An Extended Kalman Filtering Technique for Geodetic First Order Data Assimilation

Ikechukwu Kalu (✉ kaluichechukwu360@gmail.com)

University of Uyo <https://orcid.org/0000-0003-1708-4354>

Christopher E. Ndehedehe

Griffith University Australian Rivers Institute

Onuwa Okwuashi

University of Uyo

Aniekan E. Eyoh

University of Uyo

Research Article

Keywords: Data assimilation, EKF, geodetic operations, set-valued state estimator, GPS

Posted Date: December 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1179932/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An Extended Kalman filtering technique for geodetic first order data assimilation

Ikechukwu Kalu^{a,*}, Christopher E. Ndehedehe^{b,c}, Onuwa Okwuashi^d, Aniekan E. Eyoh^e

^{a,d,e}Department of Geoinformatics & Surveying, University of Uyo, P.M.B. 1017, Uyo, Nigeria

^bSchool of Environment & Science, Griffith University, Nathan, Queensland 4111, Australia

^cAustralian Rivers Institute Griffith University, Nathan, Queensland 4111, Australia

*Corresponding author email:
kaluikewu360@gmail.com

Abstract

Data assimilation allows merging of different sources of data to estimate possible states of a system as it evolves in time. This therefore supports the idea of combining classical observations with Global Positioning System (GPS) observations to improve the integrity of first order geodetic controls in Nigeria. Given that these geodetic controls, which were established using traditional techniques and whose algorithms are still in use, the task of optimizing the coordinate values of these monuments to improve efficiency and accuracy in conventional geodetic operations around Nigeria is still a challenge. This study introduces the Extended Kalman Filter (EKF) technique for the modeling of these observations and their uncertainties in addition to exogenous noise, which is handled by an approximate *set-valued state estimator*. The proposed EKF provides a feasible linearization process in merging classical and GPS data collection modes as shown in our study. For each discrete time in the analysis step, it employs the Kalman gain computation, which attempts to weigh and balance uncertainties between the estimate and observation before proceeding to the analysis step. In this setup, the EKF constrains the system state in order to balance and strengthen the integrity of these first order monuments. The relationship of the derived system state with GPS coordinates ($R^2 = 0.85$) and classical observations ($R^2 = 0.92$) over Nigeria using a multi linear regression analysis is considerably strong. This outcome provides insight to the performance of the test algorithm and builds on the usefulness of data assimilation techniques in geodetic operations.

Data assimilation; EKF; geodetic operations; set-valued state estimator; GPS

1. Introduction

Accurate point positioning plays a vital role in the improvement of geodetic reference systems and cartographic products. The authenticity of coordinate conversion between these frames is, however, proliferated with the use of different techniques, including data optimization, multiple regression analysis, least squares adjustment, neural networks and even data assimilation. In the last few decades, data assimilation has increasingly been used to improve estimation of the state of a system by utilizing the unified duple algorithmic framework of numerical models and measurements (e.g., Slater and Clark, 2006; Kalra et al., 2019; Khaki et al., 2018; Simon, 2006; Gogoi et al., 2021; de Rubio, 2020; de Jesús Rubio et al., 2021; Chiang et al., 2019; Rubio et al., 2021; Vargas, 2021; Soriano et al., 2020). With the application of functional and stochastic models, the mathematical transformations of any two reference frames can be obtained. However, in a physical situation, erroneous specifications, residual variabilities, and outliers may exist in the functional models and pseudo-observations limiting the accuracy of the estimated parameters (e.g., Yang, 1999; Vrugt et al., 2013; Van Dijk et al., 2014).

Assimilating precise measurements into models is an operative methodology in overcoming these inadequacies (e.g., Van Dijk et al., 2014; McLaughlin, 2002; Gharmamti et al., 2016; Zaitchik et al., 2008; Khaki et al., 2017; Sarkar et al., 2019). Data assimilation is a technique where observational data of one or more variables (in relation to their uncertainties) are combined with output from a numerical (real world) model to produce an optimal estimate of the evolving state of the system (e.g.,

Hoteit et al., 2012; Bertino et al., 2003). For example, Simon (2006) suggested a linearized Kalman filter for estimating the state of a non-linear system (e.g. moving aircraft), using the Kalman filter estimate as the nominal state trajectory. The Extended Kalman Filter (EKF) comprises two successive Kalman filter- like updates. The first uses measurements to update the state vector while the second update enacts the balance constraint between uncertainty in measurement and estimate via the gain matrix. Data merging algorithms along with the EKF has been applied in recent studies (see, e.g., Koch et al., 2020; Tian et al., 2020) for a more stable composition and balance between measured and estimated data. Even though the enhanced datasets have consistently resulted in better estimates over different time varying signals and analysis, the subject of assimilating a perfect set of measurements despite the resulting robust constraint still poses a major setback in the system state (e.g., Khaki et al., 2018). For example, Simon and Chia (2002) argued that the assimilation of a perfect set of observations has a practical disadvantage in the resulting system state through the modification of its covariance matrix and neglecting the residuals of the system's unstable measurements (Tangdamrongsub et al., 2017).

With the development and advent of new mathematical algorithms, previously adopted parameters are continuously being optimized to enhance accuracy and precision in the science of geodesy. The two major classifications of these developed algorithms are the analytical and iterative procedures (Huaian et al., 2016). The analytical procedure requires complex mathematical derivations and formulas to provide the solution of the system state without need of initial estimates. Procedures such as the procrustes algorithm postulated by Graferand and Awange (2003) deals with isotropic weights by attempting to solve the unconstrained Lagrangian extremum problem using a singular value decomposition method. This approach attempts to achieve an optimum process for the computation of the system rotation matrix. Han (2010) computed transformation parameters using an analytical step-wise approach, while Shen et al. (2006) postulated a quaternion-based analytical procedure using eigenvalue-eigenvector decomposition. The use of the analytical-iterative procedure by Zeng and Yi (2010a), which employed the Rodrigues matrix exhibited poor iterative convergence due to initial value problems but produced an elaborate and efficient model in any case of small rotation angle values. Awange et al., (2008) pointed out the effectiveness of sole iterative algorithms requiring initial values in geodetic coordinate transformation packages, since they are highly efficient in correcting for distortion where the translation and rotation elements that have already been established (e.g., Matthes, 2002; Frohlich and Broker, 2003; Featherstone and Vanicek, 1999).

First order controls in Nigeria were established using classical methods based on the local reference frame, i.e., Minna datum-Clarke 1880 ellipsoid. This frame, however, consists of inherent distortions that impede the efficiency of the system state through (i) inaccuracy of the ellipsoids compressed scale factor (ii) erroneous origin determination (iii) defects in distance and angle measurements (iv) difficulty in determination of transformation parameters due to the degree of heterogeneity of the entire network. Unfortunately, these controls are still in use till today. The drive to relate geodetic reference frames stems from the introduction of the Global Navigation Satellite Systems (GNSS) which aims at providing autonomous geo-spatial positioning, as well as presenting a unified system of the entire earth for easier data manipulation. Given the contemporary advancement in this technology, it becomes pertinent to optimize the coordinate of these controls with an auxiliary frame that spans global coverage (GPS data). The progress of this optimization process will also help improve the accuracy of coordinate conversion between these two frames.

Past studies have shown that assimilation of different datasets produces improved results in terms of state estimates (e.g., Lievens et al., 2017; Tian et al., 2017; Han et al., 2015; Renzullo et al., 2014; Zhang et al., 2014). These studies merged data from different sources so that their errors can be used to attain optimal weights steering to the best estimates of the system matrix. To enhance the estimation of transformation parameters between reference frames, the error matrix derived from the functional model and coordinate difference between datums must be minimum (i.e., $\Sigma e = 0$). This can be achieved only by optimizing existing models using conventional algorithms, thus the intent behind this study. This study aims to combine the classically obtained data with the conventional GPS data

through data assimilation using the EKF algorithm. This procedure anticipates producing an optimal estimate through the derived gain matrix for the optimization of the locally obtained first order controls. To combine classical observations and GPS coordinates, the covariance matrix of the system state is approximated and parsed through a time-based iterative loop.

This study therefore attempts to explore the techniques of the EKF in achieving this goal. The EKF has been widely applied in hydrological forecasts (Muluye, 2011; Hamid et al., 2005) object tracking (e.g., Yang and Baum, 2017; Santosh and Mohan, 2014), data science (e.g., David et al., 2017; Wenzel et al., 2006), earth science applications (Soroush et al., 2011), machine learning (e.g., Bishop, 2013; Belhajem, 2016; Zibar et al., 2016) and predictive analysis (Liu et al., 2006). Its application in practical engineering problems is also well-known (Bertsekas & Rhodes, 1971; Petersen & Savkin, 1999; James & Petersen, 1998; Kallapur et al., 2009), especially with the introduction of the set membership state estimation approach by Bertsekas and Rhodes (1971). An extension of this model was undertaken by Petersen and Savkin (1999) in a bid to accommodate linear uncertain continuous systems. This also formed the basis of its application in discrete-time cases with sum quadratic constraint uncertainty descriptions (Savkin & Petersen, 1998), and the solving of robust filtering problems for a class of uncertain non-linear systems (James & Petersen, 1998). This study on the other hand attempts to build on the novel discrete-time robust extended Kalman filter model developed by Kallapur et al. (2009). Its distinction in modeling uncertainties in addition to exogenous noise makes it a suitable algorithm for our proposed integration of classically-derived and GPS observations in an attempt to build a stronger framework for optimal datum interactions (i.e. between local and global frames). This therefore makes the case for the novelty of our study as the implementation of this robust filter has not been applied in the optimization of any global-local reference frames for geodetic operations in any of the numerous African datums. The EKF used in this study also satisfies the discrete time modalities, including dynamics and measurements while taking the uncertainties in the dataset into account. EKF adopts a high computational power to integrate system dynamics by approximating $h(\mathbf{x}_k, \mathbf{v}_k)$ through a Taylor's series expansion around the *a priori* estimate ($\hat{\mathbf{x}}_k^-$) value. With a promise of faster convergence in terms of iterations and good approximations for non-linear models as demonstrated by these studies (Kallapur et al., 2009; Rhudy, 2015; Alonge et al., 2014; Dakhlallah et al., 2008; Oisiović & Cruz, 2000), the EKF is implemented by merging these two sets of data to achieve a robust system matrix.

To achieve homogeneity in the datasets before assimilation, the coordinates obtained from the GPS were oriented to the Clarke 1880 reference frame using the embedded systems' Gauss-Newton method based on the least squares adjustment and the similarity transforms. The conventional least squares adjustment has been the prevalent mode of datum homogeneity over the years, and has provided good fits in datum relations due to its rigorously based theory on mathematical probability. This therefore promoted its extensive application due to its (i) rigorous adjustment for finer solutions (ii) ease of applicability (iii) enabling and making rigorous post-adjustment analysis (e.g., Charles, 2010). The motive for adopting the EKF algorithm for this analysis is due to its use of both observations and residuals. Also, its assumption that all observations are error-prone helps in achieving the significant accuracy needed for first order geodetic controls. The strength of the system matrix however, is determined using the gain matrix, which attempts to weigh the uncertainties in the systems state and tune the model to its optimal mode.

2.Data

2.1 Classical observations

The local geodetic reference frame for Nigeria is the Clarke 1880 ellipsoid whose origin is inconsistent with the Earth's center of mass. The origin is situated at the center of the primary trilateration and triangulation network in Minna, Nigeria. It is otherwise referred to as the Minna datum. Due to the obsolete mode of data establishment, this system suffers from several inherent deficiencies contributing to significant distortion in the system state. However, since these controls are still in use, we must find a way to optimize their values in order to improve the accuracy of

conventional geodetic operations in Nigeria. First-order controls of geodetic monuments spread across the country used in this study satisfy the specification for relation between frames. One hundred and fifty-seven (157) first order points were obtained from the office of the surveyor general of the federation (OSGOF). These datasets constitute national coordinates of existing reference network in the country (e.g., Wonnacot, 2007; Poku-Gyamfi and Schueler, 2008). The objective for the national reference network was to enhance the use of GNSS for land-related positioning activities such as survey and mapping operations, timing, engineering projects, meteorology, traffic and transportation, and geotechnical investigations.

2.2 GPS coordinates

The global geospatial infrastructure which is currently referenced to the World Geodetic System (WGS84) was developed in 1984 by the U.S. Defense Mapping Agency (DMA). This infrastructure is currently called the National Geospatial-Intelligence Agency (NGA) and is used to access reference coordinate systems through a precise definition of their origin and orientation at a certain epoch. These frames are three-dimensional representing coordinates in geocentric and topocentric modes, respectively. The ellipsoids oriented and adjusted to the global (WGS84) dimensions serve as apposite references to determining earth's topographic position. The GPS observation will yield coordinates of the survey stations in the World Geodetic System WGS-84, which is a geocentric datum. Therefore, for this study, we obtained the corresponding derived GPS datasets of the already established 157 first order controls using the static differential GPS approach (Fig. 1). Given that the controls are primary geodetic controls, the observations taken here lasted between 2 to 4 hours for each of the controls with the adequate observations procedures for base line of the order of 20-40km. The instrument was however set to capture data in topocentric mode. To achieve homogeneity in the system state, the instrument was referenced to the Clarke 1880 ellipsoid. The GPS in-built algorithm makes use of similarity transforms and the Gauss-Newton method based on the least squares adjustment to obtain coordinate values for topocentric frames. However, the algorithm for this is not discussed in this study. The set of data obtained here augments the classically transformed dataset for the assimilation process.

Fig. 1. Spread of first order controls whose coordinates were obtained using classical and GPS procedures

2.3 The Clarke 1880 and WGS 1984 ellipsoid

The Clarke 1880 is the local Nigerian geodetic reference ellipsoid having its center and origin at a position not coincident with the earth's center of mass. The position of the station at Minna, Nigeria is

the center of the primary triangulation network of the region. The origin of this framework is adopted given that it is a local datum with the origin of the coordinates adopted. As stated earlier, crude and classical means were employed in the establishment of this control which has caused a significant inherent deficiency causing serious distortion in the region's mathematical framework. These include, among others, inaccurate definition of the origin of the regions framework, absence of a geoidal height model, defect in the measured distances as a result of the erroneous nature of the scale factor by compression of the Clarke 1880 ellipsoid, as well as difficulty in determining transformation parameters. However, the Federal Government of Nigeria through the Office of the Surveyor-General of the Federation (OSGOF) has launched the Continuously Operating Reference Station (CORS) in an attempt to join other regions in the conventional mapping techniques and discarding the traditional geodetic passive networks which are found to be basic in any region. The coordination of the CORS station to the International Terrestrial Reference Frame (ITRF) is very crucial especially in the realization of the region's geocentric datum (Kalu et al., 2021b).

The original intent of designing the World Geodetic System (1984) was to support the United States Department of Defense as well as its armed forces. Its design was based on the Doppler observation procedures, and this framework has been the nominal datum employed by the GPS in positioning. This system has undergone significant refinements over the years, especially in a bid to align it with the ITRF so as to capture the precise variations and dynamics of the earth. Regardless, it is not a perfect fit for Nigeria, which has warranted the development of a local reference frame for more accurate positioning in the Nigerian region (Kalu et al., 2021a)

3. Methods

3.1 Observation model and assimilation

The relationship between the system state and obtained measurement (e.g., GPS) is given by the observational model. In the linear case, the measurements can be described by a system of linear equations, which depend on the state variables. Normally, the observations are made up of discrete time steps k_i . It is represented in its vector form as,

$$\mathbf{l}_k = H \cdot \mathbf{x}_k + \mathbf{w}_k \quad (1)$$

Where, $\mathbf{l}_k$ is the vector of the observations at the epoch k , H is the observation matrix (e.g., GPS derived data), and $\mathbf{w}_k$ represents the noise of the measurement process. The ultimate aim in assimilating the GPS and classical data is to obtain minimum variance between the two datasets. Its optimality is determined by the system's extent in enforcing the position balance constraints, based on the prior observations and gain matrix.

3.1.1 State initialization/Procedure

Our discrete-time system model is represented as;

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \quad (2)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k) \quad (3)$$

$$\mathbf{w}_k \sim (0, \mathbf{Q}_k) \quad (4)$$

$$\mathbf{v}_k \sim (0, \mathbf{R}_k) \quad (5)$$

Where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{y}_k \in \mathbb{R}^{n_y}$ represents the system state and observation at time k with sizes n_x and n_y , respectively. $\mathbf{w}_k \sim (0, \mathbf{Q}_k)$ and $\mathbf{v}_k \sim (0, \mathbf{R}_k)$ implies that $\mathbf{w}_k$ is a continuous-time initial state process noise of Gaussian random variables with error covariance of $\mathbf{Q}_k$ while $\mathbf{v}_k$ is a discrete-time measurement white noise with $\mathbf{E}[\mathbf{v}_k \mathbf{v}_{k1}^T] = \mathbf{R}_k \delta_{kl}$; with δ_{kl} being the kronecker delta. $\mathbf{f}_{k-1}$ is a nonlinear operator, integrating the system state from time $k-1$ to k , and $\mathbf{h}_k$ is a non-linear design

operator at time k . The known noise signals for $\mathbf{u}_k$ is given by $\mathbf{u}_k = \mathbf{f}_k(\mathbf{x}_k, 0) - \mathbf{F}_k \mathbf{x}_k$. The model time step k is regarded to be equal to the assimilation time step.

The Extended Kalman filter algorithm works well in practice for moderate non-linearities. But oftentimes, it produces large uncertainties which lead to an increased error approximation.

3.2 The Extended Kalman filter (EKF)

Kalman filters have a record of diverging in several cases, especially under the influence of uncertainties and non-linearities. To overcome these problems, several forms of KF have been developed to address the limitations of the earlier algorithm, especially with regards to non-linearities. Among the several algorithms developed for state estimation, the set membership state estimation methodology provides a deterministic elucidation of the KF with regards to a set-valued state estimator. The set membership state estimation problem entails, the individual location of all states coherent with given resultant measurements for a time varying system, with a standard bounded noise input. Several attempts have been made to extend the set membership state estimation approach to acclimatize linear uncertain continuous systems with an adjoining integral quadratic constraint (IQC, KonradReif et al., 1999). An extension of the study results was made to the discrete time case having a sum quadratic constraint (SQC) uncertainty depiction and finally extended in solving a robust filtering problem, which in turn leads to an interesting version of the EKF.

The uncertainty connected with the model (Eqns. 2-5) is described in terms of an IQC as;

$$(x(0) - x_0)^T N(x(0) - x_0) + \int_0^S [w_k^T \Pi_k w_k + v_k^T \Lambda_k v_k] \leq d + \int_0^S \|z_k\|^2 dk \quad (6)$$

With $\|\cdot\|$ defining the Euclidean custom; $x(0)$, x_0 is the initial system state and nominal initial system state respectively; A cardinal value of output uncertainty $z(t)$ or constant, d , allows a limited difference $(x(0) - x_0)$ between sets; assuming $x(0)=x_0$, $z(k)$ and $d = 0$. Then, $w(\cdot)$ and $v(\cdot)$ represents allowable uncertainties which is described by;

$$\begin{bmatrix} w(k) \\ v(k) \end{bmatrix} = \emptyset(k, x(\cdot)) \quad (7)$$

Where, $\emptyset(\cdot)$ is a non-linear time-varying dynamic error function, $x(\cdot)$ stands for the function $x(k)$ for all time instant $k \in [0, K]$. Since $w(k)$ and $v(k)$ are dependent variables of function $x(\cdot)$, equivalently, they can rely on the system state too. $N = N^T \in +\mathbf{Z}$, $x_0 \in \mathbb{R}^n$, $d \in +\mathbf{Z}$ represents given state matrices, vectors and constants; $\Pi(k)$, $\Lambda(k) > 0$ are uniform matrix functions of time.

It becomes necessary to discretize the continuous-time error model, in order to achieve a robust discrete-time state estimator. However, James and Petersen, (1998) pointed out that, it is most uncomplicated technique to derive the properties of the discrete-time set-valued state estimator, if the model is discretized in reverse time rather than in forward time. The discretization of these nonlinear uncertain systems can be achieved using standard first-order iterative procedures such as, the Runge-Kutta and Euler techniques. Upon application, a discrete nonlinear reverse-time uncertainty model described by the state equation is derived;

$$x_k = A(k, x_{k+1}, u_k) - B_k w_k \quad (8)$$

$$y_{k+1} = C(k, x_{k+1}) + v_{k+1} \quad (9)$$

$$z_{k+1} = K(k, x_{k+1}, u_k) \quad (10)$$

With $A(\cdot)$, $C(\cdot)$, and $K(\cdot)$ representing discrete-time nonlinear functions, while $B(\cdot)$ represents a given time-varying matrix.

In order to obtain the error matrix of the discrete nonlinear reverse time system (Eqns. 8-10), the integral quadratic constraint is discretized (Eqn. 6) to obtain the sum quadratic constraint.

$$(x(0) - x_0)^T N(x(0) - x_0) + \sum_0^{k-1} (w_k^T \Pi_k w_k + v_{k+1}^T \Lambda_{k+1} v_{k+1}) \leq d + \sum_0^{k-1} \|z_{k+1}\|^2 \quad (11)$$

Where, $v_{k+1} = [y_{k+1} - C(k, x_{k+1})]$, $z_{k+1} = K(k, x_{k+1}, u_k)$, and the allowable errors $w(\cdot)$, $v(\cdot)$ are defines as;

$$\begin{bmatrix} w_k \\ v_{k+1} \end{bmatrix} = \Psi(k, x(\cdot)) \quad (12)$$

With, $\Psi(\cdot)$ being a discrete nonlinear dynamic uncertainty function.

The indeterminate system (Eqns. 8-10), with its equivalent SQC error depiction (Eqn. 11) is used to derive the robust filter and Riccatti equations, which finally determines the Extended Kalman Filter in discrete time.

If there exists an uncertain input chain, $w(\cdot)$ (Eqn. 7) such that, $V_k[x_k, w(\cdot)] \leq d$, given that $V_k[x_k, w(\cdot)]$ is the cost/loss function obtained from Eqn. 11, then the uncertainty creates an optimization problem (Eqn. 13) which arises due to the non-linearity of the system control (Eqns. 8-10) as a quadratic objective function. Therefore, the EKF can be derived by the optimization of the system control.

$$\inf_{w(\cdot)} V_k[x_k, w(\cdot)] \quad (13)$$

To optimize Eqn. 13, a dynamic programming approach is adopted. The discrete time Hamilton-Jacobi-Bellman (HJB) algorithm employed for the optimization process is given by;

$$V_{k+1}^*(x_{k+1}) = \min_{w_k} \{V_k^*[A(k, x_{k+1}, u_{0k}) - B_k w_k] + w_k^T \Pi_k w_k + v_{k+1}^T \Lambda_{k+1} v_{k+1} - z_{k+1}^T z_{k+1}\} \quad (14)$$

With initial state as;

$$V_0^*(x(0)) = (x(0) - x_0)^T N (x(0) - x_0) \quad (15)$$

As a primary step towards obtaining an optimal solution to the nonlinear system control problem (Eqns. 8-10, 13), V_k^* is approximated as,

$$V_k^*(x_k) \approx (x_k - \hat{x}_k)^T X_k (x_k - \hat{x}_k) + \Phi_k \quad (16)$$

With, $\hat{x}_k$ signifying the optimized system estimate; X_k is a symmetric matrix. The evaluation of Eqn. 15 against Eqn. 16 results to a review of the initial condition to $\hat{x}(0) = x_0$, $X(0)=N$, $\Phi(0)=0$.

Applying the approximate result (Eqn. 16) to the HJB (Eqn. 14)

$$\begin{aligned} & (x_{k+1} - \hat{x}_{k+1})^T X_{k+1} (x_{k+1} - \hat{x}_{k+1}) + \Phi_{k+1} \\ = & \min_{w_k} \{ [A(k, x_{k+1}, u_{0k}) - B_k w_k \hat{x}_k]^T \} \times X_k [A(k, x_{k+1}, u_{0k}) - B_k w_k - \hat{x}_k] + v_{k+1}^T \Lambda_{k+1} v_{k+1} + \\ & \Phi_k + w_k^T \Pi_k - z_{k+1}^T z_{k+1} \} \end{aligned} \quad (17)$$

Applying completing the squares method to the minimization problem in Eqn. 17, an optimal value for w_k denoted by w_k^* is attained

$$w_k^* = [\Pi_k + B_k^T X_{k+1} B_k]^{-1} B_k^T X_k \times [A(k, x_{k+1}, u_k) - \hat{x}_{k+1}] \quad (18)$$

Substituting $w_k = w_k^*$

$$\begin{aligned} (x_{k+1} - \hat{x}_{k+1})^T X_{k+1} (x_{k+1} - \hat{x}_{k+1}) + \Phi_{k+1} = & \{A(k, x_{k+1}, u_{0k})^T \Xi_k A(k, x_{k+1}) - A(k, x_{k+1}, u_{0k}) - \\ & A(k, x_{k+1}, u_{0k})^T \Xi_k \hat{x}_k + C(k, x_{k+1})^T \Lambda_{k+1} C(k, x_{k+1}) + y_0(k+1)^T \Lambda_{k+1} y_{0k+1} - 2y_0(k+1)^T \Lambda_{k+1} C(k, x_{k+1}) - \\ & \hat{x}_k^T \Xi_k A(k, x_{k+1}, u_{0k}) + \hat{x}_k^T \Xi_k \hat{x}_k - z_k + 1^T z(k+1) + \Phi_k\} \end{aligned} \quad (19)$$

Where;

$$\Xi_k = X_k - X_k B_k [\Pi_k + B_k^T X_{k+1} B_k]^\vee \times B_k^T X_k \quad (20)$$

In Eqn. 20 above, $(\cdot)^\vee$ represents the Moore-Penrose pseudo inverse. The pseudo-inverse in Eqn. 20 can be substituted for a normal matrix inverse such that the matrix $(\Pi_k + B_k^T X_{k+1} B_k) > 0$. This condition will hold in all cases for which a proper solution is possible for the symmetric matrix (X_k).

Eqn. 19 poses a nonlinear difference function which can be solved by attempting to linearize the non-linearities of Eqns. 8-10 about the optimized systems estimate ($\hat{x}_k$)

$$\begin{aligned} A(k, x_{k+1}, u_{0k}) & \approx A(k, \hat{x}_k, u_{0k}) + \nabla_x A(k, \hat{x}_k, u_{0k}) \times x_{k+1} - \hat{x}_k \\ K(k, x_{k+1}, u_{0k}) & \approx K(k, \hat{x}_k, u_{0k}) + \nabla_x K(k, \hat{x}_k, u_{0k}) \times (x_{k+1} - \hat{x}_k) \\ C(k, x_{k+1}, u_{0k}) & \approx C(k, \hat{x}_k, u_{0k}) + \nabla_x C(k, \hat{x}_k, u_{0k}) \times (x_{k+1} - \hat{x}_k) \end{aligned} \quad (21)$$

The current state estimate ($\hat{x}_k$) maintains an intense conformity with the linearization point just as the standard Extended Kalman Filter estimate (Abhijit and Anavatti, 2009).

See Ξ_k defined in Eqn. 20. Suppose the recursive functions (Eqns. 8-10) have solutions $\hat{x}_k, X_k, \Phi_k \mid X_k, \Pi_k + B_k^T X_{k+1} B_k = +Z$ for $k=1,2,3,\dots,K$, therefore the parallel approximate set-valued state estimate is given by;

$$Z_K[x_0, u_0(\cdot) \Big|_1^K, y_0(\cdot) \Big|_1^K, d] \cong \{x_K \in \mathbb{R}^n: (x_K - \hat{x}_K)^T X_K (x_K - \hat{x}_K) \leq d - \Phi_K\} \quad (22)$$

It is pertinent to point out that if the reverse-time discrete-time error models (Eqns. 21-24) has the same property of linearity as their counterparts (Eqns.15-16), then the obtained recursive functions produces a discrete-Time Extended Kalman Filter that is in line with the solution generated in Eqns. 15, 16.

3.3 Error bounds and significance of non-linear observabilities of the EKF

The EKF projects two common formulations; a two-step recursion consisting of time and measurement updates but having to re-linearize between steps (Fig. 2), and a one-step formulation in terms of the *a priori* variables. In as much as these two formulations presents varying performances and transient behaviors, their convergence properties remain constant.

Considering a non-linear stochastic system given by Eqns. 7-10, let the following assumptions hold;

- i. There exist positive real numbers $a, c, p, P > 0$ such that bound $n \geq 0$ is fulfilled for various matrices: $\|A_n\| \leq a; \|C_n\| \leq c$
- ii. A_n and C_n are non-singular for every $n \geq 0$
- iii. There exist positive real numbers such that the nonlinear functions and estimation errors are bounded in mean squares and probability ones.

The nonlinear system presented by Eqns. 2-5 satisfies the nonlinear observability rank condition at $y \in R^q$, if the nonlinear observability matrix $U(y)$ has full rank q at y , the results can be easily stated (see Konrad et al., 1999, for more details).

Fig. 2. A schematic illustration of the EK filter's step applied for data assimilation in this study.

4. Results and discussion

4.1 Experimental setup

In Nigeria, it is occasionally required to transform point coordinates between the global (earth-centered frame based on the WGS84 ellipsoid) and local reference frame (Minna datum based on the Clarke 1880 ellipsoid). However, the possibility to efficiently transform coordinates from one datum to another is largely based on the (i) mode of determination of the two reference frames, (ii) the accuracies of the reference frames as well as their interaction between each other (i.e. homogeneity). Given the inherent dissimilarities of the Nigerian geodetic reference frame (Minna datum), it has always been near impossible to achieve a set of transformation parameters that has a ubiquitous effect of the entire country. This therefore warranted the optimization of the first order controls of the Minna geodetic reference frame using the well-known EKF technique.

The results presented in this study suggest that the inclusion of noise in the EKF through the use of perturbed observations, especially from the GPS readings may lead to a degraded performance relative to other state estimators. One potential problem with the EKF in general is the influence of spurious correlations with physically unrelated parameters (Kallapur et al., 2009). This is the motivation for limiting the impact of spatially remote observations and amplifying the notational and conceptual clarity to the systems of linear equations using the Moore-Penrose pseudo inverse (eqn.20). Fig. 3 shows the bias between the assimilated observations (x_k) against the GPS and classical datasets used in our experiment. It attempts to show the performance for the non-linear observation case described in eqns. 2-5. It can be argued however, that the high degree of correlation between the assimilated datasets and original observations may arise from the similarity of the datasets employed in the validation process. However, we must point out that the general aim being to strengthen the integrity of the local geodetic framework can only be accepted when the entire framework forms part of the test procedures. The veracity of our assimilation however, is well founded in our minimal bias between the obtained datasets (x_k) and the original datasets (GPS and classical readings). Although the classical datasets tend to align more with the assimilated datasets (Table 1), this can largely be due to the fact that the local triangulation and trilateration observations procedures aligned more with the Minna datum than the GPS-procedures which is primarily oriented to the WGS-84 reference frame, as well as the inherent transformation errors which will be encountered in attempting to transform the GPS readings to the local system for assimilation. Without the implementation of the Moore-Penrose pseudo inverse technique, the rms error of the EKF (x_k) is expected to increase. However, we report that the EKF diverges from the truth for all values of covariance inflation. The impact of the Moore-Penrose pseudo inverse technique on rms is not surprising, especially as it gives way for more observations to impact each state variable. This effect on rms error is consistent with past studies (Hamill et al., 2001; Anderson, 2001) as more observations with weak physical relations can be allowed to influence individual state variables. This behavior was largely felt in the assimilated dataset (x_k) since the noise introduced from the original observations can itself lead to spurious correlations. Fig. 5 on the other hand gives a direct spatial overlay of the two assimilated datasets (GPS and classical) and the product (x_k). The EKF was noticed to have an extra source of potential filter divergence since only the expected values of the updated sample covariance are equal to the iterations. Occasionally, it was noticed that the updated covariance is smaller than the expected value. Basically, it is expected that this results in a prior estimate parameter with reduced covariance and increased error at the next iterate, which in turn is expected to result in an even more reduced estimate after the next assimilation. In a bid to circumvent this, the upcovariance inflation technique can be employed to avoid filter divergence. But this has been reported to cause a burden on the observational dataset (Anderson, 2001) by giving it too much weight and at other times when the updated covariance estimates are not too large by chance. Generally, we deduce that the EKF's addition of random noise through perturbed observations at each assimilation step appears to be responsible in degrading the quality of assimilation derived, but still provide a sufficient enough result for our optimization (Fig. 5, Table 1). However, we must point out that the noise added to the EKF could be of additional concern in less tolerant systems.

4.1.1 Error sensitivity analysis

We first investigate the effects of the separate datasets, i.e., both the GPS coordinates and classical geodetic observations on the filter estimates (Fig. 3a,b). The integration of the pseudo observation (classical geodetic coordinates) in the update step of the EKF modifies the impact of the derived GPS coordinates on the estimation of the system state. Also, the noise matrices would also bring about a difference in the individual contribution of the GPS and pseudo-observations. The correlation between the filter estimates and both observations is shown in Table 1 as well as the variation between them. A higher correlation is noticed when the EKF having minimum error values is overlaid to the observed coordinate values. The least imbalance error is also achieved in this scenario. In general, existent anomalies in the GPS and classical data values which bring about noisy observations decrease the correlation between the derived system state and observation datasets. To further increase the veracity of the system state, outliers are removed from the analysis due to its highly corrosive nature to the accuracy of the entire model (Fig. 4).

Fig. 3. Plot showing bias in the rescaled coordinates between (a) filtered estimate and GPS coordinates and (b) filtered estimate and classical observation.

4.1.2 Outlier regulation

Outliers may be introduced due to fault in data collection, error during data entering, or due to data transmission errors. Outlier regulation in this study however, was achieved using the principal component analysis (PCA). The use of this statistical tool in outlier regulation is consistent with past studies (e.g., Chen et al., 2020). The principal component analysis-based cumulant scheme is basically used to statistically decompose a given dataset into diverse dimensions while retaining interpretability and diminishing data loss (e.g. Ndehedehe et al., 2016b; Ziehe, 2005; Jolliffe, 2002; Kalu et al., 2021c). The simplicity and proficiency in the explanation of optimized variance parameters, with a minimum number of principal components, contributed to its popularity and relevance of this statistical data analysis tool in climate science (Ndehedehe et al. 2016a, 2017).

Fig. 4. Observation outliers determined using PCA analysis used to detect outliers for the GPS and Classical observations, each having 17 and 8 outliers respectively. The relative higher outliers in GPS are due to the prevalent inherent transformation errors in relating different reference frames (Kalu et al., 2021). On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the red dot symbol.

PCA gives the best possible illustration of a p -dimensional dataset in q -dimensions ($q < p$) by attempting to maximize the variance in q dimensions. For the purpose of regularization, outliers and gross errors were eradicated from the observation matrix $\hat{y}_k$ (eqns. 2,3) in order to achieve a robust principal component analysis (e.g., Jolliffe, 2002). The need for the identification of periodic signals embedded in multi resolution datasets is increasing, and has resulted in the application of the robust PCA to several vigorous lines of research such as; image processing, machine learning, bioinformatics or web data analysis including the regionalization and analysis of global geophysical time series data (e.g. Luet al., 2017; Ndehedehe, 2019; Agutu et al., 2017; Ivits et al., 2014; Montazerolghaem et al., 2016; Dutta& Das, 2019). Its unique and simple interactions allow for its broad range of application in spatial science (Westra et al., 2010), the choice of this technique in this study is due to its ability to linearly combine centered (original) variables provided their units are the same. It is common practice to use a predefined percentage of total variance in deciding the number of PCs to be retained (70% of total unpredictability is normal, if subjective, cut-off point) (Ndehedehe and Vagner, 2019). The quality of any q -dimensional estimation can be measured by the inconsistency of the set of retained PCs. The sum of variances of the p original variables is the trace of the covariance matrix $\mathbf{S}$. The standard measure of quality of a given PC is represented by a simple matrix theory which attempts to show the proportion of total variance accounted for,

$$\pi_j = \frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{tr(\mathbf{S})}, \quad (23)$$

Where $tr(\mathbf{S})$ represents the trace of $\mathbf{S}$. The inherent incremental nature of the PC explains the proportion of total variance explained by a set U , and is often denoted as a % of total variance accounted for; $\sum_{j \in U} \pi_j \times 100\%$. So, the PCA decomposition approach enhances the perception quality of the p dimensional GPS coordinates and enables optimal eminence to its dimensionality estimation. The use of a component extraction method such as PCA to decompose raw GPS coordinates into sets of principal components might probably help address some questions such as the impacts of the clock corrections, ephemeris and time parameters in the positioning of a ground control point, in comparison with an existing classical geodetic control. However, this study did not dwell on that. PCA reduces the dimensions of multivariate data by linearizing functions of the original variables and parsing them into new variables.

The linear combinations for K principal components at a given discrete-time T is given by,

$$\begin{bmatrix} y_{T,1} = P_{11}\hat{\mathbf{x}}_{T,1} + P_{12}\hat{\mathbf{x}}_{T,2} + P_{13}\hat{\mathbf{x}}_{T,3} + \dots + P_{1K}\hat{\mathbf{x}}_{T,K} \\ y_{T,2} = P_{21}\hat{\mathbf{x}}_{T,1} + P_{22}\hat{\mathbf{x}}_{T,2} + P_{23}\hat{\mathbf{x}}_{T,3} + \dots + P_{2K}\hat{\mathbf{x}}_{T,K} \\ \dots \\ y_{T,K} = P_{K1}\hat{\mathbf{x}}_{T,1} + P_{K2}\hat{\mathbf{x}}_{T,2} + P_{K3}\hat{\mathbf{x}}_{T,3} + \dots + P_{KK}\hat{\mathbf{x}}_{T,K} \end{bmatrix} \quad (24)$$

The state matrix $\hat{\mathbf{x}}_{T,K}$ contains rows corresponding to the discrete time T , and the variables K , representing the positional values of the first order controls in both frames. In the linear arrangement above (Eqn. 24), the y values are orthogonal, while $y_{T,1}$ to $y_{T,K}$ represents the PC signifying the variance of the system matrix with the initial PC explaining the highest variability. The weight of the original variables in the PCs is provided by the coefficients of the linear observations (eigenvectors) also known as factor loadings. The factor loadings, also referred to as component coefficient is determined by dividing the factor's eigenvalue by the total number of variables. While the normalized eigenvectors (e.g., $P_{11}, P_{12}, P_{13}, \dots, P_{KK}$) are the component coefficients, each characteristic root measures the amount of variation in the total sample accounted for each factor (Ndehedehe et al., 2016a) in our case, the GPS coordinates. A factor's characteristic roots may be computed as the sum of its squared component coefficient for all the variables.

The use of PCA in this study provided a good knowledge in the decomposition of the raw classical and GPS coordinates into sets of principal components (PCs). Most importantly, it provided an avenue to regulate and manage outliers in order to achieve a robust system matrix.

Fig. 5. Spatial average observation series of the system state showing the GPS and classical in relation to the filtered estimate (the Extended Kalman Filter is modeled in discrete time; eqn. 8-10). The abscissa has been rescaled for easy visualization

4.2 Validation analysis using multi-linear regression

The result of the final system estimate is validated against the originally obtained classical data (Fig. 5). We apply the multi-linear regression model to comprehend the degree of association between the derived system matrix and the original observations.

Table 1. Multi linear regression analysis to determine the significance of the relationship between the derived system matrix and the original observations (unit: mm)

	Intercept	GPS Coordinates ($r^2 = 0.85$)	Classical Observations ($r^2 = 0.92$)
Coefficients	10.137	-2.109	77.30
RMSE	0.375	0.273	0.251
t-stat	0.007	-0.312	0.459
p-value	0.720	0.815	0.858
Lower 95%	-0.072	0.004	0.004
Upper 95%	0.0691	0.010	0.017

The multi-linear regression analysis is a robust statistical technique which applies a least squares approach in modeling geodetic datasets. In its application, the relationship between a dependent variable (e.g., derived system state) and an independent variable (e.g., GPS or classical observations) is evaluated in order to model the degree of significance and relationship between them. To achieve this, the regression model,

$$\tilde{D}_{K/R} = \gamma_0 + \gamma_{1k_*} + \sin(2\pi k_*) + \cos(2\pi k_*) + \sin(4\pi k_*) + z \quad (25)$$

is employed (see Rieser et al., 2010). Where, $\tilde{D}_{K/R}$ is a time function (k_*) of either the GPS or classical observation dataset, γ_0 and γ_1 refers to the varied constant coefficients for both GPS and classical observations, the eccentricity between the derived observations and model outputs is represented by z . The Root Mean Square Error (RMSE) is given by,

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k (obs\left(\tilde{D}_{\frac{T}{R}}\right) - sim\left(\tilde{D}_{\frac{T}{R}}\right))^2} \quad (26)$$

Where $obs\left(\tilde{D}_{\frac{T}{R}}\right)$ and $sim\left(\tilde{D}_{\frac{T}{R}}\right)$ represent GPS/classical observations and model outputs from eqn. 25 respectively for the time k . With the use of other assessment parameters such as the RMSE (Table 1), bias (Fig. 3) and so on, the multi-linear regression method maintains a unique view of accuracy assessment for geodetic studies.

Results show that the classical observations have a higher association degree to the derived system matrix than the GPS coordinates. This is largely as a result of inherent GPS errors accumulated during the origin transformation to a topocentric frame. However, the r values from both GPS and classical observations show that the regression is significant, and the exercise was successful.

5. Summary and Conclusion

GPS coordinates are assimilated through extensive filtering models to improve the reliability of positioning and satisfy the science of geodetic operations. For this reason, we proposed the discrete-Time EKF to optimize classical based data for first order geodetic operations in Nigeria. This technique fundamentally attempts to merge two homogenous data through assimilation in order to improve the veracity of the existing positional values of the 157 controls used. Higher correlation between the derived system state and the classical observations ($r^2 = 0.92$, Table 1, Fig. 5) are noticed in the study. This is to be expected since the GPS coordinates made internal transformations to align with the coordinate system of the classical data. The strength of the derived system estimate shown in

Fig. 5 indicates the imminent potential in relating heterogeneous reference frames, and still achieve high accuracy and results.

There are some key factors that can affect the performance of the discrete time EKF. Most importantly, the linearization of a discrete-time non-linear system using dynamic programming requires inverting the non-linear functions describing the non-linear discrete time system, which is almost impossible. Instead, a reverse time discrete time system dynamic is considered which is extremely cumbersome and requires very high computational power. The relative certainty of the observations and derived system state estimates is of utmost importance, as the response of the filter is majorly hinged on the gain matrix. The gain matrix served as the relative weight given to the observations and the current state estimates. Due to the inadequacy of uncertainty parameters, the gain matrix was tuned in order to achieve an optimal system matrix. Whenever a high gain matrix is encountered in the system, more weight is placed on the most recent observations, following them more responsively, whereas, with a low gain matrix, the model sticks to the filtered estimates more closely. At extreme points, a low gain matrix closes in on zero, in order to smooth out noise and reduce the reaction of the system state, and a high gain matrix, close to one will result to a jumpier system state estimation.

This study evaluates several different error scenarios and investigates their impacts on the achieved results (Figs. 3,5, Table 1). However, the implemented suppositions, especially using a static uncertainty, may reduce accuracy given that various datasets have varied performance index in distinct applications. Therefore, it is pertinent that more investigations are carried out in order to fully ascertain the filters capability especially in terms of data uncertainties and explore the possibility of the application of multiple datasets especially in other areas of spatial science. Results from our analyses show that,

- (i) GPS stations with unknown or inaccurate uncertainties tend to affect the functionality of the system state. This was however tackled through the outlier regulation technique employed in this study.
- (ii) During the process of geodetic re-analysis with the raw data, we discovered that existing controls in the southern region of Nigeria maintained a very high relationship with the derived system matrix. This was to be expected given that a large number of the dataset used in this analysis was gotten from the region (Fig. 1) as well as the availability of corresponding uncertainties of each GPS data point used in the training.
- (iii) Due to the severity of non-linearity in the network, it was necessary to implement the set membership state estimation methodology, in order to discretize the continuous-time error model. This was a necessary step to achieve a robust estimation of the system state.
- (iv) In a general examination between the EKF and the conventional least squares adjustment in the accuracy assessment, the EKF outperformed the least squares in terms of fit between expected and computed values of transformed points in the southern and eastern Nigeria. But the remaining parts of the country exhibited little or no difference when compared to the conventional least squares adjustment. This, therefore, reiterates on the necessity of more points in the estimation of a system state in order to achieve an optimal filtered system estimate for geodetic operations.

The relatively low correlation between the derived system matrix and the GPS coordinates has already proposed to be due to the inherent problems resulting from transforming coordinate systems. This is to be expected since the earth on its own has a very rugged shape and this makes it difficult to perform necessary computations on it, thus the adoption of datums and ellipsoids. This consequently motivates the deduction that, coordinate transformation may not be the ultimate solution to datum problems. Overall, the scope of this study was limited by data gaps (non-availability of GPS uncertainty dataset for a large region of Northern and Western Nigeria) and the non-uniformity of a defined GPS mode in the acquisition of data for the entire study region. This could impact on the interpretation of results for a large portion of the study area. Despite these constraints, statistical relationships developed from this data assimilation exercise are adequate to guide geodesists in performing highly accurate transformation of coordinates between the local and global datums

especially in Southern Nigeria. Support Vector Machine and Artificial Neural Network, for example, have shown potentials in improving the integrity of geodetic frameworks. The emergence of a deep learning architecture to improve the stability of system states (Gin et al. 2020) are indications of interesting future research directions that can be leveraged upon for effective optimization of geodetic frameworks.

Declarations

Funding: Not applicable

Competing Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials: One hundred and fifty-seven first-order geodetic controls spread across Nigeria were used in this study. These controls were obtained from the office of the surveyor general of the Nigerian federation (OSGOF) and made available for academic purposes and not yet available to the general public.

Author contribution: Ikechukwu Kalu: Writing – original draft, Writing – review & editing, Methodology, Formal analysis, Conceptualization, Investigation, Software. **Christopher E.**

Ndehedehe: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Software. **Onuwa Okwuashi:** Conceptualization, Writing – review & editing, Formal analysis.

Aniekan Eyoh: Resources

Acknowledgments: The authors are grateful to the office of the surveyor general of Nigeria (OSGOF), and the department of geoinformatics and surveying, University of Uyo, for all the data used in this study.

References

- Abhijit, G K., Anavatti, S G., 2009. A discrete-Time Robust Extended Kalman Filter for Uncertain Systems with Sum Quadratic Constraints. *IEEE Trans on Automatic Control*, vol. 54, No 4
Doi: 10.1109/TAC.2008.2010962
- Agutu N, Awange J, Zerihun A, Ndehedehe C, Kuhn M, Fukuda Y. 2017. Assessing multi-satellite remote sensing, reanalysis, and land surface models' products in characterizing agricultural drought in East Africa. *Remote Sens Environ* 194(0):287–302. <https://doi.org/10.1016/j.rse.2017.03.041>
- Alonge, F., Cangemi, T., D'Ippolito, F., Fagiolini, A., & Sferlazza, A. (2014). Convergence analysis of extended Kalman filter for sensorless control of induction motor. *IEEE Transactions on Industrial Electronics*, 62(4), 2341-2352.
- Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, 129(12), 2884-2903.
- Awange, J L., Bae, K H., Claessens, S J., 2008. Procrustean solution of the 9-parameter transformation problem. *Earth Planets Space*, 60, 529–537
- Belhajem I., Ben Maissa Y., Tamtaoui A., (2016) A robust low cost approach for real time car positioning in a smart city using Extended Kalman Filter and evolutionary machine learning, *4th IEEE International Colloquium on Information Science and Technology (CiSt)*, Tangier, 2016, pp. 806-811, doi: 10.1109/CIST.2016.7804998.
- Bertino, L., Evensen G., Wackernagel, H., 2003. Sequential Data Assimilation Techniques in Oceanography, *International Statistical Review*, vol. 71, No. 2 (Aug., 2003). pp. 223–241.

- Bertsekas, D., & Rhodes, I. (1971). Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16(2), 117-128.
- Charles D G. (2010) Adjustment computations: spatial data analysis John Wiley & Sons, INC., Hoboken, New Jersey. ISBN 978-0-470-46491-5
- Chen, X., Zhang, B., Wang, T., Bonni, A., & Zhao, G. (2020). Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. *BMC bioinformatics*, 21(1), 1-20.
- Chiang, H. S., Chen, M. Y., & Huang, Y. J. (2019). Wavelet-based EEG processing for epilepsy detection using fuzzy entropy and associative petri net. *IEEE Access*, 7, 103255-103262.
- Dakhlallah, J., Glaser, S., Mammar, S., & Sebsadji, Y. (2008, June). Tire-road forces estimation using extended Kalman filter and sideslip angle evaluation. In *2008 American control conference* (pp. 4597-4602). IEEE.
- David González, Alberto Badías, Icíar Alfaro, Francisco Chinesta, Elías Cueto, (2017) Model order reduction for real-time data assimilation through Extended Kalman Filters, Computer Methods in Applied Mechanics and Engineering, Volume 326, <https://doi.org/10.1016/j.cma.2017.08.041>
- de Jesús Rubio, J., Lughofer, E., Pieper, J., Cruz, P., Martinez, D. I., Ochoa, G., ... & Garcia, E. (2021). Adapting H-infinity controller for the desired reference tracking of the sphere position in the maglev process. *Information Sciences*, 569, 669-686.
- de Rubio, J. J. (2020). Stability analysis of the modified Levenberg-Marquardt algorithm for the artificial neural network training. *IEEE transactions on neural networks and learning systems*.
- Dutta, I., & Das, A. (2019). Modeling dynamics of peri-urban interface based on principal component analysis (PCA) and cluster analysis (CA): a study of English Bazar Urban Agglomeration, West Bengal. *Modeling Earth Systems and Environment*, 5(2), 613-626.
- Featherstone, W. and Vanicek P., 1999. The role of coordinate systems, coordinates and heights in horizontal datum transformations, *Aust. Surv.*, 44, 143–150.
- Frohlich, H. and G. Broker, 2003. Trafox version 2.1—3d-Kartesische Helmert-Transformation, <http://www.koordinatestransformation.de/data/trafox.pdf>.
- Gharamti, M.E., Valstar, J., Janssen, G., Marsman, A., Hoteit, I., 2016. On the efficiency of the hybrid and the exact second-order sampling formulations of the EnKF: a reality-inspired 3-D test case for estimating biodegradation rates of chlorinated hydrocarbons at the port of Rotterdam. *Hydrol. Earth Syst. Sci.* 20 (4561–4583), 2016. <https://doi.org/10.5194/hess-20-4561-2016>.
- Gin, C., Lusch, B., Brunton, S. L., & Kutz, J. N. (2021). Deep learning models for global coordinate transformations that linearise PDEs. *European Journal of Applied Mathematics*, 32(3), 515-539.
- Gogoi, R. B., Kutty, G., & Borgohain, A. (2021). Impact of INSAT-3D satellite-derived wind in 3DVAR and hybrid ensemble-3DVAR data assimilation systems in the simulation of tropical cyclones over the Bay of Bengal. *Modeling Earth Systems and Environment*, 1-11.
- Grafarend EW, Awange JL., 2003. Nonlinear analysis of the three-dimensional datum transformation [conformal group C7(3)]. *J Geod* 77:66–76
- Hamid Moradkhani, Soroosh Sorooshian, Hoshin V. Gupta, Paul R. Houser, (2005) Dual state–parameter estimation of hydrological models using ensemble Kalman filter,

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11), 2776-2790.

Han J Y ., (2010). Non-iterative approach for solving the indirect problems of linear reference frame transformations. *J SurvEng* 136(4):150–156

Han, R., Tian, X.J., Fu, Y., Cai, Z.N., 2015. Real-data assimilation experiment with a joint data assimilation system: assimilating carbon dioxide mole fraction measurements from the Greenhouse gases Observing Satellite. *Atmos. Oceanic Sci. Lett.* 9 (2), 107–113. <https://doi.org/10.1080/16742834.2016.1133070>.

Hoteit, I., Luo, X., Pharm, D.T., 2012. Particle Kalman filtering: a nonlinear Bayesian framework for ensemble Kalman filters. *Mon. Weather Rev.* 140 (2), 528-542.

HuaïenZeng, Qinglin Yi, Yue Wu., 2016. Iterative approach of 3D datum transformation with a non-isotropic weight. *ActaGeodGeophys* (2016) 51:557–570. doi: 10.1007/s40328-015-0144-2

Ivits E, Horion S, Fensholt R, Cherlet M (2014) Drought footprint on European ecosystems between 1999 and 2010 assessed by remotely sensed vegetation phenology and productivity. *Glob Chang Biol* 20(2):581– 593. <https://doi.org/10.1111/gcb.12393>

James, M R., Petersen, I R., 1998. Nonlinear state estimation for uncertain systems with an integral constraint. *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 46, no. 11, pp. 2926-2937

Kalra, S., Kumar, S., & Routray, A. (2019). Simulation of heavy rainfall event along east coast of India using WRF modeling system: impact of 3DVAR data assimilation. *Modeling Earth Systems and Environment*, 5(1), 245-256.

Kalu, I., Ndehedehe, C. E., Okwuashi, O., & Eyoh, A. E. (2021a). A comparison of existing transformation models to improve coordinate conversion between geodetic reference frames in Nigeria. *Modeling Earth Systems and Environment*, 1-14.

Kalu, I., Ndehedehe, C. E., Okwuashi, O., & Eyoh, A. E. (2021b). Integration of satellite geodetic observations for regional geoid modeling using remove-compute-restore technique. *Earth Science Informatics*, 1-19.

Kalu, I., Ndehedehe, C. E., Okwuashi, O., & Eyoh, A. E. (2021c). Assessing freshwater changes over Southern and Central Africa (2002–2017). *Remote Sensing*, 13(13), 2543.

Kallapur, A. G., Petersen, I. R., & Anavatti, S. G. (2009). A discrete-time robust extended Kalman filter for uncertain systems with sum quadratic constraints. *IEEE Transactions on Automatic Control*, 54(4), 850-854.

Khaki M., Ait-El-Fquih B., Hoteit I., Forootan E., Awange J., Kuhn M., 2018. Unsupervised ensemble Kalman filtering with an uncertain constraint for land hydrological data assimilation, *Journal of Hydrology*, 564, 175-190, ISSN 0022-1694, <https://doi.org/10.1016/j.jhydrol.2018.06.080>.

Khaki M., Ait-El-Fquih B., Hoteit I., Forootan E., Awange J., Kuhn M., 2017. A two-update ensemble Kalman Filter for land hydrological data assimilation with an uncertain constraint. *Journal of Hydrology*, 555, 447-462 ISSN 0022-1694, <https://doi.org/10.1016/j.jhydrol.2017.10.032>.

Koch, D. P., Wheeler, D. O., Beard, R. W., McLain, T. W., & Brink, K. M. (2020). Relative multiplicative extended Kalman filter for observable GPS-denied navigation. *The International Journal of Robotics Research*, 39(9), 1085–1121. <https://doi.org/10.1177/0278364920903094>

- KonradReif, Stefan Gunther, EnginYaz, Rolf Unbehauen, 1999. Stochastic stability of the Discrete-Time Extended Kalman Filter. *IEEE Trans. Automatic Control*, vol. 44, No 4
- Lievens, H., Reichle, R.H., Liu, Q., De Lannoy, G.J.M., Dunbar, R.S., Kim, S.B., Das, N.N., Cosh, M., Walker, J.P., Wagner, W., 2017. Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophys. Res. Lett.* 44, 6145–6153. <https://doi.org/10.1002/2017GL073904>.
- Liu H, Van Zuylen H, Van Lint H, Salomons M. (2006) Predicting Urban Arterial Travel Time with State-Space Neural Networks and Kalman Filters. *Transportation Research Record*.1968(1):99-108. doi:[10.1177/0361198106196800112](https://doi.org/10.1177/0361198106196800112)
- Lu, W., Atkinson, D. E., & Newlands, N. K. (2017). ENSO climate risk: predicting crop yield variability and coherence using cluster-based PCA. *Modeling Earth Systems and Environment*, 3(4), 1343-1359.
- Mathes, A., 2002. EasyTrans Pro-Edition, ProfessionelleKoordinaten transformation fur Navigation,Vermessung und GIS, ISBN 978-3-87907-367- 2, CD-ROM with manual
- McLaughlin, D., 2002. An integrate approach to hydrologic data assimilation: interpolation, smoothing, and filtering, *Adv. Water Resour.* 25, 12751286
- Montazerolghaem M, Vervoort W, Minasny B, McBratney A (2016) Long-term variability of the leading seasonal modes of rainfall in south-eastern Australia. *Weather and Climate Extremes* 13:1–14. <https://doi.org/10.1016/j.wace.2016.04.001>
- Muluye, G. Y. (2011) Improving long-range hydrological forecasts with extended Kalman filters. *Hydrol. Sci. J.* **56**(7), 1118–1128.
- Ndehedehe C., (2019) The water resources of tropical West Africa: problems, progress and prospect. *ActaGeophysica* 67(2):621–649. <https://doi.org/10.1007/s11600-019-00260-y>
- Ndehedehe C, Awange J, Agutu N, Kuhn M, Heck B (2016a) Understanding changes in terrestrial water storage over West Africa between 2002 and 2014. *Adv Water Resour* 88:211–230. <https://doi.org/10.1016/j.advwatres.2015.12.009>
- Ndehedehe C, Agutu N, Okwuashi O, Ferreira VG (2016b) Spatio-temporal variability of droughts and terrestrial water storage over Lake Chad Basin using independent component analysis. *J Hydrol* 540:106–128. <https://doi.org/10.1016/j.jhydrol.2016.05.068>
- Ndehedehe C E, Awange J, Kuhn M, Agutu N, Fukuda Y (2017) Climate teleconnections influence on West Africa’s terrestrial water storage. *Hydrol Process* 31(18):3206–3224. <https://doi.org/10.1002/hyp.11237>
- Oisiovici, R. M., & Cruz, S. L. (2000). State estimation of batch distillation columns using an extended Kalman filter. *Chemical Engineering Science*, 55(20), 4667-4680.
- Petersen, I. R., & Savkin, A. V. (1999). *Robust Kalman filtering for signals and systems with large uncertainties*. Springer Science & Business Media.
- Poku-Gyamfi Y, Schueler, T (2008) Renewal of Ghana’s Geodetic Reference Network. 13th FIG Symposium on Deformation Measurement and Analysis, 4th IAG Symposium on Geodesy for Geotechnical and Structural Engineering, LNEC, LISBON, 2008, pp 1–9

- Renzullo, L.J., Van Dijk, A.I.J.M., Perraud, J.M., Collins, D., Henderson, B., Jin, H., Smith, A.B., McJannet, D.L., 2014. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. *J. Hydrol.* 519, 2747–2762. <https://doi.org/10.1016/j.jhydrol.2014.08.008>.
- Rhudy, M. B. (2015). Increasing the convergence rate of the extended Kalman filter. In *AIAA Infotech@ Aerospace* (p. 0187).
- Rieser, D., Kuhn, M., Pail, R., Anjasmara, I. M., & Awange, J. (2010). Relation between GRACE-derived surface mass variations and precipitation over Australia. *Australian Journal of Earth Sciences*, 57(7), 887-900.
- Rubio, J. D. J., Pan, Y., Pieper, J., Chen, M. Y., & Sossa Azuela, J. H. (2021). Advances in Robots Trajectories Learning via Fast Neural Networks. *Frontiers in Neurorobotics*, 15, 29.
- Santosh D. H., Mohan P. G. K., (2014) Multiple objects tracking using Extended Kalman Filter, GMM and Mean Shift Algorithm - A comparative study, *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, Ramanathapuram, 2014, pp. 1484-1488, doi: 10.1109/ICACCCT.2014.7019350.
- Sarkar, A., Das, S., & Dutta, D. (2019). Computation of skill of a mesoscale model in forecasting thunderstorm using radar reflectivity. *Modeling Earth Systems and Environment*, 5(2), 443-454.
- Savkin, A. V., & Petersen, I. R. (1998). Robust state estimation and model validation for discrete-time uncertain systems with a deterministic description of noise and uncertainty. *Automatica*, 34(2), 271-274.
- Shen YZ, Chen Y, Zheng DH., 2006. A quaternion-based geodetic datum transformation algorithm. *J Geod* 80:233–239
- Simon D., 2006. Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches. Wiley - Interscience, ISBN-13: 978-0-471-70858-2.
- Slater, A.G., and Clark M. P., 2006. Snow Data Assimilation via an Ensemble Kalman Filter. *J. Hydrometeor.*, 7, 478-493, <https://doi.org/10.1175/JHM505.1>.
- Soriano, L. A., Zamora, E., Vazquez-Nicolas, J. M., Hernández, G., Madrigal, J. A. B., & Balderas, D. (2020). PD Control Compensation Based on a Cascade Neural Network Applied to a Robot Manipulator. *Frontiers in Neurorobotics*, 14.
- Soroush A., Richard T., Nordholm S., (2011) Estimation of the drill bit angular velocity using accelerometer measurements and extended Kalman filters, *7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Dubrovnik, 2011, pp. 68-71
- Tangdamrongsub, N., Steele-Dunne, S.C., Gunter, B.C., Ditmar, P.G., Sutanudjaja, E.H., Xie, T., Wang, Z., 2017. Improving estimates of water resources in a semi-arid region by assimilating GRACE data into the PCR-GLOBWB hydrological model. *Hydrol. Earth Syst. Sci.* 21, 2053–2074.
- Tian, S., Renzullo, L. J., Pipunic, R. C., Lerat, J., Sharples, W., and Donnelly, C. (2020). Satellite soil moisture data assimilation for improved operational continental water balance prediction, *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2020-485>, in review
- Tian, S., Tregoning, P., Renzullo, L.J., van Dijk, A.I.J.M., Walker, J.P., Pauwels, V.R.N., Allgeyer, S., 2017. Improved water balance component estimates through joint assimilation of GRACE water

storage and SMOS soil moisture retrievals. *Water Resour. Res.* 53.
<https://doi.org/10.1002/2016WR019641>.

van Dijk, A.I.J.M., Renzullo, L.J., Wada, Y., Tregoning, P., 2014. A global water cycle reanalysis (20032012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrol. Earth Syst. Sci.* 18, 29552973. <https://doi.org/10.5194/hess-18-2955-2014>.

Vargas, D. M. (2021). Superpixels extraction by an Intuitionistic fuzzy clustering algorithm. *Journal of Applied Research and Technology*, 19(2), 140-152.

Vrugt, J.A., terBraak, C.J.F., Diks, C.G.H., Schoups, G., 2013. Advancing hydrologic data assimilation using particle Markov chain Monte Carlo simulation: theory, concepts and applications. *Advances in Water Resources*, Anniversary Issue – 35 Years, 51, 457–478, doi: 10.1016/j.advwatres.2012.04.002.

WenzelT. A., BurnhamK. J., Blundell M. V., WilliamsR. A. (2006) Dual extended Kalman filter for vehicle state and parameter estimation, *Vehicle System Dynamics*, 44:2, 153171, doi: [10.1080/00423110500385949](https://doi.org/10.1080/00423110500385949)

Westra S, Brown C, Lall U, Koch I, Sharma A (2010) Interpreting variability in global SST data using independent component analysis and principal component analysis. *Int J Climatol* 30(3):333–346.
<https://doi.org/10.1002/joc.1888>

Wonnacott R (2007) A progress report on the AFREF project and its potential to support development in Africa. *Space Geodesy Workshop*, Matjiesfontein, 13–14 November

Yang S., Baum M., (2017) Extended Kalman filter for extended object tracking, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 4386-4390, doi: 10.1109/ICASSP.2017.7952985

Yang Y, 1999. Robust estimation of geodetic datum transformation. *Journal of Geodesy* 73: 268-274

Zaitchik, B.F., Rodell, M., Reichle, R.H., 2008. Assimilation of GRACE terrestrial water storage data into a land surface model: results for the Mississippi River Basin. *J. Hydrometeorol.* 9 (3), 535548.
<https://doi.org/10.1175/2007JHM951.1>.

Zeng HE, Yi QL (2010a) A new analytical solution of nonlinear geodetic datum transformation. In: *Proceedings of the 18th international conference on geoinformatics*. doi:10.1109/GEOINFORMATICS. 2010.5568048

Zhang, J., Campbell, J.R., Hyer, E.J., Reid, J.S., Westphal, D.L., Johnson, R.S., 2014. Evaluating the impact of multisensor data assimilation on a global aerosol particle transport model. *J. Geophys. Res. Atmos.* 119, 4674–4689. <https://doi.org/10.1002/2013JD020975>.

ZibarD., PielsM., Jones R.,SchäeffeC. G., (2016) Machine Learning Techniques in Optical Communication, in *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442-1452, doi: 10.1109/JLT.2015.2508502.

Ziehe A., (2005) Blind source separation based on joint diagonalization of matrices with applications in biomedical signal processing. PhD thesis, Universitat Potsdam. Retrieved from: <http://en.youscribe.com/> <http://en.youscribe.com/catalogue/reports-and-theses/knowledge/blind-source-separation-based-on-jointdiagonalization-of-matrices-1424347>. Accessed 1 December 2020