MSc thesis: Random Forest Modelling of Surface Ammonia Concentrations in the Netherlands

Name: Tycho Jongenelen
Study: MSc Industrial Ecology
Student number: 1743015 (LU), 5193575 (TUD)
Supervisors: Prof. dr. ir. J.W. Erisman
Prof. dr. ir. H.X. Lin
Dr. ir. E. Dammers

Date: March 2022

Universiteit Leiden

TU Delft

TNO innovation for life

Abstract

The emission of reactive nitrogen species increased rapidly in the twentieth century causing a significant perturbation of the nitrogen cycle. This causes many detrimental effects for both humans and ecosystems such as eutrophication, acidification and biodiversity loss. Ammonia, which is a reactive nitrogen species is difficult to quantify because the compound is hard to observe with measurement equipment such as ground-based measurement stations and remote sensing instruments such as CrIS. Therefore, this thesis aims to improve the monitoring of surface ammonia concentrations with a machine learning technique called random forests (RF). These type of models can detect complex and non-linear relationships between variables and are frequently used in other air pollution studies. In this study, a RF model has been built to estimate the ammonia surface concentration with vertical column density (VCD) datasets, meteorological variables and land-use variables. Different combinations of VCD datasets (either modelled or CrIS VCD data) have been used to train the model and to predict the surface ammonia concentration. The result of the study is that RF models are statistically more accurate at estimating the surface ammonia concentration than the LOTOS-EUROS model when validated by ground-based measurements stations from the MAN and LML network. Especially trained RF models that have been supplied with CrIS VCD data during the prediction phase show strong performance. Moreover, when comparing the RF model that has been trained without VCD data to the 'complete' RF models, the complete RF models show better performance, proving the added benefit of incorporating satellite data in RF modelling. Recommendations for further research include performing similar experiments with other input variables and other machine learning algorithms, validating the performance of the RF model in different years and considering using other datasets as the ground-truth variable for RF models – such as ground-based measurement data from MAN and LML.

Preface

Reactive nitrogen species such as ammonia can have a detrimental impact on both ecosystems and society and the emission of ammonia has increased rapidly during the twentieth century. Monitoring ammonia has proved to be difficult due to many obstacles, such as the lack of ground-based measurement stations and the uncertainty of satellite ammonia observations. Therefore, this explores the possibilities for modelling ammonia surface concentrations with random forest algorithms.

Writing my thesis has been a very pleasant experience and I look back on a period in which I have learnt – not only a lot about conducting research about ammonia, programming, and machine learning – but also a lot about myself. Before we will be diving into this thesis, I want to thank some people for their support during my thesis. My first thanks go out to Enrico: During this whole period, he motivated me to think critically about the research I was conducting and helped to keep me on track during my thesis. Also, his hospitality and humor helped me feel welcome and at ease at TNO almost immediately. I also want to thank Jan Willem for offering me this thesis topic in the first place and because of his constructive and pleasant supervision. Likewise, I am hereby thanking Hai Xiang for his kind supervision and for helping me to better understand the world of machine learning which has been completely new for me.

In general, I want to thank the Climate, Air and Sustainability department from TNO for the kindness and help they offered me during my internship. Specifically, I want to thank Arjo for helping me to get started on the supercomputer and fix all the Python modules I needed, and Astrid for her insights about getting the random forest model 'out of the black-box'.

Finally, I want to thank my friends, family and especially Dieke for supporting me and helping me to believe in myself during my thesis. I am very grateful that they are a part of my life.

Den Haag, March 2022

Tycho Jongenelen

Table of contents

List of abbreviations

CrIS:	Cross-track Infrared Sounder
EU:	European Union
IASI:	Infrared Atmospheric Sounding Interferometer
LML:	Landelijk Meetnet Luchtkwaliteit
MAE:	Mean Average Error
MAN:	Meetnet Ammoniak in Natuurgebieden
NRMSE:	Normalized Root Mean Square Error
ppb:	Parts per billion
r:	Correlation Coefficient
RF:	Random forest
RMSE:	Root Mean Squared Error
RSE:	Residual Squared Error
VCD:	Vertical column density

Section 1: Introduction

Nitrogen is an essential component for biological life as it is the key element of amino acids and nucleic acids (Denk et al., 2017). Most of the nitrogen on earth is found in the atmosphere as dinitrogen (N_2) and is classified as nonreactive nitrogen, while the remaining nitrogen species are called reactive nitrogen (N_r; Galloway et al., 2003). Examples of N_r species are nitric oxide (NO), nitric dioxide (NO_2), ammonia (NH_3), and nitrate (NO_3^-) Atmospheric nitrogen is naturally converted into N_r species by, among others, lightning, biomass burning and biological nitrogen fixation. Before the Anthropocene, the production of N_r was balanced by denitrification, which converts N_r back to N_2 (Ayres et al., 1994).

However, in the twentieth century, the N_r production rapidly expanded as the global population increased. The main processes that led to this increase are the production of artificial fertilizer (Haber-Bosch process), increased livestock activities, the combustion of fossil fuels and the increased cultivation of legumes and rice (Galloway et al., 2003). This caused a perturbation in the nitrogen cycle, leading to a wide range of detrimental effects on ecosystems such as eutrophication, acidification, biodiversity loss, global warming and the depletion of the ozone layer.

Currently, the Netherlands is undergoing a nitrogen crisis, as 118 of 162 Dutch nature reserves have exceeded nitrogen deposition thresholds (Stokstad, 2019). This crisis has far-reaching consequences for the Netherlands because it not only threatens the environment but also the Dutch economy. Due to nitrogen emission limits, housing and infrastructural projects are halted and the speed limit on the Dutch highway had to be restricted from 130 to 100 kilometres per hour. A significant portion of the emitted reactive nitrogen can be allocated to the ammonia emission by the agricultural industry (TNO, 2019).

Global regulations to mitigate ammonia emissions are limited. In contrast to nitrogen oxides, regulations aimed at controlling ammonia emissions have drawn less attention (Sutton et al., 2020). An exception to the rule is the policy in EU member states: the NEC Directive (National Emission Ceilings) sets long-term goals and emission ceilings aimed at reducing the ammonia emission in its member states (European Commission, 2001).

The quantification of the ammonia emissions, concentrations and deposition rates is challenging – even though the contribution of this compound to the perturbed nitrogen cycle is significant (Erisman et al., 2007). Ground observations are limited and there is a considerable level of uncertainty surrounding the observations made by various satellites (e.g. van Damme et al., 2015; Dammers et al., 2016; Dammers et al., 2017b; Dammers et al., 2019). Chemical transport models (CTMs) such as LOTOS-EUROS models atmospheric compounds through physical laws and estimates the emission, distribution and deposition of ammonia. While offering a high spatiotemporal resolution of ammonia emission and deposition, they generally seem to underestimate the actual ammonia concentration (van Damme et al., 2014, van der Graaf, 2022).

Given these challenges, this thesis aims to improve the monitoring of surface ammonia concentrations in the atmosphere by machine-learning modelling. With the Netherlands as study domain, random forest (RF) models will use remote-sensing, meteorological and land-use datasets to predict the ammonia surface concentration. The results will be compared to the LOTOS-EUROS model and ground-based measurements stations such from the LML and MAN network. Previous studies which have used machine learning algorithms for modelling air quality demonstrate its advantages such as the ability to detect complex and non-linear relationships, the possibility to inquire which variables are

most important in air quality modelling and how these variables influence the pollutant concentration (Breiman, 2001; Brokamp et al., 2017).

This thesis is structured as follows: In Section 2, the core concepts and the knowledge gap are further explained. Moreover, the research question and the objective are formulated in this section. In Section 3, the research methods and data requirements are described. In Section 4, the results of the research will be described, discussed and limitations and recommendations for further research will be provided. In Section 5, the thesis is concluded by answering the research question and the sub-research questions.

Section 2: Literature review

In this Section, the core concepts, such as the behaviour of ammonia in the biosphere, ammonia monitoring methods and the utilization of machine learning in trace gas modelling will be discussed. Moreover, the objective of this thesis, the research question and the sub-research questions will be formulated.

2.1: Ammonia in the biosphere

Ammonia is one of the reactive nitrogen species causing a perturbation in the nitrogen cycle. The sources, sinks, processes and pressure of ammonia on the environment will be further explained with the atmospheric dispersion box model as a theoretical framework (Figure 1). Inflow, chemical production and emission are the sources of a compound in a box, outflow, chemical loss and deposition are the sinks in the box.

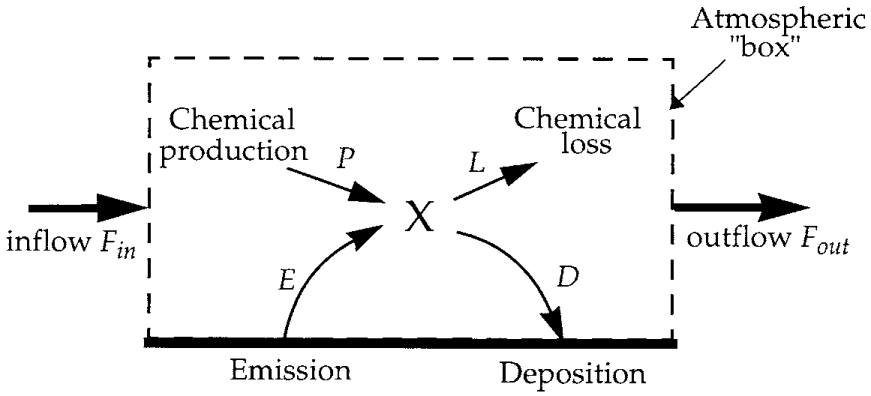

Figure 1: Graphical representation of the classic atmospheric dispersion model. The concentration of compound X increases by chemical production, emission and inflow due to the wind. In contrast, the concentration can also decrease again by chemical loss processes, deposition or by the outflow due to the wind. (Source: Jacob, 1999.)

2.1.1: Emission, chemistry, deposition and transport of ammonia

The emission of ammonia is mainly caused by the agricultural industry due to several processes such as livestock housing and grazing, manure volatilization and fertilizer application (Sutton et al., 2013). Unmetabolized nitrogen from animal feed ends up in urine and faeces and through microbial processes, ammonia is released from the excreta into the atmosphere (Gay & Knowlton., 2005). The volatilization rate of ammonia from excreta is influenced by meteorological conditions. Ammonia volatilization increases with rising temperature, relative humidity, wind speed and soil moisture (Sutton et al., 2013, Sommer et al., 1991). Ammonia emissions from livestock housing also depend on the ventilation and the building type (Koerkamp et al., 1998). Natural sources of ammonia include biomass burning (Hegg et al., 1988), ocean-atmosphere exchange of ammonia (Johnson et al., 2008), plant-atmosphere exchange (Sutton et al., 1995), volcanoes (Sutton et al., 2008; Uematsu et al., 2004), land

animals (Dentener & Crutzen, 1994) and seabird and sea lion colonies (Riddick et al., 2012; Fariña et al., 2003).

In the atmosphere, ammonia can interact with acidic gases to form solid aerosols which consist of ammonium (NH_4^+; Ferm, 1998). Examples are nitric acid and sulfuric acid which convert ammonia to respectively ammonium sulphate and ammonium nitrate, as shown in reactions R2.1 and R2.2. In contrast to Reaction 2.1, Reaction 2.2 is reversible and determined by atmospheric temperature and relative humidity (Stelson & Seinfield, 1982).

$$2\ NH_3\ (g) + H_2SO_4(g) \rightarrow (NH_4)_2SO_4\ (s) \quad \text{[R2.1]}$$

$$NH_3\ (g) + HNO_3\ (g) \leftrightarrow NH_4NO_3\ (s) \quad \text{[R2.2]}$$

Ammonia deposition can take place by either dry or wet deposition. Ammonia can be scavenged by precipitation or dissolve in clouds (Asman et al., 1998). These processes fall under wet deposition. On the other hand, through dry deposition, ammonia and ammonium are removed from the atmosphere by depositing on surfaces such as vegetation, soil and water (Farquhar et al. 1980; Duyzer, 1994; Quinn et al., 1988). This deposition is bidirectional, as ammonia can be re-released towards the atmosphere, dependent on the type of surface, surface conditions, the amount of ammonium in the surface and the amount of ammonia in the atmosphere above the surface (Nemitz et al., 2001). The soil, vegetation and water can thus be seen as a reservoir of ammonia, which if not re-released, is chemically transformed by soil processes or removed through run-off.

The inflow and outflow of ammonia take place by atmospheric transport, which is determined by wind speed and wind direction. As stated previously, ammonia can be converted in the atmosphere to aerosols such as ammonium sulphate and ammonium nitrate. The atmospheric lifetime of these aerosols is between 1-15 days, in contrast to just several hours up to two days for NH_3 (Behera et al., 2013; Lutsch et al., 2016; Dammers et al., 2019), which implies NH_3/NH_4^+ can be transported over longer distances.

Finally, indirectly, the ammonia concentration at the surface can be influenced by the boundary layer height (Stull, 1988). Assuming that the boundary layer is mixed homogeneously, an increase in the boundary layer height implies a greater volume in which atmospheric compounds can mix. If the amount of ammonia in the boundary layer remains constant, dilution would take place during the extension of the boundary layer height.

2.1.2: Detrimental effects of ammonia on society and the environment

The accumulation of ammonia (and reactive nitrogen in general) in the atmosphere has many detrimental consequences. First, an excess of ammonia in water bodies – mainly caused by leaching and surface runoff - may result in extreme algae plant growth, reducing water quality and light penetration (Chislock et al., 2013). This process is better known as eutrophication and can cause hypoxia and ultimately 'dead zones'. For humans, this implies that drinking water systems can be contaminated and that regions that depend on the fishing industry will experience economical issues (Walker, 1983; Selman et al., 2008).

Second, acidification of soils can take place due to ammonia deposition, causing damage to soil and plant organisms (Tian & Niu, 2015). This can lead to a decrease in crop yields produced by the agricultural industry. Third, as mentioned in Section 2.1.1, ammonia can be converted into aerosols,

which can cause detrimental health effects such as cardiovascular and pulmonary diseases (Wolfe & Patz, 2002; Pope III & Dockery, 2006).

Finally, excessive deposition of ammonia is a threat to biodiversity as ecosystems can be damaged by eutrophication and acidification (de Vries, 2011). Especially N_r-sensitive areas such as grasslands and peatlands are vulnerable. For society, biodiversity loss will imply the loss of a multitude of essential ecosystem services such as food supply and climate regulation (Haines-Young & Potschin, 2010).

2.2: Techniques for determining the budget of ammonia: ground-based stations, modelling and remote sensing

Ground-based measurements of ammonia with a high temporal resolution are sparse, making it difficult to understand the behaviour of ammonia in the atmosphere. This sparseness is mostly caused by the measurement of ammonia itself, which can be manpower intensive and expensive, mostly due to the physical properties of the species (Shaw et al., 1998, von Bobrutzki et al., 2010). In the Netherlands, hourly ground-based measurements are made by the miniDOAS instrument at six locations (Volten et al., 2012). Monthly averaged observations are done by the Dutch MAN network at more than 300 locations in Natura 2000-areas by passive samplers (Noordijk et al., 2020). According to Dammers (2017a), the Netherlands has the most extensive ammonia measuring network in the world but still struggles to give a good representation of ammonia dispersion and deposition in the Netherlands. This is mostly due to the limited number of instruments measuring ammonia at a sufficiently high spatiotemporal resolution, as well as a lack of instruments measuring the deposition (Schrader et al., 2018).

Another method of observing the ammonia concentration is by using satellite instruments such as the Infrared Atmospheric Sounding Interferometer (IASI) and the Cross-track Infrared Sounder (CrIS). Satellite remote sensing observations are very valuable because they can homogeneously measure vertical column densities (VCD) over large areas with a high frequency. However, there is a considerable uncertainty bound to individual satellite observations of approximately 30-50% (Dammers et al., 2019). This is caused by systematic biases and random errors that occur in the CrIS and IASI instruments and the retrievals applied to the spectra thereafter. Moreover, the availability of satellite observations is highly dependent on the cloud cover and the thermal contrast, which is the temperature difference between the surface and the atmosphere at 1.5 km (van Damme et al., 2015)

Ammonia concentrations can also be estimated by chemical transport models (CTM), which simulate the emission, deposition, transformation and transportation that take place in the atmosphere. Important inputs for CTMs are emission inventories, land-use maps and meteorological data (Manders et al., 2017). Examples of models that can simulate ammonia concentrations and deposition are LOTOS-EUROS, EMEP, GEOS, CMAQ and WRF. These models require detailed emission inventory data which is necessary for capturing the spatial and temporal variability of ammonia. Furthermore, these models are often based on land-use data and animal density maps which are too coarse to accurately model ammonia concentrations (Paulot et al., 2014).

2.3: Machine learning modelling

An addition to CTMs are machine learning models: These are statistical learning algorithms that can detect complex and non-linear relationships between predictors and the response variable (Rybarczyk & Zalakeviciute, 2018). In contrast to CTMs, statistical models do not calculate the physical and chemical processes in a certain domain. Examples of machine learning methods that have been used

in atmospheric modelling include random forests (Lu et al., 2020), artificial neural networks (Di et al., 2016), support vector regression (Dun et al., 2020) and generalized boosted machines (Zhan et al., 2017).

Random forest algorithms are one of the most frequently used methods in machine learning-based atmospheric research (Rybarczyk & Zalakeviciute, 2018). Random forests have been developed by Breiman (2001) and are a collection of many decision trees. Decision trees are machine-learning models that predict the value of a response variable inferred from the given explanatory variables (Fürnkranz, 2017).

Decision trees consist of root nodes, sub-nodes and leaf nodes. The decision tree starts with the root node, is followed by subsequent sub-nodes and ends with the leaf node. The goal of the decision tree is to split the dataset in such a manner, that a certain response variable can be predicted as good as possible. At each node, a conditional test – i.e. a test with has either 'yes' or 'no' as the outcome – is passed which further splits the dataset (Figure 2; Quinlan, 1983). These splits are based on the Gini Impurity, which determines the goodness of the split: The lower the Gini impurity, the better a conditional statement would split and categorize the dataset into unambiguous results (Robnik-Šikonja, 2004). The leaf node is the final node of the branch and supplies the user with a prediction of the response variable. A decision tree on its own is often unstable and prone to overfitting, but in an ensemble such as the random forest, it can be very powerful.

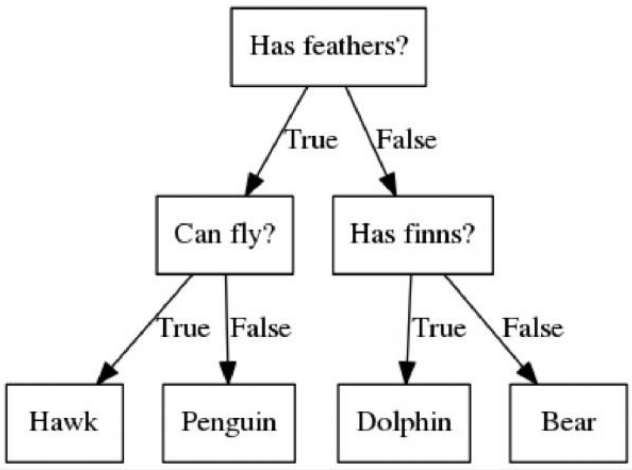

Figure 2: Example decision tree and the corresponding sample data. The response variable is the type of animal; the explanatory variables are whether the animal has feathers, can fly or has fins (Source: Towards AI, 2020).

The rationale behind using an ensemble of decision trees is to cancel out the effect of overfitting by an individual decision tree (Müller & Guido, 2016). To build a random forest model, many different decision trees must be created. The difference in each decision tree is caused by the randomness of the data used for training a decision tree and by randomly selecting explanatory variables for each test split. The data used for training each decision tree is *bootstrapped* from the total dataset. Assume we have data with 10 samples. Bootstrapping means that a random sample is drawn from the total dataset 10 times, which results in a new dataset of 10 samples. Some samples may be drawn multiple times, while other samples will not be drawn at all. Thereafter, a decision tree is built with a bootstrapped dataset. At each node, the algorithm only selects a pre-defined set of variables (instead of all variables) on which the node can be split – hence the randomness of the random forest. After all individual decision trees have been built, the random forest can start predicting: Each decision tree in the random forest makes

a prediction and for random forest regression the average of all predictions amount to the final prediction. For random forest classification, the class which has been predicted most often is chosen.

The main advantage of the random forest algorithm is that it overcomes the weaknesses of decision tree models, as it can prevent overfitting by introducing the randomness in how each decision tree is built. This is supported by the law of large numbers, which states that the average outcome of executing an experiment a large number of times is close to the expected outcome (Dekking et al., 2005). Concerning random forest modelling, a prediction becomes more accurate with a great number of decision trees. Moreover, because not all explanatory variables will be used in decision trees, multicollinearity between variables will not impair the performance of random forest models – in contrast to e.g. linear regression models (e.g. Tomaschek et al., 2018)

Another advantage of random forests is the calculation of the feature importance (Breiman, 2001). This is a score that explains how important an explanatory variable is to predict the response variable. However, the feature importance can be misleading, as they are biased towards explanatory variables with high cardinality (meaning they have many unique values). In contrast, features with low cardinality such as binary data are often perceived by feature importance as less important. An alternative to this is the permutation-based feature importance: This sensitivity procedure randomly shuffles data of one explanatory variable and tests how noticeably the model performance decreases (Breiman, 2001; Fisher et al., 2018). The shuffled variable has larger permutation importance if the model performance decreases strongly.

Another valuable tool is the partial dependency plot, which reveals the relationship between input variables and the response variable (Goldstein et al., 2015). The partial dependence must be interpreted as the marginal effect of an explanatory variable on the response variable when all other explanatory variables are kept constant (Friedman, 2001). Both the feature importance and partial dependency analysis support the user to get better insight into how random forest models function, which are often treated as black-box models.

Multiple studies have already been performed where trace gases or aerosols are modelled with random forests. For instance, Lu et al. (2020) investigated the feasibility of modelling nitrogen dioxide (NO_2) surface concentrations at a high spatial resolution with several machine learning algorithms. Inquired algorithms include several linear regression models and ensemble tree-based approaches such as random forests, stochastic gradient boosting (SGB), extreme gradient boosting (XGB). Predictors that have been used include road maps and industrial area land-use data, population counts, monthly average temperatures, wind speed, elevation, GEOS-CHEM NO_2 surface data and satellite data from OMI and TROPOMI.

In general, the ensemble tree-based models had a higher accuracy when compared to linear models in terms of root mean squared error (RMSE). Moreover, the tree-based models showed more spatial details compared to the linear models, as shown in Figure 3. The authors prefer the random forest model over XGB, as XGB models tend to create artefacts. In the feature importance analysis, it became clear that the TROPOMI VCD data and road map data had the highest feature importance.

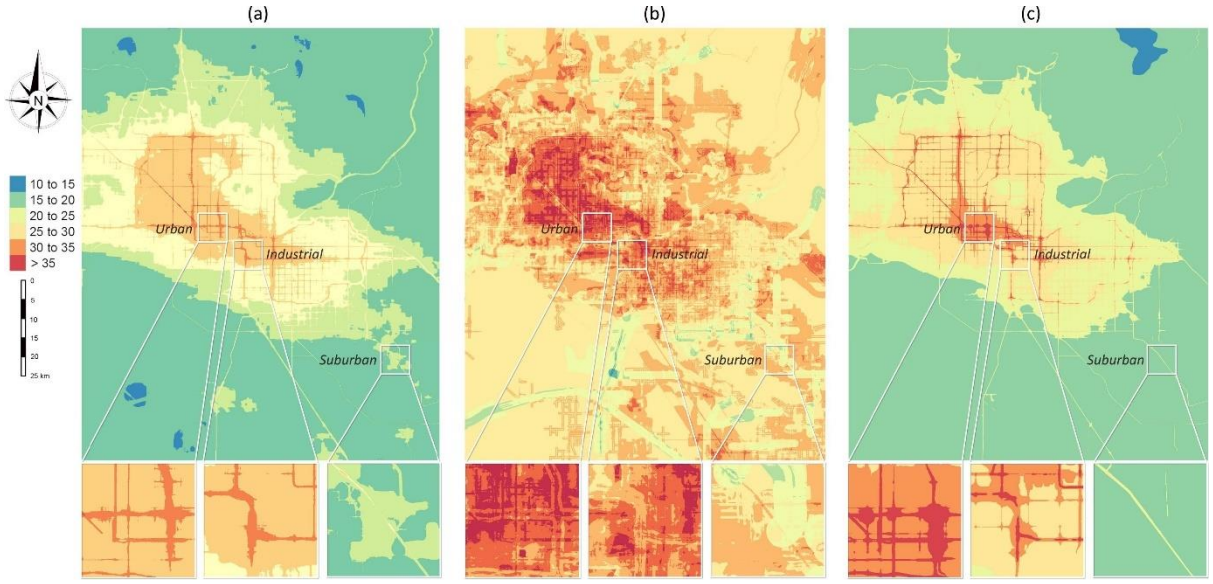

Figure 3: Annual mean prediction of NO_2 in µg m^{-3} in Phoenix, as modelled by a random forest model (a), gradient boosting machine (b) and Lasso (c) at a resolution of 25m × 25m (Source: Lu et al., 2020).

2.5: Objective and research question

Given that estimating the concentration, emission, deposition and transformation of ammonia in the atmosphere is a challenge, and that ammonia plays a central role in environmental degradation, the objective of this thesis is to use random forest machine learning to model surface ammonia concentrations by combining satellite data, meteorological data, and land-use data, to analyse how these variables influence the ammonia concentration. To the best of my knowledge, there has as of yet no research been done on modelling atmospheric ammonia concentrations based on random forests. The main research question in this thesis is as follows:

How can random forest models be applied with ammonia data from the CrIS satellite, meteorological variables and land-use based variables to improve the accuracy of ammonia surface concentration datasets, demonstrated by validation with ground-truth data?

Which can be subdivided into three sub-research questions:

1. Which of the explanatory variables used as input for the random forest model are the most important predictors for the surface ammonia concentration?
2. Which random forest model setup is most accurate when validated by ground-based measurement data from the LML- and MAN-network?
3. What is the added value of the CrIS vertical column density retrievals when estimating the surface ammonia concentration with random forest models?

The research question and the sub-research questions will be answered in the conclusion.

Section 3: Materials and methods

In this section, the data and methods used in this study will be discussed. Before going into detail on the methodology, the response variable and the explanatory variables that will be used for the random forest model are discussed. This is followed by a discussion of the standard algorithm used to acquire useful output for each experiment. Thereafter, the experimental setup will be discussed. Finally, the data that will be used for validation will be briefly discussed. The methodology must align with the following four elements:

- The problem: The biogeochemical nitrogen cycle is perturbed due to anthropogenic activities, which causes a wide range of detrimental effects on the environment and society.
- The knowledge gap: The lack of ground stations measuring atmospheric ammonia, the dependency of CTMs on emission inventory data and the moderately high uncertainty of the satellite observations by CrIS and IASI make it difficult to monitor the ammonia emission, deposition, transportation and transformation.
- The research objective: Construction of a random forest model which can model surface ammonia concentrations based on satellite, land-use and meteorological variables, and analyse how each input variable influences the ammonia surface concentrations.
- Research question: How can random forest models be applied with ammonia data from the CrIS satellite, meteorological variables and land-use based variables to improve the accuracy of ammonia surface concentration datasets, demonstrated by validation with ground-truth data?

3.1: Response variable

The response variable in this study is the LOTOS-EUROS ammonia surface concentration (in parts per billion) at CrIS overpass time (13:30 LST), which is approximately at 12:30 in the geographical domain of this study. Therefore, in this study, the time of 12:00 GMT+1 has been selected. This data is representative of the concentration in the surface layer at 2 meters height and incorporates emission and deposition processes that influence the surface concentration. The homogenous coverage of the LOTOS-EUROS model output ensures that ample data is available on which the random forest can train. Furthermore, the LOTOS-EUROS model output has been chosen as the ground-truth variable because the dataset was shown to be a relatively accurate representation (e.g. Ge et al., 2020).

It must be pointed out that in the optimal case, the RF model would be trained with in-situ measurements but due to a lack of ground-based measurement stations measuring ammonia at a high temporal resolution, the LOTOS-EUROS surface output is used for this. Also, it must be emphasized that the goal of this study is not to perfectly replicate the LOTOS-EUROS model but to create a new model that goes beyond the performance of the LOTOS-EUROS model when validated by ground-based measurements.

3.2: Explanatory variables

The candidate explanatory variables were all resampled to the boundaries and resolution of the LOTOS-EUROS model output. The target grid resolution is 0.05°× 0.025° or approximately 5.6 km × 2.8 km in horizontal and vertical direction respectively. Explanatory variables with different spatial resolutions were resampled with the GeoPandas overlay function to the target grid resolution (Jordahl et al., 2020). Moreover, all explanatory variables that have hourly data available, were set to match the

overpass time of the CrIS satellite as close as possible. As mentioned before, the time of 12:00 GMT+1 is selected.

The variables can be considered in three different categories: vertical column-based variables, meteorological variables and land-use based variables. Based on the atmospheric box-model framework discussed in Section 2.1.1, variables relevant to ammonia emission, deposition, chemical transformation and dispersion have been selected. The explanatory variables that will be used in various combinations in this study to predict the ammonia surface concentration are the following:

- LOTOS-EUROS vertical column density data
- Synthesized CrIS VCD of ammonia
- CrIS VCD of ammonia
- Surface temperature from LOTOS-EUROS
- Specific humidity from ECMWF
- Total precipitation from ECMWF
- Boundary layer height from ECMWF
- Windspeed on the horizontal plane in west-east and south-north direction from ECMWF
- Corine land cover map of 2018
- CAMS-REG v4.1 time profile map from manure application
- Cattle, pig and chicken density maps from the FAO

These input variables will be explained in greater detail in the subsection hereafter.

3.2.1: Vertical column-based variables

In this study, three different vertical column variables are used: LOTOS-EUROS VCD data, synthetic CrIS VCD data and VCD measurements from CrIS (in molecules cm^{-2}). Note that during the training phase of the model, only one of the three VCD datasets will be used as input.

The CrIS instrument has a sun-synchronous orbit and has two overpasses per day at 13:30 and 01:30 LST (Dammers et al., 2019). Although CrIS observations have been validated and show decent accuracy, lower ammonia column densities ($< 10^{16}$ molecules cm^{-2}) close to the detection limit are often overestimated (Dammers et al., 2017b). Not all CrIS observations can be used and therefore three filters have been applied to assure quality control (Shepherd et al., 2020). First, night-time observations have been filtered as these have not been validated yet. Moreover, observations with a quality flag equal to or lower than 2 and cloudy retrievals have been removed from the CrIS dataset as these retrievals are less reliable.

Besides actual CrIS VCD data, LOTOS-EUROS and synthetic CrIS VCD data have been created as follows: The LOTOS-EUROS pixel which is nearest to the satellite observations is selected. From twelve layers, the weight per layer is calculated and subsequently multiplied with the mole fraction of each matching layer to calculate the LOTOS-EUROS VCD. For the synthetic CrIS VCD, a systematic error and random noise term have been added to the LOTOS-EUROS VCD to simulate the behaviour of the CrIS satellite. For vertical column densities below 10^{16} molecules cm^{-2}, a term of 3.3×10^{15} molecules cm^{-2} with a standard deviation of 4.1×10^{15} molecules cm^{-2} has been added (Dammers et al., 2017b). For vertical column densities above 10^{16} molecules per cm^{-2}, a term of 0.4×10^{15} molecules cm^{-2} with a standard deviation of 5.3×10^{15} molecules cm^{-2} has been added.

3.2.2: Meteorological variables and land-use based variables

Five meteorological variables have been added to the random forest model to account for the influence they have on emission, deposition, transformation and transportation of ammonia (See Section 2.1.1). From the LOTOS-EUROS model, surface temperature data (in K) has been retrieved. From the European Centre for Medium-Range Weather Forecasts (ECMWF), total precipitation (in m), specific humidity in the surface layer (in kg kg^{-1}), boundary layer height (in m) and the wind speed (m s^{-1}) at 10 meters height on the horizontal plane in the west-east and south-north direction have been retrieved.

Furthermore, three types of land-use datasets have been used in this study as input in the random forest model. First, the Corine land cover map from 2018 has been used, which is a raster dataset with a 100m × 100m resolution. The land-use map consists of 44 different land-use types and is divided into five major categories: artificial surfaces, agricultural areas, forest and semi-natural areas, wetlands and water bodies (See Table A1 in appendix A for all land-use classes). For the sake of simplicity, these five Corine land-use classes will be used as input in this study. Although there is no specific land-use map for neither 2015, 2016 or 2017, it is assumed that the land-use has not changed significantly between these years and 2018. Because the output grid size of the experiments (approximately 5.6 km × 2.7 km) in this study has a considerably lower spatial resolution, the Corine land cover map is resampled to a lower spatial resolution. The resampled pixel value will be based on the most frequent land-use type that occurs in the new pixel boundary.

The second land-use based variable is time profile data of ammonia emissions from manure application. Time profile data accounts for the temporal variation in ammonia and is from the CAMS-REG v4.1 model (Kuenen et al., 2021). Finally, the third land-use dataset is gridded livestock density from the Food and Agricultural Organization (FAO, 2010). In this study, livestock density data of cattle, pigs and chickens (in the number of animals per km^2) will be supplied to the random forest model.

3.4: Machine learning algorithm

The experiments are done in the programming language of Python, and the "Sci-kit Learn" package is used to execute the random forest modelling (Pedregosa et al., 2011). To statistically validate the accuracy, the random forest model is split into a *training* subset and a *testing* subset. Using the complete dataset to train the model would be incorrect as the model cannot be validated with data that it has 'seen' before. In all experiments, the random forest model is trained in 2015 and 2016 and tested in 2017. These years have been selected due to data availability and because these were relatively normal in terms of meteorology and emission (e.g. neither long heatwaves such as in 2018 or 2019 nor potentially anomalous emissions patterns in 2020 during the COVID-19 pandemic).

Another important step is that satellite data must be filtered for NaN-values (Not-A-Number). Satellite datasets do not homogeneously cover the geographical domain and a NaN-value would be interpreted by the random forest algorithm as no ammonia in the atmosphere in a specific pixel. Therefore, all pixels without satellite observations are filtered.

Moreover, the land-use data is categorical (e.g. *pastures* or *fruit-and-berry plantations*) and must be pre-processed by one-hot-encoding because random forest algorithms only handle numerical data (Müller & Guido, 2016). The categorical variable is replaced with binary variables (0 or 1) for each individual land-use class.

The random forest consists of hyperparameters, which are parameters that influence the machine learning procedure (Müller & Guido, 2016). Through an algorithm, these hyperparameters can be optimized which can result in improved model performance. The hyperparameters that are optimized are the number of decision trees being built (*n_estimators*), the number of features used per decision tree (*max_features*), the maximum depth of each tree (*max_depth*), the minimal number of samples at which an internal node is split into another node (*min_samples_split*) and the minimal number of samples in a leaf node.

These hyperparameters are optimized with an algorithm called *Bayesian Optimization*. A simplified explanation of this process is that an external function assigns random values for the hyperparameters, builds the random forest model with the training subset and validates the model performance with the testing subset (Snoek et al., 2012). The external function remembers the outcome and searches further for the most optimal set of hyperparameters which provides the most accurate results. In this study, the optimization algorithms search and evaluate 20 sets of hyperparameters. This is less than the default setting of 100 iterations, to limit computational costs. In this study, the *gp_minimize* function by Scikit-optimize has been used for Bayesian optimization (Head et al., 2018)

For each iteration in the Bayesian optimization algorithm, *k-fold cross-validation* is conducted which splits the training subset into *k* parts. In this study, the number of folds is 4. The training subset will be split into four parts, and the model will always be trained with ¾ of the training subset and will be tested with the remaining part. This means that per optimization iteration, four different random forest models are built. The benefit of k-fold cross-validation is that splitting the training set in multiple folds is that it balances out potential folds that might influence the model performance *too positively* or *too negatively*. In other words, it will eliminate the possibility of training the model on a very hard subset and testing it on the easy-to-predict subset – or vice versa. The reason for the four folds in this study is because it assumed that enough data is available within the training set to cancel out possible biases between the training and testing subset.

After Bayesian Optimization, the random forest model is trained and fitted with the most optimal set of hyperparameters. Thereafter, the model can predict the daily surface ammonia concentration for each pixel. Because the satellite VCD data is one of the explanatory variables of the random forest model, the model can only predict at locations where satellite observations have been made.

Statistical metrics and figures are computed to assess the performance of the model. The root mean square error (RMSE), mean average error (MAE) and r are calculated to assess how well the model predicts the surface ammonia concentration when validated by ground-based measurement stations. The RMSE is a measure to explain the difference between the observed value and the predicted value, has the same unit as the response variable and is calculated as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y} - y_i)^2}{n}}$$

Besides, the normalised RMSE (NRSME) will be calculated to allow easy comparison between different model outputs. The NRMSE is calculated by dividing the RMSE by the mean of the observations:

$$NRMSE = \frac{RMSE}{\bar{y}}$$

In contrast to the RMSE, the MAE is a linear score, as all individual differences are weighted equally. This implies that the MAE measures the average difference between the observed and the predicted values. With the RMSE, the error is squared which implies that greater errors get a relatively higher weight. Together, the RMSE and the MAE can identify the variation in the errors in the dataset. The RMSE will always be equal to or greater than the MAE. If the RMSE and the MAE are equal, the average error is equal over all predictions. The MAE is calculated as follows:

$$MAE = \frac{\sum |\hat{y} - y_i|}{n}$$

The correlation r represents the strength of a relationship between two variables. In contrast to the RMSE, the r is measuring the correlation and the rule is that the higher r is, the better the fit. The r is calculated as follows:

$$r = 1 - \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}}$$

Finally, a linear regression model will be fitted to analyse the relationship between the predicted random forest model values and the ground-based measurements. The residual standard error (RSE) will be calculated, which measures the standard deviation of the residuals (the difference between the prediction and the linear regression model). The lower the RSE, the lower the noise around the linear regression model. The RSE is calculated as follows:

$$RSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n - 2}}$$

To get more information about the utility of the explanatory variables, the permutation importance will be calculated. The permutation importance will be calculated instead of the feature importance because both low- and high cardinality data are used as input variables. The permutation importance can be calculated with the *permutation_importance* function from Scikit-learn (Head et al., 2018). Moreover, partial dependency plots will be produced for each random forest model to analyse how each variable influences the surface concentration.

To summarize, the following steps are followed for each experiment:

1. The response variables and explanatory variables are loaded.
2. Data is filtered to only consist of pixels where satellite observations have been made.
3. Land-use data is transformed into dummy variables.
4. The dataset is split into a training subset with data from 2015 and 2016 and a testing subset with data from 2017.
5. With Bayesian Optimization, the most optimal hyperparameters are determined for the random forest function.
6. The random forest model is built with the most optimal hyperparameters.
7. The random forest model predicts the surface ammonia concentrations for each day in 2017.

8. Statistical metrics are computed and figures comparing the observed and predicted surface ammonia concentrations are created.
9. The permutation importance of each feature is assessed.
10. A partial dependency plot will be created to assess how each explanatory variable influences the response variable.

3.5: Experiments

The geographical boundaries remain the same in all the experiments and cover all of the Netherlands, and parts of Germany, Belgium and France (Figure 4). The longitudinal boundaries are 3.175° – 7.475° and the latitudinal boundaries are 50.6625° - 53.6875°. The target grid size is 0.05°× 0.025°, (horizontal and vertical distance respectively). The reason why this particular domain has been chosen is due to the large number of ground observations from the LML and MAN network available for validation.

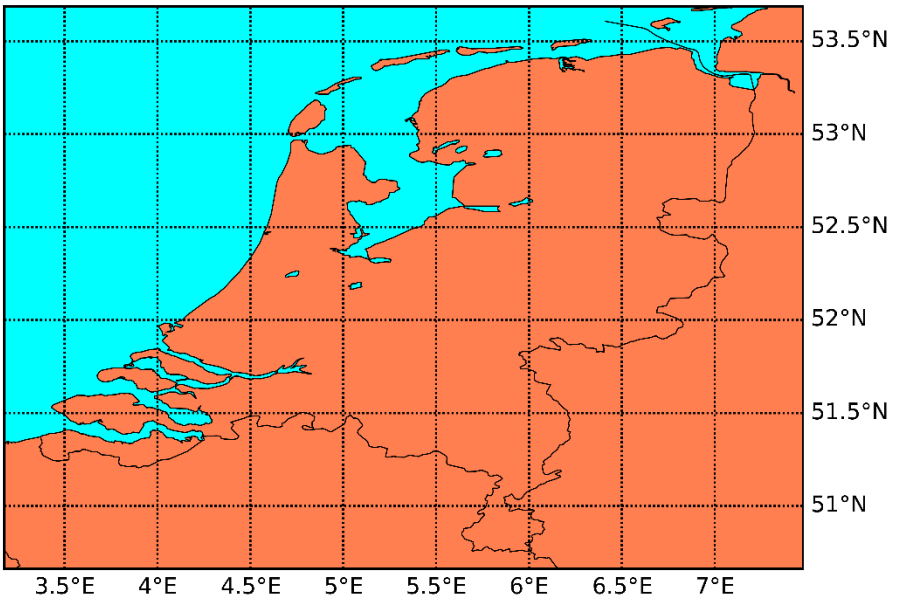

Figure 4: The geographical boundaries of this study.

In total, three main random forest models are built (see Table 1). Each random forest uses the LOTOS-EUROS surface concentration as the response variable and the explanatory features described in section 3.2. However, each of the three random forest models is trained with a different VCD dataset: either the LOTOS-EUROS VCD, the synthetic CrIS VCD or the actual CrIS VCD. The LOTOS-EUROS VCD dataset is used to train and test a random forest model with the most optimal combination of variables (because both the explanatory variable and response variable come from the same model). The synthetic CrIS values are used to mimic the CrIS satellite data as closely as possible. From this, we can analyse the influence on the outcome of the random forest when adding noise to the original LOTOS-EUROS VCD dataset. Finally, the real CrIS VCD dataset is used to analyse whether the satellite observations can already be used in random forest modelling.

After the random forest models have been fitted, they will predict the surface concentration in 2017 with the three different VCD datasets. Nine results will be computed in total by the random forest models and will be validated by the LML network and the MAN network (Table 1). Besides the three main random forest models, a fourth random forest model is trained without VCD data. This is done to understand the potential added value of VCD data when estimating the ammonia surface concentrations.

The rationale behind this experimental setup is to understand which combination of VCD dataset used for training and VCD dataset used as input during the prediction phase leads to the most accurate results when validated by ground-based measurements. A RF model is trained with LOTOS-EUROS VCD data as a baseline to understand how the RF algorithm would perform when supplied with the most optimal dataset (given that the data of the response variable is also from the LOTOS-EUROS model). Incorporating synthetic or real CrIS VCD data could add new information about the spatiotemporal distribution of the ammonia surface concentration because the CrIS satellite shows a different distribution of ammonia through time and space when compared to LOTOS-EUROS. Van der Graaf et al. (2022) discuss this phenomenon and showed that the CrIS surface concentrations are lower than LOTOS-EUROS during the spring and higher from June to October.

It can be expected that the results from RF models 2 and 3 which have been supplied with LOTOS-EUROS VCD data during the training phase will not lead to meaningful outcomes. However, for the sake of completeness, the results of all nine combinations (i.e. RF models 1 to 3, which all predict the ammonia concentration with the three different input VCD datasets) have been added in this report to portray better how the RF models work when supplied with different VCD datasets.

Table 1: Summary of the conducted experiments in this study. The 10 results in total are named as followed: RF is followed by the experiment number. Thereafter it is followed with the name of the VCD dataset which is used for the prediction phase.

		VCD dataset used as input for prediction phase		
RF no/ Exp. no	**RF model trained with**	LOTOS-EUROS VCD	Synthetic CrIS VCD	CrIS VCD
1	LOTOS-EUROS VCD	RF1-LE	RF1-Synth	RF1-CrIS
2	Synthetic CrIS VCD	RF2-LE	RF2-Synth	RF2-CrIS
3	CrIS VCD	RF3-LE	RF3-Synth	RF3-CrIS
4	No VCD data	RF4		

3.6: Validation of results

The outcome of the random forest results will be validated by the LML network (Landelijk Meetnet Luchtkwaliteit) and the MAN-network (Meetnet Ammoniak in Natuurgebieden). The LML network consists of six stations in the Netherlands which measure the hourly ammonia concentration with miniDOAS instruments (Berkhout et al., 2017). These six instruments are active differential optical absorption instruments and are stationed at low, medium and high ammonia emission areas. The measurements from the Vredepeel site will not be included for validation in this study since these are not reliable enough. The MAN-network provides monthly averaged ammonia concentrations in more than 82 low-emission natural areas with Gradko passive samplers(Lolkema et al., 2015). The measurements of MAN are calibrated by the more accurate measurements of the LML stations. In this study, 266 MAN measurement stations are used for validation.

The LOTOS-EUROS model output at overpass time will also be validated with the LML and MAN datasets. From this, there can be concluded whether the RF model performs better than the LOTOS-EUROS dataset. Additionally, CrIS surface retrieval data will also be validated by LML and MAN and serve as a reference.

Section 4: Results and discussion

In this section, the results of the four random forest (RF) models will be explained objectively by comparing the yearly averaged RF output of 2017 to the yearly averaged LOTOS-EUROS output. In Appendix B, the scatterplots which compare the LOTOS-EUROS model output with the RF model output are shown. Moreover, the permutation importance of each variable is reported. By calculating the permutation importance of each explanatory variable, the sensitivity of the model per explanatory variable is determined. Thereafter, the output of the RF models will be validated with LML and MAN-network measurements. Furthermore, in the partial dependency analysis, the marginal effect of each explanatory variable on the response variable (the surface ammonia concentration) is assessed. Finally, the results and limitations of the experiments will be discussed. Moreover, the recommendations for further research will be explained.

4.1.1: Experiment 1. Random forest model 1 trained with LOTOS-EUROS VCD

For experiment 1, the RF model has been trained with LOTOS-EUROS VCD data (Figure 5). The RF model with LOTOS-EUROS VCD as input closely resembles the actual LOTOS-EUROS model, although it seems to underestimate the ammonia concentration in high emission areas (e.g. Gelderse Vallei, east Noord-Brabant and the Dutch-German border). When the RF model is supplied with synthetic and real CrIS VCD data, the surface concentrations are much higher in most regions when compared to the LOTOS-EUROS model output.

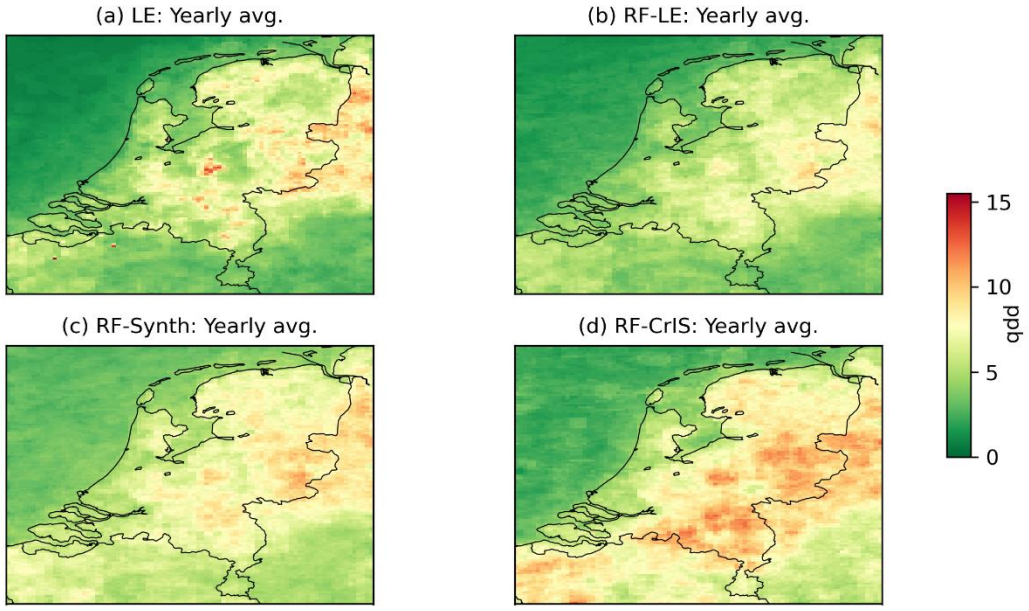

Figure 5: The yearly averaged output of the RF model in 2017 which has been trained with the LOTOS-EUROS VCD on data of 2015 and 2016. (a) The LOTOS-EUROS model output. (b) The RF model with LOTOS-EUROS data as input VCD. (c) The RF model with synthetic CrIS VCD as input VCD. (d) The RF model with CrIS VCD as input VCD. The unit of the surface concentration is in parts per billion (ppb).

The most dominant explanatory variable in RF model 1 is VCD data (Figure 6), followed by the boundary layer height, the CAMS-reg time profiles for manure application and cow density data.

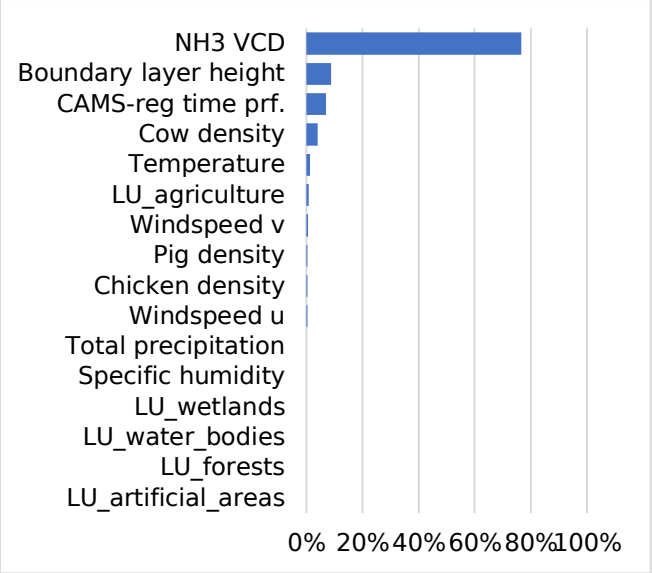

Figure 6: Permutation importance of the explanatory variables used as input for RF model 1. The permutation importance of the variables has been standardized to add up to 100%.

It is worth bearing in mind that the resolution of the satellite footprint datasets is lower than the resolution of the LOTOS-EUROS surface dataset. Therefore, it can be expected that the predictions by the RF model have less spatial detail. This explains the underestimation by the RF-LE model in high emission areas such as de Gelderse Vallei, east North-Brabant and the Dutch-German border.

4.1.2: Experiment 2. Random forest model 2 trained with synthetic CrIS VCD data

For experiment 2, the RF model has been trained with synthetic CrIS VCD data. In Figure 7, the LOTOS-EUROS output is shown next to the output of the RF model. The RF models which have been supplied with LOTOS-EUROS and synthetic CrIS VCD data as input VCD generally underestimate the surface ammonia concentration when compared to the LOTOS-EUROS output. Similarly to experiment 1, when the trained RF model is supplied with real CrIS VCD data, the ammonia concentration increases significantly in most regions. Nevertheless, the peak concentration levels in the Gelderse Vallei cannot be estimated well by the RF model.

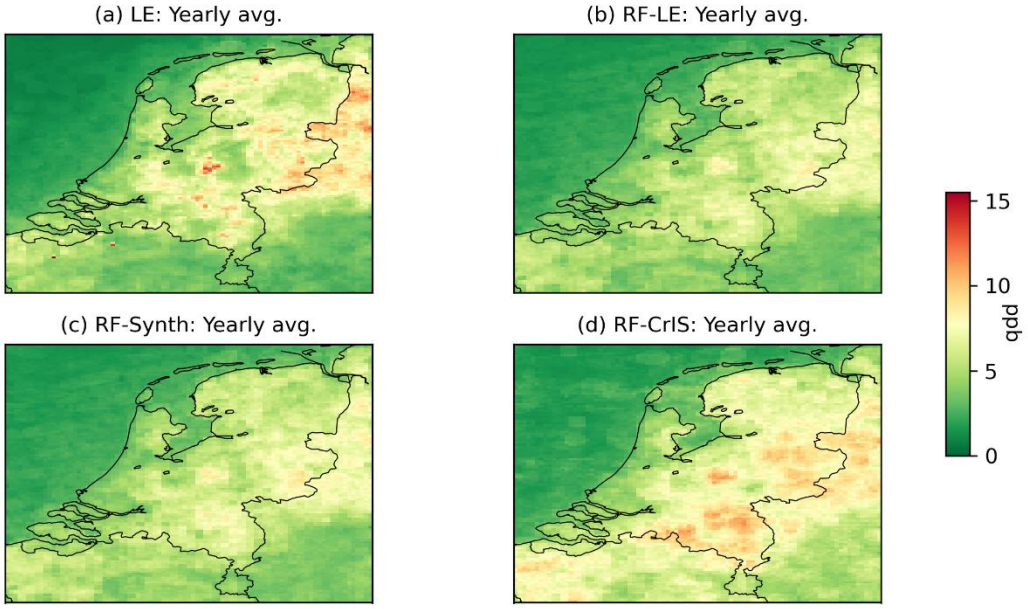

Figure 7: The yearly averaged output of the RF model in 2017 which has been trained with synthetic CrIS VCD data in the years 2015 and 2016. (a) The LOTOS-EUROS model output. (b) The RF model with LOTOS-EUROS data as input VCD. (c) The RF model with synthetic CrIS VCD as input VCD. (d) The RF model with CrIS VCD as input VCD. The unit of the surface concentration is in parts per billion (ppb).

The permutation importance of most important explanatory variables in RF model 2 (Figure 8) is almost similar to RF model 1. Compared to RF model 1, the permutation importance drops from 77% to 57%. This is likely caused by the systematic and random error terms which have been added to the LOTOS-EUROS VCD. Additionally, the CAMS-reg time profile for manure application and boundary layer height variables are more significantly present in RF model 2 with a permutation importance of 16% and 13% respectively.

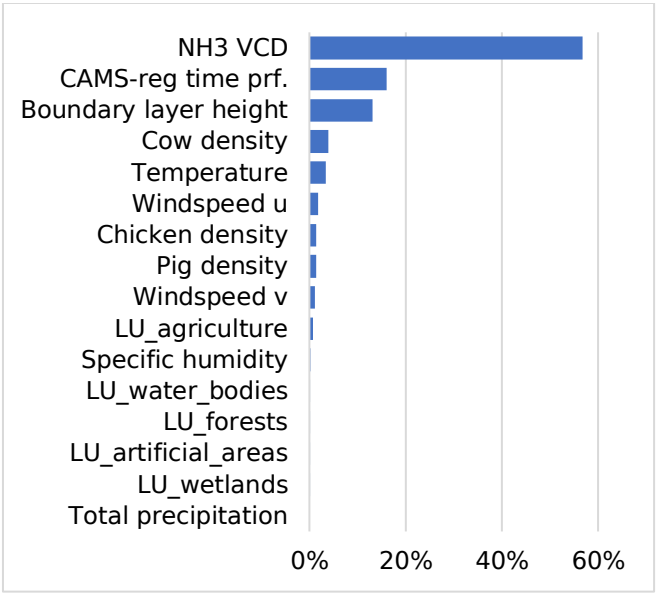

Figure 8: Permutation importance of the explanatory variables used as input for RF model 2. The permutation importance of the variables has been standardized to add up to 100%.

4.1.3: Experiment 3. Random forest 3 model trained with CrIS VCD data

For experiment 3, the RF model has been trained with CrIS VCD data (Figure 9) It is evident that there is very little difference in the RF model output when supplied with either LOTOS-EUROS VCD,

synthetic CrIS VCD or real VCD data. Although the RF model can locate areas of medium-to-high ammonia emission, it systematically underestimates the ammonia surface concentration in these regions. Moreover, it seems to overestimate the ammonia surface concentration in Flanders.

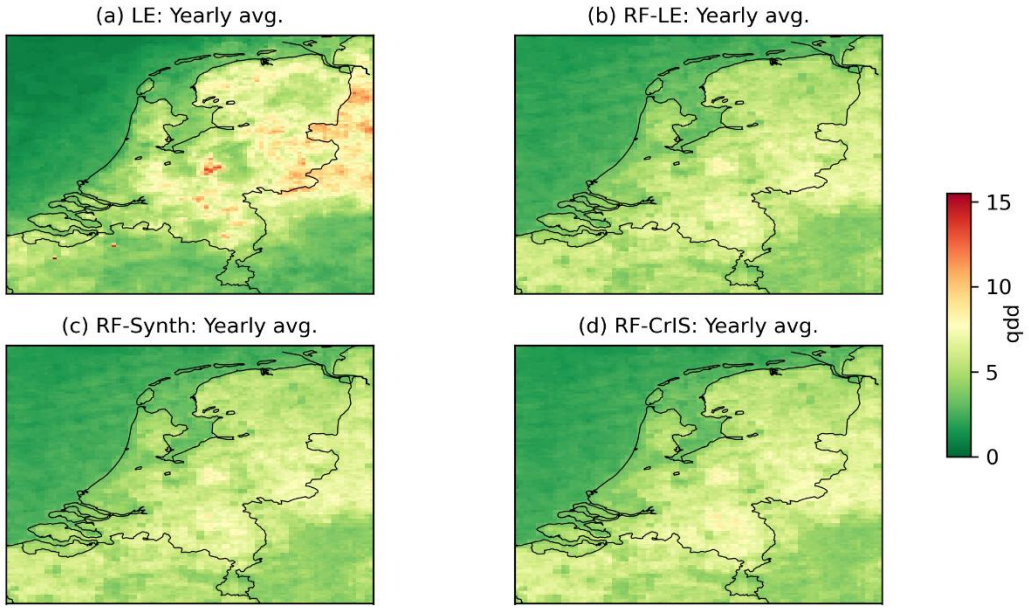

Figure 9: The yearly averaged output of the RF model in 2017 which has been trained with CrIS VCD data in the years 2015 and 2016. (a) The LOTOS-EUROS model output. (b) The RF model with LOTOS-EUROS data as input VCD. (c) The RF model with synthetic CrIS VCD as input VCD. (d) The RF model with CrIS VCD as input VCD. The unit of the surface concentration is in parts per billion (ppb).

The permutation importance of the variables reported in RF 3 (Figure 10) vary significantly with the previous two models. Where the VCD data in RF models 1 and 2 had an importance of more than 60%, it is reduced in this RF model to just 3%. This also explains why the RF model output does not change significantly when providing it with different VCD datasets. With 63%, the CAMS-reg time profiles for manure application has the highest importance, followed by meteorological variables such as windspeed in the west-east direction, surface temperature and boundary layer height.

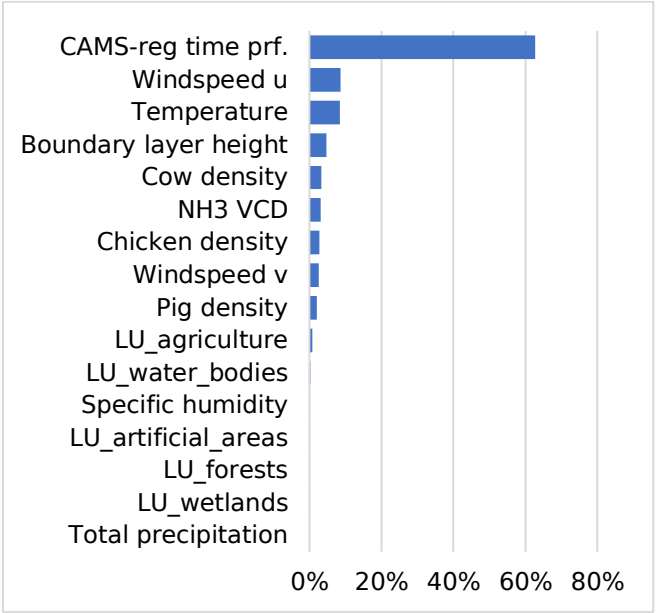

Figure 10: Permutation importance of the explanatory variables used as input for RF model 3. The permutation importance of the variables has been standardized to add up to 100%.

4.1.4: Experiment 4: Random forest model 4 trained without VCD data

For experiment 4, the RF model has been trained without VCD data. In Figure 11, the output of the LOTOS-EUROS model and the RF model is shown. The RF closely resembles the RF output of experiment 3: It can locate the medium-to-high emission areas but does underestimate the surface concentration in these regions. Also, it overestimates the surface ammonia concentration in Flanders.

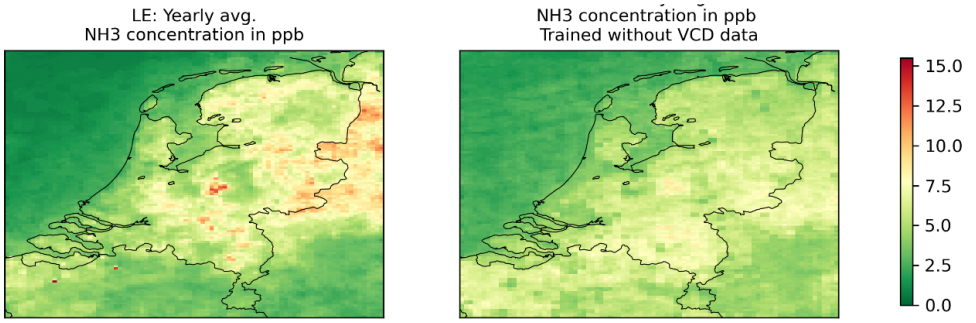

Figure 11: The yearly averaged output of the RF model in 2017 which has been trained with CrIS VCD data in the years 2015 and 2016. (a) The LOTOS-EUROS model output.

The permutation importance of the explanatory variables in Figure 12 show similar scores as the permutation importance of the variables reported in RF model 3: The CAMS-reg time profiles or manure application has an importance of ~60% and is followed to a lesser extent by variables such as temperature, windspeed and boundary layer height.

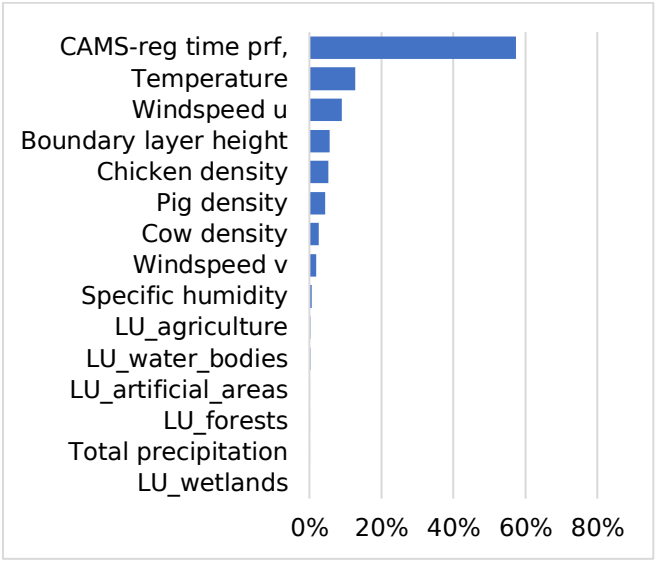

Figure 12: Permutation importance of the explanatory variables used as input for RF model 4. The permutation importance of the variables has been standardized to add up to 100%.

4.2: Validation of results with LML

In this subsection, the results of experiments 1, 2, 3 and 4 will be compared to the LML network measurements in chronological order. As a reference, the original LOTOS-EUROS model output and CrIS surface retrievals will also be validated with the LML-measurements The validation will be done by comparing the monthly average output of each RF model, with the monthly average concentration of the LML-station. The most nearby pixel of the RF model is matched with the location of each LML-station. Moreover, the monthly-averaged value for each station will be calculated from the ammonia

concentration at CrIS overpass time – which is approximately at 12:00. The daily validation of the RF output with the LML network can be found in Appendix C. Finally, the measurements from the Vredepeel station will not be included in this study because it is an unreliable station to use for validation.

For the sake of clarity, the results are referred to by their experiment number and the VCD dataset which has been used as input during the prediction phase (See Table 1 for naming convention).

4.2.1: Validation of results of experiment 1 with LML-measurements

First, the results from RF model 1 will be validated against the LML measurements (Figure 13). This RF model has been provided with LOTOS-EUROS VCD data during the training phase. Result RF1-CrIS has the smallest RMSE of 5.73 ppb – meaning the RF output varies 5.73 ppb on average with the LML measurements. RF model RF1-LE has the highest RMSE of 6.91 ppb. The correlation between the RF model output and the LML measurements is the highest for the RF1-CrIS model with $r = 0.52$ and lowest for the RF1-Synth model with $r = 0.37$.

When analysing the performance per station, RF1-LE generally reports moderate-to-high correlation coefficients in stations Wekerom, Valthermond, de Zilk and Zegveld. In RF1-Synth, the performance per station drops significantly, although stations de Zilk and Wekerom still have moderate-to-high correlation coefficients. In RF1-CrIS, stations Valthermond, Wekerom and Zegveld have moderate-to-high correlation coefficients while the correlation with de Zilk is significantly lower when compared to the other results. This seems to be caused by a strong outlier. In all three results, the station of Wieringerwerf does not correlate with the RF-output.

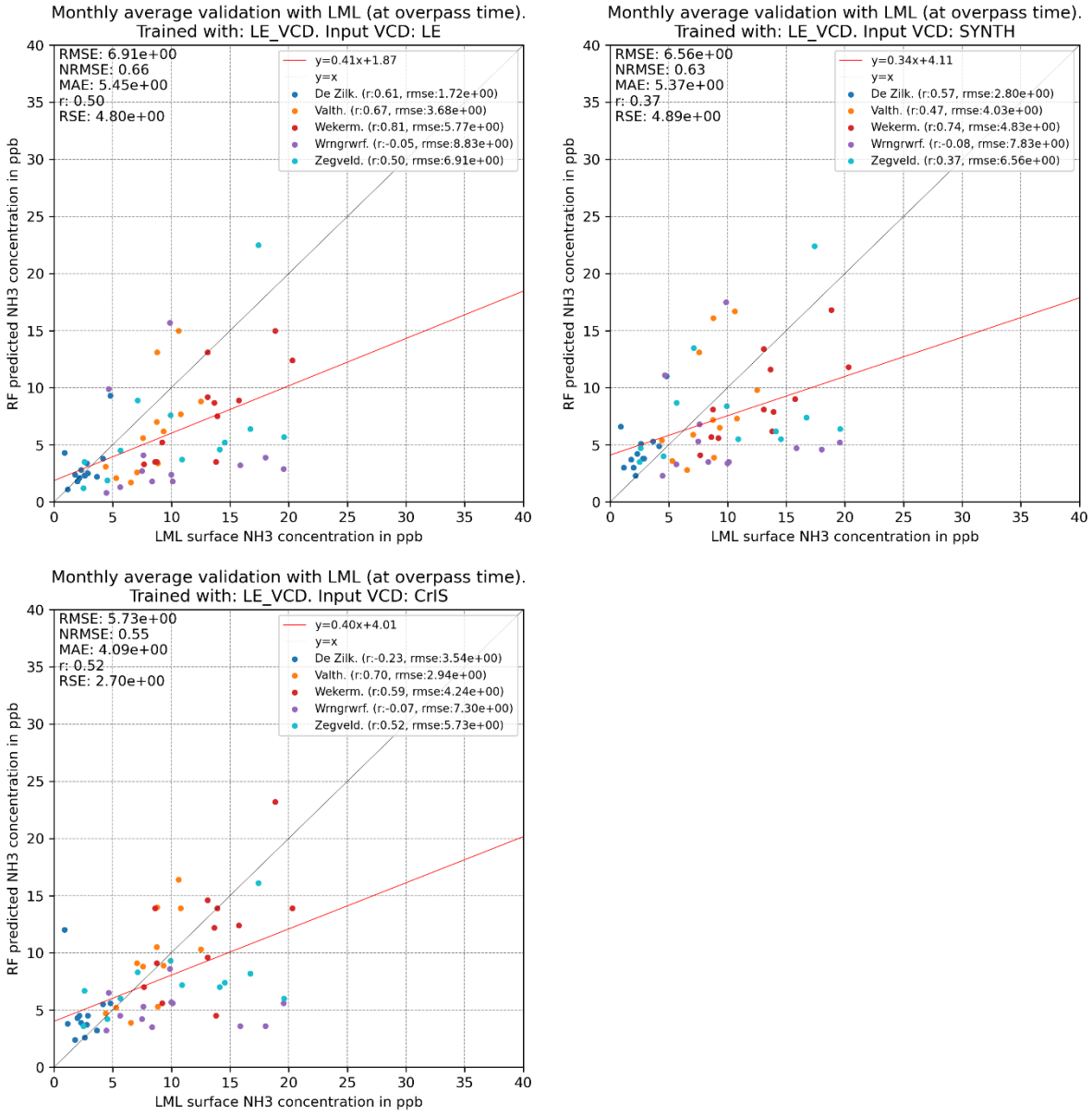

Figure 13: The monthly-averaged validation of the output of RF model 1 with LML. The red line depicts the fitted linear regression model. The grey line is the 1:1 line. Each colour depicts a different station.

4.2.2: Validation of results of experiment 2 with LML-measurements

Second, the results from RF model 2 will be validated against the LML-measurements (Figure 14). This RF model has been provided with synthetic CrIS VCD data during the training phase. From the three results, RF2-CrIS has the lowest RMSE of 6.57 ppb and the highest correlation of 0.50. RF2-LE has the highest RMSE of 7.26 ppb and RF2-Synth has the lowest correlation of 0.46.

When analysing the performance per station, stations Wekerom, De Zilk and Valthermond have the highest r for RF2-LE and RF2-Synth. For RF2-CrIS, only Valthermond, Wekerom and Zegveld have moderate-to-high correlation coefficients. There is no correlation between RF2-CrIS and de Zilk, which seems to be caused by an extreme outlier. Similarly to experiment 1, Wieringerwerf does not correlate when compared to the RF model output.

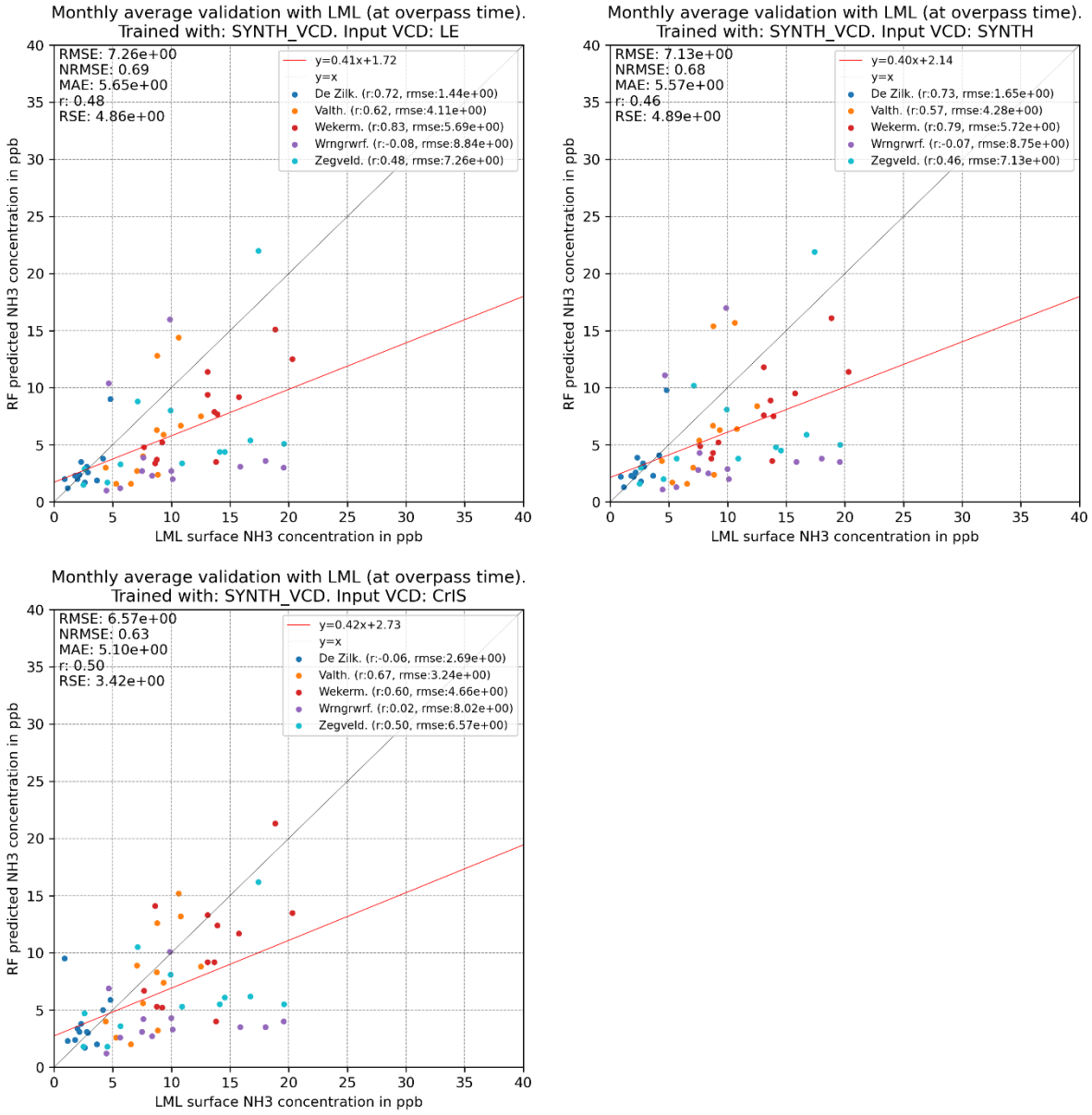

Figure 14: The monthly-averaged validation of the output of RF model 2 with LML. The red line depicts the fitted linear regression model. The grey line is the 1:1 line. Each colour depicts a different station.

4.2.3: Validation of results of experiment 3 with LML-measurements

Third, the results from RF model 3 will be validated against the LML-measurements (Figure 15). This RF model has been provided with CrIS VCD data during the training phase. From the three results, RF3-CrIS has the lowest RMSE of 6.89 ppb while RF3-LE has the highest RMSE of 7.08 ppb. Moreover, RF3-CrIS also has the highest correlation of 0.50, while the other two results have a correlation of 0.47.

Similarly to the previous two random forest models, the correlation between the RF model and the LML-measurements are the highest for Wekerom, Valthermond and De Zilk. In Zegveld the correlation is moderate and in Wieringerwerf there is no correlation with the RF model output.

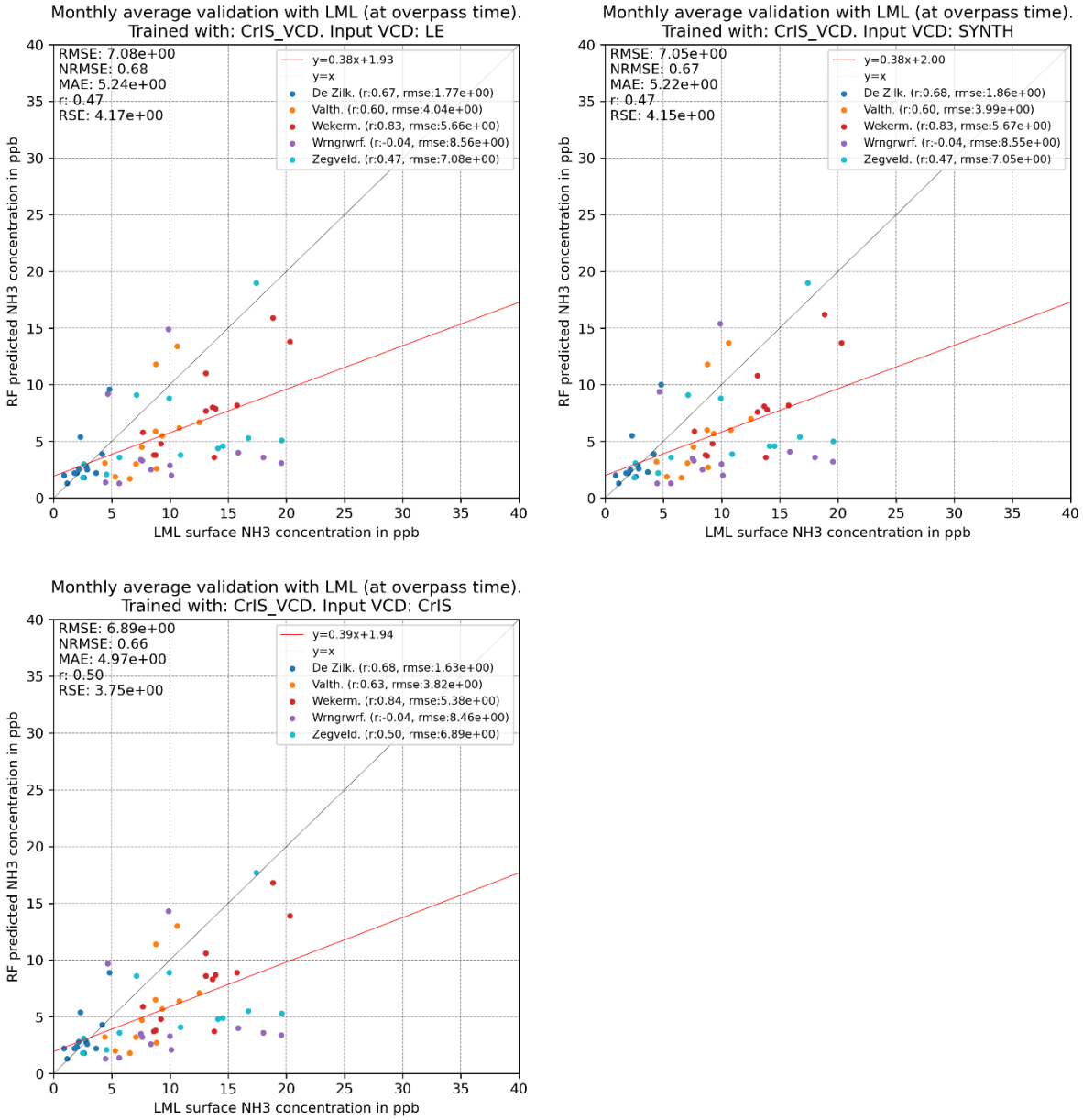

Figure 15: The monthly-averaged validation of the output of RF model 2 with LML. The red line depicts the fitted linear regression model. The grey line is the 1:1 line. Each colour depicts a different station.

4.2.4: Validation of results of experiment 4 with LML-measurements

Fourth, the results from RF model 4 will be validated against the LML-measurements (Figure 16). This RF model has not been provided with VCD data during the training phase and does not require VCD data during the prediction phase. The RMSE of the RF output of experiment 4 is 6.88 ppb and the correlation between the monthly average LML-measurements and the RF model output is 0.46. Similarly to the previous experiments, the RF model is most accurate at the stations Valthermond, de Zilk and especially Wekerom with medium-to-high correlations.

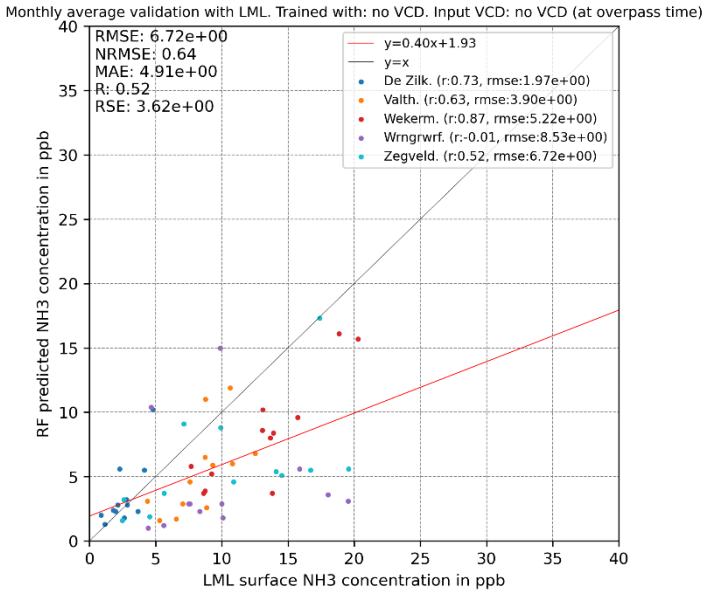

Figure 16: The monthly-averaged validation of the output of RF model 4 with LML. The red line depicts the fitted linear regression model. The grey line is the 1:1 line. Each colour depicts a different station.

4.2.5: Validation of LOTOS-EUROS model output as CrIS surface retrieval with LML

In this section, LOTOS-EUROS and CrIS surface retrievals from 2017 will be validated by the LML-measurements to serve as a reference to the validation results of experiments 1 to 4 (Figure 17). When validated by the LML-measurements, the LOTOS-EUROS model has a RMSE of 7.11 ppb and a moderate correlation of 0.48. Stations such as the Zilk, Wekerom and Valthermond have the highest correlation when validated while Zegveld has a moderate correlation and Wieringerwerf has no correlation.

The CrIS surface retrievals have a RMSE of 9.59 ppb and a moderate-to-high correlation of 0.61 when validated by the LML-measurements. In opposition to previous experiments, all stations (except Wieringerwerf) have similar correlations between 0.49 and 0.66.

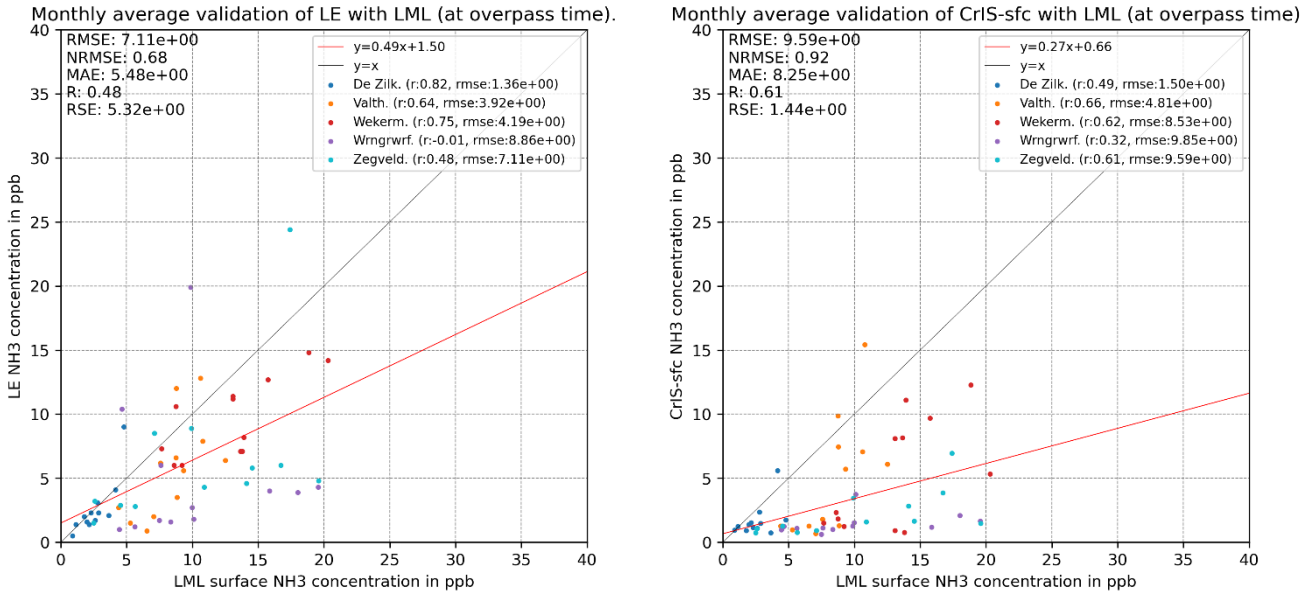

Figure 17: The monthly averaged validation of the LOTOS-EUROS ammonia surface concentrations (left) and the CrIS surface retrievals (right) with LML measurements. The red line depicts the fitted linear regression model. The grey line is the 1:1 line. Each colour depicts a different station.

4.2.6: Interim conclusions on the validation of random forest models 1-4 by LML-measurements

In this subsection, the results from the validation of the random forest models by the LML-measurements will be briefly summarized and discussed. In Table 2, the results of the validation by the LML-measurements are shown.

The CrIS surface retrievals have the highest correlation of 0.61 with the LML-measurements, which is a moderately high correlation. In return, it also has the highest RMSE of 9.59 ppb. The RF1-CrIS model output has the lowest RMSE of 5.98 ppb and also has the second-highest correlation coefficient. As a reference, the LOTOS-EUROS model output has a RMSE and correlation of 7.11 and 0.48 respectively, when validated by the LML measurements. From this, there could be concluded that the RF model has a better performance than the LOTOS-EUROS model at overpass time. However, this validation is done at just five locations and with such sample sizes, there cannot be inferred with certainty that the RF model is generally better than the LOTOS-EUROS model.

Finally, when comparing the validation results of RF model 4 which has been trained without VCD data to the other RF models which have been trained with VCD data, there can be concluded that adding VCD data to the RF models has added value. The performance of RF model 4 with a RMSE and correlation of 6.88 ppb and 0.46 respectively, is significantly less than most RF models which have been trained with VCD data.

Table 2: The RMSE and correlation of the RF experiment output, LOTOS-EUROS model output and CrIS surface retrievals when validated by LML-measurements. The best results are marked in green.

Experiment	RMSE (ppb)	r	Experiment	RMSE (ppb)	r
RF1-LE	6.91	0.50	RF3-LE	7.08	0.47
RF1-Synth	6.56	0.37	RF3-Synth	7.05	0.47
RF1-CrIS	**5.73**	0.52	RF3-CrIS	6.89	0.50
RF2-LE	7.26	0.48	RF4 (No VCD)	6.88	0.46
RF2-Synth	7.13	0.46	LOTOS-EUROS	7.11	0.48
RF2-CrIS	6.57	0.50	CrIS surface retr.	9.59	**0.61**

Moreover, when analysing the performance per station, stations such as de Zilk, Valthermond and Wekerom – which are low-, medium- and high-emission areas respectively – have a moderate-to-high correlation in most experiments (Table 3). Moreover, Zegveld only has a moderate correlation in most experiments and Wieringerwerf has almost no correlation in most experiments. When comparing the validation of the RF output with the LOTOS-EUROS model output, the RF model performs significantly better at the Valthermond and Wekerom sites and slightly better than at the Zegveld site. Only at de Zilk, the LOTOS-EUROS output is more accurate. From this analysis per station, we can again conclude that the RF model performs better at these locations than the LOTOS-EUROS model. However, we must again be careful with such conclusions because the validation is performed at just five stations.

Table 3: The correlation coefficients of each station when validated by the LML-measurements for the different experiments. The result with the best correlation per station is marked in green. The bold black line shows the demarcation between the RF experiments and the LOTOS-EUROS model output and CrIS surface retrievals which serve as a reference.

Experiment	De Zilk	Valthermond	Wekerom	Wieringerwerf	Zegveld
RF1-LE	0.61	0.67	0.81	-0.05	0.50
RF1-Synth	0.57	0.47	0.74	-0.08	0.37
RF1-CrIS	-0.23	**0.70**	0.59	-0.07	0.52
RF2-LE	0.72	0.62	0.83	-0.08	0.48
RF2-Synth	0.73	0.57	0.79	-0.07	0.46
RF2-CrIS	-0.06	0.67	0.60	0.02	0.50
RF3-LE	0.67	0.60	0.83	-0.04	0.47
RF3-Synth	0.68	0.60	0.83	-0.04	0.47
RF3-CrIS	0.68	0.63	**0.84**	-0.04	0.50
RF4 (No VCD)	0.62	0.63	0.82	-0.00	0.46
LOTOS-EUROS	**0.82**	0.64	0.75	-0.01	0.48
CrIS surface retr.	0.49	0.66	0.62	**0.32**	**0.61**

4.3: Validation of results with MAN

In this subsection, the results of experiments 1, 2, 3 and 4, and LOTOS-EUROS output and CrIS surface retrievals will be validated by the MAN network. The MAN measurements are compared to the yearly averaged ammonia concentration of each RF model at the matching pixel of the station location. For the sake of clarity, the results are referred to by their experiment number and the VCD dataset which has been used as input during the prediction phase (See Table 1 for naming convention). The scatterplots will also show in which month the observations have been done to acquire more information about how the RF model performs throughout different seasons. The yearly averaged validation plots of the RF experiments and the performance of each model per month can be found in Appendix D.

4.3.1: Validation of results of experiment 1 with MAN-measurements

In the first experiment, the RF model has been trained with LOTOS EUROS VCD data. When validated by the MAN measurements, the RF1-CrIS has the lowest RMSE of 4.21 ppb (Figure 18) and the highest correlation of 0.47. Furthermore, the RF1-LE model has the highest RMSE of 4.49 ppb and the RF1-Synth output has the lowest correlation of 0.38. Models RF1-LE and RF1-Synth have a fairly similar outcome and show that the RF models overestimate the ammonia concentration in February and March and underestimate the concentration in October, November and December. The RF1-CrIS model output differs from the two aforementioned models: Although in February and March the RF1-CrIS model still overestimates the ammonia concentration, the concentration is estimated more accurately between March and July (See Table D1)

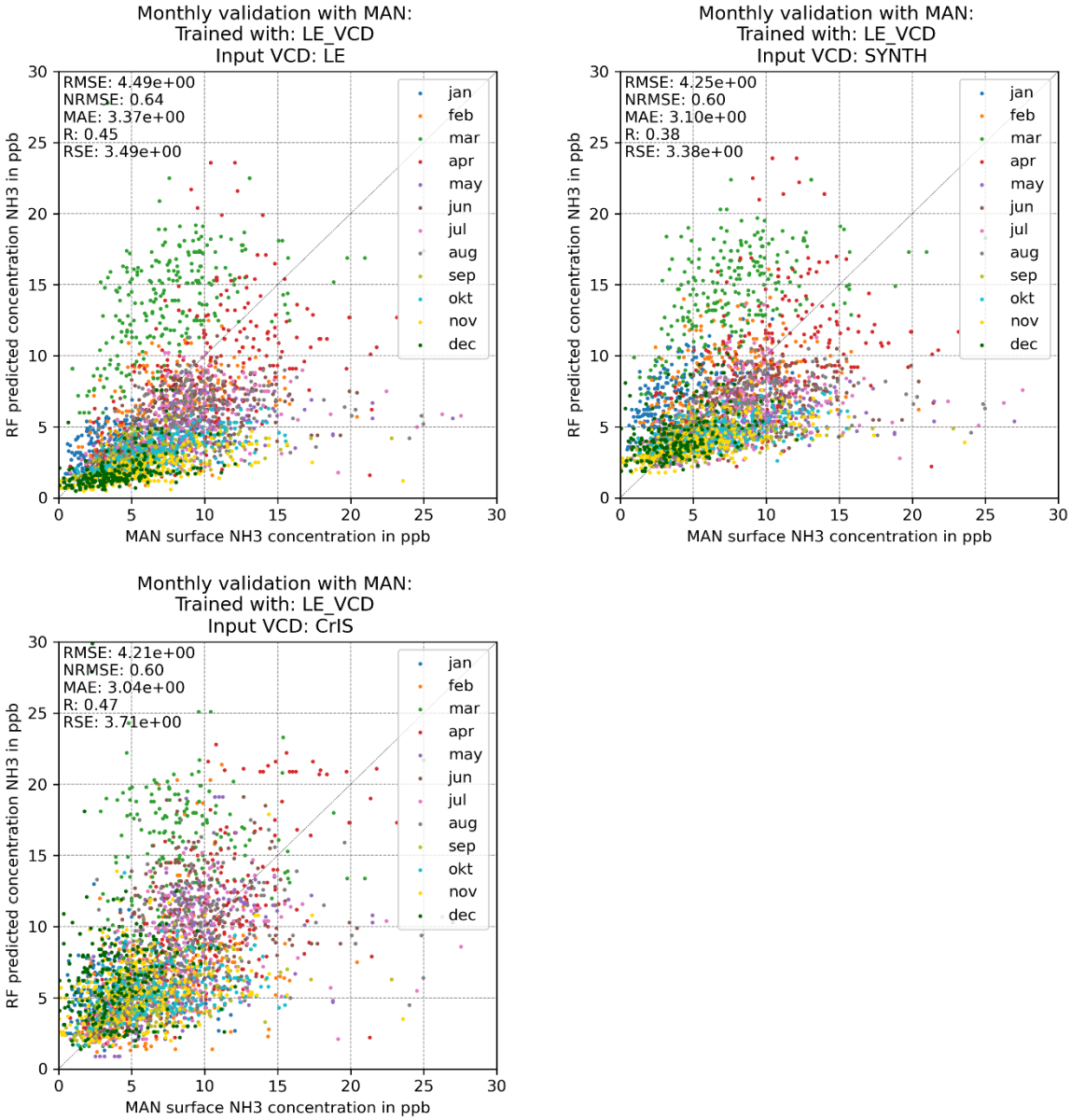

Figure 18: The monthly averaged validation of the output of RF model 1 with MAN measurements. The grey line depicts the 1:1 line and each colour depicts another month.

4.3.2: Validation of results of experiment 2 with MAN-measurements

In the second experiment, the RF model has been trained with synthetic CrIS VCD data. When validated by the MAN measurements, the RF2-CrIS has both the lowest RMSE of 4.12 ppb and the highest correlation of 0.50 (Figure 19). Moreover, the RF2-LE model has the highest RMSE of 4.54 ppb and the RF2-Synth has the lowest correlation of 0.45. Similarly to experiment 1, RF2-LE and RF2-Synth have a very similar outcome. Again, the concentrations are overestimated by these models in February and March and underestimate the ammonia concentrations in October, November and December. Moreover, the output of RF2-CrIS again overestimates the concentration in February and March, but the estimation of the ammonia concentration is more accurate between March and November (See Table D1).

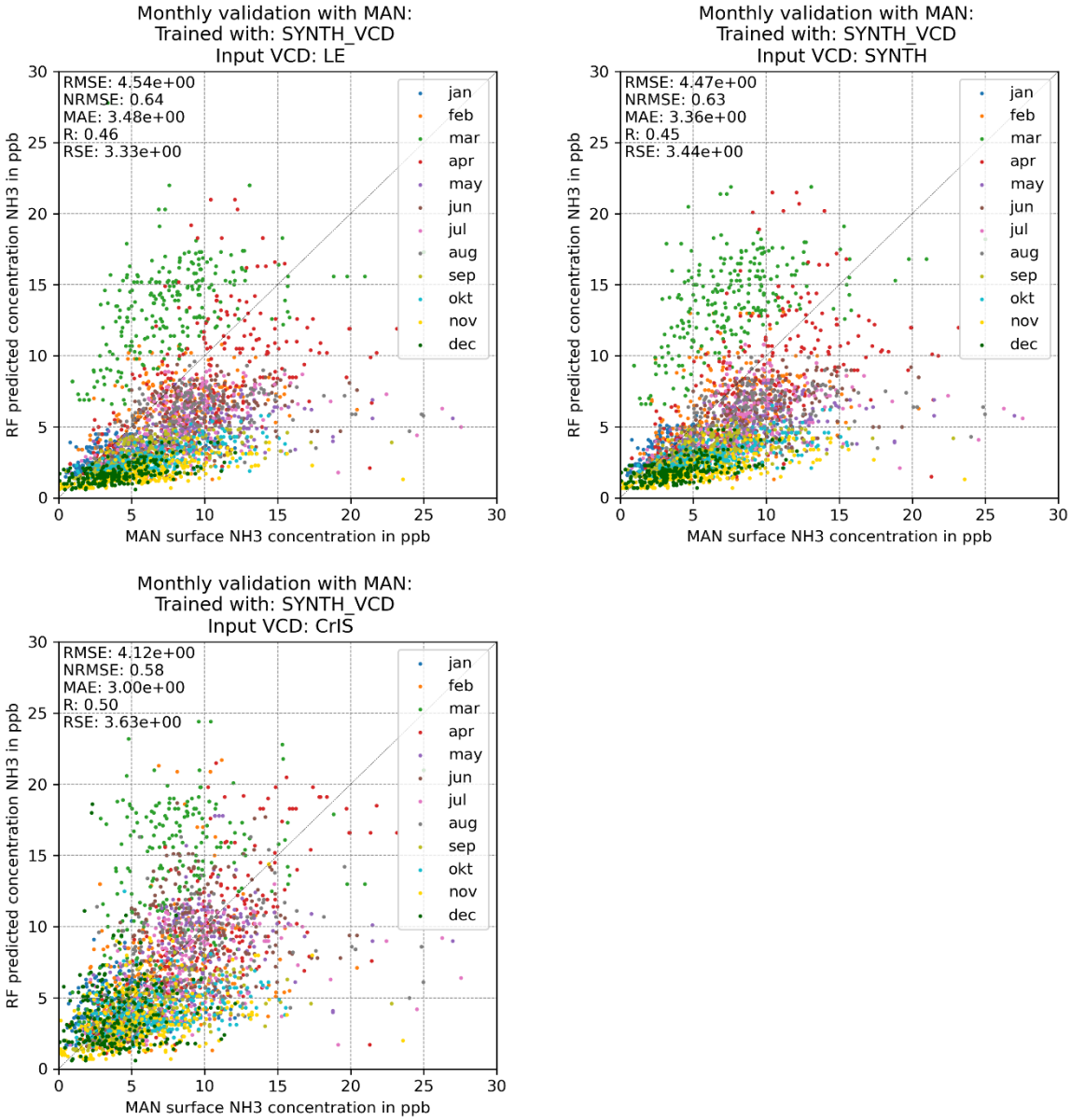

Figure 19: The monthly averaged validation of the output of RF model 2 with MAN measurements. The grey line depicts the 1:1 line and each colour depicts another month.

4.3.3: Validation of results of experiment 3 with MAN-measurements

In the third experiment, the RF model has been trained with CrIS VCD data. When validated against MAN, RF3-CrIS has the lowest RMSE of 4.32 ppb (Figure 20) and also has the highest correlation of 0.47. Furthermore, the RF3-LE model has the highest RMSE of 4.50 ppb and RF3-Synth has the lowest correlation of 0.43. The validation of the results of experiment 3 leads to similar outcomes because the permutation importance of the VCD dataset is very low, implying that varying the VCD dataset will not have much effect (See Section 4.1.3). Similar to the previous experiments, the RF models overestimate the ammonia concentration significantly in February and March and underestimate the ammonia concentration in October, November and December.

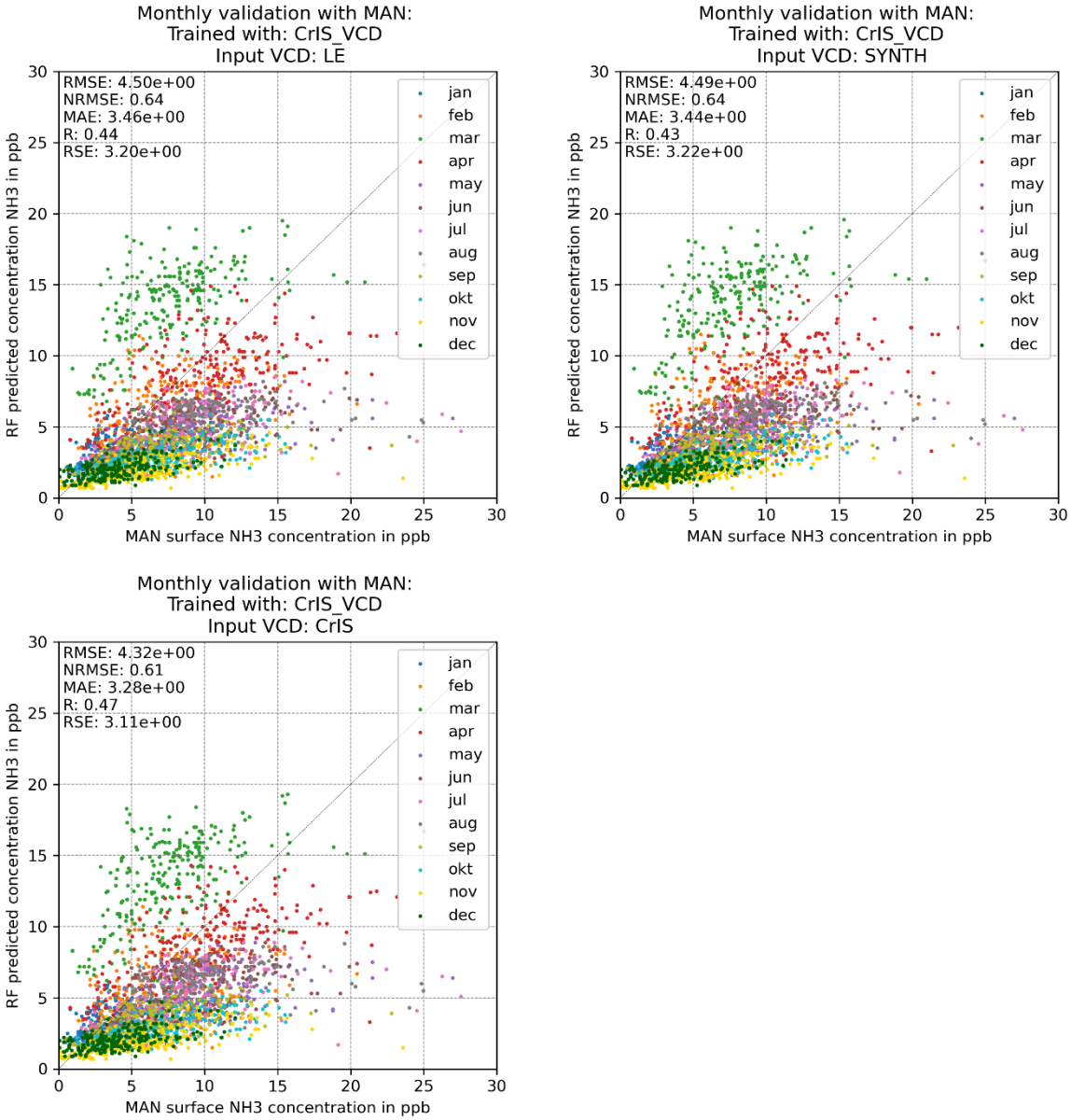

Figure 20: The monthly averaged validation of the output of RF model 3 with MAN measurements. The grey line depicts the 1:1 line and each colour depicts another month.

4.3.4: Validation of results of experiment 4 with MAN-measurements

In experiment 4, no VCD data has been added to the RF model. When this model is validated by the MAN measurements, it has a correlation of 0.43 and a RMSE of 4.43 (Figure 21). Again the RF model overestimates the ammonia concentration in February and March, and in October, November and December it underestimates the ammonia concentration.

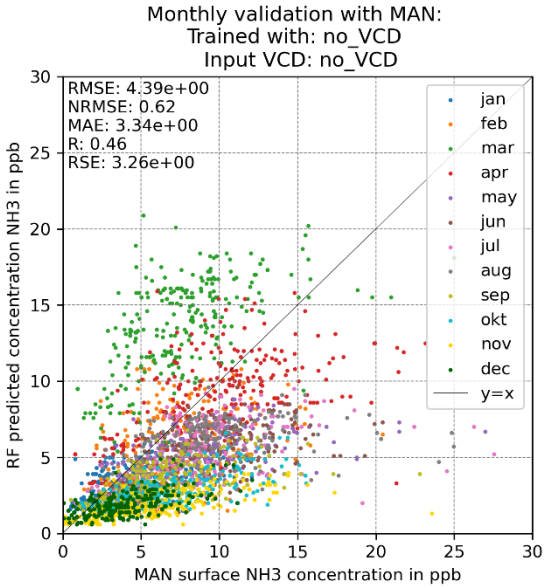

Figure 21: The monthly averaged validation of the output of RF model 4 with MAN measurements. The grey line depicts the 1:1 line and each colour depicts another month.

4.3.5: Validation with MAN of the LOTOS-EUROS model output and the CrIS surface retrieval

In this subsection, the LOTOS-EUROS model output of 2017 and the CrIS surface retrievals are validated by the MAN measurements. The LOTOS-EUROS model output has a RMSE of 4.54 ppb and a correlation of 0.43 when validated (Figure 22). In March, the LOTOS-EUROS model significantly overestimates the ammonia concentration and in November the LOTOS-EUROS model underestimates the ammonia concentration. In the other months, the mean error between the LOTOS-EUROS model and the MAN measurements is less notable.

When validating the CrIS surface retrieval with the MAN measurements, the RMSE is 5.38 ppb, and the correlation is 0.44. The CrIS instrument performs poorly in January, February, March, November and December and underestimates the ammonia concentration significantly. The other months generally have a lower error between the CrIS surface retrieval and the MAN measurements, although the spread around the 1:1 line is moderately broad.

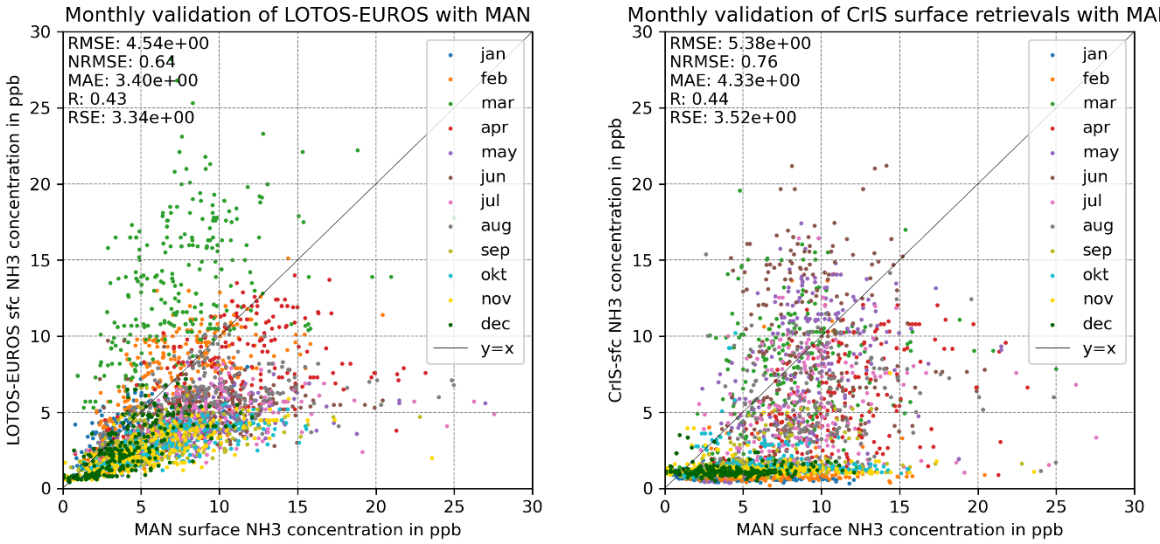

Figure 22: The monthly averaged validation of the LOTOS-EUROS ammonia surface concentrations (left) and the CrIS surface retrievals (right) with MAN measurements. The grey line is the 1:1 line. Each colour depicts a different month.

4.3.6: Interim conclusions on the validation of the random forest models 1-4 by MAN-measurements

In this section, the interim conclusions are drawn from the validation by the MAN-measurement stations (Table 4). From all experiments, RF2-CrIS had both the lowest RMSE of 4.12 ppb and the highest correlation of 0.50. In comparison with the validation results of the LOTOS-EUROS model, the RF2-CrIS model – but also many other RF models – have better performances when validated with MAN data. From this, we can also infer that these RF models could be a viable addition to the LOTOS-EUROS model. However, it must be stressed that although this validation has been performed at 266 different stations, this validation with MAN measurements only takes place in natural areas where the ammonia concentrations are generally lower.

The validation result of RF model 4 which has been trained without VCD data also shows that adding VCD data has an added value: With a RMSE of 4.43 ppb and a correlation of 0.43, it has a lower performance than most RF models which have been supplied with VCD data.

Table 4: The RMSE and correlation of the RF experiment output, LOTOS-EUROS model output and CrIS surface retrievals when validated by MAN measurements. The best results are marked in green.

Experiment	RMSE (ppb)	r	Experiment	RMSE (ppb)	r
RF1-LE	4.49	0.45	RF3-LE	4.50	0.44
RF1-Synth	4.25	0.38	RF3-Synth	4.49	0.43
RF1-CrIS	4.21	0.47	RF3-CrIS	4.32	0.47
RF2-LE	4.54	0.46	RF4 (No VCD)	4.43	0.43
RF2-Synth	4.47	0.45	LOTOS-EUROS	4.54	0.43
RF2-CrIS	**4.12**	**0.50**	CrIS surface retr.	5.38	0.44

When assessing the performance of the RF models and also the LOTOS-EUROS model over the months, the most significant phenomenon was the clear overestimation of the ammonia concentration

in February and March and the underestimation in the final three months of the year. Two general remarks can be made when comparing the performance of the RF models with the LOTOS-EUROS model over the months. In the months of January, February and December, the LOTOS-EUROS model has higher accuracy in terms of RMSE (See Table D1). However, the RF models perform better than the LOTOS-EUROS model in the majority of the months – namely from March to November.

4.4: Partial dependency analysis

In this subsection, the partial dependency analysis is performed: For each explanatory variable, the marginal effect on the response variable is calculated while all other explanatory variables are kept constant (Friedman, 2001). The partial dependency plots for each RF model will be briefly discussed to infer whether the variables have a logical relationship with the ammonia surface concentration. Note that the partial dependency plot of random forest model 4 is excluded in this section, as random forest models 3 and 4 are nearly the same. The partial dependency plot of random forest model 4 can be found in Appendix E.

An important assumption during partial dependency analysis is that the explanatory variables used as input for the RF model are not correlated. Due to collinearity, the calculation of the relationship between the explanatory variable and the response variable can be distorted. In Figure 23, the correlation matrix between all explanatory variables is shown. The correlation between the LOTOS-EUROS VCD, the synthetic CrIS VCD and the CrIS VCD can be ignored as these VCD datasets will never be used at the same time during the prediction phase.

For most explanatory variables, the absolute correlation with other variables is below 0.5. However, for the following variable pairs, a moderate to strong correlation ($r > 0.5$ or $r < -0.5$) has been found:

- Specific humidity and surface temperature.
- Cow density, chicken density and pig density.
- LOTOS-EUROS VCD with CAMS-reg time profiles for manure application.
- Synthetic CrIS VCD with CAMS-reg time profiles for manure application.
- Cow density and the land-use class agriculture
- Cow density and the land-use class water bodies
- The land-use classes water bodies and agriculture

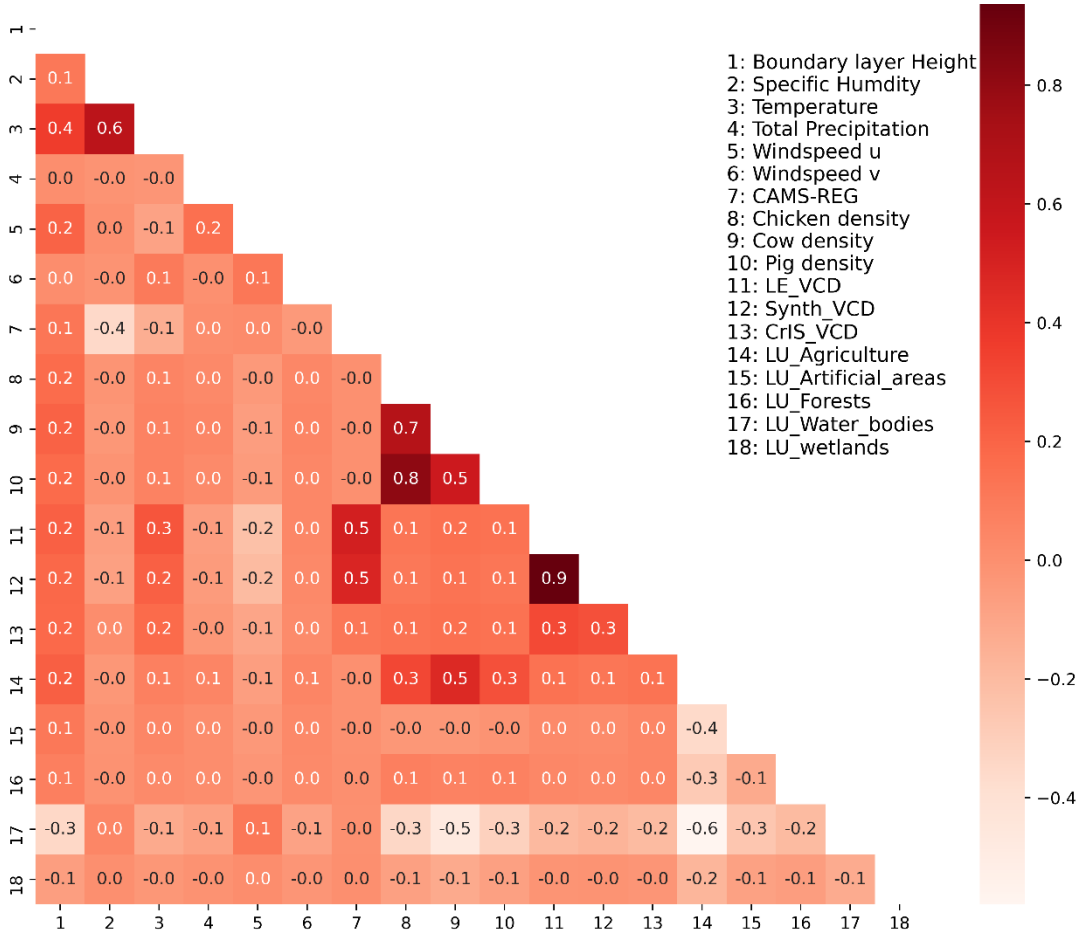

Figure 23: The correlation matrix for all explanatory variables used in this study. Note that the three VCD datasets (numbers 11, 12 and 13) will be used at the same time in an experiment as input: These correlations can be ignored.

4.4.1.: Partial dependency analysis of random forest model 1

The partial dependency plot of random forest model 1 clearly shows the great influence the VCD dataset has on the surface concentration: As the VCD increases, the surface ammonia concentration also increases (Figure 24). Other explanatory variables that show a clear effect on the ammonia surface concentration are the boundary layer height and the CAMS-reg time profiles for manure application. It is evident that as the boundary layer height increases, the surface ammonia concentration decreases. This is a logical effect, as the depth of the mixing layer increases which causes the concentration of pollutants to decrease (Stroll, 1988). Moreover, the surface concentration increases both with increasing CAMS-reg time profile values and increasing cow densities. Finally, a typical effect can be noticed for the surface temperature: As the temperature drops below ~275K, the ammonia concentration increases. Normally, the volatilization of ammonia from manure should decrease when the temperature drops. The temperature variable is likely correlated with another variable that is not included in this study. The permutation importance of temperature is 1%, so this will not have serious consequences for the model output.

The other variables do not have a significant impact on the surface ammonia concentration. This division between important and unimportant explanatory variables is also reflected in the outcome of the permutation importance of each variable of RF model 1, which has been discussed in Section 4.1.1.

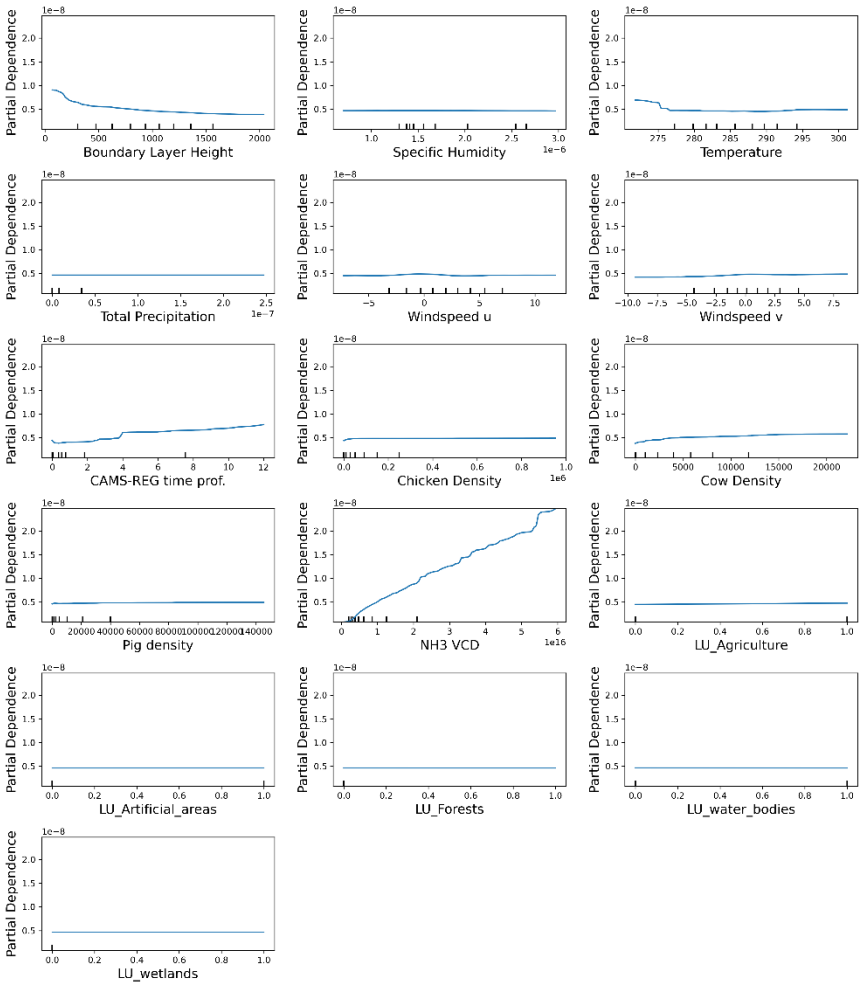

Figure 24: Partial dependency plot of RF model 1. The black lines along the x-axis show the deciles of the original dataset. The x-axis displays the values of the explanatory variable, while the y-axis displays the marginal effect of the variable on the ammonia concentration.

4.4.2: Partial dependency analysis of random forest model 2

The partial dependency plot for random forest 2 is quite similar when compared to the plot of random forest 1 (Figure 25). First, it is shown again that as the boundary layer height increases, the surface concentration decreases. Second, when the temperature increases, the surface concentration also increases. This is reasonable, as the volatilization of ammonia from manure increases with temperature (Sutton et al., 2013). Again, the dubious effect of the increasing ammonia concentrations below 275 K can be seen. The partial dependency of windspeed in horizontal (u) and vertical (v) direction on the ammonia concentration is limited but does show that the concentration is higher when the windspeeds are near 0 m s^{-1}.

The CAMS-reg time profiles for manure application show a clear positive relationship: the higher the value of the time profile, the higher the ammonia concentration. The impact of the animal density datasets by the FAO remains very limited and a plateau is reached very quickly. Considering the correlation between the datasets of chicken-, cow-, and pig density, these partial dependence plots must be interpreted with care. Therefore, thorough conclusions should not be drawn. Finally, the synthetic CrIS VCD still has the strongest (positive) effect on the ammonia surface concentration.

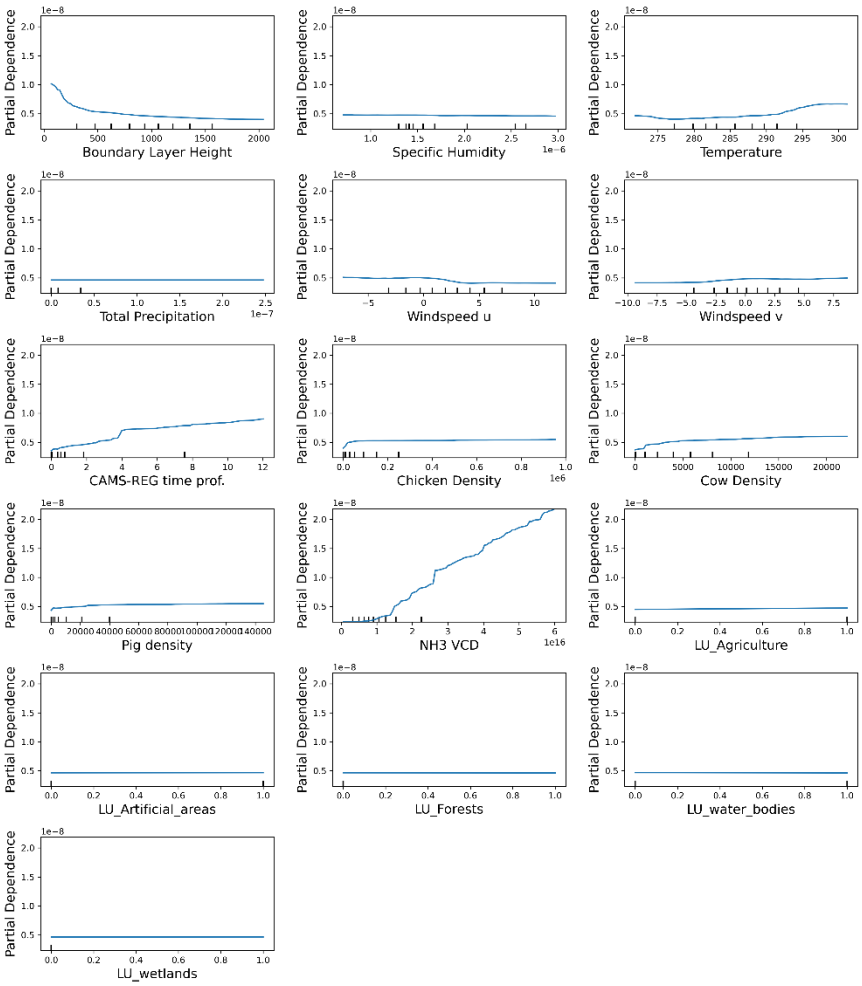

Figure 25: Partial dependency plot of RF model 2. The black lines along the x-axis show the deciles of the original dataset. The x-axis displays the values of the explanatory variable, while the y-axis displays the marginal effect of the variable on the ammonia concentration.

4.4.3: Partial dependency analysis of random forest model 3

In random forest model 3, the permutation importance of the VCD dataset dropped significantly (see Section 4.1.3) causing other explanatory variables to have a greater influence on the ammonia concentration (Figure 26). The effect of the windspeed on the surface concentration is dubious: While higher windspeeds in the eastward (positive u) and southward direction (negative v) lead to lower ammonia surface concentrations, the effect of the wind from the opposite directions remains very limited.

The CAMS-reg time profile data is the dominant variable and has the most significant effect on the surface concentration. Moreover, the marginal effect of the CrIS VCD dataset is less steep compared to the marginal effect of the LOTOS-EUROS VCD and synthetic CrIS VCD datasets (see Sections 4.4.1 and 4.4.2). At a VCD of 5 x 10^{16} molecules cm^{-2}, the increase in the ammonia surface concentration even reaches a plateau. Finally, the effects of the land-use variables 'Agricultural area' and 'Water bodies' can be seen, although the effect on the ammonia surface temperature is very limited. The relationship of these two variables with the ammonia surface concentration is rational: In agricultural areas (a value of 1 indicates the presence of agricultural area in a grid cell), the ammonia concentration increases slightly, while the ammonia concentration is slightly lower above water bodies.

There must be noted that these two land-use variables have a correlation, which could distort the calculation of the partial dependency.

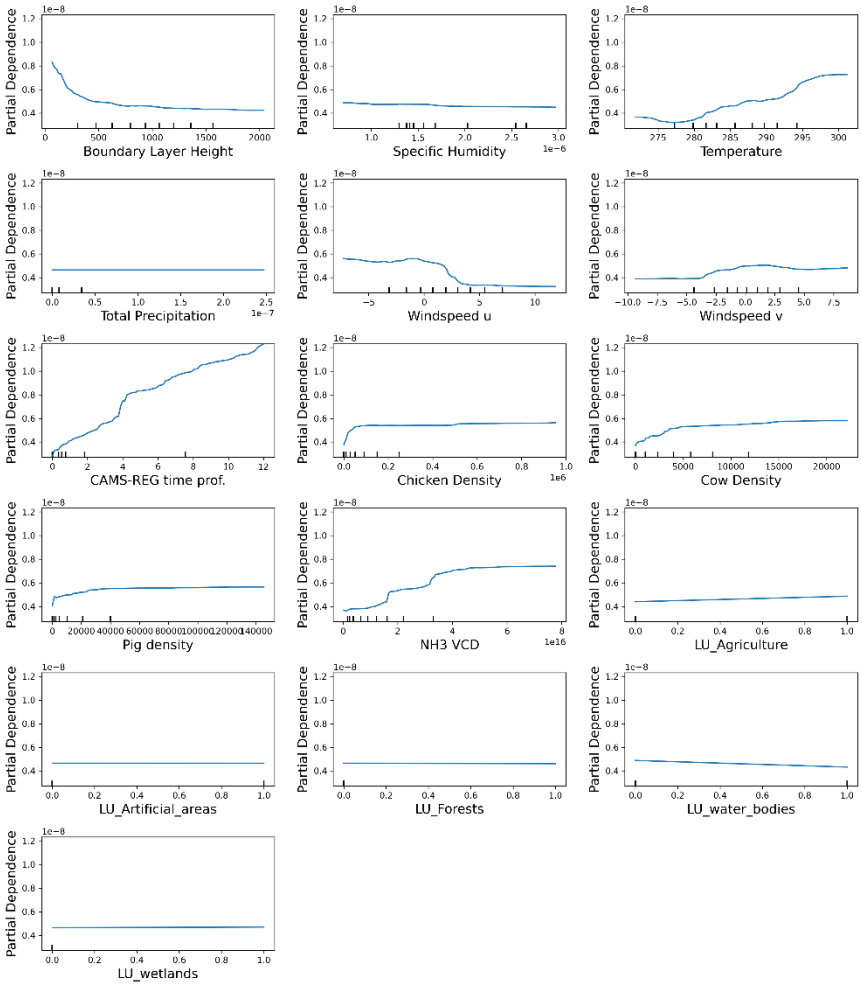

Figure 26: Partial dependency plot of RF model 3. The black lines along the x-axis show the deciles of the original dataset. The x-axis displays the values of the explanatory variable, while the y-axis displays the marginal effect of the variable on the ammonia concentration.

4.5: Discussion of results

In this section, the results will be briefly summarized and discussed. Also, the limitations and recommendations will be deliberated. In this study, four different RF models have been trained with either LOTOS-EUROS VCD, synthetic CrIS VCD, CrIS VCD or no VCD data. Consequently, each RF model (except RF model 4) has been supplied with either LOTOS-EUROS VCD, synthetic CrIS VCD or CrIS VCD data during the prediction phase. This is the phase in which the ammonia surface concentrations are predicted for 2017. Thereafter, the results of each experiment has been validated with ground-based measurements from LML and MAN.

RF models 1 and 2 were both heavily reliant on VCD data, while RF model 3 depended mainly on CAMS-reg time profiles for manure application, surface temperature, windspeed and boundary layer height. RF model 4 which has been trained without VCD data mostly relied on CAMS-reg time profile data and meteorological variables such as the surface temperature, windspeed and boundary layer height. In all random forest models, the effect on the ammonia surface concentration by Corine land-use data, FAO animal density data, total precipitation and specific humidity was very limited.

The validation of the RF experiments has been done with ground-based measurement data from LML and MAN. When validated with LML, the most accurate RF model was trained with LE VCD data and supplied with CrIS VCD. With a RMSE of 5.73 and a correlation of 0.52, this RF model had a stronger performance than the LOTOS-EUROS model output. The LOTOS-EUROS model had a RMSE of 7.11 and a correlation of 0.48. Frankly, most other RF models also had a better performance than the LOTOS-EUROS output. Also when comparing the validation scores of the LE-model and the RF model per LML measurement station, it became apparent that the RF model predicted the ammonia concentration with better accuracy in 3 out of 5 stations. Although this proves that the RF model can be a useful addition to the LOTOS-EUROS model, it should be noted that this validation is only performed at five locations, which is a low sample size.

Similarly to the validation with LML, the RF models show promising results when validated by the MAN measurement stations. The RF models show better performance than the LOTOS-EUROS model. The strongest RF model is the model which is trained with synthetic CrIS data and fed with real CrIS data during the prediction phase with a RMSE of 4.12 ppb and a correlation of 0.50. This RF model shows to be more accurate than the LE-model, which has a RMSE of 4.54 ppb and a correlation of 0.43. Moreover, when analysing the performance per month, the LOTOS-EUROS model still performed the best in the winter (December to February) while the RF models performed better in the spring, summer and fall.

Finally, it was also proven that adding VCD data to RF models is beneficial when estimating the surface ammonia concentration. The validation results of RF model 4 which has been trained without VCD data shows a weaker performance than the regular RF models. However, RF model 3 which has been trained with CrIS VCD data, is among the weaker RF models. This is caused by the inability of the RF algorithm to interpret the CrIS VCD data. Nonetheless, the low performance of this RF model should not be interpreted as the futility of the CrIS VCD dataset in general. When the CrIS VCD data is applied to the RF model which has been trained with the LOTOS-EUROS VCD or the synthetic CrIS VCD dataset, it results in the most accurate datasets when validated by LML and MAN.

During the partial dependency analysis, it was demonstrated that most explanatory variables had a relationship with the surface ammonia concentration which is consistent with basic meteorology. However, the effect of the windspeed on the surface concentrations is doubtful: While stronger eastward and southwards windspeeds led to a lower ammonia concentration, windspeeds in the westward and northward direction do not lead to lower ammonia concentrations. This does not cohere with standard meteorology because an increase in ventilation would normally decrease the ammonia concentration. Finally, from the partial dependency analysis, it also became clear once more that the influence of specific humidity, total precipitation and especially the Corine land-use data was very limited.

When comparing the results of this study with the study by Lu et al. (2020) mentioned earlier in the literature review (Section 2.3), this study has a lower performance in terms of NRMSE: Their global RF model for predicting the NO_2 concentration had a NRMSE of 0.28, while the strongest RF model in this study had a NRMSE of 0.55. There are several possible explanations for why the results by Lu et al. have a stronger performance: First, Lu et al. train their RF model with the ground station NO_2 measurements as the response variable and validate their model with this dataset, meaning that the model will be optimized to predict the surface concentration as measured by the ground stations as good as possible. On the other hand, the RF model in this thesis is trained with the LOTOS-EUROS

surface concentration as the response variable, instead of the ground-based measurement stations by which the RF output will be validated. This will partly cause the lower performance of the RF model in this thesis because it has not been made to optimize the ammonia concentration based on the ground-based measurements stations. Another factor that could explain the difference in the model performance is that Lu et al. use more than 3500 NO_2 monitoring stations with an hourly resolution. Currently, sample sizes in this order of magnitude cannot be achieved with ammonia ground stations.

Moreover, the low permutation importance of the Corine 2018 land-use classes in this study is remarkable. In the study by Lu et al. (2020), among the variables with the highest feature importance were land-use classes representing roads of different sizes. A possible explanation why the Corine land-use classes – and the land-use class 'Agriculture area' in particular – was not useful in this study could be because it is too general. The class 'Agricultural area' does not give information about whether the grid cell contains livestock housing or if it contains a pasture that will be fertilized. Moreover, it also does not specify the number of animals, the type of animals or (in the case livestock housing is present) the type of indoor ventilation in the grid cell. In contrast, in the study of Lu et al. (2020) road length is arguably a better proxy for the intensity of the NO_2 emission.

Another point that must be considered is the fact that satellite observations can only be done during clear-sky conditions. The result of this is that the RF model might be biased towards days on which the sky was relatively clear. The lack of clouds can influence other meteorological variables, such as temperature, windspeed, humidity or precipitation. This could cause a serious underrepresentation of certain meteorological conditions in the training and testing datasets and could decrease the representability of the RF model output. This potential bias has also been discussed by Schiferl et al. (2016) concerning the ammonia observations by IASI and in several other air pollution studies (e.g. Geddes et al., 2012). Another disadvantage related to the use of satellite data is that the RF model can only predict the surface concentrations where satellite observations have been made, implying that the model is very dependent on satellite retrievals.

Two final limitations that must be discussed are related to the validation of the RF output. First, the RF model that has been built in this study is only tested in 2017, which makes it uncertain how the RF model would perform in other years. Moreover, there is a lack of ground-truth data points where the RF model can be validated. Although the MAN network is extensive in space, it is placed in natural areas where the ammonia concentration is lower. Also, it has a low temporal resolution of one measurement per month. The miniDOAS equipment used for the LML network does have a high temporal resolution of one observation per hour. However, there are just six LML stations in total, which is too little for the validation of a novel ammonia product such as the one created in this study.

A few recommendations for further research can be given. First, it is recommended to experiment with other machine learning models to predict the ammonia surface concentration. This can be other ensemble tree-based approaches such as stochastic gradient boosting (SGB) or extreme gradient boosting (XGB), but also other machine learning algorithms such as support vector machines and neural networks which show to be promising for air quality modelling (Rybarczyk & Zalakeviciute, 2018).

Second, as mentioned before, the random forest algorithm in this study only predicts the surface ammonia concentration at grid cells where VCD observations are available. The result of this is that potentially valuable data will not be used in the case that no satellite observations have been made in certain grid cells. A solution to this problem can be to implement value imputation, which is the process

of using the "information and relationships among the non-missing predictors to provide an estimate to fill the missing value" (Kuhn & Johnson, 2019). Various strategies for gap-filling satellite data have already been studied (Xiao et al., 2021) and could be further researched for atmospheric ammonia remote sensing.

The third recommendation for further research would be to investigate other useful explanatory variables that can be used as input for RF models. Examples of new variables that can be added are aerosol optical depth (AOD) data (ammonia can transform into aerosols when it reacts with nitrate or sulphate), NO_2 and SO_2 concentration data, NH_3 VCD data from the IASI instrument and soil moisture data.

A fourth recommendation is to broaden the geographical boundaries of the study domain. Doing similar experiments in a bigger domain increases the amount of data on which the RF model can train, which can lead to a stronger RF model. However, modelling ammonia can also become more complex when increasing the study range because new meteorological and topological conditions will be introduced when extending the geographical boundaries (e.g. the influence of elevation on the ammonia concentration when extending beyond the Dutch borders). Additionally, the potential also lies in training a model in a certain region and using it to predict the ammonia concentration in other regions. This concept is better known as transfer learning in the machine learning discipline and the added value of this could be that a RF model can be trained and validated in one region, which can then be applied in another region.

Fifth, it is recommended to train and test the model in different years to understand how the RF performance would be affected. In the current model setup, the model is trained in 2015 and 2016 and tested in 2017 only. However, to test whether this new method for modelling the ammonia surface concentrations would be a viable addition, the model should also be trained and tested in other years as conclusions should not be drawn from just one year.

A sixth recommendation is to use the LOTOS-EUROS ammonia surface dataset which has been assimilated with CrIS data instead of the regular LOTOS-EUROS model data which has been used in this study. With a local ensemble transform Kalman filter (LETKF), the model output is altered to better match the spatiotemporal ammonia distribution of CrIS (van der Graaf et al., 2022). Using this dataset could improve the model performance when validated by ground-based measurement stations.

A seventh recommendation is to explore the capability of RF models to model the ammonia concentration at a higher spatiotemporal resolution. Especially for a highly dispersible compound such as ammonia, datasets with a higher resolution in space and time would be useful. In the research by Lu et al. (2020), an ultra-high resolution of 25 × 25 meters has already been computed for NO_2 concentrations with success.

The final recommendation is to research the possibility to use ground ammonia concentrations measured by the MAN- and LML-stations as the response variable instead of the LOTOS-EUROS surface values. This would be the same approach as the study by Lu et al. (2020). Although the number of stations measuring at a daily temporal resolution is low, a RF model can be constructed to calculate the monthly average concentrations at other locations than the MAN and LML measurement stations. Such a RF model can help with filling the gaps between the measurement stations which would create a new interpolated ammonia concentration dataset independent of the LOTOS-EUROS model.

Moreover, an advantage of such as model is that it will be optimized specifically to approach the measurements by MAN and LML as good as possible (in contrast to the RF constructed in this study).

Section 5: Conclusion

In this study, the surface concentrations of ammonia in the Netherlands have been modelled with the random forest algorithm and validated with LML and MAN measurement stations. The research question posed in this thesis is how random forest models can be applied with CrIS satellite data, meteorological variables and land-use based variables to improve the accuracy of ammonia surface concentration datasets when validated by ground-truth data.

The results of this study have shown the potential that lies in RF models: The comparison of the results of the RF models and the LOTOS-EUROS model when validated by the LML and MAN networks show that RF models have a higher correlation and a lower error. More specifically, while the LOTOS-EUROS model had a RMSE of 7.11 ppb and a correlation of 0.48 when validated by LML, the best RF model had a RMSE of 5.73 ppb and a correlation of 0.52. Likewise, when validated with MAN measurements, the LOTOS-EUROS model had a RMSE of 4.54 and a correlation of 0.43, while the best RF model had a RMSE of 4.12 and a correlation of 0.50.

The results also show that in the current experimental setup, the RF model which was trained with CrIS VCD does not perform properly yet. However, the CrIS satellite data is very beneficial when it is applied to a RF model which has been trained with either the LOTOS-EUROS VCD (RF model 1) or the synthetic CrIS VCD dataset (RF model 2). Furthermore, the validated results of the RF model which has been trained without VCD data (RF model 4), show a lower performance compared to the regular RF models. This proves the added benefit of incorporating VCD datasets in RF models for estimating the ammonia surface concentration.

For the strongest RF models (RF models 1 and 2), the VCD was by far the most important explanatory variable with a permutation importance above 50%. Thereafter, the CAMS-reg time profiles for manure application and the boundary layer height were the most important variables. Variables that had a smaller contribution were chicken-, cow- and pig density, the surface temperature and the windspeed in the vertical and horizontal direction. Explanatory variables which practically had no impact on the model performance were the Corine 2018 land use classes, the specific surface humidity and the precipitation per day.

The results in this study are promising, although there are some limitations: The RF model constructed in this study only supplies output for locations where satellite observations have been made, meaning that the daily coverage of the model is fully dependent on the ability of the CrIS instrument to capture the VCD. Moreover, there is a chance that the model is biased towards cloud-free conditions because the CrIS retrievals in this study are made during clear-sky conditions. Given the possibility that the lack of clouds can be correlated with other meteorological variables (e.g. precipitation, windspeed, surface temperature), this could potentially cause an underrepresentation of certain meteorological conditions in the RF model. Furthermore, the RF model is only tested in the year 2017, making it unknown how the model would perform over more years. Finally, there are too few ground stations measuring the ammonia concentration, which makes it hard to conduct a thorough validation of the RF model output.

Recommendations for further research include testing the capabilities of other machine learning algorithms to model the ammonia concentration, experimenting with value imputation to increase the coverage of the RF model, and investigating other explanatory variables (e.g. IASI VCD or AOD data) which can improve the model performance. Additionally, it is necessary to also test and validate the RF model in other years than 2017 to infer whether random forest modelling is a viable addition to

other ammonia monitoring methods. Finally, it is also suggested to experiment with using ground-based measurement stations as the ground-truth for the RF algorithm instead of the LOTOS-EUROS surface concentration. In line with the research by Lu et al. (2020), such an approach can create a new ammonia concentration dataset that is independent of the LOTOS-EUROS model and is optimized to approach the MAN and LML measurements as closely as possible.

Although the monitoring of ammonia concentrations and emissions have proven to be difficult and full of uncertainty, many new promising techniques are being explored and developed which can help to better understand the behaviour of this challenging compound in the biosphere. This thesis can be considered as a stepping stone towards exploring the full potential for modelling ammonia in the atmosphere with machine learning techniques such as the random forest model. As an addition to the existing methods of monitoring ammonia, it can contribute to mitigating the damage caused by reactive nitrogen emissions on both ecosystems and society.

References

Asman, W. A., Sutton, M. A., & Schjørring, J. K. (1998). Ammonia: emission, atmospheric transport and deposition. *New phytologist*, *139*(1), 27-48.

Ayres, R. U., Schlesinger, W. H., & Socolow, R. H. (1994). Human impacts on the carbon and nitrogen cycles

Behera, S. N., Sharma, M., Aneja, V. P., & Balasubramanian, R. (2013). Ammonia in the atmosphere: a review on emission sources, atmospheric chemistry and deposition on terrestrial bodies. Environmental Science and Pollution Research, 20(11), 8092-8131.

Berkhout, A. J., Swart, D. P., Volten, H., Gast, L. F., Haaima, M., Verboom, H., ... & Hoogerbrugge, R. (2017). Replacing the AMOR with the miniDOAS in the ammonia monitoring network in the Netherlands. Atmospheric Measurement Techniques, 10(11), 4099-4120.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. Atmospheric Environment, 151, 1-11.

Chislock, M. F., Doster, E., Zitomer, R. A., & Wilson, A. E. (2013). Eutrophication: causes, consequences, and controls in aquatic ecosystems. *Nature Education Knowledge*, *4*(4), 10.

Dammers, E. (2017a). Measuring atmospheric Ammonia with Remote Sensing: Validation of satellite observations with ground-based measurements.

Dammers, E., McLinden, C. A., Griffin, D., Shephard, M. W., Graaf, S. V. D., Lutsch, E., ... & Erisman, J. W. (2019). NH 3 emissions from large point sources derived from CrIS and IASI satellite observations. *Atmospheric chemistry and physics*, *19*(19), 12261-12293.

Dammers, E., Palm, M., Van Damme, M., Vigouroux, C., Smale, D., Conway, S., ... & Erisman, J. W. (2016). An evaluation of IASI-NH 3 with ground-based Fourier transform infrared spectroscopy measurements. *Atmospheric Chemistry and Physics*, *16*(16), 10351-10368.

Dammers, E., Shephard, M. W., Palm, M., Cady-Pereira, K., Capps, S., Lutsch, E., ... & Erisman, J. W. (2017b). Validation of the CrIS fast physical NH 3 retrieval with ground-based FTIR. *Atmospheric Measurement Techniques*, *10*(7), 2645-2667.

de Vries, W., Erisman, J. W., Spranger, T., Stevens, C. J., & van den Berg, L. (2011). Nitrogen as a threat to European terrestrial biodiversity. *The European nitrogen assessment: sources, effects and policy perspectives*, 436-494.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media.

Denk, T. R., Mohn, J., Decock, C., Lewicka-Szczebak, D., Harris, E., Butterbach-Bahl, K., ... & Wolf, B. (2017). The nitrogen cycle: A review of isotope effects and isotope modeling approaches. Soil Biology and Biochemistry, 105, 121-137.

Dentener, F. J., & Crutzen, P. J. (1994). A three-dimensional model of the global ammonia cycle. Journal of Atmospheric Chemistry, 19(4), 331-369.

Di, Q., Koutrakis, P., & Schwartz, J. (2016). A hybrid prediction model for PM2. 5 mass and components using a chemical transport model and land use regression. *Atmospheric environment*, *131*, 390-399.

Dun, M., Xu, Z., Chen, Y., & Wu, L. (2020). Short-term air quality prediction based on fractional grey linear regression and support vector machine. *Mathematical Problems in Engineering*, *2020*.

Duyzer, J. (1994). Dry deposition of ammonia and ammonium aerosols over heathland. *Journal of Geophysical Research: Atmospheres*, *99*(D9), 18757-18763.

Erisman, J. W., Bleeker, A., Galloway, J., & Sutton, M. S. (2007). Reduced nitrogen in ecology and the environment. Environmental pollution, 150(1), 140-149.

European Commission. (2001). Directive 2001/81/EC of the European Parliament and of the Council of 23 October 2001 on national emission ceilings for certain atmospheric pollutants. Official Journal of the European Communities, 309, 22-30.

FAO. (2010). *Global distributions | Livestock Systems*. Global distributions | Livestock Systems | Food and Agriculture Organization of the United Nations. Retrieved January 25, 2022, from https://www.fao.org/livestock-systems/global-distributions/en/

Fariña, J. M., Salazar, S., Wallem, K. P., Witman, J. D., & Ellis, J. C. (2003). Nutrient exchanges between marine and terrestrial ecosystems: the case of the Galapagos sea lion Zalophus wollebaecki. Journal of Animal Ecology, 873-887.

Farquhar, G. D., Firth, P. M., Wetselaar, R., & Weir, B. (1980). On the gaseous exchange of ammonia between leaves and the environment: determination of the ammonia compensation point. *Plant Physiology*, *66*(4), 710-714.

Ferm, M. (1998). Atmospheric ammonia and ammonium transport in Europe and critical loads: a review. *Nutrient Cycling in Agroecosystems*, *51*(1), 5-17.

Fisher, A., Rudin, C., & Dominici, F. (2018). All models are wrong but many are useful: variable importance for black-box, proprietary, or misspecified prediction models, using model class reliance. *arXiv preprint arXiv:1801.01489*, 237-246.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Fürnkranz, J. (2017). Decision tree. *Encyclopedia of Machine Learning and Data Mining*, 330–335. https://doi.org/10.1007/978-1-4899-7687-1_66

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., & Cosby, B. J. (2003). The nitrogen cascade. Bioscience, 53(4), 341-356.

Gay, S. W., & Knowlton, K. F. (2005). Ammonia emissions and animal agriculture.

Ge, X., Schaap, M., Kranenburg, R., Segers, A., Reinds, G. J., Kros, H., & de Vries, W. (2020). Modeling atmospheric ammonia using agricultural emissions with improved spatial variability and temporal dynamics. *Atmospheric Chemistry and Physics*, *20*(24), 16055-16087.

Geddes, J. A., Murphy, J. G., O'Brien, J. M., & Celarier, E. A. (2012). Biases in long-term NO2 averages inferred from satellite observations due to cloud selection criteria. *Remote Sensing of Environment*, *124*, 210-216.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *journal of Computational and Graphical Statistics*, *24*(1), 44-65.

Haines-Young, R., & Potschin, M. (2010). The links between biodiversity, ecosystem services and human well-being. *Ecosystem Ecology: a new synthesis*, *1*, 110-139.

Head, T., MechCoder, L. G., & Shcherbatyi, I. (2018). scikit-optimize/scikit-optimize: v0. 5.2. *Version v0*, *5*.

Hegg, D. A., Radke, L. F., Hobbs, P. V., & Riggan, P. J. (1988). Ammonia emissions from biomass burning. Geophysical research letters, 15(4), 335-337.

Hendriks, C., Kranenburg, R., Kuenen, J. J. P., Van den Bril, B., Verguts, V., & Schaap, M. (2016). Ammonia emission time profiles based on manure transport data improve ammonia modelling across north western Europe. *Atmospheric Environment*, *131*, 83-96.

Johnson, M. T., Liss, P. S., Bell, T. G., Lesworth, T. J., Baker, A. R., Hind, A. J., ... & Gibb, S. W. (2008). Field observations of the ocean-atmosphere exchange of ammonia: Fundamental importance of temperature as revealed by a comparison of high and low latitudes. Global Biogeochemical Cycles, 22(1).

Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride J., Gerard, J., … Leblanc, F. (2020). geopandas/geopandas: v0.8.1 (Version v0.8.1). Zenodo. http://doi.org/10.5281/zenodo.3946761

Koerkamp, P. G., Metz, J. H. M., Uenk, G. H., Phillips, V. R., Holden, M. R., Sneath, R. W., ... & Wathes, C. M. (1998). Concentrations and emissions of ammonia in livestock buildings in Northern Europe. *Journal of Agricultural Engineering Research*, *70*(1), 79-95.

Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J. P., Super, I., & Denier van der Gon, H. (2021). CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling. *Earth System Science Data Discussions*, 1-37.

Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. CRC Press.

Lolkema, D. E., Noordijk, H., Stolk, A. P., Hoogerbrugge, R., van Zanten, M. C., & van Pul, W. A. J. (2015). The measuring ammonia in nature (MAN) network in the Netherlands. Biogeosciences, 12(16), 5133-5142.

Lutsch, E., Dammers, E., Conway, S., & Strong, K. (2016). Long-range transport of NH3, CO, HCN, and C2H6 from the 2014 Canadian Wildfires. *Geophysical Research Letters*, *43*(15), 8286-8297.

Manders, A. M., Builtjes, P. J., Curier, L., Denier van der Gon, H. A., Hendriks, C., Jonkers, S., ... & Schaap, M. (2017). Curriculum vitae of the LOTOS–EUROS (v2. 0) chemistry transport model. Geoscientific Model Development, 10(11), 4145-4173.

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.".

Nemitz, E., Milford, C., & Sutton, M. A. (2001). A two-layer canopy compensation point model for describing bi-directional biosphere-atmosphere exchange of ammonia. *Quarterly Journal of the Royal Meteorological Society*, *127*(573), 815-833.

nLu, M., Schmitz, O., de Hoogh, K., Kai, Q., & Karssenberg, D. (2020). Evaluation of different methods and data sources to optimise modelling of NO2 at a global scale. *Environment international*, *142*, 105856.

Noordijk, H., Braam, M., Rutledge-Jonker, S., Hoogerbrugge, R., Stolk, A. P., & van Pul, W. A. J. (2020). Performance of the MAN ammonia monitoring network in the Netherlands. *Atmospheric Environment*, *228*, 117400.

Paulot, F., Jacob, D. J., Pinder, R. W., Bash, J. O., Travis, K., & Henze, D. K. (2014). Ammonia emissions in the United States, European Union, and China derived by high-resolution inversion of ammonium wet deposition data: Interpretation with a new agricultural emissions inventory (MASAGE_NH3). *Journal of Geophysical Research: Atmospheres*, *119*(7), 4343-4364.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association*, *56*(6), 709-742.

Quinlan, J. R. (1983). Learning efficient classification procedures and their application to chess end games. In *Machine learning* (pp. 463-482). Springer, Berlin, Heidelberg.

Quinn, P. K., Charlson, R. J., & Bates, T. S. (1988). Simultaneous observations of ammonia in the atmosphere and ocean. *Nature*, *335*(6188), 336-338.

Robnik-Šikonja, M. (2004, September). Improving random forests. In *European conference on machine learning* (pp. 359-370). Springer, Berlin, Heidelberg.

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine learning approaches for outdoor air quality modelling: A systematic review. *Applied Sciences*, *8*(12), 2570.

Schiferl, L. D., Heald, C. L., Van Damme, M., Clarisse, L., Clerbaux, C., Coheur, P. F., ... & Eilerman, S. J. (2016). Interannual variability of ammonia concentrations over the United States: sources and implications. *Atmospheric Chemistry and Physics*, *16*(18), 12305-12328.

Selman, M., Greenhalgh, S., Diaz, R., & Sugg, Z. (2008). Eutrophication and hypoxia in coastal areas: a global assessment of the state of knowledge. *World Resources Institute*, *284*, 1-6.

Shaw, W. J., Spicer, C. W., & Kenny, D. V. (1998). Eddy correlation fluxes of trace gases using a tandem mass spectrometer. Atmospheric Environment, 32(17), 2887-2898.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.

Sommer, S. G., Olesen, J. E., & Christensen, B. T. (1991). Effects of temperature, wind speed and air humidity on ammonia volatilization from surface applied cattle slurry. *The Journal of Agricultural Science*, *117*(1), 91-100

Stelson, A. W., & Seinfeld, J. H. (1982). Relative humidity and temperature dependence of the ammonium nitrate dissociation constant. *Atmospheric Environment (1967)*, *16*(5), 983-992.

Stull, R. B. (1988). An introduction to boundary layer meteorology (Vol. 13). Springer Science & Business Media.

Sutton, M. A., Reis, S., Riddick, S. N., Dragosits, U., Nemitz, E., Theobald, M. R., ... & De Vries, W. (2013). Towards a climate-dependent paradigm of ammonia emission and deposition. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1621), 20130166.

Sutton, M. A., Schjoerring, J. K., & Wyers, G. P. (1995). Plant—atmosphere exchange of ammonia. Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences, 351(1696), 261-278.

Sutton, M. A., Van Dijk, N., Levy, P. E., Jones, M. R., Leith, I. D., Sheppard, L. J., ... & Wolseley, P. A. (2020). Alkaline air: changing perspectives on nitrogen and air pollution in an ammonia-rich world. Philosophical Transactions of the Royal Society A, 378(2183), 20190315.

Tian, D., & Niu, S. (2015). A global analysis of soil acidification caused by nitrogen addition. *Environmental Research Letters*, *10*(2), 024019.

TNO. (2019). Emissies en depositie van stikstof in Nederland.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. Journal of Phonetics, 71, 249-267.

Uematsu, M., Toratani, M., Kajino, M., Narita, Y., Senga, Y., & Kimoto, T. (2004). Enhancement of primary productivity in the western North Pacific caused by the eruption of the Miyake-jima Volcano. Geophysical research letters, 31(6).

Van Damme, M., Clarisse, L., Dammers, E., Liu, X., Nowak, J. B., Clerbaux, C., ... & Coheur, P. F. (2015). Towards validation of ammonia (NH 3) measurements from the IASI satellite. *Atmospheric Measurement Techniques*, *8*(3), 1575-1591.\

Van Damme, M., Wichink Kruit, R. J., Schaap, M., Clarisse, L., Clerbaux, C., Coheur, P. F., ... & Erisman, J. W. (2014). Evaluating 4 years of atmospheric ammonia (NH3) over Europe using IASI satellite observations and LOTOS-EUROS model results. *Journal of Geophysical Research: Atmospheres*, *119*(15), 9549-9566.

van der Graaf, S., Dammers, E., Segers, A., Kranenburg, R., Schaap, M., Shephard, M. W., & Erisman, J. W. (2022). Data assimilation of CrIS NH 3 satellite observations for improving spatiotemporal NH 3 distributions in LOTOS-EUROS. Atmospheric Chemistry and Physics, 22(2), 951-972.

Volten, H., Bergwerff, J. B., Haaima, M., Lolkema, D. E., Berkhout, A. J. C., Van Der Hoff, G. R., ... & Swart, D. P. J. (2012). Two instruments based on differential optical absorption spectroscopy (DOAS) to measure accurate ammonia concentrations in the atmosphere. *Atmospheric Measurement Techniques*, *5*(2), 413-427.

von Bobrutzki, K., Braban, C. F., Famulari, D., Jones, S. K., Blackall, T., Smith, T. E. L., ... & Nemitz, E. (2010). Field inter-comparison of eleven atmospheric ammonia measurement techniques. *Atmospheric Measurement Techniques*, *3*(1), 91-112.

Walker Jr, W. W. (1983). Significance of eutrophication in water supply reservoirs. *Journal-American Water Works Association*, *75*(1), 38-42.

Wolfe, A. H., & Patz, J. A. (2002). Reactive nitrogen and human health: acute and long-term implications. *Ambio: A journal of the human environment*, *31*(2), 120-125.

Xiao, Q., Geng, G., Cheng, J., Liang, F., Li, R., Meng, X., ... & He, K. (2021). Evaluation of gap-filling approaches in satellite-based daily PM2. 5 prediction models. Atmospheric Environment, 244, 117921.

Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M. L., Shen, X., ... & Zhang, M. (2017). Spatiotemporal prediction of continuous daily PM2. 5 concentrations across China using a spatially explicit machine learning algorithm. *Atmospheric environment*, *155*, 129-139.

Appendix A: The Corine 2018 land-use classes

Table A1: A list of the Corine 2018 land-use classes. The land-use classes in bold have been used in this study. These main land-use classes can be further specified by the land-use classes listed below.

Artificial surfaces	**Agricultural areas**	**Water bodies**
Continuous urban fabric	Non-irrigated arable land	Water courses
Discontinuous urban fabric	Permanently irrigated land	Water bodies
Industrial or commercial units	Rice fields	Coastal lagoons
Road and rail networks and associated land	Vineyards	Estuaries
Port areas	Fruit trees and berry plantations	Sea and ocean
Airports	Olive groves	Sea and ocean
Mineral extraction sites	Pastures	
Dump sites	Annual crops associated with permanent crops	
Construction sites	Complex cultivation patterns	
Green urban areas Sport and leisure facilities	Land principally occupied by agriculture with significant areas of natural vegetation	
	Agro-forestry areas	
Forest and semi-natural areas	**Wetlands**	
Broad-leaved forest	Inland marshes	
Coniferous forest	Peat bogs	
Mixed forest	Salt marshes	
Natural grasslands	Salines	
Moors and heathland	Intertidal flats	
Sclerophyllous vegetation		
Transitional woodland-shrub	**Water bodies**	
Beaches dunes sands	Water courses	
Bare rocks	Water bodies	
Sparsely vegetated areas	Coastal lagoons	
Burnt areas	Estuaries	
Glaciers and perpetual snow	Sea and ocean	

Appendix B: Hexbin plots with the comparison of LOTOS-EUROS output versus random forest output

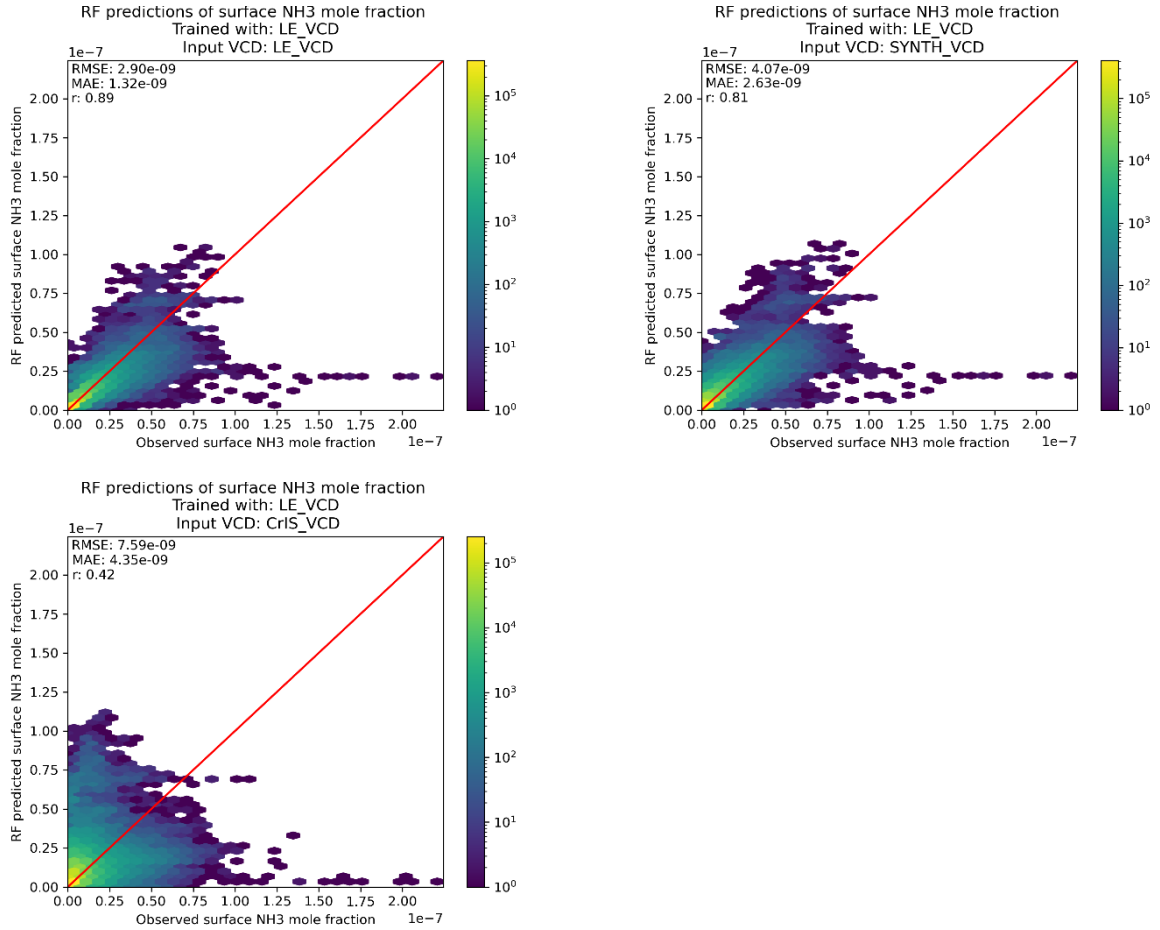

Figure B1: Hexbin plots comparing the LOTOS-EUROS surface concentration with the surface concentrations predicted by RF model 1 which has been trained with LOTOS-EUROS VCD. Each plot shows the output of RF model 1 which has been supplied with either the LOTOS-EUROS VCD, synthetic CrIS VCD or the CrIS VCD.

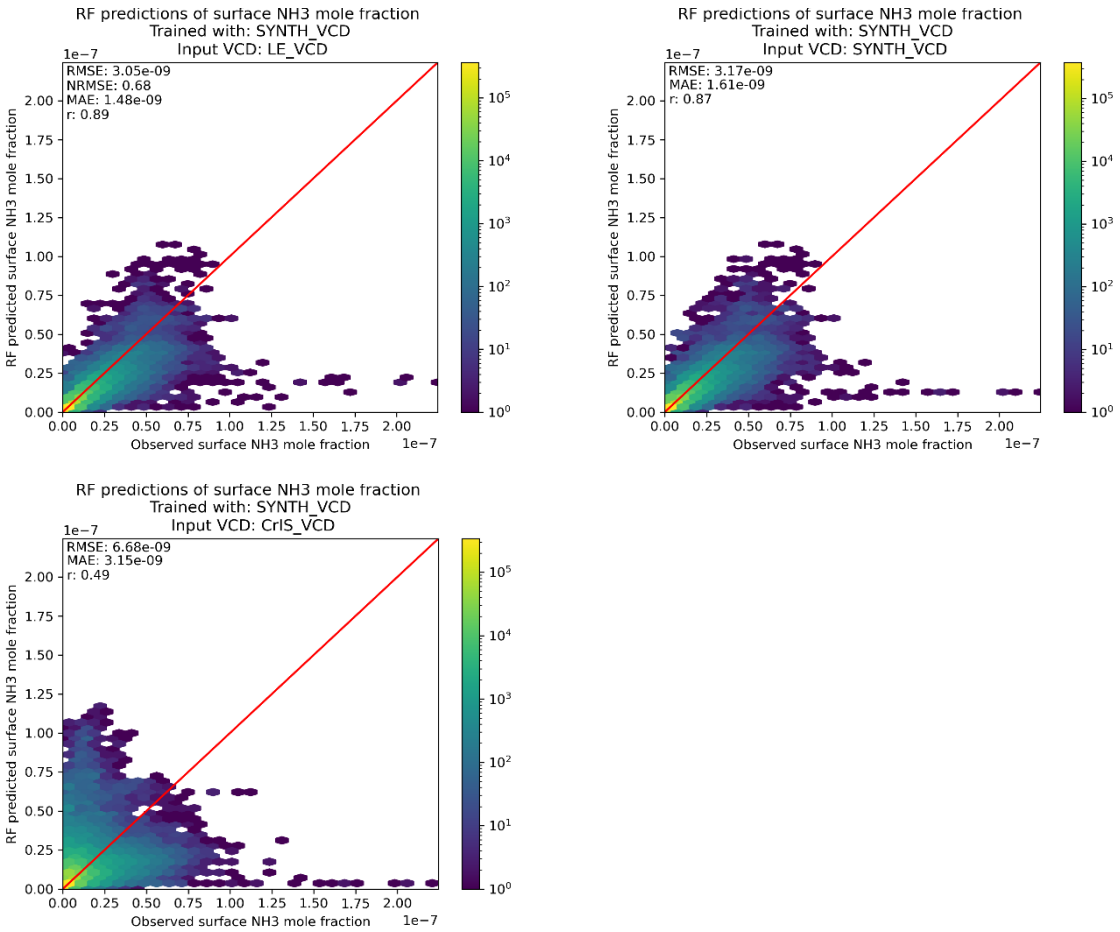

Figure B2: Hexbin plots comparing the LOTOS-EUROS surface concentration with the surface concentrations predicted by RF model 2 which has been trained the synthetic CrIS VCD. Each plot shows the output of RF model 2 which has been supplied with either the LOTOS-EUROS VCD, synthetic CrIS VCD or the CrIS VCD.

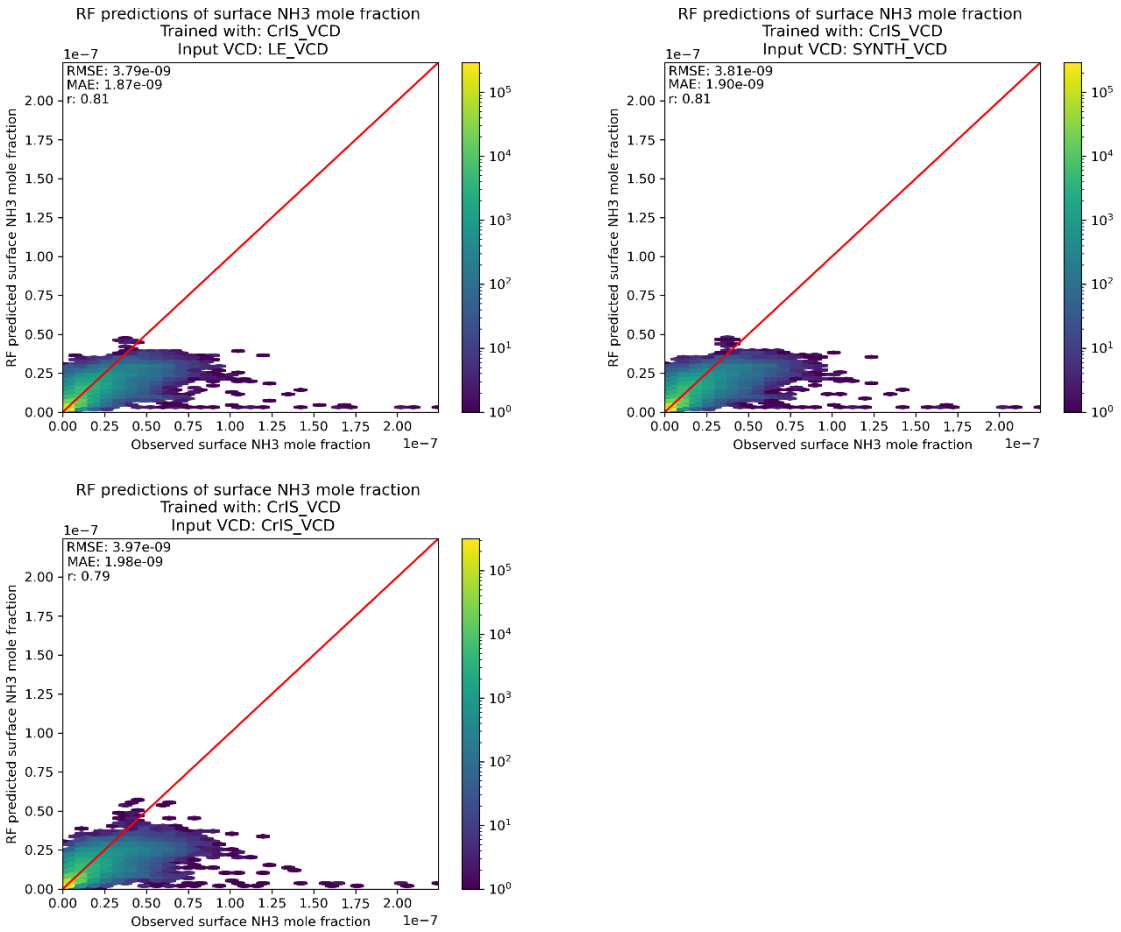

Figure B3: Hexbin plots comparing the LOTOS-EUROS surface concentration with the surface concentrations predicted by RF model 3 which has been trained with the CrIS VCD. Each plot shows the output of RF model 3 which has been supplied with either the LOTOS-EUROS VCD, synthetic CrIS VCD or the CrIS VCD.

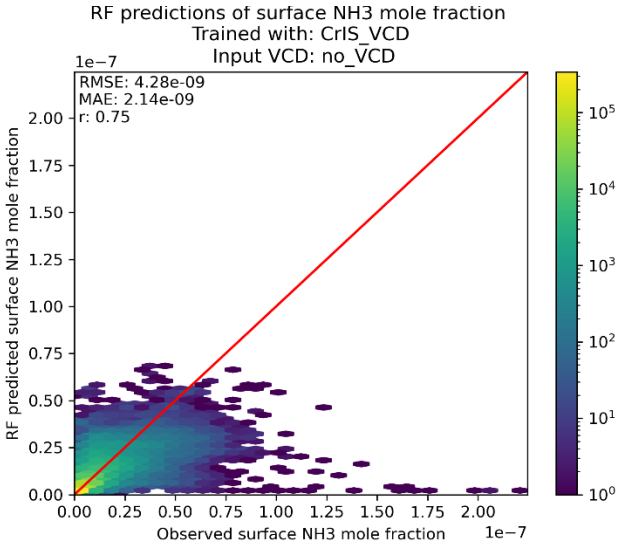

Figure B4: Hexbin plot comparing the LOTOS-EUROS surface concentration with the surface concentration predicted by RF model 4, which has been trained without VCD data.

Appendix C: Daily validation results of the random forest output with LML stations

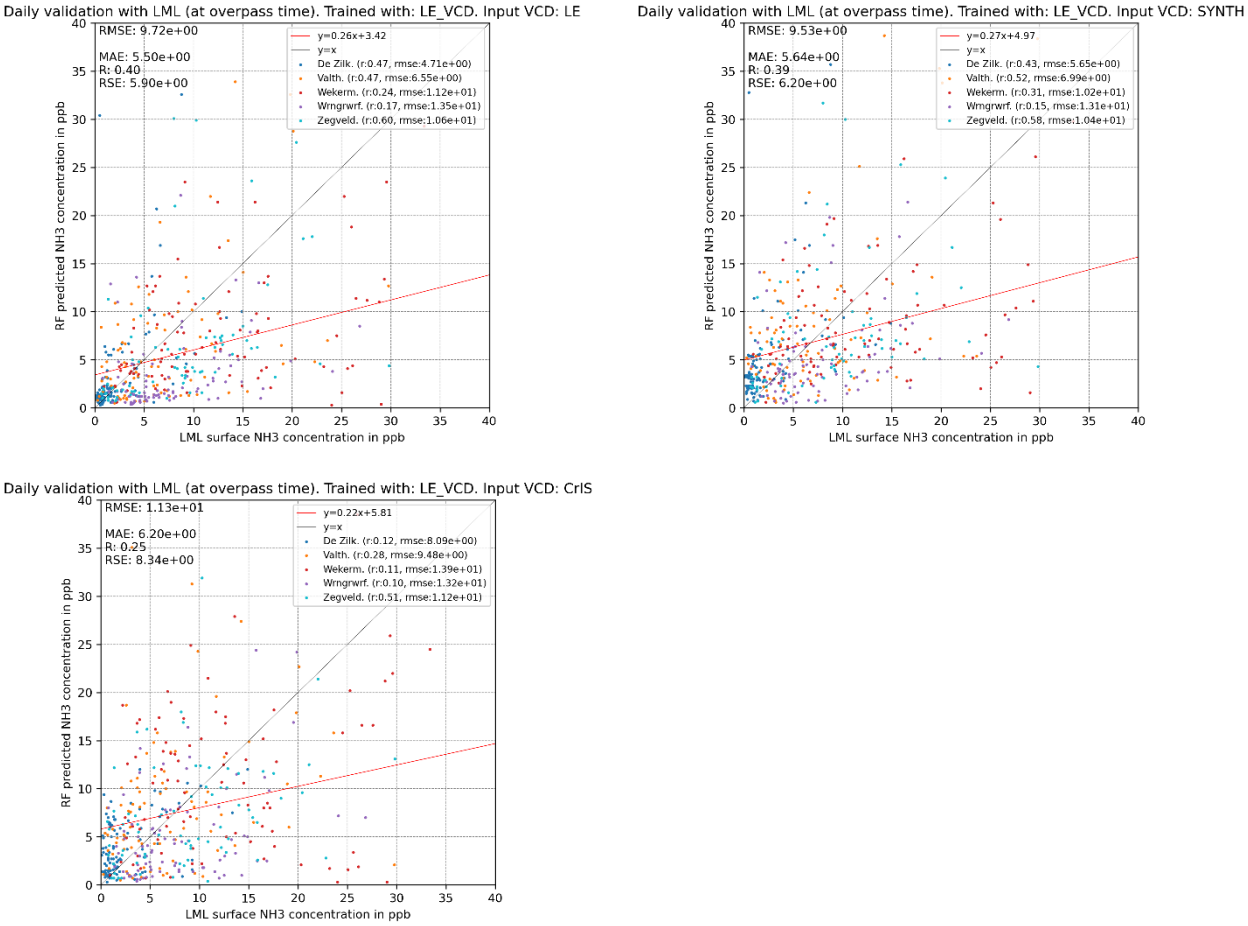

Figure C1: The daily validation with the LML measurement stations of the output of RF model 1, which has been trained with LOTOS-EUROS VCD. Each plot shows the output of RF model 1 which has been supplied with either the LOTOS-EUROS VCD, synthetic CrIS VCD or the CrIS VCD.

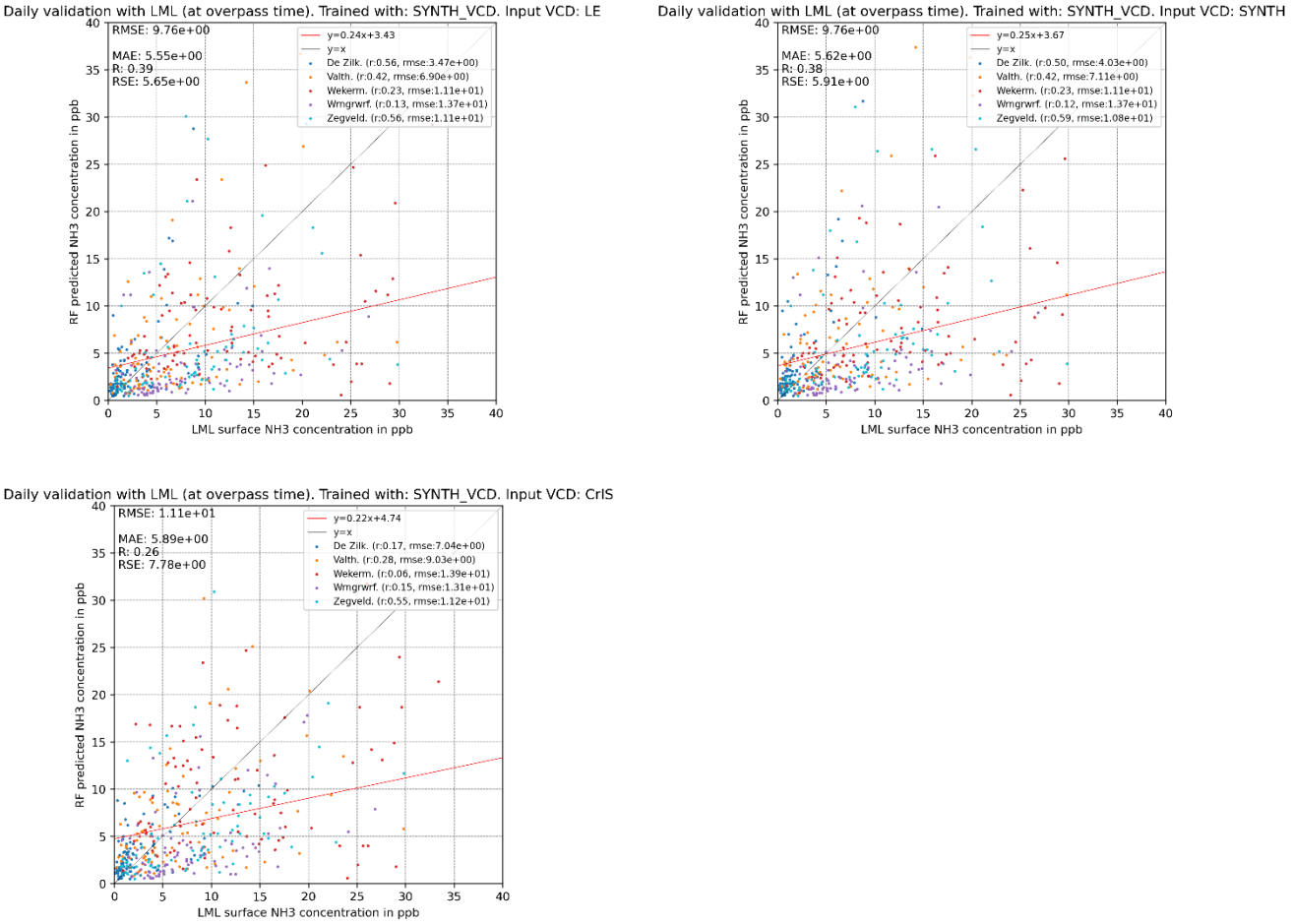

Figure C2: The daily validation with the LML measurement stations of the output of RF model 2, which has been trained with the synthetic CrIS VCD. Each plot shows the output of RF model 2 which has been supplied with either the LOTOS-EUROS VCD, synthetic CrIS VCD or the CrIS VCD.

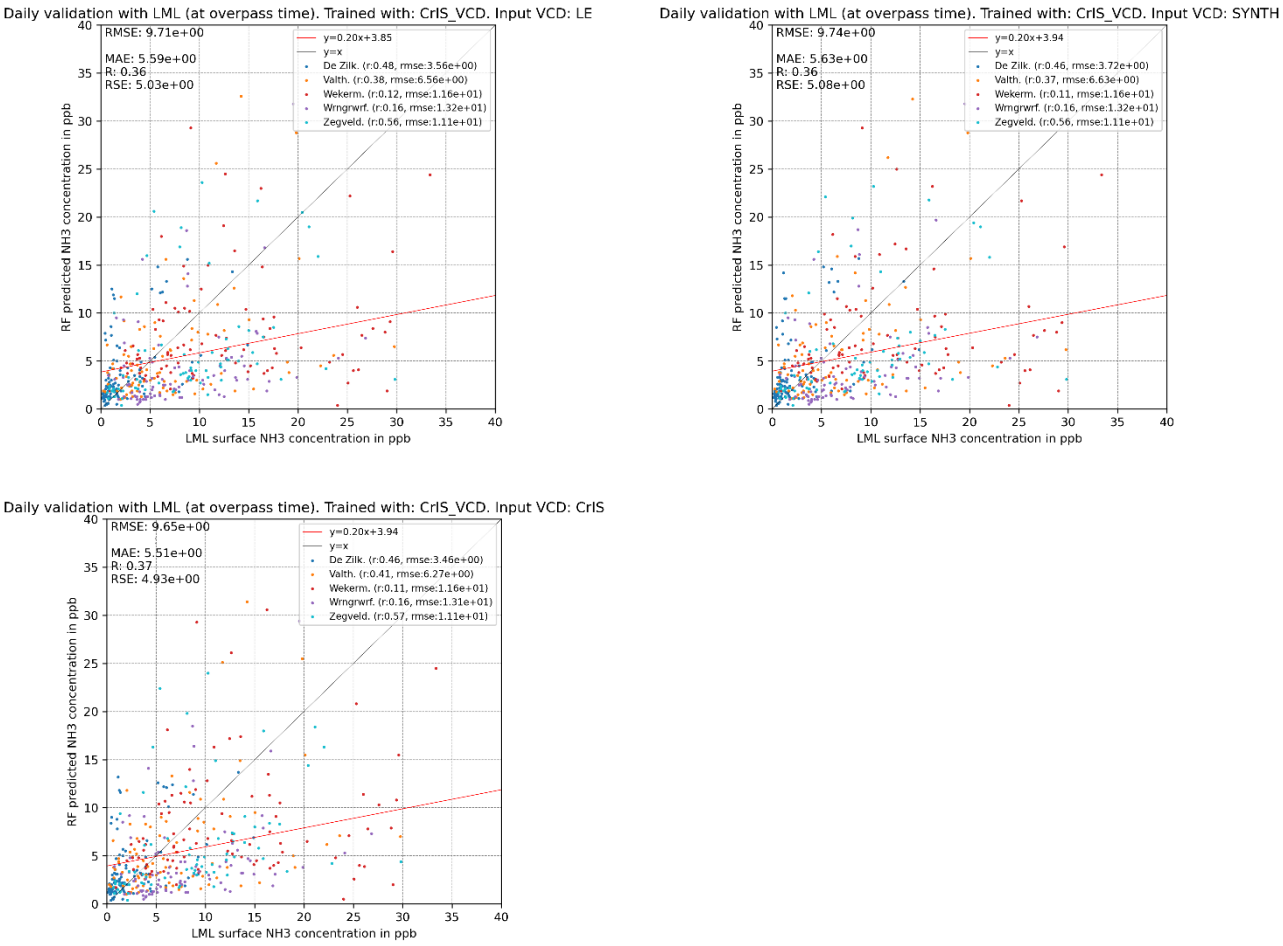

Figure C3: The daily validation with the LML measurement stations of the output of RF model 3, which has been trained with the CrIS VCD. Each plot shows the output of RF model 3 which has been supplied with either the LOTOS-EUROS VCD, synthetic CrIS VCD or the CrIS VCD.

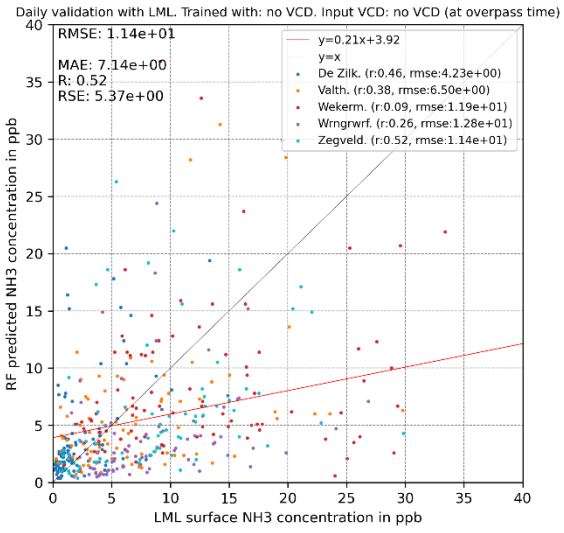

Figure C4: The daily validation with the LML measurement stations of the output of RF model 4, which has not been trained with a VCD dataset.

Appendix D: Yearly validation results of the random forest output with MAN stations and monthly validation per month of random forest output with MAN stations

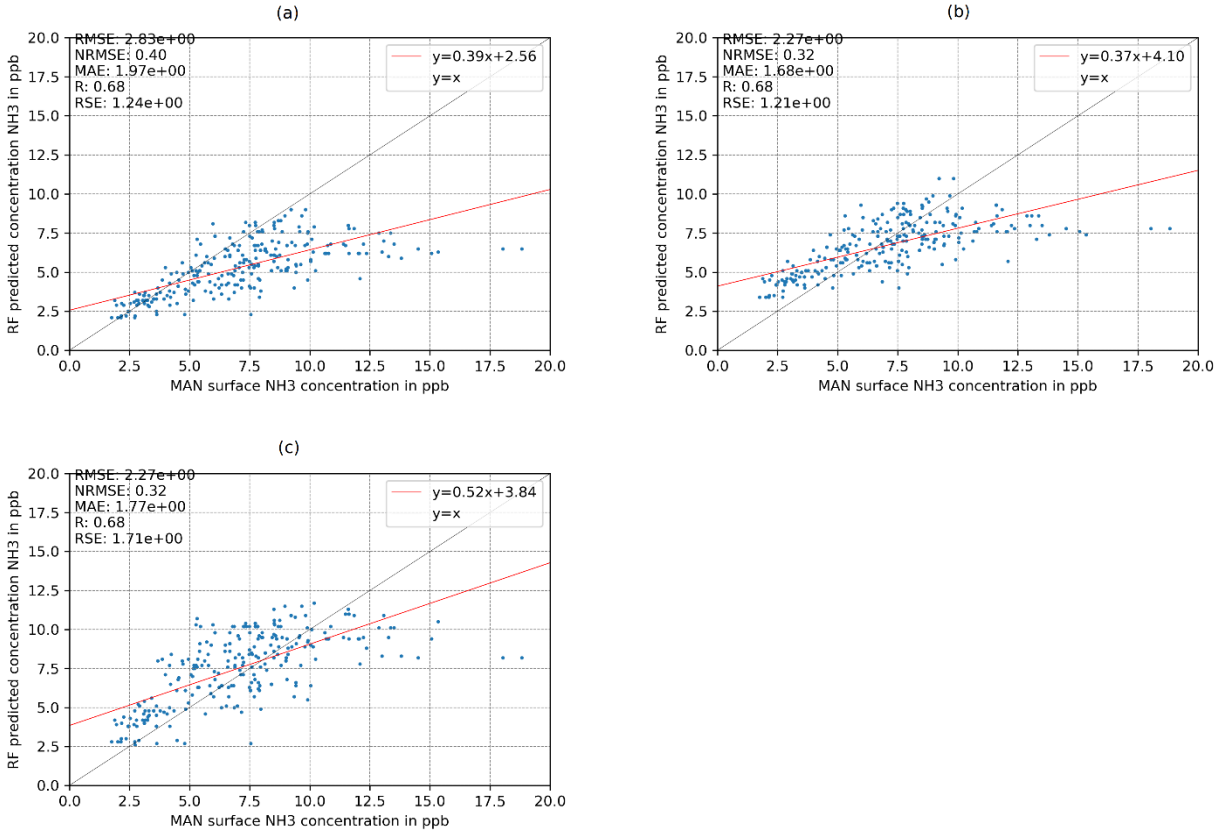

Figure D1: The yearly validation with the MAN measurement stations of the output of RF model 1, which has been trained with the LOTOS-EUROS VCD. Each plot shows the output of RF model 1 which has been supplied with either the LOTOS-EUROS VCD (a), synthetic CrIS VCD (b) or the CrIS VCD (c).

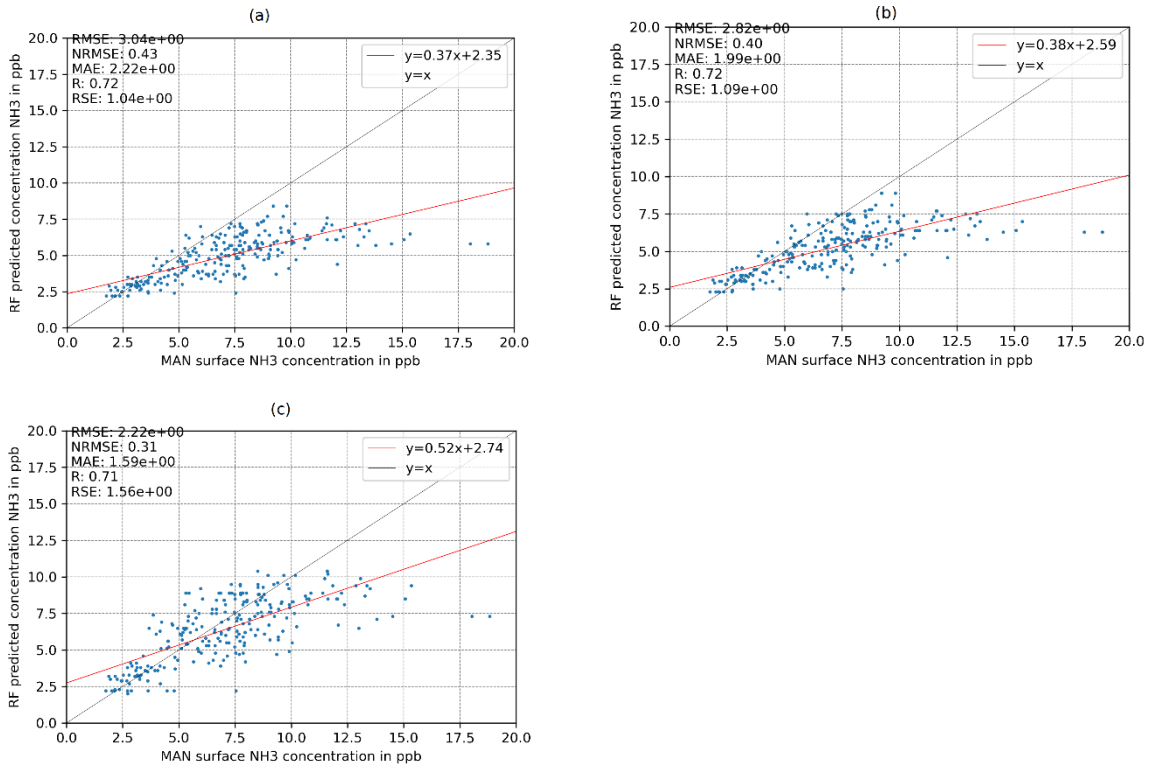

Figure D2: The yearly validation with the MAN measurement stations of the output of RF model 2, which has been trained with the synthetic CrIS VCD. Each plot shows the output of RF model 2 which has been supplied with either the LOTOS-EUROS VCD (a), synthetic CrIS VCD (b) or the CrIS VCD (c).

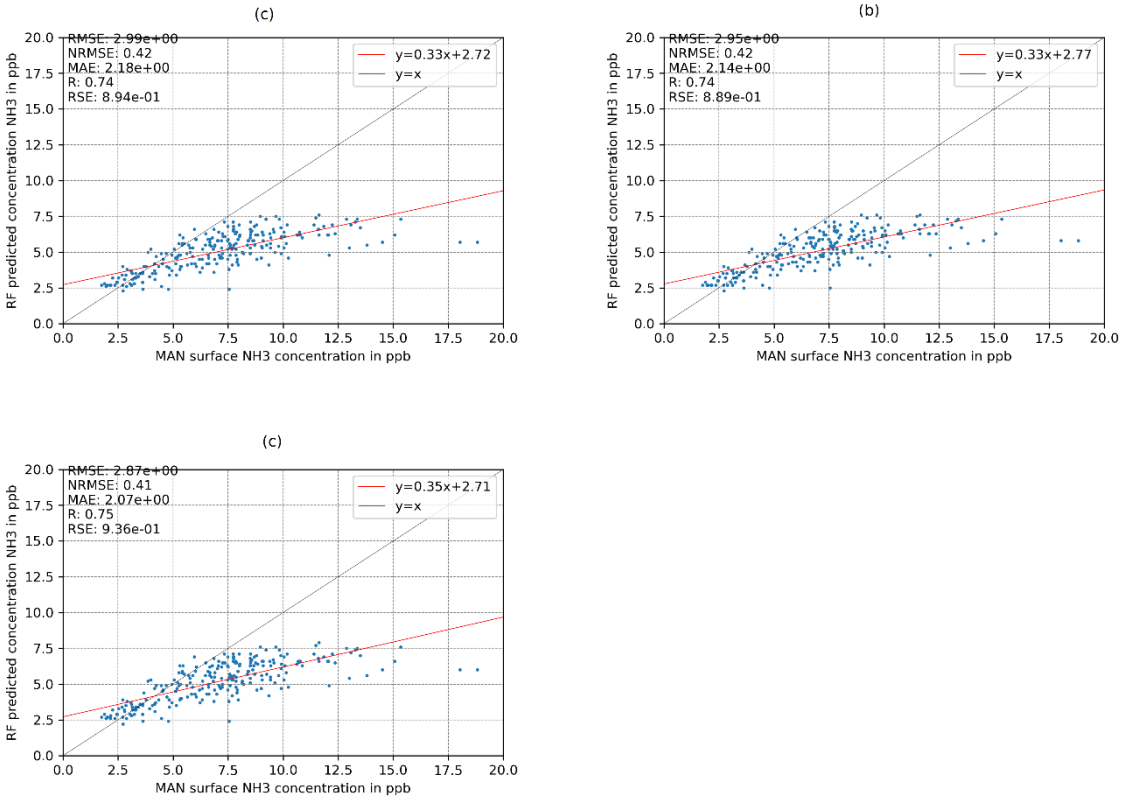

Figure D3: The yearly validation with the MAN measurement stations of the output of RF model 3, which has been trained with the CrIS VCD. Each plot shows the output of RF model 3 which has been supplied with either the LOTOS-EUROS VCD (a), synthetic CrIS VCD (b) or the CrIS VCD (c).

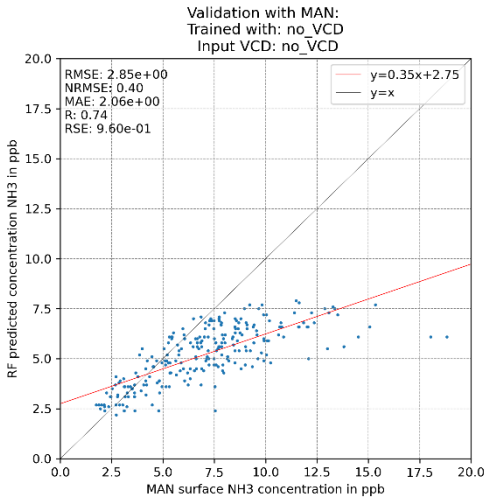

Figure D4: The yearly validation with the MAN measurement stations of the output of RF model 4, which has been trained with no VCD dataset.

Table D1: The RMSE of all MAN stations per month, per experiment. The values which have been marked in bold green show the lowest RMSE of all experiments in a certain month.

RMSE	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
RF1-LE	2.22	3.09	7.15	4.47	5.11	4.12	4.98	4.70	4.06	4.18	4.84	3.02
RF1-Synth	3.61	3.26	7.70	4.50	4.73	**3.60**	4.59	4.17	3.34	3.39	**3.39**	2.37
RF1-CrIS	3.06	3.93	7.45	3.69	**4.04**	4.06	**4.14**	**3.95**	**2.91**	**3.23**	3.50	4.72
RF2-LE	2.27	3.33	6.67	4.36	5.31	4.28	5.00	4.75	4.06	4.72	5.07	3.10
RF2-Synth	2.06	3.36	7.21	4.32	5.11	4.02	4.84	4.57	3.89	4.54	4.81	2.92
RF2-CrIS	2.42	3.87	7.09	**3.59**	4.07	3.83	4.26	3.99	3.46	3.89	4.02	3.36
RF3-LE	2.15	3.31	6.68	3.77	5.27	4.60	5.15	4.93	3.99	4.70	4.80	2.86
RF3-Synth	2.14	3.31	6.76	3.76	5.22	4.54	5.10	4.90	3.94	4.69	4.80	2.86
RF3-CrIS	2.12	3.28	6.41	3.62	4.86	4.20	4.90	4.71	3.86	4.64	4.78	2.86
RF4 (no VCD)	2.09	3.45	6.73	3.84	4.86	4.38	4.75	4.59	3.56	4.64	4.94	2.92
LE	**1.91**	**2.57**	7.23	4.39	5.66	4.94	5.49	4.46	3.99	4.45	4.28	**2.15**
CrIS-sfc	4.05	6.67	**4.49**	6.10	4.88	5.08	5.88	5.76	5.33	6.07	5.68	3.76

Appendix E: Partial dependency plot of random forest model 4

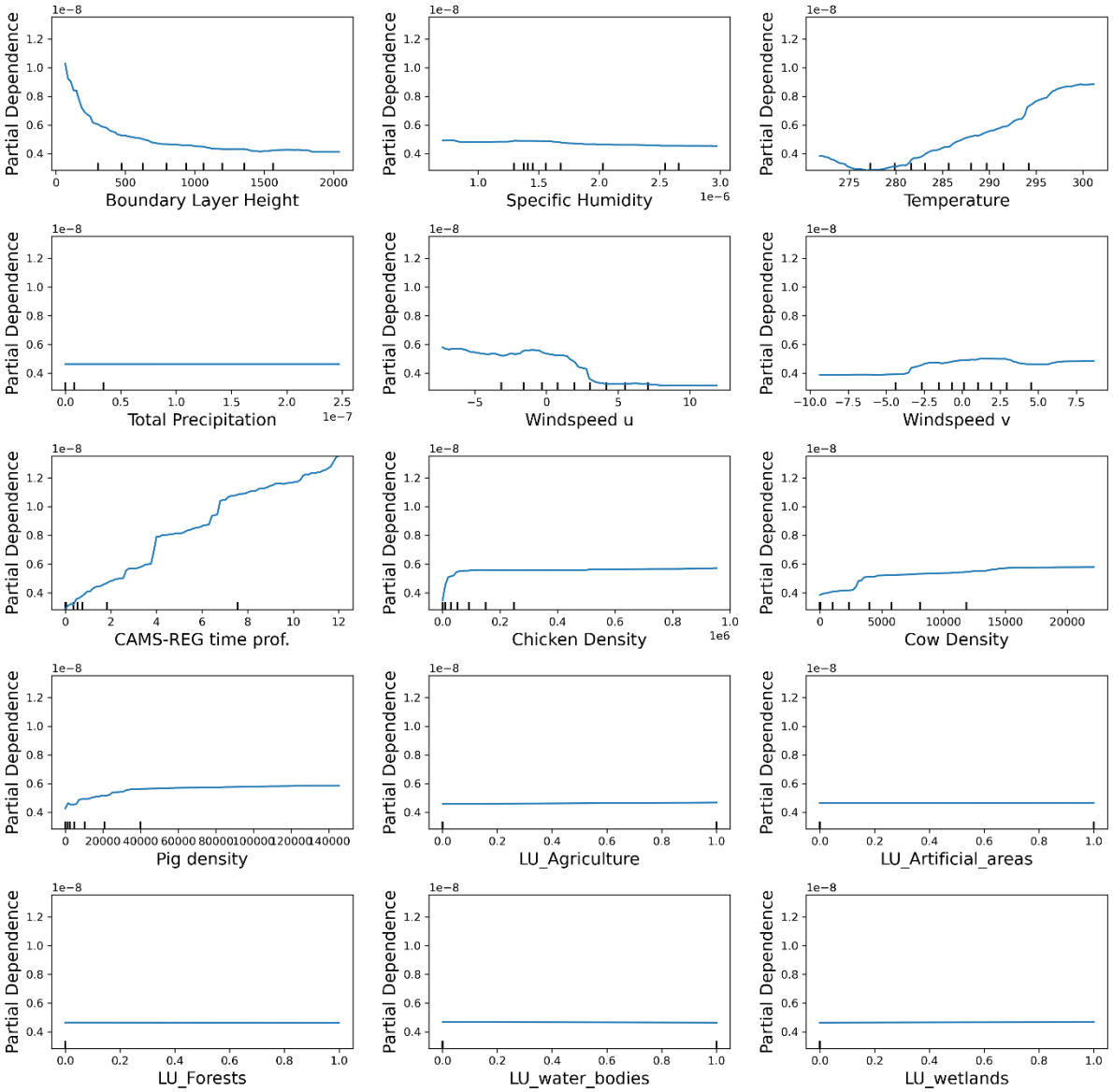

Figure E1: The partial dependency plot of random forest 4, which has been trained without VCD data.