University of Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study program/specialization: Petroleum Engineering/Reservoir Engineering	Spring 2019 open
Author: Mohamed Ahmed Ibrahim Abdrabou	..... (signature of author)
Supervisor: Professor Remus Gabriel Hanea	
Title of master's thesis: Principle Component Analysis as a Method for Error Covariance Matrix Inflation	
Credits (ECTS): 30	
Key words: History Matching Ensemble Kalman Filter (EnKF) Ensemble Smoother (ES) Filter Divergence Ensemble Collapse Principle Component Analysis (PCA) Measurements Conditioning	Number of pages: 46 Stavanger,15/06/2019

ACKNOWLEDGEMENTS

First, I would like to express my gratitude and appreciation to my supervisor Professor Remus Hanea for his technical and professional guidance, his genuine support, endless patience, motivation, enthusiasm, and immense knowledge.

To Markus Fanebust Dregi, thank you for all your help. I'm grateful for all your guidance, suggestions and support with the software instalments and using.

To Amro, Fawzi, Sebastian and my classmates at the University of Stavanger for the support and encouragement pushed me forward throughout this master program, and for all the memories that made this journey an enjoyable one.

Finally, to my family in Egypt. Without you I wouldn't have been here. To my father and my mother; my role models and heroes, for my brother and my two sisters, your endless love and support made me the person I'm today. To that special girl whose been with me through sorrow and joy, thank you.

Table of Contents

Table of Figures

Table of Tables

Introduction

An oil and gas discovery goes through a multiple-stage process to increase the understanding of the asset in hand. By increasing the understanding of the field, production plan gains extra credibility and the uncertainty associated with the plan decreases. To increase the understanding of the asset, each measurement is considered an indication of the reservoir properties. These measurements are used to update the prior proposed reservoir models. This process of model update and calibration is called history matching in the oil and gas industry.

The use of several models adds the value of considering the uncertainty associated with our understanding of the reservoir, and decreases the uncertainty in the future prediction, thus field development plan and facilities design are more reliable.

The power of the ensemble-based modelling is its ability to represent various points in the possibility space. If it happens to have identical (or near identical) models, the ensemble loses its power and adds unneeded computational cost. Nonetheless, each measurement point used to assimilate the models decreases the standard deviation of the models, therefore using redundant data leads eventually to a model collapse.

During the project, the root of the redundancy was studied and methods of eliminating or reducing redundancy is discussed and its effect on collapsing the ensemble of models (filter divergence). Few methods have been used to prevent the filter divergence. In this paper we discuss the use of Principle Component Analysis (PCA) in innovation to count for the dependency associated with the measurements. Moreover, the workflow associated use PCA and its effect on the ensemble spread is presented.

Literature Review

Ensemble Based Modelling:

During the process of reservoir modelling, the initial models are created as a set of static parameters and initial values for the dynamic parameters. Static parameters are the parameters that can be assumed constant during the reservoir life; like permeability, porosity, reservoir geometry whereas dynamic parameters are the constantly changing parameters presented in pressures and saturations.

These parameters are used as input to the model while creating it, these parameters are associated with a reasonable amount of uncertainties. Uncertainties arise from:

1. Measurement tools.
2. Empirical correlations.
3. Population of facies Houtekamer and Hersche (2005)

Taking porosity as an example, porosity can be calculated using core analysis, well logging, seismic maps (McCulloh,1967). Using core analysis to measure porosity of the given rock location is perceived as the most accurate method. However, the method is associated with the inaccuracy of rock compressibility and contamination, similarly the inaccuracy associated with measurement tools.

McCulloh stated that using well logging, the uncertainty increases significantly, since the well logging tools measure a physical, electrical or chemical property of the rock then an empirical correlation is used to estimate porosity from the measured property. The uncertainty associated from the measurement tool is considerable when the depth of investigation of the tool is few meters, therefore the measured zone is mostly invaded and altered zone.

Furthermore, core analysis and logging can be obtained at few given locations in the reservoir, which leaves the data at most locations of the reservoir to be obtained using correlations and geological understanding of the reservoir geology and geometry (Luffel & Guidry, 1992). Assisted by knowledge and seismic data, geologist populate the reservoir with petrophysical parameters. However, the precision of the seismic is constrained by the seismic wavelength of few meters long. That significantly affects the accuracy of parameters obtained. (DOLBERG et. al. 2000)

Historically, the reservoir modelling process relied initially on creating one model to represent the reservoir in hand, the use of one model ignores the uncertainty associated with the model used to represent the reservoir, where the uncertainty is not quantified. Additionally, the geological understanding in the model cannot be altered. (Rwechungura et. al., 2011)

To overcome these problems, an ensemble-based simulation was introduced. The reservoir is presented with a set of different realisations. The use of several model counts for the uncertainty present in the measurement of the parameters as well as the possibility of incorporating several geological models while representing the reservoir.

History matching is an object-based assimilation. The state variables include dynamic and static parameters. The static parameters are tailored asset allocations and used for the model calibration. The dynamic parameters are not considered for adjustment in the assimilation process. Thus, history matching is a model calibration process using a state and parameter estimation.

Assisted History Matching:

With the high uncertainty associated with building reservoir models, the need for models update appeared. The aim of updating the models is to improve the models' ability to predict the reservoir behaviour. Therefore, comparing the measurements obtained from the reservoir and those obtained from the models' results provide a clear indication of the models' accuracy and validity. (Rwechungura et. al., 2011)

Production rates, pressures can be measured in the process of production, injection, etc., then compared to the same results obtained from the reservoir models. Traditionally, History matching process relied on a trial and error method of updating the reservoir model. The method utilized the use of multipliers to update static data associated with uncertainty. During the process, the values of the selected parameter (porosity, Permeability, etc) is multiplied by the same constant number for all grid cells of the whole reservoir or in a given region of the reservoir. (Rwechungura et. al., 2011)

For Example, at a given reservoir where the uncertainty is defined to be associated with the permeability. If the oil and water production rate measurements obtained are observed to be 200 m^3/day and 400 m^3/day respectively, while the reservoir model produced oil and water production rates to be 150 m^3/day and 300 m^3/day respectively. Therefore, the permeability is multiplied by a factor of

1.5, 2, 3 in a trial and error process. By observing the new obtained results, the most accurate multiplier is selected, and the reservoir model is updated.

However, the problem of history matching is a basic problem of data assimilation where the static model parameters and initial values of dynamic model parameters are updated continuously with the measurements taken from the reservoir production data. Statistically, the update of given model from obtained measurements allows the use of Kalman filter method. Sebacher et.al. 2013 stated that Kalman filter is a Monte Carlo technique, it assesses the density of probabilities by using a set of simultaneously updated realisations. Therefore, the physical knowledge and understanding of the geology is honoured.

Kalman Filter

Mandel (2009) stated that Kalman filter is a recursive data processing algorithm. Kalman filter has the ability to incorporate all the data provided to it. The data provided to the algorithm is:

1. Initial condition Information
2. Knowledge of the system and the measurement device dynamics
3. The statistical understanding of the noises associated with the measurements and the model.

The Kalman filter can be described by the equations

$$x_k = Ax_{k-1} + w_{k-1} \quad \text{EQ 1}$$

$$z_k = Hx_k + v_k \quad \text{EQ 2}$$

where X is the state Vector, A is the State transition matrix from state at time k-1 to time k, w is the white noises of the process, z is the actual measurements and H is the connection between the measurement and the state vector (physical model), and v is the measurement white noise. (Welch & Bishop, 2006)

Welch and Bishop mentioned, Kalman filter updates the state vector through a process of updating a prior state ($\hat{x}_k^-$) into a posterior state (x_k). The estimate errors on the prior state and posterior state are

$$e_k^- = x_k - \hat{x}_k^- \quad \text{EQ 3}$$

$$e_k = x_k - \hat{x}_k \quad \text{EQ 4}$$

where the covariances of the prior and posterior estimates are:

$$P_k^- = E\left[e_k^- e_k^{-T}\right] \quad \text{EQ 5}$$

$$P_k = E\left[e_k\, e_k^T\right] \quad \text{EQ 6}$$

We then define the Kalman filter equation as an equation to estimate the posterior State vector ($\hat{x}_k$) from the prior estimated state vector (x_k^-) and the difference between the actual measurements (z_k) and the estimated measurements of the model .

The Kalman Filter equation is then defined as

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \quad \text{EQ 7}$$

where K is a matrix defined as Kalman Gain calculated by:

$$K = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad \text{EQ 8}$$

where R is measurement error covariance matrix.

Using these equations of Kalman filter, a static model can be continuously updated as new measurements are obtained. With each update the state vector (x_k) is expected to gain more accuracy.

Hence, the Kalman filter includes the physical model (H) and uses the measurements to update the model parameters. Furthermore, it respects the uncertainty associated with the measurements and the model dynamics by including noises. However, the Kalman filter works under a set of assumptions:

1. The system is linear
2. Whiteness of the noises, which means the noises on the measurements are not time dependent
3. The model is Gaussian

In oil and gas field, the Kalman filter problem is defined where the State vector is populated with the data to be updated, usually containing static parameters of the reservoir (porosity, permeability, etc), initial conditions of dynamic parameters of the reservoir (pressure, phase pressures, saturations, etc)

The measurements vector is populated with field measurements of borehole pressures, borehole phase pressures, production rates and phase production rates.

However, the Kalman filter is used to update a model problem where the measurements are a linear equation of the model equation. The reservoir model is highly nonlinear. That led to the creation of the ensemble Kalman filter method, the use of random distribution to create a set of models to be updated.

Hence, the state vector changes into a matrix representing a gaussian field distribution of the state Vector. Since a gaussian field is linear, solving the Kalman filter equation for a gaussian distribution of field parameters (state vector) solves the linearity problems. (Mandel, 2009)

Ensemble Kalman Filter

As the relation between the measured parameters in the reservoir field and the state vector is highly nonlinear, the Kalman filter equation needed modification to handle the nonlinearity in the readings. Hence the Extended Kalman Filter **(EKF)** was established to overcome the problem. (Welch & Bishop, 2006)

Like Taylor Series, the EKF uses the partial derivative of the process and the measurements around the current estimate to produce the prior estimation. If the State vector is calculated by the equation

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \quad \textit{EQ 9}$$

$$z_k = h(x_k, v_k) \quad \textit{EQ 10}$$

where the state vector is nonlinear equation of the previous step state vector. Furthermore, the measurement is a nonlinear equation of the state vector. The Extended Kalman filter linearizes the equations where it transforms into:

$$x_k \approx \acute{x}_{k-1} + A(x_{k-1} - \hat{x}_{k-1}) + Ww_{k-1} \quad \textit{EQ 11}$$

$$z_k \approx \acute{z}_{k-1} + H(x_{k-1} - \acute{x}_{k-1}) + Vv_{k-1} \quad \textit{EQ 12}$$

where $\acute{x}_{k-1} = f(x_{k-1}, u_{k-1}, 0)$ is the approximate state measurement, where the model noise is estimated to be 0 similarly the approximated estimated measurements are $\acute{z}_k = h(x_k, 0)$

Additionally, A, W, H, V are the relation conversion matrices which is obtained by the partial derivates.

$$A_{[i,j]} = \frac{\partial f_i}{\partial x_j}(\hat{x}_{k-1}, u_{k-1}, 0) \quad \textit{EQ 13}$$

$$W_{[i,j]} = \frac{\partial f_i}{\partial w_j}(\hat{x}_{k-1}, u_{k-1}, 0) \quad \textit{EQ 14}$$

$$H_{[i,j]} = \frac{\partial h_i}{\partial x_j}(\acute{x}_{k-1}, 0) \quad \textit{EQ 15}$$

$$V_{[i,j]} = \frac{\partial h_i}{\partial v_j}(\acute{x}_{k-1}, 0) \quad \textit{EQ 16}$$

Then EKF uses EQ (17), EQ (18) to predict the prior state vector $(\hat{x}_k^-)$ and the prior error covariance (P_k^-)

$$\hat{x}_k^- = f(\hat{x}_{k-1}, u_{k-1}, 0) \quad \textit{EQ 17}$$

$$P_k^- = A_k\, P_{k-1} A_k^T + W_k\; Q_{k-1} W_k^T \quad \textit{EQ 18}$$

where Q is the process noise covariance.

Then the posterior state vector $(\hat{x}_k)$ and the posterior error covariance (P_k) are updated using the EQ (19), EQ (20), EQ (21)

$$\hat{x}_k = \hat{x}_k^- + K_k\,(z_k - h(\hat{x}_k^-, 0)) \quad \textit{EQ 19}$$

$$K_k = P_k^- H_k^T (H_k\, P_k^- H_k^T + V_k\; R_k\, V_k^T)^{-1} \quad \textit{EQ 20}$$

$$P_k = (1 - K_k\, H_k)\, P_k^- \quad \textit{EQ 21}$$

However, the extended Kalman filter is faced by the problem of closure, where the Jacobian (H_k) in the Kalman Gain equation (K_k) can only propagate the relevant component of the measurement information only. Hence, if the mapping between the measurements (z_k) and the state $(\hat{x}_k)$ through function (h) are not one-to-one, the Jacobian (H_k) creates a Kalman gain where only the residual $(h(\hat{x}_k^-, 0))$ portions that affects the state is updated .Therefore, if the relationship between the measurements and the state isn't one to one the process diverges quickly. That led to the use of the Ensemble Kalman Filter **(EnKF).**

The ensemble Kalman filter has been used in several processes, it has gained high popularity due to its benefits:

1. Ease of use and implementation
2. No tangent linear operators
3. Relatively low computational cost (Mandel, 2009)

The ensemble Kalman Filter can be used in non-linear models since it depends on the linearity of randomly distributed space of state vector and measurements to establish the linearity of the equation.

During Ensemble Kalman filter, a state matrix (A) of models is created from randomly distributed state vectors (X), in which each column in the state matrix is a state vector representing a model in the given ensemble of models. The state vector of each model contains a set of parameters to be updated by the filter, each parameter represent a physical property of the reservoir at a given grid cell. Therefore, in a given row in the state matrix, the values of a parameter (porosity or permeability, etc) in a given model location is randomly distributed. (Mandel, 2009)

Similarly, the measurements vector (d) and error vector (e) are replaced with a measurement matrix (D) and error matrix (E) which is a collection of perturbated measurement vectors and error vectors simultaneously.

$$A = (x_1, x_2, x_3, x_4, \dots\dots\dots, x_N) \in R^{nxN} \quad \textit{EQ 22}$$

$$D = (d_1, d_2, d_3, d_4, \dots\dots\dots, d_N) \in R^{mxN} \quad \textit{EQ 23}$$

$$E = (e_1, e_2, e_3, e_4, \dots\dots\dots, e_N) \in R^{mxN} \quad \textit{EQ 24}$$

For an ensemble of N models, where n is the number of parameters in a model state vector (state vector size) and m is the number of measurements.

The covariance of the ensemble can be defined as

$$P^e = \frac{1}{N-1} A'(A')^T \in R^{nxn} \quad \textit{EQ 25}$$

$$R^e = \frac{1}{N-1} EE^T \in R^{mxm} \quad \textit{EQ 26}$$

where (A') represent the ensemble perturbation, (A) – mean (A). Therefore, the ensemble is updated by the equation:

$$A^a = A + P^e H^T (HP^e H^T + R^e)^T (D - HA) \quad \textit{EQ 27}$$

where (H) is the measurement operator. (H) is a linear operator that selects the predicted measurements from the ensemble by the help of the covariance matrix.

Therefore, the equation of EnKF can be written as:

$$S = HA' \quad \textit{EQ 28}$$

$$C = SS^T + EE^T = SS^T + (N-1)R^e \approx SS^T + (N-1)R \quad \textit{EQ 29}$$

$$X = I + S^T C^{-1} (D - HA) \quad \textit{EQ 30}$$

$$A^a = AX \qquad \text{EQ 31}$$

where $S \in R^{mxN}$ represents the ensemble perturbation, $C \in R^{mxm}$ represents the error on measurements using the ensemble perturbation and error covariance matrix.

The method is shown to be easily implementable, since the function doesn't rely on the type of process rather it is statistical. However, the algorithm respects the physical realm of the problem, since the measurements are both obtained through physical simulation of the reservoir prior ensemble.

Additionally, the EnkF is regarded as relatively computationally feasible method. Since EnKF is a step updating method, therefore the data stored from on step to another are represented in the ensemble of models whereas the uncertainty is propagated through the covariance matrix. Enkf avoid the storage of all the data from time zero by including the dynamic data of the reservoir in the state vector of each model. Hence, we include in the state model the static parameters of the reservoir typically porosity, permeability or any uncertain static parameters in the reservoir whereas the dynamically related reservoir parameters typically pressure of each of each grid cell, saturation, gas-oil ratio are not required to be stored at each simulation time. Therefore, it avoids storing a large amount of data which makes it computationally feasible.

Kalman smoother

EnKF provides the advantage of not storing large amount of data while assimilating the ensemble. However, it required restarting the assimilation process from time zero at each history matching time. Therefore, at larger reservoir models or reservoir models where the upscaling must be limited, the process of assimilation can be more computationally expensive than desirable. Hence the ensemble smoother (ES) is introduced by Leeuween and Eversen 1996. ES doesn't assimilate data sequentially in time, therefore it avoided rerunning the process of assimilation from time zero at each update step.

EnKF was found to yield higher performance in ocean circulation model application, van Leeuewen 1996. Nevertheless, Van Leeuwen and Evensen 2000, used Lornez equations to compare the ES and EnKF and has confirmed that EnKF yields higher performance since the recursive update of EnKF keeps the ensemble near to the true parameters. Furthermore, EnKF was shown to be equivalent to applying one Guess-Newton iteration with a full step and sensitivity matrix at each data assimilation time step. However, the sensitivity matrix is replaced by an average sensitivity matrix as discussed early. The

average sensitivity matrix is obtained from the ensemble itself. Hence EnKF updates the state matrix in a closely spaced time steps compared to the ES process which uses a one-time update of the state matrix. ES in that sense can be viewed as a single Gauss-Newton correction. Hence, it can be explained that the EnKF is reasonably of higher performance compared to the ES in history matching of reservoir.

Contrarily, ES and EnKF has been compared by Skjervheim et. al. (2011), it was concluded that both methods produce similar results with application of history matching of oil and gas reservoirs. That resulted in the clear advantage of the ES due to the highly reduced computational cost since ES avoid restarting the reservoir simulator at each update step. Furthermore, Skjervheim et. al. (2011) reported that the ES was 10 times faster than EnKF in the reservoir history matching applications.

Since the introduction of ES in 1996 by Leeuween and Eversen, many ensemble-based smoother algorithms have been introduced. Emerick and Reynolds 2012, has conducted multiple data assimilation (MDA) explained as performing the assimilation multiple times using the covariance of the measurement errors multiplied by the number of data assimilations (iterations). It concluded that single data assimilation (SDA) and multiple data assimilation (MDA) are the same for linear Gaussian case. Similarly, it was concluded that MDA can improve the estimates and data matching of EnKF in time-lapse seismic data.

Eversen, Eikrem (2018) stated that using the ES has practical advantages over EnKF including:

- During ES update the model update takes place in one step rather than sequentially, resulting in a reduced computational cost.
- The measurements aren't assumed to be independent, therefore a time dependent measurement error can be modelled.
- The dynamic data of the model (ex: Pressure, Saturation) are not updated, therefore the possibility of having unphysical model is eliminated which could hinder the simulator.
- The parameter update problem is easier to understand, since it is similar to other history matching software products.

However, it introduces challenges:

- The amount of measurements used in one update step can be extremely large. Since all measurements are simulated on simultaneously rather than sequentially as in EnKF case.

- During EnKF update the measurements are assumed to be time independent, often leading to filter divergence. Using ES it is possible to define the time-dependency of the measurements, however that leads to the creation of the nondiagonal measurement error covariance matrix which can cause computational problem in inversion while using large number of measurements.
- The possibility of including redundant data can create ill-conditioned inversion. While using ES it is possible to include similar measurements over short span in time which presents redundancy in the measurements. Contrarily in EnKF, the inversion takes place sequentially over small matrices which resolves the problem.

Skjervheim et.al (2011) conducted an experiment to compare EnKF to ES for model update where they used synthetic model and field application. They concluded that both EnKF and ES achieve an improved match for the production measurement where the uncertainty is reduced. However, ES was used with one update step where the estimated uncertainty was larger, covering higher space of possibility, giving ES the advantage of delayed filter divergence compared to EnKF. The difference was explained as EnKF uses the measurement recursively, therefore, ES reduces the impact of spurious correlations.

There have been different methods for implementing ES method. Using the standard ES method, ES with Multiple Data Assimilation ESMDA, and Iterative ES (IES) are different options. Similarly, Chan and Oliver (2012) proposed the use of Randomized Maximum Likelihood (EnRML). But rather than using all inflation coefficients (a_i) equal to the number of data assimilations, ENRML uses the inflation coefficients, (a_i), in a decreasing order resulted in only small improvements with further iterations.

Emerick and Reynolds (2012) has introduced ESMDA as an efficient iterative form of ES. It executes an iterative ES updates by dividing the measurement innovation by a higher factor moving from iteration 1 to iteration n where the sum of the inverted weighting (division) factors is 1.

Chen and oliver (2012) have conducted a study to compare the EnKF, ES, sequential EnRML and batch EnRML for a small five-spot problem and Brugge field case (Peters et al., 2010), where they suggested the use of EnRML. They concluded that the use of ES often requires iterations to achieve satisfactory matches. They stated that batch EnRML required 26 iterations to achieve a lower data mismatch objective function than the standard EnKF which conclude that the cost of batch EnRML cost is higher

than EnKF. Emrick and Reynolds (2012) has pointed out that Oliver and Chan have focused on used pseudo-observations to condition the ensemble rather than using actual observations multiple times.

Emerick and Reynolds (2012) concluded that ESMDA requires few modifications to the standard ES therefore it can be used generally as well as in reservoir engineering applications. It was concluded that ESMDA yields improved results compared to ES and EnKF at comparable computational cost while having less computational cost and better results over batch EnRML.

Measurements Conditioning

The purpose of the ensemble-based modelling techniques is to establish a method that honours uncertainty in the created models, therefore a bigger space of possibilities covered by the ensemble is considered a great advantage to the technique. By updating the ensemble of models using one of the assimilation techniques discussed before, the accuracy of the ensemble increases, and the space of possibilities narrows. However, the accuracy of the ensemble is measured by its mean's ability to predict future behaviour. Thus, it is beneficial for the ensemble to maintain a higher variance to cover more possibility spaces.

A problem that often happens using ensemble-based modelling is ensemble collapse where the ensemble collapses into one model, denying the technique of its advantage. Many Techniques has been suggested to prevent ensembles from collapsing (filter divergence).

Localization is a concept that has been used to overcome filter divergence problem. The concept of localization is based on the hypothesis that the obtained real fields measurements are influenced only by a defined area around the wells, therefore the update process is constrained to a certain space around the wells (Jung et.al ,2018).

Therefore, Localization is a weight function applied for the error covariance. Thus, it solves the filter divergence problem by increasing the degree of freedom of the measurements.

There have been two main techniques suggested for localization:

- Distance Cut-off
- Streamline Localization

Houtekamer and Mitchell (1997) have studied the distance cut-off by identifying an optimum cut-off radius to select data to be updated from the ensemble. The technique uses a radius around the producing wells of a circle shape where within the circle, parameters are updated. The shape of the cut-off zone can be altered from a circular shape depending on the reservoir heterogeneity (Emerick, Reynolds,2011).

Watanabe and Datta-Gupta (2012) proposed the use of water, oil and gas phase streamlines as means for covariance localization for water-cut and GOR data assimilation, using streamline analysis for each phase to decide the zones that influence the flow of each well. This method of localization is favourable due to its physically sound basis. Furthermore, the cut-off area is determined dynamically which provide an optimum update zone. Nonetheless, the technique offers multiple computational challenges while applying it. The computational cost of determining the flow zone of each model in the ensemble prior to each update step. Additionally, the population of the covariance matrix dynamically based on the results of the streamline simulation increases the computational challenges.

Another Technique for avoiding (delaying) filter divergence is by conditioning measurements. As mentioned before, the use of redundant data as measurements result in ill-conditioned inversion while using ES. Similarly, the use of redundant data with EnKF technique leads to overshooting, where the ensemble is over fitted into the same measurements leading to ensemble collapse (filter divergence).

Using measurements in the ensemble-based techniques is based on assumption that the measurements are independent. However, some measurements are dependent and other are redundant.

Eversen, Eikrem (2018) defined dependent measurements as measurements where its likelihood can be defined as the product of the separate likelihood of both products

$$f(d|x) \neq \prod_i f(d_i \vee x) \qquad \textit{EQ 32}$$

where (d) donates the measurements, (i) donates the index of the measurement and (x) the static model parameter.

They also defined redundant measurements, as measurements that contain a redundant amount of information. By obtaining the rate flow in a well over an infinitely short period of time, the measurements contain significant redundancy. In their paper, they proposed the use of the total accumulated production of oil, water and gas for each well at the final time of history. The results show improvement in the spread of the ensemble where it covers larger area in the probability space.

Eversen (2004) has studied the effect of square root scheme (as sampling technique) on the quality of the EnKF when applied to initial ensemble and measurement error perturbation. The square root scheme originally introduced by Anderson (2001), Whitaker and Hamil (2002) and Bishop et. al. (2001) has been simplified in their paper. In their modification of the square root scheme, they avoid the perturbation of the measurements. Additionally, the scheme avoids the approximation imposed by the assumption of uncorrelated measurements error or the knowledge of the inversion of the error covariance matrix by replacing the inversion of the error covariance matrix (C) with a low rank covariance matrix $(\mathrm{C^e})$.

$$C^e = SS^T + (N-1)R^e = SS^T + EE^T \qquad \textit{EQ 33}$$

It was concluded that by using a large initial ensemble that covers higher space span, the low rank derivation of the ensemble error covariance matrix (P) can be an accurate representation of the full rank ensemble error covariance matrix $(\mathrm{P^e})$. $(\mathrm{P^e})$ was calculated by using singular value decomposition (SVD).

$$P^e = \frac{1}{N-1}A'(A')^T = \frac{1}{N-1}U\Sigma^2U^T \qquad \textit{EQ 34}$$

Additionally, they suggested improving the rank of the matrix by ordering the first N singular vector in (U) to be similar to the first N eigenvectors in the in (Z), where (Z) is obtained from the eigenvalue decomposition of (C).

$$Z\Lambda^2Z^T = C \qquad \textit{EQ 35}$$

Similarly, by applying SVD on a large initial measurement perturbation, an analysed measurement error perturbation is produced where the accuracy of the full rank error perturbation matrix is preserved while using low rank error perturbation matrix.

However, for an ensemble covariance matrix with large dimensions, the matrix falls into the risk of being singular, therefore the use of low rank error perturbation matrix with a pseudo-inversion for the ensemble covariance matrix was used. They concluded the use of the square root scheme as a replacement for Monte Carlo method for creating perturbation ensemble and measurements matrices result in a larger spread of the ensemble.

Similarly, Luo and Hotett (2013) investigated the effect of singularity associated with the large ensemble size. They stated that to solve the problem of singularity, covariance localization (Hamill et al. 2001) can be introduced to the ensemble error matrix in order to increase its rank (Hamill et al. 2009) or using

hybrid of sample error matrix with full rank matrix to replace the ensemble error matrix. Alternatively, reducing the dimensions of the observation in the update formula, for instance, by assimilating the observation in a serial way or by assimilating the observation within the framework of a local EnKF.

Furthermore, they suggested the use of inflation factor to the covariance matrix to solve the singularity problem, hence solving the issue of filter divergence.

Principle Component Analysis

Principle Component Analysis (PCA) is a mathematical method used to reduce the dimensionality of a given set of data. PCA is used for a large set of data where the variables at different dimensions are interrelated, the result is a reduced set of data at a reduced set of dimensions. The resulting reduced data set maintains as much information as possible. The new set of data are called principle components. (Principal Component Analysis, 2002)

The resulting principle components are a result of a reduction that presents the set of uncorrelated variables, the principle components are ordered in a manner allowing the first few components to retain the most of variation existing in the original large set of data. Therefore, PCA reduction reduces the problem into eigenvalue- eigenvector problem for a positive semidefinite symmetric matrix. (Principal Component Analysis, 2002)

Eigen Vectors and Eigen Values

To explain the Eigen Vector, a mathematical example will be used

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \; x \; \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \; x \; \begin{pmatrix} 3 \\ 2 \end{pmatrix} \qquad \textit{EQ 36}$$

Eigen Vector is a special case of matric multiplication. In Eq. 28, the matrix is multiplied with a vector.

The vector (3,2) is a vector that originate from the origin point (0,0) to the point (3,2), the result of the multiplication is a vector (12,8) this vector is a direct multiplication of the original vector (3,2) meaning it is on the same axis of the original vector. In this case the vector (3,2) is an Eigen Vector to the matrix. (Smith, 2002)

The Eigen Vector has a set of properties:

- Eigen Vector can be found only for square matrices
- The magnitude of the eigen vector is irrelevant, therefore it is always possible to use the unit value scale of the eigen vector
- All eigen vectors of a matrix are perpendicular

Furthermore, Eigen values are closely related to the eigen vector. In Eq. 28, the eigen value is 4. The eigen value 4 is associated with the eigen vector (3,2). In other words, by taking any scale of the eigen vector (3,2) the result of the multiplication will be 4 times the scaled vector used in the multiplication. (Smith, 2002)

PCA

Principle Components are obtained from PCA by linear combination of variables that express maximum covariance. PCA is achieved by mathematical steps:

1. Find the mean of the variables (data)
2. Subtract the mean to obtain the data adjust
3. Calculate the covariance matrix
4. Calculate the Eigen vectors and eigen values of the covariance matrix

   The Eigen vectors in this case represent the correlation between the different dimensions in the data set, were each vector draws a pattern of corelated change.
5. Order the eigen vectors by the help of the eigen values

   The relevance of the eigenvectors isn't equal, some eigenvectors represent the data more than others, therefore using the eigen values, eigen vectors can be ordered highest to lowest. The eigen vector that corresponds to a higher eigen value carries higher significance thus represents the data more closely.

   By selecting a cutoff for the eigenvectors, a Row Feature Vector can be constructed which is a matrix built from the eigen vectors selected.
6. New data set is derived

   Final Data = Row Feature Vector x Row Data Adjust (Smith, 2002)

Methodology

As mentioned above one of the methods involved in avoiding filter divergence is getting rid of dependency and redundancy in the measurements used.

In this project, we introduce the principle component analysis in innovation (PCA). As mentioned above EnKF, ES and ESMDA relay on the assumption that the observations are independent, and they are not correlated. However, this assumption is not true. By analysing the measurements, it is observed that the measurements are correlated in space and time.

For instance, considering the production rate at three wells present at a homogenous reservoir. The first two wells are 500 meters apart while the third well is 3 km far from both wells. The production rates are the first 2 wells are subject to similar reservoir rock conditions compared to the third well. Therefore, the dependency on the measurements exist and not uniform. Similarly considering the time difference between measurements taken at a given time, the measurements are more dependent or contain more redundancy as the time span between the measuring time decreases. (Eversen & Eikrem, 2018).

By using the total phase production rates at the end of the production time, Eversen & Eikrem (2018) has counted for redundancy presented in time between measurement. However, they didn't consider the dependency presented in measurement due to the distance and the reservoir nature between wells.

To account for the dependency and redundancy in the measurements, one needs to populate observation error correlation matrix. In turn the matric will not be diagonal which can lead to an increased computational complexity. However, at the time there is no known way to effectively and automatically populate the matrix with the dependency factor.

The suggested workflow in the PCA created an inflation factor for the errors associated with the measurement to counter for the overlying dependency in the measurements. The inflation factor is calculated automatically and based on the original error covariance matrix. Below is the workflow to describe the algorithm of the method.

Step 1: using measurements vector (d), matrix D is created (d_1, d_2, d_3, ……., d_N). Similarly, using measurement error vector (e) matrix E (e_1, e_2, e_3, ……., e_N), simulated results Matrix from ensemble (HA'), Prior matrix of perturbed results is created X.

$$X= (HA'-D)/E \qquad EQ\ 37$$

Step 2: Using X, the measurement error covariance matrix (R) is calculated.

$$R = X - \bar{X} \qquad \textit{EQ 38}$$

Step 3: Applying singular value decomposition on R:

$$USV = R \qquad \textit{EQ 39}$$

Step 4: Calculating the number of PCA coefficients (n).

- Calculate the sum of Eigenvalues from the Singular values S. T= $\sum S^2$
- Define cutoff coefficient c.
- PCA No. is the number n of S^2 values (n) where $\sum_i^n S < \sum S^2 * c$

Step 5: Calculate inflation factor $I = \sqrt{\frac{no\ of\ observations}{n}}$ *EQ 40*

Finally, we use the inflation factor to inflate the measurement error covariance matrix. The cut off has minimal effect of the values since the earliest eigen vectors carry the highest relevance in representing the error covariance matrix. Therefore, using the PCA represents the value of correlation between errors and the method encounters that by inflating the error to compensate for the covariance between the errors in the matrix.

To experiment the method, an ensemble of 100 models of the known egg model was used to test the PCA effect on the update. The simulation was run using Ensemble-based Reservoir Tool (ERT) with Open Porous Media (OPM) as a simulator.

The history match process was created using ES and ESMA with 3 update iterations to measure the effect of the inflation using PCA on the quality of the update and the divergence of the filter.

Using a model as the true case, measurements of flow was recorded every 5 years at different wells. With initial error on measurement of 10%, using PCA the measurement error was inflated. The update process using ES, ESMDA was performed for PCA, Standard Assimilation process using measurements. Furthermore, the assimilation process was run by using only the last measurements to examine the effect of PCA inflation compared to the conditioning workflow suggested by Eversen & Eikrem.

Model Setup

The egg model is set up with four production wells and eight injection wells. The model is 2 phase reservoir containing water and oil. The injection phase is water, with 10 years of production.

The model is reporting Water and Oil production rates at each production well. The production wells are controlled by BHP while injection wells are controlled by rate.

MODEL	Egg Model
START TIME	15 JUN 2011
PRODUCTION WELLS	4
INJECTION WELLS	8
ENSEMBLE SIZE	100
UPDATED PARAMETERS	PERM
PHASES	Oil, Water
SUMMARY	WOPR, WWPR for all Wells
INITIAL ERROR ON MEASUREMENTS	10%
USED MEASUREMENTS	WOPR, WWPR of all Wells
END TIME	23 APRIL 2021
REPORTING TIME STEP	30 days

Table 1 Reservoir Data

	DEFAULT RUN	**PCA**	**EVERSEN COND.**
MODEL	Egg Model	Egg Model	Egg Model
UPDATE TIME	Every 2 years	Every 2 years	23 APRIL 2021
PCA CUTOFF	NA	0.95	NA

Table 2 Measurement Data

Results and Discussion

Measurements using in each case with error used on it.

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	28.6329 sm3/day	2.86329 sm3/day	9.054518 sm3/day
24 JUN 2015	7.10808 sm3/day	0.71081 sm3/day	2.247769 sm3/day
13 JUN 2017	4.22581 sm3/day	0.42258 sm3/day	1.336318 sm3/day
3 JUN 2019	2.95862 sm3/day	0.29586 sm3/day	0.935598 sm3/day
23 APRIL 2021	2.26009 sm3/day	0.22601 sm3/day	0.714703 sm3/day

Table 3 PROD 1 WOPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	91.5221 sm3/day	9.15221 sm3/day	28.941829 sm3/day
24 JUN 2015	35.9481 sm3/day	3.59481 sm3/day	11.367787 sm3/day
13 JUN 2017	18.7288 sm3/day	1.87288 sm3/day	5.922567 sm3/day
3 JUN 2019	11.9122 sm3/day	1.19122 sm3/day	3.766968 sm3/day
23 APRIL 2021	8.43347 sm3/day	0.84335 sm3/day	2.66690 sm3/day

Table 4 PROD 2 WOPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	38.9367 sm3/day	3.89367 sm3/day	12.31287 sm3/day
24 JUN 2015	9.99258 sm3/day	0.99926 sm3/day	3.15993 sm3/day
13 JUN 2017	5.52553 sm3/day	0.55255 sm3/day	1.74732 sm3/day
3 JUN 2019	3.58032 sm3/day	0.35803 sm3/day	1.13220 sm3/day
23 APRIL 2021	2.63834 sm3/day	0.26383 sm3/day	0.83432 sm3/day

Table 5 PROD 3 WOPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	59.1861 sm3/day	5.91861 sm3/day	18.71629 sm3/day
24 JUN 2015	10.8288 sm3/day	1.08288 sm3/day	3.42437 sm3/day
13 JUN 2017	5.26320 sm3/day	0.52632 sm3/day	1.66437 sm3/day
3 JUN 2019	3.46164 sm3/day	0.34616 sm3/day	1.09467 sm3/day
23 APRIL 2021	2.62917 sm3/day	0.26292 sm3/day	0.83142 sm3/day

Table 6 PROD 4 WOPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	93.9928 sm3/day	9.39928 sm3/day	29.72313 sm3/day
24 JUN 2015	134.151 sm3/day	13.41510 sm3/day	42.42227 sm3/day
13 JUN 2017	134.168 sm3/day	13.51680 sm3/day	42.42765 sm3/day
3 JUN 2019	132.958 sm3/day	13.29580 sm3/day	42.04501 sm3/day
23 APRIL 2021	132.316 sm3/day	13.23160 sm3/day	41.84199 sm3/day

Table 7 PROD 1 WWPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	259.305 sm3/day	25.93050 sm3/day	81.99944 sm3/day
24 JUN 2015	283.112 sm3/day	28.31220 sm3/day	89.52788 sm3/day
13 JUN 2017	296.824 sm3/day	29.68240 sm3/day	93.86299 sm3/day
3 JUN 2019	306.24 sm3/day	30.52400 sm3/day	96.84159 sm3/day
23 APRIL 2021	311.756 sm3/day	31.17560 sm3/day	98.58590 sm3/day

Table 8 PROD 2 WWPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	30.6714 sm3/day	3.06714 sm3/day	9.69914 sm3/day
24 JUN 2015	64.8392 sm3/day	6.48392 sm3/day	20.50079 sm3/day
13 JUN 2017	70.7263 sm3/day	7.07263 sm3/day	22.36562 sm3/day
3 JUN 2019	73.4476 sm3/day	7.34476 sm3/day	23.22617 sm3/day
23 APRIL 2021	74.6726 sm3/day	7.46726 sm3/day	23.61355 sm3/day

Table 9 PROD 3 WWPR

TIME	MEASUREMENT	ERROR INITIAL	ERROR PCA
4 JUN 2013	38.7499 sm3/day	3.87499 sm3/day	12.25379 sm3/day
24 JUN 2015	90.0338 sm3/day	9.00338 sm3/day	28.47119 sm3/day
13 JUN 2017	100.537 sm3/day	10.05370 sm3/day	31.79259 sm3/day
3 JUN 2019	101.729 sm3/day	10.17290 sm3/day	32.16953 sm3/day
23 APRIL 2021	101.311 sm3/day	10.13110 sm3/day	32.03735 sm3/day

Table 10 PROD 4 WWPR

Depending on the ensemble size, measurements number and the errors the inflation factor was calculated to be 3.16228.

As shown in the tables the error (standard deviation) of the measurements are inflated, therefore the measurement error covariance matrix is produced to cover a larger space of possibility.

Effect of PCA

In this section, results from using PCA inflated error covariance matrix and without using the PCA with initial error of 10% will be presented. The difference in the ensemble spread and mean will be discussed to show the effect of PCA on ensemble spread and quality.

Using ES

In the graphs below, the generated results after inflating the ensemble using PCA appear to cover higher space. The effect is representing the decoupling of the measurements by compensating for the dependency and the redundancy by inflating the measurement error covariance matrix.

However, the effect of the PCA on the ensemble spread using standard ES with one iteration isn't high since the original spread using uninflated measurements remains high. The use of one iteration results in a lower dependency on the measurements since it is used only once.

Using ES doesn’t president the best quality in ensemble-based history matching as discussed before. Nevertheless, it was concluded that using ESMDA with multiple iterations is required to achieve good quality match.

Therefore, the experiment was executed using ESMDA with 3 iterations.

Figure 1 WOPR PROD1 -- ES

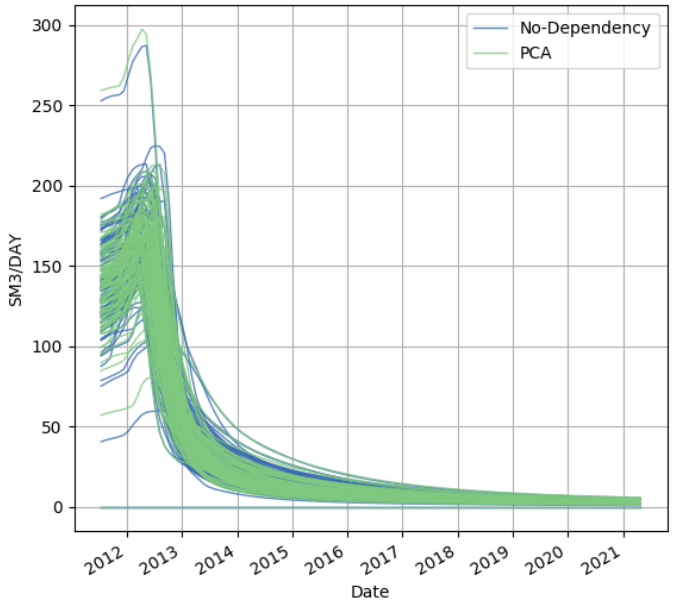


Figure 2 WOPR PROD2 -- ES

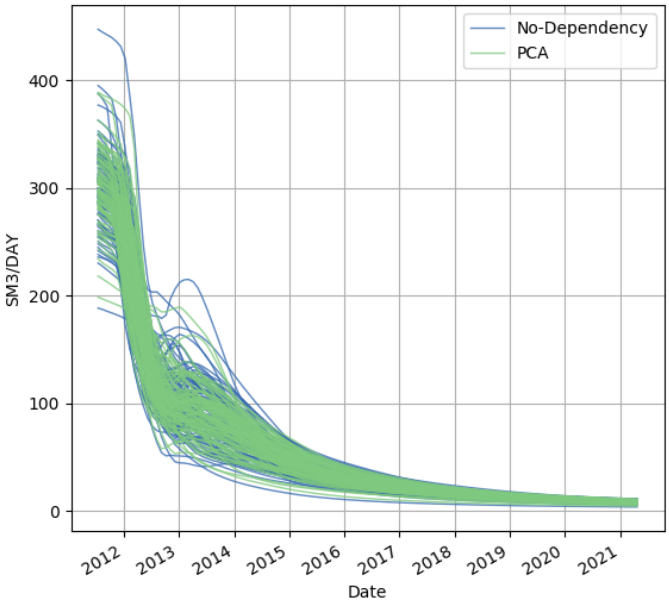


Figure 3 WOPR PROD3 -- ES

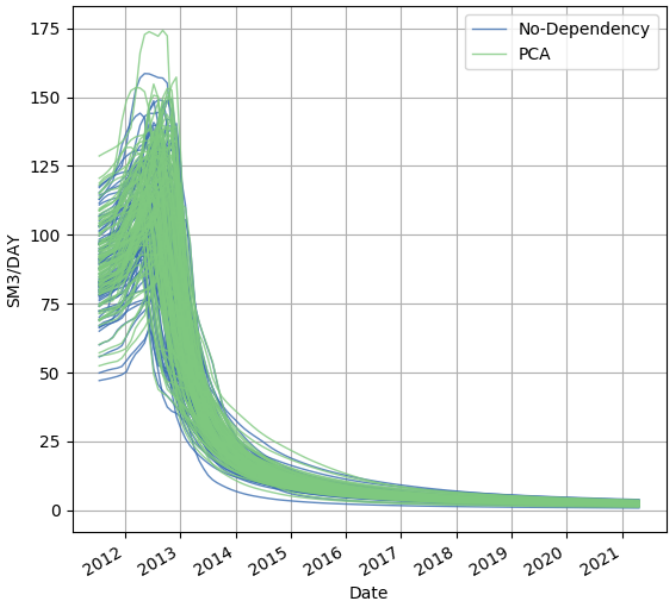


Figure 4 WOPR PROD4 -- ES

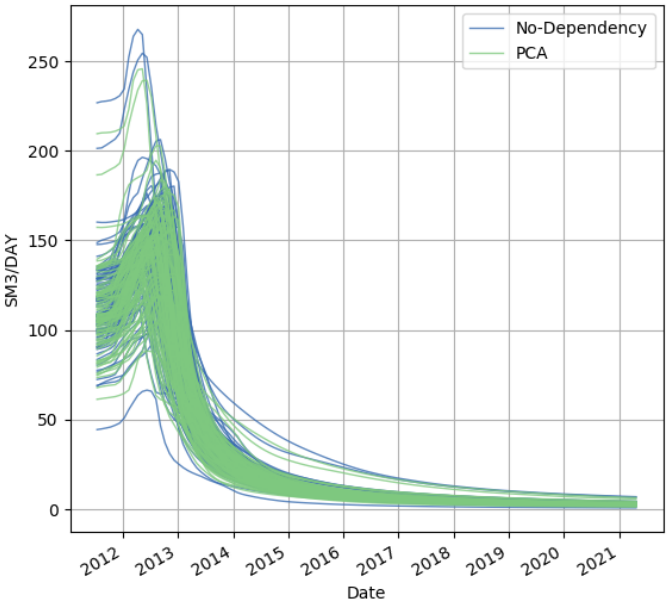


Figure 5 WWPR PROD1 -- ES

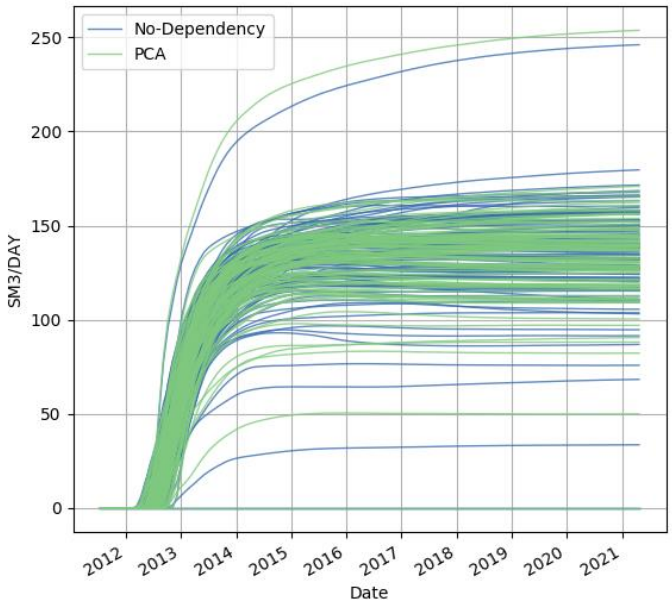


Figure 6 WWPR PROD2 -- ES

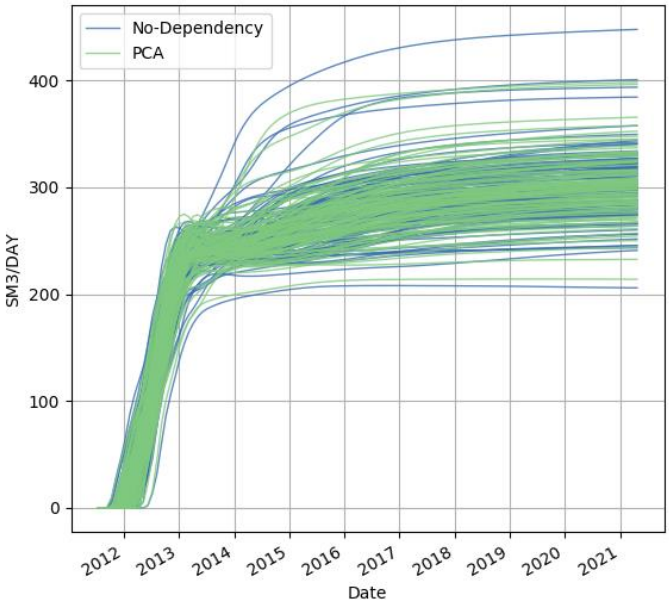


Figure 7 WWPR PROD3 -- ES

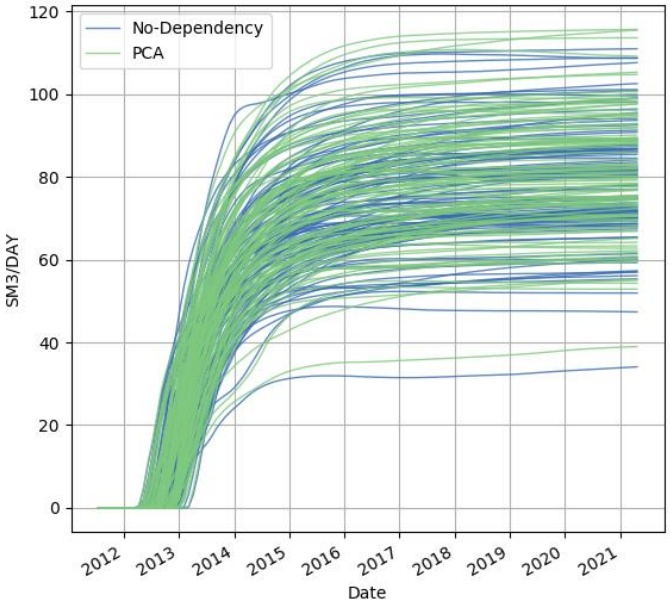
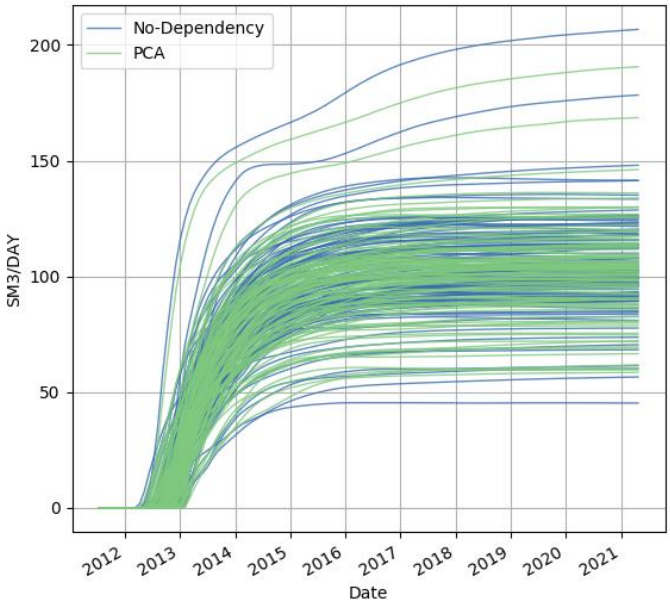


Figure 8 WWPR PROD4 -- ES



Using ESMDA

Using the ESMDA with 3 iterations requires the use of the measurements once at each iteration to assimilate the models. In the figures blow, it is shown the clear difference in the spread after using the PCA inflated results.

However, the mean of the ensemble is barely changed as show in figures (15) to (22). Thus, the quality of the match isn't affected and maintained.

Figure 9 WOPR PROD1 -- ESMDA

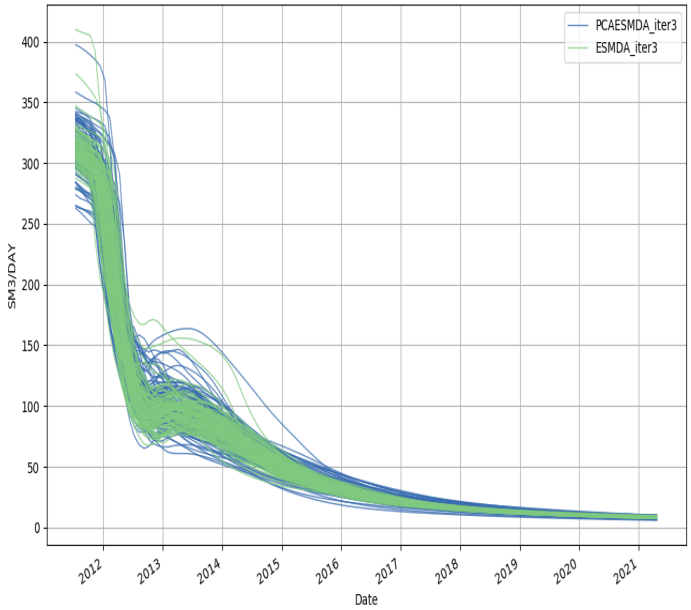


Figure 10 WOPR PROD2 -- ESMDA

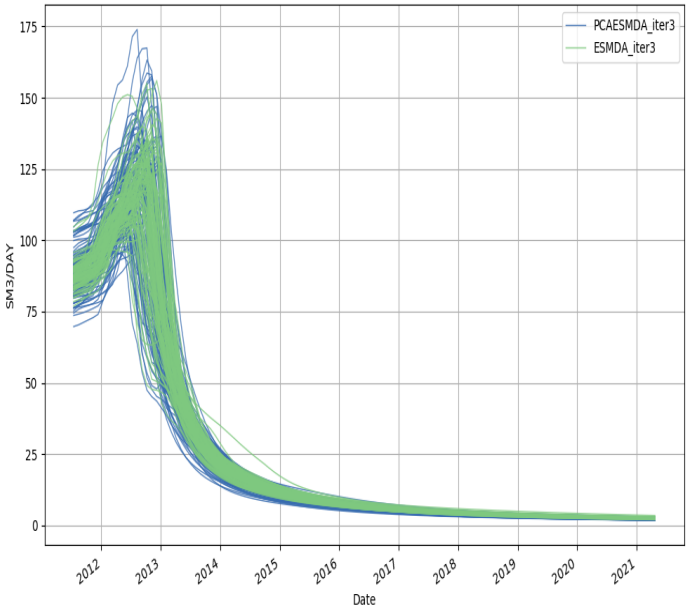


Figure 11 WOPR PROD3 -- ESMDA

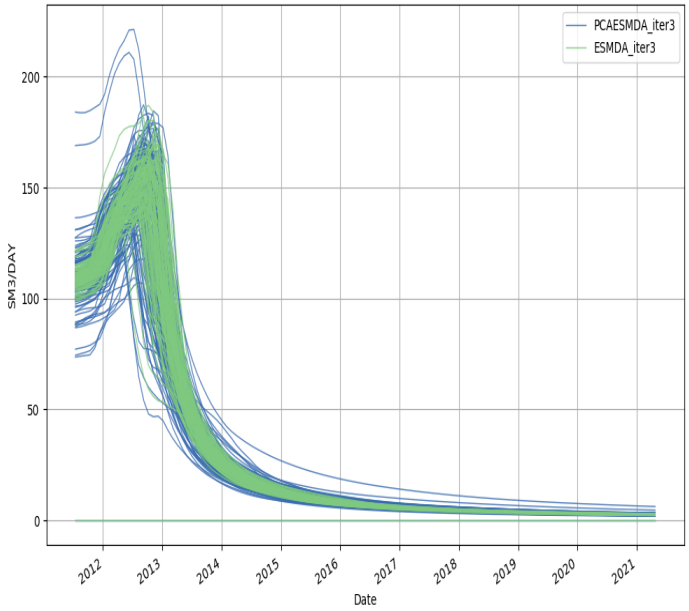


Figure 12 WWPR PROD2 -- ESMDA

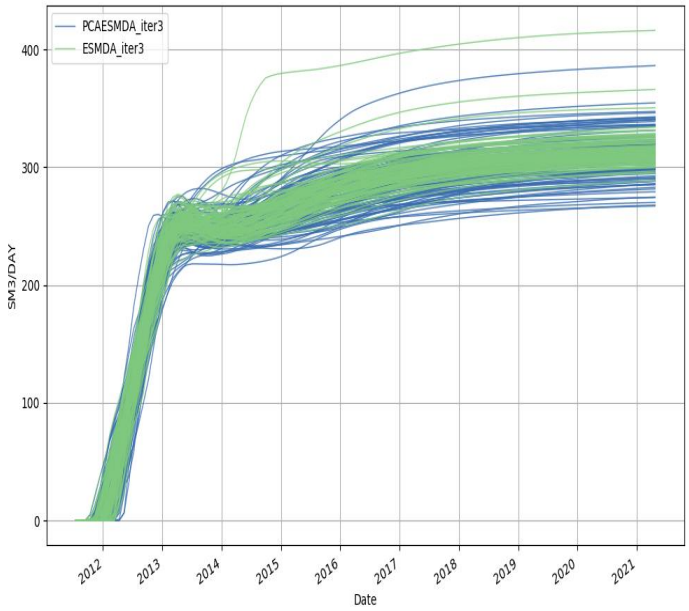


Figure 13 WWPR PROD3 -- ESMDA

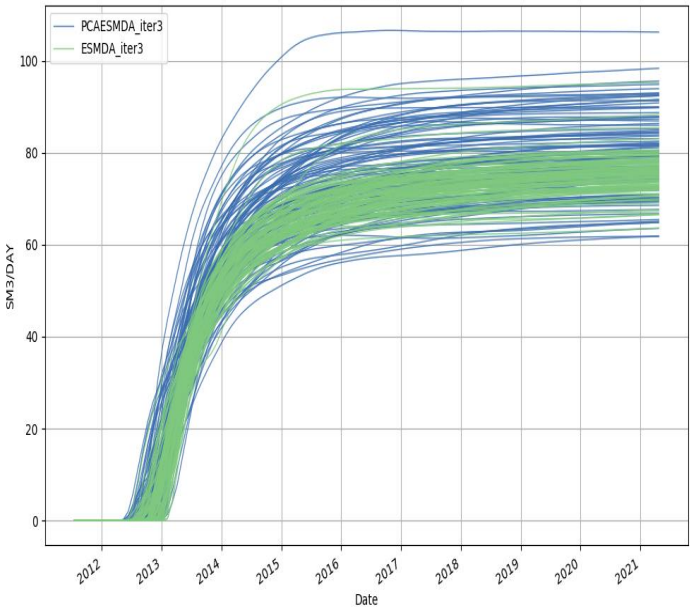
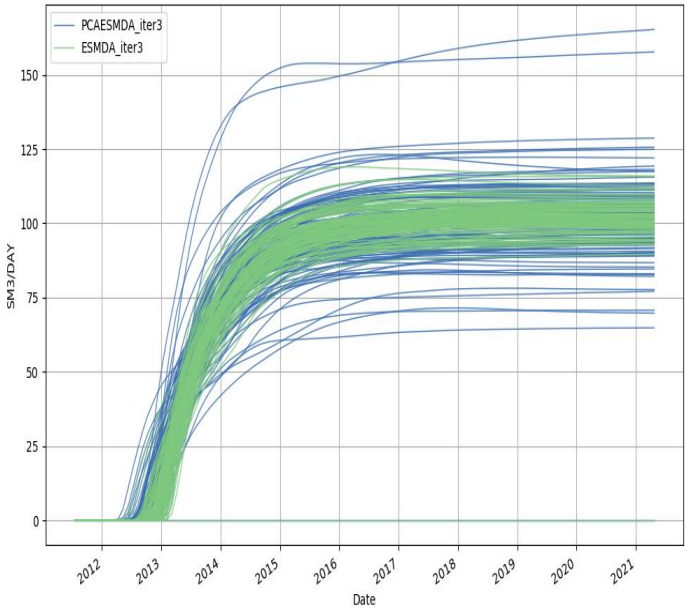


Figure 14 WWPR PROD4 -- ESMDA



The inflation of the measurements' perturbation by the PCA coefficient is produced dynamically based on the number of principle components required to represent the matrix. Thus, the process is done dynamically based on the measurements obtained.

The ensemble is shown to cover higher span from the figures (9) to (14), since the measurements are perturbated with higher covariance. However, the mean of the measurements is unchanged. This explained the little to no change in the mean of the ensemble results. Therefore, the quality of the ensemble is maintained while it spans over higher probability span, delaying the filter divergence.

Compare PCA with Conditioning

In this section, three sets of results will be displayed:

- Updated ensemble using inflated measurement error (PCA)
- Updated ensemble using last measurement time (no-Dependency)
- Updated ensemble using initial measurement error (no-conditioning)

In the plots P10 -P90 and means are compared rather than the full ensemble to represent the spread of ensemble due to the visual difficulty to distinguish the 3 different sets.

Using ES

From the graph we can conclude:

- The mean of the three different set are barely different indicating the similar quality of the different methods.
- The PCA appears to be slightly superior in some measurements in figures (9-14). However, in figures (15-22) using last measurement time yields a higher spread.
- The difference in the results appear slight due to the nature of standard ES where only 1 iteration is used

Figure 15 WOPR PROD1 – ES STAT

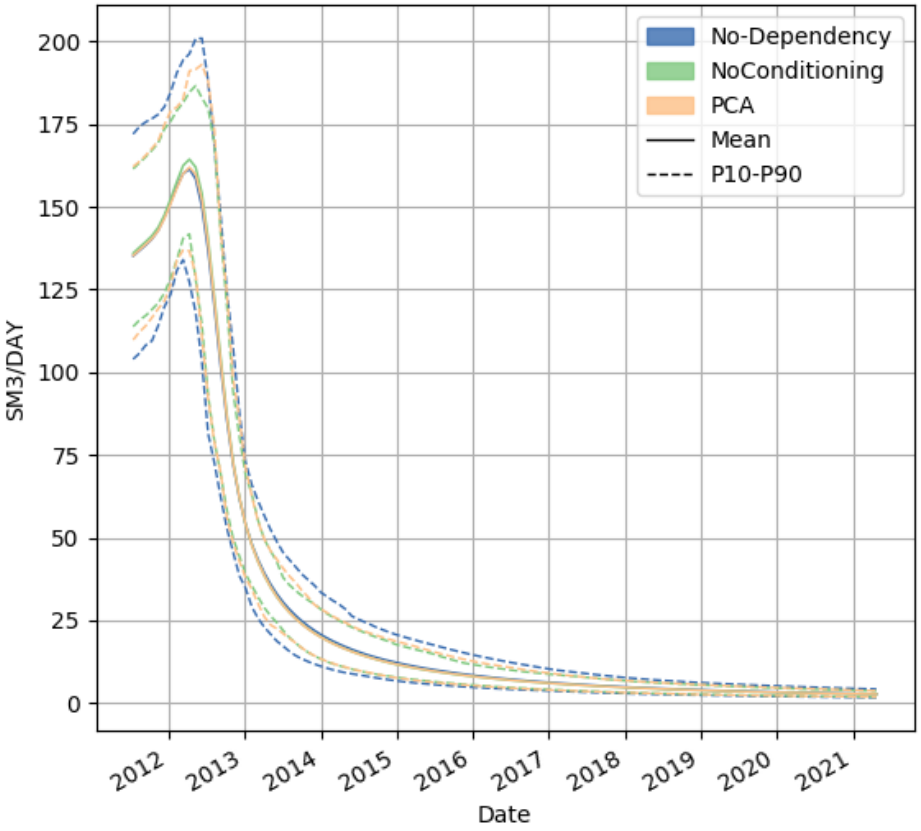


Figure 16 WOPR PROD2 – ES STAT

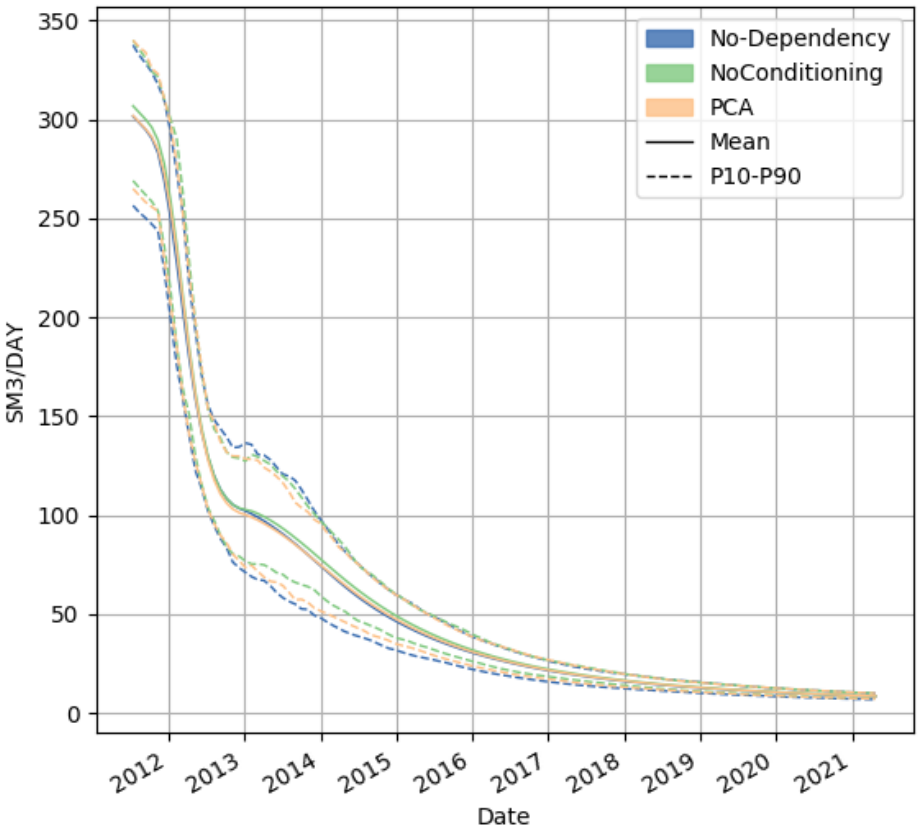


Figure 17 WOPR PROD3 – ES STAT

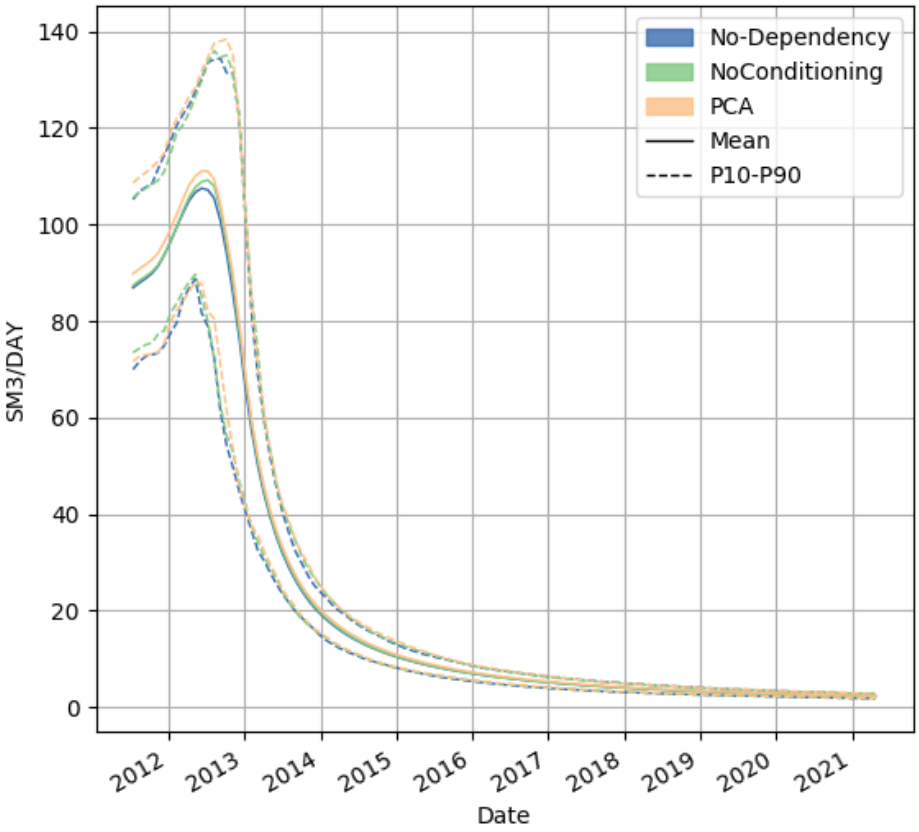


Figure 18 WOPR PROD4 – ES STAT

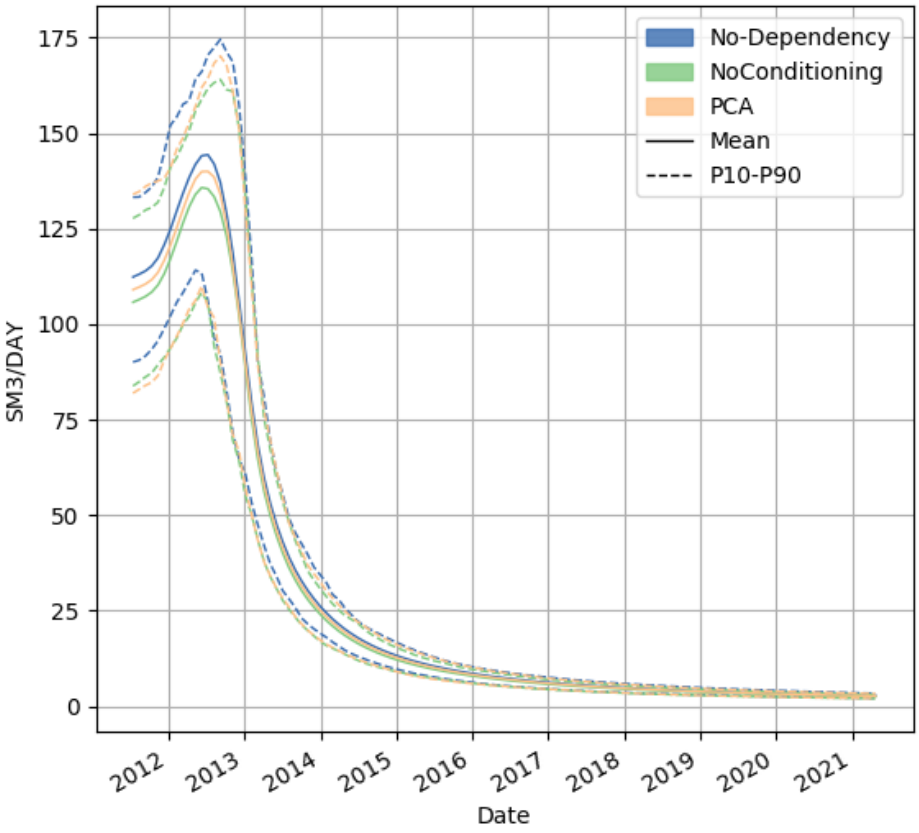


Figure 19 WWPR PROD1 – ES STAT

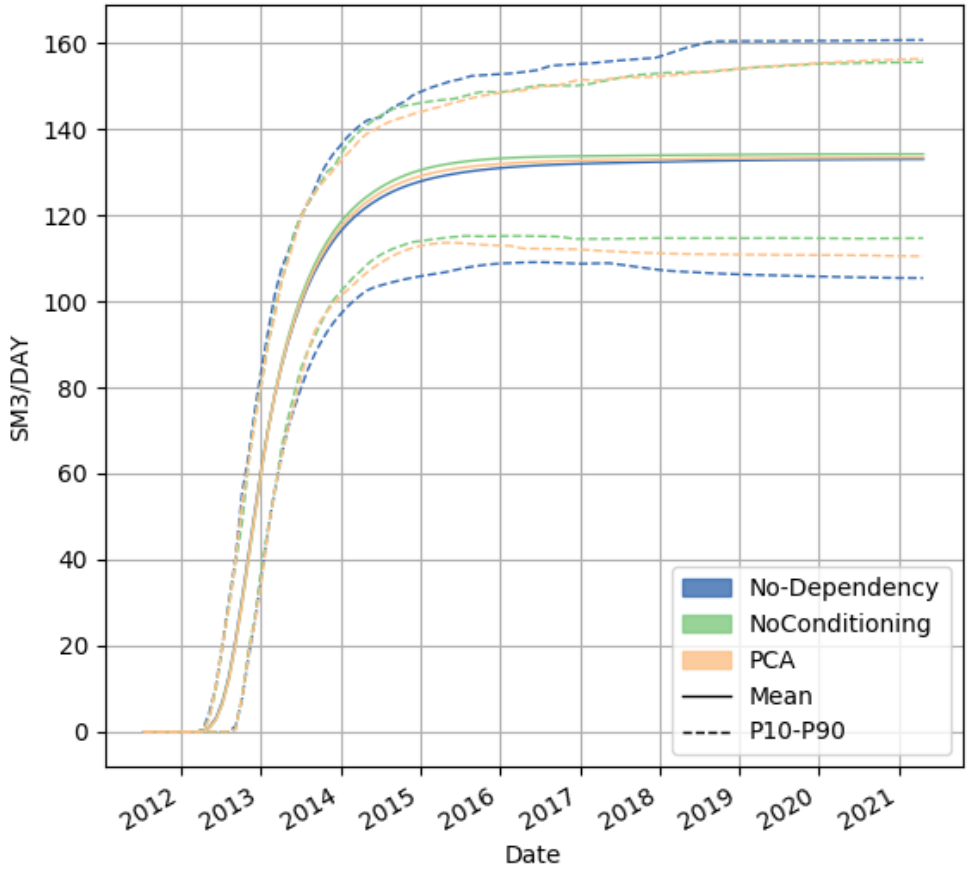


Figure 20 WWPR PROD2 – ES STAT

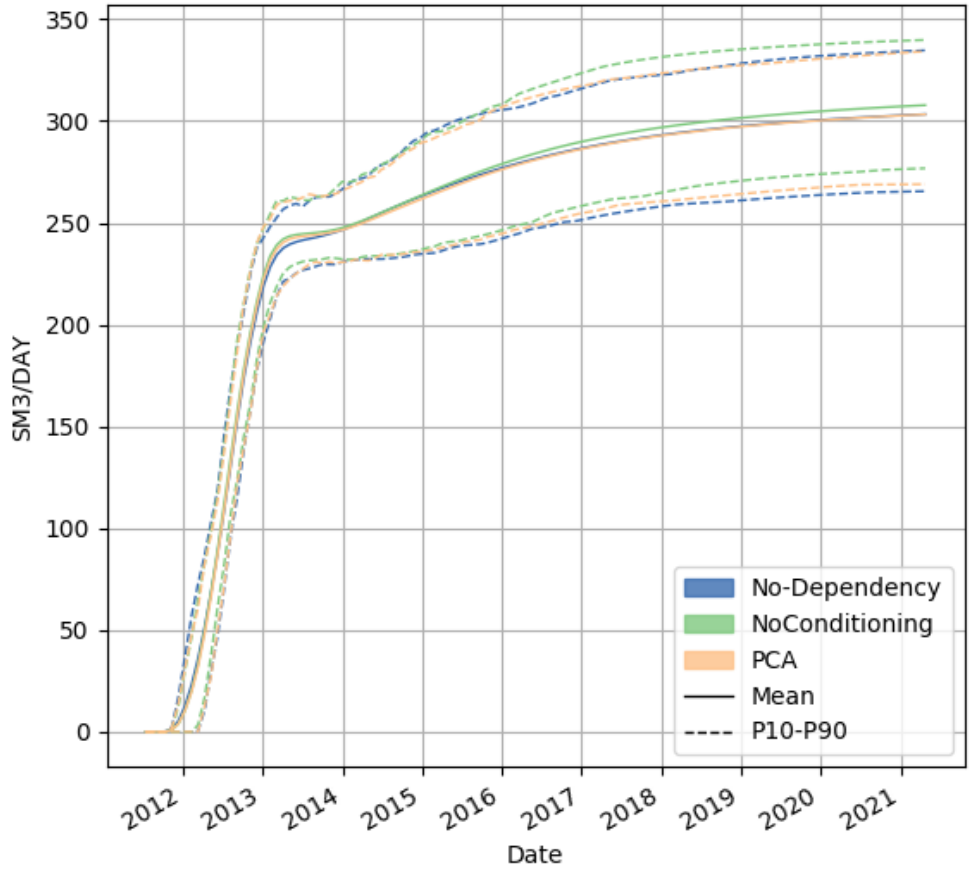


Figure 21 WWPR PROD3 – ES STAT

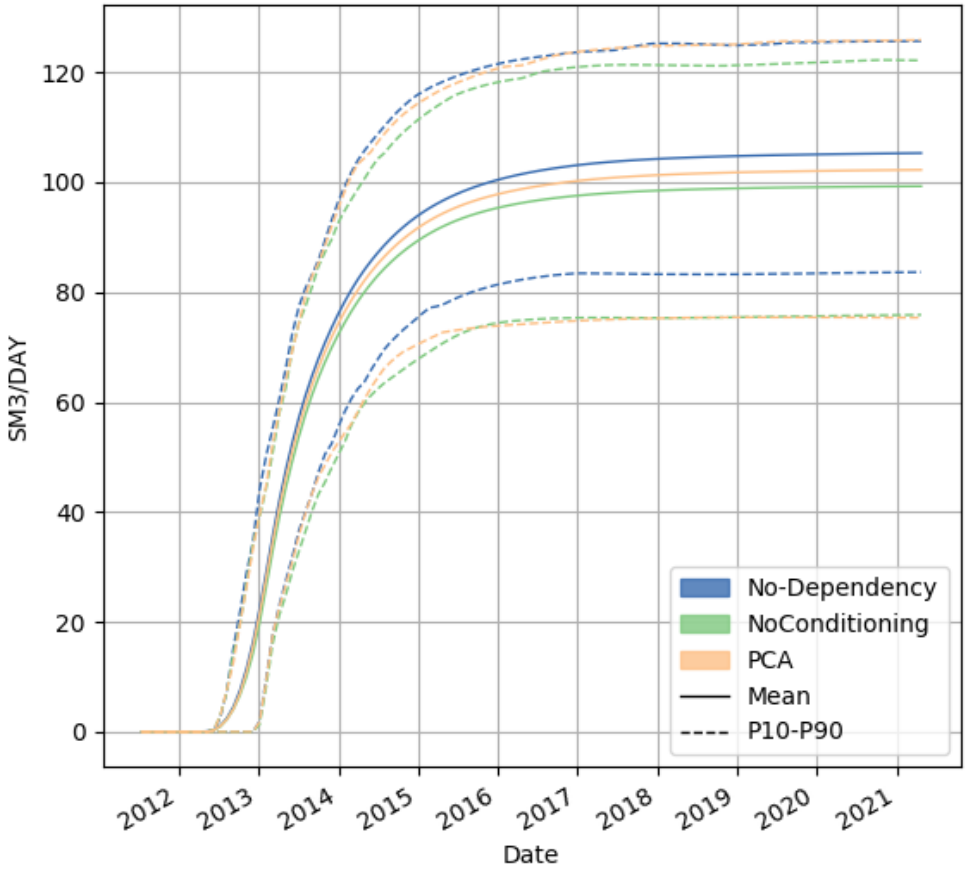


Figure 22 WWPR PROD4 – ES STAT

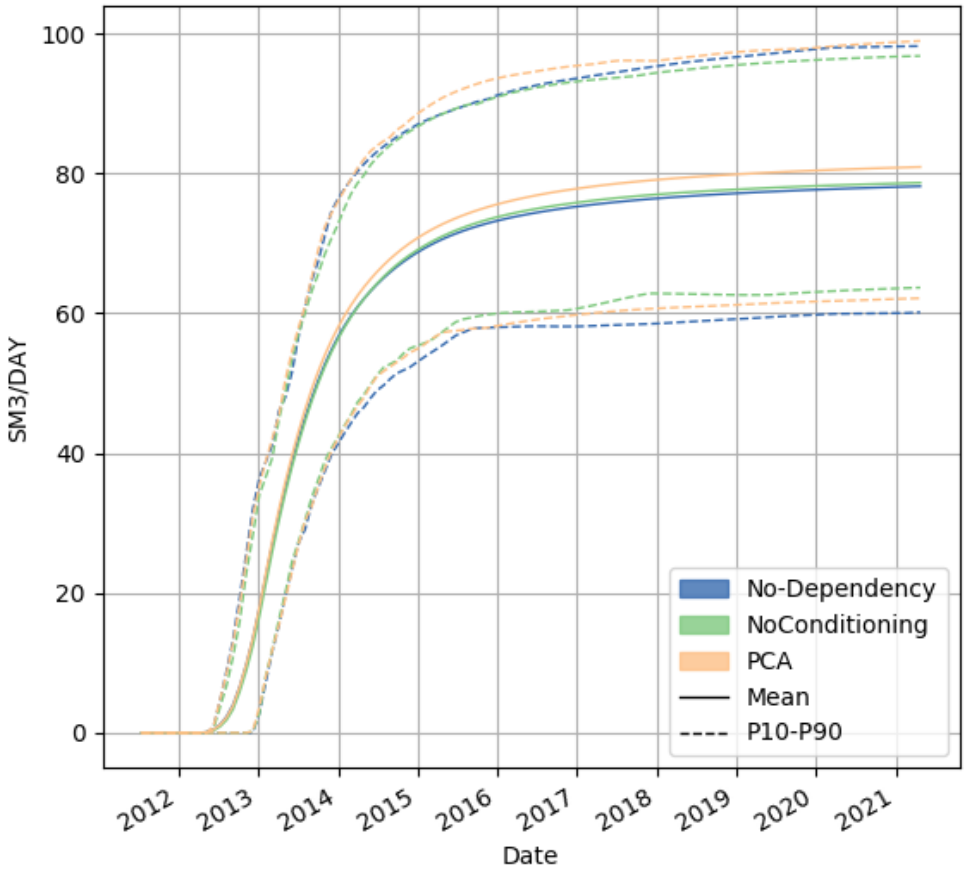


Figure 23 WOPR PROD1 – ESMDA STAT

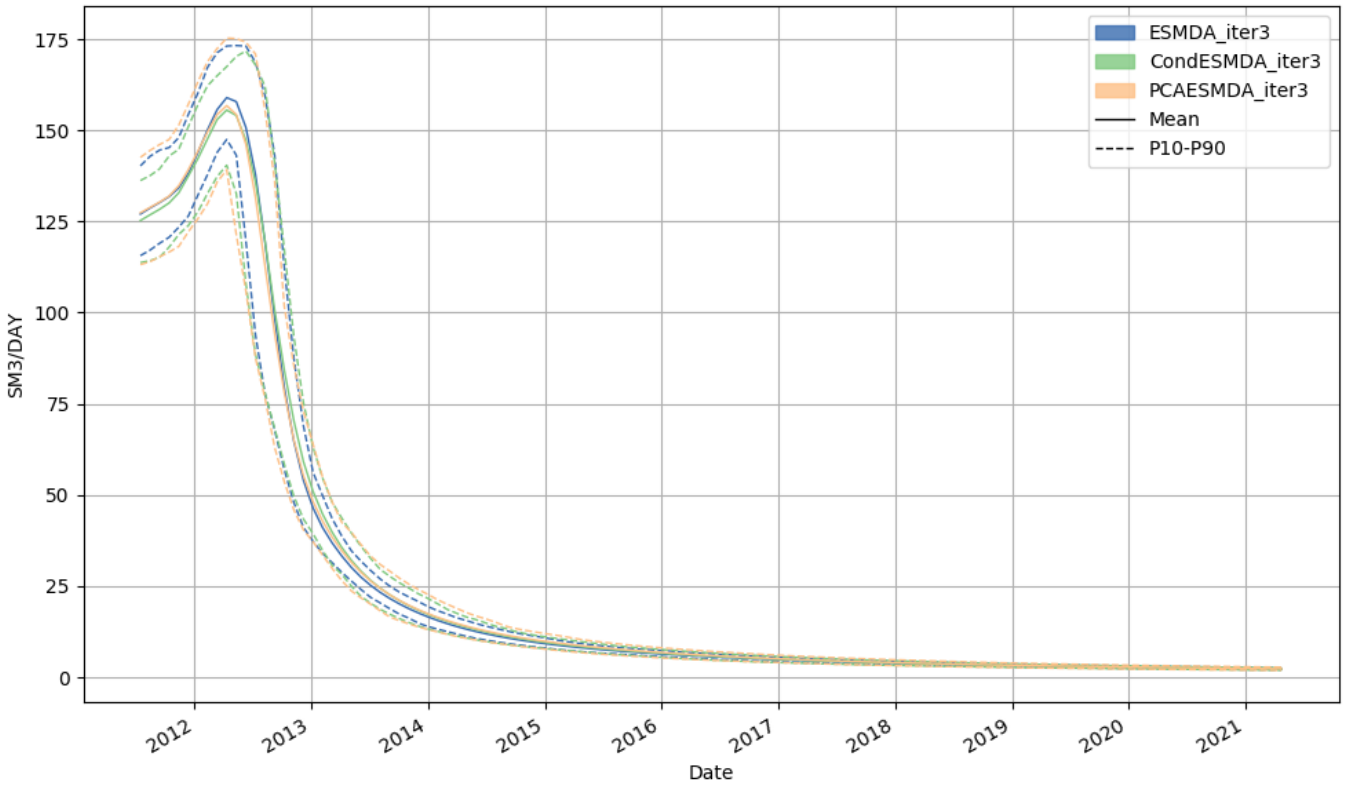


Figure 24 WOPR PROD2 – ESMDA STAT

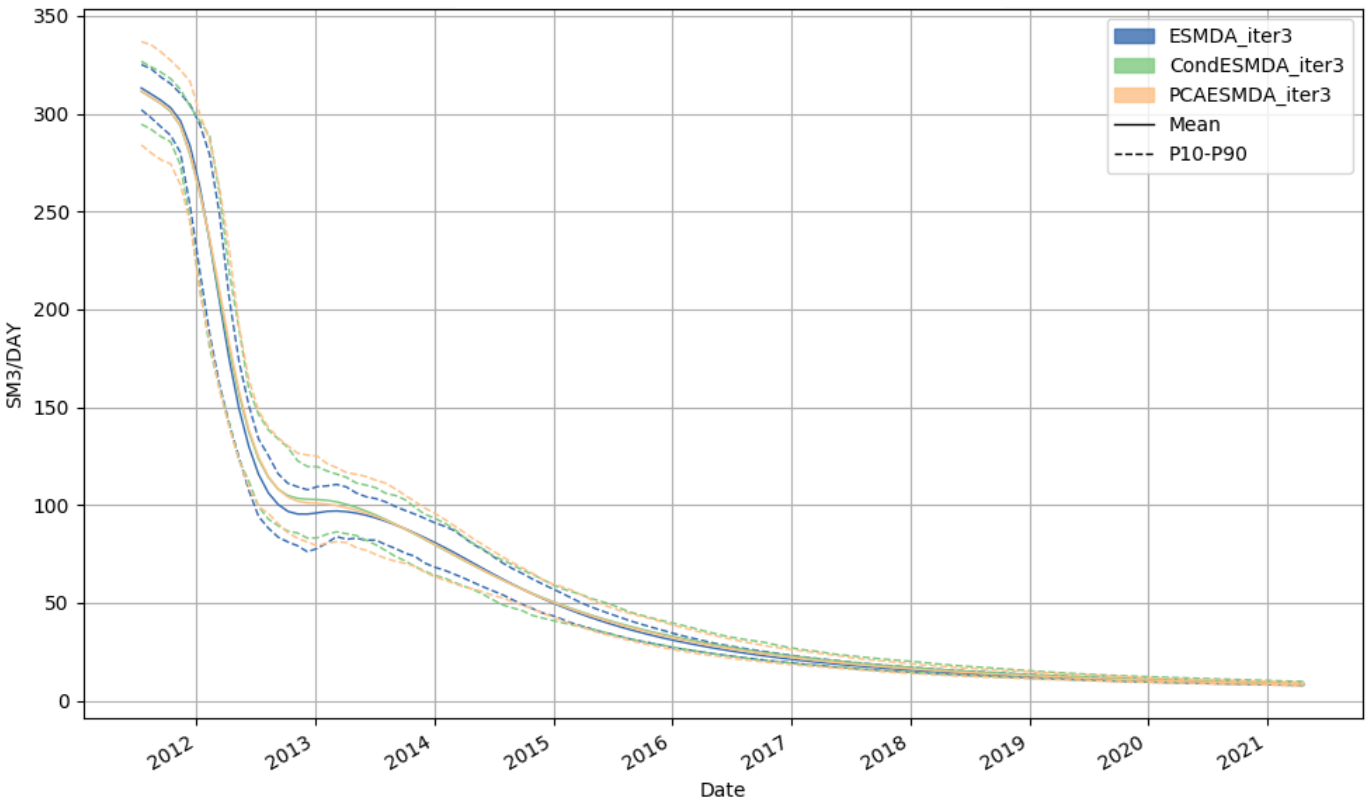


Figure 25 WOPR PROD3 – ESMDA STAT

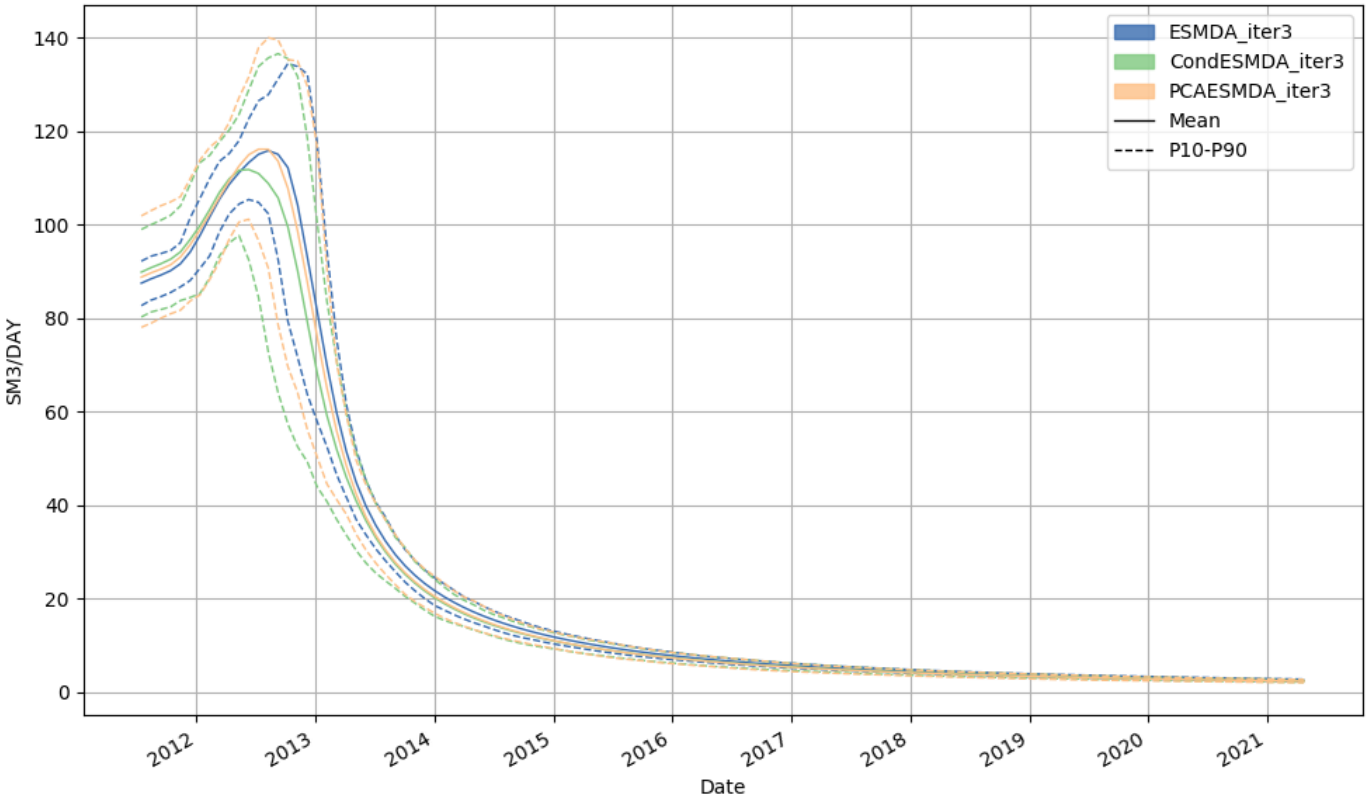


Figure 26 WOPR PROD4 – ESMDA STAT

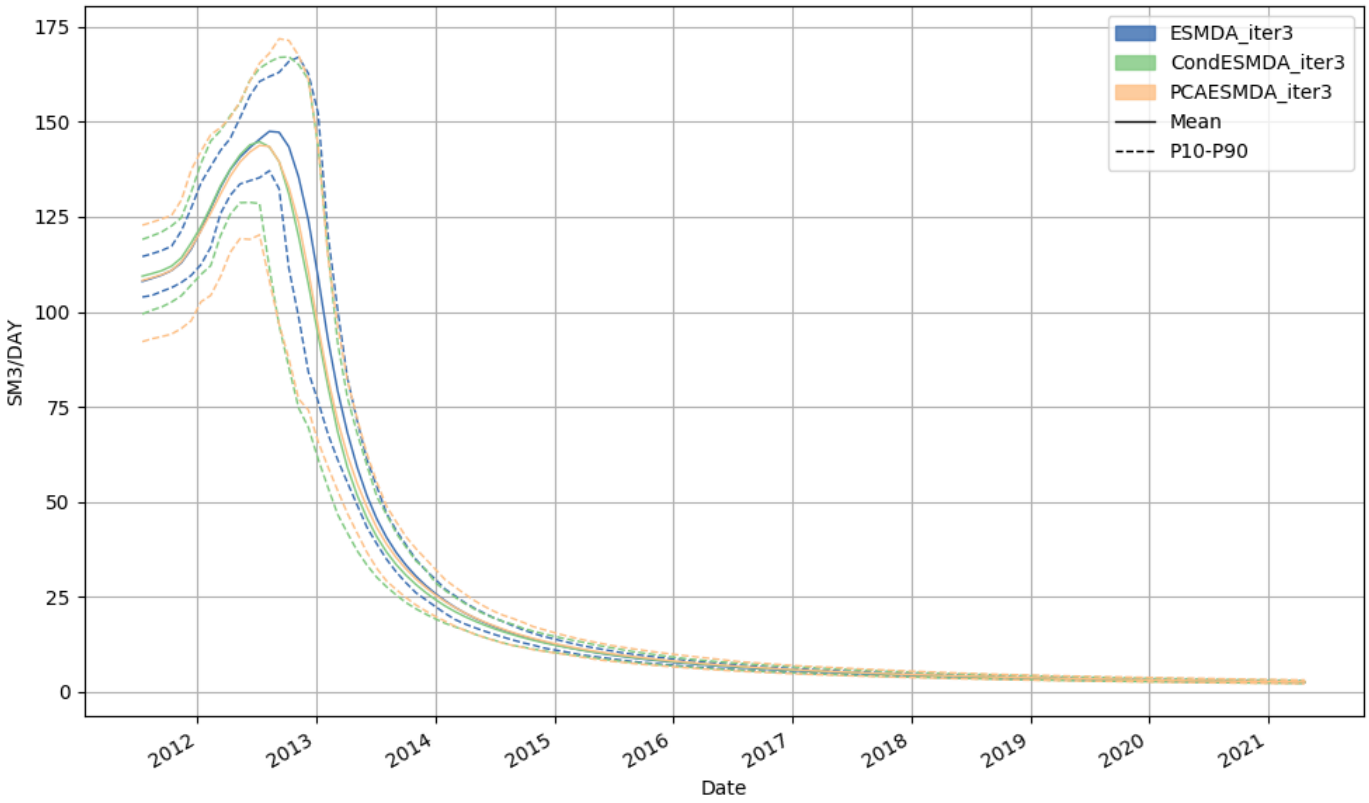


Figure 27 WWPR PROD1 – ESMDA STAT

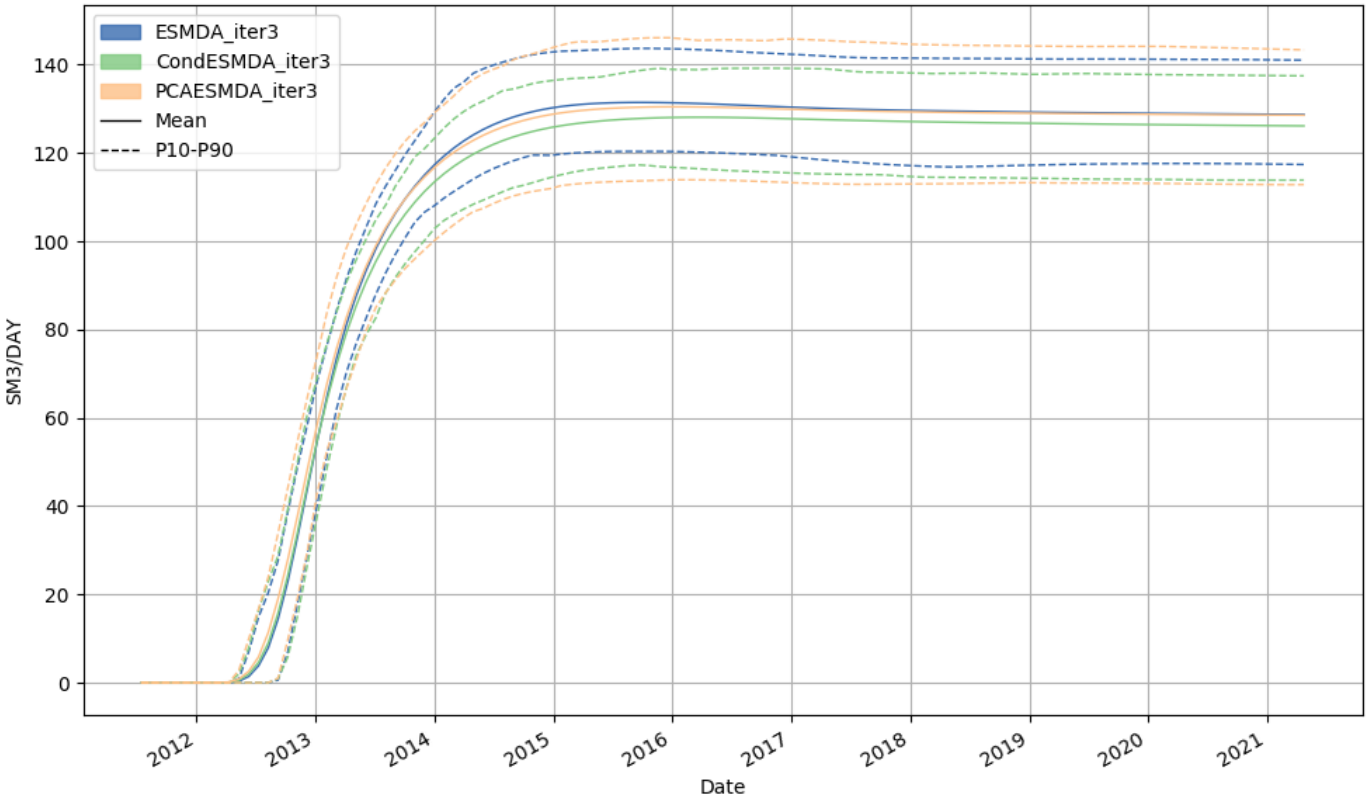


Figure 28 WWPR PROD2 – ESMDA STAT

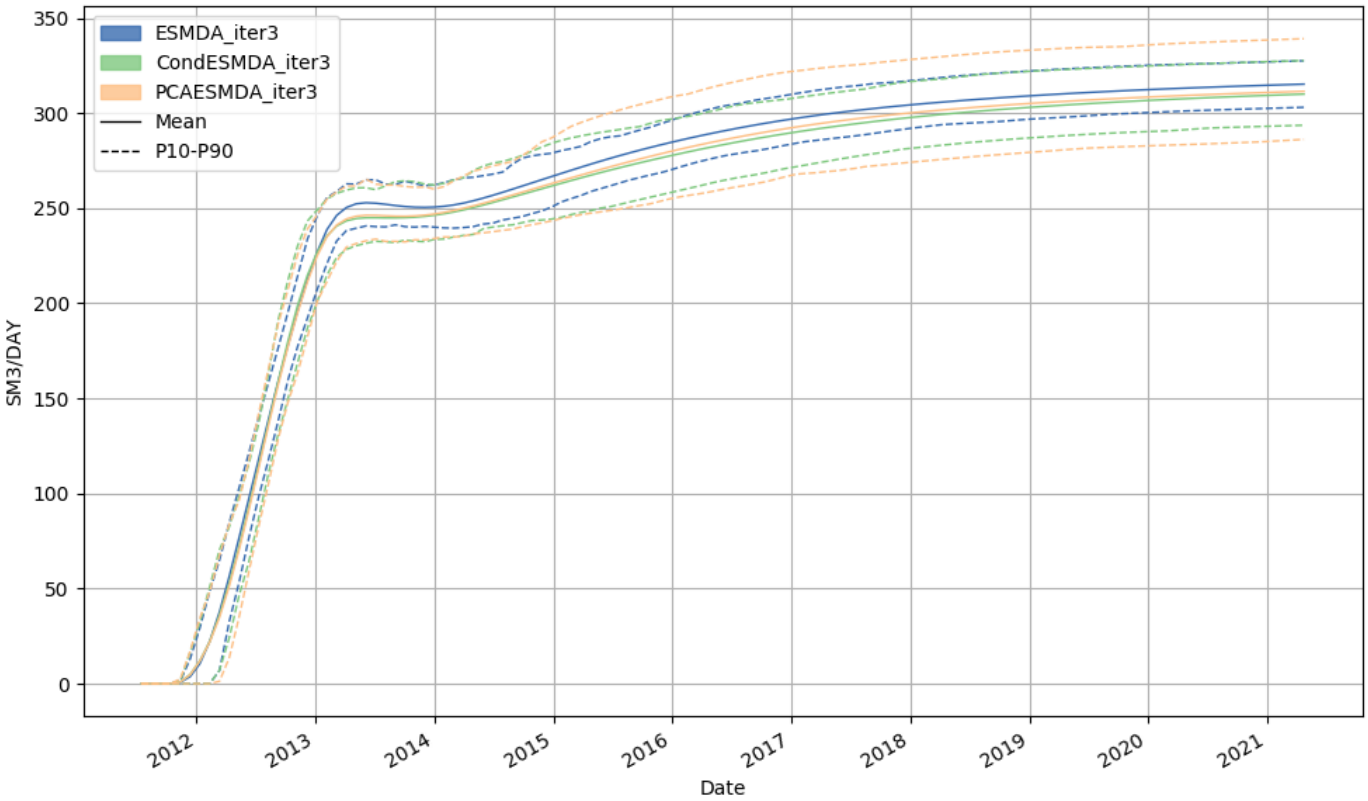


Figure 29 WWPR PROD3 – ESMDA STAT

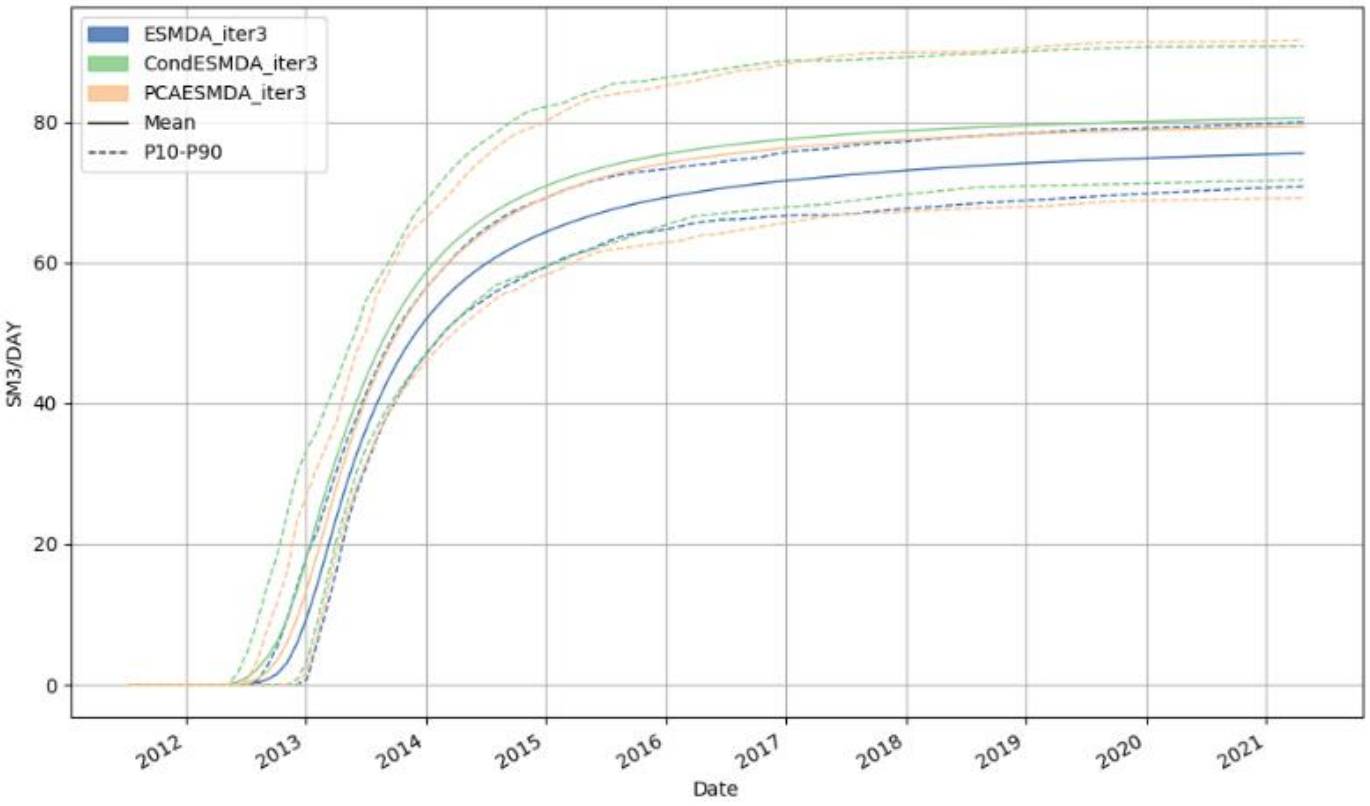
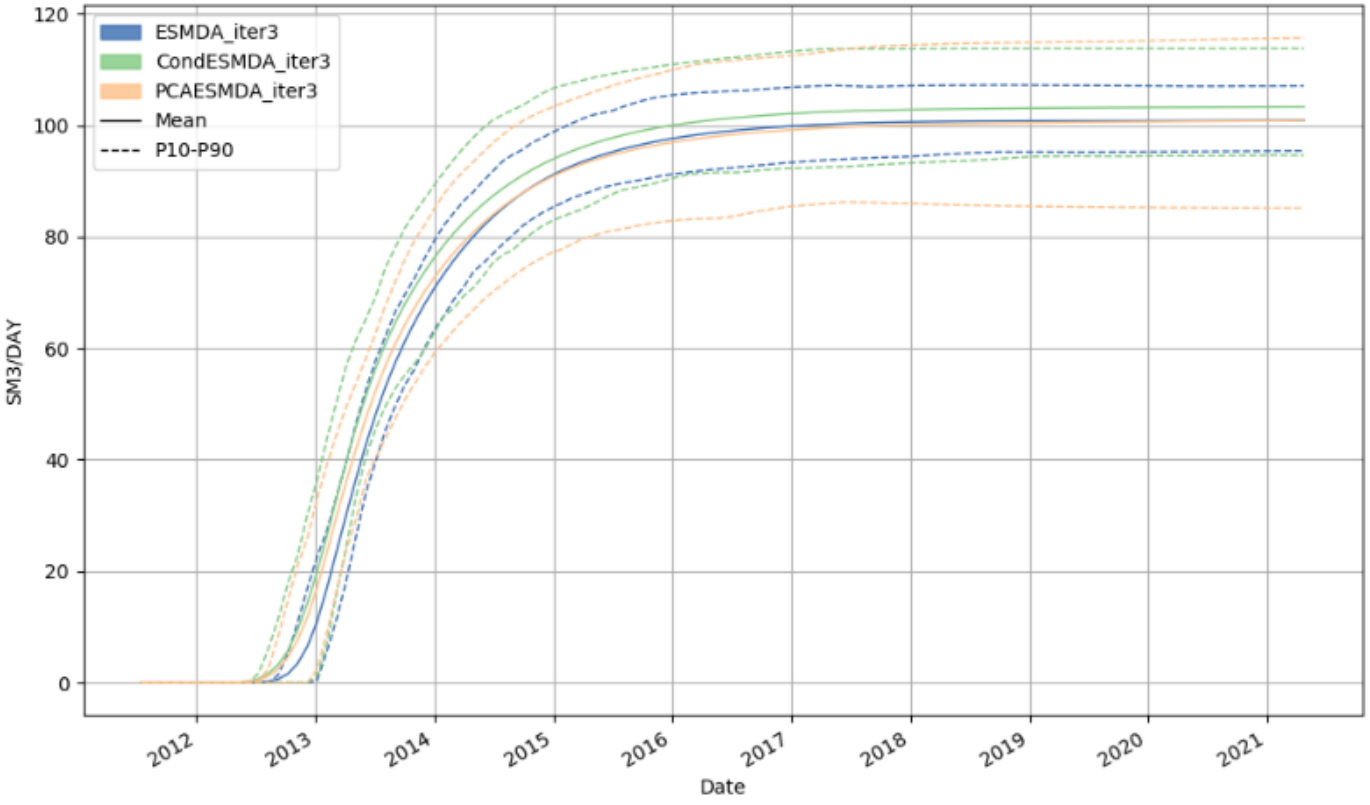


Figure 30 WWPR PROD4 – ESMDA STAT



The measurements matrix $\mathrm{D} = (\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \mathrm{d}_4, \ldots \ldots, \mathrm{d}_\mathrm{N}) \in \mathrm{R}^{\mathrm{mxN}}$ (where m is the number of measurements and N is the number of models) takes the shape

$$\begin{bmatrix} d_{1,1} & \cdots & d_{n,1} \\ \vdots & \ddots & \vdots \\ d_{1,n} & \cdots & d_{n,n} \end{bmatrix}$$

In this matric, $d_{1,1}$ to $d_{n,1}$represent the perturbated measurements of a given well reading at a given time. In the egg model case with PCA, it represents PROD 1 WOPR on 4th June 2013. At a given vector $\mathrm{d}_1$the values represent the different measurements at different times. While in the case of the egg model with Eversen's conditioning, it represents PROD 1 WOPR on 23rd April 2021. Therefore, the matrix dimension in the case of PCA is (40,100) where it is (8, 100) in the case of Eversen's conditioning.

While populating the 100 vectors in the matrix (D), the error in measurements is used as the variance. By increasing the variance value, the covariance of the measurements at the matrix inflates. Subsequently, the dependency between the measurements decreases.

Figure 31 Water Saturation Model 1 on 4rd June 2013 Iteration 3

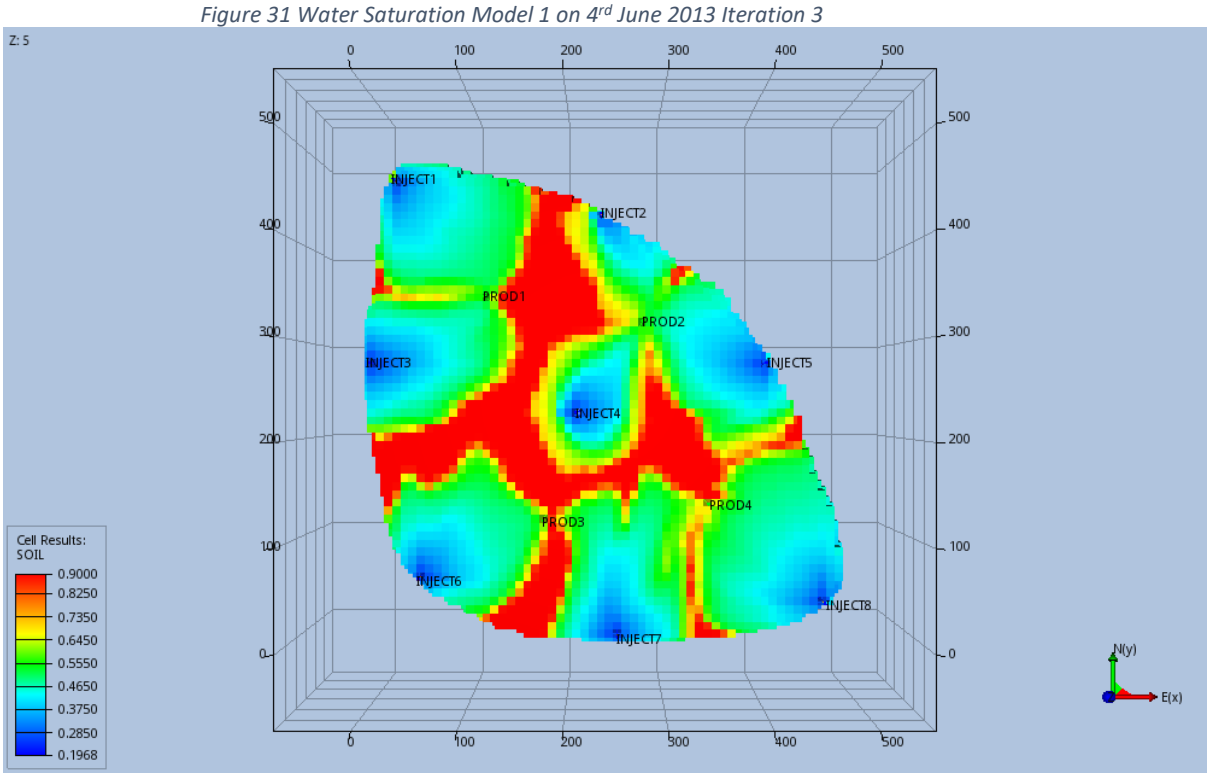


Figure 32 Water Saturation Model 15 on 4rd June 2013 Iteration 3

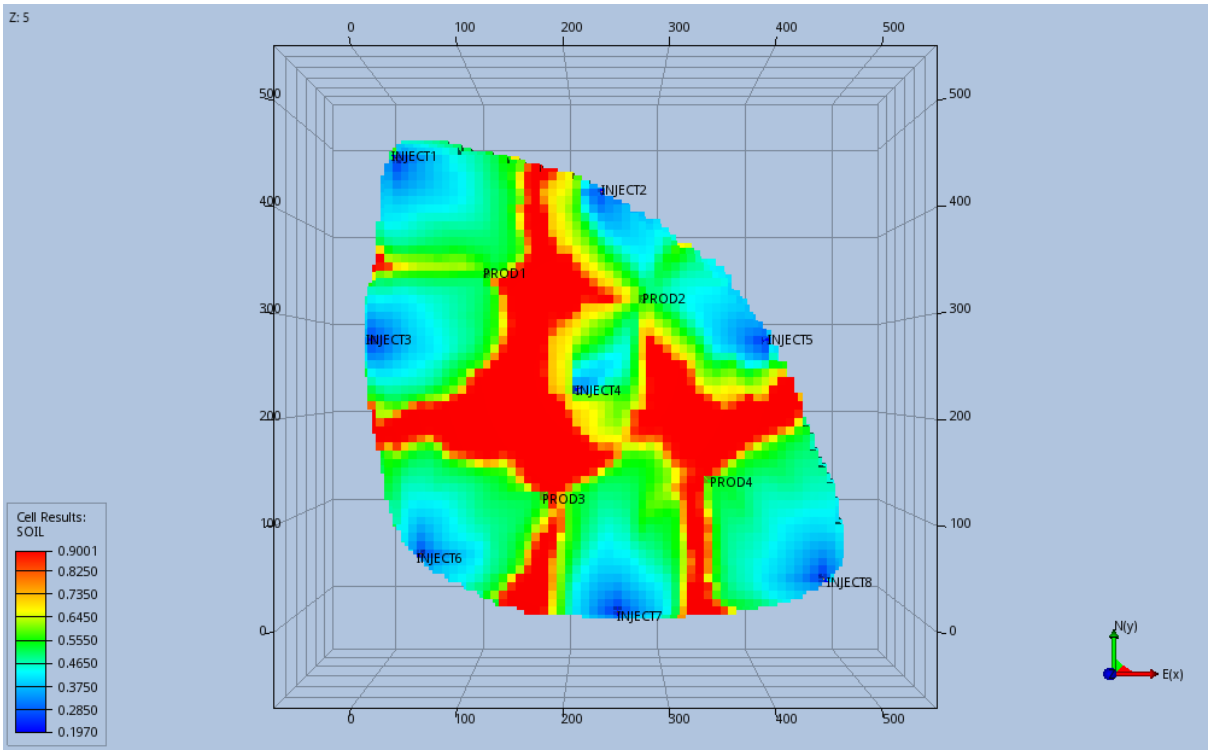
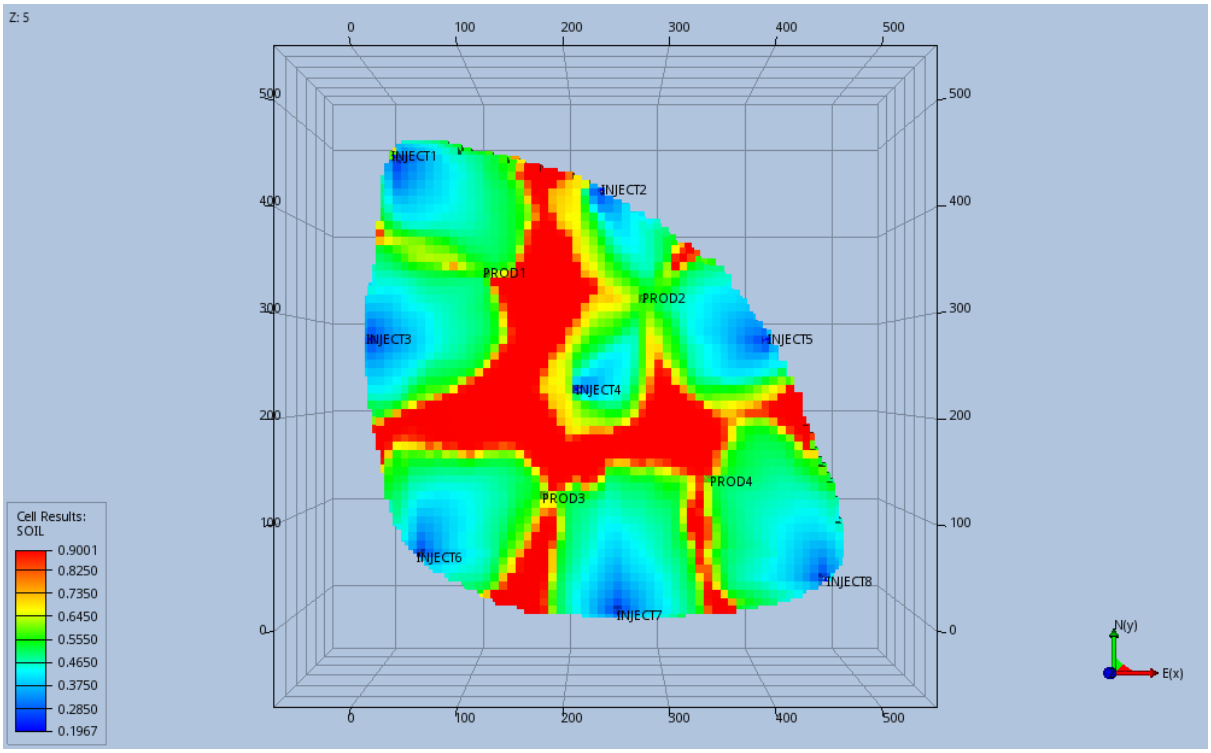


Figure 33 Water Saturation Model 35 on 4rd June 2013 Iteration 3



Using the Eversen's conditioning method the time-dependency between the measurements is removed. However, the dependency between the water and the oil production rate is not considered. For

example, the water production rate and the oil production rate at well PROD1 are governed by the bottom hole pressure (BHP) at the well. Therefore, there is a blatant dependency between the values. Similarly, the production rates at well PROD4 is more co-dependent with PROD2 and PROD 3 than production rates at well PROD1 since they share a larger portion drainage area and affected by shared Injectors as shown in figures (31), (32), (33). Therefore, the production rates are dependent on the same reservoir pressure values.

However, using the principle component analysis, the principle components are extracted from the matrix (D), the principle components representing the matrix are extracted and ordered using the eigen values of each eigen vector (principle component). Therefore, the components are ordered descending where the most representing component is first. Therefore, the number of principle components can be a representation of the level of dependency between the measurements. For example, if the measurements matrix is presented by one principle component that shows the complete dependency between the measurements. Since, the whole sum of measurements can be presented by one multidimensional vector.

Therefore, as the number of components increase, the dependency between the measurements decrease. The PCA method is completely mathematical. Hence, it incorporates all types of dependency existing in the measurements.

From figures (27),(28),(30) the P10, and P90 of the PCA results are shown to have higher spread that Eversen's Conditioning, whereas figures(23),(24),(25),(26),(29) show slight to no difference in the spread. However, the oil production rates are extremely low in the reservoir making it difficult to establish a clear difference in the production rates. Nevertheless, the dependency of the production rates of the wells can be magnified by introducing more production wells. The effect of the improvement of the PCA method can magnify as the dependency between the wells increase.

Therefore, we can conclude the Superiority for PCA due to inclusion of space dependency. Since the PCA compensates for the dependency between the measurements including the redundant information.

However, including the last measurement time doesn't account for the dependency present between measurements at different wells, nor the dependency between the measurements of different phases at the same well.

In the represented results the phases existing are oil and water. Water and Oil production rate are slightly dependent at the same well since they are governed by the well dynamics as well as the

pressures present in near well area. However, the dependencies between oil and gas are higher due to the presence of gas in the oil field at reservoir conditions. Hence, the difference in the spread between PCA and Conditioning by Eversen and Eikrem is expected to be higher in a three-phase reservoir.

Moreover, PCA is more suitable for multiple updates without saving initial ensemble distribution without starting from time 0. Eversen and Eikrem (2018) presented the idea of using the last measurement time to update the reservoir. However, in a practical history matching situation the update of the ensemble takes place continuously over the life of the reservoir. Therefore, to correctly use the method it will be required to save the initial reservoir prior to be used at each update step. Contrarily, PCA can be used regardless of the starting time of the reservoir in a continuous way. Since the error inflating can be calculated at any time by including the total number of used measurements taken over the reservoir life. This provides PCA with a clear storage and computational advantage.

Conclusion

Ensemble-based history matching presented the oil and gas industry with a smart and effective way to incorporate uncertainties associated with data obtaining process. The method to measure that is by the accuracy of the mean and the inclusion of larger probability space. Kalman Filter has been identified to be the most promising method for ensemble model assimilation, where multiple derivation of the filter has been introduced including EnKF, ES, and ESMDA, where ESMDA was concluded to be the most efficient while preserving a relatively low computational cost.

In this paper, the reasons behind ensemble collapse (Filter divergence) was discussed. The redundancy and dependency between the measurements obtained as well as the recursive use of data was associated as one of the large reasons for filter divergence.

Populating covariance matrices with variables to represent correlation is a method that is not achieved at the current time due to the lack of information on a scientific (physical) ground. However, many methods are discussed aiming to delaying or preventing the filter divergence. Localization is a promising method of delaying the filter divergence, where is restricts the assimilation process to a defined space of the reservoir. Similarly, conditioning the measurement data to counteract the effect of dependency between measurements.

Furthermore, the use of Principle Component Analysis (PCA) to inflate the error covariance matrix was discussed. The method uses Eigenvectors to represent the dependency and redundancy between the measurements and uses the values of dependency to inflate the error on measurements. The method was found to increase the spread of the ensemble delaying filter divergence. Furthermore, the method was compared to the conditioning method suggested by Eversen and Eikrem and found to represent higher spread using ESMDA with 3 iteration. The results were obtained by simulating the synthetic Egg model. However, the effect of using smaller ensemble size motivated by a larger model is recommended to be investigated further.

References

Bishop, G., & Welch , G. (2006, Juky 26). An Introduction to Kalman Filter. *UNC- Chapel Hill*.

Chen, Y., & Oliver, D. S. (2012). Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother. *Math GeoSci*, 1-26.

DOLBERG, D. M., HELGESEN, J., HANSSEN, T. H., MAGNUS, I., SAIGAL, G., & PEDERSEN, a. B. (2000, April). Porosity prediction from seismic inversion, Lavrans Field, Halten Terrace, Norway. *The Leading edge*, pp. 392-399.

Emerick, A. A., & Reynolds, A. C. (2011). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*.

Emerick, A., & Reynolds, A. (2010). Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput Geosci*.

Eversen, G. (2004). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, pp. 343-367.

Eversen, G., & Eikrem, K. S. (2018). Conditioning reservoir models on rate data using ensemble smoothers. *Computational Geosciences*, 1251- 1270.

Eversen, G., & Leeuwen, P. J. (1996, January). Assimilation of Geosat Altimeter Data for the Agulhas Current using the ensemble kalman dilter with a Quasigeostrophic Model. *American Meteorological Society*, pp. 85-124.

Eversen, G., & Leeuwen, P. J. (2000, June). An Ensemble Kalman Smoother for Nonlinear Dynamics. *Monthly Weather Review*, pp. 1852-1867.

Hamill, T. M., Hamill, T. M., Anderson, J. L., & Snyder, C. (2009). Comments on ‘‘Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems.’’. *J. Atmos. Sci.*, pp. 3498-3500 .

Hamill, T. M., Snyder, a. C., & Hamill, T. M. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman Filter. *Monthly Weather Review*, pp. 2776 - 2790.

HOUTEKAMER, P. L., & MITCHELL, H. L. (1997, June 10). Data Assimilation Using an Ensemble Kalman Filter Technique. *MONTHLY WEATHER REVIEW*, pp. 796-811.

HOUTEKAMER, P. L., & MITCHELL, H. L. (2005). Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc*, 3269-3289.

Luffel, D. L., & Guidry, F. K. (1992, November 1). *New Core Analysis Methods for Measuring Reservoir Rock Properties of Devonian Shale.* Retrieved from Society of Petroleum Engineers: https://doi.org/10.2118/20571-PA

Mandel, J. (2009). *A Brief Tutorial on the Ensemble Kalman Filter.* Denver: Center for Computational Mathematics, University of Colorado.

McCULLOH, T. H. (1967). *Mass Properties of Sedimentary Rocks and Gravimetric Effect of Petroleum and Natural-GasReservoirs.* WASHINGTON: UNITED STATES GOVERNMENT PRINTING OFFICE.

Principal Component Analysis, S. E. (2002). *I.T. Jolliffe.* New York: Springer.

Rwechungura, R. W. (2011, January 1). *Advanced History Matching Techniques Reviewed. Society of Petroleum Engineers.* Retrieved from Society of Petroleum Engineers: https://doi.org/10.2118/142497-MS

Sebacher, B., Hanea, R., & Heemink, A. W. (2013). A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). *Computational Geosciences*.

Skjervheim, J. A., Eversen, G., Hove, J., & Vabo, J. G. (2011). An Ensemble Kalman Smoother for Nonlinear Dynamics. *An Ensemble Kalman Smoother for Nonlinear Dynamics.* Woodlands, Texas: An Ensemble Kalman Smoother for Nonlinear Dynamics.

Smith, L. I. (2002). *A tutorial on Principal Components Analysis.* Dunedin: Department of Computer Science, University of Otago.

Watanabe, S., & Datta-Gupta, A. (2012, June). Use of Phase Streamlines for Covariance Localization in Ensemble Kalman Filter for Three-Phase History Matching. *SPE Reservoir Evaluation and Engineering*, pp. 273-289.