2017

Estimation Sensor Networks: Attacks, Applications, and Interdependencies

Basel R. Alnajjab
Lehigh University

ESTIMATION SENSOR NETWORKS: ATTACKS, APPLICATIONS, AND INTERDEPENDENCIES

by

BASEL ALNAJJAB

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Lehigh University

August, 2017

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Date

_____
Date Accepted

_____
Dissertation Advisor

Committee Members:

_____
Prof. Rick S. Blum
(Committee Chair)

_____
Prof. Alberto Lamadrid

_____
Prof. Shalinee Kishore

_____
Prof. Lawrence V. Snyder

Acknowledgements

I wish to recognize the many people to whom I owe great gratitude and appreciation for their efforts and contributions towards my development and success.

First and foremost, I express my utmost gratitude to my advisor, Professor Rick S. Blum. Getting to know and work with Prof. Blum has been one of the greatest pleasures I have experienced. His enduring support and belief in my abilities propelled me to explore the different research topics in this dissertation with confidence and resolve. Aiming to match his dedication and incredible work ethic has led me down a great and enjoyable path to success. Prof. Blum's honest critique and commitment to the highest research standards meant the details of my work would satisfy the harshest critic as he would let no subtlety go unnoticed, unexplained, or unexplored. I learned about many important topics under the direct supervision of Prof. Blum but perhaps more valuable were the indirect lessons I learned through observing his professionalism, honesty, and selflessness that I will continue to remember for the rest of my life.

I would also like to thank Prof. Alberto Lamadrid, Prof. Larry Snyder and Prof. Shalinee Kishore for serving on my committee and helping guide my work throughout my PhD. My many meetings with Prof. Lamadrid were invaluable to my progress and advancing my knowledge on power networks and ways of modeling them. Prof. Snyder's expertise drew me towards learning more about optimization techniques than I had imagined I would and his answers seemed to always come to my aide whenever I felt most stuck. Prof. Kishore's great knowledge and pursuit of new and innovative ideas helped me dive deeper into the different topics I studied with purpose and excitement.

I extend my appreciation to my colleagues Jiangfan Zhang and Mohsen Moarefdoost for their helpful discussions and insightful collaboration on our joint works. I would also like to express my sincere and special appreciation to my close friend and PORT lab colleague, Parth Pradhan, for our countless conversations on hydrodynamics, smart grids, and the many other topics we discussed.

My thanks go to my friends and colleagues at SPCRL Lab for their support, camaraderie, and enjoyable conversations: Anand Guruswamy, A. K. Karthik, Jake Perazzone, Ananth Narayan, and Alireza Famili. I am also grateful for our staff members David Morrisette, Diane Hubinsky, and Ruby Scott who were always very helpful and caring.

The importance of the support and encouragement I received outside of Lehigh could not be overstated. My enrollment at Lehigh as an undergraduate student was only made possible through the great generosity of James R. Tanenbaum and Elizabeth Scofield to whom I will always be indebted. My love and gratitude goes to the Marvel family for letting me into their warm hearts, giving me a place to call home here in the USA, and being my family away from my family.

All the great experiences and many opportunities I was able to take advantage of throughout my life would be impossible without my parents Maha Alsaheb and Rashid Alnajjab and my siblings Muhannad and Hana. Their love, encouragement, and the many great examples they set for me have shaped much of who I am today. I am forever grateful for their endless life-enriching contributions and I will always aspire to make them proud of my character, actions, and accomplishments.

Contents

List of Tables

List of Figures

Abstract

The presented research investigates topics relating to sensor systems focusing mainly on estimation. The first topic studies Byzantine attacks on sensor systems estimating the value of an unknown deterministic parameter based on quantized observations. The presented work initially describes aspects of the optimal processing under a practical family of attacks where the sensors employ bad data detectors to check if the observed sensor data fits the observation models assumed by the estimation algorithm under no attack. Next, the performance of any estimation approach employed by the sensor system under any general attack is described for cases where any number of observations, sensors, and quantization levels could be employed.

The second topic studies sensor networks focused on estimating ocean waveforms. Our work is the first to derive the Cramer-Rao bound (CRB) for the short-term forecasting of ocean waves. The CRB is a lower bound on the mean square estimation error for an unbiased estimator. The obtained results are general in the sense that they apply to a number of types of wave sensors. A low-complexity estimation method is presented along with numerical results demonstrating its accuracy and runtime performance. We also describe a method that relies on the CRB to calculate the expected loss in absorbed power, under optimal control, by a single or multiple WEC devices due to errors in the estimation of short-term future waveforms experienced by the devices.

The third topic focuses on developing a novel model accounting for the dependence of power grids on communication networks for their safe and economic operation. Our model is formulated as a two-settlement stochastic optimal power flow (OPF) problem where we account for errors in forecasting while also considering random failures in the communication network.

Our work jointly optimizes the topology of the communication network and the control actions in anticipation of the different random failures and errors that a given power grid may face. We present results that identify optimal topologies for a communication network supporting the operation of an IEEE standard 9-bus system under different conditions and we discuss some properties

of the optimal solutions.

Finally, the fourth topic studies topology estimation in power distribution networks. Accurate topology estimates are crucial for maintaining situational awareness, properly dispatching distributed energy sources, detecting cyber attacks, and many other key tasks. The presented work takes advantage of the ever increasing adoption of novel metering and sensing devices providing data from network locations that were traditionally unmonitored by grid operators. We present a topology estimation scheme for radial distribution networks relying on power flow measurements and nodal load forecasts. We also describe a sensor placement method that enables the presented scheme to identify all the detectable faults in the network. The performance of our detection scheme is then demonstrated through several numerical results on the IEEE 123-bus test feeder.

Chapter 1

Introduction

1.1 Attacks on Parameter Estimation Based on Quantized Observations

Sensor systems employed for parameter estimation have been extremely successful in applications ranging from inexpensive commercial systems to complex military and defense surveillance systems and have seen even greater interest in recent years [1]. Recent technological advances in coding, digital wireless communications technology and digital electronics have lead to the dominance of digital communications using quantized data in such systems. Hence, a great deal of attention has focused on parameter estimation using quantized data [2–9]. In some work [2–4] the unknown parameters are modeled as being random, while in other work the unknown parameters are treated as deterministic unknowns [5–9]. Most of the existing work focused on binary quantization employed for the estimation of a single unknown parameter. However, [8] and [9] considered some aspects of vector parameter cases for binary quantization.

Malicious attacks are widely regarded as one of the most serious problems faced in our dangerous world, leading to several recent publications on this topic, including [10–20] and references therein. The majority of the research related to sensor networks attempting to solve hypothesis testing and parameter estimation problems has focused on Byzantine attacks [18, 21], in which an attacker tampers with a subset of the sensors and forces them to send falsified messages to the fusion center (FC) with the goal of interfering with the detection or estimation task at the FC. Byzantine attacks on sensor networks attempting to perform hypothesis testing have been studied in [10,17–20], while Byzantine attacks on networks performing parameter estimation have been studied in [11–13, 22].

The presented work investigates Byzantine attacks on sensor systems estimating the value of an unknown deterministic parameter based on quantized observations. We first focus on the optimal processing under a practical family of attacks where the sensors employ bad data detectors that check if the observed sensor data fits the unattacked observation model assumed by the estimation algorithm under no attack. Afterwards, the estimation performance is considered for any general estimation approach employed by the sensor system under any general attack for cases where any number of observations, sensors, and quantization levels could be employed and expressions are derived to describe the performance of any specific estimation algorithm under such attacks.

1.2 Ocean Wave Estimation

Ocean wave energy holds the potential to become a significant source of renewable energy. Unfortunately, large costs associated with the production, deployment, and maintenance of wave energy converters (WECs) present a challenge to their economic viability. As a result, extending the lifetime of a WEC and maximizing the amount of energy it captures are of great interest. To this end, several active control strategies have been developed where the controller adapts the behavior or characteristics of a WEC in response to the prevailing ocean conditions in order to more effectively capture the energy in the waves.

Complex-conjugate control [23] was the earliest control strategy considered for WEC devices [24]. This control strategy focuses on heaving WECs and is formulated using a frequency based description of WEC dynamics. Executing complex-conjugate control in the time domain was shown in [24, 25] to require future information on the impending wave excitation forces. Alternative control approaches that account for physical constraints on the motion of the WEC have also been proposed. Latching control [26–28] is one such control strategy where the device is locked at various points in the wave cycle and released later such that the device is always oscillating "in-phase" with the incident excitation force. Since Latching control adapts to the incident excitation force which is related to the motion of the WEC through a typically non-causal impulse response function, this control strategy also requires knowledge of the future excitation forces acting on the WEC. Another approach is Model Predictive Control (MPC) which has been the focus of several recent work [29–35] on WEC control. A more comprehensive review of WEC control approaches is provided in [36]. The focus of our work is on providing wave estimates that might be required by a current or novel approach for WEC control.

The presented work investigates several aspects of the estimation and short-term prediction of ocean waves using a network of spatially distributed sensors. We present general expressions for the Fisher Information Matrix (FIM) and the Cramer Rao Bound (CRB). Our work is among the first to account for noise in the sensor observations and employ tools from estimation theory to derive expressions providing lower bounds on the estimation performance. Also presented is a low-complexity estimation approach for the case where several stationary ocean sensors are employed for estimation. We provide several numerical results to highlight the estimation accuracy of our approach relative the CRB. Further, we present a method that relates errors in estimation to the mean loss in the amount of absorbed power by WEC devices.

1.3 Stochastic Optimal Power Flow Under Forecast Errors and Failures in Communication

The reliable and economic operation of an electric grid is a complex problem involving several different tasks. Managing and executing such tasks requires collecting a great deal of information delivered to grid operators through communication networks. Upon receiving new data, an operator may seek to modify specific properties or values in the electric grid through transmitting control commands back to the grid through the communication network. Clearly, this creates a dependency where failures, malfunctions, or other unexpected changes affecting components in the communication network could destabilize or even interrupt the operation of the electric grid or some parts of it. In fact, the work in [37] showed that failures or attacks affecting the communication of sensor measurements to system operators may result in consistent and sustained economic losses due to the misinformed operation of the grid.

Most related work has focused on the study of cascading failures under an interdependent model where the communication network relies on the power grid for its operation. The study of interdependent networks using a mathematical framework was first undertaken by the authors of [38] who focused on the interdependency between two abstract randomly-generated networks having the same number of nodes. Focusing on power networks, the work in [39] applied interdependency models from [38] to a non-random graph representing the Italian power grid,which suffered a large blackout in the year 2003 [40], and its supporting communication network. The recent work in [41] and [42] incorporated Kirchoff's laws and simple control operations into their models. The work in [41] presented a load shedding algorithm that aims to mitigate the effects of failures in a random

communication network supporting the operation of a power grid while the work in [42] studied the effect of communication failures on state estimation and the delivery of control signals. In both cases, the work in [41] and [42] assumed a given communication network with pre-defined interdependencies and did not focus on identifying robust communication networks tailored to the power grids they support although the work in [41] stressed the importance of this step. In fact, the authors of [41] and [42] found that their results were sensitive to the assumed communication topology and that some of the phenomena observed by studying random abstract interdependent networks [38, 43–47] may not hold for power networks supported by well designed communication networks. Further, they cautioned that communication networks designed in isolation of the power grids they will support may be unable to withstand or mitigate the effects of interdependence on the system as a whole.

In the presented work, we focus on identifying the optimal communication network topology and the optimal choice for dispatch and re-dispatch actions while accounting for forecasting errors and failures in communication. Our problem is formulated as a stochastic optimal power flow (OPF) problem [48–51] where the objective is to minimize the average overall economic cost of operating the system under a set of constraints representing the stochastic scenarios that may result from possible forecasting errors and communication failures. We believe our work is the first to jointly consider the effects of forecast errors and failures in communication and study them under a stochastic OPF formulation where the communication topology is also optimized. Our unique formulation describes the communication network as a flow network where any node representing a sensor measuring physical values from the power grid is assigned a positive 'supply' and the node representing the system operator is assigned a negative supply. This description allows our optimization problem to capture the effect of losing the ability to communicate between any sensor and the system operator as a loss in the flow supply from the affected sensor. We present results obtained using our model under test cases that we use to validate our model and identify the best communication topologies for the power system we consider.

1.4 Topology Estimation in Power Distribution Networks

As novel controls, applications, and services continue to be integrated into distribution systems, the demand for accurate and timely estimates of the network topology is becoming increasingly more critical [52]. Maintaining situational awareness in distribution networks is imperative for

the effectiveness of many important modern tools and applications that are currently being implemented and developed. For example, having a correct estimate of the network topology is crucial for efficiently and reliably dispatching distributed energy resources [53], sectioning into microgrids [54], and providing demand response [55] capabilities. Further, many techniques for distribution system state estimation (DSSE), [56–58] for example, are based on the assumption that the correct network topology is already known. In fact, the performance of most DSSE methods under a misspecified topology is generally not well understood and may be difficult to characterize.

Due to several engineering and practical concerns, distribution networks are predominately operated as radial (tree) graphs where power flows in one direction away from the root node typically taken as the main feeder. Their special characteristics have traditionally led to fewer installed sensors and monitoring devices making distribution networks historically less observable than transmission systems which tend to have complex mesh topologies and extensive monitoring. As a result of the differences between the two systems, topology estimation techniques devised for transmission systems [59–62] have limited applicability in distribution networks and there is a need for methods and techniques designed especially for distribution networks.

Many topology estimation efforts in distribution networks are based on information obtained through phone calls from customers or expert systems [63] to identify and locate outages in the network. Other knowledge-based methods combining different types of information [64] have also been proposed to include data from advanced metering infrastructure (AMI) and supervisory control and data acquisition (SCADA) systems. Voltage sag measurements and matching has also been studied for fault detection in [65] and [66]. However, the majority of these methods were generally limited in the number of simultaneous outages they are able to detect.

Fortunately, recent advances in the development of measurement units and the increased adoption of monitoring devices will produce new and reliable data streams that may lead to improved topology estimation techniques capable of detecting several topology changes. In fact, some new phasor measurement units [67] are being designed specifically for distribution networks with promising results and new topology estimation methods [68, 69] assuming their widespread adoption are already being proposed. Relying on a limited number of sensors, the insightful work in [70, 71] uncovered some foundational properties for the topology estimation problem in distribution networks. In their work, the authors of [70, 71] show that the topology estimation problem, referred to as outage detection in [70, 71], for radial distribution networks can be decoupled into smaller

detection problems within subtrees of the same network.

In our work, we study the topology estimation and fault detection problem based on infrequent and slow-changing nodal load forecasts and power flow measurements obtained from sensing devices installed at a subset of the nodes in the network. Aided by the decoupling of the full problem, detailed in [71], we propose a novel sensor placement scheme allowing for the identification of any number of detectable faults in the system. Further, we present a novel topology estimation scheme designed to limit our dependence on unavoidable enumeration operations that could increase the numerical complexity of the scheme. We also present numerical results demonstrating the performance of our proposed estimation scheme.

1.5 Outline of The Dissertation

The reset of this dissertation is organized as follows. Chapter 2 presents our work considering attacks on systems estimating an unknown deterministic parameter based on quantized observations. Next in Chapter 3, our research regarding the estimation and short-term forecasting of ocean waves for the use by WEC devices is presented. Chapter 4 describes our work on the joint optimization of communication networks and power flow control under forecast errors and failures in communication. Finally, Chapter 5 describes our work on topology estimation and fault detection in radial distribution networks.

Chapter 2

Attacks on Systems Estimating an Unknown Deterministic Parameter Based on Quantized Observations

The existing work on Byzantine attacks on estimation systems [11–13, 22] has covered a wide variety of topics, including a game theoretic analysis and mitigation schemes, but has not investigated the topic of optimum processing of the attacked and unattacked sensor observations. The work in [72] considered the optimum processing, under an asymptotically large number of observations, for attacks that modify the quantized sensor data such that P distinct attacks are launched on P distinct sets of sensors. However, the results in [72] are restricted to cases with binary quantization and where a shift-in-mean parameter is to be estimated.

In the first section of this chapter, we consider a practical attack scenario where the sensors employ bad data detectors that check if the observed sensor data fits the unattacked observation model assumed by the estimation algorithm. Only sensor data which passes the checks made by the bad data detectors will be employed in the subsequent estimation procedure. To be more specific, we are assuming that a bad data detector will collect observations from a single sensor and compute some similarity index between the empirically observed probability density function (pdf) and the possible pdfs assumed under no attack, which are parameterized by the unknown the parameter of

interest. Then, as the number of observations per sensor becomes very large, the only way for an attacked sensor to pass the bad data detector is if the pdf of the observations at the sensor is that of observations at an unattacked sensor, but with a different value for the unknown parameter. In such cases, spoofing attacks, which modify the physical phenomenon observed by the sensors, become of considerable interest since, unlike in man-in-the-middle attacks, the attacker can launch spoofing attacks which are guaranteed to pass the bad data detectors even if the attacker has no knowledge regarding the quantizers employed by the sensor system or the true value of the parameter to be estimated. In fact, low complexity spoofing attacks capable of generating data that is guaranteed to pass bad data detectors in this way have been employed in practice, but have not been studied in previous work on attacks on estimation systems. Consider the case where we want to estimate the position of an emitter or reflector by estimating the time taken for the signal to reach an antenna. One practical example of a spoofing attack is when the attacker uses a memory device to capture and regenerate the signal in order to add an extra time delay to it when it finally reaches the antenna. As one example, this basic idea has been exploited [73] in Civilian GPS systems which were been shown to be highly vulnerable to such attacks. Radar systems have also been attacked in this way. In this chapter we assume the attacker launches one of these spoofing attacks that are guaranteed to pass the bad data detectors while considering general estimation problems.

Given that the attacked sets of sensors can be identified at the fusion center using a procedure different from the bad data detectors, we show in this chapter that the attacked sensor data cannot be employed to improve the estimation performance beyond that achieved by optimum processing of the unattacked sensor data for cases with a sufficient number of observations. These results are valid for arbitrary quantizer designs and for a large set of estimation problems.

While optimum processing is of considerable interest, the impact of attacks on both optimum and suboptimum processing is also of great interest. For example, information about estimation performance under different attacks could be used to determine the different advantages and disadvantages provided by different estimation approaches. In the second section of this chapter, we provide a general approach to characterize the after-attack estimation performance of any estimation approach, optimum or suboptimum, under any general type of attack, without any assumptions on the estimation problem, the observation models, the number of sensors, the number of observations or the independence of any observations to any others. A classification of these attacks, which are much more general than those assumed in the first section of this chapter,

that categorizes them into classes according to the information available to the attacking entity is introduced and some notable properties of these attack classes are studied and highlighted using examples. Optimization problems over these attack classes are solved to provide, for the first time, expressions which describe the performance of any specific estimation algorithm under the most devastating attacks with full information and the generally less effective information free attacks. Lastly, constraints on the different classes of attacks are considered to find the most devastating attacks under some level of attack detection carried out by the sensor system. These ideas are illustrated by considering the degradation in estimation performance under the appropriately constrained most devastating attacks with full information and the generally less effective information free attacks. Such investigations could be of significant practical importance as some legacy or even modern systems might not have been designed to detect and react to attacks.

The following assumptions and notations will be used throughout this chapter. We consider the estimation of a deterministic parameter θ where $\theta \in \Re$. The parameter θ is estimated based on quantized observations collected from N sensors, each producing K observations. The NK observations are collected in the vector $\mathbf{u} = (u_{11}, u_{12}, \ldots, u_{NK})^T$ where $u_{jk}, j = 1, 2, \ldots, N, k = 1, 2, \ldots, K$ is the quantized version of x_{jk}, the unquantized output of sensor j at discrete time k which has a pdf $f(x_{jk}|\theta)$ that is parameterized by θ. At each sensor, the number of quantization levels used is assumed to be R such that each u_{jk} can take any value from 1, 2, ..., R. Perfect, error-free communication of $\mathbf{u}$ to the fusion center is assumed. In the case of no attack, the probability of the vector $\mathbf{u}$ having a specific value $\mathbf{r}$ for a given value of θ is described by $\Pr(\mathbf{u} = \mathbf{r}|\theta)$. When attacked, the effect of the attack is captured completely by the after-attack probabilities $\Pr(\tilde{\mathbf{u}} = \mathbf{r}|\theta)$ for every possible realization of $\mathbf{r}$, where $\tilde{\mathbf{u}}$ denotes the after-attack vector of quantized observations.

The remainder of this chapter is organized as follows. Section 2.1 considers the optimum processing of the attacked and unattacked sensor data when bad data detection is employed at each sensor, provided the sets of attacked sensors are identified. Section 2.2 considers the impact of general types of attacks on general types of estimation approaches and describes the most devastating attacks which employ full information and the generally less effective information free attacks. An example where an estimation system, which uses the estimation approach it would use assuming no attack, experiences full information and information free attacks is given. Further, constraints on the considered attacks, employed to account for some level of attack detection under the described attack classes, are studied and an example showing the impact of such attacks is

presented. Finally, Section 2.3 summarizes and concludes the work presented in the chapter.

2.1 Asymptotically Optimum Processing under Sensor Bad Data Detection

In this section, we consider attacks that are launched without knowledge of the quantizers employed by the sensor system or the true value of the parameter to be estimated, but that still produce data guaranteed to pass the bad data detectors. The focus is on cases where the unattacked sensor observations are statistically independent and identically distributed (iid) across all sensors and time epochs. The following assumptions are made throughout this section.

Assumption 1. *The pdf of the unquantized observations at time k and unattacked sensor j, $f(x_{jk}|\theta)$, is a twice differentiable log-concave [74] function of θ which obeys regularity (smoothness) conditions such that interchanges involving derivatives (up to order 2) with respect to θ and integrals with respect to x_{jk} of $f(x_{jk}|\theta)$ are valid.*

Assumption 2. *The unattacked observations at different sensors and times $x_{jk}, j = 1, \ldots, N, k = 1, \ldots, K$ are iid. All x_{jk} are quantized using the same thresholds which are fixed over all $j = 1, \ldots, N$ and $k = 1, \ldots, K$.*

The common quantizer design is described by the regions $A_1, \ldots, A_R$ such that if the quantizer input falls into the region A_ℓ then the output is the symbol ℓ. Thus for $k = 1, \ldots, K, j = 1, \ldots, N$, define the probability mass function (pmf) of the quantized version of x_{jk} as

$$\Pr(u_{jk} = r_{jk}|\theta) = \int_{x_{jk} \in A_{r_{jk}}} f(x_{jk}|\theta) dx_{jk}, \quad r_{jk} = 1, \ldots, R \tag{2.1}$$

and denote the indicator function $I(r_{jk} = r'_{jk})$ as taking on the value unity when $r_{jk} = r'_{jk}$ and zero otherwise. Let $\mathcal{J}$ define a set of sensor indices that correspond to a set of selected sensors. Thus the members of $\mathcal{J}$ are selected from $1, \ldots, N$. Under no attack and by *Assumption 2*, the pmf of the vector of selected quantized observations $\mathbf{u}_\mathcal{J}$ evaluated at $\mathbf{u}_\mathcal{J} = \mathbf{r}$ is

$$\Pr(\mathbf{u}_\mathcal{J} = \mathbf{r}|\theta) = \prod_{j \in \mathcal{J}} \prod_{k=1}^{K} \prod_{r'_{jk}=1}^{R} \Pr(u_{jk} = r'_{jk}|\theta)^{I(r_{jk} = r'_{jk})}. \tag{2.2}$$

The log-likelihood function evaluated at $\mathbf{u}_{\mathcal{J}} = \mathbf{r}$, the natural logarithm of (2.2), is

$$L(\theta, \mathcal{J}) = \sum_{j \in \mathcal{J}} \sum_{k=1}^{K} \sum_{r'_{jk}=1}^{R} I(r_{jk} = r'_{jk}) \ln \Pr(u_{jk} = r'_{jk}|\theta). \tag{2.3}$$

To guarantee the attack will pass the bad data detector without knowledge of the quantizers employed and the true parameter value, the attackers employ attacks that follow the same model as in (2.1), (2.2), and (2.3) but with θ replaced by $\theta - e$ for some acceptable e, leading to the following assumption which is also employed throughout this section.

Assumption 3. *The pdf of the unquantized observations at time k and attacked sensor j is $f(x_{jk}|\theta - e)$, where $f(x_{jk}|\theta)$ follows Assumption 1*

Consider a general case where there are P such attacks that each employ a different and non-zero value for e, denoted by $e_p,\ p = 1, \ldots, P$ and the unattacked sensors are considered to be affected by the 0^{th} attack for which $e_0 = 0$. Assume that the indices of all sensors under the p^{th} attack are described by the set $\mathcal{A}_p$ and, for a sufficiently large N, the number of these sensors $|\mathcal{A}_p|$ is a fixed non-zero percentage $\mathcal{P}_p$ of the total number of sensors, N. Such an assumption is required so that as $N \to \infty$ the effect of an attack will not shrink to zero ($\mathcal{A}_p$ becoming a set of measure zero). Further, the sets $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_P$ are disjoint and cover $\mathcal{J}$ so that

$$\mathcal{J} = \bigcup_{p=0,\ldots,P} \mathcal{A}_p \quad \text{and} \quad \mathcal{A}_p \bigcap \mathcal{A}_{p'} = \emptyset \quad \text{if} \quad p \neq p'. \tag{2.4}$$

For the general case just described and from (2.2), the pmf of the quantized observations $\mathbf{u}_{\mathcal{J}}$ at the selected sensors $j \in \mathcal{J}$ evaluated at $\mathbf{u}_{\mathcal{J}} = \mathbf{r}$ becomes

$$\Pr(\mathbf{u}_{\mathcal{J}} = \mathbf{r}|\theta) = \prod_{p=0}^{P} \prod_{j \in \mathcal{A}_p} \prod_{k=1}^{K} \prod_{r'_{jk}=1}^{R} \Pr(u_{jk} = r'_{jk}|\theta - e_p)^{I(r_{jk}=r'_{jk})}, \mathcal{P}_p = \frac{|\mathcal{A}_p|}{N} \text{ fixed.} \tag{2.5}$$

Note that (2.5) includes the case where some sensors are unattacked since $e_0 = 0$.

In order to gain a better understanding of what is presented next, consider the Fisher Information obtained from a selected set of sensors, some unattacked and others attacked with known

$e_1, \ldots, e_P$, which can be shown to be

$$J(\theta) = |\mathcal{A}_0| K \sum_{r'_{jk}=1}^{R} \frac{\left(\int_{x_{jk} \in A_{r'_{jk}}} \frac{d}{d\theta} f(x_{jk}|\theta) dx_{jk} \right)^2}{\int_{x_{jk} \in A_{r'_{jk}}} f(x_{jk}|\theta) dx_{jk}} + \sum_{p=1}^{P} |\mathcal{A}_p| K \sum_{r'_{jk}=1}^{R} \frac{\left(\int_{x_{jk} \in A_{r'_{jk}}} \frac{d}{d\theta} f(x_{jk}|\theta - e_p) dx_{jk} \right)^2}{\int_{x_{jk} \in A_{r'_{jk}}} f(x_{jk}|\theta - e_p) dx_{jk}}. \quad (2.6)$$

The first term on the right of the equality sign in (2.6) represents information from the unattacked sensors and the second term represents information from attacked sensors. Now consider the following lemma, based on (2.6), which is concerned with the case of attacks that are perfectly known.

Lemma 1. *With $e_1, \ldots, e_P$ known in (2.5), an efficient (Cramer Rao bound (CRB) achieving) unbiased estimator of θ will generally utilize the attacked observations in the process of estimating θ.*

Contrast the just described Lemma with the following theorem based on the special structure of the Fisher Information Matrix (FIM) for problems of the type given in (2.5).

Theorem 1. *With the attacked sensors in (2.5) known (know all $\mathcal{A}_p$ for which $|e_p| > 0$) but the corresponding magnitudes of the attacks (e_p when $|e_p| > 0$) unknown, the mean-square error (MSE) of an efficient unbiased estimator of θ is as large or larger than the MSE of an efficient unbiased estimator that knows the magnitudes of the attacks for any finite K (from (2.6)). In fact, the MSE of the efficient unbiased estimator of θ when the attack magnitudes are unknown but the attacked sensors are known is that of an efficient unbiased estimator that only uses the unattacked observations for any value of K.*

Proof. Let J_0 denote the first sum in (2.6) and let J_p, $p = 1, 2, \ldots P$ denote the p^{th} term from the second sum in (2.6), then the FIM $\mathbf{J}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta, e_1, e_2, \ldots, e_P)'$, when the attacked sensors

are known but the corresponding magnitudes of the attacks are unknown is

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} J_0 + J_1 + \ldots J_P & -J_1 & -J_2 & \ldots & -J_P \\ -J_1 & J_1 & 0 & \ldots & 0 \\ -J_2 & 0 & J_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -J_P & 0 & 0 & \ldots & J_P \end{bmatrix}. \tag{2.7}$$

The MSE of an efficient unbiased estimator of θ is the first element on the diagonal of the inverse of (2.7) and due to the special structure of (2.7), it can be shown to be, as one might expect, given by

$$[\mathbf{J}^{-1}(\boldsymbol{\theta})]_{1,1} = \frac{1}{J_0} \geq \frac{1}{J_0 + \sum_{p=1}^{P} J_p}, \tag{2.8}$$

where $\frac{1}{J_0+\sum_{p=1}^{P} J_p}$ is the reciprocal of (2.6), which is the MSE of an efficient estimator of θ that knows the magnitudes of the attacks. The last statement in the theorem can be proved by a direct comparison between $[\mathbf{J}^{-1}(\boldsymbol{\theta})]_{1,1}$ and the MSE of an efficient unbiased estimator that only uses the unattacked observation. □

The asymptotic identification of the attacked sensors is discussed in[1] [75].

The next section considers the impact of general types of attacks on general types of estimation approaches and describes the most devastating attacks which employ full information and the generally less effective information free attacks.

2.2 Bounds on the Performance of General Estimation Approaches under General Attacks

In this section, we consider the estimation of a deterministic parameter θ such that $\theta \in [\Omega_l, \Omega_h]$ for some known finite Ω_l and Ω_h. The most general attacks will change the after-attack probabilities $\Pr(\tilde{\mathbf{u}} = \mathbf{r}|\theta)$ in an arbitrary way as opposed to the specific changes considered in Section 2.1. Since there exists a wide range for such possible attacks, there is a need to classify the various types of possible attacks and to describe their impact on systems employing general estimation approaches.

[1]The asymptotic identification of the attacked sensors was based on work by Jiangfan Zhang and is thus excluded from this dissertation

Working towards this goal, a classification for attacks is presented in this section that separates the attacks according to the knowledge or information available to the entity carrying out the attack about the estimation system it is attacking and the true value of the parameter being estimated. Having full information, which describes the most devastating type of attacks, is defined as knowing the design of the quantizer used at each sensor, the estimation algorithm employed at the fusion center, and the true value of the parameter to be estimated. The least informed attacks, described as information free attacks, are those attacks that lack any information in regards to the design of the quantizers, the estimation algorithm, or the true value of the parameter being estimated. Adopting the mean square error (MSE) as a performance metric throughout the remaining part of this chapter, the effect on the estimation performance of the different attacks just described is presented in the following theorems. The same analysis can be carried out using other metrics as well. The following theorem describes full information attacks.

Theorem 2. *Without any loss of generality, consider an attack on only the first p sensors (all times) with quantized samples $\tilde{\mathbf{u}}_p = \mathbf{r}_p$. Denote the quantized observations at the other sensors as $\mathbf{u}_o = \mathbf{r}_o$ so that the full vector of quantized observations presented to the fusion center is $\tilde{\mathbf{u}} = (\tilde{\mathbf{u}}_p, \mathbf{u}_o) = (\mathbf{r}_p, \mathbf{r}_o)$. The general estimation rule based on the actual realization of the quantized observations is denoted by $\hat{\theta}(\tilde{\mathbf{u}})$. Assume the attacker is unable to observe the realizations of $\mathbf{u}_o$ and is also unable to change them and that the attacker has full information concerning the actual value of θ and the estimation system. Then the largest MSE of any such attack when $p = N$ is*

$$MSE(\hat{\theta})_{OCU} \quad = \quad \max_{\mathbf{r}} (\hat{\theta}(\mathbf{r}) - \theta)^2. \tag{2.9}$$

An attack resulting in this MSE is called an optimum complete knowledge uncoordinated attack (OCU). When $0 < p < N$ the MSE under an OCU attack is

$$MSE(\hat{\theta})_{OCU} = \sum_{all\ \mathbf{r_o}} \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p^*, \mathbf{r}_o)) - \theta)^2 \tag{2.10}$$

where $\mathbf{r}_p^*$ is an $\mathbf{r}_p$ maximizing $\sum_{all\ \mathbf{r_o}} \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2$. If $p = 0$ then the unattacked performance is achieved.

If the attacker is able to observe the realizations of both $\mathbf{u}_o$ and $\mathbf{u}_p$ but is only able to change $\mathbf{u}_p$, while still having full information, then the attack is called the optimum complete knowledge coordinated attack (OCC). The largest possible MSE under an OCC attack, which is larger than

or equal to (2.10) becomes

$$MSE(\hat{\theta})_{OCC} = \sum_{all\ \mathbf{r_o}} \max_{\mathbf{r}_p} \left(\Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2 \right) \tag{2.11}$$

when $0 < p < N$. For $p = N$, an OCC attack provides the same largest possible MSE as an OCU attack.

Proof. In the general case, the MSE can be written as

$$MSE(\hat{\theta}) = \sum_{\text{all } \mathbf{r}} \Pr(\tilde{\mathbf{u}} = \mathbf{r}|\theta)(\hat{\theta}(\mathbf{r}) - \theta)^2 \tag{2.12}$$

$$= \sum_{\text{all } \mathbf{r}_p} \sum_{\text{all } \mathbf{r_o}} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p, \mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2. \tag{2.13}$$

If the attacker is unable to observe the realizations of $\mathbf{u}_o$ and is also unable to change them, then the attacker is forced to attack with a $\tilde{\mathbf{u}}_p$ that must be independent from $\mathbf{u}_o$. In this case (2.13) becomes

$$MSE(\hat{\theta})_{OCU} = \sum_{\text{all } \mathbf{r}_p} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) \sum_{\text{all } \mathbf{r_o}} \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2. \tag{2.14}$$

As the attacker has control over only $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta)$, (2.14) is maximized by putting all the probability mass on $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p^*|\theta)$ where $\mathbf{r}_p^*$ is an $\mathbf{r}_p$ maximizing $\sum_{\text{all } \mathbf{r_o}} \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2$ and, hence, (2.10) follows. If $p = N$ then $\mathbf{r} = \mathbf{r}_p$ so the same maximization can be applied to (2.12), now putting all probability mass on the $\mathbf{r}$ that maximizes $(\hat{\theta}(\mathbf{r}) - \theta)^2$.

For an OCC attack, (2.13) is expressed as

$$MSE(\hat{\theta})_{OCC} = \sum_{\text{all } \mathbf{r}_p} \sum_{\text{all } \mathbf{r_o}} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\mathbf{u}_o = \mathbf{r}_o, \theta) \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2. \tag{2.15}$$

Now for each realization of $\mathbf{u}_o$, the attacker can choose a different realization of $\tilde{\mathbf{u}}_p$ by setting $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\mathbf{u}_o = \mathbf{r}_o, \theta)$ in (2.15). Thus, the attacker can maximize (2.15), by picking $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\mathbf{u}_o = \mathbf{r}_o, \theta) = 1$ for the $\mathbf{r}_p$ that makes $\Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2$ largest for the given $\mathbf{r}_o$. In this case the MSE becomes (2.11). □

An OCC attack maximizes the MSE for each possible realization of $\mathbf{u}_o$ which results in the largest possible MSE values and, therefore, the MSE under an OCC attack outlines the upper

boundary of the achievable MSE by an attack. Using MSE curves produced under OCC attacks should allow system designers to quantify the worst case under-attack performance of their systems. This could be employed as one of the design criteria for estimation systems where the worst case performance is required to be smaller than a certain tolerable MSE value. Further, considering the MSE values under a given number of OCC-attacked sensors can provide system designers with information regarding the number of sensors that need to be protected in order to guarantee MSE values smaller than a certain tolerable MSE value. Another use is to obtain these curves for different estimation approaches in order to compare their performance under attacks. The following defines an information free attack and presents a theorem considering the estimation performance under such attacks.

Definition 1. *Under an information free (IF) attack,*

$$\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) = \left(\frac{1}{R}\right)^{pK}, \ \forall \mathbf{r}_p, \tag{2.16}$$

where p is the number of attacked sensors and $\tilde{\mathbf{u}}_p$ is the vector of quantized observations from the p attacked sensors.

Remark 1. *An (IF) attack is attempting to model a case where the attacker has no information regarding the true value of θ, the estimation algorithm used at the fusion center, the threshold values employed for quantizing the sensor observations, or the interpretation of the quantized sensor observations. Further, the attack is unable to accumulate any information over the duration of the attack. Here it is assumed that the sensors and fusion center can essentially employ coding to hide the true values of quantized data. As a result, each attacked sensor must be attacked in a way that appears completely random and uncoordinated when viewed over many realizations, resulting in all the possible values for any attacked quantized sensor sample being equally likely. Since we are interested in statistically characterizing the average performance of the system, such a model seems appropriate. Note that the model makes it impossible for the attackers to repeatedly produce the same attack owing to the attackers inability to interpret the attacked samples throughout their attack.*

Remark 2. *An attack which turns out to be mathematically equivalent to IF attacks was introduced in [22] where it was called a 'blinding' attack. The approach in [22] is different from the one adopted here and the authors of [22] provide an alternative argument that motivates the attacker to resort to this attack.*

Theorem 3. *Under Definition 1, and without loss of generality, consider an IF attack on only the first p sensors (all times) with quantized observations* $\tilde{\mathbf{u}}_p = \mathbf{r}_p$. *Denote the quantized observations at the other sensors as* $\mathbf{u}_o = \mathbf{r}_o$ *so that the full vector of quantized observations presented to the fusion center is* $\tilde{\mathbf{u}} = (\tilde{\mathbf{u}}_p, \mathbf{u}_o) = (\mathbf{r}_p, \mathbf{r}_o)$. *Then, the MSE when* $p = N$ *is*

$$MSE(\hat{\theta})_{IF} = \left(\frac{1}{R}\right)^{NK} \sum_{all\ \mathbf{r}} (\hat{\theta}(\mathbf{r}) - \theta)^2. \tag{2.17}$$

When $0 < p < N$ *the MSE of an IF attack is*

$$MSE(\hat{\theta})_{IF} = \left(\frac{1}{R}\right)^{pK} \sum_{all\ \mathbf{r}_p} \sum_{all\ \mathbf{r_o}} \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2. \tag{2.18}$$

If $p = 0$, *there is no attack.*

Proof. Since the attacker does not have any information about the estimation system or the true value of θ, the changes this attacker makes are completely random and unrelated to how the estimation system is operating and to the true value of θ. Thus, the IF attack can be modeled as the attacker observing a quantized sensor value r'_{jk} and then modifying r'_{jk} to each of its possible values[2] $1, \ldots, R$ with equal probability of $\frac{1}{R}$. This mapping implies the conditional pmf of the modified value $\tilde{u}_{jk} = r_{jk}$ given the unmodified value $u_{jk} = r'_{jk}$ is $\Pr(\tilde{u}_{jk} = r_{jk}|u_{jk} = r'_{jk}) = \frac{1}{R}$ so that the assumption of uncoordinated (independent) attacks on all sensors implies

$$\begin{aligned}
\Pr(\tilde{\mathbf{u}} = \mathbf{r}|\theta) &= \sum_{\text{all } \mathbf{r}'} \Pr(\tilde{\mathbf{u}} = \mathbf{r}|\mathbf{u} = \mathbf{r}')\Pr(\mathbf{u} = \mathbf{r}'|\theta) \\
&= \sum_{\text{all } \mathbf{r}'} \left(\prod_{j=1}^{N}\prod_{k=1}^{K} \frac{1}{R}\right) \Pr(\mathbf{u} = \mathbf{r}'|\theta) \\
&= \left(\frac{1}{R}\right)^{NK} \left(\sum_{\text{all } \mathbf{r}'} \Pr(\mathbf{u} = \mathbf{r}'|\theta)\right) \\
&= \left(\frac{1}{R}\right)^{NK}
\end{aligned} \tag{2.19}$$

[2]Another IF attack could be described where the attacker modifies the value of r'_{jk} to each of its possible values except for the original one with equal probability of $\frac{1}{R-1}$. The analysis would be nearly identical and is, therefore, omitted.

which is independent of θ. For an attack on only the first p sensors

$$\begin{aligned}
&\Pr(\tilde{\mathbf{u}} = \mathbf{r}|\theta) = \\
&\sum_{\text{all } \mathbf{r}'_p} \Pr(\tilde{\mathbf{u}} = \mathbf{r}_p|\mathbf{u} = \mathbf{r}'_p)\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}'_p, \mathbf{u}_o = \mathbf{r}_o|\theta) \\
&= \left(\frac{1}{R}\right)^{pK} \sum_{\text{all } \mathbf{r}'_p} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}'_p, \mathbf{u}_o = \mathbf{r}_o|\theta) \\
&= \left(\frac{1}{R}\right)^{pK} \Pr(\mathbf{u}_o = \mathbf{r}_o|\theta). \qquad (2.20)
\end{aligned}$$

Using (2.19) and (2.20) in (2.12), the theorem follows

□

Although IF attacks lack any information regarding the parameter to be estimated and the estimation algorithm employed by the system, IF attacks are still capable of causing significant increases in the after-attack MSE values in some cases. Next, an example problem showing the effect of the different attacks just introduced is presented.

2.2.1 An example: Maximum Likelihood estimation of a Constant in Truncated Gaussian Noise

Consider the case where each sensor observes a constant $\theta \in [-3, 3]$ observed in truncated zero-mean unit-variance Gaussian noise. The noise is truncated to represent the case where the sensors produce only finite valued observations in the range $[-5, 5]$ prior to quantization, which may be considered more realistic. For the analysis, we employ a system comprised of eight sensors that each transmit a single quantized observation to a fusion center with perfect transmission. Under no attack, the observations at different sensors $x_{j1}, j = 1, \ldots, N$ are assumed to be statistically independent and identically distributed (iid). Further, the sensors are assumed to use identically designed three-level quantizers having lower and upper thresholds of -1 and 1, respectively. For all the presented figures, the estimation algorithm used at the fusion center is taken to be maximum likelihood (ML) estimation under no attack, where the estimate $\hat{\theta}(\mathbf{r})_{ML}$ is defined to be the solution to $\max_\theta L(\theta, \mathcal{J})$, where $L(\theta, \mathcal{J})$ is the log-likelihood function.

2.2.2 Estimating a Constant in Truncated Gaussian Noise

Using the estimation system just described, and with four out of the eight sensors being attacked, Figure 2.1 shows the resulting MSE values for the estimation under the different attacks described in this section. Figure 2.1 also shows that, for the considered case, an attack having full information always provides an MSE which is at least twice as large when compared to the IF attack, which is the one lacking any information. System designers can use figures similar to

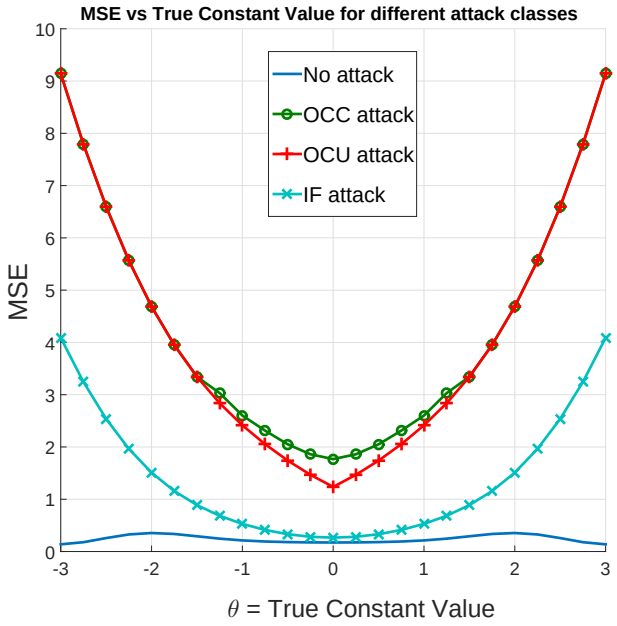


Figure 2.1: MSE for Constant in Truncated Gaussian estimation with $N = 8, K = 1$.

Figure 2.1 in order to understand the possible impacts of attacks on different estimation algorithms and quantizer designs for a specific estimation problem.

Figure 2.2 shows the MSE curves under an increasing number of OCC attacked sensors for the same estimation system. OCC attacks are the most devastating attacks since the attacker has complete information regarding the parameter being estimated and the estimation system as a whole, including the employed estimation algorithm. Therefore, system designers can use figures similar to Figure 2.2 in order to deduce the number of sensors that need to be protected in order to guarantee MSE levels below a certain tolerable value.

2.2.3 Special properties for IF attacks

In figure 2.3, the same estimation system described earlier is used with the only exception of changing the thresholds of the quantizers from -1 and 1 to $-\mathbf{Q}^{-1}(\frac{1}{3})$ and $\mathbf{Q}^{-1}(\frac{1}{3})$ respectively

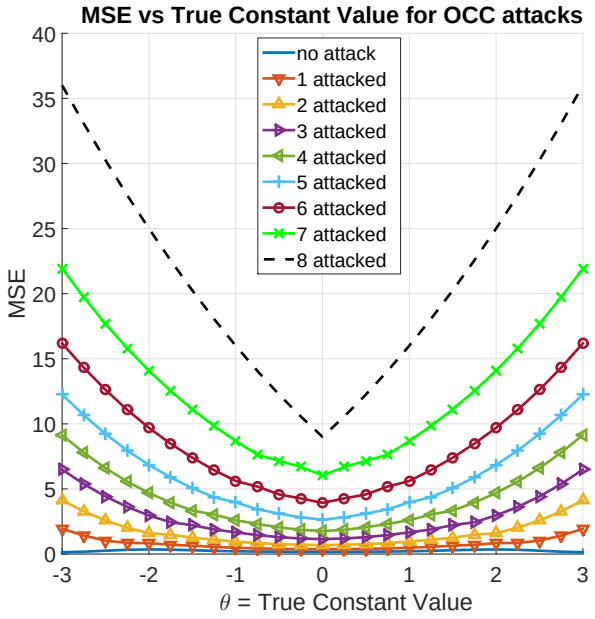


Figure 2.2: MSE for Constant in Truncated Gaussian estimation with $N = 8, K = 1$ and an increasing number of OCC attacked sensors.

where $\mathbf{Q}^{-1}$ is inverse of the $\mathbf{Q}$ function, which is defined as $\mathbf{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{\frac{-u^2}{2}} du$. This choice for the thresholds sets $\Pr(\tilde{\mathbf{u}} = \mathbf{r}|\theta) = 1/R$ for all $\mathbf{r}$ when $\theta = 0$. Such a quantizer setting disarms the IF attack when $\theta = 0$ by starting with (2.19) even before the IF attack. As a result, this creates a point (at $\theta = 0$) of insensitivity towards IF attacks regardless of the number of sensors attacked. This is shown in Figure 2.3. In general, system designers can choose the set of thresholds used in the quantizers such that a point of insensitivity towards IF attacks occurs at any given value of θ. Another interesting result shown in Figure 2.3 that is not seen in Figure 2.1 is that there exists a region for the parameter values where the after-attack MSE is smaller than the unattacked MSE. This indicates that defending against some attacks might not be necessary for some values of the parameter being estimated since the IF attack is a random attack and it can actually result in improved MSE values.

2.2.4 Bounds on the Performance of Constrained General Estimation Approaches under General Attacks

In this subsection, we consider a case where the system employs some attack detection method that constrains the passable attacks faced by the system. One way to make this possible could be through spending extra resources on a selected set of sensors to protect them from being attacked.

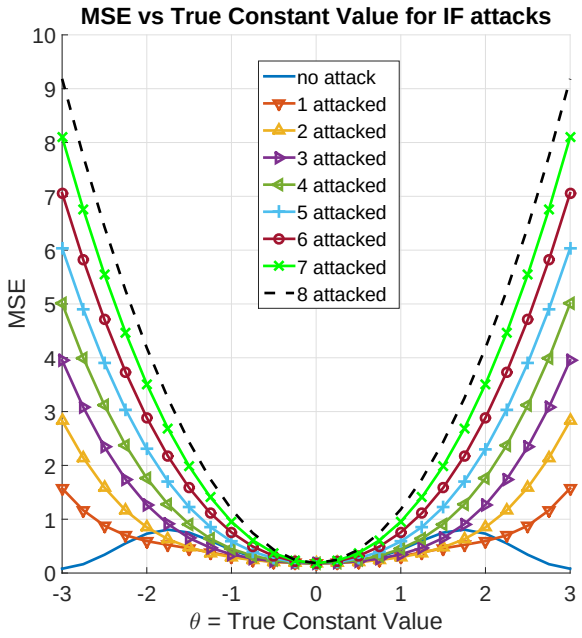


Figure 2.3: MSE for Constant in Truncated Gaussian estimation with $N = 8, K = 1$ showing IF insensitive point.

With a certain number of sensors protected against attacks, the information from those sensors could be used by the fusion center to estimate the nominal unattacked pmf of the quantized data which can then be compared to the estimated pmf of the data originating from unprotected sensors. Now, with no loss of generality, consider an attack on only the first p sensors (all times)[3] with quantized samples $\tilde{\mathbf{u}}_p = \mathbf{r}_p$ which is characterized by the attacked pmf $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta)$ and let $\Pr(\mathbf{u}_p = \mathbf{r}_p|\theta)$ denote the nominal unattacked pmf. Denote the quantized observations at the unattacked sensors as $\mathbf{u}_o = \mathbf{r}_o$ so that the full vector of quantized observations presented to the fusion center is $\tilde{\mathbf{u}} = (\tilde{\mathbf{u}}_p, \mathbf{u}_o) = (\mathbf{r}_p, \mathbf{r}_o)$. A sensor system where the fusion center rejects data when a suitable distance, for example

$$g(\mathbf{p}, \tilde{\mathbf{p}}) = \sum_{\text{all } \mathbf{r}_p} \Big(\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) - \Pr(\mathbf{u}_p = \mathbf{r}_p|\theta) \Big)^2 \tag{2.21}$$

[3]A marginally more general attack model could allow for an attack that chooses, at each time epoch, to either corrupt the measurements or leave them unaltered. We worked out expressions for such attacks and they were deemed to be a trivial extension of the work presented here. Consequently, those results were excluded.

is larger than a certain threshold, d, will have a maximum MSE given by

$$\max_{\{\Pr(\tilde{\mathbf{u}}_p=\mathbf{r}_p|\theta),\forall \mathbf{r}_p\}} \sum_{\text{all } \mathbf{r}_p} \sum_{\text{all } \mathbf{r_o}} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta)\Pr(\mathbf{u}_o = \mathbf{r}_o|\tilde{\mathbf{u}}_p = \mathbf{r}_p, \theta)(\hat{\theta}((\mathbf{r}_p, \mathbf{r}_o)) - \theta)^2$$

$$\text{subject to } \sum_{\text{all } \mathbf{r}_p} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) = 1,$$

$$g(\mathbf{p}, \tilde{\mathbf{p}}) \leq d,$$

$$0 < \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) < 1, \forall \mathbf{r}_p, \tag{2.22}$$

which, when compared to the general attacks described in this section turns out to be an OCU attack with the added constraint that $g(\mathbf{p}, \tilde{\mathbf{p}}) \leq d$. For the case where the attacker has no information, it is not possible for the attacker to optimize the attack as proposed in (2.22). Instead, the attacker could resort to some generalization of the information free attack which incorporates the constraint that $g(\mathbf{p}, \tilde{\mathbf{p}}) \leq d$. The powerful entropy concept [76], a natural measure for the randomness of some observations, suggests replacing the objective function in (2.22) with

$$\max_{\{\Pr(\tilde{\mathbf{u}}_p=\mathbf{r}_p|\theta),\forall \mathbf{r}_p\}} \left(- \sum_{\text{all } \mathbf{r}_p} \Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) \log \left(\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) \right) \right), \tag{2.23}$$

under the same constraints. The objective-maximizing solution in this case is the set of values $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta)$ for all $\mathbf{r}_p$. After this solution is obtained, the MSE value is given by substituting the appropriate value for $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta)$ in (2.14). It can be shown that (2.23) reduces to (2.18) when d is sufficiently large so (2.23) does generalize IF attacks in this sense.

There exists many suitable algorithms for solving the optimization problems in (2.22) and (2.23). One possible algorithm is to start with the solution $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) = \Pr(\mathbf{u}_p = \mathbf{r}_p|\theta), \forall \mathbf{r}_p$ and then start shifting probability mass from the $\mathbf{r}_p$ making the smallest contribution to the metric being optimized to the $\mathbf{r}_p$ making the largest contribution, until a constraint is reached. If the constraint reached is $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) = 0$ for some $\mathbf{r}_p$ then the component of the pmf $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta)$ for that $\mathbf{r}_p$ has been emptied and one should start emptying mass from the $\mathbf{r}_p$ making the next smallest contribution to the metric. The optimum solution is reached when $g(\mathbf{p}, \tilde{\mathbf{p}}) = d$ is satisfied or when $\Pr(\tilde{\mathbf{u}}_p = \mathbf{r}_p|\theta) = 1$ for some $\mathbf{r}_p$ which indicates that all the probability mass has been emptied into the $\mathbf{r}_p$ making largest contribution to the objective function. The algorithm is justified based on the fact that each step always increases the objective function without violating any constraints. In fact, each step provides the largest possible increase from the last solution with the

given amount of probability transferred.

Figure 2.4 shows the different achievable MSE curves under different values of the distance constraint value, d in (2.22). The same sensor system employed in Figures 2.1 and 2.2 was used to generate Figure 2.4 with the exception of using only three senors instead of eight. The jumps in the curves are the result of a change in the ordering of the $\mathbf{r}_p$ making the smallest contribution to the objective function for different values of θ. When the $\mathbf{r}_p$s are emptied in a different order for a different θ, a jump in the MSE curve results. From Figure 2.4, it can be seen that smaller values of d result in limiting the attack and having better MSE values.

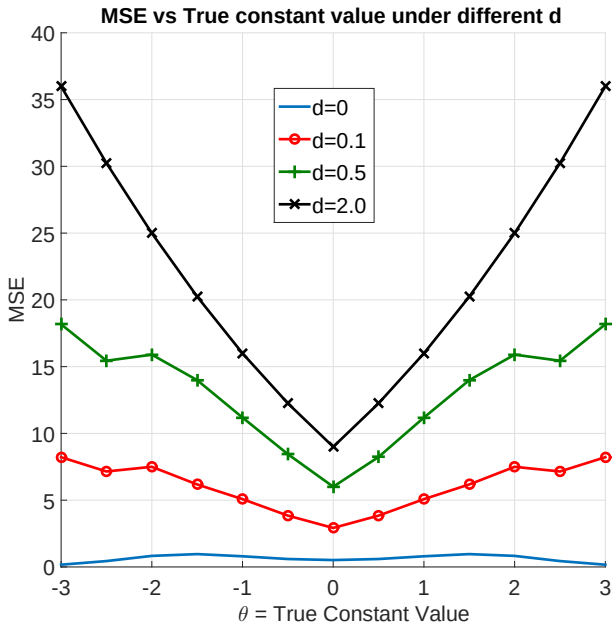


Figure 2.4: MSE for Constant in Truncated Gaussian estimation with $N = 3, K = 1$ with varying distance constraint values under an OCU attack.

2.3 Conclusion

In this chapter, attacks on systems estimating an unknown deterministic parameter based on quantized observations were studied. As bad data detectors are frequently employed in practice, initially, we considered attacks that pass bad data detectors employed at the sensors that check that the sensor data fits the unattacked observation model assumed by the estimation approach. The class of attacks considered include attacks that cannot be represented by the previously studied man-in-the-middle attacks for general estimation problems. For example, they include spoofing attacks which have not been extensively studied for attacks on sensor networks performing

parameter estimation. Assuming that all the observations are independent, we demonstrated that the optimum processing will discard all the observations from the attacked sensors and that the estimation performance cannot be improved beyond that achieved through the optimum processing of unattacked sensor data for cases with a sufficient number of observations, provided the sets of attacked sensors are identified at the fusion center. These results are valid for arbitrary quantization schemes, any estimation problem, and any attack which passes the bad data detectors.

In the second section, we provided a general approach to characterize the after-attack estimation performance of any optimum or suboptimum approach under any general type of attack, with no assumptions on the estimation problem, the observation models, the number of sensors or observations, or the dependence of the observations. The general attacks were classified into the full information OCC and OCU attacks and the generally less effective IF attacks. Expressions which describe the performance of any specific estimation algorithm under these attack classes were provided. We showed that a quantizer can be designed such that there exists a value of the parameter at which the estimation performance is insensitive to IF attacks regardless of the number of attacked sensors. Further, a case where the after-attack MSE under an IF attack is smaller than the unattacked MSE was presented. To account for some level of attack detection carried out by the sensor system, constraints on the attack classes were considered and methods to obtain the after-attack MSE in these cases were described.

Chapter 3

Ocean Wave Estimation

It is well known that the optimal control that maximizes the power absorbed by a wave energy converter (WEC) is future-dependent (non-causal). In most cases, this means that in order to enhance the efficiency of energy conversion, the system controlling the WEC ought to be guided by estimates of the impending time waveforms of the ocean waves. In fact, there are several control strategies, described in [32, 77, 78], which have shown that through relying on estimates of future waves, delivered just a few seconds ahead, it is possible to achieve a twofold increase in the conversion efficiency of WEC devices. Additionally, more recent work on novel control strategies of WEC devices, for example the work in [33, 34], has continued to rely on the use of estimates for future ocean waves.

The idea of using (short-term) estimates of future waves for enhanced WEC efficiency was originally suggested by the authors of [79, 80]. The work in [81] provided experimental results demonstrating the possible benefits of using sensors for the estimation of ocean waves. The same work also described some heuristic measures for obtaining better predictions but those measures did not account for noise in the measurements and errors in the predictions. Requirements for the provided estimates (or predictions) and how they change in relation to a device's hydrodynamic properties were studied in [82]. Some of the more recent work on ocean wave estimation, like that in [83, 84], proposed adopting purely stochastic univariate time series solutions where predictions of wave elevation at a certain point are generated based solely on past measurements of the same type at that same location. While this approach has its benefits, stemming mainly from its adoption of simplified models that do not require estimating the directionality of the waves, it may suffer from some drawbacks. One such drawback is that whenever an estimate is produced, a univariate

and purely stochastic time series approach will discard any valuable information contained in measurements collected at other nearby locations which, if properly employed, may lead to more accurate estimates. Also, such an a approach is unable to provide any information regarding the directionality of the waves, which could become important when several WECs are operating in a farm.

In this chapter, we consider an alternate approach of having a network of spatially distributed sensors where all their measurements are fused together to create a spatial and temporal description of the wave surface surrounding the sensors and the WEC devices. Having a full description of the wave surface means that we are able to provide estimates pertaining to locations that are possibly different from the locations of the sensors. Another advantage of using this approach is that the number of sensors in the network is no longer related to the number of WEC devices or the number of locations at which the estimates are needed. Therefore, we may employ a number of sensors that is larger than the number of WECs in order to collect a larger number of time samples, obtained within the same amount of time, and hence produce more accurate predictions. Thus, for robustness under failures, it may suffice to have a few extra sensors in the network that allow us to produce accurate estimates even if a small number of sensors in the network fail over time whereas univariate time series approaches would require each sensor to have its own backup. Since the approach we consider has the potential to provide more information about the ocean surface, it does suffer from having a fairly more complex wave model and is typically associated with relatively higher upfront costs when compared to the other method. However, because our approach has the potential to provide very valuable information that is especially useful for novel WEC control methods, we believe that studying this approach, where a lot of work remains to be explored and completed, is of great interest from both research and practical prospectives. The authors of [85] and [86] also considered employing a network of spatially distributed sensors but their work was purely deterministic and did not account for noise in the sesnors' measurements. As a result, the performance of the filters designed in [85] and [86] under noisy measurements is not known as they have not been compared to any bounds on performance including the Cramer-Rao Bound (CRB) which, as demonstrated by our numerical results, is shown to be nearly achievable. It is worth mentioning that the resulting wave models under our approach involve many complexities that we do not claim to have solved or completely answered and that, through the work presented here, we rather aim to contribute to the advancement of the work attempting to address some of them. We believe that having the ability to compare approaches relying on measurements from spatially

distributed sensors to those employing univariate time series approaches is of great value and that much remains to be understood about the performance-cost trade-offs that exist between the two methods. A necessary first step towards that comparison is to develop an accurate estimation approach under an array-based model which we present in this chapter.

The presented work derives general expressions for the Fisher Information Matrix (FIM) and the CRB. The CRB tightly bounds the smallest mean square error (MSE) of any unbiased estimation approach. This is the first work to provide closed form expressions providing a bound on the accuracy of the estimation for the problem of estimating ocean waves. These expressions are general in the sense that the set of collected measurements is not restricted to originating from a single type of sensor, elevation or acceleration for example, as it allows the measurements to be a mixture of any number of types from a given set of possible sensor types. In the chapter, we also present a low-complexity method for the parametric estimation of the quantity of interest (elevation for example) and provide numerical results highlighting the accuracy of the method. Although our method is suboptimal, its performance seems to come very close to the CRB for almost all the cases we considered under the assumptions specified in this chapter. Moreover, as shown by the numerical results presented in this chapter, the suggested method has a running time that greatly improves on that of more conventional methods typically employed for solving problems similar to the one addressed in this chapter. All the work in this chapter is conducted while describing the ocean as a sum of incident plane waves, each parameterized by amplitude, frequency, direction, and phase. Such a wave model is obtained under stated fluid mechanics assumptions and the assumption that the sensors employed to collect the measurements are sufficiently small to have no effect on the wave field.

In the following section the wave model used to describe the sensor measurements is formally introduced. Section 3.2 describes the CRB and includes all the general equations for calculating the FIM and the CRB under noisy observations. In Section 3.3, we present our estimation approach and describe several key ideas regarding its operation. Section 3.4 presents numerical examples demonstrating the performance of our proposed approach relative to the CRB. In Section 3.5, we propose a method to quantify the loss in mean absorbed power by WEC devices due to errors in estimation. Finally, Section 3.6 summarizes our results and concludes the work presented in this chapter.

3.1 Ocean Wave Model

For our analysis involving the plane waves, we use the well accepted standard wave equation model from [24] where we assume the ocean is an ideal incompressible fluid with no loss of mechanical energy. We also adopt the common assumptions that the fluid motion is irrotational and that the wave amplitudes are small enough so that linear theory is applicable. Moreover, the deployment area in the ocean is assumed to be of sufficient depth such that finite depth effects, other than dispersion, are small. Finally, we assume that the incident waves were created by forcing functions, distant storms for example, that were applied at sufficient distances away resulting in the observation of fully developed ocean waves. These assumptions are used extensively in the area of control for ocean WEC devices and are generally well accepted. Under the just described assumptions, a measurement of any of the quantities identified in Table 3.1 made at any location $(x, y)^T$ in the two-dimensional incident field and any time t of interest is described by the general expression[1]

$$\psi(x, y, t; \boldsymbol{\theta}) = \sum_{i=1}^{L} \sum_{j=1}^{M_i} A_{i,j} w_i^a \cos\left((\frac{w_i^2}{g}) x \cos(\beta_{i,j}) + (\frac{w_i^2}{g}) y \sin(\beta_{i,j}) - t w_i + \phi_{i,j} - b\frac{\pi}{2} \right), \quad (3.1)$$

which is parameterized by

$$\begin{aligned} \boldsymbol{\theta} = & (A_{1,1}, A_{1,2}, \ldots, A_{1,M_1}, A_{2,1}, \ldots, A_{L,M_L}, \\ & \omega_1, \ldots, \omega_L, \\ & \beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,M_1}, \beta_{2,1}, \ldots, \beta_{L,M_L}, \\ & \phi_{1,1}, \phi_{1,2}, \ldots, \phi_{1,M_1}, \phi_{2,1}, \ldots, \phi_{L,M_L})^T \end{aligned} \quad (3.2)$$

where $A_{i,j}$ is the amplitude in meters, ω_j is the frequency in radians per second, $\beta_{i,j}$ is the angular direction in radians measured relative to the x-axis, $\phi_{i,j}$ is the phase in radians and the following

[1] While L and M_i, $i = 1, \ldots, L$ are often picked based on experimental investigations by WEC manufacturers and researchers, there are studies in the signal processing community [87] that suggest methods to optimize these choices, but we omit such a discussion for brevity and assume M_i $i = 1, \ldots, L$ and L are known here.

hold true.

$$0 < \ A_{i,j} \quad \forall i \text{ and } \forall j, \tag{3.3}$$

$$0 < \ w_i \quad \forall i, \text{ and } w_i \neq w_{i'} \quad \text{for } i \neq i', \tag{3.4}$$

$$0 \leq \ \beta_{i,j} < 2\pi \quad \forall i \text{ and } \forall j, \tag{3.5}$$

$$0 \leq \ \phi_{i,j} < 2\pi \quad \forall i \text{ and } \forall j. \tag{3.6}$$

The constants 'a' and 'b' in (3.1) are integer constants whose values are set according to Table 3.1 to determine the quantity under consideration. The quantities provided in Table 3.1 are a subset of a larger group of possible types of measurements, provided in our preliminary work [88] and [89] and the work presented here could be easily extended to include the larger set of measurements.

Since the wave field, given by (3.1), is parameterized by the unknown $\boldsymbol{\theta}$, given by (3.2), it is by producing estimates of $\boldsymbol{\theta}$ that we are able to generate, at any point in space and any time, future predictions of $\psi(x, y, t; \boldsymbol{\theta})$ which could be used for control. While $\boldsymbol{\theta}$ will slowly change over long periods of time, it may be assumed to remain constant for few minutes at a time, at least over the area monitored by the sensors. This assumption was adopted in [78] and [85] and as long as the values of the parameters are continuously re-estimated several times over far shorter time periods, this assumption is a reasonable one.

Table 3.1: Integer Constant Values for Selected Predicted Quantities or Sensor Measurements

Sensor Measurement	a	b
Surface Elevation	0	0
Vertical Surface Velocity	1	1
Vertical Surface Acceleration	2	0

In the next section, we derive general expressions for the FIM and CRB.

3.2 Cramer-Rao Bound

In this section we derive the CRB [90] on the MSE for the prediction, made at location $(x_p, y_p)^T$ and time t_p, of the value $\psi(x_p, y_p, t_p; \boldsymbol{\theta})$ given values for the constants 'a' and 'b' specified according to one of the rows from Table 3.1. We consider having a network with N sensors, possibly of different types, each having known position $(x_r, y_r)^T$, $r = 1, \ldots, N$ and providing the noisy sensor

measurements

$$\Psi(x_r, y_r, t_m) = \psi(x_r, y_r, t_m; \boldsymbol{\theta}) + v_r(t_m), \quad t_m = 0, \ldots, (K-1)T_s \tag{3.7}$$

where T_s is the sampling period, K is the total number of time samples collected by each sensor, and $v_r(t_m)$ represents noise and distortion. The noise in the measurements is meant to capture errors due to electrical and thermal noise and other random measurement errors in the system which are bound to exist. For simplicity, assume

$$(v_1(0), \ldots, v_N(0), \ldots, v_1((K-1)T_s), \ldots, v_N((K-1)T_s))^T$$

is jointly Gaussian with a zero mean vector and a covariance matrix that is diagonal with $(\sigma^2_{1,0}, \ldots, \sigma^2_{N,K-1})$ along the diagonal. Then, the joint probability density function (pdf) of the observations conditioned on $\boldsymbol{\theta}$, often called the likelihood function, is

$$f_\Psi(\Psi; \boldsymbol{\theta}) = \prod_{r=1}^{N} \prod_{k=0}^{K-1} \frac{1}{\sqrt{2\pi\sigma^2_{r,k}}} \exp\left(\frac{-\big(\Psi(x_r, y_r, kT_s) - \psi(x_r, y_r, kT_s; \boldsymbol{\theta})\big)^2}{2\sigma^2_{r,k}} \right). \tag{3.8}$$

Let $\hat{\boldsymbol{\theta}}$ and $\hat{\psi}(x_p, y_p, t_p; \hat{\boldsymbol{\theta}})$ be unbiased estimators of $\boldsymbol{\theta}$ and $\psi(x_p, y_p, t_p; \boldsymbol{\theta})$, respectively. Then by [91, 92] the MSE of the estimator $\hat{\psi}(x_p, y_p, t_p; \hat{\boldsymbol{\theta}})$, which is equal to the variance in this case, must satisfy

$$\text{var}(\hat{\psi}(x_p, y_p, t_p)) \geq q(\boldsymbol{\theta})^T J(\boldsymbol{\theta})^{-1} q(\boldsymbol{\theta}), \tag{3.9}$$

where the quantity on the right hand side of (3.9) is the CRB which is computed using the vector $q(\boldsymbol{\theta})$ and the FIM, $\mathbf{J}(\boldsymbol{\theta})$. The vector $q(\boldsymbol{\theta})$ is defined as

$$\begin{aligned} q(\boldsymbol{\theta}) &= q(x_p, y_p, t_p; \boldsymbol{\theta}) \\ &= \left(\frac{\partial \psi(x_p, y_p, t_p; \boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial \psi(x_p, y_p, t_p; \boldsymbol{\theta})}{\partial \theta_{3M+L}} \right)^T, \end{aligned} \tag{3.10}$$

where $\theta_i, i = 1, 2, \ldots, 3M + L$ is the i^{th} element in $\boldsymbol{\theta}$ which was given in (3.2) and $M = \sum_{i=1}^{L} M_i$. The $\ell - n^{th}$ element of $\mathbf{J}(\boldsymbol{\theta})$ is calculated according to [91, 92] and is given by

$$J_{\ell,n}(\boldsymbol{\theta}) = E\left\{ \frac{\partial}{\partial \theta_\ell} \ln f_\Psi(\Psi; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_n} \ln f_\Psi(\Psi; \boldsymbol{\theta}) \right\}, \tag{3.11}$$

which is a function of the sensor locations and the sampling times but is not a function of $(x_p, y_p)^T$ and t_p, the location and time of prediction respectively. The FIM also allows us to obtain the CRB for the MSE in estimating the elements of $\boldsymbol{\theta}$ since the MSE for estimating the i^{th} element in $\boldsymbol{\theta}$, equal to the variance in this case, must satisfy

$$\text{var}(\hat{\theta}_i) \geq \left[J(\boldsymbol{\theta})^{-1} \right]_{i,i}, \quad i = 1, \ldots, 3M + L. \tag{3.12}$$

We note that $(x_p, y_p)^T$ is not limited to locations where sensors are available and could be any location of interest. We also note that, for almost all the considered situations, we did not observe any significant difference between the CRB values at sensor locations and the CRB values at nearby locations where no measurements (sensors) are available. Under our assumptions about the noise, (3.11) takes on the general form

$$J_{\ell,n}(\boldsymbol{\theta}) = \sum_{r=1}^{N} \sum_{k=0}^{K-1} \frac{1}{\sigma_{r,k}^2} \left(\frac{\partial}{\partial \theta_\ell} \psi(x_r, y_r, kT_s; \boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \theta_n} \psi(x_r, y_r, kT_s; \boldsymbol{\theta}) \right) \tag{3.13}$$

which is a sum over the product of two derivatives of (3.1) with respect to either amplitude, direction, frequency or phase depending on the parameters that θ_ℓ and θ_n correspond to in $\boldsymbol{\theta}$. We derived the expressions for these derivatives, and give them below as

$$\frac{\partial}{\partial A_{i,j}} \Phi(x_r, y_r, t) = w_i^{a_r} \cos\left(\frac{w_i^2}{g} \left(x_r \cos(\beta_{i,j}) + y_r \sin(\beta_{i,j}) \right) - t w_i + \phi_{i,j} - b_r \frac{\pi}{2} \right), \tag{3.14}$$

$$\frac{\partial}{\partial w_i}\Phi(x_r, y_r, t) = \sum_{j=1}^{M_i} A_{i,j} a_r w_i^{a_r - 1}$$
$$\cos\left(\frac{w_i^2}{g}\big(x_r\cos(\beta_{i,j}) + y_r\sin(\beta_{i,j})\big) - tw_i + \phi_{i,j} - b_r\frac{\pi}{2}\right) -$$
$$A_{i,j} w_i^{a_r}\sin\left(\frac{w_i^2}{g}(x_r\cos(\beta_{i,j}) + y_r\sin(\beta_{i,j})) - tw_i\right.$$
$$\left. + \phi_{i,j} - b_r\frac{\pi}{2}\right)\left(\frac{2w_i}{g}(x_r cos(\beta_{i,j}) + y_r sin(\beta_{i,j})) - t\right) \tag{3.15}$$

$$\frac{\partial}{\partial \beta_{i,j}}\Phi(x_r, y_r, t) = A_{i,j} w_i^{a_r}(\frac{w_i^2}{g})(x_r\sin(\beta_{i,j}) - y_r\cos(\beta_{i,j}))$$
$$\sin\left(\frac{w_i^2}{g}(x_r\cos(\beta_{i,j}) + y_r\sin(\beta_{i,j})) - tw_i + \phi_{i,j} - b_r\frac{\pi}{2}\right), \tag{3.16}$$

and

$$\frac{\partial}{\partial \phi_{i,j}}\Phi(x_r, y_r, t) = -A_{i,j} w_i^{a_r}\sin\left(\frac{w_i^2}{g}(x_r\cos(\beta_{i,j}) + y_r\sin(\beta_{i,j})) - tw_i + \phi_{i,j} - b_r\frac{\pi}{2}\right), \tag{3.17}$$

where the values of the constants a_r and b_r determine the type of the measurement available at $(x_r, y_r)^T$. A numerical example where the CRB is calculated using the just provided expressions is given in section 3.4 where we compare the accuracy our proposed estimation approach, which we present next, to the CRB.

3.3 The Parametric Estimation of Ocean Waves

For this section, we consider the same network setup introduced in Section 3.2 where we have a network of N distributed sensors with known locations $(x_r, y_r)^T$, $r = 1, \ldots, N$. Since the choice of the origin is arbitrary, we will assume, with no loss of generality, that $(x_1, y_1) = (0, 0)^T$. We adopt the signal model for the noisy sensor measurements given by (3.7). Further, throughout this section we assume that[2] $\sigma_{1,0}^2 = \cdots = \sigma_{N,K-1}^2 = \sigma^2$.

Under the just stated assumptions and before introducing our suggested estimation approach, let us consider the maximum likelihood (ML) estimator for $\boldsymbol{\theta}$. The ML estimate $\hat{\boldsymbol{\theta}}_{ML}$ is defined

[2]The assumption that $\sigma_{1,0}^2 = \cdots = \sigma_{N,K-1}^2 = \sigma^2$ is adopted only to simplify the presentation of our method. Once the presented method is understood, the extension to the case where it possible that $\sigma_{1,0}^2 \neq \cdots \neq \sigma_{N,K-1}^2$ is fairly straightforward.

as $\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \ln(f_\Psi(\Psi;\boldsymbol{\theta}))$ where $L(\Psi;\boldsymbol{\theta}) \triangleq \ln(f_\Psi(\Psi;\boldsymbol{\theta}))$ is the natural logarithm of the likelihood function (3.8), typically called the log-likelihood function, which in our case is given by

$$L(\Psi;\boldsymbol{\theta}) = \ln(2\pi\sigma^2)^{\frac{-NK}{2}} - \frac{1}{2\sigma^2}\sum_{r=1}^{N}\sum_{k=0}^{K-1}(\Psi(x_r,y_r,kT_s) - \psi(x_r,y_r,kT_s;\boldsymbol{\theta}))^2, \tag{3.18}$$

and since the first term in (3.18) does not depend on $\boldsymbol{\theta}$ and neither does $\frac{1}{2\sigma^2}$, it can be shown that the ML estimate $\hat{\boldsymbol{\theta}}_{ML}$ can be described as

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \sum_{r=1}^{N}\sum_{k=0}^{K-1}(\Psi(x_r,y_r,kT_s) - \psi(x_r,y_r,kT_s;\boldsymbol{\theta}))^2, \tag{3.19}$$

which is the value of $\boldsymbol{\theta}$ that minimizes the sum of the squared error over all time measurements and all sensors (least squares fit). For a simple convex problem, the ML estimate could be found using derivatives of the objective function but, unfortunately, the considered problem is non-convex which hinders the use of such methods due to the existence of local extrema.

We now present our proposed estimation approach in the following subsection.

3.3.1 A Proposed Estimation Approach

Consider the general wave model given in (3.1). If none of the component waves in (3.1) have the same frequency, then the inner sum in (3.1) would be over only a single value and we would essentially have a single sum. Now, as an approximation, if we take all the waves in (3.1) that share the same frequency and just slightly perturb the frequency values for those waves such that the resulting frequency values are all unique yet very close to the unperturbed value, then we essentially have a single sum approximation of (3.1). A step-by-step explanation for reducing the double sum to a single sum is given in the appendix for this chapter. Having the frequencies of all the waves be different provides us with great benefits when it comes to the design of the estimation algorithm, which we present next, and since the frequencies could be made arbitrarily close to their original value, the amount of loss due to the modeling error can be controlled and driven to be smaller than any predetermined acceptable value. We believe that such an approximation is reasonable, especially since the amount of modeling error introduced by this approximation is controllable. Therefore, using the just described approximation and through re-indexing the parameter values

in (3.1), we will adopt, as our wave model for the remainder of this chapter, the expression

$$\psi(x,y,t;\boldsymbol{\theta}) = \sum_{i=1}^{M} A_i w_i^a \cos\left((\frac{w_i^2}{g}) x\cos(\beta_i) + (\frac{w_i^2}{g}) y \sin(\beta_i) - t w_i + \phi_i - b\frac{\pi}{2} \right), \tag{3.20}$$

where $M = \sum_{i=1}^{L} M_i$ and the unknown parameters are

$$\boldsymbol{\theta} = (A_1, \ldots, A_M, \omega_1, \ldots, \omega_M, \beta_1, \ldots, \beta_M, \phi_1, \ldots, \phi_M)^T. \tag{3.21}$$

However, to ensure that we are able to correctly estimate the frequencies in (3.20), we will assume that K, the number of samples from a single sensor, is chosen sufficiently large such that

$$|w_i - w_{i'}| > \frac{4\pi}{K} \ \forall i \neq i',\ 1 < i < M,\ 1 < i' < M, \tag{3.22}$$

is satisfied. The right-hand side in (3.22) is equal to twice the value of what is known as the frequency or spectral resolution of the employed periodogram [93] approach. Since K is typically a design variable, there will always exist a choice of K that is sufficiently large such that (3.22) is satisfied and the approximation of (3.1) by (3.20) is sufficiently accurate. Lastly, since the ordering of the frequencies $\omega_1, \ldots, \omega_M$ is arbitrary, we will assume, without any loss of generality, that

$$\omega_1 < \omega_2 < \cdots < \omega_M. \tag{3.23}$$

Next, we present our proposed estimation method under **Algorithm 3.1**. The presented method depends on the subroutines *Estimate_Frequencies*, *Linear_LS* and *Find_Betas* given, respectively, by **Subroutine 3.1**, **Subroutine 3.2**, and **Subroutine 3.3**. The presented method employs the function *Size* which accepts an input vector and then simply returns the size, or length, of the vector. and the function *Sort* which takes in an input vector and a sorting mode (either ascending or descending) and returns as output a sorted version of the input vector as well as a vector giving the position in the original vector of the now sorted element. The arctan 2 function which is also known as the four quadrant arctan is also employed. We also employ the

p **modulo** q operation which is defined as the operation returning a value v such that

$$p \textbf{ modulo } q \triangleq v = p - qn, \tag{3.24}$$

$$\text{where } 0 \leq v < q, n \in \mathbb{Z}, \tag{3.25}$$

and $\mathbb{Z}$ is the set of integer numbers.

Algorithm 3.1: Proposed Estimation Approach - (Partial)

Input: Sensor locations $(x_1, y_1)^T$ and types $(a_r, b_r)^T$ for $r = 1, \ldots, N$ with $(x_1, y_1)^T = (0, 0)^T$. Total number of frequencies L. Sampling period T_s. Sensor measurements $\mathbf{Y}_r = (\Psi(x_r, y_r, 0), \ldots, \Psi(x_r, y_r, (K-1)T_s))^T,\ r = 1, \ldots, N$.
Output: $\hat{\boldsymbol{\theta}} = (A_1, \ldots, A_L, \omega_1, \ldots, \omega_L, \beta_1, \ldots, \beta_L, \phi_1, \ldots, \phi_L)^T$
Begin
for $r = 1$ to N **do**
    $\hat{\mathbf{W}}_{r,inital} = Estimate\_Frequencies(L, T_s, \mathbf{Y}_r)$
end for
for $i = 1$ to L **do**
    $\hat{\mathbf{W}}_{init}[i] = \frac{1}{N} \sum_{r=1}^{N} \hat{\mathbf{W}}_{r,init}[i]$
end for
for $r = 1$ to N **do**
    $[\hat{\mathbf{A}}_{r,init}, \hat{\mathbf{\Gamma}}_{r,init}] = Linear\_LS(\hat{\mathbf{W}}_{intial}, a_r, b_r, \mathbf{Y}_r)$
end for
for $i = 1$ to L **do**
    $\hat{\mathbf{A}}_{init}[i] = \frac{1}{N} \sum_{r=1}^{N} \hat{\mathbf{A}}_{r,init}[i]$
end for
for $r = 1$ to N **do**
    Define $\psi'_r(t; \boldsymbol{\theta}'_r) = \sum_{i=1}^{L} A_i \cos(\Gamma_{i,r} - w_i t - b_r \frac{\pi}{2})$ where $\boldsymbol{\theta}'_r = (\mathbf{A}, \mathbf{W}, \mathbf{\Gamma}_{\boldsymbol{r}})^T = (A_1, \ldots, A_L, w_1, \ldots, w_L, \Gamma_{1,r}, \ldots, \Gamma_{L,r})$.
    Use local search algorithm with $\boldsymbol{\theta}'_r = (\hat{\mathbf{A}}_{init}, \hat{\mathbf{W}}_{init}, \hat{\mathbf{\Gamma}}_{r,init})^T$ as initial guess to find a new $\boldsymbol{\theta}'_r = (\hat{\mathbf{A}}, \hat{\mathbf{W}}_r, \hat{\mathbf{\Gamma}}_r)^T$ that minimizes $\sum_{k=0}^{K-1} (\Psi(x_r, y_r, kT_s) - \psi'_r(t; \boldsymbol{\theta}'_r))^2$
    $\hat{\mathbf{W}}_r = (\hat{w}_1, \ldots, \hat{w}_L)^T = \boldsymbol{\theta}'_r[L+1 : 2L]$
end for
for $i = 1$ to L **do**
    $\hat{\mathbf{W}}[i] = \frac{1}{N} \sum_{r=1}^{N} \hat{\mathbf{W}}_r[i]$
end for

The presented approach relies on a few key ideas. The first key idea, utilized in **Subroutine 3.1**, is that of obtaining estimates for frequency as those values of w that correspond to the largest L peaks in the periodogram, which is given by

$$I(w) = \frac{1}{K} \left| \sum_{k=0}^{K-1} \Psi(x_r, y_r, kT_s) \mathrm{e}^{-jwk} \right|^2. \tag{3.26}$$

This idea is well-studied for estimating frequency and is used in several signal processing problems [92].

Algorithm 3.1: Proposed Estimation Approach - Continued

for $r = 1$ to N **do**
  $[\hat{\mathbf{A}}_r, \hat{\mathbf{\Gamma}}_r] = Linear\_LS(\hat{\mathbf{W}}, a_r, b_r, \mathbf{Y}_r)$
end for
for $i = 1$ to L **do**
  $\hat{\mathbf{A}}[i] = \frac{1}{N}\sum_{r=1}^{N} \hat{\mathbf{A}}_r[i]$
end for
$\hat{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}_1$
for $r = 2$ to N **do**
  $\hat{\mathbf{\Xi}}_r = (\hat{\mathbf{\Gamma}}_r - \hat{\boldsymbol{\phi}})$ **modulo** 2π
end for
for $i = 1$ to L **do**
  $\mathbf{P}_i = (\hat{\mathbf{\Xi}}_2[i], \ldots, \hat{\mathbf{\Xi}}_N[i])^T$
  $\hat{\mathbf{W}}_{e,i} = (\hat{\mathbf{W}}[i], \hat{\boldsymbol{W}}_1[i], \hat{\boldsymbol{W}}_2[i], \ldots, \hat{\boldsymbol{W}}_N[i])^T$
  $\boldsymbol{B}_i = Find\_Betas\big(\hat{\mathbf{W}}[i], \mathbf{P}_i, (x_2, y_2)^T, \ldots, (x_N, y_N)^T\big)$
end for
Define $\psi(x_r, y_r, t; \boldsymbol{\theta}) = \sum_{i=1}^{L} A_i w_i^{a_r} \cos\left((\frac{w_i^2}{g})x_r\cos(\beta_i) + (\frac{w_i^2}{g})y_r\sin(\beta_i) - tw_i + \phi_i - b_r\frac{\pi}{2}\right)$ with $\boldsymbol{\theta} = (A_1, \ldots, A_L, \omega_1, \ldots, \omega_L, \beta_1, \ldots, \beta_L, \phi_1, \ldots, \phi_L)^T$.
$\boldsymbol{\theta}_{possible} = (\hat{\mathbf{A}}, \hat{\mathbf{W}}, \boldsymbol{B}_1[1], \boldsymbol{B}_2[1], \ldots, , \boldsymbol{B}_L[1], \hat{\boldsymbol{\phi}})^T$
for $i = 1$ to L **do**
  for $j = 1$ to $Size(\boldsymbol{B}_i)$ **do**
    $\boldsymbol{\theta}_{possible}[2L + i] = B_i[j]$
    $\text{fit}[i, j] =$

$$\sum_{r=1}^{N}\sum_{k=0}^{K-1}(\Psi(x_r, y_r, kT_s) - \psi(x_r, y_r, kT_s; \boldsymbol{\theta}_{possible}))^2$$

  end for
  $j^* = \min(fit[i, j])$
  $\boldsymbol{\beta}[i] =^{j} \boldsymbol{B}_i[j^*]$
end for
$\hat{\boldsymbol{\theta}} = (\hat{\mathbf{A}}, \hat{\mathbf{W}}, \boldsymbol{\beta}, \hat{\boldsymbol{\phi}})^T$
(optional) Use local search algorithm with $\hat{\boldsymbol{\theta}}$ as initial guess to find a new $\hat{\boldsymbol{\theta}}$ that minimizes $\sum_{r=1}^{N}\sum_{k=0}^{K-1}(\Psi(x_r, y_r, kT_s) - \psi(x_r, y_r, kT_s; \boldsymbol{\theta}_{possible}))^2$
Return $\hat{\boldsymbol{\theta}}$

The second key idea, utilized in **Subroutine 3.2**, is based on repeatedly employing the trigonometric identity

$$\cos(P_{i,r} - Q_i) = \cos(P_{i,r})\cos(Q_i) + \sin(P_{i,r})\sin(Q_i), \tag{3.27}$$

in equation (3.1) with

$$P_{r,i} = (\frac{w_i^2}{g})x_r\cos(\beta_i) + (\frac{w_i^2}{g})y_r\sin(\beta_i) + \phi_i - b\frac{\pi}{2}, \tag{3.28}$$
$$Q_i = tw_i, \quad i = 1, 2, \ldots, L,$$

Subroutine 3.1: Estimate_Frequencies

Input: Total number of frequencies L. Sampling period T_s. Sensor measurements $\mathbf{Y}_r = (\Psi(x_r, y_r, 0), \ldots, \Psi(x_r, y_r, (K-1)T_s))^T$
Output: $\hat{\mathbf{W}} = (\hat{w}_1, \ldots, \hat{w}_L)^T$
Begin
for $i = 1$ to $1/2T_s$ **do**
    $\Omega[i] = 2\pi(i-1)/(T_s 2^{\lceil \log_2(Size(\mathbf{Y}_r)) \rceil})$
    $I[i] = \frac{1}{K}\left| \sum_{k=0}^{K-1} \Psi(x_r, y_r, kT_s)\exp(-j\Omega[i]k) \right|^2$
end for
for $i = 2$ to $Size(I) - 1$ **do**
    if $I[i] > I[i-1]$ and $I[i] > I[i+1]$ **then**
        peak$[i-1] = I[i]$
        peak_freq$[i-1] = \Omega[i]$
    end if
end for
[peak_sorted,index] = $Sort$(peak, 'descending')
for $i = 1$ to L **do**
    $\hat{\mathbf{W}}[i]$ =peak_freq[index[i]]
end for
[$\hat{\mathbf{W}}$, index2] = $Sort$($\hat{\mathbf{W}}$, 'ascending')
Return $\hat{\mathbf{W}}$

such that the observations from a sensor located at $(x_r, y_r)^T$, collected in the vector

$$\mathbf{Y}_r = (\Psi(x_r, y_r, 0), \ldots, \Psi(x_r, y_r, (K-1)T_s))^T, \tag{3.29}$$

where $\Psi(x_r, y_r, kT_s)$ for $k = 0, 1, \ldots, (K-1)$ is defined in (3.7), are described by the linear model

$$\mathbf{Y}_r = \mathbf{H}_r(\alpha_{r,1,1}, \alpha_{r,1,2}, \ldots, \alpha_{r,L,1}, \alpha_{r,L,2})^T + (v_r(0), v_r(T_s), \ldots, v_r((K-1)T_s))^T \tag{3.30}$$

where the noise terms $v_r(kT_s),\ k = 0, 1, \ldots, (K-1)$ are as defined in (3.7), the observation matrix $\mathbf{H}_r$ is obtained according to **Step 4** through **Step 7** in **Subroutine 3.2**, and

$$\alpha_{r,i,1} = A w_l^{a_r} \cos(P_{r,i}), \tag{3.31}$$

$$\alpha_{r,i,2} = A w_l^{a_r} \sin(P_{r,i}), \tag{3.32}$$

where $P_{i,r}$ is given by (3.28). Given this linear model we implement a linear least squares estimator, which is known to have desirable properties, in **Step 8** of **Subroutine 3.2**.

The third key idea employed in the presented approach is recognizing that the initial estimates from executing **Step 4** through **Step 15** of **Algorithm 3.1** can be employed as great starting

Subroutine 3.2: Linear_LS

1: **Input:** Frequency estimates $\hat{W} = (\hat{w}_1, \ldots, \hat{w}_L)^T$, sensor constants a_r, b_r, and sensor measurements $\mathbf{Y}_r$
2: **Output:** $\hat{\mathbf{A}}_r = (\hat{A}_{1,r}, \ldots, \hat{A}_{L,r})^T$ and $\hat{\mathbf{\Gamma}}_r = (\hat{\Gamma}_{1,r}, \ldots, \hat{\Gamma}_{L,r})^T$
3: **Begin**
4: **for** $l = 1$ to $Size(\hat{W})$ **do**
5: $\quad \hat{\mathbf{H}}_{l,r} = \begin{bmatrix} \cos(\frac{b_r\pi}{2}) & \sin(\frac{b_r\pi}{2}) \\ \cos(\hat{\omega}_l T_s - \frac{b_r\pi}{2}) & \sin(\hat{\omega}_l T_s + \frac{b_r\pi}{2}) \\ \vdots & \vdots \\ \cos(\hat{\omega}_l (K\text{-}1) T_s - \frac{b_r\pi}{2}) & \sin(\hat{\omega}_l (K\text{-}1) T_s + \frac{b_r\pi}{2}) \end{bmatrix}$
6: **end for**
7: $\hat{\mathbf{H}}_r = [\hat{\mathbf{H}}_{1,r}\ \hat{\mathbf{H}}_{2,r}\ \ldots\ \hat{\mathbf{H}}_{L,r}]$
8: $(\hat{\alpha}_{1,1},\ \hat{\alpha}_{1,2}, \ldots, \hat{\alpha}_{L,1},\ \hat{\alpha}_{L,2})^T = (\hat{\mathbf{H}}_r^T \hat{\mathbf{H}}_r)^{-1} \hat{\mathbf{H}}_r^T \mathbf{Y_r}$
9: **for** $l = 1$ to $Size(\hat{W})$ **do**
10: $\quad \hat{A}_{l,r} = \sqrt{(\hat{\alpha}_{l,1})^2 + (\hat{\alpha}_{l,2})^2}/(\hat{w}_l)^{a_r}$
11: $\quad \hat{\Gamma}_{l,r} = \arctan 2(\hat{\alpha}_{l,2}, (-1)^{b_r} \hat{\alpha}_{l,1})$ **modulo** 2π
12: **end for**
13: $\hat{\mathbf{A}}_r = (\hat{A}_{1,r}, \ldots, \hat{A}_{L,r})^T$
14: $\hat{\mathbf{\Gamma}}_r = (\hat{\Gamma}_{1,r}, \ldots, \hat{\Gamma}_{L,r})^T$
15: **Return** $\hat{\mathbf{A}}_r, \hat{\mathbf{\Gamma}}_r$

points (guesses) to standard iterative optimization algorithms. Since these starting points are fairly accurate to begin with, the optimization algorithms will typically converge relatively quickly and have favorable running times. Safeguards limiting the maximum number of iterations for the optimization algorithm could be employed in order to limit the total running time of the algorithm but this might cause losses in the accuracy of the estimates.

The fourth and most important key idea is that used in **Subroutine 3.3** which is employed to facilitate the estimation of the directions $\beta_i,\ i = 1, \ldots, L$. The problem that **Subroutine 3.3** solves stems from the fact that the estimates of $\Gamma_{i,r},\ r = 2, \ldots, N,\ i = 1, \ldots, L$, obtained in **Step 11** and **Step 25** of **Algorithm 3.1**, which contain the information regarding the wave directions are produced through an inverse tangent operation whose range is limited to $[-\pi, \pi]$. Due to this fact, each of the obtained $\Gamma_{i,r}$'s is possibly shifted by some unknown integer multiple of 2π and instead of obtaining estimates of

$$\gamma_{i,r} \triangleq (\frac{w_i^2}{g}) x \cos(\beta_i) + (\frac{w_i^2}{g}) y \sin(\beta_i) + \phi_i, \tag{3.33}$$

we are obtaining estimates of

$$\Gamma_{i,r} \triangleq \gamma_{i,r} \textbf{ modulo } 2\pi. \tag{3.34}$$

Subroutine 3.3: Find_Betas

Input: Frequency estimate $\hat{w}$ and local frequency estimates $\hat{w}_1, \ldots, \hat{w}_N$, collected into $W = (\hat{w}, \hat{w}_1, \ldots, \hat{w}_N)$. Vector of phase terms $\mathbf{P}$. Sensor locations $(x_r, y_r)^T,\ r = 2, \ldots, N$
Output: The vector *Possible_betas*
Begin
$\mathcal{L} = \{0\}$
for $r = 2$ to N **do**
    $\mathcal{L}_r = \{\}$
    for $i = 1$ to $N + 1$ **do**
        $n_r = \left\lfloor \frac{(W[i])^2\sqrt{(x_r)^2+(y_r)^2}}{2\pi g} \right\rfloor$
        for k $= -n_r$ to n_r **do**
            $\zeta_1 = \arccos\left(\frac{k(2\pi g)}{(W[i])^2\sqrt{(x_r)^2+(y_r)^2}}\right) + \arctan 2(y_r, x_r)$
            $\mathcal{L}_r = \mathcal{L}_r \cup (\zeta_1$ **modulo** $2\pi)$
            $\mathcal{L}_r = \mathcal{L}_r \cup ((\zeta_1 - \pi)$ **modulo** $2\pi)$
        end for
    end for
    $\mathcal{L} = \mathcal{L} \cup \mathcal{L}_r$
end for
$\mathcal{S} = Sort(\mathcal{L}, \text{'ascending'})$
$\mathbf{R} = \left[(x_2 \ldots x_N)^T, (y_2 \ldots y_N)^T\right]$
for $j = 1$ to $Size(\mathcal{S}) - 1$ **do**
    $\zeta_2 = 0.5(\mathcal{S}[j] + \mathcal{S}[j+1])$
    for $r = 1$ to $N - 1$ **do**
        $\boldsymbol{Q}[r, j] = floor\left(\frac{(\hat{w})^2}{2\pi g}\left(x_r \cos(\zeta_2) + y_r \sin(\zeta_2)\right)\right)$
    end for
end for
for $j = 1$ to $Size(\mathcal{S}) - 1$ **do**

$$\begin{bmatrix} \hat{\rho}_{r,1} \\ \hat{\rho}_{r,2} \end{bmatrix} = (\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T \begin{bmatrix} \mathbf{P}[1] + 2\pi\boldsymbol{Q}[1,j] \\ \mathbf{P}[2] + 2\pi\boldsymbol{Q}[2,j] \\ \vdots \\ \mathbf{P}[N\text{-}1] + 2\pi\boldsymbol{Q}[N\text{-}1,j] \end{bmatrix}$$

    $Possible\_betas[j] = \arctan 2(\hat{\rho}_{r,2}, \hat{\rho}_{r,1})$
end for
Return *Possible_betas*

Subroutine 3.3 generates a list of possible combinations of $(n_{i,2}, \ldots, n_{i,N})^T$, the number of integer multiples of 2π, that ought to be added to the obtained estimates of $\Gamma_{i,r}$ such that $\gamma_{i,r} = \Gamma_{i,r} + 2\pi n_{i,r}, i = 1, \ldots, L, r = 2, \ldots, N$. While one might expect the number of possible combinations generated to grow exponentially with N, the clever construction of these combinations by **Subroutine 3.3** results in the total number of possible combinations of $(n_{i,2}, \ldots, n_{i,N})^T$ being at most $2N^2\left(2\left\lfloor \frac{(w_i)^2 D}{2\pi g} \right\rfloor + 1\right)$ which is polynomial in N since $D \triangleq \max_{r=2,\ldots,N} \sqrt{(x_r)^2 + (y_r)^2}$. The great reduction in the number of generated combinations stems from the fact that **Subroutine 3.3** takes advantage of knowing the locations of the sensors and judiciously avoids generating combinations of $(n_{i,2}, \ldots, n_{i,N})^T$ that are not possible under all values of β_i given the locations of the

sensors. Since the number of generated combinations is relatively small, and the correct combination lies amongst the generated combinations, the proposed approach tries all the possible combinations and generates a list of possible values for each $\beta_i, i = 1, 2, \ldots, L$ and then, for each i, chooses the one that results in the best squares fit over all sensor measurements and sensors locations as the estimate of β_i.

3.4 Discussion and Numerical Results

In this section we present numerical results that demonstrate the estimation accuracy of the presented approach and also compare its runtime to that of possible global optimization methods. We first consider the case of a regular ocean environment and then consider the case where the ocean environment is described by 3 different component waves. We also consider having correlated noise for a case where the ocean environment is described by 6 components waves.

3.4.1 Estimation Performance: Simple Ocean Environment

In this subsection we consider the simple case where the ocean is described by a single component wave. That is, we assume the ocean is given by (3.20) with $L = 1$ and that the value of the unknown variable $\boldsymbol{\theta} = (A, w, \beta, \phi)^T$ is given. For the result presented in Figure 3.1, we set the value of the unknown variable to $\boldsymbol{\theta} = (0.5, 0.5, 0.5, 0.5)^T$. We assumed that 8 elevation sensors were employed in the network and that they were placed, equally spaced along the perimeter, on a circle of radius equal to $100m$ and centered at $(0, -100)^T$ such that one of the sensors is located at the origin. Further, we assumed that the sensors will collect their measurements over an interval of $60s$ starting at time $t = 0$ and that the variance of the additive noise in the sensors' measurements is equal to $\sigma^2 = (0.25)^2$ for all the sensors. Under the just stated conditions and assumptions, we employed our estimation method to produce estimates of the unknown variable $\boldsymbol{\theta}$ as we varied the sensors' sampling frequency, and hence the total number of samples, in increments of $60Hz$, from $60Hz$ to $300Hz$ and then increasing the sampling frequency to $500Hz$ and afterwards to $1000Hz$. For each sampling frequency, we ran a Monte Carlo simulation of 10000 runs where we generated a new realization of the noise for each run. In Figure 3.1, we include a plot of the obtained MSE for estimating each element of $\boldsymbol{\theta}$ and compare it to the CRB for estimating that same element of $\boldsymbol{\theta}$. The CRB is obtained according to the expressions and equations derived and presented in Section 3.2. The different curves were multiplied by different scaling factors, given in the legend of Figure

3.1, in order to include all the curves in a single figure. From Figure 3.1, we can see that the MSE obtained using our estimation approach seems to come extremely close to the CRB[3].

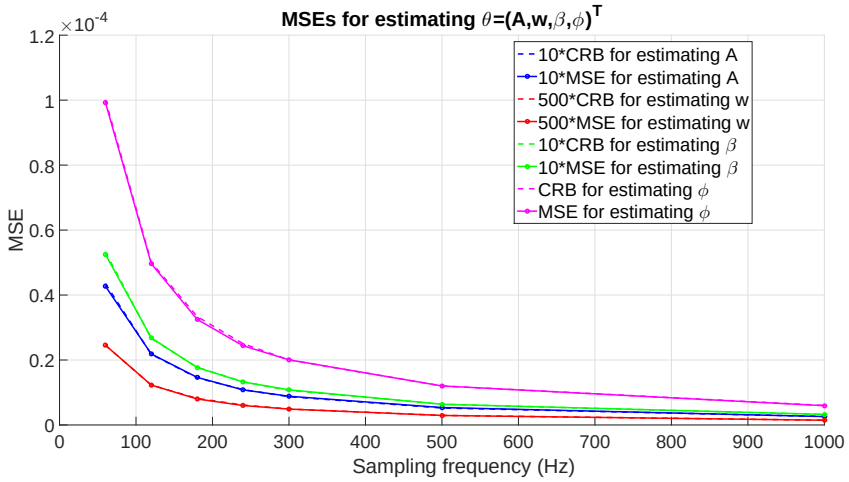


Figure 3.1: CRB and MSE curves for estimating $\theta = (0.5, 0.5, 0.5, 0.5)^T$

3.4.2 Estimation Performance: Three Component Waves

For this subsection, we consider the case where the ocean environment is described by 3 different component waves. This means that we have $L = 3$ in (3.20) and that our unknown variable is given by

$$\boldsymbol{\theta} = (A_1, A_2, A_3, w_1, w_2, w_3, \beta_1, \beta_2, \beta_3, \phi_1, \phi_2, \phi_3)^T. \tag{3.35}$$

Using the same sensor types, sensor layouts, and the same noise variance as those assumed in the previous subsection and only changing the length of the interval over which the measurements are collected to $30s$, starting at $t = 0$, we generated 100 different random ocean environments with $L = 3$ where the amplitudes were chosen uniformly randomly over a range from $0.5 - 2m$ and the directions and phases were each chosen uniformly randomly over the full range of $0 - 2\pi$. Since the frequencies must be sufficiently far apart, w_1 was chosen uniformly randomly from $0.5 - 1rad/s$, w_2 was chosen uniformly randomly from $2 - 2.5rad/s$, and w_3 was chosen uniformly randomly from $3.5 - 4rad/s$. To obtain our numerical results, we employed our presented estimation method under sampling frequencies that were increased in increments of $200Hz$ from $100Hz$ to $1100HZ$. For each

[3]While the provided results are for a single possible value of the unknown parameter $\boldsymbol{\theta}$, we attempted a very large number of different and random values for $\boldsymbol{\theta}$ and in every attempted case, the obtained results were essentially identical to the ones presented here.

sampling frequency and for each of the 100 ocean environments, we ran a Monte Carlo simulation of 300 runs where we generated a new realization of the noise for each run. Afterwards, for every sampling frequency and every environment, the MSE for each component of the unknown variable $\boldsymbol{\theta}$, given in (3.35), was obtained and compared to the CRB, calculated according to the expressions derived in Section 3.2. Then, for each of the considered ocean environments, we obtained curves of MSE and CRB similar to those presented in Figure 3.1. Since the number of such curves is very large, and although the curves were generally very similar to each other and to those in Figure 3.1, we decided to rank the curves according to the sum of the percentage difference between the MSE and CRB values at each of the considered sampling frequencies and present some of the worst scoring curves for the different types of parameters in $\boldsymbol{\theta}$. In Figure 3.2 we present the worst scoring MSE curves, on a logarithmic scale, for estimating A_1, w_2, β_3, and ϕ_1 over the 100 considered ocean environments along with the CRB associated with estimating that parameter for the ocean environment returning the worst score for that parameter.

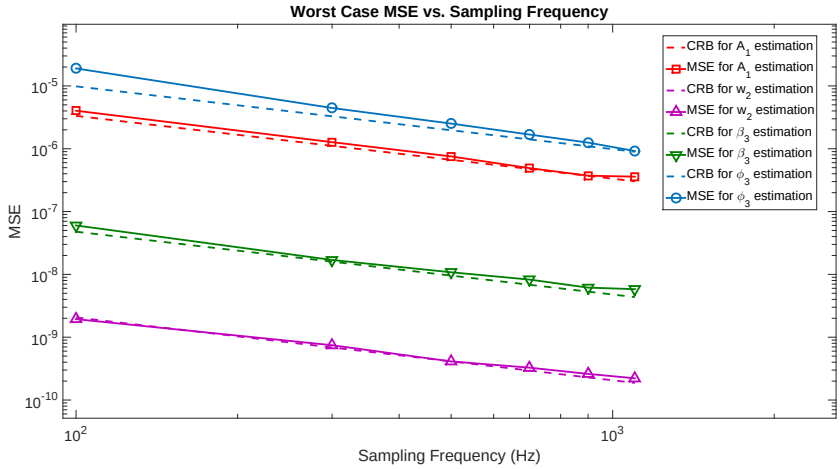


Figure 3.2: Worst scoring MSE curves according to the sum of percentage difference between MSE curve and associated CRB curve, which is also included

From Figure 3.2, we can see that even the worst scoring MSE curves came very close to the CRB which highlights the great accuracy we were continually able to achieve with our presented estimation approach not just under regular wave environments but also under complex, irregular wave environments. It is important to note that the presented results are numerical and while we have not been able to generate cases where the presented estimation approach fails to produce accurate estimates, this does not provide a mathematical or analytical guarantee regarding the performance of the estimation approach over the entire space of the unknown variables.

In the following subsection, we consider an even more complicated wave environment containing 6 component waves and where the noise in the measurements is correlated.

3.4.3 Estimation Performance: Six Component Waves and Correlated Noise

To further demonstrate the accuracy of the presented approach, in this subsection we consider an ocean environment that could be regarded as more realistic under the presence of wind waves and other real-world conditions. The considered environment was chosen to have $L = 6$ component waves. Since ocean environments are sometimes decomposed into wave systems where waves from any given system all travel in the same direction, the first three component waves were chosen to all have the same wave direction of $\pi/4$ and the last three components were chosen to all have the same wave direction of $2\pi/3$. The amplitudes were all chosen uniformly randomly over $0.5 - 1.5m$ and the phases were all chosen uniformly randomly over the full range of $0 - 2\pi$. Further, we assumed the first three components waves had frequencies of $w_1 = 0.25rads/s, w_2 = 0.5rads/s$, and $w_3 = 0.75rads/s$ and we assumed the last three component waves had $w_4 = 1.25rads/s, w_5 = 1.5rads/s$, and $w_6 = 1.75rads/s$ as their frequencies. It is also assumed that we employ the same sensor types, number, and layouts outlined earlier and we consider the case where the sensor measurements are collected over an interval of $60s$ starting at time $t = 0$ and at a sampling frequency of $200Hz$.

As for the noise, we compared the performance of our algorithm under both correlated and uncorrelated noise. For the uncorrelated case, we assumed the noise is Gaussian with zero mean and a variance of $\sigma^2 = 0.2m^2$. For the correlated case we assumed that, at each time interval, the noise in the measurements collected by the eight sensors in the system is jointly Gaussian with zero mean and a covariance matrix given by

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma^4 & \sigma^6 & \sigma^8 & \sigma^{10} & \sigma^8 & \sigma^6 & \sigma^4 \\ \sigma^4 & \sigma^2 & \sigma^4 & \sigma^6 & \sigma^8 & \sigma^{10} & \sigma^8 & \sigma^6 \\ \sigma^6 & \sigma^4 & \sigma^2 & \sigma^4 & \sigma^6 & \sigma^8 & \sigma^{10} & \sigma^8 \\ \sigma^8 & \sigma^6 & \sigma^4 & \sigma^2 & \sigma^4 & \sigma^6 & \sigma^8 & \sigma^{10} \\ \sigma^{10} & \sigma^8 & \sigma^6 & \sigma^4 & \sigma^2 & \sigma^4 & \sigma^6 & \sigma^8 \\ \sigma^8 & \sigma^{10} & \sigma^8 & \sigma^6 & \sigma^4 & \sigma^2 & \sigma^4 & \sigma^6 \\ \sigma^6 & \sigma^8 & \sigma^{10} & \sigma^8 & \sigma^6 & \sigma^4 & \sigma^2 & \sigma^4 \\ \sigma^4 & \sigma^6 & \sigma^8 & \sigma^{10} & \sigma^8 & \sigma^6 & \sigma^4 & \sigma^2 \end{bmatrix}, \tag{3.36}$$

where we assume that the correlation decreases as the distance between the sensors increases which is a typical method of modeling correlation between sensors in a network. Similar to the uncorrelated case, we assumed that $\sigma^2 = 0.2m^2$ in this case as well.

Table 3.2 summarizes the results obtained under the just described conditions and assumptions. In Table 3.2, we provide the values for each of the 24 variables describing the assumed environment with six component waves, the CRB value associated with estimating each of the 24 variables assuming the noise is uncorrelated, and the MSE values obtained using our presented estimation approach both with and without correlation in the noise. The MSE values were obtained by averaging the results over 1000 Monte Carlo runs where each run generated a new realization of the correlated and uncorrelated noise.

Table 3.2: Comparing MSE values obtained using the presented algorithm with correlated and uncorrelated noise

Unkown	**True Value**	**CRB**	**MSE with uncorr. noise**	**MSE with corr. noise**
A_1	1.282	4.181e-06	3.996e-06	5.167e-06
A_2	1.412	4.179e-06	4.186e-06	4.413e-06
A_3	0.940	4.171e-06	4.174e-06	3.970e-06
A_4	0.960	4.171e-06	4.220e-06	4.382e-06
A_5	0.787	4.176e-06	4.250e-06	4.597e-06
A_6	1.055	4.164e-06	4.396e-06	4.219e-06
w_1	0.250	8.162e-09	8.076e-09	1.186e-08
w_2	0.500	6.020e-09	6.177e-09	6.330e-09
w_3	0.750	1.160e-08	1.191e-08	1.089e-08
w_4	1.250	7.314e-09	7.448e-09	7.417e-09
w_5	1.500	8.795e-09	8.893e-09	8.953e-09
w_6	1.750	4.048e-09	4.128e-09	4.248e-09
β_1	0.785	1.249e-05	1.260e-05	1.561e-05
β_2	0.785	6.440e-07	6.826e-07	6.835e-07
β_3	0.785	2.893e-07	2.818e-07	2.835e-07
β_4	2.094	3.594e-08	3.616e-08	3.691e-08
β_5	2.094	2.578e-08	2.578e-08	2.748e-08
β_6	2.094	7.751e-09	7.714e-09	8.197e-09
ϕ_1	5.598	1.367e-05	1.372e-05	1.941e-05
ϕ_2	3.605	1.259e-05	1.308e-05	1.306e-05
ϕ_3	4.605	2.882e-05	2.909e-05	2.693e-05
ϕ_4	1.654	1.360e-05	1.375e-05	1.500e-05
ϕ_5	4.191	1.896e-05	1.806e-05	1.907e-05
ϕ_6	5.940	1.004e-05	1.005e-05	1.113e-05

As seen from Table 3.2, the presented approach seems to work very well relative to the CRB even under correlated noise. Further, the results from Table 3.2 seem to support results from the previous subsection suggesting that the presented approach works well under complicated wave

environments described by multiple component waves. We also note that a noise variance equal to $\sigma^2 = 0.2m^2$ is considered a relatively very large value especially that the largest amplitude among all component waves was $A_1 = 1.282m$ so our results seem to suggest that the presented approach is able to perform at near optimal accuracy even under very large, and possibly correlated, noise values.

Next, we present numerical results involving the runtime of our presented estimation approach.

3.4.4 Runtime Performance

In this subsection, we compare the runtime of the presented estimation approach to that of a Genetic Algorithm (GA) and a Particle Swarm Optimizer (PSO) which are well-known global optimization techniques that, similar to the presented approach, do not require an initial guess for the unknown variable $\boldsymbol{\theta}$. For the presented results, we used a population size of 100 for both GA and PSO. Included in Figure 3.3 are curves for the average runtime of the three different estimation methods obtained while assuming the ocean is described by a single component wave ($L = 1$). The same sensor types, sensor layouts, and noise variance assumed in Figure 3.1 and Figure 3.2 are employed and it also assumed that the sensors collect their measurements over an interval of $60s$ starting at time $t = 0$. In Figure 3.3, at each of the considered sampling frequencies, 100 different values of $\boldsymbol{\theta} = (A, w, \beta, \phi)^T$ were generated uniformly randomly such that the amplitude is between $0.5 - 1.5m$, the frequency is between $0.2 - 0.7rads/s$, and both the β and ϕ where between $0 - 2\pi$. Due to the complexity of the estimation problem, and although the selected population size for both GA and PSO is relatively large, our best recorded percentages for having the estimates produced by the algorithms fall within $\pm 2.5\%$ of their correct values over all the considered sampling frequencies were 94% for GA and 95% for PSO while the estimation approach presented here produced estimates that fell within the $\pm 2.5\%$ threshold 100% of the time. As a result, the running times for both GA and PSO in Figure 3.3 are averaged over only the cases where the algorithms were able to produce estimates that fell within the $\pm 2.5\%$ range around the the correct value. However, if we did include the cases where the algorithms produced poor results (outside of the 2.5% threshold), the results would have been almost identical as the times were almost the same regardless of whether the algorithms converged to estimates inside or outside the 2.5% threshold. Also, we note that in all the considered cases, both GA and PSO terminated due to convergence to a local minimum and not due to reaching the maximum number of iterations or the maximum allowable running time so the cases where they produced poor estimates most likely

resulted from converging to poor local minimums. From Figure 3.3, we can see that the presented estimation approach clearly outperforms both GA and PSO in terms of runtime performance as the number of observations is increased by increasing the sampling frequency.

We attempted to obtain a similar figure for cases where the ocean surface is described by multiple component waves ($L \neq 1$), but we were unable to obtain any results where either GA or PSO were able to produce estimates within the $\pm 2.5\%$ range for at least 90% of the cases while maintaining a runtime of less than 120s. In comparison, the runtime of our presented approach was consistently less than 20s over all attempted cases and it was again able to produce the correct estimate values 100% of the time.

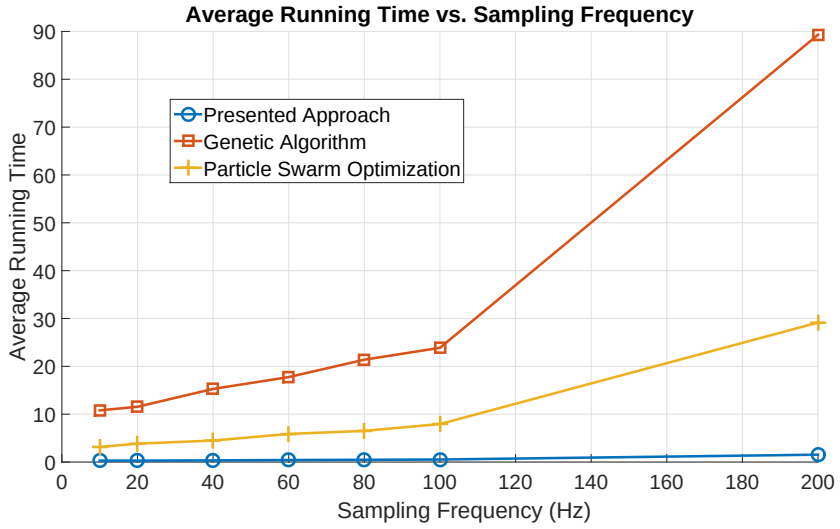


Figure 3.3: Average runtime of presented method in comparison with GA and PSO.

In the next section, we consider the impact of estimation errors on the amount of loss in the power absorbed by WEC devices.

3.5 Power Loss Due to Estimation Errors

Throughout this section, we assume that the ocean is described by a single component wave (L=1). We also focus on measurements of wave elevation which are given by setting the values of 'a' and 'b' according to the first row of Table 3.1 such that $\Phi(x, y, t)$ is given by

$$\Phi(x, y, t) = A \cos\left((\frac{w^2}{g}) x \cos(\beta) + (\frac{w^2}{g}) y \sin(\beta) - tw + \phi \right), \tag{3.37}$$

since the provided analysis can be easily applied to the remaining measurements given in Table 3.1.

From the work by Evans [94] and Falnes [95], we know that the mean (mechanical) power absorbed by N devices under monochromatic, plane incident waves is given by

$$P = \frac{1}{4}(\mathbf{U}^H\mathbf{F} + \mathbf{F}^H\mathbf{U}) - \frac{1}{2}\mathbf{U}^H\mathbf{B}\mathbf{U}, \tag{3.38}$$

where $\mathbf{U}$ is the column vector of complex velocity amplitudes, $\mathbf{F}$ is the column vector of exciting forces, $\mathbf{B}$ is the matrix of (real) damping coefficients, and $(\cdot)^H$ denotes the Hermitian transpose. The expression for power given by (3.38) may be maximized such that the maximum mean absorbed power is

$$P_{max} = \frac{1}{8}\mathbf{F}^H\mathbf{B}^{-1}\mathbf{F}, \tag{3.39}$$

which is achieved when the vector of complex velocity amplitudes, $\mathbf{U}$, is set to

$$\mathbf{U}_{optimal} = \frac{1}{2}\mathbf{B}^{-1}\mathbf{F}, \tag{3.40}$$

where we assume that $\mathbf{B}^{-1}$ exists.

From the literature on hydrodynamics [24], the wave elevation is related to the excitation force acting on device $r = 1, 2, \ldots N$ located at $(x_r, y_r)^T$ through

$$f_{ext}(x_r, y_r, t) = \int_{-\infty}^{\infty} h_{ext}(t-\tau)\Phi(x_r, y_r, t)d\tau, \tag{3.41}$$

where f_{ext} is the excitation force in Newtons, $h_{ext}(t)$ is the excitation impulse response function, and $\Phi(x, y, t)$ is the wave elevation in meters. Assuming that $\Phi(x, y, t)$ is given by (3.37), the excitation force from monochromatic, plane incident waves can be expressed as

$$f_{ext}(x_r, y_r, t) = A|H_{ext}(w)| \cos\left((\frac{w^2}{g})x\cos(\beta) + (\frac{w^2}{g})y\sin(\beta) - tw + \phi + \angle H_{ext}(w)\right), \tag{3.42}$$

where $H_{ext}(w) = \mathcal{F}\{h_{ext}(t)\}$ is the excitation force transfer function and $\mathcal{F}\{\cdot\}$ denotes the Fourier

transform. Thus, the excitation force has a complex (phasor) given by

$$F(x_r, y_r) = A|H_{ext}(w)| \exp(j\gamma), \tag{3.43}$$

where $\gamma = (\frac{w^2}{g})\binom{x_r}{y_r} \cdot \binom{\cos(\beta)}{\sin(\beta)} + \phi + \angle H_{ext}(w)$. To be able to use the power equation given in (3.38), we will assume that the value of w is known exactly. This will cause our errors to be slightly underestimated. However, in practice, we have found that one can obtain an estimate of w that is much more accurate than the other parameters so this is not an unreasonable assumption. As a result, our task of estimating the complex amplitude of the excitation force in order to set our control such that the velocity of the WEC is set to satisfy (3.40) becomes the task of estimating the value of the $\boldsymbol{\theta} = (A, \beta, \phi)^T$ using an estimator $\hat{\boldsymbol{\theta}}$ based on collected measurements. Note that (3.40) describes a control law that is optimal when the complex excitation force is known exactly so due to errors from estimating the excitation force, the resulting control will actually be sub-optimal and cause a drop in the average absorbed power.

Since we assume that the collected measurements contain a random component due to noise, the values of $\hat{\boldsymbol{\theta}}$ obtained from different sets of measurements collected under the same wave conditions will be slightly different from set to set because the realized values of the noise will not be exactly the same from one set to the other. Thus, it is possible to have a set of measurements that results in a very accurate value for $\hat{\boldsymbol{\theta}}$ which causes only a small loss in the power but it also possible that we collect a set of measurements that happens to be corrupted by noise such that the value of $\hat{\boldsymbol{\theta}}$ is not as accurate, causing a larger loss in the absorbed power. In order to handle this variability in the amount of power lost, we use the statistical description of $\hat{\boldsymbol{\theta}}$ in order to calculate the expected value for the loss which provides us with an indication of the amount of power lost on average. Given a certain realization (value) for $\hat{\boldsymbol{\theta}}$, the predicted value of the complex amplitude of the excitation force at device $r = 1, 2, \ldots, N$ will be

$$\hat{F}(x_r, y_r) = \hat{A}|H_{ext}(w)| \exp(j\hat{\gamma}), \tag{3.44}$$

where $\hat{\gamma} = (\frac{w^2}{g})\binom{x_r}{y_r} \cdot \binom{\cos(\hat{\beta})}{\sin(\hat{\beta})} + \hat{\phi} + \angle H_{ext}(w)$, and we can use this value to set the velocity amplitude according to (3.40) as

$$\hat{\mathbf{U}} = \frac{1}{2}\mathbf{B}^{-1}\hat{\mathbf{F}}, \tag{3.45}$$

where $\hat{\mathbf{F}} = (\hat{F}(x_1, y_1), \ldots, \hat{F}(x_N, y_N))^T$. As a result, the average power absorbed given an estimate $\hat{\boldsymbol{\theta}}$, according to (3.38), will be given by

$$\begin{aligned} P_{est}(\hat{\theta}) &= \frac{1}{4}(\hat{\mathbf{U}}^H\mathbf{F} + \mathbf{F}^H\hat{\mathbf{U}}) - \frac{1}{2}\hat{\mathbf{U}}^H\mathbf{B}\hat{\mathbf{U}}, \\ &= \frac{1}{8}((\mathbf{B}^{-1}\hat{\mathbf{F}})^H\mathbf{F} + \mathbf{F}^H(\mathbf{B}^{-1}\hat{\mathbf{F}})) - \frac{1}{8}(\mathbf{B}^{-1}\hat{\mathbf{F}})^H\mathbf{B}(\mathbf{B}^{-1}\hat{\mathbf{F}}), \\ &= \frac{1}{8}(\hat{\mathbf{F}}^H\mathbf{B}^{-1}\mathbf{F} + \mathbf{F}^H\mathbf{B}^{-1}\hat{\mathbf{F}}) - \frac{1}{8}\hat{\mathbf{F}}^H\mathbf{B}^{-1}\hat{\mathbf{F}}, \end{aligned} \tag{3.46}$$

and this is a random quantity depending on the value of $\hat{\boldsymbol{\theta}}$ which is a random variable due to the noise added to the sensor measurements used for the estimation. If the probability distribution for $\hat{\boldsymbol{\theta}}$ is given, we can proceed to calculate the expected value of $P_{est}(\hat{\theta})$ under the given distribution. If not, we will assume that the distribution for $\hat{\boldsymbol{\theta}}$ is a multivariate normal with mean $\boldsymbol{\theta}$ and covariance matrix equal to the Cramer Rao Bound matrix (inverse of the FIM) as justified in [92, 96]. This assumes the performance of the estimator is optimal and should therefore provide a lower bound on the loss in power. The FIM for this problem was derived in Section 3.2. After finding the expected value for $P_{est}(\hat{\boldsymbol{\theta}})$, we can divide that number by P_{max} which is given by (3.39) to determine the percentage of the maximum power that we expect to absorb under losses due errors in estimation.

In the next section, we present a numerical example where the just presented ideas and equations are employed.

3.5.1 Numerical Example: Power loss for a single device in isolation

Consider the case where we have a single device in isolation. Assuming the excitation force acting on the device is given by (3.43), the power absorbed by N devices given the estimate $\hat{\boldsymbol{\theta}}$ is given by (3.46) and, for a single device, it simplifies to

$$P_{est}(\hat{\theta}) = \frac{\hat{A}A}{4B}\cos(\gamma - \hat{\gamma}) - \frac{(\hat{A})^2}{8B}. \tag{3.47}$$

For this numerical example, we assume that the WEC is a heaving vertical cylinder having a radius of $5m$ and a length of $20m$. Also, we assume that the true value of $\theta = (A, B, \phi)^T$ is given by $\theta = (1, 1, 1)^T$ and that w is known to be equal to unity. The values for the damping coefficient B and the excitation force transfer function are obtained using WAMIT® [97] assuming the device is placed in an ocean of infinite depth. Further, we assume that the WEC is surrounded by 4 wave elevation sensors that have been placed, along the major axes, on a circle of radius $100m$. The

measurements collected by the sensors are assumed to be described by

$$\begin{aligned} \phi(t_m, x_l, y_l) &= \Phi(t_m, x_l, y_l) + v_l(t_m), \\ t_m &= T_s, \ldots, KT_s, \ l = 1, \ldots, 4 \end{aligned} \tag{3.48}$$

where T_s is the sampling period, K is the total number of time samples, $v_l(t_m)$ is noise, (x_l, y_l) is the location of the l^{th} sensor, and Φ is given in (3.37). For simplicity, we assume

$$(v_1(T_s), \ldots, v_N(T_s), \ldots, v_1(KT_s), \ldots, v_N(KT_s))^T$$

is jointly Gaussian with zero mean vector and covariance matrix which is diagonal with $(\sigma_1^2, \ldots, \sigma_{NK}^2)$ along the diagonal and $\sigma^2 = \sigma_1^2 = \ldots = \sigma_{NK}^2 = 1$. The distribution of the employed estimator $\hat{\theta}$ is taken to be a multivariate normal distribution with mean equal to θ and a covariance matrix equal to the CRB matrix (inverse of the FIM). To quantify the power loss, we calculate the percentage of the power absorbed while accounting for estimation errors relative to the maximum power as given in (3.39) as

$$P_{rel} = \frac{E(P_{est}(\hat{\theta}))}{P_{max}} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{est}(\hat{\theta}) f(\hat{\theta}) d\hat{A} d\hat{\beta} d\hat{\phi}}{P_{max}}, \tag{3.49}$$

where $f(\hat{\theta})$ is the probability density function of $\hat{\theta}$. Figure 3.4 includes a plot of the value for P_{rel} as we varied the total sampling time KT_s for different sampling frequencies, where the values of the P_{rel} we obtained through numerically approximating the triple integral in (3.49). Figure 3.4 shows that increasing both quantities decreases the power loss due to errors in estimation. Figure 3.4 also shows that increasing the sampling frequency and the sampling time will eventually provide diminishing returns with respect to the gain in absorbed power, which should be expected.

3.6 Conclusion

In this chapter, we investigated the parametric estimation of spatial and temporal ocean waveforms using a network of spatially distributed ocean sensors whose measurements are corrupted by additive white Gaussian noise. We derived and presented, for the first time, the CRB associated with the estimation of wave properties critically important to novel control strategies. The presented CRB expressions are general in that they allow for employing any combination of types of

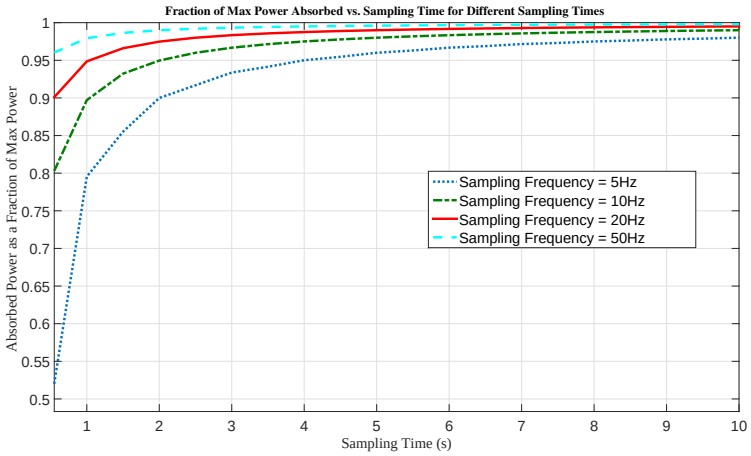


Figure 3.4: Expected power loss vs sampling time for different sampling frequency

measurements from a given set of possible ocean measurements. We also presented a suboptimal estimation approach that works to overcome possible problems associated with the non-convexity of the estimation problem and the existence of many local extrema in the objective function. We provided numerical results that demonstrate the near optimal accuracy of the presented approach under different ocean environments and also its superior runtime performance in comparison to that of reasonable possible alternative approaches. Finally, we introduced a method that quantifies the loss in the average absorbed power due to errors in estimating the elevation of the waves that are incident on a single WEC in isolation or multiple WECs in a farm. The devices are assumed to be controlled such that their velocity would be optimal had the estimates been exactly correct. A lower bound on the expected loss in the average absorbed power was be obtained by relying on the CRB of the estimates and a numerical example considering the case where a heaving cylindrical WEC is controlled optimally based on measurements collected from four sensors surrounding the device was presented. Results giving the percentage of the expected absorbed power under estimation errors relative to the maximum possible absorbed power by the device were obtained for different values of total sampling time and sampling frequency. The results clearly showed that the loss in power could be very significant if the values of the total sampling time or the sampling frequency were chosen to be too small while also showing that there are diminishing returns when increasing either variable in order to recover the otherwise lost power.

3.7 Appendix

A Method for Reducing the Double Sum to a Single Sum

Suppose we adopt

$$\psi(x,y,t;\boldsymbol{\theta}) = \sum_{i=1}^{L}\sum_{j=1}^{M_i} A_{i,j} w_i^a \cos\left((\frac{w_i^2}{g})x\cos(\beta_{i,j}) + (\frac{w_i^2}{g})y\sin(\beta_{i,j}) - tw_i + \phi_{i,j} - b\frac{\pi}{2}\right), \quad (3.50)$$

as our wave model. Then, the unknown parameters would be given by

$$\begin{aligned}\boldsymbol{\theta} =& (A_{1,1}, A_{1,2}, \ldots, A_{1,M_1}, A_{2,1}, \ldots, A_{L,M_L},\\ & \omega_1, \ldots, \omega_L,\\ & \beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,M_1}, \beta_{2,1}, \ldots, \beta_{L,M_L},\\ & \phi_{1,1}, \phi_{1,2}, \ldots, \phi_{1,M_1}, \phi_{2,1}, \ldots, \phi_{L,M_L})^T. \end{aligned} \quad (3.51)$$

Now, for each j in (3.50), slightly perturb the value of w_i to $w_{i,j} \triangleq w_i + \delta_j$ where $\delta_j \neq \delta_{j'}$ for $j \neq j'$ and let $\Phi(x,y,t;\boldsymbol{\theta}_s)$ denote the resulting wave model after the change in the values of w_i to $w_{i,j}$. Then, we have that

$$\begin{aligned}\Phi(x,y,t;\boldsymbol{\theta}_s) = \sum_{i=1}^{L}\sum_{j=1}^{M_i} A_{i,j}(w_{i,j})^a \cos\Bigg((\frac{w_{i,j}^2}{g})x\cos(\beta_{i,j})\\ + (\frac{w_{i,j}^2}{g})y\sin(\beta_{i,j}) - w_{i,j}t + \phi_{i,j} - b\frac{\pi}{2}\Bigg),\end{aligned} \quad (3.52)$$

which now has all the parameters inside the double sum indexed over both i and j, and thus can be written as the single sum

$$\Phi(x,y,t;\boldsymbol{\theta}_s) = \sum_{k=1}^{M} A_k w_k^a \cos\left((\frac{w_k^2}{g})x\cos(\beta_k) + (\frac{w_i^k}{g})y\sin(\beta_k) - tw_k + \phi_k - b\frac{\pi}{2}\right), \quad (3.53)$$

where $M = \sum_{i=1}^{L} M_i$ and A_k, β_k, w_k, and ϕ_k are obtained by relabeling the $A_{i,j}$ values, according to

$$\begin{array}{ccc} \text{New Label} & \leftarrow & \text{Old Label} \\ A_k & \leftarrow & A_{i,j} \end{array}$$

$$\begin{array}{ccc} A_1 & \leftarrow & A_{1,1} \\ A_2 & \leftarrow & A_{1,2} \\ \vdots & & \vdots \\ A_{M_1} & \leftarrow & A_{1,M_1}, \\ A_{M_1+1} & \leftarrow & A_{2,1}, \\ A_{M_1+2} & \leftarrow & A_{2,2}, \\ \vdots & & \vdots \\ A_{M_1+M_2} & \leftarrow & A_{2,M_2} \\ A_{M_1+M_2+1} & \leftarrow & A_{3,1} \\ \vdots & & \vdots \\ A_M & \leftarrow & A_{L,M_L} \end{array} , \tag{3.54}$$

and then following the same exact scheme to relabel $w_k \leftarrow w_{i,j}$, $\beta_k \leftarrow \beta_{i,j}$, $\phi_k \leftarrow \phi_{i,j}$. Note that the vector of unknown parameters is now given by

$$\begin{aligned} \boldsymbol{\theta}_s =& (A_1, A_2, \ldots, A_M, \\ & \omega_1, \omega_2, \ldots, w_M \\ & \beta_1, \beta_2, \ldots, \beta_M, \\ & \phi_1, \phi_2, \ldots, \phi_M)^T, \end{aligned} \tag{3.55}$$

rather than (3.51).

As an example, suppose we start with the (double sum) wave model given by (3.52) with $L = 2$,

$M_1 = 2, M_2 = 3$, and

$$\begin{aligned}\boldsymbol{\theta} =&(A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}, A_{2,3},\\ &\omega_{1,1}, \omega_{1,2}, \omega_{2,1}, \omega_{2,2}, \omega_{2,3}\\ &\beta_{1,1}, \beta_{1,2}, \beta_{2,1}, \beta_{2,2}, \beta_{2,3},\\ &\phi_{1,1}, \phi_{1,2}, \phi_{2,1}, \phi_{2,2}, \phi_{2,3})^T. \end{aligned} \tag{3.56}$$

Then, we have that

$$\begin{aligned}&\Phi(x, y, t; \boldsymbol{\theta}_s) = \sum_{i=1}^{L}\sum_{j=1}^{M_i}(w_{i,j})^a A_{i,j}\cos\left((\frac{w_{i,j}^2}{g})x\cos(\beta_{i,j}) + (\frac{w_{i,j}^2}{g})y\sin(\beta_{i,j}) - w_{i,j}t + \phi_{i,j} - b\frac{\pi}{2}\right)\\
&= A_{1,1}(w_{1,1})^a\cos\left((\frac{w_{1,1}^2}{g})x\cos(\beta_{1,1}) + (\frac{w_{1,1}^2}{g})y\sin(\beta_{1,1}) - w_{1,1}t + \phi_{1,1} - b\frac{\pi}{2}\right)\\
&+ A_{1,2}(w_{1,2})^a\cos\left((\frac{w_{1,2}^2}{g})x\cos(\beta_{1,2}) + (\frac{w_{1,2}^2}{g})y\sin(\beta_{1,2}) - w_{1,2}t + \phi_{1,2} - b\frac{\pi}{2}\right)\\
&+ A_{2,1}(w_{2,1})^a\cos\left((\frac{w_{2,1}^2}{g})x\cos(\beta_{2,1}) + (\frac{w_{2,1}^2}{g})y\sin(\beta_{2,1}) - w_{2,1}t + \phi_{2,1} - b\frac{\pi}{2}\right)\\
&+ A_{2,2}(w_{2,2})^a\cos\left((\frac{w_{2,2}^2}{g})x\cos(\beta_{2,2}) + (\frac{w_{2,2}^2}{g})y\sin(\beta_{2,2}) - w_{2,2}t + \phi_{2,2} - b\frac{\pi}{2}\right)\\
&+ A_{2,3}(w_{2,3})^a\cos\left((\frac{w_{2,3}^2}{g})x\cos(\beta_{2,3}) + (\frac{w_{2,3}^2}{g})y\sin(\beta_{2,3}) - w_{2,3}t + \phi_{2,3} - b\frac{\pi}{2}\right),\end{aligned} \tag{3.57}$$

which by applying the relabeling scheme given by (3.54) is equal to

$$\begin{aligned}&\Phi(x, y, t; \boldsymbol{\theta}_s) = A_1(w_1)^a\cos\left((\frac{w_1^2}{g})x\cos(\beta_1) + (\frac{w_1^2}{g})y\sin(\beta_1) - w_1t + \phi_1 - b\frac{\pi}{2}\right)\\
&+ A_2(w_2)^a\cos\left((\frac{w_2^2}{g})x\cos(\beta_2) + (\frac{w_2^2}{g})y\sin(\beta_2) - w_2t + \phi_2 - b\frac{\pi}{2}\right)\\
&+ A_3(w_3)^a\cos\left((\frac{w_3^2}{g})x\cos(\beta_3) + (\frac{w_3^2}{g})y\sin(\beta_3) - w_3t + \phi_3 - b\frac{\pi}{2}\right)\\
&+ A_4(w_4)^a\cos\left((\frac{w_4^2}{g})x\cos(\beta_4) + (\frac{w_4^2}{g})y\sin(\beta_4) - w_4t + \phi_4 - b\frac{\pi}{2}\right)\\
&+ A_5(w_5)^a\cos\left((\frac{w_5^2}{g})x\cos(\beta_5) + (\frac{w_5^2}{g})y\sin(\beta_5) - w_5t + \phi_5 - b\frac{\pi}{2}\right),\end{aligned} \tag{3.58}$$

which can be expressed as the single sum

$$\Phi(x, y, t; \boldsymbol{\theta}_s) = \sum_{k=1}^{5} A_k w_k^a\cos\left((\frac{w_k^2}{g})x\cos(\beta_k) + (\frac{w_i^k}{g})y\sin(\beta_k) - tw_k + \phi_k - b\frac{\pi}{2}\right). \tag{3.59}$$

Since we slightly perturbed the frequency values from (3.50) in order to arrive at the single

sum model (3.53), it remains to show that the error introduced by our perturbation could be controlled such that it is always within any given level of tolerable modeling error. To do so, we now show that there will always exist a choice for the values of δ_j such that the difference between $\Phi(x, y, t; \boldsymbol{\theta}_s)$ and $\psi(x, y, t; \boldsymbol{\theta})$ is made sufficiently small.

For simplicity, suppose we set $\delta_j = \delta(j-1)$ where $\delta > 0$. Then, we have that

$$\begin{aligned}
&\Phi(x, y, t; \boldsymbol{\theta}_s) = \sum_{i=1}^{L} \sum_{j=1}^{M_i} (w_i + \delta_j)^a A_{i,j} \\
&\cos\left((\frac{w_i^2}{g}) x \cos(\beta_{i,j}) + (\frac{w_i^2}{g}) y \sin(\beta_{i,j}) - w_i t + \phi_{i,j} - b\frac{\pi}{2} \right. \\
&\left. + (\frac{(\delta_j)^2}{g}) x \cos(\beta_{i,j}) + (\frac{(\delta_j)^2}{g}) y \sin(\beta_{i,j}) - \delta_j t \right) \\
&= \sum_{i=1}^{L} \sum_{j=1}^{M_i} (w_i + \delta(j-1))^a A_{i,j} \\
&\cos\left((\frac{w_i^2}{g}) x \cos(\beta_{i,j}) + (\frac{w_i^2}{g}) y \sin(\beta_{i,j}) - w_i t + \phi_{i,j} - b\frac{\pi}{2} + \right. \\
&\left. \frac{\delta^2 (j-1)^2}{g} x \cos(\beta_{i,j}) + \frac{\delta^2 (j-1)^2}{g} y \sin(\beta_{i,j}) - \delta(j-1) t \right).
\end{aligned} \tag{3.60}$$

Now, if we consider the limit of $\Phi(x, y, t; \boldsymbol{\theta}_s)$ as δ goes to zero, we get that

$$\begin{aligned}
&\lim_{\delta \to 0} \Phi(x, y, t; \boldsymbol{\theta}_s) = \sum_{i=1}^{L} \sum_{j=1}^{M_i} (w_i + 0(j-1))^a A_{i,j} \\
&\cos\left((\frac{w_i^2}{g}) x \cos(\beta_{i,j}) + (\frac{w_i^2}{g}) y \sin(\beta_{i,j}) - w_i t + \phi_{i,j} - b\frac{\pi}{2} + \right. \\
&\left. \frac{0^2 (j-1)^2}{g} x \cos(\beta_{i,j}) + \frac{0^2 (j-1)^2}{g} y \sin(\beta_{i,j}) - 0(j-1) t \right) \\
&= \sum_{i=1}^{L} \sum_{j=1}^{M_i} (w_i)^a A_{i,j} \cos\left((\frac{w_i^2}{g}) x \cos(\beta_{i,j}) + \right. \\
&\left. (\frac{w_i^2}{g}) y \sin(\beta_{i,j}) - w_i t + \phi_{i,j} - b\frac{\pi}{2} \right) \\
&= \psi(x, y, t; \boldsymbol{\theta}).
\end{aligned} \tag{3.61}$$

By the definition of limits, this means that given any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|\Phi(x, y, t; \boldsymbol{\theta}_s) - \psi(x, y, t; \boldsymbol{\theta})| < \epsilon, \tag{3.62}$$

which implies that, by choosing an appropriate value of δ, approximating $\psi(x, y, t; \boldsymbol{\theta})$ by $\Phi(x, y, t; \boldsymbol{\theta}_s)$ can be made as accurate (within a tolerable approximation error of $\epsilon > 0$) as the user desires.

Chapter 4

Stochastic Optimal Power Flow Under Forecast Errors and Failures in Communication

Nomenclature

$\mathcal{G}$	Set of all generators/loads
p_i	Power produced/demanded by generator/load $i \in \mathcal{G}$
C_i	Per unit cost of power produced/demanded
C_i^-, C_i^+	Per unit cost for negative/positive adjustment in dispatch of generator/load i
Λ_{kj}	$(k,j)^{th}$ entry of the PTDF matrix
t_k^{max}	Power flow limit for power line k
p_i^{min}, p_i^{max}	generation/consumption capacities for generator/load i
$p_{s,i}^{min}, p_{s,i}^{max}$	p_i^{min}, p_i^{max} under scenario s
$p_i^{no\_comm}$	Pre-set power generation/consumption under no communication for generator/load i
$\mathcal{S}$	Set of all possible random scenarios
π_s^{sc}	Probability that scenario $s \in \mathcal{S}$ will occur
$\Delta p_{s,i}$	Change in dispatch for generator/load i under scenario s
$\Gamma_{s,i}^-, \Gamma_{s,i}^+$	Under/over dispatch of generator/load i under scenario s
D_i^-, D_i^+	Per unit cost for under/over utilization of generator/load i under scenario s
$\mathcal{C}$	Set of all nodes in the communication network
$B_{i,j}$	Cost of leasing a comm. link between nodes i and j

Owing to the great size of most electric grids, data collected from sensors is typically transmitted to the grid operator using a communication network. The communication network is also responsible for sending control signals from the system operator to loads and generators in the power grid. In many cases, the control signals represent corrective re-dispatch commands based on the data collected from the sensors in the power grid. As a result, communication failures between the grid and the system operator could affect the ability to correct or control the dispatch level of a load or generator.

We consider the case where the system operator initially dispatches the loads and generators based on the forecasts for their capacities and then relies on sensors and the communication network to obtaining the actual load and generator capacities. Based on the measured capacities, the operator may decide to issue re-dispatch commands which must be carried to the communication network to the required load or generator. Facilitating the communication of the measured capacities and re-dispatch commands is a communication network whose links are susceptible to random failures.

We focus on identifying the optimal communication network topology and the optimal choice for dispatch and re-dispatch actions while accounting for forecasting errors and failures in communication. Our problem is formulated as a stochastic optimal power flow (OPF) problem where the objective is to minimize the average overall economic cost of the system under a large set of constraints representing the different stochastic scenarios we consider. While our study is on optimal power flow (OPF) problems, the underlying framework and analysis we employ could be easily extended to investigate the effect of failures and forecast errors on other important grid operations like unit commitment and economic dispatch for example. OPF problems sit at the core of many modern day power markets and operations with different variations of the problem solved several times a day at varying intervals that range from every hour to, in some cases, every five minutes. In general, OPF problems seek to minimize the cost of balancing power demand and generation while satisfying various physical and operational constraints relating to generation limits, electric line capacities and laws governing the flow of electricity. The generality of OPF allows the presented work to be extended to other market operations and problems as they will mostly involve the same or very similar constraints and considerations. Throughout this chapter, we employ the well known direct current (DC) approximation for power networks [98] to formulate our DC OPF problems that we simply refer to as OPF problems.

The rest of this chapter is organized as follows. In the next section, we describe our problem

setup and discuss the stochastic scenarios we consider. In section 4.2, we formulate our stochastic OPF problem and provide details on modeling communication failures and their effect on the system. Section 4.3 presents several numerical results obtained using our presented model. Finally, Section 4.4 summarizes the main findings presented in this chapter.

4.1 Model Setup and Description

We consider a stochastic, two-settlement, OPF problem for a power grid whose operation is supported by a communication network having a topology this chosen by the system operator as a first-settlement decision. In the first settlement, the system operator must rely on stochastic forecasts for the generation capacity of the system's renewable energy sources and the demand capacities of the system's loads which are all assumed to be controllable. Also included in the first settlement is the system operator's decision on the communication topology leased (activated) in anticipation of the different possible failures that some of the links might experience in between the two settlements. The stochastic events occurring between the two settlements are due to three factors:

- Errors in forecasting renewable generation capacity,
- Errors in forecasting load capacities (demands),
- Failures in the communication network.

The second settlement takes place after the actual capacity values are realized and the failure-prone communication network is employed to transmit the updated capacity values to the control center. The delivery of the actual capacity values to the control center is not guaranteed since the random link failures may result in having no active communication path connecting the control center to the sensor measuring a certain capacity.

We model the forecasts for renewable capacity and their errors by a discrete probability mass function (pmf) where the most likely outcome represents the forecast value and the remaining outcomes represent the possible realized capacity values above or below the forecast. We let N_{levels}^{renew} denote the number of possible capacity values the renewable source could take including its forecast. Similarly, we assume forecasts and errors in load capacities are also modeled by a discrete pmf with N_{levels}^{load} levels with the most probable level taken as the forecast. Capacity forecasts are typically based on historical user data, weather conditions, and other seasonal trends

which are all impossible to predict or forecast with perfect accuracy so fluctuations in capacities around their forecasts will always exist especially for intermittent renewable resources like wind and solar energy.

Failures in the communication network are modeled as a loss in any number of links from the graph describing the communication network. Failures in communication links are assumed to be independent and we model the failure of any given link by a Bernoulli random variable with probability π^{on}_{comm} such that the link fails with probability $1-\pi^{on}_{comm}$. For a real network, different communication links may be more likely to fail than others since the network may include different communication technologies or links having different characteristics. Further, the assumption of independent link failures may be inaccurate if a failure occurs due to severe weather conditions affecting links in a certain area of the network or if the failures are due to a complex cyber attack affecting multiples links at the same time. More complex probability models allowing communication links to have different failure probabilities or accounting for correlated link failures may be employed and should apply directly to our presented problem.

The number of possible link failures grows exponentially with N_{nodes} the number of nodes in the communication network since a network with N_{nodes} nodes could have up to $N_{links} \triangleq \frac{N_{nodes}(N_{nodes}-1)}{2}$ communication links and since each of those links is subject to failure, the total number of failure scenarios is $N^{comm}_{scenarios} = 2^{N_{links}}$.

Throughout this chapter, we assume that the pmfs describing load forecasts share the same N^{load}_{levels} and that if the i^{th} level is observed for a given load, the observed capacity for every load will be the one represented by the i^{th} capacity level. This assumption is meant to simplify our analysis by having a smaller number of possible scenarios and it may be removed in exchange for an increase in the computational complexity of the analysis. To further simplify our analysis, we also assume that only one generator in the grid is a renewable energy source and this assumption may also be removed in exchange for an increase in computational complexity.

Taken together, the three factors contributing to the stochasticity in the system may be used to index $\mathcal{S}$ the set of all possible random scenarios where elements of $\mathcal{S}$ may be given as a tuple (a,b,c) where $\mathcal{A} \in \{1,2,\ldots,N^{renew}_{levels}\}, \mathcal{B} \in \{1,2,\ldots,N^{load}_{levels}\}, \mathcal{C} \in \{1,2,\ldots,N^{comm}_{scenarios}\}$ such that $|\mathcal{S}| = N^{renew}_{levels} \times N^{load}_{levels} \times N^{comm}_{scenarios}$. We may also use $s \in \mathcal{S}$ as a general element of $\mathcal{S}$ to simplify the presentation when appropriate. We note that given a sufficiently robust communication network, the connectivity between sensor nodes and the control center could be maintained even under cases where many communication links have failed. For example, in Figure 4.1, all the sensor

nodes remain connected to the control center although five of the nine communication links have failed or are inactive. Next, we present our OPF formulation and we formally describe the effect of communication failures on the power grid.

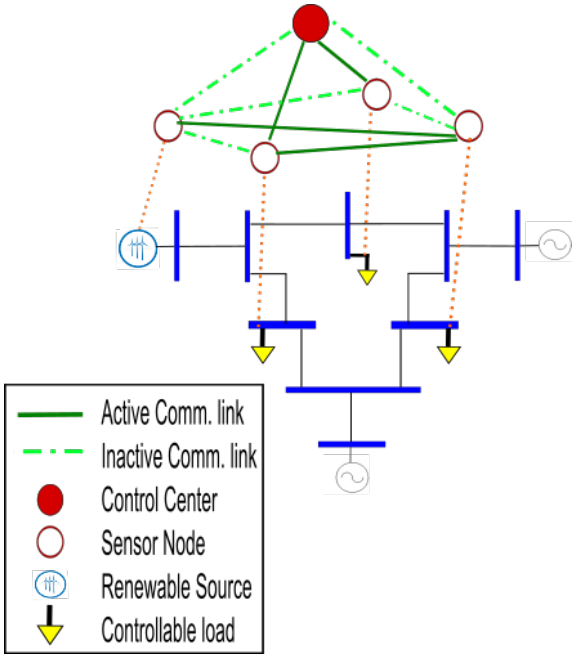


Figure 4.1: System Diagram for a communication network supporting a 9-bus system while having 5 failed links

4.2 Stochastic OPF Problem Formulation

In the stochastic DC OPF problem we consider, decision variables go beyond choosing dispatch levels for the loads and power generators as they include an added set of variables relating to the leased topology of the communication network. The optimal decisions are chosen to minimize the total expected cost for operating the power system while accounting for the different scenarios resulting from random fluctuations in load and renewable capacities and losing communication and control capabilities due to communication link failures. The objective function we consider is

$$\text{minimize} \sum_{i\in\mathcal{G}} C_i p_i + \sum_{i=1}^{N_{nodes}} \sum_{j=1}^{N_{nodes}} B_{i,j} X_{i,j} + \sum_{s=1}^{|\mathcal{S}|} \sum_{i\in\mathcal{G}} \pi_s^{sc} \Big(C_i^+ \Delta p_{s,i}^+ + C_i^- \Delta p_{s,i}^- \Big) + \sum_{s=1}^{|\mathcal{S}|} \sum_{i\in\mathcal{G}} \pi_s^{sc} \Big(D_i^+ \Gamma_{s,i}^+ + D_i^- \Gamma_{s,i}^- \Big), \tag{4.1}$$

where the minimization is over the problem decisions variables given in Table 4.1. The variables $\Delta p^+_{s,i}$ and $\Delta p^-_{s,i}$ are, respectively, the positive and negative parts of the decision variable $\Delta p_{s,i}$ which represents the adjustment in the dispatch level for generator/load $i \in \mathcal{G}$ under scenario $s \in \mathcal{S}$.

Table 4.1: Problem Decision Variables

Variable	Description
p	Vector of first settlement dispatch levels for all loads and generators
Δp_s	Vector of adjustments in the first settlement dispatch levels under scenario $s \in \mathcal{S}$
$\mathbf{X}$	Comm. links leased. A binary $N_{nodes} \times N_{nodes}$ matrix
$\mathbf{Y}^c$	Matrix of Communication flow under link failure scenario $c \in \{1, 2, \ldots, N^{comm}_{scenarios}\}$
w^c	Vector of node reachability from control center under communication link failure scenario $c \in \{1, 2, \ldots, N^{comm}_{scenarios}\}$

We formally define them by including

$$\Delta p_{s,i} = \Delta p^+_{s,i} - \Delta p^-_{s,i} \quad \forall s \in \mathcal{S}, i \in \mathcal{G}, \tag{4.2}$$

$$\Delta p^+_{s,i} \geq 0 \quad \forall s \in \mathcal{S}, i \in \mathcal{G}, \tag{4.3}$$

$$\Delta p^-_{s,i} \geq 0 \quad \forall s \in \mathcal{S}, i \in \mathcal{G}, \tag{4.4}$$

in our problem's constraints. Similarly, $\Gamma^+_{s,i}$ and $\Gamma^-_{s,i}$ are the positive and negative parts of the variable $\Gamma_{s,i}$. The variable $\Gamma_{s,i}$ represents the difference between the dispatch level and the capacity of generator/load i under scenario s. Since we take generating power to be positive and consuming it to be negative, $\Gamma^+_{s,i}$ and $\Gamma^-_{s,i}$ are defined by the constraints

$$p_i + \Delta p_{s,i} - p^{max}_{s,i} = \Gamma^+_{s,i} - \Gamma^-_{s,i} \quad \forall s \in \mathcal{S}, \tag{4.5}$$

$$\Gamma^+_{s,i} \geq 0 \quad \forall s \in \mathcal{S}, \tag{4.6}$$

$$\Gamma^-_{s,i} \geq 0 \quad \forall s \in \mathcal{S}, \tag{4.7}$$

when $i \in \mathcal{G}$ corresponds to a generator. When i corresponds to a load (4.5) is replaced with

$$p_i + \Delta p_{s,i} - p^{min}_{s,i} = \Gamma^+_{s,i} - \Gamma^-_{s,i} \quad \forall s \in \mathcal{S}. \tag{4.8}$$

The decision variables $\mathbf{Y}_c \ \forall \ c \in \mathcal{C}$ and $w_c \ \forall \ c \in \mathcal{C}$ do not appear in the objective function but

their values have a very significant role in affecting the optimal solution. We explain the role of $\mathbf{Y}_c$ and w_c next by first introducing the remainder of the constraints we consider that we categorize into

- Communication network constraints,
- Power grid constraints,
- Joint power and communication constraints,

and describe separately in the following subsections.

4.2.1 Communication Constraints

Communication constraints determine the system's ability to relay sensor measurements to the control center and the control center's ability to adjust power dispatch levels. Successful communication between sensors and the control centered is modeled by the existence of a path consisting of leased communication links connecting the control center to the load or renewable generator. We describe the initial communication topology as an undirected graph representing a 'flow' network whose topology is chosen as a first-settlement decision variable. The first-settlement constraints setting up our failure-free communication network are

$$\mathbf{X} = \mathbf{X}^{\mathbf{T}} \tag{4.9}$$

$$X_{i,i} = 0, \forall\ i \in \mathcal{N} \tag{4.10}$$

$$X_{i,j} \in \{0,1\} \forall\ i \in \mathcal{N},\ j \in \mathcal{N}, \tag{4.11}$$

where $\mathcal{N} \triangleq \{1, 2, \ldots, N_{nodes}\}$ and $\mathbf{X}$ is the matrix whose $(i,j)^{th}$ element is given by $X_{i,j}$ and is thus the adjacency matrix of the communication network.

A node in the communication network is either a *sensor node*, the *control center*, or a *relay node*. Since there can only be one control center, the size of $\mathbf{X}$ is one more than the sum of $N_{sensors}$ (the number of sensor nodes) and N_{relays} (the number of relay nodes). Each sensor node is assigned a flow supply of positive one, each relay node is assigned zero supply, and the control center is assigned a supply equal to $-N_{sensors}$. Then, for every link failure scenario, the communication network will attempt to send a unit of 'flow' along its active links from each sensor node to the control center. If an active path connecting the control center to a sensor exists, they are able to exchange realized capacity updates and control commands. If no active path exists between a

certain sensor node and the control center, we shut off that sensor's supply under that scenario by setting it to zero and also adjusting the 'flow' demand of the control center to accommodate the supply loss. We formally state these conditions by the following second-settlement communication network constraints, which are all repeated for every $c \in \mathcal{C}$

$$0 \leq Y^c_{i,j},\ Y^c_{i,j} \in \mathbb{N}\ \ \forall\ i \in \mathcal{N},\ j \in \mathcal{N}, \tag{4.12}$$

$$Y^c_{i,j} + Y^c_{j,i} \leq M Q^c_{i,j} X_{i,j},\ \ \forall\ i \in \mathcal{N},\ j \in \mathcal{N}, \tag{4.13}$$

$$\sum_{t=1}^{N_{nodes}} Y^c_{i,t} - \sum_{t=1}^{N_{nodes}} Y^c_{t,i} = 1 - w^c_i,\ \ \forall \text{ sensor } i \in \mathcal{N}, \tag{4.14}$$

$$\sum_{t=1}^{N_{nodes}} Y^c_{i,t} - \sum_{t=1}^{N_{nodes}} Y^c_{t,i} = 0,\ \ \forall \text{ relay } i \in \mathcal{N}, \tag{4.15}$$

$$\sum_{t=1}^{N_{nodes}} Y^c_{j,t} - \sum_{t=1}^{N_{nodes}} Y^c_{t,j} = \sum_{i=1}^{N_{sensors}} w^c_i - N_{sensors}, \tag{4.16}$$

$$w^c_i \geq 1 - \sum_{t=1}^{N_{nodes}} Y^c_{j,t},\ \ \forall \text{ sensor } i \in \mathcal{N}, \tag{4.17}$$

$$w^c_i \in \{0,1\} \forall \text{ sensor } i \in \mathcal{N}, \tag{4.18}$$

where the elements $Q^c_{i,j}$ in (4.13) make up the symmetric and binary $N_{nodes} \times N_{nodes}$ parameter matrix $\mathbf{Q}^C$ that specifies the failed communication links under scenario $c \in \mathcal{C}$. Constraint (4.13) is of great importance since it sets the capacity of every link (i,j) in the communication network to $M > 0$ if and only if the link (i,j) was leased in the first settlement (true when $X_{i,j} = 1$) and it was free from any failures (true when $Q^c_{i,j} = 1$) under c the considered scenario. M in constraint (4.13) is chosen as a sufficiently large positive number such that link (i,j) effectively has unlimited capacity when it is active[1]. For the remainder of the chapter, we will refer to links that have been leased by the operator and are free from failures as *active* communication links and we otherwise refer to them as *inactive* links.

Constraints (4.14)-(4.18) set up the flow problem between the sensors and the control center and adjust their supplies when no path of active communication links exists between a sensor node and the control center. Based on constraints (4.12)-(4.18), decision variables $Y^c_{i,j}$ represent the amount of 'communication flow' between nodes i and j under scenario $c \in \mathcal{C}$ and the binary decision variable w^c_i is equal to one only if no path exists connecting sensor node i to the control center and the sensor supply is therefore shut off.

[1]While M may be chosen to be arbitrarily large, we set it to $M = N_{sensors}$ since the amount of flow along any link cannot exceed the total flow supply in the network

Remark 3. *For constrains (4.12)-(4.18) to function as just explained, the cost parameters in the objective function, given in (4.1), must be chosen such that losing connectivity between a sensor node and the control center has a positive net effect on the objective function. Choosing the cost parameters in a manner that does not satisfy this condition is not reasonable as we see no possible practical situation where losing communications will be beneficial to the system operator.*

Next, we present our constraints that deal solely with the power grid before presenting our joint power and communication constraints which specify the effects of losing connectivity between a sensor and the control center.

4.2.2 Power Grid Constraints

Power grid constraints are the most typical constraints in a standard OPF problem and since they represent physical constraints, they must be satisfied in both the first and second settlement. For the first settlement, we express these constraints as

$$\sum_{i \in \mathcal{G}} p_i = 0, \tag{4.19}$$

$$\sum_{i \in \mathcal{G}} \Lambda_{k,i} p_i \leq t_k^{max}, \ \ \forall \text{ power lines } k, \tag{4.20}$$

$$p_i^{min} \leq p_i \leq p_i^{max}, \ \ \forall i \in \mathcal{G}, \tag{4.21}$$

where (4.19) is a power balance constraint requiring the total power generated to be equal to the total power consumed, $\Lambda_{k,i}$ is an element from the power transfer distribution factor (PTDF) matrix [98] such that (4.20) are line flow capacity constraints ensuring that the power flowing through lines in the power grid does not exceed the line's thermal limit. Constraint (4.21) is a capacity constraint limiting the amount of power we can draw from a generator or supply to a load. We take $p_i > 0$ to represent generating power and $p_i < 0$ to represent consuming power so we set $p_i^{min} = 0, p_i^{max} \geq 0$ for generators and $p_i^{min} \leq 0, p_i^{max} = 0$ for loads.

After the stochastic events are realized, the system operator must choose $\Delta p_{s,i} \ \forall s \in \mathcal{S}, i \in \mathcal{G}$ (the second-settlement recourse actions) and adjust the power dispatch levels according to the

realized scenario $s \in \mathcal{S}$. Therefore, for each scenario $s \in \mathcal{S}$, we will have the constraints

$$\sum_{i \in \mathcal{G}} \Delta p_{s,i} = 0, \tag{4.22}$$

$$\sum_{i \in \mathcal{G}} A_{ki}(p_i + \Delta p_{s,i}) \leq t_k^{max}, \ \ \forall k, \tag{4.23}$$

$$p_i^{min} \leq p_i + \Delta p_{s,i} \leq p_i^{max}, \ \ \forall i \text{ a conventional generator}, \tag{4.24}$$

where constraints (4.22)-(4.23) correspond to constraints (4.19)-(4.20) respectively. The constraints defined by (4.24) represent generation limits for conventional generators which we model as being independent of the communication network or capacity forecasts. The second settlement constraints corresponding to constraint (4.21) for loads and renewable generators are given in the following subsection as they are affected by both the status of the communication network and the random capacity fluctuations in the power grid.

4.2.3 Joint Power and Communication Constraints

This set of constraints specifies the possible effects of losing communications between a sensor node and the control center. When a sensor node is disconnected, the system operator will force the disconnected load/generator to draw/generate a pre-set amount of power $p_i^{no\_comm}$. For example, if $p_i^{no\_comm} = 0$, then the system operator and load/generator i have an agreement to disconnect from the grid if the they are unable to communicate with each other. In the numerical results, we explore an alternative agreement where loads and generators do no completely disconnect from the grid when their communication fails. We express our joint power and communication constraints as

$$w_{c,i} p_i^{no\_comm} \leq p_i + \Delta p_{s,i} \leq w_{c,i} p_i^{no\_comm} + (1 - w_{c,i}) p_{s,i}^{max}$$
$$\forall \ s \equiv (a, b, c) \in \mathcal{S}, \tag{4.25}$$

when i represents a renewable generator and as

$$(1 - w_{c,i}) p_{s,i}^{min} + \ w_{c,i} p_i^{no\_comm} \leq p_i + \Delta p_{s,i} \leq w_{c,i} p_i^{no\_comm}$$
$$\forall \ s \equiv (a, b, c) \in \mathcal{S}, \tag{4.26}$$

when i represents a load.

Taking everything together, our stochastic OPF problem minimizes the objective function (4.1) subject to the large set of constraints (4.2)-(4.26). In the following section, we present several numerical results and some key findings obtained using our stochastic OPF model.

4.3 Numerical Results and Discussion

In this section, we provide numerical results validating our presented model. We also present results describing the optimal communication topology under different conditions and stochastic events. For our results, we focus on the standard IEEE 9-bus system where we assume the first generator represents a renewable energy source and that all loads in the grid are controllable. Thus, we have one renewable generator, two conventional generators, and 3 controllable loads which we refer to as generators $1-6$ respectively since we may consider loads as 'negative' generators.

Throughout this section, we set $\pi_{comm}^{on} = 0.9$ so that the probability for any communication link to fail is 0.1. We further assume that the pmfs representing the forecasts for load and renewable generation are described by three capacity levels making $N_{levels}^{renew} = N_{levels}^{load} = 3$. where the middle level is the most probable outcome and, hence, represents the forecast value.

As mentioned earlier, we assume that the observed load capacities behave the same relative to their forecast value in the sense that, under any given scenario, the loads will all be either below, at, or above their own forecast level. The numerical results presented in this section all use the pmfs given in Table 4.2 to describe the possible realized capacity values.

Table 4.2: Forecast pmfs for stochastic Generators

Level	**Prob**	**Gen. 1**	**Gen. 4**	**Gen. 5**	**Gen. 6**
LOW	0.25	170	-85.5	-95	-118.75
FORECAST	0.50	200	-90	-100	-125
HI	0.25	230	-94.5	-105	-131.25

For the presented results, we adopted a strategy where $P_i^{no\_comm}$ is set to LOW forecast if i is the renewable source and $P_i^{no\_comm}$ is set to the HI forecast if the i represents a load. The adopted strategy may be regarded as a defensive, or worst-case, strategy since losing communications might force us to oversupply power to loads and under-utilize the cheapest generator in the system.

Other variables held constant for all the results we present in this section are the cost parameters given in Table 4.3. The cost values given in Table 4.3 were chosen and tuned in order for the optimal solutions obtained with our model to represent reasonable real world decisions.

Under the given cost structure, the system operator will always prefer to dispatch the free renewable generation first and will always aim to satisfy the load demands communicated to the control center. Although costs C_4, C_5, C_6 are all positive, they actually reward the system operator

Table 4.3: Values for Select Cost Parameters

Parameter	**Value**	**Parameter**	**Value**
C_1	0	C_1^+, C_1^-	0
C_2	60	C_2^+, C_2^-	120
C_3	40	C_3^+, C_3^-	80
C_4, C_5, C_6	1000		
C_4^+, C_5^+, C_6^+	200	C_4^-, C_5^-, C_6^-	200
D_1^+, D_1^-	500		
D_2^+, D_2^-	0	D_3^+, D_3^-	0
D_4^+, D_5^+, D_6^+	1000	D_4^-, D_5^-, D_6^-	5000

for delivering power to loads in the system since their power consumption is negative. Costs C_i^+, C_i^- for $i = 1, 2, \ldots, 6$ penalize the operator for the different adjustments in dispatch levels while costs D_i^+, D_i^- for $i = 1, 2, \ldots, 6$ penalize the operator if the net dispatch of the renewable generator is below its measured capacity or if the net dispatch for a load is different from its capacity. The only model parameter left to specify is $B_{i,j}$ for $i, j \in \mathcal{N}$ which represents the cost to lease links in the communication network.

For all the presented results, we set $B_{i,j} = B$ for all $i, j \in \mathcal{N}$. Figure 4.2 shows the impact of increasing B on the optimal number of leased communication links in a 5-node (4 sensors and 1 control center) communication network supporting the operation of an IEEE 9-bus system. Since we have $N_{nodes} = 5$ nodes in the communication network, the maximum possible number of communication links is $0.5N_{nodes}(N_{nodes} - 1) = 10$ which corresponds with the optimal number of leased links in Figure 4.2 when the cost is very close zero. As the cost to lease a communication link is increased, Figure 4.2 allows us determine the cutoff points where the number of leased communication links must be decreased as it will no longer be justified by the added communication paths they may provide. In fact, our model provides us with the actual optimal topology under any given cost and Figure 4.3 provides optimal topologies obtained when the optimal number of leased links was eight, six, five, and four. The optimal topology with four links is not surprising since it simply establishes direct connections between each generator and the control center. Optimal topologies with extra links, however, were not always straightforward to justify. For example, the optimal topology with six links displays a higher preference towards keeping generator six, which is the largest load in the network, connected to the control center over any of the other stochastic

generators. Of the three loads, generator six has the largest capacity deviations over its possible realized capacity values so the preference towards maintaining connection with it may be regarded to be lowering the probability that the system operator will have to incur the large cost associated with re-dispatching generator six to meet its realized demand or for over-supplying it with power when the communication with it is lost.

Figure 4.4 shows the change in the optimal objective value as B the leasing cost per link is increased for the same 5-node communication network supporting an IEEE 9-bus system we considered for Figure 4.2. As B is increased, the objective function is linear over the intervals where the optimal communication topology is unchanged since the only change in the objective function will be the number of communication links leased multiplied by the amount of increase in B. Further, since the optimal number of leased communication links is a decreasing function in B, the slope of the linear segments making up the objective function will also decrease with B as it is equal to the total number of leased communication links. As a result, the objective function is a concave function in B which can be easily observed from Figure 4.4.

The same experiments whose results are presented in Figures 4.2-4.4 were run for a communication network of size 6 where the extra communication node was taken as a relay. The results obtained with 6 nodes in the communication network were nearly identical to the ones presented in 4.2-4.4 with the main difference being that the optimal communication network would lease all 15 links when B is very close to zero and, therefore, those results were omitted.

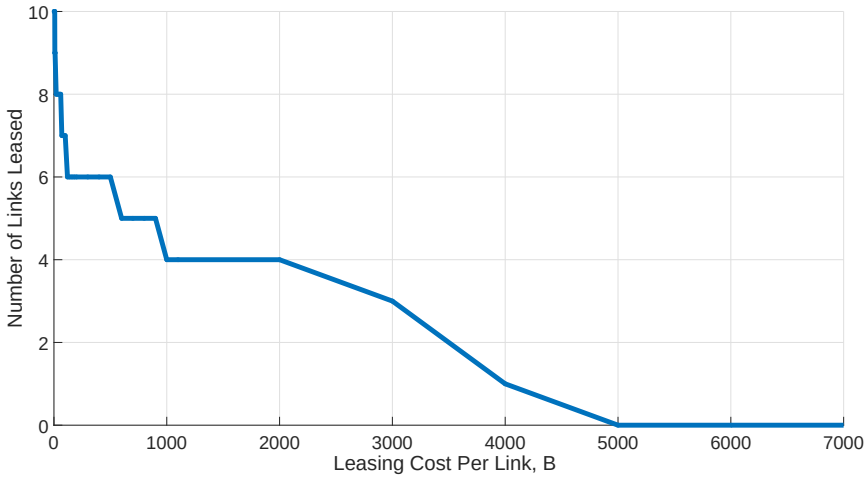
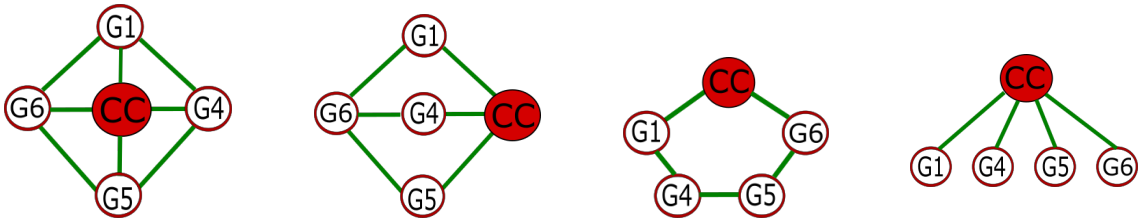


Figure 4.2: Number of links leased vs. increasing link costs for a 5-node communication network supporting an IEEE 9-Bus system

Figures 4.2-4.4 show results where we consider every possible combination of communication



(a) 8 Links Leased (b) 6 Links Leased (c) 5 Links Leased (d) 4 Links Leased

Figure 4.3: Optimal Topologies for Different Numbers of Links Leased

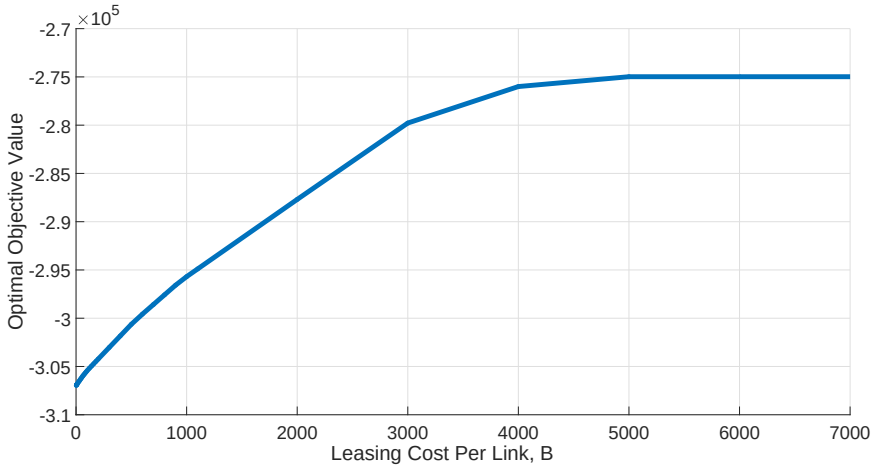


Figure 4.4: Optimal objective value vs. increasing link costs for a 5-node communication network supporting an IEEE 9-Bus system

link failures. As a result, the communication topologies obtained under such conditions take into account what might be extreme and highly unlikely scenarios where all the links in the communication network may fail. In reality, it is perhaps more reasonable to only consider scenarios where at most only a few links may fail at the same time.

Figure 4.5 shows the optimal number of leased communication links as K the maximum number of communication links that may fail at any given time is increased from 1 to its maximum value of 10 under several different values for B. Similar to the results from Figure 4.2 (where K was fixed at 10), Figure 4.5 indicates that the optimal number of leased communication links decreases as B is increased. More importantly, Figure 4.5 indicates that the optimal number of leased communication links increases with K which should be expected as more communication links would be needed to defend against the increased risk of having more links fail at the same time. With $K = 1$, the optimal communication topology leased 5 links and arranged them, as indicated in Figure 4.3c, such that the communication network is a single loop. A loop topology provides the ability for all the generators to remain connected to the control center under any single link failure

so the system operator will not gain from leasing any additional links. With $K = 2$, the optimal communication topology required 8 links arranged according to Figure 4.3a where all the sensors would remain connected under any two-link failure scenario. We note that the optimal link choice is not unique for the 5 link topology since the order of the different generators in the loop will have no affect on the scenario probabilities or any other factor in the optimization problem. The same is true for the 8 link topology where the actual leased link may be exchanged with another set of links as long as the general network structure is preserved.

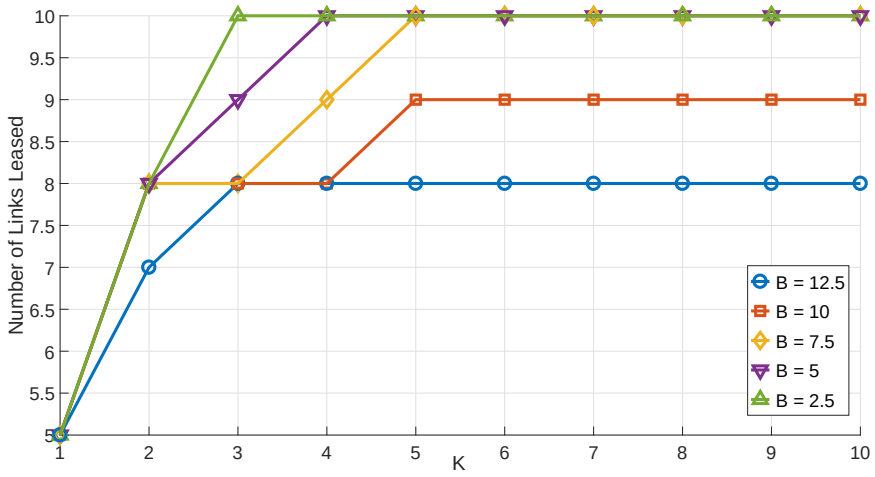


Figure 4.5: Number of links leased vs. maximum number of failed links for a 5-node communication network supporting an IEEE 9-Bus system

Finally, Figure 4.6 provides plots showing the change in the optimal objective value as K is increased under the same cost values considered for Figure 4.5. Figure 4.6 seems to indicate that the optimal objective function is also concave in K. While this is associated with the fact that the number of leased links is an increasing function of K, having higher K values changes the probabilities of the different stochastic scenarios so proving that the objective function is concave in K may be more complex. Figure 4.6 also demonstrates that accounting for a single link failure could greatly underestimate the optimal objective value for the full problem accounting for all the possible failure scenarios and that considering just a few more failures may lead us to a more accurate estimate.

Our formulation relies on enumerating all possible communication link failures and so the number of scenarios in the full problem grows exponentially with every added link in the system. As a result, applying our full model to larger power grids and their possible communication networks may not be practical from a numerical viewpoint. Fortunately, this problem in not unique to our

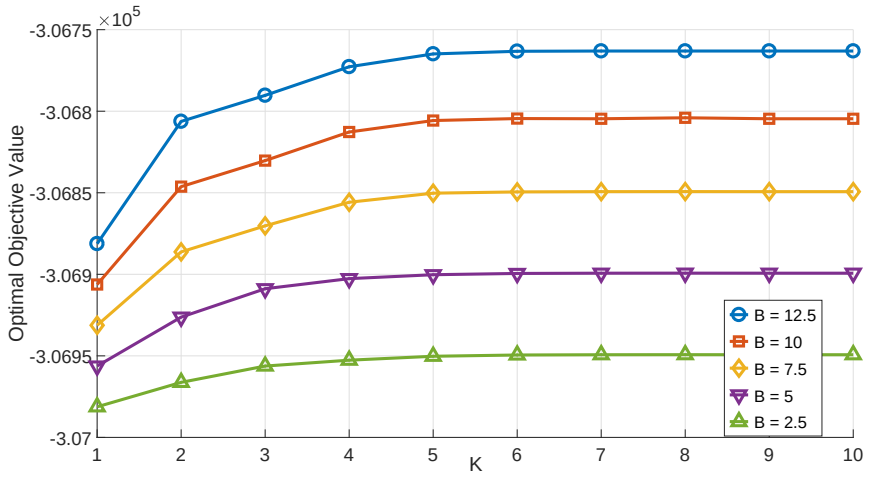


Figure 4.6: Optimal objective value vs. maximum number of failed links for a 5-node communication network supporting an IEEE 9-Bus system

model and there are many techniques in the literature discussing possible methods of reducing the size of a stochastic optimization problem in a manner that preserves its essence and closely approximates its optimal solution. One of the possible solutions was already introduced in this chapter by considering smaller values of K since that decreases the total number of communication scenarios. Other possible solutions may include well-established techniques in the literature on sampling, see [99–102] for example. Also, there may be practical considerations that limit the problem's numerical complexity. For example, establishing a communication link between nodes associated with loads or generators that are sufficiently far away from each other in the power grid may not be possible in the real world and so we may be able to exclude such links from our analysis.

4.4 Conclusion

In this chapter, we presented a novel two-settlement stochastic OPF model accounting for failures in communication as well as errors in forecasting loads and renewable generation. Our OPF model's decision variables extend beyond power dispatch and re-dispatch decisions to include the decision variables identifying the optimal topology for the communication network supporting the considered power system. We presented several numerical results on the optimal communication network supporting an IEEE 9-bus system. Based on the presented results, we are able to determine the effect that increasing the per link cost for leasing a communication network will have on the optimal communication topology and we are able to uncover preferences towards establishing more

communication paths between certain sensors in the grid and the control center that are not be obvious.

Chapter 5

Topology Estimation in Power Distribution Networks

In this chapter, we consider the case where sensors are placed at nodes in the system and we assume that endowing a node with a sensor allows us to measure the power flow along every line incident to that node. We present a novel topology estimation scheme for radial distribution networks based on slow-changing nodal demand forecasts and power flow measurements obtained from sensors installed at subset a of the nodes. Also included in this chapter is a proposed sensor placement scheme based on a deterministic treatment of the load forecasts. The sensor placement obtained by following the proposed scheme allows for the identification of any number of detectable faults in the system. Numerical results are presented demonstrating the performance of our topology estimation scheme for the case where the distribution network is the IEEE-123 bus test feeder and the employed sensors are placed at the locations obtained by following our proposed sensor placement scheme.

The remainder of this chapter is organized as follows. In the next section, we present a description of the system model and introduce some of our notation. In Section 5.2, we present the fault detection problem along with our proposed sensor placement scheme. Next in Section 5.3, we describe our novel topology estimation approach. Section 5.4 presents numerical results where we employ our proposed sensor placement scheme along with the presented topology estimation approach to detect randomly generated faults in the IEEE 123-bus distribution feeder. Finally, Section 5.5 summarizes our results and concludes this chapter.

5.1 System Description and Notation

We consider power distribution networks operated in a radial (tree) structure where the power is supplied to the feeder at the root node of the tree. We assume that all lines in the network are susceptible to faults except for the line connecting the root node to the main power grid. We focus on detecting the number and location of faults in the network by relying on forecasts of nodal power demands and power flow measurements obtained from a limited number of carefully placed sensors.

We assume the complete error-free network topology is known and described by the graph $\mathcal{T}_{full} = (V, E)$ where the nodes are given by $V = \{v_0, v_1, \ldots, v_N\}$ and the lines are given by $E = \{e_1, e_2, \ldots, e_N\}$ where $e_n \triangleq (v_m, v_n)$ and v_m is the parent node of v_n. Figure 5.1 shows a simple tree demonstrating our naming convention.

Sensor placements are defined as a set of nodes $\mathcal{P}$ where $\mathcal{P} \subset V$ and we denote by $E(\mathcal{P})$ the set of lines whose power flows are measured under placement $\mathcal{P}$. If a node is endowed with a sensor, we are able to measure the power flow along all lines incident to that node and, therefore, we may define $E(\mathcal{P})$ as

$$E(\mathcal{P}) = \{e|\ e \subset E \text{ and } e \text{ is incident on some } v \in \mathcal{P}\}. \tag{5.1}$$

Each node in the network is taken to have a power demand (load) that must be supplied through a fault-free path connecting it to the root node. Nodal power demands are set to zero if the node serves as a relay which is a common function of at least a few nodes in any real world distribution network. Nonzero nodal demands are assumed to be based on load forecasts and are denoted by the vector $\mathbf{d}^+ = (d_{i_1}, d_{i_2}, \ldots, d_{i_M})^T$ where d_{i_m} represents the load forecast for the $(i_m)^{th}$ node with $i_m \in \{1, \ldots, N\}$ and $M \leq N$ is the number of nodes having a nonzero demand forecast. Since nodal load forecasts are typically based on infrequent grid measurements, historical data, and prevailing weather conditions, they are bound to have some errors and other inaccuracies. We assume the forecasting errors may be modeled as additive noise such that $\tilde{\mathbf{d}}^+ = (\tilde{d}_{i_1}, \tilde{d}_{i_2}, \ldots, \tilde{d}_{i_M})^T$ the vector of realized nodal demands is jointly Gaussian with zero mean and a known diagonal covariance matrix $\Sigma = \sigma^2\mathbf{I}$ of size M resulting in the vector $\tilde{\mathbf{d}}^+$ being distributed as $\tilde{\mathbf{d}}^+ \sim N(\mathbf{d}^+, \Sigma)$.

Under a radial network drawing power from its root, the power flow through a line $e_n \in E$ is given as the sum of the demand for node v_n and all the nodes downstream from v_n. Thus, the

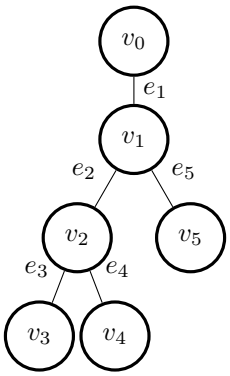


Figure 5.1: Tree showing naming convention

measured power flow through line $e_n \in E$ is given by

$$y_{\mathcal{T}_{true}}(e_n) = \sum_{i \in \mathbf{Downstream}(v_n; \mathcal{T}_{true})} \tilde{d}_i, \tag{5.2}$$

where $\mathbf{Downstream}(v_n; \mathcal{T}_{true})$ is a function returning a set comprised of n and the indices of all the nodes downstream from v_n in $\mathcal{T}_{true}$ the unknown true (actual) network topology. Based on our assumption for the noise in the load forecasts, a nonzero $y_{\mathcal{T}_{true}}(e_n)$ will be normally distributed with a mean $\bar{y}_{\mathcal{T}_{true}}(e_n)$ given by the right hand side of (5.2) where each nonzero $\hat{d}_i$ value is replaced by d_i (its forecast) and a variance equal to σ^2 multiplied by the number of downstream nodes with nonzero forecasts.

Faults are modeled as a loss in any number of lines in the network and are given by $F \subset E$ such that the resulting faulted network is a forest $\mathcal{F}_{fault} = (V, E_f)$ where $E_f \triangleq E \setminus F$ and $|F|$ are both unknown. Because $\mathcal{T}_{full}$, the error-free network, is a single connected tree, every line fault will increase the number of trees in the faulted graph describing the network by one. Thus, if we suppose that $|F| = k$ faults have occurred, the graph describing the faulted network $\mathcal{F}_{fault}$ is a forest comprised of $k+1$ components or trees $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k$.

Since power is drawn solely from the root node, only one of the trees comprising $\mathcal{F}_{fault}$ will be energized and that will be tree the containing the root node which we denote by $\mathcal{T}_0$ without any loss in generality since the numbering is arbitrary. $\mathcal{T}_0$ may be described as $\mathcal{T}_0 = (V_0, E_0)$ where $V_0 \subset V$ and $E_0 \subset E_f \subset E$ are respectively the set of nodes and lines that remain connected to the power feeder after the occurrence of all the faults. As a result, all the lines not included in E_0 will have zero power flowing through them and all sensor measurements collected under any possible placement on these lines will be equal to zero. Therefore, a fault occurring in a line (v_m, v_n) where

both $v_m \neq V_0$ and $v_n \neq V_0$ will have no effect on the grid-connected part of the system or any of the measurements collected from lines in E_0. The detection of this type of faults excluded from our analysis *Definition 1* highlights this type of faults and the reason for excluding them.

Definition 1 (Topologically Undetectable Faults)**.** *Any fault affecting a line $(v_i, v_j) \in E$ that would otherwise still be excluded from the grid-connected part of the network will have no effect on any possible measurement under any possible sensor placement. Such a fault carries no practical significance on the topology estimation problem and is called a topologically undetectable fault.*

An alternate description of a topologically undetectable fault is given in [70, 71] where it is defined as a fault occurring downstream from another fault and the authors of [70,71] also excluded these faults from their analysis.

Next, we provide a brief description of the fault detection problem and present our sensor placement scheme.

5.2 Sensor Placement and The Fault Detection Problem

To better explain our proposed sensor placement scheme, we first define the fault detection problem. An optimal detector solving the topology estimation (fault detection) problem must determine both the number of faults and the lines affected by those faults. We consider the case where the detector is based on the knowledge of $\mathcal{T}_{full}$ (the fault-free network topology), nodal load forecasts, and the sensor measurements $\mathbf{Y} = (y(e_{n_1}), y(e_{n_2}), \ldots, y(e_{n_{|\mathcal{P}|}}))^T$. One possible detection method is to employ a maximum likelihood (ML) detector that may be expressed as

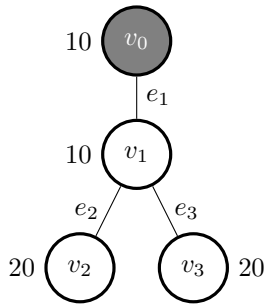
$$\hat{F} \in \underset{F \in \mathbb{P}(E)}{\arg\max} \Pr(\mathbf{Y}|\hat{\mathbf{d}}^+, F), \tag{5.3}$$

where F is a set containing the lines that are in outage, $\mathbb{P}(E)$ is the superset of E, and $\hat{F}$ is chosen after enumerating over all elements of $\mathbb{P}(E)$. Sadly, the size of $\mathbb{P}(E)$ is exponential in $|E|$ (the number of links in the network) making enumerating all the possible fault scenarios infeasible for most real distribution networks. The number of fault scenarios was shown in [71] to still grow exponentially in $|E|$ for a worst case distribution network even after excluding all topologically undetectable faults. However, the authors of [71] were able to show that the ML detection problem (5.3) may be decoupled and solved over smaller disjoint subtrees in the network as long as each subtree is rooted at a node where the amount of power it draws from its parent node is measured

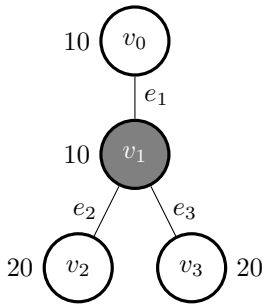
under the employed sensor placement. As a result, having a few sensors measuring power flows along several lines in the network can help us solve (5.3) by decoupling the full distribution network into smaller and more manageable problems.

Ambiguity in the ML detector (5.3) is another problem affecting the detection performance. This problem occurs when the sensor locations are such that multiple fault scenarios result in the same expected flow along the lines we measure. In most cases, adding carefully placed sensors to the network will enable us to differentiate between the different faults causing the ambiguity. This may be done by placing a sensor closer to the lines whose failure is causing the ambiguity of the detector. An example showing the effect of sensor placement on our ability to identify and differentiate between all detectable faults is demonstrated in Figure 5.2. The gray nodes in Figure 5.2 represent a node endowed with a sensor, which allows us to measure the power flow along all the lines incident to it, and the number next to each node represents the node's load or demand.

With the sensor placed at v_0 in Figure 5.2a, we are able to observe the power flow only along line e_1 where we expect the measured flow to belong to the set $\mathcal{C}(e_1) = \{0, 10, 30, 50\}$. Since the demands for nodes v_2 and v_3 are both 20, all the values in $\mathcal{C}(e_1)$ can be mapped to unique detectable fault scenarios except for 30 which represents either of e_2 and e_3 being in outage. Alternatively, placing the sensor at v_1 in Figure 5.2b allows us to measure the power flow along e_1, e_2, and e_3. This placement now enables us to distinguish between the case where e_2 is in outage and the case where e_3 is in outage since the sets of expected flow measurements are $\mathcal{C}(e_2) = \{0, 20\}$ for e_2 and $\mathcal{C}(e_3) = \{0, 20\}$ for e_3 and each value in both sets now maps to a unique fault. Note that if v_0 was connected to other nodes and having a sensor there was required so that other faults in the network remain distinguishable, we would have needed to add an extra sensor at v_1 instead of simply moving the one at v_0.



(a) Faults in e_2 and e_3 indistinguishable



(b) Faults in e_2 and e_3 are distinguishable

Figure 5.2: The effect of sensor placement on the ability to distinguish faults

Unfortunately, there are fault scenarios that result in cases where the ambiguity of the ML detector (5.3) cannot be resolved under any possible sensor placement. This type of faults exists due to having relay nodes with zero load demand. Consider for example the tree and sensor placement given in Figure 5.2b but suppose that v_1 is a relay node with zero demand instead of 10. In this case, we will be unable to distinguish between the scenario where e_1 is in outage and the scenario where both e_2 and e_3 are in outage although our sensor allows us to measure the flow along every line in the network. We refer to this type of faults as *numerically indistinguishable* since distinguishing between them is not possible by any detector under any sensor placement and they are indistinguishable due to the numerical value of a node's demand. We call a group of such faults that are all associated with the same expected flow measurements as a *set* of numerically indistinguishable faults. Throughout this chapter, we consider the occurrence of a fault from a set of numerically indistinguishable faults as being equivalent to having all the faults in that set occur and we consider the detection of a fault that is numerically indistinguishable as being equivalent to detecting all the faults from the same set.

In the next subsection, we present a sensor placement scheme where the sensor locations are chosen such that numerically indistinguishable faults are the only indistinguishable faults in the network.

5.2.1 Proposed Placement Scheme

In this subsection, we propose a sensor placement scheme for the purpose of detecting faults in a radial distribution network based on nodal load forecasts and line flow measurements. The obtained sensor locations result in all topologically detectable faults in the network being identifiable. *Definition 2* formalizes the concept of identifiability.

Definition 2 (Identifiable faults)**.** *A fault in line e_n is identifiable under a given placement and nodal load forecasts if the expected power flow along e_m, the first line endowed with a sensor along the path from v_n to v_0 (the root node), is unique under all topologically detectable faults involving e_n and any or none of the lines that are downstream from v_m along e_m. The uniqueness of the expected flow along e_m is allowed to be violated only when numerically indistinguishable faults are being considered.*

Our placement scheme is given in **Algorithm 5.1** as the recursive function **Placement**. The presented method depends on the functions **Child**, **UpFlow**, **UpdateTopology**, **Size**, **IsEmpty**, and **CombineVects** that are all briefly described in Appendix 5.6. The sensor placement for the

full distribution network is obtained by calling the function **Placement** with the inputs v_n set as the root node, $\mathcal{P}$ chosen as an empty set, $\mathcal{T}$ chosen as $\mathcal{T}_{full}$ (the fault-free distribution network), and the vector $\mathbf{d} = (d_0, d_1, \ldots, d_N)^T$ constructed such that d_i is the load forecast for node i and $d_i = 0$ when v_i is a relay node.

Since the detection problem may be decoupled over disjoint subtrees rooted at a node whose incoming power flow is measured by a sensor [71], the set of expected flow values for a line may be constructed through considering the subtree it belongs to as an isolated network independent from the remainder of the distribution network. The difficulty in taking advantage of the decoupling of the problem is that it is by placing the sensors that we create the subtrees over which the problem may be decoupled.

Our proposed placement approach starts at the root node and traverses the tree in a depth-first search manner while constructing arrays (vectors) of expected power flows for the lines it visits. The full vector of expected power flows for a line e_n is constructed only after visiting all child nodes of v_n so the complete vectors we construct start at leaf nodes of the tree. The constructed vectors continue to grow as we backtrack from leaf nodes since they will always have a unique set of expected power flows (a leaf node cannot have zero demand) and, therefore, no sensors are placed at leaf nodes. We also do not place sensors at nodes having a single child, since if v_n is a node with only one child node, the expected power flows for the line e_n will always be unique, provided the node does not have a zero demand. If node v_n has a single child and its demand is zero, the repeated value in the expected flows for line e_n is discarded since it is associated with a numerically indistinguishable fault. As a result, we only place sensors at a node v_m with more than one child node if the expected power flow values for line e_m are not all unique.

Choosing to place a sensor at v_m creates a subtree that we may ignore when continuing to traverse the remainder of the distribution network due to the decoupling in the ML detection problem. Therefore, when we backtrack from a node where we decided to place a sensor (Lines 41-45 in **Algorithm 5.1**), we return to the parent node an empty array of expected flow values. Otherwise, if the elements in the vector of expected flow measurements are all unique and no sensor is required at v_m, we continue to grow the vector until another sensor is placed or we finish traversing the remainder of the tree.

Having repeated values in the vector of expected flows for a line e_n indicates that some faults occurring downstream of the node v_n will be unidentifiable if the sensor is placed at v_n's parent. Therefore, it is by construction that our presented scheme ensures that all the detectable faults in

the network will be identifiable since we use the test for repeated values in the vector of expected flows for ever line we visit as the condition for including a sensor in the placement. Further, the obtained placement is guaranteed to have the minimum number of sensors required to achieve identifiability. This can be explained by first supposing that we remove a sensor from the obtained placement. By construction, removing any sensor from the obtained placement is guaranteed to result in having some unidentifiable faults. We could attempt to recover identifiability in the network by changing the locations of some of the sensors we did not remove. However, we are unable to move any of the remaining sensors upstream since doing so was originally prevented by the algorithm because it would have led to having unidentifiable faults. Moving a sensor downstream and closer to the location of the removed sensor will not work either since we will be required to move it all the way to the location of the removed sensor and that would lead us to a case equivalent to the one we are trying to solve my moving the sensor. Thus, no other placement scheme can guarantee the identifiability of all faults in the network while also employing fewer sensors than we require under our proposed placement scheme.

Next, *Remark 4* discusses the effect of having noisy nodal load forecasts on the detectability of identifiable faults and then we present our fault detection scheme.

Remark 4. *It is important to note that having a placement where all detectable faults satisfy Definition 2 does not guarantee their perfect detection when noise is added to the system. For example, consider the case where the distribution network and sensor placement are given in Figure 5.2a and suppose that the nodal loads forecasts for* v_2 *and* v_3 *are respectively* 19.9 *and* 20.1. *In this case, all faults in the network are identifiable but if the variance of the noise affecting the load forecasts is sufficiently large, the performance of a fault detector relying on measurements of power flow along* e_1 *will likely be unfavorable unless we are able to collect a sufficiently large number of measurements for the flow on* e_1 *in order to obtain a better estimate of the mean power along* e_1. *On the other hand, having a placement where some faults fail to satisfy Definition 2 guarantees that any employed detection method will not be able to differentiate between those fault under any number of collected measurements, and even in the absence of noise, since the mapping between the expected flow measurements and the possible fault scenarios will not be unique.*

5.3 Fault Detection Scheme

This section presents a novel fault detection scheme under the problem setup described in Section 5.1. In the absence of noise, our approach will detect all the topologically detectable faults in the network if the employed sensor placement is such that all faults in the system are identifiable. If the sensor placement is such that there are some unidentifiable faults in the system, our presented approach will detect all the possible fault scenarios that map to the measured power flows.

We first summarize the steps taken by our scheme and then provide more detailed descriptions of the steps afterwards:

1. For every node endowed with a sensor, remove any directly identifiable faults on the lines connecting the node to its children.
2. Identify all the lines measuring zero flow.
3. Adjust the expected flow along the remaining lines in $E(\mathcal{P})$ based on **Step 1** and **Step 2**.
4. Test if the flow on the lines with nonzero flow matches with the updated expected flows, starting with the deepest lines first.
5. Identify subtrees as the ones where the expected flow along the line separating the tree from the network is significantly different from its expected value and estimate the difference between the measured flow and its expected value.
6. Generate a vector of expected flows for each identified subtree with faults having expected flow smaller than the difference estimated in **Step 5**.
7. Use ML to detect faults in the subtree.
8. Check for faults on lines connected to nodes endowed with a sensor from above.
9. Generate estimate of grid-connected tree based on results from **Step 1**, **Step 7**, and **Step 8**.

Since installing a sensor at a node v_n allows us to measure the flow on every line incident on it, if we measure zero flow along a line drawing power from v_n but measure a positive flow on the line e_n, then we may directly identify that a fault has occurred on the line with zero flow. Identifying and removing such faults is done in **Step 1** of the proposed approach. **Step 2** simply identifies

the lines with zero flow and **Step 3** simply adjusts our expected values of $\mathbf{Y}$ according to **Step 1** and **Step 2**.

Next in **Step 4**, for all the lines with a nonzero measured flow, and following a decreasing depth order, we test whether the measured flow matches with our expected value for it where the test is formulated as the simple binary hypothesis

$$H_0 : y_{\mathcal{T}_{true}}(e_n) = \bar{y}_{\mathcal{T}_{updated}}(e_n),$$

$$H_1 : y_{\mathcal{T}_{true}}(e_n) < \bar{y}_{\mathcal{T}_{updated}}(e_n), \tag{5.4}$$

where $\bar{y}_{\mathcal{T}_{updated}}(e_n)$, which is obtained by using (5.2) after replacing $\tilde{d}_i$ with d_i, is our updated expected value for $y_{\mathcal{T}_{true}}(e_n)$ (the measured flow on line e_n). The hypothesis under H_1 in (5.4) is one sided since we expect the mean flow on a line to decrease under a fault. Under our assumption of additive white Gaussian noise for the errors in load forecasts, deciding between the two hypotheses may be done based on

$$z_n = \frac{\left(y_{\mathcal{T}_{true}}(e_n) - \bar{y}_{\mathcal{T}_{updated}}(e_n)\right)^2}{\sigma^2}, \tag{5.5}$$

for a chosen value of σ and later comparing z_n to a threshold value obtained under a given desired false alarm probability.

For every line e_n identified in **Step 4** as having a flow that is less than its expected value, in **Step 5** we identify all the lines and nodes downstream from v_n in our network as an isolated subtree $\mathcal{S}(e_n)$ and we update $\mathcal{T}_{updated}$ after each such step. Further, we associate with each isolated subtree a value $\Delta y(e_n) = y_{\mathcal{T}_{true}}(e_n) - \bar{y}_{\mathcal{T}_{updated}}(e_n)$ which is the difference between the measured flow on e_n and our expected value for it.

Next in **Step 6**, for each identified subtree $\mathcal{S}(e_n)$, we generate a vector of expected power flows under all the detectable faults that may occur in $\mathcal{S}(e_n)$. Generating the vector of expected flows under all possible faults could be computationally expensive. Thus, if there are single line faults that would result in a $\Delta y(e_n)$ that is significantly larger than the one we obtained in **Step 5**, we will choose to exclude it from the process of generating the vector of expected power flows for the subtree. We formulate the test determining whether to include faults involving a line e_n in the generation of the vector of expected power flows for a subtree $\mathcal{S}(e_n)$ as the simple binary

hypothesis

$$H_0 : \Delta(y_n) \leq y_{\mathcal{T}_{true}}(e_n) - \bar{y}_{\mathcal{S}(e_n)\setminus e_m}(e_n),$$

$$H_1 : \Delta(y_n) > y_{\mathcal{T}_{true}}(e_n) - \bar{y}_{\mathcal{S}(e_n)\setminus e_m}(e_n), \tag{5.6}$$

where $\bar{y}_{\mathcal{S}(e_n)\setminus e_m}(e_s)$ represents the expected power flow on line e_n under the single fault scenario where line e_m is in outage. If we accept H_1 in (5.6), we may ignore all faults involving e_m when constructing the vector of expected power flows for subtree $\mathcal{S}(e_n)$. To generate the vector of expected power flows for **Step 6**, we may traverse $\mathcal{S}(e_n)$ using a depth first search method as long as we continue to accept H_1 in (5.6). Then, if we find a line where H_1 in (5.6) is rejected, we can generate a vector of expected power flows associated with the faults involving that line and all the lines below it before backtracking and continuing to traverse the tree using depth first search as long as we accept H_1. Once all the vectors of expected power flows are generated over the different regions of $\mathcal{S}(e_n)$ where we accept H_1, they can be used to create the vector of expected flows for **Step 6** by using the function **CombineVects** which is given in the Appendix and was previously used by **Algorithm** 5.1.

In **Step 7** and for each subtree $\mathcal{S}(e_n)$, we employ the ML detector given in (5.3) after replacing $\mathbb{P}(E)$ with the set of faults that map to the values comprising the vector of expected flows obtained in **Step 6** and also replacing $\mathbf{Y}$ by the power flow measurement for line e_n. **Step 7** is where we tend to identify the majority of the faults in the network.

In **Step 8**, we focus on the faults occurring on lines $e_n \in E(\mathcal{P})$ when v_n is a node endowed with a sensor and the measured flow on e_n is zero. For every such line, if none of the faults identified in **Step 1** or **Step 7** are on the upstream path from the v_n to the root node (or the next node endowed with a sensor), we declare the line e_n as being in outage. Finally, **Step 9** combines all the faults found in **Step 1**, **Step 7**, and **Step 8** to produce a final estimate of the gird connected tree $\mathcal{T}_0$ and the set of topologically detectable faults.

The next section presents numerical results demonstrating the performance of the just presented fault detection scheme for the case where the sensors in the system are installed according to the placement scheme described in Section 5.2.

5.4 Numerical Results

This section presents results demonstrating the performance of the fault detection scheme proposed in Section 5.3 when the sensors in the network are placed according to the approach given in Section 5.2. Throughout this section, we consider the case where the distribution network is the IEEE 123-bus test distribution feeder where are all the switches are assumed to be at their nominal positions and we take the load demands to be the sum of the spot loads over all three phases.

For the IEEE 123-bus test feeder, the approach described in Section 5.2 will result in a sensor placement with 20 sensors located at nodes

$$\begin{aligned}\mathcal{P}^* =&\{1, 3, 8, 13, 18, 23, 26, 36, 40, 44, 57,\\ &67, 76, 78, 81, 89, 93, 97, 105, 110\},\end{aligned} \tag{5.7}$$

where the node numbers in (5.7) follow those given in the documentation of the test feeder. Figure 5.3 shows the IEEE-123 bus feeder with the locations of the sensors under $\mathcal{P}^*$ indicated by circles.

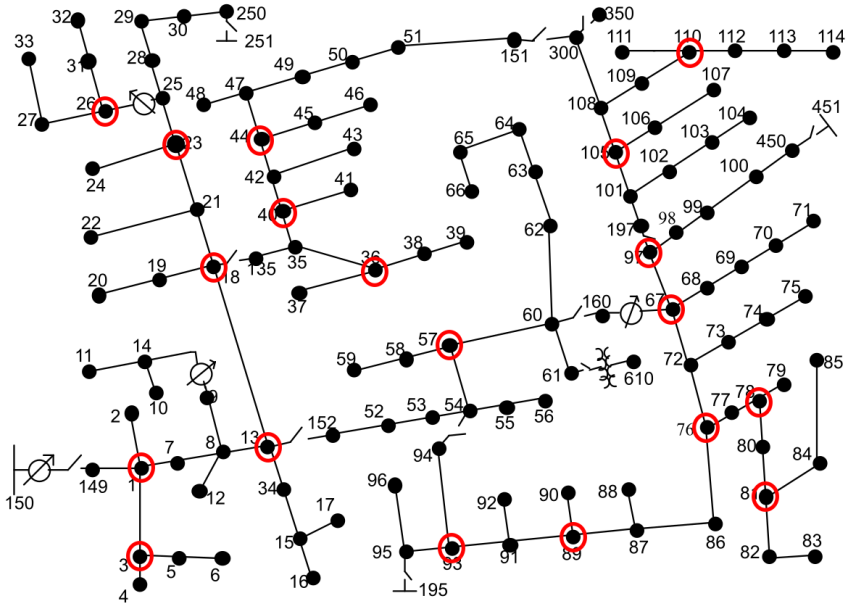


Figure 5.3: IEEE 123-bust test feeder with the sensor placement obtained using our proposed approach indicated

To test the performance of the estimation approach introduced in Section 5.3 under the sensor placement given by (5.7), we first set a value for the standard deviation σ of the additive noise in

the nodal load forecasts. Next, a number N_1 is chosen at random and then we randomly select N_1 different lines in the network to be in outage. For every chosen value for σ, we repeat the process of randomly choosing N_1 followed by randomly generating N_1 faults for 1000 different runs. Afterwards, we record the number of runs where the faults identified by our approach were all the topologically detectable faults that were randomly generated for that run. Then, we divide the number of runs where we were successful and divide that by the number of runs (1000) in order to obtain an estimate for the probability of detection of our approach.

Figure 5.4 includes plots showing the estimated probability of detection with our approach for increasing values of the noise standard deviation of $0.1, 0.25, 0.5, 1, \sqrt{2}, 2$. We note that the smallest nodal demand in the IEEE-123 feeder is 20 so a standard deviation of 2 is quite significant. The two graphs in Figure 5.4 represent the case where we only rely on measurements of real power flow ($N = 1$) for our fault detection and the case where we use both real and active power flow measurements ($N = 2$) under the assumption that the noise in the two measurements is independent and identically distributed (iid). From Figure 5.4, it easy to see that employing both real and active power flow measurements significantly increases our probability of detection, especially for higher values of σ. This should be expected as including active power flow measurements doubles the number of measurements we employ for our detection.

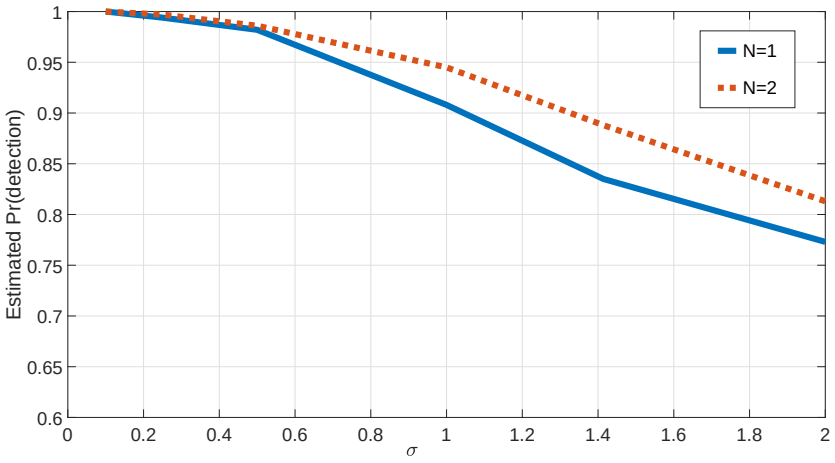


Figure 5.4: Estimated probability of detection vs increasing values of noise standard deviation under an unrestricted number of faults

The results from Figure 5.4 were obtained for the case where the number of faulted lines in the network was chosen at random from a uniform distribution over the number of lines in the feeder. Consequently, a large number of the generated faults scenarios involved cases where the

majority of the faults were undetectable since the probability that random fault is generated very close to the root is relatively high when N_1 is sufficiently large. Therefore, we decided to obtain additional results where we restricted N_1 to be chosen uniformly at random between 1 and 20 in order to generate more scenarios having topologically detectable faults at locations that are far from the root node. Figure 5.5 also includes two graphs representing the performance of our approach for the case where we rely only on measurements of real power flow ($N = 1$) for our fault detection and the case where we use both real and active power flow measurements ($N = 2$) under the assumption that the noise in the two measurements is iid. From Figure 5.5, it seems that the probability of detection for our approach remains unchanged when compared to the results from Figure 5.4 for small values of σ whereas it was slightly lower in Figure 5.5 for larger values of σ.

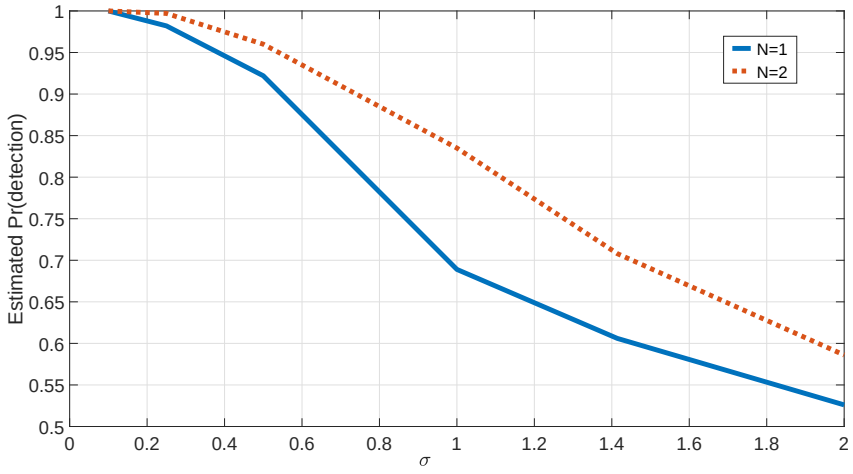


Figure 5.5: Estimated probability of detection vs increasing values of noise standard deviation with number faults at most 20

Finally, Figure 5.6 shows the effect of increasing the number of measurements collected by each sensor on our estimated probability of detection. The results for Figure 5.6 were obtained for the case where we assumed that $N = 1, 25, 50, 75, 100$ measurements were collected by real power flow measurements under fixed load forecasts having iid additive white Gaussian noise with $\sigma = 2$. The results in Figure 5.6 show that only a small number of measurements is needed in order to achieve significant increases in the probability of detection. Figure 5.6 also shows that employing more measurements will continue to improve the detection performance. This should be expected since having more measurements allows us to obtain more accurate estimates for the mean power flow along the lines we observe.

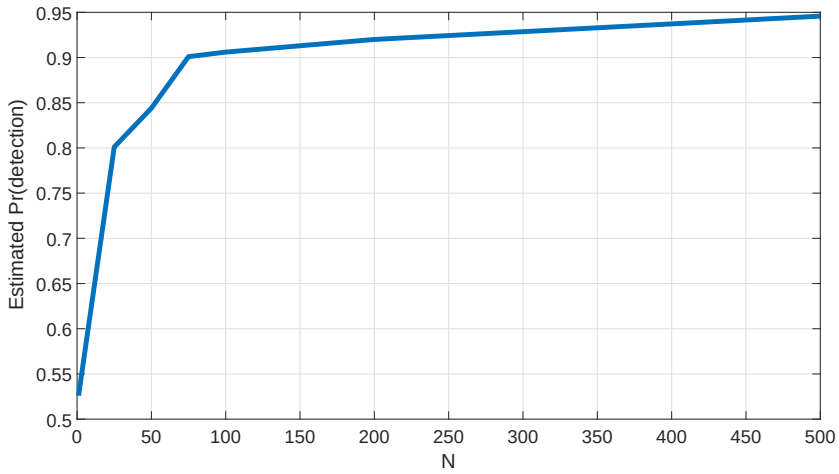


Figure 5.6: Estimated probability of detection vs increasing number of measurements per sensor

5.5 Conclusion

In this chapter, we focused on topology estimation and fault detection in radial distribution networks based on noisy nodal demand forecasts and power flow measurements collected from a subset of the lines in the network. We proposed a sensor placement scheme providing the minimum number of sensors required such that all the topologically detectable faults in the network are identifiable. Afterwards, we described a novel fault detection scheme taking advantage of the decoupling of the ML detector over subtrees in the network. Finally, we presented several numerical results where we employed our proposed sensor placement scheme along with the presented topology estimation approach to detect randomly generated faults in the IEEE 123-bus distribution feeder.

5.6 Appendix

Description of the Functions Employed by the Proposed Sensor Placement Scheme

Throughout this appendix, we take v_n to represent a node in the distribution network, $\mathbf{d}$ to represent a vector of load demands, and $\mathcal{T}$ to represent the topology of a distribution network.

1. **Child**$(v_n, \mathcal{T})$: this function returns a set containing all child nodes of v_n in the network $\mathcal{T}$.

2. **UpFlow**$(v_n, \mathbf{d}, \mathcal{T})$: The output of this function is a scalar quantity equal to the flow along line e_n where e_n is the node connecting v_n to its parent node. We may calculate the output

using equation (5.2).

3. **UpdateTopology**$(v_n, \mathcal{T})$: This function return the network $\mathcal{T}$ after removing the line e_n from it.

4. **Size**() : this function simply returns the size (or length) of the set or vector provided to it as an input.

5. **IsEmpty**$(\mathbf{x})$: this function return a logical true if all elements of the vector $\mathbf{x}$ are unique and returns a logical false otherwise.

6. **CombineVects**$(\mathbf{x}_1, \mathbf{x}_2)$: This function return a vector constructed from combining the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ as described by **Subroutine** 5.1

Algorithm 5.1: Placement$(v_n, \mathcal{P}, \mathbf{d}, \mathcal{T})$
A recursive sensor placement scheme

Input: Node v_n, old sensor placement set $\mathcal{P}$, load forecasts $\mathbf{d}$, and topology $\mathcal{T}$
Result: Updated set $\mathcal{P}$ contains the proposed sensor placement
Begin
$S =$ **Child**$(v_n, \mathcal{T})$
$s =$ **Size**(S)
if s == 0 **then**
    $flow\_vector =$ **UpFlow**$(v_n, \mathbf{d}, \mathcal{T})$
    return
else if s == 1 **then**
    $[f\_vector, \mathcal{P}, \mathcal{T}] =$ **Placement**$(S, \mathcal{P}, \mathbf{d}, \mathcal{T})$
    if $d[n] == 0$ **then**
        $flow\_vector = f\_vector$
    else
        $flow\_vector = (f\_vector,$ **UpFlow**$(v_n, \mathbf{d}, \mathcal{T}))^T$
    end if
    return
else
    //**Node** v_n **has multiple child nodes**
    $index = [\ ]$
    for $i = 1$ to s **do**
        $[f\_vect\_i, \mathcal{P}] =$ **Placement**$(S[i], \mathcal{P}, \mathbf{d}, \mathcal{T})$
        if IsEmpty$(f\_vect\_i) == 0$ **then**
            $index.append(i)$
        end if
    end for
    if $d[n] \neq 0$ **and Size**$(index) == 1$ **then**
        $flow\_vector = (f\_vect\_index,$ **UpFlow**$(v_n, \mathbf{d}, \mathcal{T}))$
    else if $d[n] == 0$ **and Size**$(index) == 1$ **then**
        $flow\_vector = (f\_vect\_index)$
    else
        $flow\_vector = [\ ]$
        for $j = 1$ to **Size**$(index)$ **do**
            $flow\_vector =$ **CombineVects**$(flow\_vector, f\_vect\_index[j])$
        end for
    end if
    //**Check if all values are unique**
    if unique$(flow\_vector) == 1$ **then**
        return
    else
        //**Place Sensor at** v_n
        $\mathcal{P} = \mathcal{P} \cup v_n$
        $flow\_vector = [\ ]$
        //**Update expected flow above sensor**
        $[\mathcal{T}] =$ **UpdateTopology**$(v_n, \mathcal{T})$
        return
    end if
end if

Subroutine 5.1: CombineVects$(\mathbf{x}_1, \mathbf{x}_2)$

1: **Input:** Vectors $\mathbf{x}_1$ and $\mathbf{x}_2$
2: **Result:** Vector $\mathbf{x}$
3: **Begin**
4: $n_1 =$ **Size**$(\mathbf{x}_1)$
5: $n_2 =$ **Size**$(\mathbf{x}_2)$
6:
7: **if** $n_1 == 0$ **then**
8: $\quad \mathbf{x} = \mathbf{x}_2$
9: $\quad$ **return**
10: **else if** $n_2 == 0$ **then**
11: $\quad \mathbf{x} = \mathbf{x}_1$
12: $\quad$ **return**
13: **end if**
14:
15: $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$
16: **for** $i = 1$ to n_1 **do**
17: $\quad$ **for** $j = 1$ to n_2 **do**
18: $\quad\quad \mathbf{x}.append(x_1[i] + x_2[j])$
19: $\quad$ **end for**
20: **end for**
21: **return**

Bibliography

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[2] W.-M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *Communications, IEEE Transactions on*, vol. 41, no. 11, pp. 1602–1605, 1993.

[3] J. A. Gubner, "Distributed estimation and quantization," *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1456–1459, 1993.

[4] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *Communications, IEEE Transactions on*, vol. 48, no. 2, pp. 181–184, 2000.

[5] K. Zhang and X. R. Li, "Optimal sensor data quantization for best linear unbiased estimation fusion," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 3. IEEE, 2004, pp. 2656–2661.

[6] P. Venkitasubramaniam, L. Tong, and A. Swami, "Minimax quantization for distributed maximum likelihood estimation," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 3. IEEE, 2006, pp. III–III.

[7] ——, "Quantization for maximin are in distributed estimation," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3596–3605, 2007.

[8] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part ii: Unknown probability density function," *Signal Processing, IEEE Transactions on*, vol. 54, no. 7, pp. 2784–2796, 2006.

[9] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed estimation in wireless sensor networks," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5682–5699, 2009.

[10] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Optimal distributed detection in the presence of Byzantines," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13), IEEE*.

[11] K. Agrawal, A. Vempaty, H. Chen, and P. K. Varshney, "Target localization in wireless sensor networks with quantized data in the presence of Byzantine attacks," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 1669–1673.

[12] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. Varshney, "Localization in wireless sensor networks: Byzantines and mitigation techniques," *Signal Processing, IEEE Transactions on*, vol. 61, no. 6, pp. 1495–1508, 2013.

[13] A. Vempaty, O. Ozdemir, and P. K. Varshney, "Mitigation of Byzantine attacks for target location estimation in wireless sensor networks," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*. IEEE, 2012, pp. 1–6.

[14] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 2, pp. 52–73, 2009.

[15] D. Liu, P. Ning, and W. K. Du, "Attack-resistant location estimation in sensor networks," in *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 2005, p. 13.

[16] X. Luo, M. Dong, and Y. Huang, "On distributed fault-tolerant detection in wireless sensor networks," *Computers, IEEE Transactions on*, vol. 55, no. 1, pp. 58–70, 2006.

[17] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attack in large wireless sensor networks," in *Military Communications Conference, 2006. MILCOM 2006. IEEE*. IEEE, 2006, pp. 1–4.

[18] ——, "Distributed detection in the presence of Byzantine attacks," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 16–29, 2009.

[19] A. Vempaty, K. Agrawal, H. Chen, and P. Varshney, "Adaptive learning of Byzantines' behavior in cooperative spectrum sensing," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*. IEEE, 2011, pp. 1310–1315.

[20] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of Byzantine attacks in cognitive radio networks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 774–786, 2011.

[21] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, pp. 382–401, 1982.

[22] V. Nadendla, Y. Han, and P. Varshney, "Distributed inference with m-ary quantized data in the presence of Byzantine attacks," *Signal Processing, IEEE Transactions on*, vol. 62, no. 10, pp. 2681–2695, May 2014.

[23] K. Budal and J. Falnes, "Optimum operation of improved wave-power converter," *Mar. Sci. Commun.;(United States)*, vol. 3, no. 2, 1977.

[24] J. Falnes, *Ocean waves and oscillating systems*. Cambridge University Press, 2004.

[25] S. Naito and S. Nakamura, "Wave energy absorption in irregular waves by feedforward control system," in *Hydrodynamics of ocean wave-energy utilization*. Springer, 1986, pp. 269–280.

[26] K. Budar and J. Falnes, "A resonant point absorber of ocean-wave power," *Nature*, vol. 256, p. 478, 1975.

[27] A. Babarit and A. Clément, "Optimal latching control of a wave energy device in regular and irregular waves," *Applied Ocean Research*, vol. 28, no. 2, pp. 77–91, 2006.

[28] F. Kara, "Time domain prediction of power absorption from ocean waves with latching control," *Renewable Energy*, vol. 35, no. 2, pp. 423–434, 2010.

[29] G. Bacelli, J.-C. Gilloteaux, and J. Ringwood, "A predictive controller for a heaving buoy producing potable water," in *Proceedings of the European Control Conference*, 2009, pp. 3755–3760.

[30] J. Cretel, A. Lewis, G. Lightbody, and G. Thomas, "An application of model predictive control to a wave energy point absorber," in *Control Methodologies and Technology for Energy Efficiency*, vol. 1, no. 1, 2010, pp. 267–272.

[31] J. A. Cretel, G. Lightbody, G. P. Thomas, and A. W. Lewis, "Maximisation of energy capture by a wave-energy point absorber using model predictive control," in *Proceedings of the 18th IFAC World Congress, Milano, Italy, Aug*, 2011, pp. 3714–3721.

[32] J. Hals, J. Falnes, and T. Moan, "A comparison of selected strategies for adaptive control of wave energy converters," *Journal of Offshore Mechanics and Arctic Engineering*, vol. 133, no. 3, p. 031101, 2011.

[33] T. Brekken, "On model predictive control for a point absorber wave energy converter," in *PowerTech, 2011 IEEE Trondheim*, June 2011, pp. 1–8.

[34] M. Richter, M. E. Magaña, O. Sawodny, and T. K. Brekken, "Nonlinear model predictive control of a point absorber wave energy converter," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 1, pp. 118–126, 2013.

[35] G. Li and M. R. Belmont, "Model predictive control of sea wave energy converters–part i: A convex approach for the case of a single device," *Renewable Energy*, vol. 69, pp. 453–463, 2014.

[36] J. V. Ringwood, G. Bacelli, and F. Fusco, "Energy-maximizing control of wave-energy converters: the development of control system technology to optimize their operation," *Control Systems, IEEE*, vol. 34, no. 5, pp. 30–55, 2014.

[37] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, Dec 2011.

[38] S. Buldyrev, R. Parshani, G. Paul, H. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks." *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.

[39] M. Parandehgheibi and E. Modiano, "Robustness of interdependent networks: The case of communication networks and the power grid," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 2164–2169.

[40] A. Berizzi, "The italian 2003 blackout," in *Power engineering society general meeting, 2004. IEEE.* IEEE, 2004, pp. 1673–1679.

[41] M. Parandehgheibi, E. Modiano, and D. Hay, "Mitigating cascading failures in interdependent power grids and communication networks," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2014, pp. 242–247.

[42] M. Rahnamay-Naeini and M. M. Hayat, "On the role of power-grid and communication-system interdependencies on cascading failures," in *2013 IEEE Global Conference on Signal and Information Processing*, Dec 2013, pp. 527–530.

[43] J. Shao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Cascade of failures in coupled network systems with multiple support-dependence relations," *Phys. Rev. E*, vol. 83, p. 036116, Mar 2011. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.83.036116

[44] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of interdependent networks under targeted attack," *Physical Review E*, vol. 83, no. 6, p. 065101, 2011.

[45] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, "Networks formed from interdependent networks," *Nature physics*, vol. 8, no. 1, pp. 40–48, 2012.

[46] A. Sen, A. Mazumder, J. Banerjee, A. Das, and R. Compton, "Identification of k most vulnerable nodes in multi-layered network using a new model of interdependency," in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2014, pp. 831–836.

[47] D. T. Nguyen, Y. Shen, and M. T. Thai, "Detecting critical nodes in interdependent power networks for vulnerability assessment," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 151–159, March 2013.

[48] T. Yong and R. H. Lasseter, "Stochastic optimal power flow: formulation and solution," in *2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134)*, vol. 1, 2000, pp. 237–242 vol. 1.

[49] H. Zhang and P. Li, "Chance constrained programming for optimal power flow under uncertainty," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2417–2424, Nov 2011.

[50] R. A. Jabr, "Adjustable robust opf with renewable energy sources," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4742–4751, Nov 2013.

[51] A. Monticelli, M. V. F. Pereira, and S. Granville, "Security-constrained optimal power flow with post-contingency corrective rescheduling," *IEEE Transactions on Power Systems*, vol. 2, no. 1, pp. 175–180, Feb 1987.

[52] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*. John Wiley & Sons, 2012.

[53] A. von Meier and G. D. Rodriguez, "Monitoring for impacts of distributed resources: Initial planning considerations," in *2013 IEEE Power Energy Society General Meeting*, July 2013, pp. 1–5.

[54] Z. Wang and J. Wang, "Self-healing resilient distribution systems based on sectionalization into microgrids," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3139–3149, Nov 2015.

[55] J. Medina, N. Muller, and I. Roytelman, "Demand response and distribution grid operations: Opportunities and challenges," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 193–198, Sept 2010.

[56] C. N. Lu, J. H. Teng, and W. H. E. Liu, "Distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 229–240, Feb 1995.

[57] R. Singh, B. C. Pal, and R. A. Jabr, "Choice of estimator for distribution system state estimation," *IET Generation, Transmission Distribution*, vol. 3, no. 7, pp. 666–678, July 2009.

[58] M. Baran and T. E. McDermott, "Distribution system state estimation using ami data," in *2009 IEEE/PES Power Systems Conference and Exposition*, March 2009, pp. 1–3.

[59] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1644–1652, Nov 2008.

[60] ——, "Double line outage detection using phasor angle measurements," in *2009 IEEE Power Energy Society General Meeting*, July 2009, pp. 1–5.

[61] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, Nov 2012.

[62] Y. C. Chen, T. Banerjee, A. D. Domnguez-Garca, and V. V. Veeravalli, "Quickest line outage detection and identification," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 749–758, Jan 2016.

[63] C. Fukui and J. Kawakami, "An expert system for fault section estimation using information from protective relays and circuit breakers," *IEEE Transactions on Power Delivery*, vol. 1, no. 4, pp. 83–90, Oct 1986.

[64] Y. Liu and N. N. Schulz, "Knowledge-based system for distribution system outage locating using comprehensive information," *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 451–456, May 2002.

[65] R. A. F. Pereira, L. G. W. da Silva, M. Kezunovic, and J. R. S. Mantovani, "Improved fault location on distribution feeders based on matching during-fault voltage sags," *IEEE Transactions on Power Delivery*, vol. 24, no. 2, pp. 852–862, April 2009.

[66] M. Kezunovic, "Smart fault location for smart grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 11–22, March 2011.

[67] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *ISGT 2014*, Feb 2014, pp. 1–5.

[68] D. Deka, M. Chertkov, and S. Backhaus, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[69] G. Cavraro, R. Arghandeh, G. Barchi, and A. von Meier, "Distribution network topology detection with time-series measurements," in *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2015, pp. 1–5.

[70] Y. Zhao, R. Sevlian, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection in power distribution networks with optimally-deployed power flow sensors," in *2013 IEEE Power Energy Society General Meeting*, July 2013, pp. 1–5.

[71] R. Sevlian, Y. Zhao, A. Goldsmith, R. Rajagopal, and H. V. Poor, "Outage detection in power distribution networks," *CoRR*, vol. abs/1503.08244, 2015. [Online]. Available: http://arxiv.org/abs/1503.08244

[72] J. Zhang, R. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *Signal Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[73] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Capkun, "On the requirements for successful gps spoofing attacks," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*. New York, NY, USA: ACM, 2011, pp. 75–86.

[74] S. B. Cambridge Univ. Press and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press,, 2004.

[75] B. Alnajjab, J. Zhang, and R. Blum, "Attacks on sensor network parameter estimation with quantization: Performance and asymptotically optimum processing," *Signal Processing, IEEE Transactions on*, vol. 63, no. 24, pp. 6659–6672, Dec 2015.

[76] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY: John Wiley and Sons, 1991.

[77] M. Belmont, "Increases in the average power output of wave energy converters using quiescent period predictive control," *Renewable Energy*, vol. 35, no. 12, pp. 2812–2820, 2010.

[78] G. Li, G. Weiss, M. Mueller, S. Townley, and M. R. Belmont, "Wave energy converter control by wave prediction and dynamic programming," *Renewable Energy*, vol. 48, pp. 392–403, 2012.

[79] K. Budal, J. Falnes, T. Hals, L. Iversen, and T. Onshus, "Model experiment with a phase controlled point absorber," in *Proceedings of Second International Symposium on Wave and Tidal Energy*, 1981, pp. 191–206.

[80] K. Budal, J. Falnes, L. C. Iversen, P. M. Lillebekken, G. Oltedal, T. Hals, T. Onshus, and A. Høy, "The norwegian wave-power buoy project," 1982.

[81] S. E. Sand, *Three-dimensional deterministic structure of ocean waves.* Institute of Hydrodynamics and Hydraulic Engineering, Technical University of Denmark, 1979.

[82] F. Fusco and J. V. Ringwood, "A study of the prediction requirements in real-time control of wave energy converters," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 1, pp. 176–184, 2012.

[83] ——, "Short-term wave forecasting for real-time control of wave energy converters," *Sustainable Energy, IEEE Transactions on*, vol. 1, no. 2, pp. 99–106, 2010.

[84] M. Schoen, J. Hals, and T. Moan, "Wave prediction and robust control of heaving wave energy devices for irregular waves," *Energy Conversion, IEEE Transactions on*, vol. 26, no. 2, pp. 627–638, June 2011.

[85] M. Belmont, J. Horwood, R. Thurley, and J. Baker, "Filters for linear sea-wave prediction," *Ocean Engineering*, vol. 33, no. 17, pp. 2332–2351, 2006.

[86] J. Tedd and P. Frigaard, "Short term wave forecasting, using digital filters, for improved control of wave energy converters," in *Proc. of Int. Offshore and Polar Eng.(ISOPE) Conf*, vol. 388, 2007, p. 394.

[87] P. Stoica and R. Moses, *Spectral Analysis of Signals (see pages 188-194).* Upper Saddle River, New Jersey: Prentice Hall, 2005.

[88] B. Alnajjab and R. S. Blum, "Optimal design of sensor networks for enhanced ocean wave energy conversion," in *Signals, Systems and Computers, 2013 Asilomar Conference on*, Nov 2013, pp. 384–388.

[89] ——, "Ocean wave prediction from noisy sensor measurements," in *Global Marine Renewable Energy Conference*, April 2013. [Online]. Available: http://www.foroceanenergy.org/wp-content/uploads/2013/07/OCEAN-WAVE-PREDICTION-FROM-NOISY-SENSOR-MEASUREMENTS.pdf

[90] C. Radhakrishna Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.

[91] H. Van Trees, *Detection, Estimation, and Modulation Theory.* Wiley-Interscience, 2001, vol. 1.

[92] S. M. Kay., *Fundementals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, New Jersey: Prentice Hall, 1993.

[93] P. Stoica and R. L. Moses, *Spectral analysis of signals.* Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[94] D. Evans, "Some analytic results for two and three dimensional wave-energy absorbers," *Power from sea waves*, pp. 213–249, 1980.

[95] J. Falnes, "Radiation impedance matrix and optimum power absorption for interacting oscillators in surface waves," *Applied Ocean Research*, vol. 2, no. 2, pp. 75–80, 1980.

[96] H. V. Poor, *An introduction to signal detection and estimation.* Springer, 1994.

[97] C. Lee and J. Newman, "Wamit® user manual, version 7. pc, wamit," *Inc., Chestnut Hill, MA, USA*, 2013.

[98] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*. John Wiley & Sons, 2012.

[99] P. Hahn and M. Jeruchim, "Developments in the theory and application of importance sampling," *IEEE transactions on Communications*, vol. 35, no. 7, pp. 706–714, 1987.

[100] M. Jeruchim, "On the application of importance sampling to the simulation of digital satellite and multihop links," *IEEE Transactions on Communications*, vol. 32, no. 10, pp. 1088–1092, 1984.

[101] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: a review of importance sampling techniques in communications systems," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 4, pp. 597–613, May 1997.

[102] G. Infanger, "Monte carlo (importance) sampling within a benders decomposition algorithm for stochastic linear programs," *Annals of Operations Research*, vol. 39, no. 1, pp. 69–95, 1992.

Vita

Basel Alnajjab was born on June 22, 1990 to Maha Alsaheb and Rashid Alnajjab. He was raised in Amman, Jordan with his older siblings Muhannad and Hana. He attended The National Orthodox School and graduated in 2008. Basel then enrolled at Lehigh University in the P.C. Rossin College of Engineering, and he later graduated with a Bachelor of Science in Electrical Engineering in May of 2012. As an undergraduate, he won the IEEE Walter B. Morton Contest and was awarded the Elisha P. Wilbur Mathematics Prize, the Philip Francis du Pont Memorial Prize in Electrical Engineering, and Dean of Student's Development of Student Life Award. In the summer of 2012, Basel joined Ocean Power Technologies in Pennington, New Jersey as an Electrical Engineering Intern where he worked on designing components and systems for an ocean wave energy converter. Afterwards, Basel joined Dr. Rick S. Blum's Signal Processing and Communication Research Lab (SPCRL) as a graduate student and a Research Assistant. During his time with SPRCL, Basel worked on ocean wave estimation, attacks on sensors networks, and the interdependence between power grids and their supporting communication networks. He was awarded a Master of Science Degree in Electrical Engineering in January of 2015. In the Fall of 2016, Basel joined Argonne National Laboratory in Illinois as a Research Aide in the Center for Energy, Environmental, and Economic Systems Analysis.