

# Deep transfer learning method for seasonal TROPOMI XCH<sub>4</sub> albedo correction

Alexander C. Bradley<sup>1,2</sup>, Barbara Dix<sup>2</sup>, Fergus Mackenzie<sup>3</sup>, J. Pepijn Veefkind<sup>4,5</sup>, and Joost A. de Gouw<sup>1,2</sup>

<sup>1</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA

<sup>2</sup>Chemistry Department, University of Colorado, Boulder, CO 80309, USA

<sup>3</sup>BlueSky Resources, Boulder CO 80302, USA

<sup>4</sup>Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

<sup>5</sup>Department of Geosciences and Remote Sensing, Delft University of Technology, Delft, the Netherlands

**Correspondence:** Joost A. de Gouw (joost.degouw@colorado.edu)

Received: 24 July 2024 – Discussion started: 20 August 2024

Revised: 16 January 2025 – Accepted: 3 February 2025 – Published:

**Abstract.** The retrieval of methane from satellite measurements is sensitive to the reflectance of the surface, and in many regions, especially those with agriculture, surface reflectance depends on the season. Existing corrections for this effect do not take into account a changing relationship between reflectance and the methane correction value over time. It is an important issue to consider, as agricultural emissions of methane are significant and other sources, like oil and gas production, are also often located in agricultural lands. In this work, we use a set of 12 monthly machine learning models to generate a seasonally resolved surface albedo correction for TROPOMI methane data across the Denver–Julesburg basin. We found that land cover is important in the correction, specifically the type of crops grown in an area, with drought-resistant-crop-covered areas requiring a correction of 5–6 ppb larger than areas covered in water-intensive crops in the summer. Additionally, the correction over different land covers changes significantly over the seasonally resolved timescale, with corrections over drought-resistant crops being up to 10 ppb larger in the summer than in the winter. This correction will allow for more accurate determination of methane emissions by removing the effect of agricultural and other seasonal effects on the albedo correction. The correction may also allow for the deconvolution of agricultural methane emissions, which are seasonally dependent, from oil and gas emissions, which are more constant in time.

## 1 Introduction

The second most significant anthropogenic greenhouse gas (GHG), methane, has important climate implications. Providing 27 times the warming potential of carbon dioxide on a 100-year timescale, but with a much shorter lifetime of less than 10 years, a reduction in methane emissions could ease global warming and potentially help achieve 1.5 or 2° goals (Boucher et al., 2009; Collins et al., 2018). Agriculture is the largest contributor to global anthropogenic methane emissions (41.0 %), followed by the energy sector (38.4 %) (IEA, 2023). Methane emissions from the energy sector are dominated by oil and gas operations which, in the United States, are still expanding following phase-out trends in coal-fired power production (U.S. Energy Information Administration (EIA), 2024). From natural gas production sites, methane emission rates are estimated to be 830 Mg h<sup>-1</sup>, with a high fraction from super-emitting sites (Omara et al., 2018).

Climate policy solutions generally rely on bottom-up inventories, which are derived from known emission rates for individual processes at a source. Bottom-up inventories are often at odds with top-down measurement techniques, which rely on measurements of atmospheric concentrations, and some have tried to reconcile these differences (Allen, 2014; Etiope and Schwietzke, 2019). Atmospheric measurements have expanded in both number and choice of platform with the advances of satellite monitoring systems (de Gouw et al., 2020; Jacob et al., 2016). Due to the prevalence of super-emitters skewing averages, both bottom-up and top-down methods have large and difficult-to-quantify uncertain-

ties (sometimes well over 100 %; Riddick et al., 2024), especially when diverse sources of methane overlap (Allen, 2016, 2014). Bottom-up inventories rely on accurate reporting of emissions and emissions factors from private companies and are extremely sensitive to super-emitting events that make up a minority of events but a majority of the emissions (Allen, 2014; Riddick et al., 2024). Top-down methane emissions measurements rely on the accuracy of the instrumentation aboard the satellite, the retrieval method, and the methods used to calculate emissions from the column densities, such as a Bayesian inversion or a flux divergence method (Liu et al., 2021; Zhang et al., 2020). Improved methane inventories are invaluable to regulators and policymakers. Understanding the extent of methane emissions would help climate policymakers set more accurate and achievable goals and would allow regulators to effectively monitor those goals.

The accuracy of the inventory depends quite significantly on the accuracy of the measurements. The Tropospheric Monitoring Instrument (TROPOMI), an imaging spectrometer aboard the Sentinel-5 Precursor satellite, is known to have significant biases in the operational methane retrievals related to surface albedo (Lorente et al., 2021). There have been several recent updates to the dataset to mitigate this albedo effect using TROPOMI retrieval data over areas without emissions, as well as by comparison with proxy retrievals from the Greenhouse Gases Observing Satellite (GOSAT), which are much less affected by surface albedo (Balasus et al., 2023; Lorente et al., 2021). When applied to methane retrievals on a seasonal basis, we show here that some residual albedo effects are still apparent and may thus bias seasonal data. This study attempts to develop a seasonal albedo correction for the area of the Denver–Julesburg (DJ) basin in Colorado to account for these effects.

Colorado ranks in the top 10 US states in total energy production (U.S. EIA, 2020). The state produced nearly 5 times more crude oil in 2022 than in 2010 largely due to the expansion of horizontal drilling and hydraulic fracturing (Cook et al., 2018; Annual Energy Outlook 2023, 2025), and production of natural gas has more than doubled since the year 2000 (Annual Energy Outlook 2023, 2025). The majority of crude oil produced in Colorado comes from the Niobrara shale formation, located mostly within Weld County, while the whole basin stretches from southern Colorado to Wyoming and from the front range uplift into Nebraska and Kansas (Pétron et al., 2014; U.S. EIA, 2023). Weld County is also one of the richest agricultural counties east of the Rocky Mountains, producing over 27 % of the entire state’s agricultural sales. A total of 80 % of the land area in Weld County is used for agriculture, with 44 % of that land used for cropland and 53 % used for pastureland (United States Department of Agriculture (USDA), 2017). Agriculture complicates the measurement and attribution of methane emissions data in two major ways: (1) cropland seasonal albedo shifts are under-compensated for in current albedo corrections due to a variable relationship between albedo and correction over time

and (2) unreported methane emissions from animal feedlots like Concentrated Animal Feeding Operations (CAFOs) occur in close proximity to oil and gas production. A seasonally resolved albedo correction would assist with both of the issues by (1) correcting the seasonal shifts in albedo while accounting for the changing relationship between albedo and correction value and (2) allowing for more accurate top-down seasonal methane emissions quantification, which may allow deconvolution of consistent O&G emissions from seasonal agricultural emissions. Co-location of large oil and gas production with massive agricultural operations makes the DJ basin and Weld County in particular a prime target for a machine-learning-based seasonal albedo correction.

Machine learning is a branch of artificial intelligence where computers are trained to recognize patterns and make decisions based on data, similar to how humans learn from experience. Some machine learning models are considered a “black box” because it can be difficult to understand how they make decisions. To address this, tools like SHapely Additive exPlanations (SHAP) help provide insights into how machine learning models arrive at their predictions (Lundberg and Lee, 2017; Rudin, 2019). Neural networks, a type of machine learning model, are inspired by the way the human brain processes information. They consist of layers of “neurons” that work together to identify patterns in data (Abadi et al., 2015).

Albedo corrections for the TROPOMI methane data have been described in the literature, with the most prominent being from Lorente et al. (2021) which is now incorporated into the TROPOMI retrieval algorithm. Another effective albedo correction is from Balasus et al. (2023), who also utilize machine learning. Lorente et al. (2021) used a B-spline interpolation of albedo dependence calculated over 2 years of data, while Balasus et al. (2023) trained a machine learning model on global data over a multi-year timescale, using the University of Leicester (UoL) Greenhouse Gases Observing Satellite (GOSAT) proxy retrievals as the target data (Balasus et al., 2023; Lorente et al., 2021) (hereafter referred to as simply Balasus et al. and Lorente et al. when referring to the corrections they designed). In this work we demonstrate that seasonal or monthly averaged methane retrievals over Colorado continue to be biased by albedo effects after the implementation of these correction algorithms. A major reason for using a seasonal- or finer-time-resolution average is for deconvolution of agricultural emissions. Presently, oil and gas operations are required to report on their emissions, but the accuracy is disputed (Zavala-Araiza et al., 2015). Meanwhile, agricultural operations are largely exempted from emissions reporting. The difficulty arises when agricultural and oil and gas operations are near to each other or co-located. Satellite methods for measuring emissions from oil and gas operations can be biased by the unaccounted-for agricultural operations. Deconvolution of oil and gas emissions, which largely remain constant through the seasons, and agricultural operations, which cycle through the seasons, could be made

more accurate if the measurements could be seasonally resolved. A seasonal albedo correction, as presented here, is a step towards making a seasonal measurement more accurate for better determination of emissions.

## 2 Methods

The two satellites used in this study, TROPOMI and GOSAT, have different spatial resolutions both in a latitude–longitude grid but also vertically, with different numbers of vertical retrieval pressure levels, known as averaging kernels. Because of this, as well as other instrument sensitivities, we do not expect TROPOMI and GOSAT to measure the same concentrations over the same places at the same time.  $\Delta(\text{TROPOMI} - \text{GOSAT})$  is an adjustment made here to place TROPOMI and GOSAT data onto common averaging kernel sensitivities and vertical profiles and determine the difference between the measurements on the same spatial scale. The calculation of this value is described in Balasus et al. and involves interpolating GOSAT vertical pressure levels to TROPOMI’s vertical pressure grid in order to calculate what GOSAT would have retrieved with TROPOMI’s vertical sensitivity (Balasus et al., 2023). This value is used as the target of the machine learning (ML) model training.

### 2.1 Satellite data

#### 2.1.1 TROPOMI

TROPOMI is the push-broom imaging spectrometer aboard the European Copernicus Sentinel-5 Precursor (S5P) satellite, capable of measuring methane among other chemicals. It has been described in detail previously (Levelt et al., 2022; Veefkind et al., 2012). In this work, TROPOMI orbit files from April 2018–December 2022 were downloaded from the ESA Copernicus open-access hub. We used level 2 reprocessed and offline version 2.4 methane column data,  $\text{XCH}_4$ , with internal TROPOMI-defined QA values of  $\geq 0.5$ , indicating good-quality retrievals and better, including good-quality snow-covered scenes, and the shortwave infrared (SWIR) surface albedo as co-retrieved with  $\text{XCH}_4$ . The bounding box for machine learning training data used was latitude 34 to 42°N, longitude 106 to 95°W, which encompasses the largest production regions of the Denver–Julesburg basin and extends into the surrounding states that also contain parts of the basin: Wyoming, Nebraska, and a small part of Kansas.

A well-known artifact in methane retrievals from TROPOMI is striping caused by small differences between across-track pixels, which can be mitigated by performing a stripe correction (Liu et al., 2021). This work utilizes an inherent stripe correction instead of a separate explicit stripe correction.

#### 2.1.2 GOSAT

The University of Leicester Full-Physics dataset (UoL-FP) proxy retrieval scheme was used (Parker and Boesch, 2020). The proxy retrieval involves retrieving the  $\text{CO}_2$  column to act as a proxy for aerosol scattering effects (Scheepers et al., 2012). This dataset has been used extensively before as a measurement that is less affected by changing surface albedo (Balasus et al., 2023; Lorente et al., 2021). The data were downloaded from the Center for Environmental Data Analysis for the years 2018–2020 on a global scale, and the code calculating the co-location of TROPOMI and GOSAT data to calculate TROPOMI–GOSAT pairs was based on that of Balasus et al., where pairs are calculated as pixel centers  $< 5$  km apart in space and  $< 1$  h apart in time (Balasus et al., 2023). As our method requires a large number of data and the region is much smaller, we loosened the criteria to any pixel overlap in space and  $< 2$  h apart in time.

### 2.2 Machine learning methods

A neural network machine learning algorithm was trained on a large subset of co-located TROPOMI and GOSAT data gridded to a  $0.1^\circ \times 0.1^\circ$  latitude–longitude square grid, totaling 17 634 points with 31 variables, described in Table S1 in the Supplement, each to develop a hybrid TROPOMI–GOSAT dataset, which combines the measurement accuracy and lack of albedo effect of the GOSAT proxy retrieval with the data coverage of TROPOMI. The variables were selected based on previous ML work on this topic with a few changes (Balasus et al., 2023). We chose to incorporate the retrieved and corrected  $\text{XCH}_4$ , which is corrected based on the on-board albedo correction from Lorente et al. We also chose to remove the surface classification variable because our relatively smaller area of study has relatively few bodies of water. Furthermore, we chose to remove wind speed variables so that we would not introduce a bias or double counting if this model were to be used with the flux divergence method of quantifying methane emissions, which requires wind speed and direction (Beirle et al., 2021). The predictor variables were normalized using  $z$ -score normalization to ensure the predictor values are on the same scale for training purposes. A neural network can be described as “deep” if it has 3 or more “hidden layers” or levels in the network. Hidden layers are the strata of neurons which receive input from above and output to below and are “hidden” because the only layers the user interacts with are the top-level input and the bottom-level output, while there may be hundreds or even thousands of layers sandwiched between. The term “transfer learning” is used to describe a model that has been trained previously and is subsequently trained again starting from the previous training endpoint. A deep transfer learning (DTL) method was used where an annual base model was trained on 80 % of the total points, randomly sampled. These same points were then separated by the month of their collection and used to

train 12 separate monthly models, starting from the annual base model; a schematic representing this training process is presented in Fig. 1. The remaining 20 % of the data were used to calculate final fit statistics for each monthly model.

There is an unequal distribution of points across the months, which introduces some seasonal bias to the annual model. This bias is subsequently removed when the monthly models are trained. DTL is especially suited to this type of learning because (1) the initial learning phase trained on all training data helps the lower levels of the model learn to generalize the task and (2) the subsequent training occurs on much smaller monthly training datasets that help train higher, more specific levels of the model. Various hyperparameters were tuned using Optuna, a hyperparameter tuning package for Python (Takuya et al., 2023<sup>TS1</sup>). To monitor against overfitting, training and validation loss for the training period of each model were calculated and are presented in Fig. S3.

### 2.3 Python

All of the computations were completed using Python, a general-use, interpreted, object-oriented programming language ideal for building and implementing machine learning models and algorithms. A number of third-party packages were useful in the computations completed in this work: Matplotlib for figure generation (Hunter, 2007); the machine learning packages TensorFlow (Abadi et al., 2015), Keras (Chollet et al., 2015), <sup>CE1</sup>XGBoost (Chen and Guestrin, 2016), LightGBM (Ke et al., 2017), and CatBoost (Dorogush et al., 2018); Pandas and GeoPandas for tabular and geospatial data organization (Jordahl et al., 2020<sup>TS2</sup>; The pandas development team, 2024); Optuna and Fast and Lightweight AutoML Library (FLAML) for tuning ML models hyperparameters (Takuya et al., 2023<sup>TS3</sup>); SciPy for scientific and statistical functions (Virtanen et al., 2020); Shapely for manipulation of geometric objects (Gillies et al., 2007); NumPy for array manipulation (Harris et al., 2020); netcdf4 for opening and reading satellite data (Whitaker, 2008); Rasterio for raster manipulation (Gillies et al., 2013); and tqdm to visualize data processing progress (da Costa-Luis, 2019<sup>TS4</sup>). Figure 1 was created using the Google Drawings suite, Fig. 2 was created using Python Matplotlib, and Figs. 3–7 were created in Igor Pro 8.04.

### 2.4 Other geospatial data

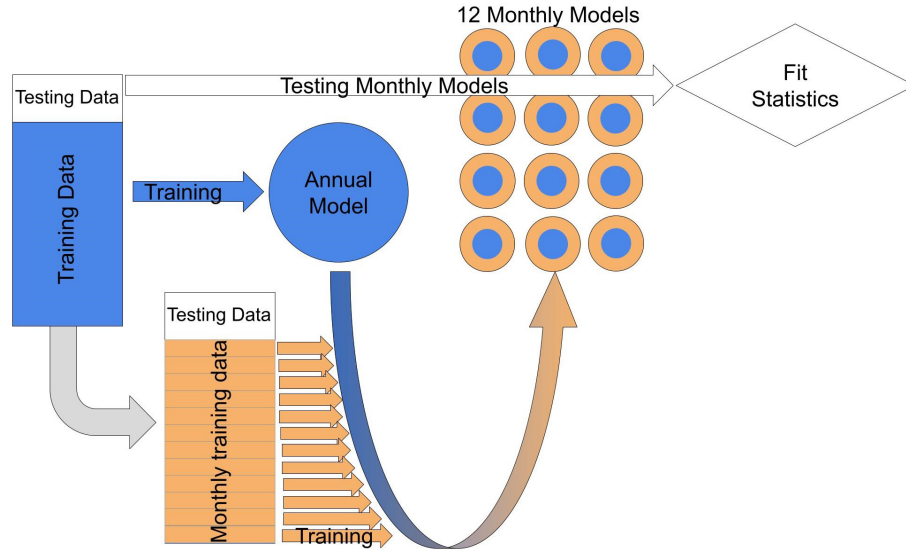
River paths and extent data for the South Platte River and North Platte River were downloaded from NOAA (National Weather Service, 2024). Crop data were downloaded from CropScape, a geospatial thematic agricultural mapping software (Han et al., 2014). Cartographic shapefiles containing state, county, and urbanized area boundary lines were downloaded from the US Census Bureau (2024). Finally, data visualizations were made to be color-accessible by Fabio Crameri’s scientific color maps (Crameri et al., 2020).

## 3 Results and discussion

The seasonal biases of the current TROPOMI operational product, which includes the albedo correction from Lorente et al., are studied in Fig. 2 for the area of interest. Figure 2 shows the ratio between co-located GOSAT and TROPOMI methane retrievals as a function of surface albedo in the shortwave infrared. In the ideal case, these ratios are equal to 1 and there is no correlation between this ratio and surface albedo ( $R=0$ ). When all data are used (Fig. 2a) the Pearson correlation is indeed calculated to be low, i.e., below a threshold of 0.1, which we chose here as a target value for minimal correlation between SWIR surface albedo and the albedo corrected methane retrieval. Though the significance of Pearson coefficients is up to interpretation, most would agree that a value of  $< 0.1$  signifies negligible correlation (Akoglu, 2018; Schober et al., 2018). When the data are shown by season, this is no longer true – Pearson correlations with an absolute value greater than 0.1 indicate that there exists some correlation between the SWIR surface albedo and the albedo corrected methane retrieval. The Lorente et al. correction algorithm does account for some seasonality because the TROPOMI-retrieved variables include the surface albedo SWIR which is used to calculate a correction value. However, the seasonal correlation reappears after the Lorente et al. correction because this correction assumes that the relationship between surface albedo SWIR and correction value is static over time. Figure 2b and c demonstrate the change in surface albedo as a function of season, with the density of counts shifting from the left side of the plot, indicating smaller albedos, to the center of the plot, indicating higher albedos on average from summer to winter. The different seasons also have different directions of change, with summers having an inverse correlation and winters having a positive correlation. The QA value used in processing the TROPOMI retrieval data retained high-quality snow-covered scenes, so some of this shift could be attributed to the SWIR reflectance of snow over bare soil. Regardless of the reason, the shifting albedo and seasonally variable albedo effect biases methane retrieval data from TROPOMI at finer timescales. In order to correct for this bias we employed a DTL neural network machine learning algorithm.

### 3.1 Model evaluation

The DTL neural network models were trained and evaluated as described in Sect. 2.2 and compared against the uncorrected methane retrieval, the Lorente et al. corrected methane retrieval, and the blended TROPOMI–GOSAT product produced by Balasus et al., commonly referred to as the “Harvard dataset”, for their effectiveness in methane correction. To evaluate against the other models, Pearson correlations were calculated and presented in Fig. 3a, where different constructions of Pearson values have been unified according to Table S2. Pearson correlations have been calculated the



**Figure 1.** Schematic representation of the data training process. Blue represents the annualized, long-term model, while orange represents the short-term monthly seasonal model and data. Transfer learning, the process by which a pre-trained model is trained again, usually on more specific data, was utilized here to generate 12 monthly models with the deeper understanding that comes from larger data quantities in the annualized model combined with the better specialization of the monthly seasonal training data, represented by the orange circles with the blue center.

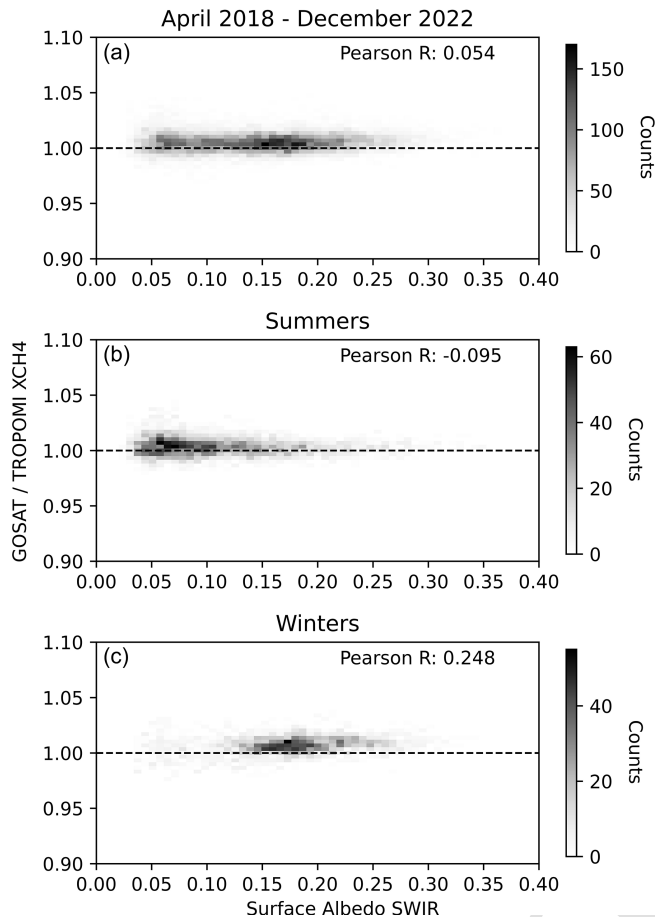
same ways as in Fig. 2 with the correlation between GOSAT-TROPOMI and surface albedo. To reiterate, a Pearson correlation of 0 is the preferred value, as the difference between the two datasets does not depend on surface albedo. The surface albedo is the SWIR albedo as retrieved by TROPOMI. Figure 3b depicts the 95 % confidence intervals about the mean of the 12 months of the Pearson values and is helpful in determining the most effective model. Dashed lines in both figures represent the ideal values indicating no correlation (Kuckartz et al., 2013), with values in Fig. 3a being between the dashed lines at  $-0.1$  and  $0.1$  Pearson correlation values and Fig. 3b being the average and center of the 95 % confidence margin of error on the line at 0 Pearson correlation value.

Only the models devised by this work entirely remove the seasonality described by the uncorrected data. Additionally, the Pearson values remain within our goal Pearson correlation value of 0.1 for each month. As expected, the uncorrected data reach the farthest outside of this range and remain outside for the greatest number of points. The Lorente et al. correction, which handles seasonality with a temporally static correction based on SWIR surface albedo, significantly improves upon the uncorrected data but preserves the seasonal trend in the data, demonstrating larger, positive correlations in the winter months and cycling through the seasons. The Balasus et al. blended dataset improves this further by reducing the seasonality of the correlation, but this dataset still displays correlations outside the 0.1 correlation threshold desired. Finally, this work's devised monthly models all ways fall between the ideal  $-0.1$  and  $0.1$  Pearson values. In

comparing the mean and 95 % confidence intervals, we observe the steady improvement in the progression of models, with this work's monthly models providing for the Pearson value closest to 0 and with the smallest 95 % confidence interval. All this demonstrates that the use of the monthly models provides a small but measurable improvement over previously designed models for albedo correction in this specific region around the Denver–Julesburg basin.

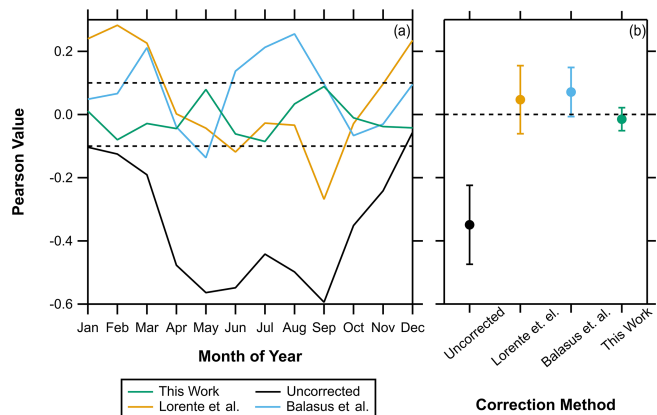
### 3.2 Model results

The Python library SHapely Additive exPlanations (SHAP) was used to determine the relative importances of the different variables incorporated into the model (Lundberg and Lee, 2017). The importance of a variable indicates how much each variable contributes to the difference between the actual model output and the average model output. The importances of the variables were calculated on a monthly basis to show how the importances change over time, and 2 representative months, 1 month for winter and 1 month for summer, are shown in Fig. 4. Figure 4a depicts the model outputs for the month of January in a decision plot. Decision plots are generally used to show how models make their determinations and what variables are affecting their decisions the most. Here the decision plot is showing that the range of correction values stretches from approximately  $-40$  to  $40$  ppb, indicating significant changes in the total methane concentrations ( $\sim 2\%$ – $4\%$ ). This change is larger than the mission specifications of bias less than 1.5 % and much larger than the measured mean bias of the corrected TROPOMI XCH<sub>4</sub> data of 0.2 % (Apit-



**Figure 2.** Albedo effect on methane retrievals on seasonally averaged TROPOMI data. TROPOMI bias-corrected methane level 2 retrieval data averaged from April 2018 to December 2022 (a) All months, (b) summer months (July–September), and (c) winter months (January–March). The TROPOMI data are co-located in space and time with UoL GOSAT proxy retrievals treated as ground truth. The dashed line represents perfect overlap and no correlation. Pearson  $R$  values represent the correlation between surface albedo and  $XCH_4$  retrievals. Our target Pearson  $R$  values are  $-0.1 < R \text{ value} < 0.1$ .

uley et al., 2022; Landgraf et al., 2023). That the corrections are larger than the biases suggests that the corrections are significant and important. Contrasting the general shapes of the decision plots, Fig. 4a appears to be more cone shaped, having a much starker taper in the less important variables, while Fig. 4b appears more cylindrical, sporting a milder taper. This indicates that the relative importance of predictor variables changes between seasons. A single model would miss this detail entirely, but the set of 12 monthly models allows for this change to occur. Additionally, the final model output value for Fig. 4b remains in the same range of approximately  $-40$  to  $40$  ppb. Together, this indicates that while outputs remain in the same range, the difference in the importance of the variables changes the method that the models

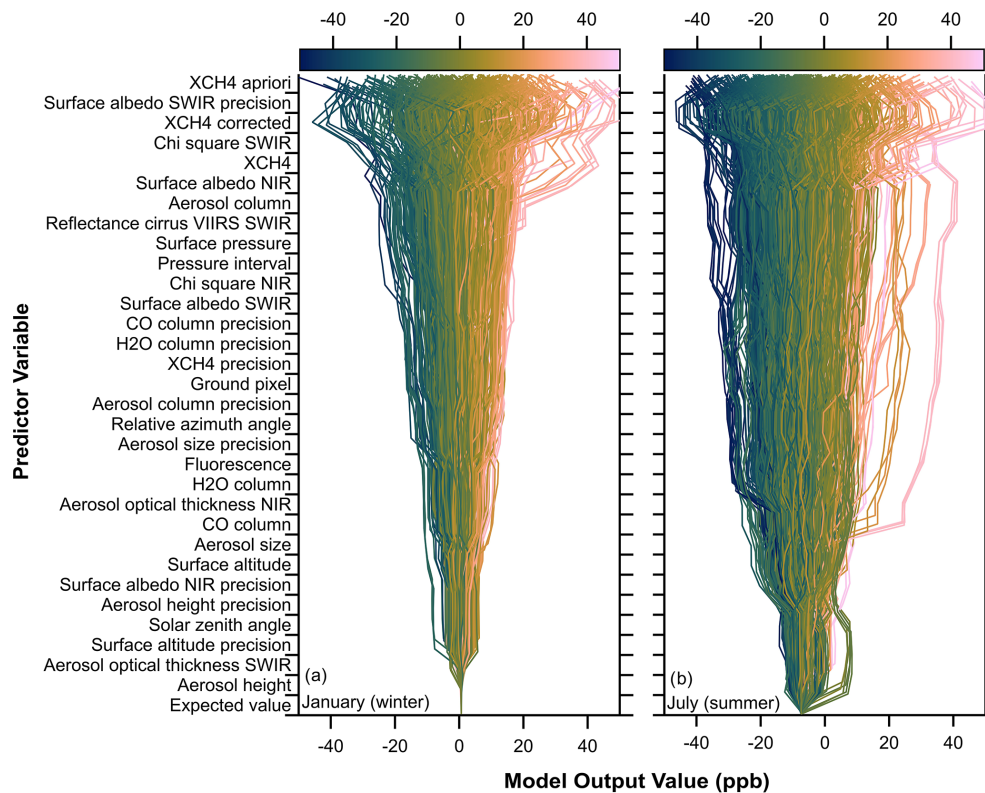


**Figure 3.** (a) Comparison of the model developed in this work with Lorente et al. and Balasus et al. corrections and uncorrected TROPOMI retrieval data. Pearson value describes the Pearson correlation value of (sensitivity-corrected GOSAT / measured or calculated TROPOMI value) and surface albedo SWIR for the ML model-predicted data, the Pearson correlation value of (raw GOSAT / calculated value) and surface albedo SWIR for scalar corrections, and the Pearson correlation value of (raw GOSAT / raw TROPOMI value) and surface albedo SWIR for the uncorrected data. (b) Points represent the average, and error bars describe 95 % confidence intervals of the 12 months.

use to predict the outcome. This difference of importance is indicative that the relationship between variables and the correction value is changing over time.

While the training process attempted to minimize the differences between TROPOMI and GOSAT data, thus effectively reducing the dependence on SWIR surface albedo, not all training iterations were successful in this due to the multitude of features to incorporate. As part of our model validation, we only considered those that reduced the correlation between  $XCH_4$  and surface albedo SWIR as viable models. Due to this validation method, we call our machine learning product an “albedo correction”. Figure 4 shows that other features may be more important than the surface albedo SWIR in the actual model calculation. “Importance” in a ML model is the magnitude of effect that variable has on the final output value of the model. The variables that appear higher on the y axis than “surface albedo SWIR” tended to be more important and should be analyzed as well. Some of these variables have clear reasonings as to why they are more important:  $XCH_4$  a priori,  $XCH_4$  corrected, and  $XCH_4$  are all the measurements of methane mixing ratio that were either priors for the TROPOMI measurement ( $XCH_4$  a priori) or direct measurements of the methane mixing ratio by TROPOMI ( $XCH_4$  and  $XCH_4$  corrected).  $XCH_4$  and  $XCH_4$  corrected directly measure methane mixing ratios via TROPOMI, serving as primary data sources for our predictive models. The reasoning for other important variables (surface albedo SWIR precision and chi square SWIR) is not so clear. The precision of the surface albedo SWIR measure-





**Figure 4.** Decision plots depicting relative importances of predictor variables on a seasonal basis. SHAP importances were calculated for (a) January and (b) July, and the contributions from each predictor variable are shown. Variables are ordered from top to bottom by importance in January. Color scale indicates the final model output value, which is the  $\Delta(\text{TROPOMI} - \text{GOSAT})$  value. Expected value is the average prediction made by the model across all possible combinations of features and is thus the same value for all trials using the same model.

ment being important was not expected but may be the result of a well-trained model successfully making the association between the SWIR albedo measurement and its precision. A less precise measurement would be less heavily relied upon for the model's predictions, so the importance may come from the association between the precision measurement and how much a particular measurement affected the model during training. Similarly, the chi square SWIR is a goodness-of-fit check that ensures that the SWIR measurements by the instrument fall within an appropriate distribution. Poor goodness of fit could allow the model to rely less heavily on that particular training data point in making future predictions. Additionally, there were some factors that appear lower on the y axis that are somewhat unexpected, such as aerosol optical thickness SWIR and solar zenith angle. Aerosol optical thickness SWIR describes the atmospheric density of aerosols that reflect in the SWIR band, which could be expected to be important for this prediction due to the importance of the other factors affecting the SWIR band that appear towards the top of the axis. Solar zenith angle is a fundamental factor in the calculation of the methane mixing ratio because it describes the angle of incident light, which is integral to remote sensing by satellites. That this factor is relatively unimportant suggests that this information is well

incorporated in the retrieval. The importances of variables here differ from the importances determined in Balasus et al. likely due to extent. This work's much smaller area focused on the Denver–Julesburg basin, which has a very limited range of surface albedo SWIR values, whereas the Balasus et al. global extent sees a range of 0.01–0.6 in some regions. The much smaller range of SWIR surface albedo here likely contributes to the lower overall importance. The extent likely also affects the importance of aerosol-related variables, which Balasus et al. also found to be significantly more important – our extent focused on the oil and gas basin with significant agricultural influence, which are two important sources of aerosols, but our proximity to sources may limit the range of aerosol-related values, making this term less important here than on a global extent as well.

This study utilized an implicit stripe correction instead of an explicit one. The UoL target data are not subjected to a striping effect, so the use of the target data and the use of the ground pixel index as a variable in the model allowed for a stripe-corrected dataset to be output from the input of non-stripe-corrected data. This process relies heavily on the ground pixel variable which finds middling importance in Fig. 4, indicating that while the stripe correction is important, other factors affect the overall output more. Other in-

formation describing the training and validation process is available in the Supplement.

### 3.3 Model corrections in practice

The trained models were then used to predict corrected  $\text{XCH}_4$  values on a monthly basis on data from April 2018 to December 2022, the correction values for which are depicted in Fig. S1 in the Supplement. The months of January and July, representing winter and summer data, respectively, are presented in Fig. 5. The model-predicted positive and negative correction values for these data appear to be seasonally dependent, with more positive corrections being made in colder months and negative corrections being made in warmer months, appearing as blue colors in the summer (Fig. 5a) and red colors in the winter (Fig. 5b). The correction values also show a specific geographic distribution: two curved lines, one curving upwards from Denver and the other curving down through Nebraska, appear to follow the South Platte River and North Platte River, respectively (dashed white lines in Fig. 5a). As Colorado has been described in the past as part of the Great American Desert, water sources like these two eventual tributaries to the Missouri River dictate where larger water-intensive agricultural operations exist. As such, larger densities of water-intensive crop farms are co-located with these rivers, bringing their albedo-influencing crops and plant life, and thus requiring an albedo correction which is not necessarily reflected in magnitude by the surrounding scrubland. It has been shown that water-intensive crops, like corn, sugar beets, and alfalfa, and drought-resistant crops, like winter wheat, millet, and dry beans, reflect SWIR light differently, allowing for identification of crops from space with the SWIR reflectance variable along with other variables (Chen et al., 2005). This effect is possibly due to water content or leaf size of the vegetable matter. The spatial extent of the water-intensive crops is much wider than the riverbed; the North Platte River and South Platte River are extremely small (average discharges of 38 and 5  $\text{m}^3 \text{s}^{-1}$  respectively; the Mississippi River is 16 800  $\text{m}^3 \text{s}^{-1}$ ) and are far less in extent than one satellite pixel, making the flagging or removing of these data due to water content unnecessary. In his book *Roughing It*, Mark Twain describes the South Platte River in 1870 as “shallow, yellow, muddy... and only saved from being impossible to find with the naked eye by its sentinel rank of scattering trees standing on either bank” (Twain, 1891).

Particularly prominent in Fig. 5a is a darker swath south of the upward bend in the South Platte River. This area also has many farms, but these farms are more likely to grow drought-resistant crops. Additionally, many more of these fields lie fallow in a given year than the ones irrigated by river water. Another area of agricultural significance is around Greeley, Colorado (white triangle). Greeley is also visible in the colder months maps, giving further indication that cropland is associated with albedo effects, but with magnitude or di-

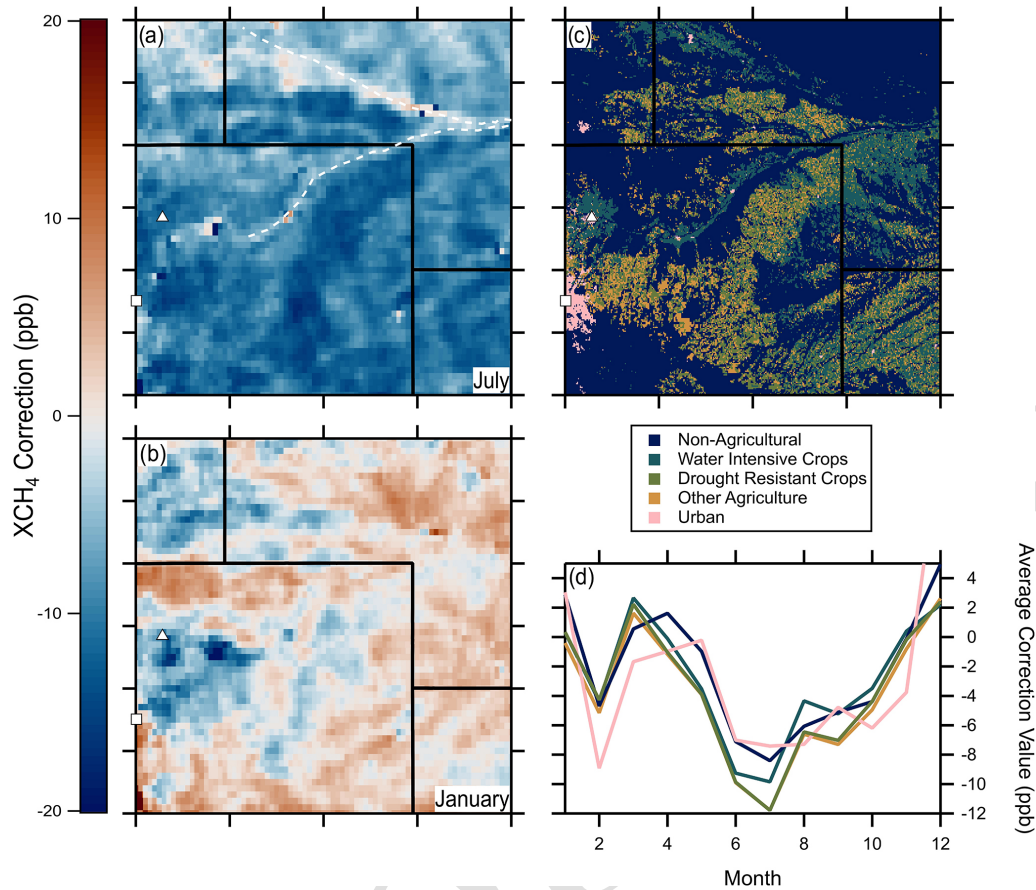
rection differing based on crop types and growing seasons. Greeley and the surrounding farms make up a large portion of the crop farming capacity within Weld County.

Figure 5c depicts the agricultural land use in the area of interest where visual comparison of the water-intensive crops and the bright-line regions of the summer seasonal albedo correction plots can be made. The numerical comparison agrees with visual inspections, as Fig. 5d depicts average albedo correction values over each kind of land cover. Over-arching seasonal trends appear, with corrections over all land covers appearing closer to 0 in the winter and fall and increasingly negative through the spring and summer. Additionally, seasonal effects over individual types of land cover are measurable. During the winter and fall, many of the land cover types appear very similar, while diverging from each other in the spring and summer, when vegetation in Colorado becomes increasingly stressed for water. That the urban points also follow the general seasonal trend is important and indicates that a driving factor in the seasonal albedo change is the relationship between surface albedo SWIR and other variables with the correction value and how that relationship changes seasonally.

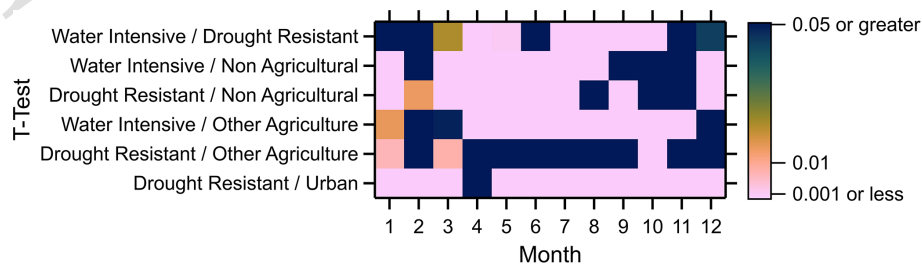
$T$  tests were performed between categories to determine the significance of the differences between the different land uses and presented in Fig. 6.  $T$  tests for each month of data on a small subset of 500 points for each land use demonstrate, for example, that drought-resistant crops and other agriculture types are not statistically significant.  $P$  values for the  $T$  tests between other land uses tend to increase and indicate no statistical significance in the winter and late fall, while indicating statistically high significance throughout the spring and summer for most land use pairs for most months. This indicates that in general the different land uses require different correction values, and this is related to the kinds of agriculture utilized. Water-intensive agriculture is likely irrigated, and soil moisture and vegetable water content can play a significant role in surface albedo SWIR, such that measurements of the like have been used to measure extents of irrigated agricultural land uses (Chen et al., 2005). This demonstrates that a seasonally resolved albedo correction, one that takes into account the changing relationship between the surface albedo SWIR and the correction value over time, is important and may be different in different parts of the world over different land cover types. Similarities between water-intensive and non-agricultural and drought-resistant and non-agricultural in the winter and fall indicate that non-agricultural land may not be as affected by the seasonal bias.

The corrected  $\text{XCH}_4$  data were calculated and averaged across the summer and winter months to demonstrate the difference the models developed here make in their corrections. Visually apparent in the uncorrected data and the Lorente et al. corrections in Fig. 7a, b, e, and f are structural features that are similar to features shown in the surface albedo SWIR retrieval (Fig. 7d and h). The corrected dataset devised here has

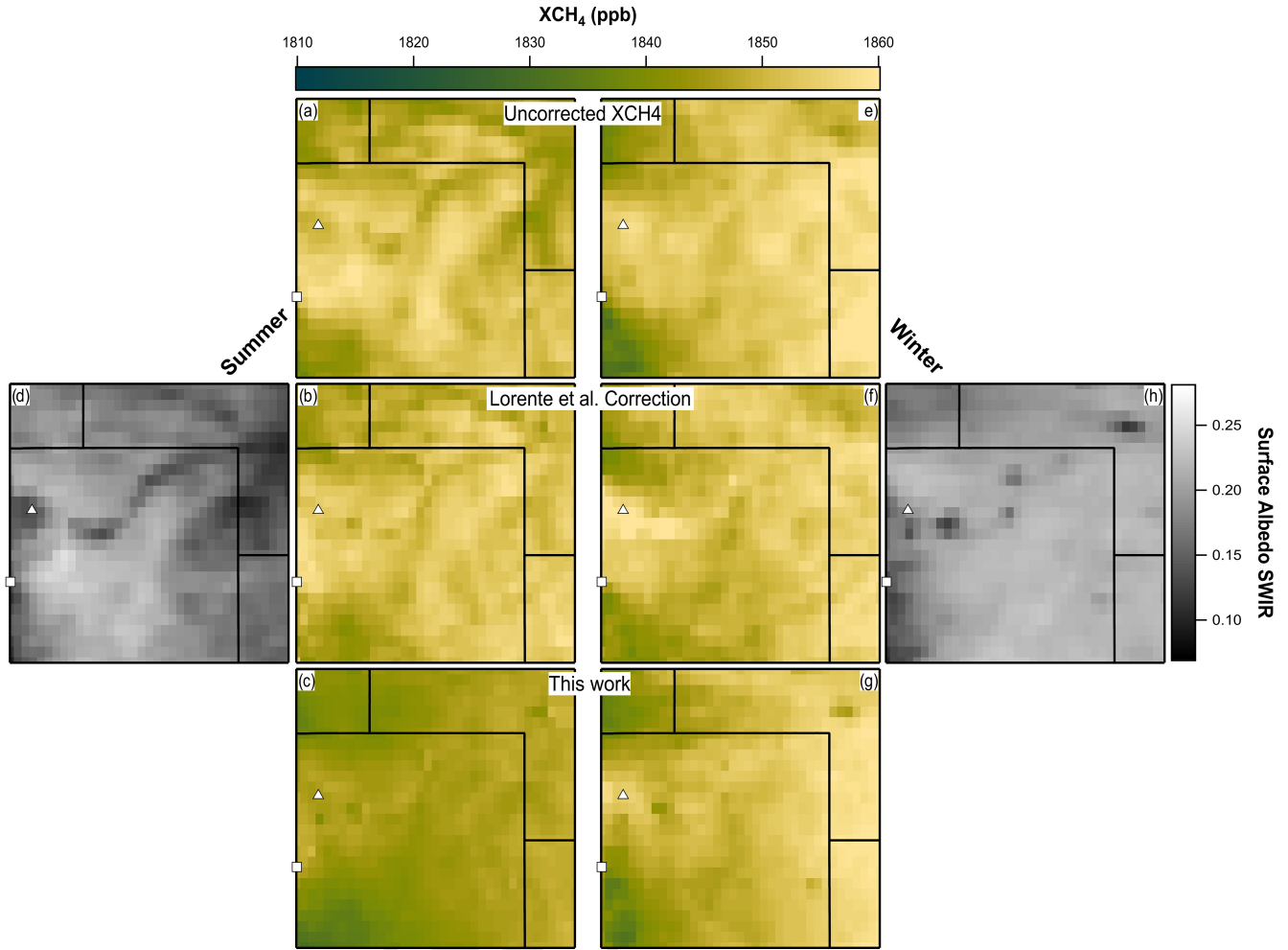




**Figure 5.** Average XCH<sub>4</sub> correction values for water-intensive vs. drought-resistant crops. XCH<sub>4</sub> correction value maps with dashed white lines representing the North Platte River and South Platte River for July (a) and January (b) representing data for the summer and winter, respectively. Locations of crop types (c) around the DJ basin. Water-intensive crops include corn, alfalfa, and sugar beets; drought-resistant crops include winter wheat, millet, and dry beans. Crops with both traits, fallow land, and other agricultural types are described as other agriculture. Grassland and other non-agricultural types, except urban areas, are described as non-agricultural. Developed land includes parts of the cities of Denver, Greeley (marked with the white square and triangle, respectively), Cheyenne, and other smaller communities. Average CH<sub>4</sub> correction values for the crop types (d) and drought-resistant crops require larger corrections throughout the summer months, while water-intensive crops are more similar to, though not the same as, the surrounding grasslands. No error bars are shown due to the large number of points making both standard error and 95 % confidence interval values too small to see. Crop data are from 2021 only and calculated using the April 2018–December 2022 correction data.



**Figure 6.** Significance tests demonstrating the statistical significance between paired datasets. All values that are not the darkest blue (equal to 0.05 or greater) are significantly different in a  $p$  critical (equal to 0.05) environment. All values that are pink (equal to 0.001 or less) indicate very significant differences. More blue early in the year and later in the year indicates that albedo corrections are more similar between different land cover types, and more pink in the summer months indicates that albedo corrections are more different between different land cover types in this time period.



**Figure 7.** Result of the methane albedo correction devised in this work. The uncorrected  $\text{XCH}_4$  data retrieved by TROPOMI (a) and the Lorente et al. correction in the summer (b) are compared against the correction devised by this work (c) and the average surface albedo SWIR retrieval map for this time period (d). This is repeated for the winter months on the right with the uncorrected retrieval (e), Lorente et al. correction (f), this work's correction (g), and the winter average surface albedo SWIR (h).

an average mixing ratio 6.9 ppb smaller than the Lorente et al. corrected data in the summer and 0.4 ppb smaller in winter, appearing slightly darker in color than the Lorente corrected data. This reduction is likely due to the new correction algorithm's dependence on the UoL GOSAT proxy retrieval target data, which on a global average measures 9.2 ppb less  $\text{XCH}_4$  than TROPOMI (Balasus et al., 2023). Notably, the Denver metropolitan area has lower average methane concentrations in our model output data than in the original TROPOMI Lorente et al. corrected data (6.4 ppb less in summer and 4.9 ppb less in winter). Figure 7 cannot be evaluated as before with a Pearson correlation because the correlation requires GOSAT–TROPOMI data to be used to account for natural correlation between surface albedo SWIR and  $\text{XCH}_4$ . There is not sufficient GOSAT data over this extent and time period to calculate such a Pearson correlation. Instead we assume that the tested model output correlations hold for these

data, making the correlations between GOSAT–TROPOMI and the surface albedo SWIR  $-0.03 \pm 0.04$  and  $0.01 \pm 0.08$  for winter and summer, respectively, for the models developed in this work and  $0.25 \pm 0.03$  and  $-0.1 \pm 0.1$  for the Lorente et al. correction values; error values are  $1\sigma$ . Overall it appears that the correction is effective in removing the albedo effect over seasonal time resolutions. This is important, as emissions calculation methods generally rely on local gradients. Fewer features in the methane distributions should coincide with lower emissions estimates.

#### 4 Conclusions

A small but significant seasonal dependence on surface albedo biases was found in TROPOMI  $\text{XCH}_4$  retrievals over Colorado even after the application of the current state-of-the-art albedo corrections. A series of deep learning ensem-

ble models specifically designed to reduce differences between TROPOMI and GOSAT while also reducing dependency on surface albedo in the SWIR have been developed to improve upon previous corrections. The output of the trained models removes the lasting seasonal dependence on surface albedo and demonstrates the fewest exceedances of a  $-0.1 < R < 0.1$  Pearson correlation with surface albedo in the TROPOMI dataset. Application of the albedo correction to the Denver–Julesburg basin reveals albedo correction dependencies on land cover, requiring larger-in-magnitude corrections in the summer months over drier, drought-resistant crops than irrigated water-intensive crops, with differences that also fluctuate seasonally. The 12 monthly models' seasonal albedo correction appears to resolve previously understudied issues surrounding long-term albedo corrections over seasonally changing areas, like cropland, making this a valuable tool for developing more accurate methane emissions inventories and models, as well as potentially deconvoluting relatively constant oil and gas emissions from seasonally dependent agricultural emissions. Methane measurements corrected utilizing this albedo correction method will be quantified in a forthcoming publication.

**Code availability.** The code used for all portions of this project is archived on Zenodo at <https://doi.org/10.5281/zenodo.12809441> (Bradley, 2024).

**Data availability.** The TROPOMI data used here are available at <https://browser.dataspace.copernicus.eu> (ESA, 2024) for April 2018–present. The GOSAT data used here are available at <https://doi.org/10.5285/18ef8247f52a4cb6a14013f8235cc1eb> (Parker and Boesch, 2020) for 2009–2021. The agricultural data used here are available at <https://nassgeodata.gmu.edu/CropScape/> (USDA NASS, 2025) [TSS](#).

**Supplement.** The supplement related to this article is available online at [the link will be implemented upon publication].

**Author contributions.** ACB and JdG designed the study. ACB performed the analysis with contributions from JdG and BD and led the writing of the paper with contributions from all co-authors.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every

effort to include appropriate place names, the final responsibility lies with the authors.

**Acknowledgements.** We would like to thank Ilse Aben, Ben Hmiel, and John Evans for useful discussions, as well as Bud Pope for the cooperation with BlueSky Resources. We would also like to thank Raquel Serrano-Calvo for her unpublished results that inspired us to perform this analysis. This work contains modified EU Copernicus Sentinel-5 Precursor TROPOMI data (2018–2024).

**Financial support.** This research has been supported by the Colorado Department of Public Health and Environment (CDPHE) (grant no. 2024\*2228) and the Cooperative Institute for Research in Environmental Sciences (CIRES) Graduate Fellowship.

**Review statement.** This paper was edited by Sandip Dhomse and reviewed by three anonymous referees.

## References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Zenodo [code], <https://doi.org/10.5281/zenodo.4724125>, 2015.
- Akoglu, H.: User's guide to correlation coefficients, *Turk. J. Emerg. Med.*, 18, 91–93, <https://doi.org/10.1016/j.tjem.2018.08.001>, 2018.
- Allen, D.: Attributing Atmospheric Methane to Anthropogenic Emission Sources, *Acc. Chem. Res.*, 49, 1344–1350, <https://doi.org/10.1021/acs.accounts.6b00081>, 2016.
- Allen, D. T.: Methane emissions from natural gas production and use: reconciling bottom-up and top-down measurements, *Curr. Opin. Chem. Eng.*, 5, 78–83, <https://doi.org/10.1016/j.coche.2014.05.004>, 2014.
- Annual Energy Outlook 2023: <https://www.eia.gov/outlooks/aeo/index.php>, last access: 16 January 2025.
- Apituley, A., Pedernana, M., Sneep, M., Veeffkind, P., Loyola, D., Hasekamp, O., Lorente, A., and Borsdorff, T.: Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Methane, SRON-S5P-LEV2-MA-001, SRON, <https://sentinels.copernicus.eu/documents/247904/2474726/Sentinel-5P-Level-2-Product-User-Manual-Methane.pdf> (last access: 14 September 2024), 2022.
- Balagus, N., Jacob, D. J., Lorente, A., Maasakkers, J. D., Parker, R. J., Boesch, H., Chen, Z., Kelp, M. M., Nesser, H., and Varon, D. J.: A blended TROPOMI+GOSAT satellite data product for atmospheric methane using machine learning to correct retrieval biases, *Atmos. Meas. Tech.*, 16, 3787–3807, <https://doi.org/10.5194/amt-16-3787-2023>, 2023.

- Beirle, S., Borger, C., Dörner, S., Eskes, H., Kumar, V., de Laat, A., and Wagner, T.: Catalog of  $\text{NO}_x$  emissions from point sources as derived from the divergence of the  $\text{NO}_2$  flux for TROPOMI, *Earth Syst. Sci. Data*, 13, 2995–3012, <https://doi.org/10.5194/essd-13-2995-2021>, 2021.
- Boucher, O., Friedlingstein, P., Collins, B., and Shine, K. P.: The indirect global warming potential and global temperature change potential due to methane oxidation, *Environ. Res. Lett.*, 4, 044007, <https://doi.org/10.1088/1748-9326/4/4/044007>, 2009.
- Bradley, A. C.: Tropomi\_seasonal\_albedo\_correction v1.0, Version v1.0, Zenodo [code] [TS6](https://doi.org/10.5281/zenodo.12809441), <https://doi.org/10.5281/zenodo.12809441>, 2024.
- Chen, D., Huang, J., and Jackson, T. J.: Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands, *Remote Sens. Environ.*, 98, 225–236, <https://doi.org/10.1016/j.rse.2005.07.008>, 2005.
- Chollet, F., [TS7](https://keras.io), et al.: Keras, <https://keras.io> (last access: 28 March 2025), 2015.
- Collins, W. J., Webber, C. P., Cox, P. M., Huntingford, C., Lowe, J., Sitch, S., Chadburn, S. E., Comyn-Platt, E., Harper, A. B., Hayman, G., and Powell, T.: Increased importance of methane reduction for a  $1.5^\circ$  target, *Environ. Res. Lett.*, 13, 054003, <https://doi.org/10.1088/1748-9326/aab89c>, 2018.
- Cook, T., Perrin, J., and Van Wagener, D.: Hydraulically fractured horizontal wells account for most of new oil and gas well, Hydraulically Fractured Horizontal Wells Account for Most of New Oil and Gas Well, <https://www.eia.gov/todayinenergy/detail.php?id=34732> (last access: 8 June 2024), 2018.
- Cramer, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, *Nat. Commun.*, 11, 5444, <https://doi.org/10.1038/s41467-020-19160-7>, 2020.
- da Costa-Luis, C. O.: tqdm: A Fast, Extensible Progress Meter for Python and CLI, *Journal of Open Source Software*, 4, 1277, <https://doi.org/10.21105/joss.01277>, 2019.
- de Gouw, J. A., Veefkind, J. P., Roosenbrand, E., Dix, B., Lin, J. C., Landgraf, J., and Levelt, P. F.: Daily Satellite Observations of Methane from Oil and Gas Production Regions in the United States, *Scientific Reports*, 10, 1379, <https://doi.org/10.1038/s41598-020-57678-4>, 2020.
- ESA: TROPOMI Methane Product, ESA [data set], <https://browser.dataspace.copernicus.eu/> (last access: 12 August 2024), 2024.
- Etiopie, G. and Schwietzke, S.: Global geological methane emissions: An update of top-down and bottom-up estimates, *Elementa: Science of the Anthropocene*, 7, 47, <https://doi.org/10.1525/elementa.383>, 2019.
- Gillies, S., van der Wel, C., Van den Bossche, J., et al.: Shapely, Version 2.1.0rc1, Zenodo [code], <https://doi.org/10.5281/zenodo.15032054>, 2025.
- Gillies, S., Perry, M., Ward, B., et al. [TS8](https://github.com/rasterio/rasterio): Rasterio: Geospatial Raster I/O for Python Programmers, GitHub [code], <https://github.com/rasterio/rasterio> (last access: 28 March 2025), 2013.
- Han, W., Yang, Z., Di, L., and Yue, P.: A geospatial Web service approach for creating on-demand Cropland Data Layer thematic maps, *T. ASABE*, 57, 239–247, <https://doi.org/10.13031/trans.57.10020>, 2014.
- Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, *Nature*, 585, 357–362, <https://doi.org/10.1038/s41586-020-2649-2>, 2020.
- Hunter, J. D.: Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.*, 9, 90–95, 2007.
- IEA: Global Methane Tracker, IEA, Paris, <https://www.iea.org/reports/global-methane-tracker-2023> (last access: 28 March 2025) [TS9](https://www.iea.org/reports/global-methane-tracker-2023), 2023.
- Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., and Frankenberg, C.: Satellite observations of atmospheric methane and their value for quantifying methane emissions, *Atmos. Chem. Phys.*, 16, 14371–14396, <https://doi.org/10.5194/acp-16-14371-2016>, 2016.
- Jordahl, K., Bossche, J. V. D., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., Maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journais, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, and Leblanc, F.: Geopandas/Geopandas: v0.8.1, Version v0.8.1, Zenodo [code], <https://doi.org/10.5281/ZENODO.3946761>, 2020.
- Kuckartz, U., Rädiker, S., Ebert, T., and Schehl, J.: Statistik: Eine verständliche Einführung, 2., überarb. Aufl., 2013th edn., VS Verlag für Sozialwissenschaften, Wiesbaden, 316 pp., ISBN: 978-3-531-19889-7, 2013.
- Landgraf, J., Lorente, A., Langerock, B., and Kumar Sha, M.: S5P Mission Performance Centre Methane [L2\_CH4\_] Readme; S5P MPC Product Readme Methane, SRON, p. 4, <https://doi.org/10.5270/S5P-3lcdqiv>, 2023.
- Levelt, P. F., Stein Zweers, D. C., Aben, I., Bauwens, M., Borsdorff, T., De Smedt, I., Eskes, H. J., Lerot, C., Loyola, D. G., Romahn, F., Stavrakou, T., Theys, N., Van Roozendael, M., Veefkind, J. P., and Verhoelst, T.: Air quality impacts of COVID-19 lockdown measures detected from space using high spatial resolution observations of multiple trace gases from Sentinel-5P/TROPOMI, *Atmos. Chem. Phys.*, 22, 10319–10351, <https://doi.org/10.5194/acp-22-10319-2022>, 2022.
- Liu, M., van der A, R., van Weele, M., Eskes, H., Lu, X., Veefkind, P., de Laat, J., Kong, H., Wang, J., Sun, J., Ding, J., Zhao, Y., and Weng, H.: A New Divergence Method to Quantify Methane Emissions Using Observations of Sentinel-5P TROPOMI, *Geophys. Res. Lett.*, 48, e2021GL094151, <https://doi.org/10.1029/2021GL094151>, 2021.
- Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velasco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., and Landgraf, J.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements, *Atmos. Meas. Tech.*, 14, 665–684, <https://doi.org/10.5194/amt-14-665-2021>, 2021.
- Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, *Adv. Neur. In.*, 30, 4765–4775, 2017.
- National Weather Service (NWS): Rivers of the U.S., NWS [TS10](https://www.weather.gov/gis/Rivers) [data set], <https://www.weather.gov/gis/Rivers>, last access: 16 February 2024.



- Omara, M., Zimmerman, N., Sullivan, M. R., Li, X., Ellis, A., Cesa, R., Subramanian, R., Presto, A. A., and Robinson, A. L.: Methane Emissions from Natural Gas Production Sites in the United States: Data Synthesis and National Estimate, *Environ. Sci. Technol.*, 52, 12915–12925, <https://doi.org/10.1021/acs.est.8b03535>, 2018.
- Parker, R. and Boesch, H.: University of Leicester GOSAT Proxy XCH<sub>4</sub> v9.0, Centre for Environmental Data Analysis [data set], <https://doi.org/10.5285/18ef8247f52a4cb6a14013f8235cc1eb>, 2020.
- Pétron, G., Karion, A., Sweeney, C., Miller, B. R., Montzka, S. A., Frost, G. J., Trainer, M., Tans, P., Andrews, A., Kofler, J., Helmig, D., Guenther, D., Dlugokencky, E., Lang, P., Newberger, T., Wolter, S., Hall, B., Novelli, P., Brewer, A., Conley, S., Hardesty, M., Banta, R., White, A., Noone, D., Wolfe, D., and Schnell, R.: A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin, *J. Geophys. Res.-Atmos.*, 119, 6836–6852, <https://doi.org/10.1002/2013JD021272>, 2014.
- Riddick, S. N., Mbua, M., Santos, A., Hartzell, W., and Zimmerle, D. J.: Potential Underestimate in Reported Bottom-up Methane Emissions from Oil and Gas Operations in the Delaware Basin, *Atmosphere*, 15, 202, <https://doi.org/10.3390/atmos15020202>, 2024.
- Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence*, 1, 206–215, <https://doi.org/10.1038/s42256-019-0048-x>, 2019.
- Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., Blavier, J.-F., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., and Aben, I.: Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, *J. Geophys. Res.-Atmos.*, 117, D10307, <https://doi.org/10.1029/2012JD017549>, 2012.
- Schober, P., Boer, C., and Schwarte, L. A.: Correlation Coefficients: Appropriate Use and Interpretation, *Anesth. Analg.*, 126, 1763, <https://doi.org/10.1213/ANE.00000000000002864>, 2018.
- Takuya, A., Shotaro, S., Toshihiko, Y., Takeru, O., and Masanori, K.: Optuna: A Next-Generation Hyperparameter Optimization Framework, in: KDD '19: The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 August 2019, Association for Computing Machinery, 2623–2631, <https://doi.org/10.1145/3292500.3330701>, 2623–2631, 2023. **TS11**
- The pandas development team: pandas-dev/pandas: Pandas, Version v2.2.3, Zenodo [code], <https://doi.org/10.5281/zenodo.13819579>, 2024.
- Twain, M.: *Roughing It*, American Pub. Co., Hartford, Conn., ISBN: 978-0451531100, 1891.
- United States Department of Agriculture (USDA): County Profile: Weld County, 2017 Census of Agriculture, USDA, 1–2, [https://www.nass.usda.gov/Publications/AgCensus/2017/Online\\_Resources/County\\_Profiles/Colorado/cp08123.pdf](https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/County_Profiles/Colorado/cp08123.pdf) (last access: 28 March 2025 **TS12**), 2017.
- U.S. Census Bureau: Cartographic Boundary Files - Shapefile, U.S. Census Bureau [data set], <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>, last access: 16 February 2024.
- USDA NASS (National Agricultural Statistics Service): CropScape - Cropland Data Layer (CDL), <https://nassgeodata.gmu.edu/CropScape/>, last access: 28 March 2025. **TS13**
- U.S. EIA: Primary Energy Production Estimates, State Energy Data System (SEDS), United States Energy Information Administration, United States, <https://www.eia.gov/state/seds/archive/> (last access: 5 June 2024), 2020.
- U.S. EIA: Drilling Productivity Report, <https://www.eia.gov/petroleum/drilling/> (last access: 5 June 2024), 2023.
- U.S. Energy Information Administration (EIA): <https://www.eia.gov/outlooks/aeo/data/browser/#?id=1-AEO2023&region=0-0&cases=ref2023&start=2021&end=2050&f=A&linechart=ref2023-d020623a.6-1-AEO2023&ctype=linechart&sourcekey=0>, last access: 4 January 2024.
- Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sens. Environ.*, 120, 70–83, <https://doi.org/10.1016/j.rse.2011.09.027>, 2012.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat. Methods*, 17, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- Whitaker, J.: Netcdf4-Python: Python/Numpy Interface to the netCDF C Library, <http://unidata.github.io/netcdf4-python> (last access: 28 March 2025 **TS14**), 2008.
- Zavala-Araiza, D., Lyon, D. R., Alvarez, R. A., Davis, K. J., Harriss, R., Herndon, S. C., Karion, A., Kort, E. A., Lamb, B. K., Lan, X., Marchese, A. J., Pacala, S. W., Robinson, A. L., Shepson, P. B., Sweeney, C., Talbot, R., Townsend-Small, A., Yacovitch, T. I., Zimmerle, D. J., and Hamburg, S. P.: Reconciling divergent estimates of oil and gas methane emissions, *P. Natl. Acad. Sci. USA*, 112, 15597–15602, <https://doi.org/10.1073/pnas.1522126112>, 2015.
- Zhang, Y., Gautam, R., Pandey, S., Omara, M., Maasakkers, J. D., Sadavarte, P., Lyon, D., Nesser, H., Sulprizio, M. P., Varon, D. J., Zhang, R., Houweling, S., Zavala-Araiza, D., Alvarez, R. A., Lorente, A., Hamburg, S. P., Aben, I., and Jacob, D. J.: Quantifying methane emissions from the largest oil-producing basin in the United States from space, *Science Advances*, 6, eaaz5120, <https://doi.org/10.1126/sciadv.aaz5120>, 2020.



### Remarks from the language copy-editor

- CE1** Please note that we can only insert minor technical and necessary changes at this stage without editor approval. Any content changes need to be approved by the editor. If you'd still like to proceed with the remaining changes to this sentence (i.e., the removal of several packages), please provide an explanation for the editor.

### Remarks from the typesetter

- TS1** Please confirm.  
**TS2** Please confirm.  
**TS3** Please confirm.  
**TS4** Please confirm.  
**TS5** Please confirm the inserted citation.  
**TS6** Please confirm.  
**TS7** Thank you very much for your feedback. However, according to our referencing standards, at least three authors have to be cited if "et al." is used. Thank you.  
**TS8** Please see the reference "Chollet" for my comment concerning the list of authors.  
**TS9** Please confirm.  
**TS10** Please confirm the repository/publisher.  
**TS11** Please confirm inserted information.  
**TS12** Please confirm.  
**TS13** Please confirm the inserted reference.  
**TS14** Please confirm.