State Space Modeling of Event Count Time Series

Sidratul Moontaha [1,*] [ID], **Bert Arnrich** [1] [ID] **and Andreas Galka** [2]

1. Digital Health—Connected Healthcare, Hasso Plattner Institute, University of Potsdam, 14482 Potsdam, Germany
2. Bundeswehr Technical Centre for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), 24340 Eckernförde, Germany
* Correspondence: sidratul.moontaha@hpi.de

Abstract: This paper proposes a class of algorithms for analyzing event count time series, based on state space modeling and Kalman filtering. While the dynamics of the state space model is kept Gaussian and linear, a nonlinear observation function is chosen. In order to estimate the states, an iterated extended Kalman filter is employed. Positive definiteness of covariance matrices is preserved by a square-root filtering approach, based on singular value decomposition. Non-negativity of the count data is ensured, either by an exponential observation function, or by a newly introduced "affinely distorted hyperbolic" observation function. The resulting algorithm is applied to time series of the daily number of seizures of drug-resistant epilepsy patients. This number may depend on dosages of simultaneously administered anti-epileptic drugs, their superposition effects, delay effects, and unknown factors, making the objective analysis of seizure counts time series arduous. For the purpose of validation, a simulation study is performed. The results of the time series analysis by state space modeling, using the dosages of the anti-epileptic drugs as external control inputs, provide a decision on the effect of the drugs in a particular patient, with respect to reducing or increasing the number of seizures.

Keywords: n onlinear state space model; iterated extended Kalman filter; Bayesian filtering; count time series; singular value decomposition

1. Introduction

Temporal data sequences resulting from counting discrete events over a given time interval represent a particular variant of time series called discrete-valued or event count time series. These count data arise in various fields, such as physics, epidemiology, finance, econometrics, or medicine. Much of the existing framework on time series analysis relies on the assumptions of Gaussianity and linearity. The non-negativity and integrity constraints inherent in count time series have led to the development of alternative modeling approaches that instead employ probability distributions of Poisson type. As prominent examples we mention generalized linear models (GLMs) [1], dynamic generalized linear models [2,3], integer-valued autoregressive (INAR) models [4], and integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) models [5–7].

However, in this paper we aim at modeling count time series within the linear Gaussian regime as long as possible, while introducing non-negativity and integrity only at the last stage of modeling, namely, at the stage of modeling the observation process. The suitable framework for this agenda is given by classical state space modeling [8]. In state space modeling, the dynamical process underlying the data is explicitly separated from the observation process, such that the former can be kept Gaussian and linear, which considerably simplifies state estimation and model identification. Further advantages of the state space approach are given by the possibility to discriminate between dynamical noise and observation noise, and by the option of straightforward generalization to the multivariate case.

It is also possible to include explanatory factors and external control inputs into state space models, and to incorporate conditional heteroscedasticity [9], with little additional effort.

Since the dynamics is kept Gaussian and linear, temporal correlations in the given data can be modeled by standard models from linear time series analysis, such as linear autoregressive moving average (ARMA) models [10]; ARMA models can be rewritten as (components of) linear state space models. Finally, non-negativity is modeled by employing a nonlinear observation function, while integrity is interpreted as the effect of a suitable additive observation noise term, which performs a kind of quantization of the non-integer output of the nonlinear observation function.

In linear state space modeling, tasks such as prediction, filtering, and smoothing may be performed by algorithms based on the linear Kalman filter (KF) [8,11]. However, in the case of a nonlinear observation function, generalizations of the linear KF need to be employed. Nowadays, a variety of algorithms is available, such as the extended KF (EKF), the iterated extended KF (IEKF), the unscented KF (UKF), and particle filtering [11,12].

The EKF is based on applying a linear KF by forming the local derivative of the nonlinear observation function (and the dynamical process, if it is also nonlinear) at the current estimate of the predicted state. The IEKF extends the EKF by an additional iteration that aims at finding consistent estimates for predicted and filtered states. It has recently been shown that the IEKF iteration can be interpreted as an application of Gauss–Newton optimization [12]. The UKF generalizes the EKF by propagating an ensemble of deterministically chosen points, thus improving Gaussian approximation and eliminating the explicit calculation of the Jacobian matrices. In principle, in Kalman filtering the error covariance matrices of the state estimates should be bounded and should converge to a steady solution, irrespective of the error in the initial state estimate and the corresponding initial error covariance matrix. However, the opposite phenomenon called *theoretical divergence* of classical KF is also well known. In order to keep the covariance matrix positive definite and improve the numerical behavior of the KF, square-root variants of the linear KF and its nonlinear generalizations have been introduced [13].

The mentioned algorithms have played a prominent role in applications in biomedical research and adjacent fields, such as public health [14]. Examples of corresponding count time series include epidemiological data (such as the famous U.S. poliomyelitis incidence time series, consisting of monthly counts starting in 1970) [15], sleep stage sequences, erythrocyte counts, infectious disease data [16], and epileptic seizure counts [17,18]. While most epilepsies respond well to anti-epileptic treatment, modeling the effects of anti-epileptic drugs (AEDs) on seizure frequency is essential for patients with difficult-to-treat or treatment-resistant epilepsies [19,20].

Several quantitative approaches to the analysis of seizure count time series have been proposed, which suffer from significant deficiencies. The *mixed-effects models* employed by Tharayil et al. cannot assess the effects of the changes in AEDs [21]. The epilepsy seizure management tool (*EpiSAT*) proposed by Chiang et al. does not account for observation errors caused by missed seizures or misinterpreted non-seizure events [22]. The Bayesian negative binomial dynamic linear model, recently proposed by Wang et al., cannot model the interaction effect between AEDs [23].

Application of state space modeling and Kalman filtering to seizure count time series has the potential to solve these deficiencies by quantifying the effect of an AED in the presence of other AEDs, describing delays in the effect of each AED, and modeling interaction effects between several AEDs. Moreover, the state space approach allows for the presence of temporal correlations in the seizure count time series that are unrelated to the current AED medication, but may result from other unknown influences on the probability of seizures. Furthermore, the state space approach is robust with respect to observation errors, such as missed seizures, events misclassified as seizures, outliers, missing data, or other observer-related errors. Therefore, the primary objective of this contribution is to explore and develop state space modeling algorithms tailored explicitly for event count time series, with a particular focus on the modeling of seizure count time series, as an

illustrative example. To the best of our knowledge, apart from our previous work [24–26], this approach is novel on its own.

This paper is organized as follows. In Section 2, "Materials and Methods", we discuss the main algorithms for Kalman filtering and for parameter estimation; we also describe the simulated data and the real data from a patient. In Section 3, "Results", we show the performance of the proposed algorithm, provide some comparison with respect to the numerical problems arising in previous algorithms, and show results from application to both simulated and real data. Section 4, "Discussion", concludes the paper. Additional information on the proposed state space models and Kalman filtering algorithms is provided in Appendices A and B.

2. Materials and Methods

2.1. Independent Components Linear State Space (IC-LSS) Models

The independent components linear state space (IC-LSS) model is a distinct category of the linear state space (LSS) models proposed by Galka et al. [27]. Let the data vector observed at time t be denoted by $\mathbf{y}_t$, where $t = 1, \ldots, T$ denotes discrete time, and let the dimension of $\mathbf{y}_t$ be denoted by n. The examples for analysis of actual data that are shown below are for scalar data only, i.e., $n = 1$, but we choose to keep the presentation of the methodology more general.

In linear state space (LSS) modeling, the observed data are linked to an unobserved m-dimensional state vector, $\mathbf{x}_t$, as described by an observation equation

$$\mathbf{y}_t = \mathsf{C}\mathbf{x}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathsf{R}) \tag{1}$$

where C and $\boldsymbol{\epsilon}_t$ represent the *observation matrix* and the *observation noise*, respectively. The observation noise is a white Gaussian noise with zero mean and *observation noise covariance matrix* R. By including this noise, the model acknowledges the limitations and uncertainties of real-world observations, and the information loss about the true state of the system. Within LSS models, the temporal evolution of the state vector, $\mathbf{x}_t$, is described by a discrete-time dynamical equation

$$\mathbf{x}_t = \mathsf{A}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathsf{Q}) \tag{2}$$

where A and $\boldsymbol{\eta}_t$ represent the *state transition matrix* and the *dynamical noise*, respectively. Also, the dynamical noise is a white Gaussian noise with zero mean and *dynamical noise covariance matrix* Q. An additional control input term, depending on a known external control input, $\mathbf{u}_t$, may be added to the dynamical equation:

$$\mathbf{x}_t = \mathsf{A}\mathbf{x}_{t-1} + \mathsf{B}_u\mathbf{u}_t + \boldsymbol{\eta}_t \tag{3}$$

where B_u represents the *control gain matrix*. The respective dimensions of the matrices and vectors are given in Table 1. The IC-LSS model is a specific subset of the LSS model family that characterizes data as a combination of independent source processes through a weighted sum and chooses specific structures of parameter matrices of the general state space model. The model depends on a set of parameters matrices, collected in the set $\Theta = \{\mathsf{C}, \mathsf{A}, \mathsf{R}, \mathsf{Q}, \mathsf{B}_u\}$. Both the state transition matrix, A, and the dynamical noise covariance matrix, Q, are constructed as block-diagonal matrices with identical sets of block dimensions, as described in Appendix A.

Our modeling approach assumes that the impact of each control input, i.e., each component of the vector $\mathbf{u}_t$, is independent of the other components, and that the corresponding processes can be modeled as deterministic first-order autoregressive models, to be denoted by AR(1). These AR(1) models are made deterministic by setting the corresponding elements of Q to zero. To account for temporally correlated fluctuations in the data caused by factors other than the control input, a stochastic process is also included. This stochastic process is modeled by an autoregressive moving average model with orders p and $p - 1$,

to be denoted as ARMA($p, p-1$). In the current implementation, we usually choose $p = 2$, which represents a trade-off between the stochasticity and stability of the model. AR(1) and ARMA(2,1) models can easily be incorporated into IC-LSS models (for technical details, see Appendix A).

Table 1. Quantities arising in state space models.

Notation	Meaning and/or Name	Dimension
$\mathbf{y}_t$	data vector	n
$\boldsymbol{\epsilon}_t$	observation noise vector	n
$\mathbf{x}_t$	state vector	m
$\boldsymbol{\eta}_t$	dynamical noise vector	m
$\mathbf{u}_t$	external control vector	u
C	observation matrix	$n \times m$
A	state transition matrix	$m \times m$
R	observation noise covariance matrix	$n \times n$
Q	dynamical noise covariance matrix	$m \times m$
B$_u$	control gain matrix	$m \times u$

2.2. Nonlinear State Space (NLSS) Models

Count time series do not follow Gaussian probability distributions, therefore the class of linear Gaussian state space models is not well suited for modeling such time series; rather it would be beneficial to employ appropriate nonlinear state space (NLSS) models. In this paper, we keep the dynamical equation linear, as in Equation (3), while defining a nonlinear observation equation by

$$\mathbf{y}_t = \mathbf{f}(\mathbf{x}_t) + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathsf{R}) \tag{4}$$

where we have assumed that the observation noise, $\boldsymbol{\epsilon}_t$, can be kept Gaussian; $\mathbf{f}(\,.\,)$ represents a nonlinear observation function. We employ two different nonlinear functions, namely

$$\mathbf{f}_1(\mathbf{x}) = \exp(\mathsf{C}\mathbf{x}) \tag{5}$$

and

$$\mathbf{f}_2(\mathbf{x}) = \frac{\mathsf{C}\mathbf{x}}{2} + \sqrt{\frac{(\mathsf{C}\mathbf{x})^2}{4} + k} \tag{6}$$

In the case of multivariate data, $n > 1$, these functions are to be applied component-wise.

The first of these observation functions, simply an exponential function, has been chosen with the intention of achieving non-negativity of the observed data, as it may also be achieved in Poisson regression. The disadvantage of this choice for the observation function is the fact that it diverges exponentially for large positive arguments. In previous work, we have occasionally encountered numerical breakdown of Kalman filtering algorithms due to the resulting extremely large values [25,26]; details will be provided below in Section 3.1. For this reason, we propose a different nonlinear observation function, to be called the "affinely distorted hyperbolic" function, as given in Equation (6); while for negative arguments it behaves like the exponential function, for positive arguments it converges to the linear function, rising with a slope of C, see Figure 1.

Recently, Weiß and coworkers [7,28] have introduced a nonlinear function called the "softplus" function, given by

$$\mathbf{f}(\mathbf{x}) = k \log\left(1 + \exp(\mathbf{x}/k)\right)$$

where k is a real positive parameter. This function could be employed instead of our "affinely distorted hyperbolic" function for the purpose of constraining observations to be non-negative, since it has similar behavior.
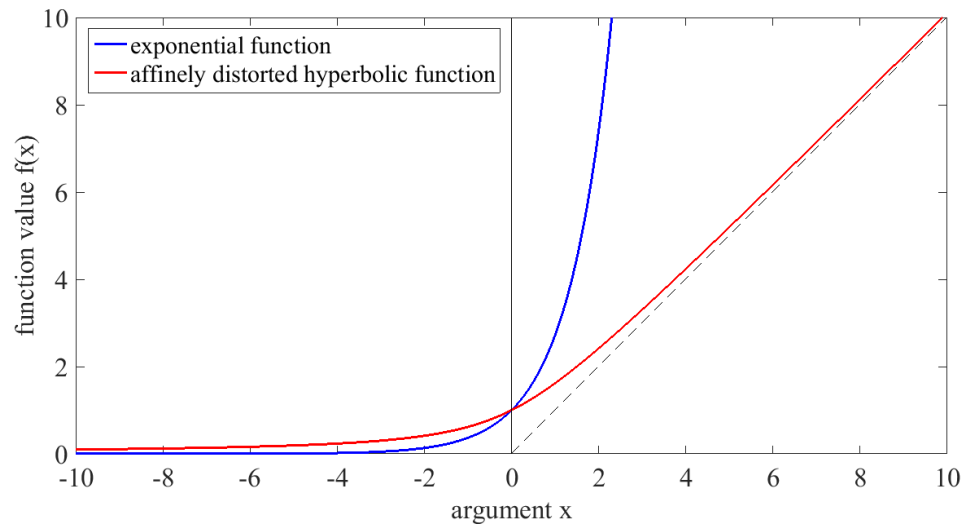
Figure 1. Nonlinear observation functions $\mathbf{f}_1(x)$ (blue) and $\mathbf{f}_2(x)$ (red), with $\mathsf{C} = 1$ and $k = 1$; the dashed vertical line on the right side represents the linear function $\mathbf{f}(x) = x$.

2.3. Iterated Extended Kalman Filter (IEKF)

In this paper, we are dealing with a two-fold estimation problem: estimation of states and estimation of parameters. The extended Kalman filter (EKF) has been used as a popular tool for estimating the states in nonlinear state space models. It was first proposed by Kalman and Bucy in 1961 [29]. The EKF results from extending the original Kalman filter developed for linear systems to nonlinear systems by linearizing the dynamical equation and the observation equation around the current state estimate. As an improved variant of EKF, the iterated extended Kalman filter (IEKF) has been proposed, in order to improve the accuracy and stability of the EKF [30]. The IEKF has been applied to various fields, including robotics, aerospace, and control systems.

For a given time series of length T and given initial state estimate $\mathbf{x}_{0|0}$ and corresponding covariance matrix $\mathsf{P}_{0|0}$, the forward temporal recursion begins at time $t = 1$ with the predicted state estimate [25], which is computed by

$$\mathbf{x}_{t|t-1} = \mathsf{A}\mathbf{x}_{t-1|t-1} + \mathsf{B}_u\mathbf{u}_t \tag{7}$$

The notation $\mathbf{x}_{t_1|t_0}$ is used throughout the paper to indicate that an estimate at time t_1 is obtained by using all data available at time t_0. The corresponding predicted state covariance is computed by

$$\mathsf{P}_{t|t-1} = \mathsf{A}\mathsf{P}_{t-1|t-1}\mathsf{A}^T + \mathsf{Q} \tag{8}$$

At each time point, the IEKF iteration starts, after Equations (7) and (8) have been evaluated, with iteration index $i = 1, \ldots, i_m$. First, the derivative of the chosen nonlinear function is computed as

$$\mathsf{H}_t^{(i)} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_{t|t}^{(i-1)}} \tag{9}$$

In the multivariate case, this derivative will be a matrix (Jacobian matrix). At each iteration, the prediction errors (also known as *innovations*), the innovation variance, the Kalman gain, and the filtered state estimates are computed, according to

$$\boldsymbol{\nu}_t^{(i)} = \mathbf{y}_t - \mathbf{f}\big(\mathbf{x}_{t|t}^{(i-1)}\big) \tag{10}$$

$$\mathsf{V}_t^{(i)} = \mathsf{H}_t^{(i)} \mathsf{P}_{t|t-1}\big(\mathsf{H}_t^{(i)}\big)^T + \mathsf{R} \tag{11}$$

$$\mathsf{K}_t^{(i)} = \mathsf{P}_{t|t-1}\big(\mathsf{H}_t^{(i)}\big)^T \big(\mathsf{V}_t^{(i)}\big)^{-1} \tag{12}$$

$$\mathbf{x}_{t|t}^{(i)} = \mathbf{x}_{t|t-1} + \mathsf{K}_t^{(i)}\left(\nu_t^{(i)} - \mathsf{H}_t^{(i)}\left(\mathbf{x}_{t|t-1} - \mathbf{x}_{t|t}^{(i-1)}\right)\right) \tag{13}$$

The iteration ends when a stopping criterion is fulfilled. We use the stopping criterion of either the norm of the relative change of $\mathbf{x}_{t|t}^{(i)}$ falling below 10^{-10}, or the iteration index reaching a maximal value of $i_m = 100$. After the iteration, the filtered state covariance is computed by

$$\mathsf{P}_{t|t} = \big(\mathsf{I}_m - \mathsf{K}_t^{(i)}\mathsf{H}_t^{(i)}\big)\mathsf{P}_{t|t-1} \tag{14}$$

where I_m denotes the m-dimensional identity matrix.

The IEKF algorithm, as presented above, is summarized in Appendix B. In an earlier paper [25], we have presented results from actual application of the IEKF, for the case of an exponential observation function.

2.4. Singular Value Decomposition Iterated Extended Kalman Filter (SVD-IEKF)

At each time step, the covariance matrices $\mathsf{P}_{t|t-1}$, V_t, and $\mathsf{P}_{t|t}$ arising in the Kalman filter recursion have to be positive definite, or at least, positive semi-definite. However, it is well known that during the recursion due to numerical effects these covariance matrices may lose the property of positive definiteness. As a consequence, computation of the likelihood becomes unreliable; furthermore, the iteration of the IEKF may converge only with delays, or it may entirely fail to converge [25].

As a remedy, one may work with matrix square roots of the covariance matrices, instead of the covariance matrices themselves. There are, at least, two ways to define the square root of a matrix: by Cholesky decomposition (CD) and by singular value decomposition (SVD).

For a given real square matrix M with dimension $m \times m$, the Cholesky decomposition is given by

$$\mathsf{M} = \mathsf{S}_\mathsf{M}^T \mathsf{S}_\mathsf{M} \tag{15}$$

where S_M denotes a $(m \times m)$-dimensional upper triangular matrix with non-negative diagonal elements. The Cholesky Decomposition is only possible for matrices that are positive (semi-)definite; for semi-definiteness, the decomposition may not be unique.

For a given real matrix, M, with dimension $m \times n$, the SVD decomposition is given by

$$\mathsf{M} = \mathsf{U}_\mathsf{M} \Sigma_\mathsf{M} \mathsf{W}_\mathsf{M}^T \tag{16}$$

where U_M and W_M denote two orthogonal matrices with dimensions $m \times m$ and $n \times n$, respectively, and Σ_M is a diagonal matrix with dimension $m \times n$, which has non-negative real numbers, σ_i, on its diagonal, called the singular values of M [31]. The singular values are the positive square roots of the eigenvalues of $\mathsf{M}^T\mathsf{M}$ [13]. SVD can be applied to any matrix, without the condition of positive (semi-)definiteness.

The SVD of a positive (semi-)definite square matrix, M, e.g., a covariance matrix, can be formulated as

$$\mathsf{M} = \mathsf{W}_\mathsf{M} \Sigma_\mathsf{M}^2 \mathsf{W}_\mathsf{M}^T \tag{17}$$

such that we have $U_M = W_M$, and the definition of the matrix of singular values has been changed to Σ_M^2 instead of Σ_M; this is a reasonable change of definition, since the singular values are non-negative.

Square-root variants of the Kalman filter that employ CD were proposed already in the 1960s. However, SVD represents a matrix decomposition that offers superior numerical properties, compared with CD. Square-root variants of the Kalman filter that employ SVD were proposed in 1992 by Wang et al. [13], and in 2017 by Kulikova and Tsyganova [32]. In an earlier paper, we proposed a square-root variant of the IEKF that employs SVD [26], based on the algorithm of Kulikova and Tsyganova.

In the proposed nonlinear Kalman filter algorithm, the initial covariance matrix, $P_{0|0}$, and the noise covariance matrices, Q and R, are factorized by SVD as follows:

$$P_{0|0} = W_{0|0}\Sigma_{0|0}^2 W_{0|0}^T$$
$$Q = W_Q\Sigma_Q^2 W_Q^T$$
$$R = W_R\Sigma_R^2 W_R^T$$

These factorizations are performed once outside of the forward recursion of the Kalman filter. The recursion then begins with the same update equation for the predicted state as before, Equation (7). However, now the corresponding predicted state covariance matrix is updated by performing a factorization step by applying SVD to a *pre-array*, as follows [26]:

$$U_1 \underbrace{\begin{bmatrix} \Sigma_{P_{t|t-1}} \\ 0 \end{bmatrix}}_{\text{singular values array}} W_{P_{t|t-1}}^T = \underbrace{\begin{bmatrix} \Sigma_{P_{t-1|t-1}} W_{P_{t-1|t-1}}^T A^T \\ \Sigma_Q W_Q^T \end{bmatrix}}_{\text{pre-array}} \tag{18}$$

$$P_{t|t-1} = W_{P_{t|t-1}}\Sigma_{P_{t|t-1}}^2 W_{P_{t|t-1}}^T \tag{19}$$

where U_1 is an orthogonal matrix that can be discarded. Also, in the SVD-IEKF algorithm, at each time point an iteration is performed, with the equations for the Jacobi matrix and the innovations given by Equations (9) and (10), respectively. However, Equation (11) is replaced by an array factorization step:

$$U_2 \underbrace{\begin{bmatrix} \Sigma_{V_t}^{(i)} \\ 0 \end{bmatrix}}_{\text{singular values array}} \left(W_{V_t}^{(i)}\right)^T = \underbrace{\begin{bmatrix} \Sigma_R W_R^T \\ \Sigma_{P_{t|t-1}}^{(i)} \left(W_{P_{t|t-1}}^{(i)}\right)^T \left(H_t^{(i)}\right)^T \end{bmatrix}}_{\text{pre-array}} \tag{20}$$

$$V_t^{(i)} = W_{V_t}^{(i)}\left(\Sigma_{V_t}^{(i)}\right)^2\left(W_{V_t}^{(i)}\right)^T \tag{21}$$

where U_2 is an orthogonal matrix which can be discarded. As the iteration proceeds, the normalized innovation, the normalized gain, the optimal Kalman gain, and the filtered state estimate are computed as follows:

$$\widehat{\boldsymbol{v}}_t^{(i)} = \left(W_{V_t}^{(i)}\right)^T \boldsymbol{v}_t^{(i)} \tag{22}$$

$$k_t^{(i)} = P_{t|t-1}\left(H_t^{(i)}\right)^T W_{V_t}^{(i)} \tag{23}$$

$$K_t^{(i)} = k_t^{(i)}\left(\Sigma_{V_t}^{(i)}\right)^{-2}\left(W_{V_t}^{(i)}\right)^T \tag{24}$$

$$\mathbf{x}_{t|t}^{(i+1)} = \mathbf{x}_{t|t-1} + k_t^{(i)}\left(\Sigma_{V_t}^{(i)}\right)^{-2}\left(\widehat{\boldsymbol{v}}_t^{(i)} - \left(W_{V_t}^{(i)}\right)^T H_t^{(i)}\left(\mathbf{x}_{t|t-1} - \mathbf{x}_{t|t}^{(i)}\right)\right). \tag{25}$$

After reaching the stopping criterion of the iteration, the filtered state covariance matrix, $P_{t|t}$, is computed by a third array factorization step:

$$
U_3 \underbrace{\begin{bmatrix} \Sigma_{P_{t|t}} \\ 0 \end{bmatrix}}_{\text{singular values array}} W_{P_{t|t}}^T = \underbrace{\begin{bmatrix} \Sigma_{P_{t|t-1}} W_{P_{t|t-1}}^T \left(I_m - K_t^{(i)} H_t^{(i)} \right)^T \\ \Sigma_R W_R^T \left(K_t^{(i)} \right)^T \end{bmatrix}}_{\text{pre-array}}
\tag{26}
$$

$$
P_{t|t} = W_{P_{t|t}} \Sigma_{P_{t|t}}^2 W_{P_{t|t}}^T
\tag{27}
$$

where U_3 denotes another orthogonal matrix that can be discarded, and i denotes the index at which the iteration had stopped. The SVD-IEKF algorithm, as presented above, is summarized in Appendix B.

2.5. Non-Dynamic Regression Model: Gaussian Case

For the purpose of comparison, we also employ two types of non-dynamic regression models. The first of these models represents the classical linear Gaussian case; it is defined as follows:

$$
\mathbf{y}_t = \tilde{B}_u \mathbf{u}_t + \mathbf{n}_t
\tag{28}
$$

where $\tilde{B}_u$ denotes a matrix of regression coefficients, and $\mathbf{n}_t$ denotes a time series of regression residuals, assumed to have a Gaussian distribution with zero mean and covariance matrix Σ_n. $\tilde{B}_u$ can be estimated by ordinary least squares; the covariance matrix, Σ_n, can then be computed as

$$
\Sigma_n = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \tilde{B}_u \mathbf{u}_t)(\mathbf{y}_t - \tilde{B}_u \mathbf{u}_t)^T
\tag{29}
$$

Note that the regression model of Equation (28) implicitly assumes that the effects of the different components of the control vector, $\mathbf{u}_t$, are uncorrelated.

From the regression model of Equation (28), a logarithmic likelihood can be computed by

$$
\log L(\tilde{B}_u) = -\frac{T}{2} \left(\log |\Sigma_n| + n \log 2\pi + n \right)
\tag{30}
$$

2.6. Non-Dynamic Regression Model: Poisson Case

The second non-dynamic regression model is an example of a generalized linear model, representing the Poisson case; it is defined as follows:

$$
\log y_t = \tilde{B}_u \mathbf{u}_t + n_t
\tag{31}
$$

where $\tilde{B}_u$ denotes another matrix of regression coefficients, and n_t denotes another time series of regression residuals; in this case, y_t is assumed to have a Poisson distribution. Note that, for simplicity, we formulate this model only for the case of scalar data.

From the regression model of Equation (31), a logarithmic likelihood can be computed by

$$
\log L(\tilde{B}_u) = \sum_{t=1}^{T} \left(y_t \tilde{B}_u \mathbf{u}_t - \exp(\tilde{B}_u \mathbf{u}_t) - \log(y_t!) \right).
\tag{32}
$$

2.7. Parameter Estimation and Ensembles of Models

As mentioned earlier, when fitting a state space model to given data we have to solve a two-fold estimation problem: estimation of states and estimation of parameters. In this paper, estimation of the model parameters, denoted by Θ in Section 2.1, is performed by numerical maximization of the logarithmic innovation likelihood, denoted by $\log L(\Theta)$, employing the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton and Nelder–Mead simplex algorithms [33]. Apart from filtered state estimates, the forward recursion

of the Kalman filter also provides the corresponding contributions to the logarithmic innovation likelihood, which may then be summed up:

$$\log L(\Theta) = -\frac{1}{2} \sum_{t=1}^{T} \left(\log |\mathsf{V}_t| + \boldsymbol{v}_t^T \mathsf{V}_t^{-1} \boldsymbol{v}_t \right) - \frac{nT}{2} \log 2\pi \tag{33}$$

In this expression, the effect of the initial state has been ignored. If the innovations, $\boldsymbol{v}_t$, have been obtained by the IEKF or SVD-IEKF algorithms, their values correspond to the final values resulting from the iteration performed at time t, until the stopping criterion is fulfilled.

In order to avoid parameter redundancy with respect to the control gain matrix, B_u, the observation matrix, C, has not been included in the set of parameters to be estimated by optimization; instead constant values of 1.0 are employed, except for the ARMA(2,1) component, for which the control gain parameters are fixed at zero.

For given data, comparison of the performance of state space models, as discussed above in Sections 2.2–2.4, with the performance of non-dynamic regression models, as discussed above in Sections 2.5 and 2.6, can be performed by comparison of the corresponding values of an information criterion, such as the (corrected) Akaike information criterion (AICc) [27]. The AICc can be computed from the logarithmic likelihood according to

$$\text{AICc} = -2 \log L(\Theta) + \frac{2N_{\text{par}} T}{T - N_{\text{par}} - 1} \tag{34}$$

where N_{par} denotes the number of data-adaptive parameters of the corresponding model; in the case of state space models, this would be the total number of parameters in Θ; if the nonlinear observation function, $\mathbf{f}(.)$, according to Equation (6), is chosen, also the parameter k becomes part of Θ.

The second term in Equation (34) represents a penalty term for the complexity of the model; through this term, the values of AICc for different models can be directly compared, while the corresponding values of the logarithmic likelihood cannot, since they would be biased in favor of the more complex model. The task of parameter estimation is then given by finding parameters that minimize the AICc. Due to the complicated dependence of the AICc on Θ, including the possible existence of numerous local minima, this task can only be approached by numerical optimization. The chosen algorithm for numerical optimization may converge to one of these local minima, instead of the global minimum, or to other minima with smaller AICc values.

This issue can be resolved by employing an ensemble of models. Each model in the ensemble is initialized with randomly selected initial parameter values and optimized using the same minimum AICc approach until convergence. Finally, the model with the smallest AICc value is retained. However, although the probability of actually finding the global minimum will be improved by this ensemble approach, no assurance of finding the global minimum is provided.

2.8. Simulated Data

To validate the performance and effectiveness of our proposed NLSS modeling approach, we conduct simulations, before applying it to the patient data. These simulations serve to demonstrate the algorithm's capability in accurately capturing the dynamics of event count time series. We assume that three anti-epileptic drugs are given, named AED1, AED2 and AED3, and that, for the particular simulated "patient", AED1 and AED3 reduce the daily number of seizures, while AED2 increases it. The (arbitrarily chosen) time-dependent dosages of the drugs during a time interval of 500 days are shown in the upper panel of Figure 2a, while the resulting time series of the daily numbers of seizures is shown in the lower panel of the figure.

The simulated time series is generated by employing Equations (1) and (3), utilizing the affinely distorted hyperbolic function $\mathbf{f}_2(\mathbf{x}_t)$, as discussed in Equation (6). In addition

to the contributions of the three anti-epileptic drugs, a stochastic ARMA(2,1) process is added. The resulting time series is rounded to integer values, representing the simulated daily seizure counts. The daily seizure counts assume values from 0 to 6, which is a realistic interval. The MATLAB code for recreating the simulated data is provided in Appendix C.
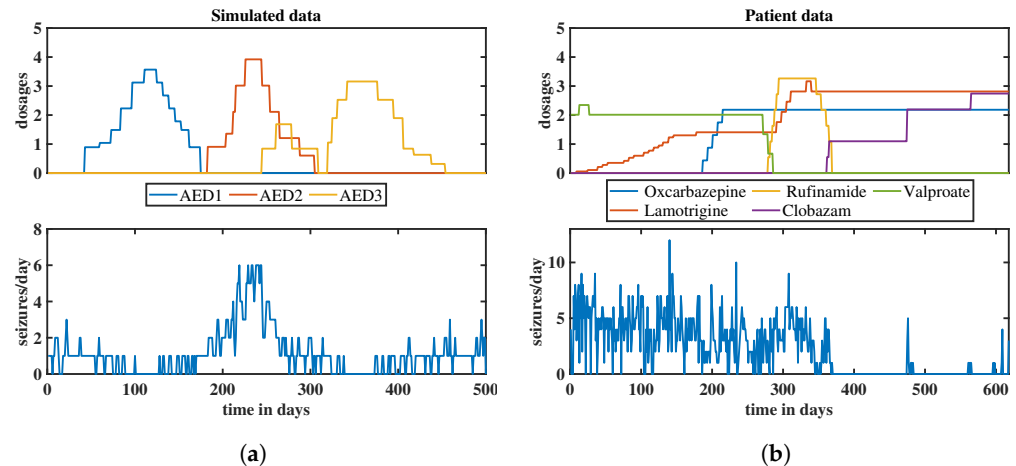
Figure 2. Time-dependent dosages of anti-epileptic drugs (upper panels) and corresponding count time series of daily epileptic seizures (lower panels) for a simulation (**a**) and for a real patient (**b**); in the upper panels, the anti-epileptic drugs are discriminated by colors (see legends below the panels, which also give the names of the drugs).

2.9. Patient Data

We demonstrate the practical application of NLSS modeling through the analysis of a real-world data set obtained from a patient suffering from symptomatic epilepsy. The data set utilized in this study was collected from an electronic seizure diary called EPI-Vista (http://www.epivista.de, accessed on 20 September 2023), which has been in routine use at the North German Epilepsy Center for Children and Adolescents in Schwentinental-Raisdorf, Germany, since 2007. EPI-Vista is a freely available therapy management tool that records information about dosages of administered anti-epileptic drugs and about seizure events.

The time-dependent dosages of the drugs during the chosen time interval of 618 days are shown in the upper panel of Figure 2b, while the recorded time series of the daily numbers of seizures is shown in the lower panel of the figure. During the chosen time interval, five different anti-epileptic drugs were administered: oxcarbazepine, lamotrigine, rufinamide, clobazam, and valproate.

3. Results

3.1. Convergence Behavior of Iteration

We will briefly comment on the convergence behavior of the iteration of the SVD-IEKF. Within the recursion of the SVD-IEKF through the data, the iteration takes place at each time point. We may plot the norm of the relative change of the state estimate as a function of the iteration index, obtaining a set of curves; examples are shown in Figures 3 and 4. These examples refer to the analysis of the simulated data, as described in Section 2.8. In Figure 3a, it can be seen that for all 500 time points the iteration converges according to a power law within, at most, 35 iterations, thereby confirming that the SVD-IEKF works properly.

However, within an ensemble of models, cases may also occur that show less favorable behavior. An example is shown in Figure 3b. In this case, we see convergence only for the first time point; the iteration loop stops after 500 iterations. State estimates obtained from this iteration were extremely large (in the order of e^{100}), and therefore the Kalman filter recursion was not continued, and no further iterations were performed. This problem can be resolved by replacing the exponential function with the affinely distorted hyperbolic function in the observation equation, Equation (1).
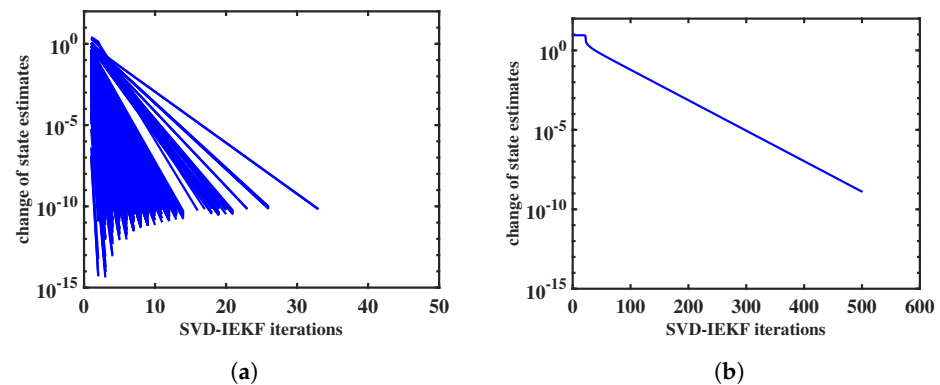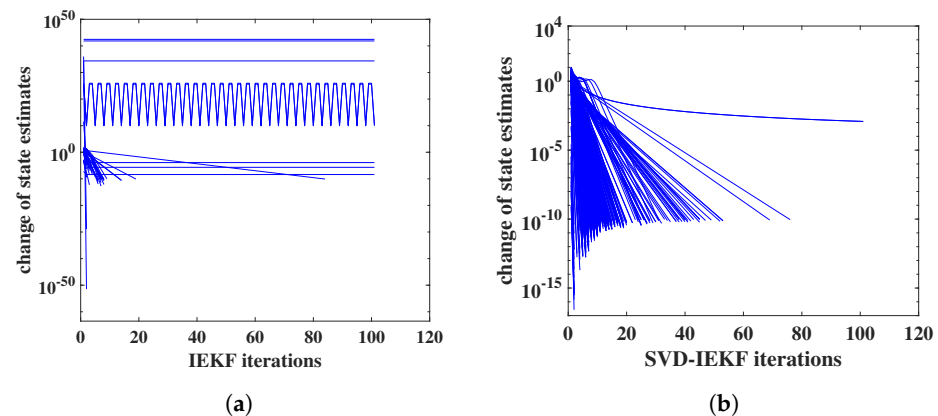
Figure 3. Norm of relative change of state estimate vs. iteration index for the application of the SVD-IEKF to a simulated count time series, using a model initialized with two different sets of random model parameters; in (**a**) iterations for all 500 time points are shown, while in (**b**) only the first iteration is shown, since the recursion of the Kalman filter was aborted afterwards due to numerical failure.

In Figure 4, we illustrate the effects of loss of positive definiteness of covariance matrices on the convergence behavior. Also, this example refers to the analysis of the simulated data. The standard IEKF was used, as described above in Section 2.3, and the affinely distorted hyperbolic function was employed.

In Figure 4a, it can be seen that for most time points the iteration completely fails to converge, instead norms of relative state changes stay approximately constant or oscillate; note the extremely large, or small, values on the vertical axis. In Figure 4b, the effect of switching from the IEKF to the SVD-IEKF is illustrated, for the same simulated data and the same set of model parameters: now, for almost all time points, good convergence within 60 iterations is obtained. For a single time point, convergence is slower and seems not to follow a power law.

Within an ensemble of 1000 random initial models, we find that for 18 models the IEKF encounters numerical problems resulting from covariance matrices losing positive definiteness or becoming singular, and another 11 models exhibit poor convergence behavior. For the SVD-IEKF, all models of the ensemble display good or satisfactory convergence behavior, without any numerical instability of the Kalman filter.

Figure 4. Norm of relative change of state estimate vs. iteration index for the application of the IEKF (**a**) or SVD-IEKF (**b**) to a simulated count time series, using a model initialized with a set of random model parameters; iterations for all 500 time points are shown.

### 3.2. Results of Ensemble Approach: Simulated Data

Following earlier work [24], we plot the control gain parameters, i.e., the diagonal elements of the control gain matrix, B_u, as defined in Equation (3), against the corresponding values of the AICc for all models of the ensemble; for each diagonal element, i.e., for each

anti-epileptic drug, a separate plot is created. For the simulated data, the resulting plots are shown in Figure 5; blue dots represent the results from the 1000 models of the ensemble. The model with the lowest AICc value is highlighted in green. In addition, the results obtained by the two non-dynamic regression models are also shown, and are represented by red (Gaussian) and deep purple (Poisson); error bars are also shown, although they are mostly very small.

From Figure 5 it can be seen that most of the models of the ensemble achieve a lower AICc, compared with the non-dynamic regression models. For AED1 and AED2, the clouds of blue dots scatter over both positive and negative values of the control gain parameter; however, the model with the smallest control gain parameter is negative for AED1, and positive for AED2. For AED3, literally all models yield negative control gain parameters.

Since the simulation was designed such that AED1 and AED3 reduce the daily number of seizures, while AED2 increases it, the ensemble approach has succeeded in retrieving the correct result. As can be seen from Figure 5, in this case also the non-dynamic regression models reproduce the correct result.

In Table 2, the correct values of the observation parameters, which were used for creating the simulated data, and the estimated values of these parameters, obtained for the model with the lowest AICc value, are shown. The table also lists estimated errors for the estimated parameters; these errors can be estimated by computing the Hessian of the local likelihood at the optimal point. However, it is obvious that, at least for AED1 and AED3, these estimated errors are much too small to describe the actual deviation of the estimated values from the correct values; the probable reason for this is that in nonlinear filtering algorithms with an iteration at each time point, such as the IEKF and SVD-IEKF, the local likelihood often has a complicated shape with discontinuous behavior, such that numerically computed Hessians do not provide reliable estimates of the estimation errors.
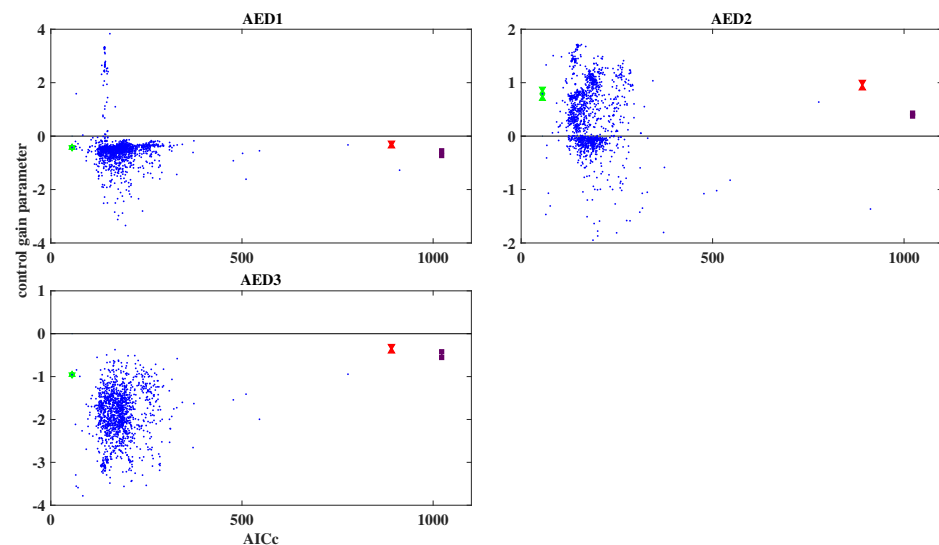
Figure 5. Estimated control gain parameters vs. AICc for simulated data, using an ensemble of 1000 randomly initialized models. The blue dots represent the results for the ensemble, with the model with the lowest AICc highlighted in green. The red (Gaussian case) and deep purple (Poisson case) dots represent the results obtained by the non-dynamic regression models. For the green, red, and deep purple dots, error bars are added, but they are mostly very small. The three panels refer to the three anti-epileptic drugs that were used in the simulation.

Table 2. Correct and estimated values of the observation parameters for the simulation study; the third row gives estimated errors for the estimated values, obtained from the Hessian of the local likelihood.

Anti-Epileptic Drug	AED1	AED2	AED3
correct values	-0.40	0.95	-0.70
estimated values (best model)	-0.43	0.79	-0.96
estimated errors (best model)	$\pm 1.86 \times 10^{-7}$	$\pm 4.57 \times 10^{-2}$	$\pm 4.33 \times 10^{-8}$

3.3. Result of Ensemble Approach: Patient Data

For the patient data, an ensemble of 700 randomly initialized models was employed. When using the SVD-IEKF and the affinely distorted hyperbolic function, optimization of all models proceeded without cases of numerical problems. The resulting plots of estimated control gain parameters against the corresponding values of the AICc for all models of the ensemble are shown in Figure 6; again, blue dots represent the results from the models of the ensemble, the model with the lowest AICc value is highlighted in green, and the results obtained by the two non-dynamic regression models are represented by red (Gaussian) and deep purple (Poisson).

From Figure 6 it can be seen that, again, most of the models of the ensemble achieve a lower AICc, compared with the non-dynamic regression models. The error bars of the red dots are somewhat larger now than in the case of the simulated data; for clobazam, the error interval includes the value of zero for the control gain parameter, such that it would become impossible to decide whether the effect of this anti-epileptic drug on the seizure count would be increasing, decreasing, or zero. Furthermore, based on the non-dynamic regression models, we would conclude that oxcarbazepine and lamotrigine would decrease the seizure count, while rufinamide and valproate would increase it.
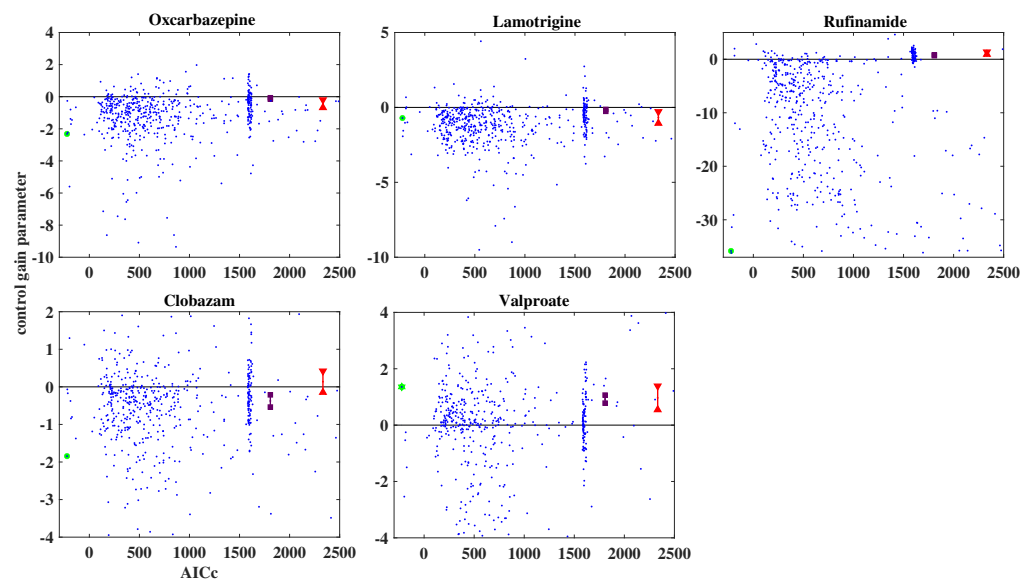
Figure 6. Estimated control gain parameters vs. AICc for patient data, using an ensemble of 700 randomly initialized models. The blue dots represent the results for the ensemble, with the model with the lowest AICc highlighted in green. The red (Gaussian case) and deep purple (Poisson case) dots represent the results obtained by the non-dynamic regression models. For the green, red, and deep purple dots, error bars are added. The five panels refer to the five anti-epileptic drugs that were administered during the chosen time interval of 618 days.

On the other hand, according to the results of the analysis by the SVD-IEKF, we would conclude that all anti-epileptic drugs, except for valproate, would decrease the seizure count; consequently, for two anti-epileptic drugs, the conclusions would differ from the conclusions based on the non-dynamic regression models.

We add a comment on the results of the ensemble approach, both for simulated and patient data, as shown in Figures 5 and 6. In this paper, we are dealing with time series of fairly short length, at most a few hundred values, since this is the typical situation for time series of a daily number of certain events, such as epileptic seizures. As a consequence of this scarcity of data, relative to the dimension of the space of model parameters, the AICc (i.e., the negative likelihood), as a function of these parameters, will display many local minima, and the task of finding the global minimum may not be well defined. For this reason, we consider it necessary to choose an ensemble approach. In the figures showing the results of the ensemble approach, the clouds of results that have a higher value of the AICc than the best model also contain some useful information.

As an example, consider again Figure 6. For each of the five anti-epileptic drugs displayed, it can be seen that the majority of the blue dots correspond to the same sign of the control gain parameter as the best model (i.e., the dot highlighted in green), thereby lending additional credibility to the resulting conclusions on the effects of these drugs. If, in such a case, the blue dots were distributed about equally over positive and negative values of the control gain parameter, we would be less inclined to regard the sign of this parameter for the best model as significant.

The underlying problem with respect to estimating model parameters from small data sets is unrelated to the particular choice of a state space model with linear dynamics and nonlinear observation; it has been observed quite similarly in an earlier study based on purely linear modeling [24]. The task to be addressed here is to draw the optimal conclusion, based on the scarce available data.

3.4. Innovation Whiteness Test

The aim of modeling a given time series by a parametric model, such as a state space model, is given by extracting temporal correlations as much as possible, such that the remaining prediction errors, i.e., the innovations, do not contain any residual correlations; this is equivalent to the innovations being white noise. In order to demonstrate that our modeling of the given data has been successful, with respect to this whiteness criterion, we will now show the autocorrelation function of the innovations of the best state space models from the ensembles.

In Figure 7, the autocorrelation functions for the innovations of the simulated data (left panel) and the patient data (right panel) are shown (green curves); for comparison, the autocorrelation functions of the raw data are also shown (blue curves). The red dashed lines correspond to one standard deviation. Note that, at lag zero, autocorrelation functions always assume a value of 1.0. By comparison of the blue and the green curves it can be seen that, both for simulated data and for patient data, very good whitening has been achieved, with only a few outliers exceeding one standard deviation. This result confirms the validity, in a statistical sense, of the chosen approach of modeling the data by a state space model with a nonlinear observation function.

(**a**)   (**b**)
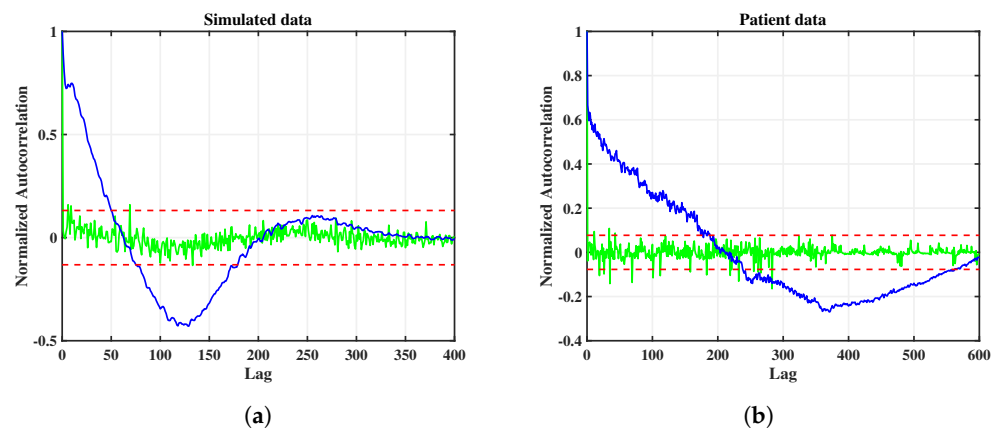
Figure 7. Autocorrelation functions of raw data (blue curves) and innovations after state space modeling (green curves) for simulated data (**a**) and patient data (**b**), for the best models of the respective ensembles. The red dashed lines correspond to one standard deviation of the innovations.

4. Discussion

In this paper, we propose an algorithm for analyzing event count time series by state space modeling and Kalman filtering. Within the larger field of time series analysis, the analysis of event count time series represents a special case, which is characterized by the need to model data that are both non-negative and integer, i.e., data given by natural numbers. The classical linear Gaussian state space model seems less suited for such data, since count data are described by Poisson distributions, rather than Gaussian distributions.

However, as we demonstrate, it is possible to keep the dynamics of the state space model Gaussian and linear, while introducing nonlinearity only at the last stage of modeling, namely, at the stage of modeling the observation process, thereby enforcing non-negativity and integrity of the observed data. The step from the non-integer output of the observation function to the integer data can then be roughly interpreted as addition of "quantization noise". By this device, Poisson distributions need not be employed explicitly. Nevertheless, due to the choice of a nonlinear observation function, it is necessary to employ nonlinear Kalman filters.

The present paper summarizes and extends earlier work, according to a sequence of steps that can be described as follows:

- In an initial study, we had proposed to analyze event count time series by purely linear Gaussian state space modeling, using the standard linear Kalman filter [24].
- Then, as a generalization, we employed nonlinear state space modeling, such that a linear dynamical equation was combined with a nonlinear observation function; the exponential function was chosen. State space modeling was performed by the iterated extended Kalman filter (IEKF) [25].
- As a further step of improvement, the standard IEKF was replaced by the numerically superior singular value decomposition variant of the IEKF [26].
- In the present paper, we replaced the exponential function with the "affinely distorted hyperbolic" function; alternatively, the "softplus function" of Weiß and coworkers could also have been employed. We have not yet systematically compared both functions, but expect that they would display similar performance.

The classical linear Kalman filter consists of a recursion in the forward direction through time, which represents the optimal state estimator for linear Gaussian state space models [8]. As soon as nonlinearities are introduced into the dynamical equation or the observation equation, no closed-form optimal recursion exists, such that approximations and additional iterations have to be employed. The iterated extended Kalman filter (IEKF) represents a well-established example of such approximative nonlinear state estimators.

It is well known that, in the practical application of both the classical linear Kalman filter and its nonlinear generalizations, numerical problems may arise, which result from

covariance matrices losing the property of positive definiteness. The usual remedy for such problems is given by expressing the recursion, not directly for the covariance matrices, but for square roots of these matrices, which is known as *square root filtering*. Matrix square roots may be defined either by Cholesky decomposition, or by singular value decomposition (SVD). Since the latter decomposition represents the more general and numerically more robust decomposition [34], we have chosen to employ it for our state estimation algorithm. The resulting algorithm is, to the best our knowledge, the first SVD variant of the IEKF that has been proposed [26].

When defining a state space model with a linear dynamical equation, the use of an exponential observation function represents a natural choice, in order to keep the data non-negative. However, the disadvantage of the exponential function is its exponential divergence for positive arguments, which may lead to numerical failure of the SVD-IEKF algorithm. For this reason, we propose a new nonlinear observation function that converges to zero for negative arguments, just like the exponential function, while converging to a linear function for positive arguments. This function can be derived by affinely distorting a (negative) hyperbolic function, such that the vertical axis becomes the desired linear function for positive arguments. As we have demonstrated in the present paper by analyzing both simulated and real data, employing the affinely distorted hyperbolic function within the SVD-IEKF algorithm finally removes the risk of numerical failure.

We emphasize that the proposed algorithm can be applied to any event count time series that one may wish to analyze under a non-negativity constraint; here, as an example, we have applied the algorithm to the analysis of time series of the daily number of seizures of drug-resistant epilepsy patients undergoing treatment with several, simultaneously administered anti-epileptic drugs. The time-dependent dosages of these drugs are inserted into the state space model as an external control input. The simultaneous presence of several drugs, their potential superposition effects, delay effects, and further unknown factors influencing the daily number of seizures make the objective analysis of seizure count time series arduous. As we have demonstrated, both in a simulation study and in the analysis of data from a real patient, state space modeling provides a powerful and flexible framework for analyzing the effects of anti-epileptic drugs in epilepsy patients. By comparison of an information criterion, such as the AICc, it can be proven that state space modeling provides a modeling of the data that is superior to modeling by conventional non-dynamic regression models.

There exist other model classes that take temporal correlations into account, such as the dynamic GLM, INAR, and INGARCH model classes. Our approach to modeling event count time series, as proposed in the present paper, represents an alternative to these model classes, but we do not intend to replace these classes, but rather provide an additional tool for the analysis of event count time series. According to a well-known proverb, "all models are wrong, but some models are useful". Then, the justification of introducing a new class of models can come only from their usefulness in practical work, which needs to be explored by application to real data. Within the likelihood framework, the comparison of different model classes should be performed by quantifying their performance in predicting the data, preferably by using an information criterion, such as the AICc. It would be very interesting to perform a systematic study of the performance of our approach to modeling event count time series, in comparison with the already available model classes; this, however, is beyond the scope of the present paper and remains a task for future work.

With respect to the chosen field of application, namely, the analysis of time series of the daily number of seizures of drug-resistant epilepsy patients, in the future we intend to apply the presented algorithm to data from larger cohorts of patients. By discriminating between different types of seizures in the same patient, it is also intended to generalize the analysis to the case of vector data. Furthermore, the analysis should be generalized such that not only the effects of individual anti-epileptic drugs are modeled, but also the interaction effects between pairs, or groups, of drugs. Finally, in order to reduce the computational time

consumption, work should be devoted to developing more efficient algorithms for fitting ensembles of nonlinear state space models by numerical maximum-likelihood procedures.

Abbreviations

The following abbreviations are used in this paper:

AED	Anti-epileptic drug
IC-LSS	Independent components linear state space
NLSS	Nonlinear state space
AR	Autoregressive
ARMA	Autoregressive moving average
IEKF	Iterated extended Kalman filter
SVD	Singular value decomposition
CD	Cholesky decomposition
BFGS	Broyden–Fletcher–Goldfarb–Shanno
AICc	corrected Akaike information criterion

Appendix A. Block Diagonal Structure of A and Q

The IC-LSS model discussed in Section 2.1 is based on a block-diagonal structure [35] of the state transition matrix, A, and the dynamical noise covariance matrix, Q, as follows [27]:

$$
A = \begin{pmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ldots & A_J \end{pmatrix}, \quad Q = \begin{pmatrix} Q_1 & 0 & 0 & \ldots & 0 \\ 0 & Q_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ldots & Q_J \end{pmatrix} \tag{A1}
$$

where J denotes the number of independent components and also represents the number of blocks. If we assume that the jth block represents an autoregressive moving average (ARMA) model with model orders $(p, p-1)$, consisting of an autoregressive part with parameters $a_\tau^{(j)}$, $\tau = 1, \ldots, p$, and a moving average part with parameters $b_\tau^{(j)}$, $\tau = 0, \ldots, p-1$, the block matrix, A_j, has left companion form, given by [27]

$$
A_j = \begin{pmatrix}
a_1^{(j)} & 1 & 0 & \dots & 0 \\
a_2^{(j)} & 0 & 1 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & 0 \\
a_{p-1}^{(j)} & 0 & 0 & \dots & 1 \\
a_p^{(j)} & 0 & 0 & \dots & 0
\end{pmatrix} \tag{A2}
$$

and the block matrix, Q_j, has outer product form, given by [27]

$$
Q_j = \begin{pmatrix}
(b_0^{(j)})^2 & b_0^{(j)} b_1^{(j)} & \dots & b_0^{(j)} b_q^{(j)} \\
b_1^{(j)} b_0^{(j)} & (b_1^{(j)})^2 & \dots & b_1^{(j)} b_q^{(j)} \\
\vdots & \vdots & \vdots & \ddots \\
b_{p-2}^{(j)} b_0^{(j)} & b_{p-2}^{(j)} b_1^{(j)} & \dots & b_{p-2}^{(j)} b_q^{(j)} \\
b_{p-1}^{(j)} b_0^{(j)} & b_{p-1}^{(j)} b_1^{(j)} & \dots & (b_{p-1}^{(j)})^2
\end{pmatrix} \tag{A3}
$$

We usually employ the scaling convention $b_0^{(j)} = 1$ for all blocks.

Appendix B. Iterated Extended Kalman Filter (IEKF) Algorithms

---

Algorithm A1 Iterated Extended Kalman Filter (IEKF)

Require: Initial state estimate: $\mathbf{x}_{0|0}$

Require: Initial covariance matrix: $P_{0|0}$

1: **for** $t = 1$ to T **do**

2:     **Prediction Step:**

3:     Compute the predicted state estimate: $\mathbf{x}_{t|t-1} = A\mathbf{x}_{t-1|t-1} + B_u \mathbf{u}_t$

4:     Compute the predicted state covariance matrix: $P_{t|t-1} = A P_{t-1|t-1} A^T + Q$

5:     **Update Step:**

6:     Initialize iteration: $i = 1$, $\mathbf{x}_{t|t}^{(0)} = \mathbf{x}_{t|t-1}$

7:     **repeat**

8:         Compute the Jacobian matrix of the observation function: $H_t^{(i)} = \left. \dfrac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}_{t|t}^{(i-1)}}$

9:         Compute the innovation: $\boldsymbol{\nu}_t^{(i)} = \mathbf{y}_t - \mathbf{f}(\mathbf{x}_{t|t}^{(i-1)})$

10:        Compute the innovation covariance matrix: $V_t^{(i)} = H_t^{(i)} P_{t|t-1} (H_t^{(i)})^T + R$

11:        Compute the Kalman gain: $K_t^{(i)} = P_{t|t-1} (H_t^{(i)})^T (V_t^{(i)})^{-1}$

12:        Update the state estimate: $\mathbf{x}_{t|t}^{(i)} = \mathbf{x}_{t|t-1} + K_t^{(i)} \left(\boldsymbol{\nu}_t^{(i)} - H_t^{(i)} \left(\mathbf{x}_{t|t-1} - \mathbf{x}_{t|t}^{(i-1)} \right) \right)$

13:        Increment iteration: $i = i + 1$

14:     **until** Stopping criterion met or maximum number of iterations (i_m) reached

15:     Compute the filtered state covariance matrix: $P_{t|t} = \left(I_m - K_t^{(i)} H_t^{(i)} \right) P_{t|t-1}$

16: **end for**

Algorithm A2 Singular Value Decomposition Iterated Extended Kalman Filter (SVD-IEKF)

Require: Initial state estimate: $\mathbf{x}_{0|0}$

Require: Initial covariance matrix: $P_{0|0} = W_{0|0}\Sigma_{0|0}^2 W_{0|0}^T$

Require: Factorized dynamic noise covariance matrix: $Q = W_Q\Sigma_Q^2 W_Q^T$

Require: Factorized observation noise covariance matrix: $R = W_R\Sigma_R^2 W_R^T$

1: **for** $t = 1$ to T **do**

2:     **Prediction Step:**

3:     Compute the predicted state estimate: $\mathbf{x}_{t|t-1} = A\mathbf{x}_{t-1|t-1} + B_u\mathbf{u}_t$

4:     Compute the predicted state covariance matrix:

$$U_1 \begin{bmatrix} \Sigma_{P_{t|t-1}} \\ 0 \end{bmatrix} W_{P_{t|t-1}}^T = \begin{bmatrix} \Sigma_{P_{t-1|t-1}} W_{P_{t-1|t-1}}^T A^T \\ \Sigma_Q W_Q^T \end{bmatrix}$$

$$P_{t|t-1} = W_{P_{t|t-1}} \Sigma_{P_{t|t-1}}^2 W_{P_{t|t-1}}^T$$

5:     **Update Step:**

6:     Initialize iteration: $i = 1$, $\mathbf{x}_{t|t}^{(0)} = \mathbf{x}_{t|t-1}$

7:     **repeat**

8:        Compute the Jacobian matrix of the observation function: $H_t^{(i)} = \dfrac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}_{t|t}^{(i-1)}}$

9:        Compute the innovation: $\boldsymbol{\nu}_t^{(i)} = \mathbf{y}_t - \mathbf{f}\big(\mathbf{x}_{t|t}^{(i-1)}\big)$

10:      Compute the innovation covariance matrix:

$$U_2 \begin{bmatrix} \Sigma_{V_t}^{(i)} \\ 0 \end{bmatrix} \big(W_{V_t}^{(i)}\big)^T = \begin{bmatrix} \Sigma_R W_R^T \\ \Sigma_{P_{t|t-1}}^{(i)} \big(W_{P_{t|t-1}}^{(i)}\big)^T \big(H_t^{(i)}\big)^T \end{bmatrix}$$

$$V_t^{(i)} = W_{V_t}^{(i)} \big(\Sigma_{V_t}^{(i)}\big)^2 \big(W_{V_t}^{(i)}\big)^T$$

11:      Compute normalized innovation: $\widehat{\boldsymbol{\nu}}_t^{(i)} = \big(W_{V_t}^{(i)}\big)^T \boldsymbol{\nu}_t^{(i)}$

12:      Compute normalized gain: $k_t^{(i)} = P_{t|t-1}\big(H_t^{(i)}\big)^T W_{V_t}^{(i)}$

13:      Compute the optimal Kalman gain: $K_t^{(i)} = k_t^{(i)}\big(\Sigma_{V_t}^{(i)}\big)^{-2}\big(W_{V_t}^{(i)}\big)^T$

14:      Update the state estimate:
$$\mathbf{x}_{t|t}^{(i+1)} = \mathbf{x}_{t|t-1} + k_t^{(i)}\big(\Sigma_{V_t}^{(i)}\big)^{-2}\Big(\widehat{\boldsymbol{\nu}}_t^{(i)} - \big(W_{V_t}^{(i)}\big)^T H_t^{(i)}\big(\mathbf{x}_{t|t-1} - \mathbf{x}_{t|t}^{(i)}\big)\Big)$$

15:      Increment iteration: $i = i + 1$

16:     **until** Stopping criterion met or maximum number of iterations (i_m) reached

17:     Compute the filtered state covariance matrix:

$$U_3 \begin{bmatrix} \Sigma_{P_{t|t}} \\ 0 \end{bmatrix} W_{P_{t|t}}^T = \begin{bmatrix} \Sigma_{P_{t|t-1}} W_{P_{t|t-1}}^T \big(I_m - K_t^{(i)} H_t^{(i)}\big)^T \\ \Sigma_R W_R^T \big(K_t^{(i)}\big)^T \end{bmatrix}$$

$$P_{t|t} = W_{P_{t|t}} \Sigma_{P_{t|t}}^2 W_{P_{t|t}}^T$$

18: **end for**

Appendix C. Matlab Code for Generating Simulated Data

Listing A1. MATLAB code for generating simulated data

```matlab
x = zeros(m, n); % state vector according to  Equation (3).
x_initial = zeros(m, 1); % initializing sate vector with zero.
k = 1; % parameter defined in Equation (6).
C = [-0.40, 0.95, -0.70, 0.75, 0];
A = [0.5, 0, 0, 0, 0;
     0, 0.25, 0, 0, 0;
     0, 0, 0.25, 0, 0;
     0, 0, 0, 0.9, 1;
     0, 0, 0, -0.5, 0];
ϵ = [zeros(1, u), randn(1), zeros(1, 1)]; % observation noise from
    Equation (6).


for(time = 1: T)  % Begin Kalman forward loop

    external_input = [u_t(:, time:-1:T, 1)' , zeros(1, (m - u))]; %
        external input defined in Equation (4)
    if t == 1
        x(:, t) = A * x_initial + external_input' + ϵ';
    else
        x(:, t) = A * x(:, t - 1) + external_input' + ϵ';
    end

    for seizure_type = 1:n % for multi-dimensional data
        srexpr = sqrt(((C(seizure_type, :) * x(:, t))^2) / 4 + k); %
            intermediate variable
        y(seizure_type, t) = round((C(seizure_type, :) * x(:, t)) /
            2 + srexpr);
    end
end % End Kalman forward loop
```

References

1. Nelder, J.; Wedderburn, R. Generalized linear models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [CrossRef]
2. West, M.; Harrison, P.J.; Migon, H.S. Dynamic generalized linear models and Bayesian forecasting. *J. Am. Stat. Assoc.* **1985**, *80*, 73–83. [CrossRef]
3. Fahrmeir, L.; Tutz, G.; Hennevogl, W.; Salem, E. *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed.; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2001; Volume 425.
4. McKenzie, E. Some simple models for discrete variate time series. *J. Am. Water Resour. Assoc.* **1985**, *21*, 645–650. [CrossRef]
5. Rydberg, T.H.; Shephard, N. *BIN Models for Trade-by-Trade Data. Modelling the Number of Trades in a Fixed Interval of Time*; Technical Report; Nuffield College: Oxford, UK, 1999.
6. Weiß, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
7. Liu, M.; Zhu, F.; Li, J.; Sun, C. A Systematic Review of INGARCH Models for Integer-Valued Time Series. *Entropy* **2023**, *25*, 922. [CrossRef] [PubMed]
8. Kailath, T.; Sayed, A.; Hassibi, B. *Linear Estimation*; Prentice Hall: Hoboken, NY, USA, 2000.
9. Galka, A.; Ozaki, T.; Muhle, H.; Stephani, U.; Siniatchkin, M. A data-driven model of the generation of human EEG based on a spatially distributed stochastic wave equation. *Cogn. Neurodyn.* **2008**, *2*, 101–113. [CrossRef]
10. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NY, USA, 1994.
11. Grewal, M.S.; Andrews, A.P. *Kalman Filtering: Theory and Practice with MATLAB*, 4th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015.
12. Skoglund, M.; Hendeby, G.; Axehill, D. Extended Kalman filter modifications based on an optimization view point. In Proceedings of the 18th International Conference on Information Fusion, Washington, DC, USA, 6–9 July 2015; pp. 1856–1861.
13. Wang, L.; Libert, G.; Manneback, P. Kalman filter algorithm based on singular value decomposition. In Proceedings of the 31st IEEE Conference on Decision and Control, Tucson, AZ, USA, 16–18 December 1992; pp. 1224–1229.
14. Zeger, S.L.; Irizarry, R.; Peng, R.D. On time series analysis of public health and biomedical data. *Annu. Rev. Public Health* **2006**, *27*, 57–79. [CrossRef]

15. Zeger, S.L. A regression model for time series of counts. *Biometrika* **1988**, *75*, 621–629. [CrossRef]
16. Allard, R. Use of time-series analysis in infectious disease surveillance. *Bull. World Health Organ.* **1998**, *76*, 327.
17. Albert, P.S. A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **1991**, *47*, 1371–1381. [CrossRef]
18. Balish, M.; Albert, P.S.; Theodore, W.H. Seizure frequency in intractable partial epilepsy: A statistical analysis. *Epilepsia* **1991**, *32*, 642–649. [CrossRef]
19. Krauss, G.L.; Sperling, M.R. Treating patients with medically resistant epilepsy. *Neurol. Clin. Pract.* **2011**, *1*, 14–23. [CrossRef] [PubMed]
20. Kanner, A.M.; Ashman, E.; Gloss, D.; Harden, C.; Bourgeois, B.; Bautista, J.F.; Abou-Khalil, B.; Burakgazi-Dalkilic, E.; Park, E.L.; Stern, J.; et al. Practice guideline update summary: Efficacy and tolerability of the new antiepileptic drugs I: Treatment of new-onset epilepsy: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the American Epilepsy Society. *Neurology* **2018**, *91*, 74–81. [PubMed]
21. Tharayil, J.J.; Chiang, S.; Moss, R.; Stern, J.M.; Theodore, W.H.; Goldenholz, D.M. A big data approach to the development of mixed-effects models for seizure count data. *Epilepsia* **2017**, *58*, 835–844. [CrossRef] [PubMed]
22. Chiang, S.; Vannucci, M.; Goldenholz, D.M.; Moss, R.; Stern, J.M. Epilepsy as a dynamic disease: A Bayesian model for differentiating seizure risk from natural variability. *Epilepsia Open* **2018**, *3*, 236–246. [CrossRef]
23. Wang, E.; Chiang, S.; Cleboski, S.; Rao, V.; Vannucci, M.; Haneef, Z. Seizure count forecasting to aid diagnostic testing in epilepsy. *Epilepsia* **2022**, *63*, 3156–3167. [CrossRef]
24. Galka, A.; Boor, R.; Doege, C.; von Spiczak, S.; Stephani, U.; Siniatchkin, M. Analysis of epileptic seizure count time series by ensemble state space modelling. In Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 5601–5605.
25. Moontaha, S.; Galka, A.; Meurer, T.; Siniatchkin, M. Analysis of the effects of medication for the treatment of epilepsy by ensemble Iterative Extended Kalman filtering. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 187–190.
26. Moontaha, S.; Galka, A.; Siniatchkin, M.; Scharlach, S.; von Spiczak, S.; Stephani, U.; May, T.; Meurer, T. SVD Square-root Iterated Extended Kalman Filter for Modeling of Epileptic Seizure Count Time Series with External Inputs. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 616–619.
27. Galka, A.; Wong, K.F.K.; Ozaki, T.; Muhle, H.; Stephani, U.; Siniatchkin, M. Decomposition of neurological multivariate time series by state space modelling. *Bull. Math. Biol.* **2011**, *73*, 285–324. [CrossRef]
28. Weiß, C.H.; Zhu, F.; Hoshiyar, A. Softplus INGARCH models. *Stat. Sin.* **2022**, *32*, 1099–1120. [CrossRef]
29. Kalman, R.E.; Bucy, R.S. New results in linear filtering and prediction theory. *J. Basic Eng.* **1961**, *83*, 95–108. [CrossRef]
30. Jazwinski, A. *Stochastic Processes and Filtering Theory*; Academic Press: Cambridge, MA, USA, 1974.
31. Björck, Å. *Numerical Methods in Matrix Computations*; Texts in Applied Mathematics; Springer: Berlin/Heidelberg, Germany, 2015; Volume 59.
32. Kulikova, M.V.; Tsyganova, J.V. Improved discrete-time Kalman filtering within singular value decomposition. *IET Control Theory Appl.* **2017**, *11*, 2412–2418. [CrossRef]
33. Nocedal, J.; Wright, S. *Numerical Optimization*; Springer: Berlin/Heidelberg, Germany, 1999.
34. Kulikova, M.V.; Kulikov, G.Y. SVD-based factored-form Cubature Kalman Filtering for continuous-time stochastic systems with discrete measurements. *Automatica* **2020**, *120*, 109110. [CrossRef]
35. Godolphin, E.; Triantafyllopoulos, K. Decomposition of time series models in state-space form. *Comput. Stat. Data Anal.* **2006**, *50*, 2232–2246. [CrossRef]