https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Improving ammonia emission predictions with dynamic machine learning models

Armand Favrot[1, 2], Sophie Génermont[1], Vincent Guigue[2], Céline Décuq[1], and David Makowski[2]

[1]Université Paris-Saclay, INRAE, AgroParisTech, UMR EcoSys, Palaiseau, 91120, France
[2]Université Paris-Saclay, INRAE, AgroParisTech, UMR MIA-PS, Palaiseau, 91120, France

Correspondence: Armand Favrot (armand.favrot@inrae.fr)

Abstract. Ammonia emissions pose significant challenges for both environmental protection and human health. A substantial portion of these emissions occurs after field fertilization. Accurate prediction of these emissions is essential for national inventories and for identifying effective mitigation strategies. Although several static machine learning models have been developed to estimate final cumulative emissions, the potential benefits of dynamic machine learning to improve these predictions remain unknown. To address this gap, we compared 13 static models (1 random forest, 12 neural networks) and 33 dynamic models (7 random forests and 26 recurrent neural networks). The best performing model was a recurrent neural network, achieving an average mean absolute error (MAE) of 4.56 kgN/ha (95% CI $= [4.17, 4.95]$), corresponding to a decrease in MAE of 13.6% and 17.7% compared to the best static neural network and the static random forest, respectively.

1 Introduction

Ammonia (NH_3) is a major atmospheric pollutant. In Europe, 93% of total emissions, including both natural and anthropogenic sources, originate from agriculture (European Environment Agency, 2024). Of these agricultural emissions, around 20% comes from the spreading of livestock manure and 35% from synthetic fertilizers (Beusen et al., 2008). NH_3 concentrations in the atmosphere have increased in recent decades on a global scale (Warner et al. (2017); Van Damme et al. (2021)), probably due to rising temperatures and increased use of fertilizers (Shen et al. (2020); He et al. (2021)). Ammonia emissions contribute to the formation of particulate matter, particularly $PM_{2.5}$ which has been linked to respiratory and cardiovascular diseases (Brook et al. (2010)). NH_3 can also be transported long distances before deposited in soils and aquatic systems, causing acidification, eutrophication, and biodiversity loss (Krupa (2003); Behera et al. (2013); Guthrie et al. (2018)). Mitigation strategies are urgently needed to reduce these environmental and health impacts.

For more than four decades, various models have been developed to predict ammonia emissions following fertilizer application. They serve multiple purposes: identifying the main drivers of volatilization, providing estimates for national emission inventories, and supporting the evaluation of mitigation strategies. Emission inventories are particularly important in the context of regulatory frameworks such as the National Emission Ceilings (NEC) Directive in Europe (European Environment

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Agency (2022)), which sets binding targets for the reduction of five major air pollutants, including ammonia.

Mechanistic models (Génermont and Cellier (1997); Gericke et al. (2012); Sommer and Olesen (2000)), empirical models (e.g, Braschkat et al. (1997); Plöchl (2001); Misselbrook et al. (2005)), and semi-empirical models such as ALFAM2 (Hafner et al. (2019); Hafner et al. (2025)) have been developed by different groups of environmental scientists. Mechanistic models aim to describe precisely the physical and physicochemical processes involved in volatilization, but are often too complex, with many parameters and input variables, limiting their practical use.

Most empirical models are based on relatively simple mathematical tools such as regression analysis and Michaelis–Menten functions. For example, Huijsmans et al. (2018) used independent logistic regressions across different time periods to model the emission curve, while Plöchl (2001) and Lim et al. (2007) applied multilayer perceptrons to estimate the coefficients of models of Michaelis-Menten type. However, these models are often overly simplistic and lack flexibility, which constrains their performance.

With the development of new algorithms and the increasing availability of large agricultural datasets such as the ALFAM2 database (Hafner et al. (2018); Hafner et al. (2025)), machine learning offers promising perspectives for modeling agricultural NH_3 emissions. Recently, static machine learning models, predicting only the final cumulative emission (e.g., Menzi et al. (1998); Xu et al. (2024); Favrot et al. (2026)) have been developed. Favrot et al. (2026) applied random forests and gradient boosting machine learning models to the ALFAM2 database and showed that these methods outperformed the semi-empirical ALFAM2 model on an independent test set. Xu et al. (2024) used random forests to estimate emission factors for a wide variety of fertilizers at the global scale and applied their model to assess emission trajectories under different climate change scenarios. Machine learning can capture complex, nonlinear interactions among multiple variables, which mechanistic models often struggle to represent. However, advanced machine learning algorithms such as random forests and neural networks have not yet been used to predict the full dynamic profile of ammonia volatilization. Our study aims to fill this gap.

In other fields of research, several studies have highlighted the relevance of machine learning for dynamic environmental prediction tasks. For example, Chelani et al. (2002) used recurrent neural networks (RNNs) to predict daily sulfur dioxide concentrations in Delhi and showed that RNNs outperformed classical multivariate regression. Hamrani et al. (2020) compared nine machine learning algorithms, including lasso, random forest, and long short-term memory (LSTM) neural networks, for predicting greenhouse gas emissions from agricultural fields. They found that LSTMs offered the best performance for forecasting carbon dioxide (CO_2) and nitrous oxide (N_2O) fluxes. Feng et al. (2019) applied both random forests and RNNs to forecast air pollutants including sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and fine particulate matter ($PM_{2.5}$, PM_{10}), 24 hours ahead in Hangzhou, China, and reported higher accuracy and computation speed compared to traditional atmospheric models. These results suggest that, potentially, the use of dynamic machine learning could improve the accuracy of ammonia

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

emission predictions.

The objective of our study is to assess whether predictions of ammonia emissions could be improved by incorporating the full emission dynamics in advanced machine learning models, compared to static machine learning . To this end, we compared 33 dynamic models based on random forests and recurrent neural networks with 13 static models derived from Favrot et al. (2026), including neural networks. The models were trained and tested using the ALFAM2 dataset including results of 699 experimental trials. This dataset compiles measurements of ammonia volatilization dynamic after application of pig or cattle slurry - two important types of organic fertilizer. A key feature of this dataset is the strong heterogeneity in measurement time steps across experiments. To address this, we developed irregular time-step models that explicitly included the time interval as a predictor, which constitutes an original contribution of our work. In addition to static machine learning models, the semi-empirical ALFAM2 model (Hafner et al. (2025), R package version: 4.2) was included as a benchmark because this model is currently often used for practical applications. The best-performing models identified from these comparisons were then implemented to evaluate the efficacy of different manure application techniques in reducing ammonia volatilization, across a wide range of climate conditions and manure characteristics. The best model was made available in a Python package.

2 Material and method

2.1 Data

2.1.1 Data description

For this study, we used a subset of the ALFAM2 database (version 2.56), which contains time series of ammonia emissions following field fertilization trials collected by 27 research institutes across 12 countries (Hafner et al. (2018); Hafner et al. (2025)). We retained most of the 722 trials used in Hafner et al. (2025), which were conducted using micrometeorological measurement methods with pig or cattle slurry. Twenty-three trials were excluded because measurements began more than 0.5 h after fertilizer application or because the emission patterns were considered unreliable (e.g., no emissions during the first intervals or atypical responses). The final dataset comprised 699 trials, identified by the 'pmid' variable in the database, and the full list is provided in the Supplementary Material.

For each trial, we had access to cumulative ammonia emissions over multiple time intervals, meteorological variables measured over the same intervals (temperature, rainfall, wind speed), and plot-level variables (application method, incorporation type, time of incorporation, total ammoniacal nitrogen application rate (TAN), manure application rate, manure origin, manure pH, and dry matter content). Figure 1 illustrates the relative emissions over time for different application methods. Among these methods, broadcast, a technique in which manure is spread uniformly over the soil surface, is typically the most emissive. In contrast, other methods that reduce the exposed manure surface area, such as trailing hose, trailing shoe, or injection techniques, are considered emission mitigation strategies. Figure 2 displays the emission reduction percentages of these techniques

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

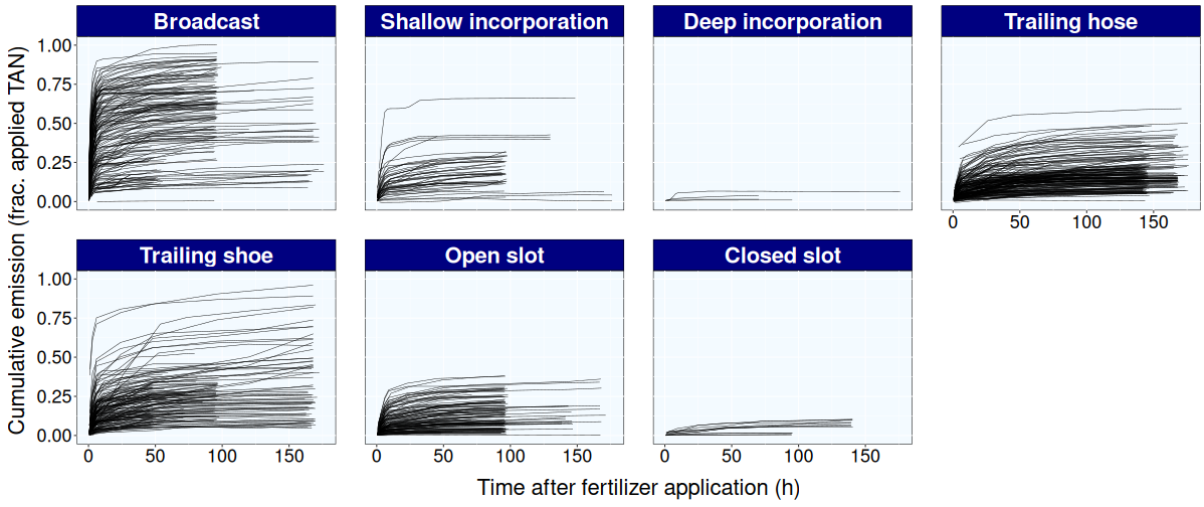

Figure 1. Cumulative relative ammonia emissions over time for different manure application and incorporation techniques. The y-axis shows the cumulative emission as a fraction of TAN, and the x-axis indicates time since manure application (hours). Subplots represent various combinations of application and incorporation methods: broadcast (no - shallow - deep incorporation), trailing hoses, trailing shoes, open slot, and closed slot. All trials are included except those combining incorporation with trailing hoses or shoes ($n = 4$).

(referred to as efficacy) compared to the broadcast method. These reductions were computed directly from the database using only experiments where at least one broadcast trial was available for comparison. An important aspect of these data is the heterogeneity in the time steps of measurements and trial durations. The observations were not recorded at regular time intervals, and the durations of the trials varied considerably. For example, the trial associated with pmid 2234 lasted 142 hours with measurements taken every 30 minutes, whereas the trial associated with pmid 3226 lasted 34.7 hours, with measurements recorded at (0.5h, 1.6h, 3.0h, 6.0h, 10.6h, 21.7h, 34.7h). A statistical summary of the dataset used is provided in Tables A1 and A2 (Appendix).

2.1.2 Data pre-processing

To mitigate discrepancies between short and long trials, measurements taken after 178 hours were excluded. For trials with measurements preceding manure application, we interpolated cumulative emissions at $t = 0$ (time of manure application) and removed all earlier measurements so that the first time interval begins at $t = 0$. In such cases, cumulative emissions at $t = 0$ were subtracted from all subsequent values, assuming that emissions due to the manure only begin after its application.

For trials with more than 50 time points, we kept only a subset of points closest to a reference time sequence, defined with a time step of 2 hours between $0 - 12$ h, 4 hours between $12 - 24$ h, and 8 hours between $24 - 178$ h. Missing values for manure pH were imputed using the average pH per manure origin. Missing values for air temperature and wind speed were linearly interpolated, while missing rainfall values were replaced with zeros.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

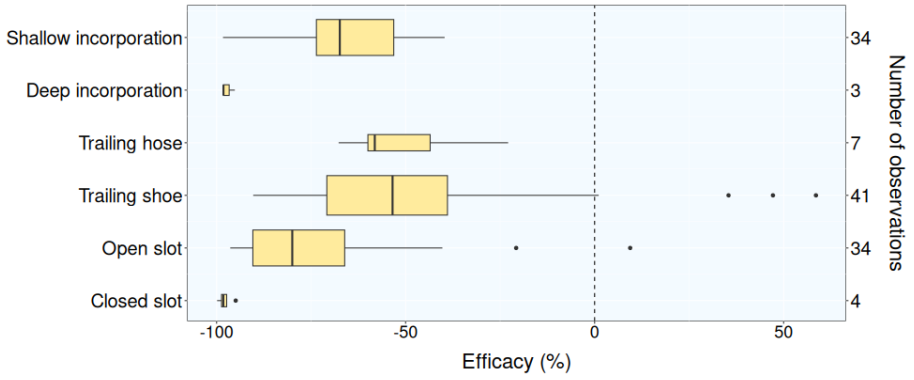

Figure 2. Percentage reduction in ammonia emissions for different application techniques compared to broadcast application, calculated directly from the database. Only experiments including at least one broadcast trial were considered. The number of trials included in the calculations for each application method is indicated on the right y-axis.

2.2 Models

We implemented two types of models: static models and dynamic models. The static models aim to predict the final cumulative emissions following fertilization and are trained only on the final cumulative values. In contrast, the dynamic models are designed to predict the entire emission trajectory from the time of fertilizer application. Importantly, our objective was not to formulate a step-ahead prediction problem (i.e., estimating emissions at time $t+1$ given emissions at time t), but rather to predict the complete temporal dynamics. Both static and dynamic models were developed using random forests and different types of neural networks.

2.2.1 Static models

Transformation of the data was necessary for the static models and was carried out in the same manner as described in Favrot et al. (2026). For each trial, the cumulative ammonia emission at the end of the last interval was used as the response variable. Intermediate observations were not retained for model training. Meteorological variables were summarized over six predefined time windows. For each of the three variables (air temperature, wind speed, and rainfall rate), six new variables were created: five corresponding to average values over regular intervals in the $[0, 20h]$ period (each lasting 4 hours), and one covering the interval from 20 hours to the end of the trial.

Random forest

Random Forest (Breiman, 2001) is an ensemble learning method that builds multiple decision trees and combines their predictions, reducing overfitting and improving accuracy through averaging. Following Favrot et al. (2026), we trained a random forest with 500 trees using final cumulative emission as the response variable. The hyperparameters mtry and nodesize, which control the diversity between trees and their size respectively, were optimized using validation sets. Computations were per-

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

formed in R using the 'randomForest' package (Liaw and Wiener, 2002).

Neural networks

For the static models, we used multilayer perceptron (MLP) neural networks (Rumelhart et al., 1986). An MLP consists of an input layer, one or more hidden layers with non-linear activation functions, and an output layer. It is a feedforward architecture in which information flows in one direction, from inputs to outputs. We defined 12 architectures differing in the number and size of hidden layers, and in whether or not an embedding layer was used for the categorical variables (manure source, incorporation method, and application method). Embedding layers are dense vector representations that map each category to a trainable low-dimensional continuous space, allowing the model to capture similarities between categories. When embedding layers were not used, categorical variables were transformed using one-hot encoding. ReLU functions were used as activation functions between layers. A detailed description of the architectures is provided in Table A5 (Appendix). Models were implemented in PyTorch and optimized using the Adam algorithm. The learning rate was set to 0.0001, and training was conducted for 12000 epochs using minibatches of size 32. A validation set was used to select the model parameters corresponding to the number of epochs minimizing the validation loss.

2.2.2 Dynamic models

The construction of the dynamic models was based on three main components: the model time step (regular or irregular), the type of algorithm used (random forest or neural networks), and the nature of the training data (measured, interpolated, or augmented, see below). In the following, y_t denotes the cumulative emission at time t, let $t_1 < t_2 < \cdots < t_F$ define a sequence of time points ($t_1 = 0$), and let $dt_k = t_{k+1} - t_k$ denote the duration between consecutive time points t_k and t_{k+1}.

Timestep

Regarding the time step, we considered two types of models: models with regular time steps and models with irregular time steps. In the case of regular time step models, the model provides predictions at regular intervals of length dt. In our study, we considered the following values for dt: 2, 4, 6, 8, or 10 hours. For irregular time step models, predictions can be made at any arbitrary sequence of time points $t_1 < t_2 < ... < t_F$.

Algorithm

We considered two types of algorithms: random forests and recurrent neural networks.

In the case of random forests, the cumulative emission at time t_k is explicitly used to predict the cumulative emission at time t_{k+1}. The model can be expressed as follows:

$$y_{t_{k+1}} = f\left(y_{t_k},\ t_{k+1},\ dt_k,\ \text{meteo}([t_k, t_{k+1}]),\ \text{plot level variables}\right)$$

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

with f being a random forest model and $y(0) = 0$. When the model is used for prediction, the cumulative emission at time t_k is unknown and is therefore replaced by its predicted value. The model is thus applied recursively to generate predictions over the entire time sequence $t_1 < t_2 < ... < t_F$. The response variable is the cumulative ammonia emissions and predictors are the cumulative emissions at the beginning of the time interval (y_{t_k}), the number of hours after manure application (t_{k+1}), the length of interval measurement (dt_k), the meteorological variables and the plot level variables. As with the static random forest model, the mtry hyperparameter was optimized using validation sets.

Recurrent Neural Networks (RNNs) are a class of artificial neural networks specifically designed to handle sequential data by incorporating temporal dependencies through loops within their architecture. Unlike feedforward networks, RNNs maintain a memory of past inputs via hidden states that are updated at each time step, making them well-suited for time series prediction tasks (Elman, 1990). We tested 24 RNNs architectures that differed in terms of the response variable (cumulative emissions, interval emissions, or both), network type (unidirectional or bidirectional), presence or absence of an additional dense layer with six neurons between the main recurrent layer and the output, and use of embedding layers for categorical variables (yes/no). When embedding layers were not used, categorical variables were transformed using one-hot encoding. A full list of all model architectures is provided in Table A6 in the Appendix. Figure 3 illustrates the architecture of model 'rnn 9': a bidirectional recurrent neural network with embedding layers for categorical variables and an additional dense layer between the recurrent layer and the output. The model predictors include the number of hours after manure application, the length of the measurement interval (dt_k), meteorological variables, and plot-level variables. The size of the recurrent hidden layer was optimized using validation sets. All computations were performed using PyTorch, and model training was carried out using the Adam optimization algorithm. The learning rate was set to 0.0001, and training was conducted for 2000 epochs using minibatches of size 32. A validation set was used to select the model parameters corresponding to the number of epochs minimizing the validation loss.

Data used for training

Several variations of the original dataset were considered for model training, depending on the model type. Regular time step models were trained using an interpolated version of the original dataset. For example, for a model with a regular time step of two, all dynamic variables (cumulative emissions, weather conditions) were linearly interpolated at a time step of two. To train irregular time step models, we used either the original dataset or an augmented version of it. The augmented dataset was constructed by merging six datasets: the original dataset and five interpolated versions generated for time steps of 2, 4, 6, 8, and 10 hours. Data augmentation, originally introduced in computer vision to reduce overfitting by artificially increasing the size of the training set (Krizhevsky et al., 2012), aims to improve model generalization by reducing the gap between the training and test data distributions (Shorten and Khoshgoftaar, 2019). It has also been applied in time series contexts such as classification, anomaly detection, and forecasting (Wen et al., 2020). A list of the datasets defined here with corresponding number of observations is given in Table A4 (Appendix).

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

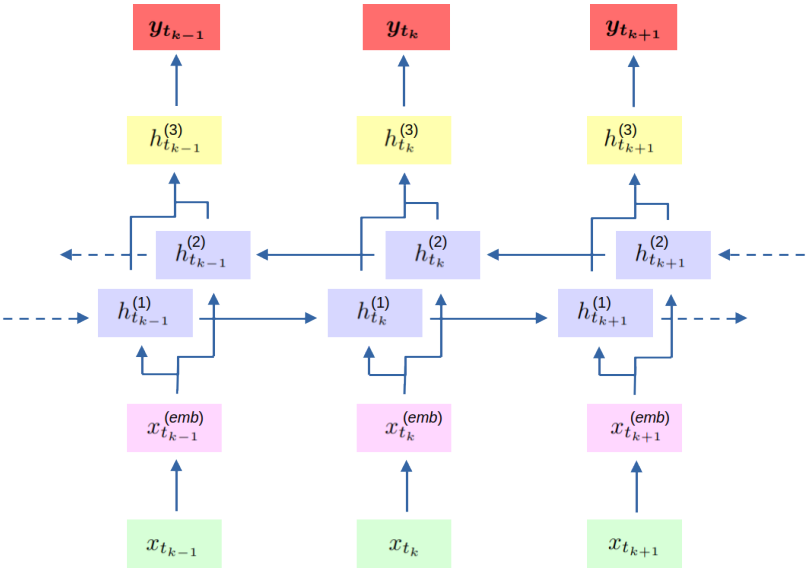

Figure 3. Architecture of the 'rnn 9' model. Bidirectional recurrent neural network with embeddings for categorical variables and an additional dense layer between the main recurrent layers and the output. Green indicates the input, pink the embedding layer, purple the main recurrent layer, yellow the additional dense layer between the recurrent layer and the output, and red the output.

For obvious computational reasons, it was not feasible to test all model combinations resulting from the three components described above. Instead, a representative subset of 33 models was developed, as summarized in Table A7 (Appendix).

2.2.3 Cross validation

To compare model performances and to optimize model hyperparameters, we employed a cross-validation procedure. Specifically, we generated 15 random partitions of the 699 available trials, while preserving the original proportion of trials across research institutes. Each partition was divided into three subsets: training, evaluation, and test, containing 450, 75, and 74 trials, respectively. These 15 partitions were then applied to the datasets described in Table A4 (Appendix), to produce the training, evaluation, and test sets used for the training, hyperparameter optimization, and performance assessment of the different models considered. In the following, we denote this collection of datasets as $(\text{train}_i,\ \text{eval}_i,\ \text{test}_i)_{i=1,\ldots,15}$.

Hyperparameter optimization

For the static random forest model, the optimal values of the hyperparameters mtry and nodesize (ns), selected from the set $\{8,9,\ldots,22\} \times \{2,3,4\}$, were those that minimized the mean absolute error (MAE) across all validation sets:

$$(\text{mtry}^*, \text{ns}^*) = \underset{x,y}{\operatorname{argmin}} \sum_{i=1}^{15} \text{MAE}\ (\text{mtry} = x,\ \text{ns} = y;\ \text{dataset} = \text{eval}_i)$$

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

The models were then retrained on the concatenated training and evaluation sets for each cross-validation partition before final evaluation on the corresponding test set.

For each of the static neural network models, the validation sets were used to optimize the number of training epochs:

$$n^*_{\text{epoch},\,i} = \underset{x \leq 12000}{\text{argmin}} \text{MAE}\ (n_{\text{epoch}} = x,\ \text{dataset} = \text{eval}_i)$$

For each of the dynamic random forest models, the optimal values for the mtry hyperparameter, selected from $\{2,3,4,5\}$, were those that minimized the MAE over all validation sets:

$$\text{mtry}^* = \underset{x}{\text{argmin}} \sum_{i=1}^{15} \text{MAE}\ (\text{mtry} = x;\ \text{dataset} = \text{eval}_i)$$

The models were then retrained on the concatenated training and evaluation sets for each cross-validation partition before final evaluation on the corresponding test set.

For the recurrent neural networks, the size of the recurrent hidden layer was optimized by evaluating the following values: $\{16, 64, 128, 256, 512\}$. The optimization was performed only using 'rnn 7' and 'rnn 8' irregular time step models: unidirectional recurrent neural networks with cumulative emissions as the response variable, without an additional dense layer before the output, and without embeddings ('rnn 7') or with embeddings ('rnn 8'). The optimal hidden size found for 'rnn 7' was then used for all models without embeddings, and the one found for 'rnn 8' was applied to all models using embeddings. The optimal sizes were determined by minimizing the mean absolute error across all evaluation sets:

$$h^*(\text{without embeddings}) = \underset{h}{\text{argmin}} \sum_{i=1}^{15} \min_{x \leq 2000} \text{MAE}\ (\text{hidden size} = h,\ n_{\text{epoch}} = x,\ \text{model} = \text{'rnn 7'};\ \text{dataset} = \text{eval}_i)$$

$$h^*(\text{with embeddings}) = \underset{h}{\text{argmin}} \sum_{i=1}^{15} \min_{x \leq 2000} \text{MAE}\ (\text{hidden size} = h,\ n_{\text{epoch}} = x,\ \text{model} = \text{'rnn 8'};\ \text{dataset} = \text{eval}_i)$$

Once the optimal hidden layer size was determined, all models were trained using the training sets, and the validation sets were used to determine the optimal number of epochs, following the same procedure as for the static neural networks.

The optimized hyperparameter values for the static random forests, dynamic random forests, and recurrent neural networks are provided in Table A3 (Appendix). Figure A1 (Appendix) presents the mean absolute error as a function of the hidden layer size for models 'rnn 7' (without embeddings) and 'rnn 8' (with embeddings).

Final testing

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

To perform the final testing, we computed the mean absolute error on the final cumulative emissions for all models across all test sets. For dynamic models, we also computed the mean absolute error over all time points of the emission dynamics. In order to obtain predictions at the observed time points with regular time step dynamic models, a post-processing step using linear interpolation was applied on the model outputs. For irregular time step dynamic models, predictions were directly computed at the observed time points, except for the 'rnn 9 – data a.' model, for which the MAE was also computed using predictions generated at regular time intervals (dt = 2, 4, 6, 8, and 10). In that case, as for the regular time step models, linear interpolation was applied on the model outputs in order to obtain predictions at the observed time points. As a benchmark, the ALFAM2 model was applied to the 15 test sets, although it was originally trained on a larger dataset comprising 722 trials, including the 699 trials used in our study (Hafner et al., 2025).

2.3 Sensitivity to time step

When making predictions with an irregular time step model, the time points at which emissions are predicted can be freely chosen. This raises the question of how sensitive the predictions are to the choice of time sequence. To address this, we used the 'rnn 9' and 'rnn 9 – data a.' models to predict emission dynamics for all trials in the 15 test sets, using both the observed time sequences and a regular time sequence with a 2-hour interval. In the latter case, meteorological data were linearly interpolated to match the 2-hour time steps, and a post-processing interpolation step was applied to obtain predictions at the observed time points. We then computed the absolute difference at the final observed time points between predictions generated with the two time sequences. Examples of such predictions are shown in Fig. A7 (Appendix), and the boxplots of the absolute differences are presented in Fig. A8 (Appendix).

2.4 Predictions on scenarios

We considered several recognized fertilizer application techniques able to reduce ammonia emissions, namely: broadcast fertilizer application followed with immediate incorporation, trailing hose, trailing shoe, open slot, and closed slot. These techniques were previously found to reduce emissions compared to the standard fertilizer application technique, broadcast without incorporation (Webb et al., 2010). We used the models ALFAM2, 'static rf', 'dynamic rf 2', 'static nn 12', and 'rnn 9 - data a.' to evaluate their efficacy by simulating ammonia emissions for the 128 scenarios previously defined in Favrot et al. (2026). These scenarios correspond to all possible combinations of two contrasting values for each model variable, except for pH, which was fixed at 7.5 (Table 1). For each scenario and application method, cumulative ammonia emissions were predicted using models trained on the full dataset. We then computed the percentage reduction in emissions relative to the broadcast method, 72 hours after fertilization. The ALFAM2 model was used with a time step of 2, and the 'rnn 9 - data a.' model was applied with different time steps (dt = 2, 4, 6, 8 and 10). Values of meteorological variables provided in Table 1 for the static models were converted to mean values over the time intervals corresponding to the different dynamic models.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Table 1. Values of variables used to define the full factorial design considered to compare manure application techniques. For each of the meteorological variables, two lists of six values were defined for the six time periods considered in the static models. These values correspond to the first and third quartiles calculated over the dataset used in Favrot et al. (2026), except for rainfall, where the first quartile and the 0.95 quantile were used in order to obtain more contrasted conditions. Each value was combined with all the others, leading to 128 scenarios.

Variable	Values
pH	7.5
TAN application rate (kg/ha)	36.7, 80.3
application rate (t/ha or m^3/ha)	18.7, 36.7
slurry dry matter (%)	3.8, 8.2
slurry source	cattle, pig
air temperature (°C)	(11.2, 10.8, 8.7, 8.0, 8.1, 9.2), (18.2, 18.5, 16.2, 14.2, 14.0, 16.0)
wind speed (m/s)	(2.3, 2.3, 1.8, 1.4, 1.4, 2.2), (5.0, 4.6, 3.8, 3.7, 3.7, 4.1)
rainfall rate (mm/h)	(0, 0, 0, 0, 0, 0, 0), (0.3, 0.2, 0.3, 0.4, 0.3, 0.1)

3 Results

3.1 Model comparison

The results for models 'rnn 1' to 'rnn 24' are presented separately in Fig. A4 and A5 (Appendix). The best-performing model was 'rnn 9', which is bidirectional, includes embedding layers for categorical variables, incorporates an additional dense layer between the recurrent layer and the output, and uses interval emissions as the response variable. It achieved an average MAE of 3.64 kgN/ha on all the time points of the emission dynamics. Results by architectural features (Fig. A5, Appendix) show that using embedding layers for categorical variables consistently improves model performance. The bidirectional structure is particularly beneficial when interval emissions are used as the response variable. Adding an extra dense layer between the recurrent layer and the output has a more moderate effect overall, except in the case of unidirectional models trained on interval emissions, where this addition significantly reduces the mean absolute error.

Figure 4 presents boxplots of the mean absolute error on final cumulative emissions, calculated for each test set, across different modeling approaches: static random forest, static neural networks, dynamic random forest, the three recurrent neural network models based on the 'rnn 9' architecture, and the ALFAM2 model. Among the static models, 'static rf' and 'static nn 12' - the latter being the most complex static neural network - yield similar performances, with average MAEs of 5.55 kgN/ha and 5.28 kgN/ha, respectively. Other static neural network models exhibit lower performance, with performance decreasing as model complexity is reduced. Dynamic random forest models produced relatively consistent and better results than the static models, despite the presence of an outlier with an error close to 8 kgN/ha across all these models. Within this group, the best-performing model was 'dynamic rf - 10', achieving an average MAE of 4.93 kgN/ha, corresponding to reductions of 11.1%

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

and 6.6% compared to 'static rf' and 'static nn 12', respectively. The three RNN models based on the 'rnn 9' architecture - the most complex among those predicting interval emissions rather than cumulative emissions - yield contrasting results. The irregular time-step model trained on the original dataset ('rnn 9') achieves an average MAE of 4.84 kgN/ha, while the version trained on the augmented dataset ('rnn 9 - data a.') reaches an average MAE of 4.56 kgN/ha, making it the best-performing model among all tested. Notably, 'rnn 9 – data a.' outperforms the other models on the majority of the test sets (Fig. A2 in the Appendix). The average MAE obtained with this model corresponds to reductions of 17.8% and 13.6% compared to 'static rf' and 'static nn 12', respectively. Additionally, 'rnn 9 - data a.' displays a narrower error distribution than 'rnn 9'. In contrast, the 'rnn 9 - 2' model with a fixed time-step of 2 performs worse, with an average MAE of 5.67 kgN/ha and a more dispersed error distribution. For comparison, the ALFAM2 model achieves an average MAE of 5.61 kgN/ha, corresponding to an increase of 1.1% and 6.2% compared to the static models 'static rf' and 'static nn 12', respectively, and a 23.0% increase relative to 'rnn 9 - data a.'. Figure 5 complements these findings by plotting predicted versus observed final cumulative emissions for ALFAM2, 'rnn 9 - data a.', 'static nn 12', and 'static rf' models. A noticeable bias is observed for high-emission trials in ALFAM2 and 'static rf', the latter systematically underestimating large emission values. In contrast, 'rnn 9 - data a.' appears visually unbiased and exhibits the highest predictive accuracy among the models compared.

Model evaluation based on all time points of the emission dynamics (Fig. 6) revealed results broadly consistent with those obtained for the final cumulative emissions. The dynamic random forest models showed similar performance overall, although a slight degradation was observed as the time step increased for models with regular time intervals. The best-performing among them was the 'dynamic rf – 2' (time step of 2) with a mean absolute error of 4.05 kgN/ha. The three RNNs based on the 'rnn 9' architecture exhibited the same pattern of contrasting performance observed for final cumulative predictions. Among them, the 'rnn 9 – data a.' model outperformed all others, achieving a mean MAE of 3.54 kgN/ha, representing a 12.6% reduction compared to 'dynamic rf – 2', and a 22.5% reduction relative to the ALFAM2 model, whose average MAE was 4.57 kgN/ha. As with the final cumulative emissions, 'rnn 9 – data a.' also outperformed the other models on the majority of the test sets (Fig. A3 in the Appendix). Figure 7 presents illustrative examples of predicted emission dynamics for four trials from the first test set, obtained using the ALFAM2, 'dynamic rf – 2', and 'rnn 9 – data a.' models. These examples highlight a variety of predictive behaviors. In trials 1527 and 3168, one of the models shows noticeably lower predictive accuracy than the other two, whereas all three models exhibit good agreement with the observed values in trial 1132. Conversely, trial 2828 illustrates a case where all models substantially deviate from the measured emissions.

3.2 Simulated efficacy of different manure application techniques

The predicted efficacy values obtained over all 128 scenarios defined by the full combination of factors presented in Table 1 are shown in Fig. 8. For the 'rnn 9 – data a.' model, results were obtained using a time step of 6, as this configuration provided the best performance for new predictions (Fig. A6 in Appendix). The predicted efficacies vary across models, both in terms of medians and associated variability. Regarding the ranking of manure application techniques based on median efficacy, the ALFAM2 model, as well as the 'static nn 12' and 'rnn 9 - data a.' models, yield consistent rankings for incorporation, open

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

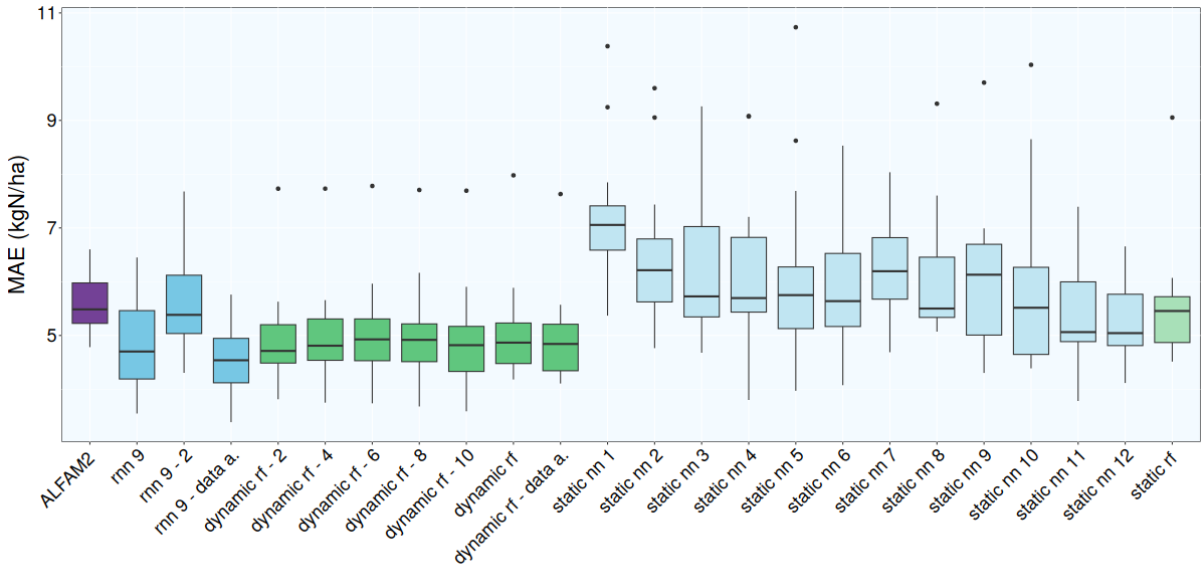

Figure 4. Mean absolute error on the final cumulative emission for the 15 test sets of the cross validation procedure, grouped by model. Colors indicate model types: blue for neural networks and green for random forests. Bright colors represent dynamic models, while pale colors represent static ones. For recurrent neural networks, only models derived from architecture 'rnn 9' are shown. See Tables A6 and A7 for model details.

slot, and closed slot methods, with relatively similar median values. For instance, the closed slot method is ranked as the most effective for all three models, with median efficacy of 91.4%, 88.6%, and 88.7% for ALFAM2, 'static nn 12', and 'rnn 9 - data a.', respectively. For these same models, trailing hose and trailing shoe exhibit comparable performance, with trailing shoe being slightly more effective than trailing hose for ALFAM2 and 'static nn 12', whereas the reverse is observed for 'rnn 9 - data a.'.

Both neural network models exhibit greater variability than ALFAM2. In some scenarios, they even predict positive efficacies, particularly with the 'static nn 12' model for the incorporation method. However, these positive efficacies are associated with low absolute emission values (below 10 kgN/ha in most cases), as shown in Fig. A11 of the Appendix.

Turning to the random forest models, it is immediately evident that their predicted median efficacy values are generally lower than those of the other models. Notably, the closed slot method is predicted to be less effective than the open slot method. Additionally, the incorporation method appears significantly less effective than the other techniques according to the 'static rf' model, while it is predicted to be more effective than trailing hose and trailing shoe in the 'dynamic rf - 2' model. In terms of variability, the dynamic random forest model ('dynamic rf - 2') displays greater variability than its static counterpart ('static rf'), whereas the opposite trend was observed for the neural networks, with the dynamic 'rnn 9 - data a.' model showing less variability than the 'static nn 12' model. Positive efficacy values were also observed for the 'dynamic rf - 2' model in some

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

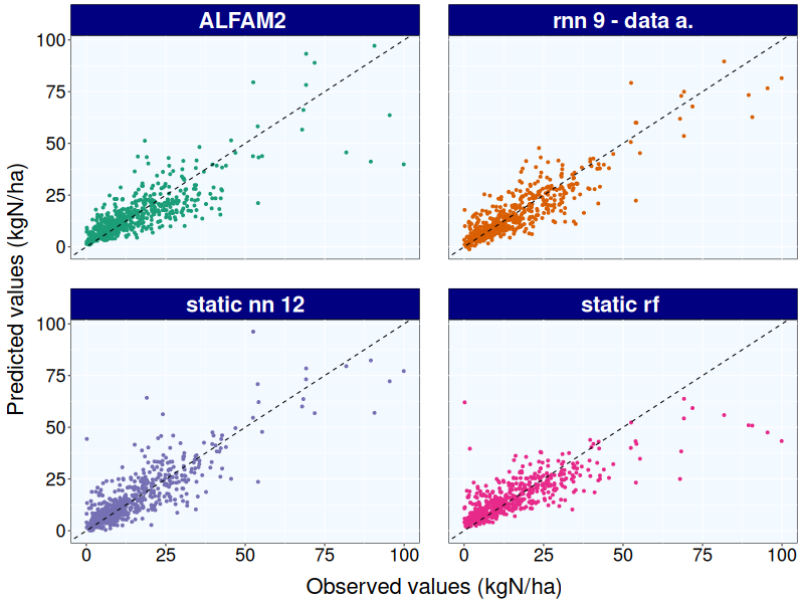

Figure 5. Observed versus predicted values of final cumulative ammonia emissions obtained from two static models ('static nn 12' and 'static rf') and two dynamic models (ALFAM2 and 'rnn 9 – data a.'), across the 15 test sets. When a trial appears in multiple test sets, the corresponding prediction represents the average across those sets. The dashed line indicates the identity line ($y = x$).

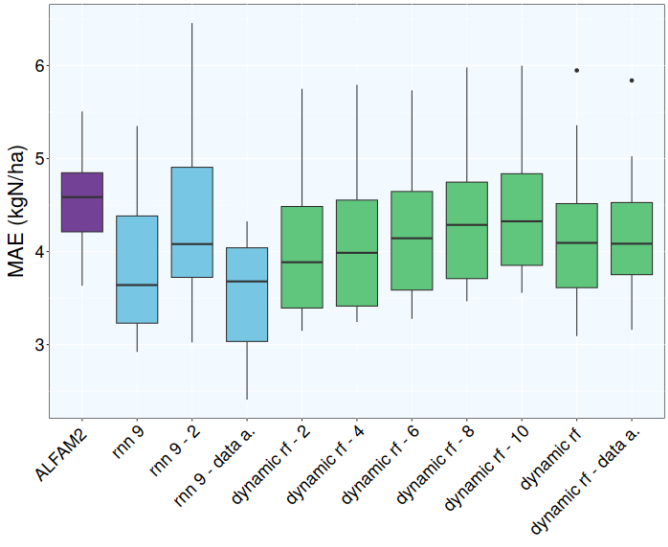

Figure 6. Mean absolute error computed over all time points of the emission dynamics for the 15 test sets of the cross validation procedure, grouped by model. Colors indicate model types: blue for neural networks and green for random forests. For recurrent neural networks, only models derived from architecture 'rnn 9' are shown. See Tables A6 and A7 for model details.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

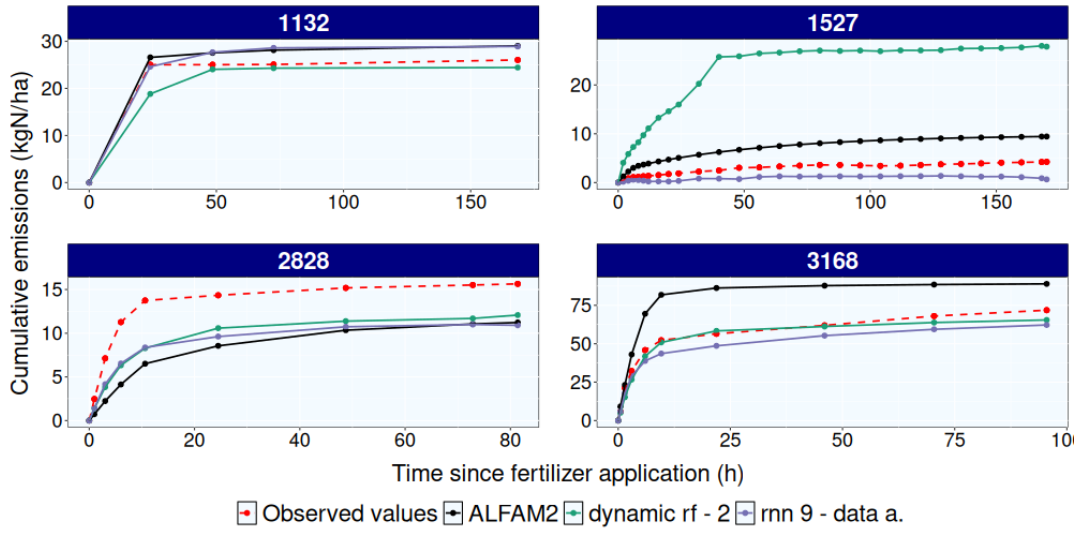

Figure 7. Example of emission dynamics predictions from the first test set. Each subplot corresponds to a trial identified by its pmid. Red representing the observed values and other colors indicate the output of different models.

cases, though to a lesser extent than 'static nn 12' and 'rnn 9 - data a.', and again primarily for scenarios involving low absolute emission values (Fig. A11 in Appendix).

The results obtained with other time steps for the 'rnn 9 – data a.' model (Fig. A10 in Appendix) indicate a relatively limited sensitivity to the time step used for predictions. In particular, the ranking of application methods based on median efficacy remains generally consistent across all time steps, except for $dt = 10$, where the incorporation method appears slightly more effective than the open slot method, whereas the opposite is observed for predictions obtained with the other time steps.

4 Discussion

To our knowledge, this study is the first to investigate the use of dynamic machine learning models to predict ammonia emissions following field fertilization, and to demonstrate the value of incorporating emission dynamics for improving the prediction of final cumulative emissions compared to static models. Specifically, we compared 33 dynamic models based on random forests and recurrent neural networks algorithms with 13 static models using random forests and multilayer perceptron neural networks. It is worth noting that we did not use long short-term memory networks because our preliminary tests indicated that LSTMs did not provide any performance improvement over simple recurrent neural networks, likely due to the relative regularity of the emission time series (Fig. 1) and the limited number of observations per series (median = 8 observations per series). Additionally, we tested data renormalization approaches, but these did not improve model performance (Fig. A9 in Appendix). Based on a cross-validation framework, our results show that the best-performing dynamic models consistently

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

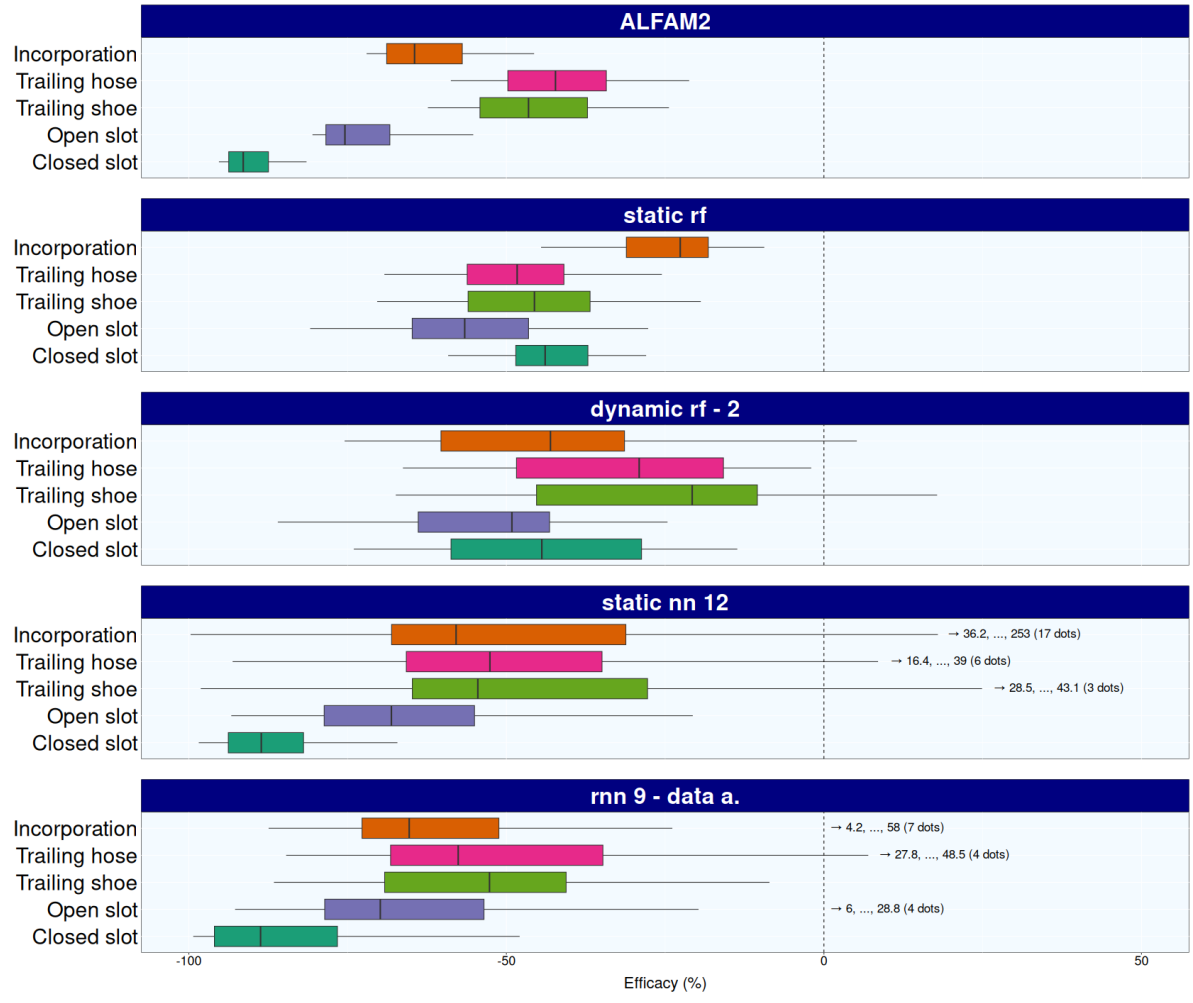

Figure 8. Percentage reduction in emissions compared to the broadcast method (efficacy) for the application techniques: incorporation, trailing hose, trailing shoe, open slot, and closed slot. Each subplot corresponds to a different model, and each boxplot represents the variability across the 128 scenarios defined by all possible combinations of the factors given in Table 1.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

outperform their static counterparts in predicting final cumulative emissions. The best recurrent neural network model ('rnn 9 – data a.') achieved a 13.6% lower average MAE than the best static neural network ('static nn 12'), while the best dynamic random forest reduced the average MAE by 11.2% compared to the static random forest model. This superiority of dynamic over static models likely stems from a better ability to capture interactions between meteorological conditions and the volatilization process, as well as from the larger effective training set available to dynamic models through the use of sequential data.

The best-performing model among all those tested was the 'rnn 9 – data a.' model, which achieved an average MAE of 4.56 kgN/ha for final cumulative emissions. By comparison, the standard deviation of measurement errors has been estimated to be at most 31% of the absolute emissions (Kamp et al., 2024). Assuming normally distributed errors, this corresponds to an expected MAE of approximately 3.85 kgN/ha for the dataset used in this study, given an observed mean final cumulative emission of 15.6 kgN/ha. Without data augmentation, the 'rnn 9' model was the most accurate among all tested RNNs. Among the models using interval emissions as response variable, its architecture was the most complex: categorical variables were embedded, the recurrent layer was bidirectional, and an additional dense layer was added before the output. With data augmentation, prediction error decreased by 5.8% for the final cumulative emissions and by 7.6% for the full emission dynamics, demonstrating the effectiveness of this approach. Another benefit of data augmentation observed with the 'rnn 9 – data a.' model was reduced sensitivity to the choice of time step, as shown in Fig. A7 and A8 in the Appendix.

For both static and dynamic models, neural networks outperformed random forests. The mean absolute error was reduced by 4.9% for 'static nn 12' compared to 'static rf', and by 8.11% for 'rnn 9 – data a.' compared to 'dynamic rf – 10'. One key difference between these two model types is their ability to extrapolate. Random forests, which rely on averaging observed target values within the terminal nodes, cannot extrapolate beyond the training domain. In contrast, the final linear layer in neural networks does not inherently constrain the prediction range. This likely explains the results shown in Fig. A11 (Appendix), where the predicted range of emissions for the broadcast method is much narrower with both static and dynamic random forests than with neural networks or the ALFAM2 model.

When predicting final cumulative emissions up to a maximum of 72 hours after manure application, Favrot et al. (2026) reported a mean absolute error of 3.3 kgN/ha and an R^2 of 0.88 using a static random forest model on an independent test subset. These values are substantially better than those obtained in our study for the static random forest model, which yielded a MAE of 5.55 kgN/ha and an R^2 of 0.63 (Table A8 in Appendix). Even our best-performing model, 'rnn 9 – data a.', produced a MAE above 3.3 kgN/ha. Two main factors may explain this difference: (i) Favrot et al. (2026) evaluated their model on a single test set, and (ii) all trials used in that previous study were limited to a maximum duration of 72 hours, whereas in our dataset, the majority of trials exceed 90 hours, with some extending up to 176 hours. Using a random forest model predicting final relative emissions, Xu et al. (2024) obtained an R^2 of 0.79 on a test set. In comparison, our static random forest model achieved an R^2 of 0.61 for the same response variable, while the 'rnn 9 – data a.' model reached an R^2 of 0.75. However, the datasets used in the two studies differ significantly. The dataset employed by Xu et al. (2024) comprises 2775 trials - com-

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

pared to 699 in our study - and includes both organic and mineral fertilizers. Their dataset also contains trials with multiple fertilizer applications, and the set of input variables differs from ours. For instance, their model incorporates soil properties such as carbon content and bulk density. These differences likely account for the discrepancies in model performance observed between the two studies. Finally, Hafner et al. (2025) reports a MAE of 0.093 and an R^2 of 0.66 for final relative emissions using parameter set 3 of the ALFAM2 model, evaluated on the calibration dataset (comprising 722 trials, including the 699 used in our study). By comparison, the 'rnn 9 – data a.' model achieved a lower MAE of 0.077 and a higher R^2 of 0.75 for the same response variable, but on independent test sets.

The results obtained for the prediction of the efficacy of mitigation techniques across different scenarios reveal non-negligible variability depending on the model used. The outcomes for both static and dynamic random forest models are somewhat unexpected, with the closed slot method predicted to be less effective on average than the open slot method. This is likely due to the limited number of trials available for the closed slot method in the dataset ($n = 12$ trials), combined with the known inability of random forests to extrapolate - particularly illustrated by the narrow range of predicted emissions for the broadcast method compared to other models (Fig. A11 in Appendix). When compared to the empirical efficacy values derived directly from the ALFAM2 database (Fig. 2), the predictions of the 'rnn 9 – data a.' model show good agreement. Notably, positive efficacy values are also observed empirically for the trailing shoe and open slot methods. Finally, compared to the results reported by Favrot et al. (2026), which were obtained on identical scenarios but using a smaller dataset for model training ($n = 538$ trials), the predicted efficacy values obtained here are generally similar for the static random forest, whereas ALFAM2 tends to estimate slightly higher mitigation effects.

Despite a few cases of predicted negative efficacy - mostly observed for scenarios with very low absolute emissions (Fig. A11 in Appendix) - the 'rnn 9 – data a.' model produces average efficacy values consistent with those reported in the literature, even though different datasets were used for model calibration. For instance, average predicted efficacies are 48.1% for trailing hose and 53.8% for trailing shoe, closely matching those reported by Häni et al. (2016), who found 51% and 53% efficacy for these two techniques, respectively, for pig and cattle slurries also. The open slot and closed slot techniques are estimated to reduce emissions by $50 - 70\%$ and $70 - 90\%$, respectively, according to European Environment Agency (2023). These ranges are consistent with the average predicted efficacies of 62.3% and 81.9% obtained with the 'rnn 9 – data a.' model. The relatively high variability in efficacy predicted across scenarios is also in line with previous studies. For example, Webb et al. (2010) reported variability ranging from 0 to 75% for trailing hose and from 23 to 99% for open slot, while European Environment Agency (2023) indicates a range of 20 to 90% for incorporation. The average efficacy predicted by the 'rnn 9 – data a.' model for incorporation is 58.3%, which falls within this reported range.

Although our study demonstrated the strong predictive capabilities of models such as recurrent neural networks for forecasting post-fertilization ammonia emissions, its scope remains limited to a narrow range of fertilizers, specifically pig and cattle slurries. Future work could extend these findings to other types of fertilizers by leveraging techniques such as transfer learning

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

(Pan and Yang, 2009), which would allow the knowledge acquired by recurrent neural networks to be transferred to new fertilizers using external datasets, even when only a limited number of trials are available. Model performance could also potentially be improved by making broader use of the ALFAM2 database, for example by including trials based on measurement methods other than micrometeorological techniques, and by incorporating the measurement method itself as a predictor, as was done by Søgaard et al. (2002) for the ALFAM model. Another direction could be to predict instantaneous fluxes rather than cumulative or interval emissions, or to explore long short-term memory neural networks architectures (Hochreiter and Schmidhuber, 1997) as alternatives to standard RNNs. Finally, generating prediction intervals using approaches such as conformal prediction (Vovk et al., 2005) would add substantial value, particularly in view of potential practical applications such as compiling national emission inventories.

5 Conclusions

Our results demonstrate that predictions of cumulative ammonia emissions after manure application could be improved by using dynamic machine learning models instead of static models. The best-performing model is a bidirectional recurrent neural network with embeddings for categorical variables and an additional dense layer between the recurrent layer and the output. This model is able to handle datasets with heterogeneous time steps, and achieves a mean absolute error of 4.56 kgN/ha across all test sets in the cross-validation, corresponding to a 13.6% reduction in error compared to the best static model. Notably, this MAE is close to the MAE value expected when considering only measurement errors (3.85 kgN/ha), revealing that it would be difficult to improve prediction accuracy further based on this type of data. Moreover, results shows that the simulated efficacy of the different slurry application methods was in good agreement with the literature. Given their predictive accuracy and ease of implementation, such models represent promising tools for integration into national emission inventories and could also be embedded in precision farming systems to improve nitrogen use efficiency.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Table A1. Summary of quantitative variables.

Variable	Min	1st quartile	Median	Mean	3rd quartile	Max
cumulative emission [kg/ha]	- 0.45	2.56	6.75	10.63	14.60	125.14
air temperature [°C]	-3.15	9.90	13.70	13.89	17.70	35.24
wind speed [m/s]	0.06	1.78	2.75	3.11	4.00	28.37
rainfall rate [mm/h]	0.00	0.00	0.00	0.02	0.00	3.48
trial duration [h]	24.00	93.70	96.55	113.41	162.75	176.27
application rate [t/ha]	6.60	19.74	28.21	29.38	35.00	132.60
TAN application rate [kg/ha]	10.90	36.66	54.28	62.07	80.30	235.40
manure dry matter concentration [%]	0.77	4.00	6.59	6.25	8.15	13.80
manure pH	4.30	7.10	7.40	7.38	7.72	8.90
incorporation time [h]	0.00	0.00	0.00	1.76	0.00	24

Table A2. Categories and number of observations per category for the categorical variables.

Variable	Levels	Number of observations
Application method	broadcast	213
	trailing hoses	197
	trailing shoes	158
	open slot	119
	closed slot	12
Incorporation	none	643
	shallow	52
	deep	4
Manure source	cattle	485
	pig	214

Appendix A

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

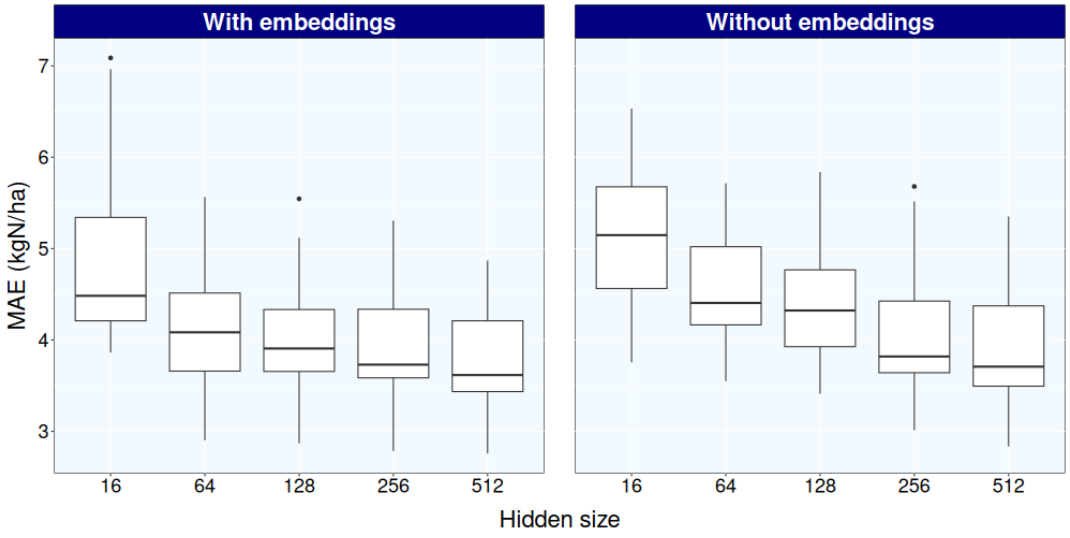

Figure A1. Impact of hidden size on model performance for RNN models. Mean absolute error as a function of hidden size for unidirectional models predicting e.cum, without an additional layer between the recurrent and output layers, with embeddings ('rnn 7') and without embeddings ('rnn 8'). Each boxplot represents variability across the 15 test sets.

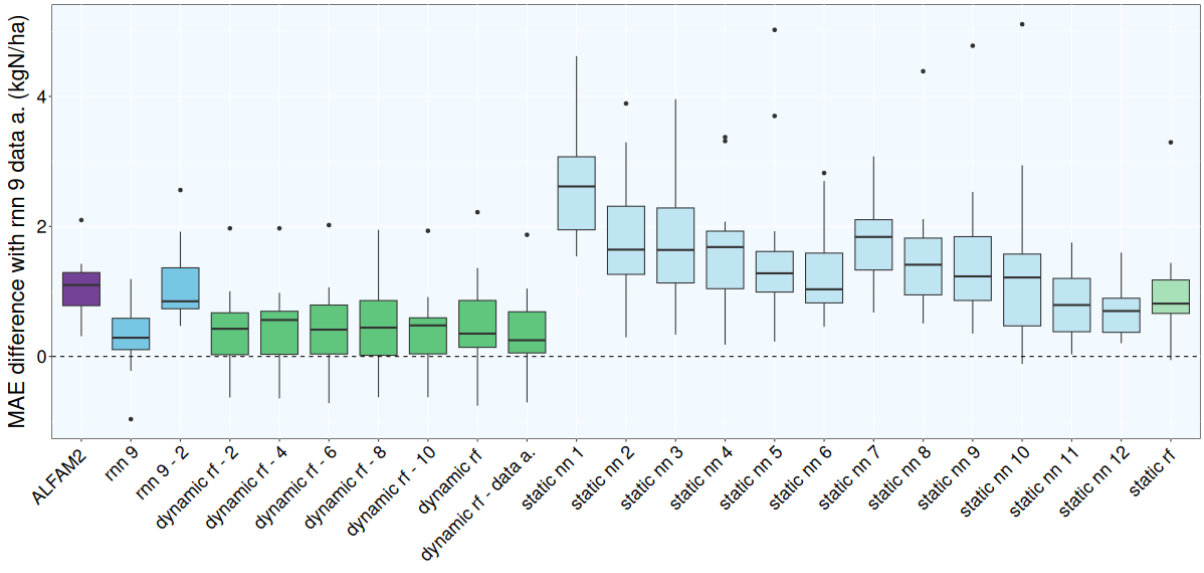

Figure A2. Difference in mean absolute error on final cumulative emissions relative to the 'rnn 9 – data a.' model, across the 15 cross-validation test sets. Colors indicate model types: blue for neural networks and green for random forests. Bright colors represent dynamic models, while pale colors represent static ones. For recurrent neural networks, only models derived from architecture rnn 9 are shown. See Tables 6 and 7 for model details.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

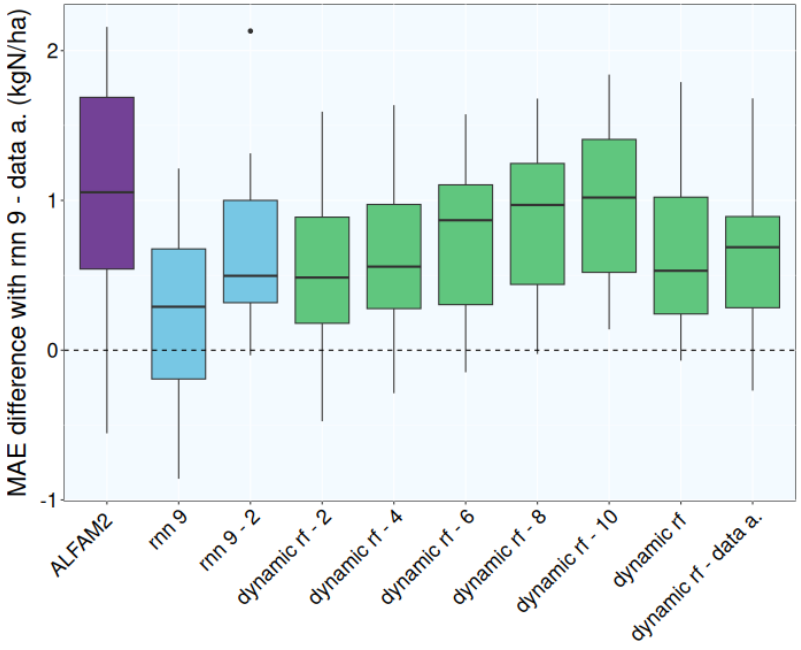

Figure A3. Difference in mean absolute error computed over all time points of the emission dynamics relative to the 'rnn 9 - data a.' model, across the 15 cross-validation test sets. Colors indicate model types: blue for neural networks and green for random forests. For recurrent neural networks, only models derived from architecture rnn 9 are shown. See Tables 6 and 7 for model details.

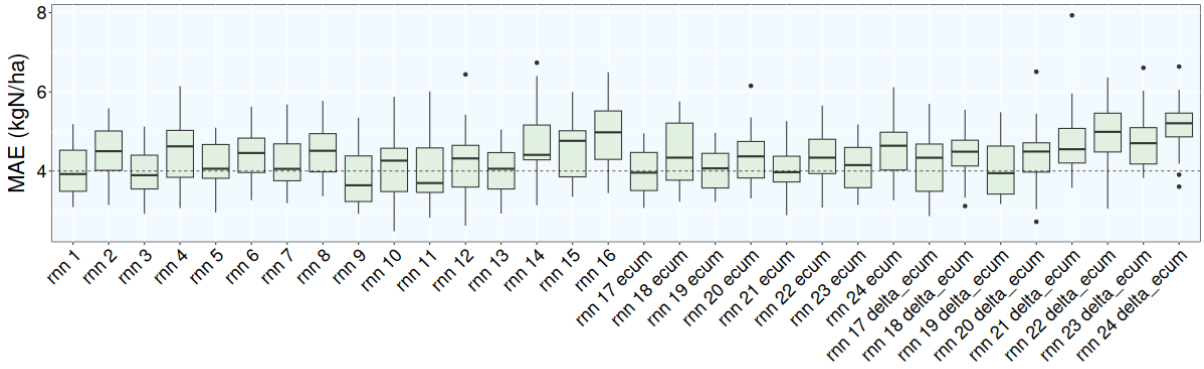

Figure A4. Mean absolute error for models rnn 1 to rnn 24, computed over all time points of the emission dynamics across the 15 cross-validation test sets. For models rnn 17 to rnn 24, which were trained to predict both cumulative and absolute emissions, final cumulative values were obtained either directly from the cumulative outputs ('rnn 17 ecum' to 'rnn 24 ecum') or by summing the absolute emissions ('rnn 17 delta_ecum' to 'rnn 24 delta_ecum').

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Table A3. Selected hyperparameters for the models.

Model	Optimized hyperparameter	Value
Static random forest	mtry	15
	nodesize	4
Dynamic random forest	mtry	2
Recurrent neural networks - without embeddings	hidden size	512
Recurrent neural networks - with embeddings	hidden size	512

Table A4. Number of observations in each dataset version used for model training.

Dataset	Corresponding models	Number of Observations
Final cumul dataset	statics	699
Original dataset	dynamics	6,169
Interpolated dataset – dt = 2	dynamics	39,968
Interpolated dataset – dt = 4	dynamics	20,144
Interpolated dataset – dt = 6	dynamics	13,537
Interpolated dataset – dt = 8	dynamics	10,219
Interpolated dataset – dt = 10	dynamics	8,240
Augmented dataset	dynamics	98,277

Table A5. Static neural network models.

Model name	Number of layers	Layer sizes	Embeddings
static nn 1	1	16	no
static nn 2	2	32 - 16	no
static nn 3	3	64 - 32 - 16	no
static nn 4	4	128 - 64 - 32 - 16	no
static nn 5	5	256 - 128 - 64 - 32 - 16	no
static nn 6	6	512 - 256 - 128 - 64 - 32 - 16	no
static nn 7	1	16	yes
static nn 8	2	32 - 16	yes
static nn 9	3	64 - 32 - 16	yes
static nn 10	4	128 - 64 - 32 - 16	yes
static nn 11	5	256 - 128 - 64 - 32 - 16	yes
static nn 12	6	512 - 256 - 128 - 64 - 32 - 16	yes

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Table A6. Recurrent neural network architectures.

Architecture name	Response	Bidirectional	MLP	Embeddings
rnn 1	cumul	yes	yes	yes
rnn 2	cumul	yes	yes	no
rnn 3	cumul	yes	no	yes
rnn 4	cumul	yes	no	no
rnn 5	cumul	no	yes	yes
rnn 6	cumul	no	yes	no
rnn 7	cumul	no	no	yes
rnn 8	cumul	no	no	no
rnn 9	Δ cumul	yes	yes	yes
rnn 10	Δ cumul	yes	yes	no
rnn 11	Δ cumul	yes	no	yes
rnn 12	Δ cumul	yes	no	no
rnn 13	Δ cumul	no	yes	yes
rnn 14	Δ cumul	no	yes	no
rnn 15	Δ cumul	no	no	yes
rnn 16	Δ cumul	no	no	no
rnn 17	(cumul, Δ cumul)	yes	yes	yes
rnn 18	(cumul, Δ cumul)	yes	yes	no
rnn 19	(cumul, Δ cumul)	yes	no	yes
rnn 20	(cumul, Δ cumul)	yes	no	no
rnn 21	(cumul, Δ cumul)	no	yes	yes
rnn 22	(cumul, Δ cumul)	no	yes	no
rnn 23	(cumul, Δ cumul)	no	no	yes
rnn 24	(cumul, Δ cumul)	no	no	no

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Table A7. Complete list of models. *For static neural networks and recurrent neural network models, see Tables A5 and A6 for detailed architectures, respectively.

Model name	Type	Algo	Timestep	Training set
static rf	static	random forest	-	final cumul
static nn 1	static	static nn 1*	-	final cumul
static nn 2	static	static nn 2*	-	final cumul
static nn 3	static	static nn 3*	-	final cumul
static nn 4	static	static nn 4*	-	final cumul
static nn 5	static	static nn 5*	-	final cumul
static nn 6	static	static nn 6*	-	final cumul
static nn 7	static	static nn 7*	-	final cumul
static nn 8	static	static nn 8*	-	final cumul
static nn 9	static	static nn 9*	-	final cumul
static nn 10	static	static nn 10*	-	final cumul
static nn 11	static	static nn 11*	-	final cumul
static nn 12	static	static nn 12*	-	final cumul
dynamic rf - 2	dynamic	random forest	2	interpolated - dt = 2
dynamic rf - 4	dynamic	random forest	4	interpolated - dt = 4
dynamic rf - 6	dynamic	random forest	6	interpolated - dt = 6
dynamic rf - 8	dynamic	random forest	8	interpolated - dt = 8
dynamic rf - 10	dynamic	random forest	10	interpolated - dt = 10
dynamic rf	dynamic	random forest	irregular	original
dynamic rf - data a.	dynamic	random forest	irregular	augmented
rnn 1	dynamic	rnn 1*	irregular	original
rnn 2	dynamic	rnn 2*	irregular	original
rnn 3	dynamic	rnn 3*	irregular	original
rnn 4	dynamic	rnn 4*	irregular	original
rnn 5	dynamic	rnn 5*	irregular	original
rnn 6	dynamic	rnn 6*	irregular	original
rnn 7	dynamic	rnn 7*	irregular	original
rnn 8	dynamic	rnn 8*	irregular	original
rnn 9	dynamic	rnn 9*	irregular	original
rnn 10	dynamic	rnn 10*	irregular	original
rnn 11	dynamic	rnn 11*	irregular	original
rnn 12	dynamic	rnn 12*	irregular	original
rnn 13	dynamic	rnn 13*	irregular	original
rnn 14	dynamic	rnn 14*	irregular	original
rnn 15	dynamic	rnn 15*	irregular	original
rnn 16	dynamic	rnn 16*	irregular	original
rnn 17	dynamic	rnn 17*	irregular	original
rnn 18	dynamic	rnn 18*	irregular	original
rnn 19	dynamic	rnn 19*	irregular	original
rnn 20	dynamic	rnn 20*	irregular	original
rnn 21	dynamic	rnn 21*	irregular	original
rnn 22	dynamic	rnn 22*	irregular	original
rnn 23	dynamic	rnn 23*	irregular	original
rnn 24	dynamic	rnn 24*	irregular	original
rnn 9 - 2	dynamic	rnn 9*	2	interpolated - dt = 2
rnn 9 - data a.	dynamic	rnn 9*	irregular	augmented

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Table A8. Models' performance on the final cumulative emissions, computed with various metrics.

Model	MAE	MSE	RMSE	Pearson's r	R^2	MBE
ALFAM2	5.61	68.51	8.16	0.82	0.66	-0.10
static rf	5.55	81.92	8.70	0.80	0.63	0.29
static nn 12	5.28	65.81	8.00	0.84	0.67	0.66
dynamic rf - 2	4.95	69.38	8.04	0.83	0.67	0.03
dynamic rf - 4	4.98	69.32	8.03	0.83	0.68	0.29
dynamic rf - 6	5.03	70.42	8.09	0.83	0.67	0.80
dynamic rf - 8	5.01	68.78	8.01	0.84	0.68	0.77
dynamic rf - 10	4.93	67.43	7.92	0.84	0.68	0.61
dynamic rf	5.05	68.04	7.99	0.84	0.68	0.07
dynamic rf - data a.	4.95	67.18	7.96	0.84	0.69	-0.99
rnn 9	4.84	51.83	7.06	0.87	0.75	-0.73
rnn 9 - 2	5.67	95.79	9.10	0.80	0.55	-0.96
rnn 9 - data a.	4.56	46.22	6.71	0.88	0.77	-0.29

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

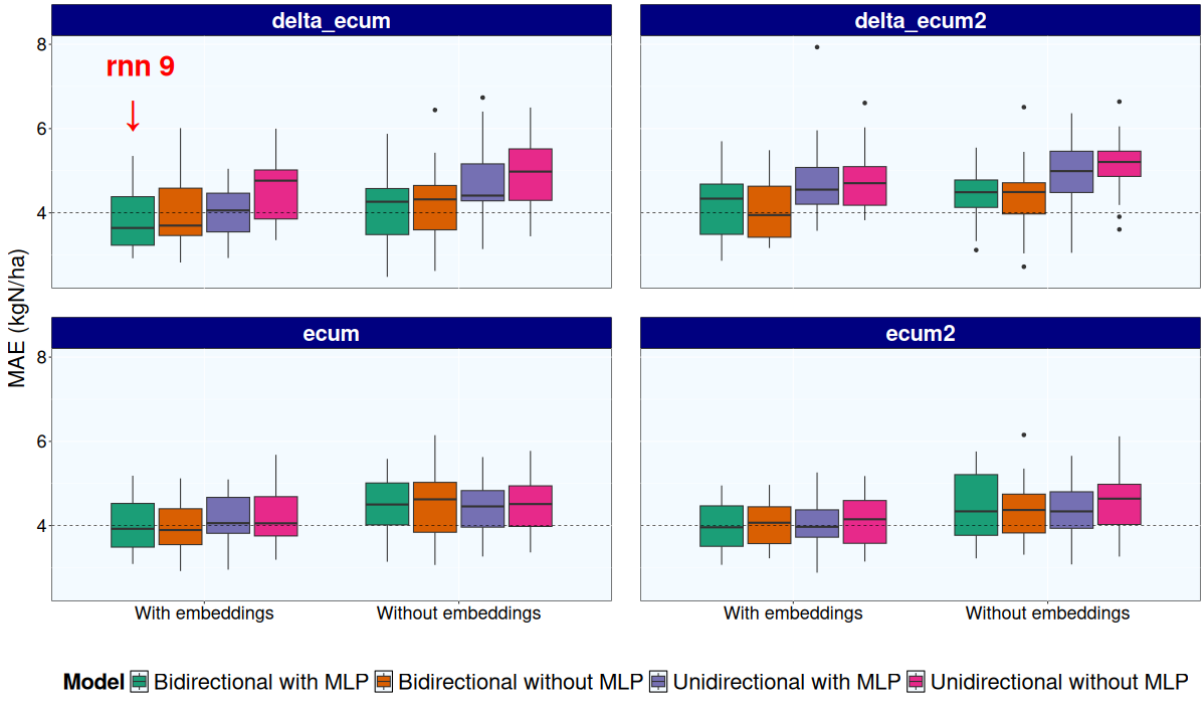

Figure A5. Mean absolute error for models rnn 1 to rnn 24, computed over all time points of the emission dynamics across the 15 cross-validation test sets, with a representation highlighting the architectural characteristics. The different subplots correspond to the target variable used for training. For models 'rnn 17' to 'rnn 24', which were trained using both cumulative and interval emissions as targets, predictions were subsequently computed using either cumulative or interval emissions. These correspond to the 'ecum2' and 'delta_ecum2' subplots, respectively. The x-axis indicates whether embedding layers were used for categorical variables. Colors represent the model architecture: unidirectional or bidirectional recurrent networks, and the presence (with MLP) or absence (without MLP) of an additional dense layer between the recurrent layer and the output.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

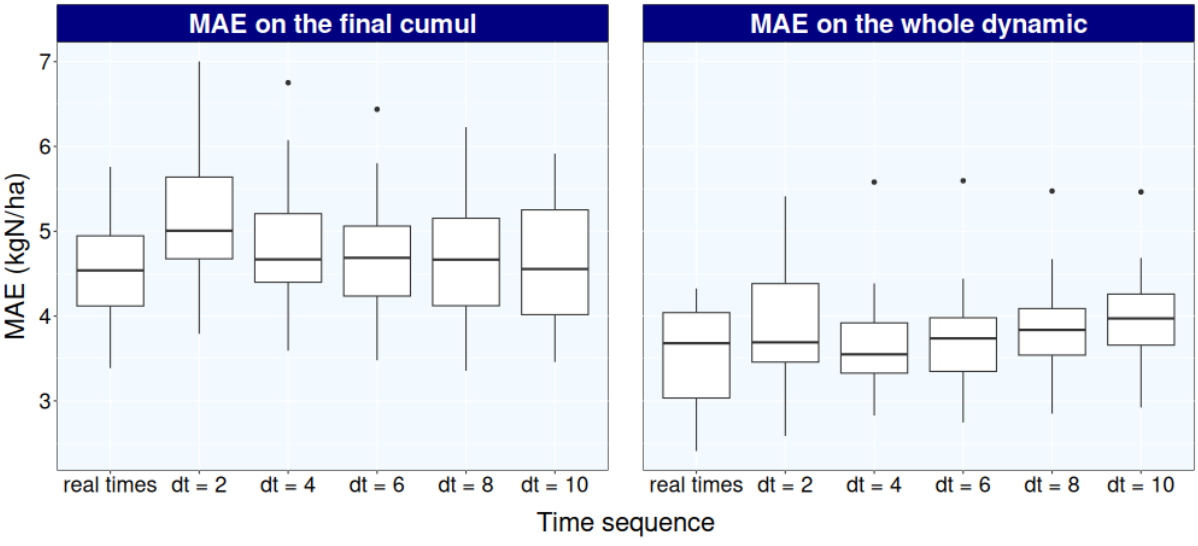

Figure A6. Mean absolute error for the 'rnn 9 – data a.' model on final cumulative emissions (left subplot) and on the full emission dynamics (right subplot), as a function of the time sequence used for prediction, across all test sets from the cross-validation procedure. 'Real times' refers to the observed time sequence, while 'dt = 2', and so on, denote regular time steps. For sequences with regular time steps, linear interpolation was applied after model prediction to obtain values at the measured time points.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

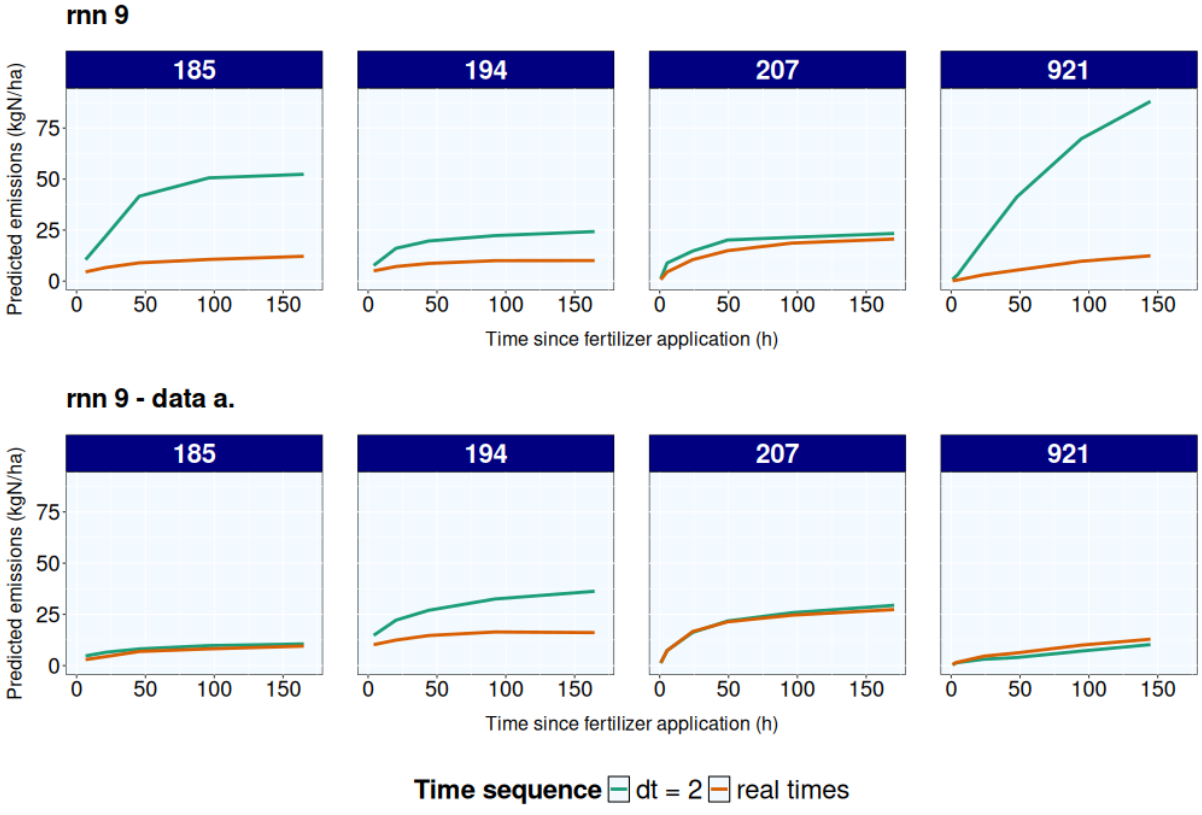

Figure A7. Effect of data augmentation on the sensitivity of predictions to the time step. Predicted emission dynamics for four trials from the first test set, using models 'rnn 9' and 'rnn 9 – data a.', with either the observed time sequence ('real times', orange curves) or a regular time step of 2 ('dt = 2', green curves). For the 'dt = 2' case, interpolation was performed after model application to obtain predictions at the measured time points.

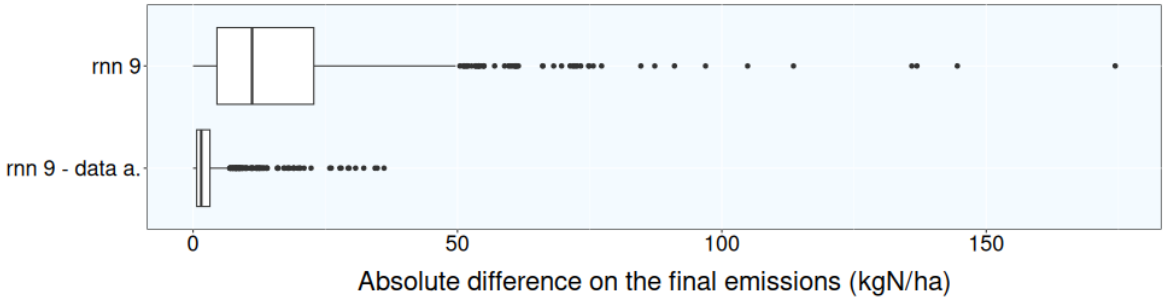

Figure A8. Absolute difference in final cumulative emissions between predictions obtained using the real time sequence and those obtained using a regular time step of 2, for models 'rnn 9' and 'rnn 9 – data a.'. Each boxplot represents variability across the 15 cross-validation test sets.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

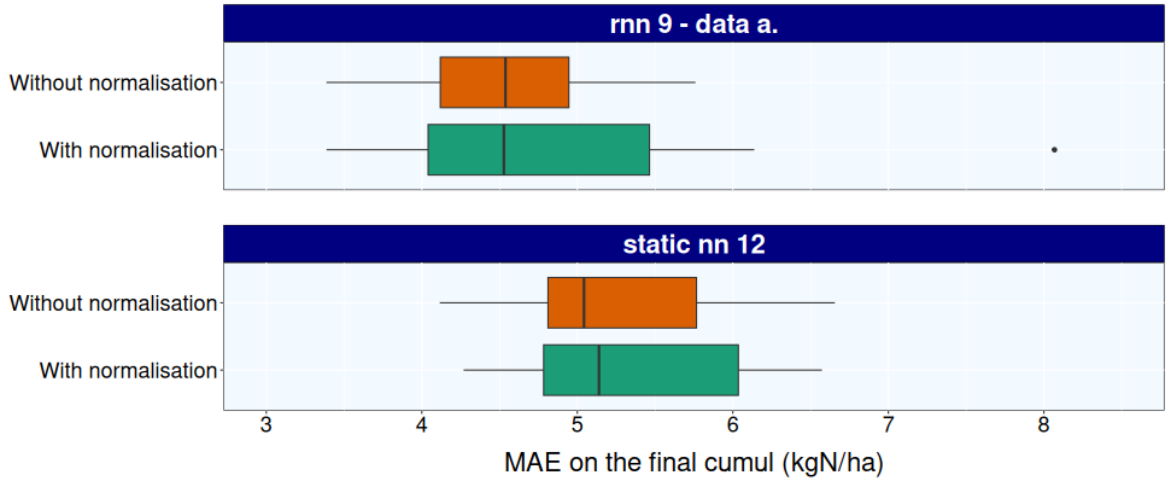

Figure A9. Mean absolute error of final cumulative emissions using either raw or normalized data, for both the "rnn 9 - data a." model and the "static nn 12" model.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

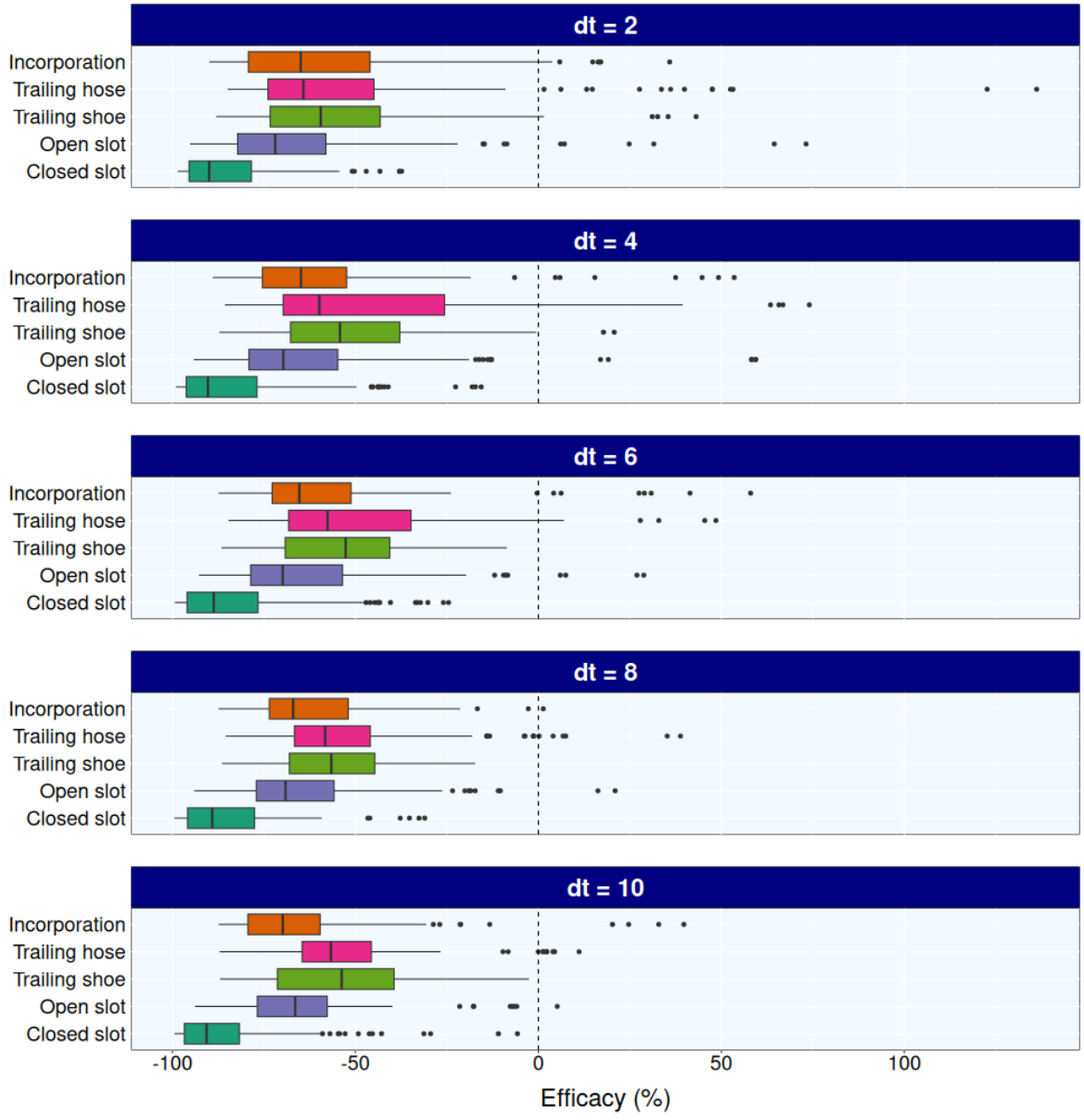

Figure A10. Percentage reduction in emissions at 72 hours for different application techniques compared to the reference broadcast method, calculated using the 'rnn 9 – data a.' model over the 128 scenarios defined by all possible combinations of the values in Table 1. Each subplot corresponds to a different time step used for the predictions.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

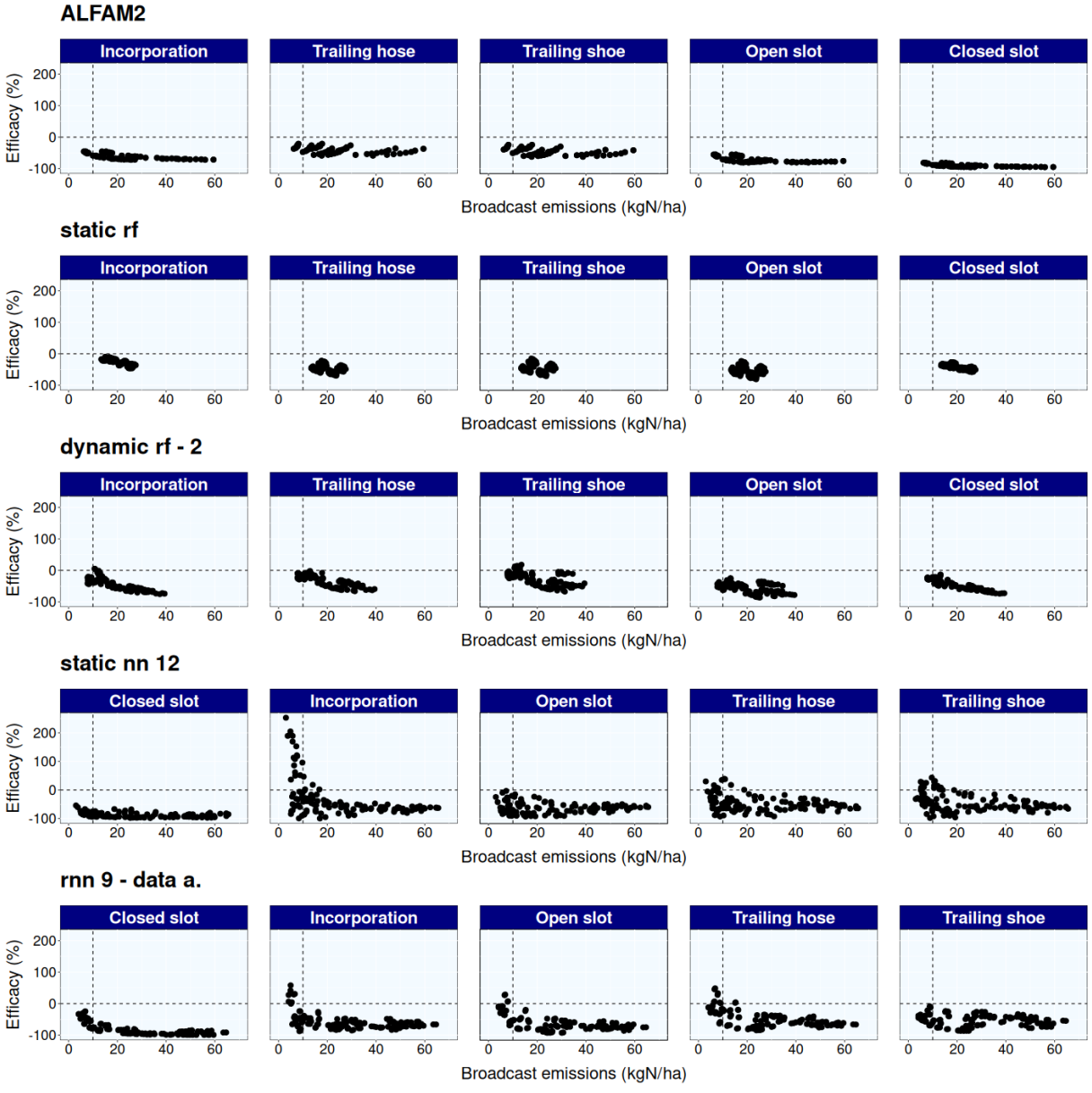

Figure A11. Efficacy of the reduction techniques as a function of emissions from the reference broadcast method at 72h, for models ALFAM2, 'static rf', 'dynamic rf – 2', 'static nn 12', and 'rnn 9 – data a.'.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Code and data availability. The source code and the data used in this research are available on github at: https://github.com/armandfavrot/Improving-ammonia-emission-predictions-with-dynamic-machine-learning-models.

The best model was implemented as a Python package available on PyPI at https://pypi.org/project/nh3pred/.

Author contributions. **Armand Favrot**: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sophie Génermont**: Writing – review and editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Vincent Guigue**: Writing – review and editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Céline Décuq**: Writing – review and editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **David Makowski**: Writing – review and editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements. This work was supported by the French ANR under the "Investissements d'avenir" programme with the reference ANR-16-CONV-0003 and from the AgroEcoSystem department of INRAE. This work was granted access to the HPC resources of IDRIS under the allocation 2024-AD011015906 made by GENCI. During the preparation of this work the authors used ChatGPT in order to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

References

Behera, S. N., Sharma, M., Aneja, V. P., and Balasubramanian, R.: Ammonia in the atmosphere: a review on emission sources, atmospheric chemistry and deposition on terrestrial bodies, Environ. Sci. Pollut. Res., 20, 8092–8131, https://doi.org/https://doi.org/10.1007/s11356-013-2051-9, 2013.

Beusen, A., Bouwman, A., Heuberger, P., Van Drecht, G., and Van Der Hoek, K.: Bottom-up uncertainty estimates of global ammonia emissions from global agricultural production systems, Atmos. Environ., 42, 6067–6077, https://doi.org/https://doi.org/10.1016/j.atmosenv.2008.03.044, 2008.

Braschkat, J., Mannheim, T., and Marschner, H.: Estimation of ammonia losses after application of liquid cattle manure on grassland, Z. Pflanzenernähr. Bodenkd., 160, 117–123, https://doi.org/https://doi.org/10.1002/jpln.19971600202, 1997.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, https://doi.org/https://doi.org/10.1023/A:1010950718922, 2001.

Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Whitsel, L., and Kaufman, J. D.: Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association, Circulation, 121, 2331–2378, https://doi.org/https://doi.org/10.1161/CIR.0b013e3181dbece1, 2010.

Chelani, A. B., Rao, C. C., Phadke, K., and Hasan, M.: Prediction of sulphur dioxide concentration using artificial neural networks, Environ. Model. Softw., 17, 159–166, https://doi.org/https://doi.org/10.1016/S1364-8152(01)00061-5, 2002.

Elman, J. L.: Finding structure in time, Cogn. Sci., 14, 179–211, https://doi.org/https://doi.org/10.1207/s15516709cog1402_1, 1990.

European Environment Agency: National Emission reduction Commitments Directive reporting status 2022, https://www.eea.europa.eu/publications/national-emission-reduction-commitments-directive-2022, (accessed 20 October 2025), 2022.

European Environment Agency: EMEP/EEA air pollutant emission inventory guidebook 2023, https://www.eea.europa.eu/en/analysis/publications/emep-eea-guidebook-2023, (accessed 20 October 2025), 2023.

European Environment Agency: Air pollution in Europe: 2024 reporting status under the National Emission reduction Commitments Directive, https://www.eea.europa.eu/en/analysis/publications/national-emission-reduction-commitments-directive-2024, (accessed 20 October 2025), 2024.

Favrot, A., Génermont, S., Décuq, C., and Makowski, D.: Machine learning for ammonia volatilization prediction and slurry application management, J. Environ. Sci., 160, 481–489, https://doi.org/https://doi.org/10.1016/j.jes.2025.04.045, 2026.

Feng, R., Zheng, H.-j., Gao, H., Zhang, A.-r., Huang, C., Zhang, J.-x., Luo, K., and Fan, J.-r.: Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in Hangzhou, China, J. Clean. Prod., 231, 1005–1015, https://doi.org/https://doi.org/10.1016/j.jclepro.2019.05.319, 2019.

Gericke, D., Bornemann, L., Kage, H., and Pacholski, A.: Modelling ammonia losses after field application of biogas slurry in energy crop rotations, Water Air Soil Pollut., 223, 29–47, https://doi.org/https://doi.org/10.1007/s11270-011-0835-4, 2012.

Guthrie, S., Giles, S., Dunkerley, F., Tabaqchali, H., Harshfield, A., Ioppolo, B., and Manville, C.: The impact of ammonia emissions from agriculture on biodiversity, Tech. rep., RAND Europe and The Royal Society, https://doi.org/https://doi.org/10.7249/RR2695, 2018.

Génermont, S. and Cellier, P.: A mechanistic model for estimating ammonia volatilization from slurry applied to bare soil, Agric. For. Meteorol., 88, 145–167, https://doi.org/https://doi.org/10.1016/S0168-1923(97)00044-0, 1997.

Hafner, S. D., Pacholski, A., Bittman, S., Burchill, W., Bussink, W., Chantigny, M., Carozzi, M., Génermont, S., Häni, C., Hansen, M. N., Huijsmans, J., Hunt, D., Kupper, T., Lanigan, G., Loubet, B., Misselbrook, T., Meisinger, J. J., Neftel, A., Nyord, T., Pedersen, S. V.,

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Sintermann, J., Thompson, R. B., Vermeulen, B., Vestergaard, A. V., Voylokov, P., Williams, J. R., and Sommer, S. G.: The ALFAM2 database on ammonia emission from field-applied manure: Description and illustrative analysis, Agric. For. Meteorol., 258, 66–79, https://doi.org/https://doi.org/10.1016/j.agrformet.2017.11.027, 2018.

Hafner, S. D., Pacholski, A., Bittman, S., Carozzi, M., Chantigny, M., Génermont, S., Häni, C., Hansen, M. N., Huijsmans, J., Kupper, T., Misselbrook, T., Neftel, A., Nyord, T., and Sommer, S. G.: A flexible semi-empirical model for estimating ammonia volatilization from field-applied slurry, Atmos. Environ., 199, 474–484, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.11.034, 2019.

Hafner, S. D., Pedersen, J., Fuß, R., Kamp, J. N., Dalby, F. R., Amon, B., Pacholski, A., Adamsen, A. P. S., and Sommer, S. G.: Improved tools for estimation of ammonia emission from field-applied animal slurry: Refinement of the ALFAM2 model and database, Atmos. Environ., 340, 120 910, https://doi.org/https://doi.org/10.1016/j.atmosenv.2024.120910, 2025.

Hamrani, A., Akbarzadeh, A., and Madramootoo, C. A.: Machine learning for predicting greenhouse gas emissions from agricultural soils, Sci. Total Environ., 741, 140 338, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.140338, 2020.

He, Y., Xu, R., Prior, S. A., Yang, D., Yang, A., and Chen, J.: Satellite-detected ammonia changes in the United States: Natural or anthropogenic impacts, Sci. Total Environ., 789, 147 899, https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.147899, 2021.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Comput., 9, 1735–1780, https://doi.org/https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Huijsmans, J., Vermeulen, G., Hol, J., and Goedhart, P.: A model for estimating seasonal trends of ammonia emission from cattle manure applied to grassland in the Netherlands, Atmos. Environ., 173, 231–238, https://doi.org/https://doi.org/10.1016/j.atmosenv.2017.10.050, 2018.

Häni, C., Sintermann, J., Kupper, T., Jocher, M., and Neftel, A.: Ammonia emission after slurry application to grassland in Switzerland, Atmos. Environ., 125, 92–99, https://doi.org/https://doi.org/10.1016/j.atmosenv.2015.10.069, 2016.

Kamp, J. N., Hafner, S. D., Huijsmans, J., van Boheemen, K., Götze, H., Pacholski, A., and Pedersen, J.: Comparison of two micrometeorological and three enclosure methods for measuring ammonia emission after slurry application in two field experiments, Agric. For. Meteorol., 354, 110 077, https://doi.org/https://doi.org/10.1016/j.agrformet.2024.110077, 2024.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems, edited by Pereira, F., Burges, C., Bottou, L., and Weinberger, K., vol. 25, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf, 2012.

Krupa, S.: Effects of atmospheric ammonia (NH3) on terrestrial vegetation: a review, Environ. Pollut., 124, 179–221, https://doi.org/https://doi.org/10.1016/S0269-7491(02)00434-7, 2003.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2, 18–22, 2002.

Lim, Y., Moon, Y.-S., and Kim, T.-W.: Artificial neural network approach for prediction of ammonia emission from field-applied manure and relative significance assessment of ammonia emission factors, Eur. J. Agron., 26, 425–434, https://doi.org/https://doi.org/10.1016/j.eja.2007.01.008, 2007.

Menzi, H., Katz, P., Fahrni, M., Neftel, A., and Frick, R.: A simple empirical model based on regression analysis to estimate ammonia emissions after manure application, Atmos. Environ., 32, 301–307, https://doi.org/https://doi.org/10.1016/S1352-2310(97)00239-2, 1998.

Misselbrook, T., Nicholson, F., and Chambers, B.: Predicting ammonia losses following the application of livestock manure to land, Bioresour. Technol., 96, 159–168, https://doi.org/https://doi.org/10.1016/j.biortech.2004.05.004, 2005.

Pan, S. J. and Yang, Q.: A survey on transfer learning, IEEE Trans. Knowl. Data Eng., 22, 1345–1359, https://doi.org/https://doi.org/10.1109/TKDE.2009.191, 2009.

https://doi.org/10.5194/egusphere-2026-2404
Preprint. Discussion started: 5 May 2026

Plöchl, M.: Neural network approach for modelling ammonia emission after manure application on the field, Atmos. Environ., 35, 5833–5841, https://doi.org/https://doi.org/10.1016/S1352-2310(01)00281-3, 2001.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533–536, https://doi.org/https://doi.org/10.1038/323533a0, 1986.

Shen, H., Chen, Y., Hu, Y., Ran, L., Lam, S. K., Pavur, G. K., Zhou, F., Pleim, J. E., and Russell, A. G.: Intense Warming Will Significantly Increase Cropland Ammonia Volatilization Threatening Food Security and Ecosystem Health, One Earth, 3, 126–134, https://doi.org/https://doi.org/10.1016/j.oneear.2020.06.015, 2020.

Shorten, C. and Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning, J. Big Data, 6, 1–48, https://doi.org/https://doi.org/10.1186/s40537-019-0197-0, 2019.

Sommer, S. G. and Olesen, J. E.: Modelling ammonia volatilization from animal slurry applied with trail hoses to cereals, Atmos. Environ., 34, 2361–2372, https://doi.org/https://doi.org/10.1016/S1352-2310(99)00442-2, 2000.

Søgaard, H., Sommer, S., Hutchings, N., Huijsmans, J., Bussink, D., and Nicholson, F.: Ammonia volatilization from field-applied animal slurry—the ALFAM model, Atmos. Environ., 36, 3309–3319, https://doi.org/https://doi.org/10.1016/S1352-2310(02)00300-X, 2002.

Van Damme, M., Clarisse, L., Franco, B., Sutton, M. A., Erisman, J. W., Kruit, R. W., Van Zanten, M., Whitburn, S., Hadji-Lazaro, J., Hurtmans, D., et al.: Global, regional and national trends of atmospheric ammonia derived from a decadal (2008–2018) satellite record, Environ. Res. Lett., 16, 055 017, https://doi.org/https://doi.org/10.1088/1748-9326/abd5e0, 2021.

Vovk, V., Gammerman, A., and Shafer, G.: On-line compression modeling I: conformal prediction, in: Algorithmic Learning in a Random World, edited by Vovk, V., Gammerman, A., and Shafer, G., pp. 189–221, Springer US, Boston, MA, ISBN 978-0-387-25061-8, https://doi.org/https://doi.org/10.1007/0-387-25061-1_8, 2005.

Warner, J. X., Dickerson, R. R., Wei, Z., Strow, L. L., Wang, Y., and Liang, Q.: Increased atmospheric ammonia over the world's major agricultural areas detected from space, Geophys. Res. Lett., 44, 2875–2884, https://doi.org/https://doi.org/10.1002/2016GL072305, 2017.

Webb, J., Pain, B., Bittman, S., and Morgan, J.: The impacts of manure application methods on emissions of ammonia, nitrous oxide and on crop response—A review, Agric. Ecosyst. Environ., 137, 39–46, https://doi.org/https://doi.org/10.1016/j.agee.2010.01.001, 2010.

Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., and Xu, H.: Time series data augmentation for deep learning: A survey, arXiv preprint arXiv:2002.12478, 2020.

Xu, P., Li, G., Zheng, y., Fung, J., Chen, A., Zeng, Z., Shen, H., Hu, M., Mao, J., Zheng, Y., Cui, X., Guo, Z., Chen, Y., Feng, L., He, S., Zhang, X., Lau, A., Tao, S., and Houlton, B.: Fertilizer management for global ammonia emission reduction, Nature, 626, 792–798, https://doi.org/https://doi.org/10.1038/s41586-024-07020-z, 2024.