# Towards a Bayesian analysis of impurity transport data

M.A. Chilenski, Y. Marzouk,[*] M. Greenwald,
N.T. Howard, J.E. Rice, and A.E. White

MIT PSFC/Alcator C-Mod
[*]MIT Aero/Astro, Uncertainty Quantification Group

First IAEA Technical Meeting on Fusion Data Processing, Validation and Analysis
Nice, France
June 3, 2015

Supported by USDoE award DE-FC02-99ER54512. Supported in part by a DOE SCGF fellowship, administered by ORISE-ORAU under contract DE-AC05-06OR23100. Computations using STRAHL were carried out on the MIT PSFC parallel AMD Opteron/Infiniband cluster Loki.

# Increasing confidence in validation studies through statistically rigorous inference of impurity transport coefficient profiles

## Motivation

- Validation of simulations requires rigorous inference of the experimental quantities used for comparison.
- Current approaches to inferring impurity transport coefficients suffer from issues with:
  - Uniqueness of solution
  - Complete accounting of uncertainty

## Outline

- Measuring impurity transport coefficients on Alcator C-Mod.
- Current approaches and their shortcomings.
- Fully Bayesian inference of impurity transport coefficients using MCMC.
- *Preliminary* results.

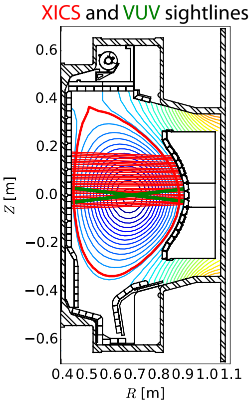
# Alcator C-Mod is uniquely equipped to make detailed measurements of impurity transport

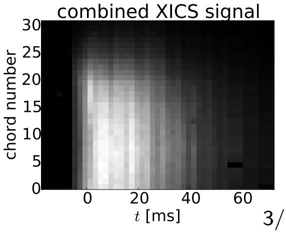
## Multipulse laser blow-off impurity injector provides controlled impurity injections [1]

- Multiple injections per shot: up to 10 Hz
- Typically inject $CaF_2$: calcium is *non-intrinsic* and *non-recycling*

## X-ray imaging crystal spectrometer [2] and VUV spectrometers [3] track the impurities

- XICS observes spatial profile of a *single charge state* ($Ca^{18+}$): more direct interpretation than unresolved soft x-rays
- Two single-chord VUV spectrometers measure $Ca^{16+}$, $Ca^{17+}$

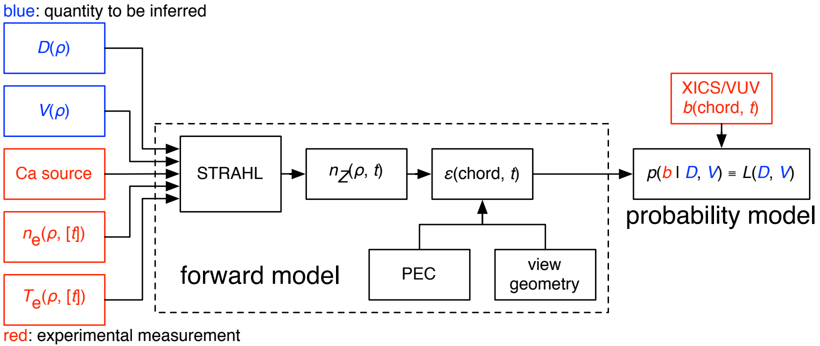




[1] Howard et al. (2011), RSI [2] Ince-Cushman et al. (2008), RSI
[3] Reinke et al. (2010), RSI

# Inferring impurity transport coefficients is a nonlinear inverse problem




- Objective is to find $D$, $V$ profiles that best reproduce the observed brightnesses $b$ on each of the diagnostics.
- Key issues are existence, uniqueness and stability of the solution.

[1] Dux (2006), IPP Report 10/30

# Current approaches: maximum likelihood estimate (MLE)

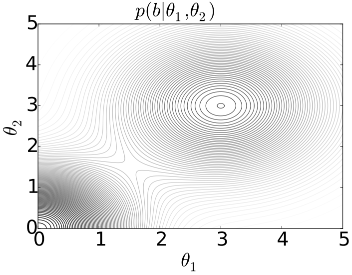
## MLE is a standard approach to handle this problem…

$$\hat{D}, \hat{V} = \arg\max_{D,V} p(b|D, V)$$

- Pick $D$, $V$ profiles which make the observations most likely.
- Use standard optimization techniques: assumption of Gaussian noise makes this a "least squares" problem.
- Need basis functions to represent the profiles with a finite number of variables: typically piecewise linear functions with fixed knots.

## …but it has some potential shortcomings

- Point estimate:
  - Risk of underestimating uncertainty.
  - Not valid when there are multiple extrema.
- Propagation of uncertainty in $n_e$, $T_e$ profiles requires an additional step.



# Bayesian statistics provides a framework to overcome the shortcomings of MLE

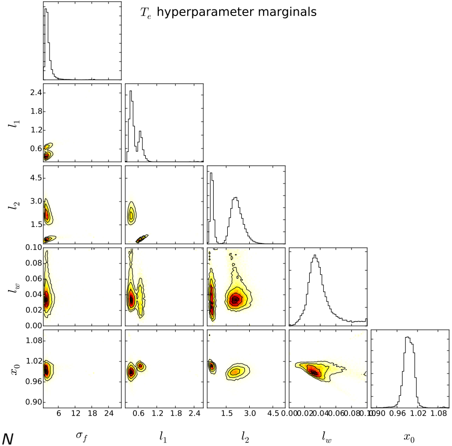
- Use Bayes' rule to obtain the posterior distribution $p(D, V|b)$, including constraints/prior knowledge $p(D, V)$:

$$p(D, V|b) \propto p(b|D, V)p(D, V)$$

- $p(D, V|b)$ represents the state of knowledge about $D$, $V$ after having accounted for the data $b$.
- Working with $p(D, V|b)$ avoids the issues of MLE.

# Markov chain Monte Carlo (MCMC) sampling enables a complete accounting of uncertainty

- MCMC draws samples from unnormalized probability distribution such as $D^{(i)}, V^{(i)} \sim p(D, V|b) \propto p(b|D, V)p(D, V)$.
- Histogram to view $p(D, V|b)$ directly: nonuniqueness can be identified immediately.
- Allows for better point estimates, such as posterior mean and variance:



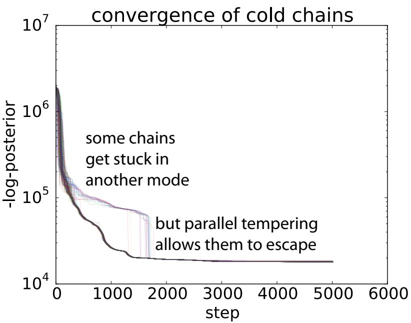

Chilenski et al. (2015), NF

$$\mathbb{E}[D|b] = \int Dp(D|b)\,\mathrm{d}D \approx \frac{1}{N}\sum_{i=1}^{N} D^{(i)}$$

$$\mathrm{var}[D|b] = \int (D - \mathbb{E}[D|b])^2 p(D|b)\,\mathrm{d}D \approx \frac{1}{N-1}\sum_{i=1}^{N}(D^{(i)} - \mathbb{E}[D|b])^2$$

# Multimodal posterior necessitates advanced MCMC

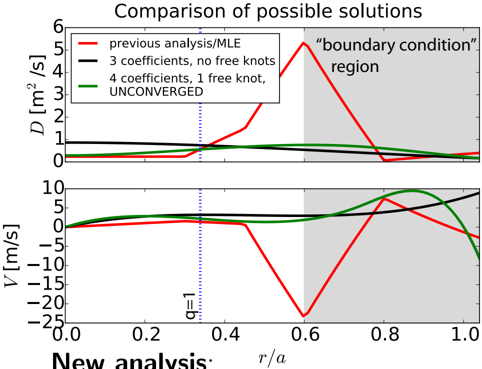
- Affine-invariant ensemble sampler (ES) [1, 2]
  - Eliminates need to tune proposal distribution.
  - But, cannot efficiently sample distributions with well-separated modes.
- Parallel tempering (PT) [3]
  - Sample from $p(b|D, V)^{1/T} p(D, V)$ for multiple values of $1 \leq T \leq \infty$.
  - Exchange of information between adjacent $T$ lets chains move between modes.
- Adaptive parallel tempering (APT) [4]
  - Automatically tune $T$ ladder.




- APT with ES in each temperature.
- 200 walkers per temperature, 25 temperatures.
- Plot shows $-\ln p(D, V|b)$ on a log scale: lower value = better fit.

[1] Goodman and Weare (2010), CAMCS [2] Foreman-Mackey et al. (2013), PASP
[3] Earl and Deem (2010), PCCP [4] Vousden et al. (2015), arXiv:1501.05823

# PRELIMINARY results do not match previous analysis



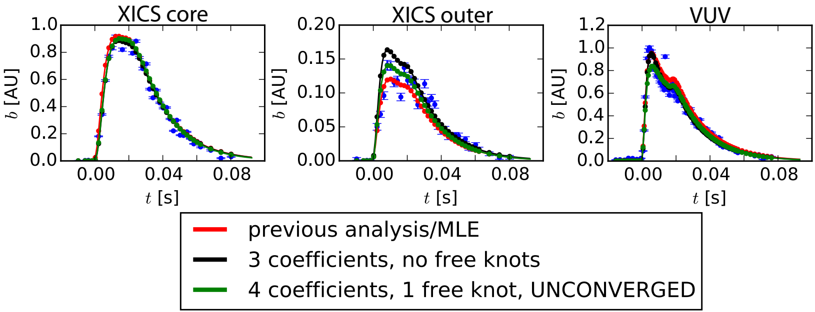

**Previous analysis**:

- Piecewise linear basis functions.
- MLE without estimate of width of posterior distribution.
- Behavior in $r/a > 0.6$ thought to be only weakly constrained.
- But, uncertainty there too small to be consistent with this.

**New analysis**:

- Cubic B-spline basis functions.
- APT to handle multiple maxima, width of posterior distribution.
- Uncertainty estimate in $r/a > 0.6$ still too small to be consistent with assumed lack of knowledge there.
- *Cases shown are likely overconstrained.*
- Models with more free parameters are running now, but have not found any reasonable maxima yet.

# Predicted brightnesses are similar between all three cases




- Agreement on core XICS chords is good in all cases.
- Agreement on outer XICS chords shows widest variation – 4 coefficient case seems to do best job.
- Agreement on VUV spectrometer is reasonable in all cases.
- *This shows the importance of accounting for the possibility of multiple solutions.*

# Next step: include uncertainty in $n_e$, $T_e$ profiles

Form joint posterior distribution, now also conditional on the profile measurements $d$:

$$p(D, V, n_e, T_e | b, d) = p(D, V | n_e, T_e, b, d) p(n_e, T_e | b, d)$$

Use Gaussian processes for $n_e$, $T_e$ [1]:

$$p(n_e | d) = \mathcal{N}(m(\rho), k(\rho, \rho))$$

Reduce dimension of parameter space by approximating this with truncated eigendecomposition:

$$n_e = Q\Lambda^{1/2} u + m(\rho), \quad u \sim \mathcal{N}(0, I), \quad k(\rho, \rho) = Q\Lambda Q^{-1}$$

Find marginal posterior distribution for $D$, $V$ using MCMC:

$$p(D, V | b, d) = \int p(D, V, n_e, T_e | b, d) \, \mathrm{d}n_e \, \mathrm{d}T_e$$

[1] Chilenski et al. (2015), NF

## Application of Bayesian inference allows rigorous estimation of impurity transport coefficient profiles, better confidence in validation studies

- The combination of XICS and LBO enables detailed studies of impurity transport on Alcator C-Mod.
- Inferring impurity transport coefficient profiles using point estimates such as maximum likelihood suffers from issues with:
  - Uniqueness of solution
  - Complete accounting of uncertainty
- New approach under development: use MCMC to find "all" physically reasonable solutions to yield a complete accounting of uncertainty.

# Backup slides

# Model selection with information criteria [1]

- Formalize the tradeoff between goodness of fit and complexity of model: picking the model which minimizes an information criterion is a way to avoid overfitting.
- Common choice: Deviance information criterion (DIC)

$$DIC = -2 \ln p(b|\,\mathbb{E}[D|b], \mathbb{E}[V|b]) + 2p_{eff}$$

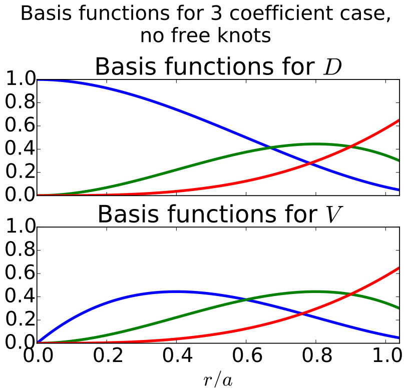
  - Effective number of parameters $p_{eff}$ has two definitions:

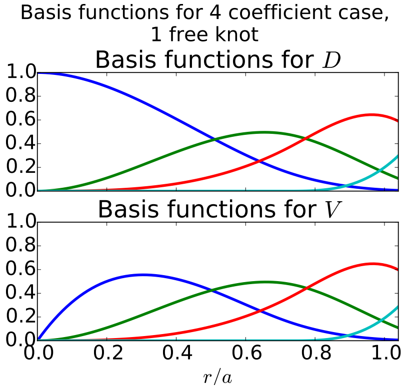
$$p_{eff,1} = 2\left[\ln p(b|\,\mathbb{E}[D|b], \mathbb{E}[V|b]) - \mathbb{E}[\ln p(b|D, V)]\right]$$
$$p_{eff,2} = 2\,\mathrm{var}[\ln p(b|D, V)]$$

  - $p_{eff,1} = p_{eff,2} =$ true # of parameters for linear models with flat priors.
  - Easy to compute from MCMC output.

[1] Gelman et al. (2014), BDA3

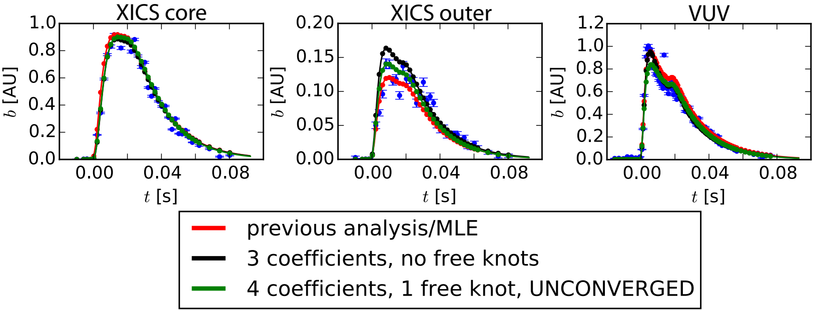
# B-spline basis functions are used to obtain a smooth profile, impose constraints



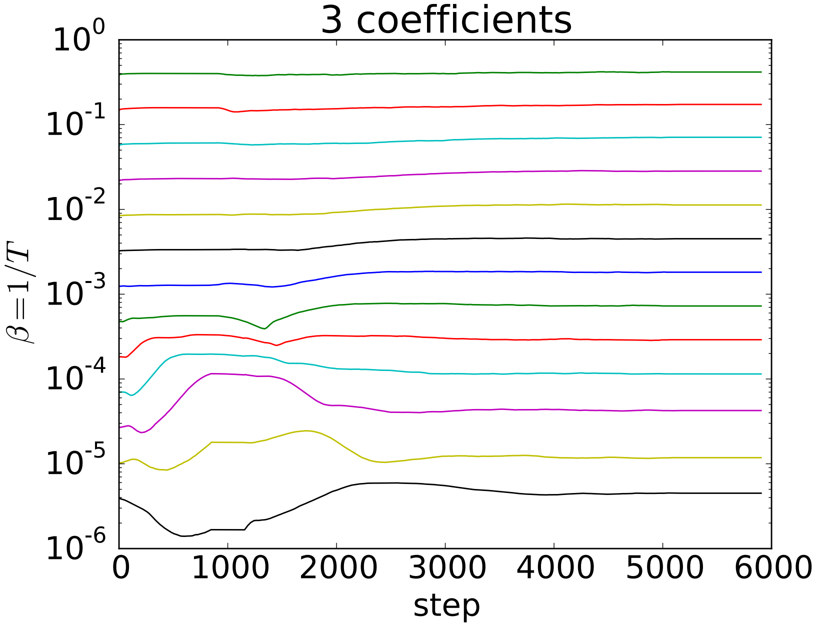

- $\mathrm{d}D/\mathrm{d}r = 0$ at $r/a = 0$
- $D \geq 0$ everywhere
- $V(0) = 0$

# Goodness of fit is comparable between all three solutions

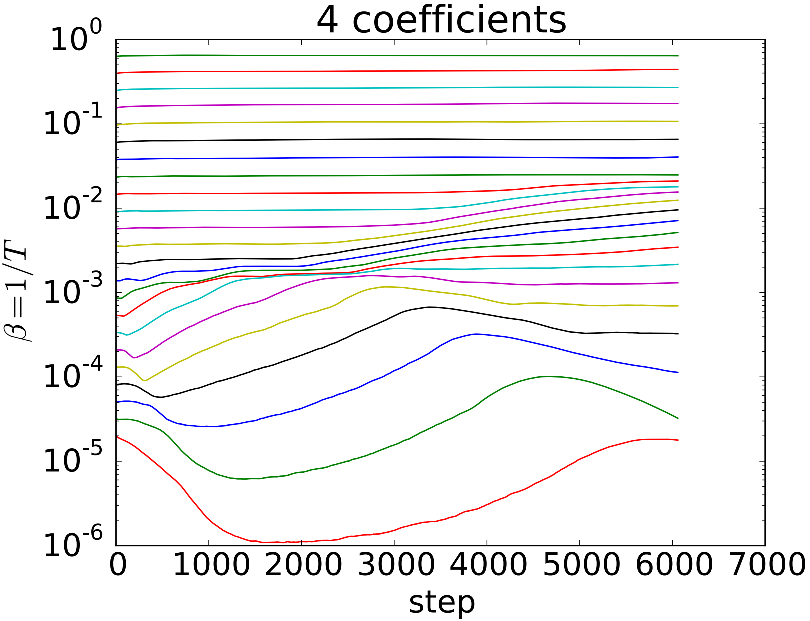



| case | $-\ln p(b\|D, V) \sim \chi^2$ |
|---|---|
| previous analysis/MLE | $3.80 \times 10^4$ |
| 3 coefficients, no free knots | $2.15 \times 10^4$ |
| 4 coefficients, 1 free knot | $1.81 \times 10^4$ |

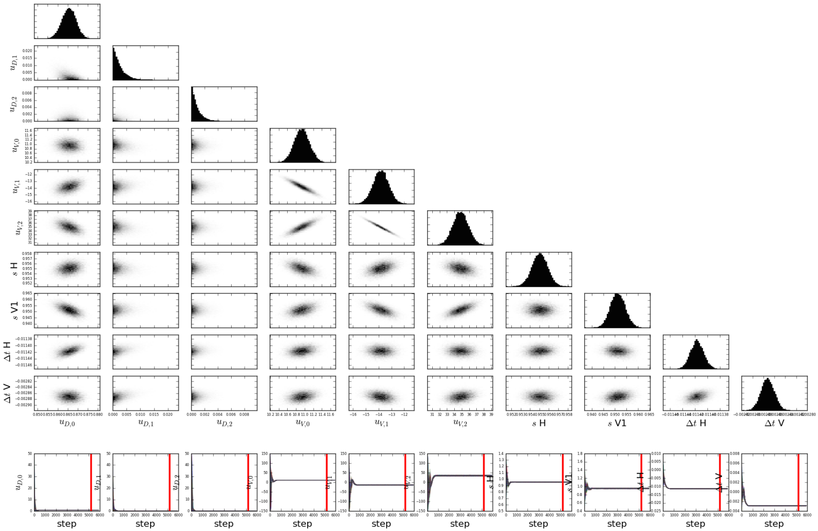
# Temperature ladder adaptation for 3 coefficient case



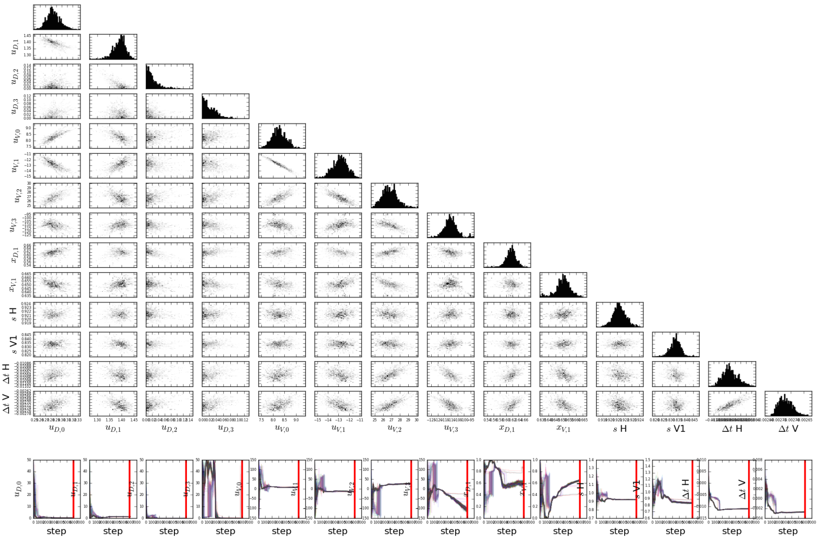
# Temperature ladder adaptation for 4 coefficient case



# Posterior distribution for 3 coefficient case

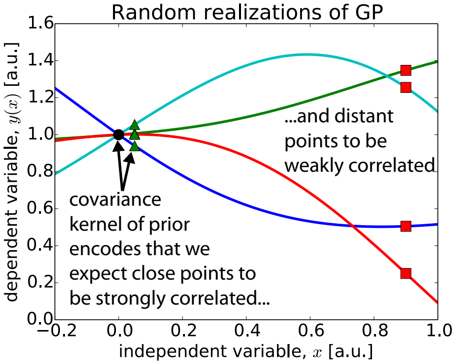


# Posterior distribution for 4 coefficient case



# Profile fitting with Gaussian process regression (GPR)

- Established statistical/machine learning technique [1].
- Expresses profile in terms of (spatial) covariance of multivariate normal (MVN) distribution.
- Selection of fit properties is **automated** and **statistically rigorous**.

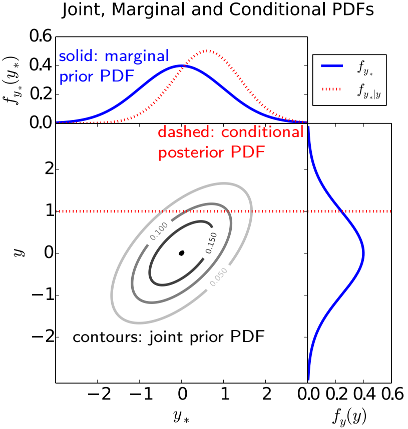


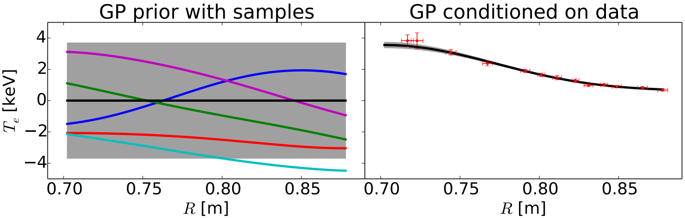


[1] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

# GPR: a probabilistic method to fit profiles

- Create multivariate normal *prior distribution* that sets smoothness, symmetry, etc.
- *Condition* on observations to yield the fit, including uncertainty estimate.
- Distribution can include derivatives, line integrals, volume averages, etc.



# gptools implements a very general form of GPR

$$f\left(\begin{bmatrix}\boldsymbol{M}_* \\ \boldsymbol{M}\end{bmatrix}\right) = \mathcal{N}\Bigg(\begin{bmatrix}\mathsf{T}_* & 0 \\ 0 & \mathsf{T}\end{bmatrix}\begin{bmatrix}\boldsymbol{\mu}(\mathsf{X}_*) \\ \boldsymbol{\mu}(\mathsf{X})\end{bmatrix},$$
$$\begin{bmatrix}\mathsf{T}_* & 0 \\ 0 & \mathsf{T}\end{bmatrix}\begin{bmatrix}\mathsf{K}(\mathsf{X}_*,\mathsf{X}_*) & \mathsf{K}(\mathsf{X},\mathsf{X}_*) \\ \mathsf{K}(\mathsf{X}_*,\mathsf{X}) & \mathsf{K}(\mathsf{X},\mathsf{X})\end{bmatrix}\begin{bmatrix}\mathsf{T}_*^T & 0 \\ 0 & \mathsf{T}^T\end{bmatrix} + \begin{bmatrix}0 & 0 \\ 0 & \Sigma_M\end{bmatrix}\Bigg)$$

$$\ln\mathcal{L} = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\left|\mathsf{T}\mathsf{K}(\mathsf{X},\mathsf{X})\mathsf{T}^T + \Sigma_M\right|$$
$$-\frac{1}{2}(\boldsymbol{M} - \mathsf{T}\boldsymbol{\mu}(\mathsf{X}))^T(\mathsf{T}\mathsf{K}(\mathsf{X},\mathsf{X})\mathsf{T}^T + \Sigma_M)^{-1}(\boldsymbol{M} - \mathsf{T}\boldsymbol{\mu}(\mathsf{X}))$$

$$f(\boldsymbol{M}_*|\boldsymbol{M}) = \mathcal{N}\big(\mathsf{T}_*\boldsymbol{\mu}(\mathsf{X}_*) + \mathsf{T}_*\mathsf{K}(\mathsf{X}_*,\mathsf{X})\mathsf{T}^T(\mathsf{T}\mathsf{K}(\mathsf{X},\mathsf{X})\mathsf{T}^T + \Sigma_M)^{-1}(\boldsymbol{M} - \mathsf{T}\boldsymbol{\mu}(\mathsf{X})),$$
$$\mathsf{T}_*\mathsf{K}(\mathsf{X}_*,\mathsf{X}_*)\mathsf{T}_*^T - \mathsf{T}_*\mathsf{K}(\mathsf{X}_*,\mathsf{X})\mathsf{T}^T(\mathsf{T}\mathsf{K}(\mathsf{X},\mathsf{X})\mathsf{T}^T + \Sigma_M)^{-1}\mathsf{T}\mathsf{K}(\mathsf{X},\mathsf{X}_*)\mathsf{T}_*^T\big)$$

- Supports data of arbitrary dimension $\boldsymbol{x} \in \mathbb{R}^n$.
- Supports explicit, parametric mean function $\mu(\boldsymbol{x})$: can perform nonlinear regression with GP fit to residuals.
- Supports arbitrary linear transformations $\mathsf{T}$, $\mathsf{T}_*$ of inputs, outputs: can perform tomographic inversions constrained with point measurements.
- Supports noise of arbitrary structure $\Sigma_M$ on observations.