**Predicting and correcting the influence of boundary conditions on inverse analyses**

**Abstract**

[To write]

**1. Introduction**

Regional simulations of greenhouse gases provide insights into greenhouse gas fluxes at relatively high spatiotemporal resolution. At local scales, such analyses support emissions quantification and mitigation efforts for urban and regional governments and for private entities (e.g., Zhang et al., 2020; Sargent et al., 2021; Jeong et al., 2016; Yu et al., 2021). At larger scales, inversions can support national emissions reduction strategies (e.g., Z. Chen et al., 2022, 2023; Nesser et al., 2024) and improve understanding of the factors driving carbon cycle variability and its response to a changing climate (e.g., Schuh et al., 2004; Byrne et al., 2023). However, errors in the boundary conditions used to drive background concentrations may have a spatially and temporally variable influence on model output, resulting in systematic biases in the inferred emissions. We characterize the sensitivity of simulated concentrations to boundary conditions to predict and correct the influence of boundary condition errors. We consider the case of inverse analyses of satellite observations used to infer greenhouse gas emissions, but our analysis is generalizable to any regional simulation of long-lived atmospheric trace gases.

Regional inverse analyses infer the emissions that best explain observed greenhouse gas concentrations by fitting the data to simulated concentrations from a chemical transport model (CTM). This optimization is often done by minimizing a Bayesian cost function regularized by a prior emissions estimate (Rodgers, 2000). Other inversions estimate fluxes using the ratio of the observed anomaly relative to the background to a modeled anomaly (e.g., Sargent et al., 2018). In all cases, regional inversions must contend with data for which the inflow dominates the observed concentrations. Often, the variability of inflow also dominates observed variability, as for many urban CO2 flux inversions. Accurate specification of the background is therefore essential to avoid propagating biases in the modeled background to the inferred fluxes.

Boundary conditions for regional simulations are often provided by coarse-resolution global simulations adjusted to be unbiased with respect to observations (e.g., Lauvaux et al., 2012, Gockede et al., 2010, Qu et al., 20xx). Other studies use in situ observations of greenhouse gas concentrations and meteorology at the edge of the inversion domain to estimate the inflow concentrations (e.g., Lauvaux et al., 2016; Sargent et al., 2018; Balashov et al., 2020; …). In both cases, the observations used to constrain the boundary condition are often spatially or temporally sparse, limiting the periods over which they provide a constraint (citation). Smoothed satellite observations have also been used to generate boundary conditions (e.g., Estrada et al., 2023), but this is limited to spatiotemporally dense data and may remove systematic variability.

Given the specified boundary conditions, residual biases may also be corrected as part an inversion. Boundary conditions can be optimized as part of the inversion (e.g., Wecht et al., 2014; Z. Chen et al., 2022), but the spatial and temporal resolution at which boundary conditions should be adjusted is poorly constrained. Optimizing the boundary condition at high spatiotemporal resolution may not be possible with the available information from the observations or within reasonable computational cost. Biases may also be corrected by allowing “buffer grid cells” at the edge of the domain to absorb any residual errors (e.g., Shen et al., 20xx; Varon et al., 20xx). These buffer grid cells increase the size of the domain and therefore the computational cost of the inversion.

Despite efforts to reduce biases in the boundary conditions used in regional inverse analyses, residual biases may have a significant effect on inferred fluxes. Göckede et al. (2010) found that a mean bias of 0.1 ppm in modeled boundary conditions could bias flux estimates for the state of Oregon by 3.5 Tg C a-1, representing ~10% of net annual uptake. Lauvaux et al. (2012) found in a regional inversion of CO2 observations from towers in the U.S. Corn Belt that potential bias in the aircraft-corrected boundary conditions resulted in an additional uncertainty of 24 Tg C, representing a 13% error on the total carbon estimate and almost doubling the original uncertainty estimate of 35 Tg C. Feng et al. (2019) and Chen et al. (2019) found that while boundary condition biases were the lowest source of model uncertainty within the planetary boundary layer, they were the largest source of errors in total column concentrations.

We present a theoretical and numerical framework to quantify, predict, and correct the influence of these biases on the improved emissions estimates. We develop an analytical framework for the influence of boundary condition biases on posterior fluxes, including the definition of a parameter to predict the influence of biases. We demonstrate the framework by applying constant, random, and periodic boundary condition perturbations to a series of inversions of increasing complexity, from a simple one-dimensional model for the transport of an inert tracer to a two-dimensional system with real transport but simulated observations. Within these inversions, we test the effect of various interventions to correct for boundary condition biases within the inversion. Finally, we apply the framework and the lessons learned to a real inversion over the Permian Basin.

**2 Framework for quantifying the effect of boundary condition biases**

We quantify the effect of boundary condition biases and inverse methods to correct for the biases using observing system simulation experiments (OSSEs) of increasing complexity. Using the true fluxes and boundary conditions, each OSSE generates pseudo-observations that are used in inversions with varying boundary condition perturbations and correction methods. We compare the posterior fluxes to those generated by an inversion solved using the true boundary condition and no correction method (the “true” inversion). After developing this framework for the analytical solution (Section 2.1), we apply it to a one-dimensional model of an inert tracer (Section 3) and to a two-dimensional OSSE over the Permian basin (Section 4). These simple models remove transport and representation errors, allowing exact quantification of the influence of boundary condition biases on the posterior solution. By successively increasing the complexity of the framework, we build understanding of the variables controlling boundary condition biases, hypothesize which results are generally transferable, and establish a set of tools for best predicting and correcting the resulting biases.

Table 1 summarizes the bias correction methods tested here. The “standard” case demonstrates the effect of no correction intervention. The “correction” method one or more elements corresponding to the boundary condition along with the fluxes (e.g., Wecht et al., 2014; Chen et al. …; Nesser et al., 2024; Estrada et al., 2024). The “buffer” approach increases the prior error standard deviation by a factor of in the first grid cells so that those grid cells “absorb” biases in the boundary condition (e.g., Shen et al.; Varon et al.). The “combined” method uses both boundary condition correction and buffer grid cells. The “sequential” approach optimizes first the boundary condition and then the fluxes in two separate inversions (e.g., Wecht et al., 2014).

**2.1 Analytical solution to the inverse problem**

Given an *n*-dimensional vector of gridded emissions and the *m*-dimensional vector of observations , both with normally distributed errors, the optimal emissions estimate is obtained by minimizing a Bayesian cost function

where and are the prior emissions estimate and error covariance matrix, respectively, is the observational error covariance matrix, and is the chemical transport model (CTM) that simulates observations as a function of emissions (Brasseur and Jacob, 2017). The model – observation difference can be written as , where gives the “true” model – observation difference attributable to the fluxes and gives the total observing system errors, including uncertainty in the boundary condition, forward model transport, instrument, and representation.

If the CTM is linear, then where is the Jacobian matrix that is a function of the species’ lifetime within a given grid cell. Indeed, for a steady state system that optimizes only gridded fluxes, the Jacobian matrix can be derived through mass balance to be lower diagonal with entries equal to the lifetime of the species in a single grid cell. The constant represents the background defined by the boundary condition so that in the case that represents gridded fluxes. If the boundary condition is optimized as part of the inversion, information about the boundary condition is instead contained in the columns of the Jacobian matrix that represent the sensitivity of the observations to the boundary condition (). In both cases, an analytical solution exists for the cost function minimum that yields the optimal (posterior) emissions estimate :

where

is the gain matrix that represents the sensitivity of the posterior emissions to the observations. The posterior solution is therefore an update to the prior flux estimate controlled by the model – observation difference as mediated by the gain matrix. Importantly, the gain matrix controls both the signal of “true” model – observation difference and the noise of the errors and biases, including those in the boundary condition. Because posterior emissions are most sensitive to nearby observations and observations near the boundary are most sensitive to the boundary, the gain matrix results in large boundary condition effects at domain boundaries. The boundary condition interventions (Table 1) use elements optimized by the inversion to correct for these biases. However, because the buffer approach optimizes emissions, its efficacy depends on the buffer grid cell prior flux and uncertainty and on grid cell lifetime, while the correction method depends only on boundary condition concentrations and the associated uncertainty.

We define the true posterior emissions generated by an inversion that uses the true boundary condition and no correction method. We then consider the effect of a boundary condition perturbation . For each of the interventions, we calculate the difference between the corrected inversion with a perturbed boundary condition and :

We define as the posterior solution and as the gain matrix for a given intervention and boundary condition perturbation. The first term of Equation 4 represents the loss in “true” information about the fluxes resulting from the correction method. The second term represents the change in the influence of the boundary condition bias.

We consider first a simple inversion of a steady state system with two observations of an inert species over two optimized flux grid cells. We assume a uniform prior with fluxes and uniform relative prior errors . Observing system errors are also assumed constant. Off-diagonal terms for both error covariance matrices are set to zero. The Jacobian matrix is lower diagonal with entries equal to the single grid cell lifetime . We compute Equation 4 for the correction and buffer methods. For the buffer approach, we scale the prior error standard deviation for the first grid cell by a factor of . For the correction method, the prior for the boundary condition has concentration and relative errors . The column of the Jacobian matrix corresponding to the boundary condition element is a vector of ones.

We compare the posterior solution for the second grid cell since the first grid cell is discarded by the buffer method. If the boundary condition perturbation is larger than the emissions signal (), the boundary condition correction approach minimizes errors relative to the buffer grid cell approach if the ratio of the errors in the boundary condition to the errors in the emitted concentrations for the buffer grid cell is greater than 1:

If the boundary condition perturbation is instead smaller than the emissions signal (), the inequality reverses direction. While the magnitude of the true signal from the emissions is unknown in real inversions, it could be approximated using the observing system error covariance estimate. Similarly, the magnitude of the boundary condition bias could be estimated by comparison to observations. Equation 5 can then be used to guide decision making about the best correction method for boundary condition biases.

**3 One-dimensional numerical solution**

We first demonstrate this analytical framework with inversions of a one-dimensional model for the transport of a passive tracer. Figure 1 shows a schematic of this model for grid boxes of length (m). The tracer sources are the boundary condition (ppb) and the emissions (ppb h-1). The domain is ventilated with wind speed (km/hr). Advection is solved using the Lax-Wendroff scheme in the first grid cells and an upstream scheme for the last grid cell (Brasseur and Jacob, 2017). The time step is defined so that the maximum Courant number equals 1. Initial conditions are given by steady state concentrations. We generate pseudo-observations by adding random noise to the simulated concentrations (ppb) for each grid cell driven by the true boundary condition. Given prior emissions , prior errors , and observing system errors , we solve the inversion using Equation 2. We specify demonstration values for each of these terms to demonstrate the effect of different correction methods on boundary condition biases (Sections 3.1 and 3.2). We generalize the results by varying the parameter choice in Section 3.3.

In the demonstration inversions, we simulate concentrations of the inert tracer over grid cells of length km with constant emissions of 30 ppb d-1 in each grid cell and a true boundary condition of 1900 ppb at all time steps. We test the effect of constant ( km/hr) and varying wind speeds. The varying wind decelerates from 5 km/hr to 0 km/hr, reverses direction, and accelerates to – 5 km/hr before repeating in reverse. After a 150-hour spin-up, we sample each grid box at 50 regular intervals for 150 hours to generate 1000 observations. Pseudo-observations are created by adding random noise with mean 0 ppb and standard deviation 8 ppb to the simulated concentrations driven by the true boundary condition.The prior emissions vector for the inversion is given by random values with mean 25 ppb d-1 and standard deviation 5 ppb d-1. We use relative diagonal prior errors of 50% and diagonal observing system errors of 10 ppb. The Jacobian matrix is constructed using a finite difference approach. Figure 2 shows the posterior emissions for an inversion solved with the true boundary condition (). The inversion is most accurate in the center of the domain where emissions have a larger relative influence than the boundary condition on simulated concentrations and where there are many downstream observations.

We apply boundary condition perturbations and test our bias correction interventions with this model. For the correction method, the prior value of the boundary condition element is given by the boundary condition perturbation with a prior error standard deviation of 50 ppb. For the buffer approach, we scale the prior error standard deviation in the first grid cell by a factor of ranging from 5 to 50. Given a grid cell lifetime of 5 hr for the constant wind speed, the range results in and < 1 (Equation 5), suggesting that the buffer and correction approaches should perform similarly in some cases.

**3.1 Constant and random boundary condition perturbations**

We first test the effect of constant boundary condition perturbations. Constant boundary condition perturbations may exist because of a region-wide bias (e.g., from errors in the atmospheric sink for methane or from a bias in a flux area source) or because of low-frequency biases in the boundary conditions of a high temporal resolution inversion. Constant perturbations also provide the simplest representation of the common case where a mean bias exists in the boundary condition.

We perturb the true boundary condition in this simple one-dimensional model by a constant value, solve the inversion, and compare the posterior emissions. Figure 3 shows the relative difference in the posterior emissions for the standard inversion compared to the true emissions (top left) and to the posterior solution generated with the true boundary condition (middle left) for a range of boundary condition perturbations. As expected from Equation 4 and the structure of the gain matrix (Section 2.1), the error increases with the boundary condition perturbation and decreases exponentially as the distance from the upstream boundary increases.

Figure 3 also shows the effect of the four bias correction interventions. None of the interventions entirely removes the influence of the boundary condition perturbation compared to the true solution. However, boundary condition correction (lower left) results in the smallest biases regardless of perturbation magnitude. Indeed, the magnitude of the residual bias is independent of the perturbation. The uncorrected bias results from the change in the signal (first term of Equation 4) resulting from the intervention: boundary condition correction decreases the information available to correct the upwind fluxes, increasing the error. The boundary condition optimized by the inversion is in all cases slightly higher than the true boundary condition (1905 ppb compared to 1900 ppb), suggesting that the inversion cannot perfectly distinguish between the increase in concentrations attributable to the boundary condition compared to the emissions.

The buffer approach (top right) produces similar results to the correction approach. It successfully limits the direct effect of the boundary condition perturbation (second term of Equation 4) to the most upstream grid cell, but it results in a large change to the emissions signal in nearby grid cells that mitigates the benefit of this approach in the subsequent grid box. Larger scale factors improve performance. When , (Equation 5) suggesting that the buffer approach should outperform the correction approach. Instead, the buffer approach generates identical results to the correction approach in all but the buffer grid cell, suggesting that the correction approach represents an upper performance bound. When we vary the inversion parameters, we find that this result is robust in cases where the boundary condition prior error is larger than the boundary condition perturbation . If is smaller than , the added cost of correcting the boundary condition improves the relative performance of the buffer approach so that it outperforms the correction approach in non-buffer grid cells. Larger buffer grid cells do not improve the performance of this approach, suggesting that smaller buffer grid cells can be used to reduce the computational cost of regional inversions.

Combining the buffer and correction approaches (center right) degrades the performance of the correction approach in the buffer grid cell but improves the performance of the buffer method in non-buffer grid cells. If there is interest in recovering information in the most upstream grid cell, it is better to use the correction approach alone, though there may still be residual biases.

Finally, the sequential approach (lower right) is least successful. The first inversion yields a too-large boundary condition estimate because the inversion is unable to distinguish between the increased concentrations resulting from the boundary condition perturbation and from the emissions. This too-large estimate propagates to the second step. We therefore remove the sequential approach from consideration.

Figure 4 shows the effect of varying wind speeds on this demonstration. Compared to the true emissions (top left), the inversions done with varying wind speeds on average outperform those done with constant wind speeds because the emitted concentrations are observed for a longer period before they are advected out of the domain. However, the influence of the boundary condition perturbations persists systematically throughout the domain. Consistent with the “sloshing” winds used here, the resulting bias resembles a damped oscillation beginning at the upstream domain edge.

Despite these systematic biases, the correction method (lower left) almost entirely removes both upstream and downstream biases. The improved performance compared to the constant wind speed inversions results from the increased information from the observations about emissions. By contrast, the buffer approach (top right) is unable to correct for downstream biases. This is because the buffer approach corrects a flux element rather than a boundary condition element, for which the uncertainty is proportional to the grid cell lifetime of the species. The systematic changes in grid cell lifetime resulting from “sloshing” wind therefore propagate more to the buffer method than to the correction approach. The combined approach (center right) performs as well as the buffer approach.

Finally, we test the effect of adding random boundary condition perturbations of mean ranging from 5 to 45 ppb and standard deviation 5 ppb. Random boundary condition errors are unavoidable in inversions whether the boundary conditions are given by model simulations or smoothed observations. The effects of the random perturbations are consistent with those found in Figures 3 and 4, suggesting that the constant boundary condition perturbations described above are adequate to represent the common case where a mean bias with random noise exists in the boundary condition of an inversion.

**3.2 Systematic boundary condition biases**

Systematic boundary condition biases may occur if, for example, there are errors in the seasonality or diurnal cycle of a model simulation used to generate boundary conditions or if variability in nearby sources driven by weekday-weekend effects or by economic drivers is smoothed out in the boundary conditions. Such systematic biases are unresolvable within inversions. We therefore demonstrate how different components of systematic biases influence the inferred fluxes to illustrate the types of boundary condition systematic biases that are most important to avoid in an inversion.

We assume that systematic biases can be represented as the sum of periodic functions and therefore simulate their effect using periodic boundary condition biases. We define a base periodic boundary condition perturbation as a function of time

where is the last simulated time. This perturbation has no mean bias relative to the true boundary condition, an amplitude of 10 ppb, and completes two complete periods within the simulated period, including the spin-up. We then separately vary the y-intercept (from 1900 ppb to 1940 ppb), amplitude (from 10 ppb to 50 ppb), frequency (from one to 5 complete periods within the simulated period), and phase (from zero to 2).

Figure 5 shows the periodic boundary condition perturbations (first column), with each row representing the varied intercept, amplitude, frequency, and phase, respectively. We also show the resulting effect on the inferred emissionsfor the standard inversion and the correction, buffer, and combined interventions (second through fifth column, respectively). We divide the effect of the periodic perturbations into two components: the upstream component and the downstream component.

The upstream errors resemble those produced by constant boundary condition perturbations: the errors are largest for the grid cells closest to the upstream boundary, increase with the intercept or magnitude of the perturbation, and decay exponentially as distance from the upstream boundary increases. Unlike the constant boundary condition case, the interventions have limited success correcting these biases, although all three interventions successfully limit the upstream biases to the most upstream grid cells. As in the constant perturbation case, the correction and combined approaches result in the smallest errors compared to the standard inversion. The remaining biases result from the decreased information available to the upstream grid cells. The relative performance of these interventions is robust when the boundary condition prior error is larger than the boundary condition perturbation . If is instead smaller than , the added cost of correcting the boundary condition degrades its performance. In these cases, the combined approach produces the smallest biases in the upstream grid cells.

The downstream errors are systematic biases in the inferred fluxes that persist beyond the upstream enhancement. While none of the interventions reduce the downstream biases. However, all downstream biases are less than 50%. These biases are effected by the amplitude, frequency, and phase but not the intercept. The amplitude of the perturbation controls the magnitude of the downstream biases so that only the largest amplitude perturbations (40 ppb and 50 ppb) produce errors larger than 25%. The distribution of the biases across the domain is controlled by the frequency and phase of the perturbation. The relative magnitude of these varying distributions is correlated with the mean bias in the boundary condition over the inversion period (i.e., post spin-up). Moreover, higher frequency perturbations tend to generate higher frequency biases in the inferred fluxes, while lower frequency perturbations tend to generate more constant biases.

**3.3 Varying inversion configurations**

Figure 6 shows the effect of varying the prior error standard deviations (left), observing system error standard deviations (center), and grid cell lifetime (right) on each correction method. For each sensitivity test, we hold all other parameters constant as described for the demonstration inversion. We apply a 10 ppb boundary condition perturbation in all cases. We quantify the error induced by the boundary condition perturbation and the correction method as the RMSE of the posterior fluxes compared to the true fluxes. We do not compare the posterior fluxes to the true posterior because there is no clear “ideal” set of inversion parameters. Error bars represent the standard deviation of the RMSE generated for each parameter set for 100 inversions solved with random prior emissions.

In general, varying the inversion parameters has a small effect on (posterior – true) emissions RMSE. Increasing the prior error standard deviations increases the RMSE, indicating that the prior plays an important regularizing role in the solution. Increasing the observing system error standard deviation causes the RMSE to move towards the (prior – true) emissions RMSE, consistent with the solution’s increased reliance on the prior emissions. Increasing grid cell lifetime decreases the RMSE because the observations provide more information on the emissions, which represent a larger fraction of the total observed concentration than for shorter lifetimes.

Across all the interventions, the correction method minimizes RMSE for all parameter combinations. For all but one set of parameters, the correction method RMSE is also smaller than the (prior – true) emissions RMSE, indicating that the correction method reduces the boundary condition bias and corrects the emissions toward the truth. In all cases, the buffer approach maximizes the RMSE, though the difference is not always significant compared to the standard inversion. This is consistent with the method’s attempts to maximize errors in the buffer grid cells to minimize errors in the remainder of the domain. The combination approach is statistically comparable to the correction approach, though it consistently produces slightly larger RMSE values. For large prior errors, the performance of the combination approach degrades, consistent with its increased reliance on prior error values compared to the correction approach alone (Section 3). We find that (Equation 5) does not correctly predict the relative performance of the correction and buffer methods.

**4 Two-dimensional solution**

[In prep]

**5 Demonstration inversion of TROPOMI observations over the Permian**

[To write]

**6 Discussion and Conclusions**

[To write]