S1 Data-availability of holdout set

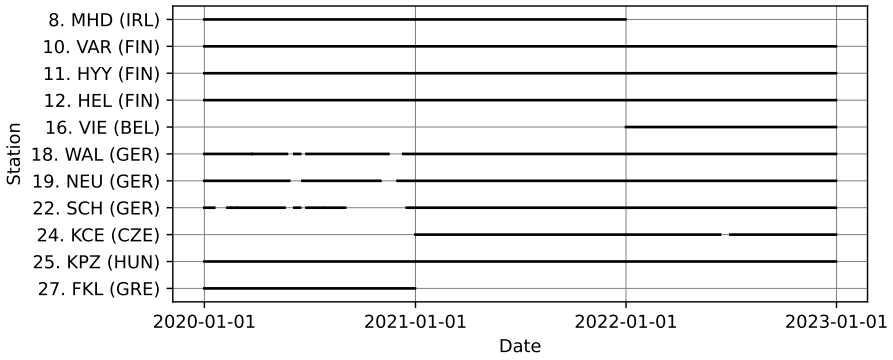

Figure S1. The data-availability of N100 measurements in the holdout set.

S2 Hyperparameters

Table S1. The values of selected XGBoost hyperparameters after tuning. In addition to these hyperparameters, we adjusted the objective of the XGBoost model to regression and squared loss. The details of hyperparameters can be found in XGBoost documentation (2022).

Hyperparameter	Value
max_depth	7
n_estimators	900
subsample	0.6
colsample_bytree	0.3
learning_rate	0.1
min_child_weight	0
gamma	0.2
reg_alpha	1e-6
reg_lambda	1e-1

S3 Representative subsample for scatterplots

In order to have comparable number of data points in the validation set from all stations for Figure 5, we sampled around 200 data points per station. This was done by dividing the station's data into bins based on season (Winter: December-February, Spring: March-May, Summer: June-August, and Fall: September-November). From each of these season bins, we sampled 50 data points. For these data points to capture the entire range of N100 values within the season bin we further stratified the data into 50 percentile bins and randomly sampled one data point from each percentile bin.

For stations where seasons that had less than 50 data points in the season, we selected all data points. As a result, a few stations had less than 200 data points sampled into the validation set: AMA (191), FKL (175), HRW (182), PRL (197), UAE (174).

S4 Station-included vs station-excluded comparison

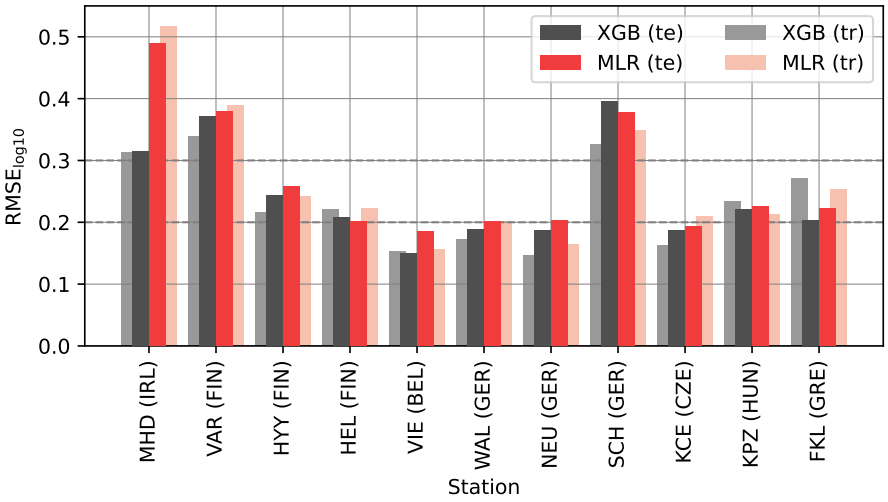

Figure S2. The comparison between MLR and XGB station-excluded model testing and training errors (RMSE calculated for log10-transformed concentrations) for the stations that had data during 2020-2022.

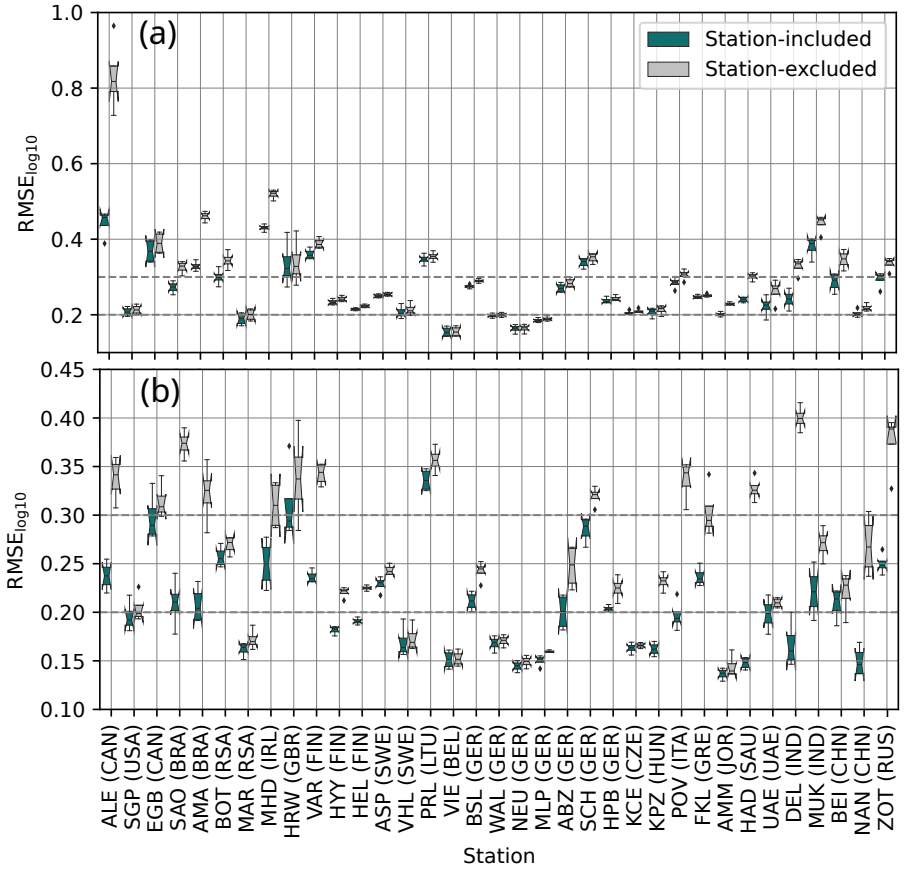

Figure S3. Boxplots comparing the model performance (RMSE calculated for log10-transformed concentrations) of station-included and station-excluded a) MLR models b) XGB models at each station. For this analysis we modified the station-excluded models to have validation sets corresponding to station-included models (Table 3). The boxes indicate the variation caused by different train-validation splits. The boxes show the quartiles and whiskers show the 1.5 interquartile range of the lower and upper quartile. Data points outside these are considered outliers and marked with individual markers. Additionally notches in the boxplots indicate the confidence interval of the median. If the notches of two boxes do not overlap, it indicates that the medians are statistically significantly different at 5 % significance level.

S5 BC and OM mixing ratio effect on models

In the training set, both the hydrophilic and hydrophobic BC and OM mixing rations correlated strongly, and after scaling their ratio followed 1:1 line closely (Fig. S4). The MLR model also assigned hydrophilic BC and OM almost equally large coefficients (Fig. 8). As a result, the closer the hydrophilic mixing ratios are the less they contribute to overall N100.

However, the training data does not represent the actual global distribution of BC and OM mixing ratios. There were locations where BC/OM ratio was not 1:1 but was a lot larger or smaller than what the MLR model was trained with (Fig. S5a). In these locations, the model estimates the overall contribution of hydrophilic BC and OM to N100 to either have a large decreasing contribution (if $OM_{h.phil}$ is much larger than $BC_{h.phil}$) or a large increasing contribution ($BC_{h.phil}$ is much larger than $OM_{h.phil}$). However, it should be noted that mostly the ratio of $BC_{h.phil}$ and $OM_{h.phil}$ is close to 1 (Fig. S5b).

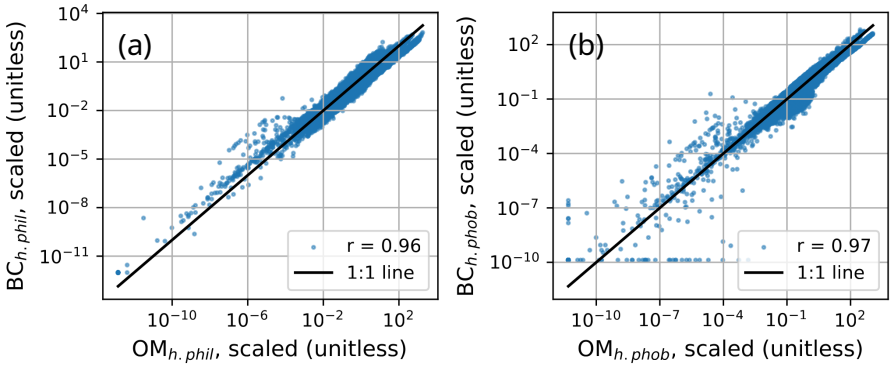

Figure S4. In the training set, the Pearson correlation coefficient (r) was very high between BC mixing ratio and OM mixing ratio both for a) hydrophilic mixing ratios and b) hydrophobic mixing ratios. The figure shows the scaled values.

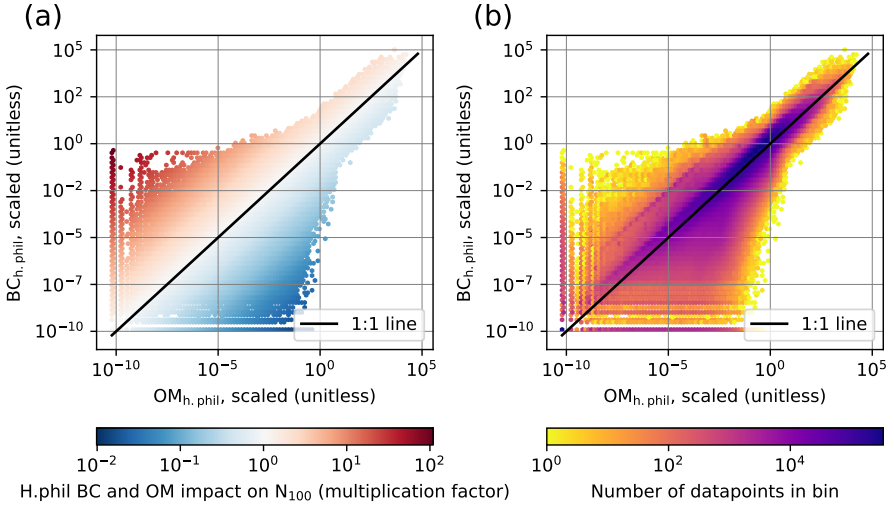

Figure S5. Panel a) shows the combined contribution of $BC_{h.phil}$ and $OM_{h.phil}$ to the total N100. Panel b) how the data points are distributed.

S6 The impact of ammonia concentration on MLR model results

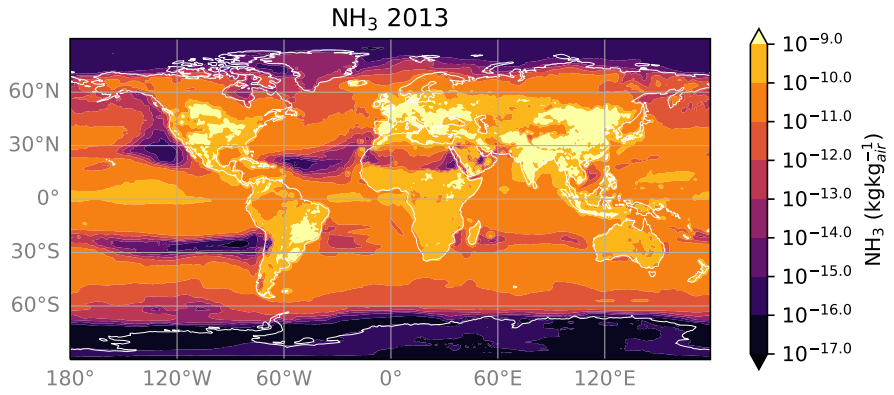

Figure S6. Average ammonia concentration in 2013. The low ammonia concentrations result in low N100 estimate with $\mathrm{MLR}_{\mathrm{global}}$ model.

S7 Comparison between single-station and station-included model results

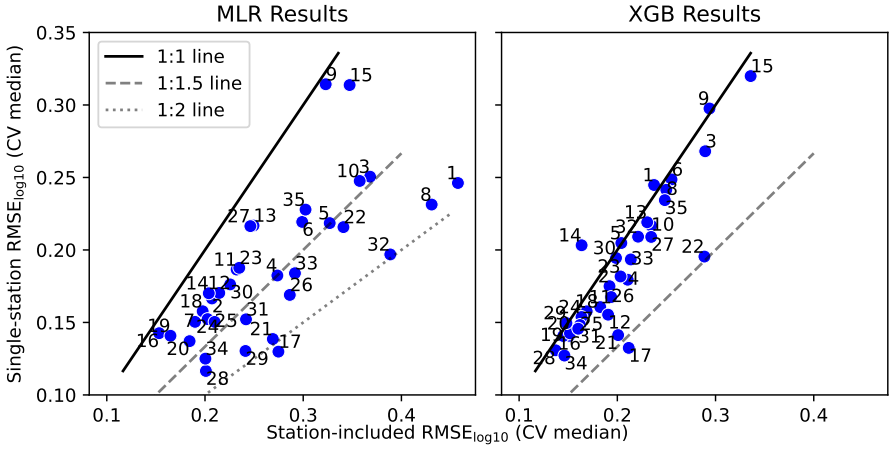

Figure S7. The comparison between the medians of cross-validation (CV) results from single-station and station-included models for a) MLR and b) XGB. The numbers correspond to stations as listed in Table 1.