Response to comments of referee #1 and #2 for "Validation and uncertainty quantification of three state-of-the-art ammonia surface exchange schemes using NH_3 flux measurements in a dune ecosystem" by T. Jongenelen, M.C. van Zanten, E. Dammers, R. Wichink Kruit, A. Hensen, L.F.G. Geers and J.W. Erisman.

We are grateful to anonymous referees #1 and #2 for their constructive comments on our manuscript, which helped us tremendously to further improve the paper. Below, we have responded to the remarks of referee #1 (indicated as RC1.) and referee #2 (indicated as RC2.) with an author comment (noted as AC). The comments of the referees are in **bold**.

Responses to referee #1

General Comments:

The authors present a comparative study of three ammonia surface exchange schemes, compared with NH_3 flux observations from a natural ecosystem. The topic is interesting, and the paper is written in a concise manner, but leaves some important details out of the main body of the text.

Thank you for the positive evaluation of our manuscript.

Specific Comments:

RC1.1: Paragraph at lines 46 – 54: The discussion of the NH_3 surface exchange schemes here is very abstract, and would benefit from including at least one illustrative figure, and/or one or more equations (for example the equation used for the compartment-specific compensation point; the parallel/in series resistances). This material is covered in the methods section as well, but should be introduced more clearly here.

AC1.1: Thank you for your suggestion. In the introduction, we included a reference to Fig. 1 where the resistance analogy and compensation point are illustrated. Additionally, we refer to Section 2.2 for further discussion on the parameterization. We added the following text:

"The NH_3 exchange schemes are illustrated in Fig. 1 and further discussed in Section 2.2."

RC1.2: Line 74 – 76: what is the temporal resolution of the sonic anemometer and the on/off-site meteorological measurements?

AC1.2: On-site, wind speed, sensible heat flux, friction velocity, and the Obukhov length are measured at 20 Hz with the Windmaster pro, Gill. The remaining meteorological variables were measured offsite with an hourly resolution at a nearby KNMI weather station. The resolution of the sonic anemometer and the off-site meteorological instruments is now mentioned in Section 2.1 as follows:

"Besides the NH_3 flux measurements, the wind speed and direction were measured with the 3D sonic anemometer Windmaster Pro, Gill. with a temporal resolution of 20 Hz at 5.15 m, from which the friction velocity u_*, and Monin-Obukhov length were inferred."

RC1.3: Section 2.2: This section discusses the parameters used to model the NH_3 fluxes in this study, but does not state how the modelling was done. I see that in the acknowledgements, thanks is given to Cor Jacobs for his assistance on the Fortran code of DEPAC 1-D, was all of the modelling done in Fortran? Did the modelling consist of setting up the system of equations from Table A-1, and calculating the net flux at each time step? Connecting with the comment for line 74 – 76, what time step was the model calculated at? The same as the fastest measurement, the slowest, or? Additionally, I believe this section would benefit from moving table A1 into the main text, and more directly explaining the parameters in the text as well. (Which may simply involve appending the appropriate equations to the existing in-text descriptions).

AC1.3: Thank you for your questions regarding the modeling, which we needed to explain more in the text. The three selected exchange schemes in this study were coded in Python. The fluxes were calculated at an hourly resolution. This information is now included in Section 2.2:

"To calculate the fluxes at Solleveld, the parameterization of the DEPAC, Massad, and Zhang scheme have been coded in Python based on the parameterization presented in van Zanten (2010), Massad et al.(2010) and Zhang et al. (2010). The modeled fluxes have an hourly temporal resolution."

Additionally, we followed your suggestion to move Table A1 and moved it to the end of Section 2.2. Moreover, we have expanded the explanation of the resistances by briefly explaining the underlying physics for each exchange pathway and adding more context on how the resistances are modeled and differ across the schemes. We have considered numbering each equation in Table 1 (previously Table A1); however, given the large number of equations in the table and the limited space available for elements in the table, we decided not to include equation numbers in the table.

RC1.4: Line 89 – 96: This paragraph might benefit from one or more equations or diagrams that summarize the relationships expressed in the text.

AC1.4: Thank you for the suggestion. This is solved in AC1.3, where we extended the explanation of the resistances and parameterization used in the schemes.

RC1.5: Line 97 – 99: I would suggest introducing the per-compartment compensation point (and as below, writing out the equation used in this study), and then extending this to the canopy compensation point, rather than the current order of introducing the canopy compensation point first and the per-compartment compensation point afterwards.

AC1.5: We agree that a bottom-up explanation of the canopy compensation point would be more beneficial. We therefore changed the order by first introducing the compartment compensation point χ_i, followed by an explanation of the canopy compensation point χ_c. Although the DEPAC, Massad, and Zhang schemes differ in the calculation of the χ_c – based on both the compensation points that are (not) included in the schemes and the configuration of R_a and R_b (in the Massad scheme R_a and R_b are not in series, see Fig. 1) – we use a general formula (Eq. 2) for χ_c which clearly describes the structure used for calculating the χ_c in exchange schemes. The following text has been added to Section 2.2:

The canopy compensation point χ_c can be interpreted as the effective NH_3 concentration in an ecosystem and is used to calculate the total flux. The formulae of χ_c in the three exchange schemes are lengthy but generally adhere to the following format (Sutton et al., 1995)*:*

$$\chi_c = \frac{\sum_i \chi_i / R_i}{\sum_i 1/R_i} \tag{2}$$

RC1.6: Line 105– 107: This might benefit from writing the equation used to calculate the canopy compensation point.

AC1.6: See AC1.5.

RC1.7: Line 117 – 118: Is Γ_s in the DEPAC scheme based on a meta analysis from Van Zanten et al (2010), or by a separate meta analysis by the authors?

AC1.7: Thank you for your comment. The Γ_s parameterization in the DEPAC scheme is based on a meta-analysis performed by Wichink Kruit et al. (2010), and has not been modified by us. Likewise, the Γ values in the Massad and Zhang scheme used in this study are based on the original formulations by Massad et al. (2010) and Zhang et al. (2010), respectively. We have now cited the work by Wichink Kruit et al. (2010) in the text. We have added the following text in Section 2.2 to clarify this:

"The stomatal emission potential Γ_s parameterization is derived from a meta-analysis of Γ_s values for multiple land use classes (Wichink Kruit et al., 2010)."

RC1.8: Line 124 – 125: Is Γ_s calculated using an empirical equation & meta-analysis done by the authors, or reported by Massad (2010)? If the author's work, more detail should be provided, otherwise add the appropriate citations.

AC1.8: The Γ_s parameterization in the Massad and Zhang scheme is according to the original formulations by Massad et al. (2010) and Zhang et al. (2010). The relevant references have been added to the text.

Additionally, given comments RC1.7 and RC1.8, we have also added the following text to Section 1, emphasizing that we are comparing the original formulations of the three exchange schemes without modification or optimization:

"*Importantly, we evaluate these schemes without altering or optimizing their parameters, ensuring that the comparison reflects the formulations as implemented in the operational models.*

RC1.9: Figure 1 (Line 141): I believe that $X_{(z0)}$, and R_{inc} were not previously defined in the text

AC1.9: Thank you for identifying this omission. The in-canopy resistance R_{inc} is now explained in Section 2.2 which has been extended. The definition of χ_{z0}, which is the modeled NH_3 concentration at height z_0 + d (Massad et al., 2010), is added to the caption of Fig. 1, as follows:

"Figure 1: Schematic of the DEPAC, Massad, and Zhang scheme. Note that at Solleveld, the DEPAC and Massad scheme do not model NH3 exchange with the soil pathway. χ_{z0} is the NH_3 concentration at height z_0 and is only calculated in the Massad scheme."

RC1.10: Line 220: Perhaps "Conversely" rather than "Contrarily"

AC1.10: Thank you, we changed "Contrarily" into "Conversely".

RC1.11: Line 196 – 200: If the timing of the DEPAC scheme closely matches the measured fluxes, but is significantly offset, while the Zhang scheme closely matches the order of magnitude of the observed fluxes, but not the diurnal cycle, could better results be obtained by using the applicable inputs for the Zhang scheme as the inputs for the DEPAC scheme (or vice versa)?

AC1.11: Thank you for this excellent question. Adjusting input parameters in the DEPAC scheme based on parameter values used in the Zhang scheme could certainly improve the agreement between the DEPAC scheme and the measurements. For example, the external resistance R_w in the Zhang scheme is significantly higher than R_w in the DEPAC scheme (see Fig. 6g and Fig. 6h), which could hint at the fact that R_w in the DEPAC scheme is too low. However, doing such modifications does not necessarily ensure that the underlying physical processes occurring at Solleveld would be represented more accurately. As mentioned in discussion sections 4.1 and 4.4, it is difficult to isolate the stomatal, external, and soil fluxes at Solleveld from each other. This would be crucial to validate (and improve) the parameterization for each separate exchange pathway of the schemes.

Since the scope of this paper is not to optimize the parameterization NH_3 exchange schemes at Solleveld, we have not included such analyses in this paper. For a study that does perform such modifications, we refer to Vendel et al. (2023) which have altered the parameterization of the DEPAC scheme to improve the alignment with the measurements. In a follow-up study, we will use more flux measurements to develop improved dry deposition parameterization.

RC1.12: Line 237: Would the DEPAC scheme result in similar results to the Massad scheme (and/or more closely match the observations) if the same parameter was used for Γ_s in each modelling scheme?

AC1.12: Thank you for this question. This also relates to the previous question RC1.11 and is for the next step in developing the deposition parameterization. A quick response would be that the Γ_s values in the Massad scheme are on average lower than the Γ_s in the DEPAC scheme, which consequently leads to a lower stomatal compensation point χ_s in the Massad scheme, which can be seen in Fig. 6d and Fig. 6e. The stomatal flux in the DEPAC scheme, which is currently positive (i.e. net emission), would decrease or even change from emission to deposition, which ultimately leads to an even stronger overestimated deposition flux.

RC1.13: Section 4.1 (line 308): Do the poorly modelled emission episodes consistently occur only/usually at or around midday/afternoon? Are there factors (e.g., time of day, meteorological variables, SO_2) that can be used to explain the difference between the minority of emission episodes that are (somewhat) properly modelled, and the majority of emission episodes that are poorly modelled?

AC1.13: Thanks for the suggestion regarding the timing of the emissions and whether these can be related to meteorological factors. We have performed a preliminary analysis to investigate whether specific meteorological variables or time-of-day effects could explain the differences between well- and poorly modeled emission episodes – and also whether these could explain emission fluxes in general. While discrepancies tend to occur more frequently around the afternoon, the overall differences remain difficult to attribute to a single factor with confidence.

Additionally, the uncertainty of the flux estimates has to be considered here: During emission events, the measured concentration differences between the highest and lowest sampling high were often very small (≤ 0.1 µg m^{-3}), making it challenging to distinguish these concentration differences from zero. While aggregated data do give more certainty that emission fluxes occur, the low concentration differences raise questions about the extent to which they can be meaningfully interpreted in relation to other variables such as the SO_2 concentration or meteorological parameters.

Given these uncertainties and the scope of the current study, we have chosen not to include a further breakdown of emission episode characteristics in this paper. However, we are planning to further analyze the mechanisms of compensation points and the modeling of emission events in future research papers.

RC1.14: 336 – 339: If the R_w parameterization is important to discuss in-depth, the equations used by each of the model schemes should be shown here, and not just in Table A1.

AC1.14: We agree that given the importance of the external pathway, it should be mentioned more explicitly. We have now added the equations for R_w of the three exchange schemes in Section 4.2.

RC1.15: Line 443 – 446: It might be interesting to prepare a companion figure to Figure 6 (if possible) comparing the modelled deposition for each of the three schemes using their original parameterization methods (e.g. for Γ_{soil}, Γ_s, etc), but a version that uses the same input values (using your best estimate) for each model—presumably these schemes vary both in how emissions/deposition are modelled, and also what inputs are used, and comparing just on the modelling approach would be interesting.

AC1.15: Although it certainly would be interesting to see how schemes would model the NH_3 exchange at Solleveld when we standardize the emission potentials for each scheme, it falls outside of the scope of the paper. We will consider such approaches in a following paper.

RC1.16: Section 4.4: Something that stood out to me in this paper was the methodological choice to have each scheme (DEPAC, Zhang, Massad) encompass not only the functions that calculate whether/at what rate NH_3 is emitted or deposited (e.g. F_w, F_{soil}, R_s, R_w, etc); but also for each scheme to use its own parameterization for the model inputs as well (primarily for Γ_s judging by Table A1). I think this choice needs to be justified in this section, because it isn't clear to me whether the conclusions—that the Zhang scheme accurately models the net flux, but not the diurnal cycle, while the DEPAC scheme captures the diurnal cycle, but over-estimates the net flux—are due to, for example, the DEPAC scheme over-estimating the mass transfer rate (e.g. Rs or Rw are too small), or because in the DEPAC the concentration gradient is too large (e.g. because Γ_s is too small).

The paper suggests that additional measurements of the underlying inputs (Γ_s, Γ_{soil}, Γ_w, etc) would improve the modelling approach (line 418), but this would also suggest that the models should be evaluated using the same inputs, rather than each model being evaluated with a separate set of inputs. Perhaps is not a significant factor as the different approaches for calculating the model inputs result in substantively similar values, but if so, that should be clarified in the text.

AC1.16: First, the methodological choice to evaluate each scheme with its original parameterization has been stated more clearly now in the introduction section of the paper to avoid confusion regarding the scope of the paper (see comment AC1.8). To clarify, we consider the Γ parameterization intrinsic to each scheme rather than an input, which is why it differs across the three schemes.

Whether the discrepancy between the modeled fluxes by the DEPAC and Zhang scheme results from the differences in the mass transfer rates or the compensation point parameterization cannot be inferred from the measurements. The reason for this is that it is not possible to fully disentangle the stomatal, external, and (possible) soil flux from each other. We also discuss this in the second paragraph of Section 4.4. Thus, while adjusting parameters such as R_w in the DEPAC scheme could bring the modeled fluxes closer to the observations, such modifications may not accurately reflect the physical processes taking place on-site.

Moreover, using in-situ Γ measurements and inserting these values as input in the exchange schemes would be a very useful method to evaluate the resistance parameterization, as we can rule out errors related to the Γ values and the compensation point parameterization. We have added the following text to Section 4.4:

"Additionally, directly inserting in-situ Γ values into the model descriptions – rather than calculating the Γ using the model equations – would allow for a more targeted validation of the resistance parameterization. By using measured Γ values, errors related to the compensation point parameterization can be ruled out, making it easier to assess the accuracy of the resistance terms – assuming that the measured Γ values themselves are accurate."

We acknowledge the reviewer's suggestion regarding using the same Γ values across all three exchange schemes. While this would lead to a more controlled comparison of the three exchange schemes, it would not necessarily lead to clearer conclusions.

Although the three exchange schemes have a stomatal compensation point and Γ_s values could be kept the same across the schemes, fundamental structural differences between the three schemes (e.g. DEPAC also has a χ_w and Zhang has a χ_{soil}) would still complicate the interpretation of results.

Technical corrections:

RC1.17: Line 21: The acronym DEPAC should be defined when used for the first time

AC1.17: A definition of the acronym DEPAC has been added to the abstract.

RC1.18: Line 60: The acronym DEPAC should be defined when used in the body for the first time.

AC1.18: A definition of the acronym DEPAC has been added to the main body.

RC1.19: Line 117: Γs (the stomatal emission potential?) used without being defined the first time

AC1.19: A definition has been added to the line 108-109 for Γ_s, Γ_w and Γ_{soil} – being the stomatal, external, and soil emission potential respectively.

RC1.20: Line 119: NH3 missing superscript

AC1.20: The subscript has been added to NH3 on line 119.

RC1.21: Table 1 (line 182): The text DEPAC scheme in the table seems to be right-justified, or is placed off-center.

AC1.21: The header for the DEPAC scheme has been centered properly now in Table 1.

RC1.22: Table A1(line 500 – 503): Parameters in the table should be defined, and where applicable, units should be given.

AC1.22: Table 1 now includes the units next to the variable abbreviation.

RC1.23: Table A2: Some, but not all parameters in the table are given with units; some of the parameters without (e.g. the emission potential) may be dimensionless, but others have previously been given with units (e.g. resistances). Not all of these parameters have been previously defined in the text.

AC1.23: Table A1 (previously Table A2) now includes a special column with the unit of each parameter, and the full name of every parameter has been given in the “Parameter” column.

Responses to referee #2

General comments:

This study assessed the performance of three ammonia bi-directional air-surface exchange schemes by (1) comparing model-estimated fluxes with one-year gradient measurement over a dune ecosystem, (2) comparing the dominant flux pathways between the models, and (3) conducting an error analysis to assess the uncertainties in modeled fluxes. The method used in this study is scientifically sound, some useful results are obtained, and the manuscript is generally well written. I have the following comments and hope the authors can consider when revising the manuscript:

Thank you for the positive evaluation of our paper.

Specific Comments:

RC2.1: Quantifying surface flux using the concentration gradient method typically requires the measurements to be conducted over a homogeneous terrain. It is not clear to me if such condition was satisfied where and when the measurement was made. Heterogeneous terrain can cause large uncertainties in the estimated flux, even changing flux directions (e.g., from deposition to emission or the other way around). Even over homogeneous terrain flux measured using gradient method can differ significantly from those measured using eddy-covariance method (as has been seen for O3 flux measurements over a forest reported by Wu et al., 2015, ACP, 15, 7487-7496). What is the uncertainty magnitude (or confidence level) of the measured ammonia flux data? This directly affects the model-measurement comparison results.

AC2.1: Thank you for this excellent question. The deposition measurements at Solleveld were done to examine how much nitrogen is taken up by nitrogen-sensitive Natura-2000 ecosystems. However, given the high biodiversity of the Natura-2000 areas like Solleveld, the terrain is heterogeneous by definition, making it technically challenging to locate a homogeneous footprint for dunelands. Additionally, back in 2014-2015, when the measurements were performed, EC systems were not easily available for flux evaluation and the gradient method was therefore our only option. Vendel et al. (2023) discuss the footprint of the measurement site, dividing the data into four wind sectors (north, east, south, and west) to account for the inhomogeneity. Figure 4c in Vendel et al. (2023) shows that two ponds are located east and south of the measurement set-up, which acknowledges the terrain was not perfectly homogeneous.

A homogeneous footprint can be achieved by applying a strict wind direction filter, using only the western quadrant and omitting data from the north, east, and south (where the ponds are located). We have considered applying this filter to this study, and Fig. S1 and Fig. S2 show the impact of this approach. With a stronger filter, the observed NH_3 flux remains net deposition, although relatively more emission takes place when compared to the original dataset. Moreover, the filtering notably affects the performance of the Massad scheme, which now aligns more closely with observations. However, other key findings remain unchanged: The DEPAC scheme predicts higher fluxes than observed, and the Zhang scheme slightly overestimates the flux now but is still accurate. Additionally, the emission events are still not captured properly by the exchange schemes, suggesting inaccuracies in the compensation point parameterization.

However, the application of this tight filter comes with a trade-off: NH_3 concentrations are significantly higher when the wind comes from the east and south (see Table S1), while the average NH_3 concentration from the west is just 0.7 µg m^{-3}. This means that removing data from the east and south would not only discard two-thirds of our dataset but also prevent the validation of the three exchange schemes under more polluted conditions. For example, the omission of the high-concentration periods, associated with stronger deposition fluxes, also explains the improvement in the performance of the Massad scheme, as conditions are filtered where the Massad scheme typically underperforms. As a result, applying such a strict filter could also give a misleading impression of the overall performance of the schemes.

Despite the theoretical demands that are hard to fulfill when considering all wind directions, we decided not to apply a stricter wind filter. To indicate this trade-off between a homogeneous footprint and retaining valuable data, we added the following text to Section 2.1:

"To the east and south of the site are ponds surrounded by reed, contributing to terrain inhomogeneity. This inhomogeneity could be resolved by filtering out data from these wind directions; However, this would significantly reduce the dataset and would remove periods with higher NH_3 concentrations, limiting the validation of the schemes under more polluted conditions. Therefore, we did not apply this filter."

Table S1: Sample size and NH_3 concentration per wind sector. Note that the values in this table are based on the dataset after filtering and therefore differ slight from those in Table 1 of Vendel et al. (2023), which were calculated before filtering.

Wind quadrant	North	East	South	West
Sample size	451	691	462	834
Mean NH_3 concentration	0.9 µg m^{-3}	3.7 µg m^{-3}	2.4 µg m^{-3}	0.7 µg m^{-3}

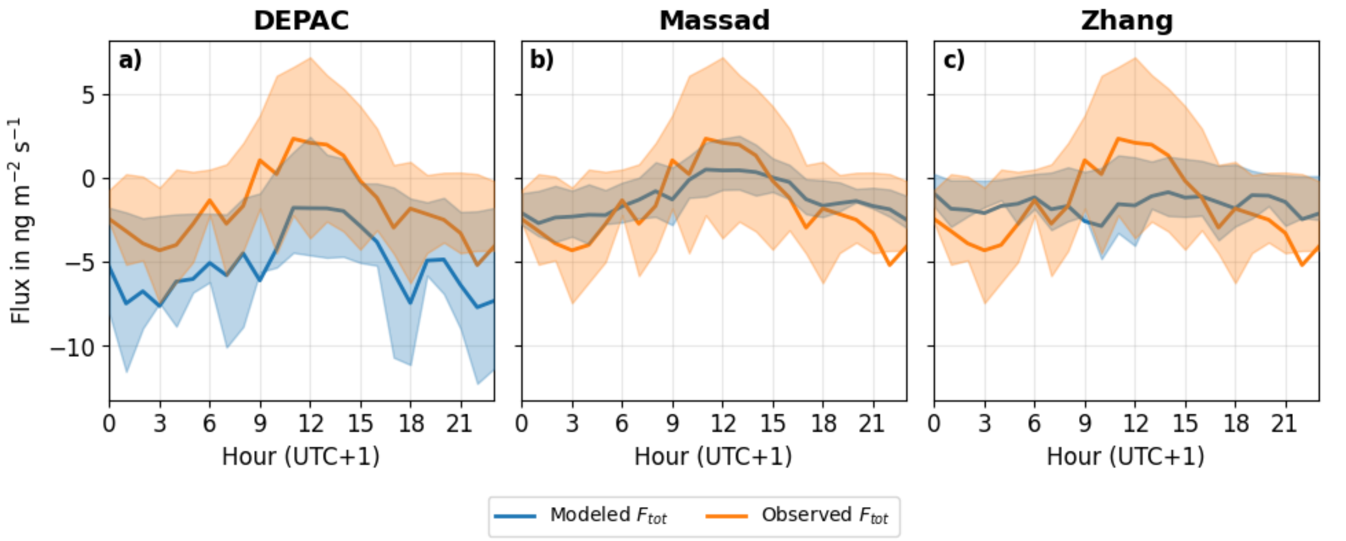

Figure S 1: Observed (orange) and modelled (blue) diurnal fluxes of the DEPAC, Massad and Zhang scheme at Solleveld, with data exclusively when the wind comes from the west (220-330°). The shading around the line indicates the 25-75 percentile range.

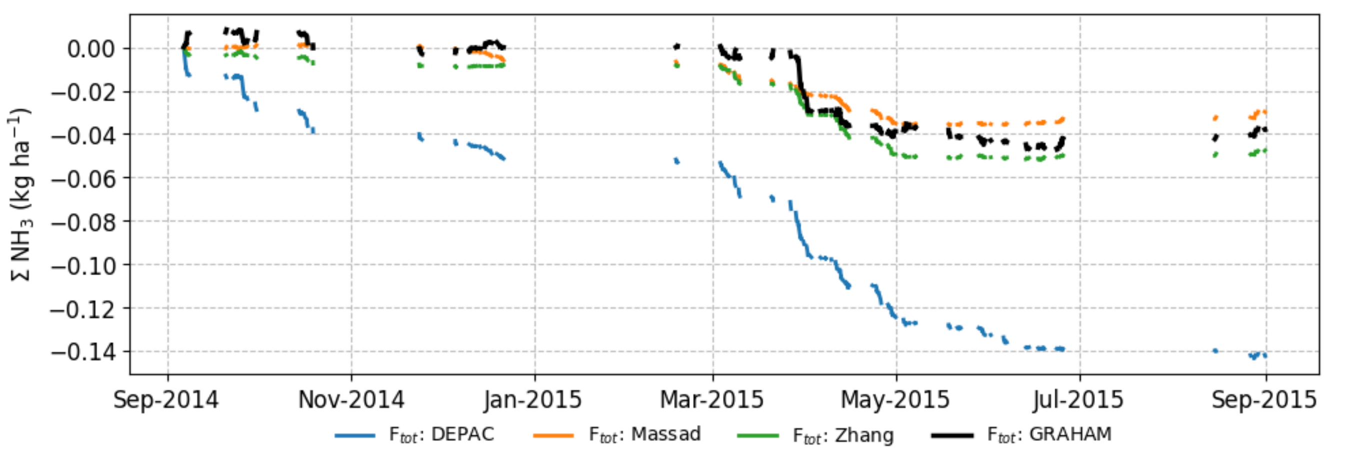

Figure S 2: Measured and modelled accumulated NH_3 flux at Solleveld, with data exclusively when the wind comes from the west (220-330°).

Finally, Vendel et al. (2023) estimated the total uncertainty for aggregated fluxes to be ± 24% (2 σ). Given that the conclusions drawn in our study are only based on aggregated statistics such as averaged diurnal cycles, and accumulated fluxes, we also applied this confidence level in our study.

RC2.2: The model-measurement comparison split the flux data into different categories (Table 1), which is a very good practice. Can the analysis also be done by looking at different wetness conditions (as was done for O3 deposition in Zhang et al., 2002, Atmospheric Environment, 36, 4787-4799). Such an analysis may help identify emission sources (e.g., evaporation of morning dew or soil emission).

AC2.2: Thank you for your suggestion. We conducted a small analysis, evaluating the performance of the three schemes under different wetness conditions. However, presenting the results as averaged fluxes for wet and dry conditions would be misleading. While net deposition took place in both cases, NH_3 emissions were frequently observed under both wet and dry conditions (see Fig. S3), which would not be captured by mean values.

Given the complexity of NH_3 emission dynamics and their representation in models, we aim to investigate these mechanisms in more detail in future papers beyond the scope of this study.

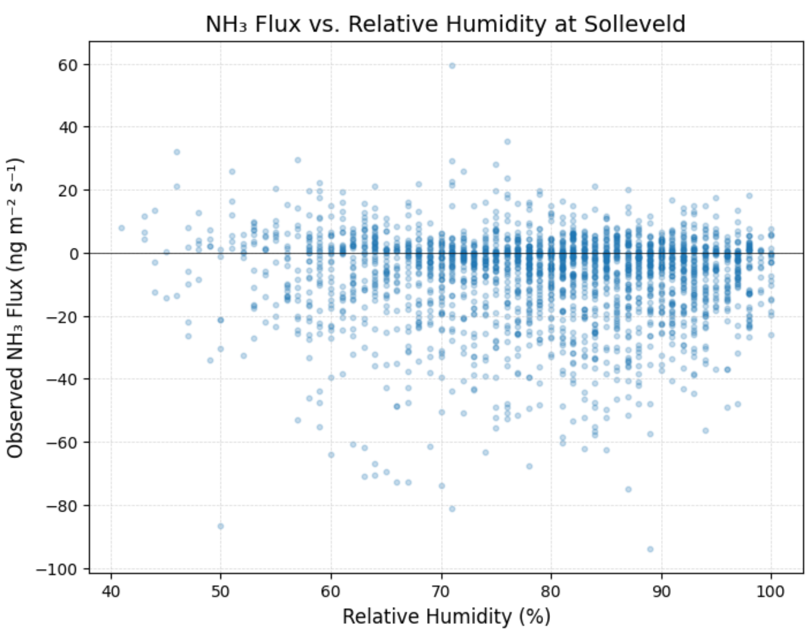

Figure S 3: Scatterplot indicating the relative humidity in % (x-axis) versus the observed NH_3 flux in ng m^{-2} s^{-1} (y-axis) at Solleveld.

RC2.3: Although using a bidirectional air-surface exchange scheme is more theoretically correct than using a traditional big-leaf dry deposition scheme, the former does not necessarily perform better than the latter in the simulated ammonia fluxes on seasonal to annual basis and in regional-scale air-quality modeling, as reported by several existing studies. This is because modelling the bi-directional flux requires additional model parameters such as the soil and canopy NH3 emission potentials, which may not be available at high spatial resolution on the reginal scale. Besides, more model parameters can introduce additional uncertainties. For example, Wen et al. (2014) compared the

unidirectional model of Zhang et al. (2003) and the bi-directional exchange model of Zhang et al, (2010) against 53 sites in southwestern Ontario, Canada, and showed that the bidirectional scheme performed best at locations with high observed NH3 concentrations, but overestimated NH3 levels for locations with low observed NH3 concentrations, and the unidirectional dry deposition generally performed better than the bidirectional scheme at sites with low observed NH3 concentrations. Another example is by Zhu et al. (2015, Atmos. Chem. Phys, 15, 12823-12843) who introduced bidirectional fluxes into GEOS-CHEM and compared with surface AMoN observations, which resulted decreasing errors in July, but increasing them in April and October. Thus, similar to Wen et al., 2014 above, even though the bidi scheme incorporates more physics into the parameterization, it still might not perform the best under all conditions. Thus, I am wondering if the authors can recommend under what conditions the unidirectional model is acceptable and under what conditions the bi-directional exchange model should be used?

AC2.3: Thank you for your interesting question. The comparison between the uni-directional scheme from Zhang et al. (2003) and the bi-directional scheme from Zhang et al. (2010) in the study by Wen et al. (2014) shows that the bi-directional scheme performs better in areas with higher agricultural activity, whereas uni-directional schemes underestimate the observed NH_3 concentration. In contrast, it performs worse in natural areas with lower observed NH_3 concentrations, overestimating the ambient NH_3 concentration, caused by the inclusion of a non-zero χ_c, which effectively inhibits deposition (or perhaps even causes net emission).

Despite the uncertainties introduced by modeling emission potentials, we would argue that ideally bi-directional schemes should be used as they provide a more mechanistic representation of the ammonia exchange but also allow the user/modeler to fine-tune the scheme based on observed physical processes. To further explain this, consider the following example related to the results by Wen et al. (2014): Correcting the bi-directional schemes in the natural areas (where it generally overestimates the NH_3 concentration) would be achieved by lowering the emission potentials, which would also be justifiable from a physical perspective. In contrast, correcting a uni-directional scheme in agricultural areas (where it mostly underestimated the NH_3 concentration) can only be done by increasing the resistances, which is physically unsatisfactory as the problem lies in the underestimated “nitrogen status” of the ecosystem (i.e. the compensation point) and not the mass transfer rate.

For bi-directional schemes to consistently outperform their uni-directional counterpart, improvements are necessary in the parameterization of the emission potentials. As land use and vegetation type alone are not necessarily strong predictors of NH_3 emission potentials, a more refined approach should incorporate factors such as nitrogen inputs from agricultural activity, or prior N deposition, similar to the DEPAC scheme or the scheme by Massad et al. (2010).

RC2.4: Considering the large uncertainties in the modeled ammonia flux between the three bi-directional exchange models investigated in this study (and likely the case for several other existing ones in literature), would an ensemble approach reduce such uncertainties?

AC2.4: Thank you for your question. We have conducted a preliminary study on the performance of a simple ensemble model, taking the average of three modeled fluxes. Figure S4d below shows the diurnal pattern of an ensemble model. In terms of order of magnitude and timing the ensemble model does perform well. However, the statistics in Table S27 (an adaptation of Table 2 in the manuscript) indicate that the ensemble model does overestimate the mean NH_3 flux and that in the majority of the cases, the Zhang scheme still outperforms the ensemble model.

Additionally, given that all three exchange schemes do not properly model the NH_3 emissions, an ensemble approach would also not reduce the error in modeling NH_3 emissions (which can be seen in Fig S4d). While ensemble modeling is interesting, its added value in the context of improving NH_3 exchange schemes remains limited. However, it is definitely a relevant question for operational models, where reducing the uncertainty in model output is the goal.

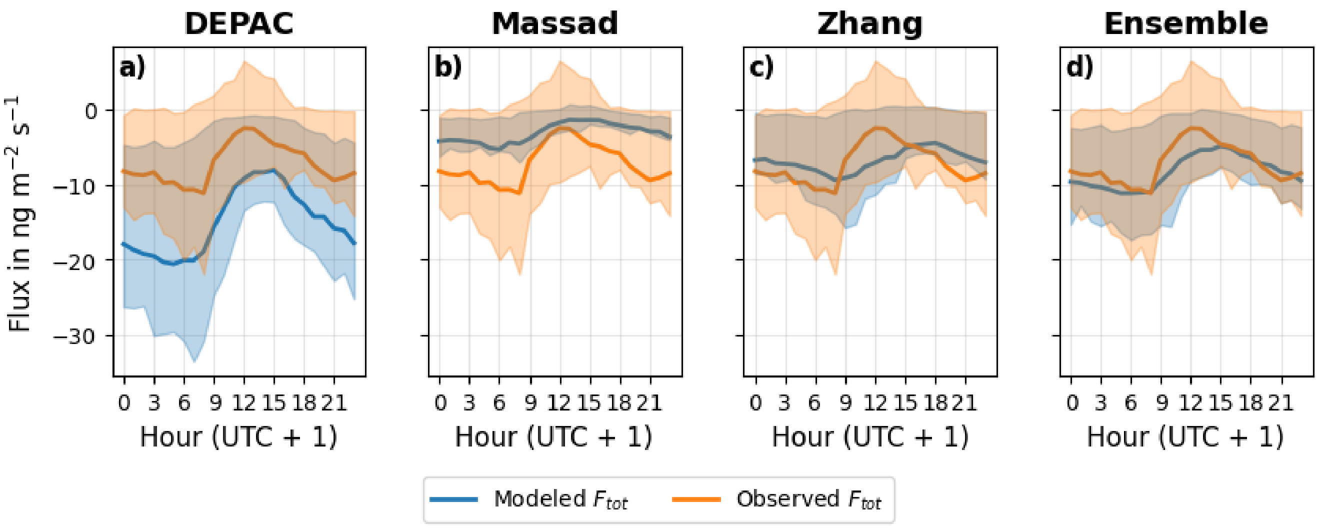

Figure S 4: Measured and modeled the average diurnal pattern of the NH_3 flux at Solleveld in ng m^{-2} s^{-1}, per model and the ensemble model. The shading around the lines depicts the 25-75% percentile range

Table S 2: Descriptive statistics of the measured and the modeled fluxes per exchange scheme. Additionally, the statistics are shown when strong deposition ($F_{tot} \leq -20$), moderate deposition ($-20 > F_{tot} < 0$), and emission ($F_{tot} \geq 0$) is observed.

	n	**GRAHAM**	**DEPAC scheme**			**Massad scheme**			**Zhang scheme**			**Ensemble**		
		Mean flux	Mean flux	RMSE	R	Mean flux	RMSE	R	Mean flux	RMSE	R	Mean flux	RMSE	R
ng m^{-2} s^{-1}	-	ng m^{-2} s^{-1}	ng m^{-2} s^{-1}	ng m^{-2} s^{-1}	-	ng m^{-2} s^{-1}	ng m^{-2} s^{-1}	-	ng m^{-2} s^{-1}	ng m^{-2} s^{-1}	-	ng m^{-2} s^{-1}	ng m^{-2} s^{-1}	-
Total dataset	2438	-7.0	-14.6	16.5	0.56	-3.0	14.7	0.18	-6.8	11.1	0.63	-8.2	11.5	0.59
F ≤ -20	368	-33.2	-34.6	22.5	0.16	-5.0	33.2	0.1	-22.1	18.8	0.40	-20.6	19.2	0.29
-20 > F < 0	1269	-6.8	-13.4	13.8	0.50	-2.9	7.9	0.14	-4.8	6.8	0.45	-7.1	6.8	0.48
F≥ 0	801	4.7	-7.4	17.1	-0.15	-2.3	10.3	-0.03	-2.9	11.8	-0.25	-4.2	12.5	-0.18

Finally, we have computed the confidence interval of this ensemble model, which is presented in Fig. S5. The confidence interval for the ensemble model was derived by aggregating the 5000 Monte Carlo simulations from each of the three models and determining the 2.5th and 97.5th percentiles of the mean flux, which define the 95% CI. The mean flux of the ensemble model is -8.2 ng m^{-2} s^{-1} with a 95% CI ranging from -26.8 to 3.4 ng m^{-2} s^{-1}. The 95% CI of the ensemble model is still large, which can be attributed to the large uncertainty ranges it inherits from the DEPAC and Massad scheme.

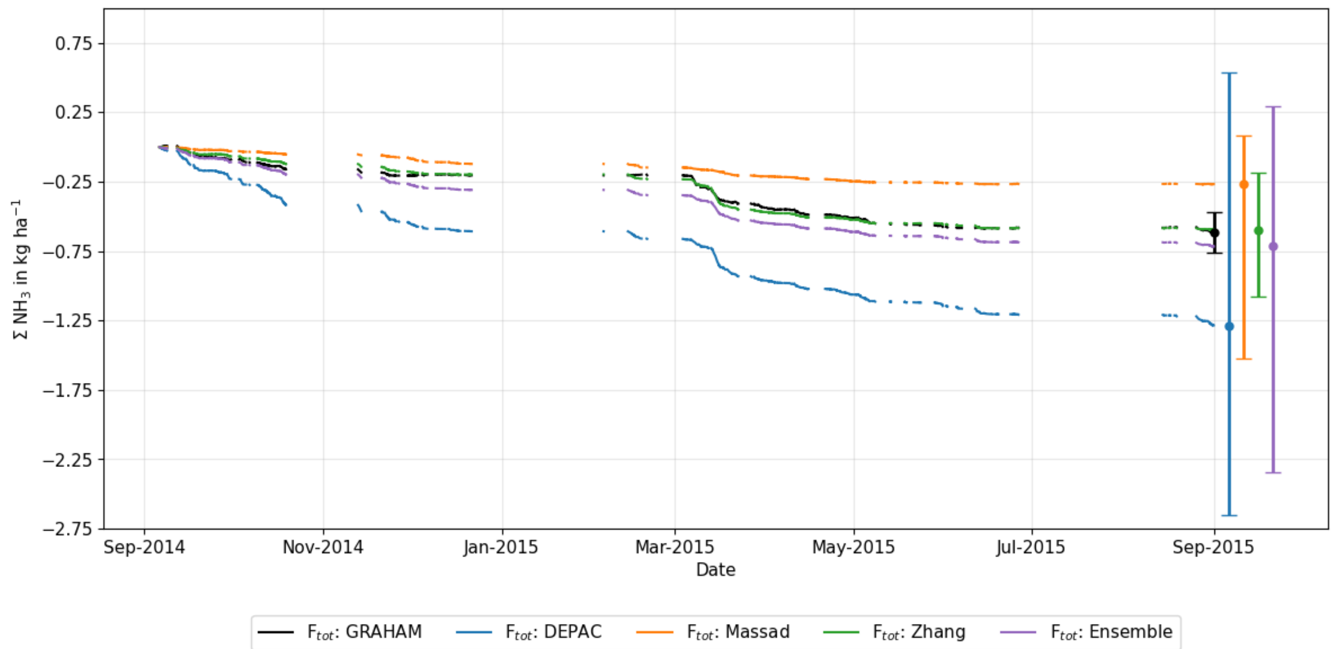

Figure S 5: Accumulated NH_3 flux at Solleveld of the three schemes, the ensemble model and the observations from the GRAHAM. The error bars on the right indicate the 95% confidence interval of the total accumulated flux per exchange scheme or model

Changes to the original manuscript

Besides some small edits on e.g. the choice of words or phrases to improve the readability of the text, we have made some additional changes to the manuscript:

First, we have chosen to apply the exact same filters to the Solleveld measurements as performed by Vendel et al. (2023). In the previous manuscript, two filters were accidentally not applied: (a) the reference concentration at 1 meter should be higher than the detection limit (0.1 µg m^{-3}) and (b) the calculated NH_3 concentration at $z = z_0$ should be higher than zero. We did filter already for (a) the friction velocity should be above 0.1 m s^{-1} and (b) the filter for measurements made when the wind direction was between 100°-120°, as the measurement equipment would obstruct the wind flow. As a consequence, the number of valid measurements decreased slightly from n = 2451 to n = 2438. The three exchange schemes have been rerun with this altered dataset, just as the uncertainty analysis and the sensitivity analysis. In terms of results, this did not lead to different conclusions than the ones drawn in the manuscript. The statistics presented in Tables 1 and 2, and all figures have been replaced to align with the updated dataset selection criteria. Additionally, we added the following text to indicate that the same filter criteria have been applied as Vendel et al. (2023):

To ensure the quality of the NH_3 gradient flux measurements, we have applied the same filters to the dataset as Vendel et al. (2023).

Second, we did not mention in the methodology section that the Massad scheme utilizes the same R_s parameterization as DEPAC. The reason for this is that Massad et al. (2010) does not specify the parameterization for R_s but only suggests several parameterizations. This information has now been added to Section 2.2 as follows:

"Similarly, the Massad scheme also adopts this parameterization as Massad et al. (2010) do not provide specific R_s parameterization."

Third, in Table A1, we did not list the co-deposition parameterization, as mentioned in Wichink Kruit et al. (2017), even though this parameterization is used in the DEPAC scheme. The co-deposition function $F(\alpha_{SN})$ has been added to Table 1 (previously Table A1). Fourth, the default maximum stomatal conductance G_s^{max} value for ozone was used in the DEPAC and Massad scheme during the sensitivity analysis and the Monte Carlo uncertainty propagation. However, the G_s^{max} value should be scaled proportionally with the diffusivity of NH_3 in the air ($Gsmax \cdot \frac{D_{NH3}}{D_{O3}}$). This scaling has now been implemented. While the updated uncertainty and sensitivity analyses show slight variations, the overall conclusions drawn from the results remain largely unchanged given the relatively small contribution of stomatal exchange in all three exchange schemes.

Fifth, The graphic design of Fig. 1 was changed to improve the readability. Additionally, the canopy compensation point χ_c of the Massad scheme was missing in the previous figure and has now been added. We have also improved the readability of Figures 2-6, changing the font size of the tick labels and the axis labels.

Sixth, in Fig. 5i, the diurnal average of the stomatal conductance of the Zhang scheme was poorly visible, due to the handling of extreme R_s values in the Zhang scheme. Specifically, when any of the Jarvis functions equaled zero, G_s effectively becomes zero, leading to R_s values being set to an extremely high value of 10^{25} s m^{-1}. In total, 1038 out of 2438 records had an $R_s = 10^{25}$ s m^{-1}, mainly occurring during the evening and night, but occasionally in the afternoon. Since even a few of these extreme values strongly affected hourly averaging, we omitted them to prevent distortion.

Finally, for Section 3.1, we have recalculated how often the exchange schemes would model emission during observed emission events. A mistake was made here which has been corrected – not related to data that has been filtered after the two extra filters have been applied – and as a result, the values changed from 6% to 14% for DEPAC, 12% to 20% for the Massad scheme, and from 23% to 31% in the Zhang scheme. Although the percentages increased, the main conclusion still holds that emission events are still poorly captured by the three exchange schemes.

Bibliography

Massad, R. S., Nemitz, E., & Sutton, M. A. (2010). Review and parameterisation of bi-directional ammonia exchange between vegetation and the atmosphere. *Atmospheric Chemistry and Physics*, *10*(21), 10359–10386. https://doi.org/10.5194/acp-10-10359-2010

Sutton, M. A., Schjorring, J., & Wyers, G. (1995). Plant atmosphere exchange of ammonia. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *351*, 261–276.

Van Zanten, M. C., Sauter, F. J., Kruit, R. J. W., J.A. van Jaarsveld, P., & Pul, W. A. J. van. (2010). *Description of the DEPAC module Description of the DEPAC module*.

Vendel, K. J. A., Wichink Kruit, R. J., Blom, M., van den Bulk, P., van Egmond, B., Frumau, A., Rutledge-Jonker, S., Hensen, A., & van Zanten, M. C. (2023). Dry deposition of ammonia in a coastal dune area: Measurements and modeling. *Atmospheric Environment*, *298*(June 2022), 119596. https://doi.org/10.1016/j.atmosenv.2023.119596

Wen, D., Zhang, L., Lin, J. C., Vet, R., & Moran, M. D. (2014). An evaluation of ambient ammonia concentrations over southern Ontario simulated with different dry deposition schemes within STILT-Chem v0.8. *Geoscientific Model Development*, *7*(3), 1037–1050. https://doi.org/10.5194/gmd-7-1037-2014

Wichink Kruit, R. J., Aben, J., De Vries, W., Sauter, F., Van Der Swaluw, E., Van Zanten, M. C., & Van Pul, W. A. J. (2017). Modelling trends in ammonia in the Netherlands over the period 1990–2014. *Atmospheric Environment*, *154*, 20–30. https://doi.org/10.1016/j.atmosenv.2017.01.031

Wichink Kruit, R. J., van Pul, W. A. J., Sauter, F. J., van den Broek, M., Nemitz, E., Sutton, M. A., Krol, M., & Holtslag, A. A. M. (2010). Modeling the surface-atmosphere exchange of ammonia. *Atmospheric Environment*, *44*(7), 945–957. https://doi.org/10.1016/j.atmosenv.2009.11.049

Zhang, L., Brook, J. R., & Vet, R. (2003). A revised parameterization for gaseous dry deposition in air-quality models. *Atmospheric Chemistry and Physics*, *3*(6), 2067–2082. https://doi.org/10.5194/acp-3-2067-2003

Zhang, L., Wright, L. P., & Asman, W. A. H. (2010). Bi-directional air-surface exchange of atmospheric ammonia: A review of measurements and a development of a big-leaf model for applications in regional-scale air-quality models. *Journal of Geophysical Research Atmospheres*, *115*(20). https://doi.org/10.1029/2009JD013589