Data assimilation on random smooth functions with applications to ensemble Kalman filter and satellite fire detection

Jan Mandel

Department of Mathematical and Statistical Sciences

University of Colorado Denver

Based on joint projects with Loren Cobb, Adam Kochanski, Martin Vejmelka, Sher Schranz.

Data Assimilation and Inverse Problems

University of Warwick, February 23, 2016

Supported by NSF grant DMS-1216481, Czech Science Foundation grant GA13-34856S,

and NASA grant NNX13AH59G

From least squares to Bayes theorem

Inverse problem $Hu \approx d$ with background knowledge $u \approx u^f$

$$|u - u^f|^2_{Q^{-1}} + |d - H(u)|^2_{R^{-1}} \to \min_u$$

u^f=forecast, what we think the state should be, d=data,
H=observation operator, $H(u)$=what the data should be given state u.

$$\underbrace{e^{-\frac{1}{2}\left(|u-u^a|^2_{(Q^a)^{-1}}\right)}}_{\substack{\propto\, p^a(u) \\ \text{analysis (posterior) density}}} = \underbrace{e^{-\frac{1}{2}|d-H(u)|^2_{R^{-1}}}}_{\substack{\propto\, p(d|u) \\ \text{data likelihood}}} \underbrace{e^{-\frac{1}{2}\left|u-u^f\right|^2_{Q^{-1}}}}_{\substack{\propto\, p(u) \\ \text{forecast (prior) density}}} \to \max_u$$

$\propto$ means proportional

u is the Maximum Aposteriori Probability (MAP) estimate.
We got the Bayes theorem for Gaussian probability densities

Bayes theorem in infinite dimension

- Forecast and analysis are probability distributions, densities p^f, p^a
- Bayes theorem: $p^a(u) \propto p(d|u)\, p^f(u)$

- In infinite dimension do not have densities, integrate over an arbitrary set A instead:

$$\mu^a(A) \propto \int_A p(d|u)\, d\mu^f(u)\,, \forall \mu^f - \text{measurable set } A$$

- Data likelihood is the Radon-Nikodym derivative: $p(d|u) \propto \dfrac{d\mu^a}{d\mu^f}$

(e.g., Stuart 2010)

- Normalize: $\mu^a(A) = \frac{\int_A p(d|u) d\mu^f(u)}{\int_V p(d|u) d\mu^f(u)}$

- **But how do we know that** $\int_V p(d|u)\, d\mu^f(u) > 0$?

Infinite dimensional Gaussian case

- data likelihood: $d = Hu + \varepsilon$, $\varepsilon \sim N(0, R)$, $p(d|u) = e^{-\frac{1}{2}|Hu-d|^2_{R^{-1}}}$
- μ^f is Gaussian measure on U, data $d \in V$
- **state space** U and **data space** V are separable Hilbert spaces
- Difficulties when the data are infinite dimensional...

Infinite-dimensional data, Gaussian measure error bad

• The simplest example: $\mu^f = N(0, Q)$, $H = I, d = 0, R = Q, U = V$. The whole state is observed, data error distribution = state error distribution. Come half-way? Wrong.

• $p(d|u) = \text{const}\, e^{-\frac{1}{2}|u|^2_{R^{-1}}} = \text{const}\, e^{-\frac{1}{2}\left\langle R^{-1/2}u, R^{-1/2}u \right\rangle}$
• data likelihood $p(d|u) > 0$ if $u \in R^{1/2}(V) = \mathcal{D}\left(R^{-1/2}\right)$
• $p(d|u) = e^{-\infty} = 0$ if $u \notin R^{1/2}(V) = Q^{1/2}(V)$

• $Q^{1/2}(V)$ is the Cameron-Martin space of the measure $N(0, Q)$

• But $\mu = N(0, Q) \Rightarrow \mu\left(Q^{1/2}(V)\right) = 0$. Thus, $\int_V p(d|u)\, d\mu^f(u) = 0$

Note. The MAP estimate can still be defined in a generalized sense (Dashti et al, 2013).

Commutative case

• State and data covariance commute $\Rightarrow$ same eigenvectors e_i

• $Qe_i = q_i e_i$, $\sum_{i=1}^{\infty} q_i < \infty$, $Re_i = r_i e_i$

• Recall that $p(d|u) = e^{-\frac{1}{2}|d-u|^2_{R^{-1}}}$, $\mu^f = N(0, Q)$

Theorem.

$$\int_V p(d|u)\, d\mu^f(u) > 0 \Leftrightarrow \sum_{i=1}^{\infty} \frac{q_i}{r_i} < \infty$$

That is, **Bayesian estimation is well posed if the eigenvalues of the state covariance decay fast enough compared to the eigenvalues of data covariance**.

• In particular $r_i = 1$, **white data noise** $R = I$, $p(d|u) = e^{-\frac{1}{2}|d-u|^2}$ is always OK because $\sum_{i=1}^{\infty} q_i < \infty$ is needed for μ^f to be a probability measure.

Infinite-dimensional data, white noise error good

- All is good when data is finite-dimensional and R not singular
- More generally, when data covariance R is bounded below:

$$
\begin{aligned}
\langle u, Ru \rangle \geq \alpha \left| u \right|^2 \quad & \forall u, \alpha > 0 \\
& \Rightarrow \left| u \right|^2_{R^{-1}} < \infty \quad \forall u \\
& \Rightarrow \ p\left(d|u\right) = e^{-\left| Hu - d \right|^2_{R^{-1}}} > 0 \quad \forall u \\
& \Rightarrow \int_U p\left(d|u\right) d\mu^f\left(u\right) > 0
\end{aligned}
$$

- But if V is infinite dimensional, then $N\left(0, R\right)$ **is not a probability measure on** V - the trace condition is violated, $\text{Tr}\left(R\right) = \sum_{i=0}^{\infty} r_i = \infty$.
- But this is not a problem. The data likelihood $p\left(d|u\right)$ is just a function of u on the state state U for a fixed d.
- For a fixed u, $p\left(\cdot|u\right)$ does not need to be a probability density.

Positive data likelihood

Theorem. If the forecast μ^f is a probability measure on the state space V, and, for a fixed realization of the data d, the function $u \mapsto p(d|u)$ is μ^f-measurable, and

$$0 \leq p(d|\cdot) \leq C \ \mu^f\text{-a.s.},$$

for some constant C, and $p(d|\cdot) > 0$ on some set of positive measure μ^f. Then the analysis measure

$$\mu^a(A) = \frac{\int_A p(d|u)\, d\mu^f(u)}{\int_V p(d|u)\, d\mu^f(u)}$$

is well defined.

Proof. Because $\mu^f(V) = 1$, from the assumption,
$0 \leq \int_V p(d|u)\, d\mu^f(u) \leq 1$.
If $\int_V p(d|u)\, d\mu^f(u) = 0$, then $p(d|\cdot) = 0$ μ^f- a.s., hence
$\int_V p(d|u)\, d\mu^f(u) > 0$.

Examples of positive data likelihood

- *White noise*: $(V, \langle \cdot, \cdot \rangle)$ is a Hilbert space and

$$p\left(d|u\right) = e^{-\frac{1}{2}\langle d - Hu, d - Hu \rangle}$$

- *Pointwise:* μ^f is a random field on $D \subset \mathbb{R}^2$ with a.s. continuous realizations u, data is a function $d{:}D \to \mathbb{R}$, and

$$p\left(d|u\right) = e^{-\frac{1}{2}\int_D g(d(x), u(x))dx}$$

(the satellite sensing application will be like that)

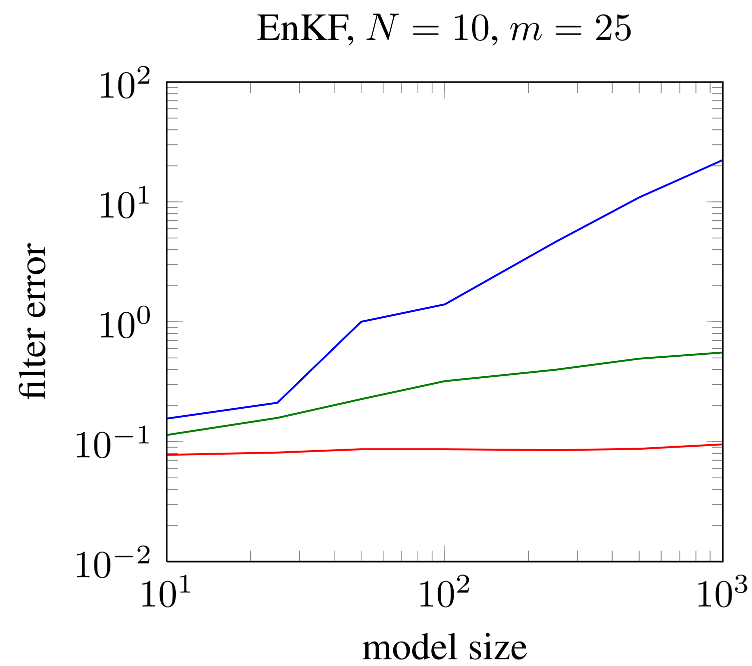
- *General case:*

$$p\left(d|u\right) = e^{-f_d(u)},$$

where $f_d\left(u\right) \geq 0$ for all u and d.

Application 1: Mean field convergence of randomized EnKF with white noise data error in infinite dimension

Curse of dimensionality? Not for probability measures!



One EnKF analysis. **Constant covariance eigenvalues** $\lambda_n = 1$ and the inverse law $\lambda_n = 1/n$ are not probability measures in the limit because $\sum_{n=1}^{\infty} \lambda_n = \infty$.

Inverse square law $\lambda_n = 1/n^2$ gives a probability measure because $\sum_{n=1}^{\infty} \lambda_n < \infty$.

m=25 uniformly sampled data points from 1D state, N=10 ensemble members.

From the thesis Beezley (2009), Fig. 4.7. **Similarly for particle filters.**

Randomized data EnKF

Lemma. *Let* $U^f \sim N(u^f, Q^f)$*,* $D = d + E$*,* $E \sim N(0, R)$ *and*

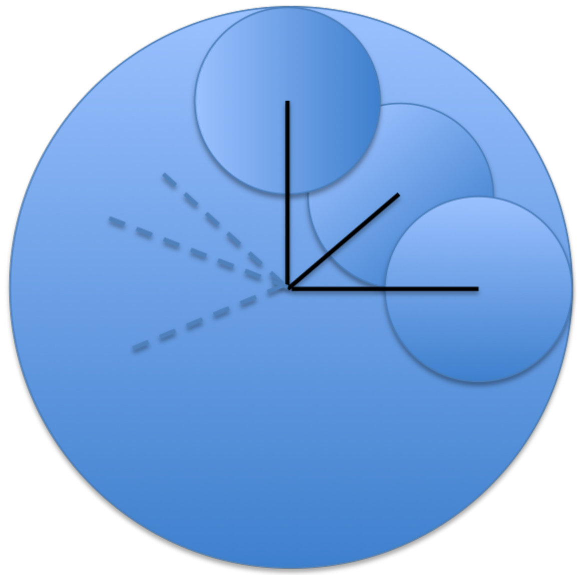
$$U^a = U^f + K(D - HU^f), \quad K = Q^f H^\mathsf{T}(HQ^f H^\mathsf{T} + R)^{-1}.$$

Then $U^a \sim N(u^a, Q^a)$*, i.e.,* U *has the correct analysis distribution from the Bayes theorem.*

Proof. Computation in Burgers at al., 1998.

• EnKF: **update every ensemble member separately** by this, replacing covariance by an approximation from the ensemble.
• $R = I$ guarantees that the posterior measure is well defined. But $D \sim N(0, I)$ is not a random element in infinite dimension. How do we know that U^a is random element? (=measurable function with values in the state space)

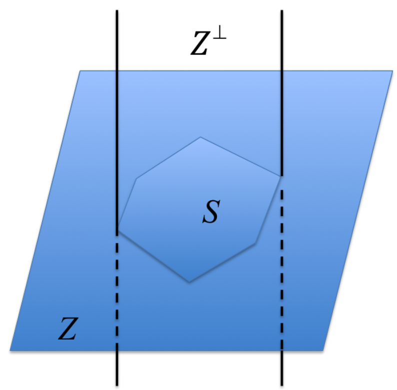
There is no probability measure on infinite dimensional Hilbert space that is translation or rotation invariant, with balls measurable



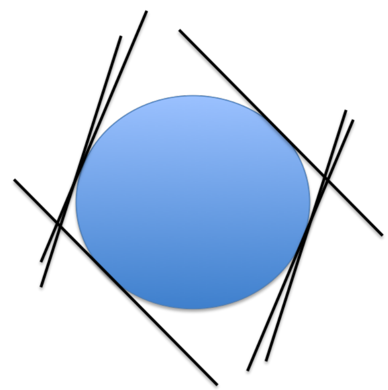
Proof: There are infinitely many orthonormal vectors. Put a ball with radius $1/2$ at the end of each, the balls all have the same positive measure and fit in a ball at zero with radius $3/2$, which, therefore, cannot have a finite measure. In particular, $N(0, I)$ cannot be a (σ-additive) probability measure such that balls are measurable.

How to define $N(0, I)$ on a separable Hilbert space

• μ (whole space Y) = 1, rotation invariant $\Rightarrow$ *balls not measurable*
• Defined on the algebra $\mathcal{C}$ of cylinder sets with finite dimensional Borel measurable base



Cylinder set



Ball $= \bigcap_{\text{countable}}$ cylinder sets

• μ cannot be extended to a σ-algebra, which would contain balls
• μ cannot be σ-additive, only finitely additive. Balls would be measurable.
Consider 1D base => weak random variable

Weak random vectors on Hilbert space

• $U : \Omega \to H$ such that $\forall v \in H$ the function $\langle U, v\rangle : \Omega \to \mathbb{R}$ is measurable

• Mean m of U defined by $\langle m, v\rangle = E\left[\langle U, v\rangle\right] \quad \forall v \in H$

• Covariance C of U defined by $\langle Cu, v\rangle = E\left[\langle U, u\rangle \langle U, v\rangle\right] \quad \forall u, v \in H$

• $U \sim N(0, I)$ (white noise) means
$\langle U, v\rangle \sim N(0, 1)$ $\forall v \in H$, and $E\left[\langle U, u\rangle \langle U, v\rangle\right] = 0$ $\forall u, v \in H, u \perp v$

Hilbert-Schmidt operators

Denote V_{HS} the space of Hilbert-Schmidt operators

$$|A|_{HS}^2 = \sum_{n=1}^{\infty} \langle Ae_n, Ae_n \rangle < \infty, \quad \langle A, B \rangle_{HS} = \sum_{n=1}^{\infty} \langle Ae_n, Be_n \rangle,$$

where $\{e_n\}$ is any complete orthonormal sequence in V. V_{HS} is Hilbert space. In finite dimension, the Hilbert-Schmidt norm becomes the Frobenius norm

$$|A|_{HS}^2 = \sqrt{\sum_{i,j} \left|a_{ij}\right|^2}.$$

Hilbert-Schmidt operators are compact, with singular values σ_k, and

$$|A|_{HS}^2 = \sum_{n=1}^{\infty} \sigma_k^2$$

If V is separable, V_{HS} is also separable.

Hilbert-Schmidt operators make weak random variables strong

Recall standard (strong) L^p -norms $\|U\|_p = \left(E\left[|U|^p\right]\right)^{1/p}$

Define **weak** L^p**-norms**: $\|U\|_{p,w} = \sup_{v \in H, |v|=1} \left(E\left[|\langle U, v\rangle|^p\right]\right)^{1/p}$

Lemma. If A is a Hilbert-Schmidt operator and U a weak random element on a Hilbert space, $\|U\|_{p,w} < \infty$, $2 \le p < \infty$ then

$$\|AU\|_p \le \|A\|_{\mathsf{HS}} \|U\|_{p,w}$$

Lemma If A is a random Hilbert-Schmidt operator and U a weak random element on a Hilbert space, $\|U\|_{p,w} < \infty$, $p \ge 2$, A and U independent, then

$$\|AU\|_p \le \left\||A|_{\mathsf{HS}}\right\|_p \|U\|_{p,w}$$

Convergence of EnKF in the large ensemble limit

• L^p laws of large numbers to guarantee that the EnKF gives correct results for large ensembles, in the Gaussian case: Le Gland et al. (2011), Mandel et al (2011).

• In general, the EnKF converges to a mean-field limit (Le Gland et al. 2011), Law et al. (2014).

- **mean-field approximation** = the effect of all other particles on any one particle is replaced by a single averaged effect, in the limit of many particles.
- **mean field limit** = large number of particles, the influence of each becomes negligible.
- Here, mean field simply means **using the covariance** of random variable passed through the model and the analysis step.

Convergence of the EnKF to mean-field limit in finite dimension

• Legland et al. (2011): analysis step as nonlinear transformation of probability measures

• Law et al. (2014): mean field convergence for model as general Markov chain

• Nonlinear tranformation of **ensemble as vector of exchangeable random variables** $[X_1, X_2, \ldots, X_N] \mapsto [Y_1, Y_2, \ldots, Y_N]$. L^p continuity of the model.

• Using exact covariance Q:
$Y_k^{\text{Mean field}} = X_k^{\text{Mean field}} + \mathcal{K}(Q)(D_k - HX_k^{\text{Mean field}})$, $Q = \text{Cov}\,(X_1)$

• Randomized EnKF:
$Y_k^{\text{Computed}} = X_k^{\text{Computed}} + \mathcal{K}(Q_N)(D_k - HX_k^{\text{Computed}})$,
Q_N =ensemble covariance

• Subtract the two...

Convergence of the EnKF to mean-field limit in infinite dimension

• Subtract, continuity of Kalman gain:
$\| |\mathcal{K}(Q) - \mathcal{K}(Q_N)|_{HS} \|_p \leq \text{const} \, \|Q - Q_N\|_{2p}$
• Same realization of white noise D_k,
$\|(\mathcal{K}(Q) - \mathcal{K}(Q_N)) \, D_k\|_p \leq \| |\mathcal{K}(Q) - \mathcal{K}(Q_N)|_{HS} \|_p \, \|D_k\|_{p,w}$
• L^p law of large numbers for sample covariance **in Hilbert-Schmidt norm**
• Apriori bound on the state $\left\| X_k^m \right\|_p \leq \text{const}\,(m)$ for all m from $\left| (HQH^* + R)^{-1} \right| \leq \frac{1}{\alpha}$ by $R \geq \alpha I$
• Induction over m: $\left\| X_1^{m,\text{Computed}} - X_1^{m,\text{Mean field}} \right\|_p \leq \frac{\text{const}_{m,p}}{\sqrt{N}}$, $1 \leq p < \infty$

Extensions of EnKF analysis

• Computational experiments confirm that EnKF converges uniformly for high-dimensional distributions that approximate a Gaussian measure on Hilbert space. (J. Beezley, Ph.D. thesis, 2009). EnKF for distributions with slowly decaying eigenvalues of the covariance converges very slowly and requires large ensembles.

• Square root EnKF (SREnKF) has no randomization, convergence in the linear Gaussian case Kwiatkowski and Mandel (2015) including infinite dimension, but controls mean and covariance only, not individual ensemble members

• SREnKF analysis in the nonlinear case requires control of ensemble members. In progress with Ivan Kasanicky and Kody Law

• Long-term convergence of the EnKF: combine with ergodic estimation. In progress with Kody Law.

References

[1] A. V. Balakrishnan. *Applied functional analysis*. Springer-Verlag, New York, 1976.

[2] Jonathan D. Beezley. *High-Dimensional Data Assimilation and Morphing Ensemble Kalman Filters with Applications in Wildfire Modeling*. PhD thesis, University of Colorado Denver, 2009.

[3] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

[4] Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Springer-Verlag, Berlin, 2006.

[5] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 27, 2013.

[6] Geir Evensen. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 (C5)(10):143–162, 1994.

[7] Evan Kwiatkowski and Jan Mandel. Convergence of the square root ensemble Kalman filter in the large ensemble limit. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1–17, 2015.

[8] Kody J. H. Law, Hamidou Tembine, and Raul Tempone. Deterministic methods for filtering, part I: Mean-field ensemble Kalman filtering. arXiv:1409.0628, 2014.

[9] F. Le Gland, V. Monbet, and V.-D. Tran. Large sample asymptotics for the ensemble Kalman filter. In Dan Crisan and Boris Rozovskiĭ, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 598–631. Oxford University Press, 2011.

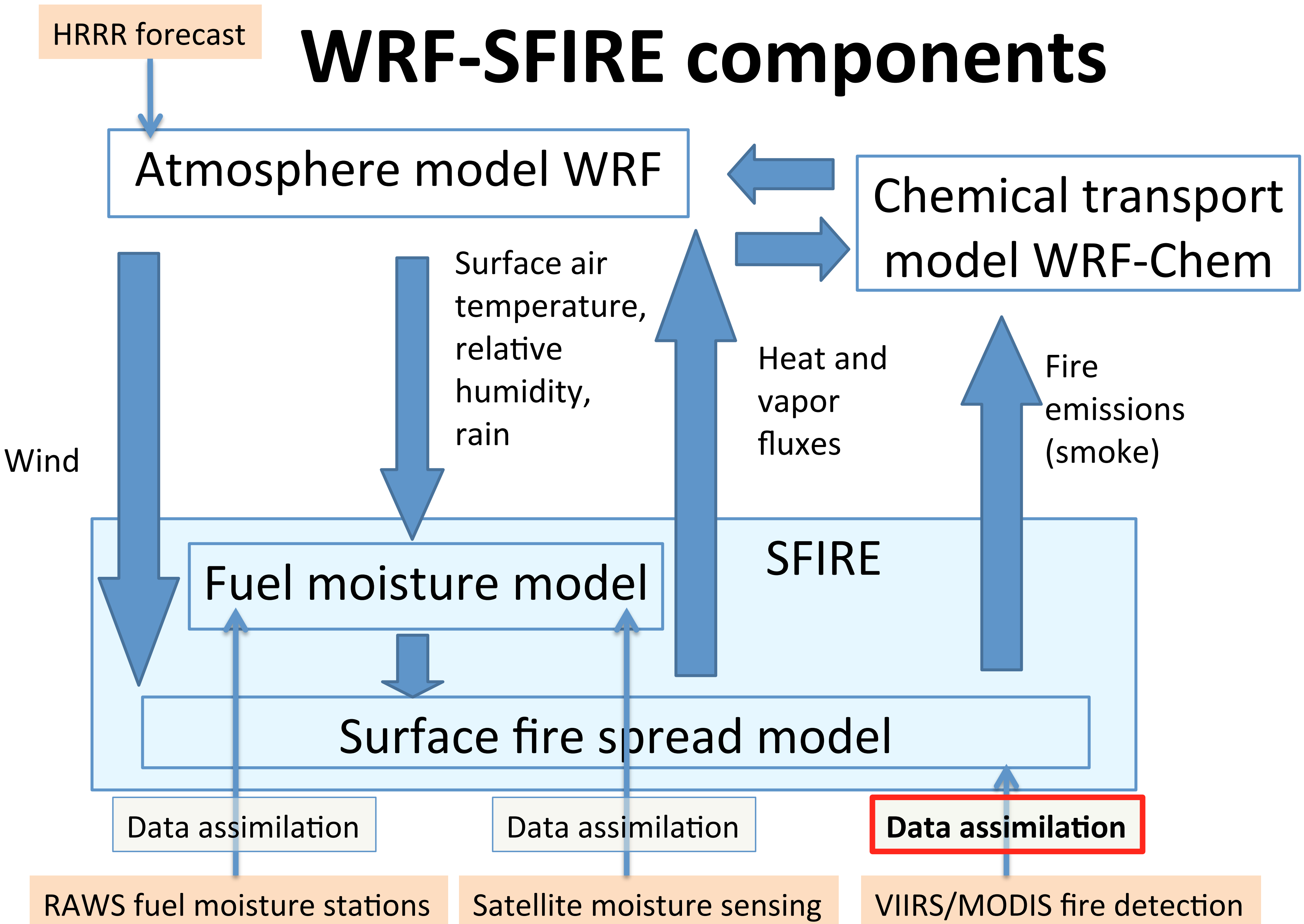
[10] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Vol. 23. Springer-Verlag, Berlin, 1991.

[11] Jan Mandel, Loren Cobb, and Jonathan D. Beezley. On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, 56:533–541, 2011.

[12] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

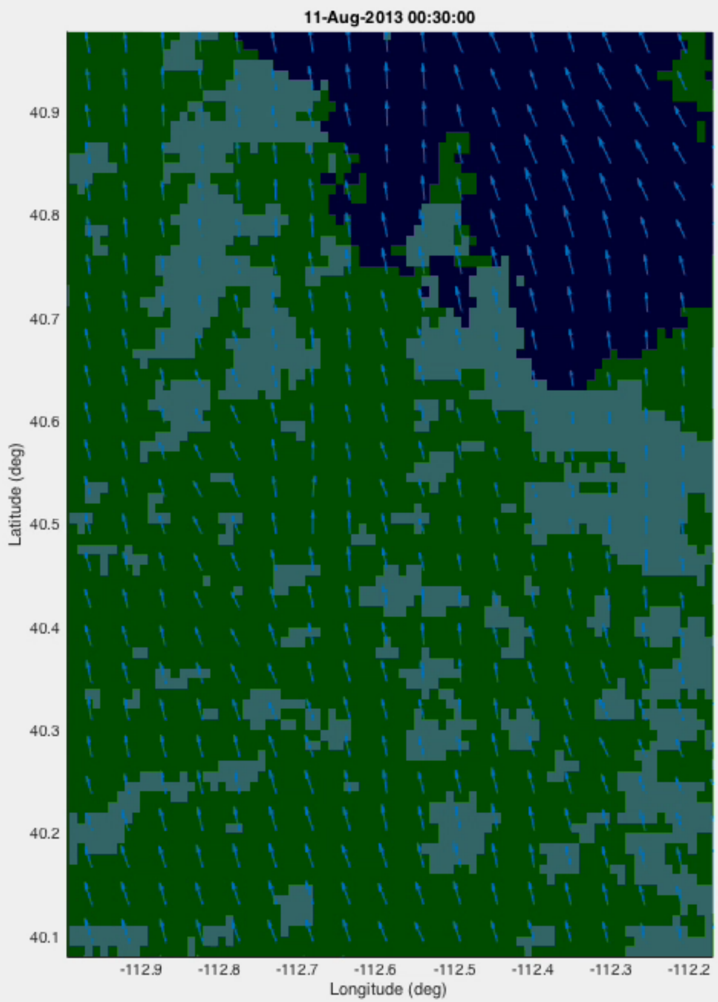
Application 2: Data Assimilation of Satellite Active Fire Detection in Wildfire Simulations

WRF-SFIRE components



2013 Patch Springs Fire

WRF-SFIRE simulation with MODIS/VIIRS active fires detection



Fireline (simulation)

Fire detection > 60hrs old
- Low confidence
- Nominal confidence
- High confidence

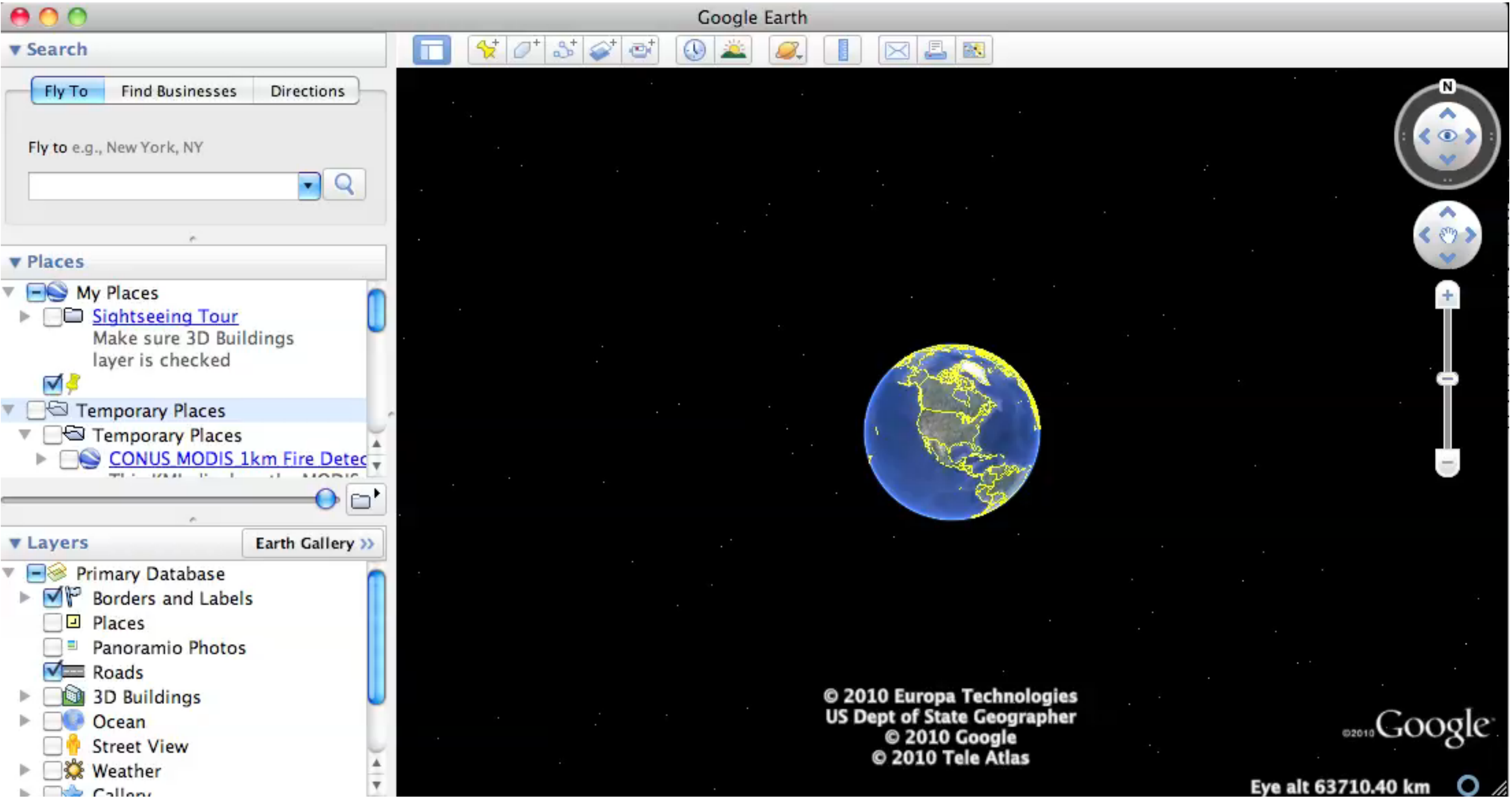
Fire detection 0-6 hrs old
- Low confidence
- Nominal confidence
- High confidence

- Water
- Ground – no fire
- Cloud – no detection

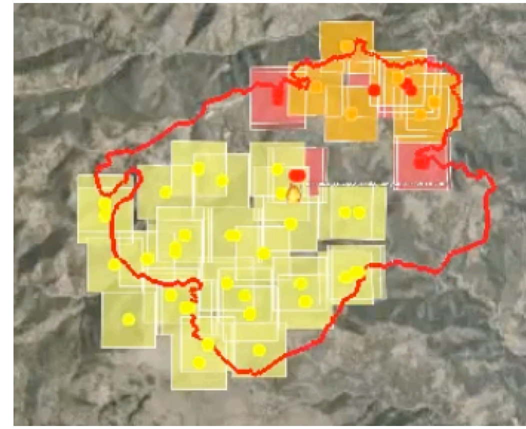
MODIS scanning

Source: NASA

Satellite Fire Detection – 2010
Fourmile Canyon Fire, Boulder, CO



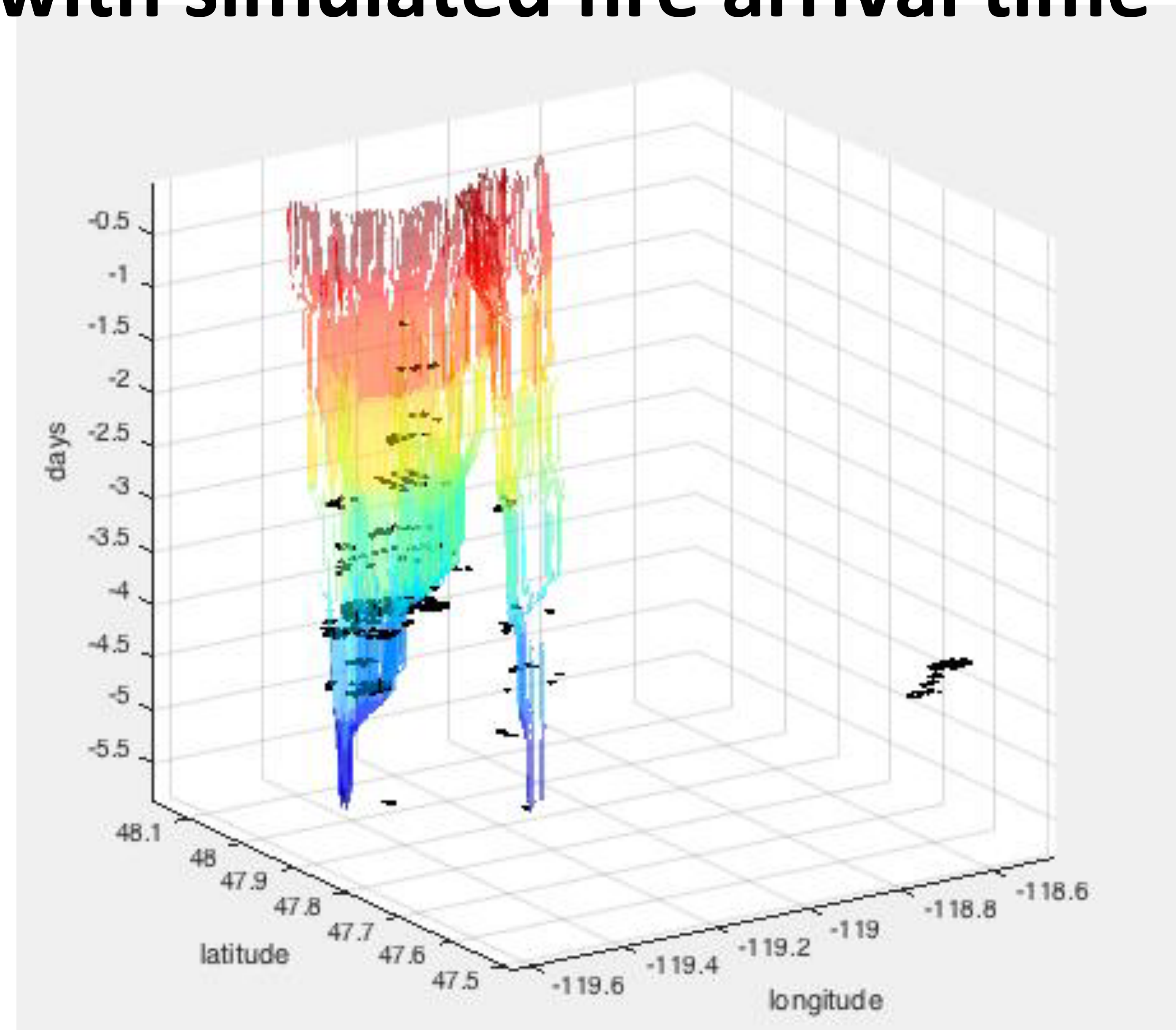
MODIS/VIIRS Active Fire Detection Data



- Detection squares - fire **sensed somewhere in the square**, not that the whole square would be burning.
- Level 3 product – 1km detection squares (used here)
- Level 2 product – 0.1deg grid, confidence levels, cloud mask

- MODIS instrument native resolution 750m at nadir to 1.6km, geo-location uncertainty up to 1.5km, VIIRS resolution 375m.
- MODIS processed to 1.1km detection squares, VIIRS 375m. **Much coarser scale than fire behavior models** (10-100m)
- **False negatives are common**. 90% detection at best. 100m^2 flaming fire has 50% detection probability (MODIS. VIIRS is better but **nothing can be ever 100% accurate**).
- No detection under cloud cover – cloud mask in Level 2 product

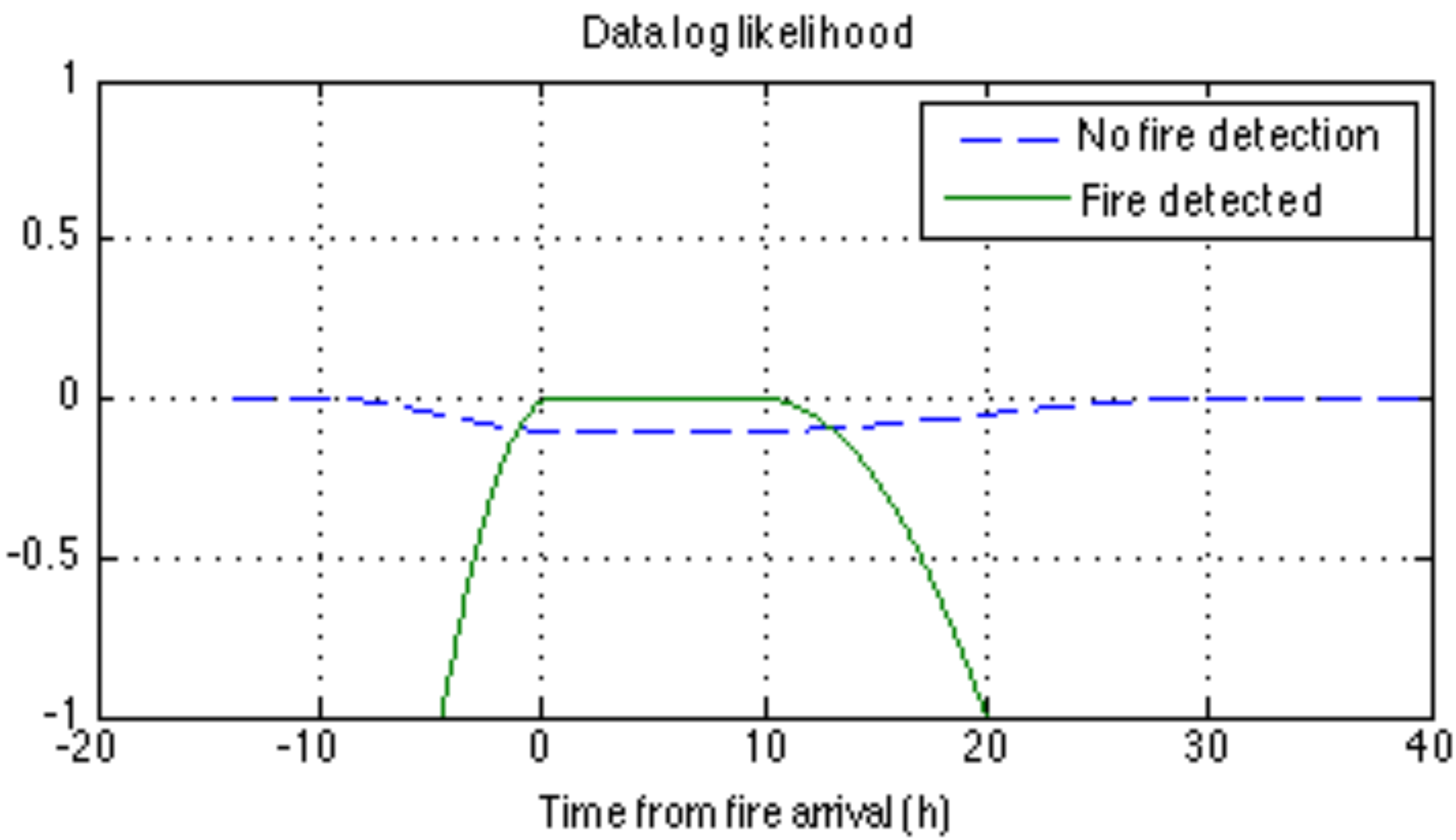
MODIS active fires detection with simulated fire arrival time



Assimilation of active fires detection

- ***Fire model state = fire arrival time***
- ***Modify the fire arrival time*** to *simultaneously minimize the change and to maximize the likelihood of the observed fire detection*.
- Need more general data likelihood than $e^{-\frac{1}{2}\left\|H(u)-d\right\|_{R^{-1}}^2}$
- Inspired by computer vision in Microsoft Kinect, which modifies a level set function for contour detection to simultaneously minimize the change and to maximize the likelihood of the observed images (A. Blake, Gibbs lecture at Joint Math Meetings, Baltimore 2014)
- Bayesian statistics view: Maximum Aposteriori Probability, found by nonlinear least squares.

$f(t,x,y)$: log of the likelihood of fire detection as a function of the time t elapsed since the fire arrival at the location (x,y)



Assimilation of MODIS/VIIRS Active Fire detection: generalized least squares

Fit the **fire arrival time** T to the forecast T^{f} and the fire detection data:

$$J(T) = -\int c(x,y)\, f(T^{S} - T, x, y)\, dx\, dy + \frac{\alpha}{2}\left\| T - T^{f} \right\|_{A^{-1}}^{2} \rightarrow \min_{C(T - T^{f}) = 0}$$

- T^{S} = satellite overpass time
- constraint $C(T\text{-}T^{s})$=0 : no change of fire arrival time at ignition points
- $f(t,x,y)$ = log likelihood of detection t hours after time arrival at x,y
- $c(x,y)$ = confidence level of the fire detection (0 = cloud)
- A = covariance operator to penalize non-smooth changes:

$$A = \left(-\frac{\partial^2}{\partial^2 x} - \frac{\partial^2}{\partial^2 y} \right)^{-a} \qquad a > 1$$

Assimilation of MODIS/VIIRS Active Fire detection as Maximum Aposteriori Probability

Fit the fire arrival time T to the forecast T^{f} and fire detection data

$$J(T) = -\int f(T^{\mathrm{S}} - T, x, y)\,dx\,dy + \frac{\alpha}{2}\left\| T - T^{\mathrm{f}} \right\|_{A^{-1}}^{2} \rightarrow \min_{T:\, C(T-T^{\mathrm{f}})=0}$$

$$\Leftrightarrow \quad e^{\int f(T^{\mathrm{S}}-T,x,y)\,dx\,dy} \; e^{-\frac{\alpha}{2}\left\|T-T^{\mathrm{f}}\right\|_{A^{-1}}^{2}} \rightarrow \max_{T:\, C(T-T^{\mathrm{f}})=0}$$

$$\Leftrightarrow \quad p(\mathrm{detection}|T) \quad p^{\mathrm{f}}(T) \quad \rightarrow \max_{T:\, C(T-T^{\mathrm{f}})=0}$$

Minimization by preconditioned steepest descent

$$J(T) = -\int f(T^{\mathrm{S}} - T, x, y)\,dxdy + \frac{\alpha}{2}\left\|T - T^{\mathrm{f}}\right\|_{A^{-1}}^{2} \to \min_{C(T - T^{\mathrm{f}})=0}$$

$$\nabla J(T) = -F(T) + \alpha A(T - T^{\mathrm{f}}), \quad F(T) = \frac{\partial}{\partial t} f(T^{\mathrm{S}} - T, x, y)$$

But $\nabla J(T)$ is a terrible descent direction, A ill conditioned

- no progress at all!

Better: preconditioned descent direction $A\nabla J(T) = \alpha(T - T^{\mathrm{f}}) - AF(T)$

$$AF(T) = \left(-\frac{\partial^2}{\partial^2 x} - \frac{\partial^2}{\partial^2 y}\right)^{-a} \frac{\partial}{\partial t} f(T^{\mathrm{S}} - T, x, y)$$

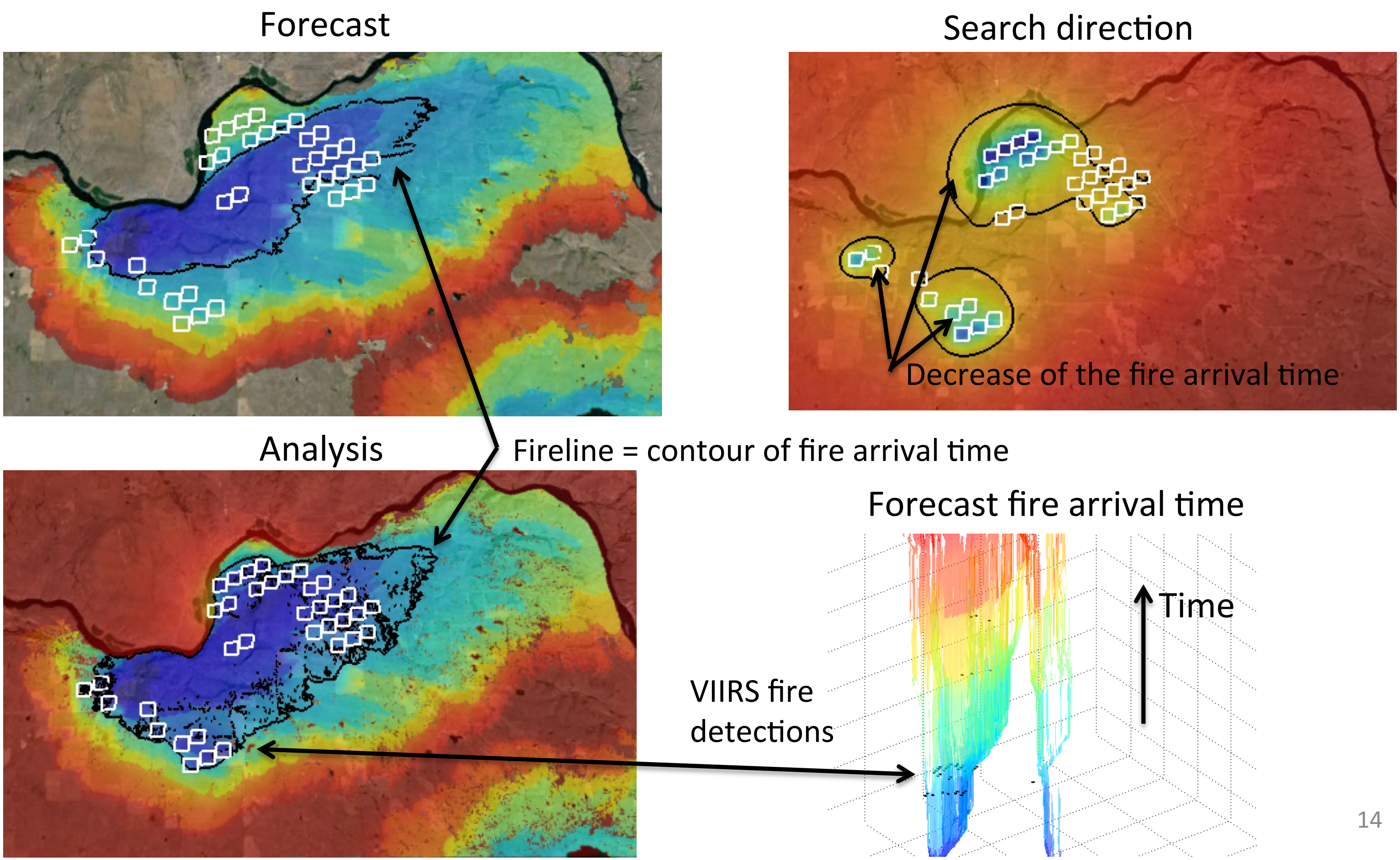
$a > 1$: spatial smoothing of the forcing by log likelihood maximization

T at ignition point does not change $\Rightarrow$ descent direction δ from

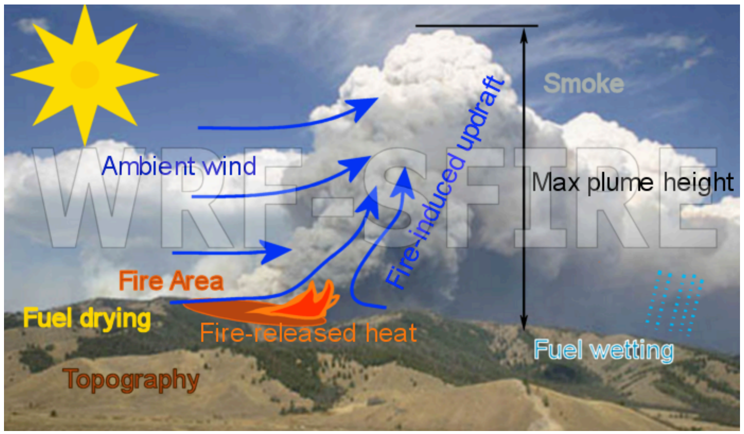
the saddle point problem $A\delta + C\lambda = \dfrac{\partial}{\partial t} f(T^{\mathrm{S}} - T, x, y),\ C^{\mathrm{T}}\delta = 0$

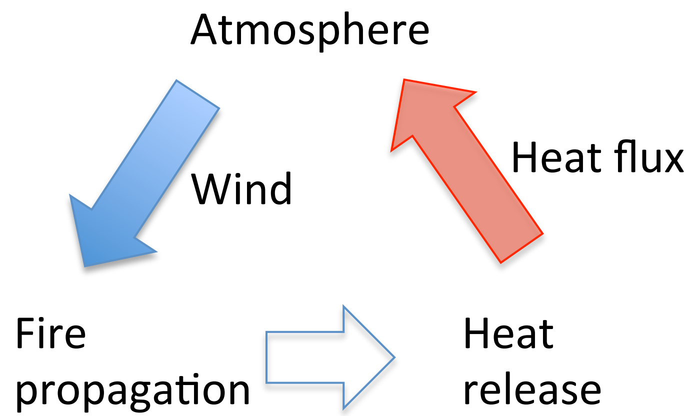
Now one descent iteration is enough.

Assimilation of the VIIRS Fire Detection into the Fire Arrival Time for the 2012 Barker Fire



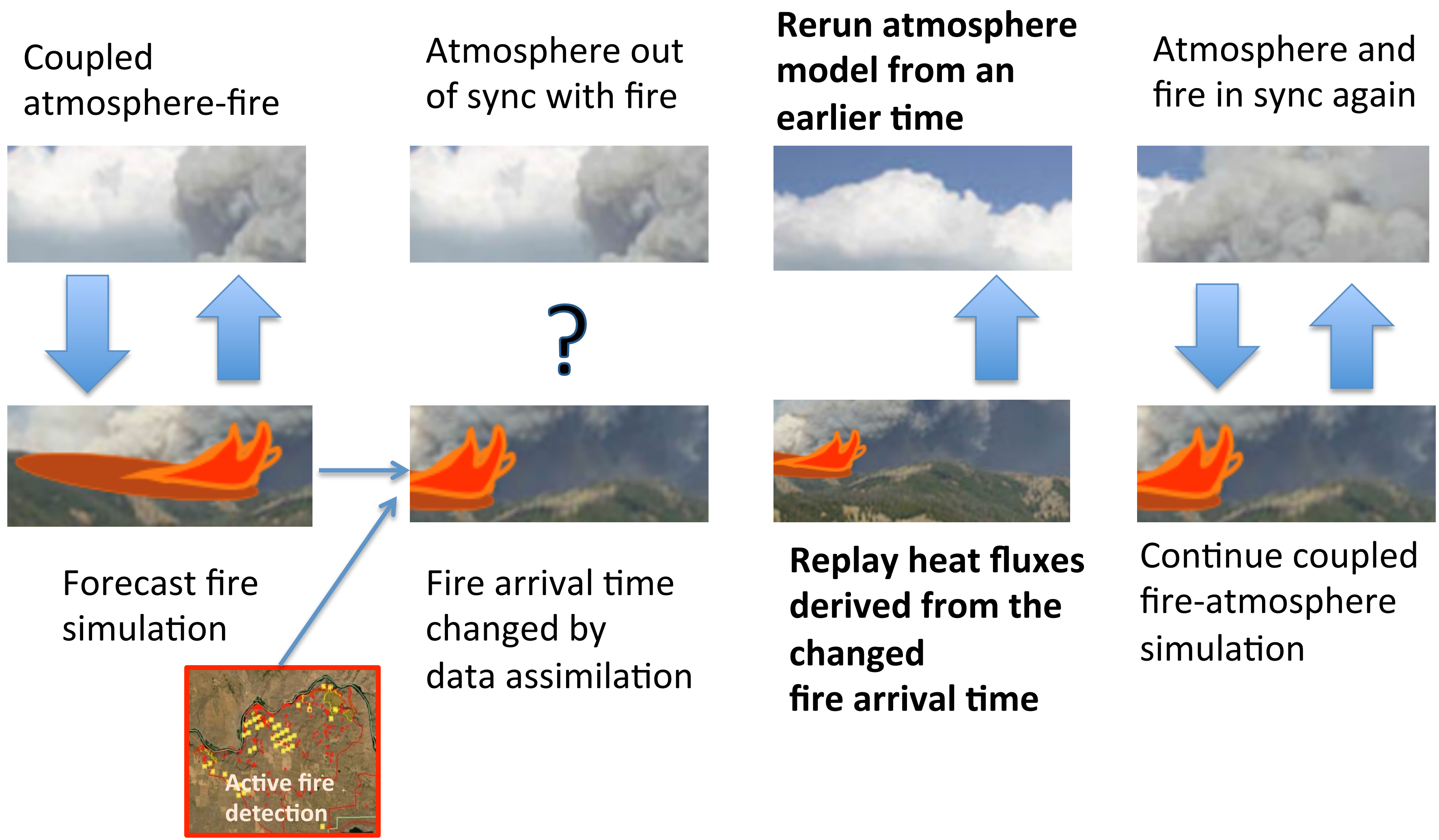
But fire is coupled with the atmosphere





- Heat flux from the fire changes the state of the atmosphere over time.
- Then the fire model state changes by data assimilation.
- The atmospheric state is no longer compatible with the fire.
- How to change the state of the atmosphere model in response data assimilation into the fire model?
- And not break the atmospheric model.

Spin up the atmospheric model after the fire model state is updated by data assimilation



Conclusion 2

- A simple and efficient method – implemented by FFT
- One iteration is sufficient to minimize the cost function in practice, further iterations do not improve much
- Pixels under cloud cover do not contribute to the cost function
- Standard Bayesian data assimilation framework: Forecast density • data likelihood = analysis density
- In progress: Active fire detection likelihood from the physics and the instrument properties
- Future: Combination with standard data assimilation into the atmospheric model, e.g., add to 4DVAR cost function