JOURNAL OF AEROSPACE INFORMATION SYSTEMS
Vol. 19, No. 6, June 2022

Real-Time Communication Control in Decentralized Autonomous Sensor Networks

Hongchuan Wei*
Duke University, Durham, North Carolina 27708
and
Keith A. LeGrand,† Andre A. Paradise,‡ and Silvia Ferrari§
Cornell University, Ithaca, New York 14850

https://doi.org/10.2514/1.I010934

Minimizing the amount of communication required by a sensor network is crucial to minimizing both energy and time consumption, as well to operating covertly and robustly in communication-contested environments. This paper presents a novel intermittent communication control approach applicable to sensor networks deployed to sense and model spatio-temporal processes by nonparametric models such as Gaussian processes (GPs). The approach relies on a novel and efficient approximation of the GP average generalization error (AGE), as well as on novel GP sensor control and regression methods presented in this paper. This novel AGE approximation allows each sensor to characterize the nominal prediction performance of the learned GP model in the absence of communications. As a result, individual sensors can update the GP hyperparameters based solely on local measurements and decide to communicate only if and when their estimate of the nominal prediction performance falls below an acceptable threshold.

I. Introduction

SIGNIFICANT progress has been made to date on the development of planning and control methods for coordinating the motion and measurements of collaborative sensor networks so as to maintain communications [1–3] and maximize information value [4–9]. These methods typically assume that the network connectivity is to be maintained by ensuring that sensors (nodes) are within a desired communication range, thus enabling pairwise or nearest-neighbor communication links. Under this assumption, sensors are able to exchange local measurements and other relevant information, such as control policies and environmental conditions, necessary for collaborative sensing tasks. In many sensing applications, however, constantly maintaining communication links over time is undesirable even when communication constraints are satisfied by the sensor location [10,11]. This is because establishing communication links may require significant resources due to contested communication environments or the need for covertness and, thus, may pose undue burden on available bandwidth and energy supply [12–18]. Example applications include ecological and environmental monitoring [19], underwater sensor networks [20], and physical security [21]. Energy conservation is crucial to extending the network lifetime in these applications, as well as in many other battery-powered wireless sensor networks [22]. This paper develops an approach for automatically controlling communication broadcasting based on the desired level of network sensing performance in spatio-temporal learning problems. While this paper considers the control of broadcast communications, the approach can also be extended to other communication protocols and network topologies, such as those discussed in [22].

Existing communication control approaches can be divided into time-driven and event-driven algorithms. In time-driven algorithms, sensors communicate periodically at a fixed time interval determined from sensing parameters. Because these parameters must be obtained from all sensing nodes, however, the policy may become obsolete in the absence of communication links [23–29]. In particular, the communication time interval is determined by the maximum allowable transfer interval between two subsequent message transmissions to ensure closed-loop stability. Event-driven algorithms trigger communications based on a predefined event that can be monitored at the sensor level [30]. In the "send-on-delta" algorithm proposed in [31], for instance, sensors communicate when the change in the observed signal exceeds a predefined threshold. The "deadband" algorithm developed in [32] determines the communication time for a network of proportional-integral-derivative (PID) controllers based on system state error bounds that can be adjusted based on system response and network traffic. Other event-based communication strategies have been developed for formation maintenance [33], and decentralized control of linear time-invariant systems [34–36].

Most of these existing communication control approaches have been developed for networks of linear or nonlinear dynamic systems characterized by a fixed number of parameters. To date, few results exist on communication control for networks engaged in sensing for learning nonparametric models, such as Gaussian processes (GPs). The advantage of nonparametric models is that they allow the sensor network to learn the dynamics of a stochastic process from data, adjusting the model dimensionality to the sensor measurements received incrementally over time [37–41]. By this approach, the dimensionality of the model can be increased or reduced based on the information underlying new sensor measurements, as reviewed in Sec. III.

This paper develops a novel approach for estimating the network's GP average generalization error (AGE) locally, in the absence of communications and ground truth. Due to the duality of the sensor inference and control problems, the predicted GP generalization error depends on the future actions of the sensors, which are unknown in the absence of communications. By deriving an approximation to the network's generalization error that assumes uniform sampling, this paper develops an automatic control method that is able to determine communication times and prevent the sensing performance from falling below an acceptable level. The generalization error provides a nominal performance baseline for the entire sensor network that can be monitored and updated based on local measurements when communication links are absent.

Received 11 November 2020; revision received 22 December 2021; accepted for publication 27 December 2021; published online Open Access 10 May 2022.

*Ph.D. Student, Mechanical Engineering and Materials Science.
†Ph.D. Student, Sibley School of Mechanical and Aerospace Engineering. Member AIAA.
‡Ph.D. Student, Sibley School of Mechanical and Aerospace Engineering.
§John Brancaccio Professor, Sibley School of Mechanical and Aerospace Engineering. Fellow AIAA.

The novel and efficient approximation of the network generalization error is presented in Sec. IV and validated numerically in Sec. VII. The computational complexity of this approximation is shown to be linear in the number of collocation points and polynomial in the number of admissible sensor positions. As a result, the approximate GP generalization error can be computed recursively and monitored over time to control communications, while optimizing the sensor network placement. The effectiveness of the communication control method developed in Sec. VI is demonstrated on a temperature modeling problem for the Conterminous United States.

The numerical simulation results presented in Sec. VII show that, by the proposed control method, climate sensors require few intermittent communication events in order to maintain the GP error above the desired threshold. Moreover, even in the presence of rare and intermittent communications, the sensor network is able to significantly improve the GP model generalization performance by means of sensor control and GP hyperparameter optimization algorithms leveraging fusion only at the chosen communication times, when all past measurements are broadcasted to the sensors.

II. Problem Formulation

Consider a network of N sensors deployed in a region of interest (ROI) in order to learn the model of a spatio-temporal process that is time invariant and possibly nonlinear. The sensors may be installed on fixed stations or unmanned vehicles in order to obtain multiple observations distributed over both space and time. Examples of relevant applications are modeling of environmental characteristics such as atmospheric temperature, gas concentration, or oceanic features [42,43], or modeling of traffic patterns and pedestrians [39,44]. Let the underlying process be represented by an unknown function $f: \mathbb{R}^n \to \mathbb{R}$ that is possibly nonlinear and, at any time t, maps any spatial inertial coordinate vector $\boldsymbol{x}$ into a hidden variable y,

$$y(k) = f[\boldsymbol{x}(t)], \qquad \boldsymbol{x} \in \mathcal{X}, \quad y \in \mathbb{R} \tag{1}$$

where the domain $\mathcal{X} \subset \mathbb{R}^n$ represents the ROI of dimensions $n \in \mathbb{Z}^+$. The function f is assumed to be of class C^1 (i.e., differentiable and with continuous derivatives) and to be bounded in $\mathcal{X}$. The domain $\mathcal{X}$ is assumed to be a connected, compact, and nonempty subset of a Euclidean space that may or may not be convex [45].

Typically the variable of interest, y, is hidden (or nonobservable) everywhere in $\mathcal{X}$. Thus, the process (1) is to be modeled based on point measurements obtained by the sensor network over time. Each sensor in the network, indexed by i, can decide where to move over time based on a set of *admissible* sensor positions, denoted by $\mathcal{S}_i$, assumed discrete, finite, and known a priori. Sensor measurements obtained at discrete moments in time, indexed by k, obey the measurement equation

$$z_i(k) = f[\boldsymbol{x}_i(k)] + \nu(k), \qquad \boldsymbol{x}_i(k) \in \mathcal{S}_i, \quad i = 1, \ldots, N \tag{2}$$

where the measurement noise ν is zero-mean, white, and Gaussian, and is characterized by a known standard deviation σ. Every point measurement (2) is obtained by a sensor positioned at $\boldsymbol{x}_i \in \mathcal{S}_i$, chosen by the sensor control algorithm from a set of positions dictated by the range of the mobile platform, or other motion constraints, such as feasible orbital or aerial trajectories. As an example, this paper considers a sensor network used to model the average monthly mean temperature over the Conterminous United States (Fig. 1), where at every time k each sensor can take a measurement at one of its admissible locations and decide whether to communicate its measurements to the rest of the network.

In communication-contested environments, establishing communication links is problematic due, for example, to their energy cost or the need for covertness. On the other hand, communication links are required for learning the spatial process (1) everywhere in $\mathcal{X}$, because of the need for fusing the decentralized sensor measurements, $z_1, \ldots, z_N$, obtained by the network over time. Let $u_i(k) \in \{0, 1\}$ denote the ith sensor's decision to communicate at time k,

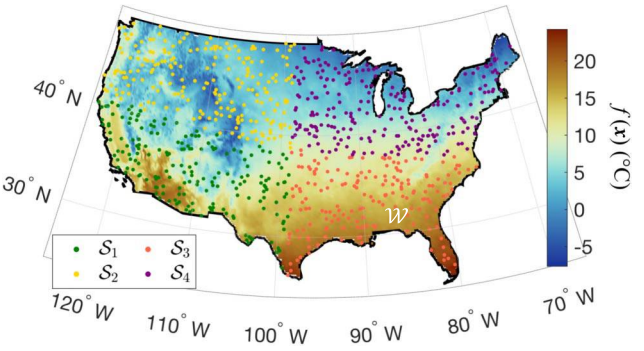

Fig. 1 Average monthly mean temperature in March 2016 for the Conterminous United States,[¶] sets of measurement locations available to $N = 4$ sensors (colored dots).

where $u_i = 1$ represents the decision to communicate, and $u_i = 0$ represents the decision to remain silent. This paper considers a network broadcasting problem in which, once any one sensor in the network decides it must communicate, a connected communication graph, $G = (V, E)$, is established over the set of vertices V, representing the set of N sensors, where the set of edges E represents the communication channels.

In the absence of communications, each sensor only has access to its local information composed of all position and temperature measurements obtained up to the present time. The *local data set* of sensor i at time k is denoted by $Q_i(k) \triangleq \{\boldsymbol{x}_i(\ell), z_i(\ell) | \ell = 1, \ldots, k\}$. Because the ground truth is always unknown, in the absence of communications every sensor i performs local learning of the spatial process (1) based only on $Q_i(k)$, and plans its future movements, $\boldsymbol{x}_i(k+1), \boldsymbol{x}_i(k+2), \ldots$, so as to optimize its local information gain. Although other variables, such as local sensor travel distance and associated cost, may also be considered for planning [4,46–49], they are not considered in this paper for simplicity.

Once the local sensing performance falls below an acceptable level, however, network communications must be established so as not waste valuable time and resources. Because each sensor is subject to different environmental conditions (e.g., noise), the goal is to establish communications as soon as any of the sensors in the network exhibits unsatisfactory performance. The decision to communicate at time k by any of the sensors in the network can be represented by the following clause:

$$\bigvee_{i=1}^{N} u_i(k) = 1 \tag{3}$$

where $\vee$ denotes the logical disjunction operator, and the positions and measurements of all N sensors, represented by the *global data set* $Q(k) \triangleq \bigcup_{i=1}^{N} Q_i(k)$, are shared via the communication graph G.

The intermittent *communication control problem* considered in this paper is to develop a decentralized policy by which each sensor i plans the sensor positions, $\boldsymbol{x}_i(k)$, and communication decisions, $u_i(k)$, to learn the spatial process (1) incrementally over time with minimal information exchange with the network.

III. Gaussian Process Model

GPs are an effective approach for learning spatio-temporal processes from noisy measurements, because they are capable of learning stochastic nonlinear processes from data without specifying the dimensionality of the model a priori [44,50–52]. The advantage over other data-driven approaches is that the dimensionality of the GP model can be adapted incrementally to data that become available over time by adjusting the GP hyperparameters and dimensionality to the underlying complexity while avoiding overfitting.

[¶]PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created March 2016, archived at https://perma.cc/KWS3-84AR.

A GP defines a multivariate Gaussian distribution $P(\cdot)$, over a function $f: \mathcal{X} \to \mathbb{R}$, that can be used to model a spatial or temporal process, as shown in Eq. (1). Let $F = \{y_i | y_i = f(\boldsymbol{x}_i), i = 1, \ldots, \eta\}$ be a set of η function evaluations in $\mathcal{X}$. Then, the multivariate distribution $P(f)$ is a GP if the marginal distribution $P(F)$ is a joint multivariate Gaussian distribution for any finite set of points in the function domain, $\{\boldsymbol{x}_i | \boldsymbol{x}_i \in \mathcal{X}, i = 1, \ldots, \eta\}$ [53]. The existence of a GP can be proven using Kolmogorov's consistency theorem, as shown in [54,55].

A GP is completely specified by its mean function

$$\theta(\boldsymbol{x}) \triangleq \mathbb{E}_y[f(\boldsymbol{x})], \quad \forall\, \boldsymbol{x} \in \mathcal{X} \tag{4}$$

and its covariance function

$$\phi(\boldsymbol{x}, \boldsymbol{x}') \triangleq \mathbb{E}_{y,y'}\{[f(\boldsymbol{x}) - \theta(\boldsymbol{x})][f(\boldsymbol{x}') - \theta(\boldsymbol{x}')]\} \tag{5}$$

for $\forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, where $\mathbb{E}_y[\cdot]$ denotes the expectation operator with respect to the random variable y [37,39,56]. Then, the notation

$$f(\boldsymbol{x}) \sim \mathrm{GP}(\theta(\boldsymbol{x}), \phi(\boldsymbol{x}, \boldsymbol{x}')), \quad \forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X} \tag{6}$$

is used to indicate that f is "distributed as" the GP with mean function $\theta(\cdot)$ and covariance function $\phi(\cdot, \cdot)$.

A. GP Regression

A major advantage of GP models is the existence of simple analytic formulas for updating the mean and covariance of the posterior distribution [57]. Let the set of all available data be denoted by

$$Q = \{(\boldsymbol{x}_i, z_i) | \boldsymbol{x}_i \in \mathcal{X}, z_i = f(\boldsymbol{x}_i) + \nu_i, i = 1, \ldots, \eta\}$$

where η is the number of data pairs collected up to the present time and all other variables are defined in Sec. II. Under the assumption of additive Gaussian noise (ν_i, $\forall\, i$), the posterior distribution of the spatial process (1) conditioned on Q is also a GP with its own mean and covariance functions.

To obtain the mean and covariance of the posterior for the hidden variable given the data Q, the data pairs are organized into a $(\eta n \times 1)$ vector $\boldsymbol{X} \triangleq [\boldsymbol{x}_1^T \ \cdots \ \boldsymbol{x}_\eta^T]^T$, and a $(\eta \times 1)$ vector $\boldsymbol{z} \triangleq [z_1 \ \cdots \ z_\eta]^T$. Then, the *unknown* function evaluations corresponding to the η data pairs can be denoted by a $\eta \times 1$ vector $\boldsymbol{y} \triangleq [y_1 \ \cdots \ y_\eta]^T$ comprising unknown random variables. Consider now any other set of m positions and measurements, say $\boldsymbol{X}'$ and $\boldsymbol{z}'$, to which there corresponds a vector of hidden variables $\boldsymbol{y}'$. Then, the $(\eta \times m)$ cross-covariance matrix of the two random vectors $\boldsymbol{y}$ and $\boldsymbol{y}'$ is given by

$$\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{X}') \triangleq \mathbb{E}_{y,y'}[(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])(\boldsymbol{y}' - \mathbb{E}[\boldsymbol{y}'])^T] = \begin{bmatrix} \phi(\boldsymbol{x}_1, \boldsymbol{x}_1') & \cdots & \phi(\boldsymbol{x}_1, \boldsymbol{x}_m') \\ \vdots & \ddots & \vdots \\ \phi(\boldsymbol{x}_\eta, \boldsymbol{x}_1') & \cdots & \phi(\boldsymbol{x}_\eta, \boldsymbol{x}_m') \end{bmatrix} \tag{7}$$

where $\phi(\cdot, \cdot)$ is defined in Eq. (5). Without loss of generality, the GP mean function is assumed zero; i.e., $\theta(\boldsymbol{x}) \equiv 0$.

Then, from GP regression [58], the joint distribution of $\boldsymbol{z}$ and $\boldsymbol{y}'$ is a multivariate Gaussian distribution, such that

$$\begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{y}' \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}_{(m+\eta)}, \begin{bmatrix} \Sigma & \boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{X}') \\ \boldsymbol{\Phi}(\boldsymbol{X}', \boldsymbol{X}) & \boldsymbol{\Phi}(\boldsymbol{X}', \boldsymbol{X}') \end{bmatrix}\right) \tag{8}$$

where

$$\boldsymbol{\Sigma} \triangleq \boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}_\eta \tag{9}$$

and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$ denotes a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{K}$. The notation $\boldsymbol{0}_n$ represents an $(n \times 1)$ vector of zeros, and $\boldsymbol{I}_n$ is an $(n \times n)$ identity matrix. Marginalizing Eq. (8) over $\boldsymbol{z}$ shows that the posterior distribution of $\boldsymbol{y}'$ given the data Q is also a multivariate Gaussian distribution with mean

$$\boldsymbol{\mu}' = \boldsymbol{\Phi}(\boldsymbol{X}', \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{z} \tag{10}$$

and covariance

$$\boldsymbol{\Sigma}' = \boldsymbol{\Phi}(\boldsymbol{X}', \boldsymbol{X}') - \boldsymbol{\Phi}(\boldsymbol{X}', \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{X}') \tag{11}$$

B. GP Hyperparameter Learning

In GP regression, a function $f(\boldsymbol{x})$ with desired properties, such as smoothness and periodicity, can be learned from data by a proper choice of covariance function [59]. For example, if $f(\boldsymbol{x})$ is stationary (i.e., the joint probability distribution of $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ does not change when $\boldsymbol{x}$ and $\boldsymbol{x}'$ are translated simultaneously) stationary covariance functions such as the squared exponential covariance function should be used. As another example, if $f(\boldsymbol{x})$ is periodic, then the periodic covariance function can be used to model the periodicity [60]. Each of these families of covariance functions typically is specified by a finite number of *hyperparameters* whose values need to be determined before GP regression [53]. The prefix "hyper" is used to distinguish these parameters from the GP regression model parameters, which are learned from data. These hyperparameters, denoted by Θ, provide added flexibility to the choice of covariance function families. A common choice of stationary covariance function is the *squared exponential*

$$\phi(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{x}')\right] \tag{12}$$

which provides smoothness characteristics specified by a diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}([\lambda_1 \ \cdots \ \lambda_n])$. In Eq. (12), the term σ_f^2 is the function output variance that determines the average distance between $f(\cdot)$ and its mean $\theta(\cdot)$, and λ_i is a length scale in the ith dimension of $\boldsymbol{x} \in \mathbb{R}^n$. The measurement noise variance σ^2, defined in Eq. (2), is also treated as a hyperparameter [53]. Then, the complete set of GP hyperparameter to be learned from data is $\Theta = \{\sigma_f, \boldsymbol{\Lambda}, \sigma\}$.

In the Bayesian inference framework, the GP model and its hyperparameters are viewed as the prior distribution of the underlying stochastic process. In order for the GP model to best match the underlying process, the optimal hyperparameters can be learned by maximizing the logarithm of the *marginal likelihood* function [61],

$$\mathcal{L} = \log p(\boldsymbol{z} | \boldsymbol{X}, \Theta) = -\frac{1}{2}\boldsymbol{z}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{z} - \frac{1}{2}\log[(2\pi)^\eta |\boldsymbol{\Sigma}|] \tag{13}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the noisy measurements $\boldsymbol{z}$ defined in Eq. (9), and $|\cdot|$ denotes the determinant of a matrix.

The maximization of the marginal likelihood function can be solved by a gradient-based algorithm, such as the conjugate gradient method [62]. In this paper, the gradient search directions are defined by the partial derivatives of the likelihood function (13) with respect to the hyperparameters Θ. For the squared-exponential covariance function (12), these partial derivatives are found by matrix calculus, as follows:

$$\frac{\partial \mathcal{L}}{\partial \sigma_f} = \frac{1}{\sigma_f}\mathrm{tr}[(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \boldsymbol{\Sigma}^{-1})\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{X})] \tag{14}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \frac{1}{\sigma}\mathrm{tr}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \boldsymbol{\Sigma}^{-1}) \tag{15}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_\ell} = \frac{1}{4\lambda_\ell^2}\mathrm{tr}[(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \boldsymbol{\Sigma}^{-1})(\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{X}) \circ \boldsymbol{D})] \tag{16}$$

where $\ell = 1, \ldots, n$, $(\circ)$ denotes the Hadamard (or elementwise) product between two matrices, $\mathrm{tr}(\cdot)$ is the trace of a matrix, and

$$\boldsymbol{\alpha} \triangleq \boldsymbol{\Sigma}^{-1}\boldsymbol{z} \tag{17}$$

$$\boldsymbol{D}_{(i,j)} \triangleq \left(\boldsymbol{e}_\ell^T \boldsymbol{x}_i - \boldsymbol{e}_\ell^T \boldsymbol{x}_j\right)^2 \tag{18}$$

$$\boldsymbol{e}_{\ell} \triangleq \left[\underbrace{0 \quad \cdots \quad 0}_{\ell-1} 1 \underbrace{0 \quad \cdots \quad 0}_{n-\ell}\right]^T \tag{19}$$

The inverse matrix $\boldsymbol{\Sigma}^{-1}$ is first evaluated based on prior available data and can be computed with a computational complexity of $\mathcal{O}(\eta^3)$. Once Σ^{-1} is obtained, the computational complexity associated with evaluating the marginal likelihood derivatives is $\mathcal{O}(\eta^2)$ for Eqs. (14–16), when $\ell = 1$, and it is $\mathcal{O}(1)$ for Eq. (16) when $\ell \neq 1$. Therefore, the GP hyperparameters can be learned efficiently, which will in turn improve the prediction performance, as demonstrated by the simulation results in Sec. VII.

The next section shows how the GP-regression AGEs can be used to estimate the network performance in the sensing problem formulated in Sec. II, and derives an efficient generalization error approximation for online implementations. Subsequently, the results presented in the next section are used in Secs. V and VI to develop motion and communication control policies, respectively.

IV. Nominal Sensor Network Performance

This paper presents a novel approach for controlling network communications based on the expected nominal sensing network performance. By definition, this is a challenging problem because when a sensor is not communicating with the rest of the network, it is unaware of the performance and policies of other sensors. GP generalization errors have been previously used in the literature to represent deviations between the model predictions and the ground truth for "unseen" data sets. Therefore, while they may be used to characterize the *generalized* performance of a model in some stochastic settings [63–65], existing generalization errors are not applicable to problems in which both the ground truth and the other sensors' policies are unknown.

This section derives an approximation of the GP AGE that can be used by individual sensors to estimate the effectiveness of recent measurements and decide whether communications are required to prevent the network sensing performance from falling below an acceptable threshold. In the absence of communications, a sensor i has no knowledge of the measurements and policies of other sensors in the network and, therefore, cannot evaluate standard generalization error or other existing performance metrics such as information gain [4]. The novel GP generalization error presented here, on the other hand, can be used to provide an estimate of *nominal* prediction performance for the sensor network described in Sec. II. The estimate can be obtained in the absence of both communications and ground truth, based solely on the local dataset $Q_i(k)$, as shown in Sec. IV.A.

By this approach, the nominal network performance can be estimated by each sensor locally at any time, and used to determine whether the sensor network performance has fallen below a minimum acceptable threshold, as shown in Sec. VI. Because this measure of nominal sensor network performance must be evaluated repeatedly over time, a computationally efficient approximation of the GP generalization error is derived in Sec. IV.B, thereby significantly reducing the computational complexity by means of a set of time-invariant collocation points adjoined to the set of admissible physical sensor positions described in Sec. II.

A. GP Generalization Error

In regression problems, the prediction performance is typically measured by the squared difference between the true value and the estimated value of the variable(s) of interest, i.e., the dependent variable y in Eq. (1) for some $\boldsymbol{x} \in \mathcal{X}$ [66]. Such a performance measure, however, is not applicable when y is latent or hidden, because its true value is never known to the sensor. Consider a GP model learned from observations of the spatial process (1), using the GP regression approach in Sec. III, under the assumptions described in Sec. II. Then, the estimated value of y at time k can be obtained from the mean of the posterior distribution function, conditioned on the available data. Therefore, although the true value of y is never available, knowledge of the network measurements can be used to improve prediction performance provided communications are available (Sec. VII).

In the absence of communications, the data available to a sensor i at time k are limited to the local data set $Q_i(k)$. Let all sensor measurements $z_i(\ell) \in Q_i(k)$, $\ell = 1, \ldots, k$, be grouped in a $(k \times 1)$ measurement vector

$$\boldsymbol{z}_i(k) \triangleq [z_i(1) \ \cdots \ z_i(k)]^T, \qquad i = 1, \ldots, N \tag{20}$$

and all prior sensor positions $\boldsymbol{x}_i(\ell) \in Q_i(k)$, $\ell = 1, \ldots, k$, be grouped into an $(nk \times 1)$ position vector

$$\boldsymbol{X}_i(k) \triangleq [\boldsymbol{x}_i^T(1) \ \cdots \ \boldsymbol{x}_i^T(k)]^T, \qquad i = 1, \ldots, N \tag{21}$$

When communications become available and G is a connected graph, all N measurement vectors can be organized into an $(Nk \times 1)$ vector containing all of the sensor network measurements,

$$\boldsymbol{z}(k) \triangleq [\, \boldsymbol{z}_1^T(k) \quad \cdots \quad \boldsymbol{z}_N^T(k) \,]^T \tag{22}$$

and all N position vectors can be organized into an $(Nkn \times 1)$ vector containing all of the sensor network positions:

$$\boldsymbol{X}(k) \triangleq [\, \boldsymbol{X}_1^T(k) \quad \cdots \quad \boldsymbol{X}_N^T(k) \,]^T \tag{23}$$

The above measurement and position vectors are used for GP learning and will also be used later to evaluate the prediction performance of GP learning in the absence of communications. To best reflect the performance of GP learning, one needs to pick a set of reference spatial points that is preferably invariant with respect to sensor locations and time. For large state spaces, such as the Conterminous United States in Fig. 2, a tractable estimate of the GP mean and covariance is obtained by considering a finite set of *collocation points* in the process domain $\mathcal{X}$ [39,41]. In this paper, the collocation set is chosen from a uniform grid over $\mathcal{X}$, and is denoted by $\Omega = \{\boldsymbol{\xi}_\iota | \boldsymbol{\xi}_\iota \in \mathcal{X}, \iota = 1, \ldots, M\}$. Then, the spatial process (1) is discretized by evaluating $f(\cdot)$ at every collocation point in Ω obtaining a set of function evaluations denoted by $F_c = \{y_\iota | y_\iota = f(\boldsymbol{\xi}_\iota), \boldsymbol{\xi}_\iota \in \mathcal{X}, \iota = 1, \ldots, M\}$.

Substituting Eqs. (22) and (23) into Eqs. (10) and (11), where now $\eta = kN$, the posterior GP mean and covariance functions conditioned on $Q(k)$ can be obtained as follows:

$$\theta_k(\boldsymbol{\xi}_\iota) = \boldsymbol{\Phi}(\boldsymbol{\xi}_\iota, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{z} \tag{24}$$

$$\phi_k(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j) = \boldsymbol{\Phi}(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j) - \boldsymbol{\Phi}(\boldsymbol{\xi}_\iota, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{\xi}_j) \tag{25}$$

for $\iota, j = 1, \ldots, M$, where $\boldsymbol{\Sigma}$ is defined in Eq. (9), and where the index k has been omitted from the covariance matrices and sensor network position matrix for brevity. Then, the GP prediction error at the collocation points is

$$\varepsilon[\boldsymbol{\xi}_\iota, Q(k)] \triangleq [y_\iota - \theta_k(\boldsymbol{\xi}_\iota)]^2 \tag{26}$$

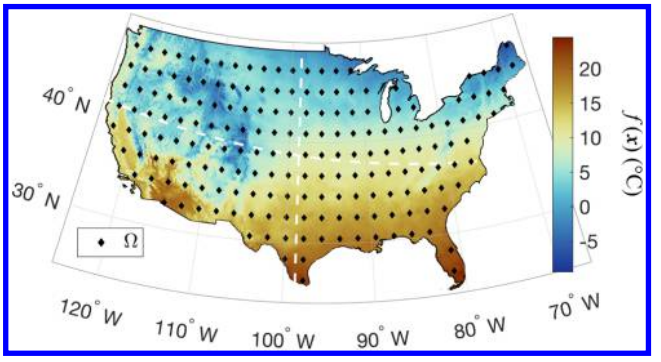

Fig. 2 Collocation points (Ω) used for GP performance evaluation in the temperature modeling problem for the Conterminous United States.

for all $\boldsymbol{\xi}_\iota \in \Omega$ and $y_\iota \in F_c$. Although Eq. (26) can provide a measure of the prediction performance of GP regressions, it cannot be evaluated in the absence of communications, or in the absence of ground truth on the hidden variable value y_ι, which is always unknown to the network (Sec. II).

Therefore, this paper develops a novel approach that adopts the Bayesian generalization error presented in [67] to evaluate the GP prediction performance in the absence of both communications and ground truth. Under the proposed novel framework, the GP generalization error is defined to be the expectation of the GP prediction error, $e[\boldsymbol{\xi}_\iota, Q(k)]$, with respect to the hidden variable y_ι. The expected GP generalization error, $e[\boldsymbol{\xi}_\iota, Q(k)]$, at a single collocation point is then obtained by taking the expectation of the GP generalization error with respect to the network measurement vector $\boldsymbol{z}$ as follows:

$$e[\boldsymbol{\xi}_\iota, Q(k)] \triangleq \mathbb{E}_{\boldsymbol{z}}\{\mathbb{E}_{y_\iota}\{\varepsilon[\boldsymbol{\xi}_\iota, Q(k)]\}\} = \phi_k(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_\iota) \tag{27}$$

for any ι, where ϕ_k is the posterior GP covariance function in Eq. (25). The proof of the above result, taken from Ref. [68], is provided in the Appendix for completeness. Then, the GP prediction performance over the entire state space is estimated by averaging the GP generalization error with respect to the network data that may not be available in the absence of communications.

If sensor trajectories are known, the expected GP AGE can be obtained from the expected average value of Eq. (27) over all collocation points in Ω:

$$\mathcal{E}(k) \triangleq \mathbb{E}_{Q(k)}\left\{\sum_{\iota=1}^{M} w_\iota e[\boldsymbol{\xi}_\iota, Q(k)]\right\} \tag{28}$$

where every collocation point is weighed by a positive parameter w_ι that represents the importance of recovering the function value at $\boldsymbol{\xi}_\iota$ ($\iota = 1, \ldots, M$) from noisy sensor measurements, and $\sum_{\iota=1}^{M} w_\iota = 1$. However, in the absence of communications, individual sensors do not have information about other sensors' future policies. Therefore, this paper addresses the fundamental challenge of reducing communications in problems in which the motion and sensing processes are inevitably intertwined. A probabilistic model of future sensor policies (positions) is introduced in order to allow for the local evaluation of the network AGE between communication times. Since the AGE in Eq. (28) only depends on the time step k, it can be evaluated by each sensor at any time, with or without communications with the rest of the network. The relationship of AGE and the number of sensor measurements (Nk) is also commonly known as the *learning curve* [69–71]. As more data are obtained by the sensors over time, GP learning improves decreasing the AGE.

In most cases, exact evaluation of the AGE in Eq. (28) is analytically and computationally intractable because it requires performing a marginalization over the joint distribution of the sensor positions and the collocation points. The marginalization process is further daunted by complexities in the sensor control algorithms. To this end, a recursive algorithm for approximating the AGE efficiently and incrementally online is presented in the next section by exploiting the fact that Ω and $\mathcal{S}$ are both discrete and time-invariant sets.

B. Recursive Approximation of GP AGE

An efficient approximation of the AGE proposed in Eq. (28) is derived in this section by using the recursive update formula of the GP covariance function from [53]. Given a new sensor position for the ith sensor, $\boldsymbol{x}_i(k+1)$, the posterior covariance function can be updated as follows:

$$\phi_{k+1}(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j) = \phi_k(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j) - \frac{\phi_k[\boldsymbol{\xi}_\iota, \boldsymbol{x}_i(k+1)]\phi_k[\boldsymbol{x}_i(k+1), \boldsymbol{\xi}_j]}{\phi_k[\boldsymbol{x}_i(k+1), \boldsymbol{x}_i(k+1)] + \sigma^2} \tag{29}$$

for $\iota, j = 1, \ldots, M$, and $i = 1, \ldots, N$ [53]. Therefore, the expected GP posterior covariance function can be obtained by taking the expectation of Eq. (29) with respect to the sensor positions. Assuming that the sensor measurements are independent, the expected GP posterior covariance can be approximated as follows:

$$\mathbb{E}[\phi_{k+1}(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j)] \approx \tilde{\phi}_{k+1}(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j) = \tilde{\phi}_k(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j) - \sum_{i=1}^{N} \sum_{\boldsymbol{s}_l \in \mathcal{S}_i} \frac{\tilde{\phi}_k(\boldsymbol{\xi}_\iota, \boldsymbol{s}_l)\tilde{\phi}_k(\boldsymbol{s}_l, \boldsymbol{\xi}_j)}{\tilde{\phi}_k(\boldsymbol{s}_l, \boldsymbol{s}_l) + \sigma^2} p_{X_i}(\boldsymbol{s}_l) \tag{30}$$

for $\iota, j = 1, \ldots, M$, where

$$p_{X_i}(\boldsymbol{s}_l) \triangleq p_{X_i}(\boldsymbol{x}_i(k+1) = \boldsymbol{s}_l) \tag{31}$$

denotes the probability that the sensor obtains measurements at position $\boldsymbol{s}_l \in \mathcal{S}_i$, which is a decision made by the sensor control algorithm presented in Sec. V. Note that Eq. (29) is nonlinear in $\phi_k(\boldsymbol{x}_i, \boldsymbol{x}_i)$, and thus Eq. (30) is only an approximation of the true expected covariance $\mathbb{E}[\phi_{k+1}(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_j)]$. Then, the AGE can be iteratively approximated as follows:

$$\mathcal{E}(k) \approx \tilde{\mathcal{E}}(k) = \sum_{\iota=1}^{M} w_\iota \tilde{\phi}_k(\boldsymbol{\xi}_\iota, \boldsymbol{\xi}_\iota) \tag{32}$$

Although Eqs. (30) and (32) present a recursive AGE calculation, they require repeated steps with complexity $O[k(M+L)^3]$ at every time stamp k, and for every new collocation point or new sensor position, where L denotes the total number of admissible sensor locations. Given that the complexity scales with the cube of the total number of admissible sensor positions, the computational complexity can be quite high for many real-world applications. Because sets Ω and $\mathcal{S}$ in many applications can be assumed to be time invariant, it is possible to derive an efficient algorithm with a complexity that does not grow with k. Let $\tilde{\boldsymbol{\Phi}}_k$ denote the approximate expected posterior cross-covariance matrix defined by substituting $\phi(\cdot,\cdot)$ with $\tilde{\phi}_k(\cdot,\cdot)$ in Eq. (7). Then, all the information needed for calculating $\tilde{\mathcal{E}}(k)$ can be stored in an $(M+L)\times(M+L)$ symmetric matrix:

$$\boldsymbol{M} \triangleq \left[\begin{array}{c|c} \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{\xi}) & \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{S}) \\ \hline \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{\xi}) & \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{S}) \end{array}\right] \tag{33}$$

where

$$\boldsymbol{\xi} \triangleq [\boldsymbol{\xi}_1^T \quad \cdots \quad \boldsymbol{\xi}_M^T]^T \tag{34}$$

$$\boldsymbol{S} \triangleq [\boldsymbol{s}_1^T \quad \cdots \quad \boldsymbol{s}_L^T]^T \tag{35}$$

Furthermore, the elements of the set $\{p_{X_i}(\boldsymbol{s}_l)\}_{l=1}^{L}$ can be placed on the diagonal of a zero matrix:

$$\boldsymbol{P}_i \triangleq \mathrm{diag}([p_{X_i}(\boldsymbol{s}_1) \ \cdots \ p_{X_i}(\boldsymbol{s}_L)]) \tag{36}$$

for $i = 1, \ldots, N$. Because sensors have bounded range and can only be placed at a subset of available facilities, it follows that $p_{X_i}(\boldsymbol{s}_l) = 0$ if $\boldsymbol{s}_l \in \mathcal{S}_i$ for $l = 1, \ldots, L$ ($l \neq i$). Now, let $\boldsymbol{P} \triangleq \sum_{i=1}^{N} \boldsymbol{P}_i$, such that the M^2 equations defined by Eq. (30) can be summarized by the following matrix operation:

$$\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{\xi}, \boldsymbol{\xi}) = \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{\xi}) - \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{S})\, \boldsymbol{Q}\, \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{\xi}) \tag{37}$$

where

$$\boldsymbol{Q} \triangleq \boldsymbol{P}\boldsymbol{D}^{-1}\boldsymbol{D} \triangleq (\mathrm{diag}\{\mathrm{diag}[\tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{S})]\} + \sigma^2 \boldsymbol{I}_L),$$

and $\mathrm{diag}(\cdot)$ denotes an operation that retrieves the principal diagonal of a matrix argument. The remaining elements of $\boldsymbol{M}$ are updated according to the relationships

$$\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{\xi}, \boldsymbol{S}) = \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{S}) - \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{S})\, \boldsymbol{Q}\, \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{S}) \tag{38}$$

$$\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{S}, \boldsymbol{S}) = \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{S}) - \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{S})\, \boldsymbol{Q}\, \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{S}, \boldsymbol{S}) \tag{39}$$

Then, the AGE can be obtained by the weighted sum of the diagonal elements of $\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{\xi}, \boldsymbol{\xi})$, such that

$$\tilde{\mathcal{E}}(k) = \mathrm{tr}[\boldsymbol{W} \circ \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{\xi})] \tag{40}$$

where $\boldsymbol{W} \triangleq \mathrm{diag}([w_1 \ \cdots \ w_M])$.

From Eqs. (37–40), it can be seen that the computational complexity of AGE evaluation is dominated by the matrix multiplications, which require $O(L(L^2 + LM + M^2))$ time. Storing the matrix $\boldsymbol{M}$ requires space complexity $O((M + L)^2)$. However, because only the diagonal of $\tilde{\boldsymbol{\Phi}}_k(\boldsymbol{\xi}, \boldsymbol{\xi})$ in the top-left block of $\boldsymbol{M}$ needs to be stored, the space complexity can be reduced to $O(L(L + M))$.

From Eq. (36), it can be seen that the predicted GP generalization error also depends on the sensor control policy. Due to the duality of the sensor inference and control problems, in fact, the GP performance depends not only on the network's past measurements but also on its future actions. In the absence of communications, individual sensors are not only unable to make optimal motion planning decisions but also to predict other sensors' future planning decisions. Therefore, a conservative estimate of the *nominal* GP prediction performance measure is obtained in this paper by assuming a uniform network control algorithm in the AGE approximation. While this policy may never be actually implemented by the sensors, it allows each sensor to locally estimate the AGE by assuming that other sensor positions are sampled uniformly at random from $\mathcal{S}_i$, such that

$$p_{X_i}(\boldsymbol{s}_l) = \begin{cases} \frac{1}{L_i}, & \boldsymbol{s}_l \in \mathcal{S}_i \\ 0, & \boldsymbol{s}_l \notin \mathcal{S}_i \end{cases} \tag{41}$$

for $i = 1, \ldots, N$, and $l = 1, \ldots, L$, where L_i is the cardinality of $\mathcal{S}_i$.

In practice, the sensors use an informed control policy, such as the greedy methodology described in Sec. V or, possibly, an information-driven controller [4]. Nevertheless, the above assumption allows for a local AGE evaluation that is both computationally efficient and conservative. In fact, from Eq. (41), the matrix $\boldsymbol{P}$ used in Eqs. (37–40) can be written as

$$\boldsymbol{P} = \mathrm{diag}\left(\left[\underbrace{\tfrac{1}{L_1} \cdots \tfrac{1}{L_1}}_{L_1 \text{ times}} \ \cdots \ \underbrace{\tfrac{1}{L_N} \cdots \tfrac{1}{L_N}}_{L_N \text{ times}}\right]\right) \tag{42}$$

and, thus, the approximate AGE, denoted by $\tilde{\mathcal{E}}_u(k)$, is obtained by substituting Eq. (42) into Eqs. (37–40). The nominal AGE obtained by assuming a uniform control policy will be denoted by $\mathcal{E}_u(k)$ hereon.

The nominal and approximate AGEs functions, $\mathcal{E}_u(k)$ and $\tilde{\mathcal{E}}_u(k)$, can be viewed as the "average" GP predicted performance among all the sensor planing algorithms. Therefore, they can be used by individual sensors to evaluate the performance of the sensor network in the absence of communications, i.e., when no information about other sensors' decisions is available. The AGE performance baselines established in this section are used in Sec. VI to develop a communication control policy. Because in contested-communication environments sensors may often be disconnected from the rest of the network, a new sensor control algorithm that can function with intermittent communications is also presented in the next section.

V. Sensor Control Policy

This section presents a greedy sensor control algorithm that maximizes the reduction of the AGE in Eq. (28) based on the local data set $Q_i(k)$. In the absence of communications, the local AGE of a sensor can be found similarly to Eq. (28) as follows:

$$\epsilon_i(k) \triangleq \mathbb{E}_{Q_i(k)}\left\{\sum_{t=1}^{M} w_t e[\boldsymbol{\xi}_t, Q_i(k)]\right\} \tag{43}$$

Although the generalization error in Eq. (43) depends only on the local data $Q_i(k)$, it is averaged over all collocation points $\boldsymbol{\xi}_t \in \Omega$, as shown in its definition (28). Then, the next sensor position at time k can be obtained by solving the maximization problem

$$\begin{aligned} \boldsymbol{x}_i^*(k+1) &= \arg\max_{\boldsymbol{x}_i \in \mathcal{S}_i} \{\epsilon_i(k) - \epsilon_i(k+1)\} \\ &= \arg\max_{\boldsymbol{x}_i \in \mathcal{S}_i} \{-\epsilon_i(k+1)\} \end{aligned} \tag{44}$$

for sensor i ($i = 1, \ldots, N$). The second equality in Eq. (44) holds because $\epsilon_i(k)$ is independent of $\boldsymbol{x}_i(k+1)$ at time step k.

The local AGE for the greedy sensor control policy can be obtained similarly to Eq. (40) as follows:

$$\epsilon_i(k+1) = \mathrm{tr}[\boldsymbol{W} \circ \boldsymbol{\Phi}_{k+1}(\boldsymbol{\xi}, \boldsymbol{\xi})] \tag{45}$$

The matrix $\boldsymbol{\Phi}_{k+1}$ can be obtained efficiently from $\boldsymbol{\Phi}_k$ by the GP covariance update formula (29), such that

$$\begin{aligned} \boldsymbol{\Phi}_{k+1}(\boldsymbol{\xi}, \boldsymbol{\xi}) = \boldsymbol{\Phi}_k(\boldsymbol{\xi}, \boldsymbol{\xi}) \\ - \frac{\boldsymbol{\Phi}_k[\boldsymbol{\xi}, \boldsymbol{x}_i(k+1)]\boldsymbol{\Phi}_k[\boldsymbol{x}_i(k+1), \boldsymbol{\xi}]}{\phi_k[\boldsymbol{x}_i(k+1), \boldsymbol{x}_i(k+1)] + \sigma^2} \end{aligned} \tag{46}$$

Then, a greedy sensor policy can be obtained by substituting Eqs. (45) and (46) into Eq. (44), such that

$$\begin{aligned} \boldsymbol{x}_i^*(k+1) &= \arg\max_{\boldsymbol{x}_i \in \mathcal{S}_i} \left\{ \mathrm{tr}\left[\boldsymbol{W} \circ \frac{\boldsymbol{\Phi}_k(\boldsymbol{\xi}, \boldsymbol{x}_i)\boldsymbol{\Phi}_k(\boldsymbol{x}_i, \boldsymbol{\xi})}{\phi_k(\boldsymbol{x}_i, \boldsymbol{x}_i) + \sigma^2} \right] \right\} \\ &= \arg\max_{\boldsymbol{x}_i \in \mathcal{S}_i} \{\boldsymbol{w}^T [\boldsymbol{\Phi}_k(\boldsymbol{\xi}, \boldsymbol{x}_i) \circ \boldsymbol{\Phi}_k(\boldsymbol{\xi}, \boldsymbol{x}_i)]\} \end{aligned} \tag{47}$$

where $\boldsymbol{w} \triangleq [\, w_1 \quad \cdots \quad w_M \,]^T$.

From Eq. (47), it can be seen that computing the ith sensor's position requires $O(LM)$ time. Because the same data structure used in Eq. (33) are also required by the sensor control policy, its space complexity also is $O(L(M + L))$. After the optimal sensor position is found, the matrices $\boldsymbol{\Phi}_k(\boldsymbol{\xi}, \boldsymbol{\xi})$, $\boldsymbol{\Phi}_k(\boldsymbol{\xi}, \boldsymbol{S})$, and $\boldsymbol{\Phi}_k(\boldsymbol{S}, \boldsymbol{S})$ are updated similarly to Eqs. (37–39), where $\boldsymbol{Q}$ is substituted by $[\phi_k(\boldsymbol{x}_i^*, \boldsymbol{x}_i^*) + \sigma^2]^{-1}$. Therefore, the total computational complexity of the sensor control policy presented in this paper is $O(L(M + L))$.

VI. Communication Control Policy

A communication control policy for the class of sensor network problems formulated in Sec. II is developed in this section by leveraging the nominal approximated expected AGE, derived in Sec. IV, and the greedy sensor control policy presented in Sec. V. The philosophy behind the communication control policy developed in this section is that a sensor decides to establish communications with other sensors in the network when the nominal performance of its GP model falls below a minimum acceptable threshold. The expected network performance, as expressed by the approximate AGE derived in Sec. IV, assumes a sensing baseline corresponding to uniform control that is conservative and does not rely on network communications. Furthermore, the communication control policy takes into account the local AGE of the greedy sensor control policy (Sec. V) that represents the network performance as will be improved by the actions of the individual sensor.

Similarly to integral compensation [72], any nonzero difference between the local AGE and predicted network AGE is summed over time and compared to the desired threshold before triggering a communication event. Let k' denote the last communication time and k denote the present time at which communication control a decision is to be made given that sensor i is going to move to the optimal position $\boldsymbol{x}_i^*(k+1)$, obtained from Eq. (47), at the next time step. Then, the communication control policy is given by

$$u_i(k) = \begin{cases} 1, & \text{if } \sum_{\ell=k'}^{k} |\epsilon_i(\ell) - \tilde{\mathcal{E}}_u(\ell)| \geq \gamma \\ 0, & \text{otherwise} \end{cases} \tag{48}$$

where $\epsilon_i(k)$ is the local AGE of the sensor control policy given by Eq. (45), γ is a user-defined threshold, and $\tilde{\mathcal{E}}_u$ is the approximate network AGE obtained by assuming a uniform control policy, as described in Sec. IV. By considering the accrued error between the local AGE and the local prediction of the network AGE, the local AGE may approach the actual network AGE in the steady state. This is because any nonzero difference in the local and predicted network AGE will eventually trigger a communication event, immediately after which the two AGE estimates coincide thanks to the broadcasted sensor trajectory information. Furthermore, although the policy proposed in Eq. (48) is based on the novel AGE approximation, $\tilde{\mathcal{E}}_u$ approaches the nominal network performance ($\mathcal{E}_u$), as shown by the results in Sec. VII.

A schematic of the communication control algorithm is shown in Fig. 3. The pseudocode of the integrated sensor and communication control method is provided in Algorithm 1, where

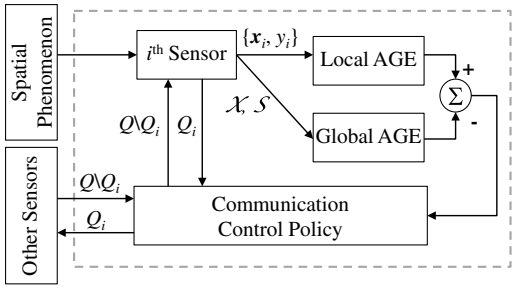

Fig. 3 Communication control architecture.

Algorithm 1 Sensing and communication control algorithm

Input: Admissible sensor positions $\{\mathcal{S}_i\}_{i=1}^N$; Collocation points Ω; GP parameters $\{\theta(\cdot), \phi(\cdot,\cdot)\}$; Communication threshold γ; Cross-covariance matrices $\boldsymbol{\Phi}_k$ and $\tilde{\boldsymbol{\Phi}}_k$

Output: Communication control signal $u_i(k+1) \in \{0, 1\}$; Cross-covariance matrices $\boldsymbol{\Phi}_{k+1}$ and $\tilde{\boldsymbol{\Phi}}_{k+1}$

1: Obtain $\boldsymbol{x}_i^*(k+1)$, by the greedy policy (47).
2: Update $\boldsymbol{\Phi}_{k+1}(\boldsymbol{\xi}, \boldsymbol{\xi})$, with respect to $\boldsymbol{x}_i^*(k+1)$ using Eq. (46).
3: Obtain local AGE $\epsilon_i(k+1)$ from Eq. (45).
4: Update $\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{\xi}, \boldsymbol{\xi})$, $\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{\xi}, \boldsymbol{S})$, $\tilde{\boldsymbol{\Phi}}_{k+1}(\boldsymbol{S}, \boldsymbol{S})$ using Eqs. (37–39).
5: Obtain approximate network AGE $\tilde{\mathcal{E}}_u(k+1)$ from Eq. (40).
6: Determine $u_i(k+1)$ from Eq. (48).

$\boldsymbol{S}_i \triangleq [\boldsymbol{s}_1^T \ \cdots \ \boldsymbol{s}_{L_i}^T]^T$ denotes the $(L_i d \times 1)$ vector that aggregates all of the L_i sensor positions in the set of allowable facilities for sensor i, namely, $\mathcal{S}_i$. Because it does not require knowledge of sensor measurements, the communication control policy (48) may be computed offline provided that all sensor trajectories are known a priori and without error. On the other hand, online evaluation is required when the sensor positions are not known a priori or when the intermittent hyperparameter optimization described in Sec. III.B is adopted in order to improve GP performance.

VII. Simulation Results

The integrated AGE-driven sensing and control method presented in the previous sections is first illustrated on a network of four sensors deployed over a uniform grid (Fig. 4a) to measure and model a spatio-temporal temperature process. The sensor and communication control results are shown at two sample moments in time in Figs. 4 and 5. The instantaneous sensor position and communication links are shown in Figs. 4a, 4b, and 5a and the time-histories of individual sensors' AGEs are shown in Figs. 4a, 4b, and 5b. The network's approximate predicted AGE and the average local AGE in the total absence of communications are also plotted for comparison by dashed red and gray lines, respectively. Corrections in the network AGE approximation are observed at the communication times, where the approximation becomes exact by communicating the true sensor position histories. These results demonstrate that, by using the proposed integrated sensing and control policy, the sensors are able to achieve a GP generalized prediction performance near that of the ideal constant communication case and below the desired threshold by means of only few communication times (brown vertical lines). It can be seen that the frequency of communication required decreases over time as the model improves and local measurements become useful even without fusion.

The overall effectiveness of the sensing and communication control policy is demonstrated by using a Monte Carlo (MC) simulation of the GP temperature modeling problem formulated in Sec. II, with nonuniform sensor locations. The MC simulation takes into account all random inputs, including the GP hyperparameters, that affect the performance of the integrated sensing and communication policy. The AGE approximation, $\tilde{\mathcal{E}}_u(k)$, is first compared to the nominal AGE, $\mathcal{E}_u(k)$, defined in Sec. IV in order to validate the theoretical result derived in Sec. IV.B. MC simulations are performed to obtain the nominal AGE, $\mathcal{E}_u(k)$, for the spatial process and the admissible sensor positions shown in Fig. 1. The collocation points plotted in Fig. 2 are used to compute $\tilde{\mathcal{E}}_u(k)$ for various length scales (λ_i) of the covariance function in Eq. (12), which play a key role in the GP generalization ability. As shown in Fig. 6, the approximate AGE closely approaches the nominal AGE for all length scales, ranging from small ($\lambda_i = 10^{-2}$) to large ($\lambda_i = 10^3$) values. Therefore, it can be concluded that the computational savings afforded by the approximation can be justified without compromising performance.

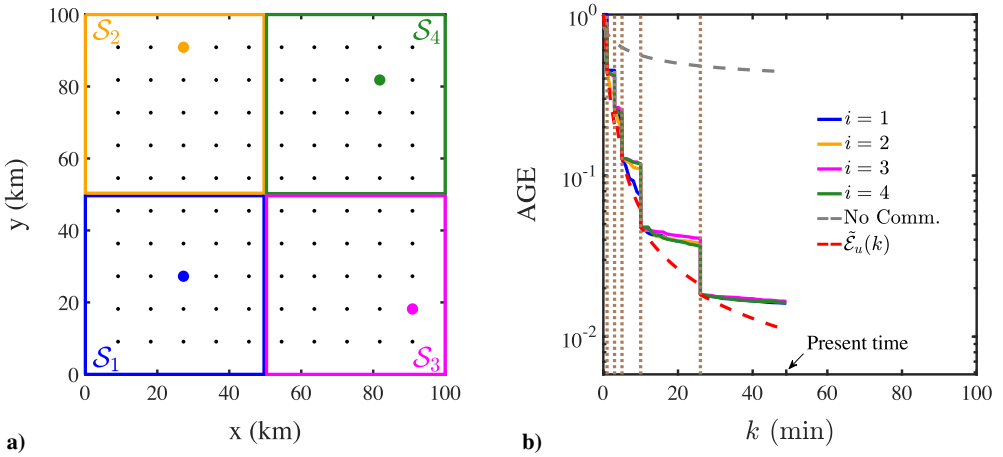

Fig. 4 Sensor positions (colored dots) (a) and AGE time histories (b) of individual sensors ($i = 1, \ldots, 4$) with controlled intermittent communications (brown dotted lines), and average sensor performance in the absence of communication up to the present time ($k = 50$).

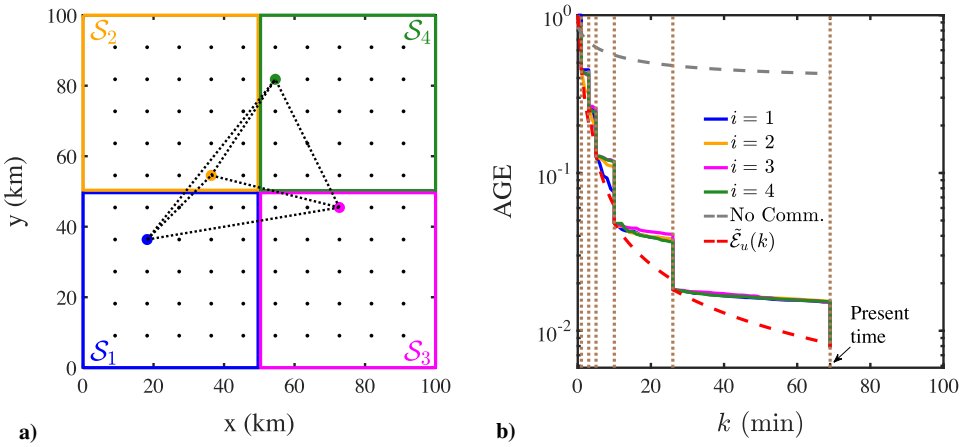

Fig. 5 Sensor positions (colored dots) and communication links (dotted lines) (a) and AGE time histories (b) of individual sensors ($i = 1, \ldots, 4$) with controlled intermittent communications (brown dotted lines), and average sensor performance in the absence of communication up to $k = 70$.

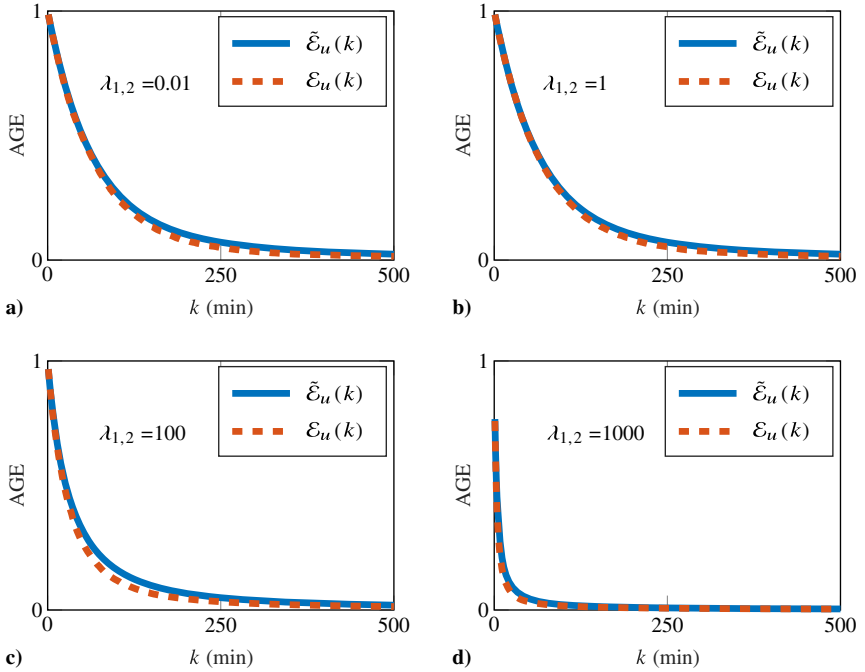

Fig. 6 Validation of novel AGE approximation via MC simulation as a function of length scale.

The performance of the communication control policy for a representative sensor ($i = 1$) is shown in Fig. 7 for a case study with $\lambda_1 = \lambda_2 = 10$ and a performance threshold $\gamma = 0.4$. The results of the communication control policy are illustrated by dotted brown lines that represent all communication times. In between these times, none of the sensors communicate to minimize cost and maximize covertness. A comparison of the local sensor AGE, $\epsilon_1(k)$, the nominal AGE, $\epsilon(k)$, and the approximate AGE, $\tilde{\mathcal{E}}_u(k)$, shows that when performance falls below the threshold the sensor initiates communications and, as a result, at these times sensor fusion improves the GP model. Since the communication times are chosen based on the accumulated AGE errors, the communication control policy (48) is able to automatically adjust the time interval between intermittent communications, reducing the individual sensor error $\epsilon_i(k)$ to $\epsilon(k)$ as a result. Also, Fig. 7 shows that at the onset of the sensing process, when the nominal AGE decreases rapidly because the first measurements are becoming available, more frequent communications are initiated by the sensors. As a result, the cost of communications is offset by a high uncertainty reduction, and the sensing and communication control algorithm learns the spatial process very efficiently. In contrast, toward the end of the sensing process, when new measurements are less informative, the control policy increases the intervals between intermittent communications because fusion brings about less gain compared to the individual sensor measurements.

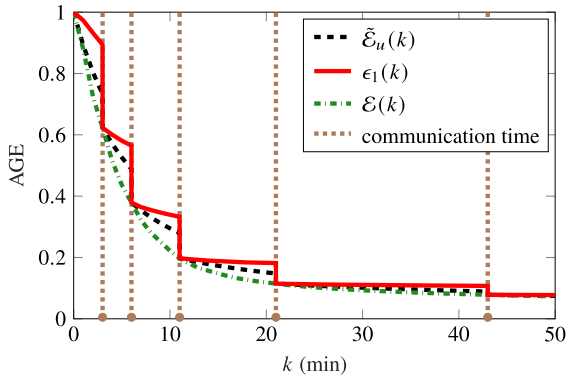

Fig. 7 Communication times and AGE values obtained by intermittent communication control policy for $\lambda_1 = \lambda_2 = 10$ and $\gamma = 0.4$.

The effect of GP length scales (λ), which reflect the assumption of the complexity of the spatial process under observation, is also investigated. Figure 8 shows that the required communication frequency decreases as the length scales of the GP increase. This behavior is expected because smaller GP length scale indicates that the underlying spatial process is more complex, and measurements of the spatial process by different sensors are less correlated. In other words, when GP length scales are small, individual sensors cannot adequately model and predict the spatio-temporal process over the entire state space based solely on their individual measurements. Therefore, as determined by the communication control algorithm, more frequent communications are needed.

The performance of the AGE-based communication control policy is compared to a random communication policy in Fig. 9. The average random communication control network AGE $\mathcal{E}_r$ is the AGE averaged over 100 MC trials, where in each trial, five communication times are randomly chosen. Note that the number of communication events is kept constant across both policies and all trials to ensure a fair comparison. As shown, the AGE-based communication policy results in a network AGE $\mathcal{E}_g$ that is several orders of magnitude lower than that of the average random communication policy. In the simulation of both policies, GP hyperparameter optimization is performed at the times of communication to further reduce the AGE, as evidenced by the sharp decreases in $\mathcal{E}_g$ at communication times.

Simultaneously to sharing and fusing measurements at communication times, the GP hyperparameter optimization described in Sec. III.B can be carried out to improve the GP model of the underlying process. This approach has shown to lead to a reduction of individual sensor AGEs, demonstrated here by considering the same sensor control problem with and without GP hyperparameters optimization at intermittent communication times. As shown in Fig. 10, the AGE obtained by intermittent GP hyperparameter optimization ($\epsilon_1^*(k)$) is significantly reduced when compared to the traditional AGE ($\epsilon_1(k)$).

A second study, carried out using a database of mean temperatures obtained in March 2016, demonstrates the performance of the AGE-based communication control policy in a large-scale sensing problem. Figure 11 shows the sequence of measurement locations by four notional sensors, such as satellites capable of making focused thermal infrared radiation measurements. The integrated sensing and communication control policy results in a significant decrease in AGE using only six broadcast events, as shown in Fig. 12. Hyperparameter optimization is also shown to

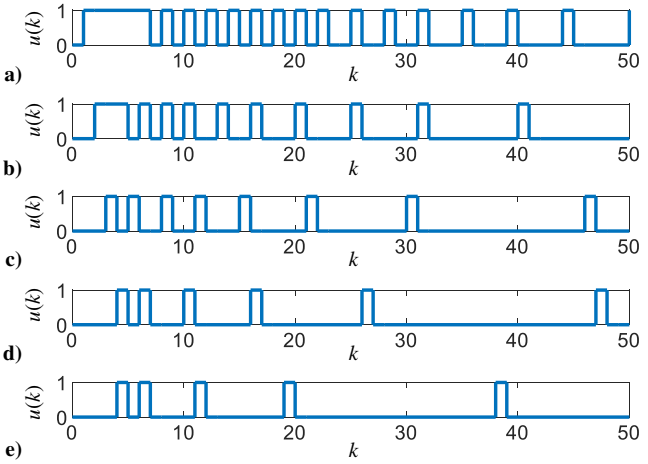

Fig. 8 Communication control policy history for various GP length scales: a) $\lambda_1 = \lambda_2 = 0.1$, b) $\lambda_1 = \lambda_2 = 1$, c) $\lambda_1 = \lambda_2 = 10$, d) $\lambda_1 = \lambda_2 = 20$, and e) $\lambda_1 = \lambda_2 = 50$.

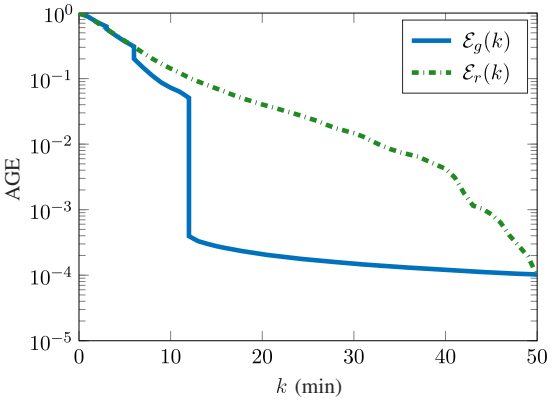

Fig. 9 Comparison of AGE-driven and random intermittent communications.

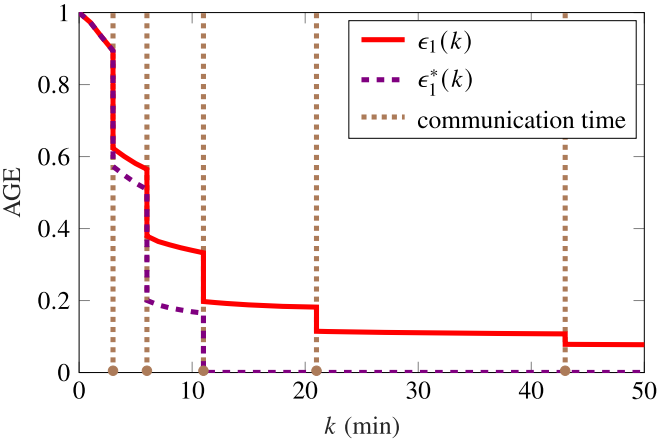

Fig. 10 Effect of GP hyperparameters optimization with controlled intermittent communications.

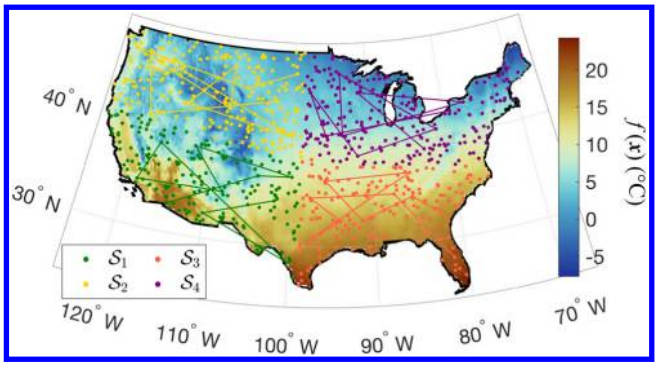

Fig. 11 Progression of measurement locations over first 10 time steps.

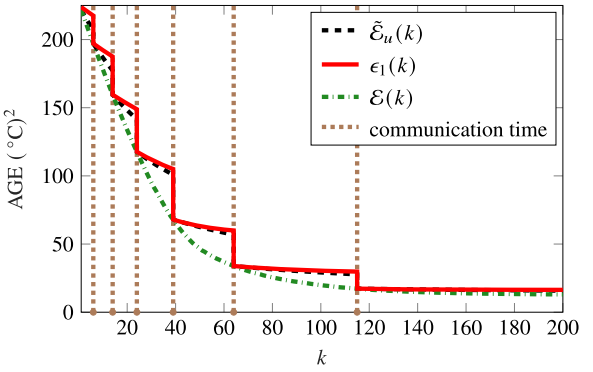

Fig. 12 Communication times and unnormalized AGE values obtained by intermittent communication control policy used in Fig. 11.

lead to improved GP regression performance. The GP regression results with and without GP hyperparameter optimization, shown in Figs. 13 and 14, respectively, indicate that the prediction errors are significantly reduced in the former. Furthermore, the final AGE with and without hyperparameter optimization is 8.41 $(°C)^2$ and 13.05 $(°C)^2$, respectively, suggesting that GP hyperparameter optimization under the intermittent communication control policy is capable of improving both the GP model generalization (AGE) and prediction performance.

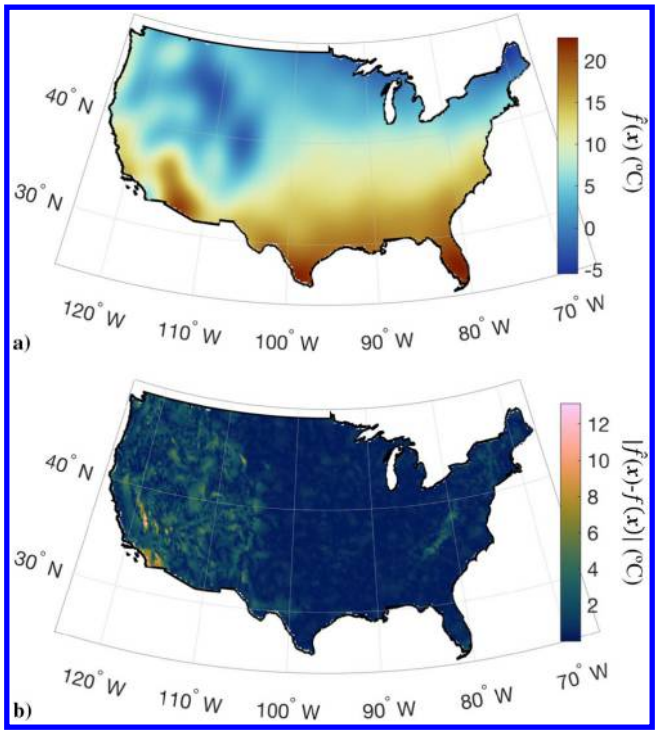

Fig. 13 Temperature prediction (a) and prediction error (b) *with* optimization of the GP hyperparameters. True value data shown in Fig. 2.

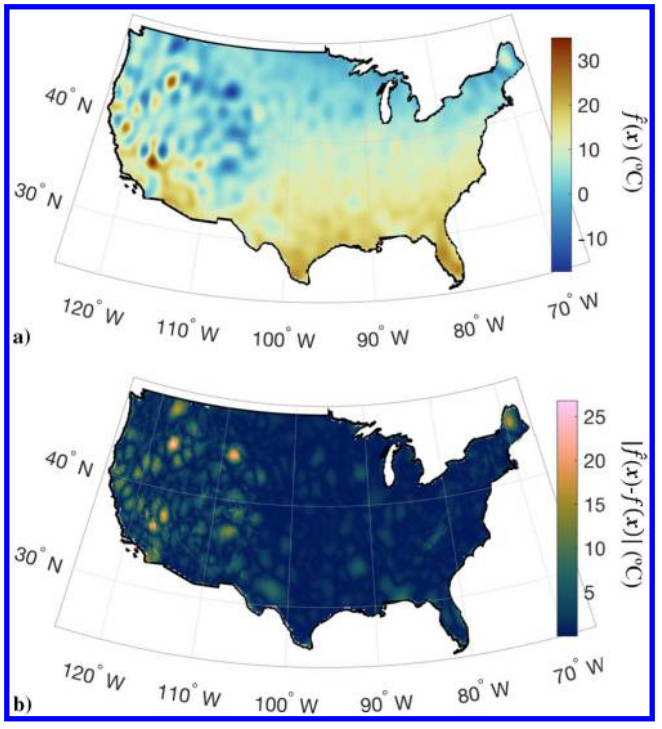

Fig. 14 Temperature prediction (a) and prediction error (b) *without* optimization of the GP hyperparameters. True value data shown in Fig. 2.

VIII. Conclusions

This paper presents a new method for controlling intermittent communications in networks of collaborative mobile sensors. Collaborative sensor networks require communications in order to fuse and learn from distributed sensor measurements, as well as to plan future actions. In many sensing applications, however, constant communications are not always possible and, furthermore, require significant resources, such as energy and time required to resurface and establish a communication link. This paper presents a new approach for controlling intermittent communications based on the AGE of the learned GP. The approach allows to decide locally if and when to establish communication links, based on the expected quality of service expressed by the AGE. By this approach, the network is able to learn the GP model of a nonlinear spatio-temporal process by means of minimal communications, as demonstrated through a temperature modeling problem for the Conterminous United States.

Appendix: GP Generalization Error

It was shown in [68] that the right-hand side of the GP generalization error in Eq. (27) can be expanded and grouped with respect to the expectations, as follows:

$$e[\boldsymbol{\xi}_t, Q(k)] = \mathbb{E}_{y_t}[f(\boldsymbol{\xi}_t)^2] - 2\boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\mathbb{E}_z\mathbb{E}_{y_t}[f(\boldsymbol{\xi}_t)z] + \mathbb{E}_z[\boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}zz^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{\xi}_t)] \tag{A1}$$

Since the third term of Eq. (49) is a scalar, it can be rearranged as follows:

$$\begin{aligned} e[\boldsymbol{\xi}_t, Q(k)] = {} & \mathbb{E}_{y_t}[f(\boldsymbol{\xi}_t)^2] - 2\boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\mathbb{E}_z\mathbb{E}_{y_t}[f(\boldsymbol{\xi}_t)z] \\ & + \mathbb{E}_z[z^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{\xi}_t)\boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1})z] \\ & + \mathbb{E}_z\left\{\mathrm{tr}[z^T(\boldsymbol{\Sigma}^{-1}\Phi(\boldsymbol{X}, \boldsymbol{\xi}_t)\Phi(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1})z]\right\} \end{aligned} \tag{A2}$$

Because the trace of a matrix is invariant under cyclic permutation, Eq. (A1) can be expressed in terms of the covariance matrices in Eq. (7), such that

$$e[\boldsymbol{\xi}_t, Q(k)] = \boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t) - 2\boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{\xi}_t) + \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\Phi(\boldsymbol{X}, \boldsymbol{\xi}_t)\Phi(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\mathbb{E}_z(zz^T)] \tag{A3}$$

Substituting $\mathbb{E}_z(zz^T) = \boldsymbol{\Sigma}$ into Eq. (A3) and applying the cyclic permutation to the third term of Eq. (A3), it can be shown that the following holds:

$$\begin{aligned} e[\boldsymbol{\xi}_t, Q(k)] &= \boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t) - \boldsymbol{\Phi}(\boldsymbol{\xi}_t, \boldsymbol{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{\xi}_t) \\ &= \phi_k(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t) \end{aligned} \tag{A4}$$

Acknowledgments

This work was supported in part by Office of Naval Research (ONR) Code 321 Grant N00014-16-1-2447, ONR Multidisciplinary University Research Initiative (MURI) Grant N00014-11-1-0688, and ONR Basic Research Challenge (BRC) Grant N00014-17-1-2175. Keith LeGrand was supported under the National Defense Science and Engineering Graduate (NDSEG) Fellowship.

References

[1] Akyildiz, I., Su, W., Sankarasubramaniam, Y., and Cayirci, E., "Wireless Sensor Networks: A Survey," *Computer Networks*, Vol. 38, No. 4, 2002, pp. 393–422.

[2] Yick, J., Mukherjee, B., and Ghosal, D., "Wireless Sensor Network Survey," *Computer Networks*, Vol. 52, No. 12, 2008, pp. 2292–2330.

[3] Kammer, D. C., "Sensor Placement for On-Orbit Modal Identification and Correlation of Large Space Structures," *Journal of Guidance, Control, and Dynamics*, Vol. 14, No. 2, 1991, pp. 251–259, Chap. 24.

[4] Ferrari, S., and Wettergren, T. A., *Information-Driven Planning and Control*, MIT Press, Cambridge, MA, 2021.

[5] Fisher, T., and Sharma, R., "Rudder-Augmented Trajectory Correction for Small Unmanned Aerial Vehicle to Minimize Lateral Image Errors," *Journal of Aerospace Information Systems*, Vol. 15, No. 9, 2018, pp. 544–552.

[6] Valasek, J., Kirkpatrick, K., May, J., and Harris, J., "Intelligent Motion Video Guidance for Unmanned Air System Ground Target Surveillance," *Journal of Aerospace Information Systems*, Vol. 13, No. 1, 2016, pp. 10–26.

[7] Wei, H., and Ferrari, S., "A Geometric Transversals Approach to Analyzing the Probability of Track Detection for Maneuvering Targets," *IEEE Transactions on Computers*, Vol. 63, No. 11, 2014, pp. 2633–2646.

[8] Wei, H., Lu, W., Zhu, P., Huang, G., Leonard, J., and Ferrari, S., "Optimized Visibility Motion Planning for Target Tracking and Localization," *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Inst. of Electrical and Electronics Engineers, New York, 2014, pp. 76–82.

[9] Li, X. R., and Jilkov, V. P., "Survey of Maneuvering Target Tracking. Part I. Dynamic Models," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 39, No. 4, 2003, pp. 1333–1364.

[10] Wolf, M., Minzlaff, M., and Moser, M., "Information Technology Security Threats to Modern E-Enabled Aircraft: A Cautionary Note," *Journal of Aerospace Information Systems*, Vol. 11, No. 7, 2014, pp. 447–457.

[11] Goppert, J., Shull, A., Sathyamoorthy, N., Liu, W., Hwang, I., and Aldridge, H., "Software/Hardware-in-the-Loop Analysis of Cyberattacks on Unmanned Aerial Systems," *Journal of Aerospace Information Systems*, Vol. 11, No. 5, 2014, pp. 337–343.

[12] Von Moll, A., Garcia, E., Casbeer, D., Suresh, M., and Swar, S. C., "Multiple-Pursuer, Single-Evader Border Defense Differential Game," *Journal of Aerospace Information Systems*, Vol. 17, No. 8, 2020, pp. 407–416.

[13] Ellingson, G., Brink, K., and McLain, T., "Cooperative Relative Navigation of Multiple Aircraft in Global Positioning System-Denied/Degraded Environments," *Journal of Aerospace Information Systems*, Vol. 17, No. 8, 2020, pp. 470–480.

[14] Bakule, L., and Papík, M., "Decentralized Control and Communication," *Annual Reviews in Control*, Vol. 36, No. 1, 2012, pp. 1–10.

[15] Chen, D., and Varshney, P. K., "QoS Support in Wireless Sensor Networks: A Survey," *International Conference on Wireless Networks*, Vol. 233, CSREA Press, Las Vegas, NV, 2004.

[16] Ulvklo, M., Nygards, J., Karlholm, J., Skoglar, P., Ahlberg, J., and Nilsson, J., "A Sensor Management Framework for Autonomous UAV Surveillance," *Proceedings of SPIE*, Vol. 5787, 2005, p. 49.

[17] Skoglar, P., "Planning Methods for Aerial Exploration and Ground Target Tracking," Doctoral Dissertation, Linköping Univ. Electronic Press, 2009.

[18] Hristu, D., and Morgansen, K., "Limited Communication Control," *Systems & Control Letters*, Vol. 37, No. 4, 1999, pp. 193–205.

[19] Chu, D., Deshpande, A., Hellerstein, J. M., and Hong, W., "Approximate Data Collection in Sensor Networks Using Probabilistic Models," *22nd International Conference on Data Engineering (ICDE'06)*, IEEE, New York, 2006, p. 48.
https://doi.org/10.1109/ICDE.2006.21

[20] Vasilescu, I., Kotay, K., Rus, D., Dunbabin, M., and Corke, P., "Data Collection, Storage, and Retrieval with an Underwater Sensor Network," *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*, Assoc. for Computing Machinery, New York, 2005, pp. 154–165.
https://doi.org/10.1145/1098918.1098936

[21] Spencer, S. J., Parikh, A., McArthur, D. R., Young, C. C., Blada, T. J., Slightam, J. E., and Buerger, S. P., "Autonomous Detection and Assessment with Moving Sensors," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, New York, 2020, pp. 8231–8238.
https://doi.org/10.1109/IROS45743.2020.9340997

[22] Anastasi, G., Conti, M., Di Francesco, M., and Passarella, A., "Energy Conservation in Wireless Sensor Networks: A Survey," *Ad Hoc Networks*, Vol. 7, No. 3, 2009, pp. 537–568, https://www.sciencedirect.com/science/article/pii/S1570870508000954.
https://doi.org/10.1016/j.adhoc.2008.06.003

[23] Walsh, G. C., Beldiman, O., and Bushnell, L., "Asymptotic Behavior of Networked Control Systems," *Proceedings of the 1999 IEEE International Conference on Control Applications, 1999*, Vol. 2, Inst. of Electrical and Electronics Engineers, New York, 1999, pp. 1448–1453.

[24] Jiang, X., and Han, Q.-L., "New Stability Criteria for Linear Systems with Interval Time-Varying Delay," *Automatica*, Vol. 44, No. 10, 2008, pp. 2680–2685.

[25] Branicky, M. S., Phillips, S. M., and Zhang, W., "Stability of Networked Control Systems: Explicit Analysis of Delay," *Proceedings of the 2000 American Control Conference, 2000*, Vol. 4, Inst. of Electrical and Electronics Engineers, New York, 2000, pp. 2352–2357.

[26] Walsh, G. C., Ye, H., and Bushnell, L. G., "Stability Analysis of Networked Control Systems," *IEEE Transactions on Control Systems Technology*, Vol. 10, No. 3, 2002, pp. 438–446.

[27] Kim, D.-S., Lee, Y. S., Kwon, W. H., and Park, H. S., "Maximum Allowable Delay Bounds of Networked Control Systems," *Control Engineering Practice*, Vol. 11, No. 11, 2003, pp. 1301–1313.

[28] Tabbara, M., Neaić, D., and Teel, A. R., "Stability of Wireless and Wireline Networked Control Systems," *IEEE Transactions on Automatic Control*, Vol. 52, No. 9, 2007, pp. 1615–1630.

[29] Liu, K., Fridman, E., and Johansson, K. H., "Networked Control with Stochastic Scheduling," *IEEE Transactions on Automatic Control*, Vol. 60, No. 11, 2015, pp. 3071–3076.

[30] Liu, Q., Wang, Z., He, X., and Zhou, D., "A Survey of Event-Based Strategies on Control and Estimation," *Systems Science & Control Engineering: An Open Access Journal*, Vol. 2, No. 1, 2014, pp. 90–97.

[31] Miskowicz, M., "Send-on-Delta Concept: An Event-Based Data Reporting Strategy," *Sensors*, Vol. 6, No. 1, 2006, pp. 49–63.

[32] Otanez, P. G., Moyne, J. R., and Tilbury, D. M., "Using Deadbands to Reduce Communication in Networked Control Systems," *Proceedings of the 2002 American Control Conference, 2002*, Vol. 4, Inst. of Electrical and Electronics Engineers, New York, 2002, pp. 3015–3020.

[33] Teixeira, P. V., Dimarogonas, D. V., Johansson, K. H., and Sousa, J., "Multi-Agent Coordination with Event-Based Communication," *2010 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, 2010, pp. 824–829.

[34] Narendra, K. S., and Mukhopadhyay, S., "To Communicate or Not to Communicate: A Decision-Theoretic Approach to Decentralized Adaptive Control," *2010 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, 2010, pp. 6369–6376.

[35] Wang, X., and Lemmon, M. D., "Event-Triggering in Distributed Networked Control Systems," *IEEE Transactions on Automatic Control*, Vol. 56, No. 3, 2011, pp. 586–601.

[36] Peng, C., Han, Q.-L., and Yue, D., "To Transmit or Not to Transmit: A Discrete Event-Triggered Communication Scheme for Networked Takagi-Sugeno Fuzzy Systems," *IEEE Transactions on Fuzzy Systems*, Vol. 21, No. 1, 2013, pp. 164–170.

[37] Wei, H., Zhu, P., Liu, M., How, J. P., and Ferrari, S., "Automatic Pan–Tilt Camera Control for Learning Dirichlet Process Gaussian Process (DPGP) Mixture Models of Multiple Moving Targets," *IEEE Transactions on Automatic Control*, Vol. 64, No. 1, 2018, pp. 159–173.

[38] Wagle, N., and Frew, E. W., "Forward Adaptive Transfer of Gaussian Process Regression," *Journal of Aerospace Information Systems*, Vol. 14, No. 4, 2017, pp. 214–231.

[39] Wei, H., Lu, W., Zhu, P., Ferrari, S., Liu, M., Klein, R., Omidshafiei, S., and How, J. P., "Information Value in Nonparametric Dirichlet-Process Gaussian-Process (DPGP) Mixture Models," *Automatica*, Vol. 74, 2016, pp. 360–368.

[40] Grande, R. C., Chowdhary, G., and How, J. P., "Experimental Validation of Bayesian Nonparametric Adaptive Control Using Gaussian Processes," *Journal of Aerospace Information Systems*, Vol. 11, No. 9, 2014, pp. 565–578.

[41] Wei, H., Lu, W., Zhu, P., Ferrari, S., Liu, M., Klein, R., Omidshafiei, S., and How, J. P., "Information Value in Nonparametric Dirichlet-Process Gaussian-Process (DPGP) Mixture Models," *Automatica*, Vol. 74, 2016, pp. 360–368.

[42] Singh, A., Ramos, F., Whyte, H. D., and Kaiser, W. J., "Modeling and Decision Making in Spatio-Temporal Processes for Environmental Surveillance," *2010 IEEE International Conference on Robotics and Automation (ICRA)*, Inst. of Electrical and Electronics Engineers, New York, 2010, pp. 5490–5497.

[43] Hoang, T. N., Hoang, Q. M., and Low, B. K. H., "A Distributed Variational Inference Framework for Unifying Parallel Sparse Gaussian Process Regression Models," *Proceedings of the 33rd International Conference on Machine Learning*, International Machine Learning Soc., Princeton, NJ, 2016, pp. 382–391.

[44] Joseph, J., Doshi-Velez, F., Huang, A. S., and Roy, N., "A Bayesian Nonparametric Approach to Modeling Mobility Patterns," *Autonomous Robots*, Vol. 31, No. 4, Nov. 2011, pp. 383–400.

[45] Simmons, G. F., *Introduction to Topology and Modern Analysis*, Vol. 3, McGraw–Hill, New York, 1963, Chaps. 2, 4, and 6.

[46] Ferrari, S., Fierro, R., Perteet, B., Cai, C., and Baumgartner, K., "A Geometric Optimization Approach to Detecting and Intercepting Dynamic Targets Using a Mobile Sensor Network," *SIAM Journal on Control and Optimization*, Vol. 48, No. 1, 2009, pp. 292–320.

[47] Ferrari, S., Fierro, R., and Tolic, D., "A Geometric Optimization Approach to Tracking Maneuvering Targets Using a Heterogeneous Mobile Sensor Network," *Proceedings of Decision and Control Conference*, IEEE, New York, 2009, pp. 1080–1087.
[48] Ferrari, S., Cai, C., Fierro, R., and Perteet, B., "A Multi-Objective Optimization Approach to Detecting and Tracking Dynamic Targets in Pursuit-Evasion Games," *Proceedings of American Control Conference*, IEEE, New York, 2007, pp. 5316–5321.
[49] Ferrari, S., Zhang, G., and Wettergren, T. A., "Probabilistic Track Coverage in Cooperative Sensor Networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 40, No. 6, 2010, pp. 1492–1504.
[50] Krause, A., Singh, A., and Guestrin, C., "Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies," *Journal of Machine Learning Research*, Vol. 9, Feb. 2008, pp. 235–284.
[51] Ny, J. L., and Pappas, G. J., "On Trajectory Optimization for Active Sensing in Gaussian Process Models," *Proceedings of the IEEE Conference on Decision and Control*, IEEE, New York, 2009, pp. 6286–6292.
[52] Wei, H., Lu, W., Zhu, P., Ferrari, S., Klein, R. H., Omidshafiei, S., and How, J. P., "Camera Control for Learning Nonlinear Target Dynamics Via Bayesian Nonparametric Dirichlet-Process Gaussian-Process (DP-GP) Models," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Inst. of Electrical and Electronics Engineers, New York, 2014, pp. 95–102.
[53] Rasmussen, C. E., and Williams, C., *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006, Chap. 2.
[54] Kolmogorov, A. N., *Foundations of the Theory of Probability*, Chelsea Publishing Co., New York, 1950, pp. 29–30.
[55] Athreya, K. B., and Lahiri, S. N., *Measure Theory and Probability Theory*, Springer Science & Business Media, New York, 2006, pp. 199–212.
[56] Rasmussen, C. E., "Gaussian Processes in Machine Learning," *Advanced Lectures on Machine Learning*, Springer, Berlin, Heidelberg, 2004, pp. 63–71.
[57] Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W., "Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting," *IEEE Transactions on Information Theory*, Vol. 58, No. 5, 2012, pp. 3250–3265.
[58] Zhu, P., Wei, H., Lu, W., and Ferrari, S., "Multi-Kernel Probability Distribution Regressions," *2015 International Joint Conference on Neural Networks (IJCNN)*, Inst. of Electrical and Electronics Engineers, New York, 2015, pp. 1–7.
[59] Seeger, M., "Gaussian Processes for Machine Learning," *International Journal of Neural Systems*, Vol. 14, No. 02, 2004, pp. 69–106.
[60] MacKay, D. J., "Introduction to Monte Carlo Methods," *Learning in Graphical Models*, Springer, Dordrecht, The Netherlands, 1998, pp. 175–204.
[61] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer–Verlag, New York, 2006, Chap. 6.
[62] Shewchuk, J. R., "An Introduction to the Conjugate Gradient Method Without the Agonizing Pain," School of Computer Science Carnegie Mellon Univ., Pittsburgh, PA, 1994, https://perma.cc/NPK3-T9GC.
[63] White, H., "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings," *Neural Networks*, Vol. 3, No. 5, 1990, pp. 535–549.
[64] Sollich, P., "Learning Curves for Gaussian Processes," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 1999, pp. 344–350.
[65] Opper, M., "Regression with Gaussian Processes: Average Case Performance," *Theoretical Aspects of Neural Computation: A Multidisciplinary Perspective*, World Scientific, Singapore, 1997, pp. 17–23.
[66] Williams, C. K., "Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond," *Learning in Graphical Models*, Springer, Dordrecht, The Netherlands, 1998, pp. 599–621.
[67] Williams, C. K. I., and Vivarelli, F., "Upper and Lower Bounds on the Learning Curve for Gaussian Processes," *Machine Learning*, Vol. 40, No. 1, 2000, pp. 77–102.
[68] Ashton, S., and Sollich, P., "Learning Curves for Multi-Task Gaussian Process Regression," *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Red Hook, NY, 2012, pp. 1781–1789.
[69] Opper, M., and Vivarelli, F., "General Bounds on Bayes Errors for Regression with Gaussian Processes," *Advances in Neural Information Processing Systems*, Vol. 11, MIT Press, Cambridge, MA, 1999, pp. 302–308.
[70] Sollich, P., "Gaussian Process Regression with Mismatched Models," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2002, pp. 519–526.
[71] Urry, M., and Sollich, P., "Exact Learning Curves for Gaussian Process Regression on Large Random Graphs," *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Red Hook, NY, 2010, pp. 2316–2324.
[72] Stengel, R. F., *Optimal Control and Estimation*, Dover, New York, 1994, Chap. 6.

J. P. How
Associate Editor