

# HMCTomo

A didactic polyglot framework for sampling high-dimensional seismicological Bayesian inverse problems

Lars Gebraad<sup>1</sup> (lars.gebraad@erdw.ethz.ch), Andrea Zunino<sup>2</sup> (zunino@nbi.dk),  
Andreas Fichtner<sup>1</sup>, Klaus Mosegaard<sup>2</sup>

<sup>1</sup> Institute of Geophysics, ETH Zürich.

<sup>2</sup> Niels Bohr Institute, University of Copenhagen.

## Executive summary

We present a **work in progress pedagogical** framework for sampling high-dimensional Bayesian inverse problems implemented in Julia and Python. The modular framework allows the user to mix-and-match various priors, inverse problems and sampling algorithms, even across the two programming languages used. This, combined with various tutorials and well-documented and annotated code, embodies the purpose of the framework; to illustrate and facilitate large-scale Bayesian inference of seismicological inverse problems using modern gradient based sampling algorithms. We illustrate sampling for two typical seismicological inverse problems.

## What does HMCTomo do?

HMCTomo samples large seismicological inverse problems with appropriate Monte Carlo algorithms and explains how these works. This is essential for non-linear uncertainty quantification of computationally expensive models geoscientists frequently use.

- × Sampling Bayesian inverse problems with a **large parameter space**.
- × Implements **gradient based sampling** algorithms (proven efficiency for seismicological inverse problems: [1], [2], [3], [4], [5]) with extensive documentation in Python and Julia.
- × **Various priors and inverse problems** relevant to seismology.
- × **Works across scales**, from laptops to computing clusters.
- × Provides visualization and analysis tools.
- × **Mix-and-match**: use any prior and any inverse problem with any algorithm.

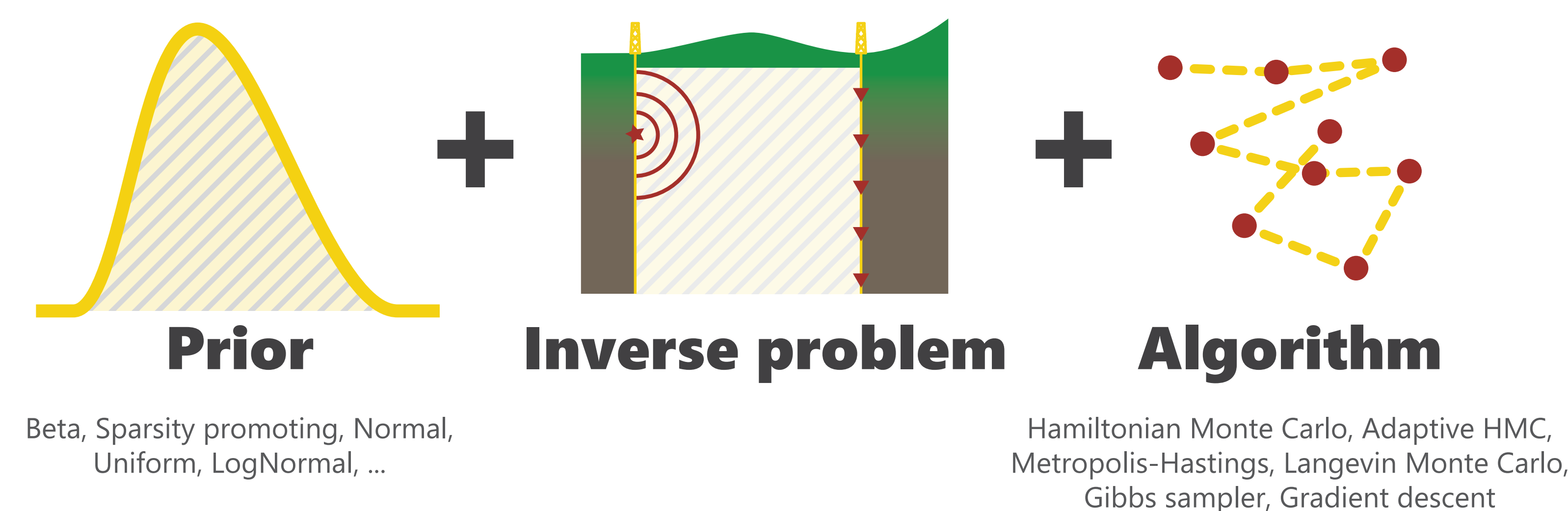


Figure 1: Diagram of possible mix-and-match combinations

## How is this different from ... ?

HMCTomo differs from existing libraries because it **specifically caters to geoscientists** who wish to tackle large seismicological Bayesian inverse problems.

- × In contrast to existing tools (e.g. [6], [7], [8]), HMCTomo focuses on seismicological inverse problems.
- × HMCTomo supplies relevant priors, visualizations and inverse problems.
- × The language used for HMCTomo's functions is similar to that of Bayesian inference.
- × HMCTomo has useful defaults and autotuners, eliminating some need for experience in tuning MCMC algorithms.

## Do you have some interesting examples?

Inverse problems that are available to the user include: **traveltime tomography** (eikonal equation) in 2D & 3D with Cartesian spherical coordinates, **acoustic full-waveform** in 2D, **AVA reflection seismic data** including rock physics models in 3D, **Earthquake hypocenter location**, **sparse quadratic functionals**, **magnetic anomaly problems** 2D and 3D.

### Example 1: Earthquake hypocenter location

- × Relatively simple seismicological inverse problem.
- × **Non-linear forward model**, synthetic data.
- × Requires sampling for true uncertainty quantification.
- × Sampling many events gives lower posterior dispersion.
- × Non-gradient based MCMC methods scale poorly with this.
- × HMC converges quickly for 100 events (**Figure 2**).
- × Metropolis-Hastings breaks down at this point.

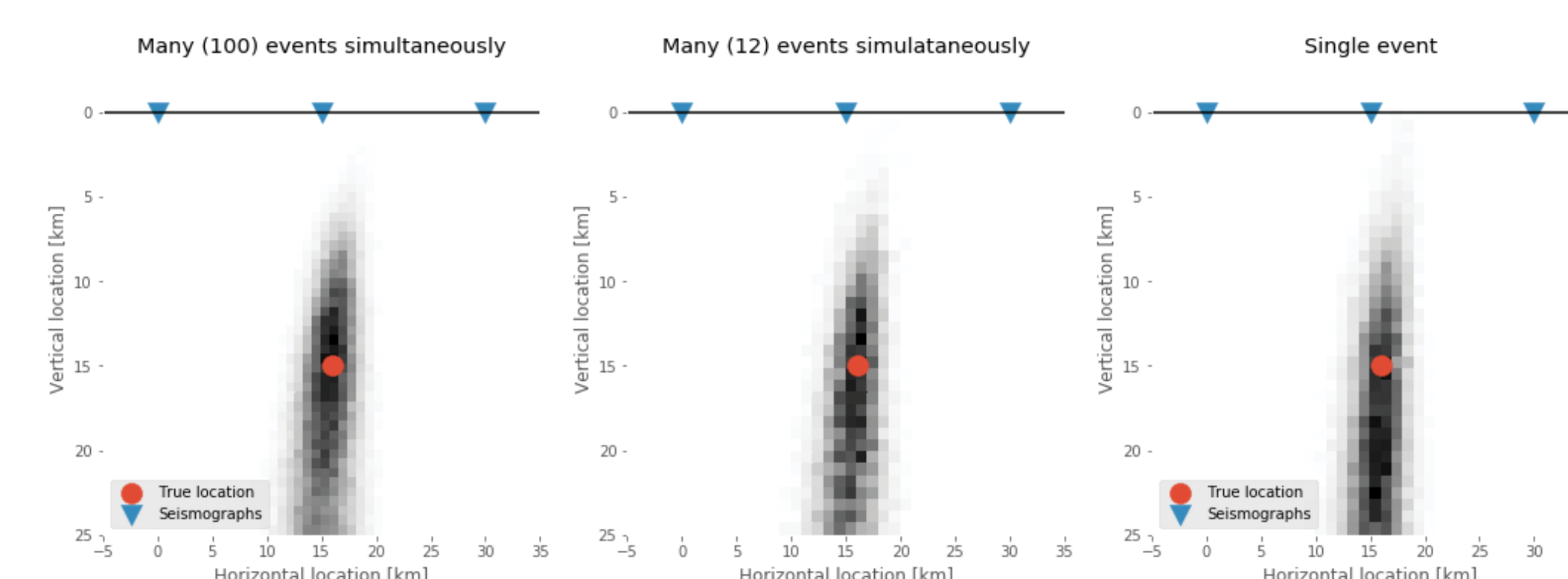


Figure 2: Posterior 2D marginals for a single hypocenter location, with decreasing number of simultaneous inferred events from left to right. Total parameters in the inference are respectively 301, 37 and 4. By inferring location on many events simultaneously the posterior is typically more informative (less dispersion), but sampling can only be handled by appropriately scaling algorithms such as HMC.

### Example 2: Acoustic full-waveform inversion

- × **Strongly non-linear forward model**, synthetic data, illustrated in [5].
- × Reflection seismology set-up.
- × 6 sources, 20 receivers.
- × Burn-in achieved after 10'000 samples.
- × Target: **SEG / EAGE overthrust model** [9].

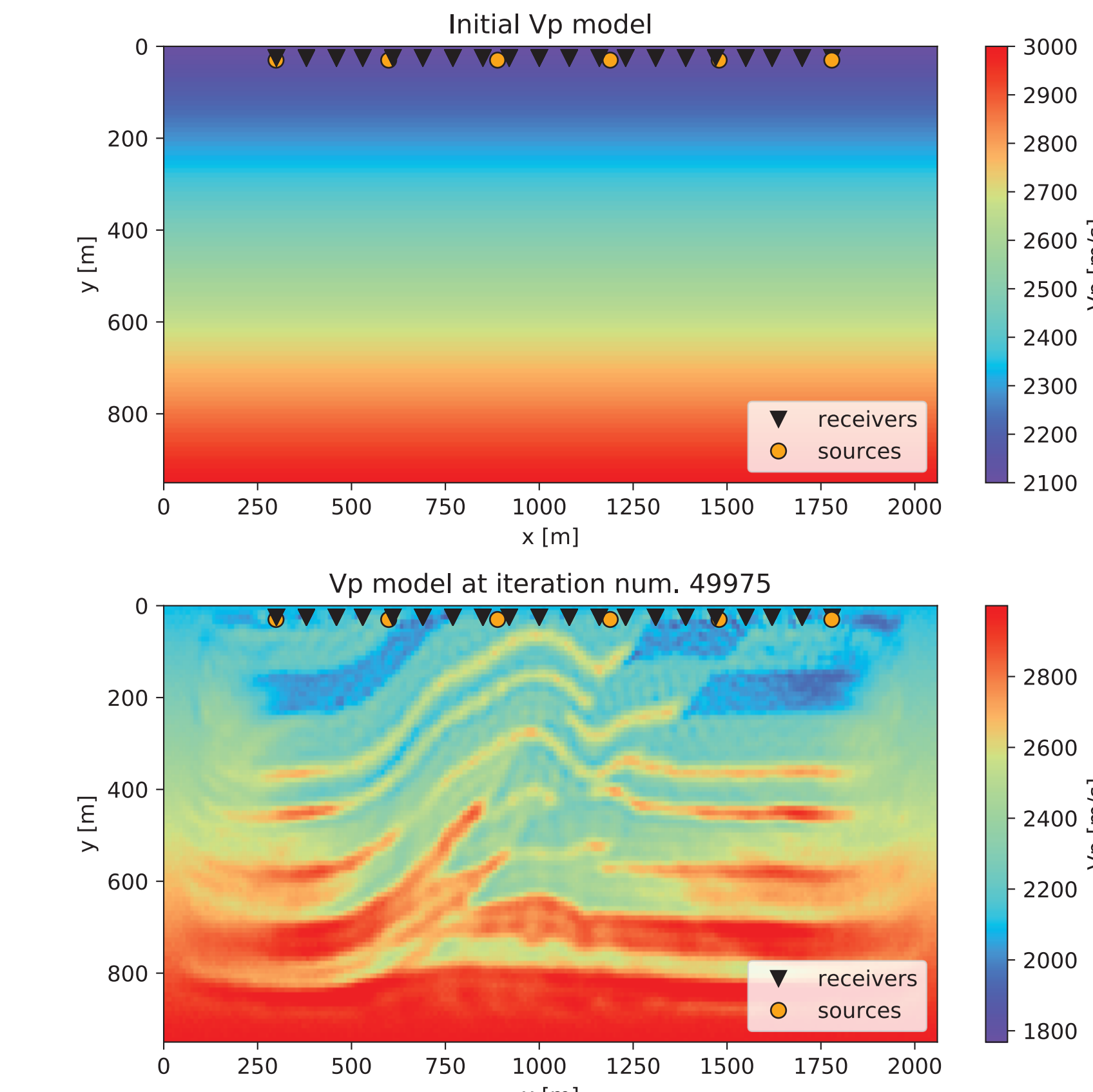


Figure 3: Two snapshots of samples generated during sampling of the acoustic FWI problem. The top panel is the initial sample at the beginning of the Markov chain. The second panel is a sample that lies well after the burn-in phase, resembling the structure (with appropriate noise) of the target model where coverage is good, and being more influenced by the prior where coverage is poor. The target model is a portion of the SEG/EAGE overthrust model.

## How can I use it?

HMCTomo supports your inversion implementation with minimal changes to your code, see **Snippet 1**. It is however, very much a **work in progress. For access at this moment, email the authors.**

**Snippet 1:** Example of a user implemented inverse problem in Python.

```
import hmc_tomography.Distributions._AbstractDistribution
class my_inverse_problem_class(hmc_tomography.Distributions._AbstractDistribution):
    def __init__(self):
        self.dimensions: int = 2

    def misfit(self, m) -> float:
        return 0

    def gradient(self, m) -> numpy.ndarray:
        return numpy.zeros((self.dimensions, 1))
```

Because HMCTomo comes with useful defaults and user prompts, a complete example producing meaningful results requires less than 10 lines of code, as illustrated in **Snippet 2**.

**Snippet 2:** Minimal example of sampling an inverse problem using HMCTomo in Python.

```
import hmc_tomography
likelihood = my_inverse_problem_class()

prior_mean = [0, 0], prior_variance = [1, 1]
prior = hmc_tomography.Distributions.Normal(prior_mean, prior_variance)

posterior = hmc_tomography.Distributions.BayesRule([prior, likelihood])

hmc_tomography.Samplers.HMC.sample(posterior, 'Output.h5')
```

## References

- [1] Fichtner, A., Zunino, A., & Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, 216(2), 1344-1363.
- [2] Simuté, S., Krischer, L., Boehm, C., & Fichtner, A. (2019). Seismic source inversion using Hamiltonian Monte Carlo and a 3-D Earth model in the Japanese Islands. *AGUFM*, 2019, S41A-06.
- [3] Gebraad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo. *Journal of Geophysical Research: Solid Earth*, 125, <https://doi.org/10.1029/2019JB018428>
- [4] Aleardi, M., & Salusti, A. (2020). Hamiltonian Monte Carlo algorithms for target-oriented and interval-oriented AVA inversions. *Geophysics*, 85(3), 1-76.
- [5] Zunino, A., and K. Mosegaard. "Integrating gradient information with probabilistic traveltime tomography using the Hamiltonian Monte Carlo algorithm." 80th EAGE Conference & Exhibition 2018 Workshop Programme. European Association of Geoscientists & Engineers, 2018.
- [6] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).
- [7] Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2, e55.
- [8] Villa, U., Petra, N., & Ghattas, O. (2018). hiPPYlib: An extensible software framework for large-scale inverse problems. *Journal of Open Source Software*, 3(30), 940.
- [9] Leconte, J.-C., Campbell, E., Letouzey, J. 1994. Building the SEG/EAGE Overthrust Velocity Macro Model. EAGE/SEG Summer Workshop - Construction of 3-D Macro Velocity Depth Models. 3D Model Representation and Visualization.