2014

3D Variational Analysis In Subsurface Contaminant Transport Model

Anup Saha
North Carolina Agricultural and Technical State University

3D Variational Analysis in Subsurface Contaminant Transport Model

Anup Saha

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department: Civil, Architectural and Environmental Engineering

Major: Civil Engineering

Major Professor: Dr. Shoou-Yuh Chang

Greensboro, North Carolina

2014

The Graduate School
North Carolina Agricultural and Technical State University
This is to certify that the Master's Thesis of

Anup Saha

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2014

Approved by:

Dr. Shoou-Yuh Chang Major Professor	Dr. Manoj Kumar Jha Committee Member
Dr. Stephanie Luster-Teasley Committee Member	Dr. Sameer Hamoush Department Chair

Dr. Sanjiv Sarin
Dean, The Graduate School

Biographical Sketch

Anup Saha was born on December 12, 1982, in Laksam, Comilla, Bangladesh. He received his Bachelor of Science Degree in Civil Engineering from Bangladesh University of Engineering and Technology in June, 2005. He is currently a graduate student and Research Assistant in the Civil, Architectural and Environmental Engineering Department at North Carolina Agricultural and Technical State University, and a candidate for the Master of Science degree in Civil Engineering.

Dedication

This thesis is dedicated to my parents and my lovely niece, Aradhaya.

Acknowledgements

I would like to express my sincere appreciation and my heartfelt gratitude to Dr. Shoou-Yuh Chang, advisor of my graduate studies, for the guidance and support he shown throughout the study and for his constructive criticism throughout the period of my research and graduate studies; Dr. Manoj Kumar Jha and Dr. Stephanie Luster Teasley for being a part of my thesis committee, and for helping me with their assistance. My parents, brother, sister and relatives whose love and blessings have made this endeavor possible. My sincere thanks to Sikder Latif, Torupallab Ghoshal, Md Sayemuzzaman, Somsubhra Chattyapadhya, Bhasker Jha and Brooks L. Budd for their kind support. I also want to thank Xi Chen for being so nice and supportive. Finally I want to thank my country, the peoples' republic of Bangladesh who made me capable for this degree.

This research was funded by the Department of Energy Samuel Massie Chair of Excellence Program under grant no. DE-NA0000718. The views and conclusions contained herein are those of the writers and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the funding agency.

Table of Contents

List of Figures

List of Tables

Abstract

Modeling of contaminant transport in a subsurface environment by a numerical model deviates from the real world environment because of the highly heterogeneous nature of the subsurface environment. In this study, the data assimilation techniques are integrated with the numerical model and are applied to the subsurface environment to predict the contaminant transport. The Forward Time Center Space (FTCS) model is used as a numerical approach to solve the classical advection-dispersion-reaction transport equation and the Kalman Filter, Ensemble Kalman Filter (EnKF) and 3D Variational (3DVAR) analysis are used for data assimilation purpose. A hybrid scheme, termed as EnKF-3DVAR is developed using EnKF and 3DVAR analysis. The EnKF is a Monte Carlo based sequential data assimilation technique that divides the state vector into N number of ensembles rather than computing one state vector. The 3DVAR analysis uses the EnKF mean state vector as the background state and uses a cost function to find out an optimal estimate of that EnKF mean state vector. The simulation is run using an ensemble size of 100 members. A Root Mean Square Error (RMSE) profile is used to evaluate the prediction accuracy of the models. This study shows that state predictions are better for both the EnKF and EnKF-3DVAR when compared to those of the numerical and KF solutions. The introduction of 3DVAR analysis with EnKF is found to be effective in the reduction of the prediction errors. The EnKF-3DVAR model shows an error reduction of 22.3% from the EnKF solution. The mean RSME for the four models numerical, KF, EnKF and EnKF-3DVAR are 159.1 mg/L, 72.9 mg/L, 17.5 mg/L and 13.6 mg/L respectively.

Key words: Contaminant Transport; Kalman Filter; Ensemble Kalman Filter; Forward Time Center Space Method; 3D Variational Analysis; Root Mean Square Error.

CHAPTER 1

Introduction

Water is essential to all the living beings. Of all the water on earth about 97.5% is saline water and only 2.5% of the total water is fresh water. Other than glaciers and ice, groundwater is the largest available source of fresh water; about 30.1% of fresh water comes from the groundwater sources (Shiklomanov, 1993). With the increase of industrialization and other man made activities the groundwater is being polluted and as a result the quality of groundwater is deteriorating day by day. Increased demands for water have stimulated the development of groundwater resources. According to the national groundwater association (NGWA) and united states geological survey report (2005) groundwater contributed 26.4% of the nations water supply and 43.8% of the total population regularly depend on groundwater for drinking water and 79.6 billion gallons of groundwater was withdrawn daily. This fresh water supply is increasingly threatened by the increasing contamination of the groundwater. According to the recent study on the wells of the United states by Ruth et al. (2005), it is found that approximately 50% of the wells are contaminated by the pathogens. Approximately 74 pesticides were cited in the groundwater of 38 states, of which most are carcinogenic; thus, it is very important to prevent and control the contamination of groundwater from these pollutants. For the better control and management of the groundwater contamination, the type and sources of pollutants and their behavior through the subsurface porous media needs to be evaluated efficiently. If we can efficiently evaluate the pollutant behavior through the porous media, we can better manage the groundwater system and prevent the pollution of groundwater.

Fate and transport of contaminants in groundwater can be referred to the physical, chemical, and biological processes that impact the movement of contaminants from one point to

another and how they change with time. Modeling of contaminant transport in a subsurface environment gives us the idea of how the contaminants transport through the medium and change with time and space. The pollutant flow occurs in two processes i.e. advection where pollutant is transported with the flow of the groundwater and diffusion or hydrodynamic dispersion where the pollutants are diffused within the flow. The groundwater flow and the transport of pollutant can be considered as a practical problem. As the soil formations vary rapidly, it is very difficult to get the exact information about the subsurface reservoir system and how the pollutant would behave in the subsurface system. Also, there is lot of uncertainties involved in solving pollutant transport problems. Therefore, the results in a pollutant transport model may become erroneous due to the randomness of the system and also if the assumptions are made inaccurately. Moreover, the pollutant prediction in the groundwater and the transport of pollutant in the subsurface system depends on the widely varying information of the heterogenic aquifer system, reaction and decay mechanism of the system and advection and dispersion mechanism of the system. Thus, it has become a challenge for the engineers to accurately predict the pollutant transport into the subsurface environment.

Various analytical and numerical approaches are used to evaluate the contaminant transport problem in the subsurface media. These analytical and numerical models however give large errors and cannot predict the transport problem correctly due to their limitations to predict the randomness and heterogeneity of the subsurface medium. Thus it is impossible to attain a perfect analytical or numerical groundwater flow and transport model, as even a little error will grow incorrigibly with time. If these predictions from numerical models are used with data assimilation techniques the efficiency of the contaminant transport models can be improved.

Data assimilation can be defined as the technique of analyzing of a state by incorporating the observation data into the models (Bouttier and Courtier, 2001). Data assimilation techniques proceed by analysis cycles. In each cycle of the analysis, the state estimates from numerical models are updated by incorporating the observations from current state. Filtering techniques and its extensions when combined with numerical approaches can give us cost-effective results. Both the Kalman Filter and its extension Ensemble Kalman Filter are used effectively in subsurface transport problems. Kalman Filter essentially consists a set of mathematical equations that uses a predictor-corrector type estimator that minimizes the estimated error covariance (Welch and Bishop, 1995). It takes a prior estimation and observation measurements into consideration to compute the best prediction for a state variable. By calculating the covariances of state and observations these equations are solved. The Ensemble Kalman Filter is an extension of the discrete Kalman Filter technique and is first introduced by Evensen (1994). It is developed to handle more dynamic and uncertain behavior of the problems. It is also a sequential based data assimilation technique that uses Monte Carlo sampling method. In weather forecasting models there are significant uses of variational analysis as a data assimilation tool. Among the variational data assimilation techniques, 3D Variational (3DVAR) analysis is used numerously in weather prediction models and is found to be effective improving the results when combined with other models. 3DVAR uses a cost minimization function to find out an approximate optimal analysis. The conceptual simplicity of the 3DVAR analysis makes it popular. It uses the background state and background error covariance to calculate the analysis. It is important to design the background error covariance matrix properly. In practice, it has to be symmetric positive definite and realistic when expressed in terms of observation parameters as it guides the analysis (Bouttier and Courtier, 2001).

The objective of this study is to develop a three dimensional subsurface contaminant transport model using 3D Variational analysis and Ensemble Kalman Filter (EnKF-3DVAR) for continuous pollutant input and to compare the results with a simulated true solution. The study also aims to compare the EnKF-3DVAR solution with Numerical, Kalman Filter and Ensemble Kalman Filter solutions and to determine the prediction accuracy of the model.

CHAPTER 2

Literature Review

Geer (1982) used Kalman Filter to simulate one-dimensional phreatic groundwater flow where he showed a relationship between interpolation confidence of groundwater levels to the observation wells distance. In his study with KF for a nuclear waste management program, Pimentel et al. (1982) identified the aquifer parameters from noisy measurement data by taking on-line measurements from a single observation well. Geer et al. (1991) used KF with a deterministic numerical model to improve the uncertainties in numerical groundwater modeling. They performed the assimilation of spatial and temporal measure of hydraulic head by combining the KF with a deterministic groundwater model. Zou and Parr (1995) estimated the optimal concentration of a two dimensional pollutant plume in subsurface environment by using the KF with the state-space estimation technique. Chang and Latif (2009) modeled the one dimensional transport of benzene leachate from industrial landfill to the subsurface by using KF and particle filtering. It was found that KF gave 80% less error in comparison to conventional numerical approach. Walker et al. (2001) evaluated a one-dimensional soil moisture and temperature profile by comparing the KF and direct insertion scheme.

Evensen and Leeuwen (1996) assimilated EnKF with the Geosat altimeter data into a two-layer quasigeostrophic ocean model to study a western boundary current flowing along the east coast of South Africa. Houtekamer et al. (1998) used EnKF technique to assimilate data for an idealized environment. It was found that results were more accurate when data were assimilated with EnKF and the error decreased with the increase of the size of the ensemble members. They concluded by saying that 100 number of ensemble members were sufficient to describe the correlation between the parameters accurately. Lisæter et al. (2003) developed a

model by coupling EnKF with ice-ocean model where they used observed variable from passive microwave sensor data. Annan and Hargreaves (2004) demonstrated in a oceanographic model that the iterative EnKF efficiently performed the estimation of multivariate parameter in the presence of noisy environment. Wen and Chen (2005) estimated the reservoir model parameters and their associated uncertainties by using EnKF. The model continuously updated the each ensemble member to match the real-time data. Their results showed that a large number of measurements were needed to obtain a stable performance of the model. Moradkhani et al. (2005) used EnKF to develop a dual state-parameter estimation approach. The EnKF technique was used to sequentially estimate the parameters and state variables of a rainfall-runoff model. Vrugt et al. (2005) implemented a joint parameter and state estimation method for improved inverse modeling in subsurface flow and transport. The new method was named as Simultaneous Optimization and Data Assimilation (SODA). They verified this new approach with a one-dimensional subsurface conceptualization and EnKF was used as a recursive data assimilation technique to update model states. Huang et al. (2008b) demonstrated the effectiveness of assimilating in situ and remotely sensed observation data with EnKF in a one-dimensional land surface model. Huang et al. (2008a) used EnKF and Common Land Model to develop a one-dimensional land data assimilation scheme and the new scheme was found to improve the estimation of soil temperature profile. Huang et al. (2009) calibrated a heterogeneous conductivity field by using EnKF approach and improved solute transport prediction in subsurface with an unknown contaminant source. They performed their numerical experiment in a two dimensional heterogeneous flow and transport field. Their study results indicated that the EnKF method significantly improved the estimation of hydraulic conductivity distribution and solute transport prediction by assimilating hydraulic head measurements with a known initial

solute condition.

Hamill and Snyder (2000) demonstrated a hybrid approach of EnKF with 3D Variational (3DVAR) analysis using the quasigeostrophic model for perfect-model assumptions. They replaced the background-error covariance matrix with weighted sum of the 3DVAR background-error covariance. They found that the system gave good results for large ensemble size when the background error covariances were calculated from ensemble members. Thornhill et al. (2012) modeled the hydrodynamic and sediment transport of Morecambe Bay, UK using the integration of a Morpho-dynamic model with the 3DVAR data assimilation technique to assess the prediction efficiency of the newly coupled model in comparison with the model stand alone performance. They found that the calculated mean square error (MSE) and brier score showed a significant improvement for the data assimilation technique with 3DVAR analysis over the model stand-alone performance. The assimilation with 3DVAR analysis reduced the error substantially and improved the brier score to 0.262 from model stand-alone performance. Gao et al. (2006) derived a mathematical design of 3DVAR land data assimilation scheme (LDAS) with ECMWF model for numerical weather and climate models in China. Assimilating a single point observational datum into the background setup they tested the capacity of this LDAS scheme. The 3DVAR LDAS showed improvement in the background estimates with the observations of air temperature and the assimilation outputs showed better agreement with the air temperature variation trend than the background. It was also found that the assimilation scheme described the air temperature more accurately. The root mean square error of assimilation scheme was found to be 1.52, lower than that of background 2.65. Wang (2011) used a hybrid ensemble transform Kalman filter 3DVAR data assimilation scheme to track two major hurricanes forecasts, Ike and Gustav over the Gulf of Mexico in a weather research and forecasting (WRF) model. The flow

dependent covariances were generated by the ensemble transform Kalman filter to be used with the hybrid assimilation scheme whereas 3DVAR used the static background covariance. The root mean square error of the tracking of the hurricanes by the hybrid data assimilation scheme was smaller than that by the 3DVAR system gaining 1-2 days of lead time for hybrid method. The improvement was achieved by using the flow dependent error covariance generated by the ensemble transform Kalman filter. The flow dependent background error covariance provided the estimate of the uncertainty of the position of the hurricane. Wang et al. (2008a, 2008b) developed a hybrid Ensemble Transform Kalman Filter (ETKF)-3DVAR data assimilation scheme for Weather Research and Forecasting Model. Unlike WRF-3DVAR that used a static covariance, the hybrid system incorporated ensemble covariance with the static covariance to calculate the flow dependent forecast error statistics. In part-1 of his study he tested the hybrid scheme with system simulated observing system experiments under the perfect model assumptions. With using 50 ensemble members the hybrid system produced more accurate analyses over the 3DVAR analyses, which was about 15%-20%. Improvement was found to be larger over the region where data ware sparsely distributed than the densely data distributed regions. In part-2 for a real life experiments, a domain of North America was considered with a grid spacing of 200km. The hybrid system produced a better results and less error than that of the 3DVAR system in a 12h forecasts. The hybrid system produced 9%-11% improvement for the wind forecasting and 3% improvement for temperature forecasting over 3DVAR system. In their study Barker et al. (2004) performed a case study to illustrate the 3DVAR response to a single surface pressure observation where a significant improvement was found for the wind forecasting though the improvement for temperature and humidity was marginal. Hsiao et al. (2012) evaluated the impact of outer loop and partial cycling with the WRF3DVAR by analyzing

78 forecasts of three typhoons during 2008 for Taiwan's Central Weather Bureau. The WRF3DVAR reduced typhoon track forecast errors by more than 30%. The use of the outer loop allowed more observations and produced more accurate analyses. On the other hand, use of the partial cycling showed a slight improvement in typhoon track. Xiao et al. (2005) evaluated the effect of Doppler velocities on heavy rainfall by assimilating radial velocities in 3DVAR analysis scheme with the fifth generation Pennsylvania State University-NCAR Mesoscale Model (MM5). The assimilation of vertical velocity component of the Doppler radial velocity observation was adopted through the introduction of vertical velocity increments and forecasts of cloud water and rainwater were used in the 3DVAR background. The developed system was found to be producing less error compared to the experimental results without radar data assimilation. The wind and vertical velocity analysis were found to be reasonable with Doppler velocity data assimilation. The results for rainfall forecasts were also better and well defined when the radar velocity data were assimilated with 3DVAR. In their study Xiong et al. (2013) developed a hybrid scheme called Breeding Growing Mode (BGM)-3DVAR data assimilation system for heavy rain forecasting. They used BGM-ensemble technique as background filed and flow dependent background error covariance for the hybrid scheme to produce a better rain forecast. They showed that successful inclusion of the background field was very important as almost 85% of the information would come from it and the transfer of the observational information form one time step to the next was very crucial as it would influence the analysis.

The intention of this paper is to apply 3DVAR analysis with EnKF to develop a hybrid three dimensional contaminant transport model and evaluate the performance of this model for a continuous pollutant source in subsurface environment.

CHAPTER 3

Methodology

3.1 Three Dimensional Contaminant Transport Equation

The classical advection-dispersion-reaction equation with flow dominantly in the x-direction is used to represent the solute transport in a three-dimensional heterogeneous saturated porous media. The advection-dispersion-reaction equation (Schwartz and Zhang, 2004) is given below.

$$R\frac{\partial c}{\partial t} = D_x\frac{\partial^2 c}{\partial x^2} + D_y\frac{\partial^2 c}{\partial y^2} + D_z\frac{\partial^2 c}{\partial z^2} - v(x)\frac{\partial c}{\partial x} - \kappa Rc \qquad (1)$$

where,

R = Retardation factor, dimensionless;

c = solute concentration, mg/l;

t = time, day;

Dx, Dy and Dz = hydrodynamic dispersions in the x, y, and z directions, respectively, m^2/day

V(x) =linear pore water velocity in X direction, m/day; and

k=decay constant of pollutants, /day.

The boundary conditions with a continuous pollutant source can be given as

$$c(x_o, y_o, z_o, t) = C_o;$$

$$\frac{\partial c}{\partial x} = \frac{\partial c}{\partial y} = \frac{\partial c}{\partial z} = 0, \text{for } x = y = z = \infty$$

Here (x_o, y_o, z_o) is the pollutant input point and C_o is the continuous input pollutant concentration, mg/l.

3.2 Analytical Solution

For the three dimensional subsurface contaminant transport model for continuous pollutant source, the analytical solution for the equation (1) is obtained from the Domenico and Schwartz (1997) equation. The solution is

$$C(x,y,z,t) = \frac{C_0}{8} e^{\frac{xv_x}{2D_x}\left[1-\left(1+4kD_x/v_x^2\right)^{1/2}\right]} erfc\left(\frac{x - v_x t / R_f\left(1+4kD_x / v_x^2\right)^{1/2}}{2\sqrt{D_x t / R}}\right)$$
$$\times\frac{1}{2}\left\{erf\left[\frac{y+Y/2}{2(D_y x / v_x)^{1/2}}\right] - erf\left[\frac{y-Y/2}{2(D_y x / v_x)^{1/2}}\right]\right\}$$
$$\times\frac{1}{2}\left\{erf\left[\frac{z+Z}{2(D_z x / v_x)^{1/2}}\right] - erf\left[\frac{z-Z}{2(D_z x / v_x)^{1/2}}\right]\right\} \qquad (2)$$

where,

Co = input pollutant concentration (mg/l)

V_x = ground-water velocity in the x direction

Y= width of waste source in the saturated zone (L)

Z= depth of waste source in the saturated zone (L)

erf= error function

erfc= complementary error function

3.3 Simulated True Solution

The analytical solution for transport model is used in this study to generate the true solution. The true solution is made by adding 5% random Gaussian error with the analytical solution. This simulated true solution is then used as a reference dataset to compare the model predictions.

3.4 Numerical Model

The Forward Time Central Space (FTCS) finite difference approach is used to solve the three dimensional advection-dispersion-reaction transport equation (Equation-1). This approach is used by Zou and Parr (1995) to predict contaminant transport in a two dimensional field. In this study the vertical terms for three-dimensional field in z-axis have been added.

Let C(i,j,k,t) = C(xi, yj, zk, t), we can get

$$\frac{\partial c}{\partial t} \approx \frac{C(i,j,k,t+1) - C(i,j,k,t)}{dt} \tag{3}$$

$$\frac{\partial c}{\partial x} \approx \frac{C(i+1,j,k,t) - C(i-1,j,k,t)}{2dx} \tag{4}$$

$$\frac{\partial^2 c}{\partial x^2} \approx \frac{C(i+1,j,k,t) - 2C(i,j,k,t) + C(i-1,j,k,t)}{dx^2} \tag{5}$$

Similarly,

$$\frac{\partial^2 c}{\partial y^2} \approx \frac{C(i,j+1,k,t) - 2C(i,j,k,t) + C(i,j-1,k,t)}{dy^2} \tag{6}$$

$$\frac{\partial^2 c}{\partial z^2} \approx \frac{C(i,j,k,+1t) - 2C(i,j,k,t) + C(i,j,k-1,t)}{dz^2} \tag{7}$$

and

$$kC = \frac{C(i,j,k,t+1) + C(i,j,k,t)}{2} \tag{8}$$

Substituting these equation into equation (2) we can get

$$C(i,j,k,t) = b_1 C(i-1,j,k,t) + b_2 C(i,j,k,t) + b_3 C(i+1,j,k,t) + b_4 C(i,j-1,k,t) + b_5 C(i,j+1,k,t) + b_6 C(i,j,k-1,t) + b_7 C(i,j,k+1,t) \tag{9}$$

where,

$$b_1 = \left(\frac{D_x dt}{Rdx^2} + \frac{V_x dt}{2Rdx}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (10)$$

$$b_2 = \left(1 - \frac{2D_x dt}{Rdx^2} - \frac{2D_y dt}{Rdy^2} - \frac{2D_z dt}{Rdz^2} - \frac{kdt}{2}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (11)$$

$$b_3 = \left(\frac{D_x dt}{Rdx^2} - \frac{V_x dt}{2Rdx}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (12)$$

$$b_4 = b_5 = \left(\frac{D_y dt}{Rdy^2}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (13)$$

$$b_6 = b_7 = \left(\frac{D_z dt}{Rdz^2}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (14)$$

$$b_2 = \left(1 - \frac{2D_x dt}{Rdx^2} - \frac{2D_y dt}{Rdy^2} - \frac{2D_z dt}{Rdz^2} - \frac{kdt}{2}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (15)$$

$$b_3 = \left(\frac{D_x dt}{Rdx^2} - \frac{V_x dt}{2Rdx}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (16)$$

$$b_4 = b_5 = \left(\frac{D_y dt}{Rdy^2}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (17)$$

$$b_6 = b_7 = \left(\frac{D_z dt}{Rdz^2}\right) \Bigg/ \left(1 + \frac{kdt}{2}\right) \quad (18)$$

After time-space discretization in the three dimensional space, the linear system of equations can be written in the following vector-matrix form

$$C_{t+1} = A_t C_t \quad (19)$$

where,

C_{t+1} = solute concentration at time step t+1 at all nodes

C_t = solute concentration at time step t at all nodes,

A_t = linear operator or state transition matrix that contains the parameter for model.

For each time period dt, the concentration vector to the next step (C_{t+1}) is calculated by multiplying the state transition matrix by the previous concentration vector (C_t). The boundary conditions are adopted in such a way that the effects of adjacent nodes are neglected in case of determining the concentration of the sides of the aquifer such as top and lower boundaries as there did not exist all directional flow. So the state transition matrix is modified for top and bottom layer as the effect of coefficient b7 is omitted while multiplication. So to adjust the state transition matrix to calculate for the effect of b7 for theses cases while multiplying the value of b2 is modified to (b2+b7). The equation can be re-expressed with a new denotation by

$$x_{t+1} = A_t \times x_t \qquad (20)$$

x_{t+1} = the state vector at time step t+1 at all nodes

x_t = the state vector at time step t at all nodes,

The stability and convergence criteria for this numerical model are as follows

$$dx < \frac{2D_x}{V_x} \quad \text{and} \quad dt < \frac{0.5R}{\left(\frac{D_x}{dx^2} + \frac{D_y}{dy^2} + \frac{D_z}{dz^2}\right)} \qquad (21)$$

dx, dy, dz are the space discretization along the X, Y and Z directions

dt is the time discretization

Dx, Dy and Dz are dispersion coefficients along the X, Y and Z directions

Vx is the velocity of groundwater movement along the X direction

For highly uncertain characteristics of the real world transport problems the state estimate for data assimilation process can be represented as numerical dynamics with added noise to it (Pham, 2001).

$$x_{t+1} = A_t x_t + p_t \qquad (22)$$

x_{t+1} = state vector at time t+1

x_t = state vector at time t

A_t = linear operator containing all the parameter values or state transition matrix which transform the state vectors from one time step to the next

p_t = random noise called process noise.

Similarly for observation data we can write,

$$z_{t+1} = H_t z_t + o_t \qquad (23)$$

z_{t+1} = observation vector at time t+1

z_t = observation vector at time t

H_t = observation data pattern matrix

o_t = random noise called measurement noise

3.5 Kalman Filter

Kalman filter essentially contains a set of mathematical equations that uses the observation data to produce the probable best analysis of a state estimate. The algorithm of Kalman filtering works in a two-step process. In first step called the prediction step, the filter estimates the current state and error covariance estimates. Then in second stage called correction, it incorporates the observation data with the current state to update the result. These estimates are updated using a weighted average called Kalman gain giving more weight to the estimates with higher certainty. Being a recursive process it uses only the present measurements and previous state estimates and does not need all the previous information to predict the state. The equation for the Kalman filtering after adjustment can be written as

$$x_t(+) = [x_t(-) + K_t(z_t - Hx_t(-)] \qquad (24)$$

where,

$x_t(+)$ = the updated state after the adjustment

$x_t(-)$ = the predicted state before the adjustment

K_t = the Kalman gain

The Kalman gain matrix is be found from the following equation

$$K_t = P_t(-)H^T[HP_t(-)H^T + R_t]^{-1} \qquad (25)$$

where,

$P_t(+)$ = the optimal state covariance matrix

()T and ()- denote the transpose and inverse of matrix

P_t is calculated from the following recursive equation,

$$P_t(+) = P_t(-) - K_t HP_t(-) = (I - K_t H)P_t(-) \qquad (26)$$
$$P_t(-) = AP_t(+)A^T Q_t$$

I = identity matrix

Figure-1 (Chang and Latif, 2009) shows the two stage operations of the Kalman Filter.

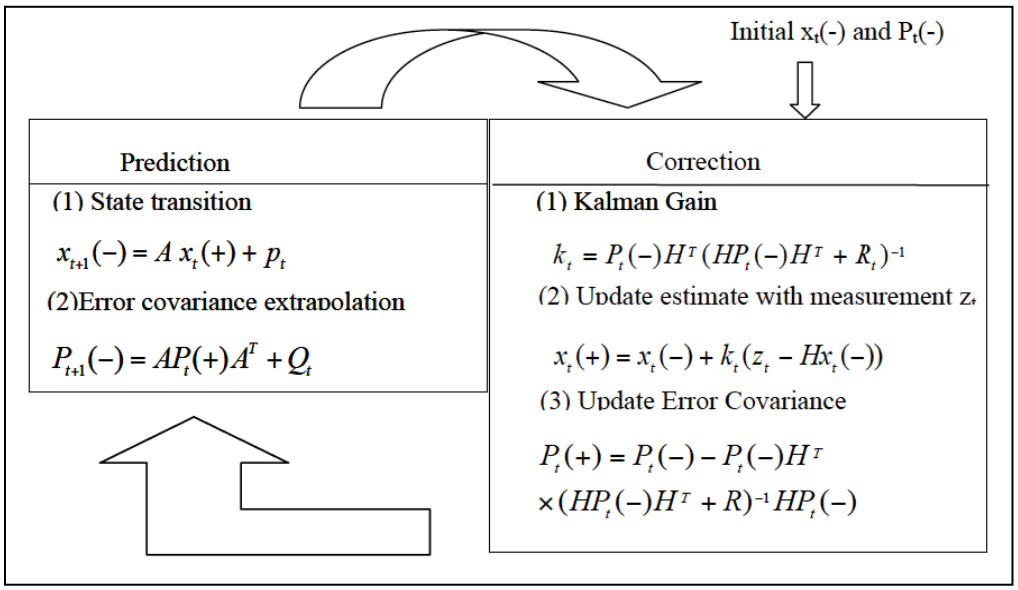

Figure 1: The two stage Kalman Filter Process

3.6 Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) is an extension of the discrete Kalman filtering technique, developed to handle more dynamic and uncertain behavior of the problems that are non-linear in characteristics. It is a sequential based data assimilation technique proposed by Evensen (1994) that uses Monte Carlo sampling method so that the traditional and computationally expensive approximate error covariance matrixes could be avoided. Other than calculating only one state vector, this approach divides the one state vector into N number of samples or ensembles, analyzes each ensemble members and then takes the mean of the ensembles as the final state vector for which the approach is termed as Ensemble technique.

According to Lisæter et al. (2003) using the information of a certain time, in this approach an ensemble of model states is run in parallel when the analysis is carried in the discrete time. As in discrete Kalman filtering observations are assimilated sequentially with model state forward in time whereas in EnKF every time a new data set is created for forward time integration and as the analysis is done recursively there is no need for storing all the previous time step data sets. In EnKF, posterior estimate of the previous step is used to calculate the prior state estimation for the next time step and then the observation of the current step is introduced to update the prior estimate. In the EnKF technique it runs on multiple state vectors, which are generated from a first guess state vector. These multiple vectors that are generated artificially called ensemble members. The mean of these artificially generated ensemble members are taken to be the first guess. The generation of these ensemble members is created using the Monte Carlo sampling method. The first guess is then perturbed with known statistical variables to generate the ensemble members. In this approach for each time sequence EnKF gives two estimates. One is called prior estimate or model prediction and the posterior estimate

or model analysis. After model prediction is calculated the observation data is incorporated to update the prior estimate and then the posterior state is calculated. The posterior state is the final analysis of the EnKF. Figure-2 shows the operation of sequential Ensemble Kalman Filter.

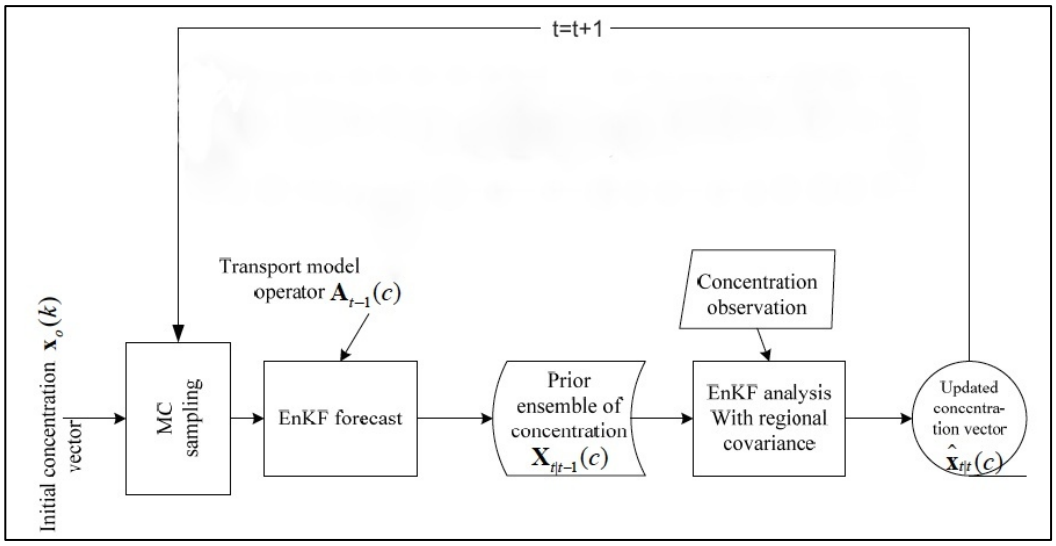

Figure 2: Ensemble Kalman Filter Operation

The steps and processes of Ensemble Kalman Filter are described as follows as described in Evensen (2003, 2004). An ensemble state matrix after Monte Carlo sampling with N number of ensembles can states as

$$x_i \in \Re^n, (i = 1,.....N)$$

$$X_{t|t-1} = [x_1, x_2, ..., x_N] \in \Re^{n \times N} \qquad (27)$$

where, n is the size of model state vector.

The ensemble mean state can be calculated as

$$\overline{X}_{t|t-1} = X_{t|t-1} B \qquad (28)$$

where,

$\overline{X}_{t|t-1}$ is the ensemble mean matrix and

$B \in \Re^{N \times N}$ is the matrix where each element is equal to 1/N.

The ensemble residual matrix can be defined as,

$$\Phi = X_{t|t-1} - \overline{X}_{t|t-1} \tag{29}$$

The prior ensemble covariance matrix $P_{e,t|t-1} \in \Re^{n \times n}$ can be defined as

$$P_{e,t|t-1} = \frac{\Phi\Phi^T}{N-1} = P_e \tag{30}$$

Now let the observation vector to be $z \in \Re^m$ with m number of observation grid points. Now we also can generate ensemble observation matrices with the addition of noises. For N number of ensemble member, the perturbed observations are,

$$z_j = z + \xi_j, j = 1,...,N \tag{31}$$

We can write,

$$Z_e = [z_1, z_2, ..., z_N] \in \Re^{m \times N} \tag{32}$$

Now the mean of the observation is

$$\overline{Z}_{t|t-1} = Z_{t|t-1} B \tag{33}$$

The observation residual can be calculated as

$$\xi = Z_{t|t-1} - \overline{Z}_{t|t-1} \tag{34}$$

The residual of observation

$$\gamma = [\xi_1, \xi_2, ..., \xi_N] \in \Re^{m \times N} \tag{35}$$

Now the ensemble observation error covariance matrix can be represented as

$$R_e = \frac{\gamma\gamma^T}{N-1} \tag{36}$$

The ensemble Kalman gain can be calculated by

$$K_e = P_e H^T S_e^{-1} \qquad (37)$$

where,

H = observation operator;

S_e = the ensemble innovation covariance matrix, can be expressed by

$$S_e = HP_e H^T + R_e \qquad (38)$$

The residual matrix is defined as

$$Z'_{t|t-1} = Z_e - HX_{t|t-1} \qquad (39)$$

Now we can get the final posterior estimate of the state matrix as follows

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + K_e Z'_{t|t-1} \\ &= X_{t|t-1} + P_e H^T S_e^{-1} Z'_{t|t-1} \\ &= X_{t|t-1} + P_e H^T (HP_e H^T + R_e)^{-1} Z'_{t|t-1} \\ &= X_{t|t-1} + \Phi\Phi^T H^T (HEE^T H^T + \gamma\gamma^T)^{-1} Z'_{t|t-1} \end{aligned} \qquad (40)$$

$\overline{X}_{t|t}$ is the posterior estimate of the prior state $\overline{X}_{t|t-1}$.

The final equation for calculating the posterior matrix involves inversion of covariance matrices, which has the potentiality of producing the singularity issues. To overcome with this singularity problems, Evensen (2003) prescribed a technique called pseudo-inverse. According to his study, the eigenvalue decomposition has been performed to find the inverse or pseudo inverse even if the matrix is singular.

$$(H\Phi\Phi^T H^T + \gamma\gamma^T)^{-1} = D\Lambda^{-1} D^T, \in \Re^{m \times m} \qquad (41)$$

The rank of $D\Lambda^{-1}D^T$ is less than or equal to N and, therefore, Λ will have N or less non-zero eigenvalues.

For uncorrelated measurement error, the following is valid (Evensen, 2003)

$$(H\Phi\Phi^T H^T + \gamma\gamma^T) = (H\Phi + \gamma)(H\Phi + \gamma)^T \qquad (42)$$

The singular value decomposition (SVD) of the $m \times N$ matrix,

$$(H\Phi + \gamma) = U\Gamma V^T$$

$$(H\Phi + \gamma)(H\Phi + \gamma)^T = U\Gamma V^T V\Gamma^T U^T = U\Gamma I\Gamma^T U^T = U\Gamma\Gamma^T U^T \qquad (43)$$

We assume the number of observation grid points is less than the ensemble members, m<N.

For m<N,

$$(H\Phi\Phi^T H^T + \gamma\gamma^T) = U\Gamma\Gamma^T U = U\Lambda U^T \qquad (44)$$

And the posterior ensemble state matrix

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + \Phi\Phi^T H^T U\Lambda^{-1}U^T Z'_{t|t-1} \\ &= X_{t|t-1} + E(H\Phi)^T U\Lambda^{-1}U^T Z'_{t|t-1} \end{aligned} \qquad (45)$$

The above equation can be divided into a more computationally efficient scheme by the following sequential steps

$$X_1 = \Lambda^{-1}U^T \qquad \in \Re^{N\times m} \qquad (46)$$

$$X_2 = X_1 Z'_{t|t-1} \qquad \in \Re^{N\times m} \qquad (47)$$

$$X_3 = UX_2 \qquad \in \Re^{m\times N} \qquad (48)$$

$$X_4 = E(H\Phi)^T \qquad \in \Re^{n\times m} \qquad (49)$$

$$X_{t|t} = X_{t|t-1} + X_4 X_3 \quad \in \Re^{n\times N} \qquad (50)$$

And the ensembles mean posterior state can be found as

$$\overline{X}_{t|t} = X_{t|t}B \qquad (51)$$

Any column of $\overline{X}_{t|t}$ gives the posterior estimate $\hat{\mathbf{x}}_{t|t}$.

The posterior ensemble covariance matrix can be calculated as

$$P_{e,t|t} = \frac{1}{N}(X_{t|t} - \overline{X}_{t|t})(X_{t|t} - \overline{X}_{t|t})^T \qquad (52)$$

This posterior state ensemble matrix at time step t, $X_{t|t}$ is then used to predict the prior state at time step t+1 by multiplying with the state transition operator A_t,

$$X_{t+1|t} = A_t X_{t|t} \qquad (53)$$

3.7 3D Variational Analysis

The concept of 3D Variational analysis is to find an approximate optimal analysis by minimizing a cost function. The principal lies in the fact that it completely avoids the calculation of computationally expensive Kalman Gain by looking for an approximate solution by the minimization function (Bouttier and Courtier, 2001). As described in Hamill and Snyder (2000) the cost function for 3D Variational analysis can be written as,

$$J(x) = \frac{1}{2}[(x_a - x_b)^T B_e^{-1}(x_a - x_b) + (z_t - Hx_a)^T R_e^{-1}(z_t - Hx_a)] \qquad (54)$$

where,

x_a = analysis at time step t

x_b = background state at time step t

B_e = background error covariance matrix

H = observation data pattern matrix

R_e = observation error covariance matrix

z_t = observation at time step t

Figure-3 below shows the minimization process of 3D VAR analysis *(Bouttier and Courtier, 2001)*.

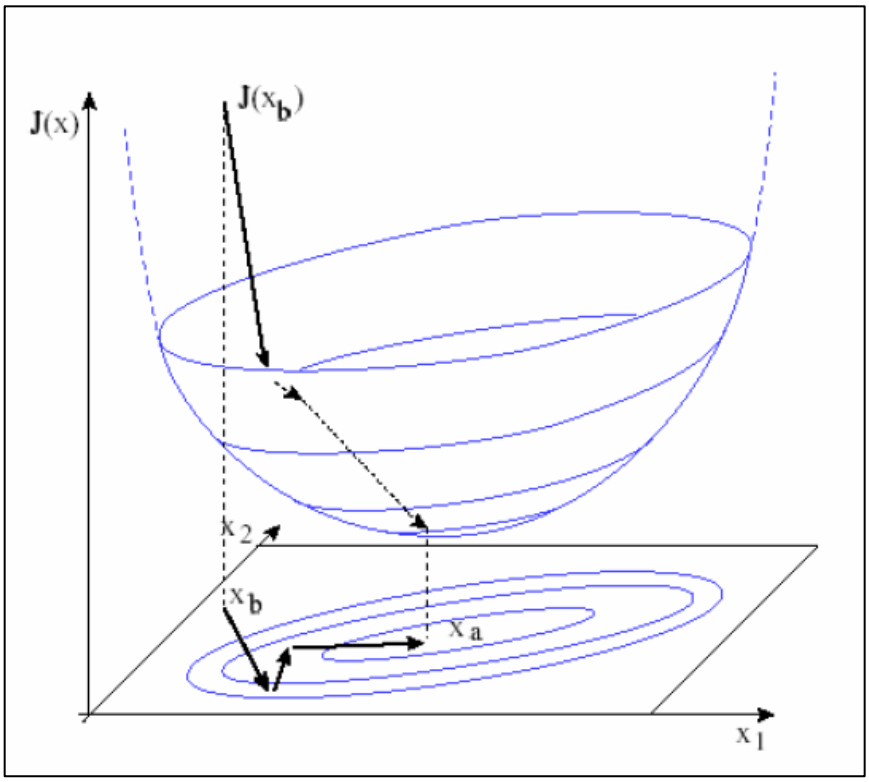

Figure 3: Schematic representation of 3D Variational cost-function minimization

The quadratic cost function is paraboloid in shape and has the optimum solution at x_a. The minimization function operates by doing several line-search to find the best possible value of x where cost function is minimum. In practice the background state is assumed to be the initial point or first guess. Although the analysis will not depend on the first guess if the minimization process progresses effectively. The difficulty with the 3D Variational analysis is that its efficiency depends on properly defining the background error covariance matrix B_e as it directs the solution for each time step and it needs to be positive definite to be invertible. In our study prior covariance matrix from Ensemble Kalman Filter solution is used as background error covariance matrix. The observation covariance matrix R_e is considered as a block diagonal matrix.

The gradient of the cost function w.r.t the x is,

$$dJ(x) = (B_e^{-1} + H^T R_e^{-1} H)(x_a - x_b) - H^T R_e^{-1}(z_t - Hx_t) \quad (55)$$

Now for minimum J(x), we can set dJ(x=xa)=0 we can write the equation as

$$(B_e^{-1} + H^T R_e^{-1} H)(x_a - x_b) = H^T R_e^{-1}(z_t - Hx_t) \quad (56)$$

After rearranging the terms we can get the final equation as

$$x_a = x_b + (B_e^{-1} + H^T R_e^{-1} H)^{-1} H^T R_e^{-1}(z_t - Hx_t) \quad (57)$$

If we write the equation in terms of time steps as

$$x_{t+1|t} = x_{t|t} + (B_e^{-1} + H^T R_e^{-1} H)^{-1} H^T R_e^{-1}(z_t - Hx_t) \quad (58)$$

In EnKF-3DVAR hybrid scheme, the ensemble mean forecast from EnKF model is updated by the 3DVAR optimization process to get the final analysis. For this study with smaller domain space we perform the direct inversion to get the final analysis rather than an iterative approach.

3.8 The Efficiency of The Models

The prediction accuracy of the models is evaluated by determining the Root Mean Square Error (RMSE). The RMSE can be calculated form the following formula.

$$RMSE = \sqrt{\frac{1}{N-1} \sum_{i,j,k} [Ct(x,y,z,t) - Cm(x,y,z,t)]^2} \quad (59)$$

where,

N= Number of nodes

Ct= True/Reference Solution

Cm= Model Solution

In this study, we consider a three dimensional space with five layers. The numbers of grid points in X, Y and Z-directions are 10, 10 and 5 respectively and a total of 10x10x5=500 nodes. The space discretization in X, Y directions are 2.5 meters and in Z direction is 2.0 meter. That is, the total domain area is 25 meter by 25 meter with a depth of 10 meter. The flow of system is considered to be only in the X-direction and flows in other directions are neglected. Dispersions in all three directions are also considered. The model simulation is run for total 30 time steps with each time step of 1 day. The initial pollutant (Co) is introduced in the first layer with the input point (x, y, z)=(1, 5, 1). The table-1 below shows the parameters used (Cheng, 2002) for the three dimensional contaminant transport model.

Table 1: Parameters for three dimensional groundwater contaminant transport model

No. of grid points along X-direction	10
No. of grid points along Y-direction	10
No. of grid points along Z-direction	5
Total number of Nodes	500
Grid interval along X-direction, dx	2.5m
Grid interval along Y-direction, dy	2.5m
Grid interval along Z-direction, dz	2.0m
Total Domain Space	25mx25mx10m
Velocity in X-direction, Vx	0.5 m/day
Dispersion in X-direction, Dx	1.0 m2/day
Dispersion in X-direction, Dy	0.5 m2/day
Dispersion in X-direction, Dz	0.7 m2/day
Retardation Factor, R	1.5
Total Assimilation Time Step, tn	30
Each Time Step Interval, dt	1 day
Total Assimilation Time	30 days
Process Noise	10%
Observation Noise	2.5%
Total Ensemble Members	100

CHAPTER 4

Results and Discussion

In this simulation the simulated results from the analytical solution is termed as analytical solution, which is used to generate our simulated reference or "True Solution". The numerical solution is termed as "Numerical Solution", Kalman Filter as "KF Solution", Ensemble Kalman Filter Solution as "EnKF Solution" and Ensemble Kalman Filter with 3D Variational analysis solution as "EnKF-3DVAR Solution". The results from our study are presented in graphical form. For this study we plot two-dimensional contour plots.

Figure. 4 to Figure-11 below show the concentration contour profile of the pollutants at beginning time steps (time step 5 and time step 10).

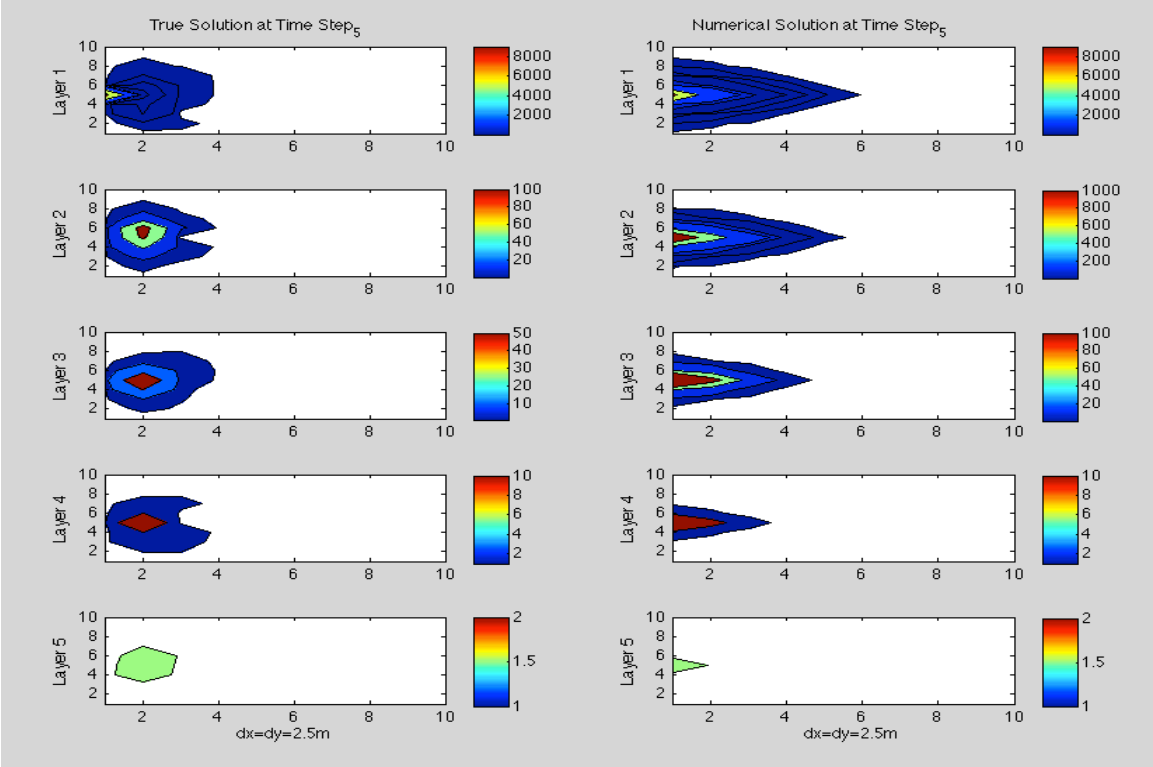

Figure 4: Contour profile for True Solution vs Numerical Solution at time step 5.

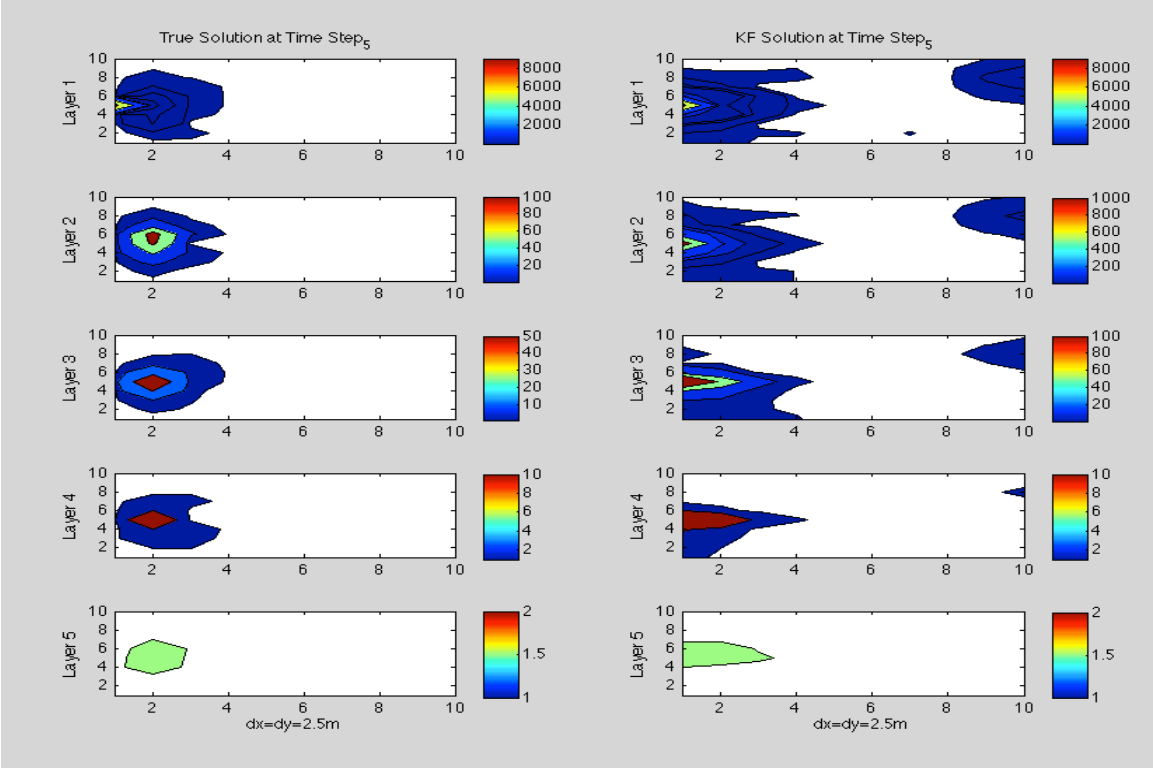

Figure 5: Contour profile for True Solution vs KF Solution at time step 5.

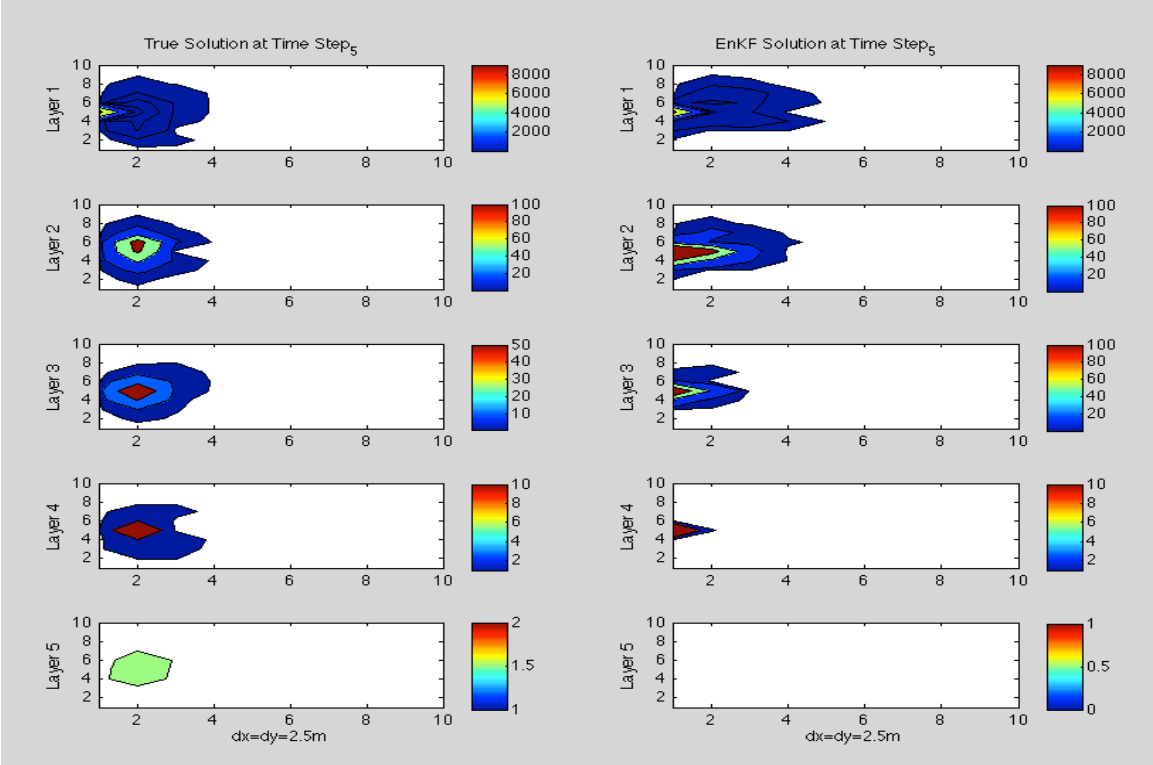

Figure 6: Contour profile for True Solution vs EnKF Solution at time step 5.

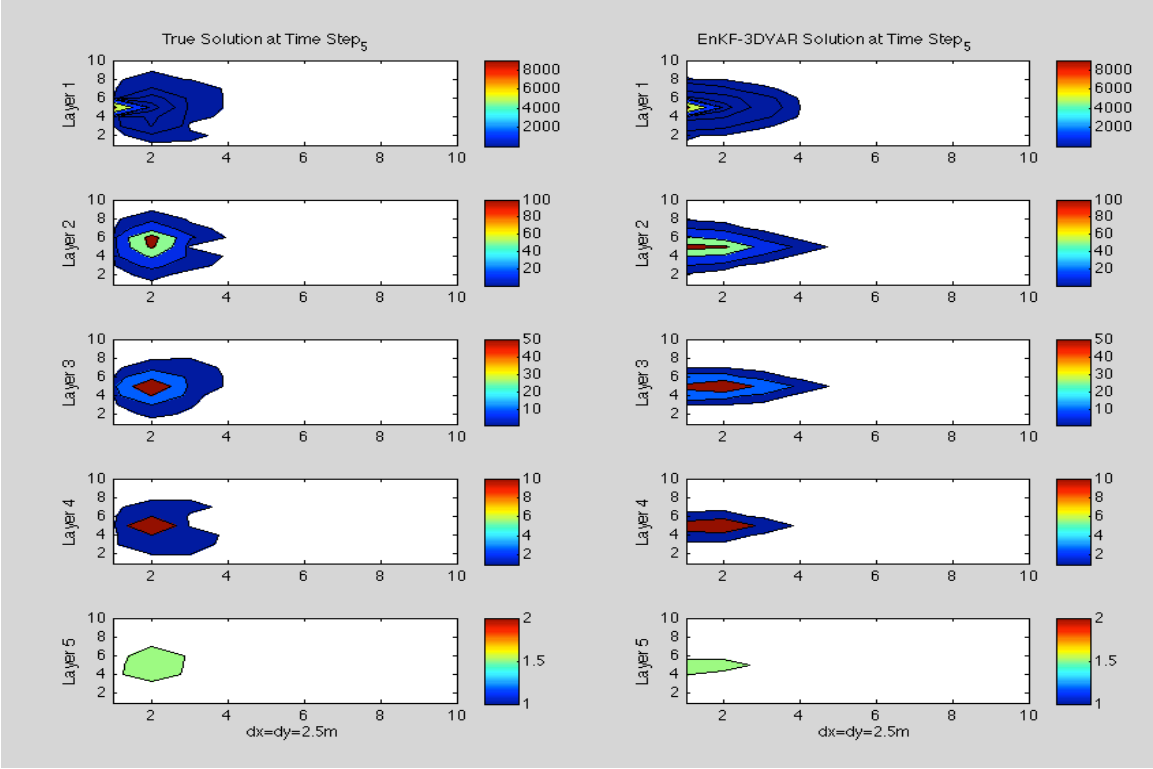

Figure 7: Contour profile for True Solution vs EnKF-3DVAR Solution at time step 5

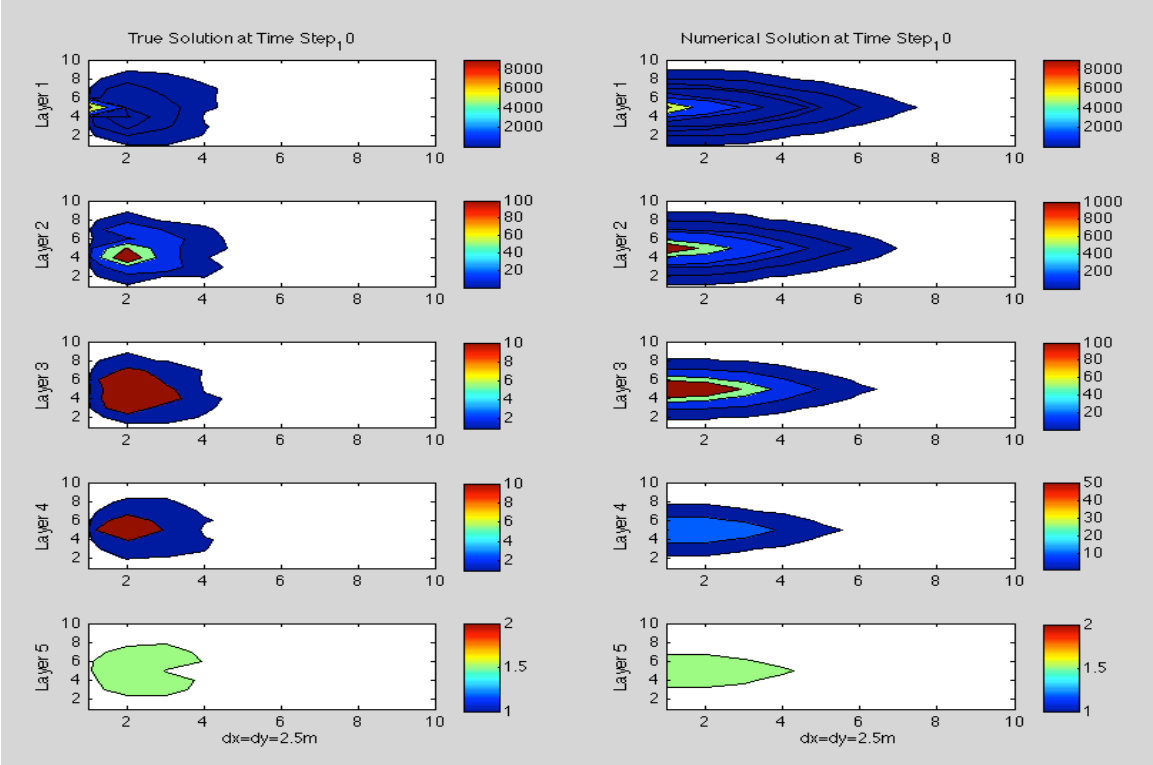

Figure 8: Contour profile for True Solution vs Numerical Solution at time step 10

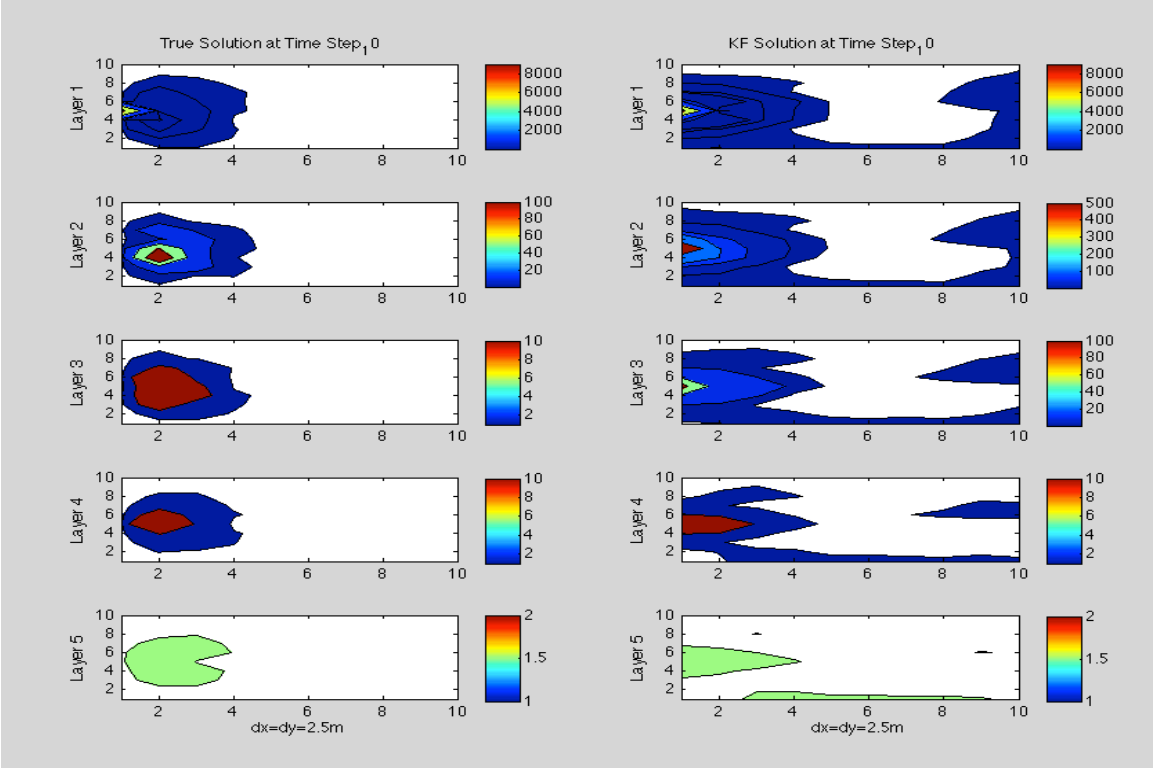

Figure 9: Contour profile for True Solution vs KF Solution at time step 10

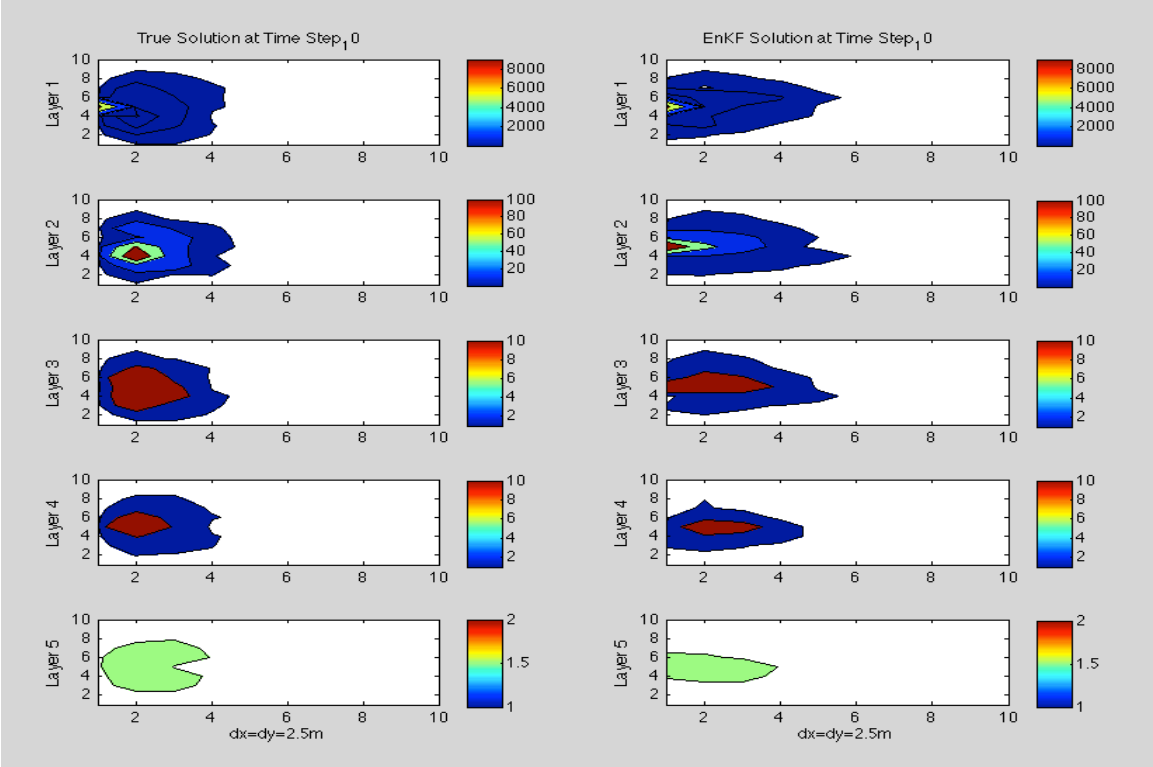

Figure 10: Contour profile for True Solution vs EnKF Solution at time step 10

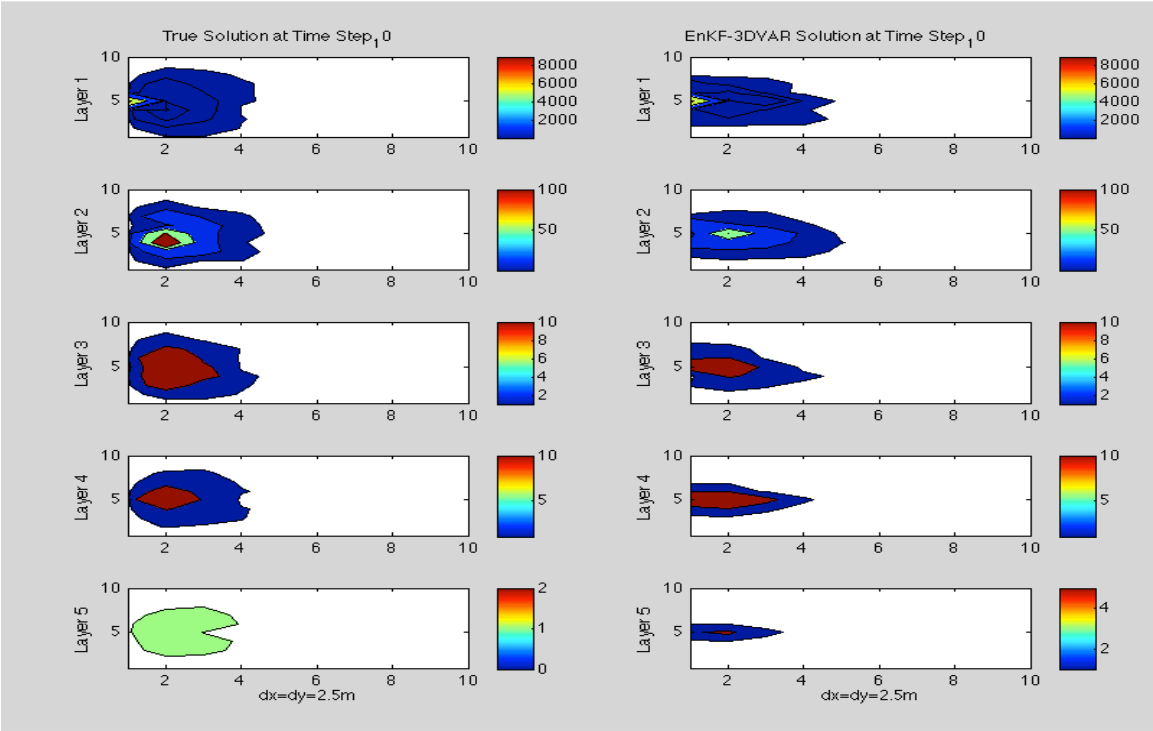

Figure 11: Contour profile for True Solution vs EnKF-3DVAR Solution at time step 10

It is found that at time step 5 the numerical solution moves faster than the simulated true solution and it also gives concentration contour values higher than the true solution. Though time step 5 is the beginning stage of the data assimilation, the KF, EnKF and EnKF-3DVAR solution start to evaluate the real situation by incorporating the observation data into the system that eventually lower the concentration values for these models. At the beginning time step the KF model shows some system noise around the domain space whereas the EnKF and EnKF-3DVAR solution give better contour shapes. To understand the change of the pollutant concentration with time, the concentration contour profile for the intermediate time steps (time step 15, 20, and 25) for all the models have been presented in figures 12 to 23.

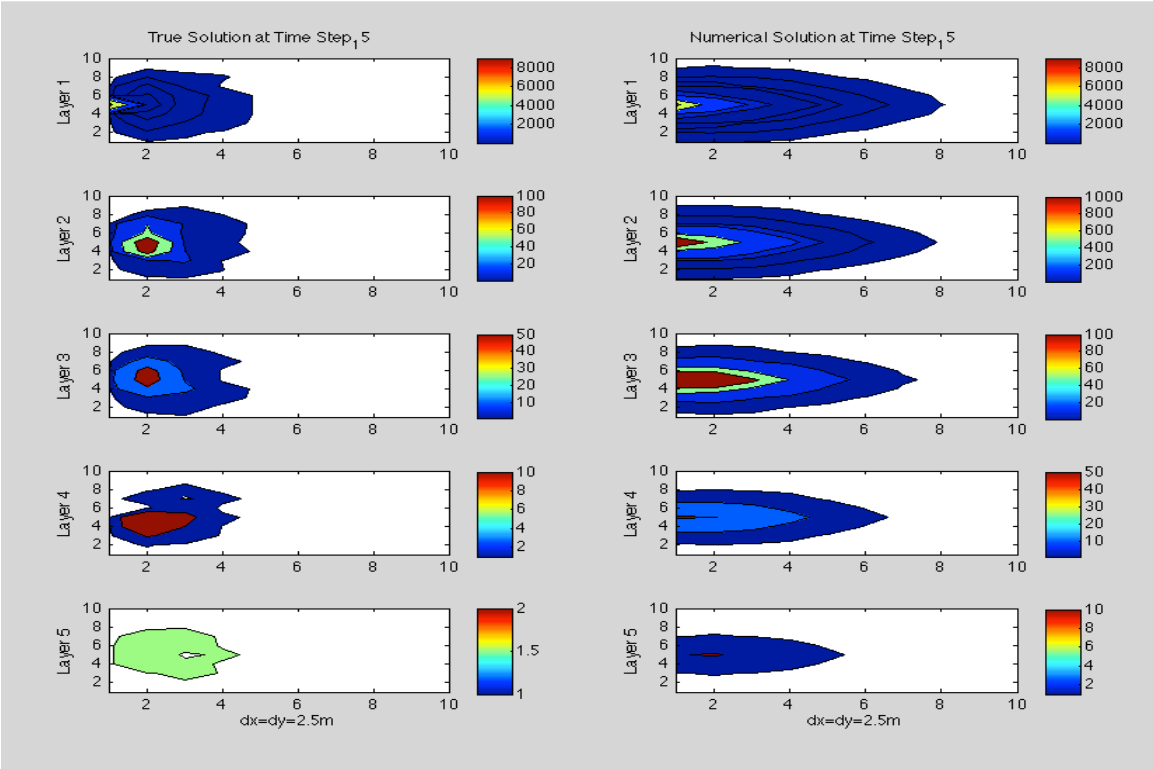

Figure 12: Contour profile for True Solution vs Numerical Solution at time step 15

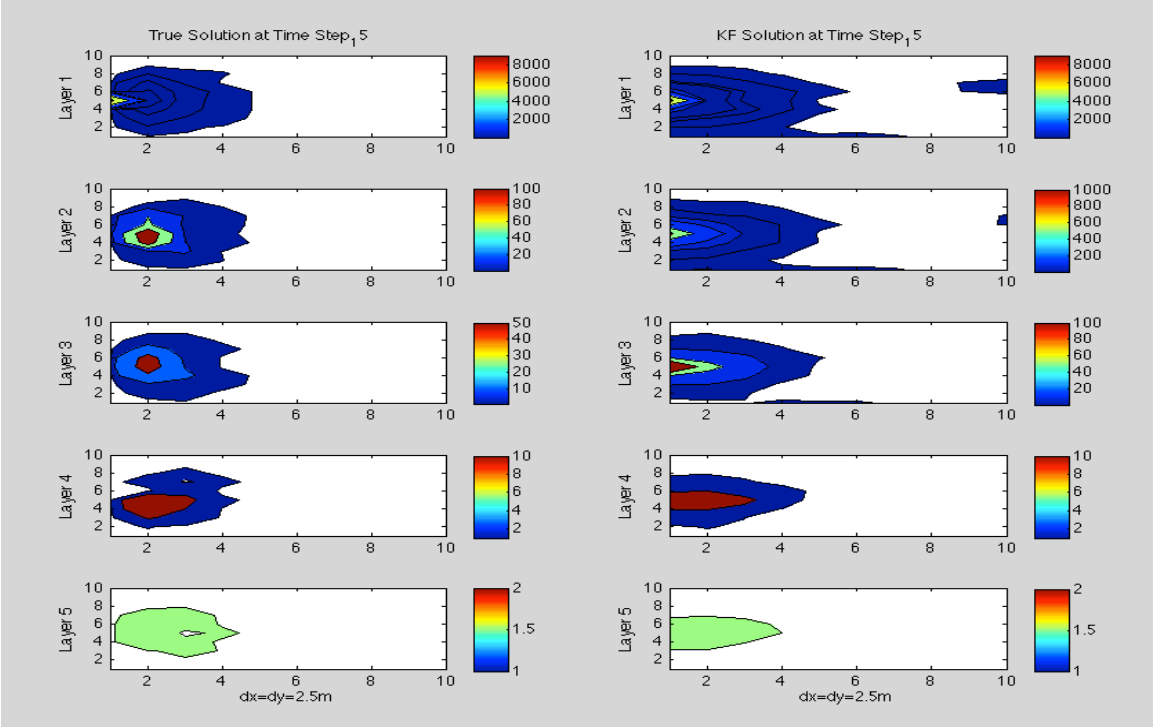

Figure 13: Contour profile for True Solution vs KF Solution at time step 15

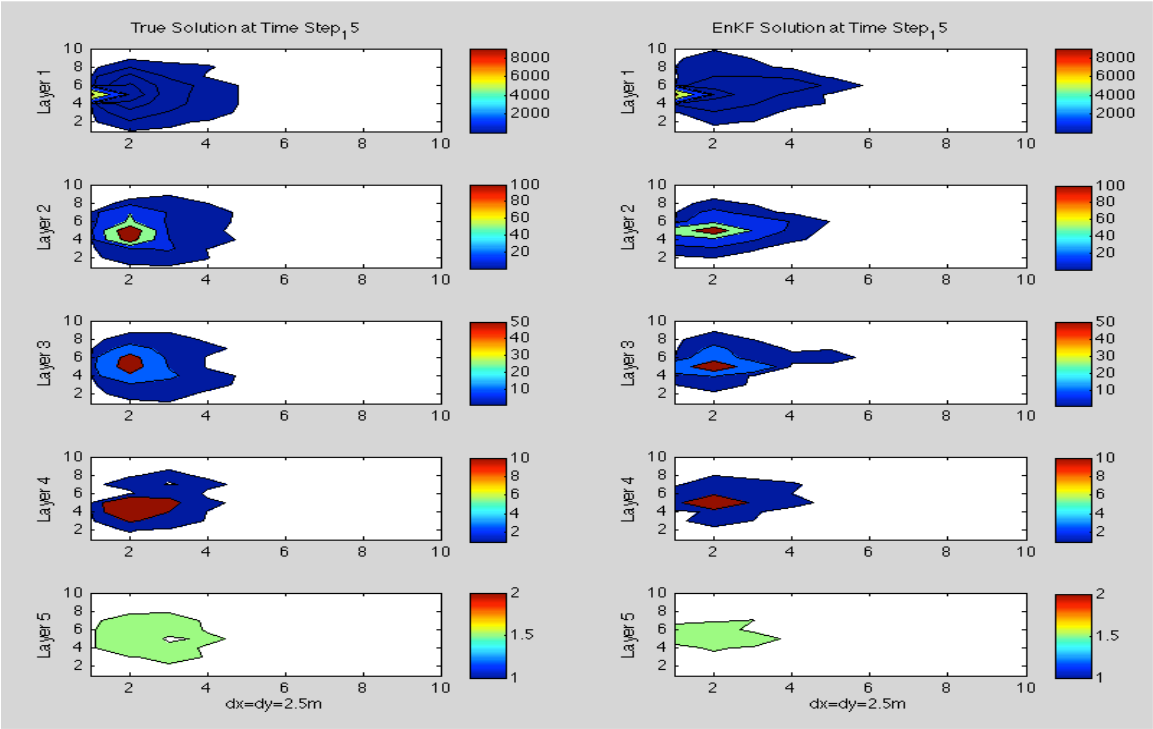

Figure 14: Contour profile for True Solution vs EnKF Solution at time step 15

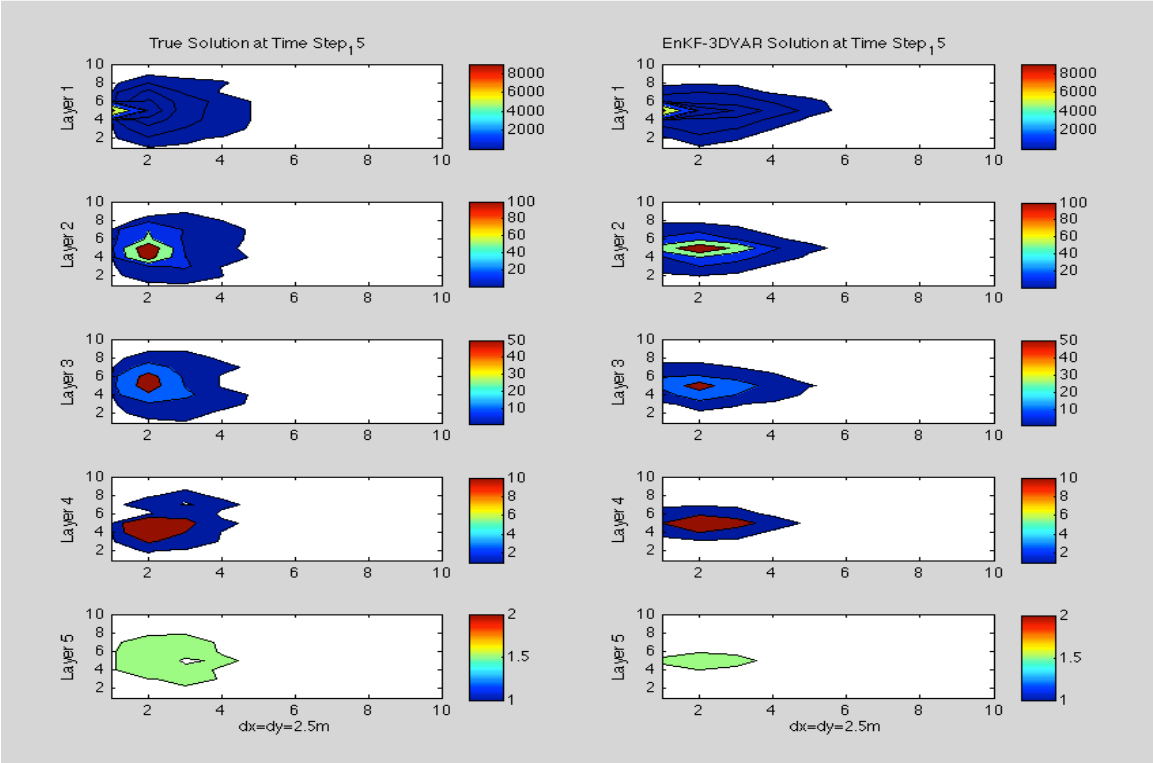

Figure 15: Contour profile for True Solution vs EnKF-3DVAR Solution at time step 15

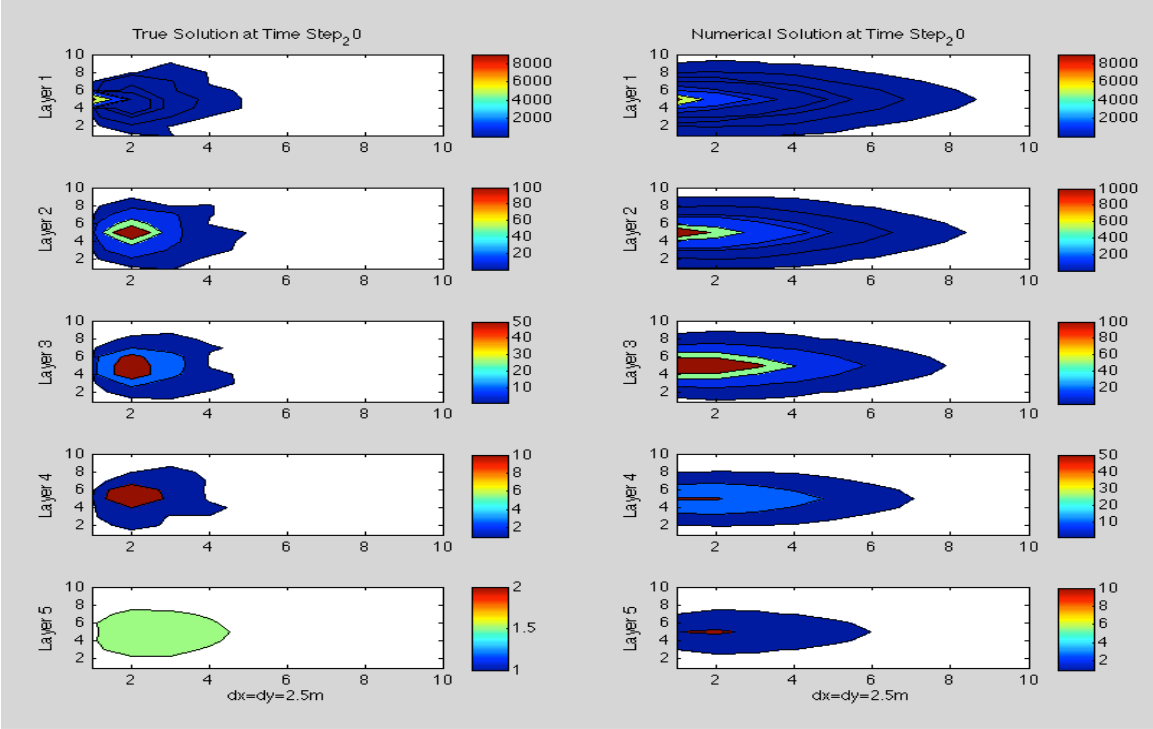

Figure 16: Contour profile for True Solution vs Numerical Solution at time step 20

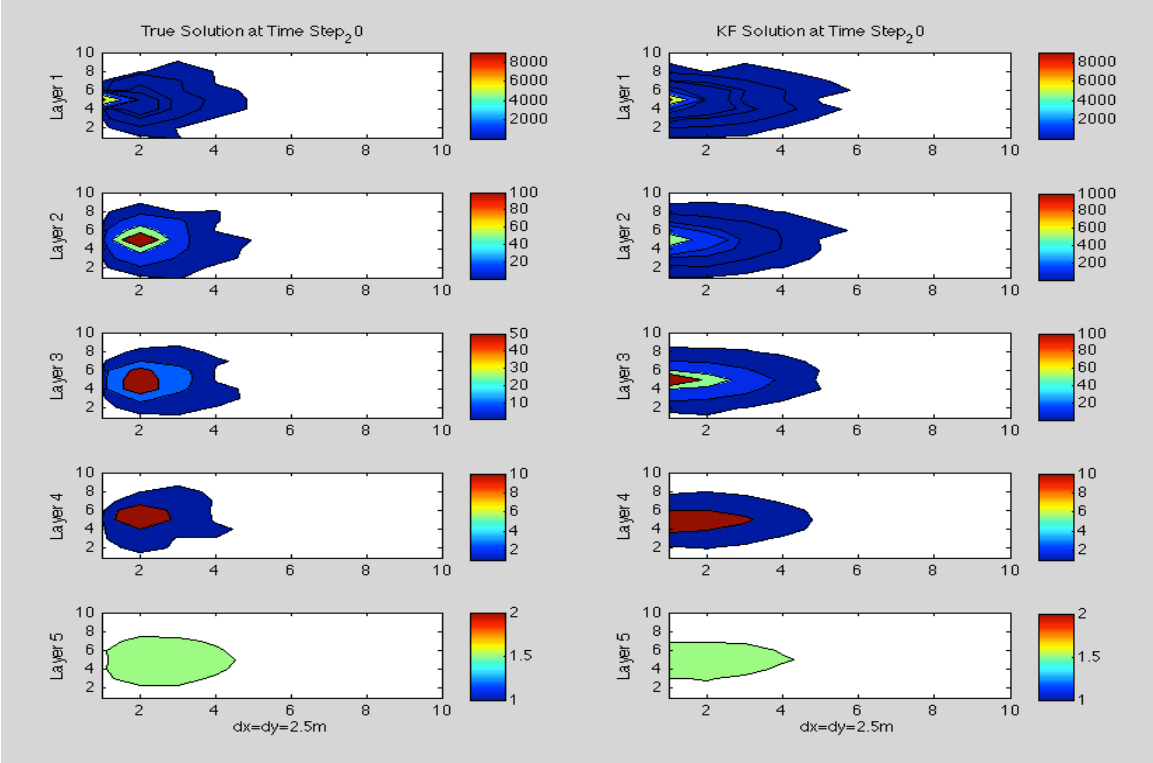

Figure 17: Contour profile for True Solution vs KF Solution at time step 20

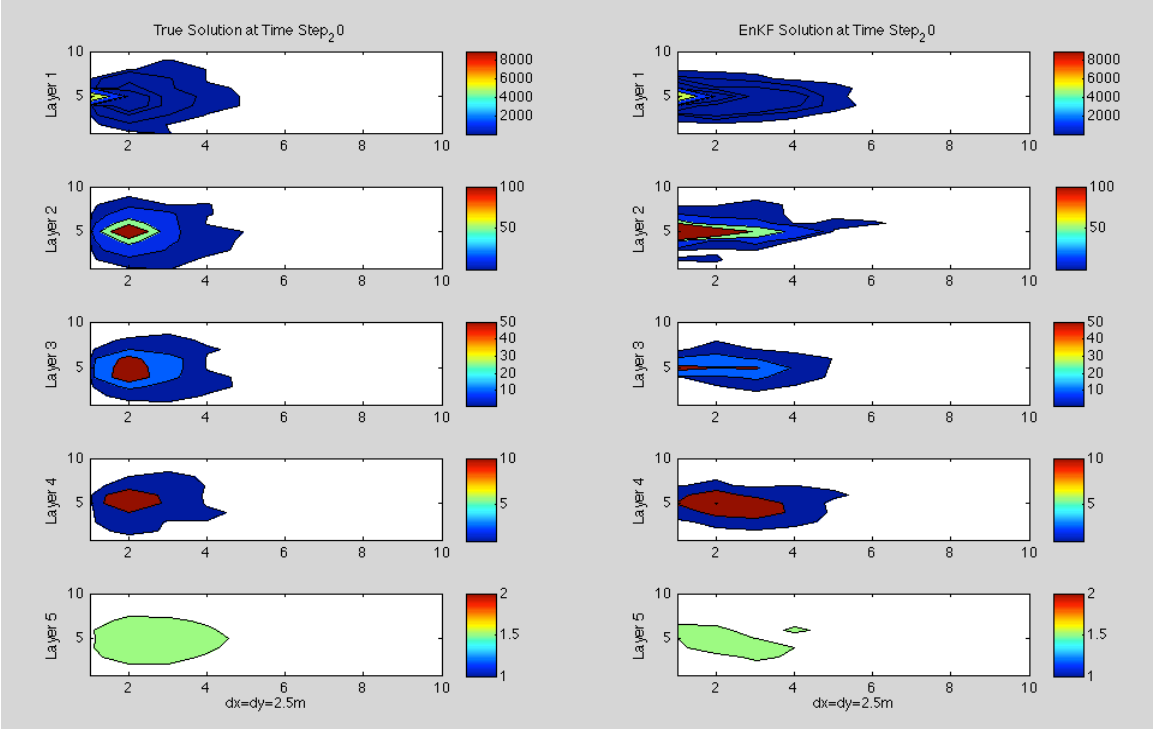

Figure 18: Contour profile for True Solution vs EnKF Solution at time step 20

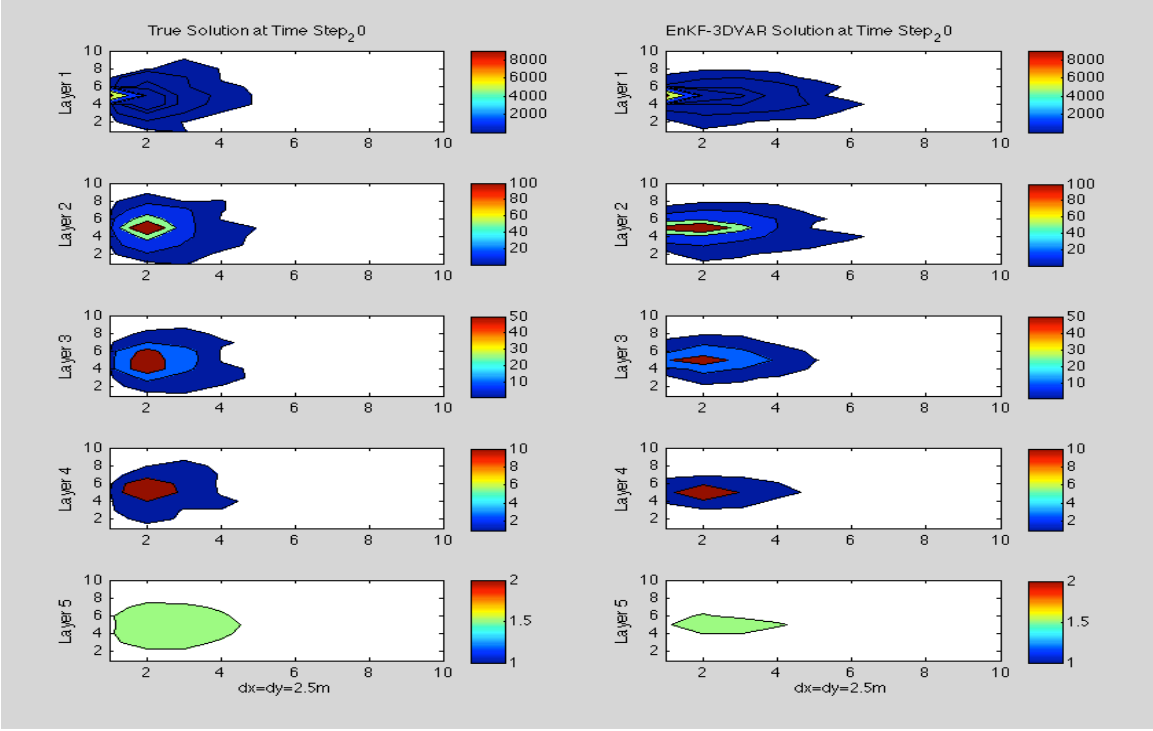

Figure 19: Contour profile for True Solution vs EnKF-3DVAR Solution at time step 20

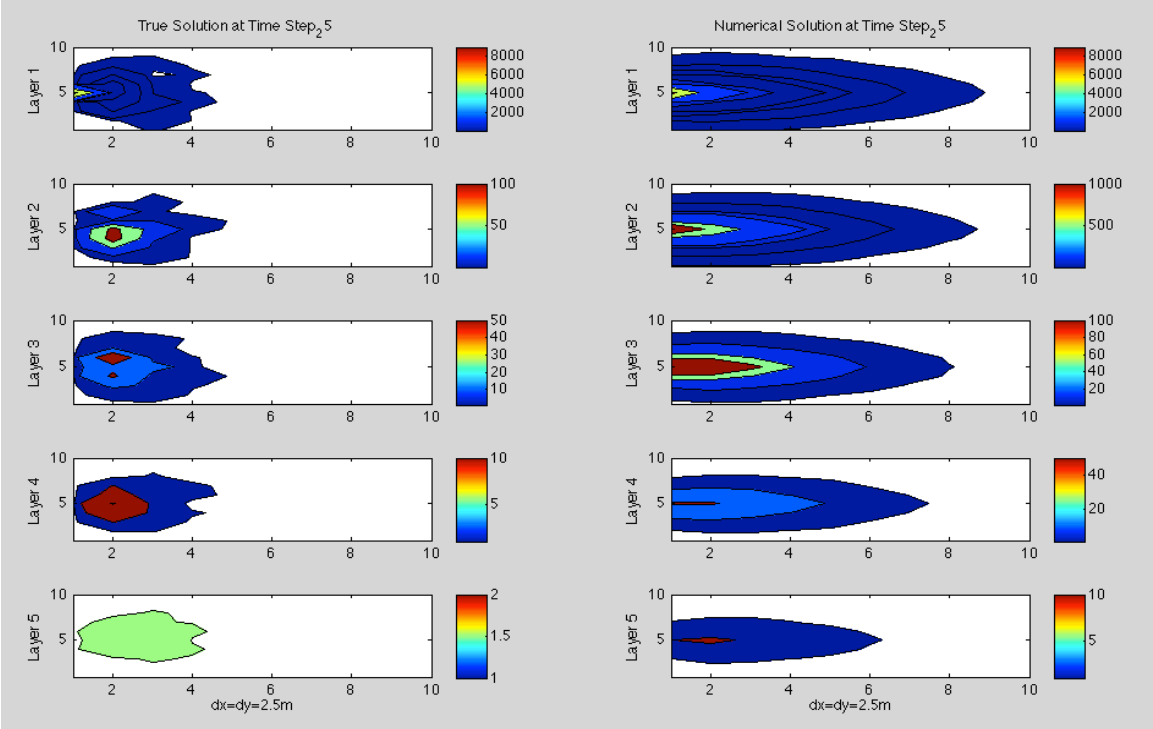

Figure 20: Contour profile for True Solution vs Numerical Solution at time step 25

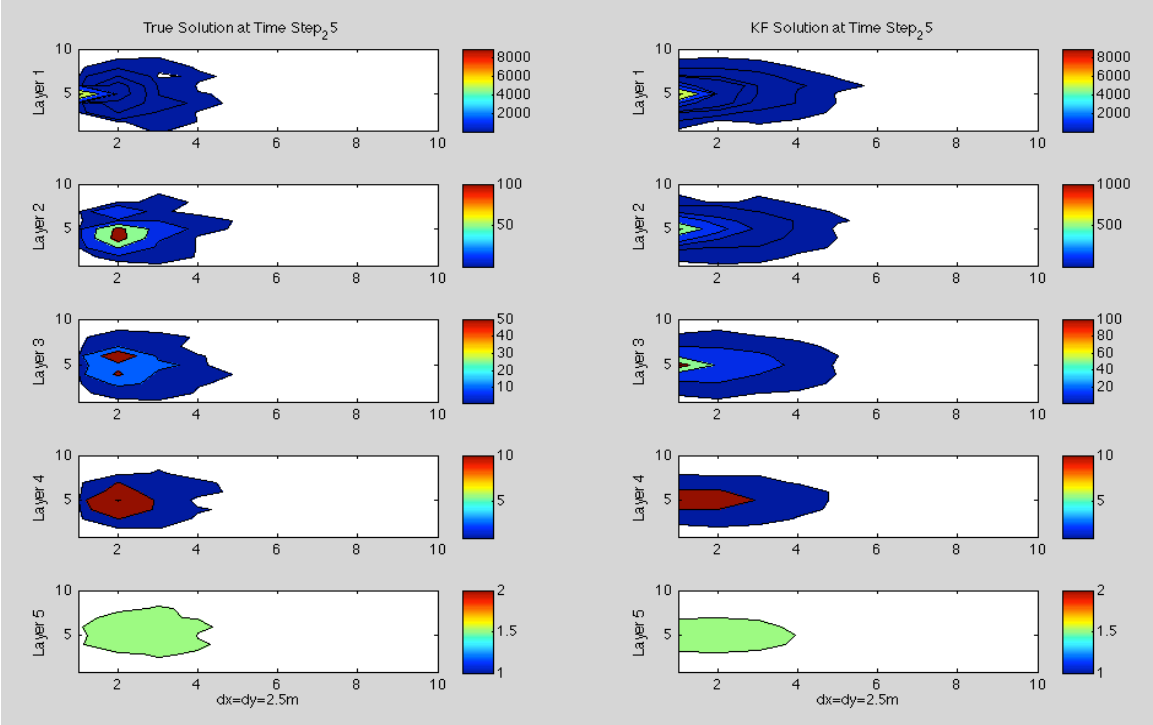

Figure 21: Contour profile for True Solution vs KF Solution at time step 25

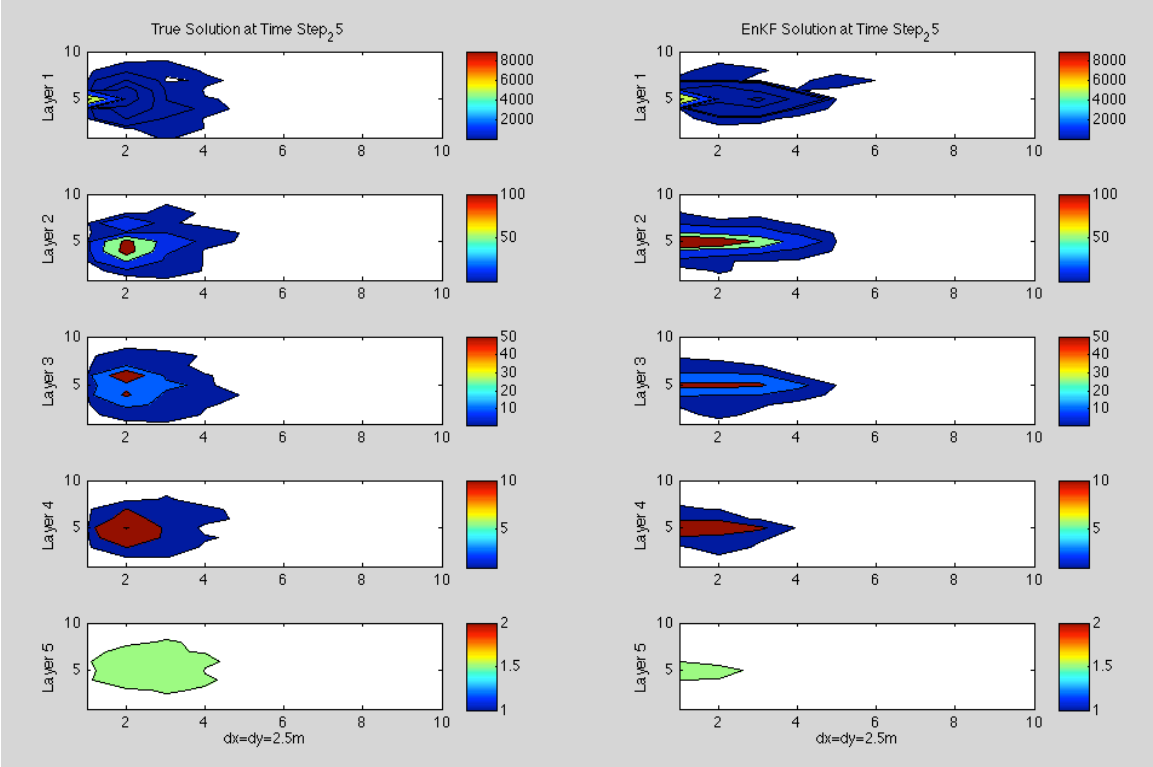

Figure 22: Contour profile for True Solution vs EnKF Solution at time step 25

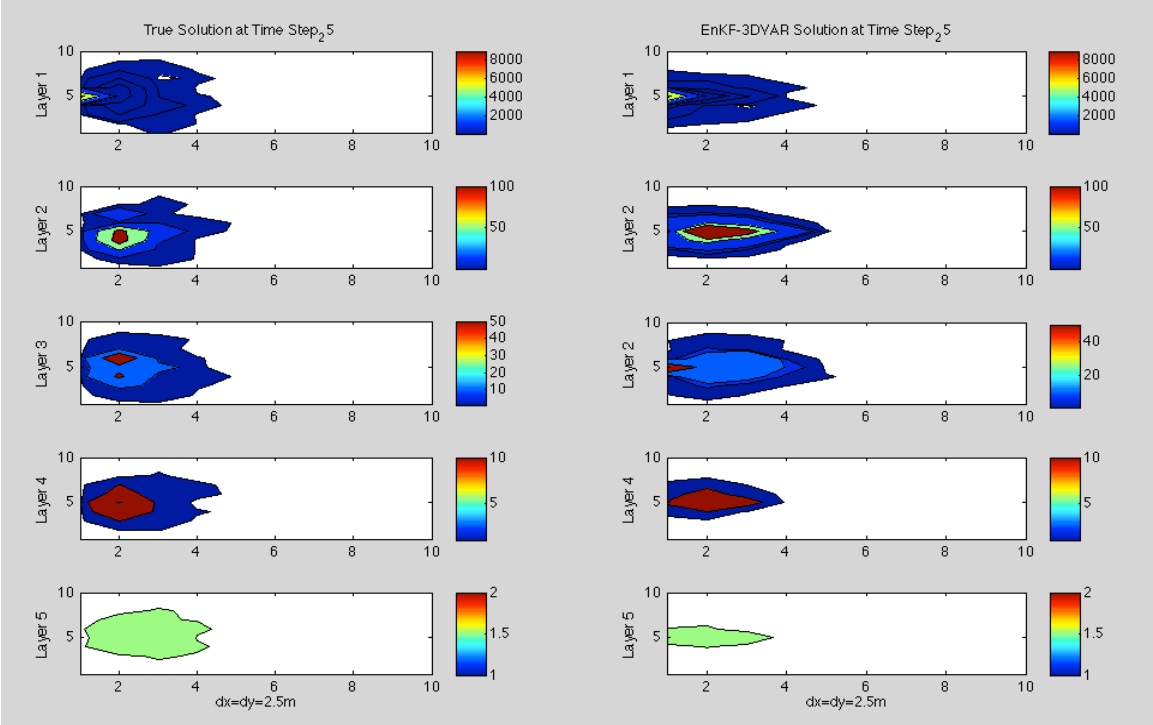

Figure 23: Contour profile for True Solution vs EnKF-3DVAR Solution at time step 25

In the intermediate time steps we can see that as the time progresses, the numerical solution moved more away than the true solution. It also gives a uniform concentration profile in comparison with the irregular shape of simulated true solution. The contour shapes are closer for KF solution in comparison with the numerical solution but the KF solution gives much higher concentration values compared to the true values at the same levels. The contour profiles of EnKF and EnKF-3DVAR solution are much closer to the simulated true solution and also the concentration values are closer to the true solution. The EnKF-3DVAR has a good improvement over the EnKF solution giving the contour shapes much better and also the concentration range is closer to true solution in comparison with the other models. This is because with the progress of time more observation data are being incorporated in these models and as the numerical solution cannot incorporate the observation data it fails to evaluate the real world condition and gives a regular pattern of the contour lines. Figures 24 to 27 show the model predictions of the final time step simulation.

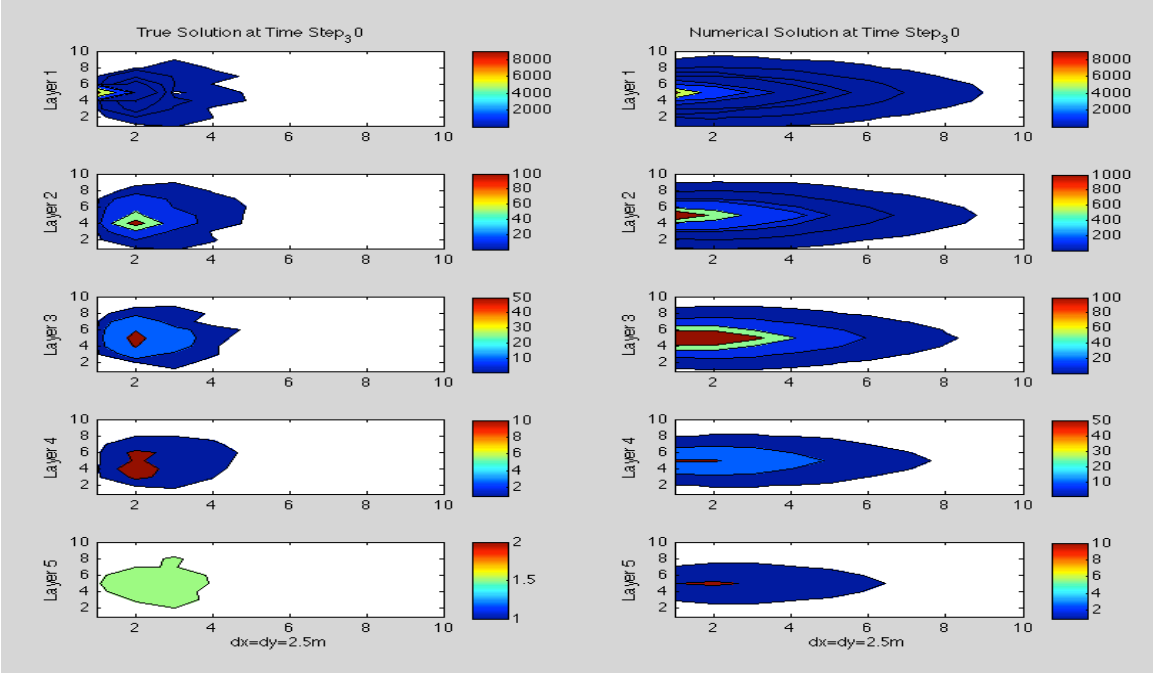

Figure 24: Contour profile for True Solution vs Numerical Solution at time step 30

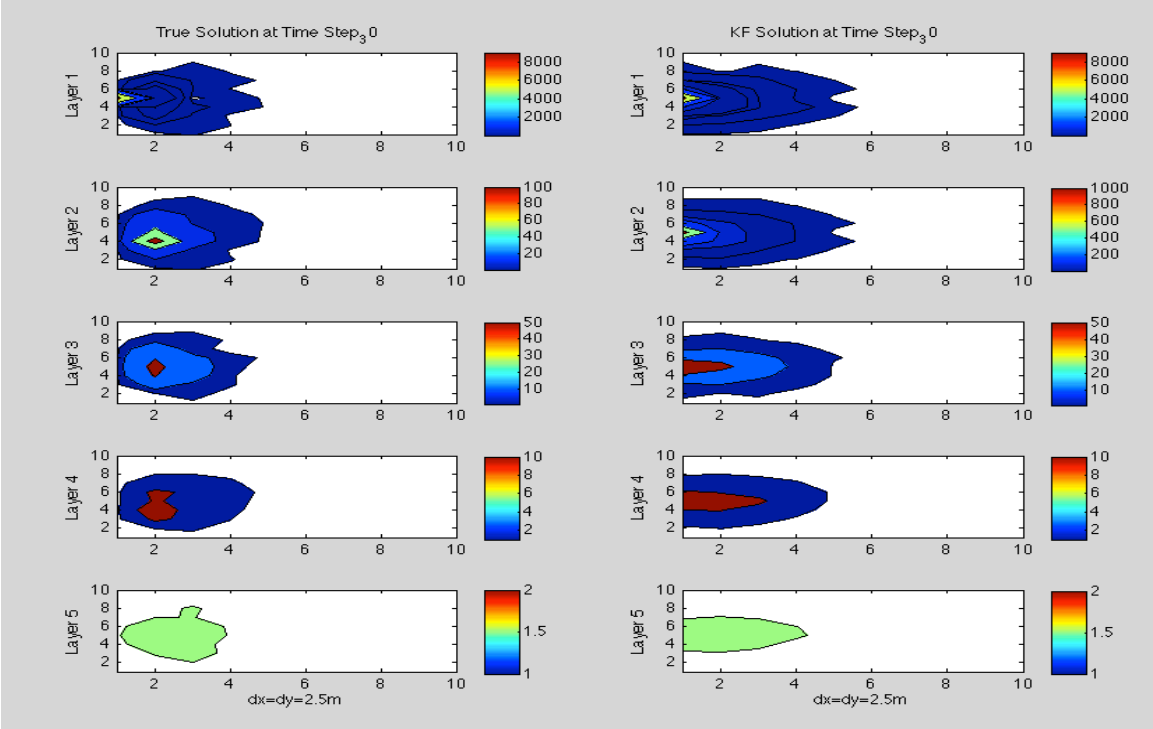

Figure 25: Contour profile for True Solution vs KF Solution at time step 30

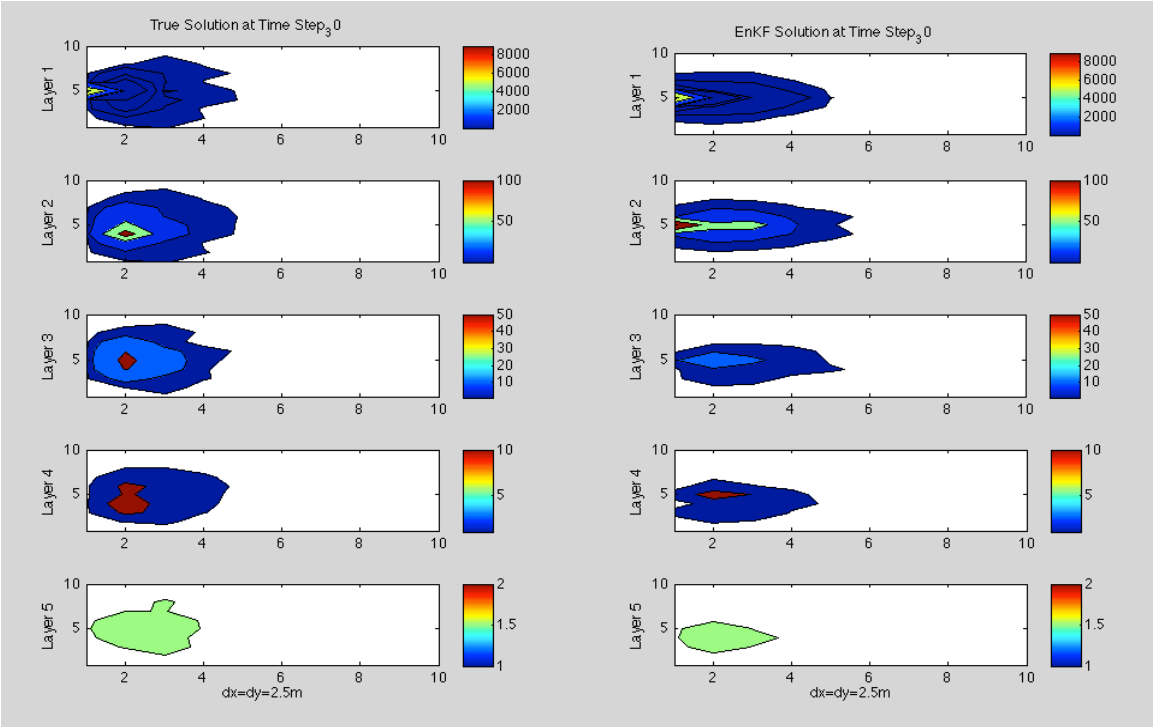

Figure 26: Contour profile for True Solution vs EnKF Solution at time step 30

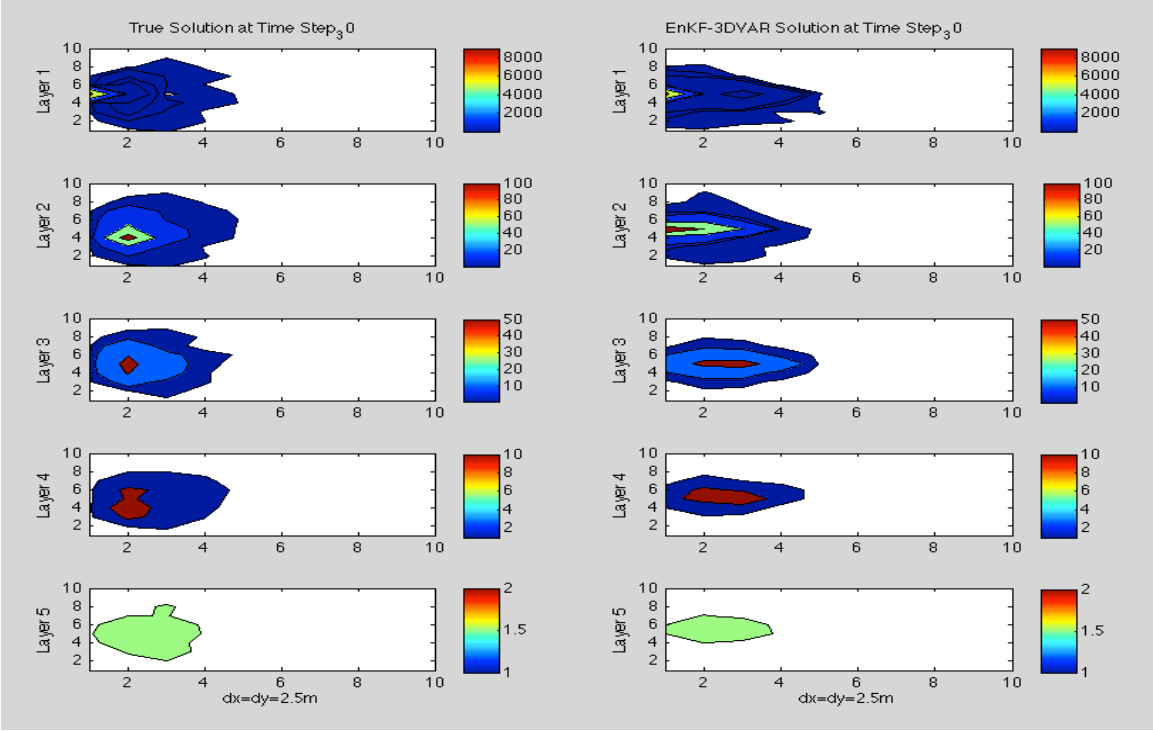

Figure 27: Contour profile for True Solution vs EnKF-3DVAR Solution at time step 30

In final time step we can see that at all five layers, EnKF-3DVAR evaluates the best contour profiles and have closer concentration values as the simulated true solution. The EnKF solution on the other hand also has good contour shapes although at layer 3 it gives some a little lower concentration values compared to the true solution. The KF solution is found to give concentration profile for lower layers (layer 3, 4 and5) that has low concentration ranges but it gives much higher predictions in top layers and also the contour profiles are more uniform. On the other hand the numerical approach gives high concentration values at all the 5 layers.

In this study we use sparse observation points. While abundant and large sampling points gives higher predictability of the model, sparse observation points are found to be economical and practical as in real life problems for subsurface systems there are very few numbers of observation wells are available and the installation of observation wells are expensive (Ferreyra et al., 2002). Graham and McLaughlin (1991) and Huang et al. (2009) used a small number of

observation points located in high concentration uncertainty zones and produced a good estimates. Unlike them and to make it more realistic we also use only 45 of 500 nodal points as observation points located sparsely (figure 28) at different concentration points and still our model produce good results as shown in the above figures.

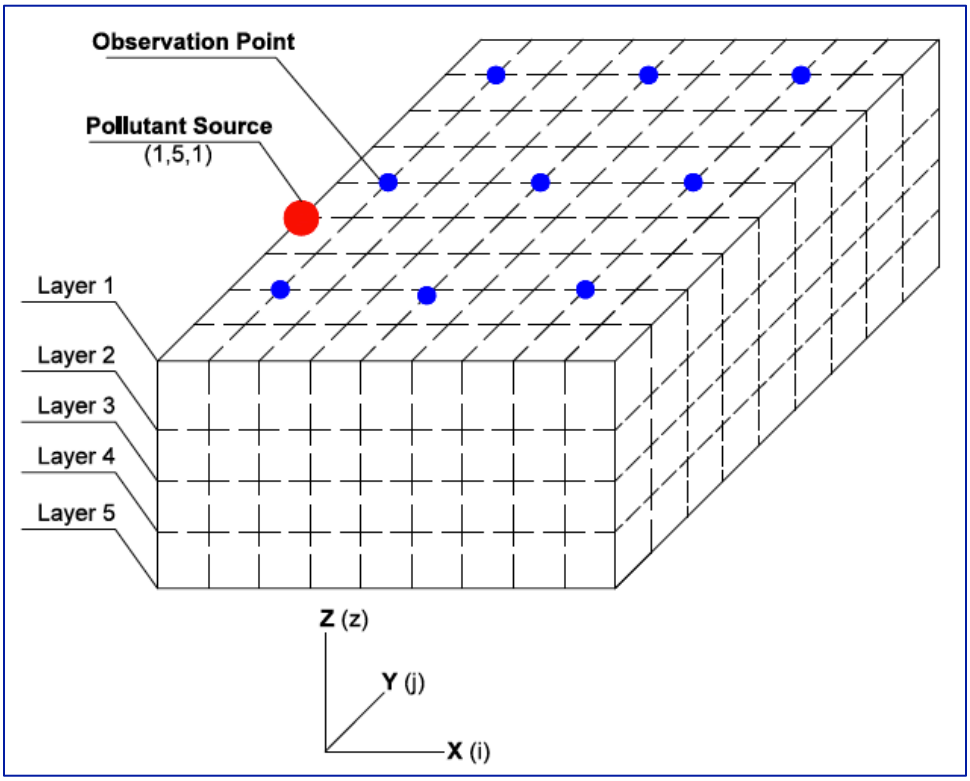

Figure 28: 3D domain space showing the observation locations and pollutant input point

To compare the accuracy of the results of different approaches, a RMSE profile is plotted. To evaluate the randomness of the groundwater and pollutant transport system the simulated true solution is generated in a way with addition of random noise such that it gives a new and different solution with each model run. Figure 29 shows the RMSE profile for all different models at all layer for all time steps. It is found that the error in the numerical approach increases with time, which is due to the fact that for a continuous pollutant source, more pollutants are introduced to the system at each time step and the numerical model fails to

incorporate the field data. But with the progress of simulation the KF, EnkF and EnKF-3DVAR solutions showed a significant reduction of error from the numerical model.

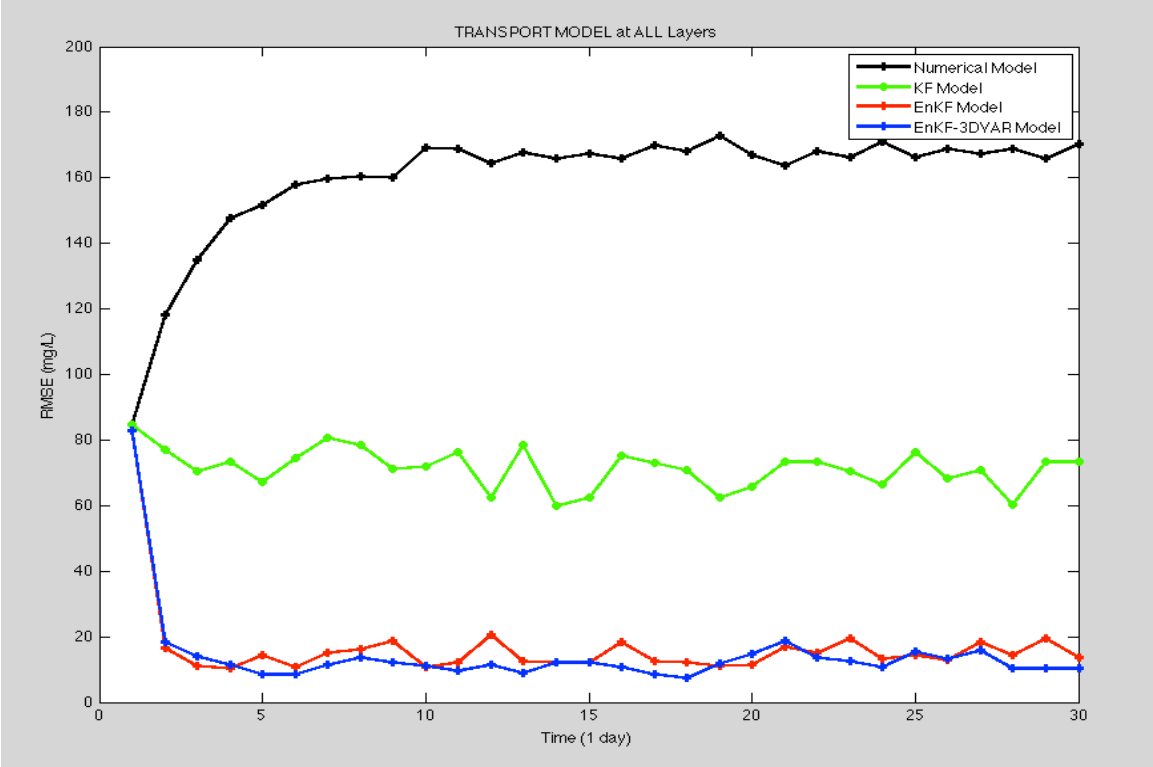

Figure 29: RMSE profile at all layers

The Kalman filter gives a lesser reduction in error in comparison with the EnKF and EnKF-3DVAR solution, which proves the robust nature of the EnKF. From figures 30-34 we can see that the errors are generally higher in the top layers (layer 1, layer 2 and layer 3) as top layers have the high concentration values compared to the lower layers (layer 4and5). At the lower layers all the models show better performance and have lesser errors compared to the top layers. The new approach EnKF-3DVAR shows a stable good performance in all layers and shows a decent improvement over the EnKF solution, which proves the capacity of 3DVAR analysis to optimize the background state prediction.

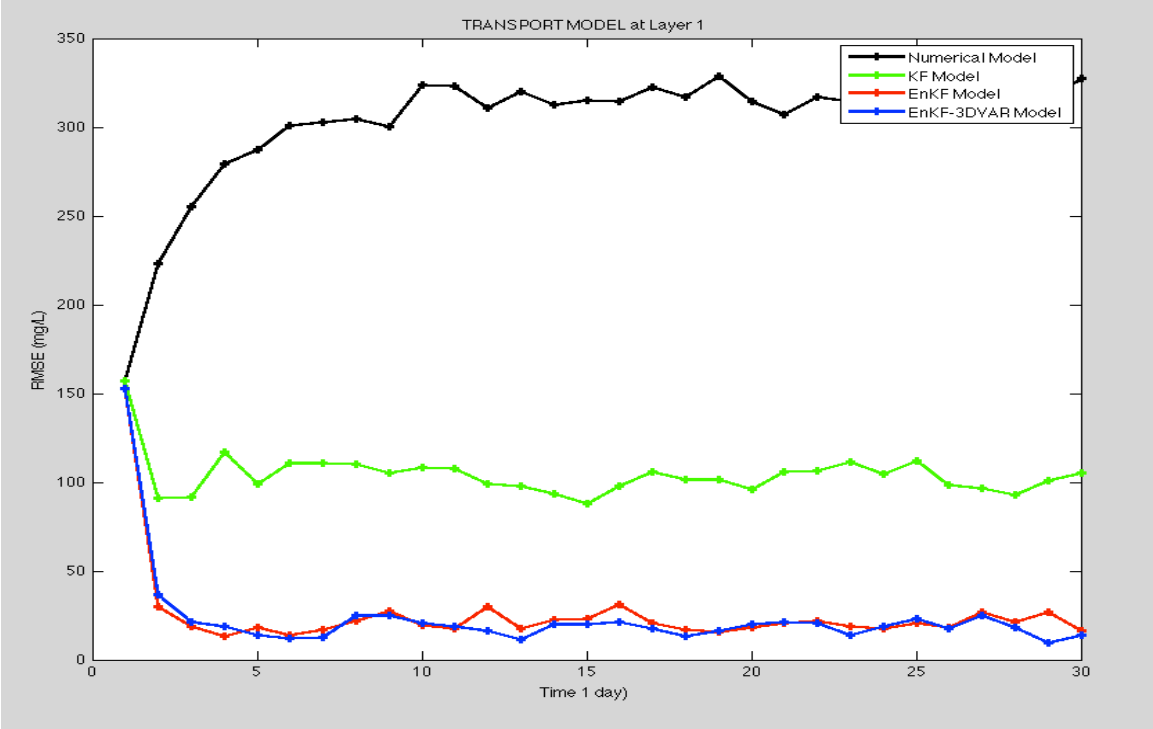

Figure 30: RMSE at layer 1

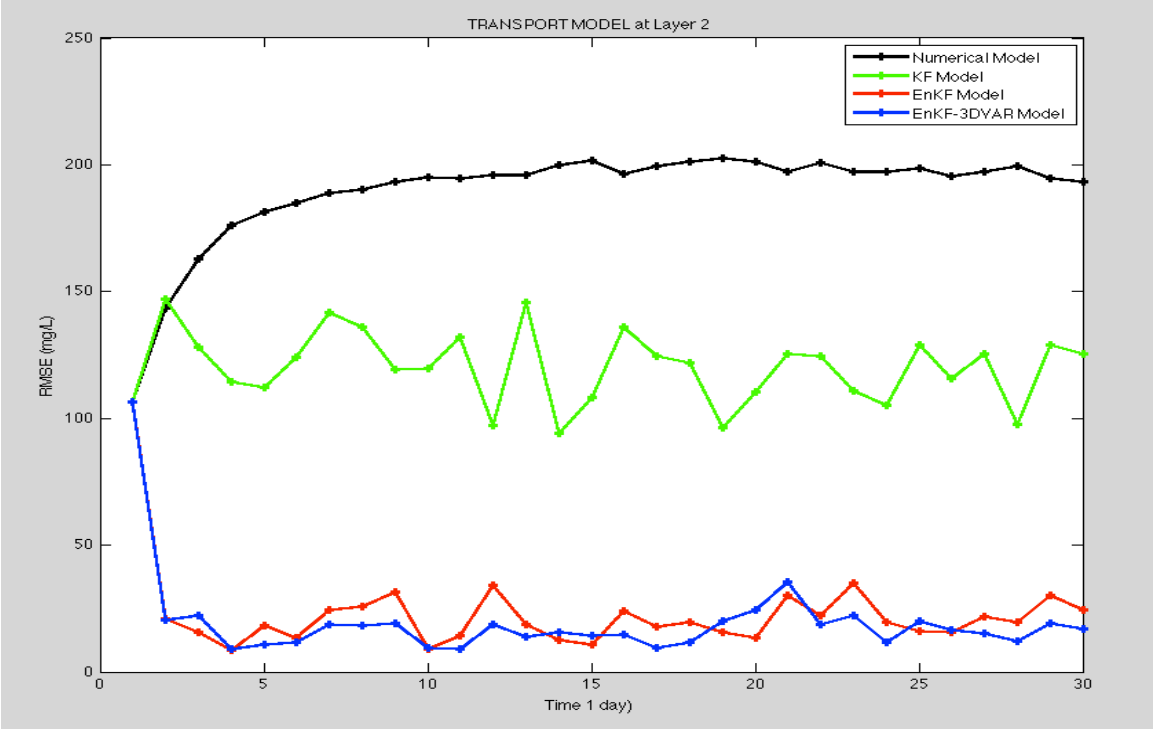

Figure 31: RMSE at layer 2

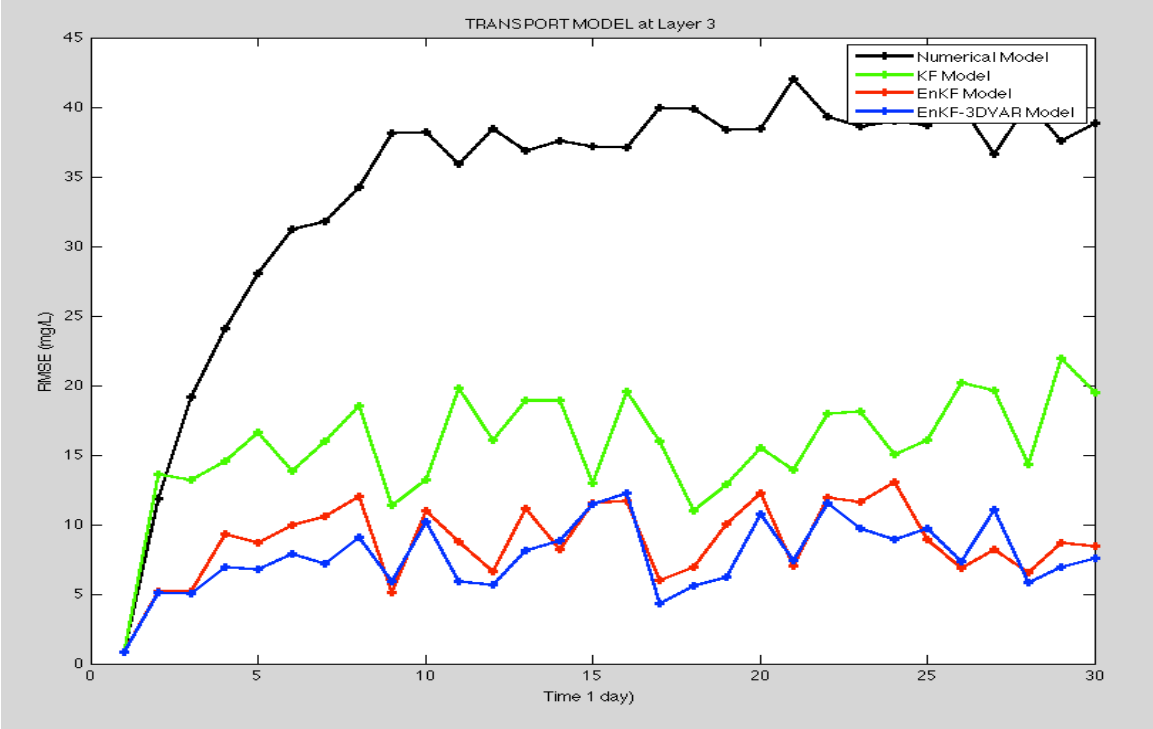

Figure 32: RMSE at layer 3

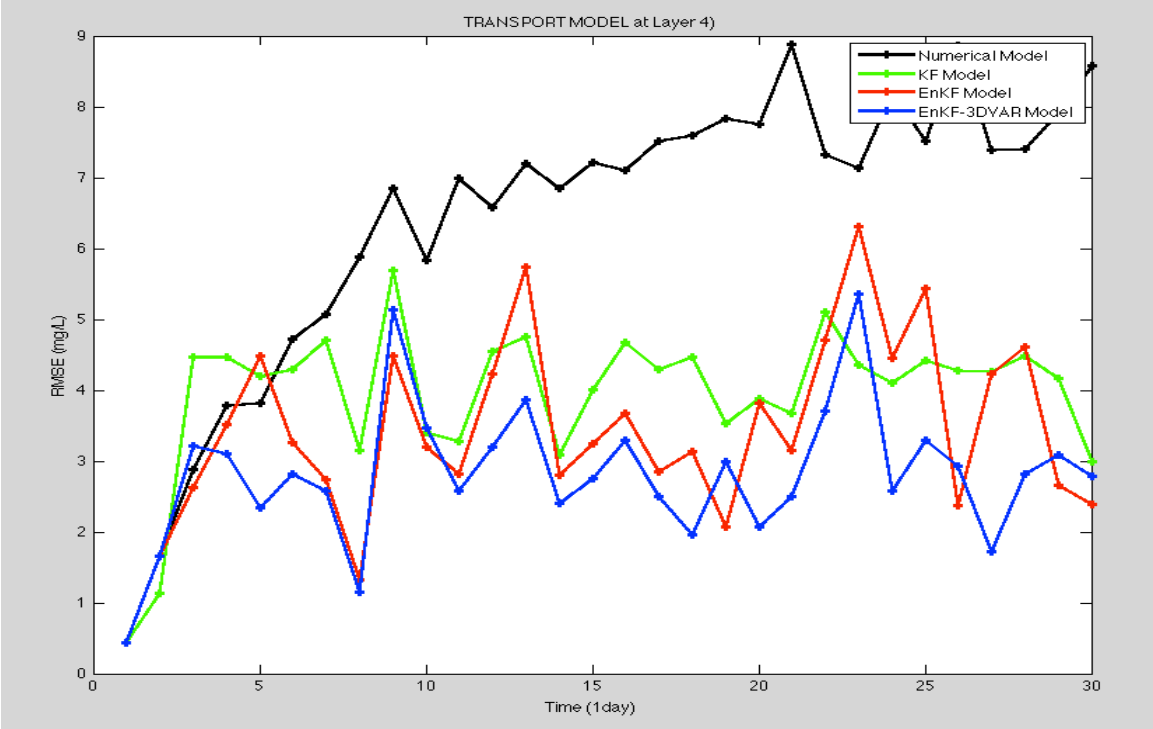

Figure 33: RMSE at layer 4

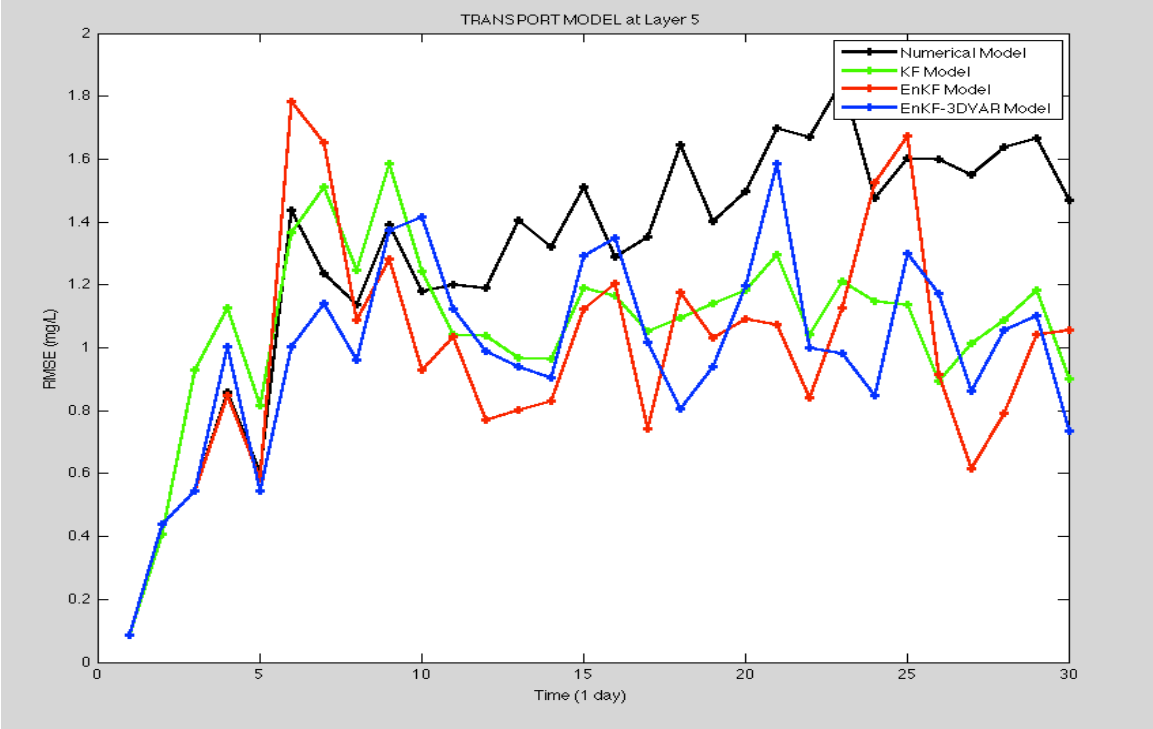

Figure 34: RMSE at layer 5

To check the performance and stability of the models, the simulation is run for five times. The shape of the RMSE profile is changed for every run and gives a total five profiles for each of the models, which indicates that in each simulation a random transport field is generated and a different true solution is obtained. Figure 35 below shows the performance stability of the models.

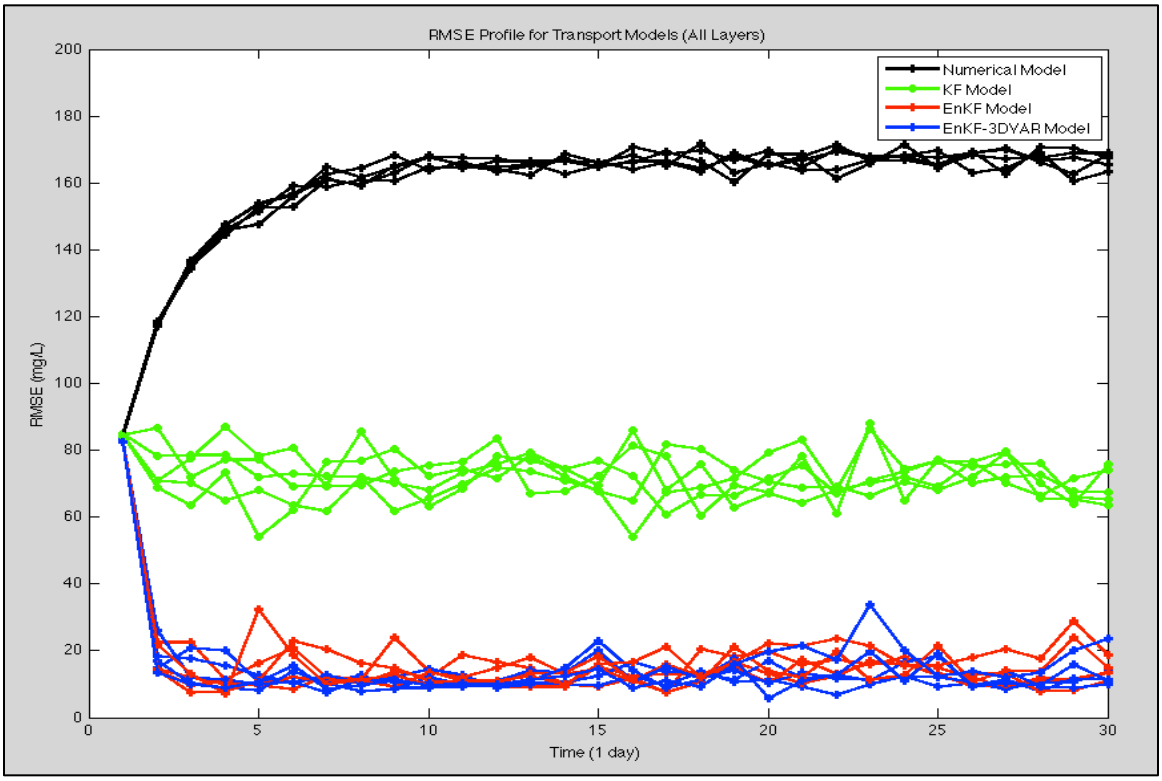

Figure 35: Stability of the models.

CHAPTER 5

Conclusion

The numerical model in the subsurface contaminant transport model can cause larger errors from the real life as it fails to incorporate the uncertainty of the nature. By using the data assimilation techniques with the numerical approaches to incorporate the information of the environment from the observation can significantly reduce the errors and improve the prediction of the model. From this study we find that Kalman Filter, Ensemble Kalman Filter and 3DVAR analysis have the capability to incorporate the observation data to update the results and to reduce the error. The EnkF-3DVAR improves the accuracy about 91.5% from the numerical model while the KF and the EnKF solution improve the accuracy of the results about 54.2% and 89% respectively than that of the numerical solution. The 3DVAR analysis scheme when integrated with EnKF, it improves the prediction accuracy of the model 22.3% from the EnKF solution. The application of 3DVAR analysis with the EnKF is found to be effective and it performs well compared to the EnKF in representation of the pollutant concentration profiles. The mean RMSE for the models at all layers are calculated as 159.1 mg/l for numerical model, 72.9 mg/l for KF model, 17.5 mg/l for EnKF and 13.6 mg/l for EnKF-3DVAR model.

References

Annan, J. D., and Hargreaves, J. C. (2004). Efficient parameter estimation for a highly chaotic system. *Tellus A, 56*(5), 520-526. doi: 10.1111/j.1600-0870.2004.00073.x.

Barker, D. M., Huang, W., Guo, Y. R., Bourgeois, A. J., and Xiao, Q. N. (2004). A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. *Monthly Weather Review, 132*(4), 897-914. doi: 10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2.

Bouttier, F., and Courtier, P. (2001). *Data assimilation concepts and methods March 1999*.

Chang, Shoou-Yuh, and Latif, Sikdar M.I. (2009). *Use of Kalman Filter and Particle Filter in a One Dimensional Leachate Transport Model.* Paper presented at the Proceedings of 2007 National Conference on Environmental Science and Technology, Greensboro, NC, USA.

Cheng, Xin. (2002). *Kalman filtering scheme for three dimensional subsurface transport simulation with a continuous input.* (MS thesis), North Carolina A&T State University, Greensboro, USA.

Domenico, P., and Schwartz, Franklin W. (1997). *Physical and Chemical Hydrogeology (2nd edition)*.

Evensen, Geir. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res., 99*(C5), 10143-10162. doi: 10.1029/94jc00572.

Evensen, Geir. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics, 53*(4), 343-367. doi: 10.1007/s10236-003-0036-9.

Evensen, Geir. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics, 54*(6), 539-560. doi: 10.1007/s10236-004-0099-2.

Evensen, Geir, and Leeuwen, Peter Jan Van. (1996). Assimilation of Geosat Altimeter Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model. *Monthly Weather Review, 124*(1), 85-96.

Ferreyra, R. A., Apezteguía, H. P., Sereno, R., and Jones, J. W. (2002). Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. *Geoderma, 110*(3-4), 265-289.

Gao, Wei, Zou, Lanjun, Gao, Wei, Wu, Tongwen, Xu, Xiaofeng, Du, Bingyu, . . . Ustin, Susan L. (2006). A 3DVAR land data assimilation scheme Part 1: Mathematical design. *6298*, 62982E-62982E-62988. doi: 10.1117/12.679994.

Geer, F. C. Van. (1982). An equation based theoretical approach to network design for groundwater levels using kalman filters. *International Association of Hydrological Sciences, 136*, 241-250.

Geer, Frans C. Van , Stroet, Chris B. M. Te , and Yangxiao, Zhou. (1991). Using Kalman Filtering to Improve and Quantify the Uncertainty of Numerical Groundwater Simulations: 1. The Role of System Noise and Its Calibration. *Water Resour. Res., 27*(8), 1987-1994. doi: 10.1029/91wr00509.

Graham, Wendy D., and McLaughlin, Dennis B. (1991). A stochastic model of solute transport in groundwater: Application to the Borden, Ontario, Tracer Test. *Water Resour. Res., 27*(6), 1345-1359. doi: 10.1029/91wr00260.

Hamill, Thomas M., and Snyder, Chris. (2000). A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Monthly Weather Review, 128*(8), 2905-2919. doi: doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Houtekamer, P. L., Mitchell, Herschel L., and Mitchell, L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review, 126*, 796-811.

Hsiao, Ling-Feng, Chen, Der-Song, Kuo, Ying-Hwa, Guo, Yong-Run, Yeh, Tien-Chiang, Hong, Jing-Shan, . . . Lee, Cheng-Shang. (2012). Application of WRF 3DVAR to Operational Typhoon Prediction in Taiwan: Impact of Outer Loop and Partial Cycling Approaches. *Weather and Forecasting, 27*(5), 1249-1263. doi: 10.1175/WAF-D-11-00131.1.

Huang, Chunlin, Hu, Bill, Li, Xin, and Ye, Ming. (2009). Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source. *Stochastic Environmental Research and Risk Assessment, 23*(8), 1155-1167. doi: 10.1007/s00477-008-0289-4.

Huang, Chunlin, Li, Xin, and Lu, Ling. (2008a). Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sensing of Environment, 112*(4), 1320-1336.

Huang, Chunlin, Li, Xin, Lu, Ling, and Gu, Juan. (2008b). Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. *Remote Sensing of Environment, 112*(3), 888-900.

Lisæter, Knut Arild, Rosanova, Julia, and Evensen, Geir. (2003). Assimilation of ice concentration in a coupled ice–ocean model, using the Ensemble Kalman filter. *Ocean Dynamics, 53*(4), 368-388. doi: 10.1007/s10236-003-0049-4.

Moradkhani, Hamid, Sorooshian, Soroosh, Gupta, Hoshin V., and Houser, Paul R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources, 28*(2), 135-147.

Pham, Dinh Tuan. (2001). Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. *Monthly Weather Review, 129*(5), 1194-1207. doi: 10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.

Pimentel, K. D., Candy, J. V., Azevedo, S. G., and Doerr, T. A. (1982). Simplified groundwater flow modeling: An application of Kalman filter based identification. *Mathematics and Computers in Simulation, 24*(2-3), 140-151.

Ruth, P., Emily, F., and John, Q. (2005). Groundwater Contamination in the United States (Third Edition ed., pp. 165-195). Philadelphia, PA University of Pennsylvania press.

Schwartz, Franklin W., and Zhang, Hubao. (2004). *Fundamentals of Ground Water*. NY, USA: John Wiley & Sons, Inc.

Shiklomanov, Igor (1993). World Fresh Water Resources. In P. H. Gleick (Ed.), *Water in Crisis: A Guide to the World's Fresh Water Resources.* . New York: Oxford University Press.

Survey, National Ground Water Association and United States Geological. (2005). National Ground Water Association and United States Geological Survey. Ground Water Statistics.

Thornhill, G. D., Mason, D. C., Dance, S. L., Lawless, A. S., Nichols, N. K., and Forbes, H. R. (2012). Integration of a 3D variational data assimilation scheme with a coastal area morphodynamic model of Morecambe Bay. *Coastal Engineering, 69*, 82-96. doi: 10.1016/j.coastaleng.2012.05.010.

Vrugt, Jasper A., Robinson, Bruce A., and Vesselinov, Velimir V. (2005). Improved inverse modeling for flow and transport in subsurface media: Combined parameter and state estimation. *Geophys. Res. Lett., 32*(18), L18408. doi: 10.1029/2005gl023940.

Walker, Jeffrey P., Willgoose, Garry R., and Kalma, Jetse D. (2001). One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. *Advances in Water Resources, 24*(6), 631-650.

Wang, Xuguang. (2011). Application of the WRF Hybrid ETKF–3DVAR Data Assimilation System for Hurricane Track Forecasts. *Weather and Forecasting, 26*(6), 868-884. doi: 10.1175/WAF-D-10-05058.1.

Wang, Xuguang, Barker, Dale M., Snyder, Chris, and Hamill, Thomas M. (2008a). A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Monthly Weather Review, 136*(12), 5116-5131. doi: 10.1175/2008mwr2444.1.

Wang, Xuguang, Barker, Dale M., Snyder, Chris, and Hamill, Thomas M. (2008b). A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Monthly Weather Review, 136*(12), 5132-5147. doi: 10.1175/2008mwr2445.1.

Welch, Greg, and Bishop, Gary. (1995). An Introduction to the Kalman Filter: University of North Carolina at Chapel Hill.

Wen, Xian-Huan, and Chen, Wen. H. (2005). *Real-Time Reservoir Model Updating Using Ensemble Kalman Filter*. Paper presented at the SPE Reservoir Simulation Symposium Houston, Texas USA.

Xiao, Qingnong, Kuo, Ying-Hwa, Sun, Juanzhen, Lee, Wen-Chau, Lim, Eunha, Guo, Yong-Run, and Barker, Dale M. (2005). Assimilation of Doppler Radar Observations with a Regional 3DVAR System: Impact of Doppler Velocities on Forecasts of a Heavy Rainfall Case. *Journal of Applied Meteorology, 44*(6), 768-788. doi: 10.1175/JAM2248.1.

Xiong, Chun-hui, Zhang, Li-feng, Guan, Ji-ping, Peng, Jun, and Zhang, Bin. (2013). Analysis and numerical study of a hybrid BGM-3DVAR data assimilation scheme using satellite radiance data for heavy rain forecasts. *Journal of Hydrodynamics, Ser. B, 25*(3), 430-439. doi: 10.1016/s1001-6058(11)60382-0.

Zou, S., and Parr, A. (1995). Optimal Estimation of Two-Dimensional Contaminant Transport. *Ground Water, 33*(2), 319-325. doi: 10.1111/j.1745-6584.1995.tb00287.x.