ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

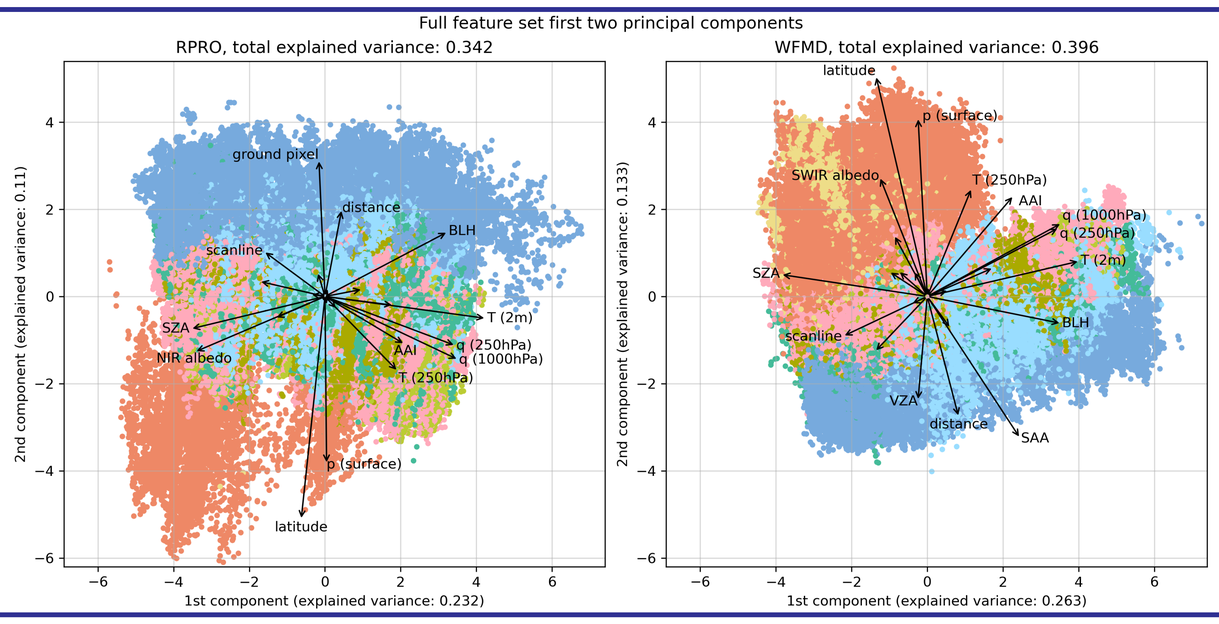
Quantifying environmental impacts on bias in Arctic TROPOMI methane retrievals using machine learning

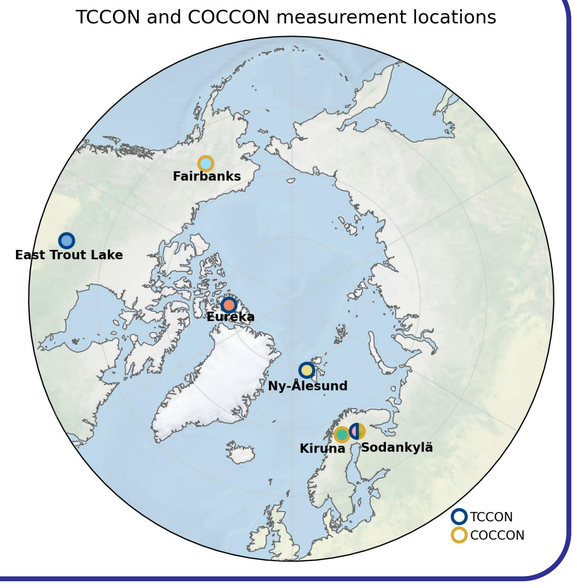
Tuomas Häkkilä[1] (*tuomas.hakkila@fmi.fi*), **Hannakaisa Lindqvist**[1], **Ella Kivimäki**[1], **Tomi Karppinen**[1], **Rigel Kivi**[1], **Johanna Tamminen**[1]

[1]Space and Earth Observation Centre, Finnish Meteorological Institute

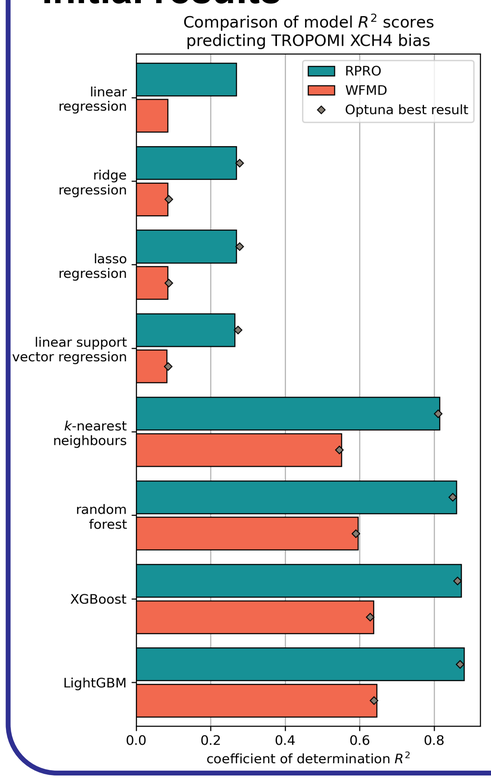
Introduction

- We explore the environmental factors affecting the bias in Arctic TROPOMI XCH_4 observations against ground-based TCCON and COCCON measurements.
- Two TROPOMI XCH_4 products, the operational reprocessed (RPRO) and the WFM-DOAS v2.0 (WFMD), are analysed, using a dataset with a total of 99 893 (RPRO) and 121 339 (WFMD) individual observations co-located near 8 ground-based instruments.

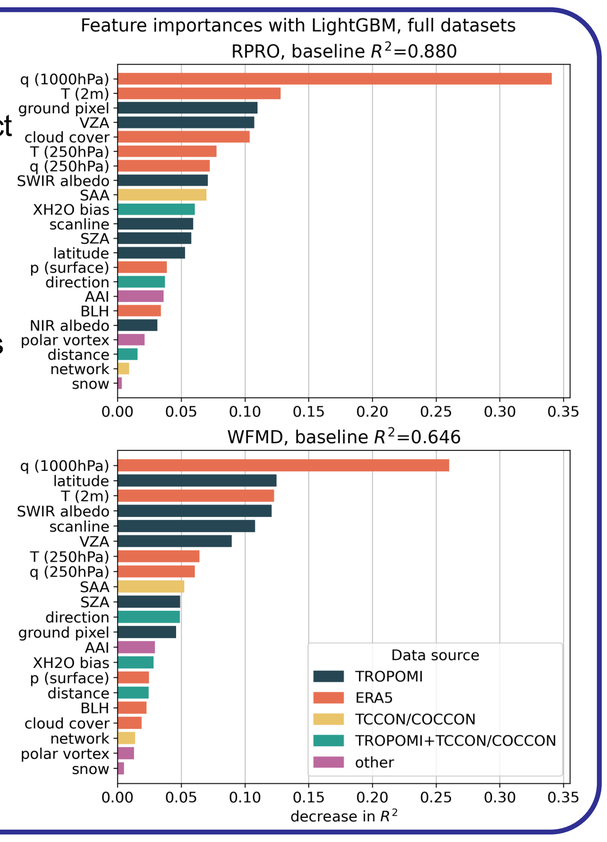


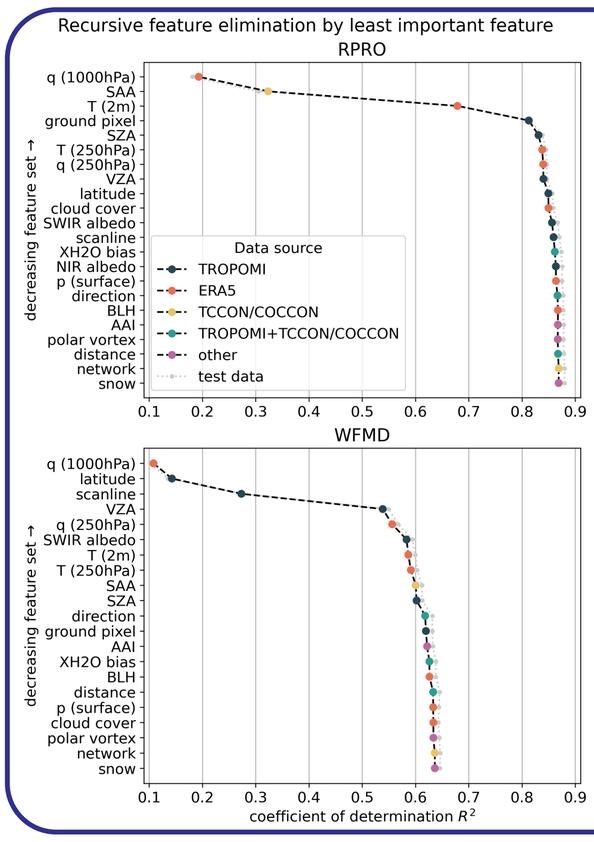


Initial results



- After an initial comparison of ML estimators, we select LightGBM as the best model in predicting the XCH_4 bias for further analysis.
- LightGBM R^2 scores are around 0.88 for RPRO and 0.64 for WFMD.
- Importances are assessed based on permutation feature importance (PFI), showing surface specific humidity dominates in importance for both RPRO and WFMD.



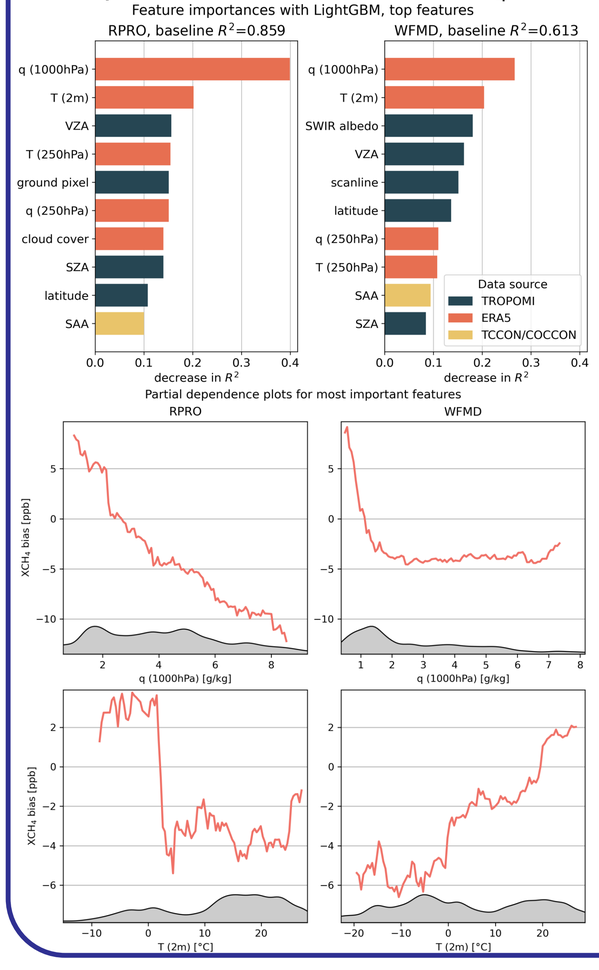


Limiting the dataset

- After calculating PFIs on the full feature set, feature selection is performed to limit the number of features.
- Recursive feature elimination (RFE) is used based on the least important feature to select the 10 top features for further analysis.
- To minimise the effect of coincidental correlation due to similar seasonal cycles, the full analysis is also performed separately on three "seasonal" datasets.

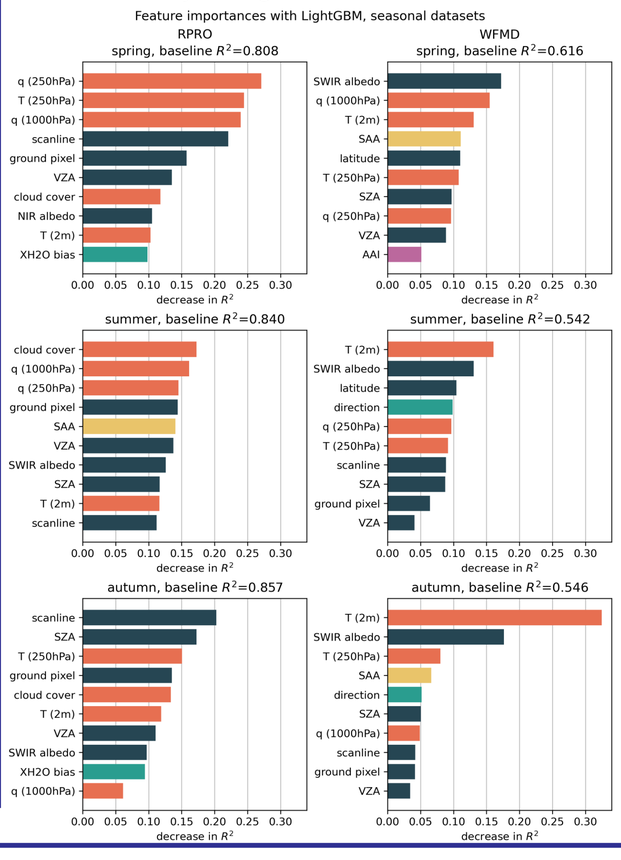
Selected features

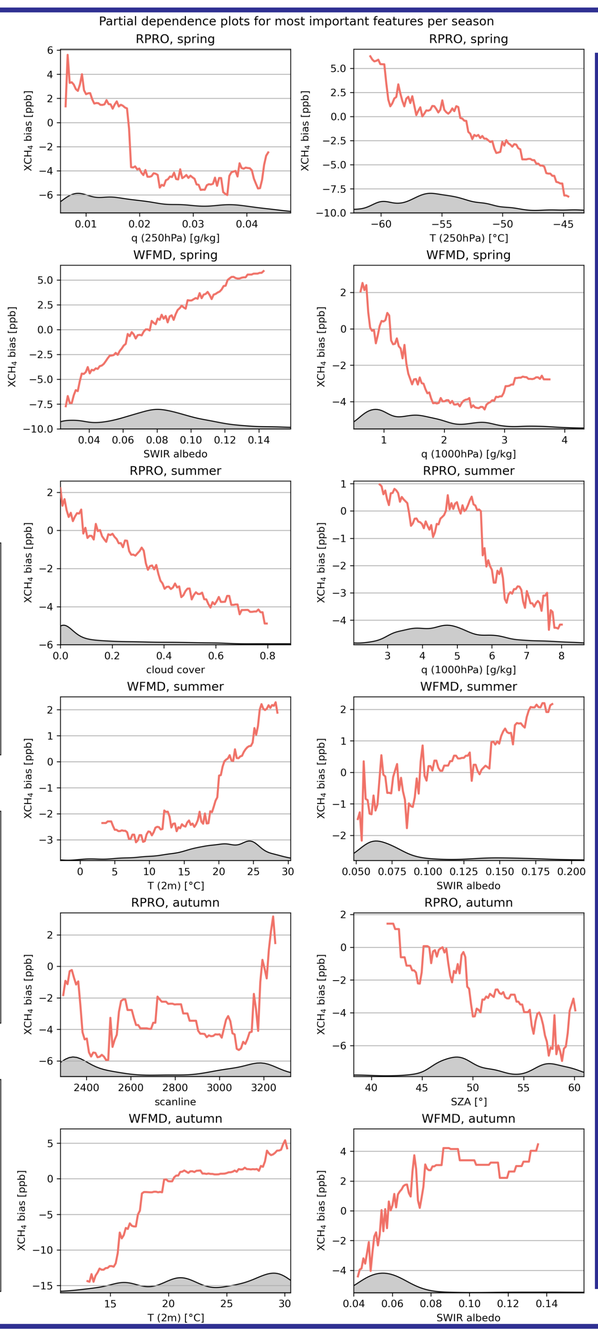
- After RFE, specific humidity at the surface persists as the most important feature, followed by surface temperature, possibly due to seasonality.
- Partial dependence plots show the average relation between the most important features and the XCH_4 bias.



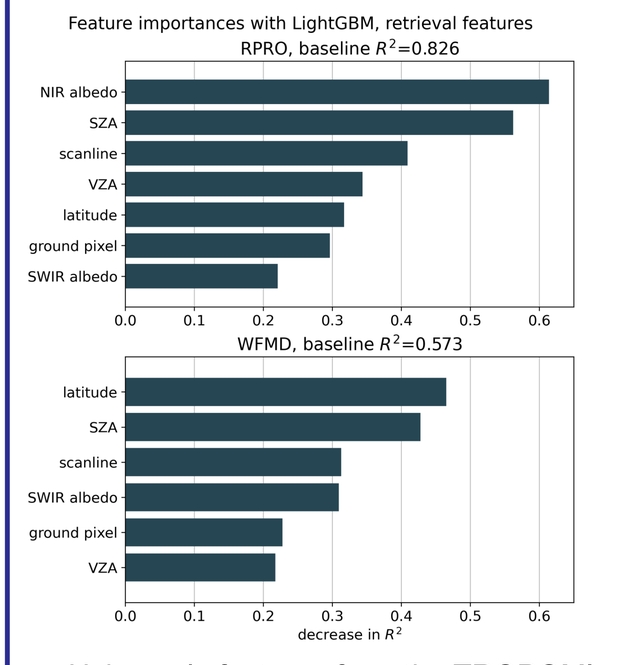
Seasonal datasets

- We select mid-summer data ±25 days of the summer solstice to limit the seasonality in the TROPOMI XCH4 bias, and consider spring and autumn as 50 days before and after this period.
- The reduction in samples results in somewhat worse model scores.
- The seasons show very different PFIs, and there are more differences between RPRO and WFMD compared to the fully merged datasets.

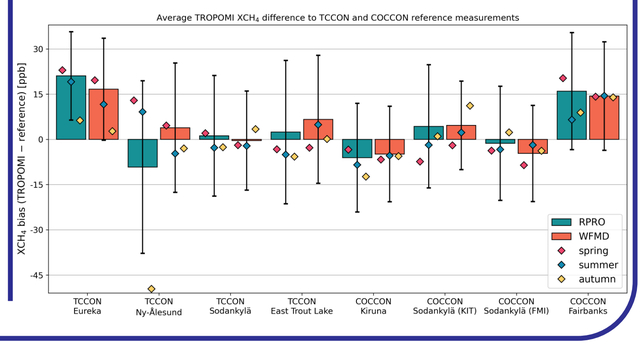




Retrieval features



- Using only features from the TROPOMI XCH_4 retrievals results in only a slight drop in model performance.
- This set of features could be used operationally to correct biases (shown below at each reference station), though further testing is required.



Acknowledgements

We would like to acknowledge all TROPOMI satellite data producers and TCCON and COCCON ground-based data producers for their efforts in making their products publicly available. This work has been supported by ESA projects MethaneCamp and AMPAC-Net as well as the FAME Flagship of the Research Council of Finland.