Approximate Bayesian Geophysical Inversion using Generative Modeling and Subset Simulation

Eliane Maalouf[1*], David Ginsbourger[2], Niklas Linde[3]

[1]University of Neuchâtel, [2]University of Bern, [3]University of Lausanne

* contact : eliane.maalouf@unine.ch

ABSTRACT

We present preliminary work on solving geophysical inverse problems by exploring the latent space of a joint Generative Neural Network (GNN) model by Approximate Bayesian Computation (ABC) based on Subset Simulation (SuS). Given pre-generated subsurface domains and their corresponding solver outputs, the GNN surrogates the forward solver during inversion to quickly explore the input space through SuS and locate regions of credible solutions. Akin to ABC methods, our methodology allows to tune the similarity threshold between observed and candidate outputs. We explore how tuning this threshold influences the uncertainty in the solutions, allowing to sample solutions with a selected diversity level. Our initial tests were carried out with data from straight-ray (linear) tomography with Gaussian priors on slowness fields and Gaussian versus Gumbel observation noise distributions. We are presently testing the methodology on non-linear physics to demonstrate its applicability in more general inversion settings.

PROBLEM SETUP

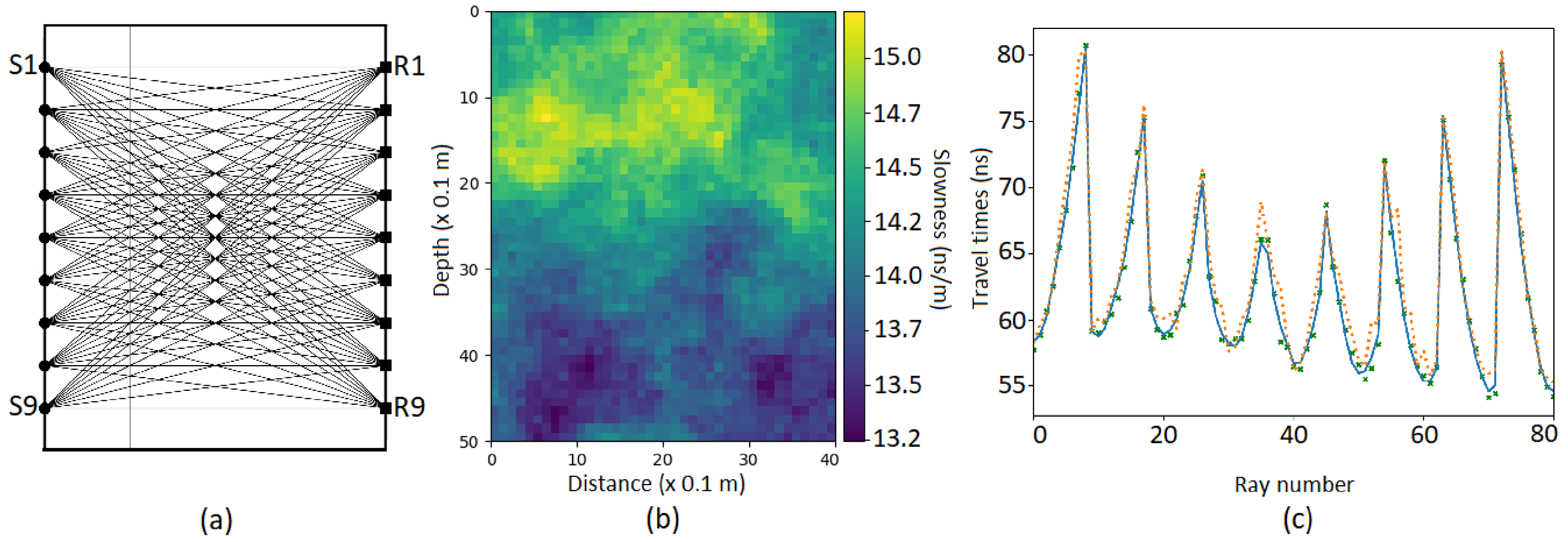

Figure 1: (a) Cross-hole Ground Penetrating Radar (GPR) sources (S1-9) emit high-frequency electromagnetic waves in one borehole, 5 m in depth, and are recorded by receivers (R1-9) in another borehole 4 m away. Tomography aims to retrieve (b) the slowness field of the geophysical domain from (c) noise-contaminated measurements of the 81 first-arrival travel times (TT), in ns, of waves that travel through the subsurface between the boreholes. In (c): blue full line, TT without noise given by the linear forward solver; green crosses, TT with noise from a multivariate standard Gaussian (std = 0.5 ns); orange dotted line, TT with noise from a multivariate Gumbel (location = 0.5 ns, scale = 1 ns). A Gaussian prior is set on the slowness field with mean 14 ns/m and exponential isotropic kernel (lengthscale = 2.5 m, variance at the origin = 0.16 $(ns/m)^2$).

METHODOLOGY

X and Y represent the subsurface domain and the response of interest respecitvely, and are defined on $\mathbb{X}$ and $\mathbb{Y}$. We assume that $Y = F(X) + E$ with E a noise with an unknown distribution, and F expresses the physics of the problem encoded by the simulator. We also assume that X and Y can be expressed as a function of a latent variable $Z \in \mathbb{Z}$ of moderate dimension compared to $\mathbb{X}$ and $\mathbb{Y}$, such that $(X, Y) = G^o(Z) = (g_1^o(Z), g_2^o(Z))$. When Z is sampled from a prescribed prior distribution, G^o is a generative function of couples (X, Y).

The inverse problem consists in retrieving a distribution of subsurface domains X having potentially generated a given vector of measurements y_{obs}.

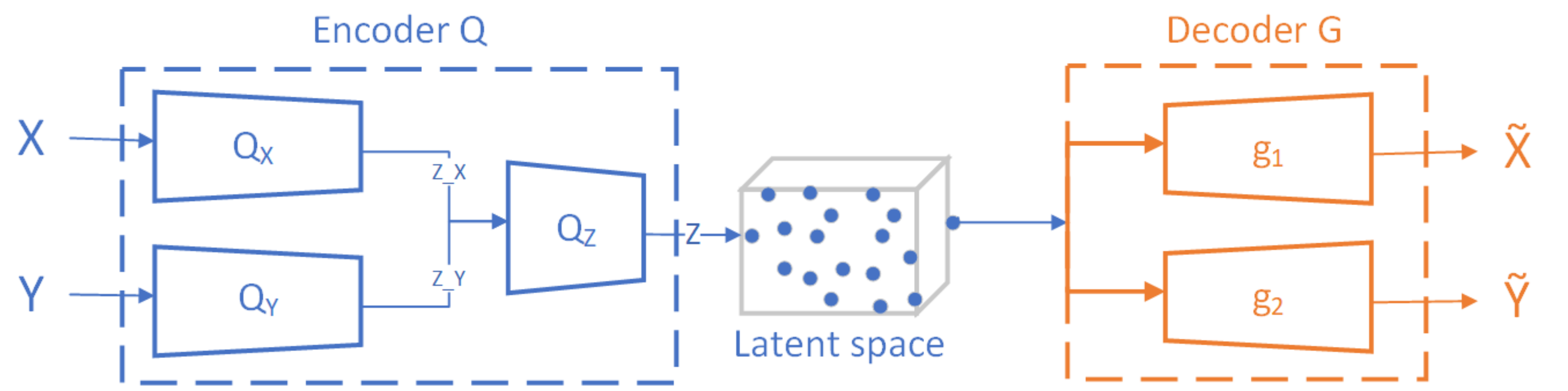

Figure 2: Schematic of the joint GNN

Phase I - Joint Generative Neural Network From couples (X, Y), we train a joint GNN based on Wasserstein Auto-Encoder [1] and Sinkhorn Auto-Encoder [2] models to estimate G^o. The training minimizes the Wasserstein distance between $P_{(X,Y)}$, the true distribution, and $P_{(\tilde{X},\tilde{Y})}$, the generated distribution, by minimizing

$$\min_G \min_Q \mathbb{E}_{XY \sim P_{(X,Y)}}[||X - g_1(Q(X,Y))||_p^p + ||Y - g_2(Q(X,Y))||_p^p] + \lambda W_p(Q_Z, P_Z)$$

with $||.||_p$ an l_p norm, $\lambda > 0$ a penalty and W_p the Wasserstein distance between the induced distribution of Z, Q_Z, and the prescribed prior, P_Z.

Phase II - ABC by Subset Simulation Given y_{obs}, and denoting by π_Z the prior density of Z, we consider a surrogate posterior density on $\mathbb{Z}$ defined by $\pi_{Z|Z \in \Gamma_\epsilon}(z) \propto 1_{\Gamma_\epsilon}(z)\pi_{\mathbb{Z}}(z)$ with $\Gamma_\epsilon = \{z \in \mathbb{Z} : d(g_2(z), y_{obs}) \leq \epsilon\}$.
We estimate Γ_ϵ by SuS [3]. SuS introduces a decreasing sequence of thresholds t_ℓ determining a sequence of nested subsets of $\mathbb{Z}$, $\Gamma_{t_\ell} = \{z \in \mathbb{Z} : d(g_2(z), y_{obs}) \leq t_\ell\}(\ell = 0, ..., m)$, with t_m set to ϵ. t_ℓ is set such that the probability of $\{Z \in \Gamma_{t_\ell}\}$ is equal to a prescribed value α. $p_\epsilon = P(Z \in \Gamma_\epsilon) = P(Z \in \Gamma_{t_0}) \prod_{\ell=1}^{m} P(Z \in \Gamma_{t_\ell} | Z \in \Gamma_{t_{\ell-1}})$ with $P(Z \in \Gamma_{t_0}) = 1$ since $\{Z \in \Gamma_{t_0}\}$ is certain. This gives $\hat{p}_\epsilon = \alpha^{m-1} \frac{N_{m-1}}{N}$ with N the sample size, and N_{m-1} the number of observations that yield $d(g_2(z), y_{obs}) \leq t_{m-1}$.

RESULTS & CONCLUSION

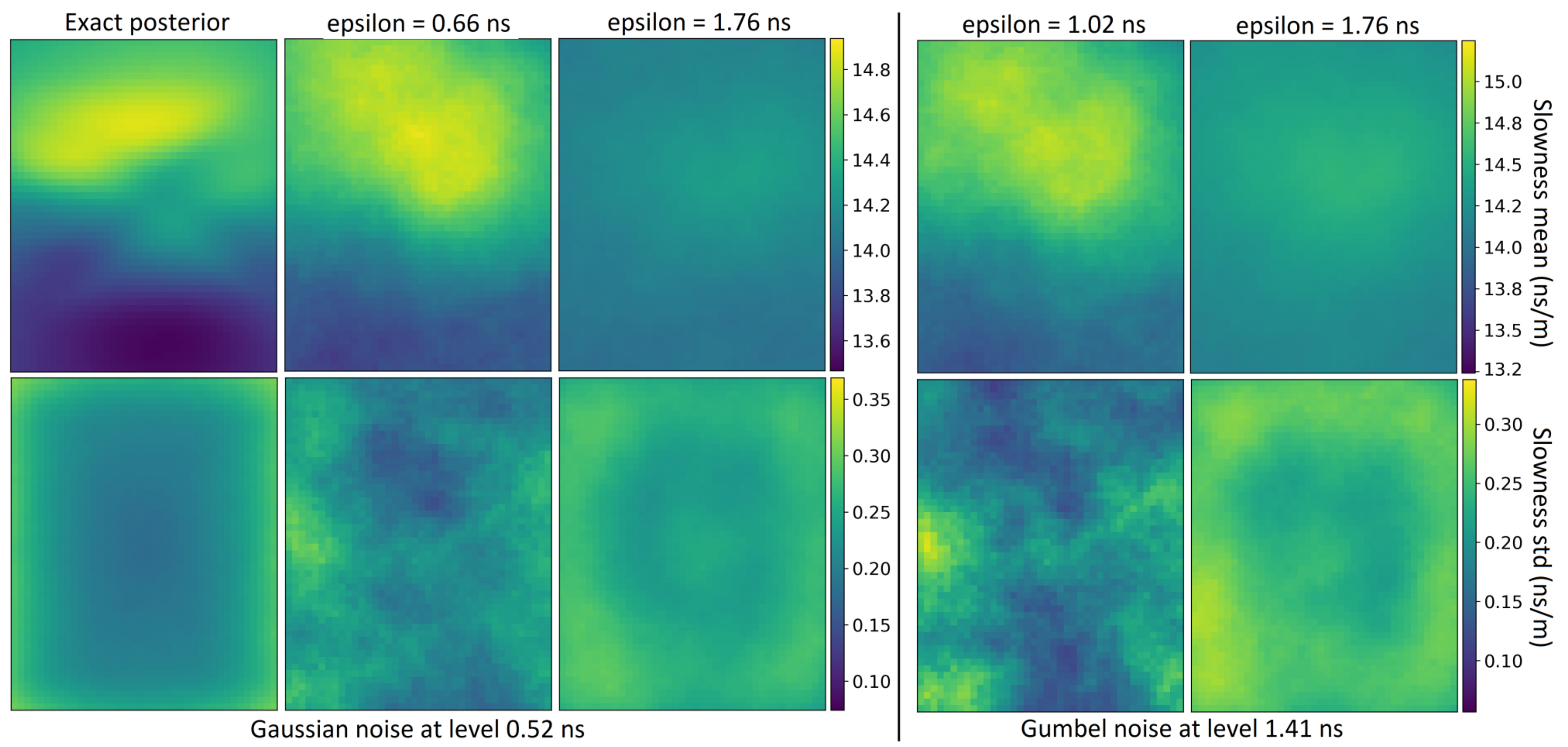

Figure 3: Proposed solutions pixel means and standard deviations at different tolerances with Gaussian and Gumbel noises

	Gaussian noise			Gumbel noise	
	Exact posterior	$\epsilon = 0.66$ ns	$\epsilon = 1.76$ ns	$\epsilon = 1.02$ ns	$\epsilon = 1.76$ ns
RMSE with ground truth	0.298 [0.220, 0.375]	0.336 [0.213, 0.460]	0.488 [0.323, 0.653]	0.381 [0.275, 0.487]	0.464 [0.295, 0.633]
RMSE with posterior mean		0.294 [0.145, 0.444]	0.454 [0.280, 0.629]		

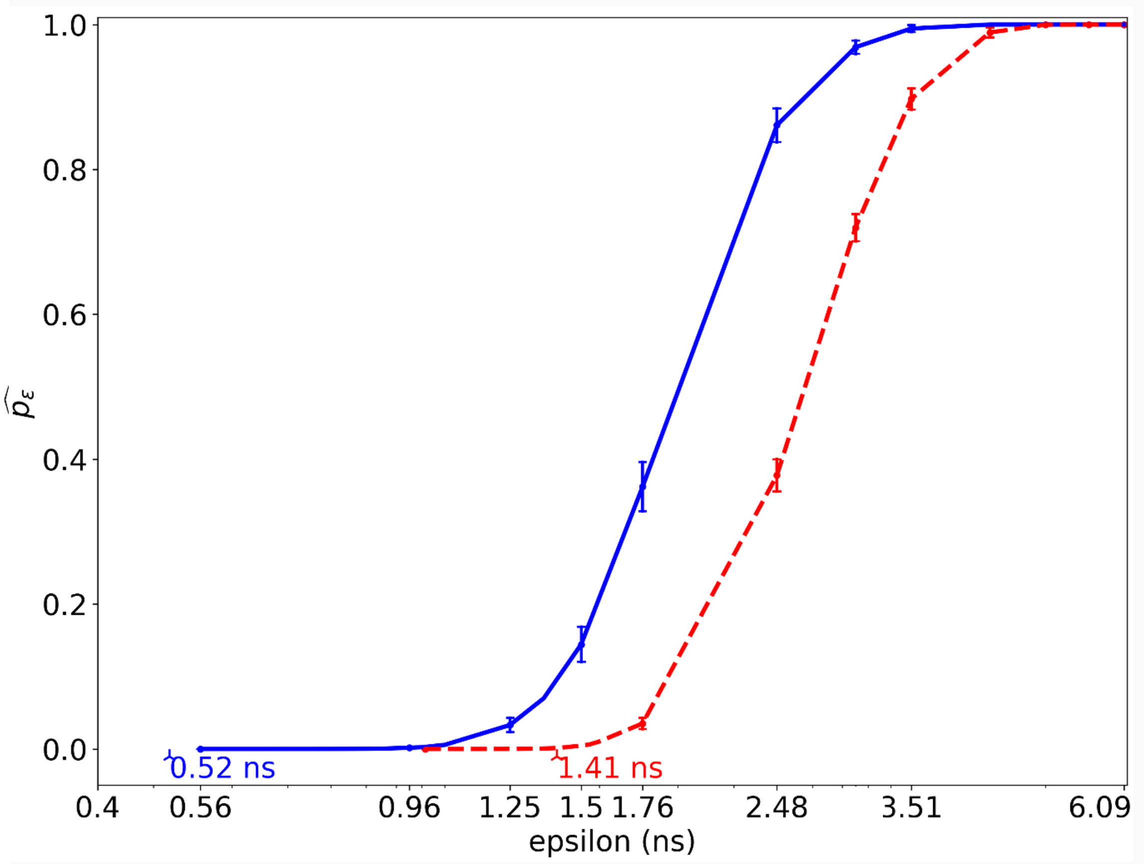

Figure 4: Probability content in $\mathbb{Z}$ vs tolerance ϵ with Gaussian noise (blue) and Gumbel noise (red)

We proposed a methodology to explore the solution space of an inverse problem while avoiding the evaluation of the forward solver. Through the estimation of the probability content on the latent space at different ABC tolerances, we explore the diversity of the solutions and we gain insight about the unknown noise. Future work includes testing with non-linear physics and more complex noise distributions to evaluate the method's potential to generalize.

REFERENCES

[1] Tolstikhin, I. et al., Wasserstein auto-encoders. ICLR (2018)
[2] Patrini, G. et al., Sinkhorn autoencoders. UAI (2020)
[3] Au, S.K. & Beck, J.L., Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic engineering mechanics (2001)