. 2 for the example Sandwich Tern). Of those counts with a rather large variation some cases can be explained: in two cases counters had to walk through the colony and could not spot it from one observation point (rhombus in Fig. 1), in some cases flying birds (after a spontaneous flight) had to be counted (black square in Fig. 1), in one case one outlier considerably raised the RSD of this count (circle in Fig. 1). Clearly, per species, only limited data are available to thoroughly test those parameters.

Colony species – potential systematic errors due to different methods

For most areas with large colonies, methods like "nest counts" and "aerial counts" are applied and best suited for yielding comparable results for monitoring purposes. For some QAM, those results are used as "control numbers". In comparison to "control numbers" the QAM results show more cases of under- than overestimation. However, the variation of the count situations, size of the colonies and habitat situation do not suffice to thoroughly analyse the results per species. Repeated counts conducted at different times of the day counts supported the already known fact, that the presence of birds during the day varies either with regard to tide or with regard to time of the day (Fig. 2).

Non-colony species – stochastic counting error

Only for Oystercatcher (*Haematopus ostralegus*), enough counts (15) are available to assess those

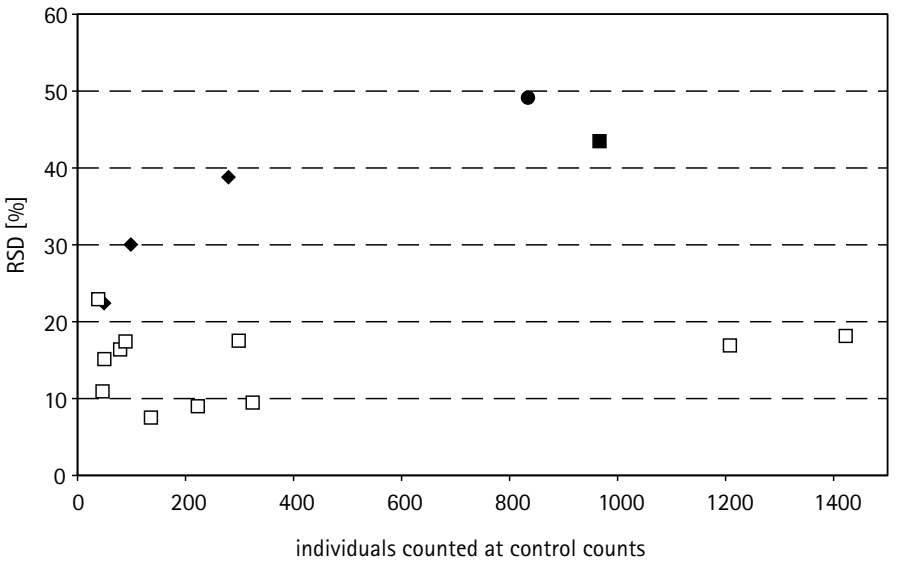

Figure 1: Counts of medium to large colonies: RSD* in relation to colony size; results of 16 species counts (open squares: 11 species counts; black squares: flying birds counted; rhombus: birds counted while walking through colony; circle: high RSD due to one outlier within the count). * RSD = ratio of standard deviation over mean (see "Material and Methods")

parameters. Here, the magnitude of the individual counting error lies within the limits of literature values (Rappoldt et al. 1985), and it is not dependent on the number of birds present; some of the high RSD might be explained by large census plots (black squares in Fig. 3). For other species like Redshank (*Tringa totanus*) (4 counts) or small breeding aggregations of Arctic Tern (*Sterna paradisaea*), the stochastic counting error is rather large, but there are not enough comparable counts at QAM to draw any conclusions.

Non-colony species – potential systematic errors due to different methods

During the QAM most of the counts at areas with non-colony species are conducted by scanning the census plot from a few points from a distance; in most cases, the "control numbers" are derived in the same manner. As with the colony counts, the QAM results of non-colony species are generally lower than the "control numbers". One reason could be that the QAM counts have in general been rather late in the season; six out of seven QAM took place in the beginning of June. Other differences between the QAM and the "control" results can be attributed to bad counting conditions or insufficient counting instructions.

At one QAM, a census plot was counted first for 15 minutes, than for another 15 minutes, and a third count was carried out after a person had walked through the plot. First of all, the results increased when more time was invested in counting. Secondly, they increased again, after a person had walked through the plot.

Discussion and Recommendations

For the monitoring of breeding birds in coastal habitats, a thorough method description exists (Hälterlein et al. 1995) and efforts are taken to internationally (trilaterally) standardize the counts (Blew 2003). Thus, for the coastal breeding birds - compared to other bird groups - reliable and fairly accurate results exist. Still, even here, many different situations exist, from large colonies in remote areas to smaller colonies, from common non-colony species to rare species. If time and effort allow, large and medium colonies are best counted from the air or with nest counts (see below). Due to habitat and location, some large or medium colonies will always be counted from the ground, and for the non-colony species ground counts will be the common method.

Over the first years of QAM, organizers and participants have learned that they are facing many different situations, sometimes lacking comparable data. Frequently, QAM were carried out well after the main counting period, trying to coordinate with the counters availability within the field season. Naturally, the group of participants (experienced counters) changed over time, their individual knowledge with regard to local species and counting plots were different. Consequently, during this first compilation of existing QAM data, it turned out that not for every situation enough data exist and for some analyses there were not sufficient data available. However, the approach was worthwhile and to further investigate and minimize both the systematic and the stochastic error, the QAM should be continued. Using the

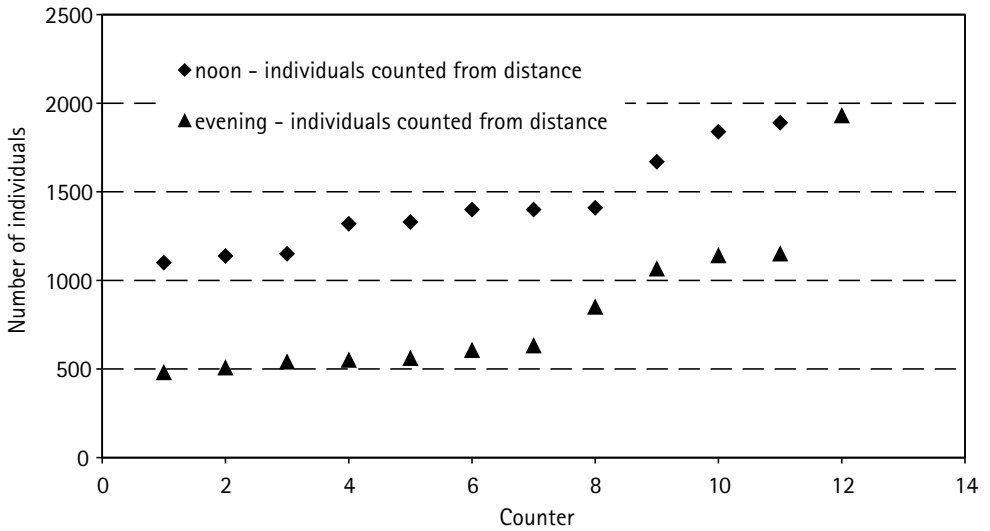

Figure 2: Sandwich Tern counted during noon and evening hours (n = 13).

experiences made during the past QAM, the following recommendations shall help to refine future QAM and shall concentrate on still open questions and species-habitat-situations.

General recommendations for future QAM

Exact instructions

In the standard variation of the results, systematic and stochastic errors cannot be separated. To exclude external influences as much as possible and to minimize the systematic error the counting conditions should be exactly explained for every QAM participant. Those are: clear instructions on how to carry out the count (observation points, time taken for counting, field protocol etc.), clear delineations of counting areas (borders of census plot, birds to be included and not included etc.), count of individuals (Method F) mandatory besides the estimation of pairs (Method E), optic used, repeated counts (if applicable).

Protocol

For each QAM count the counting conditions, the number of participants, the time of day, the status of the tide, the instructions given to the participants and any extra observations should be written into a protocol; protocols should be summarized in a short report.

"Control numbers"

Future QAM should preferably be conducted where "control numbers" exist, and those should be researched as close as possible to the QAM date. The persons who provided those numbers should participate.

For colony species, aerial and nest counts shall be used as "control numbers". For non-colony counts, the results ("counted individuals" and "estimated pairs") of the "experienced counters" most familiar with the plots shall be used. Even better as "control numbers" could be nest counts conducted just before the QAM (for method, necessary precautions and limitations of nest counts see Exo et al. 1996, Wilkens & Exo 1998).

Experienced participants

To facilitate and refine the analysis of the results, it is recommended to involve mainly "experienced counters" or classify the counters according to their own estimate, because the methods instead of the individual counter should be tested.

Recommendation for particular situations

Colony species – stochastic counting error

It is recommended, to conduct more QAM counts of medium to large colonies; thus, one should be able to draw more conclusions for additional species. Different methods of counting can be assessed for their particular stochastic counting errors. Of interest are especially those applicable in the standard monitoring. E.g. it would be of interest, whether:

- the counting of birds in flight can be improved if up to three persons synchronously count different parts of a colony;
- the numbers improve / RSD decreases, when persons are better informed about the area,
- the numbers improve / RSD decreases, when persons invest more time into the counts,

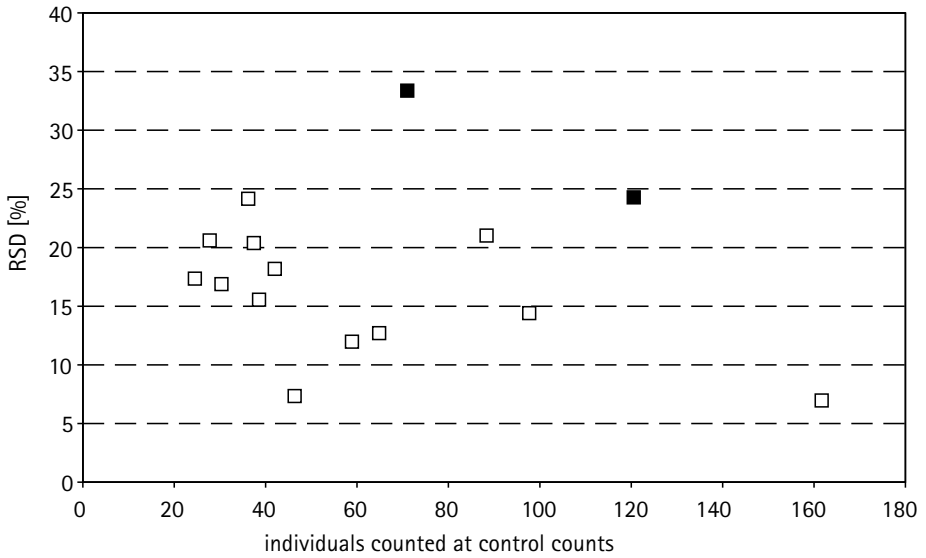

Figure 3: Counts of Oystercatcher (non-colony species): RSD* in relation to the number of individuals at the plot; results of 15 counts (black squares: results from very large census plots). * RSD = ratio of standard deviation over mean (see "Material and Methods").

- the repetition of counts does influence counting results,
- the factor to calculate "breeding pairs" from "counted individuals" is subject to counting conditions or size of the colony.

Colony species – potential systematic errors due to different methods

Large and medium colonies are best counted from the air if the conditions (habitat type, area of the colony) are suitable and certain methods are followed. Since neither aerial nor nest counts are subject of QAM, it is recommended that, for both those methods, the counters invest some time and effort in testing and examining their own results in regular intervals.

For colony ground counts recommendations for future QAM are:

- include more medium size colonies in the counts, from which "nest" or "aerial" counts are available, preferably as close as possible to the QAM date;
- test the factor (to be multiplied with the "counted individuals") applied in the standard methods, using nest counts for the estimation of the breeding pair numbers;
- analyse whether cases of under- or overestimations can be attributed to certain conditions (habitat, visibility, size of colony, species).

Non-colony species – stochastic counting error

In general, the recommendations are the same as those for "Colony species – stochastic counting error" (see above). More QAM are needed to assess counts for species other than Oystercatcher. In future QAM, it should be tested whether the RSD can be narrowed down by improving the counting instructions (see above).

Non-colony species – potential systematic errors due to different methods

From the discussions during the QAM, it seems that – with regard to counting plots with non-colony species - even the experienced counters do not always have the same opinion of which birds to include into a count. Thus, frequently birds at the border of the counting plots, especially towards the sea (mudflats) are not counted by some and counted by others, leading to high variance in the results. To minimize this systematic error, in particular the definition of the census area as well as which birds to include into the count need to be stated more precisely.

Further recommendations

Here I will briefly mention some topics which came up during the analyses and discussions.

- QAM results for the less numerous or less conspicuous species such as Ringed Plover, Redshank or ducks show large variations. Future QAM could take this topic into account by instructing the participants to pay equal attention to numerous and less numerous species. Extra counting efforts could be initiated such that after counting the numerous species an extra round is carried out for counting less numerous species.
- In many coastal areas and salt marshes the vegetation is growing higher because grazing by cattle or sheep has been decreased or

phased out completely. For most species, but especially for Redshank or duck species, numbers are grossly underestimated if the plot is just spotted from the dyke. In future QAM, one person could walk through the plot while the participants keep counting from outside the plot. Clearly, aspects of protection of birds, nests and juveniles should be obeyed. These data could test, whether walking through the plot increases counting results and whether the RSD increases due to confusion of counted and non-counted birds.

- Clearly, the "number of breeding pairs" is the number of interest for monitoring efforts. Counters are asked to estimate the "number of breeding pairs" using certain rules (Method E in Hälterlein et al. 1995). The question is: does the "estimation of pairs" introduce more variation into the data? So far QAM results show, that the values of "estimated pairs" do not have a higher variation than the values of "counted individuals". Future QAM data should help to further investigate this hypothesis and thus provide both numbers – "estimated pairs" and "counted individuals".

Concluding remarks

The Quality Assurance Meetings present some pioneer work in coastal breeding bird monitoring. Some meetings and experiences were necessary to become aware of the complexity of the task. The first analyses of their results presented here make evident that for the QAM standardized instructions and a thorough protocol should be kept to further improve these quality assessments.

Acknowledgements

Thanks to all organizers and participants of the Quality Assurance Meetings, spending additional free time to make those comparative counts possible and be willing to test their own counting results. Bernd Hälterlein, Lars-Maltha Rasmussen, Martin Schulze-Diekhoff and Ole Thorup did provide some additional data. Thanks to Lieuwe Dijksen, Bernd Hälterlein, Kees Koffijberg, Karsten Laursen, Lars-Maltha Rasmussen, Petra Potel, Bettina Reineking and Peter Südbeck for comments on several drafts.

References

Blew, J. (2003): Evaluation of the results of quality assurance activities of monitoring breeding bird 1993-2001 - technical report. Common Wadden Sea Secretariat, Wilhelmshaven.

Exo, K.-M., P. H. Becker, B. Hälterlein, H. Hötker, H. Scheufler, A. Stiefel, M. Stock, P. Südbeck & O. Thorup (1996): Bruterfolgsmonitoring bei Küstenvögeln. Vogelwelt, 117, 286-294.

Hälterlein, B., D. M. Fleet, H. R. Henneberg, T. Mennebäck, L. M. Rasmussen, P. Südbeck, O. Thorup & R. Vogel (1995): Anleitung zur Brutbestandserfassung von Küstenvögeln im Wattenmeerbereich. Wadden Sea Ecosystem No. 3, Common Wadden Sea Secretariat, Trilateral Monitoring and Assessment Group, Joint Monitoring Group for Breeding Birds in the Wadden Sea.

Hälterlein, B. (1996): Brutvogel-Bestände im Schleswig-Holsteinischen Wattenmeer. UBA-Texte, 76/97 (1998).

Rappoldt, C., M. Kersten & C. Smit (1985): Errors in large-scale shorebird counts. Ardea, 73, 13-24.

Wilkens, S. & K.-M. Exo (1998): Brutbestand und Dichteabhängigkeit des Bruterfolgs der Silbermöwe (Larus argentatus) auf Mellum. Journal für Ornithologie, 139, 21-36.

Jan Blew
Theenrade 2,
D-24326 Dersau
jan.blew@t-online.de