Accounting for uncertainties in ammonia emission from manure applied to grassland

P. W. Goedhart & J. F. M. Huijsmans
1Biometris, Wageningen University and Research, PO Box 16, 6700 AA Wageningen, The Netherlands, and
2Agrosystems Research, Wageningen University and Research, PO Box 16, 6700 AA Wageningen, The Netherlands

Abstract
A recent study has raised doubts about the ammonia emission reduction achieved in the Netherlands when applying manure to grassland by means of low-emission techniques such as narrow band and shallow injection. The critics claim that percentages of ammonia released to the atmosphere associated with low-emission techniques might even overlap with that from surface broadcast spreading given the large alleged experimental uncertainties in measurements. Consequently, the rationale behind the regulations to which farmers are exposed is questioned. In this study, it is shown that the alleged large uncertainties were obtained by means of an erroneous statistical method and that the real uncertainties are much smaller. It is also shown that, even when there is a large uncertainty in individual measurements, previous conclusions about differences in emission between different manure application techniques are still valid. It is further argued in this study that uncertainty in the percentage of applied ammonia emitted is implicitly taken into account in any comparative statistical analysis conducted in the past.

Keywords: Ammonia emission, surface broadcast spreading, narrow band application, shallow injection, statistics, parametric bootstrap

Introduction
Ammonia emission from field-applied manure contributes considerably to the total nitrogen released by manure management practices (Sommer & Hutchings, 2001). As ammonia has a negative impact on the environment (Erisman et al., 2011), during the past decades, many studies have been dedicated to the development of new low-emission application techniques. Comparison of these application techniques requires an experimental assessment of ammonia emission, and consequently, there have been many experiments on the field application of manure (Webb et al., 2010). Huijsmans & Schils (2009) summarized a large number of these experiments in the Netherlands and concluded that narrow band application and shallow injection result in considerably smaller ammonia emissions than surface broadcasting.

In a recent critical study, Hanekamp et al. (2017a,b) raised doubts about the emission reduction achieved in the Netherlands using low-emission techniques, such as narrow band and shallow injection. Their study fuelled a long-lasting debate (e.g. Huijsmans et al., 2016), which culminated in an official hearing by the Dutch parliamentary commission for Economic Affairs on 22 February 2017. Hanekamp et al. (2017a) claimed that large uncertainties in estimated emission percentages measured in individual experiments ‘at least would result in emission uncertainties that would perhaps overlap between the different manure application techniques. Thus, the widely reported dividing lines between emissions from different manure application techniques are likely blurred’. In addition, Hanekamp et al. (2017b) noted that the large uncertainties imply that ‘claims such as ‘74% of the ammonia is released when manure is applied by means of surface broadcast spreading’ are not only irresponsible but also incorrect’. These are serious comments that are addressed in this study by revisiting ammonia emission data obtained in 199 experiments conducted in the Netherlands, as reported by Huijsmans & Schils (2009).

This paper evaluates the method of Hanekamp et al. (2017a) for calculating measurement uncertainties of the percentage of applied ammonia emitted. It is shown here that their method is fundamentally flawed. This flaw is addressed, and a method is presented to obtain correct uncertainties.
values for the uncertainties. Following the reasoning of Hanekamp et al. (2017a), that large uncertainties blur differences between application techniques, observed emission percentages in pairwise experiments were given a large artificial uncertainty and subsequently re-analysed, using a meta-analysis, to identify whether their criticism could be valid. Furthermore, all the conducted experiments are re-analysed to obtain confidence intervals for mean emission percentages of the manure application methods. Measurement uncertainties were then considered from the perspective of biological variation. Finally, some remarks are made on the Ryden & McNeill (1984) model that is commonly used to estimate the percentage ammonia emission in individual experiments.

Materials and methods

Estimating ammonia emission from manure application

Total emission of ammonia after application of manure on a field can be measured without interfering with the emission process using micrometeorological methods. The micrometeorological mass balance method, as described by Denmead (1983) and Ryden & McNeill (1984), is a commonly used method to determine ammonia loss, see, for example, Bless et al. (1991), Misselbrook et al. (2002), Thompson & Meisinger (2004) and Hani et al. (2016). The method is concisely described by Ryden & McNeill (1984): ‘the mass balance is determined from the differences in the amount of a gas driven by the wind across the windward and leeward boundaries of an experimental area’. This method thus requires measurements of the wind speed and the concentration of the gas, in this case ammonia, at various heights at both a windward and leeward location relative to a, usually circular, field plot to which manure is applied. These measurements are commonly undertaken at short intervals just after manure application when emission rates are high, followed by longer intervals when emission rates are increasingly lower. A typical sequence has eight measurement intervals with durations of 1, 2, 3, 15, 24, 24 and 24 h, giving four full days of data in total. The mean wind speed and the concentration of gas in each period are usually measured at four, five or six heights, ranging from 15 to 350 cm.

Using the notation of Ryden & McNeill (1984), let \( u(z) \) and \( c(z) \) be the mean values of wind speed and concentration of ammonia in a particular period at height \( z \) at the leeward location. Furthermore, let \( c_B \) denote the mean background windward concentration of ammonia in the same period; \( c_B \) is assumed to be constant at different heights. The net flux of the ammonia loss from the manure is the difference in flux between the leeward and windward location, and is converted to the basis per unit land area by dividing by the fetch of the plot \( x \), or

\[
F = \frac{1}{x} \left[ \int_{z_0}^{z_p} u(z) c(z) \, dz - c_B \int_{z_0}^{z_p} u(z) \, dz \right] \tag{1}
\]

The integration limit \( z_0 \) is the height at which the wind speed falls to zero, and \( z_p \) is the height at which the concentration of ammonia has decreased to its background windward value \( c_B \). Empirical relationships between \( u \) and \( z \), and between \( c \) and \( z \), are then estimated from the measured leeward data, to obtain emission (kg) per unit area separately for each period. Finally, the emission is summed over the different periods to give the total emission in kg ammonia-N per hectare and this is usually expressed as the emission factor (EF), which is the percentage relative to the total ammoniacal nitrogen applied (TAN).

A crucial step in this procedure is the functional form of the empirical relationships \( u(z) \) and \( c(z) \), as different relationships will give different emissions. Ryden & McNeill (1984) used linear relationships employing the logarithm of the wind speed. The relationships are then given by

\[
u(z) = D \ln(z) + E \tag{2} \]
\[
c(z) = -A \ln(z) + B \tag{3}
\]

in which parameters \( D, E, A \) and \( B \) are estimated by linear regression. The integration limit \( z_0 \) is obtained by solving \( u(z_0) = 0 \), while \( z_p \) is obtained by solving \( c(z_p) = c_B \). Note that the integration limits are both a function of the regression parameters. The resulting ammonia emission in a certain period, that is employing equations (1), (2) and (3), is then given by equation (4), which is identical to equation (8) in Ryden & McNeill (1984).

\[
F = \frac{1}{x} \left[ -AD(z \ln^2 z - 2z \ln z + 2z) + (BD - AE - c_B D)(z \ln z - z) + (B - c_B)Ez \right] \tag{4}
\]

Emission is then, apart from the fetch, only a function \( F(A, B, D, E) \) of the four regression parameters. This model will be subsequently referred to as the R&M model.

Ammonia emission experiments and data

Hanekamp et al. (2017a) estimated the uncertainty of the emission percentages from eight experiments conducted in 2011 (Huijsmans & Hol, 2012). The wind speed and ammonia concentration data of these eight experiments, that large uncertainties blur differences between application techniques, observed emission percentages in pairwise experiments were given a large artificial uncertainty and subsequently re-analysed, using a meta-analysis, to identify whether their criticism could be valid. Furthermore, all the conducted experiments are re-analysed to obtain confidence intervals for mean emission percentages of the manure application methods. Measurement uncertainties were then considered from the perspective of biological variation. Finally, some remarks are made on the Ryden & McNeill (1984) model that is commonly used to estimate the percentage ammonia emission in individual experiments.

\[
F = \frac{1}{x} \left[ -AD(z \ln^2 z - 2z \ln z + 2z) + (BD - AE - c_B D)(z \ln z - z) + (B - c_B)Ez \right] \tag{4}
\]

Emission is then, apart from the fetch, only a function \( F(A, B, D, E) \) of the four regression parameters. This model will be subsequently referred to as the R&M model.

\[
F = \frac{1}{x} \left[ -AD(z \ln^2 z - 2z \ln z + 2z) + (BD - AE - c_B D)(z \ln z - z) + (B - c_B)Ez \right] \tag{4}
\]

Emission is then, apart from the fetch, only a function \( F(A, B, D, E) \) of the four regression parameters. This model will be subsequently referred to as the R&M model.

\[
F = \frac{1}{x} \left[ -AD(z \ln^2 z - 2z \ln z + 2z) + (BD - AE - c_B D)(z \ln z - z) + (B - c_B)Ez \right] \tag{4}
\]
either narrow band or shallow injection under the same weather, manure and field conditions. Cumulative ammonia loss percentages in each period, obtained with the R&M model for the pairwise experiments, are provided as Appendix S1. For the complete set of 199 experiments, the percentage of ammonia emitted in each period was again estimated using the R&M model. These estimates were then used, employing a saturation curve, to estimate the cumulative volatilization percentages, provided as Appendix S1. Note that some percentages for surface broadcasting are somewhat larger than 100 which is due to measurement and modelling errors. As the statistical analysis described below requires that the percentages are smaller than 100, such percentages were bounded to 99% in the statistical analysis. As a consequence, the mean emission percentages for surface broadcast spreading are slightly too small which does not invalidate our conclusions.

Uncertainty of estimated ammonia emission

Hanekamp et al. (2017a) remarked that the linear regression parameters \(D, E, A\) and \(B\) in the R&M model are estimated from data, and, as these estimates are uncertain, the resulting emission is also uncertain. They quantified the uncertainty for the eight 2011 data sets by calculating a 95% confidence interval for total emission over the full duration of each experiment. However, in doing so, they made two errors. The first error is that, for each separate period, they took the upper 95% confidence limit of the estimated regression parameters \(A^+, B^+, D^+\) and \(E^+\), where the superscript + denotes the upper confidence limit, and then calculated the emission \(F(A^+, B^+, D^+, E^+)\) which they claim is the upper 95% confidence limit for emission in that period. However, a confidence limit for a function of parameters cannot be obtained by applying the function to the confidence limits of the individual parameters. In short, what is required is \(F^+(A, B, D, E)\) rather than \(F(A^+, B^+, D^+, E^+)\). The second error is that the upper confidence limits for each period thus acquired were added to obtain an upper confidence limit for total emission. However, using a subscript for the different periods, what is required is \((F_1 + F_2 + \ldots + F_8)^+\) rather than \((F_1^+ + F_2^+ + \ldots + F_8^+)\). Consequently, the confidence intervals in Table 2 of Hanekamp et al. (2017a) are much too wide. Appendix S2 provides a simple theoretical example using samples from the normal distribution to illustrate this concept. An R script which implements the calculations of Hanekamp et al. (2017a) for the eight 2011 data sets is provided as Appendix S3.

Correct confidence intervals can be obtained by employing the parametric bootstrap (Hastie et al., 2008). This works as follows. First, estimate the regression parameters \(D, E, A\) and \(B\) separately for each period and save the variance–covariance matrix for the two pairs \((D, E)\) and \((A, B)\) which represent the uncertainty in the estimates. Standard linear regression theory states that each pair of estimates follows a bivariate normal distribution. This can then be used to quantify the uncertainty in total emission in the following way:

1. Simulate new parameters \(D^S\) and \(E^S\) for \(D\) and \(E\) employing the estimates and the accompanying variance–covariance matrix using the bivariate normal distribution.
2. Simulate in an analogous way new parameters \(A^S\) and \(B^S\) for the parameters \(A\) and \(B\).
3. Use these four simulated parameters \(D^S\), \(E^S\), \(A^S\) and \(B^S\) to calculate a new simulated emission \(F(A^S, B^S, D^S, E^S)\); note that, this involves new integration limits \(z_0\) and \(z_p\).
4. Do this separately for each period and sum the emissions to obtain total emission.
5. Perform steps 1–4 many times, for example 10 000 times. A 95% confidence interval for total emission is then obtained by the 2.5% and 97.5% percentiles of the simulated 10 000 emissions.

This method is applied to the eight 2011 experiments, and the resulting confidence intervals are compared to those of Hanekamp et al. (2017a). The parametric bootstrap method is implemented in the R script in Appendix S3.

Analysis of pairwise experiments with additional uncertainty

Huijsmans & Schils (2009) provided a test on the statistical significance of the difference between application techniques in the pairwise experiments, but they did not explicitly account for the uncertainty in the emission percentages per individual experiment. The basic wind speed and concentration data needed for such an analysis are unfortunately no longer available. Instead, here, a ‘what if the uncertainty in individual emission percentages is large’ analysis is carried out to see whether conclusions about differences between application methods would change. This was done for the pairwise experiments because these comprise a direct comparison, under the same conditions, of the application techniques.

Uncertainty in individual emission percentages was introduced in the following way. Each percentage was divided by 10, and the resulting number was then assumed to be generated by means of a (pseudo) binomial distribution with binomial total equal to 10. For example, an observed percentage of 40% was considered to be an observation of four coming from a binomial distribution with probability of success \(\pi\) and binomial total 10. The 95% confidence interval for the probability \(\pi\) then equals (0.12, 0.74), which gives, back-transformed to the percentage scale, the interval (12%, 74%). Likewise an observed percentage of 0% then gives a 95% interval of (0%, 31%). Clearly, this induces a large uncertainty in the
measurements. Note that the distribution is pseudo-binomial because a true binomial distribution can only generate integer values while percentages divided by 10 will generally not be integer. These pseudo-binomial data were then analysed using meta-analysis for pairwise experiments closely following van Houwelingen et al. (2002). The difference between two paired binomial observations \( x_1 \) and \( x_2 \), here representing two application techniques, both with binomial total \( N \) is usually expressed by the difference \( \Delta \) on the logit scale, also known as the log-odds ratio. The log-odds ratio \( \Delta \) and its approximate within-experiment standard error \( se(\Delta) \) are given by (see van Houwelingen et al., 2002):

\[
\Delta = \log \left( \frac{x_1}{N-x_1} \right) - \log \left( \frac{x_2}{N-x_2} \right) \tag{5}
\]

\[
se(\Delta) \approx \sqrt{\frac{1}{x_1} + \frac{1}{N-x_1} + \frac{1}{x_2} + \frac{1}{N-x_2}} = s_i \tag{6}
\]

A 95% confidence interval for each single pairwise experiment is then given by \( \Delta \pm 1.96 \, se(\Delta) \). These intervals can be combined into a single-point estimate for the log-odds ratio with an accompanying single confidence interval, and this is the purpose of meta-analysis. There are two variants of meta-analysis: analysis under homogeneity and analysis under heterogeneity. The first method assumes that the underlying log-odds ratio in every experiment is identical, or homogenous. Alternatively, the second method assumes that the log-odds ratios follow a normal distribution with some unknown mean and variance. The homogeneity assumption can be tested by means of the \( Q \)-test of Whitehead (2002, section 4.3.2). This test is, for both pairwise sets, not significant with \( P \)-values equal to 0.767 (surface broadcast spreading vs. narrow band) and 0.998 (surface broadcast spreading vs. shallow injection). So it is reasonable to assume homogeneous log-odds ratios. An estimate of the common difference \( \theta \) on the logit scale, and its standard error, is then given by (van Houwelingen et al., 2002):

\[
\theta = \frac{\sum_i \Delta_i}{\sqrt{\sum_i \frac{1}{\hat{N}_i}}} \quad \tag{7}
\]

\[
se(\theta) = \frac{1}{\sqrt{\sqrt{\sum_i \frac{1}{\hat{N}_i}}}} \quad \tag{8}
\]

This is used to obtain a confidence interval \( \theta \pm 1.96 \, se(\theta) \) for the common log-odds ratio. This method is implemented in the R script in Appendix S3.

**Analysis of all emission experiments**

The emission percentage in the 199 emission experiments conducted between 1989 and 2003 are re-analysed using quasi-logistic regression employing a binomial denominator of 100 (McCullagh & Nelder, 1989), see Appendix S3 for an R script. The purpose of this analysis is to report confidence intervals for the mean emission percentage for the three application techniques. Note that the fact that some experiments are part of pairwise experiments is neglected in this analysis.

**Results and discussion**

**Uncertainty of estimated emission**

Table 1 lists 95% confidence intervals for the emission percentage for the eight 2011 experiments according to the method of Hanekamp et al. (2017a) and according to the parametric bootstrap. As expected, the intervals derived by Hanekamp et al. (2017a) are wider than the bootstrap intervals as employed in this paper. The widths of the bootstrap intervals are shorter by a factor of three or more, with the exception of the already very short interval for experiment 5. Ryden & McNeill (1984) noted about the uncertainty that ‘Consequently, the maximum error in determinations of NH3 flux using the procedures adopted in the present study is about 10%’. However, the maximum

| Experiment | Percentage emission \( \% \) | 95% CI Hanekamp et al. (2017a) | 95% CI parametric bootstrap | Ratio of width of intervals | Maximum error percentage based on bootstrap |
|------------|--------------------------------|-------------------------------|-----------------------------|-----------------------------|---------------------------------------------|
| 1          | 13.9                           | (10.8, 24.8)                 | (12.3, 17.0)                | 3.0                         | 17                                          |
| 2          | 14.9                           | (11.7, 29.3)                 | (13.2, 18.3)                | 3.5                         | 17                                          |
| 3          | 18.7                           | (14.5, 43.4)                 | (16.0, 24.5)                | 3.4                         | 23                                          |
| 4          | 11.8                           | (9.3, 26.5)                  | (10.2, 15.6)                | 3.2                         | 23                                          |
| 5          | 10.0                           | (8.4, 14.9)                  | (9.0, 11.6)                 | 2.5                         | 13                                          |
| 6          | 16.3                           | (12.3, 96.1)                 | (13.9, 25.9)                | 6.9                         | 37                                          |
| 7          | 24.2                           | (18.7, 57.5)                 | (21.4, 30.3)                | 4.3                         | 18                                          |
| 8          | 17.3                           | (13.2, 41.2)                 | (15.3, 22.1)                | 4.1                         | 20                                          |
error percentage, defined here as half the width of the bootstrap confidence interval as a percentage of the emission percentage, ranges between 13% and 37% for the relatively low-emission percentages observed in these experiments.

Analysis of pairwise experiments with additional uncertainty

The ammonia emission percentages for the 37 pairwise experiments are depicted in Figure 1. Some of these values are based on the mean taken over two or four fields, whenever these fields are effectively replications. For all pairwise experiments, the emission percentage for surface broadcast spreading is larger than for the low-emission technique already indicating that there is a large difference between the application methods.

Added artificial large binomial uncertainty to these measurements results in the confidence intervals in Figure 2 for the difference on the logit scale between the application techniques. The intervals for the individual experiments are quite wide and frequently cross the vertical zero difference line indicating that there is not a significant difference between the application methods in some individual experiments. However, the single combined interval resulting from a meta-analysis under homogeneity, given by the bottom interval in Figure 2, is short. Moreover, the combined interval is far from zero indicating a very significant difference. So even with an added artificial large uncertainty in individual emission percentages, the conclusion regarding a significant difference between surface broadcast spreading and the two low-emission techniques stands. A binomial total of 10 was chosen because this results in 95% confidence intervals for individual emission percentages which are generally wider than the 95% confidence intervals of Hanekamp et al. (2017a) in Table 1. One might argue that there are many ways to add uncertainty to the observed percentages and that each particular choice yields its own confidence interval for the common difference $\theta$. While this is true, it is highly unlikely that one can reach the opposite conclusion that the two low-emission techniques do not reduce ammonia loss as compared to surface broadcast spreading. The pairwise experiments were already statistically analysed (see Table 5 of Huijsmans & Schils, 2009) where it was also concluded that surface broadcasting has a significantly larger emission than the two other techniques.

Analysis of all 199 emission experiments

Figure 3 shows a histogram of the ammonia emission percentage obtained in 199 experiments separately for the three manure application methods. Although there is some overlap, it is evident that, on average, surface broadcasting results in greater emission than the low-emission techniques. Table 2 gives estimates of the mean emission percentage for the three application methods, and a 95% confidence interval based on a quasi-binomial logistic regression analysis of the 199 emission experiments. Due to the large number of experiments, especially for surface broadcast spreading and shallow injection, the 95% intervals are not very wide. This implies that, again on average, the conclusion of Huijsmans & Schils (2009) that ‘74% of the ammonia is released when manure is applied by means of surface broadcast spreading’, which was criticized by Hanekamp et al. (2017b), is supported by the analysis in this paper. Moreover, differences between the three application methods are very significant with $P$-values smaller than 0.001. Additionally, circumstantial evidence of
reduced nitrogen losses has been provided in 14 experiments in which the nitrogen yield of grassland was increased much more by shallow injection than by surface broadcast application. The extent of this difference was in good agreement with the extra retention of ammonia-N (Huijsmans et al., 2016).

**Biological variation and measurement errors**

In biological and agricultural research, there is, next to measurement errors, also biological variation, which is generally larger than measurement errors. That is why in biological research, again generally, more emphasis is placed
on repeating experiments under different conditions rather than focusing on reducing the size of the measurement errors. Cox (1958) states in this context ‘An essential difference between the above (biological) experiments and many experiments in physics and chemistry is that in the latter, once the experimental technique is mastered and the apparatus working correctly, closely reproducible results are obtained’, and Fisher (1935) states ‘No isolated experiment, however significant in itself, can suffice for the experimental demonstration of any natural phenomenon’. Therefore, the golden rule in biological research is repeat, repeat and repeat again, and this is exactly why many emission experiments have been conducted in the Netherlands. In doing so, the average emission of a technique can be estimated with good precision by taking the mean over many different experiments. To formalize this, a model for a single observation \( Y \) is given by \( Y = \mu + B + E \) in which \( B \) represents biological variation with variance say \( \sigma_B^2 \), and \( E \) represents measurement error with variance say \( \sigma_E^2 \). The variance of \( Y \) is then given by \( \sigma^2 = \sigma_B^2 + \sigma_E^2 \), and the standard error of the mean of \( n \) experiments equals \( \sqrt{(\sigma_B^2 + \sigma_E^2)/n} \). This shows that both measurement error and biological variation are reduced in the same way. Moreover, the resulting variance \( \sigma^2 \) is used in \( t \)-tests or, more generally, in analysis of variance. The point is that in any previously conducted statistical analysis, the measurement error was already taken into account. Knowledge of the measurement error is thus not vital when comparing, for example surface broadcast spreading with shallow injection. Thus, the emphasis Hanekamp et al. (2017a) put on measurement error is unnecessary.

### The model of Ryden & McNeill

All the results presented above assume the Ryden & McNeill (1984) model to be sufficiently valid. Hanekamp et al. (2017a) showed that the linear regression model of concentration on the logarithm of the height, that is \( u(z) = D \ln(z) + E \), does a poor job for the eight 2011 data sets and that the R&M model overestimates emission percentages. Due to this misfit, application of the parametric bootstrap sometimes resulted in a small proportion of unrealistically large simulated emissions. While this did not affect the 95% bootstrap confidence intervals, it indicates that the linear regression model in equation (3) does not fit well. This was especially the case for experiment 6 (see Table 1), for which Hanekamp et al. (2017a) also found the widest intervals. We are currently working on an alternative to the R&M model that does not have this flaw. This may well lead to somewhat different, presumably lower, emission percentages for all three application methods without changing their order and the relative efficacy of low-emission application techniques.

### Conclusions

Revisiting ammonia emission data obtained in 199 experiments in the Netherlands clearly shows that narrow band application and shallow injection have, on average, much smaller ammonia emissions than surface broadcast spreading, even when accounting for large uncertainties. This finding agrees well with many international studies such as the ones reviewed in Sommer & Hutchings (2001) and Webb et al. (2010).

The critique of Hanekamp et al. (2017a), that large uncertainties imply blurred differences between application techniques, seems to be largely based on their erroneous wide intervals for total emission percentages in individual experiments. Employing the parametric bootstrap gives intervals that are at least a factor of three smaller. Moreover, for pairwise experiments, it is shown that, even when individual emission percentages are artificially given a large uncertainty, conclusions regarding significant differences between surface broadcast spreading on the one hand and narrow band and shallow injection on the other hand hold firm. Analysis of all 199 ammonia emission experiments revealed that confidence intervals for the mean emission percentage are quite narrow due to the large number of experiments. It is further argued that uncertainty in the emission percentage per individual experiment, or measurement error, is implicitly taken into account in any comparative statistical analysis conducted in the past.

The conclusions given above assume the Ryden & McNeill (1984) model to be sufficiently valid. As this model does not seem to fit the concentration data particularly well for data in the eight emission experiments conducted in 2011, there is room for improvement. We are currently working on an alternative model that provides a better fit.

### References

Bless, H.G., Beinbauer, R. & Sattelmacher, B. 1991. Ammonia emission from slurry applied to wheat stubble and rape in North Germany. *Journal of Agricultural Science*, **117**, 225–231.
Cox, D.R. 1958. Planning of experiments. Wiley, New York.

Denmead, O.T. 1983. Micrometeorological methods for measuring gaseous losses of nitrogen in the field. In: Gaseous loss of nitrogen from plant-soil systems (eds J.R. Freney & J.R. Simpson), pp. 133–158. Springer, Dordrecht.

Erisman, J.W., Van Grinsven, H., Grizzetti, B., Bouraoui, F., Powlsion, D., Sutton, M.A., Bleeker, A. & Reis, S. 2011. The European Nitrogen Problem in a Global Perspective. In: The European nitrogen assessment (eds M.A. Sutton, C.M. Howard, J.W. Erisman, G. Billen, A. Bleeker, P. Grennfelt, H. van Grinsven & B. Grizzetti), pp. 9–31. Cambridge University Press, Cambridge.

Fisher, R.A. 1935. The design of experiments. Oliver and Boyd, New York.

Hanekamp, J.C., Briggs, W.M. & Crok, M. 2017a. A volatile discourse – reviewing aspects of ammonia emissions, models and atmospheric concentrations in The Netherlands. Soil Use and Management, 33, 276–287.

Hanekamp, J.C., Crok, M. & Briggs, M. 2017b. Ammoniak in Nederland. Enkele kritische wetenschappelijke kanttekeningen. V-focusrapport. DOI: 10.13140/rg.2.2.10930.81608. (In Dutch).

Hani, C., Sintermann, J., Kupper, T., Jocher, M. & Neftel, A. 2016. Ammonia emission after slurry application to grassland in Switzerland. Atmospheric Environment, 125, 92–99.

Hastie, T., Tibshirani, R. & Friedman, J. 2008. The elements of statistical learning, 2nd edn. Springer, New York.

van Houwelingen, H.C., Arends, L.R. & Stijnen, T. 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine, 21, 589–624.

Huijsmans, J.F.M. & Hol, J.M.G. 2012. Ammoniakemissie bij mesttoediening en inwerken in aardappelruggen en bij mesttoediening in sleuven op niet beteeld gekoeld kleibouwland. Plant Research International, Wageningen U.R., rapport 445, pp. 22. Available at http://edepot.wur.nl/210562.

Huijsmans, J.F.M. & Schils, R.L.M. 2009. Ammonia and nitrous oxide emissions following field-application of manure: state of the art measurement in the Netherlands. International Fertiliser Society Proceedings 655, York, pp. 37. Available at http://library.wur.nl/WebQuery/wurpubs/fulltext/255892

Huijsmans, J.F.M., Schröder, J.J., Mosquera, J., Vermeulen, G.D., Ten Berge, H.F.M. & Neeteson, J.J. 2016. Ammonia emissions from cattle slurries applied to grassland: should application techniques be reconsidered? Soil Use and Management, 32(Suppl. 1), 109–116.

McCullagh, P. & Nelder, J.A. 1989. Generalized linear models, 2nd edn. Chapman and Hall, London.

Messelbrook, T.H., Smith, K.A., Johnson, R.A. & Pain, B.F. 2002. Slurry application techniques to reduce ammonia emissions: results of some UK field-scale experiments. Biosystems Engineering, 81, 313–321.

Ryden, J.C. & McNeill, J.E. 1984. Application of the micrometeorological mass balance method to the determination of ammonia loss from a grazed sward. Journal of the Science of Food and Agriculture, 35, 1297–1310.

Sommer, S.G. & Hutchings, N.J. 2001. Review: ammonia emission from field-applied manure and its reduction. European Journal of Agronomy, 15, 1–15.

Thompson, R.B. & Meisinger, J.J. 2004. Gaseous nitrogen losses and ammonia volatilization measurement following land application of cattle slurry in the mid-Atlantic region of the USA. Plant and Soil, 266, 231–246.

Webb, J., Pain, B., Bittman, S. & Morgan, J. 2010. The impacts of manure application methods on emissions of ammonia, nitrous oxide and on crop response-a review. Agriculture, Ecosystems and Environment, 137, 39–46.

Whitehead, A. 2002. Meta-analysis of controlled clinical trials. Wiley, Chichester.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Ammonia concentration and wind speed data measured at various heights and in different periods for eight ammonia emission experiments conducted in 2011 in the Netherlands. Ammonia emission data for 199 experiments and for 37 pairwise experiments conducted between 1989 and 2003 in the Netherlands.

Appendix S2. A confidence interval for a sum of parameters cannot be obtained by summation of the confidence limits of the individual parameters.

Appendix S3. R scripts to perform various statistical analyses.