IDEA League

Master of Science in Applied Geophysics

Research Thesis

Fault state estimation in subduction zones using a particle filter with time-lagged particle generation

Shiran Levy

August 28, 2018

Fault state estimation in subduction zones using a particle filter with time-lagged particle generation

Master of Science Thesis

for the degree of Master of Science in Applied Geophysics at
Delft University of Technology
ETH Zürich
RWTH Aachen University
by

Shiran Levy

August 28, 2018

Department of Geoscience & Engineering · Delft University of Technology
Department of Earth Sciences · ETH Zürich
Faculty of Georesources and Material Engineering · RWTH Aachen University

Delft University of Technology

Printed in The Netherlands, Switzerland, Germany

IDEA LEAGUE
JOINT MASTER'S IN APPLIED GEOPHYSICS

Delft University of Technology, The Netherlands
ETH Zürich, Switzerland
RWTH Aachen, Germany

Dated: *August 28, 2018*

Committee Members:

Dr. lr. Femke Vossepoel

Dr. Marie Bocher

Dr. Ylona van Dinther

Prof. Dr. Florian Wellmann

Dr. Deyan Draganov

Supervisor(s):

Dr. lr. Femke Vossepoel

Dr. Marie Bocher

Dr. Ylona van Dinther

Abstract

Data assimilaiton, a procedure in which observed data is combined with prior knowledge, is widely used in geophysical systems and especially popular in atmospheric and oceanic models. In this study a Monte Carlo based data assimilation method referred as a bootstrap Particle Filter (PF) and a time lag sampling technique are combined together to perform Sequential Data Assimilation (SDA) of borehole observation into a Seismo-Thermo-Mechanical model (STM). The aim of this study is to estimate the state of faults in subduction zones. The STM, a strongly non-linear model, is taken to be a perfect model and serves as a source for both observed data and model realizations termed "particles" or ensemble members. The ensemble is being generated by drawing particles out of a seismic cycle with a constant time lag. Results demonstrate that assimilation strongly depends on the choice of time lag since, small time lags provided with better results. Changing the time lag for sampling leads to a trade off between ensemble spread and resolution due to presence of trends in some of the observed state variables. Although the sampling technique in its current setup is computationally efficient, it was found to be insufficient in representing the model errors. Comparison between the current study of the PF and recent work involving the Ensemble Kalman Filter (EnKF) suggests that the success of the EnKF is related to its error covariance matrix correlating the various state variables. Based on the results and comparison to the EnKF improvements and possible next steps are discussed.

Acknowledgements

I would like to thank my main supervisors Dr. lr. Femke Vossepoel and Dr. Marie Bocher for their guidance, patience and motivation through the process. You made this project interesting and enjoyable and it was a pleasure to work with you during this period. I want to thank Dr. Ylona van Dinther for the effort and support in the last stage of the thesis. I would also like to thank Dr. Deyan Draganov for his help with bureaucratic issues and for kindly volunteering to read my thesis and to be in the examiners committee.

Lastly, I would like to thank my boyfriend Matteo Menichell and my family for their support through this Master program.

Delft, University of Technology
Zürich, ETH

Shiran Levy
August 28, 2018

Contents

Abstract v

Acknowledgements vii

Acronyms xvii

1 Introduction 1

I First Part 3

2 Theory 5
2-1 Forward model: Seismo-Thermo-Mechanical model . . . 6
2-2 Sequential data assimilation . . . 8
2-2-1 Particle Filter . . . 9
2-2-2 Ensemble Kalman Filter (EnKF) . . . 11
2-2-3 Ensemble smoother . . . 13

3 Methods 15
3-1 STM model: setup and output . . . 15
3-2 Particle Filter . . . 16
3-2-1 Ensemble generation . . . 17
3-2-2 Assimilation: calculating weights and estimation . . . 17
3-3 Particle Smoother . . . 21

II Second Part 23

4 Ensemble generation 25
4-1 Early analysis and representation of process physics . . . 25
4-2 Particle separation . . . 28
4-3 Ensemble size . . . 29

5 Results and analysis **33**
5-1 Particle Filter 33
5-2 Trend effect 39
5-3 Ensemble Smoother 42

6 Discussion and conclusions **45**

References **49**

A Ensemble generation **53**

B Results and analysis **55**

C Statistics calculation **60**

List of Figures

2-1 (a) Laboratory setting simulating the seismicity along a subduction megathrust . A visco-elastic gelatin wedge with dimensions of 60 X 11 X 34 cm3 representing a 380 km long and 70 km deep section of forearc lithosphere rotated to have a dip of 10 degrees [Corbi et al., 2013]. A backstop on the rear part of the gelatin holds it while the plate is driven in the direction of the green arrow at a constant velocity. (b) Numerical version of the analog model. A gravity acceleration of 9.81 ms^{-2} is applied on the gelatin surface with respect to a 10 degrees dip angle. Grey and red colors in (a) and (b) under the gelatin wedge represent the aseismic and seismogenic zones of the thrust fault. Adapted from van Dinther et al. (2013b). 7

2-2 Illustration of a tracking problem, possibly nonlinear with observed data d, hidden state ψ and connecting them are H and F operators. 9

3-1 Model domain consisting of 701x136 nodes. Properties differ inside the domain according to the material: Air (white), Gelatin (orange), Fault (red- seismic and gray- aseismic zones), backstop wall (magenta). Dots represent markers of interest: GPS markers (light blue) borhole location (black) and Fault markers (blue). . . . 16

3-2 Visualization of the time lagged particle sampling technique discussed. Each colored line represents a time lagged version of the seismic cycle evolution in blue. . 18

3-3 Illustration of assimilation step with the true evolution (brown line) and the true state indicated at ~ 141 seconds by the red bullet. The ensemble of particles sampled from a separate evolution (blue line) is indicated by the small colored bullets. In order to use the true state as observed data, errors are added to each assimilated state with a corresponding error size. 19

3-4 Illustration of an ensemble in subsequent five assimilation steps (a-e) for the state variable V_i. Colored bullets are the particles constituting the ensemble. 20

3-5 Illustration of an ensemble in subsequent five assimilation steps (a-e) for the state variable σ_{ii}. Colored bullets are the particles constituting the ensemble. 20

3-6 Comparison between Lorenz and Gaussian distributions of particle weights for the same ensemble of particles. 21

4-1 Particles sampled from evolution of v_i, v_j, σ_{ii}, σ_{ij} and P with a constant time lag Δt. 26

4-2 Fit of expected value to true evolution for GPS marker S1. Sampling particles forwards in time with $\Delta t = 6.63s$ . . . 27
4-3 Fit of expected value to true evolution for GPS marker S1. Sampling particles Backward and forward in time with $\Delta t = 6.63s$ . . . 27
4-4 Frequency spectrum of (a) v_i and (b) P evolutions in time. Time window for FFT is $\sim$ 144 -1989 seconds. . . . 28
4-5 My short caption . . . 31
4-6 My short caption . . . 32

5-1 Data fit at the borehole location for experiment with 300 particles and $\Delta t = 9.9$ s. Mean absolute error v_i: 0.0032 ± 0.0065 [cm/s], v_j: 0.0023 ± 0.0044 [cm/s], σ_{ii}: 6.08 ± 3.71 [Pa], σ_{ij}: 2.10 ± 1.71 [Pa] and P: 3.12 ± 3.97 [Pa]. . . . 34
5-2 Fit of expected and true evolutions in F11 marker, for experiment with 300 particles and $\Delta t = 9.9$ s. Mean absolute error v_i: 0.0040 ± 0.0076 [cm/s], v_j: $7.72 \cdot 10^{-5} \pm 1.90 \cdot 0.00018$ [cm/s], σ_{ii}: 12.41 ± 13.87 [Pa], σ_{ij}: 14.43 ± 16.71 [Pa] and P: 9.58 ± 9.83 [Pa]. . . . 34
5-3 Data fit at the borehole location for experiment with 300 particles and $\Delta t = 0.33$ s. 35
5-4 Fit of expected and true evolutions in F11 marker, for experiment with 300 particles and $\Delta t = 0.33$ s. . . . 35
5-5 The mean absolute error between true and expected evolutions resulting from assimilating 300 particles that were sampled with time lag $\Delta t = 0.33$. The MAE and corresponding SD are displayed for few chosen markers on the fault and for the borehole. F10, F11 and F12 markers situated within the seismogenic zone. . 36
5-6 Weight contribution of each state separately at the borehole location for assimilation step marked by brown in figure 5-4. Particle marked in purple, is the particle with the highest weight of combined states w_i. . . . 38
5-7 Combined posterior (w_i) with which particles are weighted at fault marker F11, for assimilation step marked by brown in figure 5-4. . . . 38
5-8 Comparison of partitioned weight distributions of particles sampled at the borehole location with $\Delta t = 9.9s$ and $\Delta t = 0.33s$. . . . 39
5-9 Posterior distribution of particles sampled with $\Delta t = 0.33s$ at the borehole location for assimilation step 80. Particles are colored by their order relative time in the evolution, from blue (early) to yellow (later). . . . 40
5-10 Posterior distribution of particles sampled with $\Delta t = 9.9s$ at the borehole location for assimilation step 80. Particles are colored by their order relative time in the evolution, from blue (early) to yellow (later). . . . 41
5-11 Particles corresponding to the F11 marker location and their attached weights as calculated at the borehole location (fig.5-10). Particles are colored by their order of sampling in the evolution, from blue (early) to yellow (late). . . . 41
5-12 Evolution of Particles' weights with data assimilation for (a) for PF and $\Delta t = 0.33$ s (b) PS and $\Delta t = 0.33$ s (c) PF and $\Delta t = 9.9$ s (d) PS and $\Delta t = 9.9$ s. The size of the bullet correspond to its weight. . . . 43

A-1 Ensemble spread at different assimilation steps with $\Delta t = 0.33$ s. Red dot is the true state, blue dot is the observation (true state+error) and the two black dots connected by a bar is the variance of the measurement error. σ_{ii} could be approximated as a Gaussian however the mean of is not close to might not be close to the observation or the true state. P does not exhibit a Gaussian prior, therefore a mean would not target the middle of the distribution as expected in a Gaussian distribution. . . . 54

B-1 Fit of expected and true evolutions in F9 marker, for experiment with 300 particles and $\Delta t = 9.9$ s. 55
B-2 Fit of expected and true evolutions in F14 marker, for experiment with 300 particles and $\Delta t = 9.9$ s. This location on the fault consistently exhibit better results than the rest of the Marker locations. 56
B-3 Fit of expected and true evolutions in F9 marker, for experiment with 300 particles and $\Delta t = 0.33$ s. 56
B-4 Fit of expected and true evolutions in F14 marker, for experiment with 300 particles and $\Delta t = 0.33$ s. Although assimilation shows an improvement, the envelope of particles trajectory which could also be regarded as an approximated error distribution of the model, does not cover well the true evolution throughout the assimilation window. 57
B-5 Weight contribution of each state separately in the borehole for assimilation step marked by green in figure 5-4. Particle marked in purple, is the particle with the highest weight of combined states w_i. 58
B-6 Combined posterior (w_i) with which particles are weighted, for assimilation step marked by green in figure 5-4. 58
B-7 Weight contribution of each state separately in the borehole for assimilation step marked by purple in figure 5-4. Particle marked in purple, is the particle with the highest weight of combined states w_i. 59
B-8 Combined posterior (w_i) with which particles are weighted, for assimilation step marked by purple in figure 5-4. 59

List of Tables

5-1 Relative errors corresponding to previous figure (5-5) showing errors resulting from assimilating 300 particles that were sampled with time lag $\Delta t = 0.33$ 36

Acronyms

PF Particle Filter

EnKF Ensemble Kalman Filter

DA Data Assimilation

KF Kalman Filter

EKF Extended Kalman Filter

ES Ensemble Smoother

LAF Lagged Average Forecast

STM Seismo-Thermo-Mechanical

ODF Ordinary Dynamical Forecast

MIC Marker In Cell

PIC Particle In Cell

MAC Marker And Cell

Chapter 1

Introduction

According to the National Geophysical Data Center (NGDC) online database, in the last decade 26 earthquakes occurred for which the number of deaths is above 100; these include the 2010 devastating earthquake in Haiti with a staggering 316,000 deaths. At least 10 out of the 26 earthquakes are situated at or near the margin of a subducting plate. In addition to death toll, earthquakes and the tsunamis they trigger, inflict a heavy price on the economy, making prediction and forecasting of earthquakes an important tool.

While earthquake prediction is considered by some impossible (due to non-linearity and sensitivity to earth's state; Geller, 1997) and therefore mainly based on precursors including seismic velocity, electric conductivity and hydrological changes, radon emission and electromagnetic signals, earthquake forecasting includes a variety of operational methods [Jordan et al., 2011]. Forecasting involves a probabilistic estimation of earthquake occurrence and each method rely on different assumptions (time-dependent\independent) and models [Jordan et al., 2011]. Forecasting is divided into long-term (years or more) and short-term (months or less). Short term forecasting concerns mainly aftershocks triggered by large earthquakes, while long term focuses on earthquake damage and plays a significant role in risk management and risk reducing engineering [Jordan et al., 2011]. Since forecasting mainly relies on physical models, knowledge about rupture processes and the physical state of faults can possibly lead to better earthquakes forecasting ability [Jordan et al., 2011; van Dinther et al., 2017]. Various models exist today, address different aspects of earthquake dynamics, from small scale rupture models to large scale and long term deformation [van Dinther, 2013]. One type of model focusing on the seismic cycle in subductiong zones is the STM model developed by van Dinther et al. (2013b). This state of the art model combines short term rupture dynamics with long term deformation. The STM model was later used by van Dinther et al., 2017 to perform Sequential Data Assimilation (SDA) using the Ensemble Kalman Filter (EnKF); their work served as a proof of concept to the ability of estimating and forecasting the state of a fault in subduction zones and results were remarkably good. Nevertheless, the underlying assumption for using the EnKF is the Gaussianity of the probability densities involved. As an alternative to the EnKF, I suggest a fully non-linear bootstrap particle filter with a unique sampling method.

Particle Filters have a wide range of applications, such as parameter and state estimation for dynamic models (e.g. atmosphere and ocean models; Vossepoel & van Leeuwen, 2007, van Leeuwen, 2010; Liu & West, 2001; Stroud et al., 2018), navigation, positioning and tracking [Gustafsson et al., 2001], robotics [Thrun, 2002], subsidence estimation and reservoir characterization [Vossepoe et al., 2017; Beers, 2018], ground water flow modeling [Hendricks Franssen & and Kinzelbach, 2008] etc. Particle filtering combines a discrete representation of the model space distribution with observed data in a Bayesian frame of work. The major advantage of the bootstrap filter (or generic particle filter) is its ability to assimilate highly non-linear and in some cases high dimensionality problems. Nonetheless, specifying a prior could be a difficult and sometimes impossible task if the prior distribution is too complex.

In 1983, Hoffman & Kalnay proposed the Lagged Average Forecast (LAF) technique which is a method based on time lagged versions of the physical model referred to as the ordinary dynamical forecast (ODF). This method was originally designed as a forecasting method in weather prediction, where ensemble members are sampled from an ODF at different time lags: $t_i = (i-1)\Delta t$, computed as different model realizations and averaged into a single forecast. This method proved to be superior to other methods such as the ODF, MCF (Monte Carlo Forecast) and PCF (Persistence Climatology Forecast).

In this thesis I combine the bootstrap particle filter together with a LAF based sampling technique, in order to assimilate observed data onto a seismic cycle model. Observations are generated using the STM Model at a predefined location corresponding to a borehole at a depth of approximately 5 km from the surface. The assimilated data consists of five state variables: horizontal velocity v_i, vertical velocity v_j, normal stress σ_{ii}, shear stress σ_{ij} and pressure P as would be observed from a borehole measurements. Particles are sampled with a time lag from a second evolution generated by the STM model. Performance of the PF will be evaluated at the borehole and at several other locations on the fault. If successful, the suggested sampling technique can serve as an efficient way to sample the search space.

Part I

First Part

Chapter 2

Theory

Contrary to traditional view, earthquakes are a frictional phenomenon and rarely occur by the emergence of a new fault [Scholz, 1998]. The relative motion between adjacent tectonic blocks builds up stress in their contact zone, eventually leading to the mechanical failure of the material, a short period of rupture and the release of accumulated elastic strain as seismic energy which is then expressed as ground movement known to us as an earthquake [Scholz, 1998; Corbi et al., 2013]. This period of material failure accompanied by a pressure drop and fast movement along the fault is termed "slip" (coseismic period) while the state of partial locking of the fault and elastic strain accumulation is termed "stick" (interseismic period; [Brace & Byerlee, 1966; van Dinther, 2013]).

There exist several approaches to model faulting and/or earthquakes: geodynamic, dynamic rupture, seismic cycle and analog models [see van Dinther, 2013 and references therein]; each computes the evolution of an earthquake-prone region at a different time and length scale, from hundreds of seconds over hundreds of kilometers in rupture dynamics to million years over thousands of kilometers in geodynamic modeling [van Dinther, 2013]. In order to express both long term deformation and short term rupture process characteristic of the earthquake cycle, van Dinther (2013) developed a Seismo-Thermo-Mechanical (STM) model based on the geodynamic model of Gerya & Yuen (2007) and a slip rate dependent friction [van Dinther et al., 2013b]. In their model, van Dinther et al., 2013b incorporated the principle of stick-slip mechanism [Brace & Byerlee in 1966] together with rate dependent friction, slow tectonic loading and visco-elastic stress relaxation in order to simulate the seismic cycle in subduction zones.

Van Dinther et al. (2017) introduced a procedure as a proof of concept, that combines the forward model of a fault slip computed by the STM model with the Ensemble Kalman Filter (EnKF) in order to assimilate data and obtain a better estimation of the system's hidden state. A possible alternative to the EnKF is the method discussed in this thesis: Particle Filtering. In the next subsections I describe the STM model and the two mentioned data assimilation methods.

2-1 Forward model: Seismo-Thermo-Mechanical model

The STM model (van Dinther, 2013) is a numerical model used to compute the evolution of the physical state of a 2D domain, using conservation equations and rheological constitutive equations with rate-dependent friction coefficient. In particular, the model can simulate the subduction of an oceanic slab under a continent [van Dinther et al., 2013a]. To account for long-term deformation processes van Dinther (2013) based her work on the I2ELVIS code: a continuum-mechanics-based dynamical model for visco-elasto-plastic material introduced by Gerya & Yuen (2007) and extended it with a slip-rate-dependent friction.

The I2ELVIS code uses a combination of conservative, implicit finite-difference scheme and a non diffusive characteristics-based marker-in-cell (MIC) technique on a staggered grid [Gerya & Yuen, 2007]. The MIC technique is a methodology that dates back to the MAC and PIC techniques of Harlow & Welch (1965) and Brackbill & Ruppel (1986). It enables tracking of complex flow features by markers discretization of grid cells in a Lagrangian framework [Duretz et al., 2011; Gerya & Yuen, 2007].

The equations of conservation of mass, momentum and energy for an incompressible material

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial y} = 0, \tag{2-1}$$

$$\frac{\partial \sigma'_{xx}}{\partial x} + \frac{\partial \sigma'_{xz}}{\partial z} - \frac{\partial P}{\partial x} = \rho \frac{Dv_x}{Dt}, \tag{2-2}$$

$$\frac{\partial \sigma'_{zx}}{\partial x} + \frac{\partial \sigma'_{zz}}{\partial z} - \frac{\partial P}{\partial z} = \rho \frac{Dv_z}{Dt} - \rho g, \tag{2-3}$$

where v_x and v_z are horizontal and vertical velocities, σ'_{ij} are the deviatoric stress tensor components, P is the pressure and $g = 9.81\ ms^{-2}$ is gravitational acceleration, are solved on an Eulerian grid, while material properties are advected according to the calculated velocity field using markers in a Lagrangian framework. The exchange of information between the two frameworks occurs by interpolation of material properties from the Eulerian nodes to Lagrangian markers and vice versa; this is especially useful for tracking non diffusive material properties like lithology and rheological parameters, even through sharply varying material properties [van Dinther et al., 2013b].

Van Dinther et al. (2013b) introduced an inertial term (density multiplied by the Lagrangian time derivative of the velocity) in eq. 2-2 and 2-3 to regularize high slip rates with decreasing time steps. Notice that since the incompressibility assumption is used, only shear waves are generated [van Dinther et al., 2013b]. The above conservation equations can be solved by using a non-linear visco-elasto-plastic relationship between the deviatoric stress and strain rate; ε'_{ij}

$$\dot{\varepsilon}'_{ij} = \frac{1}{2\eta}\sigma'_{ij} + \frac{1}{2G}\frac{D\sigma'_{ij}}{Dt} + \begin{cases} 0 & for\ \sigma'_{II} < \sigma_{yield} \\ \chi \frac{\partial g_{plastic}}{\partial \sigma'_{ij}} & for\ \sigma'_{II} = \sigma_{yield} \end{cases}, \tag{2-4}$$

where η is the effective viscosity, G is the shear modulus, χ is a plastic multiplier relating plastic strain rates to stresses, $g_{plastic}$ is the plastic flow potential and σ_{yield} is the plastic

strength of the material. The first, second and third terms on the right hand side of eq.2-4 are the viscous, elastic and plastic strains respectively. Apart from the addition of the inertial term, van Dinther et al. (2013b) incorporated a strongly rate dependent friction in the code. This key ingredient describes brittle instabilities with an effective steady state friction that is a function of the slip velocity

$$\mu = \mu_s(1-\gamma) + \mu_s \frac{\gamma}{1+\frac{V}{V_c}}, \tag{2-5}$$

where $\gamma = 1 - \frac{\mu_d}{\mu_s}$, μ_d and μ_s are the dynamic and static friction coefficients. V_c is the velocity at which friction occurrence reached its half. The above equations are only a small part of the full code. For an extensive description about the I2ELVIS and STM models, please refer to Gerya & Yuen (2007) and van Dinther et al. (2013b).

The setup of the numerical model can be seen in Figure 2-1.(b). In order to control the behavior of the topography in the upper boundary of the wedge (overriding plate), van Dinther et al. (2013) set a low viscosity body known as 'sticky air' to the upper part of the domain (white color in Figure 2-1; van Dinther et al., 2013a). Apart from GPS-recorded surface dis-

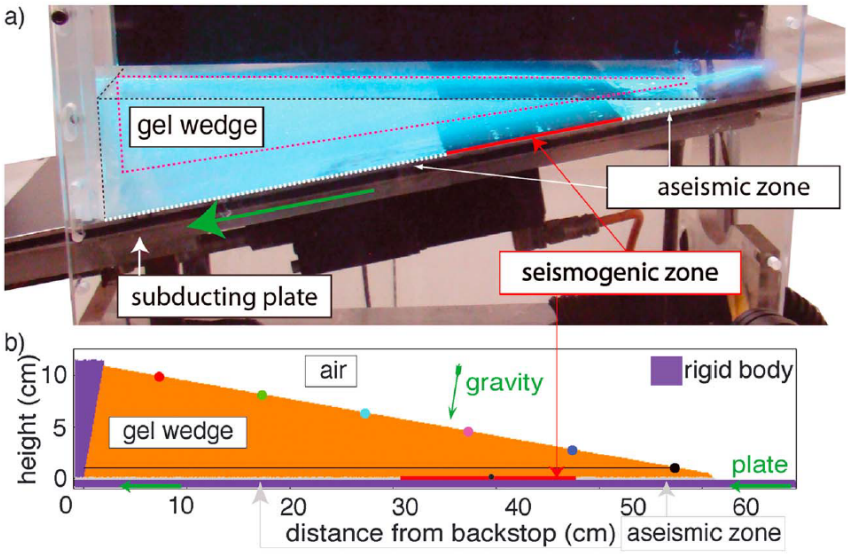

Figure 2-1: (a) Laboratory setting simulating the seismicity along a subduction megathrust . A visco-elastic gelatin wedge with dimensions of 60 X 11 X 34 cm3 representing a 380 km long and 70 km deep section of forearc lithosphere rotated to have a dip of 10 degrees [Corbi et al., 2013]. A backstop on the rear part of the gelatin holds it while the plate is driven in the direction of the green arrow at a constant velocity. (b) Numerical version of the analog model. A gravity acceleration of 9.81 ms^{-2} is applied on the gelatin surface with respect to a 10 degrees dip angle. Grey and red colors in (a) and (b) under the gelatin wedge represent the aseismic and seismogenic zones of the thrust fault. Adapted from van Dinther et al. (2013b).

placements and natural, seismological and field observations [van Dinther, 2013], van Dinther et al., 2013b also validated the STM model with an accompanying laboratory experiment (see Corbi et al., 2013). The analog model consists of a visco-elastic gelatin wedge that is placed on a flat plastic covered plate with sand paper on a portion of it. The contacts of

gelatin-plastic and gelatin-sand paper simulate aseismic and seismogenic zones (respectively) in the interface of a thrust fault (Figure 2-1; Corbi et al., 2013). Model parameters as well as time were scaled in the model to represent the process in nature [Corbi et al., 2013]. The time scale in the model varies throughout the evolution as the model was originally scaled such that 1 s of interseismic period in the model is equivalent to 286 yr in nature and 1 s of coseismic period in the model is equivalent to 800 s in nature.

2-2 Sequential data assimilation

Data assimilation (DA) is a general name for a variety of techniques integrating two sources of information: data and a model, of a physical process (either probabilistic or deterministic), with the aim of obtaining an estimation of the true state evolution of the system. Data and model spaces are inhabited by observations and mathematical/numerical model describing a physical process (respectively). DA is widely used in oceanography (e.g. ocean circulation models) and meteorology (e.g. numerical weather prediction models) and it includes three main approaches: variational DA, sequential DA or a mix of both [Lahoz et al., 2010; and Evensen, 2009]. While variational DA is looking for the best fitted estimate over the whole assimilation time window (therefore, require storage of an entire assimilation time window), sequential DA proceeds in a chronological order, by repeating steps of analysis when new observations are available and forecasts until the next observation time [Evensen, 2009; Lahoz et al., 2010; Bannister, 2017; Lang & Owens, 2018].

A tracking problem with state variable ψ and measurement data d both at discrete time k, can be defined as [Arulampalam et al., 2002],

$$\psi_k = F_k(\psi_{k-1}, v_{k-1}), \tag{2-6}$$

$$d_k = H_k(\psi_k, n_k), \tag{2-7}$$

for $k = 1, 2, 3...$. For a non linear problem, F_k and H_k are non linear functions relating ψ_{k-1} to ψ_k and ψ_k to d_k with v the process noise and n measurement noise. The process noise v associated with the model and initial conditions includes inaccuracies, numerical truncation and discretization while the measurement noise n associated with observed data includes random, systematic and representative errors. By combining observations and the model, we account for uncertainties and shortcomings of one and compensate it with the other to obtain improved estimates [Lahoz et al., 2010]. Integration of two information sources can be performed through Bayesian inference [Bayes & Price 1763], in which probability densities reflects the uncertainty in a value, may it be a model or observed data. $p(\psi_{0:k}|d_{1:k})$, the conditional probability known as the posterior, is

$$p(\psi_{0:k}|d_{1:k}) = \frac{p(d_{1:k}|\psi_{0:k})p(\psi_{0:k})}{p(d_{1:k})} \tag{2-8}$$

where $p(d_{1:k}|\psi_{0:k})$ and $p(\psi_{0:k})$ are the likelihood and prior distributions and $p(d_{1:k}) = \int p(d_{1:k}|\psi_{0:k})p(\psi_{0:k})d\psi_{0:k}$ is a normalizing factor [van Leeuwen, 2009; Särkkä, 2013]. The

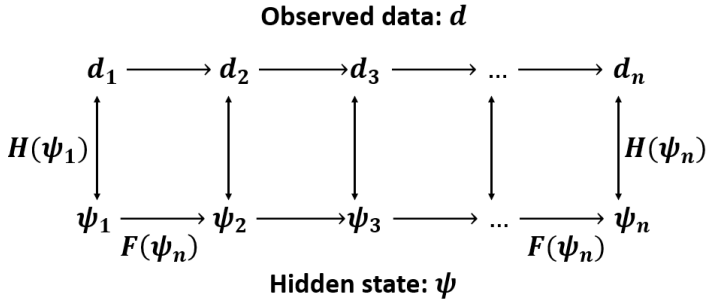

Figure 2-2: Illustration of a tracking problem, possibly nonlinear with observed data d, hidden state ψ and connecting them are H and F operators.

above equation calculates the full posterior distribution of the states for all available observations. As number of observation fed to the system increases, the calculation becomes computationally too expensive. In order to make calculation more simple and applicable, only marginal distributions are being calculated and distributions are approximated as: $p(\psi_{0:k}|d_{1:k}) \sim p(\psi_k|d_k)$, $p(d_{1:k}|\psi_{0:k}) \sim p(d_k|\psi_k)$ and $p(\psi_{0:k}) \sim p(\psi_k|\psi_{k-1})$ [Särkkä, 2013]. The latter approximation describes a transition probability distributions of a Markov chain process where ψ_k is independent of steps preceding k-1 and succeeding k. The observation vector d is assumed to be a discrete vector of measurements at $t \in [t_1, t_k]$. Recursive application of this Bayesian based approach consists of two main stages: prediction and updating at each assimilation step [Arulampalam et al., 2002]. Prediction consist of propagating the current system's state and its corresponding probability density function (pdf) in time to the next step and the update consist of modifying this predicted state using measured data and corresponding pdf [Arulampalam et al., 2002].

Within the filtering category there are optimal algorithms which compute a closed form solution such as the Kalman filter and suboptimal algorithms which approximate solutions such as the Extended Kalman Filter (EKF), the Ensemble Kalman Filter (EnKF) and Particle Filters [Arulampalam et al., 2002; Evensen, 2009; van Leeuwen, 2009; van Leeuwen, 2010; Särkkä, 2013]. The problems that are being solved by methods like KF (linear) and EKF (linearized problem) have normally distributed probability densities fully characterized by their mean and covariance. On the other hand, problems that are either non-linear or have non-Gaussian behavior, can be solved among other options using EnKF if Gaussianinity assumption hold or Particle Filter for fully non-linear and non-Gaussian distributions [Arulampalam et al., 2002].

2-2-1 Particle Filter

Particle Filtering is a recursive Bayesian filter and a type of sequential Monte Carlo method, in which model posterior is represented by a set of random particles, each of which is a different realization of the model [Arulampalam et al., 2002; Kitagawa, 1996]. Particle Filters are particularly useful in cases of non-linear problems with non-Gaussian state distributions and considerable dimensionality of ψ where various types of Bayesian filters such as Kalman filters cannot be applied due to their basic assumptions [Arulampalam et al., 2002]. In the most basic form of the Particle filter, particles are drawn randomly from a prior distribution

$$\left\{\psi_0^{(1)}, \psi_0^{(2)}, ..., \psi_0^{(s)}\right\} \sim p_0(\psi), \tag{2-9}$$

such that the model prior is represented by a sum of N_s delta functions

$$p(\psi) = \frac{1}{N_s}\sum_{i=1}^{N_s}\delta(\psi - \psi_i) = 1, \tag{2-10}$$

where δ is the Dirac function [van Leeuwen, 2009]. If the number of particles $N \to \infty$ the particle filter converges to the true distributions (under some conditions; Crisan & Doucet, 2002). By drawing samples out of a continuous distribution, the prior and eventually the posterior are being represented in a discrete manner such that closed form computation of statistical quantities are avoided and the number of computations required at each assimilation step remains constant [Särkkä, 2013]. The posterior distribution is determined through "importance sampling" where the posterior is approximated in the following way:

$$p(\psi|d) = \sum_{i=1}^{N_s} w_i\delta(\psi - \psi_i) \tag{2-11}$$

where w_i is a weighting function given by:

$$w_i = \frac{p(d|\psi_i)}{\sum\limits_{j=1}^{N_s} p(d|\psi_j)} \tag{2-12}$$

The expected value of state variable ψ would then be a weighted average of all particles:

$$E[\psi] = \frac{\sum\limits_{i=1}^{N_s} \psi_i p(d|\psi_i)}{\sum\limits_{j=1}^{N_s} p(d|\psi_j)} = \sum_{i=1}^{N_s} \psi_i w_i \tag{2-13}$$

Another type of importance sampling commonly discussed in literature uses an approximated distribution $\pi(\psi)$ (could also appear as $q(\psi)$) from which samples are drawn, such that w_i becomes an "importance weight" [Doucet et al.,2001]:

$$w_i \propto \frac{p(\psi_i)}{\pi(\psi_i)} \tag{2-14}$$

An approximated or proposed distribution is used when either $p(\psi)$ is too complex or the variance when sampling it is too large. By proposing a distribution, one can focus sampling to the important region of the distribution.

The likelihood $p(d|\psi_i)$ can be computed using a Gaussian distribution [Van Leeuwen, 2009]:

$$p(d|\psi_i) = Aexp\left\{-\frac{1}{2}[d - H(\psi_i)]^T(\sigma^2)^{-1}[d - H(\psi_i)]\right\} \tag{2-15}$$

or a Lorenz distribution [Vossepoel & Van Leeuwen, 2007]:

$$p(d|\psi_i) = \frac{1}{1 + \frac{[d-H(\psi_i)]^2}{\sigma^2}} \tag{2-16}$$

where A is an amplitude factor, $H(\psi_i)$ is the measurement operator acting on the i^{th} particle, d is the observation, $[d - H(\psi_i)]$ is called the innovation and σ^2 is the standard deviation. σ is a scalar or a Matrix depending on the number of variables being observed and it is proportional to observational errors. If the observational errors are correlated, σ becomes the error covariance as in van Leeuwen (2009). The Lorenz distribution has its shape similar to a Gaussian, though it is much broader away from its peak. This characteristics of the Lorenz distribution enables more particles to influence the estimate [Vossepoel & van Leeuwen, 2007] therefore, making it more favorable when the distribution is too narrow using a Gaussian. The final product of eq. 2-15 and 2-16 is a distribution of weights that are based on the distance of each particle to the observation. This importance sampling procedure is relatively easy to implement and it is repeating itself also in next analysis steps. Nonetheless, van Leeuwen (2009) describes two problems arising from the basic particle filter:

1) If particles are moving away from the observation, they would not be pulled back since only the weight of the particles is being modified at each analysis step and not the particles themselves.
2) Particle filters suffer from degeneracy, resulting in most particles having negligible weight after few analysis steps, turning the analysis to be statistically meaningless.

A useful way to measure a filter's degeneracy is to calculate an estimate of the effective sample size N_{eff} [Arulampalam et al., 2002],

$$N_{eff} \approx \frac{1}{\sum_{i=1}^{N_s} (w_i)^2} \tag{2-17}$$

where w_i was already introduced in eq. 2-12. To overcome degeneracy one could use very large ensemble or choose the importance density such that the variance can be decreased [Arulampalam et al., 2002]. Option 1 is impractical in most applications where computational effort will increase dramatically due to large data sets. Another option is to perform resampling: once N_{eff} reduces below a threshold value, low weight particles are eliminated from the ensemble and copies are made from high weight particles to rebuild the ensemble's original size [Arulampalam et al., 2002].

2-2-2 Ensemble Kalman Filter (EnKF)

This subsection is meant to be a short summary of the EnKF method according to how it was implemented in van Dinther et al. (2017). For a more comprehensive description the reader is referred to van Dinter et al. (2017) and Evensen (2009). The EnKF [Evensen, 1994] is a stochastic alternative of the MCMC (Markov chain-Monte Carlo) type for the Extended Kalman Filter (EKF) and it is suited for solving large dimensional non-linear problems. The EKF has two major drawbacks: 1) it uses a closure approximation scheme to the error

covariance evolution 2) it is impractical in high dimensional dynamical models due to the size of the error covariance matrices [Evensen, 2009]. The EnKF circumvent these by using an ensemble of model states. Instead of computing explicitly the evolution of the mean state and forecast error covariance $C^f_{\psi\psi}$, the EnKF computes the evolution of an ensemble of model states whose sample covariance matrix approximates the error covariance matrix [Evensen, 2009].

The two probability densities $p^f(\psi)$ (forecast states) and $p^a(\psi)$ (analyzed states resulting from the EnKF procedure) corresponding to ψ^f and ψ^a are approximated by N ensemble members [van Dinther et al., 2017]. In van Dinther et al. (2017) the ensemble was generated using small perturbations resulting from small difference in cell markers distribution. The ensemble members were running in parallel as separate models until reaching an assimilation step where each of the ensemble members is manipulated into a new value ψ^a_i as part of the analysis stage (see explanation below). The error covariance matrix $C^f_{\psi\psi}$ is initially approximated by a forecast-best estimate

$$C^f_{\psi\psi} \approx C^{f,e}_{\psi\psi} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(\psi^f_i)_p - \overline{(\psi^f)}_p\right]\left[(\psi^f_i)_q - \overline{(\psi^f)}_q\right]^T \tag{2-18}$$

where $\overline{(\psi^f)}$ is the mean ensemble states forecast and (p,q) are matrix entries. Notice that due to the true value being unknown, the covariance matrix is determined by the distance of each ensemble member to the ensemble's mean value. The model states are propagated using the dynamic model described in section 2-1 and at each assimilation step every ensemble member is analyzed and updated according to:

$$\psi^a_i = \psi^f_i + C^{f,e}_{\psi\psi}M^T(MC^{f,e}_{\psi\psi}M^T + C^e_{dd})^{-1}(d_i - M\psi^f_i) \tag{2-19}$$

where $C^e_{dd} = \varepsilon_{d,i}\varepsilon^T_{d,i}$ is the measurement error covariance matrix (with $\varepsilon_{d,i}$ being the measurement error) and M is the measurements matrix, within it locations of observation are stored and d_i is the observation corresponding to ensemble member i. As mentioned before propagation of the state in time also propagate its uncertainty in time, this enables the estimation of the error covariance by sampling the analyzed models of the ensemble:

$$C^{a,e}_{\psi\psi} = \left[(\psi^a_i)_p - \overline{(\psi^a)}_p\right]\left[(\psi^a_i)_q - \overline{(\psi^a)}_q\right]^T \tag{2-20}$$

By doing so, it is also necessary to add artificial random perturbations to the observed data in order to have a correct propagation of the error statistics [Burgers et al., 1998].

$$d_i = d + \varepsilon_{d,i} \tag{2-21}$$

The result is a vector of observed data corresponding to the different ensemble members. The best estimate is obtained by taking a mean value of the analyzed ensemble members $\overline{\psi^a}$. In contrast to the EnKF, unless a re-sampling technique is applied (then the composition of the ensemble is changed) particles do not change during the updating stage using the PF. In addition to that, the two methods differ in representation of model errors. While in the EnKF model errors are represented by the ensemble and its corresponding covariance matrix, in the PF model errors are fully represented by the choice of the ensemble.

2-2-3 Ensemble smoother

In literature smoothers differ from filters by calculating an estimate at time k based on past and future measurements (observed data) $p(\psi_k|d_{1:T})$ [Särkkä, 2013]. As opposed to the PF where information is carried only forward in time, in smoothers information is carried backward and forward in time and the best estimate for each time step is based on all observed data in the time series [Evensen & Van Leeuwen, 2000]. For that reason Smoothers are a useful tool for studying the process' characteristics and gaining further understanding. The Ensemble Smoother (ES) solution discussed here was introduced as "Direct ensemble method" in van Leeuwen & Evensen (1996) and since then it has been investigated with regards to non-linear dynamics and found a wide use in flow problem and petroleum reservoir history matching [Skjervheim & Evensen, 2011; Emerick & Reynolds, 2013; Crestani et al.,2013]. In their paper, Evensen & Van Leeuwen (2000) derive a general formulation of the smoother into a sequential smoother, that for the time interval $t \in [t_1, t_k]$ would take the form

$$p(\psi_1, ..., \psi_k|d_1, ..., d_k) = \frac{p(d_1, ..., d_k|\psi_1, ..., \psi_k)p(\psi_1, ..., \psi_k)}{\int p(d_1, ..., d_k|\psi_1, ..., \psi_k)p(\psi_1, ..., \psi_k)d\psi}. \tag{2-22}$$

Eq.2-22 assume d to be a discrete data vector. With the additional assumptions that (1) the process is a first order Markov chain and (2) a data point within d at time $t \in [t_1, t_k]$ depends only on the current state (see Evensen & Van Leeuwen, 2000), the smoother formulation becomes:

$$p(\psi_0, \psi_1..., \psi_k|d_1, ..., d_k) = \frac{p(d_k|\psi_k)p(\psi_0, \psi_1, ..., \psi_{k-1}|d_1, ..., d_{k-1})p(\psi_k|\psi_{k-1})}{\int p(d_k|\psi_k)p(\psi_0, \psi_1, ..., \psi_{k-1}|d_1, ..., d_{k-1})p(\psi_k|\psi_{k-1})} \tag{2-23}$$

If k=2 in the above equation, the data is assimilated based on the time interval $t \in [t_1, t_2]$ and as k increases more data influences the solution. van Leeuwen & Evensen (1996) and Evensen & Van Leeuwen (2000) applied the ES in the context of the EnKF, such that the addition to ψ_i^f in eq.2-19 is determined by integration of all data defined in the time interval for the smoother. For an ensemble smoother in the context of the PF, refer to section 3-3.

Chapter 3

Methods

The work of constructing a suitable particle filter was based on: 1) generating an evolution in time and space of a seismic cycle in a subduction zone, using the previously mentioned STM model (chapter 2) and 2) writing a code which reads the evolution of state variables, sample (generate) particles from it, attach a weight to each particle and eventually calculate an estimation to the true state variables.

3-1 STM model: setup and output

For running a simulation of the seismic cycle, we require initial geometry and parameters; those were chosen to be the same as in van Dinther et al. (2017), since one of the goals of this thesis is to compare the performance of EnKF with that of PF. However, it is important to mention that full comparison between the two based on a model generated by the current STM code is not possible, due to the fact that markers inside the cells of the model are distributed randomly and therefore, each time a different state evolution is obtained. As part of the simulation, files containing the model geometry and the state variables are outputted at user predefined time steps. The state variables V_i, V_j, σ_{ii}, σ_{ij} and P at a subsurface location serve as observed data for data assimilation.

The time step was set to 0.066 seconds. Van Dinther et al. (2013b) explained this choice of time step with the stability of the solution for a range of models (see Appendix A2 in their paper). The model generated by the STM code was set to contains 701x136 nodes corresponding to the x and z directions respectively. The model was set to output state variables for points of interest (markers) at each time step (can be seen in Figure 3-1): 10 markers situated on the top of the gelatin wedge (GPS markers) and 14 Markers situated on the fault (contact surface between the slabs) or above it. The point marked with a black circle is the location of the borehole which can be regarded as a five km deep borehole [van Dinther et al., 2017].

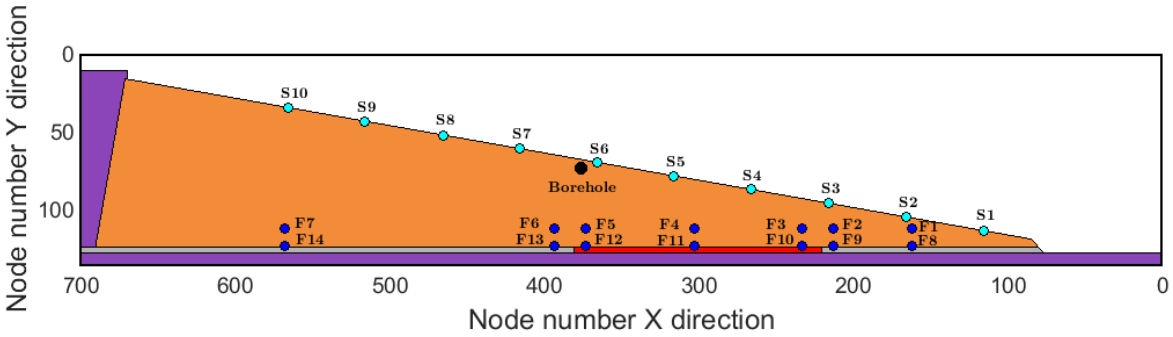

Figure 3-1: Model domain consisting of 701x136 nodes. Properties differ inside the domain according to the material: Air (white), Gelatin (orange), Fault (red- seismic and gray- aseismic zones), backstop wall (magenta). Dots represent markers of interest: GPS markers (light blue) borhole location (black) and Fault markers (blue).

3-2 Particle Filter

The particle filter introduced here, is a basic bootstrap particle filter described in detail in section 2-2-1 in which particles have been generated with a time lag sampling technique rooted in the LAF of Hoffman & Kalnay (1983). As discussed before, the basic particle filter does not require an assumption of linearity of the model or Gaussianity of the probability densities thus, it is an appropriate DA technique for the STM model. Nonetheless, there are other assumptions that had to be made in order to use the particle filter:

- The model is a Markov chain process, meaning it is stochastic and the current state depends on former state [Evensen & Van Leeuwen (2000)]:

$$d\psi = F(\psi)dt + dv \tag{3-1}$$

  where F is the non-linear model operator and dv is random noise increment.

- Observed errors are uncorrelated between state variables.

- The model has finite model space with ensemble size corresponding to the model space size.

Since the work done here is only in its early stages and serve as a proof of concept, a fourth assumption is:

- regarding the STM model as a perfect model, meaning that it truthfully and completely reflects the physical process.

The standard PF performs an importance sampling, using an importance distribution q (which might be quite different from the true distribution p). Here I use time lag sampling, which does not sample randomly out of a distribution. If the system is ergodic, meaning that a single realization averaged over time is equal to the model space average [Walters, 2000], time lag sampling of one evolution should in theory represent the model space. Lagged time sampling

presents several advantages over other techniques of particle generation: it limits the model space to a finite number of combinations of the state variables, it forces the ensemble to contain only values which are physically meaningful and exist within the modeled process and most importantly by choosing lagged versions of an already existing evolution, computation of the ensemble models is spared. However, it loses its advantage if the ensemble is not constructed correctly or if the size of the ensemble is still not large enough to have a good representation of the model distribution. Moreover, model errors might not be correctly represented due to realizations being time lagged versions of the same model. To truly stand within the limits of the second part of the 3^{rd} assumption in cases of large model space, one is required to have a very large ensemble that may not be feasible or practical if computation increases remarkably with increasing number of ensemble members.

3-2-1 Ensemble generation

The PF performs the sampling of an existing combination of state variables V_i, V_j, σ_{ii}, σ_{ij} and P out of an evolution generated by the STM model; this is done by using a time lag Δt and its multipliers. At each t_a (assimilation time step where a=1,2,3...n), a time lag $\sum_{n=1}^{N_s} n \cdot \Delta t$ is used to sample N_s particles consisting of 5 state variables and generate an ensemble. For the time being Δt is chosen to be constant, but random Δt that results in random time separation between the various particles, is also possible. The concept of doing so is to make copies of different stages of the true evolution of the seismic cycle, as simulated using the STM model and artificially force them to t_a. To illustrate the idea of time lag sampling, Figure 3-2 shows an ensemble of four time lagged particles. The blue line is a model run and the other colored lines are the time delayed versions (particles trajectories) forced to $t \approx 141$ s. For each assimilation step the ensemble will include four possible combinations of the state variables that are simply the states of the system at different times in the seismic cycle.

Referring back to eq. 2-10, the prior using a PF with time lag sampling would take the form:

$$p(\psi) = \frac{1}{N_s} \sum_{i=1}^{N_s} \delta(\psi - \psi_{t+i\Delta t}) \tag{3-2}$$

The prior is now a collection of delta functions of time lagged states existing in one model run which ideally represent the statistical uncertainty. To further demonstrate the idea of time lag sampling, Figure 3-3 shows two evolutions: evolution from which particles are sampled (blue) and the true evolution (brown). The figure also shows the true state (big red bullet) at a specific assimilation step $t_a \approx 141$ s and an ensemble consisting of four particles (small colored bullets), sampled by a constant time lag from a single evolution. Figures 3-4 and 3-5 show an ensemble of four particles in five subsequent assimilation steps for V_i and σ_{ii} respectively.

3-2-2 Assimilation: calculating weights and estimation

To be as realistic as possible for this assimilation experiment, errors were added to each state variable in the true state. The errors were created and added to each state variable separately,

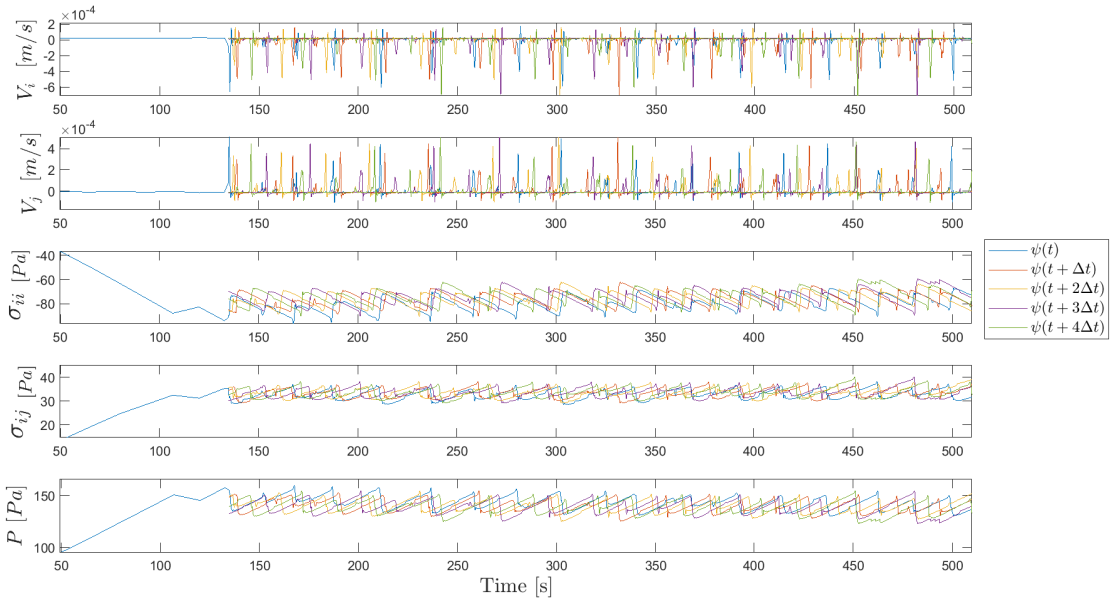

Figure 3-2: Visualization of the time lagged particle sampling technique discussed. Each colored line represents a time lagged version of the seismic cycle evolution in blue.

by randomly sampling a normal distribution around zero with a standard deviation that corresponds to the state variable uncertainty. The standard deviations used here were taken from van Dinther et al. (2017) and appear in table 2 in their paper. These errors reflect the uncertainty in the measurement.

The posterior of a marginal distribution of the model space in a time lagged form reads

$$p(\psi|d) = \sum_{i=1}^{N_s} w_i \delta(\psi - \psi_{t+i\Delta t}). \tag{3-3}$$

Recall w_i from eq.2-12. The likelihood appearing in w_i was calculated using a Lorenz distribution. The final form of the calculated posterior is given by:

$$p(\psi_{t+i\Delta t}|d) = \frac{\sum_{i=1}^{N_s} \left(1 + \frac{[d-H(\psi_{t+i\Delta t})]^2}{\sigma^2}\right) \delta(\psi - \psi_{t+i\Delta t})}{\sum_{i=1}^{N_s} \left(1 + \frac{[d-H(\psi_{t+i\Delta t})]^2}{\sigma^2}\right)} \tag{3-4}$$

where $d - H(\psi_{t+i\Delta t})$ is the innovation vector with M elements (M being the number of state variables), indicating distance/misfit between the observed data and the particle and σ^2 is

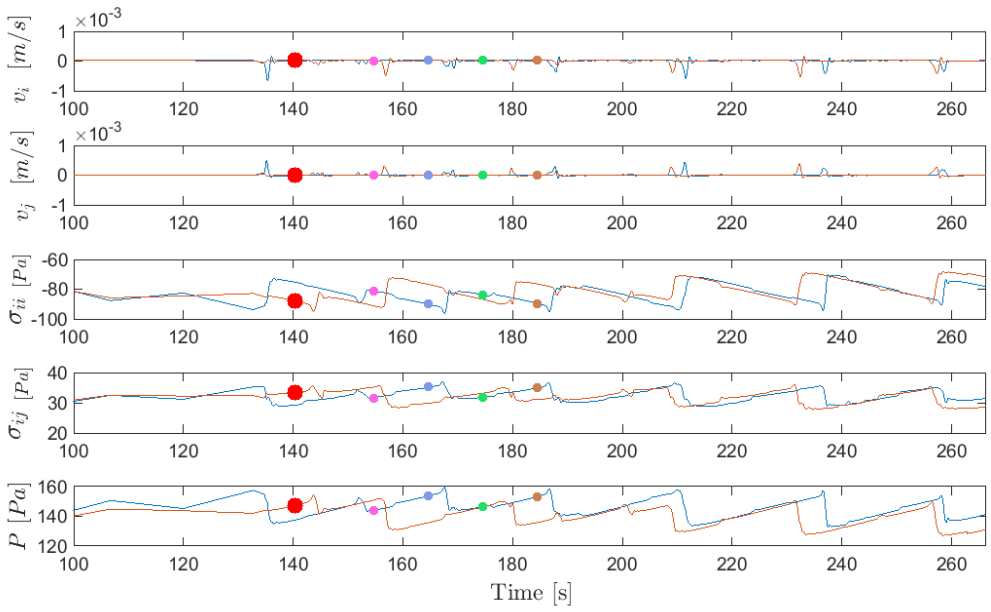

Figure 3-3: Illustration of assimilation step with the true evolution (brown line) and the true state indicated at $\sim$ 141 seconds by the red bullet. The ensemble of particles sampled from a separate evolution (blue line) is indicated by the small colored bullets. In order to use the true state as observed data, errors are added to each assimilated state with a corresponding error size.

the variance matrix with $M \times M$ dimensions. For the i^{th} particle at the a^{th} assimilation step

$$\frac{[d^a - H^a(\psi_i)]^2}{\sigma^2} = \begin{bmatrix} V^a_{i,obs} - V^a_{i,i} \\ V^a_{j,obs} - V^a_{j,i} \\ \sigma^a_{ii,obs} - \sigma^a_{ii,i} \\ \sigma^a_{ij,obs} - \sigma^a_{ij,i} \\ P^a_{obs} - P^a_i \end{bmatrix}^T \begin{bmatrix} \sigma^2_{\varepsilon_{V_i}} & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_{V_j}} & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_{\sigma_{ii}}} & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\varepsilon_{\sigma_{ij}}} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_P} \end{bmatrix}^{-1} \begin{bmatrix} V^a_{i,obs} - V^a_{i,i} \\ V^a_{j,obs} - V^a_{j,i} \\ \sigma^a_{ii,obs} - \sigma^a_{ii,i} \\ \sigma^a_{ij,obs} - \sigma^a_{ij,i} \\ P^a_{obs} - P^a_i \end{bmatrix} \tag{3-5}$$

To justify the choice to use the Lorenz function and the prove that it is indeed broader than a Gaussian distribution, the reader is refereed to Figure 3-6. The figure shows that for the same ensemble composition, a Gaussian distribution would always be narrower than a Lorenz distribution enabling less particles and thus less statistical information to be incorporated. This choice later proved to be crucial when it comes to the first two state variables V_i and V_j.

The result of eq.3-4 for a single particle is a scalar which is a weighted average of all five states in this particle. Each of the particles is assigned a weight (posterior) according to eq.3-4 and the estimated state is a weighted average of the particles.

Recall from section 2-2-2 describing the EnKF as applied in van Dinther et al. (2017), that as opposed to the PF, the EnKF does not weigh particles, but rather calculates and applies a correction of the misfit multiplied by the Kalman gain to each particle, leading to creation of a new ensemble of analyzed particles out of which the best estimate is calculated as a mean value. This difference between the two methods does not only affect the goodness of estimate

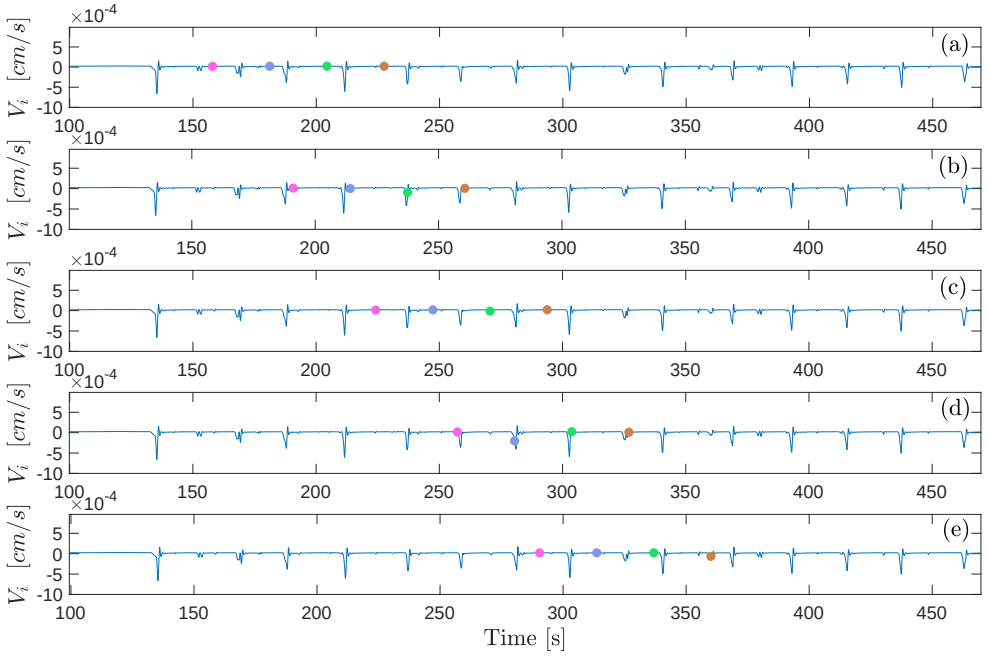

Figure 3-4: Illustration of an ensemble in subsequent five assimilation steps (a-e) for the state variable V_i. Colored bullets are the particles constituting the ensemble.

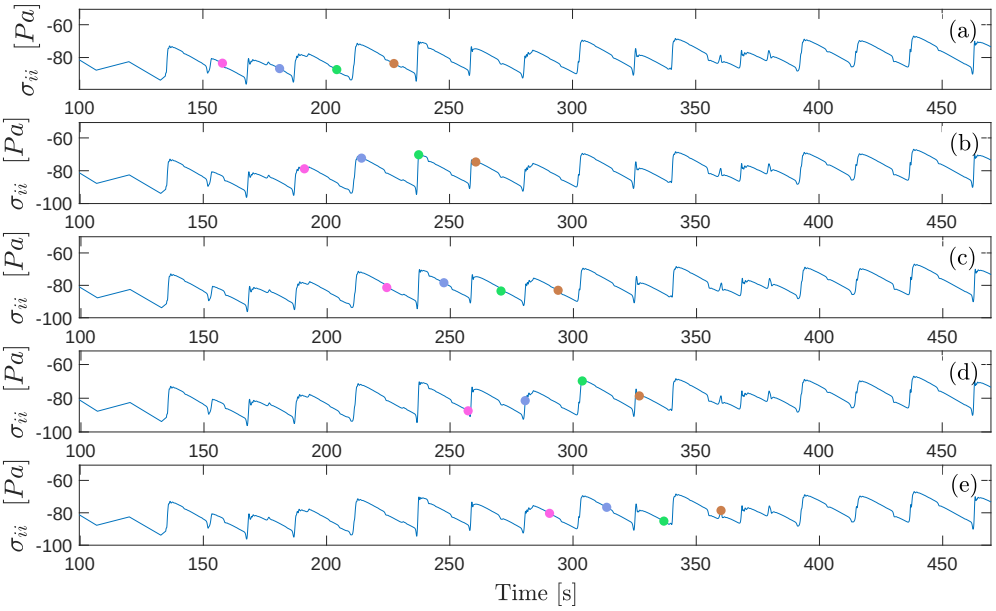

Figure 3-5: Illustration of an ensemble in subsequent five assimilation steps (a-e) for the state variable σ_{ii}. Colored bullets are the particles constituting the ensemble.

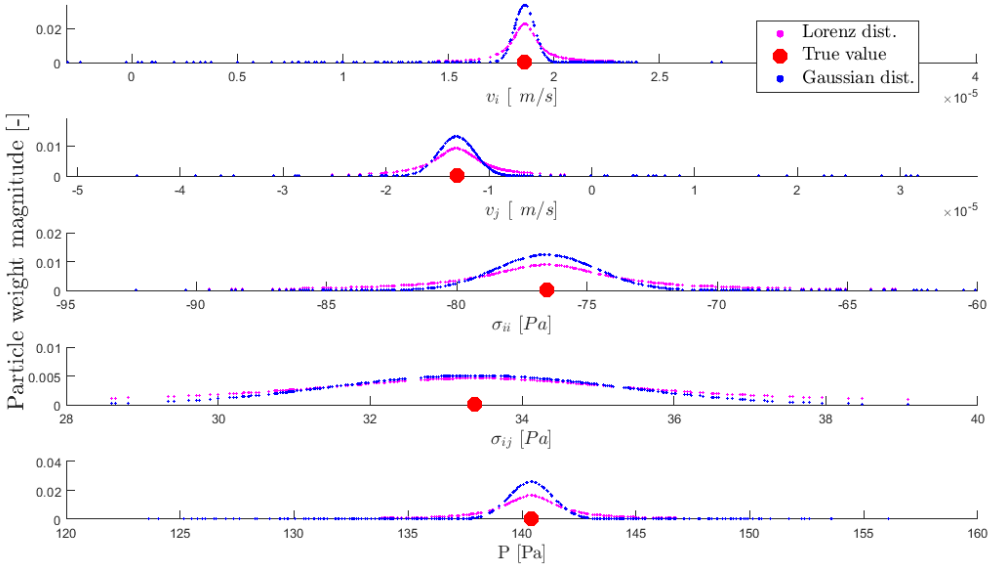

Figure 3-6: Comparison between Lorenz and Gaussian distributions of particle weights for the same ensemble of particles.

but also the computation time. PF requires only computation of the posterior while particles remain unchanged. Whereas in the EnKF particles are being updated for each of the states and future state is then predicted based on the new ensemble. Another advantage of the PF is that it does not require running the assimilation while running the STM model; this is not the case when using the EnKF where assimilation is performed online while running the STM model.

3-3 Particle Smoother

The smoother could be implemented based on existing particle filter implementation. The smoother only requires storing the observed data and particles for all assimilation steps. Storing all data as a single vector, the smoother routine is performed using

$$p(\psi_{t+i\Delta t}^{1:k})|d^{1:k}) = \frac{\sum_{i=1}^{N_s}(1+\frac{[d^{1:k}-H(\psi_{t+i\Delta t})^{1:k}]^2}{\sigma^2})\frac{1}{A}\sum_{k=1}^{A}\delta(\psi-\psi_{t+i\Delta t}^{k})}{\sum_{i=1}^{N_s}(1+\frac{[d^{1:k}-H(\psi_{t+i\Delta t})^{1:k}]^2}{\sigma^2})}, \tag{3-6}$$

where A is the total number of assimilation steps. This equation indicates that only a single weight is calculated to the i^{th} particle based on the entire assimilation sequence in time. For

the i^{th} particle, $d^{1:k} - H(\psi_{t+i\Delta t})^{1:k}$ would take the form:

$$\begin{bmatrix} V^1_{i,obs} \\ V^1_{j,obs} \\ \sigma^1_{ii,obs} \\ \sigma^1_{ij,obs} \\ P^1_{obs} \\ V^2_{i,obs} \\ \vdots \\ V^k_{i,obs} \end{bmatrix} - \begin{bmatrix} V^1_{i,i} \\ V^1_{j,i} \\ \sigma^1_{ii,i} \\ \sigma^1_{ij,i} \\ P^1_i \\ V^2_{i,i} \\ \vdots \\ P^k_i \end{bmatrix}$$

and the variance matrix σ^2 is constant for all particles

$$\begin{bmatrix} \sigma^2_{\varepsilon_{V_i}} & 0 & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & \sigma^2_{\varepsilon_{V_j}} & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_{\sigma_{ii}}} & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & \sigma^2_{\varepsilon_{\sigma_{ij}}} & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_P} & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_{V_i}} & .... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\varepsilon_P} \end{bmatrix}$$

The estimated state for each assimilation step is a weighted average of all particles in that step. As mentioned before, the i^{th} particle would be multiplied by the same weight for all assimilation steps. If we look at a matrix with size $K \times N$ (K number of assimilation steps and N number of particles) containing all weights, for the PF the elements in the rows and columns will vary, while for the smoother the rows are simply copies.

Part II

Second Part

Chapter 4

Ensemble generation

The previous chapter described the PF, PS and the time lag particle sampling technique. In order to find the proper ensemble that characterizes/describes the system, different parameters for creating an ensemble were tested. The two most important parameters which characterize the ensemble are time lag Δt (time separation between particles) and number of particles N_s (size of the ensemble). Given a simulation time step size Δt of 0.066 s, assimilation step size for all the experiments was chosen to be 2 s corresponding to 30 simulation time steps (taken to be same as in van Dinther et al. (2017)). Apart from being a measure to find a suitable ensemble, these tests are also indicators for the sensitivity of the representation of model error to the sampling technique.

4-1 Early analysis and representation of process physics

In different experiments (even when repeating an experiment) the error being added to the data is random and thus, different each time. To make sure that random errors do not significantly affect the results and that the PF is robust, an experiment was repeated several times in which only the random error (added to the data) varies. It was found that the deviation of errors from the mean error (difference between the true and the expected evolutions over 10 experiments), does not exceed 3.3% in the borehole location and 2.3% (except for v_i where velocity is of the order of 10^{-5} and deviation goes up to 8.7%) at the F11 marker location (see fig.3-1). Considering that the errors are around $10^{-3} - 10^{-5}$ [cm/s] for the velocities and $10^{0} - 10^{1}$ [Pa] for the stress and pressure state variables, deviation of the error between experiments is at least one order smaller than the mean error over all experiments. It is therefore concluded that the PF is robust and that variation of the results between different experiments due to randomization of the measurement error can be disregarded.

As a starting point and as a way to gain insight into the PF and physical system, the ensemble of particles is drawn out of an evolution considered to be the 'true evolution'. Particles are being sampled with respect to the true state at the current assimilation time step. To distinguish past and future sampling in time, sampling particles from future evolution to

current true state will be termed forward sampling and sampling particles from both past and future times to current true state will be termed Backward-Forward sampling.

Looking at figure 4-1 it is observed that σ_{ii} and P exhibit a trend with time. The term trend here is used to describe a consistent increase or decrease of the mean value with time. To illustrate the problem of sampling data with an embedded trend, Figures 4-2 and 4-3 show a time window of 199 seconds assimilated with Forward and Backward-Forward sampling and $\Delta t = 6.63$ seconds.

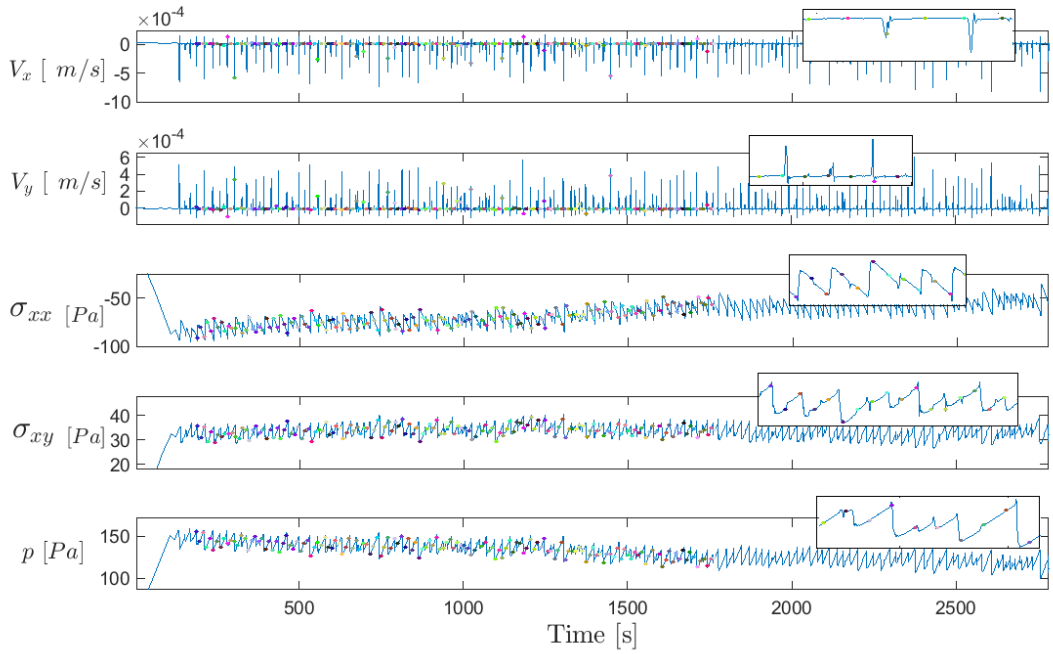

Figure 4-1: Particles sampled from evolution of v_i, v_j, σ_{ii}, σ_{ij} and P with a constant time lag Δt.

The dark pink background in figures 4-2 - 4-3 is created by plotting ensemble paths consisting of 150 particles. The blue and black lines are the true evolution (without added errors) and expected evolution respectively. The mean absolute error for σ_{ii}, σ_{ij} and P is 5 times lower when sampling past and future states as opposed to sampling only future states with respect to current assimilated state. The results for the two velocities are surprising, also exhibiting better fit even though no trend is apparent. The pink envelope of ensemble paths bounding the expected value can be also viewed as a representation of the model error. Thus, the true state has to be enveloped by the ensemble of particles in each state from above and below, in order to be properly assimilated.

In order to successfully perform DA using the PF, the ensemble has to reflect the physics of the system/process. Observing figure 4-1, one could divide the seismic cycle as simulated by the STM model into two periods: inter- and co-seismic. During the interseismic period loading of stress and pressure occur and during the short coseismic period a slip occurs, resulting in stress and pressure decrease and sharp increase in the magnitude of velocity. Since the process is a cycle, it is expected to exhibit dominant frequencies of occurrence. These frequencies can

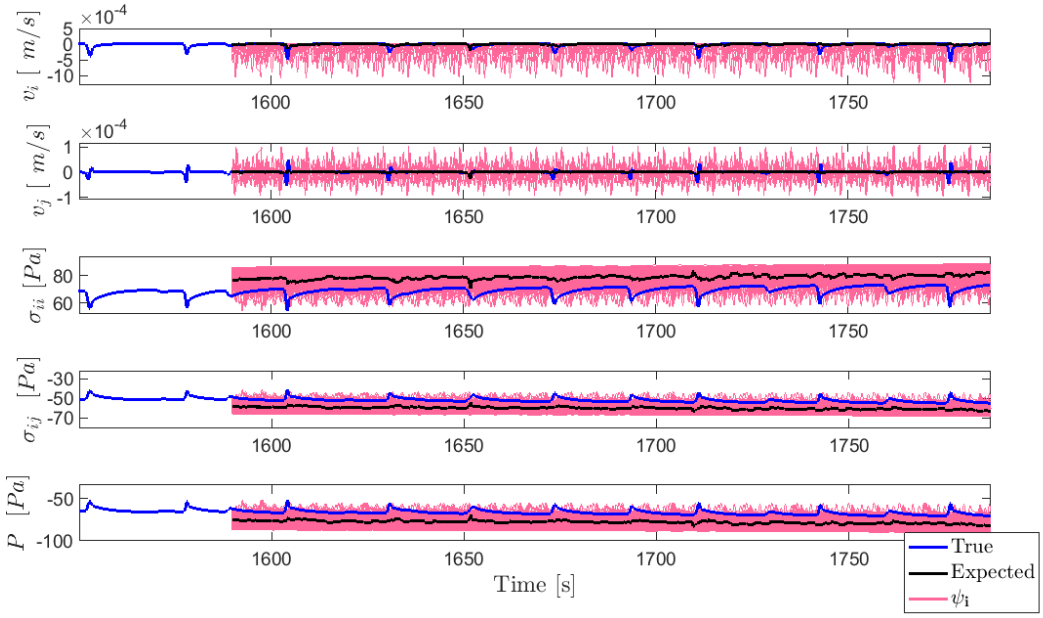

Figure 4-2: Fit of expected value to true evolution for GPS marker S1. Sampling particles forwards in time with $\Delta t = 6.63s$

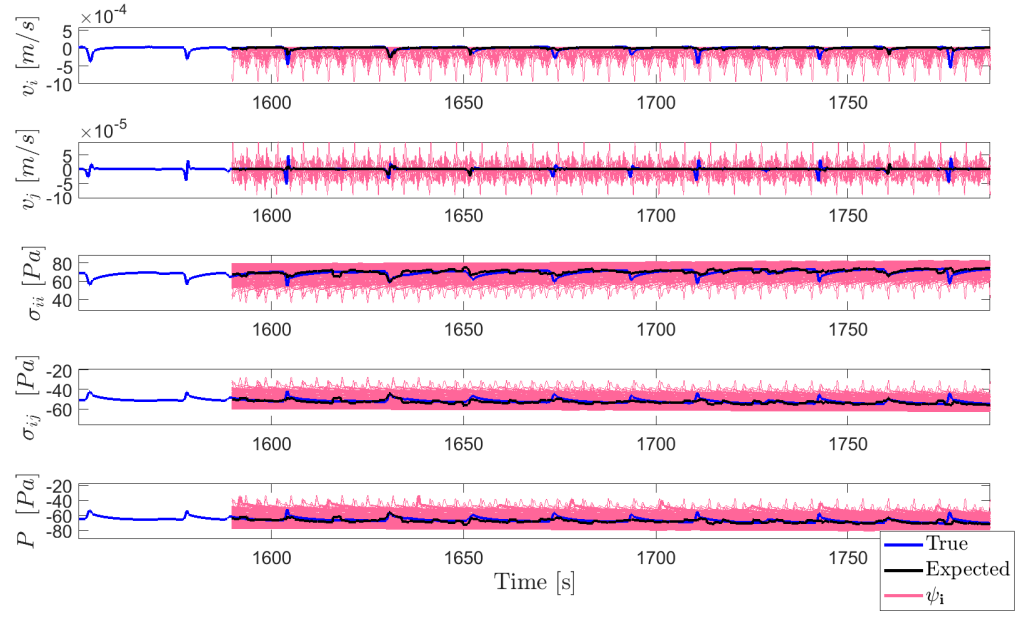

Figure 4-3: Fit of expected value to true evolution for GPS marker S1. Sampling particles Backward and forward in time with $\Delta t = 6.63s$

be obtained by performing Fast Fourier Transform (FFT) on a time window within the states' time series. Figure 4-4 shows the frequency spectrum of state variables v_i and σ_{ii} respectively.

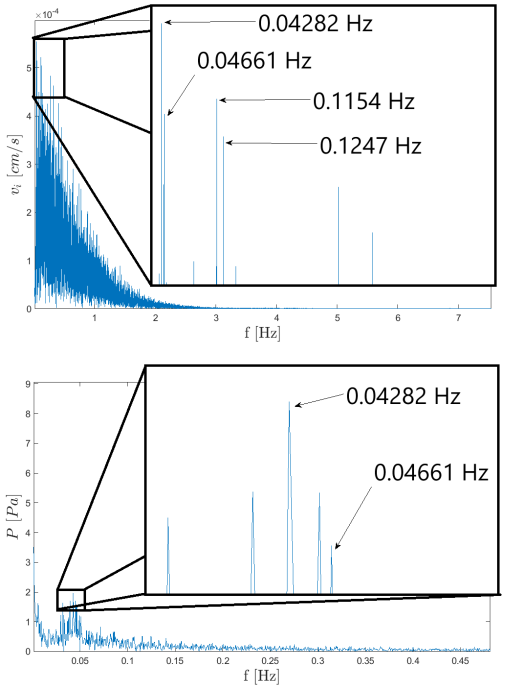

Figure 4-4: Frequency spectrum of (a) v_i and (b) P evolutions in time. Time window for FFT is $\sim$ 144 -1989 seconds.

Due to similarity of v_j to v_i and σ_{ij} and P to σ_{ii}, it is sufficient to look at the frequency spectra of v_i and P. The frequency spectra shows high amplitudes for a wide range of frequencies: the earthquake cycle is more complex than only few re-occurrence frequencies. Two frequencies are shared by both spectra are: 0.04282 and 0.04661 Hz, those are translated to $\frac{1}{f} = T = \Delta t$: 23.4 and 21.5 seconds respectively. van Dinther et al., (2013b) point out an average re-occurrence time interval of 19.3 s. In the context of time lag sampling, the FFT results suggest that in order to sample a full cycle, the sampling window $\Delta t \times Timestep \times N_s$ should be at least 20 s.

4-2 Particle separation

In a real case, the true evolution is not available. Therefore, the former example where the prior is fully consistent with the observation is not realistic. How would the PF Perform in a more realistic setting where particles are to be sampled from a different evolution to that of the true evolution? To answer that, experiments were constructed in which particles are sampled out of a different model run to that of the true state.

The time lag, or in other words the sampling frequency, determines the members of which the ensemble consists of. Therefore, it determines what stages of the seismic cycle are represented within the ensemble. So far it was suggested that the ensemble should accommodate the frequency of slip occurrence and that its range of values should be large enough to ensure

that the true value is within the ensemble's envelope of available states for assimilation. Ideally, the spread of the ensemble (minimum to maximum range) should be larger than the variance of the measurement error in order to account for it. If the ensemble's spread is too small, it might not include particles that are sufficiently close to the observation, lowering the chance of a successful data assimilation using the PF.

One way of testing the goodness of an ensemble is through looking at its prior distribution. Essentially, an ensemble is a collection of models, together these models compose a model distribution. Since it is approximated to be discrete, the model distribution should have a) a wide spread (range of values) for proper statistical representation and b) enough values that are smoothly scattered around the observation. These criteria were tested by looking at a histogram of ensemble values for increasing sampling time lag (figure 4-5).

v_i and v_j exhibited poor spreads with concentration of particles around zero velocity (interseismic period) even for increasing time lag. This is not surprising given that the interseismic period, in which velocity perturbations are fairly small, dominates the evolution. On the other hand σ_{ii} and P have shown to be very sensitive to different time lags, with approximately 50 Pa difference going from smallest to largest time lag. One possibility for this sensitivity could be the presence of trends with time (see figure 4-1). Yet, the variance of the measurement error is covered well already for time lag as small as 5 s for both σ_{ii} and P. It can be observed in figure 4-5.(c) that the spread in σ_{ij} values within the ensemble remains smaller for all tested Δt's. This suggests that the model error which depend on the variability of σ_{ij} with time is too small relative to the measurement error.

4-3 Ensemble size

It has been shown that for an ensemble containing 150 particles, the majority of particles have v_i and v_j around 0, with only few particles having velocities away from 0. In addition to that, the spread of σ_{ij} is smaller than the observational error. The main question in this subsection is whether it is possible to complement the frequency of sampling (time lag), by increasing the number of particles, in order to properly represent the interseismic and coseismic periods and avoid filter degeneracy. Figure 4-6 compares the spread of the ensemble for different number of particles. It comes as no surprise that the main difference between the different histograms of the same state is the form of the histogram rather than the variance. As the number of particles increases the ensemble takes the form of the true prior distribution. The increase in ensemble spread for σ_{ii} and P in (b) and (c) is a combination of increasing the number of particles thus, sampling later in time, and an existing trend in those state variables with time.

Figures 4-5 and 4-6 present possible contradictions between the (fitting) choice of time lag and number of particles. Looking at v_i, $\Delta t = 0.06$ s enables a higher sampling resolution of possible states during a seismic cycle, such that there are more particles away from 0 velocity. On the other hand σ_{ii}, σ_{ij} and P show larger spread as Δt increases.

As a summary, ensemble sampling window should be at least 20 s in order to capture a full cycle consisting of inter- and coseismic periods. Ensemble spread for state variables σ_{ii} and P is very sensitive to change in Δt and number of particles, as opposed to other state variable such as v_i and v_j. v_i and also v_j (not shown but exhibit the same behavior) display

very little sensitivity, especially for large Δt's due to dominant interseismic period which is characterized by only minor perturbations around the 0. σ_{ij} has variability that is lower than the measurement error, suggesting that very little information is added by assimilating it. There is a trade off between the resolution with which the ensemble can represent the seismic cycle and the spread of the ensemble when varying the time lag.

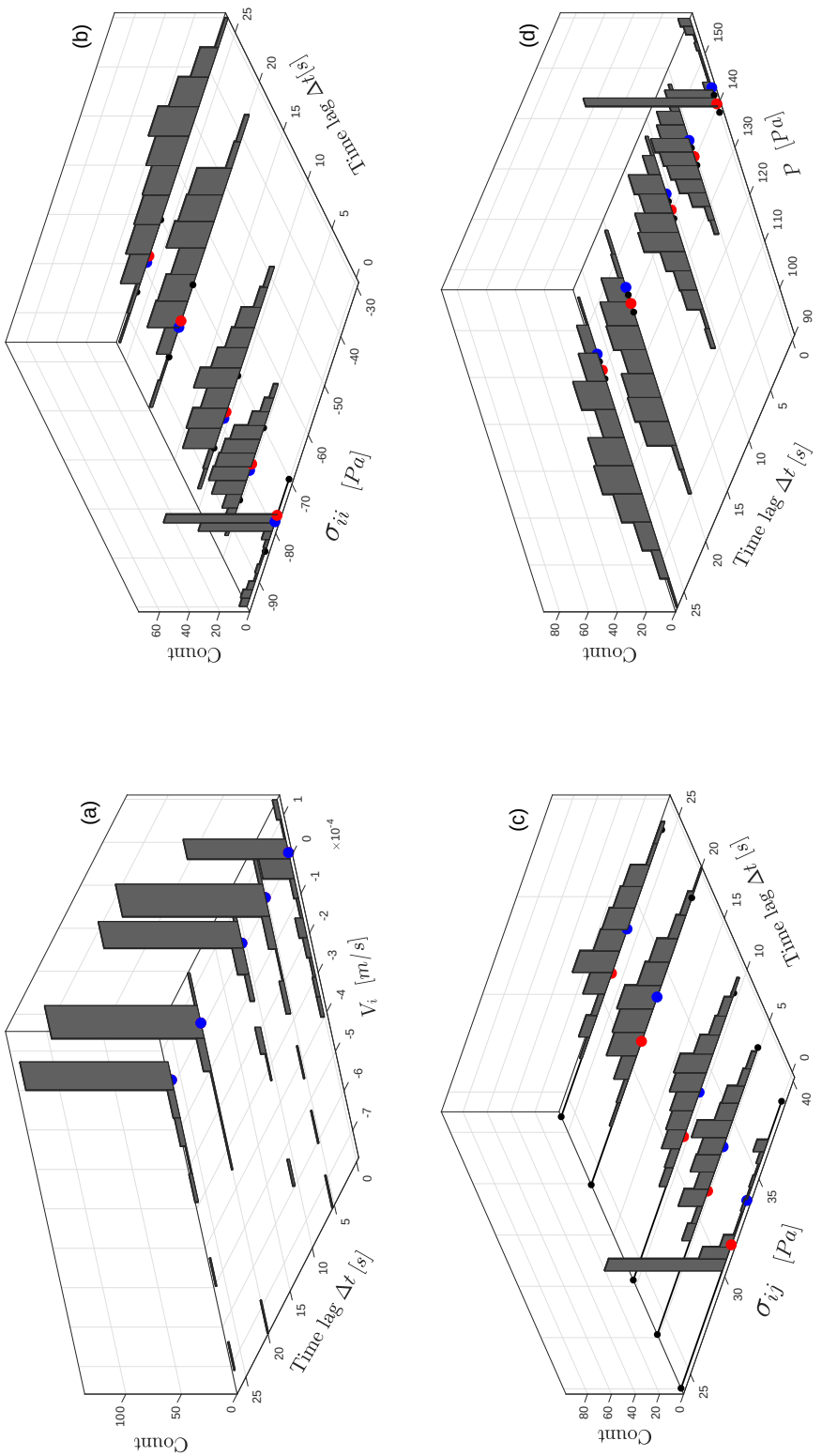

Figure 4-5: Ensemble spread for v_i, σ_{ii}, σ_{ij} and P state variables for different time lag between 150 particles. The red bullet indicate the true value for a specific state variable, blue bullet indicate the observation and black bullets indicate the variance of the observational error from the true state.

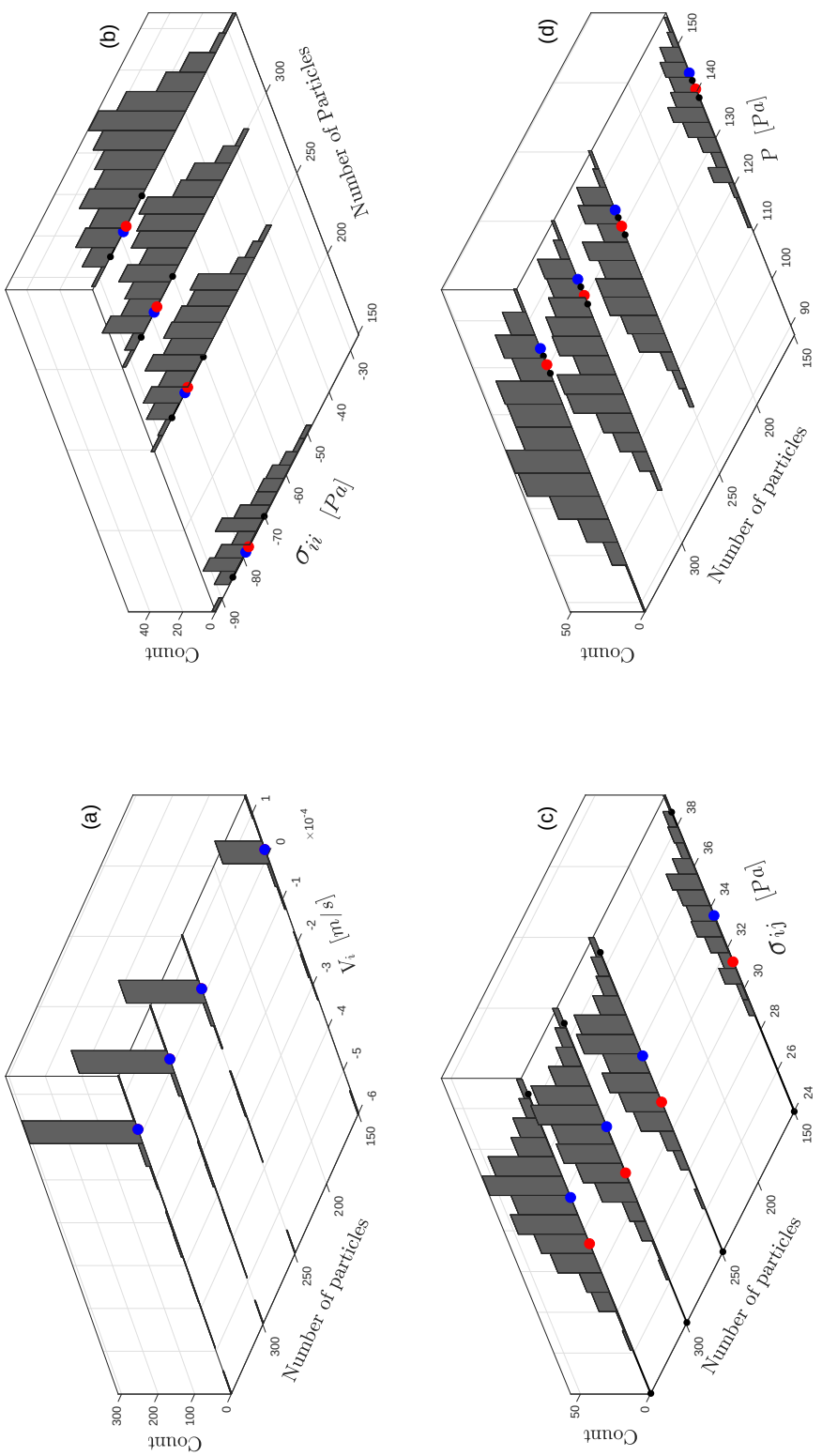

Figure 4-6: Ensemble spread for v_i, σ_{ii}, σ_{ij} and P state variables for varying number of particles and a fixed sampling time lag of 10.6 s. The red bullet indicate the true value for a specific state variable, blue bullet the observation and black bullets indicate the variance of the observational error from the true state.

Chapter 5

Results and analysis

Analysis in chapter 4 suggests that there is a trade off between ensemble coverage of possible states throughout a seismic cycle and its spread when changing the time lag with which the evolution is sampled. This trade off was tested with various combinations of time lags and ensemble sizes. In following sections two of these experiments are presented as representative examples. Both experiments were based on an ensemble consisting 300 particles in order to maximize coverage of possible states.

5-1 Particle Filter

One experiment was conducted with $\Delta t = 9.9$ s and 300 particles. Data fit and fit for marker F11 are displayed in figures 5-1 and 5-2 respectively. Recall from Figure 3-1 that the F11 marker is situated in the middle part of the seismogenic zone, where we would like to have state estimation. Although expected evolution (black) mimics the true evolution (blue) it does not fit it well and exhibits severe irregularities. The fit of the expected evolution to the true evolution is even poorer for the F11 marker, especially for the stress and pressure. The expected evolution remains around a certain value for each of the states, although smoother than for the fit of the borehole data, it is still noisy. The mean number of particles with effective weight N_{eff} for the entire assimilation window is 67.5±41.1 with 50% of assimilations steps having $Neff$ lower than 57.5. This means that for half of the assimilation steps, only fifth of the particles effectively contribute to the expected value while the rest of the particles have negligible weight.

The second experiment was conducted with $\Delta t = 0.33$ s (covering almost 5 cycles) and 300 particles. Results in figure 5-3 show an improvement from previous experiment of 300 particles and $\Delta t = 9.9$. Irregular shapes appear in the expected evolution at the borehole location as in previous experiment. For both experiments, velocity state variables exhibit the poorest fit compared to other state variables. $Neff$ has also improved with mean of 138.3 ± 50.3 and 50% of assimilation steps with N_{eff} lower than 141.6. This is a significant improvement in the effective number of particles. Results for marker F11 appearing in figure 5-4 also show an

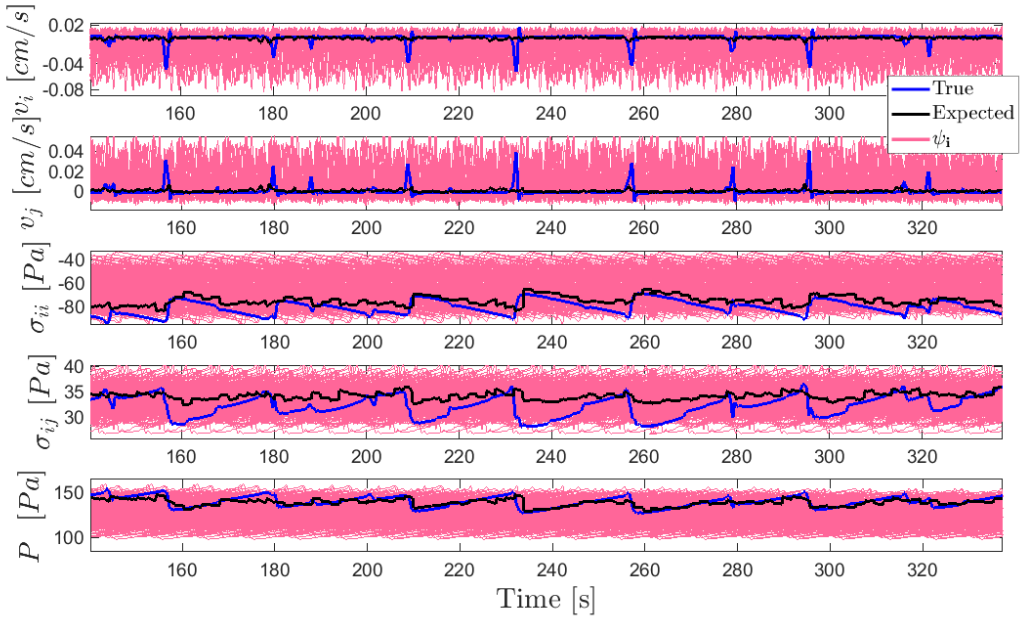

Figure 5-1: Data fit at the borehole location for experiment with 300 particles and $\Delta t = 9.9$ s. Mean absolute error v_i: 0.0032 ± 0.0065 [cm/s], v_j: 0.0023 ± 0.0044 [cm/s], σ_{ii}: 6.08 ± 3.71 [Pa], σ_{ij}: 2.10 ± 1.71 [Pa] and P: 3.12 ± 3.97 [Pa].

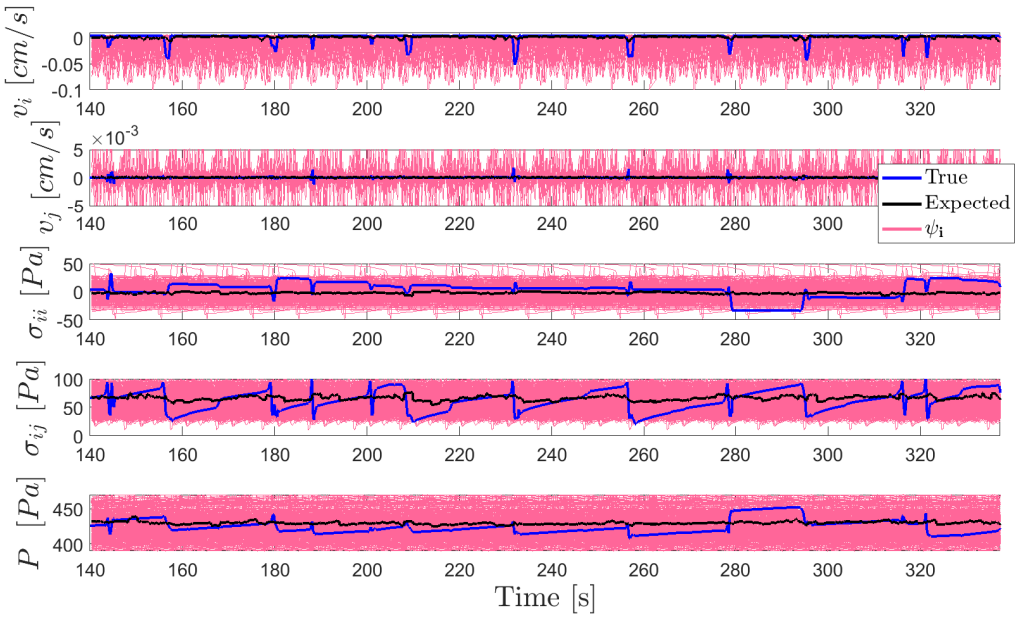

Figure 5-2: Fit of expected and true evolutions in F11 marker, for experiment with 300 particles and $\Delta t = 9.9$ s. Mean absolute error v_i: 0.0040 ± 0.0076 [cm/s], v_j: $7.72 \cdot 10^{-5} \pm 1.90 \cdot 0.00018$ [cm/s], σ_{ii}: 12.41 ± 13.87 [Pa], σ_{ij}: 14.43 ± 16.71 [Pa] and P: 9.58 ± 9.83 [Pa].

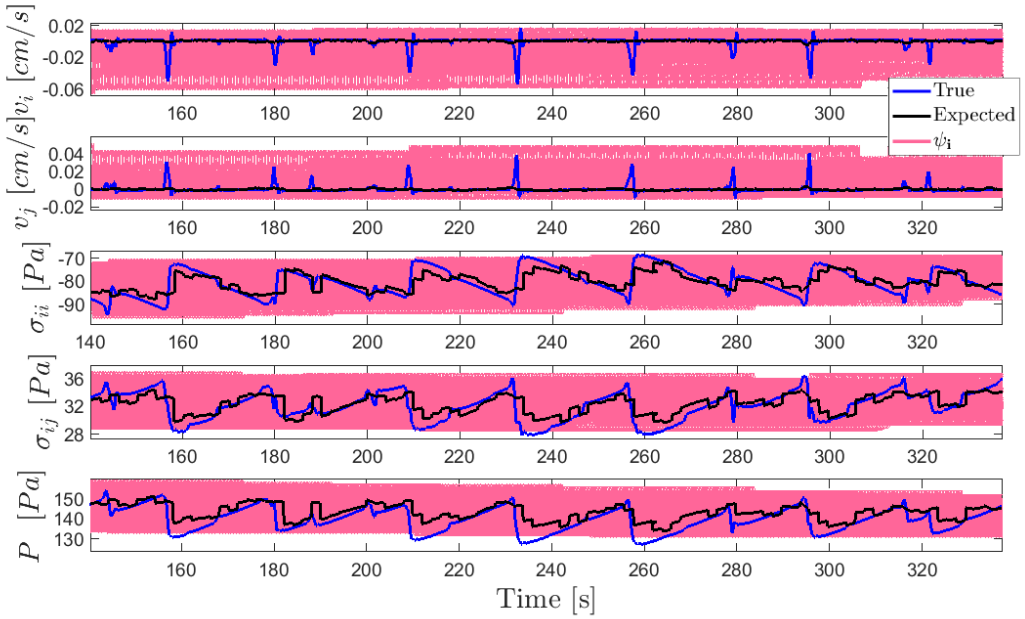

Figure 5-3: Data fit at the borehole location for experiment with 300 particles and $\Delta t = 0.33$ s.

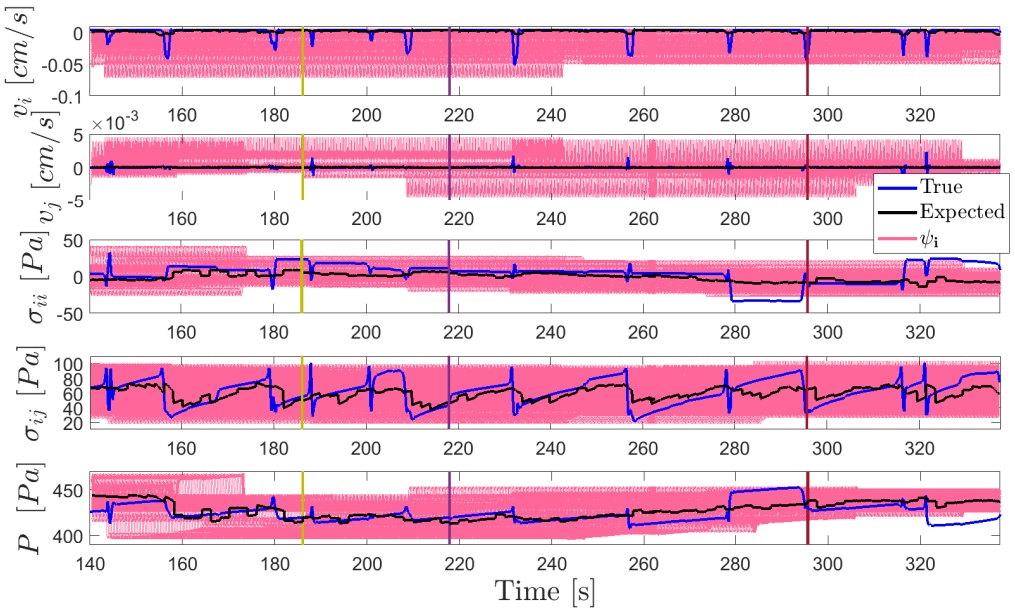

Figure 5-4: Fit of expected and true evolutions in F11 marker, for experiment with 300 particles and $\Delta t = 0.33$ s.

improvement in σ_{ij} and P fit. Improvement was also evident at other fault markers outside the seismogenic zone. MAE (mean absolute error) and SD (standard deviation) for time lag of 0.33 s are displayed in Figure 5-5 and corresponding RE (relative error) is displayed in Table 5-1. The equations for calculation of statistical values are detailed in appendix C. Comparison between errors at different marker locations reveals that the largest errors between true and expected evolutions are within the seismogenic zone while the smallest errors are observed downdip the aseismic zone (F14). This can be explained by the different rheological behavior of the down-dip aseismic part of the fault which compared to the seismogenic zone, experiences less events and weaker amplitudes.

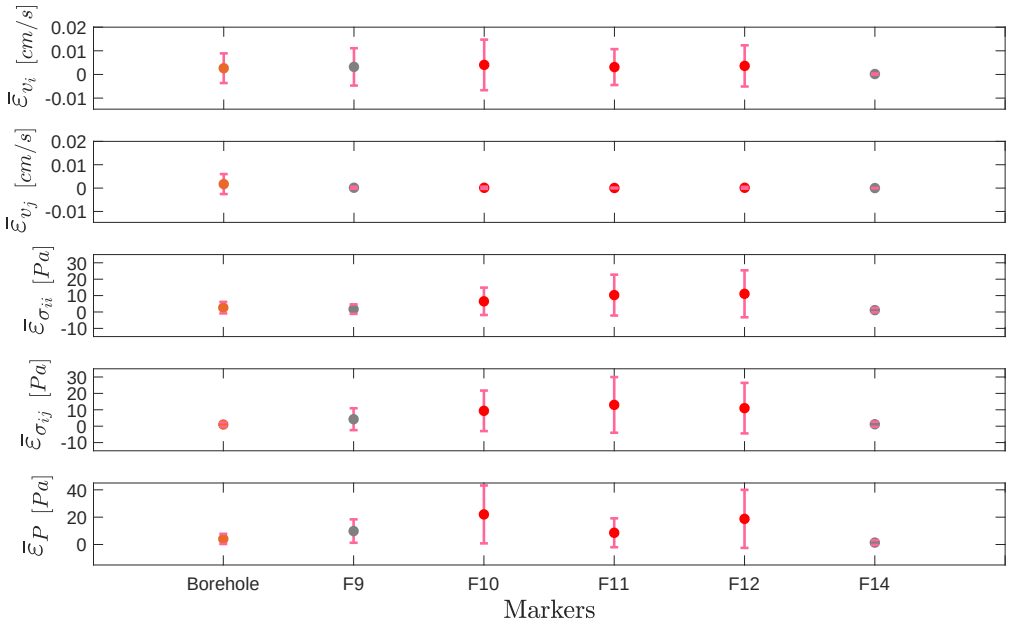

Figure 5-5: The mean absolute error between true and expected evolutions resulting from assimilating 300 particles that were sampled with time lag $\Delta t = 0.33$. The MAE and corresponding SD are displayed for few chosen markers on the fault and for the borehole. F10, F11 and F12 markers situated within the seismogenic zone.

	State variable	Borhole	F4	F10	F11	F12	F14
Relative error [%]	v_i	534.41	102.53	90.39	65.68	219.35	214.89
	v_i	153.72	1370.43	823.13	1462.39	230.50	-
	σ_{ii}	3.35	9.88	159.92	142.53	870.50	284.25
	σ_{ij}	3.23	14.53	20.55	23.56	12.60	3.04
	P	3.00	1.76	6.69	2.01	3.35	0.16

Table 5-1: Relative errors corresponding to previous figure (5-5) showing errors resulting from assimilating 300 particles that were sampled with time lag $\Delta t = 0.33$.

In order to understand how particles are distributed around each state and how misfits of multiple state variables are translated into a single posterior, the partitioned weight of each state variable was calculated and plotted separately for the 24^{th} (Fig.B-5 in appendix B), 40^{th} (Fig.B-7 in appendix B) and 80^{th} (Fig.5-6) assimilation steps (as marked by colored vertical

lines in fig.5-4). In these figures, small blue bullets are 300 particles distributed around the observation (Big blue bullet). Weight appears as partitioned weight, meaning that weights were calculated for each state variable independently. For example, the weight for state variable P was calculated for the i^{th} particle as

$$w_i(P) = \frac{\left(1 + \frac{[P^{obs} - P^{model}_{t+i\Delta t}]^2}{\sigma_P^2}\right)}{\sum_{i=1}^{N_s}\left(1 + \frac{[P^{obs} - P^{model}_{t+i\Delta t}]^2}{\sigma_P^2}\right)}. \tag{5-1}$$

Eq.5-1 simply normalizes the misfit (distance between observed and model state), scales it by the measurement error and distributes it according to a Lorenz distribution. The red bullet in these figures is the true value for each state variable and the purple bullet indicates the particle which has the maximum weight combining all states as shown in eq. 3-4.

According to figure 5-4 assimilation step 80 (marked with a reddish-brown vertical line) happens during the coseismic phase of the seismic cycle. Figure 5-6 shows that the highest posterior marked by the purple bullet is a combination of weights that are not necessarily the highest for each state variable separately. It means that in order to represent multiple states with a single weight, the states must be consistent in the particle realization. A possible reason for reduced consistency between the different states in the particle is the random observation error which differs for each state variable. Notice how velocity is poorly covered by particles; this might suggest that the ensemble is still missing with the part of the model space that represent the coseismic period. As expected by early analysis (section 4-2), the misfit of σ_{ij} appears as a very flat distribution of partitioned weight due to the variance assumed for this measurement. Overall, σ_{ij} contribute less information to the assimilation than other state variables, due to its variability in the ensemble which is small compared to the variance of the measurement error.

How would the distribution of the posterior look like for state variables on the fault? Figure 5-7 shows how the posterior distribution appears at the F11 marker location and which particles contribute the most to the expected value at that location for assimilation step 80. As in previous figure, the purple bullet indicates the highest posterior. The fit of the PF estimate to the true evolution depends on how close high weights are to the true state and although particles are distributed quite close to the true state (apart from v_i which was not captured by the ensemble) in assimilation step 80, for other assimilation steps particles are distributed far from the true state and captured by only few low weights. Another interesting point in assimilation step 80, is the weight distribution for σ_{ii}. When looking at figure 5-6, σ_{ii} is well covered by particles and the highest weight in the posterior distribution (purple bullet) is very close to σ_{ii}'s highest partitioned weight located at the middle of the distribution. But the weights plotted for the F11 marker reveal a different picture. While high weight is around 0, the true state is closer to -10. This is because the weight is determined by the combination of all state variables' misfits and not just σ_{ii}'s. It also suggests that a good fit in a variable at the borehole location does not necessarily mean that the true state is well estimated at the fault.

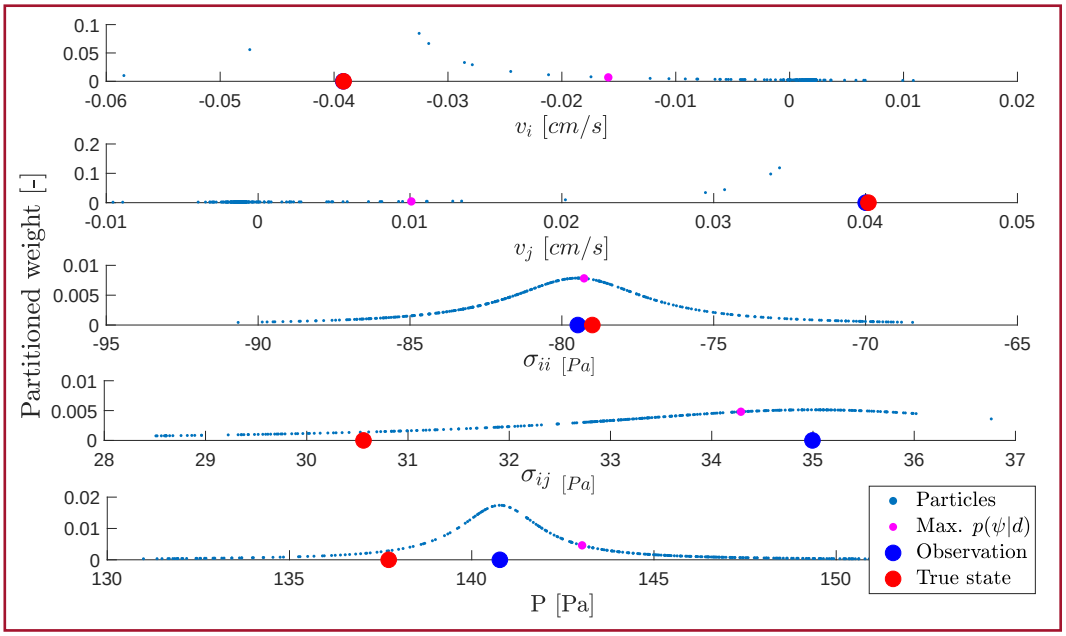

Figure 5-6: Weight contribution of each state separately at the borehole location for assimilation step marked by brown in figure 5-4. Particle marked in purple, is the particle with the highest weight of combined states w_i.

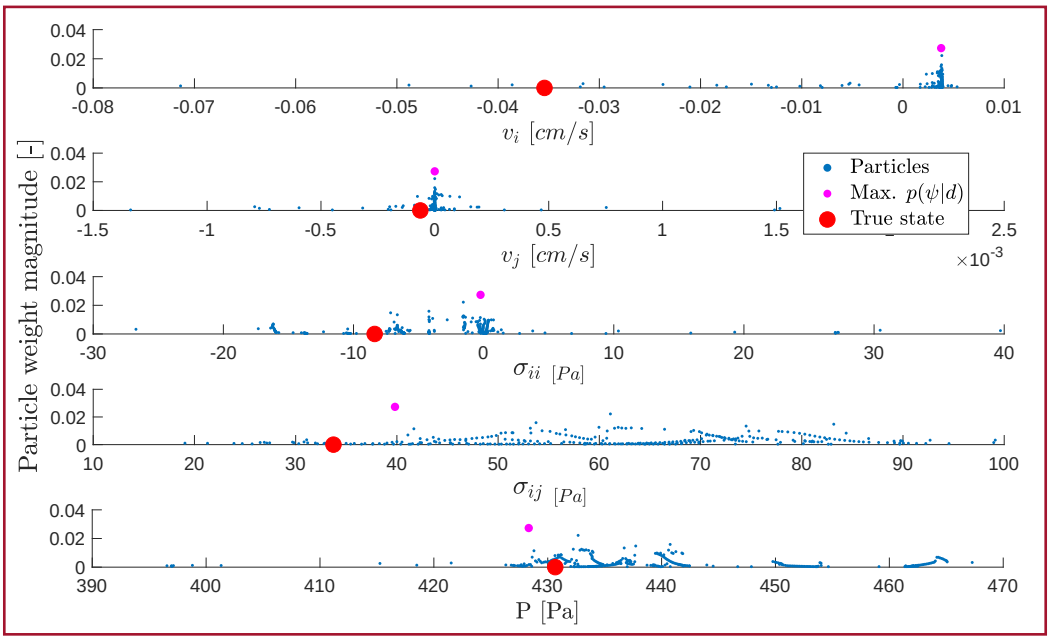

Figure 5-7: Combined posterior (w_i) with which particles are weighted at fault marker F11, for assimilation step marked by brown in figure 5-4.

5-2 Trend effect

One criterion defining a good ensemble is a constant large envelope made of particle paths that represents the model distribution for the entire assimilation window. In order to obtain a good fit to the observation, the ensemble should include many realizations with a state that is close to the observed state. With large time lags the spread of the ensemble is larger, but does the ensemble include realizations that are relevant to the observed state? and does the choice of time lag in the observation point necessarily fit other locations in the domain? To answer the former, figure 5-8 show a comparison between two distributions of particles sampled with time lags $\Delta t = 0.33$ and $\Delta t = 9.9s$. Notice that weights in this figure are **partitioned weights** (the weight distribution of each state independently) and not the weight combining all five states.

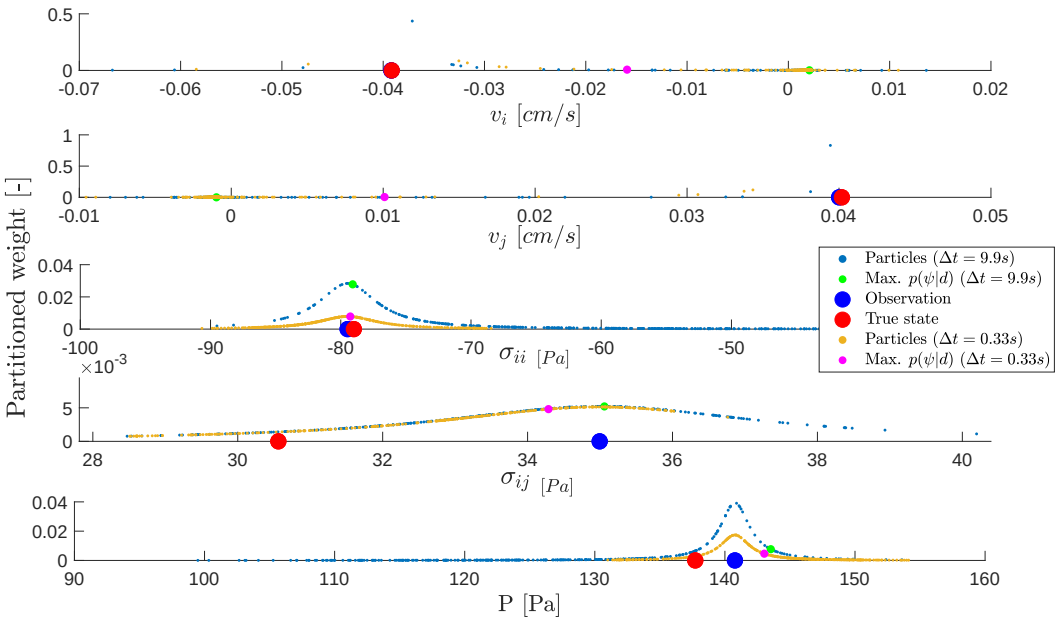

Figure 5-8: Comparison of partitioned weight distributions of particles sampled at the borehole location with $\Delta t = 9.9s$ and $\Delta t = 0.33s$.

The effective number of particles is clearly higher for small time lags where weight is distributed over a larger number of particles. Nevertheless, the larger time lag leads to a wider weight distribution in the sense that it is covering more values. Figures 5-9 to 5-11 further demonstrate the problem encountered when using large time lags. In these figures the relative sampling time of the various particles is highlighted by their color: blue (early) to yellow (late). Figure 5-9 shows a colored weight distribution (posterior, **not** partitioned weight!) as calculated from observations at the borehole, for particles that were sampled with $\Delta t = 0.33s$. As can be seen, particles sampled at different times are fairly mixed. Nonetheless, for most states the high weighted particles do not target the observation nor the true value. Figure 5-10 shows the weight distribution of particles that were sampled at the borehole location with a time lag $\Delta t = 9.9$ s. Weights are separated by their sampling time for state variables σ_{ii} and P while they are mixed for the others. High weights are restricted to the blue colored

particles. The combination of state variables is no longer consistent between early and late times of the evolution due to a presence of a trend in some state variables but not in others. The corresponding distribution at the F11 marker location, can be seen in Figure 5-11. Particles are well mixed around the true state, since trends are not present in data at the fault. The presence of a trend in the observed data leads to a mismatch between the current state at the borehole and states that are sampled later in time. Hence, particles that are sampled later in time would have low weights attached to them and their contribution to the expected value is reduced, even at locations where no trend is present and they are closer to the true state.

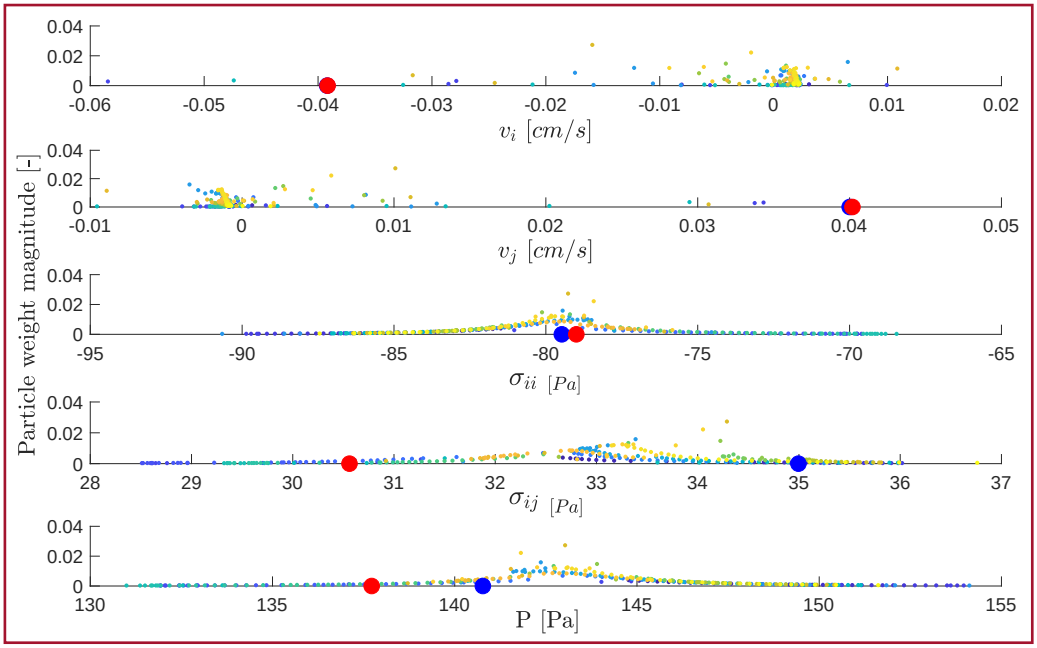

Figure 5-9: Posterior distribution of particles sampled with $\Delta t = 0.33s$ at the borehole location for assimilation step 80. Particles are colored by their order relative time in the evolution, from blue (early) to yellow (later).

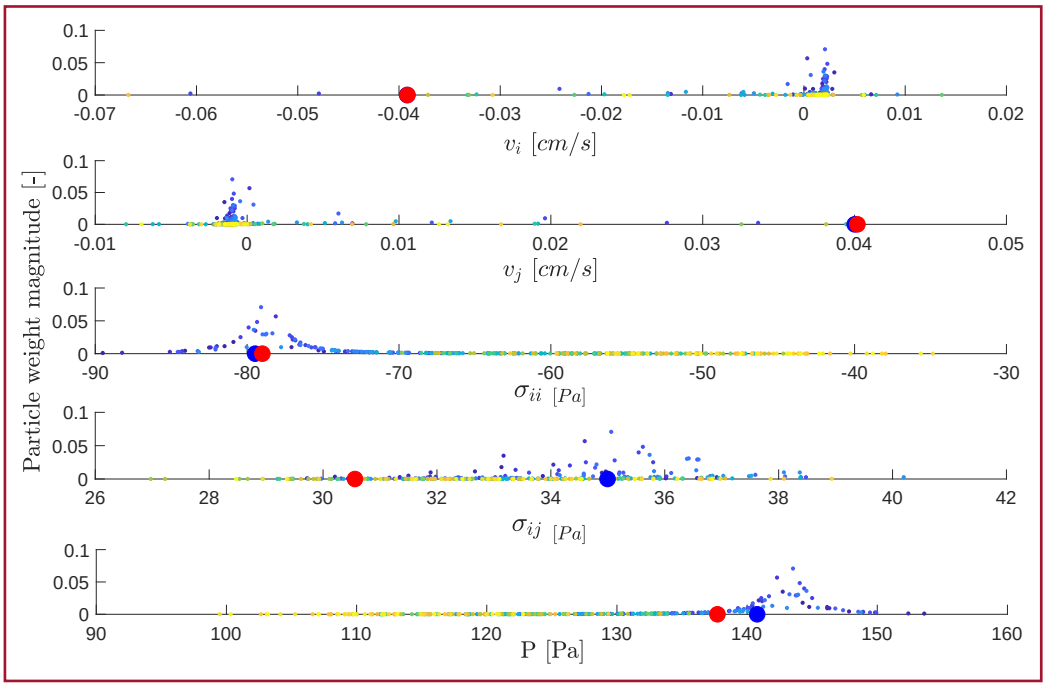

Figure 5-10: Posterior distribution of particles sampled with $\Delta t = 9.9s$ at the borehole location for assimilation step 80. Particles are colored by their order relative time in the evolution, from blue (early) to yellow (later).

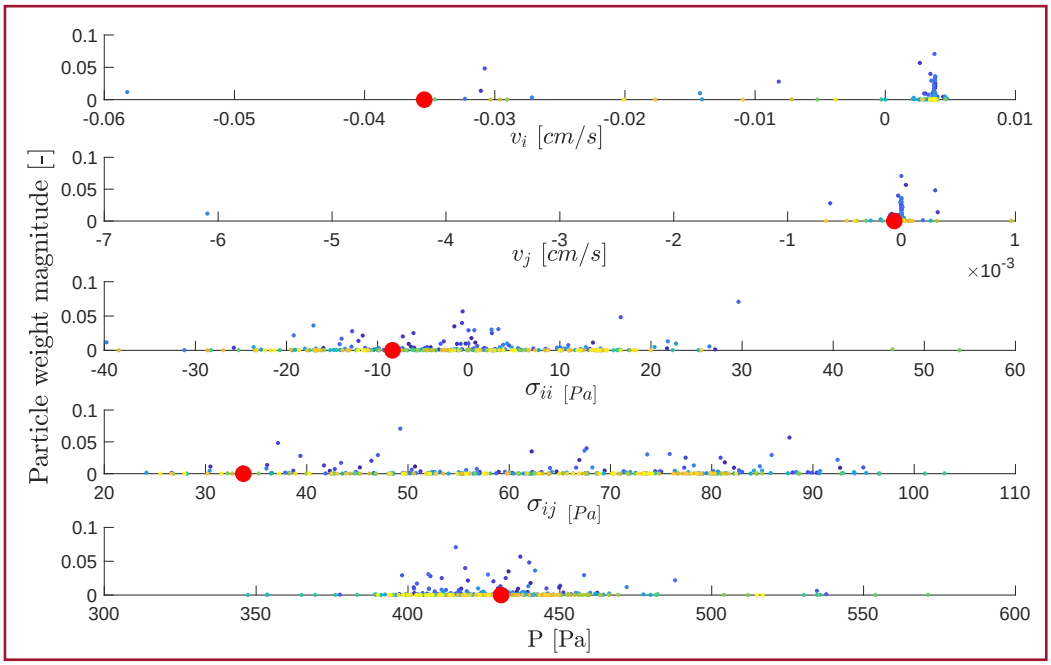

Figure 5-11: Particles corresponding to the F11 marker location and their attached weights as calculated at the borehole location (fig.5-10). Particles are colored by their order of sampling in the evolution, from blue (early) to yellow (late).

5-3 Ensemble Smoother

Generally speaking, particle filters suffer from weight degeneracy. Recall that degeneracy in particle filters is expressed as a decrease in number of efficient particle (weights) N_{eff} as data assimilation progresses. If the PF suffers from degeneracy it will be seen as concentrated weight in one or very few particles. The reason for degeneracy stems from the inevitable increase of variance in the ensemble for non-linear processes. Generating particles by sampling a single evolution with a time lag, limits particles such that they do not diverge much during assimilation and variance is kept almost constant. Although it could also mean that only small part of the model space is sampled. Figure 5-12 contains four weight diagrams which show weights for all assimilation steps using the PF (a and c) and the PS (b and d) methods for $\Delta t = 0.33$ and $\Delta t = 9.9$ s.

Diagram 5-12.(a) shows weights that are calculated at each assimilation step using the PF for time lag $\Delta t = 0.33$. Higher weights (relative to other particles) exhibit diagonal and wavy patterns. The clear diagonal patterns have repetition rate of around 20 s, which is the average time of event re-occurrence considered in section 4-1. These patterns are no longer visible when time lag is increased to 9.9 s (diagram 5-12.(c)) and weight distribution becomes more random and dispersed. Nevertheless, a transition is visible around particle 125^{th} from high and dispersed to low uniform weight. This is consistent with previous figures for which the effect of trend on weight distribution was discussed.

The particle filter as shown in figure 5-12.(a) and (c) does not exhibit any noticeable degeneracy although, figure 5-12.(c) showing the particle filter for $\Delta t = 9.9$ s can be deceiving. In this diagram the number of effective particles is relatively low with half of the assimilation steps having N_{eff} below $\sim$58 particles. Nevertheless, this is does not count as filter degeneracy since weights do not converge into a single particle.

Figure 5-12.(b) and (d) show particle weights for increasing **total** number of assimilation steps. Detailed weight diagrams consisting all assimilation steps are presented for the particle smoother in order to give further insight into the process of assimilating the seismic cycle model. As assimilation steps are added (from 1 to 100), more steps are taken into account in the calculation of the weights. However, the weights that are eventually attached to the particles are those of the **last assimilation step only**. Recall from section 3-3 that in the PS, each realization of the model has a single weight attached to it for the entire assimilation window. Since weight distribution between particles is changed from one assimilation step to the next, none of the particles "accumulate" weight as the assimilation window is lengthened. Notice how with addition of assimilation steps using the smoother, weight becomes more equally distributed between particles. Weight patterns visible at the first few assimilation steps, completely vanish as more assimilation steps are added suggesting a highly non-linear process. Distinctive particles with high weights are visible until the addition of the 10^{th} assimilation step, for which weights are becoming more evenly distributed.

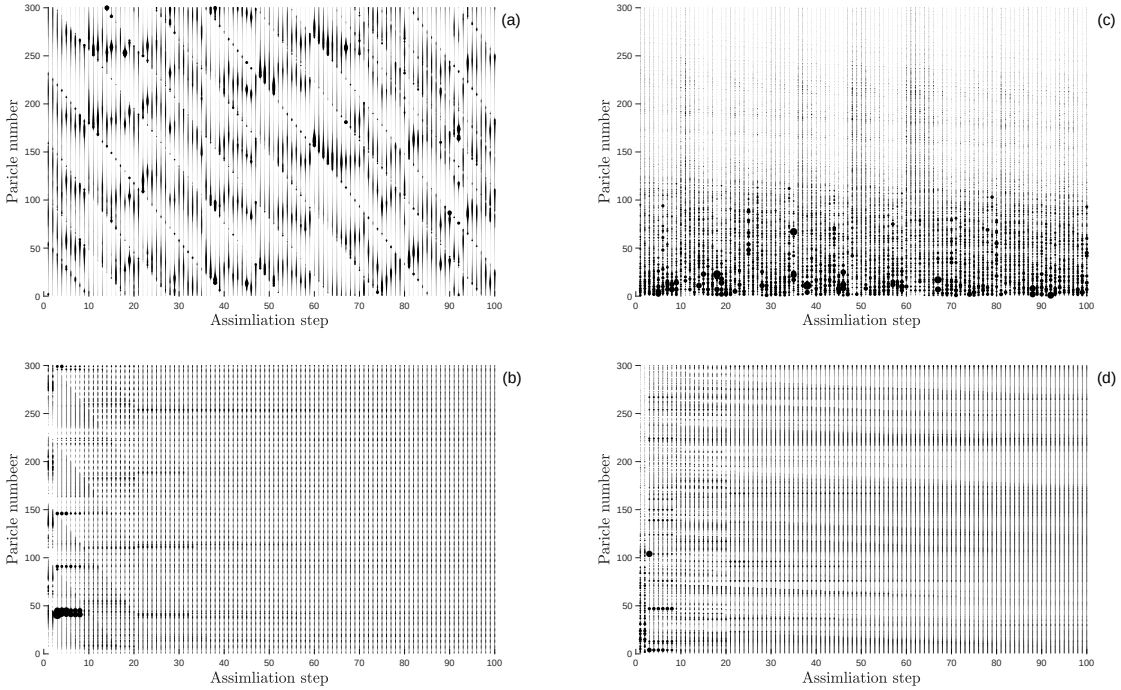

Figure 5-12: Evolution of Particles' weights with data assimilation for (a) for PF and $\Delta t = 0.33$ s (b) PS and $\Delta t = 0.33$ s (c) PF and $\Delta t = 9.9$ s (d) PS and $\Delta t = 9.9$ s. The size of the bullet correspond to its weight.

Chapter 6

Discussion and conclusions

Previous sections introduced data assimilation using an ensemble of particles, generated through sampling of state variables with a constant time lag. Assimilation of observed data with ensemble of particles has been performed using a bootstrap Particle Filter. Following an analysis of ensemble characteristic, two possible approaches were discussed: 1) small time lag for coverage of more states during the seismic cycle (resolution) or 2) large time lag for larger ensemble spread. Results of both experiments suggested that coseismic periods are not being properly captured by expected values of velocity, although results for a small time lag were showing an improvement.

In section 4-1 the effect of trends on assimilation of observed data was discussed. Plots of weight distribution at the borehole and at the F11 marker strongly suggest that the presence of trends in the observation lowers the effectiveness of the assimilation. Although sampling with a large time lag increased the spread of the ensemble and resulted in a fairly constant envelope (one of the criteria for a good ensemble), these extra values were not taken in account because of their incompatibility in σ_{ii} and P values. It was later observed for both small and large time lags that weights of individual state variables calculated into a single posterior values, might not target the closest particles to the true state for every state variable. It could indicate that the involved state variables represent a different model state than the one observed and that the ensemble still lacks the resolution. A support to the latter can be found in figures 5-6, 5-8 and B-7, where velocity distribution around the observation is very poor. An ensemble sampled using a time lag as presented here, does not seem to represent the coseismic period properly. Since the interseismic period dominates the evolution, the amount of particles representing it in the ensemble is much larger than those capturing an event for all experimented time lags. less information is fed into the system than expected due to small variability of σ_{ij} compared to the variance of the measurement error. This lack of variability is expressed in (partitioned) weight plots as a flat curve of particles with low weights. Overall these three factors: 1) Trends in the data, 2) Dominant interseismic period and 3) low variability of state variables within the ensemble, turn data assimilation of earthquake cycles (as simulated by the STM model) using a PF and time lag sampling into a hard task.

Using the EnKF, van Dinther et al. (2017) managed to estimate the same five states remarkably well. Although the model is highly nonlinear and the prior distribution does not seem to be a Gaussian, still EnKF outperformed the PF in the setup presented here. There are few major differences in the way EnKF and PF are applied: (a) van Dinther et al. (2017) generated an ensemble by running 150 models with slightly different initial conditions (markers' distribution in the cell). Here ensemble members were drawn from the same evolution with a constant time lag between particles. Particles that are sampled from different stages of the evolution are subjected to more extreme values because of accumulation of stress and pressure with time (as shown in section 4-1). Furthermore, variability within a single model run is extremely limited (as was shown for σ_{ij}) with particles possibly representing only very small part of the model space. (b) Using the EnKF, each ensemble member is analyzed according to its misfit. The misfit is multiplied by the Kalman gain which includes the measurement and model error covariance matrices that correlate the errors for the various state variables. In contrast to that, the PF weighs particles by the combination of their state variables' misfits and correlation between the various state variables is embedded in the choice of ensemble members. (c) Apart from the misfit, in the EnKF the updated ensemble members and the corresponding analyzed covariance matrix are integrated forward in time. In the PF, model errors and their integration in time are considered through the choice of ensemble which is not being actively updated. (d) EnKF uses a Gaussian approximation of the probability densities and therefore, expected values are the mean of the analyzed ensemble. PF assumes full non-linear process with non-Gaussian PDF's therefore, the expected value is a weighted average.

Considering these differences future work should be focused on better ensemble construction. One possible alternative to current sampling method is to draw particles by running them as separate models. By doing so, one introduces stochasticity into the model error represented by the ensemble and the effect of the trends in the observation location is removed without compromising the spread of the ensemble (and possibly even increasing the spread). However, this kind of implementation may result in degeneracy of the filter if particles' trajectories would continuously diverge with each assimilation step. In such case implementation might require re-sampling of the ensemble in order to reduce the variance.

Particle filtering in the current setup, does however present some significant advantages in: a) computation time (even for a very large number of particles), b) the possibility to run data assimilation off-line to the STM model. It also does not seem to suffer from filter degeneracy as discussed in sections 2-2-1 and 5-3 and appear in more detail in van Leeuwen (2009) and Arulampalam et al. (2002). Therefore I would further suggest modifications which might improve the performance of the Particle Filter:
1. Variation in borehole location and addition of observation points (multiple boreholes). This might increase the amount of data and constraints on the resulting expected value.
2. Using different combinations of state variables to assimilate the data. For example in van Dinther et al., (2017) horizontal velocity v_j exhibit very little influence on updating of shear stress.
3. Combining different model realizations using random initial distribution of markers in the cells (as done in van Dinther et al. (2017)) with time lag sampling, by sampling each separate model realization with a different time delay. The success of this approach depends heavily on the consistency in trending of different model realizations. If trends are not consistent between model runs, the effect of the trend would be removed.

4. In order to increase the number of particles with states that are within the coseismic period, one can use random time delays for sampling with an addition of a criterion that targets more states exceeding a certain threshold.

This study shows that an ensemble generated with time lagged versions of a single simulation translates into a very small part of the model space. This means that in order to obtain better results, we must increase the part of the model space that is being sampled. Possible solution is to introduce stochasticity by adding model runs as independent model realizations hence, potentially increasing the spread of the ensemble (discretized prior distribution) and removing trend effects.

References

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188.

Bannister, R. N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703):607–633.

Bayes, T., Price, R., and Canton, J. (1763). An essay towards solving a problem in the doctrine of chances.

Beers, K. (2018). Data assimilation, geomechanical parameter estimation in the groningen hydrocarbon reservoir from ps-insar measurements with a particle filter.

Brace, W. F. and Byerlee, J. D. (1966). Stick-slip as a mechanism for earthquakes. *Science*, 153(3739):990–992.

Brackbill, J. U. and Ruppel, H. M. (1986). FLIP: A method for adaptively zoned, particle-in-cell calculations of fluid flows in two dimensions. *Journal of Computational Physics*, 65(2):314–343.

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6):1719–1724.

Corbi, F., Funiciello, F., Moroni, M., Van Dinther, Y., Mai, P. M., Dalguer, L. A., and Faccenna, C. (2013). The seismic cycle at subduction thrusts: 1. Insights from laboratory models. *Journal of Geophysical Research: Solid Earth*, 118(4):1483–1501.

Crestani, E., Camporese, M., Baú, D., and Salandin, P. (2013). Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. *Hydrology and Earth System Sciences*, 17(4):1517.

Crisan, D. and Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on signal processing*, 50(3):736–746.

Doucet, A., De Freitas, N., and Gordon, N. (2001). *An introduction to sequential Monte Carlo methods.*, volume In Sequential Monte Carlo methods in practice (pp. 3-14). Springer, New York, NY.

Duretz, T., May, D. A., Gerya, T. V., and Tackley, P. J. (2011). Discretization errors and free surface stabilization in the finite difference and marker-in-cell method for applied geodynamics: A numerical study. *Geochemistry, Geophysics, Geosystems*, 12(7).

Emerick, A. A., Reynolds, A. C., et al. (2013). History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer Science and Business Media.

Evensen, G. and Van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6):1852–1867.

Geller, R. J. (1997). Earthquake prediction: a critical review. *Geophysical Journal International*, 131(3):425–450.

Gerya, T. V. and Yuen, D. A. (2007). Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems. *Physics of the Earth and Planetary Interiors*, 163 (1-4):83–105.

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. (2001). *Particle filters for positioning, navigation and tracking*. Linköping University Electronic Press.

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

Hendricks Franssen, H. J. and Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44(9).

Hoffman, R. N. and Kalnay, E. (1983). Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A: Dynamic Meteorology and Oceanography*, 35(2):100–118.

Jordan, T. H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K., and Zschau, J. (2011). Operational earthquake forecasting. state of knowledge and guidelines for utilization. *Annals of Geophysics*, 54(4).

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics*, 5(1):1–25.

Lahoz, W., Khattatov, B., and Ménard, R. (2010). *Data assimilation and information*, volume In Data Assimilation (pp. 3-12). Springer, Berlin, Heidelberg.

Lang, M. and Owens, M. (2018). A variational approach to data assimilation in the solar wind. *arXiv preprint arXiv:1806.07306*.

Liu, J. and West, M. (2001). Combined Parameter and State Estimation in Simulation-Based Filtering. In *Doucet A., de Freitas N., Gordon N. (eds) Sequential Monte Carlo Methods in Practice*, chapter 10. Springer, New York, NY. Statistics for Engineering and Information Science.

National Geophysical Data Center / World Data Service (NGDC/WDS): Significant Earthquake Database. National Geophysical Data Center, NOAA. doi:10.7289/V5TD9V7K. Accessed July 26, 2018.

Särkkä, S. (2013). *Bayesian filtering and smoothing*, volume 3. Cambridge University Press.

Scholz, C. H. (1998). Earthquakes and friction laws. *Nature*, 391(6662):37.

Skjervheim, J.-A., Evensen, G., et al. (2011). An ensemble smoother for assisted history matching. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

Stroud, J. R., Katzfuss, M., and Wikle, C. K. (2018). A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation. *Monthly Weather Review*, 146(1):373–386.

Thrun, S. (2002). Particle filters in robotics. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 511–518. Morgan Kaufmann Publishers Inc.

Van Dinther, Y. (2013). *Seismo-thermo-mechanical modeling of subduction zone seismicity*. PhD thesis, ETH Zurich.

Van Dinther, Y., Gerya, T. V., Dalguer, L. A., Corbi, F., Funiciello, F., and Mai, P. M. (2013b). The seismic cycle at subduction thrusts: 2. Dynamic implications of geodynamic simulations validated with laboratory models. *Journal of Geophysical Research: Solid Earth*, 118(4):1502–1525.

Van Dinther, Y., Gerya, T. V., Dalguer, L. A., Mai, P. M., Morra, G., and Giardini, D. (2013a). The seismic cycle at subduction thrusts: Insights from seismo-thermo-mechanical models. *Journal of Geophysical Research: Solid Earth*, 118(12):6183–6202.

Van Dinther, Y., R., K. H., and A., F. (2017). Sequential data assimilation for seismicity: Probabilistic estimation and forecasting of fault stresses. Unpublished.

Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114.

Van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653):1991–1999.

Van Leeuwen, P. J. and Evensen, G. (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, 124(12):2898–2913.

Vossepoel, F., Beers, K., Hanssen, R., and Voskov, D. (2017). On estimating geomechanical parameters from surface deformation with a particle method: Data assimilation for subsidence monitoring. 12th International EnKF Workshop, Bergen, Norway.

Vossepoel, F. C. and van Leeuwen, P. J. (2007). Parameter estimation using a particle method: Inferring mixing coefficients from sea level observations. *Monthly weather review*, 135(3):1006–1020.

Walters, P. (2000). *An introduction to ergodic theory*, volume 79. Springer Science & Business Media.

Appendix A

Ensemble generation

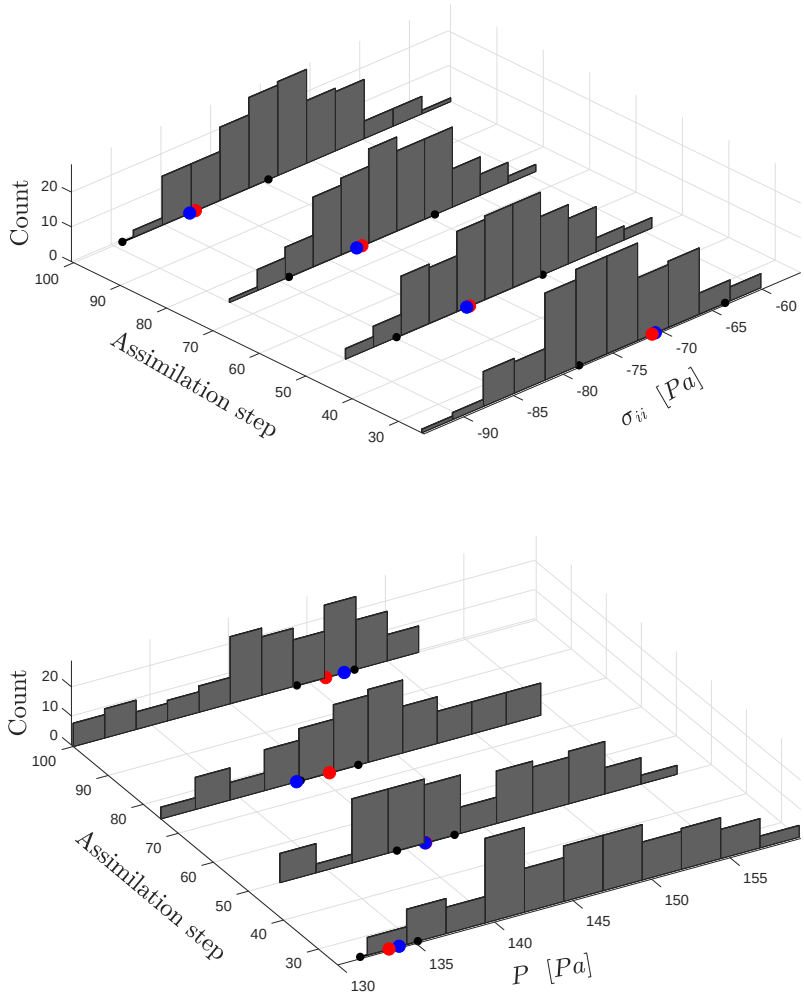

Figure A-1: Ensemble spread at different assimilation steps with $\Delta t = 0.33$ s. Red dot is the true state, blue dot is the observation (true state+error) and the two black dots connected by a bar is the variance of the measurement error. σ_{ii} could be approximated as a Gaussian however the mean of is not close to might not be close to the observation or the true state. P does not exhibit a Gaussian prior, therefore a mean would not target the middle of the distribution as expected in a Gaussian distribution.

Appendix B

Results and analysis

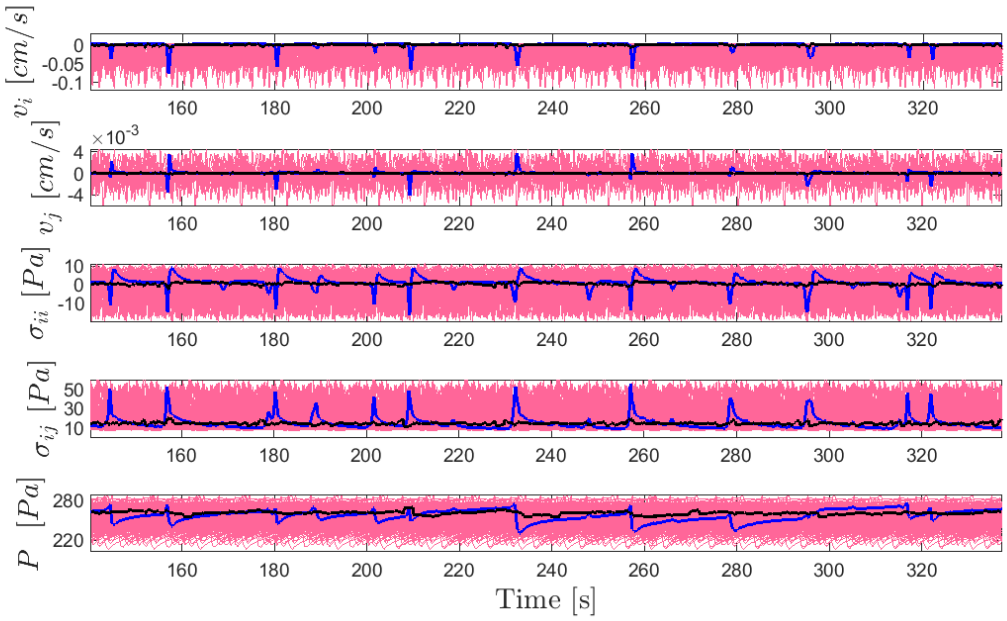

Figure B-1: Fit of expected and true evolutions in F9 marker, for experiment with 300 particles and $\Delta t = 9.9$ s.

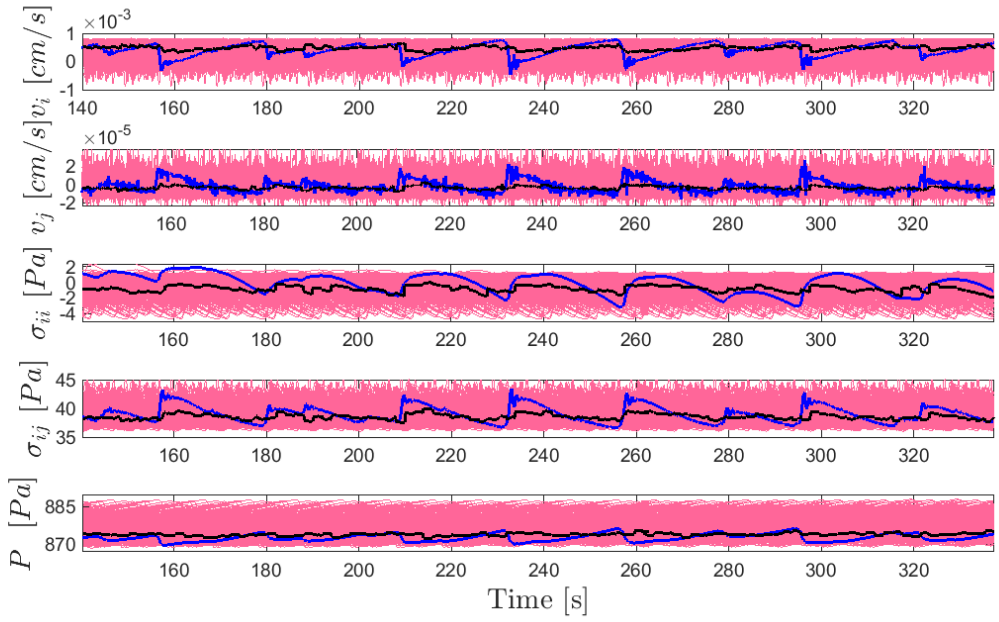

Figure B-2: Fit of expected and true evolutions in F14 marker, for experiment with 300 particles and $\Delta t = 9.9$ s. This location on the fault consistently exhibit better results than the rest of the Marker locations.

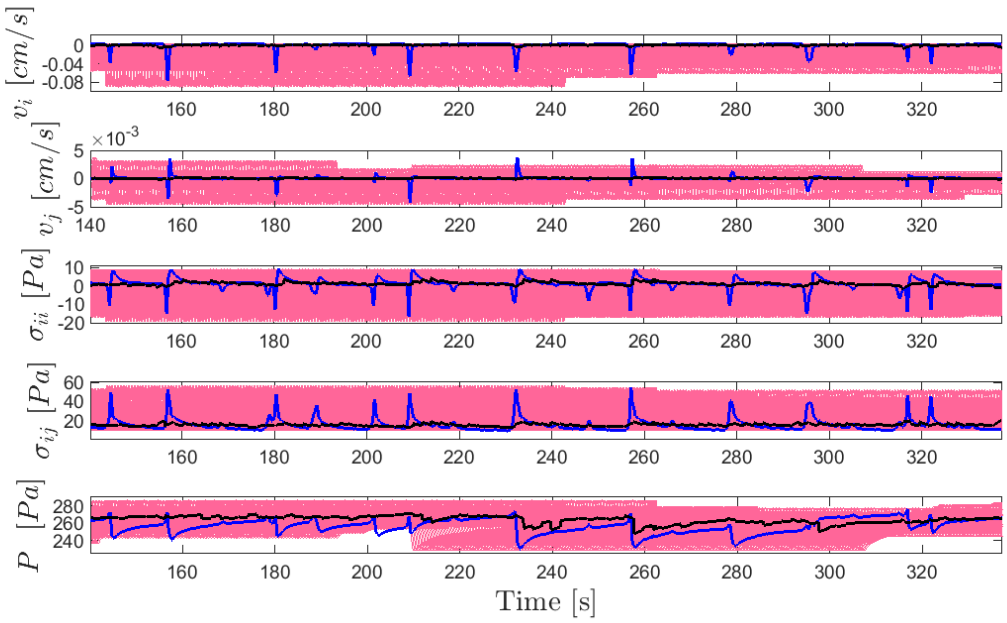

Figure B-3: Fit of expected and true evolutions in F9 marker, for experiment with 300 particles and $\Delta t = 0.33$ s.

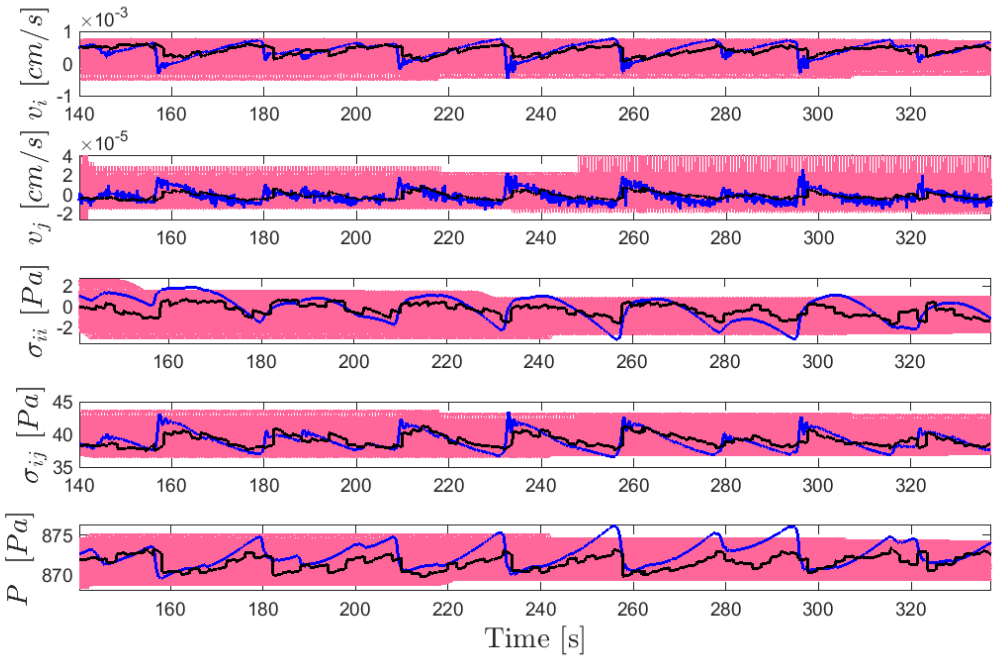

Figure B-4: Fit of expected and true evolutions in F14 marker, for experiment with 300 particles and $\Delta t = 0.33$ s. Although assimilation shows an improvement, the envelope of particles trajectory which could also be regarded as an approximated error distribution of the model, does not cover well the true evolution throughout the assimilation window.

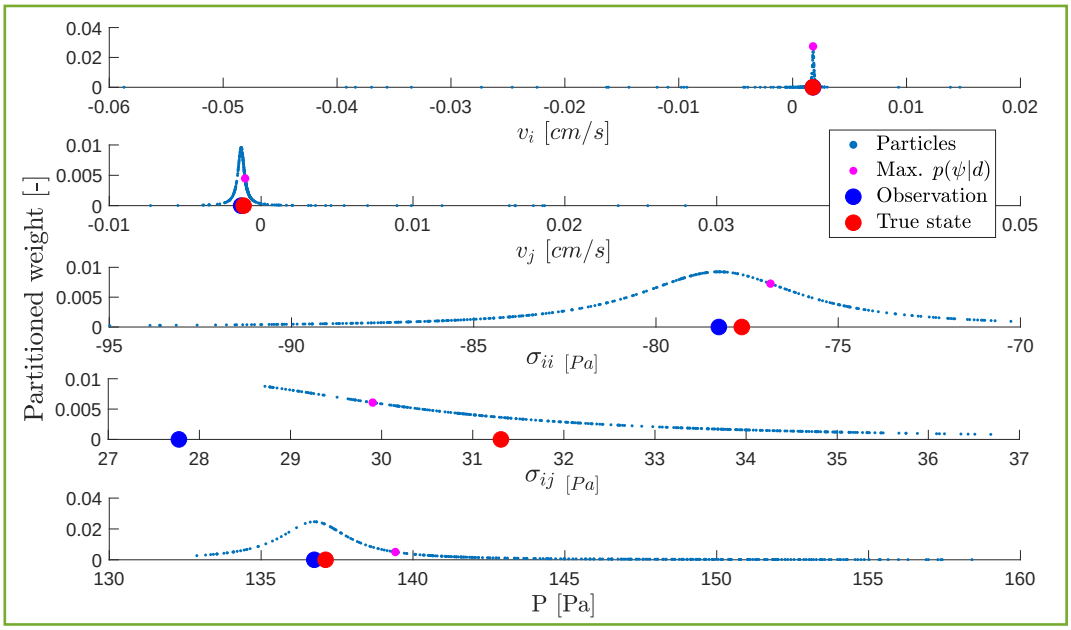

Figure B-5: Weight contribution of each state separately in the borehole for assimilation step marked by green in figure 5-4. Particle marked in purple, is the particle with the highest weight of combined states w_i.

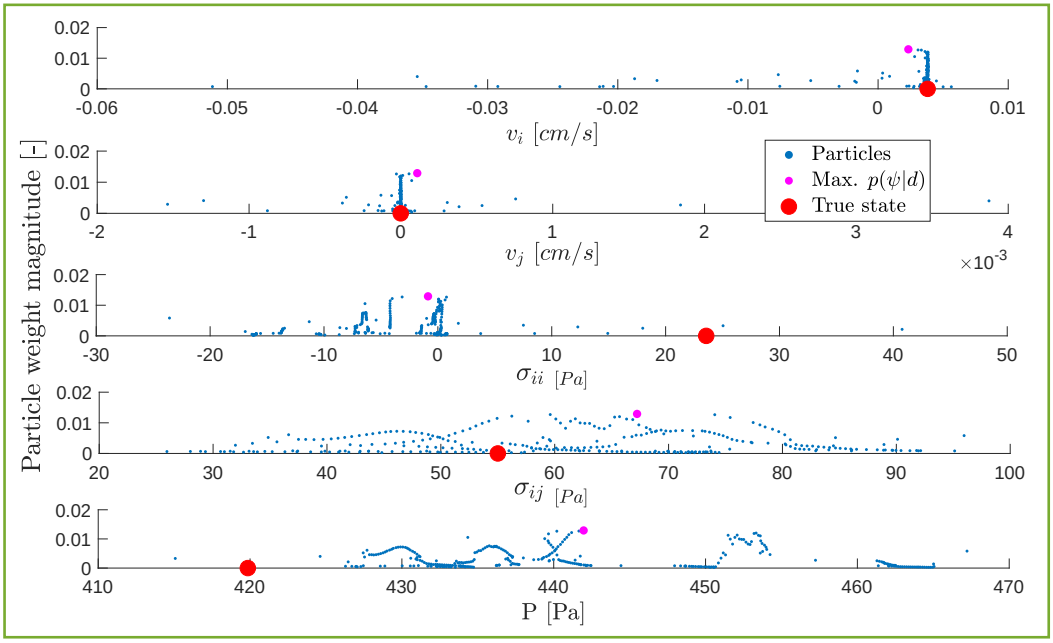

Figure B-6: Combined posterior (w_i) with which particles are weighted, for assimilation step marked by green in figure 5-4.

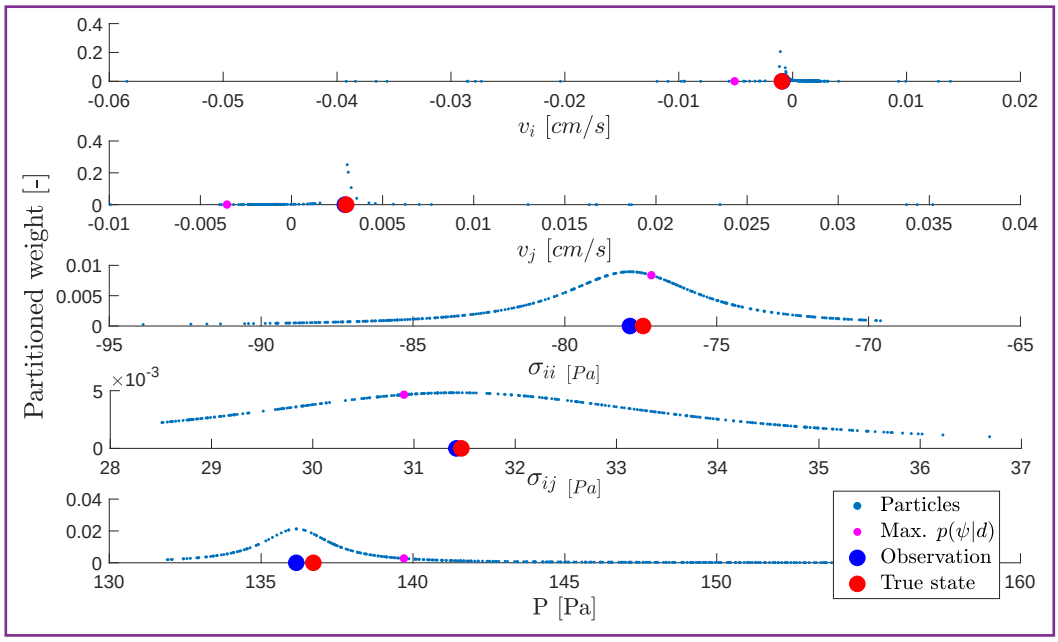

Figure B-7: Weight contribution of each state separately in the borehole for assimilation step marked by purple in figure 5-4. Particle marked in purple, is the particle with the highest weight of combined states w_i.

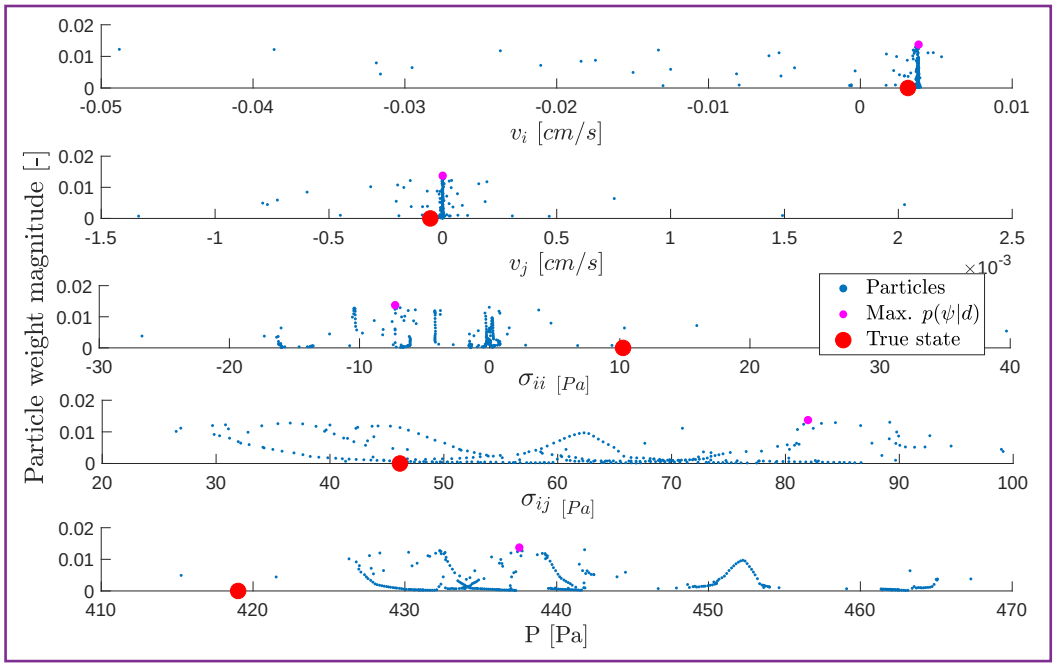

Figure B-8: Combined posterior (w_i) with which particles are weighted, for assimilation step marked by purple in figure 5-4.

Appendix C

Statistics calculation

Each of the statistical values was calculated on a matrix with the first dimension corresponding to the number of state variables assimilated (five) and the second dimension to the number of time steps in the assimilation window 'T'. For each state variable a mean value was calculated separately according to the following:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\psi_t - \bar{\psi}| \tag{C-1}$$

The same principles applies to calculating SD and RE but with the following formulas:

$$SD = \sigma = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}|\psi_t - \bar{\psi}|^2} \tag{C-2}$$

$$RE = |\frac{\psi^{expected} - \psi^{true}}{\psi^{true}}| \cdot 100\% \tag{C-3}$$