MODEL ERROR VARIANCE ESTIMATION FOR WEAK CONSTRAINT DATA ASSIMILATION*

SANDRA R. BABYALE , JODI MEAD†, DONNA CALHOUN‡, AND PATRICIA AZIKE

Abstract. State estimates from weak constraint four-dimensional variational (4D-Var) data assimilation can vary significantly depending on the data and model error variance. As a result, the accuracy of these estimates heavily depends on the correct specification of both model and observational data error variances. In this work, we assume that the data error is specified and frame weak constraint 4D-Var as a regularization inverse problem, with the scalar model error variance as the regularization parameter. We employ the representer method to reduce the 4D-Var solution search space from the state space to the data space, which provides an analytical expression for the optimal state estimates. This method allows us to derive matrix expressions for three regularization parameter selection methods: the L-curve, generalized cross-validation (GCV), and the χ^2 method, used to estimate the model error variance. We validate our approach by assimilating simulated data into a 1D transport equation modeling wildfire smoke transport under various observational noise and first guess perturbations. The results show that the estimated model error variances accurately capture the balance between the influence of observational data and model predictions on assimilated state estimates.

Key words. variational data assimilation, weak constraint, model error variance, inverse methods, regularization parameter

AMS subject classifications. 65K10, 65F22

1. Introduction.

Data assimilation is an important tool for analyzing complex physical phenomena across various scientific disciplines, such as geoscience, oceanography, and atmospheric sciences among others [6]. At its core, data assimilation merges theory, mathematical models, and observations to provide a more accurate estimation of a system's state. While dynamical models provide insights into physical interactions and observations provide point measurements, combining both through data assimilation results in a more robust and realistic depiction of the physical phenomena in play. This integration offers a clear advantage over relying solely on estimates from either the model or observations alone, as models often face numerous limitations and challenges and observations are frequently sparse and incomplete [2, 3, 16, 18, 21, 26].

Data assimilation methods can be broadly categorized into sequential and variational approaches. Sequential methods, such as the particle filter, Kalman Filter, Extended Kalman Filter (EKF), and Ensemble Kalman Filter (EnKF), are derived from stochastic filtering principles and Bayesian minimum variance estimation. These methods continually update system state statistics as new observations become available, propagating the system state forward in time [11, 14, 27]. Variational methods, on the other hand, are grounded in optimal control theory and calculus of variations. These methods include 3D-Var (three-dimensional variational) and 4D-Var (four-dimensional variational) approaches. The 3D-Var method optimizes the cost function using observations from a single time window, focusing solely on spatial dimensions while as the 4D-Var method extends this optimization to include the temporal dimension, considering how the system evolves over time. This makes 4D-Var particularly advantageous for systems with complex temporal dynamics, where sequential methods may struggle [8, 17, 28].

*Submitted to the editors DATE.

Funding: This work was funded by the National Science Foundation grant DMS-2111585.

†Department of Mathematics, Boise State University, Boise, ID.

‡Department of Mathematics, Boise State University, Boise, ID.

In 4D-Var data assimilation, the model is corrected in both spatial and temporal dimensions using either strong or weak constraint formulations. In the strong constraint formulation, it is assumed that the numerical model is perfect, meaning there are no errors in the model dynamics. The cost function is minimized under the assumption that any discrepancies between the model and the observations are solely due to observational errors. Conversely, the weak constraint formulation allows for the presence of model errors, acknowledging that the model dynamics may not perfectly represent the true system due to errors that may arise from parameterization, unresolved processes in the forcing, and other factors. In this case, the cost function includes additional terms that account for model errors, which is particularly advantageous for large-scale models that may not fully capture the intricacies of real-world systems [32, 33]. According to [35], weak constraints 4D-Var data assimilation often yields accurate results because it includes more information than the strong constraints method and is capable of altering the dynamics of a given model while fitting observations.

Optimizing the cost function for weak constraint 4D-Var continuous data assimilation results in a system of coupled Euler-Lagrange equations, which can be decoupled using the representer method. The representer method leverages the finite-dimensional nature of the observation space to express the optimal state estimates as a linear combination of representers, one per datum thereby reducing the solution search space from the state space to the data space. [4, 7, 20, 25]. State estimates obtained using weak constraint 4D-Var with representers can vary significantly depending on the choice of data and model error variances. The accuracy of these state estimates heavily relies on properly specifying model and observational data error variances, which is challenging due to large system dimensions and insufficient observations. In this work, we propose a principled approach to determining the optimal model error variance by treating weak-constraint 4D-Var as a regularized inverse problem.

Inverse methods, of which data assimilation methods are fundamentally a subset, are mathematical techniques used to estimate model parameters from observable data. Inverse methods typically invert the forward model for unknown parameters, while data assimilation typically seeks initial conditions, boundary conditions, and states from measured outputs of complex systems. Inverse problems can be classified as either well-posed or ill-posed. A problem is well-posed if a solution exists, is unique, and changes continuously with the input data, whereas a problem is ill-posed if it lacks one or more of these properties. In practice, most inverse problems are ill-posed. The most common cause of ill-posedness is non-uniqueness, which typically arises when the number of parameters to be estimated exceeds the amount of available observational data, which is often the case in data assimilation problems. Because inverse problems are often ill-posed, they cannot be solved without specifying certain assumptions or constraints in form of regularization. Therefore, the main goal of inverse methods is to develop regularization approaches that transform ill-posed problems into a well-posed ones [10, 30, 34]. Common regularization techniques include Bayesian methods, which incorporate prior knowledge about the parameters into the solution process, and Tikhonov regularization, which adds a penalty term to control the magnitude of the parameters. Both techniques are often formulated as least-squares problem with an L2-norm regularization term.

Regularization parameter selection methods are used to determine the influence of the regularization term and balance the trade-off between it and the fit to the observational data. Methods such as the L-curve, Generalized Cross-Validation (GCV), and the χ^2 method provide systematic approaches for selecting this parameter. The

L-curve method involves plotting the norm of the solution against the norm of the residuals to identify a corner point that represents their optimal balance [15]. Generalized Cross-Validation (GCV) minimizes the prediction error by systematically leaving out parts of the data and estimating the error [13]. The χ^2 method ensures that the regularized residual behaves like the expected distribution [23, 22, 24, 29]. These methods ensure that the chosen regularization parameter effectively stabilizes the solution while maintaining an accurate fit to the observational data, thereby enhancing the overall solution's stability and accuracy [1, 9, 12, 31].

We leverage the regularization parameter selection methods(L-curve, Generalized Cross-Validation (GCV), and the χ^2 method) to estimate dynamic model error variance. The elegance of the representer method seamlessly integrates with these regularization parameter selection techniques, enabling efficient computations. By calibrating the model error variance in this way, our approach paves the way for robust and reliable state estimation through weak-constraint 4D-Var data assimilation, even in the face of substantial model and observational uncertainties.

This article is structured as follows: In section 2, we provide a brief overview of weak constraint 4D-Var data assimilation with the method of representers. In section 3, we derive the reduced penalty functionals that result when using the method of representers. Given these reduced penalty functionals, in section 4, we present weak constraint data assimilation as a regularization inverse problem problem and derive regularization parameter selection methods that can be used to estimate the optimal model error variance. Numerical experiments using this approach are presented in section 5. Finally, we give conclusions in section 6.

2. Weak Constraint 4D-Var Data Assimilation with Representers. We demonstrate weak constraint 4D-Var data assimilation using the model

$$\begin{aligned} \frac{\partial q}{\partial t} + L[q(\mathbf{x},t)] &= Q(\mathbf{x},t) \quad \text{for } \mathbf{x} \in \Omega, \quad t \in [0,T] \\ q(\mathbf{x},0) &= I(\mathbf{x}) \\ q(\mathbf{0},t) &= B(t) \end{aligned} \tag{2.1}$$

where operator $L[q(\mathbf{x},t)]$ represents the dynamics and physics that are linear or nonlinear in nature, $Q(\mathbf{x},t)$ represents the forcing, $I(\mathbf{x})$ is the initial state and B(t) represents the boundary conditions.

We assume that imperfect data are available at a limited number of points $(\mathbf{x}_m, t_m)$, collected at M points in space and time. These observations are related to the "true" state $q(\mathbf{x},t)$ by

$$\mathbf{d}_m = q(\mathbf{x}_m, t_m) + \boldsymbol{\epsilon}_m \qquad m = 1, 2, \cdots, M \tag{2.2}$$

where $\boldsymbol{\epsilon}_m \sim \mathcal{N}(0, \sigma_m^2)$ is the measurement error with error variance σ_m^2.

In weak-constraint 4D-Var data assimilation, we account for errors in the model, which may arise from neglected dynamics, inaccuracies in parameterization of the physics, truncation, uncertainty in the forcing etc, i.e.

$$\begin{aligned} \frac{\partial q}{\partial t} + L[q(\mathbf{x},t)] &= Q(\mathbf{x},t) + f(\mathbf{x},t) \quad \text{for } \mathbf{x} \in \Omega, \quad t \in [0,T] \\ q(\mathbf{x},0) &= I(\mathbf{x}) + i(\mathbf{x}) \\ q(\mathbf{0},t) &= B(t) + b(t) \end{aligned} \tag{2.3}$$

where $f(\mathbf{x},t)$, $i(\mathbf{x})$ and $b(t)$ are errors in the model dynamics, in the initial and boundary conditions with variances ${\sigma_f}^2$, ${\sigma_i}^2$, and ${\sigma_b}^2$ respectively. These errors are assumed to be unbiased and mutually uncorrelated.

Since the core reason for data assimilation is to obtain the optimal state estimates that best fit the model and the observations, we minimize the weighted least-squares cost function $\mathcal{J}$ with respect to the state variable $q(\mathbf{x},t)$. We seek to find the state estimate $\hat{q}(\mathbf{x},t)$ that corresponds to the smallest values of f,i,b and $\boldsymbol{\epsilon}_m$, i.e.

$$\hat{q}(\mathbf{x},t) = \arg\min_q \mathcal{J}[q] \tag{2.4}$$

where

$$\begin{aligned}\mathcal{J}[q] = W_f \int_0^T \int_\Omega \left\{ \frac{\partial q}{\partial t} + L[q(\mathbf{x},t)] - Q(\mathbf{x},t) \right\}^2 d\mathbf{x}dt \\ + W_i \int_\Omega \big\{ q(\mathbf{x},0) - I(\mathbf{x}) \big\}^2 d\mathbf{x} + W_b \int_0^T \big\{ q(\mathbf{0},t) - B(t) \big\}^2 dt. \\ + \sum_{m=1}^{M} w_m \big\{ q(\mathbf{x}_m,t_m) - d_m \big\}^2\end{aligned} \tag{2.5}$$

The weights W_f, W_i, W_b, and w_m are assumed to be the inverses of the error variances in the model dynamics, initial and boundary conditions, and data respectively. Due to the lack of observations at all points in space and time, the error fields $f(\mathbf{x},t)$, $i(\mathbf{x})$ and $b(t)$ are undetermined. However with appropriate weights for $f(\mathbf{x},t)$, $i(\mathbf{x})$ and $b(t)$, there is a unique optimum $\hat{q}(\mathbf{x},t)$ that minimizes the errors $f(\mathbf{x},t)$, $i(\mathbf{x})$, $b(t)$, and $\boldsymbol{\epsilon}_m$ in a weighted least-squares sense. Given that $\mathcal{J}$ is non-negative and quadratic in q, we find the global minimum using calculus of variations. This results in coupled Euler-Lagrange equations (2.6).

The result in this work applies to any linear operator L; however, the Euler-Lagrange equations will change depending on the type of boundary conditions used. Therefore, we omit boundary condition errors from this point to section 3 for generalization purposes.

$$\begin{aligned} -\frac{\partial \lambda}{\partial t} - L^T\lambda(\mathbf{x},t) &= -\sum_{m=1}^{M} w_m(\hat{q}(\mathbf{x}_m,t_m) - d_m)\delta(\mathbf{x}-\mathbf{x}_m)\delta(t-t_m) \\ \lambda(\mathbf{x},T) &= 0 \\ \frac{\partial \hat{q}}{\partial t} + L\hat{q}(\mathbf{x},t) &= Q(\mathbf{x},t) + W_f^{-1}\lambda(\mathbf{x},t) \\ \hat{q}(\mathbf{x},0) &= I(\mathbf{x}) + W_i^{-1}\lambda(\mathbf{x},0) \end{aligned} \tag{2.6}$$

where $\delta(\mathbf{x})$ is the Dirac delta function and the adjont variable $\lambda(\mathbf{x},t) = W_f \left\{ \frac{\partial \hat{q}}{\partial t} + L\hat{q}(\mathbf{x},t) - Q(\mathbf{x},t) \right\}$ and we assume that the operator L has been linearized.

The coupled system of equations (2.6) requires λ to be solved backward in time and the optimal estimate $\hat{q}$ forward in time. We simplify this linear system of equation by using the the representers method [4]. The representer method decouples (2.6) and the optimal solution $\hat{q}$ is expressed as the sum of a first guess q_F and a finite linear

combination of representer functions $r_m(\mathbf{x},t)$,

$$\hat{q}(\mathbf{x},t) = q_F(\mathbf{x},t) + \sum_{m=1}^{M} \beta_m r_m(\mathbf{x},t) \qquad \in \mathbb{R}^1. \tag{2.7}$$

The first guess q_F is the solution to the error-free model (2.1) and the representer functions $r_m(\mathbf{x},t)$ form a reproducig kernel Hilbert space. Substituting (2.7) into the Euler-Lagrange equations we get decoupled systems

$$(F_m) \begin{cases} \frac{\partial r_m}{\partial t} + L r_m(\mathbf{x},t) = W_f^{-1}\alpha_m(\mathbf{x},t) \\ r_m(\mathbf{x},0) = W_i^{-1}\alpha_m(\mathbf{x},0) \end{cases} \tag{2.8}$$

and

$$(B_m) \begin{cases} -\frac{\partial \alpha_m}{\partial t} - L^T\alpha_m(\mathbf{x},t) = \delta(\mathbf{x}-\mathbf{x}_m)\delta(t-t_m) \\ \alpha_m(\mathbf{x},T) = 0 \end{cases} \tag{2.9}$$

where $\alpha_m(\mathbf{x},t)$ is the adjoint corresponding to $r_m(\mathbf{x},t)$. We solve for $\alpha_m(\mathbf{x},t)$ backward in time and then use it to solve forward in time $r_m(\mathbf{x},t)$, $1 \le m \le M$. We find the representer coefficients β_m by substituting (2.7) back into the Euler-Lagrange equations. The result is M equations for unknowns $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_m]^T$ which are written in matrix notation as

$$(\mathbf{R} + \mathbf{W}_d^{-1})^{-1}\boldsymbol{\beta} = \mathbf{h} \qquad \in \mathbb{R}^M \tag{2.10}$$

where $\mathbf{R}$ is the $M \times M$ representer matrix evaluated where there is data, i.e. $\mathbf{R}_{m_1 m_2} = r_1(\mathbf{x}_{m_2}, t_{m_2})$, $\mathbf{h} = \mathbf{d} - \mathbf{q}_{Fm}$, $\mathbf{q}_{Fm} = [q_F(\mathbf{x}_1,t_1), q_F(\mathbf{x}_2,t_2), \cdots, q_F(\mathbf{x}_M,t_M)]^T$ and $\mathbf{W}_d = \mathrm{diag}(w_1, w_2, \ldots, w_M)$ is the data weight matrix. Substituting for β_m in (2.7), the optimal state estimate can be expressed as

$$\hat{q}(\mathbf{x},t) = q_F(\mathbf{x},t) + \mathbf{h}^T\mathbf{P}^{-1}\mathbf{r}(\mathbf{x},t) \tag{2.11}$$

where $\mathbf{P} = \mathbf{R} + \mathbf{W}_d^{-1}$ and $\mathbf{r}(\mathbf{x},t) = [r_1(\mathbf{x},t), r_2(\mathbf{x},t), \cdots, r_M(\mathbf{x},t)]^T$.

The representer method finds the optimal state estimate by solving the linear transport model backward in time M times for $\alpha_m(\mathbf{x},t)$ and forward M times for $r_m(\mathbf{x},t)$. This makes weak constraint 4D-Var more computationally feasible because the optimal adjoint in the state space is determined by the optimal representer coefficients in the data space, which is much smaller than the state space.

3. Reduced Posterior Penalty Functionals. In this section we use the analytical expression for optimal estimates (2.11) found using the representer method to explicitly express the penalty functionals evaluated at the optimum. These expressions for the reduced penalty functionals are necessary for section 4 where we derive methods to estimate model error variance based on regularization parameter selections methods. The results in this section are given in [4].

$$\begin{aligned} \mathcal{J}_{mod}[\hat{q}] = W_f \int_0^T \int_\Omega \left\{ \frac{\partial \hat{q}}{\partial t} + L\hat{q}(\mathbf{x},t) - Q(\mathbf{x},t) \right\}^2 d\mathbf{x}dt \\ + W_i \int_\Omega \left\{ \hat{q}(\mathbf{x},0) - I(\mathbf{x}) \right\}^2 d\mathbf{x} \end{aligned} \tag{3.1}$$

and

$$\mathcal{J}_{data}[\hat{q}] = \sum_{m=1}^{M} w_m \{\hat{q}(\mathbf{x}_m, t_m) - d_m\}^2 \tag{3.2}$$

so that

$$\mathcal{J}[\hat{q}] = \mathcal{J}_{mod}[\hat{q}] + \mathcal{J}_{data}[\hat{q}]$$

.

LEMMA 3.1. *The data penalty satisfies*

$$\hat{\mathcal{J}}_{data} \equiv \mathcal{J}_{data}[\hat{q}] = \mathbf{h}^T \mathbf{P}^{-1} \mathbf{W}_d^{-1} \mathbf{P}^{-1} \mathbf{h} \tag{3.3}$$

Proof. Writing (3.2) in matrix form we have

$$\hat{\mathcal{J}}_{data} = \big(\hat{\mathbf{q}} - \mathbf{d}\big)^T \mathbf{W}_d \big(\hat{\mathbf{q}} - \mathbf{d}\big). \tag{3.4}$$

Subsititing (2.7) into the Euler-Lagrange equations and equating coefficients of the impulses, the representer coefficients $\hat{\boldsymbol{\beta}}$ can be written as $\hat{\boldsymbol{\beta}} = \mathbf{W}_d(\hat{\mathbf{q}} - \mathbf{d})$. Thus

$$\hat{\mathcal{J}}_{data} = \hat{\boldsymbol{\beta}}^T \mathbf{W}_d^{-1} \hat{\boldsymbol{\beta}}. \tag{3.5}$$

Using (2.10), we have

$$\hat{\mathcal{J}}_{data} = \big((\mathbf{R} + \mathbf{W}_d^{-1})^{-1}\mathbf{h}\big)^T \mathbf{W}_d^{-1} \big((\mathbf{R} + \mathbf{W}_d^{-1})^{-1}\mathbf{h}\big) \tag{3.6}$$

Noting that $\mathbf{P} = \mathbf{R} + \mathbf{W}_d^{-1}$ is symmetric, the result then follows. ∎

LEMMA 3.2. *The model penalty satisfies*

$$\hat{\mathcal{J}}_{mod} \equiv \mathcal{J}_{mod}[\hat{q}] = \mathbf{h}^T \mathbf{P}^{-1} \mathbf{R} \mathbf{P}^{-1} \mathbf{h} \tag{3.7}$$

Proof. Substituting $\hat{q}(\mathbf{x}, t)$ (2.11) into the penalty functional (3.1), we have

$$\begin{aligned} \hat{\mathcal{J}}_{mod} = W_f \int_0^T \int_\Omega \Big\{ & \frac{\partial q_F}{\partial t} + L q_F(\mathbf{x}, t) - Q(\mathbf{x}, t) \\ & + \mathbf{h}^T \mathbf{P}^{-1} \left(\frac{\partial \mathbf{r}(\mathbf{x}, t)}{\partial t} + L\mathbf{r}(\mathbf{x}, t) \right) \Big\}^2 d\mathbf{x} dt \\ & + W_i \int_\Omega \big\{ q_F(\mathbf{x}, 0) - I(\mathbf{x}) + \mathbf{h}^T \mathbf{P}^{-1} \mathbf{r}(\mathbf{x}, 0) \big\}^2 d\mathbf{x} \end{aligned} \tag{3.8}$$

Since the first guess $q_F(\mathbf{x}, t)$ is the solution to the error-free model (2.1), we have

$$\begin{aligned} \hat{\mathcal{J}}_{mod} = \mathbf{h}^T \mathbf{P}^{-1} \Big\{ & W_f \int_0^T \int_\Omega \left(\frac{\partial \mathbf{r}}{\partial t} + L\mathbf{r}(\mathbf{x}, t) \right) \left(\frac{\partial \mathbf{r}}{\partial t} + L\mathbf{r}(\mathbf{x}, t) \right)^T d\mathbf{x} dt \\ & + W_i \int_\Omega \mathbf{r}(\mathbf{x}, 0) \mathbf{r}(\mathbf{x}, 0)^T d\mathbf{x} + W_b \int_0^T \mathbf{r}(\mathbf{0}, t) \mathbf{r}(\mathbf{0}, t)^T dt \Big\} \mathbf{P}^{-1} \mathbf{h} \end{aligned} \tag{3.9}$$

Given that $\mathbf{r}(\mathbf{x}, t)$ is the solution to (2.8)

$$\begin{aligned} \hat{\mathcal{J}}_{mod} = \mathbf{h}^T \mathbf{P}^{-1} \Big\{ & W_f^{-1} \int_0^T \int_\Omega \boldsymbol{\alpha}(\mathbf{x}, t) \boldsymbol{\alpha}(\mathbf{x}, t)^T d\mathbf{x} dt \\ & + W_i^{-1} \int_\Omega \boldsymbol{\alpha}(\mathbf{x}, 0) \boldsymbol{\alpha}(\mathbf{x}, 0)^T d\mathbf{x} \Big\} \mathbf{P}^{-1} \mathbf{h} \end{aligned} \tag{3.10}$$

where $\boldsymbol{\alpha}(\mathbf{x},t) = [\alpha_1(\mathbf{x},t), \alpha_2(\mathbf{x},t), \cdots \alpha_M(\mathbf{x},t)]^T$. The adjoint representer function $\alpha_m(\mathbf{x},t)$ for a point measurement $(\mathbf{x}_m, t_m)$ is the Green's function $\gamma(\mathbf{x},t,\mathbf{x}_m,t_m)$ where $\gamma(\mathbf{x},t,\mathbf{y},s)$ satsifies

$$\mathcal{L}^*\gamma(\mathbf{x},t,\mathbf{y},s) \equiv -\frac{\partial\gamma(\mathbf{x},t,\mathbf{y},s)}{\partial t} - L^T\gamma(\mathbf{x},t,\mathbf{y},s) = \delta(\mathbf{x}-\mathbf{y})\delta(t-s) \qquad \gamma(\mathbf{x},T,\mathbf{y},s) = 0 \tag{3.11}$$

The corresponding representer functions $r_m(\mathbf{x},t) = \Gamma(\mathbf{x},t,\mathbf{x_m},t_m)$ where

$$\mathcal{L}\Gamma(\mathbf{x},t,\mathbf{y},s) \equiv \frac{\partial\Gamma(\mathbf{x},t,\mathbf{y},s)}{\partial t} + L\Gamma(\mathbf{x},t,\mathbf{y},s) = W_f^{-1}\gamma(\mathbf{x},t,\mathbf{y},s) \qquad \Gamma(\mathbf{x},0,\mathbf{y},s) = W_i^{-1}\gamma(\mathbf{x},0,\mathbf{y},s) \tag{3.12}$$

Integrating $\int_0^T \int_\Omega \gamma\mathcal{L}\Gamma \, d\mathbf{z}\, dr$ by parts and using dummy variables $\mathbf{z}$ and r, we obtain

$$\int_0^T \int_\Omega \gamma\mathcal{L}\Gamma \, d\mathbf{z}\, dr = \int_\Omega \gamma(\mathbf{z},r,\mathbf{x},t)\Gamma(\mathbf{z},r,\mathbf{y},s)\Big|_0^T d\mathbf{z} + \int_0^T \int_\Omega \Gamma(\mathbf{z},r,\mathbf{x},t)\mathcal{L}^*\gamma(\mathbf{z},r,\mathbf{y},s)\, d\mathbf{z}\, dr \tag{3.13}$$

Substituting $\mathcal{L}^*\gamma(\mathbf{z},r,\mathbf{y},s) = \delta(\mathbf{z}-\mathbf{y})\delta(r-s)$ we obtain

(3.14)

$$\Gamma(\mathbf{x},t,\mathbf{y},s) = \int_0^T \int_\Omega \gamma(\mathbf{z},r,\mathbf{x},t)\mathcal{L}\Gamma(\mathbf{z},r,\mathbf{y},s)\, d\mathbf{z}\, dr - \int_\Omega \gamma(\mathbf{z},r,\mathbf{x},t)\Gamma(\mathbf{z},r,\mathbf{y},s)\Big|_0^T d\mathbf{z}$$

Substituting for $\mathcal{L}\Gamma(\mathbf{z},r,\mathbf{y},s)$ and $\Gamma(\mathbf{z},0,\mathbf{y},s)$ and evaluating at the data points, we have

$$\Gamma(\mathbf{x}_m,t_m,\mathbf{x}_l,t_l,) = W_f^{-1}\int_0^T \int_\Omega \gamma(\mathbf{z}_m,r_m,\mathbf{x}_m,t_m)\gamma(\mathbf{z}_m,r_m,\mathbf{x}_l,t_l,)d\mathbf{z}dr + W_i^{-1}\int_\Omega \gamma(\mathbf{z}_m,0,\mathbf{x}_m,t_m,)\gamma(\mathbf{z}_m,0,\mathbf{x}_l,t_l,)d\mathbf{z} \tag{3.15}$$

$$\begin{aligned}\Gamma(\mathbf{x}_m,t_m,\mathbf{x}_l,t_l) &= W_f^{-1}\int_0^T \int_\Omega \alpha_l(\mathbf{x}_m,t_m)\alpha_l(\mathbf{x}_m,t_m)d\mathbf{x}dt \\ &\quad + W_i^{-1}\int_\Omega \alpha_l(\mathbf{x}_m,0)\alpha_l(\mathbf{x}_m,0)d\mathbf{x} \\ &= \mathbf{R}_{lm}\end{aligned} \tag{3.16}$$

If we consider for all l, m we have

$$\mathbf{R} = W_f^{-1}\int_0^T \int_\Omega \boldsymbol{\alpha}(\mathbf{x},t)\boldsymbol{\alpha}(\mathbf{x},t)^T d\mathbf{x}dt + W_i^{-1}\int_\Omega \boldsymbol{\alpha}(\mathbf{x},0)\boldsymbol{\alpha}(\mathbf{x},0)^T d\mathbf{x} \qquad \square \tag{3.17}$$

THEOREM 3.3. *The posterior functional $\hat{\mathcal{J}}$ satisfies*

$$\hat{\mathcal{J}} \equiv \mathcal{J}[\hat{q}] = \mathbf{h}^T\mathbf{P}^{-1}\mathbf{h} \tag{3.18}$$

Proof. Using lemmas (3.1) and (3.2) we have

$$\begin{aligned}\hat{\mathcal{J}} &= \mathbf{h}^T\mathbf{P}^{-1}\mathbf{W}_d^{-1}\mathbf{P}^{-1}\mathbf{h} + \mathbf{h}^T\mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-1}\mathbf{h} \\ &= \mathbf{h}^T\mathbf{P}^{-1}\big(\mathbf{R} + \mathbf{W}_d^{-1}\big)\mathbf{P}^{-1}\mathbf{h} \\ &= \mathbf{h}^T\mathbf{P}^{-1}\mathbf{h}\end{aligned} \tag{3.19}$$

∎

4. Weak Constraint as Regularization. In this section, we frame weak-constraint 4D-Var data assimilation as a Tikhonov regularization problem. This approach allows us to treat model error variance estimation as a regularization parameter selection problem. We use three established regularization parameter selection methods i.e. the L-curve[15], the χ^2 method [4, 23], and the generalized cross-validation (GCV) method [1, 13] to estimate the model error variance.

4.1. Tikhonov regularization. Consider the parameter estimation problem

$$\mathbf{d} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}, \tag{4.1}$$

where $\mathbf{d}$ is the observational data, $\mathbf{A}$ is the operator mapping the model parameters to the observation space, and $\boldsymbol{\epsilon}$ accounts for the errors in the data. We assume a first guess, $\mathbf{x_0}$, for the parameters, such that:

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{f}, \tag{4.2}$$

where $\mathbf{f}$ represents the unknown errors in the first guess. In Tikhonov regularization, the goal is to find $\mathbf{x}$ that minimizes errors in both the data and the first guess in a weighted least-squares sense. This is expressed as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \big\{(\mathbf{A}\mathbf{x} - \mathbf{d})^T\mathbf{W}_d(\mathbf{A}\mathbf{x} - \mathbf{d}) + (\mathbf{x} - \mathbf{x}_0)^T\mathbf{W}_f(\mathbf{x} - \mathbf{x}_0)\big\}. \tag{4.3}$$

Here, $\mathbf{W}_d$ and $\mathbf{W}_f$ are weighting matrices that account for the relative uncertainties in the data and the first guess respectively. We express $\mathbf{W}_f$ as $\lambda\mathbf{I}$, where λ is the regularization parameter. The optimal parameter estimates, $\hat{\mathbf{x}}$, that minimize (4.3) are given by:

$$\hat{\mathbf{x}} = \mathbf{x}_0 + (\mathbf{A}^T\mathbf{W}_d\mathbf{A} + \mathbf{W}_f)^{-1}\mathbf{A}^T\mathbf{W}_d(\mathbf{d} - \mathbf{A}\mathbf{x}_0), \tag{4.4}$$

In the context of weak constraint data assimilation, the optimal state estimate takes on the form

$$\begin{aligned}\hat{q} = \arg\min_{q} \Bigg\{ &\sum_{m=1}^{M} w_m\big[q(\mathbf{x}_m, t_m) - d_m\big]^2 + W_i\int_\Omega \big[q(\mathbf{x}, 0) - I(\mathbf{x})\big]^2 d\mathbf{x} \\ &+ W_b\int_0^T \big[q(\mathbf{0}, t) - B(t)\big]^2 dt + W_f\int_0^T\int_\Omega \left[\frac{\partial q}{\partial t} + L[q(\mathbf{x}, t)] - Q(\mathbf{x}, t)\right]^2 d\mathbf{x}dt\Bigg\}.\end{aligned} \tag{4.5}$$

This formulation can be viewed as a Tikhonov regularization problem, where the regularization term reflects the weakly constrained model dynamics. Using the representer method, we can express the optimal state estimate in a closed form, similar to the parameter estimation problem i.e.

$$\hat{\mathbf{q}}(\mathbf{x}, t) = \mathbf{q}_F(\mathbf{x}, t) + \mathbf{r}(\mathbf{x}, t)^T(\mathbf{R} + \mathbf{W}_d^{-1})^{-1}(\mathbf{d} - \mathbf{q}_{Fm}). \tag{4.6}$$

In this formulation, we assume that the weight $\mathbf{W}_d = diag\big(w_1, w_2, \cdots, w_M\big)$, to the data is know, as it is typically derived from the characteristics of the data, such as measurement errors provided by the instruments or observational systems. These errors are either measured directly or inferred based on standard error models for the specific application(e.g., Gaussian noise with a known variance). Additionally, we assume that the weights W_i and W_b, applied to the initial and boundary conditions respectively, are also known or are assumed to be exact. In future work, we plan to estimate weights W_i and W_b using regularization parameter selection methods, along with the model dynamics error. The only unknown in this case is the weight W_f, which governs the influence of the model dynamics and we find it using regularization parameter selection methods.

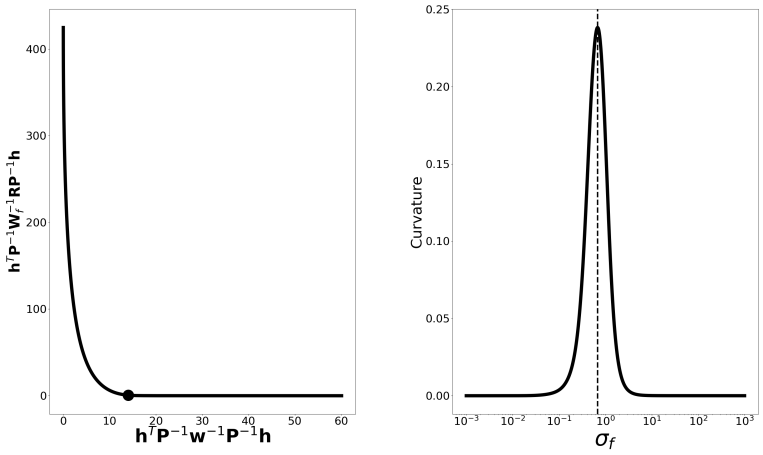
4.2. Model dynamics error estimate. When applying regularization techniques, such as Tikhonov regularization, the challenge lies in determining the optimal regularization parameter λ. This parameter controls the trade-off between fitting the observational data and ensuring that the solution remains smooth and physically meaningful. If the parameter is too small, the solution may overfit the noisy data, leading to instability and unrealistic results. On the other hand, if the parameter is too large, the solution becomes overly smooth, potentially ignoring important features in the data.

To address this, we use regularization parameter selection methods such as the L-curve, generalized cross-validation, and the χ^2 method to find the optimal regularization parameter. These methods provide a systematic way of selecting the regularization parameter that balances the data fit and the solution's smoothness, leading to more stable and accurate parameter estimates. In the standard Tikhonov regularization problem, regularization parameter selection involves finding λ such that equation (4.3) holds and in this case the data misfit is often scaled with $\mathbf{W}_d = \sigma_d^{-2}\mathbf{I}$. Given the formulation of the weak constraint data assimilation as Tikhonov regularization, we can use these regularization parameter selection methods to choose the model dynamics weights W_f in data assimilation. These weights are effective estimates of the error variance in the model dynamics since we assume $W_f = \sigma_f^{-2}$.

4.2.1. L-curve. The L-curve is a graphical technique where the norm of the regularized solution is plotted against the residual norm for a range of regularization parameters. The name 'L-curve' refers to the characteristic L-shaped curve that forms for linear problems. The optimal regularization parameter is the one that corresponds to the point of maximum curvature, i.e., the 'corner' of the L. The regularization parameter at this point is optimal in the sense that the errors in the weighted parameter misfit and data misfit are balanced.

For the standard Tikhonov regularization problem, the L-curve is obtained by plotting $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$ against $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{d}\|_2^2$, with the regularization parameter λ corresponding to the corner of the curve. **In weak constraint 4D-Var, the L-curve is obtained by plotting the regularization norm $W_f\mathcal{J}_{mod}[\hat{q}]$ against the data misfit $\mathcal{J}_{data}[\hat{q}]$.** When using the representer functions, we can use the matrix expression for the regularization norm in Lemma 3.2 and data misfit in Lemma 3.1 to plot the L-curve.

The L-curve approach provides a visually intuitive means of selecting the regularization parameter however, it may not always yield a distinct corner, and its interpretation can be subjective, relying on visual inspection.



(a) L-curve. (b) Curvature of the L-curve

Fig. 1: Illustrative example of model error estimation for 4D-Var with representers using L-curve and the corresponding curvature of the L-curve as a function of regularization parameter W_f.

Figure 1(a) shows a typical L-curve, where the regularization norm (model error) is plotted against the data misfit for different values of model error variance σ_f^2. The black dot marks the point of maximum curvature, often indicating the optimal regularization parameter. Figure 1(b) provides additional insight by plotting the curvature of the L-curve as a function of σ_f, with the peak corresponding to the optimal model error variance. However, as noted by Hansen (1999), the L-curve corner doesn't always align with the maximum curvature, especially when the regularization norm doesn't rise sharply for regularization parameters beyond the optimal σ_f^2.

4.2.2. Generalized cross-validation (GCV). The GCV method is a powerful technique for selecting the regularization parameter in Tikhonov regularization. The key idea behind GCV is to evaluate how well the regularized solution would generalize to new data by systematically leaving out individual observations and measuring the resulting prediction error. This approach ensures that the optimal regularization parameter chosen minimizes the trade-off between fitting the data and maintaining a smooth, stable solution. GCV offers an efficient way to select the optimal regularization parameter by minimizing the prediction error for unseen data.

In the standard Tikhonov regularization problem, the GCV criterion is typically formulated as a function of the regularization parameter and aims to find the value of λ that minimizes the GCV function

$$g(\lambda) = \frac{M||\mathbf{A}\hat{\mathbf{x}} - \mathbf{d}||_2^2}{\mathrm{Tr}(\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{W}_d\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{W}_d)^2} \ . \tag{4.7}$$

Extending this GCV concept to weak constraint 4D-Var data assimilation, we derive a GCV function that can be used to select the regularization parameter $W_f = \sigma_f^{-2}$ associated with the error in the model dynamics.

THEOREM 4.1. *Let* $\mathbf{W}_d = diag\big(w_1, w_2, \cdots, w_M\big)$ *and assume that* W_i *and* W_b *are specified. The generalized cross validation function for weak constraint 4D-Var with representers is*

$$g(\sigma_f) = M \frac{\mathbf{h}^T \mathbf{P}^{-1} \mathbf{W}_d^{-1} \mathbf{P}^{-1} \mathbf{h}}{Tr(\mathbf{I} - \mathbf{R}(\mathbf{R} + \mathbf{W}_d^{-1})^{-1})^2} \ . \tag{4.8}$$

Proof. Consider a modified weak constraint 4D-Var data assimilation problem where we leave out a data point d_k:

$$\min_q \Bigg\{ \sum_{m \neq k}^{M} w_m \big[q(\mathbf{x}_m, t_m) - d_m\big]^2 + W_i \int_\Omega \big[q(\mathbf{x}, 0) - I(\mathbf{x})\big]^2 d\mathbf{x} + W_b \int_0^T \big[q(\mathbf{0}, t) - B(t)\big]^2 dt + W_f \int_0^T \int_\Omega \left[\frac{\partial q}{\partial t} + L[q(\mathbf{x}, t)] - Q\right]^2 d\mathbf{x} dt \Bigg\}. \tag{4.9}$$

Let $\hat{q}^{[k]}(x, t)$ be the solution to the optimization problem (4.9), where the superscript k indicates that d_k was left out of the computation. In the leave-one-out approach, we select the regularization parameter $W_f = \sigma_f^{-2}$ that minimizes the sum of predictive errors over all k:

$$\min_{\sigma_f} g(\sigma_f) \equiv \min_{\sigma_f} \left\{ \frac{1}{M} \sum_{k=1}^{M} w_k \big(\hat{q}^{[k]}(\mathbf{x}_k, t_k) - d_k\big)^2 \right\}. \tag{4.10}$$

Computing $g(\sigma_f)$ involves solving M problems of the form (4.9) and as described in [13], we can speed up this computation by using the leave-one-out lemma. In the case of weak constraint 4D-Var, note that $\hat{q}^{[k]}(x, t)$ optimizes

$$\min_q \Bigg\{ w_k \big(q(\mathbf{x}_k, t_k) - \tilde{d}_k\big)^2 + \sum_{m \neq k} w_m \big(q(\mathbf{x}_m, t_m) - \tilde{d}_m\big)^2 + W_i \int_\Omega \big[q(\mathbf{x}, 0) - I(\mathbf{x})\big]^2 d\mathbf{x} + W_b \int_0^T \big[q(\mathbf{0}, t) - B(t)\big]^2 dt + W_f \int_{T_0}^T \int_\Omega \left[\frac{\partial q}{\partial t} + L[q(\mathbf{x}, t)] - Q(\mathbf{x}, t)\right]^2 d\mathbf{x}\, dt \Bigg\} \tag{4.11}$$

where

$$\tilde{d}_m = \begin{cases} \hat{q}^{[k]}(\mathbf{x}_k, t_k) & m = k \\ d_m & m \neq k \end{cases}. \tag{4.12}$$

Define $\hat{\mathbf{q}}_m = \begin{bmatrix} \hat{q}(\mathbf{x}_1, t_1) & \hat{q}(\mathbf{x}_2, t_2) & \cdots & \hat{q}(\mathbf{x}_M, t_M) \end{bmatrix}^T$ then

$$\hat{\mathbf{q}}_m = \mathbf{q}_{Fm} + \mathbf{R}\mathbf{P}^{-1}(\mathbf{d} - \mathbf{q}_{Fm}) \text{ and } \hat{\mathbf{q}}_m^{[k]} = \mathbf{q}_{Fm} + \mathbf{R}\mathbf{P}^{-1}(\tilde{\mathbf{d}} - \mathbf{q}_{Fm}).$$

Consider $\hat{\mathbf{q}}_m^{[k]} - \hat{\mathbf{q}}_m = \mathbf{R}\mathbf{P}^{-1}(\tilde{\mathbf{d}} - \mathbf{d})$ so that

$$\hat{q}^{[k]}(\mathbf{x}_k, t_k) - \hat{q}(\mathbf{x}_k, t_k) = (\mathbf{R}\mathbf{P}^{-1}\tilde{\mathbf{d}})_k - (\mathbf{R}\mathbf{P}^{-1}\mathbf{d})_k. \tag{4.13}$$

Since $\tilde{d}_m = d_m$ where $m \neq k$ and

$$(\mathbf{RP}^{-1}\tilde{\mathbf{d}})_k = \sum_{\substack{m=1 \\ m\neq k}}^{M} (\mathbf{RP}^{-1})_{km} d_k + (\mathbf{RP}^{-1})_{kk}\tilde{d}_k, \tag{4.14}$$

we have that

$$\hat{q}^{[k]}(\mathbf{x}_k, t_k) - \hat{q}(\mathbf{x}_k, t_k) = (\mathbf{RP}^{-1})_{kk}(\tilde{d}_k - d_k). \tag{4.15}$$

Now

$$\frac{\tilde{d}_k - d_k - \hat{q}^{[k]}(\mathbf{x}_k, t_k) + \hat{q}(\mathbf{x}_k, t_k)}{\tilde{d}_k - d_k} = 1 - (\mathbf{RP}^{-1})_{kk} \tag{4.16}$$

and since $\hat{q}^{[k]}(\mathbf{x}_k, t_k) = \tilde{d}_k$, we have

$$\frac{\hat{q}(\mathbf{x}_k, t_k) - d_k}{\hat{q}^{[k]}(\mathbf{x}_k, t_k) - d_k} = 1 - (\mathbf{RP}^{-1})_{kk}. \tag{4.17}$$

This means we do not have to solve the data assimilation problem k times because

$$\hat{q}^{[k]}(\mathbf{x}_k, t_k) - d_k = \frac{\hat{q}(\mathbf{x}_k, t_k) - d_k}{1 - (\mathbf{RP}^{-1})_{kk}}. \tag{4.18}$$

Substituting for $\hat{q}^{[k]}(\mathbf{x}_k, t_k) - d_k$ in (4.10), we find the optimal $W_f = \sigma_f^2$ by minimizing

$$g(\sigma_f) = \frac{1}{M}\sum_{k=1}^{M} w_k \left(\frac{\hat{q}(\mathbf{x}_k, t_k) - d_k}{1 - (\mathbf{RP}^{-1})_{kk}} \right)^2. \tag{4.19}$$

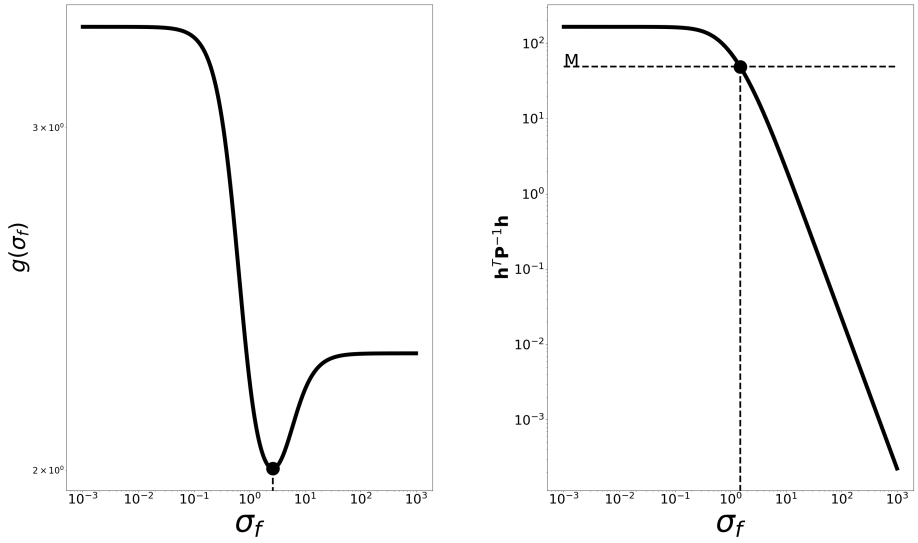
Simplifying further, we replace $(\mathbf{RP}^{-1})_{kk}$ with it's average value and we minimize

$$g(\sigma_f) = M \frac{(\hat{\mathbf{q}}_m - \mathbf{d})^T \mathbf{W}_d (\hat{\mathbf{q}}_m - \mathbf{d})}{\mathrm{Tr}(\mathbf{I} - \mathbf{RP}^{-1})^2} \tag{4.20}$$

$$= \frac{M\mathcal{J}_{\mathrm{data}}[\hat{q}]}{\mathrm{Tr}(\mathbf{I} - \mathbf{RP}^{-1})^2}. \tag{4.21}$$

Applying Lemma 3.1, the result follows. ∎

Figure 2(a) demonstrates the role of the GCV method in identifying the optimal model error variance, σ_f^2. The GCV method evaluates how well the model and observational data are balanced, minimizing prediction error by systematically adjusting σ_f^2. As seen in Figure 2(a), the GCV function $g(\sigma_f)$ is plotted against σ_f, with a clear minimum indicated by the black dot. This minimum represents the optimal value of σ_f, ensuring that the regularization is neither too strong nor too weak, resulting in a solution that effectively balances the data fit and the regularization of model dynamics.



(a) GCV curve (b) χ^2 curve

Fig. 2: Illustrative examples of model error variance estimates for 4D-Var with representers using (a) GCV where $g(\sigma_f)$ is given by (4.1) and (b) χ^2 method.

4.2.3. χ^2 method. The χ^2 method uses a statistical test to select the regularization parameter based on the assumption that the minimized objective function should approximate a chi-square distribution with M degrees of freedom, where M is the number of observations. Rather than check if the residual passes the χ^2 test after the regularization parameter is chosen, the test is used to solve for the regularization parameter. For ill-posed inverse problem $\mathbf{Ax} = \mathbf{d}$, it is shown in [23] that

$$\mathcal{J} \equiv (\mathbf{A}\hat{\mathbf{x}} - \mathbf{d})^T \mathbf{W}_d (\mathbf{A}\hat{\mathbf{x}} - \mathbf{d}) + (\hat{\mathbf{x}} - \mathbf{x}_0)^T \mathbf{W}_f (\hat{\mathbf{x}} - \mathbf{x}_0) \sim \chi^2_M, \tag{4.22}$$

for large enough M, even if the data errors are not normally distributed. In standard Tikhonov regularization, the regularization parameter is chosen so that

$$\mathcal{J}[\lambda] \approx M. \tag{4.23}$$

An efficient algorithm for the χ^2 method was developed in [24]. For weak constraint 4D-Var data assimilation with representers, $W_f = \sigma_f^{-2}$ is chosen such that the cost function $\mathcal{J}[\hat{q}]$ in Theorem 3.3 also follows a χ^2 random variable with M degrees of freedom [4], i.e.

$$\mathcal{J}[\hat{q}] = \mathbf{h}^T \mathbf{P}^{-1} \mathbf{h} \sim \chi^2_M. \tag{4.24}$$

It was shown in [24] that for a convergent optimization algorithm, the minimized cost function is monotonically decreasing as a function of σ_f. Figure 2 (b) gives an illustrative example. Note that the optimal regularization parameter does not

minimize $\mathcal{J}[\hat{q}]$ due to the errors in the dynamics and data. The optimal value of the minimized cost function is the number of data, $M = 49$. Unlike previous approaches such as in [5], where the model error covariance is first specified and then tested to see if it passes the χ^2, in this case, we select the model error variance that passes the χ^2 test and use it to weight the model dynamics during the data assimilation process.

5. Numerical Experiments. The numerical experiments estimate $q(\mathbf{x}, \mathbf{t})$, representing wildfire smoke $PM_{2.5}$ concentration, using a one-dimensional transport equation with simulated observational data. The goal is to find model error estimates that properly weigh the model given the data error weight. We use the regularization parameter selection methods described in section 4 to find the optimal values of the model error weight $W_f = {\sigma_f}^2$, which are then integrated into the data assimilation algorithm.

5.1. Model and data setup.

5.1.1. Transport model. The $PM_{2.5}$ concentration $q(x,t)$ is modeled by a 1D transport equation with an exponential source term $Q(x,t)$ representing wildfire emissions:

$$\begin{cases} \frac{\partial q}{\partial t} + u\frac{\partial q}{\partial x} = Q(x,t) \quad \text{for } x \in [30,45], \quad t \in [0,20] \\ q(x,0) = 0 \\ q(0,t) = B(t) \end{cases} \tag{5.1}$$

where u is a constant wind field. Since the main focus of this paper is estimating the error variance of the model dynamics, we assume that the initial and boundary conditions are exact. The source term $Q(x,t)$ is designed to model the emission of $PM_{2.5}$ from wildfire sources and is defined as the sum of two Gaussian functions. Each function represents a distinct wildfire source, characterized by its emission strength, spatial and temporal decay rates, and the location. Specifically, the source term is given by:

$$Q(x,t) = S_0 \exp(-\alpha_0(x - x_0)^2 - k_0 t) + S_1 \exp(-\alpha_1(x - x_1)^2 - k_1 t) \tag{5.2}$$

Here S_0 and S_1 denote the initial emission strengths of the two sources, while α_0, α_1 are the spatial decay rates, controlling the rate at which the $PM_{2.5}$ concentration diminishes with distance from the source locations $x_0 = 33$ and $x_1 = 40$. The parameters k_0 and k_1 represent the temporal decay rates, dictating how the emission intensity decreases over time. This formulation allows the model to capture the temporal and spatial dynamics of $PM_{2.5}$ dispersion resulting from multiple wildfire sources.

5.1.2. Upwind finite volume method. We solve the transport model using the upwind finite volume method. To solve the transport equation (5.1) using this method, the domain is divided into a series of control volumes or cells. Let x_i represent the center of the i-th cell width Δx and $q_i(t)$ represent the average value of $q(x,t)$ over this cell [19]. The finite volume method integrates the transport equation over each cell $\left[x_{i+\frac{1}{2}}, x_{i-\frac{1}{2}}\right]$:

$$\frac{d}{dt}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x,t)dx + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u\frac{dq}{dx}dx = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} Q(x,t)dx. \tag{5.3}$$

The first term, which represents the time derivative, can be approximated as:

$$\frac{d}{dt}\big(q_i(t)\Delta x\big) = \Delta x\frac{dq_i(t)}{dt}. \tag{5.4}$$

By applying Gauss's theorem, the advection term $u\frac{dq}{dx}$, is given as

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u\frac{dq}{dx}dx = u\bigg(q(x_{i+\frac{1}{2}},t) - q(x_{i-\frac{1}{2}},t)\bigg). \tag{5.5}$$

This scheme accounts for the direction of the wind field u which is key for solving the forward (2.8) and backward (2.9) systems that result from using the method of representers. If the wind field $u > 0$, the advection term at the interface $x_{i+\frac{1}{2}}$ is influenced by the upstream value (left side of the interface). Conversely, if $u < 0$, it is influenced by the downstream value (right side of the interface). Thus, the flux at the interface $x_{i+\frac{1}{2}}$ can be written as:

$$F_{i+\frac{1}{2}} = \begin{cases} uq_i, & \text{if } \ u > 0 \\ uq_{i+1}, & \text{if } \ u < 0. \end{cases} \tag{5.6}$$

Substituting this into the integral form of the transport equation (5.3) and dividing by Δx, the discrete form becomes:

$$\frac{dq_i(t)}{dt} = -\frac{1}{\Delta x}\bigg(F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}} + Q_i\bigg), s \tag{5.7}$$

where Q_i is the discretized source term over the control volume. Δt is chosen so that the Courant-Friedrichs-Lewy (CFL) condition is satisified.

5.1.3. Observations. Simulated observations d_m are generated at space-time locations (x_m,t_m) by adding white noise to the true state found by solving (5.1)

$$\mathbf{d}_m = q(x_m, t_m) + \boldsymbol{\epsilon}_m \qquad \boldsymbol{\epsilon}_m \sim \mathcal{N}(0, \sigma_m^2). \tag{5.8}$$

The simulated observational data noise varies in space and time, with its variance σ_m^2 proportional to the true concentration, i.e., $\sigma_m = \sigma q(x_m, t_m)$. This ensures that regions with higher $PM_{2.5}$ levels exhibit greater variability and captures the inherent uncertainty and measurement errors typical of environmental monitoring systems. The space and time locations of the observations are chosen randomly to from a uniform distribution over the domain $[30, 45]$ and time period $[0, 20]$.

5.1.4. First guess. The first guess $q_F(x,t)$ for the $PM_{2.5}$ concentration is obtained by solving the transport equation using perturbed parameter values in the source term. These perturbations are applied to the spatial decay rates, and temporal decay rates of the wildfire sources, introducing discrepancies between the first guess and the true solution. Specifically, the source term for the first guess is defined with slightly different parameters α_{F0}, α_{F1}, k_{F0} and k_{F1} compared to the true values with $\alpha_{Fi} \sim \mathcal{N}(\alpha_i, \sigma_{\alpha_i}^2)$ and $k_{Fi} \sim \mathcal{N}(k_i, \sigma_{k_i}^2)$ for $i = 0, 1$. i.e.

$$Q_F(x,t) = S_0 \exp(-\alpha_{F0}(x - x_0)^2 - k_{F0}t) + S_1 \exp(-\alpha_{F1}(x - x_1)^2 - k_{F1}t). \tag{5.9}$$

The first guess serves as a baseline from which the data assimilation process begins, highlighting the need for adjustments based on observational data to achieve more accurate state estimates.

5.2. Experiment setup. The experimental setup is designed to evaluate the performance of model error estimation in weak constraint 4D-Var data assimilation under different data noise levels and errors in the dynamical model. The experiments are classified into two categories based on whether the simulated observational data or first guess is trusted more. In the first category, experiments 1 and 2 in Table 1, the first guesses are more accurate, representing situations where the model is more trusted than the data. In the second category, experiments 3 and 4 in Table 1, the simulated observational data is closer to the true $PM_{2.5}$ solution reflecting scenarios where the data are more reliable than the model.

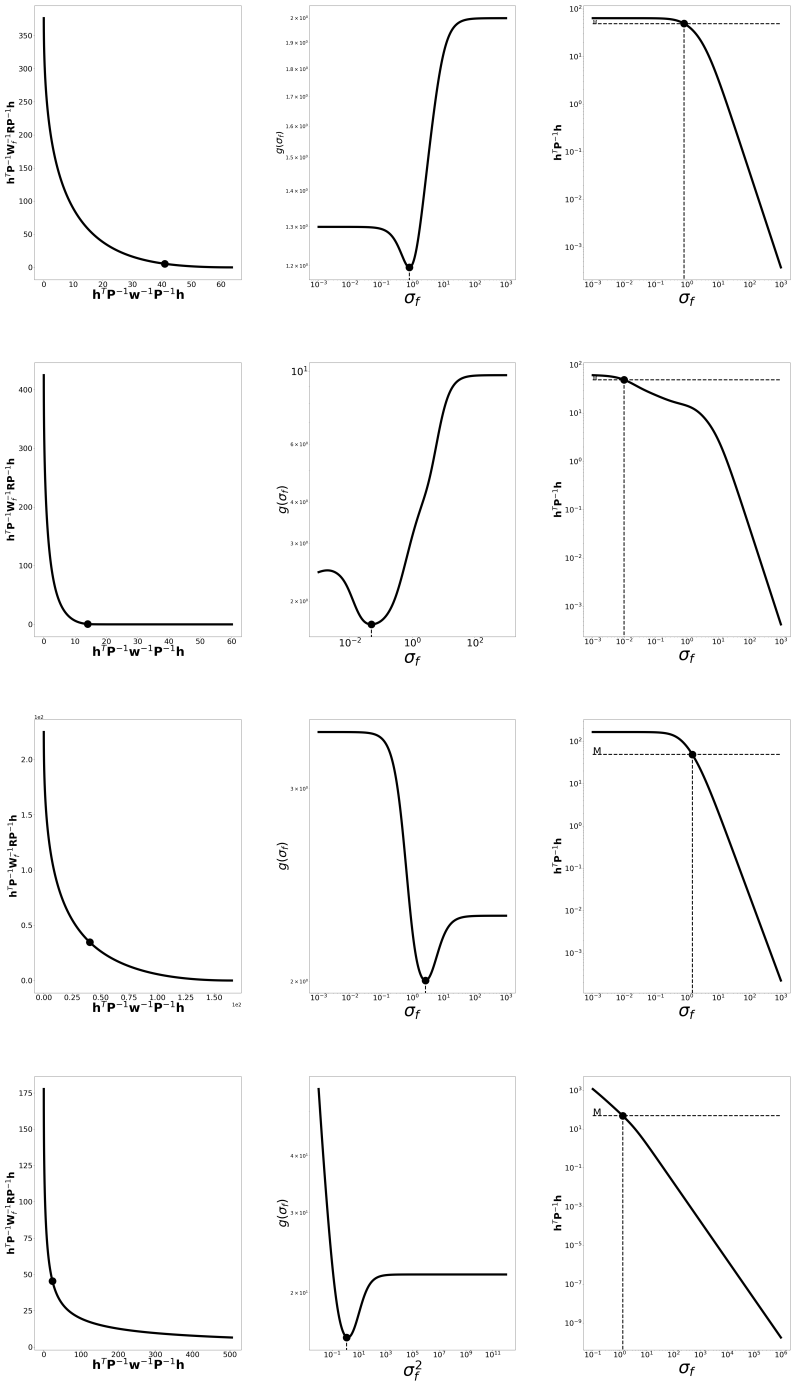
Exper.	BC	S_1	k_1	α_1	σ	σ_{k_0}	σ_{k_1}	σ_{α_0}	σ_{α_1}
1	periodic	0	0	0	0.7	0.2	0	0.2	0
2	no flux	50	0.25	5	0.6	0.2	0.2	0.2	0.2
3	periodic	0	0	0	0.3	0.5	0	0.7	0
4	no flux	50	0.25	5	0.2	0.6	0.5	0.5	0.5

Table 1: Parameters and their corresponding standard deviations used in numerical experiments 1-4.

Table 1 shows the parameter values used in the experiments, including boundary conditions(BC), source emission strength, decay constants, noise levels, and perturbation magnitudes for the first guess. The parameters for the first source of $PM_{2.5}$ emissions in (5.2) are held constant across all experiments with $S_0 = 100$, $k_0 = 0.5$, and $\alpha_0 = 10$. Perturbations are applied to the parameters k_0, k_1, α_0, and α_1 in the first guess to introduce discrepancies between the model and the true solution, with σ_{k_0} and σ_{α_0} corresponding to the perturbations in the temporal and spatial decay rates of the first source, and σ_{k_1} and σ_{α_1} corresponding to the second source. σ represents the noise level applied to the observational data.

5.3. Model error estimation. In section 4 we presented three methods for estimating the transport model error variance, σ_f^2. This parameter balances the influence of the model dynamics and the observational data in the assimilation process. Figure 3 illustrates the L-curves, GCV curves, and χ^2 method curves for each experiment, showing the process of estimating the model error standard deviation. Each column of subplots corresponds to one of the regularization methods and highlights the specific value that corresponds to estimated ${\sigma_f}^2$. Each row corresponds to an experiment with experiment 1 at the top and 4 at the bottom. The black dot indicates where σ_f^2 was chosen for each method. The corresponding values for σ_f^2 are given in Table 2.

For the L-curve method, the values of ${\sigma_f}^2$ at the corner of the L-curve are not necessarily the points of maximum curvature. As discussed in [15], the L-curve criterion only pinpoints the optimal regularization parameter at the curve's corner when the regularization norm increases immediately as σ_f^2 becomes smaller than the optimal variance ${\sigma_f}^2$ which isn't the case for our experiments. The values obtained at the point of maximum curvature provided better optimal estimates across all experiments, and these are the values shown on the L-curves in Figure 3 and listed under the L-curve row in Table 2.



(a) L-curve (b) GCV curve (c) χ^2 curve

Fig. 3: Model error standard deviations selected by regularization parameter selection methods. Each row represents an experiment with experiment 1 at the top and 4 at the bottom. The first column is L-curve, second is generalized cross validation and the third is χ^2 method.

Experiment	1	2	3	4
L-curve	0.6163	0.4546	0.9863	1.0424
GCV	0.6526	0.00237	6.9549	1.5107
χ^2	0.6764	$9.7616e^{-5}$	2.2348	1.6697

Table 2: ${\sigma_f}^2$ estimates from the three regularization parameter selection methods.

Table 2 gives the model error variances estimated by each regularization parameter selection method for each experiment. In experiments 1 and 2, the first guess is closer to the true solution than the observational data and in experiments 3 and 4 the observational data is more accurate than the first guess. In experiment 1, the values for ${\sigma_f}^2$ are relatively similar across all regularization parameter selection methods methods. In experiment 2, the ${\sigma_f}^2$ values vary significantly. The χ^2 method yields the smallest variance while the the L-curve method gives the highest variance. This difference reflects a decreased reliance on the model dynamics in favor of data correction with the L-curve method and more trust in the model dynamics with the χ^2 method. For experiment 3, the estimated model error variances differ considerably, with the L-curve method producing the lowest value and and the GCV method the highest. This suggests that the L-curve method gives more weight to the model than other methods, while the GCV method strongly favors the observational data. In experiment 4, the values for σ_f^2 are similar, but the L-curve method still provides the smallest variance followed by the GCV and χ^2 methods. There is a minimal difference between the GCV and χ^2 methods which implies a slightly greater alignment between them in weighting the model dynamics.

Across all regularization parameter selection methods, the σ_f^2 estimates from experiments 1 and 2 are generally lower than those from experiments 3 and 4. This suggests that when the model is more accurate (experiments 1 and 2), the regularization parameter selection methods assign lower σ_f^2 values, thereby correctly identifying that we should rely more on model dynamics. In contrast, when the observational data is more reliable (experiments 3 and 4), the higher σ_f^2 values correctly indicate we should rely more heavily on the data. However, when σ_f^2 becomes too large, particularly with the GCV method in experiment 3, there is a problem with over-reliance on data. There are 49 data and 89000 state estimates, so the problem is very under determined. If σ_f^2 is too large, the problem is ill-posed and we will see in subsection 5.4 that this results in unrealistic oscillations in the state estimates.

5.4. Optimal $PM_{2.5}$ estimates. After obtaining σ_f^2 in Table 2 for each experiment using the three regularization parameter selection methods, we use those model error variances to do data assimilation in order to obtain $PM_{2.5}$ concentration estimates as shown in the results below.

5.4.1. Experiment 1. This experiment was set up with the first guess closely approximating the true $PM_{2.5}$ concentration, more so than the simulated observational data. Figure 4 shows the spatial $PM_{2.5}$ concentrations at four distinct times during the transport process. In this experiment, there is only one source at $x = 33$ and periodic boundary conditions were used. We note that the assimilated estimates from the three regularization parameter selection methods behave similar across space and time.

At $t = 0.13$, the assimilated state estimates, first guess, and true concentration values are closely aligned across all spatial locations. The assimilated estimates fit

the first and last data points, which are in agreement with the first guess at their respective locations. This suggests that, early in the transport process, the assimilated estimates follow the first guess almost exactly, showing no significant deviation from it. At $t = 9.03$, the first data point still aligns well with the first guess and the true concentration. However, the assimilated estimates do not fit the first data point because with global optimization, they attempt to match the second data point. At this point in time the impulse has moved to the right and increased. The assimilated estimates correctly follow the first guess rather than fit the data. At $t = 14.97$, the assimilated estimates continue to match the first guess between $x = 30$ and $x = 34$, then try to fit the data. However, the assimilated estimates fail to fit these data points and more heavily rely on the first guess. At $t = 17.93$, we continue to see that the assimilated estimates make an attempt to fit the data but generally rely on the first guess, ultimately giving good estimates of the solution.

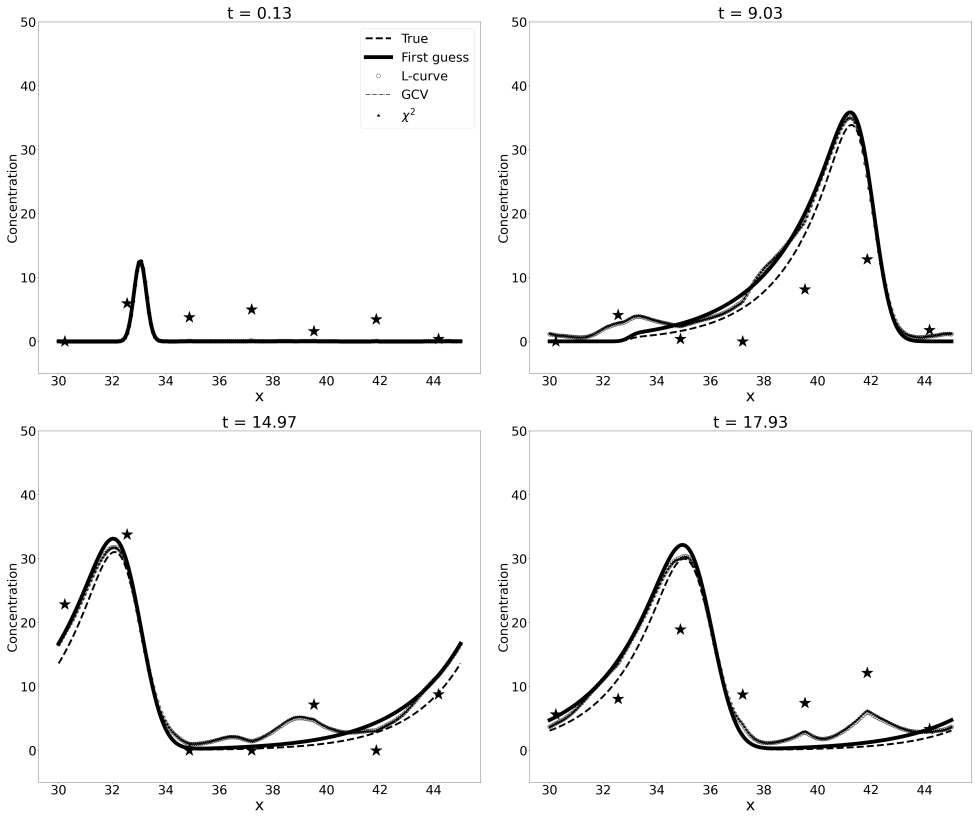


Fig. 4: $PM_{2.5}$ concentration as a function of space for experiment 1

5.4.2. Experiment 2. This experiment was set up similarly to experiment 1, but the key difference in this experiment is that it has two sources of $PM_{2.5}$ located at $x = 33$ and $x = 40$, and no flux boundary conditions were used. Figure 5 presents the spatial $PM_{2.5}$ concentration at four different times during the transport process. Again, we see that the assimilated estimates using the three regularization parameter selection methods behave similarly across space and time.

At $t = 0.09$, the assimilated state estimates, observational data, first guess, and true concentration values all align closely across the spatial domain. This close alignment early in the simulation indicates that the model, first guess, and data are in agreement at this stage, with minimal deviation among them. At $t = 5.93$, as the concentration evolves across space, discrepancies become more noticeable between the first guess and the true concentration, we note that estimates from all methods follow the first guess throughout this spatial evolution even though it is not necessarily closer to the solution than the data. At $t = 8.85$, the differences between the first guess and the true concentration become more pronounced as the concentration continues to evolve across space. Even though the data has more noise than the first guess due to the randomness in our experiments the data is closer to the true values. The estimates from all methods continue to follow the first guess and only fit the data points that align with the first guess, particularly in the earlier spatial regions. At $t = 11.78$, the L-curve estimates display a clear effort to better fit the third, fourth and fifth data points, showing a significant deviation from the first guess. This deviation is an indication of the L-curve choice of larger variance in the model. Estimates from the GCV and χ^2 methods, on the other hand remain closely aligned with the first guess, fitting only the observed data that align with the first guess.

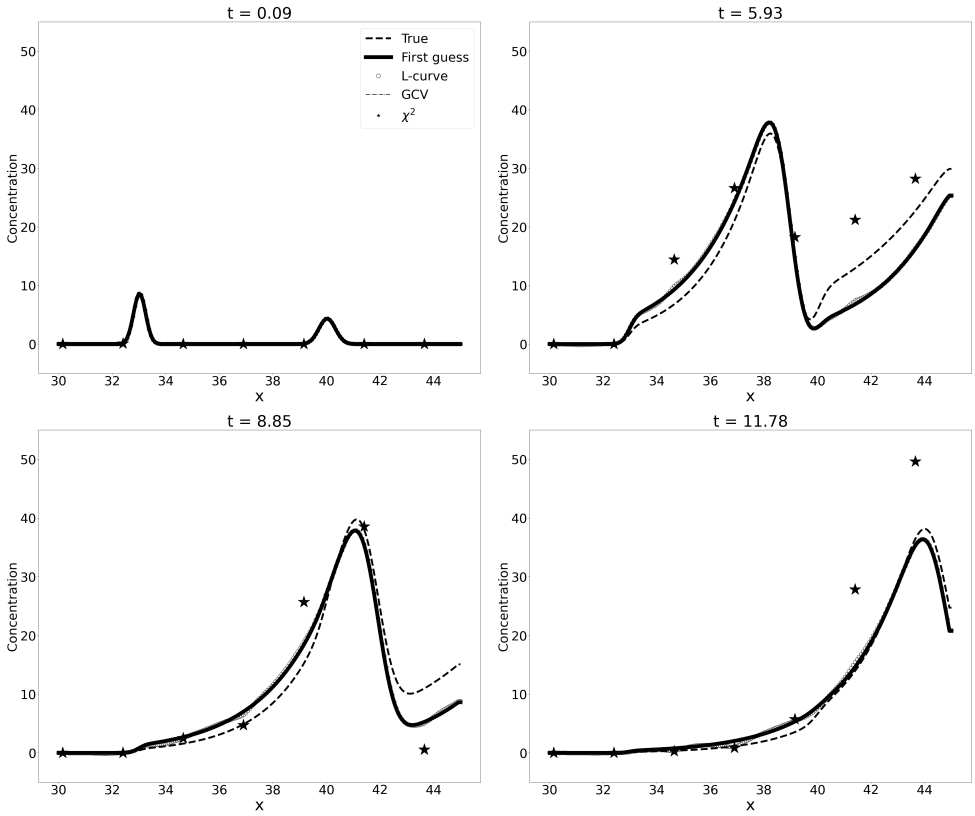


Fig. 5: $PM_{2.5}$ concentration estimates a function of space for *experiment 2*

5.4.3. Experiment 3. This experiment was designed with the simulated observational data closer to the true $PM_{2.5}$ concentration than the first guess. Figure 6 shows $PM_{2.5}$ concentration as a function of time at four different spatial locations, namely $x = 32.55, 34.88, 37.20, 41.85$. These locations were strategically chosen to capture critical moments in the transport process, focusing on positions before and after the source of the pollutant. Similar to experiment 1 there is one source and periodic boundary conditions were used.

At $x = 32.55$, the assimilated estimates attempt to fit the data points. Between $t = 5$ and $t = 15$ the estimates tend to oscillate most likely due to ill-posedness in the underdetermined problem. After the data point at $t = 17.9$, the estimates from each method become more distinct. The GCV estimates demonstrate a better overall fit to the data throughout the time series, aligning more precisely with the observed values compared to the L-curve and χ^2 methods. At $x = 34.88$, the assimilated estimates try to fit the data but at the impulse at $t = 2.5$, they more closely align with the first guess. After this point, the estimates from different methods slightly differ. The estimates from the GCV method fit the data points more closely. All three assimilated estimates deviate significantly from the first guess in an attempt to fit the data. At $x = 37.20$ and $x = 41.85$, all three assimilated estimates attempt to fit the data, showing a stronger preference to the observational data than the first guess. The GCV estimates fit the data points more closely than other methods reflecting the larger value of σ_f^2 that was chosen with this method.

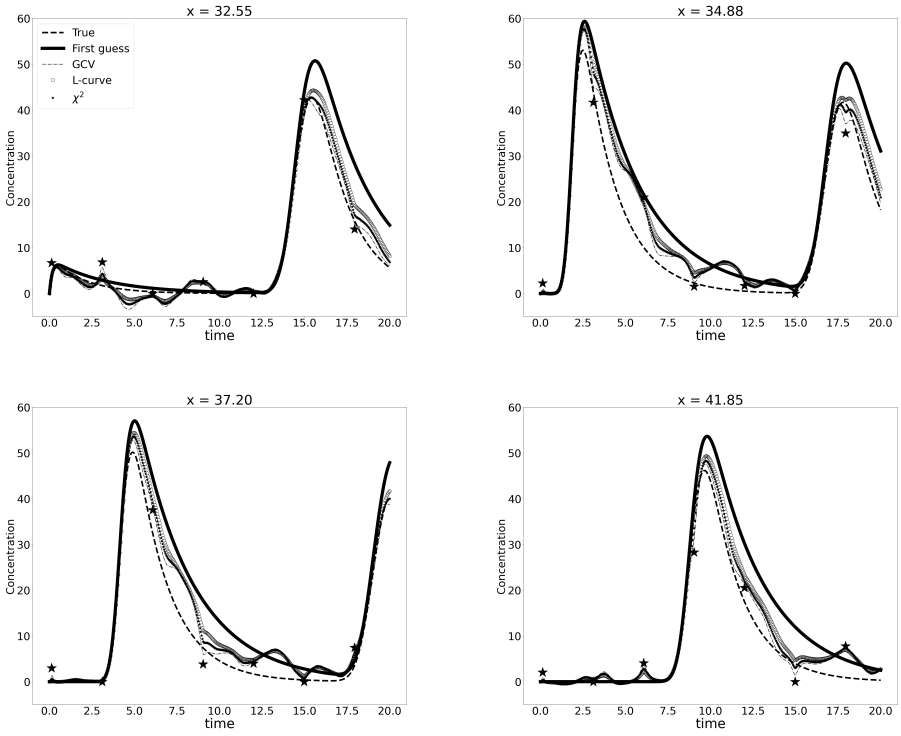


Fig. 6: $PM_{2.5}$ concentration estimates a function of time for *experiment 3*

5.4.4. Experiment 4. This experiment was set up similarly to experiment 3, but the key difference in this experiment is that it has two sources of $PM_{2.5}$ located at $x = 33$ and $x = 40$ and no flux boundary conditions were used. Figure 7 presents the temporal evolution of $PM_{2.5}$ concentration estimates at four different spatial locations: $x = 32.55, 37.20, 39.52, 41.85$.

At $x = 32.55$, a location on the left side of both emission sources, the $PM_{2.5}$ concentration is close to zero due to the constant wind field used, blowing the pollutant to the right and the assimilated estimates oscillate as they fit the data. This oscillation can be attributed to the ill-posed nature of the underdetermined problem. At $x = 37.2$ and $x = 39.52$, the assimilated estimates show a clear attempt to fit the data points but they also follow the first guess while making necessary adjustments to accommodate the observed data. This balance between following the first guess and adapting to the data points illustrates the estimates' capacity to capture the key features of the transport process at these locations. At $x = 41.85$, the assimilated estimates manage to largely follow the first guess while making necessary adjustments to accommodate the observed data at most times except for the attempt to fit data near $t = 7.5$. This balance between following the first guess and adapting to the data points illustrates the estimates' capacity to capture the key features of the transport process at this location.

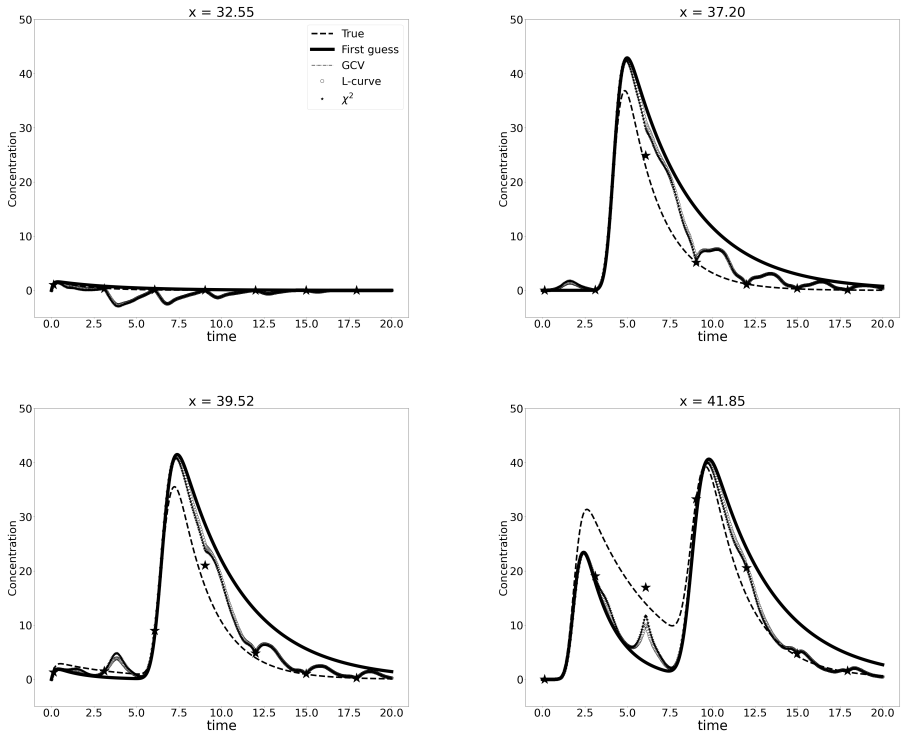


Fig. 7: $PM_{2.5}$ concentration estimates a function of space for *experiment 4*

Table 3 presents the root mean square errors (RMSEs) of the three $PM_{2.5}$ concentration estimates and the true concentrations. In experiment 1, the RMSE of the

Experiment	1	2	3	4
First guess	1.5319	2.1465	6.5241	5.8753
Data	5.7629	6.6861	3.5120	2.4087
L-curve	1.8219	2.1548	3.8177	4.1629
GCV	1.8400	2.1426	2.9801	4.0341
χ^2	1.8516	2.1459	3.2786	4.0005

Table 3: RMSEs between the various $PM_{2.5}$ estimates and the true concentration

first guess is 1.5319, while the RMSE of the data is much higher at 5.7629 through these scalar values we see how the first guess is closer to the true solution than the data. The RMSEs of the assimilated estimates using the parameter selection methods are similar to each other and slightly higher than the RMSE of the first guess. From these values, we conclude that the automated choices of σ_f^2 correctly identified the magnitude of accuracy of the model dynamics. For experiment 2, the RMSEs of the first guess and the assimilated estimates are similar indicating that the data did not mislead the estimates and the automated choice of σ_f^2 accurately reflects the model error. In experiment 3, the first guess has a higher RMSE of 6.5241, while the data RMSE is lower at 3.5120. The GCV method achieves the best RMSE, followed by the χ^2 method and the L-curve method. Since the RMSE of the assimilated estimates is near the RMSE of the more accurate data, we again see that the automated choice of larger σ_f^2 correctly reflects the larger model error.

In experiment 4, the first guess has a RMSE of 5.8753, while the data RMSE is lower at 2.4087. The RMSEs for all the three assimilated estimates are greater than that for the data but less than that for the first guess. This may indicate that the assimilated estimates are worse than the data. However, the data are sparse and the model dynamics give concentration estimates where there are no data. Since the RMSEs for the assimilated estimates are smaller than the RMSE for the first guess, we did find a reasonable compromise between data and model estimates with the automatic σ_f^2.

In all attempts, assimilated estimates using σ_f^2 found by the regularizition parameter selection methods i.e. the GCV, L-curve, and χ^2 methods have lower RMSE values compared to the largest first guess or the data RMSE.

6. Conclusion. In this work, we formulated weak constraint 4D-Var as a regularization inverse problem, with the model error variance as the regularization parameter. The main focus of our study was on applying three regularization parameter selection methods namely, the L-curve, GCV, and χ^2 methods to estimate the model error variance. A key aspect of our approach was the application of the representer method, which leverages the finite-dimensional nature of the observation space to express optimal state estimates as a linear combination of representers. This method reduces the solution search space from the state space to the data space, providing an efficient implementation for 4D-Var and enabling us to express the assimilated estimates analytically. **Additionally, the representer method facilitated the derivation of matrix expressions for each regularization parameter selection method, such as $J_{\mathbf{data}}[\hat{q}]$ in Lemma 3.1, $J_{\mathbf{mod}}[\hat{q}]$ in Lemma 3.2 for the L-curve, Lemma 3.1 for the GCV, and $J[\hat{q}]$ in Theorem 3.3 for the χ^2 method.**

We conducted numerical experiments using the transport equation to estimate

PM2.5 concentrations with simulated observational data. The results illustrated the importance of accounting for model imperfections in data assimilation. Across the four experiments, the three regularization parameter selection methods yielded consistent model error variance estimates. These estimates accurately captured the balance between the influence of observational data and model predictions on assimilated state estimates. This balance is crucial for accurate $PM_{2.5}$ estimation in complex wildfire smoke transport scenarios, where, in addition to unknown uncertainties in the forcing, there are also unknown model physics.

In situations where observational data was significantly more reliable than the model, our results showed that the assimilated estimates closely fit the data. However, as demonstrated in experiments 1 and 4, the problem is often severely underdetermined due to insufficient data, a common challenge in state estimation problems. In such instances, artificially decreasing σ_f^2 can help create a well-posed problem. Alternatively, a large value of σ_f^2 may indicate that further model improvement is necessary before using it for data assimilation.

REFERENCES

[1] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.

[2] Ross N Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703):607–633, 2017.

[3] A Benedetti, J-J Morcrette, O Boucher, A Dethof, RJ Engelen, M Fisher, H Flentje, N Huneeus, L Jones, JW Kaiser, et al. Aerosol analysis and forecast in the european centre for medium-range weather forecasts integrated forecast system: 2. data assimilation. *Journal of Geophysical Research: Atmospheres*, 114(D13), 2009.

[4] Andrew F Bennett. *Inverse methods in physical oceanography*. Cambridge university press, 1992.

[5] Andrew F Bennett, Boon S Chua, D Ed Harrison, and Michael J McPhaden. Generalized inversion of tropical atmosphere–ocean data and a coupled model of the tropical pacific. *Journal of Climate*, 11(7):1768–1792, 1998.

[6] Alberto Carrassi, Marc Bocquet, Laurent Bertino, and Geir Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535, 2018.

[7] Alberto Carrassi and Stéphane Vannitsem. Accounting for model error in variational data assimilation: A deterministic formulation. *Monthly Weather Review*, 138(9):3369–3386, 2010.

[8] William J Crawford, Polly J Smith, Ralph F Milliff, Jerome Fiechter, Christopher K Wikle, Christopher A Edwards, and Andrew M Moore. Weak constraint four-dimensional variational data assimilation in a model of the california current system. *Advances in Statistical Climatology, Meteorology and Oceanography*, 2(2):171–192, 2016.

[9] Heinz W Engl and Wilhelm Grever. Using the l–curve for determining optimal regularization parameters. *Numerische Mathematik*, 69(1):25–31, 1994.

[10] Geir Evensen. Inverse methods and data assimilation in nonlinear ocean models. *Physica D: Nonlinear Phenomena*, 77(1-3):108–129, 1994.

[11] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[12] Gene H Golub, Per Christian Hansen, and Dianne P O'Leary. Tikhonov regularization and total least squares. *SIAM journal on matrix analysis and applications*, 21(1):185–194, 1999.

[13] Gene H Golub, Michael Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[14] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE proceedings F (radar and signal processing)*, volume 140, pages 107–113. IET, 1993.

[15] Per Christian Hansen. The l-curve and its use in the numerical treatment of inverse problems.

1999.
[16] Edward J Hyer, Christopher P Camacho, David A Peterson, Elizabeth A Satterfield, and Pablo E Saide. Data assimilation for numerical smoke prediction. *Landscape Fire, Smoke, and Health: Linking Biomass Burning Emissions to Human Well-Being*, pages 105–125, 2023.
[17] Kayo Ide, Philippe Courtier, Michael Ghil, and Andrew C Lorenc. Unified notation for data assimilation: Operational, sequential and variational (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):181–189, 1997.
[18] Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.
[19] Randall J LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.
[20] Magnus Lindskog, Dick Dee, Yannick Tremolet, Erik Andersson, Gabor Radnoti, and Mike Fisher. A weak-constraint four-dimensional variational analysis system in the stratosphere. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 135(640):695–706, 2009.
[21] AC Lorenc, SP Ballard, RS Bell, NB Ingleby, PLF Andrews, DM Barker, JR Bray, AM Clayton, T Dalby, D Li, et al. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012, 2000.
[22] J.L. Mead and C.C. Hammerquist. Chi-sqaured tests for the choice of the regularization parameter in nonlinear inverse problems. *SIAM Journal on Matrix Analysis and Applications*, 34(3):1213–1230, 2013.
[23] Jodi L Mead. Parameter estimation: A new approach to weighting a priori information. *J. Inv. Ill-posed Problems*, 16(2):175–194, 2008.
[24] Jodi L. Mead. Chi-squared test for total variation regularization parameter selection. *Inverse Problems Imaging*, 14(3):401–421, 2020.
[25] Hans Ngodock, Matthew Carrier, Scott Smith, and Innocent Souopgui. Weak and strong constraints variational data assimilation with the ncom-4dvar in the agulhas region using the representer method. *Monthly Weather Review*, 145(5):1755–1764, 2017.
[26] Jason A Otkin, Roland WE Potthast, and Amos S Lawless. Nonlinear conditional model bias estimation for data assimilation. *SIAM Journal on Applied Dynamical Systems*, 20(1):299–332, 2021.
[27] Nicolas Papadakis, Étienne Mémin, Anne Cuzol, and Nicolas Gengembre. Data assimilation with the weighted ensemble kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, 62(5):673–697, 2010.
[28] F Rabier, P Courtier, and Martin Ehrendorfer. Four-dimensional data assimilation: Comparison of variational and sequential algorithms. *Quarterly Journal of the Royal Meteorological Society*, 118(506):673–713, 1992.
[29] Rosemary A Renaut, Iveta Hnětynková, and Jodi Mead. Regularization parameter estimation for large-scale tikhonov regularization using a priori information. *Computational statistics & data analysis*, 54(12):3430–3445, 2010.
[30] Daniel Sanz-Alonso, Andrew M Stuart, and Armeen Taeb. Inverse problems and data assimilation. *arXiv preprint arXiv:1810.06191*, 2018.
[31] Andre Nikolaevich Tikhonov and VIAK Arsenin. Solutions of ill-posed problems. *(No Title)*, 1977.
[32] Yannick Tr'emolet. Accounting for an imperfect model in 4d-var. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 132(621):2483–2504, 2006.
[33] PA Vidard, Andrea Piacentini, and F-X Le Dimet. Variational data analysis with control of the forecast bias. *Tellus A: Dynamic Meteorology and Oceanography*, 56(3):177–188, 2004.
[34] Haiyan Zhou, J Jaime Gómez-Hernández, and Liangping Li. Inverse methods in hydrogeology: Evolution and recent trends. *Advances in Water Resources*, 63:22–37, 2014.
[35] Dusanka Zupanski. A general weak constraint applicable to operational 4dvar data assimilation systems. *Monthly Weather Review*, 125(9):2274–2292, 1997.