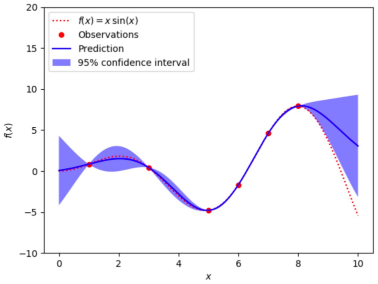
Statistical Emulator For Fast Uncertainty Quantification

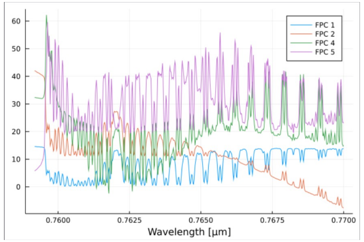
Otto Lamminpää*, Jonathan Hobbs*, Amy Braverman*, Pulong Ma**, Anirban Mondal***

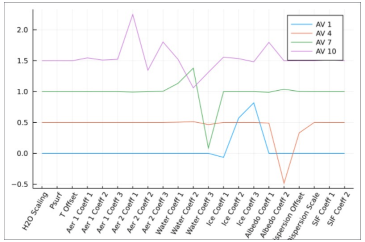
* Jet Propulsion Laboratory
** Duke University
*** Case Western Reserve University

We propose a fast Gaussian process emulator for the OCO-2 retrieval problem. Idea is to predict radiances at unknown atmospheric states based on a model trained with known radiances y_i obtained at known states x_i



Dimension reduction:
Functional Principal Components for y,
Active Subspace for x





$$X \to X[O_2],\quad X[WCO_2],\quad X[SCO_2]$$

$$X_r^{O_2} = W_r^T X[O_2] \qquad X_r^{WCO_2} = W_r^T X[WCO_2] \qquad X_r^{SCO_2} = W_r^T X[SCO_2]$$

$$z_i = GP_i(X_r^{SCO_2}) \qquad z_j = GP_j(X_r^{SCO_2}) \qquad z_k = GP_k(X_r^{SCO_2})$$

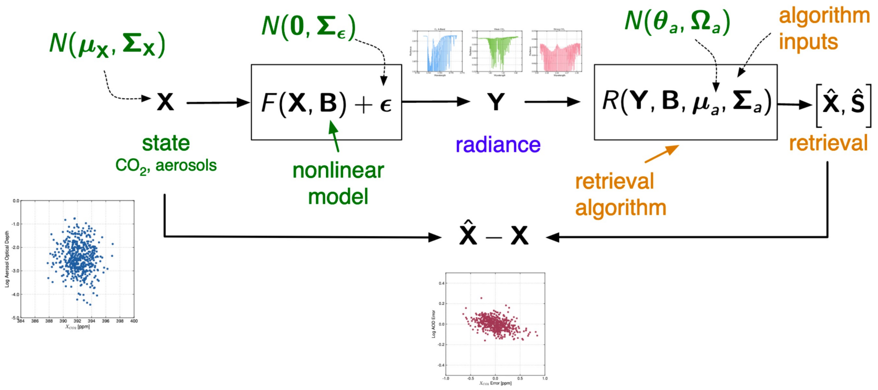
$$y[O_2] = \mu_{O_2} + \sum_{i=1}^{N_{O_2}} \eta_i z_i \qquad y[WCO_2] = \mu_{WCO_2} + \sum_{j=1}^{N_{WCO_2}} \eta_j z_j \qquad y[SCO_2] = \mu_{SCO_2} + \sum_{k=1}^{N_{SCO_2}} \eta_k z_k \quad \to y$$

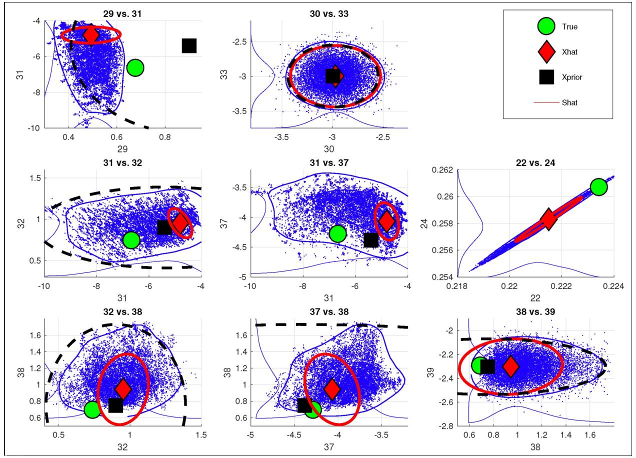
Final pipeline for an emulator predicting radiances y from atmospheric states x

Jet Propulsion Laboratory
California Institute of Technology

Motivation for an emulator

Rigorous Uncertainty Quantification (UQ) for remote sensing retrievals often relies on large amount of forward model evaluations, which becomes prohibitively expensive in the case of OCO-2 ACOS Full Physics model. We aim to replace the costly Full Physics model by a fast *Gaussian Process Emulator*.

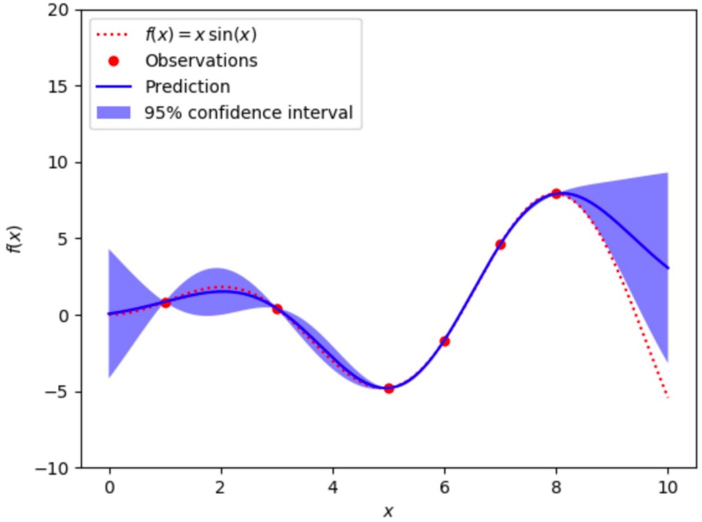




Simulation based UQ of Hobbs et al. (2017) utilizes a large amount of forward model evaluations to get radiances and consecutive retrievals to produce a distribution of true vs retrieved states.

Markov Chain Monte Carlo (MCMC) implemented for OCO-2 in Lamminpää et al. (2020) evaluates the Full Physics model numerous time to produce samples from exact posterior distribution of a retrieval.

Interpretable Machine Learning: Gaussian Processes



A *Gaussian Process* (GP) is a statistical method for learning a function *f* from a set of evaluations $y_i = f(x_i)$ at known points x_i . A GP is specified by training a Kernel function K by specifying its parameters θ. Predictions at new point x are obtained as

$$y = K(x,X)K(X,X)^{-1}Y$$

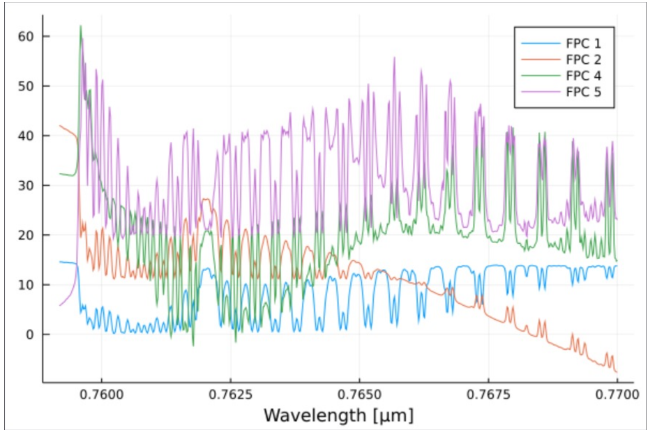
where X and Y represent all training points and corresponding training data.
A GP also provides confidence intervals for predictions, so the user can assess the credibility and performance of the predictions unlike in many other machine learning methods. The uncertainty is given by

$$\sigma = K(x,x) - K(x,X)K(X,X)^{-1}\,K(X,x)$$

Dimension Reduction

Using a spline basis, *Functional Principal Component Analysis* (FPCA) extracts a low dimensional decomposition for radiances. This approach is suited for OCO-2 output, which is obtained on a non-uniform wavelength grid

$$y_k(\omega) = \mu_k(\omega) + \sum_{i=1}^{p_k} \eta_{k,i}(\omega) z_{k,i}(x)$$



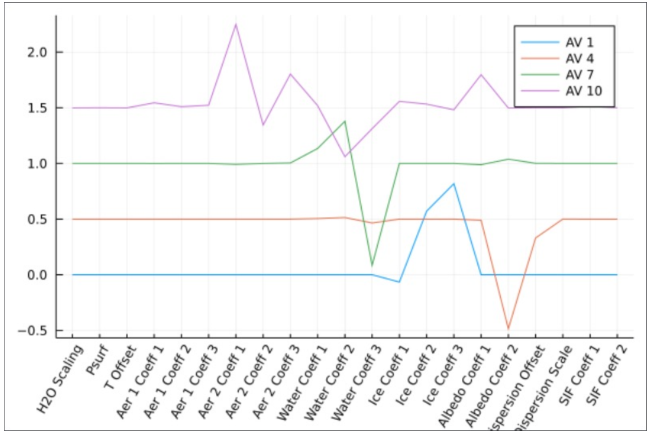
Example FPCA basis vectors for O_2 band

For each wavelength band, we find a gradient-based *Active Subspace* (AS), which captures the principal components of the state vector that the forward model is most sensitive to. This way we get a low-dimensional decomposition for the state.

$$\boldsymbol{C} = \mathbb{E}[(\nabla_x z_i(x))(\nabla_x z_i(x))^T]$$

$$\Rightarrow \boldsymbol{C} = \boldsymbol{W} \Lambda \boldsymbol{W}^T, \quad \boldsymbol{W} = [\boldsymbol{W}_r, \boldsymbol{W}_\perp]$$

$$\Rightarrow x_r = \boldsymbol{W}_r^T x, \quad x_\perp = \boldsymbol{W}_\perp^T x$$



Example AS basis vectors for O_2 band

Emulator Flowchart

X

$$X[O_2] \qquad X_r^{O_2} = W_r^T X[O_2] \qquad z_i = GP_i(X_r^{SCO_2}) \qquad y[O_2] = \mu_{O_2} + \sum_{i=1}^{N_{O_2}} \eta_i z_i$$

$$X[WCO_2] \qquad X_r^{WCO_2} = W_r^T X[WCO_2] \qquad z_j = GP_j(X_r^{SCO_2}) \qquad y[WCO_2] = \mu_{WCO_2} + \sum_{j=1}^{N_{WCO_2}} \eta_j z_j$$

$$X[SCO_2] \qquad X_r^{SCO_2} = W_r^T X[SCO_2] \qquad z_k = GP_k(X_r^{SCO_2}) \qquad y[SCO_2] = \mu_{SCO_2} + \sum_{k=1}^{N_{SCO_2}} \eta_k z_k$$

y

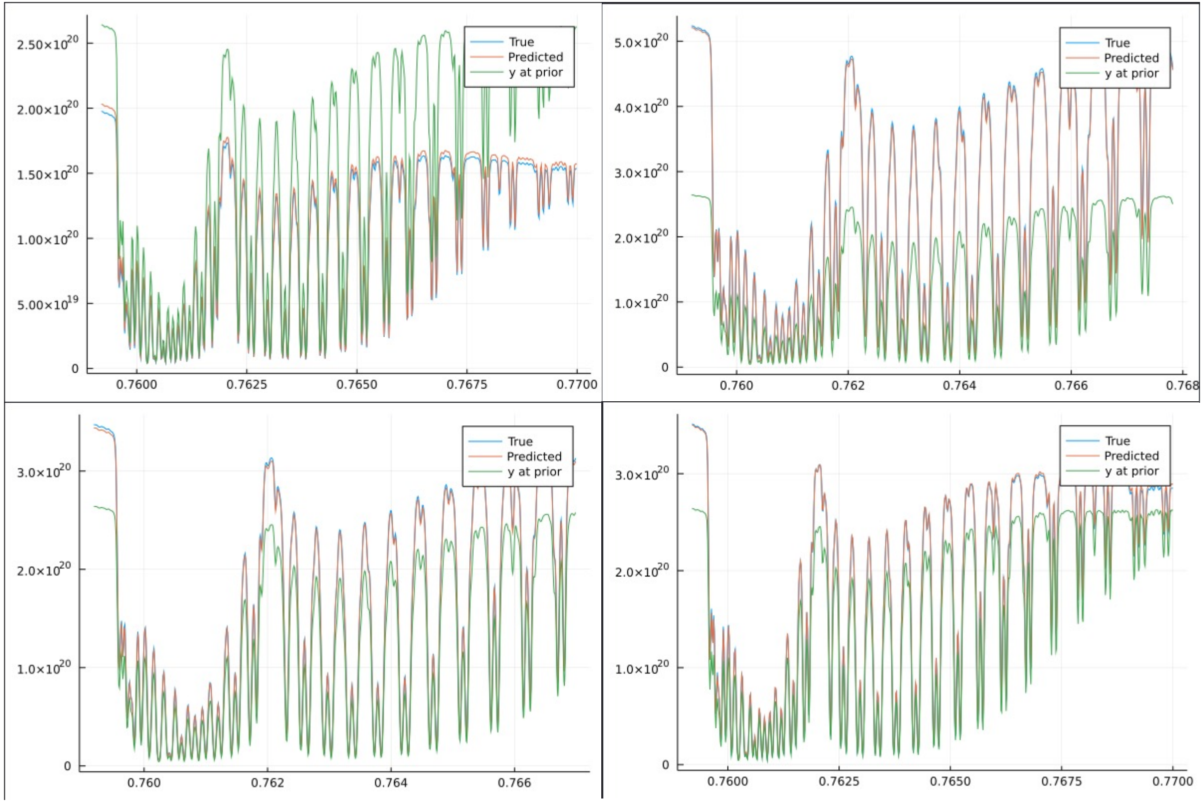
Partition X to components each band is sensitive to

Extract the AS component of X for each band

Predict FPCA coefficients z with independent Gaussian Processes

Assemble predicted coefficients, multiplied by FPCA basis functions, and mean vectors to output radiances

Results: O_2 band



Full Physics forward model evaluated at a known true state and at the prior mean, compared to the emulator evaluated at a known true state.

Our statistical emulator is thus shown to reproduce the output of the OCO-2 Full Physics model on the O_2 band with great accuracy. Although work is ongoing for WCO_2 and SCO_2 bands, these results show the potential of the emulator to be used in a retrieval setting.

The emulator is also computationally fast: an evaluation of Full Physics model takes minutes on a personal computer, while the emulator is evaluated in only milliseconds.