Variational Data Assimilation and Regularization for Ill-posed Problems: A Common Framework

Jodi Mead[*,†], Sandra Babyale[‡], Donna Calhoun[*], Patricia Azike[‡]

[*]Department of Mathematics, Boise State University
[†]National Science Foundation, Division of Mathematical Sciences
[‡]Computing PhD, Boise State University

Supported by NSF Independent Research/Development (IR/D) Program and DMS-2111585

Motivation

Wildfire smoke modeling

$$\frac{\partial q}{\partial t} + \nabla \cdot \vec{u}(\mathbf{x}, t) q(\mathbf{x}, t) = Q(\mathbf{x}, t)$$

Atmospheric winds *$PM_{2.5}$ concentration* *Wildfire plume rise*
Smoke emissions

- Fire related processes are difficult to simulate
  e.g. Pole Creek and Bald Mountain Fires, Central Utah, September, 2018

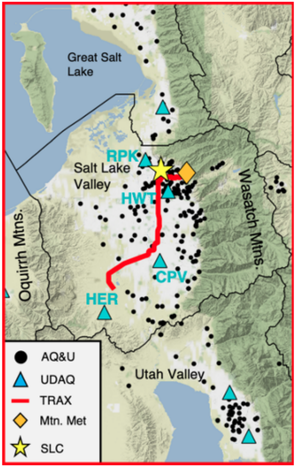
AGU ADVANCING EARTH AND SPACE SCIENCE

JGR Atmospheres

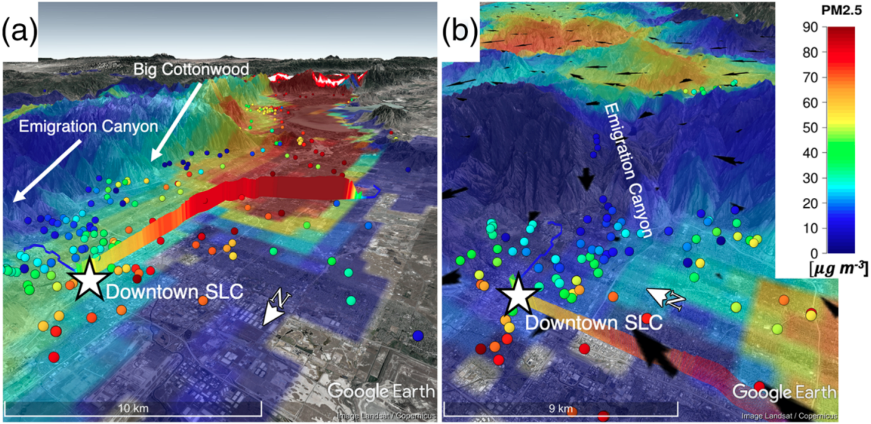
Evaluating Wildfire Smoke Transport Within a Coupled Fire-Atmosphere Model Using a High-Density Observation Network for an Episodic Smoke Event Along Utah's Wasatch Front

Derek V. Mallia[1], Adam K. Kochanski[1,2], Kerry E. Kelly[3], Ross Whitaker[4], Wei Xing[5], Logan E. Mitchell[1], Alex Jacques[1], Angel Farguell[6], Jan Mandel[6], Pierre-Emmanuel Gaillardon[7], Tom Becnel[7], and Steven K. Krueger[1]

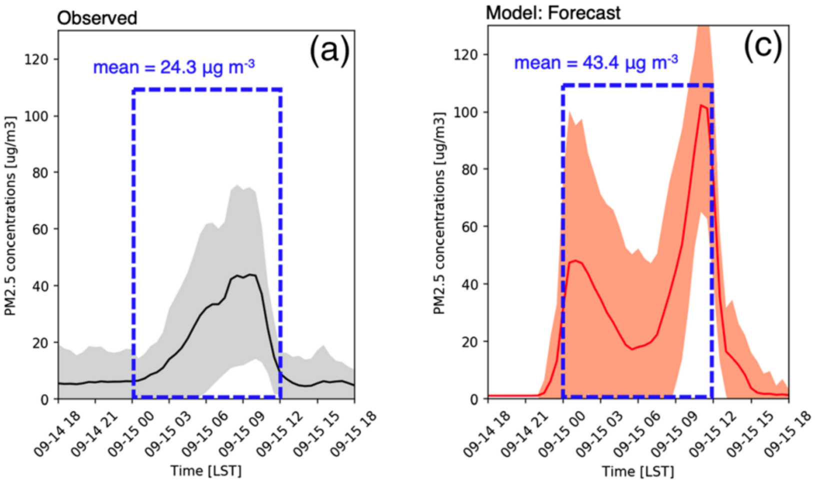
[1]Department of Atmospheric Sciences, University of Utah, Salt Lake City, UT, USA, [2]Department of Meteorology and Climate Science, San José State University, San Jose, CA, USA, [3]Department of Chemical Engineering, University of Utah, Salt Lake City, UT, USA, [4]School of Computing, University of Utah, [5]Scientific Computing and Imaging Institute, University of Utah, [6]Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO, USA, [7]Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT, USA



WRFSFC Forecast and Observations



Averaged $PM_{2.5}$ Concentrations

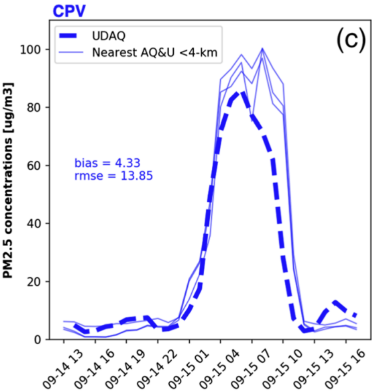


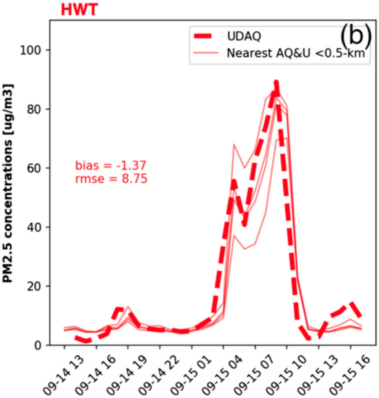
Outline

- Variational Data Assimilation (4DVAR)
  Weakly constrained dynamics and the Representer method
- Ill-posed Inverse Problems
  Regularization parameter selection
- Common Framework
  Uncertainty estimates using L-curve, GCV and χ^2 method
- Results
  1D transport model with simulated data

Observations

$$d_m = q(x_m, t_m) + \epsilon_m$$





Forward Problem

$$\frac{\partial q_F}{\partial t} + \nabla \cdot u(\mathbf{x},t) q_F(\mathbf{x},t) = Q(\mathbf{x},t)$$

$$q_F(\mathbf{x},0) = I(\mathbf{x})$$

$$q_F(\mathbf{x},t) = B(t) \qquad \text{for } \mathbf{x} \in \partial\Omega$$

Failure of the forward solution: $q_F(\mathbf{x}_m, t_m) \neq d_m$

Weakly constrained dynamics

$$\begin{cases} \dfrac{\partial q}{\partial t} + \nabla \cdot u(\mathbf{x},t) q(\mathbf{x},t) = Q(\mathbf{x},t) + f(\mathbf{x},t) \\ q(\mathbf{x},0) = I(\mathbf{x}) + i(\mathbf{x}) \\ q(\mathbf{x},t) = B(t) + b(t) \qquad \text{for } \mathbf{x} \in \partial\Omega \end{cases}$$

Normal Distribution Assumption

$$f(\mathbf{x},t) \sim \mathcal{N}(0, C_f(\mathbf{x},t))$$
$$i(\mathbf{x}) \sim \mathcal{N}(0, C_i(\mathbf{x}))$$
$$b(t) \sim \mathcal{N}(0, C_b(t))$$

$$\epsilon_m \sim \mathcal{N}(0, \sigma_{d_m})$$

e.g. probability density function of data error

$$(2\pi\sigma_{d_m}^2)^{-1/2} e^{-\frac{(d_m - q(x_m, t_m))^2}{2\sigma_{d_m}^2}}$$

Least Squares Fitting

Maximum likelihood estimate when errors are normally distributed

$$
\begin{aligned}
\mathcal{J}[q(\mathbf{x},t)] = & \int_0^T \int_\Omega \int_0^T \int_\Omega f(\mathbf{x},t) C_f^{-1}(\mathbf{x},t,\mathbf{x}',t') f(\mathbf{x}',t') \; d\mathbf{x}' dt' d\mathbf{x} dt \\
& + \int_\Omega \int_\Omega i(\mathbf{x}) C_i^{-1}(\mathbf{x},\mathbf{x}') i(\mathbf{x}') \; d\mathbf{x}' d\mathbf{x} \\
& + \int_0^T \int_0^T b(t) C_b^{-1}(t,t') b(t') \; dt' dt \\
& + \boldsymbol{\epsilon}^T \mathbf{C}_\epsilon^{-1} \boldsymbol{\epsilon}
\end{aligned}
$$

where

$$
\int_0^T \int_\Omega C_f^{-1}(\mathbf{x},t,\mathbf{x}',t') C_f(\mathbf{x}',t',\mathbf{x}'',t'') \; dx' dt' = \delta(\mathbf{x}-\mathbf{x}'')\delta(t-t'')
$$

Calculus of Variations

$\hat{q}(\mathbf{x}, t) = \arg\min_{q(\mathbf{x},t)} \mathcal{J}$ so that

$$\mathcal{J}[\hat{q} + \delta q] = \mathcal{J}[\hat{q}] + \mathcal{O}(\delta q)^2$$

Define adjoint variable

$$\lambda(\mathbf{x}, t) = \int_0^T \int_\Omega C_f^{-1}(\mathbf{x}, t, \mathbf{x}', t') f(\mathbf{x}', t')\ d\mathbf{x}' dt' \equiv C_f^{-1} \bullet f$$

and write the discrete term as $\hat{\boldsymbol{\epsilon}}^T \mathbf{C}_\epsilon^{-1} \delta \mathbf{q}$ or

$$\int_0^T \int_\Omega \sum_{m=1}^{M} (\hat{\boldsymbol{\epsilon}}^T \mathbf{C}_\epsilon^{-1})_m \delta q(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_m) \delta(t - t_m)\ d\mathbf{x} dt$$

with $\hat{\boldsymbol{\epsilon}}, \delta \mathbf{q} \in \mathbb{R}^M$ and $\mathbf{C}_\epsilon^{-1} \in \mathbb{R}^{M \times M}$

Euler-Lagrange Equations

Coupled system of equations where

$$(B)\begin{cases} -\dfrac{\partial \lambda}{\partial t} - \nabla \cdot u(\mathbf{x},t)\lambda(\mathbf{x},t) = -\sum_m \left((\hat{\mathbf{q}}-\mathbf{b})^T \mathbf{C}_\epsilon^{-1}\right)_m \delta(\mathbf{x}-\mathbf{x}_m)\delta(t-t_m) \\ \lambda(\mathbf{x},T) = 0 \\ \lambda(\mathbf{x},t) = 0 \qquad \text{for } \mathbf{x} \in \partial\Omega \end{cases}$$

is solved backward in space and time for adjoint variable $\lambda(\mathbf{x},t)$ while

$$(F)\begin{cases} \dfrac{\partial \hat{q}}{\partial t} + \nabla \cdot u(\mathbf{x},t)\hat{q}(\mathbf{x},t) = Q(\mathbf{x},t) + C_f \bullet \lambda \\ \hat{q}(\mathbf{x},0) = I(\mathbf{x}) + C_i \circ \lambda(\mathbf{x},0) \\ \hat{q}(\mathbf{x},t) = u(\mathbf{x},t)B(t) + C_b * \lambda(\mathbf{x},t) \qquad \text{for } \mathbf{x} \in \partial\Omega \end{cases}$$

is solved forward in space and time for optimal state estimate $\hat{q}(\mathbf{x},t)$

Representer Theorem

Let

$$\hat{q}(\mathbf{x}, t) = \arg\min_{q \in \mathcal{H}} \left\{ \boldsymbol{\epsilon}^T \mathbf{C}_\epsilon^{-1} \boldsymbol{\epsilon} + \mathcal{R}(q) \right\}$$

with $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ where $\mathcal{H}_0 = \{q : \mathcal{R}(q) = 0\}$ is a finite dimensional space with basis $\phi_1, \ldots, \phi_N$ and $\mathcal{H}_1$ is a Reproducing Kernel Hilbert Space (RKHS) with reproducing kernel Γ. Let $r_m(x, t) = \Gamma(\mathbf{x}, t, \mathbf{x}_m, t_m)$ be representers. Then

$$\hat{q}(\mathbf{x}, t) = \sum_{n=1}^{N} c_n \phi_n(\mathbf{x}, t) + \sum_{m=1}^{M} \beta_m r_m(\mathbf{x}, t).$$

Note that for $\mathcal{J} = \boldsymbol{\epsilon}^T \mathbf{C}_\epsilon^{-1} \boldsymbol{\epsilon} + \mathcal{R}(q)$, $\mathcal{R}(q_F) = 0$ so that

$$\hat{q}(\mathbf{x}, t) = q_F(\mathbf{x}, t) + \sum_{m=1}^{M} \beta_m r_m(\mathbf{x}, t).$$

Reproducing Kernel Hilbert Space (RKHS)

Define adjoint representer $\alpha_m(\mathbf{x}, t)$ as the Green's Function

$$(B) \begin{cases} -\frac{\partial \alpha_m}{\partial t} - \nabla \cdot u(\mathbf{x}, t)\alpha_m(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_m)\delta(t - t_m) \\ \alpha_m(\mathbf{x}, T) = 0 \\ \alpha_m(\mathbf{x}, t) = 0. \qquad \text{for } \mathbf{x} \in \partial\Omega \end{cases}$$

and let $r_m(\mathbf{x}, t)$ satisfy

$$(F) \begin{cases} \frac{\partial r_m}{\partial t} + \nabla \cdot u(\mathbf{x}, t) r_m(\mathbf{x}, t) = C_f \bullet \alpha_m(\mathbf{x}, t) \\ r_m(\mathbf{x}, 0) = C_i \circ \alpha_m(\mathbf{x}, 0) \\ r_m(\mathbf{x}, t) = u(0, t) C_b * \alpha_m(\mathbf{x}, t) \text{for } \mathbf{x} \in \partial\Omega \end{cases}$$

Then $r_m(\mathbf{x}_l, t_l) = r_l(\mathbf{x}_m, t_m)$ is a reproducing kernel.

Representer Coefficients

Substituting $\hat{q}(\mathbf{x},t) = q_F(\mathbf{x},t) + \sum_{m=1}^{M} \beta_m r_m(\mathbf{x},t)$ into E-L equations decouples them and we get

$$(\mathbf{R} + \mathbf{C}_\epsilon)\,\boldsymbol{\beta} = \mathbf{d} - \mathbf{q}_F$$

where $\mathbf{R}$ is the representer matrix with elements $\mathbf{R}_{ml} = r_m(\mathbf{x}_l, t_l)$.

This gives the optimal estimate

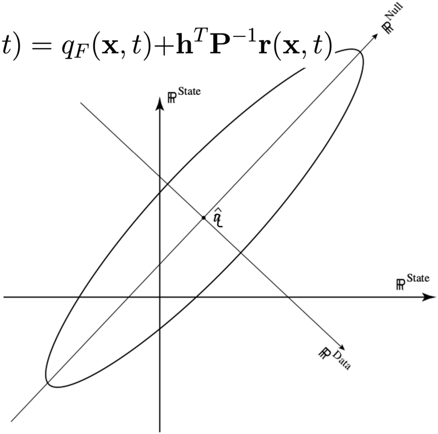
$$\hat{q}(\mathbf{x},t) = q_F(\mathbf{x},t) + (\mathbf{d} - \mathbf{q}_F)^T(\mathbf{R} + \mathbf{C}_\epsilon)^{-1}\mathbf{r}(\mathbf{x},t)$$

with $\mathbf{r}(\mathbf{x},t)$ the vector of the m representers evaluated at $(\mathbf{x},t)$.

Representer Method for 4DVAR*

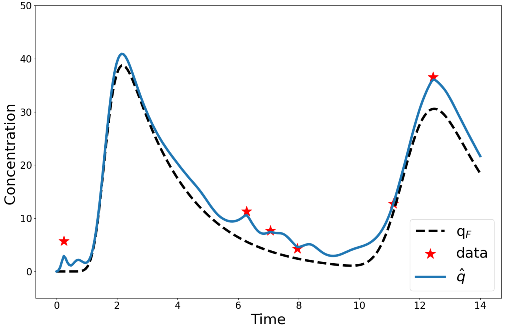
Define $\mathbf{h} = \mathbf{d} - \mathbf{q}_F$ and $\mathbf{P} = \mathbf{R} + \mathbf{C}_\epsilon$ so that

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

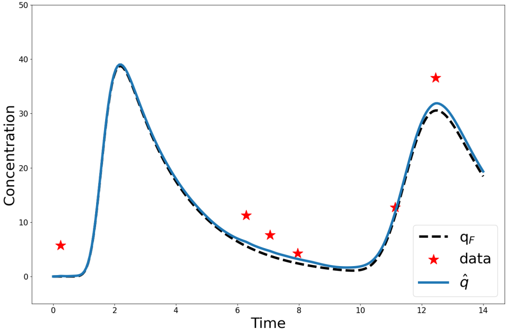


*Bennett, 2005

Effect of Error Covariance on Optimal Estimates



C_f is “larger" than $\mathbf{C}_\epsilon$



$\mathbf{C}_\epsilon$ is “larger" than C_f

Simplifying Assumptions

$$\mathcal{J}[q(\mathbf{x},t)] = \sigma_f^{-2}\int_0^T\int_\Omega\left(\frac{\partial q}{\partial t} + \nabla\cdot u(\mathbf{x},t)q(\mathbf{x},t) - Q(\mathbf{x},t)\right)^2 \, d\mathbf{x}dt$$
$$+\boldsymbol{\epsilon}^T\mathbf{C}_\epsilon^{-1}\boldsymbol{\epsilon}$$

i.e. initial and boundary conditions are exact while errors in PDE can be represented with a scalar σ_f.

Note that

$$\min_{q(\mathbf{x},t)} \ \boldsymbol{\epsilon}^T\mathbf{C}_\epsilon^{-1}\boldsymbol{\epsilon}$$

is ill-posed in continuous intervals $\mathbf{x}\in\Omega$, $t\in[0,T]$. Weakly constrained 4DVAR regularizes it as long as σ_f is "small" enough.

Well-posed Problems

The problem of finding $q(\mathbf{x}, t)$ from $\mathbf{d}$ is called well-posed (Hadamard, 1923) if the following hold

- Existence - a solution $q(\mathbf{x}, t)$ exists for any $\mathbf{d}$ in data space
- Uniqueness - the solution is unique
- Stability - continuous dependence of $q(\mathbf{x}, t)$ on $\mathbf{d}$, the inverse mapping $\mathcal{A}^{-1} : \mathbf{d} \rightarrow q(\mathbf{x}, t)$ is continuous

The first two conditions are equivalent to saying that the operator $\mathcal{A}$ has a well defined inverse $\mathcal{A}^{-1}$. Moreover, we require that the domain of $\mathcal{A}^{-1}$ is the data space.

Well-conditioned Problems

For condition number $\kappa(x)$

$$\text{variation in output} \approx \kappa(x) \text{ perturbation in input}$$

Example

Let $\mathbf{A} \in \mathbb{R}^{M\times M}$, then

$$\kappa(\mathbf{A}) = \begin{cases} \|\mathbf{A}\|\|\mathbf{A}^{-1}\| & \text{if } \mathbf{A} \text{ is invertible} \\ \infty & \text{otherwise} \end{cases}$$

Regularization

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

Existence: Least squares

$$\min_{\mathbf{x}} \|\mathbf{C}_\epsilon^{-1/2}(\mathbf{b} - \mathbf{A}\mathbf{x})\|_2^2 \Rightarrow \mathbf{A}^T\mathbf{C}_\epsilon^{-1}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{C}_\epsilon^{-1/2}\mathbf{b}$$

However, $\mathbf{A}^T\mathbf{C}_\epsilon^{-1}\mathbf{A}$ may not be invertible and so we use regularization.

<u>Tikhonov Regularization</u>

Introduce a regularization parameter $\eta > 0$ such that small η gives a problem "close" to original e.g.

$$\min_{\mathbf{x}} \{\|\mathbf{C}_\epsilon^{-1/2}(\mathbf{b} - \mathbf{A}\mathbf{x})\|_2^2 + \eta\|\mathbf{L}\mathbf{x}\|_2^2\} \Rightarrow (\mathbf{A}^T\mathbf{C}_\epsilon^{-1}\mathbf{A} + \eta\mathbf{L}^T\mathbf{L})\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{C}_\epsilon^{-1/2}\mathbf{b}$$

where $\mathbf{L}$ can be a differential operator and $\|\mathbf{L}\mathbf{x}\|$ an approximation to a Sobolev norm.

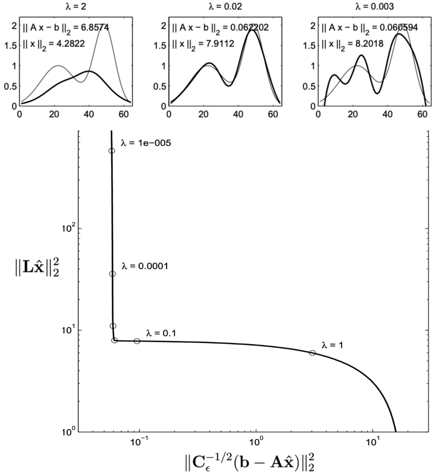
L-curve* Choice of Regularization Parameter

Methodology: Find the best trade-off between the size of a regularized estimate and its fit to the given data, as the regularization parameter varies.

Approach: For a range of values of η, create a log-log plot of the norm of a regularized solution versus the the corresponding residual norm.

*Hansen, 2000

L-curve example ($\lambda = \eta$)



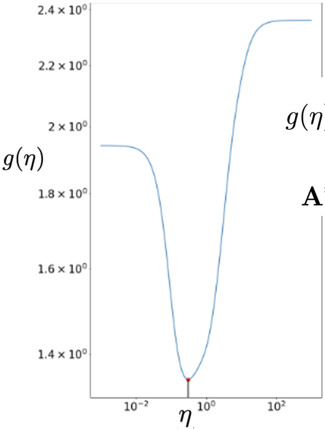
Generalized Cross Validation (GCV)* Choice of η

Methodology: Hold back data and the optimal regularization parameter is the one which minimizes the predictive error for all other data.

Approach: Minimize the sum of predictive error by applying the leave-one-out lemma to speed up computation.

*Golub et al, 1979

GCV Example



$$g(\eta) = \frac{m\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|^2}{\left(\text{trace}(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#})\right)^2}$$

$$\mathbf{A}^{\#} = (\mathbf{A}^T\mathbf{A} + \eta\mathbf{I})^{-1}\mathbf{A}^T$$

χ^{2*} Discrepancy for Choice of η

Methodology: Find the regularization parameter for which the objective function evaluated at the optimal state estimate follows a χ^2_M distribution.

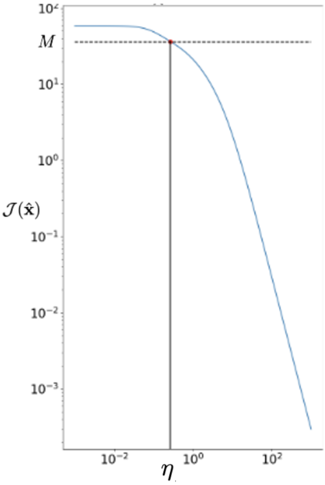
Morozov's discrepancy principle: $\|\mathbf{C}_\epsilon^{-1/2}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}})\|_2^2 \approx M$

χ^2 Method: $\mathcal{J}(\hat{\mathbf{x}}) = \|\mathbf{C}_\epsilon^{-1/2}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}})\|_2^2 + \eta\|\mathbf{L}\hat{\mathbf{x}}\|^2 \sim \chi^2_M$

Approach: For a given data set, solve the nonlinear equation $\mathcal{J}[\hat{\mathbf{x}}] = M$ for η.

*Mead, 2006

χ^2 Method Example



Common Framework - Optimal Estimates

Regularized Inverse Problem

$$\mathcal{J}[\mathbf{x}] = (\mathbf{b} - \mathbf{A}\mathbf{x})^T \mathbf{C}_\epsilon^{-1} (\mathbf{b} - \mathbf{A}\mathbf{x}) + \eta \|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\|^2$$

$$\hat{\mathbf{x}} \equiv \arg\min_{\mathbf{x}} \mathcal{J}[\mathbf{x}] = \mathbf{x}_0 + (\mathbf{A}^T\mathbf{A} + \eta \mathbf{L}^T\mathbf{L})^{-1}\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}_0)$$

Weakly Constrained 4DVAR with Representers

$$\begin{aligned}\mathcal{J}[q(x,t)] = & \sum_{m=1}^{M} \sigma_{\epsilon_m}^{-2} (d_m - q(x_m, t_m))^2 \\ & + \sigma_f^{-2} \int_0^T \int_0^L \left(\frac{\partial q}{\partial t} + \frac{\partial}{\partial x} (u(x,t) q(x,t)) - Q(x,t) \right)^2 dx dt\end{aligned}$$

$$\hat{q}(x,t) \equiv \arg\min_{q(x,t)} \mathcal{J}[q(x,t)] = q_F(x,t) + \mathbf{r}(x,t)^T (\mathbf{R} + \sigma_\epsilon^{-2}\mathbf{I})^{-1} (\mathbf{d} - \mathbf{q}_F)$$

Choice of PDE Error Estimates

The PDE error

$$\int_0^T \int_0^L \left(\frac{\partial q}{\partial t} + \frac{\partial}{\partial x} \left(u(x,t) q(x,t) \right) - Q(x,t) \right)^2 \; dx dt$$

in the cost function is weighted by its uncertainty estimate σ_f^{-2}. Existing approaches to specifying σ_f^{-2} include

- Order of numerical approximation
- Order of error in physics

Here we propose using regularization parameter selection methods. We assume the data uncertainty $\mathbf{C}_\epsilon$ is known.

Results for 4DVAR with Representers

L-curve

Plot

$$\mathcal{J}_{mod}[\hat{q}] = \sigma_f^{-2} \int_{T_0}^{T} \int_{\Omega} \left(\frac{\partial \hat{q}}{\partial t} + \nabla \cdot \vec{u}(x,t)\hat{q}(x,t) - Q(x,t) \right)^2 dx\, dt$$
$$= \sigma_f^{-2} \mathbf{h}^T \mathbf{P}^{-1} \mathbf{R} \mathbf{P}^{-1} \mathbf{h}$$

vs

$$\mathcal{J}_{data}[\hat{q}] = \sum_{m=1}^{M} \sigma_{\epsilon_m}^{-2} (d_m - \hat{q}(x_m, t_m))^2$$
$$= \mathbf{h}^T \mathbf{P}^{-1} \mathbf{C}_\epsilon \mathbf{P}^{-1} \mathbf{h}$$

and estimate the value of σ_f at the corner.

Results for 4DVAR with Representers

GCV

Optimal σ_f is the one that minimizes the GCV function

$$g(\sigma_f) = M \frac{\mathbf{h}^T \mathbf{P}^{-1} \mathbf{C}_\epsilon \mathbf{P}^{-1} \mathbf{h}}{\text{trace}(\mathbf{I} - \mathbf{R}^{\#})}$$

where the influence matrix is $\mathbf{R}^{\#} = \mathbf{R}(\mathbf{R} + \mathbf{C})^{-1}$.

χ^2 Discrepancy

Solve the nonlinear equation

$$\mathcal{J}[\hat{\mathbf{q}}] = \mathcal{J}_{data}[\hat{\mathbf{q}}] + \mathcal{J}_{mod}[\hat{\mathbf{q}}] = \mathbf{h}^T \mathbf{P}^{-1} \mathbf{h}$$

for σ_f.

Numerical Experiments

$$\frac{\partial q}{\partial t} + \nabla \cdot u(\mathbf{x},t)q(\mathbf{x},t) = S_0 exp(-kt - ar_0^2) + S_1 exp(-kt - ar_1^2)$$

where $r_i = x - x_i$. In all experiments $k = 0.5$, $x_0 = 32$ and $x_1 = 40$.

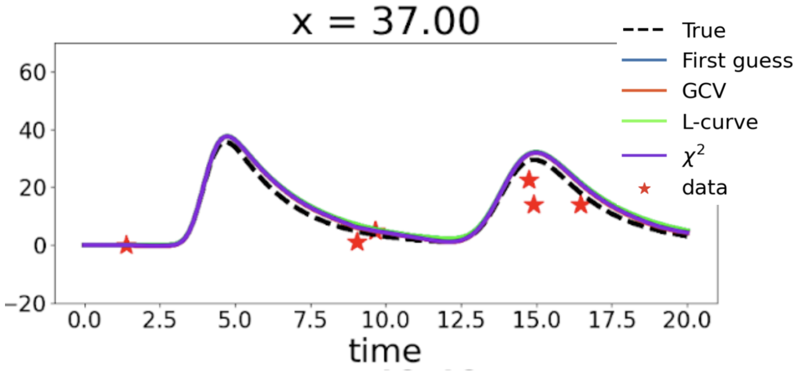
Experiment	B.C.	S_0	S_1	a
1	periodic	100	0	10
2	periodic	100	0	5
3	no flux	100	50	10
4	no flux	100	100	10

Estimates of PDE uncertainty σ_f

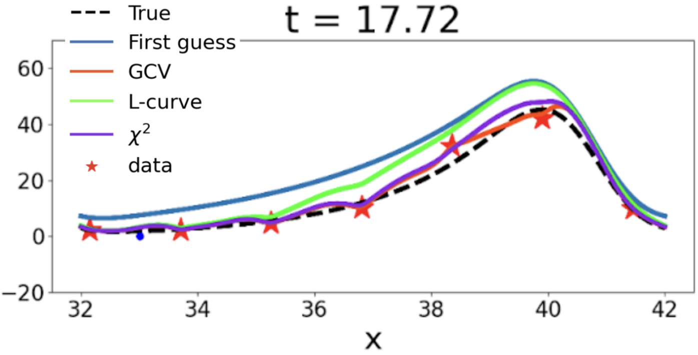
Random is noise added to k and a with standard deviation σ.
The mean of standard deviation of data noise is σ_ϵ.

Experiment	σ_ϵ	σ	L-curve	GCV	χ^2
1	0.5	0.1	2.66	0.09	0.07
2	0.15	0.5	0.04	18.7	2.0
3	0.35	0.2	0.498	0.018	0.001
4	0.22	0.6	0.0025	1.520	1.520

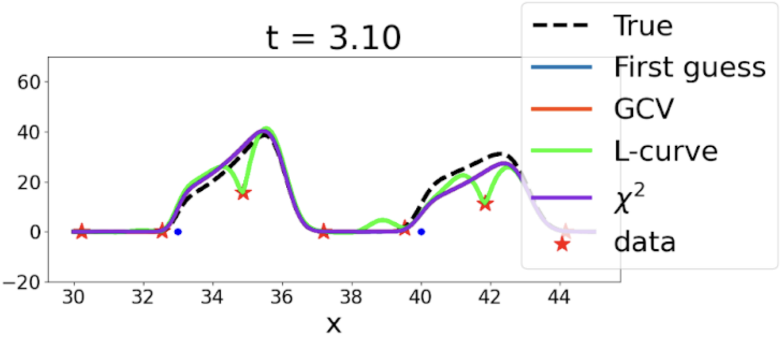
Experiment 1 - optimal concentration estimates over time



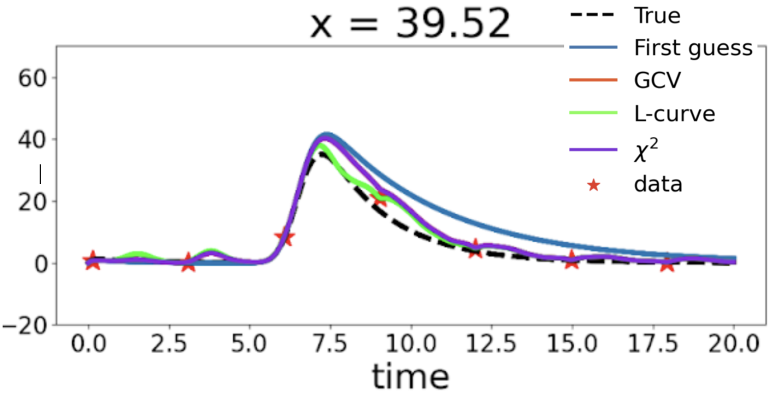
Experiment 2 - optimal concentration estimates over space



Experiment 3 - optimal concentration estimates over space



Experiment 4 - optimal concentration estimates over time



RMSE

Experiment	Data	Forward (q_F)	L-curve	GCV	χ^2
1	5.0944	2.0625	2.0625	1.7402	1.7930
2	3.4272	8.4616	3.2877	3.2040	3.5306
3	4.2152	2.1465	2.6059	2.1299	2.1431
4	2.5433	5.8943	3.0762	3.8050	3.8050

Summary and Conclusions

We present a new approach to estimating model errors or uncertainty for four dimensional variational data assimilation (4DVAR).

The approach

- utilizes the Representer Method which reduces the search space of state estimates to the size of the data space,
- formulates 4DVAR as an ill-posed inverse problem in order to use regularization parameter selection methods to estimate error or uncertainty in the forward PDE model.

Results include

- formulations for L-curve, GCV and χ^2 discrepancy principle methods for 4DVAR with Representers,
- error or uncertainty estimates for the PDE that accurately reflect if more weight should be given to the PDE or to the data,
- optimal state estimates that improve those from the forward PDE model or data.