mathematics

Article

Automatic Calibration of Process Noise Matrix and Measurement Noise Covariance for Multi-GNSS Precise Point Positioning

Xinggang Zhang [1,2,3], Pan Li [3,*], Rui Tu [1,2], Xiaochun Lu [1,2], Maorong Ge [3,4] and Harald Schuh [3,4]

[1] National Time Service Center, Chinese Academy of Sciences, Shu Yuan Road, Xi'an 710600, China; zxg@ntsc.ac.cn (X.Z.); turui-2014@126.com (R.T.); luxc@ntsc.ac.cn (X.L.)

[2] Key Laboratory of Precision Navigation Positioning and Timing Technology, Chinese Academy of Sciences, Xi'an 710600, China

[3] German Research Centre for Geosciences (GFZ), 14473 Potsdam, Germany; maorong.ge@gfz-potsdam.de (M.G.); schuh@gfz-potsdam.de (H.S.)

[4] Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, 10623 Berlin, Germany

* Correspondence: panli@gfz-potsdam.de

Received: 28 January 2020; Accepted: 12 March 2020; Published: 2 April 2020

Abstract: The Expectation-Maximization algorithm is adapted to the extended Kalman filter to multiple GNSS Precise Point Positioning (PPP), named EM-PPP. EM-PPP considers better the compatibility of multiple GNSS data processing and characteristics of receiver motion, targeting to calibrate the process noise matrix Q_t and observation matrix R_t, having influence on PPP convergence time and precision, with other parameters. It is possibly a feasible way to estimate a large number of parameters to a certain extent for its simplicity and easy implementation. We also compare EM-algorithm with other methods like least-squares (co)variance component estimation (LS-VCE), maximum likelihood estimation (MLE), showing that EM-algorithm from restricted maximum likelihood (REML) will be identical to LS-VCE if certain weight matrix is chosen for LS-VCE. To assess the performance of the approach, daily observations from a network of 14 globally distributed International GNSS Service (IGS) multi-GNSS stations were processed using ionosphere-free combinations. The stations were assumed to be in kinematic motion with initial random walk noise of 1 mm every 30 s. The initial standard deviations for ionosphere-free code and carrier phase measurements are set to 3 m and 0.03 m, respectively, independent of the satellite elevation angle. It is shown that the calibrated R_t agrees well with observation residuals, reflecting effects of the accuracy of different satellite precise product and receiver-satellite geometry variations, and effectively resisting outliers. The calibrated Q_t converges to its true value after about 50 iterations in our case. A kinematic test was also performed to derive 1 Hz GPS displacements, showing the RMSs and STDs w.r.t. real-time kinematic (RTK) are improved and the proper Q_t is found out at the same time. According to our analysis despite the criticism that EM-PPP is very time-consuming because a large number of parameters are calculated and the first-order convergence of EM-algorithm, it is a numerically stable and simple approach to consider the temporal nature of state-space model of PPP, in particular when Q_t and R_t are not known well, its performance without fixing ambiguities can even parallel to traditional PPP-RTK.

Keywords: EM-algorithm; multi-GNSS; PPP; process noise; observation covariance matrix; extended Kalman filter; machine learning

1. Introduction

Since Precise Point Positioning (PPP) emerged [1,2], people are primarily focusing on improving precise orbit and clock products, developing new algorithms to solve for ambiguities, to accelerate its convergence and expand its applications such as PPP-real-time-kinematic (PPP-RTK), triple frequency PPP [3], ionosphere-constraint PPP and low-cost receiver PPP [4–10].

Generally, PPP can be realized by the least-squares method (including sequential least-squares) or extended Kalman filter. The least-squares method is for static state estimation and thus does not reflect varying user dynamics. To work the same as Kalman filter, the process noise matrix is added to the gain matrix of the sequential least-squares method to adjust receiver clock behavior and atmospheric activity and so on, which is named as a sequential filter [2]. Hence, in the following paper, the authors will only consider the Kalman filter for PPP data processing.

Although both the process noise Q_t and observation covariance matrix R_t are the key to Kalman filter, limited attention is paid to the fundamental problem for multi-GNSS PPP. Q_t and R_t must be consistent with state dynamics and measurement accuracy, respectively. For example, if the value of Q_t is too small, the estimated state will lose its minimum mean squared error property, and if the value of Q_t is too large with respect to the correct one, the estimated state will oscillate around the true value. Moreover, because of ground deformation and specific surroundings, Q_t should not be kept fixed to calculate the optimal estimates. In other words, Q_t should evolve with time and a proper Q_t will shorten the PPP convergence time. If Q_t is improper, it may damage PPP convergence sometimes.

As for GNSS, R_t not only depends on the measurement accuracy, elevation of the satellite, orbit and clock error, atmospheric delay error, multipath, missing data, etc. but possibly deteriorate after a lapse of a period. What is more, the assumption, frequently used in geodesy, that different types of measurements have a fixed error ratio is not always true, because the ratio is closely linked to receivers and antenna types, and to the performance of satellite system itself. For example, while fusing multiple GNSS, the weight of measurements of GPS is intended to be higher, other GNSS systems to be lower. It is not easy to give a prior accurate ratio.

Generally, four methods are often applicable to calibrate Q_t and R_t. The first one is based on the innovation property of the Kalman filter, in which a moving-window recursive way is used to identify Q_t and R_t [11–14]. However, none of them can maintain the positive semi-definiteness of the estimated covariances. To solve this problem, Odelson developed the autocovariance least-squares method for estimating covariances using a lagged autocovariance function [15]. This kind of least-squares method depends on the user-defined autocovariance function.

The second scheme to recognize Q_t and R_t is the multiple model adaptive estimation (MMAE) [13,16]. MMAE runs a bank of Kalman filter in parallel, every one of them is driven by its pair of Q_t and R_t. The final Q_t and R_t are thought of as the weighted sum of the estimates of individual Kalman filter.

In the third scheme, M-estimator is introduced into an adaptive Kalman filter to increase its resistance to outliers, where an adaptive factor α to state error covariance matrix is constructed [17,18]. Yet, choosing a value for α is still very challenging. An improper α will result in biased results.

Another attractive scheme is the least-squares variance component estimation (LS-VCE) [19], which is based on least-squares principles. Similar to restricted maximum likelihood (REML), LS-VCE does not use observations directly but combine observations to exclude any fixed effects. However, LS-VCE needs to define the weight matrix on the user‘s own and increase its complexity.

In this contribution, a machine learning algorithm, the Expectation-Maximization (EM) algorithm, is developed to the extended Kalman filter to estimate PPP states, $\vec{x}_t$, together with a large number of Q_t and R_t. The EM-algorithm, which can be classified as the first scheme, works in an iterative procedure to locate maximum likelihood estimates of parameters. Its iteration consists of two steps: Expectation and Maximization. In the Expectation step, a function for the expectation of the log-likelihood is computed using the estimates of the current parameters. In the Maximization step, estimates of parameters are updated by maximizing the expected log likelihood function.

On the one hand, the main drawback of the EM-algorithm is that it converges slowly and needs heavy computation. Here the convergence refers to finding maximum likelihood estimator of parameters, not the PPP convergence time. However, for example, its convergence can be accelerated using the Aitken method or conjugate method [20].

On the other hand, it is fairly simple, and has robust convergence and deals conveniently with problems having a lot of parameters. For such problems, it is often the only algorithm to a large extent [19]. It is also capable of finding Kalman parameters even if we have missing data. In addition, it can detect outliers by introducing small weights for large outliers and can even estimate the outliers [21]. In contrast, the outliers are not removed but automatically downweighed in our article, since outliers sometimes take some useful information.

The paper is organized as follows, after clarifying the importance to calibrate Q_t and R_t, the EM algorithm is introduced in the first section. Next, the state space model of PPP is reviewed from a point of machine learning and the methodology to adapt the EM-algorithm for extended Kalman filter for PPP is explained theoretically in detail. Thirdly, we compare the EM-algorithm with other methods, the static and kinematic instances are also given to demonstrate EM-PPP performance to improve the accuracy and reliability of PPP. Finally, the results are analyzed and the conclusion is drawn.

2. State Space Model for PPP

The state-space model allows to process GNSS data in a uniformed form. It is characterized as two equations: the state equation, which comprises a series of vector $\vec{x}_t$, $(1 \leq t \leq N$, N is the number of epochs), and the observation equation. The state $\vec{x}_t$ cannot be observed directly, usually called hidden states, which is driven by hidden process noise. In this article the state-space model is described as the Kalman filter.

2.1. State Equation

The hidden state $\vec{x}_t$ of multi-GNSS PPP Kalman filter involves five types of parameters: three components of receiver coordinates, receiver clock error, system time difference w.r.t. GPS, troposphere zenith wet delay and ambiguities. Using subscript t to denote a specific time epoch, the state at time t evolves from the state at $(t-1)$ according to:

$$\vec{x}_t = \Phi_t \vec{x}_{t-1} + \vec{\omega}_t \tag{1}$$

where Φ_t is the transition matrix and $\vec{\omega}_t$ the state process noise, which is assumed to be drawn from a zero-mean multivariate normal distribution, with covariance: $\vec{\omega}_t \sim \mathcal{N}(0, Q_t)$. Initial condition $\vec{x}_0$ is assumed to be a Gaussian vector with the a priori information $E\{\vec{x}_0\} = \vec{\mu}_0, Cov(\vec{x}_0) = P_0$.

The state transition matrix and the process noise matrix in static mode is defined for the position block:

$$\Phi_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_t, \; Q_t = \begin{bmatrix} Q_x & 0 & 0 \\ 0 & Q_y & 0 \\ 0 & 0 & Q_z \end{bmatrix}_t \tag{2}$$

where Q_x, Q_y and Q_z are the random process noise in X, Y and Z direction, respectively.

In addition to position parameters, the velocity parameters are also included in the state vector for our kinematic processing, whose system model for position and velocity block in the extended Kalman filter is given as:

$$\Phi_t = \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, Q_t = \begin{bmatrix} \frac{q_x \Delta t^3}{3} & \frac{q_x \Delta t^2}{2} & 0 & 0 & 0 & 0 \\ \frac{q_x \Delta t^2}{2} & q_x \Delta t & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{q_y \Delta t^3}{3} & \frac{q_y \Delta t^2}{2} & 0 & 0 \\ 0 & 0 & \frac{q_y \Delta t^2}{2} & q_y \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{q_z \Delta t^3}{3} & \frac{q_z \Delta t^2}{2} \\ 0 & 0 & 0 & 0 & \frac{q_z \Delta t^2}{2} & q_z \Delta t \end{bmatrix} \quad (3)$$

The corresponding state vector is $\vec{x}_t = \begin{bmatrix} x & v_x & y & v_y & z & v_z \end{bmatrix}$ for position and velocity block. The process noise matrix Q_t is uniquely determined by q_x, q_y and q_z, which are named as acceleration variance.

2.2. Observation Equation

The observation equation, that we used, is the double frequency ionosphere-free combination for multi-GNSS [22]:

$$PC_r^{i,S} = \rho_r^i + cdt_r^G + T_r^S + m_r^i \cdot ztd_r + \varepsilon_{r,PC}^{i,S} \quad (4)$$

$$LC_r^{i,S} = \rho_r^i + cdt_r^G + T_r^S + m_r^i \cdot ztd_r + \lambda_{LC}^S \cdot N_{r,LC}^S + \varepsilon_{r,LC}^{i,S} \quad (5)$$

where subscript r indicates the receiver, superscript i represents the satellites, superscript S indicates GNSS constellation, following the convention of Rinex3.x (G: GPS, E: GALILEO, R: GLONASS and C: BEIDOU). $PC_r^{i,S}$ and $LC_r^{i,S}$ are the ionosphere-free combinations of code pseudo-range and phase observations (unit: m) respectively, which have already corrected satellite clock, the relativity effect, solid Earth tides, polar tides, ocean tides, phase wind up and a priori troposphere delay using troposphere model [23,24]. ρ_r^i is the geometry distance between receiver and satellite, cdt_r^G is receiver clock (unit: m), the superscript G of cdt_r^G implies that GPS time is selected as the reference time, T_r^S is the system time difference in meters of system S to GPS time. Specifically, for $S = G$, T_r^G is zero. m_r^i is troposphere mapping function and ztd_r is troposphere zenith wet delay, λ_{LC}^S is the wavelength of LC combination corresponding to system S and $N_{r,LC}^S$ is the LC ambiguity. $\varepsilon_{r,PC}^{i,S}$ and $\varepsilon_{r,LC}^{i,S}$ indicate other unmodeled errors or noise.

Equations (4) and (5) are nonlinear, the extended Kalman filter (EKF) can be used for nonlinear state estimation. For easy description, they are rewritten in a general form:

$$\vec{y}_t = h(\vec{x}_t) + \vec{v}_t \quad (6)$$

where $\vec{y}_t = \begin{bmatrix} y_1 & \cdots & y_j & \ldots & y_{k_t} \end{bmatrix}$ is the observation vector, k_t is the number of observations at epoch t, $y_j \in \{PC_r^{i,S}, LC_r^{i,S}\}$, $\vec{v}_t$ is the observation noise satisfying $\vec{v}_t \sim \mathcal{N}(0, R_t)$, R_t is the observation noise covariance matrix at epoch t.

2.3. Kalman Filter

Let $Y_m = \{\vec{y}_1, \ldots, \vec{y}_m\}$ denote all observations from epoch 1 to epoch m, and $\vec{x}_{t|m}$ represent the estimate of $\vec{x}_t$ given observations Y_m, we have predicted state estimate and predicted covariance estimate:

$$\vec{x}_{t|t-1} = \Phi_t \vec{x}_{t-1|t-1} \quad (7)$$

$$P_{t|t-1} = \Phi_t P_{t-1|t-1} \Phi_t' + Q_t \quad (8)$$

After linearization of Equation (6) at predicted state $\vec{x}_{t|t-1}$,

$$\vec{e}_{t|t-1} \approx H_t(\vec{x}_t - \vec{x}_{t|t-1}) + \vec{v}_t \quad (9)$$

where $H_t = \frac{\partial \vec{y}}{\partial \vec{x}}\Big|_{\vec{x}_{t|t-1}}$, $\vec{e}_{t|t-1} = \vec{y}_t - h\left(\vec{x}_{t|t-1}\right)$. $\vec{e}_{t|t-1}$ is called innovations or measurement residuals, then the Kalman filter is obtained:

$$K_t = P_{t|t-1}H_t'\left(H_t P_{t|t-1} H_t' + R_t\right)^{-1} \tag{10}$$

$$\vec{x}_{t|t} = \vec{x}_{t|t-1} + K_t\,\vec{e}_{t|t-1} \tag{11}$$

$$P_{t|t} = (I - K_t H_t)P_{t|t-1} \tag{12}$$

where $\vec{x}_{t|t}$ and $P_{t|t}$ are the updated Kalman estimate and the updated covariance estimate.

2.4. Kalman Smoothing

The Kalman smoother estimator could be obtained [25]:

$$\vec{x}_{t-1|N} = \vec{x}_{t-1|t-1} + J_{t-1}\left\{\vec{x}_{t|N} - \vec{x}_{t|t-1}\right\} \tag{13}$$

$$P_{t-1|N} = P_{t-1|t-1} + J_{t-1}\left(P_{t|N} - P_{t|t-1}\right)J_{t-1}' \tag{14}$$

where $J_{t-1} = P_{t-1|t-1}\phi_t{}'\left[P_{t|t-1}\right]^{-1}$, $1 \le t \le N$, N is the number of epochs.

Kalman lag-one covariance holds with the starting condition

$$P_{N,N-1|N} = (I - K_N H_N)\Phi_N P_{N-1|N-1} \tag{15}$$

for $t = N,\ N-1,\ \ldots\ ,2$

$$P_{t-1,t-2|N} = P_{t-1|t-1}J_{t-2}' + J_{t-1}\left(P_{t,t-1|N} - \Phi_t P_{t-1|t-1}\right)J_{t-2}' \tag{16}$$

3. EM-PPP

The EM-algorithm is based on the innovation of the likelihood function to compute maximum likelihood estimation [25,26]. The likelihood function describes the probability of the observations, given a set of parameters. The parameters are found such that they maximize the likelihood function. The derivative of the likelihood function or log-likelihood is not always tractable. Therefore, iterative methods like Expectation-Maximization algorithms are very effective to find numerical solutions for the parameter estimates.

Denoting $\Theta = \{\vec{\mu}_0, P_0, Q_t, R_t \big| t = 1, \ldots, N\}$, $X = \{\vec{x}_0, \vec{x}_1, \ldots, \vec{x}_N\}$, $Y = \{\vec{y}_1, \ldots, \vec{y}_N\}$. Y is thought of as incomplete data, and $\{X,\ Y\}$ as complete data. Specifically for PPP, the log likelihood of the parameters of the state space model is approximately derived (ignoring constant):

$$\begin{aligned} 2\log L_{X,Y}(\Theta) = &-\log\left|\Sigma_0\right| - (\vec{x}_0 - \vec{\mu}_0)'\Sigma_0^{-1}(\vec{x}_0 - \vec{\mu}_0) \\ &- \sum_{t=1}^{N}\log|Q_t| - \sum_{t=1}^{N}(\vec{x}_t - \Phi_t\vec{x}_t)'Q_t^{-1}(\vec{x}_t - \Phi_t\vec{x}_t) \\ &- \sum_{t=1}^{N}\log|R_t| - \sum_{t=1}^{N}(\vec{y}_t - h(\vec{x}_{t|N}))'R_t^{-1}(\vec{y}_t - h(\vec{x}_{t|N})) \end{aligned} \tag{17}$$

Since the hidden states $\vec{x}_t$ are unknown, only the expected value of the log likelihood conditioned on Y is accessible, as a result, the observation equation is expanded at smoother point $\vec{x}_{t|N}$.

The Expectation (E-step) of EM algorithm for PPP requires computing the expected log-likelihood at the *jth* iteration:

$$\Omega(\,\Theta\big|\Theta^{(j-1)}\,) = E\{2\log L_{X,Y}(\Theta)\big|Y, \Theta^{(j-1)}\} \tag{18}$$

then the parameters are recalculated at the Maximization step (M-step):

$$\Theta^{(j)} = \underset{\Theta}{\operatorname{argmax}}\, \Omega\left(\Theta \middle| \Theta^{(j-1)}\right) \tag{19}$$

The two steps are repeated until the $\Theta^{(j)}$ converges.

The EM-PPP is terminated when the following convergence criterion is reached:

$$R\text{-log} = \left| \frac{\ell^{(j)} - \ell^{(j-1)}}{\ell^{(j)}} \right| < \varepsilon \text{ or } j \ \geq \textit{maximum number of iterations} \tag{20}$$

where ε is a small predefined amount and $\ell^{(j)}$ is equal to

$$\ell^{(j)} = \sum_{t=1}^{N} \log\left|H_t P_{t|t-1} H_t' + R_t\right| + \sum_{t=1}^{N} \left(\vec{e}_{t|t-1}\right)' \left(H_t P_{t|t-1} H_t' + R_t\right)^{-1} \left(\vec{e}_{t|t-1}\right) \tag{21}$$

A flowchart of our EM-PPP procedure is shown in Figure 1. In the initialization step, GNSS data preprocessing is performed including data integrity checking, cycle slips and outliers detection, phase center offset (PCO) and phase center variations (PCV) correction, synchronization of receiver clock using only code range measurements and initialization of parameters Θ. It also initializes the hidden state $\vec{x}_0$, sharing of the same processing noise Q_t across the time step $t = 1, \ldots, N$.

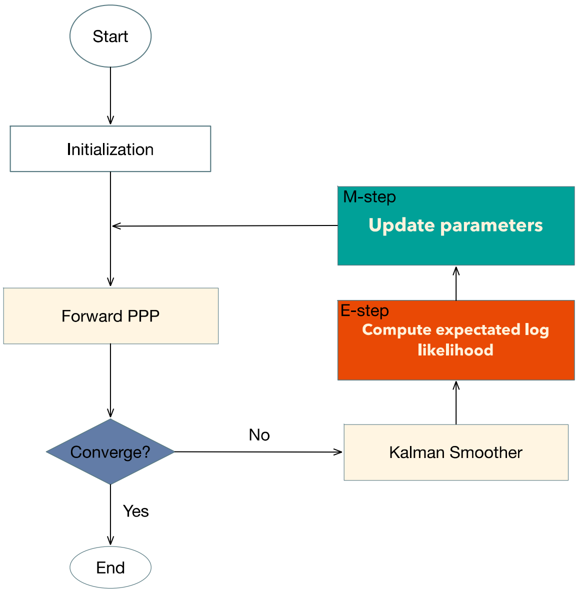

Figure 1. Flowchart of the Expectation-Maximization (EM)-Precise Point Positioning (PPP) loop.

In the next step, the extended Kalman filter Equations (10)–(12) are implemented to compute a series of hidden states and their covariance. If the convergence condition Equation (20) is not satisfied, then Kalman smoothing is used to calculate smoothed state estimates and involved covariance matrix Equations (13)–(16), which is prepared for the E-step: calculation of the expected log likelihood function Equation (22). All parameters of Θ are updated during the M-step and prepared for the next iteration Equation (27).

3.1. E-Step

Taking the expectation upon Equation (17) over conditional distribution of the latent given observed data, we find immediately:

$$\begin{aligned}\Omega\left(\Theta\middle|\Theta^{(j-1)}\right) &= E\left\{-2lnL_{X,Y}(\Theta)\middle|Y\right\}\\ &= ln|\Sigma_0| + tr[\Sigma_0^{-1}E\{(\vec{x}_0 - \vec{\mu}_0)(\vec{x}_0 - \vec{\mu}_0)'\Big|Y\}] + \sum_{t=1}^{N}\ln|Q_t|\\ &+ \sum_{t=1}^{N} tr[Q_t^{-1}E\{(\vec{x}_t - \Phi_t\vec{x}_{t-1})(\vec{x}_t - \Phi_t\vec{x}_{t-1})'\Big|Y\}] + \sum_{t=1}^{N}\ln|R_t|\\ &+ \sum_{t=1}^{N} tr\left[R_t^{-1}E\left\{\left(\vec{y}_t - h\left(\vec{x}_t\right)\right)\left(\vec{y}_t - h\left(\vec{x}_t\right)\right)'\middle|Y\right\}\right]\end{aligned} \tag{22}$$

where

$$E\{(\vec{x}_0 - \vec{\mu}_0)(\vec{x}_0 - \vec{\mu}_0)'|Y\} = P_{0|N} + (\vec{x}_{0|N} - \vec{\mu}_0)(\vec{x}_{0|N} - \vec{\mu}_0)' \tag{23}$$

$$\begin{aligned}E\{(\vec{x}_t - \Phi_t\vec{x}_{t-1}) \quad &(\vec{x}_t - \Phi_t\vec{x}_{t-1})'|Y\}\\ &= P_{t|N} + \vec{x}_{t|N}\vec{x}'_{t|N} + \Phi_t(P_{t-1|N} + \vec{x}_{t-1|N}\vec{x}'_{t-1|N})\Phi_t{}'\\ &-\Phi_t(P_{t,t-1|N} + \vec{x}_{t|N}\vec{x}'_{t-1|N})' - (P_{t,t-1|N} + \vec{x}_{t|N}\vec{x}_{t-1|N}{}')\Phi_t{}'\end{aligned} \tag{24}$$

Using Taylor series expression

$$\vec{y}_t \approx h\left(\vec{x}_{t|N}\right) + H_{t|N}\left(\vec{x}_t - \vec{x}_{t|N}\right) + \vec{v}_t, \tag{25}$$

where $H_{t|N} = \left.\frac{\partial\vec{y}}{\partial\vec{x}}\right|_{\vec{x}_{t|N}}$, and let $\vec{e}_{t|N} = \vec{y}_t - h\left(\vec{x}_{t|N}\right), \Delta\vec{x}_{t|N} = \vec{x}_t - \vec{x}_{t|N}$, we get

$$E\left\{\left(\vec{y}_t - h\left(\vec{x}_t\right)\right)\left(\vec{y}_t - h\left(\vec{x}_t\right)\right)'\middle|Y\right\} \approx \vec{e}_{t|N}\vec{e}'_{t|N} + H_{t|N}P_{t|N}H'_{t|N} \tag{26}$$

3.2. M-Step

Similar to complete-data weighted maximum likelihood estimation, from the first differential of $\Omega\left(\Theta\middle|\Theta^{(j-1)}\right)$, the maximum likelihood estimators are updated as follows:

$$\begin{aligned}\vec{\mu}_0 &= \vec{x}_{0|N}\\ \Sigma_0 &= P_{0|N}\\ Q_t &= E\{(\vec{x}_t - \Phi_t\vec{x}_{t-1})(\vec{x}_t - \Phi_t\vec{x}_{t-1})'|Y\}\\ Q_t &= E\{(\vec{x}_t - \Phi_t\vec{x}_{t-1})(\vec{x}_t - \Phi_t\vec{x}_{t-1})'|Y\}\end{aligned} \tag{27}$$

For simplicity, the initial covariance P_0, and the measurement covariance R_t are assumed to be a diagonal matrix:

$$\begin{aligned}P_0 &= diag\left(q_{01}, q_{02}, \ldots, q_{0k_0}\right)\\ R_t &= diag(r_{t1}, r_{t2}, \ldots, t_{tm_t})\end{aligned}$$

where k_0 indicates the dimension of the hidden state vector at initial epoch and m_t is the dimension of the observation vector at epoch t.

4. EM Compared to MLE and LS-VCE

In literature [19], a comprehensive comparison is demonstrated between different estimation principles such as LS-VCE, best linear unbiased estimator (BLUE), best invariant quadratic unbiased estimator (BIQUE), minimum norm quadratic unbiased estimator (MINQUE) and restricted maximum likelihood estimator (REML). As shown previously in Section 3.1, the EM algorithm may be thought of as maximum likelihood estimation (MLE), but which finds the ML estimator in an iterative way.

EM can be realized based on REML as well. Therefore, an additional comparison between EM variance estimation and LS-VCE and MLE is adequate.

To make theoretical analysis easy and consistent, in the following we first introduce how to covert Kalman filter to least-squares. Then we directly give different (co)variance estimators according to their distribution assumptions and the reason why the EM-algorithm is preferable in our solution.

4.1. From Kalman Filter to Least Squares

The linear (extended) Kalman equation can be transformed into the least-squares function model, which allows the following EM algorithm to be compared with LS-VCE on the same function model and makes the theoretical analysis easy and convenient. To do so, the state Equation (1) and the observation Equation (6) are expanded at a priori value and organized as the function model:

$$\begin{gathered}\vec{y} = A\vec{x} + \vec{w} \\ E\{\vec{w}\} = 0,\ D\{\vec{w}\} = E\{\vec{w}\vec{w}'\} = Q = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k\end{gathered} \tag{28}$$

with

$$\begin{gathered}\vec{y} = \begin{bmatrix} 0 & d\vec{y}_1{}' & 0 & \cdots & 0 & d\vec{y}_N{}' \end{bmatrix}' \\ A = \begin{bmatrix} \Phi_1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & H_1 & 0 & \cdots & 0 & 0 \\ 0 & \Phi_2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \Phi_N & -1 \\ 0 & 0 & 0 & \cdots & 0 & H_N \end{bmatrix} \\ \vec{x} = \begin{bmatrix} d\vec{x}_0{}' & d\vec{x}_1{}' & d\vec{x}_2{}' & \cdots & d\vec{x}_{N-1}{}' & d\vec{x}_N{}' \end{bmatrix}'\end{gathered}$$

where the a priori $\vec{x}_i^0, (i = 0, \ldots, N)$ value is subtracted from the original state vector $\vec{x}_i$, leading to $d\vec{x}_i = \vec{x}_i - \vec{x}_i^0$ and $d\vec{y}_i = \vec{y}_i - \vec{y}_i^0$, N is the number of epochs. The $m \times n$ matrix A is full column rank. The cofactor matrices Q_k, are assumed to be known and their weighted sum $Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$ is assumed to be positive definite and $Q_k, (k = 1, \ldots p)$ are linearly independent, and the (co)variance components σ_k are unknown parameters. Matrix Q_0 is the known part of the variance matrix [19].

4.2. Least-Squares EM

Similar to Section 3, we can calculate the non-constant part of the full log-likelihood function and then conditional expectation on the observation $\vec{y}$ and $Q^{(j)}$, given the data $\vec{y}$ and the *jth* iteration estimates of (co)variance components $\sigma_k^{(j)}$ or $Q^{(j)}$:

$$\begin{aligned}\Omega(Q|\vec{y}, Q^{(j)}) &= E\{L|\vec{y}, Q^{(j)}\} = \log Q + E\{tr((\vec{y} - A\vec{x})' \mathrm{Q}^{-1} (\vec{y} - A\vec{x}))|\vec{y}\} \\ &= \log Q + tr(\mathrm{Q}^{-1} E\{(\vec{y} - A\vec{x})(\vec{y} - A\vec{x})' |\vec{y}\})\end{aligned} \tag{29}$$

where

$$\begin{gathered}E\{(\vec{y} - A\vec{x})(\vec{y} - A\vec{x})' |\vec{y}\} = (\vec{y} - A\hat{x})(\vec{y} - A\hat{x})' + AQ_{\hat{x}}^{(j)} A' Z_{ML}^{(j)} \\ \hat{x} = (\mathrm{A}'(\mathrm{Q}^{(j)})^{-1} A)^{-1} A'(\mathrm{Q}^{(j)})^{-1} \vec{y}\end{gathered} \tag{30}$$

M-step:

Maximizing the likelihood of the completed data based on Equation (29), the new estimates $\sigma_k^{(j+1)}$ are calculated as

$$\begin{bmatrix} \sigma_1^{(j+1)} \\ \vdots \\ \sigma_p^{(j+1)} \end{bmatrix} = \left(A'_{vEM} W_{vEM} A_{vEM}\right)^{-1} A'_{vEM} W_{vEM}\ vec\left(Z_{EM}^{(j)} - Q_0\right) \tag{31}$$

where $A_{vEM} = \begin{bmatrix} vec\left(Q_1^{(j)}\right) & \cdots & vec\left(Q_p^{(j)}\right) \end{bmatrix}$, $W_{vEM} = \left(\mathrm{Q}^{(j)}\right)^{-1} \otimes \left(\mathrm{Q}^{(j)}\right)^{-1}$, $\otimes$ is the Kronecker product and vec is *vec*-operator.

Equations (29) and (31) are the EM algorithm for ML estimation. If convergence is reached, set $\sigma_k = \sigma_k^{(j)}$, otherwise increase j by one and return to E-step.

4.3. MLE

Once the general structure of probability density function is known, MLE can be simply realized and therefore used widely. If a multivariate normal distribution is given, the (co)variance components takes form:

$$\begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_p \end{bmatrix} = \left(A'_{vMLE} W_{vMLE} A_{vMLE}\right)^{-1} A'_{vMLE} W_{vMLE}\ vec(Z_{MLE} - Q_0) \tag{32}$$

where $A_{vMLE} = \begin{bmatrix} vecQ_1 & \cdots & vecQ_p \end{bmatrix}$, $W_{vMLE} = Q^{-1} \otimes Q^{-1}$.

From Equations (31) and (32), we know that least-squares EM and MLE estimators share the same design matrix and weight matrix. Their difference is mainly caused by the pseudo observation $vec\left(Z_{EM}^{(j)} - Q_0\right)$ and $vec(Z_{MLE} - Q_0)$. $Z_{EM}^{(j)}$ for EM-algorithm includes the effects of both observation post-residuals and the accuracy of the estimates $\hat{x}$. In contrast, Z_{MLE} for MLE consider only the observation post-residuals. Therefore, MLE estimator is probably over-optimistic to EM.

However, if the REML principle is used to derive the EM-algorithm, the effect of $\hat{x}$ is implicitly removed. Then, the EM-algorithm based on REML will be equivalent to the REML estimator.

4.4. LS-VCE

Another important problem is that the EM-algorithm and MLE do not take the loss of degrees of freedom from the estimation of $\vec{x}$ into account. Borrowing the idea of REML, LS-VCE overcomes this problem based on $(n-p)$ independently error contrasts. Specifically, let

$$\hat{t} = B'\vec{y}, Q_{\hat{t}} = B'QB \tag{33}$$

where $\vec{t}$ is misclosure vector, B is $m \times (m-n)$ matrix satisfying $B'A = 0$, $rank(B) = m - n = b$. Then LS-VCE estimator is obtained:

$$\begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_p \end{bmatrix} = \left(A'_{vLS-VCE} W_{vLS-VCE} A_{vLS-VCE}\right)^{-1} A'_{vLS-VCE} W_{vLS-VCE}\ vec(Z_{LS-VCE} - Q_0) \tag{34}$$

with $A_{vLS-VCE} = \begin{bmatrix} vec(B'Q_1B) & \cdots & vec\left(B'Q_pB\right) \end{bmatrix}$, $Z_{LS-VCE} = \hat{t}\hat{t}'$, $W_{vLS-VCE}$ is a user-defined weight matrix. If $W_{vLS-VCE}$ is set to $Q_{\hat{t}} \otimes Q_{\hat{t}}$, LS-VCE is the same as EM-algorithm based on REML.

LS-VCE is derived purely based on the least-squares method and we do not make any assumption on a probability density function (PDF). In contrast, EM-algorithm, MLE and REML are built upon a certain distribution, which explains why when applying LS-VCE, it is necessary for users to set weight matrix on their own.

4.5. Preference for Recursive EM

As discussed previously, the EM-algorithm can be implemented as either recursive form or batch form like MLE and LS-VCE. In our solution, we prefer the EM-algorithm based Kalman filter to other methods.

Recursive EM discriminates between the process noise and the observation noise. For GNSS, the process noise is usually different from observation noise. The process noise is directly connected to the geophysical phenomenon, which has not only linear but also non-linear variations, and suffers both time and spatial correlations [27–29]. As a result, it is relatively more difficult to estimate the process noise than the observation covariance matrix. Other batch methods mix the two types of stochastic processes with different behavior, which will bring us extra trouble.

Recursive EM can process data in not only post mode, but also in real-time mode. In both modes, data is processed epoch by epoch, allowing us to dynamically adjust the weight matrix, monitoring time-varying behavior and detecting abrupt motion.

5. Validation

5.1. Static PPP Scheme

A total of 14 IGS multi-GNSS stations are selected to assess the performance of EM-PPP (Figure 2). Those stations are evenly distributed on the Earth and track as many GNSS constellations as possible, covering not only the continent but also coastal and polar regions.

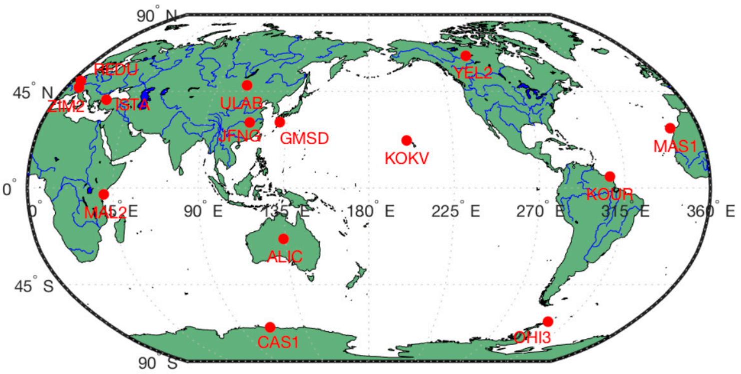

Figure 2. Distribution of selected IGS multi-GNSS stations.

Daily GNSS measurements from those IGS stations observed during DOY 119 in 2017 are used in this study. The true coordinate benchmarks are from IGS weekly solutions. The final GFZ Beidou multi-GNSS (GBM) products of the satellite orbits and clocks are applied (ftp://cddis.gsfc.nasa.gov/pub/gps/products/mgex/). The used precise orbit correction has a sampling interval of 15 min while the precise clock has a sampling interval of 30 s, both generated at GFZ. For GPS, we aligned C1 to P1 using CODE differential code bias (DCB) product (ftp://ftp.aiub.unibe.ch/CODE/2017/). The frequency bands we used are L1 and L2 for GPS, G1 and G2 for Glonass, E1 and E5a for Galileo, B1 and B2 for BeiDou [9]. Receiver and satellite PCO and PCV were corrected using igs14.atx. solid Earth tides, pole tide and ocean tides are removed according to IERS Conventions 2010. For troposphere delay estimation, the zenith dry component of tropospheric delays was corrected with the a priori Niell model [24]. The zenith wet delay (ZWD) is estimated as an unknown parameter. Then, 24-h observation data sets with a sampling interval of 30 s were processed for all stations. The elevation cut-off was set to 6 degrees.

The initial guess of receiver coordinates is intentionally deviated by 100 m from IGS station solution. A priori standard deviation (std) of PC is set to 3 m, a priori std of LC to 0.03 m for pseudo-range combination (PC) and carrier phase combination (LC) combination, respectively.

The starting values for Q_t are shown in Table 1. Random walk noise process with a spectral density equal to 1.0 mm/ $\sqrt{30}$ s is imposed on coordinates, which means a 1.0 mm disturbance very 30 s for IGS station in North, East and Up. It is not true in reality of course, but useful for test purposes. The receiver clock offset is supposed to be white noise. Zenith wet delay (ZWD) and inter-system bias (ISB) are modeled as random walk noise. Ambiguities can be considered as constant or random walk noise with very small spectral density.

Table 1. Initial guess Q_t for static PPP: ZWD is the zenith wet delay of the troposphere, ISB is the system time difference with respect to GPS time.

	X	Y	Z	cdt_r	ZWD	ISB	Ambiguities
Q_t	$\frac{1.0\ mm}{30\ s}$	$\frac{1.0\ mm}{30\ s}$	$\frac{1.0\ mm}{30\ s}$	$9.0 \times 10^{10}\ m$	$\frac{1.0\ cm}{\sqrt{h}}$	$\frac{3.0{\times}10^{-6} m}{\sqrt{s}}$	0.0

If the maximum number of iteration is reached and EM-PPP does not converge, smaller spectral density should be assigned for X, Y and Z for the next cycle.

5.2. Kinematic PPP Scheme

Another kinematic dataset was used to further validate the performance of the EM-PPP. The data was collected at Wuhan, China, in November 14, 2013. The sampling interval is 1 Hz and the observed time span is about one hour. The final CODE precise satellite orbit and 5 s clock products are used to estimate the 1 Hz GPS displacements. The ambiguity-fixed double differenced real-time kinematic (RTK) solutions are adopted as the reference to assess the performance of kinematic EM-PPP solution.

The initial acceleration variance is assumed to be $\frac{10\ \mathrm{m}^2}{\mathrm{s}^3}$ for position and velocity states (Table 2), which can be used to calculate the process noise matrix for position and velocity. The initial Q_t for receiver clock is modeled as white noise and estimated epoch-wisely. ZWD and ambiguities are also modeled as random walk processes with initial spectral density 0.01 cm/ $\sqrt{s}$ and 0.0 m/ $\sqrt{s}$, respectively.

Table 2. Initial guess Q_t for kinematic PPP: ZWD is zenith wet delay of troposphere.

	q_x	q_y	q_z	cdt_r	ZWD	Ambiguities
Q_t	$\frac{10\ m^2}{s^3}$	$\frac{10\ m^2}{s^3}$	$\frac{10\ m^2}{s^3}$	$(100\ m)^2$	0.01 cm/ $\sqrt{s}$	0.0

6. Results and Discussion

6.1. Static EM-PPP Solution

It is found that EM-PPP usually converges after the iteration counter reaches 50. The positioning errors are shown in Table 3, including PPP results at 1st iteration with the biased stochastic model, and the results after 50 iterations of calibration to assess EM-PPP performance. EM-PPP convergence in our research means that the square root of 3D positioning errors of the last 20 consecutive epochs is less than 10 cm.

Table 3 indicates that when the biased Q_t and R_t are fed in the beginning, PPP 3D errors are up to decimeters for a few stations. Horizontal errors are often greater than vertical errors, which is not consistent with the property of GNSS, because of inappropriate process noise matrix and measurement noise covariance matrix. After 50 iterations, the position errors are reduced to within 1 cm in North, East and Up direction on average using our EM-PPP algorithm. The mean 3D error is reduced to 1.77 cm without fixing ambiguities. The overall decrease percentage on average is 66.91%, 66.16%, 71.60% in North, East and Up direction, respectively, 69.95% for 3D errors.

Table 3. Statistics of EM-PPP absolute positioning errors (cm) with respect to IGS stations solution. All sites except YEL2 are processed using multiple GNSS observations. YEL2 is processed using only GPS data to verify the algorithm for single GNSS constellation. The ambiguities are not fixed.

No.	Site	1st Iteration				50th Iteration			
		North	East	Up	3D	North	East	Up	3D
1	ALIC	4.32	12.17	4.38	13.64	0.89	1.47	0.24	1.74
2	CAS1	2.49	2.23	2.97	4.48	1.58	0.80	0.91	2.00
3	GMSD	0.29	0.13	2.38	2.41	0.22	0.59	1.22	1.38
4	ISTA	2.11	1.44	1.02	2.75	2.11	0.07	0.95	2.33
5	JFNG	3.32	3.65	5.48	7.39	0.75	0.52	0.64	1.12
6	KOKV	0.88	4.97	3.43	6.11	0.54	2.13	3.20	3.89
7	KOUR	0.15	2.38	2.27	3.30	1.06	0.76	0.39	1.37
8	MAL2	1.21	2.99	5.67	6.53	1.23	1.50	0.98	2.18
9	MAS1	3.38	1.15	1.67	3.94	1.36	1.22	1.00	2.09
10	OHI3	16.47	2.94	4.83	17.42	1.13	0.44	0.55	1.34
11	REDU	0.76	0.39	5.19	5.25	0.56	0.59	0.40	0.91
12	ULAB	1.47	0.04	0.19	1.48	0.71	0.60	0.30	0.98
13	YEL2	0.61	2.17	2.72	3.54	0.34	0.93	0.12	1.00
14	ZIM2	0.15	0.49	4.15	4.19	0.03	1.02	2.25	2.47
average		2.69	2.66	3.31	5.89	0.89	0.90	0.94	1.77

To see what happened to the process noise Q_t and the observation covariance matrix R_t before and after calibration, an example of JFNG station located in China is illustrated.

The residuals for PC and the corresponding formal errors are shown in Figure 3. The residuals for LC and the corresponding formal errors against satellite elevation angles after calibration are shown in Figure 4. To be clear, PC and LC residuals for BeiDou are plotted separately from those for GPS, Glonass and Galileo.

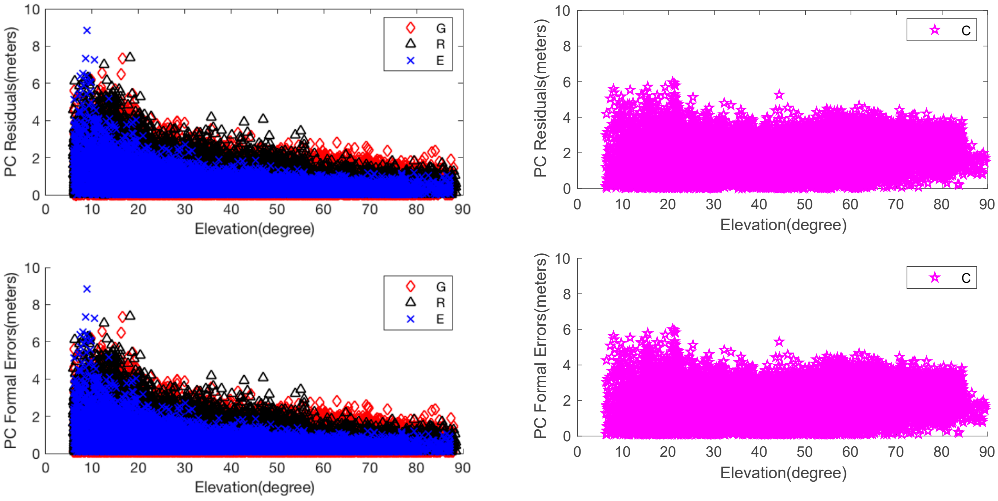

Figure 3. EM-PPP absolute pseudo-range combination (PC) residuals and formal errors of JFNG station at 50th iteration: Left two pictures are the PC residuals and their formal errors (square root of observation matrix R_t) of GPS, Glonass and Galileo, respectively. Similarly, the right two pictures show the PC residuals and formal errors of BeiDou.

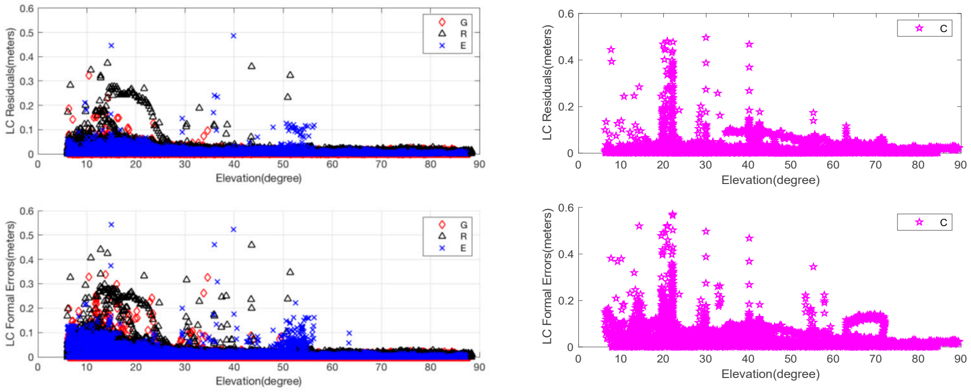

Figure 4. EM-PPP carrier phase combination (LC) residuals and formal errors of JFNG station at 50th iteration: Left two pictures are the LC residuals and formal errors (square root of the observation covariance matrix) for GPS, Glonass and Galileo, respectively. Similarly, the right two pictures show the LC residuals and formal errors for BeiDou.

The results allow us to examine the relationship between R_t and the residuals. It is clearly shown that formal errors of R_t are changed from 3 m to vary about between 0 and 10 m for PC, and from 0.03 m to vary about between 0 and 0.6 m for LC. The big outliers of LC are due to the non-convergence at the initial epoch and EM-PPP does not fully converge at some epochs. Although R_t variations, similar to the residuals-varying pattern, are dependent on the satellite elevation angle, LC formal error of BeiDou R_t at the elevation of about 20 degrees is greater than the lower degree formal error (refer to the bottom right picture in Figure 4). Clearly, it is not advisable to choose observation weight only according to the satellite elevation angle. Another example is station ISTA, where Glonass R_t for PC at satellite elevation higher than 50 degrees is almost as great as lower degree errors (not plotted here). In addition, random or systematic outliers are downweighed accordingly for both PC and LC.

It can also be observed that BeiDou PC residuals peak are almost 6 m, worse than GPS, Glonass and Galileo at JFNG. In fact, the biggest error source comes from GEO C05 and IGSO C06, C07 and C08, probably due to their poor orbit accuracy and clock offset, because the residuals of those satellites show less independence of satellite elevation angle. R_t is different among GPS, Glonass and Galileo.

Thus R_t not only reflects the accuracy of measurement type itself, elevation-dependent, characteristics of GNSS constellation, but also the quality of GNSS satellite orbit and clock products, and should be adjusted dynamically. Consequently, EM-PPP is effective to calibrate R_t automatically and suppress outliers.

In general, Q_t can be adjusted to its correct value, which is zero in our case, after less than ten iterations, as shown in Figure 5. The initial square root of Q_t in North, East and Up directions are 0.001 m/$\sqrt{30}$ s. Q_t becomes zero at the sixth iteration in all three components and position solution at the last epoch varies little.

It has to be pointed out that EM-PPP suffers local extrema problems like other alternative methods. It is not an algorithm to locate the global maxima, therefore EM-PPP is sensitive to initial guesses. To escape from this local extrema trap, several sets of initialization schemes can be used, and select the best one selected as the final result.

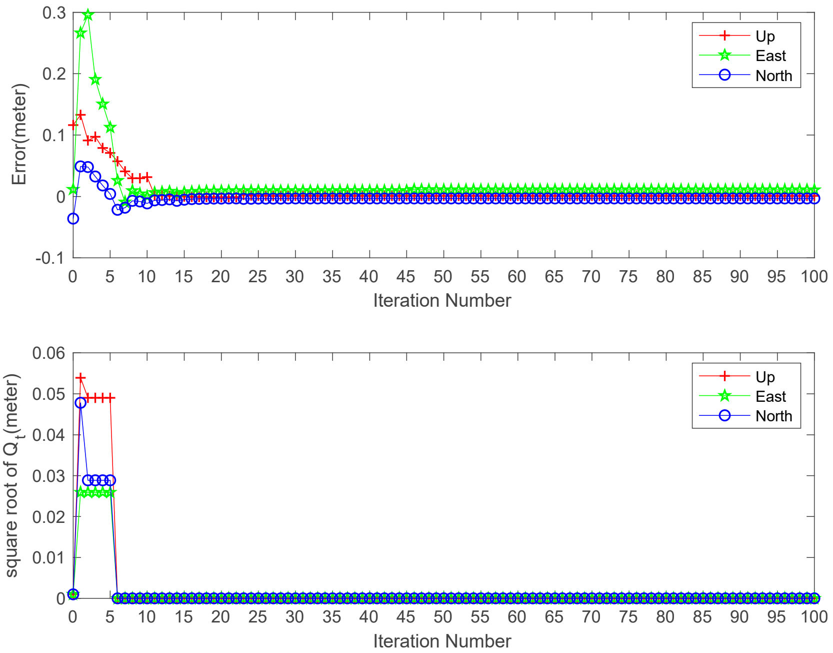

Figure 5. Effects of EM-PPP iteration number on the position errors (top) and variations of the square root of Q_t at last epoch on station JFNG.

6.2. Kinematic EM-PPP Solution

Figure 6 is the tracking route recovered by the EM-PPP at the 500th iteration. The relative log-likelihood versus iterations is plotted in Figure 7, where the results at the first iteration correspond to the solution to the traditional PPP (TPPP). The relative log likelihood decreases gradually from 1.0 to 8.665×10^{-7} and forms a concave curve. The less the relative log-likelihood, the less perturbative the EM-PPP solution becomes with respect to the RTK solution. As the iterative number grows to about 150, the EM-PPP solution converges nearly completely.

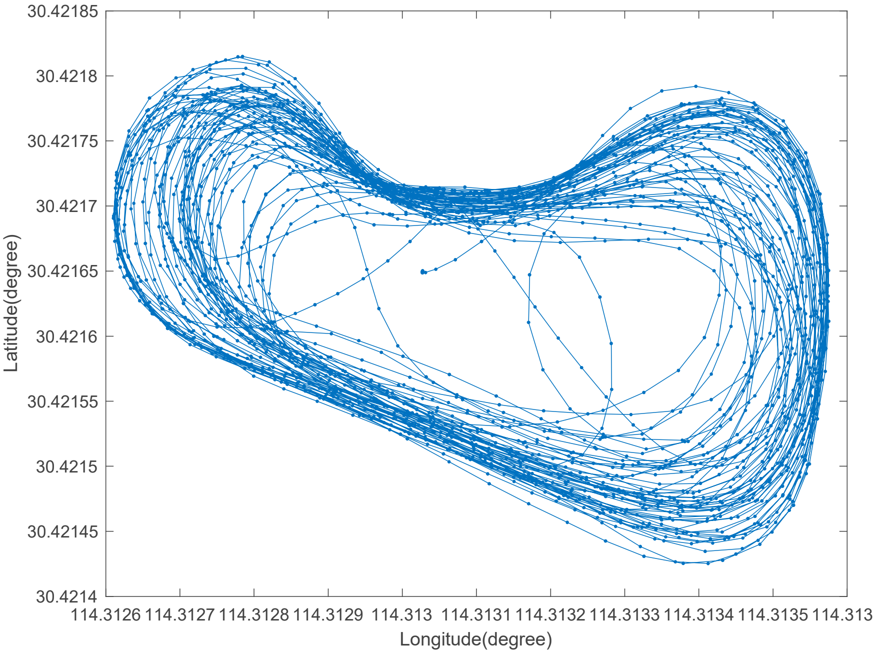

Figure 6. Driving route of moving vehicle in Wuhan, China on 14 November 2013.

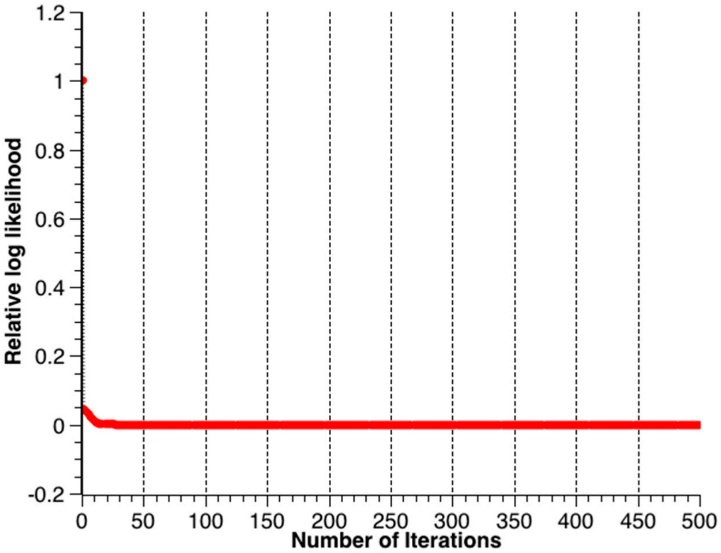

Figure 7. Relative log-likelihood of two consecutive iterations.

Illustrated in Figure 8 are the positioning errors of the traditional PPP solution and EM-PPP solution at the 500th iteration with respect to the reference coordinates in the Up, East and North directions. It is observed that EM-PPP solutions are much more stable in all directions when compared to TPPP solutions. The TPPP solution in the East changes wobbly in comparison to the North and Up due to greater acceleration in the East (Figure 9), which leads to a larger bias in the East.

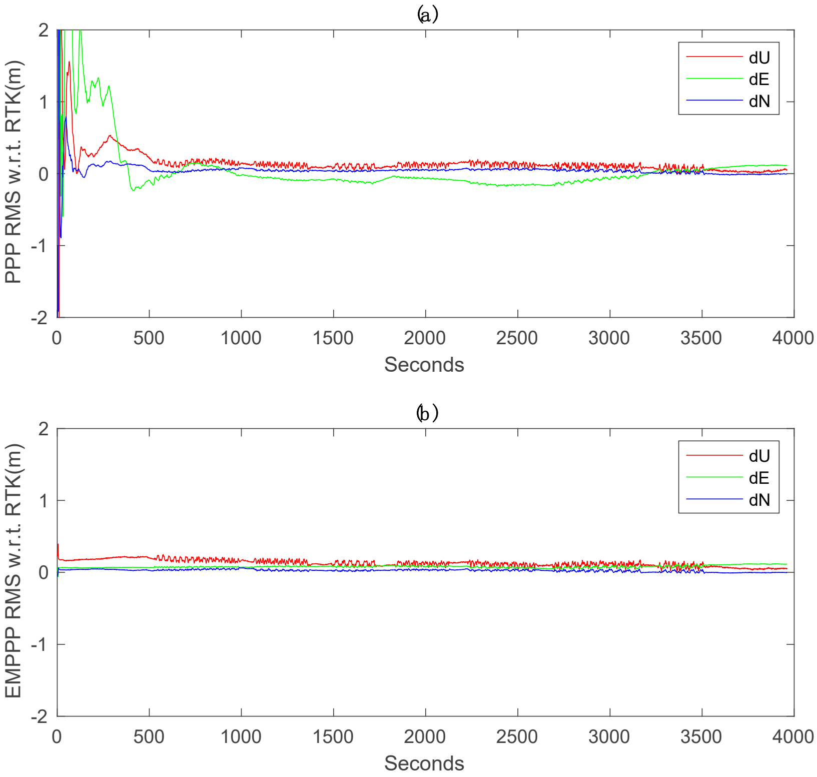

Figure 8. GPS-only kinematic positioning errors with respect to real-time kinematic (RTK): (**a**) traditional PPP errors, (**b**) EM-PPP positioning errors at the 500th iteration. Traditional PPP and EM-PPP share the same initial conditions.

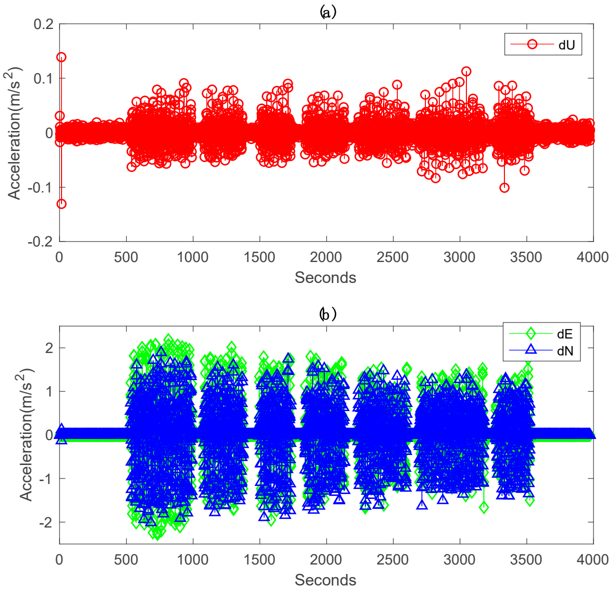

Figure 9. Acceleration time series derived from second-order differencing RTK position time series in Up, East and North: (**a**) acceleration in Up. (**b**) Accelerations in the East and North.

In fact, the position and velocity (PV) model of our kinematic state equation is assumed to be a constant velocity model, which means the acceleration is zero. However, the realistic acceleration of the vehicle is no zero, which results in systematic bias and unstable solution.

Figures 10 and 11 describe the EM-PPP RMSs and STDs with respect to RTK against the number of iterations, respectively. Obviously, after about 40 iterations, RMSs in Up and North are decreasing and the solutions are improved in those two directions. In contrast to Up and North direction, in the East direction it presents a decrease from 9.7 cm falling to 7.6 cm and then increases up to 8.5 cm for the RMS. However, the combined effects of all three directions get decreasingly 3D RMS, proving that EM-PPP does improve the positioning accuracy in the kinematic mode in our case, and apparently converges with increasing iterations, consistent with the EM theory, though there is a little disturbance because of the existing system bias and outliers of pseudo-range. It can be imagined that a better result can be expected if acceleration observations are also observed.

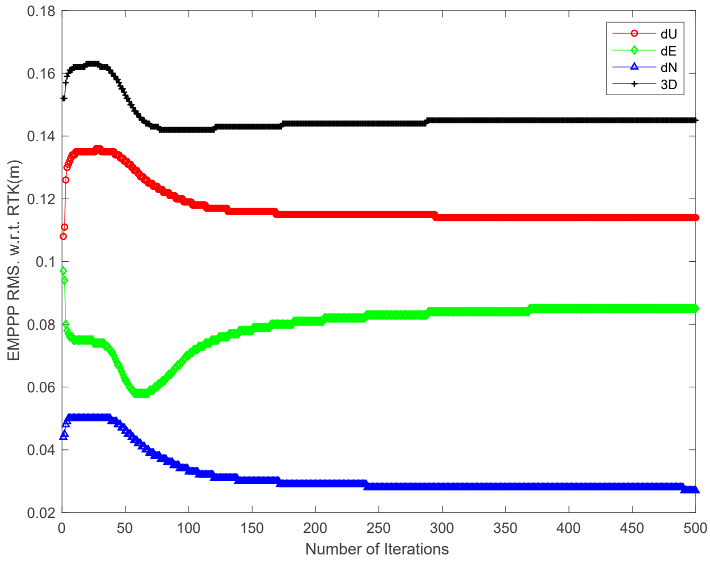

Figure 10. EM-PPP RMSs w.r.t. RTK solution in Up, East and North (unit: m).

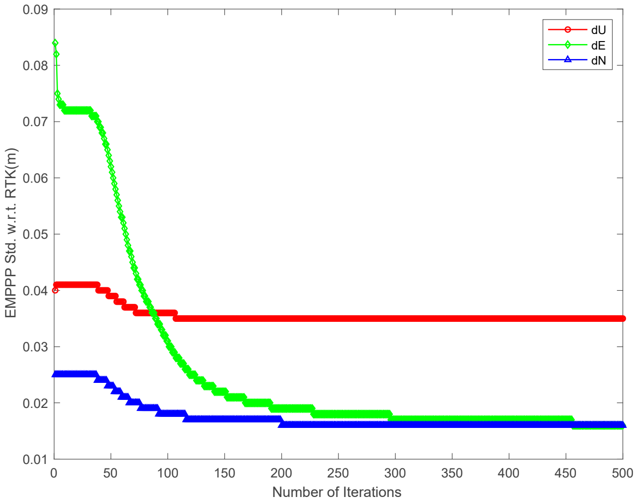

Figure 11. EM-PPP STDs w.r.t. RTK solution in Up, East and North (unit: m).

As for STDs with respect to RTK, they are consistently increasing in all three directions when EM-PPP continues its iteration. The STDs decrease from 4.0 cm, 8.4 cm and 2.5 cm to 3.5 cm, 1.6 cm and 1.6 cm in Up, East and North components, respectively. In other words, the STDs are improved by 12.5%, 80.9% and 36.0% in Up, East and North, accordingly.

Given Figure 12 is the estimates of geodetic coordinates B, L and H, and their estimates of the square root of Q_t after the 500th iteration. It is noted that the coordinates simultaneously stay stable or alternatively change sharply in all three directions, telling that the moving patterns of the vehicle switch between static and kinematic status from time to time. Theoretically, the process Q_t should change between zero and positive values. Obviously, the estimates of the square root Q_t, displayed in Figure 12b–f agree well with variations of coordinates. Time-varying Q_t is identified, which takes the concrete dynamic mode into consideration.

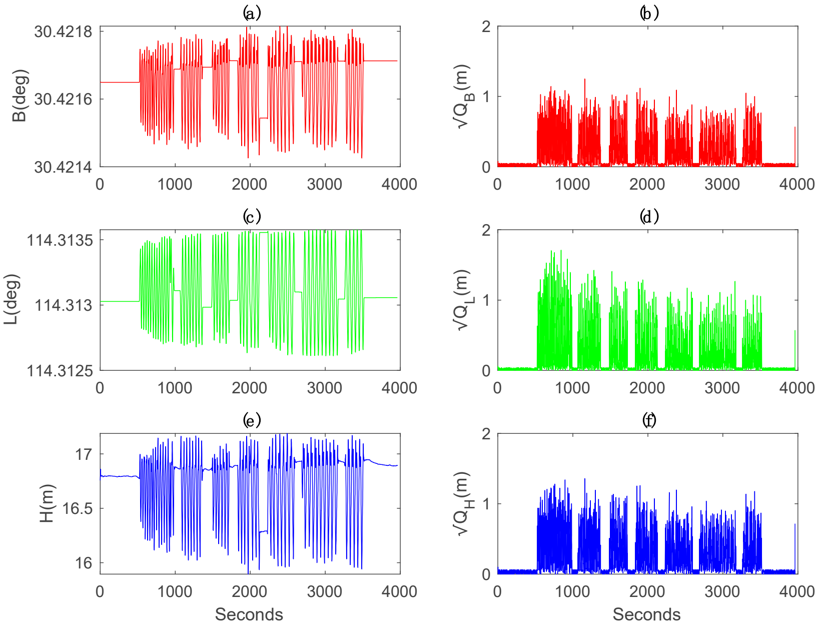

Figure 12. EM-PPP estimates of the Geodetic coordinates B, L, H and their correspondingly calibrated Q_t after the 500th iteration: (**a**) Latitude estimates B (degree). (**b**) The square root of Q_t of B (m). (**c**) Longitude estimates L (degree). (**d**) The square root of Q_t of L (m). (**e**) Height estimates H (m). (**f**) The square root of Q_t of H (m).

If much smaller initial acceleration variances are given, for example $\frac{0.01\,m^2}{s^3}$ rather than $\frac{10\,m^2}{s^3}$ for position and velocity states, it can be shown that our algorithm can still recover the process noise matrix as well and makes no difference. As a result, the user can choose values for Q_t randomly to a large extent.

7. Conclusions

A machine learning algorithm, EM algorithm, is adapted particularly to the extended Kalman filter to calibrate the process noise matrix and observation covariance noise matrix for PPP. The main advantage of EM-PPP is in fact that it is straightforward, simple, (locally) optimal and able to estimate large amounts of parameters and thus competent in calibrating the time-varying process noise and observation covariance for PPP state-space model, though its execution is time-consuming. The basic framework of EM-PPP is not limited to multi-PPP and can be applied to other fields of geodesy.

The whole procedure of EM-PPP is comprised of three parts: initialization, feedforward and backpropagation. In the beginning, the GNSS preprocess is performed to check the availability of the required data and mainly recognize cycle slips. Next, the whole process iterates between the estimation of hidden state and expectation and maximization.

The EM-algorithm is then compared with MLE and LS-VCE methods. We choose the recursive algorithm because it is superior to separate the process noise and observation variance, and to monitor time-varying behavior.

The approach was verified by selecting a global distribution of 14 IGS multi-GNSS station without fixing ambiguities. Based on the presented results, it concluded that EM-PPP is well suited for dynamically determining the time-varying process noise and observation noise. The calibrated observation variance matches the observation residuals from low satellite elevation angle to high satellite elevation angle. It resists orbit and clock errors and outliers through downweighing abnormal observations at different epochs, which is an alternative reasonable solution in contrast to the popular way that assigns weight according to the satellite elevation angle. People do not need to worry about separating observations into different categories carefully based on different GNSS constellations to estimate variance components like variance component estimation (VCE).

The spectral density of the assumed kinematic IGS station with 1 mm disturbance every 30 s in North, East and Up direction was estimated to be zero, implying that stations are static, which is consistent with reality.

An additional kinematic test was also implemented and reasonable values of Q_t are found when biased initial Q_t guess was given. The position errors are reduced in Up, East and North direction, respectively, w.r.t. RTK. In particular the STDs with respect to RTK are improved by 12.5%, 80.9% and 36.0% in Up, East and North, further showing that EM-PPP is also beneficial to kinematic PPP.

It has been confirmed that the EM-PPP is competitive for the calibration of the PPP stochastic model dynamically. The main drawback of this approach is that it converges slowly due to its first-order convergence. In the future, online EM-PPP may be derived to process GNSS data in real-time to overcome this problem if a large number of observations are available.

Author Contributions: X.Z. conceived and designed the algorithm; P.L. provided the kinematic GNSS data and helped validate the algorithm; M.G. supervised the whole procedure, and continuously discussed and analyzed the results and gave constructive suggestions; R.T. and X.L. participated in the experimental investigation; H.S. helped edit and revise the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the China Scholarship Council (CSC).

Acknowledgments: X.Z. is financially supported by the China Scholarship Council (CSC) for his study at the German Research Centre for Geosciences (GFZ). We thank the IGS and GFZ for providing GNSS observations, DCBs, precise orbit, and clock products. Our research is also partly supported by the Chinese Academy of Sciences (CAS) program of "Light of West China" (Grant No. 29Y607YR000103), Chinese Academy of Sciences, Russia and Ukraine and other countries of special funds for scientific and technological cooperation (Grant No. 2BY711HZ000101). This work is also partly supported by the National Natural Science Foundation of China (Grant No: 41674034, 41974032, 11903040), the Chinese Academy of Sciences (CAS) programs of "Pioneer Hundred

Talents" (Grant No: Y923YC1701), and the Chinese Academy of Sciences (CAS) program of "Western Youth Scholar" (Grant No: Y712YR4701, Y916YRa701), as well as "The Frontier Science Research Project" (Grant No: QYZDB-SSW-DQC028). We also thank the reviewers for their careful review.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zumberge, J.F.; Heflin, M.B.; Jefferson, D.C.; Watkins, M.M.; Webb, F.H. Precise point positioning for the efficient and robust analysis of GPS data from large networks. *J. Geophys. Res. Solid Earth* **1997**, *102*, 5005–5017. [CrossRef]
2. Kouba, J.; Héroux, P. Precise point positioning using IGS orbit and clock products. *GPS Solut.* **2001**, *5*, 12–28. [CrossRef]
3. Li, P.; Zhang, X.; Ge, M.; Schuh, H. Three-frequency BDS precise point positioning ambiguity resolution based on raw observables. *J. Geod.* **2018**, *92*, 1357–1369. [CrossRef]
4. Ge, M.; Gendt, G.; Rothacher, M.A.; Shi, C.; Liu, J. Resolution of GPS carrier-phase ambiguities in precise point positioning (PPP) with daily observations. *J. Geod.* **2008**, *82*, 389–399. [CrossRef]
5. Laurichesse, D.; Mercier, F.; Berthias, J.; Broca, P.; Cerri, L. Integer ambiguity resolution on undifferenced GPS phase measurements and its application to PPP and satellite precise orbit determination. *Navigation* **2009**, *56*, 135–149. [CrossRef]
6. Collins, P.; Bisnath, S.; Lahaye, F.; Héroux, P. Undifferenced GPS ambiguity resolution using the decoupled clock model and ambiguity datum fixing. *Navigation* **2010**, *57*, 123–135. [CrossRef]
7. Seepersad, G.; Bisnath, S. Reduction of PPP convergence period through pseudorange multipath and noise mitigation. *GPS Solut.* **2015**, *19*, 369–379. [CrossRef]
8. Banville, S.; Collins, P.; Zhang, W.; Langley, R. Global and regional ionospheric corrections for faster PPP convergence. *Navigation* **2014**, *61*, 115–124. [CrossRef]
9. Li, P.; Zhang, X.; Guo, F. Ambiguity resolved precise point positioning with GPS and BeiDou. *J. Geod.* **2017**, *91*, 25–40.
10. Li, P.; Zhang, X.; Ren, X.; Zuo, X.; Pan, Y. Generating GPS satellite fractional cycle bias for ambiguity-fixed precise point positioning. *GPS Solut.* **2016**, *20*, 771–782. [CrossRef]
11. Mehra, R. On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Autom. Control* **1970**, *15*, 175–184. [CrossRef]
12. Maybeck, P.S. Combined State and Parameter Estimation for On-Line Applications. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1972.
13. Maybeck, P.S. Moving-bank multiple model adaptive estimation and control algorithms: An evaluation. *Control Dyn. Syst.* **1989**, *31*, 1–31.
14. Mohamed, A.; Schwarz, K. Adaptive Kalman filtering for INS/GPS. *J. Geod.* **1999**, *73*, 193–203. [CrossRef]
15. Odelson, B.; Lutz, A.; Rawlings, J. The autocovariance least-squares method for estimating covariances: Application to model-based control of chemical reactors. *IEEE Trans. Control Syst. Technol.* **2006**, *14*, 532–540. [CrossRef]
16. Magill, D. Optimal adaptive estimation of sampled stochastic processes. *IEEE Trans. Autom. Control* **1965**, *10*, 434–439. [CrossRef]
17. Yang, Y.; He, H.; Xu, G. Adaptively robust filtering for kinematic geodetic positioning. *J. Geod.* **2001**, *75*, 109–116. [CrossRef]
18. Yang, Y.; Gao, W. An optimal adaptive Kalman filter. *J. Geod.* **2006**, *80*, 177–183. [CrossRef]
19. Teunissen, P.J.; Amiri-Simkooei, A.R. Least-squares variance component estimation. *J. Geod.* **2008**, *82*, 65–82. [CrossRef]
20. Jamshidian, M.; Jennrich, R.I. Conjugate gradient acceleration of the EM algorithm. *J. Am. Stat. Assoc.* **1993**, *88*, 221–228.
21. Koch, K.R. Robust estimation by expectation maximization algorithm. *J. Geod.* **2013**, *87*, 107–116. [CrossRef]
22. Cai, C.; Gao, Y. Modeling and assessment of combined GPS/GLONASS precise point positioning. *GPS Solut.* **2013**, *17*, 223–236. [CrossRef]
23. Böhm, J.; Niell, A.; Tregoning, P.; Schuh, H. Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data. *Geophys. Res. Lett.* **2006**, *33*. [CrossRef]

24. Niell, A.E. Global mapping functions for the atmosphere delay at radio wavelengths. *J. Geophys. Res. Solid Earth* **1996**, *101*, 3227–3246. [CrossRef]
25. Shumway, R.H.; Stoffer, D.S. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* **1982**, *3*, 253–264. [CrossRef]
26. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–22.
27. Benoist, C.; Collilieux, X.; Rebischung, P.; Altamimi, Z.; Jamet, O.; Métivier, L.; Chanard, K.; Bel, L. Accounting for spatiotemporal correlations of GNSS coordinate time series to estimate station velocities. *J. Geodyn.* **2020**, *135*, 101693. [CrossRef]
28. Bock, O.; Collilieux, X.; Guillamon, F.; Lebarbier, E.; Pascal, C. A breakpoint detection in the mean model with heterogeneous variance on fixed time intervals. *Stat. Comput.* **2020**, *30*, 195–207. [CrossRef]
29. Jongrujinan, T.; Satirapod, C. Improving the stochastic model for VRS network-based GNSS surveying. *Artif. Satell.* **2019**, *54*, 17–30. [CrossRef]