

# Inference of Spatiotemporal Processes over Graphs via Kernel Kriged Kalman Filtering

Vassilis N. Ioannidis<sup>†</sup>, Daniel Romero<sup>\*</sup>, and Georgios B. Giannakis<sup>†</sup>

<sup>†</sup> ECE Dept. and Digital Tech. Center, Univ. of Minnesota, Mpls, MN 55455, USA

<sup>\*</sup> ICT Dept., Univ. of Agder, Grimstad 4879, Norway

E-mails: ioann006@umn.edu, daniel.romero@uia.no, georgios@umn.edu

**Abstract**—Inference of space-time signals evolving over graphs emerges naturally in a number of network science related applications. A frequently encountered challenge pertains to reconstructing such dynamic processes given their values over a subset of vertices and time instants. The present paper develops a graph-aware kernel-based kriged Kalman filtering approach that leverages the spatio-temporal dynamics to allow for efficient on-line reconstruction, while also coping with dynamically evolving network topologies. Laplacian kernels are employed to perform kriging over the graph when spatial second-order statistics are unknown, as is often the case. Numerical tests with synthetic and real data illustrate the superior reconstruction performance of the proposed approach.

**Index Terms**—Graph signal reconstruction, time series on graphs, kriged Kalman filtering, Laplacian kernels.

## I. INTRODUCTION

A number of applications involve data that can be efficiently represented as node attributes over social, economic, sensor, communication, and biological networks [1]. An inference problem that often emerges is to predict the attributes of all nodes in the network given the attributes of a subset of nodes. Such a task is of paramount importance in applications where collecting the attributes at all nodes is prohibitive, as is the case when sampling massive graphs or when the attribute of interest is of sensitive nature such as the transmission of HIV in a social network. This problem has been formulated as extrapolation or reconstruction of a function or signal on a graph [1], [2]. Extrapolation typically leverages smoothness of the attributes with respect to the graph, meaning that connected nodes have similar attribute values. Oftentimes, the aforementioned networks and attributes evolve over time. The space-time dynamics of such time-varying graph functions should be properly modeled to achieve accurate reconstruction over space and time.

Reconstruction of time-invariant graph functions has attracted great attention in recent years. The community of signal processing on graphs mainly adopts the so-called *bandlimited model*, which postulates that the signal of interest lies in a  $B$ -dimensional subspace related to the graph topology [3], [4], or assumes that the signal can be sparsely represented on

an over-complete dictionary [5]. On the other hand, the machine learning community advocates estimators that exploit the aforementioned notion of smoothness [6], [7]. Interestingly, most estimators considered by both communities can be seen as special cases of kernel-based estimators [2].

On the other hand, reconstruction of time-varying graph functions has been typically tackled by assuming that the function of interest changes slowly over time. Distributed reconstruction methods are reported in [8] and [9]. However, they rely on the bandlimited model, whose effectiveness in capturing the dynamics of real-world graph functions may not hold. A kernel-based Kalman filter that captures multiple forms of spatiotemporal dynamics through space-time kernels was explored in [10]. But it mainly relies on smoothness and does not explicitly account for the underlying dynamics. However, there are cases where the wanted function exhibits markedly different behaviors over space and time, which existing approaches cannot account for. To circumvent this limitation, a kriged graph Kalman filter is introduced in this paper.

Kriging has been traditionally employed to interpolate stationary spatial processes that take values over subsets of the Euclidean space [11, Ch. 3]. Kriging essentially performs linear minimum mean-square error (LMMSE) estimation. To accommodate time-evolving fields, [12] introduced the kriged Kalman filter (KrKF), which affords low-complexity online spatial prediction. A reduced-dimension version of the KrKF was introduced in [13] by expanding the spatio-temporal process as a linear combination of basis functions and applying the Kalman filter (KF) to the expansion coefficients.

Kriging was extended in [14] to estimate path delays over IP networks modeled by time-evolving functions defined on the edges of a graph. Building on [14], [15] exploits temporal dynamics through the KrKF for estimating network delays. However, [15] adopts a random walk model on a static graph and therefore cannot capture general spatial dynamics. All these KrKF approaches require knowledge of the spatial statistics, which are furthermore assumed fixed over time.

The main contribution of this paper is to extend the KrKF for prediction of general spatiotemporal processes that evolve over dynamic graphs whose topology may change over time. The resulting estimator is capable of promoting smoothness

The work of V. N. Ioannidis and G. B. Giannakis was supported by ARO grant W911NF-15-1-0492 and NSF grants 1343248, 1442686, and 1514056.

over time through a state-space model, and smoothness over space through kriging. A generalization of the latter based on *Laplacian kernels* is introduced to cope with uncertainty in the spatial statistics of the process. The computational complexity of the proposed algorithm is linear in the number of time samples, rendering it attractive for online and big data applications.

The rest of the paper is structured as follows. Sec. II formulates the problem and presents the proposed model. Sec. III presents the kernel KrKF (KkrKF) for graphs. The numerical experiments in Sec. IV demonstrate the benefits of the proposed method. Finally, Sec. V provides some closing remarks.

*Notation:* Scalars are denoted by lowercase, column vectors by bold lowercase, and matrices by bold uppercase letters. Superscripts  $\top$  and  $\dagger$  respectively denote transpose and pseudo-inverse;  $\mathbb{E}[\mathbf{x}]$  stands for the expectation of the random vector  $\mathbf{x}$ ;  $\mathbf{1}_N$  for the  $N \times 1$  all-one vector;  $\delta[\cdot]$  for the Kronecker delta; and  $\text{diag}\{\mathbf{x}\}$  corresponds to a diagonal matrix with the entries of  $\mathbf{x}$  on its diagonal. Finally, if  $\mathbf{A}$  is a matrix and  $\mathbf{x}$  a vector, then  $\|\mathbf{x}\|_{\mathbf{A}}^2 := \mathbf{x}^\top \mathbf{A}^{-1} \mathbf{x}$  and  $\|\mathbf{x}\|_2^2 := \mathbf{x}^\top \mathbf{x}$ .

## II. MODELING AND PROBLEM FORMULATION

A time-varying graph is a tuple  $\mathcal{G}[t] := (\mathcal{V}, \mathbf{A}[t])$   $t = 1, 2, \dots$ , where  $\mathcal{V} := \{v_1, \dots, v_N\}$  denotes the vertex set and  $\mathbf{A}[t]$  the  $N \times N$  adjacency matrix, whose  $(n, n')$ -th entry  $A_{n,n'}[t]$ , is the nonnegative edge weight connection vertices  $v_n$  and  $v_{n'}$  at time  $t$ . The graphs in this paper are undirected and have no self-loops, which respectively imply that  $\mathbf{A}[t] = \mathbf{A}^\top[t]$  and  $A_{n,n}[t] = 0$ ,  $\forall t, n$ . The Laplacian matrix is defined as  $\mathbf{L}[t] := \text{diag}\{\mathbf{A}[t]\mathbf{1}_N\} - \mathbf{A}[t]$ , and is known to be positive semidefinite [1]. A time-varying graph function (or signal) is a map  $f : \mathcal{V} \times \mathcal{T} \rightarrow \mathbb{R}$ , where  $\mathcal{T} = \{1, 2, \dots\}$  is the set of time indices. Specifically,  $f_n[t] := f(v_n, t)$  represents an attribute value at node  $n$  and time  $t$ , e.g. the closing price of the  $n$ -th stock on the  $t$ -th day. Vector  $\mathbf{f}[t] := [f_1[t], \dots, f_N[t]]^\top \in \mathbb{R}^N$  collects the function values at time  $t$ .

At time  $t$ ,  $f_n[t]$  is observed at a subset of  $S[t]$  nodes  $\mathcal{S}[t] \subset \mathcal{V}$ . The observations  $\mathbf{y}[t] \in \mathbb{R}^{S[t]}$  can be compactly arranged as

$$\mathbf{y}[t] = \mathbf{S}[t]\mathbf{f}[t] + \mathbf{e}[t], \quad t = 1, 2, \dots \quad (1)$$

where  $\mathbf{S}[t] \in \{0, 1\}^{S[t] \times N}$  selects the rows of  $\mathbf{f}[t]$  with indices in  $\mathcal{S}[t]$ , and  $\mathbf{e}[t] \in \mathbb{R}^{S[t]}$  represents the observation error. It is assumed that  $\mathbf{e}[t]$  has zero mean  $\mathbb{E}[\mathbf{e}[t]] = \mathbf{0}$ , and is uncorrelated over time and space, meaning that  $\mathbb{E}[\mathbf{e}[t]\mathbf{e}^\top[\tau]] = \sigma_e^2 \mathbf{I}_{S[t]}$ , if  $t = \tau$ , and  $\mathbf{0}_{S[t], S[\tau]}$  otherwise.

Per slot  $t$ ,  $\mathbf{f}[t]$  will be modeled as the superposition

$$\mathbf{f}[t] = \mathbf{f}_\chi[t] + \mathbf{f}_\nu[t] \quad (2)$$

where  $\{\mathbf{f}_\nu[t]\}_t$  are temporally uncorrelated and capture only spatial dependencies, while  $\{\mathbf{f}_\chi[t]\}_t$  are spatio-temporally colored obeying the state equation

$$\mathbf{f}_\chi[t] = \mathbf{P}[t]\mathbf{f}_\chi[t-1] + \boldsymbol{\eta}[t], \quad t = 1, 2, \dots \quad (3)$$

where  $\mathbf{P}[t]$  is an  $N \times N$  transition matrix, and  $\boldsymbol{\eta}[t]$  is noise with  $\mathbb{E}[\boldsymbol{\eta}[t]] = \mathbf{0}$  and  $\mathbb{E}[\boldsymbol{\eta}[t]\boldsymbol{\eta}^\top[\tau]] = \boldsymbol{\Sigma}_\eta[t]\delta[t-\tau]$ . Moreover,  $\mathbf{f}_\chi[0]$  has mean  $\mathbb{E}[\mathbf{f}_\chi[0]] = \boldsymbol{\mu}_\chi[0]$  and covariance matrix  $\mathbb{E}[(\mathbf{f}_\chi[0] - \boldsymbol{\mu}_\chi[0])(\mathbf{f}_\chi[0] - \boldsymbol{\mu}_\chi[0])^\top] = \boldsymbol{\Sigma}_\chi[0]$ , and it is assumed uncorrelated with  $\boldsymbol{\eta}[t]$  and  $\mathbf{e}[t]$ ; that is  $\mathbb{E}[\boldsymbol{\eta}[t]\mathbf{f}_\chi^\top[0]] = \mathbf{0}_{N,N}$ , and  $\mathbb{E}[\mathbf{e}[t]\mathbf{f}_\chi^\top[0]] = \mathbf{0}_{S[t],N} \forall t$ . The model in (3) is widely used and offers flexibility in tracking multiple forms of temporal dynamics [16, Ch. 3].

On the other hand,  $\mathbf{f}_\nu[t]$  is assumed zero mean, since its mean can otherwise be incorporated into  $\mathbb{E}[\mathbf{f}_\chi[t]]$ , and has covariance matrix  $\mathbb{E}[\mathbf{f}_\nu[t]\mathbf{f}_\nu^\top[\tau]] = \boldsymbol{\Sigma}_\nu[t]\delta[t-\tau]$ . Finally,  $\mathbf{e}[t]$  and  $\boldsymbol{\eta}[t]$  are uncorrelated, meaning that  $\mathbb{E}[\mathbf{e}[t]\boldsymbol{\eta}^\top[\tau]] = \mathbf{0}_{S[t],N}$ , and also uncorrelated with  $\mathbf{f}_\nu[t]$ , i.e.,  $\mathbb{E}[\mathbf{e}[t]\mathbf{f}_\nu^\top[\tau]] = \mathbf{0}_{S[t],N}$ ,  $\mathbb{E}[\boldsymbol{\eta}[t]\mathbf{f}_\nu^\top[\tau]] = \mathbf{0}_{N,N} \forall t, \tau$ .

Given the model described by (1)-(3), the goal of this paper is to reconstruct  $\mathbf{f}[t]$  online, given  $\mathbf{y}[\tau]$ ,  $\mathbf{S}[t]$ ,  $\sigma_e^2$ ,  $\boldsymbol{\Sigma}_\eta[\tau]$ ,  $\mathbf{P}[\tau]$  and  $\mathbf{A}[\tau]$  for  $\tau = 1, \dots, t$ .

**Remark 1.** In the field of geostatistics,  $\mathbf{f}_\nu[t]$  models the so-termed small-scale *spatial fluctuations*, while  $\mathbf{f}_\chi[t]$  corresponds to the so-called *trend*. The decomposition (2) is often dictated by the sampling interval: whereas  $\mathbf{f}_\chi[t]$  captures slow dynamics relative to the sampling interval, fast variations are modeled with  $\mathbf{f}_\nu[t]$ . Examples motivating (2) include network delay prediction [15], where  $\mathbf{f}_\chi[t]$  represents the queuing delay while  $\mathbf{f}_\nu[t]$  the propagation, transmission, and processing delays. Likewise, when predicting prices across different stocks,  $\mathbf{f}_\chi[t]$  captures the daily evolution of the stock market, which is correlated across stocks and time samples, while  $\mathbf{f}_\nu[t]$  describes unexpected changes, such as the daily drop of the stock market due to political statements, which are considered uncorrelated over time.

**Remark 2.** The state transition matrix  $\mathbf{P}[t]$  can be selected in accordance with the prior information available. Simplicity in estimation motivates the random walk model, where  $\mathbf{P}[t] = \alpha \mathbf{I}_N$  with  $\alpha > 0$ . On the other hand, adherence to the graph, prompts the selection  $\mathbf{P}[t] = \alpha \mathbf{A}[t]$ , in which case (3) amounts to a graph-constrained vector autoregressive model; see e.g. [17].

## III. GRAPH-AWARE KERNEL KRIGED-KF

This section presents our KkrKF approach. After establishing that kernel ridge regression (KRR) in [2], [18] generalizes clairvoyant kriging in [11], we will introduce our novel KkrKF that does not require knowledge of  $\boldsymbol{\Sigma}_\nu[t]$ .

### A. Kriged Kalman Filter

The KrKF algorithm was first introduced for prediction of processes evolving over continuous fields, as typically occurs in geostatistics [13]. In contrast, this section reviews the KrKF for processes  $\mathbf{f}[t]$  that evolve over a graph [15], where estimation is performed in two steps. In the first step, an estimate  $\hat{\mathbf{f}}_\chi[t|t]$  is obtained from the measurements  $\{\mathbf{y}[\tau]\}_{\tau=1}^t$  using the traditional Kalman filter (KF) [16, Ch. 3] with the unknown  $\mathbf{f}_\nu[t]$  lumped in the observation noise. In the second

step,  $\mathbf{f}_\nu[t]$  is estimated through the kriging predictor [11], which is given by the LMMSE estimator

$$\begin{aligned}\hat{\mathbf{f}}_\nu[t|t] &= \mathbb{E}\{\mathbf{f}_\nu[t] | \mathbf{f}_\chi[t] = \hat{\mathbf{f}}_\chi[t|t], \mathbf{y}[t]\} \\ &= \Sigma_\nu[t] \mathbf{S}^\top[t] (\mathbf{S}[t] \Sigma_\nu[t] \mathbf{S}^\top[t] + \sigma_e^2 \mathbf{I}_{S[t]})^{-1} \boldsymbol{\psi}[t]\end{aligned}\quad (4)$$

where  $\boldsymbol{\psi}[t] := \mathbf{y}[t] - \mathbf{S}[t] \hat{\mathbf{f}}_\chi[t|t]$ . Finally, combining the component estimates yields [cf. (2)]

$$\hat{\mathbf{f}}[t|t] = \hat{\mathbf{f}}_\chi[t|t] + \hat{\mathbf{f}}_\nu[t|t].\quad (5)$$

A challenge associated with KrKF is that the kriging predictor (4) requires knowledge of  $\Sigma_\nu[t]$ . The next subsection reviews graph Laplacian kernels, and proposes a generalization of the kriging predictor that does not require knowledge of the underlying spatial statistics.

### B. Kernel Kriged Kalman Filter

After recognizing that the kriging predictor is a special case of the KRR estimator, our KKrKF algorithm is presented.

Kernel ridge regression seeks an estimate of a graph function  $\mathbf{f}_\nu$  given the observations  $\boldsymbol{\psi} = \mathbf{S} \mathbf{f}_\nu + \mathbf{e}$ . The argument  $[t]$  is dropped to reflect that a single time instant will be considered. The KRR estimate of  $\mathbf{f}_\nu$  is given by [2]

$$\begin{aligned}\hat{\mathbf{f}}_\nu &= \arg \min_{\mathbf{f}_\nu} \frac{1}{S} \|\boldsymbol{\psi} - \mathbf{S} \mathbf{f}_\nu\|_2^2 + \mu \mathbf{f}_\nu^\top \mathbf{K}^{-1} \mathbf{f}_\nu \\ &= \mathbf{K} \mathbf{S}^\top (\mathbf{S} \mathbf{K} \mathbf{S}^\top + \mu \mathbf{S} \mathbf{I}_S)^{-1} \boldsymbol{\psi}\end{aligned}\quad (6)$$

where  $\mu$  is a user-selected regularization parameter and  $\mathbf{K} > \mathbf{0}$  is a kernel matrix, whose  $(n, n')$ -th entry encodes some notion of similarity between  $v_n$  and  $v_{n'}$  [2], [6], [7], [18]. Notice that the KRR estimator (6) reduces to the kriging predictor (4) if  $\mu S = \sigma_e^2$  and  $\Sigma_\nu = \mathbf{K}$ . As a result, (6) generalizes (4) in the sense that  $\mathbf{f}_\nu$  can be deterministic, so long as it belongs to a reproducing kernel Hilbert space generated by the prescribed  $\mathbf{K}$ . Rather than minimizing the LMMSE criterion, the resulting KRR can account for the underlying graph through a judicious selection of  $\mathbf{K}$ .

Laplacian kernels have been widely used [2], [7] to promote the smoothness embodied in the graph topology. For a given Laplacian matrix with eigendecomposition  $\mathbf{L} = \mathbf{U} \text{diag}\{\boldsymbol{\lambda}\} \mathbf{U}^\top$ , a Laplacian kernel is defined as [7]

$$\mathbf{K}^{-1} := r(\mathbf{L}) := \mathbf{U} \text{diag}\{r(\boldsymbol{\lambda})\} \mathbf{U}^\top\quad (7)$$

where  $r : \mathbb{R} \rightarrow \mathbb{R}$  is a monotonically increasing function. Table I summarizes common choices of  $r(\cdot)$ , which can be selected to promote a certain structure in the so-called graph Fourier transform of  $\mathbf{f}_\nu$  [1], [2], [7]. To sum up, one can obtain  $\hat{\mathbf{f}}_\nu$  through (4) after replacing  $\Sigma_\nu$  with a Laplacian kernel.

Interestingly, for a class of random  $\mathbf{f}_\nu$  there exists a function  $r(\cdot)$  such that the LMMSE and KRR estimators yield the same estimate. These graph signals are deemed to be graph stationary in [19], [20], and their covariance matrix is diagonalizable by the eigenvectors of  $\mathbf{L}$ .

**Proposition 1.** *If  $\mathbf{f}_\nu$  is a graph stationary signal on  $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ , and the Laplacian  $\mathbf{L} = \mathbf{U} \text{diag}\{\boldsymbol{\lambda}\} \mathbf{U}^\top$  has distinct*

Kernels	Function
Diffusion [6]	$r(\lambda) = \exp\{\sigma^2 \lambda / 2\}$
Laplacian regularization [1], [7]	$r(\lambda) = 1 + \sigma^2 \lambda$
Bandlimited [2]	$r(\lambda) = \begin{cases} 1/\beta, & \lambda \leq \lambda_{\max} \\ \beta, & \text{o.w.} \end{cases}$

TABLE I: Common transformation functions.

---

#### Algorithm 1: Kernel Kriged Kalman filter (KKrKF)

---

**Initialize:**  $\hat{\mathbf{f}}_\chi[0|0] = \boldsymbol{\mu}_\chi[0]$ ,  $\mathbf{M}[0|0] = \Sigma_\chi[0]$

**for**  $t = 1, \dots$  **do**

**Input:**  $\Sigma_\eta[t] \in \mathbb{S}_+^N$ ,  $\mathbf{P}[t] \in \mathbb{R}^{N \times N}$ ,  $\mathbf{y}[t] \in \mathbb{R}^{S[t]}$ ,  
 $\mathbf{S}[t] \in \{0, 1\}^{S[t] \times N}$ ,  $\sigma_e^2$ ,  $\mathbf{K}[t] \in \mathbb{S}_+^N$ .

$\Sigma_\epsilon[t] := \mathbf{S}[t] \mathbf{K}[t] \mathbf{S}^\top[t] + \sigma_e^2 \mathbf{I}_{S[t]}$

$\hat{\mathbf{f}}_\chi[t|t-1] = \mathbf{P}[t] \hat{\mathbf{f}}_\chi[t-1|t-1]$

$\mathbf{M}[t|t-1] = \mathbf{P}[t] \mathbf{M}[t-1|t-1] \mathbf{P}^\top[t] + \Sigma_\eta[t]$

$\mathbf{G}[t] = \mathbf{M}[t|t-1] \mathbf{S}^\top[t] (\Sigma_\epsilon[t] + \mathbf{S}[t] \mathbf{M}[t|t-1] \mathbf{S}^\top[t])^{-1}$

$\hat{\mathbf{f}}_\chi[t|t] = \hat{\mathbf{f}}_\chi[t|t-1] + \mathbf{G}[t] (\mathbf{y}[t] - \mathbf{S}[t] \hat{\mathbf{f}}_\chi[t|t-1])$

$\mathbf{M}[t|t] = (\mathbf{I}_N - \mathbf{G}[t] \mathbf{S}[t]) \mathbf{M}[t|t-1]$

$\hat{\mathbf{f}}_\nu[t|t] = \mathbf{K}[t] \mathbf{S}^\top[t] \Sigma_\epsilon^{-1}[t] (\mathbf{y}[t] - \mathbf{S}[t] \hat{\mathbf{f}}_\chi[t|t])$

**Output:**  $\hat{\mathbf{f}}_\chi[t|t]$ ;  $\hat{\mathbf{f}}_\nu[t|t]$ ;  $\mathbf{M}[t|t]$ .

**end for**

---

eigenvalues, then the covariance matrix  $\Sigma_\nu$  of  $\mathbf{f}_\nu$  is a Laplacian kernel.

*Proof.* Since  $\mathbf{f}_\nu$  is graph stationary,  $\Sigma_\nu$  is diagonalizable by  $\mathbf{U}$ , meaning that  $\Sigma_\nu = \mathbf{U} \text{diag}\{\boldsymbol{\sigma}\} \mathbf{U}^\top$ , where  $\boldsymbol{\sigma} := [\sigma_1 \dots \sigma_N]^\top > \mathbf{0}$  collects the eigenvalues of  $\Sigma_\nu$ . A transformation  $r^{-1}(\cdot)$  can then be selected such that  $\sigma_i = r^{-1}(\lambda_i)$ , as long as  $\{\lambda_i\}_i$  are distinct.  $\square$

Therefore, if  $\mathbf{f}_\nu[t]$  is graph stationary, it follows that

$$\Sigma_\nu[t] = \mathbf{K}[t] = \mathbf{U}[t] (\text{diag}\{r(\boldsymbol{\lambda}[t])\})^{-1} \mathbf{U}^\top[t]\quad (8)$$

for some  $r(\cdot)$ , and the estimates (4) and (6) coincide.

The proposed KKrKF is summarized as Algorithm 1. This online estimator with complexity  $\mathcal{O}(N^3)$  per  $t$ , tracks the temporal variations of the signal of interest through (3), and promotes desired properties such as smoothness over the graph, by judiciously selecting the Laplacian kernel. Different from existing approaches, our KKrKF takes into account the underlying graph structure in estimating  $\mathbf{f}_\nu[t]$  as well as  $\mathbf{f}_\chi[t]$ . Furthermore, by using the Laplacian matrix in (8), it can also accommodate dynamic graph topologies.

## IV. SIMULATIONS

This section describes tests on synthetic and real graph functions over dynamic graphs which demonstrate the superior performance of KKrKF over competing alternatives. The tests compare the following reconstruction algorithms: (i) The least mean-squares (LMS) algorithm in [9] with step size  $\mu_{\text{LMS}}$ ; (ii) the distributed least-squares reconstruction (DLSR) algorithm [8] with step sizes  $\mu_{\text{DLSR}}$  and  $\beta_{\text{DLSR}}$  (both LMS

and DLSR can track slowly time-varying  $B$ -bandlimited graph signals); (iii) The  $B$ -bandlimited instantaneous estimator (BL-IE) which uses the estimator in [3], [4] per slot  $t$ ; and (iv) Algorithm 1 with the following configuration: a diffusion kernel (cf. Table I) with parameter  $\sigma$ ; a state noise covariance  $\Sigma_\eta[t] = s_\eta \Sigma_\eta$  with parameter  $s_\eta > 0$  and  $\Sigma_\eta := \mathbf{N}\mathbf{N}^\top$  a positive definite matrix with  $\mathbf{N} \in \mathbb{R}^{N \times N}$  a random matrix with standardized Gaussian entries; and a transition matrix  $\mathbf{P}[t] = \alpha \mathbf{A}[t]$  with parameter  $\alpha$ .

The performance of the aforementioned approaches is evaluated in terms of the normalized mean-square error

$$\text{NMSE}(\{\mathcal{S}[\tau]\}_{\tau=1}^t) := \frac{\mathbb{E}[\sum_{\tau=1}^t \|\mathbf{S}^c[\tau](\mathbf{f}[\tau] - \hat{\mathbf{f}}[\tau])\|_2^2]}{\mathbb{E}[\sum_{\tau=1}^t \|\mathbf{S}^c[\tau]\mathbf{f}[\tau]\|_2^2]}$$

where the expectation is taken over the sample locations, and realizations of  $\Sigma_\eta$ , and  $\mathbf{S}^c[\tau]$  is an  $(N - S[\tau]) \times N$  matrix comprising the rows of  $\mathbf{I}_N$  whose indices are not in  $\mathcal{S}[t]$ . For all the tests, the sampling set is chosen uniformly at random without replacement over  $\mathcal{V}$  and kept constant over time; that is  $\mathcal{S}[t] = \mathcal{S}, \forall t$ .

The first real dataset contains timestamped messages exchanged over an online social network between students at the University of California, Irvine [21] for a period of 90 days corresponding to 3 months. The sampling interval  $t$  is one day. A network was created where  $\{A_{n,n'}[t]\}_{t=30(k-1)+1}^{t=30k}$  counts the number of messages exchanged between student  $n$  and  $n'$  in the  $k$ -th month. The resulting topology changes across months. A subset of  $N = 310$  users for which  $\mathbf{A}[t]$  corresponds to a connected graph  $\forall t$  was selected. At each time  $t$ ,  $\mathbf{f}[t]$  was generated by adding a temporally uncorrelated  $B$ -bandlimited component with  $B = 5$  and a spatio-temporally correlated component. Specifically,  $\mathbf{f}[t] = \sum_{i=1}^5 \gamma_i[t] \mathbf{u}_i[t] + \mathbf{f}_\chi[t]$ , where  $\mathbf{f}_\chi[t]$  follows (3),  $\{\gamma_i[t]\}_{i=1}^5$  are standardized Gaussian distributed for all  $t$ , and  $\{\mathbf{u}_i[t]\}_{i=1}^5$  are the eigenvectors associated with the 5 smallest eigenvalues of the Laplacian matrix at time  $t$ . Function  $f_n[t]$  is therefore smooth with respect to the graph and can be interpreted e.g. as the time that the  $n$ -th student spends on the specific social network during the  $t$ -th day.

The first experiment justifies the proposed decomposition by assessing the impact of dropping each term on the right hand side of (2). Fig. 1 depicts the NMSE over the time index for KKrKF; the Kalman filter (KF) estimator, which results from setting  $\hat{\mathbf{f}}_\nu[t|t] = \mathbf{0}$  for all  $t$  in the KKrKF and therefore does not exploit spatial information, as well as kernel Kriging (KKr), which the KKrKF reduces to if  $\hat{\mathbf{f}}_\chi[t|t] = \mathbf{0}$  for all  $t$  and therefore does not exploit temporal information. As observed, the novel algorithm, which accounts for both terms, is capable of efficiently capturing the spatial as well as the temporal dynamics over time-evolving topologies.

The second dataset is provided by the National Climatic Data Center [22], and comprises hourly temperature measurements at  $N = 109$  measuring stations across the continental United States in 2010. A time-invariant graph was constructed as in [10], based on geographical distances. The value  $f_n[t]$  represents the temperature recorded at the  $n$ -th station and

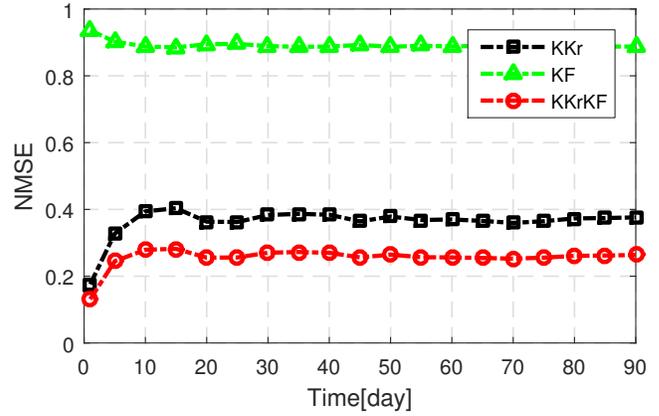


Fig. 1: NMSE of function estimates. ( $\mu_\chi[0] = \mathbf{0}$ ,  $\Sigma_\chi[0] = \mathbf{0}$ ,  $\sigma_e^2 = 10^{-4}$ ,  $\sigma = 1.5$ ,  $\alpha = 0.028$ ,  $s_\eta = 10^{-4}$ )

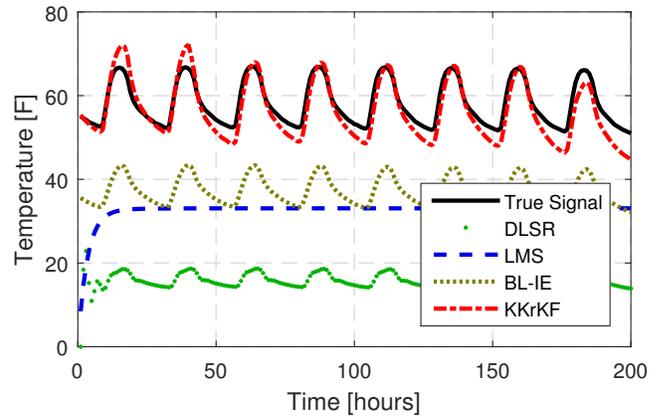


Fig. 2: True temperature values along with the estimated ones. ( $\mu_\chi[0] = \mathbf{0}$ ,  $\Sigma_\chi[0] = \mathbf{0}$ ,  $\sigma = 1.8$ ,  $B = 5$ ,  $\mu_{\text{DLSR}} = 1.2$ ,  $\beta_{\text{DLSR}} = 0.5$ ,  $\mu_{\text{LMS}} = 0.6$ ,  $\alpha = 10^{-3}$ ,  $s_\eta = 10^{-5}$ )

$t$ -th sample. The sampling interval is one hour for the first experiment and one day for the second.

Next, the performance of the different reconstruction algorithms is evaluated in tracking the temperature values. Fig. 2 depicts the true temperature value along with the estimates of the different algorithms for a station  $n$  that is not sampled, i.e.  $n \notin \mathcal{S}$ , with  $S = 40$ . Clearly, the novel algorithm accurately tracks the temperature by exploiting spatial and temporal information. On the other hand, DLSR and LMS cannot capture the fast signal variations. Finally, Fig. 3 compares the NMSE of all considered approaches, and showcases the superior NMSE performance of Algorithm 1 for  $S = 40$ . As observed, KKrKF captures the spatio-temporal dynamics and outperforms existing alternatives.

## V. CONCLUSIONS

This paper introduced an online estimator to reconstruct dynamic processes over dynamic graphs. In this context, the function to be estimated was decomposed in two parts: one capturing the spatial dynamics while being uncorrelated over time, and the other modeling jointly spatiotemporal dynamics.

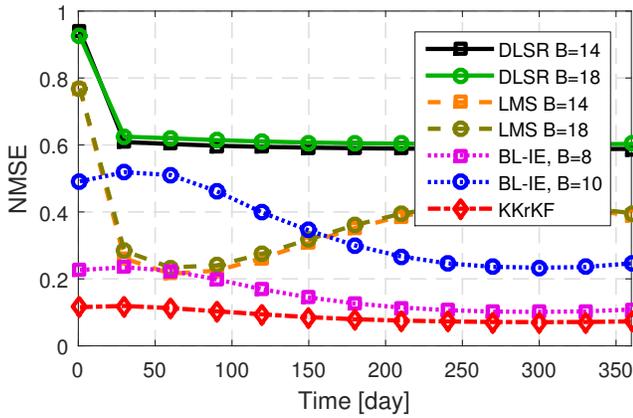


Fig. 3: NMSE of temperature estimates. ( $\mu_x[0] = \mathbf{0}$ ,  $\Sigma_x[0] = \mathbf{0}$ ,  $\sigma = 1.8$ ,  $\mu_{\text{DLSR}} = 1.6$ ,  $\beta_{\text{DLSR}} = 0.5$ ,  $\mu_{\text{LMS}} = 0.6$ ,  $\alpha = 10^{-3}$ ,  $s_\eta = 10^{-5}$ )

A novel kernel kriged Kalman filtering approach was developed that leverages Laplacian kernels for reconstructing the spatial component. The algorithm was evaluated on synthetic as well as real-data scenarios, and performed markedly better than existing alternatives. Future work includes distributed implementations, multi-kernel approaches for optimal selection of  $r(\cdot)$ , and data-driven learning of  $\mathbf{P}[t]$ .

#### REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Sig. Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [2] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Sig. Process.*, vol. 65, no. 3, pp. 764–778, Feb. 2017.
- [3] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Sig. Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.
- [4] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Global Conf. Sig. Inf. Process.*, Austin, Texas, 2013, pp. 491–494.
- [5] D. Thanou, D. I. Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Sig. Process.*, vol. 62, no. 15, pp. 3849–3862, Aug. 2014.
- [6] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proc. Int. Conf. Mach. Learn.*, Sydney, Australia, Jul. 2002, pp. 315–322.
- [7] A. J. Smola and R. I. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*, pp. 144–158. Springer, 2003.
- [8] X. Wang, M. Wang, and Y. Gu, "A distributed tracking algorithm for reconstruction of graph signals," *IEEE J. Sel. Topics Sig. Process.*, vol. 9, no. 4, pp. 728–740, Feb. 2015.
- [9] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean-square estimation of graph signals," *IEEE Trans. Sig. Info. Process. Netw.* (early access), 2016.
- [10] D. Romero, V. N. Ioannidis, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions on dynamic graphs," *IEEE J. Sel. Topics Sig. Process.* (to appear), Sep. 2017.
- [11] N. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.
- [12] K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso, "The kriged Kalman filter," *Test*, vol. 7, no. 2, pp. 217–282, Dec. 1998.
- [13] C. K. Wikle and N. Cressie, "A dimension-reduced approach to space-time Kalman filtering," *Biometrika*, pp. 815–829, Dec. 1999.
- [14] D. B. Chua, E. D. Kolaczyk, and M. Crovella, "Network kriging," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 12, pp. 2263–2272, Nov. 2006.
- [15] K. Rajawat, E. Dall'Anese, and G. B. Giannakis, "Dynamic network delay cartography," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2910–2920, May 2014.
- [16] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, vol. 2, Wiley New York, 1958.
- [17] Y. Shen, B. Baingana, and G. B. Giannakis, "Nonlinear structural vector autoregressive models for inferring effective brain network connectivity," *arXiv preprint arXiv:1610.06551v1*, 2016.
- [18] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.
- [19] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *arXiv preprint arXiv:1603.04667*, 2016.
- [20] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *CoRR*, vol. abs/1601.02522, 2016.
- [21] "Snap temporal networks: Collegemsg." [Online]. Available: <http://snap.stanford.edu/data/CollegeMsg.html>.
- [22] "1981-2010 U.S. climate normals." [Online]. Available: <https://www.ncdc.noaa.gov/data-access>.