A Study on Inferring Daytime Variations of XCO_2 from Current and Future Space-Based Missions

Calla Marchetti[1,2,*], Jonathan Hobbs[1], Peter Somkuti[3,4], Joshua L. Laughner[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2]Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, USA
[3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[4]Global Modeling and Assimilation Office, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD, USA
[*]now at: Department of Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA

Key Points:

- We attempted to train a machine learning model to reproduce daytime variations in XCO_2 over $\sim$ 6 h from sparse satellite observations.
- We found that under current satellite sampling patterns our method was unable to rigorously reproduce these variations.
- We found the optimized parameters that would allow machine learning to constrain these variations with future missions.

Corresponding author: Joshua Laughner, josh.laughner@jpl.nasa.gov

Abstract

Net ecosystem exchange (NEE) measures the net transfer of carbon between terrestrial ecosystems and the atmosphere, and is an important quantity for understanding land-atmosphere feedbacks and constraining the land carbon sink. Atmospheric inverse models and biophysical models provide regional and global NEE estimates, but validation of these models is limited by the sparsity of flux towers. NEE can also be calculated from the change in XCO_2 over the course of a day. XCO_2 is observed by the Orbiting Carbon Observatory 2 and 3 (OCO-2 and -3) satellites, which working together have the potential to observe locations between $\sim 52°$S and 52°N twice a day but at a sparse temporal frequency. Here, we investigate the possibility of using machine learning (ML) to extrapolate the variation in XCO_2 over daytime hours, which could be in turn be used to derive NEE.

We find that the current temporal sampling from OCO-2 and -3 is not ideal for this purpose, and our ML approach is not able to reliably infer either the daily patterns of XCO_2 or the difference of XCO_2 across solar noon. A thrice-daily observation pattern, such as could be achieved with a GeoCarb-like (geosynchronous) instrument, provides much better performance. It is also essential that systematic biases between observations at different times of day be minimized, as the ability to predict daily variation or morning to afternoon differences decreases when the standard error between the means of observations at different times of day exceeds ~ 0.1 ppm.

Acronyms

EOF Empirical orthogonal function
EVI Enhanced Vegetation Index
GeoCarb Geostationary Carbon Cycle Observatory
GEOS FP-IT Goddard Earth Observing System Forward Processing for Instrument Teams
GOSAT Greenhouse Gas Observing Satellite
ISS International Space Station
KDE Kernel density estimate
ML Machine Learning
MODIS Moderate Resolution Imaging Spectrometer
NEE Net Ecosystem Exchange
NDVI Normalized Difference Vegetation Index
OCO Orbiting Carbon Observatory
PC Principal Component
SIF Solar Induced Fluorescence
SMAP Soil Moisture Active and Passive
SMOS Soil Moisture and Ocean Salinity
TCCON Total Carbon Column Observation Network
XCO_2 the column average mole fraction of carbon dioxide in the atmosphere

1 Introduction

Net ecosystem exchange (NEE) describes the magnitude of carbon flux between terrestrial ecosystems and the atmosphere, calculated as the difference between ecosystem respiration (RE) and gross primary productivity (GPP, W. Zhang et al., 2023; Xiao et al., 2008). Accurate measurement and validation of NEE is crucial in understanding land-atmosphere feedbacks, constraining regional land-carbon sinks, and facilitating climate policy (W. Zhang et al., 2023; Zeng et al., 2020). There are currently multiple methods of calculating NEE. Eddy covariance flux towers provide continuous time series mea-

surements of NEE, and there are currently over 400 towers around the globe (Xiao et al., 2008). The data from flux towers is used in model validation and has given fundamental insight to carbon cycle processes. However, flux towers provide site-scale observations, meaning the fluxes captured only represent the tower footprint up to a few square kilometers. The sparsity and uneven distribution of flux tower sites makes it difficult to extrapolate eddy covariance measurements to regional or global scales (Jung et al., 2019; W. Zhang et al., 2023; Xiao et al., 2008).

Atmospheric inverse models and biophysical models are also used to provide large scale estimates of NEE. However, atmospheric inverse models are known to have biases, and biogeochemical models are dependent on site level parameterizations (Xiao et al., 2008). When ecological variables such as plant functional type or soil depth are poorly mapped, or when nonlinear relationships such as the response of NEE to light are not mathematically constrained, biophysical model accuracy is limited (Zhou et al., 2019). Additionally, both model types are evaluated and validated against flux towers, but the sparsity of flux towers that prevents extrapolation of eddy covariance measurements also hinders model validation (W. Zhang et al., 2023; Zeng et al., 2020).

Diurnal cycles of XCO_2 over the course of a day (where XCO_2 is the column average dry air mole fraction of CO_2 in the atmosphere) have also been used to calculate NEE. (Here, we use "diurnal cycle" to refer to the variation of XCO_2 during the daytime part of a 24-hour period.) The diurnal cycle of XCO_2 will, in part, reflect the balance between photosynthesis and respiration (and the resulting NEE) upwind of the observation: when photosynthesis dominates, the near-surface CO_2 will decrease over the course of a day. This pattern can also be seen in the column XCO_2; however, XCO_2 is also affected by transport of CO_2 in the free troposphere. Synoptic-scale motion of air can induce confounding changes in the XCO_2 value that do not relate to the local NEE by bringing in an air mass containing a different mole fraction of CO_2.

A 2012 study tested whether diurnal cycles of XCO_2 from the Total Carbon Column Observation Network (TCCON) could be related to NEE and found that the four hour difference in XCO_2 centered around solar noon, termed "XCO_2 drawdown," can be used to calculate NEE that correlates with NEE calculated from flux towers (Keppel-Aleks et al., 2012). They found that weekly to monthly aggregation of the inferred drawdown was necessary to separate the synoptic-scale transport effect from the effect of NEE and thus achieve clear correlation with NEE measured from a colocated flux tower. Additionally, TCCON sites, similarly to flux towers, are limited to $\sim$30 stations around the globe at time of writing.

If NEE could be similarly inferred from global space-based XCO_2 observations, this would be a valuable constraint on carbon cycle models. While GPP has been shown to correlate with solar induced fluorescence (SIF) observed from space (Frankenberg et al., 2011), NEE has proven difficult because of the ecosystem respiration component. Previous efforts to apply machine learning (ML) to this problem have focused on, for example, upscaling NEE as measured by eddy covariance using a variety of remotely sensed data (Ichii et al., 2017). In contrast, XCO_2 is a parameter that can be observed from space. Satellites like the Orbiting Carbon Observatory (OCO) 2 and 3 observe XCO_2 at a global scale, with greater spatial density but lesser temporal frequency than TCCON or flux towers. The ability to extrapolate XCO_2 diurnal cycles from temporally sparse satellite measurements, and therefore calculate NEE from satellites, would allow for model validation on a broad scale.

Of particular interest is the potential to provide a source of validation for regional differences in NEE. Cui et al. (2022) evaluated models from the OCO-2 Model Intercomparison Project (OCO-2 MIP, Crowell et al., 2019; Peiro et al., 2022) against aircraft data from the Atmospheric Carbon and Transport—America mission and found that the models underestimated the seasonal NEE amplitude in North America. In Gier et al. (2024)

the land-atmosphere carbon flux simulated by models in the sixth Coupled Model Intercomparison Project (CMIP6, Eyring et al., 2016) was evaluated against three reference data sets, all of which involved models or inversions themselves. If NEE were able to be inferred on a regional scale from space-based XCO_2 data, this would provide a valuable dataset to test improvements to earth system models' ability to simulate the different balance and timing of photosynthesis and respiration in regions across the world.

In this study, we evaluate the possibility of using ML to reconstruct diurnal cycles of XCO_2 (during the core daylit hours only, as discussed above) or XCO_2 drawdown from temporally sparse XCO_2 observations. At a minimum, the ability to infer drawdown (and therefore an estimate of NEE, given sufficient samples to average out the impact of transport) would provide a useful counterpart to transport model-based approaches to infer carbon fluxes over much of the world. We also explore the possibility of reproducing the diurnal variation of XCO_2 over a 6.5 hour period centered on solar noon, as the ability to inspect these diurnal cycles might allow users to look for patterns indicative of other processes besides fluxes impacting the XCO_2 amounts.

The structure of this paper is as follows. We first describe our ML model design and data preparation (Sect. 2), then analyze the model performance given two simulated XCO_2 observations per day at times that follow the distributions of OCO-2 and OCO-3 overpass times (Sect. 3). Section 4 examines how this approach would perform given different scenarios of denser temporal sampling, from future CO_2 observing missions and infrequent multiple crossings per day of OCO-2 and OCO-3.

2 Methods

2.1 Method overview

Our goal is to test whether a machine learning approach can infer (1) the variation of XCO_2 over a 6.5 hour period centered on solar noon for individual days and/or (2) the difference in XCO_2 between 2 hours before and 2 hours after solar noon. Throughout the remainder of this work, we refer to (1) as the "diurnal cycle" and (2) as the "drawdown" for simplicity. The latter follows the terminology from Keppel-Aleks et al. (2012). The reasoning for the 6.5 hour window in the former is explained in Sect. 2.4.

To perform this test, we need a source of both true diurnal cycles and drawdown values and simulated XCO_2 observations that mimic the temporal sampling of current XCO_2 observing satellites or hypothetical future missions. Data from the TCCON network provide a readily accessible set of true diurnal cycles and drawdowns (within the limits of retrieval error, which is small enough to treat these data as truth for the purposes of this study), as a TCCON site reports XCO_2 values for all sunlit and cloud-free hours of a day. We can then sample specific times of day from the TCCON data to serve as our simulated satellite observations.

The times of day at which to sample TCCON data to simulate satellite observations are either determined by analysis of currently operational satellite ground tracks or hypothetical plans for future missions. We use OCO-2 and OCO-3 to represent current satellite missions, as their markedly different orbits allow them to potentially observe a given location at different times of day. Once we identify such cases where their orbits cross, we use all crossings within 10° of a TCCON site to determine the distribution of simulated observation times to use for that site. This process is described in more detail in Sect. 2.5.

These simulated satellite observations, along with additional variables sourced either from TCCON data or the GEOS FP-IT meteorological model, are used as input to the machine learning model. This is what we call our "feature set" and will be described

Table 1. TCCON sites used in this study. The "Days Post EOF Filter" and "Days Post Growing Filter" give the number of days of data (obtained on 01/20/2024) for each site that passes our criteria described in Sects. 2.4 and 3.2, respectively.

Site	Lat/Lon	Days Post EOF Filter	Days Post Growing Filter	Data Citation
East Trout Lake (ETL)	54.4N/105.0W	535	316	(Wunch et al., 2022)
Park Falls (PF)	45.9N/90.3W	914	625	(Wennberg, Roehl, et al., 2022)
Lamont	36.6N /97.5W	1329	1041	(Wennberg, Wunch, et al., 2022)
Lauder	45.0S/169.7E	722	604	(Sherlock et al., 2022a, 2022b) (Pollard et al., 2022)
Izana	28.3N/16.5W	87	87	(García et al., 2022)
Nicosia	35.1N/33.4E	219	219	(Petri et al., 2024)

in detail in Sects. 2.6 and 2.7. The model is trained to reproduce the diurnal cycles as observed by TCCON.

2.2 Data Sources

TCCON is a network of ground based spectrometers that continuously measure XCO_2 throughout the sunlit hours of each day (Wunch et al., 2011; Laughner et al., 2024). As we are interested in carbon fluxes to or from the biosphere, we include TCCON data from the sites at Park Falls, East Trout Lake, Lamont, Lauder, Izana, and Nicosia. These sites were chosen because they are (1) clearly distant from anthropogenic emissions that could overwhelm photosynthetic sources, (2) have sufficient days with data across enough hours of data for our analysis, and (3) are within (or nearly within) the $\sim 52^\circ$ S to 52° N range in which OCO-3 has observations (due to the inclination of its orbit). TCCON data are the foundation of our model; we create our target variables from TCCON XCO_2 data, our feature set consists largely of subsampled TCCON variables, and we validate our model against TCCON XCO_2 diurnal cycles. The sites used, their locations, and number of days per site are listed in Table 1.

OCO-2 and OCO-3 are NASA instruments that measure XCO_2 from space. OCO-2 was launched in 2014 and flies in a sun-synchronous polar orbit. It has a local overpass time of around 13:36, or about 1 hour after solar noon, with a 16 day repeat cycle (Taylor et al., 2023; Wunch et al., 2017). OCO-3 began operation in 2019, and its spectrometer is a direct copy of the one in OCO-2, and as such both instruments have the same precision of better than 1 ppm per sounding. The main difference between the two satellites is that OCO-3 is mounted on the International Space Station (ISS) and has a varying time of day overpass (Eldering et al., 2019; O'Dell et al., 2018). Note that we are interested in when and where the OCO satellites observe the same location in a single day to construct our model training data. Thus, we use latitude, longitude, and time data to identify these crossings, but do not use XCO_2 values retrieved by OCO-2 or OCO-3 in this work.

GEOS FP-IT is a 3 dimensional reanalysis product starting from 1998 at a three-hourly frequency and 50 km spatial resolution (Lucchesi, 2018). We use GEOS FP-IT

Table 2. Hyperparameter values used in our model. The "Parameter" column names the XGBoost parameter of interest. The "Effect" column describes the parameter's effect. The "Optimized Value" column lists the resulting values from our hyperparameter tuning, where we run the model until the goodness-of-fit metrics converge.

Parameter	Effect	Optimized Value
n_estimators	Number of Boosting Rounds	800
eta	Learn Rate	0.07
max_depth	Maximum tree levels	8
min_child_weight	Minimum weight (hessian) needed in a child	8
lambda	Regularization term on weights	2
alpha	Regularization term on weights	2

to supplement our feature set. From GEOS FP-IT we add the change in potential temperature at 700 mb, which we calculate from 3d temperature and pressure files, and vapor pressure deficit (VPD), which we calculate from temperature at 10 m, specific humidity at 10 m, and surface pressure (Sect. 2.6).

2.3 Model Design Overview

We employ an XGBoost model for this project, which is an ensemble learning method used for classification and regression. XGBoost is founded on decision trees, which are trained to predict the target variable using relationships between the target variable and multiple input variables. Unlike a random forest, which creates all of the decision trees in parallel and averages their predictions, boosting methods create trees sequentially, with each tree more powerful than the previous. Using XGBoost, the residuals from each tree's prediction are calculated, and points with large residuals are given a higher weight for the next tree (Chen & Guestrin, 2016).

The hyperparameter values used in our XGBoost model are listed in Table 2. The values were chosen by iteratively changing the values until our goodness-of-fit metrics converged. The majority of our hyperparameter tuning had the effect of making the algorithm less complex and more applicable to data it hasn't seen before.

As input variables (i.e., the feature set), our model takes simulated XCO_2 observations created by sampling TCCON data at specific times of day along with ancillary variables that are a combination of additional TCCON data and values sampled from the GEOS FP-IT meteorological model. The details are given in Sects. 2.6 and 2.7. The outputs of the model are not the XCO_2 values of the diurnal cycle directly, but rather the principal components (PCs) that when multiplied by a set of empirical orthogonal functions (EOFs) approximate the "true" diurnal cycle for one day of data from one TCCON site. This has the advantage of reducing the number of output parameters the model is required to produce. Construction of the EOFs is described in Sect. 2.4.

The final diurnal cycle and drawdown are then computed from the model's predicted PCs. The diurnal cycle is approximated following Eq. (1),

$$\mathbf{x} = \sum_{i=1}^{6} \phi_i \mathbf{e}_i \tag{1}$$

where ϕ represents a principal component and $\mathbf{e}$ an EOF. We find that 6 EOFs are sufficient to capture 95% of the variability in the original TCCON data. (See Sect. 2.4 for details.) The drawdown is then computed from the reconstructed diurnal cycle ($\mathbf{x}$ in Eq. (1)) as the difference between the XCO_2 values on two hours either side of solar noon, following Keppel-Aleks et al. (2012).

2.4 EOF decomposition of TCCON data

As discussed in Sect. 2.3, we represent the diurnal cycles as a linear combination of EOFs, which are derived through decomposition of daily TCCON data into EOFs. This decomposition requires our data to be on consistent time intervals. To maximize both hours in the day and number of days available we chose quarter hour intervals over the 6.5 hours centered around solar noon. This interval was chosen such that our six TCCON sites reliably report data for this full time span in all seasons. Otherwise, EOFs would have to be derived for different seasons separately (to allow for longer summer days and shorter winter days), potentially introducing step changes in the model as it shifts from one season's EOFs to another. For each time bin we average together all points within half an hour of time; this wider window of averaging allows us to smooth out spikes caused by outliers or gaps in data. If there are less than two data points in a window, that bin is left empty, and at the end of each day, if there are more than two empty bins in a row, or if the empty bins occur at the endpoints of the day, that day is discarded. For any remaining days, the empty bins are filled via cubic spline interpolation. This method of discarding days is to ensure that interpolation does not skew the day's endpoints nor completely fill in days with sparse data.

Because we are interested in diurnal cycles, we need to remove day-to-day, seasonal, and interannual variation from the TCCON data before carrying out EOF decomposition. Without this detrending step, the dominant patterns of variability will be these longer-term ones rather than diurnal. To detrend, we fit a parabola to each day's solar zenith angle measurements, and take the time of the minimum as solar noon. We then average together all XCO_2 values within half an hour of this time to obtain the XCO_2 value at solar noon, which we then subtract off of each measurement from that day. We use the "Climate Data Toolbox for Matlab" (Greene et al., 2019) to decompose our detrended and averaged data from all TCCON sites except the one withheld for validation (see Sect. 2.9) into six EOFs, which is the number required to reach 95% explained variance, and the PCs of this decomposition make up our target variable set. Our six EOFs and an example of diurnal cycle reconstruction are shown in Fig. 1. The same EOFs (for a given iteration of the k-fold cross validation, Sect. 2.9) are thus used for model predictions for all TCCON sites used in this study. Because Fig. 1 demonstrates how the diurnal cycle for one day is reconstructed with the general EOFs, the PC values given in the insets are not ordered largest to smallest. Instead, the PC values given are those needed to reconstruct this particular day. The overall percent variance explained by the cumulative sum of the first N EOFs is given in the legend: the first EOF explains 45% of the variability in the full dataset, the second explains an addition 36% for a total of 81% explained variance for the first two EOFs, and so on.

To confirm that the EOFs are not dominated by any one TCCON site, we also performed the decomposition on each sites' data separately (Fig. S2). We see that the first two EOFs at all sites can be combined to produce a different slope in the morning and afternoon, either by having each one with one half the day with no trend and the other half with a trend, or by having a combination of one EOF with opposite morning and afternoon trends and another with morning and afternoon trends of the same direction. The third through sixth EOFs for all sites show higher frequency patterns which are similar across the sites.

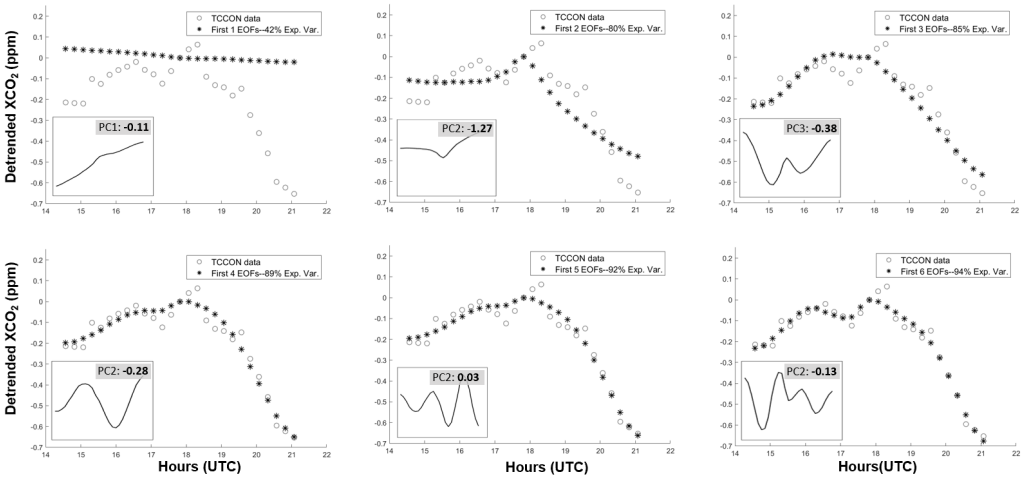

Figure 1. An example diurnal cycle from TCCON and how we use EOFs to represent it. The panels show an EOF representation of diurnal cycles as we increase the number of EOFs included. Also shown is the shape of each EOF and the PC applied to the EOF to reproduce our example day.

As a reminder, EOFs are a mathematical decomposition of data into patterns of variability. There is no requirement that they relate to physical drivers or phenomena, though that often happens to be the case. Additionally, the sign of the EOF is irrelevant, as it can be easily flipped by multiplication with a negative PC. In Fig. 1, we can see that the first EOF represents a roughly consistent trend in XCO_2 throughout the day and the second an afternoon-only trend in XCO_2. These can be summed in a linear combination to give an approximate diurnal cycle with a different morning and afternoon slope of XCO_2 versus time. The third through sixth EOFs then simply represent patterns of higher frequency variation, with no obvious relationship to physical drivers of XCO_2. We use EOFs here solely as a method of dimensionality reduction, and do not need to ascribe physical meaning to each EOF.

2.5 Sampling TCCON data to simulate satellite observations

As mentioned in Sect. 2.1, we sample TCCON data at specific times of day to simulate more temporally sparse satellite observations. This sampling comes in two types:

1. sampling designed to mimic OCO-2 and/or OCO-3, and
2. prescribed sampling to test a hypothetical future mission.

The second approach is extremely straightforward; we simply prescribe a set of times throughout the day to test. The first approach is more complicated. First, we analyze two years of OCO-2 and OCO-3 data to identify instances where an OCO-2 and OCO-3 orbit intersect or multiple OCO-3 orbits intersect themselves. The details of this algorithm are presented in the following section, 2.5.1. Once we have this set of known crossings, we construct a distribution of times for each TCCON site and randomly sample from that distribution to determine the times to sample TCCON data to mimic OCO-2 and OCO-3. The specifics of this sampling depend on whether we are simulating OCO-2/OCO-3 crossings or OCO-3 self crossings, and are detailed in Sects. 2.5.2 and 2.5.3, respectively.

For both OCO-2/OCO-3 crossings and OCO-3 self crossings, we use the crossings identified to build a probability distribution to draw from when choosing sampling times for each day of data used in the training or testing. This allows us to use days of TCCON data without an actual crossing, and thus increases the amount of training and testing data over sampling only TCCON days with a real crossing. The end result is a dataset where, for each day with sufficient TCCON data, we have subset that TCCON data to only two times of day, with times chosen to mimic the times that OCO-2 and/or OCO-3 could observe.

2.5.1 Determining OCO-2/3 Crossings

To determine times of day that are representative of when OCO-2 and -3 would observe the same area within some time window (in our case, 12 h), we need to identify cases in which OCO-2 and -3 observe the same location. Our algorithm to find such cases when the OCO-2 and OCO-3 observations cross has two steps. The first is a simple brute-force search. For each OCO-2 sounding of interest, we scan all OCO-3 soundings for the day before, day of, and day after the OCO-2 sounding and list those OCO-3 soundings within 100 km and 12 h of the OCO-2 sounding. We repeat this process for all OCO-2 soundings to produce a sequence of one-to-many maps of one OCO-2 sounding to one or more OCO-3 soundings that meet the 100 km and 12 h criteria.

Second, we group these paired soundings together into crossings. For each one-to-many map of one OCO-2 sounding to one or more OCO-3 soundings from the first step, we check if any of the OCO-3 soundings have already been assigned to a crossing. If so, we add the OCO-2 sounding and all the matched OCO-3 soundings to that existing crossing. If not, we create a new crossing with the OCO-2 sounding and its matched OCO-3 soundings. In graph theory terms, if the OCO-2 and -3 soundings are nodes and a match between an OCO-2 and -3 sounding is an edge, then a crossing is a subset of nodes for which a route between any two nodes in the subset exists in the graph. The end result is illustrated by Fig. 2. This figure shows how a crossing (represented by the gray boxes) is built up from individual pairs of OCO-2 and -3 soundings that are within the distance limit (indicated by an arrow connected a circle labeled with a "2" to one with a "3" - the letters represent different soundings). Information about these crossings are written to netCDF files for later analysis. In the final dataset, the crossings span up to ~ 300 km east to west and up to ~ 500 km north to south (Fig. S5). (Note, the crossings exceed 100 km in total length because a sequence of soundings less 100 km apart from their closest neighbors can chain together to produce a combined crossing greater than 100 km across.)

For the OCO-3 self-crossing tests (Sect. 3.3), the approach is nearly identical, simply using OCO-3 soundings in place of the OCO-2 soundings. To avoid the algorithm identifying an entire orbit as a self crossing, we add a condition to the first step that two OCO-3 soundings are considered a match only if the time difference is ≤ 12 h *and* ≥ 0.77 h. The latter value is approximately half of the orbital period of the ISS, and was chosen to ensure that two OCO-3 soundings would contribute to a self-crossing only if they were in different orbits.

2.5.2 Determining sampling times for the OCO-2/3 simulation

For the OCO-2/3 crossing simulation, we take the output of the algorithm described in Sect. 2.5.1 and find all crossings within 10 degrees of a given TCCON site from the years 2021 and 2022. For consistency across seasons, we want to know how the satellites observe relative to solar noon rather than to UTC times, so for each date of crossing we calculate the time of solar noon from latitude, longitude and date. We then fit a Burr probability distribution to the time that OCO-2 observes the site with respect to solar noon, and a kernel density estimate (KDE) to the time difference between the OCO-3

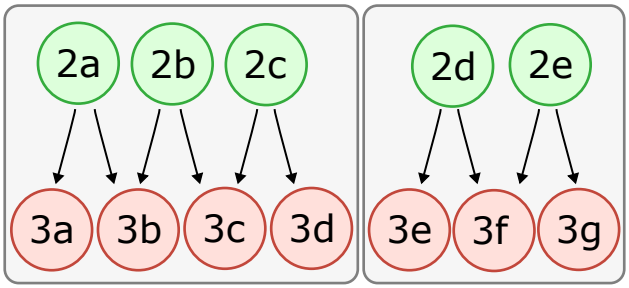

Figure 2. An illustration of how OCO-2 and -3 soundings are grouped into crossings. The circles labeled with "2" represent OCO-2 soundings, those with "3" represent OCO-3 soundings (the letters simply distinguish different soundings). The black arrows indicate which soundings fall within the 100 km and 12 h criteria to be considered paired. The grey boxes represent a crossing, that is all soundings in one box would be grouped together. Each OCO-2 sounding is shown matched to two OCO-3 soundings for simplicity; in practice, the number of OCO-3 soundings matched to an OCO-2 sounding will vary significantly.

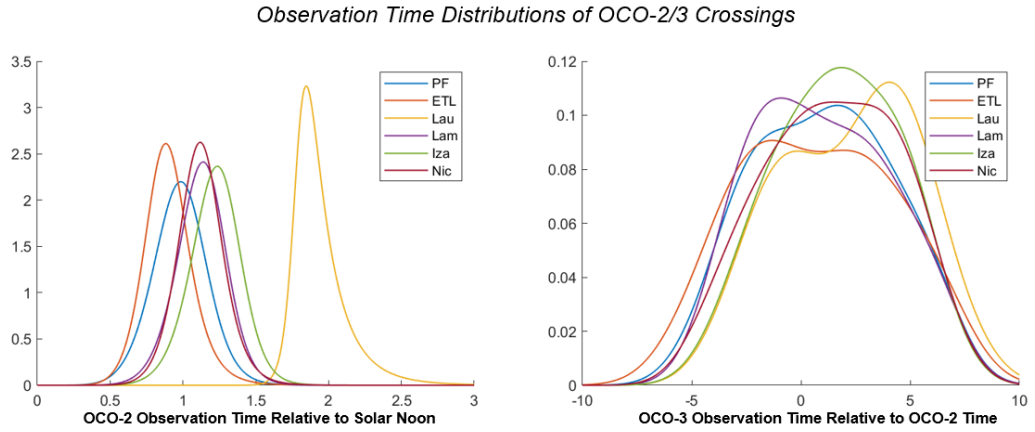

Figure 3. Probability distributions showing when the OCO-2/3 satellites observe a TCCON site on the same day; x-axis units are hours. The OCO-2 distributions show when OCO-2 observes the site with respect to solar noon, and the OCO-3 distribution shows when OCO-3 observes the site with respect to OCO-2. With the exception of Lauder, OCO-2 tends to pass over a TCCON site about an hour after solar noon. OCO-3 has a wide range of times that it observes the sites at.

observation and the OCO-2 observation. A KDE is fit to the time difference because it allows for negative values, and OCO-3 can observe a location before or after OCO-2 which tends to pass slightly after solar noon. The probability distributions defining the times the satellites observe each TCCON site are shown in Fig. 3. Combined, the distribution of OCO-2 and OCO-3 times in the crossings from the six TCCON sites covers the global distribution of OCO-2/3 relative crossing times (Fig. S7), thus, we consider these six sites a sufficiently broad test set to evaluate the general applicability of this method to OCO-2 and OCO-3 data.

For each TCCON day, we simulate an OCO-2/3 crossing by independently randomly sampling from our two probability distributions. (Random sampling is appropriate, as the times are not correlated, as shown in Fig. S6.) We find the time that OCO-2 observes

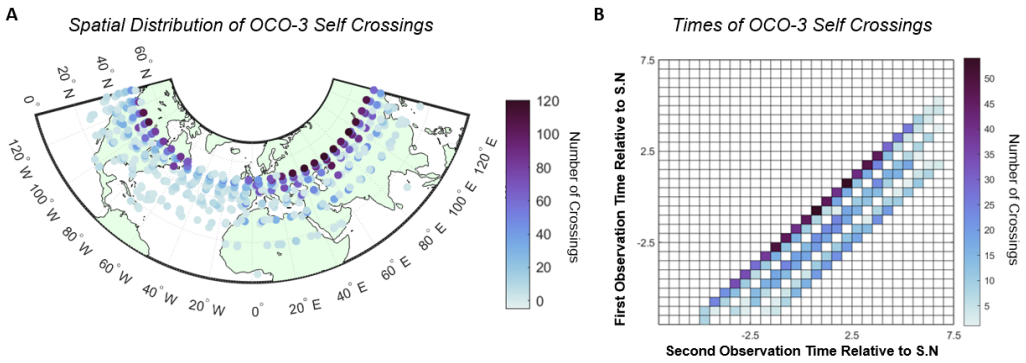

Figure 4. Locations and crossing patterns of OCO-3 self crossings. OCO-3 self crossings mainly happen around 50°N, and the second observation happens a multiple of 90 min after the first. In panel (B), "S.N." stands for "solar noon".

the site by adding the time of solar noon to the value drawn from the OCO-2 distribution, and the time that OCO-3 observes the site by adding the time of OCO-2 to the value drawn from the difference distribution. For each time, we average the XCO_2 values from within half an hour of the selected times. We follow the same process for subsampling the other TCCON fields.

We note that the OCO-3 time distribution used here incorrectly handled cases where multiple OCO-3 orbits crossed the OCO-2 orbit. Fortunately, these were a small minority of the data used to determine this distribution, so we do not believe it affects the results. Details of that analysis are given in the supplemental text (S2).

2.5.3 Determining sampling times for the OCO-3 self-crossing simulation

As in Sect. 2.5.2, we first need to construct a distribution of times of day observed in OCO-3 self crossing. Looking at all self crossings from 2020 and 2021, we find OCO-3 tends to cross itself around 50°N (Fig. 4a), due to the inclination of the ISS. To best approximate the diurnal cycles observable with OCO-3 self crossings, we only use data from ETL and PF, which are slightly north and south, respectively, of the crossing range, to construct these simulations. To identify the times of day that the satellite crosses, we filtered to the band of latitudes between 45°N and 55°N, and to the north American continent between longitudes 55°W and 125°W. This differs from the approach we used in Sect. 2.5.2 of using crossings at the specific TCCON locations to determine the times of day at which we sample; because ETL and PF are slightly outside of the self crossing region, we would not be able to identify sufficient instances of OCO-3 self crossings at those locations to construct a robust distribution of observation times. The other difference in methodology from the OCO-2/3 crossings is that when adding simulated error, we sample from the OCO-3 error distribution for both points.

In Fig. 4b, we see that there are three distinct "bands" in the sampling times. This is because, while the first observation may happen at any time of day, the second observation happens on a successive orbit and the ISS completes an orbit every ∼90 min. Following these results, we fit a KDE to the time of first observation, but for spacing between observations we select between 90 min, 180 min, and 270 min at their relative frequencies from the self crossing data.

2.5.4 Adding Simulated Error

The OCO-2/3 satellites have an uncertainty associated with each XCO_2 sounding. This error is comprised of systematic error that is a property of the instrument and retrieval used, and random error which will diminish as we average multiple soundings together for a single crossing. To best replicate the satellites' behavior in our models, we added a simulated uncertainty to the sampled TCCON data from the validation site withheld from training, keeping the model trained on 'true' TCCON data. To find how to scale the random error component, we found all crossings within 5°×4° of the site, as this is a common grid box size in global models and we are interested in whether this approach could provide additional constraints on modeled carbon fluxes. We then found the average number of soundings per crossing over each site. Finally, we calculate the standard error for a crossing over area j following:

$$s_{\overline{x}} = \sqrt{\sigma_\alpha^2 + \frac{\sigma_\gamma^2}{n_{\text{eff},j}}} \tag{2}$$

where σ_α is systematic area-wide uncertainty, which we set to 0.8 ppm, σ_γ is individual retrieval uncertainty which we set to 1 ppm, and n_{eff} is the effective sample size that accounts for potential spatial correlation in the single sounding errors. This functional form and the values for σ_α and σ_γ are derived from ongoing work to quantify the uncertainty associated with OCO-2 and -3 data (Taylor et al., 2023; Braverman et al., 2021; Kulawik et al., 2016). Since the OCO-2/3 crossings cover small areas, a modest positive correlation is typical, so we use the adjustment

$$n_{\text{eff},j} = 0.5 \cdot n_j \tag{3}$$

where n_j is the average number of soundings per crossing. For each subsampled point, we sample from a normal probability distribution with a mean of 0 and a standard deviation given by the calculated standard error. The factor of 0.5 for n_{eff} was chosen based on typical correlation lengths found in OCO-2 data (e.g., order 10 km, Sect. 3.3.4 of Mitchell et al., 2023). Work done by the OCO-2 uncertainty quantification working group found that this factor of 0.5 reduction is typical for OCO-2 data. Even in areas with large spatial correlation, the value of n_{eff} rarely falls below $n/3$.

Equation 2 assumes no systematic error cancels out. However with OCO-2 and OCO-3 being very similar instruments and using the same retrieval algorithm, we would expect the systematic error between the two satellites' observations of the same location on the same day to likewise be similar. As we use the difference between the two satellite's observations as our feature, we would expect some of the systematic error to cancel out in subtraction. We do not have a concrete estimate for how much systematic error actually cancels, so we bound the real error by the pessimistic scenario where none cancels, and the optimistic scenario where all the systematic error cancels. Note that for these calculations, n is the number of *soundings*, not crossings, and there are at least several hundred soundings per crossing for all sites. We run two simulations for adding error, one pessimistic case where σ_α stays at 0.8 ppm and one optimistic case where we set σ_α to 0 ppm. Table 3 shows standard error for the optimistic scenario where systematic error cancels. In the pessimistic scenario, systematic error dominates and every site and satellite has a standard error of 0.8 ppm. While the optimistic scenario of all systematic error canceling is likely overly optimistic, we believe that these two error scenarios provide useful bounds on the error for our analysis to understand the impact of error on the ML accuracy.

2.6 Sampling GEOS FP-IT for inclusion in the feature set

Some potentially useful information was not available from the TCCON data files; for these data, we sampled the GEOS FP-IT model at each TCCON location. Specifically, we used the inverse distance weighed average of the GEOS grid cells within 0.65°

Table 3. Standard error values by TCCON site and satellite for optimistic simulation where systematic error cancels and the average number of soundings per crossing (n_j, rounded to the neared hundred) used in Eq. (3) when calculating the standard error.

Site	Standard Error for an Ensemble Measurement		n_j	
	OCO-2	OCO-3	OCO-2	OCO-3
PF	0.07	0.03	400	2000
ETL	0.09	0.05	300	1000
Lamont	0.05	0.03	700	3200
Lauder	0.05	0.09	800	300
Nicosia	0.06	0.04	600	1600
Izana	0.07	0.05	400	800

of the TCCON latitude and longitude. GEOS FP-IT outputs our variables of interest every three hours. To ensure a consistent feature set of each variable, we interpolate the spatially-averaged GEOS data from its original 3-hourly time step to the 6 hours centered on solar noon. These 7 data points become input features in the model.

We included three variables from GEOS FP-IT in our feature set. The first variable from GEOS FP-IT is near-surface vapor pressure deficit (VPD), which is the difference between actual vapor pressure and saturation vapor pressure, and it relates to rates of evaporation. We calculate VPD from temperature, pressure, and specific humidity from the 2d files:

$$\mathrm{VPD} = v_{p_{\mathrm{sat}}} - v_{p_{\mathrm{air}}} \tag{4}$$

$$v_{p_{\mathrm{sat}}} = 6.112 \cdot \exp\left(\frac{17.67 \cdot T}{T + 243.5}\right) \tag{5}$$

$$v_{p_{\mathrm{air}}} = p\left(\frac{r}{r + \epsilon}\right) \tag{6}$$

$$r = \frac{q}{1 - q} \tag{7}$$

where T is temperature (°C), q is the specific humidity, p is pressure (Pa), $\epsilon = 0.622$ is the molecular weight ratio of vapor to dry air, $v_{p_{\mathrm{sat}}}$ is saturation vapor pressure, and $v_{p_{\mathrm{air}}}$ is the atmosphere vapor pressure (Bolton, 1980; Salby, 1995; Wallace & Hobbs, 2006).

The second and third variables relate to the change in potential temperature at 700 mb over the course of a day, because large changes can indicate moving fronts with different CO_2 concentrations. To find potential temperature at 700 mb, we linearly interpolate across the vertical pressure levels, and then find the temperature from the corresponding indices. GEOS is set at fixed UTC time intervals. We took the difference and absolute value of the difference in temperature between the first and last of our 7 averaged and time-interpolated values as two input variables to the model.

An alternative approach would be to use the change in potential temperature at 700 mb as a flag to remove days where the change in XCO_2 was dominated by synoptic motion of air, rather than (or in addition to) photosynthetic activity. We elected to include this as a feature, rather than a preprocessing flag, because it was not clear what the threshold should be for a change in 700 mb potential temperature to flag out a day. Instead, we chose to incorporate it as a feature to test if it would be a predictor of different patterns in the daily change of XCO_2.

2.7 The full model features set

Our input feature set contains our subsampled TCCON variables, the two differences in potential temperature at 700 mb from GEOS FP-IT, and the 7 interpolated VPD values calculated from GEOS FP-IT. For the TCCON variables, we have three versions of each variable:

- Subsampled: To create this set, we average TCCON data within half an hour of our simulated satellite times. For the OCO-2/3 crossing simulation, this means sampling times from our created probability distributions and averaging the variable values from within half an hour of the sampled times. The OCO-2/3 subsampling is described in detail above in Section 2.5.2.
- Δ: For many TCCON variables we include the 'Δ [variable]', which is the difference between observations divided by the time difference. For the OCO-2/3 crossing simulation this is the difference between the OCO-3 subsampled observation and the OCO-2 subsampled observation divided by the time difference between the two observations. For simulations with more than two observations, we calculate the Δ between every unique pair of subsampled points.
- Mean: this is calculated as the mean of the "subsampled" values (the first bullet of this list) for all of the sampling patterns we test.

Along with XCO_2, we include a number of other variables from TCCON. The rationale for including each is as follows:

- Hour and time before solar noon provide information about the temporal position of the other data. The number of hours between observations should inform the model about how far apart in time the XCO_2 observations are, and, combined with time before solar noon for the first observation, how they map to the EOFs.
- Solar zenith angle, solar azimuth angle, and airmass give information on time of day and time of year. Further, these should provide the model a metric for the amount of sunlight available (on average, ignoring clouds). Airmass and solar zenith angle are related, but because airmass is $\approx 1/\cos(\mathrm{SZA})$, their functional forms are different, and one may be more useful for the model than the other.
- Surface pressure, surface temperature, and wind speed give meteorological information about processes (such as synoptic scale transport) that could affect diurnal cycles. In principle, higher wind speeds might indicate stronger advection, while a change in surface pressure or temperature might indicate a passing weather front.
- Prior XCO_2 gives information about the long term trend in XCO_2, since the priors incorporate trends in CO_2 as measured at Mauna Loa and American Samoa (Laughner et al., 2023). Including the priors was intended to give the model information about the expected approximate mean XCO_2 for a given day.
- VPD should provide information about whether the local plants are more or less likely to be water stressed. This, in turn, would affect their productivity and the magnitude of the diurnal cycles.
- Potential temperature at 700 mb change and its absolute value are additional indicators of synoptic scale motion that drives changes in XCO_2 not directly related to the local biosphere. Keppel-Aleks et al. (2012) showed that potential temperature at 700 mb correlated with free tropospheric CO_2 during the INTEX-NA campaign.

For XCO_2, we only include ΔXCO_2 in our feature set. We do not keep the subsampled values themselves because we want to remove seasonal and interannual variability while keeping our model applicable to sparse satellites (like OCO) where XCO_2 data cannot be detrended by subtracting off the value at solar noon. For XCO_2 we also include the difference between observations without dividing by the time difference, for in-

Table 4. Input features used in our models along with which data source they are drawn from. The "Variable" column lists the variable name and, in parentheses, how it is referred to in the TCCON files, when applicable. The "Subsampled" column indicates the TCCON data averaged about the simulated satellite times. The "Δ" column indicates the difference between observations divided by the time difference (except for "hours", which is an unnormalized difference); for the OCO-2/3 crossing simulation this is the difference between the OCO-2 simulated observation and the OCO-3 simulated observation divided by the time difference between satellite overpasses. The "Mean" column indicates the average value over the "Subsampled" values. We use the "Other" column for variables that have a special calculation we describe in the text.

Data Source	Variable	Subsampled	Δ	Mean	Other
TCCON	XCO_2 (xco2)		x		x
	Hour (hour)		x		
	Solar Zenith (solzen)	x	x	x	
	Solar Azimuth (azim)	x	x	x	
	Surf. Pressure (pout)	x	x	x	
	Surf. Temperature (tout)	x	x	x	
	Wind Speed (wspd)	x	x	x	
	Airmass (airmass)	x	x	x	
	Prior XCO_2 (prior_xco2)	x		x	
	Time Before Solar Noon				x
GEOS FP-IT	VPD				x
	Potential Temp Change				x
	Abs Potential Temp Change				x

stances where small time differences make ΔXCO_2 large while the actual difference between points is small. Additionally, we calculate the time between the first observation and solar noon. The full list of model features are shown in Table 4. We will investigate how these feature relate to the predicted PCs in Sect. 5.1.

2.8 Features not included

There are several possible predictive metrics which we were not able to include due to the requirements of our experimental design. In particular, the normalized difference vegetation index (NDVI), leaf area index (LAI), solar induced fluorescence (SIF), soil moisture, and soil temperature could all provide useful information on the availability of light and water for photosyntheis. However, our approach requires:

1. that these data be available globally, as the end goal of this work is to produce a method that can be easily applied to space-based XCO_2 data without relying on local datasets, and
2. that these data be available over enough of the 2008 to 2023 period from which we take our TCCON data to ensure that we have sufficient training data for the machine learning model.

NDVI has been produced from the Moderate Resolution Imaging Spectrometer (MODIS) instruments on board the Terra and Aqua satellites for over 25 years. We tested including the 16-day NDVI and EVI (Enhanced Vegetation Index) MODIS-Aqua product (Didan, 2021) in our model, but saw little to no improvement over models without these data as inputs. (Figs. S8 and S9 compared to Fig. 19 and 21 show little to no improvement

in R^2 for the full models—i.e., the rightmost points in Figs. S8 and S9 have at most a ~ 0.05 greater R^2 values than the rightmost points in Figs. 19 and 21.) From these results, we decided not to include LAI as well, as we expected it to be unlikely to add more information compared to NDVI and EVI.

Global daily SIF has only been available since the launch of TROPOMI in late 2017. OCO-2 itself provides SIF for its full record back to 2014, but the OCO-2 swath is quite narrow. Gap filled products such as CSIF (Y. Zhang et al., 2018) and GOSIF (Li & Xiao, 2019) exist, but at coarser temporal resolution (4 and 8 days, respectively). This meant that we would not be able to capture the day-to-day variation in SIF to predict the productivity of plants on a given day. These products also use machine learning themselves to perform the gap filling. Because our goal is to develop a method that is primarily data-driven, we elected not to pursue SIF as an input parameter, as using these products would mean our machine learning itself was trained on machine learning output.

Finally, we investigated the possibility of using soil moisture from either the Soil Moisture Active and Passive (SMAP) or Soil Moisture and Ocean Salinity (SMOS) missions. The 36 km SMAP global product (O'Neill et al., 2023) was easiest to access, despite only covering the time period from April 2015 on. We found, however, that there was only data within our colocation criteria for 2 of the 6 TCCON sites used in this study (Lamont and Lauder). As using only these two sites would severely limit our training data, we did not pursue soil moisture further as an input feature.

2.9 Model Validation

We evaluate our models' performance using the error metrics R^2 and root mean squared error (RMSE). Higher R^2 values indicate that the model is accurately fitting the data trends: the model predicted XCO_2 increases as real XCO_2 increases. RMSE tells us the average distance between observed and reconstructed values, and represents how concentrated around the 1:1 line our data falls. For drawdown (calculated as described in Sect. 2.3), RMSE tells us on average how far off the predicted drawdown is from the drawdown calculated from the true TCCON diurnal cycle.

When training our model, we choose one of the six TCCON sites to withhold for validation and only train on the input features sampled at the other five TCCON sites. We then use a 70/30 train/test split of the data from the five training sites. That is, the model is trained on 70% of the data from these five sites, then evaluated on its ability to predict PCs for both the 30% of data withheld for testing from the five training sites *and* to predict PCs for the site from which no data was included in the model training. Withholding data from an entire site is an important test, as if this approach were to be applied to space-based XCO_2 data generally, it would have to generalize to locations which did not have a TCCON site (and therefore have no training data). When presenting results in the following sections, we refer to the model's performance on the training 70% of data from the five training sites as its "training performance," its performance on the withheld 30% of data from those same five sites as its "testing performance," and its performance on the entirely withheld site as its "extendibility performance." Additionally, the performance metrics we report are the result of a k-fold cross validation test, where we repeat the above training and evaluation repeatedy, using a different TCCON site as the fully withheld one each time. Only ETL, Park Falls, Lamont, and Lauder are used as the withheld site; Izaña and Nicosia are never the withheld site as they only have a small number of useful days (enough to add to the training data, but not enough to serve as a worthwhile validation test on its own).

In the k-fold cross validation tests, the withheld site is also kept out of EOF generation. This ensures that the k-fold cross validation results are truly indicative of the model's ability to generalize to unseen locations, without any prior knowledge about that location. However, the EOFs are generated from all of the data at the five training sites,

not just the 70% of those observations used for training. Therefore, the "testing performance" can be reasonably expected to be better than the "extendability performance," both because it represents a location for which training data was provided and because the underlying EOFs also represent the variability for the observations used in the testing performance results.

Because we are using TCCON XCO_2 values both as the simulated space-based observations and the true XCO_2 comprising our diurnal cycles, it is important to add in error to simulate the difference between what a satellite will observe and the true XCO_2 at that time. For simplicity, we only apply error to the simulated satellite XCO_2 in the extendibility tests. We use an upper and lower limit for the error as described in Sect. 2.5.4, and report the results as the "pessimistic extendibility" and "optimistic extendibility" cases, respectively.

3 Testing Applicability to OCO-2 and -3 Observations

We first study how applicable our model is to current space-based missions. The OCO-2 and OCO-3 satellites can cross and observe the same location in a single day, and OCO-3 can self cross and observe the same location twice in a single day. We test how well our model can reproduce diurnal cycles from TCCON data sampled at these simulated crossing times.

3.1 OCO-2/3 simulation

With our model output, we are interested in how well our model diurnal cycles of XCO_2 compare to TCCON cycles, as well as how well accurate our model's XCO_2 drawdown predictions are. We follow the model validation methods outlined in Sect. 2.9. Scatterplots showing the combined performance in our k-fold cross validation (i.e. the combined results from leaving each site out and training on the others) for both diurnal cycles and drawdown are shown in Fig. 5. We can see that for the testing and optimistic extendibility plots the model outputted diurnal cycles do not capture the full range of variability as seen by TCCON. TCCON diurnal cycles range from -2 to +2 ppm throughout the day, and the model outputted XCO_2 values are clustered around 0 ppm, capturing only the mean. For the pessimistic extendibility results where systematic error does not cancel the results have a wide spread about 0 ppm, meaning that the model is not capturing the mean of the data and is almost uncorrelated with TCCON data.

To understand the cause of this poor performance, we examined a sample of true and predicted diurnal cycles. Fig. 6 shows individual reconstructed diurnal cycles compared to the truth from TCCON. While there are some days that the model captures well, these results show the model is not accurately reproducing diurnal cycles. Sometimes, the model captures the right shape but not the correct magnitude of variation, but often the diurnal cycle is a completely different shape. Looking at both Fig. 5 and Fig. 6 we see that for the extendibility tests, the RMSE on drawdown is 0.48 and 0.38 ppm, which is higher than the actual drawdown values shown in Fig. 6, meaning the error is far overpowering the actual data. Additionally, in the pessimistic extendibility scenario where no systematic error cancels, the added error of 0.8 ppm is of similar magnitude to the drawdown signal of interest, and partially explains the model's poor performance; when the added error is larger than the magnitude of diurnal cycle variability the model is unable to accurately reproduce diurnal cycles.

3.2 Growing Season Simulation

We ran a second set of simulations that also used OCO-2 and -3 crossings, but focused only on capturing the growing season. We are interested in photosynthetic drawdown, but many days in winter have consistent XCO_2 diurnal cycles without significant

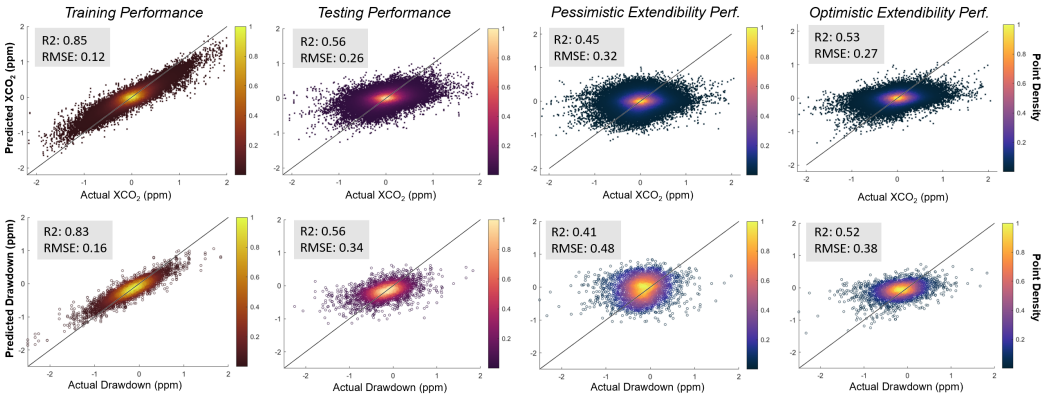

Figure 5. Scatterplots showing model performance under the OCO-2/3 Crossing Simulation. "Training performance" compares TCCON diurnal cycles and drawdown to model reconstructed diurnal cycles from data it was trained on. "Testing performance" compares TCCON diurnal cycles and drawdown to model reconstructed diurnal cycles and drawdown from data witheld from training, but from familiar sites. "Extendibility performance" simulates how the model would perform when estimating PCs from TCCON site fully withheld from training and EOF generation. Additionally "extendibility performance" cases have added simulated error; the pessimistic extendibility scenario is where systematic error between the satellites does not cancel, and the optimistic extendibility scenario is where systematic error between satellites completely cancels. Because we are interested in diurnal variation with seasonal and interannual signals removed, each XCO_2 point is the difference in value from solar noon.

drawdown. We hypothesized that those days might be limiting the possible PC combinations our model predicts, as most of our model outputs in the normal simulation capture less variation than seen by TCCON.

There are many definitions for the growing season, and we used the meteorological definition for trees, which is that the growing season starts when the daily mean temperature is above 5 °C for 5 days in a row, and the growing season ends when the daily mean temperature is first below 5 °C (Körner et al., 2023). Because TCCON has days when clouds limit the number of observations in a day, we also fit a sinusoidal function to the daily mean temperatures, and added the requirement that to begin the growing season, the sinusoid must have a positive slope, and to end the growing season sinusoid has a negative slope. Because TCCON only measures during the day, our daily average temperatures likely skew warmer than the actual daily mean temperatures and allow for more days than otherwise. However, the cutoffs can be 0 degrees or 5 degrees, so by making the requirements be the upper limit, we do our best at adhering to the definition. Filtering by growing season reduces our training set from 3806 total days to 2892.

The results from this simulation are shown in Fig. 7. Model performance from the growing season simulation is not noticeably different compared to the first simulation. However, we only see a 25% decrease in days when filtering by growing season, when we would expect the set to be halved. We hypothesize that our TCCON data preparation that requires days to have a certain number of hours represented inherently filters out more non-growing days than growing days; thus, our simulations in Sect. 3.1 are already focused on the growing season.

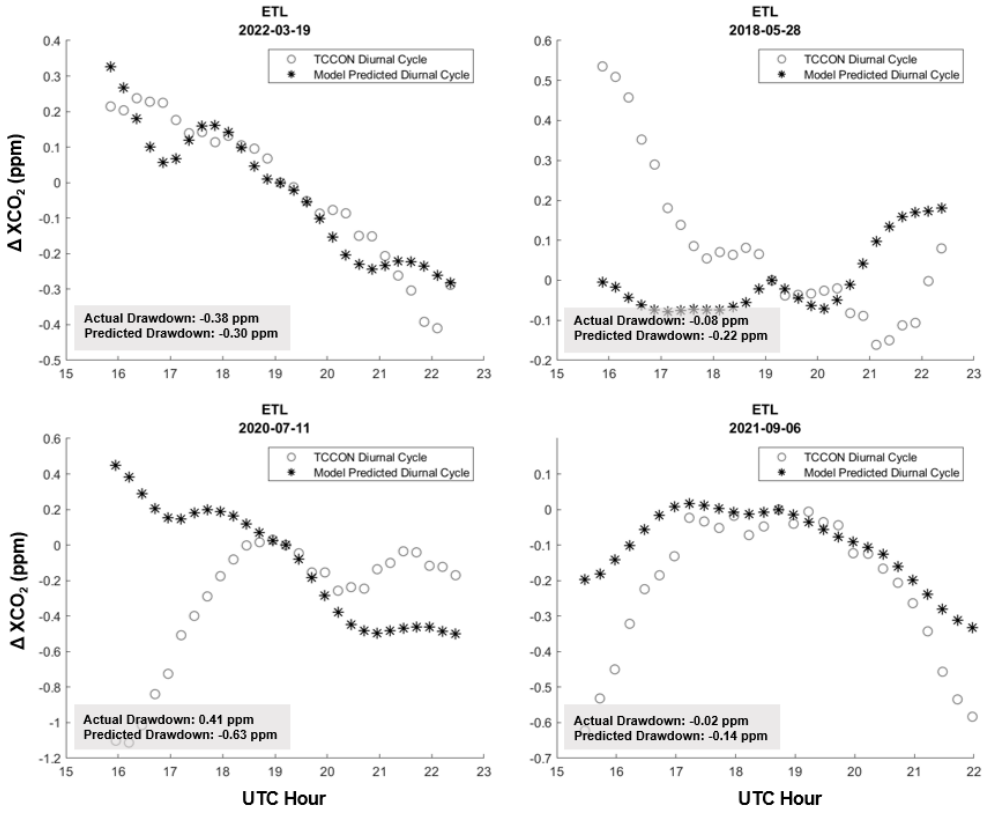

Figure 6. Model reconstructed diurnal cycles compared to the true diurnal cycles from TCCON. These are diurnal cycles from the pessimistic extendibility tests, meaning that in this case, ETL was left out of model training, and simulated error was added in under the scenario that systematic error between the two satellites does not cancel.

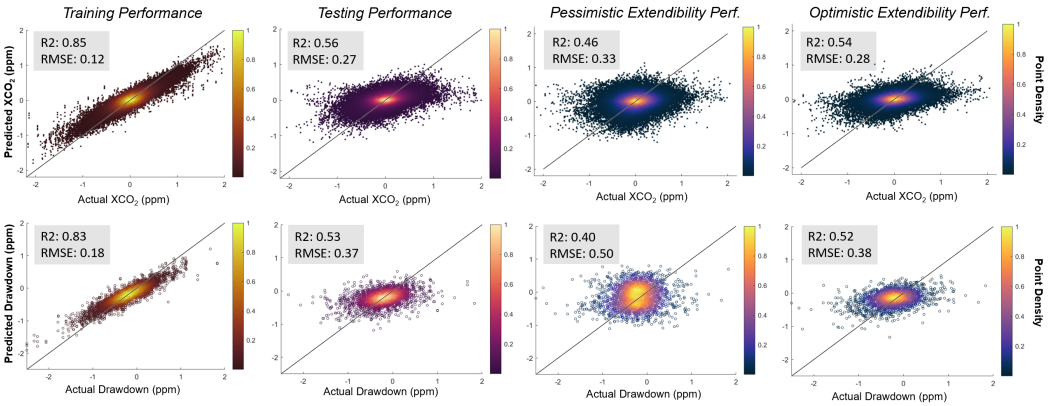

Figure 7. Scatterplots showing model performance for the Growing Season OCO-2/3 crossings simulation. For a detailed description of the columns and rows see Fig. 5.

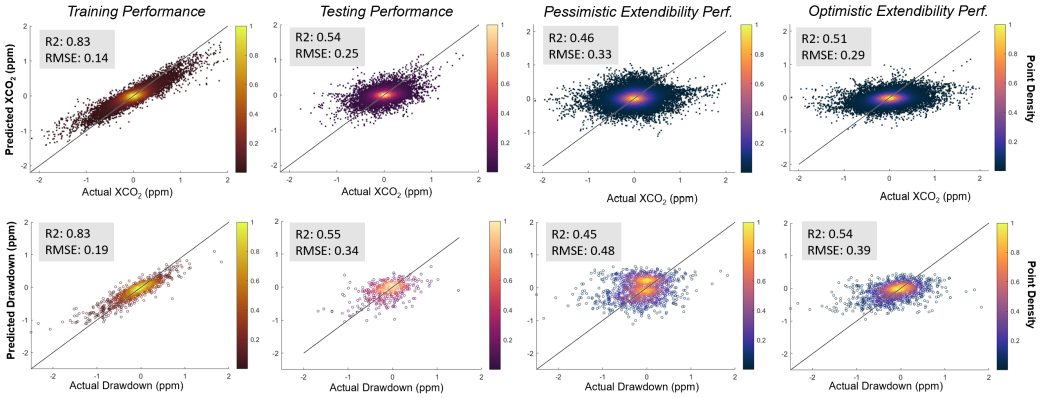

Figure 8. Scatterplots showing model performance for the OCO-3 self crossing simulation. This simulation only uses TCCON data from Park Falls and East Trout Lake to train and test the model on, as OCO-3 self crossings happen around 50°N. For a detailed description of the columns and rows see Fig. 5.

3.3 OCO-3 Self Crossing Simulation

Our third simulation utilizes just OCO-3. OCO-3's ground track can cross and observe the same location in a single day, so we ran the model subsampling TCCON at a simulated OCO-3 self crossing pattern, with self crossings identified as described in Sect. 2.5.1. For this simulation, we assume only two OCO-3 orbits cross, for a fairer comparison to Sects. 3.1 and 3.2. Crossings with multiple OCO-3 orbits will be discussed in Sect. 4.6.

The results from this simulation are shown in Fig. 8. These scatterplots show combined model predictions from a simulation that trains on ETL and uses PF as the validation site, and a simulation that trains on PF and uses ETL as the validation site. This greatly cuts down the number of training days for the model. Despite the reduced amount of training data compared to our OCO-2/3 and growing season simulations, which should inhibit model performance, the self crossing simulation performs similarly to the OCO-2/3 simulation. The training performance is marginally worse, which is likely due to the reduction in training data. The remaining results are not significantly different from the original simulation.

From these results, we conclude that two observations per day are insufficient to adequately constrain the diurnal cycles or drawdown with this method. It is possible that this could be improved with training data from more locations, giving the ML model a better foundation to generalize from. However, Figs. 5, 7, and 8 show similar performance between the "testing" and "optimistic extendibility" performance. The primary difference between these is that "testing" tries to predict diurnal cycles or drawdown for unseen data from TCCON sites that contributed training data, while both "extendibility" tests evaluate its ability to predict a site that was entirely withheld from training, as described in Sect. 2.9. That "testing" and "optimistic extendibility" have similar performance suggests that trying to generalize to a new location is not a major source of error in this approach.

4 Testing > 2 observations per day: application to present and future missions

Given the relatively poor ability of this approach to constrain diurnal cycles from crossings with only two times of day (one from OCO-2 and one from OCO-3, or two from OCO-3) data, our next step was to investigate how many additional observations per day would be required to do so reliably. We first investigate the correlation between times of observation and the information the model is attempting to recover. We next test the potential performance of our ML approach in a best-case scenario. Third, we consider three factors that impact the achievable accuracy of the diurnal cycles and drawdown values inferred by this method for consideration in future XCO_2 observing missions. Finally, we synthesize the results of these experiments to understand what utility might be gained from cases where multiple OCO-3 orbits intersect one OCO-2 orbit.

4.1 Correlation Between Sampling Times and Diurnal Information

After assessing the model on current space based satellite patterns, we evaluated how different sampling patterns would change model performance. To get a preliminary idea of where ideal sampling patterns were, we wanted to see how the times of day XCO_2 is observed correlate with drawdown and diurnal cycle reproduction. We plot correlations of ΔXCO_2 calculated between different first and second observation times with XCO_2 drawdown and the first two PCs. The correlation plots are shown in Fig. 9. Because drawdown is defined as the four hour difference in XCO_2 centered around solar noon, when drawdown is calculated as the difference between XCO_2 at 2 hours after solar noon and 2 hours before solar noon the correlation is 1. However, rather than uniformly getting weaker, sampling patterns with first observation 2 hours before solar noon or the second observation 2 hours after solar noon have strong correlations. These results indicate that in order to directly give the model information about drawdown, at least one of the subsampled points needs to be near the locations from which drawdown is defined.

For seeing how observation times correlate with diurnal cycles we look at correlations between the observed XCO_2 difference between two times of day and the first two PCs, because the first two EOFs explain 81% of diurnal cycle variance. In Fig. 10, the PCs shown are those output by the EOF decomposition of the TCCON data, i.e., the "true" PCs rather than those output by the ML. PC1 is the coefficient multiplying EOF1 in order to best represent each day; EOF1 describes XCO_2 throughout the day so it aligns with our expectations that in Fig. 10 the first PC correlates with ΔXCO_2 when it is calculated as the difference between any morning and afternoon observation. PC2 is likewise the coefficient multiplying EOF2, a function that describes XCO_2 change in the afternoon with a flat morning. Following the PC1 results, PC2 correlates with ΔXCO_2 when it is calculated as the difference between an afternoon observation and a solar noon observation. From these correlations we see that with only two observations, it is not possible to have a distribution that correlates strongly with both PC1 and PC2. The best option is to have observations 2 to 3 hours before and after noon (possible with OCO-3 self crossings), which shows a modest correlation to both EOFs. Capturing both PCs is critical to accurately reconstruct diurnal cycles, as PC1 has an explained variance of 0.46 and PC2 has an explained variance of 0.36; both are needed in order to accurately reconstruct a day but to capture both three observations would be needed.

These correlation maps also serve to validate our model's performance in the OCO-2/3 simulation and the OCO-3 self crossing simulation. We have illustrated the times observed by OCO-2 and -3 within Fig. 9; the probability distributions of each satellites' observation times are shown in the small axes beside the drawdown panel, the central 1 standard deviation of times observed by OCO-2 are highlighted in white on all panels, and the OCO-3 self crossing times (from Fig. 4) are overlaid in hatching on the correlation maps. With these times overlaid, we find the OCO-2/OCO-3 crossings cannot

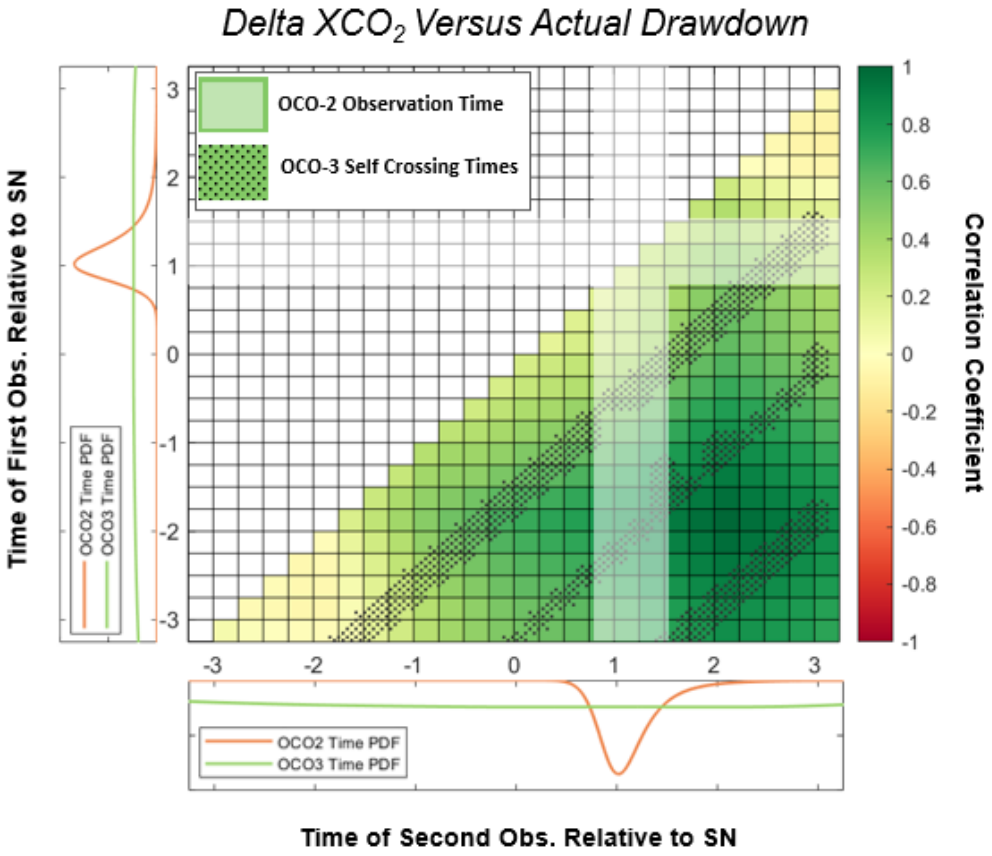

Figure 9. A plot showing the correlations between ΔXCO_2, calculated as the difference of observations at varying times of day, and XCO_2 drawdown (which we are interested in because it is the variable used to calculate NEE). Overlaid on the correlations are the observation times of OCO-2 and -3. Probability distributions for OCO-2 and -3 over all TCCON sites are in the small axes and show the likelihood of OCO-2 and OCO-3 observing at the axis times. The central one standard deviation of times observed by OCO-2 are highlighted in white on all panels, and the OCO-3 self crossing times are overlaid in hatching on the correlation maps. "SN" in the axes labels stands for "solar noon."

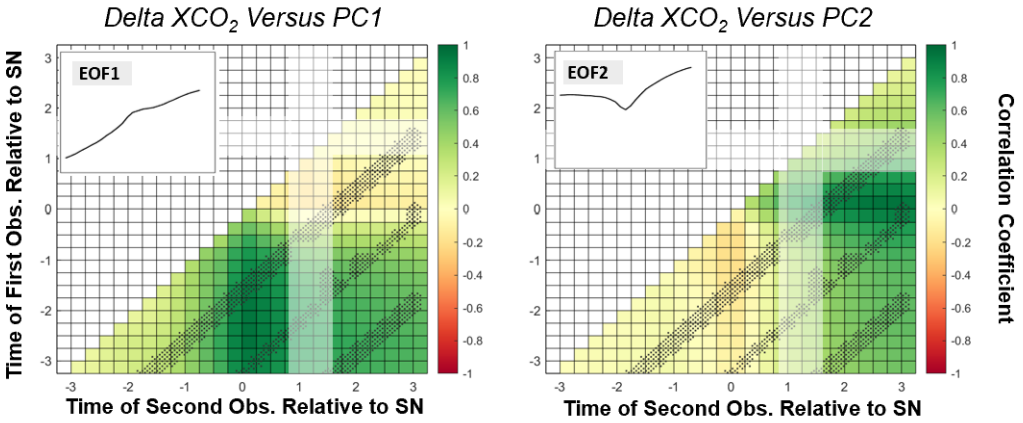

Figure 10. Plots showing how the correlations between ΔXCO_2 (calculated as the difference of observations at varying times of day) and the first two PCs. We are interested in correlations with PC1 and PC2 because they explain 81% of diurnal cycle variance. We also show the observation times of OCO-2 and -3. Following the probability distributions shown in Fig. 9, the central one standard deviation of times observed by OCO-2 are highlighted in white on all panels, and the OCO-3 self crossing times are overlaid in hatching on the correlation maps. "SN" in the axes labels stands for "solar noon."

observe at ideal times for giving the model information on diurnal cycle PCs or drawdown. OCO-3 can achieve the ideal sampling of 2 to 3 hours before and after solar noon that correlates strongly with drawdown (Fig. 9). However, this sampling depends on the orbital geometry, and so it is likely occurs somewhat infrequently. As (Keppel-Aleks et al., 2012) showed that it was necessary to average between one week and one month of data to correlate the drawdown with fluxes measured by eddy covariance towers (due to the influence of transport on the upper part of the column), these ideal self crossings are likely too infrequent to provide robust constrains on the NEE.

4.2 Upper Limit of ML Performance

We next test if ML can reproduce diurnal cycles when given sufficient information, so we set an upper limit for our model's capabilities by training our model on full diurnal cycles, without subsampling to satellite observation times. For this simulation we put all of our TCCON variables into the same quarter hour intervals as XCO_2 following the same averaging process as in Sect. 2.4. Instead of a single ΔXCO_2 feature, the model now receives as input the differences between all possible pairs of the 27 quarter-hour-average points for each variable taken from TCCON (676 inputs per Δvariable). The features from GEOS FP-IT are unchanged, and we remove the difference between the first observation and solar noon as a feature variable, because the first input for all days is 3.25 hours before solar noon. The results are shown in Fig. 11.

It is interesting to note that the R^2 for the training performance is 0.95, which is the same as the explained variance of our EOFs – with our method of EOF decomposition we can only capture 95% of diurnal cycle variation. The extendibility performance test indicates that following our methods, the optimum performance for reconstructing diurnal cycles is with an R^2 of 0.92 and RMSE of 0.09 ppm, and drawdown has a maximum R^2 of 0.93 and RMSE of 0.11 ppm.

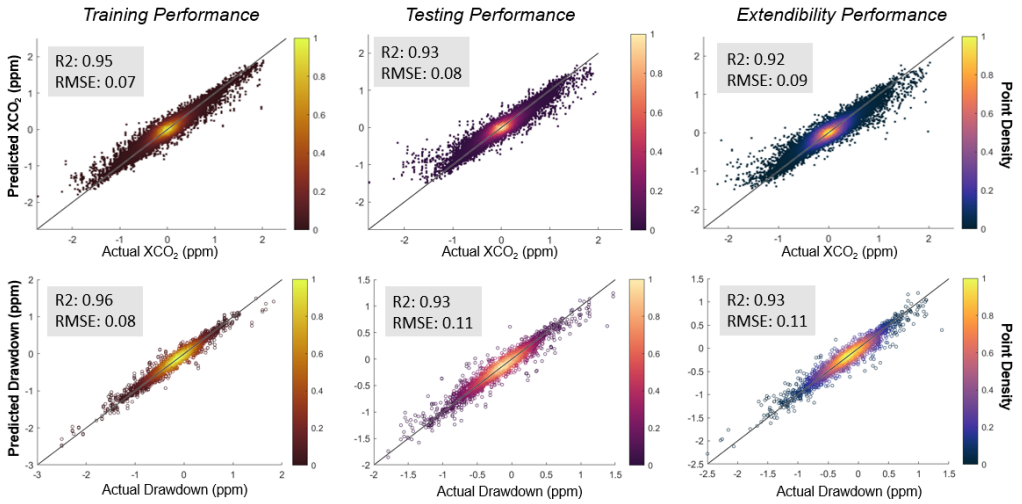

Figure 11. Scatterplots showing ML performance at reproducing diurnal cycles when TCCON data throughout the day (not only at satellite observing times) are given as inputs. For a detailed description of the columns and rows, see Fig. 5. Differing from Fig. 5, there is only one extendibility result because we do not add in simulated error for this simulation.

4.3 Optimizing Start Time and Spacing

After the 27 point run showed ML can reproduce diurnal cycles when given enough information, we wanted to find the minimum requirements to maintain good outputs. We began by investigating what sample times would optimize our model, for both two observations per day and three observations per day. To find the optimum sample times, we iterated over different start times and spacings to subsample our TCCON data at, and evaluated model performance for each combination. To account for randomness in ML, we ran the model three times for each configuration, and recorded the average R^2 values across runs. We ran these iterations for two subsampled points per day and three subsampled points per day. For three subsampled points we kept even spacings between points in order to restrict the problem space to two dimensions (i.e., start time and time between successive observations). This test and the following sensitivity tests are done with East Trout Lake as the validation site. Tables S1- S8 in the Supplement show that there is not a large performance bias across the sites, so for computational purposes we did not cross validate.

The results of these tests are shown below in Fig. 12 and Table 5. Note that the simulations in this section do not add error to the simulated observations, as the goal is to test the impact of the observations' temporal pattern and error characteristics separately.

For three points there is a clear trend in good diurnal cycle R^2: the model performs well when there is a point at solar noon and two points evenly spaced on either side. There is a drop-off in correlation for start times earlier than 3 hours before solar noon, which makes sense because our diurnal cycles begin only 3.25 hours before solar noon. Drawdown R^2 has a weaker and less uniform trend than diurnal cycle R^2. There is an arm of higher R^2 along the start time of 2 hours before solar noon, which is one of the points that defines drawdown, as well along the path of high diurnal cycle R^2. However, for drawdown there is a clear maximum when there is a point at solar noon, and two others 2 hours on either side.

Table 5. ETL Optimal Sample Times by Number of Observations and Optimized Parameter

Num Points	Parameter Optimized	Start Time	Spacing
3	Diurnal Cycle R^2	-2.75	2.75
	Drawdown R^2	-2	2
2	Diurnal Cycle R^2	-2.5	2.75
	Drawdown R^2	-2	4

In the two observation simulation, the highest R^2 for drawdown is at 2 hours on either side of solar noon, similar to the three observation test. In contrast, diurnal cycle R^2 does not appear to have a trend. There is a maximum with one point 2.5 hours before solar noon and the second 0.25 hours after, but the range across all the combinations is only 0.06, and this maximum is likely insignificant. These results agree with the correlation plots in Figures 9 and 10 that suggest two observations are not sufficient for ML to reproduce diurnal cycles.

To see how performance improves as points are added, we ran an additional set of simulations for four evenly spaced observations, which are shown in Fig. 13. The diurnal cycle results are very similar to the 3 point tests, where there is a band of good observation times. These results are also consistent with the other tests in that ideal diurnal cycles are where there are three points, one at solar noon, one in the morning and one in the afternoon, but with the four point simulation there is an additional point in the early morning. Drawdown results are optimized when there are points near the drawdown definition times, but with four points the band of good performance matches the diurnal cycle band. However, while diurnal cycle performance dramatically increases from two to three points, we only see marginal improvement in the four point simulation: with two points the optimized R^2 is around 0.55, with three points around 0.75, and with four points around 0.80. This also follows from the results seen in Fig 10, which show that the first two PCs are well captured by having three points. As those two PCs describe the majority of diurnal cycle shape, adding another point is not very beneficial with this method.

4.4 Testing Randomness in Observation Times

As shown in Figs. 3 and 4, the OCO satellites do not cross a location at a fixed time each crossing; rather, there is a range of times they sample from that we fit probability distributions to. However, in Sect. 4.3, we assumed that the temporal spacing between successive observations in a day will be consistent. In order to evaluate how ML could work with future space-missions, we assess how sensitive model performance is to this variability in observation time. For the same reason as in Sect. 4.3, no error is added to the simulated observations in this section.

To test this sensitivity, we start with the four optimal sample times listed in Table 5 and run the model with increasing spreads in observation start time, and in spacing between observations. For each start time and spacing, we varied the standard deviation across the sample times from 0 to 3 hr, with a standard deviation of 0 hr representing an observation system that always crosses at the same time of day per location, and a standard deviation of 3 hr representing crossings that happen at random times during the day without consistency. We varied the spreads in start time and spacing independently, looping over different combinations of standard deviation for both parameters.

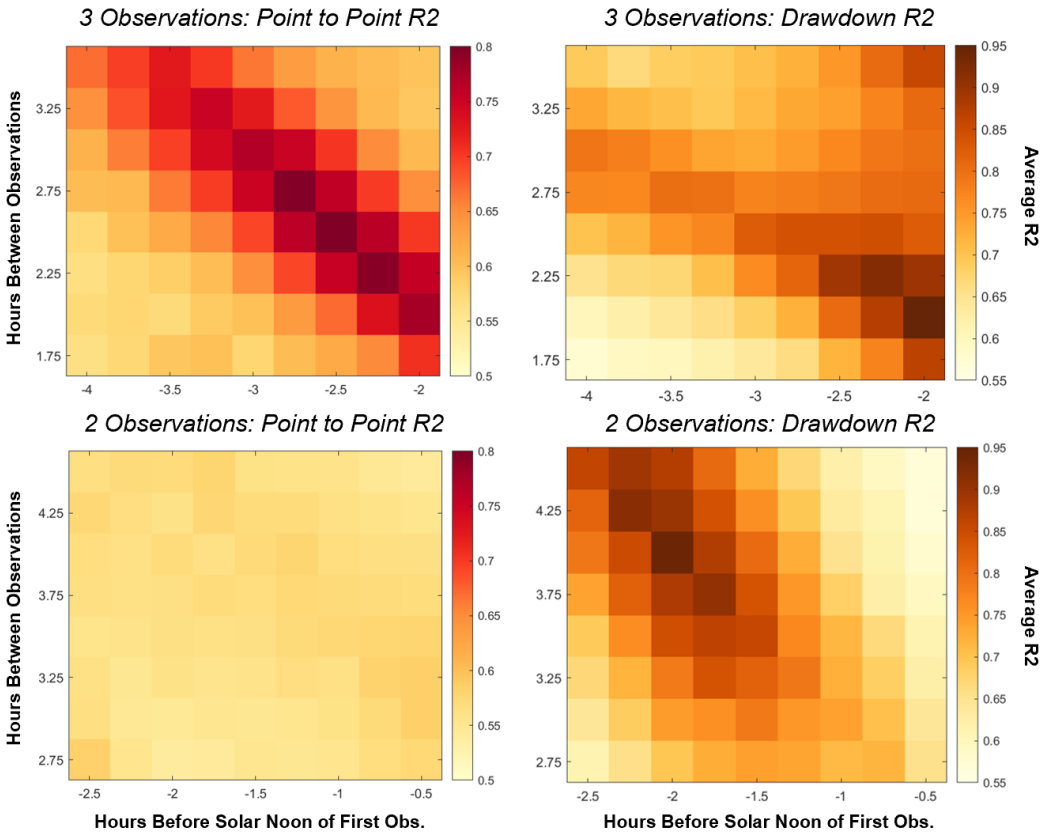

Figure 12. Plots showing model performance for different times of observation, for both 2 and 3 points per day. The color represents the average R^2 across an ensemble of 3 ML models for the given start time on the x-axis and temporal spacing between observations on the y-axis. The left two panels show the models' ability to predict the diurnal cycles as a correlation of individual XCO_2 points and the right panels show the models' ability to predict the daily XCO_2 drawdown.

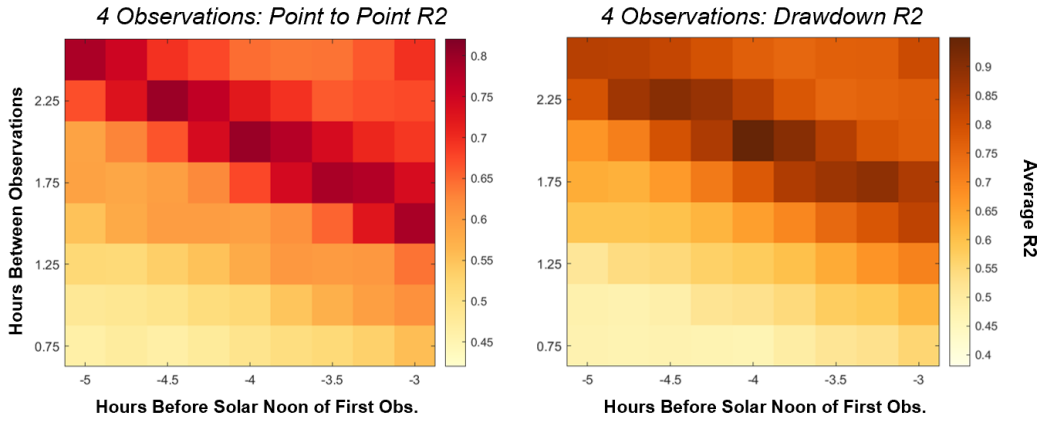

Figure 13. Plots showing model performance for different times of observation, for 4 points per day. The color represents the average R^2 across an ensemble of 3 ML models for the given start time on the x-axis and temporal spacing between observations on the y-axis. The left panel shows the models' ability to predict the diurnal cycles as a correlation of individual XCO_2 points and the right panel shows the models' ability to predict the daily XCO_2 drawdown.

Figure 14 shows the results for each set of observation times. For both 3 point simulations (diurnal cycle and drawdown, panels a and b), performance remains roughly constant until a standard deviation of 0.2 hr for both start time and spacing, before it begins to uniformly drop off with increasing variability in both parameters. For optimizing diurnal cycles, the three point simulation (panel a) always performs the same as or better than the 2 point simulation (panel c), regardless of randomness. At maximum variability the 3 point simulation has an R^2 of around 0.5 for both the diurnal cycle and drawdown. Conversely, no matter how much variability is present in the 2 point simulation, performance is never above 0.55 for either the diurnal cycle or drawdown. This agrees with the plots shown in Figs. 5 and 10 that suggest 2 points are not sufficient to capture diurnal cycles. For the 2 point configuration that optimizes drawdown (panel d), there is also a range between 0 and 0.2 hour standard deviation in which performance is near constant, with a slight drop with increased start time deviation. However, rather than dropping off uniformly, this plot shows R^2 falling below 0.75 when the std. dev. in start time is $\geq$ 0.5 hr or the std. dev. in spacing is $\geq$ 1-1.5 hr. Therefore, for capturing drawdown, variability in spacing is slightly less detrimental than standard deviation in start time. This agrees with our understanding of drawdown performance with 2 points: because the model is not reproducing accurate diurnal cycles, in order to capture drawdown the observations need to be near the locations where drawdown is defined. If the start time is fixed around -2 hr, but the spacing between points is varied, one of the points that defines drawdown is still accounted for. However, if the start time is allowed to vary, no matter if the spacing is fixed it is likely neither observation will be at the ideal time of day.

4.5 Testing the Effect of Error

In order to comprehensively study how well our ML approach could apply to future missions, we need to test how well our model could handle satellites with more or less standard error in their observations. Our third set of sensitivity tests involves varying the amount of simulated error added to our subsampled points. We took the four optimum observation configurations from Table 5 as well as the OCO-2/3 sampling pattern from Sect. 3.1 and tested how performance responded to increasing the simulated error added. For each configuration we ran the model iteratively, and at each iteration we increased the standard error about each subsampled point, with standard error ranging from 0.01 to 1 ppm. Following the other tests, to account for randomness in ML models, after adding standard error we ran the model three times and averaged statistics together.

The results from these tests are shown in Fig. 15. For all the simulations, standard error below 0.1 ppm did not substantially affect model performance for either the predicted diurnal cycles or drawdown values. Once standard error is above 0.1 ppm performance begins to drop, with the most dramatic declines being for the simulations that initially perform better. The OCO-2/3 simulation initially has the lowest R^2, but by 1 ppm of standard error it outperforms two of the other four models, and is nearly the same as the remaining two (again, for both the diurnal cycle and drawdown evaluation). This is explained in Fig. 16. In the OCO-2/3 crossing simulation, we are only giving the model the difference between 2 points at varying times, as such the model often only captures the slope between the observations, which is not always the overall slope of the day. The model is not getting enough information from XCO_2 in order to accurately reproduce diurnal cycles, so it is more reliant on other variables and less affected by error. In the model trained to predict diurnal cycles with the optimal three-observation sampling pattern, the points are at fixed times that correlate strongly with diurnal cycles. With three points at these times the model is able to learn a lot of information from our ΔXCO_2 features about diurnal cycle shape. The model is more reliant on XCO_2 over other variables, and is more affected by error.

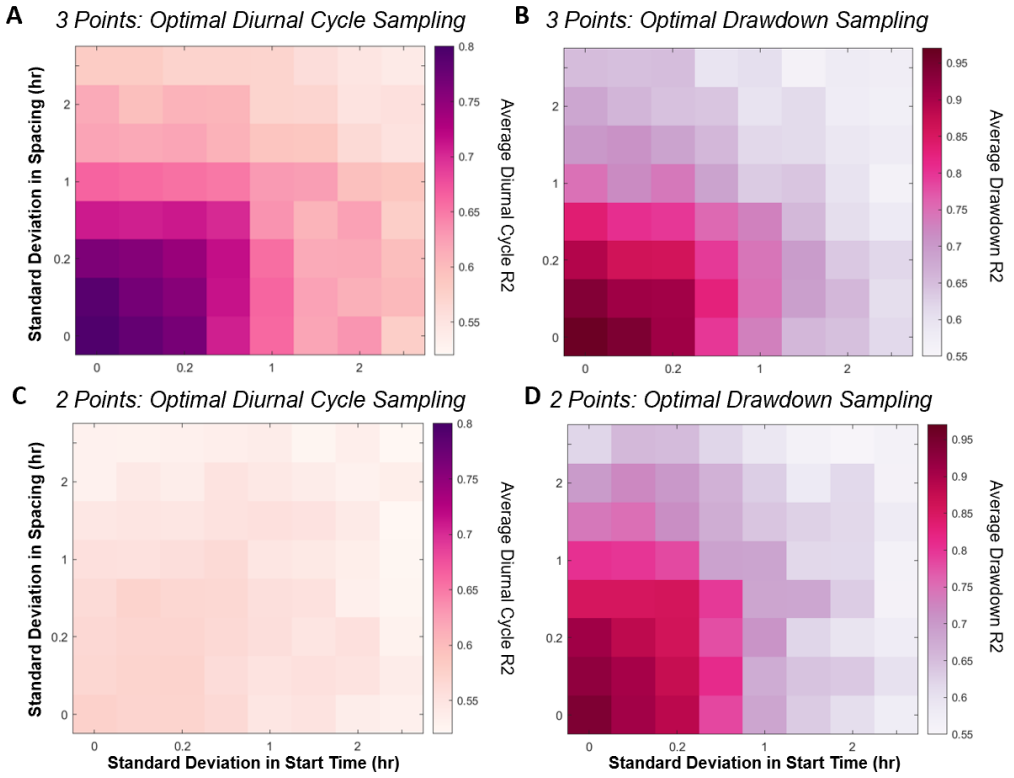

Figure 14. Results of sensitivity tests examining how variation in observation time affects model performance. We increased the standard deviation across observation times for four simulations: and 3 points ideal diurnal cycle configuration (a), 3 points ideal drawdown configuration (b), 2 points, ideal diurnal cycle configuration (c), and 2 points, ideal drawdown configuration (d). The base start time and spacing between points from which we vary are the optimal sampling patterns from Table 5.

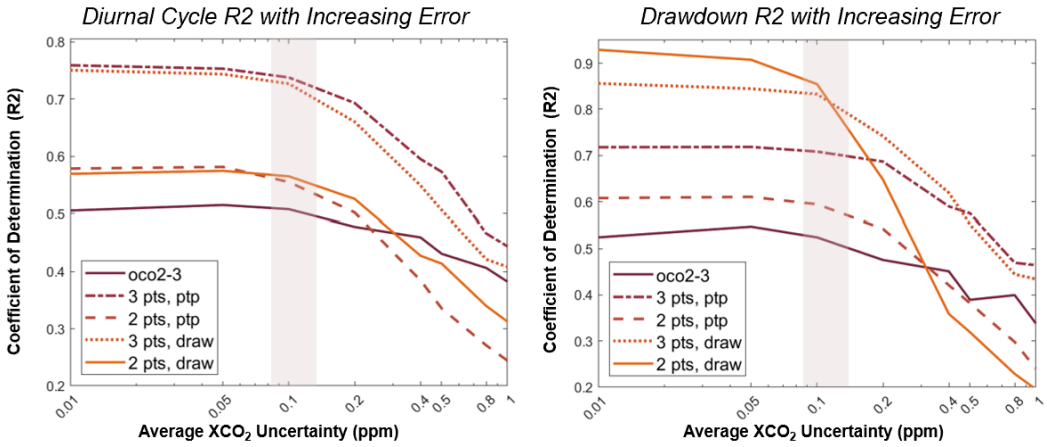

Figure 15. Results from our measurement uncertainty sensitivity tests, where for multiple different observation configurations we added increasing error to the subsampled points. Each line represents the average result from 3 tests using the OCO-2/3 sampling pattern from Sect. 3.1 ("oco2-3") or one of the simulations defined in Table 5. The highlighted region at 0.1 ppm marks the dropoff point for the majority of the simulations. In the legend, "ptp" indicates "point-to-point" correlation (i.e., the ability to reproduce the diurnal cycle) and "draw" means "drawdown" (i.e., the ability to reproduce just the drawdown value).

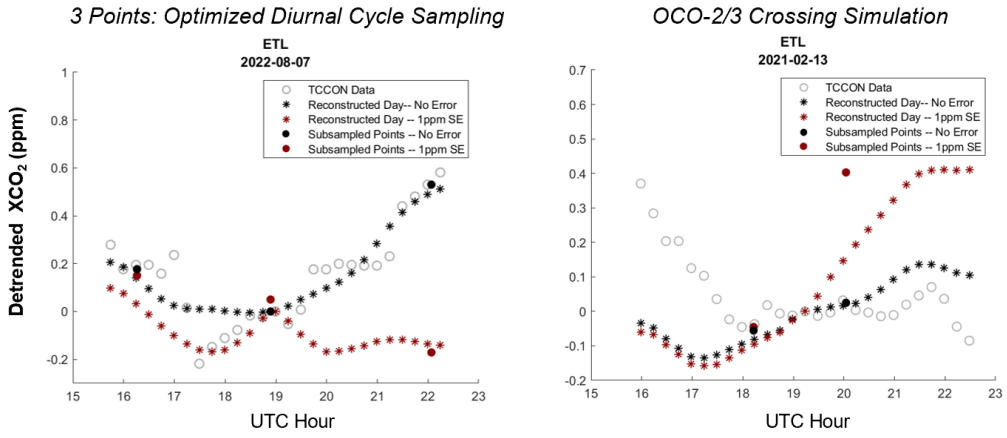

Figure 16. Examples of how error affects model predicted diurnal cycles for two different simulations: the 3 point ideal diurnal cycle sampling times (-3, 0, +3 hr) and the OCO-2/3 crossing sampling pattern. The grey hollow circles show the true diurnal cycle, as taken from the TCCON data. The filled circles show the diurnal cycles predicted by the model (with and without error) and the "subsampled points" are the points used to calculate the ΔXCO_2 the model received as input (with and without error added). For each sampling pattern we had the model predict the diurnal cycle in two cases, one in which the XCO_2 has a standard error of 0 ppm, and one with a standard error of 1 ppm.

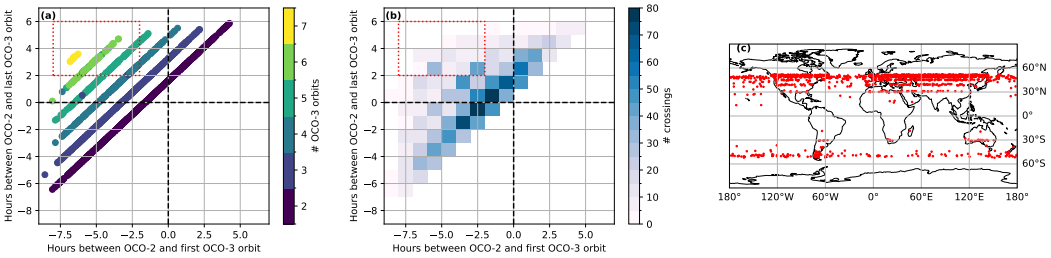

Figure 17. **(a)** The times spanned by cases where multiple OCO-3 orbits cross an OCO-2 orbit. For each crossing, the x and y values show the number of hours between OCO-2 and the first and last, respectively, OCO-3 orbits. The color indicates the number of OCO-2 orbits. **(b)** A 2D histogram of the number of crossings on the same axes as (a). In both (a) and (b), a negative x or y value indicates that OCO-3 observes earlier in the day than OCO-2. **(c)** A map showing the location of the crossings that span more than 2 hours before and after the OCO-2 orbit, which are surrounded by the red box in panels (a) and (b).

4.6 Multiple OCO-3 orbits intersecting one OCO-2 orbit

Figure 17 shows how the cases where multiple OCO-3 orbits intersect a single OCO-2 orbit near the same soundings span different times of day. We can see that a majority of such cases have all the OCO-3 orbits before or after the OCO-2 orbit (i.e., points in quadrants I and III of Fig. 17). We conducted one further set of simulations with two observations in either the morning or afternoon, with the results shown in Fig. 18. As in Sect. 4.3, no error is added to the simulated observations for this section, so as to focus on the impact of the temporal sampling pattern on the model's ability to constrain the diurnal cycle and drawdown.

From Fig. 18, we see that the best R^2 value for either diurnal cycle or drawdown is substantially less than the best performing model using three observations in Figs. 12 or 13, and are only slightly better than the "testing performance" case from Fig. 5. As this test does not include error in the simulated observations, this indicates that cases with observations in only one half of the day will not be able to adequately constrain either the diurnal cycle or drawdown. This is not surprising, as the EOFs divided the morning and afternoon variation of XCO_2 into separate modes of variation, suggesting that the morning and afternoon trends are reasonably independent. In that case, there is a clear need for a morning, noon, and afternoon observation to constrain both of those components of the diurnal cycle and even to correctly infer the drawdown, as that is a difference of XCO_2 before and after noon.

There are, however, a modest number of cases with OCO-3 orbits falling both earlier and later, compared to the OCO-2 orbit. Most of these cases span less than 1 h on either side of the OCO-2 orbit. In Fig. 12, extrapolating to a ± 1 h, 3 observation case will likely have poorer than optimal performance, but may still provide reasonable information about the diurnal cycle and/or drawdown. There are a few crossings further to the upper left of quadrant II in these plots that may perform better. In particular, there are 157 crossings that span at least 2 h before and after OCO-2 in the three years of data we used here. Based on Fig. 12, these are likely to be able to constrain both the diurnal cycle and drawdown reasonably well. However, since Keppel-Aleks et al. (2012) showed that weekly to monthly averages of drawdown-derived NEE are necessary to achieve reasonable correlation with flux towers, these crossings are too infrequent to be useful without a better method to separate transport effects from the true flux.

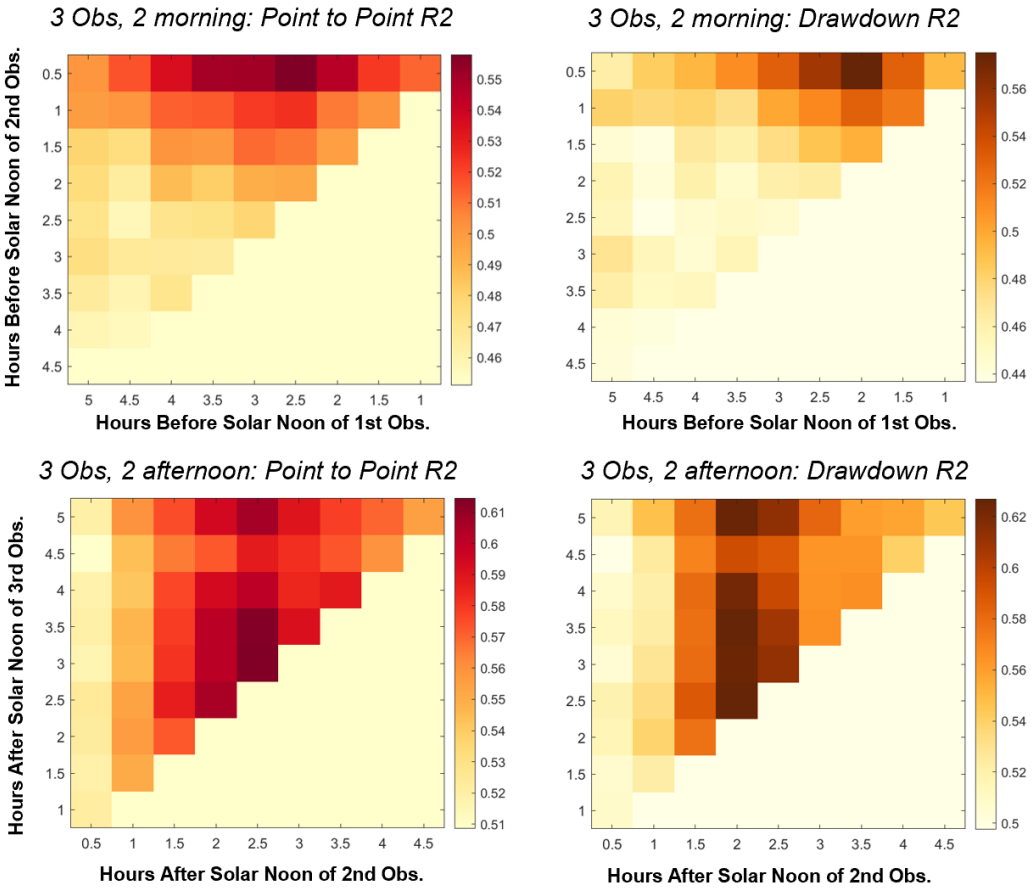

Figure 18. R^2 results for simulations using 3 observations per day mimicking the cases with multiple OCO-3 orbits. In all cases, there is one observation at solar noon. The top row shows results when the two additional observations are in the morning and the bottom row when they are in the afternoon. The left two panels show the models' ability to predict the diurnal cycles as a correlation of individual XCO_2 points and the right panels show the models' ability to predict the daily XCO_2 drawdown. In all panels, the lower-right half is empty because it would otherwise be a simple reflection of the upper-left half.

5 Discussion

5.1 Understanding the feature-prediction relationships

To provide some insight into what features are giving the model the most information, we ran a recursive feature elimination (RFE) and SHapley Additive exPlanations (Lundberg & Lee, 2017) analysis for both the OCO-2/3 crossing simulation, and the 3 observations ideal diurnal cycle simulation. Our process for running the RFE is described in section S1 in the supplement. For the Shapley analysis, we used version 0.46 of the "shap" package (`https://github.com/shap/shap/tree/v0.46.0`, last access 28 Feb 2025). The results of our RFEs are shown in Figs. 19 and 21 and the Shapley analyses are shown in Figs. 20 and 22.

In these Figs. 19 and 21, the y-axis shows the R^2 value for the model predictions versus withheld data for a model only trained on the variables to the left of that x-axis tick as input. That is, in Fig. 19, the left-most point represents a model with only "diff_xco2" as input, the second point represents a model with "diff_xco2" and "delta_xco2", and so on. This provides intuition of which variables are most important to the model's performance.

The RFE shown in Fig. 19 reveals that in the OCO-2/3 crossing simulation, the difference in XCO_2 between observations (without dividing by the time difference, "diff_xco2") is unsurprisingly the most important feature, with a model trained on just this having an R^2 of $\sim$ 0.44 for diurnal cycles and $\sim$ 0.46 for drawdown. The RFE analysis suggests that the XCO_2 change normalized by the time difference is next most important, but slightly reduces the model R^2 for both quantities compared to the model only using "diff_xco2." It is not clear why "delta_xco2" reduces the R^2 in this test. One possibility is that the model misinterprets the variation in this value from the time different in its denominator as a difference in the amount of change in XCO_2 over the day. However, this is a small decrease, and could also be an insignificant difference caused by the randomness of the ML training (particularly in the train/test split).

The next most important features in the OCO-2/3 crossing simulation are "delta solmin", which is the time difference between the first observation and solar noon, the vapor pressure deficits at the fourth and sixth hours, and the solar azimuth and zenith angles of the first observation. The time different and solar angle variables give the model information about what time of day the observations are at, which with variable observation times is necessary for the model to reproduce the diurnal cycle. The presence VPD values in the top five variables suggests the model is finding some predictive power from water availability in this data. In total, all other variables besides ΔXCO_2 improve diurnal cycle and drawdown R^2 by about 0.05 to $\sim$ 0.5 overall.

The Shapley analysis in Fig. 20 shows some similar patterns. The two variables representing the difference in XCO_2 are the top features for PCs 1 and 2. Since both EOFs have an overall positive slope, positive changes in XCO_2 between the first and second observations increase their respective PCs. The dependence of PC 1 on "azim" (solar azimuth angle) indicates that PC 1 increases as the later azimuth angle decreases. Generally, "azim 2" provides information about how far between sunrise and sunset the second measurement occurs. This relationship could either indicate a correlation with season or just when the measurement occurs in the day. We also see dependence on features such as the change in 700 mb potential temperature and pressure that we expect to indicate synoptic scale motion. The positive correlation between change in 700 mb potential temperature and PC 1 matches the relationship seen in Fig. 10a of Keppel-Aleks et al. (2012).

In Fig. 21, the 3 point ideal diurnal cycle RFE shows that having one ΔXCO_2 variable results in a diurnal cycle R^2 of 0.55, and adding a second ΔXCO_2 increases it to about 0.7. (Note that this is the second difference in XCO_2 between a second pair of points,

unlike Fig. 19 where the first two variables were the difference between the same two XCO_2 observations, just with one normalized by the time difference and one not.) This version of the model relies less on variables that give information on the location of the observations, as observations are at fixed times. Rather, all the XCO_2 values are highly ranked as giving the model important information. The non XCO_2 variables do appear to aid the model, but only marginally, as performance only increases by 0.03-0.05 in the inclusion of all of them. This suggests that with sufficient observations, most of the model performance is due to the XCO_2 values and the set of EOFs that describes diurnal cycle variability. Given that XCO_2 provides most of the information in this case, it is possible that a simpler approach, such as a multi linear regression that uses the same base set of EOFs, would perform similarly to our ML method.

In Fig. 22, the Shapley analysis for our simulation assuming 3 observations each 3 hours apart illustrates how the model can find clearer correlations with 3 observations. Both PC 1 and 2 have strong correlations with the change in XCO_2 from one observation to the next. PC 1 has positive correlation with both the differences between the first and second and second and third points ("delta xco2 1" and "delta xco2 2," respectively). This follows, as a positive sign for the delta XCO_2 values indicates an increase in XCO_2 over the day, aligning with the positive slope of EOF 1. PC 2 has a strong positive correlation with the difference in XCO_2 between the second and third XCO_2 value, which maps to the positive increase in the afternoon observed with EOF 2. The negative correlation with "delta xco2 1," which is the morning-to-noon difference, is less intuitive, but suggests that the ML is finding that large morning increases in XCO_2 tend to preclude significant afternoon decrease. For PCs 1 and 2, the other features are of much less importance. Even for PCs 3 through 6, the change in XCO_2 tends to be among the most important features.

5.2 Lessons applicable to mission design

With the current generation of satellites, our ML approach can not rigorously predict diurnal cycles from space-based observations. Two points is not sufficient to reconstruct the targeted daytime part of the diurnal cycles with our method of EOF decomposition, and the crossing patterns of OCO-2/3 do not observe at optimal times for capturing XCO_2 drawdown. Our results align with previous studies that have examined the feasibility of constraining XCO_2 diurnal cycles from OCO-3. Taylor et al. (2023) found that ΔXCO_2 from OCO-3 self crossings is not correlated with the spacing between observations nor any variables from the OCO-2/3-based XCO_2 retrievals and concluded that variability in ΔXCO_2 is dominated by uncertainty in the XCO_2 data. Additionally, Torres (2023) attempted to aggregate OCO-3 observations to constrain climatological diurnal cycles but found OCO-3 observations were not dense enough to overcome observation uncertainties and variation from transport.

There are currently other satellites besides OCO that measure XCO_2. The Greenhouse Gas Observing Satellite (GOSAT) and its follow on, GOSAT-2, observe XCO_2, but they have similar observation times as OCO-2: both are sun-synchronous and observe in the early afternoon. As GOSAT and GOSAT-2 fly north to south on the sunlit side of Earth (while OCO-2 flies south to north), there is some difference in local observation time, particularly in the southern hemisphere. However, the time difference is less than an hour at the northern hemisphere TCCON sites used in our study (Fig. S3). Given that Fig. 12 shows poor R^2 for observations less than 1 h apart, we do not expect that adding GOSAT and GOSAT-2 data would help our method perform better over at least the boreal forests, which are an important region of interest for carbon fluxes. Including GOSAT and GOSAT-2 may provide some benefit in the tropics or southern hemisphere, but we do not have the necessary TCCON data to test that hypothesis at this time. With new TCCON sites being established in the Yucatan Peninsula and Amazon, a future study could investigate this question.

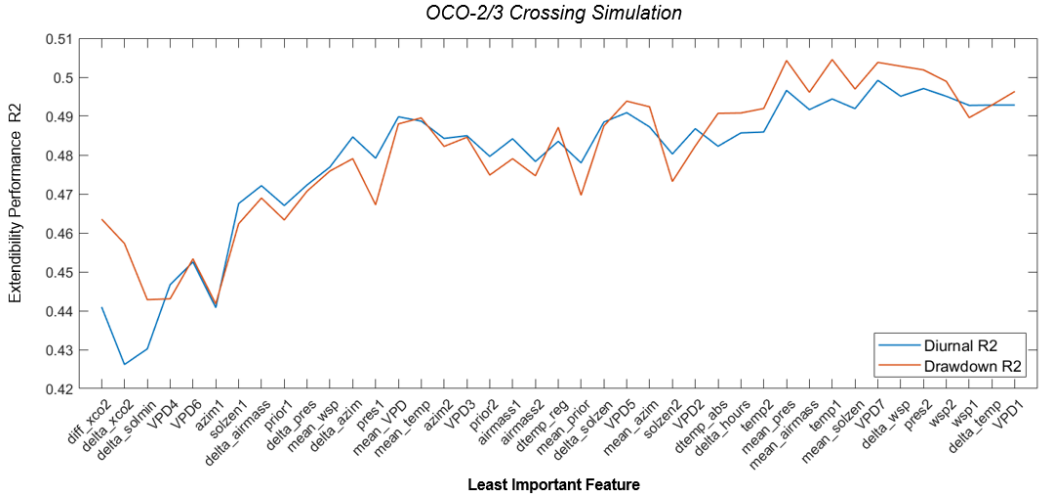

Figure 19. Recursive feature elimination for the OCO-2/3 crossing simulation. Here we record R^2 values for the model trained on a feature set with progressively more variables: the x axis lists the least important variable of the feature set containing that variable and everything to its left. Here, "delta_solmin" refers to the time difference between the first observation and solar noon, "dtemp_abs" refers to abs. potential temp change, "dtemp_reg" refers to potential temp change, "delta_xco2" refers to the difference between XCO_2 observations divided by the time difference and "diff_xco2" is the difference between XCO_2 observations without dividing by the time difference. Performance is shown for both the diurnal cycle and drawdown prediction.

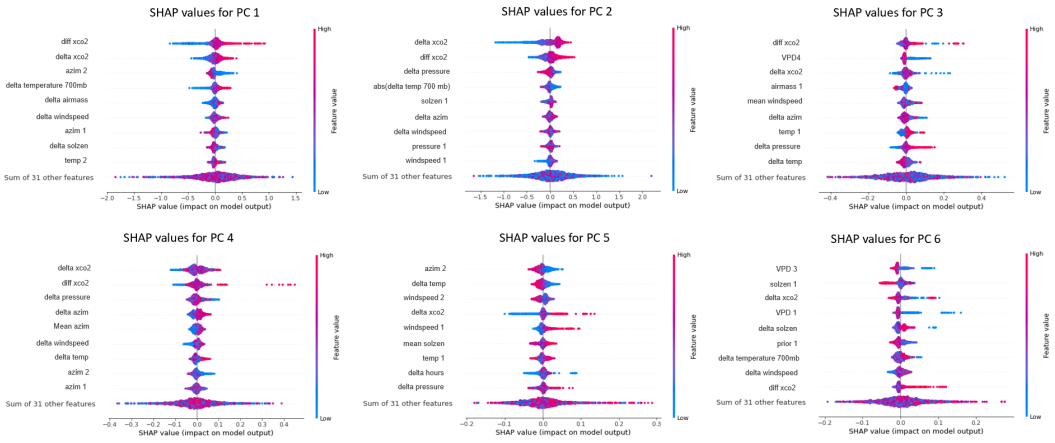

Figure 20. SHapley Additive exPlanations beeswarm plots relating the input features to their effect on the predicted PCs for the OCO-2/3 crossing simulation. The meanings of the feature labels on the y-axes are the same as in Fig. 19

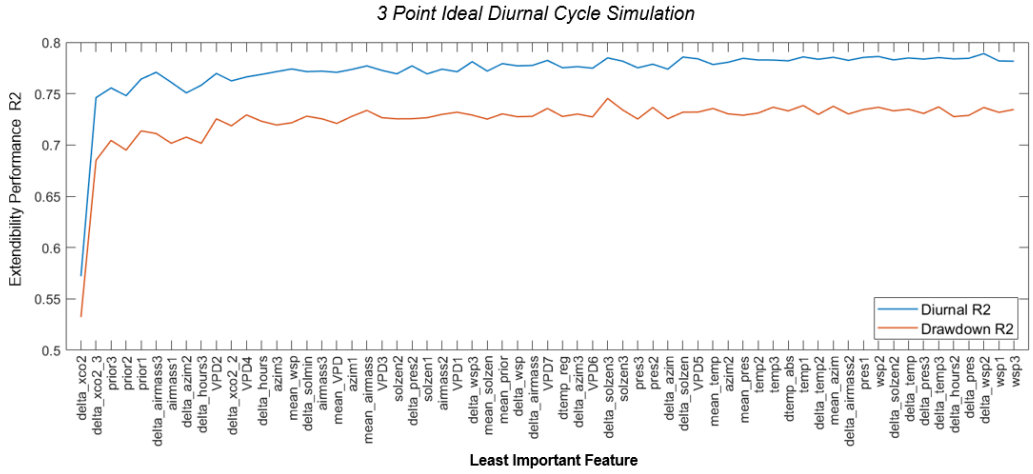

Figure 21. Recursive feature elimination for the 3 observation ideal diurnal cycle simulation. Here we record R^2 values for the model trained on a feature set with progressively more variables: the x axis lists the least important variable of the feature set containing that variable and everything to its left. Generally, "[variable]_N" indicates a variable sampled at the 1st, 2nd, or 3rd simulated observation (except for VPD, which is sampled every hour from 3 hours before solar noon to 3 hours after) while "delta_[variable]" refers to the time derivative between the second and the first observation, "delta_[variable]_2" refers to the time derivative between the third and the first observation, and "delta_[variable]_3" is the time derivative between the third and second observation. Additionally, "delta_solmin" refers to the time difference between the first observation and solar noon, "dtemp_abs" refers to abs. potential temp change, and "dtemp_reg" refers to potential temp change.

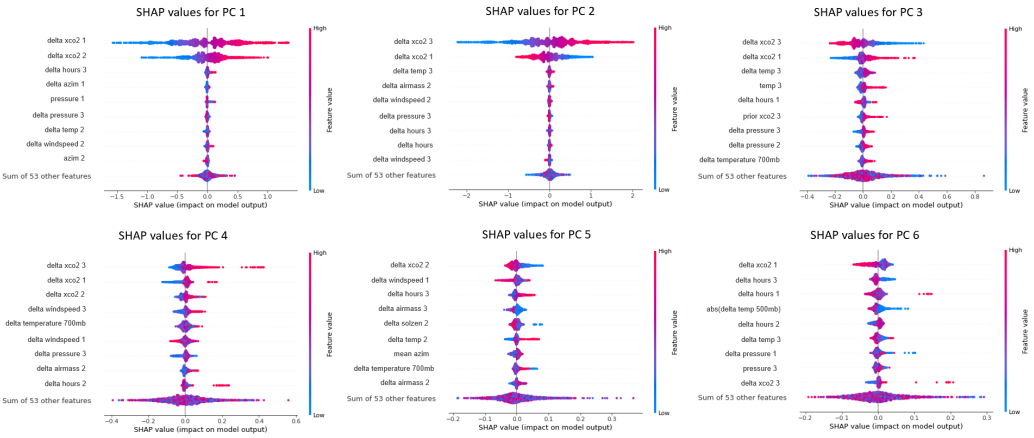

Figure 22. SHapley Additive exPlanations beeswarm plots relating the input features to their effect on the predicted PCs for the 3 observation ideal simulation. The meanings of the feature labels on the y-axes are the same as in Fig. 21

The Geostationary Carbon Cycle Observatory (GeoCarb, Polonsky et al., 2014; Moore III et al., 2018) was a planned geostationary mission that would have had an observation mode where it observes the continental US and Amazon three times in a day, with observations spaced 3-3.5 hours apart at a variable start time. Figure 12 indicates that our approach would potentially do well at inferring diurnal cycles and drawdown from GeoCarb or another satellite with a similar morning-noon-afternoon sampling pattern. Geostationary observatories, such as GeoCarb, generally have some flexibility with regards to the scanning strategy (Nivitanont et al., 2019), and could be optimized to maximize information content on the diurnal cycle.

More generally, if future missions have a goal of constraining diurnal cycles, the observation pattern is critical. The most important factor to apply this ML approach successfully is the number of observations in a day, as three points almost always outperform two. While three points results in dramatic improvement in performance over two points, adding additional points only marginally improves the model. Once the baseline of three observations is met, our results indicate that the quality of measurement, whether through instrument accuracy or consistency in observation time, is more important than quantity of observations.

We also see that the timing of these observations is key. To robustly constrain the drawdown and daytime portion of the diurnal cycle as we define these quantities requires, at a minimum, both a morning and afternoon observation. Observations only in the morning or afternoon do not provide sufficient information to constrain these quantities (Fig. 18). Additionally, this work focuses on quantities for the sunlit portion of the day out of necessity, because we use shortwave infrared instruments, which rely on sunlight. To extend this method to a full 24-hour diurnal cycle, and therefore enable analyses of nighttime respiration compared to daytime photosynthesis, would require a nighttime XCO_2 product with sub-ppm precision and minimal bias relative to daytime products (be those shortwave or thermal infrared). This could be obtained from a thermal-IR or lidar instrument, though the thermal-IR product would likely lack sensitivity near the surface and a lidar product would have a very narrow field of view (of order tenths of milliradians (Han et al., 2024), meaning care would need to be taken to ensure that these measurements are representative of the area being studied).

After number of observations, the standard error in XCO_2 observations is the next most important factor in the performance of our ML method. Our sensitivity tests indicate a standard error of below 0.1 ppm is needed for ML to perform well. (Recall that this is the standard error of the mean of a number of soundings per crossing, rather than a per-sounding error. See Sect. 2.5.4 for details on its calculation.) From our standard errors calculated from the OCO satellites and shown in Table 3, when systematic error cancels, all the standard errors were below the 0.1 ppm cutoff. However, when no systematic error cancelled, the systematic term dominated and the overall standard error in all simulations using this pessimistic assumption was well above the 0.1 ppm ideal for our approach. This leads to two conclusions. First, it is crucial that time-of-day-dependence errors in the retrieval algorithm used by the satellite be minimized, otherwise the standard error between mean XCO_2 at different times of day will likely exceed the 0.1 ppm threshold. Second, if the observations at different times of day are made by different instruments, then it is imperative that these instruments be validated against a common standard and bias among the instruments be minimized, either through consistent instrument design or post hoc bias correction.

Time of first observation, spacing between observations, and variability in observation times are more flexible parameters. Figure 12 shows that for 3 observations, there are a range of times that perform well for both diurnal cycle and drawdown accuracy. Diurnal cycle accuracy is highest when there is a point in the morning, a point around solar noon, and a point in the afternoon, and drawdown accuracy is highest when there are points near the drawdown definition times, but there are a range of times that sat-

isfy those criteria. Following from that, our ML approach performs best when the observation times are consistent across days and locations, but Fig. 14 shows that absolute accuracy is not required. Rather, we find that variations in observation times of up to 0.2 h does not affect the performance of our ML method. This flexibility agrees with the range of observation times about the ideal times that also have high performance. For diurnal cycle reconstructions, variability in start time and spacing have the same effect, but for drawdown accuracy it is more important to keep the start time as fixed as possible.

6 Conclusions

In this work, we test an ML approach for deriving diurnal cycles of XCO_2 for the 6.5 hours about solar noon from sparse space-based observations. We train our model on simulated current and potential future satellite temporal observation patterns and validate its outputs against diurnal cycles from TCCON. Our work concludes that with current satellite sampling patterns and available training data, our ML methods are not able to overcome sparse observations and reconstruct diurnal cycles, which agrees with past research using different methods with similar goals. It is possible that using alternate ML methods (e.g., neural networks) could overcome the challenges imposed by sparse data, which could be tested in future work. However, based on our results, we predict that the sparse temporal sampling of current-generation space-based XCO_2 will remain a challenge for such studies.

For this study, we chose two specific quantities to predict (quarter-hour average XCO_2 values in the 6.5 h centered on solar noon and the 10a to 2p difference in XCO_2). This allowed us to precisely quantify the performance of our chosen ML approach. For drawdown, our results suggest that morning and afternoon data are necessary for accurate estimates. Likewise, for the diurnal cycles, we saw that at least three points were necessary. As these requirements make sense from a physical standpoint, we expect that our results will generally hold true even if the exact definition of the target quantities change.

We then look to future missions and find the minimum requirements our ML approach needs in order to successfully predict diurnal cycles. Specifically, we find that three observations per day centered around solar noon are necessary to reliably constrain diurnal cycles of XCO_2 (i.e., the change in XCO_2 during the core sunlit part of the day) and that once this minimum sampling pattern is achieved, consistent timing and quality of observations are the more important characteristics. However, with such sampling patterns, it is possible that other, more direct methods that do not rely on ML could accomplish the same goal.

Other future work could investigate the possibility of 24 h diurnal cycles inferred from a combination of near-IR, thermal-IR, and/or lidar remote observations. Such an approach would potentially enable the separation of photosynthesis and respiration, and would be a powerful tool to test the fidelity of these processes in models.

7 Open Research

The TCCON data used for training and validating our model in the study are available to the public through CaltechDATA at `https://tccondata.org/` (last accessed 28 Feb 2025) (Wunch et al., 2022; García et al., 2022; Sherlock et al., 2022a, 2022b; Pollard et al., 2022; Petri et al., 2024; Wennberg, Wunch, et al., 2022; Wennberg, Roehl, et al., 2022). The GEOS FP-IT data are used to supplement our feature set in the study (Lucchesi, 2018). The OCO-2 data used for creating crossing distributions and error estimates are available to the public through NASA at `https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_11.1r/summary?keywords=OCO-2` with the newer version at `https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Standard_11.2/summary` (Science

Computing Facility, Jet Propulsion Laboratory, 2017). The OCO-3 data used for creating crossing distributions and error estimates are available to the public through NASA at https://disc.gsfc.nasa.gov/datasets/OCO3_L2_Lite_FP_10.4r/summary?keywords=OCO-3 (Science Computing Facility, Jet Propulsion Laboratory, 2022).

Data was processed and analyzed using MATLAB R2023b. The ML model was run through a Python environment with the XGBoost and sklearn packages (Chen & Guestrin, 2016). EOF decomposition and figure making was aided by the Climate Data Toolbox for Matlab, available at https://www.mathworks.com/matlabcentral/fileexchange/70338-climate-data-toolbox-for-matlab (last accessed 28 Feb 2025) (Greene et al., 2019). Shapley analyses were performed with the shap package, v0.46.0 (https://github.com/shap/shap/tree/v0.46.0, last accessed 28 Feb 2025) The data processing and ML training/evaluation code is available at https://github.com/joshua-laughner/ml-xco2-diurnal-cycles (last accessed 5 Jul 2024) and archived on CaltechDATA (Marchetti & Laughner, 2024).

Acknowledgments

A portion of this research was carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with NASA (80NM0018D0004). OCO-2 lite files were produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center OCO-3 lite files were produced by the OCO-3 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-3 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. The authors thank the OCO-2 and -3 teams for their input throughout this project. The TCCON data were obtained from the TCCON Data Archive hosted by CaltechDATA at https://tccondata.org and the individual sites used are listed in Table 1.

References

Bolton, D. (1980, July). The Computation of Equivalent Potential Temperature. *Monthly Weather Review*, *108*(7), 1046–1053. Retrieved from http://dx.doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2 doi: 10.1175/1520-0493(1980)108⟨1046:tcoept⟩2.0.co;2

Braverman, A., Hobbs, J., Teixeira, J., & Gunson, M. (2021, January). Post hoc Uncertainty Quantification for Remote Sensing Observing Systems. *SIAM/ASA Journal on Uncertainty Quantification*, *9*(3), 1064–1093. Retrieved from http://dx.doi.org/10.1137/19M1304283 doi: 10.1137/19m1304283

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2939672.2939785 doi: 10.1145/2939672.2939785

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., ... Jones, D. B. A. (2019, August). The 2015–2016 carbon cycle as seen from oco-2 and the global in situ network. *Atmospheric Chemistry and Physics*, *19*(15), 9797–9831. Retrieved from http://dx.doi.org/10.5194/acp-19-9797-2019 doi: 10.5194/acp-19-9797-2019

Cui, Y. Y., Zhang, L., Jacobson, A. R., Johnson, M. S., Philip, S., Baker, D., ... Davis, K. J. (2022, September). Evaluating global atmospheric inversions of terrestrial net ecosystem exchange co2 over north america on seasonal and sub-continental scales. *Geophysical Research Letters*, *49*(18). Retrieved from http://dx.doi.org/10.1029/2022GL100147 doi: 10.1029/2022gl100147

Didan, K. (2021). *Modis/aqua vegetation indices 16-day l3 global 0.05deg cmg*

v061. NASA EOSDIS Land Processes Distributed Active Archive Center. Retrieved from https://lpdaac.usgs.gov/products/myd13c1v061/ doi: 10.5067/MODIS/MYD13C1.061

Eldering, A., Taylor, T. E., O'Dell, C. W., & Pavlick, R. (2019). The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data. *Atmospheric Measurement Techniques*, *12*(4), 2341–2370. Retrieved from https://amt.copernicus.org/articles/12/2341/2019/ doi: 10.5194/amt-12-2341-2019

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016, May). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved from http://dx.doi.org/10.5194/gmd-9-1937-2016 doi: 10.5194/gmd-9-1937-2016

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., ... Yokota, T. (2011, September). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, *38*(17). Retrieved from http://dx.doi.org/10.1029/2011GL048738 doi: 10.1029/2011gl048738

García, O. E., Schneider, M., Herkommer, B., Gross, J., Hase, F., Blumenstock, T., & Sepúlveda, E. (2022). *TCCON data from Izana (ES), Release GGG2020.R1.* CaltechDATA. Retrieved from https://data.caltech.edu/records/20242 doi: 10.14291/TCCON.GGG2020.IZANA01.R1

Gier, B. K., Schlund, M., Friedlingstein, P., Jones, C. D., Jones, C., Zaehle, S., & Eyring, V. (2024, November). Representation of the terrestrial carbon cycle in cmip6. *Biogeosciences*, *21*(22), 5321–5360. Retrieved from http://dx.doi.org/10.5194/bg-21-5321-2024 doi: 10.5194/bg-21-5321-2024

Greene, C. A., Thirumalai, K., Kearney, K. A., Delgado, J. M., Schwanghart, W., Wolfenbarger, N. S., ... Blankenship, D. D. (2019, July). The Climate Data Toolbox for MATLAB. *Geochemistry, Geophysics, Geosystems*, *20*(7), 3774–3781. Retrieved from http://dx.doi.org/10.1029/2019GC008392 doi: 10.1029/2019gc008392

Han, G., Zhang, H., Huang, Y., Chen, W., Mao, H., Zhang, X., ... Gong, W. (2024). First global xco2 products from spaceborne lidar: Methodology, results and potentials. Retrieved from http://dx.doi.org/10.2139/ssrn.5067761 doi: 10.2139/ssrn.5067761

Ichii, K., Ueyama, M., Kondo, M., Saigusa, N., Kim, J., Alberto, M. C., ... Zhao, F. (2017, April). New data-driven estimation of terrestrial CO_2 fluxes in Asia using a standardized database of eddy covariance measurements, remote sensing data, and support vector regression. *Journal of Geophysical Research: Biogeosciences*, *122*(4), 767–795. Retrieved from http://dx.doi.org/10.1002/2016JG003640 doi: 10.1002/2016jg003640

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., ... Reichstein, M. (2019, May). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, *6*(1). Retrieved from http://dx.doi.org/10.1038/s41597-019-0076-8 doi: 10.1038/s41597-019-0076-8

Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., ... Wofsy, S. C. (2012, March). The imprint of surface fluxes and transport on variations in total column carbon dioxide. *Biogeosciences*, *9*(3), 875–891. Retrieved from http://dx.doi.org/10.5194/bg-9-875-2012 doi: 10.5194/bg-9-875-2012

Körner, C., Möhl, P., & Hiltbrunner, E. (2023, June). Four ways to define the growing season. *Ecology Letters*, *26*(8), 1277–1292. Retrieved from http://dx.doi.org/10.1111/ele.14260 doi: 10.1111/ele.14260

Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., ... Wolf, J. (2016, February). Consistent evaluation of ACOS-GOSAT, BESD-

SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON. *Atmospheric Measurement Techniques*, *9*(2), 683–709. Retrieved from http://dx.doi.org/10.5194/amt-9-683-2016 doi: 10.5194/amt-9-683-2016

Laughner, J. L., Roche, S., Kiel, M., Toon, G. C., Wunch, D., Baier, B. C., ... Wennberg, P. O. (2023). A new algorithm to generate a priori trace gas profiles for the GGG2020 retrieval algorithm. *Atmospheric Measurement Techniques*, *16*(5), 1121–1146. Retrieved from https://amt.copernicus.org/articles/16/1121/2023/ doi: 10.5194/amt-16-1121-2023

Laughner, J. L., Toon, G. C., Mendonca, J., Petri, C., Roche, S., Wunch, D., ... Wennberg, P. O. (2024, May). The Total Carbon Column Observing Network's GGG2020 data version. *Earth System Science Data*, *16*(5), 2197–2260. Retrieved from http://dx.doi.org/10.5194/essd-16-2197-2024 doi: 10.5194/essd-16-2197-2024

Li, X., & Xiao, J. (2019). A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from oco-2, modis, and reanalysis data. *Remote Sensing*, *11*(5). Retrieved from https://www.mdpi.com/2072-4292/11/5/517 doi: 10.3390/rs11050517

Lucchesi, R. (2018). File Specification for GEOS-5 FP-IT. *GMAO Office Note*, *2*(1.3), 1–60. Retrieved from http://gmao.gsfc.nasa.gov/pubs/office_notes.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Marchetti, C., & Laughner, J. (2024). *XCO2 diurnal cycles from machine learning.* doi: 10.22002/0qcyc-93t57

Mitchell, K. A., Doney, S. C., & Keppel-Aleks, G. (2023, January). Characterizing average seasonal, synoptic, and finer variability in orbiting carbon observatory-2 xco2 across north america and adjacent ocean basins. *Journal of Geophysical Research: Atmospheres*, *128*(3). Retrieved from http://dx.doi.org/10.1029/2022JD036696 doi: 10.1029/2022jd036696

Moore III, B., Crowell, S. M., Rayner, P. J., Kumer, J., O'Dell, C. W., O'Brien, D., ... Lemen, J. (2018). The potential of the Geostationary Carbon Cycle Observatory (GeoCarb) to provide multi-scale constraints on the carbon cycle in the Americas. *Frontiers in Environmental Science*, *6*, 109. doi: 10.3389/fenvs.2018.00109

Nivitanont, J., Crowell, S. M., & Moore III, B. (2019). A scanning strategy optimized for signal-to-noise ratio for the Geostationary Carbon Cycle Observatory (GeoCarb) instrument. *Atmospheric Measurement Techniques*, *12*(6), 3317–3334. doi: 10.5194/amt-12-3317-2019

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., ... Velazco, V. A. (2018). Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmospheric Measurement Techniques*, *11*(12), 6539–6576. Retrieved from https://amt.copernicus.org/articles/11/6539/2018/ doi: 10.5194/amt-11-6539-2018

O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., & Chaubell, J. (2023). *Smap l3 radiometer global daily 36 km ease-grid soil moisture, version 9.* NASA National Snow and Ice Data Center Distributed Active Archive Center. Retrieved from http://nsidc.org/data/SPL3SMP/versions/9 doi: 10.5067/4XXOGX0OOW1S

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., ... Baker, I. (2022, January). Four years of global carbon cycle observed from the orbiting carbon observatory 2 (oco-2) version 9 and in situ data and

comparison to oco-2 version 7. *Atmospheric Chemistry and Physics*, *22*(2), 1097–1130. Retrieved from http://dx.doi.org/10.5194/acp-22-1097-2022 doi: 10.5194/acp-22-1097-2022

Petri, C., Vrekoussis, M., Rousogenous, C., Warneke, T., Sciare, J., & Notholt, J. (2024). *TCCON data from Nicosia (CY), Release GGG2020.R1.* CaltechDATA. Retrieved from https://data.caltech.edu/records/9kdk2-c5881 doi: 10.14291/TCCON.GGG2020.NICOSIA01.R1

Pollard, D. F., Robinson, J., & Shiona, H. (2022). *TCCON data from Lauder (NZ), Release GGG2020.R0.* CaltechDATA. Retrieved from https://data.caltech.edu/records/20122 doi: 10.14291/TCCON.GGG2020.LAUDER03.R0

Polonsky, I. N., O'Brien, D. M., Kumer, J. B., O'Dell, C. W., & the geoCARB Team. (2014). Performance of a geostationary mission, geoCARB, to measure CO_2, CH_4 and CO column-averaged concentrations. *Atmospheric Measurement Techniques*, *7*(4), 959–981. doi: 10.5194/amt-7-959-2014

Salby, M. L. (1995). Fundamentals of atmospheric physics.. Retrieved from https://api.semanticscholar.org/CorpusID:117842004

Science Computing Facility, Jet Propulsion Laboratory. (2017). *OCO-2 Level 2 bias-corrected XCO2 and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing V11.1r.* NASA Goddard Earth Sciences Data and Information Services Center. Retrieved from https://disc.gsfc.nasa.gov/datacollection/OCO2_L2_Lite_FP_11.1r.html doi: 10.5067/8E4VLCK16O6Q

Science Computing Facility, Jet Propulsion Laboratory. (2022). *OCO-3 Level 2 bias-corrected XCO2 and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing v10.4r.* NASA Goddard Earth Sciences Data and Information Services Center. Retrieved from https://disc.gsfc.nasa.gov/datasets/OCO3_L2_Lite_FP_10.4r/summary doi: 10.5067/970BCC4DHH24

Sherlock, V., Connor, B., Robinson, J., Shiona, H., Smale, D., & Pollard, D. F. (2022a). *TCCON data from Lauder (NZ), 120HR, Release GGG2020.R0.* CaltechDATA. Retrieved from https://data.caltech.edu/records/20120 doi: 10.14291/TCCON.GGG2020.LAUDER01.R0

Sherlock, V., Connor, B., Robinson, J., Shiona, H., Smale, D., & Pollard, D. F. (2022b). *TCCON data from Lauder (NZ), 125HR, Release GGG2020.R0.* CaltechDATA. Retrieved from https://data.caltech.edu/records/20121 doi: 10.14291/TCCON.GGG2020.LAUDER02.R0

Taylor, T. E., O'Dell, C. W., Baker, D., Bruegge, C., Chang, A., Chapsky, L., ... Zong, J. (2023). Evaluating the consistency between OCO-2 and OCO-3 XCO_2 estimates derived from the NASA ACOS version 10 retrieval algorithm. *Atmospheric Measurement Techniques*, *16*(12), 3173–3209. Retrieved from https://amt.copernicus.org/articles/16/3173/2023/ doi: 10.5194/amt-16-3173-2023

Torres, A. (2023). *Using Ground- And Space-Based Observations to Quantify Within-Day to Synoptic-scale Variance Budgets of Total Column Averaged Carbon Dioxide (XCO2) Measurements.* University of Michigan Library. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/177836 doi: 10.7302/8293

Wallace, J., & Hobbs, P. (2006). *Atmospheric Science: An Introductory Survey.* Elsevier. Retrieved from http://dx.doi.org/10.1016/C2009-0-00034-8 doi: 10.1016/c2009-0-00034-8

Wennberg, P. O., Roehl, C. M., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., ... Allen, N. T. (2022). *TCCON data from Park Falls (US), Release GGG2020.R1.* CaltechDATA. Retrieved from https://data.caltech.edu/records/xa6rd-9g966 doi: 10.14291/TCCON.GGG2020.PARKFALLS01.R1

Wennberg, P. O., Wunch, D., Roehl, C. M., Blavier, J.-F., Toon, G. C., & Allen,

N. T. (2022). *TCCON data from Lamont (US), Release GGG2020.R0.* CaltechDATA. Retrieved from https://data.caltech.edu/records/20123 doi: 10.14291/TCCON.GGG2020.LAMONT01.R0

Wunch, D., Mendonca, J., Colebatch, O., Allen, N. T., Blavier, J.-F., Kunz, K., ... Strong, K. (2022). *TCCON data from East Trout Lake, SK (CA), Release GGG2020.R0.* CaltechDATA. Retrieved from https://data.caltech.edu/records/20114 doi: 10.14291/TCCON.GGG2020.EASTTROUTLAKE01.R0

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., ... Wennberg, P. O. (2011, May). The Total Carbon Column Observing Network. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1943), 2087–2112. Retrieved from http://dx.doi.org/10.1098/rsta.2010.0240 doi: 10.1098/rsta.2010.0240

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., ... Eldering, A. (2017). Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON. *Atmospheric Measurement Techniques*, *10*(6), 2209–2238. Retrieved from https://amt.copernicus.org/articles/10/2209/2017/ doi: 10.5194/amt-10-2209-2017

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., ... Torn, M. S. (2008). Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, *148*(11), 1827-1847. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192308001925 doi: https://doi.org/10.1016/j.agrformet.2008.06.015

Zeng, J., Matsunaga, T., Tan, Z.-H., Saigusa, N., Shirai, T., Tang, Y., ... Fukuda, Y. (2020, September). Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest. *Scientific Data*, *7*(1). Retrieved from http://dx.doi.org/10.1038/s41597-020-00653-5 doi: 10.1038/s41597-020-00653-5

Zhang, W., Luo, G., Yuan, X., Li, C., Xie, M., Wang, Y., ... De Maeyer, P. (2023). New data-driven method for estimation of net ecosystem carbon exchange at meteorological stations effectively increases the global carbon flux data. *Methods in Ecology and Evolution*, *14*(9), 2449-2463. Retrieved from https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14188 doi: https://doi.org/10.1111/2041-210X.14188

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., & Gentine, P. (2018). A global spatially contiguous solar-induced fluorescence (csif) dataset using neural networks. *Biogeosciences*, *15*(19), 5779–5800. Retrieved from https://bg.copernicus.org/articles/15/5779/2018/ doi: 10.5194/bg-15-5779-2018

Zhou, Q., Fellows, A., Flerchinger, G. N., & Flores, A. N. (2019, February). Examining Interactions Between and Among Predictors of Net Ecosystem Exchange: A Machine Learning Approach in a Semi-arid Landscape. *Scientific Reports*, *9*(1). Retrieved from http://dx.doi.org/10.1038/s41598-019-38639-y doi: 10.1038/s41598-019-38639-y