Post hoc Uncertainty Quantification for Remote Sensing Observing Systems [*]

Amy Braverman[†], Jonathan Hobbs[†], Joaquim Teixeira[†], and Michael Gunson[†]

Abstract. This article sets forth a practical methodology for uncertainty quantification of physical state estimates derived from remote sensing observing systems. Remote sensing instruments observe parts of the electromagnetic spectrum and use computational algorithms to infer the underlying true physical states. In current practice, many sources of uncertainty are not accounted for in this process, leading to underestimates of uncertainties on quantities of interest. We propose a procedure that combines Monte Carlo simulation experiments with statistical modeling to approximate distributions of unknown true states given point estimates of those states. Our method is carried out Post hoc; that is after the operational processing step. We demonstrate the procedure using four months of data from NASA's Orbiting Carbon Observatory-2 mission and compare to validation measurements from the Total Carbon Column Observing Network.

Key words. Uncertainty quantification, bootstrap bias correction, remote sensing, Gaussian mixture modeling, Orbiting Carbon Observatory-2 mission.

AMS subject classifications. 62G05, 62G07, 62G08, 62G15

1. Introduction.

The ability of space-borne remote sensing observations to address important Earth and climate science problems rests crucially on how well geophysical quantities of interest (QOIs) can be inferred from these data. Observing systems that collect and process this information must address uncertainties arising not only from measurement errors, but also from imperfect physical models and their parameters, computational artifacts, and potentially other unknowns that affect the conversion of observations to QOI estimates. While much of this sounds familiar in the context of the Uncertainty Quantification discipline [32], existing techniques do not address the problem in a practical way that can be applied comprehensively to very large data sets produced in routine operations.

A remote sensing observing system is an infrastructure that senses electromagnetic energy and converts it into estimates of nature's true states. In this paper, we consider a system in which observations are collected over different wavelengths as a spectrum of measured radiances. These spectra carry information about the properties of the Earth's atmosphere and surface, as encountered in each individual observational unit corresponding to a specific ground footprint (also sometimes called a "sounding"), because photons at different wavelengths are scattered and absorbed in characteristic ways, depending on the make-up, function, and properties of physical constituents with which photons interact. Inference about QOIs from noisy radiance spectra is a fundamental problem of remote sensing science. It requires knowledge of the physics of radiative transfer [4], and substantial computational resources, especially for operational satellite systems which can return terabytes of data per day corresponding to

[*] Submitted to the editors December 4, 2019.

Funding: This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

[†] Jet Propulsion Laboratory, California Institute of Technology (Amy.Braverman@jpl.nasa.gov, Jonathan.M.Hobbs@jpl.nasa.gov, Joaquim.Teixeira@jpl.nasa.gov, Michael.Gunson@jpl.nasa.gov).

millions of cases. In nearly all cases, there are only a few, sparse measurements from surface or aircraft instruments with which to validate or calibrate the satellite measurements, and these corroborating data also come with their own uncertainties. Finally, for most existing missions, uncertainty quantification must be done post hoc without rerunning computationally costly data processing algorithms.

Various authors have addressed the problem of uncertainties in satellite derived estimates of geophysical QOIs. Some papers (e.g., [31, 39, 50]) qualify as a general call to arms. Others such as [1, 20, 27] use ground-based validation data in highly restricted case studies to ascertain error characteristics under specific conditions. Numerous other examples can be found in [50]. For operational missions, [37, 36] used the linear sensitivity (first derivative) of the forward radiative transfer model to propagate radiance measurement error forward through their computations. Unlike the case-study-based methods, this method is applied in a way that produces, in principle, a nominal variance for each and every sounding estimate. [30] used a Monte Carlo simulation to quantify uncertainty of their estimates without assuming linearity. However, this analysis was performed only after aggregating to coarse spatial resolution, and only addressed uncertainty due to geographic sampling issues and to several specific methodological choices implemented in their processing stream.

More recently, Bayesian methods have become popular, as they produce probability distributions of the QOI given the observed quantities rather than point estimates alone. Markov chain Monte Carlo (MCMC) is sometimes used in small applications and case studies (e.g., [18, 38, 3, 24]). MCMC is too computationally intensive for routine operational use though. Instead, “optimal estimation” (OE) [41, 42] has been widely adopted as the de facto state-of-the-art (e.g., [51, 21, 25, 49]). OE is a computational implementation of Bayes Rule that produces (implied Gaussian) probability distributions for true QOIs given radiances.

Some argue that OE automatically achieves uncertainty quantification [49] because it produces output that can be interpreted as the moments of a distribution. However, operational implementations and limited physical knowledge result in uncertainties that impact the reliability of OE itself. For example, while the equations of radiative transfer are relatively well understood, operational codes must run quickly and usually approximate some processes (or indeed ignore them completely). Tables of spectroscopic information that describe spectral absorption patterns induced by different gases, and sensitivities of detectors and other parts of the optical system, are derived from ground-based experiments. All are assumed fixed and known, even though they are uncertain. Even the discretizations of continuous physical quantities that define the state and radiance vectors, the grids used by numerical solvers, and the optimization routines used to solve for the QOI can induce uncertainty, both individually and as a result of high-order interactions.

Neither OE nor earlier approaches quantify total uncertainty on a sounding-by-sounding basis. They do not quantify total uncertainty because they rely on enumeration of specific known sources of uncertainty to be propagated, and do not include the elusive unknown unknowns. These methods cannot be applied on a sounding-by-sounding basis because they require ground truth validation data that are not universally available. In this article, we propose a new method for estimating probability distributions for QOIs that is free of these restrictions. We derive conditional distributions of the QOI, given the operational point estimates, by fitting the parameters of a Gaussian mixture regression model to an ensemble of simulated true and estimated states. Then, we use the fitted mixture of regression functions to define the desired conditional distributions by plugging in sounding-specific operational point estimates as predictors. This is a “top-down” approach that does not require enumeration of individual uncertainty sources. Once the parameters of the Gaussian mixture are fitted, it is fast and easy to compute the sounding-specific conditional distributions.

Our method can be seen as a modified and extended version of the the bootstrap bias correction [12, 10, 23]. There, one starts with a single sample, and draws a set of resamples from it. The statistic of interest is computed from the original sample, and from the resamples. The discrepancy between (the mean of) the resampled statistics and the original statistic is used as a proxy for the relationship between the original statistic and the true parameter. Here, our analog of the resamples is a simulated joint ensemble of true and estimated QOIs. However, we go beyond correcting for bias alone with two innovations. The first is that we derive approximations for the full conditional distributions of the true states given the operational state estimates. The second is that we use both the simulated and operationally-derived information together to approximate forward model discrepancy and account for it as part of total uncertainty. As far as we know there are no comparable methods in either the uncertainty quantification or remote sensing literature that deliver such comprehensive probabilistic descriptions of uncertainties associated with operational remote sensing state estimates.

To demonstrate and evaluate our methodology, we apply it to data from NASA’s Orbiting Carbon Observatory-2 (OCO-2) mission. See [6] for an overview of the mission and statistical issues surrounding its processing and scientific value. OCO-2 uses OE to infer the distribution of its primary QOI, total column mole-fraction of CO_2 (known as XCO2 in the remote sensing community) by sounding. While our method is equally applicable to vector-valued QOIs, XCO2 is a scalar quantity; this simplifies visualization and analysis. Another reason to highlight OCO-2 is its stringent uncertainty requirements. The primary scientific application of OCO-2’s estimates is as input to flux inversion (data assimilation) models [17, 35, 48] that estimate the exchange of carbon between Earth’s surface and atmosphere. Determination of flux requires CO_2 estimates with high accuracy (less than 0.3 parts per million (ppm) in scenes with background levels of around 410 ppm), and high precision (standard errors less than 0.5 ppm). Consequently, uncertainty quantification has been a major focus of the OCO-2 science endeavor.

Members of the OCO-2 team have performed various studies that attempt to quantify uncertainties in its retrieved estimates (e.g., [5, 52]). However, many of these stay wholly within the optimal estimation, linear Gaussian framework, and are therefore not able to assess uncertainties due to failure of those assumptions. Alternatively, Cressie and co-authors [7, 8] and [19] investigated the impact of non-linearity of the forward model on both mean and variance estimates produced by OE. These analyses yield valuable insights into the performance of OE, but do not represent an attempt to evaluate or quantify total uncertainty expressed by operationally-derived distributions.

The remainder of this article is organized as follows. First, we articulate our reference statistical model for remote sensing observing systems (Section 2), and then the methodology for estimating conditional distributions of true states given their point estimates (Section 3). In Section 4 we describe how we tailor our method for the case of OCO-2, and evaluate the results through comparisons with available ground truth information. The final section contains a summary and discussion.

2. Statistical model of an observing system and its output. An observing system represents the flow of information from nature, which produces the quantities of interest, to the space-borne hardware that collects radiances, and finally to the software that performs estimation of the QOIs from these observations. Consider Figure 1. The state vector is $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_{L_\mathbf{X}})'$ and it, or some part of it, usually is the quantity of interest. Nature's forward function, F_0, converts $\mathbf{X}$ into the noiseless radiance vector, $\mathbf{Y}_0 = (\mathrm{Y}_{01}, \mathrm{Y}_{02}, \ldots, \mathrm{Y}_{0L_\mathbf{Y}})'$, which is observed by the remote sensing instrument with measurement error $\boldsymbol{\epsilon}$ (also of dimension $L_\mathbf{Y}$). The forward function also typically depends on an additional set of variables, denoted by $\mathbf{b}_0$, that are not part of the state vector but nonetheless influence the transformation of the state into noiseless radiances; e.g., spectroscopic absorption coefficients that characterize various atmospheric constituents. The distinction between F_0 and $\mathbf{b}_0$ is an academic one because they are confounded components of our model of nature, rather than properties of nature itself. Since they can't be decoupled, we will always refer to them as a unit, $F_0(\cdot, \mathbf{b}_0)$.

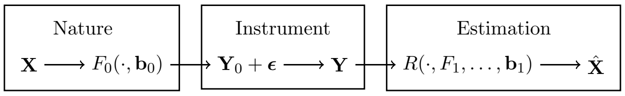
Nature
$\mathbf{X} \longrightarrow F_0(\cdot, \mathbf{b}_0)$
Instrument
$\mathbf{Y}_0 + \boldsymbol{\epsilon} \longrightarrow \mathbf{Y}$
Estimation
$R(\cdot, F_1, \ldots, \mathbf{b}_1) \longrightarrow \hat{\mathbf{X}}$

Figure 1. *Schematic diagram of a generic observing system.*

The noisy radiance vector, $\mathbf{Y} = \mathbf{Y}_0 + \boldsymbol{\epsilon}$ is ingested into the retrieval algorithm– so-named because it retrieves the true state from the observations– to produce a point estimate of the true state, denoted by $\hat{\mathbf{X}}$ in the figure. In OE, $\hat{\mathbf{X}}$ can be interpreted as a conditional mean, and is accompanied by an estimate of the covariance matrix obtained via a linear approximation to the forward function.

The retrieval algorithm depends on a forward model, F_1, and its corresponding forward model parameters, $\mathbf{b}_1$, which are the best known practical approximations to F_0 and $\mathbf{b}_0$, respectively. Arguably, F_1 and $\mathbf{b}_1$ may be considered distinct, with F_1 being implemented as algorithm computer code and $\mathbf{b}_1$ being a set of fixed, ancillary inputs provided to F_1 in addition to the radiances. The ellipses in the arguments to R in Figure 1 represent other required choices that must be made in order to run the retrieval algorithm code, and will affect the quality of the estimates. We call these "settings". Examples include convergence criteria, the grid over which the algorithm will solve for the required optimum, etc. For compactness, we subsume settings into $\mathbf{b}_1$. Despite their potential separation in implementation, we consider $F_1(\cdot, \mathbf{b}_1)$ to be a unit, as is the case with $F_0(\cdot, \mathbf{b}_0)$.

From an uncertainty quantification perspective, one might view $\mathbf{Y}$ as the primary input to a deterministic function R, and $\hat{\mathbf{X}}$ as the output for which uncertainty is to be quantified. In that case, we equate $\hat{\mathbf{X}}$ with some measure of location of $P(\mathbf{X}|\mathbf{Y})$. Alternatively, one might view $\mathbf{X}$, though not directly observed, as the input to the composite system that includes nature, the instrument, and the retrieval process. From this point of view, we seek the conditional distribution $P(\mathbf{X}|\hat{\mathbf{X}})$. This distribution reflects the uncertainty about $\mathbf{X}$ that remains after seeing $\hat{\mathbf{X}}$.

The retrieval community takes the first perspective: F_1 is fixed, and $\mathbf{b}_1$ is set by dress testing of different candidate values according to knowledge of the underlying physics. Comparisons of resulting values of $\hat{\mathbf{X}}$ to ground truth, where available, dictate the final fixed value of $\mathbf{b}_1$. The only source of uncertainty accounted for is that of the input radiance vector, $\mathbf{Y}$. By treating $F_1(\cdot, \mathbf{b}_1)$ as fixed, this procedure ignores the impact of uncertainties that may be induced by their misspecification. [19] showed that misspecifications of this sort can lead to both bias and variance in the retrieved quantities through unpredictable interactions and algorithm artifacts.

Our perspective is that the uncertainty to be quantified is that of the entire, end-to-end observing system shown in Figure 1, and so our goal is to provide the conditional distribution $P(\mathbf{X}|\hat{\mathbf{X}})$. For each $\hat{\mathbf{X}}$ we approximate $P(\mathbf{X}|\hat{\mathbf{X}})$ via a Gaussian mixture model (GMM) in which the component-wise means, variances, and mixing probabilities are functions of the realized value of $\hat{\mathbf{X}}$. Those functions are estimated from a simulation experiment that 1) incorporates a model discrepancy term to account for structural and parametric model uncertainty, and 2) borrows strength over a representative ensemble of synthetic state vectors to account for the range of conditions to which the model must apply. The simulation experiment allows us to quantify the aggregate impact of uncertainties due to both known and unknown sources because we know the "truth". Once the parameters of the GMM are estimated, the conditional distributions of the true states given actual retrieved estimates are obtained by plugging the retrieved values into the estimated regression equations. This method is applicable regardless of whether $\hat{\mathbf{X}}$ is a least squares, OE, or any other type of estimate.

The approach requires that we simulate a realistic ensemble of synthetic true states and generate corresponding ensembles of radiances and retrieved state estimates. The simulated true state ensemble need not be identical to nature's true ensemble, just realistic in the sense that it spans the range of plausible true states the observing system is likely to encounter. Procedures for creating this ensemble will vary by application and even by analyst since "plausible" is subjective. Likewise, the forward propagation of the ensemble through $F_0(\cdot, \mathbf{b}_0)$ in the simulation, is application and analyst-specific since nature's true forward function is unknown, but must be represented in some way and be distinct from $F_1(\cdot, \mathbf{b}_1)$. In Section 4, we demonstrate our solutions to these problems for OCO-2.

We stress that a major restriction on our work for OCO-2 and for other existing missions is that UQ must be performed *post hoc*. The design and implementation of the operational retrieval algorithm cannot be changed nor can it be interfered with. Uncertainty quantification must be after the fact. Consequently, our goal here is only to obtain an honest estimate of total uncertainty, not to break it down into contributing factors or to reduce it.

Our framework uses the following statistical model based on Figure 1.

$$\mathbf{X} \sim P_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}_{\mathbf{X}}), \quad \mathbf{Y}_0 = F_0(\mathbf{X}, \mathbf{b}_0), \quad \mathbf{Y} = \mathbf{Y}_0 + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}),$$

$$\boldsymbol{\theta}_{\mathbf{X}} = \{\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}}, \ldots\}, \quad \text{and} \quad \hat{\mathbf{X}}(\mathbf{Y}, \mathbf{b}_1) = R(\mathbf{Y}, F_1, \mathbf{b}_1). \tag{2.1}$$

All variables in bold are column vectors or matrices, except $\boldsymbol{\theta}_{\mathbf{X}}$ which denotes a set of parameters. The dimensions of $\mathbf{X}$, $\boldsymbol{\mu}_{\mathbf{X}}$ and $\hat{\mathbf{X}}$ are $L_{\mathbf{X}} \times 1$. The dimensions of $\mathbf{Y}_0$, $\mathbf{Y}$, $\boldsymbol{\epsilon}$, and the zero vector are $L_{\mathbf{Y}} \times 1$. The matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ is of dimension $L_{\mathbf{X}} \times L_{\mathbf{X}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ and is $L_{\mathbf{Y}} \times L_{\mathbf{Y}}$. The vector $\mathbf{b}_0$ contains both known and unknown quantities set by nature, and has unspecified length. Finally, $\mathbf{b}_1$ is a fixed, known column vector containing all parameters necessary to run the retrieval.

This model prescribes that nature draws a true state vector from $P_{\mathbf{X}}$ having parameter vector $\boldsymbol{\theta}_{\mathbf{X}}$. From the standpoint of the retrieval algorithm, the true state, its distribution, and the measurement error $\boldsymbol{\epsilon}$, are all unknown. We will, however, assume that the statistics of $\boldsymbol{\epsilon}$ are known, as would be the case from pre-launch calibration studies. Radiance measurement errors are assumed to be multivariate normal with zero mean and known covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$. The only input to R subject to randomness in Equation (2.1) is therefore $\mathbf{Y}$. All other quantities on which R depends are fixed at nominal values using expert judgement and modified, if necessary, to conform to computational requirements.

3. Approach and methods.

The foundation of our approach is the statistical model given by Equation (2.1), overlaid with our view that the computational machinery of the observing system is a complex estimator, and its performance is summarized by $P(\mathbf{X}|\hat{\mathbf{X}})$. This conditional distribution may be derived from the joint distribution $P(\mathbf{X}, \hat{\mathbf{X}})$, which contains all the information about the uncertainty of $\hat{\mathbf{X}}$ as an estimate of $\mathbf{X}$ [11, 47]. Since we do not know the true joint distribution, we appeal to resampling to provide a synthetic ensemble that is our best empirical representation of it.

3.1. Methodology.

Our general strategy is summarized in Figure 2. The first row of this flowchart roughly mirrors Figure 1, but with a number of important modifications. First, $\tilde{P}$ plays the role of nature. An ensemble of simulated true states is drawn from it. This set of M simulated state vectors is the $(M \times L_{\mathbf{X}})$ data matrix, $\{\mathbf{X}_m^{\text{sim}}\}_{m=1}^{M} \equiv (\mathbf{X}_1^{\text{sim}}, \ldots, \mathbf{X}_M^{\text{sim}})'$, and we call it the synthetic true state ensemble. It represents a set of plausible, alternative realizations of the state vector that the observing system is likely to encounter. We will usually suppress the indices on the bracket notation for brevity.

Second, Figure 1 shows that the state vector is converted by nature's true forward function, $F_0(\cdot, \mathbf{b}_0)$, into a noiseless radiance, $\mathbf{Y}_0$, which is then observed by the instrument with measurement error $\boldsymbol{\epsilon}$. Figure 2 shows the transformation by the forward *model*, $F_1(\cdot, \mathbf{b}_1)$ (highlighted in orange for emphasis). This is because in practice we do not know nature's true forward function; all we have is our best forward model which will also be used in the retrieval process in the second row of Figure 2.

It is overly optimistic to assume that the forward model used in the retrieval is identical to nature's true forward function. To compensate, we add an extra component of noise to the synthetic, noiseless radiance ensemble $\{\mathbf{Y}_0^{\text{sim}}\}$, as shown in the right-most box on the first row of Figure 2. This model discrepancy term is $\boldsymbol{\delta}^{\text{sim}}$ and is an independent draw from a multivariate Gaussian distribution with mean vector and covariance matrix that are estimated off-line from *spectral residuals* that can be produced by any retrieval algorithm, and by our simulation. Spectral residuals are the differences between the observed radiances and the radiances implied by the forward model, evaluated at the converged estimate of the state. Section 4.1.4 describes how we estimate the mean vector and covariance matrix of the distribution of $\boldsymbol{\delta}^{\text{sim}}$ from the ensembles of available spectral residuals.

The second row of Figure 2 shows how the simulated noisy radiance ensemble, $\{\mathbf{Y}^{\text{sim}}\}$, is input to the retrieval algorithm to produce the corresponding retrieved estimates. We form a synthetic training ensemble by pairing each simulated true state with its corresponding

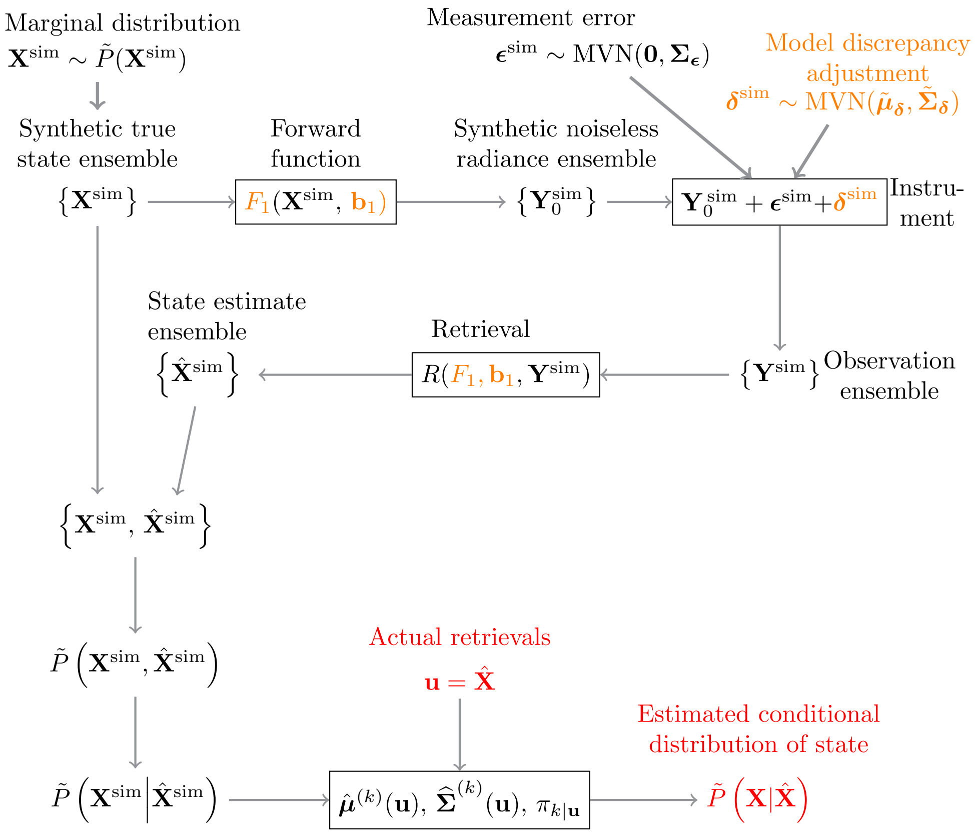
Marginal distribution
$\mathbf{X}^{\text{sim}} \sim \tilde{P}(\mathbf{X}^{\text{sim}})$
Measurement error
$\boldsymbol{\epsilon}^{\text{sim}} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$
Model discrepancy adjustment
$\boldsymbol{\delta}^{\text{sim}} \sim \text{MVN}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{\delta}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}})$
Synthetic true state ensemble
Forward function
Synthetic noiseless radiance ensemble
$\{\mathbf{X}^{\text{sim}}\}$
$F_1(\mathbf{X}^{\text{sim}}, \mathbf{b}_1)$
$\{\mathbf{Y}_0^{\text{sim}}\}$
$\mathbf{Y}_0^{\text{sim}} + \boldsymbol{\epsilon}^{\text{sim}} + \boldsymbol{\delta}^{\text{sim}}$
Instru-ment
State estimate ensemble
$\{\hat{\mathbf{X}}^{\text{sim}}\}$
Retrieval
$R(F_1, \mathbf{b}_1, \mathbf{Y}^{\text{sim}})$
$\{\mathbf{Y}^{\text{sim}}\}$
Observation ensemble
$\{\mathbf{X}^{\text{sim}}, \hat{\mathbf{X}}^{\text{sim}}\}$
$\tilde{P}\left(\mathbf{X}^{\text{sim}}, \hat{\mathbf{X}}^{\text{sim}}\right)$
Actual retrievals
$\mathbf{u} = \hat{\mathbf{X}}$
Estimated conditional distribution of state
$\tilde{P}\left(\mathbf{X}^{\text{sim}} \middle| \hat{\mathbf{X}}^{\text{sim}}\right)$
$\hat{\boldsymbol{\mu}}^{(k)}(\mathbf{u}),\ \widehat{\boldsymbol{\Sigma}}^{(k)}(\mathbf{u}),\ \pi_{k|\mathbf{u}}$
$\tilde{P}\left(\mathbf{X}|\hat{\mathbf{X}}\right)$

Figure 2. *Conceptual diagram of the uncertainty quantification strategy. F_1 and $\mathbf{b}_1$ are in orange to call attention to the fact that the retrieval uses the* same *forward function (and its parameters) as is assumed to be operative in nature. $\boldsymbol{\delta}^{sim}$ is also in orange because it adds extra uncertainty to the simulation in order to compensate. Red denotes the introduction of a single sounding's retrieved state vector in the role of a predictor after the parameters of $\tilde{P}$ are learned from simulated true state vectors and their retrieved counterparts.*

retrieval: $\left\{\mathbf{X}^{\text{sim}}, \hat{\mathbf{X}}^{\text{sim}}\right\}$. Then, we fit a Gaussian mixture model to this set, and subsequently derive 1) regression mean and variance functions for each mixture component, and 2) the conditional probabilities of component membership given the value of the predictors, $\hat{\mathbf{X}}$. Once these functions are estimated from the simulated ensemble, any new or operationally retrieved state estimate can serve as a predictor. We simply plug the predictor into the regression equations to obtain parameters of the Gaussian mixture component conditional distributions, and the conditional probability of component membership.

The next subsection briefly reviews Gaussian mixture models, and the software we use for fitting them.

3.2. Gaussian mixture models and software to fit them. The Gaussian mixture density for a multivariate random vector $\mathbf{V}$ is,

$$f_{\mathbf{V}}(\mathbf{v}) = \sum_{k=1}^{K} \pi_k \, \phi\left(\mathbf{v}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad \sum_{k=1}^{K} \pi_k = 1, \tag{3.1}$$

where $\phi\left(\mathbf{v}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$ is the multivariate normal density function with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$, evaluated at $\mathbf{v}$; π_k is the (mixing) weight of component k, and K is the total number of components [29]. We abbreviate this density by,

$$\mathbf{V} \sim \text{GMM}\left(K, \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k\}_{k=1}^{K}\right). \tag{3.2}$$

R's *densityMclust* function in the package mclust [46, 16] provides software for estimating the maximum likelihood estimates of the parameters in Equation (3.2). *densityMclust* also returns another object that is key for our purposes. It is an $N \times \widehat{K}$ matrix of conditional probabilities,

$$\hat{\pi}_{k|\mathbf{v}_n} = \tilde{P}(\kappa_n = k | \mathbf{V} = \mathbf{v}_n) = \frac{\hat{\pi}_k \, \phi\left(\mathbf{v}_n; \hat{\boldsymbol{\mu}}_k, \widehat{\boldsymbol{\Sigma}}_k\right)}{\sum_{l=1}^{\widehat{K}} \hat{\pi}_l \, \phi\left(\mathbf{v}_n; \hat{\boldsymbol{\mu}}_l, \widehat{\boldsymbol{\Sigma}}_l\right)}, \tag{3.3}$$

for $k = 1, \ldots, \widehat{K}, n = 1, \ldots, N$. Here, κ_n is a random variable that indicates component membership after realizing $\mathbf{V}_n = \mathbf{v}_n$ [47].

Equation (3.3) provides a probabilistic mapping of the $\mathbf{V}_n$ to the $\widehat{K}$ Gaussian components of the model. This is critical if, say, one fits the model to standardized data but wants to report the result on the raw scale. For example, suppose $\mathbf{V}_n$ is partitioned into two sub-vectors that are measured on very different scales so one would influence the calculation disproportionately if a model was fitted to the raw data. Suppose $\mathbf{V}$ is $d_{\mathbf{V}}$-dimensional, and let $\mathbf{V}_n = (\mathbf{W}_n', \mathbf{U}_n')'$, where $\mathbf{W}$ is $d_{\mathbf{W}}$-dimensional, and $\mathbf{U}$ is $d_{\mathbf{U}}$-dimensional, and $d_{\mathbf{V}} = d_{\mathbf{W}} + d_{\mathbf{U}}$. It would make sense to fit the model to standardized versions of these variables by converting $\mathbf{W}_n$ to $\mathbf{Z}_{1n}$ and $\mathbf{U}_n$ to $\mathbf{Z}_{2n}$:

$$\mathbf{Z}_{1n} = (\mathbf{W}_n - \mathbf{m}_{\mathbf{W}})' \, \mathbf{C}_{\mathbf{W}}^{-1/2} \quad \text{and} \quad \mathbf{Z}_{2n} = (\mathbf{U}_n - \mathbf{m}_{\mathbf{U}})' \, \mathbf{C}_{\mathbf{U}}^{-1/2}, \tag{3.4}$$

where $\mathbf{m}_{\mathbf{W}}$ and $\mathbf{C}_{\mathbf{W}}$ are the mean vector and covariance matrix of $\mathbf{W}_1, \ldots, \mathbf{W}_N$, and $\mathbf{m}_{\mathbf{U}}$ and $\mathbf{C}_{\mathbf{U}}$ are defined similarly. Denote the standardized vector $\utilde{\mathbf{V}}_n = (\mathbf{Z}_{1n}', \mathbf{Z}_{2n}')'$, and fit a GMM to $\utilde{\mathbf{V}}_1, \utilde{\mathbf{V}}_2, \ldots, \utilde{\mathbf{V}}_N$. Estimates of the component mean vectors and covariance matrices are on the standard scale ($\utilde{\hat{\boldsymbol{\mu}}}_k$ and $\utilde{\widehat{\boldsymbol{\Sigma}}}_k$), but can easily be recomputed from the raw data by calculating weighted averages and variances using the $\hat{\pi}_{k|\mathbf{v}_n}$ as weights:

$$\hat{\boldsymbol{\mu}}_k = \sum_{n=1}^{N} \mathbf{v}_n \left[\frac{\hat{\pi}_{k|\mathbf{v}_n}}{\sum_{m=1}^{N} \hat{\pi}_{k|\mathbf{v}_m}}\right], \tag{3.5}$$

$$\widehat{\boldsymbol{\Sigma}}_k = \sum_{n=1}^{N} \left(\mathbf{v}_n - \hat{\boldsymbol{\mu}}_k\right) \left(\mathbf{v}_n - \hat{\boldsymbol{\mu}}_k\right)' \left[\frac{\hat{\pi}_{k|\mathbf{v}_n}}{\sum_{m=1}^{N} \hat{\pi}_{k|\mathbf{v}_m}}\right]. \tag{3.6}$$

In practice we will use R's *weighted.mean* function to compute $\hat{\boldsymbol{\mu}}_k$, and the *cov.shrink* function from the package corpcor [44] to compute shrinkage estimates of $\boldsymbol{\Sigma}_k$.

The idea can be extended to other convenient transformations beyond simple standardization. In particular, if the dimension of $\mathbf{V}$ is large, *densityMclust* can be slow. In that case, $\underset{\sim}{\mathbf{V}}_1, \underset{\sim}{\mathbf{V}}_2, \ldots, \underset{\sim}{\mathbf{V}}_N$ may be further transformed by projecting them into the space spanned by the leading principal components of $\mathbf{V}$ estimated from the (standardized) data. Let $\mathbf{C}_{\underset{\sim}{\mathbf{V}}}$ be the (empirical) covariance matrix of $\underset{\sim}{\mathbf{V}}_1, \underset{\sim}{\mathbf{V}}_2, \ldots, \underset{\sim}{\mathbf{V}}_N$, and let $\mathbf{E}_i$ be the i-th eigenvector of $\mathbf{C}_{\underset{\sim}{\mathbf{V}}}$. The eigenvectors are arranged in order corresponding to the order of their descending eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_{d_{\mathbf{V}}}$. The leading eigenvector matrix of $\mathbf{C}_{\underset{\sim}{\mathbf{V}}}$ is $\mathbf{E}$, with columns $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_l$, where l is the smallest value such that

$$\frac{\sum_{i=1}^{l} \lambda_i}{\sum_{j=1}^{d_{\mathbf{V}}} \lambda_j} \geq \gamma, \quad 0 \leq \gamma \leq 1. \tag{3.7}$$

Finally, set

$$\underset{\sim}{\mathbf{V}} = \underset{\sim}{\mathbf{V}}\,\mathbf{E}. \tag{3.8}$$

The parameters of the model fit to $\underset{\sim}{\mathbf{V}}_1, \underset{\sim}{\mathbf{V}}_2, \ldots, \underset{\sim}{\mathbf{V}}_N$ can be rescaled using Equations (3.5) and (3.6).

The estimated joint distribution of $\mathbf{V} = (\mathbf{W}', \mathbf{U}')'$ of the form in Equation (3.1), leads directly to the estimated conditional means and covariances of $\mathbf{W}$ given $\mathbf{U} = \mathbf{u}_n$ by component:

$$\hat{\boldsymbol{\mu}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}_n) = \hat{\boldsymbol{\mu}}^{(k)}_{\mathbf{W}} + \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WU}} \left(\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UU}}\right)^{-1} \left[\mathbf{u}_n - \hat{\boldsymbol{\mu}}^{(k)}_{\mathbf{U}}\right], \tag{3.9}$$

with

$$\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{V}} = \left[\begin{array}{c|c} \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WW}} & \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WU}} \\ \hline \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UW}} & \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UU}} \end{array}\right], \tag{3.10}$$

and

$$\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}_n) = \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WW}} - \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WU}} \left(\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UU}}\right)^{-1} \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UW}}. \tag{3.11}$$

3.3. Conditional distribution of the true state given a retrieved state.

The conditional distribution of $\mathbf{W}$ for a new draw from the distribution of $\mathbf{U}$ with realization $\mathbf{u}^*$, is

$$\mathbf{W}^* \sim \mathrm{GMM}\left(\widehat{K},\ \left\{\hat{\boldsymbol{\mu}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}^*),\ \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}^*),\ \hat{\pi}_{k|\mathbf{u}^*}\right\}_{k=1}^{\widehat{K}}\right). \tag{3.12}$$

One can explore this distribution by simulating from it, and summarize it (approximately) by calculating any desired summary quantities from the ensemble of simulated realizations.

3.4. A simple example. To illustrate the main aspects of our procedure, we appeal to a simple, low-dimensional example shown in Figure 3. The left panel shows a scatterplot of two (scalar) variables that exhibit a relationship (Equation (3.13)) that could be representative of a realistic, worst-case scenario. Not only is it non-linear, but the variances are heteroskedastic.

$$W \sim N(5,1), \quad U = (1.75)^{W} + \epsilon, \quad \epsilon \sim N(1,2). \tag{3.13}$$

There are a total of $M = 5000$ (u_m, w_m), pairs in the plot. The top-left panel shows the *forward relationship* between the two variables, with w analogous to realizations of $\mathbf{W}$, and u analogous to realizations of $\mathbf{U}$ in the subsections above. The top-right panel shows the inverse problem in which the goal is to infer w from noisy and potentially biased u. The bottom-left panel shows the estimated joint distribution of u and w fitted to these data using mclust. The values of both variables are standardized prior to fitting the model, and w is transformed back to its original scale. The joint distribution is a six-component Gaussian mixture, shown superimposed on the scatterplot, as six sets of Gaussian contours (the .68 and .95 contours). The corresponding component-specific regression lines for predicting w from u are shown in the bottom-right panel. For the k-th mixture component, the conditional mean, variance, and component membership probabilities of W given $U = u$ are scalar versions of Equations (3.9), (3.10), and (3.11).

Now suppose a new value of $U = u^*$, is acquired and we wish to obtain the conditional distribution of W given $U = u^*$. Equipped with Equations (3.3) and (3.9) through (3.11), we first simulate B *iid* draws from the discrete distribution that places probability $\pi_{k|u^*}$ on the indices $k = 1, 2, 3, 4, 5, 6$, and encode these outcomes in the random vector $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_B)'$ of length B. Then, for each element of $\boldsymbol{\kappa}$ we draw a random realization, w_b^*, from the $N\left(\mu^{(\kappa_b)}(u^*), \sigma^{(\kappa_b)}(u^*)\right)$, where μ is the scalar version of $\boldsymbol{\mu}$, and σ is the scalar version of $\boldsymbol{\Sigma}$. The histogram of w_b^*, $b = 1, 2, \ldots, B$ is an approximation of the conditional distribution of W given $U = u^*$. Figure 4 shows the simulated conditional distributions for two values of u^*, 13.76 and 22.41. Visual inspection suggests that the conditional standard deviations should decrease as one moves towards higher values of u^*, which they do.

Figure 5 displays results of a cross-validation experiment that compares nominal and actual coverage probabilities for 50 percent and 95 percent confidence intervals. In this experiment, we use the same 5000 (u, w) pairs shown in the top panels of Figure 3. We randomly divide these into a training set of 2500 (u, w) pairs, and a test set of 2500 (u^*, w^*) pairs. We fit a GMM to the training set, plug each value of u^* in the test set into the Equations (3.3) and (3.9) through (3.11), and simulate 1000 draws from the posterior distributions of W^* given U^*. Finally, we compute the 0.025, 0.250, 0.750, and 0.975 quantiles of each of these empirical conditional distributions ($Q_{.025}$, $Q_{.25}$, $Q_{.75}$, and $Q_{.975}$) and determine what proportion of the 2500 test set pairs have the property that w^* lies inside the 95 percent interval defined by

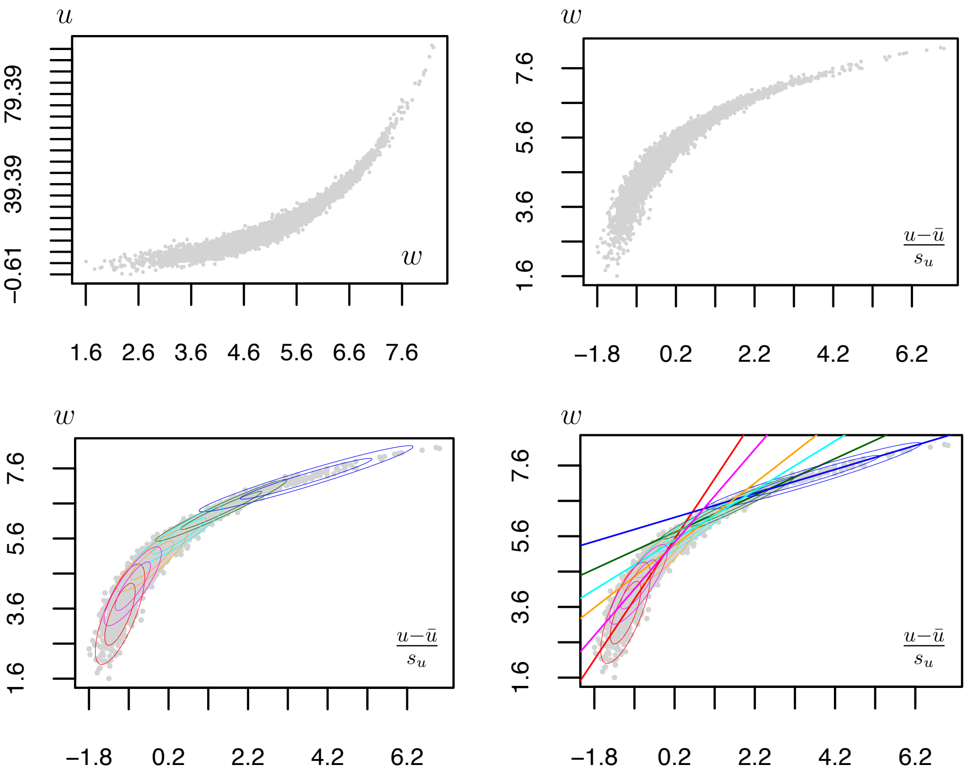

Figure 3. *Simple example of fitting a GMM to an empirical ensemble, $\{u_m, w_m\}_{m=1}^{5000}$. Top-left: scatterplot of u on w showing the forward relationship. Top-right: scatterplot of w on u showing the inverse relationship. Bottom-left: Six-component GMM fitted to the joint distribution of u and w. Two density contours capturing 68 and 95 percent of the central mass of the Gaussian components are shown. Bottom-right: component-wise regression lines for the model. Component-wise variance is the vertical dispersion around the regression line at a fixed point on the x-axis. In the latter three plots, the u values have been standardized by subtracting the sample mean, $\bar{u}$ and dividing by the sample standard deviation, s_u.*

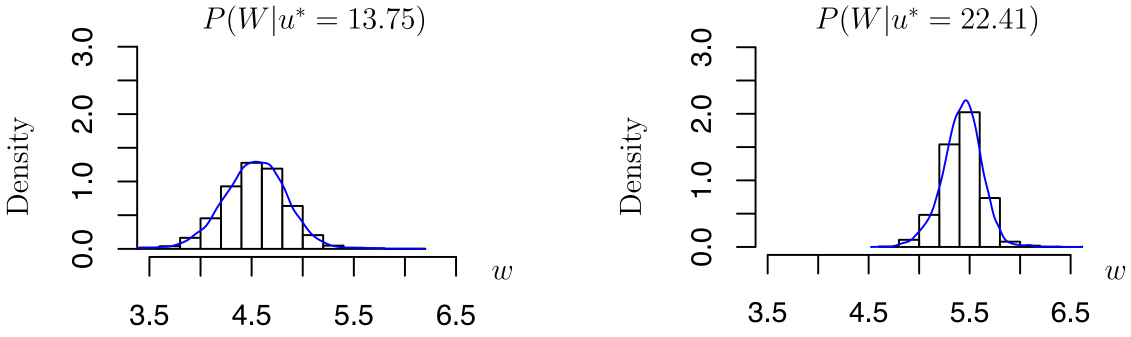

Figure 4. *Simulated conditional densities of W given $U^* = 13.76$ (left panel) and $U^* = 22.41$ (right panel). w^* is on the standardized scale.*

$[Q_{.025}, Q_{.975}]$, and similarly, inside the 50 percent interval defined by $[Q_{.25}, Q_{.75}]$.

$$p_{.95,b} = \frac{1}{2500} \sum_{i=1}^{2500} 1\left[Q_{.025} \leq w_i^* \leq Q_{.975}\right],$$

$$p_{.50,b} = \frac{1}{2500} \sum_{i=1}^{2500} 1\left[Q_{.25} \leq w_i^* \leq Q_{.75}\right]. \tag{3.14}$$

We carry out this entire simulation procedure 200 times to obtain $p_{.95,b}$ and $p_{.50,b}$, for $b = 1, 2, \ldots, B = 200$. Figure 5 shows that the actual coverage probabilities are always, or nearly always, consistent with the nominal coverage probabilities.

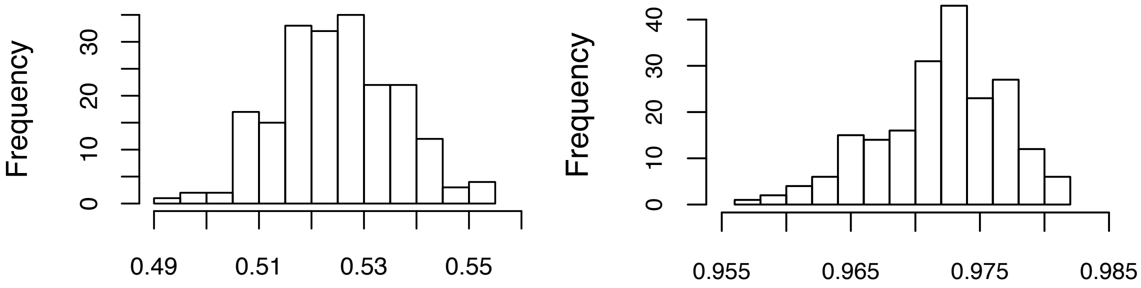

Figure 5. *Histograms of actual proportions of test set values of w contained within the central 50 percent (left) and 95 percent (right) of the estimated conditional distribution of W given U, over 200 trials of the simulation experiment. In each trail, a randomly selected half of the (u, w) pairs in Figure 3 were assigned to the training set, and other other half to the test set.*

In the next section we describe in detail how we implement our approach for our motivating application, the Orbiting Carbon Observatory-2 mission. It is considerably more complex than the simple example just presented, but nonetheless analogous in many respects. The main difference is that the predictor is high-dimensional, although the predictand remains a scalar. Other differences include the fact that we did not need to deal with model discrepancy in the simple example, but we do for OCO-2; and perhaps most importantly, how we set up the simulation experiment that allows us to learn the mechanistic properties of the retrieval estimator.

4. Application to OCO-2. NASA's OCO-2 instrument was launched into Earth orbit on July 2, 2014. Its primary scientific objective is to estimate total column concentrations (dry air mole-fractions) of carbon dioxide for use in estimating carbon fluxes between Earth's surface and atmosphere. Details of the OCO-2 mission and its retrievals can be found in [2, 14, 9, 13, 6]. OCO-2, along with Japan's GOSAT [43, 55] and GOSAT-2 missions, China's TanSAT [54], and OCO-3 [15] now form a fleet of observing systems that all use similar technology, including optimal estimation for retrievals. What we describe below for OCO-2 is also applicable to these and other missions potentially observing other variables, with suitable modifications.

The primary OCO-2 QOI is total column mole-fraction of CO_2, called XCO2, which we denote by τ. It is the number of molecules of carbon dioxide divided by the number of molecules of dry air (total air molecules minus water molecules), in a vertical column of the atmosphere. This quantity is derived on a sounding-by-sounding basis for cloud-free, trapezoidal ground footprints measuring 2.25 km along-track and 1.29 km across-track during the spacecraft's south-to-north polar orbit (see Figure 2-2 in [2]). The OCO-2 state vector, $\mathbf{X}$, includes a 20-element vertical profile of estimated CO_2 mole-fractions at various altitudes, as well as quantities describing aerosol, cloud, and surface properties. XCO2 is a scalar quantity computed by multiplying the CO_2 profile, $\mathbf{X}_{1:20}$, by a location-specific pressure weighting function, $\mathbf{h}$,

$$\tau = \mathbf{h}' \mathbf{X}_{1:20}. \tag{4.1}$$

In the next subsection, we provide details of how we applied our methodology described in Section 3 specifically for the OCO-2 case, and an evaluation of our results based on comparisons with the same ground-truth information used by the OCO-2 Valdiation Team in their activities.

4.1. Post hoc uncertainty quantification.

We built our model $\tilde{P}(\mathbf{X}^{\text{sim}}|\hat{\mathbf{X}}^{\text{sim}})$ as described in Section 3 by fitting to an empirical ensemble based on retrievals from the OCO-2 Version 7 data product [13] for week-long periods over four months spanning the four seasons: August and November 2015 and February and May 2016. This set of retrievals was obtained by OCO-2 in what is known as land-nadir mode– the ordinary operating mode of the instrument over land. Then, we plugged in actual retrieved OCO-2 state vectors acquired in *target-mode* as predictors, $\hat{\mathbf{X}}$. (See $\mathbf{u}$ in red in Figure 2.)

Target-mode is an alternative to land-nadir. In land-nadir mode, the observing mechanism is pointed straight down (nadir) and acquires eight soundings across-track at one time as the spacecraft advances over the course of a few milliseconds. The next eight soundings are geographically north of, and both disjoint from and contiguous with the previous eight, and so on. In target-mode, the instrument rotates its angle of view to repeatedly look at the same small area (roughly 0.2 degrees in both longitude and latitude) as the spacecraft passes overhead. This is typically done over locations where corroborating ground-truth is available. The purpose of target-mode is to provide large numbers of soundings nearly coincident with ground-truth information for validation. Further details can be found in [53].

Figure 6 is a modified version of Figure 2 that shows our implementation for OCO-2. The main steps are described in the following sub-sections. Green section numbers in parentheses in the figure direct the reader to relevant sections of the text.

4.1.1. Generating synthetic true state ensembles.

In Figure 6, the single marginal distribution in Figure 2 is replaced by a set of distributions, $\tilde{P}_r$, $r = \text{L1}, \ldots, \text{L11}$. The index r refers to sub-regions of the globe we call templates. Figure 7 shows the geographic domains of 41 such regions defined by colleagues in the user community who specialize in flux inversion. We asked these domain experts to define areas over which we could expect the behavior of OCO-2 state vectors, over a single calendar-week, to be representative of specific underlying physical processes generating them. This allows us to invoke the standard assumption of ergodicity in time and space. We carried out these simulations for all land and ocean regions with sufficient OCO-2 data. However, here we present results for land regions only because this is where corroborating ground-based information is available.

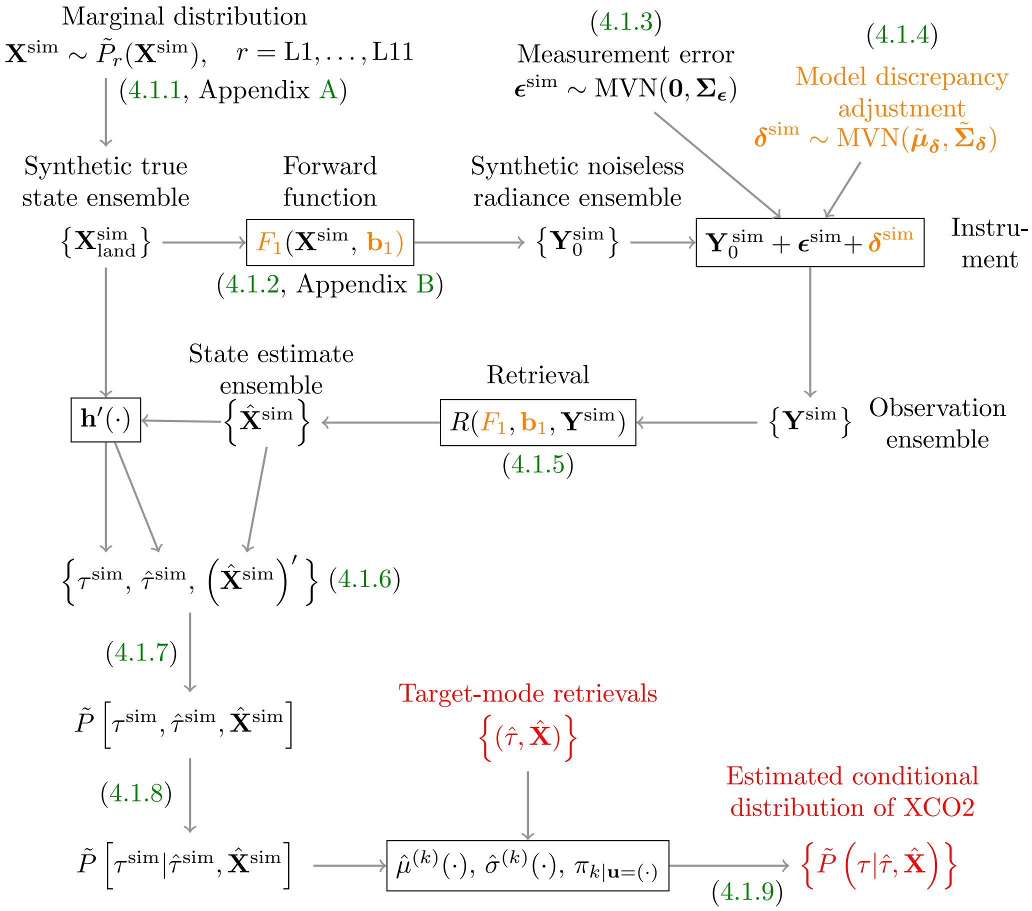
Marginal distribution
$\mathbf{X}^{\rm sim} \sim \tilde{P}_r(\mathbf{X}^{\rm sim}), \quad r = \text{L1}, \dots, \text{L11}$
(4.1.1, Appendix A)
(4.1.3)
Measurement error
$\boldsymbol{\epsilon}^{\rm sim} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon})$
(4.1.4)
Model discrepancy adjustment
$\boldsymbol{\delta}^{\rm sim} \sim \text{MVN}(\tilde{\boldsymbol{\mu}}_{\delta}, \tilde{\boldsymbol{\Sigma}}_{\delta})$
Synthetic true state ensemble
$\{\mathbf{X}^{\rm sim}_{\rm land}\}$
Forward function
$F_1(\mathbf{X}^{\rm sim}, \mathbf{b}_1)$
(4.1.2, Appendix B)
Synthetic noiseless radiance ensemble
$\{\mathbf{Y}^{\rm sim}_0\}$
$\mathbf{Y}^{\rm sim}_0 + \boldsymbol{\epsilon}^{\rm sim} + \boldsymbol{\delta}^{\rm sim}$
Instrument
State estimate ensemble
$\{\hat{\mathbf{X}}^{\rm sim}\}$
$\mathbf{h}'(\cdot)$
Retrieval
$R(F_1, \mathbf{b}_1, \mathbf{Y}^{\rm sim})$
(4.1.5)
$\{\mathbf{Y}^{\rm sim}\}$
Observation ensemble
$\{\tau^{\rm sim}, \hat{\tau}^{\rm sim}, (\hat{\mathbf{X}}^{\rm sim})'\}$ (4.1.6)
(4.1.7)
$\tilde{P}[\tau^{\rm sim}, \hat{\tau}^{\rm sim}, \hat{\mathbf{X}}^{\rm sim}]$
(4.1.8)
$\tilde{P}[\tau^{\rm sim} | \hat{\tau}^{\rm sim}, \hat{\mathbf{X}}^{\rm sim}]$
Target-mode retrievals
$\{(\hat{\tau}, \hat{\mathbf{X}})\}$
$\hat{\mu}^{(k)}(\cdot), \hat{\sigma}^{(k)}(\cdot), \pi_{k|\mathbf{u}=(\cdot)}$
(4.1.9)
Estimated conditional distribution of XCO2
$\{\tilde{P}(\tau | \hat{\tau}, \hat{\mathbf{X}})\}$

Figure 6. *Conceptual diagram of the uncertainty quantification strategy for OCO-2's OE retrieval. Orange and red quantities are explained in the caption to Figure 2. Green numbers in parentheses indicate subsections containing detailed explanations.*

Next, we fitted multivariate Gaussian mixture models to collections of actual, retrieved OCO-2 state vectors belonging to the templates, and in a single calendar-week. We sampled 5000 times from each of these distributions to generate template-week-specific ensembles of synthetic true states. Sampling from the fitted distributions is a semi-parametric analog to bootstrap resampling typically used in the bootstrap bias correction. Finally, we combined the eleven synthetically created ensembles to form the synthetic true state ensemble, $\{\mathbf{X}^{\rm sim}\}$, for the week being processed:

$$\mathbf{X}^{\rm sim}_{\rm land} \equiv \begin{pmatrix} \mathbf{X}^{\rm sim}_{\rm L1} \\ \vdots \\ \mathbf{X}^{\rm sim}_{\rm L11} \end{pmatrix}. \tag{4.2}$$

Appendix A provides additional details of model fitting and resampling. The stacked matrices in Equation (4.2) are abbreviated by $\{\mathbf{X}^{\rm sim}\}$ in Figure 6.

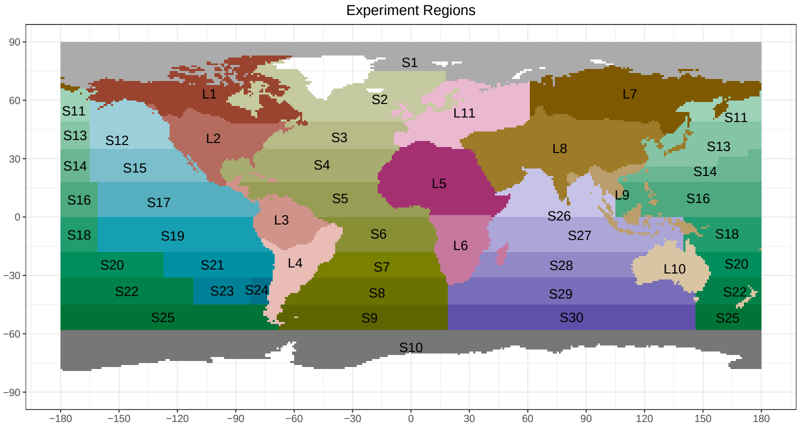

Figure 7. *Forty-one regions defining the spatial partitioning of OCO-2 data for template definition. There are 30 ocean regions with labels prefixed by "S", and 11 land regions with labels prefixed by "L".*

The reader may wonder whether generating synthetic true state vectors based on information gleaned from actual retrieved states could introduce unwanted artifacts in our ultimate uncertainty estimates. After all, the actual retrieved states are subject to the very uncertainties we seek to quantify. However, our focus is on the effects of the observing system's transformations of its inputs, not on the inherent variability of those inputs. We only require that the relationship between true state and its estimator in the simulation experiment be representative of that relationship in reality. We do not require that the synthetic true state ensemble mirror the actual true state distribution.

4.1.2. The forward function and its parameters. We applied the OCO-2 forward *model*, F_1, to each simulated true state vector in $\mathbf{X}_{\text{land}}^{\text{sim}}$. F_1 is the same forward model used operationally by OCO-2 in Version 7 of its Atmospheric Carbon Observations from Space (ACOS) retrieval algorithm [13, 34]. The ACOS forward model is often termed "full-physics" (FP). Further details are provided in Appendix B and [2].

4.1.3. Measurement error. The simulation of the measurement error,

$$\boldsymbol{\epsilon}^{\text{sim}} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\epsilon}}),$$

follows the OCO-2 instrument noise specification [14]. The noise covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ is diagonal with elements

$$\sigma^2_{\epsilon,i} = \text{var}\left(\epsilon_i^{\text{sim}}\right) = \text{b}_{\epsilon,1,i}\text{Y}_{0,i}^{\text{sim}} + \text{b}_{\epsilon,2,i}, \tag{4.3}$$

where the $\text{b}_{\epsilon,1,i}$ and $\text{b}_{\epsilon,2,i}$ are instrument calibration parameters, and i indexes elements in the radiance vector. This model suggests the noise variance is proportional to the mean signal, with an additive offset.

4.1.4. Model discrepancy adjustment. In Figure 6 we used the same forward model and forward model parameters to produce both simulated radiances and to retrieve the state estimate from them (see Subsection 4.1.2). This is overly optimistic since it implies that the retrieval's forward model is a perfect representation of nature's true forward function. On the other hand, the true forward function $F_0(\cdot, \mathbf{b}_0)$ is not known, and the best available forward model is $F_1(\cdot, \mathbf{b}_1)$. To compensate for this, we added an additional component of random error to the radiance vectors which we believe realistically degrades the radiances to account for using $F_1(\cdot, \mathbf{b}_1)$ where we should have used $F_0(\cdot, \mathbf{b}_0)$. We degraded the radiance vector by adding this "model discrepancy" adjustment, $\boldsymbol{\delta}^{\text{sim}}$, at the same time we added $\boldsymbol{\epsilon}$ to mimic measurement error of the instrument.

The model discrepancy adjustment is an $(L_{\mathbf{Y}} \times 1)$-dimensional perturbation modeled as a draw from a multivariate Gaussian distribution with mean $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\delta}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}}$. The Gaussian choice is for convenience, and may be revisited in the future. We estimated model parameters by comparing distributions of the *spectral residuals* in the simulation with those produced by the actual OCO-2 retrieval process. Spectral residuals are the differences between the observed OCO-2 radiance vectors and the radiance vectors predicted by applying the OCO-2 forward model to the retrieved state vector estimate.

To estimate $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}}$ we proceed as follows. The definition of model discrepancy is the difference between true and modeled radiances due only to the difference between $F_0(\cdot, \mathbf{b}_0)$ and $F_1(\cdot, \mathbf{b}_1)$, both evaluated at the true state, $\mathbf{X}$,

$$\boldsymbol{\delta} = F_0(\mathbf{X}, \mathbf{b}_0) - F_1(\mathbf{X}, \mathbf{b}_1).$$

However, we only have access to $F_1(\hat{\mathbf{X}}, \mathbf{b}_1)$, which motivates the approximation,

$$\boldsymbol{\delta} = F_0(\mathbf{X}, \mathbf{b}_0) - F_1(\hat{\mathbf{X}}, \mathbf{b}_1) - \Big[F_1(\mathbf{X}, \mathbf{b}_1) - F_1(\hat{\mathbf{X}}, \mathbf{b}_1)\Big], \tag{4.4}$$

$$\approx F_0(\mathbf{X}, \mathbf{b}_0) - F_1(\hat{\mathbf{X}}, \mathbf{b}_1) - \Big[F_1(\mathbf{X}^{\text{sim}}, \mathbf{b}_1) - F_1(\hat{\mathbf{X}}^{\text{sim}}, \mathbf{b}_1)\Big], \tag{4.5}$$

where $\mathbf{X}^{\text{sim}}$ and $\hat{\mathbf{X}}^{\text{sim}}$ are notional simulated true and retrieved states for the same location, time, and conditions as those which characterize $\mathbf{X}$. Then the term in square brackets on the right in Equation (4.5) is a simulated approximation to the difference between $F_1(\mathbf{X}, \mathbf{b}_1)$ and $F_1(\hat{\mathbf{X}}, \mathbf{b}_1)$.

Let $\mathbf{Y} \equiv F_0(\mathbf{X}, \mathbf{b}_0) + \boldsymbol{\epsilon}$, and $\hat{\mathbf{Y}} \equiv F_1(\hat{\mathbf{X}}, \mathbf{b}_1)$. The quantity $\Big(\mathbf{Y} - \hat{\mathbf{Y}}\Big)$ is the spectral residual, and is routinely produced as part of the retrieval process. The term in the square brackets can be computed from a simulation with no model discrepancy, where the noiseless forward model evaluation is readily available. The discrepancy can be written,

$$\begin{aligned}
\boldsymbol{\delta}^{\text{sim}} &\approx \Big(\mathbf{Y} - \boldsymbol{\epsilon} - \hat{\mathbf{Y}}\Big) - \Big(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\Big), \\
\boldsymbol{\delta}^{\text{sim}} + \boldsymbol{\epsilon} &\approx \Big(\mathbf{Y} - \boldsymbol{\epsilon} - \hat{\mathbf{Y}} + \boldsymbol{\epsilon}\Big) - \Big(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\Big), \\
&= \Big(\mathbf{Y} - \hat{\mathbf{Y}}\Big) - \Big(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\Big). \qquad (4.6)
\end{aligned}$$

Therefore the sum of the discrepancy and noise can be approximated with the operational spectral residuals and simulation forward model evaluations. Taking the expectation and variance of $\boldsymbol{\delta}^{\text{sim}} + \boldsymbol{\epsilon}$ facilitates estimation of $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}}$.

$$\begin{aligned} \mathrm{E}(\boldsymbol{\delta}^{\text{sim}} + \boldsymbol{\epsilon}) &\approx \mathrm{E}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) - \mathrm{E}\left(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\right), \\ \tilde{\boldsymbol{\mu}}_{\boldsymbol{\delta}} + \mathbf{0} &\approx \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{Y}_n - \hat{\mathbf{Y}}_n\right) - \frac{1}{M}\sum_{m=1}^{M}\left(\mathbf{Y}_{0,m}^{\text{sim}} - \hat{\mathbf{Y}}_m^{\text{sim}}\right), \end{aligned} \tag{4.7}$$

where $n = 1, \dots, N$ indexes actual OCO-2 retrievals in the template-week being studied, and $m = 1, \dots, M$ indexes trials of the simulation for that template-week. For the variance, we assume that the discrepancy and measurement noise are uncorrelated. Further, the spectral residuals from the actual OCO-2 retrievals are independent of the forward model evaluations from the simulation, by construction. Then,

$$\begin{aligned} \mathrm{cov}(\boldsymbol{\delta}^{\text{sim}} + \boldsymbol{\epsilon}) &\approx \widehat{\mathrm{cov}}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \widehat{\mathrm{cov}}\left(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\right), \\ \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} &\approx \widehat{\mathrm{cov}}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \widehat{\mathrm{cov}}\left(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\right), \\ \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}} &\approx \widehat{\mathrm{cov}}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \widehat{\mathrm{cov}}\left(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\right) - \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}, \end{aligned} \tag{4.8}$$

where $\widehat{\mathrm{cov}}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right)$ and $\widehat{\mathrm{cov}}\left(\mathbf{Y}_0^{\text{sim}} - \hat{\mathbf{Y}}^{\text{sim}}\right)$ are suitable empirical estimates of the covariance matrices of the spectral residuals and simulation forward model evaluations, respectively.

We found that outliers are often present in the spectral residuals, so we implement an estimate that combines a rank correlation matrix with robust estimates of the standard deviations for individual wavelengths. In addition, there is no guarantee that the right-hand side of Equation (4.8) yields a positive definite matrix, particularly when the variability of the model discrepancy is similar to or smaller than the measurement noise. To remedy this, we simulate model discrepancy with a low-rank approximation by retaining only the leading principal components.

4.1.5. The retrieval. The OCO-2 operational retrieval was performed for each $\mathbf{Y}^{\text{sim}}$. The retrieval's assumed statistical parameters $\mathbf{b}_1$, including the prior mean vector and covariance matrix used in OE, were set to the operational values present at a reference sounding at the geographical center of the template to which $\mathbf{Y}^{\text{sim}}$ belonged. The operational retrieval performs a numerical search for the minimum of an objective function that includes the Gaussian negative log-likelihood of the radiances and a regularization term due to a Gaussian prior distribution specified as part of OE. This was implemented with a Levenberg-Marquardt algorithm with step sizes, relaxation, and convergence criteria defined for the operational retrieval [2]. The OCO-2 inverse problem is moderately non-linear [5], so these optimization parameters can impact the overall retrieval quality. We retained those retrievals $\hat{\mathbf{X}}^{\text{sim}}$ that successfully converged within the allowed number of iterations, and discarded those that did not.

4.1.6. Assembling the empirical joint distribution. We used the set of synthetic true states and their corresponding synthetic retrieved state vectors to form an empirical sample of the joint distribution of the two. The QOIs were XCO2 values, τ^{sim}, created from the synthetic true state vectors using Equation (4.1). The predictors were the corresponding synthetic retrieved state vectors, $\hat{\mathbf{X}}^{\text{sim}}$, with retrieved XCO2 derived from them: $\hat{\tau}^{\text{sim}} = \mathbf{h}'\mathbf{X}^{\text{sim}}_{1:20}$, appended. The training ensembles used to estimate the GMM parameters for the resampling step were,

$$\left\{\tau^{\text{sim}}_m, \hat{\tau}^{\text{sim}}_m, \left(\hat{\mathbf{X}}^{\text{sim}}_m\right)'\right\}_{m=1}^{M_{\text{land}}}, \tag{4.9}$$

where $M_{\text{land}} = \sum_{r=\text{L1}}^{\text{L11}} M_r$, and M_r was the number of successful simulated retrieved state vectors in template r. It may seem that including $\hat{\tau}^{\text{sim}}_m$ as a predictor duplicates the information already in $\hat{\mathbf{X}}^{\text{sim}}_m$. It does not: $\hat{\tau}^{\text{sim}}_m$ includes the pressure weighting function. Offline experiments strongly suggested that predictions of τ^{sim}_m improve substantially when $\hat{\tau}^{\text{sim}}_m$ is included as a predictor.

We made one further modification to the training ensembles before estimating the joint distributions of predictors and predictands: six elements of the state vector were removed because they were found to degrade the predictions of τ^{sim}. The deleted elements were coefficients describing the distributions of cloud ice and liquid water in the atmospheric column.

4.1.7. Fitting the Gaussian mixture model. To estimate the joint distribution of XCO2 and the predictors from the empirical ensembles in Equation (4.9), we followed the methodology described in Section 3.2, with

$$\mathbf{V} = \left(\tau^{\text{sim}}, \hat{\tau}^{\text{sim}}, \left(\hat{\mathbf{X}}^{\text{sim}}\right)'\right)'. \tag{4.10}$$

To increase speed in the density estimation step, we reduced the dimension of $\mathbf{V}$ by converting the variables that will play the role of predictors to their corresponding values in the space of their leading principal components. To be clear, we partitioned $\mathbf{V} = (\text{W}, \mathbf{U}')'$ where

$$\text{W} = \tau^{\text{sim}}, \quad \text{and} \quad \mathbf{U} = \left(\hat{\tau}^{\text{sim}}, \left(\hat{\mathbf{X}}^{\text{sim}}\right)'\right)'. \tag{4.11}$$

In determining leading eigenvectors, we used the threshold $\gamma = 0.99$ here. (See Equation (3.7).) We used *densityMclust* function (in the R package mclust) to fit a family of Gaussian mixture models. The software requires us to specify the maximum number of allowable components, and it uses the Bayesian Information Criterion to select the best model. We set the maximum to 20 on the grounds 1) the number should exceed the number used for estimating GMM's for individual template-weeks in the resampling stage (see Appendix A), and 2) the number should be small enough to be interpretable and to achieve moderate computational speed.

The last step was to use the datum-specific weights in Equation (3.3) to convert the model's estimated parameters back to the original scale of the data via the formulas in Equation (3.6). The values of the parameters $\widehat{K}^{\text{sim}}$ and $\hat{\pi}^{\text{sim}}_1, \hat{\pi}^{\text{sim}}_2, \ldots, \hat{\pi}^{\text{sim}}_{\widehat{K}^{\text{sim}}}$ remained unchanged. The estimated mixture density for land in a given calendar-week is,

$$\left(\tau^{\text{sim}}, \hat{\tau}^{\text{sim}}, \left(\hat{\mathbf{X}}^{\text{sim}}\right)'\right)' = (\text{W}, \mathbf{U})' \sim \text{GMM}\left(\widehat{K}^{\text{sim}}, \left\{\hat{\boldsymbol{\mu}}^{\text{sim}}_k, \widehat{\boldsymbol{\Sigma}}^{\text{sim}}_k, \hat{\pi}^{\text{sim}}_k\right\}_{k=1}^{\widehat{K}^{\text{sim}}}\right), \tag{4.12}$$

where

$$\hat{\boldsymbol{\mu}}_k^{\text{sim}} = \left(\hat{\mu}_{k\text{W}}^{\text{sim}}, (\hat{\boldsymbol{\mu}}_{k\mathbf{U}}^{\text{sim}})'\right)', \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_k^{\text{sim}} = \left(\begin{array}{c|c} \hat{\sigma}_{k\text{WW}}^{\text{sim}} & \widehat{\boldsymbol{\Sigma}}_{k\text{W}\mathbf{U}}^{\text{sim}} \\ \hline \widehat{\boldsymbol{\Sigma}}_{k\mathbf{U}\text{W}}^{\text{sim}} & \widehat{\boldsymbol{\Sigma}}_{k\mathbf{UU}}^{\text{sim}} \end{array}\right), \tag{4.13}$$

4.1.8. Estimating the conditional distribution of XCO2 given retrieval predictors. We use Equations (3.3) and (3.9) through (3.11) to compute the Gaussian components' conditional means and variances of XCO2, given actual OCO-2 retrievals, $\mathbf{u}^* = \left(\hat{\tau}, \hat{\mathbf{X}}'\right)$. Thus, for each actual OCO-2 sounding, $\mathbf{u}^*$, we have the conditional distribution,

$$\text{W}^* \sim \text{GMM}\left(\widehat{K}^{\text{sim}}, \left\{\hat{\mu}_{\text{W}|\mathbf{U}}^{(k)}(\mathbf{u}^*),\ \hat{\sigma}_{\text{W}|\mathbf{U}}^{(k)}(\mathbf{u}^*),\ \hat{\pi}_{k|\mathbf{u}^*}\right\}_{k=1}^{\widehat{K}^{\text{sim}}}\right). \tag{4.14}$$

4.1.9. Simulating from sounding-specific conditional distributions of XCO2. These distributions can be approximated by simulating from Equation (4.14), and summary statistics can be computed as needed. To simulate B realizations, W_b^*, for $b = 1, \ldots, B$, from the model in Equation (4.14):

1. Let κ_b be a univariate random variable taking values in the set $\{1, \ldots, \widehat{K}^{\text{sim}}\}$ with,

$$P(\kappa_b = k) = \hat{\pi}_{k|\mathbf{u}^*}, \quad k \in \{1, \ldots, \widehat{K}^{\text{sim}}\}.$$

2. Draw B random variables,

$$\text{W}_b^* \sim N\left(\hat{\mu}_{W|\mathbf{U}}^{(\kappa_b)}(\mathbf{u}^*), \hat{\sigma}_{W|\mathbf{U}}^{(\kappa_b)}(\mathbf{u}^*)\right), \quad b = 1, \ldots, B.$$

If desired, the marginal mean and variance functions can be obtained by integrating over the mixture components.

4.2. Comparison to ground station data. In this subsection we evaluate how well our method performs by comparing ground-truth total column CO_2 at locations and times where they exist, to our conditional distributions and to distributions implied by the OCO-2 operational retrieval output. We report uncertainty quantification results for target-mode soundings' estimates of XCO2 that are within 0.01 degrees latitude and longitude (about one kilometer at the Equator and less farther poleward) of ground-based TCCON sites. Since 0.01 degrees is smaller than the OCO-2 ground footprint, these target-mode soundings literally contain the TCCON sites. This is illustrated in the left panel of Figure 8. Call this set of soundings the *OCO-2 test set* for a given overpass, and denote the number of soundings in a test set by N. TCCON acquires high-frequency time series (e.g., once per second), but there are outage periods and random drop outs. We limited the TCCON data to those acquired between the earliest and latest OCO-2 test set sounding times, among those soundings satisfying the 0.01° spatial proximity criterion. This always resulted in just a single TCCON time

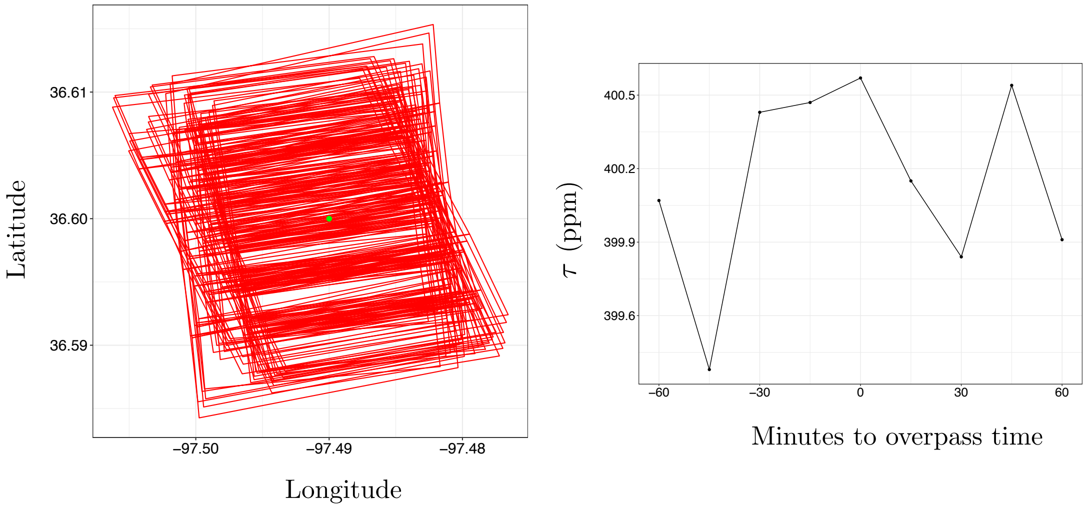

Figure 8. *Left: Spatial coincidence for a typical OCO-2 test set of target-mode footprints over the Lamont, OK TCCON site. Each red rectangle is an OCO-2 target-mode footprint. The green dot is the TCCON station. Right: Temporal coincidence with an OCO-2 test set for a typical TCCON time series at Lamont. The x-axis shows minutes before and after the OCO-2 overpass, and zero is the time of the overpass.*

point, which we call the *benchmark value* for the OCO-2 overpass. This is illustrated in the right panel of Figure 8.

For a given footprint in the OCO-2 test set, we quantified performance of the conditional distribution estimates with two simple metrics. The first is the position of the TCCON benchmark value within the conditional distribution,

$$\begin{aligned} G_{op}(\tau_T) &= P_{op}(\tau \leq \tau_T) \approx \Phi\left(\tau_T; \hat{\tau}, \hat{\mathrm{S}}\right), \\ G_s(\tau_T) &= P_s(\tau \leq \tau_T) \approx \frac{1}{B}\sum_{b=1}^{B} \mathcal{I}\left(\mathrm{W}_b^* \leq \tau_T\right), \end{aligned} \tag{4.15}$$

where *op* and *s* denote operational and simulated conditional distributions, respectively. The scalar quantity $\hat{\mathrm{S}}$ is the conditional variance of XCO2 given the observed radiance, as estimated by the retrieval algorithm through a linear approximation [2]. τ_T is the TCCON value, $\Phi(\tau; \mu, \sigma^2)$ is the value of the Gaussian cumulative distribution function with mean μ and variance σ^2 evaluated at τ. The variable B is the number of draws from the simulated distribution used to approximate it, and W_b^* is the b-th draw. Finally, $\mathcal{I}(\cdot)$ is the indicator function taking value one if its argument is true and zero otherwise. Good performance is achieved when the TCCON value is centrally located, and poor performance is indicated by a TCCON value far out in a tail. So values of $G_{op}(\tau_T)$ and $G_s(\tau_T)$ near 0.5 are judged superior to those that are close to zero or one.

The second metric is the bias of the mean of the conditional distribution relative to the

TCCON benchmark value:

$$\beta_{op}(\tau_T) = (\hat{\tau} - \tau_T) \quad \text{and} \quad \beta_s(\tau_T) = \left(\frac{1}{B} \sum_{b=1}^{B} \mathrm{W}_b^* \right) - \tau_T. \tag{4.16}$$

This is important because the primary use of OCO-2 XCO2 estimates is as input into flux inversion models. These models *assume* that XCO2 estimates are unbiased. If bias does nonetheless exist, it can lead to spurious flux estimates, particularly if there are systematic spatial patterns.

For illustration, consider four soundings' conditional distributions shown in Figure 9. These four are representative of the types of relationships between operational and simulation-based conditional distributions observed in our analysis. In the figure, the top two panels show two (of a total of $N = 132$) OCO-2 test set soundings covering the Lamont, OK TCCON site on November 2, 2015. The bottom two panels show two (of a total of $N = 77$) OCO-2 test set soundings covering the TCCON site at Tsukuba, Japan on May 13, 2016. In all panels, the red curves are the simulation-based conditional distributions' approximations based on $B = 100,000$ realizations of W_b^*, and the blue curves are Gaussian distributions with means and variances equal to the operationally retrieved moment estimates. The green vertical lines are the TCCON benchmark values for those overpasses. In both Lamont results (top row), the bias of the simulation-based estimate is smaller than that of the operational estimate. The left panel shows a case in which τ_T is more consistent (in the sense of being more centrally located) with the simulation-based distribution than with the operationally derived distribution. For Tsukuba (bottom row), the absolute value of the bias is actually lower for the operational estimate, and the biases are of different signs. The TCCON value is much more consistent with the simulation-based distribution than with the operational distribution in the left panel, but the TCCON value is consistent with neither distribution in the right panel.

In Figure 10 we summarize comparisons like those in Figure 9, for all 944 OCO-2 test set soundings in our analysis. The horizontal axis reflects the difference between the operational and simulation-based conditional distributions with respect to the centrality of the TCCON benchmark value in those distributions. The vertical axis is the difference between the absolute biases of the means of the two distributions, relative to TCCON. Points in the lower-left quadrant (colored blue) represent those soundings for which the simulation-based method is superior in both metrics. These comprise 63.1 percent of soundings used in this analysis. Points in the upper-left quadrant (colored green) represent those footprints for which the simulation-based method is superior with respect to the cumulative distribution function metric, but inferior with respect to the bias metric. These comprise 25.7 percent of soundings. For most of these, the difference in absolute bias is less than 1.25 parts per million. 11.2 percent of the footprints fall in the upper-right quadrant (colored red). In these cases the operational distribution is superior to the simulation-based distribution on both metrics. An example of such a case would be where the operational retrieval does very well, having mean very near the TCCON benchmark value and a low reported variance, while the simulation-based distribution has a long right (or left) tail. This will pull the mean of the simulation-based distribution away from the TCCON value, and cause the TCCON benchmark value to be less centrally located. However, given the tendency of operational variance estimates to be too low, such occurrences should be interpreted with caution.

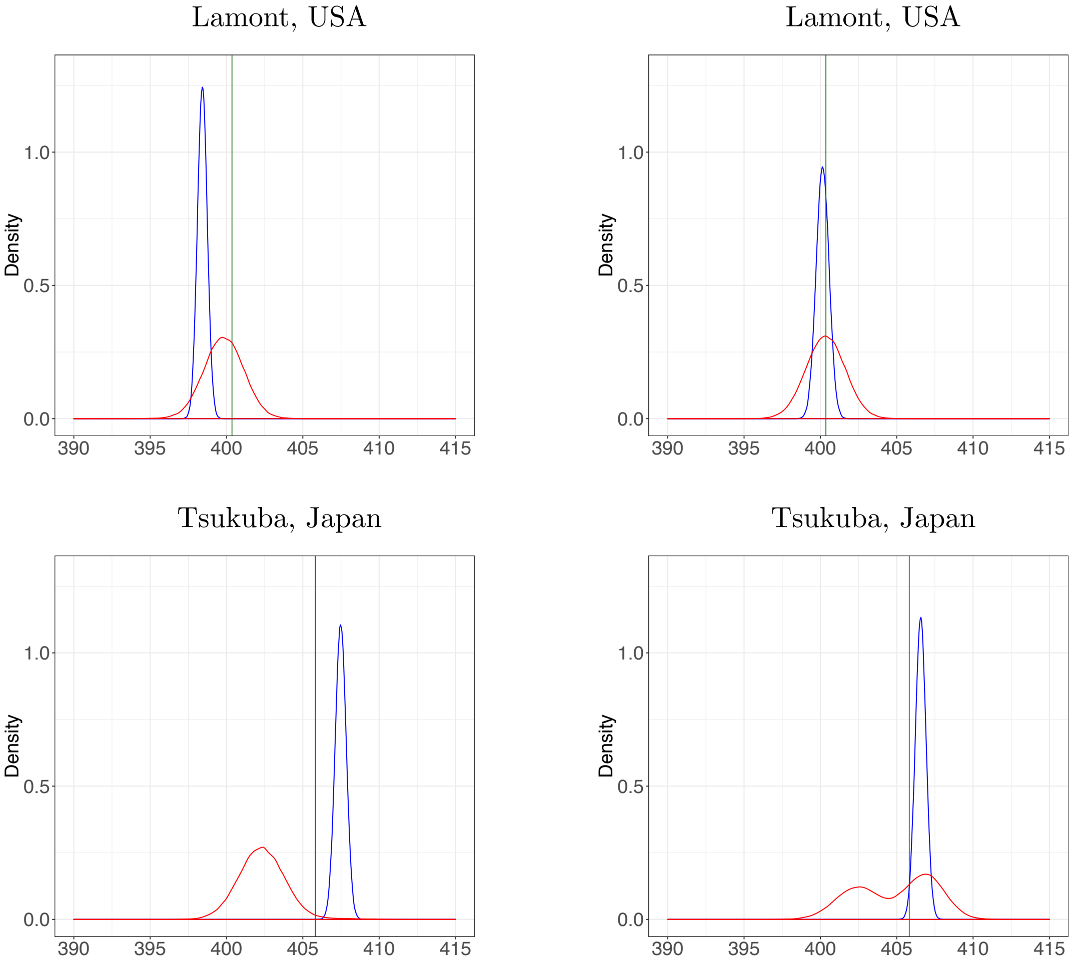

Figure 9. *Comparisons of typical operational (blue) and simulation-based (red) conditional distributions, to TCCON values (green). Top-left and top-right: Two soundings near Lamont, OK during the week of November 2, 2015. Bottom-left and bottom-right: Two soundings near Tsukuba, Japan during the week of May 13, 2016.*

As a final step in the evaluation of our methodology and comparison of it to that invoked by the operational retrieval algorithm, we examine coverage probabilities. The conditional distributions of OCO-2 test set soundings can be used to derive an ensemble of confidence intervals for the corresponding TCCON benchmark value. Approximate $(1 - \alpha)$ percent confidence intervals for TCCON XCO2 derived from the simulated and operationally-retrieved conditional distributions, respectively, are

$$\left[Q^s_{\alpha/2}\,,\,Q^s_{1-\alpha/2}\right] \quad \text{and} \quad \left[Q^{op}_{\alpha/2}\,,\,Q^{op}_{1-\alpha/2}\right], \tag{4.17}$$

where $Q_{\alpha/2}$ and $Q_{1-\alpha/2}$ are the $(\alpha/2)$ and $(1-(\alpha/2))$ empirical quantiles of the appropriate distributions.

For a given TCCON location and OCO-2 overpass, let N be the number of OCO-2 target-

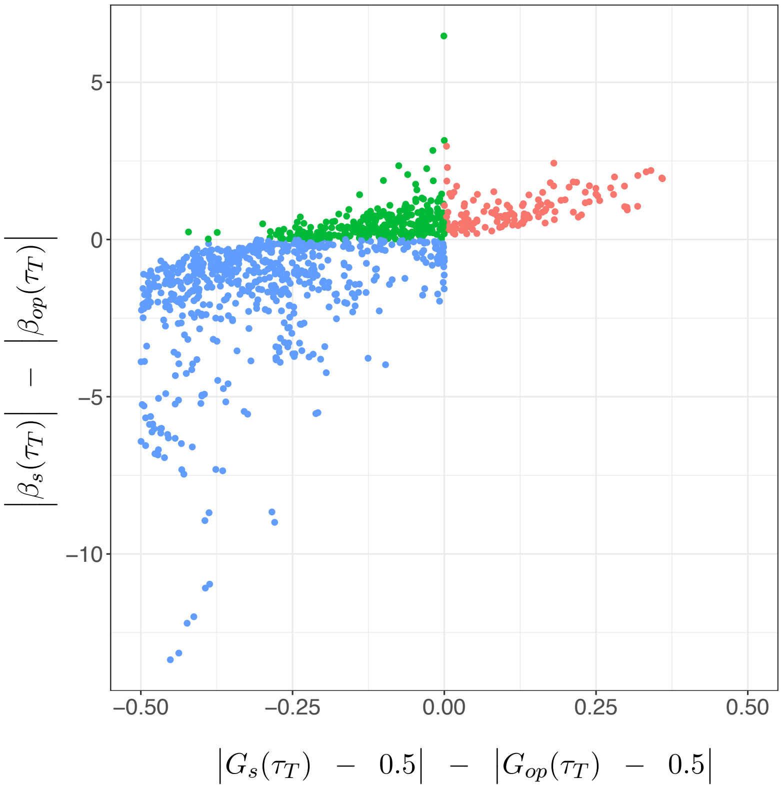

Figure 10. *Comparisons of operational and simulation-based conditional distribution estimates using two metrics. The horizontal axis is the difference in how centrally located the TCCON observations are with respect to the two distributions. The vertical axis is the difference in the distributions' means (in ppm) relative to the TCCON value. Each point represents a sounding that overlaps a TCCON site. The points' colors correspond to the quadrant of the space in which they lie.*

mode soundings in the test set, and let τ_T be the TCCON benchmark value they share. We computed,

$$p^s_{1-\alpha} = \frac{1}{N}\sum_{n=1}^{N} \mathcal{I}\left(\tau_T \in \left[Q^s_{\alpha/2}\,,\, Q^s_{1-\alpha/2}\right]\right), \text{ and} \tag{4.18}$$

$$p^{op}_{1-\alpha} = \frac{1}{N}\sum_{n=1}^{N} \mathcal{I}\left(\tau_T \in \left[Q^{op}_{\alpha/2}\,,\, Q^{op}_{1-\alpha/2}\right]\right), \tag{4.19}$$

for each overpass for $\alpha = 0.05$ and $\alpha = 0.50$. Results are displayed in Table 1.

None of the empirical coverage probabilities derived from the operational distributions agree with the nominal coverage. In contrast, the simulation-based coverages meet or exceed the nominal values for five of the eight overpasses. In addition, empirical coverage exceeds the nominal value for $p^s_{.50}$ at Orleans, and is close for Lauder at $p^s_{.95}$. Performance is poor for Sodankyla, which is at high latitude, and is a notoriously difficult site for retrievals in general. We plan further in-depth investigations to understand how different geographic conditions

Site	Overpass	$p^{op}_{.95}$	$p^{op}_{.50}$	$p^{s}_{.95}$	$p^{s}_{.50}$	N
Bialystok, Poland	2016-02-17	0.013	0.000	1.000	1.000	80
Darwin, Australia	2015-08-10	0.663	0.366	0.970	0.782	202
Lamont, OK USA	2015-11-02	0.288	0.143	1.000	0.909	132
Lauder, New Zealand	2016-02-29	0.449	0.170	0.932	0.441	118
Orleans, France	2015-11-02	0.627	0.322	0.746	0.661	59
Sodankyla, Finland	2015-08-20	0.533	0.133	0.600	0.267	15
Tsukuba, Japan	2016-05-13	0.390	0.169	0.974	0.818	77
Wollongong, Australia	2015-11-24	0.272	0.115	0.973	0.778	261

Table 1

Coverage probabilities for confidence intervals derived from operational and simulation-based conditional distributions. Values in green indicate empirical coverage at least as large as nominal values.

influence performance of the simulation-based method.

5. Summary and discussion. The primary challenge addressed by this work is that of providing uncertainties for estimates of physical states produced by remote sensing retrieval algorithms. We define uncertainty as the conditional distribution of the true state given the estimated state; this is a probabilistic quantification of what remains unknown about the QOI after seeing the estimate. The methodology we propose approximates this conditional distribution for every sounding for which a retrieved state point estimate exists.

Our method is similar in spirit to the bootstrap bias correction in using the relationship between an original sample and a set of resamples taken from it, as a proxy for the relationship between the unknown truth, and the original sample itself. Here, we extend the idea beyond bias correction alone. We model the full conditional distribution of the true state given a corresponding retrieved state, as a weighted mixture of Gaussian regressions. Then, any operationally-retrieved state estimate can be used as a predictor.

There are a number of significant benefits of this approach. First, it does not require enumerating particular sources of uncertainty in order to be complete. The mixture regression models capture what is known about the aggregated effects of all uncertainty sources, including “unknown unknowns” (e.g., higher-order interaction effects) in the operational processing chain. This is, of course, assuming that the resampling procedure produces representative ensembles of what the operational observing system encounters. Second, forward model structural and parametric uncertainties are approximated by the model discrepancy term. Identifying the role of spectral residuals in estimating model discrepancy opens the door to future experiments and empirical analyses that may help improve the forward model. Third, the entire approach is independent of how the retrieved state estimates are produced. Fourth, since we modeled the joint and conditional distributions as Gaussian mixtures, these distributions may have general non-Gaussian forms. Finally, the entire procedure is executed separately from the operational retrieval process, and can be performed without interfering with operations. Once the simulations are run and the model is fitted, the rest of the computation is very fast and easily applied to the operational output.

Looking to the future, we are exploring spatial and spatio-temporal extensions to our method. Instead of simulating ensembles of independent and identically distributed synthetic true states, we simulate entire spatial fields (e.g., [26]) at once. We are now applying that idea in the context of NASA's ECOSTRESS mission [45]. This spatial version of the post hoc UQ analysis allows us to assess not only per-sounding uncertainties in the form of conditional distributions of true states given retrieved states, but also conditional distributions of *sets* of true states at multiple locations (and eventually times), given retrievals at multiple locations. The uncertainty information therein is critical for obtaining uncertainties in computed quantities such as spatial gradients, which are crucial to understanding geophysical processes in many areas of Earth and environmental science. The spatial version also presents new and exciting computational and modeling challenges because of high dimensionality and massive data set size.

Appendix A. Generating the synthetic true state vector ensemble. This appendix gives additional details of how we generate realizations of $\mathbf{X}^{\text{sim}}_{\text{land}}$. The state vector includes a diverse collection of atmospheric, surface, and instrument properties. These include the vertical profile of CO_2, scaling factors for temperature and water vapor, surface pressure, aerosol concentration and vertical position information, surface albedo, solar-induced fluorescence (SIF), and observation wavelength offsets [22]. Due to this combination of constituents, actual OCO-2 retrievals are a logical source for reference data to inform the simulations. We assemble OCO-2 retrieval products for the region-week combinations represented in our templates. In doing so, we invoke an ergodic assumption that the collection across space and time effectively behave as replicates from a common distribution. We make a few modifications to the template sets for realism and pragmatism. First, the OCO-2 retrievals of aerosol amounts have not been reliably validated [33], so we replace them with values from the MERRA-2 reanalysis [40]. The MERRA-2 aerosol products are available globally every three hours at a grid spacing of 0.5° latitude × 0.625° longitude. We match each OCO-2 retrieval to the closest MERRA-2 location and time. The ergodic assumption is modified for the state vector elements corresponding to surface pressure and instrument dispersion. The variability across a template for these components is partially predictable due to changes in elevation, meteorology, and instrument calibration. This knowledge is captured in a variable retrieval prior mean for these quantities, so we subtract the prior mean vector used in the operational OE retrieval from the retrieved state for these elements in our template sets. A value for a reference sounding is added back when the simulation is executed, as discussed in Section 4.1.2.

We filter out any OCO-2 soundings for which the retrieval algorithm did not converge to a solution within the maximum number of iterations allowed. This information is provided in a variable called outcome_flag that is provided in the data product. We also filter out any soundings for which an additional quality indicator, called warn_level [28], does not have value less than or equal to 15. Finally, because some state vector elements are on different scales, those elements with variability on the order of 10^{-4} or less are multiplied by 1000 in order to avoid problems with covariance matrix inversion later. This rescaling is undone after fitting the mixture model.

To generate the synthetic truth ensembles for each land template, we fit a Gaussian mixture model, $\tilde{P}_r$, $r = \text{L1}, \ldots, \text{L11}$ using the *densityMclust* function in R's mclust package [46]. The procedure is described in Section 3.2. Here, we set the maximum number of components in the mixture to 15, which is the largest number that we felt we could legitimately interpret scientifically, and reduce the dimension of the input state vectors by projecting them into the space of the leading principal components (see Section 3.2) using a threshold of $\gamma = .95$. Both choices are based on balancing the quality of the GMM fit against the time it takes to fit the models. We perform this step *only* if the number of screened data vectors in the template is at least ten times the number of leading eigenvectors used for dimension reduction. We abandon any template strata which do not meet this criterion. Denote the simulation "marginal distribution" for template r by,

$$\tilde{P}_r = \mathrm{GMM}\left(\tilde{K}_r^{\mathrm{sim}}, \left\{\tilde{\boldsymbol{\eta}}_{rk}, \tilde{\boldsymbol{\Omega}}_{rk}, \tilde{p}_{rk}\right\}_{k=1}^{\tilde{K}_r^{\mathrm{sim}}}\right), \quad r = \mathrm{L1}, \ldots, \mathrm{L11}, \tag{A.1}$$

where all mean vectors and covariance matrices are on the original physical scale.

Appendix B. The forward function and its parameters. The OCO-2 full-physics (FP) forward model includes three key modules that we describe briefly, and we refer the reader to [2] for additional details. The FP forward model first solves the equation of radiative transfer (RT) at fine spectral resolution. Some discussion of the transformation from state vector elements to model inputs is provided by [19]. Next, the wavelength-dependence of the solar spectrum is applied to the solution. The final step of the FP model evaluation is the convolution of the fine-resolution spectral response with the OCO-2 instrument line shape (ILS), or spectral response function, and the result is the forward model evaluation for a generic synthetic state vector $\mathbf{X}^{\mathrm{sim}}$,

$$\mathbf{Y}_0^{\mathrm{sim}} = F_1\left(\mathbf{X}^{\mathrm{sim}}, \mathbf{b}_1\right), \tag{B.1}$$

where $\mathbf{b}_1$ are forward model parameters that are also used in the retrieval algorithm.

At least one special circumstance of the simulation experiment sets it apart from the circumstances of the actual observing system: the simulated state vectors do not have latitudes and longitudes associated with them. Thus, they have no location information other than that they were generated by a synthetic marginal distribution distribution associated with a template. Actual forward model evaluations are dependent on geographic location because this determines their observing geometry, including sun angle and other details of the observing configuration. To overcome this problem, we use the observing geometry of a template "reference sounding" for all forward model evaluations of synthetic state vectors drawn from the template's synthetic marginal distribution. Reference soundings are typically located near the geographic center of the template so as to be generally representative of observing conditions.

The observing geometry parameters are some of the components of the parameter vector $\mathbf{b}_1$. Other components include wavelength-specific parameters that describe aerosol scattering and gas absorption properties. Similarly, the instrument line shape (ILS) and solar spectrum are represented by parameters varying with wavelength that are subject to uncertainty. Connor et al. [5] assess the linear sensitivity of the operational OCO-2 retrieval to several of these parameters individually under a range of geophysical conditions. As we describe in 4.1.4, we represent forward model misspecification and parameter uncertainty collectively through an additive model discrepancy in our simulation framework.

Acknowledgments. We would like to thank the Orbiting Carbon Observatory-2 and -3 projects for sponsoring this work, and Noel Cressie, Annmarie Eldering, Maggie Johnson, Mikael Kuusela, Mike Turmon, Jouni Susiluoto, and two anonymous referees for helpful discussions and insights. We also appreciate technical assistance from Dejian Fu and James McDuffie.

REFERENCES

[1] A. AghaKouchak, N. Nasrollai, and E. Habib, *Accounting for uncertainties of the TRMM satellite estimates*, Remote Sensing, (2009), pp. 606–619, https://doi.org/10.3390/rs1030606.

[2] H. Boesch, L. Brown, R. Castano, M. Christi, B. Connor, D. Crisp, A. Eldering, B. Fisher, C. Frankenberg, M. Gunson, R. Granat, J. McDuffie, C. Miller, V. Natraj, D. O'Brien, C. O'Dell, G. Osterman, F. Oyafuso, V. Payne, I. Polonski, M. Smyth, R. Spurr, D. Thompson, and G. Toon, *Orbiting Carbon Observatory–2 Level 2 Full Physics Alogorithm Theoretical Basis Document*, Jet Propulsion Laboratory, March 2015. JPL document OCO D-55207.

[3] J. Brynjarsdóttir, J. Hobbs, A. Braverman, and L. Mandrake, *Optimal estimation versus MCMC for CO_2 retrievals*, Journal of Agricultural, Biological, and Environmental Statistics, 23 (2018), pp. 297—-316, https://doi.org/10.1007/s13253-018-0319-8.

[4] S. Chandrasekhar, *Radiative Transfer*, Dover Publications, Inc., New York, 1960.

[5] B. Connor, H. Bosch, J. McDuffie, T. Taylor, D. Fu, C. Frankenberg, C. O'Dell, V. H. Payne, M. Gunson, R. Pollock, J. Hobbs, F. Oyafuso, and Y. Jiang, *Quantification of uncertainties in OCO-2 measurements of XCO_2: simulations and linear error analysis*, Atmospheric Measurement Techniques, 9 (2016), pp. 5227—-5238, https://doi.org/10.5194/amt-9-5227/2016.

[6] N. Cressie, *Mission CO_2ntrol: A statistical scientist's role in remote sensing of atmospheric carbon dioxide*, Journal of the American Statistical Association, 113 (2018), pp. 152–168, https://doi.org/10.1080/01621459.2017.1419136.

[7] N. Cressie and R. Wang, *Statistical properties of the state obtained by solving a nonlinear multivariate inverse problem*, Applied Stochastic Models in Business and Industry, 29 (2013), pp. 424–438, https://doi.org/10.1002/asmb.1946.

[8] N. Cressie, R. Wang, M. Smyth, and C. Miller, *Statistical bias and variance for the regularized inverse problem: Application to space-based atmospheric CO_2 retrievals*, Journal of Geophysical Research: Atmospheres, 121 (2016), pp. 5526–5537, https://doi.org/10.1002/2015JDG024353.

[9] D. Crisp, *Measuring atmospheric carbon dioxide from space with the Orbiting Carbon Observatory-2 (OCO-2)*, in Earth Observing Systems, vol. 9607, SPIE, 2014, https://doi.org/10.1117/12.2187291.

[10] A. Davison and D. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, UK, 1997.

[11] R. D. de Veaux, *Mixture of linear regressions*, Computational Statistics and Data Analysis, 8 (1989), pp. 227–245, https://doi.org/10.1016/0167-9473(89)90043-1.

[12] B. Efron and R. J. Tibshirani, *Introduction to the Bootstrap*, Chapman and Hall, Boca Raton, FL, 1994.

[13] A. Eldering, C. W. O'Dell, P. O. Wennberg, D. Crisp, M. Gunson, C. Viatte, C. Avis, A. Braverman, et al., *The Orbiting Carbon Observatory-2: First 18 months of science data products*, Atmospheric Measurement Techniques, 10 (2017), pp. 549–563, https://doi.org/10.5194/amt-10-549-2017.

[14] A. Eldering, G. Osterman, R. Pollock, R. Lee, R. Rosenberg, F. Oyafuso, D. Crisp, L. Chapsky, and R. Granat, *Orbiting Carbon Observatory (OCO-2) Level 1B Algorithm Theoretical Basis*, Jet Propulsion Laboratory, 2017. JPL document OCO D-55206.

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

[16] C. Fraley and A. E. Raftery, *Model-Based Clustering, Discriminant Analysis, and Density Estimation*, Journal of the American Statistical Association, 97 (2002).

[17] K. R. Gurney and co authors, *Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models*, Nature, 415 (2002).

[18] H. Haario, M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen, *Markov chain Monte Carlo methods for high dimensional inversion in remote sensing*, Journal of the Royal Statistical Society, Series B, 66 (2004).

[19] J. Hobbs, A. Braverman, N. Cressie, R. Granat, and M. Gunson, *Simulation-based uncertainty quantification for estimating atmospheric CO_2 from satellite data*, Journal on Uncertainty Quantification, 5 (2017), pp. 956–985, https://doi.org/10.1137/16M1060765.

[20] G. C. Hulley, C. G. Hughes, and S. J. Hook, *Quantifying uncertainties in land surface temperature and emissivity retrievals from ASTER and MODIS thermal infrared data*, Journal of Geophysical Research, 117 (2012), https://doi.org/10.1029/2012JD018506.

[21] S. Illingworth, J. Remedios, H. Boesch, D. Moore, H. Sembhi, A. Dudhia, and J. Walker, *ULIRS, an optimal estimation retrieval scheme for carbon monoxide using IASI spectral radiances: sensitivity analysis, error budget and simulations*, Atmospheric Measurement Techniques, 4 (2011), p. 269–288, https://doi.org/10.5194/amt-4-269-2011.

[22] S. S. Kulawik, C. O'Dell, R. R. Nelson, and T. E. Taylor, *Validation of OCO-2 error analysis using simulated retrievals*, Atmospheric Measurement Techniques, 12 (2019), https://doi.org/10.5194/amt-12-5317-2019.

[23] M. Kuusela and V. M. Panaretos, *Statistical unfolding of elementary particle spectra: empirical Bayes estimation and bias-corrected uncertainty quantification*, The Annals of Statistics, 9 (2015), https://doi.org/10.1214/15-AOAS857.

[24] O. Lamminpää, J. Hobbs, J. Brynjarsdóttir, M. Laine, A. Braverman, H. Lindqvist, and J. Tamminen, *Accelerated MCMC for satellite-based measurements of atmospheric CO_2*, Remote Sensing, 11 (2019), https://doi.org/10.3390/rs11172061.

[25] N. Livesey, W. V. Snyder, W. Read, and P. Wagner, *Retrieval algorithms for the EOS Microwave limb sounder (MLS)*, Transactions on Geoscience and Remote Sensing, 44 (2006), https://doi.org/10.1109/TGRS.2006.872327.

[26] P. Ma, E. Kang, A. Braverman, and H. Nguyen, *Spatial statistical downscaling for constructing high-resolution nature runs in global observing system simulation experiments*, Technometrics, 61 (2017), pp. 322–340, https://doi.org/10.1080/00401706.2018.1524791.

[27] V. Maggioni, M. R. Sapiano, R. F. Adler, Y. Tian, and H. G. J., *An error model for uncertainty quantification in high-time-resolution precipitation products*, Journal of Hydrometeorology, 15 (2014), https://doi.org/10.1175/JHM-D-13-0112.1.

[28] L. Mandrake, C. Frankenberg, C. O'Dell, G. Osterman, P. Wennberg, and D. Wunch, *Semi-autonomous sounding selection for OCO-2*, Atmopsheric Measurement Techniques, Discussions, 6 (2013), pp. 5881–5922, https://doi.org/10.5194/amtd-6-5881-2013.

[29] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics, New York, 2000.

[30] C. A. Mears, F. J. Wentz, P. Thorne, and D. Bernie, *Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte Carlo estimation technique*, Journal of Geophysical Research, 116 (2011), https://doi.org/10.1029/2010JD014954.

[31] C. J. Merchant, F. Paul, T. Popp, M. Ablain, S. Bontemps, P. Defourny, H. Rainer, T. Lavergne, A. Laeng, G. de Leeuw, J. Mittaz, C. Poulsen, A. C. Povey, M. Reuter, S. Sathyendranath, S. Sandven, V. F. Sofieva, and W. Wagner, *Uncertainty information in climate data records from Earth observation*, Earth System Science Data, 9 (2017), pp. 511–527, https://doi.org/10.5194/essd-9-511-2017.

[32] National Research Council, Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification, *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*, The National Academies Press, Washington, DC, 2012.

[33] R. R. Nelson and C. W. O'Dell, *The impact of improved aerosol priors on near-infrared measurements of carbon dioxide*, Atmospheric Measurement Techniques, 12 (2019), pp. 1495–1512, https://doi.org/10.5194/amt-12-1495-2019.

[34] C. W. O'Dell, A. Eldering, P. O. Wennberg, D. Crisp, M. R. Gunson, et al., *Improved retrievals of carbon dioxide from the Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm*, Atmospheric Measurement Techniques, 11 (2018), pp. 6539–6576, https://doi.org/10.5194/amt-6539-2018.

[35] P. K. Patra, D. Crisp, J. W. Kaiser, D. Wunch, T. Saeki, K. Ichii, T. Sekiya, P. O. Wennberg, D. G. Feist, D. F. Pollard, D. W. T. Griffith, V. A. Velazco, M. De Maziere, M. K. Sha, C. Roehl, C. Abhishek, and K. Ishijima, *The Orbiting Carbon Observatory (OCO-2) tracks 2–3 peta-gram increase in carbon release to the atmosphere during the 2014–2016 El Nino*, Scientific Reports, 7 (2017), https://doi.org/10.1038/s41598-017-13459-0.

[36] S. Platnick, K. G. Meyer, M. D. King, G. Wind, N. Amarasinghe, B. Marchant, G. T. Arnold,

Z. Zhang, P. A. Hubanks, R. E. Holz, P. Yang, W. L. Ridgway, and J. Riedi, *The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua*, IEEE Transactions on Geosciences and Remote Sensing, 55 (2017), pp. 502–525, https://doi.org/10.1109/TGRS.2016.2610522.

[37] S. Platnick, R. Pincus, B. Wind, M. D. King, M. A. Gray, and P. Hubanks, *An initial analysis of the pixel-level uncertainties in global MODIS cloud optical thickness and effective particle size retrievals*, in Passive Optical Remote Sensing of the Atmosphere and Clouds, IV, S. C. Tsay, T. Yakota, and M.-H. Ahn, eds., vol. 5652, SPIE, 2004, pp. 30–40, https://doi.org/10.1117/12.578353.

[38] D. J. Posselt and G. D. Mace, *MCMC-based assessment of the error characteristics of a surface-based combined radar–passive microwave cloud property retrieval*, Journal of Applied Meteorology and Climatology, (2014), https://doi.org/10.1175/JAMC-D-13-0237.1.

[39] A. Povey and R. Granger, *Known and unknown unknowns: uncertainty estimation in satellite remote sensing*, Atmospheric Measurement Techniques, 8 (2015), pp. 4699—-4718, https://doi.org/10.5194/amt-8-4699-2015.

[40] C. A. Randles, A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, J. Hair, Y. Shinozuka, and C. J. Flynn, *The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation*, Journal of Climate, 30 (2017), pp. 6823–6850, https://doi.org/10.1175/JCLI-D-16-0609.1.

[41] C. D. Rodgers, *Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation*, Reviews of Geophysics and Space Physics, 14 (1976).

[42] C. D. Rodgers, *Inverse Methods for Atmospheric Sounding, Theory and Practice*, World Scientific, Singapore, 2000.

[43] I. Ryoichi, S. Naoko, O. Yoshifumi, and T. Shoichi, *CO_2 retrieval performance of TANSO-FTS (TIR) sensor aboard greenhouse gases observing satellite (GOSAT)*, in 15th Symposium on High-Resolution Molecular Spectroscopy, vol. 6580, SPIE, 2006, pp. 253 – 258, https://doi.org/10.1117/12.724963.

[44] J. Schafer, R. Opgen-Rhein, V. Zuber, M. Ahdesmaki, D. Silva, A. Pedro, and K. Strimmer, *corpcor: Efficient estimation of covariance and (partial) correlation*, 2017, https://CRAN.R-project.org/package=corpcor. R package version 1.6.9.

[45] D. Schimel and F. D. Schneider, *Flux towers in the sky: global ecology from space*, New Phytologist, 224 (2019), https://doi.org/10.1111/nph.15934.

[46] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, *mclust 5: clustering, classification and density estimation using Gaussian finite mixture models*, The R Journal, 8 (2016), pp. 205–233, https://journal.r-project.org/archive/2017/RJ-2017-008/RJ-2017-008.pdf.

[47] H. G. Sung, *Gaussian mixture regression and classification*, PhD thesis, Rice University, 2004, http://hdl.handle.net/1911/18710.

[48] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.

[49] D. R. Thompson, V. Natraj, R. O. Green, M. C. Helmlinger, B.-C. Gao, and M. L. Eastwood, *Optimal estimation for imaging spectrometer atmospheric correction*, Remote Sensing of Environment, 216 (2018), pp. 355–373, https://doi.org/10.1016/j.rse.2018.07.003.

[50] Y. Tian and K. Harrison, *Special issue: Uncertainties in Remote Sensing*, Remote Sensing, 9 (2017), https://www.mdpi.com/journal/remotesensing/special_issues/uncertaintiesRS.

[51] J. Warner, L. Strow, C. Barnet, L. Sparling, G. Diskin, and G. Sachse, *Improved agreement of airs tropospheric carbon monoxide products with other EOS sensors using optimal estimation retrievals*, Atmospheric Chemistry and Physics, 10 (2010), https://doi.org/10.5194/acp-10-9521-2010.

[52] J. R. Worden, G. B. Doran, S. Kulawik, A. Eldering, D. Crisp, C. Frankenberg, C. O'Dell, and K. Bowman, *Evaluation and attribution of OCO-2 XCO2 uncertainties*, Atmospheric Measurement Techniques, 10 (2017), pp. 2759–2771, https://doi.org/10.5194/amt-10-2759-2017.

[53] D. Wunch, P. O. Wennberg, G. Osterman, B. Fisher, B. Naylor, C. M. Roehl, C. O'Dell, L. Mandrake, C. Viatte, M. Kiel, D. W. T. Griffith, N. M. Deutscher, V. A. Velazco, J. Notholt, T. Warneke, C. Petri, M. De Maziere, M. K. Sha, R. Sussmann, M. Rettinger, D. Pollard, J. Robinson, I. Morino, O. Uchino, F. Hase, T. Blumenstock, D. G. Feist, S. G. Arnold, K. Strong, J. Mendonca, R. Kivi, P. Heikkinen, L. Iraci, J. Podolske, P. W.

Hillyard, S. Kawakami, M. K. Dubey, H. A. Parker, E. Sepulveda, O. E. García, Y. Te, P. Jeseck, M. R. Gunson, D. Crisp, and A. Eldering, *Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON*, Atmospheric Measurement Techniques, 10 (2017), pp. 2209—2238, https://doi.org/10.5194/amt-10-2209-2017.

[54] L. Xin, *China launches carbon-dioxide mission*, Physics World, 30 (2017).

[55] D. Yang, H. Zhang, Y. Liu, B. Chen, Z. Cai, and D. Lu, *Monitoring carbon dioxide from space: Retrieval algorithm and flux inversion based on GOSAT data and using CarbonTracker–China*, Advances in Atmospheric Sciences, 34 (2017), pp. 965–976, https://doi.org/10.1007/s00376-017-6221-4.