Uncertainty quantification in CO_2 retrieval

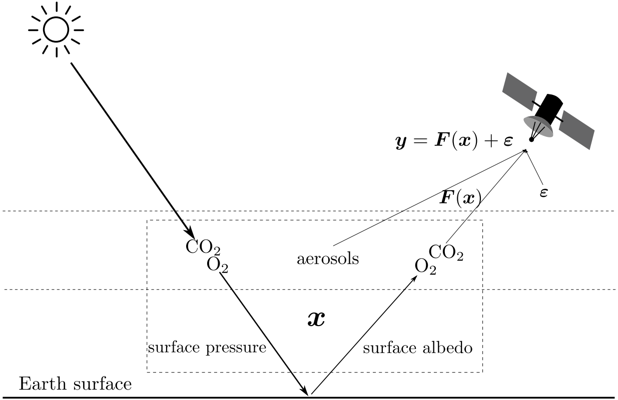
JPL Visit
October 2019

Pratik Patil, Mikael Kuusela, Jonathan Hobbs

Outlook

- Recent trend in statistics and machine learning:
  - Assumption-free/distribution-free inference
  - Safe inference
  - Robust inference
  - Agnostic inference
- Uncertainty quantification in inverse problems:
  - Different uncertainties: noise, state, model
  - Frequentist versus Bayesian measure of uncertainty
- This work on frequentist uncertainty quantification in CO_2 retrieval:
  - Potential undercoverage of operational retrieval confidence intervals
  - Ways of constructing confidence intervals using physical constraints
  - Ways of borrowing certainties from other sources or retrievals

CO_2 sensing system: general model



$\boldsymbol{x} \in \mathbb{R}^p$: state vector, $\boldsymbol{F} \in \mathbb{R}^p \to \mathbb{R}^n$: forward model,
$\boldsymbol{\varepsilon} \in \mathbb{R}^n$: instrument noise, $\boldsymbol{y} \in \mathbb{R}^n$: radiance observations
Quantity of interest: a functional of state vector $\theta(\boldsymbol{x}) \in \mathbb{R}$

CO_2 sensing system: approximated model [1]

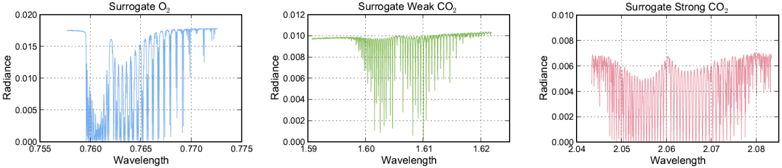
- state vector $\boldsymbol{x}$:
  - CO_2 profile (layer 1 to layer 20) [20 elements]
  - surface pressure [1 elements]
  - surface albedo [6 elements]
  - aerosols [12 elements]
- forward model $\boldsymbol{F}$:
  linearized with forward model Jacobian $\boldsymbol{K}(\boldsymbol{x}) = \frac{\partial \boldsymbol{F}(\boldsymbol{x})}{\partial \boldsymbol{x}}$
- noise $\boldsymbol{\varepsilon}$: normal approximation
- observations $\boldsymbol{y}$:
  discretized radiances in 3 near-infrared bands [1024 in each band]
  - O_2 A-band (around 0.76 microns)
  - weak CO_2 band (around 1.61 microns)
  - strong CO_2 band (around 2.06 microns)

[1] provided by Jon Hobbs [Hobbs et al., SIAM/ASA Journal on Uncertainty Quantification, 2017]

Question of interest

Input:

- radiance observations $\boldsymbol{y}$

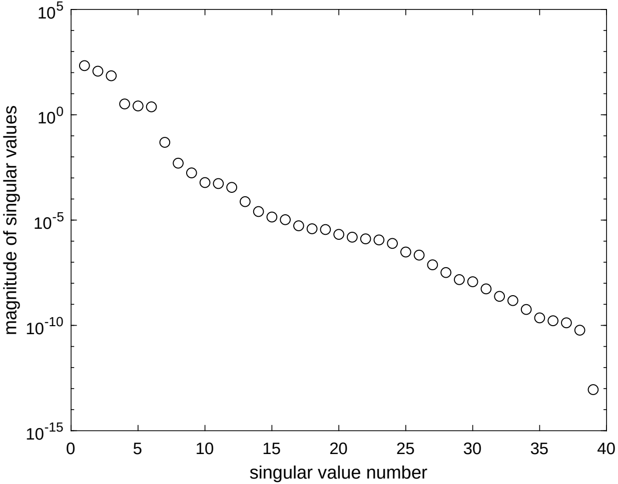


- an approximated model $\boldsymbol{y} \approx \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\varepsilon}$

Output:

- confidence interval $\left[\underline{\theta}, \overline{\theta}\right]$ for a functional $\theta(\boldsymbol{x})$ of the form $\boldsymbol{h}^T\boldsymbol{x}$ that measures column averaged CO_2 with frequentist coverage $\mathbb{P}_{\boldsymbol{\varepsilon}}(\theta \in \left[\underline{\theta}, \overline{\theta}\right]) \approx 1 - \alpha$ for any $\boldsymbol{x}$

Ill-posed inverse problem



Inverse problem severely ill-posed with exponential singular values decay
Lowest eigenvalue numerically zero leading to null space directions

Operational retrieval: outline

Key idea: let prior on $\boldsymbol{x}$ regularize the problem (Bayesian procedure)

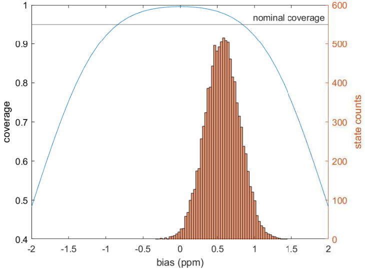
- Assume prior distribution on $p(\boldsymbol{x})$
- Combine prior with likelihood from forward model $\boldsymbol{F}(\boldsymbol{x})$ using observations $\boldsymbol{y}$ to get posterior $p(\boldsymbol{x}|\boldsymbol{y})$
- Compute MAP estimator $\hat{\boldsymbol{x}}$ maximizing $p(\boldsymbol{x}|\boldsymbol{y})$
- Use plug-in estimate as $\hat{\theta} = \theta(\hat{\boldsymbol{x}})$
- From the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$, estimate covariance $\hat{\boldsymbol{\Sigma}}$ of $\hat{\boldsymbol{x}}$
- Use plug-in estimate for posterior variance $\hat{\sigma}^2$ as $\sigma^2(\hat{\boldsymbol{\Sigma}})$
- Set the $(1-\alpha)$ credible interval as $\left[\hat{\theta} - z_{\alpha/2}\hat{\sigma}, \hat{\theta} + z_{\alpha/2}\hat{\sigma}\right]$

Potential issues: bias and undercoverage
The true uncertainty could be drastically underestimated!

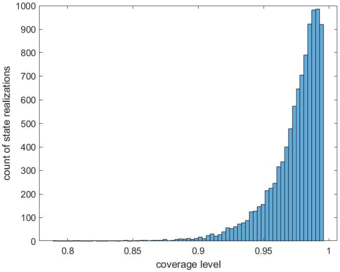
Operational retrieval: details

- Prior: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$
- Estimate: $\hat{\theta} = \boldsymbol{h}^T (\boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{\Sigma}_a^{-1})^{-1} (\boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{y} + \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\mu}_a)$
- Posterior variance: $\hat{\sigma}^2 = \boldsymbol{h}^T (\boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{\Sigma}_a^{-1})^{-1} \boldsymbol{h}$
- Standard error: $\mathrm{se}(\hat{\theta}) = \sqrt{\boldsymbol{c}^T \boldsymbol{\Sigma}_\varepsilon \boldsymbol{c}}$ for $\boldsymbol{c}^T = \boldsymbol{h}^T (\boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{\Sigma}_a^{-1})^{-1} \boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1}$
- Bias: $\mathrm{bias}(\hat{\theta}) = \boldsymbol{m}^T (\boldsymbol{x} - \boldsymbol{\mu}_a)$ for $\boldsymbol{m} = \left(\boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{K} (\boldsymbol{K}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{\Sigma}_a^{-1})^{-1} - \boldsymbol{I} \right) \boldsymbol{h}$
- Coverage: $\mathbb{P}_{\boldsymbol{\varepsilon}}(\theta \in [\underline{\theta}, \overline{\theta}]) = \Phi\left(-\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{se}(\hat{\theta})} + z_{1-\alpha/2} \frac{\hat{\sigma}}{\mathrm{se}(\hat{\theta})} \right) - \Phi\left(-\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{se}(\hat{\theta})} - z_{1-\alpha/2} \frac{\hat{\sigma}}{\mathrm{se}(\hat{\theta})} \right)$
- Length: $2 z_{1-\alpha/2} \hat{\sigma}$

Operational retrieval: bias and coverage distributions



(a) bias distribution



(b) coverage distribution

Minimum coverage: 78.9%
Fraction of cases below nominal coverage: 12.03%

Issues with operational retrieval: single sounding

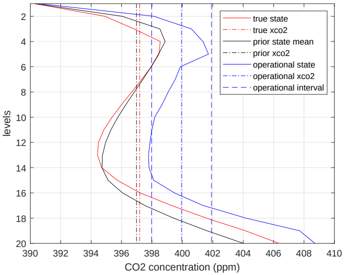
Coverage for some single soundings at Lamont, OK

x realization	operational bias	operational coverage
1	1.417	0.789
2	1.370	0.809
3	1.303	0.834
4	1.235	0.857
5	1.164	0.880
6	1.079	0.903
7	0.978	0.926
8	0.842	0.950
9	0.659	0.972
10	0.000	0.995

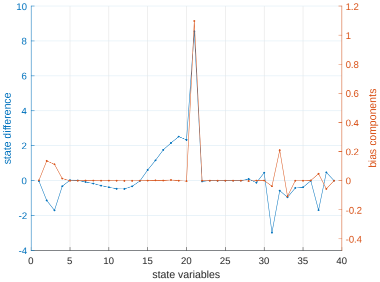
The lowest coverage sometimes drops even below 50%.

Operational retrieval illustration: single sounding

Minimum coverage:



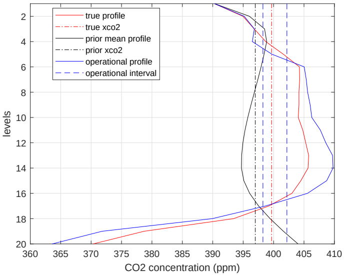
(a) profile comparisons



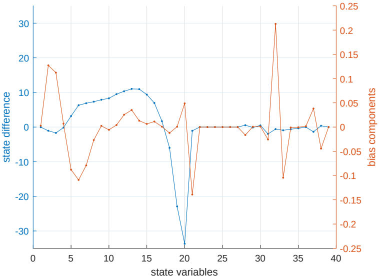
(b) bias components

Operational retrieval illustration: single sounding

Maximum coverage:



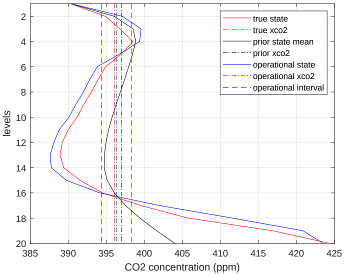
(a) profile comparisons



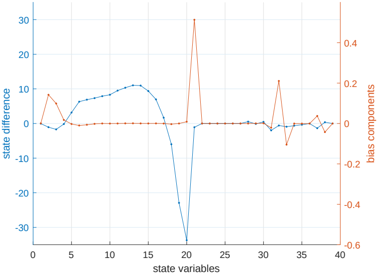
(b) bias components

Operational retrieval illustration: single sounding

Nominal coverage:

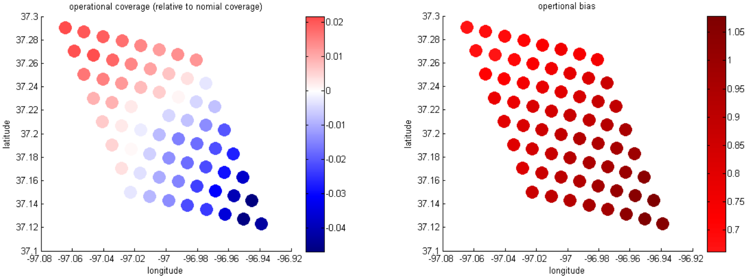


(a) profile comparisons



(b) bias components

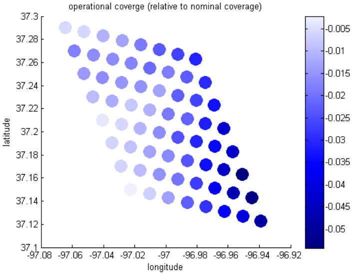
Issues with operational retrieval: grid sounding



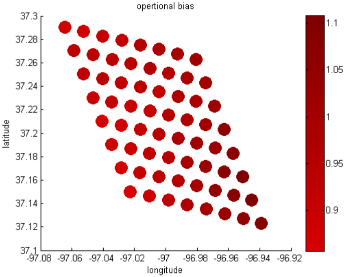
(a) coverage pattern (b) bias pattern

Fraction of soundings below nominal coverage: 0.55

Issues with operational retrieval: grid sounding



(a) coverage pattern

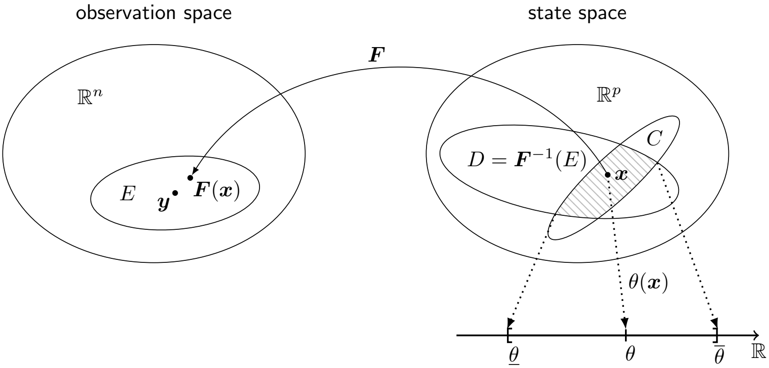


(b) bias pattern

Fraction of soundings below nominal coverage: 1

Proposed retrieval: version 1

Key idea 1: let actual physical constraints regularize the problem[2]



$$\theta = \theta(\boldsymbol{x}), \quad \underline{\theta} = \min_{\boldsymbol{x} \in C \cap D} \theta(\boldsymbol{x}), \quad \overline{\theta} = \max_{\boldsymbol{x} \in C \cap D} \theta(\boldsymbol{x})$$

[2] Stark, Journal of Geophysical Research, 1992; Kuusela and Stark, Annals of Applied Statistics, 2017

Proposed retrieval: version 2

- Version 1 is working harder than it needs to. The interval $[\underline{\theta}, \overline{\theta}]$ has correct finite-sample coverage for any functional θ. But we only care about a particular functional.
- Key idea 2: only require the procedure to satisfy *one-at-time coverage* rather than *simultaneous* coverage[3]
- One way is to restrict the set D in version 1 that still preserves the coverage guarantee for θ. For example, assume Gaussian white noise for simplicity. Then,
  - version 1 uses $D = \{\boldsymbol{x} : \|\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\|^2 \leq \chi^2_n(\alpha)\}$ which has $(1-\alpha)$ coverage set in the state space.
  - version 2 restricts it such that $D' = \{\boldsymbol{x} : \|\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\|^2 \leq z^2_{\alpha/2} + b^2\}$, where $b = \min_{\boldsymbol{x} \in C} \|\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\|$

[3]inspired by Leary and Rust, SIAM Journal on Scientific and Statistical Computing, 1986

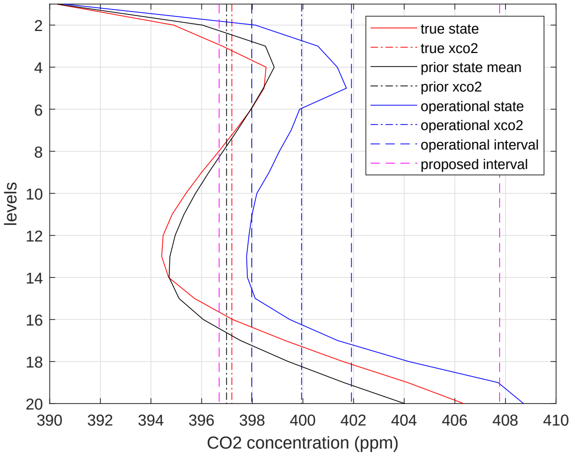
Improvements from proposed retrieval: single sounding

x realization	operational coverage	proposed coverage
1	0.789	0.951
2	0.809	0.952
3	0.834	0.952
4	0.857	0.951
5	0.880	0.951
6	0.903	0.951
7	0.926	0.950
8	0.950	0.951
9	0.972	0.952
10	0.995	0.951

Length of operational interval about 4, proposed interval about 11.

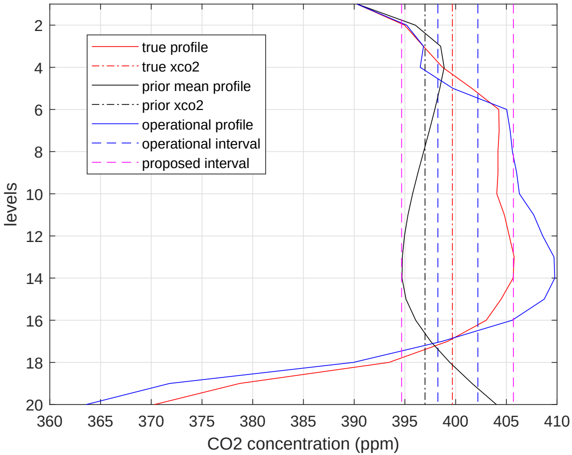
Proposed retrieval illustration: single sounding

Minimum coverage:



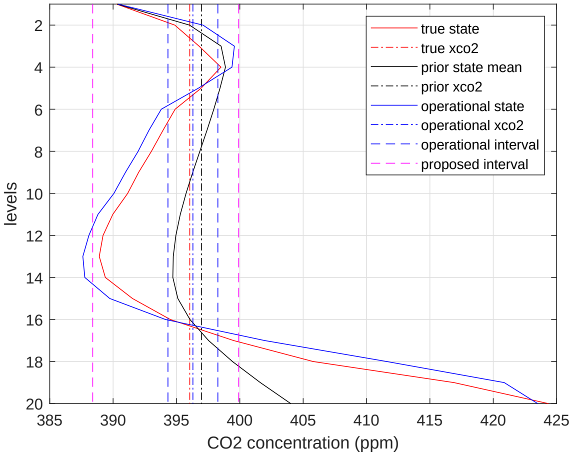
Proposed retrieval illustration: single sounding

Maximum coverage:

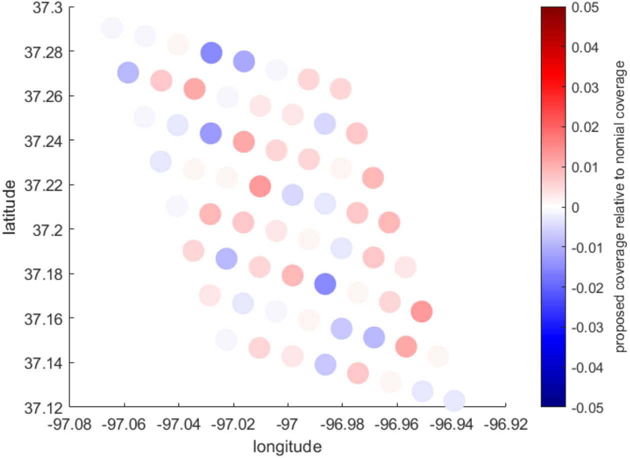


Proposed retrieval illustration: single sounding

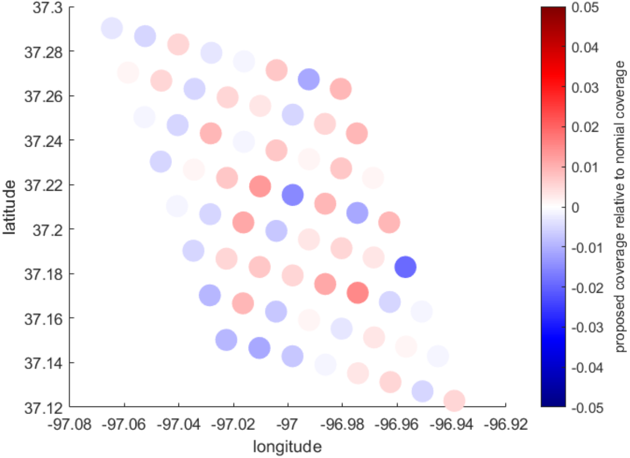
Nominal coverage:



Improvements from proposed retrieval: grid sounding



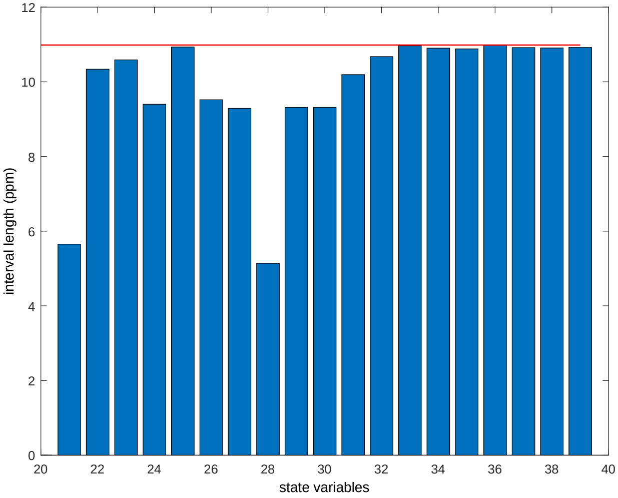
Improvements from proposed retrieval: grid sounding



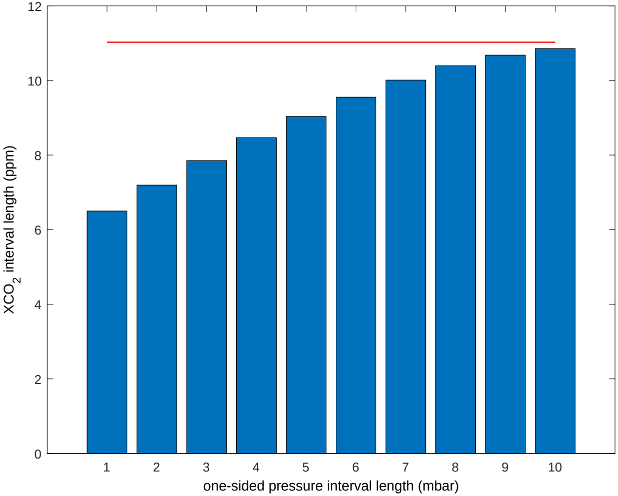
Proposed retrieval: version 3

- So far, we only used actual physical constraints on the state vector.
- But, what if we wanted to incorporate more information about state.
  - Certain ranges for some elements of state vector more likely.
  - Possibility of borrowing certainty from other sources.
- Version 3 provides a framework for incorporating additional probabilistic information and still maintaining finite-sample coverage guarantees. As an example, consider the following:
  - Individual state uncertainties
    $\mathbb{P}_{\boldsymbol{\varepsilon}}(x_i \notin \left[\underline{x_i}(\alpha_i), \overline{x_i}(\alpha_i)\right]) \leq \alpha_i$
  - Internal coverage:
    $\mathbb{P}_{\boldsymbol{\varepsilon}}(\theta \notin \left[\underline{\theta}, \overline{\theta}\right], \underline{x_i}(\alpha_i) \leq x_i \leq \overline{x_i}(\alpha_i)) \leq \gamma$
  - Final coverage:
    $\mathbb{P}_{\boldsymbol{\varepsilon}}(\theta \notin \left[\underline{\theta}, \overline{\theta}\right]) \leq \gamma + \sum_i \alpha_i$

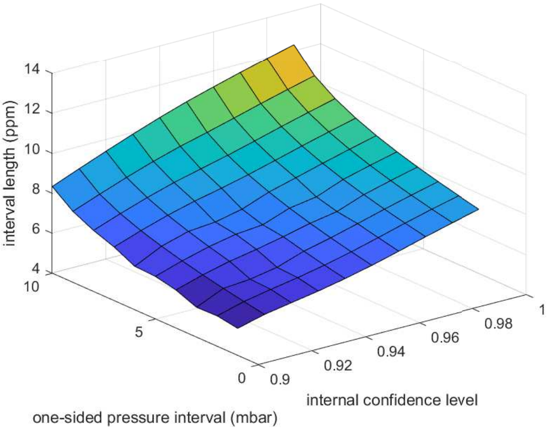
Deterministic exact information on individual elements



Deterministic range for pressure

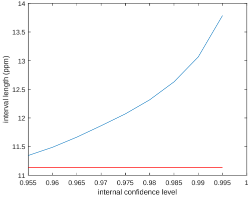


Probabilistic range for pressure

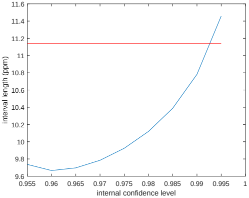


Probabilistic range for pressure

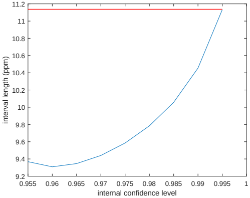
Name of the game:
Tradeoff γ and α_i keeping $\gamma + \sum_i \alpha_i \leq \alpha$ to make lengths smaller.



(c) prior



(d) intermediate



(e) truth

Conclusions and extensions

- Take-away 1: Some evidence of potential bias and undercoverage for the operational retrieval
- Take-away 2: Approach using only physical constraints to provide good coverage guarantees with reasonable confidence interval sizes
- Take-away 3: Further improvements in the size of intervals from the proposed retrieval possible using additional information
- Many extensions possible:
  - Different ways of restricting the sets for one-at-a-time intervals
  - Optimality for the size of the intervals
  - Combining information from different missions
  - Different approaches for non-linear forward models
  - Using intervals for downstream tasks instead of point estimates
  - General framework to combine different types of prior informations