# Ensemble data assimilation methods for estimating fault slip and future earthquake occurrences

Diab Montero, H.A.

**DOI**
10.4233/uuid:92ebf293-d057-4ab3-b11a-ba2204f377e8

**Publication date**
2024

**Document Version**
Final published version

**Citation (APA)**
Diab Montero, H. A. (2024). *Ensemble data assimilation methods for estimating fault slip and future earthquake occurrences*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:92ebf293-d057-4ab3-b11a-ba2204f377e8

# ENSEMBLE DATA ASSIMILATION METHODS FOR ESTIMATING FAULT SLIP AND FUTURE EARTHQUAKE OCCURRENCES

# ENSEMBLE DATA ASSIMILATION METHODS FOR ESTIMATING FAULT SLIP AND FUTURE EARTHQUAKE OCCURRENCES

## Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 24 April 2024 om 10:00 uur

door

**Hamed Ali DIAB MONTERO**

Master of Science in Applied Geophysics,
IDEA League Joint Collaboration program TU Delft, ETH Zürich, and RWTH Aachen,
geboren te Valledupar, Colombia.

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie:

| | |
|---|---|
| Rector Magnificus, | voorzitter |
| Prof. dr. ir. J.D. Jansen, | Technische Universiteit Delft, promotor |
| Dr. ir. F.C. Vossepoel, | Technische Universiteit Delft, copromotor |
| | |
| *Onafhankelijke leden:* | |
| Prof. dr. ir. E.C. Slob, | Technische Universiteit Delft |
| Prof. dr. P.J. van Leeuwen, | Colorado State University |
| Prof. dr. ir. M. Verlaan, | Technische Universiteit Delft |
| Dr. sc. nat. M. Ramgraber, | Technische Universiteit Delft |
| Dr. ir. D.S. Draganov, | Technische Universiteit Delft, reservelid |
| | |
| *Overig lid:* | |
| Dr. Y. van Dinther, | Utrecht Universiteit |

This research was carried out as part of the "InFocus: An Integrated Approach to Estimating Fault Slip Occurrence" project (grant number: DEEP.NL.2018.037) funded by NWO's (Dutch Research Council) DeepNL programme.




*Keywords:* Data assimilation · Inverse theory · Numerical modelling · Probabilistic forecasting · Earthquake interaction, forecasting, and prediction · Earthquake dynamics · Seismic cycle · Ensemble Kalman Filter

*Printed by:* Gildeprint.



ISBN 978-94-6384-568-7

An electronic version of this dissertation is available at: http://repository.tudelft.nl/.

*A mis padres, Sara y Alexander,*

# CONTENTS

# SUMMARY

In this dissertation, I explore ensemble data assimilation methods to enhance our capability to forecast earthquakes and slow slip events, focusing on the critical challenge posed by limited information on the current stress state of faults.

At the outset, the research acknowledges the inherent limitations in our current understanding of fault stress states. These limitations significantly hinder our ability to forecast seismic events accurately. The study proposes utilizing ensemble data assimilation techniques as a robust solution. Central to this dissertation, these methods enable estimating the fault's state by integrating information from physics-based models with observational data. Importantly, this approach considers the uncertainties inherent in both the models and the data, offering a more reliable framework for forecasting. The dissertation emphasizes that probabilistic forecasts represent the highest achievable goal in earthquake forecasting. However, it also recognizes the challenges that arise from limited information on critical aspects such as stress, strength, and governing parameters of seismogenic sources. These limitations can significantly impede the accuracy of forecasts.

Throughout the dissertation, I systematically examine how ensemble data assimilation can be effectively implemented to improve earthquake forecasting. This involves exploring current field measurement techniques and the data quality they produce. The study demonstrates how ensemble data assimilation can bridge the gap between empirical observations and theoretical understanding by carefully analyzing and integrating this data with advanced theoretical models.

A core component of the research is a critical evaluation of various data assimilation techniques, mainly focusing on their ability to enhance forecasting accuracy in the context of limited information on fault stress states.

Furthermore, I explore practical applications of these techniques using 1D and 2D models. This includes investigating how data assimilation can improve the forecasting of earthquake occurrences and the inherent challenges in making such estimations.

The dissertation culminates in a forward-looking discussion on the future of earthquake forecasting. It emphasizes the role of ensemble data assimilation methods in overcoming the current limitations of stress state information and proposes ways for more informed seismic forecasts.

# SAMENVATTING

In dit proefschrift onderzoek ik ensemble data-assimilatietechnieken om ons vermogen te verbeteren om aardbevingen en langzame slipgebeurtenissen te verwachting, met een focus op de uitdaging die beperkte informatie over de huidige spanningstoestand van breuken met zich meebrengt.

Aan het begin erkent het onderzoek de inherente beperkingen in ons huidige begrip van spanningstoestand van breuken. Deze beperkingen hinderen onze mogelijkheid om seismische gebeurtenissen nauwkeurig te voorspellen aanzienlijk. De studie stelt voor om ensemble data-assimilatietechnieken te gebruiken als een robuuste oplossing. Deze technieken, essentieel voor deze dissertatie, maken het mogelijk de toestand van de breuk te schatten door informatie uit fysiek gebaseerde modellen te integreren met waarnemingen. Belangrijk is dat deze benadering de onzekerheden die inherent zijn aan zowel de modellen als de gegevens in overweging neemt, en daarmee een betrouwbaarder kader voor voorspellingen biedt.

De dissertatie benadrukt dat probabilistische voorspellingen het hoogst haalbare doel vertegenwoordigen in aardbevingsvoorspellingen. Het erkent echter ook de uitdagingen die voortkomen uit beperkte informatie over kritieke aspecten zoals spanning, sterkte en bepalende parameters van seismogene bronnen. Deze beperkingen kunnen de nauwkeurigheid van voorspellingen aanzienlijk belemmeren.

Doorheen de dissertatie onderzoek ik systematisch hoe ensemble data-assimilatie effectief geïmplementeerd kan worden om de voorspelling van aardbevingen te verbeteren. Dit omvat het verkennen van huidige veldmeettechnieken en de datakwaliteit die ze produceren. De studie toont aan hoe ensemble data-assimilatie de kloof tussen empirische waarnemingen en theoretisch begrip kan overbruggen door deze gegevens zorgvuldig te analyseren en te integreren met geavanceerde theoretische modellen.

Een kerncomponent van het onderzoek is een kritische evaluatie van diverse data-assimilatietechnieken, voornamelijk gericht op hun vermogen om de nauwkeurigheid van voorspellingen te verbeteren in de context van beperkte informatie over spanningstoestanden van breuken.

Verder onderzoek ik praktische toepassingen van deze technieken met behulp van 1D- en 2D-modellen. Dit omvat het onderzoeken hoe data-assimilatie de voorspelling van aardbevingsgebeurtenissen kan verbeteren en de inherente uitdagingen bij het maken van dergelijke schattingen.

De dissertatie eindigt met een vooruitblikkende discussie over de toekomst van aardbevingsvoorspelling. Het benadrukt de rol van ensemble data-assimilatietechnieken bij het overwinnen van de huidige beperkingen van informatie over spanningstoestanden en stelt manieren voor voor meer geïnformeerde seismische verwacthing.

# RESUMEN

En esta disertación, se exploran métodos de asimilación de datos para mejorar nuestra capacidad de pronosticar terremotos y eventos de deslizamiento, enfocándose en el desafío crítico que plantea la información limitada sobre el estado actual de esfuerzos en las fallas.

Al principio, la investigación reconoce las limitaciones inherentes a nuestra comprensión actual de los estados de esfuerzo cortante. Estas limitaciones obstaculizan significativamente nuestra capacidad para pronosticar eventos sísmicos con precisión. El estudio propone utilizar técnicas de asimilación de datos como una solución robusta. Estos métodos, fundamentales para esta tesis, permiten estimar el estado de las fallas integrando información de modelos físicamente basados con datos observados. Es importante destacar que este enfoque considera las incertidumbres inherentes tanto a los modelos como a los datos, ofreciendo un marco más confiable para los pronósticos.

Esta tesis enfatiza que los pronósticos probabilísticos representan el objetivo más alto alcanzable en el pronóstico de terremotos. Sin embargo, también reconoce los desafíos que surgen de la información limitada sobre aspectos críticos como los esfuerzos, la resistencia y los parámetros que rigen las fuentes sismogénicas. Estas limitaciones pueden impedir significativamente la precisión de los pronósticos.

A lo largo de la disertación, se examina sistemáticamente cómo la asimilación de datos se puede implementar de manera efectiva para mejorar el pronóstico de terremotos. Esto implica explorar las técnicas actuales de medición in-situ y la calidad de los datos que producen. El estudio demuestra cómo la asimilación de datos puede cerrar la brecha entre las observaciones empíricas y la comprensión teórica al analizar e integrar cuidadosamente estos datos con modelos teóricos avanzados.

Un componente central de la investigación es una evaluación crítica de diversas técnicas de asimilación de datos, centrándose principalmente en su capacidad para mejorar la precisión de los pronósticos en el contexto de información limitada sobre los estados de esfuerzo de las fallas.

Además, se explora aplicaciones prácticas de estas técnicas utilizando modelos 1D y 2D. Esto incluye investigar cómo la asimilación de datos puede mejorar el pronóstico de la ocurrencia de los terremotos a pesar de los desafíos inherentes a la realización de tales estimaciones.

La disertación culmina con una discusión prospectiva sobre el futuro del pronóstico sísmico. Enfatiza el papel de los métodos de asimilación de datos para superar las limitaciones actuales de la información sobre el estado de esfuerzo y propone formas de realizar pronósticos sísmicos más informados.

# PREFACE

I arrived in the Netherlands in 2017 to obtain my master's degree in Applied Geophysics. Before starting the master's program, I had the opportunity to work as a geotechnical engineer at a civil engineering consultancy, Ingetec SA, whose main market was the design of hydroelectric dams. I greatly enjoyed working with geologists, geophysicists, and civil engineers in an interdisciplinary group. Having completed my degrees in Civil Engineering and Geosciences, I was curious about how these different areas would integrate into practice. It surprised me that specialists in Geophysics often lacked training in civil engineering and design, which made it difficult to use the information gathered by the company in civil designs due to a lack of understanding of how to interpret it. I believed I could make a significant contribution by using that information to create more robust designs, especially since soil variability is a key factor in project failures and future repairs. Exploration for construction site characterization is often limited in many projects. However, as a civil engineer, I could leverage geophysical tools to reduce costs and deepen this knowledge, benefiting my colleagues. The master's program offered a wide range of geophysical areas, such as hydrogeophysics, hydrocarbon exploration, metallurgical prospecting, infrastructure projects, and geothermal applications.

During the master's program, drawn to the topic of data, I decided to explore new areas. I completed my graduation project in machine learning to improve the detection of gravitational waves, a field I never imagined working in but which greatly inspired me due to its innovation and future potential. I soon became interested in how to integrate my mathematical and numerical modelling knowledge with advanced computational tools like neural networks. It was a pleasant surprise to discover that at TU Delft, there was a vacancy to work with data assimilation methods. This field combined my interests in uncertainty quantification, data analysis, and physical knowledge. After receiving my master's degree, I decided to spend the following years in Delft, immersing myself in seismology, climatology, laboratory experimentation, and other diverse topics that enriched my experience. Unfortunately, I did not anticipate a global pandemic or injury affecting my mobility. During this time, I understood the importance of family and having an emotional support system to achieve great goals. My father, a doctor by profession, was a fundamental pillar during this period. He not only fought tirelessly during the COVID-19 pandemic but also shared his knowledge to help me overcome my injuries. My mother was also a crucial support, providing the emotional and mental strength necessary to face such a significant challenge. They were my physical and emotional support, so I dedicated these four years of effort to them.

All this time has taught me a lot about uncertainty, how it affects engineering projects, benefits us in learning about the unknown, and brings us closer to our loved ones. In this dissertation, I have applied all the conversations I had with experts and colleagues and my findings to make progress in one of the most complex and uncertain areas: earthquake forecasting. I hope the ideas presented here inspire more scientists, colleagues, and friends so that we can forecast earthquakes someday, just as we have managed to forecast the weather.

*Hamed Ali Diab Montero*
*Delft, December 2023*

# PREFACIO

Llegué a los Países Bajos en 2017 con el objetivo de obtener mi título de máster en Geofísica Aplicada. Antes del máster, tuve la oportunidad de trabajar como ingeniero geotecnista en una consultoría de ingeniería civil, Ingetec SA, cuyo principal mercado era el diseño de presas hidroeléctricas. Disfruté mucho de trabajar en un grupo interdisciplinario con geólogos, geofísicos e ingenieros civiles. Habiendo completado mis titulaciones en Ingeniería Civil y Geociencias, tenía gran curiosidad por cómo se integrarían estas áreas en la práctica. Me sorprendió que los especialistas en Geofísica a menudo carecían de formación en ingeniería civil y diseño, lo que dificultaba el uso de la información recopilada por la compañía en diseños civiles debido a la falta de comprensión de cómo interpretarla. Creí que podría contribuir significativamente utilizando esa información para realizar diseños más robustos, especialmente porque la variabilidad del suelo es un factor clave en los fallos y reparaciones futuras de los proyectos. La exploración del sitio de construcción suele ser limitada en muchos proyectos, pero pensé que, como ingeniero civil, podría aprovechar las herramientas geofísicas para reducir costos y profundizar este conocimiento, beneficiando a mis colegas. El máster ofrecía una amplia gama de áreas geofísicas como la hidrogeofísica, la exploración de hidrocarburos, la prospección metalúrgica, proyectos de infraestructura y aplicaciones geotérmicas, entre otros.

Durante la maestría, atraído por el tema de los datos, decidí explorar nuevas áreas y realicé mi proyecto de graduación en aprendizaje automático para mejorar la detección de ondas gravitacionales, un campo en el que nunca imaginé trabajar pero que me inspiró enormemente por su innovación y potencial futuro. Pronto me interesé en cómo integrar mis conocimientos matemáticos y de modelado numérico con herramientas computacionales avanzadas como las redes neuronales. Fue una agradable sorpresa descubrir que en TU Delft había una vacante para trabajar con métodos de asimilación de datos, un campo que combinaba mis intereses en cuantificación de incertidumbre, análisis de datos y conocimiento físico. Tras obtener mi título de máster, decidí pasar los siguientes años en Delft, sumergiéndome en sismología, climatología, experimentación de laboratorio y otros temas diversos que enriquecieron mi experiencia. Desafortunadamente, no anticipé una pandemia global ni una lesión que afectaría mi movilidad. Durante este tiempo, comprendí la importancia de la familia y de contar con un sistema de apoyo emocional para alcanzar grandes metas. Mi padre, médico de profesión, fue un pilar fundamental durante este periodo. No solo luchó incansablemente durante la pandemia del COVID-19, sino que también compartió su conocimiento para ayudarme a superar mis lesiones. Mi madre también fue un apoyo crucial, proporcionando la fortaleza emocional y mental necesaria para afrontar un desafío de tal magnitud. Ellos fueron mi soporte físico y emocional durante este tiempo, por lo que les dedico este esfuerzo de cuatro años.

Todo este tiempo me ha enseñado mucho sobre la incertidumbre, cómo influye en proyectos de ingeniería, cómo nos beneficia para abordar lo desconocido y cómo nos acerca a nuestros seres queridos. En esta disertación, he intentado aplicar todas las conversaciones que tuve con expertos, colegas y mis propios descubrimientos para avanzar en una de las áreas más complejas y llenas de incertidumbre: el pronóstico de terremotos. Espero que las ideas aquí presentadas inspiren a más científicos, colegas y amigos para que, algún día, así como hemos logrado pronosticar el clima, podamos pronosticar terremotos.

*Hamed Ali Diab Montero*
*Delft, Diciembre 2023*

# 1

# INTRODUCTION

***Summary:*** *This chapter delves into integrating ensemble data-assimilation techniques in seismology, focusing on their application in earthquake forecasting. It outlines seismologists' challenges in forecasting earthquakes, highlighting the potential of data assimilation to mitigate these issues. The chapter provides a comprehensive overview of previous research efforts that have employed these techniques to estimate key variables across various phases of the seismic cycle, demonstrating both advancements and the ongoing complexities in this field.*

***Resumen:****Este capítulo profundiza en cómo se integran las técnicas de asimilación de datos en la sismología, centrándose en su aplicación para el pronóstico de terremotos. Describe los desafíos que enfrentan los sismólogos en el pronóstico de terremotos y destaca el potencial de la asimilación de datos para mitigar estos problemas. El capítulo ofrece una visión integral de los esfuerzos de investigación anteriores que han utilizado estas técnicas para estimar variables clave en distintas fases del ciclo sísmico, evidenciando tanto los avances logrados como las complejidades persistentes en este ámbito.*

Parts of this chapter have been published in the introduction sections of Hamed Ali Diab-Montero, Meng Li, Ylona van Dinther, Femke C Vossepoel, Estimating the occurrence of slow slip events and earthquakes with an ensemble Kalman filter, Geophysical Journal International, Volume 234, Issue 3, September 2023, Pages 1701–1721, https://doi.org/10.1093/gji/ggad154.
Minor changes have been applied to make the text and figures consistent with the thesis.

1

## 1.1. MOTIVATION

In seismology, the journey toward accurately understanding and forecasting earthquakes has been marked by methodological advancements. Initially, researchers employed kinematic inversion, a technique focused on analyzing past earthquakes by reconstructing their movement from observational data. This evolved into dynamic source inversion, which, while enhancing the field's capabilities by integrating sophisticated physics-based models with observational data, faces significant limitations. Dynamic source inversion is computationally expensive and primarily suited for reconstructing the history of past earthquake events. It struggles with uncertain scenarios for forecasting new events and incurs high costs in incorporating new data for sequential analysis.

Interestingly, these challenges once impeded weather forecasting as well, but the introduction of data assimilation revolutionized the field, making it operationally viable for accurate and timely weather predictions. Inspired by this success, the field of seismology is now exploring the potential of data assimilation. This promising approach aims to refine our understanding of seismic phenomena by providing a re-analysis of past earthquakes and offering the potential for future earthquake forecasting. Data assimilation seeks to estimate critical information, addressing uncertainties that have historically impeded progress in earthquake forecasting. This includes determining the current state of stress and velocities of the fault and critical parameters in earthquake models. Drawing parallels with its transformative impact in weather forecasting, data assimilation in seismology presents a promising approach for advancing our understanding and forecasting of seismic events.

## 1.2. EARTHQUAKE FORECASTING

The endeavor to predict earthquakes has been a focal point in seismological research for decades. According to a widely endorsed definition, an effective earthquake prediction necessitates accurately determining the seismic event's magnitude, timing, and epicenter within specified limits. Geller (1997) discusses the complexity of predicting earthquakes and asserts that despite numerous efforts, success in this domain has been limited, and earthquakes are inherently unpredictable due to self-organized criticality within the Earth's crust. This perspective suggests that even minor instabilities can trigger significant seismic events, complicating predictive efforts (Geller 2011). Furthermore, the reliable prediction of large earthquakes would demand an intricate understanding of many geophysical conditions across a large volume.This task is rendered unfeasible due to the sensitive nonlinear nature of earthquake rupture processes and the extensive initial condition data required. Additionally, the absence of consistent and observable earthquake precursors further hinders empirical prediction efforts. In attempting to predict earthquakes, our highest achievable goal is earthquake forecasting, which involves statistical assessments about the likelihood of seismic activity. In this dissertation, we focus on providing probabilistic statements of the onset of fault slip and, based on that, estimates of earthquake occurrences.

Earthquakes represent a significant threat due to their sudden occurrence and potential for extensive damage. In response, seismologists have developed two commonly used approaches. Probabilistic Seismic Hazard Assessments (PSHAs) epitomize the first

approach, leveraging historical seismic data, geological insights, and seismic response models to estimate the recurrence intervals of earthquakes (Bommer and Abrahamson 2006; Cornell 1968; Esteva 1967; Ordaza and Arroyo 2016). However, the accuracy of PSHAs is frequently undermined by the inherent limitations of the data, including its sparsity and associated uncertainties (Geller 2011). Complementing this, the second approach encompasses Earthquake Early Warning (EEW) systems, which capitalize on recent advancements in seismic sensor technology. These systems detect primary waves (P-waves), providing critical, albeit brief, lead time before the onset of the more destructive secondary (S-waves) and surface waves (Allen and Kanamori 2003; Allen and Melgar 2019). Despite these developments a considerable gap remains in forecasting capabilities. Current research efforts are directed towards understanding earthquake sequences and the physics of earthquake nucleation, with a particular focus on identifying reliable precursors such as foreshocks and slow slip events (SSEs), which could enhance short-term forecasting capabilities (Pritchard et al. 2020; Segall and Bradley 2012; Socquet et al. 2017; Uchida et al. 2016). Therefore, a probabilistic estimate over more extended periods is currently the highest achievable goal in earthquake forecasting.

In the realm of earthquake forecasting, two predominant theoretical models prevail. The first conceptualizes earthquakes as outcomes of a stochastic process, with the Epidemic Type Aftershock Sequence (ETAS) model as a notable example (Gualandi et al. 2020). Pioneered by Ogata (1988), this model employs a Poissonian distribution for primary shocks, supplemented by Omori's law for aftershocks. While ETAS has gained traction in operational forecasting, particularly under scenarios of high-quality data and well-known geology of the area, it is constrained by the availability and completeness of historical seismic data. It does not sufficiently address the mechanics underlying earthquake genesis (Seif et al. 2017).

Conversely, the deterministic chaotic process model, a system where small changes in initial conditions lead to vastly different outcomes despite being governed by predictable laws, emphasizes creating and utilizing sophisticated numerical models that meticulously simulate the rupture dynamics at fault lines (Gualandi et al. 2020). This approach has been instrumental in enhancing our understanding of the source-rupture processes. By integrating measurements of large displacements and geodetic data, these models have substantially advanced our knowledge of the slip history on faults via kinematic inversion techniques (Ozgun Konca et al. 2013). They are also progressively being employed to elucidate the distribution and evolution of post-seismic stress along fault lines (Minson et al. 2014). However, translating these detailed models into reliable forecasts of future seismic activity remains a significant and unresolved challenge.

## 1.3. DATA ASSIMILATION AND EARTHQUAKE FORECASTING

Data assimilation (DA) techniques are essential in forecasting geophysical systems where there is significant uncertainty in initial conditions or governing parameters. These methods merge physics-based simulations with observational data, enhancing the forecasting accuracy for complex, chaotic systems like weather, ocean currents, and hydrologic processes (Bannister 2017; Evensen 1994; Evensen 2003; Evensen et al. 2022; Liu et al. 2012; Reichle 2008; van Leeuwen 2010, 2003b; Vossepoel and Behringer 2000; Weaver et al. 2003). Notably, DA has also proven effective in various other fields such as oil and gas

production, remote sensing, and even in analyzing the SARS-CoV-2 outbreak (Aanonsen et al. 2009; Bau et al. 2014; Emerick 2018; Emerick and Reynolds 2013; Evensen et al. 2020; Evensen and Eikrem 2018; Fokker et al. 2009; Guzman et al. 2014; Oliver et al. 2011; Seiler et al. 2009; Wilschut et al. 2011). Its application in earthquake forecasting, by treating these events as deterministic chaotic systems, holds promise for enhancing the understanding and forecasting of seismic events using advanced physical models and data.

The realm of DA is broad, encompassing variational and ensemble-based methods (Carrassi et al. 2017; Evensen 2003; Evensen et al. 2022). Variational methods, like 3D Var and 4D Var, approach the estimation problem as an optimization inverse problem minimizing a cost function, but with the limitation of usually requiring the calculation of a model's adjoint operator (Daužickaitė et al. 2021; Lorenc 2003). Ensemble-based methods, conversely, focus on Bayesian inference to find the most probable posterior solutions but often demand considerable computational resources to generate extensive realizations. Examples include the Ensemble Kalman Filter (EnKF) and the particle filter (van Leeuwen 2009; van Leeuwen et al. 2019). Innovations in computing and the development of hybrid approaches are now making these methods more accessible and efficient (Chustagulprom et al. 2016; Vetra-Carvalho et al. 2017).

Several ensemble data assimilation methods have emerged as effective tools for handling non-Gaussian distributions in chaotic systems; in this thesis, I explored specifically the Adaptive Gaussian Mixture Filter (AGMF) and the Particle Flow Filter (PFF) (Hu and van Leeuwen 2021; Stordal et al. 2011). The AGMF combines elements of particle filtering by including an importance sampling step via a Gaussian mixture with Ensemble Kalman Filtering. Unfortunately, in importance sampling methods, weight degeneracy occurs, meaning that a few samples acquire disproportionately large weights, compromising the sample set's representativeness and the estimation process's efficiency. The AGMF uses resampling to ensure a balanced representation of the underlying distribution. The PFF addresses weight degeneracy by avoiding weights and using an iterative transformation process of the ensemble distributions, showcasing success in atmospheric physics models.

Expanding these ensemble data assimilation methods to earthquake physics, particularly from the context of conceptual earthquake models, such as the spring-slider model developed by Burridge and Knopoff (1967), up to complex dynamic rupture models offers an innovative approach to seismic forecasting, setting the stage for the discussion on the applications of DA in earthquake forecasting. Particularly, we expand to understanding how these methods can provide better estimates when using rate-and-state friction for the earthquake models since it is widely used in seismology to explain earthquake sequences and a wide variety of slip behaviors observed in laboratory frictional experiments, but poses challenges due to its non-linear dynamics.

In recent years, advancements in data assimilation for earthquake forecasting have seen significant strides, although with ongoing challenges. Initially, researchers focused on leveraging paleoseismic data with statistical models to forecast earthquakes. However, these efforts, as indicated by Ogata (1999) and Aalsburg et al. (2007), faced setbacks in this integration due to sparse data and high uncertainty, resulting in predictions no better than random guesses. Subsequent endeavors, integrating particle filters with

earthquake statistical models, treated as lognormal renewal processes, showed promise in applications with synthetic models (Varini 2007; Werner et al. 2011), but the application to real earthquake data remained untested.

Over the past decade, researchers have explored data assimilation in three distinct groups: tsunami early warning systems, optimal estimation of frictional parameters using the rate-and-state friction law, and probabilistic forecasting of earthquake sequences. Tsunami early warning efforts, as studied by Maeda et al. (2015), Wang et al. (2018), and Oba et al. (2020), utilized the Ensemble Kalman Filter with dense tsunameter networks, demonstrating effectiveness in real-time alerts. Another line of work by Kano et al. (2010) through Kano et al. (2020) focused on estimating frictional parameters of megathrust faults using adjoint-based models, yielding positive results in real-time corrections. However, limitations were noted in forecasting slip velocities for future seismic stages. The final group, including van Dinther et al. (2019) and Hirahara and Nishikiori (2019), employed the Ensemble Kalman Filter for modeling 2D slip-weakening patches in subduction zones. These studies showed that even minimal observations could enhance fault slip predictions, although the feasibility of applying these methods to real measurements requires further investigation.

As for data assimilation methods targeting different earthquake phases, researchers have achieved varying degrees of success. For coseismic phases of large earthquakes, seismic wavefield estimations have been prominent (Maeda et al. 2015; Oba et al. 2020). Post-seismic and slow slip events have seen developments in calculating slip rates and frictional parameters (Hori et al. 2014; Kano et al. 2020, 2013). The Ensemble Kalman Filter, in particular, has proven effective for Gaussian distributions in slow acceleration models (Hirahara and Nishikiori 2019; van Dinther et al. 2019). However, the complexity increases with fast acceleration models during coseismic phases, where non-Gaussian distributions emerge due to rapid stress and velocity changes (Diab-Montero et al. 2023). While particle filters handle non-Gaussian distributions effectively in lower-dimensional systems, as demonstrated by Banerjee et al. (2023), they struggle with the large state vectors of more comprehensive models, demanding significant computational resources (Li et al. 2022). Consequently, there's a pressing need to develop data assimilation methods that can handle high-dimensional vectors and non-Gaussian distributions efficiently, which is vital for accurate and computationally feasible earthquake forecasting.

## 1.4. THESIS OBJECTIVES

In my thesis, I investigate the application of data assimilation techniques in conjunction with physics-based models and controlled laboratory observations to advance the field of earthquake forecasting. This approach aims to improve the estimation of fault slip events by addressing the inherent uncertainties in the initial conditions and state variables that govern fault ruptures. Data assimilation is used to enhance the forecast estimates while considering these uncertainties.

The research addresses gaps commonly encountered when using data assimilation methods for earthquake forecasting. It evaluates the utility of real measurements within dynamic rupture models, moving beyond the twin experiment validations seen in prior studies like those by van Dinther et al. (2019) and Hirahara and Nishikiori (2019). An aspect of the research involves examining the impact of the timing of data assimilation

within different phases of the seismic cycle on the accuracy of fault slip forecasts. This examination draws on preliminary insights from works such as van Dinther et al. (2019) and Kano et al. (2015, 2013). Additionally, this work extends to the potential application of data assimilation to induced seismic events and smaller-scale laboratory studies, which are less extensively explored than in larger tectonic settings. The thesis also compares the effectiveness of various ensemble methods in earthquake forecasting, including Ensemble Kalman Filters, Adaptive Gaussian Mixture Filters, and Particle Flow Filters, to identify the most suitable approach for this specialized area.

The foundational hypotheses of this research are as follows:

- Data assimilation can enhance earthquake forecasting by providing more accurate estimates of the evolving state variables at fault interfaces. This technique is expected to address current method gaps, particularly measurement and model errors, and the indirect nature of available observations.
- Applying data assimilation in controlled laboratory settings, analogous to conditions such as those in reservoirs, may offer new perspectives on seismicity, employing techniques traditionally used for large-scale tectonic analysis.
- Furthermore, incorporating shear stress data and slip velocity is hypothesized to improve fault slip occurrence estimates in our data assimilation framework.

This research aims to test these hypotheses, contributing to advancing earthquake forecasting methods, and has led to the following research objective of this thesis:

**Research objective:**
Investigate how ensemble data assimilation methods can enhance the estimation of variables that control fault slip, aiming to improve the estimation of earthquake occurrences when using numerical models governed by the rate-and-state friction law.

## 1.5. THESIS OUTLINE

The outline of this thesis is as follows:

- Chapter 2, Numerical Modeling of Earthquakes.
  This chapter outlines the numerical models employed for forward modeling the earthquake sequences in the data assimilation experiments of this thesis. It delves into the underlying physics, boundary conditions, and the frictional laws applied to simulate the earthquakes. This encompasses a detailed explanation of the rate-and-state friction law, conceptual models such as the Burridge-Knopoff models, and continuum approximations that assume homogeneous elastic surrounding media, solved using finite difference methods.
- Chapter 3, Ensemble Data Assimilation Methods.
  This chapter introduces the fundamentals of data assimilation, including the specific notation used for data assimilation within the scope of this thesis. It highlights

the different ensemble data assimilation methods explored, namely the Ensemble Kalman Filter (EnKF), the Adaptive Gaussian Mixture Filter (AGMF), and the Particle Flow Filter (PFF). Additionally, the chapter presents techniques such as inflation and localization, which address issues related to spurious correlations and low-rank deficiencies. Finally, some results of perfect model experiments done with the Lorenz 96 model, a standard test bed in the data assimilation community, are shown to highlight the characteristics of the methods and their differences.

- Chapter 4, Estimating the Occurrence of SSEs and Earthquakes with an EnKF.
  This chapter explores the estimation of slow slip events and earthquake occurrences and their challenges using an Ensemble Kalman Filter (EnKF) and a 1D model. It examines critical factors such as the identification and uncertainty of key state variables that could enhance the accuracy of these estimates. Furthermore, the chapter assesses the types of necessary observations, evaluating which offers the most significant information content and the impact of their assimilation across different seismic cycle phases. Additionally, it investigates the implications of adopting Gaussian distributions in our models and how this affects the precision of fault slip occurrence estimation, offering a comprehensive insight into advanced earthquake occurrence estimation techniques.

- Chapter 5, Non-Gaussian EDA Methods for Optimized Earthquake Forecasting.
  This chapter delves into estimating earthquake occurrences, emphasizing the role of periodicity and aperiodicity in seismic events when assuming they stem from deterministic chaos. It also addresses the benefits of employing non-Gaussian methods in earthquake forecasting. Key aspects explored include determining the most suitable distribution for our prior models and the benefits of including model errors and imperfect physics for handling situations with parameter bias. The chapter further investigates the effect of underdispersion in estimating models governed by the rate-and-state friction law. It compares it with standard test beds like the Lorenz 96 presented in Chapter 3.

- Chapter 6, Towards Estimating the Occurrence of Fault-Slip in Large-Scale Frictional Experiments and at the Field Scale.
  Presents the lessons learned from perfect model experiments relevant for application to a large-scale laboratory experiment. We propose a methodology detailing how data assimilation can aid in estimating variables such as the distribution of shear stress at the contact surfaces between frictional bodies. Additionally, we conceptualize how future applications could utilize data assimilation experiments for estimating earthquake occurrences at the field scale. We report on the specific measurements and state variables that are most effective for performing data assimilation at this scale.

- Chapter 7, Conclusions and Recommendations.
  This final chapter summarizes the results, and provides general conclusions about the research presented in this thesis.

- Appendix A, Analysis of the Background Covariances, Localization, and Inflation.
  Covariance localization and inflation are techniques employed in data assimila-

tion to manage spurious correlations and underestimate forecast uncertainties, respectively. In this appendix, I estimate the covariance localization lengths for the Lorenz '96 and Burridge-Knopoff models since they differ depending on the physical model used. I also evaluate the need for covariance inflation for the perfect model experiments that used these numerical models.

- Appendix B, Hyperparameter Selection for the AGMF and the PFF.
  Using the Burridge-Knopoff model in this appendix, I derive the observation operator for systems observing shear stress and explore the implications of quasi-dynamic and fully dynamic assumptions for earthquake forecasting. This analysis studies the importance, effects, and implications of considering the system's acceleration for data assimilation of seismic events.

- Appendix C, Evaluation of the Influence of Observation Frequency and Observation Error in Estimates on a 1-D BK Model.
  This appendix focuses on selecting hyperparameters for this dissertation's non-Gaussian data assimilation methods. I evaluate the optimal bandwidth for the AGMF and the most effective pseudo-time step and kernel bandwidth for the PFF, specifically in the context of systems dominated by rate-and-state friction. This targeted analysis aims to refine these methods for enhanced accuracy in our estimates.

- Appendix D, Derivation of an Observation Operator for the Assimilation of Shear Stress Observations.
  The final appendix discusses the derivation of an observation operator for data assimilation, focusing on shear stress observations in seismic studies. It contrasts quasi-dynamic, single block, and fully dynamic shear stress models, highlighting differences in stress behaviors and the varying observational requirements for accurate state estimation.

# 2

# NUMERICAL MODELING OF EARTHQUAKES

***Summary:*** *This chapter outlines the numerical models employed for forward modeling the earthquake sequences in the data assimilation experiments of this thesis. It delves into the underlying physics, boundary conditions, and the frictional laws applied to generate the earthquakes. This encompasses a detailed explanation of the rate-and-state friction law, discrete models such as the Burridge-Knopoff models, and continuum approximations that assume homogeneous elastic surrounding media, solved using finite difference methods.*

***Resumen:*** *Este capítulo describe los modelos numéricos utilizados en esta tesis para simular las secuencias de terremotos en los experimentos de asimilación de datos. Se profundiza en la física subyacente, analizando las condiciones de frontera y las leyes de fricción empleadas para simular los terremotos. Incluye una explicación detallada de la ley de fricción rate-and-state, aborda modelos discretos como los de Burridge-Knopoff y contempla aproximaciones continuas asumiendo medios elásticos homogéneos circundantes, resueltas mediante métodos de diferencias finitas.*

Parts of this chapter have been published in the methodology sections of Hamed Ali Diab-Montero, Meng Li, Ylona van Dinther, Femke C Vossepoel, Estimating the occurrence of slow slip events and earthquakes with an ensemble Kalman filter, Geophysical Journal International, Volume 234, Issue 3, September 2023, Pages 1701–1721, https://doi.org/10.1093/gji/ggad154.
Changes have been applied to make the text and figures consistent with the thesis.

## 2.1. INTRODUCTION

One of the most important factors for generating earthquakes is the rupture and slip of the faults that produce them (Barbot et al. 2012). For natural earthquakes, this rupture is described as a periodical process called the *seismic cycle* that can be divided into three main phases: an interseismic phase, a coseismic phase, and a post-seismic phase. Fig 2.1 shows a schematic representation of these different phases. The interseismic phase corresponds to the period where strain and stress accumulate steadily, mainly due to tectonic loading. Once the strain and stress levels reach the strength of the fault, there is a sudden slip and liberation of the accumulated energy in the form of heat and seismic waves during the coseismic phase. These seismic waves travel from the faults to the surface, producing the ground motion recorded by seismometers. Finally, after the earthquake propagation comes a period of continuous deformation that can last from days to years, where different mechanical, thermal, and chemical processes restore the strength of the fault before a new seismic cycle starts; this last stage is referred to as the post-seismic phase. The coseismic phase is considerably shorter than the interseismic and post-seismic phases, which last only seconds to minutes. An additional nucleation phase is argued to happen between the interseismic and coseismic phases, based on observations (Beroza and Ellsworth 1996; Ellsworth and Beroza 1995).

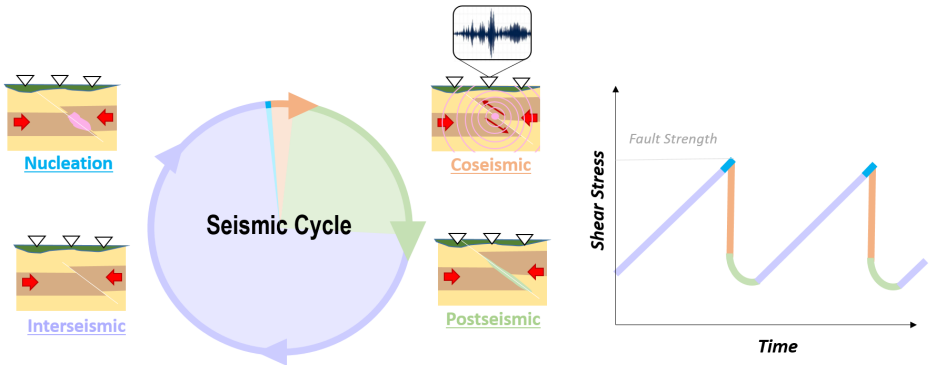


Figure 2.1: Conceptual evolution of the seismic cycle The interseismic phase is highlighted with purple, the nucleation phase is highlighted with blue, the coseismic phase is highlighted with orange, and the postseismic phase with light green.

Different types of fault slip will happen depending on factors like the faults' frictional parameters, the presence of fluids, and lubrication phenomena, among others. These different slip behaviors will produce different seismic cycles or can be aseismic, gradual, and often steady sliding of a fault without generating significant seismic waves in case of no nucleation or propagation of noticeable earthquakes. In the last 20 years, a specific type of aseismic event called Slow-Slip Events (SSEs) has been observed. SSEs are present in mega-thrust fault systems. Their main characteristics are that their stress and strains are released for hours to months instead of seconds to minutes. They are detected through GNSS tools since they are practically undetectable for traditional seismographs. These slow movements lead to a smoother transition of the state variables that dominate

friction in the surfaces where the earthquake nucleates. This smoothness makes them easier to estimate with data-assimilation methods. On the other hand, earthquakes are characterized by a short duration, fast slip rates, and a sharp transition from the interseismic to the coseismic phase. Fig. 2.2 compares the changes in the shear stress and the slip velocities of the slow-slip events with earthquakes. Significant differences are observed in the coseismic phase of the slow-slip events, which have a smoother transition to the following interseismic phase. In comparison, the transition between the seismic phases for earthquakes is faster.

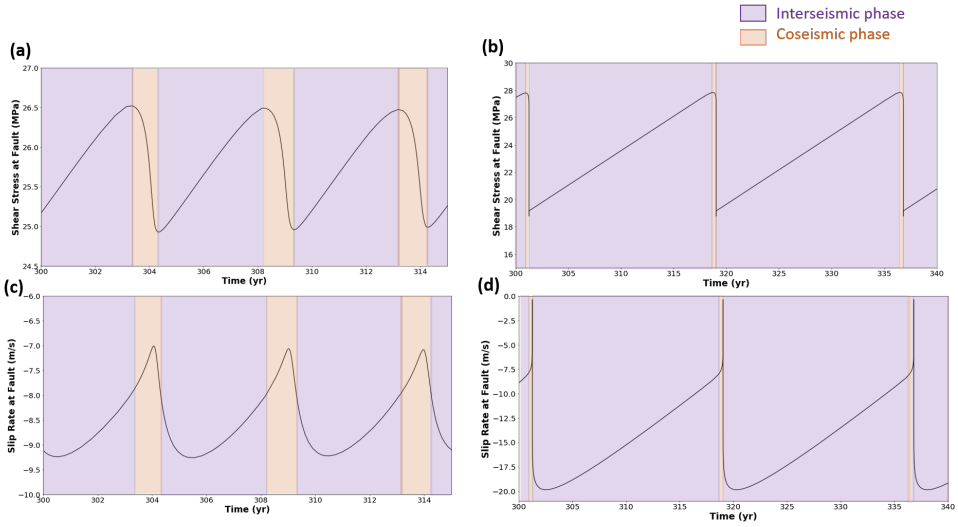


Figure 2.2: Slip rate and shear stress for two cycles: (a) Slip rate evolution and (b) shear stress evolution with time for a fault producing slow-slip events. (c) Slip rate evolution and (d) shear stress evolution with time for fault-producing earthquakes. The interseismic phase is highlighted with purple, and the coseismic phase with orange.

## 2.2. RATE-AND-STATE FRICTION LAW

Seismologists use frictional laws to explain mathematically the nature of the seismic cycle. Generally, we can describe the stress between two frictional surfaces as:

$$\tau = \mu \sigma_n \tag{2.1}$$

where $\tau$ is the shear stress, $\mu$ is the friction coefficient and $\sigma_n$ is the normal stress.

Extensive studies have focused on the dynamics of slip instabilities in laboratory experiments, which are suggested to be influenced by two phenomena: slip weakening and rate weakening. Slip weakening refers to the decrease in a fault's frictional strength as it slips, while rate weakening is the reduction in frictional resistance with an increase in sliding speed (Ruina 1983). However, further improvements to constitutive laws were necessitated by data suggesting that slip-weakening friction alone could not fully account for the relationship between static and dynamic friction, and that friction is not

solely velocity-dependent (Marone 1998). These findings led to the incorporation of a state variable into a more robust friction law, better capturing the dynamics akin to fault behavior during an earthquake rupture. This means that the friction coefficient $\mu$ could be expressed as:

$$\mu = f(V, \theta), \tag{2.2}$$

where $V$ is the slip rate and $\theta$ the "state" is the time of contact between asperities used to explain the healing of the fault and other phenomena. More specifically:

$$\tau = \left[\mu_0 + a\ln\left(\frac{V}{V_0}\right) + b\ln\left(\frac{V_0\theta}{L}\right)\right]\sigma_n, \tag{2.3}$$

where $\mu_0$, is the reference coefficient of friction when the system slips at constant velocity $V_0$, and $a$ and $b$ represent direct and evolution effects, respectively, and $L$ is the characteristic slip distance (Rabinowicz 1951). With time units, the state $\theta$ is a scalar that increases during the interseismic phase (stick), when the asperities are in contact and locked in the fault, and decreases during the coseismic phase (slip). This law is commonly called 'rate-and-state' friction in literature and can be better visualized in Fig. 2.3.

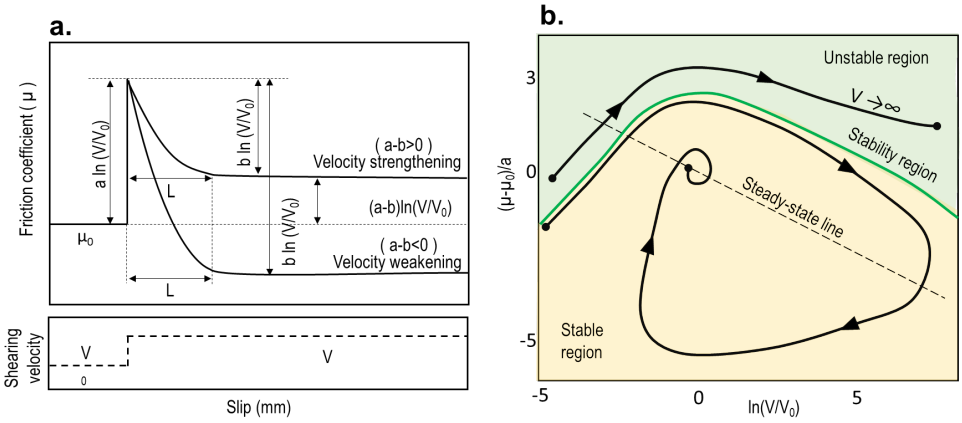


Figure 2.3: Schematic representation of (a) the rate-and-state friction law, and (b) the stability boundaries and trajectories for a system with critical stiffness, governed by the rate-and-state friction and the slip law. Adapted from Rice and Gu (1983) and Segall (2010).

The fault is *velocity-weakening* and potentially frictionally unstable when $a - b < 0$, and *velocity-strengthening* and generally frictionally stable when $a - b > 0$ as shown in in the velocity as a function of slip Fig. 2.3.a. Recent studies have shown that under particular circumstances earthquake can nucleate in the case of velocity strengthening friction ($a - b > 0$) but that is more unusual (Li et al. 2023c).

When the slider moves at constant velocity $V_{ss}$, and $\theta_{ss} = \frac{L}{V_{ss}}$, the system is considered to be in the steady state as shown in the dashed line in Fig. 2.3.b. In the steady state, the stress becomes:

$$\tau_{ss} = \mu_0 - (b - a)\ln(V_{ss}/V_0). \tag{2.4}$$

The steady-state is particularly important for understanding the non-linear dynamics of the systems dominated by rate-and-state friction. The reason is that during the stability analysis of a non-linear system, it is essential to identify and analyze equilibrium points, or *fixed points*, where the system remains constant over time. These fixed points are crucial as they help determine the system's overall behavior. By examining the stability of these points, one can discern whether a system will return to equilibrium after a small perturbation (local stability) or deviate further away (unstable region) as shown in Fig. 2.3.b.

An important concept in the stability analysis of systems governed by rate-and-state friction is the *critical stiffness*, which is the threshold stiffness of a system above which frictional sliding transitions from stable to unstable, potentially leading to sudden slip events like earthquakes (Ruina 1983). This critical stiffness ($k_{crit}$) is defined as:

$$k_{crit} = \frac{(b-a)\sigma_n}{L}. \tag{2.5}$$

If we imagine the fault as a spring-slider system, where the spring has a stiffness $k$, we can define the ratio $k/k_{crit}$. If this $k/k_{crit} > 1$ the system will respond stably to small perturbations. If $k/k_{crit} < 1$ then small perturbations will grow in amplitude, the slip-speed eventually becoming infinite like the unstable region in Fig. 2.3.b. The limit of the stability region is given by $\frac{k}{k_{crit}} = 1$.

**Rate-and-state friction regularized**

Despite empirical justifications, the non-linearity of rate-and-state friction presents difficulties in numerical simulations, leading to stiff differential equations, meaning some variables will change much more rapidly than others, posing a challenge in numerical solutions as very small time steps are required for stability. (Erickson et al. 2008). This non-linearity in the system of equations leads to varying regimes—from periodic solutions to chaotic solutions (Barbot 2019; Erickson et al. 2008; Erickson et al. 2011). Some standard strategies, like using regularized formulations of the rate-and-state friction, help with the stiffness of the system of equations and further modeling challenges at the cost of restricting the solution space to periodic behaviors under the prevalent frictional parameters present in nature (Erickson et al. 2008; Erickson et al. 2011; Lapusta and Rice 2003). However, earthquakes are in nature typically irregular in occurrence, although there is evidence also of quasi-periodic behavior in some earthquakes (Beeler et al. 2001; Williams et al. 2019), which is better explained by deterministic chaotic solutions.

The fault's temporal evolution is modeled as an initial value problem. The slip along the fault is assumed to be governed by a rate-and-state friction formulation, which was proposed based on laboratory friction experiments by Dieterich (1979, 1978) and Ruina (1983) (Eq. 2.6). We employ a regularization near zero slip rate according to Rice (1993) and Ben-Zion and Rice (1997), such that the friction formulation that defines the relation between shear stress $\tau_{\text{fault}}$ and normal stress $\sigma_n$ on the fault is given by

$$\tau = a \operatorname{arcsinh}\left\{\frac{V}{2V_0} exp\left[\frac{\mu_0}{a} + \frac{b}{a}\left(\ln\frac{\theta V_0}{L}\right)\right]\right\}\sigma_n. \tag{2.6}$$

**Evolution laws**

Evolution laws govern the state $\theta$, and the aging and slip laws are prominent (Ruina 1983). While the aging law denotes frictional strength augmentation over time, the slip law ties frictional evolution to slip magnitude (Ampuero and Rubin 2008). Studies have shown that the 1-D BK models, primarily under the slip law, can exhibit periodic, chaotic behaviors and other complex dynamical behavior (Erickson et al. 2011). The slip law is defined by

$$\dot{\theta} = -\left(\frac{V\theta}{L}\right)\ln\left(\frac{V\theta}{L}\right), \tag{2.7}$$

while the aging law (Ruina 1983) is given by

$$\dot{\theta} = 1 - \frac{V\theta}{L}. \tag{2.8}$$

**Radiation damping**

A key strategy in addressing the computational challenges of simulating earthquakes is using *quasi-dynamic* simulations. These simulations focus on slower fault movements under gradually changing stress conditions. This approach efficiently explores aseismic slip and stress buildup, excluding dynamic effects like inertia. Quasi-dynamic models, despite their simplifications, yield results remarkably similar to those of *fully-dynamic* models in low-dimensional scenarios. Fully dynamic models, which are more complex and computationally demanding, incorporate rapid stress changes and dynamic forces, including inertia and wave propagation, crucial for understanding earthquake dynamics and swift fault slip behaviors.

However, these models differ in specific aspects. For instance,quasi-static models struggle with accurately representing energy dissipation, especially in low-dimensional models. This is reflected in the differences in peak stress magnitudes and the velocities attained. To address this limitation, seismologists introduced a *radiation damping term* in the quasi-dynamic models to more accurately simulate the energy loss as seismic waves generate and propagate.

$$\tau = \tau_{RSF} + \eta V \tag{2.9}$$

In this context, the parameter $\eta$ in Eq. 2.6 denotes the radiation damping term used in the quasi-dynamic approximation of inertia (Rice 1993). This term is instrumental in earthquake cycle simulations to manage computational costs effectively (Ben-Zion and Rice 1995; Cochard and Madariaga 1994; Crupi and Bizzarri 2013; Liu and Rice 2007). Nonetheless, it's essential to recognize that incorporating this term introduces qualitative and quantitative differences compared to fully dynamic modeling outcomes (Thomas et al. 2014). The damping viscosity $\eta = G/(2c_s)$, where $G$ is the shear modulus and $c_s$ is the shear wave speed, equates to half the shear impedance of the elastic material surrounding the fault.

**Adaptive time stepping**

As mentioned before, systems governed by the rate-and-state differential equations become stiff, presenting a significant challenge in numerical simulations due to their tendency to cause instability and inaccuracies if not appropriately handled. Typically, if using a small enough time step, these equations can be solved, but using uniformly small time steps across the entire simulation can be highly inefficient, especially since the interseismic phase requires a very low temporal resolution when compared to the coseismic phase, which is instantaneous and where the smallest time steps are needed.

Adaptive time stepping emerges as an effective solution to this challenge. This method dynamically adjusts the size of the time step based on the system's current behavior. We use the formulation used in Herrendörfer et al. (2018) and Li et al. (2022), which are based on Lapusta et al. (2000), that is:

$$\Delta t = min\left\{\zeta \frac{L}{V_{max}}, 1.2\Delta t_{old}, \Delta t_{max}\right\}, \tag{2.10}$$

where $\zeta$ is a factor controlled by the material, the frictional parameters ($a, b, and L$), the normal stress $\sigma_n$, and the grid size. $V_{max}$ is the maximum slip rate, $t_{old}$ is the previous time step, and $t_{max}$ is the maximum stable time step, generally chosen for resolving the interseismic phase efficiently.

## 2.3. SEISMOLOGICAL MODELS

Significant advances have been made in recent years in the study of seismogenesis. Efforts have concentrated on understanding how fault states control the earthquake nucleation, propagation and arrest (e.g., Ellsworth and Beroza 1995; Faulkner et al. 2010; Galis et al. 2017; Rubin and Ampuero 2005; Wibberley et al. 2008). This led to more realistic, physics-based numerical models (Lapusta et al. 2019), which resolve dominant physical processes throughout different phases of the seismic cycles (Barbot 2019; Herrendörfer et al. 2018; Lapusta and Liu 2009). Due to the usually simplified modeling assumptions, it remains highly challenging to apply such physics-based models to improve short- to medium-term seismic hazard assessment; there are nevertheless progressive efforts to do so (e.g., Barbot et al. 2012; Dal Zilio et al. 2019). Earthquake simulators have been developed to model earthquake sequences over complex regional fault networks for thousands of years by assuming significant simplifications in solution procedures and physical processes. These models can (re-)produce statistical observations such as the Gutenberg-Richter and Omori laws. They also shed light on the feasibility of combining physical models with probabilistic seismic hazard assessments (Dieterich and Richards-Dinger 2010; Shaw et al. 2018). However, one of the most significant difficulties that remain in modeling upcoming sequences of slip on a fault system is that its current conditions (i.e., state of stress and strength) are unknown and cannot be directly measured (Barbot et al. 2012; Van Dinther et al. 2013; van Dinther et al. 2019).

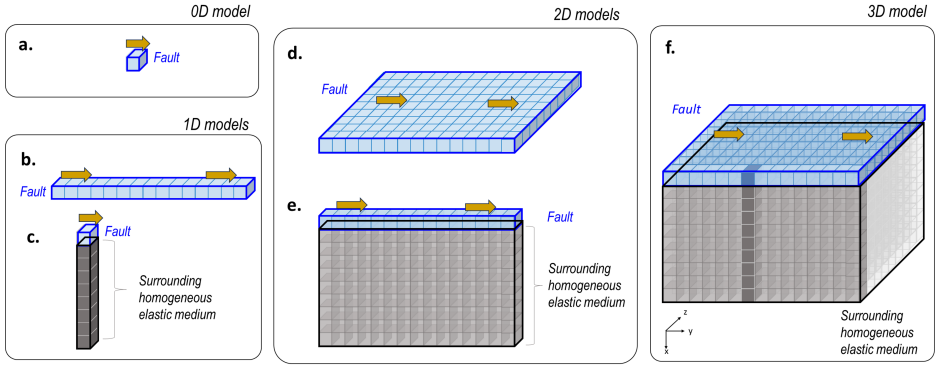


Figure 2.4: Schematic representation of the seismological models used governed by the rate-and-state friction law. The models represent a horizontal straight fault (blue) surrounded by a homogeneous elastic medium (gray) where (a) is a 0D model of the fault, (b) a 1D model along slip direction, (c) a 1D model perpendicular to the slip direction, (d) a 2D model of the fault surface, (e) a 2D model perpendicular to the fault at the face of the block, and (f) is the 3D model.

As we have seen in Section 2.2, many complex dynamics derive from the rate-and-state friction law and numerical challenges for correctly resolving the different seismic phases. For this reason, in this thesis, we focus on studying the stick-slip phenomenon in a simple geometry of a horizontal straight type of fault mimicking a direct shear setup as the ones used in laboratory experiments and explained in Chapter 6. I use models of different dimensions as shown in Fig. 2.4. We can divide these models into two types: the first type, where we only model the frictional surface(Fig. 2.4.a, b and d), and the second type, where the frictional surface, the fault, is surrounded by a homogeneous elastic medium (Fig. 2.4.c, e and f). Our investigation confines the models to a velocity-weakening regime. This particular regime results in the emergence of stable limit-cycle solutions. Table 2.1 provides a summary of relevant references of data assimilation studies that have used the geometries and types of models shown in Fig. 2.4, the table also mentions the friction law they used, the seismic events they simulated, the numerical method used for these simulations, and in which phase of the seismic cycle they focused their estimates and analysis.

### BURRIDGE-KNOPOFF MODELS

The Burridge-Knopoff (BK) model is a simplified benchmark of earthquake sequences, characterized by a spring-block slider system (Burridge and Knopoff 1967). In our study, the 1-D BK model comprises multiple blocks connected by elastic springs with stiffness $k_\mu$, depicted in Fig. 2.5. These blocks are elastically coupled (with stiffness $k_\lambda$) to a rigid plate moving at speed $v_p$ across a frictionally rough surface, serving as an analog for a 1-D earthquake fault (Carlson et al. 1991). The equation governing their movement is:

$$m\ddot{x}_i = k_\mu \left(x_{i+1} - 2x_i + x_{i-1}\right) - k_\lambda \left(x_i - v_p t\right) - F_i, \tag{2.11}$$

where $x_i$ denotes the position of the ith block. The BK model's non-linearity originates from the friction law used as a force for earthquake modeling.

Table 2.1: Relevant data assimilation studies of seismological numerical models coupled with friction laws with different dimensions

| **Dimension** | **Characteristics** | **Relevant References** | **Friction and Evolution Laws** | **Focus Seismic Phase** | **Seismic Events** | **Numerical Method** |
|---|---|---|---|---|---|---|
| 0D | 0D fault cell. Fig. 2.4.a. | Banerjee et al. (2023) | Rate-and-state and slip law | Interseismic, Coseismic | Slow slip events and earthquakes | Runge-Kutta 4th order |
| 1D | 1D fault line. Along slip direction. Fig. 2.4.b. | Chapter 5 | Rate-and-state and slip law | Interseismic, Coseismic | Slow slip events and earthquakes | Runge-Kutta 4th order |
| | 0D fault cell with 1D surrounding medium. Perpendicular to slip direction. Fig. 2.4.c. | Chapter 4, Diab-Montero et al. (2023) | Rate-and-state and aging law | Interseismic, Coseismic | Slow slip events and earthquakes | Finite Differences |
| 2D | 2D fault plane. Fig. 2.4.d. | Kano et al. (2020, 2015, 2013) | Rate-and-state and aging law | Postseismic | Slow slip events | Runge-Kutta 4th order |
| | 1D fault line with 2D surrounding medium. Fig. 2.4.e. | Chapter 6 | Rate-and-state and aging law | Interseismic, Coseismic | Slow slip events | Finite Differences |
| | | van Dinther et al. (2019) | Slip-rate-dependent friction and aging law | Interseismic, Coseismic | Slow slip events | Finite Differences |
| 3D | 2D fault plane with 3D surrounding medium. Fig. 2.4.f. | Hirahara and Nishikiori (2019) | Rate-and-state and aging law | Interseismic, Coseismic | Slow slip events | Boundary Element Method |

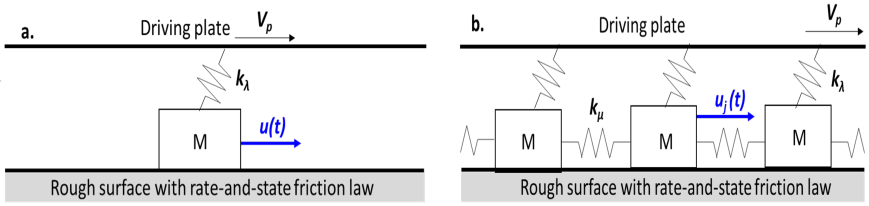


Figure 2.5: Schematic representation of the Burridge Knopoff models coupled with the rate-and-state friction law (a) a spring-loaded plate, representing the fault's other side, is coupled with a single degree of freedom slider block.(b) spring connected chain of blocks, elastically coupled to a driver plate moving at a constant velocity Vp. The friction force, which keeps the slider stationary, is controlled by a rate-state friction law, and this force acts on a rough surface where the blocks experience slipping.

This research adopts the 1-D BK system modeling methodology of Erickson et al. (2011). The system of ordinary differential equations (ODEs) used is:

$$\begin{aligned}
\dot{\bar{u}}_i &= \bar{V}_i, \\
\dot{\bar{V}}_i &= \gamma_\mu^2 (\bar{u}_{i-1} - 2\bar{u}_i + \bar{u}_{i+1}) - \gamma_\lambda^2 \bar{u}_i - \frac{\gamma_\mu^2}{\xi}\left(\bar{f} + \bar{\Theta}_i + \ln\left(\bar{V}_i + 1\right)\right), \\
\dot{\bar{\Theta}}_i &= -\left(\bar{V}_i + 1\right)\left(\bar{\Theta}_i + (1+\epsilon)\ln\left(\bar{V}_i + 1\right)\right).
\end{aligned} \tag{2.12}$$

To achieve this non-dimensional system of ODEs, several modifications and simplifications were applied, including adopting the non-dimensional variables from Erickson et al. (2008), Erickson et al. (2011), and Madariaga (1998). The variables and parameters have specific non-dimensionalized interpretations, for example, $\bar{\Theta} = \frac{b}{a}\ln\left(\frac{V_0\theta}{L}\right)$ and $\bar{V} = \frac{V}{V_0}$. A summary of all the non-dimensionalized parameters is shown in Table 2.2.

Table 2.2: Non-dimensional parameters derived from the Burridge-Knopoff models following Erickson et al. (2011)

| **Non-dimensional parameter** | **Formula** | **Description** |
|---|---|---|
| $\bar{f}$ | $\frac{\mu_0}{a}$ | Scaled steady-state friction coefficient |
| $\epsilon$ | $\frac{b-a}{a}$ | Sensitivity of the velocity relaxation |
| $\gamma_\mu$ | $\sqrt{\frac{k_\mu}{m}}\left(\frac{L}{V_0}\right)$ | Non-dimensional frequency with respect to the blocks |
| $\gamma_\lambda$ | $\sqrt{\frac{k_\lambda}{m}}\left(\frac{L}{V_0}\right)$ | Non-dimensional frequency with respect to the plate |
| $\xi$ | $\frac{k_\mu L}{a\sigma_n}$ | Non-dimensional spring constant |

The non-dimensional shear stress using the non-dimensional variables $(\bar{v}, \bar{\Theta})$ is expressed as:

$$\bar{\tau}_i = \bar{f} + \ln\left(\bar{V}_i + 1\right) + \bar{\Theta}_i \tag{2.13}$$

The logarithmic term $\ln(\bar{v}+1)$ introduces challenges, leading to numerical stiffness in the system, as indicated by exceptionally large negative eigenvalues in the local Jacobian (Erickson et al. 2008; Noda et al. 2009; Rojas et al. 2009). This stiffness requires very small time steps when using standard numerical techniques to achieve stable solutions. Even with implicit numerical methods, the time step remains limited by accuracy needs (Erickson et al. 2008; Erickson et al. 2011). A step that is too large results in an undefined logarithmic term. Consequently, we used an embedded fourth-order explicit Runge-Kutta method with a minimal step size for the ODEs.



### FINITE DIFFERENCE MODELS

In this thesis, we use a fully staggered finite differences approach for solving the schematic setups of the type shown in Fig. 2.4c and 2.4e by adopting the C++ library for non-linear coupled problems *Garnet* (Pranger 2020). Our numerical model setup is inspired by a large-scale biaxial friction apparatus consisting of two sandstone block specimens in a direct-shear configuration (Fig. 2.6). We assume a symmetric setup and model only the lower half-space. To limit computational resources, we discretize for Chapter 4 a 1D line across the block boarded by a single, 0D fault point, which provides a reasonable approximation when evaluating temporal estimates of the slip occurrences. In Chapter 6, we extend the setup to a 2D-model of the type of 2.4, and give further details about this large-scale setup used by the National Research Institute for Earth Science and Disaster Prevention in Japan (Fukuyama et al. 2014; Spiers et al. 2017).

The numerical method, conservation, and constitutive equations used are adopted from and described in detail in Li et al. (2022). Below, we summarize the equations relevant to our forward model setup which omit the effects of both gravity and inertia. The equation governing momentum balance under the assumption of static stress transfer is given by

$$\nabla \cdot \boldsymbol{\sigma} = 0, \tag{2.14}$$

where $\boldsymbol{\sigma}$ denotes the Cauchy stress tensor. In this tensor, the component $\sigma_{i,j}$ represents the stress along the $x_j$ axis on a plane orthogonal to the $x_i$ axis, for $i, j = 1,2,3$. Hooke's law connects the stress rate $\dot{\boldsymbol{\sigma}}$ with the strain rate $\dot{\boldsymbol{\epsilon}}$ as follows:

$$\dot{\boldsymbol{\sigma}} = 2G\dot{\boldsymbol{\epsilon}} + \lambda Tr(\dot{\epsilon})\boldsymbol{I}, \tag{2.15}$$

where $G$ is the shear modulus, $K$ is the bulk modulus, and $\lambda$ is defined as $K - \frac{2G}{3}$ (Lame's constant), with $\boldsymbol{I}$ representing the identity tensor. The trace of the matrix, $Tr(\dot{\epsilon})$, is given by $\dot{\epsilon}_{kk}$. The assumption of an infinitesimal strain rate $\dot{\boldsymbol{\epsilon}}$ is described by the equation

$$\dot{\epsilon} = \frac{1}{2}\left(\nabla \boldsymbol{v} + \boldsymbol{v}\nabla\right), \tag{2.16}$$

in which $\boldsymbol{v}$ symbolizes the material velocity, and $v_i$ specifies the velocity component in the $x_i$ direction (with $i = 1,2,3$). The axes are interchangeably referred to as $(x_1, x_2, x_3)$ and $(x, y, z)$ as in Fig. 2.4f.

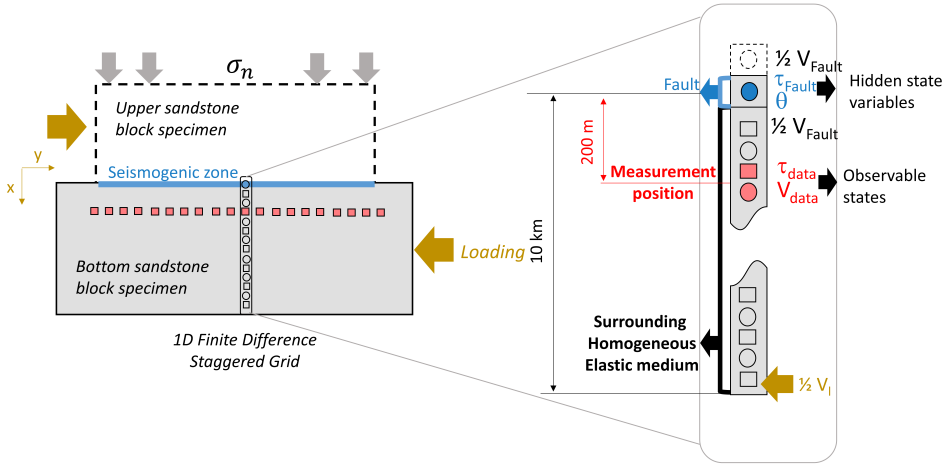


Figure 2.6: Schematic representation of the 1D model used to represent a horizontal straight fault. The shear stress ($\tau$) is estimated in the nodes represented by circles and the velocity($v_y$) in the ones represented by squares. The fault (blue) corresponds to the uppermost node of the grid and the location where the shear stress, slip rate, and state $\theta$ are estimated. The spacing between nodes is 20 m.

For a fault with unit normal vector $\hat{\boldsymbol{n}}$, the (scalar) normal stress $\sigma_n$ (positive in compression) is given by the projection $\sigma_n = -\hat{\boldsymbol{n}} \cdot \sigma \cdot \hat{\boldsymbol{n}}$, the shear traction vector $\boldsymbol{\tau}_s$ by the projection $\boldsymbol{\tau}_s = \boldsymbol{\sigma} \cdot \hat{\boldsymbol{n}} + \sigma_n \hat{\boldsymbol{n}}$, the scalar shear traction $\tau_s$ by the Euclidean norm $\tau_s = \|\boldsymbol{\tau}_s\|$, and finally the unit fault tangent $\hat{\boldsymbol{t}}$ (which defines the orientation of the scalar fault slip V) by the normalization $\hat{t} = \boldsymbol{\tau}_s / \tau_s$, such that $\tau_s = \hat{\boldsymbol{t}} \cdot \boldsymbol{\sigma} \cdot \hat{\boldsymbol{n}}$.

The two-dimensional (2D) framework, used in Chapter 6, is obtained by omitting the fault direction perpendicular to the strike. This simplification implies that material and frictional properties and boundary and initial conditions are considered homogeneous in the direction perpendicular to the strike. Consequently, this approach does not account for the perpendicular strike direction heterogeneity. Besides, we exclude the motion perpendicular to the dip, denoted by $v_z$, focusing solely on modeling the motion along the dip. As a result, this specific model for along-dip strain is required to assess only the $\sigma_{xy}$ and $\sigma_{yz}$ components of the stress tensor.

This simplified physical Eqs. 2.15 and 2.16 in 2D read:

$$\begin{aligned} \dot{\sigma_{xy}} &= G \frac{\partial v_y}{\partial x}, \\ \dot{\sigma_{yz}} &= G \frac{\partial v_x}{\partial z}, \\ \frac{\partial \sigma_{xy}}{\partial x} &+ \frac{\partial \sigma_{yz}}{\partial z} = 0. \end{aligned} \tag{2.17}$$

Adhering to the simplifications outlined in Li et al. (2022), the physical equations are

formulated directly in one dimension (1D) for Chapter 4. In this simplified model, all variables are constant along both dip and strike, leading to a scenario where only the shear stress component $\tau_{xy}$ and the velocity component $v_y$ need to be solved. Consequently, the fault is reduced to a zero-dimensional (0D) point at $x = 0$ within the computational domain $\Omega(x) = [0, H]$, where $H$ denotes the distance from the fault interface to the model's base. The fault point in this 1-D model is characterized as velocity-weakening to facilitate the nucleation of earthquake sequences. This model is obliged to fulfill the momentum balance equation and Hooke's law of elasticity, which are represented as follows:

$$\begin{aligned} \dot{\tau}_{xy} &= G\frac{\partial v_y}{\partial x}\,, \\ \frac{\partial \tau_{xy}}{\partial x} &= 0\,. \end{aligned} \tag{2.18}$$

Here, $G$ is the shear modulus.

We assume boundary conditions corresponding to the values of the velocity in both extremes of the model

$$\begin{aligned} v_y(x=0) &= \frac{1}{2}V, \\ v_y(x=H) &= \frac{1}{2}V_l. \end{aligned} \tag{2.19}$$

Our assumption about the symmetry of the setup allows us to consider the velocity in the fault interface as half of the total slip rate $V$ and the one at the bottom of the model as half the loading rate $V_l$.

The initial conditions are selected to enable the fault to undergo steady-state creep at the prescribed slip velocity $V_l$ at the initial time $t = 0$. These conditions are formally expressed as

$$\begin{aligned} V(t=0) &= V_l\,, \\ \theta(t=0) &= \frac{L}{V_l}\,, \end{aligned} \tag{2.20}$$

where $V(t = 0)$ represents the velocity at the initial time, set equal to $V_l$, and $\theta(t = 0)$ denotes the initial state variable, calculated as the ratio of characteristic length $L$ to the slip velocity $V_l$.

## 2.4. FRICTIONAL PARAMETERS, APERIODICITY AND STABILITY

In this chapter, we explore how frictional parameters are crucial in determining different slip behaviors in fault systems. These parameters are key in deciding whether a simulation will be stable or unstable, as explained by the concept of critical stiffness from a non-linear dynamic analysis perspective. They also play a significant role in the numerical modeling process, particularly in choosing the right time step for solving differential equations and affecting the criteria for adaptive time stepping. This aspect is crucial for the convergence of simulation codes, where dealing with the uncertainty and bias

of these parameters becomes a significant challenge. Especially in systems with rate-and-state friction, small changes in frictional parameters can lead to major differences in simulation results, which we will cover in this section.

**Influence of parameters a and b**

The Burridge-Knopoff model, integrated with rate-and-state friction, is a valuable tool for assessing the diverse behaviors arising from variations in frictional parameters. This model's utility is further enhanced when combined with the non-dimensional variables listed in Table 2.2, which are instrumental in understanding these behaviors. In this context, we specifically focus on the impact of the a and b parameters, utilizing the non-dimensional parameter $\epsilon$ to gauge the sensitivity of velocity relaxation. As previously discussed, the difference (a-b) dictates whether the system exhibits velocity weakening or strengthening.

For instances where $\epsilon$ is positive, indicative of velocity weakening, the system's behavior varies based on the magnitude of a, as demonstrated in Fig. 2.7. This figure illustrates the recurrence intervals from 100 simulations of a 1D Burridge-Knopoff model while altering $\epsilon$. Observations reveal that for $\epsilon$ values less than 0.2, the system tends to produce periodic events with very low velocities, similar to Slow Slip Events (SSEs). In contrast, higher $\epsilon$ values lead to deterministic chaos and aperiodicity. As $\epsilon$ approaches and exceeds 0.7, the system experiences rapid ruptures.

It's crucial to note that the *a* and *b* parameters also influence the time stepping of the simulation. A smaller step size becomes necessary, particularly as $\epsilon$ increases. Fig. 2.8 shows that higher $\epsilon$ values require a significantly smaller time step to maintain stable simulation. This relationship underscores the importance of carefully selecting these parameters, as they directly impact both the behavior of the model and the practical aspects of its computational implementation.

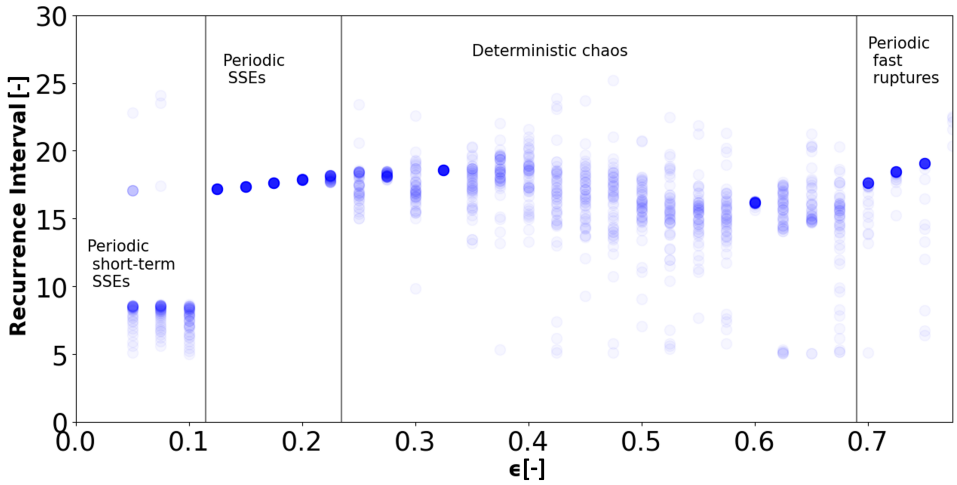


Figure 2.7: Changes on the recurrence intervals for different values of $\epsilon$. The blue circles correspond to the recurrence interval for each of the 100 events of a single simulation done with a 1-D Burridge Knopoff model for a given value of $\epsilon$. The slip behavior changes from periodic solutions of SSEs and fast ruptures to deterministic chaos depending on the ratio of parameters a and b.

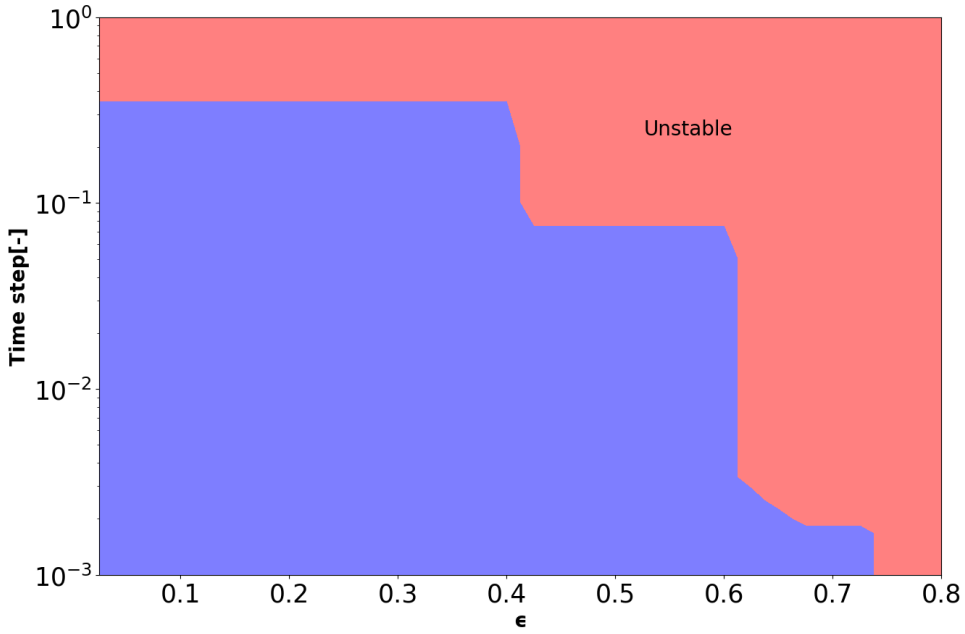


Figure 2.8: Stability analysis of simulations of a 1-D Burridge Knopoff model for different values of $\epsilon$. The blue hatched area corresponds to stable simulations for the given value of $\epsilon$, and the red hatched area corresponds to the simulations that crashed or became unstable due to the time step being too long.

**Influence of parameter L**

The critical slip distance (L) in the rate and state friction law significantly influences the frictional behavior of faults or rock interfaces. Smaller L values make the system more sensitive to slip rate variations, leading to rapid state variable ($\theta$) evolution, quicker transitions between frictional behaviors, such as stable sliding, slow slip, or unstable slip (e.g., earthquake nucleation) and potentially easier earthquake nucleation. In contrast, larger L values result in less sensitivity to slip rate variations, slower state variable evolution, and possibly more stable sliding or delayed earthquake nucleation.

The non-dimensional parameter $\xi$ is an interesting value from the Burridge-Knopoff model since it is very similar to the quantity $k/k_{crit}$ mention in Section 2.2. Therefore, a large uncertainty in this variable can bring problems in simulations given that some simulation may be unstable while the stable ones can have a great variability of behaviors due to the influence of parameter $a$ on, for example, the sensitivity to the velocity relaxation as shown in Fig. 2.7. Therefore, an analysis of the ratio $\sigma_n/L$ can bring a more controlled comparison of the influence of the parameter $L$. Herrendörfer et al. (2018) makes a sensitivity analysis by changing the values of $\sigma_n$, called $P$ in their work, in a 2D model of a strike slip fault. They observe that for that variations of $L$ can also produce different slip behaviours as shown in Fig. 2.9. We see that larger values of $L$ are required having aperiodicity and that slow slip events are just a narrow band in the wide variety of solutions, which highlights the importance of expanding the studies of data assimilation on seismic events to other systems with fast slip rates as we do in Chapter 4.

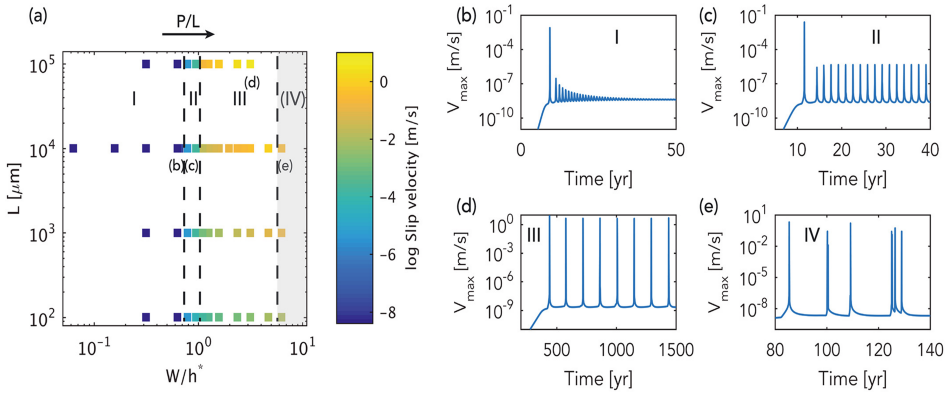


Figure 2.9: (a) Sensitivity of the slip behavior for a strike-slipe fault to variations in the parameter L, (b) stable sliding, (c) slow slip events or periodic aseismic slip transients, (d) periodic seismic events, (e) aperiodic seismic slip. Adopted from Herrendörfer et al. (2018).

## 2.5. DISCUSSION

In this chapter, we have navigated the intricate task of numerically modeling the frictional behavior of faults. This complexity persists even with simple geometries, assumptions of elasticity in the surrounding medium, and the use of low-dimensional models. The non-linear dynamics inherent to rate-and-state friction models offer significant advantages, given their ability to replicate various slip behaviors observed in laboratory experiments using a relatively small set of frictional parameters. However, this same non-linearity also introduces substantial challenges, particularly due to the stiffness of the system of differential equations and the profound impact of uncertainties in frictional parameters on the stability of simulations.

These challenges extend to data assimilation in systems modeled after earthquakes. The process of constraining physical parameters and regularizing numerical models to enhance stability and convergence can inadvertently limit the system to periodic solutions. Such a restriction might affect the effectiveness of data assimilation, a topic we delve into in Chapter 5. Furthermore, there's a clear need for more research to understand the assimilation of different slip rates, as currently undertaken and discussed in Chapter 4. This exploration is crucial for enhancing our understanding and modeling capabilities of earthquake dynamics and the broader geophysical processes they entail.

### DATA AVAILABILITY

The forward model for (a)seismic slip sequences utilizing the numerical modelling package Garnet is made accessible via repository https://bitbucket.org/cpranger/garnet/src/master/. The forward model for the 1-D Burridge Knopoff coupled with rate-and-state friction is made accessible via repository https://github.com/hdiab montero/non_gaussian_data_assim/. The data produced and analyzed in this chapter is available via 4TU.ResearchData (http://doi.org/10.4121/f0f075f2-f45c-4f8c-9d1d-bde03baeae33).

# 3

# Ensemble Data Assimilation Methods

***Summary:*** *This chapter introduces the fundamentals of data assimilation, including the specific notation used for data assimilation within the scope of this thesis. It highlights the different ensemble data assimilation methods explored, namely the Ensemble Kalman Filter (EnKF), the Adaptive Gaussian Mixture Filter (AGMF), and the Particle Flow Filter (PFF). Additionally, the chapter presents techniques such as inflation and localization, which address issues related to spurious correlations and low-rank deficiencies. Finally, we test the methods using a Lorenz 96 model with 20 cells to highlight their differences.*

***Resumen:*** *Este capítulo introduce los fundamentos de la asimilación de datos, incluida la notación específica utilizada a lo largo de esta tesis. Además, se destacan los diversos métodos de asimilación de datos de ensamble utilizados, tales como el Filtro de Kalman de ensamble (EnKF), el Filtro Adaptativo de Mezcla Gaussiana (AGMF) y el Filtro de Flujo de Partículas (PFF). El capítulo también expone técnicas como la inflación y la localización, que son empleadas para solucionar problemas asociados a correlaciones espurias y deficiencias en el rango de las matrices de covarianza. Finalmente, se comparan y evalúan las diferencias entre estos métodos utilizando un modelo Lorenz 96 de 20 celdas.*

Parts of this chapter have been published in the methodology sections of Hamed Ali Diab-Montero, Meng Li, Ylona van Dinther, Femke C Vossepoel, Estimating the occurrence of slow slip events and earthquakes with an ensemble Kalman filter, Geophysical Journal International, Volume 234, Issue 3, September 2023, Pages 1701–1721, https://doi.org/10.1093/gji/ggad154 .
Changes have been applied to make the text and figures consistent with the thesis.

## 3.1. INTRODUCTION

*Data assimilation* is an approach that helps better estimate a system's evolution by combining knowledge of a dynamic system with observations thereof. We represent variables (or parameters) that we want to estimate as a vector that we call the *state vector*,

$$\mathbf{z}^T = \left(\mathbf{z}_\psi{}^T, \mathbf{z}_\alpha{}^T\right), \tag{3.1}$$

where $\mathbf{z}$ is the state vector, $\mathbf{z}_\psi$ represent the *states* of the system, and $\mathbf{z}_\alpha$ represent the *parameters* or physical properties of the system. When applied sequentially, the optimal estimation of the variables is described as a two-step process. The first step is called the *forecast step*, in which we use our knowledge of the dynamics of the system to move it forward from a previous time $t_{c-g1}$ to a future time $t_c$ using the following formulation,

$$\mathbf{z}_c = \mathscr{M}_{c:c-1}\left(\mathbf{z}_{c-1}\right) + \boldsymbol{\eta}_c, \tag{3.2}$$

where $\mathscr{M}_{c:c-1}$ represents the forward model operator from time $t_{c-1}$ to $t_c$, $\mathbf{z}_{c-1}$, represents the state vector at time $t_{c-1}$, and $\boldsymbol{\eta}_c$ represents the model error which includes missing physics or uncertainty involved in the forward modeling (parameter uncertainties, grid resolution, uncertain initial conditions, etc).

The second step is the *analysis step*, which is formulated based on Bayes' theorem when data or observations of the system have been made:

$$p(\mathbf{z}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{z})\,p(\mathbf{z})}{p(\mathbf{d})}, \tag{3.3}$$

where $p(\mathbf{z})$ represents our prior scientific knowledge of the state vector and its uncertainties, $p(\mathbf{d}|\mathbf{z})$ represents the likelihood of the observations $\mathbf{d}$, given the prior state vector $\mathbf{z}$. $p(\mathbf{d})$ is referred to as the evidence and acts as a normalization factor. $p(\mathbf{z}|\mathbf{d})$ is the posterior estimate of the state of the system given the observed data. In this step, we update our knowledge of the system using those observations.

We represent the observation vector $\mathbf{d}$ with:

$$\mathbf{d}_c = \mathscr{H}_c(\mathbf{z}_c) + \boldsymbol{\varepsilon}_c, \tag{3.4}$$

where $\mathbf{d}_c$ is the vector containing the observations taken at time $t_c$, $\mathscr{H}_c$ is the non-linear observation operator and $\boldsymbol{\varepsilon}_c$ are the measurement errors.

One of the benefits of data assimilation is that it helps to estimate states at a location of interest that is not easily accessible to be measured directly. These states receive the name of *hidden states*. They are estimated using our physical knowledge of the interaction and relationship between the non-observable and the observable locations in the model to transfer the corresponding correction once we update our knowledge of the observable locations.

## 3.2. METHODS WITH GAUSSIAN PRIOR ASSUMPTIONS

**The Ensemble Kalman Filter**

For the state estimation in this study, we use a stochastic variant of the Ensemble Kalman Filter (e.g. Evensen 2003), an ensemble-based data assimilation and a Monte

Carlo implementation of the least-squares solution of the Bayesian update problem presented in Eq. 3.3. The EnKF optimally combines the information from the forward numerical model (prior) and its deviation to the observations (likelihood) to produce a posterior state vector estimation. The prior, likelihood and posterior probability density functions (pdfs) are approximated by an ensemble of our model's different initial conditions, states, or parameters. These distributions are assumed to be Gaussian. The ensemble representation of our state vector, which can contain the state and the parameters, is as follows:

$$\mathbf{z}_n^T = \left(\mathbf{z}_\psi^T, \mathbf{z}_\alpha^T\right)_n, \qquad 1 \leq n \leq N_m, \tag{3.5}$$



where the subscript $n$ refers to a single realization in an ensemble containing $N_m$ realizations (with $n = 1, ..., N_m$). The prior, is defined as $\mathbf{z}_n^f \sim \mathcal{N}(\overline{\mathbf{z}_n^f}, C_{zz}^f)$. The superscript $f$ stands for forecast and indicates the prior information coming from the forward numerical model, representing our knowledge about the physics of the problem. The covariance $C_{zz}^f$ quantifies the relationship between variables and the uncertainties of the given states. The covariance is defined as follows:

$$\mathbf{C}_{zz}^f = \frac{1}{N-1}\left(\mathbf{z}_n^f - \overline{\mathbf{z}_n^f}\right)\left(\mathbf{z}_n^f - \overline{\mathbf{z}_n^f}\right)^T. \tag{3.6}$$

In ensemble data, assimilation methods like the Ensemble Kalman Filter techniques, such as localization and covariance inflation, are crucial to mitigate challenges posed by a limited ensemble size and low-rank covariance matrices. Limited ensemble size can lead to spurious long-distance correlations and underestimated forecast errors, both of which degrade the quality of the assimilation. Localization techniques dampen these non-physical, long-distance correlations, ensuring a more localized and physically consistent influence of observations. There are several approaches for applying this localization. We directly use localization on the prior covariance matrix via a Schur product:

$$\mathbf{C}_{zz}^f \leftarrow \rho_{i,j} \circ \mathbf{C}_{zz}^f \tag{3.7}$$

where

$$\rho_{i,j} = exp\left\{-\left(\frac{i-j}{r_{in}}\right)^2\right\} \tag{3.8}$$

and $i, j = 1, ..., N_c$. $N_c$ is the number of cells of the model, and $r_{in}$ is the decorrelation, which depends on the model type used.

Covariance inflation, on the other hand, addresses the issue of underestimated forecast errors, which can cause over-confidence in the model forecast and potentially lead to filter divergence. By artificially increasing the analysis error covariance, inflation helps balance the weight between model forecast and observations, maintaining stability in the assimilation process. We also use inflation

$$\mathbf{C}_{zz}^a \leftarrow \rho_{infl}\, \mathbf{C}_{zz}^a \tag{3.9}$$

where $\rho_{infl}$ is the inflation factor, which is slightly larger than one.

We utilize the perturbed observations scheme, which involves updating each ensemble member with a perturbed observation. It is assumed that the observational errors

follow a Gaussian distribution represented by $\epsilon_n \sim \mathcal{N}(0, C_{dd})$. Additionally, we assume we have uncorrelated errors in our states. The perturbed observation vector for each ensemble member is calculated as follows:

$$\mathbf{d}_n = \mathbf{d} + \boldsymbol{\varepsilon}_n, \qquad 1 \leq n \leq N_m, \tag{3.10}$$

$$\mathbf{C}_{dd} = \frac{1}{N-1} \sum_{n=1}^{N} \boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^T. \tag{3.11}$$



The EnKF uses the prior distribution $\mathbf{z}_n^f$, the observation vector $\mathbf{d}_n$ and their covariance matrices for estimating the posterior distribution $\mathbf{z}_n^a$ calculated in the analysis step using the following expression:

$$\mathbf{z}_n^a = \mathbf{z}_n^f + \mathbf{K}\left[\mathbf{d}_n - \mathbf{H}\mathbf{z}_n^f\right], \qquad 1 \leq n \leq N_m, \tag{3.12}$$

where the superscript $a$ stands for analysis, $\mathbf{K}$ is the Kalman gain matrix, and $\mathbf{H}$ is the linear observation operator matrix. The Kalman gain is interpreted as the relative weight given to the observations information and current state estimate, and it is given by:

$$\mathbf{K} = \mathbf{C}_{zz}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{C}_{zz}^f \mathbf{H}^T + \mathbf{C}_{dd}^f\right)^{-1} \tag{3.13}$$

A high Kalman gain places a higher weight on the observations, making the analysis follow them more closely. A low Kalman gain imposes a higher weight on the prediction and analysis that follows the prediction more closely. Further details can be found in Evensen (2003) and Evensen et al. (2022).

## 3.3. METHODS WITH NON-GAUSSIAN PRIOR ASSUMPTIONS

**Adaptive Gaussian Mixture Filter**

The Adaptive Gaussian Mixture Filter(AGMF) is a bridging formulation that allows to smoothly transit between an ensemble Kalman filter and a particle filter analysis update (Stordal et al. 2011; Stordal and Lorentzen 2014; van Leeuwen et al. 2019). The AGMF computes a two-step update for the analysis step: in the first step we update the ensemble members and the covariance matrix using a very similar formulation as in Eq. 3.12 but with a dampened background covariance matrix as follows:

$$\mathbf{z}_n^a = \mathbf{z}_n^f + h^2 \mathbf{C}_{zz}^f \mathbf{H}^T \left(h^2 \mathbf{H}\mathbf{C}_{zz}^f \mathbf{H}^T + \mathbf{C}_{dd}^f\right)^{-1} \left[\mathbf{d}_n - \mathbf{H}\mathbf{z}_n^f\right], \qquad 1 \leq n \leq N_m, \tag{3.14}$$

where $h$ is a so-called bandwidth parameter and $h \in [0,1]$. The linear update is a function of $h$ with no update for $h = 0$ and the same update as the EnKF when $h = 1$ (Stordal and Lorentzen 2014; van Leeuwen et al. 2019). Besides, we directly apply the same localization and inflation factors used for the EnKF to the prior covariance matrix $\mathbf{C_{zz}}$ .

The second step involves implementing importance sampling in the estimation by assigning weights to the ensemble members according to:

$$\boldsymbol{w}_t^n = \Phi\left(\mathbf{d}_n - \mathbf{H}\mathbf{z}_n^f, h^2\mathbf{H}\mathbf{C}_{zz}^f\mathbf{H}^T + \mathbf{C}_{dd}^f\right)\boldsymbol{w}_{t-1}^n, \tag{3.15}$$

where $\Phi(x-\mu, P)$ denotes a multivariate Gaussian density with mean $\mu$ and covariance matrix $P$.

These weights are normalized so that the sum of the weights is equal to one:

$$\bar{\boldsymbol{w}}_t^n = \frac{\boldsymbol{w}_t^n}{\sum_n w_t^n}. \tag{3.16}$$

Then a bridging parameter $\alpha$ is introduced to avoid weight collapse. For this the weights are adaptively shrunk towards uniform weights as follows:

$$\boldsymbol{w}_t^n = \alpha_t \bar{\boldsymbol{w}}_t^n + (1-\alpha_t) N_m^{-1}. \tag{3.17}$$

We obtain then a transition between the EnKF and the particle filter by varying both $\alpha_t$ and $h$. For $\alpha_t = 0$ and $h = 1$, we obtain the uniform weights of the EnKF, while for $\alpha_t = 1$ and $h = 0$ we obtain the particle filter weights (Stordal et al. 2011; van Leeuwen et al. 2019).The optimal value of $\alpha_t$ is estimated as follows:

$$\alpha_t = \frac{N_{eff}}{N_m} = \frac{1}{N_m \sum_{n=1}^{N_m} \left(\bar{w}_t^n\right)^2}. \tag{3.18}$$

Finally, this update formulation reduces the risk of ensemble degeneracy, but cannot entirely avoid it. Therefore, resampling is necessary to avoid ensemble degeneracy. In this study, we use the same resampling technique used in (Stordal et al. 2011) which, for an adequate sample size $N_{eff}$ smaller than 80% of the ensemble size $N_m$, uses a multinomial distribution that randomly samples using the weights as the category probabilities.

**Particle Flow Filter**

The particle flow filter is an ensemble data assimilation method that iteratively transforms equally weighted samples from a prior distribution to the posterior distribution (Hu and van Leeuwen 2021). The transformation follows the solution of a differential equation of the type:

$$\frac{d}{ds}\mathbf{z}_s = \mathbf{f}_s(\mathbf{z}_s), \tag{3.19}$$

in an artificial pseudo time $s \in [0, \infty]$. The particle flow $\mathbf{f_s}$ maps the state $\mathbf{z_s}$ , which follows an intermediate pdf $q_s$,from pseudo time $s$ to $s+1$ having:

$$\begin{aligned} q_0(\mathbf{z}) &= p(\mathbf{z}), \\ q_\infty(\mathbf{z}) &= p(\mathbf{z}|\mathbf{d}), \end{aligned} \tag{3.20}$$

or, what is the same, that at $s = 0$ the pdf $q_s$ is the prior pdf $q_0$ and at the limit of $s = \infty$ the targeted pdf $q_\infty$ will correspond to the posterior pdf (Hu and van Leeuwen 2021).

The particle flow $\mathbf{f_s}$ can be found via the likelihood-factorization formalism or the minimization of a given distance between the intermediate pdf $q_s$ and the targeted pdf

$q_\infty$ (Evensen et al. 2022). In this paper, we follow the distance minimization approach of Hu and van Leeuwen (2021), where the Kullback-Leibler (KL) divergence is used as the measure of distance. The KL divergence between the PDFs above can be defined as,

$$KL\left(q_s\right) = \int q_s\left(\mathbf{z}\right) log\left(\frac{q_s(\mathbf{z})}{q_\infty(\mathbf{z})}\right) d\mathbf{z}. \tag{3.21}$$

Furthermore, we assume that $\mathbf{f_s}$ is in a reproducing kernel Hilbert space (RKHS) with a kernel $\mathbf{K}$, that maps from $\mathbb{R}^{n_z} \times \mathbb{R}^{n_z} \rightarrow \boldsymbol{M}_{n_z \times n_z}(\mathbb{R})$, where $n_z$ is the number of states $\boldsymbol{M}_{n_z \times n_z}(\mathbb{R})$ is a matrix of size $n_z - by - n_z$. Then $\mathbf{f_s}$ is defined such that the KL divergence continuously decreases concerning the pseudo time $s$ as

$$\mathbf{f_s} = \mathbf{C_{zz}} \int q_s\left(\mathbf{z}\right) \{\mathbf{K}\left(\mathbf{z}, \cdot\right) \nabla_z log\left(p\left(\mathbf{z}|\mathbf{d}\right)\right) + \nabla_z \cdot \mathbf{K}\left(\mathbf{z}, \cdot\right)\}. \tag{3.22}$$

Instead of the covariance matrix $\mathbf{C_{zz}}$ one can use any other positive-definite matrix whose dimensions ensure that the dimensions of $\mathbf{f_s}$ agree with those of the state vector. Besides, we directly apply the exact localization and inflation factors used for the EnKF to the prior covariance matrix $\mathbf{C_{zz}}$.

We then proceed to use a particle representation for the intermediate pdf $q_s$ as in:

$$q_s\left(\mathbf{z}\right) = \frac{1}{N_m} \sum_{n=1}^{N_m} \delta\left(\mathbf{z} - \mathbf{z_s}\right), \tag{3.23}$$

and replace it in Eq. 3.22 to find that:

$$\mathbf{f_s}\left(\mathbf{z}\right) = \frac{1}{N_m} \mathbf{C_{zz}} \sum_{n=1}^{N_m} \{\mathbf{K}\left(\mathbf{z_s^n}, \mathbf{z}\right) \nabla_{z_s^n} log\left(p\left(\mathbf{z_s^n}|\mathbf{d}\right)\right) + \nabla_{z_s^n} \cdot \mathbf{K}\left(\mathbf{z_s^n}, \mathbf{z}\right)\}, \tag{3.24}$$

where there are two terms on the right-hand side of the equation that follow a distribution characterized by the Fokker-Plank equation (Evensen et al. 2022), the first term, the product of the kernel with the gradient of the logarithm of the posterior pdf, is referred as the *attracting term*. It pulls the ensemble members towards the observation, driving the pdf towards regions of higher probability. The second term, the divergence of the kernel, is the *repelling term*, dispersing the ensemble members and preventing them from collapsing to a single point. Finally, we discretize the Eq. 3.19 in pseudo time and use the definition of the flow from Eq. 3.22 to estimate the evolution of the state vector from pseudo time $s$ to $s + \Delta s$:

$$\mathbf{z}_{s+\Delta s} = \mathbf{z_s} + \frac{\Delta s}{N_m} \mathbf{C_{zz}} \sum_{n=1}^{N_m} \{\mathbf{K}\left(\mathbf{z_s^n}, \mathbf{z_s}\right) \nabla_{z_s^n} log\left(p\left(\mathbf{z_s^n}|\mathbf{d}\right)\right) + \nabla_{z_s^n} \cdot \mathbf{K}\left(\mathbf{z_s^n}, \mathbf{z_s}\right)\}. \tag{3.25}$$

The kernel $\mathbf{K}\left(z_s^n, z\right)$ measures how each of the ensemble members contributes to the local particle flow of the state $\mathbf{z}$ that is being evaluated, and in the case of having an ensemble of infinite size the solution of the PFF is independent of the choice of the kernel (Lu et al. 2019). In the case of finite ensemble size, the distribution is not unique. However, the statistics of the ensembles, which are products of PFFs with different kernels, should

represent the pdf as accurately as possible (Hu and van Leeuwen 2021). This study uses the matrix-valued Gaussian kernel as in Hu and van Leeuwen (2021).

In the attracting term, by using Bayes theorem(Eq. 3.3), we can rewrite the logarithm of the posterior pdf as:

$$\nabla_{\mathbf{z}} log\ p(\mathbf{z}|\mathbf{d}) = \nabla_{\mathbf{z}} log\ p(\mathbf{z}) + \nabla_{\mathbf{z}} log\ p(\mathbf{d}|\mathbf{z}). \tag{3.26}$$

For a Gaussian observational error with error covariance matrix $\mathbf{C_{dd}}$ the gradient of the logarithm of the likelihood is:

$$\nabla_{\mathbf{z}}\ log\ p(\mathbf{d}|\mathbf{z}) = \mathbf{H}^T {\mathbf{C_{dd}}}^{-1} (\mathbf{d} - \mathbf{Hz}), \tag{3.27}$$

Moreover, when assuming the prior also follows a Gaussian distribution with $\mathbf{z_b}$ and covariance matrix $\mathbf{C^f_{zz}}$ the gradient of the prior is:

$$\nabla_{\mathbf{z}} log\ p(\mathbf{z}) = {\mathbf{C}^f_{zz}}^{-1} (\mathbf{z} - \mathbf{z_b}). \tag{3.28}$$

## 3.4. INFORMATION CONTENT AND UNDERDISPERSION

We evaluated performance through metrics like DFS and RMSE, focusing on underdispersion (Hotta and Ota 2021; Rodgers 2000).

In general, we use the degree of freedom of signal (DFS), the root-mean-square error(RMSE), the standard deviation (STD), and rank histograms to evaluate the performance of the data assimilation.

The definition of the RMSE of the state variable $z_i$, where $i = 1, ..., n_z$ and $n_z$ is the number of states, used in this thesis is:

$$RMSE\ z_i(t) = \sqrt{\frac{1}{N_c} \sum_{j=1}^{N_c} \left(z_{i,j}(t) - \bar{z}_{i,j}(t)\right)^2}, \tag{3.29}$$

where $N_c$ is the number of cells.

We have a particular interest in evaluating whether the filters suffer from underdispersion using the rank histograms and if the DFS, which measures the influence of how much information has been retrieved from the observations during the analysis step (Hotta and Ota 2021; Rodgers 2000), is high enough for the chosen influence factor. For the case of a linear Gaussian data assimilation scheme, as the EnKF, the DFS is defined as:

$$DFS = tr\,(\boldsymbol{S}) = tr\,(\boldsymbol{HK}) \tag{3.30}$$

From this definition, we can also define the *global average influence (GAI)* as the ratio between the DFS and the number of observations as:

$$I_{GS} = \frac{tr\,(\boldsymbol{HK})}{N_d}, \tag{3.31}$$

where $N_d$ is the total amount of assimilated data and offers a more fair comparison between scenarios with different numbers of observations available.

## 3.5. COMPARISON OF METHODS WITH LORENZ 96 MODEL

*Twin experiments* are evaluation techniques in data assimilation research used for understanding aspects like model sensitivity, error dynamics, and developing and benchmarking new methods in a controlled environment. They typically involve creating a model's *truth* run to generate synthetic observations. These observations are then used in the same model but with different starting conditions to test data assimilation methods. This approach allows for including model errors, which makes the testing more realistic. Perfect model experiments are a type of twin experiment where the model is assumed to be perfect, with no errors. This focuses the testing solely on the data assimilation techniques. The *Lorenz '96* model is commonly used in both experiments. It is simple, efficient, and good at showing chaotic and nonlinear behaviors seen in the atmosphere. This makes it a helpful tool for testing and comparing different data assimilation methods in atmospheric science.



### 3.5.1. LORENZ 96 MODEL

The Lorenz 96 model is a simplified yet effective representation of the chaotic behavior of atmospheric dynamics (Lorenz and Emanuel 1998) commonly used as a benchmark in testing data assimilation techniques. The equation that models Lorenz 96 is:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-1})\,x_{i-1} - x_i + F \tag{3.32}$$

with boundary conditions $x_{-1} = x_{N_c-1}$, $x_0 = x_{N_c}$, $x_{N_c+1} = x_1$ and constraint $N_c \geq 4$. Here, $x_i$ represents the state, for instance, temperature, at a sector along a latitude circle divided into $N_c$ equal sectors (Kekem 2018). The equation features advection, damping, and forcing effects. The system exhibits coherent structures and even chaotic behaviors based on parameters $F$ and $N_c$.

### 3.5.2. PERFECT MODEL EXPERIMENTS

We perform perfect model experiments for the Lorenz 96 models for different observation densities. Specifically we use Lorenz 96 models of 20 cells and used $F = 1.2$ for the periodic case and $F = 8.0$ for the chaotic case. The observation densities used were 25% (5 cells), 50% (10 cells), and 100% (20 cells). The ensemble size of 100 members which are initialized after sampling a long run of the periodic and chaotic truth. Then synthetic observations are extracted using a normal distribution of mean 0 and standard deviation 0.1. The observations are sequentially assimilated using then EnKF, AGMF and the PFF. We use a localization scheme of the covariance matrices as shown in Appendix A.

### 3.5.3. RESULTS

#### STATE ESTIMATION RESULTS

Fig. 3.1 depicts the RMSE outcomes for the variable $x_i$ over time. Notably, RMSE values for the periodic Lorenz 96 model are consistently lower than for its chaotic counterpart. In the chaotic case, all three filters yield RMSE values below the observation error for the A.1. case, which involves 100% cell coverage. In comparison, only the PFF achieves values below the observation error with 50% coverage, maintaining lower RMSE even at

25% coverage. In contrast, the EnKF shows the highest RMSE at 50% and 25% coverages, while the AGMF also has higher RMSE values than the RMSE but with less oscillation when compared to the EnKF. These findings align with the observations by Hu and van Leeuwen (2021) on the 40 variable-Lorenz 96 model, which noticed similar contrasts between the LETKF and PFF. The increased non-linearity in variable relationships during chaotic periods underscores the necessity for a non-Gaussian filter, especially when full variable coverage is unavailable.



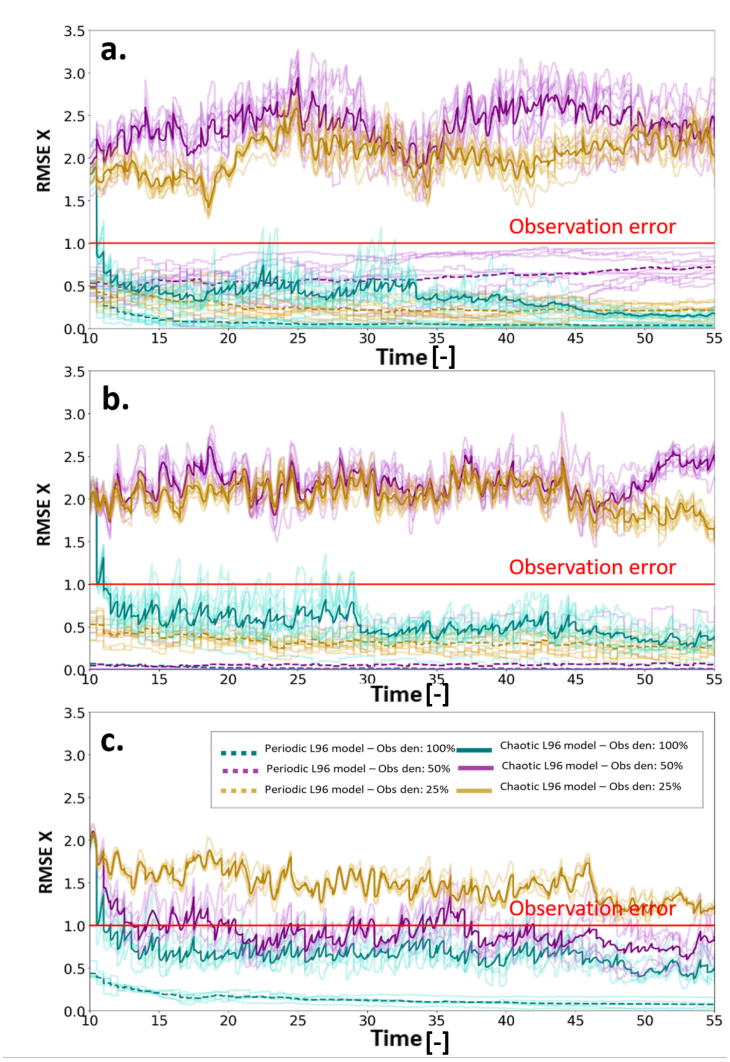


Figure 3.1: Comparison of RMSE for different observation densities for the Lorenz 96 between EnKF(a), AGMF (b), PFF (c).

### PARAMETER BIAS AND MODEL ERROR RESULTS

Fig. 3.2 shows the time series estimates of an EnKF when including a model error term for the Lorenz 96 model. This tests uses ensemble members where the parameter correspond to periodic solutions but still we want to estimate chaotic solutions. The data assimilation starts after time step 10. It is possible to see that after the assimilation of the first observation the estimates of the EnKF enter in sync with the truth despite the parameter bias. The effect of the introduction of the model error acts in this case as a type of parameter estimation due to the simplicity of the Lorenz 96 model.

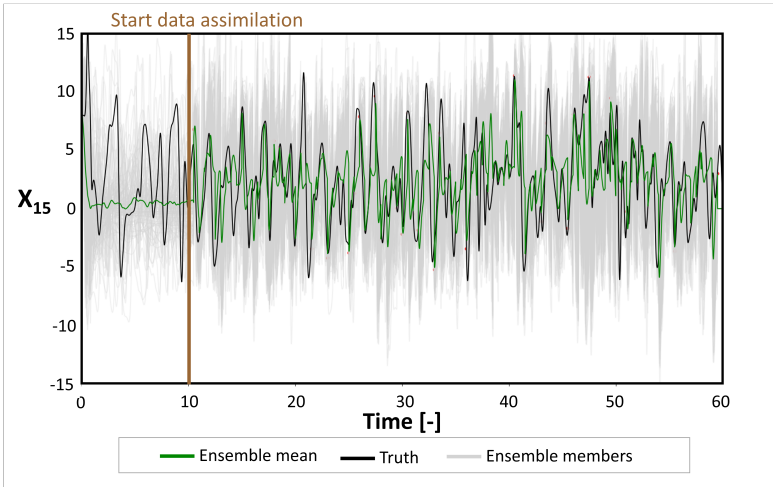


Figure 3.2: Estimates of a cell of the Lorenz 96 model when including the model error in the estimation of a chaotic model in a synthetic experiment.

Fig. 3.2 shows the time series estimates of the model error estimated.

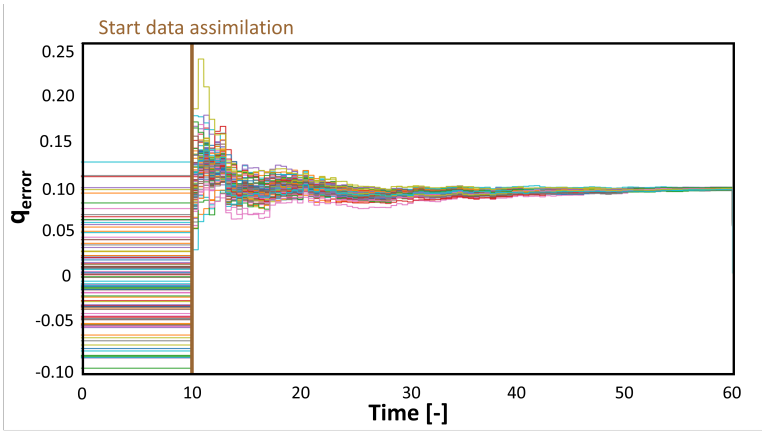


Figure 3.3: Estimates of the model error in the synthetic experiment with a chaotic model.

### PARTICLE FLOW FILTER UPDATE

As shown in Fig. 3.1, the PFF is the method with the lowest RMSE for the chaotic models, especially for the lowest observation densities. Fig. 3.4 and 3.5 show visually the non-linear update of the Particle Flow Filter an the pseudoflow that transforms iteratively the prior distribution to the posterior distribution. As is visible in the figures, the updates become more non-linear in the chaotic case, which the PFF can capture to get better estimates.



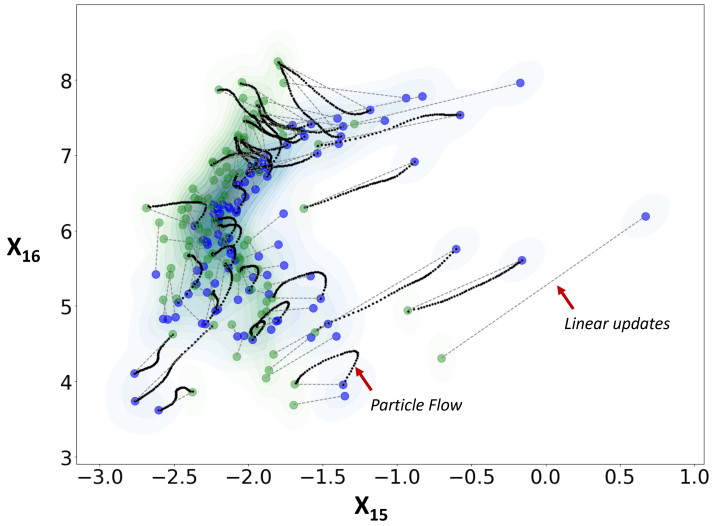


Figure 3.4: Crossplot between observed and unobserved variables. The blue dots represent the prior distribution; the green dots represent the posterior distribution of ensembles. The dashed lines represent the linear update between prior and posterior, while the black dots represent the non-linear update of the Particle Flow Filter for a single analysis step.

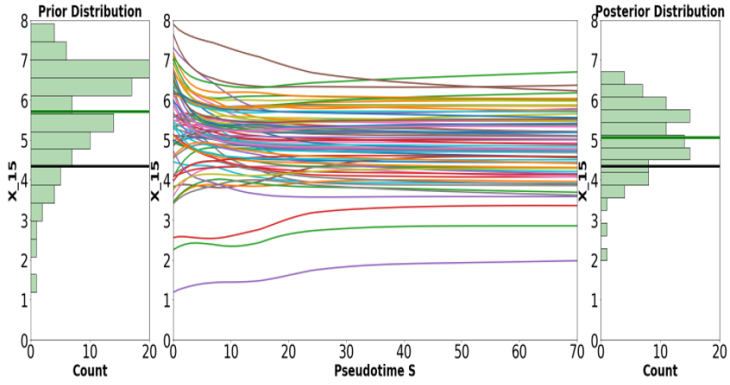


Figure 3.5: Visualization of the pseudoflow of the particle flow filter for the state $x_{15}$.

### 3.5.4. DISCUSSION

In this chapter, we introduced the ensemble data assimilation methods to be used in Chapters 4 and 5. We further used a Lorenz 96, non-linear model widely used for testing data assimilating experiments to highlight the differences between methods. We used perfect model experiments to see the differences in the errors of the estimates of the EnKF, the AGMF, and the PFF for the Lorenz model under periodic and chaotic conditions. We see that the EnKF with Gaussian prior assumptions is sufficient to capture the behavior of the systems for the periodic conditions well but that it struggles to provide good estimates for the chaotic case, especially when we have fewer observations of the system. In Chapter 5, we extend these results to evaluate if this trend also applies to systems governed with rate-and-state friction as in the 1-D Burridge Knopoff model shown in Chapter 2. As we see in some model error results, periodic solutions can estimate a chaotic system when the model error is considered in estimating the data assimilation. This is possible on the Lorenz 96 model due to its simplicity and because the model error corrects the biased forcing term. As discussed in Chapter 2, this same result is not guaranteed for the Burridge-Knopoff model, given the multiple frictional parameters that are involved in the slip behavior of the system and the sensitivity they impose on the stability of the simulations.



#### DATA AVAILABILITY

The forward model for the Lorenz 96 model is accessible via repository https://github.com/hdiabmontero/non_gaussian_data_assim/. The data produced and analyzed in this chapter is available via 4TU.ResearchData (http://doi.org/10.4121/f0f075f2-f45c-4f8c-9d1d-bde03baeae33).

# 4

# ESTIMATING THE OCCURRENCE OF SLOW SLIP EVENTS AND EARTHQUAKES WITH AN ENKF

***Summary:*** *In this chapter, we employ an Ensemble Kalman Filter (EnKF) to estimate shear stresses, slip rates, and the state $\theta$ in a 1D model with a surrounding elastic medium and a fault governed by rate-and-state friction. By assimilating noisy shear-stress and velocity data near the fault through perfect model experiments, we improve estimates in segments undergoing slow-slip events and aid slip-rate estimation for earthquakes. Shear stress observations at the medium are crucial for determining shear stress, slip rates, and the state $\theta$ at the fault. In contrast, velocity data from the medium enhances slip-rate estimates at the fault.*

***Resumen:*** *En este capítulo, utilizamos el Filtro de Kalman de Ensamble (EnKF) para estimar los esfuerzos cortantes, las tasas de deslizamiento y el estado $\theta$ en un modelo unidimensional (1D) que consta de un medio elástico y una falla, la cual se rige por la ley de fricción rate-and-state. Mediante la asimilación de datos ruidosos de esfuerzo cortante y velocidad cercanos a la falla en experimentos sintéticos, logramos mejorar las estimaciones en segmentos que experimentan eventos de deslizamiento lento, y contribuimos a la estimación de la tasa de deslizamiento en terremotos.*

This chapter has been published as Hamed Ali Diab-Montero, Meng Li, Ylona van Dinther, Femke C Vossepoel, Estimating the occurrence of slow slip events and earthquakes with an ensemble Kalman filter, Geophysical Journal International, Volume 234, Issue 3, September 2023, Pages 1701–1721, https://doi.org/10.1093/gji/ggad154 .
Minor changes have been applied to make the text and figures consistent with the thesis.

## 4.1. INTRODUCTION

Efforts to forecast earthquakes and slow slip events are significantly constrained by limited knowledge about the current stress state of faults. Ensemble data assimilation methods provide a promising approach to overcome this limitation by effectively combining physics-based models with observational data, accounting for their uncertainties. This chapter focuses on applying these methods to estimate the stresses and velocities of slow slip events and earthquakes with fast slip rates. This latter application represents a novel aspect of our research, distinguishing it from much of the existing literature in this field.

In Section 1.3, we briefly mentioned various studies that have employed data assimilation in seismic research. These include advancements in tsunami early warning systems and efforts in estimating frictional parameters of megathrust faults, exemplified by researchers such as Maeda et al. (2015), Wang et al. (2018), and Oba et al. (2020). Additionally, studies like those of van Dinther et al. (2019) and Hirahara and Nishikiori (2019) have explored the Ensemble Kalman Filter in modeling subduction zone dynamics, demonstrating the potential of minimal observations to enhance fault slip estimations. In this chapter, we expand upon the foundational insights from these studies, mainly focusing on applying ensemble-based data assimilation methods for earthquakes characterized by high-velocity slip events.

In their work van Dinther et al. (2019) conducted a perfect model experiment in a simulated subduction zone, using synthetic data on velocities and stresses near the surface. This data is processed with an Ensemble Kalman Filter(EnKF) to update velocity and stress states across multiple ensemble members governed by a seismic cycle model. The visco-elasto-plastic forward model uses slip-rate-dependent friction to simulate earthquakes with slow slip rates. Drawing from this work, we extend to the use of numerical models based on a rate-and-state friction formulation and adaptive time stepping. We aim to rigorously evaluate the Ensemble Kalman Filter's effectiveness in forecasting slow slip events and, more critically, earthquakes with fast slip rates.

## 4.2. METHODOLOGY

We use a numerical model setup inspired by a large-scale biaxial friction apparatus in a direct-shear configuration (Fig. 4.1). We discretize a 1-D line across the block boarded by a single, 0-D fault point, as described in Section 2.3.

We assimilate synthetic observations using an EnKF to evaluate how our fault estimates compare to the known truth. We assume our model is perfect, i.e., that parameters shown in Table 5.1 are correct. The synthetic observations are samples of the simulated shear stress and the velocity as if measured at a single location at a short distance from the fault, with noise added to represent measurement error. The selection of observation error amplitudes was based on widely accepted instrument sensitivity values (van Dinther et al. 2019) and the maximum observed variations in stress and velocity during the coseismic phase. For slow slip events, observation error values were set at 0.25 MPa for shear stress measurements and 0.25 for the logarithm of velocity observations. For earthquakes, the observation errors were assigned 0.75 MPa for shear stress measurements and 0.75 for the logarithm of velocity observations.

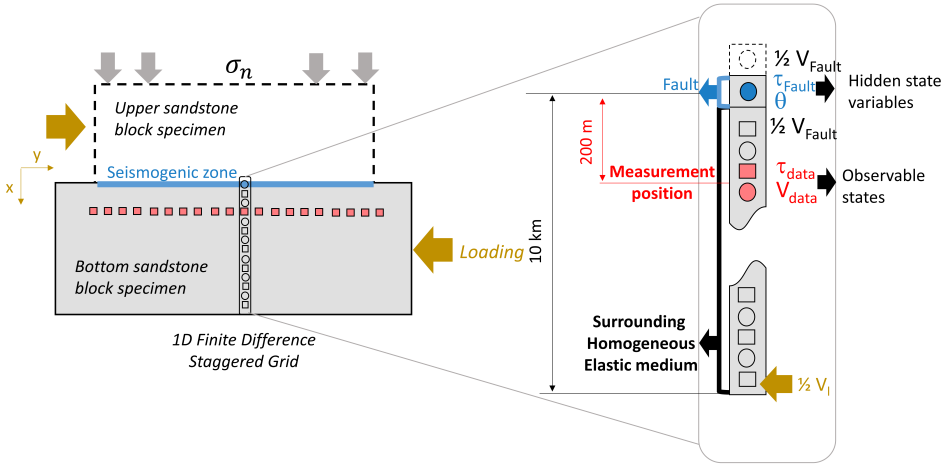


Figure 4.1: Schematic representation of the 1D model used to represent a horizontal straight fault. The shear stress ($\tau$) is estimated in the nodes represented by circles and the velocity($v_y$) in the ones represented by squares. The fault (blue) corresponds to the uppermost node of the grid and the location where the EnKF estimates shear stress, slip rate and state $\theta$. Shear-stress and velocity observations (red) at 200 m from the fault in the surrounding homogeneous elastic medium are assimilated into a 10-km model. The spacing between nodes is 20 m.

### 4.2.1. PERFECT MODEL EXPERIMENTS

In the data assimilation experiments, we use an ensemble of 50 members. Each member is initialized with a different value of initial shear stress at the fault node following a Gaussian distribution with a standard deviation of 2.5 MPa. The synthetic true value of initial shear stress is set to 20 MPa, and the mean of the ensemble has a bias of 3 MPa concerning that true value (Fig. 4.2). The total simulation time is 1500 years, from which we sample observations at 2.5-year intervals for slow slip events and at 5-year intervals for earthquakes. The workflow is illustrated in Fig. 4.3. To compare the results with and without data assimilation, we run the 50 members up to 200 years where the first observations are available. Data is assimilated then and subsequently whenever sampled synthetic observations are available.

We follow the workflow presented in Fig. 4.3 for the perfect model experiments. In our case, the state vector $\mathbf{z}_n$ includes the shear stress of the fault, the shear stress in the medium, the slip rate, the velocity in the medium, and the state $\theta$:

$$\mathbf{z}_n^T = \left(\tau_{\text{Fault}}, \tau_{xy}^T, \ln(\text{V}), \ln(\mathbf{v}_\text{y})^T, \ln(\theta)\right)_n, \qquad 1 \le n \le N, \tag{4.1}$$

After the analysis step the updated shear stress and state $\theta$ are used to calculate the slip rate $V$. This is done by using an implicit solver that finds the corresponding slip rate that satisfies Eq. 2.6. The velocity values in the medium $v_y$ are calculated for the next step solving the system of Eqs. 2.8-2.19, where the estimated posterior values become part of the velocity history of the model.

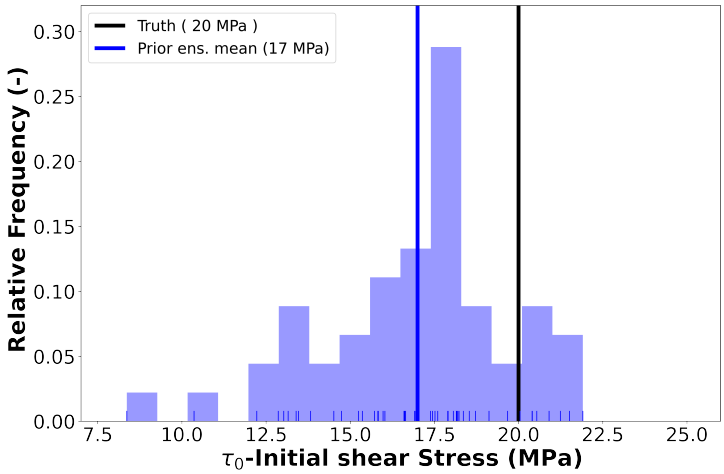


Figure 4.2: Histogram illustrating the initial distribution of shear stress at the fault used to initialize the ensemble. We define a synthetic truth with an initial shear stress at the fault of 20 MPa. The vertical blue line is the mean of the Gaussian distribution, the black vertical line is the truth, and the marks in the horizontal axis represent the 50 different initial stress values used for the ensemble.

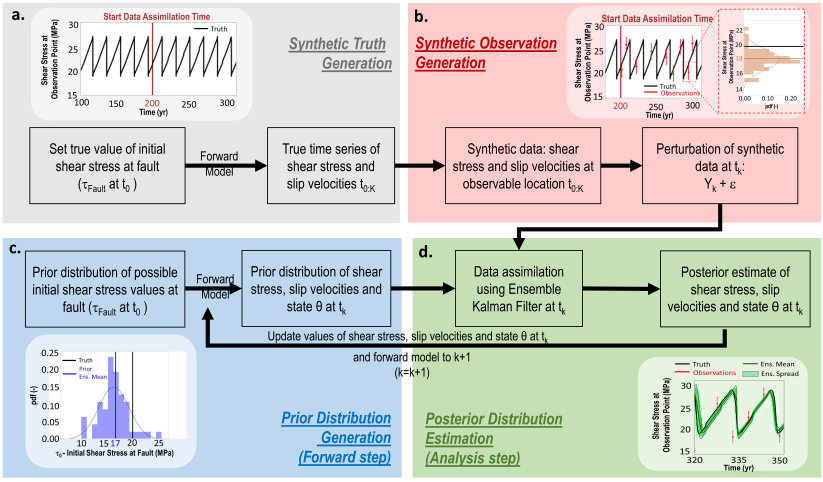


Figure 4.3: Schematic representation of a perfect model experiment and sequential data assimilation. The flowchart describes four steps of the synthetic experiment. (a) First, in gray the steps for generating the synthetic truth to be estimated. (b) Second, in red the creation of the synthetic observations to be assimilated. (c) Third, in blue the creation of the ensemble of realisations that form the prior distribution. (d) Fourth, in green the data assimilation step where a data-assimilation method (in this case the EnKF) is used to estimate a posterior distribution of the state using the synthetic observations and the prior ensemble distribution. The last two steps are done sequentially until the last time step K where observations are available. In our case, we assimilate shear stress and slip velocity observations measured in the medium.

Table 4.1: Material and rate-and-state friction parameters for the 1D model setup for both slow slip events and earthquakes

| Parameter | Symbol | Slow Slip Events | Earthquake Events |
|---|---|---|---|
| Shear modulus | G | 32 GPa | 32 GPa |
| Density | $\rho$ | 2670 kg/m$^3$ | 2670 kg/m$^3$ |
| Initial mean stress | $\sigma_n$ | 40 MPa | 40 MPa |
| Static friction coefficient | $\mu_0$ | 0.6 | 0.6 |
| Reference slip rate | $V_0$ | 1 $\mu m/s$ | 1 $\mu m/s$ |
| Characteristic slip distance | L | 0.24 m | 0.18 m |
| RSF Direct Effect | a | 0.0060 | 0.0060 |
| RSF Evolution Effect | b | 0.0158 | 0.0160 |
| Depth model | H | 10 km | 10 km |
| Loading slip rate | $V_l$ | 10 nm/s | 10 nm/s |
| Grid spacing | $\Delta x$ | 20 m | 20 m |



## 4.3. RESULTS AND ANALYSIS

The results of the EnKF estimation are analyzed at the observation location at 200 m from the fault to analyze how observations affect the assimilation and at the unknown target location, the fault.

### 4.3.1. STATE ESTIMATION IN THE HOMOGENEOUS ELASTIC MEDIUM

The time evolution of the shear stress with respect to the grid $\tau_{xy}$ at the observation location shows that the estimates of the EnKF (represented in green) for both the slow slip events and the earthquakes are in sync with the truth (Figs. 4.4a and 4.4b, resp.). The EnKF analysis resolves both the interseismic and coseismic phases well despite the large errors in the observations. For example, we observe large errors around 834 years for the slow slip events and around 980 years for the earthquakes. The EnKF is especially effective for the slow slip events (Fig. 4.4a), where the ensemble mean captures the coseismic phase reasonably well. The shear-stress estimates for the earthquakes during the coseismic phase are comparatively less precise (Fig. 4.4b). The very large and fast stress drop during earthquakes is inherently difficult for any probabilistic or averaging method to reconstruct precisely. Nonetheless, the timing of earthquakes is generally anticipated by a drop in the mean shear stress about two to three years prior to the earthquake occurrence and an increase in spread of the ensemble with an approximate standard deviation of 2 MPa.

As depicted in Figs. 4.4c and 4.4d, the uncertainty of the time evolution of the velocity at the observation location differs considerably between slow slip events and earthquakes. For both the slow slip and the earthquake events, the analysis of the EnKF captures the true velocity in the interseismic phase very well. In the coseismic phase, it is more challenging to provide an accurate estimate due to fast changes in velocity. For the slow slip events (Fig. 4.4c), the EnKF captures the evolution of the velocity very well, and, remarkably, it even accurately captures the peak velocities in terms of magnitude and timing. However, for the earthquakes, the magnitude of the estimated peak velocity still has a large uncertainty (Fig. 4.4d). This results from averaging over 50 ensemble members, each showing only 7-9 orders of magnitude higher velocities for seconds. However, the ensemble mean correctly traces the increases in velocity up to the loading rate. This

is important to realize for estimating the timing of nucleation since after the increase in slip rates is identified, its maximum is capped by the activation of inertia and thus is already known to be in the order of meters per second.

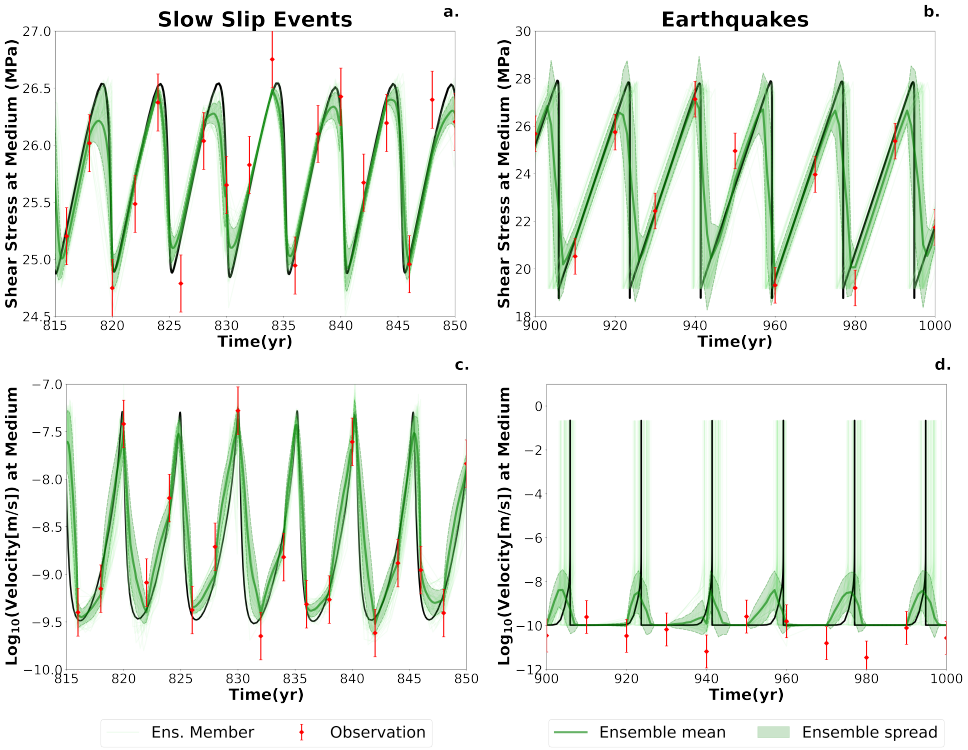


Figure 4.4: Estimated evolution of (a,b) shear stress and (c,d) slip rate at the observation location for (a,c) slow slip events and (b,d) earthquakes. The solid black line represents the true evolution; the red markers are the observations with error bars indicating the standard deviation of the observational error; the green solid line is the ensemble mean, and the light green lines represent the ensemble members. The ensemble spread is shown as the light green hatched area, which is one standard deviation below and above the ensemble mean for normal distribution and corresponds to the ensemble members between the percentiles P16 and P84 of the ensemble distribution.

### 4.3.2. STATE ESTIMATION AT THE FAULT

The EnKF estimates the shear stress and slip velocities relatively accurately at the locations where we have observations (Fig. 4.4). Nevertheless, good estimates are not guaranteed for locations where measuring devices do not have easy access, such as the fault (Fig. 4.5). This is especially difficult if the state variables experience rapid changes. Therefore, we evaluate the performance of the EnKF in the same experiments from Fig. 4.4 but at the location where the estimation is most challenging, i.e., at the fault. At the same time, this is the location where correct estimates are critical and most needed, as movements in the medium respond to displacements at the fault, and the state variables are unobserved there.

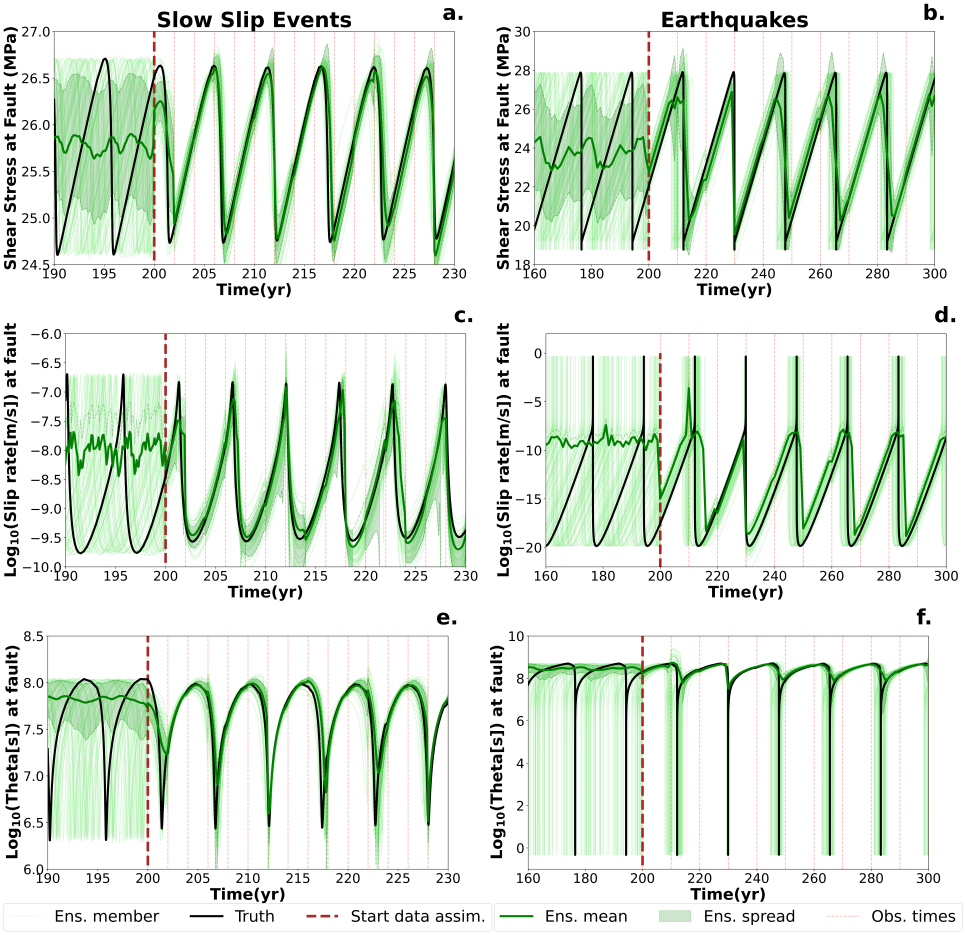


Figure 4.5: Estimated evolution of (a,b) shear stress, (c,d) slip rate, and (e,f) state $\theta$ at the fault for (a,c,e) slow slip events and (b,d,f) earthquakes. The brown line is the start of the data assimilation. The black solid line is the true evolution of the variables. The red dashed lines indicate the observation and assimilation times. The coloring of the lines is as in Fig. 4.4.

Prior to data assimilation, ensemble members are not aligned, so their mean can not separate between interseismic and coseismic phases and almost becomes a constant value (Fig. 4.5). After 200 years, there is a significant decrease in the spread of the ensemble for all hidden states due to the assimilation of the first observations. As shear stress is increased in many ensemble members, the occurrence of the first slow slip event is synchronized even after the assimilation of a single observation (Fig. 4.5a). The first earthquake is preceded by the assimilation of a second observation, which introduces uncertainties in the shear stress estimates but does allow the ensemble to accurately predict the timing of the stress drop (Fig. 4.5b). Assimilation of a rapid slip rate is also apparent from the slip rate jump at t=210 yr (Fig. 4.5d), which is immediately corrected by the forward model using shear stress to estimate the next slip rate. Assimilation dur-

ing the next slip sequences shows that the estimates of the EnKF are very well synchronized with the truth in our case of state estimation with known and constant parameters. Generally, for a large part of the interseismic period, slip rates and states from all ensemble members show that a slip event is not expected (Fig. 4.5c-f). Interestingly, all three hidden states are tracked with extreme accuracy for some sequences, e.g., around the earthquake at about 230 years (Fig. 4.5b,d,f).

The peak shear stress has a larger spread in the posterior of the earthquake model when compared to the slow slip events when approaching the coseismic phase (Figs. 4.5a and 4.5b). The ensemble spread around the coseismic phase of the earthquakes has a large spread around the peak slip rate compared to the slow slip events (Figs. 4.5c and 4.5d). The ensemble mean does not reach the peak slip rates or the peak shear stresses due to the spread of the ensemble in the interseismic and coseismic phases, which is related to the short duration of the coseismic phase. However, individual ensemble members give a good indication of peak slip rates and shear stresses. The ensemble provides good state $\theta$ estimates, whose timing is synchronized with the truth after the data assimilation starts for both fault-slip events (Figs. 4.5e and 4.5f). The ensemble prior PDF of the three hidden states shows a change in spread around 2 to 3 years before and after the coseismic phase. This distinct increase in the spread of the ensemble prior to the event may be an indication of an upcoming earthquake.



### 4.3.3. NON-GAUSSIANITY

The EnKF assumes that the prior distributions of the state vector's components and observation errors are Gaussian distributed. Additionally, the EnKF produces a low-rank representation of the prior covariance matrices using a finite ensemble of members, making the EnKF very sensitive to outliers. This is particularly true for cases where the ensemble size is small.

In this section, we evaluate how well preserved the assumption of Gaussian distributions is in the different phases of the cycle of slow slip events and earthquakes (Figs. 4.6 and 4.7). For both types of events, we provide the histograms of the ensembles during the interseismic phase (Figs. 4.6b and 4.7b) and shortly before the start of the coseismic phase (Figs. 4.6c and 4.7c). The ensemble distributions for the slow slip events, are very close to a Gaussian. The preservation of the Gaussian assumption allows the EnKF to have estimations as shown in Fig. 4.6a. The range of the differences between the ensemble mean, and the truth corresponds to the expected range of uncertainty. The maximum errors are around 0.4-0.8 MPa, which are between 25% and 50% of the average stress drop of the slow slip events. Regarding the uncertainties on the timing of slip occurrence, it is about one year, which is 20% of the recurrence interval of 5 years of the slow slip events.

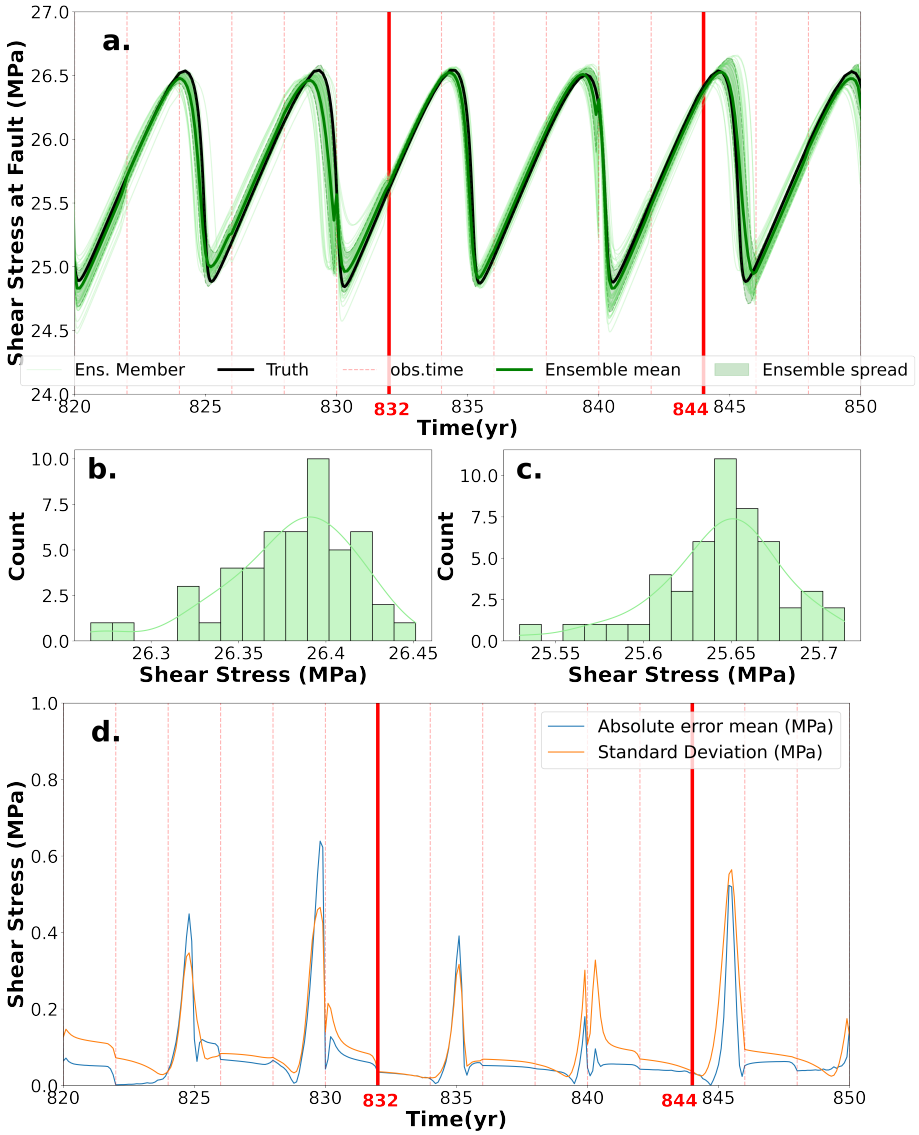


Figure 4.6: (a) Shear stress estimation of slow slip events with prior distributions during the (b) interseismic and (c) just prior to the coseismic phase. (d) Evolution of shear stress errors with respect to the ensemble spread. The vertical red solid lines indicate the time of the ensemble distributions shown in b and c. The red dashed lines indicate the observation and assimilation times The blue line represents the absolute value of the difference between the truth and the ensemble mean estimate. The orange line represents the standard deviation of the ensemble distribution.

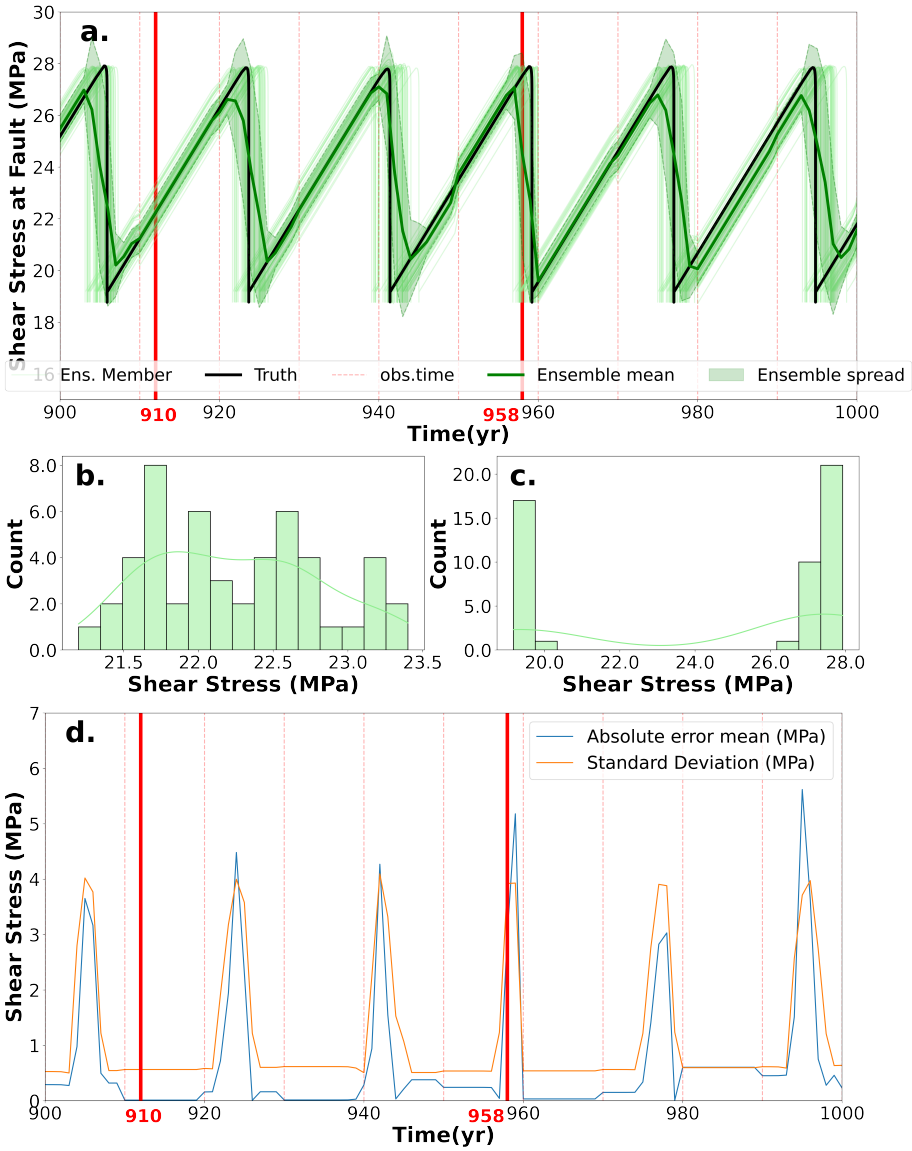


Figure 4.7: As in Fig. 4.6 but for earthquakes instead of slow slip events.

In contrast, for earthquake sequences, the Gaussian distribution in the prior is only well preserved during the interseismic phase (Fig. 4.7a,b). This is because the members that experience the sudden large change of shear stress in the coseismic phase deviate strongly from the ensemble mean (Fig. 4.7c). Some ensemble members are finishing the interseismic phase of the previous cycle and entering the coseismic phase, while others are starting the next cycle. This results in a skewed distribution of the prior shear

stress, translating into a bimodal distribution for the prior (Fig. 4.7c). This effect is also seen in Fig. 4.7d in the evolution of the absolute errors and the standard deviation of the ensemble. Similar to the slow slip events, the filter errors are in the same range of the values of the standard deviation of the ensembles and the range of expected uncertainties. The maximum errors are around 50 % of the stress drop of 10 MPa. The errors in the earthquake occurrence are about five years, which is 25% of the recurrence interval of 20 years of the earthquake sequences. This similarity in performance suggests that the non-Gaussianity of prior estimates may not significantly limit the effectiveness of the EnKF in our case of state estimation with known and constant parameters.

### 4.3.4. SENSITIVITY OF ANALYSIS TO OBSERVATION TYPE

We observe that the EnKF gives reasonable estimates of the shear stress, slip rate, and state $\theta$ at the fault despite the rapid changes in the stresses and velocities in the coseismic phase. In this section, we assess how the estimation of these rapid changes is realized by the EnKF and what the influence of the different types of observations on this is. We do this by analyzing the elements of the Kalman gain matrix at every analysis step of the data assimilation.



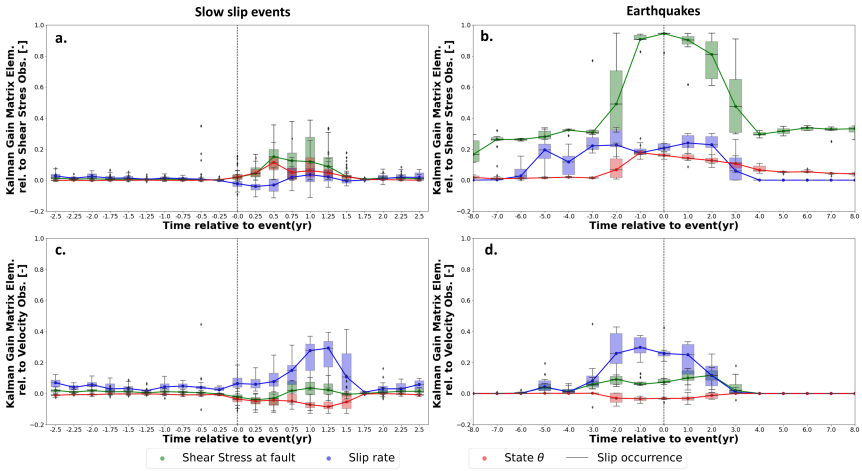


Figure 4.8: Kalman gain matrix elements vs. the relative time to the slip occurrence for slow slip events (a,c) and earthquakes (b,d). The top row (a,b) shows the Kalman gain matrix elements concerning the shear stress observations, and the bottom row (c,d) shows the Kalman gain matrix elements concerning the velocity observations. The green boxplots are the gain values for the shear stress, the blue boxplots for the slip rate, and the red boxplots for the state $\theta$ at the fault. The solid lines connect the median values for all the boxplots for a given variable.

As explained in Sec. 3.2, a high Kalman gain places a higher weight on the observations information when estimating the posterior distribution of the ensemble. Every element of the matrix maps the *innovation*, *i.e.*, the difference at the observation location between the observed values and the estimated values from the forward model to an update of a variable in the state vector. From this, we can assess, for example, the

relative influence of the shear-stress observations at the fault on the estimated slip rate. Fig. 4.8 illustrates the Kalman gain matrix elements from all analysis steps of the slow slip events and the earthquakes that relate the shear-stress observations and the velocity observations to the estimates of shear stress, slip rate, and state $\theta$ at the fault. By comparing the time when the observations are assimilated relative to the slip occurrence, we evaluate how these Kalman gain elements change after every occurrence of a slow slip event or earthquake.

In both types of events, the Kalman-gain values that relate the shear-stress estimates to the shear-stress observations are higher than those that relate to the velocity observations, suggesting that the shear-stress estimates are more strongly influenced by the shear-stress observations than by the velocity observations. This is especially noticeable for earthquakes, where the Kalman gain matrix elements that relate to the shear-stress observations are between 4 and 8 times higher than the elements that relate to the velocity observations. In contrast, the Kalman gain matrix elements of the slip rate estimates related to velocity observations for the slow slip events are 2 to 4 times higher than those related to the shear-stress observations. This suggests that the velocity observations influence the slip-rate estimates the most, and the shear-stress estimates are influenced the most by the shear-stress observations.



We observe a difference in the relative influence of the slip-rate observations on the slip-rate estimates between the slow slip events and the earthquakes. For slow slip events (Fig. 4.8c), the slip-rate estimates are mostly influenced by the velocity observations, but for both observations appear to have equal influence in the period before the slip occurrence. However, after the slip occurrence, the elements that relate the velocity observations to the slip-rate estimates are higher than the ones that relate to the shear-stress observations.

For the state $\theta$, both observations seem to influence the estimates for the slow slip events equally. In contrast, for earthquakes, the Kalman gain elements related to the shear-stress observations are two times higher than those related to the velocity observations.

All Kalman gain-matrix elements for the slow slip events tend to be higher after the slip occurrence. The higher values for earthquakes occur shortly before and after the event. The Kalman gain is a function of the observational error and the prior covariance of the state vector as shown in Eq. 3.13. In our perfect model experiments shown here, we assume that the uncertainties of the synthetic observations are constant in time. Therefore, as indicated by the boxes in Fig. 4.8, the Kalman gain values become a function of the ensemble spread. Since the ensemble spread is high shortly before and after the earthquake, the higher Kalman gain values will effectively result in a stronger influence of the observation information on the analysis during these periods. For the slow slip events, a large ensemble spread occurs mainly after the coseismic phase because of the smoother transition of the phases. We observe higher Kalman gain values after the seismic event. Interestingly, the slip rate (blue line) changes more noticeably than the other variables before the earthquake event as illustrated by Figs. 4.8b and 4.8d. This suggests that the slip rate is more sensitive to data assimilation before the earthquake event than other variables.

### 4.3.5. FORECASTABILITY

We evaluate the forecastability of the EnKF for slow slip events and earthquakes by comparing the time series of the estimated relative frequency of fault-slip occurrence to the synthetic true occurrences (Fig 4.9). Relative frequencies indicate the percentage of ensemble members that estimate an earthquake within five different rolling window lengths (1, 2, 3, 4, and 5 years) and a lag of 1 year. The lag in this scenario means we shift our window's interval each time by one year. We observe that the relative frequencies of occurrence for the rolling windows of 1-year length are typically relatively small, i.e., smaller than 0.2, although, for both fault-slip types, exceptions up to 0.5 exist. However, for longer windows from 2 up to 5 years, it is possible to reach relative frequencies from 0.75 up to 1.0.



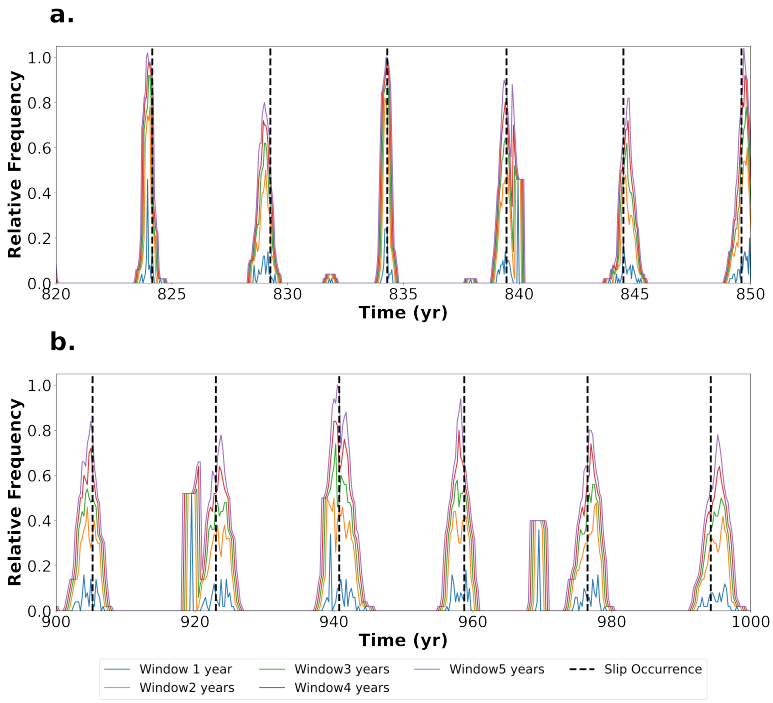


Figure 4.9: Time series of the probability of fault-slip occurrence for (a) slow slip and (b) earthquakes. These estimates are calculated using a rolling time window using different window lengths. Probabilities are given as the relative frequency of the ensemble members that forecast the earthquakes. The light gray line indicates the time of the slip occurrence. The solid blue, orange, green, red, and purple lines are the estimations for a window of 1,2,3,4 and 5 years, respectively.

Interestingly, particularly for rolling windows of 5 years, the peak of relative frequencies in most cases occurs at the same time as the true slip events, with maximum deviations of less than one year. Moreover, the maximum relative frequencies are comparable for both fault-slip types, indicating that the forecastability of earthquakes is comparable to those of slow slip events.

We also evaluate the forecastability of the data-assimilation framework by using a Molchan or error diagram to analyze what is known just prior to the earthquake, instead of forecasting a slip event once it has happened, i.e., hind-casting (Fig 4.10). The diagram considers 72 earthquakes and 259 slow slip events in 1300 years. We assume that an alarm is ringed once a percentage of the ensemble (10%,20%, and 30%) has reached its peak shear stress. We calculate the period between the moment that the alarm rings and the actual slip occurrence and divide this time by the recurrence interval of the event to obtain the alarm duration. We estimate the failure rate as the fraction of events not forecasted from the total number of events that occurred.

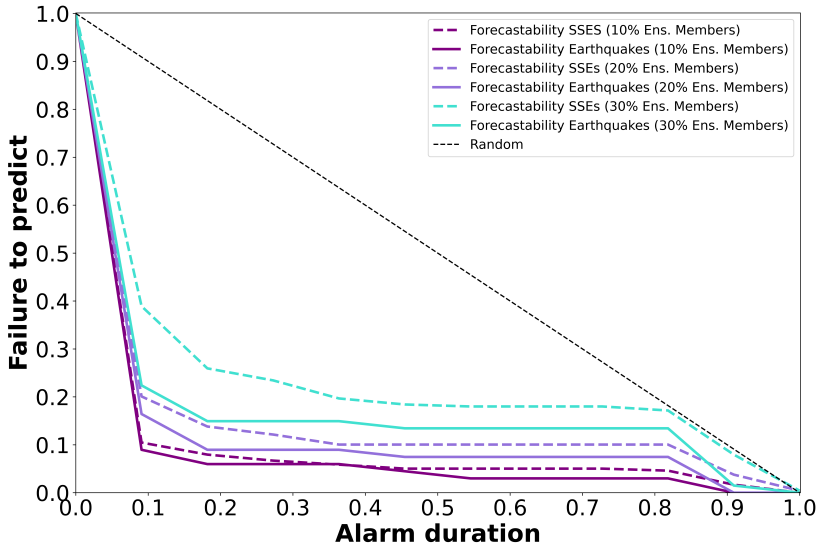


Figure 4.10: Molchan diagram comparing the forecasting potential of the EnKF for slow slip events and earthquakes. The horizontal axis is the alarm duration calculated as a percentage of the recurrence interval of the seismic event. The vertical axis is the failure to predict events and was calculated considering the percentage of slip occurrences, which were correctly forecast during the alarm duration. The dashed curves correspond to slow slip events, and the solid line curves are earthquake estimates.

In the Molchan diagram, a forecast located at (0,0) corresponds to having a zero prediction failure, while ringing the alarm of an upcoming slip event for a very short period. Theoretically, if we ring the alarm all the time we will not miss any slip events (i.e., 1,0); if we never ring the alarm we will miss all events (0,1). An optimal forecast would follow a curve that comes as close to the origin and both axes. The results show that the forecastability of the EnKF for both slow slip events and earthquakes is very similar when ringing an alarm at times when 10% of the ensemble members have reached their peak stress. Almost 90% of the slow slip events and earthquakes are forecast when ringing the alarm just for 10% of the recurrence interval duration. That means the slow slip events were forecast as early as half a year before the slip occurrence and around 2 years before for the earthquakes. We also evaluate how forecastability changes when considering a different proportion of ensemble members (20% and 30%). The EnKF shows lower

earthquake failure rates for the same alarm duration than for slow slip events. This may be connected to the larger ensemble spread experienced when estimating earthquakes, which may lead to earlier alarms compared to the slow slip events that tend to stick closer to the truth diagram.

## 4.4. DISCUSSION

The results show that the EnKF provides good estimates and the lowest deviations from the truth during the interseismic phase of both the slow slip events and the earthquakes. The state estimates are most uncertain during and around the coseismic phase of the earthquakes. This is more pronounced for the slip rates and state variable $\theta$'s estimates, which experience variations of many orders of magnitude between interseismic and coseismic values. This is unsurprising, as the estimate is constructed by averaging 50 ensemble members with varying slip rates and state variables (e.g., Fig. 4.4). This aspect of the data assimilation may differ when using a different assimilation method (see, for example, Chapter 19 in Evensen et al. (2022)).



### 4.4.1. IMPACT OF NON-LINEARITY AND NON-GAUSSIANITY

We imposed a rate-and-state friction formulation (Eq. 2.6) on the fault to generate earthquake sequences. This dynamical friction model is nonlinear because small changes in one variable (e.g., stress) trigger disproportionally large and nonlinear changes in another variable (e.g., slip rate) during the coseismic phase. This type of non-linearity can be visualized by analyzing the hidden state variables in a phase diagram (Fig. 4.11), while its impact on data assimilation is analyzed by visualizing the trajectories of the ensemble members during the analysis step in both phases of the seismic cycle (Fig. 4.12).

The phase diagrams of the slow slip events have a smoother transition between the interseismic and coseismic phases, presenting almost no sharp corners in the variables' relationships (Figs. 4.11a,c). As shown in Fig. 4.6, the ensemble distributions are Gaussian, and in these phase diagrams, there is a smaller lag in time between the truth and the ensemble mean observed by the closeness of the members to the truth. Besides, the ensemble better captures the truth in this smoother trajectory. In contrast, the earthquake phase diagrams reveal very pronounced corners and abrupt transitions, which lead to a split of the ensemble members' distribution into two groups, corresponding to time snapshots prior to and after the coseismic phase (Fig. 4.11d). This bi-modality of the ensemble is also seen in Fig. 4.7c, and although in the interseismic phase, the ensemble members surround the truth, they have a more significant lag in time when compared to the slow slip events since the transitions are sharper and restrictive. This time lag can be visualized as the distance between the members that stay further away from the truth and that will stay in the interseismic phase of a previous earthquake once the coseismic phase of the truth has already occurred.

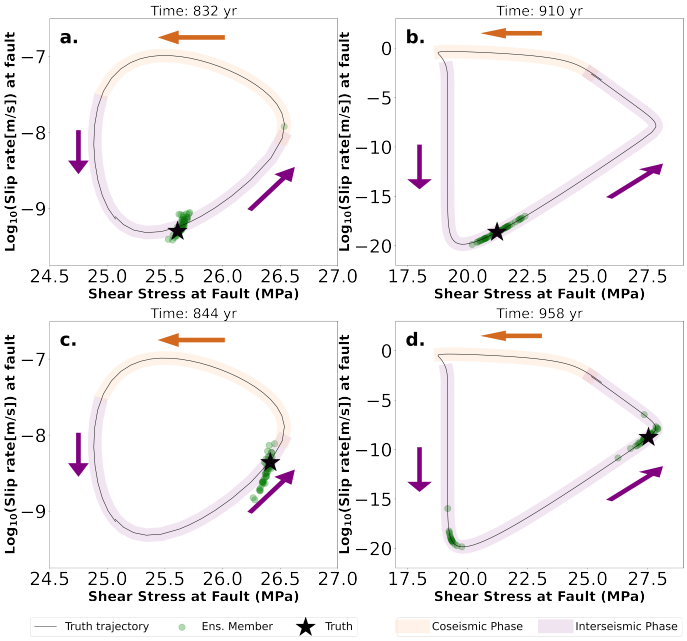


Figure 4.11: Phase diagrams showing the evolution of slip rate versus shear stress for a complete cycle (black line) of (a,c) slow slip events and (b,d) earthquakes. The top row shows the ensemble members (green circles) and the truth (black star) during the interseismic phase: (a) at time 832 years, and (b) at time 910 years, as shown in Figs. 4.6a and 4.7b, resp. The bottom row shows the ensemble members and truth before the coseismic phase: (c) at time 844 years and (d) at time 958 years as shown in Figs. 4.6a and 4.7b, resp. The purple hatched area represents the interseismic phase, and the light-orange hatched area the coseismic phase.

As mentioned before, the movements in the medium are a response to the displacements at the fault. This means that sharp transitions in state variables at the fault should translate into sharp transitions in the medium. We further analyze these relationships and effects between variables by cross-plotting the estimated hidden states and the observed variables during a single analysis step just before a coseismic period (Fig. 4.12). This shows how the non-linearities identified in the phase diagrams translate into a disruption of the estimates of the EnKF (Fig. 4.12). We see that the ensemble members stay close to each other for the slow slip events, and the updates from the EnKF are not as large as in the earthquake estimates. We also observe a deviation of the ensemble members from the true cycle of the slow slip events in Figs. 4.11a and 4.11c. This effect can be explained if we consider the 1D earthquake model as a 1D spring-slider model (e.g. Burridge and Knopoff 1967). Suppose we use this model to simulate slow slip events. In that case, the ratio between the stiffness of the spring and the so-called critical stiffness of the 1D spring-slider model will determine if the events will have a decaying, increasing, or no effect on their trajectories. This ratio is slightly less than 1 for the slow slip events shown in Fig. 4.11, resulting in cycles with slightly growing orbits in the phase diagram. In the earthquake models, this effect does not occur because of the added effects of inertia and radiation damping that force the ensemble members to stay in the same cycle.

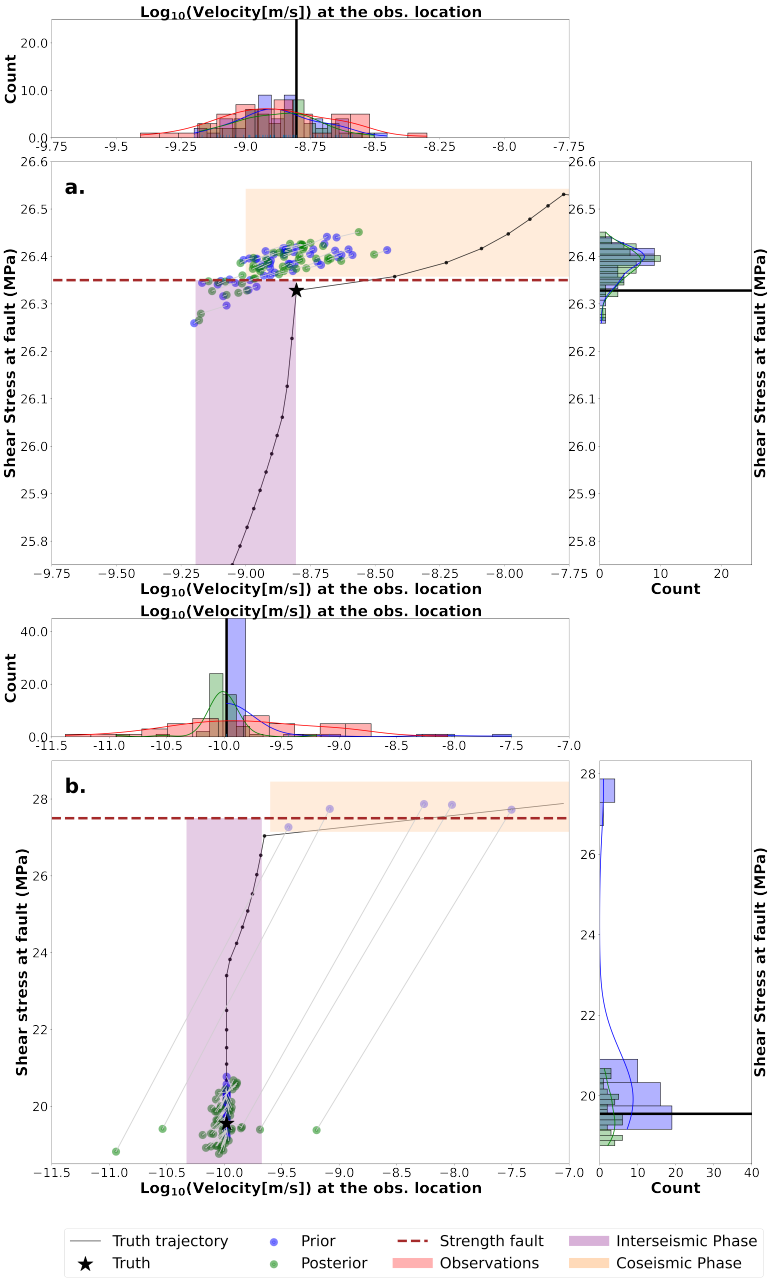


Figure 4.12: Ensemble results explaining how the update of the analysis step occurs when getting close to the coseismic phase at (a) 844 years for the slow slip events and (b) 958 years for the earthquakes. The horizontal axes show observed velocity, and the vertical axis shows the unobserved shear stress at the fault. The ensemble members are represented with circles. The blue circles are the prior values before the analysis step. The green circles are the posterior values after the analysis step. The light gray lines connecting the circles are the updates whose slope is the Kalman gain component corresponding to each ensemble member. The histograms on top and the right of the cross-plot correspond to the marginal distributions of the prior (blue) and the posterior (green). The red distribution in the upper histogram is the perturbed observations of velocity measured at the medium. The black lines show the trajectory that the truth follows during the earthquake cycle for the current event that is being estimated. The kink in this trajectory indicates the sudden transition from the interseismic phase toward the coseismic phase. The black star is the current shear stress at the fault and velocity in the medium for the truth.

The cross-plot from the earthquake models shows that a few members are already experiencing a higher velocity because they are accelerating toward an earthquake. These accelerating members behave as outliers in the sense that when corrected, they lie outside the overall trend observed for slip rates as illustrated by the prior ensemble members and truth trajectory. Nonetheless, despite the clear presence of bi-modality in the prior (e.g., Fig. 4.11d), the estimated shear stresses at the fault are projected back to end up close to the true shear stress (black star). A good shear-stress estimate is more important than a good slip-rate estimate because the forward model re-calculates slip rates using the estimated shear stresses (sec. 4.2.1). This means that the slip-rate estimate of the data assimilation will be overwritten in the next propagation time step.

Interestingly, the comparison of the analysis for slow slip events and earthquakes shows that the update of the ensemble members in the earthquake experiments brings the ensemble members' shear stress and velocity closer to the truth than in the case of the slow slip events. Despite a more angular trajectory of the true shear stress in the earthquake experiment, the average slope represented in the Kalman gain (grey lines) suggests an effective update of the the shear stresses at the fault. These findings suggest that the data assimilation is almost equally effective in estimating the occurrence of earthquakes as it is in estimating the occurrence of slow slip events. This interpretation is supported by the evaluation of the fit of the shear stress evolution of the ensemble members to the truth (Fig. 4.4), and by the quantification of the shear stress error (Figs. 4.6 and 4.7), and forecastability (Fig. 4.10).

Finally, the above analyses and the relatively minor difference in performance between slow slip events and earthquakes also suggest that the non-Gaussianity of the prior and the strong non-linearity of the forward model do not significantly hamper the effectiveness of the EnKF.

### 4.4.2. ANALYSIS OF NON-PERIODIC EVENTS

In previous sections, we assumed that the parameters were known and constant in our perfect model experiments. This assumption produces a very periodic behavior in the earthquake sequences. In this section, we further study the performance of the EnKF for estimating non-periodic earthquake sequences. Li et al. (2022) derive an equation that theoretically predicts the recurrence interval $T$ of an earthquake sequence for 1D, 2D and 3D models. For a 1D model, the equation is as follows

$$T = \frac{\Delta\tau_{dyn}}{\dot{\tau}_{h*}} = \frac{\sigma_n(b-a)H}{2GV_l}\ln\frac{V_{dyn}}{V_l}, \tag{4.2}$$

where $\Delta\tau_{dyn}$ is the stress drop, $\dot{\tau}_{h*}$ is the stress rate at the fault location, and the dynamic slip velocity $V_{dyn}$ is approximated as 1 m/s for simplicity. We perturb the loading rate $V_l$ to generate a non-periodic truth and keep the other variables shown in Eq. 4.2 constant. We assume a multiplicative noise $\beta$ to follow a Gaussian distribution $\beta \sim \mathcal{N}(1, 0.1)$ and apply it to the loading rate $V_l$ in the boundary conditions shown in Eq. 2.19.

The new boundary conditions are:

$$v_y(x=0) = \frac{1}{2}V,$$
$$v_y(x=H) = \beta\frac{1}{2}V_l. \tag{4.3}$$

Fig. 4.13 shows the variability of the recurrence interval for a synthetic truth from a non-periodic model generated by the perturbation of the loading rate. The events in the periodic model shown in the main text have a recurrence interval of 17.8 years. The non-periodic events generated using the multiplicative noise on the loading rate have a mean recurrence interval of about 18 years with a standard deviation of approximately 5 years, as shown in Fig. 4.13.



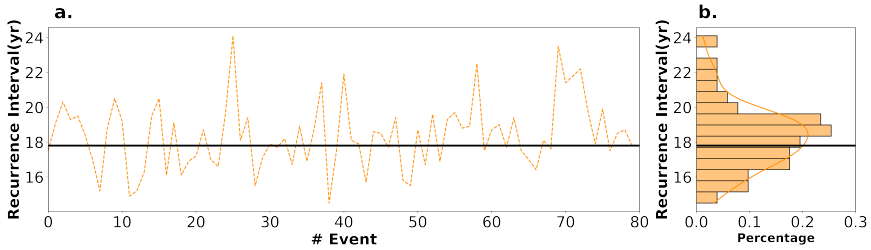


Figure 4.13: (a) Variability of the recurrence interval of the synthetic truth generated using a perturbation of the loading rate for a sequence of 80 earthquakes and (b) a histogram showing the distribution of the recurrence interval for the earthquakes from the sequence. The orange dashed line shows the variability of the recurrence interval for the earthquakes in the non-periodic model. The solid black line shows the periodic model's recurrence interval of 17.8 years.

We perform perfect model experiments where we assimilate synthetic observations of shear stress and velocities from the non-periodic model and estimate the shear stress, slip rate and state $\theta$ at the fault as shown in Fig. 4.14. The ensemble size is 50, and the ensemble members are periodic. The ensemble members also have the same loading rate with a recurrence interval of 17.8 years, but each has a different initial shear stress value at the fault at $t = 0$.

The accuracy of the EnKF is lower for the non-periodic events compared to the periodic events shown in Figs. 4.5. The EnKF gives relatively accurate estimates of the shear stresses, velocities, and state $\theta$ for earthquakes with a shorter recurrence interval than 17.8 years. However, there are large errors during the interseismic phase for earthquakes with a longer recurrence interval than 17.8 years. Fig. 4.15 shows similar results, indicating that the state space is better covered when the members have a longer recurrence interval than the truth. In contrast, the ensemble members tend to enter the coseismic phase too early, when the members have a shorter recurrence interval than the truth, leading to incorrect estimates, especially during the interseismic phase. For some of the earthquakes with longer recurrence intervals, the coseismic phase is well captured by the EnKF estimates when the following earthquake cycle corresponds to a shorter recurrence interval earthquake.

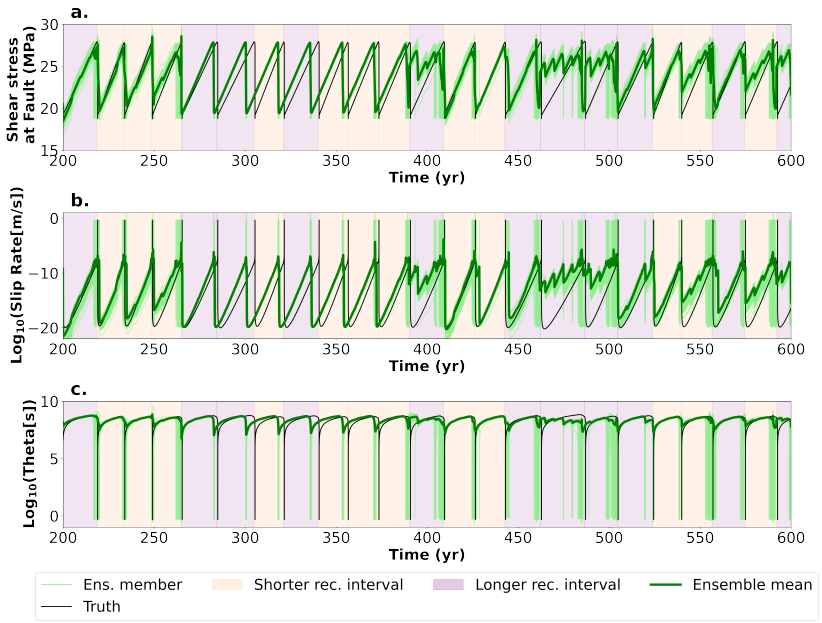


Figure 4.14: Estimated evolution of (a) shear stress, (b) slip rate, and (c) state $\theta$ at the fault for the non-periodic earthquake sequences. The black solid line is the true evolution of the variables. The estimates of the EnKF are shown in green. The color of the hatched zones differentiates between the events with a recurrence interval shorter than 17.8 years (light orange) and longer than 17.8 years (light purple).

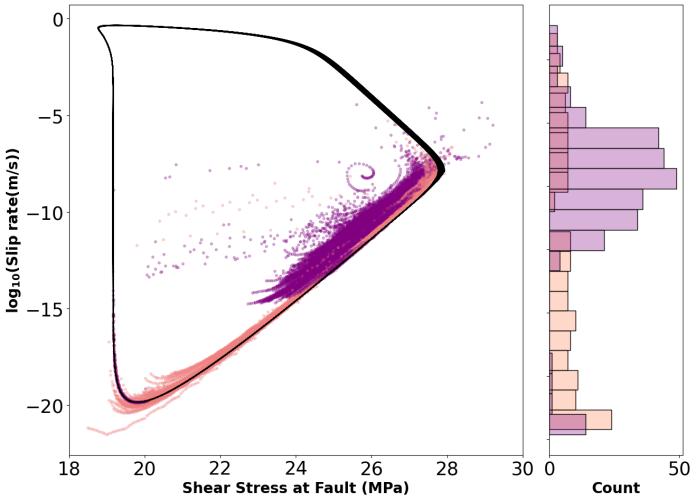


Figure 4.15: Phase diagram for the non-periodic earthquake sequences. The black solid line is the true evolution of the variables. The light orange dots are the ensemble member estimates for events with a recurrence interval shorter than 17.8 years, and the light purple dots are the estimates for events with a reccurence interval longer than 17.8 years (light purple).

Fig. 4.16 illustrates how the RMSE for the shear stress, slip rate, and state $\theta$ is larger for the non-periodic case than for the periodic case, with the difference between the slip-rate RMSE for these two cases being relatively small compared to the difference in RMSE of the other two variables.

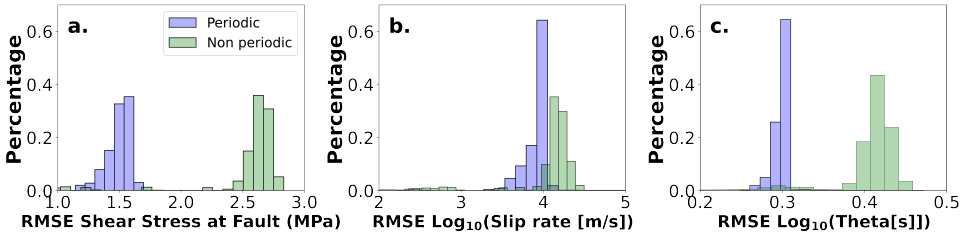


Figure 4.16: Comparison of the RMSE for the shear stress (a), slip rate (b) and state $\theta$ (c) for the periodic (blue) and non-periodic events (green).



### LIMITATIONS

We use a 1D model that is inherently periodic as we consider our frictional properties to be constant in time, and there are no spatial heterogeneities. One of the advantages of working with such a simple setup is that we can understand the impact of the rate-and-state friction law in a simple case. Another advantage is that this type of perfect model experiments with such simplifications also helps to evaluate how the assumed settings affect the assimilation results. This is very useful when moving to a more realistic application since it helps to evaluate the impact of factors such as the amplitudes of the observation errors, observation intervals, the observed variables, and measurement location. However, we expect that to fully exploit the capability of the physics-based model to estimate the shear stresses at the fault level through data assimilation; we should use higher dimensional models. These limitations are tackled by the use of higher dimensional models in 2D and 3D that are already available in Garnet (Li et al. 2022).

The use of 3D models especially helps to evaluate the benefits of the assimilation of observations in estimating the spatial distribution of the shear stress at the fault. However, they are too computationally expensive. In Chapter 6, our work is oriented into using ensemble data assimilation methods for assimilating data observed in a meter-scale biaxial shear stress experiments using 2D models that can still reproduce well some of the friction macroscopic measurements at a lower computational costs (Ji et al. 2022a; Spiers et al. 2017). These experiments have the advantage of offering data over a wide range of loading rates in which, in every single step, the parameters are relatively constant. This allows us to evaluate the effect of the uncertainty in the frictional parameters separately from the complexity of the current stress field experienced by the fault.

The study presented here is also a first step toward applying data assimilation with real data. The synthetic experiments presented here assumed that the shear stress and velocities close to the fault are directly observed. In reality, one of the likely difficulties is that we cannot observe the stress and velocity directly in a real case. This means that we cannot use a linear observation operator. Including more complex observation operators may introduce an additional source of nonlinearity that complicates the estimation of earthquake occurrences as shown in Appendix D. An alternative approach

to overcome this difficulty would be to estimate the shear stress and the slip rate along the plate interface based on crustal deformation through kinematic inversion and assimilate this information as a direct observation. However, it is important to consider that the kinematic inversion can introduce additional errors that may be hard to trace given the system's complexity. Another important simplification of our state estimation approach for a real application is that we assume that parameters are correct, known, and constant. As studied by Banerjee et al. (2023) and Hirahara and Nishikiori (2019), having a bias in models with constant frictional parameters affects the accuracy of the shear stress and velocity estimates even in the case of seismic events of long duration, such as slow slip events. As we have shown in this study and further explore in Chapter 5, the nonlinearities in seismic events of short duration, such as earthquakes, already challenge the accuracy of the estimates. Therefore, an approach with state and parameter estimation or that includes model error could be an alternative for a real-data application where the recurrence intervals are more variable and parameters are uncertain. The estimation of parameters helps the reanalysis of past earthquakes. Also, it gives additional degrees of freedom to the ensemble of shear stresses, slip rates, and state $\theta$ as defined by the rate-and-state friction law. However, it is important to note that including uncertainties in the parameters may lead to different nucleation lengths across models and, consequently, different time-step lengths in the model, among others difficulties as discussed in Chapter 2. This would make the data assimilation more computationally expensive and reduce the members' comparability. Comparisons of the effectiveness of the EnKF for non-periodic cases have been made by van Dinther et al. (2019), who benchmarked the forecastability of event alarms using ensemble data assimilation of non-periodic events compared to periodic recurrence models. Their results suggest that data assimilation in models of earthquakes with slow slip rates outperforms the periodic recurrence models, especially for moderately large events. Their study highlighted the advantages of ensemble data assimilation for including physics-based information in probabilistic hazard assessment. We further tested the state estimation of a non-periodic truth using an ensemble of periodic models (Appendix A). For a non-periodic case, we see that the accuracy of the EnKF estimates is less than in the state estimation of a periodic truth. The results suggest that when the earthquake recurrence interval of the truth is shorter than the earthquake recurrence interval of the prior, the EnKF gives a relatively accurate estimate of the shear stresses and velocities. Finally, other data assimilation methods than the EnKF can be implemented, like the particle filter, which has no assumptions of the Gaussianity of the state variable distributions. The particle filter or other data assimilation methods for non-linear models can be explored to better estimate the coseismic phase (see, for example, Chapter 19 in Evensen et al. (2022)).

### 4.4.3. IMPLICATIONS

As mentioned in Chapter 1, there is an increased interest in dynamic source inversion to combine information from physics-based models and observations. Multiple advances have been made in kinematic inversion and dynamic inversion of past earthquakes. However, these techniques are limited to past earthquakes. Data assimilation has the potential for combining both sources of information and providing useful reanalysis of past events and forecasting future earthquakes. The results from this study suggest that

ensemble data assimilation effectively estimates both slow slip events and earthquakes. In particular, the forecastability results show that with a very short alarm duration, many future slow slip events and earthquake occurrences can be estimated with our setup. With this proof of concept, we hope to catch seismologists' attention for using data assimilation to advance the field of earthquake forecasting.

## 4.5. CONCLUSIONS

In this chapter, an Ensemble Kalman Filter (EnKF) on a 1D model representing a 0D fault point loaded by a displaced, elastic medium assimilates synthetic, noisy, and indirect observations of shear stress and velocity. Assuming that the parameters of our physical model are perfect and using an ensemble with 50 members, we estimate the shear stress, slip rate, and state variable $\theta$ at the fault. We further evaluate the forecastability of the filter to estimate future occurrences of both slow slip events and earthquakes. Our results suggest that the EnKF is a useful and promising method for quantifying the uncertainty of the current state of stress, slip rates, and strength of faults.



We conclude that the estimates of the EnKF are most accurate during the interseismic phase of both the slow slip- and earthquake cycle. As an example, the absolute errors in the shear stress estimates are around 3 to 5% of the stress drop during the interseismic phase, while the standard deviation is slightly higher, between about 4 to 7 % of the stress drop. In contrast, the largest estimation errors are found during and around the coseismic phase of the earthquakes, where the shear stress errors, for example, can reach about 20 to 25 % of the stress drop shortly before and after the coseismic phase. The fast changes in the state of stress and velocity in this phase result in a sudden change in the distributions of the estimated variables. The distribution of variables becomes broader, and in the case of earthquakes, it becomes bimodal, which can introduce biases in the mean estimates of the variables.

The EnKF effectively estimates the occurrence of earthquakes that last only seconds, while observations are available over decadal time scales. An analysis of the influence of the observations during the analysis step shows that the assimilation of shear stress observations is very useful for the system to better estimate the shear stress on the fault for slow slip events and earthquakes. The velocity observations are most influential on estimates of slip rate for slow slip events. For earthquakes, both types of observations are relevant for the estimation of the slip rate. For the estimates of the state $\theta$, both types of observations are equally important for slow slip events, but for earthquakes, the shear stress is the most influential. Finally, a comparison of the evolution of the influence of the observations in time indicates that observations taken after the slip occurrence are especially important for the estimates of the filter. This is more evident in earthquakes than in slow slip events and shows the importance of the assimilation of observations recorded from previous earthquakes for better estimating the next ones.

An additional analysis of the forecastability of the EnKF for slow slip events and earthquakes shows that for both types of events, there is a very low forecasting failure rate of about 10 % when ringing very short alarms of just 10 % of the recurrence interval of the events. That means that most slow slip events could be forecasted half a year before their occurrence and around 2 years before the earthquakes with an occurrence interval of approx. 20 years. Our results suggest that data assimilation has the capacity

to improve estimates of fault-slip occurrence for both slow slip events and earthquakes and the potential to eventually advance the field of earthquake forecasting.

### DATA AVAILABILITY

The forward model for (a)seismic slip sequences utilizing the numerical modelling package Garnet is made accessible via repository https://bitbucket.org/cpranger/garnet/src/master/. The data produced and analyzed in this chapter is available via 4TU.ResearchData (http://doi.org/10.4121/20260932).

# 5

# NON-GAUSSIAN ENSEMBLE DATA ASSIMILATION METHODS FOR OPTIMIZED EARTHQUAKE FORECASTING

***Summary:*** *This chapter explores the potential of the Adaptive Gaussian Mixture Filter (AGMF) and the Particle Flow Filter (PFF) in earthquake modeling, addressing the limitations of the Ensemble Kalman Filter (EnKF) in handling non-Gaussian distributions in seismic cycles governed by rate-and-state friction (RSF) laws. We examine the AGMF and PFF using a Burridge-Knopoff 1D model. Our tests include periodic and aperiodic conditions to evaluate the impact of Gaussian prior assumptions on ensemble methods. The study also investigates whether the improved estimation of unobserved variables, previously demonstrated with the AGMF and PFF in the Lorenz 96 model in Chapter 3, can be effectively applied to earthquake forecasting and modeling.*

***Resumen:*** *Este capítulo explora el potencial del Filtro de Mezcla Gaussiana Adaptativa (AGMF) y del Filtro de Flujo de Partículas (PFF) en la modelación de terremotos. Se abordan las limitaciones del Filtro de Kalman de Ensamble (EnKF), especialmente en su capacidad para manejar distribuciones no gaussianas en ciclos sísmicos regidos por las leyes de fricción rate-and-state (RSF). En nuestro análisis, empleamos el AGMF y el PFF utilizando un modelo unidimensional de Burridge-Knopoff. Las pruebas realizadas incluyen tanto condiciones periódicas como aperiódicas, con el fin de evaluar el impacto de las suposiciones de distribuciones gaussianas inherentes a los métodos de ensamble.*

## 5.1. INTRODUCTION

In this chapter, we concentrate on the intricate behaviors arising from the rate-and-state friction in earthquake models, mainly focusing on the emergence of non-Gaussian distributions under certain conditions. As discussed in previous sections, the rate-and-state friction model can yield various outcomes based on initial conditions and system parameters. As highlighted in Section 4.3.3, a key observation is the presence of non-Gaussian prior distributions during earthquakes characterized by fast slip rates and sharp, rapid stress drops. These distributions, predominantly observed during the coseismic phase, present significant challenges to traditional data assimilation methods like the Ensemble Kalman Filter (EnKF).

Expanding upon the concepts introduced in Chapter 3, we explore various non-Gaussian data assimilation techniques better equipped to handle these complexities. A critical aspect of our study is exploring the impact of aperiodicity on the filter estimates. Attempts to induce aperiodicity, as described in Section 4.4.2, involved perturbing the loading rate of the system, a key boundary condition. We investigate how parameters such as the sensitivity of velocity relaxation ($\epsilon$) can lead to aperiodicity and even the formation of strange attractors, as explained by Erickson et al. (Erickson et al. (2011)) and explained in section 2.4. This chapter examines both regularized and non-regularized rate-and-state friction models and their influence on these dynamics, and how, even in chaotic behaviors, the data assimilation methods continue to provide accurate estimates.

The primary objective of this chapter is a comprehensive evaluation of the Adaptive Gaussian Mixture Filter (AGMF) and the Particle Flow Filter (PFF) as advanced non-Gaussian data assimilation methods. This evaluation focuses on their application in systems influenced by rate-and-state friction, assessing how they compare and potentially improve upon the performance of the Ensemble Kalman Filter (EnKF) performance. We scrutinize their performance in estimating key variables such as shear stress, velocity, and the unobservable state $\theta$ under both periodic and chaotic conditions.

A vital aspect of our investigation is examining the adequacy of assuming Gaussian priors, a common practice in data assimilation, mainly when using rate-and-state friction models. We aim to understand the implications of restricting models to quasi-periodic solutions, a common occurrence in seismological modeling, instead of embracing more chaotic solutions, which are standard in other fields like meteorology. Furthermore, we assess the impact of assimilating different observations — including shear stress, velocity, or both — on estimating all state variables, even those unobserved or 'hidden.' Through this comprehensive analysis, we hope to contribute significantly to developing more accurate and efficient methodologies for earthquake forecasting.

## 5.2. METHODOLOGY

In this section, we use perfect model experiments to evaluate whether the good estimates observed in Chapter 4 under periodic and quasi-periodic conditions, meaning seismic cycles with little variation in their recurrence intervals, can extend to an aperiodic system. We also do these perfect model experiments to evaluate if we obtain more accurate results, using data assimilation methods with non-Gaussian prior assumptions such as the AGMF and the PFF, especially in the coseismic phase. We further evaluate whether having ensembles where all the members have parameters that force them to stay in a periodic behavior impacts the filters.

### 5.2.1. PERFECT MODEL EXPERIMENTS

In this chapter, we use perfect model experiments to evaluate the performance of the data assimilation methods. In these experiments, we generate a synthetic true solution and synthetic observations, and evaluate how well the filters estimate the state variables. We specifically used 1D Burridge-Knopoff models of 20 blocks presented in Chapter 2, and ensembles of 100 members. For generating the initial conditions for these 100 members we run a very long simulation of the truth of 10.000 time steps and randomly select the state of the system in 100 different time steps randomly selected, where every timestep has an equal probability of being picked. These 100 initial conditions are stored and used for the perfect model experiments for each data assimilation method tested, for making a fair comparison between methods. We base this on the strange attractor found by Erickson et al. (2011).

For the 1-D Burridge-Knopoff models the state vector is

$$\mathbf{z}_n^T = \left(\bar{\boldsymbol{\tau}}^T, \bar{\mathbf{u}}^T, ln\left(\bar{\mathbf{V}}+1\right)^T, \bar{\boldsymbol{\Theta}}^T\right)_n, \qquad 1 \leq n \leq N_m, \tag{5.1}$$

where we employ $ln(\bar{\mathbf{V}}+1)$ instead of $\bar{\mathbf{V}}$ to impose a positivity constraint. We solve the ODEs, presented in Chapter 2, using an embedded fourth order explicit Runge–Kutta method following the advice of (Erickson et al. 2011) with a time step $\Delta t$ of 0.01 time units. The frictional non-dimensional parameters used are shown in Table 5.1.

Table 5.1: Non-dimensional rate-and-state friction parameters for 1-D Burridge-Knopoff model coupled with rate-and-state friction

| Parameter | Symbol | Periodic | Chaotic |
|---|---|---|---|
| Sensitivity of the velocity relaxation | $\epsilon$ | 0.3 | 0.5 |
| Non-dimensional spring constant | $\xi$ | 0.5 | 0.5 |
| Non-dimensional frequency | $\gamma_\mu$ | 0.5 | 0.5 |
| Non-dimensional frequency | $\gamma_\lambda$ | $\sqrt{0.2}$ | $\sqrt{0.2}$ |
| Scaled steady-state friction coefficient | $\bar{f}$ | 3.2 | 3.2 |

An example of the temporal evolution and phase diagrams of the models used is shown in Fig. 5.1. The transition of the Lorenz 96 models from periodic to chaotic is generated by changing the forcing term in Eq. 3.32, as shown in section 3.5. In the case of the BK-RSF 1D we choose specific combinations of $\epsilon$, $\xi$, $\gamma_\mu$, $\gamma_\lambda$, and $\bar{f}$ which are among the ranges mentioned in Erickson et al. (2011). We use the non-regularized rate-and-state

friction law and for the evolution of the state $\bar{\Theta}$ we use the slip law. We do this to evaluate periodic and chaotic behavior in the rate-and-state friction since it has been reported that the regularized rate-and-state friction imposes mainly periodic behaviors (Erickson et al. 2011; Lapusta and Rice 2003). The specific parameter values used in the experiments are shown in Table 5.1.

Using the periodic solutions we define the cycle duration. The BK-RSF 1D's cycle spans approximately 18 time units (or 1800 steps). We used a default rate of 8 observations per cycle or which is the same 2.25 time units (225 timesteps) between observations. We assume Gaussian uncorrelated observation errors with diagonal matrices $C_{dd}$. We define the uncertainties using typical observation uncertainties used in other works (Banerjee et al. 2023). For $\bar{\tau}$ and $ln(\bar{v}+1)$ of the seismological models we used 0.6. We extract synthetic observations of the generated truth using these uncertainties and use the same synthetic observations for the perfect model experiments done with the different data assimilation methods.



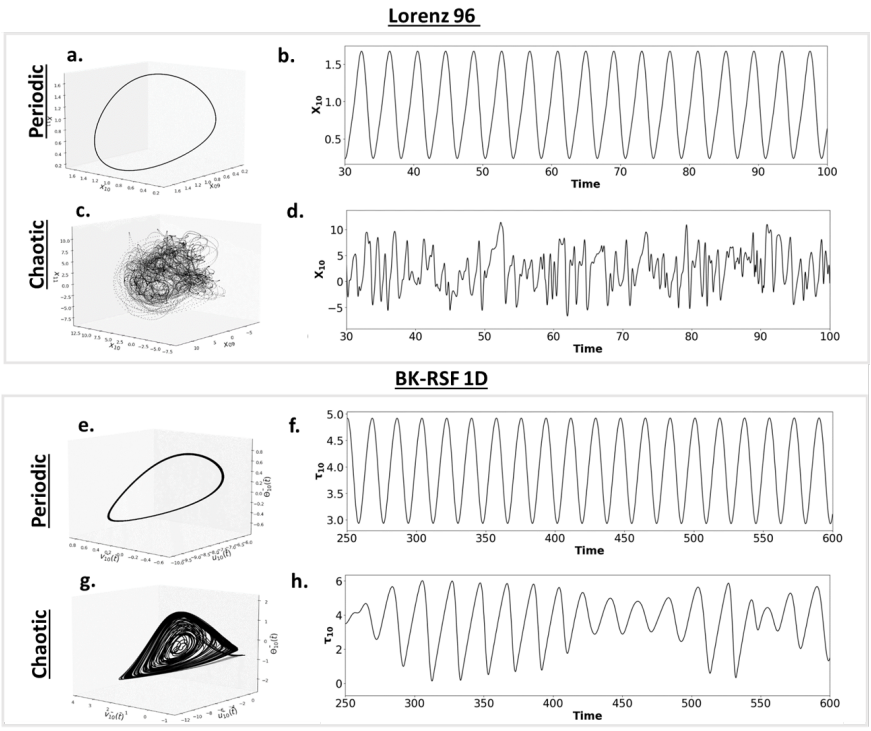


Figure 5.1: Schematic representation of the evolution of the Lorenz 96 (a-d) and 1-D Burridge-Knopoff model coupled with rate-and-state friction (e-h) use for creating the synthetic truth of the perfect model experiments. Phase diagrams of the Lorenz 96 models under (a) periodic (F=1.2) and (c) chaotic (F=8.0) conditions, and the evolution of the state of the 10-th cell through time for the (b) periodic and (d) chaotic case respectively. Phase diagrams of the BK-RSF 1D model under (e) periodic and (c) chaotic conditions, and the evolution of the shear stress for the 10-th block through time for the (b) periodic and (d) chaotic case respectively.

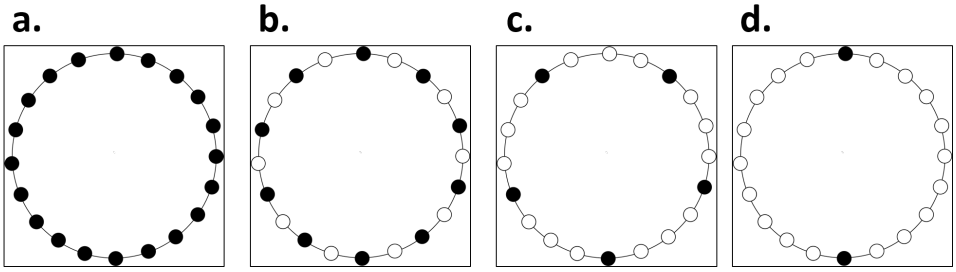


Figure 5.2: Schematic representation of the observation density for the data assimilation experiments for 4 different coverage of the blocks. Scenarios with a 100% coverage(a), with a 50 % (b), with a 25 % and with a 10%.

A higher proportion of observed model cells in data assimilation improves the assimilation results by reducing uncertainties and better capturing spatial patterns and variability. This increased coverage allows for better constraint of the model state and generates more accurate estimates. However, complete spatial coverage or observing all cells can lead to ensemble collapse, where the ensemble members become too similar, causing a lack of diversity in estimation. This convergence towards a single state limits the exploration of alternative states and underestimates uncertainties.

Our examination considered synthetic observations using different spatial coverages. We sample synthetic observations and perform perfect model experiments for the different data assimilation experiments using different spatial observation densities with coverages of 100%, 50%, 25 % and 10 % as shown in Fig. 5.2. These synthetic observations are extracted from the reference solution and perturbed using the uncertainties aforementioned. The observation vector is:

$$\mathbf{d}^T = \left(\bar{\boldsymbol{\tau}}^T, ln\left(\bar{\mathbf{V}}+1\right)^T\right). \tag{5.2}$$

For the localization of the prior covariance matrices, and defining the best hyperparameters for each filter, we followed the steps shown in Appendix A. Once the optimal hyperparameters have been chosen for the PFF and AGMF, we proceed to compare the results of the filters under the different cases with the EnKF to evaluate the advantages of data assimilation with non-Gaussian priors, as shown in the next section.

Other key considerations in data assimilation include observation frequency and error. The frequency of assimilation steps influences effectiveness; larger steps save computational resources but the estimation method may then overlook rapid changes, while smaller steps enhance temporal accuracy at increased computational costs. Observation error, the inherent uncertainty in measurements, affects the reliability of the assimilation process. High errors increase uncertainties, possibly leading to notable discrepancies between estimated and actual states. Conversely, very low errors might cause underdispersion, limiting the model's generalization capabilities. For a detailed discussion on varying observation frequencies and errors, see Appendix C.

## 5.3. RESULTS AND ANALYSIS

In this section, we present the outcomes from perfect model experiments, focusing on comparing the three data assimilation methods: EnKF, AGMF, and PFF. Our aim is to evaluate these methods in terms of accuracy, using RMSE values, and ensemble spread, assessed through rank histograms. This comparison extends across different phases of the seismic cycle, emphasizing how the ensemble spread changes with each phase of the seismic cycle.

### 5.3.1. ANALYSIS OF ERRORS AND UNDERDISPERSION

Fig 5.3 shows the RMSEs of the EnKF, the AGMF, and the PFF for the slip-rate $\bar{V}$,, which is an observed variable. The results show the comparison of the RMSEs when observing all the blocks (left column) and when only observing half of them (right column) for the periodic and the chaotic case. The results show that the three methods have estimates with errors lower than the observation error, as expected. The EnKF shows the lowest errors when having access to the observations of all the blocks. Interestingly, the AGMF has lower errors than the EnKF when fewer observations are available. The case with fewer observations presents a more challenging condition for the estimation, which makes the importance sampling step of the AGMF useful to capture the distributions of the variables better. The RMSE results show the same trend for the estimates of the shear stress $\bar{\tau}$.



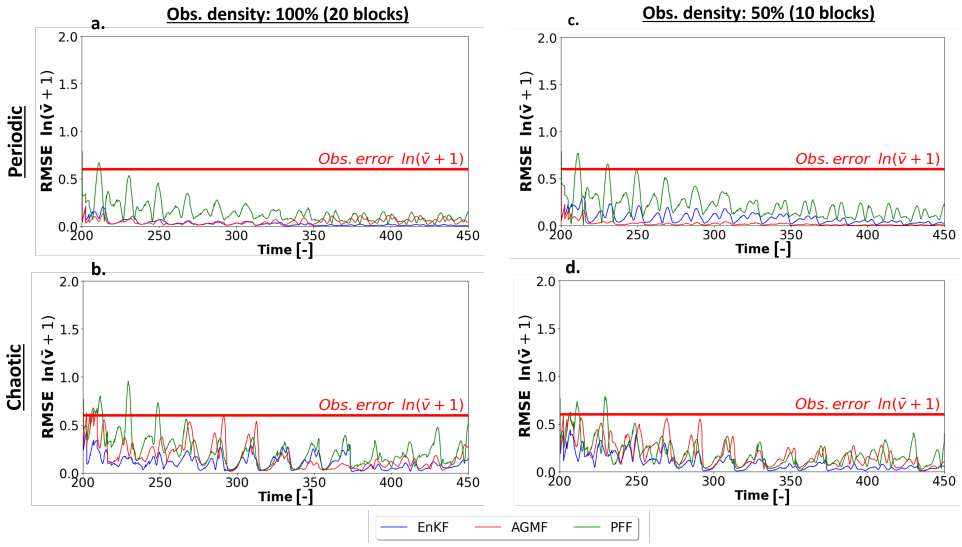


Figure 5.3: Comparison of the RMSE for the estimated of the logarithm of the velocity ($ln(\bar{V}+1)$) of the EnKF (green), the AGMF (red) and the PFF (blue) for the 1-D Burridge Knopoff model coupled with rate-and-state friction. The upper row shows the comparison for the periodic case (a,c), and the lower row for the chaotic case (b,d). The result correspond to an observation density of 100% of the blocks for the left column (a,b) and 50% of the blocks in the right column(c,d).

Fig. 5.4 show the RMSEs of the EnKF, the AGMF and the PFF for the state $\bar{\Theta}$ which is not observed also called a hidden state. The results show that for all methods the error decreases as more observation are assimilated with time. For the periodic case the ENKF and the AGMF have the lowest errors. However, for the periodic case the differences between the RMSE values for the different methods are less noticeable.

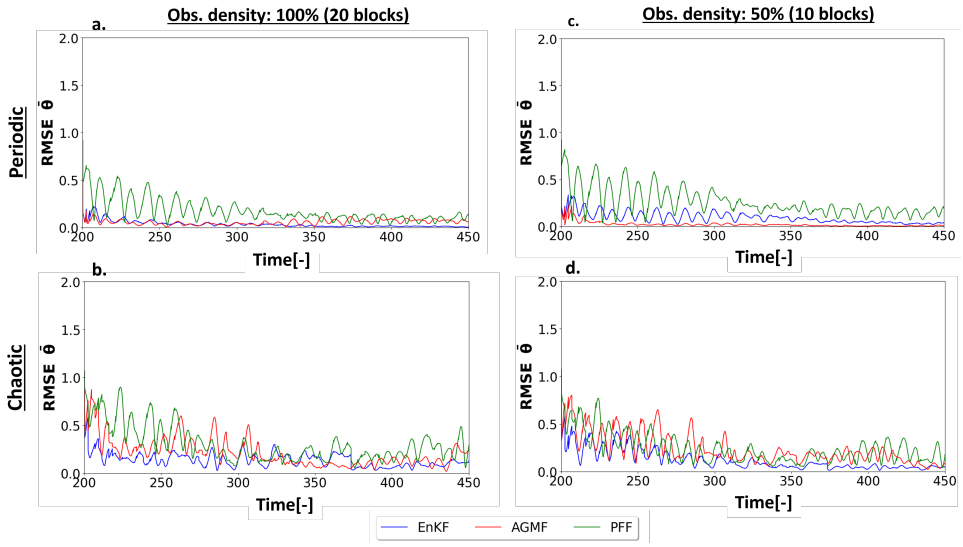


Figure 5.4: Comparison of the RMSE for the estimated of the state $\bar{\Theta}$ of the EnKF (green), the AGMF (red) and the PFF (blue) for the 1-D Burridge Knopoff model coupled with rate-and-state friction. The upper row shows the comparison for the periodic case (a,c), and the lower row for the chaotic case (b,d). The result correspond to an observation density of 100% of the blocks for the left column (a,b) and 50% of the blocks in the right column(c,d).

The DFS values are generally extremely low for the EnKF after the first observations have been assimilated. We find that the ensemble spread greatly decreases after these first assimilation windows. This indicates a problem of underdispersion, also called overconfidence. A possible remedy to this problem is the use of covariance inflation. However, very high inflation factors (2 to 5) are needed to have less underdispersion on the rank histograms. Interestingly, as we will see, the Particle Flow Filter does not experience this sudden decrease in the ensemble spread.

The rank-histograms (Fig. 5.5) highlight that the filters have problems of underdispersion. The resampling step of the AGMF can help to keep a wider ensemble, especially in the periodic case, but this resampling seems insufficient for reducing the underdispersion in the spread. Further refinement of the PFF's hyperparameters, such as bandwidth and learning rate, could yield more accurate and precise results while preserving an ensemble spread wide enough to correspond to the posterior uncertainties.

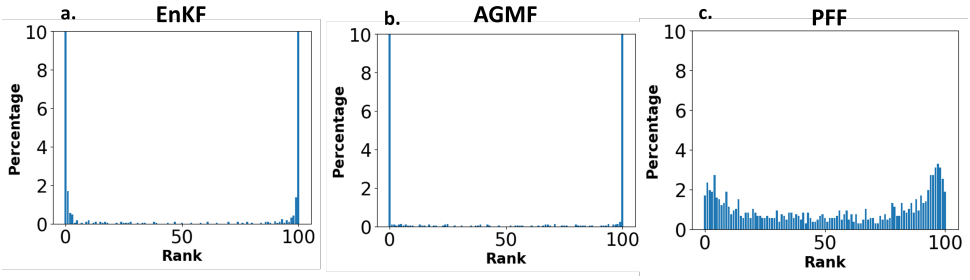


Figure 5.5: Rank histogram for the estimates of the (a) EnKF, (b) AGMF, and the (c) PFF for the periodic case

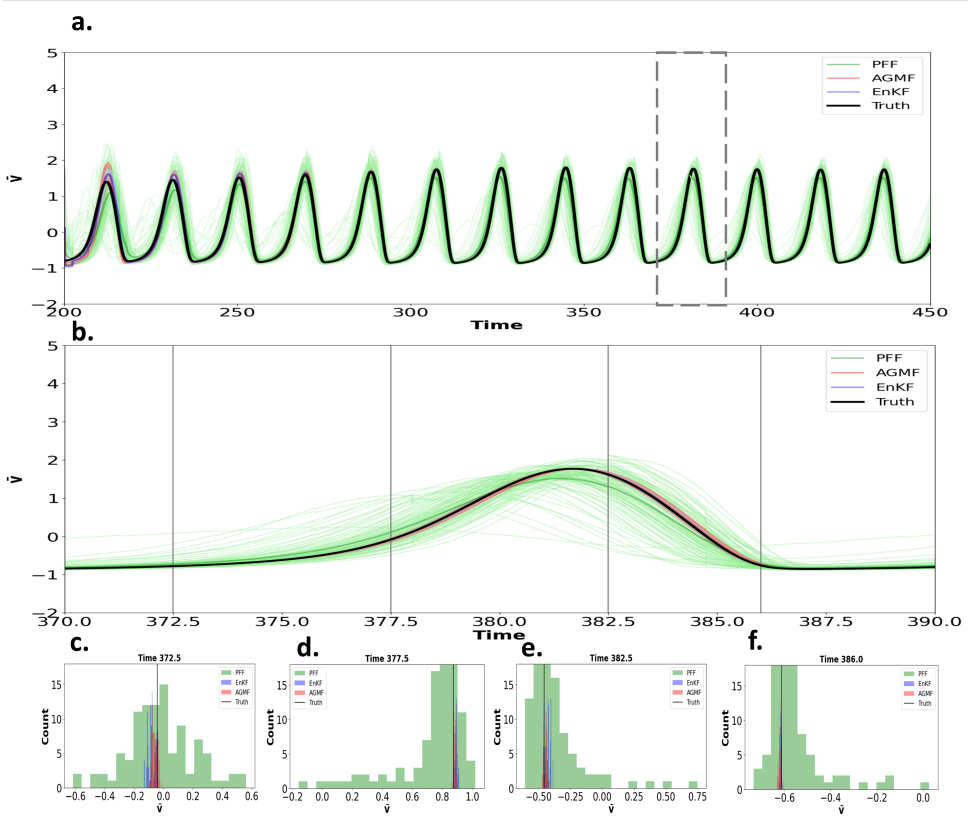


Figure 5.6: Comparison of the estimates of the slip-rate ($\tilde{V}$) of block 10 for the EnKF (blue), the AGMF (red), and the PFF (green) for estimating an earthquake occurrence for a periodic event. The true time series of the slip-rate is shown with a solid black line. We compare the ensemble distribution for the interseismic phase (c,d) during the coseismic phase (e), and at the end of the coseismic phase (f).

Figs 5.6 and 5.7 show a comparison of the time series estimates of the slip-rate($\bar{V}$) for the EnKF, AGMF, and PFF ensemble members. The results correspond to a sequence of periodic and chaotic events respectively. The histograms of the ensemble distribution of the different methods show that the PFF maintains a broad posterior distribution distribution in both cases. In contrast, the AGMF and the EnKF have very narrow ensemble distributions. Despite these narrow distributions, both methods have estimates that are very close to the truth. However, a consequence of the very narrow distributions is that the EnKF and AGMF ensemble will not cover the true state in certain phases.

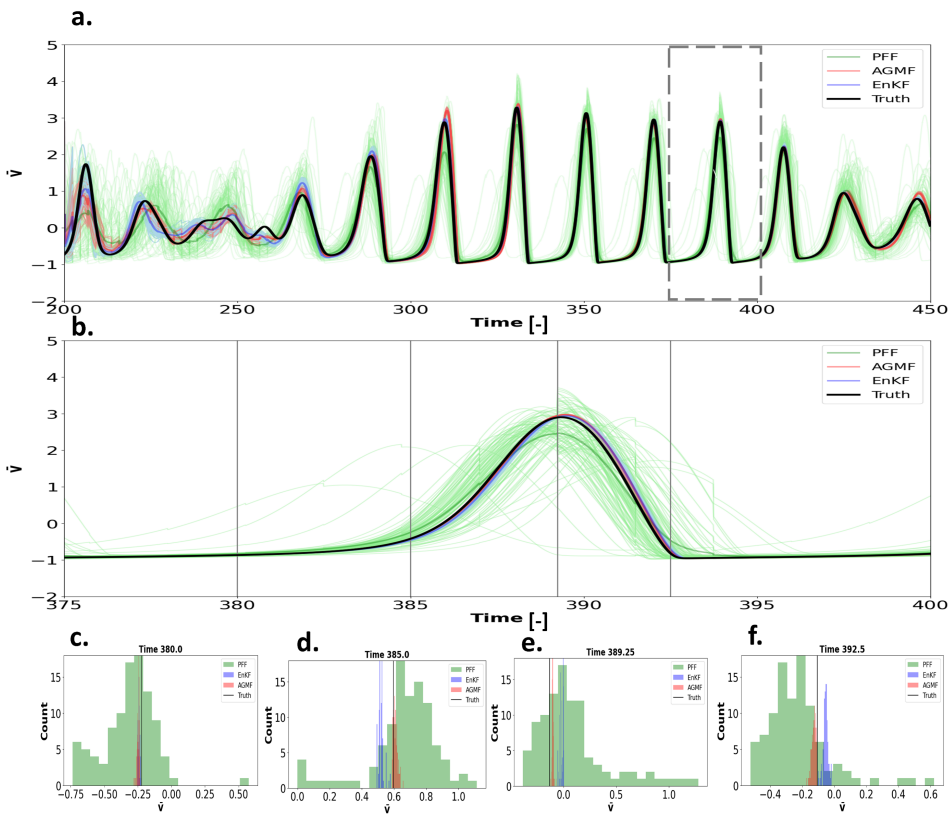


Figure 5.7: Comparison of the estimates of the slip-rate ($\bar{V}$) of block 10 for the EnKF (blue), the AGMF (red), and the PFF (green) for estimating an earthquake occurrence for a chaotic event. The true time series of the slip-rate is shown with a solid black line. We compare the ensemble distribution for the interseismic phase (c,d) during the coseismic phase (e), and at the end of the coseismic phase (f).

### 5.3.2. SENSITIVITY ON MODEL ERROR

Recent findings, e.g. (Gualandi et al. 2023), show how having a deterministic model representing a laboratory setup of a direct shear type of machine can constrain the solutions to just a set of possible states of the system. Gualandi et al. (2023) showed that even for a laboratory experiment with controlled conditions, introducing stochastic terms in the

system of ordinary differential equations was the most accurate approach for explaining the system's behavior.

We can achieve a similar result of this stochastic term by including model error terms in the state vector of the data assimilation. These model errors can account for missing physics or errors in the dynamical forward model. In this context, a model error can be used to better estimate the dynamics of the system and maintaining an ensemble spread that can help with the underdispersion problem that regularized rate-state-friction formulation imposed in methods like the EnKF.

To evaluate the effect of introducing a stochastic term in the equations, we visualize the effect of using such a term in a forward simulation of the 1-D Burridge Knopoff model. We aim to verify that we can use values of $\epsilon$ that produce periodic solutions and still estimate aperiodic behavior. The advantage of maintaining $\epsilon$ fixed is that we avoid further instability issues or changes in the frictional behavior of the system as described in section 2.4. For this, we perturbed the shear stress $\bar{\tau}$ as follows:

$$\bar{\tau}_i = \bar{f} + \bar{\Theta}_i + ln\left(\bar{V}_i + 1\right) + q, \tag{5.3}$$

where $q$ is the stochastic term that follows a distribution $q \sim$N(0,C$_{qq}$). We assume that the covariance matrix $C_{qq}$ is diagonal with $\sigma_q \in [0,1]$. Fig. 5.8 shows the evolution of the phase diagram of block 10 of a 1-D Burridge Knopoff with $\epsilon = 0.3$ in the periodic regime when increasing $\sigma_q$. We can see how the phase diagrams become more and more similar to the chaotic case shown in Fig. 5.1g, with an $\epsilon = 0.5$.

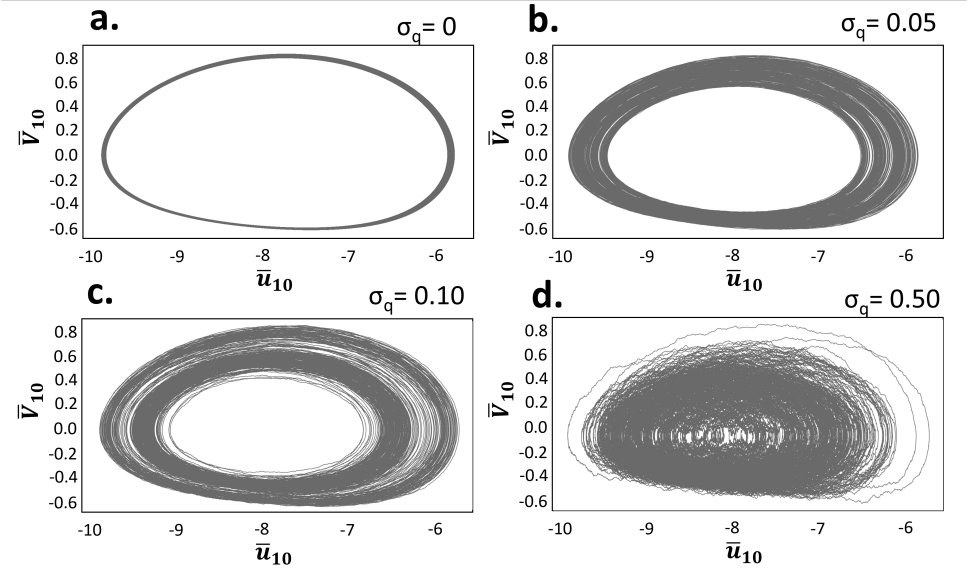


Figure 5.8: Phase diagrams for different model errors: (a) $\sigma_q = 0$,(b) $\sigma_q = 0.05$, (c) $\sigma_q = 0.1$ and (d) $\sigma_q = 0.5$ .

We see that Eq. 5.3 has a very similar form to Eq. 2.9, where the radiation term would be related to the model error term introduced. As explained before, the radiation damp-

ing term compensates for the loss of energy caused by the seismic waves after the fault's slip. The introduction of this extra term acts as a model error since it compensates for the quasi-dynamic model's missing physics. This term is, however, constant during the simulation and estimated based on the elastic properties of the rocks in frictional contact. I propose to make the model error a function of the slip-rate when using a radiation damping term and estimate the parameter $\eta$ during the data assimilation, calling it $q_{error}$. Then,

$$q = f(V) = q_{error}\bar{V}, \tag{5.4}$$

and also the new state vector would be,

$$\mathbf{z}_n^T = \left(\bar{\boldsymbol{\tau}}^T, \bar{\mathbf{u}}^T, ln\left(\bar{\mathbf{V}}+1\right)^T, \bar{\boldsymbol{\Theta}}^T, \boldsymbol{q_{error}}^T\right)_n, \qquad 1 \le n \le N_m. \tag{5.5}$$

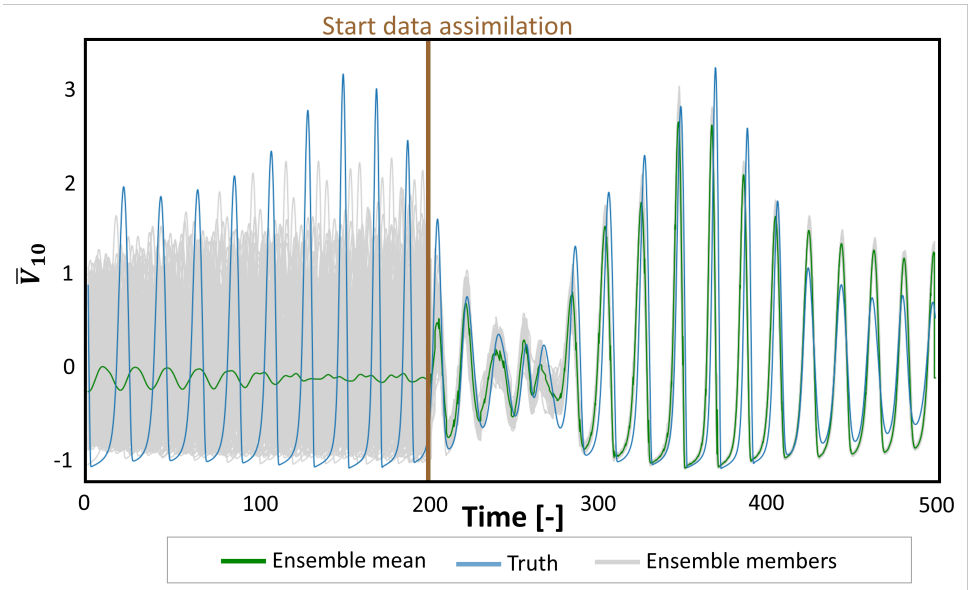


Figure 5.9: Comparison of the estimates of the slip-rate $\bar{v}$ at the block 10 of an EnKF with model error as part of the state vector. The mean values are in green. The truth, in blue, corresponds to the slip-rate of the chaotic model with $\epsilon = 0.5$. The individual ensemble members in gray are periodic with $\epsilon = 0.3$ and model error. The synthetic observations are extracted from the chaotic synthetic truth. The results show that despite the parameter bias the EnKF estimates are in sync with the truth especially in the occurrences of the seismic events, but with differences in the magnitude (amplitude of the signal).

The advantage of reformulating the assimilation this way, which is more similar to a parameter estimation exercise, is that knowing $q_{error}$ also helps us to investigate which processes could be missing/wrongly represented in the forward model, and use it to improve this model for forecasting applications. Fig. 5.9 shows the slip-rate estimates of an EnKF with periodic ensemble members that assimilate synthetic observations obtained from a chaotic truth. For the ensemble $\epsilon_n \sim \mathcal{N}(0.3, 0.02)$ while for the chaotic

truth $\epsilon = 0.5$. Fig. 5.10 shows the estimates of $q_{error}$ with time. We see that despite the parameter bias in $\epsilon$, the EnKF provides good estimates of the occurrences of the events in time. However, we see that similarly to the results of Section 4.4.2, the main differences between the truth and the estimates are in the amplitudes of the signals. This can be explained as the ensemble members with model error having a wider state space in the phase diagrams and, therefore, being able to estimate the occurrences of the earthquake as the truth will be in a smaller orbit covered by the ensemble. This explains why the best estimates of the EnKF occur when the amplitudes of the estimates of the filter are higher than the values of the truth. In contrast, the less accurate estimates occur when the filter underestimates the events' magnitude and the truth values are higher than the EnKF estimates.

These results are valuable since the correction with model error can improve the accuracy of estimating the occurrence of seismic events, even in the presence of parameter bias. Additionally, it allows simulation with a parameter that gives periodic and stable solutions with regularized formulations and still simulates and gives good estimates of aperiodic behavior.



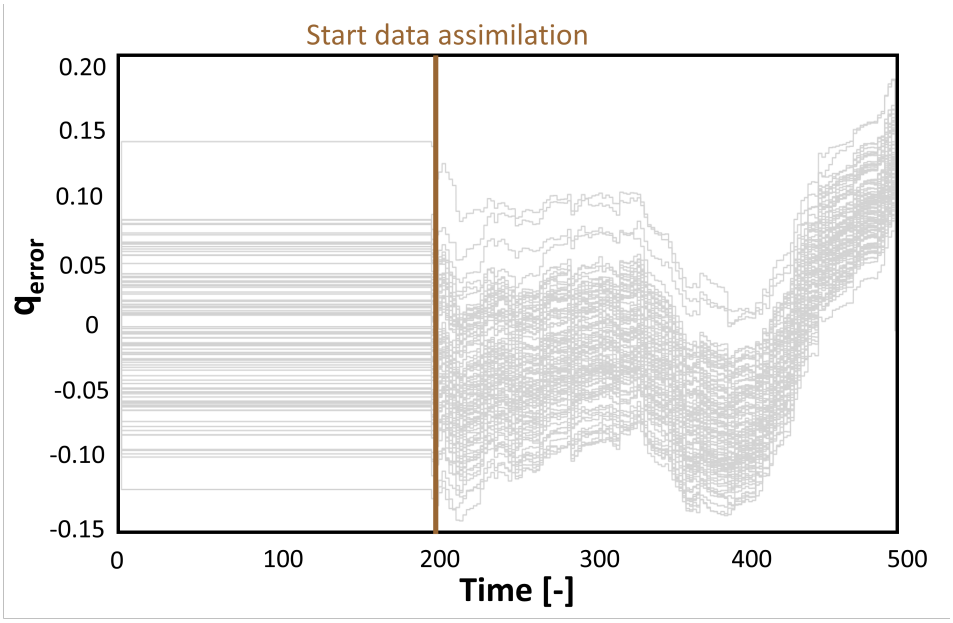


Figure 5.10: Time series estimation of the $q_{error}$

## 5.4. DISCUSSION

This chapter explored the application of non-Gaussian data assimilation methods on the 1-D Burridge-Knopoff model used in seismology. All methods tested yielded low RMSEs in perfect model experiments under periodic conditions for both models—the variation in results between the models links to their inherent chaotic and non-linear behaviors.

We also identified and further analyzed the role of prior knowledge in updates and the impact of including a model error term for better estimates in cases of parameter bias.

### 5.4.1. PFF'S SENSITIVITY TO HYPERPARAMETERS AND PRIOR KNOWLEDGE

The PFF was also tested on small chaotic dynamical systems by Stordal et al. (2021) including the Lorenz 96 model. Their results showed that the EnKF outperforms the PFF for intermediate ensemble sizes and the Particle Filter for large ensemble sizes. We observe similar results for an ensemble size of 100 members where the EnKF and PFF have very similar RMSE for the same ensemble size. The advantage of our results is that we use the 1-D Burridge Knopoff models that are not driven by noise as mentioned in Stordal et al. (2021) for the case of the Lorenz 96 model. Fig. 5.11 shows a comparison of RMSE results smoothed in time for the shear stress $\bar{\tau}$, slip velocity $\bar{v}$ and the state $\bar{\Theta}$ of two PFFs. The results presented with a dashed line correspond to a PFF whose attractive term (Eq. 3.26) only includes information from the likelihood. In contrast, the continuous line results include information from both the prior and the likelihood. Since the results are almost indistinguishable, it may lead to the conclusion that the filter becomes data-driven. Undesired behaviours like this required further study to apply this type of filter.



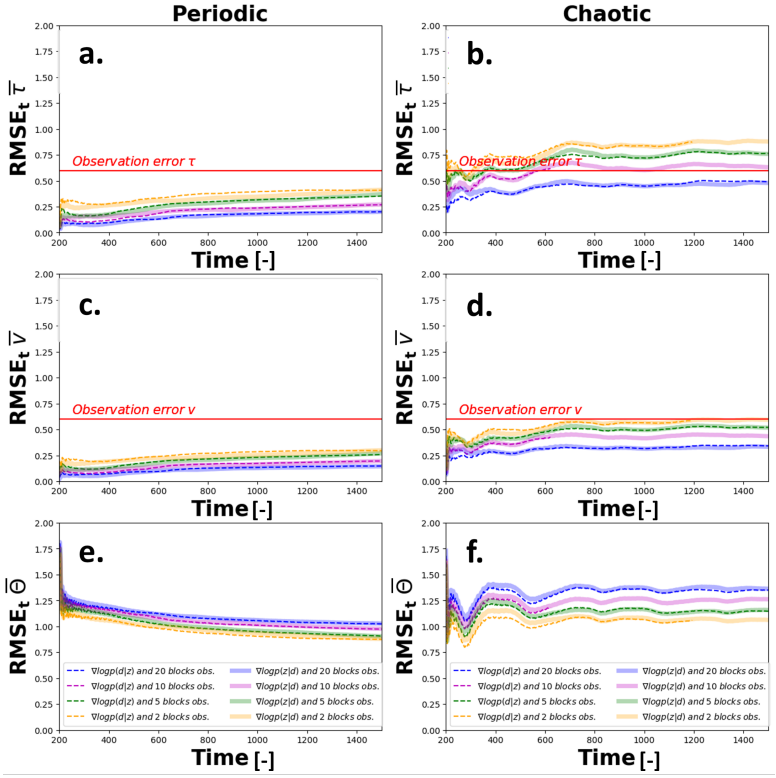


Figure 5.11: Evaluation of the effect of the prior information in the gradient of the log posterior.

### 5.4.2. LIMITATIONS

In this chapter, we employed 1D seismological models, which only simulate the seismogenic zone and neglect the surrounding medium. The lower computational cost of 0D and 1D models is beneficial for understanding the effects of the rate-and-state friction law on data assimilation. However, more complex and advanced 2D and 3D models are estimated better for the evolution of stress of the seismogenic zone and in the surrounding medium (Li et al. 2022). The 3D models are especially pertinent in determining shear stress distributions at faults and the nucleation process. We further discuss the benefits of these 2D and 3D models in Chapter 6.

We simplified our state estimation by having fixed parameters. However, as highlighted by Banerjee et al. (2023) and Hirahara and Nishikiori (2019), having biased friction parameters affects the accuracy of the velocity and shear stress estimates. Addressing these discrepancies is essential, possibly through combined state and parameter estimation or model error assessment. It is important to highlight that parameter estimation, while beneficial, can also inflate computational demands by requiring smaller time steps to maintain stability in the simulations and challenge model consistency, as explained in chapter 2.



### 5.4.3. IMPLICATIONS

Dynamic source inversion, primarily used for past earthquake inversion, is now complemented by data assimilation to analyze past and potential future earthquakes. Our research suggests that ensemble data assimilation can accurately estimate the evolution of shear stresses, velocities and state $\theta$ of the rate-and-state friction laws in earthquake models characterized by chaos, aperiodicity, and varied recurrence intervals. Regularized versions of rate-and-state friction, usually yielding periodic solutions, face criticism due to origins in small-scale lab experiments. However, recent findings affirm the validity of these small-scale observations for larger setups, up to a meter (Ji et al. 2022b). Avoiding underdispersion when using periodic simulations in ensemble data assimilation and addressing model errors as proposed in this chapter is crucial for better estimates, especially in real-world scenarios.

## 5.5. CONCLUSIONS

In this chapter, we have conducted a detailed examination of the performance of the Ensemble Kalman, Adaptive Gaussian Mixture, and Particle Flow Filters applied to 1-D Burridge-Knopoff models under periodic and chaotic conditions. The Ensemble Kalman and Adaptive Gaussian Mixture Filters faced significant underdispersion issues, necessitating substantial adjustments to their prior covariance matrices. Under periodic conditions, meaning periodic seismic cycles, the Ensemble Kalman Filter achieved the lowest RMSE, yet underdispersion remained a problem for both it and the Adaptive Gaussian Mixture Filter.

Notably, particle flow filters proved more robust against underdispersion, particularly with integrating regularized frictional laws that lead to quasi-periodic behavior. Additionally, they offered more precise estimates for unobserved variables such as the state variable $\bar{\Theta}$ in the Burridge-Knopoff models. This advantage is valuable given the

brevity and scarcity of historical seismological data relative to the low frequency of significant tectonic earthquakes. Nevertheless, it's important to consider that the tuning of the bandwidth in particle flow filters can have a substantial impact on their performance. For example, certain very wide bandwidth may affect sample separation, influencing the kernel's behavior. Hence, it's advisable to adjust the bandwidth hyperparameter thoughtfully.

Our results highlight the potential of ensemble data assimilation techniques to reliably estimate the evolution of shear stresses, velocities, and the state variable $\bar{\Theta}$ in earthquake models governed by chaotic dynamics and irregular recurrence intervals. Regularized versions of rate-and-state friction laws, typically associated with periodic outcomes, have been scrutinized for being derived from small-scale laboratory experiments. However, recent evidence supports the relevance of these laboratory observations to larger-scale scenarios (Ji et al. 2022a). Since these periodic simulations are used to explain also large-scale experiments, it is important to consider model errors and under dispersion within ensemble data assimilation frameworks.

We have also highlighted the challenges the rate-and-state friction law poses, which can cause abrupt system behavior changes due to uncertainties in frictional parameters. These uncertainties can lead to convergence issues, ensemble degeneracy, and complications in data assimilation when parameters are incorporated into the state vector of a high-dimensional system. We proposed incorporating stochastic model error terms into data assimilation as a solution, providing the necessary flexibility to accommodate a range of stable solutions and enabling the estimation of aperiodic behaviors amid predominantly periodic solutions. This approach introduces additional stochasticity in the behavior to capture earthquake dynamics more accurately with data assimilation.

Finally, we discussed how the selection of numerical models and rate-and-state friction laws can predispose systems to quasi-periodic behaviors, potentially causing underdispersion problems that compromise the reliability of estimations from methods predicated on the assumptions of Gaussianity and linearity. We demonstrated that the Particle Flow Filter can maintain adequate variance in its estimates, which is crucial for applying laboratory or field data where the accuracy of the estimates in relation to the true state is often challenging to determine.



### DATA AVAILABILITY

The forward model for the 1-D Burridge Knopoff coupled with rate-and-state friction is made accessible via repository https://github.com/hdiabmontero/non_gaussian_data_assim/. The data produced and analyzed in this chapter is available via 4TU.Research Data (http://doi.org/http://doi.org/10.4121/f0f075f2-f45c-4f8c-9d1d-bde03baeae33).

# 6

# TOWARDS ESTIMATING THE OCCURRENCE OF FAULT-SLIP IN LARGE-SCALE FRICTIONAL EXPERIMENTS AND AT THE FIELD SCALE

***Summary:*** *This chapter presents the application of lessons learned from perfect model experiments to a large-scale laboratory experiment. We propose a methodology detailing how data assimilation can aid in estimating variables such as the distribution of shear stress at the contact surfaces between frictional bodies. Additionally, we conceptualize how future applications could utilize data assimilation experiments for estimating earthquake occurrences at a field scale.*

***Resumen:*** *Este capítulo aborda la aplicación de las lecciones aprendidas de experimentos con modelos sintéticos en un experimento de laboratorio de gran escala. Se propone una metodología que explica cómo la asimilación de datos puede ser útil para estimar variables como la distribución del esfuerzo cortante en las superficies en contacto. Además, se conceptualiza cómo futuras aplicaciones podrían emplear la asimilación de datos en experimentos para estimar la probabilidad de ocurrencia de terremotos a escala de campo.*

## 6.1. INTRODUCTION

In previous chapters, we have investigated the use of data assimilation for estimating earthquake sequences using 1D seismological models. While these models have been beneficial for a basic understanding of the rate-and-state friction law, their simplicity has limited broader applicability. Additionally, we have utilized synthetic observations in our experiments, assuming that our numerical models are error-free. These experiments were instrumental in understanding the influence of assimilating observations of specific variables (see Chapter 4) and the assumptions regarding prior distributions on our estimates (see Chapter 5).

One of the major challenges in the study of faults is their location deep underground, which makes direct observations difficult, as shown in Fig. 6.1. Laboratory experiments offer a unique advantage in this context. While laboratory experiments are limited in replicating the exact conditions of significant faults, they are invaluable for understanding seismic mechanisms under various conditions. They allow control over variables that influence friction, like pressure, stresses, and velocities (Di Toro et al. 2011) and enable the strategic placement of sensors at various points to closely monitor the system's evolution. This controlled environment is essential for a detailed study of the mechanics of seismic activities.



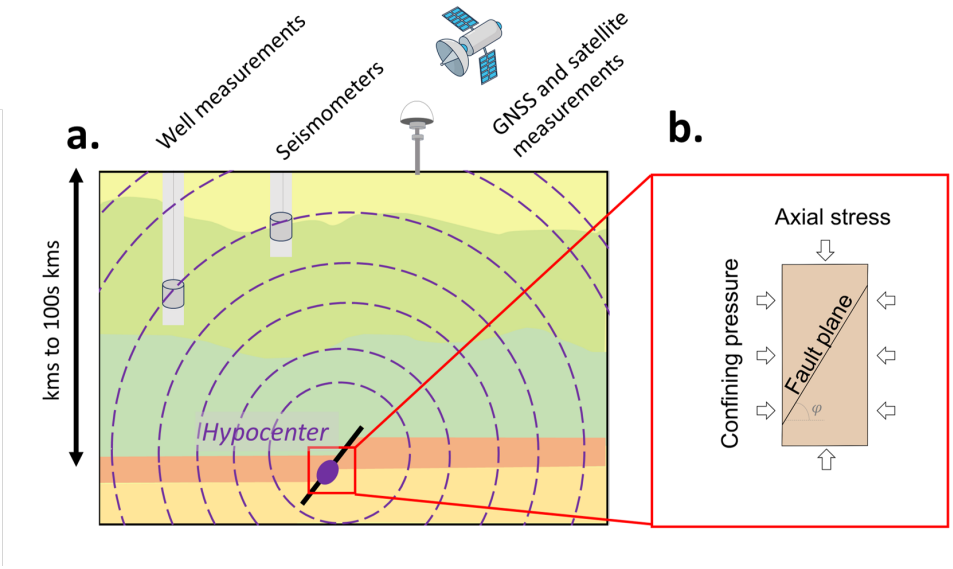


Figure 6.1: Comparison of the distance between seismic monitoring sensors and faults at different scales. (a) Field-scale seismic monitoring with distant well measurements and seismometers relative to the earthquake hypocenter, complemented by GNSS and satellite surface observations. (b) Laboratory fault slip experiment with sensors near the fault plane in a controlled setting.

This chapter focuses on two main goals:

- **Propose a data assimilation framework for laboratory experiments**: We propose a framework incorporating commonly used sensor data. We show some of the steps of this framework to evaluate the information content of some measurements in large-scale laboratory experiments and how the locations of the sensors with respect to the seismogenic zones affect the information content. For this evaluation, we use 2D models, which are key to overcoming some limitations of the 1D models used previously. These 2D models enable us to understand better the dynamics and information content of large-scale laboratory experiments. They allow us to evaluate the influence of having a discontinuous shear stress profile in the fault, as discussed in Chapter 5, and observations in the elastic homogeneous medium away from the fault, as in Chapter 4.
- **Propose a data assimilation approach for field-scale observations**: We introduce a method for deriving indirect stress observations applicable in a data assimilation framework. Moreover, we lay the groundwork for future research on the data assimilation of field-scale observations, especially in case of high slip rates. This chapter conceptualizes the integration of field measurements into a data assimilation scheme, acknowledging the challenges posed by the inability to observe stress and velocity in real-world scenarios directly. This approach is exemplified through a Groningen gas field case study, shedding light on some of the steps undertaken in this direction.



## 6.2. DATA ASSIMILATION AT THE LABORATORY SCALE

The conceptual data assimilation framework for estimating fault slip events in laboratory experiments, as detailed in Fig. 6.2, comprises three main steps. First is the forward simulation of an ensemble of numerical simulations, which accounts for the dominant physical phenomena influencing fault slip at the frictional surface (Fig. 6.2a). Second, is the derivation of laboratory experiment observations, considering their uncertainties (Fig. 6.2b). Finally, we generate posterior estimates of the current states of the stresses, velocities, and slip at the fault to estimate earthquake occurrences (Fig. 6.2c). The estimates from the data assimilation are then used to update the ensemble of numerical models, with an additional forward step possible for forecasting the fault slip.

As mentioned before, laboratory experiments have the advantage that sensors can be placed close to the fault, which is normally impossible in the field. A diverse range of sensors is used to measure critical physical quantities essential for understanding seismic mechanics. Some commonly used sensors are Linear Variable Differential Transformers (LVDTs), strain gauges, and laser displacement sensors for displacement measurement, offering different methods for monitoring movement. Stress and velocity are not directly measured variables but are derived from the forces and pressures measured using load cells, pressure transducers, or force gauges and the setup's geometry. Velocity is commonly derived from displacement measurements, and in more specialized setups equipped with cameras, techniques like Digital Image Correlation (DIC) and Photogrammetry can be applied. Wavefields are recorded using acoustic sensors, like piezoelectric transducers, transforming vibrations into electrical signals for moment tensor

inversion and amplitude change studies. A challenge when including these measurements is that it is difficult to determine their uncertainty since they initially involve electrical quantities like voltage and current, and post-processing and calibration are key to deriving mechanical quantities such as displacement, pressure, and force. Additionally, environmental conditions can affect sensor accuracy, and the derivation of critical model variables like velocity and stress from these measurements involves complex post-processing, adding to the uncertainties.

Data assimilation is vital in this context, helping to address these challenges by enhancing our ability to provide better estimates of variables like stress and slip rates, even with imperfect models. It can assist, for example, by helping to better quantify the uncertainties of the observations. Also, it can help to incorporate the data before the post-processing via observation operators and evaluate how estimates change before incorporating additional correlations between observations. This is crucial since multiple measurements influence the derived stress and velocity observations, and observation errors are commonly assumed to be uncorrelated.

In the following section, we apply some of the steps from the proposed flowchart in a large-scale direct shear setup.

### 6.2.1. CASE STUDY: LARGE SCALE LABORATORY EXPERIMENT



The large-scale experiment featured in this section stands out due to its specialized instrumentation, which provides extensive data often unavailable in standard frictional laboratory experiments. This research's laboratory setup involves a "direct shear test" assembly, consisting of two quadrangular prismatic sandstone specimens placed parallel to each other to represent a simplified horizontal fault. This setup is depicted in Figure 6.3. The measurements were carried out using a second-generation large-scale biaxial friction apparatus at the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan. This work was part of a collaborative research project between NAM and Utrecht University (UU) researchers (Spiers et al. 2017).

Figure 6.3: Large Scale Earthquake Simulator experiments at the National Institute for Earth Sciences Research and Disaster Resilience, Japan (Collaboration NAM & UU). Taken from (Fukuyama et al. 2014)

In this setup, the upper quadrangular sandstone specimen is 1.5 m long, and the lower one is 2 m long, with both having a thickness of 0.5 m. Three vertical load cells apply and record normal stresses from 1.5 to 9 MPa. The upper specimen is secured using a rigid frame and a reaction force bar, while the lower specimen is moved using a shaking table with a loading rate of up to $3x10^2 m/s$. The shear force applied is recorded by a horizontal load cell. The relative displacement of the specimens is measured using

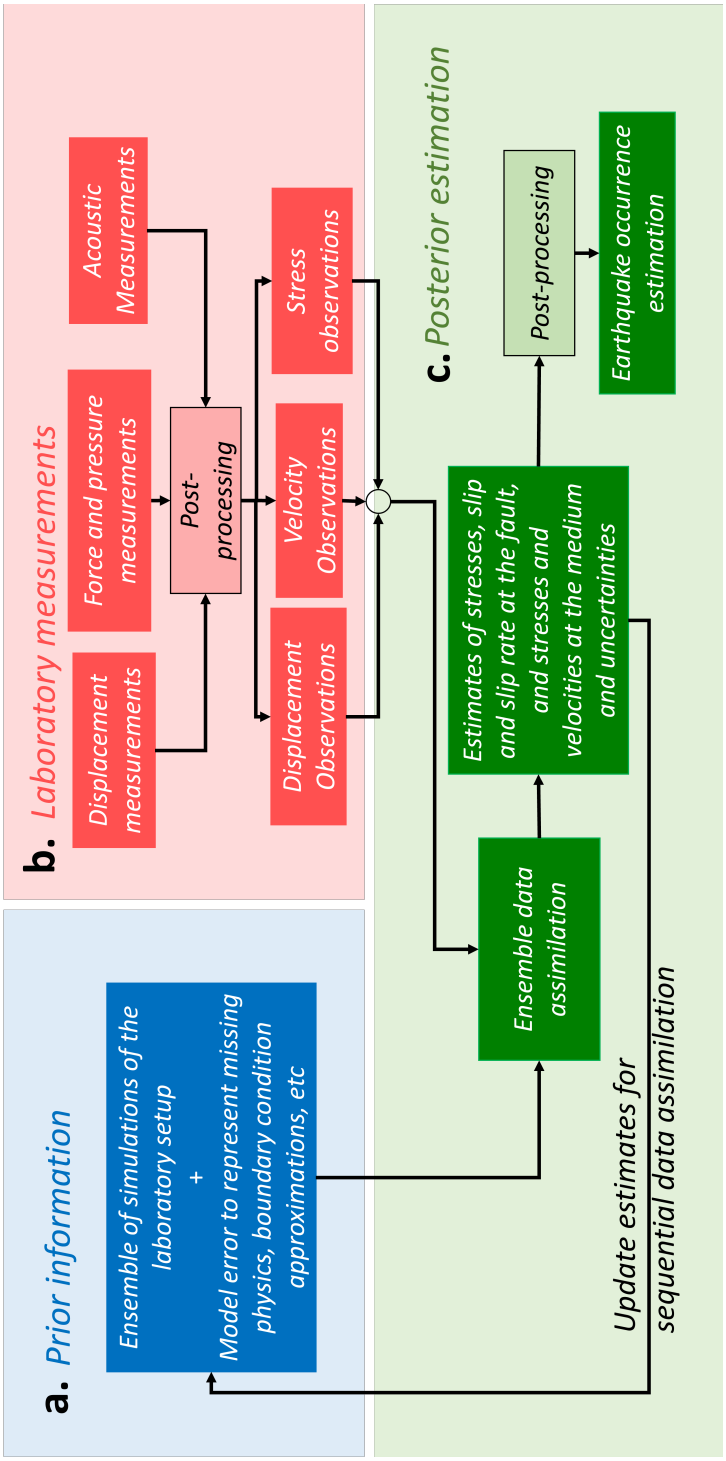


Figure 6.2: Flowchart showing the proposed data assimilation framework for estimating fault slip occurrence in laboratory experiments

lasers. An array of shear strain gauges and piezoelectric transducers are placed at a 20 mm distance from the contact surface (fault) of both sandstone blocks, specifically in the lower specimen (Fukuyama et al. 2014; Spiers et al. 2017; Yamashita et al. 2015). For measuring the displacements, 40 Rosette strain gauges are deployed on the back surface of the lower sandstone block, 2 cm under the rock-gouge interface, and every 4 cm to cover the maximum 1.5 m fault length. Additionally, pressure-sensitive films are used to measure the stress in the contact between the block specimens before starting the shearing.

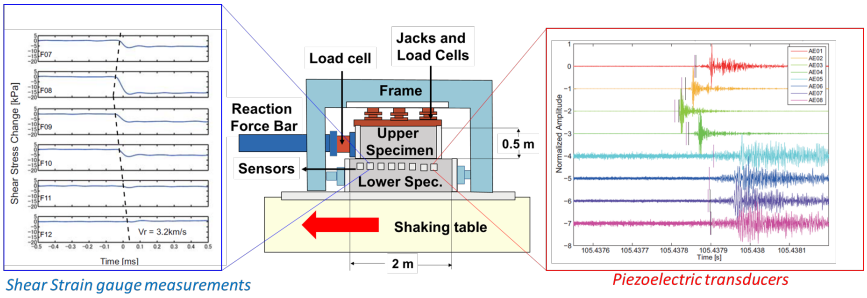


Figure 6.4: Example of measurements of the shear strain gauges and piezoelectric transducers part of the measurement arrays of the meter-scale experiment set-up. Modified from (Fukuyama et al. 2014)

6

Furthermore, piezoelectric transducers were placed close to the contact between blocks to record the wavefield during the shearing experiment. An example of such recordings, conducted in a separate experiment using Gabbro samples instead of sandstones, is shown in Fig. 6.4. During the entire time span of the experiments, high-resolution photographs were taken at a frequency of 1 MHz, but had to be downsampled during post-processing to 1 KHz. These photographs were used for Digital Image Correlation (DIC) analysis to provide information about the displacements and velocities in the sandstones during the experiments.

### NUMERICAL MODEL

In our approach to numerically modeling the slip evolution within the large-scale apparatus, we choose a symmetric setup, focusing on simulating only the lower half-space to represent the lower block in the large-scale biaxial friction apparatus accurately. This approach is detailed in Fig.6.5. To simulate the slip behavior observed in the sensor data, we employ a 2-D model. Following an exhaustive analysis, we determine that the configuration shown in Fig.6.5b most effectively replicates the observed behavior. This model configuration incorporates a gouge layer on top and a boundary condition simulated by springs. These springs are designed to emulate the equivalent stiffness of the reaction bars, actuators, and other assembly elements that affect the setup's overall stiffness. The equivalent stiffness, denoted as $k_1$, is conceptualized as a series of springs arranged in parallel, which is further illustrated in Fig.6.5a. The parameters we use in this numerical model are comprehensively summarized in Table6.1.

For a practical comparison, we juxtaposed the macroscopic friction measured in the laboratory for experiment LB 18-21, as detailed by Ji et al. (2022a), against the results of

our numerical modeling, which employed the configuration from Fig.6.5b. This comparison is visually presented in Fig.6.6.

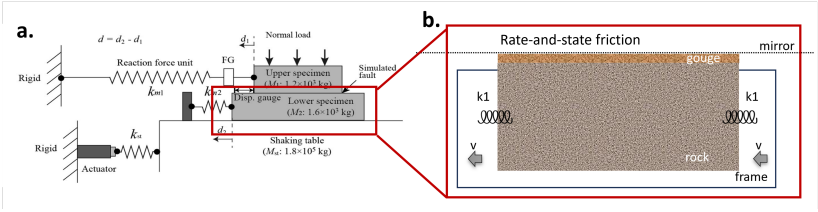


Figure 6.5: (a) Schematic representation of the large Scale Earthquake Simulator experiments at the National Institute for Earth Sciences Research and Disaster Resilience, Japan representing the actuators, reaction bars, and other sources of stifness of the machines Togo et al. (2015) (b) Numerical model representation of the lower block model including a gouge layer and a boundary condition to represent the equivalent stiffness and elastic forcing imposed by the machine. Adopted from Li et al. (2023b).

Table 6.1: Material and rate-and-state friction parameters for the 2D model setup of the laboratory experiment

| Parameter | Symbol | Values |
|---|---|---|
| Young modulus sandstone block | $E_r$ | 20.51 GPa |
| Young modulus gouge | $E_g$ | 4 GPa |
| Poisson ratio | $\nu$ | 0.16 |
| Density gouge | $\rho_g$ | 2000 kg/m$^3$ |
| Density sandstone block | $\rho_r$ | 2450 kg/m$^3$ |
| Static friction coefficient | $\mu_0$ | 0.671 |
| Reference slip rate | $V_0$ | 100 $\mu$m/s |
| Characteristic slip distance | $L$ | 18.2 $\mu$m |
| RSF Direct Effect | $a$ | 0.0024 |
| RSF Evolution Effect | $b$ | 0.0047 |
| Depth model | $H_x$ | 50 cm |
| Length model | $H_z$ | 150 cm |
| Loading slip rate | $V_l$ | 0.10 $\mu$m/s |
| Grid spacing | $\Delta x$ | 1 cm |

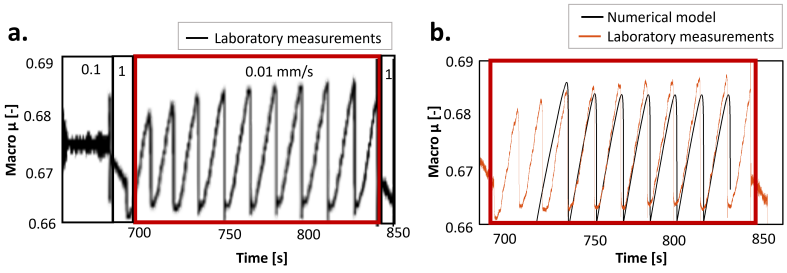


Figure 6.6: Comparison laboratory data and numerical model (a) **J. Yi, pers. commun., December, 2022** (b) Adopted from Li et al. (2023b).

### INFORMATION CONTENT EVALUATION

In assessing the influence of observations on the analysis step of data assimilation, we utilize the concept of degrees of freedom of signals (DFS), introduced in Chapter 3. To calculate the DFS, we construct prior covariance matrices using an ensemble of 10 simulations with varying profiles of normal stress. The state vector for the ensemble members, $\mathbf{z}_n$, is defined as follows:

$$\mathbf{z}_n^T = \left(\epsilon_{xz}^T, \epsilon_{yz}^T, \ln(\mathrm{V_y}), \ln(\mathbf{v}_\mathrm{y})^T, \ln(\mathbf{v}_\mathrm{x})^T, \ln(\theta)\right)_n, \qquad 1 \leq n \leq N. \tag{6.1}$$

The measurements from the pressure-sensitivity films are used to generate the ensemble members for data assimilation. Fig.6.7 outlines the three-step process for ensemble generation. Firstly, we process the pressure sensitivity film data to a resolution of 10 cells x 150 cells (Fig.6.7a), equivalent to a cell size of 1cm x 1cm based on the dimensions of the block specimens. Then, we extract stress profiles per row (1 cell x 150 cells) in the slip direction to create an average profile, depicted with a black line in Fig.6.7b. This average profile is then used to generate a synthetic profile (Fig.6.7c), to which we apply a bias of 2.5 MPa. This biased profile serves as the mean for generating random field perturbations, depicted as colored line profiles in Fig. 6.7c, which are then utilized to initialize the different ensemble members.



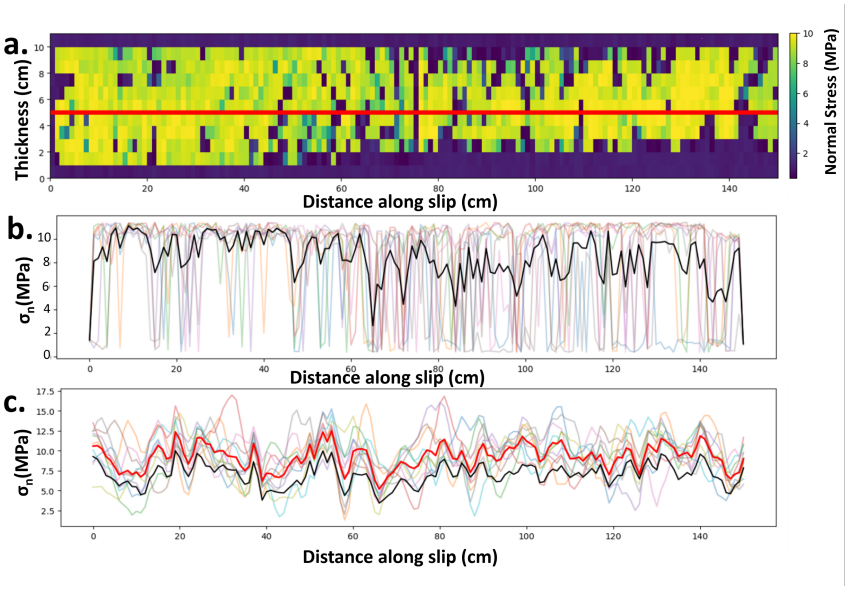


Figure 6.7: (a) Post-processed measurements of the pressure-sensitivity film for the experiment LB18-21 **J. Yi, pers. commun., December, 2022**), (b) Stress profiles of each row of the pressure sentivity film in slip direction, the black line represents the average profile when considering each row of the film and the colored lines represent each of the profiles (1 profile per cm) use to estimate the average profile. (c) Ensemble generated for the data assimilation, the black curve represent the true synthetic truth for the perfect model experiment based on the measurements, the red line represents the mean used for generating the ensemble members that has of 2.5 MPa from the truth, the ensembles are generated perturbing the red profile with random fields.

We analyze the impact of strain gauge measurements by constructing an observation operator based on sensor locations shown in Fig.6.8. An observational error of $1e-3$ is assumed for analyzing the DFS evolution over time, as presented in Fig.6.9. We estimate the DFS at specific times marked with red lines in Fig. 6.9, chosen to capture moments before and after a fault slip occurrence, thus reflecting changes in DFS throughout the seismic cycle. This analysis also varies the sensor distance from the fault, examining 10 different distances at 5 cm intervals.

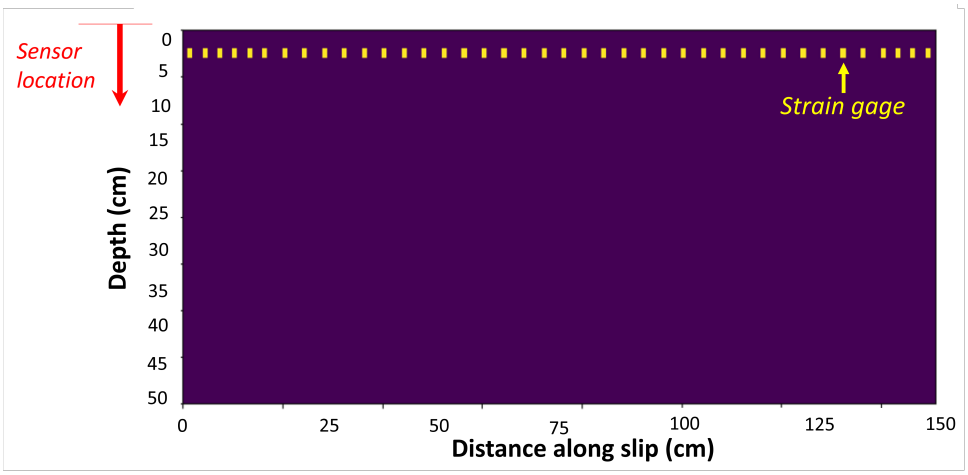


Figure 6.8: Diagram depicting the arrangement of strain gauge sensors. The sensors, represented by yellow dashes, are uniformly spaced along the horizontal axis, and only their position along the vertical axis is changed for the information content analysis.

Given that this analysis was conducted with only ten ensemble members, we acknowledge that our conclusions are preliminary. The small ensemble size may not adequately represent the covariance matrices. Consequently, this exploratory study highlights the need for further evaluation with a larger ensemble size to confirm these findings and understand the significance of strain measurements near the seismogenic zone for data assimilation. As shown in Fig.6.9, the DFS remains relatively low for most of the interseismic phase, with sensor location showing minimal impact during this period. However, a notable increase in information content is observed as the slip rate accelerates, peaking around the maximum slip rate. This increase aligns with the findings in Chapter4 regarding the Kalman gain elements in EnKF estimates. Post-coseismic phase, the DFS diminishes gradually, indicating the sustained value of information post-earthquake, potentially aiding in estimating subsequent events.

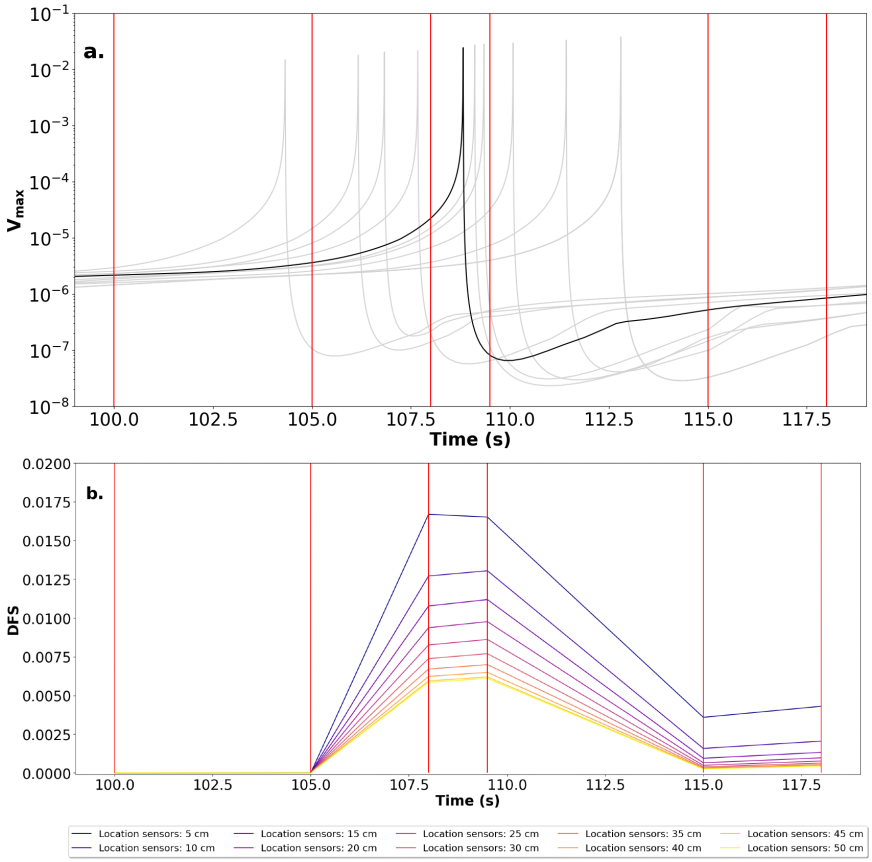


Figure 6.9: (a) Time series evolution of the maximum slip rate for a single seismic cycle in the seismogenic zone for an ensemble of 10 members(gray lines) and the synthetic truth (black lines) (b) Variations in time of the DFS for 10 different locations of the strain gauge sensors

### 6.2.2. DISCUSSION OF ASSIMILATION OF LABORATORY DATA

A significant challenge in data assimilation for fault slip analysis is accurately replicating behaviors observed in laboratory experiments, often due to complex boundary conditions such as those in large-scale laboratory experiments. Initially, numerical models were developed to capture basic frictional behaviors, but they often need help with incorporating additional phenomena like plastic deformation. This inclusion adds to the complexity and increases computational demands, leading to compromises such as smaller ensemble sizes in data assimilation efforts. Understanding the transition from stable to unstable motion in faults requires a deep understanding of frictional dynamics under different stress conditions. Research by Gualandi et al. (2023) shows the challenges in aligning experimental observations with numerical models. This challenge grows when adding a stochastic term to the model equations, especially in the context of data assimilation. Studies by Chen et al. (2021) and Mehranpour et al. (2021) focus on integrating microphysical behaviors into geomechanical models. Chen's work extends

the rate-and-state friction framework with grain-scale deformation mechanisms, while Mehranpour's research in reservoir compaction using the Discrete Element Method (DEM) looks at complex behaviors like nonlinear elasticity and grain breakage, which are more easily studied in laboratory experiments.

At the same time, the application of machine learning to analyze acoustic data in fault slip analysis is gaining attention. Changes in acoustic signals before seismic events, linked to fault grain rearrangements, have been noted in studies like those of Rouet-Leduc et al. (2017). These changes relate to the microphysical principles in the works of Chen et al. (2021) and Mehranpour et al. (2021). This is important for improving earthquake forecasting. Integrating this data into data assimilation frameworks could greatly improve forecasting abilities. However, using machine learning to analyze acoustic data for detecting fault rupture or slip comes with challenges, mainly due to potential biases toward specific experiment types and rock heterogeneity. Projects like those by Elbertsen et al. (2023), which use various rock types, try to address these biases. Data assimilation methods, effective in parameter estimation and managing uncertainty, are beneficial in this scenario, because they allow to systematically incorporate uncertain frictional parameters linked to diverse lithologies, enhancing the generalizability of the estimates. A promising strategy is a combined approach using acoustic data interpretation for shear stress information and data assimilation techniques, potentially overcoming the issue of not having direct shear stress measurements. We investigate this approach in our proposed data assimilation method for the field scale, which we consider a viable solution when direct fault measurements are unavailable. However, there is abundant information from seismic networks.



## 6.3. DATA ASSIMILATION AT THE FIELD SCALE

Like laboratory experiments, the data assimilation framework for estimating fault slip at the field scale involves forward simulation of an ensemble of numerical models, derivation of observations, and generation of posterior estimates of the current states of stresses, velocities, and slip at the fault. These estimates are then used to update the numerical model ensemble, with an additional forward step possible for forecasting fault slip. This process is illustrated in Fig. 6.10.

Understanding seismicity and fault dynamics at the field scale relies on various observational methods. Important tools include GNSS (Global Navigation Satellite System) for accurate location and displacement data and InSAR (Interferometric Synthetic Aperture Radar) for mapping surface deformation with high-resolution radar images. However, close to faults, the interferometric fringes of InSAR are densely packed, making them less reliable and difficult to interpret (Simons et al. 2002). Ground motion recordings measured using seismometers provide abundant data but fall short of directly revealing fault stress states. The sparsity of measurements, their large uncertainties, and the distance between observed and modeled areas make estimating the current stress in faults extremely challenging.

We propose including commonly used displacement measurements (GNSS) for kinematic inversion and dynamic source inversion (Lee et al. 2011; Ohta et al. 2012) and deriving indirect stress observations using information from seismic networks. We offer two methods for incorporating this information: (1) a method based on geophysical and

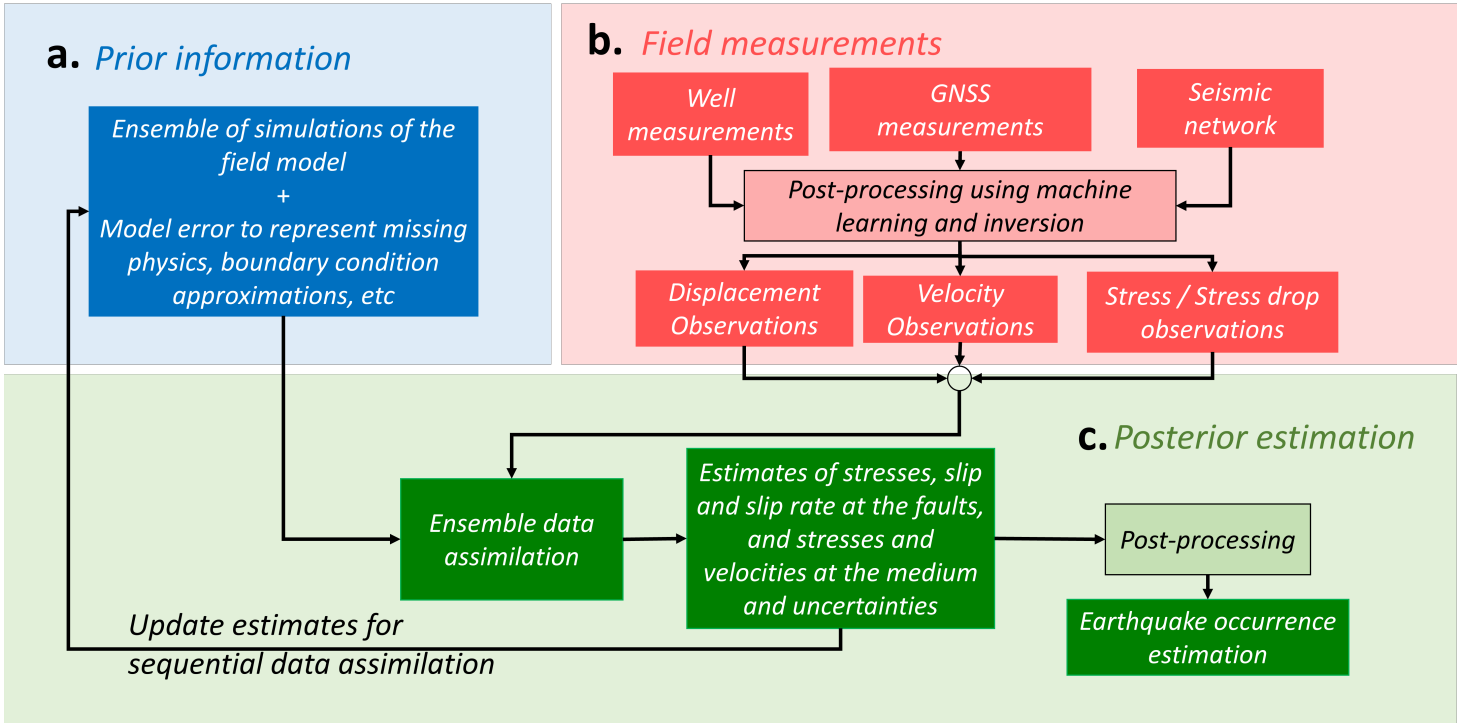


Figure 6.10: Flowchart showing the inclusion of the data assimilation of field observations

seismological knowledge to derive stress drop information from seismic recordings. The challenge here is that since our derived observation is the stress drop rather than the absolute stress value, it is difficult to include these observations in the data assimilation framework directly. Alternatively, (2) machine learning can derive shear stress observables from seismic waveforms. We illustrate some of these steps in the machine learning direction using neural networks that Deventer (2021) and Verwijs (2021) developed in their master theses, utilizing recorded and synthetic seismic data from the Groningen gas field.

### 6.3.1. PROPOSED EXTRACTION OF STRESS OBSERVATIONS FOR DATA ASSIMILATION

#### DERIVING STRESS DROP INFORMATION FROM SEISMIC RECORDINGS

The *stress drop* is defined as the average difference between the shear stress on a fault before an earthquake happens and the stress after the earthquake. Seismologists have derived analytical solutions for estimating the stress drop when a fault is embedded within a homogeneous material. Shearer (2019) presents some methodologies for estimating the stress drop using the general formulation:

$$\Delta\tau_{dyn} = CG\frac{\bar{D}}{\tilde{L}}, \tag{6.2}$$

where $C$ is a non-dimensional constant that depends upon the geometry of the rupture, $G$ is the shear modulus, $\bar{D}$ is the average slip of the fault, and $\tilde{L}$ is a characteristic rupture dimension which also depends on the geometry of the fault. This expression is based on the works of Eshelby (1957), who derived a similar expression for a fault with circular geometry, and Knopoff (1958), who derived the formulation for a rectangular fault. In general, these analytical expressions for the stress drop depend on the *seismic moment* ($M_0$):

$$M_0 = GA\frac{\bar{D}}{\tilde{L}}, \tag{6.3}$$

where $A$ is the rupture area along the fault. The seismic moment is an important quantity used to measure the size of earthquakes. It was first introduced by Keiiti (1966), and it is the basis used for the moment magnitude scale, the most widely accepted magnitude scale for ranking earthquakes by size (Hanks and Kanamori 1979). This quantity is estimated routinely by seismologists while analyzing seismic recordings of earthquakes and during the inversion of moment tensors.

The quantity $\tilde{L}$ can be estimated from seismic recordings methodologies like in Madariaga (1976) where analytical expressions are found for estimating the fault radius based on the observed corner frequency $f_c$ of the amplitude spectrum of the seismic waves. This allows us to derive expressions such as:

$$\Delta\tau_{dyn} = f(f_c,, M_0). \tag{6.4}$$

The advantage of such a formulation as Eq. 6.5 is that it allows the derivation of stress information directly from the post-processing of seismic recordings. For example, for a

circular fault, we have that:

$$\Delta\tau_{dyn} = \frac{7}{16}\left(\frac{f_c}{kV_s}\right)^3 M_0, \tag{6.5}$$

where $V_s$ is the shear velocity in the vicinity of the fault, and the factor $k$ is derived from theoretical considerations of the stress distribution and slip on the fault during the earthquake (Shearer 2019).

Now, for using the stress drop observation derived for the seismic recording during the data assimilation, we propose to assume that the fault is moving at a constant velocity after the earthquake; in other words, that it has reached the steady state as described in Section 2.2 and Fig. 2.3. We use some of the simulations of Chapter 5 to evaluate the validity of this assumption in a low-dimensional system. Fig. 6.11 shows a comparison of the periodic and chaotic solutions of the simulations for the middle block of the 1-D Burridge Knopoff model(in black) with the values of

$$\bar{\Theta}_{ss,i} = -(1+\epsilon)\ln\left(\bar{V}_i + 1\right), \tag{6.6}$$

that describes the steady state (in cyan).

Fig. 6.11 shows that for the periodic case, the system remains close to the steady state, as indicated by the overlapping lines in the graph, while for the chaotic case, we see that during the interseismic phase, the assumption is not valid. However, the assumption that the system is steady-state seems valid in the coseismic phase and after the events.



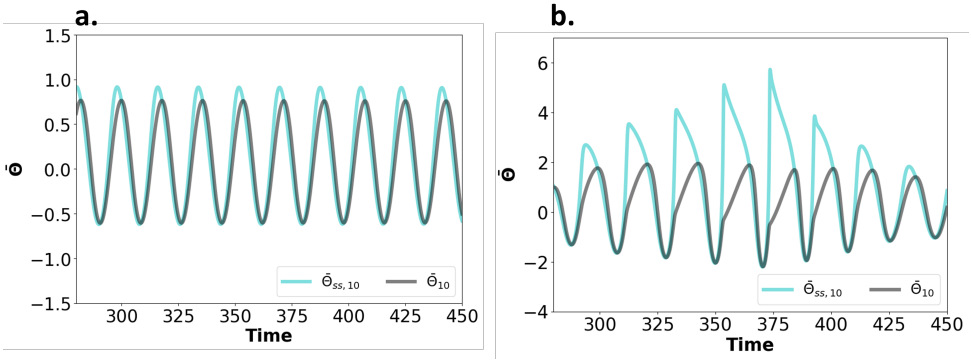


Figure 6.11: Comparison of the values of $\bar{\Theta}$(black) and the values of considering the system in a steady-state $\bar{\Theta}_{ss}$(cyan) for the (a) periodic case (b) chaotic case at the block 10 of the Burridge-Knopoff model.

In the steady state, we can use then Eq. 2.4 and Fig. 2.3a to define the following expression:

$$\tau = \tau_{peak} - \Delta\tau_{dyn} = \left[\mu_0 + aln\left(\frac{V}{V_0}\right)\right]\sigma_n - \Delta\tau_{dyn}. \tag{6.7}$$

This equation allows us to express the shear stress value as the difference between the peak stress before the earthquake($\tau_{peak}$) and the stress drop we estimated using the seismic recordings. However, this approach has not been tested, and perfect model experiments are still needed to evaluate whether Eq. 6.7 holds even in synthetic cases.

#### DERIVING STRESS INFORMATION FROM SEISMIC RECORDINGS USING MACHINE LEARNING

Alternative to the abovementioned approach, we propose a data assimilation approach with field data that utilizes artificial intelligence to extract observable information, following a method similar to the one presented in Fig. 6.12. The process begins with using seismic waveforms as input for the networks to determine the timing of events. Following this, we employ post-processing of the waveform for moment tensor inversion, also facilitated by artificial intelligence. This approach speeds up the process and offsets the high computational costs of forward modeling. Finally, the moment tensor inversion helps us define the seismic moment and, with it, the average slip and possibly stress information. We further explain the details of this proposed workflow using, as an example, neural networks trained using synthetic and real seismic recordings of the Groningen gas field.

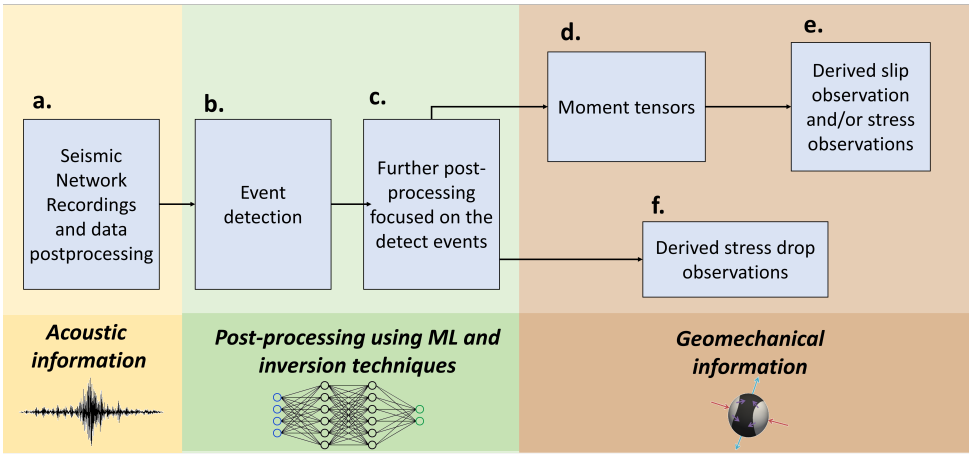


Figure 6.12: Proposed workflow for extracting stress and slip observations for data assimilation

### 6.3.2. CASE STUDY: INDUCED SEISMICITY IN THE GRONINGEN GAS FIELD

The Groningen gas field, is extensively instrumented, making it an ideal subject for data assimilation. This process integrates field measurements with numerical models to enhance understanding of the field, refine estimates of past seismic activities, and forecast future developments. However, challenges arise due to the field's complex and nonlinear physics, the computational demands of simulation models, and the difficulty in obtaining stress measurements that meaningfully inform about fault slip variables.

Discovered in the 1960s, the Groningen gas field has been one of the largest gas fields in Europe (Lele et al. 2016). The Royal Netherlands Meteorological Institute (KNMI) reports that since 1991, earthquakes have been linked to gas production in the Groningen gas field (Dost et al. 2017). Consequently, a monitoring system for induced seismicity became operational in 1995. This network was designed to detect and locate earthquakes with magnitudes of 1.5 and larger (Dost et al. 2017). Despite typically low magnitudes ($M_L < 2.5$), residents have reported structural damage to their houses (Lele et

al. 2016). The strongest recorded event, with a local magnitude of 3.6, occurred near Huizinge in 2012, causing significant damage (Dost et al. 2017; Jagt et al. 2017). Following this event, the monitoring network was expanded to improve earthquake location accuracy and lower the detection threshold (Dost et al. 2017). Additionally, gas production was reduced in the field's central area to decrease induced seismicity, leading to an immediate reduction in seismic events in the following year 6.13.

Multiple geomechanical (Niemeijer et al. 2012; Spiers et al. 2017; van Wees et al. 2014; Verberne et al. 2015), seismological (Chen et al. 2015; van Eck et al. 2006; van Eijs et al. 2006; Verberne et al. 2014) and probabilistic studies (Bourne et al. 2015; Bourne et al. 2014; Dost et al. 2004; Nepveu et al. 2016) have been conducted to better understand the different mechanisms producing the induced seismicity observed at Groningen. Three key factors have been identified that show a good correlation with the occurrence of these earthquakes: pressure drop, fault density of the reservoir, and stiffness ratio between seal and reservoir rock (Jagt et al. 2017). Geomechanical studies show that most of the seismic events are localized on pre-existing fault structures. Therefore, seismicity in the Groningen gas field is thought to be caused by the reactivation of faults. The reservoir's depletion produces changes in the stress field in the subsurface due to the pressure drop. An increase in total strain due to the producing gas layer's differential compaction translates into deformation that the faults accommodate as slip, and produces the seismicity observed (Thienen-Visser and Breunese 2014; Wassing et al. 2016). This means that production changes will influence observed seismicity, and the earthquake generation process is non-stationary (Bourne et al. 2014).



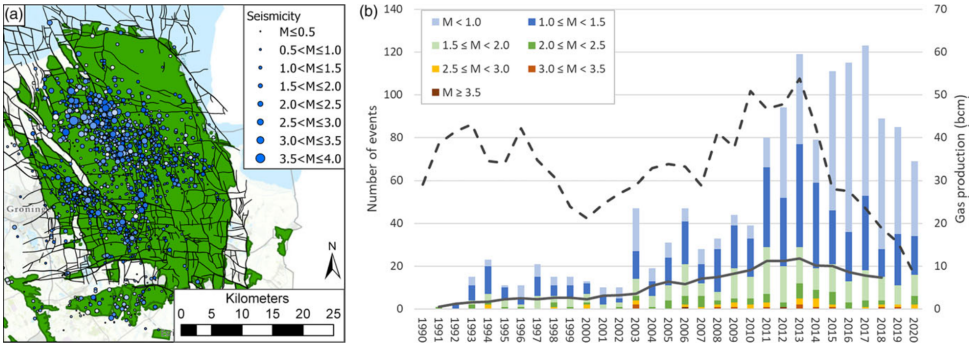


Figure 6.13: Overview of the seismicity in Groningen. (a) Map of seismic events in the Groningen gas field, with color-coding indicating the timing of events (light blue for initial, dark blue for later events) and sizes corresponding to the local magnitude ($M_L$). (b) Temporal pattern of seismicity, featuring a dark solid line for the 5-year moving average of all events with $M_L \geq 1.5$, alongside a dashed line representing annual gas production in billion cubic meters (bcm), and color bars showing the annual number of earthquakes across various magnitude classes. Adopted from Muntendam-Bos et al. (2022)

.

Estimating the recurrence of seismic events in Groningen is more complex than for natural tectonic earthquake activity because it cannot be treated as stationary, a standard assumption in probabilistic seismic hazard analysis (PSHA). The probabilistic models developed need to accommodate earthquakes over smaller magnitude ranges than

those considered by conventional PSHA for natural seismicity (Dahm et al. 2015; Zbinden et al. 2016). While addressing this difficulty through proxies, some time-dependent models need to integrate critical insights from geomechanical models. These insights include realistic friction laws (Wassing et al. 2014), time-dependent compaction, and suitable constitutive relations, all crucial for a deeper understanding of the physical processes such as stress changes and fault interactions driving seismicity. Previous models using compaction strains as a proxy for seismicity recurrence produced estimates which were inconsistent with the reduction in events experienced after the production cut (Bourne et al. 2014). Consequently, new studies employing models considering more complex rheology and time-dependent compaction have been consistent with the actual seismicity post closure in the gas field (van Wees et al. 2017). This indicates a need for further development of appropriate numerical models, experiments, and probabilistic assessments to understand induced seismicity in the Groningen gas field.

#### EVENT DETECTION USING TEMPLATE MATCHING AND NEURAL NETWORKS

As depicted in Fig. 6.12b, event detection is the initial step after processing seismic recordings from various stations. Traditionally, seismic event detection and moment tensor estimation have heavily depended on template matching and inversion methods. Template matching in seismology requires a template event, comprising two critical elements: waveforms and the moveout of arrival times across a network. The template event defines a specific spatial source and mechanism. The process involves comparing the template event to continuous data at every sample, with the degree of similarity measured using a correlation coefficient. This method effectively extracts events from background noise, as each set of events detected by a given template represents a family of events originating from a single source.

These techniques have provided accurate estimates, aiding in understanding seismic mechanisms in areas like Groningen. However, they are computationally demanding and are less effective in environments with very low signal-to-noise ratios (SNRs), particularly when detecting low-magnitude events common in Groningen's earthquake recordings. Machine learning presents a less resource-intensive alternative, capitalizing on the large amounts of data seismic networks collect. Detecting small, often unnoticed earthquakes is crucial, as they provide valuable insights into regional seismic activity. These low-magnitude events, which are more frequent than larger ones, serve as important indicators of seismic behavior. By detecting these smaller earthquakes, we can monitor seismicity more precisely and understand the impacts of production level changes on seismic activities earlier. Developing methods to identify these low-magnitude events is therefore of significant importance.

In machine learning for seismic event detection, various types of neural networks have demonstrated significant potential. Phase diagrams, for instance, have been effectively utilized for event detection (Ross et al. 2018). Perol et al. (2018) specifically highlighted the capabilities of convolutional neural networks in this field, showcasing their effectiveness in identifying seismic events. Additionally, research conducted in Groningen, such as the work by Mousavi et al. (2019), underscores the critical role played by the signal-to-noise ratio (SNR) in detecting these events. Deventer (2021), in her master thesis, delved deeper into this aspect by focusing on Groningen. She investigated the proficiency of neural networks in maintaining high performance levels even with data

characterized by very low SNR. Her evaluation centered on neural networks trained on extensive volumes of synthetic data, offering insights into the adaptability and effectiveness of these advanced machine learning tools in the specific context of seismic event detection.

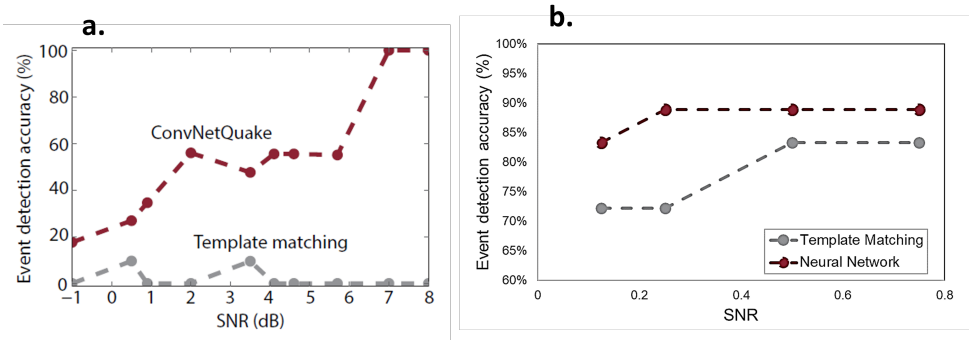


Figure 6.14: Event detection accuracy of neural networks compared to template matching for different signal-to-noise ratios (SNR). (a) Results of a convolutional neural network and template matching for detecting earthquakes product of the induced seismicity in Oklahoma,USA, adopted from Perol et al. (2018). (b) Results of a convolutional neural network and template matching for detecting earthquakes product of the induced seismicity in Groningen,the Netherlands, adopted from Deventer (2021)



Fig.6.14 compares the accuracy between neural networks and template matching across various signal-to-noise ratios (SNR) in two distinct cases of induced seismicity. Fig.6.14a illustrates the performance of a convolutional neural network and template matching in detecting earthquakes resulting from induced seismicity in Oklahoma, USA, as adapted from (Perol et al. 2018). On the other hand, Fig.6.14b displays similar results for Groningen, the Netherlands. Despite the differing mechanisms driving the induced seismicity in these two regions, a consistent observation emerges: convolutional neural networks generally achieve higher accuracy than template matching. This advantage persists even in scenarios of lower SNR, where the accuracy gap narrows, yet neural networks still outperform template matching. Deventer (2021) observed that while neural networks demonstrated superiority in accuracy, they encountered significant challenges with false-positive rates. In contrast, template matching yielded a lower rate of false positives than neural networks. Consequently, Deventer (2021) recommends combining both template matching and convolutional neural networks to achieve more robust and reliable seismic event detection.

### MOMENT TENSOR ESTIMATION OF EARTHQUAKE MECHANISMS

Following detecting seismic events, as illustrated in Fig. 6.12d, it's possible to estimate the moment tensor of an earthquake. The moment tensor is a mathematical representation of fault movement during an earthquake, related to other source characteristics such as stress drop (Madariaga 1979), magnitude, and stress (Abdel-aal and Yagi 2017; Matsumoto et al. 2015) and strain fields. These estimates are valuable as observations in a data assimilation scheme, potentially enhancing earthquake estimation and forecasting accuracy.

A common approach to understanding fault dynamics is through full waveform inversion of seismic data, which yields the moment tensor describing the earthquake mechanism. However, this method is notably time and computationally-intensive. Verwijs (2021) explored in her master thesis to estimate individual moment tensor components using two types of neural networks. The neural networks trained include fully connected neural networks (FFNNs), where each neuron in one layer connects to all neurons in the next, facilitating complex pattern recognition and data processing, and mixture density networks (MDNs), which output mixtures of probability densities, useful for modeling complex distributions and multiple possible outputs for a single input.

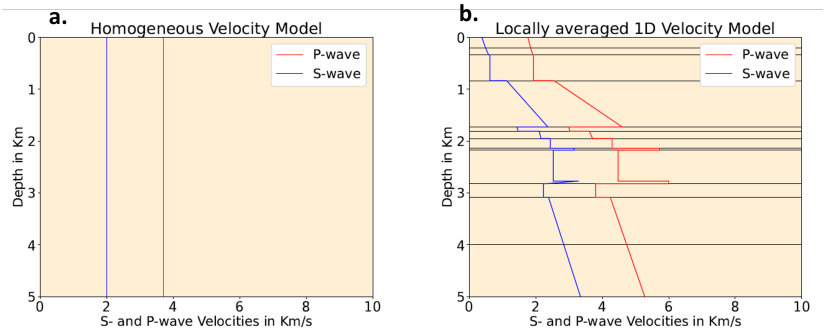


Figure 6.15: Velocity models used for generating the data for training the neural networks based on Kühn et al. (2020). (a) homogeneous model using a single averaged values of p- and s- wave velocities of the Groningen gas field, and (b) a 1D heterogeneous model with averaged velocity model considering the geology of the Groningen gas field. Modified from Verwijs (2021).

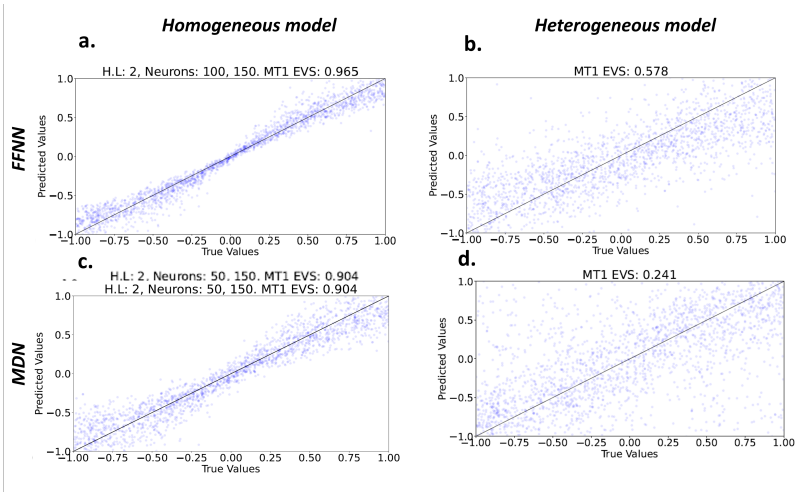


Figure 6.16: Comparison of the estimates of the FFNNs(a,b) and MDNs(c,d) for a moment tensor component for the homogeneous(a,c) and heterogeneous model(b,d). Modified from Verwijs (2021).

A significant challenge in training neural networks for moment tensor inversion is the often insufficient amount of preprocessed recorded data available for training. To address this, Verwijs (2021) utilized synthetic data, performing forward modeling of seismic recordings with assumed velocity models based on Kühn et al. (2020), as shown in Fig.6.15. This modeling used the Pyrocko software package (Heimann et al. 2017), employing two velocity models: a homogeneous profile with uniform p-wave and s-wave velocities and a heterogeneous model representing the Groningen gas field's geology. Fig.6.16 compares the predicted and true values for the first component of the moment tensor's diagonal (MT1) for both models using FFNNs and MDNs. The closer the estimates are to the diagonal, the better the networks' performance. For the homogeneous model, both network types show good estimates, while in the heterogeneous model, the complexity of the geometry appears to complicate the estimates.

### 6.3.3. DISCUSSION OF ASSIMILATION OF FIELD SCALE DATA

In field-scale simulations, capturing the complexities of geological systems effectively necessitates a multiphysics approach, encompassing aspects such as fluid flow, mechanical behavior, and frictional laws. These simulations face particular challenges due to their high-dimensional nature and the heterogeneity of geological models. Difficulties in convergence, stemming from applying frictional laws and integrating various physical phenomena, call for innovative methods to manage stress fields and ensure stability. The significance of fault interactions and their geometrical configurations adds another layer of complexity, underscoring the need for advanced methods to handle heterogeneity and anisotropy in flow and mechanics equations. In this context, developing and applying sophisticated data assimilation methods become essential to manage high-dimensional problems and nonlinear dynamics, particularly vital for maintaining accuracy and stability when models are updated or refined.

The Groningen gas field, with its extensive instrumentation, including a comprehensive seismic network and pressure measurements from wells, allows for diverse applications of the seismic data. Beyond its use in modeling stress and slip-rate evolution on faults, a key focus is seismicity rate. Probabilistic Seismic Hazard Assessments in Groningen, incorporating the Epidemic Type Aftershock Sequence (ETAS) model and a Poisson distribution for earthquake occurrences, adapt this focus within the context of the mechanical behavior of faults governed by rate-and-state friction (Baki and Lieshout 2022; Lieshout and Baki 2023; Osinga and Kraaijpoel 2020). These assessments employ the formulation of Dieterich (1994) for seismicity rate estimation. However, this approach does not account for individual earthquake magnitudes or event clustering (Dieterich 1994; Lieshout and Baki 2023), leading to a growing interest in alternative models for seismicity rate.

Data assimilation techniques using ensemble smoothers have also been explored in Groningen for seismicity rate estimation, considering the region's physical processes (Candela et al. 2022). This involves modifying the rate-and-state friction formulation and incorporating the MACRIS method (Candela et al. 2019; Wees et al. 2018, 2019), thus framing it as a physics-based statistical model. While MACRIS offers fast computations and accounts for key aspects like heterogeneous pressure distributions and fault offsets, it does not fully address factors like fault pressure evolution or the spatial heterogeneity

of lithologies' elastic properties (Candela et al. 2022). Addressing the heterogeneity and uncertainty in the frictional parameter $a$ is vital for explaining seismicity rates in such complex geological settings.

## 6.4. CONCLUSIONS

In this chapter, we delved into data assimilation's role in laboratory experiments for seismic research. The methodology proposed bridged theoretical models with practical applications, demonstrating how insights from perfect model experiments could be adapted to large-scale laboratory settings. This approach helps estimating variables like shear stress distribution at the contact surfaces between frictional bodies, a key element in understanding seismic activities. The chapter further elaborated on the specific measurements and state variables that were most effective in these laboratory experiments. By offering a sensitivity analysis of how laboratory measurements influenced the estimation of fault slip, and exploring the role of sensor placement during various phases of the seismic cycle, the chapter provided a conceptual guide for applying data assimilation in controlled experimental environments.

The second part of the chapter shifted focus to the possible application of data assimilation at a field scale, particularly in the context of earthquake forecasting. Here, the conceptualization of using data assimilation techniques for estimating earthquake occurrences in real-world scenarios was emphasized. The chapter discussed the scarcity of stress measurements in faults. We proposed a novel workflow integrating machine learning techniques and seismological knowledge to derive indirect stress observations from seismic recordings. This approach was exemplified through the Groningen gas field case study, highlighting some of the practical challenges, results, and lessons learned. The emphasis on the importance of various measurements for estimating fault-slip occurrence underlined the potential of data assimilation in enhancing our understanding and forecasting capabilities in field-scale seismic research.



### DATA AVAILABILITY

The forward model for (a)seismic slip sequences utilizing the numerical modelling package Garnet is made accessible via repository https://bitbucket.org/cpranger/garnet/src/master/. The data produced and analyzed in this chapter is available via 4TU.ResearchData (http://doi.org/10.4121/fda02e8f-00bc-4c5b-a92e-3bae47f00e97).

# 7

# CONCLUSIONS AND RECOMMENDATIONS

## 7.1. CONCLUSIONS

The main aim of this thesis was to investigate how ensemble data assimilation methods can enhance the estimation of variables that control fault slip, aiming to improve the estimation of earthquake occurrences when using numerical models governed by the rate-and-state friction (RSF) law. In the beginning of the thesis, I present the challenges imposed by the numerical modeling of earthquake sequences. The selected models were then used for data assimilation in different configurations to assess the impact of the observations in the estimation of variables like stress and velocities and their potential use for earthquake forecasting, drawing the conclusions mentioned below.

In **Chapter** 2, I outlined the numerical models employed for forward modeling the earthquake sequences in the data assimilation experiments of this thesis. We explained the seismic cycle and how the rate-and-state friction law is used for understanding and simulating it. We delved into the different types of seismological models that could be used for performing data assimilation when considering a horizontal strike-slip fault with a surrounding elastic homogeneous medium. We highlighted how Burridge-Knopoff models are helpful to understand the non-linear dynamics of the rate-and-state friction and how the non-dimensional parameters derived during the formulation of the non-dimensional ODEs are helpful to understand phenomena like aperiodicity. We also discussed and mentioned some limitations of the finite difference methods and the challenges posed by the stiffness of the equations we are solving. We also set some suggestions for the data assimilation in subsequent chapters and studies. Furthermore, we performed a sensitivity analysis of the Burridge-Knopoff model to analyze the non-linear dynamics and their effect on the regime of the model, the recurrence interval, and the stability of the simulation. This analysis highlighted the challenges and difficulties when using models governed by the rate-state restriction for data assimilation applications. It illustrated how essential frictional parameters are in the estimation of the system.

In **Chapter** 3, we introduced the fundamentals of data assimilation, including the specific notation used for data assimilation within the scope of this thesis. We also introduced the Ensemble Kalman Filter (EnKF),as well as the Adaptive Gaussian Mixture Filter (AGMF) and the Particle Flow Filter (PFF) as alternative methods with the advantage of handling non-Gaussian priors. Also, we discussed the challenges posed by low-rank approximations and spurious correlations in methods like the EnKF and how techniques like localization and inflation can help cope with these issues and still obtain good estimations. Finally, we highlighted the difference between the methods using perfect experiments with a Lorenz 96 model of 20 cells. We used the three methods to estimate the system's variables under periodic and chaotic conditions. We compared the errors of the methods using the RMSE and saw that the PFF provides the best estimates when having fewer observations and chaotic conditions. These conditions of low number of observations and aperiodicity are very common in seismological applications, which showcases the potential of using methods such as the PFF on them. Additionally, we visualized how the non-linear update of the PFF works and explored cases with uncertainty on the parameters and forcing of the system. Finally, we used these findings to compare these results with the Lorenz 96 model, with the ones obtained in perfect model experiments with a 1-D Burridge Knopoff model governed by rate-and-state friction (RSF) in Chapter 5.

In **Chapter** 4, we explored the EnKF applied to a 1D model that simulated a 0D fault point in a 1D elastic medium. The EnKF assimilated synthetic, noisy observations of shear stress and velocity. Utilizing a 50-member ensemble with our physics-based model, we estimated the shear stress, slip rate, and the state variable $\theta$ at the fault. Besides, we assessed the EnKF's ability to forecast slow slip events and earthquakes. Our findings indicated the EnKF's effectiveness in quantifying stress, slip rates, and fault strength uncertainties. It performed best during the interseismic phase, with absolute errors in shear stress estimates being around 3 to 5% of the stress drop, though standard deviations were slightly higher (4 to 7%). However, estimation errors spiked during and around the coseismic phase of earthquakes, reaching 20 to 25% of the stress drop, attributed to rapid changes in stress and velocity. The EnKF accurately estimated earthquakes that lasted seconds over decadal observation periods. The assimilation of shear stress greatly improved stress estimation for slow slips and earthquakes, while velocity observations mainly influenced slip rate estimates for slow slips. For earthquakes, both types of observations were crucial. The state variable $\theta$ estimation depended equally on both observations for slow slips, but shear stress was more influential for earthquakes. Observations taken post-slip were particularly vital, especially for estimating earthquakes. Further analysis showed the EnKF's low forecasting failure rate ( 10%) for both slow slips and earthquakes, predicting slow slips half a year and earthquakes about two years in advance, considering a 20-year occurrence interval. These results demonstrated data assimilation's potential in enhancing fault-slip occurrence estimates.

In **Chapter** 5, we evaluated and compared the estimates of the EnKF, AGMF, and PFFs for estimating earthquake occurrences in a 1-D Burridge Knopoff model governed by RSF in periodic and chaotic settings. We explored how non-regularized versions of the RSF allow aperiodic solutions by changing the frictional parameters as explained in

**Chapter** 2. These experiments allow us to understand a larger variety of regimes than the mostly quasi-periodic formulation of the RSF used to guarantee convergence of the models and further stability problems during the simulation. The results illustrated that the three methods have good estimates of the shear stress($\bar{\tau}$) and velocity($\bar{V}$), which are lower than the observation error and also of the state ($\bar{\Theta}$) which is a hidden state. In general, the EnKF has the lowest RMSE values. However, when evaluating its ensemble spread using rank histograms, we observed problems of underdispersion that hint at a problem of ensemble degeneracy. This underdispersion and low ensemble spread could not be fixed even with inflation factors as high as 2 to 5. The AGMF dealt better with the underdispersion using a resampling step after calculating the weights of the Gaussian mixture. However, this solution was not enough to solve the collapse of the ensemble spread in the subsequent assimilation windows. Conversely, the PFF showed the highest ensemble resilience against underdispersion, maintaining a healthy ensemble spread. In this chapter, we also explored introducing model errors in the estimation of the filters to deal with the problems of parameter bias in some of the frictional parameters. Using the model error has the advantage that it allows transit from periodic solutions to aperiodic ones without disturbing the system's stiffness and the simulations' overall stability. These results illustrate the potential of performing data assimilation with model errors using numerical models governed with regularized formulations of RSF that can still produce aperiodic and chaotic solutions.

In **Chapter** 6, we presented an overview of the laboratory experiments used for analyzing fault-slip in faults. We mentioned the most common laboratory measurements acquired in these types of experiments and what information they can provide for the data assimilation of stress and velocities. Then, we introduced a conceptual workflow for assimilating these measurements in a data assimilation framework and discussed how machine learning techniques could enhance this assimilation. Later, we made a more detailed analysis of the impact of these lab measurements in the estimation of fault slip in a large-scale laboratory experiment. We use the degree of freedom of signal (DFS) to evaluate the information content in these measurements. We study how this information content changes per variable depending on the current phase of the seismic cycle and the spatial coverage and location of the sensors. The results show that in the interseismic phase, the location of the sensors does not have a large impact on the DFS. Briefly, before and after the coseismic phase, the information content is maximum the closer the observations are to the fault. However, the change in the information content is visible even for the measurements taking the furthest away. These findings highlight the forecastability potential evaluated in **Chapter** 4 and the possibility to ring alarms briefly before the main events.

Furthermore, in **Chapter** 6, we showed the measurements available for data assimilation of fault-slip at the field scale. In **Chapters** 4, 5, and 6, we highlighted that stress measurements are very important for the estimation of the state variables at the fault. However, stress measurements are scarce, with valuable information about the stress in the faults. We proposed a conceptual workflow to derive indirect stress observations from seismic recording using machine learning techniques and our seismological knowledge to derive information about the stress drop that could be assimilated. Finally, we shared some experiences of the implementation of various steps of this workflow in the case

of the Groningen gas field. We mentioned the challenges, primary results, and lessons learned. This chapter then highlighted the potential of ensemble data assimilation and the importance of various measurements for estimating fault-slip occurrence.

Thus, this thesis

- Showed that ensemble data assimilation can be done on models of earthquakes with fast slip rates and still get good estimates in low-dimensional models, showing the potential to scale it to more complex models and move forward the advancements that have been done on slow-slip events models to estimates the evolution of the fault-slip. However, attention is required since in low-dimensional models, non-Gaussian priors start appearing, which need care on the data assimilation method used

- Illustrated how the choice of the numerical model and rate-and-state friction law can impose quasi-periodic behavior, which can translate into underdispersion issues that undermine the validity of the estimates of methods that assume Gaussianity and linearity. We showed how methods like the Particle Flow Filter (PFF) can still have healthy spreads for applications with laboratory or field data where it is difficult to assess how good the estimates are with respect to the truth

- We explained how there can be a sudden change in the system's behavior when frictional parameters or boundary conditions are uncertain when using rate-and-state friction law. We mention how this has convergence problems and can end in ensemble degeneracy or make data assimilation difficult when including the parameters in the state vector and dealing with the highly dimensional system. We showed how an alternative is to use model error terms in the data assimilation, which gives freedom to the set dynamics imposed by having a stable solution for a particular set of parameters. Still, we predominantly use periodic solutions to estimate aperiodic ones. This model error can be seen as a source of additional stochasticity needed to represent earthquake behavior better.

- Showed the importance of having stress observations for a better estimate of the shear stress at faults and explained the difficulties in the field. We presented alternatives to potentially derive indirect shear stress observations that can be included in the data assimilation for estimating the fault slip and possibly future earthquake occurrences.

- Showed how the EnKF is a suitable data assimilation method for the most common behaviors in laboratory experiments, like stick-slip behavior or slow-slip events. For the most common frictional parameters found in nature, despite the low-rank approximations, good estimates can be obtained in systems with homogeneous elastic media since a small ensemble can still derive good results, which is very useful for heavy computational models with discontinuous behaviors like faults, and stiff equations like the imposed by rate-and-state friction.

## 7.2. RECOMMENDATIONS

### 7.2.1. COMPUTATIONAL COSTS NUMERICAL MODELS USED

Finite difference methods are highly computationally demanding. We utilize Garnet for our simulations as it has been benchmarked with the SEAS challenges (Jiang et al. 2022). Additionally, as part of the project, the models in Garnet were intended to be expanded to include fluid dynamics. However, an alternative approach is beneficial for problems like those explored in this thesis, which are dominated by RSF in the fault and a surrounding homogeneous medium. One such alternative is the use of boundary element methods (Luo et al. 2017). These methods are faster, enabling larger ensemble simulations. However, the assumptions inherent in boundary element methods (BEMs) make it challenging to implement heterogeneities or changes in boundary conditions, such as those needed to modify the equivalent stiffness of the setup and obtain the results shown in Fig. 6.6 in **Chapter** 6 and in (Li et al. 2023a). BEMs efficiently solve quasi-dynamic or fully-dynamic earthquake ruptures but commonly rely on periodic boundary conditions and free surface approximations (Jiang et al. 2022). Finite element methods can handle different geometries and heterogeneity, but the computational costs are too high for large ensembles.

### 7.2.2. COVARIANCE MATRICES IN THE ENKF APPLIED TO SYSTEMS DOMINATED BY RSF

In **Appendix** A, we describe the localization process for the covariance matrices of the 1D-Burridge Knopoff model with rate-and-state friction. These matrices often have a high condition number, which makes them ill-conditioned. This ill-conditioning may arise from the rapid changes in the variables and their effects on the values of the different components of the matrix. As an alternative, localization based on correlations could be employed to produce better-behaved covariance matrices since inverting ill-conditioned matrices often requires alternative approaches, such as during the calculation of the gradient of log of the prior for the PFF (Eq. 3.28). We bypassed this problem by leveraging the fact that the workflow requires multiplication by the covariance matrix in the final step, thus avoiding the need for inversion. However, it is advisable to avoid inverting covariance matrices in systems dominated by rate-and-state friction, especially in regimes where slip rates are very high and the transitions of the state variables are rapid, as discussed in **Chapter** 4.

### 7.2.3. FLUID-INJECTION DOMINATED SEISMICITY AND MULTIPHYSICS

As discussed in **Chapter** 6, especially in cases of induced seismicity resulting from the depletion or injection of fluids, the dynamics of fault-slip can be dominated by phenomena such as changes in pore pressure, which are not currently modeled in our setup. The dynamics in our setup are primarily governed by the mechanical behavior of the deformation of the elastic medium and the sudden changes in slip behavior as explained by the frictional law. Data assimilation into models of systems influenced by fluid injection or depletion requires additional pore pressure measurements and the coupling of the fluid flow, a frictional law, and mechanical deformation. This integration presents a significant challenge, adding more non-linear behavior to the complexities of rate-and-

state friction. These challenges, for example, may generate convergence problems and higher computational costs associated with the required time-stepping schemes needed to resolve the coseismic phase accurately.

Usually, numerical models that fully couple fluid dynamics in porous media with mechanical deformation assume a slip-weakening law. Therefore, more research is needed into implementing data assimilation methods in applications that require the coupling of fluid flow, mechanical deformation, and rate-and-state friction. This complexity is even higher when adding the thermodynamics required for applications like induced seismicity in geothermal energy production or including plastic behavior in the mechanics.

Additionally, incorporating the variables associated with fluid phenomena introduces more state variables into the estimation problems, leading to high-dimensional systems. Achieving satisfactory results with these systems is more challenging, especially for particle approximation methods, which become more inefficient with increasing dimensions due to problems like weight degeneracy and requiring a large number of particles to sample the more complex distributions appropriately.

### 7.2.4. HETEROGENEOUS FRICTIONAL PARAMETERS AT THE FAULT

In **Chapters** 5 and 6, we explored the benefits of data assimilation for estimating the stress on the fault. We assumed a heterogeneous or discontinuous normal stress profile, affecting the shear stress field. Under propitious conditions, these discontinuities can facilitate the nucleation of earthquakes (Li et al. 2023a). However, as explained in **Chapter** 2, the contrast in the parameters $a$ and $b$ can create zones that are either velocity-strengthening or velocity-weakening (Li et al. 2022). When modeling asperities in velocity-weakening zones surrounded by velocity-strengthening areas, the nucleation of earthquakes is typically constrained to the borders between these zones, where stress differences are larger than elsewhere along the fault. Therefore, the nucleation process is predominantly governed by the contrast in rock properties, as reflected by the frictional parameters, rather than by the heterogeneity of the stress state. Another factor that plays an important role in induced seismicity is the fault geometry. Especially displaced faults result in strongly peaked Coulomb stress patterns (Buijze et al. 2019; Buijze et al. 2017; Jansen and Meulenbroek 2022; Jansen et al. 2019). Further experiments, which also estimate the spatial distribution of the frictional parameters and consider different fault geometries, are necessary to improve the representation of the slip process with data assimilation. This is challenging due to the complexities introduced by varying frictional parameters in maintaining the stability of the simulation, especially in critical factors like nucleation length or system stiffness.

# A

# Analysis of the Background Covariances, Localization, and Inflation

## A.1. Localization of covariance matrices

Localization techniques were introduced in Chapter 3. A common model in which localization is necessary for obtainineg accurate estimates with the Ensemble Kalman Filter (ENKF) is the Lorenz 96 model, tested in Chapter 3 and compared with the Burridge-Knopoff 1D model coupled with rate-and-state friction in Chapter 5. In the case of the Lorenz 96 model, localization can be easily achieved by using the Schur product of the background covariance matrix and imposing a given correlation length. However, for the Burridge-Knopoff model, this becomes a more complex task as applying a Schur product can have an undesired effect on the cross-covariance elements that connect the unobservable and observable variables. For this reason, we explore whether localization is necessary.

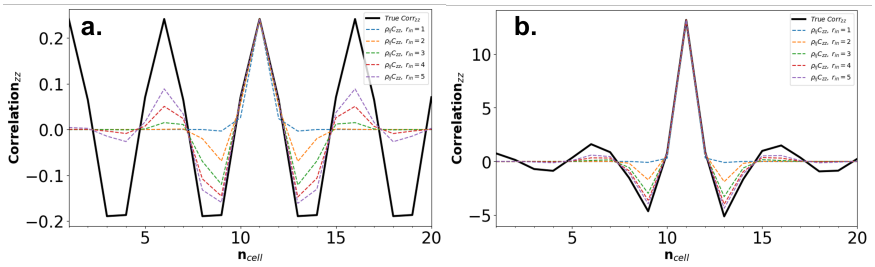


Figure A.1: Correlation coefficient analysis for the Lorenz 96 model using different localization lengths. (a) depicts coefficients for periodic solutions, and (b) depicts chaotic solutions. Each graph shows the coefficient before localization (solid black) against the correlation coefficients when applying five different localized lengths (colored lines), with a localization length of approximately 3 providing the best localization.

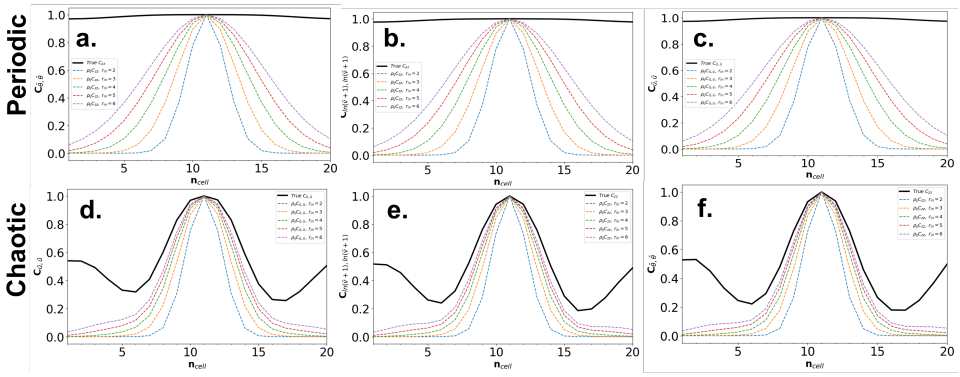


Figure A.2: Correlation coefficient analysis for the 1-D Burridge-Knopoff Model using different localization lengths. The top row (a-c) depicts coefficients for periodic solutions, and the bottom row (d-f) depicts chaotic solutions. Each graph shows the coefficient before localization (solid black) against the correlation coefficients when applying five different localized lengths (colored lines), with a localization length of approximately 5 providing the best localization.

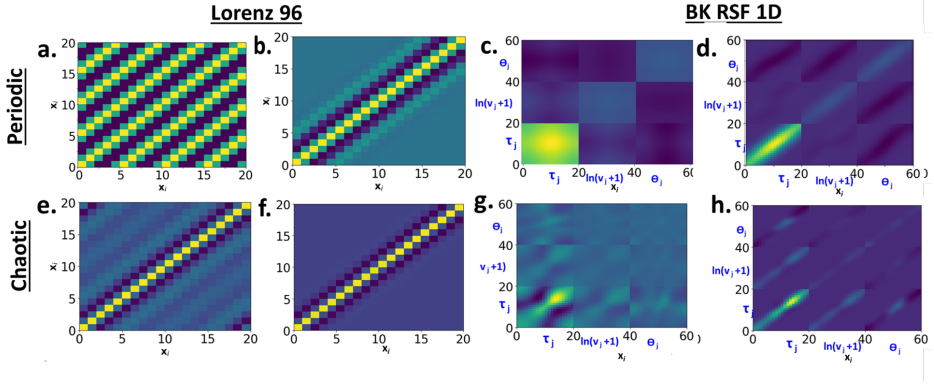


Figure A.3: Comparison of the background covariance matrices ($\mathbf{C}^{\mathbf{f}}_{\mathbf{zz}}$) of the Lorenz 96 model and the 1-D Burridge-Knopoff model before (a,c,e,g) and after (b,d,f,h) localization via a Schur product with a localization distance of 3 and 5 cells, respectively. The upper row (a-d) compares the background covariance before and after localization for periodic solutions. The lower row (e-h) compares before and after localization for chaotic solutions.

We use the singular value decomposition (SVD) of the prior covariance matrices for the Lorenz 96 and BK models to assess the effect of the localization in increasing their effective rank (Fig. A.4). We assess the effective rank as the number of components that explained 95 % of the variance of the matrices. In general, we observe that the localization helps increase the effective rank of the covariance matrices. However, it is most effective in the Lorenz 96 model compared to the Burridge Knopoff model. For the Lorenz 96 models, the effective rank before localization is about 2 in the periodic case and around 18 in the chaotic case. After the localization, the effective rank of the covariance matrices

for the periodic case increases to 14, while for the chaotic case stays around 18. For the 1-D Burridge-Knopoff models coupled with rate-and-state friction, we see that the effective rank for the periodic case is about three before the localization, which coincides with the rank expected for a single block model, while it is around 6 for the chaotic case. After the localization, the effective rank for the periodic case increases to 8, while for the chaotic case, it increases to 15.

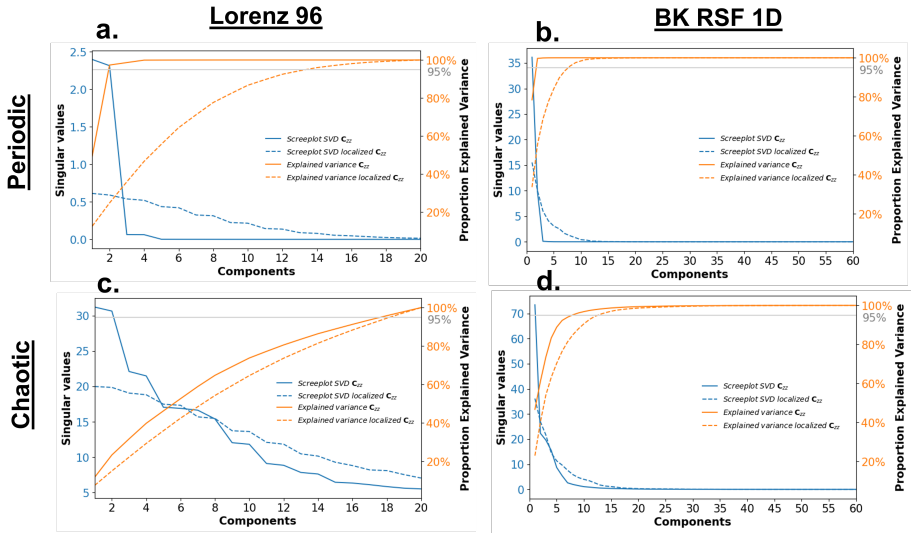


Figure A.4: Scree plot of the singular value decomposition of the prior covariance covariance matrices ($\mathbf{C}^{\mathbf{f}}_{\mathbf{ZZ}}$) of the 1-D Burridge-Knopoff before and after localization for the Lorenz 96 (a,c) and Burrdige-Knopoff model (b,d). The solid lines correspond to the decomposition of the matrices before localization, while the dashed lines correspond to the decomposition of the matrices after the localization. The blue lines represent the distribution of singular values, while the orange lines show the proportion of cumulative variance explained until that component.

## A.2. INFLATION OF COVARIANCE MATRICES

Covariance inflation techniques were introduced in Chapter 3. Here, we discuss how we assessed the need for the inflation of the background covariance matrices for the perfect model experiments in Chapter 5, where we employed seven distinct ensemble sizes (10, 20, 50, 100, 200, 500). Our approach focused on examining the variance alteration across different variables, which served as an indicator of the potential requirement for an inflation factor. This methodological consideration stems from the understanding that a smaller ensemble size often leads to underestimated variances. Hence, the implementation of an inflation factor can serve to rectify this shortfall. The comparative analysis of these ensemble sizes and their subsequent variance alterations are visually represented in Fig. A.5. Fig. A.5. shows that the variance is the same for the ensemble size, indicating that there is no need to apply an inflation factor to the covariance matrices of the 1-D Burridge-Knopoff coupled with the rate-and-state friction model. However, when considering the problems of underdispersion shown in the Section 5.3.1 under the pe-

riodic conditions, we see that we need large inflation factors to avoid the filters being underdispersed.

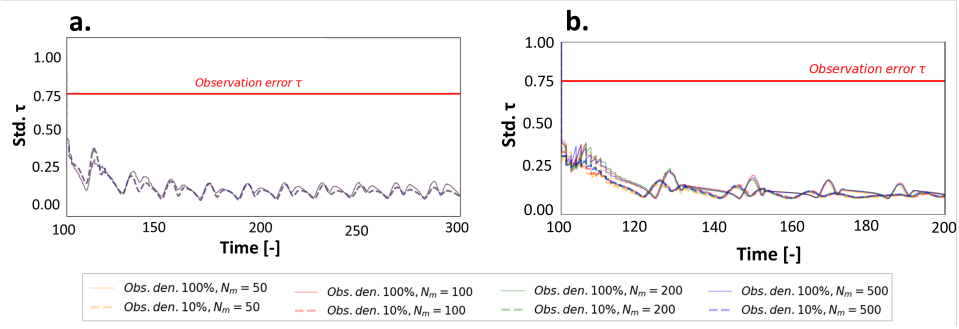


Figure A.5: Evaluation of the inflation factor using various ensemble sizes.(a) compares the standard deviation of the estimates of the shear stress $\bar{\tau}$ when using an observation density of 10% and 100% for the periodic solutions and (a) for the chaotic solutions.

# B

# HYPERPARAMETER SELECTION FOR THE AGMF AND THE PFF

## B.1. HYPERPARAMETER SELECTION FOR THE ADAPTIVE GAUSSIAN MIXTURE FILTER

For the AGMF, we tested different bandwidths for the Gaussian mixtures, denoted as $h$. We used a default value of 0.6. An analysis of the RMSE, STD, and the rank histogram showed a lower error for lower $h$ values (around 0.2) in the periodic case of the Lorenz 96 and values closer to 0.6 in the chaotic case. Such low values of $h$ are inconsistent with the high inflation factors needed to avoid underdispersion. For this reason, we adhered to a value of 0.6. This approach was also applied to the Burridge-Knopoff models.

## B.2. HYPERPARAMETER SELECTION FOR THE PARTICLE FLOW FILTER

The particle flow filter is defined by two crucial hyperparameters: the kernel bandwidth ($\alpha$) and the step size of the pseudo-time ($\Delta s$). Our experimentation focused on identifying optimal values for these parameters in the context of the 1-D Burridge-Knopoff models. For the kernel bandwidth $\alpha$, we discovered that 0.05 was most effective in periodic cases, while 0.005 suited chaotic cases. This finding was intriguing as it diverged from the AGMF's expectations, particularly since narrower bandwidths for chaotic scenarios contradict the anticipated need for higher inflation in periodic cases. Moreover, we observed that wider kernels led to more underdispersed filters in periodic conditions.

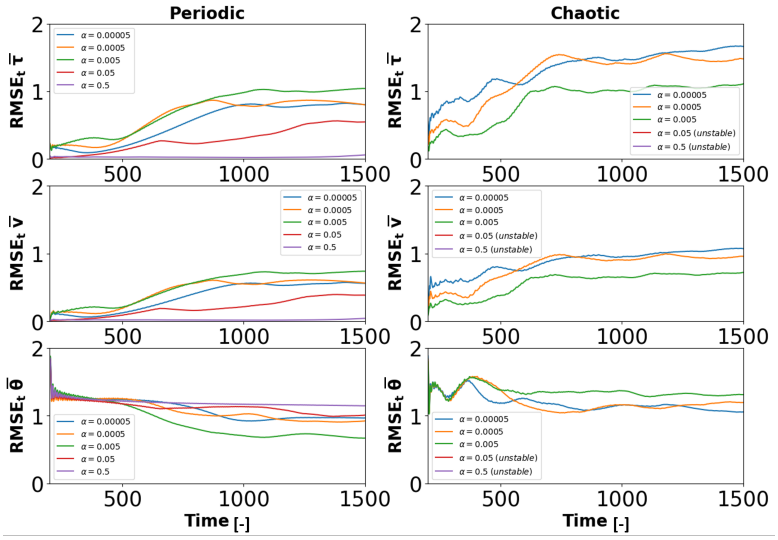


Figure B.1: Sensitivity of the hyperparameter bandwidth of the kernel ($\alpha$) for the Particle Flow Filter used on the BK-RSF 1D model. The left column shows the results for the periodic conditions of the BK-RSF 1D, while the right column shows the results for the chaotic condition. The bandwidths that produced unstable results are indicated in the legend.

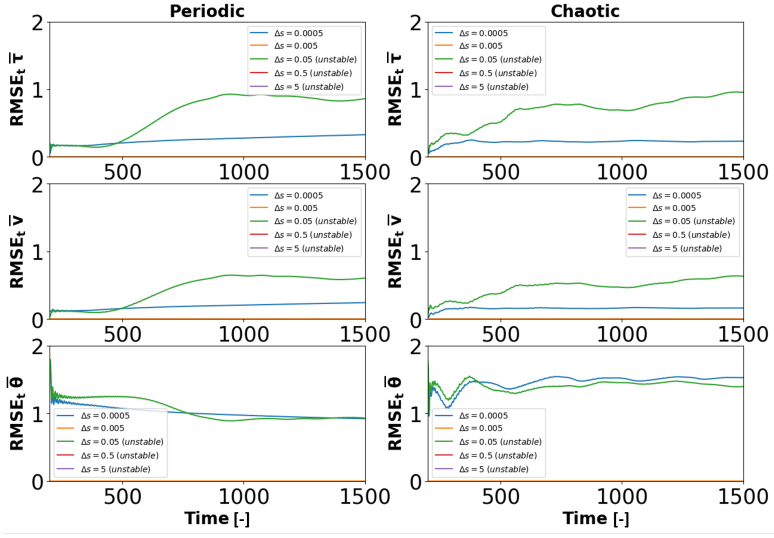


Figure B.2: Sensitivity of the hyperparameter step size of the pseudo-time ($\Delta s$) for the Particle Flow Filter used on the BK-RSF 1D model. The left column shows the results for the periodic conditions of the BK-RSF 1D, while the right column shows the results for the chaotic condition. The step sizes that produced unstable results are indicated in the legend.

Expanding on this analysis, we rigorously tested a range of bandwidths (0.00005, 0.0005, 0.005, 0.05, and 0.5) alongside various pseudo-time step sizes (0.0005, 0.005, 0.05, 0.5, and 5). Our objective was to determine the most stable and error-efficient combination. As detailed in Fig. B.1, the results guided our selection of 0.05 bandwidth for periodic and 0.0005 for chaotic conditions in the BK RSF 1D model. The criteria for this selection are based on achieving the lowest RMSE for the three variables in the periodic case while ensuring that the bandwidth did not reduce to zero, which might signify a filter collapse or an overly precise fit. Additionally, some bandwidth can produce estimates that bring stability problems to the simulations, as explained in Chapter 2. For chaotic conditions, our choice was the narrowest bandwidth that maintained stability.

The pseudo-time step was chosen for the PFF to be 0.0005 for the periodic and chaotic cases based on the results of Fig. B.2. Similarly to the case of the bandwidth of the kernel, the pseudo time size step of 0.0005 offers the lowest RMSE for the shear stress, the velocity, and the state $\theta$ that does not collapse to an RMSE of 0, which may hint to a collapse of the filter and the variance of the members.

The kernel's bandwidth impacts how the prior information is handled, as discussed in Section 5.4.1 and in Stordal et al. (2021).

# C

# Evaluation of the Influence of Observation Frequency and Observation Error in Estimates on a 1-D BK Model

## C.1. Impact of observation frequency

The assimilation step size, or frequency, in data assimilation, impacts its performance. Large assimilation steps reduce computational resources and time requirements but may result in temporal discontinuities and inaccuracies in capturing rapid changes. Small assimilation steps allow for more frequent updates, better capture temporal variations, and improve overall estimates. However, they increase computational costs, posing challenges in real-time applications.

We also sample synthetic observations at different sampling frequencies by defining a given number of observations per cycle. We define different cases as shown in Table C.1. For determining the number of observations per cycle, we used a time step of $\Delta t$ of 0.01-time units and defined the cycle duration using the periodic case. The cycle for the periodic case lasts about 18-time units or 1800 time steps.

Table C.1: Case scenarios for the observation frequency

| Case | Timesteps btw obs. | Time units btw obs. | Num. obs./ cycle |
|---|---|---|---|
| B.1. | 45 | 0.45 | 40 |
| B.2. | 90 | 0.9 | 20 |
| B.3. | 225 | 2.25 | 8 |
| B.4. | 450 | 4.5 | 4 |
| B.5. | 900 | 9 | 2 |

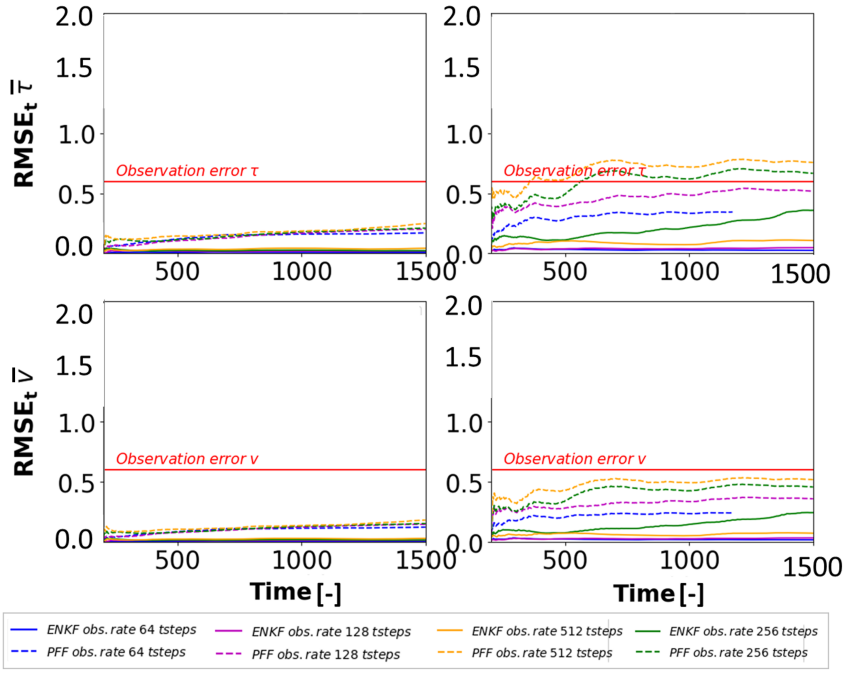


Figure C.1: Comparison of the results of the EnKF (solid lines) and the PFF (dashed lines) for different observation frequencies (observation rate in time). The left column results under a periodic condition, and the right column under chaotic conditions.

Fig. C.1 shows the RMSE for the shear stress($\bar{\tau}$) and the slip rate ($\bar{V}$) for the EnKF and the PFF under periodic and chaotic conditions for different observation frequencies. The results suggest that for the periodic case, all the observation frequencies have similar errors; we only need a few observations per observation cycle to represent the system's behavior well. In contrast, for the chaotic case, we see a larger difference in the errors for different frequencies. As discussed in Chapter 5, the EnKF has problems of underdispersion, which may explain why the changes do not influence the results in observation in Comparison to the PFF. The results show that most observation frequencies are appropriate to capture the behavior of the velocities but that more frequent observations per cycle are needed to estimate the shear stress well.

## C.2. IMPACT OF OBSERVATION ERROR

Observation error refers to the inherent uncertainty or noise associated with the measurements or observations in the data assimilation process. Higher errors introduce uncertainties, potentially leading to larger discrepancies between estimated and true states. In twin experiments, very low errors cause overfitting, as the assimilation method excessively relies on precise observations, adjusting the model state to match them closely. Overfitting limits generalization and may deviate from the true state.

In this thesis, we assumed uncorrelated observation errors, leading us to employ diagonal matrices $C_{dd}$ to express the uncertainty of our observations. We introduced five scenarios to evaluate the potential impact of varying error levels, as depicted in Table C.2. Our error level choice is guided by the uncertainties reported by (Banerjee et al. 2023) for shear stress (0.6) and slip velocity (1.15). Rather than using the slip velocity $\bar{V}$, our experiment's assimilation process utilized the transformed variable $ln\left(\bar{V}+1\right)$. Upon performing a first-order Taylor expansion of this variable, we determined that the estimated uncertainty for typical values in the chaotic BK RSF 1D model was also approximately 0.6.

Furthermore, we assess the ratio of this uncertainty (0.6) with the standard deviation of the slip $\sigma_{\bar{u}}$ in the chaotic case of the 1-D BK-RSF model. This standard deviation is about 2.42, the highest found among the modeled variables. The ratio between but standard deviations yielded a value of 0.25. Based on these findings and to simplify the analysis of our results, we opted to apply a single uncertainty value for the observations of the different variables from the 1-D BK-RSF model. Further, we chose standard deviations that were half and a quarter of the middle uncertainty values for the lower error levels and twice and four times for the higher levels, respectively.

Table C.2: Case scenarios for the observation error

| **Case** | $\sigma_\varepsilon$ | $\sigma_\varepsilon/\sigma_u$ |
|---|---|---|
| C.1. | 0.15 | 0.06 |
| C.2. | 0.30 | 0.12 |
| C.3. | 0.60 | 0.25 |
| C.4. | 1.20 | 0.50 |
| C.5. | 2.40 | 0.99 |

Fig. C.2 shows the RMSE for the shear stress($\bar{\tau}$) and the slip rate ($\bar{V}$) for the EnKF and the PFF under periodic and chaotic conditions for different observation errors. Like the results in Chapter 5, the EnKF has little influence on the observation error,r, most probably due to ensemble degeneracy and underdispersion. The results of the PFF show that all observation errors are adequate to have good estimates in the periodic case and that, as expected, larger values of RMSE are obtained when there is higher uncertainty in the observations. In the chaotic case, the shear stress is the variable that experiences the larger RMSEs in their estimates. The velocity observation errors are still adequate for good estimates with both filters. Therefore, after the results changing the observation density in Chapter 5 and the observation frequencies and errors in this appendix, we can conclude that the shear stress is the variable that requires more density and fre-

quency to be well estimated, and the lower uncertainties to better estimate the state of the system. This highlights the importance of having information about the stress to get good estimates dominated by the rate-and-state friction. The implications of this finding in data assimilation for laboratory experiments and the field scale are further explored in Chapter 6.



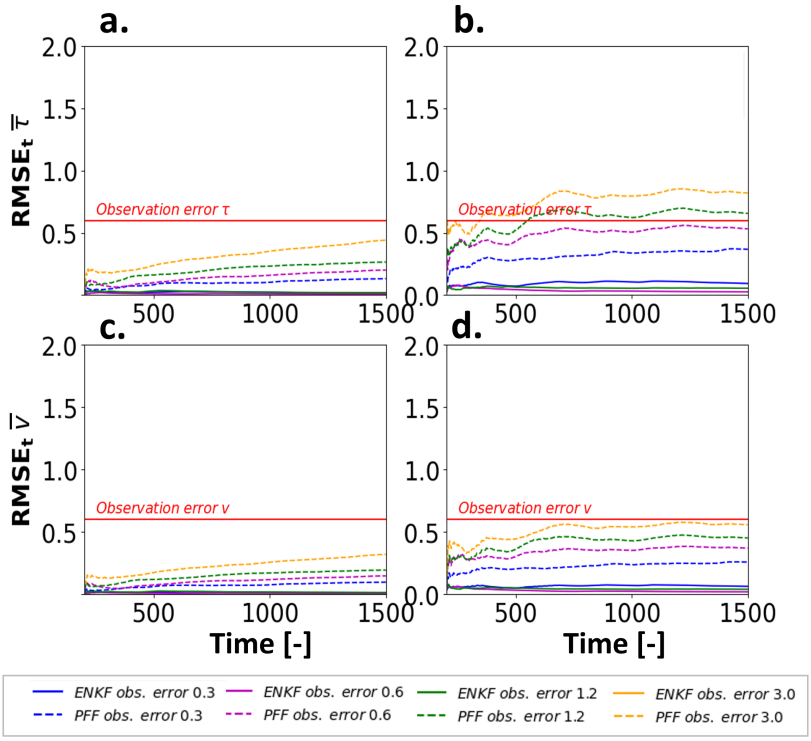


Figure C.2: Comparison of the results of the EnKF (solid lines) and the PFF (dashed lines) for different observation errors. The left column results under a periodic condition, and the right column under chaotic conditions.

# D

# Derivation of an Observation Operator for the Assimilation of Shear Stress Observations

For the numerical models used in this thesis, we use quasi-dynamic simulations. The advantage of using this approach is that we have faster simulations by omitting the acceleration term $\dot{\bar{V}}_i$, which is the inertia of the system and contains the effect of the seismic waves and the wave-field. As explained in Chapter 2, some seismologists compensate for the effect of the inertia term by replacing it with a radiation damping term ($\eta \bar{V}$), which represents the energy loss by shear wave radiation. An advantage of considering this damping term instead of including the inertia term is that the relationship that the shear stress $\bar{\tau}$ depends on the current slip $\bar{u}$ and the current slip rate $\bar{V}$ which makes the sequential data assimilation easier to define. If we consider the fully dynamic definition, using Eq. 2.12 and 2.13, we have that:

$$\bar{\tau}_{i,FD} = -\left(\frac{\xi}{\gamma_\mu^2}\right)\left[-\gamma_\mu^2\,\bar{u}_{i-1} + \left(\gamma_\lambda^2 + 2\gamma_\mu^2\right)\bar{u}_i - \gamma_\mu^2\,\bar{u}_{i+1} + \dot{\bar{V}}_i\right], \tag{D.1}$$

where we refer to $\bar{\tau}_{i,FD}$ as the fully-dynamic shear stress as it considers the term $\bar{V}_i$. We define the quasi-dynamic shear stress $\bar{\tau}_{i,QD}$ by assuming that $\dot{\bar{V}}_i \to 0$, which means that:

$$\bar{\tau}_{i,QD} = -\left(\frac{\xi}{\gamma_\mu^2}\right)\left[-\gamma_\mu^2\,\bar{u}_{i-1} + \left(\gamma_\lambda^2 + 2\gamma_\mu^2\right)\bar{u}_i - \gamma_\mu^2\,\bar{u}_{i+1}\right]. \tag{D.2}$$

where we refer to $\bar{\tau}_{i,QD}$ as the quasi-dynamic shear stress.

Interestingly, there is a very strong coupling between the blocks of the system when assuming the quasi-dynamic assumption that makes the system behave as a single block

under periodic conditions. In this situation the shear stress $\bar{\tau}$ simplifies to:

$$\bar{\tau}_{i,SB} = -\left(\xi \frac{\gamma_\lambda^2}{\gamma_\mu^2}\right)[\bar{u}_i]\,. \tag{D.3}$$

Fig. D.1 shows a comparison on the upper row of the use of the three expressions for the shear stress to describe the same system for the periodic case, while the lower row shows the comparison for the chaotic case. As we can see, the shear stress for the quasi-dynamic case and the single block are almost indistinguishable. The effect of this is that we would require fewer observations to estimate the system's state correctly, as we would have enough information by only observing one block. However, the stress behavior is very different when considering the term inertia, especially in the coseismic phase. This means more observations are needed to resolve the fully-dynamic shear stress compared to a periodic quasi-dynamic model. We can see that the strong coupling between blocks seen in the periodic case is lost in the chaotic case. A single block's shear stress is very different from the quasi-dynamic approximation. Noteworthy is that the timing of the peak stress is very similar for the three definitions of shear stress, which means that any of them can be used to estimate the occurrence of an earthquake. The main differences are the stress drop during the coseismic phase and the peak stress reached before that drop.

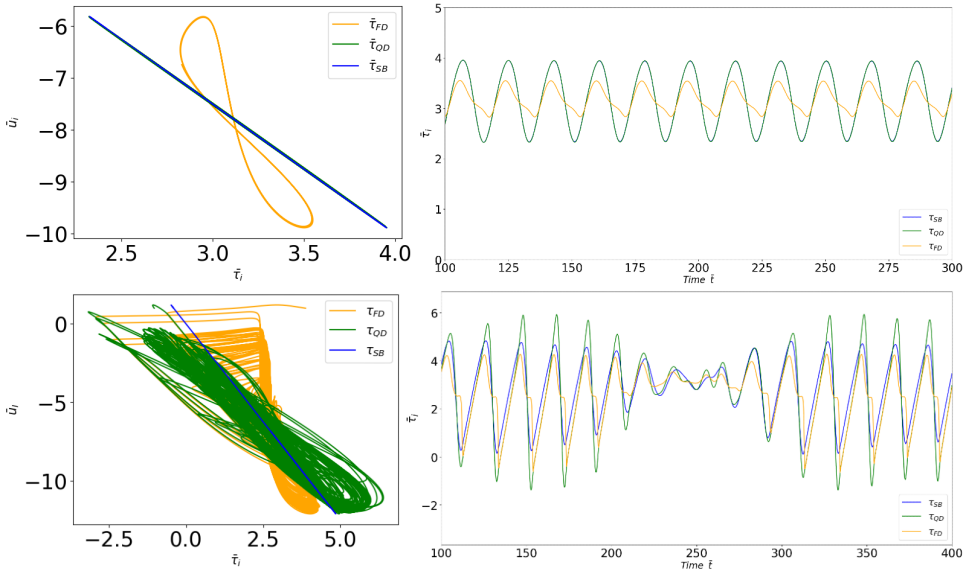


Figure D.1: Comparison shear stress single block, quasi dynamic, and fully dynamic. The upper row shows the comparison for the periodic case of the 1-D Burridge Knopoff model, while the lower row shows the comparison for the chaotic case.

To avoid further instabilities in the update step of the data assimilation, coming from the non-linearity imposed by the logarithmic terms in the ordinary differential equations

from Burridge-Knopoff models governed by rate-and-state friction, we decided to include the shear stress in the state vector and the observation vector and use a simplified observation operator. In this section, we try to derive the definition of an observation operator for an assimilation in which the shear stress is part of the observation operator but not in the state vector. The reason is that numerically, as shown in the previous equations, there is a known physical relationship between the shear stress and the slip and the slip rate. We derive the observation operator for $\bar{\tau}_{i,QD}$ and $\bar{\tau}_{i,SB}$ by stating that:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_u & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}, \tag{D.4}$$

where $\mathbf{H}_{u,k}$ is

$$\mathbf{H}_u = \left[H_{i,j}\right]_{N_c \times N_c} = \begin{bmatrix} \beta_1 & \beta_2 & 0 & 0 & \cdots & 0 & 0 & \beta_2 \\ \beta_2 & \beta_1 & \beta_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \beta_2 & \beta_1 & \beta_2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \beta_2 & \beta_1 & \beta_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \beta_1 & \beta_2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \beta_2 & \beta_1 & \beta_2 \\ \beta_2 & 0 & 0 & 0 & \cdots & 0 & \beta_2 & \beta_1 \end{bmatrix}, \tag{D.5}$$

Moreover, relates the state of the slip of the block $i$ and the surrounding blocks $i-1$ and $i+1$ to the shear stress observation at $i$. Therefore, it carries information about spatial relationships. In the single block case $\beta_1 = -\xi\frac{\gamma_\lambda^2}{\gamma_\mu^2}$, $\beta_2 = 0$, and for the quasidynamic case $\beta_1 = -\xi\left[\frac{\gamma_\lambda^2+2\gamma_\mu^2}{\gamma_\mu^2}\right]$, $\beta_2 = \xi$.

In Chapter 5, we avoided using this observation operator because it caused several stability issues in the data assimilation in the perfect model experiments that we believe are caused by ignoring the inertia term. However, we presented it here for other practitioners who would like to introduce it in their assimilation to better respect the relationships between the state variables or better formulate an observation operator for the fully-dynamic case.

# REFERENCES


[1] Aalsburg, J., Grant, L.B. van, Yakovlev, G., Rundle, P.B., Rundle, J.B., Turcotte, D.L., and Donnellan, A. "A Feasibility Study of Data Assimilation in Numerical Simulations of Earthquake Fault Systems, Physics of the Earth and Planetary Interiors". In: *Physics of the Earth and Planetary Interiors* (2007). DOI: 10.1016/j.pepi.2007.04.020.

[2] Aanonsen, S. I., Naevdal, G., Oliver, D. S., C., Reynolds A., and Valles, B. "Ensemble Kalman filter in reservoir engineering—A review". In: *SPE Journal* 14 (2009), pp. 393–412. DOI: 10.21188/117274-PA.

[3] Abdel-aal, Abdel-aziz Khairy and Yagi, Yuji. "Earthquake source characterization, moment tensor solutions, and stress field of small-moderate earthquakes occurred in the northern Red Sea Triple Junction". In: *Geosciences Journal* 21 (2017), pp. 235–251.

[4] Allen, Richard M and Kanamori, Hiroo. "The potential for earthquake early warning in southern California". In: *Science* 300.5620 (2003), pp. 786–789.

[5] Allen, Richard M. and Melgar, Diego. "Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs". In: *Annual Review of Earth and Planetary Sciences* 47 (2019), pp. 361–388. DOI: 10.1146/annurev-earth-053018-060457.

[6] Ampuero, Jean-Paul and Rubin, Allan M. "Earthquake nucleation on rate and state faults–Aging and slip laws". In: *Journal of Geophysical Research: Solid Earth* 113.B1 (2008).

[7] Baki, Zhuldyzay and Lieshout, MNM van. "The influence of gas production on seismicity in the Groningen field". In: *Proceedings of the 10th International Workshop on Spatio-Temporal Modelling METMA X*. 2022, pp. 163–167.

[8] Banerjee, A., van Dinther, Y., and Vossepoel, F. C. "On parameter bias in earthquake sequence models using data assimilation". In: *Nonlinear Processes in Geophysics* 30.2 (2023), pp. 101–115. DOI: 10.5194/npg-30-101-2023. URL: https://npg.copernicus.org/articles/30/101/2023/.

[9] Bannister, R. N. "Review article: A review of operational methods of variational and ensemble-variational data assimilation". In: *Quarterly Journal of the Royal Meteorological Society* 143 (2017), pp. 607–633. DOI: 10.1002/qj.2982.

[10] Barbot, Sylvain. "Slow-slip, slow earthquakes, period-two cycles, full and partial ruptures, and deterministic chaos in a single asperity fault". In: *Tectonophysics* 768 (2019), p. 228171. ISSN: 0040-1951. DOI: https://doi.org/10.1016/j.tecto.2019.228171.

[11] Barbot, Sylvain, Lapusta, Nadia, and Avouac, Jean-Philippe. “Under the Hood of the Earthquake Machine: Toward Predictive Modeling of the Seismic Cycle”. In: *Science* 336.6082 (2012), pp. 707–710. DOI: 10.1126/science.1218796.

[12] Bau, D., Ferronato, M., Gambolati, G., Teatini, P., and Alzraiee, A. “Ensemble smoothing of land subsidence measurements for reservoir geomechanical characterization”. In: *International Journal for Numerical and Analytical Methods in Geomechanics* 39 (2 2014). DOI: 10.1002/nag.2309.

[13] Beeler, NM, Lockner, DL, and Hickman, SH. “A simple stick-slip and creep-slip model for repeating earthquakes and its implication for microearthquakes at Parkfield”. In: *Bulletin of the Seismological Society of America* 91.6 (2001), pp. 1797–1804.

[14] Ben-Zion, Y. and Rice, J. R. “Dynamic simulations of slip on a smooth fault in an elastic solid”. In: *Journal of Geophysical Research: Solid Earth* 102 (B8) (1997), pp. 17771–17784.

[15] Ben-Zion, Y. and Rice, J. R.. “Slip patterns and earthquake populations along different classes of faults in elastic solids”. In: *Journal of Geophysical Research: Solid Earth* 100 (B7) (1995), pp. 12959–12983.

[16] Beroza, Gregory C. and Ellsworth, William L. “Properties of the seismic nucleation phases”. In: *Tectonophysics* 261 (1996), pp. 209–277.

[17] Bommer, Julian J and Abrahamson, Norman A. “Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?” In: *Bulletin of the Seismological Society of America* 96 (2006), pp. 1967–1977. DOI: 10.1785/0120060043.

[18] Bourne, S. J., Oates, S. J., Bommer, J. J., Dost, B., van Elk, J., and Doornhof, D. “A Monte Carlo Method for Probabilistic Hazard Assessment of Induced Seismicity due to Conventional Natural Gas Production published”. In: *Netherlands Journal of Geosciences* 105 (3 2015). DOI: 10.1785/0120140302.

[19] Bourne, Stephen, Oates, S. J., Elk, Jan Folkert van, and Doornhof, D. “A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir”. In: *Journal of Geophysical Research: Solid Earth* 119 (12 2014). DOI: 10.1002/2014JB011663.

[20] Buijze, L, Van den Bogert, PAJ, Wassing, BBT, and Orlic, B. “Nucleation and arrest of dynamic rupture induced by reservoir depletion”. In: *Journal of Geophysical Research: Solid Earth* 124.4 (2019), pp. 3620–3645.

[21] Buijze, Loes, Bogert, Peter AJ van den, Wassing, Brecht BT, Orlic, Bogdan, and Veen, Johan ten. “Fault reactivation mechanisms and dynamic rupture modelling of depletion-induced seismic events in a Rotliegend gas reservoir”. In: *Netherlands Journal of Geosciences* 96.5 (2017), s131–s148.

[22] Burridge, R. and Knopoff, L. “Model and theoretical seismicity”. In: *Bulletin of the Seismological Society of America* 57.3 (1967), pp. 341–371. DOI: 10.1785/BSSA0570030341.

[23] Candela, Thibault, Osinga, Sander, Ampuero, Jean-Paul, Wassing, Brecht, Pluymaekers, Maarten, Fokker, Peter A, Wees, Jan-Diederik van, Waal, Hans A de, and Muntendam-Bos, Annemarie G. "Depletion-induced seismicity at the Groningen gas field: Coulomb rate-and-state models including differential compaction effect". In: *Journal of Geophysical Research: Solid Earth* 124.7 (2019), pp. 7081–7104.

[24] Candela, Thibault, Pluymaekers, Maarten, Ampuero, Jean-Paul, Van Wees, Jan-Diederik, Buijze, Loes, Wassing, Brecht, Osinga, Sander, Grobbe, Niels, and Muntendam-Bos, Annemarie G. "Controls on the spatio-temporal patterns of induced seismicity in Groningen constrained by physics-based modelling with Ensemble-Smoother data assimilation". In: *Geophysical Journal International* 229.2 (2022), pp. 1282–1308.

[25] Carlson, Jean M, Langer, James S, Shaw, Bruce E, and Tang, Chao. "Intrinsic properties of a Burridge-Knopoff model of an earthquake fault". In: *Physical Review A* 44.2 (1991), p. 884.

[26] Carrassi, A., Bocquet, Marc, Bertino, Laurent, and Evensen, Geir. "Data Assimilation in the Geosciences - An overview on methods, issues and perspectives". In: *Wiley Interdisciplinary Reviews: Climate Change* (Sept. 2017). DOI: 10.1002/wcc.535.

[27] Chen, J., Verberne, B.A., and Spiers, C.J. "Effects of healing on the seismogenic potential of carbonate fault rocks: Experiments on samples from the Longmenshan Fault, Sichuan, China". In: *Journal of Geophysical Research: Solid Earth* (2015). DOI: 10.1002/2015JB012051.

[28] Chen, Jianye, Niemeijer, AR, and Spiers, Christopher J. "Microphysical modeling of carbonate fault friction at slip rates spanning the full seismic cycle". In: *Journal of Geophysical Research: Solid Earth* 126.3 (2021), e2020JB021024.

[29] Chustagulprom, Nawinda, Reich, Sebastian, and Reinhardt, Maria. "A Hybrid Ensemble Transform Particle Filter for Nonlinear and Spatially Extended Dynamical Systems". In: *SIAM/ASA Journal on Uncertainty Quantification* 4 (Jan. 2016), pp. 592–608. DOI: 10.1137/15M1040967.

[30] Cochard, A. and Madariaga, R. "Dynamic faulting under rate-dependent friction". In: *Pure and applied geophysics* 142 (3) (1994), pp. 419–445.

[31] Cornell, C. A. "Engineering seismic risk analysis". In: *Bulletin of the Seismological Society of America* 63 (1968), pp. 1583–1606.

[32] Crupi, P. and Bizzarri, A. "The role of radiation damping in the modeling of repeated earthquake events". In: *Annals of Geophysics* 56 (1) (2013).

[33] Dahm, Torsten, Cesca, Simone, Hainzl, Sebastian, Braun, Thomas, and Krüger, Frank. "Discrimination between induced, triggered, and natural earthquakes close to hydrocarbon reservoirs: A probabilistic approach based on the modeling of depletion-induced stress changes and seismological source parameters". In: *Journal of Geophysical Research: Solid Earth* 120.4 (2015), pp. 2491–2509.

[34] Dal Zilio, Luca, van Dinther, Ylona, Gerya, Taras, and Avouac, Jean-Philippe. "Bimodal seismicity in the Himalaya controlled by fault friction and geometry". In: *Nature communications* 10.1 (2019), pp. 1–11.

[35] Daužickaitė, Ieva, Lawless, Amos S, Scott, Jennifer A, and Van Leeuwen, Peter Jan. "Randomised preconditioning for the forcing formulation of weak-constraint 4D-Var". In: *Quarterly Journal of the Royal Meteorological Society* 147.740 (2021), pp. 3719–3734.

[36] Deventer, Julia van. "Earthquake Detection in Zeerijp: A Study on the Usage of Template Matching and Neural Networks for Detection of Small Earthquakes in Zeerijp". In: (2021).

[37] Di Toro, Giulio, Han, Raehee, Hirose, Takehiro, De Paola, Nicola, Nielsen, Stefan, Mizoguchi, Kazuo, Ferri, Fabio, Cocco, Massimo, and Shimamoto, Tosihiko. "Fault lubrication during earthquakes". In: *Nature* 471.7339 (2011), pp. 494–498.

[38] Diab-Montero, Hamed Ali, Li, Meng, van Dinther, Ylona, and Vossepoel, Femke C. "Estimating the occurrence of slow slip events and earthquakes with an ensemble Kalman filter". In: *Geophysical Journal International* 234.3 (2023), pp. 1701–1721.

[39] Dieterich, J. H. "Modeling of rock friction: 1. Experimental results and constitutive equations". In: *Journal of Geophysical Research* 84 (1979). DOI: 10.1029/JB084iB05p02161.

[40] Dieterich, J. H. "Time-dependent friction and the mechanics of stick-slip". In: *Pure and Applied Geophysics* 116 (1978). DOI: 10.1007/BF00876539.

[41] Dieterich, James. "A constitutive law for rate of earthquake production and its application to earthquake clustering". In: *Journal of Geophysical Research: Solid Earth* 99.B2 (1994), pp. 2601–2618.

[42] Dieterich, James H and Richards-Dinger, Keith B. "Earthquake recurrence in simulated fault systems". In: *Seismogenesis and Earthquake Forecasting: The Frank Evison Volume II*. Springer, 2010, pp. 233–250.

[43] Dost, B., van Eck, T., and Haak, H. "Scaling of peak ground acceleration and peak ground velocity recorded in the Netherlands". In: *Bollettino di Geofisica Teorica ed Applicata* 45 (3 2004).

[44] Dost, Bernard, Ruigrok, Elmer, and Spetzler, Jesper. "Development of seismicity and probabilistic hazard assessment for the Groningen gas field". In: *Netherlands Journal of Geosciences* 96 (Dec. 2017), s235–s245. DOI: 10.1017/njg.2017.20.

[45] Elbertsen, Rens, Niemeijer, André Niemeijer, and Vasconcelos, Ivan. *Dimension Reduction Applied to Laboratory Earthquake Acoustic Emissions for Visual Analysis and Machine Learning*. Tech. rep. Copernicus Meetings, 2023.

[46] Ellsworth, William L. and Beroza, Gregory C. "Seismic Evidence for an Earthquake Nucleation Phase". In: *Science* 268 (1995), pp. 851–855.

[47] Emerick, A. A. "Deterministic ensemble smoother with multiple data assimilation as an alternative for history matching seismic data". In: *Computational Geosciences* 22 (2018). DOI: 10.1007/s10596-018-9745-5.

[48] Emerick, Alexandre and Reynolds, Albert. “Investigation of the Sampling Performance of Ensemble-based Methods with a Simple Reservoir Model”. In: *Computational Geosciences* 17 (Apr. 2013). DOI: 10.1007/s10596-012-9333-z.

[49] Erickson, B., Birnir, B., and Lavallée, D. “A model for aperiodicity in earthquakes”. In: *Nonlinear Processes in Geophysics* 15.1 (2008), pp. 1–12. DOI: 10.5194/npg-15-1-2008.

[50] Erickson, Brittany A., Birnir, Björn, and Lavallée, Daniel. “Periodicity, chaos and localization in a Burridge–Knopoff model of an earthquake with rate-and-state friction”. In: *Geophysical Journal International* 187.1 (Oct. 2011), pp. 178–198. DOI: 10.1111/j.1365-246X.2011.05123.x.

[51] Eshelby, John Douglas. “The determination of the elastic field of an ellipsoidal inclusion, and related problems”. In: *Proceedings of the royal society of London. Series A. Mathematical and physical sciences* 241.1226 (1957), pp. 376–396.

[52] Esteva, L. “Criteria for the construction of spectra for seismic design”. In: *Panamerican Symposium of Structures* (1967).

[53] Evensen, G. “Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics”. In: *Computational Geosciences* 99 (1994), pp. 10143–10162. DOI: 10.1029/94JC00572.

[54] Evensen, G., Amezcua, J., Bocquet, M., Carrassi, A., Farchi A.and Fowler, A., L., Houtekamer P., Jones, C. K., Moraes, R. J. de, Pulido, M., Sampson, C., and Vossepoel, F. C. “An international initiative of predicting the SARS-CoV-2 pandemic using ensemble data assimilation”. In: *Foundations of Data Science* 65 (2020). DOI: 10.3934/fods.2021001.

[55] Evensen, G. and Eikrem, K. S. “Strategies for conditioning reservoir models on rate data using ensemble smoothers”. In: *Computational Geosciences* 22 (2018), pp. 1251–1270. DOI: 10.1007/s10596-018-9750-8.

[56] Evensen, Geir. “The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation”. In: *Ocean Dynamics* 53 (2003), pp. 343–367. DOI: 10.1007/s10236-003-0036-9.

[57] Evensen, Geir, Vossepoel, Femke C., and van Leeuwen, Peter Jan. *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. 1st ed. Open access. Springer, 2022, p. 254. ISBN: 978-3-030-96708-6. DOI: 10.1007/978-3-030-96709-3.

[58] Faulkner, D.R., Jackson, C.A.L., Lunn, R.J., Schlische, R.W., Shipton, Z.K., Wibberley, C.A.J., and Withjack, M.O. “A review of recent developments concerning the structure, mechanics and fluid flow properties of fault zones”. In: *Journal of Structural Geology* 32 (2010), pp. 1557–1575. DOI: 10.1016/j.jsg.2010.06.009.

[59] Fokker, P. A., Wassing, B. B. T., van Leijen F. J.and Hanssen, R. F., and Nieuwland, D. A. “Data assimilation of PS-InSAR movement measurements applied to the Bergermeer gas field”. In: *International Workshop on Geomechanics and Energy* (2009). DOI: 10.3997/2214-4609.20131971.

[60] Fukuyama, Eiichi, Mizoguchi, Kazuo, Yamashita, Futoshi, Togo, Tetsuhiro, Kawakata, Hiroyuki, Yoshimitsu, Nana, Shimamoto, Toshihiko, Mikoshiba, Tadashi, Sato, Makoto, Minowa, Chikahiro, et al. "Large-scale biaxial friction experiments using a NIED large-scale shaking table". In: *Rep Nat'l Res Inst Earth Sci Disas Prev* 81 (2014), pp. 15–35.

[61] Galis, Martin, Ampuero, Jean Paul, Mai, P., and Cappa, Frederic. "Induced seismicity provides insight into why earthquake ruptures stop". In: *Science Advances* 3 (Dec. 2017). DOI: 10.1126/sciadv.aap7528.

[62] Geller, Robert. "Earthquake prediction: a critical review". In: *Geophysical Journal International* 131 (Apr. 1997), pp. 425–450. DOI: 10.1111/j.1365-246X.1997.tb06588.x.

[63] Geller, Robert. "Shake-up time for Japanese seismology". In: *Nature* 472 (Apr. 2011), pp. 407–9. DOI: 10.1038/nature10105.

[64] Gualandi, Adriano, Avouac, J-P, Michel, Sylvain, and Faranda, Davide. "The predictable chaos of slow earthquakes". In: *Science advances* 6.27 (2020), eaaz5548.

[65] Gualandi, Adriano, Faranda, Davide, Marone, Chris, Cocco, Massimo, and Mengaldo, Gianmarco. "Deterministic and stochastic chaos characterize laboratory earthquakes". In: *Earth and Planetary Science Letters* 604 (2023), p. 117995.

[66] Guzman, Jorge, Babaei, Masoud, Shi, Ji-Quan, Korre, Anna, and Durucan, Sevket. "Coupled flow-geomechanical performance assessment of CO2 storage sites using the Ensemble Kalman Filter". In: *Energy Procedia* 63 (Dec. 2014), pp. 3475–3482. DOI: 10.1016/j.egypro.2014.11.376.

[67] Hanks, Thomas C and Kanamori, Hiroo. "A moment magnitude scale". In: *Journal of Geophysical Research: Solid Earth* 84.B5 (1979), pp. 2348–2350.

[68] Heimann, Sebastian, Kriegerowski, Marius, Isken, Marius, Cesca, Simone, Daout, Simon, Grigoli, Francesco, Juretzek, Carina, Megies, Tobias, Nooshiri, Nima, Steinberg, Andreas, et al. "Pyrocko-An open-source seismology toolbox and library". In: (2017).

[69] Herrendörfer, Robert, Gerya, Taras, and van Dinther, Ylona. "An invariant rate- and state-dependent friction formulation for visco-elasto-plastic earthquake cycle simulations". In: *Journal of Geophysical Research: Solid Earth* 123 (May 2018). DOI: 10.1029/2017JB015225.

[70] Hirahara, Kazuro and Nishikiori, Kento. "Estimation of frictional properties and slip evolution on a long-term slow slip event fault with the ensemble Kalman filter: numerical experiments". In: *Geophysical Journal International* 219 (3 2019), pp. 2074–2096. DOI: 10.1093/gji/ggz415.

[71] Hori, Takane, Miyazaki, S, Hyodo, M, Nakata, R, and Kaneda, Y. "Earthquake forecasting system based on sequential data assimilation of slip on the plate boundary". In: 62 (Jan. 2014), pp. 179–189. DOI: 10.11345/nctam.62.179.

[72] Hotta, Daisuke and Ota, Yoichiro. “Why does EnKF suffer from analysis overconfidence? An insight into exploiting the ever-increasing volume of observations”. In: *Quarterly Journal of the Royal Meteorological Society* 147.735 (2021), pp. 1258–1277.

[73] Hu, C.-C. and van Leeuwen, P. J. “A particle flow filter for high-dimensional system applications”. In: *Q. J. Roy. Meteor. Soc.)* 147 (2021), pp. 2352–2374. DOI: 10.1002/qj.4028.

[74] Jagt, Lisanne, Ruigrok, Elmer, and Paulssen, Hanneke. “Relocation of clustered earthquakes in the Groningen gas field”. In: *Netherlands Journal of Geosciences* 96 (2017). DOI: 10.1017/njg.2017.12.

[75] Jansen, Jan-Dirk and Meulenbroek, Bernard. “Induced aseismic slip and the onset of seismicity in displaced faults”. In: *Netherlands Journal of Geosciences* 101 (2022), e13.

[76] Jansen, JD, Singhal, P, and Vossepoel, FC. “Insights from closed-form expressions for injection-and production-induced stresses in displaced faults”. In: *Journal of Geophysical Research: Solid Earth* 124.7 (2019), pp. 7193–7212.

[77] Ji, Yuntao, Niemeijer, André, Baden, Dawin, Yamashita, Futoshi, Xu, Shiqing, Hunfeld, Luuk, Pijnenburg, Ronald PJ, Fukuyama, Eiichi, and Spiers, Christopher. “Friction law for earthquake nucleation: size doesn’t matter”. In: (2022).

[78] Ji, Yuntao, Niemeijer, André, Baden, Dawin, Yamashita, Futoshi, Xu, Shiqing, Hunfeld, Luuk, Pijnenburg, Ronald PJ, Fukuyama, Eiichi, and Spiers, Christopher. “Friction law for earthquake nucleation: size doesn’t matter”. In: (2022).

[79] Jiang, Junle, Erickson, Brittany A, Lambert, Valère R, Ampuero, Jean-Paul, Ando, Ryosuke, Barbot, Sylvain D, Cattania, Camilla, Zilio, Luca Dal, Duan, Benchun, Dunham, Eric M, et al. “Community-driven code comparisons for three-dimensional dynamic modeling of sequences of earthquakes and aseismic slip”. In: *Journal of Geophysical Research: Solid Earth* 127.3 (2022), e2021JB023519.

[80] Kano, Masayuki, Miyazaki, Shinichi, Ishikawa, Yoichi, and Hirahara, Kazuro. “Adjoint-based Direct Data Assimilation of GNSS Time Series for Optimizing Frictional Parameters and Predicting Postseismic Deformation Following the 2003 Tokachi-Oki Earthquake”. In: *Journal of Geophysical Research: Solid Earth* (2020). DOI: 10.21203/rs.3.rs-39393/v1.

[81] Kano, Masayuki, Miyazaki, Shinichi, Ishikawa, Yoichi, Hiyoshi, Yoshihisa, Ito, Kosuke, and Hirahara, Kazuro. “Real data assimilation for optimization of frictional parameters and prediction of afterslip in the 2003 Tokachi-Oki earthquake inferred from slip velocity by an adjoint method”. In: *Geophysical Journal International* (2015). DOI: 10.1093/gji/ggv289.

[82] Kano, Masayuki, Miyazaki, Shinichi, Ito, Kosuke, and Hirahara, Kazuro. “An adjoint data assimilation method for optimizing frictional parameters on the afterslip area”. In: *Earth, Planets and Space* 65 (2013), pp. 1575–1580. DOI: 10.1029/2019JB019047.

[83] Kano, Masayuki, Miyazaki, Shinichi, Ito, Kosuke, and Hirahara, Kazuro. “Estimation of Frictional Parameters and Initial Values of Simulation Variables Using an Adjoint Data Assimilation Method with Synthetic Afterslip Data”. In: (2010).

[84] Keiiti, Aki. “Generation and Propagation of G Waves from the Niigata Earthquake of June 16, 1964: Part 1. A statistical analysis”. In: = *Bulletin of the Earthquake Research Institute, University of Tokyo* 44.1 (1966), pp. 23–72.

[85] Kekem, Dirk Leendert van. “Dynamics of the Lorenz-96 model: Bifurcations, symmetries and waves”. In: (2018).

[86] Knopoff, Leon. “Energy release in earthquakes”. In: *Geophysical Journal International* 1.1 (1958), pp. 44–52.

[87] Kühn, Daniela, Heimann, Sebastian, Isken, Marius P, Ruigrok, Elmer, and Dost, Bernard. “Probabilistic moment tensor inversion for hydrocarbon-induced seismicity in the Groningen gas field, The Netherlands, Part 1: Testing”. In: *Bulletin of the Seismological Society of America* 110.5 (2020), pp. 2095–2111.

[88] Lapusta, N. and Liu, Y. “Three-dimensional boundary integral modeling of spontaneous earthquake sequences and aseismic slip”. In: *Journal Geophysical Research: Solid Earth* (2009). DOI: 10.1029/2008JB005934.

[89] Lapusta, Nadia et al. “Modeling Earthquake Source Processes: from Tectonics to Dynamic Rupture”. In: *Report to the National Science Foundation* (2019).

[90] Lapusta, Nadia and Rice, James R. “Nucleation and early seismic propagation of small and large events in a crustal earthquake model”. In: *Journal of Geophysical Research: Solid Earth* 108.B4 (2003).

[91] Lapusta, Nadia, Rice, James R, Ben-Zion, Yehuda, and Zheng, Gutuan. “Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate-and state-dependent friction”. In: *Journal of Geophysical Research: Solid Earth* 105.B10 (2000), pp. 23765–23789.

[92] Lee, Shiann-Jong, Huang, Bor-Shouh, Ando, Masataka, Chiu, Hung-Chie, and Wang, Jeen-Hwa. “Evidence of large scale repeating slip during the 2011 Tohoku-Oki earthquake”. In: *Geophysical Research Letters* 38.19 (2011).

[93] Lele, S.P., Hsu, S.Y., Garzon, J.L., De Dontney, N., Searles, K.H., Gist, G. A., Sanz, P.F., Biediger, E.A.O., and Dale, B.A. “Geomechanical modeling to evaluate production-induced seismicity at Groningen Field”. In: *Society of Petroleum Engineers* (Nov. 2016). DOI: 10.2118/183554-MS.

[94] Li, Meng, Dinther, Ylona van, and Niemeijer, Andre. “Earthquake nucleation location and slip behavior altered by normal stress heterogeneities”. In: *AGU Fall Meeting Conference Abstracts*. 2023, T53A–04.

[95] Li, Meng, Niemeijer, Andre, and Dinther, Ylona van. “Earthquake nucleation location and slip behavior altered by normal stress heterogeneities”. In: *AGU23* (2023).

[96] Li, Meng, Niemeijer, Andre, and Van Dinther, Ylona. “Why induced earthquakes occur on conventionally stable faults: frictional healing explains”. In: (2023). DOI: 10.21203/rs.3.rs-3317840/v1.

[97] Li, Meng, Pranger, Casper, and van Dinther, Ylona. "Characteristics of earthquake cycles: A cross-dimensional comparison of 0D to 3D numerical models". In: *Journal of Geophysical Research: Solid Earth* (2022), e2021JB023726. DOI: https://doi.org/10.1029/2021JB023726.

[98] Lieshout, MNM van and Baki, Zhuldyzay. "Exploring Seismic Hazard in the Groningen Gas Field Using Adaptive Kernel Smoothing". In: *Mathematical Geosciences* (2023), pp. 1–22.

[99] Liu, Y. and Rice, J. R. "Spontaneous and triggered aseismic deformation transients in a subduction fault model". In: *Journal of Geophysical Research: Solid Earth* 112 (B9)) (2007), pp. 17771–17784.

[100] Liu, Y., Weerts, A. H., Clark, M., Franssen, Hendricks, H. J. Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., Dijk, van, M., A. I. J., Velzen, N. van, He, M., Lee H., Noh, S. J. Rakovec, O., and Restrepo, P. "Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities". In: *Hydrology and Earth System Sciences* 16 (2012), pp. 3863–3887. DOI: 10.5194/hess-16-3863-2012.

[101] Lorenc, Andrew C. "The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var". In: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 129.595 (2003), pp. 3183–3203.

[102] Lorenz, Edward N. and Emanuel, Kerry A. "Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model." In: *Journal of Atmospheric Sciences* 55.3 (Feb. 1998), pp. 399–414. DOI: 10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

[103] Lu, Jianfeng, Lu, Yulong, and Nolen, James. "Scaling limit of the Stein variational gradient descent: The mean field regime". In: *SIAM Journal on Mathematical Analysis* 51.2 (2019), pp. 648–671.

[104] Luo, Yingdi, Ampuero, Jean Paul, Galvez, Percy, Ende, Martijn van den, and Idini, Benjamin. "QDYN: a Quasi-DYNamic earthquake simulator (v1. 1)". In: *Zenodo* (2017).

[105] Madariaga, R. "Study of an oscillator of single degree of freedom with dieterich-ruina rate and state friction, Laboratoire de Géologie, Ecole Normale Supérieure". In: *Unpublished Notes* (1998), pp. 643–696.

[106] Madariaga, Raul. "Dynamics of an expanding circular fault". In: *Bulletin of the Seismological Society of America* 66.3 (1976), pp. 639–666.

[107] Madariaga, Raul. "On the relation between seismic moment and stress drop in the presence of stress and strength heterogeneity". In: *Journal of Geophysical Research: Solid Earth* 84.B5 (1979), pp. 2243–2250.

[108] Maeda, Takuto, Obara, Kazushige, Shinohara, Masanao, Kanazawa, Toshihiko, and Uehirao, Kenji. "Successive estimation of a tsunami wavefield without earthquake source data: A data assimilation approach toward real-time tsunami forecasting". In: *Geophysical Research Letters* (2015). DOI: 10.1002/2015GL065588.

[109] Marone, Chris. “Laboratory-derived friction laws and their application to seismic faulting”. In: *Annual Review of Earth and Planetary Sciences* 26.1 (1998), pp. 643–696.

[110] Matsumoto, Satoshi, Katao, Hiroshi, and Iio, Yoshihisa. “Determining changes in the state of stress associated with an earthquake via combined focal mechanism and moment tensor analysis: Application to the 2013 Awaji Island earthquake, Japan”. In: *Tectonophysics* 649 (2015), pp. 58–67.

[111] Mehranpour, MH, Hangx, SJT, and Spiers, CJ. “Compaction of the Groningen gas reservoir sandstone: Discrete element modeling using microphysically based grain-scale interaction laws”. In: *Journal of Geophysical Research: Solid Earth* 126.9 (2021), e2021JB021722.

[112] Minson, Sarah E, Simons, M, Beck, JL, Ortega, F, Jiang, J, Owen, SE, Moore, AW, Inbal, A, and Sladen, A. “Bayesian inversion for finite fault earthquake source models–II: the 2011 great Tohoku-oki, Japan earthquake”. In: *Geophysical Journal International* 198.2 (2014), pp. 922–940.

[113] Mousavi, S Mostafa, Zhu, Weiqiang, Sheng, Yixiao, and Beroza, Gregory C. “CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection”. In: *Scientific reports* 9.1 (2019), p. 10267.

[114] Muntendam-Bos, Annemarie G, Hoedeman, Gerco, Polychronopoulou, Katerina, Draganov, Deyan, Weemstra, Cornelis, Zee, Wouter van der, Bakker, Richard R, and Roest, Hans. “An overview of induced seismicity in the Netherlands”. In: *Netherlands Journal of Geosciences* 101 (2022), e1.

[115] Nepveu, Manuel, van Thienen-Visser, Karin, and Sijacic, Danijela. “Statistics of seismic events at the Groningen field”. In: *Bulletin of Earthquake Engineering* 14 (77 2016). DOI: 10.1007/s10518-016-0007-4.

[116] Niemeijer, A.R., Di Toro, G., Griffith, W.A., Bistacchi, A., Smith, S.A.F., and Nielsen, S. “Inferring earthquake physics and chemistry using an integrated field and laboratory approach”. In: *Journal of Structural Geology* (2012). DOI: 10.1016/j.jsg.2012.02.018.

[117] Noda, Hiroyuki, Dunham, Eric M, and Rice, James R. “Earthquake ruptures with thermal weakening and the operation of major faults at low overall stress levels”. In: *Journal of Geophysical Research: Solid Earth* 114.B7 (2009).

[118] Oba, Atsuki, Furumura, Takashi, and Maeda, Takuto. “Data Assimilation-Based Early Forecasting of Long-Period Ground Motions for Large Earthquakes Along the Nankai Trough”. In: *Journal of Geophysical Research: Solid Earth* (2020). DOI: 10.1029/2019JB019047.

[119] Ogata, Y. “Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes”. In: *Journal Geophysical Research* 104 (B8 1999), pp. 17995–18014. DOI: 10.1029/1999JB900115.

[120] Ogata, Yosihiko. “Statistical models for earthquake occurrences and residual analysis for point processes”. In: *Journal of the American Statistical association* 83.401 (1988), pp. 9–27.

[121] Ohta, Yusaku, Kobayashi, Tatsuya, Tsushima, Hiroaki, Miura, Satoshi, Hino, Ryota, Takasu, Tomoji, Fujimoto, Hiromi, Iinuma, Takeshi, Tachibana, Kenji, Demachi, Tomotsugu, et al. "Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: Application to the 2011 Tohoku-Oki earthquake (Mw 9.0)". In: *Journal of Geophysical Research: Solid Earth* 117.B2 (2012).

[122] Oliver, Dean S., Zhang, Yanfen, Phale, Hemant A, and Chen, Yan. "Distributed parameter and state estimation in petroleum reservoirs". In: (2011).

[123] Ordaza, Mario and Arroyo, Danny. "On Uncertainties in Probabilistic Seismic Hazard Analysis". In: *Earthquake Spectra* 32 (2016), pp. 1405–1418.

[124] Osinga, S and Kraaijpoel, DA. "TNO Model Chain Groningen: Update and quick scan comparison of 2020 HRA model". In: (2020).

[125] Ozgun Konca, Ali, Kaneko, Yoshihiro, Lapusta, Nadia, and Avouac, Jean-Philippe. "Kinematic inversion of physically plausible earthquake source models obtained from dynamic rupture simulations". In: *Bulletin of the Seismological Society of America* 103.5 (2013), pp. 2621–2644.

[126] Perol, Thibaut, Gharbi, Michaël, and Denolle, Marine. "Convolutional neural network for earthquake detection and location". In: *Science Advances* 4.2 (2018), e1700578.

[127] Pranger, C.C. "Unstable physical processes operating on self-governing fault system, Improved Modeling Methodology". In: *Ph.D. thesis, ETH Zurich* (2020).

[128] Pritchard, Matthew E, Allen, Richard M, Becker, Thorsten W, Behn, Mark D, Brodsky, Emily E, Bürgmann, Roland, Ebinger, Cindy, Freymueller, Jeff T, Gerstenberger, Matt, Haines, Bruce, et al. "New opportunities to study earthquake precursors". In: *Seismological Research Letters* 91.5 (2020), pp. 2444–2447.

[129] Rabinowicz, Ernest. "The nature of the static and kinetic coefficients of friction". In: *Journal of applied physics* 22.11 (1951), pp. 1373–1379.

[130] Reichle, Rolf H. "Data assimilation methods in the Earth Sciences". In: *Advances in Water Resources* 31 (2008). DOI: 10.1016/j.advwatres.2008.01.001.

[131] Rice, James R and Gu, Ji-cheng. "Earthquake aftereffects and triggered seismic phenomena". In: *Pure and Applied Geophysics* 121 (1983), pp. 187–219.

[132] Rice, James R. "Spatio-temporal complexity of slip on a fault". In: *Journal of Geophysical Research: Solid Earth* 98 (B6) (Nov. 1993), pp. 9885–9907.

[133] Rodgers, Clive D. *Inverse methods for atmospheric sounding: theory and practice*. Vol. 2. World scientific, 2000.

[134] Rojas, Otilio, Dunham, Eric M, Day, Steven M, Dalguer, Luis A, and Castillo, Jose E. "Finite difference modelling of rupture propagation with strong velocity-weakening friction". In: *Geophysical Journal International* 179.3 (2009), pp. 1831–1858.

[135] Ross, Zachary E, Meier, Men-Andrin, Hauksson, Egill, and Heaton, Thomas H. "Generalized seismic phase detection with deep learning". In: *Bulletin of the Seismological Society of America* 108.5A (2018), pp. 2894–2901.

[136] Rouet-Leduc, Bertrand, Hulbert, Claudia, Lubbers, Nicholas, Barros, Kipton, Humphreys, Colin J, and Johnson, Paul A. "Machine learning predicts laboratory earthquakes". In: *Geophysical Research Letters* 44.18 (2017), pp. 9276–9282.

[137] Rubin, A. M. and Ampuero, J.-P. "Earthquake nucleation on (aging) rate and state faults". In: *Journal of Geophysical Research: Solid Earth* (2005). DOI: https://doi.org/10.1029/2005JB003686.

[138] Ruina, A. "Slip instability and state variable friction laws". In: *Journal of Geophysical Research* 88 (1983).

[139] Segall, Paul. *Earthquake and volcano deformation*. Princeton University Press, 2010.

[140] Segall, Paul and Bradley, Andrew M. "Slow-slip evolves into megathrust earthquakes in 2D numerical simulations". In: *Geophysical Research Letters* 39.18 (2012).

[141] Seif, Stefanie, Mignan, Arnaud, Zechar, Jeremy Douglas, Werner, Maximilian Jonas, and Wiemer, Stefan. "Estimating ETAS: The effects of truncation, missing data, and model assumptions". In: *Journal of Geophysical Research: Solid Earth* 122.1 (2017), pp. 449–469.

[142] Seiler, Alexandra, Evensen, Geir, Skjervheim, Jan-Arild, Hove, Joakim, and Vabø, Jon. "Advanced Reservoir Management Workflow Using an EnKF Based Assisted History Matching Method". In: 62 (Jan. 2009). DOI: 10.2118/118906-MS.

[143] Shaw, Bruce E, Milner, Kevin R, Field, Edward H, Richards-Dinger, Keith, Gilchrist, Jacquelyn J, Dieterich, James H, and Jordan, Thomas H. "A physics-based earthquake simulator replicates seismic hazard statistics across California". In: *Science advances* 4.8 (2018). DOI: 10.1126/sciadv.aau0688.

[144] Shearer, Peter M. *Introduction to seismology*. Cambridge university press, 2019.

[145] Simons, Mark, Fialko, Yuri, and Rivera, Luis. "Coseismic deformation from the 1999 M w 7.1 Hector Mine, California, earthquake as inferred from InSAR and GPS observations". In: *Bulletin of the Seismological Society of America* 92.4 (2002), pp. 1390–1402.

[146] Socquet, Anne, Valdes, Jesus Piña, Jara, Jorge, Cotton, Fabrice, Walpersdorf, Andrea, Cotte, Nathalie, Specht, Sebastian, Ortega-Culaciati, Francisco, Carrizo, Daniel, and Norabuena, Edmundo. "An 8 month slow slip event triggers progressive nucleation of the 2014 Chile megathrust". In: *Geophysical Research Letters* 44.9 (2017), pp. 4046–4053.

[147] Spiers, Christopher, Hangx, Suzanne, and Niemeijer, Andre. "New approaches in experimental research on rock and fault behaviour in the Groningen gas field". In: *Netherlands Journal of Geosciences* 96 (Dec. 2017), s55–s69. DOI: 10.1017/njg.2017.32.

[148] Stordal, Andreas S, Karlsen, Hans A, Nævdal, Geir, Skaug, Hans J, and Vallès, Brice. "Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter". In: *Computational Geosciences* 15 (2011), pp. 293–305.

[149] Stordal, Andreas S and Lorentzen, Rolf J. "An iterative version of the adaptive Gaussian mixture filter". In: *Computational Geosciences* 18 (2014), pp. 579–595.

[150] Stordal, Andreas S, Moraes, Rafael J, Raanes, Patrick N, and Evensen, Geir. "p-Kernel Stein variational gradient descent for data assimilation and history matching". In: *Mathematical Geosciences* 53.3 (2021), pp. 375–393.

[151] Thienen-Visser, Karin van and Breunese, J. N. "Induced seismicity of the Groningen gas field: History and recent developments". In: *The leading edge* 34 (6 2014). DOI: 10.1190/tle34060664.1.

[152] Thomas, M. Y., Lapusta, N., and Noda H.and Avouac, J.-P. "Quasi-dynamic versus fully dynamic simulations of earthquakes and aseismic slip with and without enhanced coseismic weakening". In: *Journal of Geophysical Research: Solid Earth* 119 (3) (2014), pp. 1986–2004.

[153] Togo, Tetsuhiro, Shimamoto, Toshihiko, Yamashita, Futoshi, Fukuyama, Eiichi, Mizoguchi, Kazuo, and Urata, Yumi. "Stick–slip behavior of Indian gabbro as studied using a NIED large-scale biaxial friction apparatus". In: *Earthquake Science* 28 (2015), pp. 97–118.

[154] Uchida, Naoki, Iinuma, Takeshi, Nadeau, Robert M, Bürgmann, Roland, and Hino, Ryota. "Periodic slow slip triggers megathrust zone earthquakes in northeastern Japan". In: *Science* 351.6272 (2016), pp. 488–492.

[155] Van Dinther, Y, Gerya, TV, Dalguer, LA, Mai, Paul Martin, Morra, G, and Giardini, D. "The seismic cycle at subduction thrusts: Insights from seismo-thermomechanical models". In: *Journal of Geophysical Research: Solid Earth* 118.12 (2013), pp. 6183–6202.

[156] van Dinther, Ylona, Künsch, Hans R, and Fichtner, Andreas. "Ensemble data assimilation for earthquake sequences: probabilistic estimation and forecasting of fault stresses". In: *Geophysical Journal International* 217 (2019). DOI: 10.1093/gji/ggz063.

[157] van Eck, Torild, Goutbeek, Femke, Haak, Hein, and Dost, Bernard. "Seismic hazard due to small-magnitude, shallow-source, induced earthquakes in The Netherlands". In: *Engineering Geology* 87 (1 2006). DOI: 10.1016/j.enggeo.2006.06.005.

[158] van Eijs, R.M.H.E., Mulders, Frans, Nepveu, Manuel, Kenter, C.J., and Scheffers, B.C. "Correlation between hydrocarbon reservoir properties and induced seismicity in the Netherlands". In: *Engineering Geology* 84 (May 2006), pp. 99–111. DOI: 10.1016/j.enggeo.2006.01.002.

[159] van Leeuwen, P. J. "Nonlinear data assimilation in geosciences: An extremely efficient particle filter". In: *Quarterly Journal of the Royal Meteorological Society* 136 (2010), pp. 1991–1999. DOI: 10.1002/qj.699.

[160] van Leeuwen, P. J. "Nonlinear ensemble data assimilation for the ocean. In Recent Developments in Data Assimilation for Atmosphere and Ocean". In: *ECMWF Seminar, September 8–12* (2003b), pp. 265–286.

[161] van Leeuwen, P.J. "Particle filtering in geophysical systems". In: *Monthly Weather Review* (2009). DOI: 10.1175/2009MWR2835.1.

[162] van Leeuwen, Peter Jan, Künsch, Hans R, Nerger, Lars, Potthast, Roland, and Reich, Sebastian. "Particle filters for high-dimensional geoscience applications: A review". In: *Quarterly Journal of the Royal Meteorological Society* 145.723 (2019), pp. 2335–2365. DOI: 10.1002/qj.3551.

[163] van Wees, J., Fokker, Peter, Thienen-Visser, Karin, Wassing, Brecht, Osinga, Sander, Orlic, Bogdan, Ghouri, Saad, Buijze, Loes, and Pluymaekers, Maarten. "Geomechanical models for induced seismicity in the Netherlands: Inferences from simplified analytical, finite element and rupture model approaches". In: *Netherlands Journal of Geosciences* 96 (Dec. 2017), s183–s202. DOI: 10.1017/njg.2017.38.

[164] van Wees, J.D., Buijze, L., Thienen-Visser, K. van, Nepveu, M., Wassing, B.B.T., Orlic, B., and Fokker, P.A. "Geomechanics response and induced seismicity during gas field depletion in the Netherlands". In: *Geothermics* (2014). DOI: 10.1016/j.geothermics.2014.05.0043.

[165] Varini, E. "A Monte Carlo method for filtering a marked doubly stochastic Poisson process". In: *Statistical Methods and Applications* 17 (2 2007), pp. 183–193.

[166] Verberne, B.A., Niemeijer, A.R., Bresser, J.H.P. de, and Spiers, C.J. "Mechanical behavior and microstructure of simulated calcite fault gouge sheared at 20-600c: Implications for natural faults in limestones". In: *Journal of Geophysical Research: Solid Earth* 120 (12 2015). DOI: 10.1002/2015JB012292.

[167] Verberne, B.A., Spiers, C.J., Niemeijer, A.R., Bresser J.H.P. andde Winter, D.A.M. de, and Pluemper, O. "Frictional properties and microstructure of calcite-rich fault gouges sheared at sub-seismic sliding velocities, Pure and Applied Geophysics". In: *Pure and Applied Geophysics volume* 171 (2014).

[168] Verwijs, Roos. "Machine Learning Techniques for Moment Tensor Estimation". In: (2021).

[169] Vetra-Carvalho, Sanita, van Leeuwen, Peter Jan, Nerger, Lars, Barth, Alexander, Altaf, M. Umer, Brasseur, Pierre, Kirchgessner, Paul, and Beckers, Jean-Marie. "State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems". In: *Tellus A: Dynamic Meteorology and Oceanography* 70 (1 2017). DOI: 10.1080/16000870.2018.1445364.

[170] Vossepoel, F. C. and Behringer, D. W. "Impact of sea level assimilation on salinity variability in the western equatorial Pacific". In: *Journal of Physical Oceanography* 30 (2000), pp. 1706–1721. DOI: 10.1175/1520-0485(2000)030<1706:IOSLAO>2.0.CO;2.

[171] Wang, Yuchen, Satake, Kenji, Maeda, Takuto, and Gusman, Aditya Riadi. "Data assimilation with dispersive tsunami model: a test for the Nankai Trough". In: *Earth, Planets and Space* (2018). DOI: 10.1186/s40623-018-0905-6.

[172] Wassing, B.B.T., van Wees, J.D., and Fokker, P.A. "Coupled continuum modeling of fracture reactivation and induced seismicity during enhanced geothermal operations". In: *Geothermics* 52 (2014), pp. 153–164. DOI: 10.1016/j.geothermics.2014.05.001.

[173] Wassing, BBT, Buijze, L, and Orlic, B. "Modelling of fault reactivation and fault slip in producing gas fields using a slip-weakening friction law". In: *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA. 2016, ARMA–2016.

[174] Weaver, A. T., Vialard, J., and Anderson, D. L. T. "Three- and four-dimensional variational assimilation in a general circulation model of the tropical Pacific Ocean. Part 1: Formulation, internal diagnostics and consistency checks". In: *Monthly Weather Review* 131 (2003), pp. 1360–1378. DOI: 10.1175/1520-0493.

[175] Wees, Jan-Diederik van, Osinga, Sander, Van Thienen-Visser, Karin, and Fokker, Peter A. "Reservoir creep and induced seismicity: inferences from geomechanical modeling of gas depletion in the Groningen field". In: *Geophysical Journal International* 212.3 (2018), pp. 1487–1497.

[176] Wees, Jan-Diederik van, Pluymaekers, Maarten, Osinga, Sander, Fokker, Peter, Van Thienen-Visser, Karin, Orlic, Bogdan, Wassing, Brecht, Hegen, Dries, and Candela, Thibault. "3-D mechanical analysis of complex reservoirs: a novel mesh-free approach". In: *Geophysical Journal International* 219.2 (2019), pp. 1118–1130.

[177] Werner, Maximilian, Ide, Kayo, and Sornette, Didier. "Earthquake forecasting based on data assimilation: Sequential Monte Carlo methods for renewal point processes". In: *Nonlinear Processes in Geophysics* 18 (Feb. 2011). DOI: 10.5194/npg-18-49-2011.

[178] Wibberley, Christopher A.J., Yielding, Graham, and Di Toro, Giulio. "Recent advances in the understanding of fault zone internal structure: A review". In: *Geological Society Special Publication* 299 (2008), pp. 5–33. DOI: 10.1144/SP299.2.

[179] Williams, Randolph T, Davis, Joshua R, and Goodwin, Laurel B. "Do large earthquakes occur at regular intervals through time? A perspective from the geologic record". In: *Geophysical Research Letters* 46.14 (2019), pp. 8074–8081.

[180] Wilschut, F., Peters, E., Visser, K., Fokker, P.A., and van Hooff, P.M.E. "Joint history matching of well data and surface subsidence observations using the ensemble Kalman filter: a field study". In: *SPE Reservoir Simulation Symposium* (2011). DOI: 10.2118/141690-MS.

[181] Yamashita, Futoshi, Fukuyama, Eiichi, Mizoguchi, Kazuo, Takizawa, Shigeru, Xu, Shiqing, and Kawakata, Hironori. "Scale dependence of rock friction at high work rate". In: *Nature* 528 (Dec. 2015), pp. 254–257. DOI: 10.1038/nature16138.

[182] Zbinden, Dominik, Rinaldi, Antonio Pio, Urpi, Luca, and Wiemer, Stefan. "On the physics-based processes behind production-induced seismicity in natural gas fields". In: *Society of Petroleum Engineers* 122 (5 Apr. 2016). DOI: 10.1002/2017JB014003.

# ACKNOWLEDGEMENTS

This dissertation was part of the 'InFocus: An Integrated Approach to Estimating Fault Slip Occurrence' project (grant number: DEEP.NL.2018.037), funded by the DeepNL program of the Dutch Research Council (NWO), which aims to improve the fundamental understanding of the deep subsurface dynamics under the influence of human interventions. One of the aspects that I like the most about my name (**Hamed**) is that one of its meanings is *grateful*. Without the help and support of many of my friends, colleagues, superiors, and family, it would have been impossible to achieve this dissertation. I want to thank many people who helped me reach this milestone in my life.

Firstly, I would like to express my sincere gratitude to my closest supervisors. To **Femke Vossepoel**, thank you for your support throughout the entire doctoral process. I have learned a lot from you over these past years. One of the lessons I have tried to apply most in my career by your side is the "less is more" principle. I have learned to apply this principle increasingly over time, but I will apply it more professionally. Thank you very much for your advice in writing, your support during the project's progress, and for maintaining the motivation to bring this work to a successful conclusion. I have great admiration for how you managed to balance your professional and personal life and also your passion for sports. I hope you have many more successes in all these areas and continue reaping triumphs. I also thank **Ylona van Dinther** for her support in developing this project. Your rigour in research and the project was a great balance with the freedom that Femke's perspective offered me to become an independent researcher. Thank you very much for the work tools and the knowledge you provided us regarding seismological aspects. I am happy that your health is much better, and also for all the achievements and academic milestones you have achieved during your career despite the obstacles. I wish you the best in your professional career and that you discover much more in the world of seismology, for which I know you have a great passion.

The guidance from my promotor, **Jan Dirk Jansen**, was also essential, especially during our annual meetings, to provide direction and purpose to this dissertation. I am very grateful for your advice and, especially, your perspective on decision-making and being more pragmatic in achieving our goals. I also appreciate the attention given, considering the load that came with holding the position of dean. I also hope you can enjoy research more peacefully in the coming years.

Likewise, I am very grateful to my colleagues from the InFocus project. Many thanks to my colleague **Meng Li** for his assistance in the project with numerical models and for helping me understand topics more practically, such as the rate-and-state friction law. I am incredibly grateful for your willingness to learn about data assimilation and for closely accompanying me in integrating the results of numerical simulations into the data assimilation framework. In this direction, I also want to thank **Casper** for allowing us to use the GARNET library, fundamental for the numerical modelling of our earthquake sequences, and to **Mohsen**, who accompanied us in expanding the capabilities of

physical modelling.

In InFocus, we had the support of experts in the seismic networks of the Netherlands, such as **Elmer Ruigrok**. Elmer, thank you very much for all your knowledge and for helping **Femke** and me with the co-supervision of the graduation projects of **Roos Verwijs** and **Julia van Deventer** in the use of machine learning for the rapid processing of seismic measurements. We also had the support of **André Niemeijer** and **Yuntao Ji** with knowledge and laboratory data to test the data assimilation concepts we developed on information acquired from state-of-the-art equipment. Thank you very much for your contributions and collaboration to understand better the work done in the laboratory and guide us in making our findings applicable.

In data assimilation, I learned a lot from both **Femke** and a large group of colleagues at TU Delft and other institutions around the world. Thanks to **Martin Verlaan** and the data assimilation research group of the mathematics department (**Arnold, Avelon, Amey, Kian, Henrique, Andrés, Jeffrey, and Elías**). Also, thanks to our data assimilation research group in the department of Geoscience and Engineering (**Gabriel, Mohsan, Arundhuti, Siamak, Dieter, Max Ramgraber, Samantha, Celine, Liang, and Jenny**). It was a pleasure to learn alongside you and share in our discussions. Thank you for addressing my multiple questions and expanding my knowledge in this fascinating area.

Also, in the direction of data assimilation, I want to thank **Peter Jan van Leeuwen** for the time I spent in Fort Collins at the University of Colorado. Thank you for allowing me to visit you and attend your data assimilation classes. Seeing how the research was done at a meteorological institute like CIRA, learning about the work of **Chih-Chi, Michael, Michelle, and Steven**, as well as our discussion sessions on the board in your office were critical for me to understand the methods of data assimilation better. Someone also very important was **Andreas Stordal**. Thank you very much for all your support in Bergen and your visits to Delft to understand better how to make the most of data assimilation methods' capabilities in your projects. Thank you for your patience and motivation; they have left a lasting impact. I wish you many successes in your sporting events and different sports fronts. The EnKF workshops in Bergen were delightful thanks to **Geir, Laurent, Remus, Rolf, Patrick, Xiaodong, Feda, Lars, and Michael**. Thanks to all of you guys for those conversations and having a beer about life and data assimilation; they were instrumental in advancing my understanding and resolving many doubts.

In the Department of Geoscience and Engineering, I also had great support in overcoming various obstacles along the way. A special thank you to **Anne Pluymakers** for discussing the challenges encountered in laboratory experiments related to seismicity and envisioning how data assimilation could help with specific difficulties. Likewise, the support of **Evert Slob** over the last six years has been a pillar during my time in the Applied Geophysics family of the IDEA League. His perspectives on education, science, research, and life have inspired me, and I hope his wisdom continues to enlighten future generations. Also, I am very grateful to **Jan Willem Thorbecke** for his assistance with supercomputing challenges and to **Liliana Vargas** for her advice on the storage, management, and processing of our research data.

Being part of the DeepNL project, I had the opportunity to work with great colleagues at TU Delft, including **Aukje**, **Sara**, **Faezeh**, **Aleks**, **Iban**, **Milad**, **Jingming**, and **Eddy**. Thank you for all our conversations about geomechanics, numerical modelling, labo-

ratory experiments, and inversion. It was a pleasure to spend time with you in the office, preparing scientific articles, presentations, abstracts, and more. It was a pleasure to continue the path to a doctorate after the master's with **Iban**, **Aukje**, **Menno**, and **Marat**. Thanks to **Milad** for being a great friend and supporting me in my best and worst moments. I sincerely hope you have much success in all your professional projects as well as in your personal life. Thanks also to **Jingming** for our dinners and to **Samantha** for encouraging me to participate in sports and constantly strive for improvement.

Likewise, I want to thank **David**, **Maria**, **Rigoberto**, **Billy**, **Eddy**, **Amanda** and **Florencia** for being like family support during these years, especially being far from home. Thank you for being a great source of support and being by my side in the highs and lows. Sharing trips, dinners, dances, movies, and many more activities has been a pleasure. I hope our friendship continues to thrive and that you succeed in every project you undertake—many blessings to each of your families. You truly made me feel understood and at home. Some of my other Latin friends to thank are **Valentina**, **Luisa Orozco**, **Diana**, **Valeria**, **Rogelio**, **Linnaea**, **Daniel Aramburo**, and the entire group of Colombians in Delft who also helped me navigate my way in the Netherlands.

Even from a distance, the support of **Sofia Michail**, **Ana Pardo**, **Nilgün Güdük**, and **Dorde Petrovic** has been critical. Thank you for opening your homes' doors in Europe and becoming invaluable friends. Thank you for the conversations about life, showing me my mistakes, and genuinely supporting me during my successes. Your advice, patience, time, and support were great strengths in overcoming difficulties and seeing a happier and brighter future. I wish you and your parents and siblings many blessings in your lives. I extend these wishes to **Maria Carrizo** and her family, as well as to my friends in Colombia **Ana María Toro**, **Alina Vásquez**, and **Lina Arrieta** who provided immense support, especially during my visits home.

I also extend my gratitude to **Christian Reinicke**, **Rutger van Limborgh**, **Entela**, **Zhenja**, and **Aydin**, with whom I have developed a beautiful friendship. Thank you for your professional advice and for sharing your special moments with me. I extend my most profound appreciation to the DOGS (**Matteo**, **Max**, **Joeri**, **Johno**, and **Myrna**) for organizing activities for doctoral students and imparting wisdom about doctoral life. And to my colleagues in Geology, Geophysics, and Reservoir Engineering: **Andrea**, **Jasper**, **Mahmoud**, **Parvin**, **Sverre**, **Nazife**, **Sepideh**, **Elahe**, **Yuan**, **Camille**, **Alesso**, **Emilio**, **Shihao**, **Amin**, **Ilshat**, **Fardad**, **Willemijn**, **Kiarash**, and **Kishan**. Thank you for being so special and for the pleasant coffee chats.

I am incredibly grateful to **Marlijn**, **Ingrid**, and **Ralf** for their assistance with the logistics of the doctorate and for helping me practice my Dutch. Your efforts made my life significantly easier amidst this journey's bureaucracy and organizational aspects.

Finally, my deepest gratitude goes to my family. My father **Alexander Diab Rincón** and my mother **Sara E. Montero Martínez**, your emotional support and encouragement have been the foundation of my achievements. Reaching this point in my life would have been impossible without all your sacrifices to give me the best. I thank you for never denying me any opportunity, protecting and advising me, and giving me your strength even from a distance. You know that this has been a great effort for everyone, especially in these last years, and I hope you know that each prayer and effort has been part of writing each of these letters to achieve this dream. Thanks also to my grandmothers,

**Esther M. Martínez Betancourt** and **Liduvina Rincón de Kamar** for their prayers, and to my cousin **Angélica García**, for your advice on how to keep my Colombian spirit alive in Europe, it has been invaluable.

# RECONOCIMIENTOS

Esta disertación fue parte del proyecto 'InFocus: Un Enfoque Integrado para Estimar la Ocurrencia de Deslizamiento de Fallas' (número de subvención: DEEP.NL.2018.037), financiado por el programa DeepNL del Consejo de Investigación Holandés (NWO), que tiene como objetivo mejorar la comprensión fundamental de la dinámica del subsuelo bajo la influencia de intervenciones humanas. Uno de los aspectos que más me gusta de mi nombre (**Hamed**) es que uno de sus significados es *agradecido*. Sin la ayuda y el apoyo de muchos de mis amigos, colegas, superiores y familia, habría sido imposible lograr esta disertación. Quisiera agradecer a muchas de esas personas que ayudaron directa o indirectamente a alcanzar este hito en mi vida.

Mi sincera gratitud va dirigida primeramente a mis supervisores más cercanos. A **Femke Vossepoel**, gracias por su apoyo durante todo el proceso del doctorado. He aprendido mucho de ti durante estos últimos años. Creo que una de las lecciones que más he intentado aplicar en mi carrera a tu lado es la de "menos es más". Creo que he aprendido a aplicar este principio cada vez más con el pasar del tiempo, pero intentaré aplicarlo más en mi vida profesional. Muchas gracias por tus consejos en la escritura, tu apoyo durante el avance del proyecto y el mantener la motivación para llevar a un feliz término este trabajo. Siento gran admiración por cómo lograste balancear tu vida profesional, personal y también tu pasión por el deporte. Espero que tengas mucho más éxitos en todos estos ámbitos y sigas cosechando triunfos. También agradezco a **Ylona van Dinther** por su apoyo para el desarrollo de este proyecto. Tu rigurosidad en la investigación y en el proyecto fueron un gran balance con la libertad que me ofreció la perspectiva de **Femke** para convertirme en un investigador independiente. Muchas gracias por las herramientas de trabajo y el conocimiento que nos proveíste en los aspectos sismológicos. Me alegra mucho que tu salud esté mucho mejor, y también por todos los logros e hitos académicos que has logrado durante tu carrera a pesar de los obstáculos. Te deseo lo mejor para tu carrera profesional y que descubras mucho más en el mundo de la sismología por la que sé que tienes una gran pasión.

La orientación que recibí de mi promotor, **Jan Dirk Jansen**, fue también muy importante, especialmente durante nuestras reuniones anuales, para proporcionar dirección y propósito a esta disertación. Agradezco mucho tus consejos y, especialmente, tu perspectiva para la toma de decisiones y el ser más pragmático en alcanzar nuestros objetivos. Agradezco también la atención prestada, dada la carga que presentó también el tener la posición de decano. Espero también que en los próximos años puedas disfrutar más tranquilamente de la investigación.

Asimismo, estoy muy agradecido con mis compañeros del proyecto InFocus. Muchas gracias a mi colega **Meng Li** por su asistencia en el proyecto con los modelos numéricos y por ayudarme a comprender de una manera más práctica temas como la ley de fricción rate-and-state. Estoy especialmente agradecido por tu disposición a aprender de la asimilación de datos y acompañarme de cerca en la integración de los resultados

de las simulaciones numéricas en el marco de asimilación de datos. En esta dirección, también quiero agradecer a **Casper** que nos permitió usar la biblioteca GARNET, fundamental para nuestra modelización numérica de nuestras secuencias de terremotos y a **Mohsen** quien nos acompañó en la expansión de las capacidades de modelación física.

En InFocus, contamos con el apoyo de expertos en las redes sísmicas de los Países Bajos, como lo es **Elmer Ruigrok**. Elmer, muchas gracias por todo tu conocimiento, y por ayudarnos a **Femke** y a mí con la cosupervisión de los proyectos de graduación de **Roos Verwijs** y **Julia van Deventer** en el uso de aprendizaje automático para el procesamiento rápido de mediciones sísmicas. También, contamos con el apoyo de **André Niemeijer** y **Yuntao Ji** con conocimientos y datos de laboratorio para probar los conceptos de asimilación de datos que desarrollamos en información adquirida de equipos de última generación. Muchas gracias por sus aportes y su colaboración para poder entender mejor el trabajo hecho en laboratorio y guiarnos en darle aplicabilidad a nuestros hallazgos.

En el ámbito de asimilación de datos aprendí mucho tanto de **Femke** como de un gran grupo de colegas tanto en TU Delft como en otras instituciones alrededor del mundo. Gracias a **Martin Verlaan** y al grupo de investigación de asimilación de datos del departamento de matemáticas (**Arnold, Avelon, Amey, Kian, Henrique, Andrés, Jeffrey, y Elías**). También gracias a nuestro grupo de investigación de asimilación de datos en el departamento de Geociencias e Ingeniería (**Gabriel, Mohsan, Arundhuti, Siamak, Dieter, Max Ramgraber, Samantha, Celine, Liang, y Jenny**). Fue un placer aprender a su lado y compartir con ustedes de nuestras discusiones. Muchas gracias también por atender a mis múltiples preguntas y expandir mi conocimiento en esta área tan fascinante del conocimiento.

También en la dirección de la asimilación de datos quiero agradecer a **Peter Jan van Leeuwen**, por el tiempo que pasé en Fort Collins, en la Universidad de Colorado. Muchas gracias por brindarme la oportunidad de visitarte y asistir a tus clases de asimilación de datos. El ver cómo se hacía investigación en un instituto meteorológico como CIRA, conocer del trabajo de **Chih-Chi, Michael, Michelle y Steven**, así como nuestras sesiones de discusión en el tablero de tu oficina fueron muy importantes para mí para entender mejor los métodos de asimilación de datos. Alguien también muy importante fue **Andreas Stordal**. Muchas gracias por todo tu apoyo en Bergen y en tus visitas a Delft para entender mejor cómo aprovechar las capacidades de los métodos de asimilación de datos en tus proyectos. Gracias por tu paciencia y motivación, han dejado un impacto duradero. Te deseo muchos éxitos en tus eventos deportivos así como en los diferentes frentes deportivos. Los talleres de EnKF en Bergen fueron muy agradables gracias a **Geir, Laurent, Remus, Rolf, Patrick, Xiaodong, Feda, Lars y Michael**. Gracias a todos muchachos también por esas conversaciones, tomando una cerveza, tanto de la vida como de asimilación de datos, fueron instrumentales para avanzar en mi comprensión y resolver muchas dudas.

En el departamento de geociencias e ingeniería también tuve un gran apoyo para sortear varios obstáculos en el camino. Un agradecimiento especial a **Anne Pluymakers** por nuestras conversaciones acerca de los desafíos encontrados en experimentos de laboratorio relacionados con la sismicidad y visualizar cómo la asimilación de datos podría ayudar con ciertas dificultades. Asimismo, el apoyo de **Evert Slob** durante los últimos

seis años ha sido un pilar durante mi tiempo en la familia de Geofísica Aplicada de la IDEA League. Sus perspectivas sobre educación, ciencia, investigación y vida me han inspirado, y espero que su sabiduría continúe iluminando a las futuras generaciones. También, estoy muy agradecido con **Jan Willem Thorbecke** por su asistencia con los desafíos de la supercomputación y a **Liliana Vargas** por su asesoramiento en el almacenamiento, manejo y tratamiento de los datos de nuestra investigación.

Al ser parte del proyecto DeepNL tuve la oportunidad de trabajar con grandes colegas en TU Delft, entre ellos **Aukje**, **Sara**, **Faezeh**, **Aleks**, **Iban**, **Milad**, **Jingming** y **Eddy**. Gracias por todas las conversaciones que tuvimos sobre geomecánica, modelización numérica, experimentos de laboratorio e inversión. Fue un placer pasar tiempo con ustedes en la oficina, preparando artículos científicos, presentaciones, abstracts y demás. Fue un placer continuar el camino al doctorado después de la maestría con **Iban**, **Aukje**, **Menno** y **Marat**. Un agradecimiento especial a **Milad** por ser un gran amigo y apoyarme tanto en mis mejores como en mis peores momentos. Espero de todo corazón que tengas mucho éxito en todos tus proyectos profesionales así como en tu vida personal. Gracias también a **Jingming** por nuestras cenas y a **Samantha** por animarme a participar en deportes y esforzarme constantemente por mejorar.

Asimismo quiero darles las gracias a **David**, **Maria**, **Rigoberto**, **Billy**, **Eddy** y **Florencia**, por ser como el apoyo de una familia durante estos años, especialmente al estar lejos de casa. Gracias por haber sido una gran fuente de apoyo y estar a mi lado tanto en las altas como en las bajas. Ha sido un placer compartir viajes, cenas, bailes, películas y muchas más actividades. Espero que nuestra amistad siga prosperando y también verlos triunfar en cada proyecto que emprendan. Muchas bendiciones para cada una de sus familias. Realmente me hicieron sentir comprendido y como en casa. Algunos de mis otros amigos latinos a agradecer son **Valentina**, **Luisa Orozco**, **Diana**, **Valeria**, **Rogelio**, **Linnaea**, **Daniel Aramburo**, y todo el grupo de colombianos en Delft, quienes también me han ayudado a sortear mi camino en los Países Bajos.

Aun desde la distancia también han sido muy importantes el apoyo de **Sofia Michail**, **Ana Pardo**, **Nilgün Güdük**, y **Dorde Petrovic**, gracias por abrirme las puertas de sus hogares en Europa y por convertirse en amigos invaluables. Gracias por las conversaciones acerca de la vida, por hacerme ver mis errores pero también por apoyarme genuinamente durante mis éxitos. Sus consejos, su paciencia, su tiempo y su apoyo realmente fueron una gran fortaleza para sortear las dificultades y ver un futuro más feliz e iluminado. Les deseo muchas bendiciones en sus vidas, a sus padres y hermanos. Estos deseos también los extiendo tanto a **Maria Carrizo** y a su familia, como a mis amigos en Colombia **Ana María Toro**, **Alina Vásquez** y **Lina Arrieta** quienes brindaron un inmenso apoyo, especialmente durante mis visitas a casa.

Así también extiendo mi gratitud a **Christian Reinicke**, **Rutger van Limborgh**, **Entela**, **Zhenja** y **Aydin** con quienes he desarrollado una bonita amistad. Gracias por sus consejos profesionales y por compartir conmigo en sus fechas especiales. Extiendo mi más profunda apreciación a los DOGS (**Matteo**, **Max**, **Joeri**, **Johno**, y **Myrna**) por organizar actividades para estudiantes de doctorado e impartir sabiduría sobre la vida de doctorado. Y a mis colegas de Geología, Geofísica y Ingeniería de Reservorios: **Andrea**, **Jasper**, **Mahmoud**, **Parvin**, **Sverre**, **Nazife**, **Sepideh**, **Elahe**, **Yuan**, **Camille**, **Alesso**, **Emilio**, **Shihao**, **Amin**, **Ilshat**, **Fardad**, **Willemijn**, **Kiarash**, y **Kishan**. Gracias por ser tan

especiales y por las agradables charlas de café.

Estoy extremadamente agradecido a **Marlijn**, **Ingrid** y **Ralf** por su asistencia con la logística del doctorado y por ayudarme a practicar mi holandés. Sus esfuerzos hicieron mi vida significativamente más fácil en medio de la burocracia y los aspectos organizativos de este viaje.

Finalmente, mi más profunda gratitud va para mi familia. Mi padre **Alexander Diab Rincón** y mi madre **Sara E. Montero Martínez**, su apoyo emocional y aliento han sido la base de mis logros. El haber llegado hasta este punto de mi vida habría sido imposible sin todos los sacrificios que han hecho en sus vidas para darme lo mejor. Les agradezco por no haberme negado nunca ninguna oportunidad, por protegerme, aconsejarme y por darme su fortaleza aun desde la distancia. Saben que este ha sido un gran esfuerzo de todos, especialmente en estos últimos años, y espero que sepan que cada oración y esfuerzo ha hecho parte de la escritura de cada una de estas letras para lograr este sueño. Gracias también a mis abuelas, **Esther M. Martínez Betancourt** y **Liduvina Rincón de Kamar** por sus oraciones, y a mi prima **Angélica García**, por tus consejos sobre cómo mantener vivo mi espíritu colombiano en Europa, ha sido invaluable.

# CURRICULUM VITÆ

## Hamed Ali DIAB MONTERO

04-09-1992 Born in Valledupar, Colombia.

## EDUCATION

2019-2023 PhD. Reservoir Engineering
TU Delft, Delft, the Netherlands
*Thesis:* Ensemble data assimilation methods for estimating fault slip and future earthquake occurrences
*Promotor:* Prof. dr. ir. Jan-Dirk Jansen
*Co-promotor:* Dr. ir. Femke C Vossepoel

2017–2019 M.Sc. IDEA League in Applied Geophysics
TU Delft, Delft, the Netherlands
ETH Zürich, Zürich, Switzerland
RWTH Aachen, Aachen, Germany
Distinction: *Cum Laude*

2011–2016 B.Sc. in Earth Sciences
Universidad de los Andes, Bogota, Colombia

2010–2014 B.Sc. in Civil Engineering
Universidad de los Andes, Bogota, Colombia
Distinction: *Cum Laude*

2004–2009 Secondary school
Gimnasio del Norte, Valledupar, Colombia

## EXCHANGE & RESEARCH VISITS

2023 Research visit in Data Assimilation
Colorado State University, Colorado, USA

2015 Exchange semester in Applied Geophysics
The University of Oklahoma, Oklahoma, USA

## INTERNSHIPS & WORK EXPERIENCE

2018 Internship, **ETH Zürich, Switzerland**
**Project:** Influence of mechanical deformation on flow and transport processes in discontinuities

2016-2017 Professional Assistant Engineer, **INGETEC SA, Colombia**
**Projects:** Engineering services for BEC Spillway design, Gatun, Colon - Panamá, and Optimization of Guaquivilo Melado hydroelectric project designs – Chile

2015 B.Sc. project, **Ecopetrol SA, Hocol SA & Universidad de los Andes, Colombia**
**Project:** Methodology for the acquisition of capillary curves form Nuclear Magnetic Resonance (NMR) logs in wells with heavy oil presence, Los Llanos Orientales Basin

## HONORS & AWARDS

2018-2019 Dean's list of top 5% of students of RWTH Aachen

2017-2019 Shell comprehensive scholarship of the IDEA League Applied Geophysics joint collaboration program

2015 Second place in the Latin America & Caribbean region Imperial Barrel Award (IBA) competition

2015 Ramón de Zubiria distinction for the highest GPA in the Earth Science department of Universidad de los Andes

2014 Excellence distinction for the highest GPA in the Earth Science department of Universidad de los Andes

# LIST OF PUBLICATIONS

## JOURNAL PUBLICATIONS

3. **Diab-Montero, H.A.**, Li, M., van Dinther, Y. and Vossepoel, F.C.(2024) *Estimating the Occurrence of Fault-slip in Large Scale Frictional Experiments.* In preparation.

2. **Diab-Montero, H.A.** , Stordal, A., van Leeuwen, P.J. and Vossepoel, F.C. (2024) *Ensemble Kalman, Adaptive Gaussian Mixture, and Particle Flow Filters for Optimized Earthquake Forecasting.* In preparation.

1. **Diab-Montero, H.A.**, Li, M., van Dinther, Y. and Vossepoel, F.C. (2023) *Estimating the Occurrence of Slow Slip Events and Earthquakes with an Ensemble Kalman Filter.* Geophysical Journal International, Volume 234, Issue 3, September 2023, Pages 1701–1721.

## CONFERENCE ABSTRACTS

7. **Diab-Montero, H. A.** , Hu, C.C., and Vossepoel, F. C. (2023, May). *A Particle Flow Filter for Estimating Future Earthquake Occurrences. Poster presentation.* (EnKF Workshop 2023)

6. **Diab-Montero, H. A.**, Li, M., van Dinther, Y., and Vossepoel, F. C. (2022, June). *Estimating Future Earthquakes with Data Assimilation: 1D and 2D Model Experiments of a Laboratory Setup. Poster presentation.* (EnKF Workshop 2022)

5. **Diab-Montero, H. A.**, Li, M., van Dinther, Y., and Vossepoel, F. C. (2022, May). *An Ensemble Kalman Filter for Estimating Future Slow Slip Events and Earthquakes on 1D, 2D and 3D Synthetic Experiments.* In EGU General Assembly Conference Abstracts (pp. EGU22-10324).

4. **Diab-Montero, H.A.**, M. Li, van Dinther, Y., Vossepoel, F.C., *Improving Earthquake Occurrences Estimation using an Ensemble Kalman Filter: a Synthetic Experiment.* The International EnKF Workshop 2021 (EnKF Workshop 2021)

3. **Diab-Montero, H. A.**, Li, M., van Dinther, Y., and Vossepoel, F.C. (2021, April). *Ensemble Kalman Filter estimates shear stress and seismic slip rates in synthetic laboratory experiment.* In EGU General Assembly Conference Abstracts (pp. EGU21-6836).

2. **Diab-Montero, H. A.**, Li, M., van Dinther, Y., and Vossepoel, F. C. (2021, April). *Estimation of earthquake occurrences with an Ensemble Kalman Filter: a synthetic experiment.* Nederlands Aardwetenschappelijk Congres (NAC)

1. **Diab-Montero, H.A.**, Li, M., van Dinther Y., and Vossepoel F.C. (2020, March). *Exploring Ensemble Data Assimilation methods for induced seismicity in the Groningen gas Field: framework validation with a perfect model setup* Nederlands Aardwetenschappelijk Congres (NAC).