ADVANCED NUMERICAL TECHNIQUES FOR INVERSE PROBLEMS IN GEOPHYSICS

Olga Ortega Gelabert

UPC

Doctoral Thesis
Advisors: Sergio Zlotnik and Pedro Díez
Barcelona, November 2020

Departament d'Enginyeria Civil i Ambiental (DECA)
Programa de Doctorat en Enginyeria Civil

“ Everything is possible. The impossible just takes longer. ”

***Dan Brown**, Digital Fortress*

ABSTRACT

Advanced numerical techniques for inverse problems in geophysics

Olga Ortega Gelabert

This thesis presents an efficient methodology to couple Model Order Reduction techniques within the framework of geophysical probabilistic inversion problems. Accurate models of the interaction between Earth inner processes and surface features are essential to make reliable predictions of the observables which are a fundamental part of Bayesian inference. Markov Chain Monte Carlo (MCMC) methods have become standard in dealing with probabilistic inversions and they rely on sampling strategies that require solving forward problems many times. Computationally expensive large-scale forward problems are the principal bottleneck that can limit the capabilities and potential of multi-observable geophysical probabilistic inversions. In particular, dynamical effects arising from the sub-lithospheric mantle flow are not usually taken into account in the estimation of surface elevation due to the high computational cost of the associated 3D Stokes flow problem.

The main idea of this thesis is to use the Reduced Basis (RB) method as a surrogate of the true forward problem (3D Stokes flow) to provide fast and accurate approximations. The surrogate is then used to generate samples of the posterior distribution at a much lower computational cost. RB strategies are based on expressing the solution of a problem in a low dimensional space, i.e. a reduced basis. Taking advantage of the convergence nature of the MCMC, we propose a greedy strategy that builds the reduced basis *on the fly* and as required by the inverse problem. In doing so, the basis is specifically tailored to the posterior features of the problem. In addition, to guarantee an accurate surrogate we define a goal-oriented error estimator which focuses on a particular Quantity of Interest of the problem and, therefore, it guides the basis to achieve the required accuracy in such particular features. All this translates into a problem-shaped basis that is more compact and smaller than if it had to be accurate everywhere in the domain. Moreover, to deal with the costly assembly of matrices, we use the specific parametrization of the problem and sampling strategy to define an assembly procedure that efficiently updates the matrices only with the contribution of the elements that changed between successive inversion steps.

The benefits and limitations of the method are illustrated through several numerical examples. Finally, to demonstrate the applicability of the method two more realistic inverse problems are presented. The first one uses dynamic topography to infer the Lithosphere-Asthenosphere Boundary depth of a spherical domain representing a portion of Earth and the second one is applied to a larger problem in which the African lithospheric structure is discretized in 1225 inversion parameters.

ACKNOWLEDGMENTS

I would like to start by expressing my sincere gratitude to my supervisors Sergio Zlotnik and Pedro Díez, for their support, advice, useful discussions and, above all, their high level of patience during all these years. I have certainly enjoyed and learned a lot from them and this thesis would not have been possible without their guidance. I would also like to thank Juan Carlos Afonso for giving me the opportunity to work with him at Macquarie University, for his dedication and time and for showing me all the potential behind probabilistic inversions.

I would also like to thank all people from LaCaN, who I had the pleasure to meet and also share office with. Also a huge thanks to the *esmorzing time* people, for a nice and triangular way to start the day. To Paff, Arash, Hasini, Alex, Julcha and all the caffeine intake during these years. To David and Alberto for opening the door to climbing and the many more experiences shared afterwards. To Karim who introduced the concept *funny in your head*, but surprisingly managed to understand it. And, finally, but not the least, to Jere and that priceless secreto!

Going to Sydney also gave me the opportunity to meet Coti and struggle, learn and enjoy together the reduced basis adventure. The down under experience would not have been the same without Montgarri, Rosa, Marina and everyone I met there.

Not to forget lifelong friends, Rodri, Isa, Quiti, Cle, Laia, Pastor, Gerard, Sergi, Lota for reaching the 3rd floor all together. Specially to Isàkit and Laia who made the lockdown writing experience far more bearable. A big thanks to everyone who in one way or another contributed to this outcome.

Finally I owe much to my family who are always there, dealing with my stubbornness and decision problems no matter what!

This project was financially supported by Ministerio de Economía, Industria y Competitividad (MINECO) with the project DPI2014-51844-C2-2-R and the associated grant BES-2015-074911. It has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 777778

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

One of the main challenges in modern lithospheric research is the understanding and characterization of the present-day physical state of the thermal and compositional structure of the Earth's lithospheric and sub-lithospheric mantle. This is essential to develop any evolution model of the Earth as well as to understand the relation between surface features and inner processes. Current knowledge comes from two main sources of information: geophysical observables (e.g., elevation, gravity anomalies, travel time data, surface heat flow, etc.) and mantle samples brought to the surface. They all can be used to make inferences about the upper mantle. Geophysical observables have a larger spatial and temporal coverage, but inferring composition and temperature structures presents more difficulties. Instead, exhumed mantle samples represent a direct evidence of the inner structure but their coverage is discontinuous and limited which implies large uncertainties. It is well known that independent models constraining single data sets typically fail at providing satisfactory fits to other observables (Afonso et al., 2013b, 2016b; Forte, 2007). Fitting simultaneously more than one observable certainly allows to obtain more consistent models since, for instance, different geophysical observables exhibit different sensitivities to variations of temperature, composition as well depth and, therefore, it can help to distinguish and constrain the model (Afonso et al., 2008; Fullea et al., 2009).

Joint inversions of long-wavelength gravity and seismic data at global scale represent a classic example pioneered by Forte et al. (1994) and more recently expanded in Simmons et al. (2010). Despite being one of the most advanced approaches to date, these global studies were necessarily based on linearized deterministic inversions and therefore not well suited for quantifying the associated uncertainties, nonlinearities, and/or rigorously considering the stochastic nature of the data noise. Moreover, although all inversions aim at minimizing the misfit between predicted and observed data, deterministic or traditional approaches look for a single best model, i.e. the model that best fits the data. Therefore, they can not account for the possible non-uniqueness of the solution inherent in geophysical problems. Probabilistic inversion schemes (Gregory, 2005; Kaipio and Somersalo, 2007; Mosegaard and Tarantola, 2002; Tarantola, 2005) represent an attractive option that can overcome some or all of these difficulties. Unlike deterministic strategies, they look for a statistical description of the model parameters so that the inverse solution is not a single model but a probability density function (PDF) over the space of model parameters. This PDF, the so-called posterior distribution, represents our best state of knowledge of the model after combining data, physical theories and prior information.

Probabilistic inversions typically rely on sampling techniques that require solving the forward problem many times, thus posing a real challenge when the forward problem is computationally expensive (as in the case of 3-D Stokes flow with varying viscosity). Recently, Baumann et al. (2014) and Baumann and Kaus (2015) demonstrated the viability of a probabilistic formulation for the geodynamic inverse problem incorporating realistic assumptions about the mechanical behavior of the lithosphere and upper mantle. These authors used this approach to constrain rheological parameters and subsurface density at lithospheric scales. However, due to the high computational cost of the forward evaluations they kept the domain relatively small and, moreover, their implementation relied heavily on a priori knowledge of the temperature and compositional structure of the model, something that is not straightforward in most regions of the world. Integrated approaches that jointly invert a number of data sets also sensitive to the thermochemical structure of the Earth (e.g., Afonso et al., 2013a,b, 2016b; Khan et al., 2008, 2011) represent a more general approach which can also help reducing the number of forward evaluations. The recent work of Afonso et al. (2013a,b, 2016b) presents a multi-observable probabilistic inversion method that simultaneously inverts the most appropriate data sets (with the necessary complementary sensitivities) for the temperature and compositional structure of the lithosphere and upper mantle: Rayleigh wave dispersion data, teleseismic P and S traveltimes, gravity anomalies, geoid height, satellite-derived gravity gradients, surface heat flow, and absolute elevation; P wave receiver functions have also been implemented recently (Tork Qashqai et al., 2016, 2018).

Dynamic contributions to the absolute elevation are included in Afonso et al. (2016b). Their rationale is based on decoupling the lithospheric and sublithospheric contributions (cf. Molnar et al. (2015) for a discussion about dynamic topography). Surface elevation is predominantly controlled by density variations in the lithosphere. The lithosphere is assumed to be cold enough to be considered essentially rigid, and, therefore its “static” contribution to surface elevation is accounted for by a simple lithospheric isostatic balance. Instead, density variations in the sublithospheric mantle combined with its lower viscosity, can evolve into flow which is transferred to the surface via viscous normal stresses, the so-called dynamical effects. The dynamic contribution from the instantaneous sublithospheric flow is superimposed to the static contribution to obtain a better prediction of the absolute elevation. However, their implementation of the associated Stokes forward problem was inefficient and based on a number of simplifying assumptions to make the problem tractable in the probabilistic framework. In order to exploit the full capabilities of joint geophysical-geodynamic probabilistic inversions, a more efficient implementation of the Stokes flow problem for large-scale models is required. This establishes the starting point and purpose behind this thesis.

1.2 State of the art

There is a large variety of problems whose governing equations are represented by parameters (e.g. material properties, initial or boundary conditions, geometry parameters, etc). Many engineering and science problems usually require the evaluation of models many times and for different values of the input parameters (Benner et al., 2015). They are often called many-query applications or problems (Rozza et al., 2007; Peherstorfer et al., 2018). Solving many large-scale and complex models usually demands huge computational resources which in some situations are unattainable. The ideal solution would be to generate a low-cost parametric reduced model that is able to approximate the original and complex full order model over a wide range of parameters.

The design of surrogates or low-fidelity models is an ongoing research topic. Low-fidelity models can be classified in three groups: *hierarchical models*, *data-fit models* and *reduced-order models*. All of them have advantages and disadvantages and when to use each of them is problem dependent. Hierarchical models can be used any time the problem allows to. They use the physics of the problem or implementation details to lower the accuracy and reduce the cost. For instance, losening tolerances, simplifying physics (Tarvainen et al., 2010), coarse grids (Arridge et al., 2006; Kolehmainen et al., 2009; Efendiev et al., 2006), polynomial chaos (PC) expansions (Yan and Zhang, 2017; Yan and Zhou, 2019). Instead, data-fit models are not based on the physics of the problem and rely on interpolation strategies from high-fidelity models to link input parameter and output variables (Kennedy and O'Hagan, 2001). Therefore, they can be built using the high-fidelity model as a black box. Building reduced-order models is more intrusive since they are obtained by projecting the governing equations of the problem onto a subspace of reduced dimension. However, as an advantage, they retain the structure of the underlying physical model which allows deriving rigorous error bounds and estimates (Prud'homme et al., 2002; Rozza et al., 2007; Grepl and Patera, 2005; Patera and Rozza, 2006) which are essential to guarantee accurate approximations. In this sense, data-fit models are more limited.

Model Order Reduction (MOR)

Reduced-order techniques, aka Model Order Reduction (MOR), are based on the idea of expressing the solution of a problem in a low dimensional space. They are specially suitable to solve large number of problems with similar characteristics. Classical approaches are divided into an *offline* stage where all costly computations are performed once and an *online* stage where fast solutions are obtained.

The Reduced Basis (RB) method is a MOR technique based on the idea of the snapshots. Snapshots are a collection of full or high-fidelity solutions evaluated at specific locations of the parameter space which combined together form a basis (Patera and Rozza, 2006; Rozza et al., 2007; Quarteroni et al., 2016; Hesthaven et al., 2016). Although the construction of the basis is straightforward, the main challenge lies on where to select the snapshots and how many are required to generate a basis that accurately represents the parameter space (without ill-conditioning the reduced system). Structured and random sampling are suitable for small number of parameters, whereas for larger values (> 10) more sophisticated sampling strategies are required. The greedy sampling approach is an iterative strategy that uses an initial training set of parameters and an error estimator of the reduced model to adaptively decide which is the next sampling point, i.e. the one with the highest error (Prud'homme et al., 2002; Grepl and Patera, 2005; Veroy et al., 2003). The advantage of the greedy approach is that it uses the underlying structure of the problem to guide the search. A wide range of error estimators have been successfully used, from the classical minimum-residual approach (Rozza and Veroy, 2007), to goal-oriented error estimators to target relevant locations or features of the problem (Florentin and Díez, 2012; Larion et al., 2020). Variations from the standard greedy approach have also been developed. For instance, Hesthaven et al. (2014) proposed a greedy strategy that adaptively enriches the initial parameter training set whereas Maday and Stamm (2013) use the greedy search to locally adapt the basis. In Bui-Thanh et al. (2008), a model-constrained optimization problem over the continuous space of parameters is solved to find the new snapshot location and a similar strategy is proposed by Lieberman et al. (2010) whose optimization problem also accounts for prior information. The so-called Proper Orthogonal Decomposition (POD) can be understood as a particular case of RB, or a suitable complement, aimed at eliminating possible redundancies of the basis. It generates an orthonormal basis which is optimal in a least-square sense. Reduced Basis methods relying on snapshots have been applied to a wide range of fields, for instance, control of fluid flows (Ravindran, 2000; Ito and Ravindran, 1998), parametrized steady incompressible Navier-Stokes equations (Veroy and Patera, 2005), heat conduction problems (Wang and Zabaras, 2005), nonlinear combustion problems (Galbally et al., 2009), electromagnetic problems (Hess and Benner, 2014), thermo-hydro-mechanical coupled problems (Larion et al., 2020), magnetothellurics (Manassero et al., 2020), among others.

Unlike classical offline/online approaches, other attractive strategies propose to use the online stage to update the reduced model to fit possible changes or variations in the problem. Peherstorfer and Willcox (2015) proposed a data-driven reduced order model that incorporates information from sensors of the online stage to dynamically adapt the reduced model. It avoids new expensive offline computations from scratch which makes it suitable for real-time decision making problems. Adaptation during the online stage is also used in Maday and Stamm (2013) to select the specific set of snapshots that participate in each reduced approximation. Florentin and Díez (2012) proposed a strategy to automatically generate and enrich the basis with snapshots during a Monte Carlo process and as needed to guarantee a certain accuracy.

Another kind of MOR is the so-called Proper Generalized Decomposition (PGD). Unlike RB or POD, it does not base its solution on snapshots but it relies on a separate representation of parametric basis functions to provide an explicit parametric solution of the problem (Chinesta and Ladevèze, 2014; Chinesta et al., 2014). Since the online part does not imply solving any reduced problem, it has been successfully applied to real-time monitoring or decision making problems, for instance electrical power flow problems (García-Blanco et al., 2018), thermal processes (Aguado et al., 2014), but also to design lattice parametric materials (Sibileau et al., 2018), in water agitation in harbours (Modesto et al., 2015), heat conduction geophysical applications (Zlotnik et al., 2015) or inverse problems (Berger et al., 2016).

Inverse problems and surrogates

Usually, in many-query situations, the large number of model evaluations can be seen as part of a loop where each iteration consists in receiving an input, evaluating the model and providing an output. When the main goal or result of the application is not the individual evaluations, but an outer-loop result obtained at the end, they are called outer-loop applications. For instance inverse problems, optimization problems, uncertainty propagation, sensitivity analysis among others (Peherstorfer et al., 2018).

The Bayesian approach to solve an inverse problem considers the input parameters of the model as random variables and the goal is to find the probability density function (PDF) that characterizes them. It requires sampling the parameter space, for instance by means of Markov Chain Monte Carlo (MCMC) methods and performing forward evaluations of the model at the chosen locations in order to approximate the distribution. Two main problems can arise related to the high cost of probabilistic inverse problems. First, sampling high-dimensional input parameter spaces is complicated and, second, it requires many large-scale forward evaluations whose cost easily becomes computationally unaffordable (Frangos et al., 2010). Frangos et al. (2010) classified the approaches to reduce the computational cost in three different groups:

- *Reducing the cost of each individual forward evaluation* by means of using surrogates or low-fidelity models. For instance, Wang and Zabaras (2005); Berger et al. (2016); Lieberman et al. (2010); Cui et al. (2014); Manzoni et al. (2016).
- *Reducing the input parameter space.* Sometimes the parameter space is linked to the spatial discretization of the problem and a different representation using the inherent structure of the field could reduce the dimension. For instance using the truncated Karhunen-Loeve (K-L) expansion (Marzouk and Najm, 2009) or the approach proposed by Lieberman et al. (2010) that applies the reduced basis strategy typical from MOR techniques to the parameter space.
- *Reducing the number of forward evaluations.* Such methods are aimed at improving the efficiency of the sampling algorithm so that a lower number of high-fidelity evaluations are required. This is typically achieved by combining low fidelity models (any of the ones above) with high-fidelity ones (Christen and Fox, 2005; Cui et al., 2014; Florentin and Díez, 2012; Lieberman et al., 2010; Efendiev et al., 2006).

Different strategies are devised aiming at reducing the computational cost in any of the aspects above or a combination of them. The simplest strategy is a surrogate-based MCMC in which the high-fidelity model is directly replaced by the low-fidelity one (Wang and Zabaras, 2005; Galbally et al., 2009). In doing so, the speedup is remarkable, but at expenses of a lower accuracy of the outer-loop result since the low-fidelity is simply providing an approximation of the input-output relationship which may yield to a biased posterior distribution. The accuracy of the outer-loop result is strongly determined by the accuracy of the low-fidelity model. Using surrogates inevitably brings the common trade-off between reducing computational cost and guaranteeing a certain accuracy.

One option to control the bias induced in the posterior distribution by the use of surrogates is by means of error bounds or a posteriori error estimators (Rozza et al., 2007; Grepl and Patera, 2005; Florentin and Díez, 2012). In some situations, if properly controlled, it is benefitial to accept some bias or inaccuracy if it brings an improvement in variance reduction of the posterior for a given computational effort (Cui et al., 2014). Or in other words, the bias introduced by an approximation is small with respect to the variance introduced by the length of the MCMC chain. Some authors studied the error between the true posterior distribution (i.e. the one obtained if sampling only uses the high-fidelity model) and the posterior approximated by the surrogate (Cui et al., 2014; Yan and Zhang, 2017; Yan and Zhou, 2019). They proved error bounds on the Kullback-Leibler and Hellinger distance and showed that they can be bounded in terms of the error of the low-fidelity forward model with respect to the high-fidelity.

Another approach to correct the bias is to introduce an error model to account for the error of the surrogate in the inversion. Kaipio and Somersalo (2007) proposed the so-called approximation error model (AEM), where the low-fidelity model performs the inference, but includes an additional term that quantifies the error between the low and high-fidelity models. The error term is considered a random variable and is modeled as an additive Gaussian noise. The AEM is also used in Arridge et al. (2006) for a diffuse optical tomography problem where the low-fidelity model is a coarse grid approximation. Tarvainen et al. (2010) applied the AEM to a simplyfied physics low-fidelity model and Berger et al. (2016) to a PGD reduced model for a heat transfer problem. Cui et al. (2011), instead, proposes to adapt the AEM during the MCMC inversion. Alternative reduction error models have also been proposed. For instance, Yan and Zhou (2019) use a second surrogate to approximate the differences between the full model and the first surrogate, whereas Manzoni et al. (2016) presents a comparison between three reduced error methods: the AEM, a radial basis interpolation of the errors and a log-linear regression model using error bounds.

A complete different approach to control the bias are the multifidelity methods. They use the low-fidelity model to speed up the problem, but keep the high-fidelity to retain accuracy and convergence to the true posterior distribution. A very nice and detailed survey about multifidelity methods can be found in Peherstorfer et al. (2018). Some methods use the low-fidelity as a filter to indicate whether the high-fidelity has to be solved, either because it fulfills some criteria or because the low-fidelity is not accurate enough. Multistage MCMC methods are widly used aiming at increasing the number of samples for a given amount of computational time (Christen and Fox, 2005; Cui et al., 2011, 2014; Efendiev et al., 2006; Laloy et al., 2013). Christen and Fox (2005) proposed a two-stage delayed acceptance method to sample the true posterior distribution. A candidate sample is first tested by the acceptance criteria using the low-fidelity model and, only if accepted, it is then evaluated with the high-fidelity model to decide whether the candidate is finally accepted or rejected. The main advantage is that the first stage discards a lot of samples and it increases the acceptance probability in the second stage. However, the number of high-fidelity models needed in the second stage to guarantee some accuracy is still large which implies a significant computational effort. A similar approach is the one presented by Efendiev et al. (2006) that uses coarse-scale meshes to precondition the MCMC and avoid the cost of a high-fidelity solution on proposals that would be rejected. Cui et al. (2014) uses a two-stage delayed acceptance strategy with a RB method as a low-fidelity model. Unlike the previous case, an initial RB is used to generate many samples using a standard Metropolis Hastings algorithm with the idea of decreasing correlations between samples. The last sample is then used as a candidate for the second stage where the high-fidelity model is used to compute the acceptance probability.

Finally, an interesting strategy which can be used in combination with any of the above is the data-driven approach. Normally surrogates are build offline and are defined on the span of the prior information of the parameters (Wang and Zabaras, 2005; Galbally et al., 2009). Although improvements have been achieved in terms of greedy sampling strategies, for instance by solving model-constrained optimization problems to account for the physics of the forward model (Bui-Thanh et al., 2008; Lieberman et al., 2010) or by refining the training set (Hesthaven et al., 2014), they are still based on prior information, i.e. prior-based surrogates. Since the posterior distribution essentially includes the gain in information with respect to the prior after considering the data, posteriors are expected to concentrate in a much smaller region of the parameter space (Cui et al., 2014; Li and Marzouk, 2014; Yan and Zhou, 2019). Some authors proposed to use the posterior information to adapt surrogates during the MCMC so that accuracy is focused on the posterior region. Therefore, unlike *offline/online* strategies, the low-fidelity models are constructed/refined *on the fly* (Florentin and Díez, 2012; Cui et al., 2014; Yan and Zhou, 2019; Li and Marzouk, 2014; Zhang et al., 2020; Manassero et al., 2020). Cui et al. (2014) implements a data-driven approach based on an error estimator to decide whether the surrogate is accurate enough. If the error is larger than a specified tolerance, the initial reduced basis is updated with the snapshot obtained from sampling the posterior distribution using a Gram-Schmidt procedure. In doing so, the model reduction strategy is tailored to inverse problems since exploration of the posterior and model reduction are pursued simultaneously. Instead, Yan and Zhou (2019) initially generates a certain amount of samples with the surrogate model. Then, the acceptance probability of the last sample is computed with the high-fidelity model and the real error between the low- and high-fidelity model is evaluated to decide whether enrichment is required. If so, enrichment is performed by generating random samples around the candidate and using their high-fidelity solutions to refine the surrogate with local approximations. The adaptive strategy of Zhang et al. (2020) is slightly different. Initially random samples are generated from the prior distribution to build a first surrogate which is used to run the MCMC and approximate the posterior. This approximated posterior is far from the true one, but closer than the prior. Therefore, some random samples are generated from it to update the surrogate and run MCMC to obtain a more accurate posterior approximation. This refinement process is repeated a certain number of times before obtaining an accurate final posterior approximation.

1.3 Objectives and outline

The main goal of this thesis is to couple Bayesian probabilistic inverse methods with Model Order Reduction (MOR) techniques within the context of high-dimensional and large-scale geophysical inverse problems. The role of the MOR is to provide fast and accurate solutions to cope with the large number of costly forward evaluations, a major bottleneck in probabilistic inversions. In particular, the main objectives of this thesis are:

- Understand from a methodological point of view the efficiency and behaviour of the Reduced Basis (RB) acting as a surrogate in the probabilistic formulation of inverse problems that use Markov Chain Monte Carlo (MCMC) methods, specifically, the Metropolis-Hastings algorithm.
- Implement the coupling between the MCMC and RB to reduce the computational cost of probabilistic inversions.
- Apply the methodology to a realistic geophysical inverse problem. The RB is used to approximate the solution of 3D Stokes flow problems which allow the estimation the dynamic topography and, from it, infer the lithospheric structure.

This thesis is organized as follows. **Chapter 2** presents the geophysical problem and the approximations considered in the corresponding modelling equations. **Chapter 3** is devoted to the forward problem. Section 3.1 describes the full order model in the context of Finite Elements (FE) and Section 3.2 introduces the reduced order techniques. In particular, it describes the Reduced Basis method, strategies to quantify the RB error and the classical approaches to construct the basis within an offline/online scheme. **Chapter 4** is devoted to the inverse problem and the probabilistic framework to solve it using the Metropolis-Hastings algorithm. It also includes a first inversion example using the full order model that will be used as a reference in later examples. **Chapter 5** focuses on the coupling between the MCMC and RB. In Section 5.1 a new strategy to build the basis *on the fly* and tailored to the inverse problem is presented and the effect of the RB error in the inversion results is studied with some examples. The remaining sections of the chapter analyse the main aspects involved in the cost of obtaining RB approximations: the basis size and the assembling efficiency of the matrices. Section 5.2 studies the influence of goal-oriented error estimators, Section 5.3 the characteristics of the inverse problem and Section 5.4 focuses on the influence of the MCMC solver and presents additional strategies to control the basis size. Finally, Section 5.5 describes procedures to efficiently assembly the stiffness matrices based on the idea of updating local changes. In **Chapter 6**, the MCMC+RB method is applied to two more realistic examples. First, the dynamic topography is used as observable in a spherical domain representing a portion of Earth. Second, the mantle velocity field is used in a larger inversion problem to determine the lithospheric structure and sublithospheric upper mantle flow beneath the African continent. **Chapter 7** summarises the conclusions obtained in previous chapters and discusses further improvements and possible future work.

Chapter 2

Geophysical problem statement

2.1 Lithosphere and upper-mantle

Plate tectonics is a unifying framework in which to study geodynamic processes. It divides the outer shell of the Earth into a number of thin and rigid plates which are in relative horizontal motion one with respect to the other. They are continuously created at ocean ridges and consumed at approximately the same rate at ocean trenches (Turcotte and Schubert, 2002). At ocean ridges, the plates diverge (seafloor spreading) and the hot mantle ascends to fill the gap, cools and is added to the plates to create new area. Conversely, at ocean trenches, the plates converge and one bends and descends beneath the other (subduction). Nevertheless, determining the detailed present-day physical state of the thermal and compositional structure of the Earth's interior is still an open research topic. Most of the information from the inside must be inferred from indirect observational quantities, physical experiments or by studying rocks which have been brought to the surface due to different internal processes.

The three major divisions on the Earth interior according to the composition are the crust, the mantle and the core. The crust is the first layer that goes from the surface down to an average of 7km below the oceans and 42km below the continents, receiving the name of oceanic and continental crust, respectively. Oceanic crust is young (between 1 and 180Myr) and is constantly created at spreading centers and subducted back to the mantle, whereas continental crust is older (around 2000Myr) and does not usually subduct. The first compositional discontinuity appears below the crust and is called Moho (Mohorovičić discontinuity). It separates the crust which contains silicic (continental) and basaltic (oceanic) rocks from the ultramafic rocks in the mantle (Schubert et al., 2004; Ranalli, 1995). Since the crustal rocks are less dense than the mantle rocks, the crust is gravitationally stable with respect to the mantle. The mantle goes from the Moho down to $\sim$ 2900km. It presents two major seismic discontinuities that separate it in three regions. The upper mantle which goes from the Moho down to 410km. Followed by a transition region ranging from 410km until the second discontinuity at 660km. At such depth there is a phase change with large contrasts of viscosities and densities. This discontinuity defines the lower boundary of the seismicity of the Earth. Below 660km there is no seismic activity and we find the lower mantle which extends down to 2900km. In addition, the subducting slabs have difficulties crossing from the transition zone to the lower mantle which suggests that there must be two different convecting regions in the mantle. Finally, the core is the last part and continues until the center of the Earth.

It is interesting to describe the region composed by the crust and the first hundreds of kilometers of the mantle from a mechanical point of view. In this sense, we can distinguish the lithosphere and the asthenosphere. The lithosphere is the outer shell of the Earth which is made of cool and stiff rocks. The low temperature of the lithosphere allows rocks to resist deformation on time scales of up to 10^9 years and, therefore, it is able to transmit elastic stresses without significant deformation. In contrast, below the lithosphere the rocks are hot enough to deform and flow viscously when subjected to long-time forces. This second part is called asthenosphere. The isotherm at approximately 1400K defines the boundary between the lithosphere and the asthenosphere and is usually named LAB (Lithosphere-Asthenosphere Boundary). Similarly to the crust, the lithosphere can be split into the continental lithosphere which has an average thickness of 200km and the oceanic lithosphere with an average thickness of 100km. The lithosphere includes the crust and the uppermost mantle and, therefore, it is composed of both crustal and mantle rocks. Although the crust is gravitationally stable with respect to the mantle due to its buoyant rocks, the thickness of the oceanic crust is not enough to prevent the subduction of the oceanic lithosphere. Therefore, the oceanic lithosphere can be seen as the upper thermal layer of the mantle convection system. Instead, the thickness of the continental crust together with its buoyancy (lower density) is enough for the continental lithosphere to remain above the mantle (Turcotte and Schubert, 2002).

There is a characteristic decoupling on the Earth behaviour between short and long timescales. Over long timescales (geological timescales, $> 10^4$ years) the mantle behaves as a viscous fluid and its internal deformation mechanism is the so-called solid-state creep. Instead, on short timescales (human timescales, seconds to days), the mantle exhibits an elastic behaviour which allows the propagation of seismic waves. Regarding the upper lithosphere, due to its low temperature and pressure, it shows elastic rheology even over long time scales. Therefore, the big picture of the dynamics of the Earth at geological timescales is of blocks of lithosphere floating on a fluid mantle.

The solid-state creep is the main deformation mechanism of the mantle of the Earth. It is a thermally activated process which consists in the motion of atoms and ions in a crystalline lattice when subjected to stresses resulting in a slow deformation (like a fluid). There are two types: diffusion creep and dislocation creep. Difussion creep occurs predominantly at very low stress levels. It is the result of diffusion of atoms through the interior and along the boundaries of crystalline grains due to the presence of vacancies (empty sites in the crystalline lattice). At any given nonzero temperature, in a crystalline lattice there is an equilibrium concentration of vacancies and this concentration is temperature dependent. When thermally activated, atoms migrate due to the movement of adjacent vacancies and, therefore, grains deform leading to rock strain (Schubert et al., 2004; Ranalli, 1995). Diffusion creep corresponds to a Newtonian fluid behaviour, in which strain rate and shear stress are linearly related. Dislocation creep is dominant at high stresses and is the result of migration of dislocations which are imperfections (line or one-dimensional) in the crystalline lattice structure. Since dislocation density strongly depends on stresses, it results in a non-linear (non-Newtonian) relationship between the strain rate and stress.

The physical properties of rocks which control the mechanical behaviour of the mantle and lithosphere (density and viscosity) strongly depend on temperature. Therefore, it is essential to characterise its internal thermal state. Observations of the surface heat flow from the Earth and of plate movements, imply that by conservation laws there must be large-scale movements in the mantle with internal heat transfer. Such heat transfer cannot only be explained by conduction and it requires convection to be the dominating mechanism of heat transfer (Ranalli, 1995; Turcotte and Schubert, 2002). The thermal convection of the mantle is a consequence of the non-equilibrium density configurations within the mantle which produce gravitational instabilities. In other words, density variations produced by thermal expansion or contraction originate buoyancy forces that drive convection. A material which is heated from below expands and ascends due to its lower density with respect to the surrounding. Then, it releases heat to the surface, cools down and due to its now higher density descends to close the cycle. Generally, inside a convecting fluid, the mean temperature increases with depth following approximately an adiabatic gradient, i.e. the temperature increases due to the compression (by the weight from the overlying rocks), but there is no heat transfer with the surrounding during this process. The temperature gradients near the surface are very high around 25K/km and the temperature distribution is mostly by conduction. All these processes produce very complex structures which are difficult to characterise.

2.2 Conservation and constitutive equations

The starting point to study the behaviour of a body is by stating the appropriate conservation equations. In the case of thermal convection: mass conservation and momentum conservation. The mass conservation leads to the continuity equation and the momentum conservation results in the equation of motion. However, the equation of motion is general for any continuous media and, therefore, rheological properties are required in order to completely characterise the behaviour. This is precisely the role of the rheological equation of state or constitutive equation. The constitutive equation relates dynamic (e.g. forces of various types) with kinematic (e.g. displacement, velocity and acceleration) quantities or states of a material through some characteristic material parameters in order to characterise its flow or deformation behaviour.

Mass conservation. According to the laws of Newtonian mechanics, in any closed volume the mass must be conserved. The mass balance can be expressed with the *continuity equation*,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0, \tag{2.1}$$

with ρ being the density, $\boldsymbol{u}$ the velocity field and t the time. In the case of an incompressible fluid, the density is constant and its dependency with time vanishes. Therefore, equation (2.1) reduces to the divergence free of the velocity,

$$\nabla \cdot \boldsymbol{u} = 0. \tag{2.2}$$

Momentum conservation. The second law of Newtonian mechanics applied to a fluid states that in a defined portion of volume of fluid, the sum of all forces acting on it equals its change of momentum. Two different kind of forces can act in a fluid: volume and surface forces. *Volume* or *body forces* correspond to long-range external forces which penetrate matter and act equally in all points within an element of fluid, for instance the gravity. *Surface forces* are short-range molecular forces which are internal to the fluid and their net effect is on a thin surface layer. The force balance yields to the equation of motion of a fluid

$$\rho \left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \boldsymbol{u} \right) = \rho \boldsymbol{g} + \nabla \cdot \boldsymbol{\sigma}, \tag{2.3}$$

where $\boldsymbol{\sigma}$ is the stress tensor and $\boldsymbol{g}$ the gravity vector and they account for surface and body forces per unit volume, respectively. The term on the left is the acceleration of an element of fluid times the density. The derivation of equation (2.3) can be found elsewhere (Ranalli, 1995; Batchelor, 2002; Schubert et al., 2004; Donea and Huerta, 2003).

Constitutive equation. As previously mentioned, at geological timescales the mantle behaves as a highly viscous fluid. Since it implies that the deformation changes continuously with time, it is natural to use the velocity as a kinematic quantity instead of the displacement (as in elasticity). Therefore, the dynamic and kinematic states are characterised with the stress tensor and strain rate tensors, respectively, and the material parameter is the viscosity. The viscosity is a measure of the resistance of a fluid to flow or, in other words, of the internal friction between contiguous layers of fluid when moving one with respect to the other.

Stress tensor. The surface forces acting on a fluid are divided in two: *hydrostatic forces* associated with fluid pressure which are perpendicular to the surface and characterise the magnitude of extension/compression across the surface; and *deviatoric forces* which act both perpendicular and parallel to the surface and are caused by the velocity gradients in the fluid due to viscosity. When a fluid is at rest, the only surface force is the pressure and, therefore, the stress tensor is isotropic with only normal stresses,

$$\sigma_{ij} = -p\delta_{ij}, \tag{2.4}$$

where $p = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33})$ is the static fluid pressure (mean normal stress) and δ_{ij} the Kronecker delta. When the fluid is in motion, deviatoric forces appear and the *stress tensor* can now be decomposed into the isotropic and deviatoric parts

$$\sigma_{ij} = -p\delta_{ij} + \tau_{ij}, \tag{2.5}$$

with τ_{ij} known as the *deviatoric stress tensor*. The deviatoric stresses acting parallel to a surface are known as *shear stresses* and they coincide with the off-diagonals of the total stress, since $-p\delta_{ij} = 0$ for $i \neq j$.

Strain rate. The dynamics of the changes of the internal deformation, i.e. rate of deformation, are defined in terms of the spatial derivatives of the velocity. The gradient of the velocity $\nabla \boldsymbol{u}$ is a second-order tensor defined as,

$$\nabla \boldsymbol{u} = \begin{pmatrix} \dfrac{\partial u_1}{\partial x_1} & \dfrac{\partial u_1}{\partial x_2} & \dfrac{\partial u_1}{\partial x_3} \\ \dfrac{\partial u_2}{\partial x_1} & \dfrac{\partial u_2}{\partial x_2} & \dfrac{\partial u_2}{\partial x_3} \\ \dfrac{\partial u_3}{\partial x_1} & \dfrac{\partial u_3}{\partial x_2} & \dfrac{\partial u_3}{\partial x_3} \end{pmatrix}, \tag{2.6}$$

where u_i are the components of the velocity field and x_j are the spacial dimensions. It can be decomposed into its symmetric and skew-symmetric parts, $\nabla \boldsymbol{u} = \nabla^{\mathsf{S}} \boldsymbol{u} + \nabla^{\mathsf{W}} \boldsymbol{u}$ with,

$$\nabla^{\mathsf{S}} \boldsymbol{u} := \frac{1}{2}(\nabla \boldsymbol{u} + \nabla^{\mathsf{T}} \boldsymbol{u}) \qquad \text{and} \qquad \nabla^{\mathsf{W}} \boldsymbol{u} := \frac{1}{2}(\nabla \boldsymbol{u} - \nabla^{\mathsf{T}} \boldsymbol{u}). \tag{2.7}$$

The symmetric part is the so-called strain rate tensor, whereas the skew-symmetric part is the spin tensor. The components of the *strain rate tensor*, $\dot{\varepsilon}_{ij}$, are defined as,

$$\dot{\varepsilon}_{ij} = \left[\nabla^{\mathsf{S}} \boldsymbol{u}\right]_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{2.8}$$

As explained in the previous section, in geodynamics there are two main mechanisms of rock deformation: diffusion creep and dislocation creep. Each of them has its constitutive equation. Difussion creep corresponds to a Newtonian fluid behaviour in which strain rate is linearly proportional to stress. The most general rheological law for *Newtonian fluids* is,

$$\sigma_{ij} = -p\delta_{ij} + \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + \lambda\frac{\partial u_k}{\partial x_k}\delta_{ij} \tag{2.9}$$

where $\lambda = \mu^* - \frac{2}{3}\mu$, and μ is the dynamic shear viscosity and μ^* is the bulk or volume viscosity of the fluid. Using equation (2.8), the stress tensor can be rewritten in a more compact form as,

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\dot{\varepsilon}_{ij} + \lambda\dot{\varepsilon}_{kk}\delta_{ij} \tag{2.10}$$

where it is interesting to note that $\dot{\varepsilon}_{kk}$ corresponds to $\nabla \cdot \boldsymbol{u}$. If the fluid is incompressible, $\dot{\varepsilon}_{kk} = 0$, the last term of previous equation vanishes and the total stress yields to the so-called *Stokes law*,

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\dot{\varepsilon}_{ij}. \tag{2.11}$$

The dynamic viscosity can be obtained from the ratio of shear stresses and shear strain rates, or equivalently from the corresponding deviatoric parts, as,

$$\mu = \frac{\sigma_{ij}}{2\dot{\varepsilon}_{ij}} \qquad i \neq j \qquad \text{or} \qquad \mu = \frac{\tau_{ij}}{2\dot{\varepsilon}'_{ij}}, \tag{2.12}$$

where $\dot{\varepsilon}'_{ij}$ is the deviatoric strain rate tensor defined as $\dot{\varepsilon}'_{ij} = \dot{\varepsilon}_{ij} - \frac{1}{3}\dot{\varepsilon}_{kk}\delta_{ij}$. For an incompressible fluid, the last term vanishes and $\dot{\varepsilon}'_{ij} = \dot{\varepsilon}_{ij}$.

On the other hand, dislocation creep strongly depends on stresses and it results in a non-Newtonian behaviour. There is no unique nonlinear rheological equation, but the most important one is the so-called power-law creep equation in which the strain rate is proportional to the nth-power of the stress ($n > 1$). The reader is referred to Ranalli (1995) and Schubert et al. (2004) for detailed explanations. In this work we will focus on the Newtonian case.

A viscosity can analogously be defined for non-Newtonian fluids. However, unlike the Newtonian case, it is not only function of temperature, pressure and material parameters, but it is also function of the stress. Therefore, the viscosity of a Non-Newtonian fluid can not be fully characterised by just fixing parameters, but it has to be defined at a given stress or strain rate.

2.3 Quasi-static Stokes flow

The conservation of mass and momentum given by equations (2.1) and (2.3) respectively define the so-called Navier-Stokes equations. However, depending on the process to be modeled, some simplifications can be performed according to the relative importance of the terms involved in the equations. A dimensional analysis using nondimensional numbers to justify the following approximations can be found in Schubert et al. (2004), Ranalli (1995) or Anderson (1989). The mechanical behaviour of the upper and transition zone of the mantle is modeled as a quasi-static flow of a Newtonian, incompressible, viscous fluid by means of the Stokes equations.

In the long timescales and in absence of phase transformations, the density variations of the rocks are small enough for the fluid to be considered incompressible. However, the density variations caused by thermal expansion are the ones responsible for the buoyancy forces that drive convection and they must be considered. Consequently, the incompressibility condition must be relaxed, it cannot be applied in the usual way (neglecting all density dependencies). In this sense, incompressibility is applied in terms of the *Boussinesq* approximation. It states that density variations are only accounted for in the buoyancy terms, i.e. the gravitational body force term of the momentum equation, everywhere else density is considered constant. Or in other words, density variations are so small that only when combined with the gravity field become significant. The general continuity equation (2.1) for an incompressible fluid reduces to (2.2); the divergence free of the velocity, $\nabla \cdot \boldsymbol{u} = 0$.

The mantle is also considered to be a Newtonian fluid. Therefore, the constitutive relation between stress and strain rate is linear and given by equation (2.10). Since the fluid is incompressible, it simplifies to the Stokes law from equation (2.11) that in matrix form is,

$$\boldsymbol{\sigma} = -p\mathbf{I} + 2\mu\nabla^{\mathrm{s}}\boldsymbol{u}. \tag{2.13}$$

Replacing the linear Stokes law in the momentum equation (2.3) gives,

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t}+(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right)=\rho\boldsymbol{g}+\nabla\cdot\big(2\mu\nabla^{\mathrm{s}}\boldsymbol{u}\big)-\nabla p. \tag{2.14}$$

This equation represents a balance between inertial forces, viscous forces, pressure forces and the body force due to the gravity field for an incompressible, Newtonian, viscous fluid.

In the case of highly viscous flow, the convective term $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}$ is neglected with respect to the viscous and gravitational forces (Batchelor, 2002; Donea and Huerta, 2003). Moreover, since the fluid behaviour of the mantle is exhibited at such long time scales (creeping flow), the flow is slow enough to assume that at a given instant in time the problem is static and remains in internal equilibrium. Consequently, the time derivative of the velocity (acceleration) can also be neglected and the mantle is approximated as a quasi-static flow. This approximation is called the *infinite Prandtl number* (Schubert et al., 2004; Turcotte and Schubert, 2002). The resulting equation is elliptic and its lack of time dependency implies that changes in a point of the domain have an immediate influence everywhere else. Finally, the whole inertia force vanished and the resulting quasi-static Stokes flow is given by,

$$-\nabla\cdot\big(2\mu\nabla^{\mathrm{s}}\boldsymbol{u}\big)+\nabla p=\rho\boldsymbol{g}. \tag{2.15}$$

In the case of the Earth, viscosity is not constant since it depends on temperature, pressure and (in nonlinear rheological models) on the gradient of the velocity and, hence it is a function of position.

2.4 Dynamic topography

In general, characteristics of the rheology of the mantle can be inferred from any internal process that generates observable quantities at the surface which are function of the rheology (Ranalli, 1995). For instance, with postglacial rebound studies it is possible to infer an average value of the mantle viscosity. During glaciations periods there were large loads of ice covering the Earth surface. When glaciers melted and the load was removed, the isostatic equilibrium was not compensated anymore and the surface of the Earth raised in order to recover equilibrium. From rebound data and assuming the mantle is a viscous fluid, Haskell (1935) was able to infer an average mantle viscosity of $\mu = 10^{21}$Pa·s. After years of studies this value prevailed.

The topography or elevation of the Earth is one of such observables. It is mainly explained by the isostasy principle which assumes a series of rigid vertical columns of lithosphere floating freely on a fluid asthenosphere or sublithospheric mantle. It considers that there is a depth within the Earth, known as compensation depth, at which the vertical stresses due to an overlying column of rocks are equal. In other words, at this depth, all columns weight the same. According to that, the elevation is controlled by the density structure of the upper, conductive and thermal boundary layer, i.e. by the buoyancy of the lithosphere. Most of the topography on the Earth can be attributed to isostatic compensation (Turcotte and Schubert, 2002; Stüwe, 2007).

Although most of the lithosphere is in isostatic equilibrium, some regions cannot be explained only by isostasy and dynamical effects must be taken into account. The dynamic topography is the elevation produced by the Earth's mantle flow. Due to the low viscosity of the sublithospheric mantle (compared to the viscosity in the lithosphere), the density variations in the sublithospheric mantle produce lateral pressure gradients which can evolve into flow over much shorter time scales than the ones in the lithosphere (they can last much longer). Such mantle flow is transferred to the surface via normal viscous stresses which are known as dynamical effects. This creates a mechanical and temporal decoupling between the long-term rigid lithosphere and much less viscous sub-lithospheric mantle. A possible example is the African superswell. Lithgow-Bertelloni and Silver (1998) showed that the excess of elevation of the Africa superswell can be dynamically supported by the mantle flow upwelling induced by the density variations beneath the African plate.

Stresses are assumed to be transmitted instantaneously and equilibrium is reached. Therefore, the forces acting on a column of fluid must balance: the normal viscous stresses at the surface σ_{nn} coming from the mantle flow are balanced with the lithostatic stresses (or lithostatic pressure) of the overlaying material column,

$$\sigma_{nn} = \rho g w \tag{2.16}$$

where w is the elevation or topography, ρ the density of the column of material above and g the gravity. The convection of the mantle also produces shear stresses at the base of the lithosphere that would contribute to the dynamic topography. However, we will not consider them since their contribution is typically less than 5% of the topography generated by normal stresses (Marquart and Schmeling, 1989). According to Marquart and Schmeling (1989), the elasticity of the lithosphere is not important in thinner regions like the oceanic lithosphere, whereas in the case of thick continental areas, it may significantly contribute to decrease the topography.

Computing the surface topography is a common geodynamic problem and different implementations are widely applied. Typically, the normal stress method is used which imposes a free slip condition on the upper surface and the generated stresses are used to compute the topography that would balance them (Mckenzie, 1977; Zhong et al., 1993; Blankenbach et al., 1989). Body-fitting methods consider the topography a true free surface and follow its deformation with the numerical mesh. This strategy has the advantage that the natural zero normal stress can be applied on the deforming surface (Poliakov and Podladchikov, 1992; Fullsack, 1995). Other methods use an Eulerian grid for the flow problem and a different discretization for tracking the free surface (Harlow and Welch, 1965; Braun et al., 2008). Another approach to approximate a free surface is the so-called "sticky-air" method which mimics a free-surface by extending the domain and including a fluid layer of a low density and low viscosity on top so that it allows free movements of the interface (Crameri et al., 2012). See Crameri et al. (2012) for a detailed comparison between different approaches and additional references. We will use the normal stress method since it is simpler and computationally cheaper. An explanation of the modelling implementation to compute the dynamic topography is detailed in Section 6.1.

Chapter 3

The Forward Problem and Reduced Basis Method

This chapter is devoted to the strategies to solve the forward problem that describes the dynamical behaviour of mantle flow. It is divided in two parts: the full order model (Section 3.1) and the reduced order model (Section 3.2).

In the first part, the full order model is presented and solved with the Finite Element Method (FEM) which is a general technique to approximate solutions of partial differential equations. Its main idea consists in dividing the domain of the solution into a finite number of subdomains, the so-called finite elements, and use a weak statement to build an approximate solution over the set of subdomains. The second part focuses on the reduced order model and its low-cost computation with respect to the full order model. It describes the Reduced Basis (RB) strategy: how to construct a reduced approximation, how to asses the error of the RB solution with respect to FE and some classic strategies to construct the reduced basis in an offline fashion. Finally, the need of a strategy that takes advantage of the inversion scheme to build a tailored basis by combining the offline and online stages is here briefly introduced and described in detail in Chapter 5.

3.1 Problem statement and full order solution

A first-order description of the dynamic behaviour of the mantle and its effect on elevation starts with the formal statement of a Stokes problem (cf. Schubert et al., 2004)

$$-\nabla \cdot (2\mu\nabla^{\mathrm{s}}\boldsymbol{u}) + \nabla p = \rho\boldsymbol{g} \qquad \text{in } \Omega \tag{3.1a}$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad \text{in } \Omega \tag{3.1b}$$

$$\boldsymbol{u} = \boldsymbol{u}_D \qquad \text{on } \Gamma_D \tag{3.1c}$$

$$-p\boldsymbol{n} + 2\mu\boldsymbol{n} \cdot \nabla^{\mathrm{s}}\boldsymbol{u} = \boldsymbol{t} \qquad \text{on } \partial\Omega \setminus \Gamma_D \tag{3.1d}$$

where $\boldsymbol{u}$ and p are velocity and pressure, respectively, $\Omega \subset \mathbb{R}^3$ is the computational domain and $\partial\Omega$ its boundary. The latter is partitioned into Dirichlet, Γ_D, and Neumann boundary, Γ_N, types. Body forces are given by $\rho\boldsymbol{g}$, being ρ the density and $\boldsymbol{g}$ the gravity vector. Dirichlet and Neumann boundary conditions are set by $\boldsymbol{u}_D$ and $\boldsymbol{t}$. Dynamic viscosity μ is also considered part of input data and, as mentioned before, both density and viscosity vary in space. In what follows, as in (3.1), we indicate vectors with lowercase bold letters.

Some numerical difficulties arise due to the incompressibility condition in (3.1b), i.e. the velocity field must be divergence free. This constrain on the velocity, forces the pressure to be considered a variable not related to any constitutive equation. The purpose of pressure in the momentum equation is to introduce a new degree of freedom to satisfy the incompressibility constrain. In other words, the pressure is meant to adjust itself instantaneously to satisfy the condition of divergence free of the velocity. It is acting as a Lagrange Multiplier of the incompressibility constrain and, therefore, velocity and pressure unknowns are coupled (Donea and Huerta, 2003).

The variable formulation that is presented next keeps velocity and pressure as unknowns resulting in the so-called *mixed finite element methods*. They entail some numerical difficulties, e.g LBB compatibility condition, that we will address later. Other formulations have been proposed, for instance penalty methods that allow to uncouple velocity and pressure fields. They eliminate the pressure variable from the momentum equation through a relaxation of the incompressibility condition. However, they introduce a penalty parameter that may cause loss of accuracy for very large values and prevent convergence for insufficient large ones.

3.1.1 Weak formulation

In order to obtain the weak form, we need to define a set of weighting functions as well as trial solutions for the pressure and velocity fields. For the velocity,

$$\mathcal{S} := \{u_i \in \mathcal{H}^1(\Omega) | u_i = u_{Di} \text{ on } \Gamma_D\} \qquad \text{(trial solutions)} \tag{3.2a}$$

$$\mathcal{V} := \{w_i \in \mathcal{H}^1(\Omega) | w_i = 0 \text{ on } \Gamma_D\} \qquad \text{(weighting functions)} \tag{3.2b}$$

where u_i and w_i, for $i = 1, ..., n_{sd}$, are the components of $\boldsymbol{u}$ and $\boldsymbol{w}$, respectively, and each of them belongs to $\mathcal{H}^1(\Omega)$. For $n_{sd} = 3$ spatial dimensions, $\boldsymbol{\mathcal{S}} := [\mathcal{S}]^3$ is an affine space and $\boldsymbol{\mathcal{V}} := [\mathcal{V}]^3$ a vectorial one. Instead, for the pressure, both trial solutions and weighting functions belong to the same space $\mathcal{Q}$,

$$\mathcal{Q} := \{q \in \mathcal{L}_2(\Omega)\} \qquad \text{(trial and weighting functions)} \tag{3.3}$$

We multiply the momentum equation (3.1a) by the test function $\boldsymbol{w}$ and integrate over the domain Ω,

$$-\int_\Omega \boldsymbol{w} \cdot (\nabla \cdot (2\mu \nabla^{\mathrm{s}} \boldsymbol{u}))\, d\Omega + \int_\Omega \boldsymbol{w} \cdot \nabla p\, d\Omega = \int_\Omega \boldsymbol{w} \cdot \rho \boldsymbol{g}\, d\Omega. \tag{3.4}$$

We integrate by parts the first two terms: applying the product rule yields

$$\begin{cases} \nabla \cdot (\boldsymbol{w} \cdot 2\mu \nabla^{\mathrm{s}} \boldsymbol{u}) = \nabla \boldsymbol{w} : 2\mu \nabla^{\mathrm{s}} \boldsymbol{u} + \boldsymbol{w} \cdot (\nabla \cdot (2\mu \nabla^{\mathrm{s}} \boldsymbol{u})) \\ \nabla \cdot (\boldsymbol{w} p) = (\nabla \cdot \boldsymbol{w}) p + \boldsymbol{w} \cdot \nabla p \end{cases} \tag{3.5}$$

where the following notation has been introduced

$$\nabla \boldsymbol{w} : \nabla^{\mathrm{s}} \boldsymbol{u} = \frac{1}{2} \sum_{i=1}^{n_{sd}} \sum_{j=1}^{n_{sd}} \frac{\partial w_i}{\partial x_j} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$$

and replacing them into equation (3.4) gives,

$$\int_\Omega \nabla \boldsymbol{w} : 2\mu \nabla^{\mathrm{s}} \boldsymbol{u}\, d\Omega - \int_\Omega \nabla \cdot (\boldsymbol{w} \cdot 2\mu \nabla^{\mathrm{s}} \boldsymbol{u})\, d\Omega - \int_\Omega (\nabla \cdot \boldsymbol{w}) p\, d\Omega + \int_\Omega \nabla \cdot (\boldsymbol{w} p)\, d\Omega = \int_\Omega \boldsymbol{w} \cdot \rho \boldsymbol{g}\, d\Omega.$$

The divergence theorem is then applied to the second and fourth terms. Moreover, since we defined the test functions $\boldsymbol{w}$ such that they vanish on the Dirichlet boundary, only the Neumann part of the second and fourth terms remains,

$$\int_\Omega \nabla \boldsymbol{w} : 2\mu \nabla^{\mathrm{s}} \boldsymbol{u}\, d\Omega - \int_{\Gamma_N} (\boldsymbol{w} \cdot 2\mu \nabla^{\mathrm{s}} \boldsymbol{u}) \cdot \boldsymbol{n}\, d\Gamma - \int_\Omega (\nabla \cdot \boldsymbol{w}) p\, d\Omega + \int_{\Gamma_N} \boldsymbol{w} p \cdot \boldsymbol{n}\, d\Gamma = \int_\Omega \boldsymbol{w} \cdot \rho \boldsymbol{g}\, d\Omega.$$

Finally, rearranging terms and recalling the Neumann boundary conditions,

$$\int_\Omega \nabla \boldsymbol{w} : 2\mu \nabla^{\mathsf{S}} \boldsymbol{u} \; d\Omega - \int_\Omega p(\nabla \cdot \boldsymbol{w}) \; d\Omega = \int_\Omega \boldsymbol{w} \cdot \rho \boldsymbol{g} \; d\Omega + \int_{\Gamma_N} \boldsymbol{w} \cdot \boldsymbol{t} \; d\Gamma. \tag{3.6}$$

Similarly, the incompressibility equation (3.1b) is multiplied by the pressure test function q and integrated over the domain Ω,

$$\int_\Omega q(\nabla \cdot \boldsymbol{u}) \; d\Omega = 0. \tag{3.7}$$

The weak form for problem (3.1) reads: *find a velocity field* $\boldsymbol{u} \in \boldsymbol{\mathcal{S}}$ *and a pressure field* $p \in \mathcal{Q}$ *such that,*

$$a\big(\boldsymbol{w}, \boldsymbol{u}\big) + b\big(\boldsymbol{w}, p\big) = l\big(\boldsymbol{w}\big) \qquad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{V}}, \tag{3.8a}$$

$$b\big(\boldsymbol{u}, q\big) = 0 \qquad \forall q \in \mathcal{Q}, \tag{3.8b}$$

with the operators defined as,

$$a\big(\boldsymbol{w}, \boldsymbol{u}\big) = 2 \int_\Omega \nabla^{\mathsf{S}} \boldsymbol{w} : \mu \nabla^{\mathsf{S}} \boldsymbol{u} \; d\Omega \tag{3.9a}$$

$$b\big(\boldsymbol{u}, q\big) = - \int_\Omega q(\nabla \cdot \boldsymbol{u}) \; d\Omega \tag{3.9b}$$

$$\ell\big(\boldsymbol{w}\big) = \int_\Omega \boldsymbol{w} \cdot \rho \boldsymbol{g} \; d\Omega + \int_{\Gamma_N} \boldsymbol{w} \cdot \boldsymbol{t} \; d\Gamma. \tag{3.9c}$$

Note that the gradient $\nabla \boldsymbol{w}$ has been replaced by its symmetric part $\nabla^{\mathsf{S}} \boldsymbol{w}$. The double contraction of a skew-symmetric tensor with a symmetric one vanishes. Hence, in replacing $\nabla \boldsymbol{w}$ by $\nabla^{\mathsf{S}} \boldsymbol{w} + \nabla^{\mathsf{W}} \boldsymbol{w}$ only the symmetric part remains.

Sometimes, it is useful to rewrite the bilinear form $a\big(\boldsymbol{w}, \boldsymbol{u}\big)$ in terms of the strain rate vector $\dot{\boldsymbol{\epsilon}}(\boldsymbol{u})$ and the constitutive matrix $\mathbf{C}_\mu$ as,

$$a\big(\boldsymbol{w}, \boldsymbol{u}\big) = \int_\Omega \dot{\boldsymbol{\epsilon}}(\boldsymbol{w})^{\mathsf{T}} \mathbf{C}_\mu \dot{\boldsymbol{\epsilon}}(\boldsymbol{u}) d\Omega \tag{3.10}$$

where in 3D, the strain rate vector is defined as,

$$\dot{\boldsymbol{\epsilon}}(\boldsymbol{u})^{\mathsf{T}} = \left(\frac{\partial u_1}{\partial x_1}, \frac{\partial u_2}{\partial x_2}, \frac{\partial u_3}{\partial x_3}, \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}, \frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2}, \frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3} \right) \tag{3.11}$$

and the constitutive matrix as,

$$\mathbf{C}_\mu = \mu\mathbf{C} = \mu \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.12}$$

The strain rate vector makes use of the symmetry of the strain rate tensor (2.8) to express it in a vector form that only contains the six relevant components (Donea and Huerta, 2003).

3.1.2 Space discretization

The weak form is now discretized by means of the Galerkin finite element method. When applied to the Stokes problem, with velocity and pressure as unknowns, it leads to a mixed finite element method. The domain Ω is discretized in smaller domains Ω^e (finite elements), resulting in a mesh composed of nodes and elements. The velocity $\boldsymbol{u}$ and pressure p are approximated inside each element as well as their corresponding weighting functions $\boldsymbol{w}$ and q. Such approximation is performed in terms of basis functions and their associated nodal values. The finite dimensional spaces $\mathcal{S}^h$, $\mathcal{V}^h$ and $\mathcal{Q}^h$ are subsets that approximate the spaces $\mathcal{S}$, $\mathcal{V}$ and $\mathcal{Q}$ previously defined in (3.2) and (3.3). Like $\boldsymbol{\mathcal{S}}$, the discretized version $\boldsymbol{\mathcal{S}}^h := [\mathcal{S}^h]^3$ is also an affine space. Similarly, the vectorial space $\boldsymbol{\mathcal{V}}$ is approximated by $\boldsymbol{\mathcal{V}}^h := [\mathcal{V}^h]^3$. For the sake of simplifying notation, we will consider homogeneous Dirichlet boundary conditions, $\boldsymbol{u}_D = 0$, and hence $\boldsymbol{\mathcal{S}}^h = \boldsymbol{\mathcal{V}}^h$. The non-homogeneous case does not bring any other complication, see for instance the derivation in Donea and Huerta (2003).

The approximation of the velocity is constructed using a large number of basis functions, the so-called *shape functions*, $N_i(\boldsymbol{x})$ for $i = 1, \dots, n_u$, that generate the discrete sub-space $\mathcal{V}^h \in \mathcal{V}$,

$$\mathcal{V}^h = \text{span}\{N_1, \dots, N_{n_u}\} \tag{3.13}$$

with n_u being the number of velocity nodes in the mesh. The vector field $\boldsymbol{u}(\boldsymbol{x})$ is approximated by $\boldsymbol{u}^h(\boldsymbol{x}) \in [\mathcal{V}^h]^3$, in such a way that each component of the velocity field belongs to the space $\mathcal{V}^h$. The approximated field is represented by the vector of nodal values $\mathbf{u} \in \mathbb{R}^{3n_u}$,

$$\boldsymbol{u}(\boldsymbol{x}) \approx \boldsymbol{u}^h(\boldsymbol{x}) = \sum_{i=1}^{n_u} N_i(\boldsymbol{x}) \begin{bmatrix} [\mathbf{u}]_{3(i-1)+1} \\ [\mathbf{u}]_{3(i-1)+2} \\ [\mathbf{u}]_{3(i-1)+3} \end{bmatrix} \tag{3.14}$$

where $[\mathbf{u}]_j$, for $j = 1, \ldots, 3n_u$, are the nodal velocities. Note that $\boldsymbol{u}$ is a function whereas $\mathbf{u}$ stands for a vector. Similarly, the shape functions to approximate the pressure are $\tilde{N}_i(\boldsymbol{x})$ for $i = 1, ..., n_p$ and they generate the discrete sub-space $\mathcal{Q}^h \in \mathcal{Q}$,

$$\mathcal{Q}^h = \operatorname{span}\{\tilde{N}_1, \ldots, \tilde{N}_{n_p}\}. \tag{3.15}$$

with n_p the number of pressure nodes. The scalar field $p(\boldsymbol{x})$ is approximated by $p^h(\boldsymbol{x}) \in \mathcal{Q}^h$ and represented by the vector of nodal values $\mathbf{p} \in \mathbb{R}^{n_p}$,

$$p(\boldsymbol{x}) \approx p^h(\boldsymbol{x}) = \sum_{i=1}^{n_p} \tilde{N}_i(\boldsymbol{x})[\mathbf{p}]_i \tag{3.16}$$

where $[\mathbf{p}]_i$ are the nodal pressures. In the Galerkin formulation, the same basis functions used to interpolate the space are also used as weighting functions. Therefore, $\boldsymbol{w}(\boldsymbol{x})$ and $q(\boldsymbol{x})$ are approximated as $\boldsymbol{w}^h(\boldsymbol{x}) \in [\mathcal{V}^h]^3$ and $q^h(\boldsymbol{x}) \in \mathcal{Q}^h$, respectively. Again, each component of the weighting function belongs to $\mathcal{V}^h$.

Introducing these approximations into the weak form and choosing as many weighting function as unknowns results in a linear system of algebraic equations. Note that for the velocity field, the unknowns are vectors and each unknown component has its own equation. The matrix system of algebraic equations is,

$$\begin{bmatrix} \mathbf{K} & \mathbf{G} \\ \mathbf{G}^\mathsf{T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \tag{3.17}$$

where $\mathbf{K}$ is the symmetric stiffness or viscosity matrix with dimensions $(n_u n_{sd}) \times (n_u n_{sd})$, since the velocity vector $\mathbf{u}$ also accounts for the $n_{sd} = 3$ spatial dimensions (see equation (3.14)). $\mathbf{G}$ is the discrete gradient operator and has dimensions $(n_u n_{sd}) \times n_p$ since the pressure vector $\mathbf{p}$ represents a scalar field and $\mathbf{G}^\mathsf{T}$ is its transpose, the discrete divergence operator. Finally $\mathbf{f}$ is the discrete body force vector with dimensions $(n_u n_{sd}) \times 1$. This matrix system is symmetric, with 2×2 blocks and a null submatrix on the diagonal. This particular shape is caused by the saddle-point nature of the variational problem. In order to solve this system, the spaces for velocity and pressure interpolation cannot be freely chosen, they must satisfy a compatibility condition known as LBB or inf-sup condition (LBB stands for Ladyzhenskaya (1969), Babuska (1970) and Brezzi (1974), who determined it). Choosing the appropriate pairs of spaces guarantees the stability of a mixed method. A detailed discussion can be found elsewhere (e.g. Brezzi and Fortin, 1991).

One interesting feature of the integral or weak form is that it has the additive property. Therefore, when the domain Ω is discretized in smaller domains Ω^e, the integral form in (3.4) can be rewritten as the sum of elemental integrals,

$$-\sum_e \int_{\Omega^e} \boldsymbol{w} \cdot (\nabla \cdot (2\mu \nabla^{\mathsf{S}} \boldsymbol{u}))\, d\Omega + \sum_e \int_{\Omega^e} \boldsymbol{w} \cdot \nabla p \, d\Omega = \sum_e \int_{\Omega^e} \boldsymbol{w} \cdot \rho \boldsymbol{g} \, d\Omega. \tag{3.18}$$

This is the basis of the assembly process in which global matrices $\mathbf{K}$, $\mathbf{G}$ and $\mathbf{f}$ are obtained from the assembly of elemental matrices $\mathbf{K}^e$, $\mathbf{G}^e$ and $\mathbf{f}^e$, respectively:

$$\mathbf{K} = \mathsf{A}^e \mathbf{K}^e, \quad \mathbf{G} = \mathsf{A}^e \mathbf{G}^e, \quad \mathbf{f} = \mathsf{A}^e \mathbf{f}^e. \tag{3.19}$$

The job of the assembly operator A^e is nothing else but to add up the coefficients of the elemental matrices in the corresponding positions in the global matrices.

The elemental matrix $\mathbf{K}^e$ is a square symmetric positive definite matrix with dimension $(n_{eu}n_{sd}) \times (n_{eu}n_{sd})$ with n_{eu} being the number of elemental velocity nodes. Using the expression of bilinear form in terms of the strain rate vector from (3.10), the elemental stiffness matrix can be computed as,

$$\mathbf{K}^e = \int_{\Omega^e} \mathbf{D}^{e^\mathsf{T}} \mathbf{C}_{\mu^e} \mathbf{D}^e \, d\Omega \tag{3.20}$$

with

$$\mathbf{D}^{e^\mathsf{T}} = \begin{bmatrix} \vdots \\ \mathbf{D}_i^\mathsf{T} \\ \vdots \end{bmatrix} \quad \text{where} \quad \mathbf{D}_i^\mathsf{T} = \begin{bmatrix} \dfrac{\partial N_i}{\partial x} & 0 & 0 & \dfrac{\partial N_i}{\partial y} & 0 & \dfrac{\partial N_i}{\partial z} \\ 0 & \dfrac{\partial N_i}{\partial y} & 0 & \dfrac{\partial N_i}{\partial x} & \dfrac{\partial N_i}{\partial z} & 0 \\ 0 & 0 & \dfrac{\partial N_i}{\partial z} & 0 & \dfrac{\partial N_i}{\partial y} & \dfrac{\partial N_i}{\partial x} \end{bmatrix} \tag{3.21}$$

where $i \in \eta_{eu}$ and η_{eu} are all the velocity nodes belonging to element e and $\mathbf{C}_{\mu^e}$ is the constitutive matrix from equation (3.12). The elemental matrix $\mathbf{G}^{e^{\mathsf{T}}}$ is computed as,

$$\mathbf{G}^{e^{\mathsf{T}}} = -\int_{\Omega^e} \tilde{\mathbf{N}}^{\mathsf{T}} (\nabla \cdot \mathbf{N}) \, d\Omega \tag{3.22}$$

with

$$\tilde{\mathbf{N}}^{\mathsf{T}} = \begin{bmatrix} \vdots \\ \tilde{N}_j \\ \vdots \end{bmatrix} \qquad \text{and} \qquad \mathbf{N} = \begin{bmatrix} \cdots \mathbf{N}_i \cdots \end{bmatrix} \quad \text{where} \quad \mathbf{N}_i = \begin{bmatrix} N_i & 0 & 0 \\ 0 & N_i & 0 \\ 0 & 0 & N_i \end{bmatrix} \tag{3.23}$$

where $j \in \eta_{ep}$ and η_{ep} are all the pressure nodes belonging to element e. Finally, using the same notation, the elemental force vector $\mathbf{f}^e$ is computed as,

$$\mathbf{f}^e = \int_{\Omega^e} \mathbf{N}^{\mathsf{T}} \rho \mathbf{g} \, d\Omega \tag{3.24}$$

where $\mathbf{g}$ is a column gravity vector. It is important to remark that the order in which the elemental nodes η_{eu} and η_{ep} appear in the previous expressions will result in different elemental matrices. However, the global matrices will be the same provided the assembly operator places them in the correct position.

The so-called mixed FE problem (3.17), with velocity and pressure as separate unknowns, is computationally expensive in large 3-dimensional domains, $n_{\text{FE}} = 3n_u + n_p$ degrees of freedom. Several numerical schemes and preconditioners have been proposed to solve it efficiently. However, the efficiency of such solvers is still insufficient for large-scale probabilistic inversions where the number of problems to be solved is in the order of 100K (or higher). The Reduced Basis technique described next attempts to overcome this problem and is independent of the numerical technique chosen to solve the forward problem.

3.2 Reduced Basis method

Reduced order techniques, aka Model Order Reduction (MOR), are a family of approximation methodologies based on the common idea of expressing the solution of a problem in terms of a basis of reduced size (Chinesta and Ladevèze, 2014; Quarteroni et al., 2016; Hesthaven et al., 2016; Ito and Ravindran, 1998; Ravindran, 2000). “Reduced” in this context refers to a basis (i.e. a family of basis functions) with a cardinal smaller than that resulting from a traditional FE discretization. These techniques are very useful when trying to obtain approximated solutions to a large number of problems with similar charateristics. For instance in probabilistic inversions where a large number of forward problems with small variations in the properties needs to be solved. The so-called Reduced Basis (RB) method (cf. Florentin and Díez, 2012; Quarteroni et al., 2016; Rozza et al., 2007, 2013), is particularly attractive for our purposes as it is easy to implement and well-suited to be coupled with bayesian methods as will be shown in Chapter 5. In the context of probabilistic inverse problems a large number of evaluations of the forward problem (typically solutions to PDEs) are required to approximate the full posterior PDF (Tarantola, 2005; Gregory, 2005; Gilks et al., 1996; Afonso et al., 2013a,b). While at the beginning of a Markov Chain Monte Carlo (MCMC) inversion the parameter space may be large, a well-formulated MCMC algorithm will converge quickly to a reduced region of the parameter space where all the forward problems will have similar input parameters and predictions. At this stage, the problems previously solved can be reused to improve the efficiency of the solver. It is here where a well-trained surrogate model performs the best and makes the MCMC simulation extremely efficient.

The *basis* of a FE procedure is the set of all $N_i(\boldsymbol{x})$ functions used to construct the solution $\boldsymbol{u}^h(\boldsymbol{x})$ as shown in (3.14) (and equivalently for pressure (3.16)). Note that each function $N_i(\boldsymbol{x})$ is associated with a node of the FE mesh and therefore the number of functions is usually very large and grows when the mesh is refined. One key aspect of the functions $N_i(\boldsymbol{x})$ is their local character. This means that the generated matrices are large and sparse; a useful feature exploited by FE solvers.

The main idea of MOR techniques, and in particular of the RB approach, is to approximate the solution $\mathbf{u}$ (and that for $\mathbf{p}$ in mixed formulations) in terms of a smaller set of functions $\mathbf{b}_m$ as

$$\mathbf{u} \approx \mathbf{u}_{\text{RB}} = \sum_{m=1}^{n_{\text{RB}}} \alpha_m \mathbf{b}_m \tag{3.25}$$

where n_{RB} is the number of basis functions and α_m the corresponding unknowns. The new basis $\{\mathbf{b}_1, \dots \mathbf{b}_{n_{\text{RB}}}\}$ is no longer associated with mesh nodes and thus no longer local. Therefore, for any new parameter (e.g. viscosity and density) input into (3.1), we seek a solution to (3.17) in a lower-dimension space $n_{\text{RB}} \ll n_{\text{FE}}$. While simple in principle, this idea rises three important questions:

1. Given a set of basis vectors $\mathbf{b}_m$, how is the solution $\mathbf{u}_{\text{RB}}$ computed?
2. How accurate is the reduced solution $\mathbf{u}_{\text{RB}}$ compared to the full-order FE solution $\mathbf{u}$?
3. How are the basis vectors $\mathbf{b}_m$ actually obtained?

We answer these three questions in the next three sections. We will first describe the simpler problem of obtaining reduced solutions $\mathbf{u}_{\text{RB}}$ and assessing their error assuming we already have an available set of basis vectors $\mathbf{b}_m$.

3.2.1 Reduced Basis approximation

Let us start with answering the first of the three questions presented above: *how is the solution* $\mathbf{u}_{RB}$ *computed?* Consider, for instance, the linear system of equations of the Stokes problem described in (3.17) which has n_{FE} unknowns (number of degrees of freedom of the mesh). The key idea behind the RB strategy is to create a basis of n_{RB} independent FE solutions $(\mathbf{u}_m, \mathbf{p}_m)$ of different realizations of the same problem (e.g. by varying the material properties). This group of solutions, the so-called *snapshots*, will be used as a basis. For the velocity, $\mathbf{b}_m = \mathbf{u}_m$, represents a subset of $\mathbb{R}^{3n_u}$,

$$\operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_{\text{RB}}}\} \subset \mathbb{R}^{3n_u}. \tag{3.26}$$

Similarly for pressure, the group of solutions $\mathbf{p}_m$ represents a subset of $\mathbb{R}^{n_p}$,

$$\operatorname{span}\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{n_{\text{RB}}}\} \subset \mathbb{R}^{n_p}. \tag{3.27}$$

Despite the smaller number of elements in the basis, it is assumed that $\mathbb{R}^{n_{\text{RB}}}$ represents the space $\mathbb{R}^{n_{\text{FE}}}$ accurately enough. Every time a new problem has to be solved, its solution $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{n_{\text{RB}}}]^{\mathsf{T}}$, is sought in $\mathbb{R}^{n_{\text{RB}}}$ instead of in $\mathbb{R}^{n_{\text{FE}}}$ thus reducing dramatically the computational cost since $n_{\text{RB}} \ll n_{\text{FE}}$. The reduced solution is found as a linear combination of the elements of the basis

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} \approx \begin{bmatrix} \mathbf{u}_{\text{RB}} \\ \mathbf{p}_{\text{RB}} \end{bmatrix} = \sum_{m=1}^{n_{\text{RB}}} \alpha_m \begin{bmatrix} \mathbf{u}_m \\ \mathbf{p}_m \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{\mathbf{u}} \\ \mathbf{B}_{\mathbf{p}} \end{bmatrix} \boldsymbol{\alpha} \tag{3.28}$$

where $\boldsymbol{\alpha}$ is the vector of unknowns and $\mathbf{B_u} = [\mathbf{u}_1, \ldots, \mathbf{u}_{n_{\text{RB}}}]$ and $\mathbf{B_p} = [\mathbf{p}_1, \ldots, \mathbf{p}_{n_{\text{RB}}}]$ are matrices representing the reduced basis for velocity and pressure, respectively. Each column of $\mathbf{B_u}$ and $\mathbf{B_p}$ correspond to a FE solution. Note that in our implementation, two separated basis have been defined for velocity and pressure, but both share the same number of functions n_{RB} and are interpolated using the same $\boldsymbol{\alpha}$.

The solution to the Stokes problem is obtained by replacing $\mathbf{u}$ and $\mathbf{p}$ with their reduced approximations from (3.28) into (3.17) and multiplying both sides by the basis to obtain,

$$\begin{bmatrix} \mathbf{B_u} \\ \mathbf{B_p} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{K} & \mathbf{G} \\ \mathbf{G}^{\mathsf{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{B_u} \\ \mathbf{B_p} \end{bmatrix} \boldsymbol{\alpha} = \begin{bmatrix} \mathbf{B_u} \\ \mathbf{B_p} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}. \tag{3.29}$$

Performing the multiplications, the system above is written as

$$\left[\mathbf{B}_\mathbf{u}^{\mathsf{T}}\mathbf{K}\mathbf{B_u} + \mathbf{B}_\mathbf{u}^{\mathsf{T}}\mathbf{G}\mathbf{B_p} + \mathbf{B}_\mathbf{p}^{\mathsf{T}}\mathbf{G}^{\mathsf{T}}\mathbf{B_u} \right] \boldsymbol{\alpha} = \mathbf{B}_\mathbf{u}^{\mathsf{T}}\mathbf{f}. \tag{3.30}$$

Since all elements in $\mathbf{B_u}$ are FE solutions, they all satisfy the incompressibility condition, namely $\mathbf{G}^{\mathsf{T}}\mathbf{u}_m = 0$, for $m = 1, ..., n_{\text{RB}}$. Therefore,

$$\mathbf{G}^{\mathsf{T}}\mathbf{B_u} = 0 \tag{3.31}$$

is always satisfied (independently of $\boldsymbol{\alpha}$) and the second and third terms in the left-hand side of equation (3.30) vanish. In other words, any reduced solution $\mathbf{u}_{\text{RB}}$ obtained as a linear combination of the elements of the basis also satisfies the incompressibility condition. Therefore, the final reduced system of equations reads,

$$\mathbb{K}_{\text{RB}}\boldsymbol{\alpha} = \mathbb{f}_{\text{RB}} \tag{3.32}$$

where the reduced stiffness matrix and reduced force vector are $\mathbb{K}_{\text{RB}} = \mathbf{B}_\mathbf{u}^{\mathsf{T}}\mathbf{K}\mathbf{B_u}$ and $\mathbb{f}_{\text{RB}} = \mathbf{B}_\mathbf{u}^{\mathsf{T}}\mathbf{f}$, respectively. Note that $\mathbb{K}_{\text{RB}}$ has the size of the reduced basis, $n_{\text{RB}} \times n_{\text{RB}}$, which is significantly smaller than the traditional FE nodal basis. For example, in the numerical examples presented below, n_{RB} is of the order of 10^2 while the full FE basis is $> 10^4$, even for the smallest example. It is worthy to remark that the final reduced system of equations does not depend on the pressure. It allows us to only use the velocity part of the basis, $\mathbf{B_u}$, and thus reduce even more computational resources and time. For simplicity, from now on we will drop the subscript $\mathbf{u}$ from the basis.

Remark. *Note that when using iterative methods to solve the system* (3.17)*, equation* (3.31) *is no longer satisfied exactly and, consequently, the terms including* **G** *in equation* (3.30) *do not vanish and should be explicitly considered. However, the additional computational cost is insignificant, since not only the matrix* **G** *is much smaller than* **K***, but also the terms including* **G** *only change when a new solution is added to the basis. In this thesis, however, we will use direct solvers and* (3.31) *holds exactly.*

3.2.2 Error of the RB with respect to the FE solution

In this section we will answer the second question from Section 3.2: *how accurate is the reduced solution* $\mathbf{u}_{RB}$ *compared to the full-order FE solution* $\mathbf{u}$*?* Two different error indicators will be presented, one global and one goal-oriented. First of all, a measure of the global error will be introduced based in the energy norm. Secondly, a goal-oriented error estimator which focuses on a specific feature of the problem, i.e. a Quantity of Interest, will be obtained by means of introducing an adjoint problem.

Error in the energy-norm

The discrete space described by the reduced basis $\mathbb{R}^{n_{\mathrm{RB}}}$ is an approximation of the FE discrete space $\mathbb{R}^{n_{\mathrm{FE}}}$, and as such, the RB solution $\mathbf{u}_{\mathrm{RB}}$ has an associated error with respect to the high-fidelity FE solution $\mathbf{u}$. We must keep in mind that $\mathbf{u}$ is neither an exact solution of the problem, but also an approximation coming from the FE discretization. The estimation of the errors associated with FE solutions is a well-known procedure and it has been described in detail elsewhere (Zienkiewicz et al., 2005; Hughes, 1987). Here, instead of measuring the error with respect to the analytic solution, we will focus on assessing the error of the RB solution with respect to the high-fidelity FE one. Specifically, we would like to measure how well the reduced basis performs, assuming that the best possible solution is that given by the FE solver. Within this context, we define the errors introduced by the RB solution with respect to the FE solution as:

$$\mathbf{e}_{\mathbf{u}} := \mathbf{u} - \mathbf{u}_{\mathrm{RB}} \qquad \text{and} \qquad \mathbf{e}_{\mathbf{p}} := \mathbf{p} - \mathbf{p}_{\mathrm{RB}}. \tag{3.33}$$

The error equation associated with the RB solution is

$$\begin{bmatrix} \mathbf{K} & \mathbf{G} \\ \mathbf{G}^{\mathsf{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}_{\mathbf{u}} \\ \mathbf{e}_{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{K} & \mathbf{G} \\ \mathbf{G}^{\mathsf{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{\mathrm{RB}} \\ \mathbf{p}_{\mathrm{RB}} \end{bmatrix}. \tag{3.34}$$

Therefore, the residual associated with the RB solution reads,

$$\mathbf{r} = \mathbf{f} - \mathbf{K}\mathbf{u}_{\mathrm{RB}} - \mathbf{G}\mathbf{p}_{\mathrm{RB}}, \tag{3.35}$$

since due to the incompressibility condition in (3.31), the part of the residual corresponding to the pressure vanishes identically, $-\mathbf{G}^{\mathsf{T}}\mathbf{u}_{\mathrm{RB}} = 0$. Note, however, that both the velocity and the pressure RB solutions contribute to the velocity part of the residual. The residual is usually a good indicator for the energy norm of the error (Quarteroni et al., 2016; Hesthaven et al., 2016). If the solution $\mathbf{u}_{\mathrm{RB}}$ and $\mathbf{p}_{\mathrm{RB}}$ were exact (i.e. equal to the FE solution), the residual would be zero, and so would the error. It follows that, in general, a large residual implies a large error in the RB approximation. This error has a global character since it considers the whole domain of the problem and it is also directly related to the output. A scalar relative version of the error indicator in the energy norm e_r is obtained with,

$$e_r = \frac{\|\mathbf{r}\|}{\|\mathbf{f}\|}. \tag{3.36}$$

Goal-oriented error estimator

Instead of using the full residual in equation (3.35) as a measure of the error introduced by the RB approximation, here we introduce a more suitable *goal-oriented criterion*. The key idea is that such criterion is based on a Quantity of Interest (QoI) which is a specific feature or region of the problem in which we are interested. In this way, the error indicator is not focused on getting a good global solution, but in assigning more weight or importance to a specific QoI of the problem and, therefore, control the maximum admissible error on the chosen QoI (Florentin and Díez, 2012). This way of measuring the error of the RB approximation helps to maintain the overall size of the RB small, as the basis is specifically trained to accurately represent (within admissible errors) only the relevant part of the solution, that is, the QoI. In terms of probabilistic inversions, the QoI is typically a model prediction that can be compared against data or that controls an important output of the model. Therefore, we do not seek to obtain accurate solutions for the entire numerical domain or for all model predictions, but rather to guarantee accurate RB solutions to the quantity or quantities that are of interest for the particular purpose of the inversion.

Two remarks follow. First, the RB approach discussed in this thesis is independent of how the QoI is defined and it can accept any other error measure. Second, the fact that the accuracy of a global solution is not assessed does not imply that the approximation of the solution will be of poor quality. In the limiting case where the QoI is of "global" character (e.g. if the residual is used to estimate the error), the error will be a measure of accuracy for the entire domain and the resulting RB approximation will be trained to reproduce the solution in any region of the domain with the same accuracy.

We start with the derivation of a goal-oriented estimator for a linear QoI and, later, an extension to a non-linear QoI will be described. The QoI is a scalar quantity Q, computed as a linear function, l^o, that extracts the relevant information from the global solution $\boldsymbol{u}$,

$$Q = l^o(\boldsymbol{u}). \tag{3.37}$$

The discrete version of the linear QoI can be expressed as a product of one vector $\mathbf{f}^o$ describing the function and the FE solution $\mathbf{u}$ as,

$$Q_{\text{FE}} = \mathbf{f}^{o\mathsf{T}}\mathbf{u}. \tag{3.38}$$

In order to obtain an error representation based on the QoI, a dual (or adjoint) problem is typically defined (e.g. Florentin and Díez, 2012). The weak form of the dual problm for the Stokes equation reads: *find* $(\boldsymbol{v}, d) \in \boldsymbol{\mathcal{S}} \times \mathcal{Q}$ *such that,*

$$a\big(\boldsymbol{w}, \boldsymbol{v}\big) + b\big(\boldsymbol{w}, d\big) = l^o\big(\boldsymbol{w}\big) \qquad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{V}}, \tag{3.39a}$$

$$b\big(\boldsymbol{v}, q\big) = 0 \qquad \forall q \in \mathcal{Q}. \tag{3.39b}$$

This dual problem is discretized in the same FE spaces as the direct problem (3.17), which gives

$$\begin{bmatrix} \mathbf{K} & \mathbf{G} \\ \mathbf{G}^{\mathsf{T}} & \mathbf{0} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{v} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{f}^o \\ \mathbf{0} \end{bmatrix}. \tag{3.40}$$

For the Stokes problem discussed here, the stiffness matrix is symmetric and therefore the dual problem differs from the direct one only in the source term, which corresponds to the discretized linear operator representing the QoI.

Finally, using equations (3.17), (3.38) and (3.40), a representation of the error in the QoI associated with the RB approximation can be defined as,

$$\begin{aligned} E_Q = Q_{\text{FE}} - Q_{\text{RB}} &= \mathbf{f}^{o\mathsf{T}}(\mathbf{u} - \mathbf{u}_{\text{RB}}) \\ &= [\mathbf{K}\mathbf{v} + \mathbf{G}\mathbf{d}]^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_{\text{RB}}) \\ &= \mathbf{v}^{\mathsf{T}}\mathbf{K}(\mathbf{u} - \mathbf{u}_{\text{RB}}) + \mathbf{d}^{\mathsf{T}}\mathbf{G}^{\mathsf{T}}(\mathbf{u} - \mathbf{u}_{\text{RB}}) \\ &= \mathbf{v}^{\mathsf{T}}(\mathbf{f} - \mathbf{K}\mathbf{u}_{\text{RB}}) \end{aligned} \tag{3.41}$$

where, in the last step, the incompressibility condition in (3.31) is used to cancel the term involving $\mathbf{G}^{\mathsf{T}}$ and the equality $\mathbf{v}^{\mathsf{T}}\mathbf{K}\mathbf{u} = \mathbf{v}^{\mathsf{T}}\mathbf{f}$ holds because $\mathbf{v}^{\mathsf{T}}\mathbf{G}\mathbf{p} = 0$ (incompressibility of the dual problem). A residual-like vector is defined as,

$$\mathbf{r}^* := \mathbf{f} - \mathbf{K}\mathbf{u}_{\text{RB}}, \tag{3.42}$$

and, although different from the residual $\mathbf{r}$ in (3.35), $\mathbf{v}^{\mathsf{T}}\mathbf{r}^* = \mathbf{v}^{\mathsf{T}}\mathbf{r}$ holds because $\mathbf{v}^{\mathsf{T}}\mathbf{G}\mathbf{p}_{\text{RB}} = 0$. Similar to the reduced system of equations (3.32), the representation of the error using a goal-oriented criterion does not require the reduced approximation of the pressure term. Thus, upholding the reduction in the resources and computational cost.

According to equation (3.41), the two ingredients needed to obtain a representation of the error are: i) the solution to the dual problem $\mathbf{v}$, and ii) the residual-like vector $\mathbf{r}^*$. A critical point here is that if the solution of the dual problem requires as much computational effort as the solution of the primal problem, then the method becomes impractical. We can therefore replace the exact solution of the adjoint problem with an approximate (and much more efficient) solution, $\hat{\mathbf{v}}$, and define an error estimator for Q as

$$E_Q \approx \hat{E}_Q = \hat{\mathbf{v}}^{\mathsf{T}}\mathbf{r}^* \tag{3.43}$$

or, its relative version as,

$$\hat{e}_Q = \left| \frac{\hat{E}_Q}{Q_{\text{RB}}} \right|. \tag{3.44}$$

Adjoint solution approximation

There is no unique approach to obtain an approximation of the solution of the adjoint problem and avoid its explicit and expensive computation. The strategy that best suits a problem needs to be examined in each specific case. The best (but no feasible) option would be to compute a full FE solution at each inversion step so that the error in the QoI is explicitly computed and not estimated. We have tested different approaches which will be presented ranging from more accurate but computationally expensive to less accurate but more efficient. However, since the approaches are strongly coupled with the inverse problem and the enrichment strategy for the basis, only a brief idea will be given here. The reader is referred to the section 5.2.2 where a detailed description together with illustrative examples are provided.

The first and more computationally expensive option is to build a Reduced Basis for the adjoint problem. It is built using the same procedure we use for the primal problem (Section 5.1), except for the error indicator of the adjoint solution which is computed using the norm of the residual. With this strategy, every time the error needs to be assessed the adjoint is approximated using its reduced basis. The second approach is based on one important property of the adjoint problem: adjoints are typically less sensitive to perturbations of the parameters than the primal problem (Florentin and Díez, 2012; García-Blanco et al., 2017; Serafin et al., 2017) and therefore the same adjoint solution can be used to estimate the error of many similar primal solutions. With the idea of having a relevant solution for the problem at hand, the second approach can be divided in three sub-approaches depending on the frequency of updating the solution of the adjoint problem: a) the adjoint solution is recomputed after a fixed number of inversion steps, b) the adjoint solution is recomputed every time the primal reduced basis has been enriched a certain number of times and, finally, c) the most efficient, but less accurate option and yet providing good estimators, uses a single solution of the adjoint problem to estimate the errors of all subsequent primal problems.

Extension to a non-linear QoI

Depending on the problem at hand, and as we will see later in Section 5.2.1, we may be interested in representing the error on a non-linear QoI. Let us define a simple non-linear QoI as,

$$Q = l^o(\boldsymbol{u}) = \|\boldsymbol{u}\|^2. \tag{3.45}$$

The discrete version of the non-linear QoI is,

$$Q_{\text{FE}} = \mathbf{u}^{\mathsf{T}}\mathbf{u}. \tag{3.46}$$

Recalling the error of the reduced basis approximation with respect to the finite elements solution defined in (3.33), $\mathbf{u} = \mathbf{u}_{\text{RB}} + \mathbf{e}_{\mathbf{u}}$, and dropping the subindex from $\mathbf{e}_{\mathbf{u}}$ for simplicity, we can express the non-linear QoI as,

$$l^o(\boldsymbol{u}_{\text{RB}} + \boldsymbol{e}) = (\mathbf{u}_{\text{RB}} + \mathbf{e})^{\mathsf{T}}(\mathbf{u}_{\text{RB}} + \mathbf{e}) = {\mathbf{u}_{\text{RB}}}^{\mathsf{T}}\mathbf{u}_{\text{RB}} + 2{\mathbf{u}_{\text{RB}}}^{\mathsf{T}}\mathbf{e} + \cancelto{0}{\mathbf{e}^{\mathsf{T}}\mathbf{e}} \tag{3.47}$$

where we have assumed that the norm of the error is much smaller than the rest of the terms so that it can be neglected. Therefore, the linearised error representation is,

$$l^o_{\text{L}}(\boldsymbol{e}) \simeq l^o(\boldsymbol{u}_{\text{RB}} + \boldsymbol{e}) - l^o(\boldsymbol{u}_{\text{RB}}) = 2{\mathbf{u}_{\text{RB}}}^{\mathsf{T}}\mathbf{e}. \tag{3.48}$$

The discrete version of $l^o_{\text{L}}(\cdot)$ is the operator $\mathbf{f}^o$ and it is defined as,

$$\mathbf{f}^o = 2\mathbf{u}_{\text{RB}}. \tag{3.49}$$

It is sometimes called "extractor" since it extracts the important information from the vector to whom it is applied. Once $\mathbf{f}^o$ is defined, there is no other difference with respect to the linear case and the error is estimated in the same way. An adjoint problem is defined as in equation (3.40) which leads to the expression of the error representation in terms of the solution of the adjoint problem and the residual of the RB solution, equation (3.41).

3.2.3 Classic approaches to build a Reduced Basis

Finally, we answer the last question: *how are the basis vectors* $\mathbf{b}_m$ *actually obtained?* Typical MOR implementations consist of two stages: 1) the *offline* stage, where all costly computations (involving a high-fidelity solver) that depend on $\boldsymbol{\mathcal{V}}^h$ are performed; and 2) the *online* stage, where the reduced order solver or "surrogate" is used to obtain fast and computationally inexpensive solutions independent of $\boldsymbol{\mathcal{V}}^h$ (Quarteroni et al., 2016; Hesthaven et al., 2016; Ito and Ravindran, 1998; Prud'homme et al., 2002; Veroy and Patera, 2005; Rozza and Veroy, 2007; Rozza et al., 2013). The process of building a basis is typically part of the offline stage, since each element of the basis is a high-fidelity solution that needs to be computed. The basis functions are built based on snapshots corresponding to different realizations of the problem when varying the parameters. For instance, for the Stokes problem here presented, viscosities and densities would be the parameters that change among snapshots. The idea of the Reduced Basis is that the basis must be able to represent, up to a certain accuracy, the solution of any problem within the parameter space. Since the parameter space can not be entirely sampled and solved, some points are chosen to be a good representation of it. Consequently, the quality of the basis will strongly depend on the choice of parameters whose solution is to be included as a basis function. Sampling the parametric space adequately is therefore of the highest importance. If the parameter space is poorly sampled, the reduced basis might provide poor approximations. Moreover, some orthogonality procedure must be applied in order to guarantee that the elements are not linearly dependent. A description of the most classic approaches to build an offline reduced basis will be presented next.

The most simple and intuitive strategy would be to sample the parameter space with uniformly distributed points (e.g. Quarteroni et al., 2016). In doing so, all regions of the parameter space will be explored (up to the chosen discretization). However, not only it may entail solving large numbers of high-fidelity problems, but also not all parameters affect the solution in the same way. In other words, the solution may not be equally sensitive to the whole parameter space and, with this strategy, all regions are sampled alike. It would be better to have the basis functions or snapshots distributed according to the importance of the parameter space. For large dimensional parameter spaces, regular lattices become impractical and other sampling strategies should be applied: Monte Carlo random sampling, latin hypercube sampling, sparse grids among others (e.g. Cochran, 1977; Gerstner and Griebel, 1998; Quarteroni et al., 2016; Bui-Thanh et al., 2008). Once a finite dimensional set of parameters is selected, the corresponding high-fidelity solutions are computed to obtain a basis of snapshots. However, as previously mentioned, correlations may exist among snapshots since orthogonality is not enforced and some orthogonalization strategy must be applied a posteriori. The Proper Orthogonal Decomposition (POD) method allows reducing the dimension of a data set by means of expressing it in an orthonormal basis which is optimal in a least-square sense. A correlation matrix is computed from the high-fidelity snapshots and its eigenvalues and eigenvectors (POD modes) are obtained (Brunton and Kutz, 2019; Quarteroni et al., 2016; Hesthaven et al., 2016). The reduction is performed truncating the new basis by only keeping the largest eigenvalues (up to a chosen tolerance) which are the ones having most of the information of the original snapshots. The elements of the new basis are the

POD modes whose eigenvalues survived. Although this new basis is orthonormal, the computational offline cost is still very high, not only due to the compression part, but also due to the large number of high-fidelity problems that need to be solved to finally only keep a few of them.

The next strategy uses a Greedy approach to build the Reduced Basis (Grepl and Patera, 2005; Quarteroni et al., 2016; Hesthaven et al., 2016; Larion et al., 2020). It consists in an iterative sampling of the parameter space in which at each step a posterior error estimate is computed to decide using an optimality criterion which is the next sampling point. Initially a training sample of points in the parameter space is chosen which will be used to select the RB space. The high-fidelity solution of a starting point is computed and added as a first element in the basis. Then, the reduced solution of all the other points in the training sample is computed with the reduced basis and their estimated error evaluated a posteriori. The point whose solution exhibits the largest error is the next one to be solved with the high fidelity solver and included in the basis. This process is repeated until the reduced basis reaches a maximum number of elements or a stopping tolerance is achieved. In doing so, the Greedy strategy is able to lead the sampling process in the direction which is worst approximated. In order to keep the basis orthonormal, when a new element is added to the basis it might be orthonormalized through a Graham-Schmidt procedure.

In both strategies described above, the selection of the initial set of points in the parameter space or training sample is crucial for a good and reliable basis within the parameter space. In the context of a probabilistic inverse problem, it is not always straightforward to predict which snapshots should be computed *a priori* during the offline stage, as there is no way of knowing in advance which solutions are most relevant as the MCMC converges towards a reduced parameter space (i.e. the high probability regions). Determining the reduced parameter space is exactly the goal of an inverse problem! Moreover, the larger is the number of parameters to recover by the inversion, the more difficult is to predict their relevant high-probability regions. For this reason, a different strategy will be devised to combine the offline and online stages into a single data-driven stage where the reduced basis is created *on the fly* and as needed during the inversion process. It results in a basis tailored to the specific problem that is being solved. This strategy will be explained in detail in section 5.1, since it is strongly coupled with the inverse scheme.

Chapter 4

The Inverse Problem and Markov Chain Monte Carlo

Physical theories represent the link between parameters and physical quantities to completely describe a physical system. A *forward* or *direct problem* is one that allows to make predictions of some quantities given the values of the parameters characterizing the system. They are usually modelled with a set of partial differential equations which, given the parameters, provide a unique solution. An *inverse problem*, instead, consists in inferring the values of the parameters for a given set of experimental data or measurements. The main goal of an inverse problem is to find the physical model, i.e. the values of a set of parameters, that best fits the data. Inverse problems typically arise when the parameters or quantities we are interested in cannot be directly measured or the system is very complex. In this situation, inferring them from other variables that can be measured is the only available option. That is precisely the case of geophysical inverse problems, where most of the available data is on the surface and the involved physical systems exhibit large complexity. Unlike forward problems, inverse problems are typically under-determined and, therefore, more than one configuration of parameters is able to explain the same data.

This chapter is focused on introducing the idea of inverse problems and some strategies to solve them. The first part presents the general Bayesian framework to deal with probabilistic inverse problems. The second part is devoted to the Markov

Chain Monte Carlo methods and in particular, the Metropolis-Hastings algorithm. Finally, in the last part, one inversion example combining the full order model and the inversion strategy is shown.

4.1 Inverse problems

A physical system is described by physical laws that express the relationship between physical variables. In the context of inverse problems it is interesting to separate such variables in two types: model parameters and observable parameters.

- **Model parameters**. They are a finite set of parameters m_i that approximate the properties of interest of a physical system. This approximation is called parametrization and, once a system is parametrized, giving a value to each model parameter completely characterizes the system. The set of n_m model parameters is represented together in a vector denoted by $\mathbf{m}$,

  $$\mathbf{m} = [m_1, \dots, m_{n_m}] \tag{4.1}$$

  In the context of an inversion problem, model parameters are the ones that can not be directly measured. In this thesis, we will use "model" to refer to a particular set of values for an already chosen parametrization.

- **Observable parameters**. They correspond to parameters that can directly be measured, even though it may require complex experiments. The set of n_d observable parameters is represented by the vector $\mathbf{d}$,

  $$\mathbf{d} = [d_1, \dots, d_{n_d}] \tag{4.2}$$

  An observable could be for instance the elevation, and the observable parameters d_i the different points where it is measured. We will usually refer to $\mathbf{d}$ as "observations" or "data". When a measurement is performed, the values of the observable parameters are collected in the data vector $\mathbf{d}^{\text{obs}} = [d_1^{\text{obs}}, \dots, d_{n_d}^{\text{obs}}]$. Not to be confused with different realizations of the same experiment.

A physical theory is developed to connect the model parameters $\mathbf{m}$ with the observable ones $\mathbf{d}$. The relation maps the model space with the data space typically through a nonlinear and complex function that, when explicit, can be represented as,

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) \tag{4.3}$$

This defines a forward problem which allows us to predict the values of $\mathbf{d}$ from a given model $\mathbf{m}$. Therefore, observable parameters can both be *measured* by means of experiments and *predicted* by solving a forward problem. In an ideal case there would be a $\mathbf{m}_{\text{true}}$ satisfying $\mathbf{d}^{\text{obs}} = \mathbf{g}(\mathbf{m}_{\text{true}})$. However, there are two sources of error which prevent predicted values from being identical to the observed ones: uncertainties in the measurements and imperfections in the modelization (Tarantola, 2005). We will address them during this section.

An inverse problem is stated as: given a set of experimental data, $\mathbf{d}^{\text{obs}}$, find the model $\mathbf{m}$ that best represents the data. A *classical* or *deterministic* approach to solve an inverse problem leads to an optimization problem where one is looking for a unique solution, i.e. a set of values for the model parameters that minimizes the misfit between predicted and measured data. A *probabilistic* approach for inverse problems assumes that both experimental data and model parameters have an associated uncertainty and, thus, their true values can not be known for certain. Consequently and with this idea in mind, we are not just interested in the model that best fits the data but in having a complete statistical description of the state of knowledge about the model parameters. In doing so, the solution of a probabilistic inverse problem is not a unique model but a *probability density function (PDF)* over the space of model parameters. Such idea of state of knowledge allows, among others, to deal with the non-uniqueness of the solution of inverse problems. Probabilistic strategies also outplay deterministic ones in accounting for propagation of uncertainties, both from measurements and modelization, as well as in dealing with cases where data is scarce or combines different datasets.

The Bayesian approach reformulates the inverse problem in terms of statistical inference by using prior information on the parameters and observational data together with a forward model that links them (Kaipio and Somersalo, 2005; Tarantola, 2005; Mosegaard and Tarantola, 2002; Gregory, 2005). Thus, the solution of an inverse problem can be thought as the gain in information when prior knowledge and data are combined. Such approach requires that all the variables included in the model are defined as random variables and, consequently, they all have an associated PDF representing its current state of information or the knowledge we have about them.

4.1.1 A priori information and observations

The *a priori* information on model parameters is all the information on model parameters that is obtained independently of the measurements. It describes how we think a priori that the space of parameters is, i.e. the probability of model parameters $\mathbf{m}$ without considering the observations $\mathbf{d}^{\mathrm{obs}}$. The state of knowledge on the model parameters is represented by $\rho_{\mathrm{M}}(\mathbf{m})$, the prior PDF on the model space.

The *observations* are the result of experimental measurements. As in any experiment, there is a source of uncertainty due to the measurement errors. Therefore, the result of a measurement is not simply a set of values $\mathbf{d}^{\mathrm{obs}}$, but it is better understood as a state of information acquired on some observable parameter. The knowledge gained after some measurements is expressed with a probability density over the data space, $\rho_D(\mathbf{d})$.

Since the prior information on model parameters should be by definition independent of the observations, in most of the cases the *joint prior PDF* can be written as the product,

$$\rho(\mathbf{d}, \mathbf{m}) = \rho_{\mathrm{D}}(\mathbf{d})\rho_{\mathrm{M}}(\mathbf{m}) \tag{4.4}$$

However, there may be more general situations in which information on $\mathbf{m}$ and $\mathbf{d}$ is not independent. In such cases, (4.4) does not hold and we must deal with $\rho(\mathbf{d}, \mathbf{m})$. Assuming independence, we can also express the joint homogeneous probability density $\mu(\mathbf{d}, \mathbf{m})$ as the product of two independent marginal homogeneous densities

$$\mu(\mathbf{d}, \mathbf{m}) = \mu_{\mathrm{D}}(\mathbf{d})\mu_{\mathrm{M}}(\mathbf{m}) \tag{4.5}$$

where $\mu_{\mathrm{D}}(\mathbf{d})$ and $\mu_{\mathrm{M}}(\mathbf{m})$ are the homogeneous probability densities on the data and model space respectively. The homogeneous probability density is the one that assigns to each region of the space a probability proportional to the volume of the region (Mosegaard and Tarantola, 2002; Tarantola, 2005).

4.1.2 Correlations and uncertainties of the physical theory

The second source of error are the modelization imperfections, which may come either from a not accurate parametrization or from a poor knowledge of the physics itself. Since the resulting physical theories are not exact, a model $\mathbf{m}$ does not exactly predict a single $\mathbf{d}$ but shows a probability for $\mathbf{d}$. This is expressed as a conditional probability density $\theta(\mathbf{d}|\mathbf{m})$, which assigns uncertainties to the predictions made from the physical theories. Therefore, the exact relation in equation (4.3) is better understood in terms of probabilistic correlations. The joint probability density describing correlations that correspond to our physical theory together with the inherent uncertainties of the theory is written as the product,

$$\Theta(\mathbf{d},\mathbf{m}) = \theta(\mathbf{d}|\mathbf{m})\mu_{\mathrm{M}}(\mathbf{m}) \tag{4.6}$$

where $\theta(\mathbf{d}|\mathbf{m})$ is the aforementioned conditional probability density of obtaining $\mathbf{d}$ given the model $\mathbf{m}$.

4.1.3 Posterior PDF on model parameters

The most general solution of a probabilistic inverse problem is represented by a probability density function known as *posterior PDF* that contains all the information of the problem at hand. Following Tarantola (2005), the probability density representing the a posteriori information can be expressed as,

$$\sigma(\mathbf{d},\mathbf{m}) = k\frac{\rho(\mathbf{d},\mathbf{m})\Theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})} \tag{4.7}$$

where $\sigma(\mathbf{d},\mathbf{m})$ is the joint posterior PDF of the model and observable parameters and k is a normalizing constant. The actual solution of the forward problem is the posterior information in the model space, i.e the marginal PDF in the model space, which is the result of integrating over the data space,

$$\sigma_{\mathrm{M}}(\mathbf{m}) = k\int \sigma(\mathbf{d},\mathbf{m})d\mathbf{d} \tag{4.8}$$

$\sigma_{\mathrm{M}}(\mathbf{m})$ describes the probabilities of all possible $\mathbf{m}$ considering our knowledge on the physical theory and observations. Using equations (4.4), (4.5) and (4.6) the *posterior PDF of the model parameters* is expressed as,

$$\sigma_{\mathrm{M}}(\mathbf{m}) = k\rho_{\mathrm{M}}(\mathbf{m})L(\mathbf{m}) \tag{4.9}$$

where $L(\mathbf{m})$ is the likelihood function,

$$L(\mathbf{m}) = \int \frac{\rho_{\mathrm{D}}(\mathbf{d})\theta(\mathbf{d}|\mathbf{m})}{\mu_{\mathrm{D}}(\mathbf{d})}d\mathbf{d} \tag{4.10}$$

that measures how good a particular set of model parameters $\mathbf{m}$ is at reproducing the observations $\mathbf{d}^{\mathrm{obs}}$. The larger the value of the likelihood, the better it reproduces the observations. Figure 4.1 illustrates from a geometrical point of view the gain in knowledge from combining initial information $\rho(\mathbf{d}, \mathbf{m})$ and physical correlations $\Theta(\mathbf{d}, \mathbf{m})$ into a posterior distribution $\sigma(\mathbf{d}, \mathbf{m})$. Specifically, comparing $\rho_{\mathrm{M}}(\mathbf{m})$ with $\sigma_{\mathrm{M}}(\mathbf{m})$ the benefit of incorporating data $\rho_{\mathrm{D}}(\mathbf{d})$ and physical correlations $\Theta(\mathbf{d}, \mathbf{m})$ is evident.

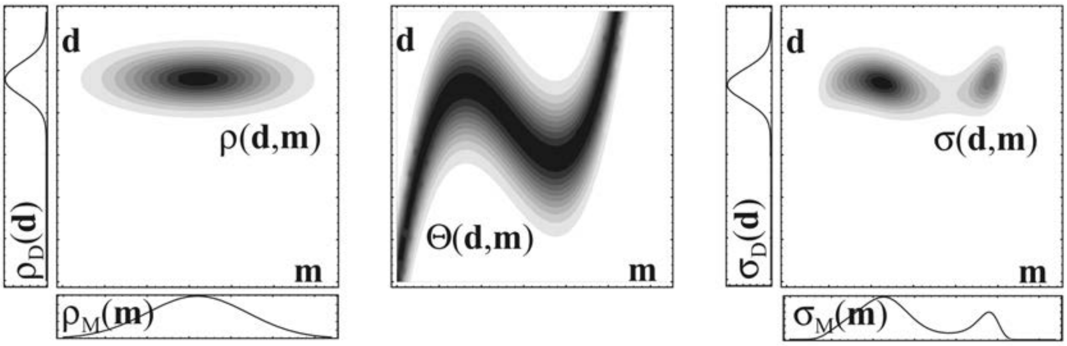

Figure 4.1: Left: $\rho(\mathbf{d}, \mathbf{m})$, joint probability density representing the initial state of information both on data and model parameters. Center: $\Theta(\mathbf{d}, \mathbf{m})$, shows the correlation between model parameters and data given by a physical theory that is not exact and, therefore, exhibits some uncertainties (thickness). Right: $\sigma(\mathbf{d}, \mathbf{m})$, shows the final state of information when combining $\rho(\mathbf{d}, \mathbf{m})$ and $\Theta(\mathbf{d}, \mathbf{m})$. The marginal posterior distributions for data $\sigma_{\mathrm{D}}(\mathbf{d})$ and model parameters $\sigma_{\mathrm{M}}(\mathbf{m})$ are also shown. Taken from Tarantola (2005).

The result of equation (4.9) is of great importance since it shows that the posterior PDF on model parameters $\sigma_{\mathrm{M}}(\mathbf{m})$ is proportional to the product of their prior PDF $\rho_{\mathrm{M}}(\mathbf{m})$ times the likelihood $L(\mathbf{m})$. It will be the base of the method explained in the next section to solve the inverse problem. As we will see later, in such context the constant k is irrelevant and it will not be computed.

Likelihood and Gaussian uncertainties

The likelihood $L(\mathbf{m})$ from equation (4.10) accounts for both sources of uncertainty; in the observations with the term $\rho_{\mathrm{D}}(\mathbf{d})$ as well as in the physical theory or modelization with $\theta(\mathbf{d}|\mathbf{m})$. When the uncertainties of the observations follow a Gaussian distribution, the probability density on the observable parameters is written as,

$$\rho_D(\mathbf{d}) = k \,\exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{d}^{\mathrm{obs}})^{\mathsf{T}} \mathrm{C}_{\mathrm{d}}^{-1} (\mathbf{d} - \mathbf{d}^{\mathrm{obs}})\right) \tag{4.11}$$

where $\mathrm{C_d}$ is a covariance matrix describing observation uncertainties. When the uncertainties in the physical theory also follow a Gaussian error model, the conditional probability is,

$$\theta(\mathbf{d}|\mathbf{m}) = k\ \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{g}(\mathbf{m}))^{\mathsf{T}}\mathrm{C_t^{-1}}(\mathbf{d}-\mathbf{g}(\mathbf{m}))\right) \tag{4.12}$$

with C_t the covariance matrix describing theoretical uncertainties. As shown in Tarantola (2005), the convolution of two Gaussian distributions also gives a Gaussian distribution, such that the likelihood from equation (4.10) can be rewritten as,

$$L(\mathbf{m}) = k\ \exp\left(-\frac{1}{2}(\mathbf{g}(\mathbf{m})-\mathbf{d}^{\mathrm{obs}})^{\mathsf{T}}\mathrm{C_D^{-1}}(\mathbf{g}(\mathbf{m})-\mathbf{d}^{\mathrm{obs}})\right) \tag{4.13}$$

where $C_\mathrm{D} = C_\mathrm{d}+C_\mathrm{t}$ is the covariance matrix accounting for both observation and theoretical uncertainties. According to this result, neglecting either of them is achieved by simply removing the corresponding covariance matrix. It is useful to understand the argument of the exponential as a misfit error function, $e(\mathbf{m})$, between predicted and observed data. In doing so,

$$L(\mathbf{m}) = k\ \exp\left(-e(\mathbf{m})\right) \tag{4.14}$$

with,

$$e(\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m})-\mathbf{d}^{\mathrm{obs}})^{\mathsf{T}}\mathrm{C_D^{-1}}(\mathbf{g}(\mathbf{m})-\mathbf{d}^{\mathrm{obs}})\ . \tag{4.15}$$

When the model $\mathbf{m}$ is not good at explaining the data $\mathbf{d}^{\mathrm{obs}}$, i.e. the difference between $\mathbf{g}(\mathbf{m})$ and $\mathbf{d}^{\mathrm{obs}}$ is large, the misfit error function exhibits a large value and, consequently, a smaller likelihood than those models in which the difference $\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\mathrm{obs}}$ is smaller. Note that in all equations k denotes a normalizing constant which is not necessarily the same in each expression.

In the case of further assuming uncorrelated uncertainties, the misfit expression can be rewritten as,

$$e(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{n_d}\left(\frac{g_i(\mathbf{m})-d_i^{\mathrm{obs}}}{\sigma_{D_i}}\right)^2 \tag{4.16}$$

where σ_{D_i} are the diagonal entries of C_D i.e., the standard deviations accounting for observations and modelization uncertainties, $\sigma_{D_i}^2 = \sigma_{d_i}^2 + \sigma_{t_i}^2$.

Priors on model parameters

As previously defined, the prior PDF encodes everything we know about the parameter space (i.e. probabilities of $\mathbf{m}$) prior to considering the data $\mathbf{d}^{\mathrm{obs}}$ and it represents an important part of any Bayesian inference problem. If the prior information is vague then the prior distribution is a very broad/flat distribution. Conversely, narrow distributions indicate that prior knowledge on the respective parameters is accurate. Gaussian, Laplacian, Poisson, uniform and Jeffreys' priors are the most common options in the literature (Mosegaard and Tarantola, 2002; Kaipio and Somersalo, 2005). Given n_m independent model parameters m_i the prior PDF accounting for all parameters simultaneously can be written as the product of the individual priors $\rho_i(m_i)$,

$$\rho_{\mathrm{M}}(\mathbf{m}) = \prod_{i=1}^{n_m} \rho_i(m_i). \tag{4.17}$$

Uniform distribution. When our knowledge on model parameters is very poor, we may typically only know the boundaries of our parameters. For instance, the density can only take positive values and we usually consider it can not be infinity. In such cases, a uniform distribution is used since it assigns equal probabilities to all values within the possible region. Given $m_i^{\min}$ and $m_i^{\max}$ as the two boundaries for each parameter m_i, it reads

$$\rho_i(m_i) = \begin{cases} \dfrac{1}{|m_i^{\max} - m_i^{\min}|} & \text{for } m_i^{\min} < m_i < m_i^{\max} \\ 0 & \text{otherwise.} \end{cases} \tag{4.18}$$

Gaussian distribution. Assuming that priors for all parameters follow a Gaussian distribution, $m_i \sim \mathrm{N}(\nu_i, \sigma_{m_i}^2)$, it follows that,

$$\rho_{\mathrm{M}}(\mathbf{m}) \propto \exp\left(-s(\mathbf{m})\right) \tag{4.19}$$

with

$$s(\mathbf{m}) = \frac{1}{2}\left(\mathbf{m} - \boldsymbol{\nu})^t \mathrm{C}_{\mathrm{M}}^{-1} (\mathbf{m} - \boldsymbol{\nu}\right) \tag{4.20}$$

where $\boldsymbol{\nu}$ is the vector of mean values for the parameters and C_{M} a covariance matrix describing prior uncertainties. Similarly to equation (4.15), $s(\mathbf{m})$ can be understood as a misfit function on the space of model parameters. In this sense, models $\mathbf{m}$ which are close to $\boldsymbol{\nu}$ exhibit a higher probability than those models in which the difference

$\mathbf{m} - \boldsymbol{\nu}$ is higher. If uncertainties are uncorrelated, the expression for the sum of squares simplifies to

$$s(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{n_m} \left(\frac{m_i - \nu_i}{\sigma_{m_i}} \right)^2 \tag{4.21}$$

where n_m is the number of parameters and ν_i and σ_i are the mean and standard deviation of the prior for parameter m_i, respectively.

4.2 Markov Chain Monte Carlo: Metropolis-Hastings algorithm

As we have seen in the previous section, the most general solution of an inverse problem is a probability distribution over the model parameters space. However, only when the probability density is simple analytical techniques can be used. For instance in problems where the relation between model and data parameters is linear or with Gaussian statistics with a single maximum. Otherwise, a different strategy must be adopted. A general approach in order to characterise a posterior distribution is by generating samples of it and using their relative occurrences to approximate it. Obviously, sampling requires a thoroughly exploration of the space of model parameters. If the number of parameters is small enough, such exploration can be done systematically. Nevertheless, for high-dimensional problems a systematical approach is not feasible since the number of required sampled points grows too fast with the number of dimensions of the model space and, instead, a random exploration must be performed. A general and efficient way of exploring a large space is by using Monte Carlo methods.

The idea of Markov Chain Monte Carlo (MCMC) algorithms is to generate a random walk on the parameters space $\{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n\}$ such that its transition kernel, i.e. the probability of a jump $p(\mathbf{m}_i|\mathbf{m}_{i-1})$, only depends on the previous jump (Markov chain) and generates samples of $\mathbf{m}$ with a probability density equal to the posterior distribution. In other words, the transition kernel must be such that after an initial burn-in period and independently of the initial sample, the Markov chain will eventually converge to a unique stationary distribution which is the posterior distribution (Tarantola, 2005; Gregory, 2005). So, how to construct a Markov chain such that its stationary distribution is precisely the posterior distribution we are looking for? A common procedure to construct the desired transition kernel is the Metropolis-Hastings algorithm.

Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is a very well known and simple strategy to generate a random walk that approximates the posterior distribution. It is specially powerful for sampling high-dimensional distributions, i.e. physical models with high number of parameters.

From a known probability distribution it is quite simple to create a random walk by generating random samples of it. The Metropolis basic idea is to tune this random walk with some acceptance criteria such that some proposed samples are rejected and the remaining ones approximate the posterior distribution. A complete deduction and proof can be found in Mosegaard and Tarantola (2002) and Kaipio and Somersalo (2005). To implement this approach in the context of an inverse problem, let us first recall equation (4.9),

$$\sigma_{\mathrm{M}}(\mathbf{m}) = k\ \rho_{\mathrm{M}}(\mathbf{m}) L(\mathbf{m})$$

where the posterior distribution on the model parameters $\sigma_{\mathrm{M}}(\mathbf{m})$ is proportional to its prior $\rho_{\mathrm{M}}(\mathbf{m})$ times the likelihood $L(\mathbf{m})$. Then, random samples of $\mathbf{m}$ are generated from the prior distribution to generate a random walk in the parameter space in which each sample only depends on the immediate previous one. In this case, the random walk would simply approximate the prior $\rho_{\mathrm{M}}(\mathbf{m})$. To sample $\sigma_{\mathrm{M}}(\mathbf{m})$ instead, each proposed model is combined with its respective likelihood $L(\mathbf{m})$ to decide whether it is accepted according to a criterion that compares it with the previous accepted sample. In doing so, the random walk is shaped according to the misfit between predicted and observed data. The algorithm can be summarized in the following two steps:

1. *Proposal step.* A random sample $\mathbf{m}_{\text{new}}$ is generated either directly from its prior distribution $\rho_{\mathrm{M}}(\mathbf{m})$ or by means of a proposal distribution $q(\mathbf{m}_{\text{new}}|\mathbf{m}_{\text{old}})$ which should be easy to evaluate.

2. *Acceptance step.* The proposed sample is subjected to an acceptance test based

on the value of the so-called *Metropolis ratio*, r_m,

$$r_m = \frac{\sigma_{\mathrm{M}}(\mathbf{m}_{\mathrm{new}})\ q(\mathbf{m}_{\mathrm{old}}|\mathbf{m}_{\mathrm{new}})}{\sigma_{\mathrm{M}}(\mathbf{m}_{\mathrm{old}})\ q(\mathbf{m}_{\mathrm{new}}|\mathbf{m}_{\mathrm{old}})} \tag{4.22}$$

If $r_m \geqslant 1$, the proposed sample $\mathbf{m}_{\mathrm{new}}$ is accepted; otherwise, it is accepted with probability r which depends on how similar the two posteriors are (the more similar, the more likely to be accepted). If a proposed sample is not accepted, the random walk remains at the current value $\mathbf{m}_{\mathrm{old}}$ and a new proposed sample is generated. This second step can be summarized by defining an *acceptance probability*,

$$\alpha_{\ \mathbf{m}_{\mathrm{old}} \leftarrow \mathbf{m}_{\mathrm{new}}} = \min\left(1, r_m\right) = \min\left(1, \frac{\sigma_{\mathrm{M}}(\mathbf{m}_{\mathrm{new}})\ q(\mathbf{m}_{\mathrm{old}}|\mathbf{m}_{\mathrm{new}})}{\sigma_{\mathrm{M}}(\mathbf{m}_{\mathrm{old}})\ q(\mathbf{m}_{\mathrm{new}}|\mathbf{m}_{\mathrm{old}})}\right) \tag{4.23}$$

When this process is repeated many times, i.e. the number of accepted samples is large enough, the Metropolis algorithm guarantees that the relative ocurrence of a parameter m_i in the random walk is proportional to its posterior density. Although simple, this algorithm is very powerful since it can explore the space while preventing the random walk from getting stuck in a local maximum or high probability area by just including the case $r_m < 1$. Note that for a symmetric proposal distribution, the terms $q(\cdot|\cdot)$ cancel and the Metropolis ratio just compares the corresponding posteriors. This was the original algorithm that was later generalized by Hastings (1970) to account for non-symmetric proposals. Moreover, one important consequence of using a ratio as an acceptance criterion is that probability densities do not need to be normalized and, therefore, computing the constant k can be avoided. It is also important to mention that randomness is the main idea behind Monte Carlo methods where the generated samples are ideally expected to be independent. However, this is not the case in the explained algorithm, where samples of the random walk exhibit strong correlations since, by definition of Markov Chain, each sample depends on the previous one. To guarantee independence, not all accepted samples are then considered to estimate the posterior PDF, but only every l^{th} sample, being l the chosen interval to ensure samples are not related.

The *proposal distribution* $q(\cdot|\cdot)$ has a key role in the Metropolis algorithm. Although any proposal used to sample the prior will converge to the same stationary distribution, its choice completely determines the efficiency of the algorithm, i.e. the rate of convergence. Let us for instance imagine a proposal which is a multivariate Gaussian distribution with mean the current sample, $\mathbf{m}_{\text{new}}$, and standard deviation σ_p. The size of σ_p governs how efficiently the space of parameters is explored: if σ_p is large (big jumps) the search is explorative; whereas if σ_p is small (small jumps) the search is exploitative. None of them is good on its own and some problems can arise. With an explorative search, we are rapidly moving around the whole space of parameters, but it is highly unlikely to accept new movements (very low acceptance probability). Instead, in the exploitative case, the random walk is moving very slowly which increases the acceptance probability, but at the same time it may require much longer chains to explore the whole space (poor mixing). A compromise is essential (Tarantola, 2005; Gilks et al., 1996). From all the possible random walks that will converge to the posterior distribution, the goal is to design one that efficiently explores the whole space getting neither lost nor stuck. This is achieved by generating samples whose perturbation on the likelihood is as small as possible (to increase the acceptance rate and therefore likely to be accepted) but large enough in the parameter space to rapidly explore it. In other words large jumps in the model space that produce small perturbations in the data space. According to Gelman et al. (1996); Roberts et al. (1997), an optimal acceptance rate is approximately 50% for problems with one or two dimensions and decreases to 23% for high-dimensional problems. Finding the proposal that optimizes this behaviour is not straightforward and depends on the problem at hand. However, one thing is known, the more similar the proposal to the posterior, the more efficient the sampling.

An efficient sampling is particularly important in high-dimensional spaces (which is often the case for inverse problems). The main and intrinsic challenge of sampling high-dimensional spaces is that they are almost empty. In other words, the chances of "hitting" a high-probability region are very scarce and this is precisely the goal of sampling the posterior density. What is more, not only the posterior represents a small part of the whole space, but also not much is known about its location. That is the reason why the acceptance probability tends to be very low when sampling high-dimensional spaces and an efficient proposal is crucial. Although plenty of methods are developed focused on improving the sampling strategies (delayed-rejection, delayed-acceptance, parallel tempering, adaptive Metropolis and a long etcetera) we will mostly work with a simple Metropolis algorithm in this thesis, so that any additional improvement in the sampling will mostly have a positive impact on the whole algorithm. However, in section 5.4.1 the effect of the sampling strategy on the order reduction is analysed. Additionaly, an adaptive Metropolis will be presented in section 5.4.2 together with some examples to see its influence on the order reduction.

The following pseudo-code summarizes the Metropolis-Hastings algorithm explained above.

Algorithm 4.1 Metropolis-Hastings algorithm

```
 1: $\mathbf{m}_{\text{old}} \leftarrow$ get an initial model parameter $\mathbf{m}_{\text{o}}$
 2: repeat
 3:     Generate new sample $\mathbf{m}_{\text{new}}$ from a proposal distribution $q(\mathbf{m}_{\text{new}}|\mathbf{m}_{\text{old}})$
 4:     Compute $\sigma_{\text{M}}(\mathbf{m}_{\text{new}})$ as the product $\rho_{\text{M}}(\mathbf{m}_{\text{new}})L(\mathbf{m}_{\text{new}})$
 5:     Compute metropolis ratio $r_m$
            $r_m = \frac{\sigma_{\text{M}}(\mathbf{m}_{\text{new}})\ q(\mathbf{m}_{\text{old}}|\mathbf{m}_{\text{new}})}{\sigma_{\text{M}}(\mathbf{m}_{\text{old}})\ q(\mathbf{m}_{\text{new}}|\mathbf{m}_{\text{old}})}$
 6:     if $r_m \geqslant 1$ then
 7:         sample $\mathbf{m}_{\text{new}}$ is accepted: $\mathbf{m}_{\text{old}} \leftarrow \mathbf{m}_{\text{new}}$
 8:     else $r_m < 1$
 9:         sample $\mathbf{m}_{\text{new}}$ is accepted with probability $r_m$
10:     end if
11: until the number of samples is long enough
```

Lines 6-10 are equivalent to: $\mathbf{m}_{\text{new}}$ is accepted with probability $\alpha_{\ \mathbf{m}_{\text{old}} \leftarrow \mathbf{m}_{\text{new}}}$, as it directly combines the two cases. In general, in order to implement how to take a decision according to a probability one uses a uniform distribution. A random number u is obtained from a uniform distribution $U(1, 0)$; if $u < r_m$, the condition is true; otherwise it is false.

Despite its many advantages relative to traditional matrix-based inversions (Tarantola, 2005; Afonso et al., 2013a, 2016a), the use of MCMC methods to solve inverse problems has the drawback of having to evaluate $L(\mathbf{m})$ at each inversion step, which involves computing a full forward problem, thousands to millions of times. Consequently, the possibility of using a probabilistic approach to solve an inverse problem relies entirely on having efficient and accurate solvers for the forward problems (Section 3.2).

4.3 Application to a geophysical problem

The goal of all the inversions presented in this thesis is to infer the thickness of the lithosphere or equivalently the depth of the Lithosphere-Asthenosphere Boundary (LAB) from some synthetic observations: either the topography or the vertical mantle velocities. The forward problem associated to the inversion links the LAB values with the topography. More specifically, from the values of the LAB we can compute the viscosities and densities within the domain to characterize a Stokes problem. The solution of the Stokes problem is a velocity field which allows us to compute the normal stresses at the surface and, hence, the topography.

In this section a small synthetic example of an inversion is presented. The Metropolis algorithm is used as inversion algorithm and the Stokes flow associated to the forward problem is solved with a full order Finite Elements Method. The purpose of this example is to outline the basic setup and discuss several numerical considerations in order to characterize the problem prior to the coupling between the Model Order Reduction and the inversion. Moreover, the need for Model Order Reduction will be clearly grasped from the intrinsic high cost of solving a large number of forward problems with full order methods.

4.3.1 Model discretization and parametrization

In all the examples of this thesis the domain represents a portion of the Earth and is discretized with Taylor-Hood hexahedral elements. They are elements of order two for velocity and order one for pressure (Q2-P1) and satisfy the LBB condition (Figure 4.2). This element requires 8 vertex nodes for the pressure, and 20 nodes for the velocity. Unlike the 27 nodes hexahedral, the 20 nodes one does not include a node in the middle of the faces.

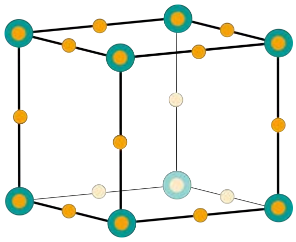

Figure 4.2: Taylor-Hood hexahedral element (Q2-P1)

This discretization generates a numerical domain that can be seen as a collection of nonoverlapping adjacent rectangular columns (Figure 4.3). On the boundary of the domain we apply free slip boundary conditions; the box is confined (normal velocity is zero) and there is no friction on the boundary (tangential stresses are zero). Denoting the boundary surfaces as Γ_S the free slip boundary conditions read,

$$\begin{cases} \boldsymbol{u} \cdot \boldsymbol{n} = 0 & \text{on } \Gamma_S \\ \boldsymbol{n} \cdot \sigma \boldsymbol{\tau}_i = 0 & \text{on } \Gamma_S \end{cases} \tag{4.24}$$

where $\boldsymbol{n}$ is the normal vector and $\boldsymbol{\tau}_i$, for $i = 1, 2$, are the two corresponding tangential directions such that they build an orthonormal system of vectors $\{\boldsymbol{n}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2\}$. When σ is replaced by the Stokes law (2.13), we obtain $\boldsymbol{n} \cdot (\nabla^{\mathrm{S}} \boldsymbol{u}) \boldsymbol{\tau}_i = 0$ since the pressure term vanishes due to the orthonormality between $\boldsymbol{n}$ and $\boldsymbol{\tau}_i$. However, due to the nature of the Finite Elements method, the Neumann boundary conditions are automatically enforced and, therefore, only the first equation needs to be implemented.

A unique model parameter is assigned to each of the columns of the domain representing the LAB depth at that location (Figure 4.3). Therefore, the vector of model parameters $\mathbf{m}$ to be retrieved by the inversion is composed of the LAB depths for all columns in the numerical domain. In doing so, the LAB is completely characterized.

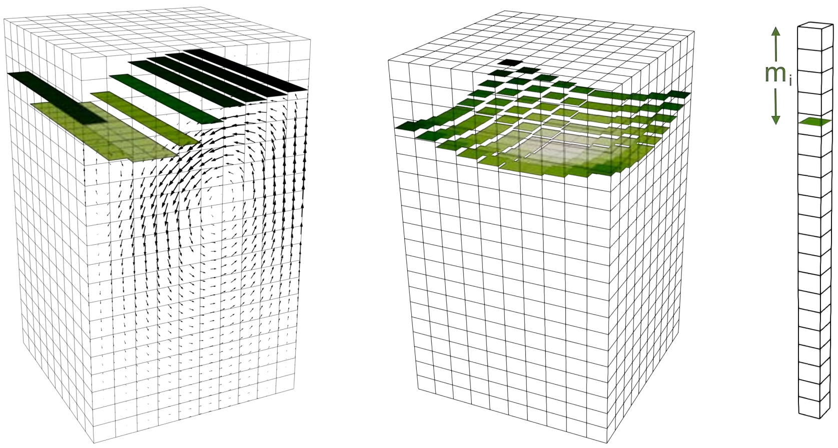

Figure 4.3: Parameters of the inversion describing the LAB depth for two different examples. The example on the left also shows the corresponding velocity field (black arrows)

Viscosity and density

The LAB depth is the main parameter controlling the flow pattern inside the numerical box as it defines the internal temperature distribution, which in turn controls the viscosity and internal buoyancy forces. For illustrative purposes, the temperature inside each column is assumed to follow a linear profile defined by the temperature at the surface ($T_0 = 293$ K) and that at the LAB depth ($T_{\mathrm{LAB}} = 1523$ K) for the listhosphere. In the asthenosphere, we also assume a constant linear gradient between T_{LAB} and a fixed bottom temperature ($T_{660} = 1873$ K). Absolute pressure is assumed to follow a lithostatic profile,

$$p = \rho_0 g z, \tag{4.25}$$

where g and z are gravity acceleration and depth, respectively.

The buoyancy forces that drive convection are produced by density variations of thermal character. Therefore, the density is computed as a deviation from a reference value with,

$$\rho = \rho_0(1 - \alpha(T - T_0) + \beta(p - p_0)) \tag{4.26}$$

where ρ_0 is the reference density at T_0 and p_0, and α and β are the coefficients of thermal expansion and compressibility, respectively.

Both experiments and theory indicate that a general form of relationship between strain rate and deviatoric stress valid for diffusion and dislocation creep is given by,

$$\dot{\varepsilon} = A\left(\frac{\tau}{G}\right)^n \left(\frac{b}{d}\right)^m \exp\left(-\frac{E + pV}{RT}\right) \tag{4.27}$$

where A is the pre-exponential factor, G is the shear modulus, d is the grain size, b is the lattice spacing, E is the activation energy, V is the activation volume, R is the universal gas constant, n is the stress or flow law exponent and m the grain size exponent, (Schubert et al., 2004). As we can see, with $n = 1$ the relation between strain rate and deviatoric stress is lineal resulting in a Newtonian rheology (e.g. difussion creep), whereas for $n > 1$ the relation is nonlinear resulting in a nonlinear viscous rheology (e.g. dislocation creep). This parametric relationship can be calibrated experimentaly both for diffusion and dislocation creep mechanisms.

Generalizing it into a full constitutive relation by making use of invariants, we can identify from equation (2.12) the viscosity function as a power-law of the form,

$$\mu = A_D^{-1/n} \dot{\varepsilon}_{\mathrm{II}}^{(1/n)-1} \exp\left(\frac{E + pV}{nRT}\right) \tag{4.28}$$

where $\dot{\varepsilon}_{\mathrm{II}} = (1/2)\varepsilon_{ij}\varepsilon_{ij}$ is the second invariant of the strain rate and A_D a constant depending on the material. As usual in geodynamic simulations, a threshold (maximum) viscosity, $\mu_{\max}$, is imposed. All values used are listed in Table 4.1. Because the methodology presented here is suited for linear problems we linearize the Stokes equations by assuming a constant strain rate $\dot{\varepsilon}_{\mathrm{II}} = 10^{-15}$ s^{-1}. We note, however, that the methodology can be adapted to nonlinear problems and it is independent of the specific parametrization used here.

Symbol	Name	Value	Units
g	gravity	9.8	m s^{-2}
ρ_0	reference density	3300	kg m^{-3}
p_0	reference pressure	0	MPa
T_0	reference temperature	293	K
T_{LAB}	LAB temperature	1523	K
T_{660}	bottom domain temperature	1873	K
α	thermal expansion coefficient	10^{-5}	K^{-1}
β	compressibility coefficient	10^{-5}	MPa^{-1}
$\mu_{\max}$	threshold for maximum viscosity	10^{24}	Pa s
A_D		1.1×10^5	MPa^{-n} s^{-1}
$\dot{\varepsilon}_{\mathrm{II}}$	second invariant of the strain rate	10^{-15}	s^{-1}
E	activation energy	5.3×10^5	J mol^{-1}
V	activation volume	14	J MPa^{-1} mol^{-1}
R	universal gas constant	8.314	J mol^{-1} K^{-1}
n	power flow exponent	3.5	

Table 4.1: Values of the parameters used to compute material properties

Finally, due to the high contrasts between viscosities and densities, in geophysical problems it is typical to scale them. In doing so, their orders of magnitude are brought closer and matrix conditioning problems reduced.

4.3.2 Synthetic observables

Two different kinds of synthetic measurements or observations will be used in this thesis: vertical mantle velocities and topography. Actually, mantle velocities cannot be directly observed and other indirect observables such as the topography need to be used. However, a number of additional assumptions and secondary quantities need to be considered in order to obtain topography from the forward problem. Therefore, in order to focus on the intrinsic features and performance of the proposed method, we prefer starting with velocity as an observable and avoid the extra complexity introduced by the topography model. The synthetic measurements used on first examples will be the vertical mantle velocities at specific points within a predefined observation region Γ. In addition, velocities allow us to better isolate and analyse the main RB effects within the MCMC simulations that are presented in Chapter 5. A more complex example using topography as observable is studied in Section 6.1 and all details thoroughly explained.

We create a dataset of synthetic observables to be used in our MCMC simulations from a reference or “true” model, i.e. a particular set of m_i values. The vertical components of the velocity field are computed at n_d points and define the dataset of synthetic observables, $\mathbf{d}^{\mathrm{obs}}$. The number of data points is specified for each example and vary depending on the number of unknown parameters in the inversion. A Gaussian noise, $r \sim \mathrm{N}(0, \sigma_d)$, is used to introduce some controlled error in the synthetic observables. The standard deviation of the data errors, σ_d, is chosen to be a percentage of the maximum absolute value in the dataset $\mathbf{d}^{\mathrm{obs}}$,

$$\sigma_d = \sigma_e \max |\mathbf{d}^{\mathrm{obs}}|. \tag{4.29}$$

Since here we are interested in studying the method rather than the actual solution to the inverse problem, in the following examples the error in the data is only included in the likelihood function; i.e. no noise is actually added to the synthetic data. The reason for this choice is that when perturbations are added to the synthetic data, the posterior PDFs will not necessarily converge to the reference model, but to a “perturbed” model that best fits the contaminated observables, thus complicating the performance analysis of the method. In the same way, since we want to be sure that the synthetic data are within the interpolating space, the mesh used to compute the synthetic observables is the same as the one used during the inversion to predict them. This is the so-called inverse crime (Kaipio and Somersalo, 2007). Although we acknowledge it can lead to too optimistic results, our main goal is not to get accurate results but to study the features and main factors controlling the coupling between MCMC and RB (Chapter 5). One example including perturbations to the observables and another one using a different mesh (twice finer) to create the synthetic observables are included in Appendix A.2 and A.1, respectively.

The observation region Γ is defined as a box of 100km along the true model of parameters (±50km on each side). This region is defined in accordance with the *a priori* information on the parameters used in the MCMC simulation which is chosen to cover as well ±50km around the true model. In doing so, the observation region is placed exactly where the main changes in the velocity field are expected or, in other words, where the model parameters will have the largest influence. Moreover, as explained in Section 5.2.1, such region also matches the region of interest used to compute the QoI and controls the accuracy of the RB solution.

4.3.3 Priors, sampling strategy and likelihood function

In order to test the inversion in adverse conditions, a relatively uninformative initial scenario is defined. As previously mentioned, the prior PDF for all parameters is set as uniform within a range of 100km (i.e. ±50 km from the reference LAB value). With these priors, all gain of information, as contained in the posterior PDF, will be controlled exclusively by the likelihood since prior terms will cancel in the Metropolis ratio. This allows a readily interpretation of the results.

The prior PDF is sampled using a multivariate Gaussian proposal distribution centered at the current state of the chain $\mathbf{m}_{\text{old}}$,

$$\mathbf{m}_{\text{new}} \sim \text{N}(\mathbf{m}_{\text{old}}, \sigma_p^2). \tag{4.30}$$

and with σ_p determining the size of the move. Its value will be specified independently for each example. For each proposed move $\mathbf{m}_{\text{new}}$, only one parameter m_i, $i \in [1, n_m]$, is modified with respect to $\mathbf{m}_{\text{old}}$. In other words, each parameter has its own one-dimensional proposal distribution from which it is sampled, $m_{i,\text{new}} \sim \text{N}(m_{i,\text{old}}, \sigma_p^2)$. Therefore, we modify the LAB structure by changing only one column per iteration. Symmetric proposal distributions lead to the initial Metropolis algorithm which only includes posterior terms and proposal ones do not appear.

For the likelihood function, so far we neglect modeling errors and assume that observation uncertainties are uncorrelated. Moreover, we assume that all observations exhibit the same standard deviation. The misfit function is then,

$$e(\mathbf{m}) = \frac{1}{2\sigma_d^2} \sum_{i \in \Gamma} \left(u_{zi}(\mathbf{m}) - d_i^{\text{obs}} \right)^2 \tag{4.31}$$

where i ranges from 1 to the number of points n_d in the region of interest Γ and d_i^{obs} and $u_{zi}(\mathbf{m})$ are the "observed" (synthetic) and predicted vertical velocities at point i, respectively. The standard deviation σ_d is related to the error in the measurements and is defined in equation (4.29).

An important numerical aspect to consider about the likelihood is its possibility to underflow. If the number of observables is very large, the misfit function from equation (4.16) (also (4.15)) can exhibit very large values causing the likelihood to be very small and in some situations to numerically underflow to zero. Or equivalently, since the likelihood is a product of terms, if such terms are smaller than one, their multiplication gives even smaller values which may lead to underflow. To avoid such situation, sometimes it is useful to define the *log-likelihood*. Working with logarithms transforms the products in sums avoiding, therefore, getting values which are too small to be represented by a floating point number. The log-likelihood equivalent to the likelihood from equation (4.14) is defined as,

$$\log L(\mathbf{m}) = \log k - e(\mathbf{m}) \tag{4.32}$$

where the logarithm corresponds to the natural logarithm, i.e. the logarithm with base e. In order to use the log-likelihood in the Metropolis ratio, equation (4.22) must be also rewritten in terms of logarithms. Therefore, the logarithmic Metropolis ratio (considering uniform priors and symmetric proposals) reduces to,

$$\log r_m = \log \left(\frac{L(\mathbf{m}_{\text{new}})}{L(\mathbf{m}_{\text{old}})} \right) = e(\mathbf{m}_{\text{old}}) - e(\mathbf{m}_{\text{new}}) \tag{4.33}$$

The acceptance condition is now: if $\log r_m \geqslant 0$, the proposed samples $\mathbf{m}_{\text{new}}$ is accepted; otherwise, it is accepted with probability $\log r_m$. Although in most of the cases the use of the likelihood or the log-likelihood produce exactly the same results, the log-likelihood formulation is safer.

4.3.4 Example results

The domain represents a portion of the Earth of 400km × 400km going from the surface down to 660km depth and is discretized in 10×10×20 elements. The LAB shape to be recovered (Figure 4.4a) is described with 25 parameters in such a way that each parameter covers four columns of finite elements. The initial model (starting point of the Markov chains) has a constant LAB at 130 km depth, except in one column, where the LAB is at 100 km depth (Figure 4.4b). This is simply to avoid an initial problem with zero velocity field. Regarding the number of synthetic observables, we fixed $n_d = 675$ data points within the observation region Γ (Figure 4.7a).

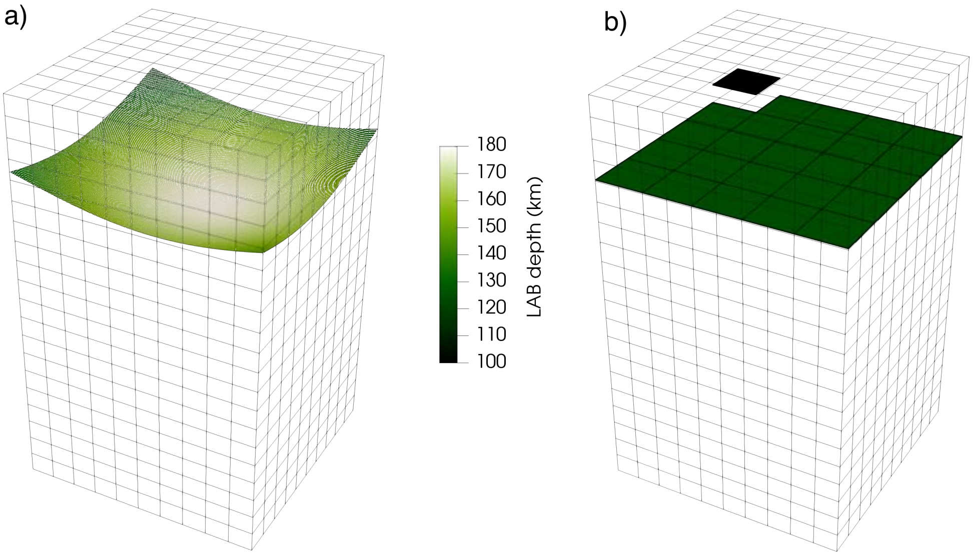

Figure 4.4: 3D cartesian domain discretized in 10×10×20 mixed elements representing a portion of Earth of 660km depth and 400km in the two directions along the surface. (a) Synthetic reference LAB and (b) initial model configuration.

We run a total of 10^5 steps in the MCMC simulation. The first 10^4 values considered burn-in and thus discarded for later analyses. The standard deviation of the proposal distribution in (4.30) is set at $\sigma_p = 5$ km and the percentage of the introduced noise in (4.29) is $\sigma_e = 10^{-1}$. The results of this inversion are summarized in Figure 4.5. Despite the uninformative initial scenario (uniform priors and flat initial state), these results demonstrate that the MCMC+FE inversion algorithm succeeded in obtaining an accurate representation of the posterior PDF for all parameters. The gain in information is clear. The trace plots also show that the inversion converged quickly to the true posterior PDF (LAB values + their associated uncertainties). At the beginning most of the parameters started far from the true value, but after the burn-in period they all converged to the true value and remained oscillating around. The posterior distributions for each parameter exhibit mean values which coincide well with the reference synthetic LAB depths; all within one standard deviation (Table B.1 from Appendix B). The acceptance rate of the inversion is around 40%.

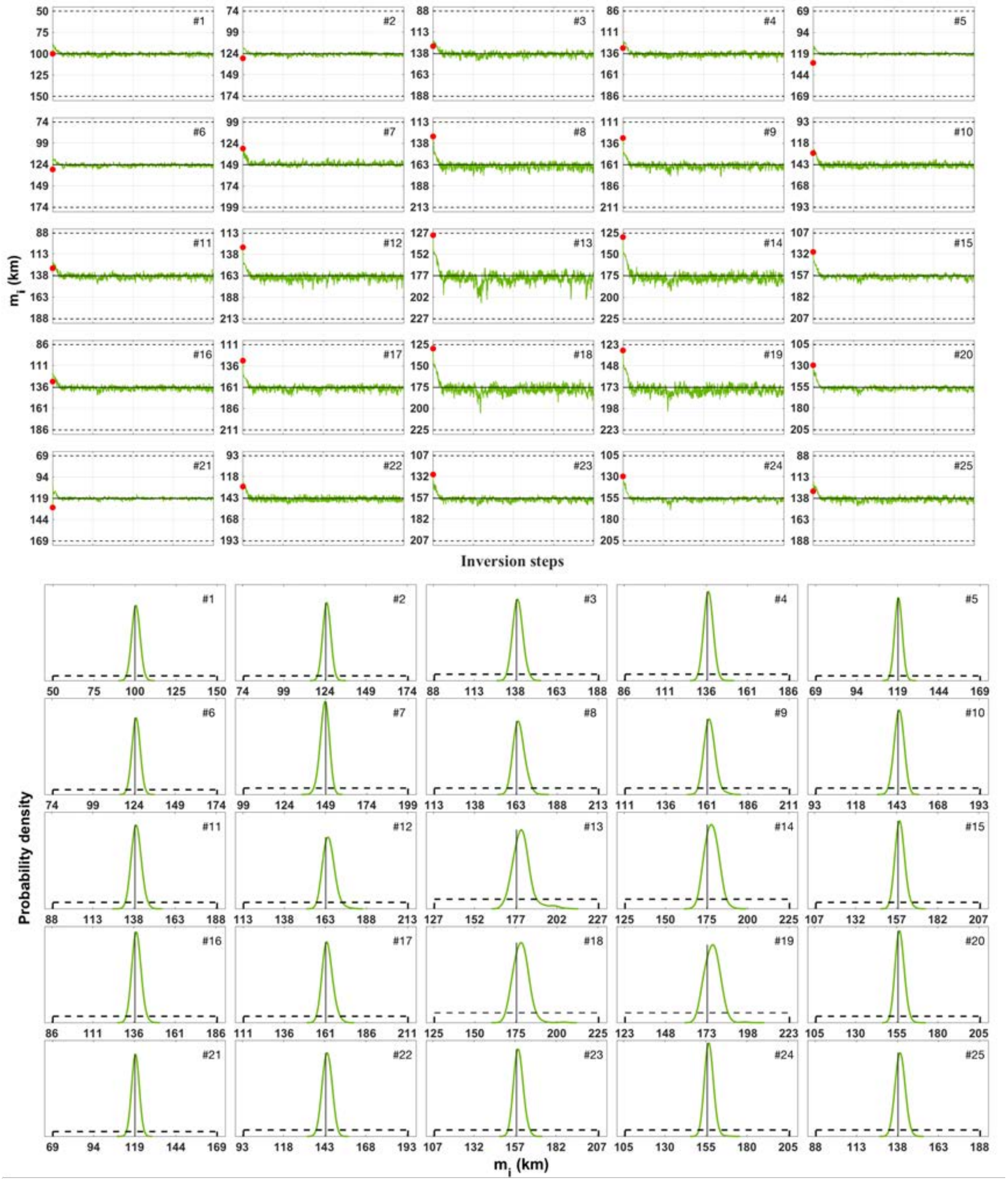

Figure 4.5: Random walk (top) and posterior PDF (bottom) for each of the 25 parameters. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot). The burn-in period has been discarded in the estimation of the posterior PDF. Values of the mean and the standard deviation for each posterior PDF are found in Table B.1 from Appendix B.

An inherent problem of all Monte Carlo methods is where to stop the search being sure that all space has been sufficiently explored and no high probability regions are missed. These convergence issues are problem dependent and we do not discuss them here (cf. Gelman and Rubin, 1992; Brooks and Gelman, 1998). Similarly, the length of the burn-in period is also problem dependent and we do not pursue to implement any specific method (cf. Geweke, 1991). Instead, given the synthetic character of the example, both the length of the simulations (as a convergence criterion) and the burn-in period are chosen by visual inspection of the chains during the simulation. In this particular example we reduced the number of inversion steps due to the high cost of FE.

Another interesting way to see convergence is by plotting the evolution of the misfit $e(\mathbf{m})$ between predicted and observed data, equation (4.15). The misfit evolution is shown in Figure 4.6. The initial misfit is around 2300 but we made zoom in the y-axis for a better visualization.

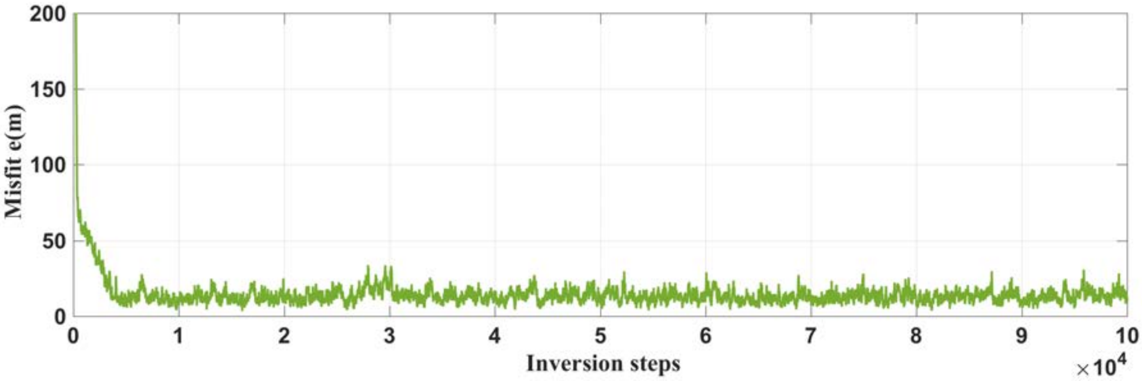

Figure 4.6: Evolution of the misfit $e(\mathbf{m})$ during the inversion.

Initially the difference $\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\text{obs}}$ is very large because the model $\mathbf{m}$ is not good at explaining the data $\mathbf{d}^{\text{obs}}$ and the misfit exhibits large values. However, as the random walk converges to the true value, the misfit quickly decreases until it stops and remains oscillating around. Although plotting the misfit helps checking the convergence of the inversion as a whole, it does not guarantee that convergence is achieved for each parameter individually. To guarantee that, each random walk must be inspected individually (Figure 4.5).

It is also very important to check whether the proposed models of parameters are truly predicting the synthetic data values. Figure 4.7a shows the location of the 675 data points where the synthetic observables are computed. They all lay within the observable region Γ (±50km along the true model parameters, Section 4.3.2) and are distributed in 3 layers. We selected 49 points of the second layer whose location is shown in Figure 4.7b. However, we checked all of them to ensure their behaviour is the same.

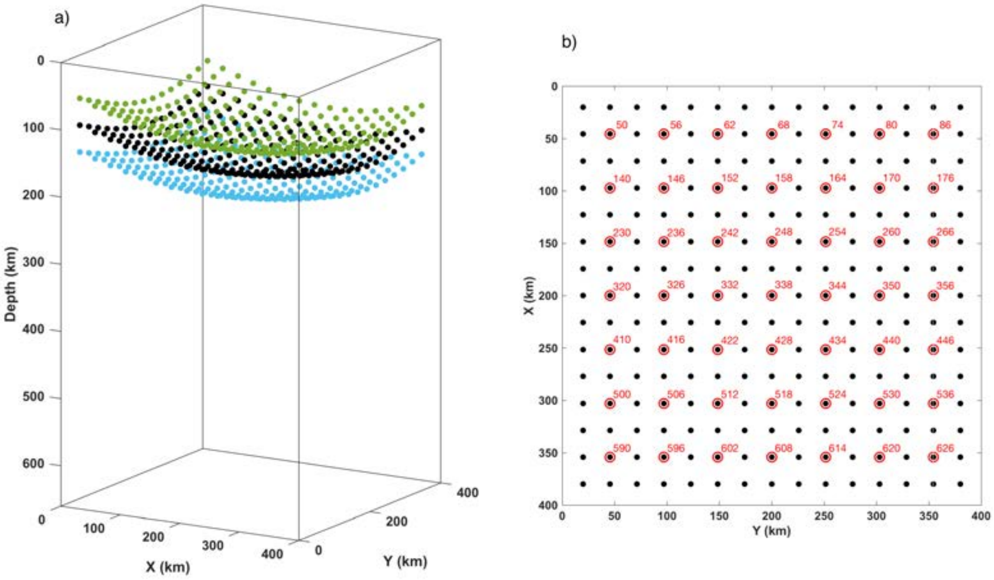

Figure 4.7: (a) Location of the 675 data points within the observation region Γ and (b) location and ID of the 49 data points (red circle) chosen to plot their predicted values in Figure 4.8.

Figure 4.8 shows histograms of the predicted values during the inversion for the 49 selected observable data points. The fact that all histograms include the synthetic value is a good indicator, since it means that each synthetic observable can actually be predicted by a certain configuration of model parameters. Moreover, not only histograms include the synthetic value, but they are also centered around them, which indicates that mean values are successfully recovered by the inversion.

The width of these distributions indicates how sensitive each observable is to perturbations of the parameters. Those observables exhibiting wider distributions are more sensitive since the same perturbations are able to produce a larger range of data values. Instead, a very insensitive observable would exhibit a very narrow distribution because no matter what is the size of the perturbation, but the predicted value

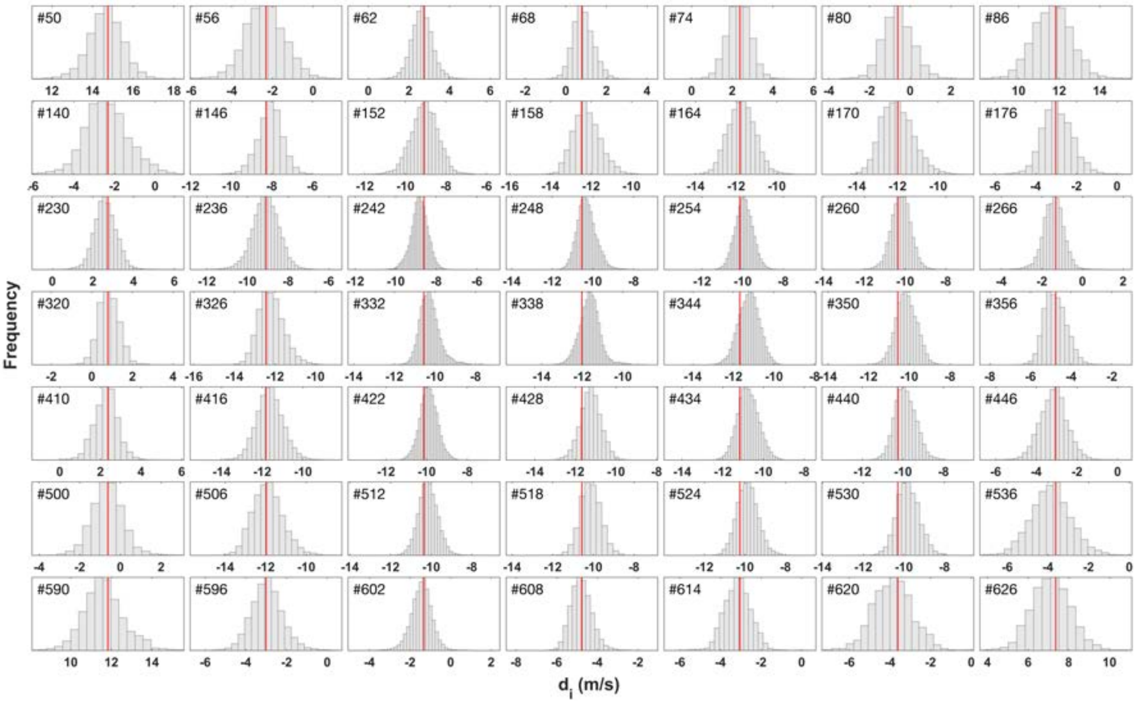

Figure 4.8: Histograms of the predicted data values for 49 data points (out of 675). In each plot: synthetic data value (red line) and histogram of the predicted values. The burn-in period is also discarded. All values of the x-axis should be multiplied by 10^{-13}. The physical location of each observable is shown in Figure 4.7.

would not be affected.

The MCMC+FE inversion scheme required to solve as many full forward problems as inversion steps. In this particular example, we solved 10^5 FE each of them taking around 6 seconds. This corresponds to a total of 166 hours or approximately 7 days. Consider now a large-scale high-dimensional geophysical inversion, where each forward solution is much more expensive and longer simulations are required to obtain reliable posteriors for each parameter. The total computational cost turns unaffordable and the MCMC+FE approach is not feasible anymore. It is obvious that to deal with such problems, an efficient solver for the forward problem is crucial.

Chapter 5

Coupling Markov Chain Monte Carlo with Reduced Basis Method

In this chapter the Markov Chain Monte Carlo (MCMC) is coupled with the Reduced Basis (RB) approach for the solution of inverse problems and the main features are studied. The probabilistic approach of solving inverse problems requires the solution of a large number of high-fidelity forward models whose high computational cost is usually an important bottleneck for the inversions. Considering the characteristics and similarities between the forward problems to be solved during the inversion, the RB strategy is an excellent approach at reducing the cost. In addition, the converging nature of MCMC-based inversion makes the option of RB solvers extremely attractive, as it is a situation where RB performs at its best.

In Section 5.1 the strategy MCMC+RB is explained with a description of the procedure to build a basis and its main characteristics followed by an illustrative example. The remaining sections of this chapter are devoted to analyse the aspects controlling the cost of solving a RB forward problem within the MCMC+RB scheme. Generally, the cost of obtaining a RB solution is mainly determined by the time required to create the reduced system (specifically $\mathbb{K}_{\text{RB}}$), since solving the reduced problem has no further complication due to its low dimension. The reduced stiffness matrix $\mathbb{K}_{\text{RB}} = \mathbf{B}^\mathsf{T}\mathbf{K}\mathbf{B}$ requires a fast computation and it depends on two factors: the size of the reduced basis $\mathbf{B}$ and the assembly process. Sections 5.2, 5.3, 5.4 are focused on controlling the basis size, whereas 5.5 is based on the assembling efficiency. Some results presented in this chapter are published in Ortega-Gelabert et al. (2020).

5.1 Building a basis tailored to the inverse problem

Classical Reduced Order methods include an *offline* stage where all costly computations are performed and an *online* stage where the reduced order solver is used to obtain fast an computationally inexpensive solutions (Quarteroni et al., 2016; Hesthaven et al., 2016; Ito and Ravindran, 1998; Prud'homme et al., 2002). In classical approaches, the process of building the reduced basis is usually part of the offline stage since it requires computing many high-fidelity solutions. The success of such strategies depends essentially on the sampling of the parameter space in order to create a reliable and accurate reduced basis. While in principle a similar splitting can be used in the context of a probabilistic inverse problem (Galbally et al., 2009; Wang and Zabaras, 2005; Lieberman et al., 2010), it is not always straightforward to define the sampling, that is, to predict which high-fidelity solutions (snapshots) should be computed *a priori* during the offline stage to be used later during the inversion, as there is no way of knowing in advance neither the high probability region nor the path the MCMC will take to find it. This is precisely the goal of the inverse problem! For this reason, we prefer to combine the offline and online stages into a single stage where the reduced basis is created *on the fly* as a response to the need of the inversion process. In other words, the process of building the reduced basis is coupled within the inversion scheme.

We propose a greedy approach, in which the basis $\mathbf{B}$ is initially empty and is subsequently enriched with FE (high-fidelity) solutions as the inversion progresses. The first forward problem required by the inversion is solved with the high-fidelity FE solver and its solution is added to the basis. From then on, every time a new forward problem needs to be solved, the solution is sought in the RB space $\mathbb{R}^{n_{\text{RB}}}$ and the error of the RB approximation is assessed. If the error is larger than a predefined tolerance, e_{RB}, the problem is solved using FE and the new high-fidelity solution is added to $\mathbf{B}$ (i.e. the basis is enriched). Therefore, the basis is only enriched when the accuracy of the RB solution is below the specified tolerance. Figure 5.1 shows the flowchart of the building process of a reduced basis simultaneously coupled within an inversion framework.

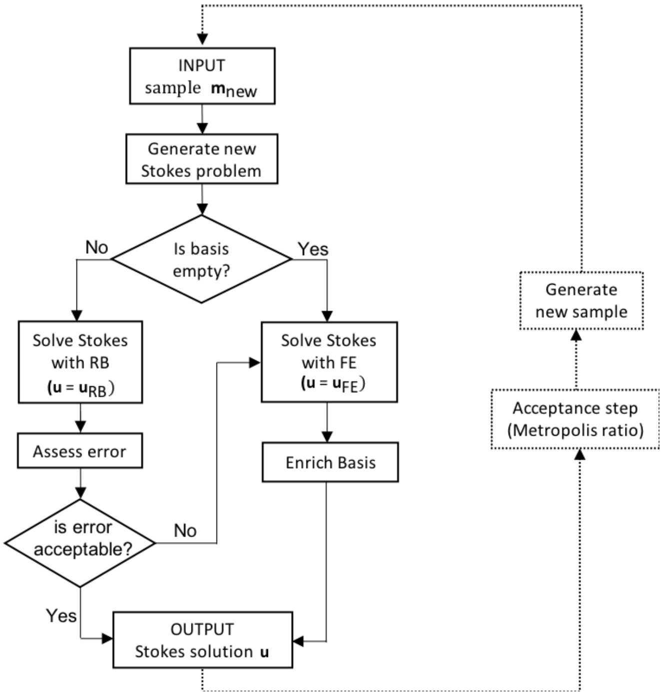

Figure 5.1: Flowchart of the greedy approach to build a reduced basis within a probabilistic inversion scheme

Since a MCMC inversion gradually converges to an equilibrium distribution which usually spans only a limited region of the initial parameter space, it is expected that the basis will initially grow until it becomes rich enough to provide accurate approximations (within a tolerance) to any subsequent forward evaluation. In this way, the basis does not need to provide accurate solutions for the entire parameter space, but it is automatically tailored to provide accurate solutions within the region of high probability as sampled by the MCMC. A similar data-driven strategy to build the reduced basis is implemented by Cui et al. (2014). Unlike classical approaches where the choice of snapshots is based on a priori information, with this greedy strategy the exploration of the parameter space to choose the snapshots is lead by the posterior distribution. Consequently, the size of the basis to attain the same accuracy (in the high-probability regions) is smaller in comparison with the offline approach.

One slight drawback of this approach is that instead of relying on fast RB solutions since the beginning of the inversion, some computational cost must be spent in creating the basis *on the fly*. In most practical cases, however, this is not a problem. The overall cost involved in computing the basis during the inversion process is significantly less than that invested in creating sufficiently accurate basis in an offline stage, except when the parameter space is relatively small (an uncommon case in geophysical inversions). Moreover, the basis created in one inversion can always be reused in subsequent inversions of similar characteristics. Either by using it as a starting basis (instead of an initially empty basis) which is enriched as the inversion requires or as a in classic offline/online stage approach.

The proposed greedy approach to build the RB does not guarantee orthogonality between the members of the basis. Redundant (linearly dependent) basis elements may still be included in the basis, which would result in an ill-conditioned reduced linear system to be solved (3.32). In practice, however, we have not observed this behaviour. The condition number of the reduced system in our problems remains almost constant as the basis grows, indicating that the basis is not deteriorated by the greedy character of the algorithm. In fact, by construction, solutions which are linearly dependent (according to a predefined tolerance) should not have been added to the basis. Nevertheless, in the case of having redundant information in the basis, a Singular Value Decomposition procedure (Quarteroni et al., 2016; Brunton and Kutz, 2019) can be applied on the fly to obtain an orthogonal basis. Other possibilities include orthogonalizing the reduced basis every time a new snapshot is added, as in Lieberman et al. (2010), Cui et al. (2014), and Bui-Thanh et al. (2008).

Finally, as the Markov chain moves to a more restricted parameter space, the space $\mathbb{R}^{n_{\mathrm{RB}}}$ can be inspected at regular intervals and processed to remove bases that are not contributing with significant information (Section 5.4.3). The result is a more compact and data-driven basis tailored for the specific problem at hand. However, care must be taken, since it may happen that the element of the basis that is taken as irrelevant is proposed again in the future. In such situation and if the element was previously removed from the basis, an extra FE would be required since the reduced basis would no longer be able to represent it. Therefore, there exists a trade-off between a slightly larger basis and the possibility of future additional high-fidelity computations.

In the next three subsections the greedy approach to build the basis is explained.

The first one presents a small but important modification to guarantee a fair coupling between the MCMC and RB. Then, the inversion example from Section 4.3 is repeated with the MCMC+RB coupled approach. And, finally, the effect of the additional error introduced by the RB approximation is analysed.

5.1.1 Metropolis ratio and a tailored basis

The success of the Metropolis algorithm depends on the acceptance criterion which is the responsible of adding new samples to the chain. It is based on the Metropolis ratio, equation (4.22), that compares the value of the posterior for the current sample $\sigma(\mathbf{m}_{\text{old}})$ with the posterior of the proposed one $\sigma(\mathbf{m}_{\text{new}})$. The evaluation of the posterior requires the solution of a forward problem which has an associated error/accuracy. For this comparison to be fair, both posteriors should be evaluated with solutions computed with methods exhibiting the same level of accuracy. In other words, it is not fair to compare a model whose solution is obtained with FE with a model whose solution is obtained with RB, as the space in which each solution in sought is different.

This is precisely what happens in the greedy procedure to build the reduced basis explained above. Consider the case when the error of the RB solution for a proposed model is above the predefined threshold and thus a high-fidelity FE solution is required. If the forward problem of the current state in the chain was accurately solved in a previous step with the RB surrogate, the comparison in (4.22) is now performed between a RB approximation (current) and a high-fidelity FE solution (new proposed model). Since both of these solutions have different intrinsic accuracies/errors, the comparison is no longer objective, and the Metropolis criterion is corrupted. In practice, significant differences in their respective errors are not uncommon and may result in a divergent chain and/or an inaccurate representation of the posterior. When this is the case, all random walks typically become “stuck” as new proposed models tend to have higher misfits and therefore are rejected (Figure 5.2). Not only the random walks stop oscillating, but they also generate high-probability peaks in the posterior PDF at usually wrong values. The richer is the basis, the lower are the chances of getting stuck, since RB solutions are closer to FE ones. Now consider the opposite case, where the proposed model is accurately solved with the RB surrogate and the current model was solved with a high-fidelity FE solver. The comparison in (4.22) remains nonobjective. However, unlike the previous case, the random walk does not get stuck since, here, the current model is the most accurate one, i.e. solved with the high-fidelity solver.

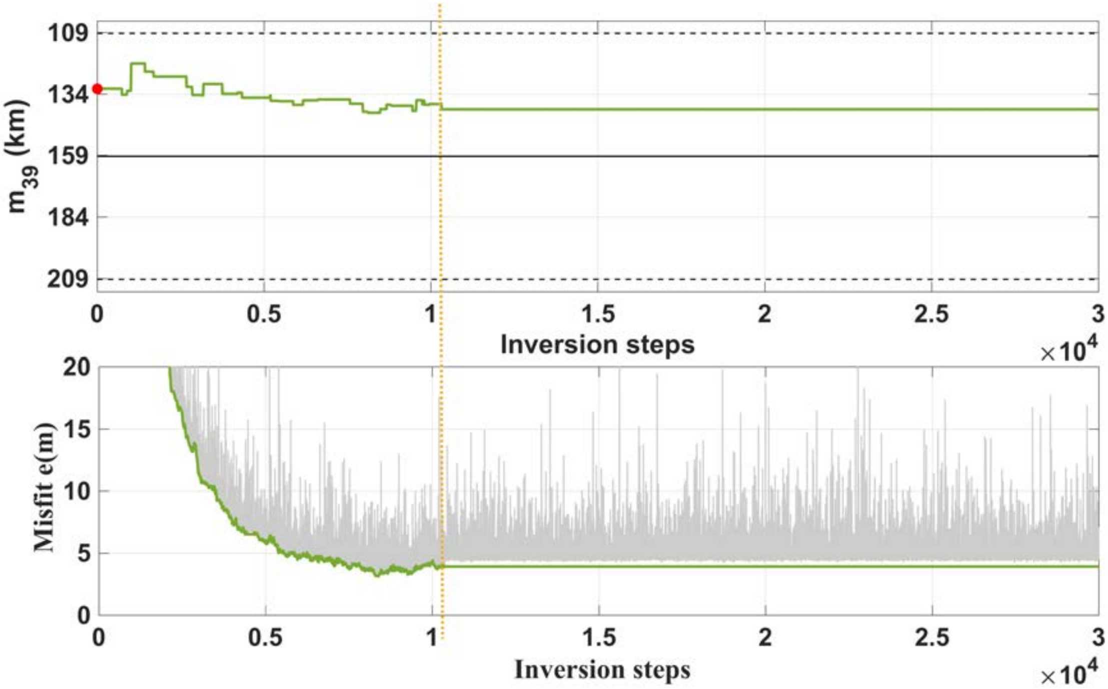

Figure 5.2: Example of a "stuck" random walk for a representative parameter (top) and evolution of the misfit (bottom). From around step 10^4 onward, all proposed models exhibit significantly larger misfits (gray) than the already accepted current value (green). Consequently, the models are rejected and the random walk cannot move to any other value.

It is important to remark that this unfair scenario mainly takes place during the building process of the reduced basis since it is when high-fidelity solutions are present. Once the basis is rich enough, all successive problems are fairly solved using the same RB and, therefore, relying on fair comparisons. The number of "wrong" samples that may end up accepted (or not) is statistically insignificant in comparison with the length of the random walk and even more considering that they mostly occur during the burn-in period which is always discarded for the estimation of the posterior PDF. Finally, the chances of getting stuck minimise as the basis gets richer, since RB solutions are closer to FE ones and, therefore, comparisons tend to be more objective. However, in order to guarantee a fair comparison and to prevent the inversion from getting stuck, two modified procedures to build the basis are proposed.

The problem described above can be avoided with a small modification to the original algorithm. Given the Markovian property of the chain (i.e., dependence only on the current state of the chain), the only strong requirement is that the comparison in (4.22) remains objective. One possible solution is to duplicate FE solutions: every time a proposed model is solved with FE, we also recompute the previous model with FE (unless it was already FE). In doing so, the comparison is always performed with solutions in the same space and exhibiting the same accuracy. The recomputed FE solution of the previous model is only used for the comparison in (4.22) and is not added to the basis; otherwise, it would provide redundant information to it. While this simple approach guarantees that the comparison is objective, it has the drawback of requiring additional FE solutions.

Another modification to circumvent the comparison problem consists on excluding all high-fidelity FE solutions from the comparison. When the error of a RB solution is above the tolerance, the high-fidelity solution is computed and the basis is enriched. However, no comparison is performed and a new sample is proposed for the same inversion step. Additionally, the RB solution of the current sample is recomputed with the new enriched basis. In doing so, high-fidelity solutions are relegated to only build the basis and comparison is always fair. Both strategies have been implemented and no differences were observed in the inversion results. However, following the idea of reducing the number of high-fidelity solutions required during the inversion, the second modification is the one used in all the examples presented in this thesis. Figure 5.3 shows the flowchart of the procedure.

This modification might affect the ergodicity and detailed balance condition of the chain. However, it only occurs during the building process of the basis which essentially coincides with the burn-in period and those samples are directly discarded. Therefore, convergence properties for the remaining part of the chain remain intact. Other approaches can certainly be devised; a complete description of them, as well as the effect on the convergence and ergodicity of the chain are not part of this thesis and is left as future work. In Manassero et al. (2020), the Metropolis ratio is modified using the idea of the Delayed Rejection algorithm proposed by Mira (2001) so that ergodicity of the method is guaranteed.

This difficulty does not exist in simpler approaches where a single surrogate is

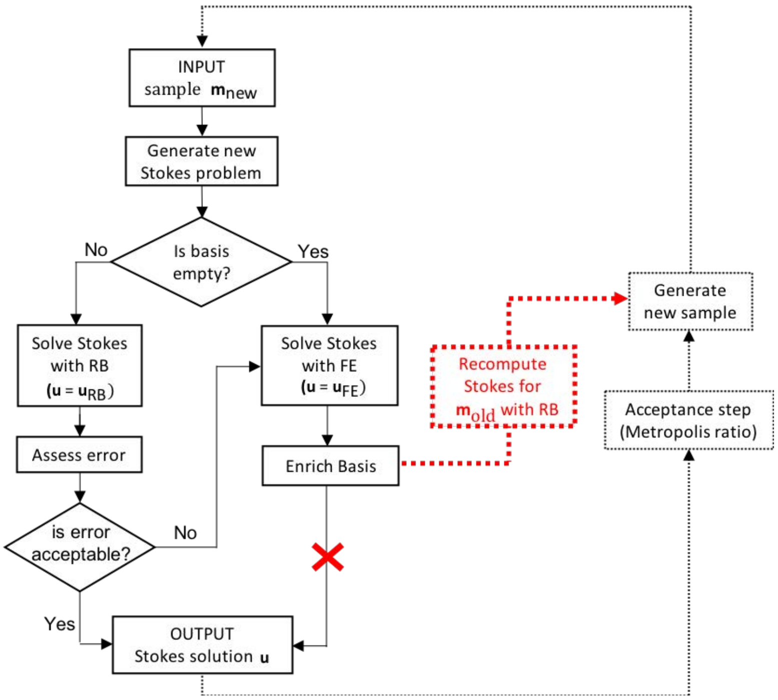

Figure 5.3: Modification of the greedy approach to build a reduced basis within a probabilistic inversion scheme to avoid an unfair RB-FE comparison in the Metropolis ratio.

created offline and used unmodified during the inversion (e.g. Galbally et al., 2009; Lieberman et al., 2010; Wang and Zabaras, 2005). Having a precomputed basis guarantees that the comparison will be always between RB approximations. However, it is at the expense of a larger and more general basis which does not take advantage of the convergence feature of the MCMC inversion to shape it to the specific problem at hand.

5.1.2 MCMC+RB application to a 3D geophysical problem

The goal of this section is to provide a simple example of the coupling between the MCMC method and the simultaneous procedure of building the reduced basis. The inversion example from Section 4.3.4 is repeated by implementing the procedure described above in order to create a tailored basis that allows us to reduce the number of high-fidelity solutions required during the inversion. The process of the basis enrichment requires assessing the error of the RB solution with respect to the FE one, so that it is enriched when the error is below a predefined tolerance. This strategy is independent of how this error is assessed. To start with and to keep the coupling simple, we will measure the error in the energy norm by using the residual as in equation (3.36) which will give us a global error of the whole RB solution. In the next sections we will introduce goal-oriented error estimators. The chosen tolerance in this problem is $e_{\mathrm{RB}} = 10^{-2}$, or in other words, RB solutions are allowed to have an error in the energy norm up to 1%.

The evolution of the basis size as function of the inversion steps is shown in Figure 5.4 (left axis). The red dashed line represents the cost of an inversion using FE only (Section 4.3.4). In the same figure we include the trace plot or random walk for a representative parameter (right axis), showing rapid convergence and stationarity of the chain.

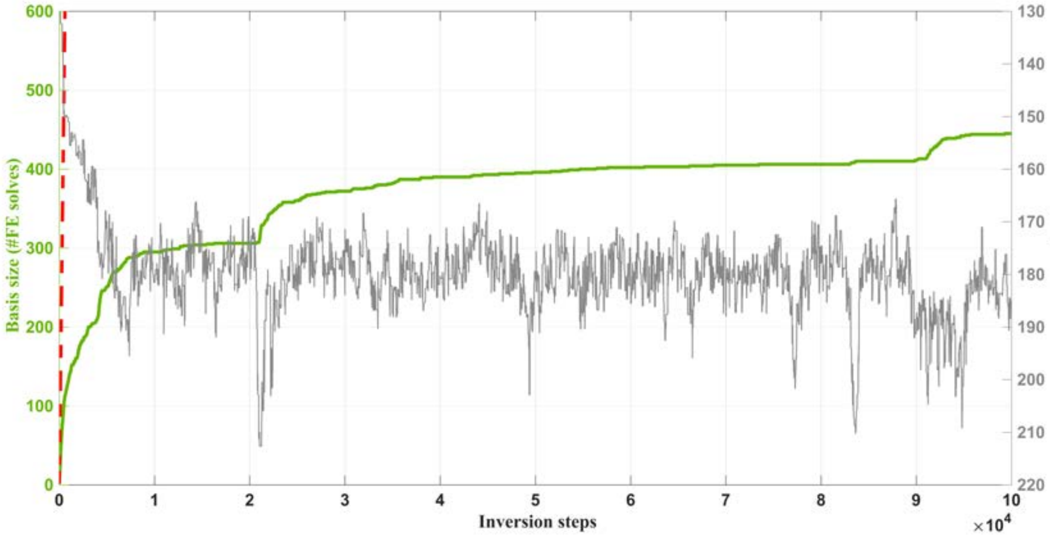

Figure 5.4: Differences between the number of full FE problems solved during the inversion using the RB approach (green solid line) and using standard FE (red dashed line) are shown on the left axis. The right axis shows the random walk for a parameter. The flattening of the basis size curve shows that most inversion steps are being solved without the need to solve the full FE problem. Once the basis is rich enough to represent the space tested by the inverse solver, its size stops increasing as any solution is properly addressed by the RB and almost no new full FE solver are required.

During the initial steps of the inversion, the high-probability regions of the parameter space are poorly known and large changes in the parameter space exploration are expected. This, together with the fact that the basis is very small, makes that the initial number of FE solutions required is relatively large. As the random walk converges, the size of the basis stabilizes. It is interesting to note the two moments in which the basis behaviour slightly deviates from its tendency to flatten, at around steps 2×10^4 and 9×10^4. Looking at the same instants on the right axis, we can see they coincide with moments when the random walk is moving to new regions that are probably not yet explained by the basis and, therefore, FE solutions are required. However, the basis behaviour is not only determined by the exploration of the space of a single parameter but by the simultaneous combination of all them. The behavior shown in Figure 5.4 illustrates the key benefit of the coupled MCMC + RB scheme; namely, the flattening of the basis size curve means that most inversion steps use RB solutions rather than full FE solutions. Once the basis is rich enough to approximate well forward problems within the high-probability space, its size stops increasing. Strictly speaking, the parameter space is not “fully” explored until all combinations of model parameters have been visited by the random walk. This is obviously unattainable and it is the main reason for MCMC algorithms. Therefore, although small, there will always exist the possibility of sporadic enrichment.

For the present example only 445 full FE solutions are required to perform 10^5 inversion steps, this is a 0.4% of the total number of forward evaluations, making the RB approach extremely efficient. Moreover, the longer the MCMC simulation, the more favorable/efficient the RB strategy becomes, since the fraction of RB solutions will become smaller. With the available computational resources and for this particular problem, a FE solution takes around 6s whereas the time for a RB solution with 500 basis elements is on average 0.12s (see Section 5.5 for details). This constitutes a staggering gain in efficiency, since the CPU time of using the MCMC+RB strategy is only 2.4% of the time required for the full MCMC+FE option (i.e., one FE solution per MCMC step). Although the metric in Figure 5.4 is not actually comparing CPU time, and it may give the impression that the RB solution is obtained at zero cost, the small RB times show that the real cost is little significant compared to the FE and, therefore, the metric used can be trusted. In addition, not only the MCMC+RB scheme implies a gain in efficiency, but it also opens the possibility of much longer simulations resulting in more reliable statistics for the inversion parameters.

Regarding the result of the inversion, it is important to validate whether the random walks and posterior PDFs using a MCMC+RB strategy coincide with those generated with a MCMC+FE. Figure 5.5 shows the results for both strategies.

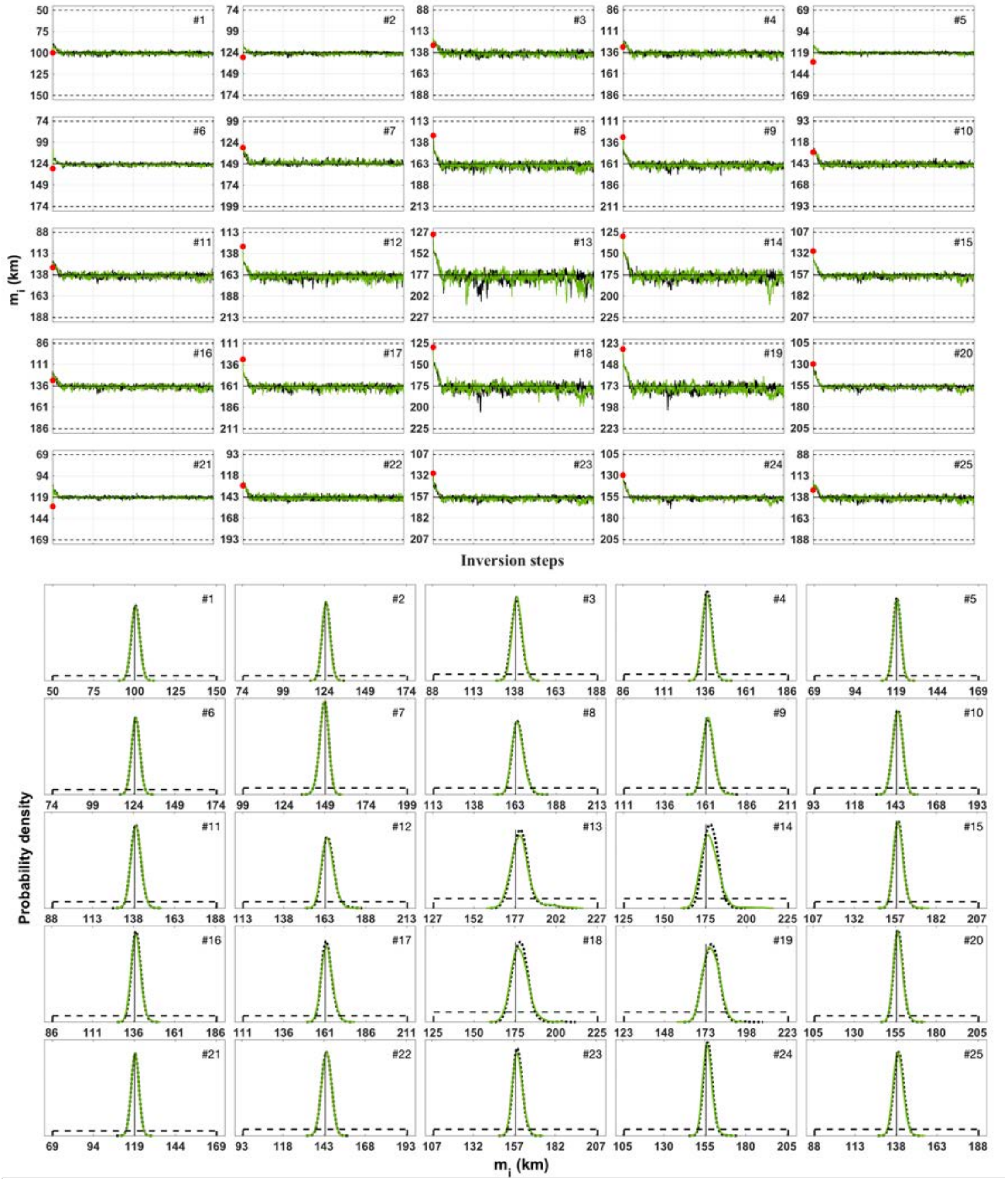

Figure 5.5: Random walk (top) and posterior PDF (bottom) for each of the 25 parameters. The results using the RB approach (green) are plotted together with the ones previously obtained in Figure 4.5 using only FE (black). In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot). The burn-in period has been discarded in the estimation of the posterior PDF. Values of the mean and the standard deviation for each posterior PDF are found in Table B.1 from Appendix B.

It can be seen convergence to the true value for the parameters is achieved and there is an excellent agreement between strategies. Values of the mean and standard deviation of each of the 25 posterior PDFs can be found in Table B.1 from Appendix B. To allow their comparison and due to the expensiveness of the MCMC+FE strategy, both simulations run for only 10^5 inversion steps. However, longer simulations would definitely improve the small discrepancies among the posterior distributions.

5.1.3 Error of the RB approximation

The solutions sought in the reduced space introduce an additional error to the approximation of each forward problem. Consequently, the random walk does not sample the "true" posterior distribution (here defined as the full order model), but an approximated one induced by the reduced space. How similar are both posteriors depends on the accuracy of the RB approximations and, ultimately, on the error tolerance set for the construction of the basis. This results in a trade-off between accuracy and computational cost and the goal is to create low-cost surrogates that do not compromise the inversion results.

To test to which extent the posterior distribution is affected by the RB error, we run four identical MCMC inversions: three of them using a MCMC+RB approach with different error tolerances $e_{\text{RB}} = 10^{-1}$, 10^{-2} and 10^{-3} and one using full FE solutions at every step of the inversion. To be consistent in the comparison we limited all simulations to 10^5 inversion steps. Figure 5.6 compares the results from these four strategies for a representative model parameter (all other parameters behave in a similar way, and therefore, are not shown here).

Results demonstrate that all MCMC simulations are in excellent agreement. The MCMC+RB strategy is not significantly affected by the RB error (up to these tolerances) and performs properly as a surrogate of a FE forward model both in terms of accuracy and, particularly, in efficiency. It can also be seen that for $e_{\text{RB}} = 10^{-1}$, the random walk took longer to converge to the true value. Since the tolerance e_{RB} is larger, the additional error in the RB approximation makes the search of the high-probability region more difficult. However, once it is found, the behaviour among all simulations is exactly the same. In addition, the initial differences correspond to the burn-in period which is discarded for the estimation of the posterior PDF. Values of the mean and standard deviation of the 25 posterior PDFs for each case are found

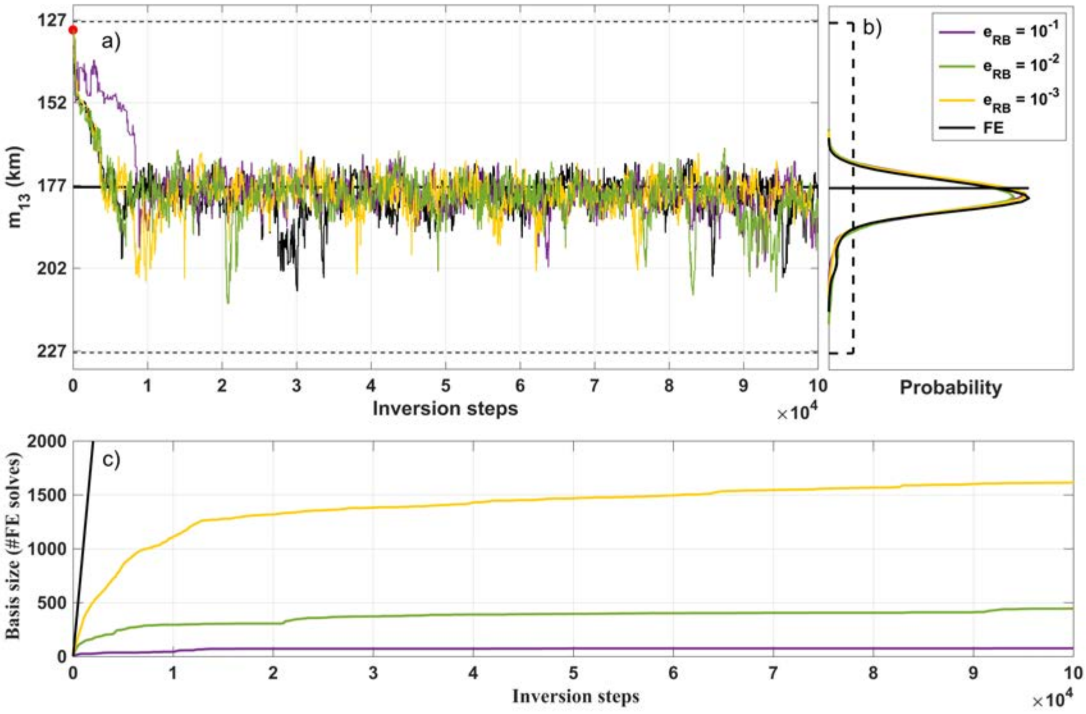

Figure 5.6: Influence of the error of the RB approximation on the (a) random walk, (b) posterior PDF and (c) basis size. The same simulation is performed with $e_{\text{RB}} = 10^{-1}$ (purple), $e_{\text{RB}} = 10^{-2}$ (green), $e_{\text{RB}} = 10^{-3}$ yellow and one using only FE (black). Plots (a) and (b) show one representative parameter. All the others behave in the same way. Values of the mean and standard deviation of each of the 25 posterior PDFs are shown in Table B.1 from Appendix B.

in Table B.1 from Appendix B. The basis size strongly depends on the tolerance assigned to the RB error indicator, e_{RB}. In the examples shown here, the basis sizes are around 1600, 445 and 76 for errors of 10^{-3}, 10^{-2} and 10^{-1}, respectively. According to these values and with the idea of keeping the basis as small as possible (within the limits of the observables error, detailed below), we decided to use $e_{\text{RB}} = 10^{-2}$ for the remaining examples. Some tests performed with larger tolerance values showed a significant posterior bias with respect to the FE case. In those tests, not only the RB error ($e_{\text{RB}} > 10^{-1}$) exceeded the noise level introduced in the data ($\sigma_d = 10^{-1}$), but also the large errors may have prevented the RB approximations from explaining most of the features of the FE solution.

The error introduced by the RB approximation is considered in the inversion in the form of modelling/theoretical uncertainties as described in Section 4.1.2. In most of the cases they are negligible compared to the uncertainties in the observations. In our case, as a first-order correction to indirectly account for them, we decided to always keep the error of the RB approximation one order of magnitude smaller than the uncertainty in the data. We acknowledge that other strategies to account for modelling uncertainties in Bayesian inversions could also be considered (see Linde et al. (2017) for a review), for instance, via a full covariance matrix defining model errors (see also a discussion in Afonso et al. (2013a,b)). Another option could be to assign priors to these errors and let them be modelled as part of a hierarchical Bayesian inversion (Titus et al., 2017; Malinverno and Briggs, 2004).

Another interesting way to show agreement between MCMC simulations is by plotting the evolution of the misfit $e(\mathbf{m})$ between predicted and observed data during the inversion. Figure 5.7 shows the misfit for each of the cases described above. Note that the initial misfit for all of them is around 2300 but we made zoom in the y-axis for a better visualization.

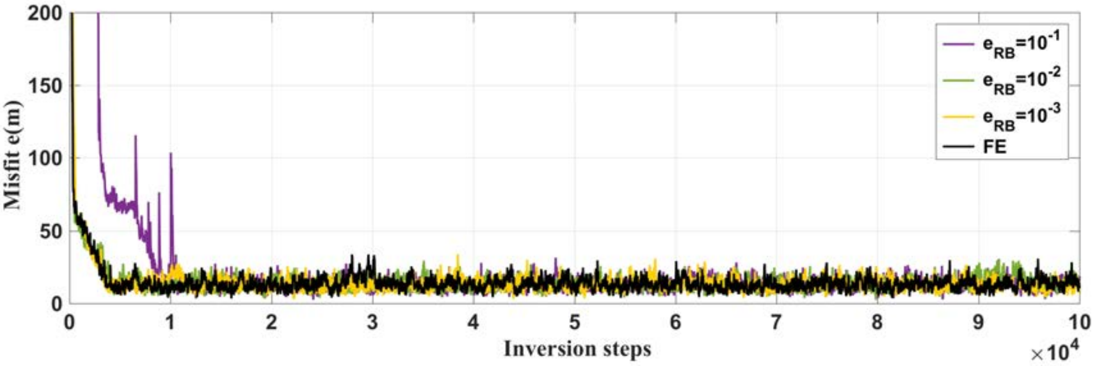

Figure 5.7: Evolution of the misfit $e(\mathbf{m})$ during the inversion for four different cases: MCMC+RB with $e_{\text{RB}} = 10^{-1}$ (purple), $e_{\text{RB}} = 10^{-2}$ (green), $e_{\text{RB}} = 10^{-3}$ yellow and finally MCMC+FE (black).

After a burn-in period, the misfit for all inversions successfully decreased and remained oscillating around the same value. Again, a slightly different behaviour is observed initially for the case with larger error in the RB approximation, $e_{\text{RB}} = 10^{-1}$. Its misfit needs more inversion steps to decrease to the same level as the other inversions. However, once there, the RB error does not seem to play any other role and all misfits behave similarly.

5.2 Influence of the error indicator on the order reduction

In the previous section the coupling between the RB method and the MCMC strategy was described. The procedure used to build the reduced basis requires an error indicator of the RB approximation with respect to the full FE solution. Although the MCMC+RB procedure does not depend on the type of error indicator, the final basis size and, hence, the computational time to obtain a RB solution does. In this section the influence of the type of error indicator on the order reduction is analysed. In particular, unlike the energy norm used in the previous section, here a goal-oriented error estimator based on a Quantity of Interest (QoI) is adopted aiming at reducing the basis size, and tailoring the basis to reproduce the relevant features only.

In the first place, the QoI is defined to obtain a goal-oriented error estimator by means of an adjoint problem and the time reduction is shown with an inversion example. Secondly, the adjoint solution accuracy is studied to see its influence in the error estimator. And, finally, an extension using a non-linear QoI is presented.

5.2.1 Goal-oriented error estimator

In order to further reduce the basis size, we want to enhance and shape the error indicator for the specific problem at hand. Taking advantage of the fact that within an inversion framework, the solution of the forward problem is used to compute an output quantity, i.e. the observables, and, therefore, our main interest is to predict accurately the observables rather than obtain a globally accurate solution. A goal-oriented error estimator is the excellent instrument for this situation, since it is able to characterise the error and drive the basis towards some particular region or quantity of interest. Since the basis is built to represent accurately such feature, its size will be smaller than if it had to be accurate everywhere.

As detailed in Section 3.2.2, we define a goal-oriented measure of the RB error, that is, we quantify the errors in some particular region of interest and for some particular QoI. The QoI summarizes the important part of the solution into one scalar number, so it is typically an integrated quantity. Considering that the observables of the inversion are the vertical velocities within the interest region Γ (see Section 4.3.2), a possible definition of QoI would be the integral of the vertical velocity within the region of interest Γ. However, for the incompressible Stokes problem, this integral vanishes identically for some configurations (e.g. a horizontal LAB). This is an undesired property, as the estimated error might be zero independently of the solution provided by the basis. Therefore, considering future (real) inversions where surface elevation and GPS velocities will be used as observables, it is more appropriate to define an error estimator that measures the gradient of the vertical velocity along the vertical direction (as this is a quantity from which dynamic topography effects can be estimated, equation (6.1)). Having this in mind, we define the QoI as,

$$Q = \int_{\Gamma} \frac{\partial u_z}{\partial z} d\Omega. \tag{5.1}$$

The QoI allows us to define the adjoint or dual problem to get a representation of the error via equation (3.41). Two ingredients are required to get an error representation: the solution of the adjoint problem and the residual of the RB solution. Since the error has to be computed at each inversion step and solving the adjoint problem requires as much computational effort as solving the direct one, we replace the exact adjoint solution with an approximation to obtain an error estimator, equation (3.44). Several approaches to estimate the adjoint solution are proposed in Section 3.2.2. The simplest one is to solve the adjoint problem once at the beginning of the MCMC simulation and reuse its solution for assessing the error in all subsequent iterations (see Section 5.2.2 for accuracy details). This strategy is based on the fact that the adjoint problem has a weaker parametric dependence than the original problem. Again, the same simulation is repeated with a goal-oriented estimator and its evolution is shown in Figure 5.8.

Figure 5.8 shows a reduction in the basis size when the goal-oriented error estimator is used instead of a measure of the error in the energy-norm. The reduced basis only needed 116 full FE solutions compared to the 445 FE required with the energy-norm. Since the basis is targeting the QoI, it does not need to provide solutions exhibiting the same accuracy everywhere in the domain. Therefore, the required number of basis elements is much smaller. It is important to remark that the use of a goal-oriented estimator does not mean that, as a consequence, the solution of the forward problem is badly approximated in the rest of the domain. In other words, the accuracy is guaranteed in the region or QoI, and elsewhere the solution is as accurate as needed to fulfill the accuracy in the QoI region.

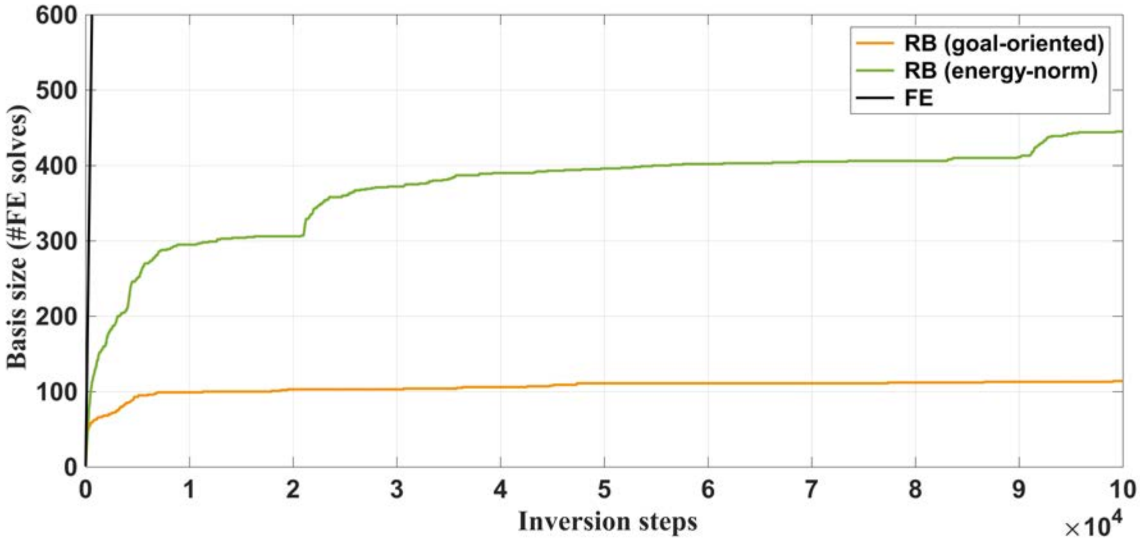

Figure 5.8: Evolution of the basis size during the inversion for three different cases: using a goal-oriented error estimator (orange), assessing the RB error in the energy-norm (green) and solving a full FE at each inversion step (black).

The posterior results of the inversion are shown in Figure 5.9. Again, convergence to the true values is achieved and random walks exhibit the same behaviour. The values of the mean and standard deviation (Table B.2 from Appendix B) indicate excellent agreement with the results obtained using any of the previous strategies: MCMC+FE or MCMC+RB with an energy-norm error. In all cases, the incorporation of the information coming from the likelihood, notoriously narrowed the prior distributions towards the true value. The gain in information on the model parameters after the inversion is significant.

The use of goal-oriented error estimators allows us to tailor the basis not only to the high-probability regions of the parameter space, but also to a specific output for the inverse problem. Consequently, the basis size is remarkably reduced. Cui et al. (2014) proposed a similar strategy using a delayed acceptance algorithm with a criteria to enrich the basis directly based on the observables (i.e. based on the scaled error of the reduce order model in the data space). When the estimated error of the reduced order model is greater than one standard deviation of the measurement noise, the RB solution is discarded.

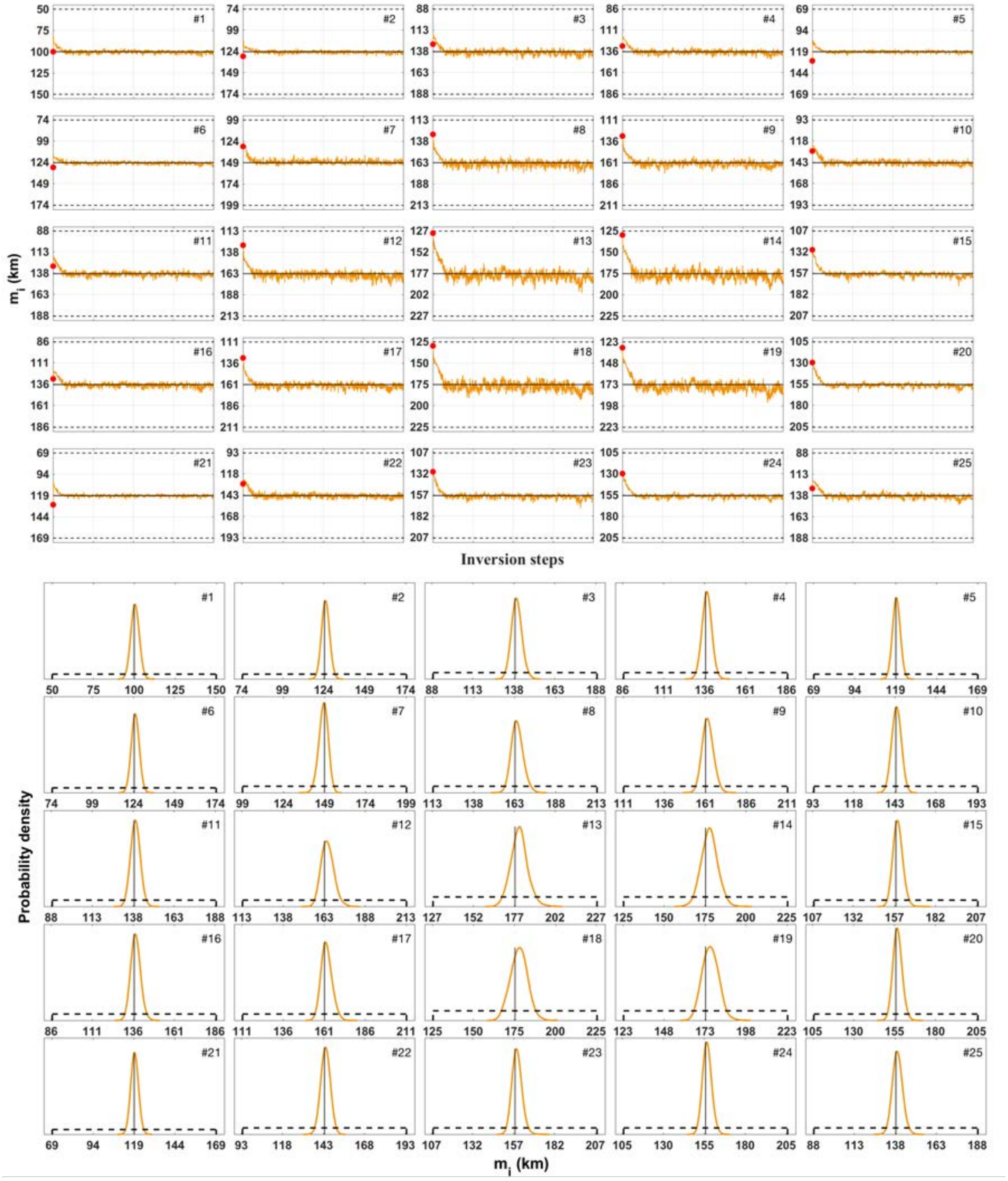

Figure 5.9: Random walk (top) and posterior PDF (bottom) for each of the 25 parameters. The results using the RB approach with a goal-oriented error estimator are plotted. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot). The burn-in period has been discarded in the estimation of the posterior PDF. Values of the mean and the standard deviation for each posterior PDF are found in Table B.1 from Appendix B.

5.2.2 Accuracy of the adjoint solution

As previously mentioned, obtaining an error representation requires both the solution of the adjoint problem, $\mathbf{v}$, and the residual of the RB approximation, $\mathbf{r}^*$. Since computing the adjoint solution at each inversion step is not feasible (same computational cost as solving the primal problem), we approximate it. In the previous example the adjoint solution is solved once at the beginning of the simulation and reused for assessing the error in the successive steps. Our goal here is to analyse how the accuracy of the adjoint solution affects the error estimator and, in particular, to which extent it affects the basis size. The more accurate is the estimator the more computationally expensive it is, so that some compromise should be sought. Some options are presented next exhibiting a variety of accuracies and costs. They can be classified in two groups:

1. The solution of the adjoint problem is approximated with a Reduced Basis. A basis $\mathbf{B}_{\text{adj}}$ is built for the adjoint problem using the same procedure that is used for the primal problem. Analogously, the reduced system for the adjoint problem is,
$$\mathbf{B}_{\text{adj}}^{\mathsf{T}}\mathbf{K}\mathbf{B}_{\text{adj}}\boldsymbol{\beta} = \mathbf{B}_{\text{adj}}^{\mathsf{T}}\mathbf{f}^{o} \tag{5.2}$$
and the RB adjoint solution is obtained as $\mathbf{v} \simeq \mathbf{v}_{\text{RB}} = \mathbf{B}_{\text{adj}}\boldsymbol{\beta}$. Creating a basis for the adjoint problem also requires an error measure and in this case we use the energy-norm.

2. The adjoint problem is solved with FE and its solution is reused for a certain number of inversion steps. This option is based on one important property of the adjoint problem: adjoints are typically less sensitive to perturbations of the parameters than the primal problem (Florentin and Díez, 2012; García-Blanco et al., 2017; Serafin et al., 2017) and the same adjoint solution can be used to estimate the error of many similar primal solutions. Due to the convergence nature of the inversion, a relevant adjoint solution may not be relevant for the problem at hand after a certain number of inversion steps during the burn-in period. As the inversion progresses, the adjoint solution is updated after a certain number of inversion steps in order to guarantee that it remains relevant. The following three options only differ in the updating frequency.
    a) The adjoint solution is recomputed after a fixed number of inversion steps.

  b) The adjoint solution is recomputed every time the primal reduced basis has been enriched a certain number of times.
  c) The adjoint problem is solved once. (Examples of previous Section 5.2.1.)

In order to check the accuracy of the goal-oriented error estimator, we compute its effectivity index, η_{eff}, defined using equations (3.41) and (3.43) as,

$$\eta_{\text{eff}} = \frac{\text{estimated error, } \hat{E}_Q}{\text{exact error, } E_Q} = \frac{|\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{r}^*|}{|l^o(\boldsymbol{u}) - l^o(\boldsymbol{u}_{\text{RB}})|}, \tag{5.3}$$

where $\hat{\mathbf{v}}$ is the approximated adjoint solution. If $\eta_{\text{eff}} > 1$, the estimator has overestimated the error of the RB solution, whereas if $\eta_{\text{eff}} < 1$ the error has been underestimated. Therefore, we are interested in an error estimator whose effectivity index is close to 1. As a reference, values of η_{eff} between 0.1 and 10 are usually taken as acceptable. If the adjoint solution was computed with FE, then $\eta_{\text{eff}} = 1$. This can be used to check that the implementation is correct. Computing the effectivity index involves the FE solution $\mathbf{u}$ and therefore it can only be computed for small test cases.

We run five inversions to study how the accuracy of the adjoint solution influences the goal-oriented error estimator of the direct RB solution and its basis size. We use the linear QoI from equation (5.1) and the error tolerance for the basis is set at $e_{\text{RB}} = 10^{-2}$ (1% of error). Details of the five strategies are summarized next:

1. Adjoint problem is approximated with RB for every inversion step. The error tolerance for the adjoint RB solution is:
    a) $e_{\text{adj}} = 10^{-2}$ (1% of error)
    b) $e_{\text{adj}} = 10^{-1}$ (10% of error)
2. Adjoint problem is solved with FE and reused for a certain number of steps
    a) Recomputed every 2000 inversion steps
    b) Recomputed every time the primal basis is enriched 10 times.
    c) Solved once

Since we are more interested in the magnitude of the effectivity index than the value itself, we define a new effectivity index as the logarithm, $\widetilde{\eta} = \log_{10}(\eta_{\text{eff}})$. In doing so, if $\widetilde{\eta} > 0$ the error has been overestimated, if $\widetilde{\eta} < 0$ it is underestimated and when $\widetilde{\eta} = 0$ the error estimator is exact. Acceptable values for $\widetilde{\eta}$ are between -1 and 1. From now on, we refer to $\widetilde{\eta}$ when talking about the effectivity index. Figure 5.10 shows the evolution of the effectivity index of the error estimator of the direct solution, the error of the adjoint solution and the size of the basis for the strategies 1a and 1b. Results for strategies 2a, 2b and 2c are shown in Figure 5.11.

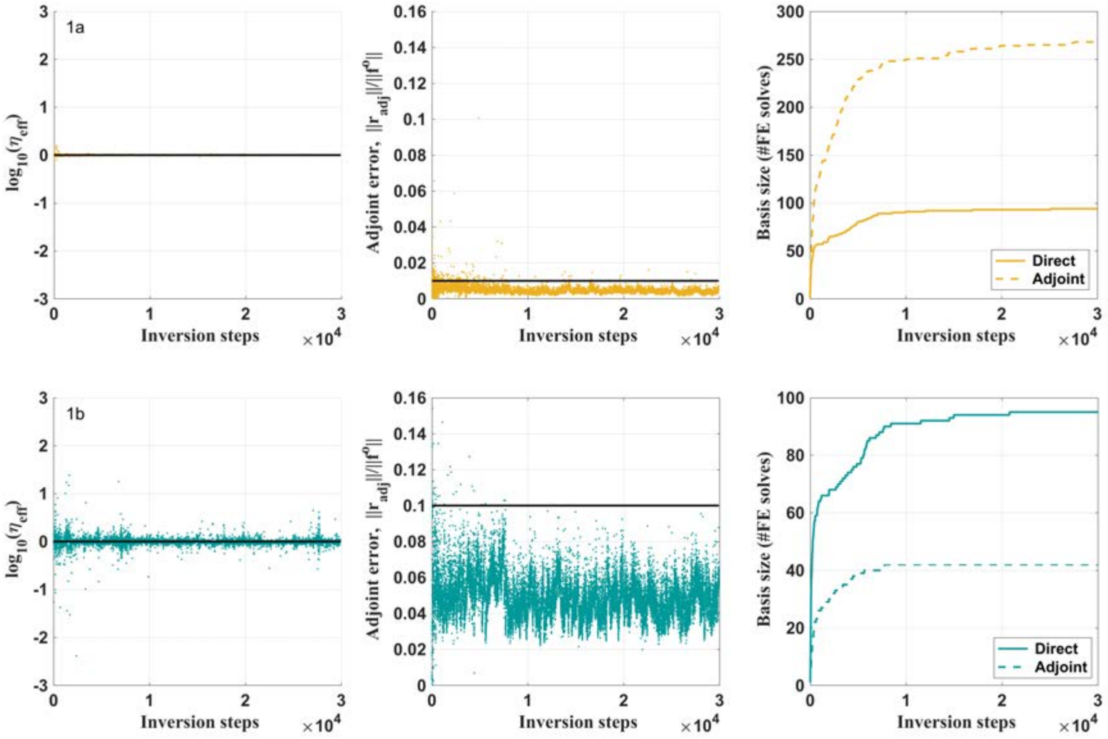

Figure 5.10: Evolution of the logarithm of effectivity index of the error estimator of the direct solution, error of the adjoint solution and basis size during 3×10^4 inversion steps. In both cases the adjoint solution is approximated with a RB: $e_{\text{adj}} = 10^{-2}$ (**1a**, top); and $e_{\text{adj}} = 10^{-1}$ (**1b**, bottom).

In both strategies from Figure 5.10 the adjoint solution is approximated for every inversion step. The error of the adjoint is always kept below the tolerance and when it is larger, the problem is solved with FE and the basis is enriched. Looking at the effectivity index plot it is clear that for the case where the adjoint solution is more accurate (1a) the effectivity index is closer to 0, i.e. the estimated error is closer to the exact error. However, the last plot shows there is no relevant difference in the basis size of the direct problem (94 and 95 basis elements for cases 1a and 1b, respectively). Contrarily, the basis size of the adjoint problem is considerably different and this is what determines the actual cost of computing the error estimator. The basis size of the adjoint problem for case 1a is 268, whereas for case 1b is only 42. Moreover, in case 1a, although both direct and adjoint solutions have the same error tolerance ($e_{\text{RB}} = e_{\text{adj}} = 10^{-2}$) the final adjoint basis is larger than the direct one. The reason is that the error for the adjoint is measured in the energy-norm (global) whereas the error for the direct is goal-oriented (local).

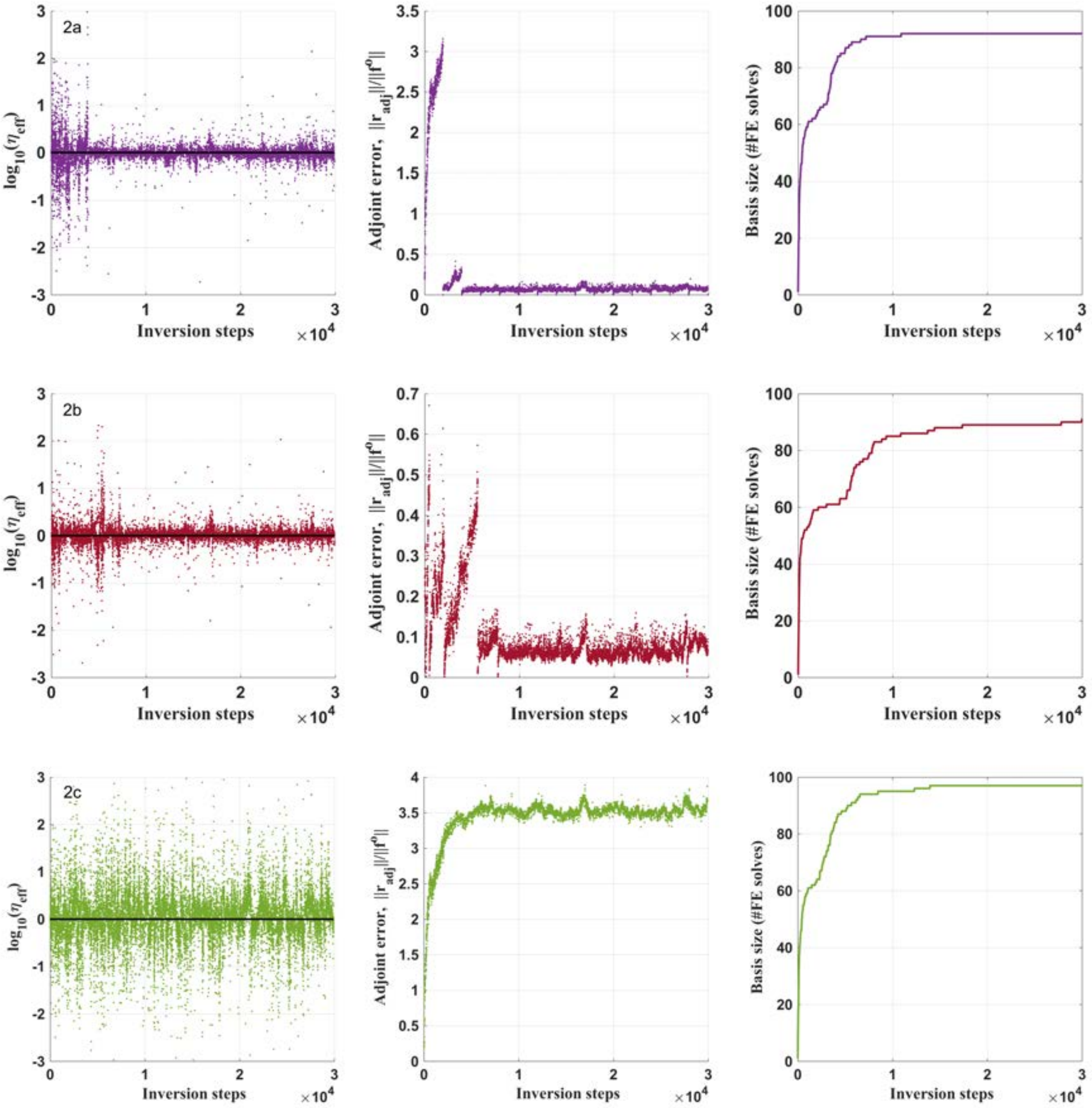

Figure 5.11: Evolution of the logarithm of the effectivity index of the error estimator of the direct solution, error of the adjoint solution and basis size during 3×10^4 inversion steps. Adjoint solution: recomputed after 2000 inversion steps (**2a**, top), recomputed every time the primal RB is enriched 10 times (**2b**, center) and only computed once (**2c**, bottom).

The three options reusing the adjoint solution are analysed next and their results presented in Figure 5.11. In this case, the behaviour of the adjoint error is slightly different. The error is initially zero and it grows while the random walks evolve from the initial point until the adjoint is recomputed. The error then decreases to zero and starts growing again. Since the adjoint solutions are reused for successive problems, their errors can be large. In case 2c, although there is only one solution to be reused, the error stabilizes as the random walk finds the high probability regions of the parameter space. The greedy approach from case 2b uses information from the basis of the direct problem to know when to recompute the adjoint solution. Therefore, they mostly occur at the beginning of the inversion, when the basis growth is faster, until they, eventually, are no longer necessary. Instead, case 2a recomputes the adjoint solution indistinctly every 2000 steps, which results in a computational cost that increases with the length of the inversion. Regarding the effectivity index, the three cases show values around 0. The wider dispersion of values exhibited by case 2c coincides with the also larger adjoint errors. Finally, the sizes of the reduced basis of the direct problem coincide with previous cases using a RB, all between 91 and 97 elements. Values of the mean $\widetilde{m}$ and standard deviation $\widetilde{\sigma}$ of the effectivity index as well as basis sizes and number of adjoint FE solutions for each of the five cases are summarized in Table 5.1.

Case	Basis size (#FE)	Adjoint (#FE)	$\widetilde{m}$	$\widetilde{\sigma}$
1a	94	268	0.0001	0.004
1b	95	42	-0.0001	0.06
2a	92	16	-0.02	0.20
2b	91	10	0.01	0.16
2c	97	1	0.05	0.51

Table 5.1: Results with a linear error estimator. Basis size of the direct problem, number of full FE adjoint solutions, mean $\widetilde{m}$ and standard deviation $\widetilde{\sigma}$ of the effectivity index for the five cases. Adjoint solution: RB with $e_{\text{adj}} = 10^{-2}$ (**1a**), RB with $e_{\text{adj}} = 10^{-1}$ (**1b**), recomputed every 2000 inversion steps (**2a**), recomputed every time primal RB is enriched 10 times (**2b**) and only computed once (**2c**).

We try next to quantify the behaviour of the different approaches to compute the adjoint solutions and the corresponding error estimators. The error estimator can predict the exact error, overestimate it or underestimate it. Although overestimation is not the ideal situation, underestimation is more problematic. It makes us believe a solution has a certain accuracy when it actually has not. In the effectivity index plots from Figures 5.10 and 5.11, all values below the horizontal black line have been underestimated to some extent. It is interesting to show the extent of the underestimation, i.e. a measure of the dispersion of underestimated errors with respect to 0. Having N^- underestimated values we define,

$$\widetilde{m}^- = \frac{1}{N^-} \sum_{i \in \{I^-\}} \widetilde{\eta}_i \qquad \text{and} \qquad \widetilde{s}^{-2} = \frac{1}{N^-} \sum_{i \in \{I^-\}} (\widetilde{\eta}_i - 0)^2 \tag{5.4}$$

Note that if dispersion was computed with respect to $\widetilde{m}^-$, it would be a variance. Equivalently, for the N^+ overestimated values we can define $\widetilde{m}^+$ and $\widetilde{s}^{+2}$. These values are computed at the beginning of the inversion (first 2000 values) and at the end (last 2000 values) and results are summarized in Table 5.2.

Case	$\widetilde{m}^+$		$\widetilde{s}^+$		$\widetilde{m}^-$		$\widetilde{s}^-$		Critical underestimation
	first	last	first	last	first	last	first	last	
1a	0.003	0.001	0.01	0.001	-0.002	-0.001	0.01	0.002	0.01 %
1b	0.05	0.02	0.10	0.05	-0.05	-0.02	0.12	0.04	0.03 %
2a	0.20	0.07	0.36	0.12	-0.27	-0.07	0.45	0.12	0.08 %
2b	0.12	0.06	0.22	0.10	-0.16	-0.06	0.28	0.10	0.09 %
2c	0.36	0.30	0.52	0.43	-0.38	-0.27	0.56	0.42	0.11 %

Table 5.2: Results with a linear error estimator. Values of the parameters characterizing the overestimation ($\widetilde{m}^+$, $\widetilde{s}^+$) and the underestimation ($\widetilde{m}^-$, $\widetilde{s}^-$) for the 2000 first and last values for the five cases. Also, percentage of the critical underestimations. Adjoint solution: RB with $e_{\text{adj}} = 10^{-2}$ (**1a**), RB with $e_{\text{adj}} = 10^{-1}$ (**1b**), recomputed every 2000 inversion steps (**2a**), recomputed every time primal RB is enriched 10 times (**2b**) and only computed once (**2c**).

As expected, in terms of both overestimation and underestimation the error estimator that performs the best is 1a, followed by 1b, 2b, 2a and finally 2c. Moreover, in all cases there is a clear improvement as the inversion progresses: mean values get closer to 0 and dispersion values get smaller. At the beginning (first 2000 values) underestimation seems to have a slightly larger effect than overestimation, but this behaviour disappears, and in some cases reverses, as the inversion continues (last 2000 values).

In general, previous tables and figures may give the impression the error is underestimated many times. However, in the context of building a reduced basis, the ones that do really present a problem are those whose error estimator indicates the error is below the given tolerance, but the real error is not. We call this situation critical underestimation. In critical underestimations, if we had known the exact error, the RB solution would not have been considered accurate, a FE solution would have been required and the basis would have been enriched. Instead, the solution is wrongly considered accurate and some error is likely to be introduced in the likelihood. This situation is rare. For the less accurate error estimator (2c), the percentage of critical underestimations (with respect to the total number of estimations) is 0.11%, whereas for the more accurate case (1a), the percentage is 0.01%. The rest of underestimated errors do not present a problem since they are within the tolerance threshold, i.e. allowed error.

Since all direct basis have similar sizes, the computational cost is basically determined by the number of adjoint solutions that must be computed with FE during the inversion. Looking at Table 5.2, the error estimator that shows an optimal cost-accuracy relationship is 2b. It performs a greedy update of the adjoint solution while keeping the cost of each inversion step controlled. Regarding the accuracy of the error estimator, although we may be interested in achieving the highest accuracy (case 1a), the error estimator is only used to determine if the basis needs to be enriched. Since the percentage of underestimated errors that skip the enrichment is very small (even smaller for longer simulations) and it mostly occurs at the beginning of the inversion (discarded as burn-in period), the corresponding “wrong” samples that appear in the random walk become statistically insignificant. In addition, in all the options presented here, no differences are observed in terms of MCMC results. Or in other words, means and standard deviations of the posterior PDFs coincide among the five cases and with the MCMC+FE reference approach. This is a clear indicator that, in all tested cases, the RB approximation errors are under control and do not significantly bias the posterior approximation while allowing to reduce its computational cost.

5.2.3 Extension to a non-linear QoI

Within our inversion framework, the goal of the QoI is to characterise the relevant part of the solution which is to be used later in the likelihood estimation. That is, the observables. In the misfit expression of the likelihood (4.31), the observable terms are not linear, i.e. they appear to the second power. With the idea of correlating the misfit and QoI as much as possible, another interesting option is to define a non-linear

QoI as,

$$Q = l^o(\boldsymbol{u}) = \|\mathbf{obs}(\mathbf{u})\|^2 \tag{5.5}$$

with $\mathbf{d} = \mathbf{obs}(\mathbf{u})$, where $\mathbf{obs}$ is an operator that computes the predicted observables $\mathbf{d}$ from the velocity field $\mathbf{u}$. In doing so, although the observables are the vertical velocities, the QoI does not vanish anymore. We can follow the same strategy than in Section 3.2.2 to obtain a linearlised error representation as,

$$l^o_{\mathrm{L}}(\boldsymbol{e}) \simeq l^o(\boldsymbol{u}_{\mathrm{RB}} + \boldsymbol{e}) - l^o(\boldsymbol{u}_{\mathrm{RB}}) = 2\ \mathbf{obs}(\mathbf{u}_{\mathrm{RB}})^{\mathsf{T}}\mathbf{obs}(\mathbf{e}) = \mathbf{f}^{o\mathsf{T}}\mathbf{e} \tag{5.6}$$

with $\mathbf{f}^o$ being the discrete version of $l^o_{\mathrm{L}}(\cdot)$. Since $\mathbf{f}^o$ depends on $\mathbf{u}_{\mathrm{RB}}$, the source vector of the adjoint problem will change at each inversion step and the new problem will have to be solved to assess the error. Again, this is not feasible. As the inversion proceeds, the predicted observables are closer to $\mathbf{d}^{\mathrm{obs}}$ and the approximation $\mathbf{obs}(\mathbf{u}_{\mathrm{RB}}) \simeq \mathbf{d}^{\mathrm{obs}}$ can be used,

$$l^o_{\mathrm{L}}(\boldsymbol{e}) \simeq 2\ \mathbf{d}^{\mathsf{T}}_{\mathrm{obs}}\mathbf{obs}(\mathbf{e}) = \mathbf{f}^{o\mathsf{T}}\mathbf{e}. \tag{5.7}$$

Once $\mathbf{f}^o$ is defined and analogously to the linear case, an adjoint problem is used to obtain an error representation in terms of the solution of the adjoint problem and the residual of the RB solution. Similar to the linear example above, the adjoint problem is solved once at the beginning of the MCMC simulation and its solution is reused for assessing the error in all subsequent iterations.

We repeated the inversion from Section 5.2.1 using the non-linear QoI for 10^6 inversion steps. The results of the posterior PDFs show excellent agreement with previous cases. Specific values for means and standard deviations of each of the posterior PDFs are shown in Table B.2 from Appendix B. The final size of the basis is of 129 elements, which is just slightly larger than the 116 of the linear case.

Regarding the accuracy of the non-linear error estimator, we implemented three of the five cases presented above: (1b) adjoint problem approximated with RB with $e_{\mathrm{adj}} = 10^{-1}$, (2b) adjoint problem recomputed every time the primal basis is enriched 10 times and (2c) adjoint solution is computed once. We define the effectivity index $\eta^{\mathrm{nl}}_{\mathrm{eff}}$ as,

$$\eta^{\mathrm{nl}}_{\mathrm{eff}} = \frac{\text{estimated error, } \hat{E}_Q}{\text{exact error, } E_Q} = \frac{|\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{r}^*|}{|l^o_{\mathrm{L}}(\boldsymbol{u} - \boldsymbol{u}_{\mathrm{RB}})|}. \tag{5.8}$$

Again, we are more interested in the magnitude than the value itself, so that we use its logarithm, $\widetilde{\eta}^{\mathrm{nl}} = \log_{10}(\eta^{\mathrm{nl}}_{\mathrm{eff}})$. In the definition from (5.8) the linearised error is considered as the exact one, so that the linearization is already assumed a valid and accurate approximation. Therefore, the effectivity index is only accounting for the error introduced by the approximation of the adjoint solution $\hat{\mathbf{v}}$ and the approximation $\mathbf{obs}(\mathbf{u}_{\mathrm{RB}}) \simeq \mathbf{d}^{\mathrm{obs}}$. Since the latter is only accurate when the parameters model $\mathbf{m}$ is close to the real one, we expect that the beginning of the inversion the non-linear error estimator estimates the error very poorly, whereas as the inversion converges and model parameters get closer to the true ones, the approximation $\mathbf{obs}(\mathbf{u}_{\mathrm{RB}}) \simeq \mathbf{d}^{\mathrm{obs}}$ applies and the estimation improves. Summarizing, if we want the non-linear error estimator to be exact ($\widetilde{\eta}^{\mathrm{nl}} = 0$), we need two things: an exact adjoint solution and that $\mathbf{m}$ is the real one.

Figure 5.12 shows the evolution of the effectivity index of the nonlinear error estimator of the direct solution, the error of the adjoint solution and the size of the basis for the strategies 1b, 2b and 2c during 3×10^4 inversion steps. Values of the mean $\widetilde{m}$ and standard deviation $\widetilde{\sigma}$ of the effectivity index as well as basis size and number of adjoint FE solutions are summarized in Table 5.3. Finally, parameters characterising overestimation and underestimation as well as the percentage of critic underestimations are shown in Table 5.4.

Case	Basis size (#FE)	Adjoint (#FE)	$\widetilde{m}$	$\widetilde{\sigma}$
1b	95	148	-0.04	0.27
2b	105	11	-0.10	0.45
2c	123	1	-0.06	0.63

Table 5.3: Results with a non-linear error estimator. Basis size of the direct problem, number of full FE adjoint solutions, mean $\widetilde{m}$ and standard deviation $\widetilde{\sigma}$ of the effectivity index for the three cases. Adjoint solution: RB with $e_{\mathrm{adj}} = 10^{-1}$ (**1b**), recomputed every time primal RB is enriched 10 times (**2b**) and only computed once (**2c**).

Looking at the effectivity index plots from Figure 5.12, it can be seen that all values are around 0. However, their dispersion is larger than in the previous linear cases and it is confirmed by the standard deviation values from Table 5.3. Although plots do not seem to show a clear improvement in the effectivity index as the inversion progresses, values of the mean and dispersion of both the overestimation and underestimation from Table 5.4 show a reduction for the 2000 last values compared to the

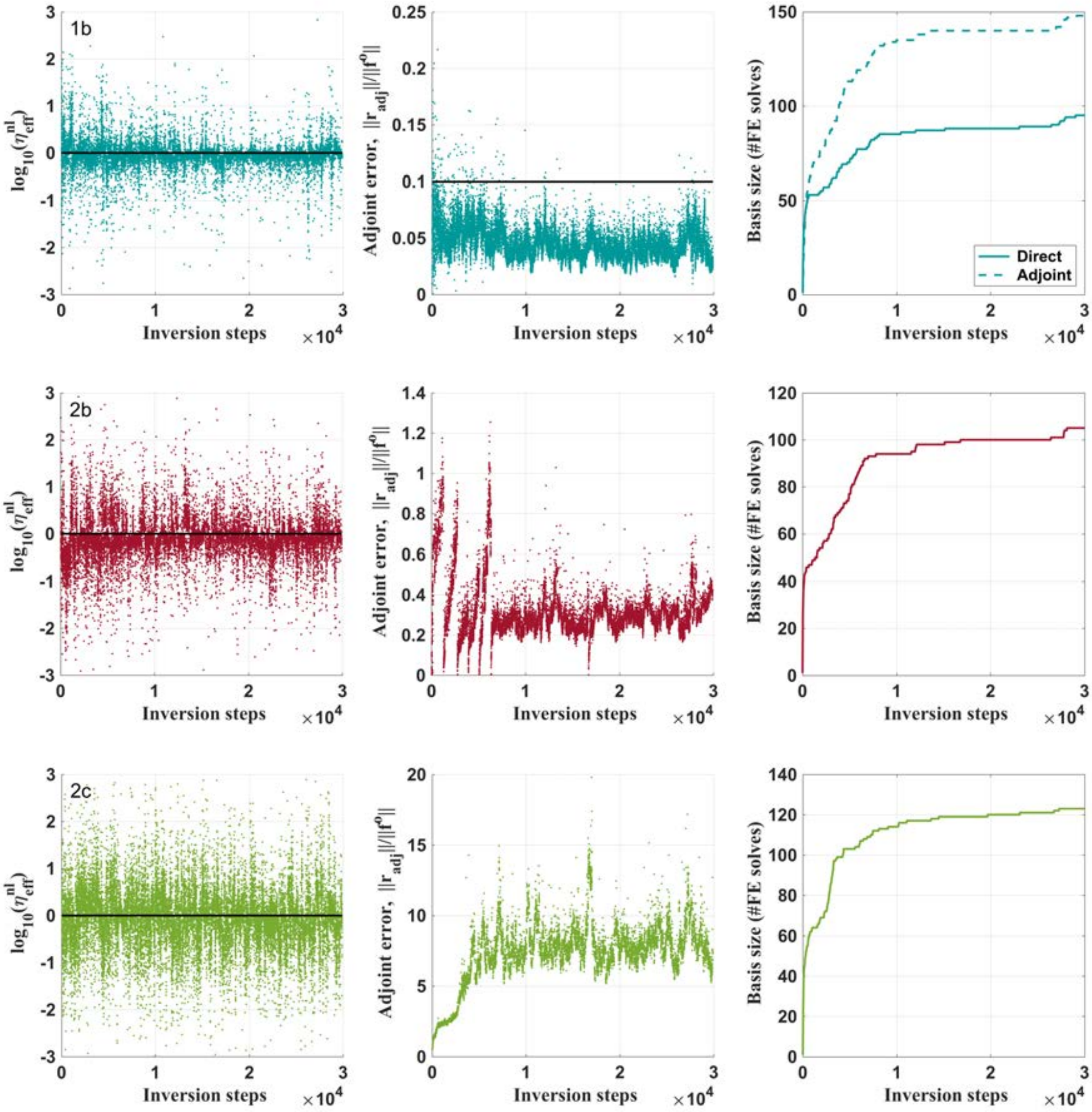

Figure 5.12: Evolution of the logarithm of the effectivity index of the error estimator of the direct solution, error of the adjoint solution and basis size during 3×10^4 inversion steps. Adjoint solution: RB with $e_{\text{adj}} = 10^{-1}$ (**1b**, top), recomputed every time the primal RB is enriched 10 times (**2b**, center) and computed once (**2c**, bottom).

first 2000. For the case 2c, the reduction is less relevant than for 1b and 2b since the same adjoint solution is reused for the whole inversion and, therefore, does not play any role in improving the error estimator. As we explained, the adjoint solution is not the only responsible term in obtaining a good error estimator. The approximation $\mathbf{obs}(\mathbf{u}_{\text{RB}}) \simeq \mathbf{d}^{\text{obs}}$ is only valid when the random walk is actually close to the true

Case	$\widetilde{m}^{+}$		$\widetilde{s}^{+}$		$\widetilde{m}^{-}$		$\widetilde{s}^{-}$		Critical underestimation
	first	last	first	last	first	last	first	last	
1b	0.26	0.11	0.43	0.25	-0.22	-0.14	0.38	0.28	0.31 %
2b	0.43	0.26	0.60	0.42	-0.49	-0.26	0.67	0.41	1.15 %
2c	0.48	0.45	0.65	0.63	-0.53	-0.47	0.71	0.64	0.64 %

Table 5.4: Results with a non-linear error estimator. Values of the parameters characterizing the overestimation ($\widetilde{m}^{+}$, $\widetilde{s}^{+}$) and the underestimation ($\widetilde{m}^{-}$, $\widetilde{s}^{-}$) for the 2000 first and last values for the five cases. Also, percentage of the critic underestimations. Adjoint solution: RB with $e_{\text{adj}} = 10^{-1}$ (**1b**), recomputed every time primal RB is enriched 10 times (**2b**) and only computed once (**2c**).

parameter model which is certainly not true at the beginning of the inversion. This explains why the effectivity index in general terms is worse than for the corresponding linear cases. It is clearly seen in the underestimation dispersion values as well as in the percentage of critic underestimated errors. All this is translated in larger reduced basis for the direct problem. In case 1b, the basis size for the adjoint problem is even larger than the basis for the direct problem which is something that did not happen for the linear case with these error tolerances. Not only this nonlinear QoI performs poorer than the linear options, but it also has a higher computational cost.

For the nonlinear case, it also makes sense to define the effectivity index as $\eta_{\text{eff}}^{\text{nl}} = |\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{r}^{*}|/|l^{o}(\boldsymbol{u} - \boldsymbol{u}_{\text{RB}})|$. In doing so, the error introduced by the linearitzation is also considered in the effectivity index. However, results are similar and we do not show them.

In all cases, the results of the inversions are in excellent agreement with the MCMC+FE and all other MCMC+RB options presented so far and do not seem to be visibly affected by the accuracy of the error estimator. However, if available, we certainly prefer a more reliable error estimator like the corresponding linear options, which, on top of that, have a lower computational cost.

5.3 Influence of the inverse problem definition on the order reduction

In previous sections we analysed how to create a reduced basis which is tailored to the high-probability regions of the parameter space and to a specific quantity of interest simultaneously. Since it is specifically shaped, its size is much smaller than if it had to reproduce the whole domain. These two strategies allowed us to reduce drastically the computational cost of solving forward problems. The goal of this section is to study how the variables defining the inverse problem affect the basis size. In particular, the number of inversion parameters and the uncertainty in the observables.

5.3.1 Number of inversion parameters

One of the main things defining the inverse problem is the number of parameters to be retrieved by the inversion. So far, in all the simulations the LAB was characterized by 25 parameters. We will now increase them and study the basis behaviour. The inversion will certainly have more difficulties in converging to the true parameters and the number of required inversion steps will be larger in order to have enough samples to approximate the posterior. Intuitively, it is expected that the size of the basis would be proportional to the size of the parameter space. This is because more parameters allow to describe a richer family of solutions and, therefore, the amount of information required to span all possible solutions grows as well.

Like in previous examples, the domain represents a portion of Earth of 400km × 400km going down to 660km depth and is discretized in 10×10×20 finite elements. The LAB structure is now discretized with 100 parameters and the number of synthetic observables is increased to 2700 to make the overall sensitivity comparable to that in the example with 25 parameters. Each LAB depth (inversion parameter) is associated with one column of the finite elements mesh, whereas previously each LAB value covered four columns. The setup of the MCMC simulation is identical to that of the previous examples: same priors and starting configuration, Gaussian proposal with $\sigma_p = 5$ km, percentage of the introduced noise $\sigma_e = 10^{-1}$ and tolerance for the RB approximation is set at $e_{\text{RB}} = 10^{-2}$. Regarding the error estimator, we used the goal-oriented approach with the linear QoI and one reference solution for the adjoint problem. We run a total of 10^6 inversion steps.

The basis needed about 262 elements (0.03% of the inversion steps required full FE solutions) to create an accurate surrogate for all simulation steps, which again represents a drastic reduction. In this case, the time to obtain a RB approximation is around 0.05 seconds which translates into a 0.8% of the overall CPU cost if using FE only. The difference between RB times with respect to the 25 parameter case (0.1 seconds for a basis of around 250 elements) is due to the parameter-column relationship since the number of columns modified per inversion step is here smaller (see Section 5.5 for a complete explanation). Figure 5.13 summarizes the posterior PDF for the LAB structure as the mean of each of the parameters.

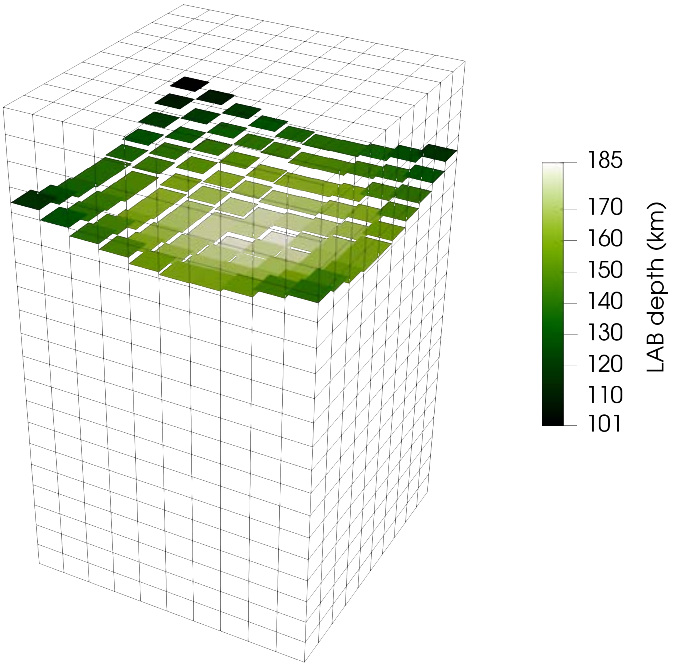

Figure 5.13: Result of the inverse problem. Mean values of the posterior PDFs for each of the 100 parameters of the unknown LAB structure.

The algorithm succeeded in obtaining an accurate representation of the solution. Differences between mean values of the posterior PDF and the reference synthetic LAB are again within one standard deviation. Plots of the random walks (Figure B.1) and posteriors PDFs (Figure B.2) for each of the 100 parameters and values of means and standard deviations (Table B.3) can be found in Appendix B. The standard deviations of the posteriors range approximately between 2km and 6km.

These values are very close to the ones obtained in the previous examples. This is to be expected, since we increased the number of observables to make the overall sensitivity comparable in both examples. If the number of data points is kept fixed, the posterior distributions become wider as the number of parameters increase. This suggests that the sensitivity of the velocity field to perturbations of a single parameter is reduced as the number of parameters increases. Or in other words, when the number of parameters increases, their individual influence area is reduced. This behaviour is well-known in inverse problems, where a trade-off always exists between the number of parameters needed to represent a model versus the number that can be well-constrained by the data. Less sensitivity to model parameters requires exploring larger areas of the parameter space, which typically results in longer chains with wider posteriors.

Figure 5.14 shows the evolution of the basis size for different number of parameters: 25, 100 and 225. The number of synthetic observables is increased proportionally so that sensitivity is not a controlling factor: 675, 2700 and 6075 data points, respectively. For the 225 parameters case the same domain is discretized in 15×15×20 finite elements, whereas for cases with 25 and 100 parameters it remains 10×10×20.

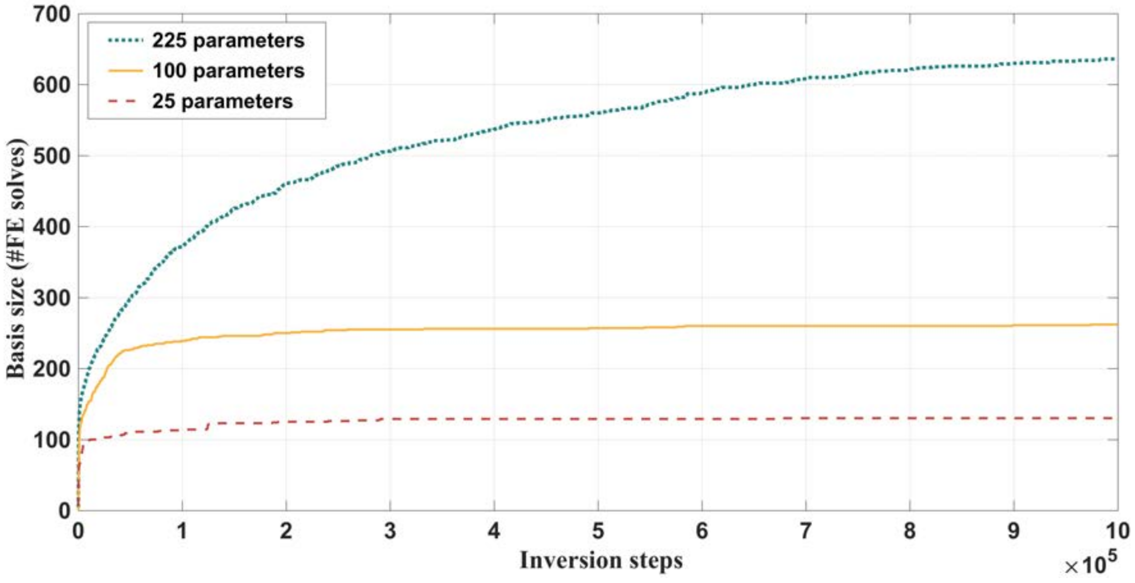

Figure 5.14: Evolution of the basis size during the inversion for three different number of model parameters: 25 (dashed red), 100 (solid yellow) and 225 (turquoise dotted).

In all cases, the growth of the basis (i.e. the number of full FE problems) reaches a “plateau” or a “saturation” level after which all MCMC steps are computed with the RB approximation. As expected, the final basis size increases with the number of parameters. To solve 10^6 inversion steps the inversion needed to solve 138, 262 and 636 full FE solutions for cases with 25, 100 and 225 parameters, respectively. The reduction in computational cost with respect to the unfeasible MCMC+FE case is outstanding. This example illustrates the scalability of the method. Although the size of the basis grows with the number of parameters, the method remains practical.

5.3.2 Uncertainty in the observables

An inverse problem uses data to obtain information about the model parameters that are more likely to explain such data. Definitely, the larger the uncertainty in the data, the less constrained the parameter space will be. Since the basis size is mainly determined by the high-probability region of the parameter space, a poorly constrained space will require a larger basis to accurately reproduce all possible solutions.

In the following example we see how the amount of error or uncertainty in the data affects the results of the inversion and the basis size. Three identical inversions are performed with data errors: 1%, 5% and 10%. Figure 5.15 shows the effect of the error level in (a) the random walk, (b) the posterior distribution for a representative parameter and (c) the evolution of the basis size.

The simulations run for 10^6 inversion steps and accurately recovered the posterior PDFs for 100 parameters. As expected, the range of possible model parameters explaining the data increases with larger errors. When the knowledge about the data is very precise (1% of error), the algorithm is more strict in accepting models to the random walk. Instead, when we are not so sure about data (10% of error), the algorithm is more permissive and the range of variability of models is larger which translates in wider chains and posterior PDFs. If the number of models whose solution needs to be explained by the basis is larger, the basis size will increase in order to include them.

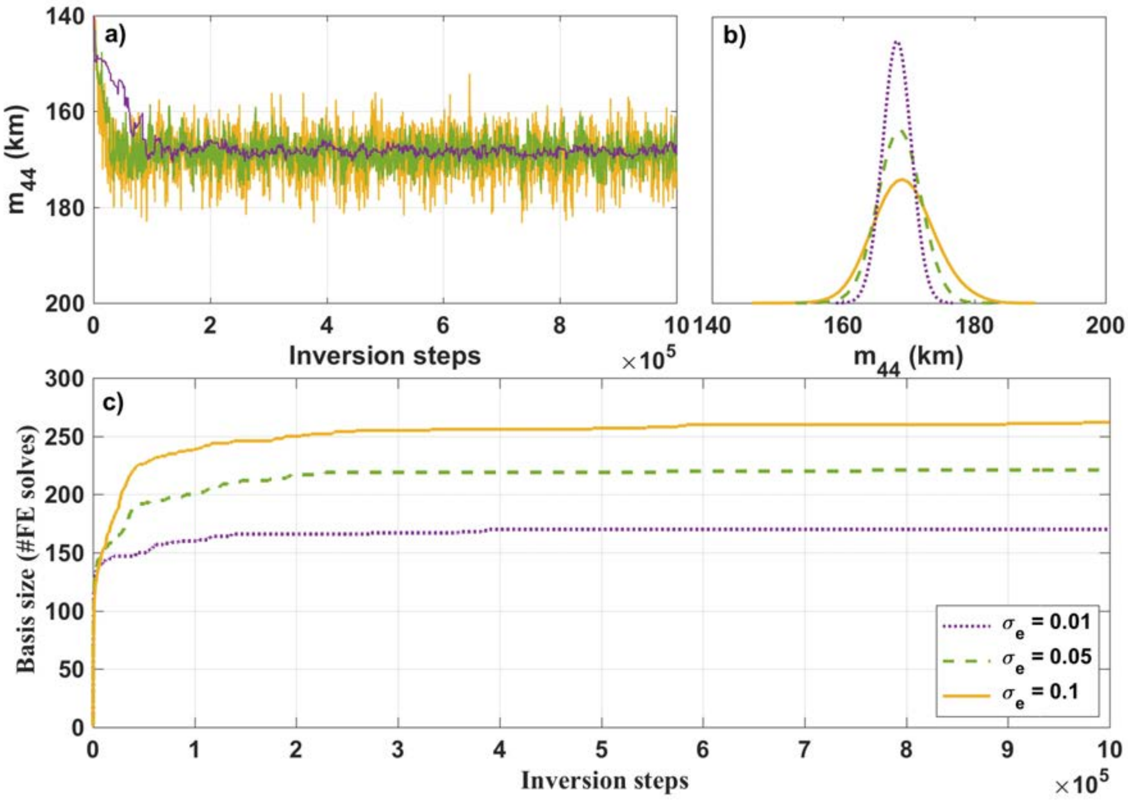

Figure 5.15: Influence of noise size in (a) random walk, (b) posterior distributions and (c) basis evolution size for three different noise levels: $\sigma_e = 0.01$ (dotted purple), $\sigma_e = 0.05$ (dashed green) and $\sigma_e = 0.1$ (solid yellow).

5.4 Influence of the MCMC solver on the order reduction

In previous sections we studied how the error indicator required in the process of building the basis as well as the characteristics of the inverse problem definition affected the basis size. In this section we analyse some aspects of the inverse MCMC solver which can be used to have a better control of the basis. Firstly, we study the influence of different sampling strategies. Then, an adaptation procedure is implemented to automatically explore the parameter space in an efficient way. And, finally, two strategies to dynamically inspect and control the basis are presented.

5.4.1 Sampling strategy

The sampling strategy of the parameter space is a key factor in solving an inverse problem and becomes crucial in high-dimensional spaces. An optimized and efficient search of the parameter space can determine whether an inversion is feasible. With an explorative sampling, the random walk is rapidly moving around the whole space of parameters but it is unlikely to accept new movements (very low acceptance rate). Instead, if the sampling is exploitative and detailed, the random walk is moving very slowly which increases the acceptance probability, but at the same time it may require longer chains to explore the whole space (poor mixing) (Tarantola, 2005; Gregory, 2005). From an inversion point of view, the goal is to design a random walk that efficiently explores the whole space getting neither lost nor stuck. This is achieved by generating samples whose perturbation on the likelihood is as small as possible (to increase the acceptance rate and, therefore, likely to be accepted) but large enough in the parameter space to rapidly explore it. In other words large jumps in the model space that produce small perturbations in the data space.

In our MCMC+RB coupled approach the sampling of the parameter space is performed simultaneously with the generation of the reduced basis. More specifically, the process of adding snapshots or elements in the basis is led by the exploration of the parameter space. Therefore, an efficient strategy from the inversion point of view, may be counterproductive in terms of order reduction. Imagine for instance the situation in which an efficient strategy requires a very large reduced basis (i.e. a lot of FE solves). Although the search is optimal, every likelihood evaluation is very expensive in a computational sense, which means inefficiency in the inversion as a whole.

Many methods have been developed with the idea of improving the sampling strategy and solve the problems described above among many others. Since our goal is not to find the best strategy but to study their relation with the order reduction, we will not attempt complex strategies. Instead, we will keep it simple and focus in two characteristics of the sampling strategies that will help us understand its main influence in the order reduction. Firstly, we will see the effect of sampling the prior by means of a proposal distribution. Secondly, the influence of the number of parameters m_i that are sampled at each inversion step. For all the examples presented next, the prior PDF about model parameters is relatively uninformative and it is represented by a uniform distribution covering a range of 100km (i.e. ±50 km from the reference LAB value). The joint prior PDF is the product of the individual priors $\rho_i(m_i)$ for each model parameter m_i.

Sampling with a proposal distribution

Generating a sample from the prior can be performed in two different ways: directly generating a random value or using a proposal distribution to sample it. Let us start with the former and simpler case. At each inversion step the proposed sample is directly generated from the corresponding prior distribution. In our case, since the priors are uniform, all values are equally likely to be proposed. Consequently, each proposed sample is independent of the previous one. This independence allows a rapid exploration of the space, but at the same time reduces the acceptance rate because perturbations can be very large. Additionally, this kind of exploration results in a space equally sampled, without giving different importance to high and low probability regions. The other option is to use a proposal distribution to sample the prior. Let us assume a one-dimensional Gaussian proposal distribution (used in previous examples): centered at the current state of the chain $m_{i,\text{old}}$ with standard deviation σ_p determining the size of the move. In doing so, each proposed sample depends on the previous one and the exploration is more controlled. Using a proposal allows a more exhaustive exploration (more samples) of high probability regions of the parameter space and less exhaustive (less samples) in low probability regions. In terms of the basis size, this is essential since it allows generating a basis tailored to the high probability regions.

The design of efficient proposals is crucial for the success of any MCMC algorithm and an important research topic in the MCMC literature (Roberts et al., 1997; Roberts and Rosenthal, 2001). Next, the effect that different proposals have on the basis size is analysed. Three identical inversions are run with different jump sizes: (a) $\sigma_p = 5$km, (b) $\sigma_p = 10$km and (c) $\sigma_p = 20$km. The behaviour of the random walk and the basis size is shown in Figure 5.16.

The simulations run for 10^6 inversion steps and accurately recovered the posterior PDFs for 100 parameters. In Figure 5.16, the values of the samples that are accepted as part of the random walk according to the Metropolis ratio r_m are shown in grey, whereas all the samples proposed during the inversion are indicated in colors. Unsur-

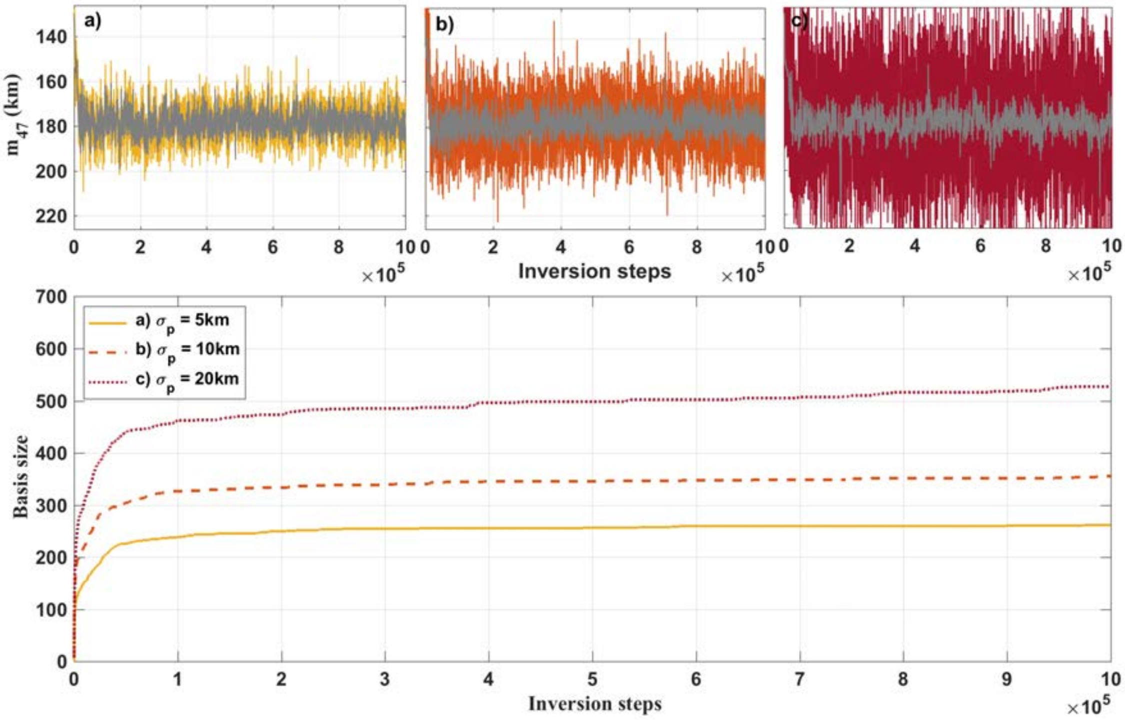

Figure 5.16: Influence of the proposal distribution in the random walk (top) and in the basis evolution size (bottom) for three different proposals: (a) $\sigma_p = 5$km, (b) $\sigma_p = 10$km and (c) $\sigma_p = 20$km. All proposed samples by the inversion (yellow, orange, red) and samples which are accepted (grey).

prisingly, the width of the random walk of accepted moves (gray) and, therefore, the width of the corresponding posterior PDFs coincide. This confirms that, independently of the chosen proposal, the random walk will eventually converge to the same stationary distribution. The difference, however, is how efficiently this convergence will be. As expected, the acceptance rate values decrease with increasing jump size: 50%, 20% and 16% for values 5km, 10km and 20km, respectively. For comparison, if no proposal distribution is used and random samples are directly generated from the prior, the acceptance rate reduces to 8%. According to Gelman et al. (1996) an optimal acceptance is approximately 50% for one dimension and decreases to 23% as dimensions increase.

We also see a significant influence in terms of basis size. Similar to the effect of increasing σ_e, larger values of σ_p (i.e. large proposed moves) result in a larger domain for which accurate solutions are required from the surrogate, which in turn requires larger basis sizes. The size of the basis for each σ_p case is also shown in Figure 5.16 (bottom): 262, 356 and 528 for 5km, 10km and 20km, respectively. The key point is that, even if a sample is not accepted, its forward problem has to be solved and, therefore, it may contribute to increase the basis size. In the extreme case of directly sampling the prior (equivalent to a very large σ_p), the whole space is sampled alike which results in an even larger basis, 628 FE.

Our main point here is that care must be taken with the trade-off between efficient sampling in terms of inversion and in terms of order reduction. In the example presented here, according to optimal efficiency rates we should go for a σ_p around 10km, which represents a larger basis size than the 5km option. If an increase in basis size implies a substantial increase in computing time for the RB solutions, it may even be favourable to have longer random walks (less sampling efficiency), but faster RB solutions (more efficient forward evaluations). This situations is problem dependent and should be analysed for each particular case.

Sampling the prior distribution by blocks

In previous examples just one parameter m_i was modified at a time and the rest remained at their current value. In other words, between $\mathbf{m}_{\text{new}}$ and $\mathbf{m}_{\text{old}}$ only one value was different. This is itself a sampling approach designed to increase the acceptance rate since the corresponding perturbations in the data space are smaller than if the whole vector is changed at the same time. However, for high-dimensional parameter spaces this may lead to a very slow exploration and, therefore, an inefficient inversion. Here, with the idea of moving faster and improving the mixing of the chain, we study the influence of modifying a few parameters simultaneously.

The results presented next are based on ideas from Section 5.5 (advice: read it first!). Three identical inversions are performed with $\sigma_p = 5$km by varying the number of randomly chosen parameters to be sampled per inversion step: 1, 4 and 10. Note that each parameter corresponds to a column of 20 finite elements. Table 5.5 shows final values of acceptance rates as well as the time required to evaluate the stiffness matrices $\mathbf{K}$ and $\mathbb{K}_{\text{RB}}$ per inversion step.

On one hand, the acceptance rate decreases with increasing number of modified parameters, 50%, 10% and 0.9%, because they generate larger perturbations in the

#modified parameters	#modified elements	$\mathbf{K}$	$\mathbb{K}_{\text{RB}}$	Acceptance rate
1	20	0.030s	0.009s	50 %
4	80	0.045s	0.036s	10 %
10	200	0.080s	0.113s	0.9 %

Table 5.5: Values of the acceptance rate and time required to evaluate the stiffness matrices $\mathbf{K}$ and $\mathbb{K}_{\text{RB}}$ for cases in which 1, 4 and 10 parameters are modified per inversion step.

data space. On the other hand, values from Table 5.5 show that the time required to evaluate the stiffness matrices $\mathbf{K}$ and $\mathbb{K}_{\text{RB}}$ also increases as more parameters are modified. At first glance it my not seem a large increase in time, however, matrices must be evaluated at each inversion step. For instance, imagine an inversion that run for 10^6 inversion steps, the total time required to evaluate the matrices sampling 1 parameter at a time is around 10 hours, whereas the same situation modifying 10 parameters takes 53 hours. Therefore, although the mixing of the chain can be improved by modifying more than one parameter at a time, it is at expenses of longer matrix evaluations.

The conclusion drawn from all the examples presented in this section is that in order to increase the computational efficiency of the inversion as a whole, the traditional acceptance rate optimal values ($\sim 35\%$) should be reconsidered. Higher acceptance rates will certainly require longer random walks, but the resulting smaller basis and effortless matrix evaluations can represent a huge difference.

5.4.2 Adaptation (SCAM)

The width of the posterior distribution is not the same for all parameters (Figures 4.5, 5.5, 5.9), indicating that the sensitivity of the velocity field to perturbations varies. Some parameters have a small influence in the velocity field, which implies that they are less constrained by the data and, therefore, exhibit wider distributions. Instead, parameters which produce large perturbations, show narrower distributions. To increase the efficiency of the sampling, the standard deviation of the proposal distribution, σ_p, should be specifically chosen for each parameter.

The adaptive metropolis method proposed by Haario et al. (1999, 2001, 2005) tunes the proposal distribution during the inversion according to the information of previous states of the chain. In this way the variance of the proposal automatically adapts to the posterior or target distribution. The process loses its Markovian character, but Haario et al. (2001) proved that, under certain conditions, the process is ergodic and has the desired stationary distribution. There are two main variants of the algorithm: the Adaptive Metropolis (AM) and the Single Component Adaptive Metropolis (SCAM). The main difference is that AM uses a multivariate proposal distribution and it is adapted using the covariance matrix of the generated chain (Haario et al., 2001). However, for high-dimensional spaces, sampling all parameters at the same time can be time consuming both for the covariance evaluation and, as described in the previous section, the forward problem itself. Instead, SCAM algorithm samples one parameter at a time using one-dimensional proposal distributions which are individually adapted using the variance of each parameter chain (Haario et al., 2005). The proposal distribution is a Gaussian distribution centered at the current value of the chain and has a variance that depends on time. Renaming the current state of the chain for parameter i as m_i^t, the variance of its proposal distribution v_i^{t+1} for time $t+1$ is adapted as,

$$v_i^{t+1} = \begin{cases} v_i^0 & t \leq t_0 \\ s\mathrm{Var}(m_i^1, \ldots, m_i^t) + s\epsilon & t > t_0 \end{cases} \tag{5.9}$$

where v_i^0 is an initial proposal variance, $\mathrm{Var}(m_i^1, \ldots, m_i^t)$ is the sample variance of the posterior values of the chain up to the current step, ϵ is a small constant to prevent the variance from being zero and $s = 2.4$ is a scaling factor. Its value is set according to the results of Gelman et al. (1996) in which it is shown that such value optimizes the mixing properties of the search in the case of Gaussian posteriors and proposals. The parameter t_0 is used as a burn-in period before adaptation starts. Defining $g_i^t = \mathrm{Var}(m_i^1, \ldots, m_i^t)$, the variance of the chain values can be efficiently updated with new samples using the following recursive expression,

$$g_i^{t+1} = \frac{1}{t}\left[(t-1)g_i^t + \frac{t+1}{t}(\overline{m}_i{}^{t+1} - m_i^{t+1})^2\right] \tag{5.10}$$

where $\overline{m}_i{}^{t+1}$ is the mean of $t+1$ samples which also satisfies a recursive expression

$$\overline{m}_i{}^{t+1} = \frac{1}{t+1}\left[t\ \overline{m}_i{}^t + m_i^{t+1}\right]. \tag{5.11}$$

We have implemented the SCAM algorithm and studied its influence in the order reduction. To be consistent with previous examples we prefer to deal with standard deviations, i.e. $\sigma_i^0 = \sqrt{v_i^0}$. Two simulations of 100 parameters are run with initial values $\sigma_i^0 = 2\text{km}$ and $\sigma_i^0 = 30\text{km}$. Proposal distributions are updated at intervals of 50.000 inversion steps which is also the chosen value for t_0. The following results show the evolution of the acceptance rate and basis size compared to the situation in which adaptation is not performed.

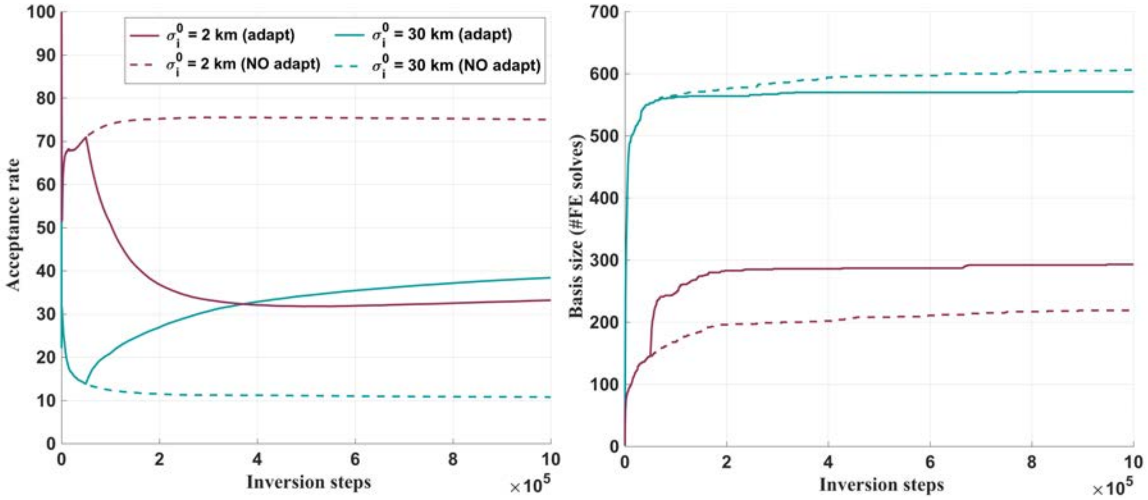

Figure 5.17: Evolution of the acceptance rate and basis size with and without adaptation with initial standard deviations $\sigma_i^0 = 2\text{km}$ and $\sigma_i^0 = 30\text{km}$ for for $i = 1, \ldots, n_m$.

Figure 5.17 shows two different behaviours. For the case with a small $\sigma_i^0 = 2\text{km}$, the acceptance rate is initially very high (around 70%) indicating that most of the jumps are accepted and the space is slowly explored. The SCAM algorithm is able to correct it by increasing the values of σ_i^0 (Figure 5.18) so that the acceptance rate decreases to more optimal values. Instead, if the initial value of $\sigma_i^0 = 30\text{km}$ is large, the behaviour is the opposite. The acceptance rate is initially very small (around 15%) and, once adaptation starts, the jump sizes reduce to increase the acceptance rate. After a certain number of adaptations, the acceptance rate in both cases is around 35%. Dashed lines show the resulting acceptance rate curves if no adaptation was performed. Adaptation also controls the evolution of the basis size. It adapts as required by the new values of σ_i^t. For $\sigma_i^0 = 2\text{km}$, the basis is initially small, but as adaptation takes place and σ_i^t values increase, the basis needs to provide accurate solutions to a now larger region of the parameter space and, therefore, a richer basis is required. A too large initial basis defined by a too large initial σ_i^0 can also be controlled with adaptation, but, unlike the previous case, adaptation must take place before the basis stabilizes. Although the shorter the burn-in time t_0, the more evident the effect in the basis size, care must be taken to have a relevant number of samples before adaptation. Finally, it is important to remark that the resulting posterior distributions are not affected by adaptation and are in excellent agreement with previous examples.

Figure 5.18 shows the initial and final values of the standard deviation (σ_i^0 and σ_i^t, respectively) for each of the 100 parameters.

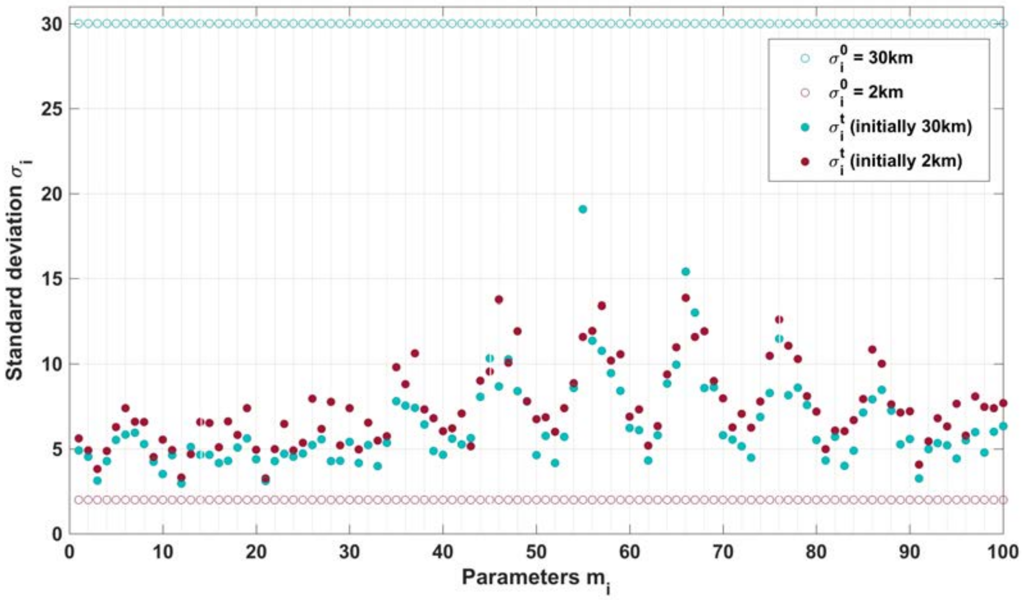

Figure 5.18: Standard deviation values of each of the 100 proposal distributions. Initial σ_i^0 and values of σ_i^t after the last adaptation for the two cases shown in previous Figure 5.17: 2km and 30km.

For each parameter, the final values of σ_i^t are very similar independently of the starting point and, in fact, they would eventually coincide for a long enough number of inversion steps. It is also interesting to observe that, as mentioned at the beginning, the final σ_i^t values are different among parameters: larger values are exhibited for those parameters whose posterior distributions are wider. Tuning each proposal individually would be very difficult without adaptation.

Adaptation is a good approach to deal with the typical compromise between the wish of a fast exploration of the parameter space and the need of generating samples which can be accepted. The proposal distribution is automatically adapted to improve the mixing of the chain and, therefore, convergence is faster. In addition, the basis size can also be controlled and adapted as the inversion requires, either if initially is too small or too large.

5.4.3 Dynamic control of the basis size

There are two main reasons to perform a dynamic control of the basis size: 1) orthogonalize the basis and 2) remove irrelevant information as the Markov chain moves to a more restricted parameter space, i.e. high-probability regions. In doing so, the idea is to obtain a smaller, compact and data-driven basis.

As explained in Section 5.1, our greedy approach to build the basis does not guarantee orthogonality and it allows some level of linearity due to the predefined tolerance. Although our tests show it is not relevant in terms of conditioning, redundant information can be eliminated and the basis reduced accordingly. The Singular Value Decomposition (SVD) is excellent at reducing dimension and eliminating small dependencies allowed during the construction of the basis. Without getting into details, the SVD is a numerical tool that factorizes a matrix $\mathbf{B} \in \mathbb{R}^{m \times n}$ into the product of 3 matrices,

$$\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}} \tag{5.12}$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthonormal matrices and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ a diagonal matrix with the singular values of $\mathbf{B}$ ordered from largest to smallest. Columns of $\mathbf{U}$ and $\mathbf{V}$ are called left and right singular vectors (see for example Brunton and Kutz (2019) or Quarteroni et al. (2016) for details). Following this idea, our basis $\mathbf{B}$ can be decomposed similarly and the resulting matrix $\mathbf{U}$ used as an orthonormal basis. Since we are also interested in reducing the size, not all columns of $\mathbf{U}$ are considered. Intuitively, columns of $\mathbf{U}$ capture the correlations in columns of $\mathbf{B}$ and the magnitude of the corresponding singular value indicates their contribution at explaining $\mathbf{B}$ (Figure 5.19). Therefore, only such columns with a significant contribution are retained. Deciding where to truncate the matrix is not straightforward and depends on factors which are problem dependent (maximum desired size, level of noise, etc). Typically an energy criteria in the accumulated variance is used. Other strategies inspect the distribution of singular values to identify certain shapes (“elbows” o “knees”) that help deciding where to truncate (Brunton and Kutz, 2019).

However, the SVD does not take into account the problem at hand. In other words, it is aimed at keeping as much independent information as possible regardless of its relevance in terms of high-probability regions. For instance, let us consider a basis element that was added to the basis at the beginning of the inversion from a region that is never visited again. This element is representing a portion of the parameter space that is no longer important for the inversion, but it provides independent information since it is the only one explaining that particular region of the space. Therefore, its singular value will be considerably high and it will not be discarded.

Instead, if our interest is to get rid of those basis vectors whose contribution at explaining the high-probability regions is no longer relevant, a different approach must be devised. A possible strategy is to check the solution vector $\boldsymbol{\alpha}$ of the reduced system, equation (3.32). Such coefficients indicate the contribution of each basis element. A representative reduced system ($\mathbb{K}_{\text{RB}}$ and $\mathbb{f}_{\text{RB}}$) is created for a parameter vector which is the current average of the chains, $\overline{\mathbf{m}} = [\overline{m}_1, \ldots, \overline{m}_{n_m}]$, and the system is solved. Then, basis vectors with coefficients below a given threshold are discarded. In doing so, the approach filters those solutions which are outside the high-probability regions of the parameter space and, hence, obtain a more compact and tailored basis.

In our case, since the basis is constructed during the inversion, if relevant information is removed by any reduction strategy, extra high-fidelity FE computations will be required to re-learn such solutions. Although it guarantees that accuracy is always restored, one must be careful of the resulting additional cost if the basis is trimmed too much. In other words, although a smaller and compact basis can be achieved, it is at expenses of extra high-fidelity solutions which imply an additional cost that should not be underestimated. Therefore, some trade-off should be sought in order to remove redundancies while keeping all the information.

The two strategies are applied to the 100 parameters inversion (Section 5.3.1) after 2×10^5 inversion steps. The coefficients of the resulting linear combination as well as the singular values obtained from applying a SVD procedure are shown in Figure 5.19 ordered from largest to smallest.

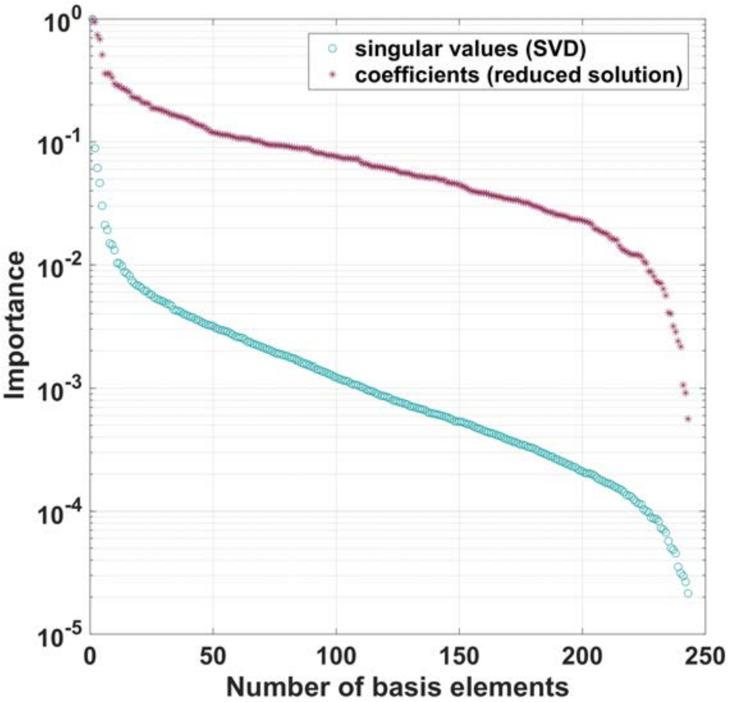

Figure 5.19: Singular values (SVD) and $\boldsymbol{\alpha}$ coefficients of the representative reduced solution for each basis vector after 2×10^5 inversion steps. Values are normalized with respect to the maximum.

Figure 5.19 shows a difference in importance between both approaches. Values represented with red asterisks characterise the relative importance or contribution of each basis element (with respect to the element that contributes the most) at explaining the high probability regions of the parameter space. Instead, blue circles express the contribution of each element considering dependencies and correlations among them. In both curves, the elements exhibiting the lowest contributions are candidates to be removed. The shape of the curve shown by the singular values (circles) characterises the amount of linearity allowed during the construction of the basis (determined by the RB tolerance e_{RB}). This particular shape suggests that the basis enrichment algorithm is working because, although not fully orthogonal, linearity is controlled.

Next we will run two identical inversions applying reductions at intervals of 2×10^5 inversion steps. Two different cut-off tolerances are used for each reduction approach. First, the basis evolution using the SVD strategy is shown in Figure 5.20.

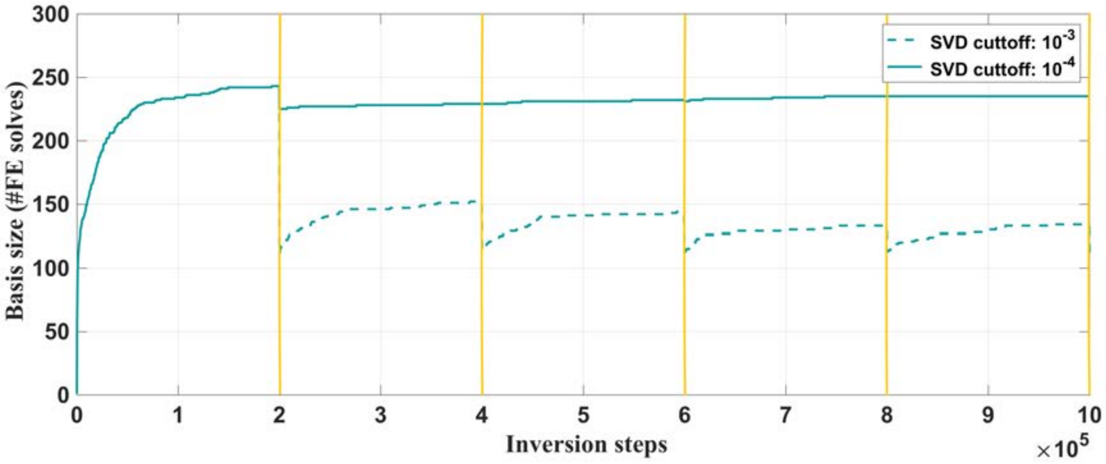

Figure 5.20: Evolution of the basis size using a SVD reduction approach with two different cut-off tolerances: 10^{-3} (dashed line) and 10^{-4} (solid line). Reduction is performed every 2×10^5 inversion steps (vertical lines).

It is not trivial to know beforehand which should be the proper cut-off tolerance that does not imply a subsequent enrichment period, i.e. additional high-fidelity FE solutions. According to the values from Figure 5.20, it seems a 10^{-4} cut-off tolerance is able to get rid of the linearity without implying too many extra solutions, whereas for the 10^{-3} case, after reduction there is always an enrichment period. This suggests that reduction is not only removing correlations, but also important information that eventually needs to be replaced. Although some of the dependencies removed in the first reduction are not recovered again, the subsequent enrichment periods confirm that the algorithm to enrich the basis is working since there is not much redundant information that can be eliminated without needing it again. According to our tests, the slight reduction achieved in the basis size does not significantly reduce the cost of obtaining a RB solution and, in addition, it brings the possibility of extra FE solutions.

Second, Figure 5.21 shows the basis evolution where the reduction approach aims to remove basis elements which are outside the high-probability regions.

Figure 5.21 shows that after reduction the basis grows again, suggesting that some of the removed basis elements were actually relevant. For this approach to be practical, a fine-tuning of the cut-off tolerance is essential. Since our basis results are good enough and in our tests the computational gains are not significant, we did not attempt it. However, if reduction was necessary, it seems reasonable to perform only

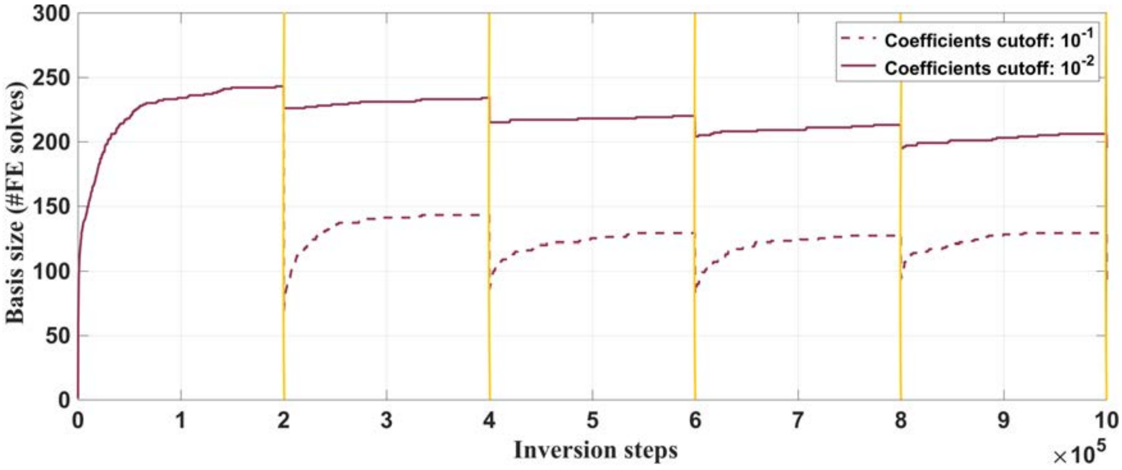

Figure 5.21: Evolution of the basis size using the coefficients of a RB reference solution to reduce it with cut-off tolerances 10^{-1} (dashed line) and 10^{-2} (solid line). Reduction is performed every 2×10^5 inversion steps (vertical lines).

one reduction to discard the uninformative elements added during the burn-in period, since the chances to need them again are low. In addition, the more advanced is the inversion before reduction, the more representative is $\overline{\mathbf{m}}$ of the high-probability areas and, therefore, more accurate is the reduction. This explains the tendency of the solid curve to converge to a smaller basis size.

More sophisticated strategies could be devised to adaptively control the reduction of the basis. For instance, evaluating the required time to compute a RB solution and apply a reduction if it is larger than a certain threshold (e.g. related to the cost of solving a high-fidelity FE problem). Also, a strategy described in Manassero (2019) in which reduction is performed when the basis size reaches a certain value. A maximum basis size is initially fixed and its value is redefined after every SVD reduction to progressively allow the basis to increase if required.

5.5 Efficient assembly of stiffness matrices

Having a reduced basis is a great improvement in reducing the order of the model and obtaining fast RB approximations. However, a fast approximation is not only determined by the basis size (i.e. the size of the reduced linear system), but it also depends on the assembling efficiency of the matrices involved. The computational cost of the assembly of the FE stiffness matrix $\mathbf{K}$ and the computational cost of reducing it via the multiplication $\mathbf{B}^\mathsf{T}\mathbf{K}\mathbf{B}$ are significant and cannot be neglected. Note that although the size of the reduced system is small, $n_{\text{RB}} \times n_{\text{RB}}$, the multiplication $\mathbf{B}^\mathsf{T}\mathbf{K}\mathbf{B}$ depends on n_{FE} and can be computationally expensive. All RB methods should efficiently handle these operations since a high cost could render them impractical. This section describes the procedures used to perform these two operations in an efficient way.

Table 5.6 shows the final basis sizes for 9 inversions varying the number of inversion parameters and the mesh discretization.

Mesh size	Length $\mathbf{u}$	Basis size	
		25 params	100 params
2000	28743	118	262
4500	61728	115	-
8000	107163	119	327
18000	235383	104	332
32000	413403	108	314

Table 5.6: Final basis sizes obtained for the inversion of 25 and 100 parameters for five different mesh discretizations. For the 25 parameters case the inversion run for 2×10^5 inversion steps, whereas for the 100 parameters they run longer, 10^6 inversion steps. It also shows the length of the basis vectors for each mesh discretization.

Two conclusions are drawn from here. First, the number of elements in the basis does not directly depend on the discretization of the forward problem (if the mesh is fine enough to describe the solution). In other words, for a fixed number of parameters and guaranteeing a certain accuracy for the RB problem, a finer discretization of the mesh does not affect the number of elements in the basis. Second, although the basis size does not significantly change, the size of each element $\mathbf{u}$ of the basis increases with the discretization. Therefore, finer discretizations increase the computational time of performing the multiplication $\mathbf{B}^\mathsf{T}\mathbf{K}\mathbf{B}$.

5.5.1 Efficient assembly of the FE stiffness matrix K

The assembly strategies described next are based on two main ideas. On one hand, they rely on the additive property of the weak form. Essentially, the global stiffness matrix $\mathbf{K}$ can be thought as a sum of elemental contributions $\mathbf{K}^e$. On the other hand, the performed assembly strategy takes advantage of the specific parametrization of the problem and sampling strategy, i.e. at every inversion step, the modification of the viscosity field is local, only a few elements change.

Combining both ideas, we devise an assembling strategy that, at each inversion step, only needs to update the part of the global stiffness matrix that has changed with respect to the previous step. Initially, the global stiffness matrix is computed by assembling the contribution of all elements in the mesh, whereas in the successive iterations, only the elemental stiffness matrices of the elements that have been modified are computed and their contribution is updated in the global matrix. In doing so, the computational cost of evaluating the stiffness matrix for each inversion step is remarkably reduced. The required time to evaluate or update **K** is determined by the number of elements that have been modified and need to be updated. Figure 5.22 shows the required time to update the stiffness matrix **K** with respect to the number of modified elements with a mesh of 2000 finite elements (Q2-P1).

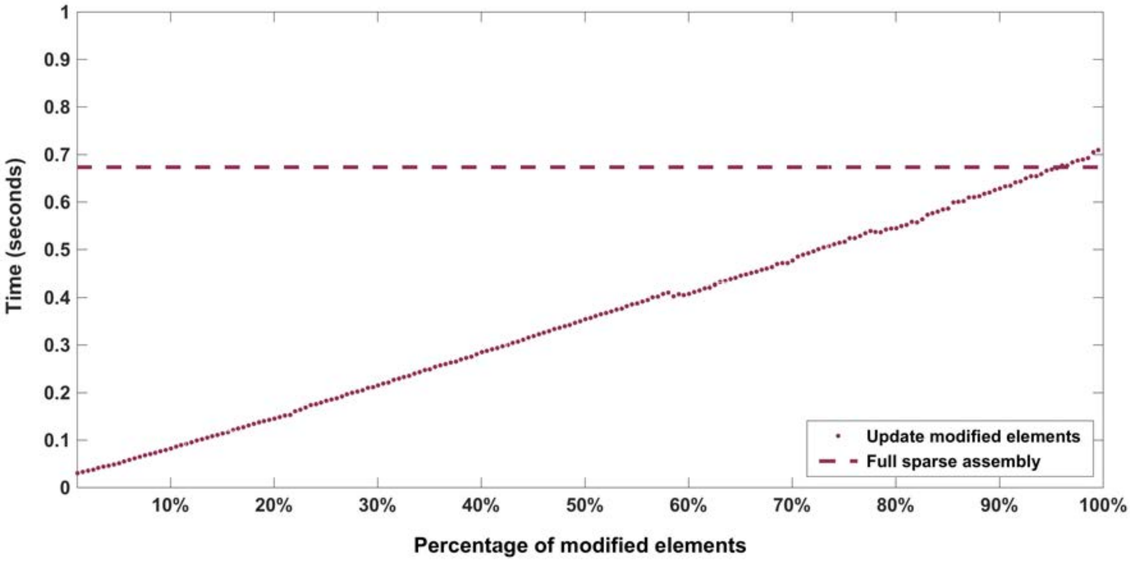

Figure 5.22: Time required to assemble the stiffness matrix **K** by only updating the contribution of the elements that have changed (each point is an average of 100 different **K** updates). For comparison, the time required to perform a full sparse assembly of **K** is shown with an horizontal dashed line. The mesh has 2000 finite elements (Q2-P1)

Figure 5.22 indicates that the reduction in computational time is significant when the modification is local in space. By only updating the contribution of the elements that have been modified, the required computational time to evaluate **K** is adapted to the amount of change and grows linearly with it. Instead, the original assembly strategy is always 0.7s per $\mathbf{K}$ evaluation regardless of the number of elements that changed. This behaviour maintains for larger matrices.

5.5.2 Efficient assembly of the reduced stiffness matrix $\mathbb{K}_{\text{RB}}$

The reduced stiffness matrix $\mathbb{K}_{\text{RB}}$ must be evaluated at each inversion step. Performing the multiplication $\mathbf{B}^{\mathsf{T}}\mathbf{K}\mathbf{B}$ can be time-consuming. To avoid such computation and efficiently obtain the reduced stiffness matrix $\mathbb{K}_{\text{RB}}$, the elemental-updating idea used for $\mathbf{K}$ is also applied to compute $\mathbb{K}_{\text{RB}}$.

In the normal situation, the reduction is applied at the global level, i.e. once we have the global matrix via a matrix product: $\mathbb{K}_{\text{RB}} = \mathbf{B}^{\mathsf{T}}\mathbf{K}\mathbf{B}$. Here, instead, the reduced approach takes place at the elemental level, which means that the elemental matrix $\mathbf{K}^e$ is the one reduced,

$$\mathbb{K}_{\text{RB}}^e = \mathbf{B}^{e\mathsf{T}}\mathbf{K}^e\mathbf{B}^e \ . \tag{5.13}$$

In doing so, we defined a new basis $\mathbf{B}^e$ which is essentially $\mathbf{B}$ restricted to the degrees of freedom (nodes) that belong to the element e. Since all reduced elemental matrices $\mathbb{K}_{\text{RB}}^e$ have the dimension of the final reduced system $\mathbb{K}_{\text{RB}}$, all elemental contributions can be added together directly (no assembly is required),

$$\mathbb{K}_{\text{RB}} = \sum_e \mathbb{K}_{\text{RB}}^e \ . \tag{5.14}$$

Following this approach, the full multiplication only needs to be performed once at the beginning. For all successive iterations the elemental reduced matrix $\mathbb{K}_{\text{RB}}^e$ is only computed for the elements that are modified and the global reduced matrix $\mathbb{K}_{\text{RB}}$ is updated with their contribution. This strategy drastically reduces the cost of the full multiplication since, essentially, only the relevant parts are actually computed. Similar to the stiffness matrix, the more elements that need to be modified, the longer it takes. In addition, not only this strategy is much faster, but it also allows us to avoid evaluating $\mathbf{K}$ at each inversion step.

However, there is an implementation detail that needs to be mentioned. If the elements that have been modified are adjacent, some of their nodes are shared. Consequently, since the reduced implementation takes place at elemental level, the reduction (basis multiplication) for the nodes that belong to more than one element is performed separately for each elemental contribution and then added. This implies that for such nodes, there is an extra number of operations with respect to the global scenario where the stiffness contribution of these nodes is first assembled (“added”) and then reduced (“multiplied”). If the number of adjacent elements that are modified is very large, the total cost will eventually be higher than performing the full multiplication at a global level. The optimal number of operations can be achieved if the reduction is performed at nodes level. The stiffness contribution of all modified elements is first added together in a matrix $\mathbf{K}^{\mathrm{dof}}$, and then the reduction with $\mathbf{B}^{\mathrm{dof}}$ is performed. The reduced stiffness matrix accounting for all modified elements is,

$$\mathbb{K}_{\mathrm{RB}}^{\mathrm{modif}} = \mathbf{B}^{\mathrm{dof}\mathsf{T}}\mathbf{K}^{\mathrm{dof}}\mathbf{B}^{\mathrm{dof}} \tag{5.15}$$

where “dof” refers to the nodes or, in the case of vectorial unknowns, to the degrees of freedom that have been modified. Note that $\mathbf{K}^{\mathrm{dof}}$ is (dof) × (dof) and a special assembly must be performed to correctly place the modified elemental contributions.

Figure 5.23 shows the time required to update the reduced stiffness matrix $\mathbb{K}_{\mathrm{RB}}$ depending on the size of the basis. It is shown for different percentages of modified elements: 0.05%, 1%, 4% and 100% which correspond to 1, 20, 80 and 2000 modified elements, respectively.

Similar to the assembly of $\mathbf{K}$, only updating the contribution of the elements that have been modified represents a drastic reduction in computational time. The time required to update 1, 20 and 80 elements is always significantly less than performing the full multiplication $\mathbf{B}^{\mathsf{T}}\mathbf{K}\mathbf{B}$ and it grows linearly with the basis size. The time growth of the full multiplication is also linear with the basis size and reaches around 3s for a basis of 1600 elements (not shown for a better visualization). When all elements are modified, updating them or doing the full multiplication have a similar cost. The observed difference is due to the extra operations required for the updating strategy, but it becomes irrelevant for larger basis.

Let us exemplify the benefit of the described strategies with some numbers. Consider, for instance, the inversion from Section 5.3.1: 100 parameters which required a basis of 262 elements for 10^6 inversion steps, and sampling 1 parameter (20 elements)

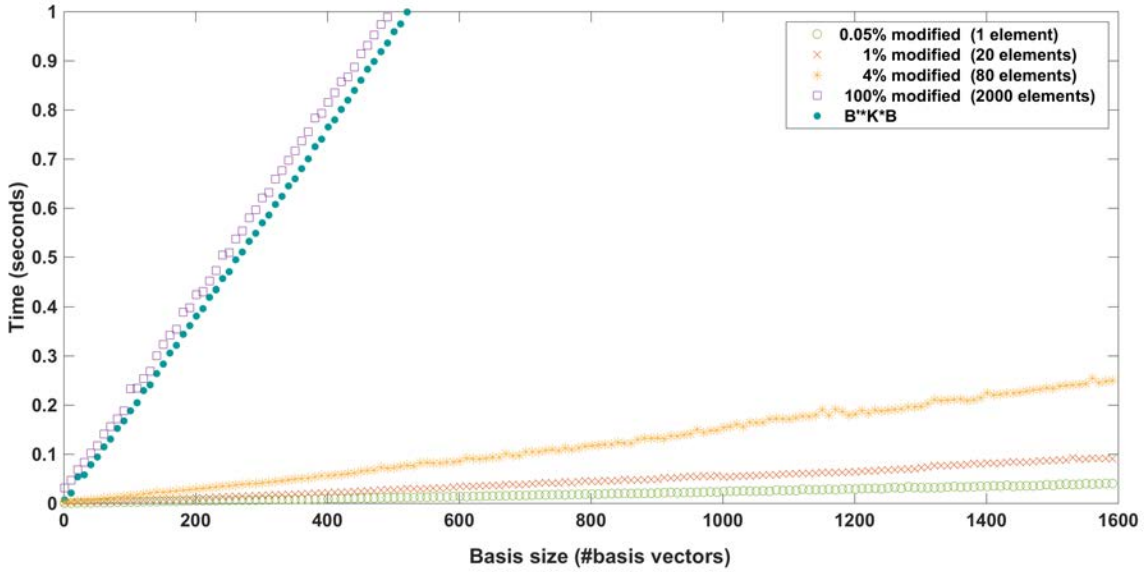

Figure 5.23: Time required to update the reduced stiffness matrix $\mathbb{K}_{\text{RB}}$ by only updating the contribution of the elements that have changed as function of the basis size. For comparison, the time required to perform the full multiplication $\mathbf{B}^{\mathsf{T}}\mathbf{K}\mathbf{B}$ is also shown with blue filled circles. The mesh has 2000 finite elements (Q2-P1)

per inversion step. The total time required to evaluate $\mathbf{K}$ and $\mathbb{K}_{\text{RB}}$ with the full assembly and doing the full multiplication would be around 132 hours. Instead, the same number of evaluations taking advantage of the elemental-updating strategies only needed 4 hours. This represents a 3% of the total full assembly time, a drastic reduction!

Basis enrichment and $\mathbb{K}_{\mathbf{RB}}$ assembly

In our coupled approach, the reduced basis is built on the fly during the inversion process. Consequently, it does not have a fixed size, but it changes as the enrichment takes place. If the basis is enriched, the elemental-updating strategy described above can not be performed, because the reduced matrix would be updated with elemental contributions of a larger dimension. Therefore, whenever the basis needs to be enriched we have two alternative ways of proceeding: either the whole multiplication is performed or a new row and column are added to $\mathbb{K}_{\text{RB}}$ to include the information of the new basis element. Both strategies have been tried and according to our results, performing the multiplication is more efficient than enlarging the matrix $\mathbb{K}_{\text{RB}}$ to account for the new element. When the basis is rich enough and FE solutions are no longer required, the matrix $\mathbb{K}_{\text{RB}}$ can be reused and updated at a very low cost for the rest of the MCMC inversion.

The results of this section demonstrate the benefit of the elemental-updating strategies both for evaluating $\mathbf{K}$ and $\mathbb{K}_{\text{RB}}$. The updating strategies always perform better than the corresponding full assembly or full multiplication regardless of the number of modified elements. The more elements that need to be modified, the closer they are with the full strategies, until they become equivalent when all elements are modified. Moreover, although the basis size is still something to take into account, its impact can be, to some extent, controlled with the elemental-updating approach. The sampling strategies described in Section 5.4.1 perfectly suit this approach since only some parameters are modified per inversion step.

Chapter 6

Inversion examples

In this chapter the coupled MCMC+RB method is applied to two realistic inversion examples. To better represent a portion of the Earth, the problem domain is no longer rectangular but spherical. In the first example the method is applied to a problem in which the topography is used as synthetic observables to infer the LAB structure. In the second example, the velocity field is again used in a larger inversion problem to determine the lithospheric structure and sublithospheric upper mantle flow pattern beneath the African continent and surroundings. While realistic in nature, we emphasize that this example is for illustrative purposes only.

6.1 Spherical 3D cartesian model using topography

In this example the surface dynamic topography is used as observable to determine the LAB structure of a spherical domain representing a portion of Earth. As explained in Section 2.4, in equilibrium, all forces acting on a column of fluid must balance (their sum must be zero) so that no movement is allowed. Therefore, the normal viscous stresses at the surface, σ_{nn}, coming from the mantle flow are balanced with the lithostatic stresses of the overlaying material column, $\rho g w$, of density ρ and height w, equation (2.16). It is important to remark that equation (2.16) just balances the dynamic effects. Therefore, the "overlaying material column" does not refer to the whole topography, but to the extra part that cannot be explained isostatically.

In our model implementation, the mesh is fixed and we impose that the normal velocity on the surface is zero (shear stresses are also zero). In doing so, we do not allow the top boundary of the domain to deform according to the normal viscous stresses generated below and, as a consequence, the stress σ_{nn} on the surface can not be zero. Instead, the reaction stresses $-\sigma_{nn}$ on the surface indicate which would be the topography generated to balance the system if movement was allowed. The stresses on the surface are computed using the velocity field obtained from solving the Stokes equation so that the final expression for the instantaneous dynamic topography w is,

$$w = \frac{2\mu}{\rho g}\frac{\partial u_n}{\partial \mathbf{n}}, \tag{6.1}$$

where u_n is the normal component of the velocity and $\mathbf{n}$ the normal direction which, in our spherical domain, coincides with the radial direction. Note that since $u_n = \mathbf{u} \cdot \mathbf{n}$, the normal derivative is obtained as, $\partial u_n / \partial \mathbf{n} = \nabla u_n \cdot \mathbf{n} = (\nabla \mathbf{u} \cdot \mathbf{n} + \mathbf{u} \cdot \nabla \mathbf{n}) \cdot \mathbf{n}$.

The interest region Γ of the QoI also needs to be modified. It was defined in accordance with the a priori information on the parameters and it was a box of 100km centered in the prior. In previous examples where the observable was the velocity, it was within this region where the approximated RB solution was sought to be particularly accurate. Instead, now the topography is computed on the surface so that the accuracy of the RB solution should not be just within the box, but it should also account for the surface. Therefore, recalling the QoI from equation (5.1), now rewritten considering the radial direction,

$$Q = \int_{\Gamma} \frac{\partial u_n}{\partial \mathbf{n}} d\Omega, \tag{6.2}$$

the interest region Γ is modified by expanding the box upwards to reach the surface. Note that, in this case, the choice of the QoI coincides with the observables since both require the gradient of the velocity along the vertical direction.

Next, we compare the behavior of the inversion scheme when using velocity and topography as observables. The first example, having 25 parameters, uses the LAB structure from Figure 4.4a and it starts from the same flat configuration, Figure 4.4b. Topography observations are measured in a 20×20 regular grid on the surface of the domain. The same simulation is repeated using the velocity as observables adjusting the number of observations also to 400. Both simulations run for 2×10^5 steps. Results of the random walks are presented in Figure 6.1.

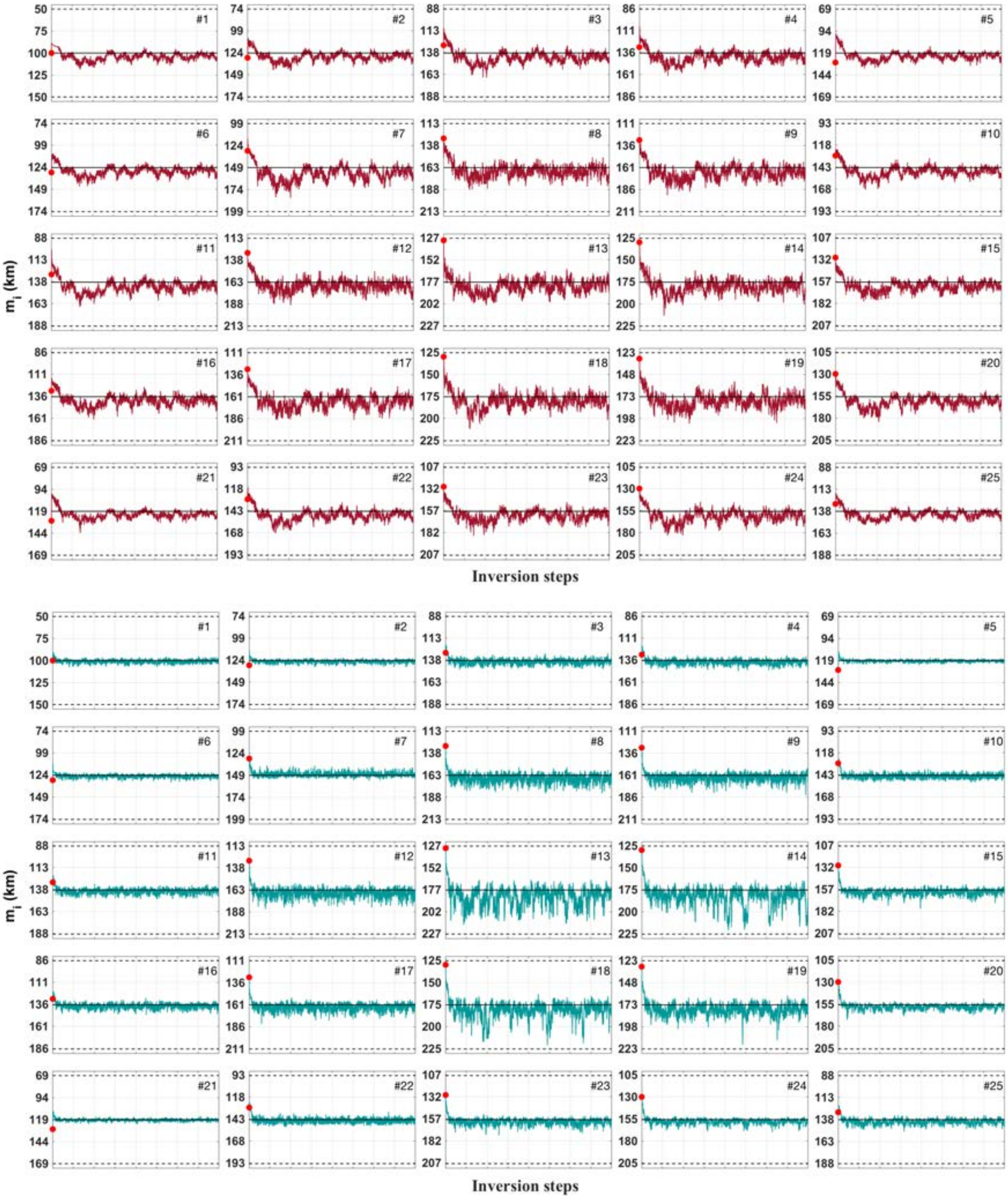

Figure 6.1: Random walk for each of the 25 parameters using topography (top) and velocity (bottom) as observables. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic observables (solid black line) and initial value (red dot).

Despite the initial uninformative scenario, the random walks in both cases managed to converge to the true LAB values and remained oscillating around. The proposal standard deviation used is $\sigma_p = 5$km which gives an acceptance rate of approximately 47% in both cases. The number of full FE solutions required for the inversion is 199 for the topography and 153 for the velocity. The larger basis for the topography also coincides with the mostly wider random walks with respect to the velocity plots. This suggests that, in general terms, topography is less sensitive to perturbations than velocity. In addition, velocity does not constrain all parameters equally, whereas topography seems to have a similar sensitivity among parameters. Other differences are observed between both cases: the length of the burn-in period, the random walk behaviour, etc. Before getting into details, we will present an example which will help to clarify and explain such differences.

This simple example is used to study the dependence of the misfit function between observed and predicted data, $e(\mathbf{m})$, with respect to the parameters value (LAB depth). We show that, in this case: i) the misfit has many local minima, ii) the local minima when the data is the velocity are “deeper” than for the topography and, iii) the global minimum observed for the topography is “flatter” than for the velocity. This indicates that, during an inversion, the random walk is likely to find local minima and the efficiency in escaping from it will be determined by the jump size in each situation. The evolution of the misfit with the LAB depth shown in Figure 6.2 is achieved using the following strategy. The observed data is obtained from the LAB structure from Figure 4.4a. Starting from the same LAB structure shifted 50km upwards (i.e. shallower), at each step we increase the average depth by increasing the depth of one parameter in 5km. In doing so, after 25 steps all parameters are increased in 5km and we recover the initial LAB structure shifted 5km deeper. This process is repeated as many times as needed to achieve a certain depth. The full forward problem is solved at each step and the misfit is computed with both observables.

In Figure 6.2, the misfit values for the velocity are larger than for the topography, indicating again that velocity is more sensitive to perturbations. The evolution of both misfits shows oscillations with local minima at intervals of 5km. This value is due to the fact that the increasing step is 5km. As designed, every 5km the same LAB structure is recovered and the misfit exhibits a minimum. If the moving step was 20km, the minima would be exhibited at intervals of 20km. This oscillating

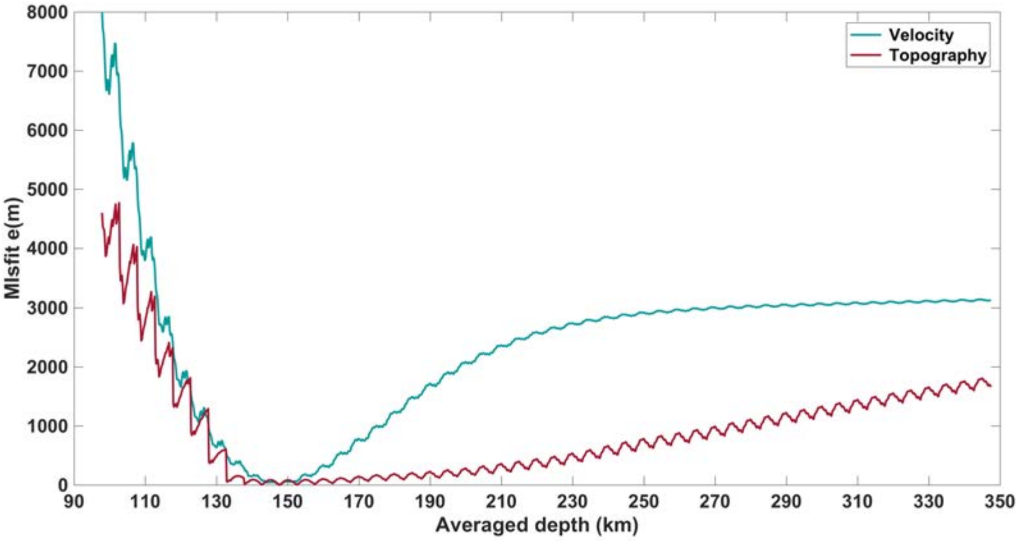

Figure 6.2: Evolution of the misfit with respect to the LAB depth for both observables: vertical velocity and topography. The x-axis shows the averaged depth of the 25 parameters. At each step one single parameter is modified 5km.

behaviour, although not truly representing the path of a normal inversion, indicates that local minima do exist. Two characteristics of these plots can be used to compare the previous inversion examples (Figure 6.1). On one hand, the flatter shape and lower misfit exhibited for the topography gives more freedom to the random walk to explore regions near the absolute minimum than for the velocity case. On the other hand, the amplitude of the oscillations is slightly larger for the topography than the velocity which suggests that, once in a minimum, it is more difficult to escape from it for the topography. Their combination makes the topography more prone to getting stuck in local minima. The shape of the random walks from Figure 6.1 is in accordance with such behaviour: the random walk for the topography seems to move more slowly within the high probability regions than for the velocity case, although both show similar acceptance rates. This effect is magnified when the initial guess is worse (larger initial misfit), since the chances to find minima in the way increase and, therefore, the convergence gets slower. Summarizing, the misfit behaviour explains the longer random walks needed for the topography. Moreover, its lower sensitivity demands for a more sophisticated or refined approach, for instance individually setting proposals for each parameter, ideally with an adaptive strategy.

Another important insight learned from Figure 6.2 is that the observables are more sensitive to shallow perturbations than deeper ones. Consequently, the observables are not able to constrain the deeper parameters as much as the shallow ones. This is translated in wider random walks and, therefore, deeper parameters tend to have wider posterior PDFs than shallower ones. The decreasing sensitivity with depth explains the tendency of the posterior PDFs to loose symmetry and get wider in the deeper side. As a consequence, their mean is slightly shifted towards deeper values. This does not mean that such distributions are wrong or inaccurate, they are, indeed, the actual solution to the probabilistic inverse problem.

Next two inversion examples are aimed at recovering the posterior PDF of 100 inversion parameters using topography values to constrain them. The reference LAB and the synthetic topography generated for examples A and B are shown in Figure 6.3. Note that on the reference LAB figures (bottom), darker colors represents deeper LAB values. In the topography figures (top), red colors indicate an uplift of the surface whereas blue ones correspond to depressed areas.

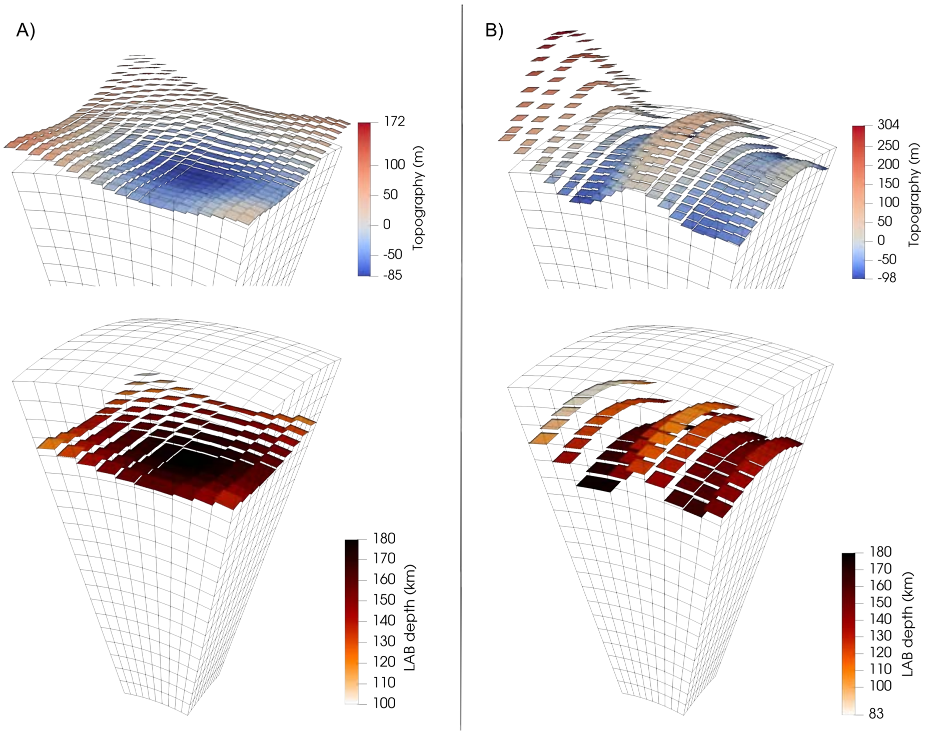

Figure 6.3: 3D spherical domain discretized in 10×10×20 mixed elements representing a portion of Earth of 660km depth and 400km in the two directions along the surface. Synthetic examples A and B: reference LAB (bottom) and the corresponding generated topography (top). Topography is not to scale.

Both inversions have $n_d = 400$ topography data points with an introduced noise of $\sigma_e = 10^{-1}$. In order to better control the error in the observables and considering that for the topography the dispersion of observable values is not as large as with the velocity, we use their mean value to compute the synthetic data errors (instead of the maximum value as in equation (4.29)). The prior information of the parameters is a uniform PDF of 100km (±50km from the reference LAB value). The initial LAB configuration is obtained by randomly perturbing the reference LAB between 10% and 20%. The standard deviation of the proposal distribution is set at $\sigma_p = 5$ km for all parameters and the tolerance for the RB error in the QoI is $e_{\text{RB}} = 10^{-2}$ (1% of error). Both inversions run for 10^7 steps and the first 10^6 are considered burn-in period. Since the obtained posterior PDFs exhibit quite simple forms, plotting their mean values and standard deviations is informative enough. Results for examples A and B are shown in Figures 6.4 and 6.5, respectively.

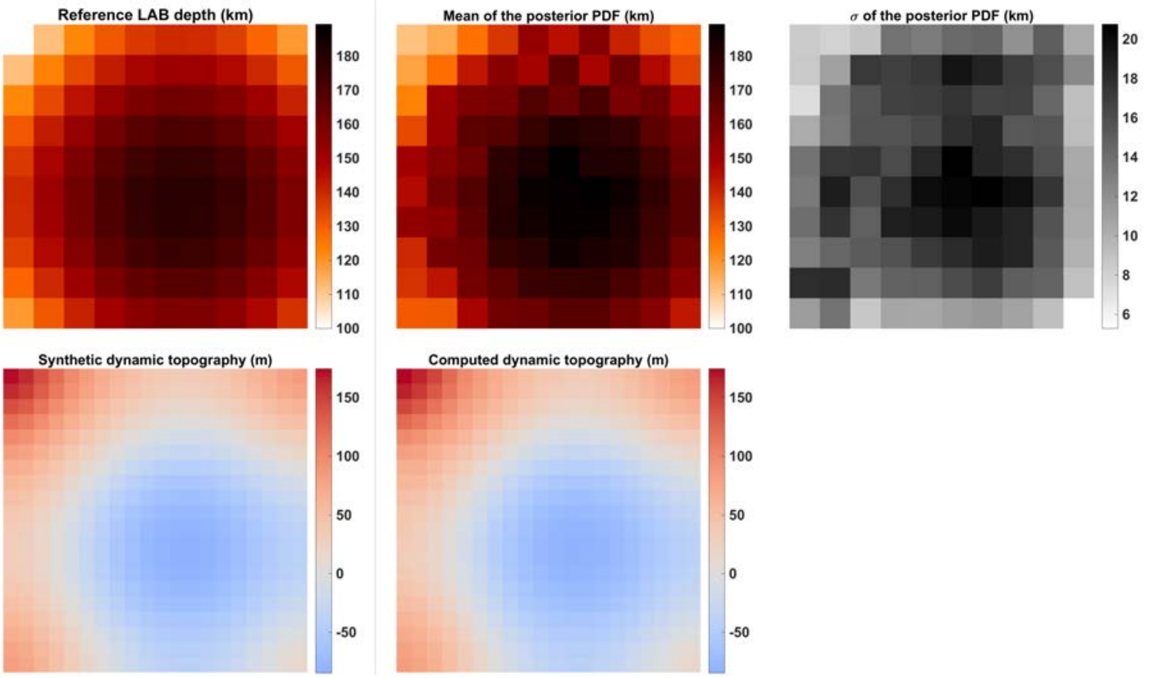

Figure 6.4: Top view of the results for inversion A. Top images represent the LAB structure: reference LAB depth (left), mean of the recovered posterior PDF for the LAB depth (center) and the corresponding standard deviation (right). Bottom images represent the topography associated for the LAB structure above: synthetic dynamic topography (left) and dynamic topography computed from the mean posterior (center). Detailed plots of each of the 100 random walks and posterior PDFs are shown in Figures B.3 and B.4 from Appendix B, respectively.

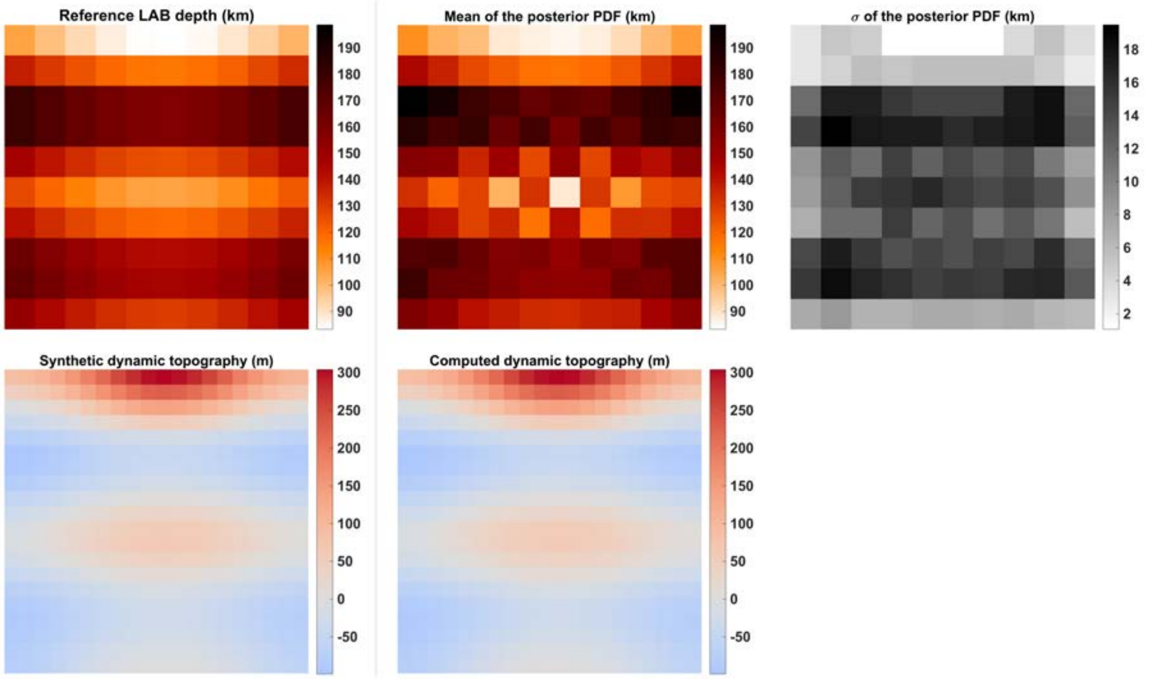

Figure 6.5: Top view of the results for inversion B. Top images represent the LAB structure: reference LAB depth (left), mean of the recovered posterior PDF for the LAB depth (center) and the corresponding standard deviation (right). Bottom images represent the topography associated for the LAB structure above: synthetic dynamic topography (left) and dynamic topography computed from the mean posterior (center). Detailed plots of each of the 100 random walks and posterior PDFs are shown in Figures B.5 and B.6 from Appendix B, respectively.

Despite the very uninformative initial scenario, both inversions managed to recover a global LAB structure that matches the reference one and most of the reference values are within $\pm\sigma$ of the posterior PDFs. As expected, since topography is less sensitive than velocity, the inversion has more difficulties in recovering the exact reference values and some slight discrepancies are observed within the general shape. A detailed description and possible solutions are discussed below. However, topography plots show excellent agreement between synthetic and computed observables, the largest difference is around 5 meters. Deeper LAB values exhibit larger standard deviations and, thus, wider posterior PDFs. Such behaviuour is in accordance with the fact that topography sensitivity decreases with depth and, consequently, deeper parameters are less constrained. The inversion acceptance rates are 48% and 45% for examples A and B, respectively. Regarding the basis size, example A needed 918 full FE solutions whereas example B needed 1233, i.e. 0.009% and 0.012% of the total number of forward evaluations. In both cases the gain in computational time of the MCMC+RB approach with respect to the full MCMC+FE is outstanding. This gain makes feasible to include the dynamic component of the topography within a multi-observable inversion.

It is worth mentioning a compensation effect that is observed in the inversions with topography. This compensation occurs when the random walks of adjacent parameters oscillate around values which are not the reference ones, but their generated topography fits the data similarly, i.e. their misfit is within the error of the observables. Figure 6.6 illustrates such behaviour.

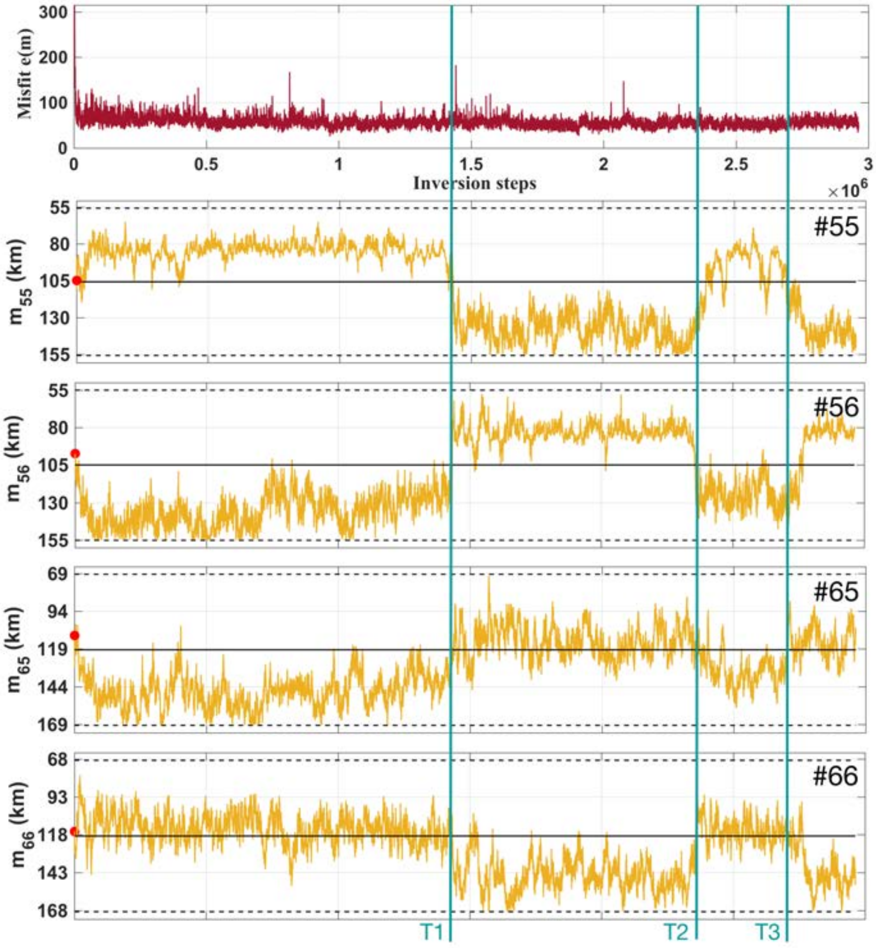

Figure 6.6: Evolution of the misfit and random walk for 4 adjacent parameter during the first 3×10^6 inversion steps for example B. Vertical lines indicate a compensation change. In each random walk plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic topography (solid black line) and initial value (red dot).

Figure 6.6 shows the random walk of four adjacent parameters (their adjacent neighbours behave similarly) during the first 3×10^6 inversion steps for example B. Initially, parameter 66 oscillates around the true value and, since it is constrained, it gives more freedom to the other parameters to move further from their true values without affecting the misfit. During this time, shallow and deeper values are compensated in terms of misfit. The rest of parameters also try jumps which bring them closer to their true values. Around T1, parameter 65 moves closer to its true value and successfully remains there. Consequently, it removes some of the constrain on the other parameters, including parameter 66, which can now jump to other regions. Between T1 and T2, it is parameter 65 who oscillates around its true value whereas the others move freely. Again, their attempts are not successful until T2. It is important to remark that reaching a high probability region does not depend on a single parameter but it is a combination of all of them. Therefore, it may happen that during the path to a high-probability region, a parameter is forced to go through a low-probability region. Or in other words, the misfit has to increase before being able to decrease to lower values. The efficiency in converging to the true solution and the good mixing of the chain completely depend on the proper combination of jump sizes. Nevertheless, the effect of compensations in the posterior PDFs is reduced with longer inversions.

In results from Figures 6.4 and 6.5, the compensation effect is responsible of the chessboard pattern in some parts of the LAB structure. Additional tests performed with MCMC+FE also present compensations which indicate that it is not an effect introduced by the MCMC+RB approach, although the RB may intensify it. In fact, it is a natural effect of probabilistic inversions which is magnified due to the low sensitivity of the observables combined with their uncertainties and a simplistic sampling strategy. If observables are not sensitive enough, their uncertainties allow for such compensations without significantly affecting the misfit. An additional example is included in Appendix A.3 in which observables sensitivity is higher and, hence, the effect of compensations is much lower. As previously mentioned, the jump size also plays an important role in the duration of the compensations. The fact that we are using the same proposal for each parameter is considerably limiting the movement of the random walks, since some parameters may need larger (or smaller) jumps than others to efficiently explore the space (Section 5.4.2). In this case, adaptation is an interesting approach to improve the mixing of the chain and, therefore, reduce the effect of compensations without having to rely on much longer random walks.

Inversion example B is run again with an adaptation strategy. Initially the proposal distribution is set at $\sigma_p = 30$ km and adaptation is performed at intervals of 5×10^4 steps. The inversion run for 5×10^6 (half the steps of the previous inversion) and results are shown in Figure 6.7.

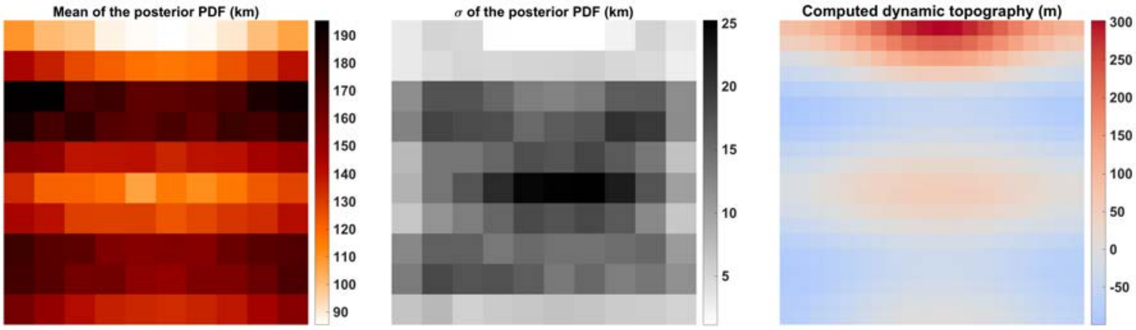

Figure 6.7: Top view of the results for inversion B with adaptation. Mean of the recovered posterior PDFs for the LAB depth (left), the corresponding standard deviations (center) and the dynamic topography computed from the mean posteriors values (right).

Adaptation helps to considerably reduce the chessboard pattern and after a much lower number of inversion steps. Posteriors are sampled with more accurate proposals that are in agreement with posterior widths, which is confirmed by looking at the final values of σ_p of each proposal distribution from Figure 6.8.

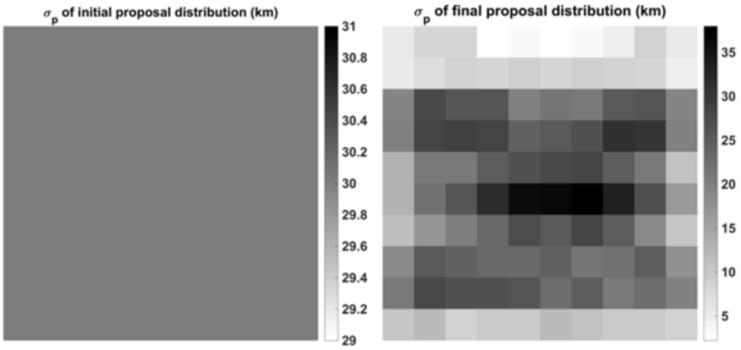

Figure 6.8: The initial σ_p for all proposal distributions is set at 30km (left). During the inversion proposals adapt to efficiently sample their corresponding posteriors (right)

Generally, wider posteriors require larger jumps to move rapidly around their space, whereas narrower ones need smaller jumps to increase the chances to hit a high probability region without getting lost. However, it is important to have in mind that such behaviour is true for an efficient sampling of individual parameters, whereas an efficient simultaneous sampling can be slightly different.

Another possibility to control and reduce compensations would be to include in the sampling strategy correlations between parameters, since in doing so, parameters would be somehow linked to their neighbours and compensations among neighbours would not be possible.

The results presented here show the feasibility of inverting the LAB using topography observations only. The proposed method works and makes possible computations that were previously unaffordable. This encourages the incorporation of the dynamic component into multi-observable inversions. In that situation, many of the difficulties studied here will be probably smoothed by the extra information provided by other observables.

6.2 Large-scale probabilistic inversion: African lithospheric structure

In this section we apply our method to a larger example to determine the lithospheric structure and sublithospheric upper mantle flow pattern beneath the African continent and surroundings. The mantle vertical velocities are used again as observables to infer a LAB depth described by 1225 parameters. While realistic in nature, we emphasize that this example is for illustrative purposes only.

The numerical domain is spherical and discretized with $35\times35\times20$ finite elements from the surface down to a depth of 1000km depth. The lithospheric structure (LAB depth and temperatures) of the reference model is based on the global lithospheric model by Afonso et al. (2019). Sublithospheric temperatures are computed based on the work of Stixrude and Lithgow-Bertelloni (2012) for a reference adiabatic gradient, with anomalies with respect to that gradient from the work of Steinberger and Becker (2018). Density and viscosity are computed as a function of temperature and pressure using equations (4.26) and (4.28), respectively. Figure 6.9 shows the reference LAB and two vertical profiles of the generated velocity field obtained by solving the Stokes problem. Blue indicates flow is going downwards, whereas red moves upwards.

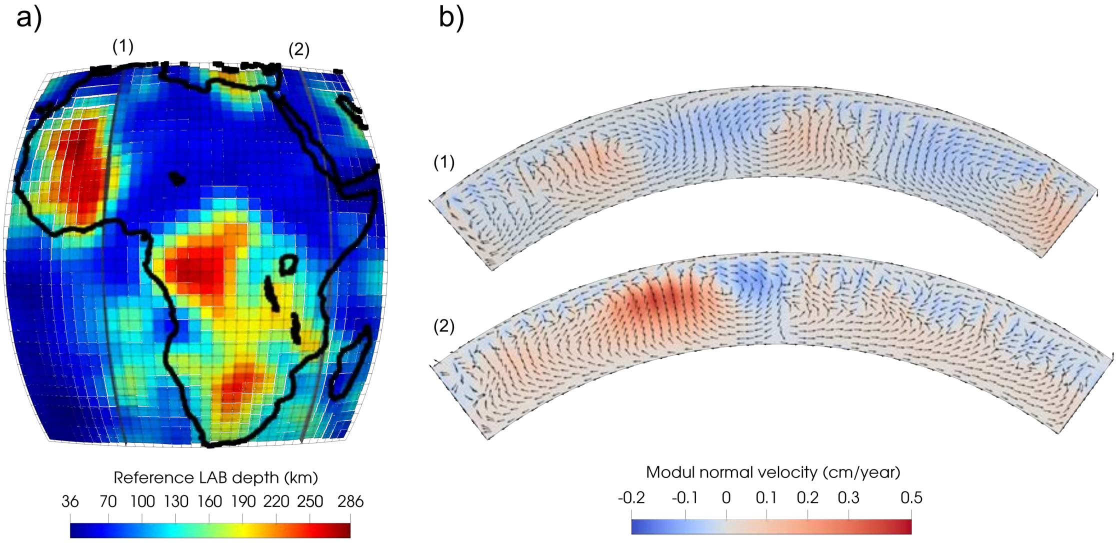

Figure 6.9: (a) Reference LAB discretized with 1225 parameters. (b) Two vertical profiles showing the flow (velocity) structure.

Since we are attempting to solve a considerably larger problem by only using the velocity to constrain the parameters, we will slightly relax the uninformative initial scenario, so that the error introduced in the synthetic observables is reduced to 5% ($\sigma_e = 0.05$). Moreover, unlike previous velocity examples, the initial LAB configuration is not flat, but it is obtained by randomly perturbing the reference one between 10% and 20%. The prior information of the parameters is maintained as a uniform PDF of 100km (±50km from the reference LAB). We run the inversion algorithm for 4×10^6 steps, from which the first 10^5 samples are discarded as burn-in. The standard deviation of the proposal distribution is set at $\sigma_p = 10$ km and the threshold for the error estimator of the RB is set at $e_{\mathrm{RB}} = 10^{-2}$. Mean values and standard deviations of the posterior PDFs resulting from the inversion are shown in Figure 6.10.

The results in Figure 6.10 compared to the reference LAB from Figure 6.9 demonstrate that the inversion succeeded in recovering the mean values of the 1225 true parameters. Wider posteriors are obtained in areas where the lithosphere is thicker, indicating that the observables are less sensitive to perturbations beneath these columns. These regions coincide with broad slow downwellings where small local perturbations

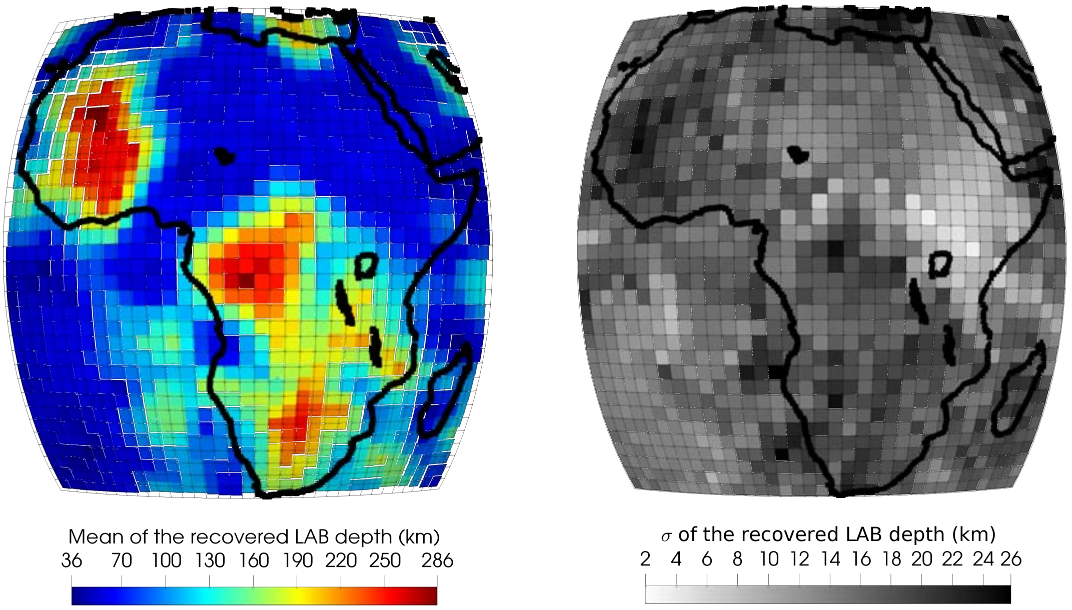

Figure 6.10: (a) Mean and (b) standard deviation of the posterior PDF for each of the 1225 parameters representing the LAB depth recovered during the inversion.

of the LAB do not change the main velocity pattern or magnitude in any significant way. The narrowest distributions have standard deviations of around 2km whereas the widest show values of 26km. A similar example with 225 parameters is shown in Appendix B with detailed plots of the posteriors and random walks. That example is also published in Ortega-Gelabert et al. (2020).

The efficiency of the method is again outstanding; to perform 4×10^6 inversion steps, only 1916 FE solves where required (0.05% of the total number of steps). It is worth emphasizing that in a real inversion context, more than 4×10^6 steps would be typically run, which will increase even further the efficiency of the method.

Chapter 7

Conclusions and future work

This work presents the coupling between the Reduced Basis method and a Metropolis algorithm (MCMC method) to efficiently solve geophysical probabilistic inverse problems. The MCMC+RB method has been applied to two different synthetic observables, mantle velocities and topography, in order to infer properties of the interior of the Earth, in particular the LAB structure. The solution of the probabilistic inverse problem is not unique, but characterised by a posterior PDF which is approximated by means of generating random samples of it using the Metropolis algorithm. Each sample requires a forward evaluation of the Stokes problem which is efficiently performed using the RB method. The RB consists of a collection of FE solutions that form a reduced basis, such that every new forward evaluation is obtained as a linear combination of the elements of the reduced space $\mathbb{R}^{n_{\text{RB}}}$ instead of the full space $\mathbb{R}^{n_{\text{FE}}}$. Since $n_{\text{RB}} \ll n_{\text{FE}}$, solving the reduced Stokes system is noticeably faster.

7.1 Conclusions

The main achievements and conclusions of this thesis are summarised below:

- Unlike classical offline/online RB strategies, here a greedy approach is proposed which constructs the reduced basis *on the fly* and guided by the inversion. Such approach requires an error estimator to control the maximum error allowed to the reduced solutions. If the error is larger than a given tolerance, the corresponding full FE problem is solved and the basis is enriched with the new solution. In doing so, the generation of the basis is coupled with the sampling of the posterior PDF in such a way that the basis is specifically tailored to the high-probability regions of the parameter space. Due to the converging nature of the MCMC inversions, the enrichment of the basis also converges so that, when it is rich enough, it stops growing and stabilizes. This greedy approach solves the typical problems of offline/online strategies regarding where to sample the parameter space so that the snapshots are representative enough. Consequently, the basis size required in order to attain a certain accuracy in the solution is smaller than using classical strategies.

- The efficiency of the MCMC+RB method is outstanding. The proposed method is able to successfully solve a probabilistic inverse problem at a largely reduced cost compared with the MCMC+FE approach and without compromising the results. Since the time required to obtain a RB solution is remarkably smaller than a FE one, the number of required full FE gives us an idea of the cost of the method. For instance, to solve 10^5 forward problems only 116 were required with FE, all the rest were obtained at the RB low-cost, this represents only 0.1% of the total number of evaluations. In fact, the longer the inversion, the more efficient becomes the method. The efficiency of the forward solver allows solving a larger amount of problems which is translated into a larger number of posterior samples and, therefore, better accuracy in the posterior results. The MCMC+RB strategy opens the door to inversion problems that were unfeasible until now.

- The coupled MCMC+RB method proposed in this thesis is independent of the solver used to obtain the full FE solutions that form the reduced basis. For simplicity, here we have used direct solvers. Direct solvers allow reducing even more the computational cost since the incompressibility condition of the RB solutions is automatically satisfied and equation 3.32 can be used. Instead, if using iterative solvers, the terms involving the reduced $\mathbf{G}$ matrix do not strictly vanish and the full reduced system from equation 3.30 is required. However, the only additional cost is saving the pressure FE solutions into $\mathbf{B_p}$, since the reduced matrix $\mathbf{G}$ is fixed among inversion steps and only needs to be recomputed if the basis is enriched.

In order to guarantee fast RB approximations in comparison with the FE ones, some aspects must be carefully considered. The cost of obtaining a RB solution is mainly determined by: the basis size and the efficiency in assembling the stiffness matrices. In this sense, the following points are highlighted:

- The accuracy of the RB solutions is a key factor in determining the basis size. The smaller is the error requested for the RB approximation, the larger is the basis. With this idea in mind, we have developed a goal-oriented error estimator which centers the accuracy of the RB solution on a specific QoI relevant for the inversion problem. In doing so, the basis is considerably reduced since it only needs to be accurate for representing the relevant part of the solution. Such error representation implies the solution of an adjoint problem whose accuracy does not seem to be a determinant factor for the final basis size. Its accuracy improves the error estimator, but the difference does not significantly affect the basis enrichment criterion.

- Ideally, the basis size is expected to be determined by the high probability regions of the parameter space. Therefore, the more uncertainty in the data, the less constrained is the parameter space and larger basis are required to explain it. The basis also grows with the number of inversion parameters. More parameters imply more combinations which translates into a richer family of solutions to be represented by the basis. However, we have seen that the basis growth scales with the number of parameters in a tractable way (up to 1225 parameters)

- In practice, since the high probability regions of the parameter space are to be determined during the inversion, the basis size is not strictly determined by them, but also by the sampling strategy. That is, it does not matter whether a proposed sample is finally accepted to the random walk, since once it is proposed, it has to be solved. Therefore, its solution can potentially contribute to increase the basis size. It is crucial to devise efficient sampling strategies that take into account the basis size. For instance we have shown that a direct sampling of the prior PDF leads to a very large basis since all regions are sampled alike. Instead using a proposal distribution with a very small jump size generates a considerably smaller basis, although at expenses of a very slow exploration of the parameter space. Adaptation is a useful tool to automatically adapt the proposal distribution to efficiently explore the space while controlling the basis size. With the aim of a small basis, we also proposed a SVD strategy to remove redundancies and an approach which identifies and eliminates the basis elements which are not contributing significantly to the high probability regions. However, it is important to highlight that a too strict reduction may imply additional FE solutions.

- Finally, we also proposed efficient assembly strategies for the stiffness matrices $\mathbf{K}$ and $\mathbb{K}_{\text{RB}}$. We take advantage of the parametrization of the inverse problem and the sampling strategy. Since only the viscosity of some elements change between inversion steps, the matrices are not computed from scratch, but they are updated with only the contribution of the elements that have changed. Therefore, the assembly time grows linearly with the number of modified elements. In addition, for the reduced stiffness matrix we also avoid performing the full multiplication $\mathbf{B}^{\mathsf{T}}\mathbf{K}\mathbf{B}$ at a global level and, instead, the reduction is performed at the elemental one and the reduced elemental matrices are added together. Again, only the contribution of the modified elements is updated.

To conclude, it is important to remark that in the context of MCMC inversions, the typical trade-off between fast exploration and the need to be accepted determines the efficiency of the inversion. However, with the inclusion of reduced order methods, such trade-off should be rethought, since an efficient sampling in terms of parameter space does not directly imply efficiency in terms of order reduction. It could happen that a slightly worse sampling, improves the global efficiency due to faster forward evaluations.

7.2 Future work

Several aspects of the proposed MCMC+RB method can be further developed:

- The present work deals with a linear rheology since the second invariant of the strain rate that appears in the constitutive equation (4.28) is kept constant. It would be interesting to also account for nonlinear rheological effects in which viscosity is defined depending on stress or strain rate. The MCMC+RB methodology presented in this thesis is still valid for a nonlinear rheological law, but some aspects must be carefully analysed. For instance, the need for iterative solvers both for the full and reduced system or the adaptation/rethinking of the fast assembly strategies of the stiffness matrices for the nonlinear case.

- The *on-the-fly* strategy to build the reduced basis presented in this thesis requires discarding all samples whose solution is obtained via full FE (or, alternatively, recomputing additional FE solutions). This is to guarantee a fair comparison in the Metropolis ratio. It would be interesting to develop an enhanced algorithm which is able to use the FE solutions in a fair RB-FE comparison without compromising the results (see for instance the approach in Manassero et al. (2020)). Another interesting option to be explored is to consider the error of the RB solutions part of the inversion unknowns, i.e. an inversion parameter.

- More complex strategies could also be devised with the idea of controlling the basis size during the inversion in a dynamic way. For instance, tracking the time required to obtain a RB solution and if higher than a threshold (related to the alternative high-fidelity FE cost) reduce the basis, either with SVD, with the reduced coefficients approach or even a combination of both.

- It would be interesting to develop more efficient sampling strategies which account for correlations between parameters and study the possible effects they may have in the basis size.

- Finally, the main goal behind the MCMC+RB approach presented in this thesis is to provide a fast Stokes solver which can be coupled with the multiobservable probabilistic inversion framework (e.g. LitMod3D from Afonso et al. (2013a,b, 2016b)), to improve the estimation of the elevation by means of including the contribution of the dynamic topography. To achieve this goal, there are two important points that need to be adapted. First, the MCMC+RB code developed in this thesis is fully written in MATLAB and should be rewritten in Fortran so that it can used within LitMod3D. Second, for larger problems and meshes like the ones LitMod3D deals with, a direct solver is no longer useful and an iterative and fast FE solver for the Stokes problem is crucial for the generation of the FE solutions that comprise the reduced basis.

7.3 Research dissemination

Some of the presented scientific contributions are published in the following article:

- Ortega-Gelabert, O., Zlotnik, S., Afonso, J. C., & Díez, P. (2020). Fast stokes flow simulations for geophysical-geodynamic inverse problems and sensitivity analyses based on reduced order modeling. *Journal of Geophysical Research: Solid Earth*, 125, e2019JB018314. `https://doi.org/10.1029/2019JB018314`

The scientific contributions have also been presented at the following national and international conferences:

- Ortega-Gelabert O., Zlotnik S., Afonso J.C. and Díez P. Incorporating dynamical information into 3D inversion problem to determinate lithospheric structure: efficient solvers for Stokes problems via Model Order Reduction. Subduction Interface Processes, Castelldefels, Spain, 2017.
- Ortega-Gelabert O., Zlotnik S., Afonso J.C. and Díez P. Model Order Reduction for the fast solution of 3D Stokes problems and its application in geophysical inversion. EGU General Assembly, Vienna, Austria, 2017.
- Ortega-Gelabert O., Zlotnik S., Afonso J.C. and Díez P. Ultra-fast solution of Stokes problem via Model Order Reduction within 3D inversion of lithospheric structure. NMASE + AdMoRe annual research meeting, Castelldefels, Spain, 2018.
- Ortega-Gelabert O., Zlotnik S., Afonso J.C. and Díez P. Ultra-fast solution of Stokes problem via Model Order Reduction within 3D inversion of lithospheric structure. EGU General Assembly, Vienna, Austria, 2018.
- Ortega-Gelabert O., Zlotnik S., Afonso J.C. and Díez P. Incorporation of dynamic effects into lithospheric inversion by using a very efficient Reduced Basis solver for 3D Stokes problems. GeoMod, Barcelona, Spain, 2018.
- Ortega-Gelabert O., Zlotnik S., Afonso J.C. and Díez P. Coupling Reduced Basis with Markov Chain Monte Carlo for the study of lithospheric structure. MathRocks Workshop Series, Barcelona, Spain, 2019.

Appendix A

Additional examples

A.1 Avoiding an inverse crime

In this additional example, the same MCMC+FE simulation as in Section 4.3 has been performed with the only difference that, in this case and in order to avoid the so-called inverse crime (Section 4.3.2), the synthetic observables are computed with a mesh twice finer than the mesh used during the inversion. Therefore, the mesh used to compute the synthetic observables has 20×20×40 finite elements. Figure A.1 shows the histograms of the predicted values for 49 of the 675 synthetic observables. All plots are centered at the synthetic data value (vertical red line) and their x axis ranges from $-\sigma_d$ to $+\sigma_d$.

Unlike Figure 4.8, here we can see that the simulation has more difficulties in predicting the synthetic values since the true synthetic values might be outside the interpolating space that is later used during the inversion to predict them. It might even be impossible to exactly predict them. Even though all the predicted values are within $\pm\sigma_d$, their means do not always coincide with the synthetic ones. This particular example clearly shows that modelization errors do exist. This is what actually happens in the real life, where there is no finite interpolating space able to exactly represent nature and the best we can do is approximate it. Consequently, the random walks are not expected to converge to the true model of parameters but to the model that best represents the "achievable" synthetic observables. Figure A.2 shows the random walks and posterior PDFs obtained for each of the 25 parameters.

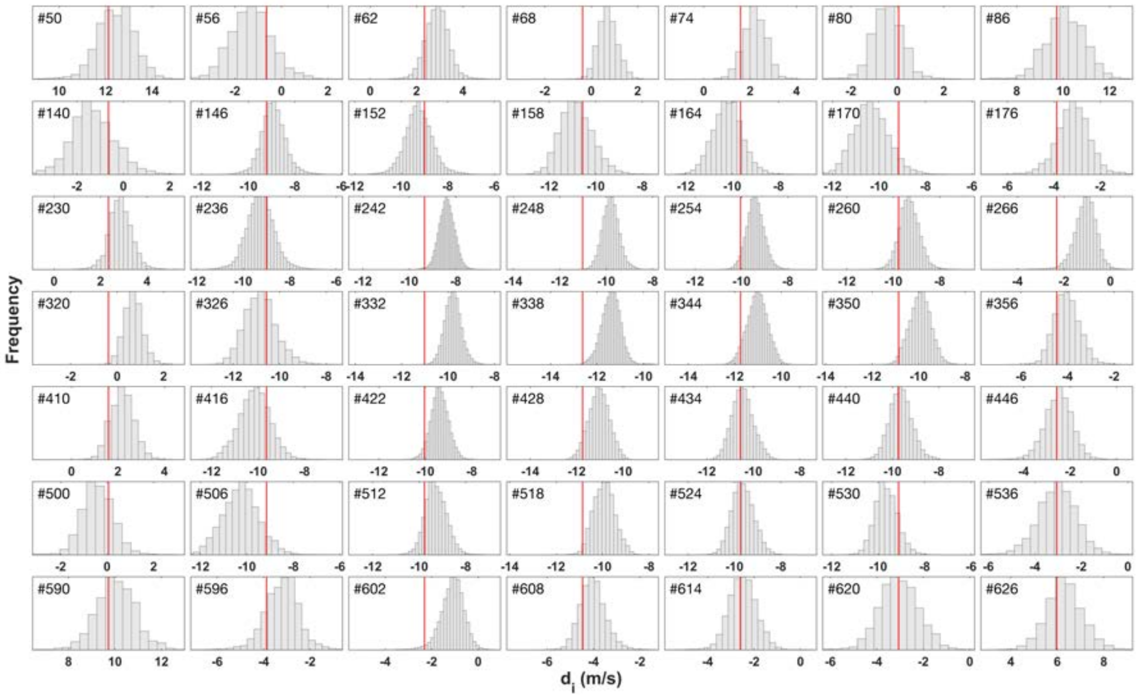

Figure A.1: Histograms of the predicted data values for 49 data points (out of 675). In this example, the synthetic observables are computed with a mesh of 20x20x40 finite elements. In each plot: synthetic data value (red line) and histogram of the predicted values. The burn-in period is discarded. All values of the x-axis should be multiplied by 10^{-13}. The plots are organized in the same order as shown in Figure 4.7.

It can be seen that although the mean values of the posterior PDFs do not always coincide with the true model, in almost all cases they are within one standard deviation of the posterior PDFs. However, to properly include and guarantee that the true values are within the posterior PDFs, modelization errors should be taken into account. In the example from Section 4.3, the fact that the same mesh has been used both for generating the synthetic observables and for predicting them is in accordance with the decision of neglecting modelization errors, since in such scenario the true model is achievable, i.e. it is within the interpolating space. In addition, as it has already been mentioned, guaranteeing that the synthetic observables are within the interpolating space allows us to better study the features and main factors controlling the coupling MCMC+RB. In other words, any additional source of error in the convergence is removed in order to isolate the MCMC+RB scheme and simplify its analysis.

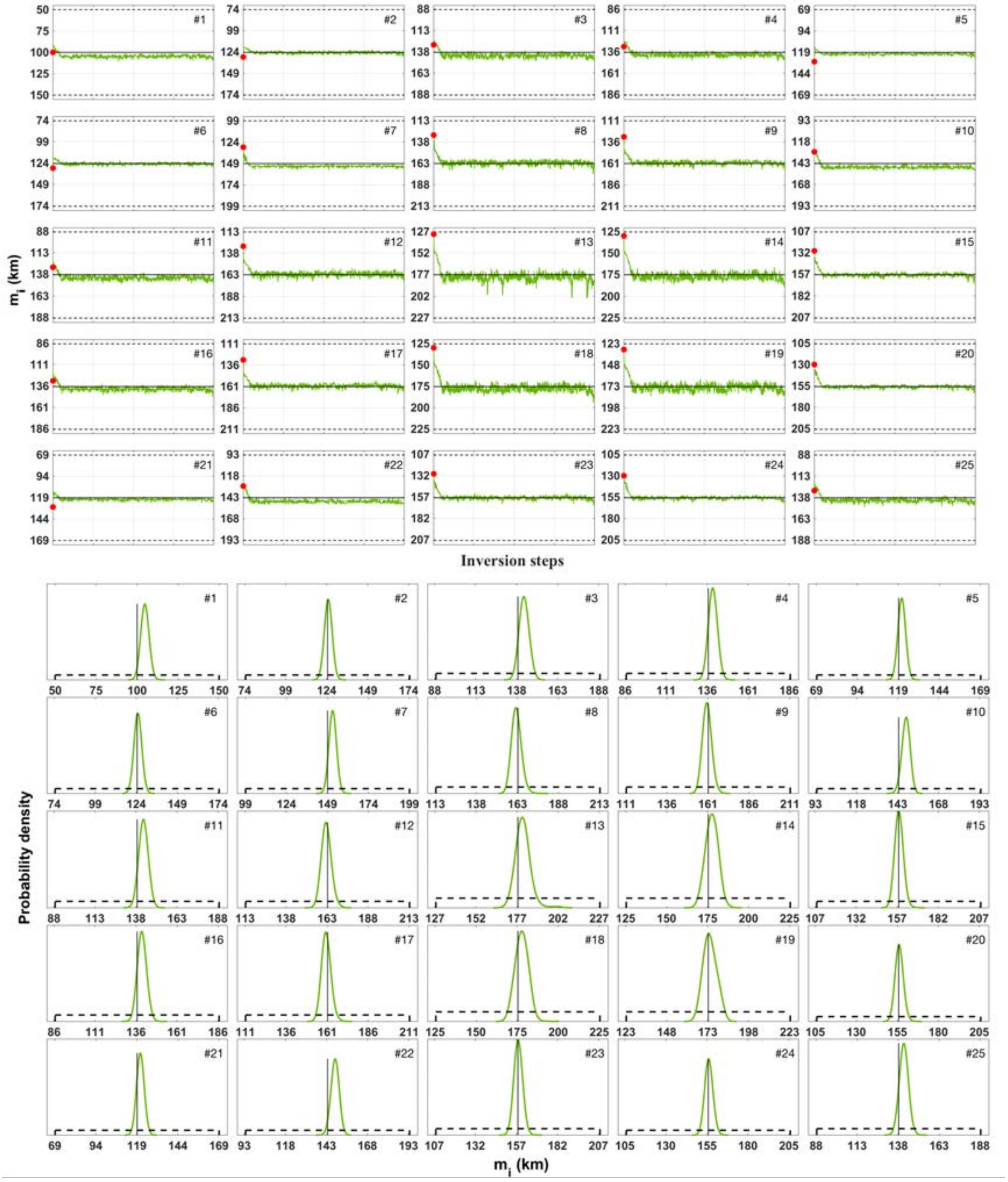

Figure A.2: (Top) random walk for each of the 25 parameters. (Bottom) posterior PDF for each of the 25 parameters. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot). The burn-in period has been discarded in the estimation of the posterior PDF.

A.2 Adding noise to the synthetic observables

In this additional example, the same MCMC+FE simulation as in Section 4.3 has been performed with the only difference that, in this case the noise or error is actually added to the synthetic observables. Figure A.3 shows the histograms of the predicted values for 49 of the 675 synthetic observables. All subplots are centered at the noisy synthetic observables (black line) and their x axis ranges from $-2\sigma_d$ to $+2\sigma_d$ (except for observable #356 whose bounds are $\pm 4\sigma_d$). The noiseless synthetic observables are represented by a vertical red line.

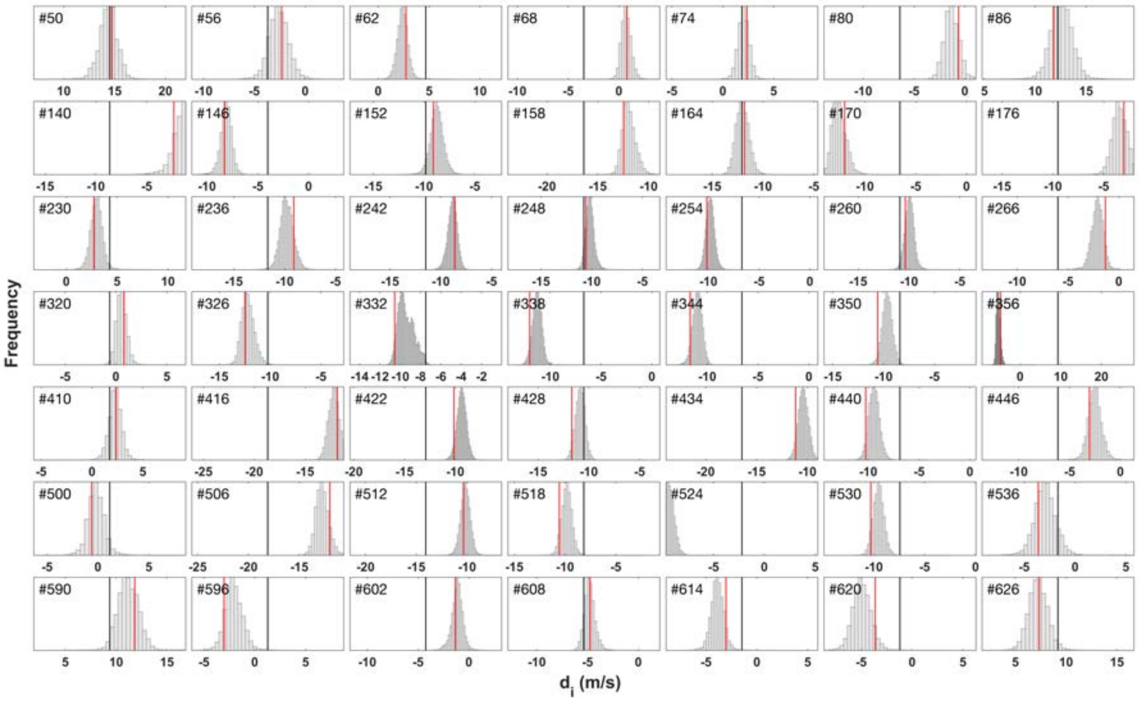

Figure A.3: Histograms of the predicted data values for 49 data points (out of 675). In this example, a noise is added to the synthetic data values. In each plot: synthetic data value (red line), perturbed synthetic data value (black line) and histogram of the predicted values. The burn-in period is discarded. All values of the x-axis should be multiplied by 10^{-13}. The plots are organized in the same order as shown in Figure 4.7.

In Figure A.3 it is seen that although the synthetic observables used as data in the inversion are perturbed (black line), the inversion predicts values which are closer to the unperturbed field (red line). The clearer case is the observable #356 whose predicted values coincide with the unperturbed synthetic and are almost $4\sigma_d$ shifted from the perturbed one. This situation can be explained by fact that the unperturbed velocity field is by far smoother than the perturbed one which is, due to the noise, hardly possible to be exactly predicted by any model of parameters. Therefore, since the introduced noise has zero mean, the general “structure” of the field is conserved in the perturbed observables and the inversion tends to predict it. It is a similar situation than Annex A.1 example in which the synthetic observables used in the inversion are outside the interpolating space. Figure A.4 shows the random walks and posterior PDFs obtained for each of the 25 parameters.

Since the simulation predicts values close to the unperturbed field, the random walks and posterior PDFs are not significantly affected by the noise addition. However, some small discrepancies appear in the mean posterior values compared with the true ones. Since our goal is to study relations and features in the MCMC+RB scheme, it is important to isolate them from any other factor. Therefore, we prefer to remove any possible additional source of error in the convergence of the random walks by not adding the noise to the synthetic observables and only consider data uncertainty in the likelihood with the σ_d.

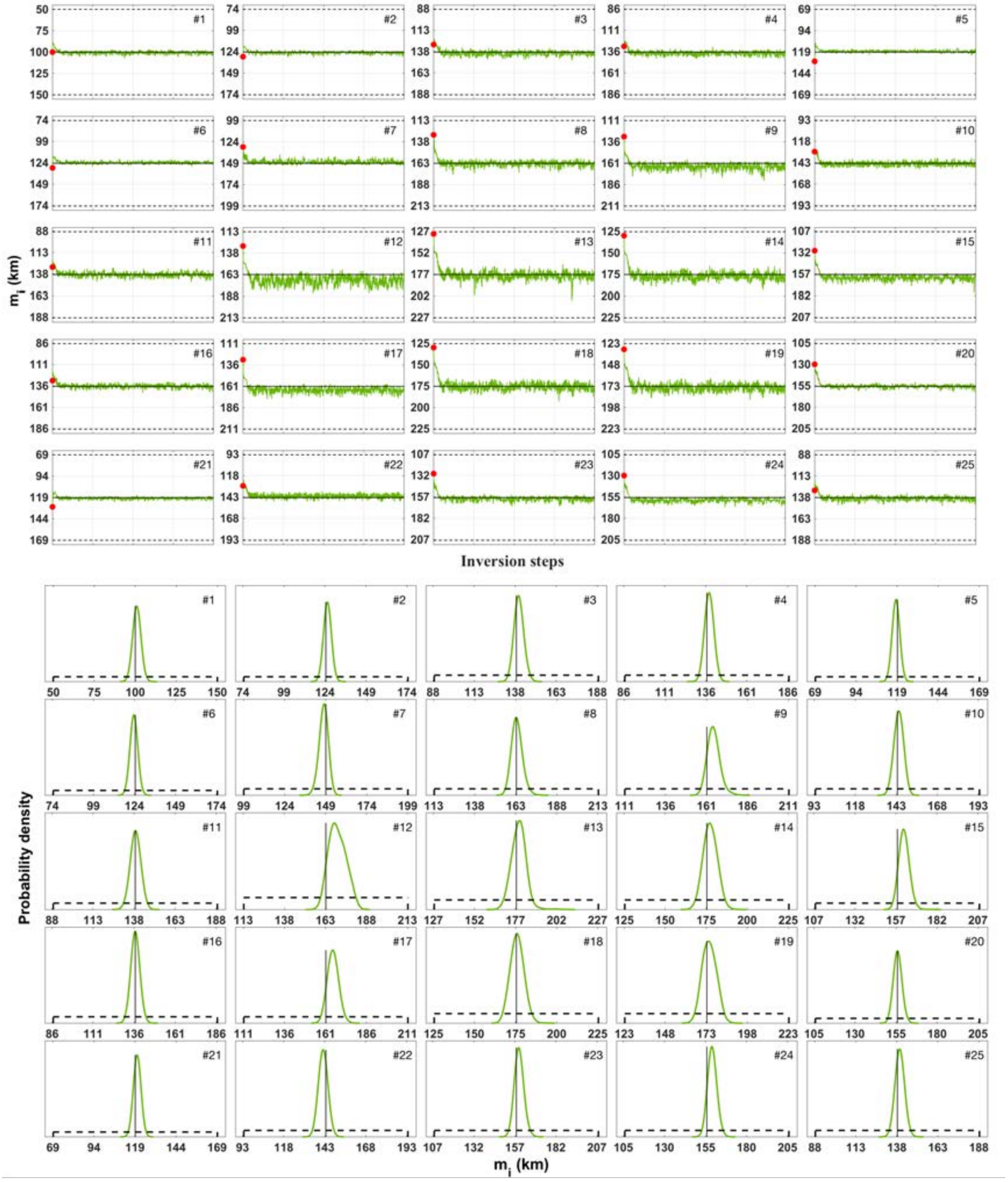

Figure A.4: (Top) random walk for each of the 25 parameters. (Bottom) posterior PDF for each of the 25 parameters. In each plot: prior distribution (black dashed line), reference LAB value used to generate the perturbed velocity field (solid black line) and initial value (red dot). The burn-in period has been discarded in the estimation of the posterior PDF.

A.3 Higher topography sensitivity

This additional example is meant to show that the low sensitivity of the topography is the main responsible of the compensation effect observed in examples A and B from Section 6.1. Here, the reference LAB structure from example B is artificially exaggerated, i.e. the same shape but larger differences between values. In doing so, the generated dynamic topography shows also larger differences which increase its sensitivity. In example B, the shallowest and deepest LAB values were 83km and 180km respectively, compared to the 67km and 208km shown here. Regarding topography, example B showed values ranging from -98m to 304m, whereas such range is now from -162m to 606m. All inversion settings are exactly the same as in example B, except that this one run for a much lower number of inversion, 2×10^6 (instead of 10^7). The burn-in period is 10^5

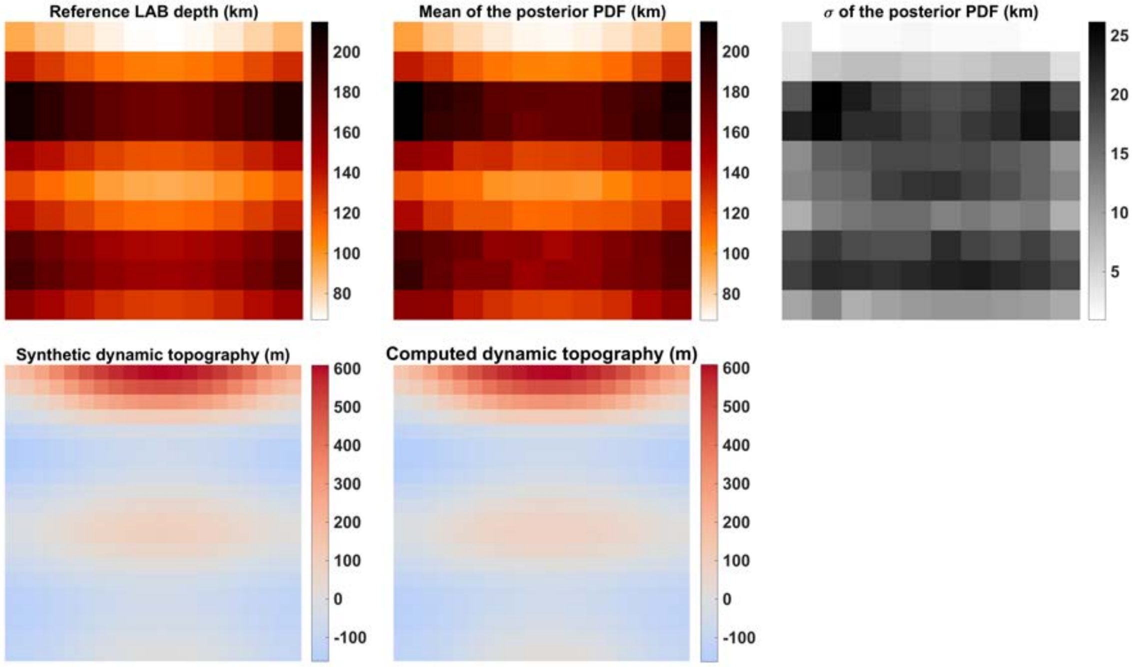

Figure A.5: Top view of the results for an inversion with a higher topography sensitivity. Top images represent the LAB structure: reference LAB depth (left), mean of the recovered posterior PDF for the LAB depth (center) and the corresponding standard deviation (right). Bottom images represent the topography associated for the LAB structure above: synthetic dynamic topography (left) and dynamic topography computed from the mean posterior (center).

Unlike Figure 6.5, the top center image from Figure A.5 shows no significant chessboard pattern which confirms that compensations had a lower effect. Even though again the same simple sampling strategy has been used, the now higher sensitivity of the observables together with their uncertainties is not allowing for such compensations. In other words, “wrong” configurations of model parameters are now affecting the misfit in a significant way.

Figures A.6 and A.7 show the random walks and posterior PDFs obtained for each of the 100 parameters in the inversion with a higher topography sensitivity.

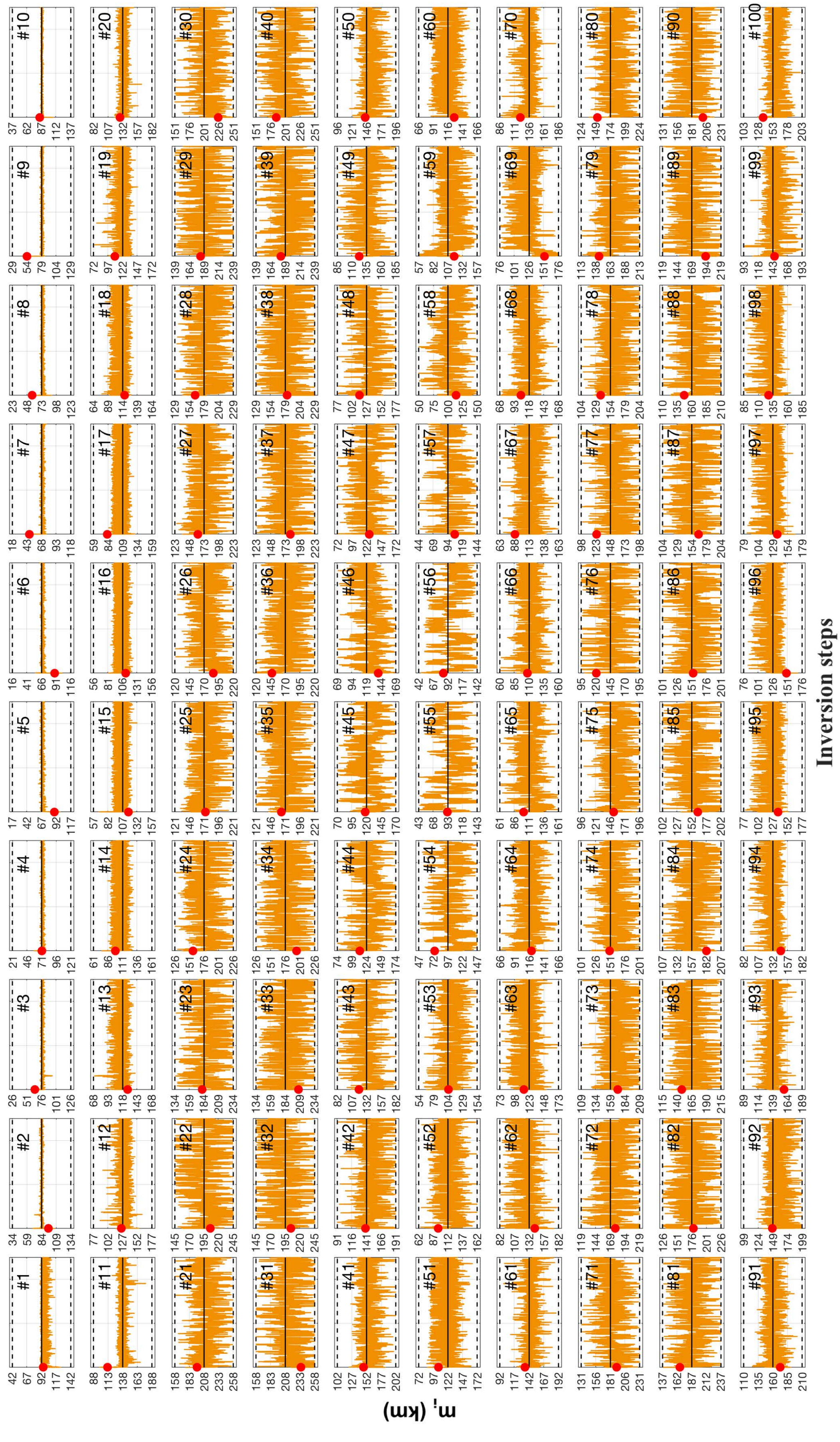

Figure A.6: Random walk for each of the 100 parameters for an inversion with a higher topography sensitivity. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot).

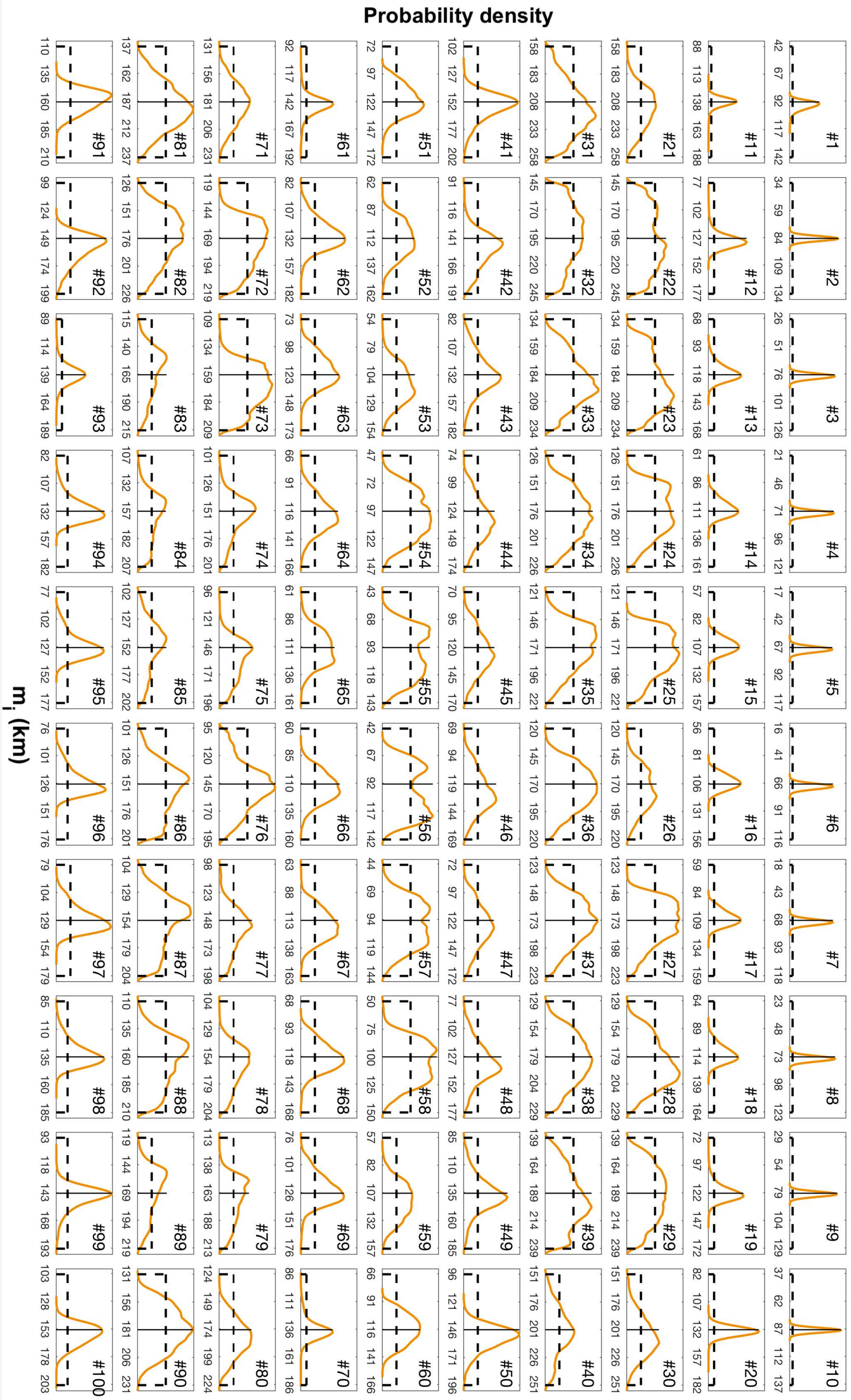

Figure A.7: Posterior PDF for each of the 100 parameters for an inversion with a higher topography sensitivity. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line). The burn-in period has been discarded in the estimation of the posterior PDF.

A.4 African lithospheric structure (225 parameters)

In this additional example, the same African lithospheric structure from Section 6.2 is now described with 225 parameters (instead of 1225). In this way, we are able to show detailed plots of each random walk and posterior PDF. Unless specifically stated, everything remains the same as described in Section 6.2.

The numerical domain is spherical and discretized with 30×30×20 finite elements from the surface down to a depth of 1000km depth. The vector of model parameters contains 225 elements describing the LAB depth (i.e. the FE discretization is twice as fine as the LAB one). The reference LAB and two vertical profiles of the generated velocity field are shown in Figure A.8.

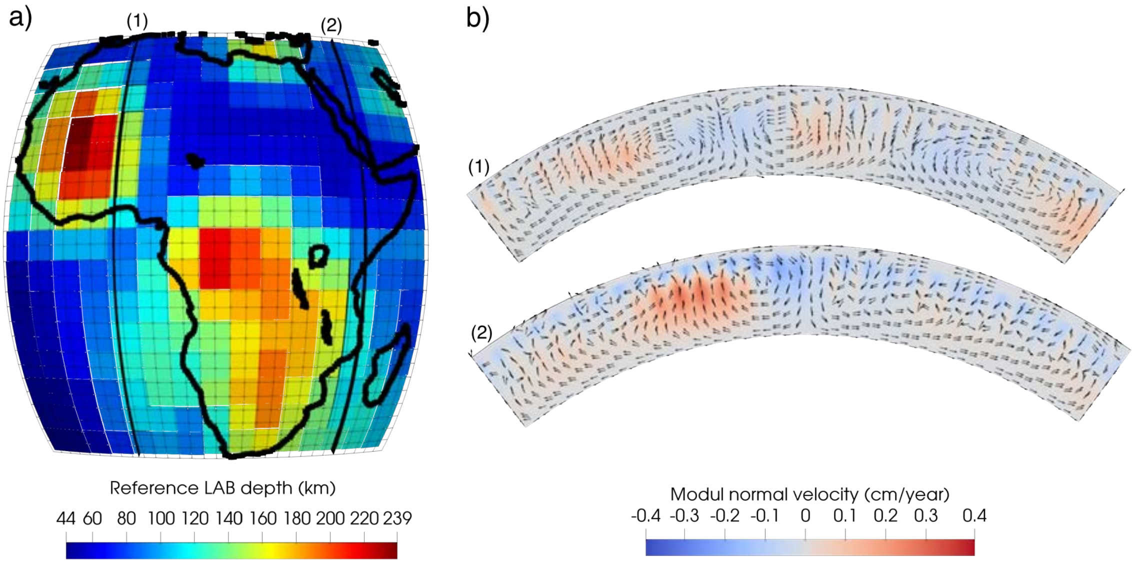

Figure A.8: (a) Reference LAB discretized with 225 parameters. (b) Two vertical profiles showing the flow (velocity) structure.

The input data are again the vertical velocities within the region of interest Γ which for this example is larger, ±100 km from the reference LAB value. The synthetic observables are contaminated with a Gaussian noise with a larger standard deviation, 15% ($\sigma_e = 0.15$). The initial LAB configuration is obtained by randomly perturbing the reference one between 10% and 20%. We run the inversion algorithm for 10^6 steps, from which the first 10^4 samples are discarded as burn-in. The standard deviation of the proposal distribution is set at $\sigma_p = 5$ km and the threshold for the error estimator of the RB is set at $e_{\mathrm{RB}} = 10^{-2}$. Mean values and standard deviations of the posterior PDFs resulting from the inversion are shown in Figure A.9. Plots of the random walks and posterior PDF for each of the 225 parameters are found in Figures A.11 and A.10.

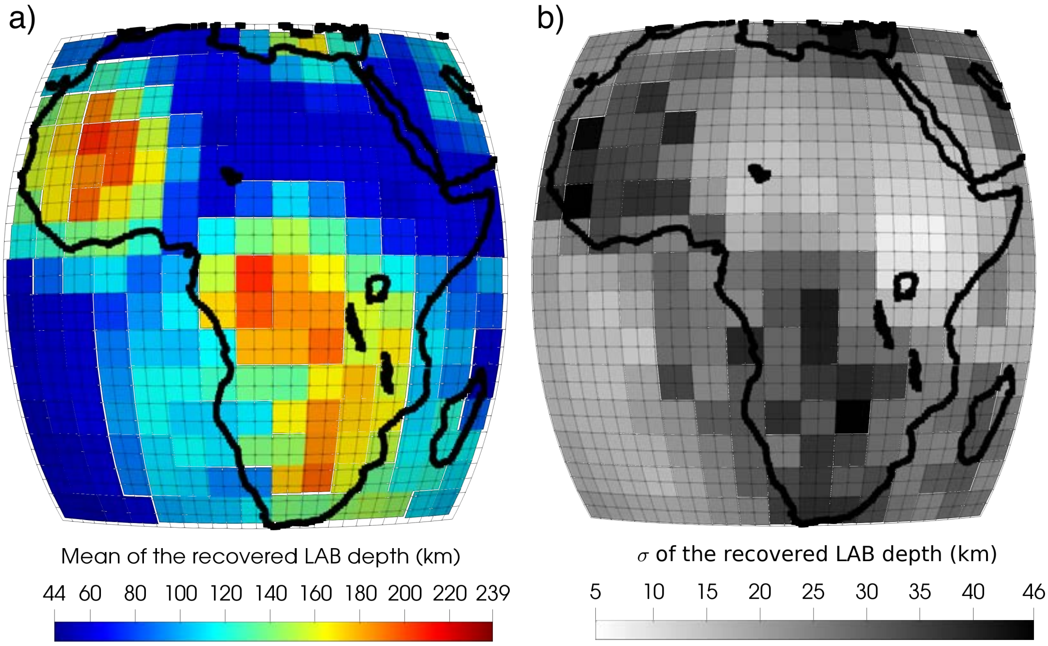

Figure A.9: (a) Mean and (b) standard deviation of the posterior PDF for each of the 225 parameters representing the LAB depth recovered during the inversion.

The inversion managed to recover the African LAB structure. Again wider posteriors are obtained in areas where the lithosphere is thicker since the velocity field is less sensitive to deep perturbations. The required number of FE solutions is 2360 to perform 10^6 inversion steps (0.24% of the total number of steps), or equivalently, one FE solution for every 400 inversions). The larger basis size with respect to the 1225 parameters case is due to the fact that the error of the data is also considerably larger for this example.

Figure A.10: Random walk for each of the 225 parameters in the Africa example. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot).

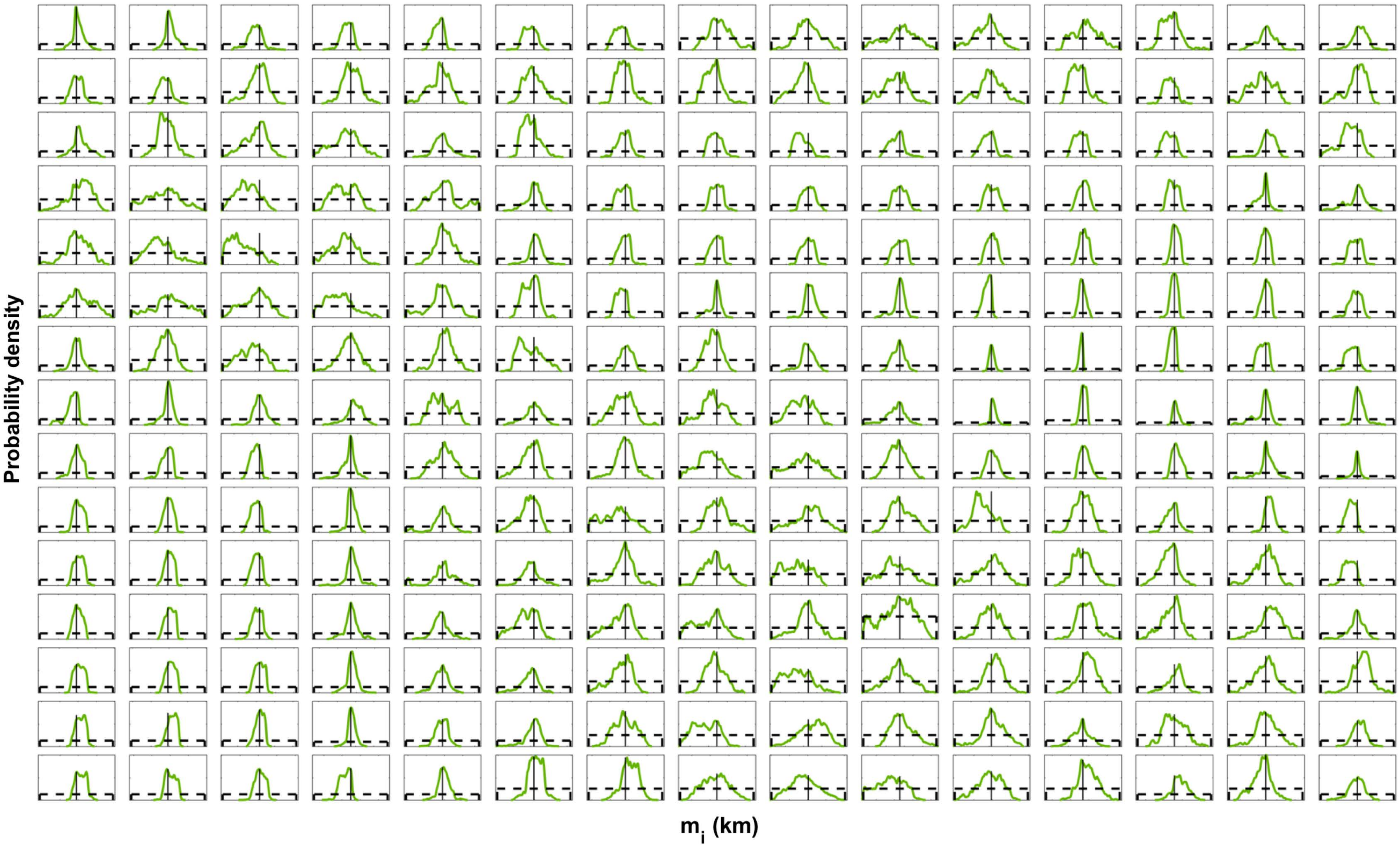

Figure A.11: Posterior PDF for each of the 225 parameters in the Africa example. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line). The burn-in period has been discarded in the estimation of the posterior PDF.

Appendix B

Posterior PDFs results

Table B.1: Mean and standard deviation of the posterior PDFs for each of the 25 parameters describing the LAB. Strategies MCMC+FE and MCMC+RB with e_{RB} values of 10^{-3}, 10^{-2} and 10^{-1}. Inversions run for 10^5 steps and the first 10^4 inversion steps are considered burn-in period and, hence, discarded.

		Mean (km)				σ (km)			
$\mathbf{m}_i$	true LAB	FE	energy-norm			FE	energy-norm		
			10^{-3}	10^{-2}	10^{-1}		10^{-3}	10^{-2}	10^{-1}
#1	100.0	100.6	100.5	100.5	100.5	1.6	1.6	1.6	1.6
#2	124.5	125.2	125.2	125.2	125.1	1.4	1.4	1.4	1.5
#3	138.8	139.6	139.5	139.8	139.5	2.5	2.3	2.3	2.2
#4	136.7	137.4	137.5	137.4	137.2	2.1	2.0	2.2	2.0
#5	119.2	119.5	119.5	119.5	119.5	1.2	1.2	1.3	1.2
#6	124.5	125.3	125.2	125.3	125.2	1.5	1.4	1.5	1.5
#7	149.1	148.5	148.3	148.5	148.4	2.0	2.0	2.0	2.0
#8	163.3	164.9	165.0	164.8	165.0	3.0	3.0	3.2	3.1
#9	161.2	162.8	162.5	162.7	162.7	2.9	2.5	2.7	2.9
#10	143.7	144.4	144.6	144.5	144.4	2.2	2.2	2.3	2.1
#11	138.8	139.7	139.4	139.8	139.5	2.4	2.2	2.4	2.3
#12	163.3	165.2	165.3	165.0	165.3	3.2	3.0	3.1	3.2
#13	177.5	181.0	179.9	181.1	180.1	5.8	4.7	6.5	4.9
#14	175.4	178.1	178.1	178.2	178.2	4.0	4.4	5.7	4.8
#15	157.9	159.2	159.1	159.2	159.1	2.3	2.0	2.5	2.2
#16	136.7	137.7	137.4	137.6	137.7	2.1	2.1	2.2	2.2
#17	161.2	162.4	162.4	162.4	162.1	2.7	2.6	2.9	2.8
#18	175.4	178.3	177.9	178.1	178.4	4.5	4.4	5.0	4.5
#19	173.4	176.6	176.3	176.7	176.3	4.3	3.9	4.5	4.1
#20	155.9	157.0	156.8	157.0	156.9	2.0	1.9	2.3	2.2
#21	119.2	119.5	119.5	119.5	119.4	1.2	1.2	1.2	1.2
#22	143.7	144.6	144.4	144.6	144.6	2.2	2.2	2.2	2.3
#23	157.9	159.2	159.1	159.3	159.1	2.2	2.1	2.5	2.3
#24	155.9	157.0	156.8	156.9	156.8	2.0	1.9	2.2	1.8
#25	138.4	139.7	139.5	139.5	139.5	2.3	2.2	2.4	2.4

Table B.2: Mean and standard deviation of the posterior PDFs for each of the 25 parameters describing the LAB. Strategies MCMC+FE and MCMC+RB with error indicators: energy-norm, and goal-oriented with linear and non-linear QoI. Inversions run for 10^5 steps and the first 10^4 inversion steps are considered burn-in period and, hence, discarded.

		Mean (km)				σ (km)			
$\mathbf{m}_i$	true LAB	FE	energy norm	QoI (linear)	QoI (nonlinear)	FE	energy norm	QoI (linear)	QoI (nonlinear)
#1	100.0	100.6	100.5	100.5	100.5	1.6	1.6	1.6	1.6
#2	124.5	125.2	125.2	125.2	125.3	1.4	1.4	1.4	1.4
#3	138.8	139.6	139.8	139.8	139.6	2.5	2.3	2.6	2.4
#4	136.7	137.4	137.4	137.5	137.6	2.1	2.2	2.2	2.0
#5	119.2	119.5	119.5	119.5	119.6	1.2	1.3	1.3	1.1
#6	124.5	125.3	125.3	125.2	125.3	1.5	1.5	1.4	1.4
#7	149.1	148.5	148.5	148.4	148.5	2.0	2.0	2.1	1.9
#8	163.3	164.9	164.8	164.7	165.1	3.0	3.2	3.1	3.1
#9	161.2	162.8	162.7	162.6	162.8	2.9	2.7	2.9	2.7
#10	143.7	144.4	144.5	144.2	144.4	2.2	2.3	2.2	2.2
#11	138.8	139.7	139.8	139.5	139.9	2.4	2.4	2.2	2.4
#12	163.3	165.2	165.0	165.3	165.4	3.2	3.1	3.3	3.0
#13	177.5	181.0	181.1	180.3	180.9	5.8	6.5	5.0	5.4
#14	175.4	178.1	178.2	177.8	178.8	4.0	5.7	4.9	4.3
#15	157.9	159.2	159.2	159.2	159.2	2.3	2.5	2.3	2.2
#16	136.7	137.7	137.6	137.6	137.7	2.1	2.2	2.2	2.0
#17	161.2	162.4	162.4	162.5	162.4	2.7	2.9	2.7	2.6
#18	175.4	178.3	178.1	177.9	179.1	4.5	5.0	4.9	4.7
#19	173.4	176.6	176.7	176.5	176.7	4.3	4.5	4.7	4.0
#20	155.9	157.0	157.0	156.9	157.0	2.0	2.3	2.0	2.1
#21	119.2	119.5	119.5	119.5	119.5	1.2	1.2	1.2	1.3
#22	143.7	144.6	144.6	144.4	144.8	2.2	2.2	2.1	2.2
#23	157.9	159.2	159.3	159.2	159.3	2.2	2.5	2.4	2.2
#24	155.9	157.0	156.9	156.8	157.1	2.0	2.2	2.0	1.9
#25	138.4	139.7	139.5	139.4	139.7	2.3	2.4	2.5	2.2

Table B.3: Means and standard deviations of the posterior PDF for each of the 100 parameters recovered in the inversion example in Section 5.3.1 Values of the true LAB used to generate the synthetic velocity field are also included. Inversion run for 10^6 steps and the first 10^5 inversion steps are considered burn-in period and, hence, discarded.

$\mathbf{m}_i$	true LAB	Mean	σ
#1	100.0	101.3	3.2
#2	111.6	111.9	2.9
#3	122.2	122.4	1.8
#4	130.8	131.7	2.7
#5	136.9	138.5	3.5
#6	139.8	140.9	3.9
#7	139.3	140.1	3.8
#8	135.4	137.2	3.5
#9	128.5	128.9	2.8
#10	119.2	119.5	2.1
#11	111.6	111.2	3.1
#12	123.1	123.4	1.8
#13	133.7	134.8	2.9
#14	142.4	143.4	3.4
#15	148.5	147.3	3.3
#16	151.4	151.8	2.4
#17	150.9	151.4	2.6
#18	147.0	145.9	3.5
#19	140.1	141.8	3.8
#20	130.7	131.1	3.2
#21	122.2	122.4	1.9
#22	133.7	134.8	2.8
#23	144.3	144.6	3.2
#24	153.0	153.6	2.3
#25	159.0	159.9	2.6
#26	161.9	163.0	3.3
#27	161.4	162.5	3.4
#28	157.6	158.9	2.6
#29	150.7	150.3	2.6
#30	141.3	143.2	3.8
#31	130.8	131.7	2.9
#32	142.4	143.4	3.3
#33	153.0	153.7	2.1
#34	161.7	162.8	3.1
#35	167.7	169.2	4.0
#36	170.6	173.5	4.8
#37	170.1	172.8	4.7
#38	166.3	167.7	4.3
#39	159.4	160.8	3.2
#40	150.0	149.4	2.9
#41	136.9	138.5	3.6
#42	148.5	146.9	3.3
#43	159.0	159.8	2.8
#44	167.7	169.2	4.2
#45	173.8	177.2	4.6
#46	176.7	179.6	4.7
#47	176.2	178.7	4.9
#48	172.3	175.3	5.2
#49	165.4	167.0	4.7
#50	156.1	157.3	2.8
#51	139.8	141.2	3.7
#52	151.4	151.6	2.5
#53	161.9	162.9	3.6
#54	170.6	173.9	4.7
#55	176.7	178.9	4.6
#56	179.6	182.1	4.8
#57	179.1	181.0	4.6
#58	175.2	178.1	4.9
#59	168.3	170.8	4.9
#60	159.0	160.2	3.2
#61	139.3	140.3	3.7
#62	150.9	151.4	2.4
#63	161.4	162.6	3.3
#64	170.1	172.7	5.0
#65	176.2	179.4	5.0
#66	179.1	180.4	5.5
#67	178.6	181.3	5.4
#68	174.7	177.2	5.1
#69	167.8	170.6	4.8
#70	158.5	159.6	3.4
#71	135.4	137.1	3.8
#72	147.0	146.1	3.3
#73	157.6	158.7	2.5
#74	166.3	168.2	4.0
#75	172.3	175.0	4.9
#76	175.2	178.2	4.8
#77	174.7	176.7	5.3
#78	170.8	174.8	5.3
#79	163.9	165.3	4.1
#80	154.6	156.0	2.7
#81	128.5	129.3	2.9
#82	140.1	141.7	3.8
#83	150.7	150.4	2.6
#84	159.4	160.4	3.0
#85	165.4	167.6	4.1
#86	168.3	170.7	4.8
#87	167.8	171.3	4.8
#88	163.9	165.2	4.4
#89	157.0	158.8	3.0
#90	147.7	146.4	3.3
#91	119.2	119.2	2.3
#92	130.7	132.0	3.4
#93	141.3	142.8	3.5
#94	150.0	149.8	2.8
#95	156.1	156.7	2.6
#96	159.0	160.4	3.4
#97	158.5	159.4	3.3
#98	154.6	156.0	3.0
#99	147.7	146.3	3.6
#100	138.4	141.6	4.2

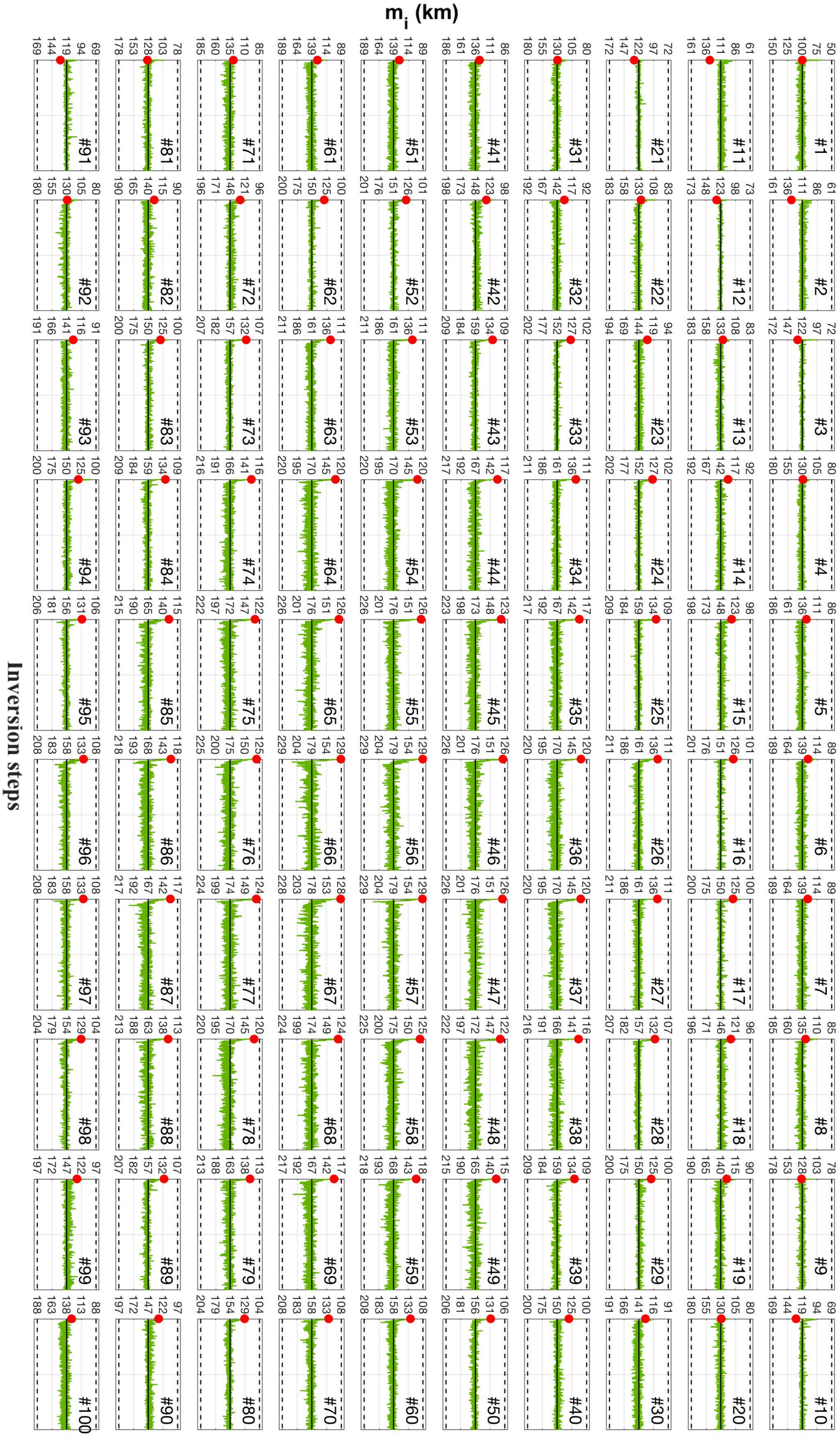

Figure B.1: Random walk for each of the 100 parameters. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot).

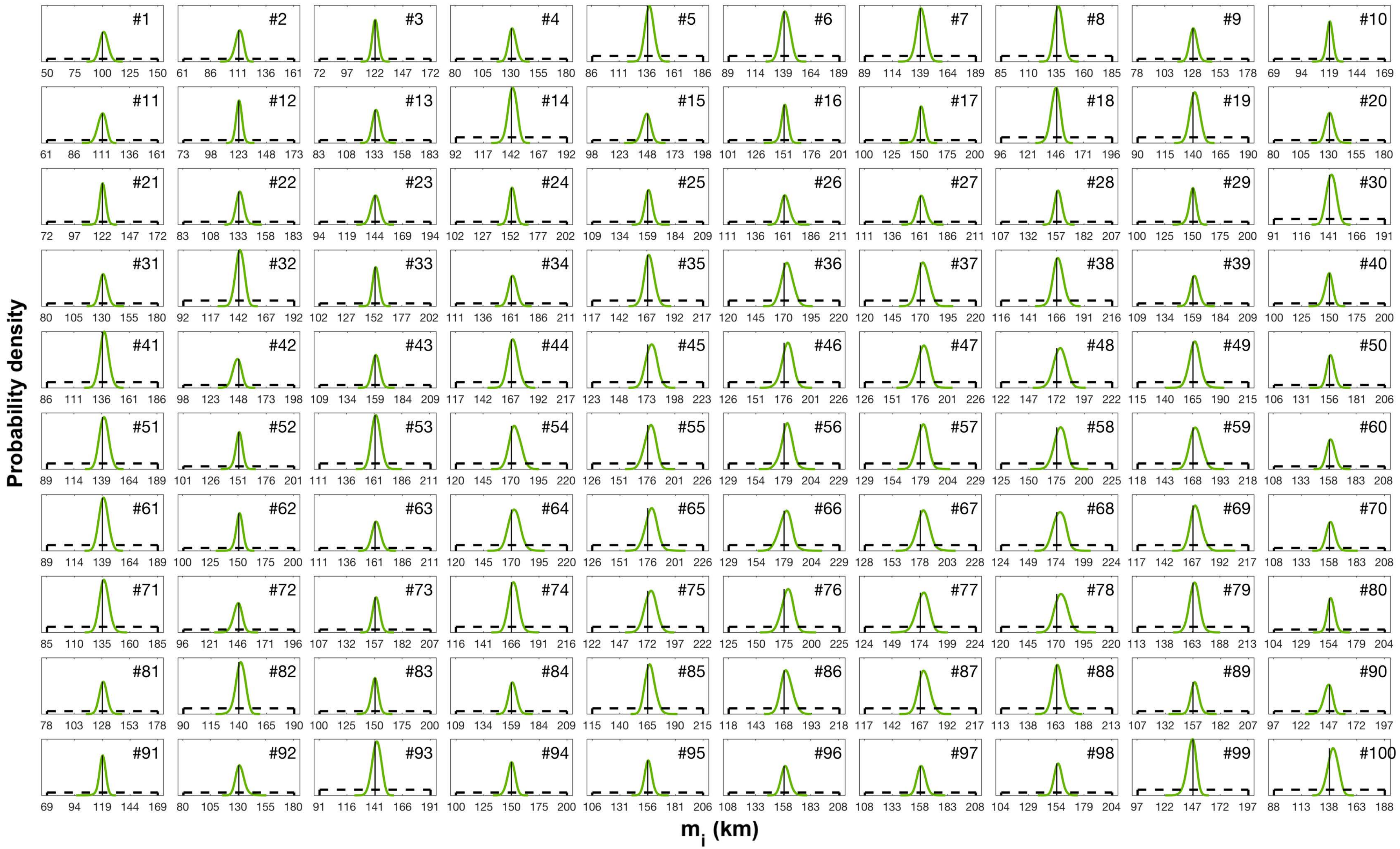

Figure B.2: Posterior PDF for each of the 100 parameters. In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line). The burn-in period has been discarded in the estimation of the posterior PDF.

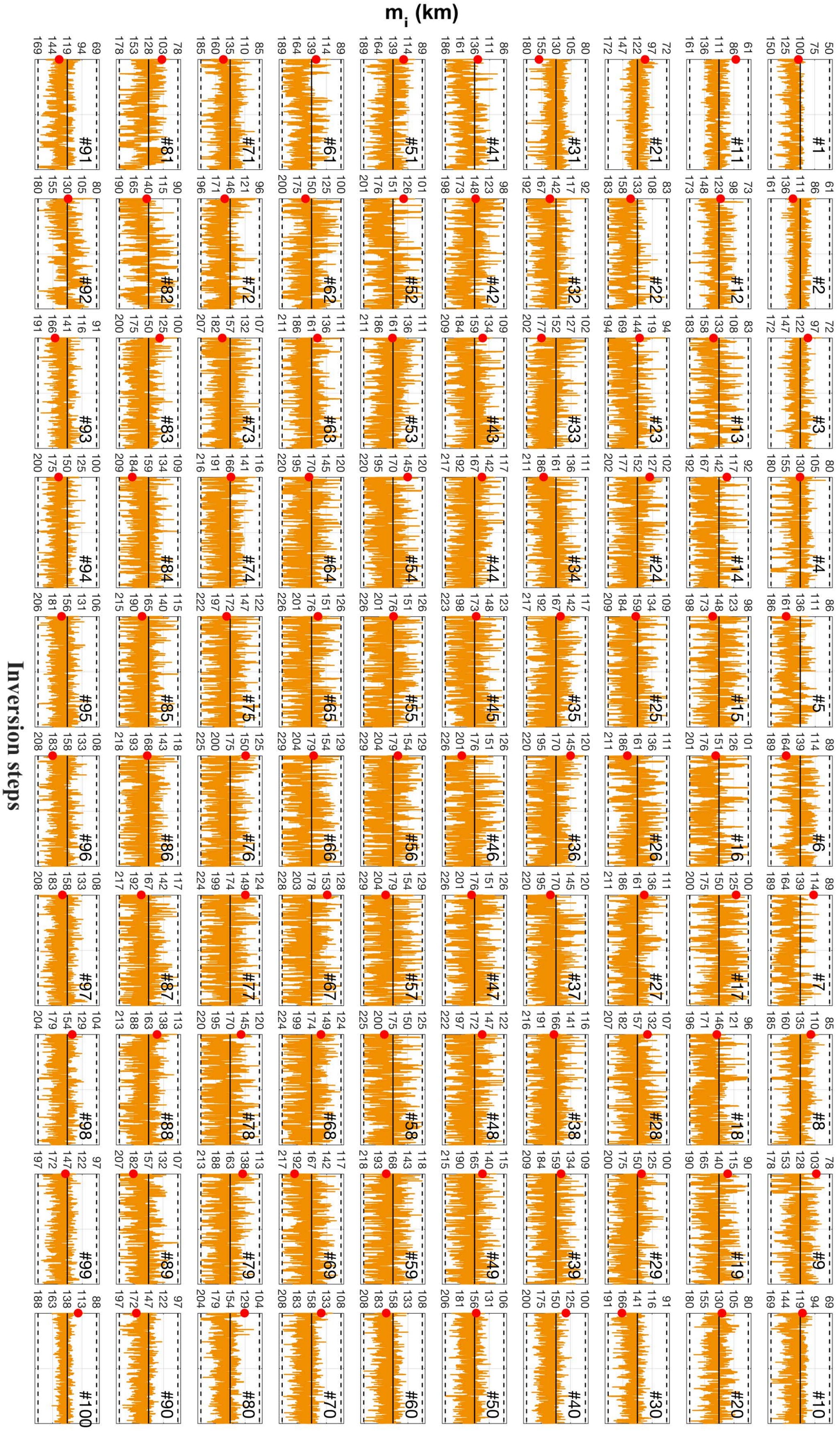

Figure B.3: Random walk for each of the 100 parameters using topography as observables (example A). In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot).

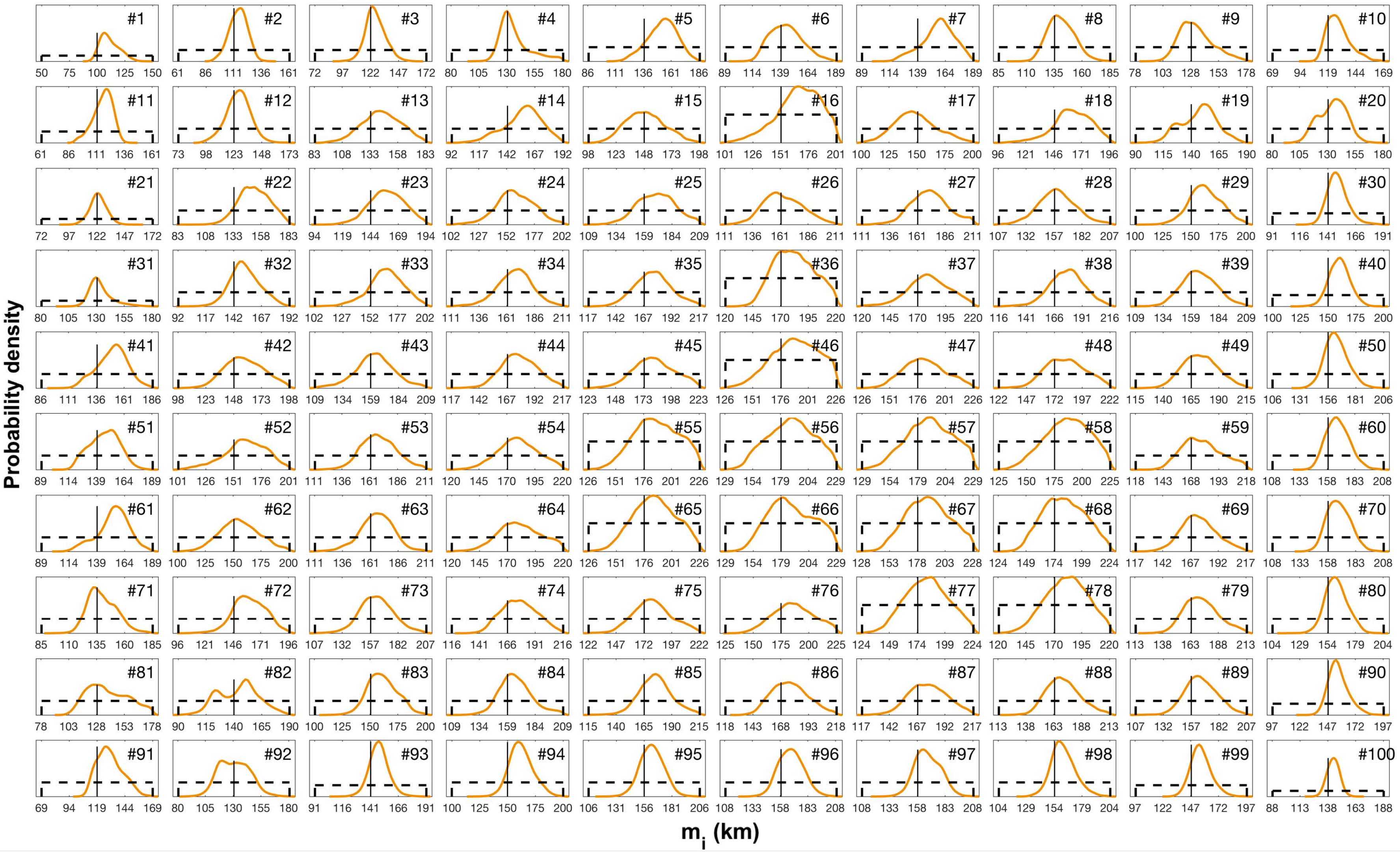

Figure B.4: Posterior PDF for each of the 100 parameters using topography as observables (example A). In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line). The burn-in period has been discarded in the estimation of the posterior PDF.

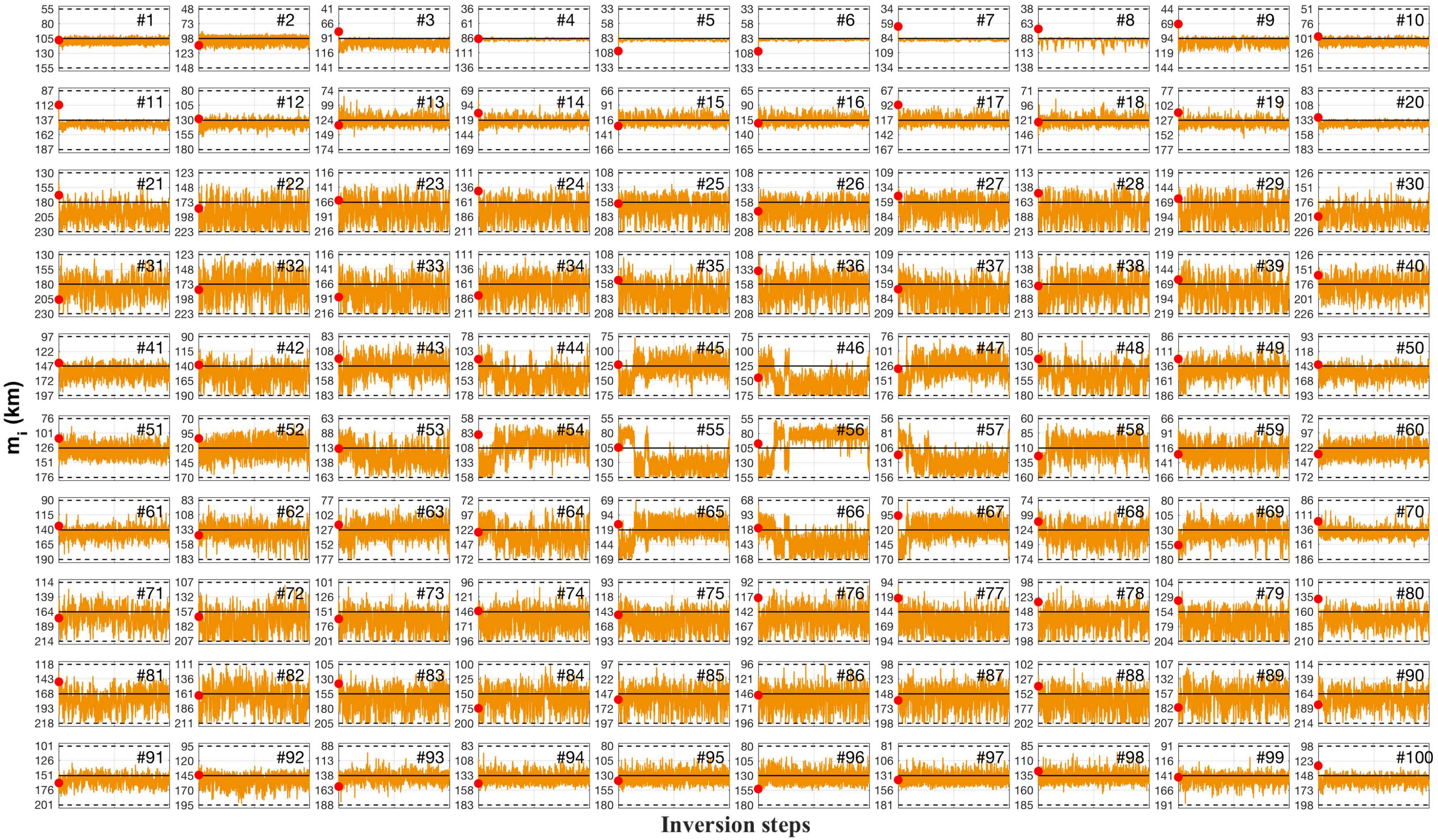

Figure B.5: Random walk for each of the 100 parameters using topography as observables (example B). In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line) and initial value (red dot).

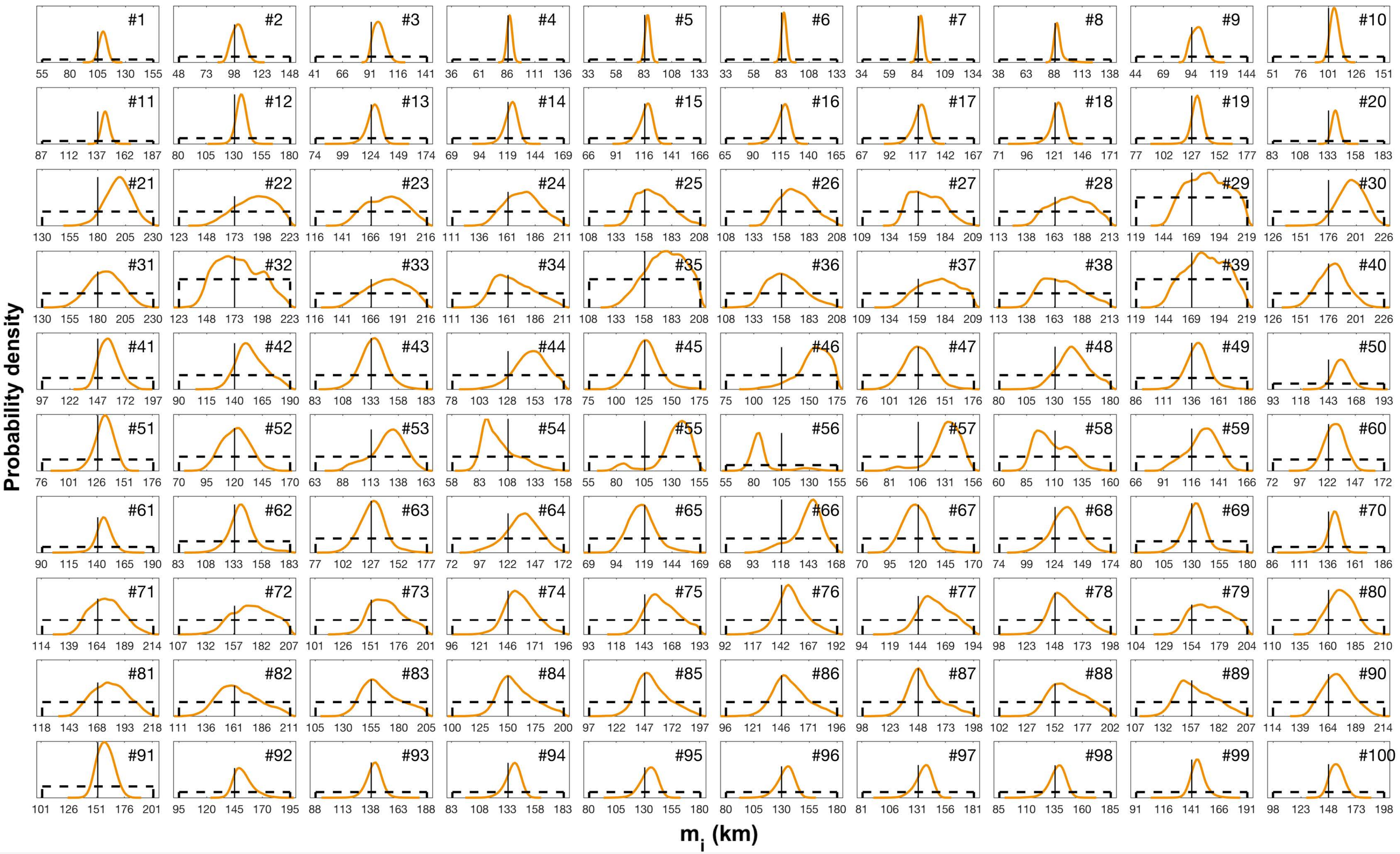

Figure B.6: Posterior PDF for each of the 100 parameters using topography as observables (example B). In each plot: prior distribution (black dashed line), reference LAB value used to generate the synthetic velocity field (solid black line). The burn-in period has been discarded in the estimation of the posterior PDF.

Bibliography

Afonso, J. C., Fernàndez, M., Ranalli, G., Griffin, W. L., and Connolly, J. A. D. (2008). Integrated geophysical-petrological modeling of the lithosphere and sublithospheric upper mantle: Methodology and applications. *Geochemistry, Geophysics, Geosystems*, 9(5).

Afonso, J. C., Fullea, J., Griffin, W. L., Yang, Y., Jones, A. G., Connolly, J. A. D., and O'Reilly, S. Y. (2013a). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. I: a priori petrological information and geophysical observables. *Journal of Geophysical Research: Solid Earth*, 118(5):2586–2617.

Afonso, J. C., Fullea, J., Yang, Y., Connolly, J. A. D., and Jones, A. G. (2013b). 3-D multi-observable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. II: General methodology and resolution analysis. *Journal of Geophysical Research: Solid Earth*, 118(4):1650–1676.

Afonso, J. C., Moorkamp, M., and Fullea, J. (2016a). Imaging the Lithosphere and Upper Mantle: Where We Are At and Where We Are Going. In *Integrated Imaging of the Earth: Theory and Applications*, pages 191–218. John Wiley & Sons, Inc.

Afonso, J. C., Rawlinson, N., Yang, Y., Schutt, D. L., Jones, A. G., Fullea, J., and Griffin, W. L. (2016b). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle: III. Thermochemical tomography in the Western-Central U.S. *Journal of Geophysical Research: Solid Earth*, 121(10):7337–7370.

Afonso, J. C., Salajegheh, F., Szwillus, W., Ebbing, J., and Gaina, C. (2019). A global reference model of the lithosphere and upper mantle from joint inversion and analysis of multiple data sets. *Geophysical Journal International*, 217(3):1602–1628.

Aguado, J. V., Huerta, A., Chinesta, F., and Cueto, E. (2014). Real-time monitoring of thermal processes by reduced-order modeling. *International Journal for Numerical Methods in Engineering*, 102(5):991–1017.

Anderson, D. L. (1989). *Theory of the Earth.* Blackwell Scientific Publications.

Arridge, S. R., Kaipio, J. P., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., and Vauhkonen, M. (2006). Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems*, 22(1):175–195.

Batchelor, G. K. (2002). *An Introduction to Fluid Dynamics.* Cambridge University Press.

Baumann, T. S. and Kaus, B. J. (2015). Geodynamic inversion to constrain the non-linear rheology of the lithosphere. *Geophysical Journal International*, 202(2):1289–1316.

Baumann, T. S., Kaus, B. J., and Popov, A. A. (2014). Constraining effective rheology through parallel joint geodynamic inversion. *Tectonophysics*, 631(C):197–211.

Benner, P., Gugercin, S., and Willcox, K. (2015). A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. *SIAM Review*, 57(4):483–531.

Berger, J., Orlande, H. R. B., and Mendes, N. (2016). Proper Generalized Decomposition model reduction in the Bayesian framework for solving inverse heat transfer problems. *Inverse Problems in Science and Engineering*, 25(2):260–278.

Blankenbach, B., Busse, F., Christensen, U., Cserepes, L., Gunkel, D., Hansen, U., Harder, H., Jarvis, G. T., Koch, M., Marquart, G., Moore, D., Olson, P., Schmeling, H., and Schnaubelt, T. (1989). A benchmark comparison for mantle convection codes. *Geophysical Journal International*, 98(1):23–38.

Braun, J., Thieulot, C., Fullsack, P., DeKool, M., Beaumont, C., and Huismans, R. (2008). DOUAR: A new three-dimensional creeping flow numerical model for the solution of geological problems. *Physics of the Earth and Planetary Interiors*, 171(1-4):76–91.

Brezzi, F. and Fortin, M. (1991). *Mixed and Hybrid Finite Element Methods*, volume 15 of *Springer Series in Computational Mechanics.* Springer-Verlag, New York.

Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Brunton, S. L. and Kutz, J. N. (2019). *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems and Control.* Cambridge University Press.

Bui-Thanh, T., Willcox, K., and Ghattas, O. (2008). Model Reduction for Large-Scale Systems with High-Dimensional Parametric Input Space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288.

Chinesta, F., Keunings, R., and Leygue, A. (2014). *The Proper Generalized Decomposition for Advanced Numerical Simulations*. SpringerBriefs in Applied Sciences and Technology. Springer International Publishing, Cham.

Chinesta, F. and Ladevèze, P. (2014). *Separated Representations and PGD-Based Model Reduction*, volume 554 of *CISM International Centre for Mechanical Sciences*. Springer Vienna, Vienna.

Christen, J. A. and Fox, C. (2005). Markov chain Monte Carlo Using an Approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810.

Cochran, W. G. (1977). *Sampling Techniques*. John Wiley & Sons, Inc., third edition.

Crameri, F., Schmeling, H., Golabek, G. J., Duretz, T., Orendt, R., Buiter, S. J. H., May, D. A., Kaus, B. J. P., Gerya, T. V., and Tackley, P. J. (2012). A comparison of numerical surface topography calculations in geodynamic modelling: an evaluation of the 'sticky air' method. *Geophysical Journal International*, 189:38–54.

Cui, T., Fox, C., and O'Sullivan, M. J. (2011). Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resources Research*, 47(10).

Cui, T., Marzouk, Y., and Willcox, K. (2014). Data-driven model reduction for the Bayesian solution of inverse problems. *International Journal for Numerical Methods in Engineering*, 102(5):966–990.

Donea, J. and Huerta, A. (2003). *Finite Element Methods for Flow Problems*. John Wiley & Sons, Ltd.

Efendiev, Y., Hou, T., and Luo, W. (2006). Preconditioning Markov Chain Monte Carlo Simulations Using Coarse-Scale Models. *SIAM Journal on Scientific Computing*, 28(2):776–803.

Florentin, E. and Díez, P. (2012). Adaptive reduced basis strategy based on goal oriented error assessment for stochastic problems. *Computer Methods in Applied Mechanics and Engineering*, 225-228:116–127.

Forte, A. M. (2007). Constraints on seismic models from other disciplines – implications for mantle dynamics and composition. In *Treatise on Geophysics*, pages 805 – 858. Elsevier.

Forte, A. M., Woodward, R. L., and Dziewonski, A. M. (1994). Joint inversions of seismic and geodynamic data for models of three-dimensional mantle heterogeneity. *Journal of Geophysical Research: Solid Earth*, 99(B11):21857–21877.

Frangos, M., Marzouk, Y. M., Willcox, K., and van Bloemen Waanders, B. (2010). Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems. In *Large-Scale Inverse Problems and Quantification of Uncertainty*, pages 123–149. John Wiley & Sons, Ltd, Chichester, UK.

Fullea, J., Afonso, J. C., Connolly, J. A. D., Fernàndez, M., García-Castellanos, D., and Zeyen, H. (2009). LitMod3D: An interactive 3-D software to model the thermal, compositional, density, seismological, and rheological structure of the lithosphere and sublithospheric upper mantle. *Geochemistry, Geophysics, Geosystems*, 10(8):1–21.

Fullsack, P. (1995). An arbitrary Lagrangian-Eulerian formulation for creeping flows and its application in tectonic models. *Geophysical Journal International*, 120(1):1–23.

Galbally, D., Fidkowski, K., Willcox, K., and Ghattas, O. (2009). Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International Journal for Numerical Methods in Engineering*, 81(12):1581–1608.

García-Blanco, R., Borzacchiello, D., Chinesta, F., and Diez, P. (2017). Monitoring a PGD solver for parametric power flow problems with goal-oriented error assessment. *International Journal for Numerical Methods in Engineering*, 111(6):529–552.

García-Blanco, R., Díez, P., Borzacchiello, D., and Chinesta, F. (2018). Algebraic and Parametric Solvers for the Power Flow Problem: Towards Real-Time and Accuracy-Guaranteed Simulation of Electric Systems. *Archives of Computational Methods in Engineering*, 25(4):1003–1026.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis Jumping Rules. *Bayesian Statistics*, 5:599–607.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequence. *Statistical Science*, 7(4):457–511.

Gerstner, T. and Griebel, M. (1998). Numerical integration using sparse grids. *Numerical Algorithms*, 18(3/4):209–232.

Geweke, J. F. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Staff Report 148, Federal Reserve Bank of Minneapolis.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Springer Science+Business Media, B.V.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press.

Grepl, M. A. and Patera, A. T. (2005). A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 39(1):157–181.

Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14(3):375.

Haario, H., Saksman, E., and Tamminen, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli*, 7(2):223–242.

Haario, H., Saksman, E., and Tamminen, J. (2005). Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20(2):265–273.

Harlow, F. H. and Welch, J. E. (1965). Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. *Physics of Fluids*, 8(12):2182–2189.

Haskell, N. A. (1935). The motion of a viscous fluid under a surface load. *Physics*, 6:265–269.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their Applications. *Biometrika*, 57(1):97–109.

Hess, M. W. and Benner, P. (2014). A reduced basis method for microwave semiconductor devices with geometric variations. *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 33(4):1071–1081.

Hesthaven, J. S., Rozza, G., and Stamm, B. (2016). *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. SpringerBriefs in Mathematics. Springer International Publishing, Cham.

Hesthaven, J. S., Stamm, B., and Zhang, S. (2014). Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(1):259–283.

Hughes, T. J. R. (1987). *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Prentice-Hall.

Ito, K. and Ravindran, S. (1998). A Reduced-Order Method for Simulation and Control of Fluid Flows. *Journal of Computational Physics*, 143(2):403–425.

Kaipio, J. P. and Somersalo, E. (2005). *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York.

Kaipio, J. P. and Somersalo, E. (2007). Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian Calibration of Computer Models. *Society*, 63(3):425–464.

Khan, A., Connolly, J. A. D., and Taylor, S. R. (2008). Inversion of seismic and geodetic data for the major element chemistry and temperature of the Earth's mantle. *Journal of Geophysical Research*, 113(B9):B09308.

Khan, A., Zunino, A., and Deschamps, F. (2011). The thermo-chemical and physical structure beneath the North American continent from Bayesian inversion of surface-wave phase velocities. *Journal of Geophysical Research*, 116(B9):B09304.

Kolehmainen, V., Schweiger, M., Nissilä, I., Tarvainen, T., Arridge, S. R., and Kaipio, J. P. (2009). Approximation errors and model reduction in three-dimensional diffuse optical tomography. *Journal of the Optical Society of America A*, 26(10):2257.

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49(5):2664–2682.

Larion, Y., Zlotnik, S., Massart, T. J., and Díez, P. (2020). Building a certified reduced basis for coupled thermo-hydro-mechanical systems with goal-oriented error estimation. *Computational Mechanics*.

Li, J. and Marzouk, Y. M. (2014). Adaptive Construction of Surrogates for the Bayesian Solution of Inverse Problems. *SIAM Journal on Scientific Computing*, 36(3):A1163–A1186.

Lieberman, C., Willcox, K., and Ghattas, O. (2010). Parameter and State Model Reduction for Large-Scale Statistical Inverse Problems. *SIAM Journal on Scientific Computing*, 32(5):2523–2542.

Linde, N., Ginsbourger, D., Irving, J., Nobile, F., and Doucet, A. (2017). On uncertainty quantification in hydrogeology and hydrogeophysics. *Advances in Water Resources*, 110(October):166–181.

Lithgow-Bertelloni, C. and Silver, P. G. (1998). Dynamic topography, plate driving forces and the African superswell. *Nature*, 395(6699):269–272.

Maday, Y. and Stamm, B. (2013). Locally Adaptive Greedy Approximations for Anisotropic Parameter Reduced Basis Spaces. *SIAM Journal on Scientific Computing*, 35(6):A2417–A2441.

Malinverno, A. and Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, 69(4):1005–1016.

Manassero, M. C. (2019). *A Reduced Order Approach for Probabilistic Inversions of 3D Magnetotelluric Data*. PhD thesis.

Manassero, M. C., Afonso, J. C., Zyserman, F., Zlotnik, S., and Fomin, I. (2020). A Reduced Order Approach for Probabilistic Inversions of 3D Magnetotelluric Data I: General Formulation. *Geophysical Journal International*.

Manzoni, A., Pagani, S., and Lassila, T. (2016). Accurate solution of Bayesian inverse uncertainty quantification problems combining reduced basis methods and reduction error models. *SIAM-ASA Journal on Uncertainty Quantification*, 4(1):380–412.

Marquart, G. and Schmeling, H. (1989). Topography and Geoid Undulations Caused By Small-Scale Convection Beneath Continental Lithosphere of Variable Elastic Thickness. *Geophysical Journal International*, 97(3):511–527.

Marzouk, Y. M. and Najm, H. N. (2009). Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902.

Mckenzie, D. P. (1977). Surface deformation , gravity anomalies and convection. *Geophysical Journal International*, 48(2):211–238.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Mira, A. (2001). On Metropolis – Hasting algorithms with delayed rejection. *Metron*, 59(3).

Modesto, D., Zlotnik, S., and Huerta, A. (2015). Proper generalized decomposition for parameterized Helmholtz problems in heterogeneous and unbounded domains: Application to harbor agitation. *Computer Methods in Applied Mechanics and Engineering*, 295:127–149.

Molnar, P., England, P. C., and Jones, C. H. (2015). Mantle dynamics, isostasy, and the support of high terrain. *Journal of Geophysical Research : Solid Earth*, 120:1932–1957.

Mosegaard, K. and Tarantola, A. (2002). Probabilistic Approach to Inverse Problems. In *International Handbook of Earthquake and Engineering Seismology (Part A)*, pages 237–265. Academic Press.

Ortega-Gelabert, O., Zlotnik, S., Afonso, J. C., and Díez, P. (2020). Fast Stokes Flow Simulations for Geophysical-Geodynamic Inverse Problems and Sensitivity Analyses Based On Reduced Order Modeling. *Journal of Geophysical Research: Solid Earth*, 125(3):1–25.

Patera, A. T. and Rozza, G. (2006). *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. Copyright MIT, version 1. edition.

Peherstorfer, B. and Willcox, K. (2015). Dynamic data-driven reduced-order models. *Computer Methods in Applied Mechanics and Engineering*, 291:21–41.

Peherstorfer, B., Willcox, K., and Gunzburger, M. (2018). Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. *SIAM Review*, 60(3):550–591.

Poliakov, A. and Podladchikov, Y. (1992). Diapirism and topography. *Geophysical Journal International*, 109(3):553–564.

Prud'homme, C., Rovas, D., Veroy, K., Machiels, L., Maday, Y., Patera, A. T., and Turinici, G. (2002). Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods. *Journal of Fluids Engineering*, 124(1):70.

Quarteroni, A., Manzoni, A., and Negri, F. (2016). *Reduced Basis Methods for Partial Differential Equations*, volume 92 of *UNITEXT*. Springer International Publishing, Cham.

Ranalli, G. (1995). *Rheology of the Earth*. Chapman & Hall, second edition.

Ravindran, S. (2000). A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 34(5):425–448.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367.

Rozza, G., Huynh, D. B. P., and Manzoni, A. (2013). Reduced basis approximation and a posteriori error estimation for Stokes flows in parametrized geometries: roles of the inf-sup stability constants. *Numerische Mathematik*, 125(1):115–152.

Rozza, G., Huynh, D. B. P., and Patera, A. T. (2007). Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):1–47.

Rozza, G. and Veroy, K. (2007). On the stability of the reduced basis method for Stokes equations in parametrized domains. *Computer Methods in Applied Mechanics and Engineering*, 196(7):1244–1260.

Schubert, G., Turcotte, D. L., and Olson, P. (2004). *Mantle Convection in the Earth and Planets.* Cambridge University Press.

Serafin, K., Magnain, B., Florentin, E., Parés, N., and Díez, P. (2017). Enhanced goal-oriented error assessment and computational strategies in adaptive reduced basis solver for stochastic problems. *International Journal for Numerical Methods in Engineering*, 110(5):440–466.

Sibileau, A., García-González, A., Auricchio, F., Morganti, S., and Díez, P. (2018). Explicit parametric solutions of lattice structures with proper generalized decomposition (PGD). *Computational Mechanics*, 62(4):871–891.

Simmons, N. A., Forte, A. M., Boschi, L., and Grand, S. P. (2010). GyPSuM: A joint tomographic model of mantle density and seismic wave speeds. *Journal of Geophysical Research*, 115(B12):B12310.

Steinberger, B. and Becker, T. W. (2018). A comparison of lithospheric thickness models. *Tectonophysics*, 746:325–338.

Stixrude, L. and Lithgow-Bertelloni, C. (2012). Geophysics of Chemical Heterogeneity in the Mantle. *Annual Review of Earth and Planetary Sciences*, 40(1):569–595.

Stüwe, K. (2007). *Geodynamics of the Lithosphere: An Introduction.* Springer-Verlag Berlin Heidelberg.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation.* SIAM.

Tarvainen, T., Kolehmainen, V., Pulkkinen, A., Vauhkonen, M., Schweiger, M., Arridge, S. R., and Kaipio, J. P. (2010). An approximation error approach for compensating for modelling errors between the radiative transfer equation and the diffusion approximation in diffuse optical tomography. *Inverse Problems*, 26(1):0–18.

Titus, W. J., Titus, S. J., and Davis, J. R. (2017). A Bayesian approach to modeling 2D gravity data using polygons. *Geophysics*, 82(1):G1–G21.

Tork Qashqai, M., Afonso, J. C., and Yang, Y. (2018). Physical State and Structure of the Crust Beneath the Western-Central United States From Multiobservable Probabilistic Inversion. *Tectonics*, 37(9):3117–3147.

Tork Qashqai, M., Carlos Afonso, J., and Yang, Y. (2016). The crustal structure of the Arizona Transition Zone and southern Colorado Plateau from multiobservable probabilistic inversion. *Geochemistry, Geophysics, Geosystems*, 17(11):4308–4332.

Turcotte, D. L. and Schubert, G. (2002). *Geodynamics*. Cambridge University Press, second edition.

Veroy, K. and Patera, A. T. (2005). Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: rigorous reduced-basisa posteriori error bounds. *International Journal for Numerical Methods in Fluids*, 47(8-9):773–788.

Veroy, K., Prud'homme, C., Rovas, D. V., and Patera, A. T. (2003). A Posteriori Error Bounds for Reduced-Basis Approximation of Parametrized Noncoercive and Nonlinear Elliptic Partial Differential Equations. (June):1–18.

Wang, J. and Zabaras, N. (2005). Using Bayesian statistics in the estimation of heat source in radiation. *International Journal of Heat and Mass Transfer*, 48(1):15–29.

Yan, L. and Zhang, Y.-X. (2017). Convergence analysis of surrogate-based methods for bayesian inverse problems. *Inverse Problems*, 33(12):125001.

Yan, L. and Zhou, T. (2019). Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems. *Journal of Computational Physics*, 381:110–128.

Zhang, J., Zheng, Q., Chen, D., Wu, L., and Zeng, L. (2020). Surrogate-Based Bayesian Inverse Modeling of the Hydrological System: An Adaptive Approach Considering Surrogate Approximation Error. *Water Resources Research*, 56(1).

Zhong, S., Gurnis, M., and Hulbert, G. (1993). Accurate determination of surface normal stress in viscous flow from a consistent boundary flux method. *Physics of the Earth and Planetary Interiors*, 78(1-2):1–8.

Zienkiewicz, O. C., Taylor, R. L., and Zhu, J. Z. (2005). *The Finite Element Method: its Basis and Fundamentals*. Elsevier.

Zlotnik, S., Díez, P., Modesto, D., and Huerta, A. (2015). Proper generalized decomposition of a geometrically parametrized heat problem with geophysical applications. *International Journal for Numerical Methods in Engineering*, 103(10):737–758.