Robust Kriged Kalman Filtering

Brian Baingana[†], Emiliano Dall' Anese[*], Gonzalo Mateos[⋆], Georgios B. Giannakis[†]

[†] Dept. of ECE and Digital Technology Center, University of Minnesota, Minneapolis, MN, USA

[*] National Renewable Energy Laboratory, Golden, CO, USA

[⋆] Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

***Abstract*—Although the kriged Kalman filter (KKF) has well-documented merits for prediction of spatial-temporal processes, its performance degrades in the presence of outliers due to anomalous events, or measurement equipment failures. This paper proposes a robust KKF model that explicitly accounts for presence of measurement outliers. Exploiting outlier sparsity, a novel ℓ_1-regularized estimator that jointly predicts the spatial-temporal process at unmonitored locations, while identifying measurement outliers is put forth. Numerical tests are conducted on a synthetic Internet protocol (IP) network, and real transformer load data. Test results corroborate the effectiveness of the novel estimator in joint spatial prediction and outlier identification.**

***Index terms*— Robust estimation, kriging, Kalman filter, sparsity, IP path delay monitoring.**

I. Introduction

The kriged Kalman filter (KKF) has been shown useful for prediction of spatial-temporal processes arising in settings as diverse as environmetrics [1], [2], path delay monitoring in IP networks [10], and spectrum sensing [9]. It combines the merits of the Kalman filter (KF), the "work-horse" for state estimation and prediction in dynamic processes, as well as kriging [3], a linear spatial interpolation technique whose origins can be traced back to the mining community. Although the KKF is provably optimal when noise terms are Gaussian, due to its reliance on a squared-error risk function, its predictive performance degrades when contaminated with outliers. Outliers deviate significantly from nominal models for spatial-temporal processes, and typically arise as a result of anomalous events, or failure of measurement equipment e.g., faulty sensors in environmental monitoring applications.

Several approaches have been proposed for developing KF estimators that are robust to outliers in measurements, the state model, or both [5], [7]. Although these can be deployed for robustification of the Kalman step in the KKF, a systematic approach for jointly robustifying both the Kalman and kriging steps has not been investigated. The present paper extends the benefits of these prior robust estimators to the KKF, by deriving an estimator that reliably identifies both the locations and values of measurement outliers.

This work explicitly models outliers by introducing an additive vector in the KKF measurement equation. The crux of the proposed approach lies in recognizing that the outlier vector is inherently sparse; that is, non-zero entries only exist at indices corresponding to outlier-contaminated measurements. Capitalizing on this prior knowledge, a link is drawn with ℓ_1-minimization approaches, that have recently been advocated for sparsity-promoting regularized estimators. This approach lies within the broader context of recently proposed robust models that exploit sparsity to facilitate outlier-rejection in canonical learning and estimation tasks; see e.g., [5], [6]. The robust KKF developed here extends the benefits of this framework to prediction of spatial-temporal processes, especially in scenarios where the measurement process is known to be prone to outlier contamination.

The rest of this paper is organized as follows. The robust KKF model is presented, and a formal statement of the problem is given in Section II. Section III derives a novel estimator that capitalizes on outlier sparsity. Practical applications involving path delay monitoring for IP networks, and spatial prediction of transformer loads in energy grids are identified in Section IV. Results of numerical tests conducted on synthetic IP path delay data, and real transformer load data are presented in Section V, while conclusions and future directions are given in Section VI.

II. Model and Problem Statement

Consider a spatial-temporal process with states $\{\mathbf{x}(t) \in \mathbb{R}^M\}_{t=1}^T$ evolving over time intervals $t = 1, \ldots, T$, as follows

$$\mathbf{x}(t) = \mathbf{H}\mathbf{x}(t-1) + \boldsymbol{\eta}(t) \tag{1}$$

where the noise $\boldsymbol{\eta}(t)$ is assumed $\mathbf{0}$-mean Gaussian with covariance matrix $\mathbf{C}_\eta$, i.e., $\boldsymbol{\eta}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\eta)$. Suppose the process is observed at K spatial locations $\{\mathbf{s}_k\}_{k=1}^K$ per interval t, and each observation $y(\mathbf{s}_k, t)$ admits the following KKF measurement model [1],

$$y(\mathbf{s}_k, t) = \mu(\mathbf{s}_k, t) + \nu(\mathbf{s}_k, t) + \epsilon(\mathbf{s}_k, t), \quad k = 1, \ldots, K. \tag{2}$$

The so-termed trend in (2) admits the following linear model,

$$\mu(\mathbf{s}_k, t) = \sum_{m=1}^{M} \alpha_m(\mathbf{s}_k) x_m(t) \tag{3}$$

with $\{\alpha_m(\mathbf{s}_k)\}_{m=1}^M$ denoting spatially-varying coefficients. Small-scale spatial variations are captured by the term $\nu(\mathbf{s}_k, t)$ whose values are correlated in space, while measurement noise is captured through $\epsilon(\mathbf{s}_k, t)$, and it is assumed spatially and

The work of B. Baingana and G. B. Giannakis was supported by NSF grants 1343248, 1423316, 1442686, 1508993, 1509040, and grant ARO W911NF-15-1-0492. The work of E. Dall'Anese was supported by the Laboratory Directed Research and Development Program at the National Renewable Energy Laboratory.

temporally white. The KKF predicts values $\{y(\mathbf{s}_k,t)\}_{k=K+1}^{K+J}$ at J unobserved locations $\{\mathbf{s}_k\}_{k=K+1}^{K+J}$ for $t=1,\ldots,T$ by: 1) estimating the state $\hat{\mathbf{x}}(t)$ during each time interval via ordinary Kalman filtering (KF) equations; and 2) predicting $\{\hat{y}(\mathbf{s}_k,t)\}_{k=K+1}^{K+J}$ in a "kriging" step that essentially boils down to linear minimum mean square estimation (LMMSE).

The present paper proposes an outlier-aware measurement model by augmenting (2) with the outlier term $o(\mathbf{s}_k,t)$, that is,

$$y(\mathbf{s}_k,t)=\mu(\mathbf{s}_k,t)+\nu(\mathbf{s}_k,t)+\epsilon(\mathbf{s}_k,t)+o(\mathbf{s}_k,t) \quad (4)$$

for $k=1,\ldots,K$. Critical to the success of the novel approach is recognizing that $o(\mathbf{s}_k,t)\neq 0$ only if $y(\mathbf{s}_k,t)$ is contaminated by an outlier.

Let $\mathbf{y}_o(t):=[y(\mathbf{s}_1,t),\ldots,y(\mathbf{s}_K,t)]^\top\in\mathbb{R}^K$ collect measurements at the K observed locations. Similarly, let the $K\times 1$ vectors $\mathbf{o}(t)$, $\boldsymbol{\nu}(t)$ and $\boldsymbol{\epsilon}(t)$ collect outliers at time t, the spatially correlated noise component, and measurement noise, respectively. Equation (2) can now be written in vector form as

$$\mathbf{y}_o(t)=\mathbf{A}_o\mathbf{x}(t)+\boldsymbol{\nu}(t)+\boldsymbol{\epsilon}(t)+\mathbf{o}(t) \quad (5)$$

where $\mathbf{A}_o:=[\boldsymbol{\alpha}^\top(\mathbf{s}_1),\ldots,\boldsymbol{\alpha}^\top(\mathbf{s}_K)]^\top$ and the $M\times 1$ vector $\boldsymbol{\alpha}(\mathbf{s}_k):=[\alpha_1(\mathbf{s}_k),\ldots,\alpha_M(\mathbf{s}_k)]^\top\in\mathbb{R}$. In addition, let $\mathbf{C}_\nu$ and $\mathbf{C}_\epsilon$ respectively denote the covariance matrices of $\boldsymbol{\nu}(t)$, $\boldsymbol{\epsilon}(t)$. Furthermore, define the $M\times J$ matrix $\mathbf{A}_u:=[\boldsymbol{\alpha}^\top(\mathbf{s}_{K+1}),\ldots,\boldsymbol{\alpha}^\top(\mathbf{s}_{K+J})]^\top$, and $\mathbf{y}_u(t):=[y(\mathbf{s}_{K+1},t),\ldots,y(\mathbf{s}_{K+J},t)]^\top\in\mathbb{R}^J$ whose entries correspond to instances of the unmonitored spatial-temporal process at time t.

Given $\{\mathbf{y}_o(t)\}_{t=1}^T$, $\mathbf{H}$, $\mathbf{A}_o$, $\mathbf{A}_u$, $\mathbf{C}_\nu$, $\mathbf{C}_\epsilon$, and $\mathbf{C}_\eta$, the goal of the robust KKF is to optimally identify the measurement outliers $\{\mathbf{o}(t)\}_{t=1}^T$, while predicting the values of the unobserved process $\{\mathbf{y}_u(t)\}_{t=1}^T$.

III. Exploiting outlier sparsity

Assume $\mathbf{o}(t)$ is known, and denote the outlier-compensated measurements by $\tilde{\mathbf{y}}_o(t):=\mathbf{y}_o(t)-\mathbf{o}(t)$, then the KF step recursively estimates the state per t as

$$\hat{\mathbf{x}}(t)=\mathbf{H}\hat{\mathbf{x}}(t-1)+\mathbf{K}(t)[\tilde{\mathbf{y}}_o(t)-\mathbf{A}_o\mathbf{H}\hat{\mathbf{x}}(t-1)] \quad (6)$$

with error covariance matrix

$$\begin{aligned}\mathbf{M}(t) &:= \mathbb{E}\left\{(\mathbf{x}(t)-\hat{\mathbf{x}}(t))(\mathbf{x}(t)-\hat{\mathbf{x}}(t))^\top\right\}\\ &= (\mathbf{I}-\mathbf{K}(t)\mathbf{A}_o)(\mathbf{H}\mathbf{M}(t-1)\mathbf{H}^\top+\mathbf{C}_\eta). \quad (7)\end{aligned}$$

Correspondingly, the Kalman gain matrix $\mathbf{K}(t)$ turns out to be

$$\begin{aligned}\mathbf{K}(t) &= (\mathbf{H}\mathbf{M}(t-1)\mathbf{H}^\top+\mathbf{C}_\eta)\mathbf{A}_o^\top(\mathbf{C}_\nu+\mathbf{C}_\epsilon\\ &\quad +\mathbf{A}_o[\mathbf{H}\mathbf{M}(t-1)\mathbf{H}^\top+\mathbf{C}_\eta]\mathbf{A}_o^\top)^{-1}. \quad (8)\end{aligned}$$

The "kriging" step leverages LMMSE, predicting the sequence $\{\mathbf{y}_u(t)\}_{t=1}^T$ via

$$\hat{\mathbf{y}}_u(t)=\mathbf{A}_u\hat{\mathbf{x}}(t)+\mathbf{A}_u\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{y}}_o(t)-\mathbf{A}_o\hat{\mathbf{x}}(t)) \quad (9)$$

per t, with $\boldsymbol{\Sigma}:=\mathbf{A}_o\mathbf{C}_\nu\mathbf{A}_o^\top+\mathbf{C}_\epsilon$. Rearranging (6) by collecting the coefficients of $\hat{\mathbf{x}}(t-1)$ yields

$$\hat{\mathbf{x}}(t)=\mathbf{K}(t)\tilde{\mathbf{y}}_o(t)+\{\mathbf{H}-\mathbf{K}(t)\mathbf{A}_o\mathbf{H}\}\hat{\mathbf{x}}(t-1) \quad (10)$$

which upon substitution into (9), one obtains

$$\begin{aligned}\hat{\mathbf{y}}_u(t) &= \{(\mathbf{A}_u-\mathbf{A}_u\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}_o)\mathbf{K}(t)\\ &\quad +\mathbf{A}_o\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}\}\tilde{\mathbf{y}}_o(t)\\ &\quad +\{(\mathbf{A}_u-\mathbf{A}_u\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}_o)\\ &\quad (\mathbf{H}-\mathbf{K}(t)\mathbf{A}_o\mathbf{H})\}\hat{\mathbf{x}}(t-1). \quad (11)\end{aligned}$$

Note that (11) can be conveniently written as

$$\hat{\mathbf{y}}_u(t)=\mathbf{B}(t)\tilde{\mathbf{y}}_o(t)+\mathbf{D}(t)\hat{\mathbf{x}}(t-1) \quad (12)$$

where $\mathbf{B}(t):=\mathbf{A}_u(\mathbf{I}-\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}_o)\mathbf{K}(t)+\mathbf{A}_o\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}$, $\mathbf{D}(t):=\mathbf{A}_u(\mathbf{I}-\mathbf{C}_\nu\mathbf{A}_o^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}_o)\mathbf{F}(t)$, and $\mathbf{F}(t):=\mathbf{H}-\mathbf{K}(t)\mathbf{A}_o\mathbf{H}$. Initializing $\hat{\mathbf{x}}(0)=\mathbf{0}$, $\{\hat{\mathbf{y}}_u(t)\}_{t=1}^T$ can be estimated entirely from outlier-compensated measurements, thus eliminating the intermediate step of determining $\{\hat{\mathbf{x}}(t)\}_{t=1}^T$. For example,

$$\begin{aligned}\hat{\mathbf{y}}_u(1) &= \mathbf{B}(1)\tilde{\mathbf{y}}_o(1)\\ \hat{\mathbf{y}}_u(2) &= \mathbf{B}(2)\tilde{\mathbf{y}}_o(2)+\mathbf{D}(2)\mathbf{K}(1)\tilde{\mathbf{y}}_o(1)\\ \hat{\mathbf{y}}_u(3) &= \mathbf{B}(3)\tilde{\mathbf{y}}_o(3)+\mathbf{D}(3)\mathbf{K}(2)\tilde{\mathbf{y}}_o(2)\\ &\quad +\mathbf{D}(3)\mathbf{F}(2)\mathbf{K}(1)\tilde{\mathbf{y}}_o(1). \quad (13)\end{aligned}$$

for $t=1,2,3$. In general, setting $\hat{\mathbf{x}}(0)=\mathbf{0}$, one can write (12) in matrix-vector form as

$$\hat{\mathbf{y}}_u=\boldsymbol{\Phi}_T\tilde{\mathbf{y}}_o \quad (14)$$

where $\hat{\mathbf{y}}_u:=[\hat{\mathbf{y}}_u^\top(1),\ldots,\hat{\mathbf{y}}_u^\top(T)]^\top$, and $\tilde{\mathbf{y}}_o:=[\tilde{\mathbf{y}}_o^\top(1),\ldots,\tilde{\mathbf{y}}_o^\top(T)]^\top$. Matrix $\boldsymbol{\Phi}_T$ is determined recursively per t as

$$\boldsymbol{\Phi}_t:=\begin{bmatrix}\mathbf{0} & \boldsymbol{\Phi}_{t-1}\\ \mathbf{B}(t) & \boldsymbol{\Psi}_t\end{bmatrix} \quad (15)$$

where $\boldsymbol{\Phi}_1=\mathbf{B}(1)$, $\boldsymbol{\Psi}_t:=[\boldsymbol{\Psi}_{t,1},\ldots,\boldsymbol{\Psi}_{t,t-1}]$, and

$$\boldsymbol{\Psi}_{t,l}:=\mathbf{D}(t)\left[\prod_{i=1}^{l-1}\mathbf{F}(t-i)\right]\mathbf{K}(t-l). \quad (16)$$

Letting $\mathbf{o}:=[\mathbf{o}^\top(1),\ldots,\mathbf{o}^\top(T)]^\top$ denote the sparse $KT\times 1$ outlier vector, substituting $\tilde{\mathbf{y}}_o=\mathbf{y}_o-\mathbf{o}$ into (14) leads to

$$\boldsymbol{\Phi}_T\mathbf{y}_o=\hat{\mathbf{y}}_u+\boldsymbol{\Phi}_T\mathbf{o}. \quad (17)$$

Exploiting the sparsity inherent to $\mathbf{o}$, one can recover the unknowns $\hat{\mathbf{y}}_u$ and $\mathbf{o}$ in (17), by solving a sparsity-promoting regularized least-squares problem. Indeed, this is tantamount to solving the following batch optimization problem

$$\text{(P0)}\quad \underset{\hat{\mathbf{y}}_u,\mathbf{o}}{\arg\min}\ \frac{1}{2}\|\boldsymbol{\Phi}_T\mathbf{y}_o-\hat{\mathbf{y}}_u-\boldsymbol{\Phi}_T\mathbf{o}\|_2^2+\lambda\|\mathbf{o}\|_1. \quad (18)$$

where the sparsity-promoting penalty is only applied with respect to $\mathbf{o}$, with the sparsity-controlling regularization parameter $\lambda>0$. (P0) can be efficiently solved using several off-the-shelf solvers, and details are omitted in the present paper for brevity.

IV. Applications

This section highlights two motivating practical applications for prediction of spatial-temporal processes, using outlier-contaminated observations.

A. Packet delay prediction in IP networks

Monitoring end-to-end packet delays in IP networks is crucial for operators to assess network health, diagnose faults, and plan for future upgrades. However, this is quite costly as it amounts to monitoring every end-to-end path in the network. In this case monitoring only a fraction of paths, and predicting delays on the remainder is well motivated; see e.g., [8], [10]. Consider a directed IP network encoded by the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ respectively denote constituent node and link sets. Let $\mathcal{P}$ denote the set of end-to-end paths, and define the corresponding cardinalities $V = |\mathcal{V}|$, $L = |\mathcal{E}|$, and $P = |\mathcal{P}|$. Define the binary routing matrix $\mathbf{R} \in \{0,1\}^{P \times L}$ with entry (i, j)

$$r_{ij} = \begin{cases} 1, & \text{if path } i \text{ contains link } j \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Let $x_l(t)$ denote the queuing delay on link l at time t, and collect per-link queuing delays into $\mathbf{x}(t) := [x_1(t), \ldots, x_L(t)]^\top$. Temporal random fluctuations observed in queueing delays of typical networks can be captured via a random walk model [10], that is,

$$\mathbf{x}(t) = \mathbf{x}(t-1) + \boldsymbol{\eta}(t). \tag{20}$$

Define $\boldsymbol{\kappa}(t) := [\kappa_1(t), \ldots, \kappa_L(t)]^\top$ where $\kappa_l(t)$ captures the propagation delay along link l. Actual path delays sum up their constituent link delays, as encoded by $\mathbf{R}$, leading to

$$\bar{\mathbf{y}}(t) = \mathbf{R}(\mathbf{x}(t) + \boldsymbol{\kappa}(t)) \tag{21}$$

where $\bar{\mathbf{y}}(t) := [\bar{y}_1(t), \ldots, \bar{y}_P(t)]^\top$, and $\bar{y}_p(t)$ denotes the actual packet delay incurred by path p.

Suppose $\mathcal{P}_o$ consists of the set of monitored paths, partition $\bar{\mathbf{y}}(t)$ and $\mathbf{R}$ as $\bar{\mathbf{y}}(t) = [\bar{\mathbf{y}}_o^\top(t), \bar{\mathbf{y}}_u^\top(t)]^\top$, and $\mathbf{R} = [\mathbf{R}_o^\top, \mathbf{R}_u^\top]^\top$ respectively. Entries of $\bar{\mathbf{y}}_o(t)$ (resp. rows of $\mathbf{R}_o$) are indexed by $\mathcal{P}_o$, while those of $\bar{\mathbf{y}}_u(t)$ (resp. rows of $\mathbf{R}_u$) are indexed by $\mathcal{P} \setminus \mathcal{P}_o$. Accounting for outliers, the observed path delay measurements adhere to the measurement model in (5),

$$\mathbf{y}_o(t) = \mathbf{R}_o\mathbf{x}(t) + \boldsymbol{\nu}(t) + \boldsymbol{\epsilon}(t) + \mathbf{o}(t) \tag{22}$$

where $\boldsymbol{\nu}(t) = \mathbf{R}_o\boldsymbol{\kappa}(t)$, and $\bar{\mathbf{y}}_o(t) = \mathbf{R}_o\mathbf{x}(t) + \boldsymbol{\nu}(t)$. Unmonitored path delay predictions $\{\mathbf{y}_u(t)\}_{t=1}^T$ and outlier estimates $\{\mathbf{o}(t)\}_{t=1}^T$ can be obtained by solving (P0) upon setting $\mathbf{H} = \mathbf{I}$ and $\mathbf{A}_o = \mathbf{R}_o$.

B. Load prediction in power grids

Modern power grids are equipped with meters that periodically record electric power consumption at specific points within the system, e.g., consumer premises and substations [11]. These measurements constitute the so-termed *load curve data*, and are critical assets for facilitating operational decisions in the envisioned smart grid system. However, exhaustive monitoring of loads at all locations is a costly proposition that may not be practical for most operators. It is then of interest to monitor the load profiles at a few optimally selected points, and then make predictions about profiles at unmonitored locations. This approach hinges in an intrinsic correlation in the load consumption of households in both spatial and temporal domains. It should be also pointed out that inferring missing data from a sub-set of measurements is critical in order to gain full visibility of the power system state in case of communication failures of the advance metering infrastructure (AMI) [12]. In this context, it is important to "cleanse" the data of measurement outliers that are often attributed to unscheduled maintenance, leading to shutdown of heavy industrial loads, weather constraints, holidays, and major sporting events, to name a few [13].

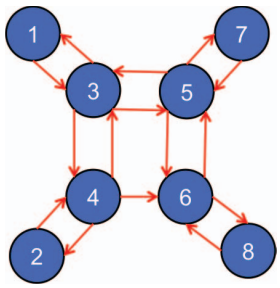

Fig. 1. Synthetic IP network with $L = 15$, and $P = 56$.

Consider a power grid comprising L distribution transformers, each serving a residential neighborhood (i.e., a group of households). Let $\mathbf{x}(t) := [x_1(t), \ldots, x_L(t)]^\top$ collect the transformer-specific load profile. Assume that $\mathbf{x}(t)$ adheres to the random walk model (21). Most distribution systems serve relatively homogenous regions with customers exhibiting similar energy consumption patterns. Letting $\boldsymbol{\mu}(t)$ capture the contribution of this correlated spatial variability, then one can model the total transformer load per t as

$$\mathbf{z}(t) = \mathbf{x}(t) + \boldsymbol{\mu}(t). \tag{23}$$

Acquiring load measurements on a few transformers is tantamount to selecting entries of $\mathbf{z}(t)$. Let the selection matrix $\mathbf{S}$ comprise rows of the identity matrix $\mathbf{I}$ corresponding to indices of selected entries in $\mathbf{z}(t)$. Incorporating measurement noise and outliers, one obtains the following observation model:

$$\begin{aligned} \mathbf{y}_o(t) &= \mathbf{S}\left(\mathbf{x}(t) + \boldsymbol{\mu}(t)\right) + \boldsymbol{\epsilon}(t) + \mathbf{o}(t) \\ &= \mathbf{S}\mathbf{x}(t) + \boldsymbol{\nu}(t) + \boldsymbol{\epsilon}(t) + \mathbf{o}(t). \end{aligned} \tag{24}$$

Similarly, load predictions and outlier estimates can be obtained by solving (P0) with $\mathbf{H} = \mathbf{I}$ and $\mathbf{A}_o = \mathbf{S}$.

V. Numerical Tests

A. Synthetic IP delay data

The robust KKF was tested on a synthetic network comprising 8 nodes and 15 directed links; see Figure 1. End-to-end paths ($P = 56$) were identified using shortest-path criteria, and $\mathbf{R}$ was accordingly constructed. Synthetic link

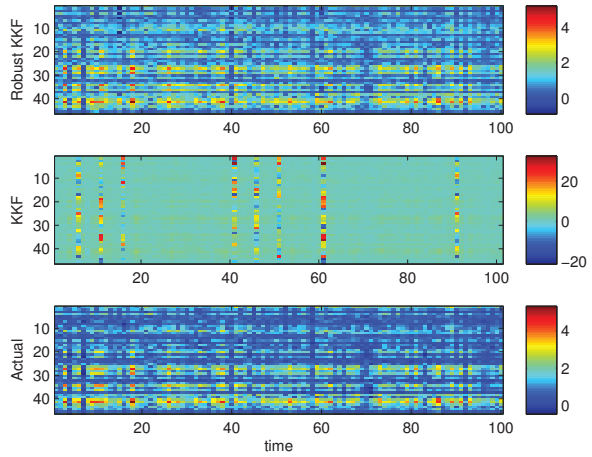

Fig. 2. Heatmaps depicting both the observed delays (paths 1 to 5) and the predicted delays on unobserved paths (paths 6 to 56) as well as the corruptive effect of measurement outliers on KKF-based predictions.

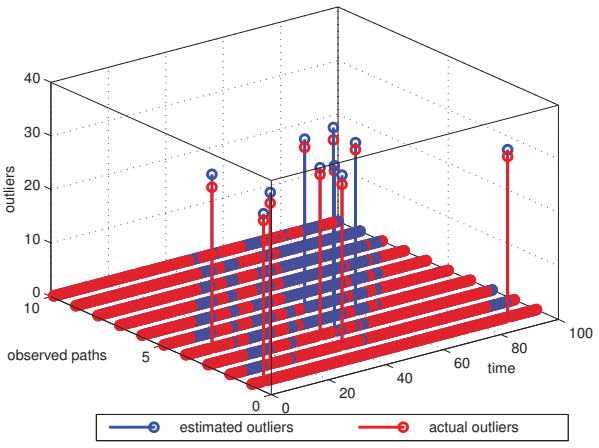

Fig. 3. Actual and estimated outliers with $\lambda = 2.25$.

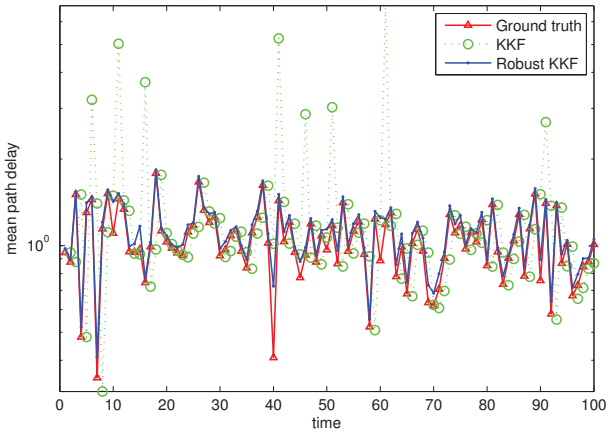

Fig. 4. Mean predicted delays on unmonitored paths, highlighting the adverse effect of outliers on the plain KKF.

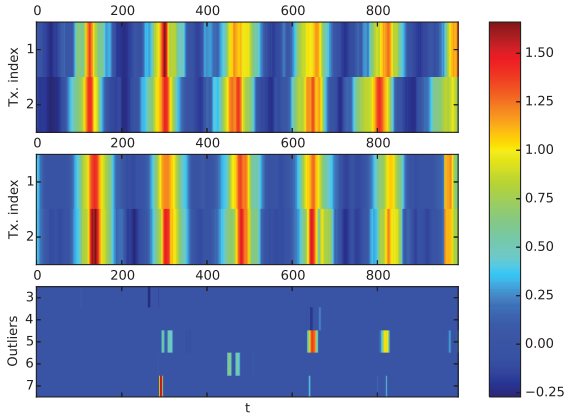

Fig. 5. Actual standardized loads (top) on unmonitored transformers, predicted transformer loads (middle), and identified outliers (bottom).

delays (in milliseconds) were generated over $T = 100$ time slots via (20), with $\boldsymbol{\eta}(t) \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 \mathbf{I})$ and $\sigma_\eta = 0.02$. To generate path delays, $\mathbf{C}_\nu$ was set to $\mathbf{R}^\top \mathbf{R} / \max_{i,j} [\mathbf{R}^\top \mathbf{R}]_{ij}$, and $\boldsymbol{\nu}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\nu)$. Using the criterion in [8], 10 paths were selected, and $\mathbf{R}$ was partitioned into $\mathbf{R}_o$ and $\mathbf{R}_u$.With $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$ and $\sigma_\epsilon = 0.01$, 10 outliers at known locations were added to the observed path measurements. Estimates of outlier locations and magnitudes were obtained via the robust KKF. Upon removal of the identified outliers, the ordinary KKF was run on the cleansed data to obtain more accurate predictions.

Figure 2 compares prediction results from both the outlier-agnostic KKF, and those obtained after outlier rejection with the robust KKF ($\lambda = 2.25$). In this case, predictions by the outlier-agnostic KKF are severely hurt by presence of outliers. As demonstrated in the plots, presence of outliers severely distorts the predictive performance of KKF, while cleansing the data with the developed robust KKF leads to predictions closely reflecting actual path delays. Figure 3 plots outliers identified by the robust KKF against the ground-truth. Remarkably, with $\lambda = 2.25$, the robust estimator is able to successfully identify all known outliers.

Network operators are often interested in aggregate metrics e.g., network-wide mean path delay. Figure 4 plots the mean predicted path delays over unmonitored paths per t, compared with the ground-truth averages. Clearly, predictions from the robust KKF were more reliable in tracking the actual mean path delay, as compared to the outlier-agnostic KKF, whose profile exhibits spikes due to the presence of outliers.

B. Transformer load data

Real load data were measured from seven distribution transformers located in Anatolia, California, over a 7-day period

in August 2012 [14]. With each transformer serving $10-12$ houses, measurements were acquired every 5 seconds. For this experiment, five of the transformers were designated for load monitoring, and predictions were made for the remaining unmonitored transformers. First, the load data were standardized, and then empirical Bayesian estimates for the covariance matrices $\mathbf{C}_\nu$ and $\mathbf{C}_\eta$ were evaluated along the guidelines in [9]. This was accomplished by setting $\mathbf{C}_\epsilon = 10^{-3}\mathbf{I}$.

Figure 5 plots heatmaps of predicted loads on the unmonitored transformers, as well as the identified measurement outliers with $\lambda = 0.015$. The load predictions give a reasonably accurate depiction of the true unmonitored loads. Moreover, it turns out that identified outliers coincide with significant spikes in the load data for the 5 monitored transformers.

VI. Conclusions and Future Work

This paper developed a robust KKF for spatial-temporal prediction, when the process measurements obtained at monitored locations are contaminated by outliers. Leveraging sparsity inherent to measurement outliers, a batch ℓ_1-norm regularized least-squares estimator was put forth. Building upon standard KKF recursions, it turns out that the resulting optimization problem is convex, and can be readily solved by standard off-the-shelf solvers. Experiments on synthetic IP path delays, and transformer load data corroborated the effectiveness of the advocated approach. This work opens up a number of directions for future research. Recognizing the limitations of batch estimators, an online robust KKF that is capable of operating in real-time streaming settings will be developed. Furthermore, a more efficient weighted least-squares estimator that incorporates "kriging" error covariance information will be derived.

References

[1] N. Cressie and C. K. Wikle, "Space–time Kalman filter," *Encyclopedia of Envirometrics*, vol. 4, pp. 2045-2049, 2002.

[2] K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso, "The kriged Kalman filter," *Test*, vol. 7, pp. 217–282, Dec. 1998.

[3] N. Cressie, "The origins of kriging," *Mathematical Geology*, vol. 22, no. 3, pp. 239–252, 1990.

[4] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.

[5] S. Farahmand, D. Angelosante, and G. B. Giannakis, "Doubly robust Kalman smoothing by exploiting outlier sparsity," *Proc. of Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2010.

[6] G. B. Giannakis, G. Mateos, S. Farahmand, V. Kekatos, and H. Zhu, "USPACOR: Universal sparsity-controlling outlier rejection," *Proc. of ICASSP*, Prague, Czech Rep., May 2011.

[7] C. Masreliez, and R. Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. on Automatic Control*, vol. 22, pp. 361–371, Jun. 1977.

[8] D. B. Chua, E. D. Kolaczyk, and M. Crovella, "Network kriging," *IEEE Journal on Selec. Areas in Communications*, vol. 24, pp. 2263–2272, Dec. 2006.

[9] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering," *IEEE J. of Sel. Topics in Sig. Proc.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.

[10] K. Rajawat, E. Dall'Anese, and G. B. Giannakis, "Dynamic network delay cartography," *IEEE Trans. on Information Theory*, vol. 60, no. 5, pp. 2910–2920, May 2014.

[11] W. H. Kersting, *Distribution System Modeling and Analysis*. 2nd ed., Boca Raton, FL: CRC Press, 2007.

[12] W. Wang, Y. Xu, and M. Khanna, "A survey on the communication architectures in smart grid," *Computer Networks*, Volume 55, Issue 15, Oct. 2011.

[13] J. Chen, W. Li, A. Lau, J. Cao, and K. Wang, "Automated Load Curve Data Cleansing in Power Systems," *IEEE Transactions on Smart Grid*, vol. 1, n. 2, 2010.

[14] J. Bank and J. Hambrick, "Development of a high resolution, real time, distribution-level metering system and associated visualization modeling, and data analysis functions," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-5500-56610, May 2013.