Evaluation of a new inference method for estimating ammonia volatilisation from multiple agronomic plots

Benjamin Loubet[1,*], Marco Carozzi[1#], Polina Voylokov[1], Jean-Pierre Cohan[2], Robert Trochard[2], Sophie Génermont[1]

1 INRA, UMR ECOSYS, INRA, AgroParisTech, Université Paris-Saclay, 78850, Thiverval-Grignon, France
2 ARVALIS-Institut du Végétal, **Station expérimentale de La Jaillière, La Chapelle Saint Sauveur, 44370 Loireauxence**, France
now at: Agroscope Research Station, Climate and **Agriculture**, Zurich, Switzerland

* Corresponding author: Benjamin.Loubet@inra.fr

Abstract. Tropospheric ammonia (NH_3) is a threat to the environment and human health and is mainly emitted by agriculture. Ammonia volatilisation following application of nitrogen in the field accounts for more than 40% of the total **NH_3** emissions in France. This hence represents a major loss of nitrogen use efficiency which needs to be reduced by appropriate agricultural practices. In this study we evaluate a novel method to infer **NH_3** volatilisation from small agronomic plots made of multiple treatments with repetition. The method is based on the combination of a set of **NH_3** diffusion sensors exposed for durations of 3 hours to 1 week, and a short-range atmospheric dispersion model, used to retrieve the emissions from each plot. The method is evaluated by mimicking **NH_3** emissions from an ensemble of 9 plots with a resistance-analogue-compensation-point surface exchange scheme over a yearly meteorological database separated into 28-days periods. A multi-factorial simulation scheme is used to test the effects of sensor number and heights, plot dimensions, source strengths and background concentrations, on the quality of the inference method. We further demonstrate by theoretical considerations in the case of an isolated plot that inferring emissions with diffusion sensors integrating over daily periods will always lead to underestimations due to correlations between emissions and atmospheric transfer. We evaluated these underestimations as -8% ± 6% of the emissions for a typical western European climate. For multiple plots, we find that this method would lead to median underestimations of -16% with an interquartile [-8% -22%] for two treatments differing by a factor of up to 20 and a control treatment with no emissions. We further evaluate the methodology for varying background concentrations and **NH_3** emission patterns and demonstrate the low sensitivity of the method to these factors. The method was also tested in a real case and proved to provide sound evaluations of **NH_3** losses from surface applied and incorporated slurry. We hence showed that this novel method should be robust and suitable for estimating **NH_3** emissions from agronomic plots. **We believe that the** method **could be further improved by using Bayesian inference and inferring surface concentrations rather than surface fluxes. Validating against controlled source is also a remaining challenge**.

Keywords: NH_3 emission, multiple sources, dispersion modelling, experimental design, diffusive samplers

Introduction

Tropospheric ammonia (NH_3) is mainly emitted by agriculture and has great environmental impacts (atmospheric pollution, eutrophication, reduction of biodiversity) which are increasingly taken into account in

European and international regulations (Council, 1996; Council, 2016; UNECE, 2012). Ammonia losses also have great agronomic and economic impacts for farmers, as it reduces nitrogen use efficiency. The varying prices of mineral fertilizers and concerns about environmental and health threats demand improvements in the efficiency of nitrogen utilisation, and especially in recycling nitrogen through organic fertilization (Sutton et al., 2011). Indeed, NH_3 volatilization during storage of manure and slurry and following their field application is the main source of NH_3 in Europe (55% of the emissions) while farm buildings emissions represent **45%. In France, crop farming represent 35% of the emission and animal farming represent 65%** (CITEPA, 2017; ECETOC, 1994; EUROSTAT, 2012; Faburé et al., 2011). Reducing NH_3 losses from this agricultural sector is therefore a major objective for applied research.

While NH_3 emissions from farm buildings and storage can be handled by engineering solutions, losses during organic fertilisation are much more dependent on the combination of application methods (splash plate, band spreading, pressurised injection, open and close slot injection, trailing hose and trailing shoe), soil type and occupation, and environmental conditions (soil humidity, air temperature, wind speed, solar radiation) (Sommer et al., 2003). For instance, Sintermann et al. (2012) report NH_3 losses following cattle and pig slurry application in the field ranging from a few percent to 50% over large fields and up to 100% over medium fields. Evaluating ammonia losses from field fertilisation over a range of practices, soil and climatic conditions is therefore key in evaluating the best application methods.

However, characterising these emissions at the field scale requires complex experimental design and most of the time **also requires the use of** large fields (Ferrara et al., 2016; Ferrara et al., 2012; Flechard and Fowler, 1998; Loubet et al., 2012; Milford et al., 2009; Sintermann et al., 2011b; Spirig et al., 2010; Sun et al., 2015; Whitehead et al., 2008). Especially useful for measuring ammonia losses are methods that can deal with small and medium-scale fields **(**20**-**50 m on the side**) that** are **commonly used in agronomic trials**. Indirect estimation methods (soil nitrogen balance or ^{15}N balance) are not well adapted to evaluate gaseous ammonia losses, mainly because of the soil heterogeneity and also because the method relies on evaluating small variations of large numbers (McGinn and Janzen, 1998). Among existing methods for measuring NH_3 emissions, the integrated horizontal flux method (Wilson and Shum, 1992) is well adapted, but is a subject of debate in its practical application since it seem to be systematically biased towards higher estimates (Häni et al., 2016; Sintermann et al., 2012). Alternatively, enclosure methods proved to be not representative for a sticky compound such as ammonia (Pacholski et al., 2006), but more concerning is the fact that ammonia fluxes result from an air-surface equilibrium which is disturbed by the confined environment offered by the chamber. Inverse dispersion modelling approaches either based on backward Lagrangian Stochastic models (Flesch et al., 1995) or Eulerian models (Kormann and Meixner, 2001; Loubet et al., 2001), based on the Philip equation (Philip**,** 1959) have been demonstrated to be adapted for estimating NH_3 volatilization from **strong** sources (Loubet et al., 2010; Sommer et al., 2005).

These approaches are well adapted to small or medium fields ($\leq 50\times50$ m^2) but typically require hourly **NH_3 concentration measurements.** Long term concentration measurements of NH_3 are now well handled by the use of short path passive samplers developed by Sutton, et al. (2001), or active denuders, which have both been used for concentration monitoring for years (Tang et al., 2001; Tang et al., 2009). These active denuders can be adapted for measuring fluxes based on conditional sampling like the conditional time averaged gradient method COTAG (Famulari et al., 2010), which is a useful method but only adapted for large fields ($\geq$ 0.5 ha). The

passive samplers have also been shown to be adapted for inverse modelling estimations of NH_3 sources for large fields (Carozzi et al., 2013b; Ferrara et al., 2014).

In another field of research, solutions to the multiple source **inference** problem, which consists of inferring multiple sources based on measured concentrations at multiple points in space and time, have been developed especially since 2008 (Crenna et al., 2008; Gao et al., 2008; Gericke et al., 2011; Mukherjee et al., 2015; Vandré and Kaupenjohann, 1998). They have chiefly been used over regional scales (Flesch et al., 2009; Lushi and Stockie, 2010; Yee and Flesch, 2010), and have been shown to be very dependent on the source-sensor geometry (Crenna et al., 2008; Flesch et al., 2009; Wang et al., 2013). Mukherjee et al. (2015) highlighted the dependency of the inferred source to background concentration and plot disposition, by means of an inverse footprint approach. Yee et al. (2008) have shown how to retrieve the number, location and intensity of multiple sources with dispersion models coupled with Bayesian inference methods. Yee and Flesch (2010) have evaluated the inversion and inference methods for determining 4 points sources using several laser transects. Flesch et al. (2009) have shown that source-receptor geometry is critical in determining whether a multiple-source inversion problem can provide realistic solutions or not. Flesch et al. (2009) have moreover shown that if the geometry is well chosen the accuracy of the method for 15 min **integration time** can reach 10% to 20%. These studies have also shown that the multiple source **inference** problems can be solved if not ill-conditioned (ill-conditioning depends on the location of sources and concentration sensors and is characterised by a conditioning number κ).

In this study, we pose the following research questions: **"Can inverse dispersion modelling approaches be used for inferring NH_3 emissions from multiple small plots (agronomic trials) using passive samplers, and to which degree of accuracy?"** The answer is given through the investigation of the optimal design in terms of field dimensions, plots **location** and size, passive sampler locations and their duration of exposure. Throughout this study, agronomic trials are considered as adjacent multiple small fields with repetitions of treatments. A typical trial would consist of three repetitions of three treatments. Hence the double challenge that we face in this study is (i) to consider together the multiple source **inference** issue (adjacent small fields) and the (ii) time-integration issue (using passive samplers).

To answer these questions, we use a 4 step approach: (1) The ammonia emissions are first modelled on each source using prescribed NH_3 emission potential dynamics coupled with a simple soil-vegetation-atmosphere exchange scheme to mimic realistic seasonal, daily and hourly variations in NH_3 emissions. (2) These prescribed emissions are then used to estimate the concentration at each target location using short-range atmospheric dispersion modelling over half hourly periods. (3) The obtained concentrations are then averaged over several integration periods to simulate the behaviour of passive samplers. Finally, (4) the sources are evaluated by inference with dispersion modelling based on the averaged concentrations.

Two dispersion models and several inference methodologies are evaluated. The effect of the size of the source, the locations of targets, the dynamics and magnitude of each source**,** the meteorological conditions **and the background concentration variability** are evaluated and discussed. The feasibility of the method is finally evaluated over a real case with two repetitions of three treatments (slurry spreading, injection and a reference without fertilisation).

2. Materials and methods

At first we present the theoretical background of source inference by optimisation for single and multiple sources with time averaging concentration sensors. Then the method used to generate a realistic ammonia source is **introduced before** the **description of the** dispersion models used for **both** generating the concentration fields and inferring back the sources**.** The geometry of the sources**, sensor locations** and the meteorological data used **for this analysis** are then shown, and finally the real test case used for evaluating the method is detailed.

2.1 The theory of the source inference method

At first we will recall some important theoretical features of the inverse dispersion modelling approach which is actually an inference method.

2.1.1 Case of a single area source and a single concentration sampler

We first consider the case of a single area source with a single concentration sampler (target). The source is varying with time. The method is based upon the general superimposition principle (Thomson et al., 2007), which relates the concentration at a given location $C(x,t)$ to the source strength $S(t)$ and the background concentration $C_{bgd}(t)$ using a transfer function $D(x,t)$, which has the dimensions of a transfer resistance (s m^{-1}).

$$C(x,t) = D(x,t) \times S(t) + C_{bgd}(t) \quad (1)$$

Here x denotes the location of the sensor and t the time. **The concentration and source units are in µg N-NH_3 m^{-3} and µg N-NH_3 m^{-2} s^{-1}, respectively.** The superimposition principle implies that the studied tracer must be conservative, which is a reasonable hypothesis for NH_3 whose reaction time with acids in the atmosphere is below the transport time for spatial scales below 1000 m (Nemitz et al., 2009). Moreover, in **Eq. (1)**, we assume a spatially homogeneous area source with strength $S(t)$. The spatial homogeneity of the source is less trivial for NH_3 **than other gas released in agriculture** as the source itself depends on the concentration at the surface. However (Loubet et al., 2010) have shown that the heterogeneity of the source can be neglected as long as the dimension of the source is larger than 20 m. Hence, this study is limited to source areas with fetch larger than 20 m and a spread of the concentration samplers over a domain smaller than 1000 m. Moreover, it is interesting to note that for infinitely spread fields, the transfer resistance is linearly linked to the transfer matrix (see supplementary material S1)

2.1.2 Effect of time averaging sensors on source inference for a single source

Since we consider time averaging concentration samplers, we develop the time-averaged equation of **Eq. (1)** over **an integration** time period τ :

$$\overline{C(x)} = \overline{D(x) \times S} + \overline{C_{bgd}} \quad (2)$$

where the overbars denote a time average over the period τ. Similarly as what is done in turbulent flux calculations, the first part of the right hand side of **Eq. (2**) is decomposed using the Reynolds decomposition of a random variable (Kaimal and Finnigan, 1994), giving:

$$\overline{C(x)} = \overline{D(x)} \times \bar{S} + \overline{C_{bgd}} + \overline{D'(x)S'} \quad \textbf{(3)}$$

where $\overline{D(x)'S'}$ is the time covariance between $D(x,t)$ and $S(t)$. If the averaged background concentration $\overline{C_{bgd}}$ is a known quantity, **Eq. (3)** can be easily manipulated to give an estimation of the averaged source strength $\bar{S}$, the quantity we want to infer:

$$\bar{S} = \frac{\overline{C(x)} - \overline{C_{bgd}}}{\overline{D(x)}} - \frac{\overline{D'(x) \times S'}}{\overline{D(x)}} \quad \textbf{(4)}$$

(I) (II)

In the right hand side of **Eq. (4)**, (I) can be calculated from measured $\overline{C_{bgd}}$ and $\overline{C(x)}$ and $\overline{D(x)}$ which is itself calculated with dispersion models. On the contrary (II) is *a priori* unknown and depends on the correlation between the source strength and the transfer function $\overline{D(x)'S'}$**.** Hence, if (II) is neglected, the inferred source $\bar{S}$ is biased. The relative bias of the method is then:

$$\frac{\delta \bar{S}}{\bar{S}} = \frac{\overline{D'(x)S'}}{\overline{D(x)} \times S} \quad (5)$$

Hence we show in **Eq. (5)** that time-averaging leads to a relative bias which can be quantified by the time covariance between the transfer function and the source strength. However this quantity is by nature unknown since the dynamics of $S(t)$ is unknown. Determining $\overline{D(x)'S'}$ requires knowledge of the source dynamics which can be obtained from measurements with a micrometeorological method. It can alternatively be approached by modelling using the state of the art of ammonia exchange processes as we do here.

Additionally to the bias, which is term (II) in **Eq. (4)**, evaluating term (I) is encompassed with errors related to the uncertainties in $\overline{C_{bgd}}$**,** $\overline{C(x)}$ and $\overline{D(x)}$**.** In particular, cases when $\overline{D(x)}$ is small may lead to large errors in inferring the source term S. This is linked to the conditioning of the inverse problem and is discussed in supplementary material S2.

2.1.3 Case of multiple sources and multiple concentration samplers with time averaging

If we generalise the approach to multiple sources and multiple receptors, then the transfer function becomes a matrix $D(x_i, S_j, t)$, which is the contribution of source S_j to concentration at target located at x_i. For reading purposes we simplify the matrix notation to $\boldsymbol{D_{ij}}$. **Eq (3)** then becomes:

$$\overline{\begin{bmatrix} C_1 \\ \vdots \\ C_M \end{bmatrix}} = \overline{\begin{bmatrix} D_{1,1} & \cdots & D_{1,M} \\ \vdots & \ddots & \vdots \\ D_{N,1} & \cdots & D_{N,M} \end{bmatrix}} \times \overline{\begin{bmatrix} S_1 \\ \vdots \\ S_M \end{bmatrix}} + \overline{C_{bgd}} + \overline{\begin{bmatrix} D'_{1,1} & \cdots & D'_{1,M} \\ \vdots & \ddots & \vdots \\ D'_{N,1} & \cdots & D'_{N,M} \end{bmatrix}} \times \overline{\begin{bmatrix} S'_1 \\ \vdots \\ S'_M \end{bmatrix}} \quad \textbf{(6a)}$$

Which in condensed notation gives:

$$\overline{C(x_i)} = \overline{D_{i,J}} \times \overline{S_J} + \overline{C_{bgd}} + \overline{D'_{i,J} \times S'_J} \quad (6b)$$

If the number of targets is equal to the number of sources, the problem can be solved by inversion of a linear system. If the number of targets is larger than the number of sources, the problem is a multiple linear regression type with unknowns $\overline{S_j}$ and $\overline{C_{bgd}}$. The third term on the right hand side of the **Eq. (6b)** is a bias which is *a priori* unknown and which we will evaluate in this study.

2.1.4 Source inference methods

The inferred sources, $\overline{S_i^{inferred}}$, were derived from **Eqns. (3)** or **(6)** assuming the covariance term (last term on right hand side) was null. The method used to infer the source was either a simple division (**Eq. (3)**) or an optimisation of the linear system using the linear model function *lm* in R (package stats, R version 3.2.3), with either $M = 1$ (single source) or $M = 9$ (multiple sources):

$$\overline{\begin{bmatrix} D_{1,1} & \cdots & D_{1,M} \\ \vdots & \ddots & \vdots \\ D_{N,1} & \cdots & D_{N,M} \end{bmatrix}} \times \overline{\begin{bmatrix} S_1^{inferred} \\ \vdots \\ S_M^{inferred} \end{bmatrix}} = \overline{\begin{bmatrix} C_1 \\ \vdots \\ C_N \end{bmatrix}} - \overline{C_{bgd}} \quad (7)$$

The bias δS_i was then evaluated as the difference between the inferred sources $\overline{S_i^{inferred}}$ and the modelled sources $\overline{S_i^{obs}}$ averaged over each period:

$$\delta S_i = \overline{S_i^{inferred}} - \overline{S_i^{obs}} \quad (8)$$

As shown in **Eqns. (3)** and **(6)** the overall mean bias δS_i contains (i) a bias term due to the inference method which is dependent mainly on the conditioning of the matrix D_{ij} (see supplementary material S2) and (ii) a bias term which is intrinsically linked to the covariance between D_{ij} and S_j (**Eqns. 3** et **6**). Thus, with **Eq. (8)** we evaluate the sum of the two biases without distinction. In order to infer the sources, the elements of the dispersion matrix D_{ij} need to be determined. The next part details how these were estimated with a dispersion model.

2.2 The dispersion model used for determining the transfer matrix D_{ij}

The elements of the transfer matrix $D_{ij} = D(x_i, S_j, t)$, **that is** by definition the concentration at location x_i and time t generated by a source S_j of strength $S_j(t) = 1$**, were calculated using a dispersion model..** The FIDES-3D model ("FIDES**",** Loubet et al., 2010), based on the analytical solution of the advection-diffusion equation of Philip (1959) was used for that purpose. This model was first compared with a backward Lagrangian Stochastic dispersion model (**bLS,** the "WindTrax" software, Thunder Beach Scientific, Nanaimo, Canada, Flesch et al.,

1995)**, and successively tuned to mimic the bLS.** The two models and how the FIDES model was tuned are briefly described hereafter and detailed in the supplementary material sections S3 and S4.

The FIDES model is based on the Philip (1959) solution of the advection-diffusion equation, which assumes power law profiles for the wind speed $U(z)$ and the vertical diffusivity K_z(z**) at height z.** This approach also assumes no chemical reactions in the atmosphere and spatial horizontal homogeneity of roughness length (z_0), wind speed (U), vertical and lateral diffusivity **(K_z and K_y).** The dispersion model is detailed in Huang (1979), and Loubet (2010). The details of the model and the way the transfer function $D(x_i, S_j, t)$ was estimated is detailed in the supplementary material S2.

The Schmidt number which is the ratio of momentum to scalar vertical diffusivity $Sc = Km_z / K_z$ is key in dispersion modelling, as it determines the vertical diffusion rate of scalars. Wilson (2015) demonstrated that bLS and dispersion models like FIDES give different values of Sc by constitution. In order to assure consistency of the **Philip** (1959) approach with bLS models, considered as references in dispersion modelling, we chose to tune the Philip (1959) model to get the same Sc number as in WindTrax as described by Flesch et al. (1995). The details are given in supplementary material S4. The comparison showed that the tuned FIDES model gives very similar concentrations to WindTrax at measurement heights lower than 2 m above the source, although slightly overestimated under stable and neutral conditions and slightly underestimated under unstable conditions. The correlation between the two models is however very high ($R^2 \geq \sim 0.96$) meaning that using the tuned FIDES model to characterise source inference performance, will lead to results **comparable** to WindTrax. Moreover since in this study the same model is used for predicting and for inferring the fluxes the results are self-consistent.

2.3 Ammonia sources from simple SVAT modelling and prescribed emission potentials

In order to evaluate the bias introduced by time averaging the concentrations when inferring single or multiple sources (third term in **Eqns. 3** and **6**), we generated NH_3 emission patterns mimicking the behaviour of real sources as closely as possible. In that prospect, we used the SurfAtm-NH_3 model developed by Personne et al. (2009) for two purposes: (i) evaluating the turbulence parameters (the friction velocity u_*, and the Monin Obukhov length L) from the meteorological datasets to parameterise the dispersion models, and (ii) providing the surface temperature $T(z_0)$ and the surface resistances in order to calculate ammonia emission patterns.

The SurfAtm-NH_3 model is a one-dimensional, bi-directional surface-vegetation-atmosphere-transfer (SVAT) model, which simulates the latent (LE) and sensible (H) heat fluxes, as well as the NH_3 fluxes between the biogenic surfaces and the atmosphere. It is a resistance analogue model separately treating the vegetation layer and the soil layer, and coupling a slightly modified (Choudhury and Monteith, 1988) model of energy balance and the two-layer bi-directional NH_3 exchange model of (Nemitz et al., 2000) with a water balance model. Unless otherwise stated, the surface was considered a bare soil with z_0 = 5 mm, **displacement height (**d**)** = 0 m, and **leaf area index (**LAI**)** = 0.

The ammonia emission patterns were modelled using the resistance approach and assuming atmospheric concentration was zero, which is a reasonable assumption following nitrogen application and leads to patterns mimicking reality, which is what we are seeking here:

$$F = \frac{C_{\mathrm{pground}}}{R_a(z_{ref}) + R_{bNH_3}} \qquad (9)$$

Where $R_a(z_{ref})$ is the aerodynamic resistance at the reference height $z_{ref} = 3.17$ m, and R_{bNH_3} is the soil boundary layer resistance for ammonia as described in Personne et al. (2009). The ground surface compensation point concentration (C_{pground}) was expressed as a function of Γ, the ratio of NH_4^+ to H^+ **concentrations** in the soil **liquid phase** at the surface, as in Loubet et al. (2012):

$$C_{\mathrm{pground}} = K_h\{T(\mathrm{z_0})\} \times K_d\{T(\mathrm{z_0})\} \times \Gamma = \Gamma \times 10^{-3.4362+0.0508\,T(\mathrm{z_0})} \quad (10)$$

where K_h and K_d are the Henry and the dissociation constant for NH_3 **respectively**, and $T(\mathrm{z_0})$ is the soil surface temperature. Since we wanted to evaluate the correlation between the transfer function D_{ij} and the source strength S_j, which is the bias in the inference problem (**Eq. 6**), the NH_3 volatilisation was modelled as to reproduce the variety of existing kinetics of NH_3 emissions from fields. In that prospect, three Γ patterns were simulated:

1. a constant $\Gamma = \Gamma_0$, which would mimic background NH_3 emissions from soils;
2. an exponentially decreasing $\Gamma = \Gamma_0 \exp(-4.6\, t / \tau_0)$, which best represents NH_3 emissions following slurry application ;
3. a Gaussian $\Gamma = N(\Gamma_0,\sigma_\Gamma)$, which would represent the typical NH_3 emissions following urea application.

Here Γ_0 is the maximum Γ during the period, t is the time in days, τ_0 is the duration of the emission in days. The factor 4.6 was chosen so that when $t = \tau_0$, Γ goes down to 1% of Γ_0. The duration of the emissions was chosen to be four weeks, $\tau_0 = 28$ days. **The time scale of the exponential decrease we used here was around 6 days, which is twice as large as the one reported by Massad et al. (2010) for slurry application (2.9 days).** While these Γ patterns gave the weekly trend of NH_3 emissions, the daily patterns were produced by the thermodynamical and turbulence drivers of NH_3 emissions which were explicitly taken into account through the compensation point (**Eq. 10**). To facilitate understanding, in most of the manuscript only the constant Γ was considered, and the effect of modifying the source strength was evaluated in a sensitivity study.

2.4 Spatial set up of the sources, concentration sensors

The sources (plots) were considered as squares with width x_{plot} and aligned south-north. Two configurations were considered: (1) a single source configuration and (2) a multiple-sources configuration which mimics typical agronomic trials with 9 sources (plots) placed next to each other, with three treatments times three repetitions. Each treatment was assigned a value of Γ_0 different from the others, while the three repetitions of the same treatment were assigned the same value of Γ. The concentration sensors (receptors) locations, x_i, were set in the middle of each plot, at several heights z_i. (**Figure 1**).

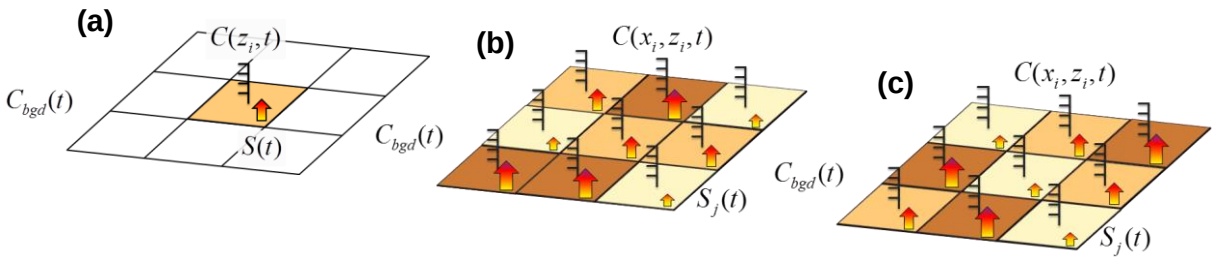

Figure 1. General scheme of the source receptor locations for (a) a single source, and (b) multiple-sources. (c) "optimum" plot layout used for the multiple-source configuration.

A number of plot sizes (x_{plot} = 25, 50, 100 and 200 m on the side), and receptor heights (z_i = 0.25, 0.5, 1 and 2 m), were tested successively. Several source strengths and dynamics were also tested: Γ was first considered constant with time (pattern 1) in all the plots , and the Γ_0 of each of the three treatments were either chosen to be significantly different in strength (10^4, 10^5, 10^6), or of the same order of magnitude (1000, 2000, 4000). Then the three Γ patterns ("constant", "exponential" and "Gaussian") were randomly assigned to the treatments for each simulation period. The ammonia background concentration, C_{bgd}, was considered constant and equal to 1 ppb except when studying the sensitivity of the inference method to the background concentration, where it was set as unknown. Throughout this study, an "optimum" block configuration was considered (shown in **Figure 1c**), which avoided trivial configurations like aligned blocks and maximised the mean distance between blocks **as in a Latin-square design.**

2.5 Simulation details

2.5.1 Meteorological data and fertiliser application periods

A range of meteorological conditions were simulated based on the half-hourly meteorological data of the FR-Gri ICOS site in 2008. In total 13 periods of 28 days were considered which spanned the whole year except the last two days of the year. Each period consisted of 1344 half-hourly data.

2.5.2 Concentration sensor integration periods

In order to evaluate the influence of the concentration averaging period on the source inference, several integration periods τ were tested: 0.5h (no integration), 3h, 6h, 12h, 24h, 48h, 168h (7 days). In practice the concentrations were computed at each sensor location using **Eq. (6)** over 0.5h: at that **time scale, which corresponds to the spectral-gap**, the covariance term is assumed to be negligible (Van der Hoven, 1957)**.** Then the averaged concentrations were computed for all integration periods.

2.5.3 Sensitivity to inferential methods scenarios

Several **scenarios** were considered and summarized in Table 1:

1) the background concentration $\overline{C_{bgd}}$ was either supposed known and fixed to the prescribed values (**C1-C4**) or was inferred (**C5-C7**);
2) the three repetitions of each treatment were either supposed to have the same source strength (**C2**, **C4**, **C5**, **C6**) or they were inferred independently (**C1**, **C3**, **C7**). In **C2**, **C4**, **C5** and **C6**, $S_i = S_m$ for all i and m belonging to the same treatment. In practice a new dispersion matrix was calculated by averaging

together all columns belonging to the same treatment (matrix dimension $N \times 3$). Three strength values of S were inferred to be tested;

3) either one concentration sensor at each source location (z_i) was considered (**C1**, **C2**, **C5**) or two sensors positioned at two heights were considered (**C3**, **C4**, **C6**, **C7**). All the measurement heights and their combinations were considered.

Table 1. Scenarios tested for inferring the sources and background concentration.

Strategy	Number of sensors	Plots# have same source strength in a given treatment	Background concentration	Note
C1	1	No	known	Each block is considered independently
C2	1	Yes	known	Each block is considered equal
C3	2	No	known	Identical to C1 except for the number of sensors
C4	2	Yes	known	Identical to C2 except for the number of sensors
C5	1	Yes	unknown	Identical to C2 except for the background concentration estimation
C6	2	Yes	unknown	Identical to C4 except for the background concentration estimation
C7	2	No	unknown	Identical to C3 except for the background concentration estimation

Each treatment have 3 plots (repetitions).

2.6 Statistical indicators

For each run the mean bias (BIAS) and the normalised mean bias (NBIAS), were calculated as: $BIAS_i = \frac{1}{N_\tau}\sum_\tau \delta cumS_i$, $NBIAS_i = BIAS_i / \left(\frac{1}{N_\tau}\sum_\tau cumS_i^{obs}\right)$, where N_τ is the number of the time averaged samples over each 28-day period and $cumS_i$ and $cumS_i^{obs}$ are the inferred and observed cumulated fluxes over the same period. The medians and interquartile of these statistical indicators were then calculated over the 13 periods of 28-days for 2008.

2.7 Real experimental test case

In order to evaluate the feasibility of the method we applied it to a real test case (**Figure 2**). The trial was located at La Chapelle Saint-Sauveur in France (47°26'44.1"N, 0°58'50.7'W') and performed from 5th April to 26th April 2011. Soil texture was loamy with a pH in water of 6.2 and a bulk density of 1.4 t m^{-3} in the first 15 cm. The experimental unit was composed by 6 squared sub-plots of 20 m wide with 2 repetitions of 3 treatments: (1) surface application of cattle slurry, (2) surface application and incorporation of the same slurry and (3) no application. Slurry pH was 7.5 with dry matter (DM) content of 6.05%, C:N ratio of 10.4 and contained 38.4 g N kg^{-1} (DM) as total nitrogen and 13.2 g N-NH_4 kg^{-1} (DM) as ammoniacal nitrogen. Slurry was applied on 5th April 2011 at a rate of 49 m^3 ha^{-1} which led to 119 kg N ha^{-1} and 41 kg N-NH_4 ha^{-1}. The application was identical between the two repetitions with a small standard deviation (< 0.2 kg N ha^{-1}). The incorporation was performed in two sub-plots one hour after the end of the slurry spreading with a disc harrower at a depth of 0.10 m. The soil humidity between 0 and 5 cm depth was homogeneous over the blocks and decreased from 20±1% to 17±1% w/w between the start and the end of the experiment. Meteorological data were measured at less than 50 m from the central plots (**Figure 2**). Air temperature, relative humidity, global solar radiation, wind velocity and direction were recorded every 30 minutes at 2 m height. The turbulence parameters (u_* and L), input of the dispersion models, were evaluated with a simple energy balance model of Holtslag and Van Ulden (1983) assuming a Bowen ratio of 0.5 and a deep soil temperature equal the averaged ambient temperature. Ammonia concentration was measured with diffusive samplers (ALPHA), (Sutton et al., 2001;

Tang et al., 2001; Tang et al., 2009), which were placed at the centre of each sub-plot at two heights (0.32 and 0.87 m from the ground) as well as next to the assay at three location (5 m away from the plots) at 3 m height. The ALPHA samplers were set in place just after slurry application and incorporation (between 14:20 and 14:50) and left exposed subsequently for **3h, 22h, 23h, 23h, 71h** (3 days) and **359h** (15 days) hence spanning 21 days. The diffusive samplers were prepared prior to the experiment, stored at 4°C in a refrigerator and analysed by colorimetry. Since no background concentrations were measured at a reasonable distance from the field, the background concentration was assumed as the minimum over the whole period of the concentrations measured on the 3 m height masts.

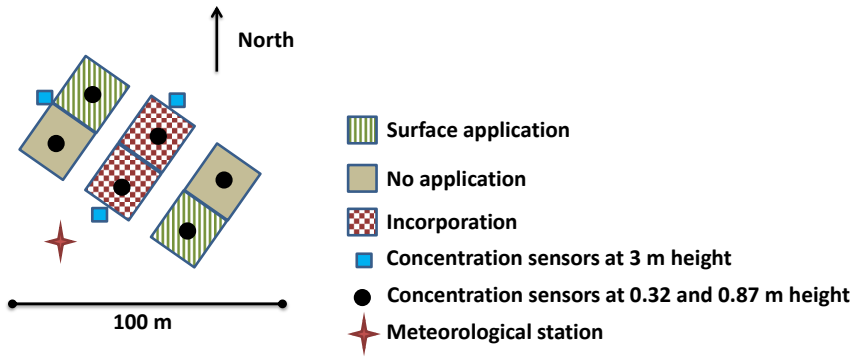

Figure 2. Scheme of the real experimental test case performed on 6 sub-plots with three treatments and two repetitions. Cattle slurry was either applied on the surface or incorporated. The concentration sensor and meteorological station locations are shown on the scheme.

3 Results and discussion

3.1 Meteorological data range and simulated ammonia sources

The meteorological conditions over the 13 periods represented a good sample of temperate climate conditions. **The friction velocity u_* varied between 0.024 and 1.181 m s^{-1},** and the stability parameter z/L **at 1m height varied between** -49 **and** 21 (**Figure 3**). It is noticeable that u_* showed greater variability during the winter than during the summer, while it was the opposite for z/L. The surface temperature also showed a structure varying between periods, with a larger temperature range during the summer (from 5.7 to 50.4°C) than during the winter (from -5.2 to 22.9°C). This surface temperature variability is an essential feature to representing real case ammonia sources (Sutton et al., 2009), which shows a variability reflecting both the surface temperature and the resistances variations (**Eqns. 9** and **10**).

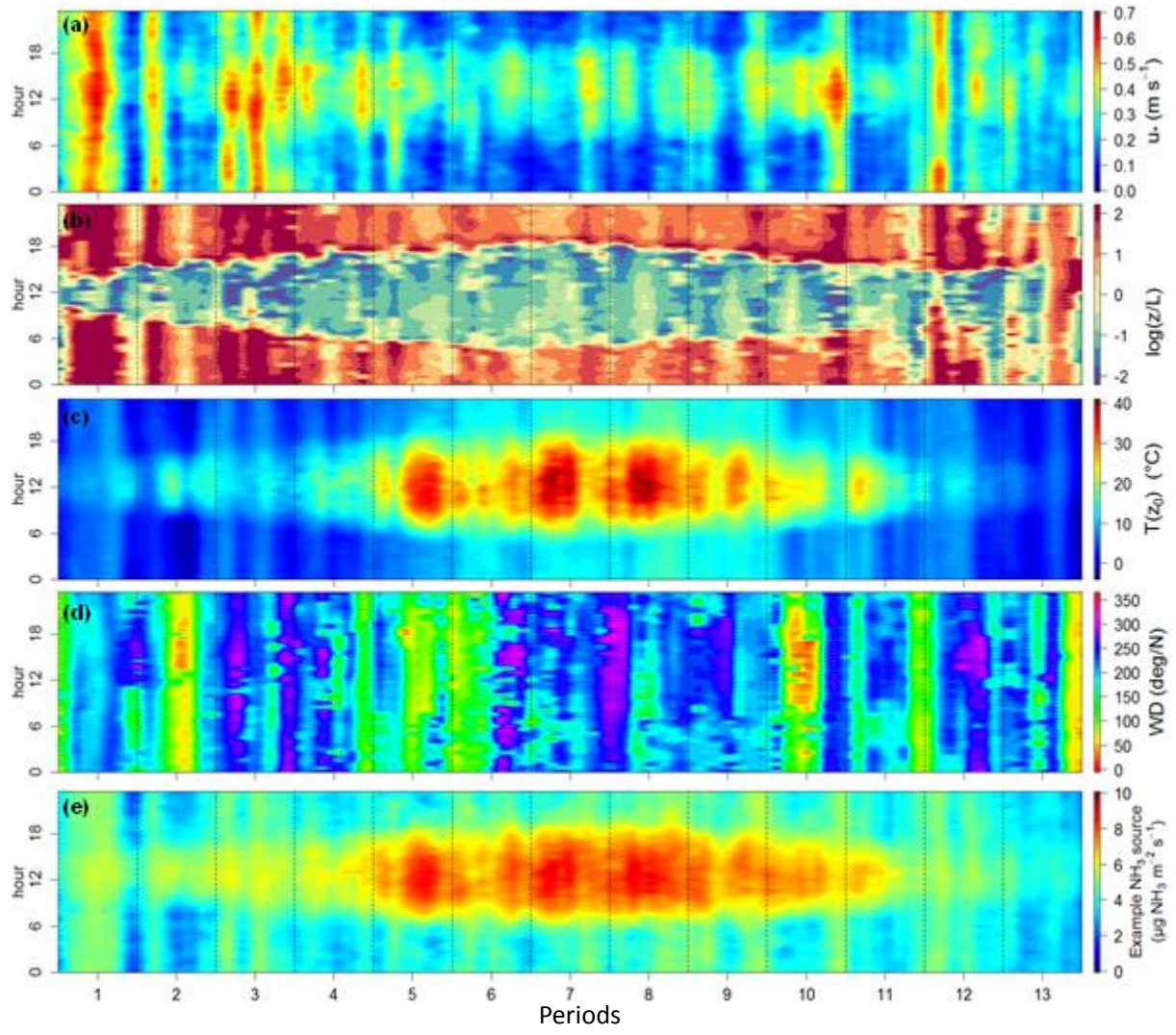

Figure 3. Footprints of measured u_* (a), *z/L at 1 m height* (b), $T(z_0)$ (c), and wind direction (d) for the hour of the day and the 13 considered periods over year 2008 in the FR-GRI ICOS site. The modelled ammonia source is also reported (e) according to Eqns. (9) and (10) over the same period with an emission potential Γ = 10000.

3.2 Example ammonia concentration dynamics modelled with the tuned FIDES model

The modelled ammonia concentrations reproduced typical patterns measured above field following nitrogen application well, with maximum concentrations during the day and minimum concentrations at night (**Figure 4**). These patterns are a consequence of daily variations of the sources driven by surface temperature combined with variations in the aerodynamic transfer function D_{ij}, which behaves similarly as a transfer resistance (see supplementary material S1). The integration periods are also shown in **Figure 4,** which illustrates the progressive loss of information of the pattern structure with integration periods. Particularly, it can be seen that the day-to-night variation is captured up to an integration period of 6h. Moreover, it should be noted that averaging also means overestimating lower concentrations and underestimating higher concentrations.

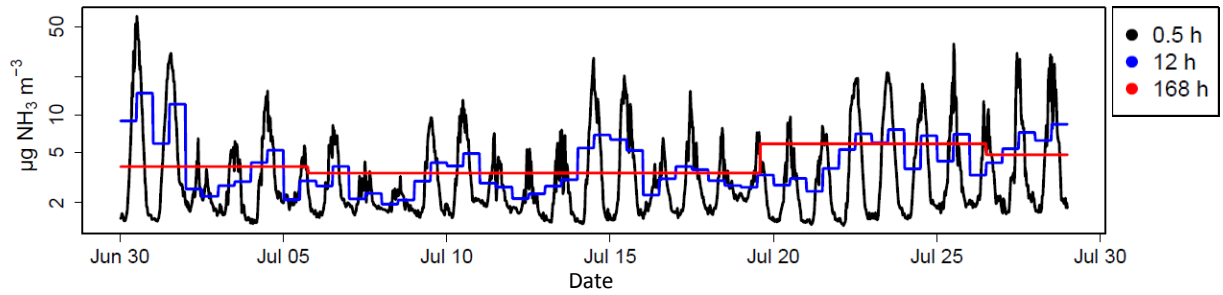

Figure 4. Example modelled concentration pattern at 1 m above a single 50 m width source for several averaging periods (0.5h, 12h and 168h) for the month of July 2008. The source Γ was set to 10^5. The y-axis is log scaled.

3.3 Evaluation of the inference method for a single source and a single sensor

At first we evaluate the bias of the inference method for the simpler case of a single source and a single sensor placed in the centre of the source field at several heights, assuming we know the background concentration (strategy C1; **Figure 1a.**). This case has the advantage of having a condition number equal to 1 (**Supplementary material section S2 and Eq. S1)** and a bias δS which is well defined and equal to $-[\overline{D}]^{-1} \times \left[\overline{D'S'}\right]$ (**Eq. (8)**). This section hence **focuses** on evaluating the influence of sensor height, time integration, and source dimension on the bias without dealing with the complexity of the interactions between multiple fields.

3.3.1 Example inferred source dynamics

Figure 5 reports an example source inference, which shows the progressive smoothing of the source with integration period. We first see that the source strength corresponding to $\Gamma = \mathbf{10^5}$ leads to ammonia emissions ranging from 0 to ~1 µg NH_3 m^{-2} s^{-1} in the winter, which corresponds to 0.71 kg N ha^{-1} day^{-1}. Over the entire year, the maximum emission occurs during the hottest days and reaches up to 7.1 kg N ha^{-1} day^{-1}. Regarding the inference method, it can be seen in that example that up to 24 hours the variability in emissions over the period is captured quite well.

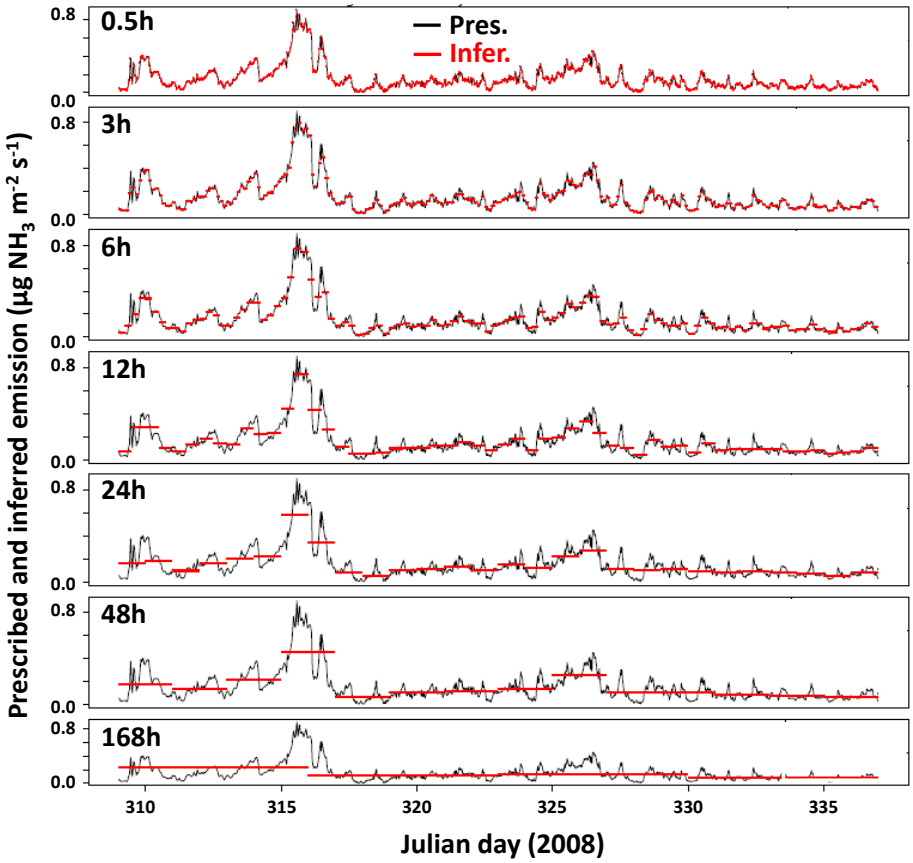

Figure 5. Example source inference for a 25 m width square field and a concentration sensor placed at 0.5 m above ground. Here Γ = 10^5 and is set to constant (pattern 1). The 7 integration periods are shown: 0.5h to 168h. The x-axis shows the day of year and corresponds to a span over November. The prescribed source is in black (Obs.) and the inferred one in red (Pred.)

3.3.2 Effect of target height, source dimension and integration period on the bias δS for a single source

In this simpler case shown in **Figure 6**, the fractional bias of the inferred emission is mostly negative for the combination where the ratio sensor height / plot dimension is small and integration times are larger than 6h. According to **Eq. (5)**, this means that the covariance term $\overline{D'S'}$ is negative for these conditions, meaning that any increase in source strength *S* at a time *t* is correlated with a decrease of the transfer function *D(t)* and vice versa. This is expected as *S(t)* increases with the surface temperature (**Eq. (10)**) and is proportional to $\left[R_a\left(z_{ref}\right) + R_{bNH_3}\right]^{-1}$ (**Eq. (9)**), while *D*(*t*) is proportional to the aerodynamic resistance $R_a(z_{ref})$, as shown in supplementary material S1. Hence, over daily periods, *S* and *D* are negatively correlated: *S* increases during the day and decreases at night (due to temperature and wind speed daily patterns), while *D* decreases during the day and increases at night (mainly due to wind speed patterns). This is expected to be a general feature for NH_3 surface fluxes as the daily variability reproduced by the model used in this study is representative of most situations from mineral and organic fertilisation, to urine patches or seabird colonies (Ferrara et al., 2014; Flechard et al., 2013; Milford et al., 2001; Moring et al., 2016; Personne et al., 2015; Riddick et al., 2014; Sutton et al., 2013).

The median bias δS_i tends to increase in magnitude with the sensor height for large fields (x_{plot} =100 and 200 m) whilst decreases for smaller fields (x_{plot} = 25 and 50) when sensor height gets close to the field boundary layer

height. Furthermore, δS_i becomes positive and very large when sensors get above the field boundary layer height (**Figure 6**). For large fields, the increase of the magnitude of the bias with lower sensor height is expected as *D* decreases with height in absolute value. For small fields, the decrease of the bias corresponds to a loss of information as *D* gets close to zero when the sensor gets closer to the field boundary layer height. For heights above this limit, we observe a change in sign of the bias which can be explained by the fact that the sensor concentration footprint is not in the source during stable conditions (at night) while it is in the source under unstable conditions during the day. The inference method will hence not work if at least one sensor is not below the plot boundary layer height.

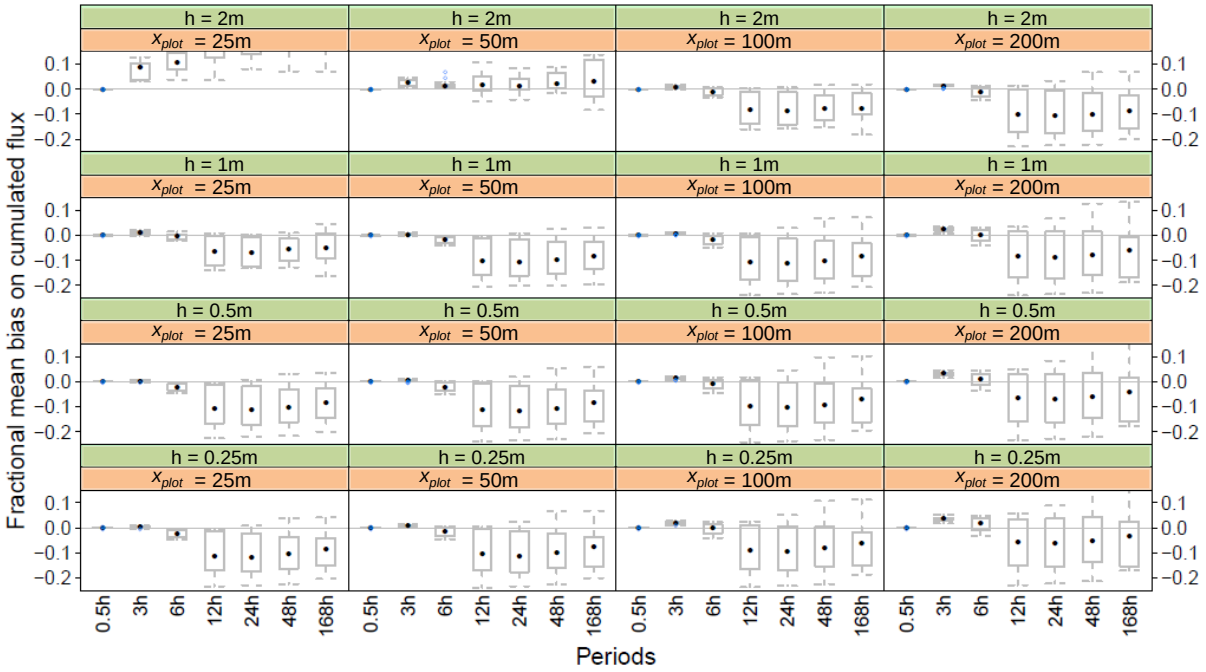

Figure 6. Fractional bias of inferred cumulated ammonia emission for a single squared field of side (x_{plot}) 25, 50, 100 and 200 m and sensors heights (h) 0.25, 0.5, 1 and 2 m, as a function of sensors integrating periods. The points show the median, the boxes the interquartile and the whiskers the maximum and minimum over the 13 application periods.

We also notice that for integration periods **equal or** below 3h, the fractional bias is slightly positive, which can be explained by the positive correlation between *S* and *D* at small time scales. This is because of the influence of u_* on $T(z_0)$: for a given solar radiation and air temperature over small time scales (< 3h), an increase in u_* leads to a decrease in $T(z_0)$, which leads to an exponential increase of the surface compensation point according to **Eq. (10)**. However, at the same time, $R_a(z)^{-1}$ decreases, but linearly with u_*. The resulting ammonia emission calculated with **Eq. (9)** nevertheless increases because the exponential effect of temperature overcomes the linear effect of the exchange velocity (data not shown). This effect is more visible for large fields than small fields because over small fields an additional effect is that when u_* decreases, the footprint increases and the source "seen" by the targets hence decreases because it incorporates a fraction of zero emission sources.

Overall, the median fractional bias for weekly integrated emissions over a 25 m field and sensor heights below 0.5 m was overall -8% with an interquartile (-14% to -2%). We can conclude that the bias of the NH_3 emissions is reproducible within ± 6%. We can also conclude that it would be better to place the concentration sensor at a low height to minimise the bias of the method.

3.3.3 Effect of surface boundary layer turbulence on the inference method for a single source

The inference method depends on the turbulence at the site and especially on the main drivers of the dispersion which are the friction velocity and the stability regime. Indeed **Figure 7** shows that the relative root mean square residual of the inferred source (RRMSR) decreases with increasing u_* at long integration periods and is larger in slightly stable than near-neutral or slightly unstable conditions. **Figure 7** also shows that the under stable conditions or low u_* the RRMSR increases by more than an order of magnitude (up to 50%) when integration periods increase from **6h** to **12h**, which catches most of the source variance. We also see that under near-neutral or high u_* conditions, the 3rd quartile of the RRMSR remains below 10% for all integration periods. Finally, we also see that the larger 3rd quartiles at short integration periods are obtained with intermediate u_* values or slightly unstable conditions. A similar response of the bias to u_* and $1/L$ was reported by Figure 6 in (Flesch et al., 2004) and Figure 3 in Gao et al. (2009) in controlled source experiments. While Gao et al. (2009) attributed the bias of the inference method to parameterisation of the stability dependence of the turbulent parameters (z/L), in this study this cannot happen since we use the same parameterisation for prescribing the concentration and inferring it. In our case, the interpretation is to be linked with **Eq. (5)**: the smaller u_* or the most stable conditions also correspond to the larger time-derivatives of source strength (driven by surface temperature and surface exchange resistances) as well as the larger time-derivatives of transfer function D. We hence expect that under such conditions, the covariance between the transfer function and the source strength will be larger than under near-neutral conditions. In a more heuristic view, under low turbulence, large time-derivatives of concentrations are expected above a source due to low mixing (small changes in mixing lead to large variations in concentrations).

We conclude that the inference method with a long integration period will lead to very moderate biases for locations with near-neutral conditions and high wind speed, but may lead to much larger bias under stable conditions and low wind speed as soon as the integration period gets up to **12h**.

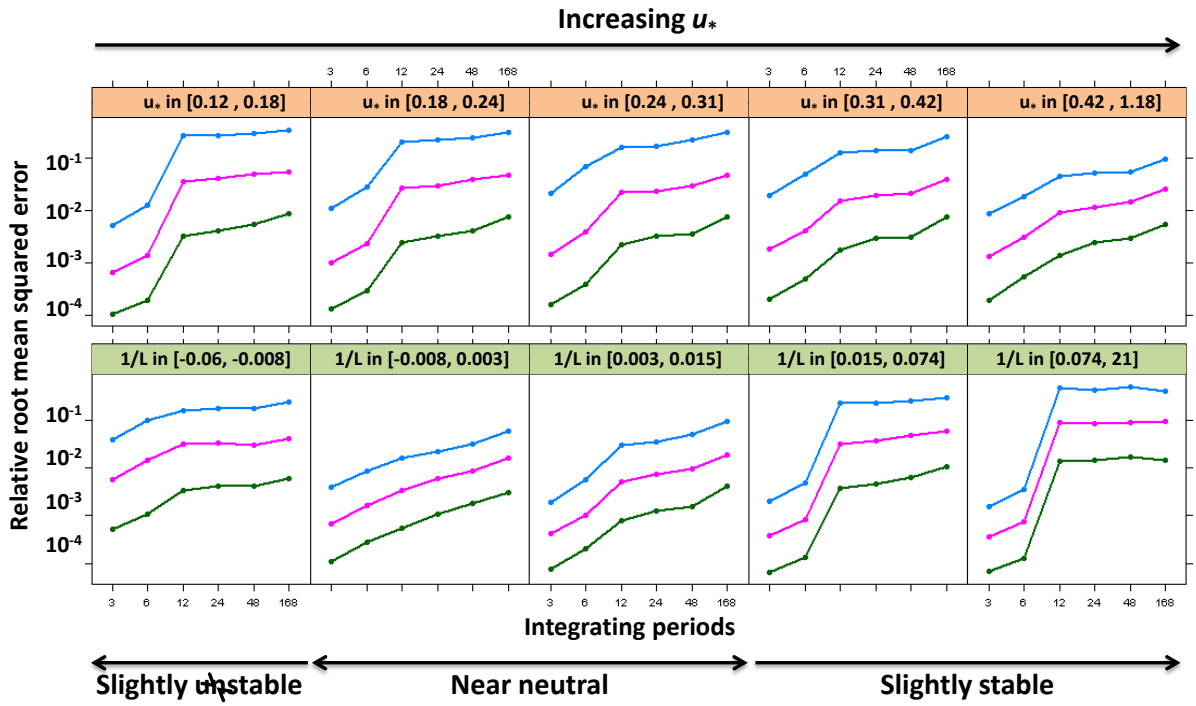

Figure 7. Relative root mean squared error as a function of integration period for stability factor and friction velocity classes for a single 25 m side field. Medians and quartiles are given for equally sized bins of u_* and $1/L$ and for the lowest sensor height (0.25 m). The blue, pink and green curves are the 3rd, 2nd and 1st quartiles, respectively.

3.4 Multiple source case

In contrast to the single source case, with multiple sources (see **Figure 1b**) the inference method leads to biases at small integration times as can be seen in the example reported in **Figure 8**. In that specific case, the emissions of treatments-2 **($\Gamma = 10^5$)** and 3 **($\Gamma = 10^6$)** are 10 times and 100 times larger than that of treatment-1 **($\Gamma = 10^4$),** respectively. This leads to concentrations over plots of treatment-1 (and to a lesser extent over those of treatment-2) being highly correlated to emissions from plots of treatment-3 (and hence less with sub-plots of treatment-1). As a result, inferring emissions of plots of treatment 1 becomes harder as soon as averaging periods become larger or equal to 3h. This can be viewed as a progressive loss of information of the treatment-1 contribution to concentrations due to the overweighing contribution of treatment 3 plots. However, we also see that treatments 2 and 3 seem quite correctly inferred for integration times smaller than 48h.

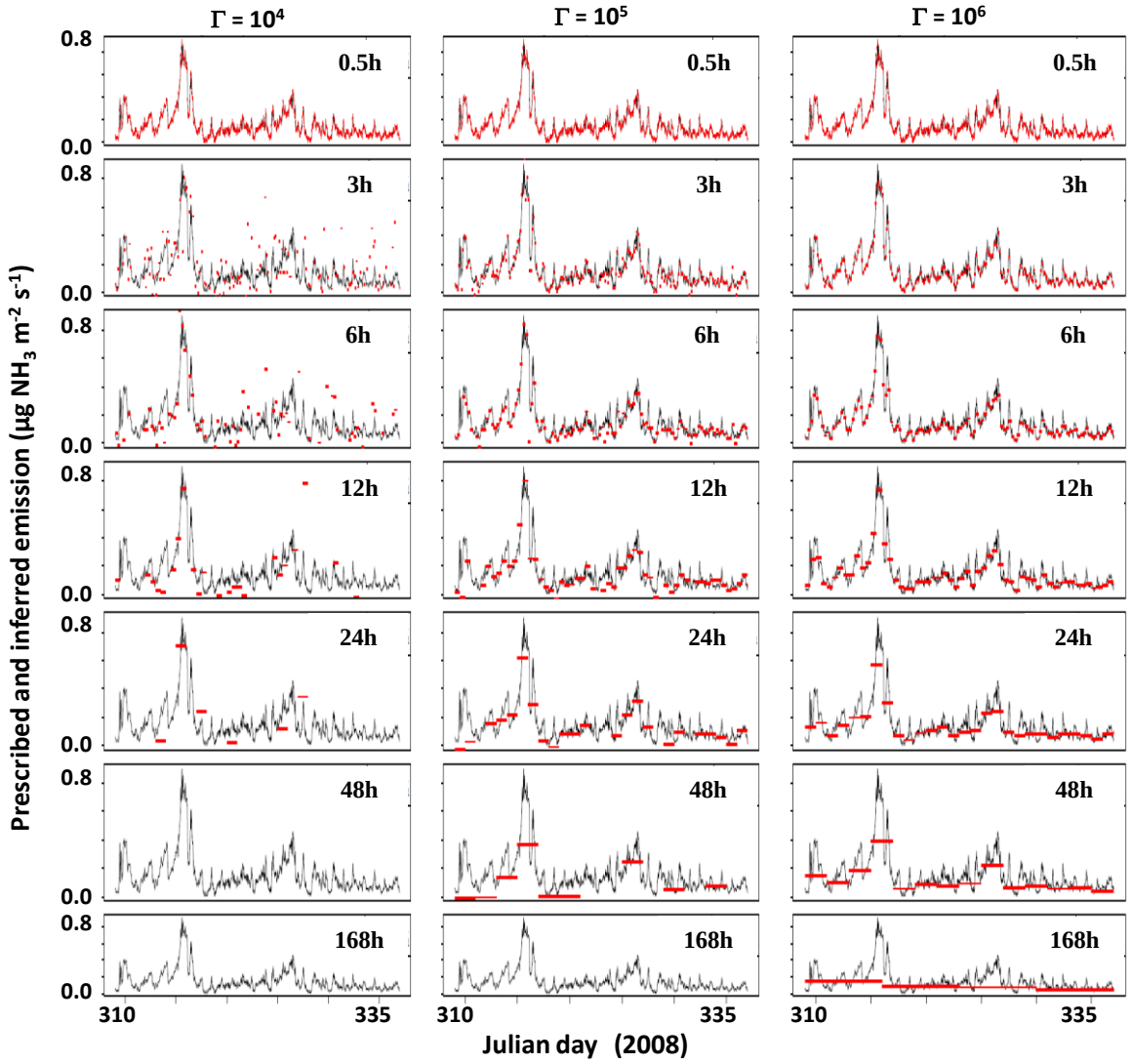

Figure 8. Example result of multiple plot case inference. Black curves: observations; red dots: inferred sources. Left: treatment-1, Γ = 10^4. Middle: treatment-2, Γ = 10^5. Right: treatment-3, Γ = 10^6. Missing red dots are out of the y-scale boundaries. Example plots from treatments 1, 2 and 3 are shown from left to right. The period is the same as in Figure 7 (November 2008 for the FR-Gri ICOS site), and emissions are up to 1, 10 and 100 µg NH_3 m^{-2} s^{-1}, for the three emission potentials. Strategy C7 with target heights 0.25 and 2 m, and source width 25 m on a side.

In the following we will first evaluate the influence of the length of integration periods, sensor heights and plots dimensions on the fractional biases made when inferring the source. Each factor will be evaluated independently of the others in order to understand the processes behind it. For these evaluations background concentration was kept constant at 1 µg NH_3 m^{-3}. Strategy C1 was used except when testing sensor heights for which strategy C3, which uses two targets, was also used. These two strategies assume that the background concentration is known which avoids any compensating effects between source and background concentration inferences. Then the sensitivity of the methodology to the (i) emission ratios between two of the three treatments and (ii) the variability in the background concentration were evaluated. Finally, seven inversion strategies were compared to determine which was the most robust (**Table 1**).

3.4.1 Effect of integration periods on the bias

We first consider strategy C1, which is the simplest configuration, in which plots are independent, background concentration is known and one target is used above each plot. **Figure 9** shows that for the given treatment range (~1-10-100 µg NH_3 m^{-2} s^{-1}), the fractional mean bias is lower than 0.2 in magnitude for the treatment emitting the most (treatment 3, **$\Gamma = 10^6$**), lower than 0.4 for the intermediate treatment (treatment-2, **$\Gamma = 10^5$**) and up to 8 for the treatment emitting the least (treatment-1, **$\Gamma = 10^5$**); here we considered the 0.25-0.75 quantiles. The bias of the highest treatment (treatment -3) actually behaves similarly to a single source case (**Figure 6**), with a median bias around 10% for 48h integration periods. This is expected because treatment-1 and treatment-2 have much smaller emission strength and hence little influence on the concentration above the treatment-3 plots, which therefore behaves in a similar manner to a single source. As a consequence, this bias in treatment-3 is mainly due to the anti-correlation between *D* and *S* which increases with integration periods. The fractional **mean** bias is very large for treatment-1 even for small integration periods. The bias can either be positive or negative showing that this method does not allow for a correct estimation of the smallest sources.

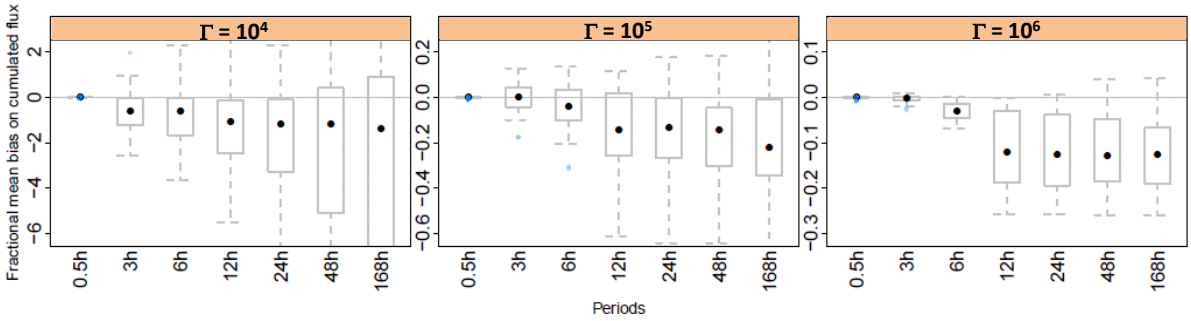

Figure 9. Effect of integration period on source inference in a multiple-plot setup. The fractional mean bias of the source is shown for each treatment. Inference strategy C1 was used (single sensor, independent blocks, background concentration known). Statistics for runs with target heights 0.25 and 0.5 m and source width = 25 m are calculated. All application periods are considered. Filled points show medians, boxes show interquartiles and bars show minimums and maximums. Outliers are points to 1.5 times away from boxes limits.

3.4.2 Effect of target heights on the bias

Figure 10 shows that the bias remains **low** as long as sensor heights are low enough to catch a sufficient part of the field footprint. When only a single height is used (strategy C1) this means that the sensor should be placed at 0.5 m or below for the field size we have tested here (25 m**). The result is similar for a pair of sensors (strategy C3). For the lowest treatment though, the bias (and its variability) remain high whatever the heights. It is interesting to notice that the heights which were found to provide an optimal inference of NH_3 sources (below 0.5 m) are smaller than ZINST reported by Wilson et al.** (1982) **(which were 0.9 m for 40 m diameter circular sources, and which we estimate as 0.65 m based on a power law extrapolation as in Laubach et al.,** 2012)**. It is also important to notice that this height should vary with both the roughness length *z0* and displacement height as was showed by Wilson et al.** (1982) **for ZINST.**

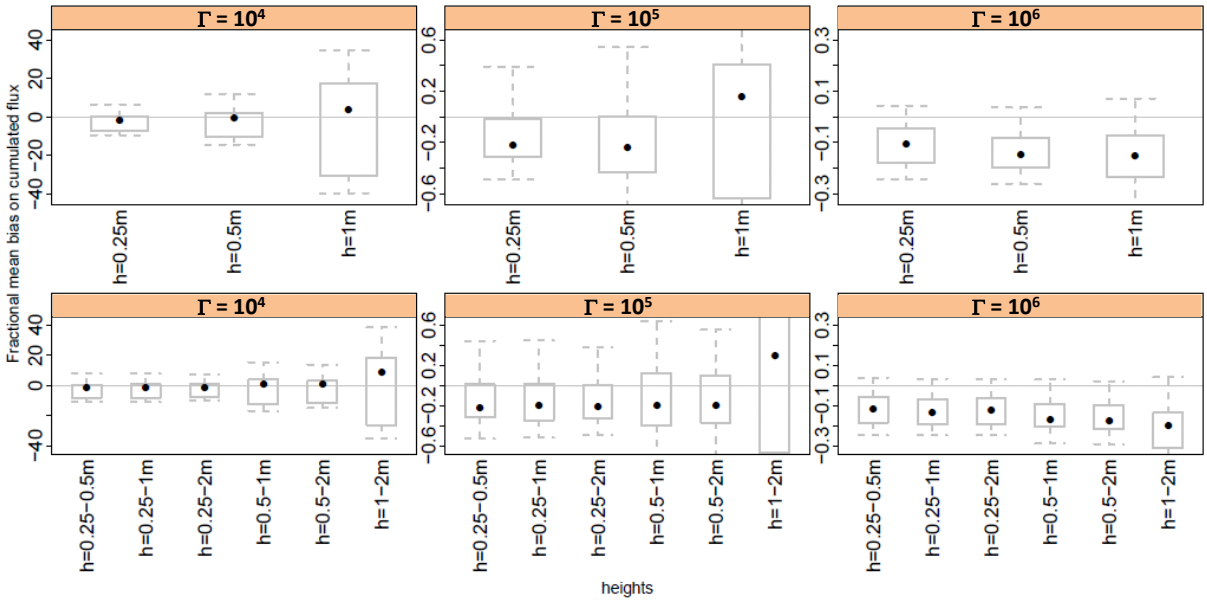

Figure 10. Effect of target heights on source inference in a multiple-plot setup for integration periods of one week (168h). Same as the case reported for Figure 9 except that strategies C1 (with a single sensor, top graphs) and C3 (with two heights, bottom graphs) are compared here (the background is assumed known in both strategies).

3.4.3. Effect of plot size on the bias

Increasing the plot size from 25 to 200 m width reduces the bias of the two **highest** source treatments for which the median bias reaches values around 10%, while the interquartiles remain stable (**Figure 11**). On the contrary, in treatment-1 **($\Gamma = 10^4$)**, the bias increases. It is expected that the bias in a multiple-source configuration never becomes smaller than the bias in a single source problem which is a limit linked to the time-integration (covariance between the source and the concentration, see **Eqns. 3 and 6**). It is also expected that the biases remain higher than the single source case until the source size increases sufficiently so that the concentration generated by a block on the neighbour fields become negligible compared to the concentration generated by the source below. This is what we observe in treatment-2 **($\Gamma = 10^5$)** and treatment-3 **($\Gamma = 10^6$), with treatment**-2 showing a median bias of -13% (larger than in the single source case) for the 200 m **wide** field, while the bias of the **largest** source tends to be -10% [-17%, -1%], which is the range observed for a single source.

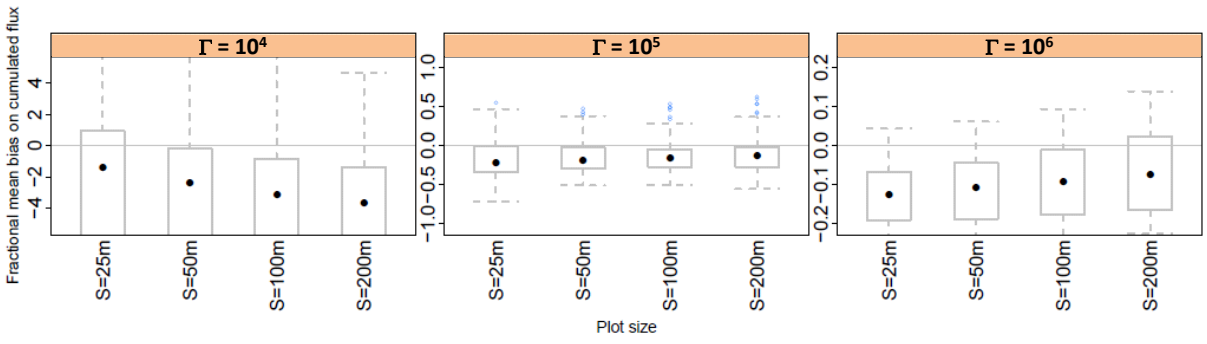

Figure 11. Effect of plots size on source inference in a multiple-plot setup for integration periods of 168h and target heights 0.25 and 0.5 m. Same as in Figure 8.

3.4.4 Sensitivity of the method to ratios of emission potentials between treatments

A central question is the capability of the inference method to resolve small or large differences in emissions from the nearby blocks. Indeed, we can speculate that small differences will be hard to resolve while large differences will lead to large bias. In order to determine the resolution power of the method, we compared the performance of the inference method with a set of three treatments: the first treatment had $\Gamma = 0$ to mimic a reference field receiving no nitrogen**:** the second treatment had a constant $\Gamma = 1000$ corresponding to a small emission (0.7 kg N ha^{-1} day^{-1}), while in the third treatment Γ was successively set to increasing values from 1500 to 10^5 (70 kg N ha^{-1} day^{-1}). In this section we consider that the background is known (sensitivity to the background concentration will be evaluated in the next section).

Figure 12 shows the median and interquartile biases of the cumulated emissions for the longest integration period 168h over the ratio of the high-to-low source treatments. The bias of the **largest** source always remained around 14%, which is larger than the single source case. The bias of the lowest source increased with increasing inter-treatments source ratio from 13% to 40%. In fact we find that the fractional bias increased approximately as a power function of the ratio of the two predicted sources (dotted lines, $0.11\ x^{0.256}$).

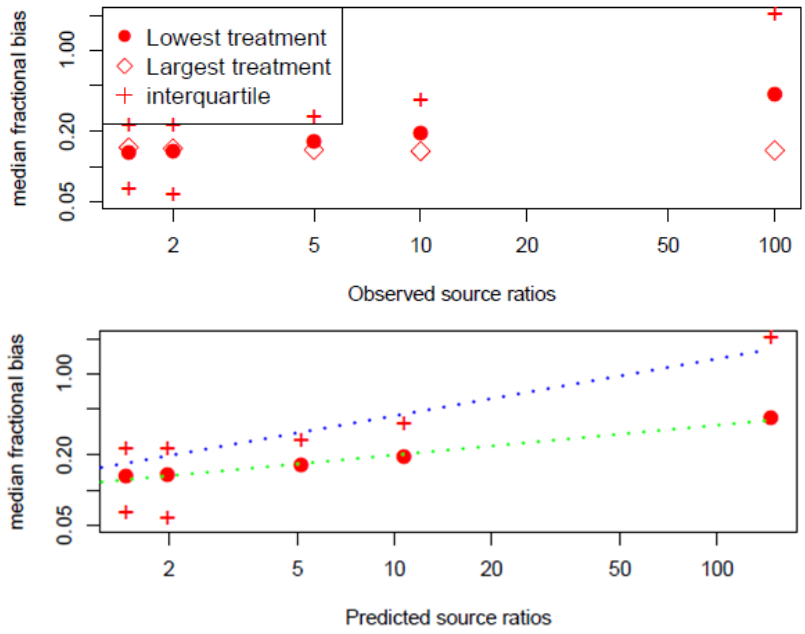

Figure 12. Median fractional bias of cumulated emissions as a function of the ratio of the high-to-low source treatments for a 7 days integration period. Top: bias as a function of the theoretical source ratios. Bottom: bias as a function of the predicted source ratios. Dotted lines show power functions regressions on medians (green) and interquartile (blue). Strategies C1 and C3 are pooled together with all runs including sensor heights 0.25 and 0.5 m

3.4.5 Quality of background concentration estimations

As pointed out by Flesch et al. (2004), the knowledge of the background concentration is essential in a source inference problem. Retrieving the background necessitates having at least $N_{sources}+1$ sensors. Hence only strategies with two heights per plot or which assume identical emissions in treatment repetitions can be evaluated in their capacity of retrieving the background (strategy C2 to C7). In order to evaluate the sensitivity of the method when the background concentration varies with time, we set a realistic background concentration as a

linear combination of u_* and air temperature (T_a) with a mean of 6 µg NH_3 m^{-3} and a standard deviation of 0.1 µg NH_3 m^{-3}. This test was performed with a range of treatments in order to elucidate the correlations between varying background and varying treatments. We see in **Figure 13** that the concentration, which follows a realistic pattern, is well retrieved even over the longest **integration** period **of 168h**. However, we see that for the treatments with the largest source contrast ($\Gamma = 1000$ and 10^5), the background concentration can be overestimated even for small integration periods (6h). The median residual of the background concentration was smaller in magnitude than 0.05 µg **NH_3** m^{-3}, except for the case with very large differences between treatments (0, 1000, 10000), for which the residual reached 0.1 and 0.5 **µg NH_3 m^{-3}** for the 6h and 24h/168h integration periods. Furthermore, the background concentrations were overestimated for the largest source ratios and underestimated for the lowest source ratios and longer integration periods (24h and 168h).

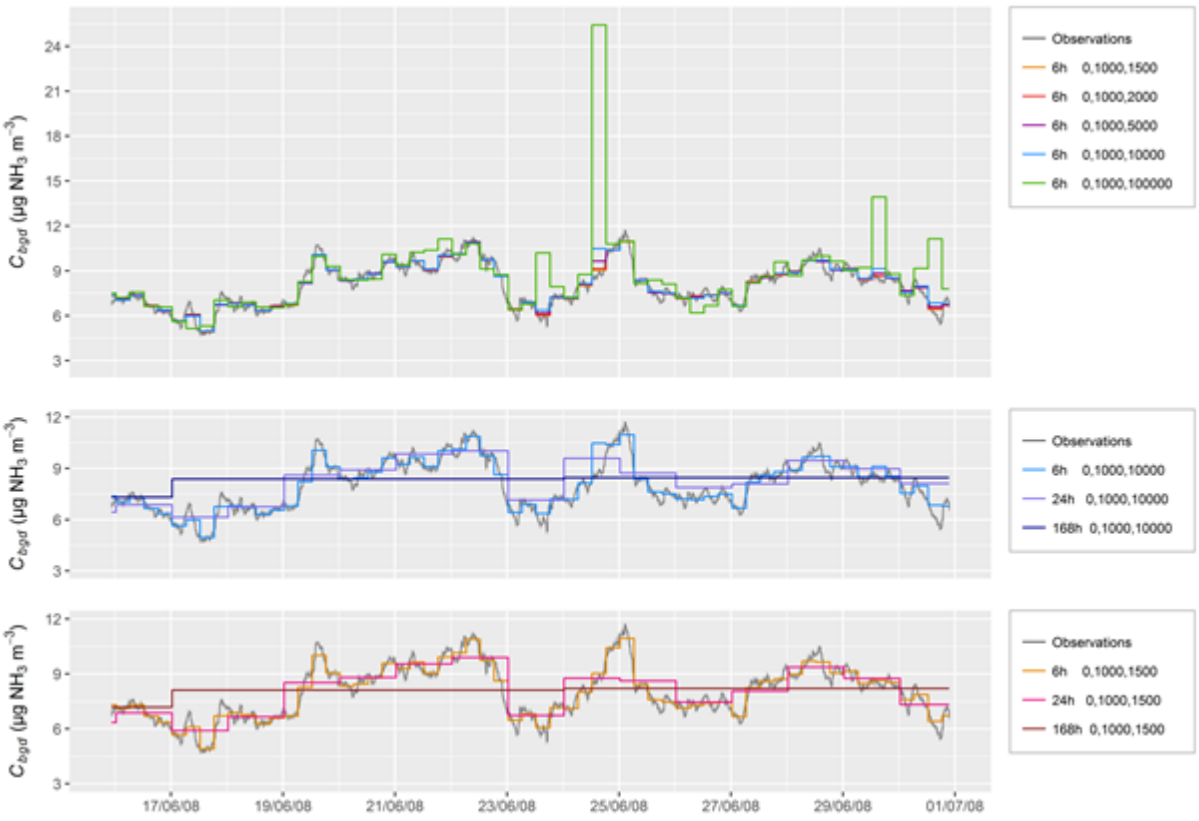

Figure 13. Background concentrations prescribed (Observation) and inferred using strategy C7 and height combination (0.25 m, 2 m): (a) effect of the treatment contrasts for a short integration period of 6h (treatments 1, 2 and 3 are given; (b) effect of integration period for contrasted treatments (Γ = 0, 1000, 10000); (c) effect of integration period for similar treatments (Γ = 0, 1000, 1500).

3.4.6 Identifying the most robust strategy

Finally to identify which strategy is the most suitable for retrieving the emissions **from the multiplot configuration, we compared** all strategies on a simulation with a variable background (set as in the previous section) and two sources ratios of 2 and 20 between treatments 2 and 3 (treatment-1 **b**eing a zero source reference). We found, as expected, that strategies with known backgrounds have low biases compared to strategies that calculate the background**,** except for the strategy C7 which provided biases similar to strategy C3

which is the strategy equivalent to C7 but with known background (**Figure 14**). We also see that incorporating some knowledge of the sources by assuming plots from the same treatment have the same emissions, gave slightly better estimates when the background is known (strategies C2 and C4 compared to C3). This is however not true when the background is unknown, in which case the magnitude of the bias increases up to a median of 0.7 (strategies C5 and C6 compared to C7). It is due to compensation between background concentration and source strength as we have seen in **Figure 14,** that the background concentration was overestimated in such cases. We also see, as expected, that the strategies with two sensors placed at different heights above each plot lead to better evaluations of the emissions. Overall, the strategy based on two sensors above each plot, which also assumes that sources are independent, seems to be the most robust (strategy C7). This strategy does not assume the background is known, nor does it assume the plots have similar emissions, which is more adapted to reality. Indeed, even though the same amount of nitrogen is applied in each repetition plot, the emission may vary due to soil heterogeneity and advection. We finally get a median bias for strategy C7 which is -16% with an interquartile [-8% -22%]. It is important to stress though that the minimums and maximums are further away, which indicates that under some rarer circumstances, the method may overestimate the sources by 12% or underestimate them by 40%. These cases correspond to integration periods with very low wind speeds and stable conditions.

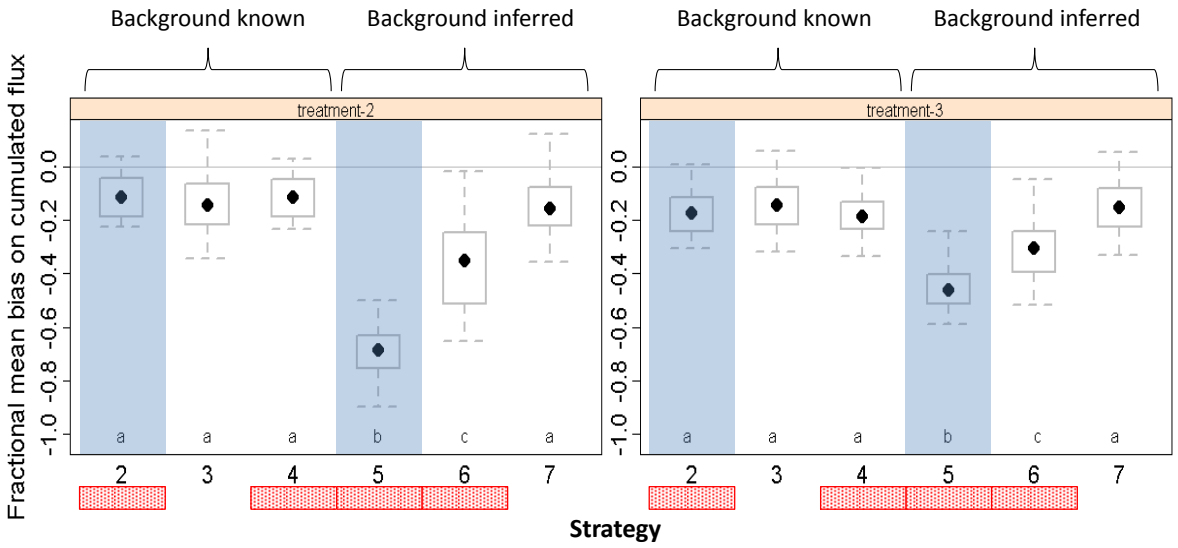

Figure 14. Comparison of biases for all source inference strategies. In strategies C2, C3 andC4 we hypothesize that we have perfect knowledge of the background concentrations, while in strategies C5, C6 and C7 background concentrations are inferred together with the sources. In strategies C2, C4, C5 and C6 (red rectangles) we suppose that plots from the same treatment have the same emissions, while in strategy C3 and C7 we infer each plot separately. In strategy C2 and C5 we assume single sensors are placed above each plot (blue shades), while in strategies C3, C4, C6, C7 we assume two sensors are placed above each plot.

3.5 Application of the methodology to a real test case with multiple treatments

The evaluation of the methodology on a real test case is shown in **Figures 15-17**. The concentration measured above the different treatments shows a much higher concentration above the surface applied slurry (up to 200 µg $N\text{-}NH_3$ m^{-3}) than above the two other treatments (below 50 µg $N\text{-}NH_3$ m^{-3}), (**Figure 15**).

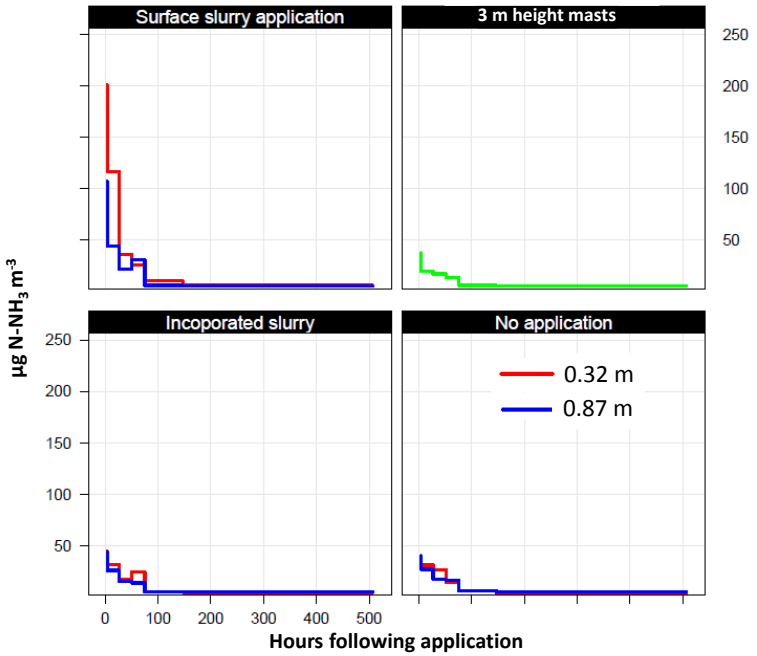

Figure 15. Concentrations measured in a real test case with 6 blocks composed of three treatments and two repetitions. Here the mean concentration for the repetition and the three replicates ALPHA samplers are shown at two heights above ground. The concentration measured at 3 m height at 5 m away from the plots is also shown in green. The background concentration, evaluated as the minimum of the green curve was 5 µg N-NH_3 m^{-3}.

The inference method gives very consistent results both in terms of comparison between repetitions **(B1 and B2)** of a given treatment and in terms of comparison between treatments (Strategy C7 shown in **Figure 16**). Surface slurry application **showed the largest emissions: 9 ± 0.3 kg N ha^{-1} in B1 and 10 ± 0.2 kg N ha^{-1} in B2 (median and confidence interval).** This corresponds to an emission factor around **24**% of the N-NH_4 applied and 8% of the total N applied, which is in-line with agronomic references (Sintermann et al., 2011a; Sommer et al., 2006). In contrast, the incorporated slurry showed much smaller **emissions: 0.3** ± 0.2 kg N ha^{-1} **in B1 and 0.6 ± 0.2 kg N ha^{-1} in B2. It is noticeable that** the no-application **showed slight deposition, especially in B2: -0.26 ± 0.2 kg N ha^{-1} in B1 and -1.7 ± 0.2 kg N ha^{-1} in B2.**

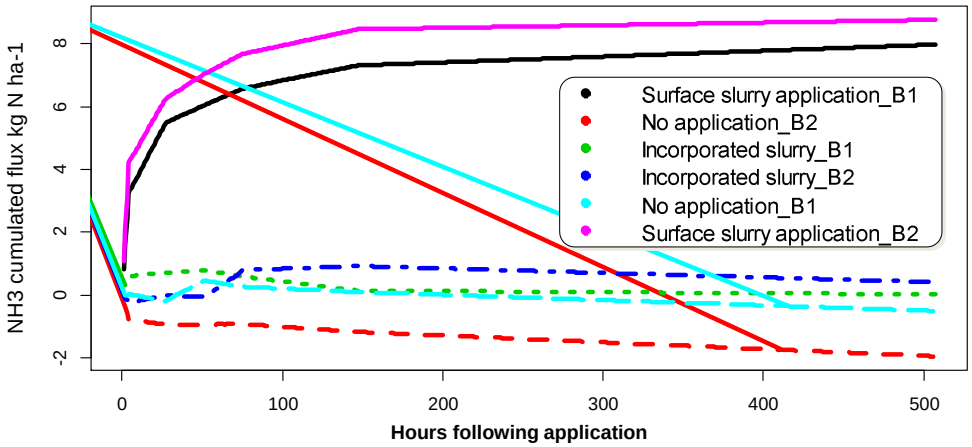

Figure 16. Cumulated fluxes estimated with the inference method on the real test case with strategy C7. Three treatments with two repetitions are compared (b1 and B2).

Comparing the inference strategies is instructive (**Figure 17**). We see that in methods which assume a known background (strategies C3 and C4), the inferred emissions are **slightly** higher than when background is assumed unknown. We should remind that we set the background concentration to the minimum concentration measured on the 3 m height masts because these were located too close to the plots to be considered as real background masts. This explains why strategies C3 and C4 lead to higher estimates compared to strategies C6 and C7, as the background may have been underestimated. We also find that all methods consistently infer a deposition flux to the blocks with no application, which is consistent with our knowledge of ammonia exchange between the atmosphere and the ground (Flechard et al., 2013). Indeed, the concentration in the atmosphere, which is enriched by the nearby sources is expected to be higher than near the ground, due to a low soil pH (6.1), a low nitrogen content in the soil surface (6-9.5 g N kg^{-1} DM), and a 20% humid soil surface, hence leading to a flux from the air to the ground.

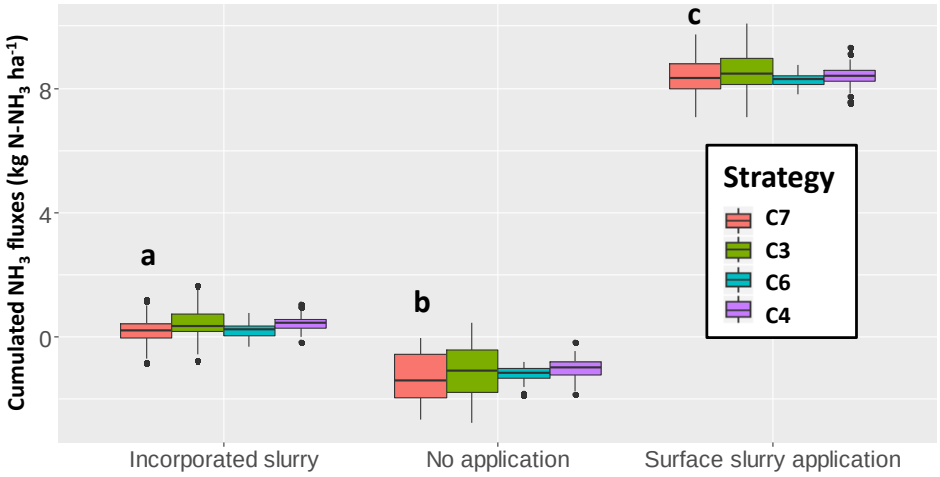

Figure 17. Same as Figure 16 but grouped by treatments and with additional strategies C4 and C6 which consider that replicates have the same surface flux. The variability in the boxplot aggregates the uncertainty on the inference method (the standard deviation on the flux estimate in the least-square model, which accounts for the variability in the replicated concentration measurements), and the variability between the repetitions in each treatment. Letters a, b and c shows significant differences between treatments for the C7 strategy, according to a Tukey test (95% family-wise confidence level).

From our theoretical study we know that strategy C7 should give a bias around **-16% ± ~7**%. Therefore, we could expect that the real flux is the one measured with C7 times 1.15 **(±** 0.08), hence **would be 10.9 ± 1.**3 kg N ha^{-1}. This corresponds to **28 ± 3**% of the N-NH_4 applied and **~9 ± 1**% of the total N applied. For the incorporated slurry, the **emissions are around 20 times smaller than the emissions from the surface applied slurry. Under these conditions, the bias on the emission would be around -20%, which means that the corrected emissions would range from 0.5% to 2.5% of the N-NH_4 applied** and **0.2 and 0.8% of the total N applied.** We should bear in mind that the theoretical **correction** is based on the median of the simulations done with the 2008 dataset in Grignon which had similar meteorological conditions to this trial. It would be much more relevant **though for** future developments to evaluate the bias based on the same method as developed here but **with** emissions and meteorological conditions **taken from** the real case.

3.6 Comparison with previous work

Several studies have reported methodologies for evaluating multiple sources using dispersion models. These were mostly based on backward Lagrangian modelling (Crenna et al., 2008; Flesch et al., 2009; Gao et al.,

2008). There were several inference methods reported: the methods based on the inversion of the dispersion matrix D_{ij} or singular value decomposition of least-square optimisation (Flesch et al., 2009), which optimise the conditioning of the dispersion matrix and one based on Bayesian inference (Yee and Flesch, 2010). Yee et al. (2010) showed that the Bayesian approach would avoid unrealistic source estimates which could appear when the matrix conditioning was poor. Unrealistic source estimates were for instance reported by Flesch et al. (2009), with negative emission sources.

In Ro et al. (2011), they evaluated the bLS technique to infer two controlled methane surface sources with laser measurements. They found 0.6 recovery ratios (ratio of inferred to known source) if the fields were not in the footprint of the sensor but with adapted filters, they found a high degree of recovery with of 1.1 ± 0.2 and 0.8 ± 0.1 for the two sources respectively. They found that in contradiction to Crenna et al. (2008) and Flesch et al. (2009), even with large conditioning numbers they had high recovery rates.

Misselbrook (2005) **compared different methodologies and showed that under high concentrations diffusion samplers may lead to overestimation of up to 70% of the concentration. They suggest potential issues related to the deformation of the Teflon membrane which would modify the distance between coated filters and the membrane itself that could cause sampler saturation. There is hence some concern on the quality of diffusion samplers to measure concentrations at heights close to large sources which would necessitate field validations.**

3.6.1 Sensor positioning and conditioning number

Crenna et al. (2008) have clearly shown that the optimal sensor positioning should be so that each sensor sees preferentially a single source, and reversely, each source should preferentially influence a single sensor. In this study the sources-sensors geometry was especially designed in a way that minimises the condition number *CN*, by placing the sensors in the middle of each plot. For the smallest source (x_{plot} = 25 m), the conditioning number ranged from 1.97 to 3.01 (median 2.42) for sensors located at 0.25 m, and increased to 2.6-6.9 (median 3.2) for sensors at 0.5 m, 4.7-150 (median 21) for sensors at 1.0 m, and 40-165000 (median 640) for sensors at 2 m. This shows that including at least one sensor per block at heights lower than the field width divided by 20 would ensure that the conditioning number remains lower than in most trials **reported by** Crenna et al. (2008).

By comparing different strategies we have found that the strategies using two sensors over each source systematically led to improved performances (C3 versus C1 and C6 versus C5, **Figure 14**). This is also in line with the results of Crenna et al. (2008), who showed that using more sensors separated spatially improves the performance of the inference method. Hence we can conclude that the inference method we used is based on a well-conditioned system which leads to robust results of the least-square optimisation. This is further illustrated by the real case example (**Figures 15-17**) which shows a good reproducibility between block repetitions. Indeed, good reproducibility between repetitions is a check for evaluating the quality of the inference method in real test cases. The use of Bayesian inference method would however also be valuable in the setup we propose here.

3.6.2 Effect of time integrating sensors on the source inference quality

The use of time averaging sensors for estimating ammonia sources was already reported by Sanz et al. (2010), Theobald et al. (2013), Carozzi et al. (2013a; 2013b), Ferrara et al. (2014) and Riddick et al. (2016a; 2014). All these studies have shown the feasibility of these measurements, however only a few of them allow estimating the

impact of averaging: Riddick et al. (2014) measured emissions from a bird colony in the Ascension Island with WindTrax using both several ALPHA samplers in a transect across the colony and a continuous analyser for ammonia (AiRRmonia, Mechatronics, NL) downwind. They also averaged the continuous sampler concentrations to evaluate the effect of averaging on the emissions estimates. They found as we do here that averaging over monthly periods would lead to systematic underestimations from -9% to -66%. They also found that estimations from **diffusive samplers** would lead to average underestimations of -12%. This is very close to what we find here for a single source over one week (**Figure 6**). In a similar comparison Riddick et al. (2016b) found that time-integration led to slight overestimations with integration approach, which is within the range of statistics of the bias we have found for the larger area sources (3rd quartile in **Figure 6**).

3.6.3 Dependency to meteorological conditions

We should bear in mind that the use of time averaging sensors in the inference method is also highly dependent on the surface layer turbulent structure as shown by **Figure 7**. We find, as expected, that stable conditions or low wind speed conditions are those that lead to the highest potential bias (as shown by the 3rd quartile under stable conditions in **Figure 7** bottom). This is a well-known limitation of inverse dispersion modelling which was reported by Flesch et al. (2009; 2004) and which suggested that inverse dispersion would be inaccurate for $u_* < 0.15$ m s^{-1} and $|z/L| < 1$. However, both our study and the studies of Riddick et al. (2014; 2016b) show that this is not as much of an issue for ammonia emissions. Indeed, this is due to the fact that ammonia emissions follow a daily cycle with low emissions at night and high emissions during the day. This is firstly because the ground surface compensation point concentration (C_{pground}) has an exponential dependency on surface temperature as assumed in **Eq. (10)** based on known thermodynamical equilibrium constants (Flechard et al., 2013). This is secondly due to the fact that ammonia emission is a diffusion-based process which is limited by the surface resistances, as modelled in **Eq. (9)**, which leads to small fluxes when $R_a(z_{ref})$ and R_{bNH_3} get large, which happens during low wind speeds (they are both roughly inversely proportional to wind speed) and stable conditions, which also happens at night (Flechard et al., 2013). In real situations, the combination of small turbulence and high surface concentration leads to a further decrease of the flux which is dependent on the difference between C_{pground} and the concentration in the atmosphere above (a feature which was not accounted for in this study as this would imply a higher degree of complexity in the modelling approach). This means that the results we found in this study would not apply for species having an emission pattern with a different temporal dynamics (either constant or anti-correlated with surface temperature or wind speed).

4. Conclusions

In this study we have demonstrated that it is possible to infer with reasonable biases ammonia emissions from multiple small fields located near each other using a combination of a dispersion model and a set of passive diffusion sensors which integrate over a few hours to weekly periods. We found that the Philip (1959) analytical model **in FIDES** gave similar concentrations as the backward Lagrangian Stochastic model WindTrax **at 2 m** above a small source, **under neutral and stable stratification** as long as the stability correction functions used in both models are similar **and the Schmidt number is identical (here set to 0.64). Under unstable conditions FIDES gave 20% smaller concentrations at 2 m compared to WindTrax.**

We demonstrated by theoretical considerations that passive sensors always lead to the underestimation of ammonia emissions for an isolated source because of the negative time correlation between the ammonia emissions and the transfer function. Using a yearly meteorological dataset typical of the oceanic climate of western Europe we found that the bias over weekly integration times is typically -8±6%, which is in line with previous reports. Larger biases are expected for meteorological conditions with stable conditions and low wind speeds as soon as the integration period is larger than 12 hours.

We showed that the quality of the inference method for multiple sources was dependent on the number of sensors considered above each plot. The most essential technique to minimise the bias of the method was to place a sensor in the middle of each source within the boundary layer. The quality of the sensor positioning was evaluated using "condition numbers" which ranged from 2 to 3 for a sensor placed at 25 cm above the ground to much higher values (40-1.6×10^5) for a sensor at 2 m height above 25 m width sources. **Although the lowest sensors have the best condition number, we would rather recommend using heights of 50 cm above the canopy in order to reduce uncertainty in positioning the sensors close to the ground as well as avoid non-diffusive transfer conditions. Similarly,** although the highest sensors had low condition numbers, they were shown to improve the robustness of the sources inference especially for evaluating the background concentrations. Using replicates of each treatment was found to be essential for evaluating the quality of the inference and derive robust statistical indicators for each treatment.

When considering a system, characteristic of agronomic trials, composed of a low and a high potential source and a reference with no nitrogen application, we found that the fractional bias remained smaller than around 25% for ratios between the largest to the smallest sources lower than factor 5 and increased as a power function of the ratio. Furthermore, the dynamics of the emissions were found not to strongly affect the fractional bias. As expected, we also found that the fractional bias decreased with increasing source dimensions, especially for the lowest source strength in a multiple source trial.

Finally, a test on a practical trial proved the applicability of the method in real situations with contrasted emissions. We indeed calculated ammonia emissions of around **27 ± 3**% of the total ammoniacal nitrogen applied for surface applied slurry while we found less than 1% emissions for the treatments with incorporated slurry.

This method could also be improved by incorporating knowledge of the surface source dynamics into the inference procedure. Further work is required however, for validating the method, for instance using prescribed emissions, and to evaluate it for growing crops using real measurements with diffusion samplers close to the ground.

Acknowledgements

This study was supported by EU FP7 NitroEurope-IP (grant number 017841) and ECLAIRE (grant number 282910), French national projects CASDAR VOLAT'NH3 (grant number 0933), ADEME EVAPRO (grant number 1560C0036), ADEME EVAMIN (grant number 1660C0012). The data sets used in this paper can be obtained from the authors upon request. The meteorological dataset used in this study are from the ICOS site FR-GRI which can be obtained from http://fluxnet.fluxdata.org/. We thank Erwan Personne for the use of the SurfAtm-NH_3 model**, and the technical team of the ARVALIS research station of "La Jaillière" for their involvement in the conduct of the "real test case" experiment.**

Supplementary material

See supplementary material manuscript

Model availability

The model is available as an R package upon request to the authors.

References

Carozzi, M., Ferrara, R.M., Rana, G. and Acutis, M., 2013a. Evaluation of mitigation strategies to reduce ammonia losses from slurry fertilisation on arable lands. Sci Total Environ, 449: 126-33.

Carozzi, M., Loubet, B., Acutis, M., Rana, G. and Ferrara, R.M., 2013b. Inverse dispersion modelling highlights the efficiency of slurry injection to reduce ammonia losses by agriculture in the Po Valley (Italy). Agric. For. Meteorol., 171: 306-318.

Choudhury, B.J. and Monteith, J.L., 1988. A four-layer model for the heat budget of homogeneous land surfaces. Q.J.R. Meteorol. Soc., 114: 373-398.

CITEPA, 2017. Inventaire des émissions de polluants atmosphériques en France métropolitaine, format CEE-NU. CITEPA 494 / Convention MATE 26 / 2001, Centre Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique.

Council, E., 1996. Directive 96/61/EC of 24th September 1996 concerning integrated pollution prevention and control. , European Council, Brussels, Belgium.

Council, E., 2016. Directive (EU) 2016/2284 of the European parliament and of the council of 14 December 2016 on the reduction of national emissions of certain atmospheric pollutants, amending Directive 2003/35/EC and repealing Directive 2001/81/EC, European Council, Brussels, Belgium.

Crenna, B.R., Flesch, T.K. and Wilson, J.D., 2008. Influence of source-sensor geometry on multi-source emission rate estimates. Atmos. Environ., 42(32): 7373-7383.

ECETOC, 1994. Ammonia emissions to air in Western Europe, European Centre for Ecotoxicology and Toxicology of Chemicals, Avenue E Van Nieuwenhuyse 4, Brussels.

EUROSTAT, 2012. Agroenvironmental indicator - ammonia emission, Eurostat, Luxemburg.

Faburé, J. et al., 2011. Synthèse bibliographique sur la contribution de l'agriculture à l'émission de particules vers l'atmosphère : identification de facteurs d'émission, ADEME / INRA.

Famulari, D. et al., 2010. Development of a low-cost system for measuring conditional time-averaged gradients of SO2 and NH3. Environ Monit Assess, 161(1-4): 11-27.

Ferrara, R.M. et al., 2016. Dynamics of ammonia volatilisation measured by eddy covariance during slurry spreading in north Italy. Agriculture, Ecosystems & Environment, 219: 1-13.

Ferrara, R.M. et al., 2014. Ammonia volatilisation following urea fertilisation in an sorghum crop in Italy irrigated. Agric. For. Meteorol., 195: 179-191.

Ferrara, R.M. et al., 2012. Eddy covariance measurement of ammonia fluxes: Comparison of high frequency correction methodologies. Agric. For. Meteorol., 158(0): 30-42.

Flechard, C.R. and Fowler, D., 1998. Atmospheric ammonia at a moorland site. II: Long-term surface-atmosphere micrometeorological flux measurements. Q.J.R. Meteorol. Soc., 124(547): 759-791.

Flechard, C.R. et al., 2013. Advances in understanding, models and parameterizations of biosphere-atmosphere ammonia exchange. Biogeosciences, 10(7): 5183-5225.

Flesch, T.K., Harper, L.A., Desjardins, R.L., Gao, Z.L. and Crenna, B.P., 2009. Multi-Source Emission Determination Using an Inverse-Dispersion Technique. Boundary-Layer Meteorology, 132(1): 11-30.

Flesch, T.K., Wilson, J.D., Harper, L.A., Crenna, B.P. and Sharpe, R.R., 2004. Deducing ground-to-air emissions from observed trace gas concentrations: A field trial. J. Appl. Meteorol., 43(3): 487-502.

Flesch, T.K., Wilson, J.D. and Yee, E., 1995. Backward-Time Lagrangian Stochastic Dispersion Models and Their Application to Estimate Gaseous Emissions. J. Appl. Meteorol., 34(6): 1320-1332.

Gao, Z.L., Desjardins, R.L., van Haarlem, R.P. and Flesch, T.K., 2008. Estimating Gas Emissions from Multiple Sources Using a Backward Lagrangian Stochastic Model. Journal of the Air & Waste Management Association, 58(11): 1415-1421.

Gao, Z.L., Mauder, M., Desjardins, R.L., Flesch, T.K. and van Haarlem, R.P., 2009. Assessment of the backward Lagrangian Stochastic dispersion technique for continuous measurements of CH4 emissions. Agric. For. Meteorol., 149(9): 1516-1523.

Gericke, D., Pacholski, A. and Kage, H., 2011. Measurement of ammonia emissions in multi-plot field experiments. Biosystems Engineering, 108(2): 164-173.
Häni, C., Sintermann, J., Kupper, T., Jocher, M. and Neftel, A., 2016. Ammonia emission after slurry application to grassland in Switzerland. Atmospheric Environment, 125(Part A): 92-99.
Holtslag, A.A.M. and Vanulden, A.P., 1983. A Simple Scheme for Daytime Estimates of the Surface Fluxes from Routine Weather Data. Journal of Climate and Applied Meteorology, 22(4): 517-529.
Huang, C.H., 1979. A theory of dispersion in turbulent shear flow. Atmos. Environ., 13: 453-463.
Kaimal, J.C. and Finnigan, J.J., 1994. Atmospheric Boundary Layer Flows, Their structure and measurement. Oxford University Press., New York, 289 pp.
Kormann, R. and Meixner, F.X., 2001. An analytical footprint model for non-neutral stratification. Boundary Layer Meteorol., 99(2): 207-224.
Laubach, J., Taghizadeh-Toosi, A., Sherlock, R.R. and Kelliher, F.M., 2012. Measuring and modelling ammonia emissions from a regular pattern of cattle urine patches. Agric. For. Meteorol., 156: 1-17.
Loubet, B. et al., 2012. Investigating the stomatal, cuticular and soil ammonia fluxes over a growing tritical crop under high acidic loads. Biogeosciences, 9(4): 1537-1552.
Loubet, B. et al., 2010. An inverse model to estimate ammonia emissions from fields. European Journal of Soil Science, 61(5): 793-805.
Loubet, B., Milford, C., Sutton, M.A. and Cellier, P., 2001. Investigation of the interaction between sources and sinks of atmospheric ammonia in an upland landscape using a simplified dispersion-exchange model. J. Geophys. Res.-Atmos., 106(D20): 24183-24195.
Lushi, E. and Stockie, J.M., 2010. An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources. Atmos. Environ., 44(8): 1097-1107.
McGinn, S.M. and Janzen, H.H., 1998. Ammonia sources in agriculture and their measurement. Canadian Journal of Soil Science, 78(1): 139-148.
Milford, C., Hargreaves, K.J., Sutton, M.A., Loubet, B. and Cellier, P., 2001. Fluxes of NH3 and CO2 over upland moorland in the vicinity of agricultural land. J. Geophys. Res.-Atmos., 106(D20): 24169-24181.
Milford, C. et al., 2009. Ammonia fluxes in relation to cutting and fertilization of an intensively managed grassland derived from an inter-comparison of gradient measurements. Biogeosciences, 6(5): 819-834.
Misselbrook, T.H., Nicholson, F.A., Chambers, B.J. and Johnson, R.A., 2005. Measuring ammonia emissions from land applied manure: an intercomparison of commonly used samplers and techniques. Environmental Pollution, 135(3): 389-397.
Moring, A. et al., 2016. A process-based model for ammonia emission from urine patches, GAG (Generation of Ammonia from Grazing): description and sensitivity analysis. Biogeosciences, 13(6): 1837-1861.
Mukherjee, S., McMillan, A.M.S., Sturman, A.P., Harvey, M.J. and Laubach, J., 2015. Footprint methods to separate N2O emission rates from adjacent paddock areas. International Journal of Biometeorology, 59(3): 325-338.
Nemitz, E. et al., 2009. Aerosol fluxes and particle growth above managed grassland. Biogeosciences, 6(8): 1627-1645.
Nemitz, E., Sutton, M.A., Schjoerring, J.K., Husted, S. and Wyers, G.P., 2000. Resistance modelling of ammonia exchange over oilseed rape. Agricultural and Forest Meteorology, 105(4): 405-425.
Pacholski, A. et al., 2006. Calibration of a simple method for determining ammonia volatilization in the field - comparative measurements in Henan Province, China. Nutrient Cycling in Agroecosystems, 74(3): 259-273.
Personne, E. et al., 2009. SURFATM-NH3: a model combining the surface energy balance and bi-directional exchanges of ammonia applied at the field scale. Biogeosciences, 6(8): 1371-1388.
Personne, E. et al., 2015. Investigating sources and sinks for ammonia exchanges between the atmosphere and a wheat canopy following slurry application with trailing hose. Agricultural and Forest Meteorology, 207: 11-23.
Philip, J.R., 1959. The Theory of Local Advection .1. J Meteorol, 16(5): 535-547.
Riddick, S. et al., 2016a. Estimate of changes in agricultural terrestrial nitrogen pathways and ammonia emissions from 1850 to present in the Community Earth System Model. Biogeosciences, 13(11): 3397-3426.
Riddick, S.N. et al., 2014. Measurement of ammonia emissions from tropical seabird colonies. Atmos. Environ., 89: 35-42.
Riddick, S.N. et al., 2016b. Measurement of ammonia emissions from temperate and sub-polar seabird colonies. Atmos. Environ., 134: 40-50.
Ro, K.S., Johnson, M.H., Hunt, P.G. and Flesch, T.K., 2011. Measuring Trace Gas Emission from Multi-Distributed Sources Using Vertical Radial Plume Mapping (VRPM) and Backward Lagrangian Stochastic (bLS) Techniques. Atmosphere, 2(3): 553-566.

Sanz, A., Misselbrook, T., Sanz, M.J. and Vallejo, A., 2010. Use of an inverse dispersion technique for estimating ammonia emission from surface-applied slurry. Atmos. Environ., 44(7): 999-1002.
Sintermann, J. et al., 2011a. Determination of field scale ammonia emissions for common slurry spreading practice with two independent methods. Atmos. Meas. Tech., 4(9): 1821-1840.
Sintermann, J. et al., 2012. Are ammonia emissions from field-applied slurry substantially over-estimated in European emission inventories? Biogeosciences, 9(5): 1611-1632.
Sintermann, J. et al., 2011b. Eddy covariance flux measurements of ammonia by high temperature chemical ionisation mass spectrometry. Atmos. Meas. Tech., 4(3): 599-616.
Sommer, S.G. et al., 2003. Processes controlling ammonia emission from livestock slurry in the field. European Journal of Agronomy, 19(4): 465-486.
Sommer, S.G., Jensen, L.S., Clausen, S.B. and SØGaard, H.T., 2006. Ammonia volatilization from surface-applied livestock slurry as affected by slurry composition and slurry infiltration depth. The Journal of Agricultural Science, 144(3): 229-235.
Sommer, S.G., McGinn, S.M. and Flesch, T.K., 2005. Simple use of the backwards Lagrangian stochastic dispersion technique for measuring ammonia emission from small field-plots. European Journal of Agronomy, 23(1): 1-7.
Spirig, C., Flechard, C.R., Ammann, C. and Neftel, A., 2010. The annual ammonia budget of fertilised cut grassland - Part 1: Micrometeorological flux measurements and emissions after slurry application. Biogeosciences, 7(2): 521-536.
Sun, K. et al., 2015. Open-path eddy covariance measurements of ammonia fluxes from a beef cattle feedlot. Agricultural and Forest Meteorology, 213: 193-202.
Sutton, M.A. et al., 2001. Comparison of low cost measurement techniques for long-term monitoring of atmospheric ammonia. J Environ Monit, 3(5): 446-53.
Sutton, M.A. et al., 2009. Dynamics of ammonia exchange with cut grassland: synthesis of results and conclusions of the GRAMINAE Integrated Experiment. Biogeosciences, 6(12): 2907-2934.
Sutton, M.A. et al., 2011. Too much of a good thing. Nature, 472(7342): 159-161.
Sutton, M.A. et al., 2013. Towards a climate-dependent paradigm of ammonia emission and deposition. Philos Trans R Soc Lond B Biol Sci, 368(1621): 20130166.
Tang, Y.S., Cape, J.N. and Sutton, M.A., 2001. Development and types of passive samplers for monitoring atmospheric NO2 and NH3 concentrations. TheScientificWorldJournal, 1: 513-29.
Tang, Y.S. et al., 2009. European scale application of atmospheric reactive nitrogen measurements in a low-cost approach to infer dry deposition fluxes. Agriculture Ecosystems & Environment, 133(3-4): 183-195.
Theobald, M.R., Crittenden, P.D., Tang, Y.S. and Sutton, M.A., 2013. The application of inverse-dispersion and gradient methods to estimate ammonia emissions from a penguin colony. Atmos. Environ., 81: 320-329.
Thomson, L.C. et al., 2007. An improved algorithm for locating a gas source using inverse methods. Atmos. Environ., 41(6): 1128-1134.
UNECE (Editor), 2012. 1999 Protocol to Abate Acidification, Eutrophication and Ground-level Ozone to the Convention on Long-range Transboundary Air Pollution, as amended on 4 May 2012 (http://www.unece.org/env/lrtap/multi_h1.html). UNECE, Brussels, Belgium.
Van der Hoven, I., 1957. Power Spectrum of Horizontal Wind Speed in the Frequency Range from 0.0007 to 900 Cycles Per Hour. Journal of Meteorology, 14(2): 160-164.
Vandré, R. and Kaupenjohann, M., 1998. In Situ Measurement of Ammonia Emissions from Organic Fertilizers in Plot Experiments. Soil Science Society of America Journal, 62(2): 467-473.
Wang, W., Liu, W.Q., Zhang, T.S. and Ren, M.Y., 2013. Evaluation of backward Lagrangian stochastic (bLS) model to estimate gas emissions from complex sources based on numerical simulations. Atmos. Environ., 67: 211-218.
Whitehead, J.D. et al., 2008. Evaluation of laser absorption spectroscopic techniques for eddy covariance flux measurements of ammonia. Environ Sci Technol, 42(6): 2041-6.
Wilson, J.D., 2015. Computing the Flux Footprint. Boundary Layer Meteorol., 156(1): 1-14.
Wilson, J.D. and Shum, W.K.N., 1992. A Reexamination of the Integrated Horizontal Flux Method for Estimating Volatilization from Circular Plots. Agric. For. Meteorol., 57(4): 281-295.
Wilson, J.D., Thurtell, G.W., Kidd, G.E. and Beauchamp, E.G., 1982. Estimation of the rate of gaseous mass transfer from a surface source plot to the atmosphere. Atmospheric Environment (1967), 16(8): 1861-1867.
Yee, E., 2008. Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference. Boundary Layer Meteorol., 127(3): 359-394.
Yee, E. and Flesch, T.K., 2010. Inference of emission rates from multiple sources using Bayesian probability theory. J Environ Monit, 12(3): 622-34.

Supplementary material: Evaluation of a new inference method for estimating ammonia volatilisation from multiple agronomic plots

Benjamin Loubet[1,*], Marco Carozzi[1,#], Polina Voylokov[1], Jean-Pierre Cohan[2], Robert Trochard[2], Sophie Génermont[1]

1 INRA, UMR ECOSYS, INRA, AgroParisTech, Université Paris-Saclay, 78850, Thiverval-Grignon, France
2 ARVALIS-Institut du Végétal, **Station expérimentale de La Jaillière, La Chapelle Saint Sauveur, 44370 Loireauxence**, France
now at: Agroscope Research Station, Climate and **Agriculture**, Zurich, Switzerland

* Corresponding author: Benjamin.Loubet@inra.fr

S1. Analogy between dispersion equation and flux-resistance approaches

It is interesting to note that **Eq. (1)** is essentially similar to resistance analogy approaches, where the flux F is evaluated as a concentration difference divided by a transfer resistance between two heights z_1 and z_2, $F = -(C(z_2) - C(z_1))/R(z_1, z_2)$. Indeed, assuming, as is done in the resistance analogy that the source is infinitely expanded in x, then computing **Eq. (1)** for heights z_1 and z_2 and recombining leads simply to $R(z_1, z_2) = D(z_1) - D(z_2)$. Hence the transfer function D is equivalent to a transfer resistance. In particular, for infinitely expanded sources, the resistance between two heights equals the difference between the transfer function between these two heights and the ground.

S2 Condition number to identify suitable source-receptor geometry

A major issue when trying to infer sources from **atmospheric** concentrations is the fact that under some circumstances, the problem is ill-conditioned, which means that a small change in the concentration or the transfer matrix D_{ij} will induce large changes on the sources strength estimates. A measure of the conditioning of the problem is therefore an important indicator for determining whether the source-receptor geometry can lead to realistic solutions. The condition number is a measure of ill-conditioning and is defined as (Crenna et al., 2008):

Code de champ modifié

$$CN = \|D_{ij}\| \times \|D_{ij}^{-1}\| \quad \text{(S1)}$$

Where $\|.\|$ denotes a norm of **the** matrix, one definition of which being the maximum of the sum of the rows. The higher *CN*, the larger the uncertainty on the solution of **Eqns. (3)** and **(6)** (Flesch et al., 2009). To evaluate the conditioning state of each set-up, we considered the simplified case where the background concentration is zero and the number of receptors equals the number of sources. In such a case, the matrix D_{ij} is squared and D_{ij}^{-1} is defined.

Considering the single source case eases the understanding of the **meaning of** condition number. Indeed, in that case $D_{ij} = D_j$ is a vector and *CN* is simply: $\max(D(x_i))/\min(D(x_i))$**.** In physical terms, this means that if some concentration samplers are well exposed to the source and others are not, *CN* is large. In such a case, **Eq.**

(4) shows that the a small error in $\overline{C(x_i)} - \overline{C_{bgd}}$ will lead to a large error in $\bar{S}$. Therefore we see here that using several concentration samplers may lead to increasing the error on $\bar{S}$ if their locations are not chosen with care. This was also showed by Crenna et al. (2008) and Flesch et al. (2009), who showed that the condition number *CN* should be minimised in order to keep this error minimal; in this regards, Gao et al. (2008) suggest that *CN* should be smaller than 10. In practice, minimising *CN* would mean minimising the range of $\overline{D(x_i)}$. **We can interpret this in terms of footprint:** the source area should represent a reasonable footprint fraction of each concentration sensor. This holds for multiple sources also: in that case each source should represent a large fraction of each sensor footprint placed above it. The setup we propose in this study is, by construction, minimising *CN* as the sensors are placed in the middle of each plot, provided they are placed low enough to catch a significant part of the field footprint. If the plots are in a non-squared configuration, the *CN* is simply calculated as in **Eq. S1**, where the second term in the right hand **side** is the pseudo inverse of the matrix D_{ij}. **In practice,** the calculation of *CN* was performed **using** the kappa function in R (version 3.2.3).

S3. Details of the FIDES model based on a solution of Philip (1959) of the advection diffusion equation

In the FIDES model, the transfer function $D(x_i, S_j, t)$ was estimated by first translating and rotating the x-y plan to locate the source S_j at the centre coordinates (0,0) and set the wind direction *WD* to 0 (align the *x*-axis with the wind vector. This was done by setting the following coordinate transformation $X_{ij} = \left(x_i - x_{s_j}\right)\sin(WD) - \left(y_i - y_{s_j}\right)\cos(WD)$, and $Y_{ij} = \left(x_i - x_{s_j}\right)\cos(WD) - \left(y_i - y_{s_j}\right)\sin(WD)$. Moreover, all heights are considered as heights above displacement height *d* ($Z = z - d$). In such conditions, the Philip (1959) solution reads:

$$U(Z_i) = aZ_i^{\,p} \tag{S2}$$

$$K_z(Z_i) = bZ_i^{\,n} \tag{S3}$$

$$D(x_i, S_j, t) = \frac{1}{\sigma_y(X_{ij})\sqrt{2\pi}} \exp\left(-\frac{(Y_{ij})^2}{2\sigma_y^2}\right) \times \frac{(Z_i Z_s)^{(1-n)/2}}{b\alpha X_{ij}} \times \exp\left(-\frac{a(Z_i^{\,\alpha} + Z_s^{\,\alpha})}{b\alpha^2 X_{ij}}\right) \times I_{-\upsilon}\left(\frac{2a(Z_i Z_s)^{\alpha/2}}{b\alpha^2 X_{ij}}\right) \tag{S4}$$

$$\sigma_y = \frac{1}{\sqrt{2}} C_y X_{ij}^{\,\frac{2-m}{2}}$$

where $\alpha = 2 + p + n$, $\nu = (1 - n) / \alpha$, and $I_{-\nu}$ is the modified Bessel function of the first kind of order -ν, and Cy and *m* were taken from Sutton (1932). The values of *a*, *b*, *p* and *n* were inferred by linear regression between ln(*U*), ln(K_z) and ln(*Z*), over the height range $2 \times z_0$ to 20 m, using *U*(z) and K_z(z) estimated from the Monin-Obukhov similarity theory as $K_z(Z) = ku_*Z[Sc\emptyset_H(Z/L)]^{-1}$. Here $\emptyset_H(Z/L)$ is the universal stability correction function as in Kaimal and Finnigan (1994), which is $\emptyset_H(Z/L) = (1 + 5.2\,Z/L)$ for $Z/L \geq 0$ and $\emptyset_H(Z/L) = (1 - 16\,Z/L)^{0.5}$ for $Z/L \leq 0$. Following Loubet et al. (2001), to ensure **Eq. (S4)** exists, the source height is taken as $Z_s = 1.01\ z_0$. FIDES is essentially the same model as the one reported by Kormann and Meixner (2001). The only difference resides in the way *a*, *b*, *p* and *n* are determined: in Kormann and Meixner (2001) these constants are determined by equating *U* and K_z from Monin-Obukhov similarity theory to **Eqns. (S2 and S3)** at the reference height (*H*), while in FIDES a range of heights ($2 \times z_0$ to 20 m) is used to compute these values.

However, Wilson **(2015)** shows that under neutral stratification, any choice of $H / z_0 \gg 10$ should return an adequate concentration profile near the surface at fetches $1 \ll x / z_0 \ll 10^5$, hence FIDES and Korman and Meixner models can be considered equivalent in the range of dimensions considered in this study.

S4. Insuring coherency between WindTrax and Philip (1959) models (tuning FIDES with WindTrax)

S4.1. Insuring comparable Schmidt numbers

The WindTrax software combines the backward Lagrangian stochastic (bLS) dispersion model described by (Flesch et al., 2004) with an interface where sources and sensors can be mapped. The transfer function $D(x_i, S_j, t)$ is calculated by releasing N trajectories upwind from each sensor location x_i for each time step and recording the vertical velocity (w_0) of those that intersect the ground (N_{source}, or "touchdowns"). The transfer function is computed as:

$$D(x_i, S_j, t) = \frac{1}{N} \sum_{N_{source}} \left| \frac{2}{w_0} \right| \quad \text{(S5)}$$

In practice N = 50000 trajectories were used to compute D_{ij}. In WindTrax the Schmidt number (Sc, see 2.2) tends to 0.64 in the neutral limit as discussed by Wilson (2015).

S4.2. Insuring comparable Schmidt numbers

Most bLS models, and especially WindTrax assume $Sc = 0.64$, while models based on the eddy diffusion analogy, and hence FIDES and the Korman and Meixner model, lead to a Sc which was calculated in Carozzi et al. (2013) to be:

$$Sc = \frac{u_*^2}{abp} Z^{1-p-n} \quad \text{(S6)}$$

Hence constitutively, the **Philip** (1959) model does not lead to a constant Schmidt number in the surface layer, unless $1 - p - n \sim 0$, which was found to be the case under near neutral conditions (Carozzi et al., 2013). Note that the Korman and Meixner approach lead to $Sc = 1$ at the reference height in all conditions by construction. Furthermore, the stability correction functions are different in the Philip (1959) model and in WindTrax. Hence in order to compare the two approaches, the vertical diffusivity $K_z(Z)$ in FIDES was set as to reproduce the far field diffusivity of **Flesch et al.** (1995)**.** Indeed, in bLS, the far-field diffusivity is $K_z = \sigma_w T_L$, where σ_w is the standard deviation of the vertical component of the air velocity, and T_L is the Lagrangian time scale. Replacing by their expression as in Flesch et al. (1995), leads to the following far-field diffusivity:

$$K_z(Z) = 0.5\sqrt{1.7} u_* Z / \left(1 + 5\frac{Z}{L}\right) \quad \text{for } L > 0 \quad \text{(S7)}$$

$$K_z(Z) = 0.5\sqrt{2.2} u_* Z \times \left(1 - 6\frac{Z}{L}\right)^{0.25} \left(1 - 3.3\frac{Z}{L}\right)^{\left(\frac{0.67}{2}\right)} \quad \text{for } L \leq 0 \quad \text{(S8)}$$

It is noticeable that in **Eqns. (S5 and S6)** there is a step change between stable and unstable conditions. Indeed, when $L \to +\infty$ $K_z(Z) \to ku_*Z \times 0.63^{-1}$, while when $L \to -\infty$ $K_z(Z) \to ku_*Z \times 0.55^{-1}$. This means that in WindTrax, the *Sc* number is set to 0.63 under stable conditions and 0.55 under unstable conditions and that in near-neutral conditions *Sc* steps from 0.63 to 0.55 when passing from $L > 0$ to $L \leq 0$.

In FIDES, to ensure compatibility **with Flesch et al. (1995),** *Sc* was set to 0.64 and parameters b / *Sc* and n where adjusted so that $K_z(Z)$ in **Eq. (S3)** fits that in **Eqns. (S7and S8)** over a logarithmically spaced vector of 30 heights from z_0×1.01 to 2 m. **We should stress here that the expressions from Flesch et al. (1995) are slightly different from those in Windtrax (Flesch et al., 2004):**

$$K_z(Z) = 0.5 \times 1.25 \times u_* Z / \left(1 + 5\frac{Z}{L}\right) \quad \textbf{for L > 0} \quad \textbf{(S9)}$$

$$K_z(Z) = 0.5 \times 1.25 \times u_* Z \times \left(1 - 6\frac{Z}{L}\right)^{0.25} \left(1 - 3\frac{Z}{L}\right)^{1/3} \quad \textbf{for L} \leq \textbf{0} \quad \textbf{(S10)}$$

From the Eqns. (S9 and S10) is noticeable that under near neutral situations ($L \to +\infty$ or $L \to -\infty$), the diffusivity $K_z(Z)$ is converges to $ku_*Z \times 0.64^{-1}$ (where k = 0.41 is the von Karman's constant and 0.64 represents the Schmidt number) and is continuous for all L. **Figure S1 based on Eqns. (S9 and S10)** shows that our approach insures a coherency between the diffusivity of the bLS and Philip approach but small differences remain which are height dependent. We should also notice that lateral dispersion was treated separately in the two models, which will also lead to differences in the modelled concentration, especially for larger fields.

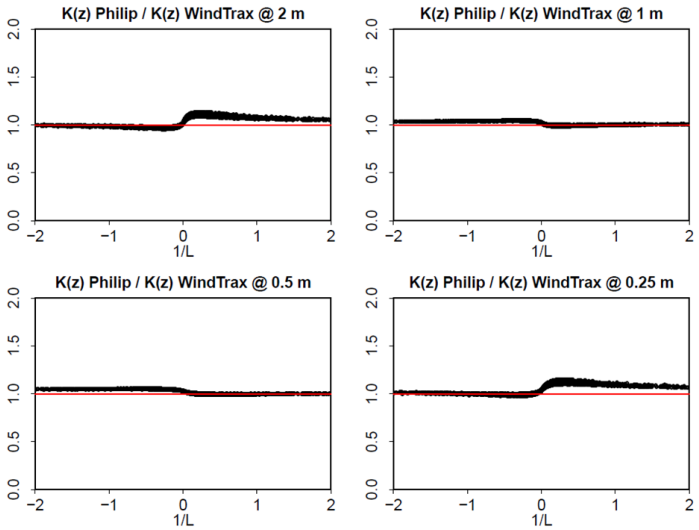

Figure S1. Ratio of the "tuned" FIDES ("Philip") to WindTrax vertical diffusivity for scalars ($K_z(z)$) as a function of the inverse of Obukhov length (*1/L*) at 0.25, 0.5, 1 and 2 m heights. The tuned diffusivity corresponds to Eqns. (S9 and S10).

S4.3. Comparison of FIDES and WindTrax models for predicting concentrations above a single source

A first step in the study was to compare the two dispersion models. **Figure S2** shows that the "tuned" FIDES model leads to the same concentration pattern as WindTrax **at 0.5 m above the source,** although systematically underestimating the maximum concentration under unstable conditions. **This behaviour is clearly visible in Figure S3 where the concentrations modelled with the "tuned" FIDES at 2 m above the surface (right graph) are underrated by about 15% for unstable conditions compared to WindTrax, while matching for stable and neutral conditions. We** further see that the concentration modelled with the original FIDES (Philip, 1959) are similar at 25 cm above the surface (left graphs) but differ substantially at 2 m above the surface (right graphs). This is expected as the longer the travel distance, the larger the expected difference in dilution if the two models' diffusivity differ. In the original FIDES, the diffusivity is lower than in WindTrax by a factor of roughly **one and half** (Sc^{Philip} = 1 and Sc^{WT} = 0.**64**). In a first order approach (over an infinitely homogeneous source), the concentration difference between z_0 and 2 m would be proportional to the aerodynamic resistance (itself proportional to the inverse of the vertical diffusivity) times the height above ground (see e.g. Flechard et al., 2013), which explains the differences observed in **Figure S3**.

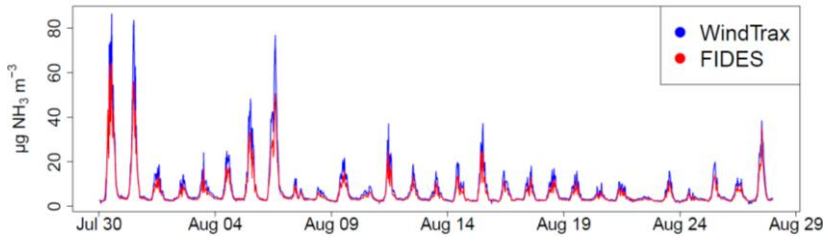

Figure S2. Example concentration modelled above a single ammonia source using two dispersion models WindTrax and FIDES with K_z as in Philip (1959), at 0.5 m above a simulated squared ammonia source of 25 by 25 m in the FR-Gri ICOS site during August 2008.

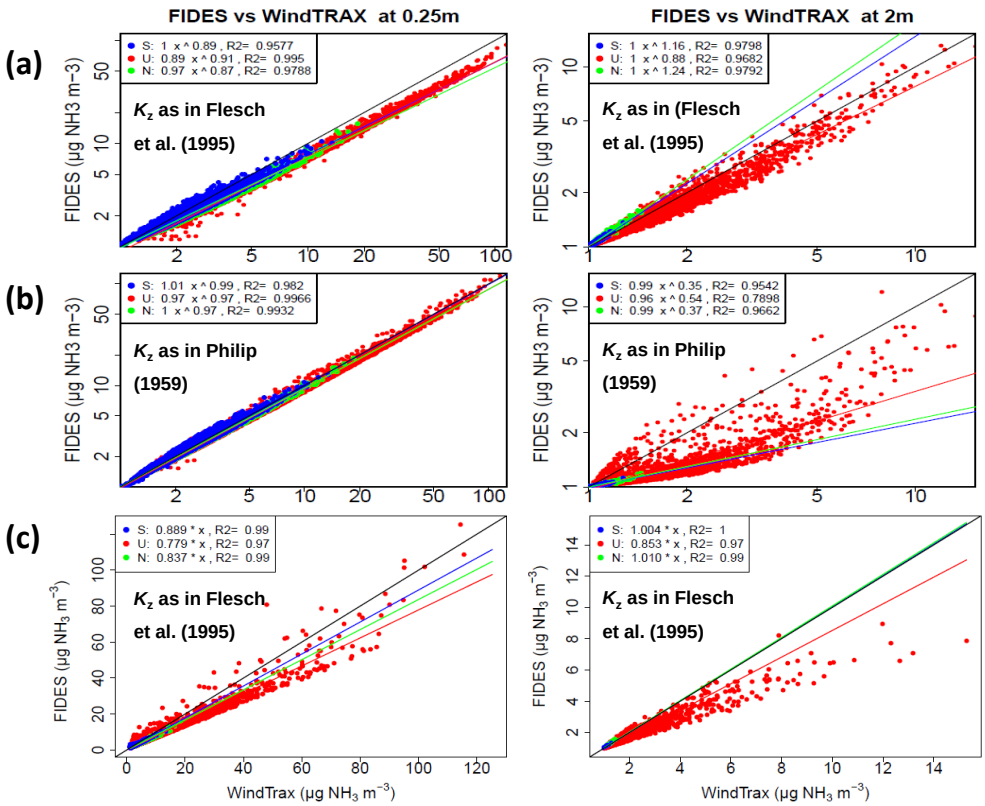

Figure S3. FIDES versus WindTrax concentration modelled above an ammonia source of 25 × 25 m at 0.25 and 2 m heights. In these graphs the FIDES vertical diffusivity K_z is either as in Philip (1959) (b) or fitted to Flesch et al. (1995) (a) and (c) as explained in S4.2. The comparison is made over the entire year of 2008 in the FR-Gri ICOS site. S, U and N stand for stable, unstable and neutral atmospheric conditions. The linear regression equation is given for each condition together with the R^2 of that regression. The black line is the 1:1 line. (a) and (b) show log-log axes an power law fits while (c) shows linear axes and linear fits.

Figure S3 also shows that the "tuned" FIDES modelled concentrations (top graphs) do not perfectly fit to the **WindTrax** ones (top graphs in **Figure S3**). At height of 25 cm, the "tuned" FIDES concentration does lead to a worse regression score than the original FIDES. Although **Figure S3** is focussing on a 25 m × 25 m field, the results are similar for larger fields (data not shown). This is explained by the difference in Z-dependency of K_z in the WindTrax and FIDES model, which is highlighted in **Figure S1**: under stable conditions ($1/L > 0$), "tuned" FIDES K_z**(Z)** is larger than WindTrax at 0.25 and 2 m, but smaller at 0.5 and 1 m, and the opposite under unstable conditions ($1/L < 0$). This means that constitutively the two models may never fit perfectly, showing a bias that will depend on height. Nevertheless, the correlation between the two models is very high as shown by large $R^2 \geq \sim 0.96$.

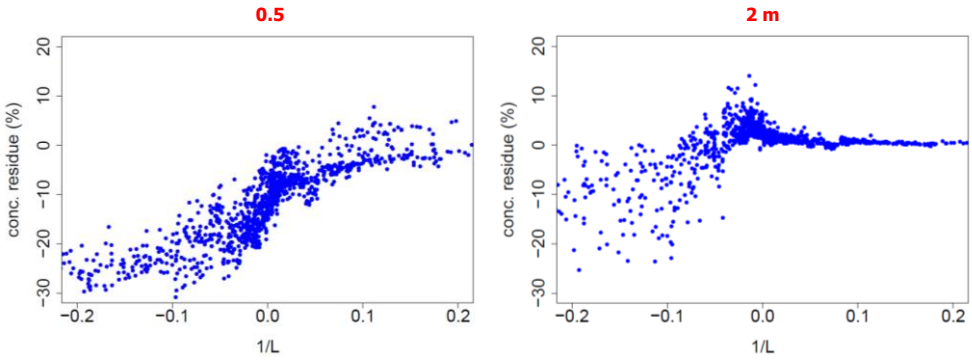

Figure S4. Relative difference between FIDES and WindTrax concentrations as a function of the stability parameter (1/L). Data refer to the same conditions reported in Figure S3.

Figure S4 reports the relative difference between the NH_3 concentration calculated by the two models at 0.5 and 2 m height, as a function of the stability parameter (*1/L*). As previously stated, under unstable conditions FIDES clearly underestimates the concentrations up to 30% at 0.5 m and lower heights, while this gap is reduced and more scattered at 2 m height. Moving towards neutral conditions the two models tends to agree notwithstanding an overestimation to 10% by WindTrax at 2 m height concurrently with an underestimation of the same magnitude by FIDES at 0.5 m. Under stable conditions there is a clear agreement at 2 m height, while this correspondence remains unbalanced to lower heights.

Supplementary figures

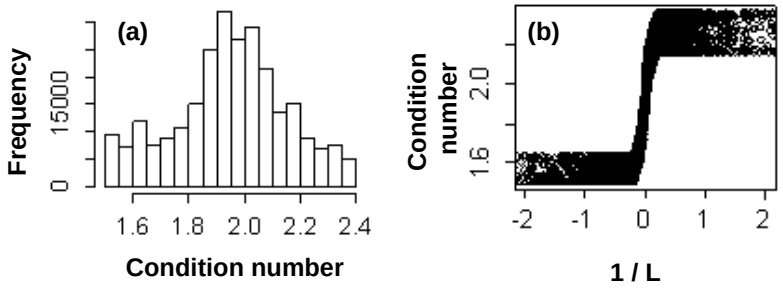

Figure S5. (a) Distribution of condition numbers for the 0.25 m height sensor and the 25 m width plots, for integration periods of 6h and 24h, and (b) condition number as a function of 1/*L*, where *L* is the Obukhov length.

References quoted in the supplementary material

Carozzi, M., Loubet, B., Acutis, M., Rana, G. and Ferrara, R.M., 2013. Inverse dispersion modelling highlights the efficiency of slurry injection to reduce ammonia losses by agriculture in the Po Valley (Italy). Agric. For. Meteorol., 171: 306-318.

Crenna, B.R., Flesch, T.K. and Wilson, J.D., 2008. Influence of source-sensor geometry on multi-source emission rate estimates. Atmos. Environ., 42(32): 7373-7383.

Flechard, C.R. et al., 2013. Advances in understanding, models and parameterizations of biosphere-atmosphere ammonia exchange. Biogeosciences, 10(7): 5183-5225.

Flesch, T.K., Harper, L.A., Desjardins, R.L., Gao, Z.L. and Crenna, B.P., 2009. Multi-Source Emission Determination Using an Inverse-Dispersion Technique. Boundary-Layer Meteorology, 132(1): 11-30.

Flesch, T.K., Wilson, J.D., Harper, L.A., Crenna, B.P. and Sharpe, R.R., 2004. Deducing ground-to-air emissions from observed trace gas concentrations: A field trial. J. Appl. Meteorol., 43(3): 487-502.

Flesch, T.K., Wilson, J.D. and Yee, E., 1995. Backward-Time Lagrangian Stochastic Dispersion Models and Their Application to Estimate Gaseous Emissions. J. Appl. Meteorol., 34(6): 1320-1332.

Gao, Z.L., Desjardins, R.L., van Haarlem, R.P. and Flesch, T.K., 2008. Estimating Gas Emissions from Multiple Sources Using a Backward Lagrangian Stochastic Model. Journal of the Air & Waste Management Association, 58(11): 1415-1421.

Kaimal, J.C. and Finnigan, J.J., 1994. Atmospheric Boundary Layer Flows, Their structure and measurement. Oxford University Press., New York, 289 pp.

Kormann, R. and Meixner, F.X., 2001. An analytical footprint model for non-neutral stratification. Boundary Layer Meteorol., 99(2): 207-224.

Loubet, B., Milford, C., Sutton, M.A. and Cellier, P., 2001. Investigation of the interaction between sources and sinks of atmospheric ammonia in an upland landscape using a simplified dispersion-exchange model. J. Geophys. Res.-Atmos., 106(D20): 24183-24195.

Philip, J.R., 1959. The Theory of Local Advection .1. J Meteorol, 16(5): 535-547.

Sutton, O.G., 1932. A Theory of Eddy Diffusion in the Atmosphere. Proceedings of the Royal Society of London. Series A, 135(826): 143-165.

Wilson, J.D., 2015. Computing the Flux Footprint. Boundary Layer Meteorol., 156(1): 1-14.

bg-2017-424-RC1: answer to referee 1 comments.

We would like to thank referee 1 for his helpful comments which we have answered below.

General Comments

This study carried out are providing answers to the much discussed question about the effect of having many plots in the field on measured ammonia emission from manure applied on the plots. Exploring the effect of measuring average ammonia concentration for increasing time intervals, the numbers of measuring heights and the best heights for measuring the emission. The answers to these questions are most important and the issue is discussed by scientist in Europe especially after the publications of Sinterman et al. questioned the existing design of measuring ammonia emission

The authors have developed a model for calculating emission of ammonia from as it varies over the day and year as affected by surface soil temperature, wind and atmospheric stability. Then, as I understand the paper, they calculate how much the emitted ammonia will contribute to atmospheric ammonia concentration at different heights above the soil surface as it is affected by climate and plot size i.e. the loss pattern over time after volatilization start is assessed using decay curves of source strength.

The atmospheric NH3 concentration data, climate data are then used as input to model calculation of the emission from a plot and plots in a field as affected a range of different management of measuring ammonia concentration, height of the ammonia concentration measurements, number of plots affecting ammonia concentration in plots downwind a plot, plot size, etc.

This reviewer is not a specialist in micrometeorology so I cannot evaluate the quality of the model calculations. In the following is my impression of the presentation and interpretation of the data.

Abstract

Line 9 NH3 is presented but later the authors write ammonia – should be NH3

This is a sound remark; we agree that we can use NH_3 throughout the manuscript once it has been defined as “tropospheric ammonia” except when it is the first word of a sentence.

Line 10: the abbreviation N for nitrogen should be given and N used in the text.

We thank the referee for the suggestion; we however think we could stick with “nitrogen” to avoid too many abbreviations in the abstract.

I am not familiar with the term inference method, the term inferring, inferred in this context? May be because my native language is not English.

We agree with the referee that “to infer” may not be a very commonly used term. It is a synonym of “to deduce”. We hence speak about a “source inference method” in the sense that the method is used to “deduce” the ammonia source.

L68: What is an intensive source?

We thank the reviewer for spotting this term which we might have mis-used. We rather wanted to mean an intense or strong source. We propose to change term the “intensive” to “strong” in the manuscript.

L69-70: require hourly concentrations of what????

Of NH_3. We propose to add this precision in the text.

L87 Multiple-source inversion problem?

We thank the reviewer for spotting this incoherency. Actually, we defined what we meant in lines 76-77, as “the multiple source problem, which consists of inferring multiple sources based on measured concentrations at multiple points in space and time,…”. We hence propose to use the term “multiple-source inference problem” throughout the manuscript to keep it coherent.

L121-124: Units are missing

Thanks for spotting that the units for concentration and emissions were missing. The concentration and source are in µg N-NH_3 m^{-3} and µg N-NH_3 m^{-2} s^{-1}, respectively. We propose to add these precision in the manuscript.

L324-327: Has the data from this experiment been used in previous articles, reports, proceedings?

Results from this experiment have been used jointly with other experimental trials in a poster proceeding, but not for testing the multiple-source inference methodology presented here. The poster was presented at a French meeting on fertilisation. The objective was to compare the emissions potential from several treatments based on the use of a “gradient” method applied to badges. The objective of the poster presentation was to show the potential of using alpha badges to differentiate nitrogen application methods in terms of potential ammonia losses. The link to the poster abstract in French is here:

http://www.comifer.asso.fr/images/pdf/11emes_rencontres/Interventions/Session%201/5%20-%20Jean-Pierre%20COHAN/Article%20Jean-Pierre%20COHAN.pdf.

L355-356: Rewrite

We propose to change to the following simplified sentence: “The friction velocity u* varied between 0.024 and 1.181 m s-1, and the stability parameter z/L varied between -49 and 21 m-1 (Figure 3).”

383: Condition number – what is this – referring to an equation S1 in annex, it is a number often used so a presentation of how it is calculated should be given in the article

The condition number is indeed an important indicator of the geometry of the multiple sources inference problem. We feel that it was well defined and discussed in the supplementary material section S2. The way it is calculated in practice is mentioned at line 46 of the supplementary material. We propose to slightly modify this sentence to make it even clearer: “In practice, the calculation of CN was performed using the kappa function in R (version 3.2.3).” We also propose to add the phrase “(see supplementary material section S2)” to lines 384 and 684 of the manuscript where this indicator is mentioned.

P519: When discussing the effect of height for measuring the horizontal then the authors should relate the outcome of their study to that of Wilson et al. who showed on basis of micro-met. Calculations that there is a best height for measuring the horizontal flux at one height (This Zinst height is higher that the height recommended here)

Although the ZINST method has no link with our approach, it is an interesting remark that brought back to our attention that the heights at which the alpha badges should be placed would depend on roughness length and displacement height. Indeed, the ZINST method is a method that uses the finding of Wilson et al. (1982) that the ratio of the source strength to the horizontal flux at height ZINST is somewhat constant whatever the stability conditions. Interestingly, Wilson et al. showed that ZINST was an exponential function of z_0 for a given source diameter. We hence propose to add the following text after Line 519: “It is interesting to note that the height which was found to provide an optimal inference of NH_3 sources (below 0.5 m) is smaller than the ZINST reported by Wilson et al. (1982) (which was 0.9 m for 40 m diameter circular sources, and which we estimate as 0.65 m based on a power law extrapolation as in Laubach et al. (2012)). It is also important to note that this height should vary with both the roughness length z0 and displacement height as was showed by Wilson et al. (1982) for ZINST.”

L555: What is the highest source?

We actually meant the largest source. We propose to change the text in L555 and L540 and also Figure 12 accordingly.

Figures:

The font size of the Y and X axis and some of the legends are too small on most figures. On some figures there are too many lines (7 lines on fig 4) making it very difficult to see the individual lines.

We thank the reviewer for his suggestion. We have looked at the figures thoroughly again and we agree indeed that some figures may be difficult to read, but most figures look good to us. We propose to improve some figures as explained below but we would like to have the editor's point of view for the other figures.

- Figure 4: we propose to reduce the number of integration periods and to keep only 0.5h, 24h and 168h which are sufficient to show the variability that is lost by integrating concentration measurements.

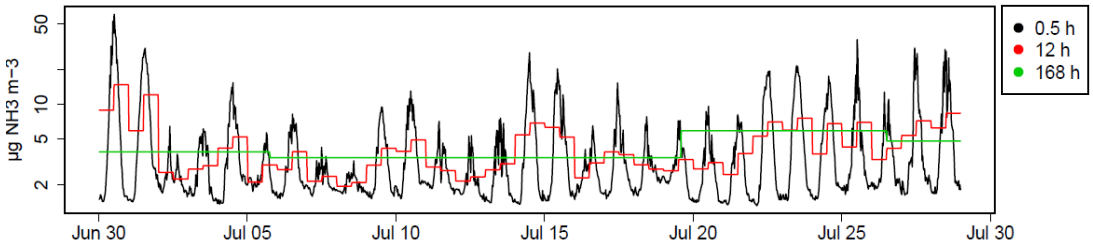

Figure 4. Example modelled concentration pattern at 1 m above a single 50 m width source for several averaging periods (0.5h, 12h and 168h) for the month of July 2008. The source Γ was set to 10^5. The y-axis is log scaled.

- Figures 5: there were some legends that we left in the right corner but these are not useful as the main legend was written at the top of the graph. We propose to erase these legends:

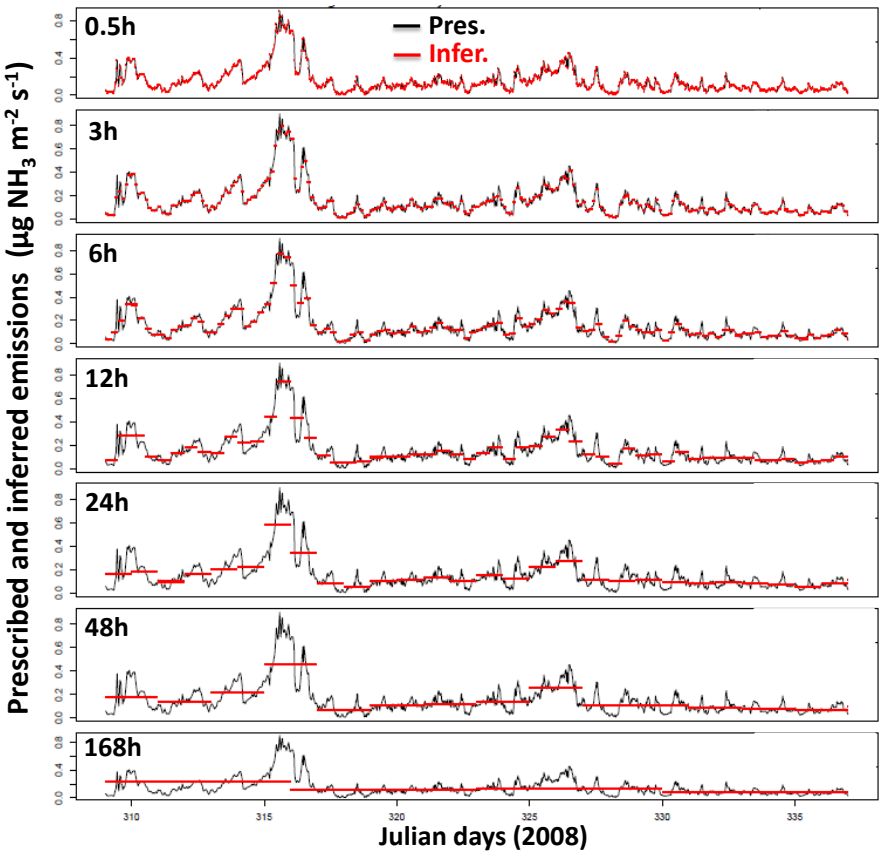

Figure 5. Example source inference for a 25 m width square field and a concentration sensor placed at 0.5 m above ground. Here Γ = 10000 and is set to constant (pattern 1). The 7 integration periods are shown: 0.5h to 168h. The x-axis shows the day of year and corresponds to a span over November. The prescribed source is in black (Obs.) and the inferred one in red (Pred.)

- Figure 7: we propose to change the font size and the Y axis label. We will also change the caption for u_* and $1/L$ classes:

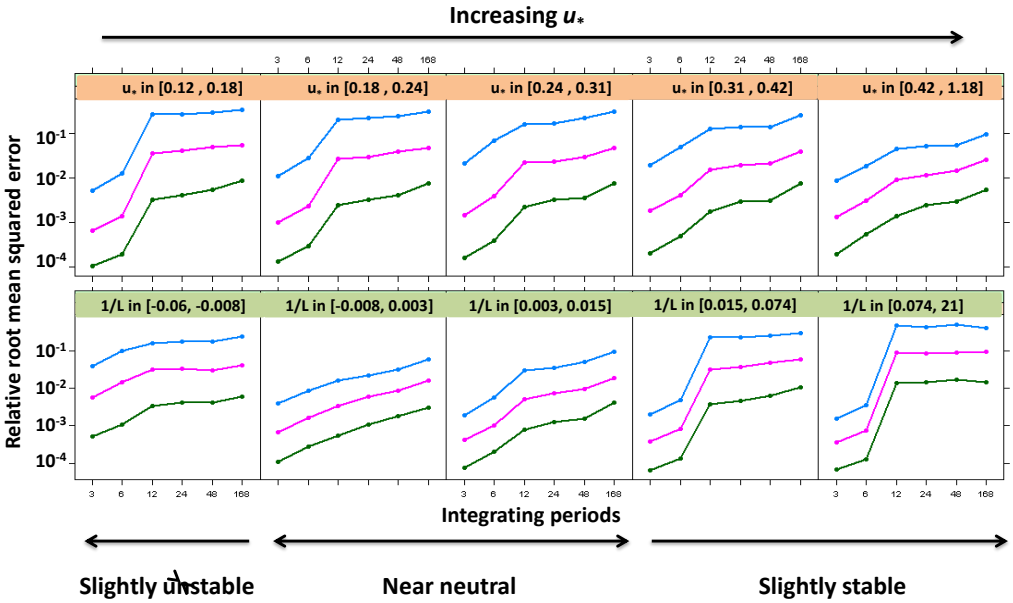

Figure 7. Relative root mean squared error as a function of integration period for stability factor and friction velocity classes for a single 25 m side field. Medians and quartiles are given for equally sized bins of u_* and $1/L$ and for the lowest sensor height (0.25 m). The blue, pink and green curves are the 3rd, 2nd and 1st quartiles, respectively.

Figure 5 & 8: I assume that prescribed is the emission data provided by calculation and inferred is emission calculated by knowing NH3 concentration at 0.5 m and weather conditions.

Indeed prescribed emissions are calculated using equations (9) and (10) with a constant Gamma, and inferred are those based on measured concentration at 0.5 m height and transfer coefficient using equation (7).

Fig. 7: Need improvement

We agree. See previous section for our proposition.

Fig 9; Why not mention the emission strength of the source instead of Treatment 1-3 (what is the units?)

This is indeed a sound remark. We propose indeed to use the emission potential Γ, which actually has no units. Figure 9 would look like this (Figures 10-12 would be changed accordingly):

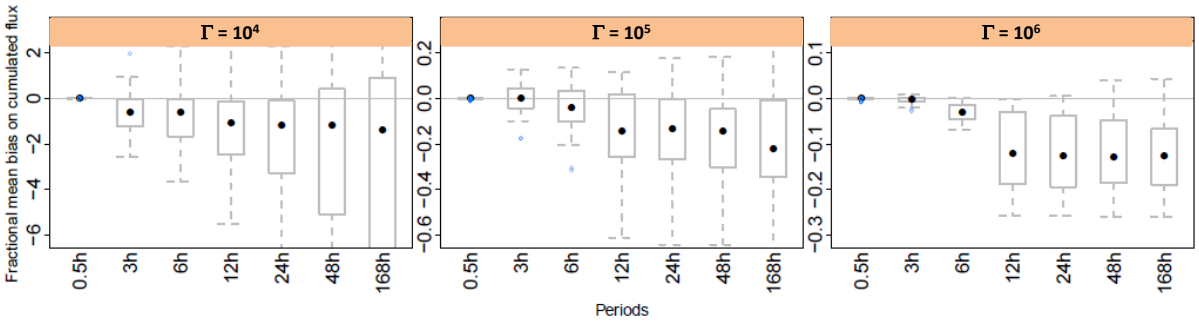

Figure 9. Effect of integration period on source inference in the multiple-plot setup. The fractional mean bias of the source is shown for each treatment. Inference strategy C1 was used (single sensor, independent blocks, and background concentration known). Statistics for runs with target heights 0.25 and 0.5 m and source side = 25 m are calculated. All application periods are considered. Filled points show medians, boxes show interquartiles and bars show minimums and maximums. Outliers are points to 1.5 times away from boxes limits.

Fig 16: Is it correct that measured emissions are not included – if so then the measured results should be included?

There might be a misunderstanding of that figure; indeed Figure 16 reports inferred emissions using the multiple-source inference method that was presented and discussed in this manuscript. But in that experiment no other method than this one was used to "measure" emissions. This is actually our point to show that this inference method is appropriate for estimating the emissions under real situations. In a way our inference method gives a measurement of the ammonia emissions.

References quoted in the answer to reviewer 1

Laubach, J., Taghizadeh-Toosi, A., Sherlock, R.R. and Kelliher, F.M., 2012. Measuring and modelling ammonia emissions from a regular pattern of cattle urine patches. Agric. For. Meteorol., 156: 1-17.

Wilson, J.D., Thurtell, G.W., Kidd, G.E. and Beauchamp, E.G., 1982. Estimation of the rate of gaseous mass transfer from a surface source plot to the atmosphere. Atmospheric Environment (1967), 16(8): 1861-1867.

Answer to Referee 2 comments

We thank Referee 2 comments which we hope will help us improving our manuscript.

General Comments

Loubet and others study a new method for inferring ammonia loss from small agricultural plots. I found the modelling analysis to make for an interesting case study regarding the applicability of field experiments. The approach treats bias errors carefully. As a consequence, I feel that the manuscript makes an earnest effort to quantify biases associated with passive ammonia sampling over small agronomic field plots and will be a valuable contribution to the literature.

Minor comments:

'Further work should anyway be produced for validating this method in real conditions' at the end of the abstract does not sound hopeful. Rather, the authors should try to discuss strategies for further improving the method and reducing uncertainties.

We thank the reviewer for this comment. We have indeed identified two strategies for further improving the method: (1) using Bayesian inference which has the potential for constraining the emissions and avoiding unrealistic sources inference as shown by Yee and Flesch (2010), and (2) changing the cost function (also called objective function); instead of inferring the emission strength, we could infer the emission potential Γ (a strictly positive number). This last method has the advantage of avoiding non-plausible deposition fluxes, because the flux is calculated as the concentration above the source minus the concentration at the ground, divided by transfer resistance. With such an approach negative fluxes (deposition) can occur within the limit of plausible transfer resistances but not above. We believe that the combination of these two strategies has the potential to improve the method substantially.

We think that "calibration" of the method against controlled sources is a remaining challenge that needs to be tackled (as also suggested by the comments of A. Neftel and C. Hanni in the interactive discussion).

We hence propose to add this more positive statement at the end of the abstract: "We believe that the method could be further improved by using Bayesian inference and inferring surface concentrations rather than surface fluxes. Validating against controlled sources is also a remaining challenge."

Line 41: 55.3% sounds remarkably specific given uncertainties in measuring NH3 flux.

This is a sound comment. We propose to change to 55%.

53: 'most of the time large fields' is awkward wording.

Indeed. We propose to change to “most of the time also requires the use of large fields”

57: agronomic trials are not necessarily of those dimensions.

This is a sound remark indeed and we should not be as general as we were. We would propose to change to: “Especially useful for measuring ammonia losses are methods that can deal with small and medium-scale fields (20-50 m on the side) that are commonly used in agronomic trials.”

Parentheses on line 67.

Thanks for spotting this. We have withdrawn the left parenthesis.

118: quotes are unnecessary.

We agree and have withdrawn them.

On 130, what is the typical reaction time (and thereby Damkoehler number?)

Typical Damkhöler numbers showed by Nemitz et al. (2009) above a cut grassland canopy fertilised with ammonium nitrate were from 0.001 to 1. Values greater than 0.1 only occurred marginally, and usually during night-time conditions (Figure 6 in Nemitz et al. 2009). We would of course expect larger Damkhöler number values for slurry application which may generate larger concentrations than those reported by Nemitz et al. (2009), or with surface canopies having larger residence times. But in any case we expect the chemical depletion of ammonia to remain small at the spatial scale we are focussing on (around 200-300 m).

I find the tau near the overbar in 2 and other equations to be a bit distracting because it could be confused with an exponential term.

This is a sound remark. We propose to remove the *taus* and just leave an explanation in the text that the overbars denote averages over the period tau.

Equation 4 could be rearranged to reflect that only the numerator of the second term on the right hand side is unknown.

It is true that the numerator of the second term is the only unknown. However we can’t see how to isolate this term apart from multiplying by D(x). Moreover, leaving the equation as it is now has the advantage of explicitly showing this term which is the bias. We hence propose to keep equation (4) as it is.

251: why is z_{ref} 3.17 m? The curly braces in $R_b\{NH_3\}$ I find to be a bit distracting.

Regarding z_{ref}, we subjectively choose to use the reference height z_{ref} as the height where our ultrasonic anemometer was placed in the field, which simplified the calculation of the aerodynamic resistance for us. This does not have much importance anyway as we assumed that atmospheric ammonia concentration was zero.

Regarding Rb, we propose to change Rb{NH3} to R_{bNH_3}.

263: is there a justification for the model in simulation 2?

Exponential decrease in emission potential is representative of strong NH_3 emissions like those happening following slurry application. The value of 4.6 and the time scale τ_0 were chosen arbitrarily and would represent emissions a little bit less intense than those for nitrogen applications reviewed by Massad et al. (2010). In fact the equation we used here would be equivalent to a time scale equal to 6 days while in Massad et al. (2010) they report a time scale of 2.88 as being representative of slurry application. We propose to add the following text in Line 270: "The time scale of the exponential decrease we used here was around 6 days, which is twice as large as the one reported by Massad et al. (2010) for slurry application (2.9 days)."

265: what are typical parameters for the Gaussian model? Also, what mechanism causes it? The urea spreader?

The Gaussian model is rather representative of urea application. Indeed, NH_3 emissions result from combined processes: first the urea is hydrolysed by urease enzymes which release ammonium which can be volatilised but can also be nitrified or absorbed by roots. This leads to typical emissions starting a few days following application and showing a maximum up to 15 days following application but also a slower decrease of the emissions following the peak.

The Gaussian model was centred on day 14 with a standard deviation of 8.4 days.

267: I understand why 4.6 now in simulation 2. . .but why does this 'best' represent NH3 emissions?

As explained in previous paragraphs and following Massad et al. (2010) this model best represents slurry applications.

302: why is the covariance term negligible at the half hourly period? The spectral gap in eddy covariance studies?

The covariance term is indeed negligible at that time scale because of the so-called spectral gap in eddy covariance studies. This gap corresponds to time scales at which there is little energy in the

turbulence and surface flux spectra (see e.g. Van der Hoven {, 1957 #25437}). We propose to replace the sentence at line 302 by “In practice the concentrations were computed at each sensor location using Eq. (6) over 0.5h: at that time scale, which corresponds to the spectral-gap, the covariance term is assumed to be negligible (Van der Hoven, 1957).”

303: in 2.5.3, these are not hypotheses as they cannot be falsified, even in the model.

This is indeed an interesting remark. We propose to change to the term “scenario” instead.

327: extra point

Thanks for spotting this. We have removed it.

336: how close is ‘nearby’? From the figure it looks like it was part of the larger setup.

The meteorological data were measured at around 25 m away from the edge of the central plots (Figure 2). We propose to change the sentence for clarification: “The meteorological data were measured at less than 50 m from the central plots (Figure 2)”.

355: results should be written in the past tense.

Thanks for the comment. We propose to change this sentence also to clarify its meaning : “The friction velocity u* varied between 0.024 and 1.181 m s-1, and the stability parameter z/L varied between -49 and 21 m-1 (Figure 3)”

365: define Gamma for the reader in the figure legend.

Thanks for the comment. We propose to change the last part of the legend to “…with an emission potential = 10000”

Please avoid using red and green simultaneously in Figure 4. This figure appears to be made using R, and gray is also a default color. And honestly yellow is never a good choice on a white background.

The comment that Figure 4 was hard to read was also made by reviewer 1. We have hence simplified the Figure and we further propose to change the colors as suggested by reviewer 2:

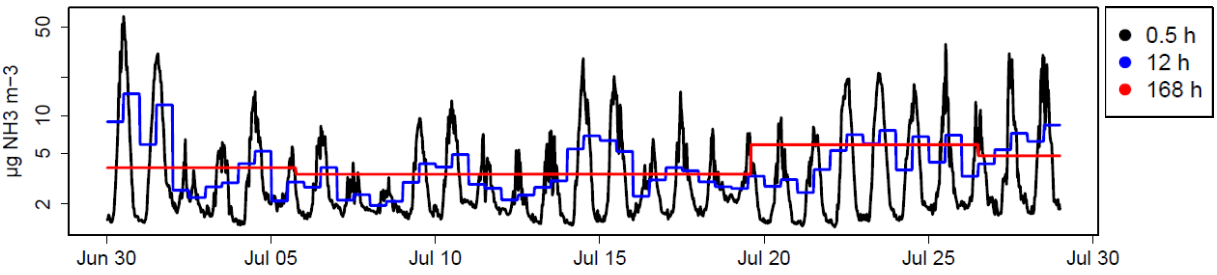

Figure 4. Example modelled concentration pattern at 1 m above a single 50 m width source for several averaging periods (0.5h, 12h and 168h) for the month of July 2008. The source Γ was set to 10^5. The y-axis is log scaled.

384: focuses

Thanks for spotting this typo. We have corrected it.

Figure 6 confuses me a bit because the 13 periods vary so strongly in their meteorological conditions from summer to winter, why are they grouped? The bars also leave the figure in the upper left subplot.

The idea for grouping the periods in Figure 6 but also in Figures 9-11 and 14 is actually to evaluate the variability of the bias due to meteorological conditions: ideally, if the method shows little variability in the bias, this bias could be characterised and even withdrawn. In Figure 6 we try to give a broad view of how the bias changes with sensor height and plot width. Figure 7 actually shows the variability of the bias due to meteorological conditions.

Regarding the scale, we chose to have a single scale for all panels to ease the comparison between heights and plot size, and we also chose to get the scale focussed enough to better see biases in the range -0.2 to 0.1. What we conclude from the upper left subplot is that the bias is much larger than all other cases which shows that that combination height-plot size is not satisfactory.

464-466: the attribution of stability with respect to continental vs. oceanic sites is too much of an approximation. There are many continental sites that are consistently windy, often due to orography.

We agree that we might have been too approximate in this statement, although we might still agree on the fact that oceanic conditions are typically windy. We propose to withdraw the reference to continental or oceanic climate to make it more general and replace the sentence for the following one: “We conclude that the inference method with a long integration period will lead to very moderate biases for locations with near-neutral conditions and high wind speed, but may lead to much larger bias under stable conditions and low wind speed as soon as the integration period gets up to 12 hours.”

There is a strange x on line 468. Font sizes for figure 7 should be increased.

Thanks for spotting the x. It came from a problem when pasting Figure 7. We propose to modify Figure 7 to increase font size and improve as follows:

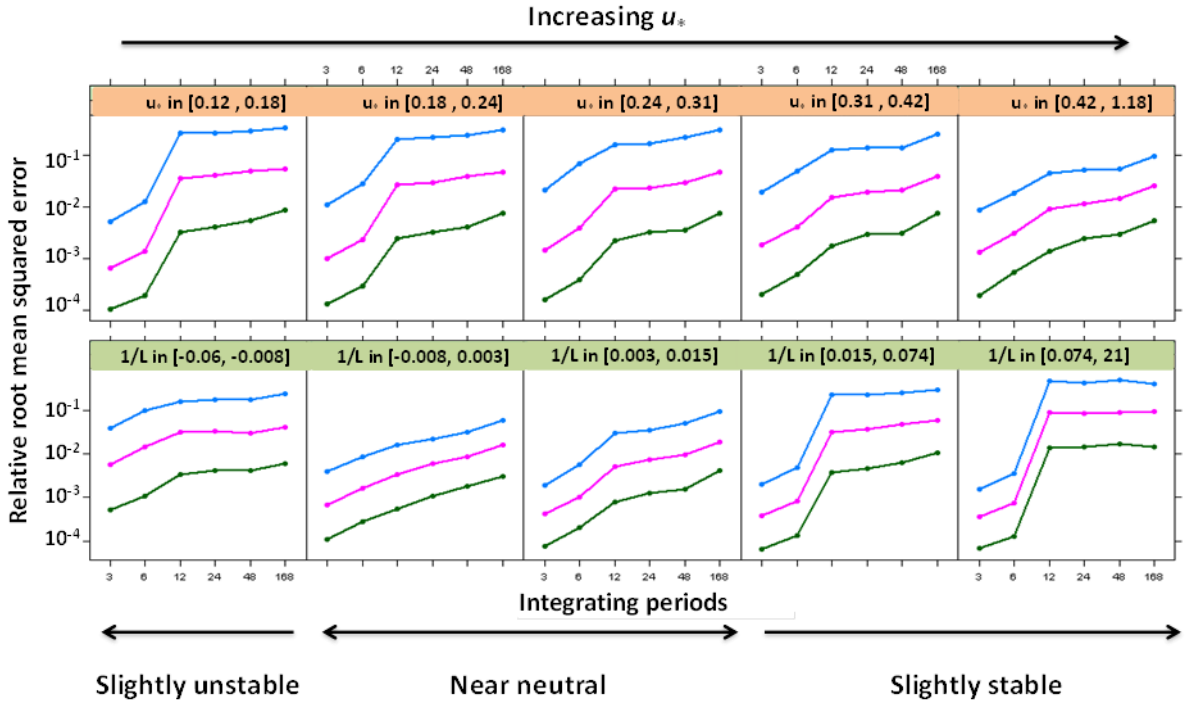

Figure 7. Relative root mean squared error as a function of integration period for stability factor and friction velocity classes for a single 25 m side field. Medians and quartiles are given for equally sized bins of u_* and $1/L$ and for the lowest sensor height (0.25 m). The blue, pink and green curves are the 3rd, 2nd and 1st quartiles, respectively.

741: why bird colonies?

Actually the only references we found where this bias was evaluated were those from emission estimates from bird colonies. We propose to withdraw the mention to bird colonies as this does not add much to the conclusion statement.

742: again, continental does not imply low wind speeds.

As in previous comment we propose to withdraw the reference to continental climate.

References

Massad, R.S., Nemitz, E. and 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.

Nemitz, E. et al., 2009. Aerosol fluxes and particle growth above managed grassland. Biogeosciences, 6(8): 1627-1645.

Yee, E. and Flesch, T.K., 2010. Inference of emission rates from multiple sources using Bayesian probability theory. J. Environ. Monit., 12(3): 622-634.

Answer to Comment to the paper „Evaluation of a new inference method for estimating ammonia volatilisation from multiple agronomic plots "by Benjamin Loubet et al. by Albrecht Neftel (Neftel Research Expertise, CH-3033 Wohlen b. Bern, CH, Switzerland) and Christoph Häni (School of Agricultural, Forest and Food Sciences, Bern University of Applied Sciences, CH-3052, Zollikofen, CH, Switzerland)

This paper presents an appealing and low-cost approach to determine NH3 losses from adjacent multiple agronomical plots by a combination of concentration measurements using passive sampler devices and a dispersion model that is driven by turbulent parameters inferred from standard 30 minutes meteorological data.

The aim of the paper is described as: "Can inverse dispersion modelling approaches be used for inferring NH3 emissions from multiple small plots (agronomic trials) using passive samplers, and to which degree of accuracy?"

The overall answer is encouraging with the statement in the conclusions "In this study we have demonstrated that it is possible to infer with reasonable biases ammonia emissions from multiple small fields located near each other using a combination of a dispersion model and a set of passive diffusion sensors which integrate over a few hours to weekly periods".

According to our judgement the accuracy will be mainly determined by two aspects addressed below.

a) Bias related to the applied dispersion modelling

Dispersion models are a simplified mathematical representation of the turbulent motion in the surface layer and will always deviate from reality. Systematic biases can be expected when modelling the lower heights of the measurements that are discussed in this paper. For concentration sensors place close to the ground (e.g. 25cm above ground) transfer functions are likely to be biased due to e.g. the needed simplifications that must be made to describe the exchange process at the ground, the natural heterogeneity on a small scale at the surface or the violation of model assumptions such as the failure of K-theory close to the canopy (Raupach and Legg, 1984). Furthermore, the translation of the sensor height in a model framework is challenging for very low heights since the sensor height value (and with that the resulting value of D at that location) becomes very sensitive to sensor height measurement errors as well as to the absolute values of z0 and d. To us, a sensor height of 25 cm seems too close to the surfaces.

We do agree with A. Neftel and C. Häni concern that model representations may be biased close to the ground, and especially Gaussian like models, since they do not intrinsically account for near-field dispersion as shown by Raupach and Legg (1984) and subsequent publications from M.R. Raupach (Raupach, 1987; Raupach, 1989b). This is not the case of the Langevin models that account for near-field dispersion (Raupach, 1989b; Thomson, 1987). However, as exposed by Raupach (1989a), the height at which the near-field effect is sensible would be equal to $\sigma_w T_L \sim 0.3\ \sigma_w h / u_* \sim 0.3 \times 1.25 \times h \sim 0.4\ h$ where σ_w is the vertical air velocity standard deviation, h is canopy height and T_L is the Lagrangian time scale. Numerical values were derived from Raupach (1989a), Figure 1. Hence we would expect this far field effect to be small on situations with small canopies (or by extension with

small roughness height for a bare soil). Typically this would correspond to about 5 cm above a 10 cm canopy and ~1 cm above a 3 cm roughness. Hence we agree with A. Neftel and C. Häni that this would represent a quite important fraction of the sensor height if this sensor would be placed lower than 50 cm (20% for 25 cm). This would especially be critical for canopies that are taller than 10 cm.

Regarding the uncertainty in determining the height of the sensor close to the ground, this is a very sound remark. We however see in Figures 6 and 10 of the manuscript that the method gives similar biases for sensors placed at 25 cm and 50 cm above ground (and also close biases for h = 1 m), for plots of 25 m x 25 m. We hence agree with the concern of A. Neftel and C. Häni that 25 cm would be too low and we should rather target heights of 50 cm. We propose to add this statement in the conclusions: *"Although the lowest sensors have the best condition number, we would rather recommend in practice using heights of 50 cm above the canopy in order to reduce uncertainty in positioning the sensors close to the ground as well as avoid being too close to the roughness layer close to the canopy which is characterised by non-diffusive transfer."*

The authors are using their FIDES-3D model that is based on an analytical solution of the advection-diffusion equation. This model is compared with the backward Lagrangian Stochastic dispersion (bLS) model described in Flesch et al. (2004) (the "WindTrax" software, Thunder Beach Scientific, Nanaimo, Canada). For the presented analysis the FIDES model Kz was adopted to match the far field approximation of Kz of the bLS model. We are missing an explanation, why this was done.

The aim of matching the far-field diffusivity of the two models was to make the FIDES model consistent with the bLS approach, which is a commonly used method nowadays for estimating ammonia emissions with inversion techniques. As exposed in the manuscript, one major difference arises from the fact that the Phillip (1959) approximation of the advection-diffusion equation (which is identical to the approach of Korman and Meixner (2001)) has a Schmidt number equal to 1, while bLS approaches have a Schmidt number equal to 0.64. Wilson (2015) showed that the choice of the Schmidt number has a great effect on footprint, and hence on concentration above a small source and should therefore be explicitly given. Wilson further showed that the difference in footprints predicted by diffusion and Langevin models (like bLS) are small under neutral and stable conditions provided they have similar Schmidt numbers, although the difference in footprints remains large under unstable conditions even with identical Schmidt number (with the Langevin models diffusing less than the Eulerian ones). We hence chose to use an approach that was as close as possible to the bLS approach. To do so, we matched the far field diffusivities of the two models, as this would ensure that the two models would provide similar concentrations at heights larger than a few decimetres. Moreover there was no point in matching near-field dispersion, as FIDES does not account for near-field dispersion.

In the supplement, a detailed investigation is presented how the two models differ in their formulation of the vertical diffusivity Kz. The assumed far field vertical diffusivity in the bLS model is approximated by parametrizations provided in Flesch et al. (1995). We would like to remark that WindTrax uses slightly different default parametrizations of σ_w than provided in Flesch et al. (1995) (see e.g. the manual on the WindTrax homepage). This is resulting in vertical diffusivities given as:

$$Kz\,(Z) = 0.5 * 1.25 * u*Z/(1 + 5Z/L) \text{ for } z/L \geq 0 \qquad \text{(WT1)}$$

$$Kz\,(Z) = 0.5 * 1.25 * u*Z \times (1 - 6Z/L)^{0.25}\,(1 - 3Z/L)^{(1/3)} \text{ for } z/L < 0 \qquad \text{(WT2)}$$

with a Schmidt number value of $Sc \cong 0.64$ for near-neutral stabilities with a smooth transition from L = ∞ to L = -∞. These equations differ from the equations S7 and S8, and imply a different interpretation of the differences between FIDES and WindTrax, though without changing the numeric results of the comparison.

We would like to thank C. Häni and A. Neftel for providing the exact expression used in Windtrax. We indeed only referred to the work of Flesch et al. (1995). These equations WT1 and WT2 are more consistent than those we reported since, as opposed to equations S7 and S8, they insure continuity in $Kz(Z)$ when $z/L \to 0$. We see from the set of equations WT1, WT2, S7 and S8 that $Kz(Z)$ is similar under non-neutral conditions in S7 and WT1 though 4% smaller in Windtrax but that $Kz(Z)$ is 16% smaller in Windtrax (WT2) than in the "tuned FIDES" (S8) under unstable conditions. This therefore explains better Figure S3 which shows a good fit between the "tuned FIDES" and WindTrax under stable and neutral conditions but a lower concentration modelled with the "tuned FIDES" at 2 m above ground under unstable conditions. Indeed, since the far-field diffusivity is larger in the tuned FIDES, this model lead to larger diffusion and hence lower concentrations away from the source. However, another difference comes from constitutive differences between Eulerian and Langevin models under unstable conditions as shown by Wilson (2015).

We have checked that the results reported in this manuscript remain mostly unchanged since already in line with the most important feature of $Sc \sim 0.64$. We however quantified a difference of around -18% ± 10% in the concentration modelled using equations S7 and S8 (Flesch et al., 1995) compared to equations WT1 and WT2 (Flesch et al., 2004) for a single source of 25 m x 25 m. Since this difference is systematic and since we use the same model for forward and backward modelling we do not expect any impact on the conclusions we have drawn from this study. Indeed, we checked for a single source of 25 m x 25 m that the biased inferred using eq. S7 and S8 and WT1 and WT2 were similar within less than 1% for most cases. Noticeably, the biases for the highest sensors were diminished with WT1 and WT2. We hence propose to leave equations s7 and S8 in section S4.2 as they are but to stipulate explicitly that these correspond to Flesch et al. (1995) and not to WindTrax. We also propose to modify Figure S3 to explicitly mention this (see new Figure 3 below).

In the supplement Figure S3 presents a comparison between evaluated concentrations with FIDES and WindTrax respectively using the prescribed emission sources with the SVAT model. This figure is hiding the apparent differences as a double logarithmic representation is used and the concentrations are shown using an emission source that shows a positive correlation between the meteo input parameters of the models and the source strength. E.g. for neutral conditions the regression of ratio of the concentrations calculated with FIDES and WindTrax at a height of 0.25m is indicated as $cFIDES = 0.97 \cdot cWindtrax^{0.87}$. For a concentration of 1 the ratio is 0.97 and for a concentration of 100 the ratio becomes 0.53. As the transfer function D in FIDES and WindTrax are only depending on the prevailing turbulent parameters it would be more instructive to use a constant unit emission of 1 and show the ratio on a linear scale as function of u_* and L in a similar way as the authors have done in a previous paper (Carozzi et al., 2013).

We thank A. Neftel and C. Häni for this very useful comment. We propose to modify Figure S3 and section S4.3. Below is given the proposed updated Figure S3 showing the comparison between the two models using linear regressions forced to zero and graphs with linear scales. We propose to change the text in the supplementary material to the following: *In Figure 3, we notice that the concentration modelled with the tuned FIDES at 2 m height was lower by roughly 15% compared to WindTrax under unstable conditions but is comparable under stable and neutral conditions. Lower down at 0.25 m height, the tuned FIDES systematically underestimated the concentration by 15-22% whatever the stability.*

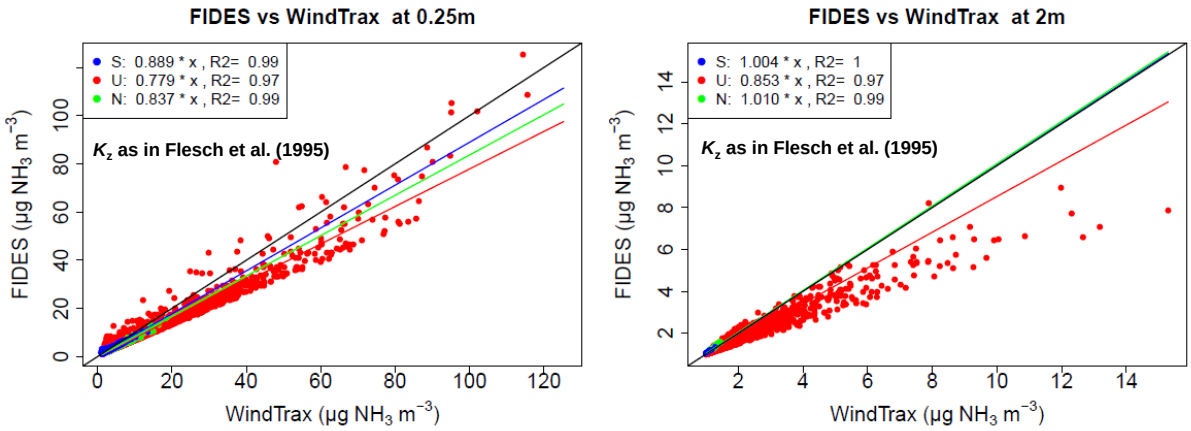

Figure S3. FIDES versus Windtrax concentration modelled above an ammonia source of 25 × 25 m at 0.25 and 2 m heights. The FIDES vertical diffusivity K_z fits the WindTrax K_z. The comparison is made over the entire year of 2008 in the FR-Gri ICOS site. S, U and N stand for stable, unstable and neutral atmospheric conditions. The linear regression equation is given for each condition together with the R^2 of that regression. The black line is the 1:1 line.

Below we also computed the graphs showing the concentration residual as a function of 1/L, as suggested by C. Häni and A. Neftel (Figure S4). We propose to include Figure 4 and the following text in the supplementary material: *Figure 4 shows that under unstable conditions FIDES underestimated the concentrations up to 30% at 0.5 m compared to WindTrax, while this gap was reduced and more scattered at 2 m height. Moving towards neutral conditions the two models tend to agree notwithstanding an overestimation of 10% by WindTrax at 2 m height concurrently with an underestimation of the same magnitude by FIDES at 0.5 m. Under stable conditions there was a good agreement at 2 m height, while this agreement remains poorer at lower heights.*

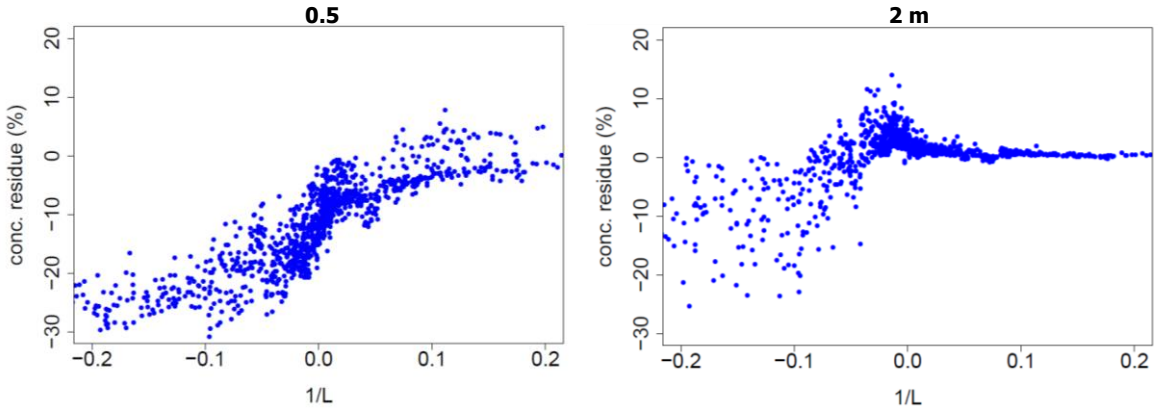

Figure S4. Relative difference between FIDES and WindTrax concentrations as a function of the stability parameter (1/L). Data refer to the same conditions reported in Figure S3.

b) Bias related to the concentration measurements

The use of passive diffusive samplers is a challenging business. Within different networks the reliability of PS such as ALPHA samplers or Radiellos have been proven, but the use of them close to emitting source showed major deviations compared to other measurements. E.g. Misselbrook et al. (2005) found severe overestimations of passive diffusive samplers. The latest investigation stems from the Dronten experiment and is discussed in a paper by Michael Bell et al. (submitted to AFM). In this experiment the ALPHA samplers were affected by a positive bias in the order of 50% relative to the other devices. We speculated that the exposure of the PS with the protection hat above them cached eddies from below loaded with higher NH3 concentrations but shielded eddies with lower concentrations from above. Figure 1 illustrates the NH3 dynamic that occur over an emitting surface. The concentration was measured with a fast device described in Sintermann et al. (2011). Immediately after application of slurry with a splash plate the NH3 concentration was measured at a height of 1m above ground with an ionization technique and a strongly heated inlet line to avoid as much as possible damping effects.

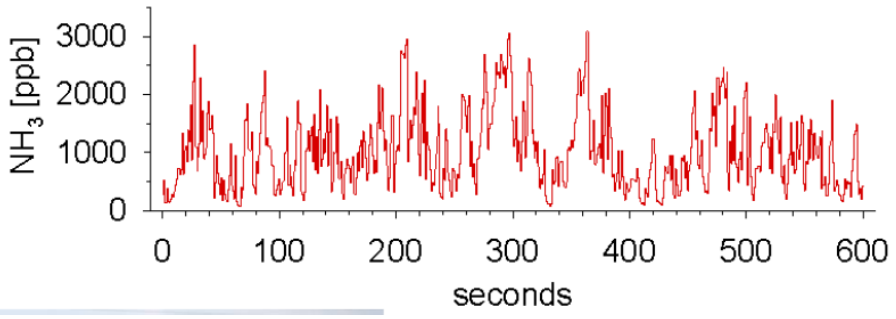

Figure 1: NH3 concentration time series measured 1m above ground over a manured surface with splash plate.

First we notice as A. Neftel and C. Häni that ALPHA badges are very reliable for network measurements concentrations, as for instance showed by the recent Met-NH3 project (http://www.metnh3.eu). We are however very much aware that the use of ALPHA badges close to emitting sources may be biased. The reason for that bias is however unclear and would need comparisons with fast and unbiased sensors. The assessment by Misselbrook et al. (2005) shows that with high concentrations diffusion samplers may lead to overestimation of up to 70% of the concentration. They suggest potential issues related to the deformation of the Teflon membrane which would modify the distance between coated filters and the Teflon membrane that could cause sampler saturation.

The speculations from A. Neftel and C. Häni are interesting. One could indeed speculate that sweeps, which dominate turbulent transport near the top of the canopy and are characterised by lower wind speed with positive vertical velocity (Poggi and Katul, 2007), could lead to an artificial build-up of the concentration underneath the protecting caps. We could speculate that ejections would not be efficient in "purging" the volume underneath the cap and hence letting over time the concentration being higher in this area. One could also speculate on adsorption-desorption of ammonia on the walls of the ALPHA badges that would be non-linear in response to NH_3 concentrations and lead to possible over estimations under highly fluctuating concentrations as shown in Figure 1 above. This issue necessitates experimental validation of the methodology anyway. We hence propose to add the following text in the discussions section 3.6: *Misselbrook et al. (2005) compared different methodologies and showed that under high concentrations diffusion samplers may lead to overestimation of up to 70% of the concentration. They suggest potential issues related to the deformation of the Teflon membrane which would modify the distance between coated filters and the Teflon membrane that could cause sampler saturation. There is hence some concern on the quality of diffusion samplers to measure concentrations close to large sources which would necessitate field validations.*

We also propose to add a sentence at the end of the conclusion to emphasize this issue for future work: *"Special care should be taken in validating the use of ALPHA samplers near very strong sources"*

Concluding comments:

We judge that the most important potential biases of the proposed multiplot approach are related to biases of the concentration measurements and the used dispersion coefficient. It would be instructive to calculate probability density functions of the estimated emissions with a dataset that reflect the distributions of the measurements and the turbulence parameters that drive the dispersion model.

This would indeed be very instructive but we feel that this issue is rather a work to be for a next study. Indeed, in this study we have explored the first order variability which is driven by the change in meteorological conditions observed in the 13 periods over a year in typical western European climate. The next step could be to extend this assessment to other datasets part of Fluxnet network to incorporate more continental climate conditions. The script we developed for this study actually incorporates measurement noise but we disabled this feature for calculation time reasons.

The authors have tested their setup in field trial in April 2011 applying slurry with a DM content of 6% and an application rate of 41 kg N-NH3/ha. According to the details given in the text, we assume that broadband application was used and was compared to fast incorporation and no application. The cumulated loss amounted to 8 to 10% of the applied NH3. For broadband application, this is a loss on the low side (see e.g. Häni et al.,2016). We would not be astonished if the real emissions would be double as high.

We would like to thank very much A. Neftel and C. Häni for their question and to grant them for their guess. Indeed, we double checked the calculation script used for the real test case and we found one bug in the calculation of the cumulated NH_3 emissions: the multiplicative constant to account for the number of second per time step was set to 30 min while the time step of that particular test case was 60 min. Since a 60 min time step is unusual we did not spot this error in the first place, but we have since then changed the script to calculate the time step from the meteorological dataset. We used this new script and found this bug. This change results in a doubled emission compared to what was given in the discussion manuscript. We hence find, as guessed by A. Neftel and C. Häni that the cumulated losses represented around 20% of the nitrogen applied. We propose to change Figures 16 and 17 as below:

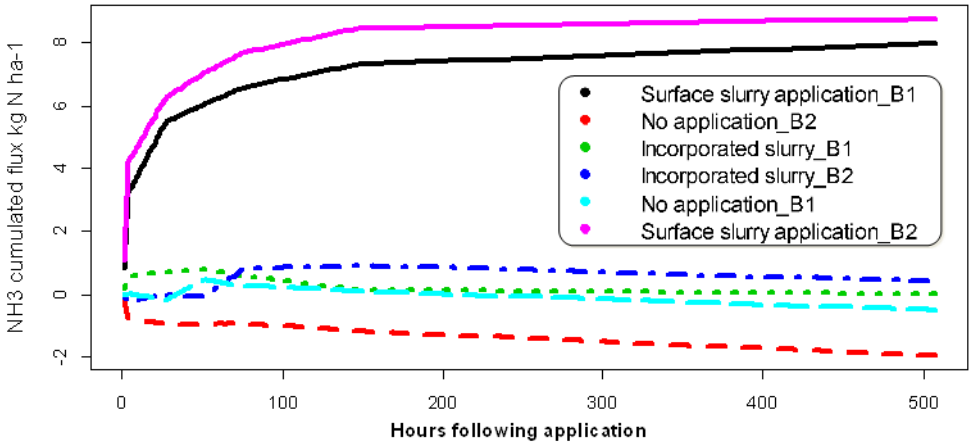

Figure 16. Cumulated fluxes estimated with the inference method on the real test case with strategy C7. Three treatments with two repetitions are compared.

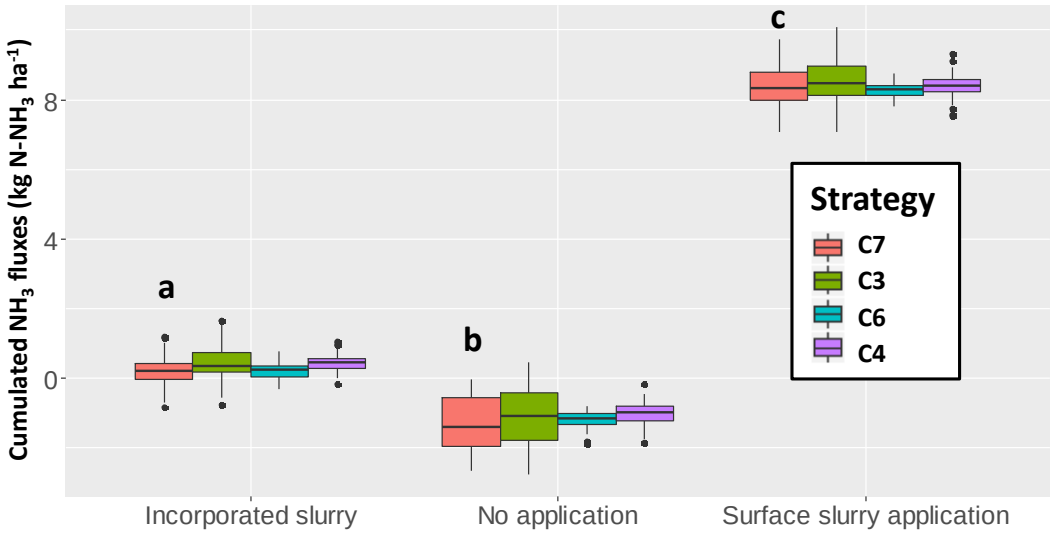

Figure 17. Same as Figure 16 but grouped by treatments and with additional strategies C4 and C6 which consider that replicates have the same surface flux. The variability in the boxplot aggregates the uncertainty on the inference method (the standard deviation on the flux estimate in the least-square model, which accounts for the variability in the replicated concentration measurements), and the variability between the repetitions in each treatment. Letters *a, b* and *c* show significant differences between treatments for the C7 strategy, according to a Tukey test (95% family-wise confidence level).

We also propose to modify the text in section 3.5 (lines 630-636) to: *Surface slurry application showed the largest emissions: 9 ± 0.3 kg N ha-1in B1 and 10 ± 0.2 kg N ha^{-1} in B2 (median and confidence interval). This corresponds to an emission factor around 23% of the N-NH$_4$ applied and 8% of the total N applied, which is in-line with agronomic references (Sintermann et al., 2011a; Sommer et al., 2006). In contrast, the incorporated slurry showed much smaller emissions: 0.3 ± 0.2 kg N ha^{-1} in B1 and 0.6 ± 0.2 kg N ha^{-1} in B2. It is noticeable that the no-application showed slight deposition, especially in B2: -0.26 ± 0.2 kg N ha^{-1} in B1 and -1.7 ± 0.2 kg N ha^{-1} in B2.*

We also propose to change lines 659-663 as follows: *Therefore, we could expect that the real flux is the one measured with C7 times 1.15 (± 0.08), hence would be 10.9 ± 1.3 kg N ha^{-1}. This corresponds to 27 ± 3 % of the N-NH$_4$ applied and ~9 ± 1% of the total N applied. For the incorporated slurry, the emissions are around 20 times smaller than the emissions from the surface applied slurry. Under these conditions, the bias on the emission would be around -20%, which means that the corrected emissions would range from 0.5% to 2.5% of the N-NH$_4$ applied and 0.2 and 0.8% of the total N applied.*

The presented approach to perform NH3 emission measurements in a multiplot arrangement is encouraging and goes in a good direction. To make the approach more robust, the employed ALPHA NH3 sampling systems should be validated under real conditions, i.e. over an emitting source in comparison with e.g. MiniDOAS systems (Sintermann et al., 2016).

We completely agree with A. Neftel and C. Häni about this issue. Although the ALPHA badges have shown to be very precise in comparisons under laboratory and field conditions, it is worth comparing

them with an independent technique in a situation where the source is small and intense and where the sensor is placed near the ground (e.g. 50 cm above ground).

Finally, we would like to invite the authors to collaborate with us to compare the FIDES and WindTrax approach. We have an extensive dataset from field trials where we released CH4 or a mixture of NH3 and CH4 from a circular artificial source with a diameter of 20 meters (Häni et al., 2017).

We are honoured by this invitation to collaborate and would be very happy to compare our approach with WindTrax on CH4 and NH3 emissions.

References:

Bell M.W., A. Hensen, A. Neftel, B. Loubet , P. Robin , Y. Fauvel , Y. Hamon , M. Haaima, A.J.C. Berkhout , D.P.J. Swart, W.C.M. van den Bulk, B.F. van Egmond, D. van Dinther, A. Frumau, B. Esnault, C. Decuq, C.R. Flechard Quantifying ammonia emissions from plot-scale and farm-scale sources using integrated mobile measurements and inverse dispersion modelling, (submitted to AFM).

Carozzi, M., Loubet, B., Acutis, M., Rana, G., Ferrara, R.M., 2013. Inverse dispersion modelling highlights the efficiency of slurry injection to reduce ammonia losses by agriculture in the Po Valley (Italy). Agric. For. Meteorol. 171, 306–318. 10.1016/j.agrformet.2012.12.012.

Flesch, T.K., Wilson, J.D., Harper, L.A., Crenna, B.P., Sharpe, R.R., 2004. Deducing ground-to-air emissions from observed trace gas concentrations: A field trial. J. Appl. Meteorol. 43 (3), 487–502.

Flesch, T.K., Wilson, J.D., Yee, E., 1995. Backward-Time Lagrangian Stochastic Dispersion Models and Their Application to Estimate Gaseous Emissions. J. Appl. Meteorol. 34 (6), 1320–1332.

Häni, C., Sintermann, J., Kupper, T., Jocher, M. and Neftel, A., 2016. Ammonia emission after slurry application to grassland in Switzerland. Atmospheric Environment, 125: 92-99.

Häni, C., Voglmeier, K., Jocher, M., Ammann C., 2017. Recovery rates from line-integrated NH3 and CH4 measurements using backward Lagrangian stocahstic dispersion modelling. Geophysical Research Abstracts, Vol. 19, EGU2017-19557, 2017

Kormann, R. and Meixner, F.X., 2001. An analytical footprint model for non-neutral stratification. Boundary Layer Meteorol., 99(2): 207-224.

Misselbrook, T.H., Nicholson, F.A., Chambers, B.J. and Johnson, R.A., 2005. Measuring ammonia emissions from land applied manure: an intercomparison of commonly used samplers and techniques. Environmental Pollution, 135(3): 389-397.

Poggi, D. and Katul, G., 2007. The ejection-sweep cycle over bare and forested gentle hills: a laboratory experiment. Boundary Layer Meteorol., 122(3): 493-515.

Raupach, M.R., 1987. A Lagrangian Analysis of Scalar Transfer in Vegetation Canopies. Q.J.R. Meteorol. Soc., 113(475): 107-120.

Raupach, M.R., 1989a. Applying Lagrangian Fluid-Mechanics to Infer Scalar Source Distributions from Concentration Profiles in Plant Canopies. Agric. For. Meteorol., 47(2-4): 85-108.

Raupach, M.R., 1989b. Stand Overstorey Processes. Philos T Roy Soc B, 324(1223): 175-190.

Raupach, M.R., Legg, B.J., 1984. The uses and limitations of flux-gradient relationships in micrometeorology. Agricultural Water Management 8 (1-3), 119–131. 10.1016/0378-3774(84)90049-0.

Sintermann, J., Dietrich, K., Häni, C., Bell, M., Jocher, M., Neftel, A., 2016. A miniDOAS instrument optimised for ammonia field measurements. Atmos. Meas. Tech. 9 (6), 2721–2734. 10.5194/amt9-2721-2016.

Sintermann, J., Spirig, C., Jordan, A., Kuhn, U., Ammann, C., and Neftel, A.: Eddy covariance flux measurements of ammonia by high temperature chemical ionisation mass spectrometry, Atmos. Meas. Tech., 4, 599–616, doi:10.5194/amt-4-599-2011, 2011.

Thomson, D.J., 1987. Criteria for the selection of stochastic models of particle trajectories in turbulent flows. Journal of Fluid Mechanics, 180: 529-556.

Wilson, J.D., 2015. Computing the Flux Footprint. Boundary Layer Meteorol., 156(1): 1-14.