# Robust Variational-based Kalman Filter for Outlier Rejection with Correlated Measurements

Haoqing Li, Daniel Medina, Jordi Vilà-Valls, Pau Closas

## HAL Id: hal-03108748
## https://hal.science/hal-03108748

Submitted on 13 Jan 2021

**To cite this version :**

Li, Haoqing and Medina, Daniel and Vila-Valls, Jordi and Closas, Pau Robust Variational-based Kalman Filter for Outlier Rejection with Correlated Measurements. (2021) IEEE Transactions on Signal Processing, 69. 357-369. ISSN 1053-587X

# Robust Variational-Based Kalman Filter for Outlier Rejection With Correlated Measurements

Haoqing Li ⓘ, Daniel Medina ⓘ, *Student Member, IEEE*, Jordi Vilà-Valls ⓘ, *Senior Member, IEEE*,
and Pau Closas ⓘ, *Senior Member, IEEE*

*Abstract*—State estimation is a fundamental task in many engineering fields, and therefore robust nonlinear filtering techniques able to cope with misspecified, uncertain and/or corrupted models must be designed for real-life applicability. In this contribution we explore nonlinear Gaussian filtering problems where measurements may be corrupted by outliers, and propose a new robust variational-based filtering methodology able to detect and mitigate their impact. This method generalizes previous contributions to the case of multiple outlier indicators for both independent and dependent observation models. An illustrative example is provided to support the discussion and show the performance improvement.

*Index Terms*—Robust filtering, variational Bayes, outliers, heavy-tailed noise, correlated measurements.

## I. INTRODUCTION

STATE estimation is a fundamental task in a plethora of engineering fields, ranging from robotics, guidance and navigation systems, to information fusion or time-series analysis [1]–[3]. For linear dynamic state-space models, the linear minimum mean square error (MSE) estimate is given by the well known Kalman filter (KF), which is optimal under nominal conditions, that is, *known system matrices*, *known noise statistics* and *perfect initialization* [4]. For nonlinear systems, the most popular approaches are the family of sigma-point Gaussian filters (SPGF) under the Gaussian assumption [5], [6] and sequential Monte Carlo (SMC) methods [7], [8] for general non-Gaussian models. A fundamental problem of all these well established techniques is their lack of robustness in case of model mismatch (i.e., misspecified noise statistics parameters, unexpected impulsive/heavy-tailed behaviours, or different types of outliers in the state and/or observations), which induces a bias and MSE degradation [9], or in high dimensional

Haoqing Li and Pau Closas are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: li.haoq@northeastern.edu; closas@northeastern.edu).

Daniel Medina is with the Institute of Communications and Navigation, German Aerospace Center (DLR), 17235 Neustrelitz, Germany (e-mail: daniel.ariasmedina@dlr.de).

Jordi Vilà-Valls is with the ISAE-SUPAERO/University of Toulouse, 31055 Toulouse, France (e-mail: jordi.vila-valls@isae-supaero.fr).

systems [10]. For the latter, a possible solution is to resort to marginalization strategies (i.e., Rao-Blackwellization) to reduce the dimension of the space to be explored by the nonlinear filter [11]. That is the main reason why there is a continued effort and an actual need to develop robust filtering techniques for real-life applicability.

To counteract the lack of knowledge for a correct filter initialization one can set a distortionless constraint, leading to the minimum variance distortionless response estimator [12], [13]. In the context of constrained filtering, a linearly constrained KF has been recently proposed to cope with process and measurement matrices model mismatch [14]. Regarding the problems related to high-dimensional spaces, a possible solution is to use a divide and conquer strategy, concept which has been applied to both SPGF [15] and SMC methods [16], or the use of Rao-Blackwellization techniques in case of mixed linear/nonlinear systems [11].

In this contribution, we are interested in nonlinear/Gaussian filtering problems with outlier (correlated) measurements. That is, state-space models involving Gaussian distributions with known parameters under nominal situations ($\mathcal{M}_0$) and with additional outliers under non-nominal conditions ($\mathcal{M}$). More precisely, we consider models of the form of

$$\boldsymbol{x}_t = \boldsymbol{f}\left(\boldsymbol{x}_{t-1}\right) + \boldsymbol{\epsilon}_t \tag{1}$$

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{h}\left(\boldsymbol{x}_t\right) + \boldsymbol{\eta}_t, & \text{under } \mathcal{M}_0 \\ \boldsymbol{h}\left(\boldsymbol{x}_t\right) + \boldsymbol{\eta}_t + \boldsymbol{o}_t, & \text{under } \mathcal{M} \end{cases} \tag{2}$$

where $\boldsymbol{x}_t \in \mathbb{R}^{n_x}$ represents hidden (i.e. to be estimated) state of the system; $\boldsymbol{y}_t = (y_t^{(1)}, \ldots, y_t^{(n_y)})^\top \in \mathbb{R}^{n_y}$ is the corresponding observation, whose elements conditioned on $\boldsymbol{x}_t$ might be correlated; $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t)$ is the process noise; $\boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_t)$ is the measurement noise; $\boldsymbol{o}_t \in \mathbb{R}^{n_y}$ represents outliers on some or all observations in $\boldsymbol{y}_t$; and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The mappings $\boldsymbol{h}(\cdot)$ and $\boldsymbol{f}(\cdot)$ are referred to as the process and measurement functions. Several solutions can be found to tackle this problem under different perspectives, namely:

- Gaussian: consider that the noise statistics are Gaussian but with unknown covariance matrices. In that case one can resort to standard innovations-based techniques [17] or variational Bayesian (VB) inference [18]–[20].
- Non-Gaussian: consider that the noise is heavy-tailed and obtain a robust filtering solution via VB approximations [21], [22], or exploiting a hierarchically Gaussian

formulation to obtain approximate Gaussian filtering solutions. For the latter it is possible to use Rao-Blackwellization [23], conjugate prior analysis [24] or for linear systems to resort again to VB approximations [25]–[29].

- Nonparametric: consider that the noise distribution is unknown (in contrast to the previous parametric appoaches) and resort to nonparametric Bayesian estimation techniques such as Dirichlet Process Mixtures [30].
- Robust statistics: consider the standard contamination model arising from robust statistics where a proportion $1 - \varepsilon$ of observations follows a nominal Gaussian noise distribution, and another proportion $0 \leq \varepsilon \leq 1$ of observations is contaminated by an unknown distribution [31]. Within this framework, the KF can be reformulated as a regression problem and then solved via iterative M-estimation techniques [32]–[34].
- Detect-and-reject: recently a probabilistic outlier rejection method was proposed in [35], where the goal is to have a binary indicator to decide whether a (vector) measurement arises from the nominal noise distribution or it is a contaminated observation (i.e., outlier). This approach relies on VB inference to jointly estimate the state vector and an indicator which informs whether the observed data is legitimate or contaminated.

In this contribution we further elaborate on the methodology proposed in [35] and extend it to more general problems. One of the main limitations of that work is that even if it allows treatment of multi-dimensional observations, a scalar indicator is considered for the entire observation vector, thus accepting/rejecting in its totality regardless of which entry is corrupted. In this article, the method is generalized to the more comprehensive case of multiple outlier indicators, accounting for both independent and correlated observation models (a common problem for RADAR processing [36] or data gathering in sensor networks [37]–[39]).

We first introduce the proposed new framework in Section II-A for the general case of correlated data, following a VB approximation. This is then particularized to independent data in Section II-B, showing that it is indeed a generalization of the work in [40]. Remarkably, the relation to the original method [35] where a single scalar indicator was used is detailed in Section II-C, showing as well the generalization of this contribution. An illustrative example is provided in Section III where correlated data is used. More precisely, a code-based Global Navigation Satellite System (GNSS) positioning problem [41] is considered where double-difference range estimates are used (causing the observations to become correlated). Finally, conclusions are discussed in Section IV.

*Notation:* Italic indicates a scalar quantity, as in $c$; lower case boldface indicates a column vector quantity, as in $\boldsymbol{a}$; upper case boldface indicates a matrix quantity, as in $\boldsymbol{A}$. The matrix/vector transpose is indicated by a superscript $(\cdot)^\top$ as in $\boldsymbol{A}^\top$. The trace operator is denoted as $\mathrm{Tr}(\boldsymbol{A})$ and $\langle \boldsymbol{A} \rangle$ represents the expectation of $\boldsymbol{A}$. $\mathbb{E}_b\{g(a)\}$ represents the expectation of $g(a)$ over the distribution of $b$, and we use equivalently $\langle g(a) \rangle$ for simplicity where $b$ is omitted. $||\mathbf{b}||^2_{\mathbf{A}^{-1}} = \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$ denotes the weighted inner norm. The $i-$th row and $j-$th column element of the matrix $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{i,j}$. The $i-$th element of the column vector $\boldsymbol{a}$ will denoted by $a^{(i)}$, $\boldsymbol{a}^{(-i)}$ is the vector of all elements in $\boldsymbol{a}$ except for $a^{(i)}$, and $[\boldsymbol{A}]_{-i,-j}$ is defined as a matrix with all elements in matrix $\mathbf{A}$ except for the $i$-th row and $j$-th column.

## II. VARIATIONAL BAYES KALMAN FILTER FOR OUTLIER REJECTION IN GENERAL NONLINEAR/GAUSSIAN SYSTEMS

In this section, we describe the proposed Variational Bayes Kalman filter (VBKF) method for outlier detection and mitigation, which generalizes the work in [35] to the case of multiple outlier indicators and correlated observations. As in the original method, the new VBKF estimates the probability of having outliers in the observation at every time step, and then down-weights that particular observation. Notice that, in the original algorithm, a single indicator $z_t$ was used for all observations gathered at the same time, $\boldsymbol{y}_t$. In practice, outliers might affect differently the elements in $\boldsymbol{y}_t$ which ignited the contribution in this paper.

### A. Generalized VBKF for Correlated Data

Following the approach in [35] to derive a VBKF algorithm, the likelihood distribution is modified to incorporate the effect of an outlier indicator vector $\boldsymbol{z}_t = (z_t^{(1)}, \dots, z_t^{(n_y)})^\top \in \mathcal{Z} = \{0, 1\}^{n_y}$, such that $z_t^{(i)} = 0$ if there is an outlier on the $i$-th (corrupted) element of $\boldsymbol{y}_t$, i.e., $y_t^{(i)}$, and $z_t^{(i)} = 1$ if the $i$-th element is otherwise clean (not corrupted). In the latter, the nominal Gaussian modeling would apply, whereas in the former the wrong information brought by $y_t^{(i)}$ must be down-weighted. Therefore, the observation model in (2) is modified to incorporate those indicators, becoming

$$p\left(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{z}_t\right) = \frac{1}{c(\boldsymbol{z}_t)} e^{-\frac{1}{2}||\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)||^2_{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)}} \quad (3)$$

where $\boldsymbol{\Sigma}_t(\boldsymbol{z}_t)$ is

$$\boldsymbol{\Sigma}_t(\boldsymbol{z}_t) = \begin{bmatrix} \sigma_{1,1}^2/z_t^{(1)} & \sigma_{1,2}^2 & \cdots & \sigma_{1,n_y}^2 \\ \sigma_{2,1}^2 & \sigma_{2,2}^2/z_t^{(2)} & \cdots & \sigma_{2,n_y}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_y,1}^2 & \sigma_{n_y,2}^2 & \cdots & \sigma_{n_y,n_y}^2/z_t^{(n_y)} \end{bmatrix}, \quad (4)$$

which corresponds to the original matrix $\boldsymbol{R}_t$ with (independent) indicators $\{z_t^{(i)}\}_{i=1}^{n_y}$ dividing its diagonal terms, such that $\boldsymbol{R}_t = \boldsymbol{\Sigma}_t(\mathbf{1})$. The dependence on time of the various elements, $[\boldsymbol{R}_t]_{i,j} \triangleq \sigma_{i,j}^2$, has been omitted for notation convenience. Contrary to what could be intuitively expected, the division by zero when an indicator is zero does not cause any numerical issue. As explained in this section, this is a consequence of the algorithm operating on the precision matrix. In that case, it can be shown using basic algebra manipulations that the term in the exponent of the likelihood can be equivalently reinterpreted as

$$||\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)||^2_{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)} = ||\mathcal{T}(\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t), \boldsymbol{z}_t)||^2_{\mathbf{C}_t^{-1}(\boldsymbol{z}_t)} \quad (5)$$

where $\mathcal{T}(\cdot, \boldsymbol{z}_t) : \mathbb{R}^{n_y} \mapsto \mathbb{R}^{n_y'}$ is an operator that removes the elements in the input vector corresponding to indicators valued zero, we define the limiting case $\mathcal{T}(\cdot, \mathbf{0}) = 1$. Matrix $\mathbf{C}(\boldsymbol{z}_t)$
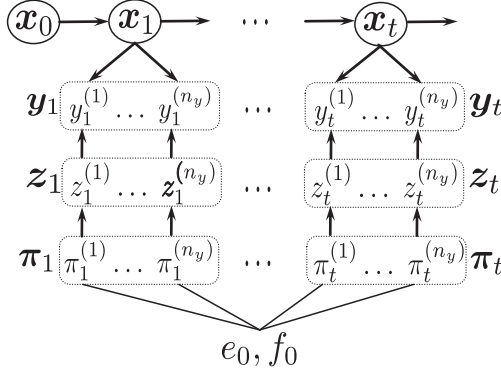
Fig. 1. Graphical model considered in the VBKF algorithm.

is a transformation of $\boldsymbol{\Sigma}_t(\boldsymbol{z}_t)$ where the rows/columns corresponding to those $z_t^{(i)} = 0$ are removed. Special cases are $i)$ $\mathbf{C}(\mathbf{1}) = \boldsymbol{\Sigma}_t(\mathbf{1}) = \boldsymbol{R}_t$ (resulting in the original model without indicators), and $ii)$ $\mathbf{C}(\mathbf{0})$, corresponding to the absence of measurements and defined as $\mathbf{C}(\mathbf{0}) = \mathbf{I}$. Therefore, the dimension of the resulting multivariate normal variable ($n_y'$) is effectively reduced by the amount of zero indicators: $n_y' = \sum_{i=1}^{n_y} z_t^{(i)} \leq n_y$, with equality when all indicators are one. Notice that $c(\boldsymbol{z}_t)$, the proportionality term in (3), depends on the indicators and can be readily computed from (5) as

$$c(\boldsymbol{z}_t) = \sqrt{(2\pi)^{n_y'} |\mathbf{C}(\boldsymbol{z}_t)|}. \tag{6}$$

Therefore, the likelihood distribution (3) can be expressed as $\mathcal{N}(\mathcal{T}(\boldsymbol{y}_t, \boldsymbol{z}_t); \mathcal{T}(\boldsymbol{h}(\boldsymbol{x}_t), \boldsymbol{z}_t), \mathbf{C}(\boldsymbol{z}_t))$.

As a consequence of introducing the indicators $\boldsymbol{z}_t$, we need to estimate those jointly with the state trajectory $\boldsymbol{x}_{0:t}$. To accomplish this task in a Bayesian sense, we impose a beta-Bernoulli hierarchical prior to each individual indicator,

$$p\left(z_t^{(i)} | \pi_t^{(i)}\right) = \left(\pi_t^{(i)}\right)^{z_t^{(i)}} \left(1 - \pi_t^{(i)}\right)^{1 - z_t^{(i)}}, \tag{7}$$

where $\pi_t^{(i)}$ is a beta distributed random variable which is parameterized by (unknown shape hyper-parameters)[1] $e_0^{(i)}$ and $f_0^{(i)}$,

$$p\left(\pi_t^{(i)}\right) = \frac{\left(\pi_t^{(i)}\right)^{e_0^{(i)}-1} \left(1 - \pi_t^{(i)}\right)^{f_0^{(i)}-1}}{\beta\left(e_0^{(i)}, f_0^{(i)}\right)}, \tag{8}$$

and $\beta(\cdot, \cdot)$ is the beta function. Notice that we are assuming that the indicators are mutually independent

$$p(\boldsymbol{z}_t, \boldsymbol{\pi}_t) = \prod_{i=1}^{n_y} p\left(z_t^{(i)} | \pi_t^{(i)}\right) p\left(\pi_t^{(i)}\right), \tag{9}$$

as well as independent from the observations since the underlying statistics modeling the outliers do not depend on the actual values of the data. These assumptions determines the graphical model shown in Fig. 1.

According to the Variational Inference (VI) principle [42], [43], to estimate the posterior distribution of the latent variables

[1]Notice that a beta distribution is generically defined as $p(x; \alpha, \gamma) \propto x^{\alpha-1}(1-x)^{\gamma-1}$, where $\alpha$ and $\gamma$ are two shape parameters. For simplicity we drop the dependence on $\alpha$ and $\gamma$ and thus we write $p(x)$.

$\boldsymbol{\theta} = \{\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t\}$, that is $p(\boldsymbol{\theta} | \boldsymbol{y}_{1:t})$, we can resort to an auxiliary distribution $q(\boldsymbol{\theta})$ such that:

$$q(\boldsymbol{\theta}) = q(\boldsymbol{x}_t) q(\boldsymbol{\pi}_t) q(\boldsymbol{z}_t)$$
$$= q(\boldsymbol{x}_t) \prod_{i=1}^{n_y} q\left(\pi_t^{(i)}\right) q\left(z_t^{(i)}\right). \tag{10}$$

According to Fig. 1, $\boldsymbol{y}_t$ is conditionally independent on $\boldsymbol{\pi}_t$; $\boldsymbol{x}_t$ is conditionally independent on $\boldsymbol{z}_t$ and $\boldsymbol{\pi}_t$; $\boldsymbol{y}_{1:t-1}$ is conditionally independent on $\boldsymbol{z}_t$, $\boldsymbol{\pi}_t$ and $\boldsymbol{y}_t$. Thus, the various marginal distributions, $q(\cdot)$, are then obtained from the mean-field VI method applied to the joint distribution

$$p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t}) \propto p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1}) p(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{z}_t) p(\boldsymbol{z}_t, \boldsymbol{\pi}_t), \tag{11}$$

such that

$$\ln[q(\boldsymbol{x}_t)] = \mathbb{E}_{\boldsymbol{\pi}_t, \boldsymbol{z}_t}\{\ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})]\}, \tag{12}$$

$$\ln[q(\boldsymbol{\pi}_t)] = \mathbb{E}_{\boldsymbol{x}_t, \boldsymbol{z}_t}\{\ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})]\}, \tag{13}$$

$$\ln[q(\boldsymbol{z}_t)] = \mathbb{E}_{\boldsymbol{\pi}_t, \boldsymbol{x}_t}\{\ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})]\}. \tag{14}$$

Within the Gaussian filtering framework, the first term $p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1})$ on the right-hand side of (11) is a predictive density, which can be approximated as $p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1}) \approx \mathcal{N}(\hat{\boldsymbol{x}}_{t|t-1}, \boldsymbol{P}_{t|t-1})$, where the corresponding mean and covariance are [6]

$$\hat{\boldsymbol{x}}_{t|t-1} = \int \boldsymbol{f}(\boldsymbol{x}_{t-1}) p(\boldsymbol{x}_{t-1} | \boldsymbol{y}_{1:t-1}) d\boldsymbol{x}_{t-1}, \tag{15}$$

$$\boldsymbol{P}_{t|t-1} = \int \left(\boldsymbol{f}(\boldsymbol{x}_{t-1}) - \hat{\boldsymbol{x}}_{t|t-1}\right) \left(\boldsymbol{f}(\boldsymbol{x}_{t-1}) - \hat{\boldsymbol{x}}_{t|t-1}\right)^{\top}$$
$$\times p(\boldsymbol{x}_{t-1} | \boldsymbol{y}_{1:t-1}) d\boldsymbol{x}_{t-1} + \boldsymbol{Q}_t, \tag{16}$$

with $\hat{\boldsymbol{x}}_{t-1|t-1}$ and $\boldsymbol{P}_{t-1|t-1}$ the mean and covariance of the filtering posterior at $t-1$, that is $p(\boldsymbol{x}_{t-1} | \boldsymbol{y}_{1:t-1}) \approx \mathcal{N}(\hat{\boldsymbol{x}}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1})$. In the sequel we derive the update of the terms in (10).

*1) Update $q(\boldsymbol{x}_t)$:* According to the mean-field VI method, $q(\boldsymbol{x}_t)$ is obtained from (12) as

$$\ln(p(\boldsymbol{x}_t, \boldsymbol{\pi}, \boldsymbol{z}_t, \boldsymbol{y}_{1:t}))$$
$$= -\frac{1}{2}||\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}||_{\boldsymbol{P}_{t|t-1}^{-1}}^2 - \frac{1}{2}||\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)||_{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)}^2 + \kappa \tag{17}$$

$$\ln(q(\boldsymbol{x}_t))$$
$$= \mathbb{E}_{\boldsymbol{\pi}_t, \boldsymbol{z}_t}\{\ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})]\} \tag{18}$$

$$= -\frac{1}{2}||\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}||_{\boldsymbol{P}_{t|t-1}^{-1}}^2 \tag{19}$$
$$- \frac{1}{2} \sum_{\boldsymbol{z} \in \mathcal{Z}} q(\boldsymbol{z}_t = \boldsymbol{z}) ||\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)||_{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)}^2 + \kappa$$

$$= -\frac{1}{2}||\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}||_{\boldsymbol{P}_{t|t-1}^{-1}}^2$$
$$- \frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t))^{\top} \langle \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t) \rangle (\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)) + \kappa, \tag{20}$$

where the term $\kappa$ gathers the logarithm of those factors that are constant in the expression, and $\boldsymbol{z}$ represents one of the $2^{n_y}$ possible combinations of $\{z_t^{(i)}\}_{i=1}^{n_y}$ binary values; the set of all those possible values is given by $\mathcal{Z} = \{0,1\}^{n_y}$ such that $|\mathcal{Z}| = 2^{n_y}$; The expectation of $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)$ with respect to $q(\boldsymbol{z}_t)$ is defined as

$$\langle \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t) \rangle = \sum_{\boldsymbol{z} \in \mathcal{Z}} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}) \, q(\boldsymbol{z}_t = \boldsymbol{z}), \tag{21}$$

and $q(\boldsymbol{z}_t = \boldsymbol{z}) = \prod_{i=1}^{n_y} q(z_t^{(i)} = z^{(i)})$.

Thus, the estimation of $q(\boldsymbol{x}_t)$ is given by

$$q(\boldsymbol{x}_t) \propto \exp\Big( -\frac{1}{2}\|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}\|^2_{\boldsymbol{P}_{t|t-1}^{-1}}$$
$$-\frac{1}{2}\|\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)\|^2_{\langle \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)\rangle} \Big) \tag{22}$$

Therefore, after manipulating (22), we can identify that $q(\boldsymbol{x}_t) \approx \mathcal{N}(\hat{\boldsymbol{x}}_{t|t}, \boldsymbol{P}_{t|t})$, with

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t \left( \boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1} \right), \tag{23}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t \left( \boldsymbol{S}_t + \langle \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)\rangle^{-1} \right) \boldsymbol{K}_t^\top, \tag{24}$$

$$\boldsymbol{K}_t = \boldsymbol{C}_t \left( \boldsymbol{S}_t + \langle \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t)\rangle^{-1} \right)^{-1}, \tag{25}$$

where

$$\hat{\boldsymbol{y}}_{t|t-1} = \int \boldsymbol{h}(\boldsymbol{x}_t) \, p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1}) \, d\boldsymbol{x}_t \tag{26}$$

$$\boldsymbol{S}_t = \int \left( \boldsymbol{h}(\boldsymbol{x}_t) - \hat{\boldsymbol{y}}_{t|t-1} \right) \left( \boldsymbol{h}(\boldsymbol{x}_t) - \hat{\boldsymbol{y}}_{t|t-1} \right)^\top$$
$$\times \, p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1}) \, d\boldsymbol{x}_t \tag{27}$$

$$\boldsymbol{C}_t = \int \left( \boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1} \right) \left( \boldsymbol{h}(\boldsymbol{x}_t) - \boldsymbol{y}_{t|t-1} \right)^\top$$
$$\times \, p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1}) \, d\boldsymbol{x}_t \tag{28}$$

where the integrals can be solved analytically in linear systems (as in the celebrated KF) or through numerical integration in nonlinear systems where closed-form solutions are not possible (approach taken in the case of SPGFs).

*2) Update $q(\boldsymbol{z}_t)$:* Recall that due to the mutually independence assumption, we operate on the marginal distributions $q(z_t^{(i)})$ in (10). Each is obtained following (14) as

$$\ln\left[ q(z_t^{(i)}) \right] = \mathbb{E}_{\boldsymbol{\pi}_t, \boldsymbol{x}_t, \boldsymbol{z}_t^{(-i)}} \{ \ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})] \}$$
$$= \mathbb{E}_{\pi_t^{(i)}, \boldsymbol{x}_t, \boldsymbol{z}_t^{(-i)}} \{ \ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})] \}, \tag{29}$$

where notice that the expectation is over $\{\pi_t^{(i)}, \boldsymbol{x}_t, \boldsymbol{z}_t^{(-i)}\}$. For the sake of convenience, we reorganize the measurement vector such that the $i$-th element is swap at the end of the vector,

$$\boldsymbol{y}_{t,i} = \begin{pmatrix} \boldsymbol{y}_t^{(-i)} \\ y_t^{(i)} \end{pmatrix}, \quad \boldsymbol{h}_i(\boldsymbol{x}_t) = \begin{pmatrix} \boldsymbol{h}^{(-i)}(\boldsymbol{x}_t) \\ h^{(i)}(\boldsymbol{x}_t) \end{pmatrix}, \tag{30}$$

and the corresponding covariance matrix (4) is then reorganized accordingly as

$$\boldsymbol{\Sigma}_{t,i}(\boldsymbol{z}_t) = \begin{bmatrix} \boldsymbol{\Sigma}_{-i,-i} & \boldsymbol{\Sigma}_{-i,i} \\ \boldsymbol{\Sigma}_{i,-i} & \sigma_{i,i}^2 / z_t^{(i)} \end{bmatrix}. \tag{31}$$

Operating on the precision matrix $\boldsymbol{\Lambda}_{t,i} = \boldsymbol{\Sigma}_{t,i}(\boldsymbol{z}_t)^{-1}$ and using the Schur complement, this can be expressed as,

$$\boldsymbol{\Lambda}_{t,i} = \begin{bmatrix} \boldsymbol{\Lambda}_{-i,-i} & \boldsymbol{\Lambda}_{-i,i} \\ \boldsymbol{\Lambda}_{i,-i} & \lambda_{i,i} \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{M}_{-i} & -\boldsymbol{M}_{-i}\boldsymbol{\Sigma}_{-i,i}\frac{z_t^{(i)}}{\sigma_{i,i}^2} \\ -\frac{z_t^{(i)}}{\sigma_{i,i}^2}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{M}_{-i} & M_i \end{bmatrix} \tag{32}$$

where

$$\boldsymbol{M}_{-i} = \left( \boldsymbol{\Sigma}_{-i,-i} - \boldsymbol{\Sigma}_{-i,i}\frac{z_t^{(i)}}{\sigma_{i,i}^2}\boldsymbol{\Sigma}_{i,-i} \right)^{-1}, \tag{33}$$

$$M_i = \left( \frac{\sigma_{i,i}^2}{z_t^{(i)}} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i} \right)^{-1}. \tag{34}$$

Applying the matrix inversion lemma, (33) can be further extended as,

$$\boldsymbol{M}_{-i} = \boldsymbol{\Sigma}_{-i,-i}^{-1} + \frac{\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\frac{\sigma_{i,i}^2}{z_t^{(i)}} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}. \tag{35}$$

Notice that for independently-distributed observations, the cross-covariance in $\boldsymbol{\Sigma}_t(\boldsymbol{z}_t)$ is zero and $M_i = z_t^{(i)}/\sigma_{i,i}^2$. For correlated data, $z_t^{(i)}$ is part of a fraction which would make it tedious in terms of computing the Bernoulli probabilities for the indicator. Fortunately, $z_t^{(i)}$ is a binary variable $z_t^{(i)} = \{0,1\}$ which allows the following equivalent expression

$$M_i = \frac{z_t^{(i)}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}, \tag{36}$$

which yields to identical result as in (34) regardless of the value $z_t^{(i)}$ takes. Thus, if $z_t^{(i)} = 1$, the multivariative Gaussian with all $n_y$ observations would be obtained; if $z_t^{(i)} = 0$, the corresponding diagonal elements and off-diagonal elements in precision matrix would be 0 according to (32)–(34), and the influence of those contaminated observations would be removed according to (5), thus only keeping observations deemed clean in a multivariate Gaussian of reduced dimension $n_y'$. In order to model $q(z_t^{(i)})$ in the form of a Bernoulli distribution, we use

$$\ln[p(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t})] = -\frac{1}{2}\|\boldsymbol{y}_{t,i} - \boldsymbol{h}_i(\boldsymbol{x}_t)\|^2_{\boldsymbol{\Sigma}_{t,i}(\boldsymbol{z}_t)^{-1}}$$
$$+ z_t^{(i)} \ln[\pi_t^{(i)}] + \left(1 - z_t^{(i)}\right)$$
$$\times \ln[1 - \pi_t^{(i)}] - \ln[c(\boldsymbol{z}_t)] + \kappa,$$

in

$$\ln\left[q(z_t^{(i)})\right]$$

$$= \mathbb{E}_{\pi_t^{(i)}, \boldsymbol{x}_t, \boldsymbol{z}_t^{(-i)}} \left\{\ln\left[p\left(\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t, \boldsymbol{y}_{1:t}\right)\right]\right\}$$

$$= \mathbb{E}_{\pi_t^{(i)}, \boldsymbol{x}_t, \boldsymbol{z}_t^{(-i)}} \left\{-\frac{1}{2}||\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)||_{\boldsymbol{\Lambda}_{-i,-i}}^2\right.$$

$$- \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top \boldsymbol{\Lambda}_{-i,i}\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)$$

$$- \frac{1}{2}||y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)||_{\lambda_{i,i}}^2 + z_t^{(i)}\ln[\pi_t^{(i)}]$$

$$+ \left(1 - z_t^{(i)}\right)\ln[1 - \pi_t^{(i)}] - 0.5 z_t^{(i)}\ln[2\pi]$$

$$- 0.5 z_t^{(i)}\ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])|$$

$$\left. - 0.5(1 - z_t^{(i)})\ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])| + \kappa\right\} \tag{37}$$

where in expanding the terms due to the normalizing constant (6) we factorized the determinant of $\mathbf{C}$ (cf. (5)) as

$$|\mathbf{C}(\boldsymbol{z}_t)| = |\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])|^{z_t^{(i)}}|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])|^{(1 - z_t^{(i)})} \tag{38}$$

After additional manipulations (cf. Appendix A) we obtain

$$q\left(z_t^{(i)}\right) = \exp\{-0.5 z_t^{(i)}\,\mathrm{Tr}\left(\boldsymbol{B}_{-i,-i}\langle\boldsymbol{\Lambda}_{t1}\rangle\right)$$

$$- 0.5 z_t^{(i)}\,\mathrm{Tr}\left(-\boldsymbol{B}_{i,-i}\langle\boldsymbol{\Lambda}_{t2}\rangle\right)$$

$$- 0.5 z_t^{(i)}\,\mathrm{Tr}\left(-\boldsymbol{B}_{i,-i}\langle\boldsymbol{\Lambda}_{t3}\rangle\right) - 0.5 z_t^{(i)}\,\mathrm{Tr}\left(b_{i,i}\langle\lambda_{t4}\rangle\right)$$

$$+ z_t^{(i)}\langle\ln[\pi_t^{(i)}]\rangle + \left(1 - z_t^{(i)}\right)\langle\ln[1 - \pi_t^{(i)}]\rangle$$

$$- 0.5 z_t^{(i)}\ln[2\pi]$$

$$- 0.5 z_t^{(i)}\langle\ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])|\rangle$$

$$- 0.5(1 - z_t^{(i)})\langle\ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])|\rangle + \kappa\} \tag{39}$$

where $\boldsymbol{B}_{-i,-i}$, $\boldsymbol{B}_{i,-i}$, and $b_{i,i}$ are given by

$$\boldsymbol{B}_{-i,-i} = \int \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)$$

$$\times \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top q\left(\boldsymbol{x}_t\right) d\boldsymbol{x}_t, \tag{40}$$

$$\boldsymbol{B}_{i,-i} = \int \left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)$$

$$\times \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top q\left(\boldsymbol{x}_t\right) d\boldsymbol{x}_t,$$

$$b_{i,i} = \int \left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)^\top q\left(\boldsymbol{x}_t\right) d\boldsymbol{x}_t, \tag{41}$$

and the expectations (with respect to $q(\boldsymbol{z}_t^{(-i)})$) are

$$\langle\boldsymbol{\Lambda}_{t1}\rangle = \sum_{\boldsymbol{z}^{(-i)} \in \mathcal{Z}_{-i}} \frac{\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}$$

$$\times q(\boldsymbol{z}_t^{(-i)} = \boldsymbol{z}^{(-i)}) \tag{42}$$

$$\langle\boldsymbol{\Lambda}_{t2}\rangle = \frac{2}{\sigma_{i,i}^2} \sum_{\boldsymbol{z}^{(-i)} \in \mathcal{Z}_{-i}} \boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}q(\boldsymbol{z}_t^{(-i)} = \boldsymbol{z}^{(-i)}) \tag{43}$$

$$\langle\boldsymbol{\Lambda}_{t3}\rangle = \sum_{\boldsymbol{z}^{(-i)} \in \mathcal{Z}_{-i}} \frac{2\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}{\sigma_{i,i}^4 - \sigma_{i,i}^2\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}$$

$$\times q(\boldsymbol{z}_t^{(-i)} = \boldsymbol{z}^{(-i)}) \tag{44}$$

$$\langle\lambda_{t4}\rangle = \sum_{\boldsymbol{z}^{(-i)} \in \mathcal{Z}_{-i}} \frac{1}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}q(\boldsymbol{z}_t^{(-i)} = \boldsymbol{z}^{(-i)}) \tag{45}$$

where the sum over $\boldsymbol{z}^{(-i)}$ is indeed the sum over all possible $2^{n_y - 1}$ values for the vector $\boldsymbol{z}^{(-i)}$, whose associated probability is $q(\boldsymbol{z}_t^{(-i)} = \boldsymbol{z}^{(-i)})$. The set of all those possible values is given by $\mathcal{Z}_{-i} = \{0,1\}^{n_y - 1}$ such that $|\mathcal{Z}_{-i}| = 2^{n_y - 1}$. Notice that $\boldsymbol{z}_t^{(-i)}$ appears in the diagonal elements of $\boldsymbol{\Sigma}_{-i,-i}$, dependence that we omitted for the sake of simplicity.

Finally, we can identify from $q(z_t^{(i)})$ that $z_t^{(i)}$ is a Bernoulli-distributed random variable, whose distribution can be characterized as

$$p\left(z_t^{(i)} = 1\right) \propto \exp\{-0.5\,\mathrm{tr}\left(\boldsymbol{B}_{-i,-i}\langle\boldsymbol{\Lambda}_{t1}\rangle\right)$$

$$- 0.5\,\mathrm{Tr}\left(\boldsymbol{B}_{i,-i}\langle\boldsymbol{\Lambda}_{t2}\rangle\right) - 0.5\,\mathrm{Tr}\left(\boldsymbol{B}_{i,-i}\langle\boldsymbol{\Lambda}_{t3}\rangle\right)$$

$$- 0.5\,\mathrm{Tr}\left(b_{i,i}\langle\lambda_{t4}\rangle\right) + \langle\ln[\pi_t^{(i)}]\rangle - 0.5\ln[2\pi]$$

$$- 0.5\,\langle\ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])|\rangle\} \tag{46}$$

$$p\left(z_t^{(i)} = 0\right) \propto \exp\{\langle\ln[1 - \pi_t^{(i)}]\rangle$$

$$- 0.5\langle\ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])|\rangle\}. \tag{47}$$

Using (46) and (47), the expectation of a Bernoulli $z_t^{(i)}$ can be readily computed as

$$\langle z_t^{(i)}\rangle = \frac{p\left(z_t^{(i)} = 1\right)}{p\left(z_t^{(i)} = 1\right) + p\left(z_t^{(i)} = 0\right)}, \tag{48}$$

a quantity that would be required in updating $q(\pi_t^{(i)})$ next.

*3) Update $q(\pi_t^{(i)})$:* Similar to the derivation in [35], [40], $q(\pi_t^{(i)})$ is updated for each indicator as,

$$q\left(\pi_t^{(i)}\right) \propto \exp\left(e_t^{(i)}\ln[\pi_t^{(i)}] + f_t^{(i)}\ln[1 - \pi_t^{(i)}]\right) \tag{49}$$

where

$$e_t^{(i)} = e_0 + \langle z_t^{(i)}\rangle \tag{50}$$

$$f_t^{(i)} = f_0 + 1 - \langle z_t^{(i)}\rangle \tag{51}$$

such that $f_0$ and $e_0$ are two initial hyper-parameters for distribution $\pi_t^{(i)}$, assumed the same for all $i$.

*4) Practical Aspects:* At this point, we would like to comment on an algorithmic choice that should be taken in order to make the proposed filter practical. More precisely, the algorithm requires the evaluation of multivariate expectations $\langle \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}_t) \rangle$, $\langle \boldsymbol{\Lambda}_{t1} \rangle$, $\langle \boldsymbol{\Lambda}_{t2} \rangle$, $\langle \boldsymbol{\Lambda}_{t3} \rangle$, and $\langle \lambda_{t1} \rangle$ which are of the form $\mathbb{E}_{\boldsymbol{z}_t}\{\cdot\}$ or $\mathbb{E}_{\boldsymbol{z}_t^{(-i)}}\{\cdot\}$. These expectations over discrete random variables involve sums over their $2^{n_y}$ and $2^{n_y-1}$ possible combinations, respectively. Additionally, the expectation requires solving a matrix inverse (with cubic complexity on the dimension) for each of the combinations, resulting on $\mathcal{O}(n_y^3 2^{n_y})$ and $\mathcal{O}((n_y-1)^3 2^{n_y-1})$ computational complexities,[2] respectively. As a consequence, it is impractical to exactly evaluate the expectation when $n_y$ grows. Alternatively, we propose two practical methods based on the knowledge and factorization of $q(\boldsymbol{z}_t = \boldsymbol{z}) = \prod_{i=1}^{n_y} q(z_t^{(i)} = z^{(i)})$.

First, we propose a maximization approach in which we select the value of each $\{z_t^{(i)}\}_{i=1}^{n_y}$ as the one with highest probability. That is for all $i$ of interest, if $p(z_t^{(i)} = 1) > p(z_t^{(i)} = 0)$ then $z_t^{(i)} = 1$ is selected, and $z_t^{(i)} = 0$ otherwise.

Secondly, a sampling approach is also proposed in which a binary value for $z_t^{(i)}$ is drawn from $q(z_t^{(i)})$ with probabilities $p(z_t^{(i)} = 1)$ and $p(z_t^{(i)} = 0)$. This is done independently for all the $i$ indicators of interest (depending on the target expectation). In general, this independent sampling can be done $N$ times, such that

$$\mathbb{E}_{\boldsymbol{z}_t}\{g(\boldsymbol{z}_t)\} \approx \frac{1}{N} \sum_{n=1}^{N} g(\boldsymbol{z}_{t,n}) \tag{52}$$

$$\mathbb{E}_{\boldsymbol{z}_t^{(-i)}}\{g(\boldsymbol{z}_{t,n}^{(-i)})\} \approx \frac{1}{N} \sum_{n=1}^{N} g(\boldsymbol{z}_{t,n}^{(-i)}) \tag{53}$$

where $g(\cdot)$ is an arbitrary function of the corresponding random variable; and the samples are independently sampled from $\boldsymbol{z}_{t,n} \sim q(\boldsymbol{z}_t)$ and $\boldsymbol{z}_{t,n}^{(-i)} \sim q(\boldsymbol{z}_t^{(-i)})$.

Finally, similarly to [35], since the VI process detailed here is an iterative procedure, in practice we need to specify a stop criteria. For instance, one could compare two consecutive state estimates (i.e. through the norm of the difference) and choose an arbitrary small threshold (e.g. $10^{-9}$ in the simulations of this paper) below which the algorithm is said to converge.

### B. Particular Case I: Independent Data and Multiple Outlier Indicators

As mentioned at the beginning of this section, the Generalized VBKF method described earlier can be particularized to obtain other methods available in the literature. Mostly, this involves a different arrangement of the latent variables $\boldsymbol{\theta}$ and the corresponding auxiliary distributions used in approximating $q(\boldsymbol{\theta})$ through the VI principle.

---

[2]The asymptotic time complexity of an algorithm can be interpreted as the number of required basic operations, denoted as $\mathcal{O}(d)$ with $d$ being the number of operations. A function $p(d)$ is $\mathcal{O}(g(d))$ if and only if there exist a real, positive constant $C$ and a positive integer $d_0$ such that $p(d) \le Cg(d),\ \forall d \ge d_0$.

Particularly, we notice that the variant in [40], where an indicator associate to each element in $\boldsymbol{y}_t$ was considered under the assumption that data was uncorrelated, can be derived from the general VBKF. In this case, $\boldsymbol{\theta} = \{\boldsymbol{x}_t, \boldsymbol{\pi}_t, \boldsymbol{z}_t\}$ remains the same, with the update of $q(\boldsymbol{\pi}_t)$ as in Section II-A3, since the model assumed for those hyper-parameters is the same in both methods. However, in [40] the assumption that all measurements are uncorrelated was used in order to simplify the model and the corresponding method's derivation. Thus, the covariance matrix $\boldsymbol{\Sigma}_t(\boldsymbol{z}_t)$ in (4) becomes:

$$\boldsymbol{\Sigma}_t(\boldsymbol{z}_t) = \begin{bmatrix} \sigma_{1,1}^2/z_t^{(1)} & 0 & \cdots & 0 \\ 0 & \sigma_{2,2}^2/z_t^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{n_y,n_y}^2/z_t^{(n_y)} \end{bmatrix}, \tag{54}$$

and the measurement model in (3) can be further derived as:

$$\begin{aligned} p\left(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{z}_t\right) &= \mathcal{N}\left(\mathcal{T}(\boldsymbol{y}_t, \boldsymbol{z}_t); \mathcal{T}(\boldsymbol{h}(\boldsymbol{x}_t), \boldsymbol{z}_t), \mathbf{C}(\boldsymbol{z}_t)\right) \\ &= \prod_{i=1}^{n_y} \mathcal{N}\left(\mathcal{T}(\boldsymbol{y}_t^{(i)}, z_t^{(i)}); \mathcal{T}(\boldsymbol{h}^{(i)}(\boldsymbol{x}_t), z_t^{(i)}), [\mathbf{C}(z_t^{(i)})]_{i,i}\right) \\ &= \prod_{i=1}^{n_y} \mathcal{N}\left(y_t^{(i)}; \boldsymbol{h}^{(i)}(\boldsymbol{x}_t), [\boldsymbol{R}_t]_{i,i}\right)^{z_t^{(i)}}, \end{aligned} \tag{55}$$

Notice that with $\boldsymbol{R}_t$ diagonal, $c(\boldsymbol{z}_t) = \sqrt{(2\pi)^{n_y'} |\mathbf{C}(\boldsymbol{z}_t)|} = \prod_{i=1}^{n_y} \sqrt{(2\pi)^{z_t^{(i)}} [\mathbf{C}(z_t^{(i)})]_{i,i}} = \prod_{i=1}^{n_y} c(z_t^{(i)})$. The normalization factor $c(z_t^{(i)})$ being the standard one for a univariate Gaussian distribution, yielding to the last equality in (55). This simplification circumvents the general algorithm complexity and computation but degrades the performance of the VBKF when correlated data is observed.

According to (12), with the same model for $\boldsymbol{z}_t$ and its parameter $\boldsymbol{\pi}_t$, $q(\boldsymbol{x}_t)$ is updated as:

$$\begin{aligned} q(\boldsymbol{x}_t) \propto \exp\Big( &- \frac{1}{2}||\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}||_{\boldsymbol{P}_{t|t-1}^{-1}}^2 \\ &- \frac{\langle \boldsymbol{Z}_t \rangle}{2}||\boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)||_{\boldsymbol{R}_t^{-1}}^2 \Big) \end{aligned} \tag{56}$$

where $\boldsymbol{Z}_t$ is a diagonal matrix with $\boldsymbol{z}_t$ as its elements, $\boldsymbol{Z}_t = \mathrm{diag}(z_t^{(1)}, \ldots, z_t^{(n_y)})$.

Using (29), the update of $q(z_t^{(i)})$ would be:

$$\begin{aligned} q\left(z_t^{(i)}\right) \propto \exp\Big( &- 0.5 z_t^{(i)}\left(b_t^{(i)}/\sigma_{i,i}^2\right) + z_t^{(i)}\langle \ln[\pi_t^{(i)}] \rangle \\ &+ \left(1 - z_t^{(i)}\right)\langle \ln[1 - \pi_t^{(i)}] \rangle \Big) \end{aligned} \tag{57}$$

where $b_t^{(i)}$ is given by

$$b_t^{(i)} = \int \left(y_t^{(i)} - h^{(i)}(\boldsymbol{x}_t)\right)\left(y_t^{(i)} - h^{(i)}(\boldsymbol{x}_t)\right)^\top q(\boldsymbol{x}_t)\, d\boldsymbol{x}_t \tag{58}$$

Thus, the update of $q(\boldsymbol{x}_t)$, $q(\boldsymbol{z}_t)$ and $q(\boldsymbol{\pi}_t)$ in this particular case would be the same as in [40], showing the generalization of the method proposed in this contribution.

## C. Particular Case II: Scalar Outlier Indicator

Here we notice that the seminal method in [35] can be readily obtained from the Generalized VBKF. In this case, $\boldsymbol{\theta} = \{\boldsymbol{x}_t, \pi_t, z_t\}$ since a single indicator is used for the entire vector $\boldsymbol{y}_t$, thus accepting or rejecting the entire observation vector. Assuming a similar model for $\pi_t$, the update of $q(\pi_t)$ will be as in Section II-A3 with exception that just one indicator is required. However, the likelihood distribution in (3) will change in this case since the *quality* of all measurements is modeled by a binary scalar indicator $z_t$:

$$\boldsymbol{\Sigma}_t(z_t) = \begin{bmatrix} \sigma_{1,1}^2/z_t & \sigma_{1,2}^2/z_t & \dots & \sigma_{1,n_y}^2/z_t \\ \sigma_{2,1}^2/z_t & \sigma_{2,2}^2/z_t & \dots & \sigma_{2,n_y}^2/z_t \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_y,1}^2/z_t & \sigma_{n_y,2}^2/z_t & \dots & \sigma_{n_y,n_y}^2/z_t \end{bmatrix}, \quad (59)$$

thus the measurement model in (3) becomes

$$\begin{aligned} p\left(\boldsymbol{y}_t|\boldsymbol{x}_t, z_t\right) &= \mathcal{N}\left(\mathcal{T}(\boldsymbol{y}_t, z_t); \mathcal{T}(\boldsymbol{h}(\boldsymbol{x}_t), z_t), \mathbf{C}(z_t)\right) \\ &= \mathcal{N}\left(\boldsymbol{y}_t; \boldsymbol{h}(\boldsymbol{x}_t), \boldsymbol{R}_t/z_t\right) \qquad (60) \\ &= \mathcal{N}\left(\boldsymbol{y}_t; \boldsymbol{h}(\boldsymbol{x}_t), \boldsymbol{R}_t\right)^{z_t}, \end{aligned}$$

where notice that $\boldsymbol{R}_t$ can be colored. In the general notation, $\boldsymbol{z}_t = z_t \cdot \mathbf{1}$, and since $c(z_t) = \sqrt{(2\pi)^{n_y z_t}|\boldsymbol{R}_t|}$ then: $c(1) = \sqrt{(2\pi)^{n_y}|\boldsymbol{R}_t|}$ and $c(0) = 1$, yielding to the last equality in (60). According to (22), the update of $q(\boldsymbol{x}_t)$ is:

$$\begin{aligned} q\left(\boldsymbol{x}_t\right) \propto \exp\Big( &- \frac{1}{2}||\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}||^2_{\boldsymbol{P}^{-1}_{t|t-1}} \\ &- \frac{\langle z_t \rangle}{2}||\boldsymbol{y}_t - \boldsymbol{h}\left(\boldsymbol{x}_t\right)||^2_{\boldsymbol{R}^{-1}_t}\Big) \qquad (61) \end{aligned}$$

and $q(z_t)$ is updated according to (14):

$$\begin{aligned} q\left(z_t\right) \propto \exp\{ &- 0.5 z_t \operatorname{Tr}\left(\boldsymbol{B}_t \boldsymbol{R}^{-1}_t\right) \\ &+ z_t \langle \ln[\pi_t] \rangle + (1 - z_t) \langle \ln[1 - \pi_t] \rangle \} \qquad (62) \end{aligned}$$

where $\boldsymbol{B}_t$ is given by

$$\boldsymbol{B}_t = \int \left(\boldsymbol{y}_t - \boldsymbol{h}\left(\boldsymbol{x}_t\right)\right)\left(\boldsymbol{y}_t - \boldsymbol{h}\left(\boldsymbol{x}_t\right)\right)^\top q\left(\boldsymbol{x}_t\right) d\boldsymbol{x}_t \qquad (63)$$

$$= \begin{bmatrix} \boldsymbol{B}_{-i,-i} & \boldsymbol{B}_{-i,i} \\ \boldsymbol{B}_{i,-i} & b_{i,i} \end{bmatrix}, \qquad (64)$$

the intuition being that when a single indicator is used, the covariance $\boldsymbol{\Sigma}_{t,i}(z_t)^{-1}$ in (37) becomes $\boldsymbol{\Sigma}_t(z_t) = \boldsymbol{R}_t/z_t$. Therefore, we have that

$$\begin{aligned} \ln\left[q(z_t)\right] &= \mathbb{E}_{\pi_t, \boldsymbol{x}_t}\{\ln\left[p\left(\boldsymbol{x}_t, \pi_t, z_t, \boldsymbol{y}_{1:t}\right)\right]\} \\ &= \mathbb{E}_{\pi_t, \boldsymbol{x}_t}\Big\{ -\frac{z_t}{2}||\boldsymbol{y}_t - \boldsymbol{h}\left(\boldsymbol{x}_t\right)||^2_{\boldsymbol{R}^{-1}_t} \\ &\qquad + z_t \ln[\pi_t] + (1 - z_t)\ln[1 - \pi_t]\Big\} + \kappa, \end{aligned} \qquad (65)$$

matching the result in the original article, resulting in the same $q(z_t)$ update. Thus, the update of $q(\boldsymbol{x}_t)$, $q(z_t)$ and $q(\pi_t)$ in this particular case would be the same as in [35]. Remarkably, the original VBKF in [35] can deal with both correlated and uncorrelated data. However, since it accepts or rejects the entire
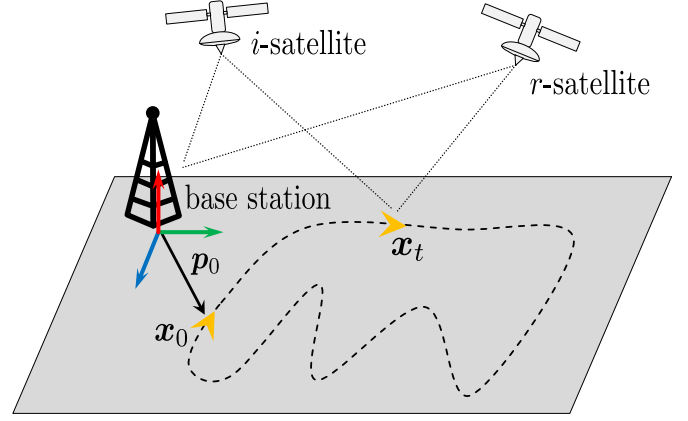


Fig. 2. Illustration of the RTK positioning system: the trajectory followed by the vehicle (in yellow) depicted with the dashed black line. The localization problem is resolved with respect to the position of the base station combining the observations of the satellites between the base and vehicle receivers.

observation vector jointly, it is potentially inefficient in dealing with independent outliers among the elements in $\boldsymbol{y}_t$ whereby a single outlier can corrupt the observed vector. In practice, the VBKF would only operate using predictions, thus eventually losing track of the estimated state even when a single element is contaminated.

## III. ILLUSTRATIVE EXAMPLE: GPS LOCALIZATION

In order to put into practice the new generalized VBKF and show the performance improvement with respect to previous contributions, we consider an illustrative positioning example where by construction the data is correlated.

### A. GPS RTK Problem Formulation

Real Time Kinematic (RTK) is a well-known differential GNSS-based positioning procedure, for which the unknown location of a vehicle is determined with respect to a georeferenced base station [44]. Since satellite observations are influenced by atmospheric delays and receiver clock offset effects, RTK applies the so-called observation double-differencing to eliminate the nuisance parameters. Double-differencing involve taking differences between the observations at the target and the base station, and then with respect to a reference satellite [45]. Fig. 2 provides a pictorial example of the agents involved in the RTK positioning: $i = 1, \dots, n_s$ satellites and $r$ (for reference) satellite, the base station and the trajectory of the moving vehicle over time.

Hereinafter, we refer to observations as the result of double-differencing the original measurements. The observation covariance matrix becomes then fully populated, since the noise of the pivot satellite is present across all observations. RTK implies the use of code and carrier phase observations, with the later introducing an additional estimation complexity, since a number of integer ambiguities is to be found. Given the fact that ambiguities are related to wavelength, double difference observations needs to be calculated among satellites observations in the same frequency band. In each frequency band, we have $\tilde{n}_s$ satellite observation and $\tilde{n}_y$ double difference observations,

$\tilde{n}_y = \tilde{n}_s - 1$. Without loss of generality and for the sake of simplicity, this example disregards the exploitation of carrier phase observations. The dynamical constant velocity model for RTK is written as

$$\underbrace{\begin{bmatrix} \boldsymbol{p}_t \\ \boldsymbol{v}_t \end{bmatrix}}_{\boldsymbol{x}_t} = \underbrace{\begin{bmatrix} \boldsymbol{I} & \Delta t \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{F}_t} \underbrace{\begin{bmatrix} \boldsymbol{p}_{t-1} \\ \boldsymbol{v}_{t-1} \end{bmatrix}}_{\boldsymbol{x}_{t-1}} + \boldsymbol{\epsilon}_t \quad (66)$$

where $\boldsymbol{p}_t$ and $\boldsymbol{v}_t$ are the three-dimensional vectors of position and velocity, respectively, $\boldsymbol{x}_t$ is the state to be inferred and $\boldsymbol{\epsilon}_t$ is a zero-mean Gaussian noise vector with known covariance $\boldsymbol{Q}_t$. Having the base station as center of the coordinate frame and the positions of the satellites and the vehicle w.r.t. to the base, the observation model for the $i$-th measurement is

$$y_t^{(i)} = \underbrace{\left\| \boldsymbol{p}_t^{(i)} - \boldsymbol{p}_t \right\| - \left\| \boldsymbol{p}_t^{(r)} - \boldsymbol{p}_t \right\| - \left\| \boldsymbol{p}_t^{(i)} \right\| + \left\| \boldsymbol{p}_t^{(r)} \right\|}_{h^{(i)}(\boldsymbol{x}_t)}$$
$$+ \underbrace{n_t^{(i)} + n_{b,t}^{(i)} + n_t^{(r)} + n_{b,t}^{(r)}}_{\eta_t^{(i)}}, \quad (67)$$

where the superscript in $\boldsymbol{p}$ indicates the satellite and the subscript $b$ in the noise $n$ refers to the noise observed at the base station receiver. Notice that, the observation model (67) is nonlinear and the observations' noise vector $\boldsymbol{\eta}_t$ becomes inter-correlated, since the noise for the reference satellite on the base and moving receiver is common for the $n_y$ observations. A linearized version of (67) subject to outliers is

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{G}_t \boldsymbol{p}_t + \boldsymbol{\eta}_t, & \text{under } \mathcal{M}_0 \\ \boldsymbol{G}_t \boldsymbol{p}_t + \boldsymbol{\eta}_t + \boldsymbol{o}_t, & \text{under } \mathcal{M} \end{cases} \quad (68)$$

where $\boldsymbol{y}_t = (y_t^{(1)}, \ldots, y_t^{(n_y)})^\top \in \mathbb{R}^{n_y}$ is the vector of double difference (DD) observations for the $n_y$ tracked satellites (plus the reference one) and $\boldsymbol{G}_t \in \mathbb{R}^{n_y \times 3}$ is the geometry matrix containing the unit line-of-sight steering vectors to the satellites, as in [45, Ch. 26]. The observation noise $\boldsymbol{\eta}_t$ has a fully populated covariance matrix $\boldsymbol{R}_t \in \mathbb{R}^{n_y \times n_y}$, as described in [46]. Multipath and non-line-of-sight (NLOS) effects lead to gross biases in the range observations to the satellites. Such errors are collected in vector $\boldsymbol{o}_t$, following the initial model (2) to generate the data while the method is derived considering the alternative model (3).

## B. Simulation Setup

The root mean square error (RMSE) over a trajectory and the percentage of misdetected outliers are taken as measures of performance. The RMSEs are obtained from 50 independent Monte Carlo runs for a 100 seconds trajectory, sampled at $\Delta t = 1$ s, while 500 Monte Carlo runs were used in the experiments comparing generalized VBKF with $N = \{1, 10\}$ to show the tiny difference between them. The present simulation, simulated in MATLAB, considers $n_s = 12$ tracked satellites at a specific time, distributed according to a realistic sky plot, as shown in Fig. 3. In practice, satellites belonging to different constellations (i.e., and depending on the satellite generation) may broadcast a different number of signals at different frequency bands. In the
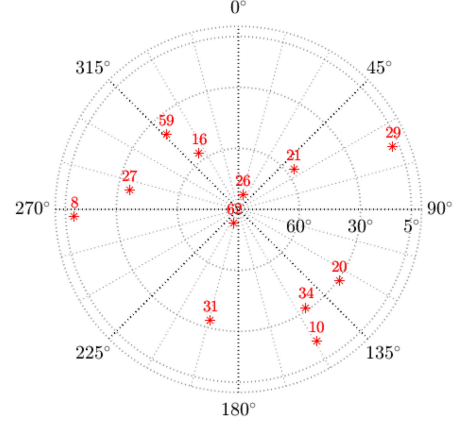


Fig. 3. Sky plot of the tracked satellites from GPS L1 and Galileo E1 for the simulation.

TABLE I
PARAMETERS FOR THE MARKOV CHAIN MODEL

| | |
|---|---|
| Initial probability for NLOS state | 0.7 |
| Initial probability for LOS state | 0.3 |
| Transition probability of staying in NLOS state | 0.9 |
| Transition probability from NLOS state to LOS state | 0.1 |
| Transition probability of staying in LOS state | 0.3 |
| Transition probability from LOS state to NLOS state | 0.7 |

simulation, we consider $n_y = 2n_s - 4$ DD observations from a dual-band receiver. Outliers (i.e., NLOS) are simulated based on a Markov Chain model according to [47] with parameters shown in Table I. The motivation of such model is to account for the time correlation of multipath (and other spatially correlated errors), in contrast to the less realistic case of assuming totally independent statistics between consecutive time instances.

Notice that from $t = 10$ s, outliers are always injected into both pivot satellite observations (i.e., at the base station and the corresponding one at the receiver side), and to the observations for 6 out of the $n_y = 20$ DD observations, therefore outliers are correlated. Besides, since outliers are simulated from the corresponding carrier-to-noise density ratio ($C/N_0$ in dB-Hz), which is a common measure in GNSS receivers, the random outlier amplitudes are in general correlated to the noise power, which is also modelled from the $C/N_0$ as [48],

$$\sigma = \frac{c \cdot 3.444 \cdot 10^{-4}}{\sqrt{C/N_0 B T_{obs}}}, \quad (69)$$

where $\sigma$ is the standard deviation for the observation noise, $c$ is the speed of light, $B = 2$ MHz is the receiver bandwidth, and $T_{obs} = 20$ ms is the coherent integration time. We consider the performance comparison of different methods: 1) a standard Extended Kalman Filtering (EKF) that does not account for possible outliers; 2) the original VBKF [35] using a single outlier indicator; 3) a VBKF with independent outlier indicators (see Section II-B); 4) the generalized VBKF, considering two different values $N = \{1, 10\}$, which are used to estimate (52)–(53); 5) an ideal EKF that accounts for all outliers; 6) a Cubature KF [6] with measurement gating by threshold equalling to triple of standard deviation of innovations [49]; and 7) a Cubature KF with measurement gating by threshold equalling to standard deviation
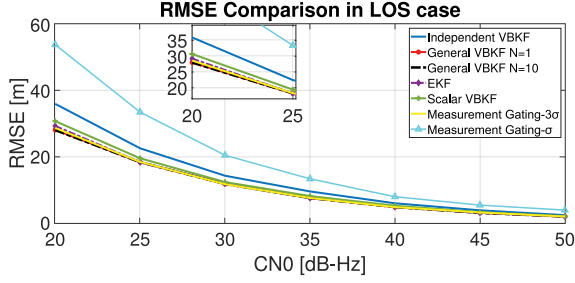
Fig. 4. RMSE performance of the VBKF with independent outliers indicator, general VBKF, standard EKF and VBKF with single outlier indicator as a function of $C/N_0$ in the LOS scenario.



Fig. 5. RMSE performance of the VBKF with independent outlier indicators, general VBKF and standard EKF as a function of $C/N_0$ in the NLOS scenario.

of innovations. Note that, when computation of expectations is needed in these methods, we used numerical integration based on Gaussian cubature rules.

It is worth noting that the original VBKF [35] is taken as the performance benchmark. In the original article [35], this method was already shown to provide improved performances with respect to other robust state-of-the-art filtering techniques (i.e., IGG-based cubature information filter (CIF), Huber-based CIF, Hampel-based CIF, and Student's t-based robust cubature KF), therefore we do not include the comparison to such techniques in the results.

### C. Results

When assessing the performance of new robust filtering techniques, and because of the optimality versus robustness trade-off, two important points are: $i)$ the so-called loss of efficiency (LoE) [31], that is, the loss in performance with respect to the optimal method under nominal conditions; and $ii)$ under non-nominal conditions, the performance gain with respect to the optimal solution that does not accounting for a possible model mismatch, that is, outliers in our case.

*1) Loss of Efficiency:* the LoE is assessed by comparing the performance of the different methods under nominal line-of-sight (LOS) conditions, i.e., without outliers. The LoE results are shown in Fig. 4. It can be seen that the general VBKF method (i.e., regardless of $N$) exhibits a pretty similar performance compared to the standard EKF and measurement gating method in terms of RMSE, therefore the proposed method has a very low LoE. The original VBKF with a single scalar outlier indicator shows a minor degradation at low $C/N_0$. In contrast, the VBKF with independent outlier indicators performance is degraded compared to the EKF, that is, it exhibits a larger LoE. This is mainly induced by the model mismatch, because this method does not take into account that observations are correlated. The uncorrelation of the observations is the main assumption for its derivation (see Section II-B).

*2) Robustness:* to assess the robustness of the proposed approach we show the performance of the different methods under non-nominal NLOS conditions, i.e., with outliers, in Fig. 5. The mean RMSE is given as a function of the $C/N_0$. When comparing both robust methods to the standard EKF, it is obvious that while the latter is severely influenced by outliers the robust solutions are able to cope with outliers and improve the overall performance. Notice that the general VBKF
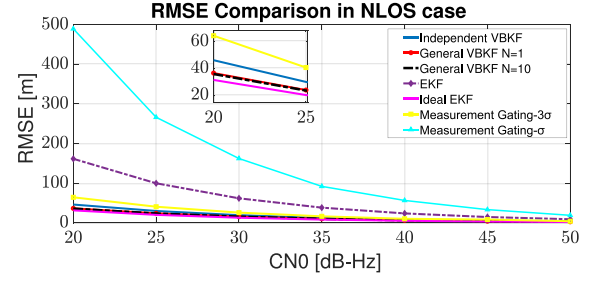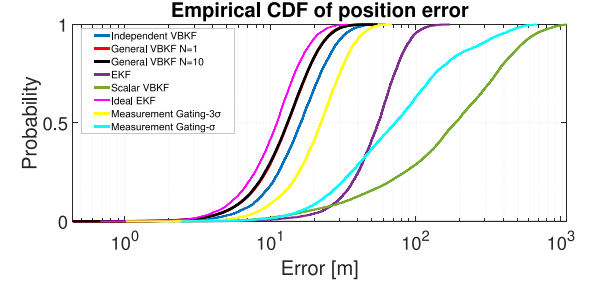


Fig. 6. Empirical CDF of position estimation error performance of the VBKF with independent outlier indicators, general VBKF, standard EKF and the VBKF with single outlier indicator in logarithmic scale.

improves the performance obtained with the independent outlier indicator VBKF. Such performance gain is expected to be even larger in applications where the correlation among observations increases. The performance of the original VBKF is not shown in Fig. 5 because it is substantially worse than the rest of methods (i.e., orders of magnitude larger). The gating method works reasonably well against outliers. Although, there are still meter level degradations compared with the VBKF method in certain configurations, for instance when the signal-to-noise ratio is small. Additionally, in other applications where data becomes more correlated, it is expected that the generalized VBKF will outperform the gating approach more clearly. On the other hand, the measurement gating strategy features a much lighter implementation, which could suffice in certain applications where data correlation is not severe.

The original VBKF with a single outlier indicator (as in Section II-C and [35]) was also simulated, exhibiting a RMSE of large magnitude. The reason being that the original VBKF, once it determines that the observation contains an outlier, discards the complete observation vector. In outlier-rich situations, as the NLOS case simulated here, this implies that few observations are indeed used in the KF update. To further support this statement and provide a meaningful comparison, Fig. 6 shows the empirical CDF of position estimation error of all VBKF versions as well as the standard EKF and measurement gating method under $C/N_0 = 30$ dB-Hz in NLOS case. It is clear that the VBKF with a single outlier indicator presents much larger errors compared to the other methods, even with respect to the standard EKF, due to the fact the it mostly relies on KF predictions as outliers are being detected.

Again, notice that the performance of the generalized VBKF in terms of RMSE for different values of $N$ is very similar. This
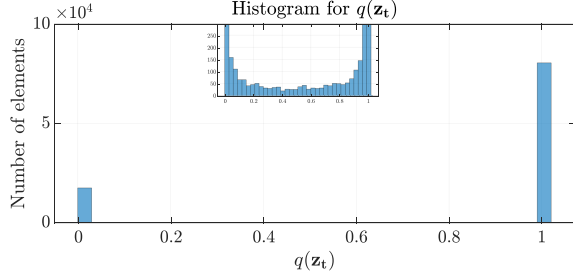
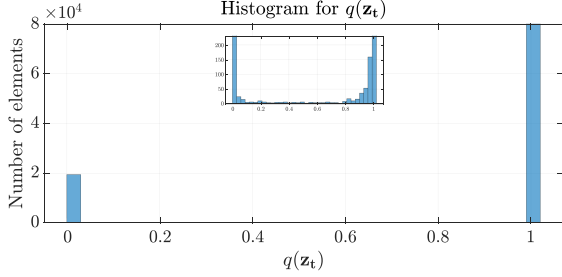Fig. 7. Histogram for $q(\boldsymbol{z}_t)$ in the transient regime.



Fig. 8. Histogram for $q(\boldsymbol{z}_t)$ in the convergence regime.

behaviour can be explained by the fact that the estimation of $q(z_t^{(i)})$ is likely to converge to either 0 or 1 when the general VBKF works properly. This can be further supported by the results in Fig. 7 and Fig. 8 (along with an inner plot showing a zoom for the lower values) for the case $N = 1$. In the iterative VBKF process, the estimation of $q(\boldsymbol{z}_t)$ experiences a transient regime and a convergence regime, which we investigate in the aforementioned figures, respectively. Whereas in the transient regime there are some values in the open interval (0,1), when the algorithm converged, almost all the values are concentrated in 0 and 1. In any case, the algorithm correctly estimates $q(\boldsymbol{z}_t)$ most of the times. As a consequence, when approximating (52)–(53) through sampling, one sample is typically sufficient (in this particular application). Similar results were obtained with $N = 10$, which are omitted for the sake of clarity.

*3) Outlier Misdetection:* we notice from the previous analysis on the estimation of $q(\boldsymbol{z}_t)$ that not all values are either 0 and 1, and that there is some probability associated with values in (0,1). Therefore, this uncertainty may lead to an outlier misdetection (i.e., contaminated observations which are regarded as clean ones). The percentage of misdetected outliers for the different methods (i.e., not for the standard EKF which do not allow to detect outliers) is shown. In this case we see that increasing the value $N$ from 1 to 10 within the general VBKF reduces the misdetection, because this value controls the accuracy of the expectation approximation in (52)–(53). Therefore, increasing $N$ leads to a lower misdetection and in turn lower RMSE. Regardless of $N$, it is shown that taking into account the correlated observations in the filter formulation as expected improves the results compared to the independent outlier indicator VBKF. That is one of the main reasons of the performance loss in terms of RMSE previously shown in Fig. 9. Notice that the VBKF with scalar outlier indicator is not shown for the similar reasons as in Fig. 5.
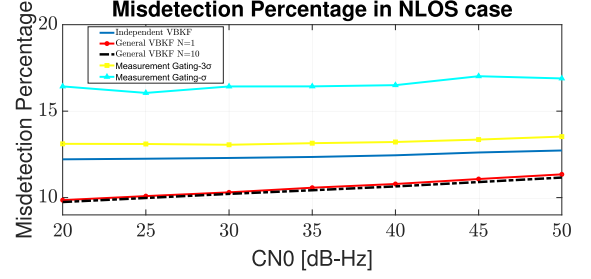


Fig. 9. Misdetection performance for the VBKF with independent outlier indicator and the general VBKF method as a function of $C/N_0$ in the NLOS scenario.

## IV. CONCLUSION

This article presented a generalization of the Variational Bayes Kalman filter (VBKF) for correlated measurement models. The VBKF acts as a method for outlier detection and mitigation in the observations in filtering problems. The original method, proposed in [35], considered a single indicator for the vector-valued observations, therefore rejecting the entire observation vector when a single element in it is faulty. The present contribution generalized VBKF to having multiple indicators (one per observation) and considering the more comprehensive case of correlated observations. Practical implementation solutions are proposed to avoid the otherwise growing complexity with the observations' dimension. The method is validated on a precise-positioning problem, namely the popular real-time kinematic (RTK) procedure, where correlated data appears. Results in this context show the benefits of the generalized VBKF method, as compared to non-robust filtering or the single-indicator VBKF.

### APPENDIX A
#### COMPUTATION OF $q(z_t^{(i)})$

To keep $q(z_t^{(i)})$ Bernoulli distributed, (32) is substituted in (37) and the equation is derived as:

$$
\begin{aligned}
\ln\left[q(z_t^{(i)})\right] = \mathbb{E}_{\pi_t^{(i)}, \boldsymbol{x}_t, \boldsymbol{z}_t^{(-i)}} \Bigg\{ &-\frac{1}{2}\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}(\boldsymbol{x}_t)\right)^\top \\
&\times \left(\boldsymbol{\Sigma}_{-i,-i}^{-1} + \frac{\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\frac{\sigma_{i,i}^2}{z_t^{(i)}} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right) \\
&\times \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}(\boldsymbol{x}_t)\right) - \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}(\boldsymbol{x}_t)\right)^\top \\
&\times \left(-\left(\boldsymbol{\Sigma}_{-i,-i}^{-1} + \frac{\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\frac{\sigma_{i,i}^2}{z_t^{(i)}} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right) \boldsymbol{\Sigma}_{-i,i} \frac{z_t^{(i)}}{\sigma_{i,i}^2}\right) \\
&\times \left(y_t^{(i)} - h^{(i)}(\boldsymbol{x}_t)\right) - \frac{1}{2}\left(y_t^{(i)} - h^{(i)}(\boldsymbol{x}_t)\right)^\top \\
&\times \left(\frac{\sigma_{i,i}^2}{z_t^{(i)}} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\right)^{-1}\left(y_t^{(i)} - h^{(i)}(\boldsymbol{x}_t)\right) \\
&+ z_t^{(i)}\ln[\pi_t^{(i)}] + \left(1 - z_t^{(i)}\right)\ln[1 - \pi_t^{(i)}]
\end{aligned}
$$

$$-0.5z_t^{(i)} \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])|$$

$$-0.5z_t^{(i)} \ln[2\pi] - 0.5(1 - z_t^{(i)}) \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])| + \kappa \Bigg\}. \tag{70}$$

Following the transform from (34) to (36), we can write that:

$$\ln\left[q(z_t^{(i)})\right] = \mathbb{E}_{\pi_t^{(i)},\boldsymbol{x}_t,\boldsymbol{z}_t^{(-i)}} \Bigg\{ -\frac{1}{2}\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top$$

$$\times \left(\boldsymbol{\Sigma}_{-i,-i}^{-1} + \frac{z_t^{(i)}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)$$

$$\times \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right) + \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top$$

$$\times \left(\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\frac{z_t^{(i)}}{\sigma_{i,i}^2} + \frac{z_t^{(i)}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}{\sigma_{i,i}^4 - \sigma_{i,i}^2\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)$$

$$\times \left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right) - \frac{1}{2}\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)^\top$$

$$\times \frac{z_t^{(i)}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)$$

$$+ z_t^{(i)} \ln[\pi_t^{(i)}] + \left(1 - z_t^{(i)}\right)\ln[1 - \pi_t^{(i)}]$$

$$-0.5z_t^{(i)} \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])| \tag{71}$$

$$-0.5z_t^{(i)} \ln[2\pi] - 0.5(1 - z_t^{(i)}) \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])| + \kappa \Bigg\}. \tag{72}$$

Given the expectation over $\{\boldsymbol{x}_t, \pi_t^{(i)}, \boldsymbol{z}_t^{(-i)}\}$, we can regard $-\frac{1}{2}(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}(\boldsymbol{x}_t))^\top\boldsymbol{\Sigma}_{-i,-i}^{-1}(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}(\boldsymbol{x}_t))$ in the first term as a normalization coefficient and thus remove it from the previous equation to further obtain that:

$$\ln\left[q(z_t^{(i)})\right] = \mathbb{E}_{\pi_t^{(i)},\boldsymbol{x}_t,\boldsymbol{z}_t^{(-i)}} \Bigg\{ -\frac{1}{2}\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top$$

$$\times \left(\frac{z_t^{(i)}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)$$

$$\times \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right) + \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top$$

$$\times \left(\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\frac{z_t^{(i)}}{\sigma_{i,i}^2} + \frac{z_t^{(i)}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}{\sigma_{i,i}^4 - \sigma_{i,i}^2\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)$$

$$\times \left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right) - \frac{1}{2}\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)^\top$$

$$\times \frac{z_t^{(i)}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)$$

$$+ z_t^{(i)} \ln[\pi_t^{(i)}] + \left(1 - z_t^{(i)}\right)\ln[1 - \pi_t^{(i)}]$$

$$-0.5z_t^{(i)} \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])| - 0.5z_t^{(i)} \ln[2\pi]$$

Taking advantage of the properties of the trace of matrix,

$$\ln\left[q(z_t^{(i)})\right] = \mathbb{E}_{\pi_t^{(i)},\boldsymbol{x}_t,\boldsymbol{z}_t^{(-i)}} \Bigg\{ -\frac{1}{2}\mathrm{Tr}\left[\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)\right.$$

$$\times \left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top \left(\frac{z_t^{(i)}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)\right]$$

$$+ \mathrm{Tr}\left[\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top\right.$$

$$\times \left(\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\frac{z_t^{(i)}}{\sigma_{i,i}^2} + \frac{z_t^{(i)}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}{\sigma_{i,i}^4 - \sigma_{i,i}^2\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)\right]$$

$$-\frac{1}{2}\mathrm{Tr}\left[\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)^\top\right.$$

$$\left.\times \frac{z_t^{(i)}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right]$$

$$+ z_t^{(i)} \ln[\pi_t^{(i)}] + \left(1 - z_t^{(i)}\right)\ln[1 - \pi_t^{(i)}]$$

$$-0.5z_t^{(i)} \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])| - 0.5z_t^{(i)} \ln[2\pi]$$

$$-0.5(1 - z_t^{(i)}) \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])| + \kappa \Bigg\}. \tag{74}$$

With $z_t^{(i)}$ being a scalar we can finally obtain:

$$\ln\left[q(z_t^{(i)})\right] = \mathbb{E}_{\pi_t^{(i)},\boldsymbol{x}_t,\boldsymbol{z}_t^{(-i)}} \Bigg\{ -\frac{1}{2}z_t^{(i)}$$

$$\times \mathrm{Tr}\left[\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top\right.$$

$$\left.\times \left(\frac{\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)\right]$$

$$+ z_t^{(i)}\mathrm{Tr}\left[\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)\left(\boldsymbol{y}_t^{(-i)} - \boldsymbol{h}^{(-i)}\left(\boldsymbol{x}_t\right)\right)^\top\right.$$

$$\left.\times \left(\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}\frac{1}{\sigma_{i,i}^2} + \frac{\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}{\sigma_{i,i}^4 - \sigma_{i,i}^2\boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right)\right]$$

$$-\frac{1}{2}z_t^{(i)}\mathrm{Tr}\left[\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)\left(y_t^{(i)} - h^{(i)}\left(\boldsymbol{x}_t\right)\right)^\top\right.$$

$$\left.\times \frac{1}{\sigma_{i,i}^2 - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i}}\right]$$

$$+ z_t^{(i)} \ln[\pi_t^{(i)}] + \left(1 - z_t^{(i)}\right)\ln[1 - \pi_t^{(i)}]$$

$$-0.5z_t^{(i)} \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 1])| - 0.5z_t^{(i)} \ln[2\pi]$$

$$-0.5(1 - z_t^{(i)}) \ln|\mathbf{C}([\boldsymbol{z}_t^{(-i)}, z_t^{(i)} = 0])| + \kappa \Bigg\}. \tag{75}$$

REFERENCES

[1] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, 2nd ed. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, 2012.

[2] D. Dardari, P. Closas, and P. M. Djurić, "Indoor tracking: Theory, methods, and technologies," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1263–1278, Apr. 2015.

[3] J. Dunik, S. K. Biswas, A. G. Dempster, T. Pany, and P. Closas, "State estimation methods: Overview and application in navigation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 35, no. 12, pp. 16–31, Dec. 2020.

[4] B. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, New Jersey, NJ, USA: Prentice-Hall, 1979.

[5] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[6] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[7] P. Djurić *et al.*, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 140, no. 2, pp. 19–38, Sep. 2003.

[8] A. Doucet, N. D. Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York, NY, USA: Springer, 2001.

[9] J. Vilá-Valls, E. Chaumette, F. Vincent, and P. Closas, "Modelling mismatch and noise statistics uncertainty in linear MMSE estimation," in *Proc. 27th Eur. Signal Process. Conf.*, Sep. 2019, pp. 1–5.

[10] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2003, vol. 4, pp. 1979–1993.

[11] T. Schon, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state–space models," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2279–2289, Jul. 2005.

[12] E. Chaumette *et al.*, "Minimum variance distortionless response estimators for linear discrete state-space models," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 2048–2055, Apr. 2017.

[13] E. Chaumette *et al.*, "On LMVDR estimators for LDSS models: Conditions for existence and further applications," *IEEE Trans. Autom. Control*, vol. 64, no. 6, pp. 2598–2605, Jun. 2019.

[14] J. Vilá-Valls, D. Vivet, E. Chaumette, F. Vincent, and P. Closas, "Recursive linearly constrained wiener filter for robust multi-channel signal processing," *Signal Process.*, vol. 167, Feb. 2020, Art. no. 107291.

[15] J. Vilà-Valls, P. Closas, and A. F. García-Fernández, "Uncertainty exchange through multiple Quadrature Kalman filtering," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1825–1829, Dec. 2016.

[16] P. Closas and M. F. Bugallo, "Improving accuracy by iterated multiple particle filtering," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 531–534, Aug. 2012.

[17] J. Dunik, O. Straka, O. Kost, and J. Havlik, "Noise covariance matrices in state-space models: A survey and comparison of estimation methods—Part I," *Int. J. Adaptive Control Signal Process.*, vol. 31, pp. 1505–1543, 2017.

[18] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 596–600, Mar. 2009.

[19] T. Ardeshiri, E. Ozkan, U. Orguner, and F. Gustafsson, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2450–2454, Dec. 2015.

[20] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrics," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 594–601, Feb. 2018.

[21] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[22] G. Agamennoni and E. Nebot, "Robust estimation in non-linear state-space models with state-dependent noise," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 2165–2175, Apr. 2014.

[23] S. Saha, "Noise robust online inference for linear dynamic systems," Apr. 2015. [Online]. Available: http://arxiv.org/abs/1504.05723

[24] J. Vilà-Valls and P. Closas, "NLOS mitigation in indoor localization by marginalized Monte Carlo Gaussian smoothing," *EURASIP J. Adv. Signal Process.*, vol. 62, Aug. 2017.

[25] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1898–1902, May 2015.

[26] Y. Huang, N. L. Y. Zhang, and J. Chambers, "A novel robust student's-t-based Kalman filter," *IEEE Trans. Aerosp. Elec. Syst.*, vol. 53, no. 3, pp. 1545–1554, Jun. 2017.

[27] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Skew-*t* filter and smoother with improved covariance matrix approximation," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5618–5633, Nov. 2018.

[28] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, and J. Chambers, "Robust Kalman filters based on Gaussian scale mixture distribution with application to target tracking," *IEEE Trans. Syst., Man. Cybern. : Syst.*, vol. 49, no. 10, pp. 2082–2096, Oct. 2019.

[29] Y. Huang, Y. Zhang, Y. Zhao, and J. Chambers, "A novel robust Gaussian–Student's t mixture distribution based Kalman filter," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3606–3620, Jul. 2019.

[30] F. Caron, M. Davy, A. Doucet, E. Duflos, and P. Vanheeghe, "Bayesian inference for linear dynamic models with Dirichlet process mixtures," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 71–84, Jan. 2008.

[31] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[32] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[33] L. Chang and K. Li, "Unified form for the robust Gaussian information filtering based on M-estimate," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 412–416, Apr. 2017.

[34] H. Wang, H. Li, W. Zhang, J. Zuo, and H. Wang, "A unified framework for M-estimation based robust Kalman smoothing," *Signal Process.*, vol. 158, pp. 61–65, May 2019.

[35] H. Wang, H. Li, J. Fang, and H. Wang, "Robust Gaussian Kalman filter with outlier detection," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1236–1240, Aug. 2018.

[36] Y. Bar-Shalom, "Negative correlation and optimal tracking with Doppler measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 3, pp. 1117–1120, Jul. 2001.

[37] Y. Zhu, R. S. Blum, Z.-Q. Luo, and K. M. Wong, "Unexpected properties and optimum-distributed sensor detectors for dependent observation cases," *IEEE Trans. Autom. Control*, vol. 45, no. 1, pp. 62–72, Jan. 2000.

[38] H. Ma and Y. Liu, "Correlation based video processing in video sensor networks," in *Proc. IEEE Int. Conf. Wireless Netw., Commun. Mobile Comput.*, 2005, pp. 987–992.

[39] K. Yuen, B. Liang, and B. Li, "A distributed framework for correlated data gathering in sensor networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 578–593, Jan. 2008.

[40] H. Li, D. Medina, J. Vilà-Valls, and P. Closas, "Robust Kalman filter for RTK positioning under signal-degraded scenarios," in *Proc. 32nd Int. Tech. Meet. Satell. Division Inst. Navigation (ION GNSS+ 2019)*, Miami, FL, USA, Sep. 2019, pp. 3717–3729.

[41] D. Dardari, E. Falletti, and M. Luise, *Satellite and Terrestrial Radio Positioning Techniques: A Signal Processing Perspective*. Cambridge, MA, USA: Academic Press, 2011.

[42] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[43] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*. New York, NY, USA: Springer-Verlag, 2005.

[44] R. B. Langley, "RTK GPS," *GPS World*, vol. 9, no. 9, pp. 70–76, 1998.

[45] P. J. Teunissen and O. Montenbruck, Eds., *Springer Handbook of Global Navigation Satellite Systems*. Cham, Switzerland: Springer International Publishing, 2017. [Online]. Available: https://doi.org/10.10072F978-3-319-42928-1

[46] D. Medina, V. Centrone, R. Ziebold, and J. García, "Attitude determination via GNSS carrier phase and inertial aiding," in *Proc. 32nd Int. Tech. Meet. Satell. Division Inst. Navigation (ION GNSS 2019)*, Oct. 2019, pp. 2964–2979.

[47] N. Viandier, D. Nahimana, J. Marais, and E. Duflos, "GNSS performance enhancement in urban environment based on pseudo-range error model," in *Proc. IEEE/ION Position, Location Navigation Symp.*, 2008, pp. 377–382.

[48] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation, and Integration*. Marblehead, MA, USA: Wiley, 2007.

[49] P. D. Groves, "Principles of GNSS, inertial, and multisensor integrated navigation systems, [book review]," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 30, no. 2, pp. 26–27, Feb. 2015.
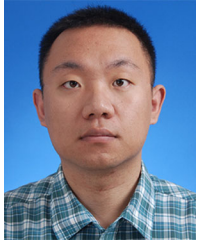
**Haoqing Li** received the B.S. degree in electrical engineering from Wuhan University, China, in 2016 and the M.S. degree in electrical and computer engineering from Northeastern University, Boston, MA, in 2018, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests include GNSS signal processing, antijamming technology, and robust statistics.

**Jordi Vilà-Valls** (Senior Member, IEEE) received the M.S. degree in electrical engineering from the Universitat Politècnica de Catalunya, Spain, and the Grenoble Institute of Technology (INPG), France, in 2006, and the Ph.D. degree in signal processing from INPG in 2010. He is currently an Associate Professor with the Institut Supèrieur de l'Aèronautique et de l'Espace (ISAE-SUPAERO), University of Toulouse, France. His primary areas of interest include statistical signal processing, estimation and detection theory, nonlinear Bayesian inference and robust filtering, with applications to localization, tracking, navigation, and remote sensing. He is an Elected Member of the EURASIP Technical Area Committee on Theoretical and Methodological Trends in Signal Processing, the Student Activities Chair for EUSIPCO'19 and EUSIPCO'21, and the Finance Chair for IEEE CAMSAP'19.
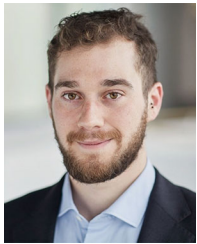
**Daniel Medina** (Student Member, IEEE) received the B.S. degree in electrical engineering from the University of Malaga in 2014 and the M.S. degree in computer science from the Charles III University of Madrid in 2016. He is currently working toward the Ph.D. degree in multisensor fusion for precise positioning and attitude determination. Since 2016, he has been a Research Fellow with the Institute of Communications and Navigation, German Aerospace Center.

**Pau Closas** (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering and the M.S. degree in advanced mathematics and mathematical engineering from the Universitat Politècnica de Catalunya, Spain, in 2003, 2009, and 2014, respectively. He is currently an Assistant Professor in electrical and computer engineering with Northeastern University, Boston, MA, USA. His primary areas of interests include statistical signal processing, stochastic filtering, robust filtering, and machine learning, with applications to positioning and localization systems. He was the recipient of the EURASIP Best Ph.D. Thesis Award 2014, the 9th Duran Farell Award for Technology Research, the 2016 ION's Early Achievements Award, and the 2019 NSF CAREER Award. He has volunteered in editorial roles for *Navigation*, PROCEEDINGS OF THE IEEE, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and *IEEE Signal Processing Magazine*, and has been actively involved in organizing committees of a number of conference such as EUSIPCO (2011, 2019–2022), IEEE SSP'16, IEEE/ION PLANS'20, and IEEE ICASSP'20.