The Statistical Analysis of Satellite Retrievals

Noel Cressie

(ncressie@stat.osu.edu)

Program in Spatial Statistics

and Environmental Statistics

(www.stat.ohio-state.edu/∼sses)

The Ohio State University

Acknowledgment: Rui Wang, Amy Braverman, Emily Kang, Tao Shi, and the ACOS team

Sample of Size One?

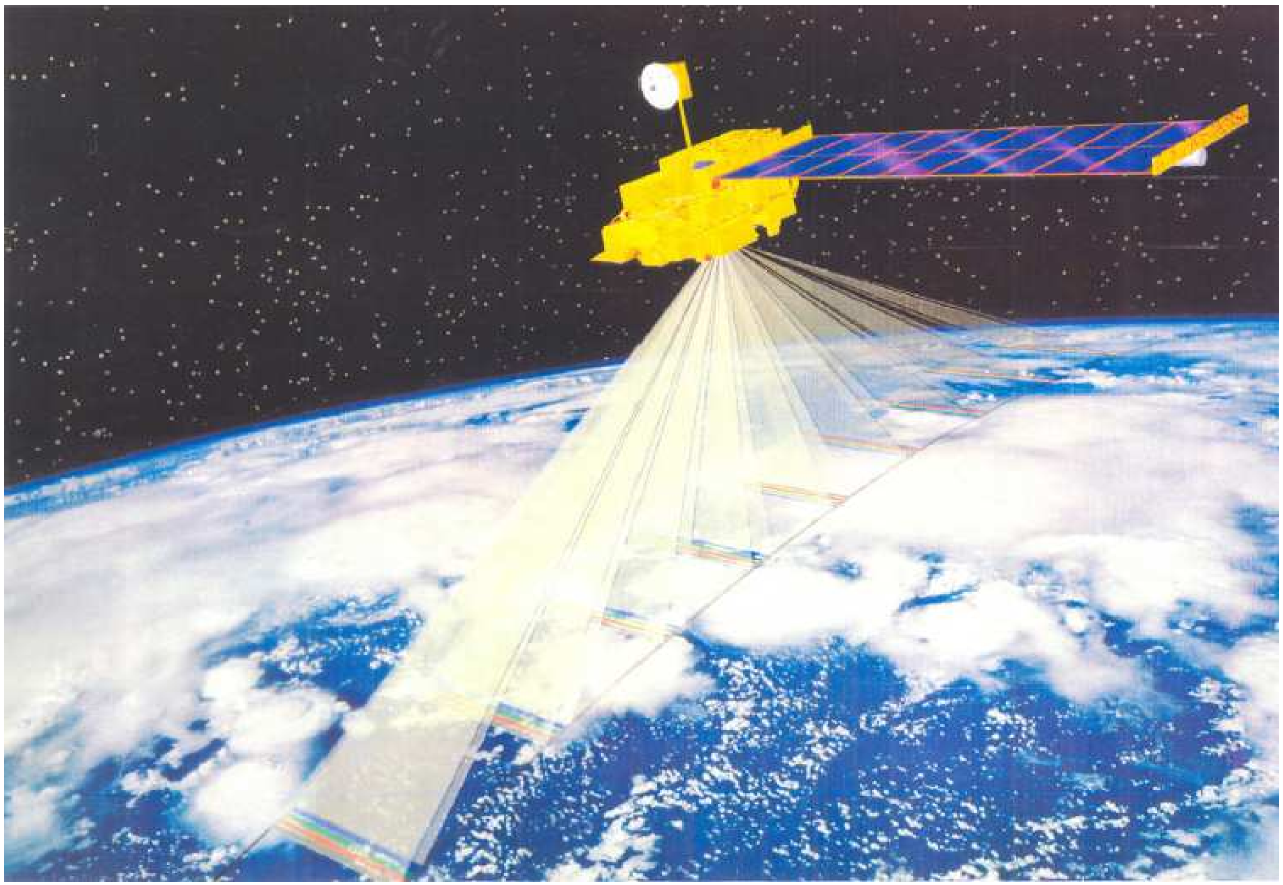
Where is the probability space to enable Uncertainty Quantification (UQ)?

Satellite Launch

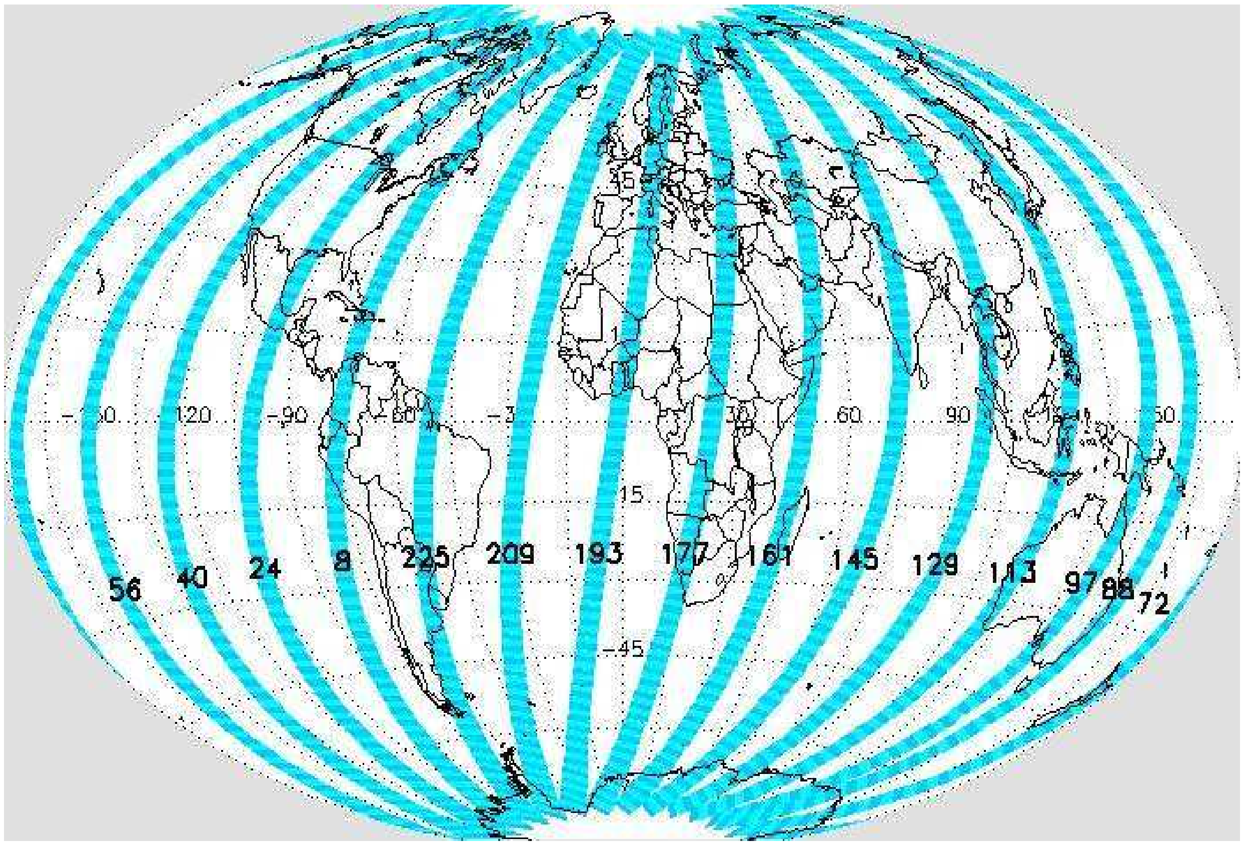
Putting small, delicate instruments in nose cones of large, powerful rockets and launching them, is a risky business

Terra Satellite (MISR Instrument Featured)

Physical properties (e.g., atmospheric AOD) are detected indirectly through spectral radiances



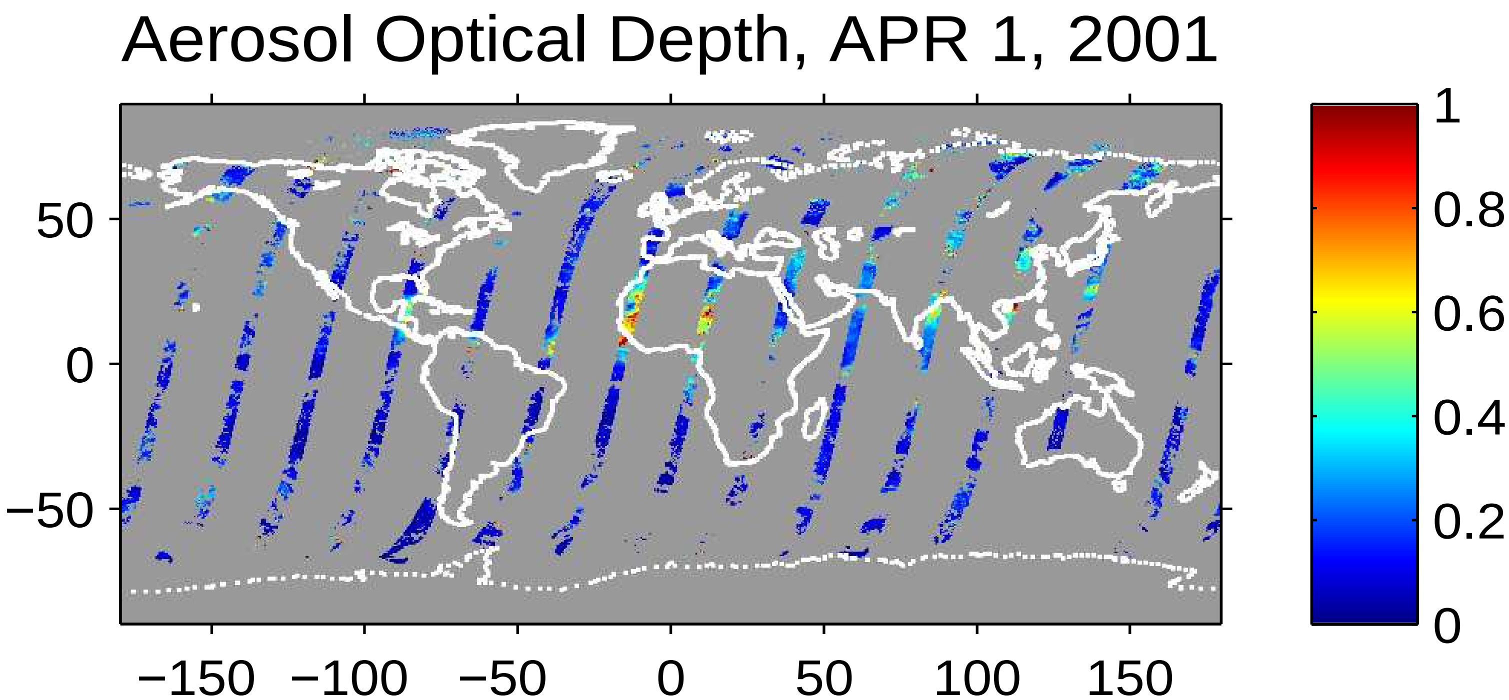
Multi-angle Imaging SpectoRadiometer (MISR) Path and Coverage



MISR repeats the same path (360km wide) every 16 days

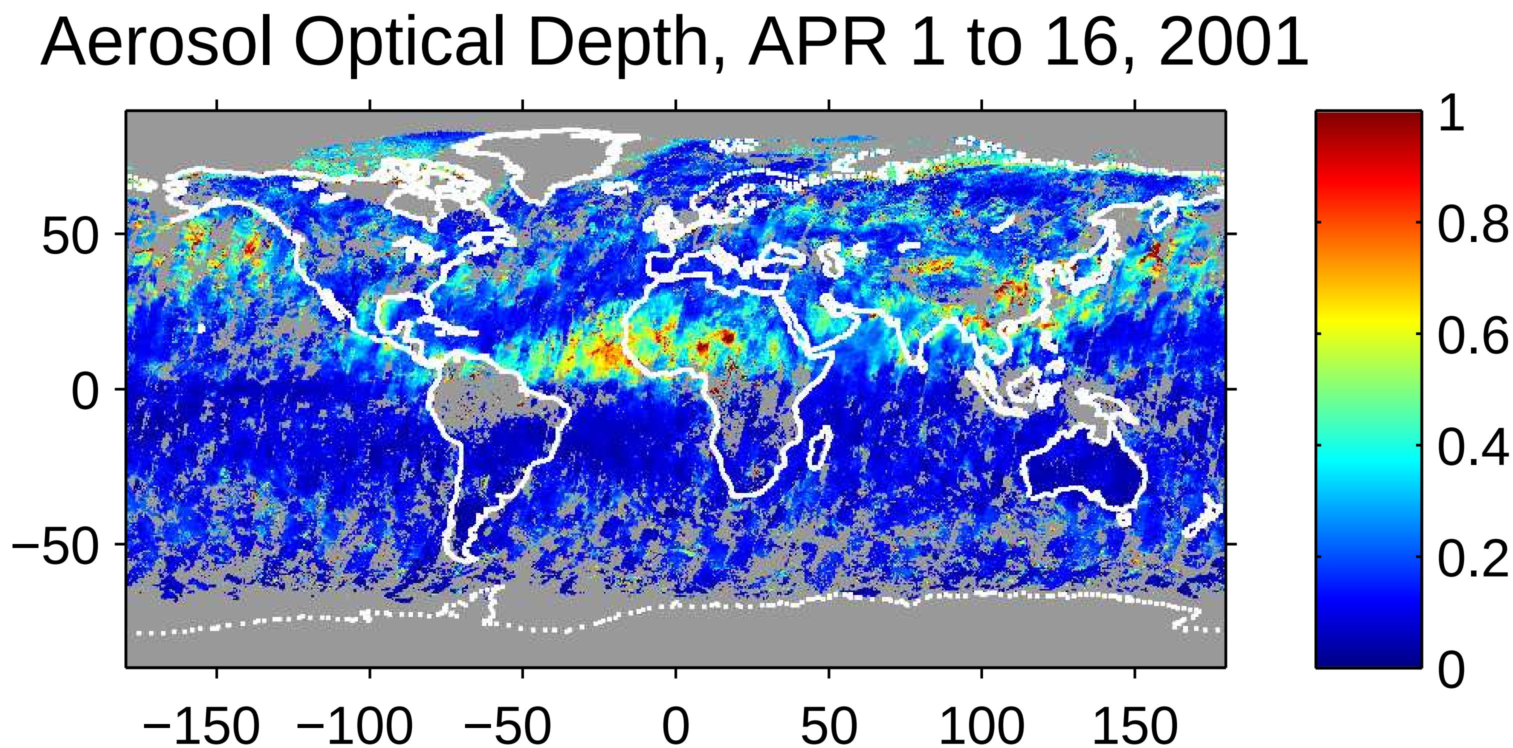
MISR Level 2 Aerosol Optical Depth (AOD) Coverage on 4/1/02

Even accounting for the orbit geometry, retrievals everywhere are not possible



MISR Level 3 AOD (Band 3,558nm) for 16-days repeat cycle

Missing data are due to orbit geometry, cloud cover, non-convergent retrieval algorithms



720×360 pixels on a $0.5° \times 0.5°$ global map

Generic Retrieval

- Aerosols: Band 3 (centered on 558nm wavelength) records **radiances $\mathbf{y}$**; large **aerosol optical depth (AOD)** means that the aerosol fraction is high
- Forward model:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}$$

- Want to estimate the **state $\mathbf{x}$** or some function $g(\mathbf{x})$; for example, AOD
- $\widehat{\text{AOD}} = g(\widehat{\mathbf{x}})$

Generic Retrieval: Linear Forward Model

- **Linear forward model**: Consider solving for the state vector $\mathbf{x}$ in the statistical model,

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon}\,,$$

  where $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are n_ε-dimensional, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{S}_\varepsilon$
- State equation: Assume

$$\mathbf{x} = \mathbf{x}_\alpha + \boldsymbol{\alpha}\,,$$

  where $\mathbf{x}$, $\mathbf{x}_\alpha$, and $\boldsymbol{\alpha}$ are n_α-dimensional, $E(\boldsymbol{\alpha}) = 0$ and $\text{var}(\boldsymbol{\alpha}) = \mathbf{S}_\alpha$
- the **optimal estimator** $\widehat{\mathbf{x}}$, maximizes the posterior distribution, $P(\mathbf{x}|\mathbf{y})$
- For the linear forward model, $\widehat{\mathbf{x}}$ has a closed-form solution:

$$\widehat{\mathbf{x}} = \mathbf{x}_\alpha + \{\mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{S}_\alpha^{-1}\}^{-1}\mathbf{K}'\mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_\alpha)$$

Bias and MSPE of Retrieval: Linear Forward Model

- For the linear forward model, $\mathbf{y} = \mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon}$. Consequently,

$$E(\widehat{\mathbf{x}} - \mathbf{x}) = \mathbf{0}\,,$$

and the MSPE is:

$$\begin{aligned} E((\widehat{\mathbf{x}} - \mathbf{x})(\widehat{\mathbf{x}} - \mathbf{x})') &= \text{var}(\mathbf{x}|\mathbf{y}) \equiv \widehat{\mathbf{S}} \\ &= \{\mathbf{S}_\alpha^{-1} + \mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K}\}^{-1} \\ &= \mathbf{A}\{\mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K}\}^{-1} \\ &= (\mathbf{A} - \mathbf{I})\mathbf{S}_\alpha(\mathbf{A} - \mathbf{I})' + \mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}'\,, \end{aligned}$$

where
$\mathbf{G} \equiv \{\mathbf{S}_\alpha^{-1} + \mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K}\}^{-1}\mathbf{K}'\mathbf{S}_\varepsilon^{-1}$, is the *gain* matrix
$\mathbf{A} \equiv \mathbf{G}\mathbf{K}$, is the averaging-kernel matrix;
the vector $\mathbf{A}(1, \ldots, 1)'$, plotted as a function of row number, is known as the **averaging kernel**

Nonlinear Forward Model

- For the non-linear case, Rodgers (2000) recommends the Levenberg-Marquardt iteration scheme: Put $\mathbf{K}_\ell = \mathbf{K}(\mathbf{x}_\ell)$; then $\widehat{\mathbf{x}} = \mathbf{x}_\infty$ in

$$\mathbf{x}_{\ell+1} = \mathbf{x}_\ell + \{(1+\gamma_\ell)\mathbf{S}_\alpha^{-1} + \mathbf{K}_\ell'\mathbf{S}_\varepsilon^{-1}\mathbf{K}_\ell\}^{-1}[\mathbf{K}_\ell'\mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}_\ell)) - \mathbf{S}_\alpha^{-1}(\mathbf{x}_\ell - \mathbf{x}_\alpha)]$$

- This approach is sometimes called **optimal estimation** in the remote-sensing literature
- The bias vector and mean squared prediction error matrix of the state vector $\widehat{\mathbf{x}}$ are approximated using the **“delta method”**
- Define

$$\begin{aligned} \mathbf{K}(\mathbf{x}) &\equiv \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \\ \mathbf{G}(\mathbf{x}) &\equiv \{\mathbf{S}_\alpha^{-1} + \mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1}\mathbf{K}(\mathbf{x})\}^{-1}\mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1} \\ \mathbf{A}(\mathbf{x}) &\equiv \mathbf{G}(\mathbf{x})\mathbf{K}(\mathbf{x}) \end{aligned}$$

Bias and MSPE of Retrieval: Nonlinear Forward Model

- Calculating the matrix of approximate mean squared prediction errors is more straightforward than calculating the vector of approximate biases:

$$\begin{aligned} E\{(\widehat{\mathbf{x}}-\mathbf{x})(\widehat{\mathbf{x}}-\mathbf{x})'\} &\simeq \widetilde{\mathbf{MSPE}}(\mathbf{x}_\alpha) \\ &\equiv (\mathbf{A}(\mathbf{x}_\alpha)-\mathbf{I})\mathbf{S}_\alpha(\mathbf{A}(\mathbf{x}_\alpha)-\mathbf{I})' + \mathbf{G}(\mathbf{x}_\alpha)\mathbf{S}_\varepsilon\mathbf{G}(\mathbf{x}_\alpha)' \end{aligned}$$

- On the next slide, we obtain $E(\widehat{\mathbf{x}}-\mathbf{x}) \simeq \widetilde{\mathbf{bias}}(\mathbf{x}_\alpha)$
- In deriving the vector of approximate biases, we need the following expressions, obtained using the delta method:

$$\begin{aligned} \text{var}(\widehat{\mathbf{x}}) &\simeq \widetilde{\text{var}}(\widehat{\mathbf{x}}) \equiv \mathbf{A}(\mathbf{x}_\alpha)\mathbf{S}_\alpha\mathbf{A}(\mathbf{x}_\alpha)' + \mathbf{G}(\mathbf{x}_\alpha)\mathbf{S}_\varepsilon\mathbf{G}(\mathbf{x}_\alpha)' \\ \text{cov}(\widehat{\mathbf{x}},\mathbf{x}) &\simeq \widetilde{\text{cov}}(\widehat{\mathbf{x}},\mathbf{x}) \equiv \mathbf{A}(\mathbf{x}_\alpha)\mathbf{S}_\alpha \\ \text{cov}(\widehat{\mathbf{x}},\boldsymbol{\varepsilon}) &\simeq \widetilde{\text{cov}}(\widehat{\mathbf{x}},\boldsymbol{\varepsilon}) \equiv \mathbf{G}(\mathbf{x}_\alpha)\mathbf{S}_\varepsilon \end{aligned}$$

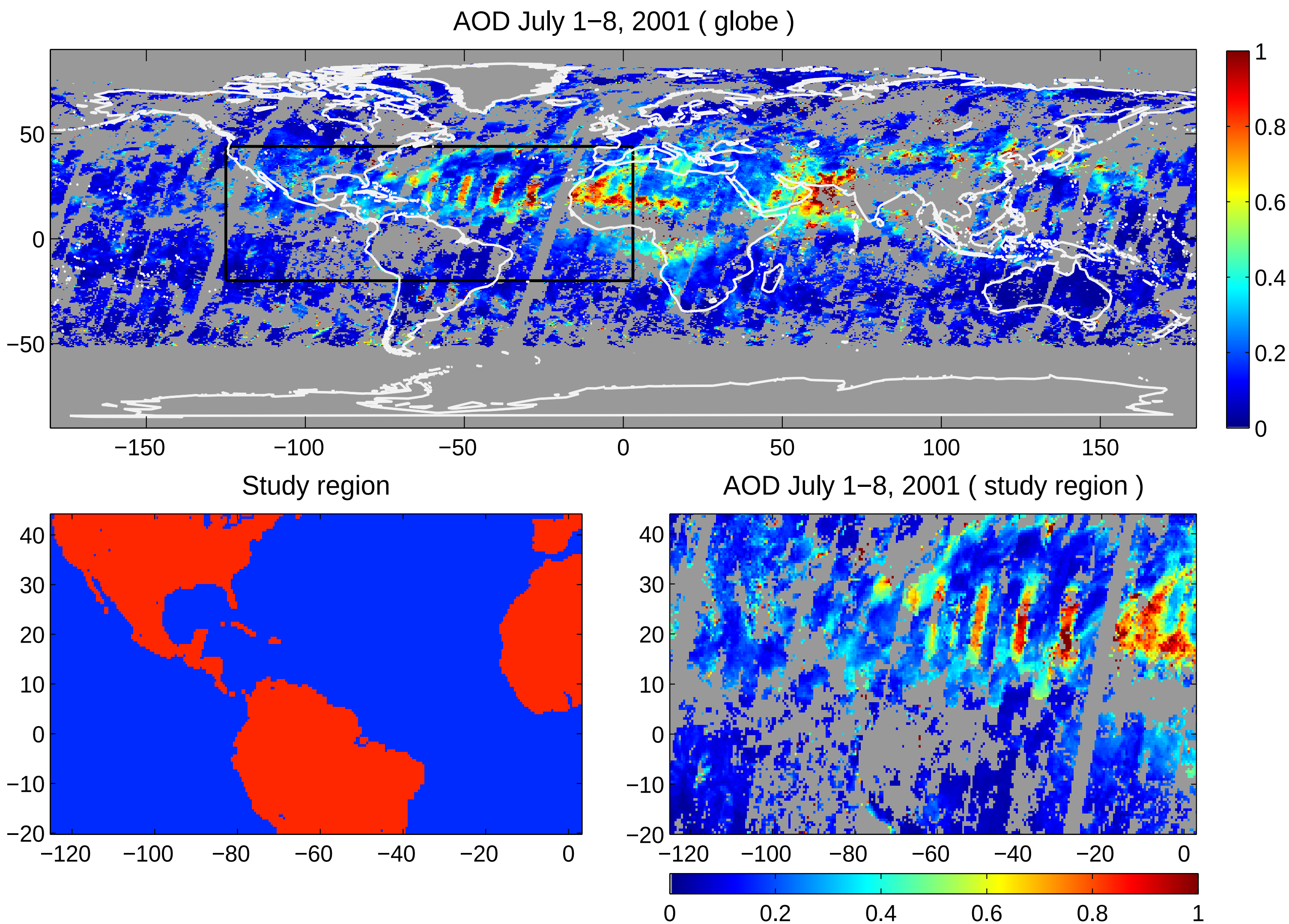
Bias and MSPE of Retrieval: Nonlinear Forward Model, ctd

$$E(\widehat{\mathbf{x}} - \mathbf{x}) = E(\widehat{\mathbf{x}}) - \mathbf{x}_\alpha \simeq \widetilde{\mathbf{bias}}(\mathbf{x}_\alpha)$$

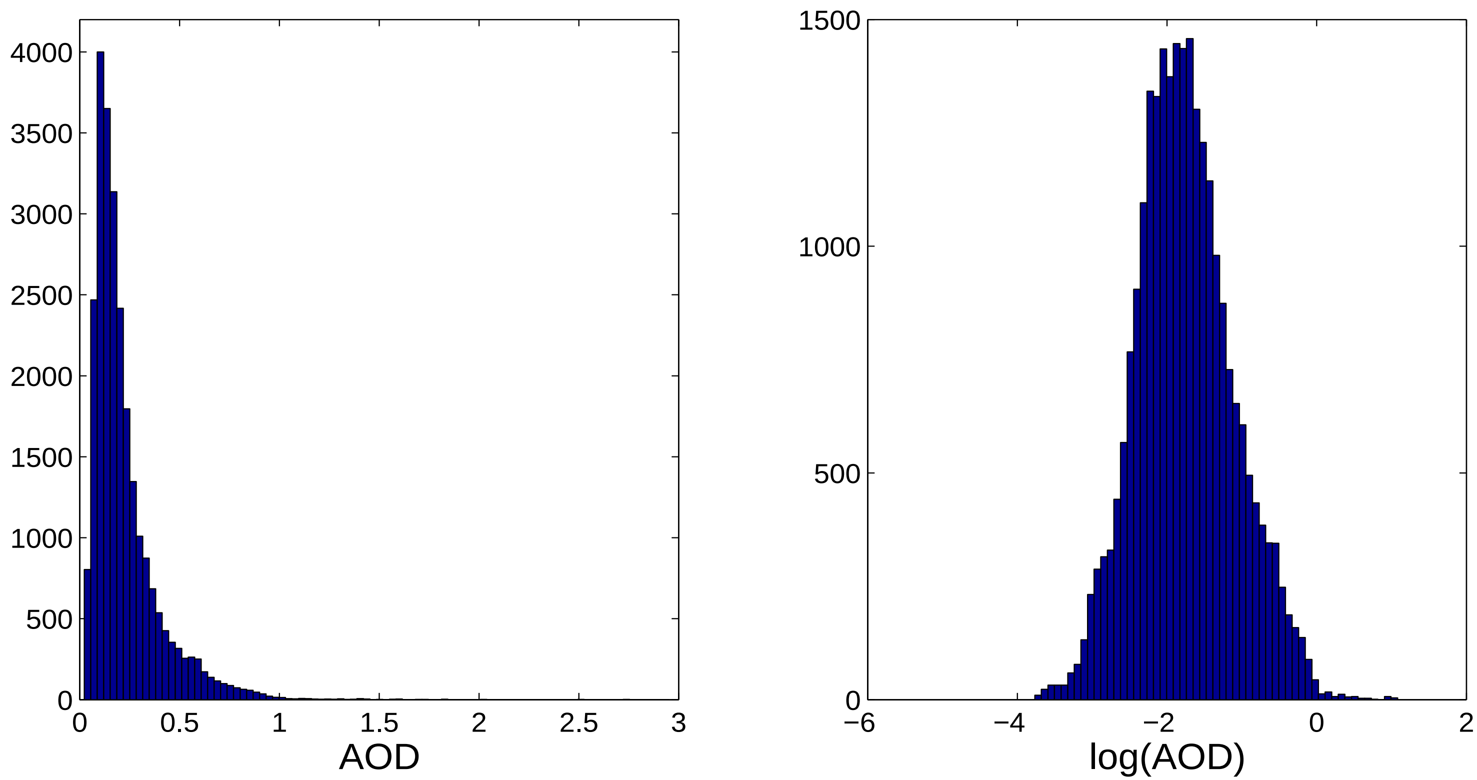
$$\equiv (1/2) \begin{pmatrix} \left(\text{vec}\left(\frac{\partial \mathbf{A}(\mathbf{x}_\alpha)_{\text{1st-row}}}{\partial \mathbf{x}_\alpha}\right)\right)' \\ \left(\text{vec}\left(\frac{\partial \mathbf{A}(\mathbf{x}_\alpha)_{\text{2nd-row}}}{\partial \mathbf{x}_\alpha}\right)\right)' \\ \vdots \\ \left(\text{vec}\left(\frac{\partial \mathbf{A}(\mathbf{x}_\alpha)_{n_\alpha\text{th-row}}}{\partial \mathbf{x}_\alpha}\right)\right)' \end{pmatrix} \cdot \text{vec}\left(\widetilde{\text{MSPE}}(\mathbf{x}_\alpha) + 2\mathbf{S}_\alpha \mathbf{A}(\mathbf{x}_\alpha)'\right)$$

$$- (1/2) \begin{pmatrix} \left(\text{vec}\left(\frac{\partial \mathbf{G}(\mathbf{x}_\alpha)_{\text{1st-row}}}{\partial \mathbf{x}_\alpha}\right)\right)' \\ \left(\text{vec}\left(\frac{\partial \mathbf{G}(\mathbf{x}_\alpha)_{\text{2nd-row}}}{\partial \mathbf{x}_\alpha}\right)\right)' \\ \vdots \\ \left(\text{vec}\left(\frac{\partial \mathbf{G}(\mathbf{x}_\alpha)_{n_\alpha\text{th-row}}}{\partial \mathbf{x}_\alpha}\right)\right)' \end{pmatrix} \cdot \text{vec}\left(\mathbf{K}(\mathbf{x}_\alpha) \cdot \widetilde{\text{MSPE}}(\mathbf{x}_\alpha) - 2\mathbf{S}_\varepsilon \mathbf{G}(\mathbf{x}_\alpha)'\right)$$

Study Region: 128×256 grid ($0.5^\circ \times 0.5^\circ$ pixels)



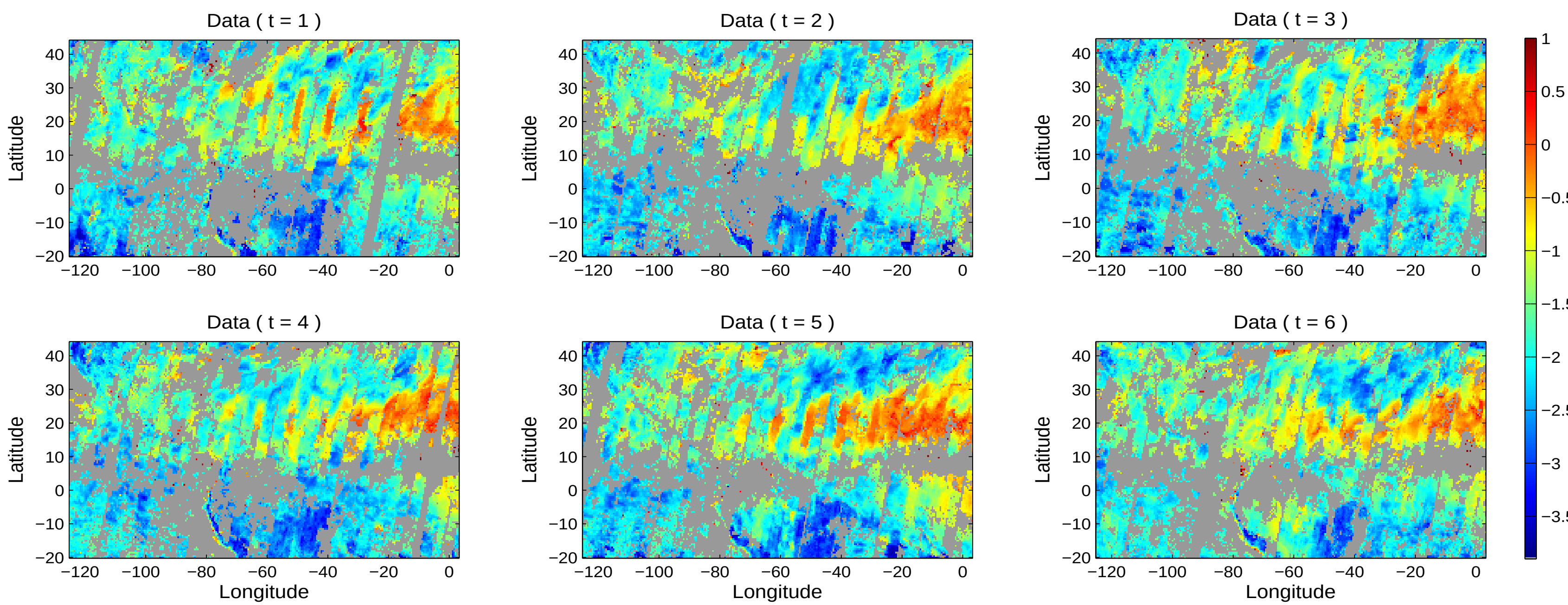
AOD Histograms (July 1-8, 2001)



Conclusion: Analyze log**(AOD)**

Spatio-Temporal AOD Data

Spatial data arrive sequentially: July 1-8, 2001 ($t = 1$), ..., August 2-9, 2001 ($t = 6$). At each time point, we have a spatial "snapshot." Large AOD means that the aerosol fraction is high; **spatio-temporal data** on log(AOD) is shown



Spatio-Temporal Statistical Model: Data Model

- **Data model**:

$$\begin{aligned} Z(\mathbf{s};t) &= Y(\mathbf{s};t) + \varepsilon(\mathbf{s};t) \\ &= \mu(\mathbf{s};t) + \nu(\mathbf{s};t) + \varepsilon(\mathbf{s};t) \end{aligned}$$

  - **Measurement-error process** $\{\varepsilon(\mathbf{s};t)\}$ is independent of $\{Y(\mathbf{s};t)\}$. Assume $\{\varepsilon(\mathbf{s};t)\}$ is Gaussian, with mean zero and $\text{var}(\varepsilon(\mathbf{s};t)) = \sigma_\varepsilon^2 v(\mathbf{s};t) > 0$; $E(\varepsilon(\mathbf{s};t)\varepsilon(\mathbf{q};u)) = 0$, unless $\mathbf{s} = \mathbf{q}$ and $t = u$

- **Noisy** and **incomplete** data at time u, for $u = 1, \ldots, t$:

$$\mathbf{Z}(u) \equiv (Z(\mathbf{s}_{1,u};u), \ldots, Z(\mathbf{s}_{n_u,u};u))'\,,$$

$$\text{var}(\mathbf{Z}(u)) \equiv \Sigma_u = S_u K_u S_u' + D_u\,,$$

  where $D_u \equiv \sigma_\xi^2 I_{n_u} + \sigma_\varepsilon^2 V_u$ and $V_u \equiv diag(v(\mathbf{s}_{1,u};u), \ldots, v(\mathbf{s}_{n_u,u};u))$

- Define $n_{+t} \equiv \sum_{u=1}^t n_u$; then $\text{var}(\mathbf{Z}(1)', \ldots, \mathbf{Z}(t)')$ is $n_{+t} \times n_{+t}$. Statistical **gap-filling and smoothing** uses dimension reduction

Aside: Spatial Random Effects (SRE) Model

- $\mathbf{S}(\cdot) \equiv (S_1(\cdot), \ldots, S_r(\cdot))'$ are spatial basis functions (known)
- $\boldsymbol{\eta} \equiv (\eta_1, \ldots, \eta_r)'$ are coefficients (random)
- $\xi(\cdot)$ is a fine-scale spatial process (random; e.g., white noise)

Then

$$\nu(\mathbf{s}) \equiv \mathbf{S}(\mathbf{s})'\boldsymbol{\eta} + \xi(\mathbf{s})\,; \qquad \mathbf{s} \in D$$

is an **SRE model** (Cressie and Johannesson, 2006, 2008)

Aside: Spatio-Temporal Random Effects (STRE) Model

- One approach is to fit each spatial “snapshot” with an SRE model; this does not take the temporal dependence into account
- **Alternatively, let the SRE model evolve dynamically.** Recall the random coefficients $\boldsymbol{\eta}$. Subscript them with time t and write:

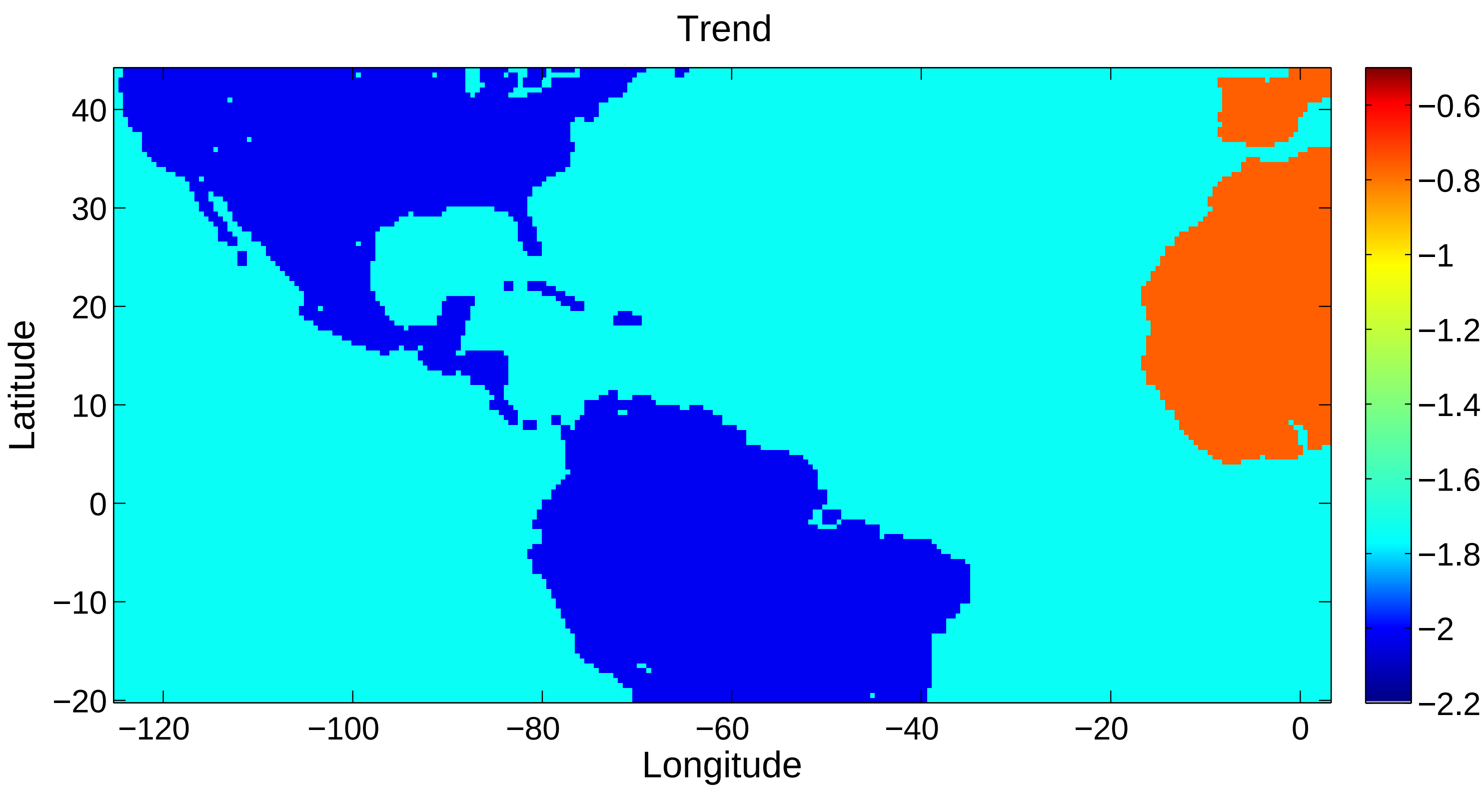
$$\boldsymbol{\eta}_{t+1} = H_{t+1}\boldsymbol{\eta}_t + \boldsymbol{\zeta}_{t+1}\,; \qquad t = 1, 2, \ldots\,.$$

See Wikle and Cressie (1999) and Cressie, Shi, and Kang (2010) for more details. This results in the **Spatio-Temporal Random Effects (STRE)** model,

$$\nu(\mathbf{s};t) \equiv \mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t + \xi(\mathbf{s};t)\,; \quad \mathbf{s} \in D_t\,,\ t = 1, 2, \ldots$$

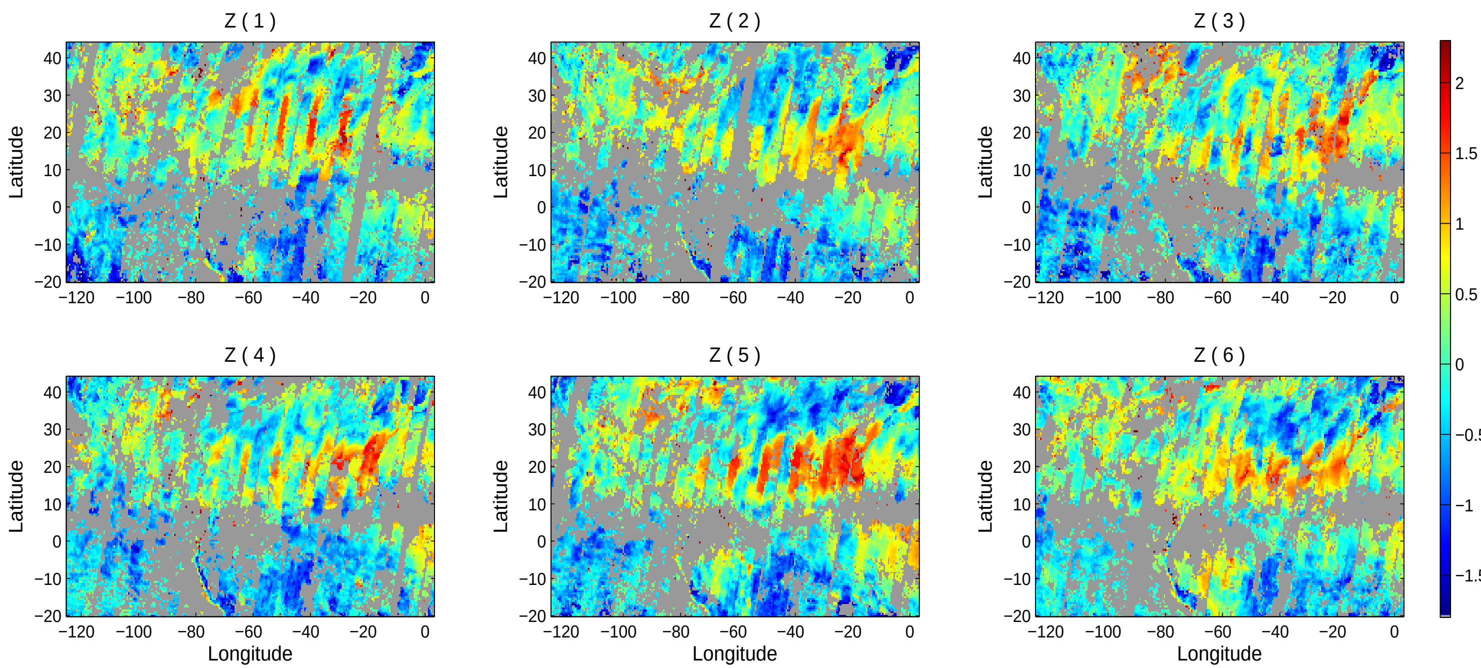
Trend

AOD over the land is different from AOD over the sea



Detrended Data

$\mathbf{Z}(1), \mathbf{Z}(2), \ldots, \mathbf{Z}(6)$, after OLS detrending; each is an approx. $20,000$-**dimensional** vector



Think of $\mathbf{Z}(t)$ as a noisy, incomplete version of a mean-zero hidden process, $\{Y(\mathbf{s};t) : \mathbf{s} \in D\}$

Spatio-Temporal Statistical Model: Data Model

- **Data model**:

$$\begin{aligned} Z(\mathbf{s};t) &= Y(\mathbf{s};t) + \varepsilon(\mathbf{s};t) \\ &= \mu(\mathbf{s};t) + \nu(\mathbf{s};t) + \varepsilon(\mathbf{s};t) \end{aligned}$$

  - **Measurement-error process** $\{\varepsilon(\mathbf{s};t)\}$ is independent of $\{Y(\mathbf{s};t)\}$. Assume $\{\varepsilon(\mathbf{s};t)\}$ is Gaussian, with mean zero and $\text{var}(\varepsilon(\mathbf{s};t)) = \sigma_\varepsilon^2 v(\mathbf{s};t) > 0$; $E(\varepsilon(\mathbf{s};t)\varepsilon(\mathbf{q};u)) = 0$, unless $\mathbf{s} = \mathbf{q}$ and $t = u$

- **Noisy** and **incomplete** data at time u, for $u = 1, \ldots, t$:

$$\mathbf{Z}(u) \equiv (Z(\mathbf{s}_{1,u};u), \ldots, Z(\mathbf{s}_{n_u,u};u))' \,,$$

$$\text{var}(\mathbf{Z}(u)) \equiv \Sigma_u = S_u K_u S_u' + D_u \,,$$

  where $D_u \equiv \sigma_\xi^2 I_{n_u} + \sigma_\varepsilon^2 V_u$ and $V_u \equiv diag(v(\mathbf{s}_{1,u};u), \ldots, v(\mathbf{s}_{n_u,u};u))$

- Define $n_{+t} \equiv \sum_{u=1}^{t} n_u$; then $\text{var}(\mathbf{Z}(1)', \ldots, \mathbf{Z}(t)')$ is $n_{+t} \times n_{+t}$. Statistical **gap-filling and smoothing** uses dimension reduction

(Hidden) Process Model

- $Y(\mathbf{s};t) = \mu(\mathbf{s};t) + \nu(\mathbf{s};t)$; $\mathbf{s} \in D$, $t = 1, 2, \ldots$ (mean μ is deterministic, $= 0$ for detrended process)
- Spatio-temporal process $\{\nu(\mathbf{s};t) : \mathbf{s} \in D \subset \mathbb{R}^d; t = 1, 2, \ldots\}$, mean zero
- **Spatio-temporal random effects (STRE)** model: At any fixed time t,

$$\nu(\mathbf{s};t) = \mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t + \xi(\mathbf{s};t)\,;$$

cf. Wikle and Cressie (1999) and Cressie, Shi, and Kang (2010)

  - **Fix** $r \ll n_u$, the number of multi-resolution basis functions
  - Consider a set of spatial basis functions (not necessarily orthogonal),

$$\mathbf{S}_t(\cdot) \equiv (S_{1,t}(\cdot), \ldots, S_{r,t}(\cdot))'$$

  - $\boldsymbol{\eta}_t \equiv (\eta_{1,t}, \ldots, \eta_{r,t})'$ is a zero-mean random vector; $\mathrm{var}(\boldsymbol{\eta}_t) = K_t$
  - ξ is fine-scale variability modeled as white noise

Process Model: Dynamics on STRE Coefficients

Introduce **temporal dynamics** through a vector autoregressive (VAR) process of order 1 for $\{\boldsymbol{\eta}_t\}$:

$$\boldsymbol{\eta}_{t+1} = H_{t+1}\, \boldsymbol{\eta}_t + \boldsymbol{\zeta}_{t+1}\, ; \quad t = 1, 2, \ldots$$

- H_{t+1}: first-order **autoregressive** (or propagator) matrix
- $\boldsymbol{\zeta}_{t+1}$: r-dimensional **innovation** vector independent of $\boldsymbol{\eta}_t$
- $\{\boldsymbol{\zeta}_{t+1}\}$: Gaussian process, temporally **independent**, with

$$E(\boldsymbol{\zeta}_{t+1}) = \mathbf{0}\,,$$

and

$$\text{var}(\boldsymbol{\zeta}_{t+1}) = U_{t+1}$$

Fixed Rank Filtering (FRF)

Goal: **Predict $Y(\mathrm{s}_0;t)$ given noisy and incomplete data up to and including time t,** that is, given $\mathbf{Z}(1),\ldots,\mathbf{Z}(t)$.

- **Kalman filter** (e.g., Kalman, 1960; Shumway and Stoffer, 2006, Sect. 6.2): Express the optimal predictor of $\boldsymbol{\eta}_t$ given $\mathbf{Z}(1),\ldots,\mathbf{Z}(t)$, recursively:

$$\begin{aligned}\widehat{\boldsymbol{\eta}}_{t|t} &\equiv E(\boldsymbol{\eta}_t \,|\, \mathbf{Z}(1),\ldots,\mathbf{Z}(t)) \\ &= \widehat{\boldsymbol{\eta}}_{t|t-1} + G_t \left\{ \mathbf{Z}(t) - \boldsymbol{\mu}_t - S_t \widehat{\boldsymbol{\eta}}_{t|t-1} \right\} ,\end{aligned}$$

  with $r \times r$ mean-squared-prediction-error (MSPE) matrix,

$$P_{t|t} \equiv E\left[(\widehat{\boldsymbol{\eta}}_{t|t} - \boldsymbol{\eta}_t)(\widehat{\boldsymbol{\eta}}_{t|t} - \boldsymbol{\eta}_t)' \right] = P_{t|t-1} - G_t\, S_t\, P_{t|t-1} \, ,$$

  - $\widehat{\boldsymbol{\eta}}_{t|t-1}$ and $P_{t|t-1}$: one-step-ahead-forecast and its $r \times r$ MSPE matrix, resp.
  - G_t**:** $r \times r$ **Kalman gain matrix**

FRF: Kalman Gain Matrix

- G_t: $r \times r$ Kalman gain matrix

$$
\begin{aligned}
G_t &= P_{t|t-1}\, S_t' \{D_t + S_t\, P_{t|t-1}\, S_t'\}^{-1} \\
&= P_{t|t-1}\, S_t' \left\{ D_t^{-1} - D_t^{-1} S_t \right. \\
&\quad \left. \times \left[P_{t|t-1}^{-1} + S_t' D_t^{-1} S_t \right]^{-1} S_t' D_t^{-1} \right\},
\end{aligned}
$$

where the following **Sherman-Morrison-Woodbury identity** is applied:

$$
(I + PKP')^{-1} = I - P(K^{-1} + P'P)^{-1} P' ,
$$

for any $n \times r$ matrix P.

- Only involves **inversion of fixed rank $r \times r$ matrices** and **of $n_t \times n_t$ diagonal matrices**
- Computing is really fast and the required storage space is small

Filtering of $Y(\cdot;t)$ (Cressie, Shi, and Kang, 2010)

Optimal prediction of $\nu(\mathbf{s}_0;t)$ given the data $\mathbf{Z}(1),\ldots,\mathbf{Z}(t)$ yields:

$$\mathbf{S}_t(\mathbf{s}_0)'\widehat{\boldsymbol{\eta}}_{t|t} + \widehat{\xi}_{t|t}(\mathbf{s}_0)\,,$$

where

$$\begin{aligned}
\widehat{\xi}_{t|t}(\mathbf{s}_0) &= E(\xi(\mathbf{s}_0;t)|\mathbf{Z}(1),\ldots,\mathbf{Z}(t)) \\
&= \mathbf{c}_t(\mathbf{s}_0)'(D_t + S_t P_{t|t-1} S_t')^{-1}(\mathbf{Z}(t) - \boldsymbol{\mu}(t) - S_t\widehat{\boldsymbol{\eta}}_{t|t-1}) \\
\mathbf{c}_t(\mathbf{s}_0) &= \sigma_\xi^2(I(\mathbf{s}_0 = \mathbf{s}_{1,t}),\ldots,I(\mathbf{s}_0 = \mathbf{s}_{n_t,t}))'\,.
\end{aligned}$$

Then the Fixed Rank Filtering (FRF) equations are:

$$\begin{aligned}
\widehat{Y}(\mathbf{s}_0;t)^{FRF} &= \mu(\mathbf{s}_0;t) + \mathbf{S}_t(\mathbf{s}_0)'\widehat{\boldsymbol{\eta}}_{t|t} + \widehat{\xi}_{t|t}(\mathbf{s}_0) \\
\sigma(\mathbf{s}_0;t)^{FRF} &= \{E(Y(\mathbf{s}_0;t) - \widehat{Y}(\mathbf{s}_0;t)^{FRF})^2\}^{1/2}
\end{aligned}$$

FRF: Discussion

- Efficient Computation
  - Computational complexity is reduced from $O(n_{+t}^3)$ to $O(n_{+t})$
  - **Small storage space** is required due to the fixed rank r: store an $r \times r$ matrix $P_{t|t}$ and an r-dimensional vector $\boldsymbol{\eta}_{t|t}$
- Given data $\mathbf{Z}(1), \ldots, \mathbf{Z}(t)$, **fixed rank smoothing** and **fixed rank forecasting** can be similarly obtained
  - **Smoothing**: prediction of $Y(\mathbf{s}_0; u)$, where $u \in \{1, \ldots, t-1\}$
  - **Forecasting**: prediction of $Y(\mathbf{s}_0; u)$, where $u \in \{t+1, t+2, \ldots\}$
- This has been applied to **AOD** (see below) and the **CO_2 product** from the Atmospheric InfraRed Sounder (AIRS) instrument (http://www.stat.osu.edu/~sses/collab_co2.html)

Spatio-Temporal Statistical Mapping of Satellite Data

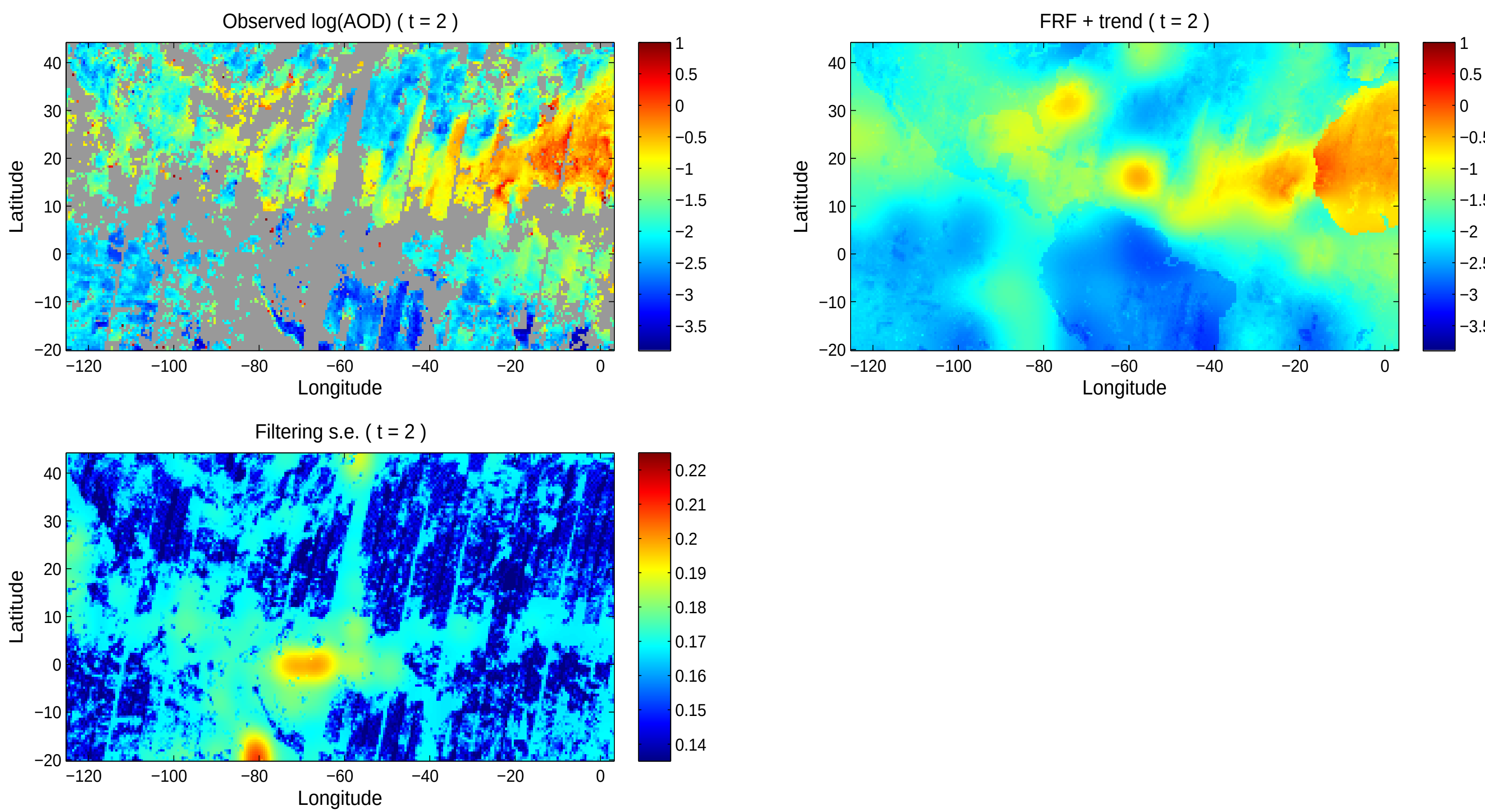
- Data $\mathbf{Z}(1), \ldots, \mathbf{Z}(6)$ are large, noisy, and spatially incomplete
- Sample sizes:

$$n_1 = 20970, \ n_2 = 19398, \ n_3 = 20819\,,$$

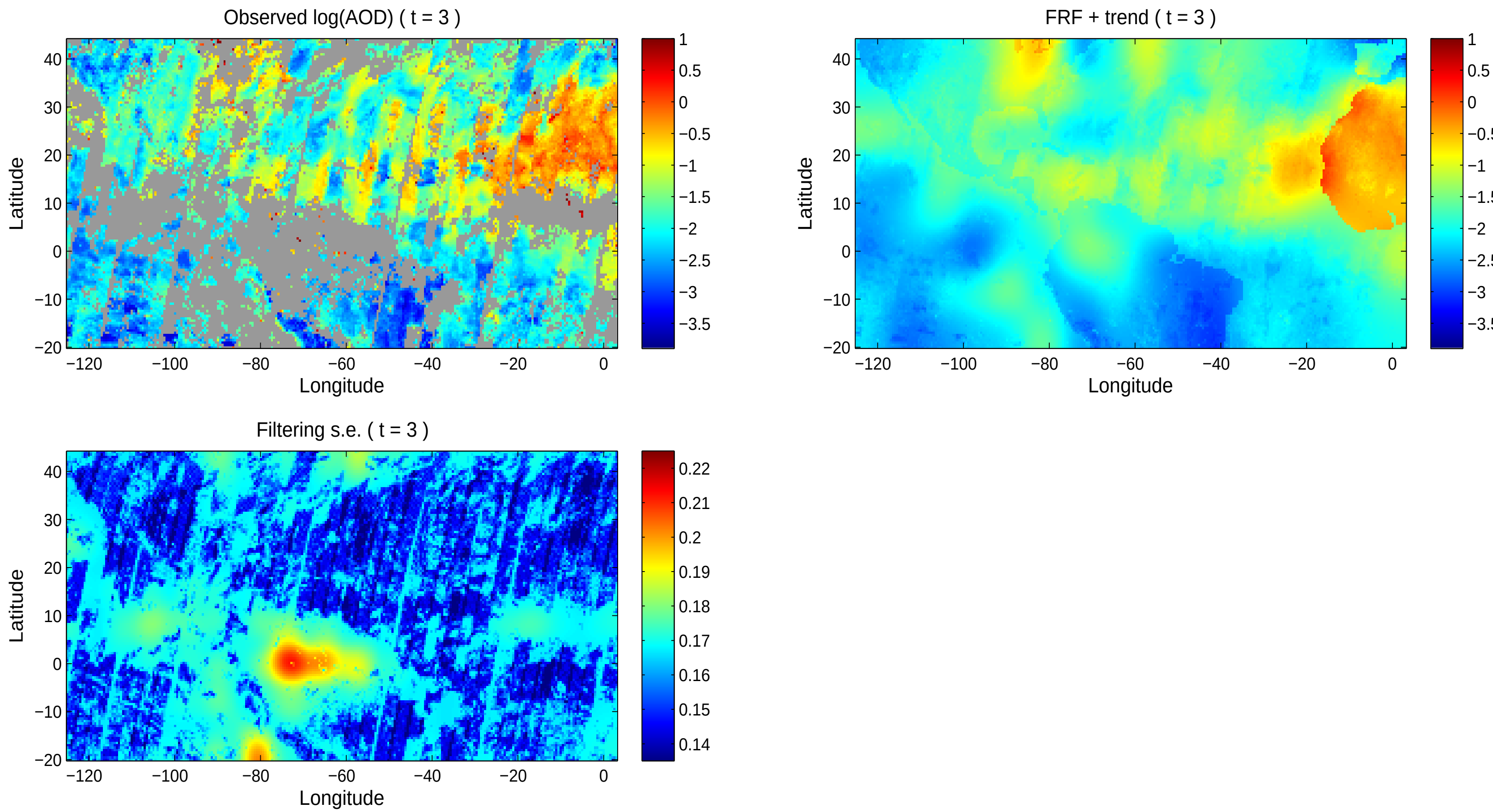
$$n_4 = 20167, \ n_5 = 21759, \ n_6 = 20625\,.$$

- Use W-wavelet functions for basis functions
- Fix $r = 94$
- See Kang et al. (2010) and Wikle and Cressie (2011), Sec. 9.2, for a more complete analysis, including methodology for **estimating the model's parameters**

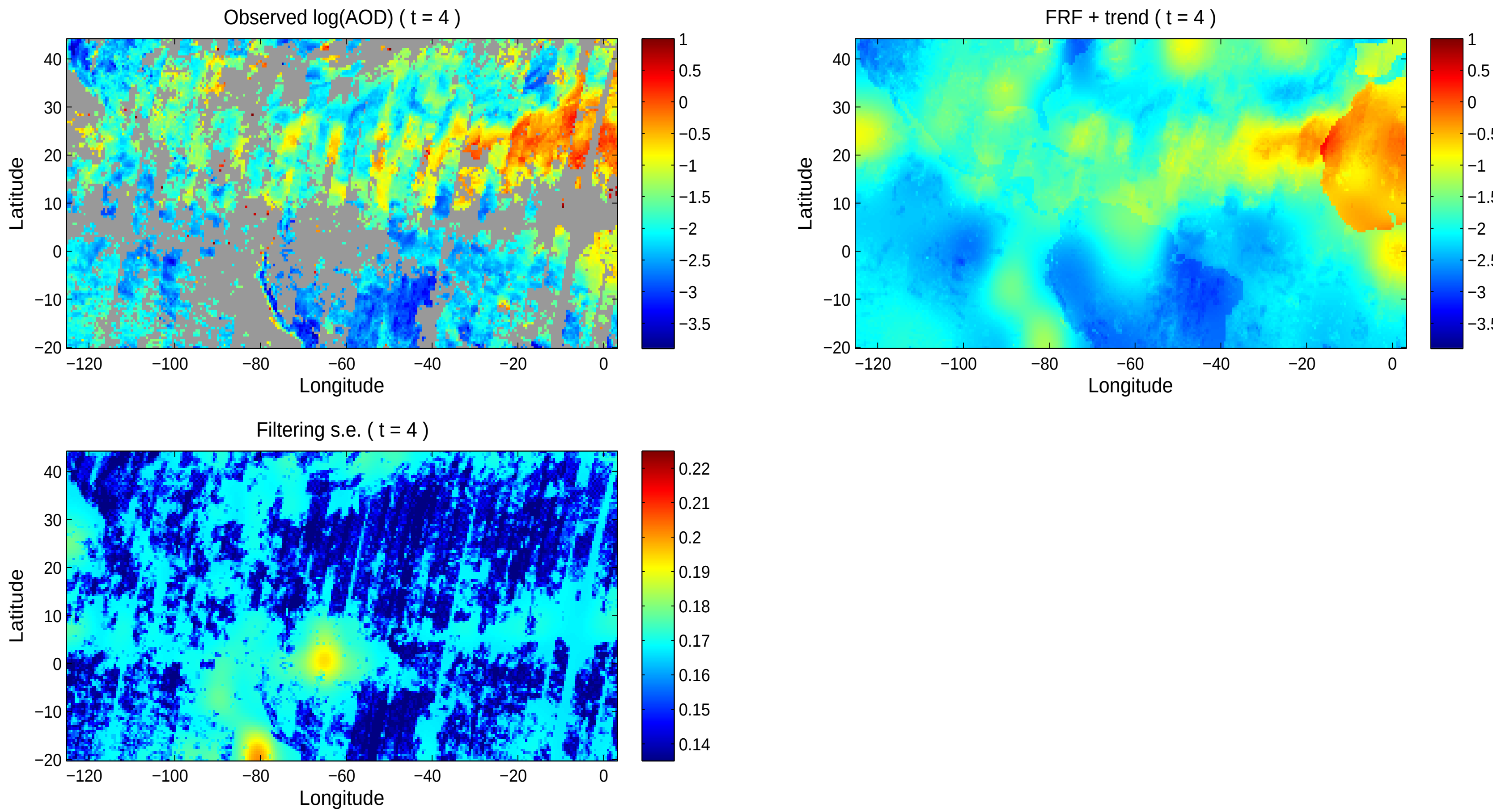
Final Result ($t = 2$)



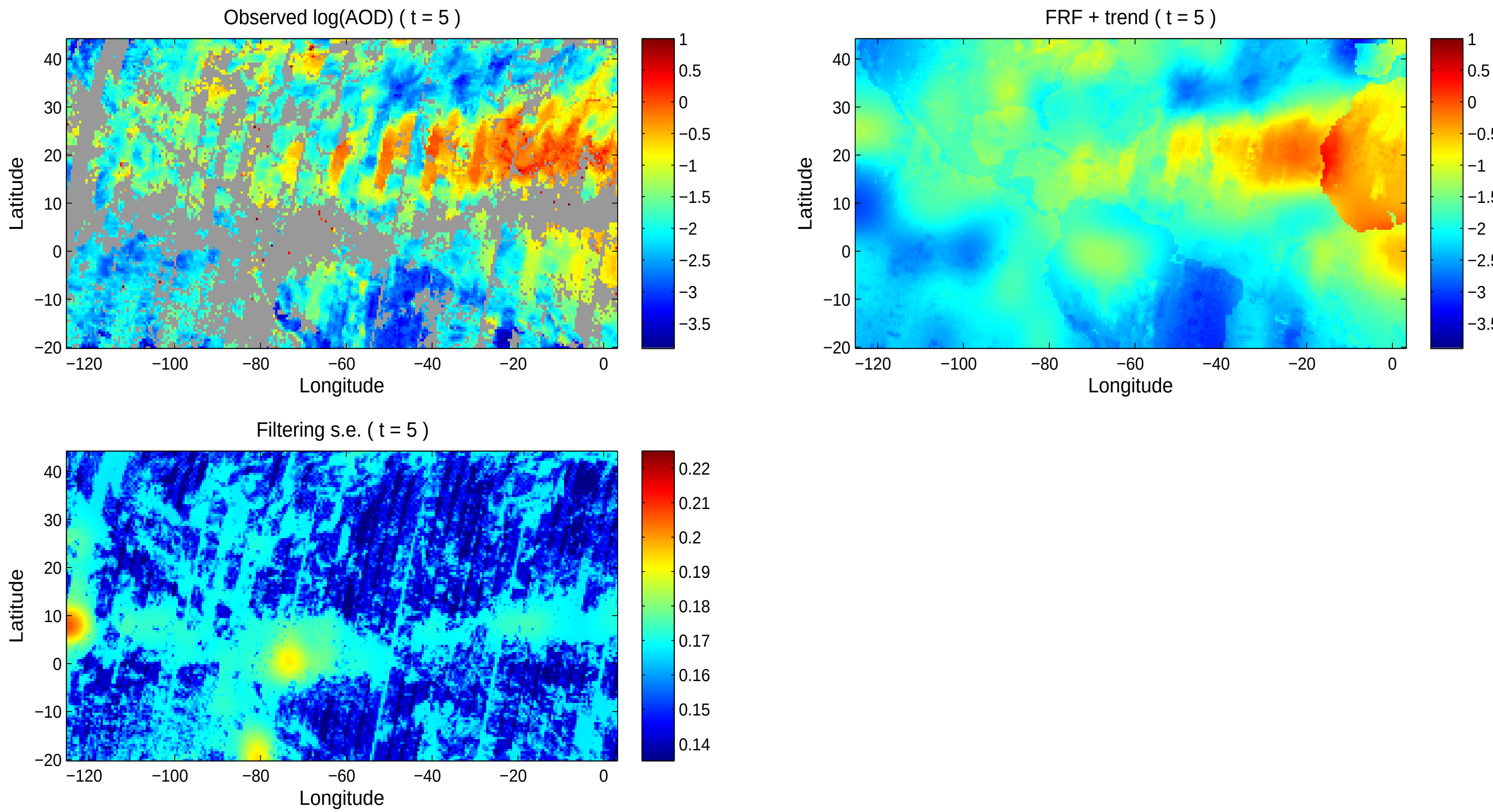
Final Result ($t = 3$)



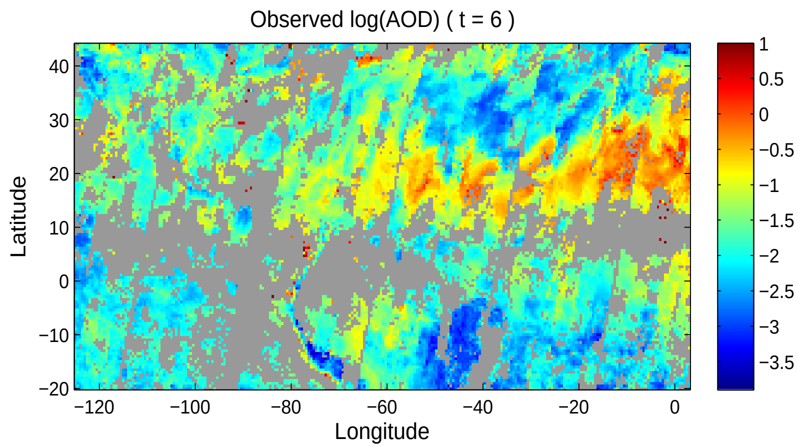
Final Result ($t = 4$)

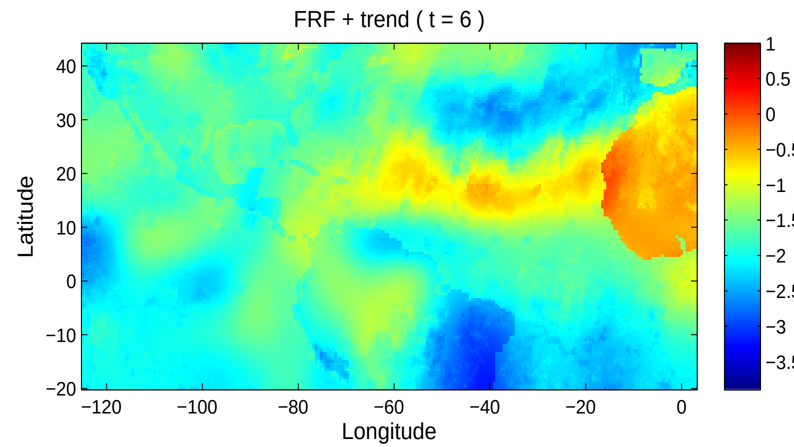


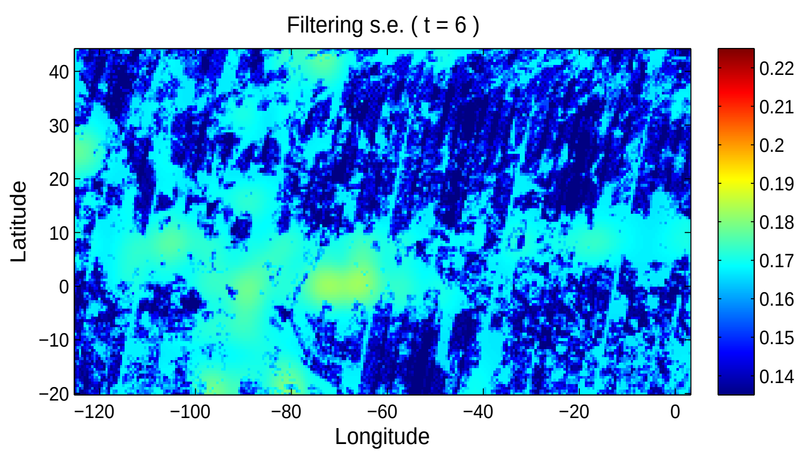
Final Result ($t = 5$)



Final Result ($t = 6$)







Computation

- Computations are carried out in Matlab on a Windows laptop with a dual core 2.0 GHz processor and 3GB memory
- **Timing**:
  - **64.4 seconds** to fit the **parameters** (including binning)
  - 77.3 seconds to compute the predictions and standard errors from time $t = 2$ through $t = 5$
  - Only **19.7 seconds** to **update the filter** at, for example, time $t = 5$, given the filtering results at $t = 4$ and the data at $t = 5$
- **Storage**: 94×94 matrix $P_{t|t}$ and a 94-dimensional vector $\boldsymbol{\eta}_{t|t}$

References

Cressie, N. and Johannesson, G. (2006). In *Proceedings of the Australian Academy of Science Elizabeth and Frederick White Conference*, pp. 1-11. Australian Academy of Science, Canberra.

Cressie, N. and Johannesson, G. (2008). *Journal of the Royal Statistical Society, Series B*, **70**, 209-226.

Cressie, N., Shi, T., and Kang, E.L. (2010) *Journal of Computational and Graphical Statistics*, **19**, 724-745.

Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ.

Kalman, R.E. (1960). *Transactions of the ASME - Journal of Basic Engineering Statistics*, **14**, 5-25.

Kang, E.L., Cressie, N., and Shi, T. (2010). *Canadian Journal of Statistics*, **38**, 271-289.

Rodgers, C.D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific, Singapore.

Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications, with R Examples*, 2nd edn. Springer, New York, NY.

Wikle, C.K. and Cressie, N. (1999). *Biometrika*, **86**, 815-829.