MIT EARTH RESOURCES LABORATORY
ANNUAL FOUNDING MEMBERS MEETING 2020

Assessing inference quality under model misspecification

An application to seismic inversion.

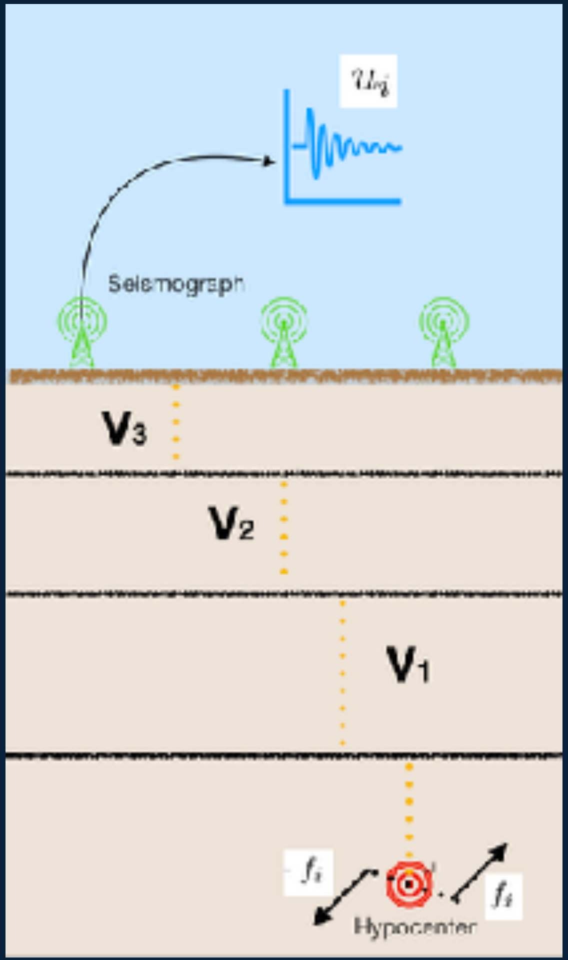
Andrea Scarinci
PHD CANDIDATE, AERONAUTICS AND ASTRONAUTICS

Supervisors: Youssef Marzouk, Michael Fehler.

What makes a good posterior?

The problem

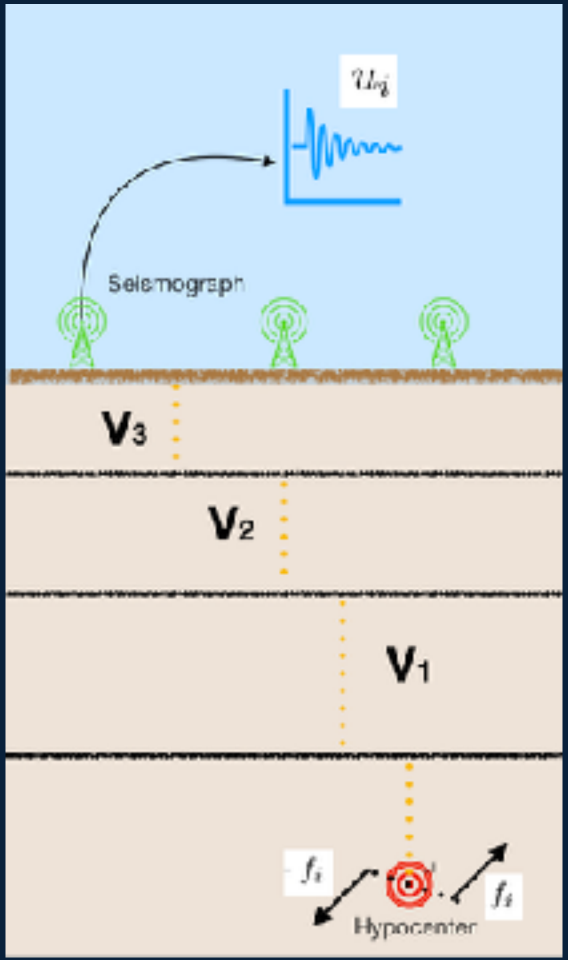
APPLICATION: BAYESIAN MOMENT TENSOR INVERSION



The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

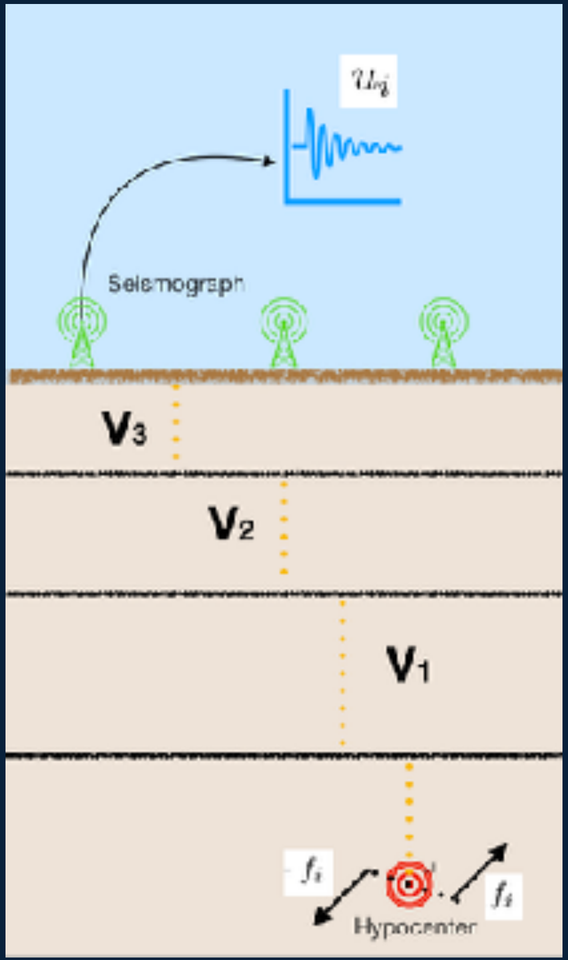


The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$



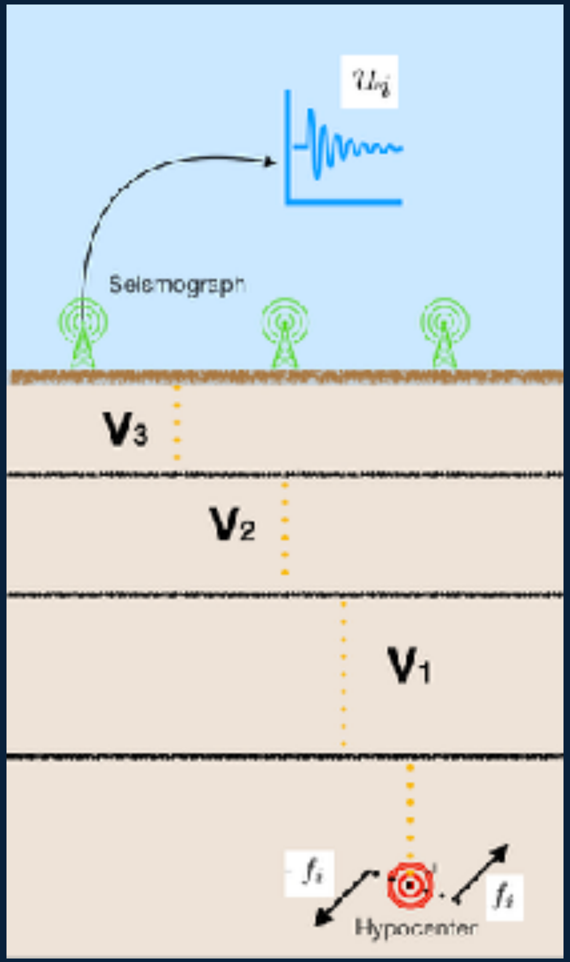
The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

$\mathbf{X}$ = location of the earthquake



The problem

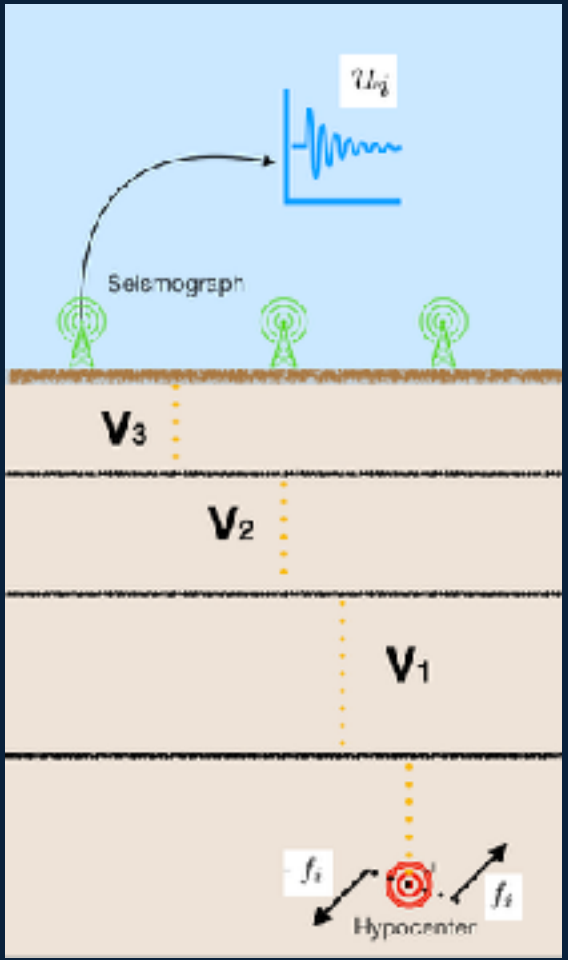
APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

$\mathbf{X}$ = location of the earthquake

$\mathbf{V}$ = velocity model



The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

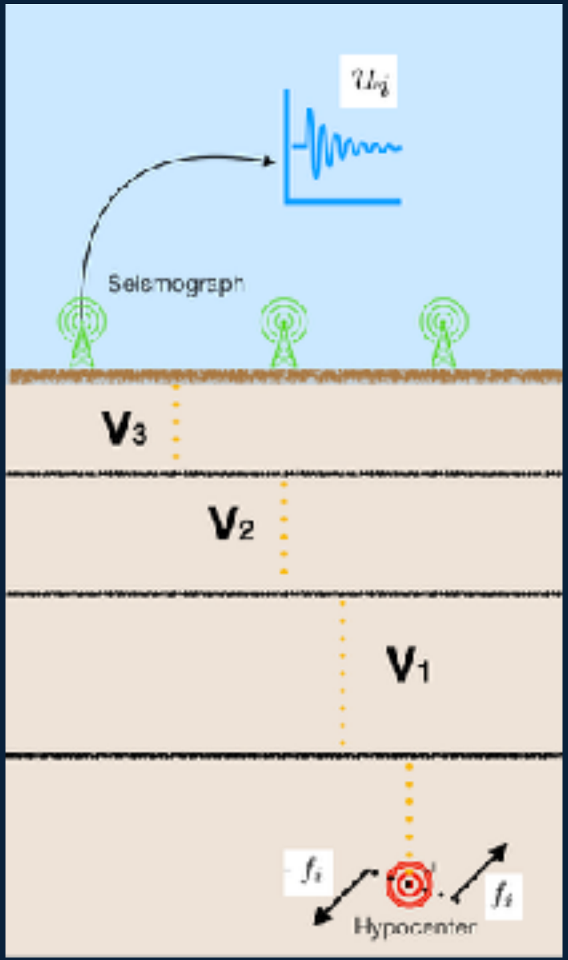
MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

$\mathbf{x}$ = location of the earthquake

$\mathbf{V}$ = velocity model

$\mathbf{m}$ = moment tensor



The problem

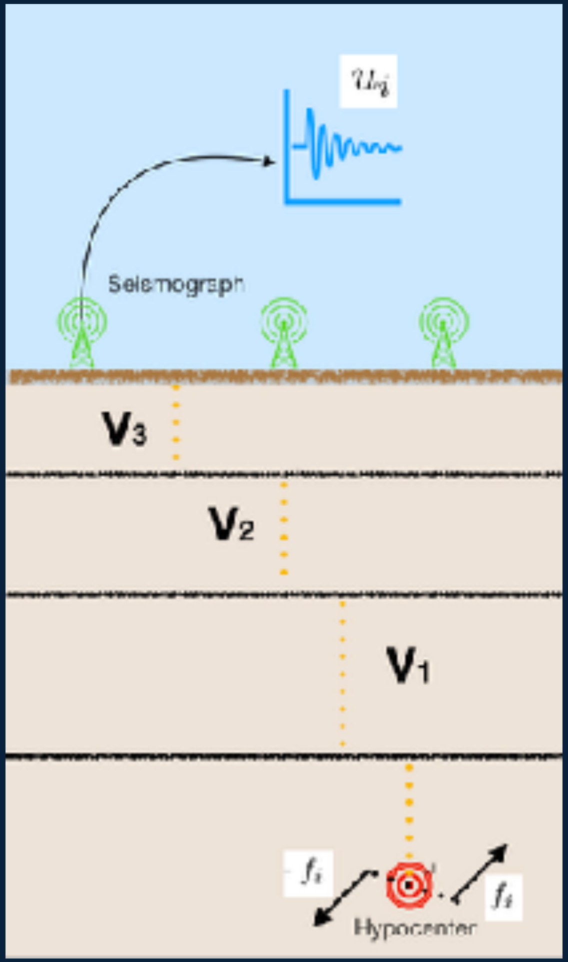
APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

$\mathbf{V}$ = velocity model

$\mathbf{m}$ = moment tensor



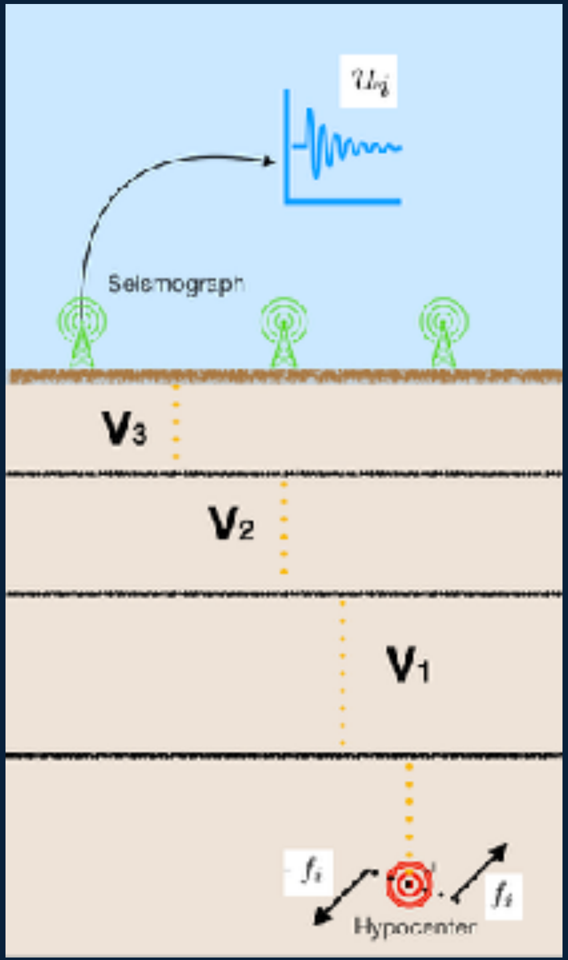
The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

$\mathbf{m}$ = moment tensor

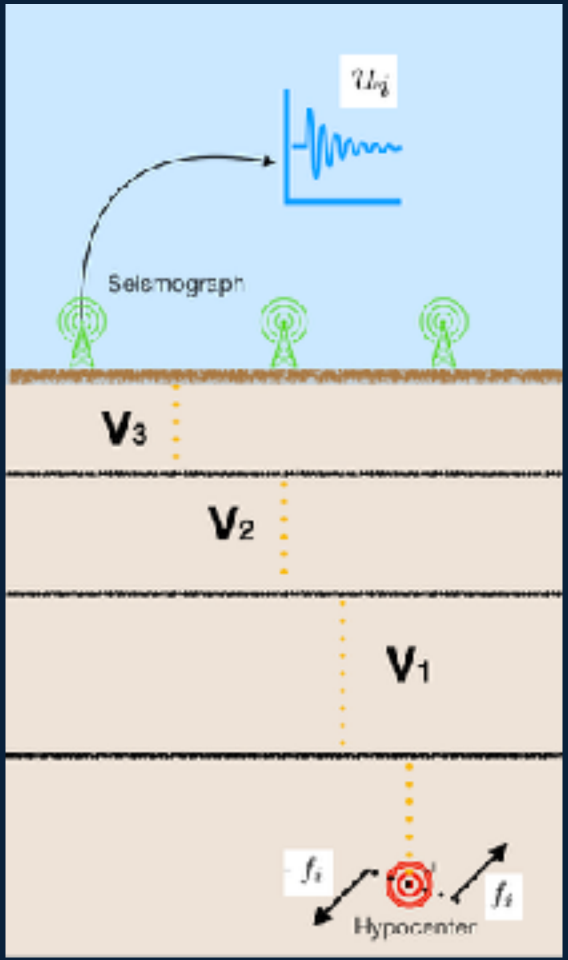


The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$



The problem

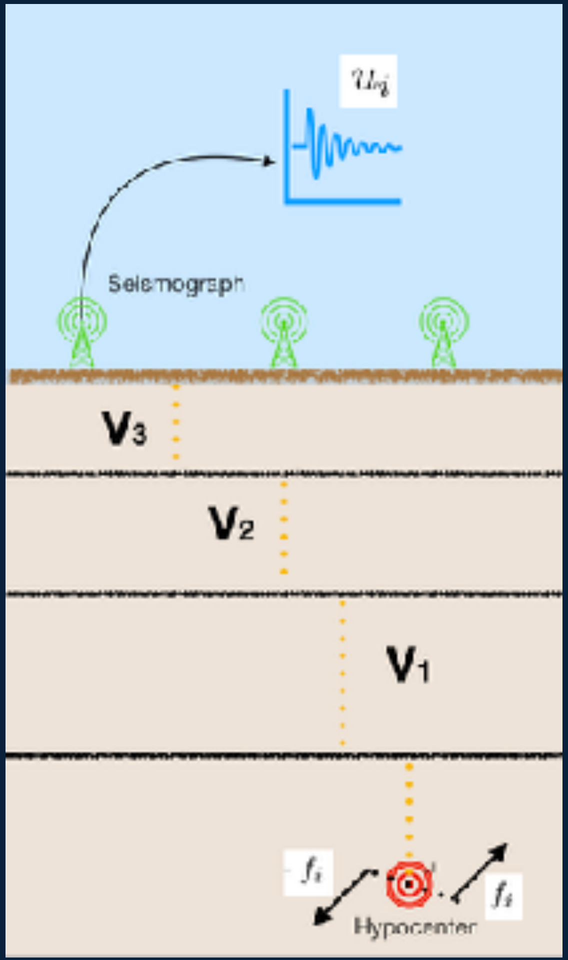
APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

DATA

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V}_0, \mathbf{x}_0, t) \cdot \mathbf{m}^T + \mathbf{e} \quad \text{with: } \mathbf{e} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$



The problem

APPLICATION: BAYESIAN MOMENT TENSOR INVERSION

INVERSE PROBLEM

MODEL

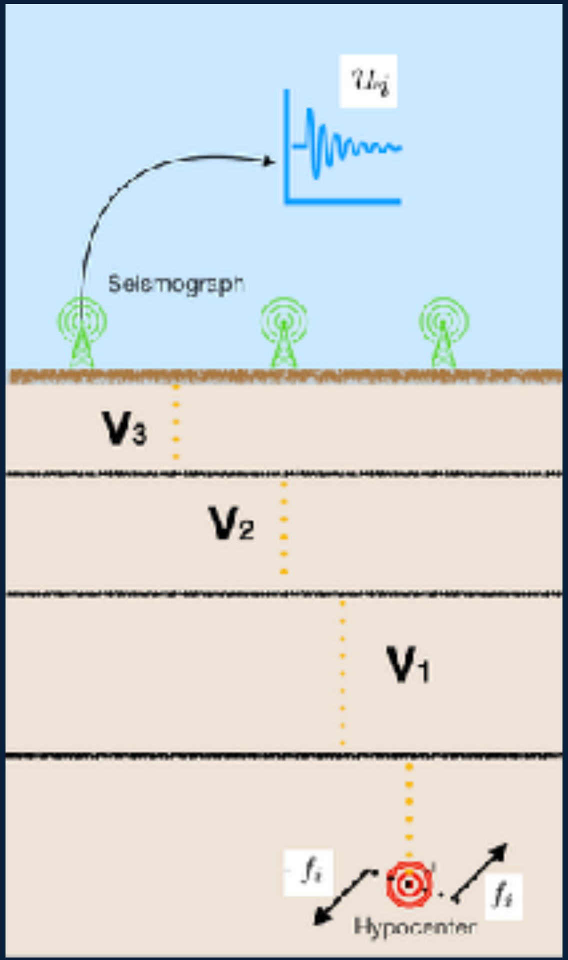
$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}, \mathbf{x}, t) \cdot \mathbf{m}^T$$

DATA

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V}_0, \mathbf{x}_0, t) \cdot \mathbf{m}^T + \mathbf{e} \quad \text{with: } \mathbf{e} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

QUANTITY OF INTEREST

$\mathbf{m} =$ moment tensor



A Bayesian framework

BAYESIAN INFERENCE

Full characterization of the uncertainty in $\mathbf{m}$

$$p(\mathbf{m} \,|\, \mathbf{y}) \propto p(\mathbf{y} \,|\, \mathbf{m}) \cdot p(\mathbf{m})$$

A Bayesian framework

BAYESIAN INFERENCE

Full characterization of the uncertainty in $\mathbf{m}$

$$p(\mathbf{m} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{m}) \cdot p(\mathbf{m})$$

PRIOR $p(\mathbf{m})$ probability distribution — encodes prior knowledge

A Bayesian framework

BAYESIAN INFERENCE

Full characterization of the uncertainty in $\mathbf{m}$

$$p(\mathbf{m}\,|\,\mathbf{y}) \propto p(\mathbf{y}\,|\,\mathbf{m}) \cdot p(\mathbf{m})$$

PRIOR $p(\mathbf{m})$ probability distribution — encodes prior knowledge

LIKELIHOOD $p(\mathbf{y}\,|\,\mathbf{m})$ function — a statistical model involving $\mathbf{u}(t)$

A Bayesian framework

BAYESIAN INFERENCE

Full characterization of the uncertainty in $\mathbf{m}$

$$p(\mathbf{m} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{m}) \cdot p(\mathbf{m})$$

PRIOR $p(\mathbf{m})$ probability distribution — encodes prior knowledge

LIKELIHOOD $p(\mathbf{y} \mid \mathbf{m})$ function — a statistical model involving $\mathbf{u}(t)$

POSTERIOR $p(\mathbf{m} \mid \mathbf{y})$ probability distribution — encodes data-updated knowledge

Misspecification

CAN WE ASSUME?

- $\mathbf{V}$ WELL-SPECIFIED $\quad \mathbf{V}_* = \mathbf{V}_0$
- $\mathbf{V}$ MISSPECIFIED $\quad \mathbf{V}_* \neq \mathbf{V}_0$

DATA

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V}_0, \mathbf{x}_0, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \text{with: } \mathbf{e} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

MODEL

$$\mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

Misspecification robust-likelihood

$\mathbf{V}$ MISSPECIFIED $\qquad$ Σ UNKNOWN

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V_0}, \mathbf{x_0}, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

LIKELIHOOD Not a natural way to express a Bayes update

Misspecification robust-likelihood

$\mathbf{V}$ MISSPECIFIED $\quad\quad$ Σ UNKNOWN

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V}_0, \mathbf{x}_0, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

LIKELIHOOD

Not a natural way to express a Bayes update

Use of Gibbs Posterior

$$p(\mathbf{m}, s \,|\, \mathbf{y}) \propto s^N \exp\left(-s\,\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)\right)$$

Misspecification robust-likelihood

$\mathbf{V}$ MISSPECIFIED

Σ UNKNOWN

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V}_0, \mathbf{x}_0, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

LIKELIHOOD

Not a natural way to express a Bayes update

Use of Gibbs Posterior

$$p(\mathbf{m}, s \,|\, \mathbf{y}) \propto s^N \exp\left(-s\,\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)\right)$$

$$\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)$$

Misspecification robust-likelihood

Earth Resources Laboratory

$\mathbf{V}$ MISSPECIFIED $\quad\quad$ $\boldsymbol{\Sigma}$ UNKNOWN

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V_0}, \mathbf{x_0}, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

LIKELIHOOD $\quad$ Not a natural way to express a Bayes update

Use of Gibbs Posterior

$$p(\mathbf{m}, s \,|\, \mathbf{y}) \propto s^N \exp\left(-s\,\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)\right)$$

$$\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)$$

Misspecification robust-likelihood

V MISSPECIFIED **Σ** UNKNOWN

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V_0}, \mathbf{x_0}, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V_*}, \mathbf{x_*}, t) \cdot \mathbf{m}^T$$

LIKELIHOOD Not a natural way to express a Bayes update

Use of Gibbs Posterior

$$p(\mathbf{m}, s \,|\, \mathbf{y}) \propto s^N \exp\left(-s\, \mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)\right)$$

$$\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right) \longrightarrow \ell_2^2$$

Misspecification robust-likelihood

$\mathbf{V}$ MISSPECIFIED $\quad\quad$ Σ UNKNOWN

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V_0}, \mathbf{x_0}, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

LIKELIHOOD Not a natural way to express a Bayes update

Use of Gibbs Posterior

$$p(\mathbf{m}, s \,|\, \mathbf{y}) \propto s^N \exp\left(-s\,\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)\right)$$

$$\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right) \to \ell_2^2$$

Misspecification robust-likelihood

$\mathbf{V}$ MISSPECIFIED **Σ UNKNOWN**

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{V_0}, \mathbf{x_0}, t) \cdot \mathbf{m}^T + \mathbf{e} \qquad \mathbf{u}(t) = \mathbf{G}(\mathbf{V}_*, \mathbf{x}_*, t) \cdot \mathbf{m}^T$$

LIKELIHOOD Not a natural way to express a Bayes update

Use of Gibbs Posterior

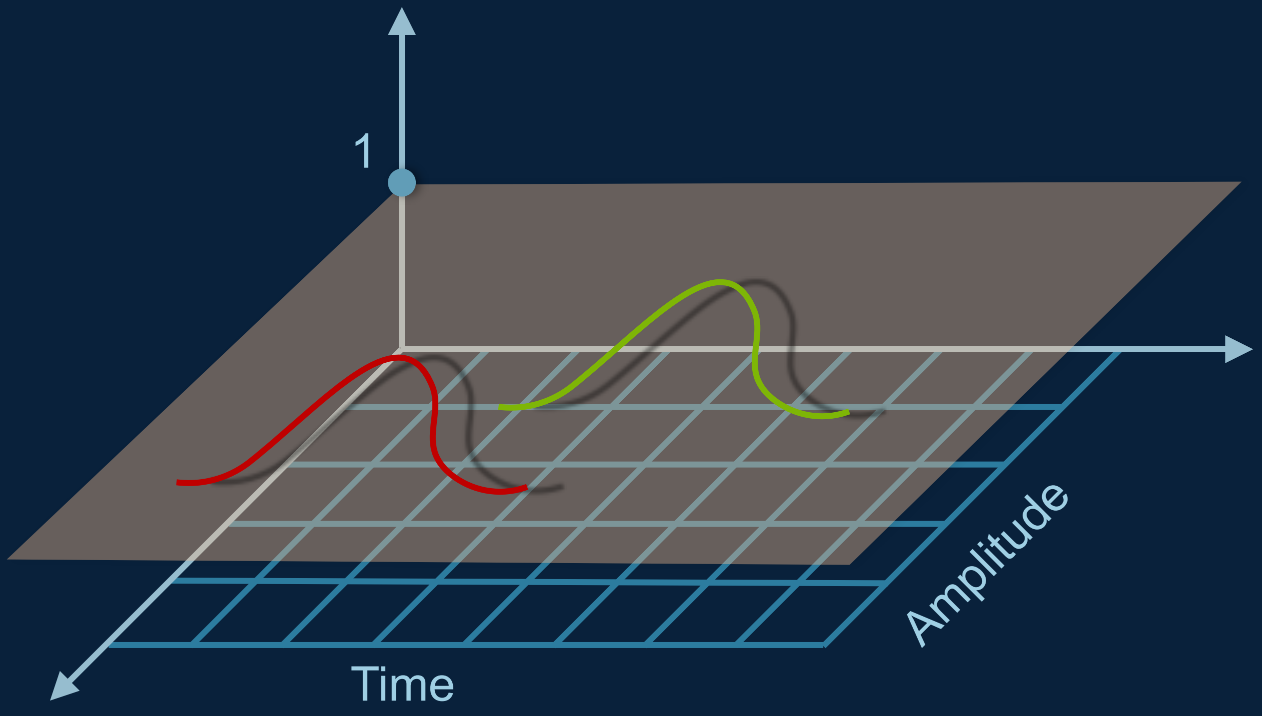
$$p(\mathbf{m}, s \,|\, \mathbf{y}) \propto s^N \exp\left(-s\,\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right)\right)$$

$$\mathscr{L}\left(\mathbf{y}(t), \mathbf{u}(t, \mathbf{m})\right) \to \ell_2^2, \quad TL_2^2$$

An alternative misfit measure

$$TL_2^2$$

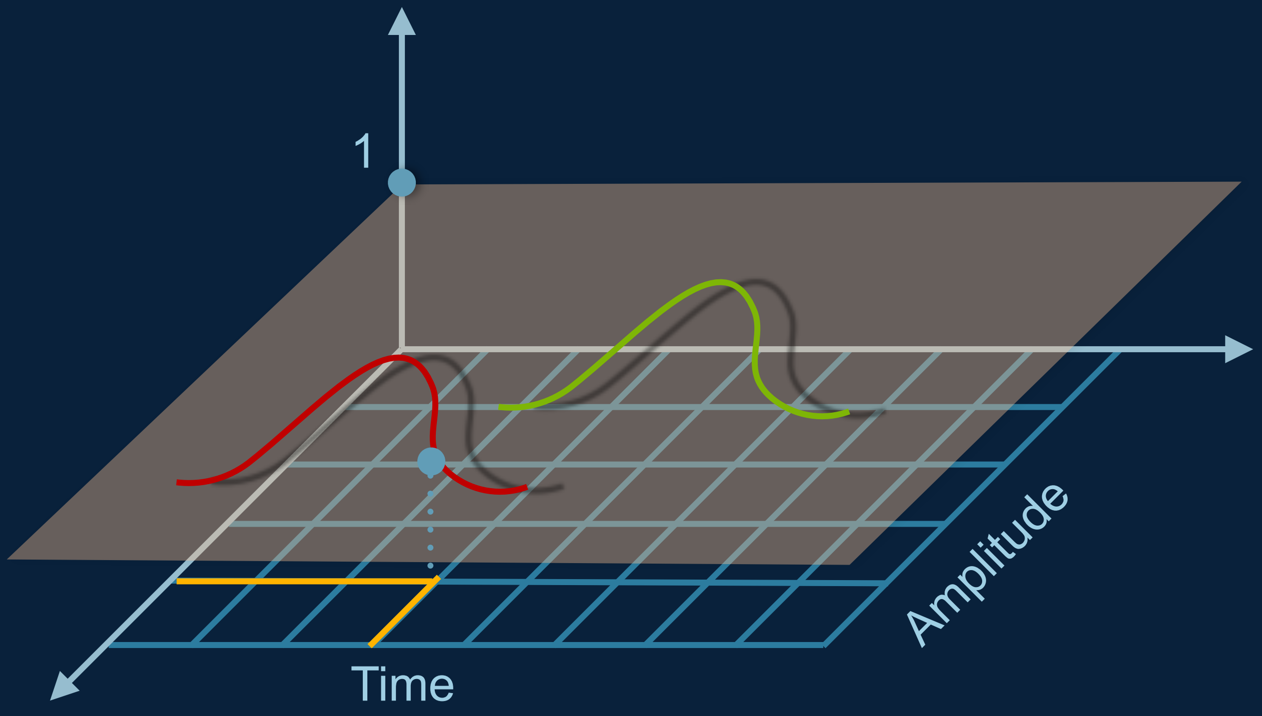
Distance between graphs of **u** and **y**



An alternative misfit measure

TL_2^2

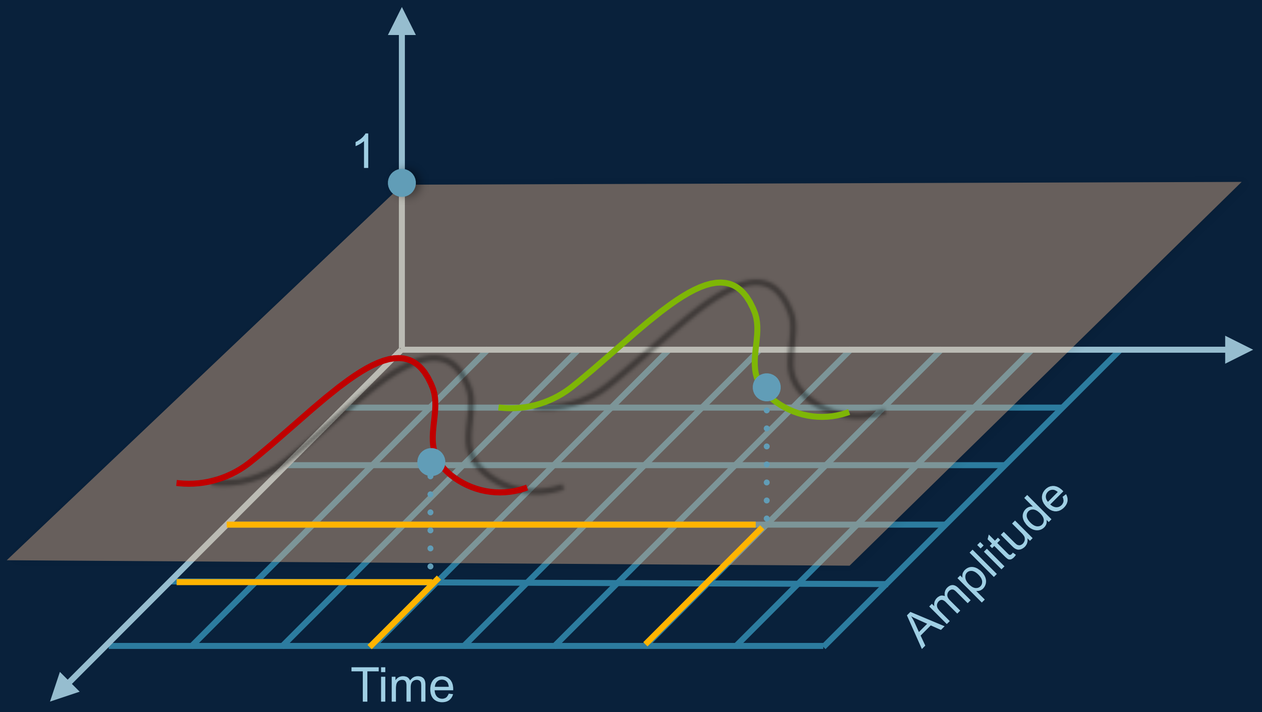
Distance between graphs of **u** and **y**



An alternative misfit measure

TL_2^2

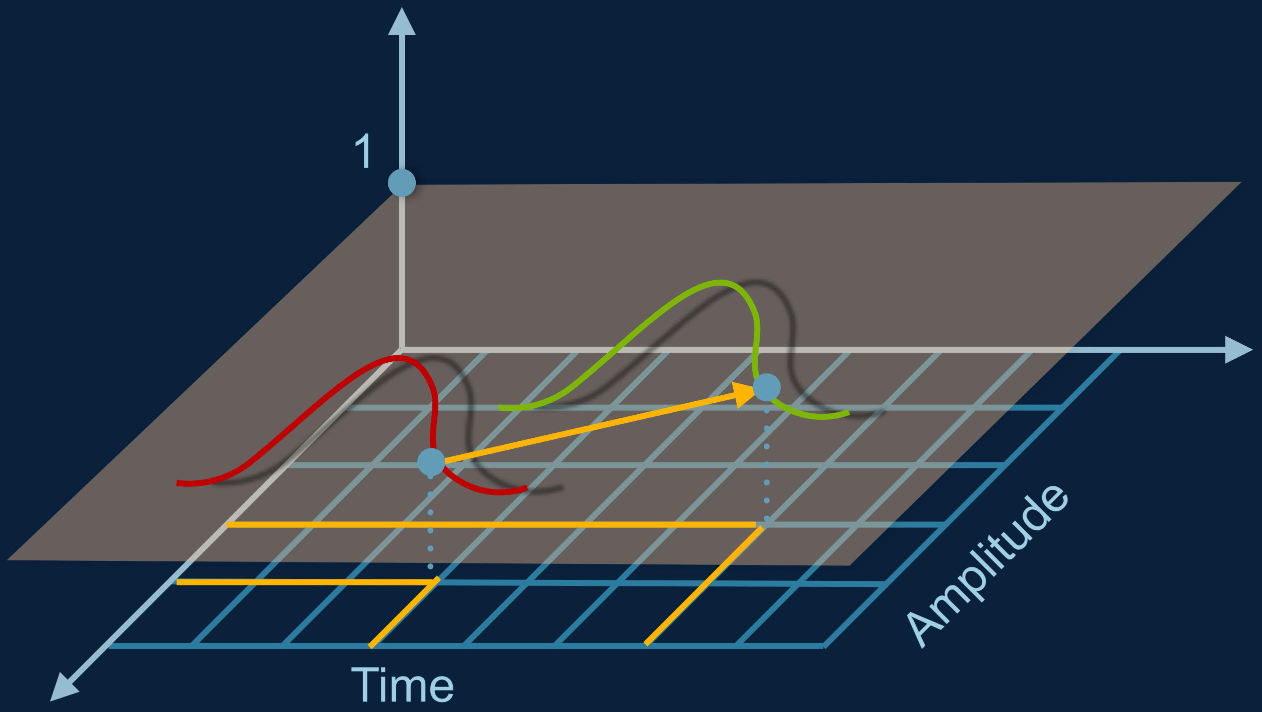
Distance between graphs of **u** and **y**



An alternative misfit measure

$$TL_2^2$$

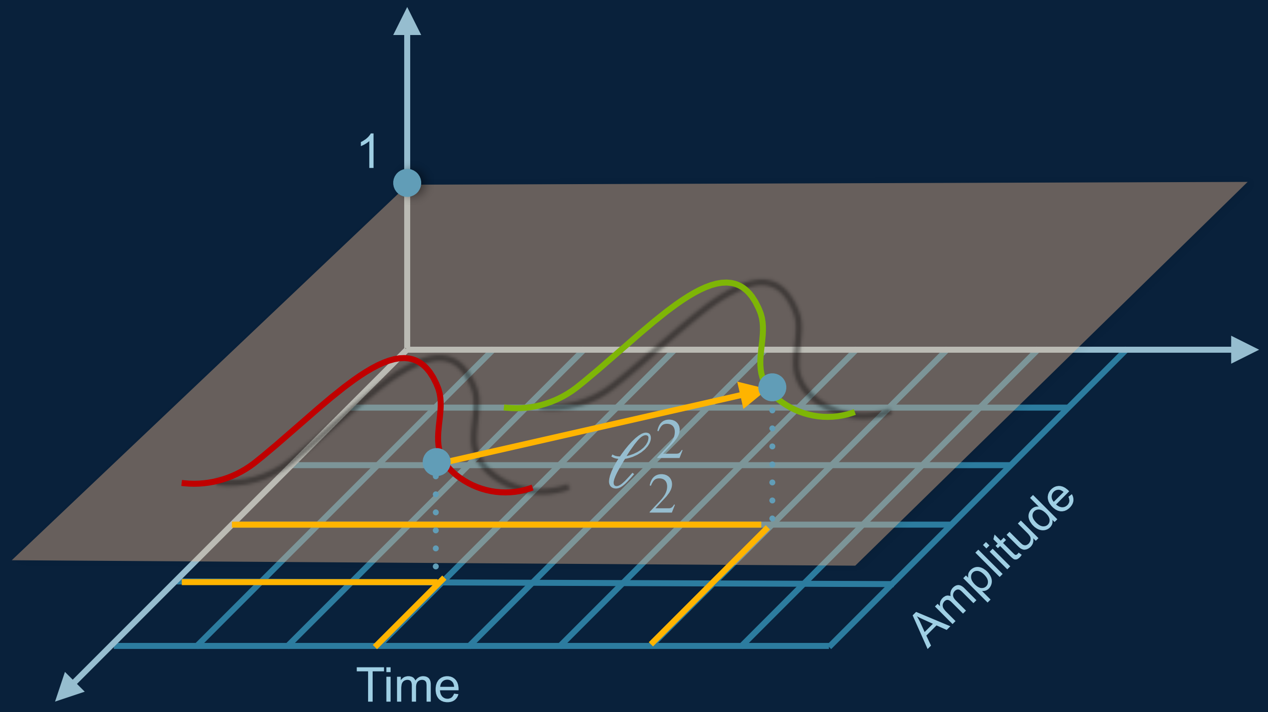
Distance between graphs of **u** and **y**



An alternative misfit measure

TL_2^2

Distance between graphs of **u** and **y**



One problem, many models

	V WELL-SPECIFIED	V MISSPECIFIED
Σ UNKNOWN		

One problem, many models

	V WELL-SPECIFIED	V MISSPECIFIED
Σ UNKNOWN		ℓ_2^2 Gibbs Posterior Hierarchical solution TL_2^2 Gibbs Posterior Hierarchical solution

One problem, many models

	V WELL-SPECIFIED		V MISSPECIFIED	
Σ UNKNOWN	ℓ_2^2	Gibbs Posterior Hierarchical solution	ℓ_2^2	Gibbs Posterior Hierarchical solution
	TL_2^2	Gibbs Posterior Hierarchical solution	TL_2^2	Gibbs Posterior Hierarchical solution

One problem, many models

	V WELL-SPECIFIED		V MISSPECIFIED	
Σ UNKNOWN	ℓ_2^2	Normal-Gamma Hierarchical solution	ℓ_2^2	Gibbs Posterior Hierarchical solution
	TL_2^2	Gibbs Posterior Hierarchical solution	TL_2^2	Gibbs Posterior Hierarchical solution

One problem, many models

UNCERTAINTY →

	V WELL-SPECIFIED	V MISSPECIFIED
Σ UNKNOWN	ℓ_2^2 Normal-Gamma Hierarchical solution	ℓ_2^2 Gibbs Posterior Hierarchical solution
	TL_2^2 Gibbs Posterior Hierarchical solution	TL_2^2 Gibbs Posterior Hierarchical solution

One problem, many models

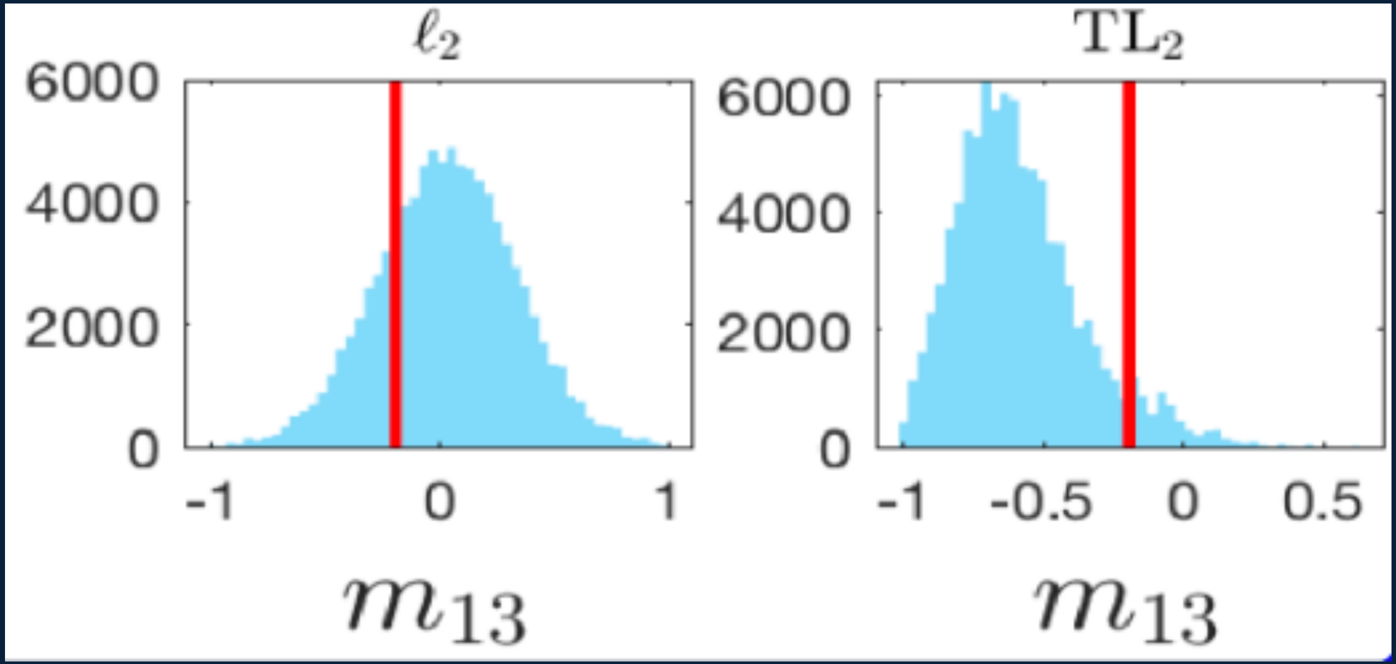
UNCERTAINTY →

	V WELL-SPECIFIED	V MISSPECIFIED
Σ UNKNOWN	ℓ_2^2 Normal-Gamma Hierarchical solution TL_2^2 Gibbs Posterior Hierarchical solution	ℓ_2^2 Gibbs Posterior Hierarchical solution TL_2^2 Gibbs Posterior Hierarchical solution

*HOW CAN WE **QUANTITATIVELY** COMPARE POSTERIORS COMING FROM ALL THESE DIFFERENT SETTINGS?*

One problem, many models

WHICH ONE IS BETTER?
IS IT EVEN WORTH TO USE TL_2 ?



Posterior scoring

EXPERIMENTAL SET-UP

Posterior scoring

EXPERIMENTAL SET-UP

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

Posterior scoring

EXPERIMENTAL SET-UP

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

(2) Generate $\mathbf{y}_{obs} \sim p(\mathbf{y} \,|\, \mathbf{m}_{true})$

Posterior scoring

EXPERIMENTAL SET-UP

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

(2) Generate $\mathbf{y}_{obs} \sim p(\mathbf{y} \mid \mathbf{m}_{true})$

(3) Calculate posterior $p(\mathbf{m} \mid \mathbf{y}_{obs})$

Posterior scoring

EXPERIMENTAL SET-UP

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

(2) Generate $\mathbf{y}_{obs} \sim p(\mathbf{y} \,|\, \mathbf{m}_{true})$

(3) Calculate posterior $p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

(4) Sample N $\mathbf{m}_i \sim p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

Posterior scoring

EXPERIMENTAL SET-UP

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

(2) Generate $\mathbf{y}_{obs} \sim p(\mathbf{y} \,|\, \mathbf{m}_{true})$

(3) Calculate posterior $p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

(4) Sample N $\mathbf{m}_i \sim p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

(5) Score posteriors

Posterior scoring

EXPERIMENTAL SET-UP

REPEAT L TIMES

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

(2) Generate $\mathbf{y}_{obs} \sim p(\mathbf{y} \,|\, \mathbf{m}_{true})$

(3) Calculate posterior $p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

(4) Sample N $\mathbf{m}_i \sim p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

(5) Score posteriors

Posterior scoring

EXPERIMENTAL SET-UP

REPEAT L TIMES

(1) Draw $\mathbf{m}_{true} \sim p(\mathbf{m})$

(2) Generate $\mathbf{y}_{obs} \sim p(\mathbf{y} \,|\, \mathbf{m}_{true})$

(3) Calculate posterior $p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

(4) Sample N $\mathbf{m}_i \sim p(\mathbf{m} \,|\, \mathbf{y}_{obs})$

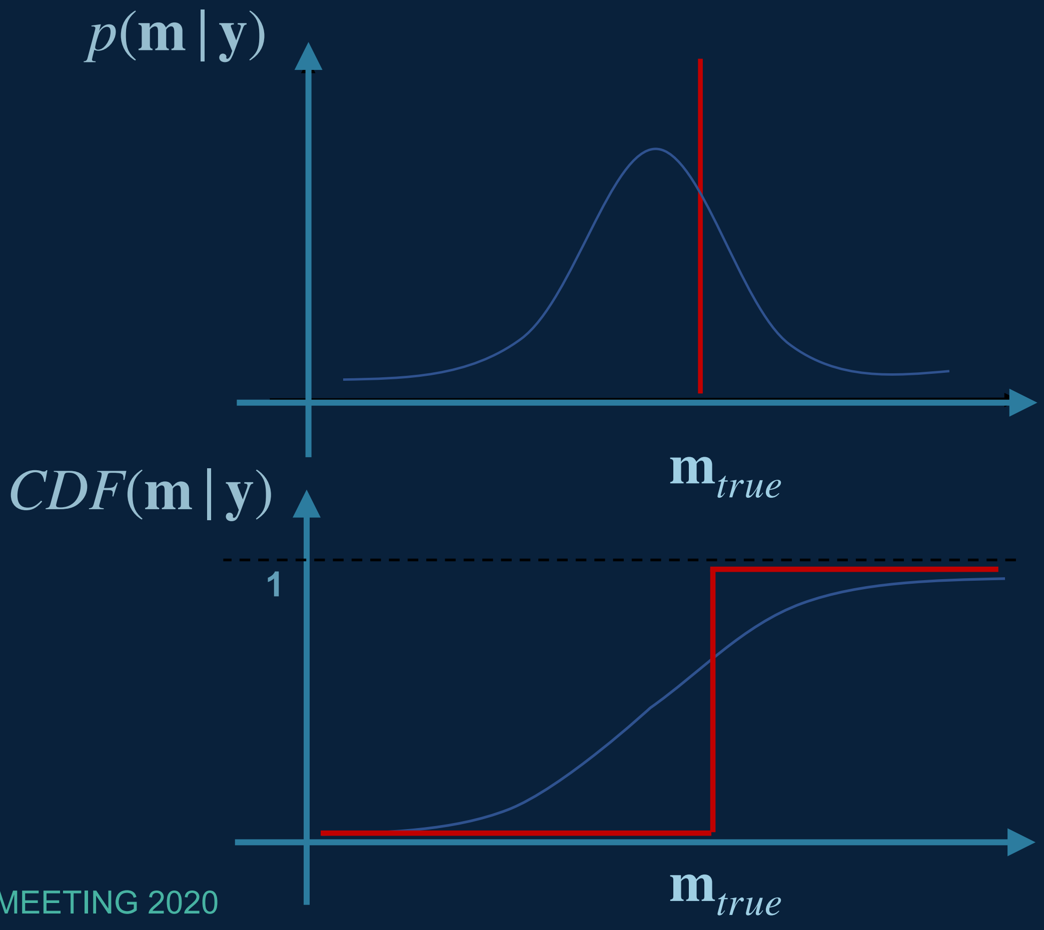
(5) Score posteriors

Report summaries for scores

One score

CRPS - CONTINUOUS RANKED PROBABILITY SCORE

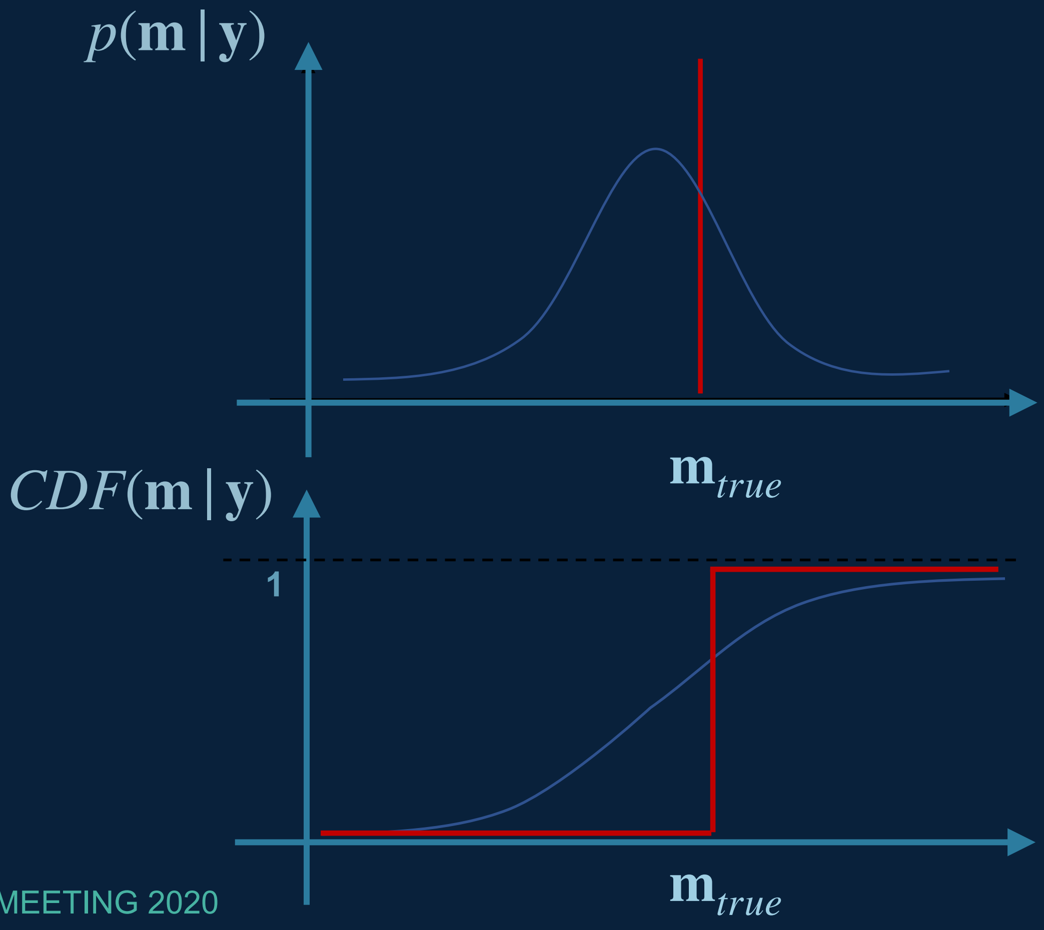
$$CRPS = \frac{1}{N}\sum_{i}^{N}\left(CDF(\mathbf{m}_i \mid \mathbf{y}_{obs}) - \mathbf{1}_{\mathbf{m}_i > \mathbf{m}_{true}}\right)^2$$



One score

CRPS - CONTINUOUS RANKED PROBABILITY SCORE

$$CRPS = \frac{1}{N}\sum_{i}^{N}\left(CDF(\mathbf{m}_i \,|\, \mathbf{y}_{obs}) - \mathbf{1}_{\mathbf{m}_i > \mathbf{m}_{true}}\right)^2$$



MEASURE OF FORECASTING CAPABILITY

-

THE LOWER THE BETTER

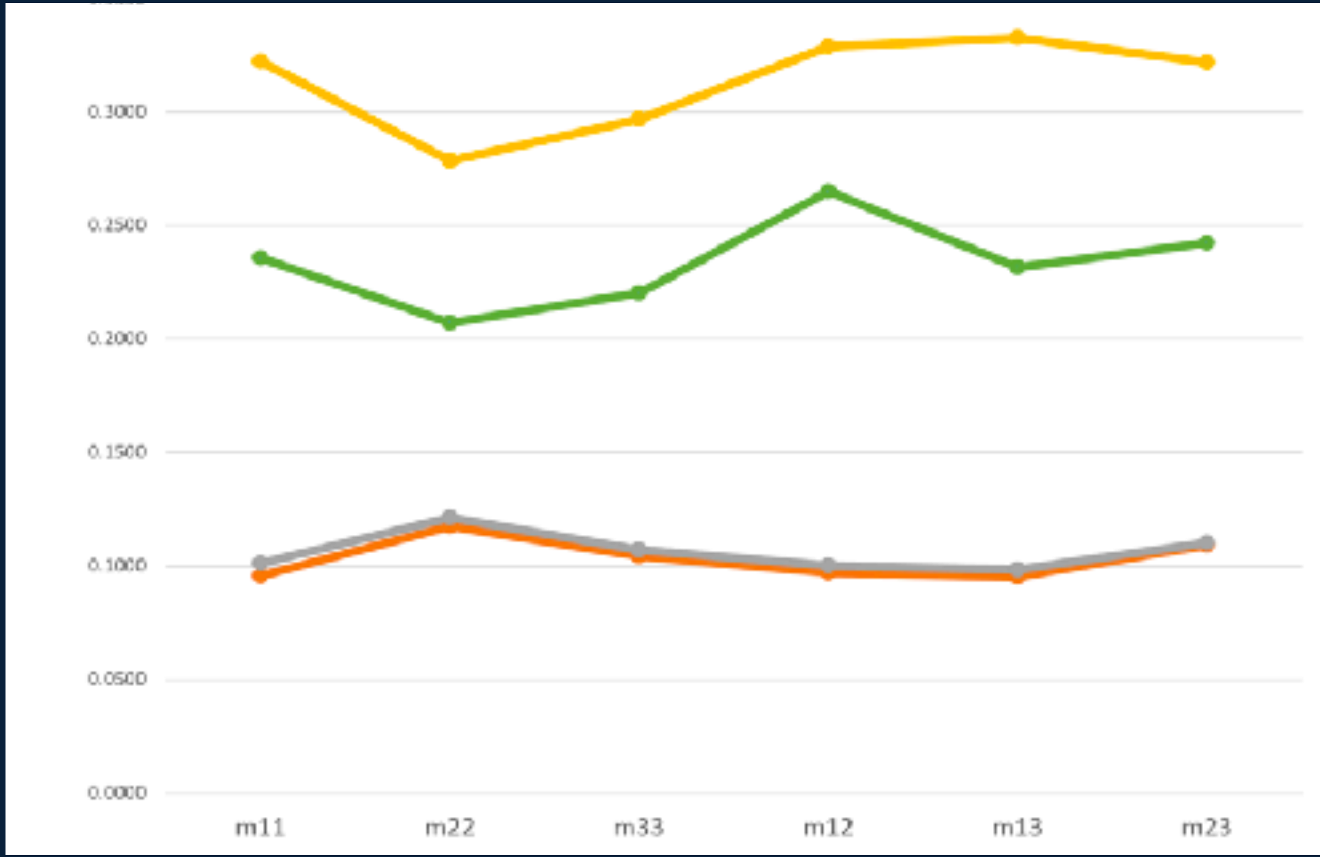
Results

MEAN CRPS SCORES

MS - GIBBS - L2

MS - GIBBS - TL

WS - HIER - TL
WS - HIER - L2



Results

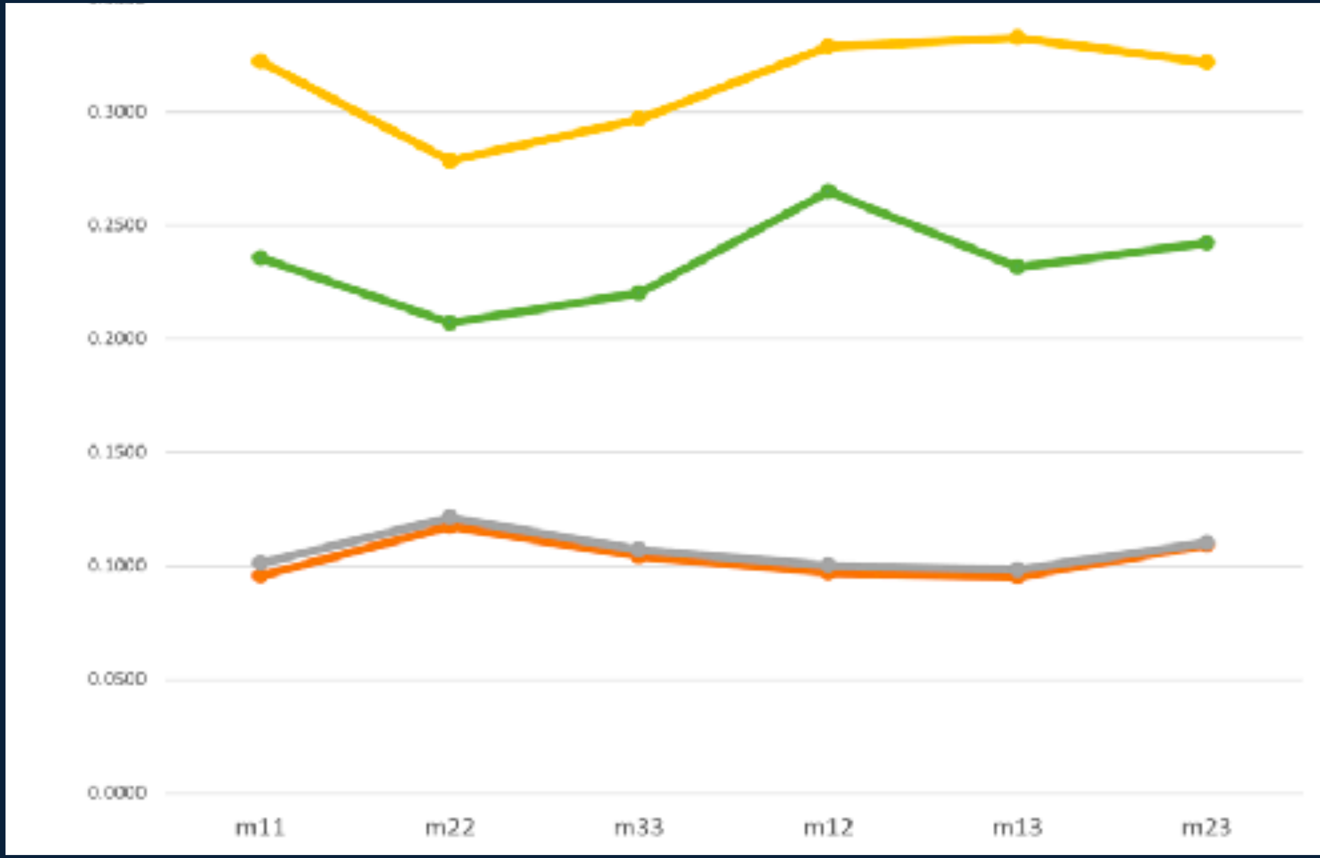
MEAN CRPS SCORES

MS - GIBBS - L2

MS - GIBBS - TL

WS - ANALYTICAL - L2
Σ KNOWN

WS - HIER - TL
WS - HIER - L2



Results

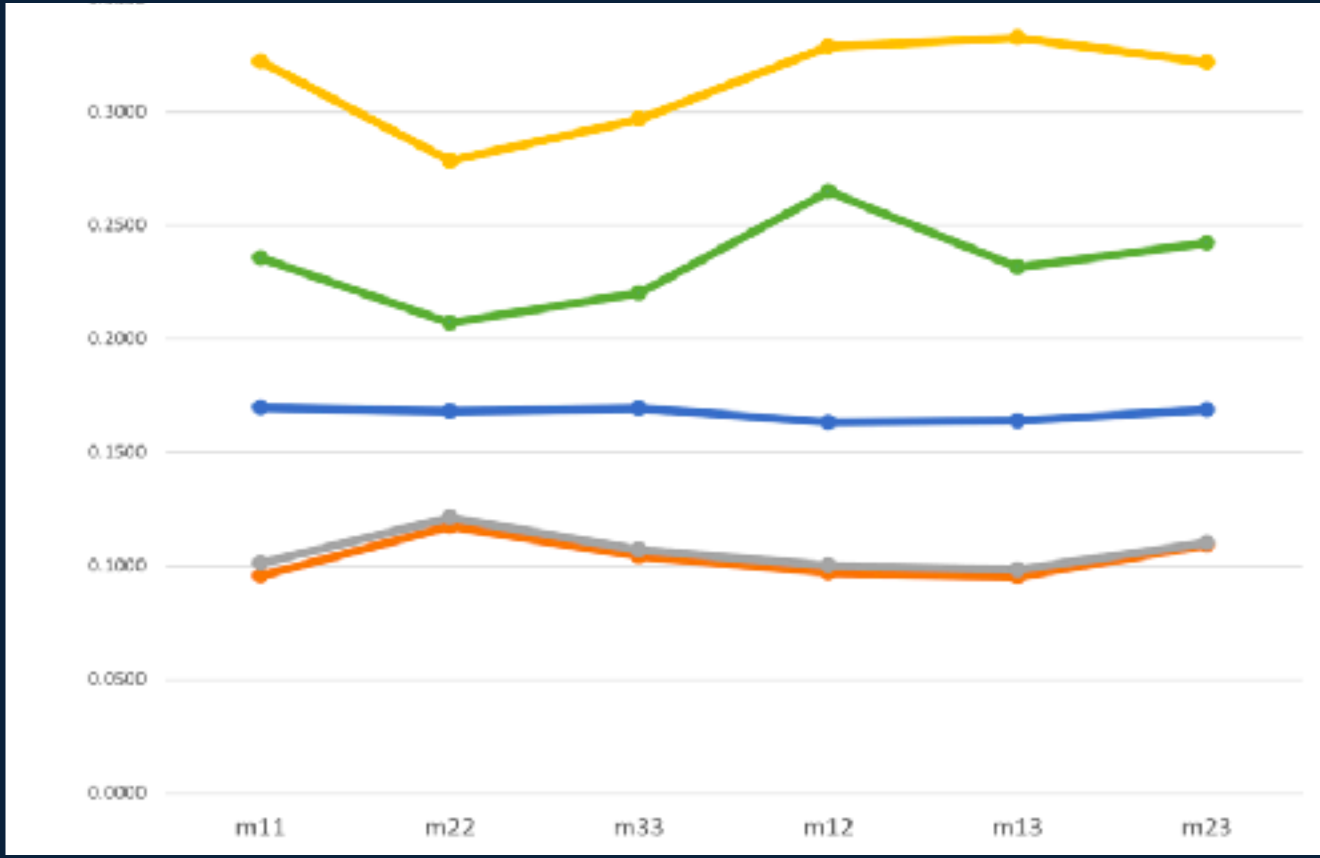
MEAN CRPS SCORES

MS - GIBBS - L2

MS - GIBBS - TL

WS - ANALYTICAL - L2
Σ KNOWN

WS - HIER - TL
WS - HIER - L2



Conclusions

1. We **quantitatively** proved that the TL_2-based likelihood provides better forecasters for different realizations of $\mathbf{m}_{true}$

2. We observed that a model with known noise level (less uncertainty) does not necessarily provide for a better forecaster

Ongoing work

More than a contradiction, a different purpose:

BETTER FORECASTER ⟷ STATISTICALLY CONSISTENT FRAMEWORK

MORE THAN…
"WHAT MAKES A GOOD POSTERIOR"

What makes a good posterior *for a given purpose*