Bayesian geophysical inversion with Gaussian process machine learning and trans-D Markov chain Monte Carlo

Anandaroop Ray* 
Geoscience Australia 
ACT, Australia 
ea2ray@ucsd.edu

David Myer
BlueGreen Geophysics 
Encinitas, CA, USA 
dmyer@bluegreengeophysics.com

SUMMARY

A key aspect of geophysical inversion is the ability to model the earth with a low dimensional representation. There exist various approaches to solve the inverse problem. However, most methods do not automatically adapt inverse model complexity or the number of active model parameters as dictated by data noise and sparse receiver coverage, do not quantify inverse model uncertainty or do not work equally well for 1D, 2D or 3D earth models. Low frequency electromagnetic (EM) inversion for example, can require for 3D problems upward of 10^6 cells to forward model. Only a small fraction of these cells is effectively resolvable and there are significant trade-offs between them. To address these limitations, we present a novel approach to earth model parametrization by using a Gaussian Processes (GP) machine learning (ML) technique, coupled with a parsimonious Bayesian trans-dimensional (trans-D) Markov chain Monte Carlo (MCMC) sampling scheme. One aspect that sets our approach apart from recent spatial dimension agnostic algorithms in the trans-D or ML literature is the ability to specify inversion property priors directly, as opposed to doing so in a transform domain of the property. Finally, we note that our method falls in the category of ML approaches that do not attempt to learn the physics of the process, but instead learn the representation of parameter values through a misfit function. We apply the trans-D-GP method to a 1D controlled source electromagnetic (CSEM) and 2D non-linear regression problem, using actual field data from the Northwest Australian Shelf for the former. The key advantages in using our method are the simplicity of prior specification, parsimonious low dimensional representations, and ease of representing large-scale models in 1D, 2D or even 3D with the same parametrization and computer code.

Key words: inverse theory, Bayesian probability, electrical properties, machine learning

INTRODUCTION

Geophysical electromagnetic (EM) inversion with linearized, gradient-based methods are efficient, well understood and have been extensively used (e.g. Constable et al., 1987; Newman and Alumbaugh, 2000; MacGregor and Sinha, 2000; de Groot-Hedlin and Constable 2004; Abubakar et al., 2008; Key, 2009; Mittet and Gabrielsen, 2013; Myer et al., 2015). To stabilize matrix inversions required to converge to a solution, keep the solution close to a preferred model and enforce smoothness in the solution, some form of regularization must be imposed for the inverted solution to be meaningful. In a Bayesian framework, regularization is looked at as a means of incorporating additional information to arrive at a desired solution (see Calvetti and Somersalo (2018) for a detailed discussion). This requires we interpret the solution model as a random variable instead of the solution possessing one single value.

In a Bayesian framework, given observed data with a description of the data noise, we aim to find the distribution of data-compatible solution values, through a forward model and prior knowledge (or belief) about the solution. This distribution, known as the posterior distribution, encapsulates our state of knowledge (and hence uncertainty) about the solution space (in our case, the earth’s subsurface conductivity). Bayes’ theorem bridges posterior and prior knowledge through the acquired EM data. This specification of prior knowledge (e.g. Hansen and Minsley, 2017) and its parametrization is often overlooked in Bayesian inversions of geophysical data (see Pasquale and Linde (2016) for a discussion). Designing an informative prior which accurately reflects the earth’s spatial character given the resolution we expect our data to possess is key to drawing meaningful inferences about subsurface geology. While this may sound like a chicken-and-egg situation, a Bayesian perspective lays bare the fact that we must make choices in designing an inversion scheme, whether it be regularized or otherwise implemented. With choices based on the physics of the problem, well-designed Bayesian algorithms can infer the resolution with which we can ‘see’ into the earth. Recent geophysical work highlighting the importance of choosing a priori appropriate basis functions in a Bayesian framework can be found in Hawkins and Sambridge (2015); Pasquale and Linde (2016) and Ray et al. (2017). In the field of hydrogeophysics, Cordua et al. (2012); Lochbuhler et al. (2015) and Laloy et al. (2017) have used Training Image (TI) based priors for this purpose.

For low frequency EM inversion, we propose to use Gaussian Processes (GPs) for prior representation. Gaussian Processes have well understood qualities of spatial variability – see Rasmussen and Williams (2006) for a thorough review. By using GPs in conjunction with the trans-dimensional (trans-D) McMC method (Green, 1995; Malinverno, 2002; Bodin and Sambridge, 2009) for model solution dimension reduction, we can effectively model high dimension, but keep the number of inverted parameters small (i.e., achieve parsimony, Malinverno, 2002). It is small model dimension that makes McMC tractable or gradient inversion stable (Laloy et al., 2017). As is usual with trans-D methods, the posterior model ensemble quantifies the non-linear spatial resolution of the data.
METHOD AND RESULTS

Gaussian Processes

A Gaussian Process is a stochastic process that is completely determined by its mean and covariance. It is defined by priors and posteriors over functions. Broadly speaking, GPs are a method of non-parametric regression that do not require a fixed discretization, providing both a prediction and uncertainty around the prediction. To gain insight into the workings of GPs we follow the Bayesian exposition of Williams and Rasmussen (1996) through an example shown in Figure 1. First, we specify prior notions of spatial smoothness through a covariance, defined by a similarity kernel. This kernel ensures that spatially close locations have similar values. Training observations are then brought in and lead to an updated, posterior multivariate Gaussian. Test outputs at all unobserved points are then simply conditional realizations from the posterior Gaussian.

Trans-D-GP

With trans-D-GP we can add and subtract GP training points via trans-D ‘birth’ and ‘death’ MCMC steps (Geyer and Møller, 1994; Green, 1995; Sambridge et al., 2006) to represent an earth property model, say conductivity, using a posterior GP mean. The key advantage of this approach lies in the fact that with the same formalism we can represent a 1D, 2D or 3D earth – instead of layers for a 1D earth, Voronoi cells for a 2D earth or other parametrizations in 3D. It follows that the same inversion code and GP parametrization can be used without the computational geometry overhead required to go from 1D to 2D or 3D earth models. Unlike in geostatistics (e.g. Krige, 1952), we are not drilling the earth to obtain more training samples of conductivity to improve the GP. Instead we use EM data, which at best only smoothly resolve the earth’s subsurface, to update the GP mean (see Ray and Myer (2019) for details).

Figure 1. Top left: A prior covariance $C_{prior}$, defined by a smoothly decaying stationary similarity kernel, such that values differing by up to 0.1 units in $x$ have high correlation. Top right: 50 random realizations from a Gaussian with zero mean and covariance $C_{prior}$. Bottom left: A GP posterior covariance, formed by modifying the prior covariance by bringing in 10 training observations. Bottom right: The 10 training observations are shown with magenta circles. The unknown true function being approximated is shown in black. The inferred posterior GP mean is shown in dashed blue. Also shown are 50 random test realizations from a Gaussian with mean set to the posterior mean, and covariance $C_{posterior}$. We should note that regions in $x$ with fewer training points have high posterior variance and vice versa.

Figure 2. CSEM inversion from the Scarborough gas field in the Northwest Australian shelf. Top: Off reservoir posterior resistivity density functions (PDFs) and their accompanying cumulative distribution functions (CDFs). Darker regions of the PDF plot are more probable. Each colour in the CDF plot corresponds to a quantile at that depth. The region between the dashed lines is the 98% credible interval. Bottom: On reservoir posterior resistivity distributions. Note how the resistivity PDFs near and above 2000 m depth shift in bulk to more resistive for the on-reservoir case. This is indicative of the resistive hydrocarbon bearing reservoir, as also found by Ray et al. (2014).

The CSEM method

The marine controlled source electromagnetic (CSEM) method is an active source sounding technique. It has been in use for over three decades for the detection of geology with high resistivity contrasts (Young and Cox, 1981; Chave and Cox, 1982). Conductive media such as seawater or brine filled sediments have a characteristic electromagnetic scale length (skin depth) $\delta = (2\rho/\mu_0)^{0.5}$ that is dependent on both the medium resistivity $\rho$ and the frequency of propagation $\omega$, where $\mu$ is the permeability of the medium. Owing to the fact that $\delta$ is smaller in conductive (low $\rho$) media, marine geophysical EM methods operate in the lower frequency quasi-static regime with physics that is more diffusive than wave like (Loseth et al., 2006). To first order, it is this diffusive decay which can characterize the conductivity of a given medium. For hydrocarbon bearing geology, it is the high resistivity of the hydrocarbon accumulation with respect to its surroundings that produces a detectable EM signature. However, reliable inferences from CSEM can only be made by means of an inversion, and a Bayesian inversion is ideal to quantify the uncertainty inherent in the inversion process (e.g. Hou et al., 2006; Chen et al., 2007; Gunning et al., 2010; Buland and Kolbjørnsen, 2012). The aforementioned references, while Bayesian, use a fixed number of dimensions as dictated by the user, and not by the data misfit. Trans-D Bayesian methods do not suffer from this limitation.
and have been used to invert CSEM data with both 1D and 2D parametrizations (e.g. Ray et al., 2014). Though the theory was similar in both cases, the implementation of the trans-D method required parametrization with layers/interfaces for a 1D earth and Voronoi cells for 2D.

Scarborough Field CSEM Data Inversion

We apply trans-D-GP to data from the Scarborough gas field, which lies inside the Exmouth Plateau in the North West Australian Shelf. The reservoir itself is between 20 to 30 m thick at a depth of ~2000 m below sea level. The bathymetry, also quite flat is at a depth of ~950 m. Resistivity at reservoir level is moderate at 25 ohm-m and the reservoir is overlain by several thin 5-10 ohm-m layers. We invert data from two sites located in the ‘off reservoir’ and ‘on reservoir’ parts of a CSEM tow-line. The posterior resistivity with depth for the both sites is shown in Figure 2. In the off-reservoir part there is evidence of weak 8-10 ohm-m anomalies, with accompanying changes in the CDF above 2000 m depth. Contrast this with the on-reservoir posteriors indicating high probability of moderately resistive material of 10-25 ohm-m at similar depths. Our results are in line with the previous findings of Ray et al. (2014), who show that the posterior PDFs of resistivity (not just the mode) near 2000 m depth move en masse to more resistive values as we tow the transmitter from off-reservoir to on-reservoir sites.

2D Inversion: A Non-Linear Regression Application

As an example of extending to higher spatial dimensions, we solve a non-linear regression problem with ‘non-function data.’ The data to be fit (Figure 3) are not the outcome of a single valued function as a function of a distance coordinate (see Criminisi et al., 2011, for further examples). We could also think of this as a 2D spatial regression problem. Geoscientific applications of this type using trans-D methods have been investigated by Gallagher et al. (2011) and Bodin et al. (2012a). Depending on the specifics of the problem, they used interfaces for one spatial dimension and Voronoi cells for two. However, with trans-D-GP, the exact same theory and computer code can be used, no matter the number of spatial dimensions.

The objective is to find 2D representations (and also their uncertainty) which approximate the true image at locations not sampled. Naturally, one could use kriging methods to solve this problem, but we are interested in one further property that a standard kriging methodology cannot ensure. We would like parsimonious representations of this image, as we would require for geophysical applications over a spatially vast part of the earth, forward modelling the physics for which would require many pixels. Further, we demand that the data coverage and noise levels should determine the complexity of the model representation(s) in concert with our prior knowledge.

One converged sample from a trans-D-GP MCMC chain is shown in the left of Figure 4, with the accompanying 55 training points needed to define the model. The mean of converged samples is shown on the right. In Figure 5 we show how our method adapts to both model complexity as well as the manner in which the data have been sampled, a hallmark of trans-D methods that we have preserved in our algorithm. Where there is less data, there should be high posterior uncertainty, as we can see in the figure to the left. On the right, we can see that where the data are informative, there is a dense nucleation of GP points.

CONCLUSIONS

We have developed a new methodology which incorporates the well-known Gaussian Process ML technique into a parsimonious trans-D framework, demonstrating its use in 1D, 2D and field applications (Ray and Myer, 2019). We have shown that ML techniques can be easily incorporated into a Bayesian geophysical inversion framework through the
specification of prior information (e.g. Laloy et al., 2017). The key advantage in using our method is the simplicity of prior specification and ease of representing large-scale models in 1D, 2D or even 3D. We contend that for low frequency geophysical inversion, trans-D-GP is simple enough to implement from scratch without the use of an ML library and provides the scalability for inverting large 2D or 3D earth models with a small number of effective parameters. Last but not least, it provides uncertainty estimates on inverted earth properties.

ACKNOWLEDGEMENTS

All calculations were carried out using the Julia language (Bezanson et al., 2017), available under the MIT license. The authors thank BHP Billiton Petroleum and the Seafloor Electromagnetic Methods Consortium at the Scripps Institution of Oceanography for making the Scarborough data publicly available.

REFERENCES


Key, K., 2009. 1D inversion of multicomponent, multifrequency marine CSEM data: Methodology and synthetic studies for resolving thin resistive layers, Geophysics, 74(2), F9.


