Reliability Engineering and System Safety 154 (2016) 219–233

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Value of information for spatially distributed systems: Application to sensor placement

Carl Malings, Matteo Pozzi *

Department of Civil and Environmental Engineering, Carnegie Mellon University, 107b Porter Hall, 5000 Forbes Ave., Pittsburgh, PA 15213-3890, USA

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form
21 May 2016
Accepted 24 May 2016
Available online 25 May 2016

Keywords:
Pre-posterior analysis
Sensor placement
Gaussian random fields

ABSTRACT

This paper investigates how the value of information (VoI) metric can guide information collection and optimal sensor placement in spatially distributed systems. VoI incorporates relevant features to decision-making, such as uncertainty about the state of the system, precision of measurements, the availability of intervention actions, and the overall cost of managing the system. Spatially distributed systems also allow for information propagation, i.e. measurements collected at one location can be used to update knowledge at other related locations. In this paper, while restricting our attention to Gaussian random field and binary state models, we illustrate first how sensor placements depend on the decision-making problem to be addressed, as encoded in a problem-specific loss function, and second how the complexity of VoI computations is impacted by this loss function's characteristics. In doing so, we consider several loss functions and present computational techniques for evaluating VoI under them. Finally, we apply these techniques to efficiently optimize sensor placements by the VoI metric in two example applications.

1. Introduction

Civil infrastructures consist of multiple components distributed over a spatial domain, which act together to fulfill the function of the system. The behaviors of these components are affected by spatially varying quantities, which can be probabilistically described using random fields. Fig. 1 gives examples of random fields and spatially distributed systems. Managers of these systems make decisions under uncertainty in these fields, trading off the cost of maintenance activities against the risk and consequences of component and system malfunctioning. Observations collected by measuring these fields reduce uncertainty and improve decision outcomes, but are expensive to acquire. Dependencies between field variables in a distributed system present opportunities for efficient information gathering, allowing observations on one part of the system to reduce uncertainties on related components. Optimal selection of these measurements, taking information propagation into account, is therefore important for efficient monitoring and management of distributed infrastructure systems.

Many factors impact the selection of these measurements; prior uncertainty in random fields, dependencies between variables in these fields, prior risks and consequences of component and system failures, the cost of possible actions to avoid these failures, and the relative costs of sensing activities all have an impact. In general, to address the problem of optimal sensor placement (where the term "sensor placement" refers broadly to the selection of a set of measurements), a metric is needed which quantifies the utility of information gathering, allowing these to be compared and optimized.

Several such metrics have been proposed. Metrics based on modal identification have been investigated by e.g. [1,2] for application to sensor placement for the monitoring of dynamic structures. The seminal work of Krause applies information-theoretic concepts like conditional entropy to more general sensor placement problems, where the sub-modular properties of this metric allow for provable near-optimal sensor placements by simple greedy optimization approaches [3]. In particular, these methods are efficiently applied in systems where variables and measurements can be jointly described by Gaussian process models [4—6]. Recently, similar techniques have been applied to efficient level set estimation and exploration/exploitation tradeoffs in online sensing [7,8].

Another general metric for sensor placement is the decision-theoretic value of information (VoI). VoI quantifies the benefit of obtaining information to a decision-maker, in terms of guiding the choice of actions to minimize expected losses specified by a predefined loss function [9]. VoI provides a rational metric for optimizing sensor placements for supporting decision-making by

* Corresponding author.
E-mail addresses: cmalings@andrew.cmu.edu (C. Malings), mpozzi@cmu.edu (M. Pozzi).

http://dx.doi.org/10.1016/j.ress.2016.05.010
0951-8320/

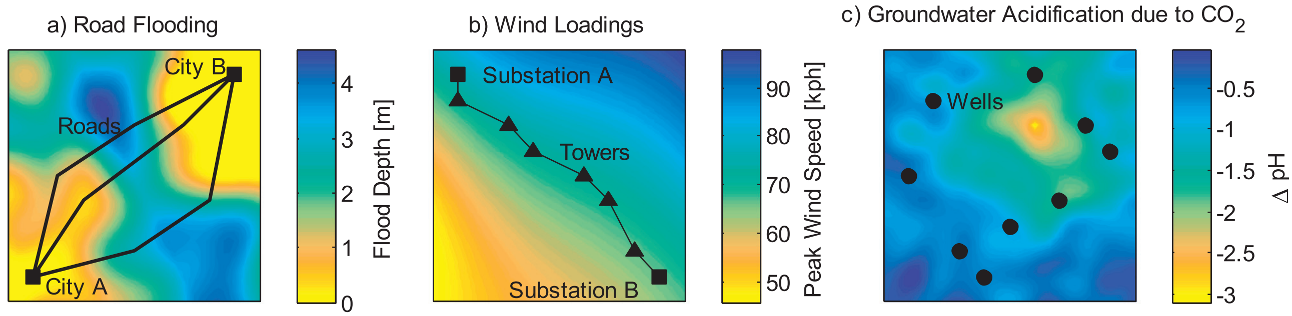

Fig. 1. Examples of distributed systems subjected to random field stresses: a) roadways crossing a floodplain with varying depth; b) an electrical transmission line affected by spatially varying wind loads; c) a group of wells subjected to groundwater acidification from leakage of a carbon sequestration reservoir.

directly assessing the benefits of sensing in terms of expected cost reduction. Efficient evaluation and optimization of this metric in chain graphical models of discrete random variables is investigated by [10]. Previous work has quantified the long-term benefits of structural health monitoring through VoI [11,12]. The state of the art in applications of VoI to infrastructures is illustrated in e.g. [13—17].

A major limitation of VoI as a sensor placement optimization metric is its computational complexity. Evaluation of VoI requires pre-posterior analysis, involving prediction of sensing outcomes, inference to update probabilistic models, optimization of decision-making, and integration across potential measures. Each aspect of this problem can be computationally challenging in itself, and thus VoI evaluation and optimization can be intractable in general. Previous work involving VoI assessment in infrastructures has often restricted the class of problems examined to those with discrete variables and observations and organized system models in such a way as to allow for more efficient VoI evaluation, e.g. [18,19]. Alternatively, VoI has been used as the basis for a heuristic approach to inspection optimization in distributed systems by [20]. This paper expands on these previous results by investigating spatially distributed continuous-valued variables and developing an approach to VoI-based sensor placement optimization in distributed systems.

Following [21], in this paper, we investigate VoI and its computational complexity in the context of Gaussian static random field models of spatially distributed systems, under one-stage decision making. Here “static” means that variable values do not change between information collection and decision-making, and “random” means that these values are uncertain. We investigate a set of assumptions on the problem structure under which VoI can be efficiently computed. Where these assumptions do not apply, we make use of approximate techniques for evaluating VoI. This paper extends the results of [21] to a more general set of problems, including those with non-decomposable loss functions and investigates the computational complexity of VoI evaluation for these loss functions. The paper begins with a general statement of the problem of sensor placement in distributed systems, including an overview of spatial random field models, the VoI metric, and the greedy algorithm for optimizing sensor placement in Section 2. In Section 3, we describe methods for tractably computing VoI using Gaussian process random field models and specific loss functions. Section 4 presents numerical investigations of the VoI metric, as well as demonstrative applications of this metric to optimal sensor placement. General discussions and conclusions regarding VoI and sensor placement are given in Section 5.

2. Pre-posterior analysis for sensor placement

2.1. Random field models of spatially distributed phenomena

Define Ω_X to be the spatial domain over which the system of interest is distributed. Vector $\mathbf{x}$ identifies a point within this domain. For computational purposes, the continuous domain is discretized at a set of m points $\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_m\}$. These coordinates can in general represent any discretization of the field, e.g. a set of grid points over a two-dimensional region, or the coordinates of the components of the infrastructure system.

Variables affecting the system are described by model F of a spatially distributed random field f, according to prior probabilistic distribution p_F [22]:

$$f(\mathbf{x})\sim p_F \tag{1}$$

In general, $f(\mathbf{x})$ may be multivariate, describing the values of multiple co-located random fields for different features of interest. Vector $\mathbf{f}$ denotes the random quantities at discrete locations $\mathbf{X}$, concatenating the variables for co-located random fields into this single vector for notational simplicity.

2.2. Observation and inference

Sensors measure subsets of random fields F, potentially with noise. The set of these measurements is denoted as Y, and a specific observation of the set as $\mathbf{y}$. While in many cases observations are made at discrete points, generally observations are of any random variables within F, as well as sets or functions of these variables. For a proposed set of measurements Y, the probabilistic distribution over potential observations conditional to the random field values $\mathbf{f}$ is:

$$\mathbf{y}\,|\,\mathbf{f}\sim p_{Y|\mathbf{f}} \tag{2}$$

The prior distribution over observations can be defined using the law of total probability:

$$p_Y=\int p_{Y|\mathbf{f}}\,p_F(\mathbf{f})\,d\mathbf{f} \tag{3}$$

Bayesian inference allows observation $\mathbf{y}$ to define a posterior probability distribution over the random fields conditional to this observation:

$$\mathbf{f}\,|\,\mathbf{y}\sim p_{F|\mathbf{y}}\propto p_{Y|\mathbf{f}}\,p_F \tag{4}$$

2.3. Actions and losses

Denote by A the set of actions which the manager of an infrastructure system can take. Vector $\mathbf{a}$ refers to the set of chosen