ISR Data Analysis & Fitting 2

John Swoboda
Ryan Volz

July 25, 2018

MIT
HAYSTACK
OBSERVATORY

Statistics

Least Squares & Maximum Likelihood

ISR Fitting

Resolution

Advanced Processing

Estimating functions

MIT HAYSTACK OBSERVATORY

Outline

Statistics

Least Squares & Maximum Likelihood

ISR Fitting

Resolution

Advanced Processing

Estimating functions

MIT HAYSTACK OBSERVATORY

Random Variables

- The probability of an event is a number between 0 and 1 that is to represent the outcomes of the event divided by the number of experiments [PP02]
- Random variables: A number that is assigned to the outcome of every experiment
- R.V. can be described using distributions
- CDF $F_x(x) = P\{X \leq x\}$
- PDF $f_x(x) \triangleq \frac{dF_x(x)}{dx} = \lim_{\Delta x \to 0} \frac{P\{x \leq X \leq x + \Delta x\}}{\Delta x}$

MIT HAYSTACK OBSERVATORY

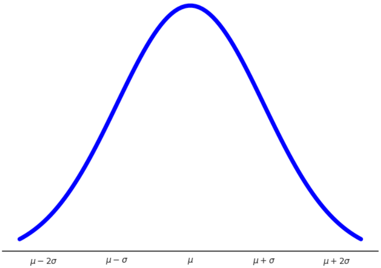
Normally Distributed R.V.

- Normally distributed R.V. have the following PDF:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- μ is mean, σ^2 is variance

- CLT: Sum of independent R.V.s converges to normally distributed R.V.



MIT HAYSTACK OBSERVATORY

Multiple R.V.

- R.V. can include information on other R.V. which can be expressed through joint distributions

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left\{-\frac{1}{2}\mathbf{x}\mathbf{C}^{-1}\mathbf{x}^T\right\}$$

- If independent $f(x_1, x_2, ..., x_n) = f(x_1)f(x_2)...f(x_n)$
- Instead of finding whole PDF often use correlation instead $E\{X_1, X_2\}$

MIT HAYSTACK OBSERVATORY

Estimation

- Often need to estimate statistics on R.V.
- Determine the mean of n of sequence $X_1, X_2 ..., X_n$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Estimation theory can give bounds on uncertainty
- Correlation estimation

$$\widehat{C_{xy}} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i$$

MIT HAYSTACK OBSERVATORY

Outline

Statistics

Least Squares & Maximum Likelihood

ISR Fitting

Resolution

Advanced Processing

Estimating functions

MIT HAYSTACK OBSERVATORY

Why least squares?

Least squares problem

$$\underset{\theta}{\text{minimize}} \qquad \frac{1}{2}\|y - h(\theta)\|_2^2$$

Why least squares and not another error metric?

Why least squares?

Least squares problem

$$\underset{\theta}{\text{minimize}} \qquad \frac{1}{2}\|y - h(\theta)\|_2^2$$

Why least squares and not another error metric?

- ~~Everybody uses it~~

MIT HAYSTACK OBSERVATORY

Why least squares?

Least squares problem

$$\underset{\theta}{\text{minimize}} \qquad \frac{1}{2}\|y - h(\theta)\|_2^2$$

Why least squares and not another error metric?

- ~~Everybody uses it~~
- It's easy: closed-form linear solution for linear least-squares

MIT HAYSTACK OBSERVATORY

Why least squares?

Least squares problem

$$\underset{\theta}{\text{minimize}} \qquad \frac{1}{2}\|y - h(\theta)\|_2^2$$

Why least squares and not another error metric?

- ~~Everybody uses it~~
- It's easy: closed-form linear solution for linear least-squares
- It punishes large errors more than small errors

MIT
HAYSTACK
OBSERVATORY

Why least squares?

Least squares problem

$$\underset{\theta}{\text{minimize}} \qquad \frac{1}{2}\|y - h(\theta)\|_2^2$$

Why least squares and not another error metric?

- ~~Everybody uses it~~
- It's easy: closed-form linear solution for linear least-squares
- It punishes large errors more than small errors
- It gives the maximum likelihood solution when errors follow the Normal distribution

MIT HAYSTACK OBSERVATORY

Likelihood function

With $Y \sim f(y \mid \theta)$, the likelihood function is defined as:

$$\mathcal{L}(\theta) \equiv f(y \mid \theta)$$

for parameters θ and a realization y.

Measurements with zero-mean Gaussian noise

$$Y = h(\theta) + N \quad \text{with} \quad N \sim \mathcal{N}(0, \Sigma) \quad \Longrightarrow \quad Y \sim \mathcal{N}(h(\theta), \Sigma)$$

Likelihood:

$$\mathcal{L}(\theta) = f(y \mid \theta) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}(y-h(\theta))^\top \Sigma^{-1}(y-h(\theta))}$$

Log-likelihood with $\Sigma = \sigma \mathrm{I}$:

$$l(\theta) = -\frac{k}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|y - h(\theta)\|_2^2$$

$$\underset{\theta}{\text{maximize}}\ \mathcal{L}(\theta) \iff \underset{\theta}{\text{minimize}}\ \frac{1}{2}\|y - h(\theta)\|_2^2$$

Likelihood function

With $Y \sim f(y \mid \theta)$, the likelihood function is defined as:

$$\mathcal{L}(\theta) \equiv f(y \mid \theta)$$

for parameters θ and a realization y.

Measurements with zero-mean Gaussian noise

$$Y = h(\theta) + N \quad \text{with} \quad N \sim \mathcal{N}(0, \Sigma) \quad \implies \quad Y \sim \mathcal{N}(h(\theta), \Sigma)$$

Likelihood:

$$\mathcal{L}(\theta) = f(y \mid \theta) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}(y-h(\theta))^\top \Sigma^{-1}(y-h(\theta))}$$

Log-likelihood with $\Sigma = \sigma \mathrm{I}$:

$$l(\theta) = -\frac{k}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|y - h(\theta)\|_2^2$$

$$\underset{\theta}{\text{maximize}} \ \mathcal{L}(\theta) \iff \underset{\theta}{\text{minimize}} \ \frac{1}{2}\|y - h(\theta)\|_2^2$$

Maximum likelihood

A useful framework

- Turns parameter estimation problem into optimization problem
- Many R.V.s are Gaussian (central limit theorem)
  - Least squares is nice!
- Estimates come with error bars governed by the curvature of the log-likelihood function (see Fisher information)

A Bayesian perspective

- Maximum a posteriori (MAP) estimate maximizes

$$P(\theta \mid y) = \frac{f(y \mid \theta) P(\theta)}{P(y)}$$

- With uniform prior $P(\theta)$, MAP = ML
- Other priors yield regularized ML problems
  - e.g. Laplace prior yields l_1-regularization

Under-determined systems of equations

Not enough measurements to constrain unknown values:

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad & A & \quad \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \end{bmatrix}$$

Measurement Model Unknown

- Infinite number of solutions
- Often have prior information about the true solution (e.g. sparsity) that can make the problem well-conditioned

Theory of compressed sensing

Finding the sparsest solution is hard in general.

Definition

Compressed sensing is a theory to guarantee solution of an under-determined set of equations.

Approximate guidelines for application

- Solution known to be sparse
- Measurements capture the effects of all parameters
- Minimum number of measurements on the order of the solution sparsity (number of nonzeros)

Benefit

Can solve an easy convex optimization problem instead of a hard combinatorial problem.

Equivalent convex optimization problem

Sparsest solution to noisy measurements

Find sparsest x

subject to $\|y - A(x)\|_2 < \eta$

$$\|x\|_2^2 = \sum_k |x_k|^2$$

l_1-regularized least-squares (convex)

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}\|y - A(x)\|_2^2 + \lambda\|x\|_1$$

$$\|x\|_1 = \sum_k |x_k|$$

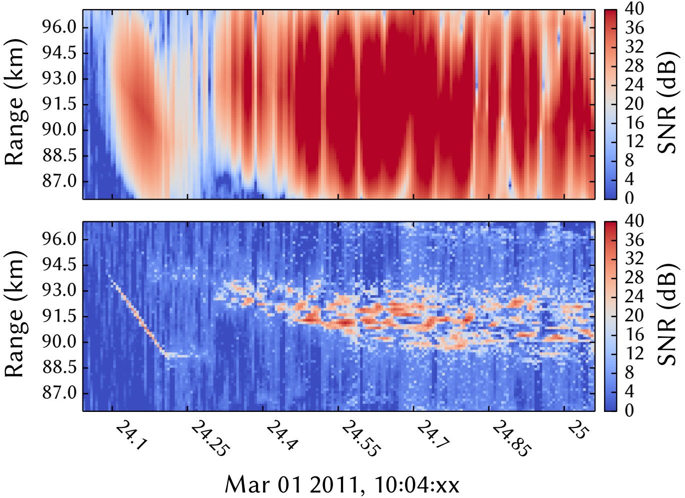
The l_1-norm promotes sparsity!

MIT HAYSTACK OBSERVATORY

Example: waveform inversion for meteor echo

Matched Filter

Waveform Inversion



MIT
HAYSTACK
OBSERVATORY

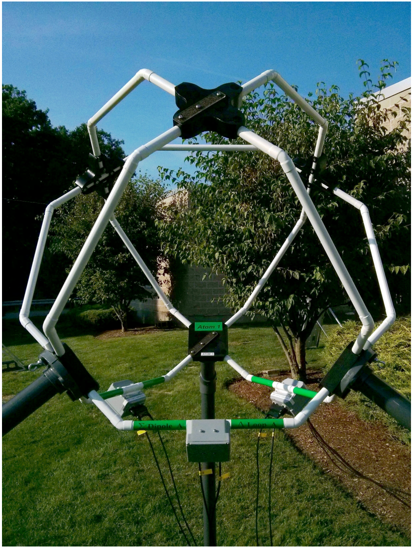
ML application: electromagnetic vector sensor

Six elements

- 3 orthogonal dipole and loop elements with common phase center

Maximum information

- Measures complete electromagnetic field at a point
- Sensitive to all directions and polarizations



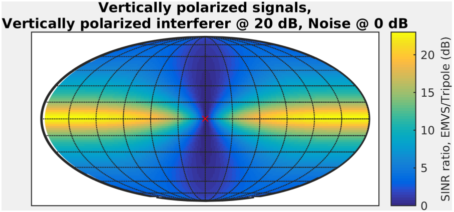
Atom antenna

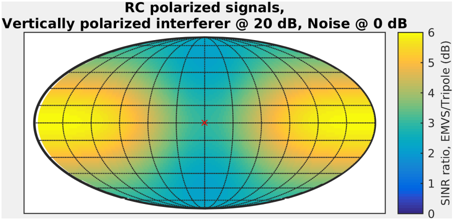
MIT HAYSTACK OBSERVATORY

Vector sensor benefits

Vector sensor benefits

- Magnitude/direction/polarization of multiple sources in single snapshot
- Frequency-independent beamforming
- Null out interfering direction/polarization





Tripole comparison (right)

- Increased sensitivity, especially in case of interference

MIT
HAYSTACK
OBSERVATORY

Measurements

Measurement equation

- Collection of independent point sources
- Sources distributed equally in solid angle on surrounding sphere
- Arbitrary polarization in horizontal/vertical basis
- $r_n = \begin{bmatrix} A_h & A_v \end{bmatrix} \begin{bmatrix} h_n \\ v_n \end{bmatrix} + w_n$

Measurement vector	Direction steering vectors	Source magnitudes/phases	Noise

Second-order statistics

- Sufficient statistic using sample covariance: $S = \frac{1}{N} \sum_{n=0}^{N-1} r_n r_n^*$

MIT HAYSTACK OBSERVATORY

Imaging problem formulation

- Assume zero-mean complex normal:

$$\begin{bmatrix} h_n \\ v_n \end{bmatrix} \sim \mathcal{CN}(0, \Sigma) \qquad w_n \sim \mathcal{CN}(0, \sigma \mathrm{I}) \qquad \forall n$$

- Entries of Σ give magnitude/polarization for source directions
- Solve covariance estimation problem:

$$\begin{array}{ll} \underset{\Sigma}{\text{minimize}} & H(\Sigma) \\ \text{subject to} & \Sigma \succeq 0 \end{array}$$

Maximum likelihood objective

$$H_{ml}(\Sigma) = \log\det(A\Sigma A^* + \sigma \mathrm{I}) + \mathrm{tr}((A\Sigma A^* + \sigma \mathrm{I})^{-1} S)$$

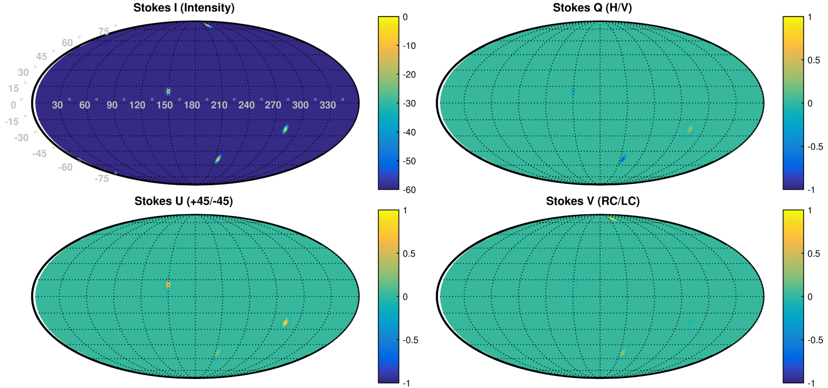
MIT HAYSTACK OBSERVATORY

Sky map using Stokes parameters

- Σ is composed of diagonal blocks (independent sources)

$$\Sigma = \begin{bmatrix} \Sigma_{hh} & \Sigma_{hv} \\ \Sigma_{hv}^* & \Sigma_{vv} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \text{diag}(I+Q) & \text{diag}(U-iV) \\ \text{diag}(U+iV) & \text{diag}(I-Q) \end{bmatrix}$$

- Can write Σ in terms of Stokes parameters I, Q, U, V



Source sky map, covariance described with Stokes parameters

Outline

MIT
HAYSTACK
OBSERVATORY

Statistics with ISR Data

- ISR data can be modeled as a Normally distributed random variable
  - The value of the parameter is an estimate of the mean of that process
  - The error bar value is the estimate standard deviation of the for random variable.
- Issues with this model
  - Correlation between different parameters
  - Bias with the measurement
  - Correlation between different measurements, e.g. close range gates
- Need to be careful in applying assumptions
  - Can impact on how you do your analysis

Fitted Data

- Least Squares

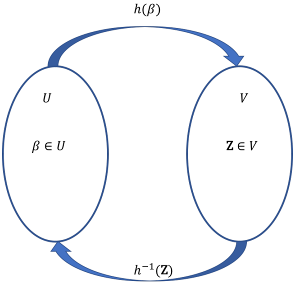
$$\hat{\beta}_{LS} = \arg\min_{\beta} [h(\beta) - \mathbf{Z}]^T \mathbf{C}^{-1} [h(\beta) - \mathbf{Z}]$$

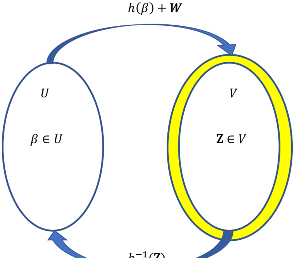
- β the plasma parameter vector
  - $[N_e, T_e, T_i, \text{etc}]^T$
- $\mathbf{Z}$ the data
  - The measured ACF or spectra
- $\mathbf{C}$ the covariance matrix of the data
- h function between parameters and data
  - e.g. [KM11]

MIT HAYSTACK OBSERVATORY

ISR Fitting

- Find parameters for a given set of data
- Find a function to move from one space to another
- Noise can increase size of space





MIT HAYSTACK OBSERVATORY

Outline

Statistics

Least Squares & Maximum Likelihood

ISR Fitting

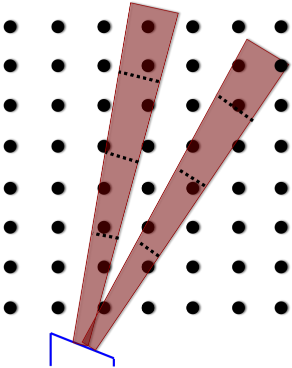
Resolution

Advanced Processing

Estimating functions

Spatial Ambiguities

- ISRs average over space as well
- The spatial averaging is dependent on the pulse type used and the beam pattern



MIT HAYSTACK OBSERVATORY

Spatial Ambiguities from Pulse Shape

- Along the beam the spatial averaging is due to the pulse pattern
- Different pulse types yield different along range ambiguities

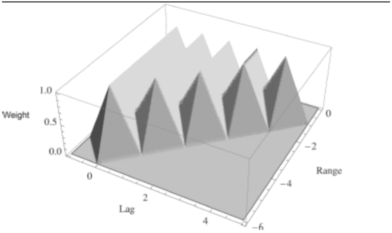
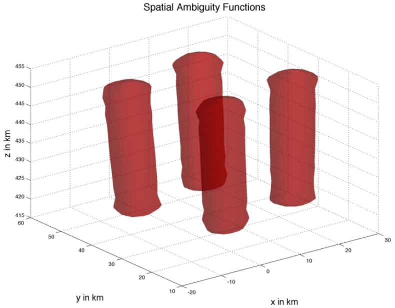


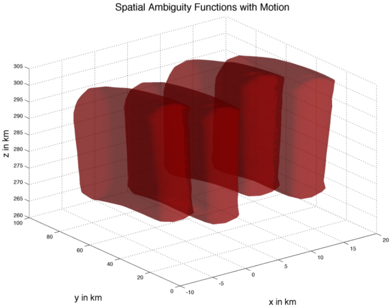
Figure: [Hys18]

MIT HAYSTACK OBSERVATORY

Spatial Ambiguities from Beam Pattern

- Motion of the plasma can increase this apparent ambiguity

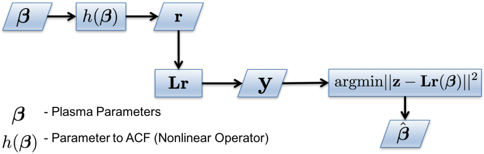




MIT HAYSTACK OBSERVATORY

Ambiguities

$$y(\tau_s, \mathbf{x}_s, t_s) = \int L(\tau_s, \mathbf{x}_s, t_s, \tau, \mathbf{x}, t) R(\tau, \mathbf{x}, t|\beta) dV dt d\tau$$



$\boldsymbol{\beta}$ - Plasma Parameters

$h(\boldsymbol{\beta})$ - Parameter to ACF (Nonlinear Operator)

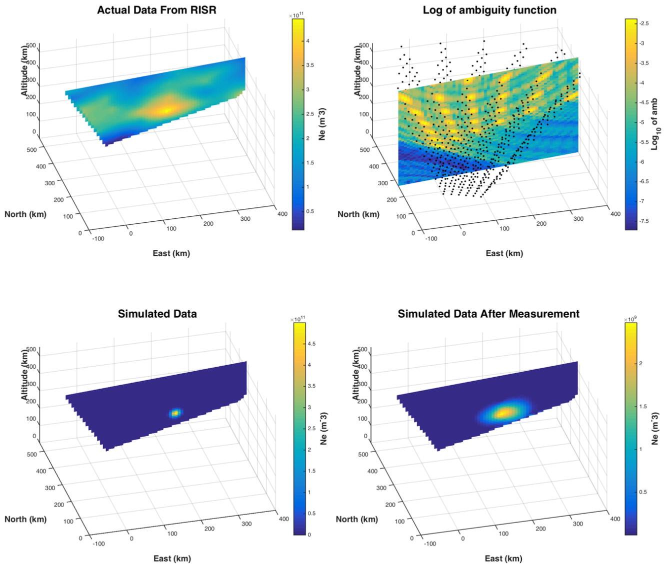
$\mathbf{r}$ - ACF

$\mathbf{Lr}$ - Space-Time Ambiguity (Linear Operator)

$\mathbf{y}$ - Blurred ACF

$\mathbf{z}$ - Blurred and Noisy ACF

Ambiguities: Data Example



MIT
HAYSTACK
OBSERVATORY

Outline

Advance Methods of Fitting

- Active area of research to improve reconstruction of the parameters
  - Better resolution
  - Enable new measurements
    - e.g. use an extremely long pulse to measure top-side
- Data Based Inversion (Lag Profile Analysis)
  - Linear Inversion, easier computationally
- Parametric Inversion (Full Profile Analysis)
  - Non-Linear Inversion, more complex computationally

Data Based Inversion (Lag Profile Analysis)

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} \|\mathbf{z} - \mathbf{L}\mathbf{r}\|_2^2 + \gamma \cdot \mathbf{f}(\mathbf{r}),$$

- Inversion of linear space-time ambiguity and then fit lags
- Constraints are generally functions of lags, does not connect directly to physics
- Many computational tools available as its similar to problems in other fields
  - e.g. deconvolution and image reconstruction
  - Examples: [VLN$^+$08], [NKKS08]

Parametric Inversion (Full Profile Analysis)

$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{z} - \mathbf{Lh}(\beta)\|_2^2 + \alpha \cdot \mathbf{f}(\beta)$$

- Inversion of linear space-time ambiguity and parameter-to-lag operator ($h(\beta)$)
- Constraints are generally functions of plasma parameters
  - e.g. $\langle d^2T_e/dz^2 \rangle, \langle d^2T_i/dz^2 \rangle$ [HRCH08]
- Not as many computational tools available
- Examples: [HRCH08], [HRTvE92], [LHP97]

MIT HAYSTACK OBSERVATORY

Outline

MIT
HAYSTACK
OBSERVATORY

Gaussian processes

Gaussian process

"A collection of random variables, any finite number of which have a joint Gaussian distribution" [Rasmussen and Williams, 2006]

- For a function $f(\mathbf{x})$, we write

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

- Fully defined by mean and covariance functions

$$m(\mathbf{x}) = \mathbf{E}[f(\mathbf{x})]$$
$$k(\mathbf{x}, \mathbf{x}') = \mathbf{E}\Big[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\Big]$$

- Evaluating at points leads to a Gaussian random vector

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_N) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \right)$$

Gaussian processes

Gaussian process

"A collection of random variables, any finite number of which have a joint Gaussian distribution" [Rasmussen and Williams, 2006]

- For a function $f(\mathbf{x})$, we write

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

- Fully defined by mean and covariance functions

$$m(\mathbf{x}) = \mathbf{E}[f(\mathbf{x})]$$
$$k(\mathbf{x}, \mathbf{x}') = \mathbf{E}\Big[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\Big]$$

- Evaluating at points leads to a Gaussian random vector

$$\mathbf{f}(\mathrm{X}) \sim \mathcal{N}(\mathbf{m}(\mathrm{X}), \mathrm{K}(\mathrm{X}, \mathrm{X}))$$

Gaussian process regression: specification

Step 1: Select forms for mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$

- Functions will typically have parameters $\boldsymbol{\theta}$

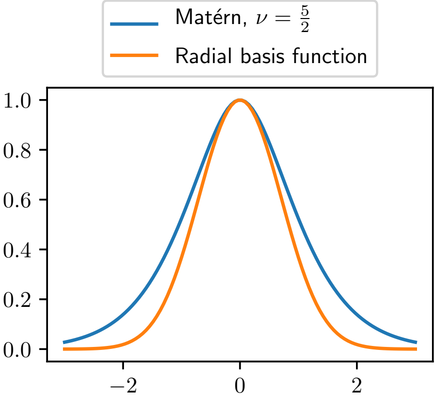
Matérn covariance, $\nu = \frac{5}{2}$

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2\left(1 + \sqrt{5}r + \tfrac{5}{3}r^2\right)e^{-\sqrt{5}r}$$

with

$$r = \left\|\frac{\mathbf{x}-\mathbf{x}'}{\boldsymbol{\delta}}\right\|_2$$

$$\boldsymbol{\theta} = \begin{bmatrix}\sigma^2 & \delta_1 & \cdots & \delta_d\end{bmatrix}^\top$$



MIT HAYSTACK OBSERVATORY

Gaussian process regression: training

Step 2: Train/fit parameters θ using measurements of $f(\mathbf{x})$

- Noisy measurements at a collection of points X

$$\mathbf{y} = \mathbf{f}(\mathrm{X}) + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\, \sigma_n^2 \mathrm{I})$$

- Maximize marginal likelihood

$$p(\mathbf{y} \mid \mathrm{X}) = \int p(\mathbf{y} \mid \mathbf{f}, \mathrm{X}) p(\mathbf{f} \mid \mathrm{X})\, d\mathbf{f}$$

$$\mathbf{y} \mid \mathrm{X} \sim \mathcal{N}\Big(\mathbf{m}(\mathrm{X}),\, \mathrm{K}(\mathrm{X}, \mathrm{X}) + \sigma_n^2 \mathrm{I}\Big)$$

$$l(\boldsymbol{\theta}) = \log p(\mathbf{y} \mid \mathrm{X}) = -\frac{1}{2}(\mathbf{y} - \mathbf{m})^\top \big(\mathrm{K} + \sigma_n^2 \mathrm{I}\big)^{-1} (\mathbf{y} - \mathbf{m}) - \frac{1}{2} \log \det\big(\mathrm{K} + \sigma_n^2 \mathrm{I}\big) - C$$

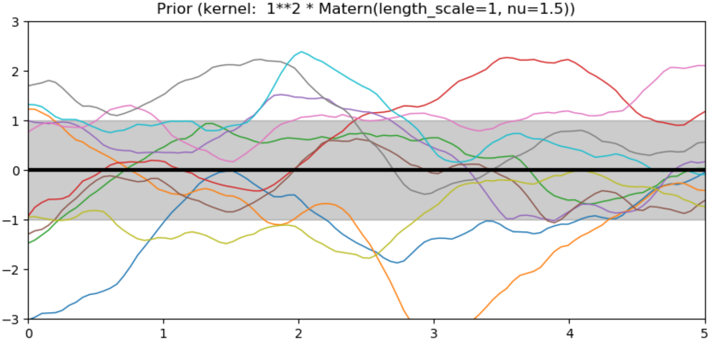
Gaussian process regression: training

Step 2: Train/fit parameters θ using measurements of $f(\mathbf{x})$

- Noisy measurements at a collection of points X

$$\mathbf{y} = \mathbf{f}(\mathrm{X}) + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\, \sigma_n^2 \mathrm{I})$$

- Maximize marginal likelihood



Gaussian process regression: prediction

Step 3: Predict f_* at a collection of test points X_*

- Joint distribution

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathrm{X}) \\ \mathbf{m}(\mathrm{X}_*) \end{bmatrix}, \begin{bmatrix} \mathrm{K}(\mathrm{X},\mathrm{X}) + \sigma_n^2 \mathrm{I} & \mathrm{K}(\mathrm{X},\mathrm{X}_*) \\ \mathrm{K}(\mathrm{X}_*,\mathrm{X}) & \mathrm{K}(\mathrm{X}_*,\mathrm{X}_*) \end{bmatrix} \right)$$

- Posterior distribution by conditioning on $\mathbf{y}$

$$\mathbf{f}_* \mid \mathbf{y}, \mathrm{X}, \mathrm{X}_* \sim \mathcal{N}(\mathbf{m}_*, \mathrm{K}_*)$$

$$\mathbf{m}_* = \mathbf{m}(\mathrm{X}_*) + \mathrm{K}(\mathrm{X}_*,\mathrm{X})\big(\mathrm{K}(\mathrm{X},\mathrm{X}) + \sigma_n^2 \mathrm{I}\big)^{-1}(\mathbf{y} - \mathbf{m}(\mathrm{X}))$$
$$\mathrm{K}_* = \mathrm{K}(\mathrm{X}_*,\mathrm{X}_*) - \mathrm{K}(\mathrm{X}_*,\mathrm{X})\big(\mathrm{K}(\mathrm{X},\mathrm{X}) + \sigma_n^2 \mathrm{I}\big)^{-1}\mathrm{K}(\mathrm{X},\mathrm{X}_*)$$

Use mean for predicted value and variance for confidence interval

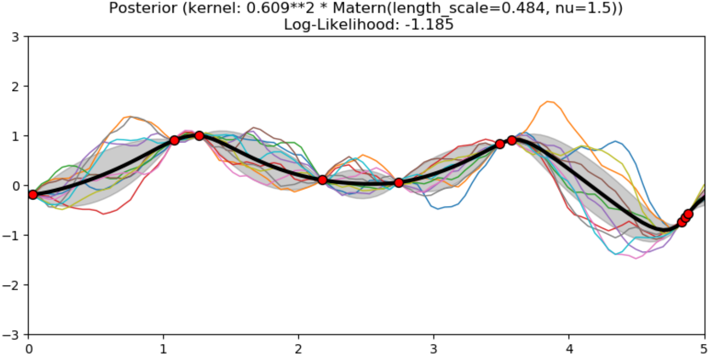
Gaussian process regression: prediction

Step 3: Predict f_* at a collection of test points X_*

- Joint distribution

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathrm{X}) \\ \mathbf{m}(\mathrm{X}_*) \end{bmatrix}, \begin{bmatrix} \mathrm{K}(\mathrm{X},\mathrm{X}) + \sigma_n^2 \mathrm{I} & \mathrm{K}(\mathrm{X},\mathrm{X}_*) \\ \mathrm{K}(\mathrm{X}_*,\mathrm{X}) & \mathrm{K}(\mathrm{X}_*,\mathrm{X}_*) \end{bmatrix} \right)$$

- Posterior distribution by conditioning on $\mathbf{y}$



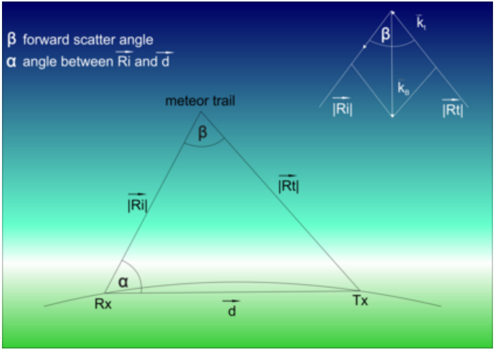
Meteor wind measurements

- Doppler shift in Bragg direction for a single location/time

$$f(x,y,z,t) = \frac{1}{2\pi}\begin{bmatrix} k_x & k_y & k_z \end{bmatrix}\begin{bmatrix} u(x,y,z,t) \\ v(x,y,z,t) \\ w(x,y,z,t) \end{bmatrix}$$

where

  - k_x, k_y, k_z are Bragg vector components
  - u, v, and w are the unknown wind components



MIT HAYSTACK OBSERVATORY

Vectorized Doppler measurement equation

$$f(x,y,z,t) = \frac{1}{2\pi}\begin{bmatrix}k_x & k_y & k_z\end{bmatrix}\begin{bmatrix}u(x,y,z,t)\\ v(x,y,z,t)\\ w(x,y,z,t)\end{bmatrix}$$

- Measure at a set of points X with noise $\boldsymbol{\epsilon}$

$$\mathbf{y}(\mathrm{X}) = \mathbf{a}_u \odot \mathbf{u} + \mathbf{a}_v \odot \mathbf{v} + \mathbf{a}_w \odot \mathbf{w} + \boldsymbol{\epsilon}$$

$$\mathrm{X} = \begin{bmatrix}\mathbf{x}_1^\top\\ \vdots\\ \mathbf{x}_M^\top\end{bmatrix} = \begin{bmatrix}x_1 & y_1 & z_1 & t_1\\ \vdots & \vdots & \vdots & \vdots\\ x_M & y_M & z_M & t_M\end{bmatrix} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0,\ \Sigma_n)$$

$$\mathbf{a}_u = \frac{1}{2\pi}\begin{bmatrix}k_{x\,1}\\ \vdots\\ k_{x\,M}\end{bmatrix} \qquad \mathbf{a}_v = \frac{1}{2\pi}\begin{bmatrix}k_{y\,1}\\ \vdots\\ k_{y\,M}\end{bmatrix} \qquad \mathbf{a}_w = \frac{1}{2\pi}\begin{bmatrix}k_{z\,1}\\ \vdots\\ k_{z\,M}\end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix}u(\mathbf{x}_1)\\ \vdots\\ u(\mathbf{x}_M)\end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix}v(\mathbf{x}_1)\\ \vdots\\ v(\mathbf{x}_M)\end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix}w(\mathbf{x}_1)\\ \vdots\\ w(\mathbf{x}_M)\end{bmatrix}$$

Gaussian process prior for winds

- Model each wind component as a Gaussian process

$$
\begin{aligned}
u(\mathbf{x}) &\sim \mathcal{GP}(m_u(\mathbf{x}), k_u(\mathbf{x}, \mathbf{x}')) \\
v(\mathbf{x}) &\sim \mathcal{GP}(m_v(\mathbf{x}), k_v(\mathbf{x}, \mathbf{x}')) \\
w(\mathbf{x}) &\sim \mathcal{GP}(m_w(\mathbf{x}), k_w(\mathbf{x}, \mathbf{x}'))
\end{aligned}
$$

- Choose prior mean and covariance for u, v, w

Constant means

$$
\begin{aligned}
m_u(\mathbf{x}) &= u_0 \\
m_v(\mathbf{x}) &= v_0 \\
m_w(\mathbf{x}) &= w_0
\end{aligned}
$$

Common Matérn covariance

$$
\begin{aligned}
k_u(\mathbf{x}, \mathbf{x}') = k_v(\mathbf{x}, \mathbf{x}') &= k_w(\mathbf{x}, \mathbf{x}') \\
&= k_{\text{Matérn}, \nu=\frac{5}{2}}(\mathbf{x}, \mathbf{x}'; \sigma^2, \delta_x, \delta_y, \delta_z, \delta_t)
\end{aligned}
$$

- Can fit parameters and/or apply physical knowledge

MIT HAYSTACK OBSERVATORY

Gaussian prior for Doppler measurements

$$u(\mathbf{x}) \sim \mathcal{GP}(m_u(\mathbf{x}), k_u(\mathbf{x}, \mathbf{x}'))$$

$$v(\mathbf{x}) \sim \mathcal{GP}(m_v(\mathbf{x}), k_v(\mathbf{x}, \mathbf{x}'))$$

$$w(\mathbf{x}) \sim \mathcal{GP}(m_w(\mathbf{x}), k_w(\mathbf{x}, \mathbf{x}'))$$

$$\mathbf{y}(\mathrm{X}) = \mathbf{a}_u \odot \mathbf{u} + \mathbf{a}_v \odot \mathbf{v} + \mathbf{a}_w \odot \mathbf{w} + \epsilon$$

- Multivariate Gaussian for winds at measurement points

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_u(\mathrm{X}) \\ \mathbf{m}_v(\mathrm{X}) \\ \mathbf{m}_w(\mathrm{X}) \end{bmatrix}, \begin{bmatrix} \mathrm{K}_u(\mathrm{X},\mathrm{X}) & 0 & 0 \\ 0 & \mathrm{K}_v(\mathrm{X},\mathrm{X}) & 0 \\ 0 & 0 & \mathrm{K}_w(\mathrm{X},\mathrm{X}) \end{bmatrix} \right)$$

- Resulting multivariate Gaussian for Doppler measurements

$$\mathbf{y} \sim \mathcal{N}(\mathbf{m}_y(\mathrm{X}), \mathrm{K}_y(\mathrm{X},\mathrm{X}))$$

$$\mathbf{m}_y(\mathrm{X}) = \mathbf{a}_u \odot \mathbf{m}_u(\mathrm{X}) + \mathbf{a}_v \odot \mathbf{m}_v(\mathrm{X}) + \mathbf{a}_w \odot \mathbf{m}_w(\mathrm{X})$$

$$\mathrm{K}_y(\mathrm{X},\mathrm{X}) = (\mathbf{a}_u \mathbf{a}_u{}^\top) \odot \mathrm{K}_u(\mathrm{X},\mathrm{X}) + (\mathbf{a}_v \mathbf{a}_v{}^\top) \odot \mathrm{K}_v(\mathrm{X},\mathrm{X}) + (\mathbf{a}_w \mathbf{a}_w{}^\top) \odot \mathrm{K}_w(\mathrm{X},\mathrm{X}) + \Sigma_n$$

MIT HAYSTACK OBSERVATORY

Gaussian prior for Doppler measurements

$$u(\mathbf{x}) \sim \mathcal{GP}(m_u(\mathbf{x}), k_u(\mathbf{x}, \mathbf{x}'))$$
$$v(\mathbf{x}) \sim \mathcal{GP}(m_v(\mathbf{x}), k_v(\mathbf{x}, \mathbf{x}'))$$
$$w(\mathbf{x}) \sim \mathcal{GP}(m_w(\mathbf{x}), k_w(\mathbf{x}, \mathbf{x}'))$$

$$\mathbf{y}(\mathrm{X}) = \mathbf{a}_u \odot \mathbf{u} + \mathbf{a}_v \odot \mathbf{v} + \mathbf{a}_w \odot \mathbf{w} + \boldsymbol{\epsilon}$$

- Multivariate Gaussian for winds at measurement points

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_u(\mathrm{X}) \\ \mathbf{m}_v(\mathrm{X}) \\ \mathbf{m}_w(\mathrm{X}) \end{bmatrix}, \begin{bmatrix} \mathrm{K}_u(\mathrm{X},\mathrm{X}) & 0 & 0 \\ 0 & \mathrm{K}_v(\mathrm{X},\mathrm{X}) & 0 \\ 0 & 0 & \mathrm{K}_w(\mathrm{X},\mathrm{X}) \end{bmatrix} \right)$$

- Resulting multivariate Gaussian for Doppler measurements

$$\mathbf{y} \sim \mathcal{N}\big(\mathbf{m}_y(\mathrm{X}),\, \mathrm{K}_y(\mathrm{X},\mathrm{X})\big)$$

$$\mathbf{m}_y(\mathrm{X}) = \mathbf{a}_u \odot \mathbf{m}_u(\mathrm{X}) + \mathbf{a}_v \odot \mathbf{m}_v(\mathrm{X}) + \mathbf{a}_w \odot \mathbf{m}_w(\mathrm{X})$$
$$\mathrm{K}_y(\mathrm{X},\mathrm{X}) = (\mathbf{a}_u\mathbf{a}_u{}^\top) \odot \mathrm{K}_u(\mathrm{X},\mathrm{X}) + (\mathbf{a}_v\mathbf{a}_v{}^\top) \odot \mathrm{K}_v(\mathrm{X},\mathrm{X}) + (\mathbf{a}_w\mathbf{a}_w{}^\top) \odot \mathrm{K}_w(\mathrm{X},\mathrm{X}) + \Sigma_n$$

MIT HAYSTACK OBSERVATORY

Doppler measurement fitting

$$\mathbf{m}_y(\mathrm{X}) = \mathbf{a}_u \odot \mathbf{m}_u(\mathrm{X}) + \mathbf{a}_v \odot \mathbf{m}_v(\mathrm{X}) + \mathbf{a}_w \odot \mathbf{m}_w(\mathrm{X})$$

$$\begin{aligned}\mathrm{K}_y(\mathrm{X},\mathrm{X}) = {} & (\mathbf{a}_u\mathbf{a}_u{}^\top) \odot \mathrm{K}_u(\mathrm{X},\mathrm{X}) + (\mathbf{a}_v\mathbf{a}_v{}^\top) \odot \mathrm{K}_v(\mathrm{X},\mathrm{X}) \\ & + (\mathbf{a}_w\mathbf{a}_w{}^\top) \odot \mathrm{K}_w(\mathrm{X},\mathrm{X}) + \Sigma_n\end{aligned}$$

- Maximize likelihood

$$l(\boldsymbol{\theta}) = -\frac{1}{2}(\mathbf{y} - \mathbf{m}_y)^\top \mathrm{K}_y{}^{-1}(\mathbf{y} - \mathbf{m}_y) - \frac{1}{2}\log\det \mathrm{K}_y - C$$

Resulting parameters: March 14, 2016, 08:00 - 12:00

$u_0 = -20$ m/s

$v_0 = -10$ m/s

$w_0 = -2$ m/s

$\sigma^2 = 500$ (m/s)^2

$\delta_x = 40$ km

$\delta_y = 20$ km

$\delta_z = 2$ km

$\delta_t = 30$ min

Wind estimation

- Can write posterior distribution for winds at prediction points X_*

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u}_* \\ \mathbf{v}_* \\ \mathbf{w}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_y(\mathrm{X}) \\ \mathbf{m}_u(\mathrm{X}_*) \\ \mathbf{m}_v(\mathrm{X}_*) \\ \mathbf{m}_w(\mathrm{X}_*) \end{bmatrix}, \kappa \right)$$

$$\kappa = \begin{bmatrix} \mathrm{K}_y(\mathrm{X},\mathrm{X}) & \mathbf{a}_u \odot \mathrm{K}_u(\mathrm{X},\mathrm{X}_*) & \mathbf{a}_v \odot \mathrm{K}_v(\mathrm{X},\mathrm{X}_*) & \mathbf{a}_w \odot \mathrm{K}_w(\mathrm{X},\mathrm{X}_*) \\ \mathrm{K}_u(\mathrm{X}_*,\mathrm{X}) \odot \mathbf{a}_u & \mathrm{K}_u(\mathrm{X}_*,\mathrm{X}_*) & 0 & 0 \\ \mathrm{K}_v(\mathrm{X}_*,\mathrm{X}) \odot \mathbf{a}_v & 0 & \mathrm{K}_v(\mathrm{X}_*,\mathrm{X}_*) & 0 \\ \mathrm{K}_w(\mathrm{X}_*,\mathrm{X}) \odot \mathbf{a}_w & 0 & 0 & \mathrm{K}_w(\mathrm{X}_*,\mathrm{X}_*) \end{bmatrix}$$

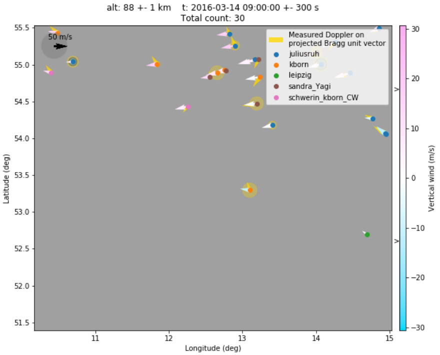
- Estimate given by mean (just linear algebra)

$$\mathbf{E}[\mathbf{u}_* \mid \mathbf{y}] = \mathbf{m}_u(\mathrm{X}_*) + \mathrm{K}_u(\mathrm{X}_*,\mathrm{X}) \odot \mathbf{a}_u \odot \mathrm{K}_y(\mathrm{X},\mathrm{X})^{-1}(\mathbf{y} - \mathbf{m}_y(\mathrm{X}))$$
$$\mathbf{E}[\mathbf{v}_* \mid \mathbf{y}] = \mathbf{m}_v(\mathrm{X}_*) + \mathrm{K}_v(\mathrm{X}_*,\mathrm{X}) \odot \mathbf{a}_v \odot \mathrm{K}_y(\mathrm{X},\mathrm{X})^{-1}(\mathbf{y} - \mathbf{m}_y(\mathrm{X}))$$
$$\mathbf{E}[\mathbf{w}_* \mid \mathbf{y}] = \mathbf{m}_w(\mathrm{X}_*) + \mathrm{K}_w(\mathrm{X}_*,\mathrm{X}) \odot \mathbf{a}_w \odot \mathrm{K}_y(\mathrm{X},\mathrm{X})^{-1}(\mathbf{y} - \mathbf{m}_y(\mathrm{X}))$$

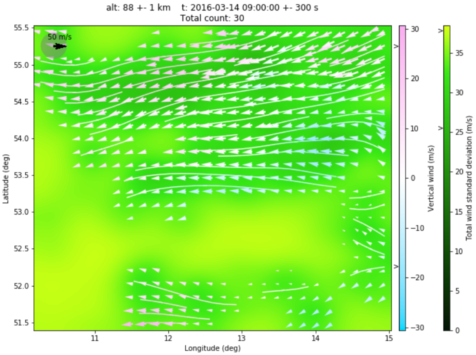
- Posterior covariance can be calculated as well

Example wind field estimates

Measurements



Estimates



References I

- D L Hysell, F S Rodrigues, J L Chau, and J D Huba, Full profile incoherent scatter analysis at Jicamarca, Annales Geophysicae **26** (2008), no. 1, 59–75.
- J M Holt, D A Rhoda, D Tetenbaum, and A P van Eyken, Optimal analysis of incoherent scatter radar data, Radio Science **27** (1992), no. 3, 435–447.
- David Hysell, Overspread Targets, Antennas and Radar for Environmental Scientists and Engineers, Cambridge University Press, 2018, p. 264–287.
- E Kudeki and M A Milla, Incoherent Scatter Spectral Theories: Part I: A General Framework and F
IEEE Transactions on Geoscience and Remote Sensing **49** (2011), no. 1, 315–328.

MIT HAYSTACK OBSERVATORY

References II

- M S Lehtinen, A Huuskonen, and J Pirttilä, First experiences of full-profile analysis with GUISDAP, Annales Geophysicae (1997).
- Romina Nikoukar, Farzad Kamalabadi, Erhan Kudeki, and Michael Sulzer, An efficient near-optimal approach to incoherent scatter radar paramete
Radio Science **43** (2008), no. 5.
- A Papoulis and S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4 ed., McGraw-Hill Higher Education, 2002.
- I I Virtanen, M S Lehtinen, T Nygrén, M Orispaa, and Vierinen, Lag profile inversion method for EISCAT data analysis, Annales Geophysicae **26** (2008), no. 571-581.

MIT
HAYSTACK
OBSERVATORY