University of Thessaly

School of Engineering

Department of Mechanical Engineering

BAYESIAN OPTIMAL SENSOR PLACEMENT FOR IRREGULARITY IDENTIFICATION IN DYNAMICAL SYSTEMS

Systems Dynamics Laboratory

Diploma Thesis of

Konstantina Ntarladima

Thesis supervisor: **Prof. Costas Papadimitriou**

Submitted in partial fulfillment of the requirements for the Mechanical Engineering Diploma from University of Thessaly

Volos, June 2019

Thesis Committee

Prof. Costas Papadimitriou

Professor of Structural Dynamics, Department of Mechanical Engineering, University of Thessaly

Prof. Georgios Kozanidis

Professor of Optimization Methods of Production/Service Systems, Department of Mechanical Engineering, University of Thessaly

Prof. Alexis Kermanidis

Professor of Mechanical Behavior of Metallic Materials, Department of Mechanical Engineering, University of Thessaly

Acknowledgments

For the development of my diploma thesis I would like to express my deep gratitude to my supervisor Prof. Costas Papadimitriou for his invaluable guidance throughout my learning process.

I would also like to thank the PhD students Tulay Ercan and Xinyu Jia for their useful advice during the performance of this task.

Abstract

A Bayesian framework for solving two Optimal Sensor Placement problems is the cornerstone of the current thesis. Firstly, the question of detecting a disturbance source in a continuous medium using arrival times of the disturbance at the sensor locations when travelling through the medium is posed. Next, the spread of an infection over the population of a number of highly interacting communities is illustrated by an epidemiology network and an abstract mathematical model is displayed to simulate its behavior. Having discussed both systems dynamics a common Bayesian approach for finding the optimal sensor placement to identify the origins of the irregularities is presented. The approach is based on minimizing an information entropy of the posterior distribution of the model parameters derived from a Bayesian learning methodology. For the continuous optimization problem of locating sensors on an elastic medium a stochastic algorithm, CMA-ES is used. For the discrete optimization problem concerning the epidemiology network heuristic forward and backward Sequential Sensor Placement algorithms are implemented. The optimal sensor locations given a fixed number of sensors are derived for both cases. In addition, by monitoring the information entropy as a function of the number of sensors placed at their optimal locations, one can estimate the optimal number of sensors to be used, since additional sensors do not provide any gain in information. The demonstrated results, which are sufficiently discussed, claim the validity of our methodologies.

Table of Contents

CHAPTER 1

Introduction

1.1 Problem Definition and Motivation

The issue of irregularity detection has received much attention over the recent years. When it comes to fault detection in a construction an efficient and timely localization of the fault origin can play a significant role both for safety and economic reasons. Early fault identification reduces damage potential and decreases maintenance costs. Especially, dealing with such a general problem as the one of crack identification in a continuous medium a first contact with the subject is obtained and foundations for further study are built.

When the concerned issue is related to human health things get even more serious. While epidemics eradication is one of the most vital issues of contemporary world, as it is directly linked to human health and quality of life, numerous interdisciplinary studies are being carried out in this attempt. In this sense, the factor that information about the spread of a disease would greatly assist targeted intervention strategies and disease forecasting can play a significant role. Thus modern epidemics studies are trying to introduce real-world, noisy data into abstract mathematical models in order to derive reliable estimations about the epidemic characteristics. Such data can be the number of symptomatic patients admitted to hospital by location and date.

In practice, the data are used to estimate model parameters and their uncertainties. For instance, predicting next week's state of an epidemic requires accurate knowledge of how quickly the disease spreads, which can be inferred processing real data.

At the same time, the problem of making robust predictions for parameter values and disease spread can become computationally very expensive depending on the complexity of the mathematical model which is used. To deal with this challenge we apply a Bayesian inference framework to an epidemic network designed in a manner to simulate a real-world problem and to make as simple as possible the computational procedure.

In general, a network can be defined as a graph that represents relations between discrete objects. These objects consist its nodes and they communicate through edges. In our case, the utilized network is a mathematical construct which models a collection of N distinct

communities. Each node is a distinct population and edges, that is the connections between communities, are a convenient framework to dictate transfer between populations.

As the most important part of the current thesis discusses the behavior of a graph consisted from nodes and edges one can say that this study lies in the broader field of Network Science and could be related with a wide range of scientific investigations into structures that can be illustrated with a similar layout. Consequently, although an epidemiology network is considered, the developed methodologies can be applied to other engineering and science networks.

The profound importance of network analysis can be clearly denoted from the variety of the disciplines in which it has application. To cite few instances, engineering networks include transportation networks, water and gas distribution networks and electricity networks. Furthermore science networks include epidemiology, biological such as blood flow network and respiratory networks, and Social Networks. To be more specific, characteristic applications of network theory include logistical networks, the World Wide Web, Internet, gene regulatory networks, metabolic networks and epistemological networks. So nodes and edges can depict relationships between social entities, that is, persons or organizations when we have to do with a social network analysis, while they can readily represent the interactions between physiological systems like brain, heart or eyes when we talk about a physiological network.

1.2 Literature review

As far as the problem of placing sensors in a continuous medium in order to detect the fault source is concerned the utilized methodologies have been already applied in previous works. The implementation of Bayesian approach in estimating acoustic emission sources has been discussed in Sen et al. work [1]. A Bayesian framework that takes into account the uncertainty of system parameters caused from measurement noise and model error is applied to estimate the origin of acoustic waves. The simplicity and efficiency of their techniques encouraged us to adopt some of their methodologies.

Another Bayesian framework have been implemented to find the optimal sensor location for detecting a crack origin in a structure using strain measurements [2]. The finite elements methodology applied at the current thesis have been previously proposed in the work by Papadimitriou and Papadimitriou [3]. The work is dealing with the problem of deriving the

optimal way of sensor placement for the estimation of turbulence model parameters in Computational Fluid Dynamics.

Furthermore, several studies have been carried out in the field of epidemics' spread forecasting. In particular, mathematical models have been proposed for predicting the spread of epidemics in real-world populations allowing for the development of policies for effectively managing disease spread through organized intervention. As for the extracted results that represent the parameters' estimation and the epidemic's spread prediction, robustness and reliability are of great essence, especially if we consider their clinical and societal impact. These two factors are highly depended on the mathematical model which is utilized. Two important related approaches are detailed bellow.

In the first place, the most popular approach is the one of Kermack and McKendrick's compartmental SIR model and its extensions, such as SIRS and SEIR [4],[5]. This is a mathematical model which decomposes the population into groups considering their relation to the disease. For example the susceptible people consist one group, while the recovered are another. It's a differential equation model which represents the interaction of these groups over time evolution. This model's simplicity and efficiency for some common diseases have been the reasons of its wide use.

Another work has considered topological factors as well, such as demography, land use and climate change [6]. Such mathematical models, incorporating in their frameworks a more holistic approach of diseases dynamics, have marked a wide range of applications. Specifically, they have been used in the research for the confrontation of several historical and modern epidemics, including HIV [7],[8] and malaria [9], [10], [11].

1.3 Organization of the thesis

The first part of this thesis deals with a continuous medium in which a fault starts to occur. The initial objective is the localization of the disturbance source. The applied methodology includes the placement of sensors in the examined area in order to record informative measurements. The measured quantity is the time in which an emitted signal travels from the fault source to the sensors. In the second place a methodology is developed for the optimal sensor placement over the medium for identification of the disturbance origin. A grid of points is used to divide the area of the initial fault and the stochastic optimization method of CMA-ES is used.

The problem is positioned in such a general way that it can depict a wide variety of physical systems. In particular the developed layout can readily illustrate a thin plate with a crack. Moreover a geographic area where an earthquake is recorded can also be considered. In that case the location of the seismic source is to be figured out. The proportion between the two cases can get clear if earthquakes physics is taken into consideration. In addition to that we have to remark the suitability of the utilization of the measurements described above in both problems. On the one hand, ultrasonic guided-waves are used to detect cracks in continuous materials. On the other hand, records made of seismic waves created when an earthquake occurs are used for the estimation of its epicenter.

Furthermore, in the current approach an epidemic model is simulated. Interactions between two distinct populations and the communities included in each of them are simulated via appropriate adjacent matrices. A system of differential equations which relates the time evolution of the number of healthy and infected people in each community is solved numerically through a software. In that way we extract the number of infected people in each community at every time point. We also develop the formulation to describe analytically the gradients of the healthy and infected people with respect to model parameters. Such gradients are needed in Bayesian inference and Optimal Sensor Placement process.

As far as our Bayesian Analysis is concerned, Optimal Sensor Placement is occurred through an appropriately constructed function which quantifies the information gained from the data. Inputs of this function consist of our uncertain parameters and the location of utilized sensors. The minimization of the aforementioned function takes place through two alternative heuristic methods. These are the Forward Sequential Sensor Placement and the Backward Sequential Sensor Placement. The derived results are compared in order to state the reliability of our approach.

The fluctuation of the uncertain parameters is simulated by two sample methods. Firstly, a sparse grid point method is implemented. Furthermore, a Markov Chain Monte Carlo algorithm is applied to generate a larger number of samples which are utilized on multiples program’s runs to extract reliable averages of the values of our interest.

CHAPTER 2

Source identification in a continuous medium

2.1 System Presentation

In the current chapter a methodology developed for detecting the origin of a disturbance is presented. A continuous homogeneous elastic medium is considered. At a time t_0 a structural disturbance occurs at a specific point of its domain. We presume that the disorder starts at a point $\underline{x} = (x_0, y_0, z_0)$ of its area. The objective is to identify the disturbance location and its initial time using measured data.

In order to extract this data, sensors are placed at specific points $\underline{s}_k = (x_k, y_k, z_k)$ of the three dimensional domain. The measured quantity is the time at which the disturbance arrives at each of the placed sensors. Let's call this time $\hat{\underline{t}}_k$. Additionally we assume the speed of signal's distribution to be v. Because of the assumption of material's homogeneity the distribution speed appears to have the same value in all directions.

It is clear that a signal that travels with a constant and even speed through all directions having started at time t_0 at the point $\underline{x} = (x_0, y_0, z_0)$ arrives at a point $\underline{s}_k = (x_k, y_k, z_k)$ at the time given by the following equation.

$$\underline{t} = t_0 + \frac{1}{v}\sqrt{(\underline{x}_s - x_0)^2 + (\underline{y}_s - y_0)^2 + (\underline{z}_s - z_0)^2} \qquad (1)$$

which is the expected time at which the signal reaches the placed sensors.

Let's define a vector $\underline{\theta}$ of the parameters that describe the disturbance location, namely, the $\underline{x} = (x_0, y_0, z_0)$ coordinates of the disturbance and the time t_0 that the disturbance started to take place. Therefore the number of parameters is equal to four and the problem of disturbance identification is equivalent to the problem of estimating $\underline{\theta} = \{x_0, y_0, z_0, t_0\}^T$. One could also include the speed v in the parameter set.

It can be thought that the parameters of our interest can be computed by solving the system of the nonlinear equations given by the relation (1) for multiple sensor points. Concerning the 2 dimensional space and that the time t_0 is known the problem can be visually illustrated as shown in Figure 1. It can be seen that one requires at least 3 sensors in order to identify the location of the disturbance.

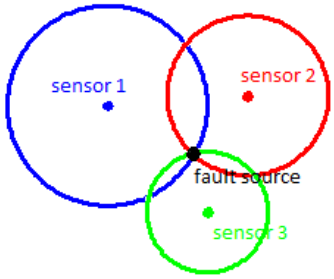


Figure 1: Three sensors used to detect a fault source

The disadvantage of this deterministic approach is that it doesn't take into account the uncertainties caused by measurements' error and the inadequacy of our model to accurately depict the real problem. Specifically, in the problem of the signal which travels throw a material, the speed cannot be determined by a specific value, while it has a fluctuation over its space and according to the direction of movement. Moreover regarding seismic waves transportation, different speeds occur in a rocky area and in a sandy one. To incorporate this effect a Bayesian approach is implemented.

2.2 Implementation of Bayesian Theory in Disturbance Source Identification

A Bayesian parameter estimation framework is used to estimate the values in $\underline{\theta}$ and their associated uncertainties, using information from the data set $D(\underline{d}) = \{\underline{\hat{g}}(\underline{d})\} = \{\hat{g}_k(\underline{d}), k = 1,\ldots,N\} = \{\hat{t}_k(\underline{s}_k), k = 1,\ldots,N\}$ that consists of N measurements inferred as described above. $\underline{d}$ is a matrix consisted of the coordinates of the located sensors. Specifically $\underline{d}$ is taken in the form $\underline{d} = \begin{bmatrix} \underline{s}_1^{\mathrm{T}} \\ \underline{s}_2^{\mathrm{T}} \\ \vdots \\ \underline{s}_N^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ & \vdots & \\ x_N & y_N & z_N \end{bmatrix}$

Let *M* represent a model parametrized by $\underline{\theta}$, simulating the behavior of the propagation of the disturbance in the medium. This model predicts the times $\underline{g}(\underline{d},\underline{\theta})$ at the locations $\underline{s}_k = (x_k, y_k, z_k)$ for given value of the parameter set $\underline{\theta}$, according to the equation (1). In attempt to proceed with the Bayesian formulation for parameter estimation, one introduces the prediction error equation:

$$\underline{\hat{g}}(\underline{d}) = \underline{g}(\underline{d},\underline{\theta}) + \underline{e}(\underline{d}) \tag{2}$$

or equivalently,

$$\underline{\hat{t}}(\underline{d}) = \underline{t}(\underline{d},\underline{\theta}) + e(\underline{d}) \tag{3}$$

where $e(\underline{d})$ is the prediction error vector that describes the discrepancy between the model predictions and experimental data. The prediction error is due to modeling errors and measurement noise and is commonly treated as a Gaussian zero-mean random vector with covariance matrix Σ. Herein Σ is assumed to be a diagonal covariance matrix given as

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathrm{I} \tag{4}$$

Thus Σ is a N by N matrix, where N is the number of measurements.

According to the Bayesian system identification methodology, the values of $\underline{\theta}$ are modeled by a Probability Density Function (PDF) that quantifies the plausibility of each possible value, given the data set $\underline{\hat{g}}(\underline{d})$. From the Bayesian theorem, this PDF is the posterior PDF given by the following:

$$p\left(\underline{\theta}\middle|\underline{\hat{g}}(\underline{d})\right) = \frac{p\left(\underline{\hat{g}}(\underline{d})|\underline{d},\underline{\theta}\right)\pi(\underline{\theta})}{p\left(\underline{\hat{g}}(\underline{d})\right)} \tag{5}$$

where $p\left(\underline{\hat{g}}(\underline{d})|\underline{d},\underline{\theta}\right)$ is the likelihood and $\pi(\underline{\theta})$ is the prior PDF of the parameter set $\underline{\theta}$. The evidence term $p\left(\underline{\hat{g}}(\underline{d})\right)$ is a constant, independent of $\underline{\theta}$. Therefore, the posterior PDF can be written as follows:

$$p\left(\underline{\theta}\middle|\underline{\hat{g}}(\underline{d})\right) = cp\left(\underline{\hat{g}}(\underline{d})|\underline{d},\underline{\theta}\right)\pi(\underline{\theta}) = cN\left(\underline{\hat{g}}(\underline{d})\middle|\underline{g}(\underline{d},\underline{\theta}),\sigma^2 I\right)\pi(\underline{\theta}) \tag{6}$$

$$= c\exp\left(-\frac{1}{2}\left(\underline{\hat{g}}(\underline{d}) - \underline{g}(\underline{d},\underline{\theta})\right)^T \Sigma^{-1}(\sigma)\left(\underline{\hat{g}}(\underline{d}) - \underline{g}(\underline{d},\underline{\theta})\right)\right)\pi(\underline{\theta}) \tag{7}$$

where c is a constant selected such that the posterior PDF integrates to 1 and the notation $N\left(\underline{x}|\mu,\Sigma\right)$ denotes that the random vector $\underline{x}$ follows the multivariate normal distribution with mean vector $\underline{\mu}$ and covariance matrix Σ. We see that the model evaluation enters the formulation through the likelihood function. It is clear that we can reach our objective to

estimate $\underline{\theta}$ by finding the value $\underline{\theta}_0$ that maximizes the posterior PDF or, equivalently, minimizes the minus log of the posterior PDF given by

$$L(\underline{\theta}) = -\ln p\left(\underline{\theta}\middle|\underline{g}\left(\underline{d}\right)\right) = -\ln c + J\left(\underline{\theta}\right) - \ln \pi\left(\underline{\theta}\right) \tag{8}$$

where $J\left(\underline{\theta}\right)$ is the discrepancy between experimental data and model prediction, given by

$$J = \frac{1}{2}\left(\underline{g}\left(\underline{d},\underline{\theta}\right) - \underline{\hat{g}}\left(\underline{d}\right)\right)^T \Sigma^{-1}\left(\underline{g}\left(\underline{d},\underline{\theta}\right) - \underline{\hat{g}}\left(\underline{d}\right)\right) \tag{9}$$

Thus finding the most probable value that maximizes the posterior PDF is equivalent to minimizing the measure of fit $J\left(\underline{\theta}\right)$.

Specifically in our case we have to minimize the function

$$J = \frac{1}{2}\left(\underline{t} - \underline{\hat{t}}\right)^T \Sigma^{-1}\left(\underline{t} - \underline{\hat{t}}\right) \tag{10}$$

For large number of data, the posterior PDF can be approximated by a Gaussian PDF with mean the most probable value and covariance matrix equal to the inverse of the Hessian of the function $L(\underline{\theta})$ evaluated at the most probable value.

The model predictions which correspond to arrival times at the sensor point due to a disturbance at location $\underline{x} = \left(x_0, y_0, z_0\right)$ are evaluated as follows. The arrival time $\underline{t}\left(\underline{\theta}\right)$ has to be calculated at the k -th sensor point as follows

$$t_k = t_0 + \frac{1}{v}\sqrt{\left(x_{sk} - x_0\right)^2 + \left(y_{sk} - y_0\right)^2 + \left(z_{sk} - z_0\right)^2} \tag{11}$$

where $k = 1,\ldots,N_s$ and N_s is the number of sensors.

Elaborating the function (10) using the diagonal covariance matrix we get:

$$J = \frac{1}{2\sigma^2}\sum_{k=1}^{N_s}\left(t_0 - \hat{t}_k + \left(\frac{1}{v}\sqrt{\left(x_{sk} - x_0\right)^2 + \left(y_{sk} - y_0\right)^2 + \left(z_{sk} - z_0\right)^2}\right)\right)^2 \tag{12}$$

As mentioned above, an optimization process is occurred. Gradient-based methods are used. The gradients of the objective function are needed to be calculated.

The gradient of J which is a 4 by 1 matrix

$$\underline{G} = \left[\frac{\partial J}{\partial x_0} \quad \frac{\partial J}{\partial y_0} \quad \frac{\partial J}{\partial z_0} \quad \frac{\partial J}{\partial t_0}\right]^T \tag{13}$$

Let's move forward by elaborating each element of $\underline{G}$

$$\frac{\partial J}{\partial x_0} = -\sum_{k=1}^{N_s} \frac{(2x_{sk} - 2x_0)\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}\right)}{2\sigma^2 v\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}}$$

$$= -\frac{1}{\sigma^2 v}\sum_{k=1}^{N_s} \frac{1}{s_k}(x_{sk} - x_0)\left(t_0 - \hat{t}_k + \frac{s_k}{v}\right)$$

where $s_k = \sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}$

Because of the symmetry of function J, the gradient of J with respect of y_0 and z_0 is given by replacing x_0, x_{sk} by y_0, y_{sk} and z_0, z_{sk}, respectively.

$$\frac{\partial J}{\partial y_0} = -\sum_{k=1}^{N_s} \frac{(2y_{sk} - 2y_0)\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}\right)}{2\sigma^2 v\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}}$$

$$= -\frac{1}{\sigma^2 v}\sum_{k=1}^{N_s} \frac{1}{s_k}(y_{sk} - y_0)\left(t_0 - \hat{t}_k + \frac{s_k}{v}\right)$$

$$\frac{\partial J}{\partial z_0} = -\sum_{k=1}^{N_s} \frac{(2z_{sk} - 2z_0)\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}\right)}{2\sigma^2 v\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}}$$

$$= -\frac{1}{\sigma^2 v}\sum_{k=1}^{N_s} \frac{1}{s_k}(z_{sk} - z_0)\left(t_0 - \hat{t}_k + \frac{s_k}{v}\right)$$

Finally, the gradient of J with respect to time t_0 can be written as follows

$$\frac{\partial J}{\partial t_0} = \sum_{k=1}^{N_s} \frac{\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{(x_{sk}-x_0)^2 + (y_{sk}-y_0)^2 + (z_{sk}-z_0)^2}\right)}{\sigma^2 v} = \frac{N_s}{\sigma^2 v}\left[t_0 - \bar{t} + \frac{\bar{s}}{v}\right]$$

where $\bar{t} = \frac{1}{N_s}\sum_{k=1}^{N_s} t_k$ and $\bar{s} = \frac{1}{N_s}\sum_{k=1}^{N_s} s_k$.

The covariance of our parameters $\theta = \{x_0, y_0, z_0, t_0\}^T$ can be computed as the square root of the inverse of the Hessian matrix. So the calculation of Hessian is necessary.

The Hessian of the function J is a 4 by 4 symmetric matrix given by

$$H=\begin{bmatrix} \frac{\partial^2 J}{\partial {x_0}^2} & \frac{\partial^2 J}{\partial x_0 \partial y_0} & \frac{\partial^2 J}{\partial x_0 \partial z_0} & \frac{\partial^2 J}{\partial x_0 \partial t_0} \\ \frac{\partial^2 J}{\partial y_0 \partial x_0} & \frac{\partial^2 J}{\partial {y_0}^2} & \frac{\partial^2 J}{\partial y_0 \partial z_0} & \frac{\partial^2 J}{\partial y_0 \partial t_0} \\ \frac{\partial^2 J}{\partial z_0 \partial x_0} & \frac{\partial^2 J}{\partial z_0 \partial y_0} & \frac{\partial^2 J}{\partial {z_0}^2} & \frac{\partial^2 J}{\partial z_0 \partial t_0} \\ \frac{\partial^2 J}{\partial t_0 \partial x_0} & \frac{\partial^2 J}{\partial t_0 \partial y_0} & \frac{\partial^2 J}{\partial t_0 \partial z_0} & \frac{\partial^2 J}{\partial {t_0}^2} \end{bmatrix}$$

Let's continue by elaborating its terms.

$$\frac{\partial^2 J}{\partial {x_0}^2}=\sum_{k=1}^{N_s}\left[\begin{array}{l} \frac{\left(t_0-\hat{t}_k+\frac{1}{v}\sqrt{\left(x_{sk}-x_0\right)^2+\left(y_{sk}-y_0\right)^2+\left(z_{sk}-z_0\right)^2}\right)}{\sigma^2 v\sqrt{\left(\left(x_{sk}-x_0\right)^2+\left(y_{sk}-y_0\right)^2+\left(z_{sk}-z_0\right)^2\right)}}+ \\ \frac{\left(2x_{sk}-2x_0\right)\left(x_{sk}-x_0\right)}{2\sigma^2 v^2\left(\left(x_{sk}-x_0\right)^2+\left(y_{sk}-y_0\right)^2+\left(z_{sk}-z_0\right)^2\right)}+ \\ \frac{\left(2x_{sk}-2x_0\right)\left(x_{sk}-x_0\right)\left(t_0-\hat{t}_k+\frac{1}{v}\sqrt{\left(x_{sk}-x_0\right)^2+\left(y_{sk}-y_0\right)^2+\left(z_{sk}-z_0\right)^2}\right)}{\sigma^2 v\left(\left(x_{sk}-x_0\right)^2+\left(y_{sk}-y_0\right)^2+\left(z_{sk}-z_0\right)^2\right)^{3/2}} \end{array}\right]$$

$$=\frac{1}{\sigma^2 v}\sum_{k=1}^{N_s}\left[\frac{t_0-\hat{t}_k+\frac{s_k}{v}}{s_k}+\frac{\left(x_{sk}-x_0\right)^2}{v s_k^2}+\frac{\left(x_{sk}-x_0\right)^2\left(t_0-\hat{t}_k+\frac{1}{v}s_k\right)}{{s_k}^3}\right]$$

Evidently the second derivative of the objective function with respect to y_0 and z_0 results from the relation above if we replace x_0, x_{sk} by y_0, y_{sk} and z_0, z_{sk} respectively.

$$\frac{\partial^2 J}{\partial {y_0}^2} = \sum_{k=1}^{N_s} \left[\begin{array}{c} \frac{\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2} \right)}{\sigma^2 v \sqrt{\left(\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2\right)}} + \\ \frac{\left(2y_{sk}-2y_0\right)\left(y_{sk}-y_0\right)}{2\sigma^2 v^2 \left(\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2\right)} + \\ \frac{\left(2y_{sk}-2y_0\right)\left(y_{sk}-y_0\right)\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2} \right)}{\sigma^2 v \left(\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2\right)^{3/2}} \end{array} \right]$$

$$= \frac{1}{\sigma^2 v} \sum_{k=1}^{N_s} \left[\frac{t_0 - \hat{t}_k + \frac{s_k}{v}}{s_k} + \frac{\left(y_{sk}-y_0\right)^2}{v s_k^2} + \frac{\left(y_{sk}-y_0\right)^2 \left(t_0 - \hat{t}_k + \frac{1}{v} s_k \right)}{{s_k}^3} \right]$$

$$\frac{\partial^2 J}{\partial {z_0}^2} = \sum_{k=1}^{N_s} \left[\begin{array}{c} \frac{\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2} \right)}{\sigma^2 v \sqrt{\left(\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2\right)}} + \\ \frac{\left(2z_{sk}-2z_0\right)\left(z_{sk}-z_0\right)}{2\sigma^2 v^2 \left(\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2\right)} + \\ \frac{\left(2z_{sk}-2z_0\right)\left(z_{sk}-z_0\right)\left(t_0 - \hat{t}_k + \frac{1}{v}\sqrt{\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2} \right)}{\sigma^2 v \left(\left(x_{sk}-x_0\right)^2 + \left(y_{sk}-y_0\right)^2 + \left(z_{sk}-z_0\right)^2\right)^{3/2}} \end{array} \right]$$

$$= \frac{1}{\sigma^2 v} \sum_{k=1}^{N_s} \left[\frac{t_0 - \hat{t}_k + \frac{s_k}{v}}{s_k} + \frac{\left(z_{sk}-z_0\right)^2}{v s_k^2} + \frac{\left(z_{sk}-z_0\right)^2 \left(t_0 - \hat{t}_k + \frac{1}{v} s_k \right)}{{s_k}^3} \right]$$

The second derivative of the objective function with respect to t_0 appears to have a constant value independent of variables $\theta = \left\{x_0, y_0, z_0, t_0\right\}^T$.

$$\frac{\partial^2 J}{\partial {t_0}^2} = \sum_{k=1}^{N_s} \frac{1}{\sigma^2 v} = \frac{N_s}{\sigma^2 v}$$

$$\frac{\partial^2 J}{\partial y_0 x_0}=\frac{\partial^2 J}{\partial x_0 y_0}=\sum_{k=1}^{N_s}\left[\begin{array}{l}\dfrac{(2y_{sk}-2y_0)(x_{sk}-x_0)}{2\sigma^2\nu^2\left((x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2\right)}+\\ \dfrac{(2y_{sk}-2y_0)(x_{sk}-x_0)\left(t_0-\hat{t}_k+\dfrac{1}{\nu}\sqrt{(x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2}\right)}{\sigma^2\nu\left((x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2\right)^{3/2}}\end{array}\right]$$

$$=\frac{1}{\sigma^2\nu}\sum_{k=1}^{N_s}\left[\frac{(y_{sk}-y_0)(x_{sk}-x_0)}{\nu s_k^2}+\frac{2(y_{sk}-y_0)(x_{sk}-x_0)\left(t_0-\hat{t}_k+\dfrac{s_k}{\nu}\right)}{s_k^3}\right]$$

$$\frac{\partial^2 J}{\partial z_0 x_0}=\frac{\partial^2 J}{\partial x_0 z_0}=\sum_{k=1}^{N_s}\left[\begin{array}{l}\dfrac{(2z_{sk}-2z_0)(x_{sk}-x_0)}{2\sigma^2\nu^2\left((x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2\right)}+\\ \dfrac{(2z_{sk}-2z_0)(x_{sk}-x_0)\left(t_0-\hat{t}_k+\dfrac{1}{\nu}\sqrt{(x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2}\right)}{\sigma^2\nu\left((x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2\right)^{3/2}}\end{array}\right]$$

$$=\frac{1}{\sigma^2\nu}\sum_{k=1}^{N_s}\left[\frac{(z_{sk}-z_0)(x_{sk}-x_0)}{\nu s_k^2}+\frac{2(z_{sk}-z_0)(x_{sk}-x_0)\left(t_0-\hat{t}_k+\dfrac{s_k}{\nu}\right)}{s_k^3}\right]$$

$$\frac{\partial^2 J}{\partial t_0 x_0}=\frac{\partial^2 J}{\partial x_0 t_0}=-\sum_{k=1}^{N_s}\frac{(x_{sk}-x_0)}{\sigma^2\nu^2\sqrt{(x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2}}=-\frac{1}{\sigma^2\nu^2}\sum_{k=1}^{N_s}\frac{x_{sk}-x_0}{s_k}$$

$$\frac{\partial^2 J}{\partial z_0 y_0}=\frac{\partial^2 J}{\partial y_0 z_0}=\sum_{k=1}^{N_s}\left[\begin{array}{l}\dfrac{(2z_{sk}-2z_0)(y_{sk}-y_0)}{2\sigma^2\nu^2\left((x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2\right)}+\\ \dfrac{(2z_{sk}-2z_0)(y_{sk}-y_0)\left(t_0-\hat{t}_k+\dfrac{1}{\nu}\sqrt{(x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2}\right)}{\sigma^2\nu\left((x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2\right)^{3/2}}\end{array}\right]$$

$$=\frac{1}{\sigma^2\nu}\sum_{k=1}^{N_s}\left[\frac{(z_{sk}-z_0)(y_{sk}-y_0)}{\nu s_k^2}+\frac{2(z_{sk}-z_0)(y_{sk}-y_0)\left(t_0-\hat{t}_k+\dfrac{s_k}{\nu}\right)}{s_k^3}\right]$$

$$\frac{\partial^2 J}{\partial t_0 y_0}=\frac{\partial^2 J}{\partial y_0 t_0}=-\sum_{k=1}^{N_s}\frac{(y_{sk}-y_0)}{\sigma^2\nu^2\sqrt{(x_{sk}-x_0)^2+(y_{sk}-y_0)^2+(z_{sk}-z_0)^2}}=-\frac{1}{\sigma^2\nu^2}\sum_{k=1}^{N_s}\frac{(y_{sk}-y_0)}{s_k}$$

$$\frac{\partial^2 J}{\partial t_0 z_0} = \frac{\partial^2 J}{\partial z_0 t_0} = -\sum_{k=1}^{N_s} \frac{\left(z_{sk} - z_0\right)}{\sigma^2 v^2 \sqrt{\left(x_{sk} - x_0\right)^2 + \left(y_{sk} - y_0\right)^2 + \left(z_{sk} - z_0\right)^2}} = -\frac{1}{\sigma^2 v^2} \sum_{k=1}^{N_s} \frac{\left(z_{sk} - z_0\right)}{s_k}$$

2.3 Applications

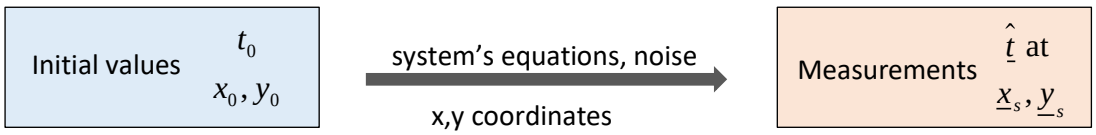
We assume that we have a plane elastic medium. A disturbance starts at $\underline{x} = (4,1)$ at time $t_0 = 3$. To begin with, we generate a sample of measurements. Taking into account the initial point and time, we calculate the time at which the signal arrives at the sensor locations assuming that the signal's velocity has a fluctuation around its mean. That takes place by adding at the mean value a random number generated by a normal distribution, with zero mean and standard deviation equal to one, multiplied by a quantity a times the mean value of spread. That is

$$v_k = v + vae_k$$

where, $e_k \square N(0,1)$ and a quantifies the size of the uncertainty. It's value can vary from 0,001 to 0,03.

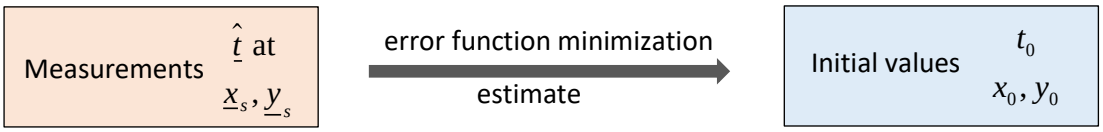
This procedure is occurred for each sensor location. Given the sensor coordinates we calculate the time at which the signal arrives at each sensor. These values consist our sample, t_k .

The generation of our sample can be illustrated schematically as follows:



After that, considering as data the measurements and the coordinates of our sensors and assuming the spread to have a constant value, we compute the time at which the signal is expected to reach the sensors, $\underline{t}$.

Finally, we minimize the objected function elaborated at the previous chapter to extract the initial characteristics of the disturbance $\theta = \{x_0, y_0, t_0\}^T$.



We have three parameters x_0, y_0, t_0 to be estimated from the data. So we begin with placing three sensors on the plane medium. The sensors' coordinates are given as

$$s_1 = (2,3)$$
$$s_2 = (3,4)$$
$$s_3 = (1,5)$$

The result that we take through the procedure described above are

$$\underline{\theta} = \begin{bmatrix} 3.1656 \\ 2.3335 \\ 4.4914 \end{bmatrix}$$

and the covariance matrix appears to be

$$\text{covariance} = \begin{bmatrix} 0.0633 \\ 0.0997 \\ 0.1083 \end{bmatrix}$$

It is clear that the inferred results are not sufficiently precise. So we add one extra sensor on the plane mean with spatial coordinates given bellow

$$s_4 = (6,8)$$

The updated results are

$$\underline{\theta} = \begin{bmatrix} 3.7884 \\ 1.3607 \\ 3.3699 \end{bmatrix}$$

which are much closer to the actual values.

This fact is also defined from the covariance of the parameters, which appear to have smaller values. The new covariance matrix is

$$\text{covariance} = \begin{bmatrix} 0.0210 \\ 0.0726 \\ 0.0672 \end{bmatrix}$$

Finally we place the sensors along a straight line that can be illustrated by the function $x = 2$.

$$s_1 = (2,3)$$
$$s_2 = (2,4)$$
$$s_3 = (2,5)$$
$$s_4 = (2,8)$$

Applying the same methodology the inferred results are

$$\underline{\theta} = \begin{bmatrix} 4.0423 \\ 1.0575 \\ 2.9920 \end{bmatrix}$$

which are very close to the actual values. However the covariance matrix has similar values to the previous case.

$$\text{covariance} = \begin{bmatrix} 0.0244 \\ 0.0837 \\ 0.0783 \end{bmatrix}$$

In conclusion we can point out that 4 sets of measurements are adequate to elicit reliable results.

CHAPTER 3

Simulation of Infection Spread in an Epidemiology Network

3.1 SIR Model Presentation

A Bayesian framework is applied to an epidemiology network in order to infer robust estimations about parameters of the system and to identify the origin of the infection. To deal with this complex and difficult task a mathematical model that illustrates the system's dynamics is used. This is the SIR model. Its workableness is due to the fact that it divides the population into three groups. These are the people who are susceptible to the disease, people who are infected and contagious, the infective group, and those who are neither susceptible nor infected, either because they gained immunity from recovery or due to a vaccine, quarantine policies, or disease-related death, the removed group.

Let us use the notations $S(t)$, $I(t)$, and $R(t)$ to denote the size of the susceptible, infective, and removed groups, respectively, as functions of a continuous time t. The SIR model is based on three main assumptions. Firstly since the timescale on which the disease evolves is assumed to be much shorter than the timescale on which the population may evolve via, e.g., births or natural deaths, the population Y is assumed constant, and so

$$S(t)+I(t)+R(t)=Y \tag{14}$$

over time t. It must be marked here that individuals killed by the disease are considered part of the removed group. Secondly, members of the population are assumed to come into contact uniformly at random and at a constant rate β which is the parameter that governs the rate at which an infection can spread. It is assumed that when a susceptible individual come into contact with an infective the former is immediately infected. If this assumption is not desired, the chance of disease transfer can be incorporated in β. Finally, infective people recover, or are otherwise removed from the infective population because of death for example, at a constant rate γ. These assumptions can be visualized as

$$S \xrightarrow{\beta} I \xrightarrow{\gamma} R$$

It can be said that at a particular time t, $S(t)$ susceptible individuals and $I(t)$ infective come into contact at a rate β, resulting in βIS transitions from susceptible to infective. At the same time, $I(t)$ infective people are removed at a rate γ. It follows that γI transitions from

infective to removed are taking place. A set of ordinary differential equations can be derived from the assumptions above:

$$\frac{dS(t)}{dt} = -\beta IS$$
$$\frac{dI(t)}{dt} = \beta IS - \gamma I \quad (15)$$
$$\frac{dR(t)}{dt} = \gamma I$$

The SIR model is readily generalized to a graph with *N* vertices, a mathematical construct which can be thought of as modeling a collection of *N* distinct communities. In particular, each node is assumed to be a distinct unit of population, community, whose dynamics evolve according to equation (15). Communities are connected through edges. That structure consists a convenient framework to dictate transfer between populations.

Additionally each population itself has three groups the susceptible, the infective and the removed, which implies that three quantities are needed to describe movement. We define the quantities λ_{ij}, η_{ij}, g_{ij} to describe the rates of movement from node i to node j on the susceptible, infective, and removed groups, respectively. In that way each transition rate is identified as the weight of the edge connecting i to j. These quantities can be written as weighted adjacency matrices, where each element depicts the corresponding transition rate. Let's define the three transition matrices as Λ, H and G. The SIR model for a network, namely a system of *N* nodes that communicate to each other can be written as [12]

$$\frac{dS_i(t)}{dt} = -\beta I_i S_i + \sum_{j=1}^{N} \lambda_{j,i} S_j - \sum_{j=1}^{N} \lambda_{i,j} S_i$$
$$\frac{dI_i(t)}{dt} = \beta I_i S_i - \gamma I_i + \sum_{j=1}^{N} \eta_{j,i} I_j - \sum_{j=1}^{N} \eta_{i,j} I_i$$
$$\frac{dR_i(t)}{dt} = \gamma I_i + \sum_{j=1}^{N} g_{j,i} R_j - \sum_{j=1}^{N} g_{i,j} R_i$$

or equivalently in matrix form:

$$\frac{dS}{dt} = -\beta I \cdot S + \Lambda^{\mathrm{T}} S - (\Lambda F) \cdot S$$
$$\frac{dI}{dt} = \beta I \cdot S - \gamma I + H^{\mathrm{T}} I - (HF) \cdot I \quad (16)$$

$$\frac{dR}{dt} = \gamma I + G^{\mathrm{T}} R - (GF) \cdot R$$

where, $F = (1,1,\ldots,1)^T$ is a vector of ones which simplifies the notation. The notation $I \cdot S$ represents a multiplication element by element, while $\Lambda^{\mathrm{T}} S$ is a matrix multiplication.

It should be noticed here that S, I and R are $N \times 1$ vectors whose i^{th} element corresponds to the i^{th} population. It is obvious that in case that there is no movement between populations, namely $\lambda_{ij} = \eta_{ij} = g_{ij} = 0$ for every i, j, each model reduces to the single population model as featured in the system of equations (15).

3.2 Network description

After having illustrated the system dynamics for a general network consisted by *N* nodes interacting each other, we have to examine the behavior of a specified graph. In that sense two distinct populations, each with many highly interacting sub-communities, are modeled. The two populations communicate through a route, modeled by a connecting edge. This constitutes a "barbell graph", that is two complete $m/2$-node graphs connected by an edge, where m is the total number of nodes. The transition matrices Λ, H, G that dictate the communication between nodes are written in the following form

$$\Lambda = \begin{bmatrix} 0 & \lambda_{1,2} & \cdots & \lambda_{1,\frac{m}{2}} & 0 & 0 & \cdots & 0 \\ \lambda_{2,1} & 0 & \cdots & \lambda_{2,\frac{m}{2}} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{\frac{m}{2},1} & \lambda_{\frac{m}{2},2} & \cdots & 0 & \lambda_{\frac{m}{2},\frac{m}{2}+1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \lambda_{\frac{m}{2}+1,\frac{m}{2}} & 0 & \lambda_{\frac{m}{2}+1,\frac{m}{2}+2} & \cdots & \lambda_{\frac{m}{2}+1,m} \\ 0 & 0 & \cdots & 0 & \lambda_{\frac{m}{2}+2,\frac{m}{2}+1} & 0 & \cdots & \lambda_{\frac{m}{2}+2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_{m,\frac{m}{2}+1} & \lambda_{m,\frac{m}{2}+1} & \cdots & 0 \end{bmatrix}$$

$$\mathrm{H}=\begin{bmatrix} 0 & \eta_{1,2} & \cdots & \eta_{1,\frac{m}{2}} & 0 & 0 & \cdots & 0 \\ \eta_{2,1} & 0 & \cdots & \eta_{2,\frac{m}{2}} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \eta_{\frac{m}{2},1} & \eta_{\frac{m}{2},2} & \cdots & 0 & \eta_{\frac{m}{2},\frac{m}{2}+1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \eta_{\frac{m}{2}+1,\frac{m}{2}} & 0 & \eta_{\frac{m}{2}+1,\frac{m}{2}+2} & \cdots & \eta_{\frac{m}{2}+1,m} \\ 0 & 0 & \cdots & 0 & \eta_{\frac{m}{2}+2,\frac{m}{2}+1} & 0 & \cdots & \eta_{\frac{m}{2}+2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \eta_{m,\frac{m}{2}+1} & \eta_{m,\frac{m}{2}+1} & \cdots & 0 \end{bmatrix}$$

$$\mathrm{G}=\begin{bmatrix} 0 & g_{1,2} & \cdots & g_{1,\frac{m}{2}} & 0 & 0 & \cdots & 0 \\ g_{2,1} & 0 & \cdots & g_{2,\frac{m}{2}} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{\frac{m}{2},1} & g_{\frac{m}{2},2} & \cdots & 0 & g_{\frac{m}{2},\frac{m}{2}+1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & g_{\frac{m}{2}+1,\frac{m}{2}} & 0 & g_{\frac{m}{2}+1,\frac{m}{2}+2} & \cdots & g_{\frac{m}{2}+1,m} \\ 0 & 0 & \cdots & 0 & g_{\frac{m}{2}+2,\frac{m}{2}+1} & 0 & \cdots & g_{\frac{m}{2}+2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & g_{m,\frac{m}{2}+1} & g_{m,\frac{m}{2}+1} & \cdots & 0 \end{bmatrix}$$

Elements of the diagonal denoting the transition from the each node to itself, are equal to zero. The elements with indicators $\frac{m}{2}+1,\frac{m}{2}$ and $\frac{m}{2},\frac{m}{2}+1$ depict the transitions between the two populations.

To specify further our problem uniform transition rates between adjacent vertices are assumed. Namely, $\lambda_{ij}=0,02$, $\eta_{ij}=0,3$ and $g_{ij}=0,05$ for every $i,\ j$.

Moreover, we consider two 20-node graphs connected by a single edge. An illustration of our network is given in Figure 2.

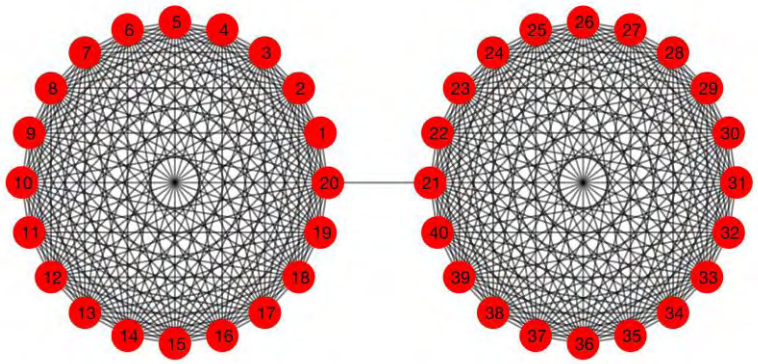


Figure 2: A barbell graph of two populations consisted of 20 nodes

We assume that at the starting point all nodes are fully susceptible with configuration $S_i(0)=100,\ I_i(0)=R_i(0)=0$ apart from one node where the infection begins. We assume that at this node, let's call it a, we have: $S_a(0)=5,\ I_a(0)=95,\ R_a(0)=0$.

After having captured our network, the objective is to find out the time evolution of the susceptible, infective and recovered population. In order to deal with it, the set of ordinary equations stated above is solved through an appropriate software which calculates numerically the values of $S(t),\ I(t),\ R(t)$ for every time step. This software integrates the system of differential equations $\frac{dy_i}{dt} = f(t, y_i)$ from an initial time t_0 to a final t_f with initial conditions $y_i(0)$.

We assume that the infection begins at node 8 which implies that $S_8(0)=5,\ I_8(0)=95,\ R_8(0)=0$ and $S_i(0)=100,\ I_i(0)=R_i(0)=0$ for every other node. It appears that 10 time units is an adequate period of time to observe our system's response after introducing the initial disorder.

Consequently we derive the quantities of susceptible, infective and removed group as a function of time evolution. Related plots are displayed in Figures 3 to 5 for four nodes of our network.

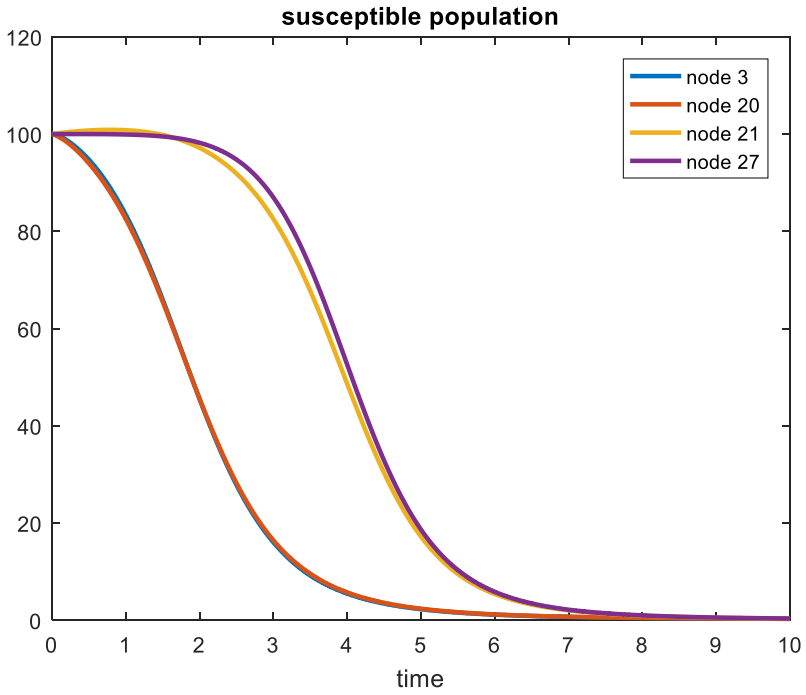


Figure 3: Time evolution of susceptible population for nodes 3, 20, 21 and 27

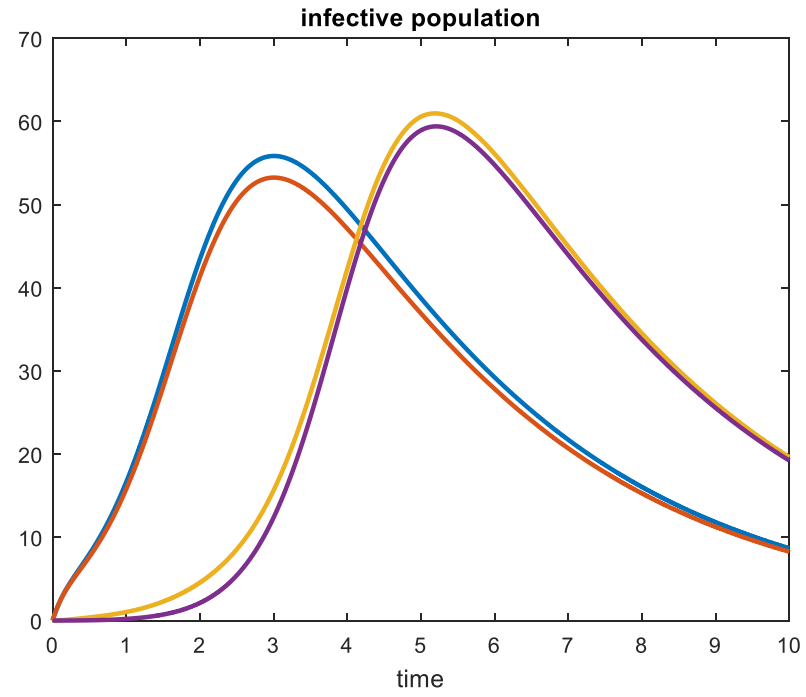


Figure 4: Time evolution of infective population for nodes 3, 20, 21 and 27

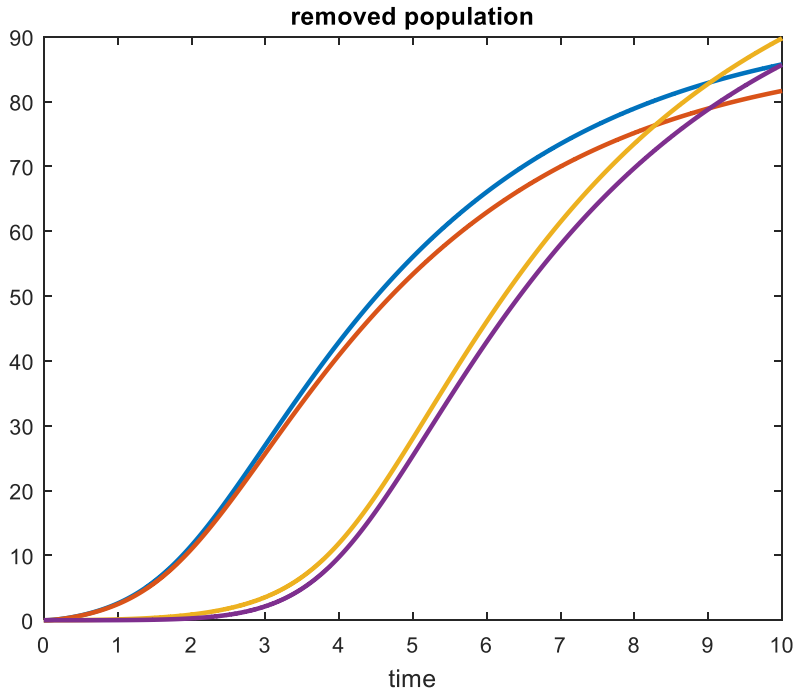


Figure 5: Time evolution of removed population for nodes 3, 20, 21 and 27

CHAPTER 4

Bayesian Optimal Sensor Placement Methodology

4.1 Method Presentation

4.1.1 Optimal Sensor Placement for Estimation of Model Parameters

The optimal sensor placing for the estimation of a problem parameters, for instance β and γ for the epidemiology network or the place of an irregularity source (parameters x_0, y_0, t_0), is a part of wider field of research the one of optimal experimental design. In general, when an experiment for a system is designed, it is of primary essence that the measured data provide sufficient information about the condition of the system. In that purpose it is important to optimize certain characteristics of the experimental set up.

Having defined system's characteristics and formed a mathematical model to represent its behavior an optimal experimental design allows us to opt for appropriate control parameters of the experiment. Appropriate experimental design is the one that provide us with the most informative data for estimating our mathematical model parameters with the least uncertainty. The control parameters of an experiment that are to be optimized can include the excitation characteristics, e.g. frequency content, amplitude and sampling frequency, the type of the outputs of sensors, e.g. displacement, acceleration and force, the location and number of sensors and the characteristics of the executed measurements, e.g. monitoring period and sampling frequency.

Especially, in the context of the network epidemic model, this subject is of considerable practical interest while the mathematical model that describes the system dynamics leans on logical assumptions and not on proved physical laws. In addition to that, observations, specifically the number of infected patients at a particular set of community health centers, are inherently noisy due to sampling effects. Optimal experimental design should compensate for that issues.

In this chapter we are particularly interested in obtaining the following objective. Given a system and its parameterized mathematical model, select the optimal sensor configuration, namely the number and location of sensors, so that the measured data obtained from the sensor system are most informative for estimating the parameters of a mathematical model of the system. In model parameter estimation, most informative measurements are those

that result in the least uncertainty in the parameters' estimations. This uncertainty is quantified by information entropy. Using information entropy as a unique scalar measure of the uncertainty in a parameter set $\underline{\theta}$, the problem of optimizing the number and location of sensors can be mathematically expressed as a problem of minimization of the information entropy of the posterior distribution of the model parameters.

The theoretical background of the described procedure is demonstrated next.

4.1.2 Bayesian Analysis

A Bayesian framework was developed in section 2.2 of the current thesis for the identification of the source of a disturbance started in a continuous mean. Similarly, the applied methodology can be used in any mathematical model in order to estimate the parameters of our interest based on experimental data. To give an abstract illustration of Bayesian Theory on which optimal sensor placement methodology is based, we will next outline the theory.

The parameters to be estimated are written in a vector $\underline{\theta} \in \mathbf{R}^{N_\theta}$. Measurements which take place at locations $\underline{\delta}$, producing a set of measured data $\underline{d} \equiv \underline{d}(\underline{\delta}) \in \mathbf{R}^N$. The location vector $\underline{\delta}$ contains the coordinates of the sensors with respect to a coordinate system. Let $\underline{g}(\underline{\theta};\underline{\delta}) \in \mathbf{R}^N$ be the vector of the values of the same output quantities predicted by a model for specific values of the parameter set $\underline{\theta}$. The following prediction error equation is introduced

$$\underline{d} = \underline{g}(\underline{\theta},\underline{\delta}) + \underline{e} \tag{17}$$

where $\underline{e}$ is the additive prediction error term due to model and measurement error. The prediction error is modeled as a Gaussian vector, whose mean value is equal to zero and its covariance is equal to $\Sigma(\underline{\sigma}) \in \mathbf{R}^{N\times N}$, where $\underline{\sigma}$ contains the parameters that define the correlation structure of Σ. Applying the Bayesian theorem, the posterior probability density function (PDF) of $\underline{\theta}$, given the measured data $\underline{d}$, is given by

$$p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta}) = c\frac{1}{\left(\sqrt{2\pi}\right)^N\sqrt{det\Sigma(\underline{\sigma})}}exp\left[-J(\underline{\theta};\underline{\sigma},\underline{d},\underline{\delta})\right]\pi(\underline{\theta}) \tag{18}$$

where

$$J(\underline{\theta};\underline{\sigma},\underline{d},\underline{\delta}) = \frac{1}{2}\left[\underline{d}-\underline{g}(\underline{\theta};\underline{\delta})\right]^T\Sigma^{-1}(\underline{\sigma})\left[\underline{d}-\underline{g}(\underline{\theta};\underline{\delta})\right] \tag{19}$$

expresses the deviation between the measured and model predicted quantities. The PDF $\pi(\underline{\theta})$ is the prior distribution for $\underline{\theta}$, and c is a normalization constant guaranteeing that the posterior PDF $p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta})$ integrates to one.

4.1.3 Information Entropy Quantification

The PDF $p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta})$ given by (18) quantifies the posterior uncertainty in the parameter values $\underline{\theta}$ based on the information contained in the measured data. The information entropy given by the expression

$$h_{\underline{\theta}}(\underline{\delta};\underline{\sigma},\underline{d}) = E_{\underline{\theta}}\left[-\ln p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta})\right]$$

$$= -\int \ln p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta})\, p(\underline{\theta}|\underline{\sigma},\underline{d},\underline{\delta})\, d\underline{\theta} \qquad (20)$$

is a scalar measure of the uncertainty of the model parameters $\underline{\theta}$. It depends on the location vector $\underline{\delta}$ of the sensors, the correlation structure of the prediction error and the details in the data $\underline{d}$. The multi-dimensional integral in equation (20) is a Laplace-type integral that can be asymptotically approximated, for large number of data, by the expression:

$$h_{\underline{\theta}}(\underline{\delta};\underline{\sigma},\underline{d}) \sim H(\underline{\delta};\underline{\theta}_0,\underline{\sigma})$$

$$= \frac{1}{2}\mathrm{N}_\theta \ln(2\pi) - \frac{1}{2}\ln det\left[Q(\underline{\delta};\underline{\theta}_0,\underline{\sigma}) + Q_\pi(\underline{\theta}_0)\right]$$

(21)

where $\underline{\theta}_0$ are the values of $\underline{\theta}$ that minimize $J(\underline{\theta};\underline{\sigma},\underline{d},\underline{\delta})$, $Q(\underline{\delta};\underline{\theta},\underline{\sigma})$ is the Fisher information matrix, a semi-positive definite matrix asymptotically given by

$$Q(\underline{\delta};\underline{\theta},\underline{\sigma}) = \nabla_{\underline{\theta}}\underline{g}(\underline{\theta};\underline{\delta})^{\mathrm{T}}\,\Sigma^{-1}(\underline{\sigma})\,\nabla_{\underline{\theta}}\underline{g}(\underline{\theta};\underline{\delta}) \qquad (22)$$

computed at the N locations where the sensors are placed, and $Q_\pi(\underline{\theta}_0) = -\nabla_{\underline{\theta}}{}^{\mathrm{T}}\nabla_{\underline{\theta}}\ln\pi(\underline{\theta})$ evaluated at the value $\underline{\theta}_0$, with $\nabla_{\underline{\theta}} = \left[\partial/\partial\theta_1, \cdots, \partial/\partial\theta_{\mathrm{N}_\theta}\right]$, represents the negative of the Hessian of the natural logarithm of the prior distribution of the model parameters. For uniform prior distribution the term $Q_\pi(\underline{\theta}_0) = 0$ and the optimal sensor placement is based only on the Fisher information matrix. There must be noticed that for small number of sensors, the matrix $Q(\underline{\delta};\underline{\theta},\underline{\sigma})$ can be ill-conditioned and the determinant could tend to zero independent of the location of sensor. Such cases arise from unidentifiability issues due to the

insufficient information provided by the data to estimate the number of model parameters involved. Non-informative uniform prior distribution do not provide any information to correct this problem. A solution could be to use non-uniform distributions to remove the ill-conditioning in $Q(\underline{\delta};\underline{\theta},\underline{\sigma})$ due to the extra information provided by the prior distribution about the uncertainty in the model parameters.

In addition, from equation (22) it can be concluded that the information entropy depends on the derivatives of the output quantities predicted by the model at the sensor locations with respect to the model parameters. The higher the derivatives, the higher the information entropy value. The computation of these derivatives is based on the differentiation of the model with respect to the parameters.

4.1.4 Optimal Sensor Location Methodology

The sensor configuration should be designed in such a way that the measured data are as much informative as possible about the model parameters to be estimated. The information entropy, defined by (21), measuring the uncertainty in these parameters, gives the amount of useful information contained in the measured data. The most informative test data are the ones that give the least uncertainty in the parameter estimates or the ones that minimize the information entropy or the change of information entropy. In other words, the sensors should be located at the places that minimize the information entropy. The problem of finding the optimal sensor configuration is formulated as an optimization problem where the objective function is the information entropy or the change of information entropy in the robust case and the design variables are the locations of sensors. Specifically, the optimal sensor location $\underline{\delta}_{best}$ is given by

$$\underline{\delta}_{best} = \arg\min \mathrm{H}(\underline{\delta}) \tag{23}$$

The minimization of (21) is equivalent to the minimization of

$$U(\underline{\delta};\underline{\theta}_0,\underline{\sigma}) = -\frac{1}{2}\ln\left(det\left[Q(\underline{\delta};\underline{\theta}_0,\underline{\sigma}) + Q_{\pi}(\underline{\theta}_0)\right]\right) \tag{24}$$

Given that the quantity $Q(\underline{\delta};\underline{\theta}_0,\underline{\sigma})$ defined in (22) depends on the sensitivity of output quantities, computed from the model at the measured locations, with respect to the parameters, the sensors tend to be placed at locations where the output quantities are most sensitive to parameter changes. This is consistent with intuition since sensors places at

locations where the quantities are insensitive to parameter changes do not provide information to estimate the values of the parameters.

4.2 Implementation for Source Identification in a Continuous Medium

A continuous elastic mean is considered. In a region of its area a disturbance starts to occur. The exact location of disorder source is unknown. The initial time of disturbance is unknown as well. Our objective is to place sensors in the area which extends beyond the area of the initial disorder, the purple region in the scheme below. The sensors have to be placed in the more efficient way in order to extract the maximum information about the initial point of the disturbance.

The optimum sensor placement methodology presented at the previous chapter is applied on the continuous mean.

It is obvious that two uncertain parameters is the exact location of fault origin and the starting time. Moreover the speed at which the signal is transmitted cannot be sufficiently pre-defined. In reality there are no perfectly isotropic materials, so that a signal can be transmitted at the same speed in all direction. Especially as far as the case of earth is examined this speed has an important fluctuation according to the soil morphology.

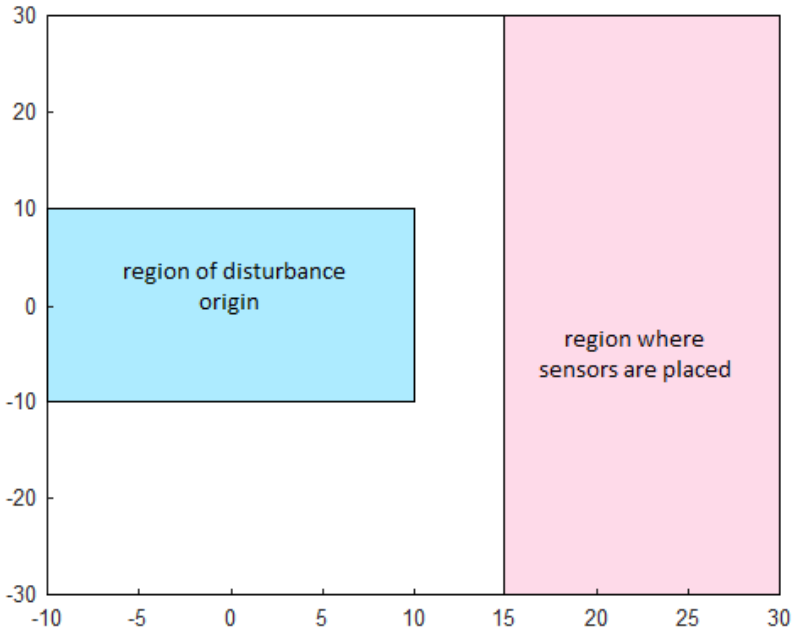


Figure 6: Continuous material with an initial disturbance

4.2.1 Sensitivities

As previously described information entropy is highly dependent on the gradient of the quantity of our interest, in the current section the derivatives of the time that the signal arrives at the sensor location with respect to the uncertain parameters are calculated. These derivatives are called sensitivities. It is clear that the larger values the sensitivities have the biggest information entropy we get.

The uncertain parameters of the present layout are the specific coordinates of the initial point of the disturbance, the initial time and the speed. That leads us on the calculation of the derivative of function given by $t_s(\underline{\theta};\underline{\delta})$ with respect to initial time t_0 as follows

$$\frac{\partial t_s(\underline{\theta};\underline{\delta})}{\partial t_0} = 1 \tag{25}$$

We are moving forward by computing the derivatives with respect to spatial coordinates x_0, y_0.

$$\frac{\partial t_s(\underline{\theta};\underline{\delta})}{\partial x_0} = \frac{-2\left(x_s - x_0\right)}{v\sqrt{\left(x_s - x_0\right)^2 + \left(y_s - y_0\right)^2}} = \frac{-2\left(x_s - x_0\right)}{vd_s} \tag{26}$$

$$\frac{\partial t_s(\underline{\theta};\underline{\delta})}{\partial y_0} = \frac{-2\left(y_s - y_0\right)}{v\sqrt{\left(x_s - x_0\right)^2 + \left(y_s - y_0\right)^2}} = \frac{-2\left(y_s - y_0\right)}{vd_s} \tag{27}$$

Finally the derivative with regard to speed can be written as:

$$\frac{\partial t_s(\underline{\theta};\underline{\delta})}{\partial v} = -\frac{1}{v^2} d_s \tag{28}$$

where

$$d_s = \sqrt{\left(x_s - x_0\right)^2 + \left(y_s - y_0\right)^2} \tag{29}$$

4.2.2 Identification of disturbance origin

4.2.2.1 Objective Function Formulation

In order to obtain the optimization of our experimental design the minimization of information entropy is necessary. Its quantification occurs as fully described in section 4.1. In

this way the problem is shifted to the minimization of the Utility function given by the expression (24). Consequently, we have to compute the equation bellow:

$$U\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right)=-\ln\det\left[\nabla_{\underline{\theta}}\underline{g}\left(\underline{\theta};\underline{\delta}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{\sigma}\right)\nabla_{\underline{\theta}}\underline{g}\left(\underline{\theta};\underline{\delta}\right)+Q_{\pi}\left(\underline{\theta}_{0}\right)\right]$$

where $\nabla_{\underline{\theta}}\underline{g}(\underline{\theta};\underline{\delta})=\nabla_{\underline{\theta}}\underline{t}(\underline{\theta};\underline{\delta})$

Taking into account the sensitivities derided previously we can state that

$$\nabla_{\underline{\theta}}t_{s}\left(\underline{\theta};\underline{\delta}\right)=\left[1\quad\frac{x_{s}-x_{0}}{vd_{s}}\quad\frac{y_{s}-y_{0}}{vd_{s}}\quad-\frac{1}{v^{2}}d_{s}\right]^{\mathrm{T}} \tag{30}$$

4.2.2.2 Discretization of Disturbance Area Using Finite Elements

For the formulation of Utility function (24), $U\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right)$ must be evaluated at $\underline{\theta}_{0}$, namely the values of $\underline{\theta}$ that minimize function (19). In the case that the experiment we are designing aims to identify the starting point of the disturbance we have that $\underline{\theta}_{0}=\left(x_{0},y_{0}\right)$. In the examined layout we know the broad area LxL where the fault starts to occur but not the specific coordinates. As a result, $U\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right)$ must be calculated at every potential starting point and then the average must be computed in order to infer a reliable result. The issue that we have to deal with in this point is how to consider every point of this continue area. In this attempt the area is divided into a grid of NxN points as the one that is depicted in Figure 7. We consider that the parameters x_{0}, y_{0} fluctuate into a range given as follows

$$-a\le x_{0}\le a \tag{31}$$

$$-b\le y_{0}\le b \tag{32}$$

The integral of the function $U\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right)$ must be calculated over x_{0}, y_{0}

$$U\left(\underline{\delta};\underline{\theta}_{0},\underline{\sigma}\right)=\int_{x_{0}}\int_{y_{0}}-\ln\det\left[Q\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right)+Q_{\pi}\left(\underline{\theta}\right)\right]dx_{0}dy_{0}=\int_{-a}^{a}\int_{-b}^{b}-\ln\det\left[Q\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right)+Q_{\pi}\left(\underline{\theta}\right)\right]dx_{0}dy_{0} \tag{33}$$

To estimate the integral above the area of our interest is divided creating grid of points. It is known that when this grid of points is used the accuracy is dependent to the number of grid points used. For an adequate number of grid points the integral can be estimated sufficiently.

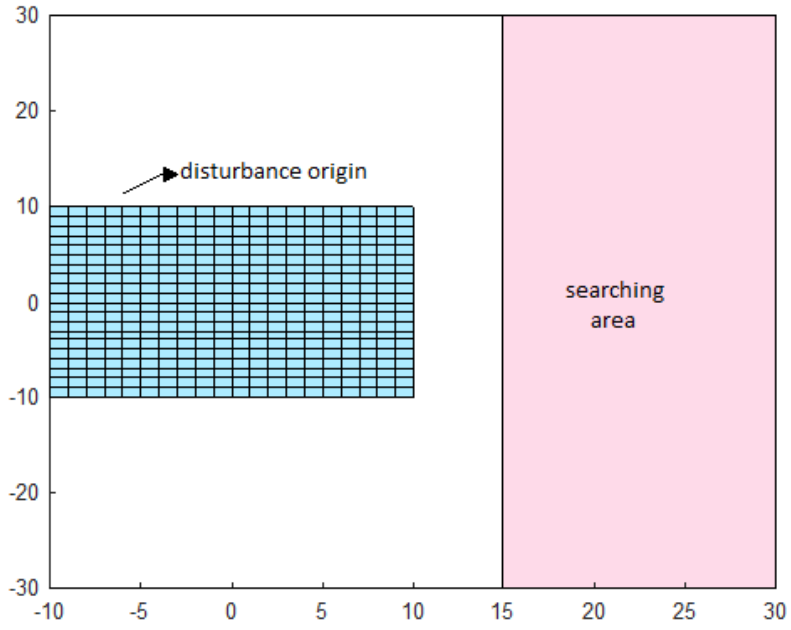


Figure 7: The examined material after discretizing the disturbance area

Thus the double integral in (33) can de written as:

$$U\left(\underline{\delta};\underline{\theta}_0,\underline{\sigma}\right)=\frac{1}{\left(N+1\right)^2}\sum_{k=1}^{\left(N+1\right)^2}-\ln\det\left[Q\left(\underline{\delta};\underline{\theta}_0^{(k)},\underline{\sigma}\right)+Q_\pi\left(\underline{\theta}_0^{(k)}\right)\right]$$

$$\frac{1}{\left(N+1\right)^2}\sum_{k=1}^{\left(N+1\right)^2}-\ln\det\left[\nabla_{\underline{\theta}}\underline{t}\left(\underline{\theta}_0^{(k)};\underline{\delta}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{\sigma}\right)\nabla_{\underline{\theta}}\underline{t}\left(\underline{\theta}_0^{(k)};\underline{\delta}\right)+Q_\pi\left(\underline{\theta}_0^{(k)}\right)\right]$$

where, N in the number of finite elements used on each coordinate. Thus, our calculation is repeated for $\left(N+1\right)^2$ points.

4.2.2.3 Unidentifiability Problem - Prediction Error Correlated Model

In order to estimate the number of uncertain parameters involved we need to use an adequate number of sensors. Otherwise, unidentifiability issues are performed, as already described in the theory bellow. The objective function cannot be calculated and it is not possible for us to infer results about the optimal sensor placement.

Similar problems are provoked when the sensors are placed at approximately the same location. This effect is thoroughly documented at a work of Papadimitriou et al. [2] and a solution is proposed.

To begin with, we have to consider the covariance matrix $\Sigma(\underline{\sigma})$ of the prediction error. Assuming uncorrelated prediction errors the covariance matrix can be defined as follows

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{N_s}^2 \end{bmatrix} \tag{34}$$

However it has been shown in a previous work [13] that the upper assumption is not convenient when we implement optimal sensor placement framework on the continuous space. That can be captured if we take into account that when the framework places the sensors at the same points or at neighboring ones we do not extract adequate information. To address that issue correlated prediction error is considered as follows

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_N\rho_{1N} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \sigma_N\sigma_1\rho_{N1} & \cdots & & \sigma_N^2 \end{bmatrix} \tag{35}$$

where ρ_{ij} indicates the spatial correlation structure, which in our case depends on distance of the location of the corresponding pair of sensors $\underline{s}_i$, $\underline{s}_j$. Spatial correlation structure is given by the expression

$$\rho_{ij} = \exp\left\{-\frac{d_{ij}}{\lambda}\right\} \tag{36}$$

where λ is a measure of spatial correlation length. It can be shown that the correlation length of the prediction error controls the minimum distance between the sensors. Thus, correlation length must be selected conveniently as very small values are not adequate while too big values related to the layout dimension lead to unreliable results.

4.2.2.4 CMA Optimization Technique

In this work a stochastic optimization technique is employed, the covariance matrix adaptation evolution strategy (CMA-ES) [14]. An evolution strategy samples new candidate solutions according to a multivariate normal distribution in $\mathbf{R}^n$. A recombination takes place in order for a new mean value for the distribution to be selected. Then, a random vector is added which is a perturbation with zero mean. A covariance matrix depicts the pairwise

dependencies between the variables in the distribution. CMA method updates the covariance matrix of this distribution to generate better and better values for the objective function. The CMA-ES has been empirically successful in numerous applications and is thought to be particularly useful on non-linear, ill-conditioned objective functions.

The advantage of this method is that it doesn't use gradients and consequently entrapments to local minima are avoided. A previous optimal sensor placement study applied in continuous space opted for the implementation of the specific method [3], a fact that encouraged us to use this technique.

4.3 Implementation for Parameter Estimation and Infection Origin Identification in Epidemiology Networks

When we are referred to optimal sensor placement in an Epidemiology Network consisted of populations each one including a number of interacting communities we mean the best way to extract information about the state of the disease. For example, from which communities to derive daily information about the number of symptomatic patients admitted to hospitals. Specifically, the measured quantity on which the following computations are based is the infective population $I(t)$, where t is a time unit that in real world can represent a week.

4.3.1 Sensitivities

In order to accomplish the optimization procedure we have to quantify the information entropy of the PDF. This is the quantity that has to be minimized to infer our results. The first step is to extract the derivatives of susceptible, infective and removed populations with respect to the parameters that we want to optimize, β, γ and I_0. This occurs by solving the relevant system of differential equation that can be derived from the initial set of equations (16).

To be more specific derivatives with respect to β can be calculated by solving computationally the following system:

$$\frac{dS_\beta}{dt} = -I \cdot S - \beta I_\beta \cdot S - \beta I \cdot S_\beta + \Lambda^{\mathrm{T}} S_\beta - (\Lambda F) \cdot S_\beta$$

$$\frac{dI_\beta}{dt} = I \cdot S + \beta I_\beta \cdot S + \beta I \cdot S_\beta - \gamma I_\beta + H^{\mathrm{T}} I_\beta - (HF) \cdot I_\beta \qquad (37)$$

$$\frac{dR_\beta}{dt} = \gamma I_\beta + G^{\mathrm{T}} R_\beta - (GF) \cdot R_\beta$$

The notations S_β, I_β, R_β represent the derivatives we are interested to extract. Having computed the quantities of susceptible, infective and removed populations $S(t),\ I(t),\ R(t)$ through the procedure discussed in section 3.2 it is simple to continue by solving the system above through numerical integration. What is left is to define the initial condition. At $t = 0$ the quantities $S,\ I,\ R$ have a specific value independent of parameter β so it is easy to understand that their derivatives at the initial time are equal to zero. That is, $S_{\beta i}(0) = 0,\ I_{\beta i}(0) = R_{\beta i}(0) = 0$ for every node.

Accordingly derivatives with respect to γ will be extracted by solving the system of equations that follows:

$$\frac{dS_\gamma}{dt} = -\beta I_\gamma \cdot S - \beta I \cdot S_\gamma + \Lambda^{\mathrm{T}} S_\gamma - (\Lambda F) \cdot S_\gamma$$

$$\frac{dI_\gamma}{dt} = \beta I_\gamma \cdot S + \beta I \cdot S_\gamma - I - \gamma I_\gamma + H^{\mathrm{T}} I_\gamma - (HF) \cdot I_\gamma \tag{38}$$

$$\frac{dR_\gamma}{dt} = I + \gamma I_\gamma + G^{\mathrm{T}} R_\gamma - (GF) \cdot R_\gamma$$

With initial conditions $S_{\gamma i}(0) = 0,\ I_{\gamma i}(0) = R_{\gamma i}(0) = 0$ for every node.

Finally, as far as derivatives with respect to I_0 are concerned, the system of equation to be solved is:

$$\frac{dS_{I_0}}{dt} = -\beta I_{I_0} \cdot S - \beta I \cdot S_{I_0} + \Lambda^{\mathrm{T}} S_{I_0} - (\Lambda F) \cdot S_{I_0}$$

$$\frac{dI_{I_0}}{dt} = \beta I_{I_0} \cdot S - \beta I \cdot S_{I_0} - \gamma I_{I_0} + H^{\mathrm{T}} I_{I_0} - (HF) \cdot I_{I_0} \tag{39}$$

$$\frac{dR_{I_0}}{dt} = \gamma I_{I_0} + G^{\mathrm{T}} R_{I_0} - (GF) \cdot R_{I_0}$$

With initial conditions $S_{I_o i}(0) = 0,\ I_{I_0 i}(0) = R_{I_0 i}(0) = 0$ for every node except for the nodes where the infection begins where the derivative of I_0 with respect to I_0 is equal to one. Namely, $S_{I_0 a}(0) = 0,\ I_{I_0 a}(0) = 1,\ R_{I_0 a}(0) = 0$.

Interpolation Scheme

It must be reminded here that the quantities $S(t),\ I(t),\ R(t)$ have been computed numerically for sufficiently small time step and stored in $(m \times N_t)$ matrices. Where m denotes the number of nodes and N_t represents the number of time steps. To carry out the calculation of gradients with regard to the parameters $\beta,\ \gamma$ and I_0 an interpolation scheme is used to compute the values at the points between the given ones. In specific linear interpolation is implemented to approximate these values.

4.3.2 Estimation of parameters β and γ

The first problem that we have to deal with is where to place sensors in order to measure the infective population in each node, I_k and how many sensors are needed so that to estimate the parameters β and γ more efficiently.

In order to formulate the utility function we have to take into consideration the sensitivities that have been developed in the previous unit. As utility function we define the function that quantifies the information entropy and is given by the general form cited in the section 4.1 as

$$U\left(\underline{\delta};\underline{\theta}_0,\underline{\sigma}\right) = -\frac{1}{2}\ln\left(det\left[Q\left(\underline{\delta};\underline{\theta}_0,\underline{\sigma}\right)+Q_\pi\left(\underline{\theta}_0\right)\right]\right) \tag{40}$$

where,

$$Q\left(\underline{\delta};\underline{\theta},\underline{\sigma}\right) = \nabla_{\underline{\theta}}\underline{g}\left(\underline{\theta};\underline{\delta}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{\sigma}\right)\nabla_{\underline{\theta}}\underline{g}\left(\underline{\theta};\underline{\delta}\right) \tag{41}$$

The term $Q_\pi\left(\underline{\theta}_0\right)$ can be ignored as it doesn't play a significant role in the optimization procedure. The notation $\nabla_{\underline{\theta}}\underline{g}^T$ represent the sensitivity matrix which is consisted by the derivatives computed at the N locations where sensors are placed.

$$\nabla_{\underline{\theta}}\underline{g}^T = \nabla_{\underline{\theta}}\underline{I} = \begin{bmatrix} \frac{\partial I_1}{\partial\beta} & \frac{\partial I_2}{\partial\beta} & \cdots\frac{\partial I_m}{\partial\beta} \\ \frac{\partial I_1}{\partial\gamma} & \frac{\partial I_2}{\partial\gamma} & \cdots\frac{\partial I_m}{\partial\gamma} \end{bmatrix}_{(t)} \tag{42}$$

$\underline{\theta}$ denotes the parameters. In this case, $\underline{\theta} = \left(\beta,\gamma\right)$.

In this point we have to take into account that the infective population I is a function of time. We assume that measurements take place periodically. We consider a period T. So the number of time steps is $\kappa = {t_{total}}/{T}$. As a result, the utility function can be inferred from the general form of $U\left(\underline{\delta};\underline{\theta}\right)$ if we sum the quantity in the det over the time points κ.

Since the parameters $\beta,\ \gamma$ are presumed to be uncertain parameters, the utility function can be computed as the integral of the information entropy conditioned on the nominal values of the parameters.

$$U\left(\underline{\delta};\underline{\theta}_0,\underline{\sigma}\right)=\int_{\beta}\int_{\gamma}\left(-\frac{1}{2}\ln det\sum_{\kappa=1}^{\bar{\kappa}}\left[\nabla_{\underline{\theta}}\underline{g}\left(\underline{\theta};\underline{\delta}\right)^{\mathrm{T}}\cdot\Sigma^{-1}\cdot\nabla_{\underline{\theta}}{}^{\mathrm{T}}\underline{g}\left(\underline{\theta};\underline{\delta}\right)+Q_{\pi}\left(\underline{\theta}_0\right)\right]\right)p\left(\beta,\gamma\right)d\gamma d\beta \tag{43}$$

where $p\left(\beta,\gamma\right)$ is the prior PDF of β and γ. Herein a uniform PDF is assumed.

This double integral can be estimated using a sparse grid point technic [15, 16]. In particular, there is created a sum over the values of information entropy at specific points of our parameter space weighed by the appropriate contributors. Consequently, it appears that the utility function which is to be optimized is

$$U\left(\underline{\delta};\underline{\theta}_0,\underline{\sigma}\right)=\sum_{j=1}^{N}w_j\left(-\frac{1}{2}\ln det\sum_{\kappa=1}^{\bar{\kappa}}\left[\nabla_{\underline{\theta}}\underline{g}\left(\underline{\theta}_0^{(j)};\underline{\delta}\right)^{\mathrm{T}}\cdot\Sigma^{-1}\cdot\nabla_{\underline{\theta}}{}^{\mathrm{T}}\underline{g}\left(\underline{\theta}_0^{(j)};\underline{\delta}\right)+Q_{\pi}\left(\underline{\theta}_0\right)\right]\right) \tag{44}$$

where, $\underline{\delta}$ is a vector that denotes sensors' locations.

N is the number of sparse grid points that are used and w_j are the corresponding weights.

$$\Sigma=\left[\begin{pmatrix}\sigma_1^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma_v^2\end{pmatrix}\right] \tag{45}$$

where, v is the number of sensors that are used.

It is assumed that $\sigma_1=\sigma_2=\ldots=\sigma_v=0.1$.

4.3.3 Origin Identification assuming constant parameters

A second task we have to deal with is to detect where to place sensors and measure I_k and how many sensors are needed so that to estimate where the infection starts if we know that it starts in a region of nodes.

In this case the parameter set to be inferred is $\underline{\theta} = \underline{I_0}$.

As for the sensitivities' matrix it is given by:

$$\nabla_{\underline{\theta}} \underline{g}^T = \nabla_{\underline{\theta}} \underline{I} = \left[\frac{\partial I_1}{\partial I_0} \quad \frac{\partial I_2}{\partial I_0} \quad \cdots \frac{\partial I_m}{\partial I_0} \right]_{(t)} \tag{46}$$

The objective is again to minimize the utility function. We know the region of nodes where the infection started, but not the specific node. In this case it is needed to compute the sum over the potential places of infection origin weighted by appropriate contributor which are assumed to be equal to $1/N$ for each point, where N is the number of the potential starting points.

$$U\left(\underline{\delta}; \underline{\theta}_0, \underline{\sigma}\right) = \sum_{j=1}^{N} \frac{1}{N} \left(-\frac{1}{2} \ln det \sum_{\kappa=1}^{\bar{\kappa}} \left[\nabla_{\underline{\theta}} \underline{g}\left(\underline{\theta}_0^{(j)}; \underline{\delta}\right)^{\mathrm{T}} \cdot \Sigma^{-1} \cdot \nabla_{\underline{\theta}}{}^{\mathrm{T}} \underline{g}\left(\underline{\theta}_0^{(j)}; \underline{\delta}\right) + Q_{\pi}\left(\underline{\theta}_0\right) \right] \right) \tag{47}$$

4.3.4 Origin Identification for uncertain parameters β, γ

In an attempt to make our information entropy more robust we assume our parameters β and γ to be uncertain parameters. While β and γ represent two physical quantities those of the rate at which the members of the population come into contact and the rate at which the infective people recover respectively, assuming that they have a specific constant value is not very realistic. In order to elicit more reliable results, in this point of our analysis, we assume that β and γ are uncertain. The uncertainty is quantified using normal distribution with $\mu_\beta = 0.02$ and $\mu_\gamma = 0.3$ and diagonal covariance matrix $diag[\sigma_\beta^2, \sigma_\gamma^2]$.

It goes without saying that we are introducing uncertainties in parameters that concern our model and are not included in the parameters to be optimized. Let's consider that the vector $\underline{\varphi_0}$ is augmented to also include these model related quantities. Papadimitriou et al. [17] have related the divergence of model parameters from their presumed values with the robustness

of the quantification of information entropy. The change of information entropy initially defined by the expression (21) is given by

$$\Delta h\left(\underline{\delta}\right)=\mathrm{E}_{\underline{\theta},\underline{\varphi}_0}\left[-\ln p\left(\underline{\theta},\underline{\varphi}_0|\underline{\delta}\right)\right]-\mathrm{E}_{\underline{\varphi}_0}\left[-\ln p\left(\underline{\varphi}_0\right)\right]$$

$$=\int H\left(\underline{\delta};\underline{\varphi}_0\right)\pi(\underline{\varphi}_0)d\underline{\varphi}_0$$

$$=\frac{1}{2}\mathrm{N}_\theta\ln\left(2\pi\right)-\frac{1}{2}\int\ln det\left[Q\left(\underline{\delta};\underline{\varphi}_0\right)+Q_\pi\left(\underline{\varphi}_0\right)\right]\pi\left(\underline{\varphi}_0\right)d\underline{\varphi}_0 \quad (48)$$

To approximate the integral above one can use sampling techniques. Two such method have been proposed from related existing works [15], [16]. These are Monte Carlo Integration or Sparse Grid points methods. Consequently, the integral in (48) can be written as:

$$\int\ln\det\left[Q\left(\underline{\delta};\underline{\varphi}_0\right)+Q_\pi\left(\underline{\varphi}_0\right)\right]\pi\left(\underline{\varphi}_0\right)d\,\underline{\varphi}_0\approx\sum_{i=1}^{n}w_i\ln det\left[Q\left(\underline{\delta};\underline{\varphi}_0^{(i)}\right)+Q_\pi\left(\underline{\varphi}_0^{(i)}\right)\right] \quad (49)$$

where the vector $\underline{\varphi}_0{}^{(i)}$, $i=1,\cdots,n$, is consisted from values for the model related parameters generated by the already referred sample techniques. The notation w_i represents the sample's weight. In the case of sparse grid points' approach they are defined appropriately with respect to the number of model parameters and the level of accuracy someone wants to obtain. For the Monte Carlo method they are equal to $w_i=1/M$, where M is the number of samples. It should be noted that expression (49) requires the sensitivities of the output quantities to be computed for all samples $\varphi_0^{(i)}$, $i=1,\ldots,M$.

Thus our utility function can be formulated as:

$$U\left(\underline{\delta};\underline{\varphi}_0\right)=\sum_{i=1}^{M}\frac{1}{w_i}\left(\sum_{j=1}^{N}\frac{1}{N}\left(-\frac{1}{2}\ln det\sum_{\kappa=1}^{\bar{\kappa}}\left[\nabla_{\underline{\theta}}\underline{g}\left(\underline{\varphi}_0^{(j,i)};\underline{\delta}\right)^{\mathrm{T}}\cdot\Sigma^{-1}\cdot\nabla_{\underline{\theta}}{}^{\mathrm{T}}\underline{g}\left(\underline{\varphi}_0^{(j,i)};\underline{\delta}\right)+Q_\pi\left(\underline{\varphi}_0^{(j,i)}\right)\right]\right)\right) \quad (50)$$

Having to deal with two uncertain parameters, β and γ, sparse grid points method is implemented.

4.3.5 Origin Identification for uncertain transition rates

In this point we introduce uncertainty in the transition rates $\Lambda,\ \mathrm{H},\ \mathrm{G}$. The elements of the three matrices express the rate of movement from one node to another for the susceptible, infective and removed group. We assume that the transition rates $\Lambda,\ \mathrm{H},\ \mathrm{G}$ follow a normal

distribution with means their previous values and standard deviation 0.2 . Same methodology as in the previous section is followed to extract robust predictions for the parameters of our interest including the uncertainty of these model parameters in the quantification of information entropy.

Transition rates are initially thought to be equal for all movements between nodes that can occur in our network. When introducing in these quantities a fluctuation one option is to consider that all the elements of each transition matrix change uniformly. This assumption implies that we have three changing quantities. The integral of our utility function over these quantities can be sufficiently approximated numerically through sparse grid points' method.

$$U\left(\underline{\delta};\underline{\varphi}_0\right)=\sum_{i=1}^{M}\frac{1}{w_i}\left(\sum_{j=1}^{N}\frac{1}{N}\left(-\frac{1}{2}\ln det\sum_{\kappa=1}^{\bar{\kappa}}\left[\nabla_{\underline{\theta}}\underline{g}\left(\underline{\varphi}_0^{(j,i)};\underline{\delta}\right)^{\mathrm{T}}\cdot\Sigma^{-1}\cdot\nabla_{\underline{\theta}}{}^{\mathrm{T}}\underline{g}\left(\underline{\varphi}_0^{(j,i)};\underline{\delta}\right)+Q_{\pi}\left(\underline{\varphi}_0^{(j,i)}\right)\right]\right)\right) \tag{51}$$

The other case that we examine makes our approach more realistic as we take into account that transition rates can change independently. This fact results in having $\frac{m\times m}{2}-m+2$ changing quantities to deal with, where m is the number of our nodes. Let us presume that the elements of $\Lambda,\ \mathrm{H},\mathrm{G}$ matrices fluctuate within a normal distribution with means their previous values and standard deviation $0,2$.

Sparse grid method cannot be implemented and infer reliable results in a problem of that high dimension. Consequently we are led to Monte Carlo Integration. Utility function in this case has the same form as in relation (51).

4.3.6 Optimization procedure - FSSP and BSSP algorithms

Having created an appropriate software that calculates the value of the utility function for each sensors' placement we have to deal with the optimization problem. That is, we have to find the sequence of sensor places that minimize the utility function. This will be the more efficient way to use our sensors and realize our measurements. One option would be to calculate the value of Utility function for all the potential sequences of sensor placements and opt for the sequence which resulted in the minimum value. However, this process which is basically the exhaustive search is extremely computationally expensive. We are led on accomplishing that task using two heuristic methods. In specific the Forward Sequential

Sensor Placement (FSSP) and Backward Sequential Sensor Placement (BSSP) are both implemented and their obtained results are compared to validate their correctness.

FSSP is a simple algorithm which places one sensor at a time at the location with the highest reduction in information entropy. In particular, it initially considers all the potential sensor locations and computing the Utility Function for one sensor at node of our network. The first sensor is placed at the node with the lowest value of Utility. Then it follows the same process considering as known the place of the first sensor. In the same way it continues by placing the third sensors given the places of the first and second sensor and so on.

On the other hand, BSSP appears to realize the inverse procedure. BSSP initializes by placing sensors at every node. After this it removes one sensor at a time from the position that results in the maximum Utility Function which is equivalent to removing a sensor from the location that produces the lowest increase in the information entropy. That occurs by omitting each time one node and identifying the omission one that leads to the Maximum Utility function. The corresponding node is removed. The procedure is repeated until we find the most "informative" nodes.

Sequential Sensor Placement algorithms do not guarantee that the global minimum is found. However, they consist a very good approach and they decrease importantly the computational effort in comparison with exhaustive search. Previous studies lying on the field of Optimal Sensor Placement proposed the employment of these two methods [18], [19], [20].

4.3.7 Sample methods

4.3.7.1 Sparse Grid Points

Sparse Grid Points Technique is used multiples times in our approach. In general, this method is a numerical discretization technique for multivariate problems. It constructs a multidimensional multilevel basis by a special truncation of the tensor product expansion of a one-dimensional multilevel basis. Discretizations on sparse grids involve only $O(N(\log N^{d-1}))$ degrees of freedom, where d is the problem dimension and N denotes the number of degrees of freedom in one coordinate direction. However, the number of basis functions or nodes (grid points) that have to be stored and processed depend exponentially on the number of dimensions. Even with today's computational power it is not possible to process functions with more than 4 or 5 dimensions.

Taking into account the methodology of the specific technique, its functional range and its inadequacies we utilize it to approximate integrals of 2 (β, γ) and 3 $(\Lambda, \mathrm{H}, \mathrm{G})$ dimensions. This procedure is realized through software programming. An appropriate code that generates sparse grid points is widely accessible in the Internet.

4.3.7.2 Monte Carlo Integration

Monte Carlo Integration methods is based on the Law of Large Numbers according to which the average of the results obtained from a large number of trials is close to the expected value and tends to become closer as more trials are performed.

The need of implementing this method derives from the fact that we want to insert uncertainties in each element of the transition matrices and make them change independently. To deal with this an adequate number of samples is demanded. Let's give an illustration of the applied framework. Firstly the non-zero elements of the transition matrices are detected. At each of them a random value from a normal distribution with zero mean and appropriately defined covariance is given. This procedure is repeated for a number of iterations equal to the number of samples we will to generate. For each sample the value of utility function is computed. Finally the average of these values is calculated.

CHAPTER 5

Results

5.1 Optimal Sensor Placement for Disturbance Identification in a Continuous Medium

The objective function thoroughly described in chapter 4 is initially constructed considering a grid of 20x20 points for the region of disturbance origin. The area that sensors can be placed ranges from 15 to 30 in x-coordinate and from -30 to 30 in y-coordinate.

To give a visual illustration of the objective function that we want to minimize we assume that the x coordinates of sensors' locations have fixed values both equal to 15 and we derive the dependence of the objective function, U , on the y vector. That occurs dividing the y area into finite elements of 1 unit length and computing the function at each point of grid. A 3-dimensional figure is generated to depict the dependence of U on the location of two sensors on the line x=15.

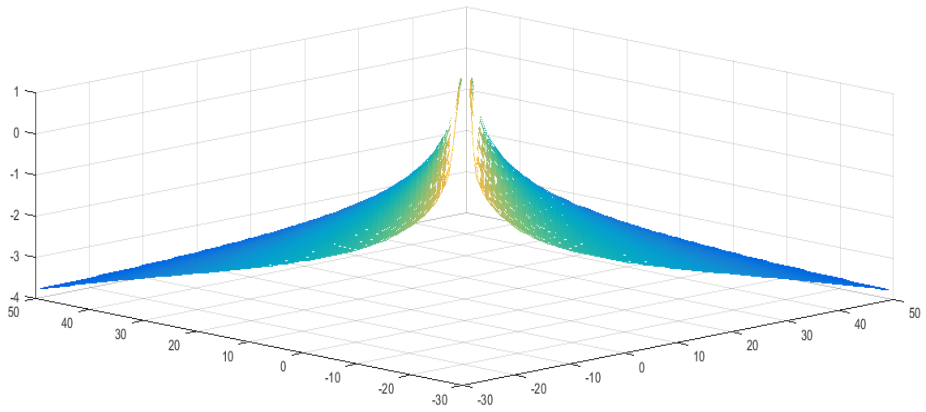


Figure 8: The objective function over y_1, y_2, considering x_1,x_2 fixed

The discontinuity that can be observed on the figure arises when the coordinates of the two sensors coincide. Given that we assumed the x-coordinates to be equal, when the y_1 , y_2 take the same value the objective function tends to infinity. We have considered that particularity by a methodology described in section 4.2.2 which controls the minimum distance at which the two sensor can approach each other. We initially define the spatial correlation length to be equal to 1.

The results in Figure 9 are derived from the CMA implementation on the minimization problem.

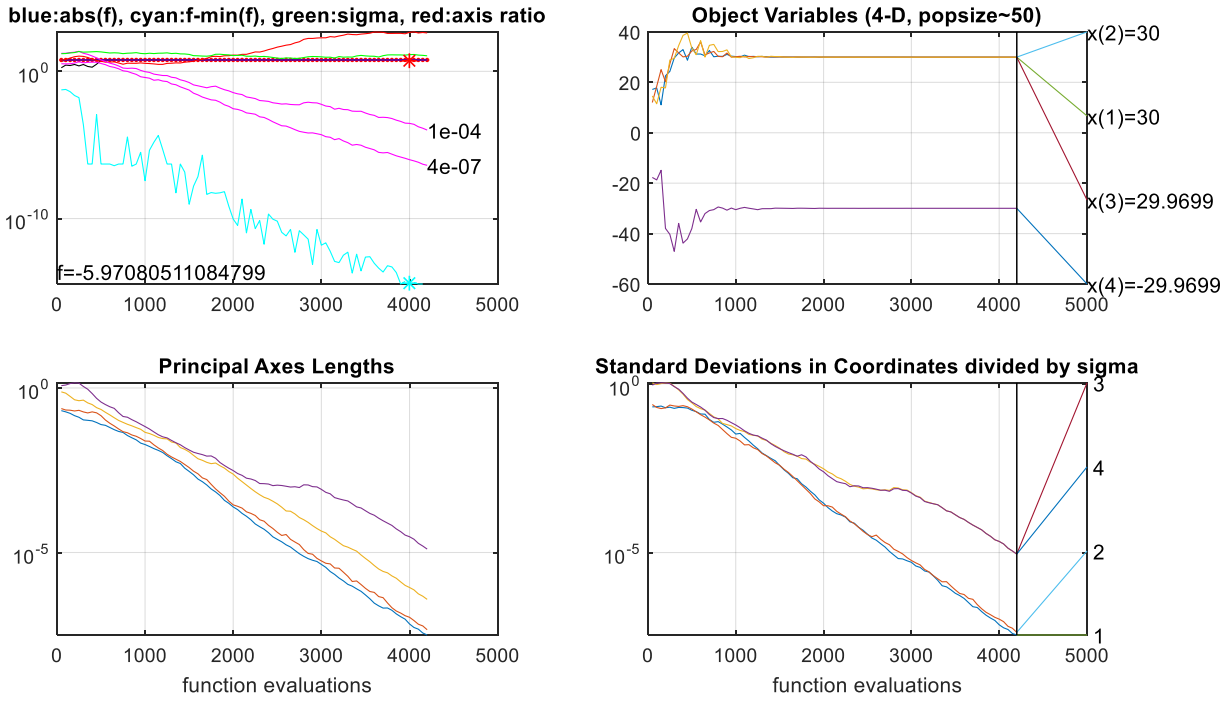


Figure 9: CMA results for 2 sensor placements

Figure 9 is generated from a CMA-ES optimization software and claims that the system converges. This is denoted by the upper right diagram where the curves that depict the variables x, y end up to straight lines. The cyan line in the upper left figure denotes the divergence between the value of objective function and its minimum value. This quantity have to decrease over function evaluations. That occurs with some fluctuation which can be considered normal. In addition, the standard deviations of our variables are represented in the lower right figure after being divided by sigma whose value is given to be close to 2 (green line on the upper left figure). Thus, standard deviations are decreasing to get very low values.

The optimal sensor places appears to be $s_1 = (30, 29.97), s_2 = (30, -29.97)$. In figure 9, x(1) and x(2) account for the x-coordinates, while x(3) and x(4) go for the y-coordinates. We can state that the optimization process proposes the placement of the sensors at the most remoted points of the examined region.

Changes in the initial values of the parameters or in the population size CMA-ES uses do not result in any discrepancies for the final results both for the parameters values and the objective function. Shifting the value of spatial correlation length from 1 to 5 we get identical

results. However it takes more function evaluations for the system to converge. Furthermore it appears that even if we run the software considering a diagonal covariance matrix, the problem converges to the same values.

In a purpose to examine the accuracy of our results we divide the distribution space into a larger number of elements. A dense grid of 200x200 is created. However, the elicited results do not differ to an important extend.

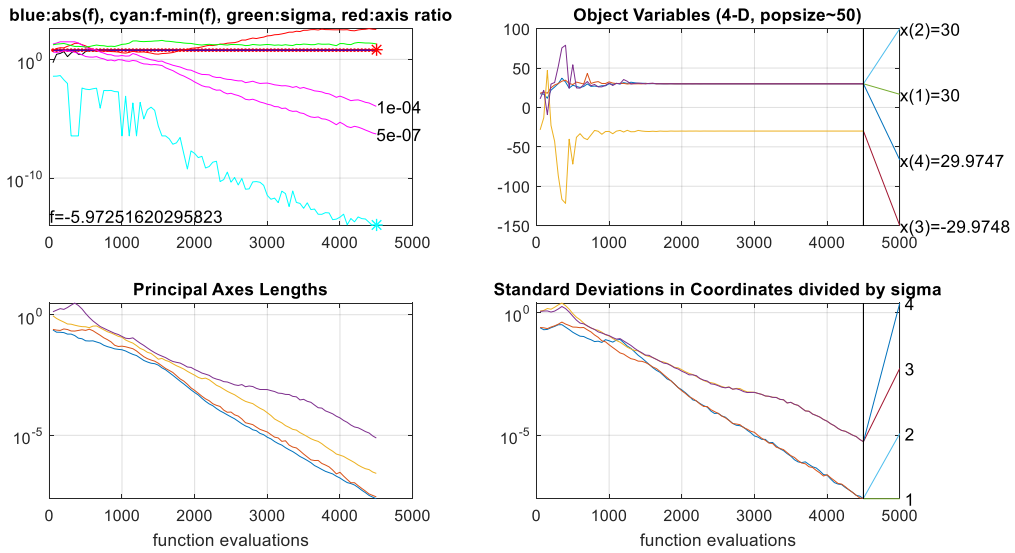


Figure 10: CMA results considering a dense grid for the initial disturbance area

The results appear to be very close to the previous ones.

Table 1: Two cases' solutions comparison

	Sparse Grid (20x20)	**Dense Grid (200x200)**	**Percentage Change**
x_1	30	30	0
y_1	29,9699	29,9747	0,016%
x_2	30	30	0
y_2	-29,9699	-29,9748	0,016%
U	-5,9708	-5,9725	0,028%

We move forward to detect the optimal location of a third sensor. Given the coordinates of the first two sensor we repeat the procedure to locate an additional one.

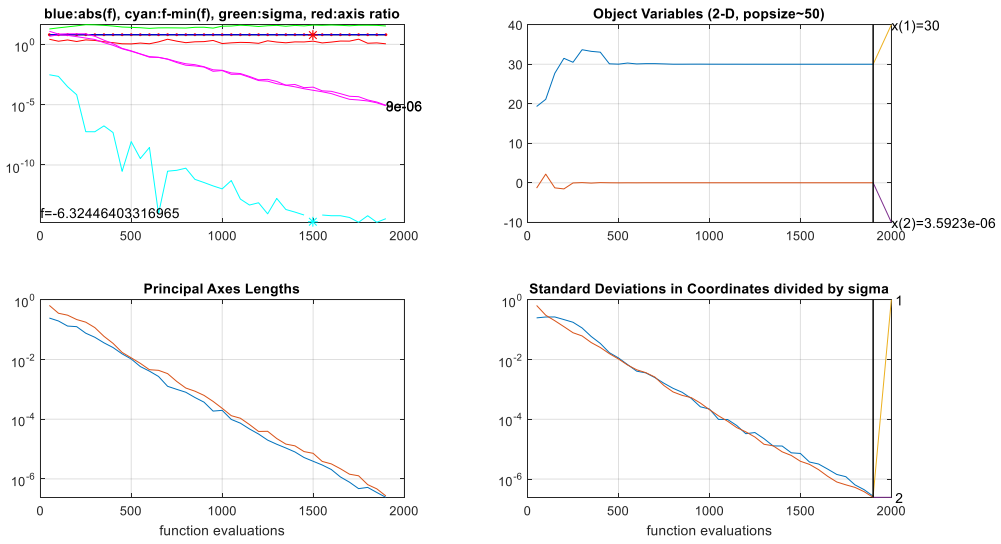


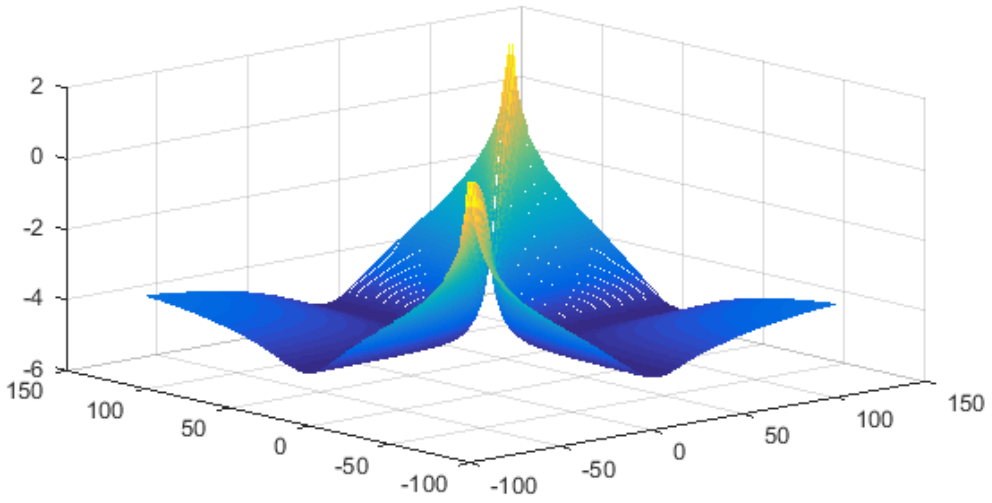
Figure 11: CMA results for a third sensor placement

It is clear that the system converges. The selected place is cited in the farthermost point of x-coordinate and in the middle of the two already placed sensors in the y- coordinate. We can note that the value of the Objective Function decreases after the placement of the 3rd sensor. That means that the utilization of three sensors provide us with more information for the identification of disturbance origin. The results for the sensors' coordinates and the Objective Function for the three sensor placements are given in Table 2.

Table 2: Sensor Placements

Procedure	Sensors added	Objective Function
1	(30, 29.97) (30, -29.97)	-5,9708
2	(30, 0)	-6,3245

The tendency of the applied optimization methodology to opt for places removed from the region of initial disorder triggers our curiosity. A figure similar to figure 8 is produced to demonstrate the objective function for an ample area of y_1, y_2 presuming again that $x_1 = x_2 = 15$.



It appears that the objective function is being reduced for increasing y_1, y_2 until they reach a critical value and then it is getting larger. However, on the examined region of our problem y-coordinates get lower values than the critical ones. That can explain why the farthermost points of the available space are selected first.

5.2 Optimal Sensor Placement for Parameter Estimation and Infection Origin Identification in an Epidemiology Network

5.2.1 Sensitivities

Derivatives of Infection Population with respect to uncertain parameters appear to have the form demonstrated bellow in figure 13. Let's clarify here that the network is consisted of 2 communities, each of them composed by 4 nodes. The initial conditions can be stated as $I_m(0)=0,\ m\neq 2$ and $I_2(0)=95$. A graph of the studied network is given in Figure 12. Node 2 is marked with red color to dictate that an infection starts there.

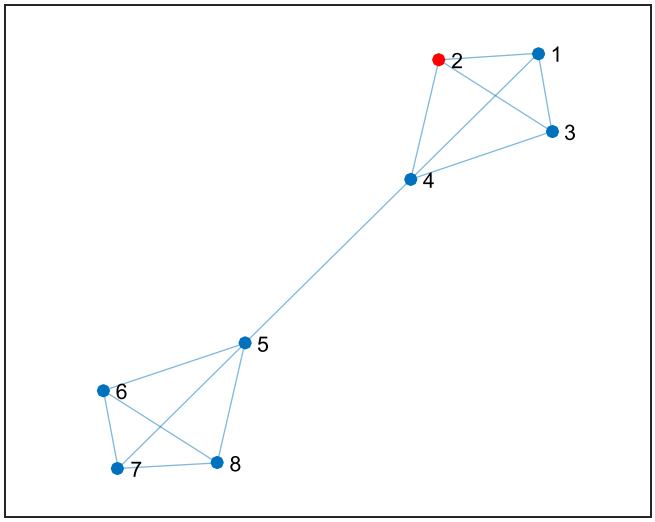
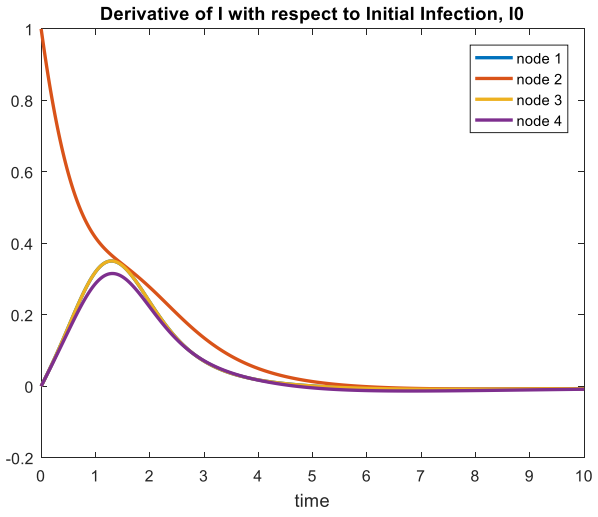
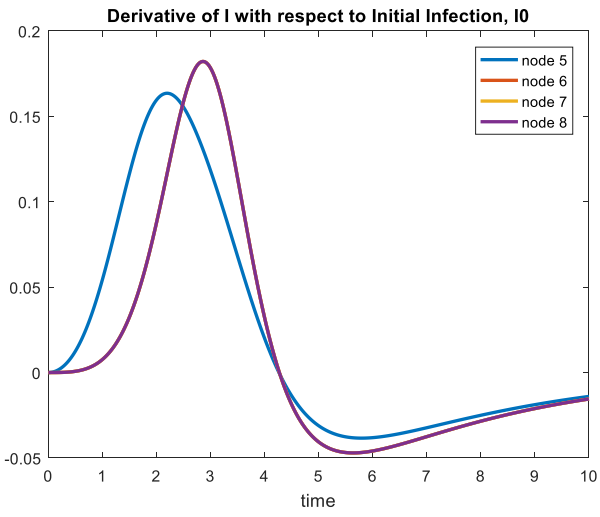


Figure 12: Examined Graph

The time evolution of gradients of infective population are displayed for each node. On the left, figures that concern the nodes of the first community are displayed, while on the right, figures for the second community are presented.





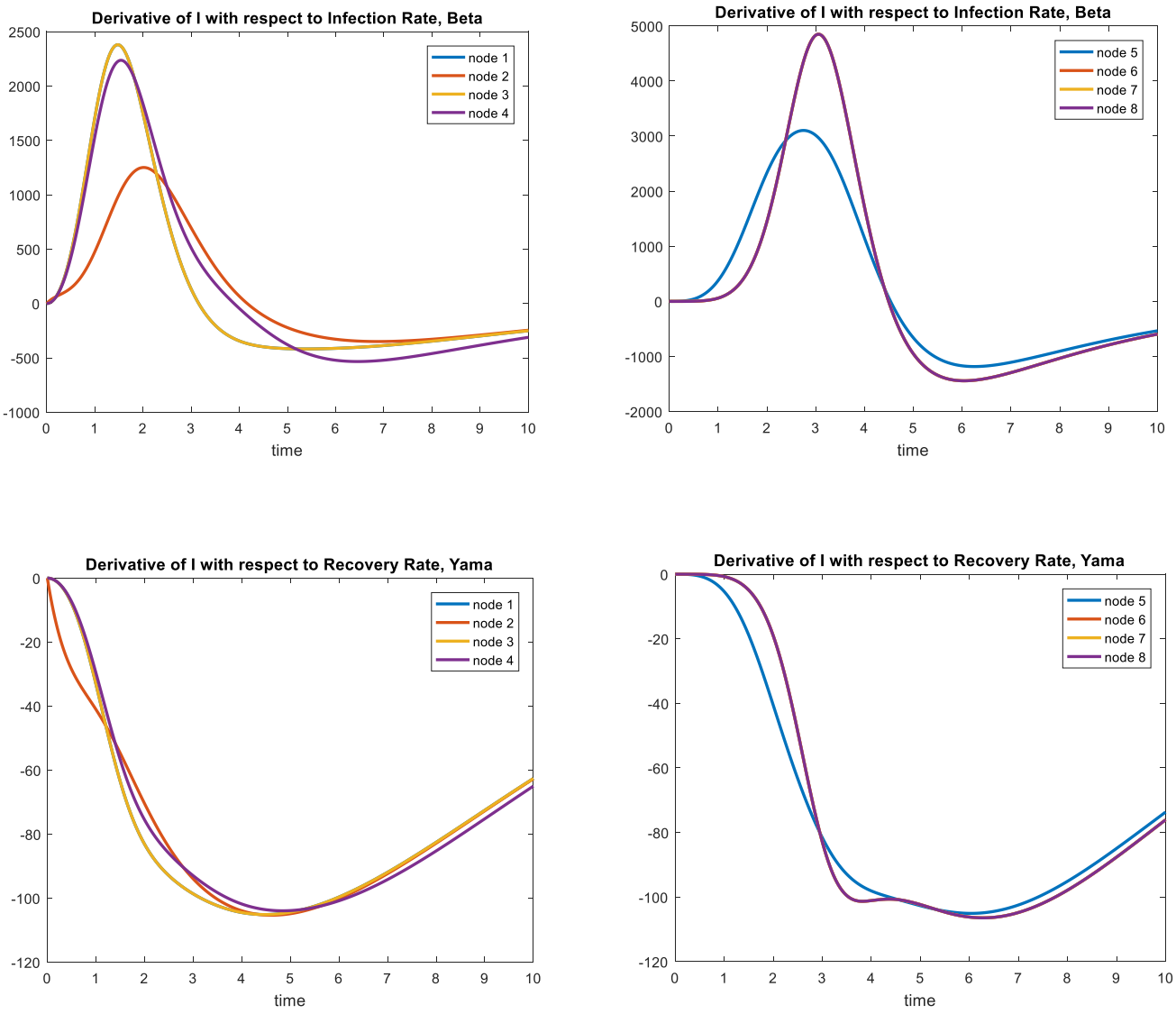


Figure 13: Sensitivities of I in regard to uncertain parameters, I_0, β and γ

As long as sensitivities were numerically computed, it is important to mark that they are continuous and smooth which indicates that the time step we used was small enough to approximate the analytical solution. Moreover their forms are consistent with intuition while curves of symmetrically placed nodes coincide. Specifically, observing the graph one can notice that while the infection starts at node 2 and transition rates are uniform, places of nodes 1 and 3 are equivalent as well as those of nodes 6, 7, 8.

5.2.2 Optimal Sensor Placement for Estimation of Parameters β and γ

Initially results concerning optimal sensor placement for the estimation of parameters β and γ are carried out. Since we have 2 uncertain parameters β , γ we use the Sparse Grid Points method. The accuracy level of this sparse grid method is defined by a parameter k which can

vary from 1 to 4. To take an adequate number of samples we apply k=3 which means that 13 points are used on the parameter space. Details of the points with the weights are given in Table 3.

Table 3: Sparse Grid Points for β and γ

β	γ	weights
0,0027	0,3000	0,1667
0,0100	0,1500	0,2500
0,0100	0,3000	-0,5000
0,0100	0,4500	0,2500
0,0200	0,0402	0,1667
0,0200	0,1500	-0,5000
0,0200	0,3000	1,3333
0,0200	0,4500	-0,5000
0,0200	0,5598	0,1667
0,0300	0,1500	0,2500
0,0300	0,3000	-0,5000
0,0300	0,4500	0,2500
0,0373	0,3000	0,1667

In Figures 14 and 15 the inferred sequence of optimal and worst sensor placements, respectively, is presented. Results are obtained from both the FSSP and BSSP and are presented in these figures. In Figure 16, the utility function (robust information entropy) as a function of the number of sensors placed at the optimal and worst sensor locations is presented. Results are also tabulated in Table 4.

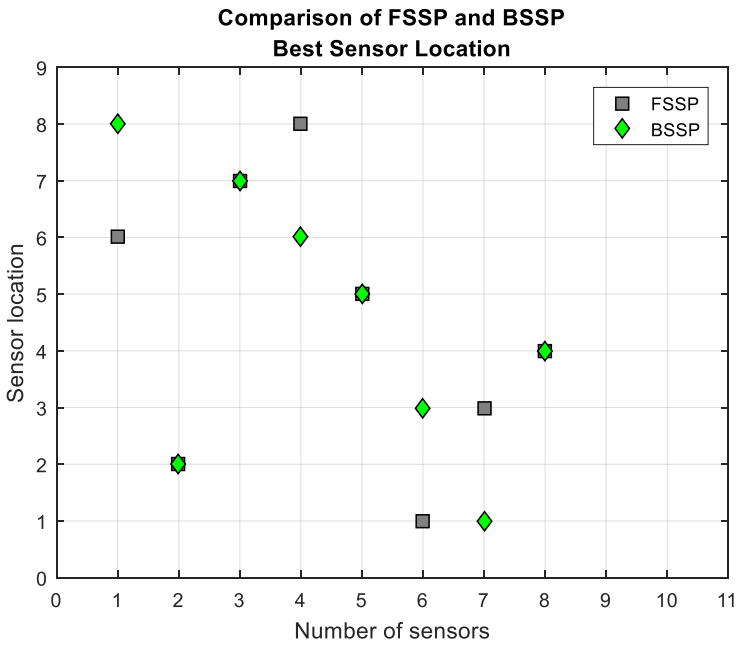


Figure 14: Best sensor placements

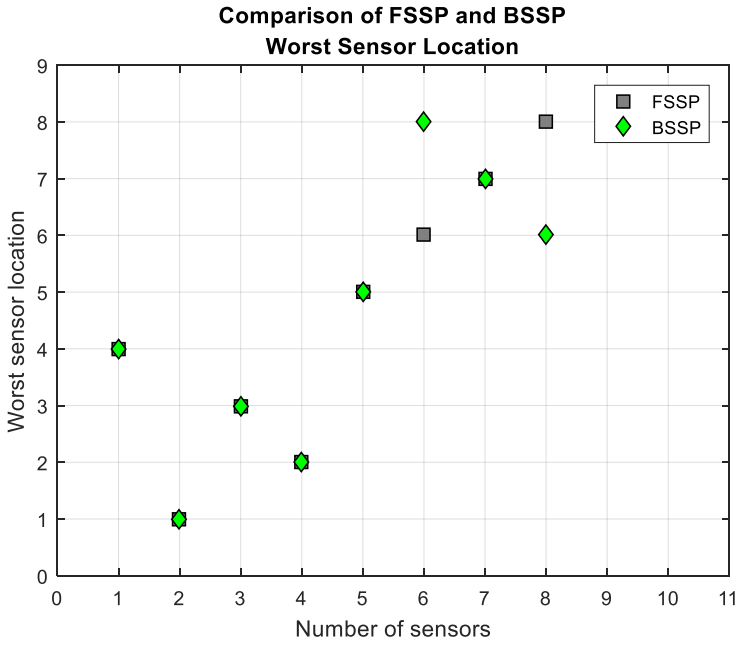


Figure 15: Worst sensor placements

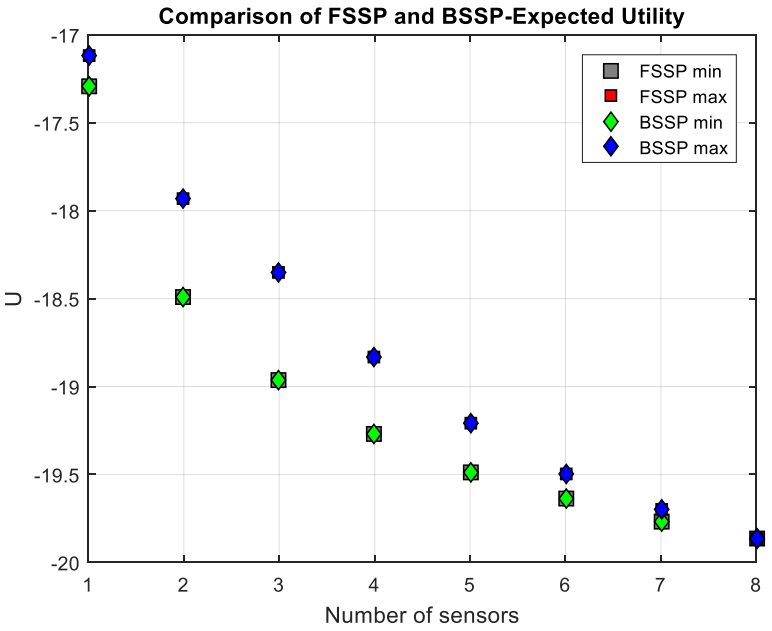


Figure 16: Maximum and Minimum Utility Function for each sensor placement

Table 4: Utility function results derived from the two methods

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-17,2917	-17,2917	-17,1141	-17,1141
-18,4944	-18,4944	-17,9324	-17,9324
-18,9632	-18,9632	-18,3539	-18,3540
-19,2674	-19,2674	-18,8361	-18,8361
-19,4870	-19,4870	-19,2081	-19,2081
-19,6403	-19,6403	-19,4933	-19,4933
-19,7644	-19,7644	-19,6999	-19,6999
-19,8643	-19,8643	-19,8643	-19,8643

- We can observe that FSSP and BSSP analysis result to identical values for the expected Utility function.
- However as far as the sensor placements preferences are concerned there exist some discrepancies. That indicates that there exist equivalent sequences of sensor placements. This effect can be explained if we consider the symmetries that appear in the examined network. The two inferred sequences differ because FSSP prefers to place a sensor at node 6 while BSSP selects the node 8, also FSSP appears to opt the node 1 while BSSP the node 3 and the opposites. It has been already stated that these specific places are equivalent.
- Finally, one can notice that for the last sensor placement the values of maximum and minimum Utility function are identical. This fact could be expected as only one place is remained, so the optimizing function can get only a unique value.

5.2.3 Optimal Sensor Placement for Origin Identification assuming constant parameters

Here it is assumed that the infection starting point is uncertain. That means that it can take any value from a certain region of nodes, which are nodes 1 and 2 as marked on Figure 17. On the other hand, parameters β and γ are assumed to have constant values equal to their means 0.02 and 0.3 accordingly.

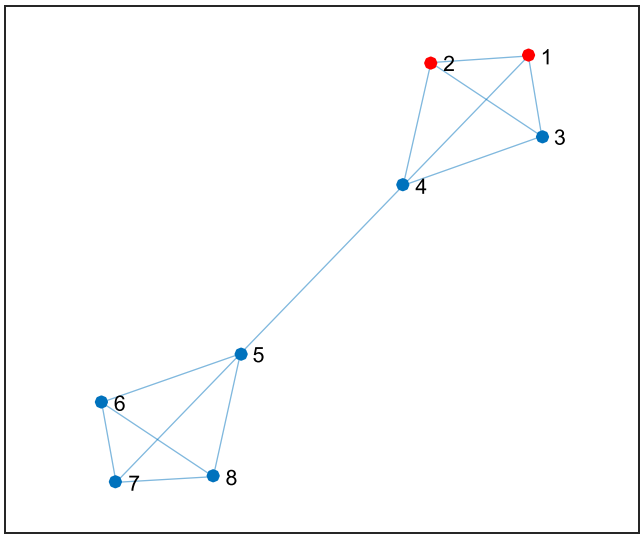


Figure 17: Examined Graph

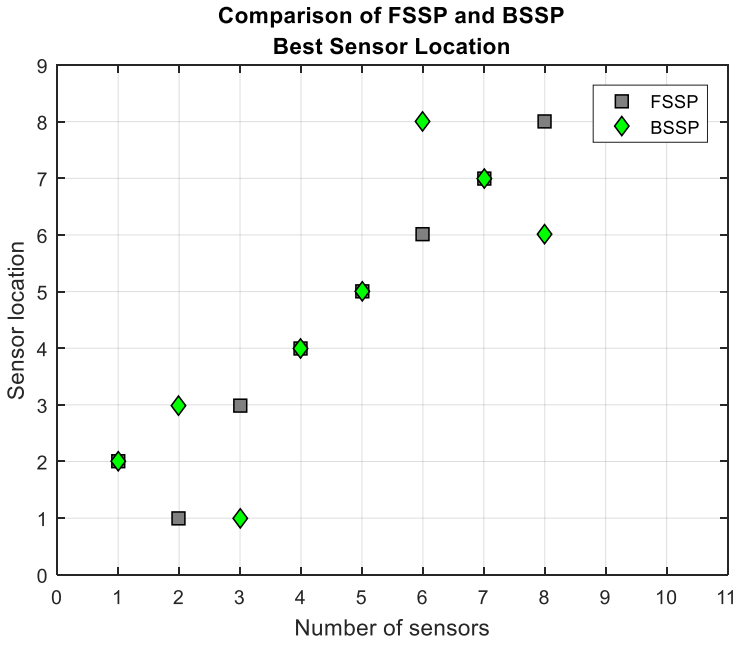


Figure 18: Best sensor placements for constant parameters β, γ

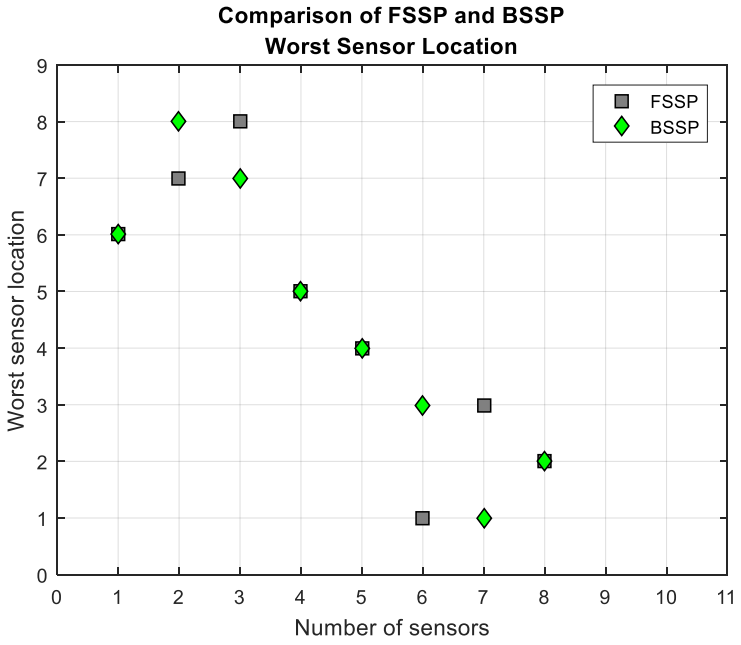


Figure 19: Worst sensor placements for constant parameters β, γ

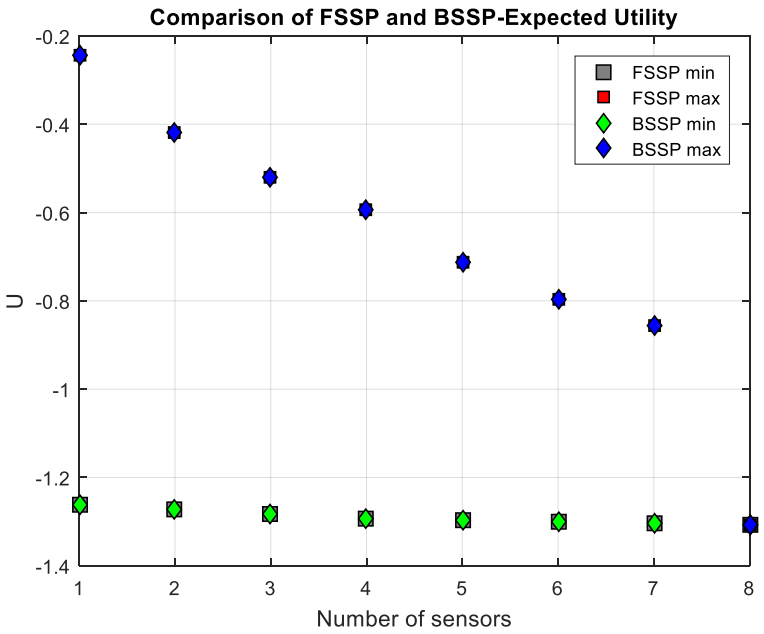


Figure 20: Maximum and Minimum Utility function for constant parameters β, γ

- Observing the diagrams one can see that the values of the expected utility function are identical for the two optimization sequential methods.
- As for the best and worst sensor location at each sensor placement we can state that in some cases they are the same while in other they differ. Specifically, both methods indicate the node 2 as the first place preferred. However for the placement of the second sensor FSSP recommends node 1, while BSSP recommends node 3. Since the values of the minimum expected utility functions are equals the two placements are equivalent.
- We also have to notice that the first node to be preferred is the one of the potential starting points. FSSP places the second sensor at the other potential starting point, whereas BSSP places there the third sensor.

5.2.4 Optimal Sensor Placement for Origin Identification for uncertain parameters β, γ

In this process β and γ are assumed to be uncertain. Sparse grid point method is used. The accuracy level is the one used in the section 5.2 (k=3).

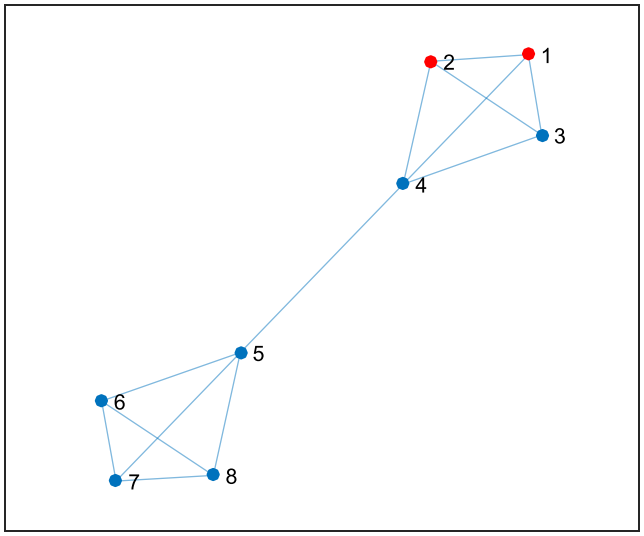


Figure 21: Examined graph

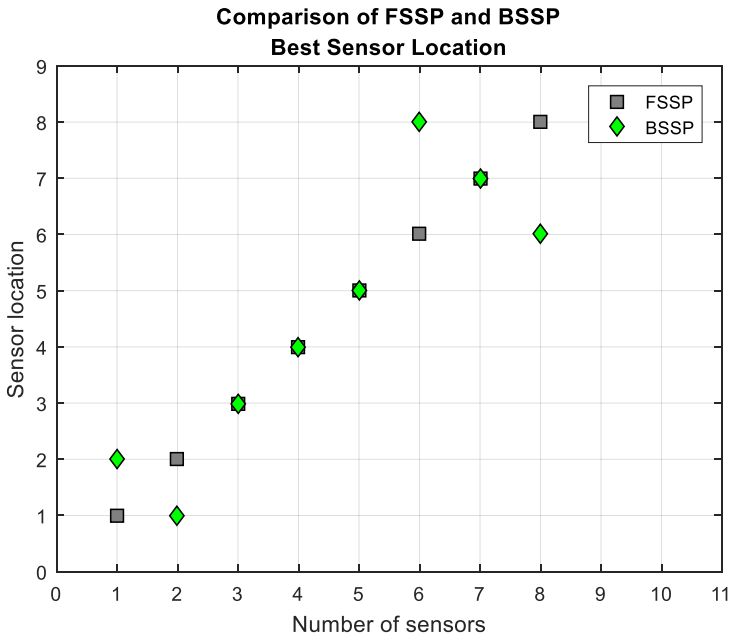


Figure 22: Best sensor placements for constant parameters β, γ

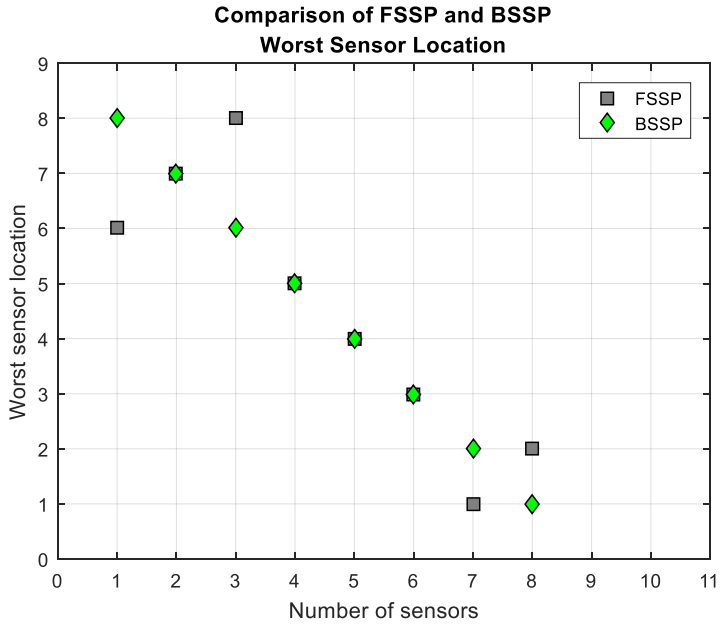


Figure 23: Worst sensor placements for constant parameters β, γ

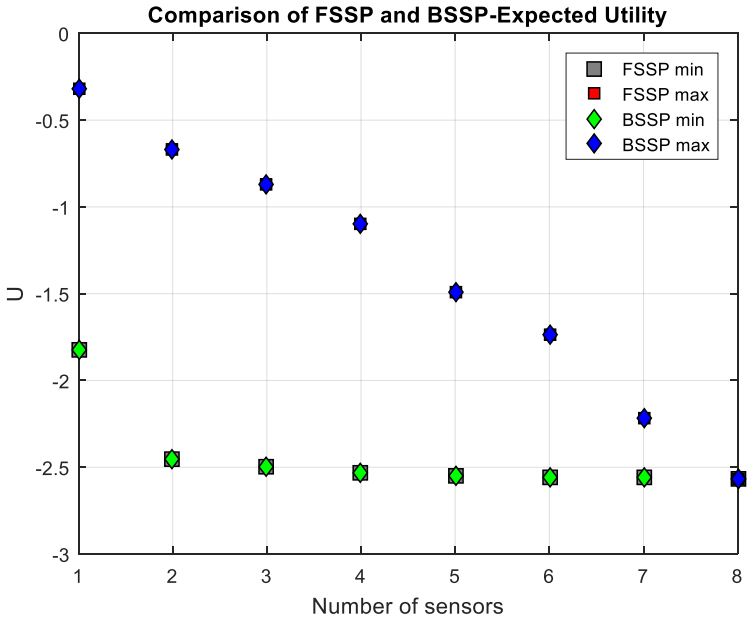


Figure 24: Maximum and Minimum Utility function for constant parameters β, γ

Table 5: Utility function results derived from the two methods

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-1,8256	-1,8256	-0,3217	-0,3217
-2,4531	-2,4531	-0,6683	-0,6683
-2,5009	-2,5009	-0,8710	-0,8710
-2,5363	-2,5363	-1,1012	-1,1012
-2,5469	-2,5469	-1,4891	-1,4891
-2,5548	-2,5548	-1,7348	-1,7348
-2,5624	-2,5624	-2,2136	-2,2136
-2,5699	-2,5699	-2,5699	-2,5699

- Implementing this methodology the two sequential algorithms propose the potential sources of infection for the first two sensor placements. Following they prefer their neighboring nodes and finally the farther ones.
- Discrepancies between the two methods' preferences are performed again because of the system symmetries.
- Figures of best and worst sensor placements appear to be antisymmetric.
- Moreover, it must be pointed out that the value of minimum expected utility function decreases considerably during the placement of the second sensor. Oppositely, afterward it appears to be approximately constant. This implies that the use of more than three sensors is senseless as it doesn't provide us with further information.

5.2.5 Optimal Sensor Placement for Origin Identification for uncertain transition rates

We consider uncertainty in the transition rates between communities, namely in the elements of Λ, H and G matrices. In this section we assume that all the elements of each matrix are perfectly correlated yielding three uncertain parameters. We implement Sparse Grid Points' method. We use accuracy level k=3. 25 sets of transition rates are produced.

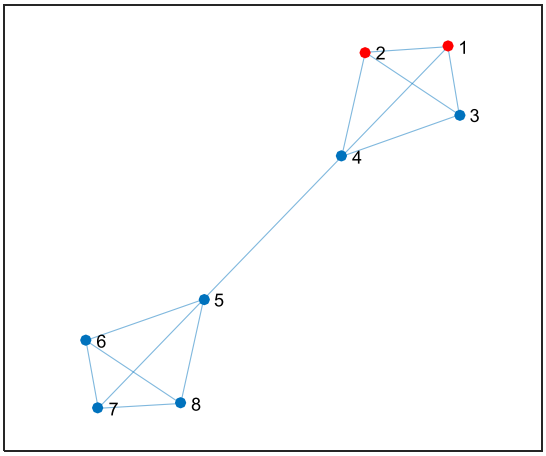


Figure 25: Examined graph

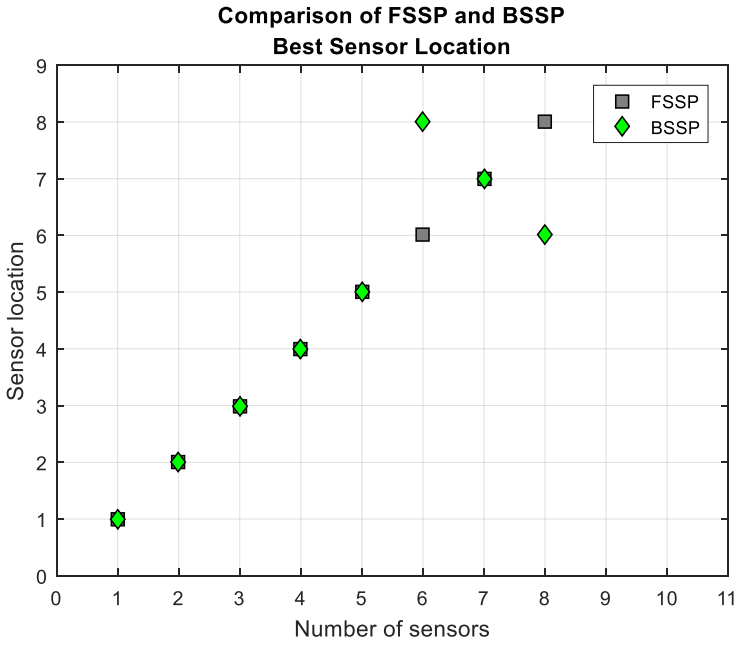


Figure 26: Best sensor placements for uncertain transition rates

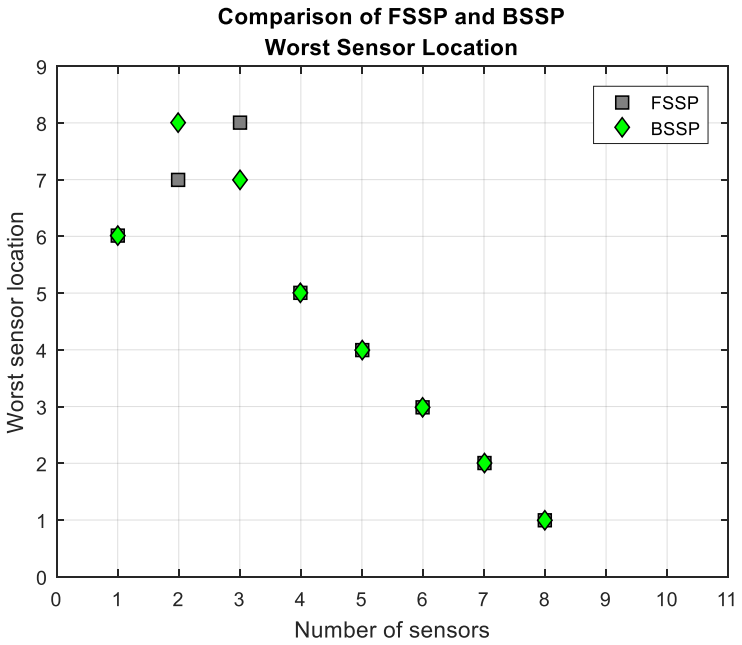


Figure 27: Best sensor placements for uncertain transition rates

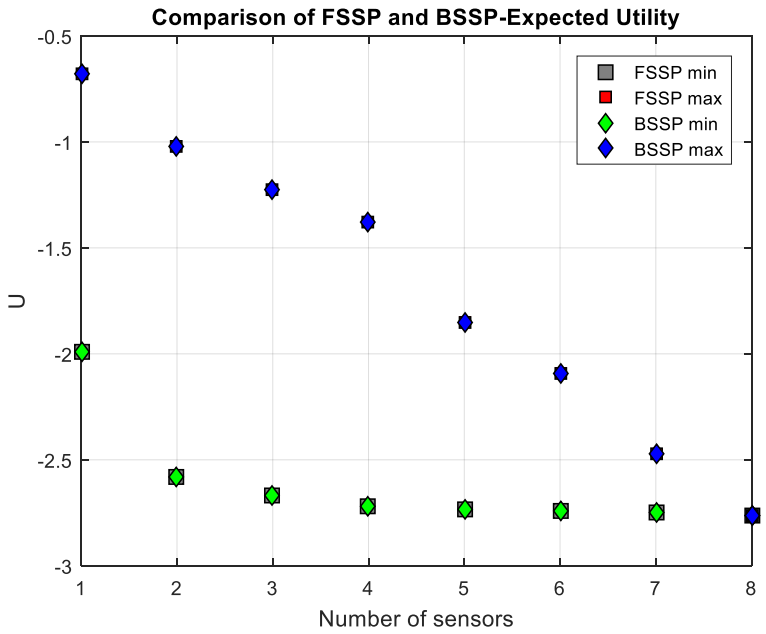


Figure 28: Maximum and Minimum Utility function for uncertain transition rates

Table 6: Utility function results derived from the two methods

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-1,9940	-1,9940	-0,6776	-0,6776
-2,5782	-2,5782	-1,0241	-1,0241
-2,6675	-2,6675	-1,2269	-1,2269
-2,7220	-2,7220	-1,3817	-1,3817
-2,7322	-2,7322	-1,8541	-1,8541
-2,7417	-2,7417	-2,0947	-2,0947
-2,7511	-2,7511	-2,4732	-2,4732
-2,7603	-2,7603	-2,7603	-2,7603

- Maximum and minimum values for the Utility function inferred from the two sequential methods coincide.
- The places preferred from both methods are in accordance apart from 2 cases, which can easily be explicated regarding problem's symmetries.
- Both methods select firstly the potential starting points and following the neighboring ones.
- The evolution of minimum Utility according to the number of sensors used indicates that two or three are enough to elicit information about the state of the system. Let's remind here that the Utility quantifies the information entropy. When its minimum value stays approximately constant after a sensor placement it is implied that the usage of that sensor is meaningless.

To validate our results a complementary process takes place. Transition rates are concerned to change independently. Monte Carlo Integration method is applied. Covariance is chosen to

be 0,2. To elicit reliable results 10000 sets of transition rates are generated. Two of them are displayed below to give an illustration to of parameters' fluctuation:

1st set

G_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0512	0,0506	0,0506	0	0	0	0
2	0,0493	0	0,0512	0,0511	0	0	0	0
3	0,0474	0,0445	0	0,0511	0	0	0	0
4	0,0470	0,0515	0,0498	0	0,0508	0	0	0
5	0	0	0	0,0495	0	0,0500	0,0547	0,0513
6	0	0	0	0	0,0510	0	0,0500	0,0508
7	0	0	0	0	0,0466	0,0485	0	0,0508
8	0	0	0	0	0,0492	0,0515	0,0497	0

H_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,2977	0,2840	0,3087	0	0	0	0
2	0,2870	0	0,2911	0,2877	0	0	0	0
3	0,2940	0,2920	0	0,2986	0	0	0	0
4	0,2943	0,3078	0,2906	0	0,2915	0	0	0
5	0	0	0	0,3043	0	0,3098	0,3068	0,3031
6	0	0	0	0	0,2991	0	0,2918	0,2961
7	0	0	0	0	0,2934	0,3063	0	0,2992
8	0	0	0	0	0,2833	0,2894	0,2874	0

Λ_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0192	0,0211	0,0201	0	0	0	0
2	0,0195	0	0,0199	0,0206	0	0	0	0
3	0,0221	0,0194	0	0,0214	0	0	0	0
4	0,0194	0,0201	0,0200	0	0,0201	0	0	0
5	0	0	0	0,0213	0	0,0193	0,0206	0,0187
6	0	0	0	0	0,0192	0	0,0193	0,0203
7	0	0	0	0	0,0200	0,0214	0	0,0191
8	0	0	0	0	0,0181	0,0193	0,0196	0

2nd set

G_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0499	0,0514	0,0492	0	0	0	0
2	0,0502	0	0,0474	0,0473	0	0	0	0
3	0,0498	0,0489	0	0,05136	0	0	0	0
4	0,0479	0,0519	0,0524	0	0,0514	0	0	0
5	0	0	0	0,0495	0	0,0513	0,0495	0,0510
6	0	0	0	0	0,0457	0	0,0486	0,0488
7	0	0	0	0	0,0501	0,0504	0	0,0513
8	0	0	0	0	0,0490	0,0514	0,0471	0

H_{ij}	1	2	3	4	5	6	7	8
1	0	0,3135	0,2999	0,3025	0	0	0	0
2	0,2869	0	0,2912	0,3212	0	0	0	0
3	0,2948	0,3131	0	0,2874	0	0	0	0
4	0,2978	0,2981	0,3122	0	0,2964	0	0	0
5	0	0	0	0,2971	0	0,2914	0,2933	0,3157
6	0	0	0	0	0,3011	0	0,2831	0,3057
7	0	0	0	0	0,2919	0,3021	0	0,3134
8	0	0	0	0	0,3226	0,2978	0,3049	0

Λ_{ij}	1	2	3	4	5	6	7	8
1	0	0,0199	0,0203	0,0196	0	0	0	0
2	0,0187	0	0,0207	0,0187	0	0	0	0
3	0,0186	0,0185	0	0,0196	0	0	0	0
4	0,0208	0,0202	0,0190	0	0,02024	0	0	0
5	0	0	0	0,0218	0	0,0202	0,0213	0,020
6	0	0	0	0	0,0191	0	0,0194	0,0206
7	0	0	0	0	0,0200	0,0198	0	0,0195
8	0	0	0	0	0,0208	0,0196	0,0203	0

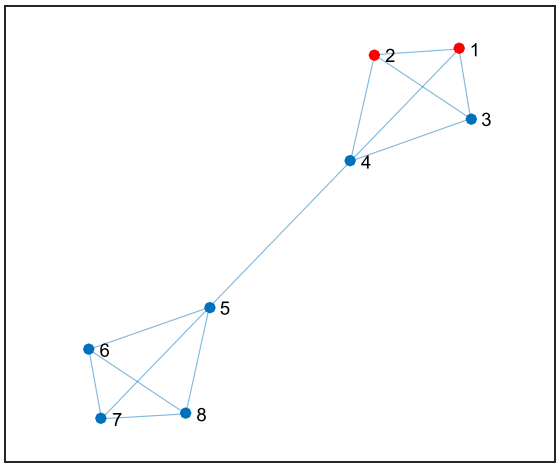


Figure 29: Examined graph

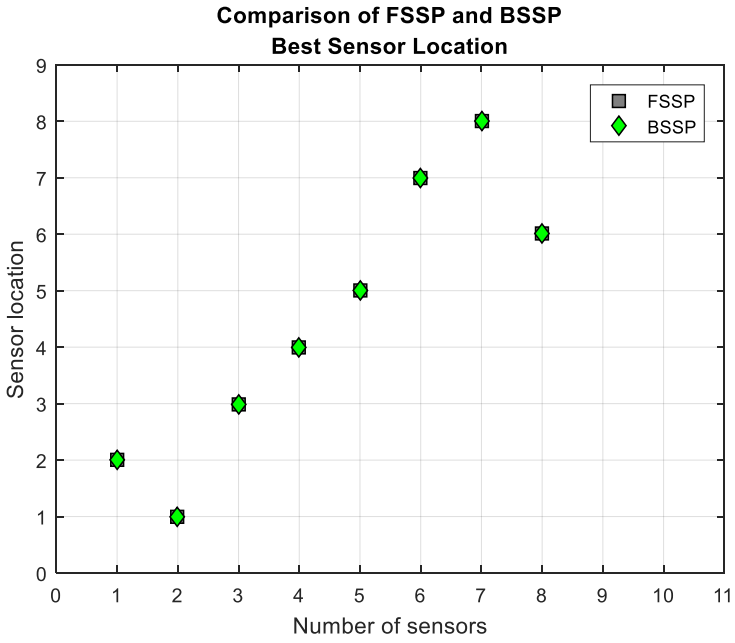


Figure 30: Best sensor placements for uncertain transition rates

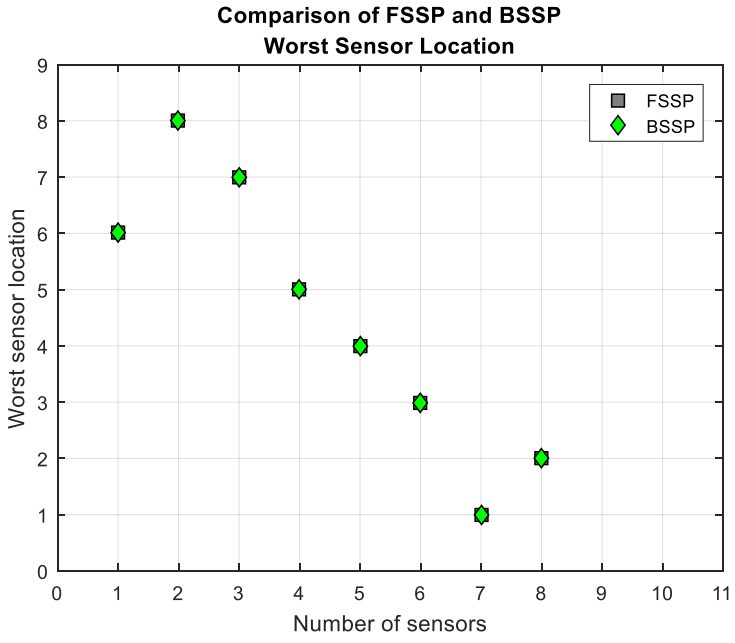


Figure 31: Worst sensor placements for uncertain transition rates

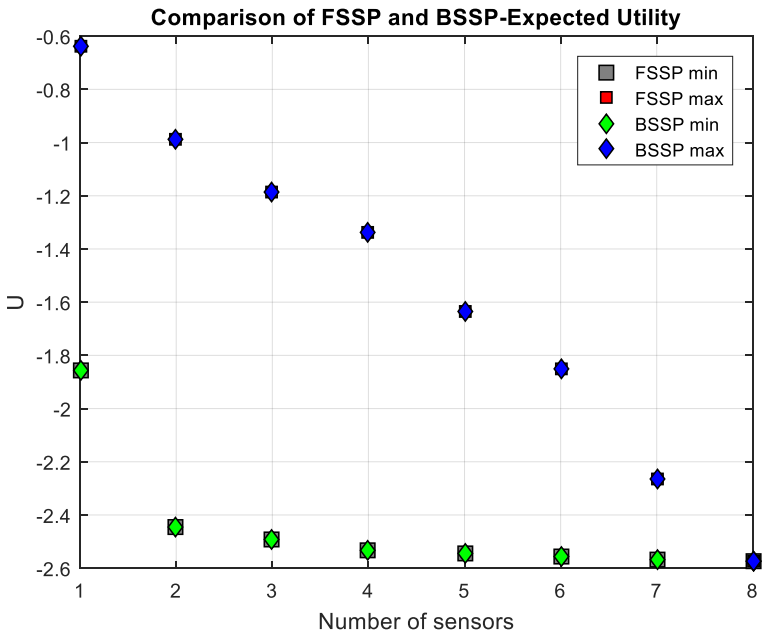


Figure 32: Maximum and Minimum Utility function for uncertain transition rates

Table 7: Utility function results derived from the two method

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-1,8553	-1,8553	-0,6389	-0,6389
-2,4437	-2,4437	-0,9857	-0,9857
-2,4942	-2,4942	-1,1886	-1,1886
-2,5326	-2,5326	-1,3361	-1,3361
-2,5442	-2,5442	-1,6325	-1,6325
-2,5551	-2,5551	-1,8480	-1,8480
-2,5659	-2,5659	-2,2659	-2,2659
-2,5764	-2,5764	-2,5764	-2,5764

- Results derived from BSSP and FSSP are identical.
- Contrasting the two sequential procedures one can point out that identical results are inferred.
- The results appears to be very reasonable. The potential starting points are opted first, then the adjacent nodes and finally the most remoted ones.
- The Utility does not significantly decrease after the second sensor placement, which implies that 2 sensors are adequate to infer sufficient information.
- Comparing Sparse Grid method and Monte Carlo Integration, one can noticed that equivalent sensor placement sequences are resulted. Discrepancies like the fact that the first technique selected firstly the node 1, while the second technic the node 2 can be attributed to system symmetries.

Following we consider the case that the infection can have started at any node of the two populations. Moreover we presume that there is uncertainty about the transition rates only in the first community. The point of this experiment is to observe how the preference for sensor places is dependent on parameters' uncertainty. Independent transition rates are assumed. **Monte Carlo Integration** is implemented. We use 10000 samples.

1st set

G_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0495	0,0538	0,0451	0	0	0	0
2	0,0476	0	0,0505	0,0474	0	0	0	0
3	0,0524	0,0506	0	0,0459	0	0	0	0
4	0,05090	0,0489	0,0516	0	0,0500	0	0	0
5	0	0	0	0,0500	0	0,0500	0,0500	0,0500
6	0	0	0	0	0,0500	0	0,0500	0,0500
7	0	0	0	0	0,0500	0,0500	0	0,0500
8	0	0	0	0	0,0500	0,0500	0,0500	0

H_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,3119	0,3100	0,2870	0	0	0	0
2	0,2805	0	0,2977	0,2987	0	0	0	0
3	0,2911	0,2937	0	0,2809	0	0	0	0
4	0,2979	0,3042	0,2928	0	0,3000	0	0	0
5	0	0	0	0,3000	0	0,3000	0,3000	0,3000
6	0	0	0	0	0,3000	0	0,3000	0,3000
7	0	0	0	0	0,3000	0,3000	0	0,3000
8	0	0	0	0	0,3000	0,3000	0,3000	0

Λ_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0197	0,0196	0,0194	0	0	0	0
2	0,0185	0	0,0203	0,0197	0	0	0	0
3	0,0202	0,0192	0	0,0200	0	0	0	0
4	0,0181	0,0210	0,0196	0	0,0200	0	0	0
5	0	0	0	0,0200	0	0,0200	0,0200	0,0200
6	0	0	0	0	0,0200	0	0,0200	0,0200
7	0	0	0	0	0,0200	0,0200	0	0,0200
8	0	0	0	0	0,0200	0,0200	0,0200	0

2nd set

G_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0517	0,0459	0,0502	0	0	0	0
2	0,0475	0	0,0483	0,0507	0	0	0	0
3	0,0461	0,0515	0	0,0503	0	0	0	0
4	0,0503	0,0486	0,0508	0	0,0500	0	0	0

5	0	0	0	0,0500	0	0,0500	0,0500	0,0500
6	0	0	0	0	0,0500	0	0,0500	0,0500
7	0	0	0	0	0,0500	0,0500	0	0,0500
8	0	0	0	0	0,0500	0,0500	0,0500	0

H_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,2888	0,2907	0,2894	0	0	0	0
2	0,2831	0	0,3012	0,3104	0	0	0	0
3	0,2953	0,2897	0	0,3049	0	0	0	0
4	0,3079	0,3010	0,2979	0	0,3000	0	0	0
5	0	0	0	0,3000	0	0,3000	0,3000	0,3000
6	0	0	0	0	0,3000	0	0,3000	0,3000
7	0	0	0	0	0,3000	0,3000	0	0,3000
8	0	0	0	0	0,3000	0,3000	0,3000	0

Λ_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0200	0,0195	0,0193	0	0	0	0
2	0,0188	0	0,0189	0,0190	0	0	0	0
3	0,0201	0,0210	0	0,0196	0	0	0	0
4	0,0207	0,0188	0,0201	0	0,0200	0	0	0
5	0	0	0	0,0200	0	0,0200	0,0200	0,0200
6	0	0	0	0	0,0200	0	0,0200	0,0200
7	0	0	0	0	0,0200	0,0200	0	0,0200
8	0	0	0	0	0,0200	0,0200	0,0200	0

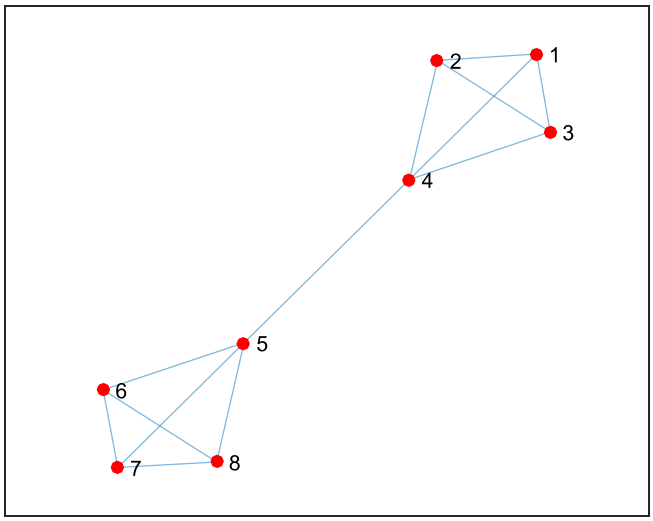


Figure 33: Examined graph

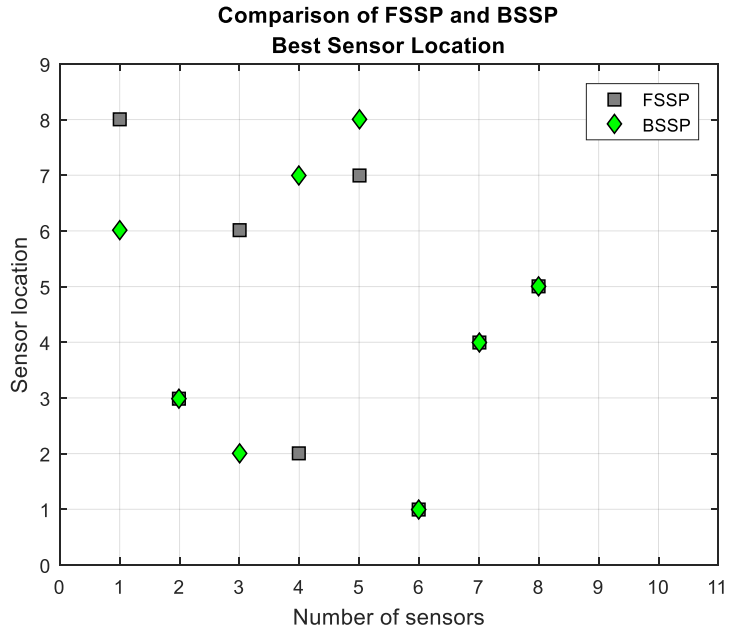


Figure 34: Best sensor placements

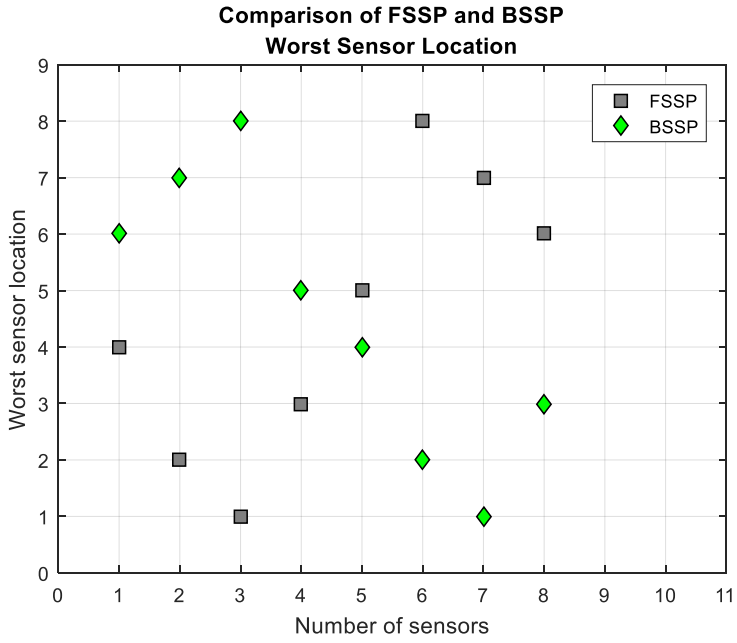


Figure 35: Worst sensor placements

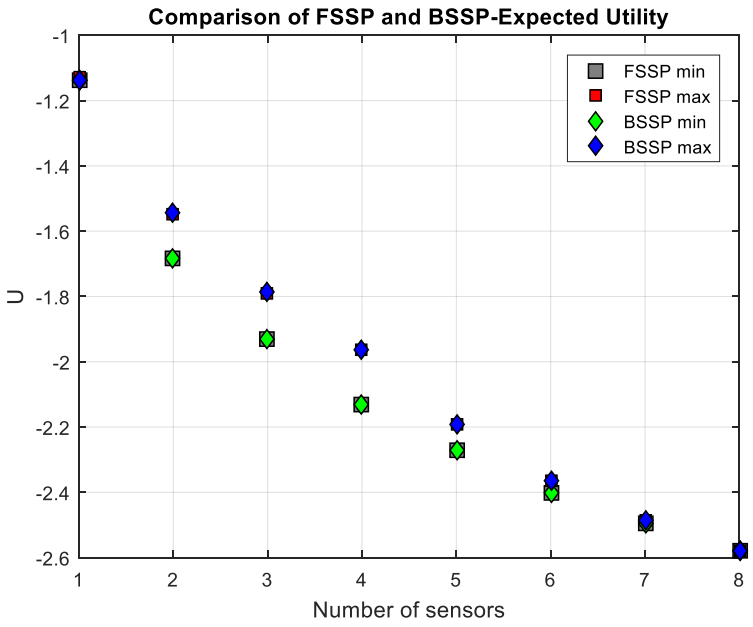


Figure 36: Maximum and Minimum Utility

Table 8: Utility function results derived from the two methods

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-1,1373	-1,1373	-1,1274	-1,1373
-1,6841	-1,6841	-1,5463	-1,5443
-1,9296	-1,9296	-1,7913	-1,7867
-2,1326	-2,1326	-1,9646	-1,9647
-2,2732	-2,2732	-2,1911	-2,1912
-2,3997	-2,3997	-2,3638	-2,3638
-2,4948	-2,4948	-2,4853	-2,4853
-2,5790	-2,5790	-2,5790	-2,5790

- Sensors are placed once in the second and once in the first population. Results of both sequential sensor placement methods coincide when sensor are placed both at the first population, while they differ when the second population is preferred.
- Minimum Utility function is constantly decreasing which implies that the more sensors we are using the more information we infer.
- From the results in Figure 36 (compare with Figure 32), the difference between best and worst utility function is small, indicating that the optimizing the location of a given number of sensors does not significantly improve results in this case. However, adding more sensors is important since the improvement in the utility function is significant.

We continue by repeating the same example, but this time with the assumption that the values of the elements of each matrix are perfectly correlated. We assume again that **uncertainty** related to the transition rates exists **only in the first community**. As long as the

values of the elements of transition matrices are perfectly correlated, we have uncertainties at three parameters regarding to our model characteristics. We apply the **sparse grid points' method** to create our samples. The utilized accuracy level is k=3.

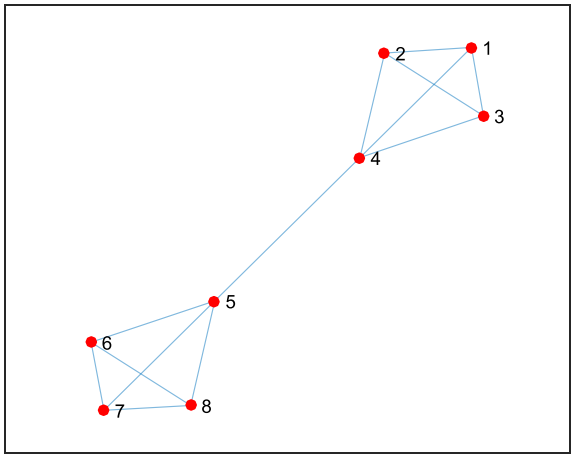


Figure 37: Examined graph

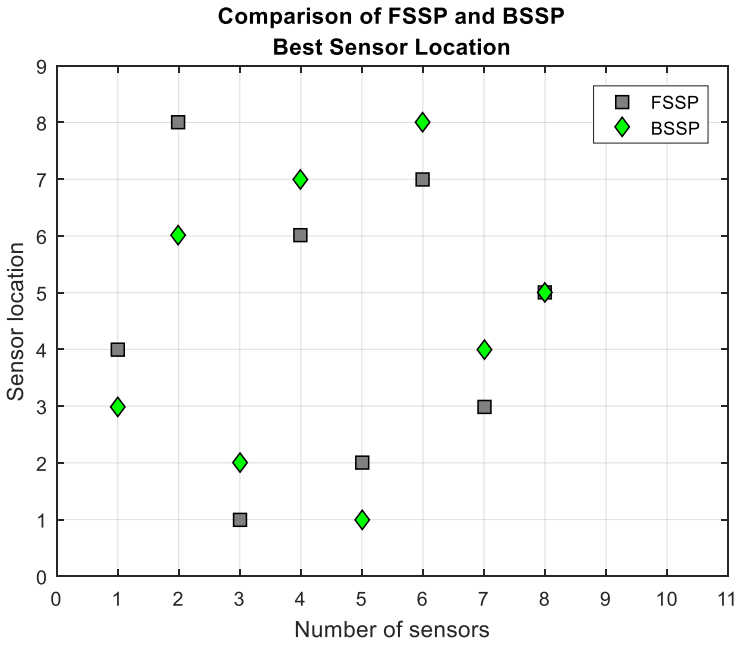


Figure 38: Best sensor placement

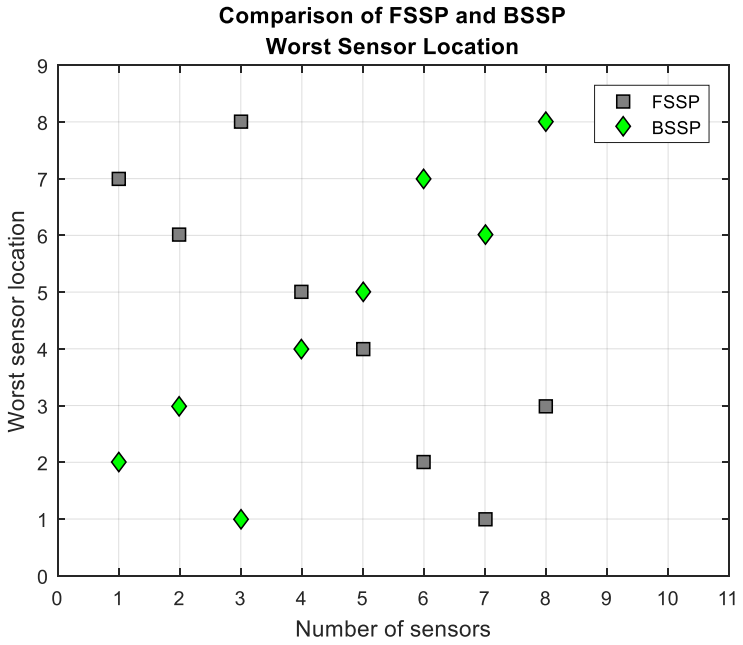


Figure 39: Worst sensor placement

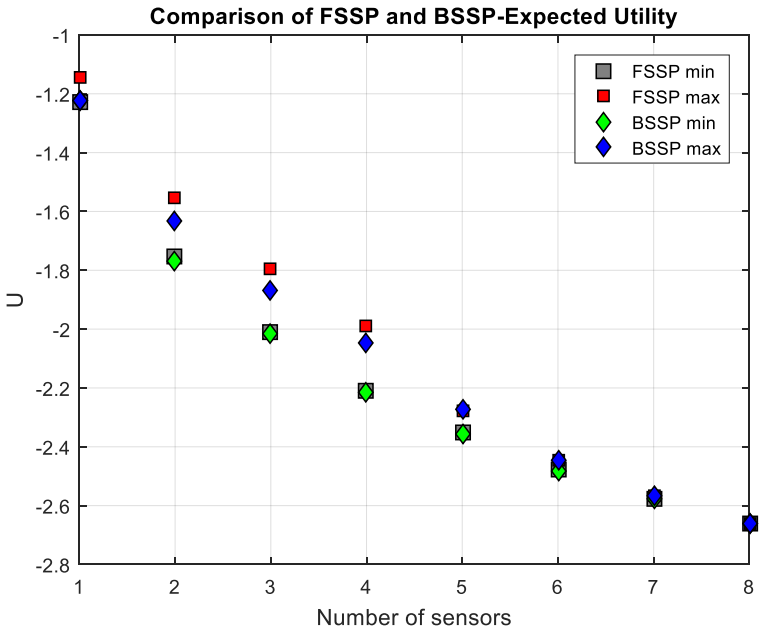


Figure 40: Maximum and Minimum Utility function

Table 9: Utility function results derived from the two methods

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-1,2284	-1,2250	-1,1463	-1,2250
-1,7553	-1,7678	-1,5533	-1,6308
-2,0106	-2,0148	-1,7957	-1,8706
-2,2079	-2,2157	-1,9890	-2,0486
-2,3524	-2,3554	-2,2760	-2,2740
-2,4757	-2,4806	-2,4463	-2,4457
-2,5758	-2,5758	-2,5666	-2,5664
-2,6593	-2,6593	-2,6593	-2,6593

- Best sensor locations are cited once in the first population and once in the second.
- One can notice that FSSP and BSSP are in accordance on choosing each time nodes from the same population, either from the one where we assumed uncertainty or from the other.
- Concerning the fact that supposing all the nodes to be potentially the infection source, we understand that the only factor of asymmetry is that not all transition rates are assumed to be uncertainty. Inconsistencies between best and worst sensor location figures as well as between Monte Carlo Integrations' and Sparse Grid Points Methods' results can be attributed to this effect.
- It is also important that the results of both methods, Monte Carlo Integration and Sparse Grid Points Method, do not perform any preference to one of the two populations. Neither to the one with the uncertain transition rates nor to the other one.
- One can notice that Figures 28 and 32 are completely different from Figure 36 or 40. This fact can be easily explained if we take into account that in the first two cases we assume that the region where the infection has started consists of two nodes, while in the last two cases we assume that the infection can have started in any node. Figures 28 and 32 suggest that for given number of sensors it is important to find the optimal sensor configuration. They also both suggest that two sensors placed at their optimal location are enough to get most of the information. In contrast, figures 36 and 40 suggest that for given number of sensors it is not important to find the optimal sensor location because the difference between the best and the worst utility value is small. However it is important to use more and more sensors.

5.2.6 Optimal Sensor Placement for Origin Identification for uncertain transition rates in an asymmetric Network

Having observed that the two methods, FSSP and BSSP, tend to prefer nodes that are symmetrically placed in the network we consider to break that symmetry. To address that issue **we assume that the transition rates** given from the values of the elements of matrices **Λ, H, G differ.** Moreover we assume that they fluctuate independently around theirs means. **Monte Carlo Integration** is applied and 10000 samples are generated. The modified set of transition rates is given by the matrices bellow:

$$G = \begin{bmatrix} 0 & 0,0500 & 0,1000 & 0,0200 & 0 & 0 & 0 & 0 \\ 0,0500 & 0 & 0,0250 & 0,0500 & 0 & 0 & 0 & 0 \\ 0,1000 & 0,0250 & 0 & 0,1000 & 0 & 0 & 0 & 0 \\ 0,0200 & 0,0500 & 0,1000 & 0 & 0,0050 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0050 & 0 & 0,0300 & 0,0600 & 0,0150 \\ 0 & 0 & 0 & 0 & 0,0300 & 0 & 0,1000 & 0,0600 \\ 0 & 0 & 0 & 0 & 0,0600 & 0,1000 & 0 & 0,0750 \\ 0 & 0 & 0 & 0 & 0,0150 & 0,0600 & 0,0750 & 0 \end{bmatrix}$$

$$H = \begin{bmatrix} 0 & 0,3000 & 0,6000 & 0,1200 & 0 & 0 & 0 & 0 \\ 0,3000 & 0 & 0,1500 & 0,3000 & 0 & 0 & 0 & 0 \\ 0,6000 & 0,1500 & 0 & 0,6000 & 0 & 0 & 0 & 0 \\ 0,1200 & 0,3000 & 0,6000 & 0 & 0,0300 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0300 & 0 & 0,1800 & 0,3600 & 0,0900 \\ 0 & 0 & 0 & 0 & 0,1800 & 0 & 0,6000 & 0,3600 \\ 0 & 0 & 0 & 0 & 0,3600 & 0,6000 & 0 & 0,4500 \\ 0 & 0 & 0 & 0 & 0,9000 & 0,3600 & 0,4500 & 0 \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} 0 & 0,0200 & 0,0400 & 0,0080 & 0 & 0 & 0 & 0 \\ 0,0200 & 0 & 0,0100 & 0,0200 & 0 & 0 & 0 & 0 \\ 0,0400 & 0,0100 & 0 & 0,0400 & 0 & 0 & 0 & 0 \\ 0,0080 & 0,0200 & 0,0400 & 0 & 0,0020 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0020 & 0 & 0,0120 & 0,0240 & 0,0060 \\ 0 & 0 & 0 & 0 & 0,0120 & 0 & 0,0400 & 0,0240 \\ 0 & 0 & 0 & 0 & 0,0240 & 0,0400 & 0 & 0,0300 \\ 0 & 0 & 0 & 0 & 0,0060 & 0,0240 & 0,0300 & 0 \end{bmatrix}$$

The transition rates from one node to another and the opposite are assumed to be equals.

1st set

G_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0530	0,1005	0,0195	0	0	0	0
2	0,0498	0	0,0258	0,0522	0	0	0	0
3	0,0987	0,0242	0	0,0982	0	0	0	0
4	0,02045	0,05054	0,1066	0	0,0049	0	0	0
5	0	0	0	0,0052	0	0,0301	0,0584	0,0143
6	0	0	0	0	0,0283	0	0,1020	0,0636
7	0	0	0	0	0,0595	0,0992	0	0,0783

8	0	0	0	0	0,0147	0,0586	0,0697	0

H_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,2300	0,6104	0,1270	0	0	0	0
2	0,2992	0	0,1437	0,2956	0	0	0	0
3	0,5819	0,1411	0	0,6509	0	0	0	0
4	0,12711	0,3074	0,6300	0	0,0290	0	0	0
5	0	0	0	0,0299	0	0,1866	0,3624	0,0944
6	0	0	0	0	0,1675	0	0,6196	0,3828
7	0	0	0	0	0,3391	0,579	0	0,4247
8	0	0	0	0	0,0950	0,3626	0,4562	0

L_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0214	0,0396	0,0073	0	0	0	0
2	0,0194	0	0,0099	0,0197	0	0	0	0
3	0,0398	0,0094	0	0,0387	0	0	0	0
4	0,0079	0,021	0,0386	0	0,0020	0	0	0
5	0	0	0	0,0019	0	0,0123	0,0242	0,0063
6	0	0	0	0	0,0115	0	0,0377	0,0244
7	0	0	0	0	0,0228	0,0390	0	0,0297
8	0	0	0	0	0,0059	0,0231	0,0299	0

2nd set

G_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,0498	0,1024	0,0192	0	0	0	0
2	0,0462	0	0,0259	0,0486	0	0	0	0
3	0,0971	0,0265	0	0,1036	0	0	0	0
4	0,0205	0,0488	0,1036	0	0,0053	0	0	0
5	0	0	0	0,0045	0	0,0290	0,0593	0,0134
6	0	0	0	0	0,0306	0	0,1005	0,0616
7	0	0	0	0	0,0595	0,1026	0	0,0732
8	0	0	0	0	0,0153	0,0576	0,0709	0

H_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	0,3099	0,58729	0,1137	0	0	0	0
2	0,3028	0	0,1436	0,3114	0	0	0	0
3	0,6000	0,1486	0	0,6074	0	0	0	0
4	0,1158	0,2864	0,6301	0	0,0304	0	0	0
5	0	0	0	0,0291	0	0,1770	0,3493	0,0921
6	0	0	0	0	0,1735	0	0,5990	0,3733
7	0	0	0	0	0,3664	0,5481	0	0,4611
8	0	0	0	0	0,0869	0,3634	0,4490	0

Λ_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

1	0	0,0209	0,0405	0,0080	0	0	0	0
2	0,0208	0	0,0102	0,0202	0	0	0	0
3	0,0399	0,0098	0	0,0385	0	0	0	0
4	0,0082	0,0199	0,0387	0	0,0021	0	0	0
5	0	0	0	0,0021	0	0,0116	0,0244	0,0056
6	0	0	0	0	0,0125	0	0,0403	0,0246
7	0	0	0	0	0,0224	0,0411	0	0,0273
8	0	0	0	0	0,0060	0,0259	0,0288	0

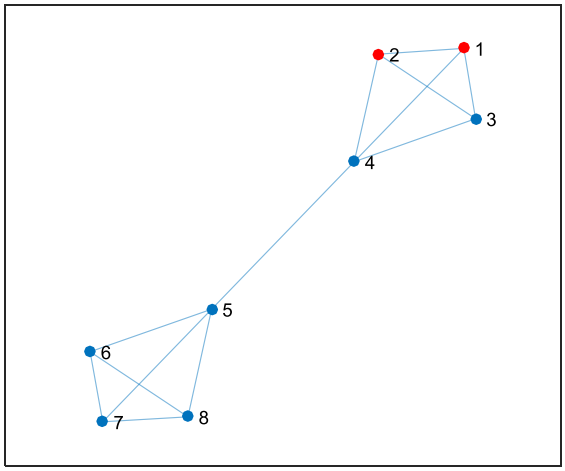


Figure 41: Examined graph

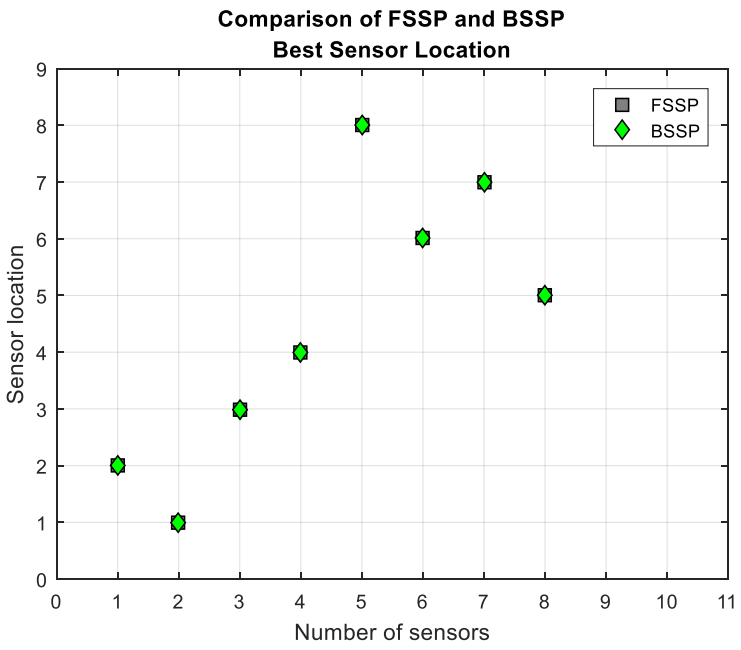


Figure 42: Best sensor placements

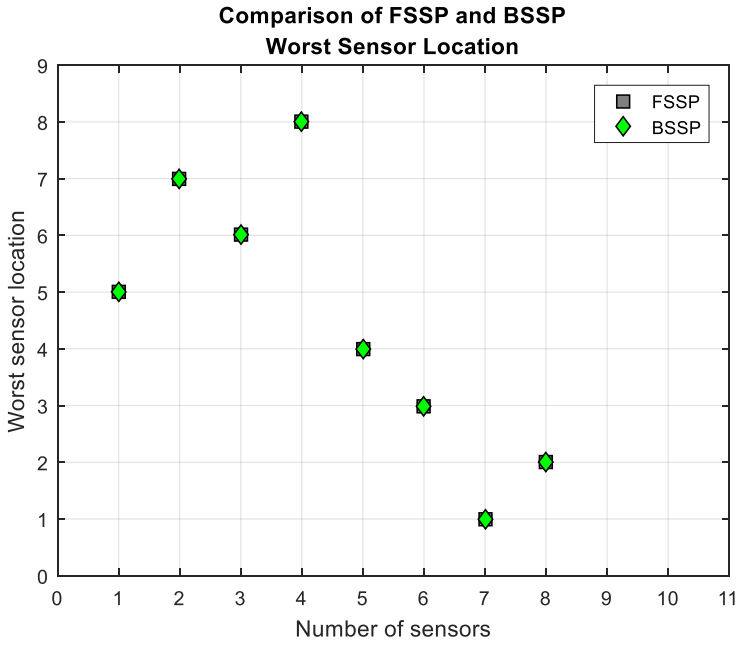


Figure 43: Worst sensor placements

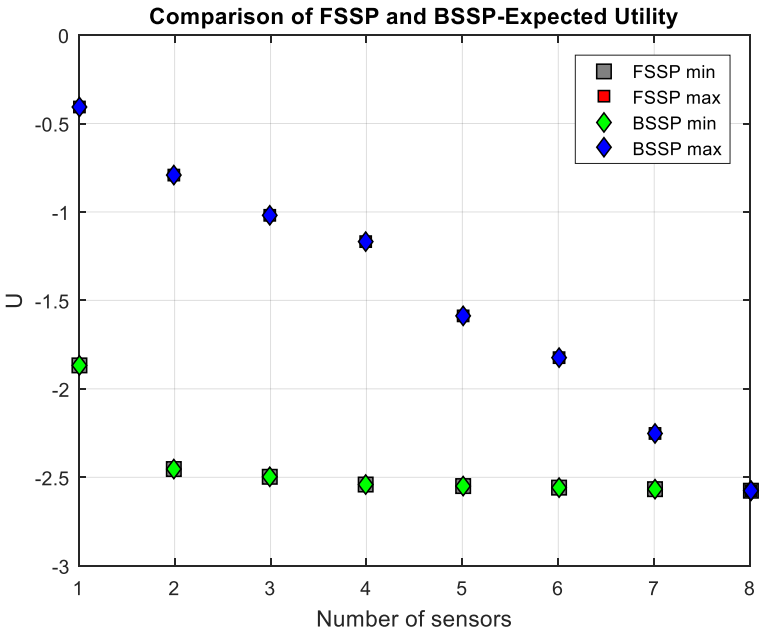


Figure 44: Maximum and Minimum Utility function

Table 10: Utility function results derived from the two methods

Umin - FSSP	Umin - BSSP	Umax - FSSP	Umax - BSSP
-1,8638	-1,8638	-0,4042	-0,4042
-2,4497	-2,4497	-0.7950	-0.7950
-2,5006	-2,5006	-1,0161	-1,0161
-2,5439	-2,5439	-1,1710	-1,1710
-2,5524	-2,5524	-1,5859	-1,5859
-2,5606	-2,5606	-1,8219	-1,8219
-2,5685	-2,5685	-2,2511	-2,2511

-2,5751	-2,5751	-2,5751	-2,5751

- It is obvious that FSSP and BSSP results coincide.
- The figures that illustrate the best and worst sensor location are antisymmetric.
- The results are particularly reasonable. The fact that no discrepancies between the results of the sequential algorithms certifies our assumption that it was caused by the existence of equivalent placements.
- Finally the evolution of minimum Utility values over extra sensor placements indicates that at least 2 sensor should be used.

CHAPTER 6

Conclusions

Two different dynamical systems where an irregularity initiates were studied throughout this thesis. Both layouts were examined from the common scope of Bayesian Analysis. However, different methods and computational tools were implemented for each case.

Firstly, a continuous medium was considered. A Bayesian methodology was applied for the detection of a fault using simulated sets of measurements. Next, the more perplex layout of an infection spread over two communities was studied. After the consideration of some logical assumptions a mathematical model was proposed to simulate its behavior.

After having presented both systems and extracted the mathematical models that describe their dynamics, a Bayesian Optimal Sensor Placement Methodology for the identification of the irregularities' origins was performed. The uncertainty which characterizes the models was quantified via the construction of an appropriate function called as Utility Function. In this point, the utility function depends in terms of the sensitivities of response quantities at the measured locations with respect to the parameters to be inferred. The sensitivities of the examined quantities were derived for the two systems. Analytical expressions for the sensitivities were derived for both systems examined. A numerical solution was taken place for the calculation of sensitivities for the second layout. In this case, a graphical illustration of inferred derivatives indicated that we approximated very well the analytical solution.

As for the implementation of Optimal Sensor Placement methodology on the problem of the continuous material, a stochastic optimization strategy, CMA-ES, was effectually applied to estimate the spatial coordinates for the optimal locations of two and three sensors. The values of the Utility function, that is the objective function formulated on the basis of Bayes Theory, appeared to decrease after the placement of the third sensor.

Furthermore, two Sequential Sensor Placement Techniques, the FSSP and BSSP, were implemented for the problem of an infection spread. Optimal sensor locations were derived. In addition to the optimal sensor locations, the computation of the Utility function indicated the optimal number of sensors that are meaningful to be placed on the system. It was clear that both FSSP and BSSP techniques gave equivalent results. The locations which were firstly preferred for the sensor placement were those in the region where the infection had started. In addition, the indicated number of sensors was in accordance with the number of the potential starting points of the infection. It should be underlined here that the inferred results

are particularly consistent to intuition, a fact that provides us with confidence for the reliability of our methodologies.

We have to note as well that through the study of two relatively simple systems we set the foundations for further investigation. Specifically, for the problem of an infection spread over a Network representing a system of interacting communities, a general software was constructed, capable to adjust easily on bigger and more perplex Networks. Moreover the applied methodologies can be readily used on a large class of dynamical systems which can be illustrated with a similar layout, namely a Network consisted of interacting elements.

Last but not least, we have to remark that the successful employment of a common approach on two altogether different layouts made clear the adaptability and usefulness of Bayesian Optimal Sensor Placement implementation on numerous real-world scenarios.

References

1. Sen, D., K. Erazo, and S. Nagarajaiah, *Bayesian estimation of acoustic emissions source in plate structures using particle-based stochastic filtering.* Structural Control and Health Monitoring, 2017. **24**(11): p. e2005.
2. Argyris, C., et al., *Bayesian optimal sensor placement for crack identification in structures using strain measurements.* Structural Control and Health Monitoring, 2018. **25**(5): p. e2137.
3. Papadimitriou, D.I. and C. Papadimitriou, *Optimal sensor placement for the estimation of turbulence model parameters in CFD.* International Journal for Uncertainty Quantification, 2015. **5**(6).
4. Kermack, W.O. and A.G. McKendrick, *A contribution to the mathematical theory of epidemics.* Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 1927. **115**(772): p. 700-721.
5. Hethcote, H.W., *The mathematics of infectious diseases.* SIAM review, 2000. **42**(4): p. 599-653.
6. Woolhouse, M., *How to make predictions about future infectious disease risks.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2011. **366**(1573): p. 2045-2054.
7. Cuadros, D.F., et al., *Mapping the spatial variability of HIV infection in Sub-Saharan Africa: Effective information for localized HIV prevention and control.* Scientific reports, 2017. **7**(1): p. 9093.
8. Shen, M., et al., *Early antiretroviral therapy and potent second-line drugs could decrease HIV incidence of drug resistance.* Proceedings of the Royal Society B: Biological Sciences, 2017. **284**(1857): p. 20170525.
9. Brady, O.J., et al., *Role of mass drug administration in elimination of Plasmodium falciparum malaria: a consensus modelling study.* The Lancet Global Health, 2017. **5**(7): p. e680-e687.
10. Eckhoff, P.A., et al., *Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics.* Proceedings of the National Academy of Sciences, 2017. **114**(2): p. E255-E264.
11. Challenger, J.D., et al., *Assessing the impact of imperfect adherence to artemether-lumefantrine on malaria treatment outcomes using within-host modelling.* Nature communications, 2017. **8**(1): p. 1373.
12. Chen, Z., et al., *Data-driven prediction and origin identification of epidemics in population networks*. 2018, Submitted.
13. Papadimitriou, C. and G. Lombaert, *The effect of prediction error correlation on optimal sensor placement in structural dynamics.* Mechanical Systems and Signal Processing, 2012. **28**: p. 105-127.
14. Hansen, N., S.D. Müller, and P. Koumoutsakos, *Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES).* Evolutionary computation, 2003. **11**(1): p. 1-18.
15. Gerstner, T. and M. Griebel, *Numerical integration using sparse grids.* Numerical algorithms, 1998. **18**(3-4): p. 209.
16. Bungartz, H.-J. and M. Griebel, *Sparse grids.* Acta numerica, 2004. **13**: p. 147-269.
17. Papadimitriou, C., J.L. Beck, and S.-K. Au, *Entropy-based optimal sensor location for structural model updating.* Journal of Vibration and Control, 2000. **6**(5): p. 781-800.
18. Papadimitriou, C., *Optimal sensor placement methodology for parametric identification of structural systems.* Journal of sound and vibration, 2004. **278**(4-5): p. 923-947.

19. Yuen, K.-V., et al., *Optimal sensor placement methodology for identification with unmeasured excitation.* Journal of dynamic systems, measurement, and control, 2001. **123**(4): p. 677-686.
20. Papadimitriou, C., *Pareto optimal sensor locations for structural identification.* Computer methods in applied mechanics and engineering, 2005. **194**(12-16): p. 1655-1673.