Nationaal Onderzoeksprogramma Broeikasgassen Veenweiden

Chapter 10 NOBV year report 2022

Airborne measurements of regional carbon dioxide exchange of Dutch fen meadows

PH-WUR taking measurements over the Groene Hart area

Airborne measurements of regional carbon dioxide exchange of Dutch fen meadows

Authors

Ronald Hutjes, Wietse Franssen, Laura van der Poel, Laurent Bataille, Bart Kruijt, Hanne Berghuis, Jan Biermann, Wilma Jans[1]

Review

Alex Buzacott[2]

Affilitation

[1]Wageningen University, WAGENINGEN

[2]Vrije Universiteit, Amsterdam

Abstract

In support of development of emission reduction policies for the Dutch fen meadow areas, the National Research programme on Greenhouse Gas Dynamics in Peatlands and Organic Soils (in Dutch NOBV: https://www.nobveenweiden.nl/) aims to investigate the effects of various mitigation measures on the total greenhouse gas balance of the targeted areas.
Complementing multi-site ground-based measurements using various techniques, we deploy repeated airborne surveys to measure in-situ turbulent CO_2 exchange. The push propellor aircraft is equipped with a turbulent wind filed sensor, in combination with an open-path gas analyser for eddy covariance (EC) fluxes of momentum, sensible and latent heat and CO_2, augmented by onboard PAR (photosynthetic active radiation) and net radiation sensors. Survey altitude is 200ft/60m nominally, guaranteeing minimal flux divergence between surface and flight level. Covariances were spatially integrated over 2 km.
From 2020 till 2022 flights were made twice weekly, weather permitting, to cover three major fen meadow landscapes in the Netherlands: the so-called 'Groene Hart' area in the west, the 'Kop van Overijssel' and the South West of the province of Friesland more to the north. Flight patterns ensure full spatial coverage of the respective areas. Thus we created a unique airborne flux dataset, comprising 129 flights (till December 2022) that produced 11451 data records (2km integrated flux estimates, spatially distributed). In addition, data from two EC towers were included in the analysis. The present study focusses on the Groene Hart region only. We analysed CO_2 fluxes in relation to potential predictors from vegetation and soil characteristics, land and water management (EO and map based) and weather, using machine learning algorithms. We optimised three Boosted Regression Tree (BRT) models using either tower-based EC measurements only, airborne EC measurement only, or a combined dataset. The BRT model based on the combined complementary data sets outperformed the other two models in terms of correlation between observed and simulated CO_2 fluxes. The main drivers identified using Shapley values were not surprisingly and in order of importance PAR, humidity, temperature and available water storage capacity, the first two driving photosynthesis the latter two ecosystem respiration. Other explanatory variables include NDVI and specific land cover and soil classes that further modulate the CO_2 flux. Isolating the thus modelled effects of ground water on regional CO_2 emissions, we calculate that every 10 cm in ground water level rise causes a 3.3 ± 0.3 up to 4.0 ± 0.1 $tCO_2.ha^{-1}.yr^{-1}$ reduction in emissions, the variation depending on the calculation method chosen and boundary conditions set. Training the ML model on the full dataset we hope to further reveal and quantify additional predictors. We aim to ultimately provide data driven regional greenhouse gas balances for the different fen meadow areas of the Netherlands.

1 Introduction

Globally, peatlands store around 25% of all terrestrial carbon and are the most carbon-dense ecosystems of the terrestrial biosphere (Loisel, Gallego-Sala et al. 2021). In natural fens and bogs generally the emissions of methane from slow anaerobic decomposition of the accumulated organic material, is more than compensated for by the carbon dioxide uptake of its vegetation and subsequent sequestration in its peat soils (Frolking and Roulet 2007). Worldwide, organic soils have been and are exploited for fuel extraction, involving their drainage, followed by extracting, drying and burning the peat, and for agricultural use, generally requiring drainage leading to fast decomposition of the stored organic material. Thus, 250,000–290,000 km^2 of peat are drained for use as cropland and grassland, releasing from 680 to 1,030 Mt CO_2 annually (Evans, Peacock et al. 2021).

In the Netherlands extensive peat areas were transformed from a carbon sink to a carbon source due to centuries of peat extraction and of intensive agriculture and livestock farming that require low water tables (Akker et al., 2008). As a result of groundwater abstraction, the oxidation of organic material leads to increased carbon dioxide emissions – currently accounting for 2-3% of all Dutch emissions or 4.25 Mt carbon dioxide annually – and land subsidence, which, in turn, increases the need for further drainage (Kwakernaak, van den Akker et al. 2010, van den Akker, Kuikman et al. 2010). To counter this spiral, the Dutch government set a specific mitigation target for fen meadows: yearly emissions must be reduced by 1 Mton by 2030 (2019).

Counter measures to reduce such emissions generally include raising the water table somehow, deploying various techniques, such as pressurized drainage, ditch infiltration and others. Several meta studies have summarized the effects of water table depth manipulations on CO_2 and CH_4 emissions. (Evans, Peacock et al. 2021) combined data from 57 sites in the UK and Ireland with an additional 49 sites elsewhere in temperate and boreal climates, and fitted a single linear relation between effective water table depth and CO_2 emissions, and an exponential relation between the water table depth and methane emissions. Every 10 cm increase in water table depth (down to -100 cm) below about -20 cm, thus leads to 3.4 tCO_2 ha^{-1} extra annual emissions. (Tiemeyer, Freibauer et al. 2020) did the same for 118 sites in Germany, but fitted a saturating, non-linear (Gompertz) curve to their data, with an average slope between water table depth and CO_2 emissions of 6.7 tCO_2 ha^{-1} extra annual emissions for every 10 cm increase in water table depth (between 0 and -60cm, levelling off for deeper water tables). Neither author found any clear effect of land use on these relations, justifying the lumping of all their data together. At the same time this left a lot of variability unexplained.

All the datasets used in these meta studies have been based on site specific observations, using either chambers or flux towers on plots with well-defined, fairly homogeneous soil and vegetation characteristics and well-known water table management. However, they do not tell us the emissions at landscape level, where all three factors generally vary widely. Also, the effectiveness of any measure taken to manipulate emissions, probably well-proven in tightly controlled research settings, once scaled up may be undermined by relatively poor implementation or maintenance (Asselen, Jansen et al. 2023). In support of the development of a Tier 2 or even Tier 3 UNFCCC LULUCF reporting mechanism for emissions from organic soils (IPCC 2019) and as a means of verifying thus reported emissions there is a need for regional scale emission quantification methods (see Glossary for an explanation of abbreviations).

In-situ regional scale emissions can be directly measured through either of two airborne strategies, direct flux measurements or mass balance based on concentration measurements, possibly augmented by inversion methods (Meesters, Tolk et al. 2013, Butterworth, Desai et al. 2021, Shaw, Allen et al. 2022) (Vellinga, Gioli et al. 2010). Relating these to the underlying surface can be done by building environmental response functions by either more classical statistical methods (Hutjes, Vellinga et al. 2010) or by machine learning approaches (Metzger, Junkermann et al. 2013), both depending on overlaying the footprint of all flux measurements over maps of vegetation, land use

and soils and/or direct satellite derived products such as NDVI. However, airborne flux measurements are generally limited to daytime conditions, because at night mixing at some height often is insufficient to reflect surface conditions, and for aeronautical restrictions. Unlike airborne studies of energy fluxes that can be self-contained because night-time fluxes are small (sensible heat) or virtually zero (evaporation), night-time fluxes of CO_2 and other greenhouse gases are important, both for diurnal balances and arguably even more so for interpretation purposes, i.e. for flux partitioning approaches to separate Net Ecosystem Exchange (NEE) into Gross Primary productivity (GPP) and ecosystem respiration (R_{eco}) (Reichstein, Falge et al. 2005). Therefore, we consider complementary use of tower and airborne flux estimates essential to jointly assess regional scale fluxes, the first being continuous in time but spatially limited, the latter spatially extensive but temporarily discrete.

Thus the purpose of the present paper is to use a combination of airborne and tower based flux measurements to analyse CO_2 emissions from fen meadow *landscapes* in the Netherlands in relation to water table management.

2 Methodology

2.1 Study area

The Dutch peat areas have entirely been formed during the Holocene, reaching their maximum extent (about 50% of present Netherlands) around 4000 BP. Having originated largely in lagoons behind growing beach ridges, close to the nutrient supplying rivers (present or former channels) woody peat dominates on alluvial clays, while further away from these in nutrient poor conditions on marine clays (sphagnum) moss peat dominates. Much smaller areas of sedge or phragmites peat remain. Between 2000-1000BP large tracts eroded away again by a rising and repeatedly intruding sea, and since the medieval times it has been extracted by humans. Presently, some 325000ha of fens remain, mostly in the (south) west, where we are targeting the so-called Groene Hart region, in the north where we target southwest Friesland (Fryslân in the west Frisian language), and south of that where we target the Kop van Overijssel. The following descriptions refers to the area covered by the aircraft trajectories, i.e. by the measurement footprints (see next section).

2.1.1 Groene |Hart

The Groene Hart (the 'Green Heart' of the most urbanized region of the Netherlands) is the region we focus on in the present paper. It is roughly limited by Amsterdam in the northwest, Utrecht in the East and Rotterdam in the southwest. See figures 1 and 2. A large part of this rural area is below sea level, while the larger towns and cities are generally at or above sea level, historically founded on the levees of the rivers in the area. The area north of the Nieuwkoopse plassen (the crescent shaped lake north west of Woerden on the map in Figure 1) is among the deepest in the Netherlands at 5 - 6m below mean sea level (*bmsl*; more precisely 'below NAP' NAP being the national zero reference). The rest of the area covered by the aircraft is mostly 1.5 - 2m bmsl and the town and villages along the rivers 'Hollandse IJssel' and the 'Oude Rijn' are at 0-2m above mean sea level (amsl).

Peat dominates the Groene Hart area, especially towards the west and north. Towards the east alluvial clay deposits dominate, while in the far northwest sandy soils dominate. Few peaty soils can be found in the north, figure 1a and 2b. Median peat thickness is 2-3 m (Fig 2d) with equal areas with thicker (upto 6m) or thinner peat. Peat types include Weideveen[1], Koopveen (both woody peat with thin mineral top layer, the first more clayey, the latter more sandy) and Waardveen (moss peat with mineral top layer), with small remnants of Vlierveen (moss peat without mineral layer, see Fig 2c).

Of the three areas the Groene Hart is the most densely populated, as it is surrounded by the major Dutch urbanisations. Most of the built-up areas covered by the aircraft measurements are small towns and rural buildings, small enterprises and industry, roads and some greenhouses (towards the north west). Small arable areas can be found towards the north west and south east of the area, but grasslands used for dairy farming dominate the area. Open water in the form of many ditches, canals, rivers and some lakes covers a significant fraction of the area. Of the three areas, the Groene Hart has the smallest area of more or less natural fens or other designated nature areas.

[1] These class names are left untranslated from the Dutch National Soil Classification. See *Auxiliary data* section.

2.2 Airborne flux measurements

2.2.1 Equipment and processing

The aircraft used in this study is a SkyArrow 650 TCNS, equipped with a set of fast response turbulence instruments (3D wind by the Best Airborne Turbulence -BAT- probe and water vapor and carbon dioxide by and open-path gas analyser) complemented with a set of scalar instruments (net radiation, incoming and reflected photosynthetic active radiation, air and surface temperature). An Inertial GPS system combined with accelerometers recorded position and attitude of the aircraft. The aircraft and its data processing are described in detail in (Vellinga, Dobosy et al. 2013), a brief summary follows here.

The aircraft is operated under SPO and VFR regulations[2]. Measurement transects were flown at an altitude of 60 m (200 ft) above mean ground level, at a nominal airspeed of about 35 m/s. Banking never exceeded 15^{o} or data were filtered out if otherwise. Generally, the pilot flew alone, initially forced so by the corona pandemic, after all data acquisition was automated providing visual status indicators on the most crucial science instruments on a small (7") video screen on the aircraft instrument panel. Occasionally, also a task specialist (= scientific staff) flew in the passenger seat, taking notes of the landscape (e.g. fraction flooded, fraction of grasslands mown, or special events such biomass burning smoke plumes crossed) and/or the measurements on 17"monitor with real time data. When no task specialist flew, a 360^{o} sports camera mounted below the fuselage was flown.

Most instruments were sampled at either 50 or 20Hz and all signals stored in netcdf files, as controlled by a Raspberry Pi. Post flight processing started with de-spiking raw data (if needed at all) due to e.g. occasional rain droplets on the open-path trace gas analyser. Next, 50 Hz air pressure and temperature measured by the BAT probe were converted to 3D wind fields and corrected for all aircraft motions. In the final step covariances between wind and water vapor and CO_2 concentrations were calculated, spatially integrating these over 2000 m windows, and converted to fluxes using standard corrections for frequency response (Moncrieff, Massheder et al. 1997, Moncrieff 2004)and density effects (Webb, Pearman et al. 1980). Data were quality assessed and filtered according to the framework by (Foken 2004). In addition the most important meteorological scalar variables, potentially influencing CO_2 and heat fluxes (incoming and reflected photosynthetic active radiation (PAR), net radiation, surface temperature -from a longwave radiometer- and air temperature and humidity) were aggregated over the same spatial windows as the EC measurements.

2.2.2 Aircraft flight strategy

In the framework of the NOBV we adopted a flight strategy aimed at a complete spatial coverage of the respective landscapes (compare e.g. (Mauder, Desjardins et al. 2008)). To that end we designed a flight pattern with parallel transects more or less perpendicular to both the prevailing wind direction and to the main soil gradients. The transects were about 2 -3 km apart, though not necessarily strictly parallel as our nominal flying altitude forced us to avoid major built-up areas, 'Natura 2000' nature reserves, powerlines and wind turbines. This transect spacing was chosen such that the footprint (calculated following (Kljun, Calanca et al. 2015)) of the measurements cover the area in between the transects. As shown in Figure 1 the footprints do indeed cover most of the area due to varying wind directions. More importantly, all the major soil classes are covered, not only in mixed footprints but also in footprints covering a single soil class, either pure organic (peat), mixed (peaty) or mineral (clay soils to the north-west, sandy soils to the south-east). Average peat thickness in the footprints is around 2m, though also thin remnants and much thicker peat layers occur. Similarly, all important land use classes in the region occur in the footprints, though only 'grass' (pastures used in dairy farming) or open water (lakes) ever dominate the weighted footprint. Half the footprints are

[2] VFR = visual flight rules, allowing only daylight flights with good visibility; SPO = specialised operations documenting and certifying all non-standard flight operations

covered by more than 80% grass, a quarter by more than 90% and only 3% of the footprints are covered by more than 99% grass, see also Figure 2 (and SM1-4 for the other regions).

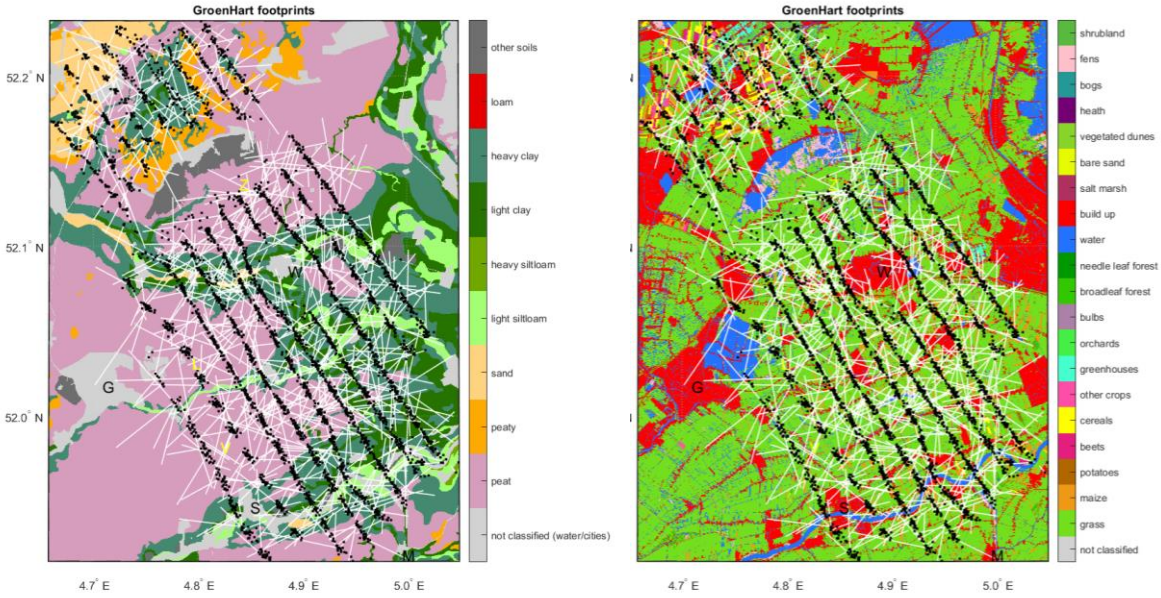

Figure 1: Soil (a, left) and Land use map (b, right) of the study area in the "Groene Hart", with linear representations of every fifth (in order to retain background map visibility) footprint of individual flux measurements. Black dots indicate the location of the aircraft, white lines extend upwind from there till the point where the footprint weight integral reaches 80%. Capital letters indicate the cities of G - Gouda, M - Meerkerk, S – Schoonhoven, W – Woerden.

Conditions and pilot availability permitting, 2 flights per week were made almost year-round. In the research period covered in this paper (2000-2022; measurements are still continuing) in total 129 flights were made distributed over the three areas as presented in **Error! Reference source not found.**. In **Error! Reference source not found.** and in the Supplementary Material more overviews are presented of typical observations. These figures clearly show a strong seasonal cycle in energy and CO_2 fluxes. Also within seasons considerable spread in e.g. light conditions (e.g. overcast vs. cloud free conditions), temperature and wind speed and direction was sampled.

Table 1 Statistics of all flights 2020-2022

Fen meadow area	Trajectory length	no of flights	no of observations
Groene Hart	262 km	54	4728
Kop van Overijssel	166 km	41	2972
Fryslân	312 km[1)]	34	3751

[1)] generally flown in two sections due to long ferry flight

Please note: Though descriptive data statistics cover all data, only data collected between March 2020 and December 2021 in the "Groene Hart" region, were used in the present ML analysis (given their processing and quality assessment status at the moment the present study was started). This dataset comprised 2639 airborne data records. For a definite paper of course we will use all available flight data for the respective regions.

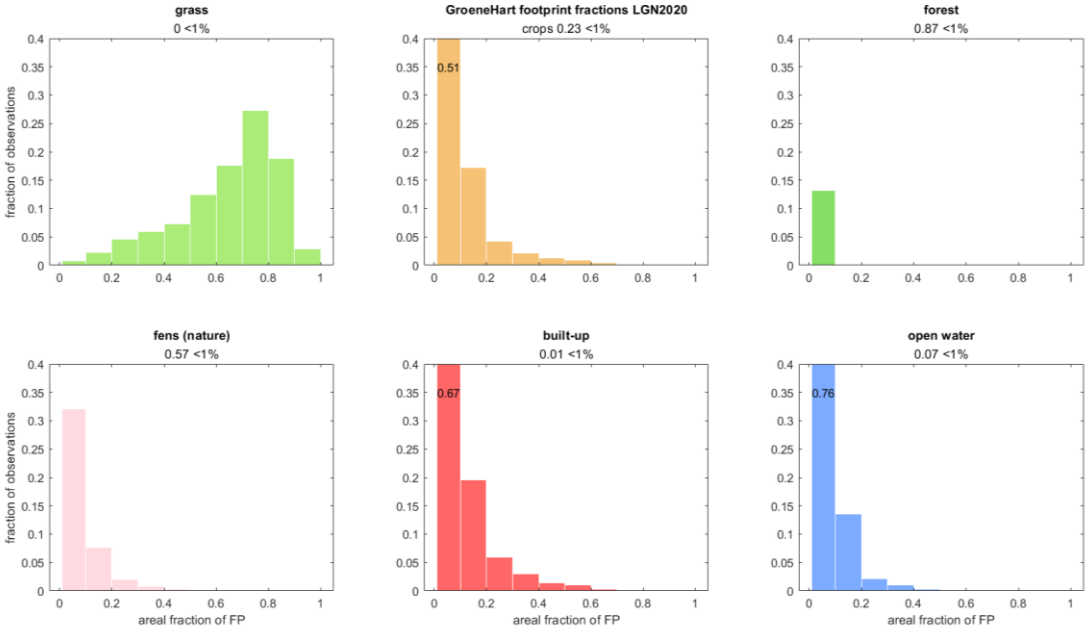

Figure 2a Histograms of footprint fractions of (merged [3]) land use classes in the "Groene Hart". Horizontal: fraction of each footprint covered by respective classes; vertical: fraction of all 4320 footprints of flights between March 2020 and December 2022. Bins: <1%, 1-10%, 10-20 %, 20-30%, etc.

[3] See section *Auxiliary data*

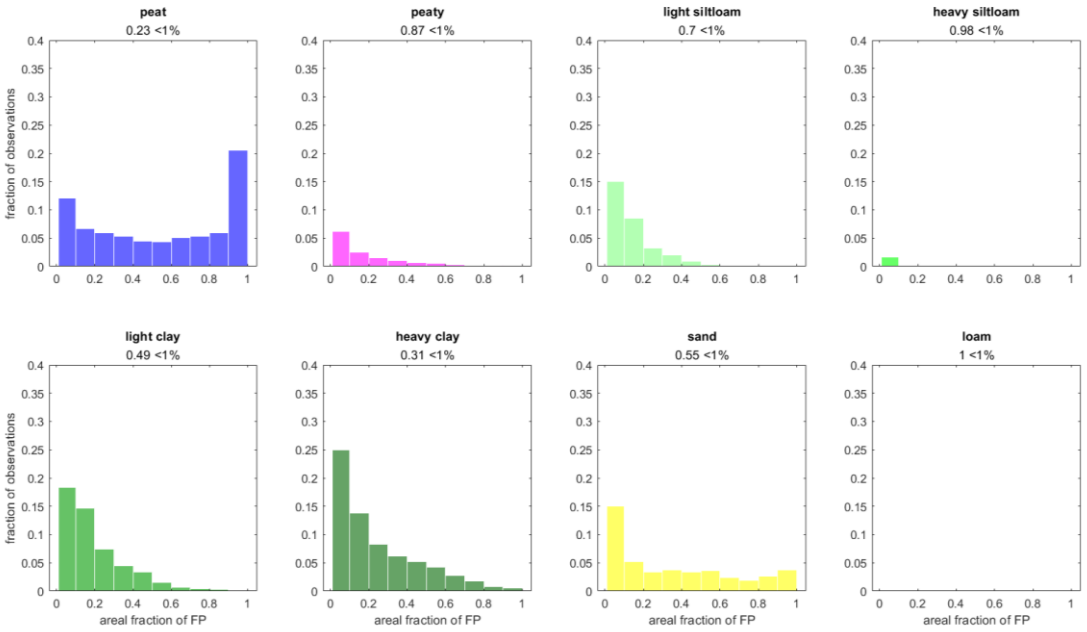

Figure 2b. Histograms of aircraft footprint fractions of (merged2) soil classes in the "Groene Hart". Otherwise as in Figure 2a.

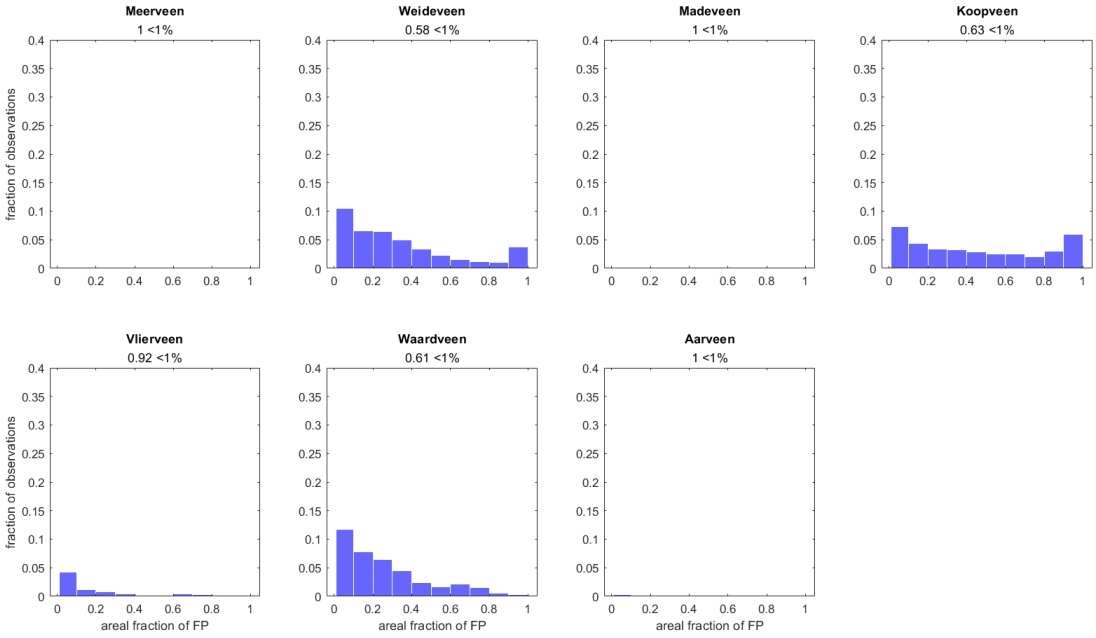

Figure 2c Histograms of aircraft footprint fractions of peat soil classes in the "Groene Hart". Otherwise as in Figure 2a

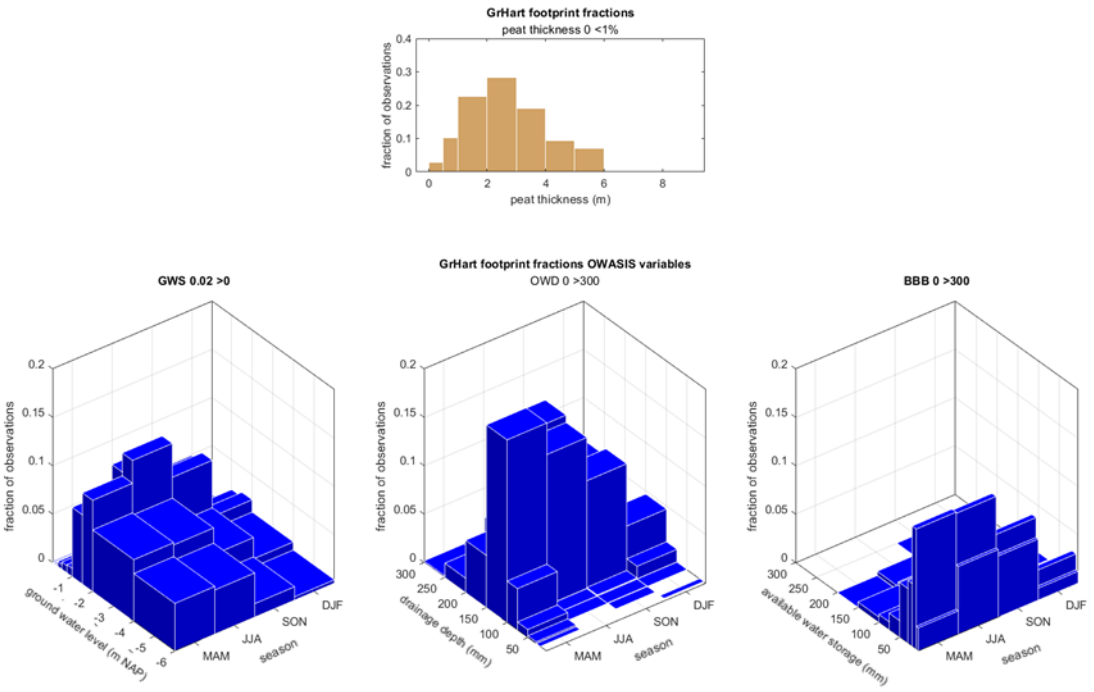

Figure 2d Histograms of aircraft footprint fractions of time varying OWASIS variables (Blue) and weighted footprint-average peat depth (brown) in the "Groene Hart". Otherwise as in Figure 2a. Note: Ground Water Levels (left graph) are relative to NAP: much of the area is 2-3 m below sea level, the area north of the Nieuwkoopse Plassen is 5-6 m below sea level.

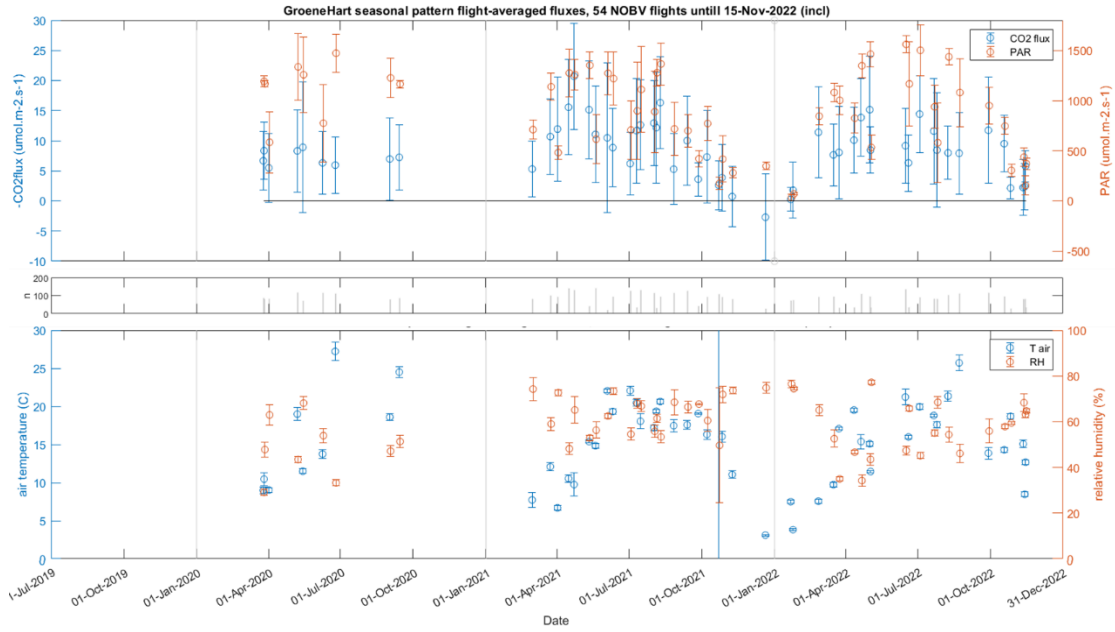

Figure 2 Flight overviews produced for the "Groene Hart" 1 Jan 2020 - 31 Dec 2022. Points in each graph present flight-averaged observations, plus standard deviation. Top graph: CO2 flux (blue, left axis; note we plot -CO2flux to show the co-evolution with PAR, so positive values indicate CO_2 uptake by the land scape) and PAR (brown-red, right axis). Grey bars in second graph present the number of data points in each flight. Third graph: air temperature (at flying altitude, blue, left axis) and relative humidity (brown-red, right axis).

2.3 Tower based flux data

In our analysis we combined aircraft data with tower data. The latter are highly complementary to the former in the sense that they are localised in space and continuous in time, which is especially important as they include also nighttime data (aircraft is only daytime!), which we expect to be important to parametrise the ecosystem respiration. In contrast aircraft data are (almost) continuous in space, but highly intermittent in time.

In the "Groene Hart" area two towers have been used for the present study: Langeweide and Zegveld (pasture, measure plot, not the typha plot on the same farm). In the present study only data between September 1st 2021 and January 31st 2022 were used (given their processing and quality assessment status, at the moment the present study was started). No attempts at gap filling were made; outliers defined as 0.5% highest and lowest values for CO_2 flux were removed, resulting in 7937 half hourly data records. For a full description of tower based flux data and their interpretation in this area see (Bart Kruijt 2023) for CO_2 and (Buzacott, Kruijt et al. 2023) for CH_4.

At these towers, fluxes of CO_2 (and CH_4, but not considered here), evaporation and sensible heat are measured with the eddy covariance method, alongside weather station measurements of PAR, 4 component radiation (shortwave and longwave, incoming and reflected), air temperature and humidity and rainfall and soil moisture and soil temperature.

2.4 Auxiliary data potentially explaining CO_2 flux variations

We want to relate the observed CO_2 fluxes to the most promising explanatory variables related to land and water management. The Netherlands is a very data rich country, so we have static maps available of

- Land use and land cover (Hazeu, Vittek et al. 2020): the LGN product is annually updated and maps land cover at 5 m resolution, comprising 48 classes. We reclassified those to 12 aggregated classes, see supplementary material. Fourteen classes were never in our footprints, and many more (e.g. crops other than maize) only in small fractions in a small number of footprints.
- Soil maps: the Dutch soil classification discerns 302 distinct soil classes based on complete soil profile information. (see https://app.pdok.nl/viewer/, then 'Overige kaarten – 'Basisregistratie Ondergrond (BRO)' – 'BRO Bodemkaart – Bodemvlakken'). The 302 original soil classes were reclassified to 21, distinguishing 9 different peat 'archetypes' and one peaty class, see supplementary material. For figure 1 (and SM1 and SM4) those were further merged into 1 peat class, only for display purposes.
- A digital elevation model is available for the Netherlands based on LIDAR altimetry at very high horizontal (0.5 m) and vertical resolution (0.05 m), the AHN3 (Lisa Keurentjes 2020), see https://www.ahn.nl/ahn-viewer. In this study a 5 m horizontal resolution product version was used.
- Peat thickness (F. de Vries 2014 and follow on reports)

In addition we use dynamic, time varying maps, that are generally earth observation (EO) based:

- The OWASIS product provides daily three related parameters at 250m resolution: Drainage depth (in Dutch 'ontwateringsdiepte – OWD, units: m), Available water storage capacity ('beschikbare bodemberging' – BBB: mm) and Groundwater Depth ('grondwater stand' - GWS: m relative to NAP). The data are based on radar precipitation, microwave and other EO data assimilated into the National Hydrological Model – LHM (van den Brink 2019, Spijker 2020). See also https://www.hydrologic.com/projects/owasis/
- NDVI derived from MODIS (MOD13Q1), at 250 m resolution and 16-day intervals (Didan 2015)

2.5 Machine learning training and optimisation

There are only a few other studies analysing airborne flux measurements with machine learning (Metzger, Junkermann et al. 2013, Serafimovich, Metzger et al. 2018, Vaughan, Lee et al. 2021). These studies focus on heat fluxes, relying only on aircraft data, and using static EO and map information. Yet, their methods serve as inspiration. All these studies promoted the usefulness and capabilities of the boosted regression trees (BRT) algorithm from the R Package by (Elith 2008), here we use the more modern Python package XGBoost (Extreme Gradient Boosting) (Chen 2016). BRT's can fit complex non-linear relationships, automatically handle interaction effects between features and are non-parametric, which means no prior assumption on the function is needed (Metzger, Junkermann et al. 2013). XGBoost potentially achieves high predictive performance (Nielsen 2016). Many more studies exist in which machine learning was used to simulate tower based fluxes (Knox, Sturtevant et al. 2015, Irvin, Zhou et al. 2021). In the present study we will use both airborne and tower based datasets to train and evaluate 3 distinct BRT models: one each for the two datasets used exclusively, and one using the combined data set. Thus we can assess the added value, if any, in terms of model performance of combining the complementary nature of these two data sets or not.

It is important to tune the model to the optimal level of flexibility for the present dataset. This was done by a) reducing the parameter space: 'feature selection', and b) optimizing the model's settings: 'hyperparameter tuning'.

In the current study, a hybrid feature selection approach was used, as is frequently done by studies that use XGBoost (Ogunleye 2019, Sang, Xiao et al. 2020, Prabha 2021). Similar to these studies, three feature selection methods were selected to determine the best subset of features: Pearson Correlation filter, Feature Importances in XGBoost, and Sequential Backward Floating Selection. First, Pearson correlation analysis (filter method) was done to make a rough selection of features. Here, (expected) correlations between the features and the response variable CO_2 flux are of interest, but strong correlations among the predictors should be prevented.

Second, Feature Importances were computed. XGBoost calculates the importance of features based on their share in important split-decisions, by averaging over the gain scores per split. The higher the gain average, the more important and effective the feature. Based on these feature selection methods, it is still possible that mutually correlated features score highly. Therefore, these first two steps serve as a pre-selection of features for the third method: Sequential Backward Floating Selection (SBFS, a wrapper method). SBFS includes an extra element compared to the more standard and widely used Sequential Backward Selection (SBS), and is known to give good results (Chandrashekar and Sahin 2014, Rodríguez-Pérez and Bajorath 2020). SBFS was run with 10-fold cross validation, the model used was a XGBoost tree with n estimators = 1000, learning rate = 0.05, max depth = 6, and subsample = 1. To avoid unequal representation in different folds of the k-fold cross validation, all datasets were shuffled beforehand. To evaluate which subset of features is optimal, the R^2, MSE (mean squared error), bias and variance of each model proposed by SBFS were computed on the test set. The R^2 was used as the final scoring metric.

After optimization of the feature subset, the following hyperparameters were optimized: number of trees (n estimators); maximum depth (tree complexity, max depth); learning rate (learning rate); and subsample ratio (subsample), the same as in Metzger et al. (2013). For every hyperparameter, multiple potential values were constructed and using a grid search with 10-fold cross validation (GridSearchCV from Scikit Learn) on the training set, the optimal combination of hyperparameters was found. Again, R^2 was used as scoring metric.

Throughout all steps, the data was split into a training set and a test set, containing 90% and 10% of the data, respectively. It is important that the test set is always the same, which can be attained by setting a 'random seed' in the splitting function. This way, the model is never evaluated using then same data it was trained on. Furthermore, all features in the training and test sets were normalized to have a mean of 0 and a standard deviation of 1.

2.6 Machine learning and model interpretation approach

Understanding the behaviour of a ML model in more physical terms is not a trivial task. Here we use two approaches: the SHAP framework and simulations along prescribed physical gradients.

The unified framework SHAP (SHapley Additive exPlanations), was developed to address the (often) difficult interpretation of 'black box' machine learning models (Rodríguez-Pérez and Bajorath 2020). By computing the marginal contribution of each feature value to the predicted output, SHAP gives insight in the model's behaviour (Lundberg and Lee 2017). SHAP is based on Shapley values; they describe the contribution of every feature to the final model outcome, given the combined contribution of all other features. One by one each feature is denied its contribution to the model by assigning random values to it, as this results in no added predictive power. Comparison of models output with and without the contribution of this feature thus isolates its contribution to the model outcome. In the current study, Shapley values thus analyse the importance and contribution of features of the best performing BRT model to the simulated CO_2 fluxes: a negative Shapley value indicates a negative contribution to the flux, meaning more uptake (more photosynthesis) or less emissions (less respiration), and vice versa for a positive Shapley value.

In the context of drained fen meadows, it is well established that the most important drivers of emissions are water table related (see Introduction). So, to further interpret model behaviour the best performing BRT model was used to simulate CO_2 emissions along a gradient of the OWASIS – BBB parameter (available water storage capacity) for several distinct and otherwise constant meteorological settings (PAR, air humidity and surface temperature).

3 Results

The following are tentative results, for the Groene Hart region only, based on partial datasets available at the start of this analysis. Analysis will be repeated and extended with more complete datasets covering multiple years, and for the other two NOBV regions.

3.1 Feature selection and model optimisation

First, correlations between features are examined by use of correlation matrices(Fig SM 7). These show, not surprisingly, considerable correlation between the meteorological variables, mostly due to their seasonal co-variation. Nevertheless, the primary ones were retained and only Vapour Pressure Deficit was omitted as it represents (almost) the same, though inverse information as contained in Relative Humidity. Next, the three OWASIS parameters also show strong mutual correlation and based on considerations explained below only available water storage capacity (BBB) was retained. Grassland dominated most footprints, and since the presence of any other land cover class will be at the cost of Grassland, it (inversely) correlates somewhat with the other landcover classes, see also **Error! Reference source not found.**a. Similarly, peat dominates the area and its subclasses (Waardveen - kV, Weideveen – pV and Koopveen - hV) anti correlated for the same reason, see also Figure 2c. Nevertheless, in the Airborne and Merged models several classes of each land cover and soil were retained; obviously not so in the Tower model as these do not vary for a fixed location.

Next, the XGBoost based Feature Importance for the predictand CO2 flux is examined, see **Error! Reference source not found.**. Firstly, it should be noted for the Tower dataset only the time evolving parameters have any importance. Also the static parameter the fraction of Grassland (Grs) in the footprint of the two towers apparently differs somewhat and thus exhibits some importance. Any other land use or soil classes show up only in the Airborne and Merged datasets. The three OWASIS parameters score very high, preceded only by PAR, confirming already their importance for the CO_2 flux, i.e. the respiration flux. Also NDVI ranks high in all datasets explaining variability in GPP. Thus the first 9 features, except GWS and VPD (because they are highly correlated to other parameters already) and Grs, were passed on to the SBFS selection algorithm for the Tower dataset and a further 14 parameters for the Airborne and Merged dataset.

The Sequential Backward Feature Selection retained only 5 parameters for the Tower dataset, dropping more features compromised the model performance. The R^2 peaked for 11 features in the Airborne datasets. The R^2 peaked for either 6 or 13 features in the Merged datasets, but MSE for the latter is slightly lower. To retain maximum meaningful surface information, but also to maintain the most parsimonious model, we continue with two feature subsets for the Merged model, with 6 and 13 features respectively. See Figure SM8.

Hyperparameter tuning involved four parameters, of which one optimized to the same value for all four models, i.e. the Tree Max depth was set to 9. For the other three hyperparameters that have been optimized for each model variant, see Table 2.

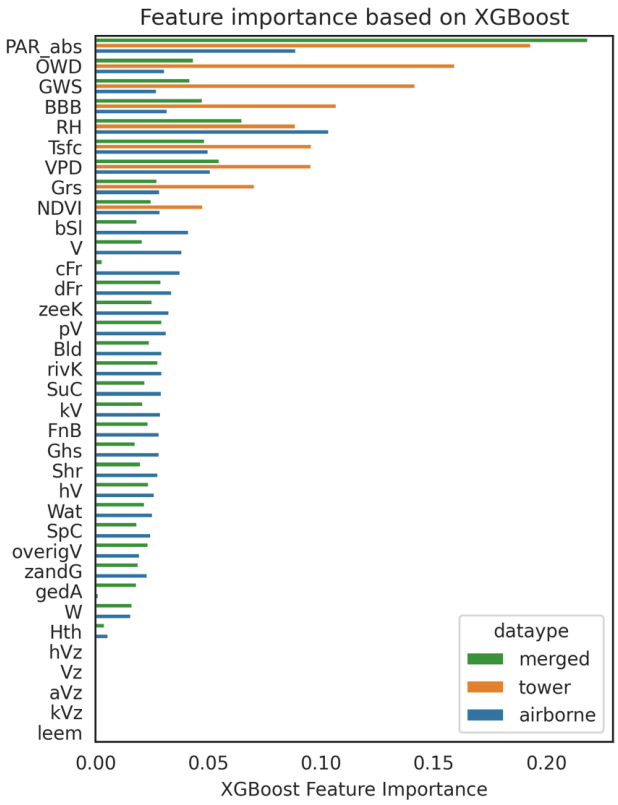

Figure 4 Feature importance for all three data sets: Tower only, Airborne only, or Merged.

Table 2 Final selected features after SBFS, the tuned hyperparameters of optimized models and the performance of the best model in each category.

Model	**Selected features after SBFS**	**Hyperparameters**			**Metrics**	
		learning rate	sub sample	no. of trees	R^2	MSE
Tower	PAR_abs, Tsfc, RH, NDVI, BBB	0.001	0.8	4000	0.47	30.9
Airborne	PAR_abs, Tsfc, RH, NDVI, Bld, dFr, rivK, hV, zeeK, kV, W	0.001	0.55	4000	0.35	44.0
Merged 6	PAR_abs, Tsfc, RH, BBB , Grs, SpC	0.005	0.8	1000	0.61	31.0
Merged 13	PAR_abs, Tsfc, RH, NDVI, BBB, SuC, SpC, Wat, dFr, rivK, pV, zeeK, V	0.001	0.8	7000	0.61	30.8

3.2 Interpretation of best performing model behaviour

Considering the highest R2 scores, but also model parsimony, the merged model with 6 features is preferred and therefore used for all subsequent analyses. Figure 5 shows an overview of all Shapley values for the merged model with 6 features. According to these Shapley values, the importance of the features in the model is ranked as follows: PAR abs, Tsfc, RH, BBB, Grs and SpC, identical to their importance ranking in Figure 4. It shows that PAR abs has the widest range of Shapley values,

from -25 to 15, where low values of PAR_abs have a positive contribution to the flux and high values of PAR abs a negative contribution, as expected for the main driver of GPP. The same pattern, but within a smaller range, is visible for RH. For Tsfc, on the other hand, the opposite is the case: higher Tsfc values lead to more positive contributions, consistent with it controlling ecosystem respiration more than it does control GPP. For BBB there is no clearly visible pattern from this particular visualisation, but the next one will elucidate why. Low values of Grs result in a positive contribution to the flux (apparently alternative land use is generally a CO_2 source, i.e. build-up area or open water), whereas low values of fractional area of spring crops (SpC: cereals found in the north of the domain, see Figure 1) result in no contribution.

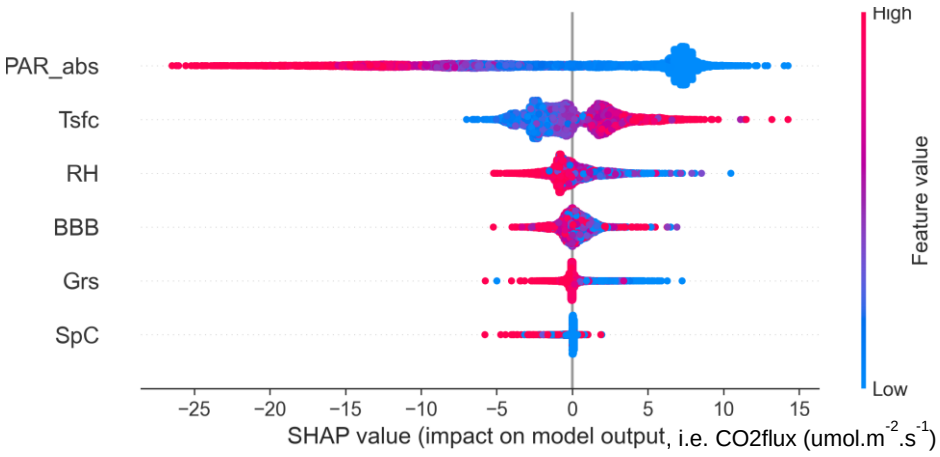

Figure 5 Shapley values for the Merged-6 feature model, showing their importance and contribution to the flux. The colour represents the magnitude of the feature value, whereas the x-axis shows the corresponding Shapley value.

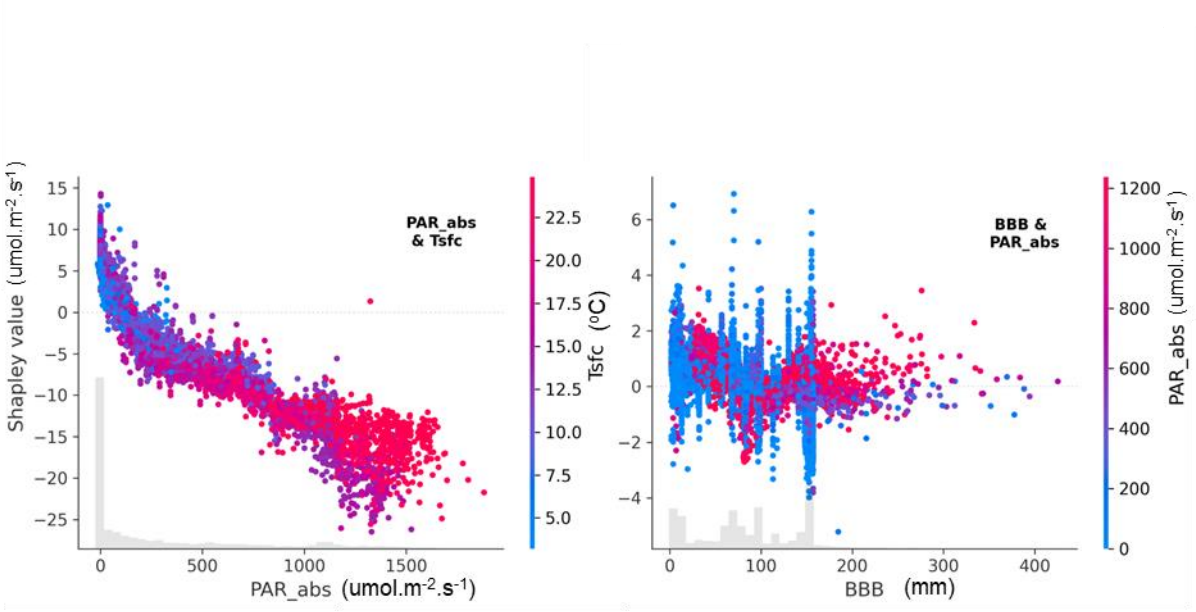

Figure 6 Shapley scatter plots for the Merged-6 model: Shapley value on the y-axis indicating the contribution to the flux of the feature on the x-axis, coloured by a third second feature. Left: the relation between PAR_abs and CO2flux coloured by Tsfc. Right: the relation between BBB and CO2 flux, coloured by PAR_abs.

Figure 6 shows Shapley scatter plots of PAR abs and BBB. The PAR curve on the left is very similar to well-known light response curves, here a kind of regionally effective light response curve. It should be noted that the Shapley value is much better constrained by PAR than a direct plot of raw data of the CO_2 flux on PAR, the latter exhibiting much more spread (not shown). It can be seen that with higher PAR, the contribution to the predicted flux is more negative and higher Tsfc values increase the contribution in the positive direction. The Shapley values of BBB remain spread around zero contribution, independent of the BBB value. This graph is hard to interpret, although it is already clear that trends may differ in sign along various subranges of observed BBB values, possibly explaining the lack of patterns in Figure 5.

The relation becomes clearer as we filter the dataset for higher PAR values. Figure 7 shows Shapley values of BBB when filtered for PAR values larger than 800, coloured by temperature. BBB has a clearer influence on the CO_2 flux now. However, a positive correlation between BBB and Tsfc apparent: low BBB occurs at low Tsfc and vice versa, indicative of their covariation over the course of autumn (BBB decreasing as the water table rises, while temperatures drop). The variation in Shapley value is not simply due to the variation in Tsfc: if all variance in flux would be captured by Tsfc, the Shapley value of BBB would be zero, which is clearly not the case.

For BBB up to 90mm, increasing BBB leads to more uptake, shown by negative contribution to the flux, probably because it covaries with better seasonal growing conditions. For BBB between 90 and 150 mm, increasing BBB leads to more emissions. For BBB above 150mm, the relation is ill defined, if only due to a low number of data records in this BBB range. To quantify the trend between BBB = 90 - 150mm, a linear regression was fitted. The t-statistics and p-values indicate that both intercept and slope are significantly different from 0; the slope indicates that for every mm more water storage capacity available in the soil, additionally 0.024 ± 0.002 µmol m^{-2} s^{-1} of CO_2 is emitted, or 333 ± 28 kg CO_2 ha^{-1} yr^{-1}.mm^{-1}.

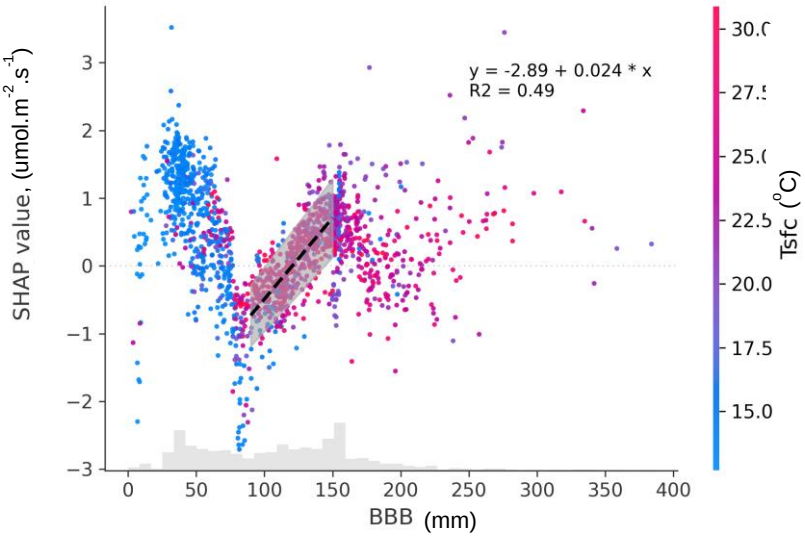

Figure 7 Shapley values of BBB when PAR > 800 for the Merged-6 model.

Another way to interpret the model is to make simulations for a set of prescribed conditions and thus factoring out the various relationships, otherwise obscured in a black-box BRT. We did so for BBB values varying from 0-300mm and for different but otherwise constant values of PAR and Tsfc. Shown in Figure 8 at left are the results for PAR_abs = 0 and three relatively low Tsfc values of 0, 5 and 10oC respectively. The same figure at right shows results for PAR_abs values of 1200 and 1600 µmol.m-2.s-1, combined with temperatures of 15 and 20oC respectively. In both figures a linear regression was fitted to all data points with 0 mm < BBB < 150 mm at left and to all data points for 90 mm < BBB < 270 mm in the right side of the figure. The upper limits in these data ranges are

set because the model is ill constrained for these values due to lack of data (hence the straight horizontal lines at the right-end of these graphs). All regression coefficients significantly ($p < 0.001$) differed from zero. The 95% interval for the slope in the figure at left is 0.027 ± 0.001 μmol.m^{-2}.s^{-1}.mm^{-1}, or 37 5 ± 14 kgCO2.ha-1.yr^{-1}.mm^{-1} . For the slope in the figure at right the slope is 0.029 ± 0.001 μmol.m^{-2}.s^{-1}.mm^{-1}, or 403 ± 14 kgCO2.ha^{-1}.yr-1.mm^{-1}.

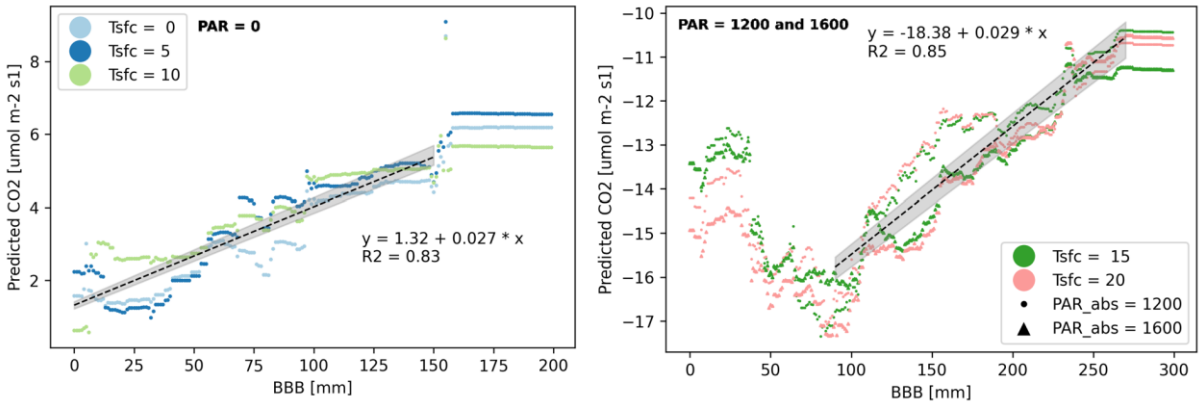

Figure 3 Simulations for the Merged-6 model for PAR = 0 (left) and PAR = 1200 and 1600 (right) for different temperatures. See text for further explanation.

4 Discussion

The analysis presented in this paper has been made on a limited dataset only. Though the results demonstrate the power of the approach, much may change when applied to the full dataset that is, longer in time and also includes the other two regions and towers in different localities. Part of the following discussion will already influence this future analysis on the extended dataset.

4.1 Methods

4.1.1 Airborne flux measurements

In the present study airborne fluxes were calculated by direct spatial integration of covariances (Crawford, Mcmillen et al. 1993, Gioli, Miglietta et al. 2006, Vellinga, Dobosy et al. 2013). While we used a fixed 2 km integration window, in some conditions other window lengths may be preferable (Sun 2018). Using typical averaging times, tower-based measurements suffer from spectral losses in lower frequencies, and basically the 2 km averaging windows for the aircraft fluxes potentially suffers from the same limitation. However, spatially resolved measurements such as our airborne eddy covariance measurements in principle can detect also larger scale (meso-β, meso-γ) transport. Alternative to block averaging, fluxes can be separated into smaller-scale turbulent and larger-scale mesoscale contributions using wavelet scalogram like approaches (Mauder, Desjardins et al. 2007) (Metzger, Junkermann et al. 2013, Paleri, Desai et al. 2022), analysing full transects or even flights, instead of short windows of these. One interpretation is that the turbulent component then reflects better the surface flux, that needs to be separated from the mesoscale flux, before relating it to any surface characteristics. All these studies analysed sensible and latent heat fluxes though, and it remains to be seen how relevant such approaches are to fluxes of CO_2 or other trace gases. In a follow-on study we intend to pay further attention to this aspect.

4.1.2 Data handling and model optimisation

Airborne and tower data have different qualities: airborne data is spatially continuous but temporally limited, and tower data is spatially stationary but temporally continuous. However, while this complementarity should be beneficial in principle, its practicalities may lead to some spurious correlations among features and predictands. Both datasets differ in temporal coverage, 6 months for the tower data, 18 months for the airborne data, yet the latter has only 2624 records and the former 7937. The land use and soil classes are fixed for the tower data and variable for the aircraft data. To demonstrate the possible artifact this may create in the model consider the following. A non-negligible correlation is apparent (see Fig SM 7) between soil class hV (koopveen) and the CO_2 flux. This is probably due to the fact that one of the towers is on this soil type, and the fact that only tower data include positive, i.e. night time fluxes. Aircraft data on the other hand generally represent only small fractions of hV soils, but are always negative, because they include day-time fluxes only. Two options are available to potentially overcome such artifacts and to increase the likelihood of identifying relations between soil and or land use classes and the CO_2 flux. First, the inclusion of data from more towers on different soil and/or land use types will be beneficial. Secondly, after all percentages of soil classes (or land use classes) in the footprints have been obtained, a clustering method such as K-means clustering or hierarchical clustering can be used to create often-occurring combinations of soil (land use) classes, potentially increasing their predictive power. Still, on one hand in the feature selection the threshold for model improvement, set to prevent un-parsimonious models, should be further optimised. On the other hand, a requirement for mandatory features can be set to maintain interpretability.

Data filtering, while desirable for the analysis, may have similarly unwanted effects. For example, aircraft measurements are ‘polluted’ by data whose footprint contains significant fractions of built-up areas. We could simply filter these out. Similarly, we are interested in peat oxidation, so we could

filter out data whose footprints contains significant fractions of mineral soils. However, both conditions would filter out aircraft data only and none of the tower data, thus greatly limiting the models potential to learn spatial relationships. A filter can have the advantage of specifying or constraining the model, but the disadvantage of decreasing the size of the training data. We prefer to leave either data set as much as possible intact, and use the trained model to separate out, i.e. simulate such idealised conditions of e.g. grassland on peat only.
To make both datasets comparable in size, we can reduce the tower dataset by random sampling, possibly by stratification of e.g. day or night time fluxes, to make it similar in size to the aircraft dataset. More careful considerations of such analysis and/or data handling alternatives, with consequently trade-offs between interpretability vs performance, will surely be part of future work.

4.1.3 Additional data

Above we already mentioned the very real possibility that the aircraft data are literally polluted by the presence of fossil fuel based CO_2 emissions from domestic heating, transport, greenhouses and small industries (and probably in that order of importance in our aircraft footprints). While the ML algorithm might be able to identify these sources because they feature on the land use map, the temporal emission profiles (both diurnal and seasonally) between these four categories differ and therefore may prohibit proper attribution. Adding to the predictors time varying fossil fuel based emission maps may ease this problem and improve the relations found for the sources of our interest, i.e. the peat oxidation. Possibilities are gridded 1 km resolution emission maps, together with temporal emission profiles per category (Guevara, Petetin et al. 2022),
Also, for temporal vegetation dynamics arguably better alternatives are available to the MODIS product used in the present version of this paper. Based on the GroenMonitor a special product for grasslands has been developed, the Grassland Monitoring Service (Roerink 2021), providing parcel-level, 10 m resolution of NDVI identifying mowing events. It is based on the DMC satellite constellations providing 3.5m resolution data with a 1 day revisit time.

4.1.4 Model evaluation

The final model, combining data from both aircraft and towers explains 61% of the observed variance. This seems acceptable given the complex interactions analysed and random noise levels typical for Eddy Covariance observations. Compared to studies also modelling CO2 fluxes but by traditional methods, this R2 is in the same range (Jung, Reichstein et al. 2011, Zulueta, Oechel et al. 2011, Dou 2018). However, studies exploiting machine learning approaches tend to have higher R2 (Metzger, Junkermann et al. 2013, Serafimovich, Metzger et al. 2018, Vaughan, Lee et al. 2021), but these all focussed on simulating heat fluxes, arguably a simpler process to analyse. Also, it appears that neither of these studies used separate data subsets for learning and evaluation, respectively, like we did. Evaluating the model on the same (complete) dataset it was trained on, would also increase the R2 for our best model from 0.61 to 0.86.

4.1.5 Model interpretation and analysis

The Shapley based relationship between BBB and CO_2 flux are not straightforward to interpret, unlike the PAR and Tsfc based Shapley values. Only filtering for certain conditions reveals more clear patterns (**Error! Reference source not found.**), though still not a monotonous relation. The decreasing trend for BBB<90mm is probably due to another co-variate that dominates in this range. Also the relation for BBB>150mm though based on fewer data points (and therefore presently neglected), if real might hint at a more saturating relation between emissions and increasing BBB like suggested by (Tiemeyer, Freibauer et al. 2020). Some of this complex behaviour may also result from the fact that R_{eco}, does not only include respiration of old peat material, the prime interest of the NOBV programme, but also of young above and below ground litter, exudates and simply autotrophic respiration. The latter component is significant and strongly driven by GPP (Kruijt et al. 2023).
The simulations to quantify the same relation were constrained to several PAR, respectively Tair values. For night-time conditions (PAR=0) a monotonous relation emerges, unlike for daytime

conditions, the latter being more similar in shape to the Shapley's and probably for the same reasons. The relations in Figure 3 show a kind of staircase like pattern, probably reflecting a series of splits the BRT algorithm makes to the data set. It might be useful to further analyse these trees to understand the reason for these splits and possibly replacing them with the aim to create more monotonous linear relations between soil moisture and fluxes. In principle, simulations for at least a full annual cycle of actual meteorology, and preferably for several decades, can be used to quantify a more climatological mean relation between the same emissions and BBB.

4.2 Implications

4.2.1 The ground water table as a driver of emissions

The major finding of the present tentative analysis is the quantitative relation between BBB and CO_2 emissions we found for the area analysed.

In order to relate the numbers found to parameters that are influenced by water management practices and to numbers reported in literature, we need to translate the BBB values to (effective) water table depth. The OWASIS framework itself provides at least a first step in that process. It's OWD parameter (ontwateringsdiepte) claims to represent ground water, including effects like bulging (winter) or hollow (summer) groundwater tables in parcels surrounded by ditches and thus is not necessarily identical to the more readily measurable and manageable ditch water level itself. For now we assume that OWD is a good proxy for the effective water table depth.

BBB and OWD are strongly correlated, and their correlation at first sight seems not to depend on soil classes as may have been expected, given known differences in porosity for different soil textures. Although the relation appears to be non-linear, we assume it to be linear since we do not know yet how to interpret the differences in the BBB interval over which the linear relationships to CO_2 flux hold, as depicted in figures 7 and 8.

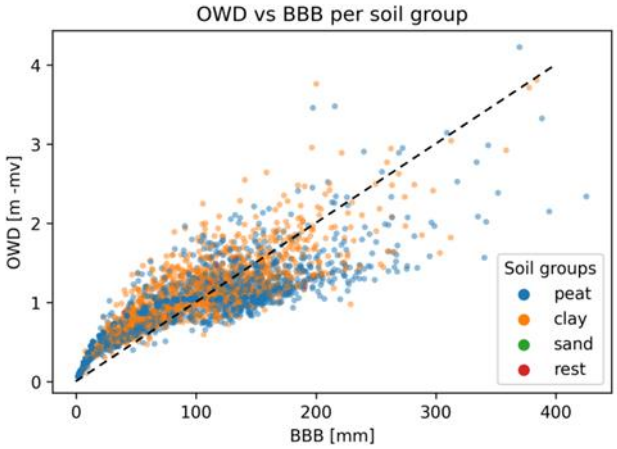

Figure 9 Correlation between all observed values of BBB (mm) and OWD (m below surface) in the merged tower and airborne datasets. Colouring by major soil groups, showing no apparent differentiation in the relation between the two. The straight line is the 1:10 relationship, i.e. 100 mm BBB equals 1 m OWD.

Assuming a 1:10 relationship between the two, i.e. 100 mm BBB equals 1 m OWD, implies that the slope of the regression lines of CO_2 flux on BBB, from figures 7 and 8, translate to an emission increase of 33.3 tCO_2 ha^{-1} yr^{-1} m^{-1} drainage depth up to 40.3 tCO2 ha^{-1} yr^{-1} m^{-1} drainage depth. These numbers are very comparable to those derived by other international studies based on annual budgets, e.g. 34.0 tCO_2 ha^{-1} yr^{-1} m^{-1} over a 100 cm WT range (Evans, Peacock et al. 2021), 22

tCO_2 ha^{-1} yr^{-1} m^{-1} over a 50 cm WT range assuming saturation occurs for WTR below 50 cm (Tiemeyer, Freibauer et al. 2020). They also are in line with those derived from closely related NOBV datasets, but using entirely different analytical and data aggregation methods (Boonman et al, 2023), 30 tCO2 ha^{-1} yr^{-1} m^{-1} over a 150 cm WT range from only tower based EC observations (Bart Kruijt 2023), as well as instrumentally different data sources, i.e. 38.02 tCO_2 ha^{-1} yr^{-1} m^{-1} from chamber data only (Ralph Aben 2023), Fritz et al. 2023). Noteworthy to realize that the Evans and Tiemeyer studies are based on annual mean fluxes, claiming that the influence of seasonal dynamics of water levels on the CO_2 flux is limited. Thus their relations represent spatial dependencies. In the present study, the relationship between BBB and the CO_2 flux does not discriminate between spatial or temporal variations of BBB. Further analysis may clarify which dimension contributes mostly to the results. It then also remains to be seen whether any further optimized model can discriminate between different relations for e.g. different peat types.

4.2.2 Use of the current approach in the context of mitigation policies

One objective of NOBV has not been addressed with the current approach, namely the assessment of the effectiveness of different interventions to manipulate the water table, alternative land uses or other emission reducing measures like clay additions. As long as these are experimental and are implemented on small plots only, the present approach is not applicable. If otherwise evaluated positively and implemented at large scale, however, then airborne observations may be valuable to assess their effectiveness in real life situations that may differ from well controlled research plots.

In the present study we analysed only the effect of one driver on CO_2 emissions in more detail. However we strongly believe the current approach, i.e. data sets and extended model, is potentially much more valuable and might reveal also effects of e.g. peat depth (in relation to water table), or peat type, or land use. Apart from using a ML model to evaluate the effect of certain drivers as done here, another important possibility is to use that same model to simulate spatially distributed fluxes and thus create a flux map of the area either for some idealized constraints, or potentially also simulating a full year using some meteorological reanalysis product, such as ERA5. The results of such work may serve to verify LULUCF reported emissions, or support further development of spatially explicit reporting methods. These will be objectives of future work.

5 Conclusions

Complementing multi-site ground-based measurements using various techniques, we deploy repeated airborne surveys to measure in-situ turbulent CO_2 exchange over Dutch fen meadow landscapes. We thus created a unique airborne flux dataset, comprising 129 flights (till December 2022) that produced 11451 data records (2km integrated flux estimates, spatially distributed. The present study focusses on the Groene Hart region only. The other areas will follow soon.
In an exploratory analysis,

- we optimised three Boosted Regression Tree models using either tower based EC measurements only, airborne EC measurement only, or a combined dataset. The BRT model based on the combined complementary data sets outperformed the other two models in terms of correlation between observed and simulated CO_2 fluxes: R^2 = 0.61 for the model trained on the combined dataset, compared to R^2=0.35 for the model trained solely on airborne data or R^2 = 0.47 for the model trained solely on tower data from 2 sites.
- we identified the main drivers of CO_2 exchange using Shapley values were not surprisingly and in order of importance PAR, humidity, temperature and available water storage capacity, the first two driving photosynthesis the latter two ecosystem respiration. Other explanatory variables include NDVI and specific land cover and soil classes that further modulate the CO_2 flux.
- we calculate that every 10 cm in ground water level rise causes a 3.3 ± 0.3 up to 4.0 ± 0.1 $tCO_2.ha^{-1}.yr^{-1}$ reduction in emissions, isolating the thus modelled effects of ground water on regional CO_2 emissions. The variation depends on the calculation method chosen and boundary conditions set. This relation seems to hold over an approx. 150 cm range in groundwater levels.

Acknowledgements

This study was part of the Netherlands Research Programme on Greenhouse Gas Dynamics of Peatlands and Organic Soils (NOBV), which was launched in 2019 by the Dutch ministry of Agriculture, Nature management and Food quality (LNV) as part of the Climate Agreement. Its objective is to research the effectiveness of measures in peatland areas and to be able to better predict emission levels. The effect on subsidence is also researched. The programme is directed by the Foundation for Applied Water Research (STOWA). The research is conducted by Wageningen University (WU), Wageningen Environmental Research (WENR), Vrije Universiteit Amsterdam (VU), Utrecht University (UU), Radboud University, Deltares research institute.

References

- Asselen, S. v., S. Jansen, R. Aben, R. Hessel, J. J. H. v. d. Akker, H. Massop and P. Gerritsen (2023). "Effects of subsoil water infiltration systems on phreatic groundwater levels in peat meadows." in prep.
- Bart Kruijt, A. B., Tom Heuts, Hanne Berghuis, Jan Biermann, Merit van den Berg, Ron Lootens, Reinder Nouta, Laurent Bataille, Ronald Hutjes, Wilma Jans, Quint van Giersbergen, Wietse Franssen, Niek Bosma, Ype van de Velde, Christian Fritz (2023). "Carbon dioxide emissions from peatlands in The Netherlands: drivers of variability in Eddy covariance fluxes " in prep.
- Butterworth, B. J., A. R. Desai, P. A. Townsend, G. W. Petty, C. G. Andresen, T. H. Bertram, E. L. Kruger, J. K. Mineau, E. R. Olson, S. Paleri, R. A. Pertzborn, C. Pettersen, P. C. Stoy, J. E. Thom, M. P. Vermeuel, T. J. Wagner, D. B. Wright, T. Zheng, S. Metzger, M. D. Schwartz, T. J. Iglinski, M. Mauder, J. Speidel, H. Vogelmann, L. Wanner, T. J. Augustine, W. O. J. Brown, S. P. Oncley, M. Buban, T. R. Lee, P. Cleary, D. J. Durden, C. R. Florian, K. Lantz, L. D. Riihimaki, J. Sedlar, T. P. Meyers, D. M. Plummer, E. R. Guzman, E. N. Smith, M. Sühring, D. D. Turner, Z. Wang, L. D. White and J. M. Wilczak (2021). "Connecting Land–Atmosphere Interactions to Surface Heterogeneity in CHEESEHEAD19." Bulletin of the American Meteorological Society **102**(2): E421-E445.
- Buzacott, A., B. Kruijt, C. Fritz, Q. v. Giersbergen, M. v. d. Berg, Y. v. d. Velde and K. v. Huissteden (2023). "Methane emission on agricultural and semi-natural peatland sites in the Netherlands." in prep.
- Chandrashekar, G. and F. Sahin (2014). "A survey on feature selection methods." Computers & Electrical Engineering **40**(1): 16-28.
- Chen, T., & Guestrin, C. (2016). (2016). "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 785–794.
- Crawford, T. L., R. T. Mcmillen, R. J. Dobosy and I. Macpherson (1993). "Correcting Airborne Flux Measurements for Aircraft Speed Variation." Boundary-Layer Meteorology **66**(3): 237-245.
- Didan, K. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid, University of Arizona, Alfredo Huete - University of Technology Sydney and MODAPS SIPS - NASA.
- Elith, J., Leathwick, J. R., Hastie, T. (2008). "A working guide to boosted regression trees." Journal of Animal Ecology **77**: 802-813.
- Evans, C. D., M. Peacock, A. J. Baird, R. R. E. Artz, A. Burden, N. Callaghan, P. J. Chapman, H. M. Cooper, M. Coyle, E. Craig, A. Cumming, S. Dixon, V. Gauci, R. P. Grayson, C. Helfter, C. M. Heppell, J. Holden, D. L. Jones, J. Kaduk, P. Levy, R. Matthews, N. P. McNamara, T. Misselbrook, S. Oakley, S. E. Page, M. Rayment, L. M. Ridley, K. M. Stanley, J. L. Williamson, F. Worrall and R. Morrison (2021). "Overriding water table control on managed peatland greenhouse gas emissions." Nature **593**(7860): 548-552.
- F. de Vries, D. J. B., B. Kempen, F. Brouwer, A.H. Heidema (2014). Actualisatie bodemkaart veengebieden; Deelgebied 1 en 2 in Noord-Nederland. Basis Registratie Ondergrond (BRO, Actualisatie bodemkaart Wageningen, Alterra.
- Foken, T., M. Goeckede, M. Mauder, L. Mahrt, B. Amiro, and W. Munger, (2004). Post-field data quality control. . Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis. W. M. X. Lee, and B. Law, Kluwer Academic Publishers. **29:** 181-208.
- Frolking, S. and N. T. Roulet (2007). "Holocene radiative forcing impact of northern peatland carbon accumulation and methane emissions." Global Change Biology **13**(5): 1079-1088.
- Gioli, B., F. Miglietta, F. P. Vaccari, A. Zaldei and B. De Martino (2006). "The Sky Arrow ERA, an innovative airborne platform to monitor mass, momentum and energy exchange of ecosystems." Annals of Geophysics **49**(1): 109-116.
- Hazeu, G. W., M. Vittek, R. Schuiling, J. D. Bulens, M. H. Storm, G. J. Roerink, W. M. L. Meijninger, G. W. Hazeu, M. Vittek, R. Schuiling, J. D. Bulens, M. H. Storm, G. J. Roerink and W. M. L. Meijninger (2020). LGN2018 : een nieuwe weergave van het grondgebruik in Nederland. Wageningen, Wageningen Environmental Research.
- Hutjes, R. W. A., O. S. Vellinga, B. Gioli and F. Miglietta (2010). "Dis-aggregation of airborne flux measurements using footprint analysis." Agricultural and Forest Meteorology **150**(7–8): 966-983.
- IPCC (2019). 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories,. E. Calvo Buendia, Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize S., Osako,

A., Pyrozhenko, Y., Shermanau, P. and Federici, S. (eds). Switzerland, The Intergovernmental Panel on Climate Change (IPCC).

• Irvin, J., S. Zhou, G. McNicol, F. Lu, V. Liu, E. Fluet-Chouinard, Z. Ouyang, S. H. Knox, A. Lucas-Moffat, C. Trotta, D. Papale, D. Vitale, I. Mammarella, P. Alekseychik, M. Aurela, A. Avati, D. Baldocchi, S. Bansal, G. Bohrer, D. I. Campbell, J. Chen, H. Chu, H. J. Dalmagro, K. B. Delwiche, A. R. Desai, E. Euskirchen, S. Feron, M. Goeckede, M. Heimann, M. Helbig, C. Helfter, K. S. Hemes, T. Hirano, H. Iwata, G. Jurasinski, A. Kalhori, A. Kondrich, D. Y. F. Lai, A. Lohila, A. Malhotra, L. Merbold, B. Mitra, A. Ng, M. B. Nilsson, A. Noormets, M. Peichl, A. C. Rey-Sanchez, A. D. Richardson, B. R. K. Runkle, K. V. R. Schäfer, O. Sonnentag, E. Stuart-Haëntjens, C. Sturtevant, M. Ueyama, A. C. Valach, R. Vargas, G. L. Vourlitis, E. J. Ward, G. X. Wong, D. Zona, M. C. R. Alberto, D. P. Billesbach, G. Celis, H. Dolman, T. Friborg, K. Fuchs, S. Gogo, M. J. Gondwe, J. P. Goodrich, P. Gottschalk, L. Hörtnagl, A. Jacotot, F. Koebsch, K. Kasak, R. Maier, T. H. Morin, E. Nemitz, W. C. Oechel, P. Y. Oikawa, K. Ono, T. Sachs, A. Sakabe, E. A. Schuur, R. Shortt, R. C. Sullivan, D. J. Szutu, E.-S. Tuittila, A. Varlagin, J. G. Verfaillie, C. Wille, L. Windham-Myers, B. Poulter and R. B. Jackson (2021). "Gap-filling eddy covariance methane fluxes: Comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands." Agricultural and Forest Meteorology **308-309**: 108528.

• Knox, S. H., C. Sturtevant, J. H. Matthes, L. Koteen, J. Verfaillie and D. Baldocchi (2015). "Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO2 and CH4) fluxes in the Sacramento-San Joaquin Delta." Global Change Biology **21**(2): 750-765.

• Kwakernaak, C., J. J. H. van den Akker, E. M. Veenendaal, J. van Huissteden and P. Kroon (2010) "Mogelijkheden voor mitigatie en adaptatie Veenweiden en klimaat." Bodem 2010 (2010) juni; ISSN: 0925-1650.

• Lisa Keurentjes, M. d. J., Khaled Alhoz, Kristof Kenesei, Manos Papageorgiou (2020). GEO1101.2020 – AHN3. Synthesis Project in Geomatics. Delft, Delft University of Technology.

• Loisel, J., A. V. Gallego-Sala, M. Amesbury, G. Magnan, G. Anshari, D. Beilman, J. Benavides, J. Blewett, P. Camill and D. Charman (2021). "Expert assessment of future vulnerability of the global peatland carbon sink." Nature climate change **11**(1): 70-77.

• Lundberg, S. M. and S.-I. Lee (2017). "A unified approach to interpreting model predictions." Advances in neural information processing systems **30**.

• Mauder, M., R. L. Desjardins and I. MacPherson (2007). "Scale analysis of airborne flux measurements over heterogeneous terrain in a boreal ecosystem." JOURNAL OF GEOPHYSICAL RESEARCH **112**: D13112.

• Meesters, A. G. C. A., L. F. Tolk, W. Peters, R. W. A. Hutjes, O. S. Vellinga, J. A. Elbers, A. T. Vermeulen, S. van der Laan, R. E. M. Neubert, H. A. J. Meijer and A. J. Dolman (2013). "Inverse carbon dioxide flux estimates for the Netherlands." J. Geophys. Res. **117**(D20306).

• Metzger, S., W. Junkermann, M. Mauder, K. Butterbach-Bahl, B. Trancón y Widemann, F. Neidl, K. Schäfer, S. Wieneke, X. H. Zheng, H. P. Schmid and T. Foken (2013). "Spatially explicit regionalization of airborne flux measurements using environmental response functions." Biogeosciences **10**: 2193–2217.

• MinEZK (2019). National Climate Agreement of the Netherlands. 00000001003214369000. D. G. f. C. a. E. Ministery of Economic Affairs and Climate. The Hague.

• Moncrieff, J. B., J. M. Massheder, H. deBruin, J. Elbers, T. Friborg, B. Heusinkveld, P. Kabat, S. Scott, H. Soegaard and A. Verhoef (1997). "A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide." Journal of Hydrology **189**(1-4): 589-611.

• Moncrieff, M. W. (2004). "Analytic Representation of the Large-Scale Organization of Tropical Convection." Journal of the Atmospheric Sciences **61**(13): 1521-1538.

• Ogunleye, A., & Wang, Q.-G. (2019). "XGBoost Model for Chronic Kidney Disease Diagnosis." Transactions on Computational Biology and Bioinformatics **5963**(100): 1-10.

• Paleri, S., A. R. Desai, S. Metzger, D. Durden, B. J. Butterworth, M. Mauder, K. Kohnert and A. Serafimovich (2022). "Space-Scale Resolved Surface Fluxes Across a Heterogeneous, Mid-Latitude Forested Landscape." Journal of Geophysical Research: Atmospheres **127**(23): e2022JD037138.

• Prabha, A., Yadav, J., Rani, A., Singh, V. (2021). "Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier." Comput. Biol. Med **136**: 104664.

• Reichstein, M., E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havránková, H. Ilvesniemi, D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, J.-M. Ourcival, J. Pumpanen, S. Rambal, E. Rotenberg, M. Sanz, J. Tenhunen, G. Seufert, F. Vaccari, T. Vesala, D. Yakir and R. Valentini

(2005). "On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm." Global Change Biology **11**(9): 1424-1439.

- Rodríguez-Pérez, R. and J. Bajorath (2020). "Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions." Journal of computer-aided molecular design **34**: 1013-1026.
- Sang, X., W. Xiao, H. Zheng, Y. Yang and T. Liu (2020). "HMMPred: Accurate Prediction of DNA-Binding Proteins Based on HMM Profiles and XGBoost Feature Selection." Computational and Mathematical Methods in Medicine **2020**: 1384749.
- Serafimovich, A., S. Metzger, J. Hartmann, K. Kohnert, D. Zona and T. Sachs (2018). "Upscaling surface energy fluxes over the North Slope of Alaska using airborne eddy-covariance measurements and environmental response functions." Atmos. Chem. Phys. **18**(13): 10007-10023.
- Shaw, J. T., G. Allen, P. Barker, J. R. Pitt, D. Pasternak, S. J.-B. Bauguitte, J. Lee, K. N. Bower, M. C. Daly, M. F. Lunt, A. L. Ganesan, A. R. Vaughan, F. Chibesakunda, M. Lambakasa, R. E. Fisher, J. L. France, D. Lowry, P. I. Palmer, S. Metzger, R. J. Parker, N. Gedney, P. Bateson, M. Cain, A. Lorente, T. Borsdorff and E. G. Nisbet (2022). "Large Methane Emission Fluxes Observed From Tropical Wetlands in Zambia." Global Biogeochemical Cycles **36**(6): e2021GB007261.
- Sun, Y. J., L.; Chen, Q.; Zheng, C. (2018). "Optimizing Window Length for Turbulent Heat Flux Calculations from Airborne Eddy Covariance Measurements under Near Neutral to Unstable Atmospheric Stability Conditions." Remote Sensing **10**(670).
- Tiemeyer, B., A. Freibauer, E. A. Borraz, J. Augustin, M. Bechtold, S. Beetz, C. Beyer, M. Ebli, T. Eickenscheidt, S. Fiedler, C. Förster, A. Gensior, M. Giebels, S. Glatzel, J. Heinichen, M. Hoffmann, H. Höper, G. Jurasinski, A. Laggner, K. Leiber-Sauheitl, M. Peichl-Brak and M. Drösler (2020). "A new methodology for organic soils in national greenhouse gas inventories: Data synthesis, derivation and application." Ecological Indicators **109**: 105838.
- van den Akker, J. J. H., P. J. Kuikman, F. de Vries, I. E. Hoving, M. Pleijter, R. F. A. Hendriks, R. J. Wolleswinkel, R. T. L. Simões and C. Kwakernaak (2010) "Emission of CO2 from agricultural peat soils in the Netherlands and ways to limit this emission." Proceedings of the 13th International Peat Congress After Wise Use – The Future of Peatlands, Vol. 1 Oral Presentations, Tullamore, Ireland, 8 – 13 june 2008; ISBN: 9780951489048.
- Vaughan, A. R., J. D. Lee, S. Metzger, D. Durden, A. C. Lewis, M. D. Shaw, W. S. Drysdale, R. M. Purvis, B. Davison and C. N. Hewitt (2021). "Spatially and temporally resolved measurements of NOx fluxes by airborne eddy covariance over Greater London." Atmos. Chem. Phys. **21**(19): 15283-15298.
- Vellinga, O. S., R. J. Dobosy, E. J. Dumas, B. Gioli, J. A. Elbers and R. W. A. Hutjes (2013). "Calibration and Quality Assurance of Flux Observations from a small Research Aircraft." Journal of Atmospheric and Oceanic Technology **30**: 161-181.
- Vellinga, O. S., B. Gioli, J. A. Elbers, A. A. M. Holtslag, P. Kabat and R. W. A. Hutjes (2010). "Regional carbon dioxide and energy fluxes from airborne observations using flight-path segmentation based on landscape characteristics." Biogeosciences **7**(4): 1307-1321.
- Webb, E. K., G. I. Pearman and R. Leuning (1980). "Correction of flux measurements for density effects due to heat and water vapour transfer." Quarterly Journal of the Royal Meteorological Society **106**(447): 85-100.

Glossary

amsl	above mean sea level
BAT probe	Best Air Turbulence, a device measuring 3D wind fields from differential pressures
BBB	available water holding capacity (an OWASIS product, in Dutch: beschikbare bodem berging)
Bld	built-up areas (class in land cover map)
bmsl	below mean sea level
BRT	boosted regression trees, a catogory of machine learning algorithms
dFr	deciduous forests (class in land cover map)
EC	eddy covariance, a micrometeorological technique to measure turbulent fluxes
EO	earth observation/ satellite data
GPP	gross primary productivity
Grs	grasslands (class in land cover map)
GWS	ground water level (relative to NAP, an OWASIS product, in Dutch grondwater stand)
hV,	woody peat with thin mineral top layer (class in soil map; from Dutch Koopveen)
kV	moss peat with mineral top layer (class in soil map ; from Dutch Waardveen)
LULUCF	Landuse Landuse Change and Forestry, a category of sources in emission inventories
ML	machine learning
NAP	Dutch national topographical zero altitude reference
NEE	net ecosystem exchange, i.e. net CO2 flux, equals GPP-R_{eco}
NOBV	Dutch National Research Programme on Greenhouse Gases in Peatlands
NDVI	normalised difference vegetation index, a proxy for amount of vegetation
OWD	drainage depth (relative to field level, an OWASIS product, in Dutch ontwaterings diepte)
PAR	photosynthetic active radiation
PAR_abs	absorbed photosynthetic active radiation (i.e. incoming minus reflected PAR)
pV	woody peat with thin clayey mineral top layer (class in soil map; in Dutch Weideveen)
RH	relative humidity
rivK	alluvial clays (class in soil map ; from Dutch rivier klei)
R_{eco}	ecosystem respiration
SBFS	Sequential Backward Floating Selection, a feature selection method to build ML models
SpC	spring crops (class in land cover map)
SuC	summer crops (class in land cover map)
Tsfc	surface temperature
UNFCCC	United Nations Framework Convention on Climate Change
V	moss peat without mineral layer (class in soil map ; from Dutch Vlierveen)
VPD	vapour pressure deficit, a measure of air humidity
W	peaty soils (class in soil map)
Wat	surface water (class in land cover map)
zeeK	marine clays (class in soil map; in Dutch zee klei)

6 Supplementary Material

6.1 Description other flight areas

6.1.1 Fryslân

The Frisian peat areas studied in NOBV can be found mostly between Wolvega in the south east, Drachten in the north east, Sneek in the north, Koudum in the west and Lemmer in the south. See figures SM1 and SM2. Again, much of the area is below sea level. The western area is mostly at 0.5 -1.5m bmsl. East of Joure, between Heerenveen and Drachten much of the area is 1.5 – 2.5m bmsl.

Peat in the study area can be found in two broad areas: a first 10km wide zone running from south west to north east, bordering the marine clays to the north west. Then a second area to the south of our domain that actually continues into the “kop van Overijssel” area, described in the next section. In-between these to peat zones as well towards to east we find sandy soils with considerable areas of peaty soils, see Figure SM1and SM2b. Hardly anywhere peat thickness exceeds 2m, Figure SM2d . Koopveen dominates the peat areas with also some Weideveen. Waardveen is found here much less than in the Groene Hart. Instead we do find small areas of Meerveen (a sandy humus rich mineral layer overlying eutrophic woody peat) and Madeveen (a peaty top layer overlying eutrophic woody peat).

Of the three areas this region is probably the most water rich with a large number of substantial lakes, Figure SM1 andSM2a. It is less dense populated than the Groene Hart reflected in smaller built-up areas, with the city of Heerenveen in the centre of the flight tracks. More than in the Groene Hart but less than in in the Kop van Overijssel we find still natural fens or other designated nature areas. Of the three areas Fryslân is the most extensively and deepest drained, Figure SM2D.

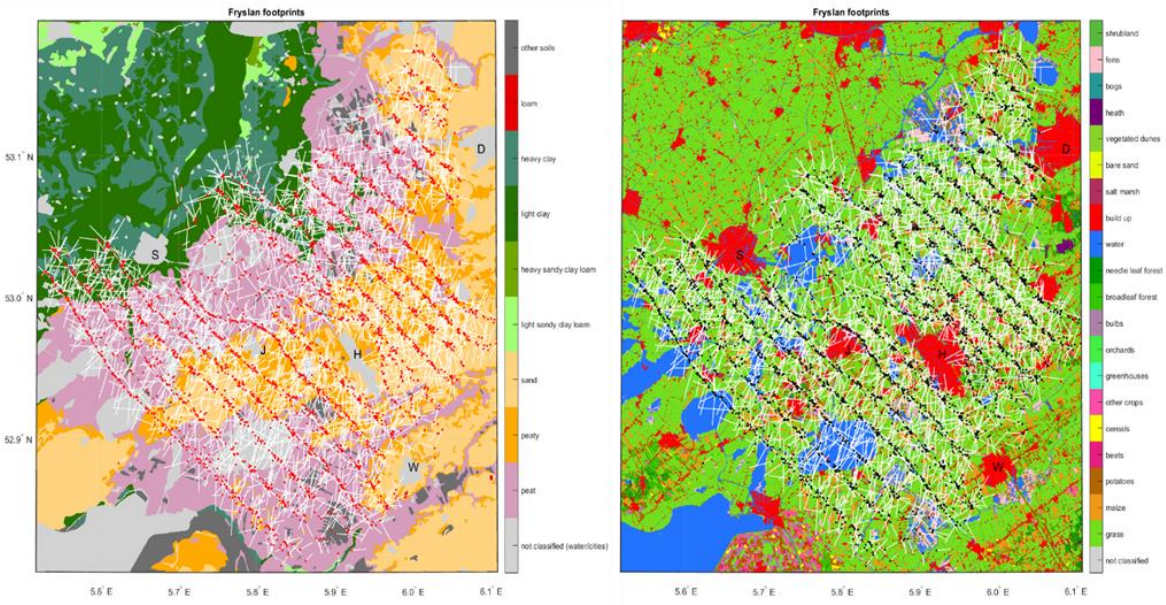

Figure SM1: Soil (left) and Land use map (right) of the study area in the Province of “Fryslân”, with linear representations of footprints of individual flux measurements. Red / black dots indicate the location of the aircraft, white lines extend upwind from there till the point where the footprint weight integral reaches 80%. Capital letters indicate the cities of D – Drachten, H – Heerenveen, J – Joure, S – Sneek, W – Wolvega.

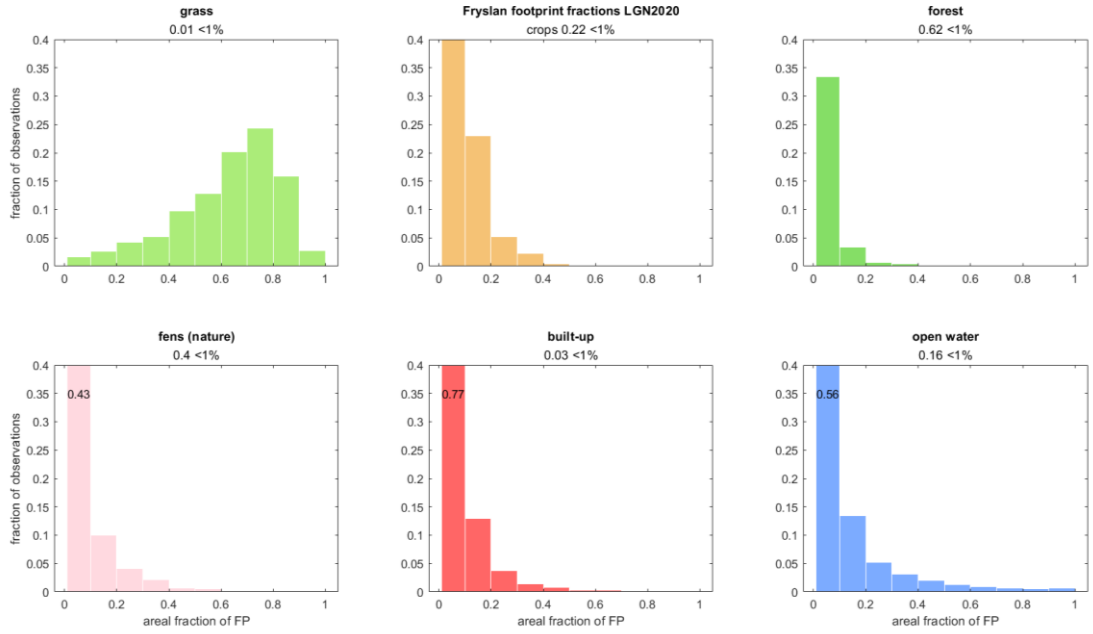

Figure SM2a Histograms of footprint fractions of (merged) land use classes in "Fryslân". Horizontal: fraction of each footprint covered by respective classes; vertical: fraction of all 3751 footprints. Bins: <1%, 1-10%, 10-20 %, 20-30%, etc.

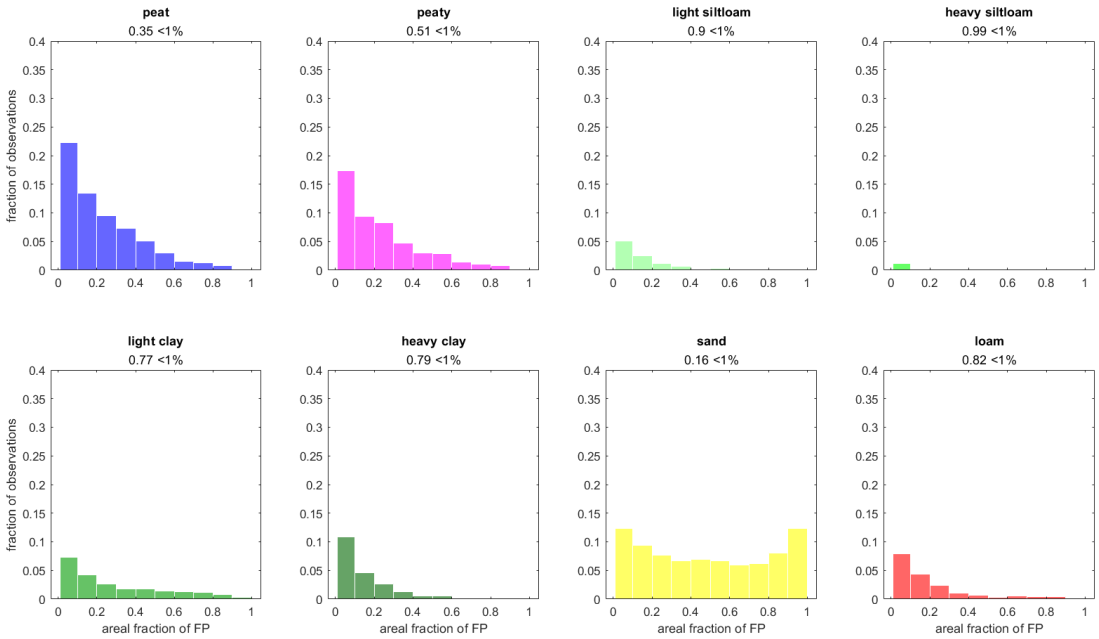

Figure SM2b Histograms of footprint fractions of (merged) land use classes in "Fryslân". Otherwise as in Figure SM2a

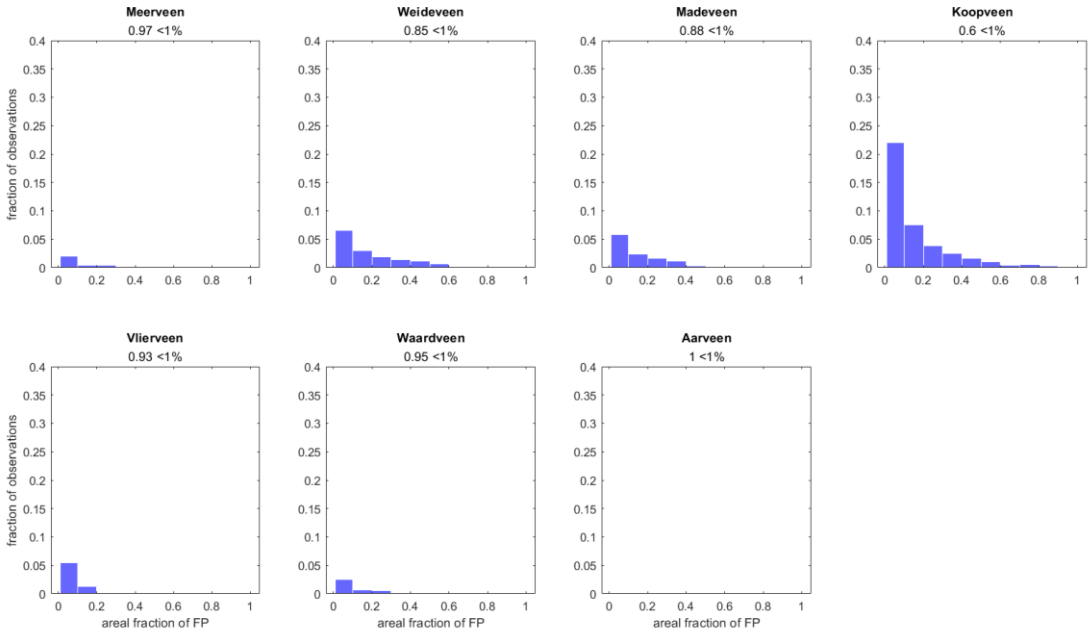

Figure SM2c Histograms of footprint fractions of peat type in "Fryslân". Otherwise as in Figure SM2a

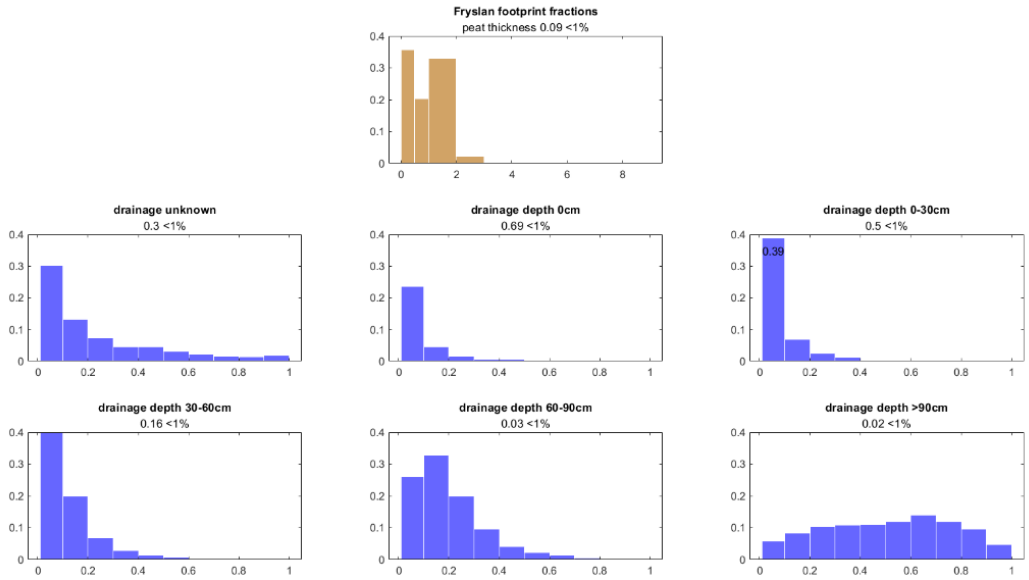

Figure SM2d Histograms of footprint fractions of drainage classes (Blue) and weighted footprint-average peat depth (brown) in the "Fryslân". Otherwise as in Figure SM2a.

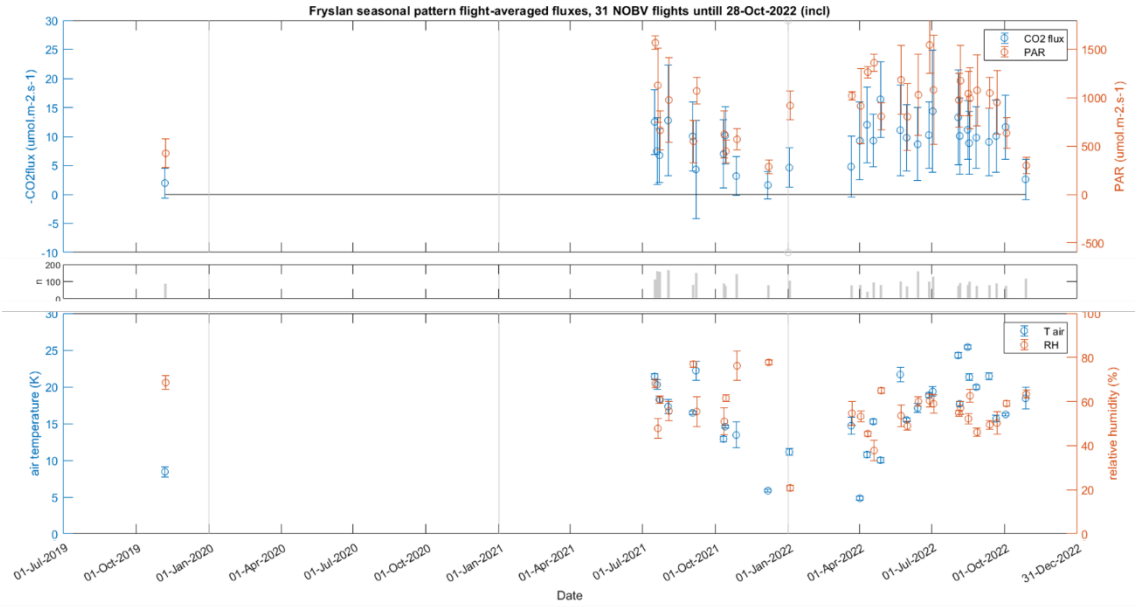

Figure SM3. Flight overviews produced for Fryslân[4] . Points in each graph present flight-averaged observations, plus standard deviation. Top graph: CO2 flux (blue, left axis, note we plot minus CO2flux to show the co-evolution with PAR, so positive values indicate CO2 uptake by the land scape) and PAR (brown-red, right axis). Grey bars in second graph present the number of data points in each flight. Third graph: air temperature (at flying altitude, blue, left axis) and relative humidity (brown-red, right axis).

[4] Data acquisition started latest of the three areas, due to delays in obtaining flight permissions.

6.1.2 Kop van Overijssel

This area is roughly limited by Meppel and Staphorst in the north and west, by Zwolle in the south and by Kampen in the west. See figures SM3 and SM4. Much of the area is below sea level, though less so than the Groene Hart. A 5 km circular depression east of IJsselmuiden is as deep as 3m bmsl, much of the Weerribben is around 1.5-2 m bmsl (not covered by the aircraft measurements, but represented by an EC tower), while most of the area covered by the aircraft is 0.5-1 m bmsl.

The research area under the flight tracks is dominated by peat. Towards the east a narrow zone of peaty soils marks the transition to the sandy soils of eastern Netherlands. Towards the south and west we find alluvial clays along the delta branches of the river “IJssel” and “het Zwarte Water”. Most of the peat is less than 2m thick, though some 10% of the footprints has peat upto 3m thick. Of the three areas this one has the largest fractions of Madeveen and Meerveen.

A few large lakes can be found in the north west of the domain surrounded by extensive nature areas of fens and cap holes, where large tracts of reed fields are harvested each winter to be used for thatched roofs. Also this area is less densely populated than the Groene Hart reflected in fewer built-up areas.

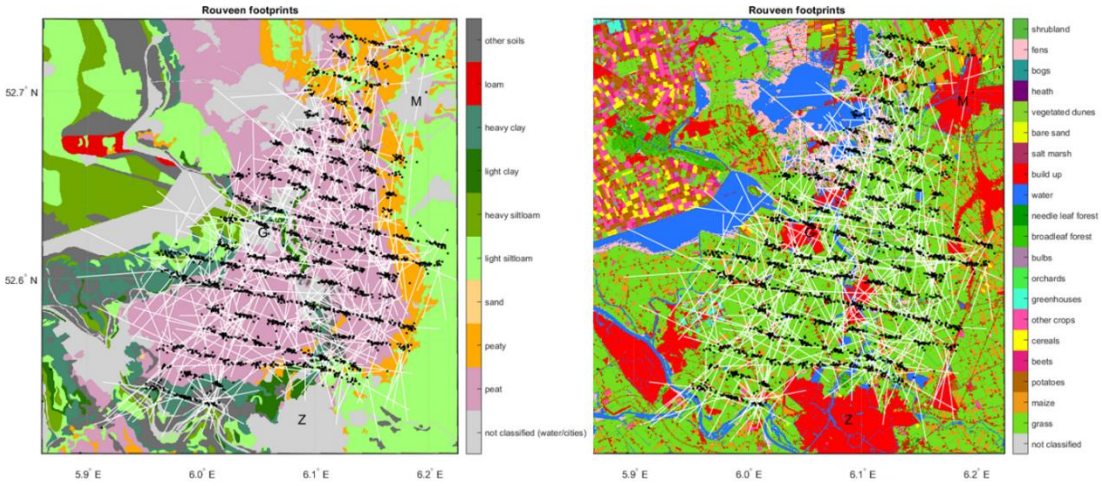

Figure SM4: Soil (left) and Land use map (right) of the study area in the Province of “Kop van Overijssel”, with linear representations of footprints of individual flux measurements. Red / black dots indicate the location of the aircraft, white lines extend upwind from there till the point where the footprint weight integral reaches 80%. Capital letters indicate the cities of G – Genemuiden, M - Meppel, Z - Zwolle.

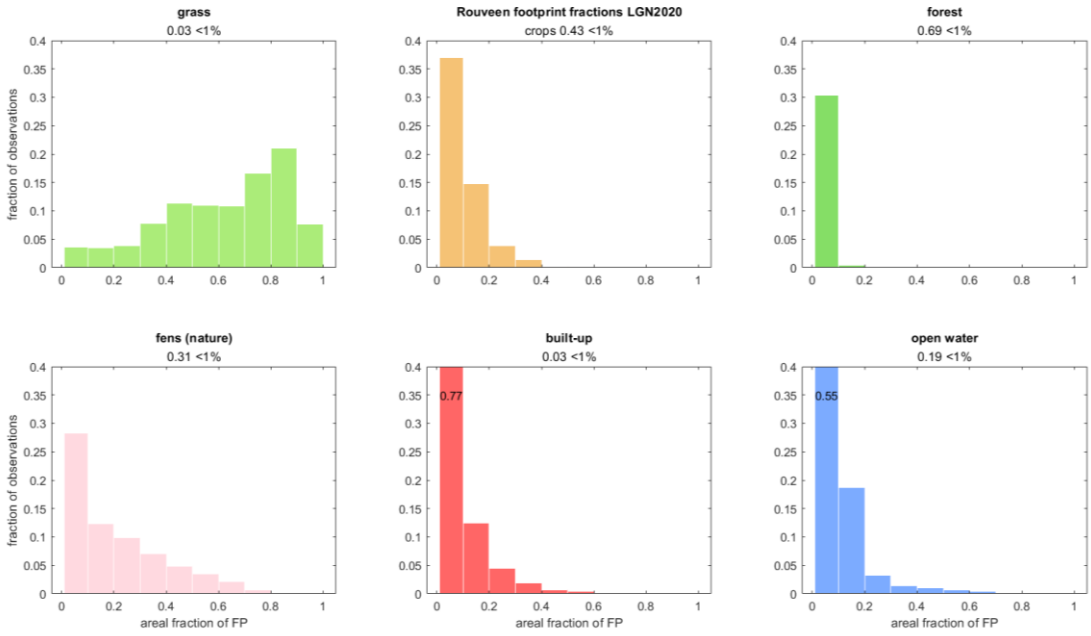

Figure SM5a Histograms of footprint fractions of (merged) land use classes in "Kop van Overijssel" ("Rouveen"). Horizontal: fraction of each footprint covered by respective classes; vertical: fraction of all 3751 footprints. Bins: <1%, 1-10%, 10-20 %, 20-30%, etc.

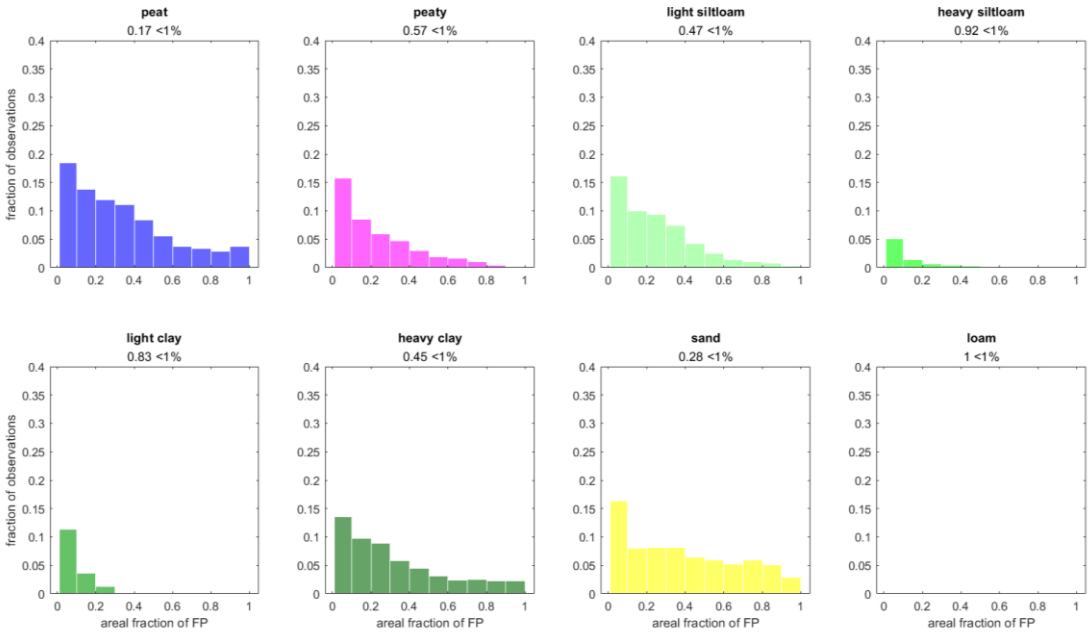

Figure SM5b Histograms of footprint fractions of (merged) soil classes in "Kop van Overijssel" ("Rouveen"). Otherwise as in Figure SM5a

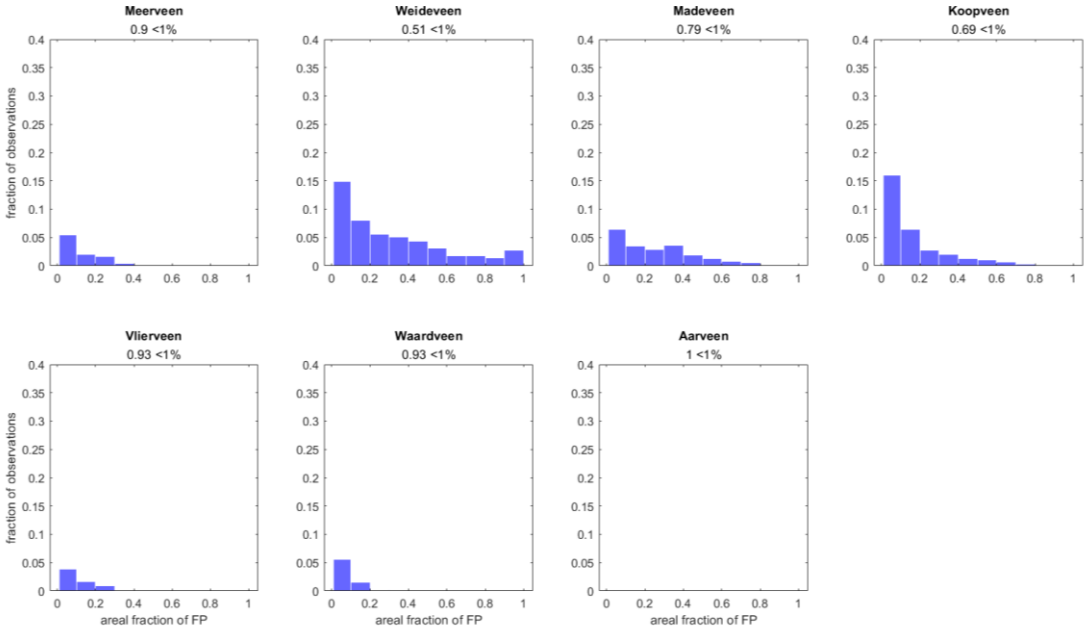

Figure SM5c Histograms of footprint fractions of peat types in the "Kop van Overijssel". Otherwise as in Figure SM5a.

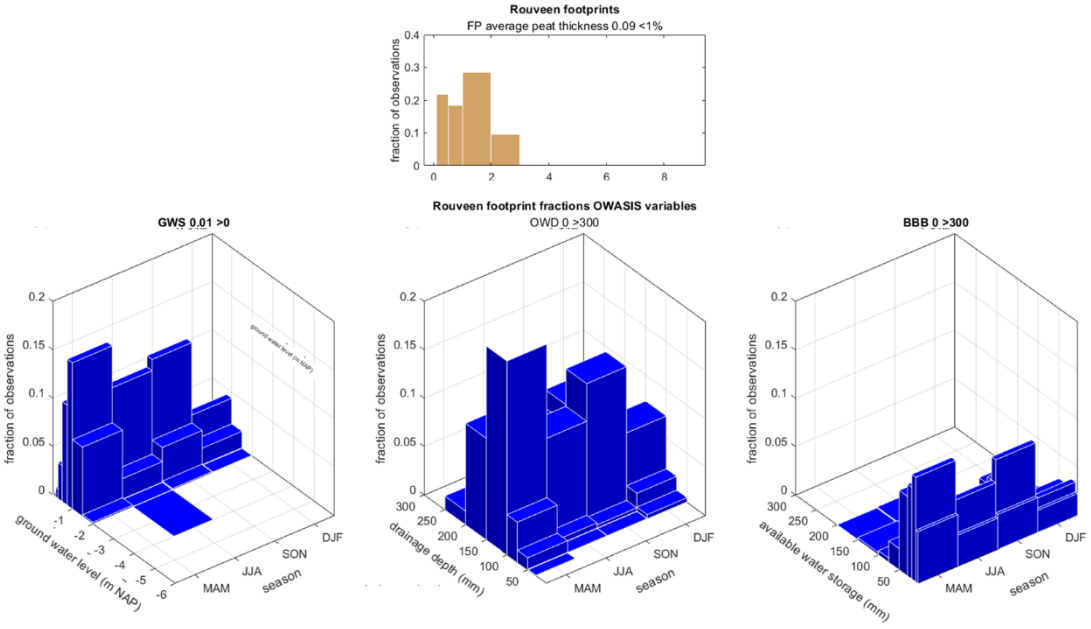

Figure SM5d Histograms of aircraft footprint fractions of time varying OWASIS variables (Blue) and weighted footprint-average peat depth (brown) in the "Kop van Overijsselt". Otherwise as in Figure SM5a. Note: Ground Water Levels (left graphs) are relative to NAP.

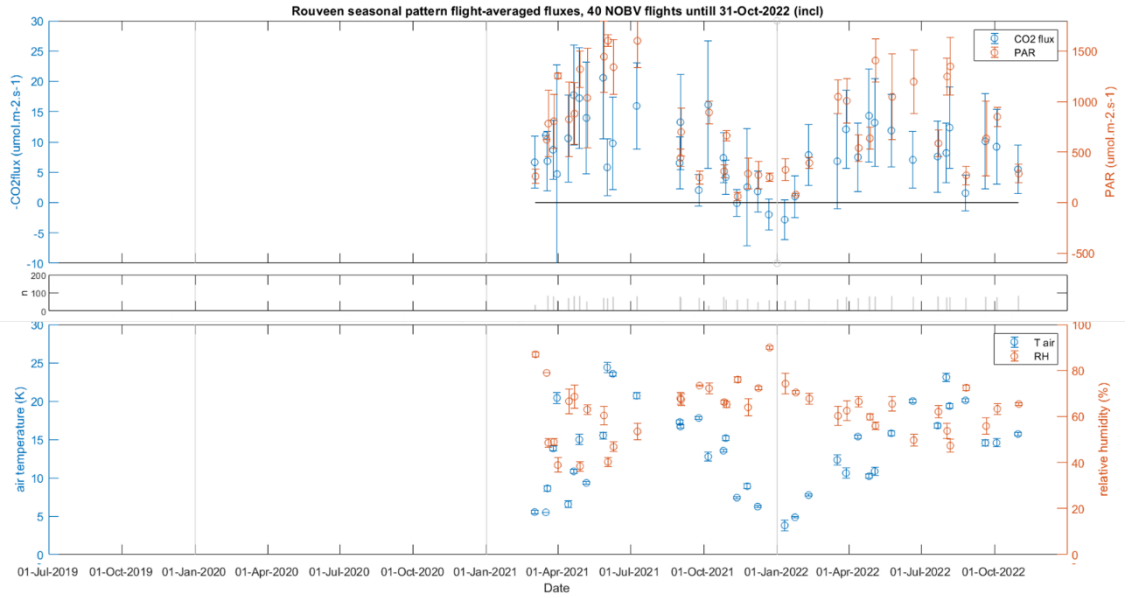

Figure SM6. Flight overviews produced for the "Kop van Overijssel" ("Rouveen")[5]. Points in each graph present flight-averaged observations, plus standard deviation. Top graph: CO_2 flux (blue, left axis, note we plot minus CO2flux to show the co-evolution with PAR, so positive values indicate CO_2 uptake by the land scape) and PAR (brown-red, right axis). Grey bars in second graph present the number of data points in each flight. Third graph: air temperature (at flying altitude, blue, left axis) and relative humidity (brown-red, right axis).

[5] Data acquisition started later than for the "Groene Hart" due to delays in obtaining flight permissions.

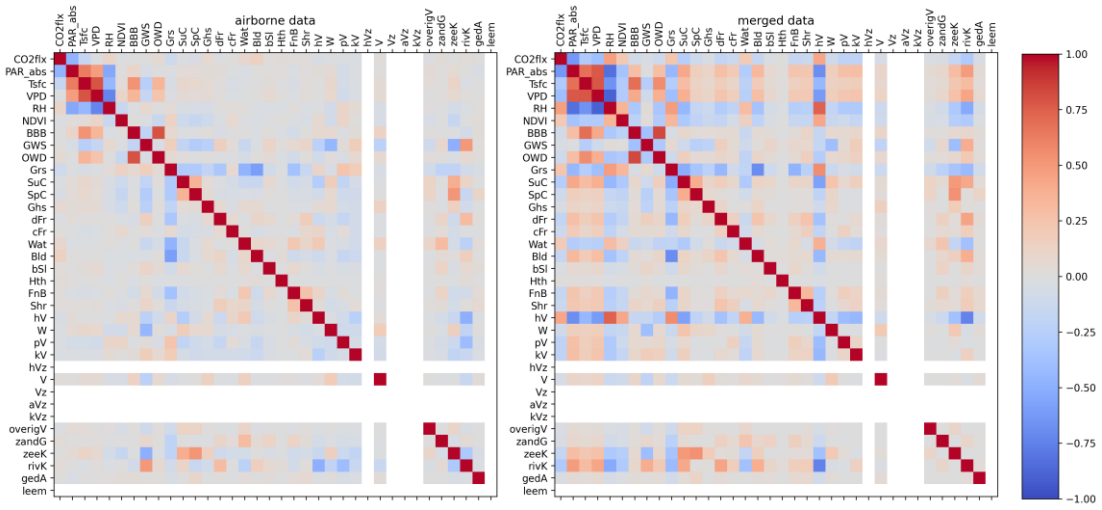

Figure SM7. Pearson correlation matrices for airborne data (left) and merged data (right). Correlations between all features with all features are shown. Features hVz, Vz, aVz, kVz and leem were not present in the datasets, and therefore show no correlation (white rows/columns).

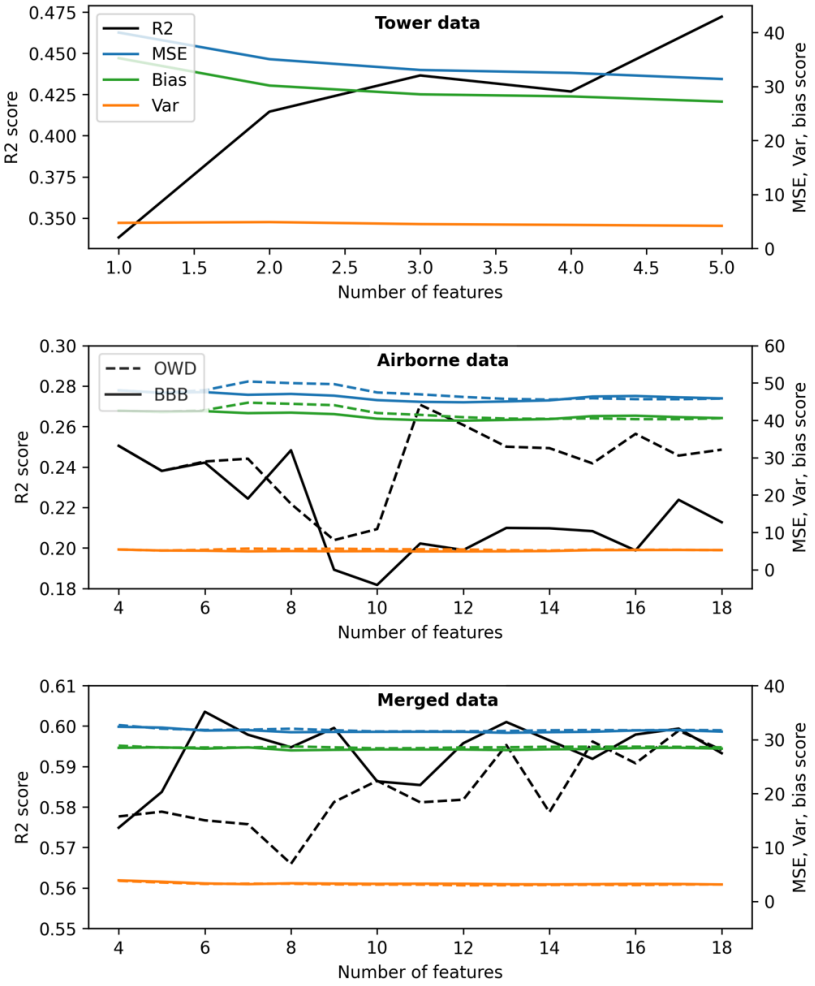

Figure SM 8. Results of Sequential Backward Feature Selection for tower, airborne and merged data. Performance of each model is shown by R2, MSE, bias and variance. Following the SBFS process, where features are sequentially omitted, figures should be read from right to left.