The University of Sydney
The School of Aerospace, Mechanical and Mechatronic Engineering

Nonlinear State Estimation Applications within Defence

A thesis submitted in fulfilment of the requirements for the degree of Masters of Philosophy

Jonathan Mitchell

2018

Abstract

State estimation has become an indispensable tool for both Engineering and Science, providing the capability to estimate internal states of a process that cannot be directly observed. Given the ubiquity of nonlinear dynamic systems, nonlinear state estimation continues to be an important research field with application to a wide variety of industries.

This thesis explores the application of nonlinear state estimation within a Defence context. Motivated by the 2016 Australian Defence White Paper, the application of nonlinear state estimation methods contribute to the key capability streams which will enable the Australian Defence Force (ADF) to achieve their Strategic Defence Objectives. Advances in nonlinear state estimation and the enabling technologies will improve mission effectiveness and lead to a more capable, agile and potent ADF.

Two application areas are explored. First, a nonlinear state estimator was used to develop a fault detection system for a military vehicle switched battery system. Second, a nonlinear state estimator was used to implement a target tracking algorithm for a bearing only buoy field sensor network.

The fault detection system for a military vehicle switched battery system employed and compared two nonlinear state estimation methods, an Extended Kalman Filter (EKF) and a Moving Horizon Estimator (MHE). Conclusions regarding the fault detection system developed are drawn and recommendations for future work outlined. Additionally, an overview of the benefits which can be realised through the application of Fault Detection and Isolation (FDI) systems to military platforms is also given.

The target tracking algorithm for a bearing only buoy field sensor network implemented and compared two nonlinear state estimation methods, an EKF and an Unscented Kalman Filter (UKF). A dynamic simulation is used to analyse the estimator performance with varying buoy field geometries. Finally, optimisation of two buoy fields is conducted which are shown to provide superior target tracking performance when compared with typical buoy field geometries. Conclusions regarding the buoy field geometry and estimator performance are drawn and recommendations for future work outlined.

Acknowledgements

This thesis has been both an engaging and challenging journey filled with a rich diversity of experiences and knowledge gained. It has given me an understanding and appreciation of the challenges involved with academia and industry collaboration; whilst reinforcing the immense benefits, both tangible and intangible, that come from such collaborations. The completion of this journey was enabled thanks to the support of the people around me and the encouragement of my friends and family.

First I'd like to express my gratitude for my supervisor, Professor Ian Manchester. His guidance, tuition and advice have enabled me to conduct and complete my research. He is a source of inspiration with an enthusiasm and passion for his fields of research, and importantly an open mind. It is these qualities which I attribute as essential for successful research and innovation. I also appreciate the many discussions that we had over the years that went beyond just theory; sharing your experiences and insight into industry and academic collaboration, and the academic environment as a whole.

With my Masters focused on nonlinear state estimation applications in Defence, I'd like thank my Thales colleagues who supported or provided input into the problems researched. Bruno Lambla, Paul Grocutt, Michael Clark, Patrick Cooper and Nicolas Josso, I am very grateful for the support and contributions made which for many were secondary to your main responsibilities.

I'd also like to thank my employer Thales Australia for providing the flexibility in working arrangement conducive for me to complete this Masters of Philosophy part-time. Specifically, I'd like to thank the Thales Training and Simulation leadership team, Thomas Birouste, Troy Stephens, Phil Swadling and Peter Newnham.

A deep gratitude goes to the Wedgetail Operational Flight Trainer project team who afforded me both patience and understanding as to why I was never in the office on Tuesdays. Craig Shepherd, Greg Nangle, Shani de Leeuw, Chrishan Fernando, Matthew Scott, William Cooper, Paul Chambers, Gerard Cathcart, Nicolas Houston, Kelvin Burchell, Kevin Wood and Mike Rice, thank you for accommodating this aspiration of mine.

Thank you to my parents, Colin and Jenny, and my sister Ellyse who have supported me remotely from Victoria. They have provided a continual stream of encouragement and support, and an empathetic ear when required.

Finally, and most importantly, I'd like to thank my wife Sarah. When I started on this path four years ago I couldn't comprehend the length and difficulty I'd face on this journey, not just with the Masters but all of life's challenges along the way. You have provided a foundation of support and stability in my life that I deeply appreciate. Thank you for being patient and understanding of the significant demand this took on my time. You mean the world to me and I look forward to the future we share.

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Jonathan Mitchell

Contents

Chapter 1

Introduction

This thesis is concerned with the application of nonlinear state estimation in Defence. Advances in nonlinear state estimation provide an important technological capability that stands to deliver solutions critical to enabling the key capability streams that will ensure the Australian Defence Force (ADF) can achieve their Strategic Defence Objectives. This thesis explores nonlinear state estimation and two applications related to these Defence capability streams, first nonlinear state estimation for a fault detection system, and secondly nonlinear state estimation for a target tracking system.

Chapter 1 provides an introduction to this thesis, giving a motivation for nonlinear state estimation applied in Defence; an overview on the problem statement nonlinear state estimation looks to address; a history on the development of state estimation and leading methods in the literature; and finally a summary of the key contributions made in this thesis.

Chapter 2 provides a comprehensive background on Fault Detection and Isolation (FDI), the problem statement FDI looks to address, and the leading model-based methods used in the literature.

Chapter 3 constitutes the main body of work in this thesis with respect to nonlinear state estimation for fault detection; with the development and evaluation of a fault detection system for a military vehicle battery system. Two methods are explored and compared to develop this fault detection system, an Extended Kalman Filter (EKF) approach and a Moving Horizon Estimator (MHE) approach.

Chapter 4 explores the implementation and evaluation of a nonlinear state estimator for target tracking using a bearing only buoy field sensor network. Two nonlinear state estimation approaches are implemented and compared, an EKF and an Unscented Kalman Filter (UKF).

Chapter 5 investigates the development and analysis of an optimal bearing only buoy field deployment for target tracking. A genetic algorithm is employed to optimise buoy field deployments against two objective functions, with the resulting deployments compared against some typical geometries.

Chapter 6 provides the concluding remarks of this thesis, summarising the key results and providing reference to the various sections where they are derived.

1.1 Motivation for Nonlinear State Estimation in Defence

The Commonwealth of Australia (CoA) recognises, as identified in the 2016 Defence White Paper [1], that Australia is facing greater security uncertainty and complexity globally. The Indo-Pacific region is in a period of significant economic transformation which will account for almost half the world's economic output by 2050. This presents an opportunity for Australia's economy and security capabilities to grow in step as the Indo-Pacific increases in both economic and strategic weight.

The security uncertainty and complexity faced by Australia is a culmination of several factors. The increasing importance of the Indo-Pacific region, military modernisation in the local region, territorial disputes between claimants in the East China and South China Seas creating tension in our region, the Democratic People's Republic of Korea continuing as a source of instability, and the rise of Daesh terrorists in the Middle East with incidents across the world demonstrating the pervasive nature and threat of terrorism.

To address these challenges and ensure continued peace and stability in the Indo-Pacific region the Australian Defence Force (ADF) needs to be more capable, agile and potent [1]. To support a more capable, agile and potent force, the CoA have recognised high defence preparedness as one of the underpinning capabilities; This capability is required to support increased ADF activity in the region while maintaining the ADF's ability to make meaningful contributions to global security operations where our interests are engaged.

The CoA also recognises technology and innovation as a core driver to maintaining their capability edge. The following six capability streams have been identified which will deliver a potent high-technology force that can achieve the Strategic Defence Objectives.

1. Intelligence, Surveillance, Reconnaissance, Space, Electronic Warfare and Cyber Capabilities to ensure superior Situational Awareness.
2. The Maritime and Anti-Submarine Warfare capabilities to enable operations in challenging maritime threat environments
3. The Strike and Air Combat capabilities providing forces with greater flexibility in responding to threats
4. The Land Combat and Amphibious Warfare capabilities to provide greater capacity for both combat and non-combat operations
5. Key enablers to support the operation and sustainment of Defence
6. The Air and Sea Lift capabilities to overcome the vast distances which the ADF is deployed and to be supplied.

Advances in nonlinear state estimation potentially stand to benefit all six capability streams. The benefits are typically recognised through the application of nonlinear control of systems which rely on accurate state information. However, state estimation methods provide further benefit to any problem where accurate state information is required but may not be directly observable and has to be inferred through noisy measurements. This application of nonlinear state estimation directly benefits capability streams one and five. These applications, which are explored further in this thesis are Fault Detection and Isolation (FDI) (Stream: Key enablers to support the operation and sustainment of Defence), and target tracking (Stream: Intelligence, Surveillance, Reconnaissance, Space, Electronic Warfare and Cyber Capabilities to ensure superior Situational Awareness). The application of nonlinear state estimation and how it benefits the other four capability streams is not explored in this Thesis.

FDI, employed in Health Usage and Monitoring Systems (HUMSs), utilises state estimation to provide fault diagnosis in dynamic processes. This capability acts as a key enabler for improving the support and sustainment of Defence assets leading to increased platform availability and reductions in Through Life Support (TLS) costs. The adoption of HUMS has been actively pursued within a military context. The UK Ministry of Defence (MoD) have recognised HUMS as an important addition to their asset maintenance policy [2]. By incorporating FDI methodologies, the UK MoD seeks to improve reliability and reduce cost of ownership by detection and diagnosis of potential and actual failures, monitoring usage and providing alerts for maintenance actions.

These capabilities facilitate a shift towards a Condition Based Maintenance (CBM) paradigm, allowing the optimisation of platform servicing. This results in reduced TLS costs and increased platform availability.

The adoption of this approach was championed by the GenHUMS programme for the UK MoD's Chinook assets. A review of the GenHUMS program was conducted in 2002 and identified the benefits and savings realised by the program [3]. FDI is recognised as a direct enabler of CBM and there are increasing opportunities to apply FDI systems cost effectively. One of the leading methods for FDI systems is enabled by state estimation theory, and thus advances in nonlinear state estimation theory naturally leads to more capable FDI systems. FDI is a promising field that has already produced proven results in its numerous applications in aviation, industrial processes and the automotive industry and delivers improved platform safety, reductions in TLS costs and increased platform availability. With the adoption of FDI in Defence, empowered by nonlinear state estimations, the CoA can recognise similar benefits.

The second application explored in this thesis is target tracking. Target tracking utilises state estimation, fusing data from multiple noisy sensors and a maneuvering model to predict and refine estimates on target motion. The benefits in the application of state estimation are significant with respect to a Defence environment given the need for Intelligence Surveillance and Reconnaissance (ISR), which is improved by accurate target tracking leading to superior situational awareness. Similarly, state estimation can be used to estimate characteristics of the target and lead to improved threat classification. Besides its usage in ISR, target tracking has diverse applications including use in computer vision, autonomous vehicles, biomedical research, oceanography, satellites [4] and generally any application where localisation or tracking is important.

Both these use cases benefit from nonlinear state estimation methods. The continuing miniaturisation of electronics and the increase in computational power embedded in platforms enables advanced algorithms for nonlinear state estimation that were previously not practical. Combined with continued research in optimal nonlinear state estimation, improved methods for nonlinear state estimation are increasingly applicable to better enhance the mission effectiveness of Defence.

1.2 Overview of State Estimation

1.2.1 Problem Definition

State estimation has become an indispensable tool for both Engineering and Science. More generally, any field interested in mathematical modelling of processes benefits from state estimation. At its core, state estimation provides the capability to estimate internal states of a process that cannot be directly observed by using a model of the physical process and observed outputs.

More formally, we can define a model of a nonlinear process as follows [5]:

$$\begin{aligned} x_{t+1}|x_t &\sim f(x_{t+1}|x_t, u_t) \\ y_t|x_t &\sim h(y_t|x_t, u_t) \end{aligned} \tag{1.2.1}$$

Where x_t is the process state at time t, $f(\cdot)$ is a stochastic process describing the process model, y_t is the measurement at time t, $h(\cdot)$ is a stochastic process describing the measurement model.

This representation is usually simplified in the literature with the assumptions of a discrete nonlinear deterministic model with additive noise:

$$
\begin{aligned}
x_{t+1} &= f(x_t, u_t) + w_t \\
y &= h(x_t, u_t) + v_t
\end{aligned} \tag{1.2.2}
$$

where $f(\cdot)$ is a nonlinear function describing the process model, $h(\cdot)$ is a nonlinear function describing the measurement model, $w_t \sim \varphi(0, \sigma_Q^2)$ and $v_t \sim \varphi(0, \sigma_R^2)$ are zero mean random variables for the process noise and sensor noise respectively.

State estimation is concerned with inferring information about the internal state of a process x_t based on available measurements $y_{1:T}$. There are three classifications of state estimation based on the relative time at which you are trying to infer information about state x_t versus available measurements $y_{1:T}$. These classifications are:

Table 1.1: State Inference Classification

Class	**Probability Density Function**
Filtering	$p(x_t \vert y_{1:t})$
Prediction	$p(x_{t+k} \vert y_{1:t}), k > 0$
Smoothing	$p(x_t \vert y_{1:T}), t < T$

The key challenge in nonlinear state estimation is in determining the Probability Density Functions (PDFs) given in Table 1.1 for a system. For nonlinear systems these calculations are intractable and current research is concerned with how to overcome this intractability.

1.2.2 History of State Estimation

The earliest known formal methods for estimation were undertaken by Gauss and Legendre whilst studying astronomical phenomena.[1] They were attempting to estimate the position of planets and comets using telescopic measurements, with Legendre first publishing the least squares method in 1805. Gauss claimed to have independently derived the least squares method in 1795 at age 18, but didn't publish these results until 1809. This caused dispute between Legendre and Gauss as to the inventor of the least squares method and has been the subject of much analysis and debate [8].

The theory of state estimation first began to coalesce in the 1940's with the pioneering work from Norbert Wiener. Weiner was looking to address the problem of extracting a signal corrupted by noise. His approach utilised the frequency domain and assumed a stationary stochastic process. This work was developed as part of the World War II effort of the United States and published in 1942 but only declassified in 1949 and released to the public in his seminal text "Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications" [9]. Independently, Andrey Kolmogorov studied the same problem and made similar discoveries on methods for extracting a signal corrupted by noise and predates Wieners work by a year (1941) [10]. However, Kolmogorov's work did not become well known until later since it was published in Russian, reducing its accessibility by western audiences. Both solutions are based on restrictive assumptions, assuming access to infinite amounts of data and that the signals involved can be described by stationary stochastic processes. During the 1950's research was directed at trying to relax these restrictive assumptions of the Wiener-Kolmogorov filtering theory [11–14].

With the emergence of State-Space Representation in the automatic control community in the late 1940's and 1950's [15], Rudolf Kalman had the idea to recast the Wiener-Kolmogorov Filtering Theory in state-space form. This transformed the filter into a much more natural representation,

[1] For a further account on the History of State estimation refer to Simon [6]. For a treatise on the history of the Kalman Filter and its application in aerospace refer to Grewal et al. [7].

simplifying the statistical calculations needed and its use for multivariable problems by being posed in the time domain as opposed to frequency domain. With the Wiener-Kolmogorov filter recasted using state-space form, with the assumption the state-space is finite dimensional, it was recognised by Richard Bucy that this equation is equivalent to a nonlinear Ordinary Differential Equation (ODE) studied by Jacopo Riccati [7]. Riccati had proven that ODEs of this form, now called the Riccati equation, can be transformed into a series of linear equations.

Using the solution to the Riccati equation, Kalman and Bucy arrived at an optimal filter for a linear model assuming Gaussian noise that minimises the variance of the estimation error between the estimated state $\hat{x}$ and the true state x. Kalman transformed this result from continuous time into discrete time and published this much celebrated result, what is now known as the Kalman Filter, in his seminal paper "A New Approach to Linear Filtering and Prediction Problems" in 1960 [16]. A year later, Kalman and Bucy published the analogous filter in continuous time [17].

The utility of the Kalman Filter was cemented through its use in the navigational system for the Apollo missions to send man to the moon [7]. It solved the problem of trajectory estimation and by 1961 NASA had implemented the Kalman Filter and began validating its performance using computer simulations. It was through this initial application with NASA that the development of the Extended Kalman Filter (EKF) originated, extending the applicability of the Kalman Filter to nonlinear models [18]. This was a necessary evolution of the method as linear systems simply don't exist in nature. However, many systems can adequately be approximated by linear systems which was achieved in the EKF by linearising the problem about the estimation point. It was shown through much analysis that using an EKF resulted in excellent performance for estimating trajectories, given accurate models and understanding of the variances of the subsequent noise.

The application of the EKF to the navigational system for the Apollo missions highlighted key challenges in the application of the EKF to nonlinear systems. Successful implementation of the EKF for nonlinear systems required a thorough understanding of the stochastic properties of the process and measurement noise and that the process and measurement models be linearisable. Model design for nonlinear processes to be used in an EKF were difficult if the nonlinearities weren't accommodating to linearisation about the estimation point. Similarly, the determination of the stochastic properties for processes is challenging and requires iterative tuning and validation to have confidence in filter performance [7].

Nonetheless, the success of the Kalman Filter as applied in aerospace soon led to its application within industrial automation. The Kalman Filter was looked towards to improve the control and monitoring of industrial processes to bring about a competitive commercial edge. However the pursuit of identifying models and the characteristics of the stochastic properties for industrial processes for use in the Kalman Filter was expensive [6]. Existing models lacked the requisite accuracy for the Kalman Filter to provide the benefits seen in the aerospace industry. The need emerged for a new class of filter that relaxed the requirement of accurate models and an understanding of the stochastic nature of the involved processes. State estimators that tolerate these higher uncertainties are called "Robust". Research in this area to develop such a filter led to what is known as the H_∞ Filter.

The H_∞ Filter was designed for robustness and takes the approach that it does not make any assumptions about the statistics of the measurement or process noise. It instead minimises the worst case estimation error while maintaining stability, as opposed to minimising the variance of the estimation error. This resulted in a conservative filter that is more tolerant of uncertainty. The H_∞ concept was originally formulated by George Zames in 1981 with the release of his paper "Feedback and optimal sensitivity: model reference transformation, multiplicative semiforms and approximate inverse" [19]. The H_∞ Filter was further developed [20–23] and by the early 1990's the general standard approach accepted today was common in the literature. Solving what it set out to do, the H_∞ Filter provides a rigorous method for dealing with model uncertainty and provides a natural way to limit the frequency response of the estimator. Similar to the Kalman Filter, the H_∞ Filter is sensitive to design parameters and requires tuning for the application. It is also subject to the same challenges that arise with nonlinearities, and likewise there are pursuits to extended H_∞ Filters to account for nonlinear models [24, 25].

Given the well developed theory behind optimal linear filters, it's an unfortunate reality that linear systems simply don't exist in nature. However, many systems can be approximated very well as a linear system about an operating point. Such results have been demonstrated repeatedly in the literature. Nevertheless, most cases encountered in practice are of a nonlinear nature and are poorly approximated as a linear system. Thus the motivation to pursue an optimal filter for nonlinear systems persists.

Given the general description of a nonlinear system by Equation 1.2.1, it has been shown that the solution for an optimal filter is analytically intractable due to the multidimensional integrals involved [26, 27]. Only in certain special cases can an analytical solution be found. One such special case is when the dynamic model is linear and the stochastic properties are Gaussian, the solution for which is the Kalman Filter. Given it is intractable to derive an analytical solution for the general nonlinear stochastic case, we are forced to instead find methods that approximate the solution.

Revisiting the EKF, it has limitations in that it is only a local approximation of an optimal nonlinear filter about an operating point. This method works well if the nonlinear systems are well approximated locally and there is a thorough understanding of the stochastic properties of the system. However, typically this isn't always the case. There have been many developments in the literature to address the limitations of the EKF and reduce the errors introduced by a local linearisation for highly nonlinear systems. These methods include the Iterated EKF and the Second Order EKF [6, 28].

Given the limitations local approximation methods face in highly nonlinear systems, it is desirable to seek a global approximation approach instead for such problems. In the literature there are many method towards global approximation of an optimal filter. The following approaches are those that have become notably wide spread.

An early approach (1972) for a global optimal filter of a nonlinear stochastic system is the Gaussian Sum Filter, conceived by Daniel Alspach and Hardold Sorenson [29]. They looked at constructing an optimal filter for a nonlinear system with non-Gaussian stochastic properties. The key principle of their approach is that a non-Gaussian PDF can be approximated by a sum of Gaussian PDFs.

The 1990's saw the rise of two methods, the Unscented Kalman Filter (UKF) and the Particle Filter for nonlinear stochastic filtering that have since become widespread in both literature and application. The UKF, introduced by Julier et al. [30,31] in 1995, is a modification to the Kalman Filter where using a fixed number of parameters it attempts to approximate the mean and covariance of the system states as it propagates through a nonlinear transform. It achieves this utilising a set of sigma points in state-space, chosen by the unscented transform, that approximate the system state PDF. This approximated PDF is then transformed through the nonlinear dynamics. The key underlying principle of the UKF is that it is easier to approximate a PDF than it is to approximate an arbitrary nonlinear function [32]. The UKF is an improvement on the EKF, but is still subject to issues when dealing with highly nonlinear systems with non-Gaussian noise [33,34]

The Particle Filter, first introduced by Gordon et al. [35] in 1993, provides a general procedure for solving nonlinear stochastic estimation problems. The key principle behind the approach is to use a number of random samples to approximate the system state PDF. This approximation is constructed initially using *a priori* information and then propagated through the nonlinear transform and resampled based on its fit against the observed system measurements. It has been shown that as the number of particles in a Particle Filter tend towards infinity, the Particle Filter estimate converges to the true optimal state estimate [36]. The Particle Filter does require significant computational power due to the Sequential Monte Carlo method used to approximate the key integrals fundamental to the solution of the optimal nonlinear state estimation equations. The Particle Filter is a phenomenal result, but is also subject to some issues such as path degeneracy from resampling and intractable computational demands when problems are sufficiently high-dimensional spaces with complex distributions [37]. Despite these issues, the Particle Filter remains a strong contender for a general optimal nonlinear state estimator and given the ubiquity of computational resources and the vast amount of continuing research in the field, the Particle Filter looks to be a promising solution [5, 36, 38].

The 90's also saw another disparate method, the Moving Horizon Estimator (MHE), for nonlinear state estimation become widespread. The MHE is the dual of Model Predictive Control (MPC), and uses the same underpinning approach of a sliding window of measurements to optimise a cost function derived from both the state model and measurement model of the nonlinear system [39]. This approach is different from the aforementioned optimal filter approaches in that MHE assumes a deterministic system model and doesn't incorporate any stochastic properties. MHE instead provides weighting in the cost function between the arrival cost (how close the initial condition for the sliding window matches the previous iteration), state model cost (how closely the state matches the state transition model) and measurement model cost (how closely the state matches the measurements). Natural to the MHE approach is the concept of constrained state estimation which lends itself for use in FDI. It is also interesting to note that for an unconstrained MHE with a window size $N = 1$ applied to a linear Gaussian system you can derive the Kalman Filter [40]. In general MHE is considered sub-optimal, since influences from prior measurements may not necessarily be adequately handled correctly by the arrival cost [26]. MHE is subject to the same issues of nonlinear optimisation and the search for a global optimum, with such methods currently being computationally intractable given moderate high dimensionality.

State estimation is a tool of great significance for Engineering and Science. It provides the capability to estimate internal states of a process which cannot be directly observed. Its application and utility is diverse from localisation, tracking, fault monitoring, or providing improved state estimation for nonlinear control. Given its importance, there is much continued research into optimal state estimation for nonlinear stochastic systems with the literature focused on addressing the aforementioned limitations with the current widespread methods and the pursuit of a general optimal nonlinear estimator.

For a listing of the nonlinear state estimation equations used in this thesis, please refer to Appendix A.

1.3 Contribution

This thesis explores two application areas for nonlinear state estimation, fault detection for a military vehicle battery system and target tracking using a bearing only buoy field. Subsequently, the contributions of this thesis are split between these two application areas and are as follows:

Fault detection for a military vehicle battery system

- Development of a test bench to generate data to explore fault detection for a military vehicle switched battery system
- Development of a diagnostic observer for a military vehicle switched battery system, including a comparison between an EKF and a MHE based approach.

These results are summarised in Section 3.6 and are presented in a conference paper by the Author [41]. Additionally, resultant from this work a report was written and provided to Thales Australia with findings and recommendations on fault detection in a switched battery system and more generally opportunities for FDI application within military vehicles.

Target tracking using a bearing only buoy field

- Development of a Simulation to investigate a dynamic case for target tracking using a bearing only buoy field sensor network.
- Implementation of a nonlinear state estimator for target tracking, including a comparison between an EKF approach and a UKF approach.
- Optimisation of two buoy fields shown to provide superior target tracking performance, inline with the design intent given by the objective functions, when compared against typical buoy field geometries.

These results are summarised in Section 5.6.

Chapter 2

Background on Fault Detection and Isolation

This chapter of the thesis provides the requisite background information on Fault Detection and Isolation (FDI). The field of FDI is extensive, with applications to industrial process, automotive, aerospace and other industries that either have mission critical aspects where failures result in catastrophe, or benefit through improved system availability and the cost reductions in Through Life Support (TLS) that FDI systems afford.

This chapter is broken down into the following sections: Section 2.1 provides a brief chronological history on the development of FDI systems, with a particular focus on model-based methods. Section 2.2 provides a formalisation of the general problem definition FDI systems try to solve. Section 2.3 provides an overview of model-based fault detection methods. Section 2.4 provides an overview of fault isolation methods.

2.1 History of Fault Detection and Isolation

FDI in dynamic systems is a field that is motivated by attractive practical benefits. Enabled by the advent of electronics, computing and system modelling theory, FDI stands to reduce the risk of adverse failure modes through early detection and mitigation. Additionally, FDI enables statistical inferences to be made regarding the life of components within dynamic processes. These two core outcomes reduce TLS costs and reduce risk of catastrophic failures providing strong incentives for application within industry.

More generally, FDI fits under the umbrella of Fault Prognosis. Fault Prognosis can be distilled into three layers; a low level Fault Detection and Isolation (FDI) layer; a middle level fault prediction layer; and a high level causal fault analysis layer. Both the middle and high level layers are enabled by the low level FDI layer.

FDI has seen continued focus and development in the control theory literature and through specialised symposiums such as the International Federation of Automatic Control (IFAC) symposium SAFEPROCESS. FDI systems can best be described as the early discovery (detection) and localisation (isolation) of a fault in a system; where a fault is defined as an unpermitted deviation of a characteristic parameter or property of a system.

Fault detection has its roots in the architecture of fault tolerant systems. Fault tolerant designs gave rise to the development of triplication and voting (Triple Modular Redundancy, circa 1960s), an approach related to FDI that relies on hardware redundancy [42]. Here the use of hardware redundancy provides a mechanism to compare nominal behaviour against potential faulty behaviour. Signals from the hardware redundant system are compared with each other and any one erroneous signal is eliminated by consensus from the functioning hardware. This method of fault detection is highly prevalent in the aviation industry where high levels of hardware redundancy already exist to provide robustness against system failures that could lead to catastrophe. However, this method is not without its limitations; Discrepancies between like instruments, inability to detect failures that affect both instruments in the same way, and inability to detect further faults once a single instrument has failed are some of the limitations of voting systems [43].

A natural progression from fault tolerant systems was the pursuit of FDI systems. As it is not always practical to rely on or implement hardware redundancy, analytical redundancy is instead utilised through model-based FDI approaches.

Development in the field of model-based FDI first began in the early 1970's. Early contributors are Beard [44] and Jones [45] for their developments in observer-based fault detection in linear systems. Through the use of observers they developed a failure-sensitive filter that could be used to construct a residual signal capable of detecting faults in dynamic processes. This approach received much focus in the literature due to its similarity with state estimators. These methods, in the FDI literature, became known as diagnostic observers and are still widely researched with examples of successful extension to nonlinear systems and applications within the automotive industry [46] and other industries.

Another method developed involved hypothesis testing of a multi-filter array of nominal and faulty behaviour. Here the development of several filters each describing a behaviour case are tested against the measured system. The filter most likely describing the current system process determines the likely state of the system. If the filter is found not to be the nominal case, a fault has occurred. A summary of these early developments is given by Willsky [43].

Fault detection through analytical redundancy became an attractive prospect for dynamic processes that featured automatic control as these processes were already well described mathematically. One of the first books on model-based FDI was written in 1978 with its application to chemical processes in mind [47]. In this work, Himmelblau laid the foundation for what led to the parity relation method.

The parity relation method was formalised by Chow and Willsky [48] in 1984 as an approach for robust residual generation for a linear dynamic system by extracting linearly independent parity relationships from the dynamic model. These relationships are checked for consistency against the observed output of the system. As such, these relationships form the residuals used for fault detection. Chow and Willsky give an exposition on developing parity relationships for robust residual generation in the presence of parameter uncertainty and noise. They adopt an approach of minimizing the sensitivity of the FDI system to uncertainty by building in the notion of parameter uncertainty and noise into the dynamic model. The parity relations method was further developed by Gertler [49] in 1997 who also gives his own account of the history and development of parity relations.

Interestingly the parity relation method can be transformed into the diagnostic observer form and is therefore structurally equivalent although the design process differs [50]. Gertler [51] first demonstrated that for every diagnostic observer there is an equivalent set of parity equations. This provides the flexibility to use diagnostic observer or parity relations methods dependant on the application suitability, with insights gained in one method informing the other indirectly.

The FDI literature has also seen a lot of focus on parameter estimation methods for fault detection [52–54] with early development of the method for fault detection attributed to Himmelblau [47]. These methods differ from diagnostic observers or parity relation methods in that they identify the system parameters and monitor these values for unpermitted deviations in the system parameters which are indicative of a fault. Parameter estimation methods for FDI leverage system identification techniques. For a treatise on system identification methods refer to Ljung [55].

Up until the 1990's FDI was treated in special sessions of symposia and congresses; Developments in these methods were also featured routinely within the context of control literature. To address the growing interest in FDI and with the objective to consolidate the literature on fault management, the IFAC SAFEPROCESS conference was created. The first Symposium for SAFEPROCESS was hosted in Baden-Baden, Germany (1991). A Technical Board was established to organise and manage continued conferences in the field of fault management. Since their first conference, SAFEPROCESS has been hosted routinely with conferences held in Finland (1994), United Kingdom (1997), Hungary (2000), United States of America (2003), China (2006), Spain (2009), Mexico (2012) and France (2015).

IFAC SAFEPROCESS has also established the working group Industrial Application of Advanced FDI/Fault Tolerant Control (FTC) Technology with the general objective of promoting the usage of fault prognosis within industry. With the formation of the IFAC SAFEPROCESS Technical Committee, one of their actions was to consolidate and clarify the terminology used in the fault management literature to provide clear definitions for common terms. A summation of this terminology is provided in Appendix B.

During the 80's and 90's the FDI literature saw a shift in focus. The development of model-based fault detection began to explore and leverage advancements in nonlinear methods [50] rather than focus on extending established linear modelling techniques. FDI also began to be explored from the perspective of machine learning and pattern recognition. Methods such as fuzzy logic and neural networks were increasingly applied to provide solutions [50, 56].

During this time academia also began applying robustness (the insensitivity to uncertainties and disturbances) to FDI. The motivation for this shift was a realism of modelling, that no model perfectly and accurately describes a physical system. A model will fall victim to model uncertainty, external disturbances and noise; resulting in discrepancies between the physical system and the model. These limitations cause difficulty in the application of FDI, often leading to false alarms and missed detections for faults. Producing an FDI solution with a robust model these limitations can be reduced, given the robust model is still sensitive to faults whilst maintaining its insensitivity to uncertainties and disturbances. As such, the application of robust control theory to the field of FDI became a logical progression [57].

The modern approach to FDI can be split into two sub problems; the accurate detection of symptoms (fault detection), and the classification of these symptoms into faults (fault isolation). Towards the end of the 90's there were now well established methods for model-based fault detection [58] and numerous examples of industry application [59]. These established methods are diagnostic observers, parity-relation methods and parameter estimation methods.

Fault isolation methods can be categorised into two approaches; Classification methods and Inference methods [60]. Classification methods include the use of statistical methodologies, geometric methods and neural networks. Alternatively, inference methods make use of knowledge based reasoning and as such apply Boolean logic, approximate reasoning, and fuzzy logic.

The early 2000s through to today has seen continued momentum in the research and application of FDI. Current areas of research interest are Robust FDI schemes (e.g. [61]), Adaptive FDI schemes (e.g. [62]), Neural Networks applied to FDI (e.g. [63, 64]), and FDI methods for hybrid systems (e.g. [65, 66]). FDI can greatly leverage advances in nonlinear state estimation techniques for robust and accurate fault detection as accurate state estimation of processes naturally lends to better performing FDI systems.

For a comprehensive understanding of the the history in FDI methods, the following survey papers have summarised the literature at each respective point in time [50, 51, 51, 52, 60, 67–69]. Similarly, the State of the Art in FDI methods have been routinely detailed in the following books [47, 57, 70–73].

The development of FDI is stimulated by the continuing trend of automation and the demands of higher availability and safety. To meet these demands autonomous systems are growing in complexity facilitated by the adoption of modern control theory. It is through modern control theory that we have the tools for mathematical modelling, state estimation and parameter estimation that enable FDI and the exploitation of analytical redundancy in complex systems. Coupled with the increasing computational power available and miniaturisation of both computing architectures and sensors, the application of FDI becomes ever increasingly feasible.

2.2 Fault Detection and Isolation Problem definition

The problem definition for FDI can be split into two sub problems; residual generation and residual evaluation. The former relates to the challenge of constructing residuals to facilitate fault detection and the generation of symptoms for fault isolation. The latter relates to the challenge of, given a set of symptoms, inference and classification of faults.

2.2.1 Residual Generation

The problem for model-based fault detection is to generate a residual that is informative with respect to system faults, yet insensitive in the presence of model uncertainties and disturbances.

Commonly in industry, instead of residuals, output signals are evaluated and compared against a given threshold for the detection of faults. The main deficiency with this approach is that successful fault detection is at the mercy of the variability in the signal. Large thresholds have to be applied to account for system dynamics to reduce false alarm rates. This imposes significant restrictions on the type and magnitude of detectable faults, often leading to an unacceptable missed detection rate. Similarly, setting aggressive thresholds for output signals results in an increase in false alarms and leads to unnecessary action and cost.

Assume a set of measurements $y_{1:T}$ generated by a process described by the Nonlinear State Space Model (NSSM) (Equation 1.2.2), repeated here for convenience.

$$\begin{aligned} x_{t+1} &= f(x_t, u_t) + w_t \\ y &= h(x_t, u_t) + v_t \end{aligned} \tag{2.2.1}$$

where $f(\cdot)$ is a nonlinear function describing the process model, $h(\cdot)$ is a nonlinear function describing the measurement model, $w_t \sim \varphi(0, \sigma_Q^2)$ and $v_t \sim \varphi(0, \sigma_R^2)$ are zero mean random variables for the process noise and sensor noise respectively.

If fault detection is achieved through monitoring the outputs of a process, a conservative threshold to avoid false detections is set by:

$$\inf(y_{1:T}) \leq h(x_t, u_t) + v_k \leq \sup(y_{1:T}) \tag{2.2.2}$$

This illustrates an immediate problem; to avoid any false detections the thresholds for fault detection are dependant on the absolute variability of the measurements. This limitation can be overcome by modelling the dynamic process and using the difference between the measured system output y_t and the estimated output $\hat{y}_t$. This difference between model and actual is known as the residual, $r(t)$, and under nominal system behaviour $r(t) \approx 0$. This is the fundamental premise behind model based fault detection.

$$
\begin{aligned}
r_t &= y_t - \hat{y_t} \\
r_t &= y_t - h(\hat{x_t}, u_t) + v_k
\end{aligned} \tag{2.2.3}
$$

Thus a fault $\mathbf{F}$ is detectable only if

$$
\inf(r_{1:T}) \leq y_t - h(\hat{x_t}, u_t) + v_k \leq \sup(r_{1:T}) \tag{2.2.4}
$$

With accurate state estimation $y_t - h(\hat{x_t}, u_t) \approxeq 0$, thus the threshold of the residual is only governed by the process and measurement noise w_t and v_t. Therefore, constructing a residual removes the limitations faced by thresholding a process output. The objective of a residual is to provide detection of a fault $\mathbf{F}_t$ from the output and input signals of a process that is both robust in the presence of parameter uncertainty and noise, whilst sensitive to system faults.

The problem of residual generation and residual evaluation are coupled and design changes to one will typically influence the performance metrics of the other. The design of an FDI system must take a holistic approach with respect to residual generation and residual evaluation for a practical implementation.

2.2.2 Residual Evaluation

The second aspect to FDI is fault isolation. This is the decision making process that in the event of a detected fault, identifies and distinguishes the type and magnitude of fault based on the available information provided. In an ideal world, residual generators are perfectly decoupled from model uncertainties and disturbances whilst only sensitive to specific faults. In this case, if all faults are perfectly decoupled, a simple application of boolean logic based on which threshold was tripped will result in accurate fault isolation.

However, in practise this is rarely the case. Model uncertainties and process noise will confound the process of FDI, thus a robust residual evaluation is necessary to mitigate the negative effects of these uncertainties and disturbances. Instead residuals can be thought of as symptoms of a problem, and then using these symptoms $\boldsymbol{S}$ to map to faults $\boldsymbol{F}$.

The objectives for residual evaluation is to uniquely identify a fault whilst minimising both the missed detection rates and false alarm rates with the available information provided. This is a complex problem to solve, particularly as the missed detection rate and false alarm rate are in direct competition. Increasing the sensitivity of the residual evaluation strategy may decrease missed detection rates, but often increases false alarm rates.

Therefore, the challenge in Residual Evaluation is to create a mapping between a a set of symptoms $\boldsymbol{S}$ to a set of faults $\boldsymbol{F}$ to reliably isolate faults in the system.

$$
\boldsymbol{S} \rightarrow \boldsymbol{F} \tag{2.2.5}
$$

2.3 Model-based Fault Detection Methods

There are various methods for fault detection. These include such methods as limit checking [52], trend checking, multi-variate data analysis [74], and model-based fault detection [60]. The focus of this section is to review the basic approach to model-based fault detection and the leading methods for residual generation.

The current focus in the fault diagnosis literature has been to extend the current known methods of residual generation through the use of nonlinear models. Similarly, there has been constant work to improve the robustness of these residual generators to handle greater levels of model uncertainties and disturbances. This section gives an overview of the fundamental methods for model-based fault detection and provides reference to area's of current research focus.

The general scheme for model-based FDI is given by Isermann [60]. The approach taken in this thesis is given in Figure 2.1.

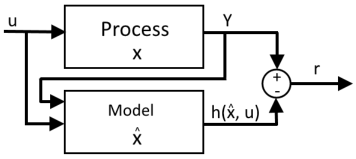

Figure 2.1: Model-based fault detection and isolation scheme

2.3.1 Diagnostic Observers

Diagnostic observers have received a large proportion of research focus for fault detection methods. This is largely due to the attention and familiarity of state estimators, also known as state observers, in general by the control community who are significant contributors to the fault diagnosis literature. State estimators have garnered substantial focus in the control literature [75], with extensions for nonlinear models, using reduced state information and greater robustness in the face of model uncertainties and disturbances. For an overview of state estimation, refer to Section 1.2.

To derive a diagnostic observer we first assume a process described by the following NSSM.

$$\begin{aligned} x_{t+1} &= f(x_t, u_t, F_t) + w_t \\ y_t &= h(x_t, u_t, F_t) + v_t \end{aligned} \tag{2.3.1}$$

where $f(\cdot)$ is a nonlinear function describing the process model, $h(\cdot)$ is a nonlinear function describing the measurement model, $w_t \sim \varphi(0, \sigma_Q^2)$ and $v_t \sim \varphi(0, \sigma_R^2)$ are zero mean random variables for the process noise and sensor noise respectively, and Faults $F_t = \{F_t^1, ..., F_t^n\}$.

Generally, a state estimator can be constructed such that [76]:

$$\begin{aligned} \hat{y_t} &= h(\hat{x_t}, u_t) + v_t \\ \hat{x}_{t+1} &= f(\hat{x_t}, u_t) + w_t + L(y_t - \hat{y_t}) \end{aligned} \tag{2.3.2}$$

Where L is is an appropriate gain to ensure stability of the estimator. It is then posited that deviations between y_t and $\hat{y_t}$ are caused by faults F_t, given the model and noise assumptions used for the observer hold.

Using the state estimator, we can construct a diagnostic observer to generate a residual as follows:

$$\begin{aligned} r_t &= y_t - \hat{y}_t \\ r_t &= y_t - h(\hat{x_t}, u_t) + v_t \end{aligned} \tag{2.3.3}$$

Where y_t is the actual measured process output, and $\hat{y}_t$ is the estimated process output calculated from the process model, measurement model and estimated internal state. This residual, r_t, is then monitored for deviations from the process described by the model given by Equation 2.3.1.

It is also important to note that the gain L is in competition with the fundamental function of a diagnostic observer. A diagnostic observer informs of a fault when the residual r_t exceeds a predetermined threshold. Given the residual is calculated as $r_t = y_t - h(\hat{x_t}, u_t)$, in faulty conditions L can erode the observer error resulting in an unrepresentative internal estimated state for the model. To account for such undesired effects, the incorporation of state constraints into the diagnostic observer may be required or alternatively the monitoring and thresholding against known nominal bounds of the estimated internal states. In the event that the estimated system state maintains representative values, both the estimated internal state and outputs provide the indicative information required for fault detection and generation of symptoms for fault isolation.

It should also be mentioned that for fault detection we are typically interested in estimating the output of a process from which we can compare directly to measured signals. This contrasts with the traditional objective of state estimation to estimate the internal state of a process. This distinction isn't always noted and as such has often misled to the erroneous opinion that state space theory is indispensable for diagnostic observers. So long as an estimate of the process output can be constructed from physical knowledge and process measurements, this can be considered a diagnostic observer [50]. Nonetheless, state estimation is a very useful tool for the construction of diagnostic observers.

There have been many methods for observer-based residual generation proposed in the literature. Such methods include banks of observers [43], fault sensitive filters [44, 45], Kalman filters, sliding mode observers, etc. A comprehensive review of methods employed to detect faults in civil aviation is given by Marzat [69]; this includes summaries of the advantages and limitations of these methods. Similarly the prominent texts written by Patton [71] and Ding [57] both provide detail and reference to the various methods for diagnostic observer based residual generation.

2.3.2 Parity Relations

The parity relations methods is another model-based method for fault detection. With early work laid down by Himmelblau [47], it was formalised further by Chow and Willsky [48] as an approach for robust residual generation for a linear dynamic system. In essence the parity relations methods takes a fixed model and runs it in parallel to the process. The output from both the model and the process are subtracted against each other thereby forming a residual. The general process for the parity relation method is given by Gertler [49].

Given a linear discrete process described by:

$$y(k) = M(q)u(k) + S(q)p(k) \tag{2.3.4}$$

where q is the backward shift operator, $M(q)$ and $S(q)$ are rational transfer functions, $u(k)$ are measured inputs, $y(k)$ are measured outputs and $p(k)$ are additive disturbances and faults. Using a model of the process residuals can be generated by the parity relation

$$r(k) = W(q)[y(k) - \hat{M}(q)u(k)] \tag{2.3.5}$$

Substituting equation 2.3.4 into our generated residual, we see we can form a residual that captures disturbances and faults.

$$\begin{aligned} r(k) &= W(q)[M(q)u(k) + S(q)p(k) - \hat{M}(q)u(k)] \\ r(k) &= W(q)[S(q)p(k)] \end{aligned} \tag{2.3.6}$$

The challenge is to then select a transfer function $W(q)$ that produces a residual that is robust to disturbances but sensitive to faults. Accurate residual generation also depends on accurate estimation of the model $M(q) = \hat{M}(q)$. Similar to diagnostic observers, the challenges to overcome with this approach is to construct residuals that are robust in the presence of model uncertainty and disturbances.

It is worth noting that there is an innate relationship between diagnostic observers and parity relations. It has been shown by Gertler [51] that any parity relation can be transposed into an observer format and visa versa. As such parity relations are structurally equivalent but have a different design procedure compared with observer based residual generation.

2.3.3 Parameter Estimation Methods

Fault detection via parameter estimation utilises the principle that some faults in the process can be associated with deviations in the parameters of a process [54]. Given the dynamics of a system are governed by physical laws, process parameters have value and tangible meaning, thus we can monitor these parameters for deviations which can be indicative of a fault.

For instance, a mechanical spring-damper system can be described using Netwon's laws and the parameters of stiffness for the spring and damping for the dashpot. Deviations in the values of these parameters can be indicative of a fault in a system. If the stiffness of the spring decreases, this can indicate component deterioration and potentially provide forewarning of failure. This is the underlying principle of Parameter Estimation Methods (PEMs) for fault detection, using the deviations of the process parameters as symptoms for FDI.

Given a process described by a NSSM as follows:

$$\begin{aligned} x_{t+1} &= f_\Theta(x_t, u_t) + w_t \\ y_t &= h_\Theta(x_t, u_t) + v_t \end{aligned} \tag{2.3.7}$$

where $f_\Theta(.)$ is a nonlinear function describing the process model , $h_\Theta(\cdot)$ is a nonlinear function describing the measurement model, Θ is the set of model parameters, $w_t \sim \varphi(0, \sigma_Q^2)$ and $v_t \sim \varphi(0, \sigma_R^2)$ are zero mean random variables for the process noise and sensor noise respectively.

The process model parameters,

$$\Theta = [\theta_1, \theta_2, ..., \theta_n] \tag{2.3.8}$$

encapsulate a combination of the physical properties of the process; e.g. mass, length, stiffness, electrical resistance, drag, friction coefficient, etc. Faults which influence these physical properties can be detectable using PEMs. To establish a PEM for fault detection we perform online estimation — using a suitable estimation procedure [55] — of the model parameters Θ from measurements y_t and u_t. These are the current estimates of the parameters of the process we are monitoring. These model parameters Θ do not always map directly to the physical process coefficients. Thus a determination of the relationship between the model parameters Θ and the physical process coefficients p may have to be performed.

$$\Theta = g(p) \tag{2.3.9}$$

so long as this relationship exists and is invertible, we can calculate the process coefficients from the model.

$$p = g^{-1}(\Theta) \tag{2.3.10}$$

by detecting the changes Δp in the process, we can use these as symptoms for FDI. Fault detection can be made simply by checking the process parameters against predetermined threshold levels or by using other means, such as adaptive thresholds, statistical decision theory or pattern recognition techniques.

For further reference on parameter estimation for fault detection see Pouliezos et al. [54]. For a more general treatment on parameter estimation see the distinguished text on systems identification by Ljung [55].

2.4 Fault Isolation Methods

The other aspect of an FDI system is fault isolation. Fault isolation entails the determination of the type of fault with as much detail as possible, such as size, location and time of detection. The diagnosis is based on the detected analytical and heuristic symptoms produced by the residual generators.

The inputs to a fault isolation system are the available symptoms $\boldsymbol{S}$ and the fault relevant knowledge about the process. Typically the symptoms are presented as a boolean (i.e. either 0 or 1), but can also take more quantitative qualities as afforded by the method used.

Given the fault symptoms of a system $\boldsymbol{S}$, we want to diagnose the system faults $\boldsymbol{F}$, a relationship has to be developed such that:

$$\boldsymbol{S} \rightarrow \boldsymbol{F} \tag{2.4.1}$$

The approaches taken to develop this mapping can be broadly divided into two categories, Classification Methods and Inference Methods [60].

2.4.1 Classification Methods

If there is no *a priori* knowledge about relationships between a system's symptoms and faults then classification methods provide the best approach for fault diagnosis. Typically the relationships between symptoms $\boldsymbol{S}$ and faults $\boldsymbol{F}$ will be experimentally determined and the relationship learned. One such classification method that has seen wide application in FDI is Artificial Neural Networks (ANNs).

2.4.1.1 Artificial Neural Networks

ANNs are a family of models that have been inspired by the neural networks of the biological central nervous systems. ANNs simulate these neuron networks to solve complex, mathematically ill-defined problems, nonlinear problems or stochastic problems in a generalised manner. They contrast to conventional computational algorithms in that they employ very simple computational operations (addition, multiplication and logical elements) in a generalised architecture.

Whereas a conventional computer algorithms will use a set of complex equations and will apply to only a given problem or subset of problems, the ANN architecture is versatile enough to provide solutions for almost any problem that can be reduced to a mathematical function given adequate training data can be supplied.

ANNs demonstrate the following two key characteristics [77]:

1. computationally and algorithmically simple
2. self organising, allowing it to be relevant for a wide range of problems

ANNs can be applied directly to the problem of fault isolation by "learning" a function to relate between symptoms $\boldsymbol{S}$ and faults $\boldsymbol{F}$.

$$\boldsymbol{F} = g(\boldsymbol{S}) \tag{2.4.2}$$

An ANN architecture consists of multiple neural layers; an input layer, an output layer and typically some hidden layers. Each layer contains a number of neurons that map to neurons on the previous layer. The input layer neurons map to the network inputs, in the case of FDI these would be the symptoms $\boldsymbol{S}$. Similarly, the output layer maps to the network outputs and in the case of FDI these are the faults $\boldsymbol{F}$. Each link between neurons has an associated weight and an activation logic. The activation logic is typically a design parameter and the associated links and weights between neurons are learned parameters achieved through neural network training. This training is performed against a training data set and results in an ANN converging to the desired function $\boldsymbol{F} = g(\boldsymbol{S})$.

ANNs have seen application within FDI [46] to classify faults $\boldsymbol{F}$ based on symptoms $\boldsymbol{S}$. One such example is the application and comparison of three ANN architectures for FDI in Robotic Manipulators [78]. In this study ANNs were used for both residual generation and evaluation with success. They were demonstrated to provide accurate fault detection and the versatility of being easily extendible to detect new classes of faults. A generic architecture of an ANN for fault isolation is given in Figure 2.2.

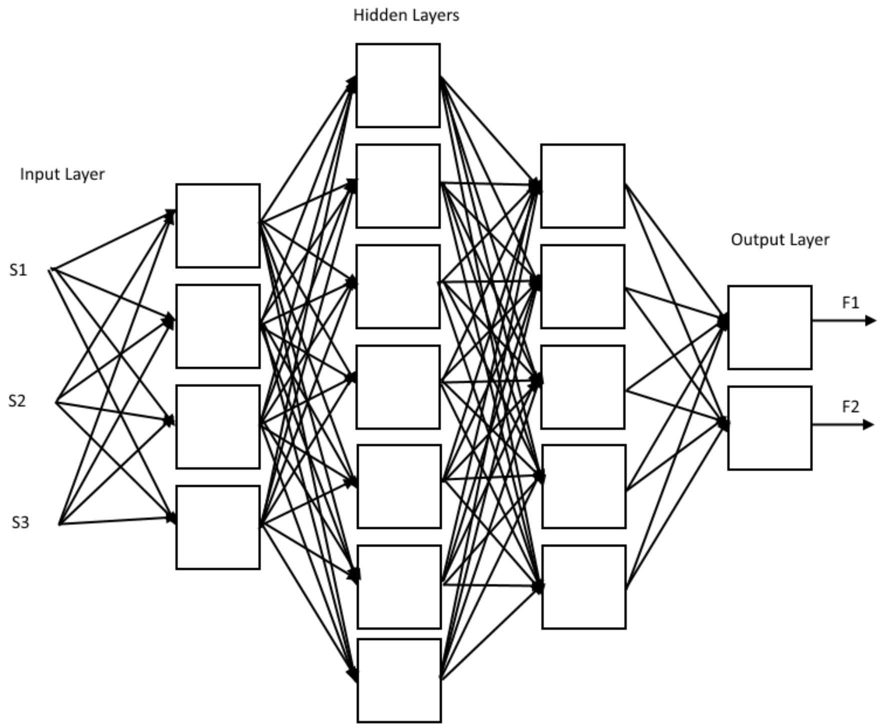

Figure 2.2: Example architecture of an ANN for fault isolation

The simplicity and generality afforded by ANNs make their application to solve complex classification problems a promising endeavour. There is currently much research ongoing in ANNs under the banner of Machine Learning. It is also worth noting that ANNs have also been applied in the construction of residual generators; providing a black box representation of complex nonlinear models in the build up of an accurate system model for process monitoring [46].

2.4.1.2 Statistical Decision Making Methods

Another set of classification based methods for Fault Isolation is through the use of statistical theory for decision making. These methods are appropriately called statistical decision making and an overview of these methods in the context of FDI is given in the following survey paper [79].

Statistical division making formulates the problem of evaluating a fault as follows. Given a set of measurements, observations and symptoms, a decision at each time step is made to test multiple hypotheses.

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \\ H_2 &: \theta = \theta_2 \\ &\vdots \\ H_N &: \theta = \theta_N \end{aligned} \tag{2.4.3}$$

So long as the data supports hypothesis H_0, that there is no fault, then monitoring continues. If the data supports one of the other hypotheses, a fault is declared by the system. Essentially this is a problem of detecting change in signals and has remained an area of strong research.

For a full treatise on change detection in the context of FDI, Gustafsson's book "Adaptive Filtering and Change Detection" [80] serves as an excellent introductory text covering both model-based filtering techniques and statistical decision making methods.

2.4.2 Inference Methods

For some systems the basic relationships between symptoms and faults are understood, at least partially. We can then use this *a priori* knowledge to develop causal relationships between detected symptoms, events and faults [60]. These methods are refered to as inference methods, and two common approaches that have seen application in FDI are Fault Tree Analysis and Fuzzy Logic.

2.4.2.1 Fault Tree Analysis

Fault Tree Analysis (FTA) was first conceived by H. R. Watson at Bell Telephone Laboratories in the early 1960s. It was used as a technique to perform safety evaluation of complex systems. It soon became apparent that the method used for describing the logical flow in their system could be used for analysing component faults how how its effects cascade through a system [81].

FTA has been further developed by both industry and academia and is now a widely used analytical tool for safety analysis. Given its ubiquitous use, its application in FDI is an obvious progression.

Here the logical rules built up for each failure mode in the FDI are used to infer a mapping between events, the symptoms $\boldsymbol{S}$ and how these progress to faults $\boldsymbol{F}$ and failure modes. This can be observed in Figure 2.3, where for example, when symptoms S1 or S2 occur fault F1 results. Further, if fault F2 has actualised and symptom S5 occurs, the system deteriorates further to fault F4, and so on. These rules allow for the build up of failure descriptions of a system using *a priori* knowledge, creating a graphical logical transition of how symptoms and faults cascade through a system.

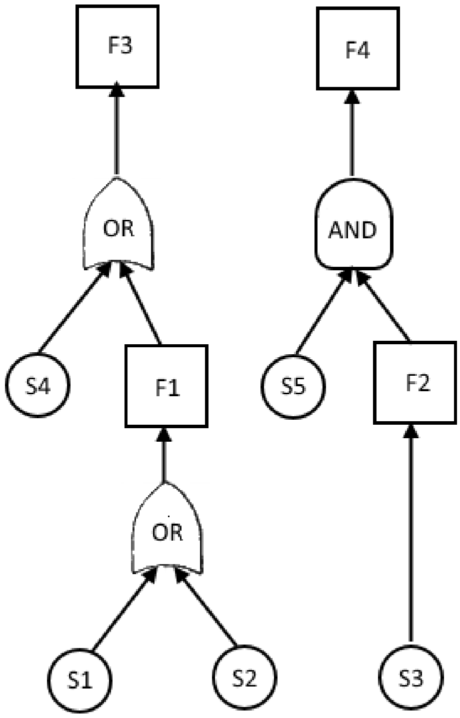

Figure 2.3: Example of a FTA for fault isolation

2.4.2.2 Fuzzy Logic

Fuzzy logic has its origins when the first developments of an alternative approach to classification sets was being explored. The objective was to construct a new paradigm to address levels of membership to sets, such that levels of uncertainty of membership could be expressed. In 1965, Lotfi Zadeh first introduced this idea, through the continuous valued logic that he called Fuzzy Sets [82]. This was a first step towards a structured way of modelling uncertainty in what has become known as Fuzzy Logic.

Uncertainty can be thought of as being the inverse of information. Information about a particular engineering problem can be incomplete, imprecise, vague, unreliable, contradictory or deficient in some other way. It is with Fuzzy Logic we have the tools to model uncertainty in all forms and build Fuzzy Systems. Several sources have shown and proven that fuzzy systems are universal approximators [83]. Fuzzy systems, like algebraic functions, maps input variables to output variables and in the context of fault diagnosis provide $\boldsymbol{S} \rightarrow \boldsymbol{F}$.

However, it is not just this mapping where Fuzzy systems provide value for FDI. The primary benefit of fuzzy system theory lies in its ability to approximate system behaviour where analytical functions or numerical relations do not exist. Fuzzy systems provide a structure to capture *a priori* knowledge and model complex systems approximately through deductive reasoning. For a structured introduction to Fuzzy systems please refer to the comprehensive text “Fuzzy Logic with Engineering Applications” written by Timothy Ross [83].

Fuzzy logic has been widely applied to FDI in the literature and continues to be actively researched today. An example of its utilisation is its application to provide a fault diagnosis system for an anaerobic digester; using a fuzzy system to map directly from raw data to system faults [84]. Similarly fuzzy logic has also been applied to augment diagnostic observers for residual generation [85].

Chapter 3

Fault Detection for a Military Vehicle Battery System

This chapter of the thesis concerns itself with the development and evaluation of a fault detection system for a military vehicle battery system. A model based diagnostic observer approach is taken, utilising an Extended Kalman Filter (EKF) with hard constraints. This is contrasted against a Moving Horizon Estimator (MHE) approach for a diagnostic observer with performance between the two algorithms compared. To facilitate the development and evaluation between the algorithms, a test bench was constructed to provide a representative system of the military vehicle battery system.

This chapter is broken down into the following sections: Section 3.1 provides a examination on the general benefit of FDI as applied to military vehicles, and more specifically the benefit of a FDI system for a military vehicle battery system. Section 3.2 gives an overview of the switched battery system to which fault detection was applied. Section 3.3 provides an overview of the fundamentals of battery dynamics and the model developed for use in the diagnostic observer. Section 3.4 describes the design and subsequent data collection from the test bench constructed to represent the military vehicle switched battery system. Section 3.5 gives the diagnostic observer formulations for the fault detection system, one based upon a constrained EKF and the other based upon a MHE, then compares their performance.

3.1 Benefit

With the the rapid pace of miniaturisation of electronics and sensors, their diminishing cost, and growing ubiquity, military vehicles are becoming increasingly instrumented. Such instrumentation enables new methods for the TLS of military vehicles, such as the application of FDI systems. FDI systems have proven results in their application for the detection and diagnosis of faults and are recognised for enabling Condition Based Maintenance (CBM) [2, 3]. It is through these new paradigms, such as CBM, that FDI delivers reductions in TLS costs and an increase to platform availability as the maintenance of military vehicles can be optimised. FDI has already been widely employed in industry, including commercial automotive [86], aerospace [87] and industrial processes [88] to name a few, with the benefits enjoyed by these industries capable of being extended to military vehicles.

In fact, equipping military vehicles with FDI systems provides further benefit than just TLS cost reductions and increased platform availability. Having access to real-time data on the condition of the military vehicle will also inform commanders with the health of their convoy, allowing for the tailoring of asset selection for missions. Additionally, having highly instrumented platforms is a natural precursor to higher levels of automation, providing the foundation for development in vehicle platooning and other autonomous capabilities.

With the general benefit of FDI understood, an application of an FDI system for military vehicles was explored. The system chosen was the battery system given its importance on platform availability and impact to TLS costs. FDI in battery systems is an important capability for modern vehicles. This need is prevalent in vehicles with electric and hybrid electric drives [89]. Health estimation in battery systems is also of importance for the power industry for the storage of energy from renewable sources. Such systems contribute to ensuring a reliable power supply uninterrupted by faults. The storage and retrieval of energy in secondary batteries has become ubiquitous as electronics and their energy demands have become integrated into all aspects of society.

This growth of embedded electronics has also been seen in military vehicles over the years. Growing electrical demands have largely been due to the modernisation of equipped Command Control Communications Computers Intelligence Surveillance and Reconnaissance (C4ISR) systems supporting Network Centric Warfare (NCW) [90]. As such, military vehicles exhibit a need, just as electric vehicles do, for robust battery systems to support these energy demands. A failure in a battery for a military vehicle can lead to its inability to complete its mission, and potentially strand its occupants. Due to the varying nature of vehicle use and mission profile, engine hours are typically not a reliable indicator of battery health. Accurate health and monitoring of military vehicle batteries is thus an important enabler for ensuring mission effectiveness and optimising vehicle TLS.

To maximise both the performance and usable life of a battery it is important to have an understanding of key battery parameters, such as State of Health (SoH) and State of Charge (SoC). Unchecked faults and deterioration in batteries can lead to irreversible damage causing reduced component life and in the case of military vehicles, platform unavailability. Lead-Acid batteries are commonly used on military vehicles and are considered here in the battery system. The military vehicle battery system is used to supply electrical energy for engine cranking and powering system equipment when the engine is off.

Typically in trucks or other heavy vehicles, Lead-Acid batteries are primarily used to crank engines and then only provide small amounts of power to the vehicle electronics when the engine is off. Due to the operational demands of military vehicles and the addition of modern C4ISR systems, there is an increased electrical load demand on military vehicle battery systems. This increased electrical load causes excessive, deep discharge of the batteries due to operational demands, such as silent watch. This can leave the batteries in depleted states that cause deterioration at a faster rate than anticipated.

To avoid such circumstances, it is important to monitor the batteries and provide early detection of faults and accurate estimations of the SoH. By having accurate SoH estimation, deteriorating batteries can be identified and recovered before permanent damage is caused, maintaining high component life. This monitoring also reduces the risk of platform unavailability, especially during silent watch operations when the batteries undergo the most amount of stress.

Increasing battery life and reducing the risk of platform unavailability are motivating cases for accurate health estimation of batteries on military vehicles. This can be achieved through the development and integration of an FDI system. This Chapter explores the development of two candidate fault detection systems for a military vehicle battery system and evaluates their effectiveness based upon the results observed.

3.2 Military Vehicle Battery System

To explore the application of FDI methods in the context of military vehicles, a subsystem of the platform was chosen. The chosen system was a military vehicle battery system. The battery system is a system that typically contributes a high impact on support costs through early replacement of batteries, has high impact on vehicle availability in the event of a failure, and can be readily equipped with low cost sensors to provide information for the application of FDI.

A typical battery system on some military vehicles is given in Figure 3.1. Here we can see the battery system provides power to the starter motor, vehicle electronics and onboard systems. There are two battery banks on the vehicle, the crank battery and the system battery. These two battery banks are charged via the engine alternator and the circuit can be divided by three isolator switches. The battery banks consist of two 12 Volts batteries in series to provide 24 Volts to the platform systems.

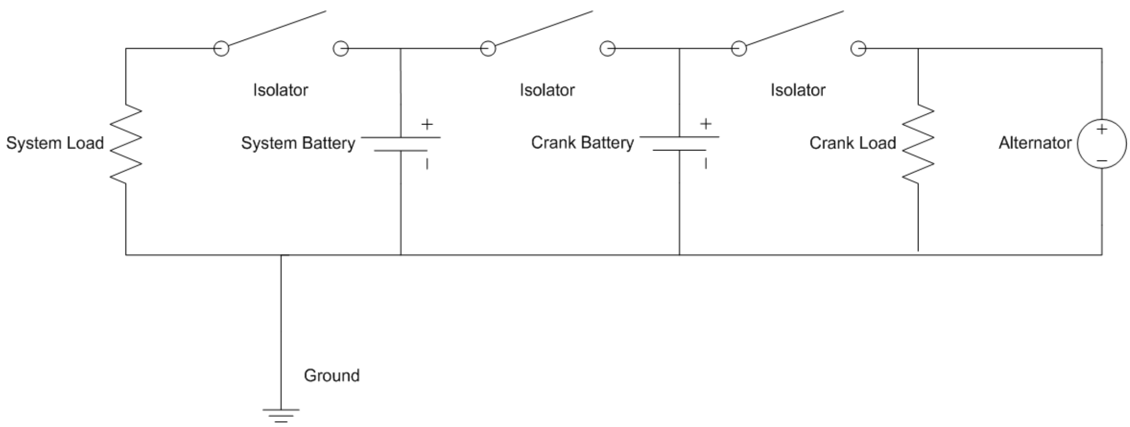

Figure 3.1: Generalised diagram of a military vehicle battery system

Through the three isolators, the battery system can be configured into eight connection states. The most common connection states are as follows:

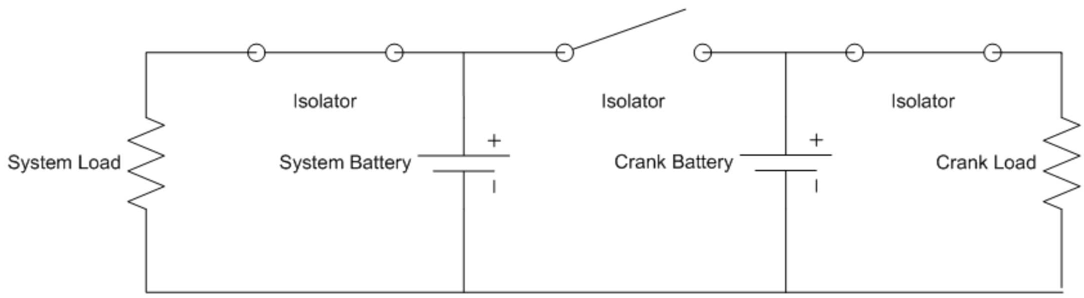

Figure 3.2: Independent battery discharge

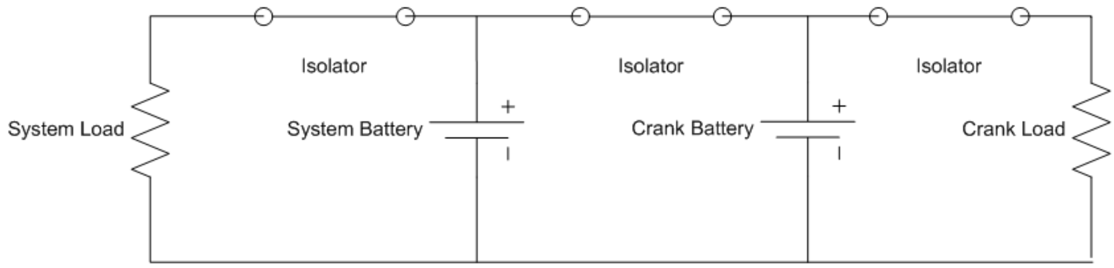

Figure 3.3: Coupled battery discharge

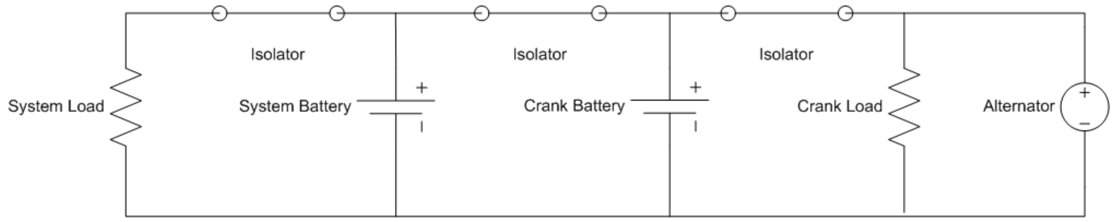

Figure 3.4: Coupled battery charge

3.3 Dynamic Model for Military Vehicle Battery System

With the system described, we wish to develop a diagnostic observer that can provide fault detection in the battery system. First we must define what faults we wish to detect. Our definition is general, a battery fault is when a battery is incapable of providing or accepting a sufficient level of charge for the system independent of its SoC. With such a definition of a fault we can assess the health of the battery system on a sliding scale as the deterioration of batteries cross a predetermined threshold as to what we decide is faulty. This allows for tuning of this threshold for health assessment to optimise TLS costs and platform availability.

With the definition of the fault we wish to detect, we now must construct a dynamic model of the system concerned for use in the diagnostic observer. The model is constructed from our understanding of the dynamics of batteries. It is through this understanding that we can exploit analytical redundancy in the model and measurements. There are typically two approaches to modelling a battery; an Electrochemical Model that relies on chemical laws and dynamics or an Equivalent Circuit Model (ECM) that relies on electrical laws and dynamics. The modelling approach taken here is to model the battery system using electrical laws and dynamics by employing an ECM based model. This approach has been used widely in the literature with successful results [89, 91–96].

3.3.1 Overview of the fundamentals of battery dynamics

To develop a model based FDI system for the battery system, an understanding of the battery dynamics is required. It is through this understanding we can best construct a model that captures the underlying mechanisms at work in the battery system; whilst utilising this knowledge to construct a diagnostic observer that is sensitive to faults and insensitive to process and measurement noise. Here we give an overview of the fundamentals of battery dynamics based on Linden's

Handbook of Batteries [97] and Jossen's work [98] in fundamental battery dynamics for the Journal of Power Sources. For a full treatise on battery dynamics, please refer to the prominent text Linden's Handbook of Batteries [97].

3.3.1.1 Battery dynamics and time domain

The time scale on which certain dynamic behaviours occur in a battery vary from microseconds all the way up to several years. The general time scales for these different physical phenomena is given in Figure 3.5. It should be noted that this is only indicative, and that actual time scales for battery physical effects are strongly dependent on the battery chemistry, design, temperature and SoC. The key phenomena we are concerned with are electric double layer effects and mass transport effects. With aging effects and cycling effects responsible in part for the degradation of the battery performance and overall health, a diagnostic observer will indirectly track these effects as it detects a divergence from a nominal performance given by the observer.

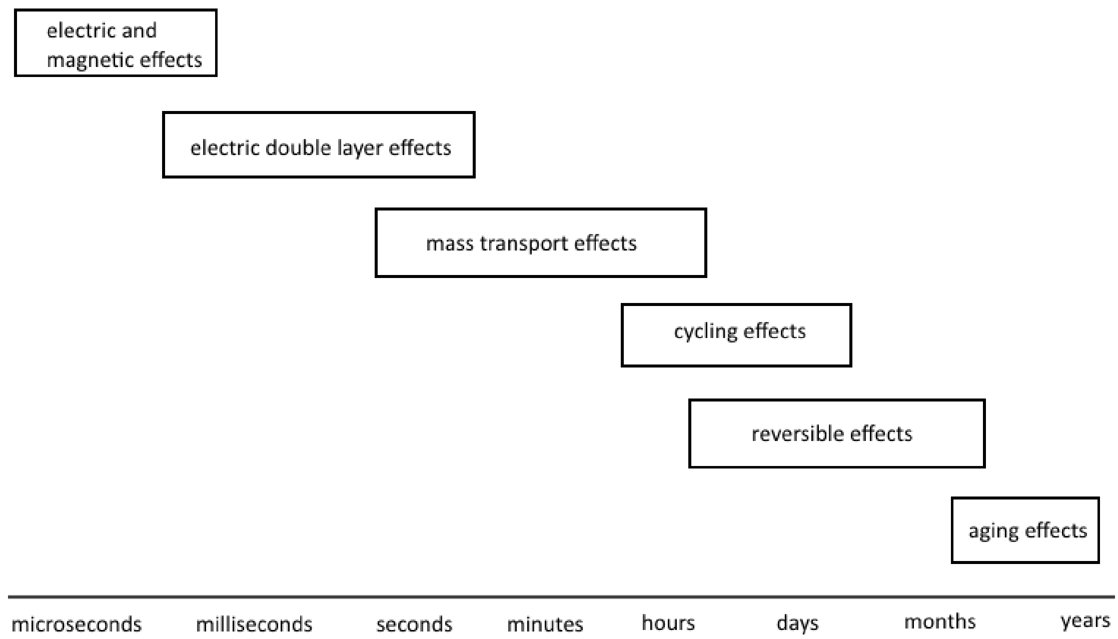

Figure 3.5: Battery dynamics time domain

3.3.1.2 Thermal effects

The temperature of a battery has a pronounced effect on its useful capacity and voltage characteristics. This is due to the effect of temperature on the battery's diffusion rate and internal resistance. At higher temperatures, the battery experiences an accelerated rate of diffusion and a decrease in internal resistance. Conversely, lowering a battery's temperature causes the rate of diffusion to slow and an increase in internal resistance. Lowering a battery's temperature also results in the battery reaching its cut off voltage sooner and thus reduces their useful capacity. Figure 3.6 shows the general effect temperature has on battery capacity. Temperature also has an impact on the rate of self-discharge for a battery. Lowering a battery's temperature will minimise battery self-discharge, whilst increasing a battery's temperature will exacerbate self-discharge.

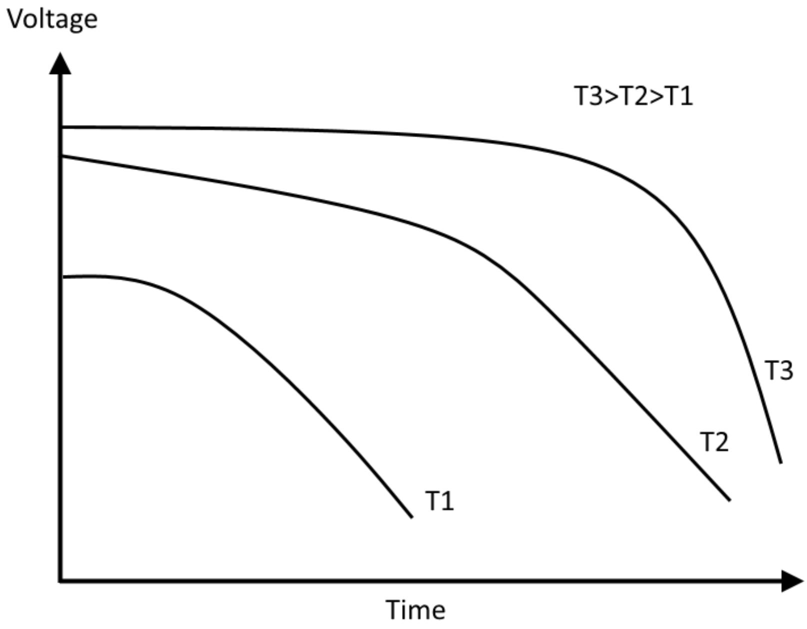

Figure 3.6: Effect of temperature on battery capacity — $T1$ to $T3$ increasing temperatures

3.3.1.3 Ageing

During a battery's lifetime, its performance degrades gradually due to irreversible physical and chemical changes that take place with its usage. Important factors that exacerbate the deterioration of a battery is its storage and discharge conditions. The time domain of battery ageing is typically in the range of months to years, with ageing effects typically predicted through the use of empirical relationships between battery age and performance. This measure of age is commonly discharge cycles. It is thus important to be conscious of the operating conditions of batteries and their respective usage history, as this information can be used to as a predictor of expected performance.

3.3.1.4 Reversible effects

Battery's also exhibit reversible effects, such as acid stratification in lead-acid batteries which can be removed by an extended charge. These effects occur during cyclic operation of the battery. Specific charge or discharge regimes can regenerate the battery, reversing these effects. Another example of a reversible effect is the memory-effect observed in nickel cadmium and nickel-metal hydride batteries. Here the batteries gradually lose their capacity if they are repeatedly recharged after being only partially discharged. Applying a full cycle to the battery can recover the batteries full capacity. The time ranges for reversible effects vary widely, from some hours to as long as a year. Reversible effects provide motivation for adequate condition monitoring of battery systems, as gradual deterioration can be ascribed to these reversible effects for some battery systems. In such cases, early identification of such deterioration can lead to preventative maintenance action to recover these batteries and extend their service life.

3.3.1.5 Cycling / State of Charge

The SoC of a battery is a dynamic characteristic that changes as a battery is charged or discharged. The time domain of SoC change depends on the battery capacity, load demand and charge profile of the battery. Typically, the time domain ranges from minutes up to several days. The terminal voltage of a battery is directly related to the SoC of a battery, with the terminal voltage decreasing as the SoC decreases. A generalisation of the relationship between SoC and terminal voltage is shown in Figure 3.7. The relationship between SoC and terminal voltage is dependant on battery chemistry and the overall design.

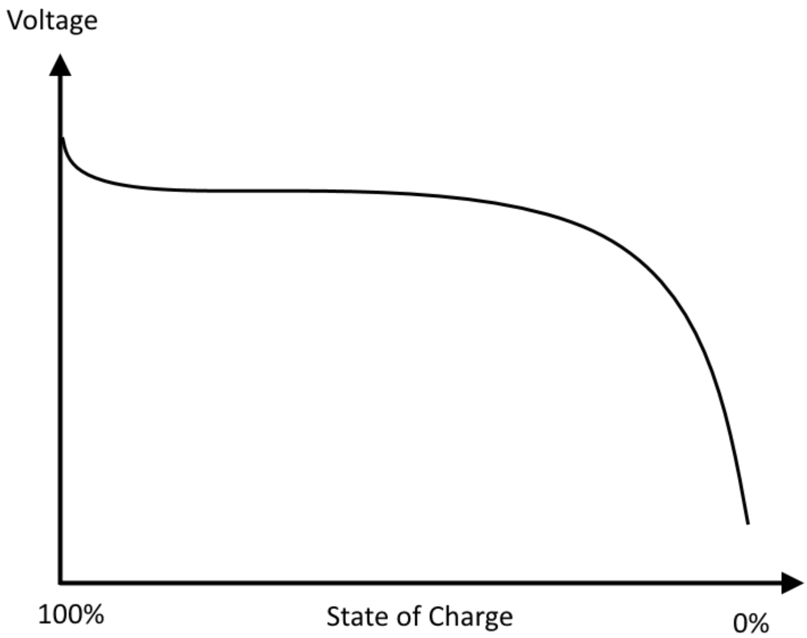

Figure 3.7: Typical relationship between SoC and terminal voltage

A cycle is defined as a discharge and charge event on the battery, SoC 100% $\rightarrow$ 0% $\rightarrow$ 100%. As a battery cycles, chemical energy is released and stored. The process in which energy is stored in a secondary (rechargeable) battery is through the exchange of ions through an electrolyte, reducing the positive material, whilst oxidising the negative material. Similarly, the process in which a secondary battery provides charge is through the reverse of this process, ion exchange causing oxidisation of the positive material whilst reducing the negative material.

Through this process of discharge and charge, deterioration to the positive and negative material structures take place resulting in reduced performance in the battery's ability to store and provide charge. This deterioration is why battery cycling has such a strong relationship with battery degradation and aging.

Additionally, as a battery cycles waste heat is generated. Typically the heat generated by a battery can be modelled as ohmic resistance, and thus waste heat generated by a battery can be calculated as:

$$P = RI_{\text{eff}}^2 \tag{3.3.1}$$

It is important to note that I_{eff} is not the same as the measured current. I_{eff} is dependant on the discharge profile of the battery. In the case of a Global System for Mobile (GSM) pulsed discharge regime, the effective current was 1.73 times greater than that of the average current [98].

3.3.1.6 Polarisation

Electrochemical reactions are the process which drives the production or storage of energy inside of a battery cell. These reaction kinematics are responsible for the rate of corrosion of a metal exposed to an electrolyte, with a reaction rate depending on the rate of electron flow to or from a metal-electrolyte interface. The maximum electrical energy that can be delivered by the electrochemical reaction inside a cell depends on the change in thermodynamic free energy ΔG of the electrochemical couple [97].

Generally, the reaction at one electrode can be represented by

$$aA + ne \rightleftharpoons cC \tag{3.3.2}$$

where a molecules of A take up n electrons e to form c molecules of C. At the other electrode, the reaction can be represented by

$$bB - ne \rightleftharpoons dD \tag{3.3.3}$$

where b molecules of B shed n electrons to form d molecules of D. The overall reaction can be formulated by adding these two half-cell reactions.

$$\begin{aligned} aA + bB - ne + ne &\rightleftharpoons cC + dD \\ aA + bB &\rightleftharpoons cC + dD \end{aligned} \tag{3.3.4}$$

The change in free energy ΔG of this reaction is given by

$$\Delta G = -nFE^o \tag{3.3.5}$$

where

$$F = \text{Faraday constant} \tag{3.3.6}$$

$$E^o = \text{Standard electromotive force of electrode reaction (Volts)} \tag{3.3.7}$$

This change in free energy ΔG is the driving force that enables a battery to deliver electrical energy to an external circuit.

However, not all of this chemical energy can be converted into useful electric energy due to losses caused by both polarisation and internal battery resistances, given off as waste heat. Losses caused by polarisation can be categorised as follows:

(1) Activation polarisation is a retarding factor that is inherent in all electrochemical reactions and drives the electrochemical reaction at the electrode surface. This phenomenon is also known as the double-layer effect and is detailed further in Section 3.3.1.7

(2) Concentration polarisation arises from the concentration differences in the reactants and products at the electrode surface and in the bulk as a result of mass transfer. This phenomenon arises from the diffusion process in the battery being unable to reliably transport ions and is known in the literature as the Mass Transport effect and is detailed further in Section 3.3.1.8

Accounting for internal resistance and polarisation losses, the terminal voltage can be calculated with the following equation given by Linden [97]. A graphical representation of the losses typical in a battery is given in Figure 3.8.

$$E = E_o - [(\eta_{ct})_a + (\eta_c)_a] - [(\eta_{ct})_c + (\eta_c)_c] - iR_i \tag{3.3.8}$$

where

$$\begin{aligned} E_o &= \text{open circuit voltage of cell (V)} \\ (\eta_{ct})_a + (\eta_c)_a &= \text{activation polarisation at anode and cathode (V)} \\ (\eta_{ct})_c + (\eta_c)_c &= \text{concentration polarisation at anode and cathode (V)} \\ i &= \text{current (A)} \\ R_i &= \text{internal resistance of cell } (\Omega) \end{aligned}$$

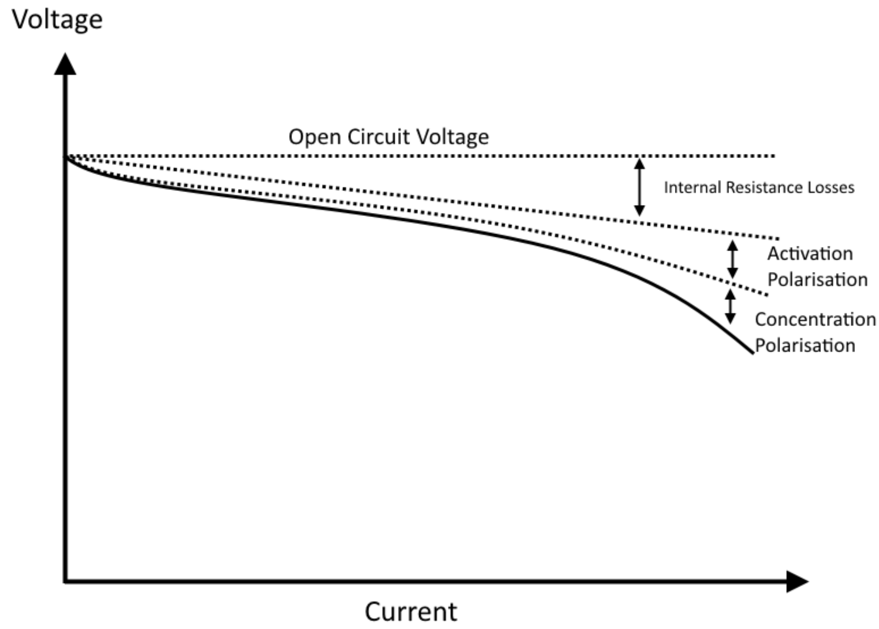

Figure 3.8: Terminal voltage versus current

Polarisation effects can be calculated by several theoretical equations, but in practice it is difficult to determine the parameters for these equations due to the complicated physical structure of battery electrodes [97]. In practise approximations of the dynamic impacts these effects have on the battery are incorporated into the models instead of exact solutions for this phenomena.

3.3.1.7 Activation polarisation effects

When an electrode is immersed in an electrolyte, the charge on the electrode attracts ions of opposite charge. With this attraction, a layer of charge between the electrode and electrolyte interface coalesces [97]. This interface becomes a boundary, with two layers of ions with opposing polarity forming. This causes a build up of electric charge stored at this interface and is referred to as the double layer effect. The capacitance stored by this effect is proportional to the applied voltage and is dependant on the electrode surface area.

Activation polarisation under a steady-state current flow can be described by polarisation curves using the Butler-Vollmer equation (Equation 3.3.9) [99]. This equation involves energy barriers known as activation energies. The change in activation energy drives the reduction and oxidation reactions in the electrochemical kinematics for the battery cell. For an in depth treatise on electrochemistry, refer to Perez [99].

$$j = j_0(\exp \frac{\alpha_a n F \eta}{RT} - \exp \frac{\alpha_c n F \eta}{RT}) \tag{3.3.9}$$

where

$$\begin{aligned}
j &= \text{electrode current density (A/m}^2\text{)} \\
j_0 &= \text{exchange current density (A/m}^2\text{)} \\
E &= \text{electrode potential (V)} \\
E_{eq} &= \text{equilibrium potential (V)} \\
T &= \text{absolute temperature (K)} \\
n &= \text{number of electrons in reaction} \\
F &= \text{Faraday constant} \\
R &= \text{Universal gas constant} \\
\alpha_c &= \text{cathodic charge transfer coefficient (dimensionless)} \\
\alpha_a &= \text{anodic charge transfer coefficient (dimensionless)} \\
\eta &= \text{Activation over-potential (defined as } E - E_{eq}\text{)}
\end{aligned}$$

Double layer effects can be approximated by an RC network in an ECM [98] and is shown in Figure 3.9. In the ECM, the current flows through the battery and is divided at the electrolyte / electrode boundary into a part that flows into the charge transfer reaction and a part that flows into the double layer capacitor. As the capacitance of the double layer effect can only store a limited amount of charge, it is charged and consumed in the first instant of current reversal and is a transient dynamic.

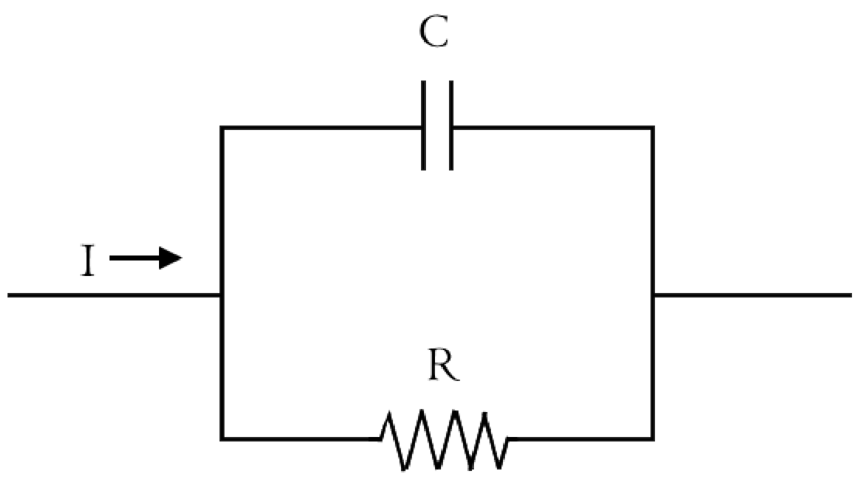

Figure 3.9: Simplified ECM of the double-layer effect

It is important to note that the parameters C and R for the equivalent circuit model are, in reality, not constant and are impacted by the State of Charge, temperature and current. There will also be variations in the parameters of the double layer effect between the positive electrode and the negative electrode due to material and design differences between these two electrodes. These parameters also change with respect to battery age and as such, could be used to predict State of Health if accurately estimated.

3.3.1.8 Mass transport effects

Within batteries the transport of ions, called mass transport, is what generates an electric current. Mass transport processes to and from electrode surfaces is one of the most important aspects to battery dynamics. Mass transport can occur by three processes: (1) convection and stirring, (2) electrical migration in an electric potential gradient, and (3) diffusion in a concentration gradient [97]. The dominant mass transport mechanism in a battery is the third process, diffusion in a concentration gradient. The diffusion process can be described through the use of Fick's law of diffusion. Fick's law of diffusion mathematically describes the diffusion flux as being proportional to the concentration gradient and is represented by:

$$J = -D\frac{\partial \phi}{\partial x} \tag{3.3.10}$$

where

$$\begin{aligned} J &= \text{diffusion flux} \\ D &= \text{diffusion coefficient} \\ \phi &= \text{concentration} \\ x &= \text{position} \end{aligned}$$

Because diffusion processes are the typical mass-transport mechanism in battery systems, phenomena that inhibit the reliable transport of ions to and from reaction sites will affect the useful energy a battery can supply. Differences in the concentrations of reactive materials in the electrode and electrolyte bulk result in a phenomenon called Concentration Polarisation. Concentration polarisation is a detrimental effect that causes losses in the useful electrical energy a battery can supply. Any losses due to concentration polarisation can be added to the resistance in the RC ECM used to approximate the double layer effect.

3.3.2 Model development

With an understanding of the underlying principles of the dynamics of batteries, we can construct an ECM for use in a diagnostic observer. There are many variations of ECMs for batteries, including simple voltage source / resistance models, RC network models and Thevenin equivalent models to name a few [100]. Typically, ECMs for batteries use a capacitor to model the energy storage of the battery, RC networks to model the time dependant transients due to polarisation and resistors to model the voltage response to current draw [92, 101, 102].

Much of the literature on FDI applied to battery systems has been heavily focused on Lithium-Ion batteries due to their associated safety issues with high current applications, such as hybrid and electric vehicles. Nonetheless, the FDI methods applied to Lithium-Ion based battery systems and their results are equally applicable to Lead-Acid batteries, or other battery chemistries for that matter, with the methods typically requiring simply a different parameterisation of the underlying models describing the system. An overview of a few approaches taken in the literature for FDI methods applied to battery systems is given here.

Plett has written a 3-part series on battery management systems for Lithium-ion batteries used in hybrid-electric vehicles. His approach is based on an EKF, with his first paper in the series [103] establishing the essential background material on the problem domain and EKF method utilised. In part two of his series [104], he addresses the approach taken for modelling the battery system and identifying the parameters of this model. Plett provides overviews of the several common methods used in the literature for modelling batteries and then presents five models of increasing complexity describing the Li-ion battery dynamics. The parameters of these models are identified and the subsequent performance of each of the models compared.

Plett finds that adding complexity to the model improves performance, but at an added cost due to the increase in complexity. The final paper in Plett's series [105] details the implementation of the battery management system for the Li-ion battery targeting the problem definition and model developed in the previous papers. Plett explores the use of an EKF for the battery management system and observes that the state estimator exhibits robustness and produces good estimates of the battery system, with SoC error being within a few percent. The author concludes that EKF methods are a good approach for battery management systems.

Chen et al. [93] developed an FDI system for a Li-ion battery string by using a set of Reduced Order Leunberger Observers. Reformulating the state space equations, the authors were able to generate a model that produced a residual per Li-ion battery in the string that was sensitive to only faults in each respective battery. This method was successfully tested against a simulation of the battery system. Their key outcome was that their method had combined the steps of both fault detection and fault isolation, removing the need for a separate fault isolation system. They also observed that the design overall was simpler due to the battery subsystems of the FDI being of reduced order.

Gould et al. [95] sets out to address the problem of a FDI system for an electric vehicle with a battery system that exhibits high dynamics. They begin with a standard ECM for their system, which they call the Randles' lead-acid ECM. The objective is to build a model for accurate SoC and SoH estimation. The key challenge in this case is that for a highly dynamic battery system, static model parameter values using a Randles' lead-acid ECM were not sufficient for describing the battery dynamics or performing online parameter estimation. The authors proposed a remapping of the Randles' lead-acid ECM which they demonstrate in this paper provides better modelling performance for the high dynamics of the battery system. The authors acknowledge that a SoH assessment is a qualitative assessment, typically informed by a deterioration in the estimated battery parameters. Thus the focus of their FDI system is on using online parameter estimation to inform on the SoH of the battery. The authors implemented online parameter estimation against their remapped ECM and observed successful detection of model parameter deterioration against experimental data; concluding that the model provides adequate information to inform a SoH estimation.

Periodically survey papers covering the current state of the art with respect to battery management systems, usually concerned with Lithium-Ion batteries, are released. One such survey of note is by Zhang et al. "A review on prognostics and health monitoring of Li-ion battery" [89]. In this survey the authors cover off important results on battery management systems for Li-ion batteries, including current methods for SoC and SoH estimation.

The modelling approach taken for our diagnostic observer is to use a Thevenin model [100]. A voltage source $E_o(\text{SoC})$, an RC Network (R_o, C_o) and a resistance R are used to construct a model for the battery system. These model parameters E_o, R_o, C_o and R are assumed static but in reality will change over time, dependant on both SoH, SoC and temperature of the battery. Given this relationship between the model parameters and SoH, there is an opportunity to also employ parameter estimation methods for the application of FDI to supplement or succeed a diagnostic observer. However, this approach is not explored in this thesis. The ECM for approximating the battery system is shown in Figure 3.10.

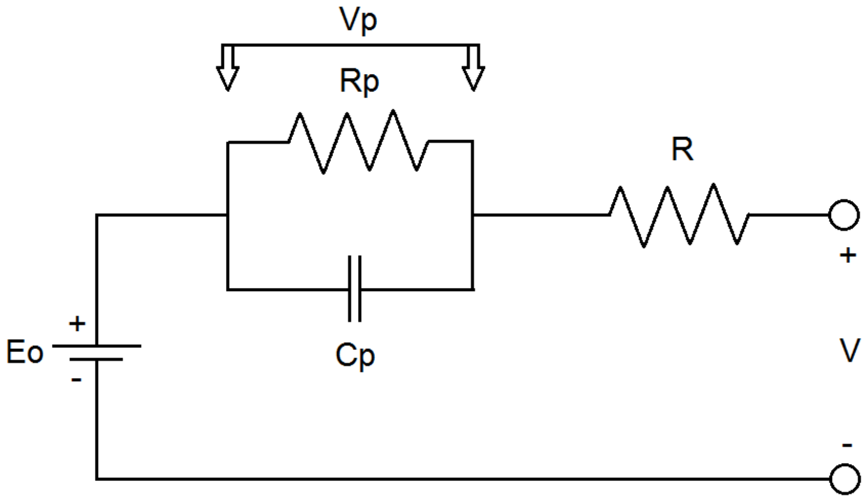

Figure 3.10: Equivalent Circuit Model for a Battery

To begin constructing the model of the battery dynamics, we apply Kirchoff's circuit laws to the ECM.

$$\begin{aligned} \frac{dV_p}{dt} &= -\frac{1}{R_pC_p}V_p + \frac{1}{C_p}i \\ V &= E_o - Ri - V_p \end{aligned} \tag{3.3.11}$$

Where V_p is the voltage across the RC network, R_p and C_p are the resistance and capacitance respectively of the polarisation effects in the battery, R is the internal battery resistance, E_o is the open circuit voltage, V is the terminal voltage and i is the battery current.

We can then discretise the battery dynamic equation (3.3.11) with a timestep Δt using the forward Euler method and obtain the following equations:

$$\begin{aligned} V_{p,k+1} &= V_{p,k} + \Delta t(\frac{1}{R_pC_p}V_{p,k} - \frac{1}{C_p}i_k) \\ V_k &= E_{o,k} - Ri_k - V_{p,k} \end{aligned} \tag{3.3.12}$$

We are also interested in the SoC of the battery and want to include these dynamics into the model. To construct the SoC model, we use the current integration method which is a common approach amongst battery modelling literature, but not without its limitations [106]. The core limitation is drift in the estimated SoC due to additive errors accumulated by the integral. To overcome this limitation, routine recalibration points for the SoC is recommended to counteract drift from the integration of noisy and possibly imprecise current measurements. Such a method of recalibration has not been implemented in this thesis and is considered future work to support a practical implementation.

The SoC dynamics are described by the following equation:

$$S = S_o - \frac{\eta_i}{C_n} \int i(t)\, dt \tag{3.3.13}$$

Where $i(t)$ is the current, S_o is the initial SoC, C_n is the nominal battery capacity and η_i is the coulombic efficiency, which is $\eta_i = 1$ for discharge and $\eta_i \leq 1$ for charge. Discretising the SoC model with a timestep Δt, via the forward Euler method we obtain the following:

$$S_{k+1} = S_k + \Delta t(\frac{-\eta_i}{C_n})i_k \tag{3.3.14}$$

We have now built a model for the battery dynamics. However, the term E_o for the open circuit voltage in our model still needs to be accounted for. The open circuit voltage E_o is a function of SoC, $E_o = f(\text{SoC})$. This relationship between SoC and open circuit voltage is nonlinear, and is represented by a curve fit given by Plett [104]. This curve was fitted to the lead-acid SoC curves developed based on *a priori* knowledge from Linden's Handbook of Batteries [97] and measurements from the physical batteries. Figure 3.11 shows the curve as fitted. A voltage of 18.5v was chosen as the 0% SoC and a voltage of 24.8v chosen as the 100% SoC. The function $E_o = f(\text{SOC})$ is piece-wise to ensure the function equates to reasonable values as $SOC \to 0$ and $SOC \to 1$ as shown in Equation 3.3.15.

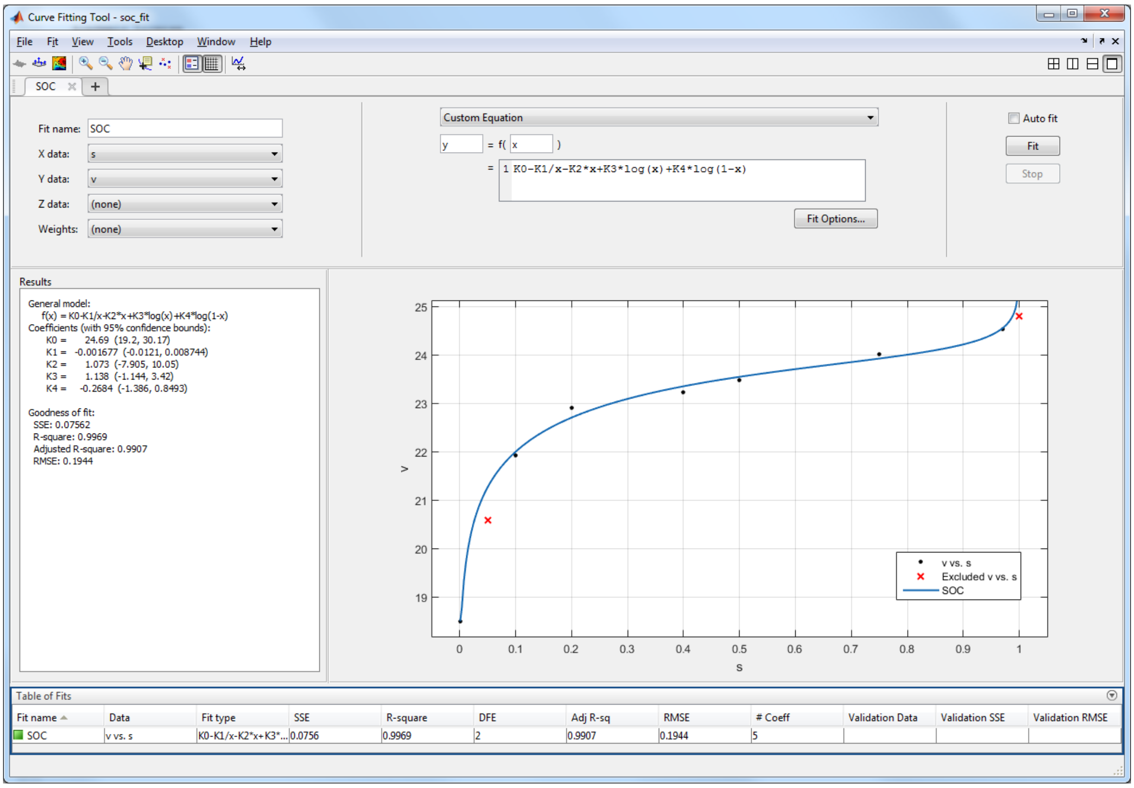

Figure 3.11: $E_o = f(\text{SoC})$ curve fit

$$E_o(\mathrm{SoC}(t)) = \begin{cases} 18.5v & (\mathrm{SoC}(t) \leq 0.01) \\ 24.8v & (\mathrm{SoC}(t) \geq 0.99) \\ K_0 - \frac{K1}{\mathrm{SoC}(t)} + K_2\mathrm{SoC}(t) + & \\ \quad K_3 \ln(\mathrm{SoC}(t)) + K_4 \ln(1 - \mathrm{SoC}(t)) & (0.01 \leq \mathrm{SoC}(t) \leq 0.99) \end{cases} \tag{3.3.15}$$

As we have now described the core dynamics for a battery, we can construct our model as a discrete NSSM. Here we incorporate additive gaussian noise for both the process and measurement dynamics into the model, w_k and v_k respectively.

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k \\ y_k &= h(x_k, u_k) + v_k \end{aligned} \tag{3.3.16}$$

$$\begin{aligned} f(x_k, u_k) &= Ax_k + Bu_k \\ h(x_k, u_k) &= E_o(x_k(2)) - Ru_k - x_k(1) \end{aligned}$$

$$\begin{aligned} w_k &= \mathcal{N}(0, Q) \\ v_k &= \mathcal{N}(0, R) \end{aligned}$$

$$x_k = \begin{bmatrix} V_p \\ \mathrm{SoC} \end{bmatrix} \qquad u_k = \begin{bmatrix} i \end{bmatrix}$$

$$A = \begin{bmatrix} (1 + \frac{-\Delta t}{R_p C_p}) & 0 \\ 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{-\Delta t}{C_p} \\ \frac{-\eta_i \Delta t}{C_n} \end{bmatrix}$$

3.4 Military Vehicle Battery System Test Bench

To support the investigation of FDI methods on the battery system a test bench was constructed. This provides a representative environment where the system dynamics can be identified and FDI methods can be applied in known conditions for evaluation and comparative purposes. The test bench also allows control over various design choices, for instance sensor placement and sampling rates.

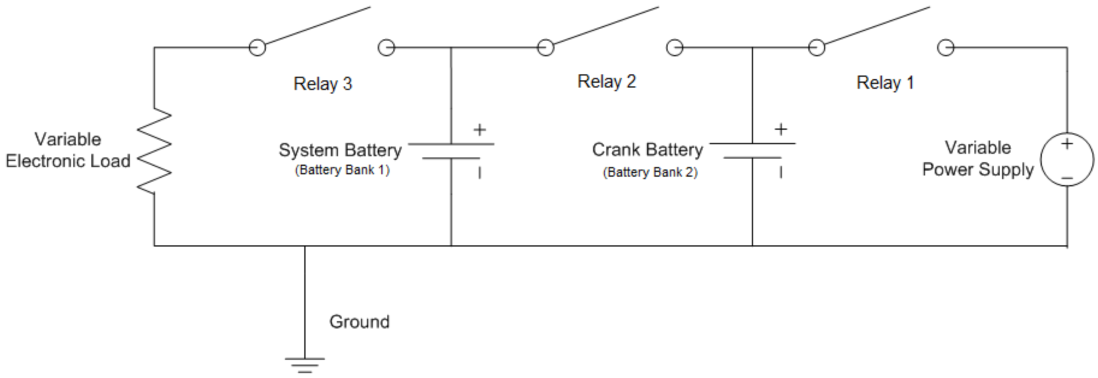

Figure 3.12: Diagram of the military vehicle battery system test bench

It is acknowledged that the test bench, given in Figure 3.12, is only representative of the Military Vehicle Battery System given in Figure 3.1. The limitations of the test bench are that it is not capable of reproducing complex dynamic loads such as engine cranking. Similarly the test bench consists of only one load as opposed to the two in the military vehicle. The test bench is also limited by its inability to influence environmental temperature experienced by the battery. However, what the test bench excels at is in replicating the steady state loads experienced by a military vehicle. It also replicates and provides control over the electrical isolation configurations experienced in the military vehicle. Finally, it provides a platform to rotate between healthy and faulty batteries to produce a data set in which FDI methods can be evaluated against.

3.4.1 Design

The test bench has been constructed as depicted in Figure 3.12 and set up at the University of Sydney in the Australian Centre for Field Robotics (ACFR) laboratory (see Figure 3.13). The electrical schematic for the test bench is given in Figure 3.14. The test bench consists of the following components:

Century 89T Deep Cycle Truck Batteries
These batteries were selected due to their use on some military vehicle platforms and the availability of faulty batteries that had been removed from service as they were incapable of supplying or accepting sufficient charge. These batteries are 12v deep cycle lead-acid, which are connected in serial to provide a total of 24v to the electrical systems on a military vehicle. These batteries have a total capacity of 125 AH at a 20 hr discharge rate.

National Instruments PXIe-1071 Chassis
Data collection is provided by a National Instruments PXIe-1071 Chassis housing a PXI-6123 Data Acquisition Card. The PXIe-1071 Chassis features four PXIe slots on a high-bandwidth backplane allowing for high-performance test and measurement applications. The PXIe-1071 accepts PXI Express modules and standard PXI hybrid-compatible modules. The chassis communicates to a laptop through the LabVIEW software which drives that data collection hardware.

National Instruments PXI-6123 DAQ
The NI PXI-6123 S Series multifunction Data Acquisition (DAQ) module has a dedicated analog-to-digital converter (ADC) per channel for high throughput and increased multichannel accuracy. The NI PXI-6123 is capable sampling rates of 500,000 samples per second per channel with four input ranges from 1.25 V to 10 V, has two 24-bit counter/timers, and eight hardware-timed digital I/O lines. The PXI-6123 is used to sample the analogue sensors, digital sensors and drive the relays in the test bench.

National Instruments SCB-68A
The SCB-68A is a shielded Input / Output connector block with 68 screw terminals which allows easy signal connection to a National Instruments 68-pin or 100-pin Data Acquisition (DAQ) device. In the test bench, the terminal block is used to interface the sensors and relays to the 6123 NI PXI-6123 DAQ. The SCB-68A features a general breadboard for custom circuitry and through hole pads for interchanging electrical components. Typically this breadboard area is used for signal conditioning. In this instance, no signal conditioning was applied, instead relying on the sensors and software to provide the signal conditioning necessary.

ACS758 Current Sensor
The current sensor is a Hall Effect based linear ACS758 Integrated Circuit. This Integrated Circuit is broken out on the DFRobot Current Sensor breakout board (SEN0098). This current sensor is capable of measuring peak voltages of up to 500VDC with a measurement range of ±50A. The sensor sensitivity is 40mV/A with an operating temperature range of -40^{o}C to +150^{o}C. The sensor features an output step rise time of 3μ s with a typical bandwidth of 120 kHz.

Phidgets Precision Voltage Sensor
The voltage sensor is a Phidgets Precision Voltage sensor (1135_0). This sensor uses a voltage divider to provide a linear analogue out signal proportional to the voltage difference across the sensor. The sensor measures a voltage range of ±30V and has an error margin, when calibrated, of ±0.7%. The sensor is Non-Ratiometric and as such, the sensor output does not rely on the sensor input voltage.

DS18B20 Temperature Sensor
The temperature sensor is a digital DS18B20 Integrated Circuit. The assembly is a sealed probe that uses a 1-Wire interface. The DS18B20 Integrated Circuit temperature sensor is powered via the data line and is capable of measuring temperature ranges from -55^{o}C to +125^{o}C. The sensor has an accuracy of 0.5^{o}C accuracy from -10^{o}C to +85^{o}C and due to its digital nature is less susceptible to noise. The sensor converts a temperature into a 12-bit digital word with a latency of 750ms worst case. These temperature sensors are used to measure the ambient temperature, and the temperature of battery banks 1 and 2. To connect the 1-wire DS18B20 temperature sensor a DS9490R USB connector is used and integrated with LabVIEW using the OneWire Utilities addon from Interface Innovations. Given the inability to control the operating temperature of the battery system on the test bench, the sensors were used to ensure temperatures of the batteries were consistent between tests such that temperature was isolated as a non-factor for comparison between healthy and faulty configurations.

HE1AN12 Relays
The Relays are Panasonic High Contact capacity relays (Model: HE1AN12). These relays are switched using a 10V supply across the solenoid, controlled via a Darlington Array and PXI-6123 DAQ. The relays are capable of switching 30A and up to 277V.

ULN2803A Darlington Array
The relays are driven by a Darlinton Array as shown in Figure 3.14. The Darlington Array inputs are controlled by Digital Outputs from the PXI-6123 DAQ.

E3620A Dual Power Supply
An Agilent Dual power Supply (E3620A) is used to provide 5V and 10V rails to the sensors and Darlington Array Relay Controller on the Test bench.

HP 6674A Power Supply
A HP 60V 35A Power Supply (6674A) is used to provide a 27V rail for the Test Bench, acting as a representative Alternator for a Military Vehicle Battery System.

B&K 8514 Electronic Load
The resistive load for the test bench was a B&K 8514 electronic load capable of providing a load of 1200W. This was used to create some time varying load dynamics which helped in the identification of model parameters for the battery system. It was also used to provide a representative steady

state load on the test bench for the purposes of providing a data set representative of what would be experienced on a military vehicle.

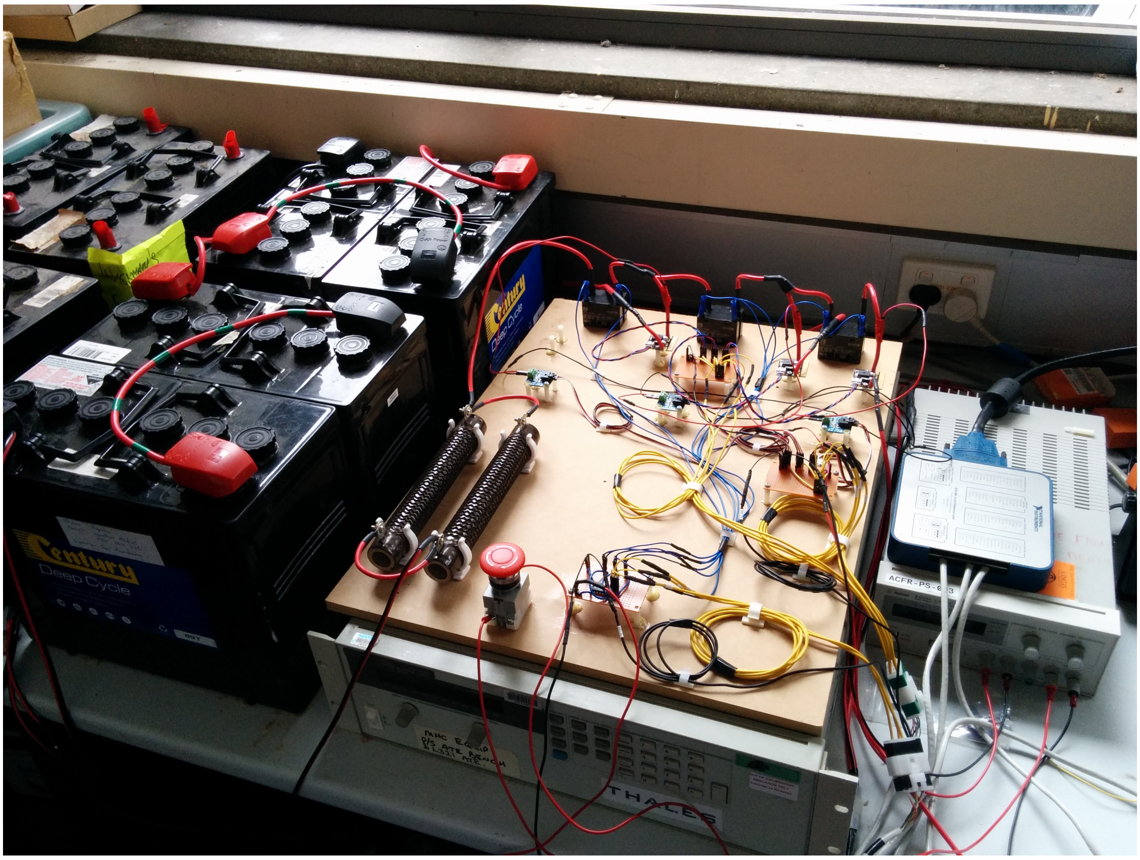

Figure 3.13: Military Vehicle Battery System Test Bench in the ACFR laboratory at the University of Sydney

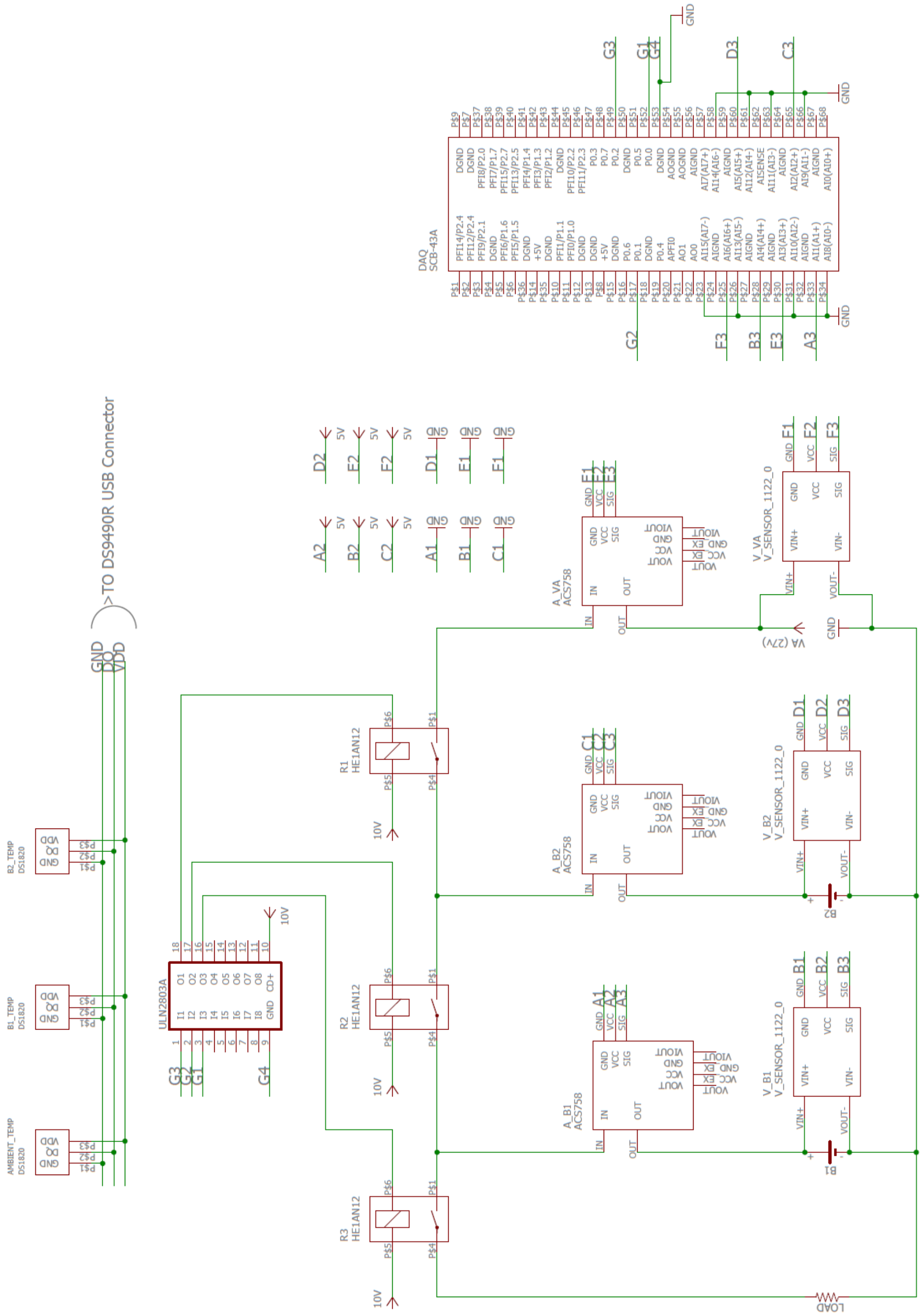

Figure 3.14: Military Vehicle Battery System Test Bench Schematic

3.4.2 Safety

The safety of the military vehicle battery system test bench has been assessed using a standard Job Safety and Environmental Analysis (JSEA) safe work method. the results from this hazard analysis are captured in Table 3.1. The risk ratings assigned for all identified hazards are using the risk rating reference provided in Figure 3.15.

Table 3.1: Test Bench Safety Analysis

Activity	Hazard	Initial Risk Rating	Risk Control Measure	Final Risk Rating
Storage of battery	Exposed terminals, leading to accidental shorting and electrocution risk / fire risk	C3	Caps for terminals of batteries that aren't in use. This will ensure terminals aren't exposed.	E4
	Damaged batteries leading to gassing	D3	Routine inspections of the battery casing to ensure no damage or gassing present. Ensure no overcharging of battery occurs. Operate batteries in a well-ventilated area	E4
	Damaged batteries leading to sulphuric acid spill	D3	Routine inspections of the battery casing to ensure no damage present. Ensure a spill kit resides nearby.	E4
Use of test environment	Electrocution risk whilst using test environment	C2	Ensure all wiring is insulated. Area is cordoned off when in use. When not in use circuit is isolated via emergency stop.	E2
	Over voltage leading to gassing	C2	Ensure all wiring is insulated. Circuits are checked by ACFR lab representative. Emergency stop is available to disconnect circuit in the case of over voltage.	D3
	Fire from shorting or catastrophic failure	C2	Ensure all wiring is insulated. When not in use circuit is isolated via emergency stop. Appropriately rated fire extinguisher at hand in case of fire	D3
Transport of batteries	Injury due to carrying heavy weight of individual batteries (approx. 30kg)	C4	Ensure all transport of batteries is done with a trolley to reduce weight being carried	E4

	Minor (5)	Serious (4)	Major (3)	Critical (2)	Catastrophic (1)
Almost Certain (A)	High	High	Very High	Extreme	Extreme
Likely (B)	Medium	High	High	Very High	Extreme
Possible (C)	Medium	Medium	High	High	Very High
Unlikely (D)	Low	Medium	Medium	High	High
Rare (E)	Low	Low	Medium	Medium	Medium

Figure 3.15: Risk rating reference

3.4.3 Data Collection

With the test bench constructed as described in Section 3.4.1, a LabVIEW model was created to interface with the National Instruments data acquisition hardware. A Virtual Instrument was created (shown in Figure 3.16) to operate the test bench. The LabVIEW model reads the analogue sensor data (voltage, current), the digital sensor data (temperature) and drives the relay circuit. The LabVIEW model also writes the sensor data to a log file which is later processed in MATLAB for the development and evaluation of the FDI schemes. The LabVIEW model is broken into two parts, a one-wire interface to read the digital temperature sensor data (see Figure 3.17) and an analog/Digital Out interface that reads the analog sensor data and drives the relays (see Figure 3.18).

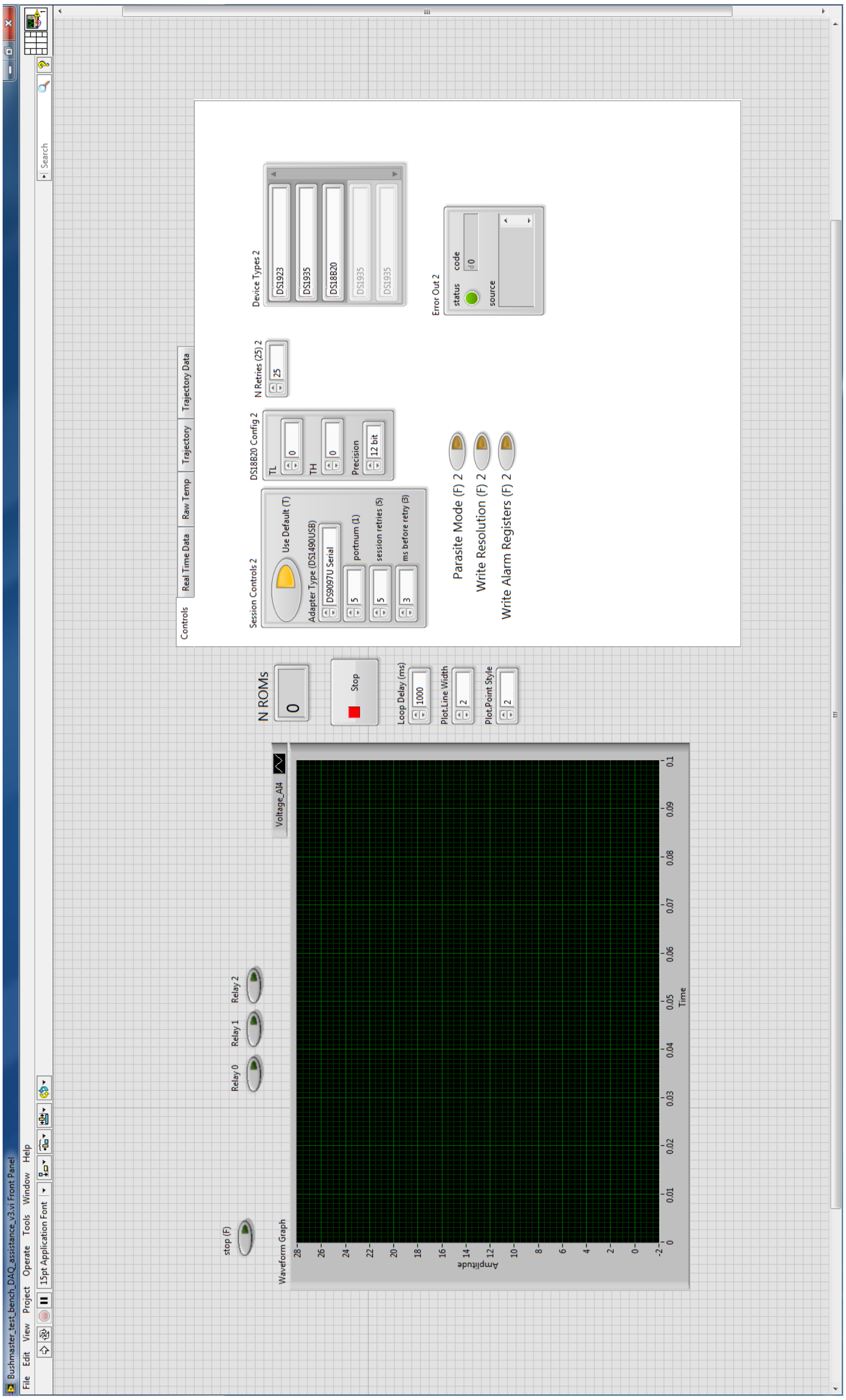

Figure 3.16: LabVIEW virtual instrument for the military vehicle battery system test bench

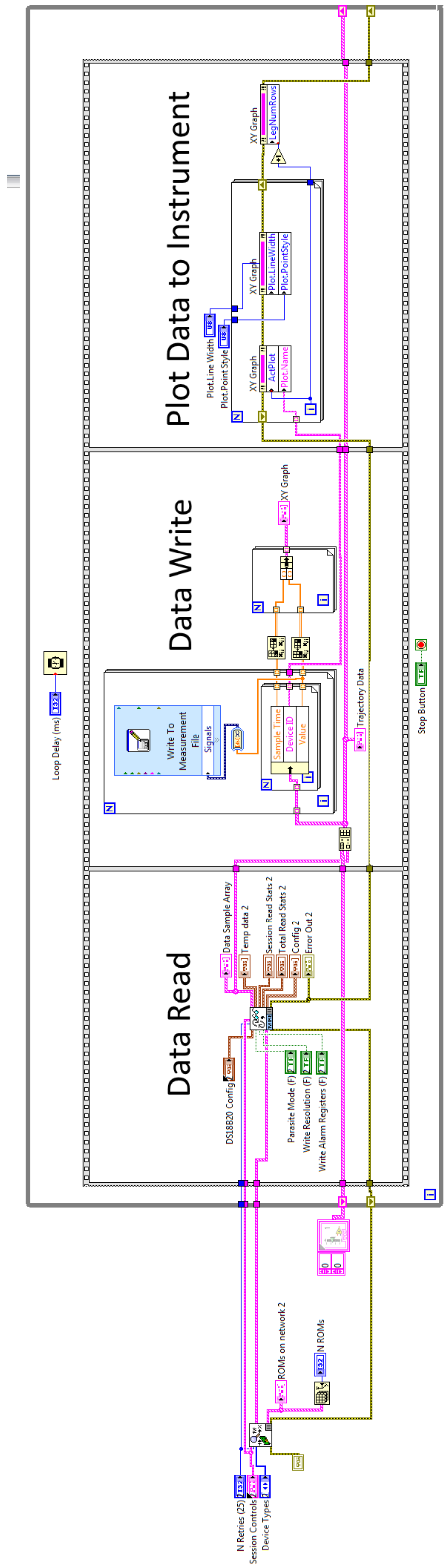

Figure 3.17: LabVIEW digital sensor collection program for the military vehicle battery system test bench

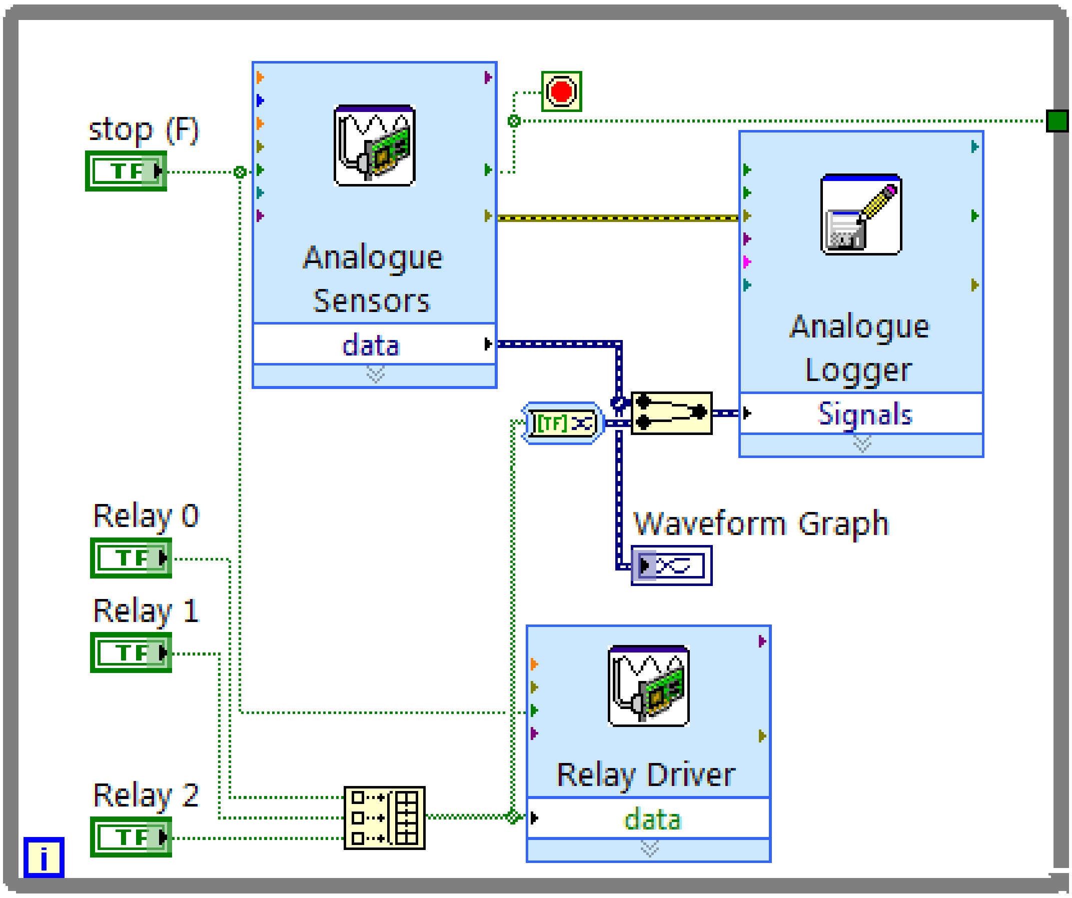

Figure 3.18: LabVIEW analogue sensor collection and relay driver program for the military vehicle battery system test bench

With the LabVIEW model constructed, a series of tests were performed and recorded using the test bench. An 18 minute test sequence was performed several times which featured an alternating relay progression that stimulated all possible configurations of the test bench. This relay progression is given in Table 3.2. The relay numbering is identified in the test bench model and schematic (see Figures 3.12 and 3.14 respectively). This test was executed with a rotation of healthy and faulty batteries integrated into the test bench as shown in Table 3.3. These tests resulted in a data set where the varied parameter was battery health, with data collected over all possible relay configurations for the test bench.

Table 3.2: Dataset Relay Progression

Relay 1	Relay 2	Relay 3	Time (s)
OFF	OFF	OFF	0
ON	OFF	OFF	120
OFF	ON	OFF	240
OFF	OFF	ON	360
OFF	ON	ON	480
ON	ON	OFF	600
ON	OFF	ON	720
ON	ON	ON	840
OFF	OFF	OFF	960
OFF	OFF	OFF	1080

Table 3.3: Test Bench Datasets

	Battery Bank 1		Battery Bank 2	
Dataset ID	Battery 1	Battery 2	Battery 1	Battery 2
1	Healthy	Healthy	Healthy	Healthy
2	Healthy	Not Healthy	Healthy	Healthy
3	Healthy	Healthy	Healthy	Not Healthy
4	Healthy	Healthy	Not Healthy	Not Healthy
5	Not Healthy	Not Healthy	Healthy	Healthy

The LabVIEW data was imported into MATLAB using the `lvm_import` script which was created by M. A. Hopcroft and available for download on the MATLAB File Exchange (`http://www.mathworks.com/matlabcentral/fileexchange/19913-lvm-file-import/content/lvm_import.m`).

A sample of the data imported into MATLAB from the military vehicle battery system test bench, in this case Dataset 1, is given in Figure 3.19.

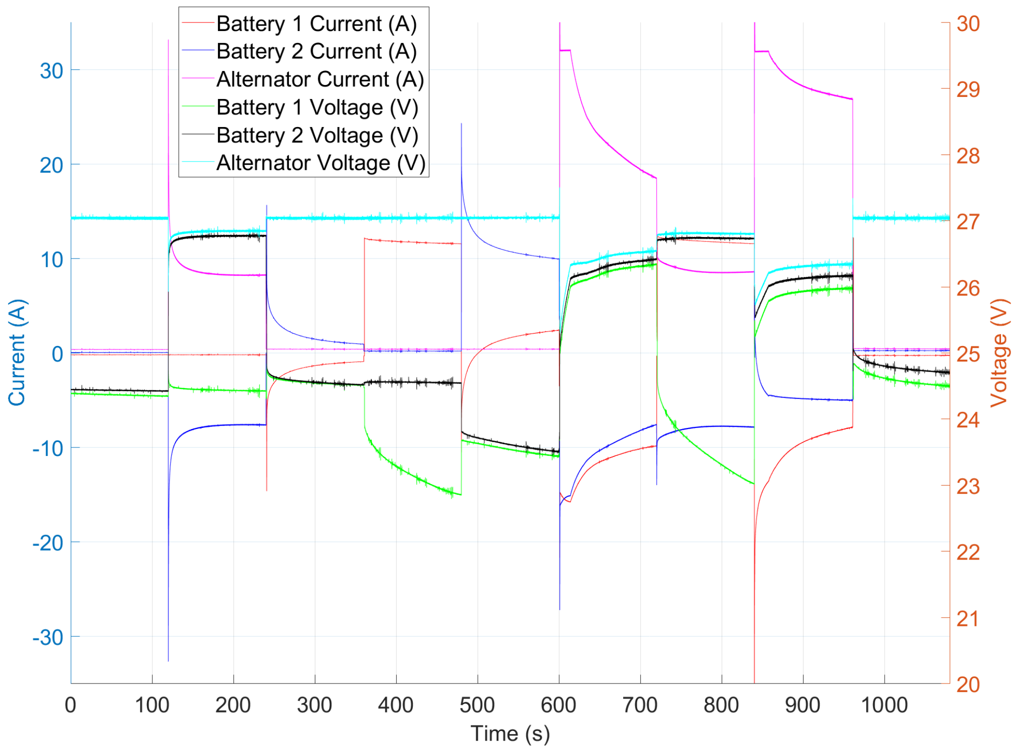

Figure 3.19: Military vehicle test bench dataset 1

3.5 Diagnostic Observer

With a model defined for a battery given by Equation 3.3.16 we can now construct a diagnostic observer. Two approaches for constructing a diagnostic observer of the military vehicle battery system have been considered here. An approach utilising an EKF with hard constraints and an approach utilising a MHE. These approaches were both implemented and executed on the data collected from the military vehicle battery system test bench to compare and evaluate their effectiveness for fault detection.

3.5.1 Extended Kalman Filter with hard constraints

Given the model for the battery (refer to Equation 3.3.16) is nonlinear in the measurement function, we employ an EKF to linearise about the estimated state trajectory $\hat{x}_k$.

We initialise the EKF as follows:

$$\begin{aligned} \hat{x}_0^+ &= \mathbb{E}[x_0] \\ P_0^+ &= \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \tag{3.5.1}$$

For $k = 0, 1, 2, \ldots$ we calculate the following.

(a) Compute the time update for both the state estimate and the estimation error covariance:

$$\begin{aligned}\hat{x}_{k+1}^{-} &= A\hat{x}_k^{+} + Bu_k \\ P_{k+1}^{-} &= AP_k^{+}A^T + Q_k\end{aligned} \tag{3.5.2}$$

(b) Compute the following partial derivative matrices:

$$\begin{aligned}H_{k+1} &= \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_{k+1}^{-}} \\ M_{k+1} &= \left.\frac{\partial h}{\partial v}\right|_{\hat{x}_{k+1}^{-}}\end{aligned} \tag{3.5.3}$$

(c) Compute the measurement update of the state estimate and the estimation error covariance:

$$\begin{aligned}L &= P_k^{-}H_{k+1}^T(H_{k+1}P_{k+1}^{-}H_{k+1}^T + M_{k+1}R_{k+1}M_{k+1}^T)^{-1} \\ \hat{x}_{k+1}^{+} &= \hat{x}_{k+1}^{-} + L[y_{k+1} - h(\hat{x}_{k+1}^{-}, u_{k+1})] \\ P_{k+1}^{+} &= (I - LH_{k+1})P_{k+1}^{-}\end{aligned} \tag{3.5.4}$$

We know based on the definition of SoC ($0 \leq \text{SoC} \leq 1$) and the physical limitations of the double-layer effect [97] we can apply hard state constraints to ensure our model state is constrained to a realistic operating region. This will ensure our diagnostic observer residual r_t diverges in the presence of faults, and the EKF won't converge to inconsistent states as it tries to reconcile measurements from a faulty system.

(d) Apply hard constraints on state estimates using the projection approach [107]:

$$\tilde{x}^{-} = \arg\min_{\tilde{x}^{-}}\left(\frac{1}{2}\tilde{x}^{-T}H\tilde{x}^{-} - \hat{x}^{+T}H\tilde{x}^{-}\right) \tag{3.5.5}$$

such that

$$C_{lb} = \begin{Bmatrix}-2.5 \\ 0\end{Bmatrix} \leq \tilde{x} = \begin{bmatrix}V_c \\ \text{SOC}\end{bmatrix} \leq C_{ub} = \begin{Bmatrix}2.5 \\ 1\end{Bmatrix}$$

$$H = \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}$$

For the hard constraint, setting the weighting matrix $H = I$ is equivalent to projecting the unconstrained state to the limits C_{lb} and C_{ub}. This allows for a practical implementation to instead saturate the states at the limits C_{lb} and C_{ub}, avoiding the need to solve an optimisation problem. This reduces the computational complexity significantly. An implementation the EKF based diagnostic observer using the above formulation is given in Algorithm 1.

Algorithm 1 EKF Diagnostic Observer

procedure EVALUATE BATTERY HEALTH
    $\hat{x}_0^+ = \mathbb{E}(x_0)$
    $\Sigma_{\tilde{x},0}^+ = \mathbb{E}[(x_0 - \hat{x}_o^+)(x_0 - \hat{x}_o^+)^T]$

    for k = 0,1,2,... **do**
        $\hat{x}_{k+1}^- = A\hat{x}_k^+ + Bu_k$
        $P_{k+1}^- = AP_k^+A^T + Q$

        $H_{k+1} = \frac{\partial h}{\partial x}\Big|_{\hat{x}_{k+1}^-}$
        $M_{k+1} = \frac{\partial h}{\partial v}\Big|_{\hat{x}_{k+1}^-}$

        $L = P_{k+1}^- H_{k+1}^T [H_{k+1} P_{k+1}^- H_k^T + M_{k+1} R M_{k+1}^T]^{-1}$
        $\hat{x}_{k+1}^+ = \hat{x}_{k+1}^- + L[y_{k+1} - h(\hat{x}_{k+1}^-, u_{k+1})]$
        $P_{k+1}^+ = (I - LH_{k+1})P_{k+1}^-$

        if $\hat{x}_{k+1}^+ > C_{ub}$ **then**
            $\tilde{x}_{k+1} = C_{ub}$
        else if $\hat{x}_{k+1}^+ < C_{lb}$ **then**
            $\tilde{x}_{k+1} = C_{lb}$
        else
            $\tilde{x}_{k+1} = \hat{x}_{k+1}^+$
        end if

        $r_k = y_k - h(\tilde{x}_{k+1}, u_k)$

        if $|r(t)| \geq \gamma$ **then**
            $B_{fault,k} = 1$
        end if

        if $\sum(B_{fault,k-Z:k})/Z == 1$ **then**
            Report Battery Fault
        end if
    end for
end procedure

Algorithm 1 has been implemented in MATLAB and executed against the datasets (Table 3.3) generated from the military vehicle battery system test bench.

3.5.2 Moving Horizon Estimator

A MHE approach for our battery system diagnostic observer is attractive due to its ability to readily incorporate constraints, such as constraining the SoC between its defined values, and voltage losses due to polarisation (V_p) based on our physical knowledge of batteries. We can construct a MHE estimator for our diagnostic observer with a sliding window of size N as follows:

$$\arg \min_{\hat{x}_{\{k-N:k\}}} \Phi(\bar{x}_{k-N}, y_{\{k-N:k\}}) \tag{3.5.6}$$

$$\Phi = ||\hat{x}_{k-N} - \bar{x}_{k-N}||_P^2 + \sum_{i=k-N+1}^{k-1} ||\hat{x}_{i+1} - f(\hat{x}_i, u_i)||_Q^2 + \sum_{i=k-N+1}^{k-1} ||y_i - h(\hat{x}_i, u_i)||_R^2$$

subject to:

$$|V_p| \leq 2.5; \quad 0.001 \leq \mathrm{SoC} \leq 0.999; \quad \mathrm{SoC}_{k+1} = \mathrm{SoC}_k + \frac{-\eta_i \Delta t}{C_n}$$

Where $\bar{x}_{k-N}$ is the arrival prediction estimated in the previous timestep $k-1$. P is a weighting matrix penalising the distance in the arrival cost from the previous sliding window estimate in our cost function. Q is a weighting matrix penalising the error in the state update component in our cost function. R is a weighting matrix penalising the error in the measurement update component of our cost function. Through out this thesis, given a symmetric positive definite matrix M and a vector z; $||z||_M \triangleq (z^T M z)^{1/2}$.

The constraints placed on SoC is due to its definition, being between 0 and 1, and its update is constrained to the coulomb counting method defined in Equation 3.3.13. It should also be noted that the piece-wise function $Eo(\cdot)$ (3.3.15) was simplified to remove the hard saturation at 0 and 1. This was due to limitations with our implementation of our optimisation solver, instead the hard constraints for SoC were set to 0.001 and 0.999 so that the values of $Eo(\cdot)$ at the hard limits are consistent with the saturation limits for the piece-wise $Eo(\cdot)$ function. The constraint placed on V_p, the voltage lost due to polarisation effects, has been determined based on nominal polarisation dynamics of the lead-acid batteries used in this Battery System [97].

With the above formulation, we can implement a diagnostic observer based on a MHE approach as given in Algorithm 2.

Algorithm 2 MHE Diagnostic Observer

procedure EVALUATE BATTERY HEALTH
  for k = N+1, N+2, N+3, N+4, ... **do**
    if Uninitialised **then**
      SoC $= E^{-1}(V_k)$
      $V_{p,k} = 0$
    end if

    Solve $\arg\min_{\hat{x}_{k-N:k}} \Phi(\bar{x}_{k-N}, y_{\{k-N:k\}})$

    $r(t) = Y_k - h(\hat{x_k}, u_k)$

    if $|r(t)| \geq \gamma$ **then**
      $B_{fault,k} = 1$
    end if

    if $\sum(B_{fault,k-Z:k})/Z == 1$ **then**
      Report Battery Fault
    end if
  end for
end procedure

Algorithm 2 has been implemented in MATLAB using YALMIP [108] and the `fmincon` nonlinear

Table 3.4: Diagnostic Observer Results

	MHE		EKF	
Dataset ID	Detection Rate	Missed Detection Rate	Detection Rate	Missed Detection Rate
1	N/A	N/A	N/A	N/A
2	65.94%	34.06%	65.75%	34.25%
3	33.13%	66.87%	33.13%	66.87%
4	99.60%	0.40%	99.59%	0.41%
5	99.60%	0.40%	99.60%	0.40%

solver. The algorithm was executed against the datasets (Table 3.3) generated from the military vehicle battery system test bench.

3.5.3 Comparison

Running the diagnostic observers over the dataset defined in Table 3.3 we obtained the results listed in Table 3.4.

For the control test data (Dataset ID 1), both diagnostic observers successfully estimated the system state and recorded no faults. As we progress through increasingly faulty configurations of the switched battery system our detection rates improved.

The diagnostic observers were capable of inferring battery faults through the divergence caused by the battery's inability to both accept and deliver charge due to its degraded state. This led to a lesser current being supplied to or delivered by the faulty battery which resulted in a drop in nominal voltage, causing a discrepancy between the diagnostic observer and measurement of the actual battery, producing a non zero residual.

We observe that the fault detection performance between the MHE diagnostic observer and the EKF diagnostic observer are near identical for all cases. This is largely expected given both diagnostics observers are driven via the same nonlinear model.

It is also noted that without the implementation of hard constraints, the performance of the EKF was unsatisfactory as the model incorrectly converged to unrepresentative internal states as shown in Figure 3.20. This is due to the battery model being incapable of reconciling the internal state into anything meaningful given sensor data from a faulty battery. This was seen in datasets 4 and 5 where the polarisation voltage V_p became arbitrarily large in an attempt to reconcile the faulty battery sensor data. Without the application of a hard constraint on the EKF, the residual would not deviate due to a faulty battery.

Alternatively, in the absence of hard constraints for the EKF a tolerance could instead be applied to the internal state V_p in line with the physical bounds understood for the polarisation voltage. This would result in identical fault detection performance as the fault threshold applied to V_p in essence is effectively acting as the hard constraint, signaling that the battery is faulty. But given the implementation of a hard constraint is trivial, the adoption of one for the battery diagnostic observer allows us to be consistent with the approach for model-based FDI as shown in Figure 2.1. We conclude that both diagnostic observers were able to successfully detect battery faults given adequate stimulus to the battery system.

We also compared the EKF and MHE diagnostic observers for sensitivity to faults by comparing the residuals r_t and how tight the tolerance γ can be set. The log values of the residuals for both diagnostics observers were plotted in Figure 3.21. We can see from this comparison that the

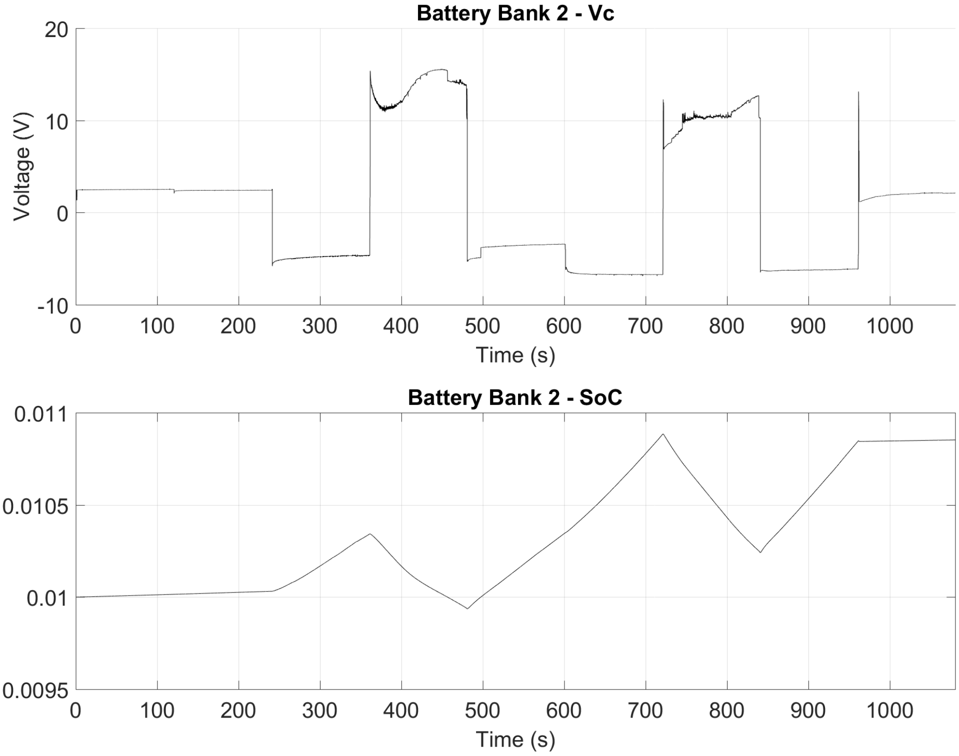

Figure 3.20: EKF (no constraints) - Dataset 4 - Unrepresentative polarisation voltage V_p state

MHE diagnostic observer can have an overall tighter tolerance γ. The MHE tolerance can be set to $\gamma = e^{-5.1} = 0.0061$ versus a tolerance for the EKF of $\gamma = e^{-4.2} = 0.0149$. Thus the MHE diagnostic observer is more sensitive to faults, given an accurate estimate of SoC is achieved. This didn't show in our results due to the faulty batteries used in the military vehicle battery system test bench being so severe that they weren't on the threshold to see a difference between the MHE and EKF diagnostic observers.

A general observation we make is that the diagnostic observers perform worse when attempting to detect a single faulty battery in a bank whilst connected in parallel with a healthy battery, but not the alternator. This is due to the healthy battery not providing an adequate potential difference to expose the faulty battery's inability to accept charge, along with the healthy battery masking the faulty battery by providing the required charge given a connected load. This is observed through the lower detection rates in Dataset IDs 2 and 3.

A comparison of the execution time for each of the diagnostic observers is given in Figure 3.22. Here the frame timing, the time taken to execute the algorithm through one iterative loop, is presented for each observer in a histogram. Here we can observe that the EKF outperforms the MHE by two orders of magnitude, which is to be expected given the differences in computational complexity. The EKF on average took 0.34 ms to execute, whilst the MHE took on average 66.2 ms to execute, with worst case performance 211.7 ms. The EKF had a worst case scenario of 0.64 ms. We also observe that the EKF algorithm was more consistent in execution time due to the relative computational simplicity compared with the MHE.

Given that the ability to detect a battery fault is dependant on having adequate stimulus to excite the system to compare against the nominal model, it would be advisable to regularly poll the batteries and measure their response for performance during periods of low current draw. This

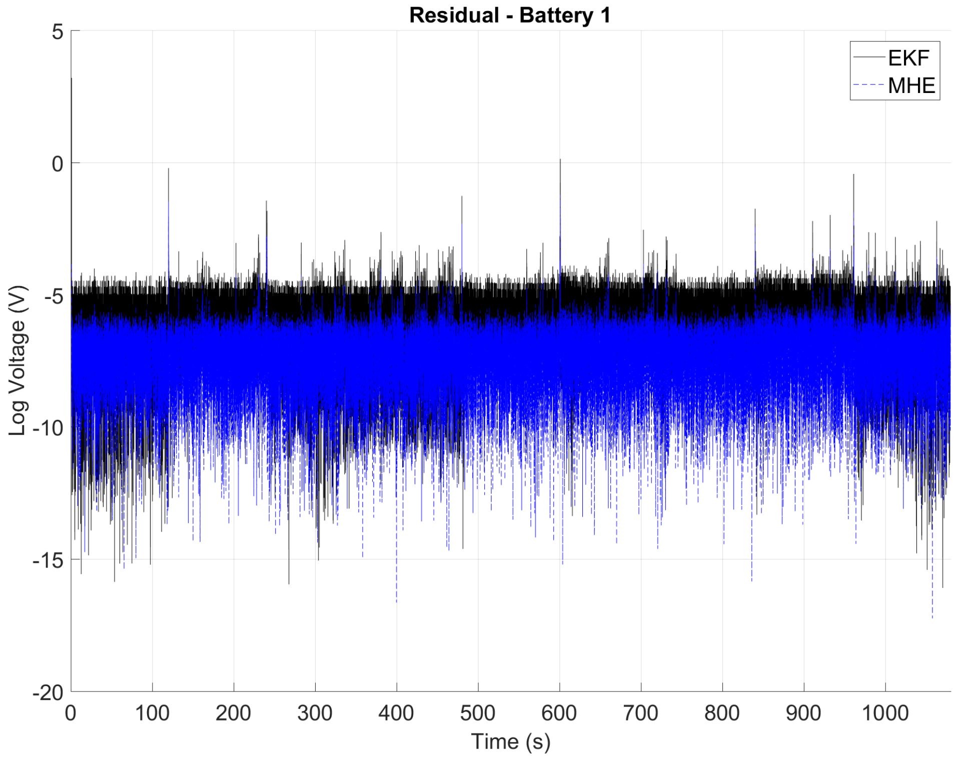

Figure 3.21: Residual analysis between EKF and MHE on Dataset 1

provides a routine estimation of battery health and would provide superiour fault detection in the case of inadequate system stimulus. By polling the batteries during no current draw, or even during steady current draw, this opens the paradigm for designing an optimal probe signal to stimulate the battery adequately to provide maximum information for the evaluation of its health. Such an approach has been used successfully for fault detection in other systems [109].

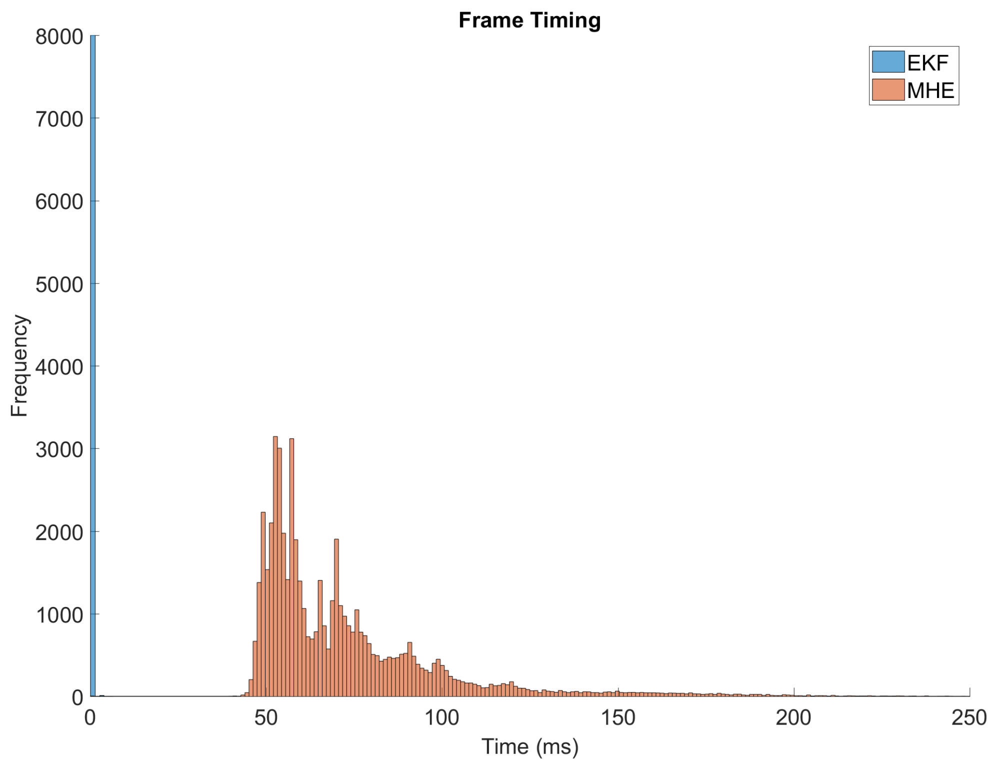

Figure 3.22: Histogram of the frame timing for the implemented fault detection algorithms

3.6 Conclusions and Future Work

Here we summarise the key results from the work presented under Chapters 2 and 3. Chapter 2 provided a comprehensive overview of FDI, first introducing a chronological review on the history of model based FDI methods, then followed by an overview of the leading FDI methods in the literature. Chapter 3 concerned itself with the development of a fault detection system for a military vehicle switched battery system and constitutes the main contribution from the first half of this thesis.

A dynamic model was developed, refer to Section 3.3, for a military vehicle switched battery system, defined in Section 3.2, using an ECM approach. The model developed is a design which features routinely in the literature and the results herein further validate its applicability as a model which adequately performs for use in fault detection.

A test bench was developed, refer to Section 3.4, which allowed for a controlled exploration of a fault detection system for a military vehicle switched battery system. This provided the means to generate data sets with known faults in the battery system.

Subsequently, we explored two approaches for developing a diagnostic observer for the switched battery system, an EKF approach (refer to Section 3.5.1), and a MHE approach (refer to Section 3.5.2). These two diagnostic observers were compared against the data sets produced by the switched battery system test bench.

From the comparison, it was found (refer to Section 3.5.3) that the EKF diagnostic observer and the MHE diagnostic observer performance for fault detection in the switched battery system were near identical across all data sets. The MHE approach provided a more natural form for describing the constraints applied to the model states. However, the EKF approach was equally capable of being extended to incorporate constraints and provides a significantly lesser computational burden than the MHE.

Without constraints applied to the EKF the diagnostic observer would converge to unrepresentative internal states with the model being incapable of reconciling sensor data from a faulty battery. This led to situations where the residual would not deviate and hence be incapable of signalling a fault. In such scenarios, thresholding the estimated states, using knowledge on the limits of these values, as opposed to the residual of the observer would be an adequate solution. However, to ensure consistency with the FDI literature constraints were placed on the estimated states for the EKF to ensure the diagnostic observer residual would deviate adequately when presented with signals from a faulty battery.

The sensitivities to faults between the EKF and MHE were also compared and from an analysis the MHE is able to maintain a much tighter tolerance on the residual for fault detection. This would need to be further explored with a data set which would allow the MHE to exploit these tighter tolerances to validate whether this is the case.

It was also noted that the ability to detect a faulty battery was dependant on having adequate stimulus to excite the system to compare against the nominal model. It was observed in Dataset 2 and Dataset 3 that there was a higher missed detection rate because of inadequate stimulus applied to the faulty battery to excite the system for fault detection. This was due to certain isolation switch configurations leading to the healthy battery masking the faulty battery. Such a limitation could be overcome with the application of a probing signal to stimulate the battery for health evaluation in situations of inadequate stimulus. This is certainly worth further investigation as the use of probing signals to stimulate fault detection has shown promise in other dynamic systems [109].

This work was summarised and presented in a conference paper by the Author [41]. Additionally, resultant from this work a report was written and provided to Thales Australia's Protected Vehicles business unit with findings and recommendations on fault detection in a switched battery system, and more generally opportunities for FDI application in military vehicles.

Future work with respect to the military vehicle switched battery system would look to extend the model to incorporate temperature effects on the battery parameters. Similarly, a longer duration of tests whereby the batteries's eventual deterioration from controlled deep discharges and cycling should be explored to evaluate and refine the diagnostic observers. The EKF and MHE diagnostic observers should then be subsequently compared with a data set wherein the battery deterioration is on the limit of what would be considered a fault. This would inform if one method provided earlier detection of a faulty battery over the other.

Additionally, online parameter estimation of the battery model is also worth investigation as it would augment the performance of the fault detection algorithm as this has been shown in the literature to provide an excellent vector of information relating to the battery SoH [95]. The exploration of probing signals to stimulate the battery for health evaluation in periods of poor utility, as experienced in certain configurations of the switch battery system in our test bench, is also an interesting prospect worth investigation.

Chapter 4

Nonlinear State Estimation in Bearing Only Target Tracking

This chapter of the thesis concerns itself with the implementation and evaluation of a nonlinear state estimator for the tracking of a target using a bearing only buoy field sensor network. A simulation is established to provide a representative scenario to explore the performance of the nonlinear state estimator for target tracking. Two state estimation approaches are implemented and compared, an EKF and an Unscented Kalman Filter (UKF) using the same maneuver and measurement model.

This chapter is broken down into the following sections: Section 4.1 provides a brief history on the origins of target tracking algorithms. Section 4.2 details the simulation developed and the assumptions made, it then provides a derivation of both state estimation approaches implemented, the EKF and the UKF. Section 4.3 compares the performance of the EKF and UKF state estimators resultant from the simulation.

4.1 Brief history of target tracking algorithms

Modern target tracking algorithms have their origins in the development of radar systems. With the need driven by the air threat posed to Britain by Germany in the lead up of World War II, Britain began researching the application of electromagnetic waves for threat detection. In 1935 a proof of concept radar, known as the Daventry Experiment, demonstrated the efficacy of the concept. By 1940 Britain had developed and deployed both surveillance radars and mobile rotary radars as part of the war effort to maximise their situational awareness and provide early warning of air raids. To reduce the complexity of interpreting the radar results a cathode-ray-tube coupled to the radar antenna rotation was developed into what is known as the Plan Position Indicator (PPI) [110].

This provided the radar operator the means to visualise the detected target and observe its position over time on the PPI. Operators would manually connect these observations on the PPI producing a target track over time to monitor the targets path and range [111]. It was apparent that in a complex environment with multiple targets and erroneous signal returns, the task of manually tracking a target using the PPI was challenging. The need for Multi-Target Tracking (MTT) algorithms was evident [112].

It was first recognised by Wax [113] in 1955 that the radar target tracking problem posed was similar to the problem of identifying and tracking particle positions in nuclear physics. Wax approached the issue of radar target tracking using the terminology and concepts of particle physics with three distinct steps, birth, maintenance and the death of a particle. In this instance the birth (track formation) of the track, observation of the life of the track (track maintenance), and then the death of the track (track rejection). Wax then proposed mathematical models to facilitate this process.

A major break through in MTT was contributed by Sittler [114] in 1964 with the development of a Bayesian formulation of the MTT problem that provided background for later developments. Similarly, the arrival of the Kalman filter [16] for optimal state estimation provided the tools needed for modern MTT systems. However, It was not until the early 1970s that the combination of data associativity methods and Kalman filtering theory was applied to MTT [115].

Independent to the development of radar, a similar need arose for the capability to detect and track targets in an underwater environment. This need was due to the emergence of the submarine in World War I which presented a significant asymmetric threat to established naval fleets [116]. This led to the development of sonar systems, which is the underwater analogue to radar. However, instead of using electromagnetic energy to detect and track targets, sonar uses acoustic energy due to its long propagation distances through water as the medium [117]. The efficacy of sonar systems was enabled by the development of the hydrophone, a transducer that converts underwater acoustic energy into an electrical signal. The hydrophone was invented in Boston, Massachusetts by Reginald A. Fessenden in 1912 whilst working for the Submarine Signal Company.

During World War 1, the threat posed by submarines to the allied naval capability and supply lines led to the formation of the Allied Submarine Detection Investigation Committee (ASDIC) which had the mandate, unsurprisingly, of researching the detection of submarines. It was through the ASDIC shipboard sonar systems emerged and development on refining these sonar systems continued through the 1920s, 1930s and 1940s. Interestingly, the term sonar (Sounding Navigation and Ranging) didn't emerge until 1942 when it was coined by Frederick Vinton Hunt, then director of the Harvard Underwater Sound Laboratory [116]. Both World War II and the resulting Cold War saw a continued need for sonar systems and the improvement of detection methods for underwater threats. This need continues today with modern militaries possessing sophisticated underwater capabilities which present asymmetric threats to Australia's national security [1]. For a more in-depth account on the development of active sonar refer to D'Amico's "A Brief History of Active Sonar" [116].

Sonar systems can be categorised by two different types, active and passive. Active sonar uses an acoustic pulse and then records the reflections to detect, range and track targets. This approach is similar to radar but differs in its use of acoustic energy as opposed to electromagnetic energy. The second category of sonar is passive. Passive sonar listens for acoustic radiation and provides a bearing reading to the source. This approach results in no indication of range on the signal source.

Passive bearing only target tracking presents the challenge of nonlinearities in the measurement space, requiring conversion from a bearing measurement to cartesian coordinates as the target track is typically maintained in cartesian form. Additionally in the case of the one sensor bearing only problem, typically experienced by sonar equipped ships, there are unobservability issues in the target state [118]. In this particular case, these unobservabilities are overcome with a technique called Target Motion Analysis (TMA) [118,119], where known motion of the bearing only sensor is compared with the observations about the target and using the relative motions the target velocity and range becomes observable. For the case of passive bearing only target tracking using multiple sensors, unobservabilities in the target state are overcome through triangulation.

4.2 Target tracking using bearing only sensors

The scenario that motivates the study of this particular nonlinear state estimation application is target tracking of a threat that is attempting to intercept a ship. Typically the ship, upon detection of the threat, will turn astern to the threat and withdraw at full speed. This typically leaves the ship blind as its engines at full power create substantial noise in the direction of the threat.

Thus the questions being explored in this section of the thesis are “What accuracy can the target be tracked to with a deployment of inexpensive bearing only sonar buoys?” and “What would an optimal deployment look like to minimise target tracking error for monitoring and interception of the threat?”

To this end and to explore this scenario of target tracking using bearing only sensors, a simulation was set up such that a ship detects a threat and turns such that the target is astern. The ship then deploys a buoy field and continues to sail away from the target. The simulation ends when the target reaches the ship, having passed through the buoy field.

To facilitate the study of this problem we have made some simplifying assumptions to reduce the complexity of the simulation.

1. The problem space is constrained to 2 dimensions
2. The target will pursue the ship using non-complex maneuvers (adjusting heading so it is always orientated towards the ship)
3. The target starts north of the ship and will pursue the ship south
4. The sensors exhibit no bias
5. The sensors are homogeneous and modelled with gaussian noise $\mathcal{N}(0, \sigma)$
6. The buoys are capable of instantaneous communication
7. The buoy field is static when deployed.
8. The target is moving with a constant unknown velocity
9. The ship possesses a detection system that will alert it of the incoming threat and provide a very uncertain initial estimate of the target position.

Future work would be to relax these simplifications and iteratively introduce complexity into the simulation to explore their impact on the results generated here.

With a simulation set up as described, we wish to track a target using a bearing only sensor buoy field. We use the information from our sensor observations and what we know about the likely trajectory of the target to construct a state estimator. Like all estimation problems, having a good model of the process is indispensable. Various mathematical models of target motion have been proposed in the literature, of which the most significant models are given by Li et al. [120] in their survey paper “Survey of maneuvering target tracking. Part I: dynamic models”. The model adopted here to construct the target tracking estimator is a discrete time standard nonmaneuver model given by Li et al.

$$\begin{aligned} x_{k+1} &= f(x_k, w_k) \\ f(x_k) &= Ax_k + w_k \\ w_k &= \mathcal{N}(0, Q) \end{aligned} \tag{4.2.1}$$

Where

$$x_k = \begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix} \tag{4.2.2}$$

$$A = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.2.3}$$

With a model for the target trajectory established, we construct our measurement model. With a buoy field of N buoys, $B_{1...N}$, we obtain N measurements $\theta_{1...N}$.

$$\begin{aligned} y_k &= h(x_k, B_{1...N}) \\ h(x_k, B_i) &= \theta_i = arctan((x_k[1] - B_i, x)/(x_k[3] - B_i, y)) + v_k \\ v_k &= \mathcal{N}(0, R) \end{aligned} \tag{4.2.4}$$

Here we define the angle θ_i, as shown in Figure 4.1, is measured from the positive y-axis on the cartesian plane, with clockwise being positive and counter clockwise being negative. θ_i takes on values $\pm 180°$. We also define the positive y-axis as north, and subsequently, the positive x-axis as East, the negative y-axis as South and the negative x-axis as west.

With the model defined we construct an EKF, first initialising as follows:

$$\begin{aligned} \hat{x}_0^+ &= \mathbb{E}[x_0] \\ P_0^+ &= \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \tag{4.2.5}$$

For $k = 0, 1, 2, ...$ we calculate the following.
(a) Compute the time update for both the state estimate and the estimation error covariance:

$$\begin{aligned} \hat{x}_{k+1}^- &= A\hat{x}_k^+ \\ P_{k+1}^- &= AP_k^+ A^T + Q_k \end{aligned} \tag{4.2.6}$$

(b) Compute the following partial derivative matrices:

$$\begin{aligned} H_{k+1} &= \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_{k+1}^-} \\ M_{k+1} &= \left.\frac{\partial h}{\partial v}\right|_{\hat{x}_{k+1}^-} \end{aligned} \tag{4.2.7}$$

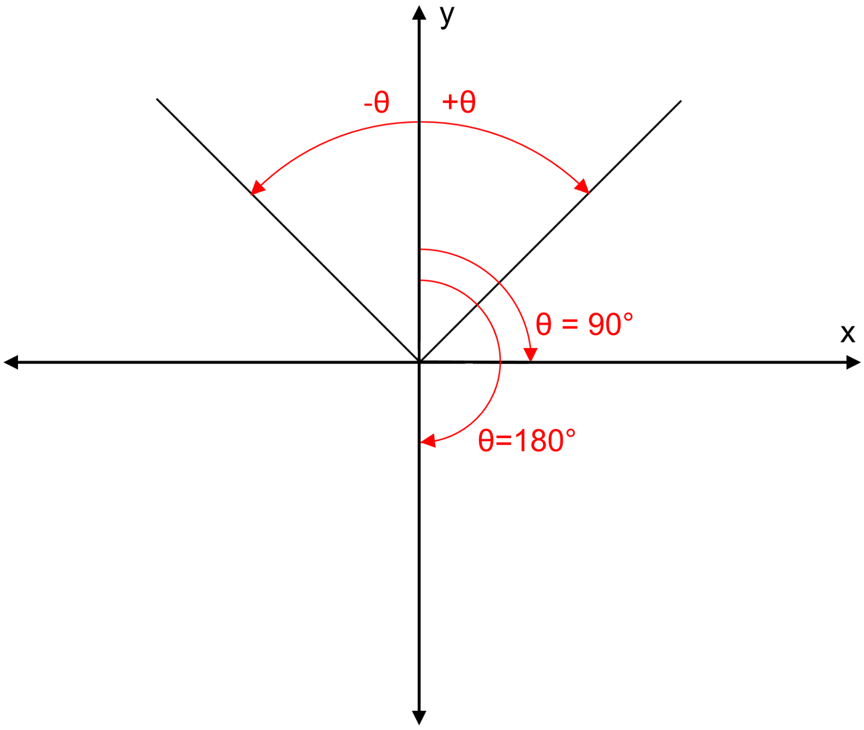

Figure 4.1: θ_i definition for bearing only passive sonar buoy

(c) Compute the measurement update of the state estimate and the estimation error covariance:

$$\begin{aligned} L &= P_k^- H_{k+1}^T (H_{k+1} P_{k+1}^- H_{k+1}^T + M_{k+1} R_{k+1} M_{k+1}^T)^{-1} \\ \hat{x}_{k+1}^+ &= \hat{x}_{k+1}^- + L[y_{k+1} - h(\hat{x}_{k+1}^-)] \\ P_{k+1}^+ &= (I - L H_{k+1}) P_{k+1}^- \end{aligned} \tag{4.2.8}$$

Given the nonlinearities in the measurement space, a UKF was also implemented to compare target tracking performance against the EKF.

With the model defined, we construct a UKF, first initialising as follows:

$$\begin{aligned} \hat{x}_0^+ &= \mathbb{E}[x_0] \\ P_0^+ &= \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \tag{4.2.9}$$

For $k = 0, 1, 2, ...$ iterate as follows:

(a) Calculate Sigma points based on previous distribution knowledge

$$\begin{aligned}
\hat{x}_k^i &= \hat{x}_k^+ + \tilde{x}^i \qquad i = 1...2n \\
\tilde{x}^i &= \left(\sqrt{nP_k^+}\right)_i^T \qquad i = 1...2n \\
\tilde{x}^{n+i} &= -\left(\sqrt{nP_k^+}\right)_i^T \qquad i = 1...2n
\end{aligned} \tag{4.2.10}$$

For an explanation on the notation $\left(\sqrt{nP_k^+}\right)_i$ refer to Appendix A.2.

(b) Propagate Sigma points through nonmaneuver model

$$\hat{x}_{k+1}^i = A\hat{x}_k^i \qquad i = 1...2n \tag{4.2.11}$$

(c) Calculate state mean and covariance from transformed Sigma points

$$\begin{aligned}
\hat{x}_{k+1}^- &= \frac{1}{2n}\sum_{i=1}^{2n}\hat{x}_{k+1}^i \\
P_{k+1}^- &= \frac{1}{2n}\sum_{i=1}^{2n}(\hat{x}_{k+1}^i - \hat{x}_{k+1}^-)(\hat{x}_{k+1}^i - \hat{x}_{k+1}^-)^T + Q_{k+1}
\end{aligned} \tag{4.2.12}$$

(d) Recalculate Sigma points based on updated Gaussian distribution from steps (b + c)

$$\begin{aligned}
\hat{x}_{k+1}^i &= \hat{x}_{k+1}^- + \tilde{x}^i \qquad i = 1...2n \\
\tilde{x}^i &= \left(\sqrt{nP_{k+1}^-}\right)_i^T \qquad i = 1...2n \\
\tilde{x}^{n+i} &= -\left(\sqrt{nP_{k+1}^-}\right)_i^T \qquad i = 1...2n
\end{aligned} \tag{4.2.13}$$

(e) Propagate Sigma points through nonlinear measurement model

$$\begin{aligned}
\hat{y}_{k+1}^i &= h(\hat{x}_{k+1}^i, B_{1...N}) \\
&= arctan((\hat{x}_{k+1}^i[1] - B_{1...N}, x)/(\hat{x}_{k+1}^i[3] - B_{1...N}, y)) + v_k \qquad i = 1...2n
\end{aligned} \tag{4.2.14}$$

(f) Calculate state mean and covariance from Sigma points and cross covariance between $\hat{x}_{k+1}^-$ and $\hat{y}_{k+1}^i$

$$\begin{aligned}
\hat{y}_{k+1} &= \frac{1}{2n}\sum_{i=1}^{2n}\hat{y}^i_{k+1} \\
P_y &= \frac{1}{2n}\sum_{i=1}^{2n}(\hat{y}^i_{k+1}-\hat{y}_{k+1})(\hat{y}^i_{k+1}-\hat{y}_{k+1})^T + R_{k+1} \\
P_{xy} &= \frac{1}{2n}\sum_{i=1}^{2n}(\hat{x}^i_{k+1}-\hat{x}^-_{k+1})(\hat{y}^i_{k+1}-\hat{y}_{k+1})^T
\end{aligned} \tag{4.2.15}$$

(g) Calculate Kalman gain and update the state estimate

$$\begin{aligned}
K &= P_{xy}P_y^{-1} \\
\hat{x}^+_{k+1} &= \hat{x}^-_{k+1} + K(y_{k+1}-\hat{y}_{k+1}) \\
P^+_{k+1} &= P^-_{k+1} - KP_yK^T
\end{aligned} \tag{4.2.16}$$

4.3 Comparison between the EKF and UKF

With both an EKF and UKF target tracking algorithm implemented in the simulation, we executed both against three basic deployments of a buoy field; a horizontal line of buoys, a semi circle of buoys and a wedge of buoys. These deployments are shown in Figure 4.2. We chose a lateral distribution distance of 1000m relative to the target trajectory for these three basic deployments. Such lateral distribution distances could be achieved practically through the deployment of sonar buoys via a projectile launch mechanism. The practical considerations for such a deployment mechanism are not explored in this Thesis.

It should be noted that the implementation of both the EKF and UKF utilised an observation rate of 50 hz to execute. Both filters were initialised with a large uncertainty about the target position assuming a distance of 4 km and a speed of 25 m/s from the ship with the threat astern at buoy field deployment. The variance of this initial estimate of the target's position (x, y) was set to $1e^6$ and $1e^2$ for the target's velocity $(\dot{x}, \dot{y})$.

The target's actual speed is closer to 26 m/s. It is assumed that the speed of the target can be closely predicted as it will be closing at max velocity. The initial position of the target is unknown and has significant error upon initialisation.

To compare the two filters, a Monte Carlo simulation was conducted where each filter was executed 100 times against the buoy field deployment and the mean and variance of the filter error was calculated. The Random Number Generator used for the Monte Carlo simulation was seeded with the same 100 seeds between filters to ensure a meaningful performance comparison could be made between filters on individual simulation runs.

A comparison of the EKF and UKF self perception of performance from one of the simulated runs, given by the determinant of the covariance of the state estimator, is shown in Figure 4.3. To aid evaluation, we define a set of three time intervals that are of interest for comparative purposes. These definitions are formally introduced in Section 5.1, but for now are shown on Figure 4.3. These time intervals are acquisition, track and intercept.

According to the estimator's own self perception of performance, there is little to no significant differences between the three basic buoy field deployments, with Figure 4.3 showing the covariance performance from one of the simulated runs of the horizontal line buoy field deployment.

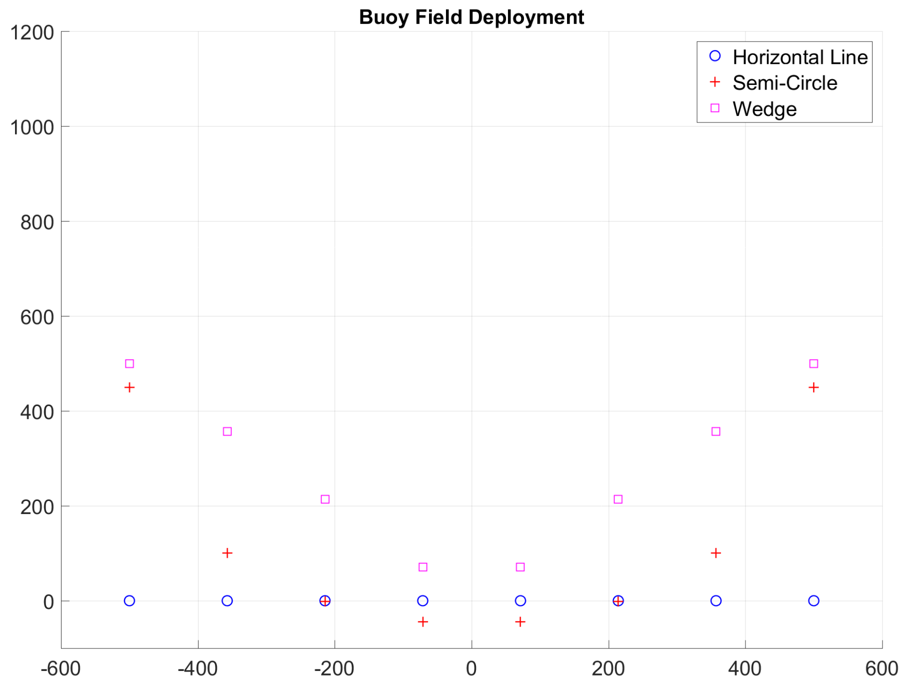

Figure 4.2: Buoy Field Deployments (N=8)

However, we wish to know the actual performance of the state estimators. First we compare the state estimators self perception of performance to the actual performance calculated from the Monte Carlo simulation results. Figures 4.4 and 4.5 show the actual mean positional error and 95% confidence interval for the two estimators. Overlayed on these figures is the 95% confidence interval as reported by the state estimators from one of the simulation runs. Figures 4.6 and 4.7 show the actual mean error and 95% confidence interval of the two filters over the track and intercept key time intervals. Similarly, these figures also have the 95% confidence interval as reported by the state estimators overlayed.

Here we compare the estimator's self perception of performance with the actual state estimator performance and can observe that both state estimators self perceived accuracy are good approximations of true performance for this buoy field configuration. We also observe that both the EKF and UKF perform identically. Figures 4.6 and 4.7 which focus on the track and intercept time intervals, best illustrate the near identical performance between the EKF and UKF over the time intervals that are of most interest.

Common to both state estimators, we observe a rapid correction in the y positional state and an initial bias during the acquisition time interval. This error converges towards zero over approximately 20 to 30 seconds after tracking has commenced. Thus during the acquisition interval, the state estimators self perception of accuracy is incorrect due to this error bias. As the target approaches the buoy field, we observe that the error converges towards zero and the variance shrinks as the target gets closer.

For each state estimator we can also review and compare their 95% confidence ellipses, calculated from their covariance matrices, over the key time intervals we're interested in. This allows us to comment on the eccentricity of these ellipses as a function of time. Figures 4.8, 4.9 and 4.10 give the mean 95% confidence ellipses over the acquisition, track and intercept time intervals for the two state estimators from one of the simulation runs. Here we observe elongation in the y-axis

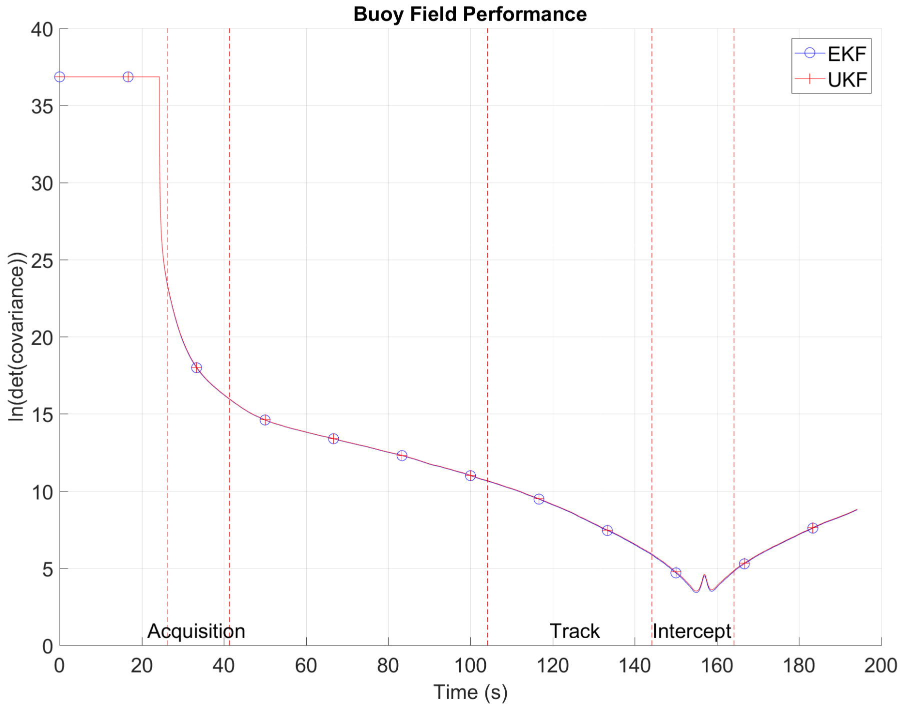

Figure 4.3: EKF and UKF Filter Covariance Performance - Lateral Distribution Distance 1000m

the further the target is from the buoy field. This is a result from the poor angular diversity from the buoy field which leads to higher positional uncertainty in the y-axis. As the target approaches the buoy field, this poor angular diversity decreases resulting in a more circular 95% confidence ellipse.

To further compare the EKF and UKF the buoy field configuration was adjusted. Through these adjustments it was found that the lateral distribution distance of the buoy field relative to the target trajectory has a some effect on the differences between the EKF and UKF estimator performance.

If we tighten the lateral distribution distance of the buoy field, for instance having N=2 with a lateral distance of 200m, we can see that the UKF provides a slight performance improvement over the EKF with respect to positional error during the acquisition time interval. This is shown in Figure 4.11. As the target closes in on the buoy field, which increases the relative lateral distance between the buoy field and target, we see the performance of the EKF and UKF converge. We also observe from Figure 4.12 that the state estimators self perceived accuracy differs from the true performance calculated from the Monte Carlo simulation. The actual variance in the state estimate is worse than the self perceived accuracy. This is attributed to the amplification of the nonlinearities of the measurement sensors due to the poor angular diversity of the buoy field deployment. We can observe as the target approaches the buoy field the actual and self perceived performance converge as the angular diversity from the sensors increase due to the target being closer to the buoy field over the intercept time interval.

Reviewing the mean 95% confidence ellipses of the state estimators from the tightened lateral buoy field distribution for one of the simulation runs, shown in Figures 4.13, 4.14 and 4.15, we observe an increase in eccentricity with significant elongation in the y-axis. This is consistent with our prior results where poor angular diversity, which is increased in the tighter buoy field configuration, led to higher positional uncertainty in the y-axis. Interestingly, over the intercept

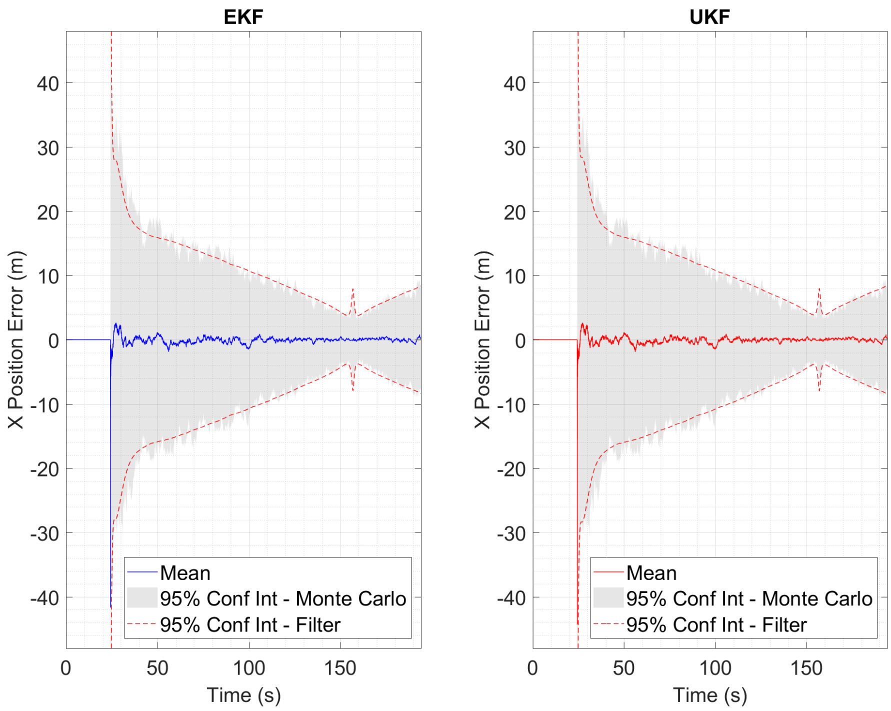

Figure 4.4: EKF and UKF Filter X Position Error - Lateral Distribution Distance 1000m

time interval when we observed the EKF and UKF performance to improve we note the error ellipses are near circular with almost no eccentricity.

To contrast, if we have a wide lateral distribution distance in the buoy field, for instance having N=2 with a lateral distribution distance of 5000m, we can see that the UKF and EKF perform identically as shown in Figures 4.18, 4.19, 4.16 and 4.17. We observe the same initial error bias during the acquisition time interval, which then converges towards zero as the target approaches the buoy field. The state estimators self perceived performance is inline with the actual performance calculated from the Monte Carlo simulation.

Reviewing the mean 95% confidence ellipses of the state estimators from the wider lateral buoy field distribution for one of the simulation runs, given in Figures 4.20, 4.21 and 4.22, we observe almost no eccentricity in the acquisition time interval. This is due to such a wide buoy field deployment providing good angular diversity in the measurements over the acquisition time interval. Then as the target approaches the buoy field we observe elongation in the x-axis in the 95% confidence ellipses as this angular diversity decreases relative to the x-axis leading to positional uncertainty along it.

To summarise, the marginal performance improvement of the UKF over the EKF is intuitive based upon how the UKF maintains and propagates the Gaussian distribution of the state estimate. With tight lateral spacing, the noise in the bearing measurements from the sensor causes the EKF's linearisation of the measurement model to be a poor representation of the true distribution of the measurements. The UKF in contrast, maintains sigma points that represent the state estimate distribution and propagates these sigma points through the nonlinear measurement model in this case.

In fact, the issue of converting coordinates from polar to cartesian and the associated nonlinearities which degrade the performance of the standard EKF is the introductory example used

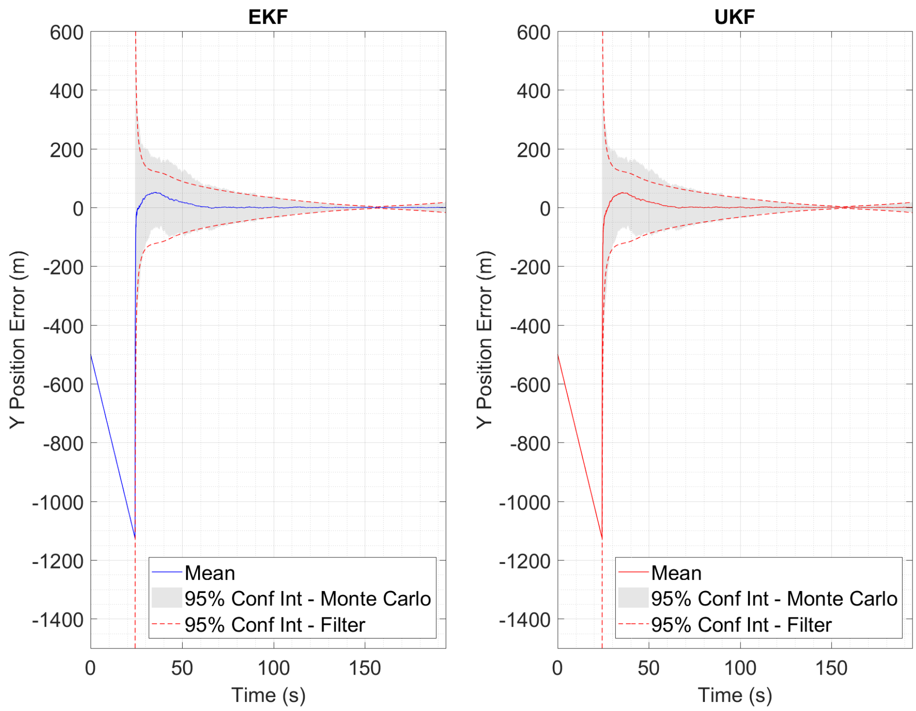

Figure 4.5: EKF and UKF Filter Y Position Error - Lateral Distribution Distance 1000m

to demonstrate the superior performance of the UKF given in the seminal paper by Julier and Uhlmann [121] which first introduced the UKF. However, in the buoy field configurations explored to compare the EKF and UKF the performance was found to be near identical. As such, we have used the EKF in Chapter 5 due to its superior computational simplicity to explore the optimisation of a buoy field to minimise target tracking error over the track and intercept time intervals.

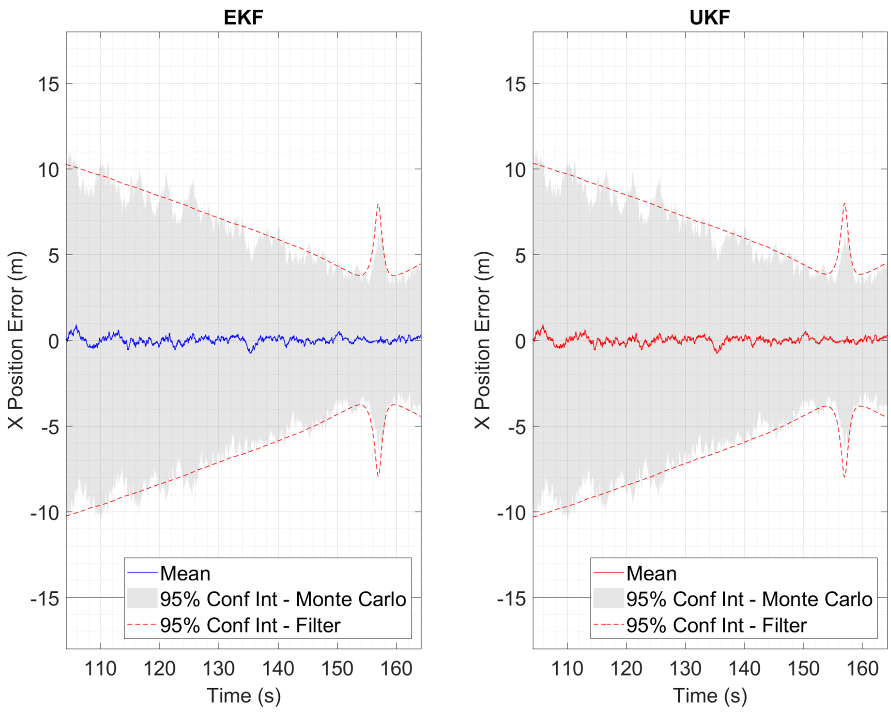

Figure 4.6: EKF and UKF Filter X Position Error - Track and Intercept Time Intervals - Lateral Distribution Distance 1000m

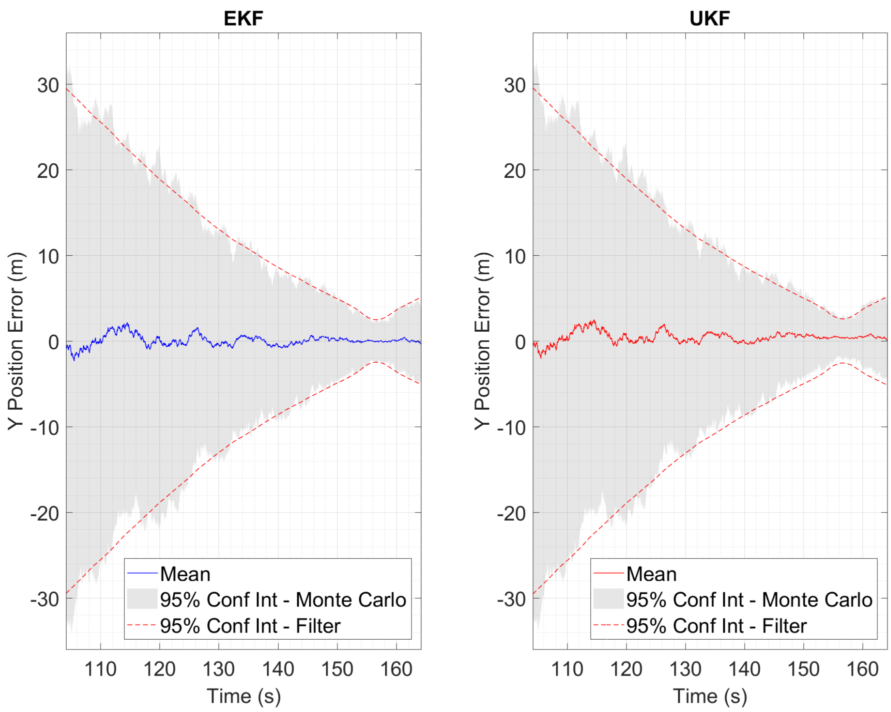

Figure 4.7: EKF and UKF Filter Y Position Error - Track and Intercept Time Intervals - Lateral Distribution Distance 1000m

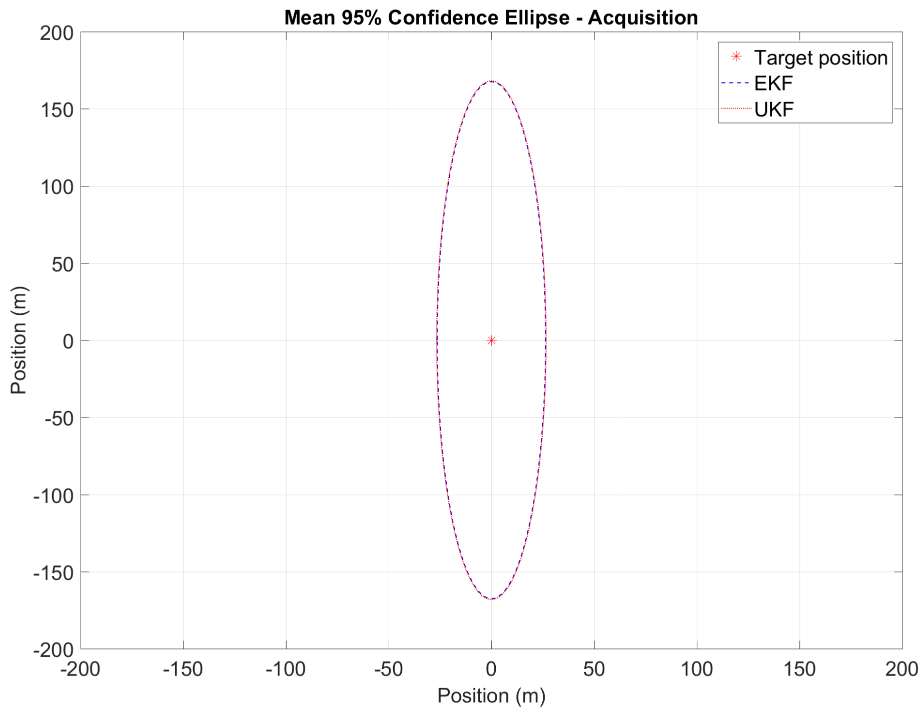

Figure 4.8: EKF and UKF Mean 95% Confidence Ellipse - Acquisition Time Interval - Lateral Distribution Distance 1000m

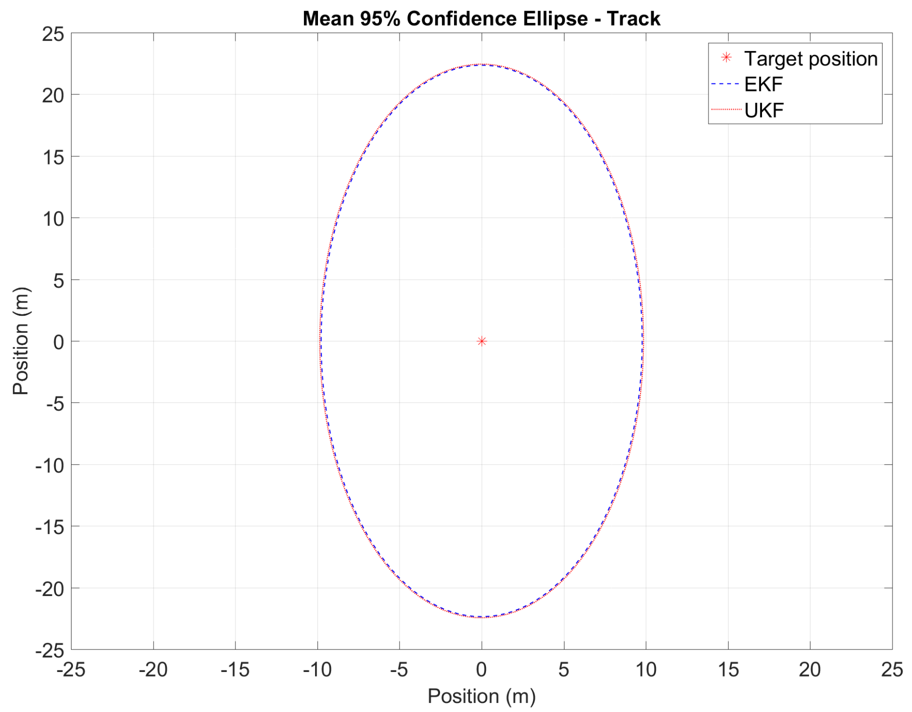

Figure 4.9: EKF and UKF Mean 95% Confidence Ellipse - Track Time Interval - Lateral Distribution Distance 1000m

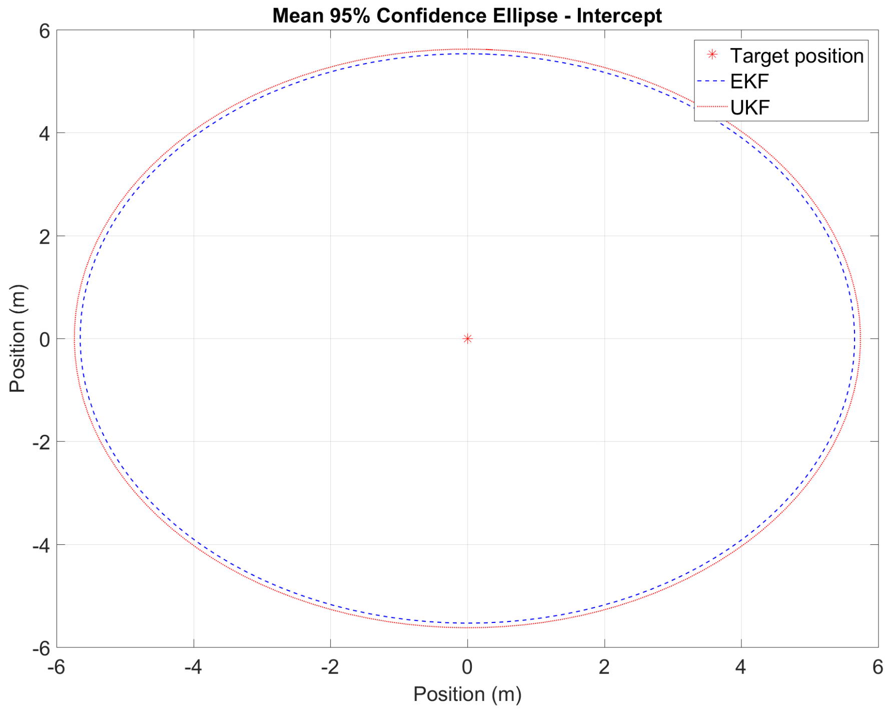

Figure 4.10: EKF and UKF Mean 95% Confidence Ellipse - Intercept Time Interval - Lateral Distribution Distance 1000m

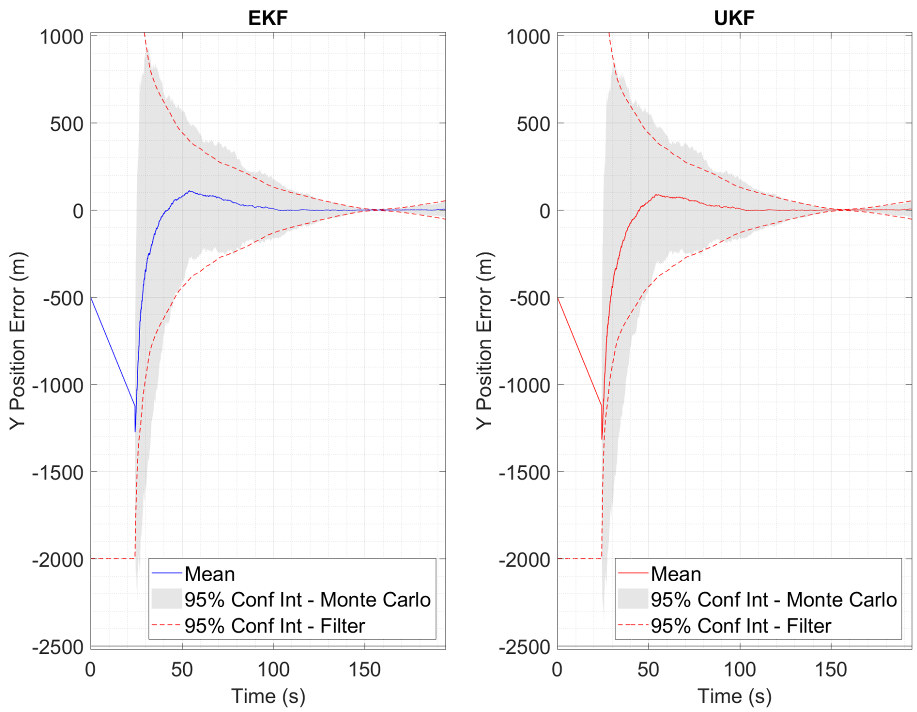

Figure 4.11: EKF and UKF Filter Y Position Error - Lateral Distribution Distance 200m

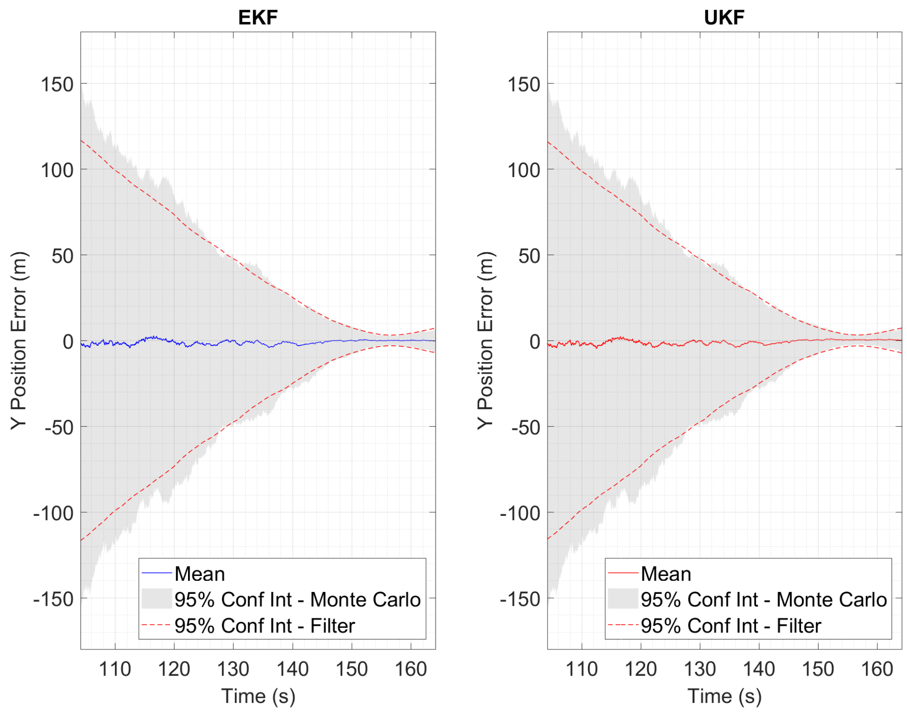

Figure 4.12: EKF and UKF Filter Y Position Error - Track and Intercept Time Intervals - Lateral Distribution Distance 200m

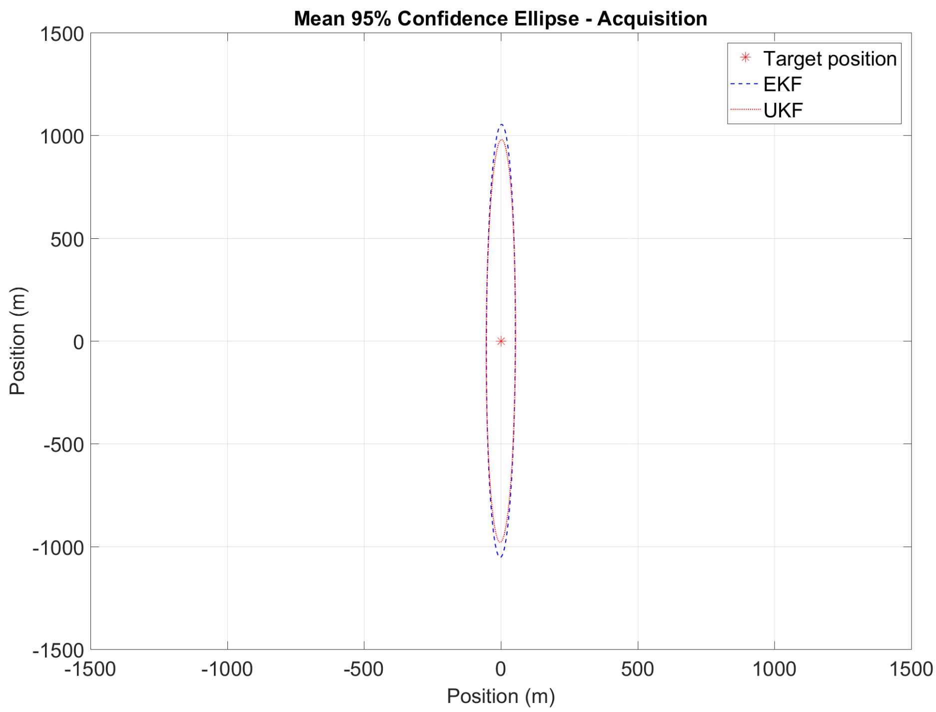

Figure 4.13: EKF and UKF Mean 95% Confidence Ellipse - Acquisition Time Interval - Lateral Distribution Distance 200m

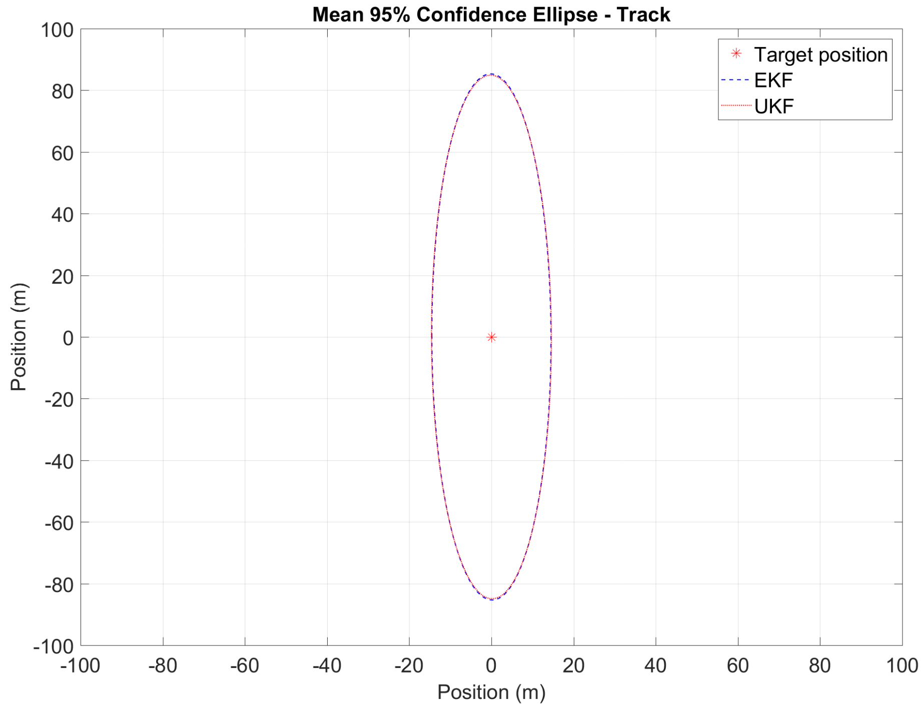

Figure 4.14: EKF and UKF Mean 95% Confidence Ellipse - Track Time Interval - Lateral Distribution Distance 200m

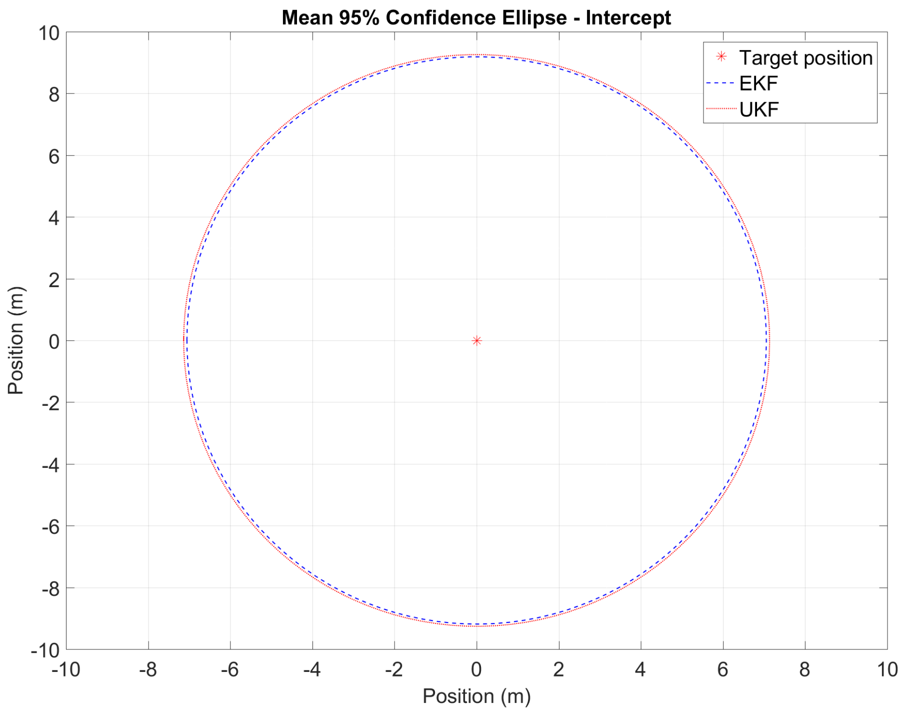

Figure 4.15: EKF and UKF Mean 95% Confidence Ellipse - Intercept Time Interval - Lateral Distribution Distance 200m

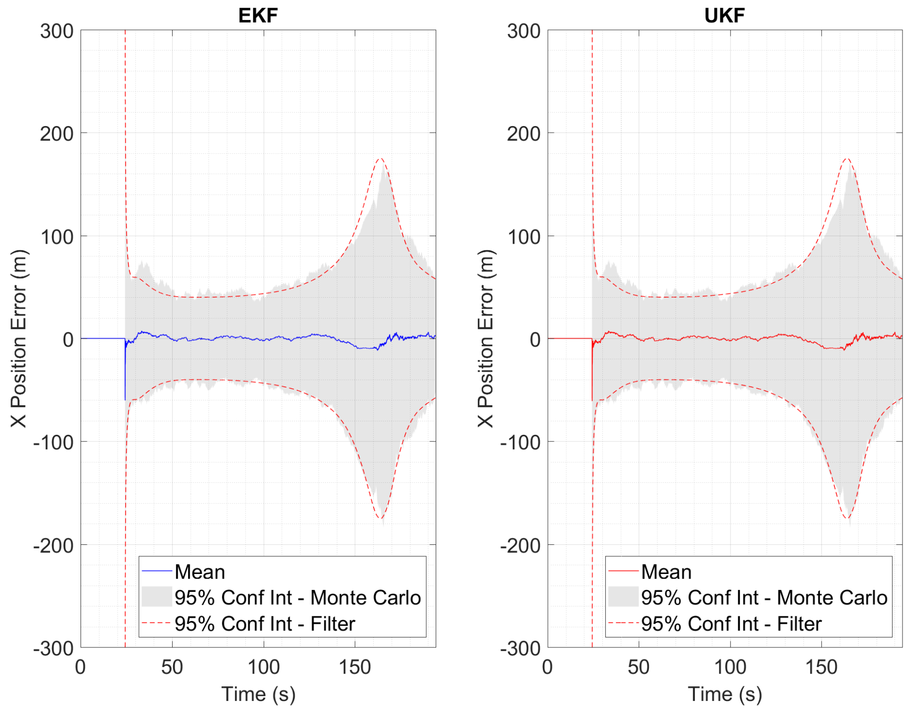

Figure 4.16: EKF and UKF Filter X Position Error - Lateral Distribution Distance 5000m

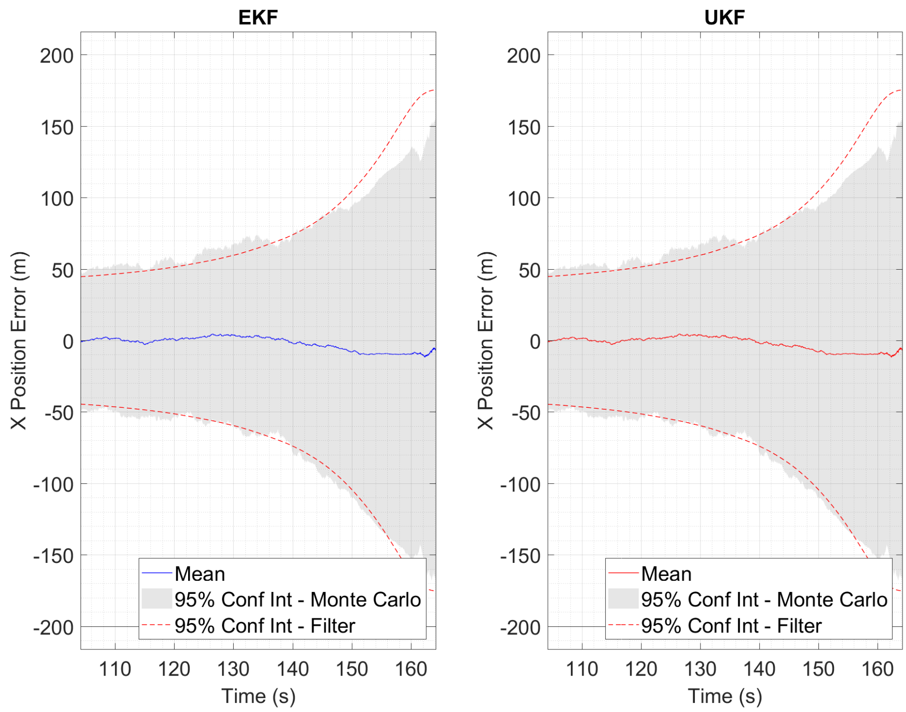

Figure 4.17: EKF and UKF Filter X Position Error - Track and Intercept Time Intervals - Lateral Distribution Distance 5000m

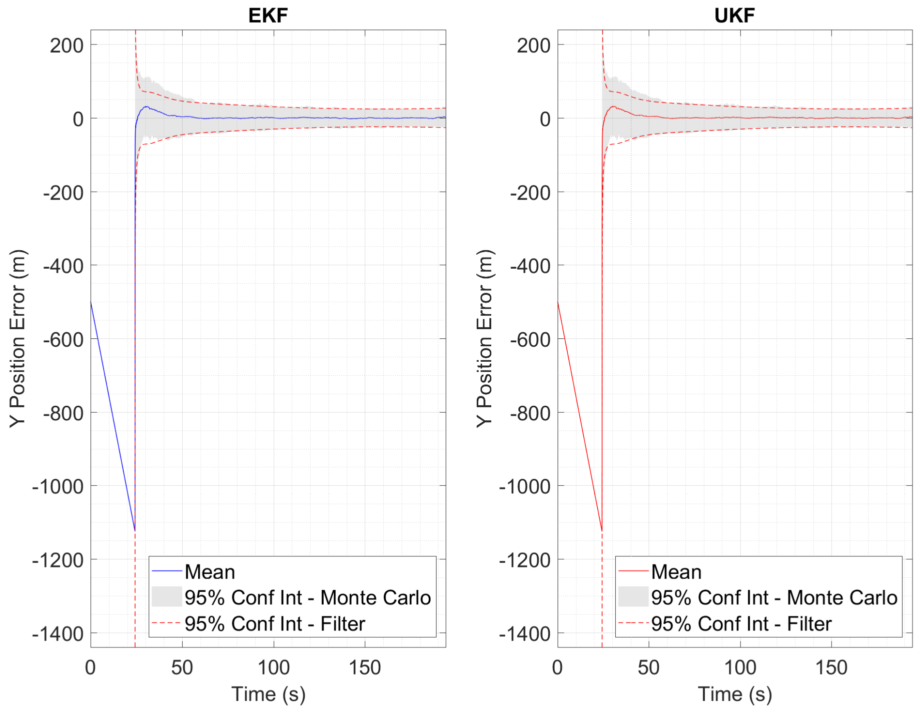

Figure 4.18: EKF and UKF Filter Y Position Error - Lateral Distribution Distance 5000m

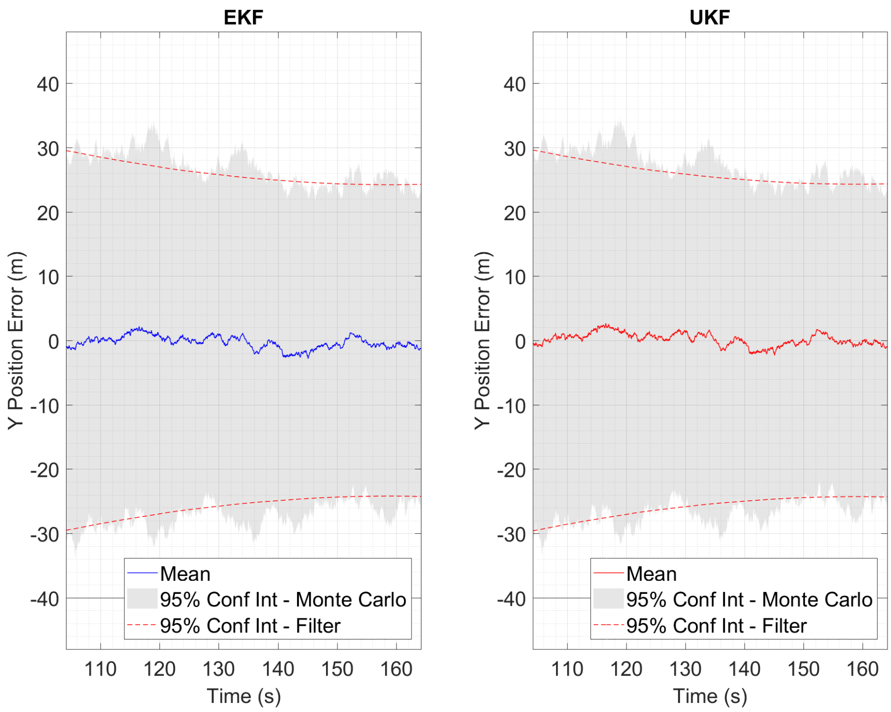

Figure 4.19: EKF and UKF Filter Y Position Error - Track and Intercept Time Intervals - Lateral Distribution Distance 5000m

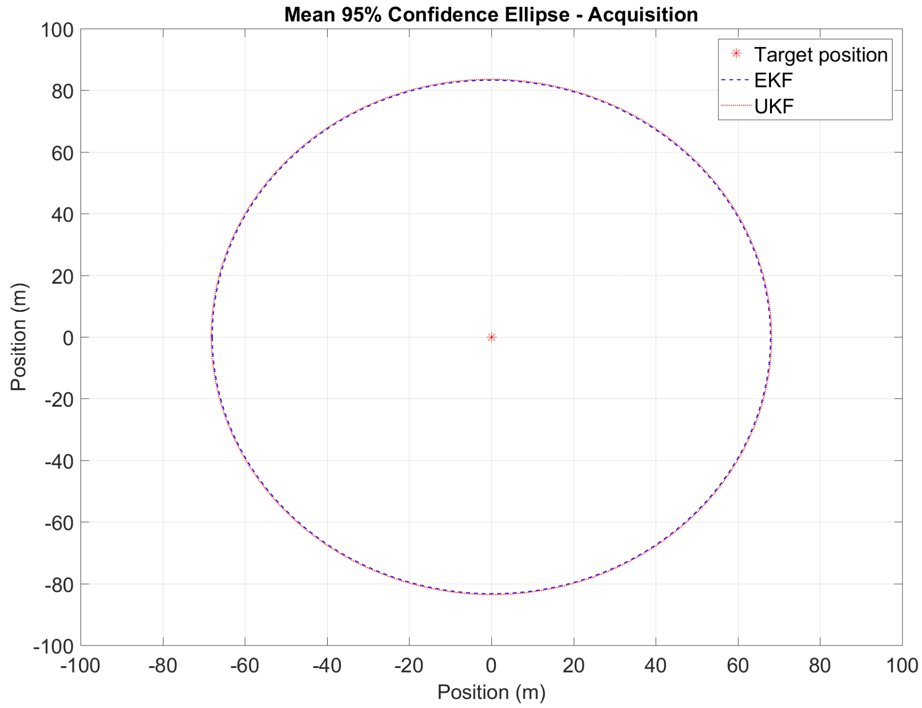

Figure 4.20: EKF and UKF Mean 95% Confidence Ellipse - Acquisition Time Interval - Lateral Distribution Distance 5000m

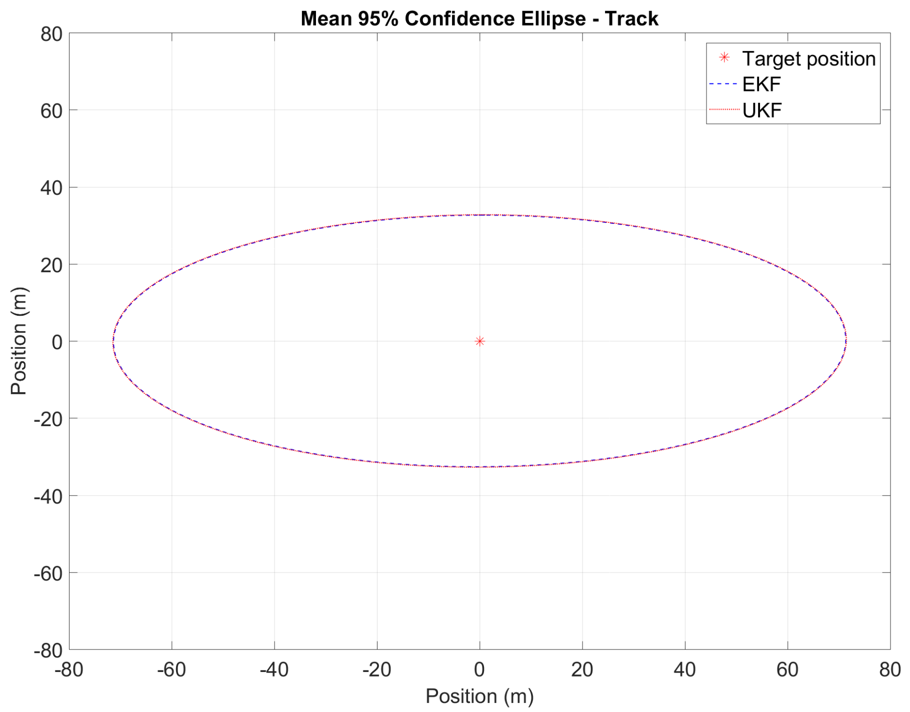

Figure 4.21: EKF and UKF Mean 95% Confidence Ellipse - Track Time Interval - Lateral Distribution Distance 5000m

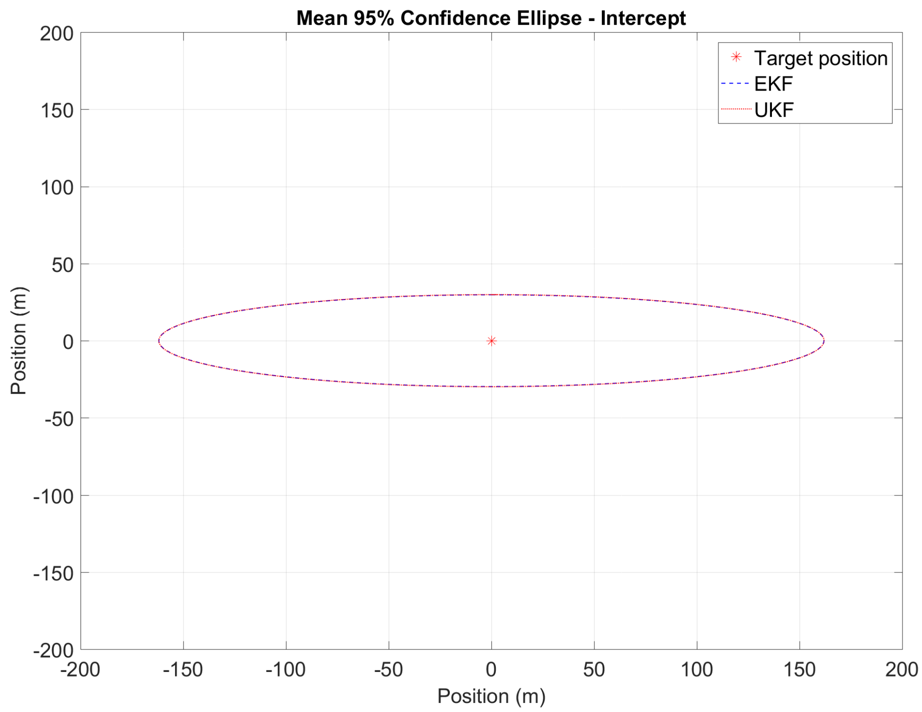

Figure 4.22: EKF and UKF Mean 95% Confidence Ellipse - Intercept Time Interval - Lateral Distribution Distance 5000m

Chapter 5

Bearing Only Target Tracking using a Buoy Field

This chapter of the thesis concerns itself with the challenge of defining an optimal deployment for a bearing only buoy field sensor network for target tracking. Informally, the problem statement is, how accurately can we track a target given realistic noise parameterisation of the sensors, and what is an optimal deployment of a buoy field for target tracking and interception. The problem statement can also be formalised as follows:

Given a target with state

$$T_r = \begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix} \tag{5.0.1}$$

and a series of N stationary buoys B_i

$$B_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{5.0.2}$$

What is the optimal static deployment of the buoy field $B_{1...N}$ such that we maximise the detection confidence $\Phi(T_r(t), B_{1...N})$ of the target T_r over a period of time $t_{s...e}$.

This chapter is broken down into the following sections: Section 5.1 reviews several approaches in the literature at developing an optimal deployment of a sensor network. This section also establishes the definition of optimal used in this thesis and defines the cost functions and the key time intervals which will be used to compare buoy field performance for target tracking. Section 5.2 explores and compares target tracking performance against some standard geometries for the buoy field sensor network. Section 5.3 explores and compares the impact the number of sensors available to the network has on the performance of a target tracking algorithm. Section 5.4 utilises a genetic algorithm to optimise the buoy field sensor network against the cost functions defined earlier in this chapter, it then compares and analyses the performance of the various resulting deployments. Section 5.6 provides conclusions about the buoy field deployments explored in this Chapter, general recommendations based upon the results observed and recommendations for future work.

5.1 Optimality of a sensor network

The performance of a state estimator tracking a target is determined by the fidelity of the motion model, the quality of the measurements from the sensor network and the sensor network geometry [122]. Having implemented a nonlinear state estimator for the target state using a bearing only buoy field and given the importance sensor network geometry has on the performance of a localisation algorithm, we now consider the problem of optimal buoy field deployment.

Sensor geometries continue to be investigated extensively in the literature [122–125]. The typical approach taken in the literature for analysing and optmising the performance of sensor networks is to utilise the Cramer-Rao Lower Bound (CRLB). The CRLB, an important result in statistics, expresses a lower bound on the variance for an unbiased estimator [126].

In Yang et. al [122], they considered the problem of an optimal strategy for the placement of heterogeneous sensors assuming Gaussian priors. They used the determinant of the Fisher Information Matrix, of which the CRLB is the reciprocal of the Fisher Information Matrix. Thus, maximising the Fisher Information is equivalent to minimising the CRLB. They considered a network consisting of a mixture of bearing only sensors, ranging sensors, time of arrival and time difference of arrival sensors. They developed strategies for maximising the Fisher Information using the approach that the largest reduction in uncertainty of the target position is achieved by rendering the variance matrix to have equal eigenvalues, and thus a circular variance for a sensor network in 2-D space.

Similar to Yang [122], Meng et. al [124] considered a heterogeneous sensor network, looking at optimal placement of mixed sensors including ranging sensors, bearing only sensors and time difference of arrival. They noted difficulty in obtaining an explicit solution for an arbitrary large amount of sensors, but generated some explicit results for heterogenous sensor networks when the number of sensors is small.

Meng also made the observation that for an angular sensor, the quality of the measurement degrades as the distance to the target increases because of the reduction in measurement sensitivity to the cross-range quantity [124].

Tekdas et. al [125] considers the problem of localisation using bearing only sensors and triangulation. They recognise the pursuit of an optimal sensor placement as an NP-hard problem and present an optimisation approach based upon integer linear programming with constraints for minimum sensing qualities for each sensor added. The uncertainty function used by Tekdas for the sensor network was a heuristic that rewards sensor positions closer to the target and relative triangulation angles between sensors of 90°.

Bishop et. al [123] considered a static localisation problem, involving multiple sensors located in 2-D space. They also used the minimisation of the CRLB as the optimality evaluation of sensor placement. They considered optimal deployments for range-only, time of arrival and bearing only sensors. They identified a number of necessary and sufficient conditions for the sensor-target angular positions around the stationary target and also showed that an optimal sensor deployment is generally not unique.

A key difference with the approach taken here in this thesis, is to instead use the simulation constructed for the evaluation of the target state estimator to evaluate various buoy field deployments. A D-Optimality approach is utilised where the determinant of the covariance of the state estimator is minimised against a desired detection performance curve. This provides evaluation of a non-static scenario and contrasts with the static analysis of a buoy field using the CRLB typically used in the literature.

To evaluate the performance of a buoy field, we first define two time periods that are of interest. A tracking time period, and an interception time period. The interception time period, $t_{int\_s...int\_e}$, is the time period which is defined as the ideal time over which to intercept the target. The

Table 5.1: Target Tracking Key Time Intervals

Time Period	Duration (s)
$t_{trk\_s...trk\_e}$	40
$t_{int\_s...int\_e}$	20
$t_{s...e}$	60

track time period, $t_{trk\_s...trk\_e}$, leads the interception time period, and is the time over which greater certainty of the target position is required but not to a level required for interception. We've defined $t_{int\_s...int\_e} < t_{trk\_s...trk\_e}$ as we've assumed that interception occurs over a shorter time frame, relative to general tracking of the target. Adding the two time intervals together we get the time interval in which we attempt to minimise the covariance of the buoy field over, $t_{s...e} = t_{int\_s...int\_e} + t_{trk\_s...trk\_e}$. The durations of these key time periods used in our simulation are defined in Table 5.1.

These time intervals can be seen for reference on the covariance performance figures from the simulation as vertical dotted red lines. This puts the buoy fields performance into perspective of these key time intervals. It should be noted that these time intervals are chosen arbitrarily, along with all the performance values used, in the context of this simulation and is not indicative of any performance characteristics for target interception.

With the key time intervals defined, we constructed two determinant vs. time curves to define the desired detection performance of the buoy field. A weighted least squares of the difference between actual performance of the buoy field versus the desired performance was used as the objective function.

Two objective functions $\Phi_1(S(t_{s...e}, B_{1...N}))$ and $\Phi_2(S(t_{s...e}, B_{1...N}))$ have been considered, both derivative of a D-Optimality objective (minimising the determinant of the covariance matrix of the estimator). Where $S(t_{s...e}, B_{1...N})$ is the simulation data over the time interval $t_{s...e}$ using the buoy field $B_{1...N}$. The simulation data includes the estimated position of the tracked target over time $x_{s...e}$, and the covariance of the state estimate over time $P_{s...e}$.

The first objective function $\Phi_1(.)$ (refer to Equation 5.1.1), simply minimises the determinant of the covariance matrix of the estimator consistently over a duration $t_{s...e}$.

$$\Phi_1(S(t_{s...e}, B_{1...N})) = \sum_{t=t_s}^{t_e} (\det(P(t)) - D)^2 \tag{5.1.1}$$

where

$$P = \text{covariance matrix from simulation using buoy field } B_{1...N}$$
$$D = \text{desired determinant value per time instant}$$

The second objective function $\Phi_2(.)$ (refer to Equation 5.1.2), minimises the determinant of the covariance matrix of the estimator over the two key time intervals. The track stage, $t_{trk\_s...trk\_e}$, where the target localisation uncertainty is reduced. Then the intercept stage, $t_{int\_s...int\_e}$, where the target localisation uncertainty is further minimised.

$$\Phi_2(S(t_{s...e}, B_{1...N})) = \sum_{t=t_{trk\_s}}^{t_{trk\_e}} (\det(P(t)) - D_{trk})^2 + \sum_{t=t_{int\_s}}^{t_{int\_e}} (\det(P(t)) - D_{int})^2 \tag{5.1.2}$$

where

$$
\begin{aligned}
P &= \text{covariance matrix from simulation using buoy field } B_{1\ldots N} \\
D_{trk} &= \text{desired determinant value per track time instant} \\
D_{int} &= \text{desired determinant value per intercept time instant}
\end{aligned}
$$

The concept behind this approach is to allow a longer track stage to get a balance between detection of the target and duration that this level can be maintained; then over a shorter period, tighten the localisation confidence to facilitate guiding an intercept vehicle to the target. This provides a compromised approach to the problem, and reduces the number of buoys required to satisfy the problem.

The idea of minimising the distance to a desired performance D is to provide a design parameter that can be used to balance the number of buoys used, their placement and target tracking performance across the target trajectory. It is made use of more significantly in Equation 5.1.2 where we can set two tiers of performance, thus accepting lesser accuracy initially and minimising the required number of buoys, then demanding a higher accuracy for the intercept stage of the target trajectory.

To assist with the analysis of the performance of the buoy fields an additional time interval was defined, the target acquisition interval $t_{acq\_s \ldots acq\_e}$. This time interval is defined as the interval $t_{\text{deploy}} + 2s$ after deployment of the buoy field until $t_{\text{deploy}} + 17s$. It provides a summary of the initial performance of a buoy field when acquiring the target from the initial prior used to initialise the state estimator.

5.2 Typical Buoy Field Deployments

Before we attempt to find an optimal buoy field deployment, we've explored four buoy field geometries that one may typically consider if posed with the task of defining a buoy field deployment. These four buoy field deployments are a horizontal line, a semi circle, a wedge and a vertical line and are shown in Figure 5.1. Figure 5.1 also depicts the nominal trajectory of a target through our buoy fields, with the trajectory divided into the Track time interval and the Intercept time interval.

Using these four buoy field deployments the performance of the state estimator, implemented as an EKF, was evaluated and compared using a Monte Carlo simulation of 100 runs against the simulation described in Section 4.2. To evaluate performance, both the Monte Carlo simulation results of the state estimator's mean error and variance was compared along with the estimator's self perceived performance as informed by the covariance matrices. From the results in Section 4.3, we saw that the EKF's self perceived accuracy was a reasonable approximation compared to the actual performance calculated from the Monte Carlo simulation. We continue to compare the state estimators self perceived performance against actual performance calculated from the Monte Carlo simulation for the different buoy field configurations. Given the confidence in the EKF's self perceived accuracy, graphing the estimator's covariance determinant is a useful measure for relative performance of the buoy field deployments. Figure 5.2 shows the state estimator performance with respect to time. Also note the segmentation of the time axis into the three key time intervals, with the segmentation represented by red dotted lines. We observe that the Wedge provides the best overall performance over the acquisition and intercept phases; with the semi-circle a close second. The horizontal line performs similarly in the intercept phase, but lacks over the acquisition and track phase. Finally the vertical line performs the worst overall, but has superior performance over the track phase.

This assessment of the performance is also consistent with the state error over time for the Y

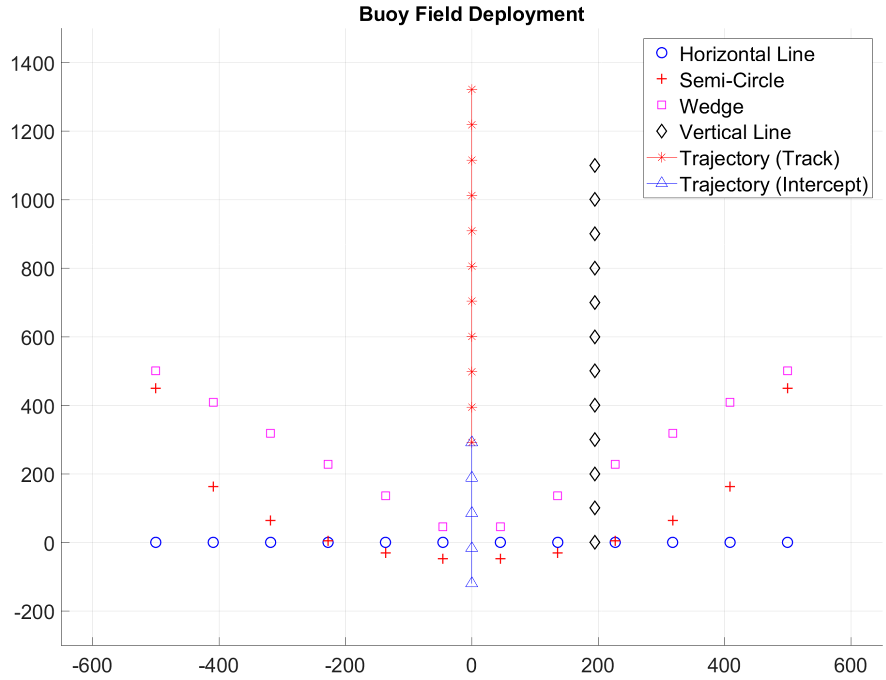

Figure 5.1: Typical Buoy Field Deployments (N=12)

position, the state with the largest error in our simulation, shown in Figures 5.7 and 5.8. In these figures we can also confirm that the EKF's self perceived accuracy is consistent with the Monte Carlo results. It is worth noting that the EKF's self perceived accuracy for the velocity components of the target state are quite conservative, with the actual variance calculated from the Monte Carlo simulation being less.

Figures 5.3 to 5.10 show the EKF error for the target states over the simulation. It can be seen here that the vertical line exhibits the worst performance, until 100 seconds into the simulation where it exhibits the lowest positional error, then from 150 seconds onwards it regresses to the worst accuracy. Similarly, we observe that the performance of the semi-circle and wedge buoy field deployments are similar, and the horizonal line buoy field deployment is marginally inferior over the acquisition and track time intervals.

As before we can review the mean 95% confidence ellipses from one of the simulated runs for each of the key time intervals, acquisition, track and intercept. These are shown in Figures 5.11, 5.12 and 5.13 respectively. Here we observe a general elongation towards the y-axis for the horizontal line, semi-circle and wedge deployments for the acquisition and track time intervals. The vertical line deployment exhibits high levels of skewness in the x-y plane resultant from a lack of diversity in the relative angle between sensors and the target. This equates to poor triangulation of the target at long distances.

It was also found that the effectiveness of the vertical line buoy field deployment is heavily dependant on its distance from the target trajectory. Exploring this dependency, vertical line buoy field deployments were simulated with varying distances from the target trajectory. Decreasing the distance of the vertical line deployment to the target trajectory, the worse the performance the buoy field exhibited when the target was at extreme distance. This was due to a severe lack of angular diversity in the sensors.

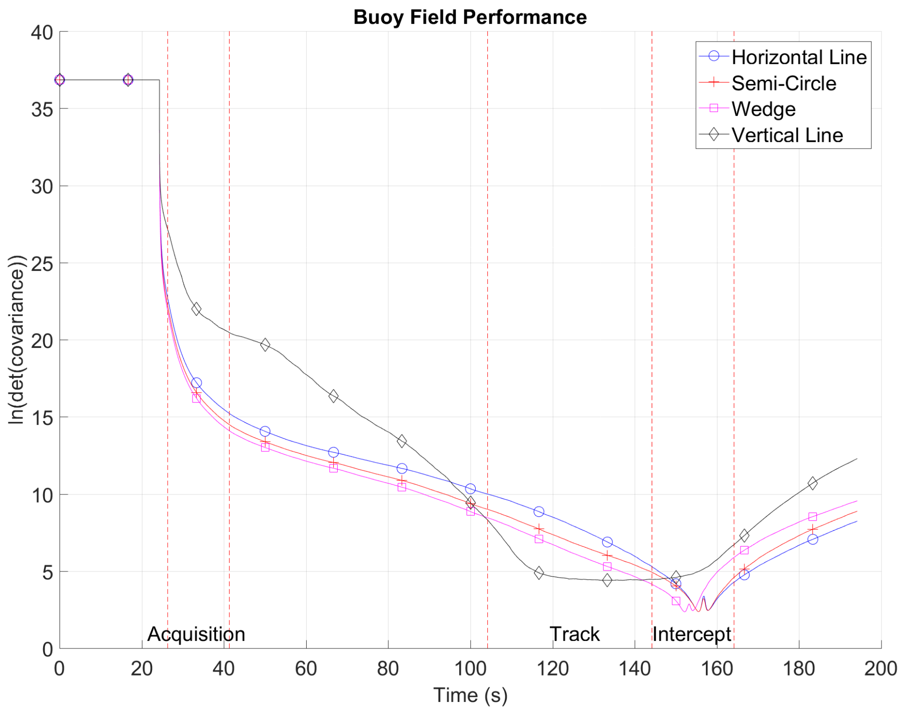

Figure 5.2: Typical Buoy Field Deployments - Covariance vs. Time (N=12)

As the target then moves across the face of the vertical line buoy field, the performance of the state estimation increases relative to the distance from the target trajectory. Increasing the distance of the vertical line deployment from the target trajectory led to superior performance initially, but a reduction in performance over the track and intercept time intervals. From these observations it is inferred that the performance of a buoy field is dependant on the rates of change and diversity in the angular measurements from the sensors. These results corroborate Meng's observation that the quality of the measurement for an angular sensor degrades as the distance to the target increases [124]. Figure 5.14 show the vertical line buoy field performance, as calculated from the state estimators covariance matrices, for distances of 200m, 400m, 600m and 800m from the target trajectory.

Reviewing the Y positional error results for the vertical line buoy fields from the Monte Carlo simulation, given in Figures 5.15 and 5.16, we observe performance consistent with the the determinant of the covariance plot given by Figure 5.14. Reviewing the 95% confidence ellipses for the key time intervals (refer to Figures 5.17 and 5.18) from one of the simulated runs we observe that the ellipses have a high eccentricity with a vertical line buoy field configuration. This occurs due to poor angular diversity of the sensor network over the target trajectory. During the acquisition time interval, the vertical line buoy field has the worst angular diversity amongst the sensors, and exhibits high levels of asymmetry in the 95% confidence ellipse in the x-y plane. This is shown in Figure 5.17. As the target crosses the face of the vertical line buoy field the asymmetry reduces as the angular diversity in the sensor network increases, shown in Figure 5.18.

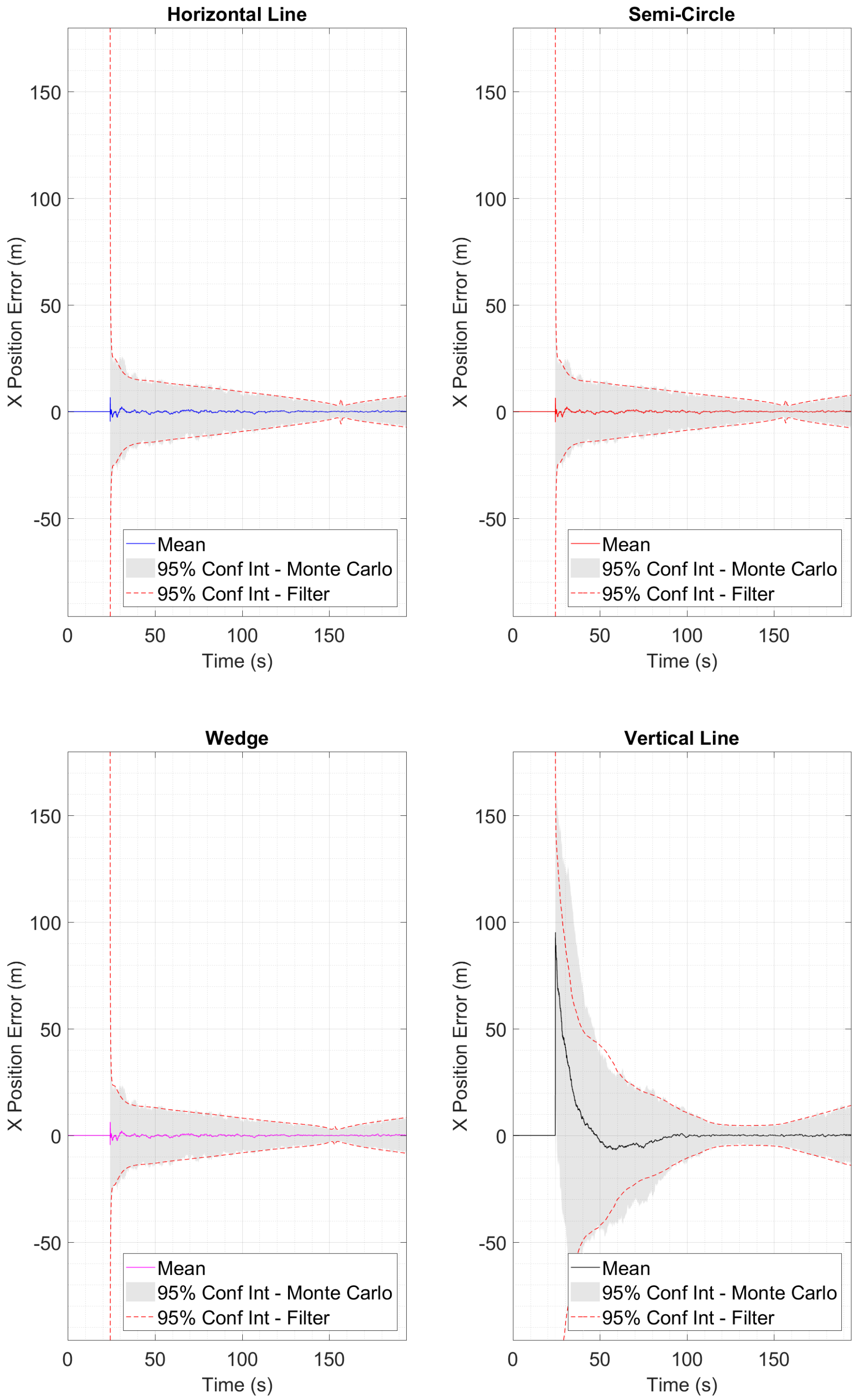

Figure 5.3: Typical Buoy Field Deployments - X Position Error (N=12)

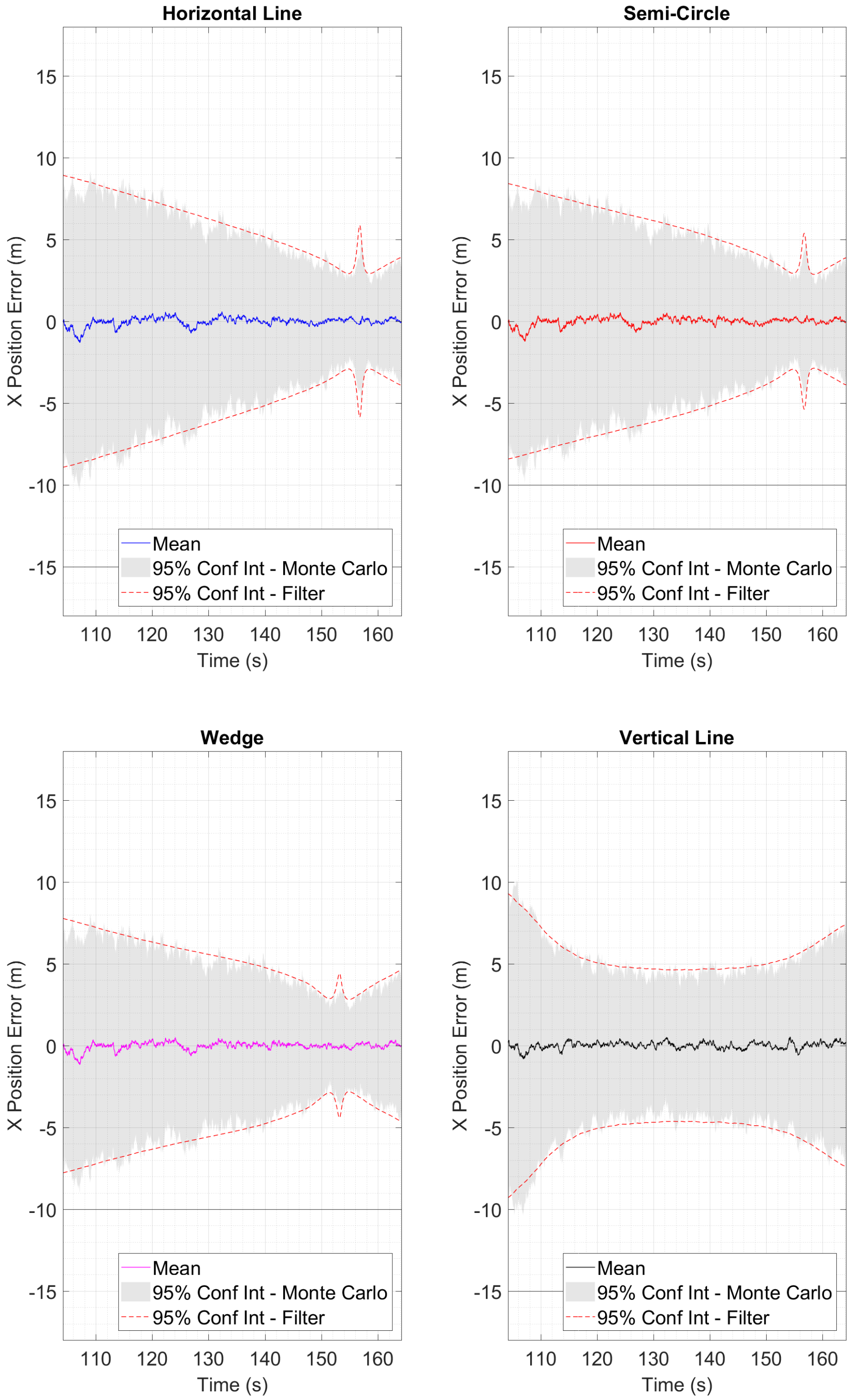

Figure 5.4: Typical Buoy Field Deployments - X Position Error - Track and Intercept Time Intervals (N=12)

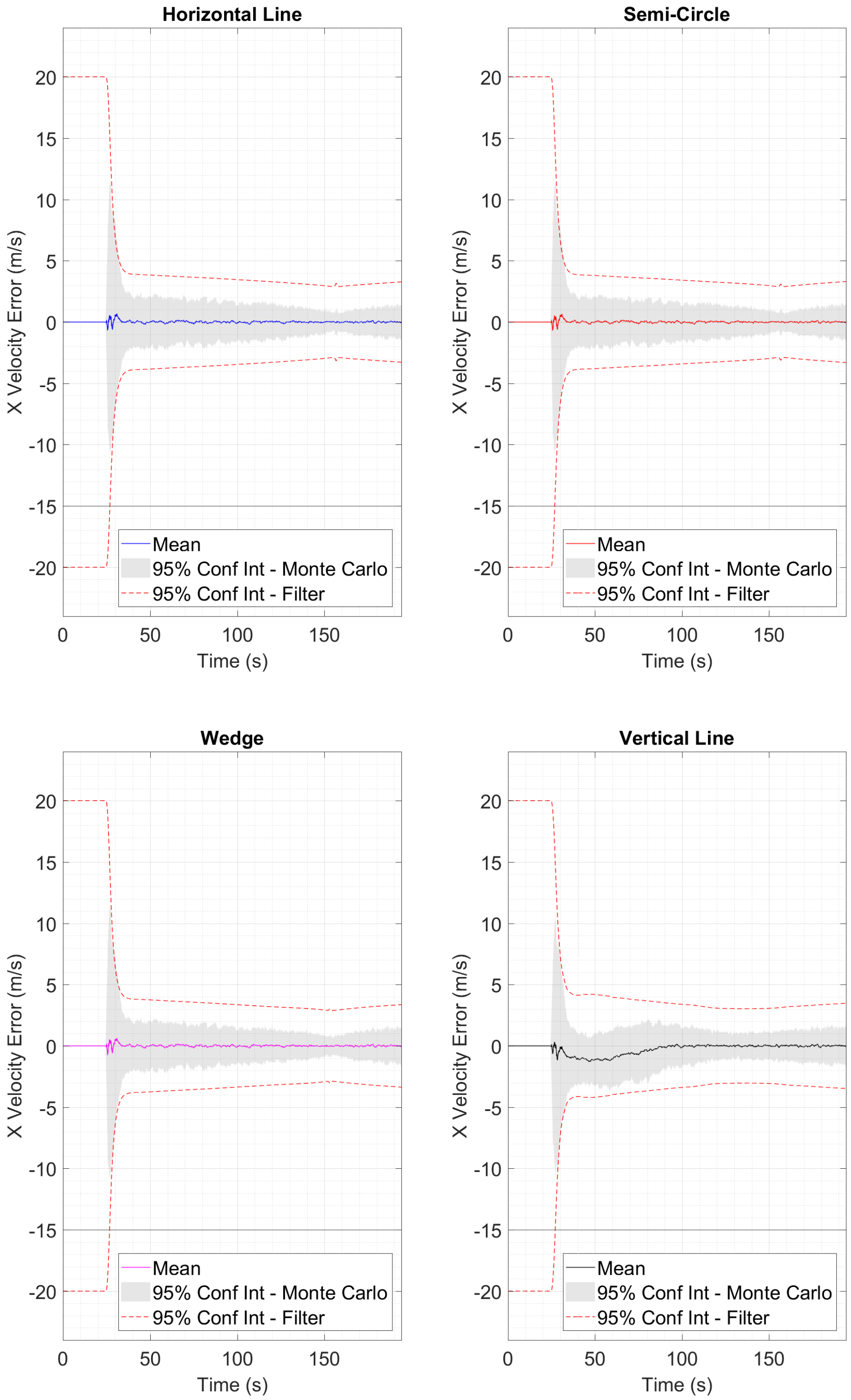

Figure 5.5: Typical Buoy Field Deployments - X Velocity Error (N=12)

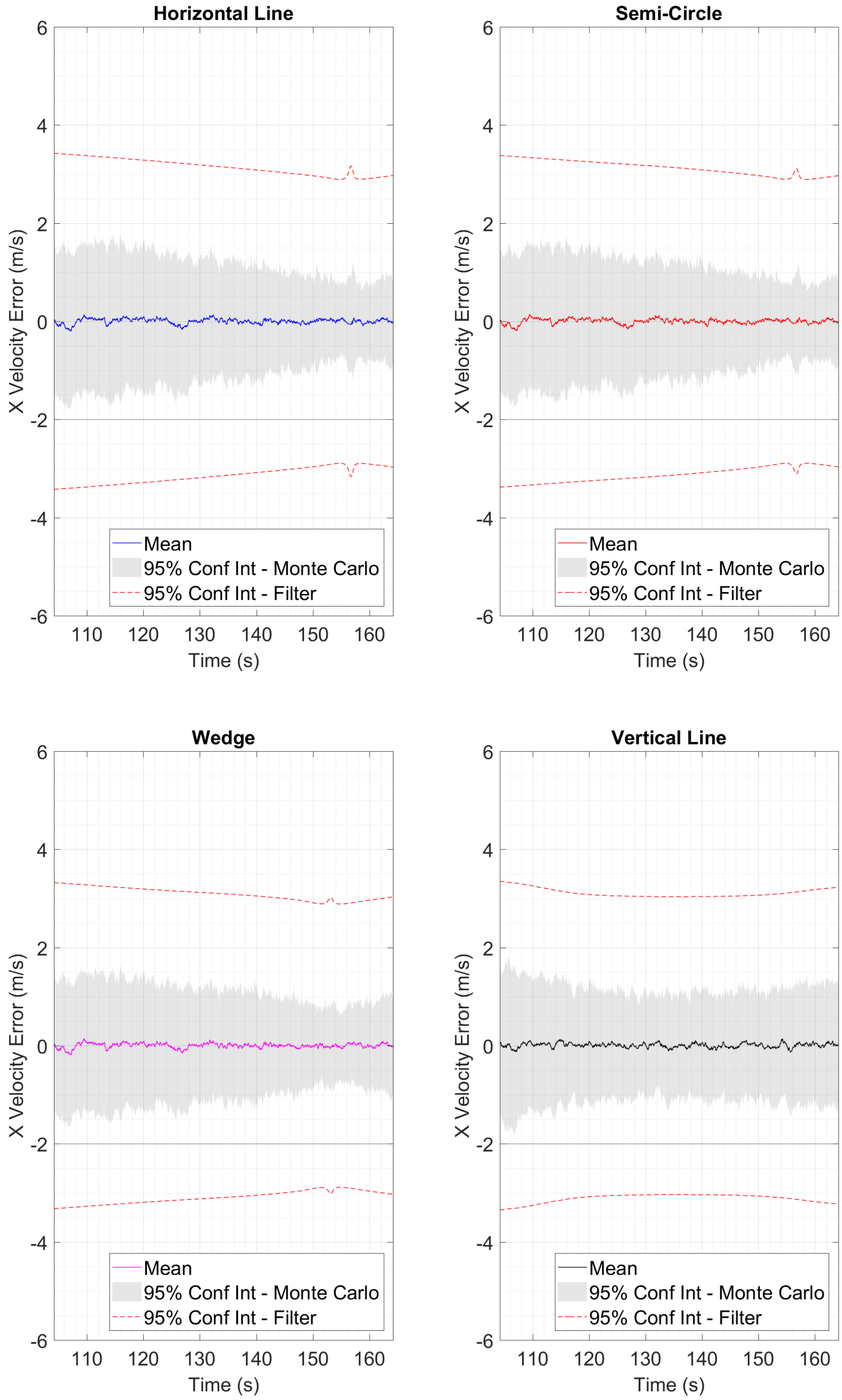

Figure 5.6: Typical Buoy Field Deployments - X Velocity Error - Track and Intercept Time Intervals (N=12)

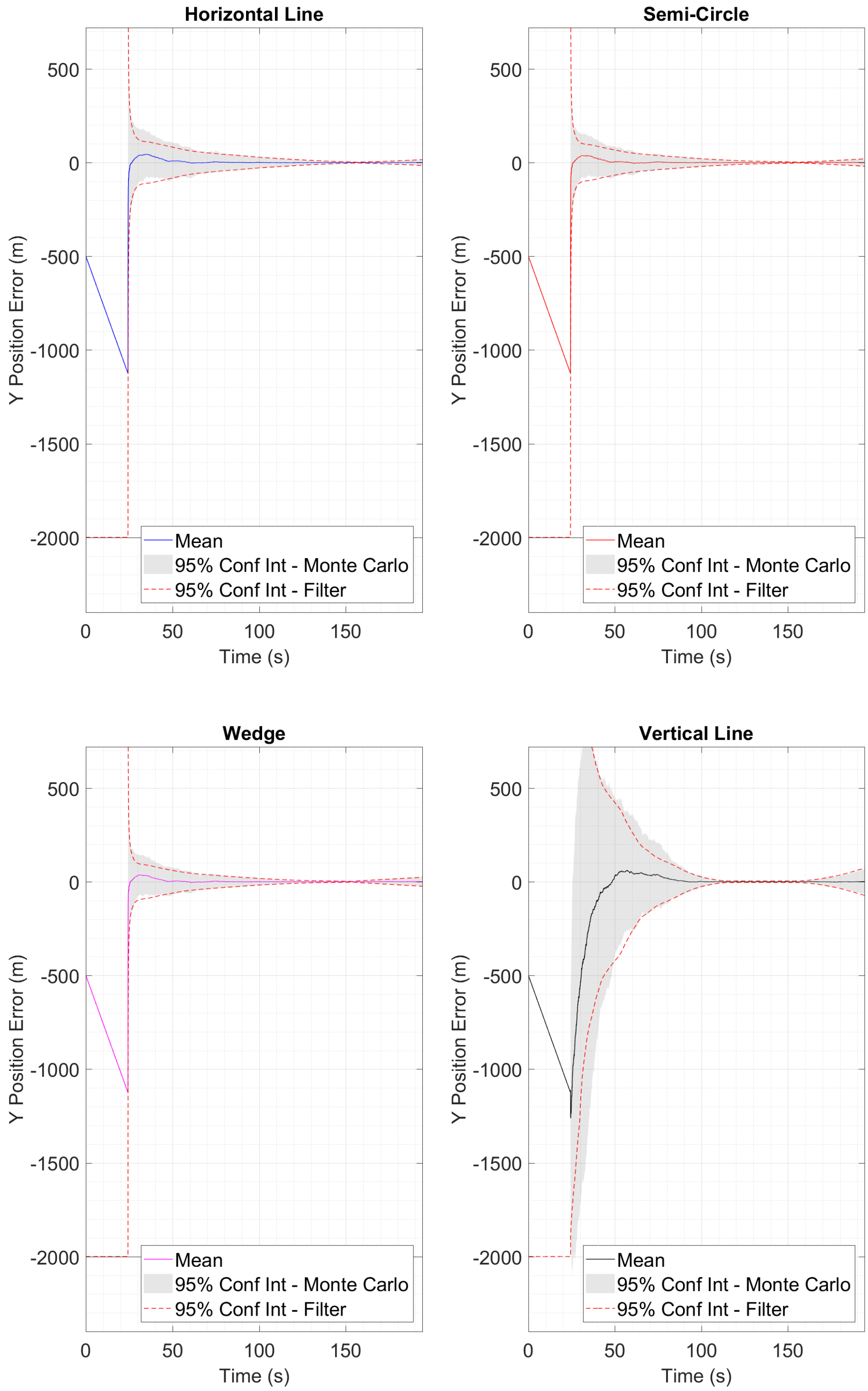

Figure 5.7: Typical Buoy Field Deployments - Y Position Error (N=12)

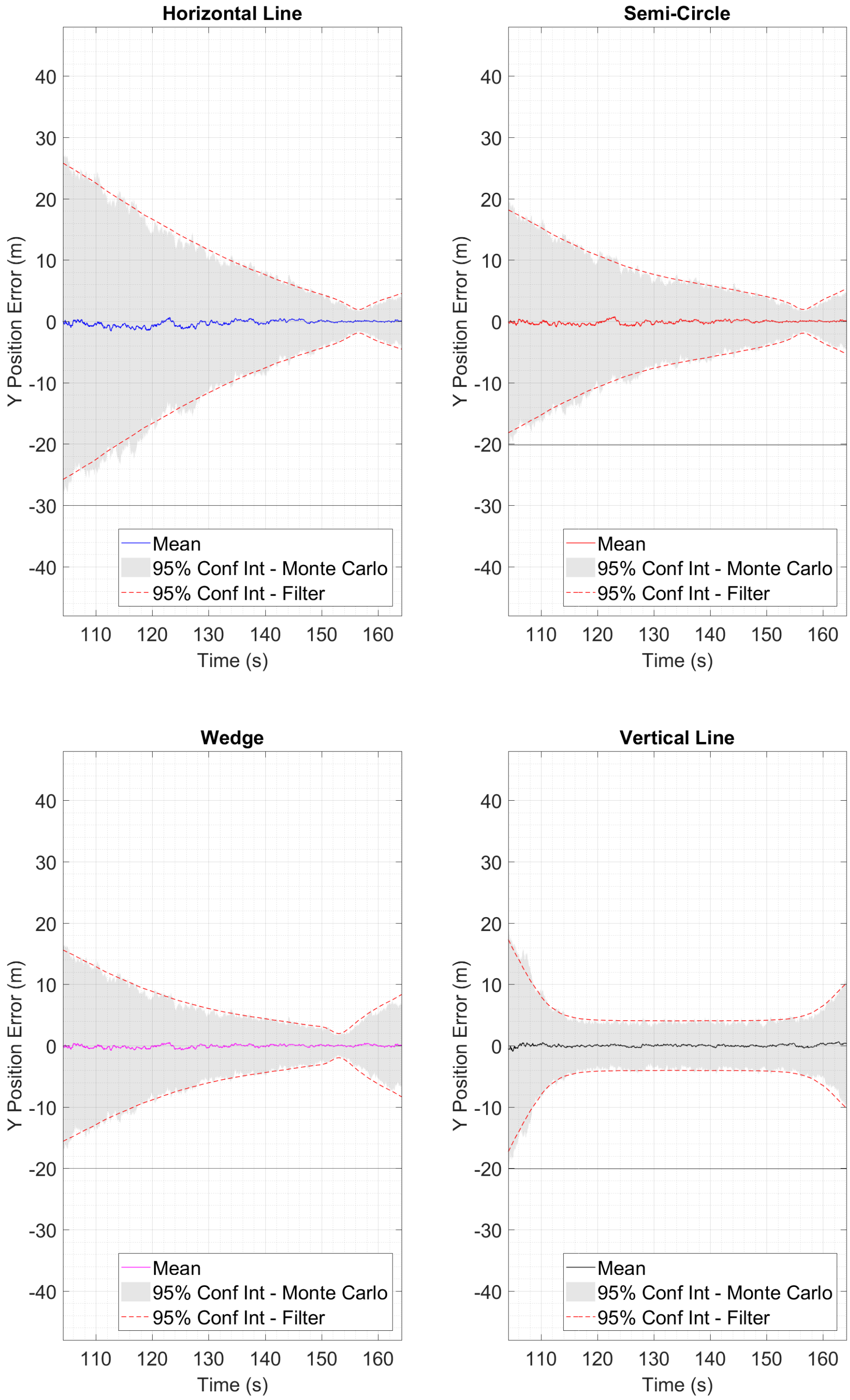

Figure 5.8: Typical Buoy Field Deployments - Y Position Error - Track and Intercept Time Intervals (N=12)

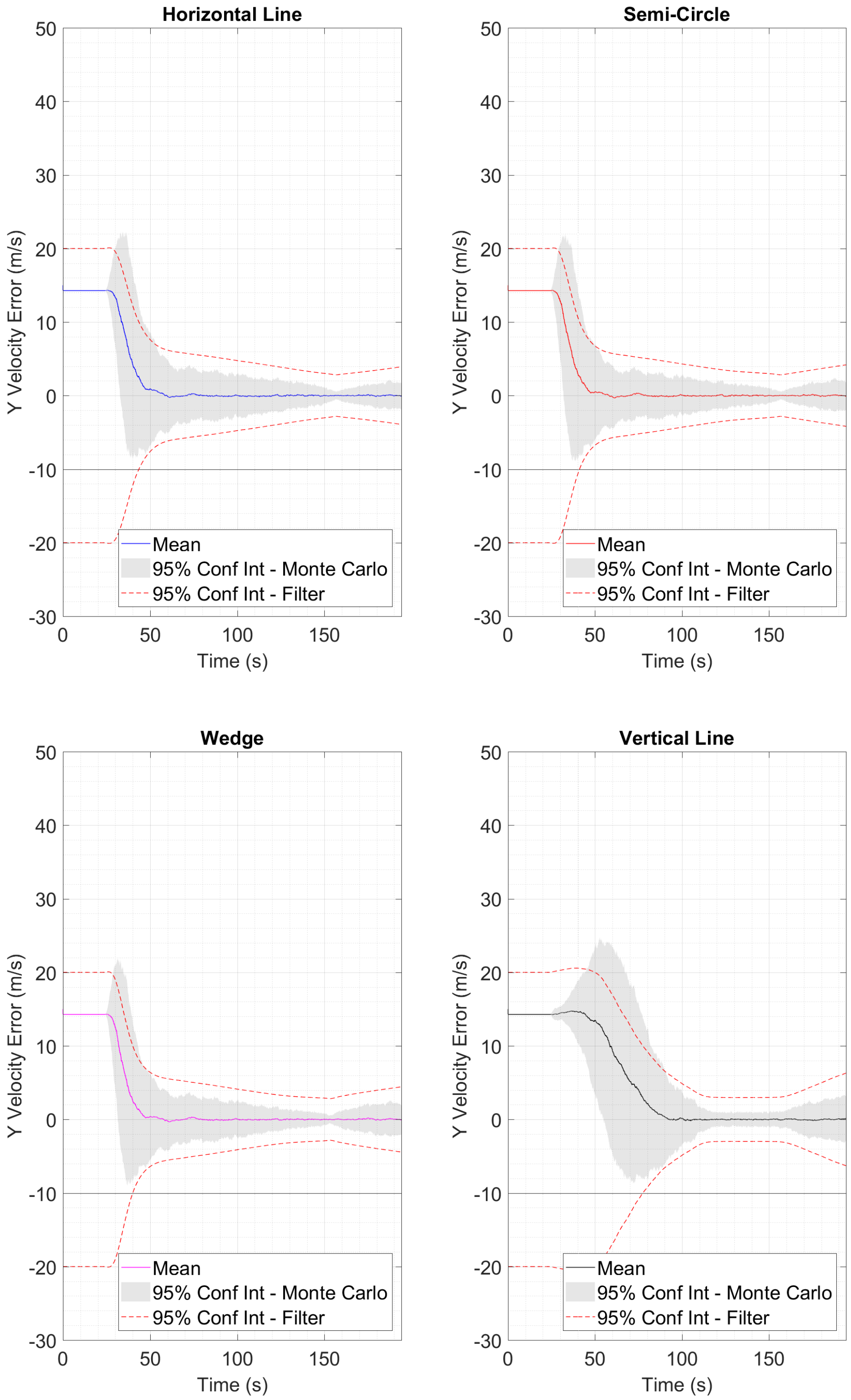

Figure 5.9: Typical Buoy Field Deployments - Y Velocity Error (N=12)

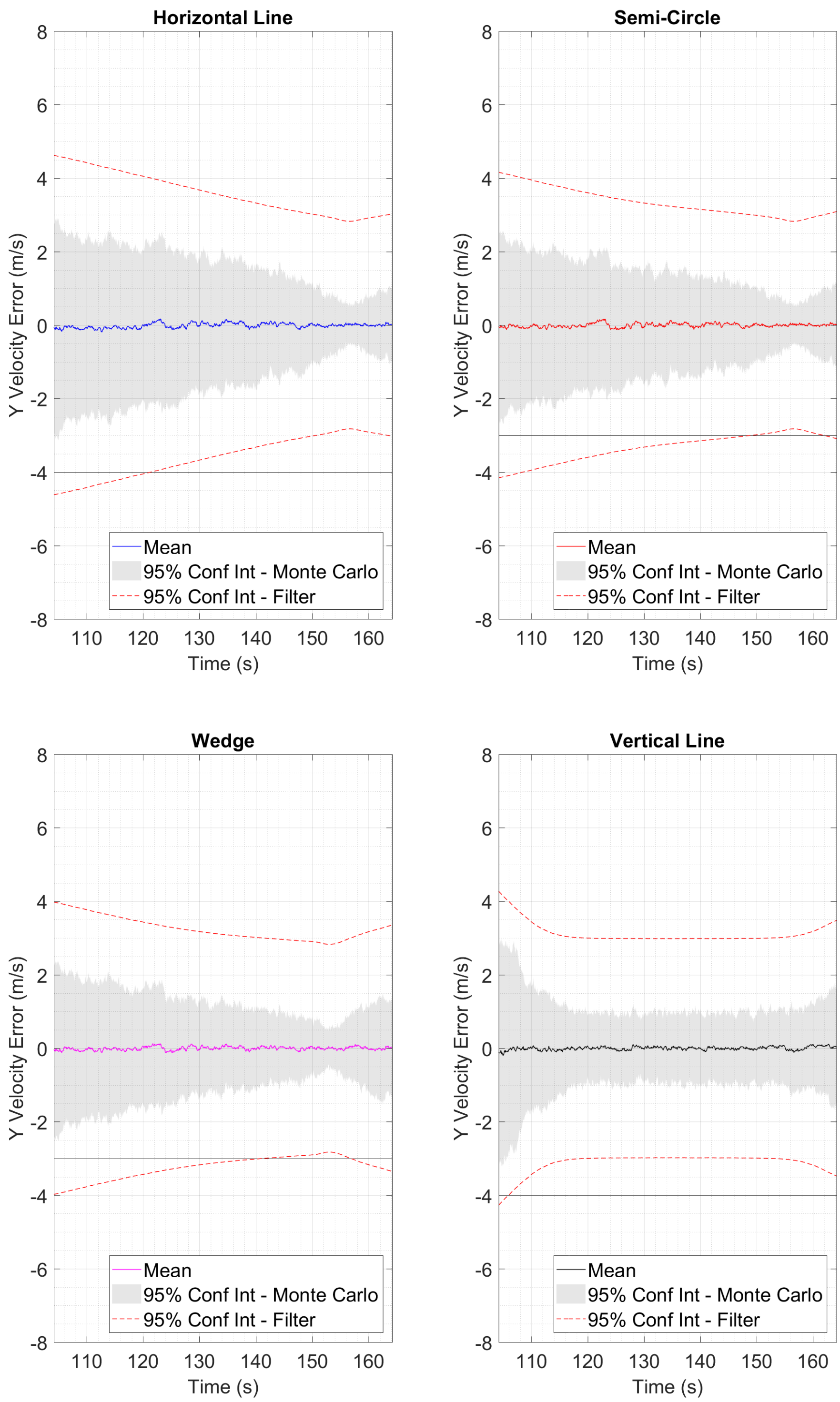

Figure 5.10: Typical Buoy Field Deployments - Y Velocity Error - Track and Intercept Time Intervals (N=12)

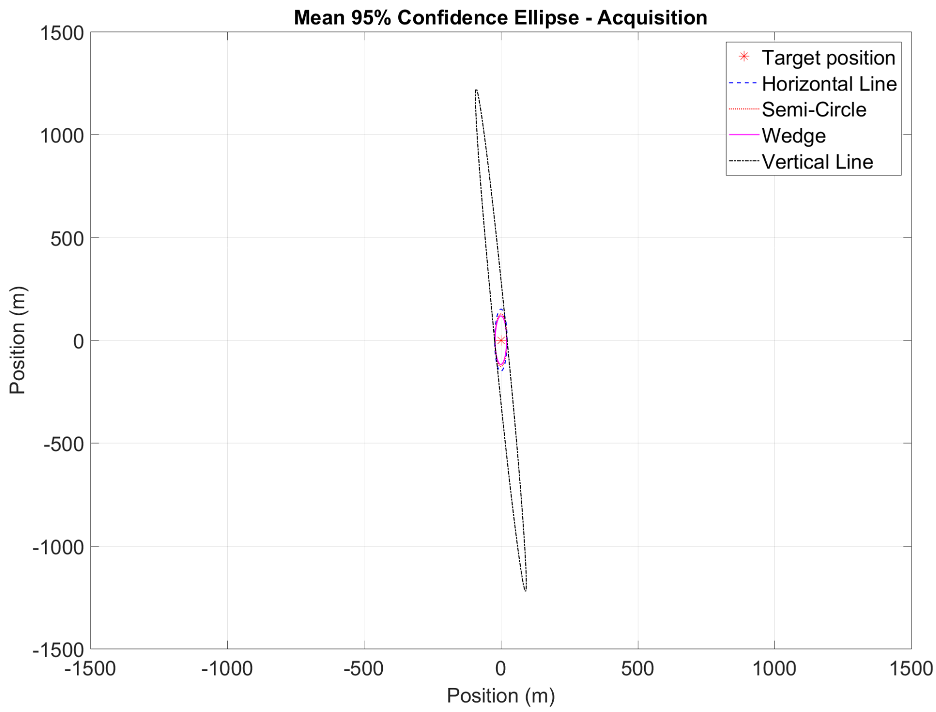

Figure 5.11: Typical Buoy Field Deployments - Mean 95% Confidence Ellipse - Acquisition Time Interval (N=12)

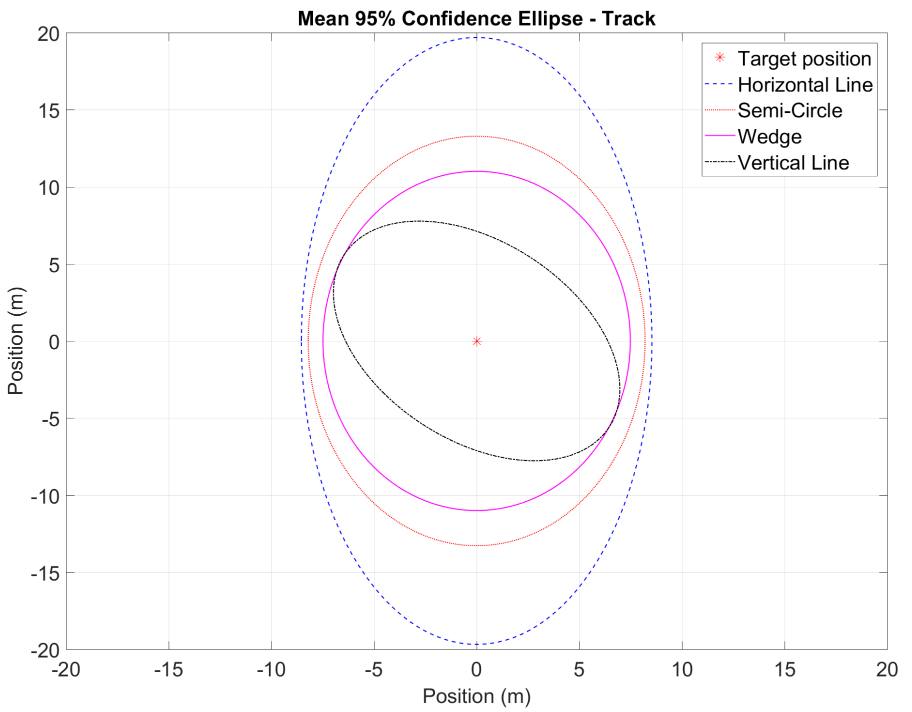

Figure 5.12: Typical Buoy Field Deployments - Mean 95% Confidence Ellipse - Track Time Interval (N=12)

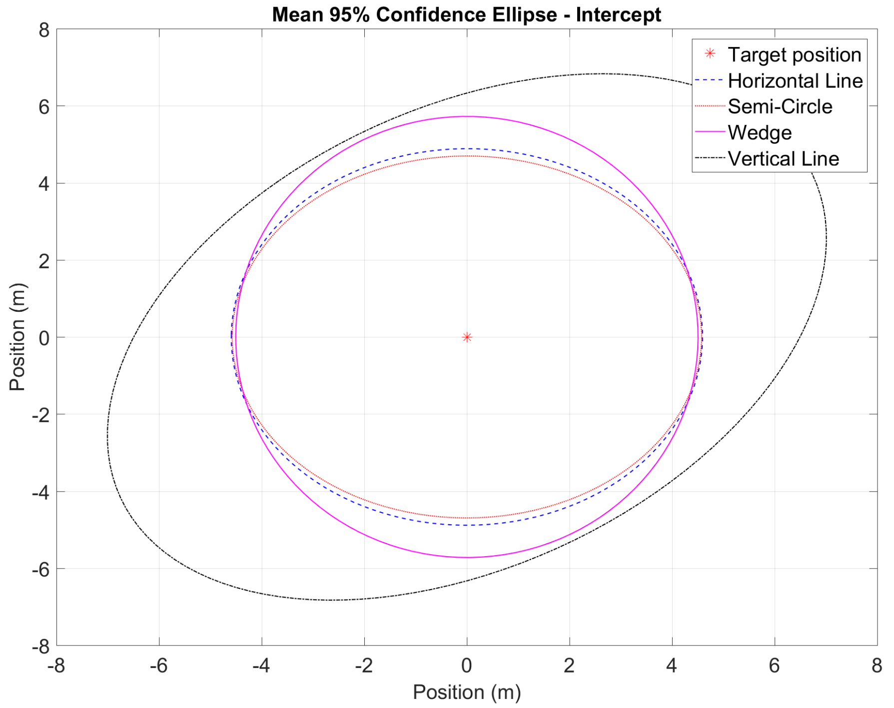

Figure 5.13: Typical Buoy Field Deployments - Mean 95% Confidence Ellipse - Intercept Time Interval (N=12)

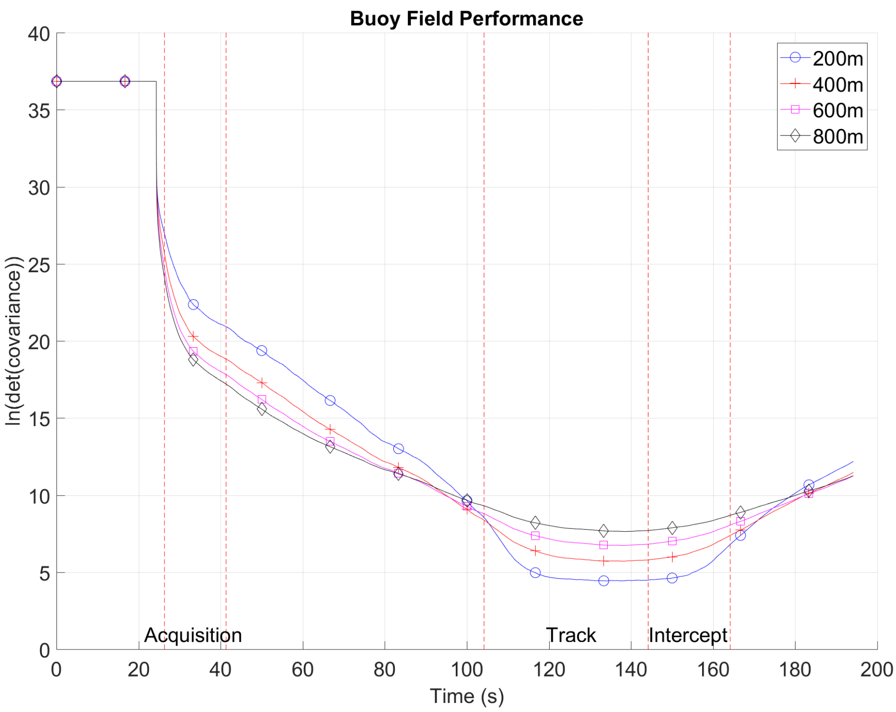

Figure 5.14: Vertical Line Buoy Field Deployments with varying orthogonal distances from target trajectory - Covariance vs. Time

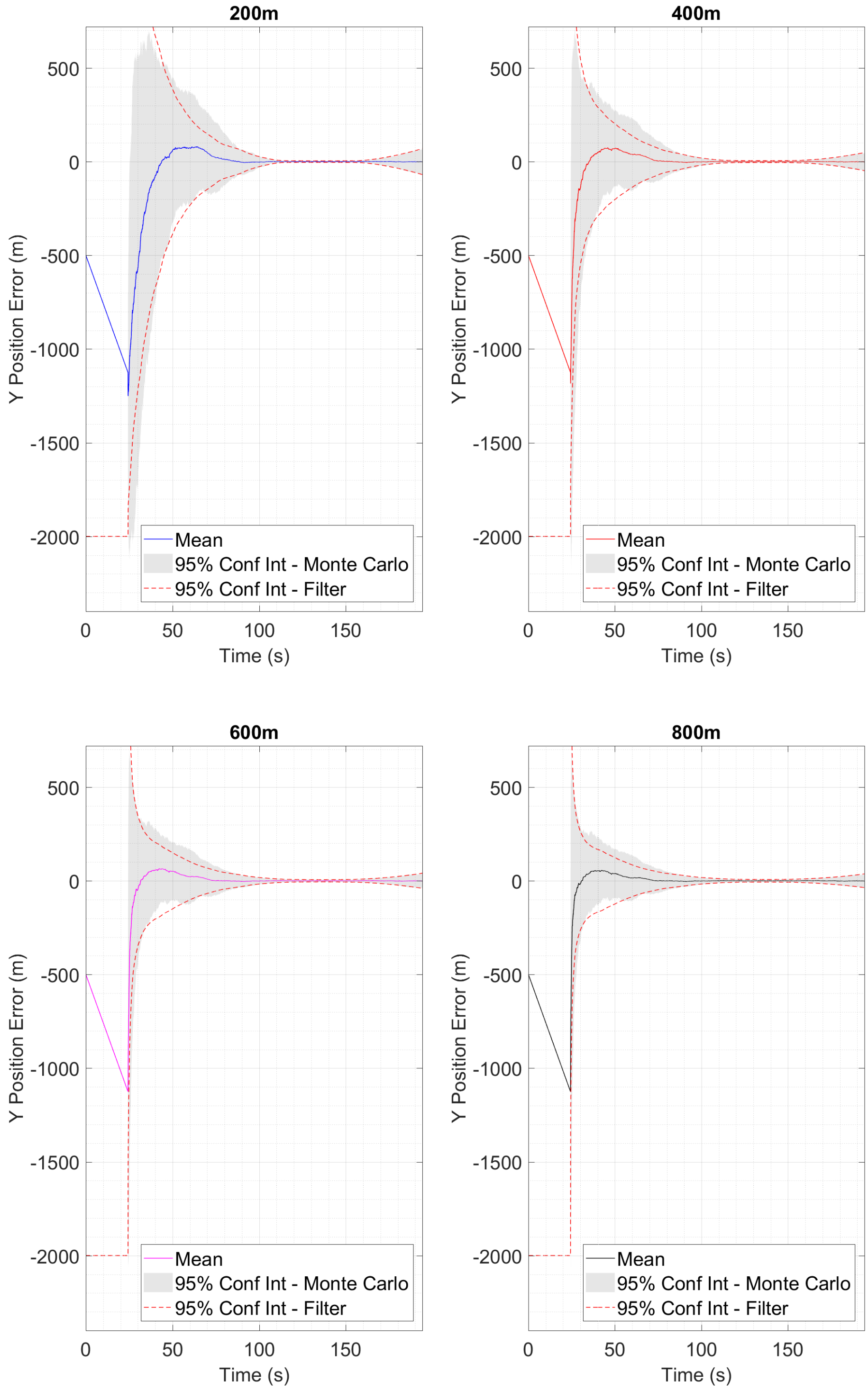

Figure 5.15: Vertical Line Buoy Field Deployments with varying orthogonal distances from target trajectory - Y Position Error (N=12)

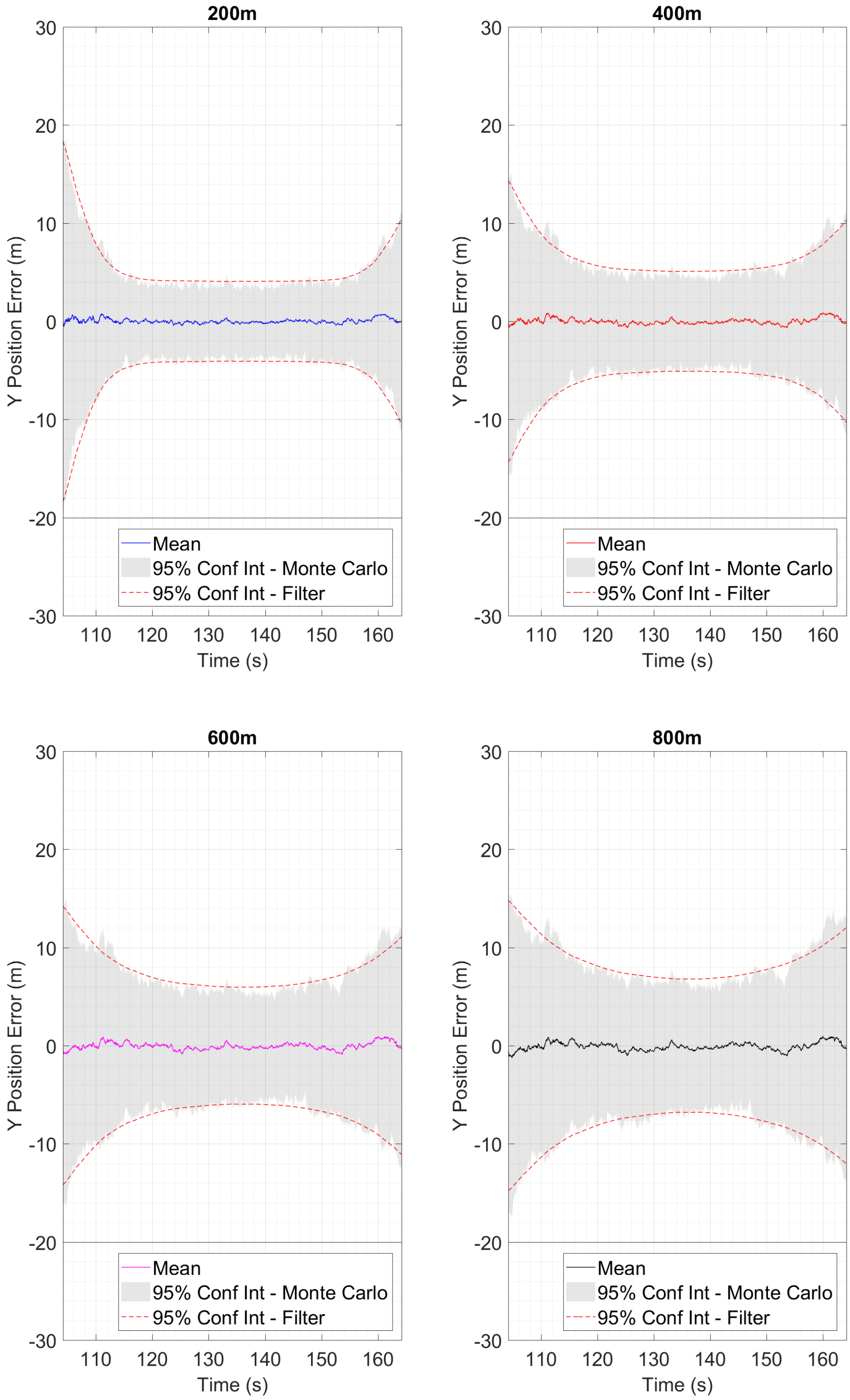

Figure 5.16: Vertical Line Buoy Field Deployments with varying orthogonal distances from target trajectory - Track and Intercept Time Intervals - Y Position Error (N=12)

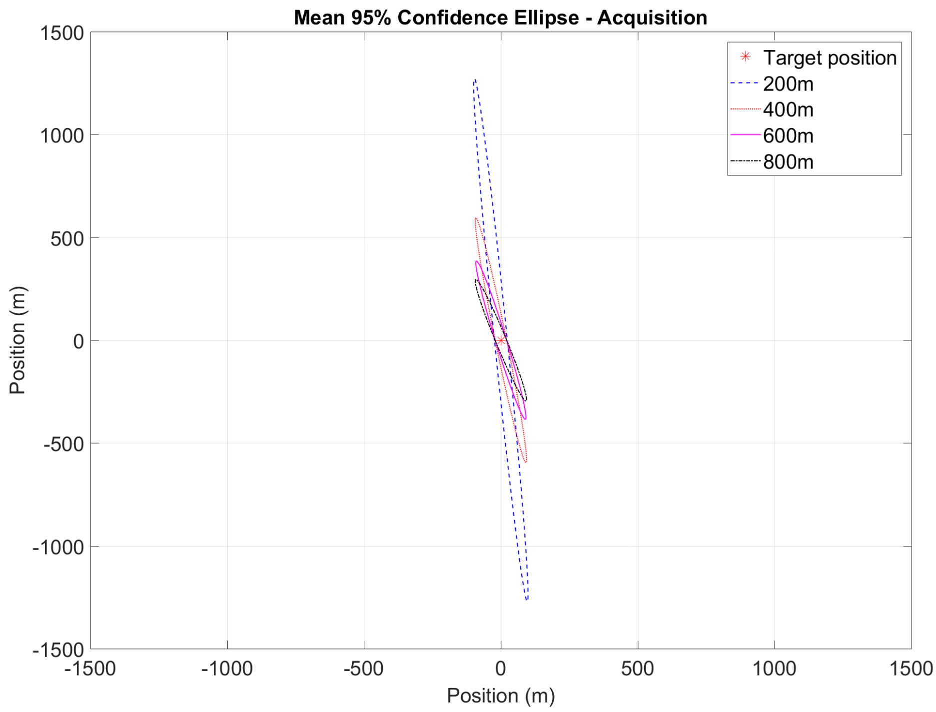

Figure 5.17: Vertical Line Buoy Field Deployments with varying orthogonal distances from target trajectory - Mean 95% Confidence Ellipse - Acquisition Time Interval

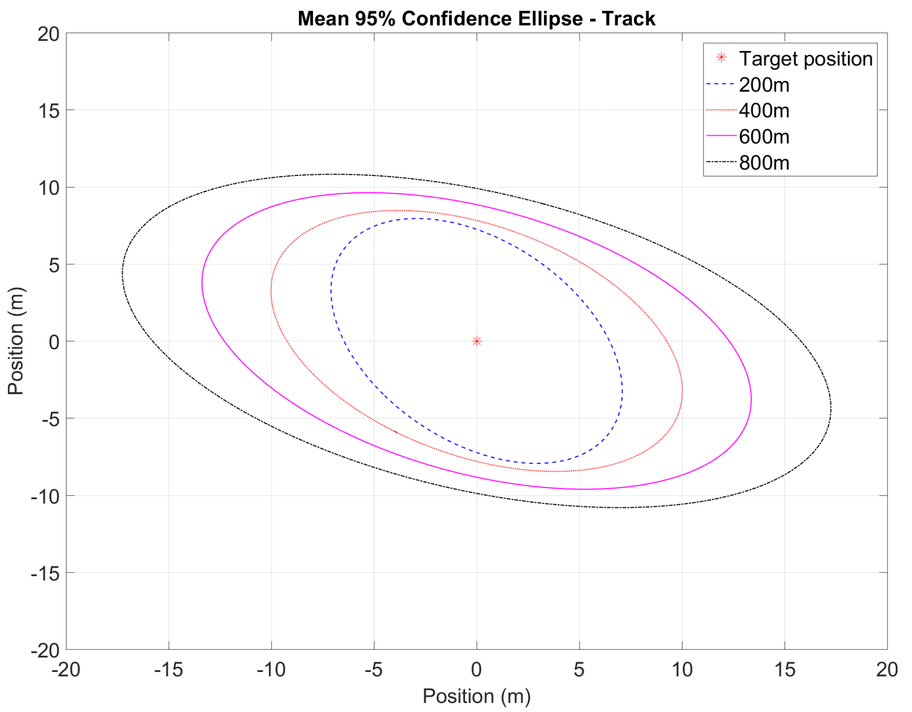

Figure 5.18: Vertical Line Buoy Field Deployments with varying orthogonal distances from target trajectory - Mean 95% Confidence Ellipse - Track Time Interval

5.3 Impact number of buoys has on buoy field

With a better understanding of the impact the buoy field geometry plays with the target tracking state estimator performance, we now wish to explore the impact the number of buoys has. Intuition would lead us to presume the greater number of sensors, the better the performance of the state estimator so long as they are adequately placed. To explore the effect the number of buoys has on performance, we used the Wedge deployment as this was found to be the best performing deployment from Section 5.2.

A Monte Carlo simulation of 100 runs for each buoy field configuration was executed. The buoy field configurations was a wedge deployment with an increasing amount of buoys, N=2,4,6,8,10 and 12. For N=2, the two buoys were placed at (-167, 167) and (167,167). For N=4, the four buoys were placed at (-500,500), (-167, 167), (167, 167) and (500, 500). For N=6 through to N=12, the buoys were evenly distributed between (-500,500) and (500,500) in a wedge arrangement according to the formula $x = -500 + \frac{1000}{N-1}(n-1)$ and $y = \left|-500 + \frac{1000}{N-1}(n-1)\right|$. This deployment arrangement is shown n Figure 5.19 for N=2,4,6 and 8.

The results were as expected. The more buoys in the buoy field led to better performance of the state estimator. There were diminishing returns with an increase in buoys, with the difference between N=12 verses N=8 seeing only a marginal performance gain as shown in Figure 5.27. What we did see is that increasing the number of buoys in the wedge deployment allowed for tighter "gates" for the target to pass through, which just prior to passing is when the confidence of the position estimate peaks. Nominally, the wedge deployment acts as a funnel, with the target passing through the inner most pair of the wedge.

Additional buoys to a fixed deployment pattern such as a wedge with a fixed spread provided little benefit past a nominal amount, in this case N=6. To get best performance for target tracking, additional buoys are better placed elsewhere in the environment to contribute higher value sensor information about the target. Such buoy placement are explored in Section 5.4.

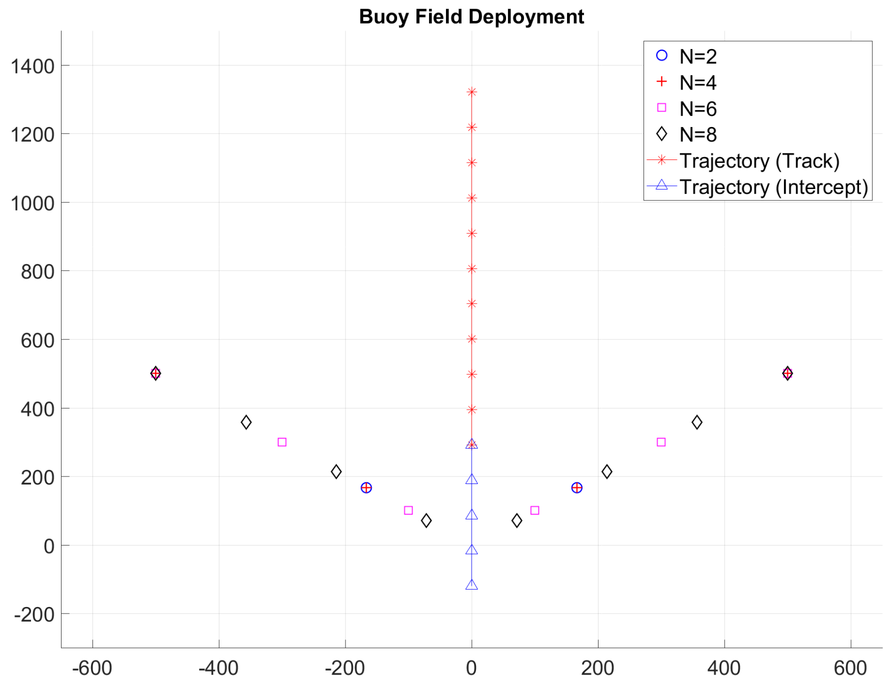

Figure 5.19: Wedge Buoy Field Deployments ($N = 2, \ldots, 8$)

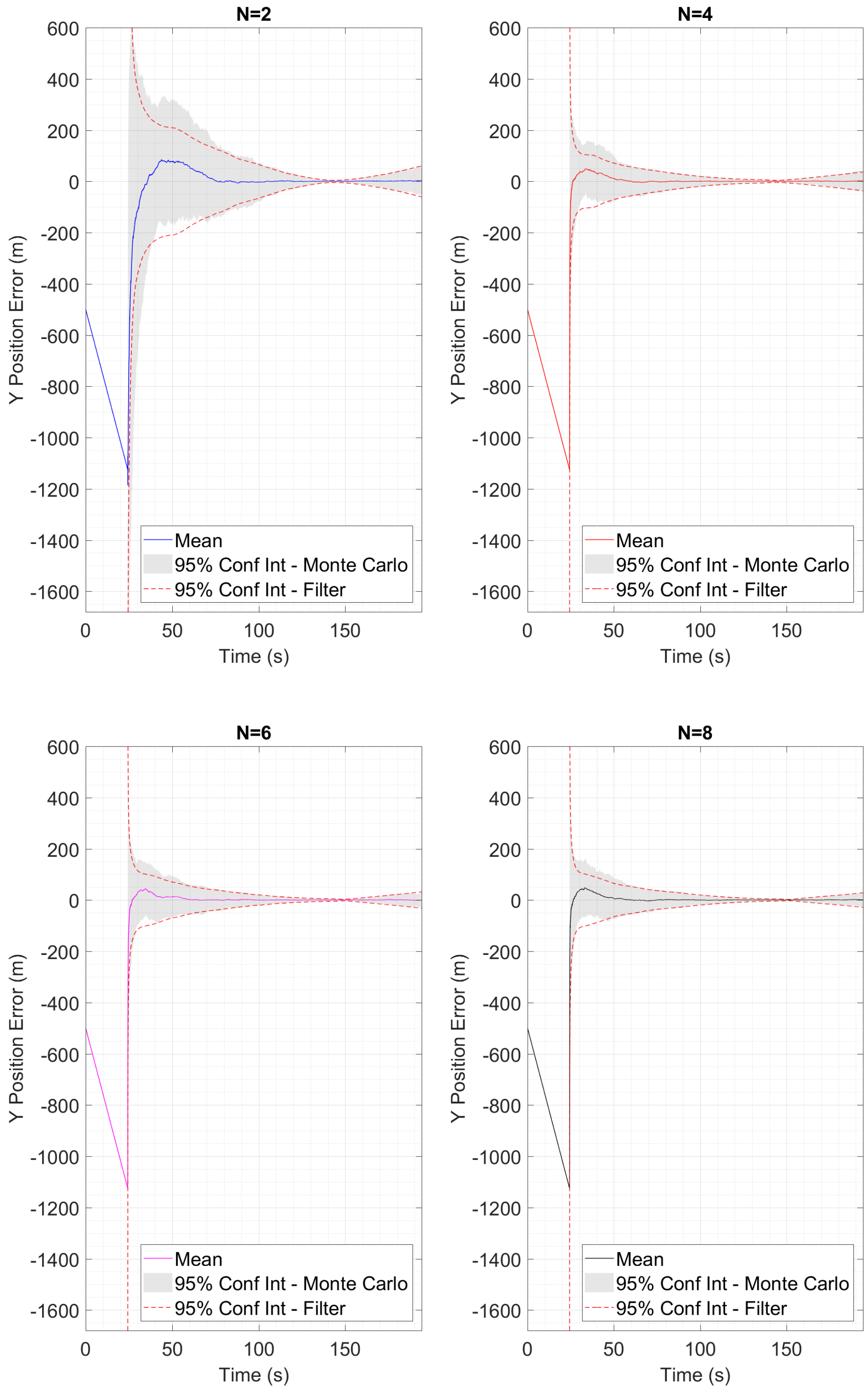

Figure 5.20: Wedge Buoy Field Deployments ($N = 2, \ldots, 8$) - Y Position Error

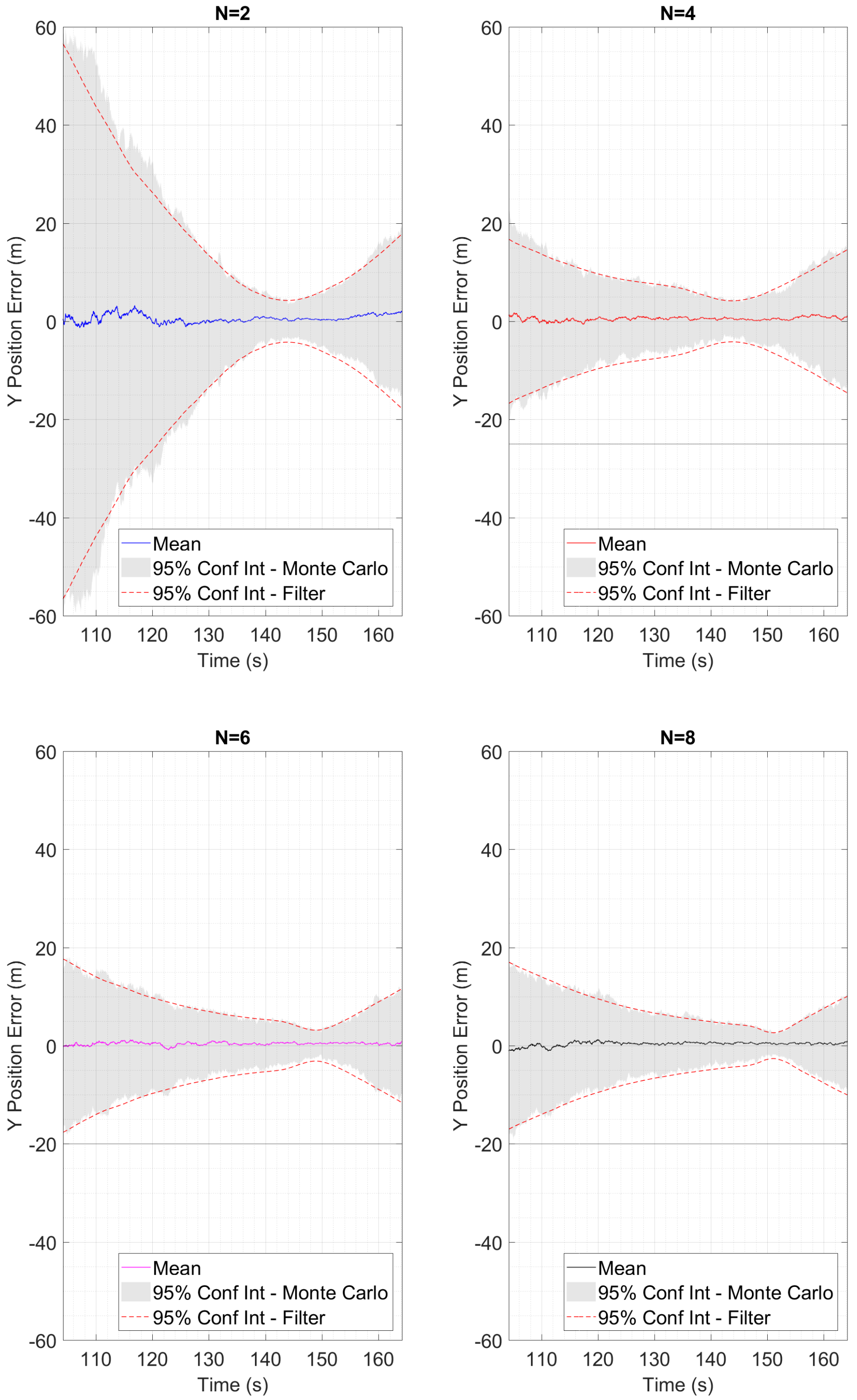

Figure 5.21: Wedge Buoy Field Deployments ($N = 2, \ldots, 8$) - Track and Intercept Time Intervals - Y Position Error

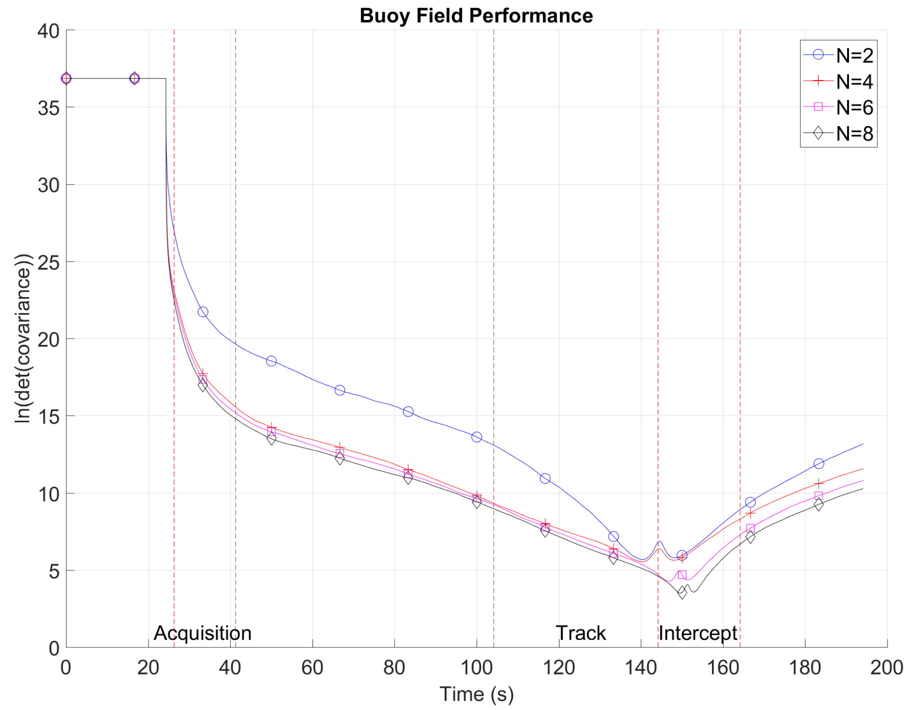

Figure 5.22: Wedge Buoy Field Deployments ($N = 2, \ldots, 8$) - Covariance vs. Time

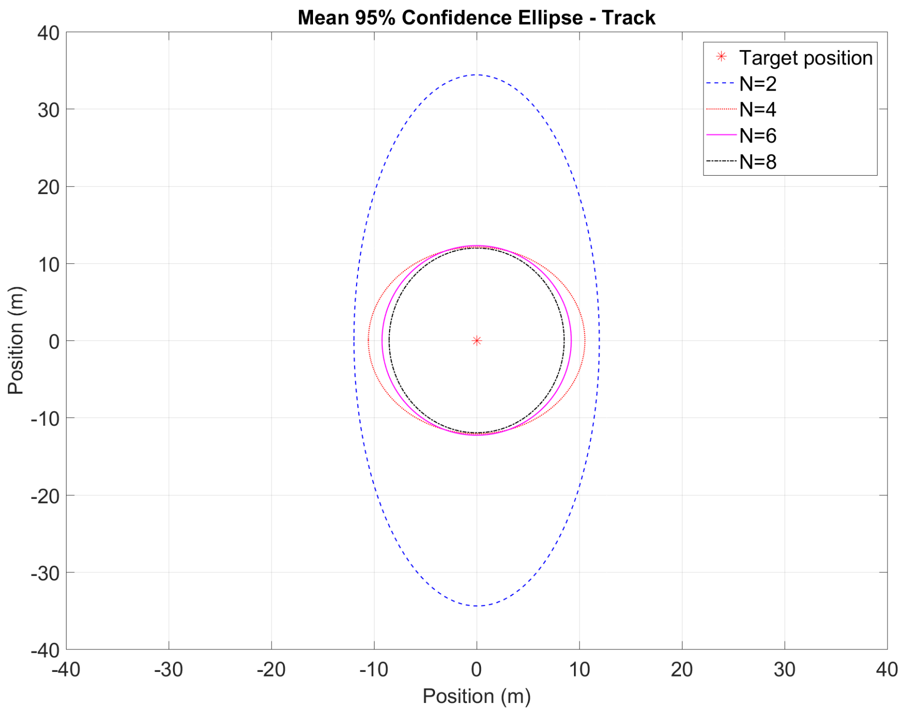

Figure 5.23: Wedge Buoy Field Deployments ($N = 2, \ldots, 8$) - Mean 95% Confidence Ellipse - Track Time Interval

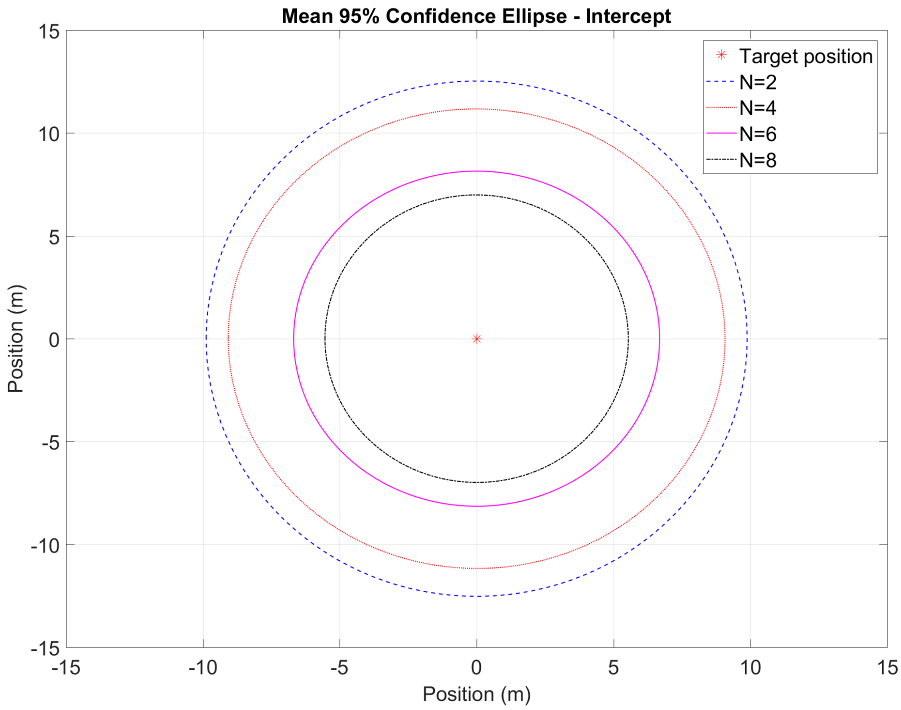

Figure 5.24: Wedge Buoy Field Deployments ($N = 2, \ldots, 8$) - Mean 95% Confidence Ellipse - Intercept Time Interval

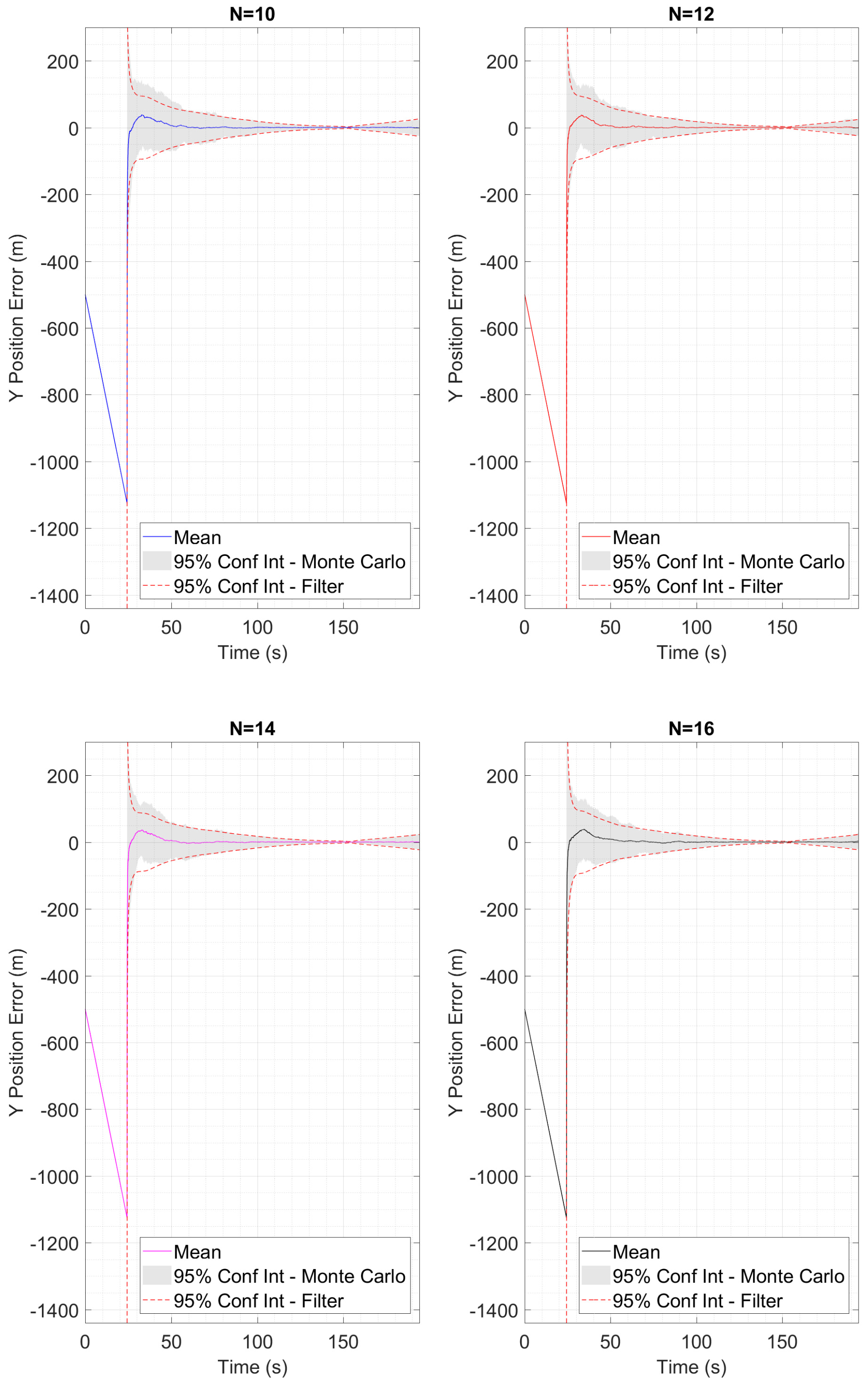

Figure 5.25: Wedge Buoy Field Deployments ($N = 10, \ldots, 16$) - Y Position Error

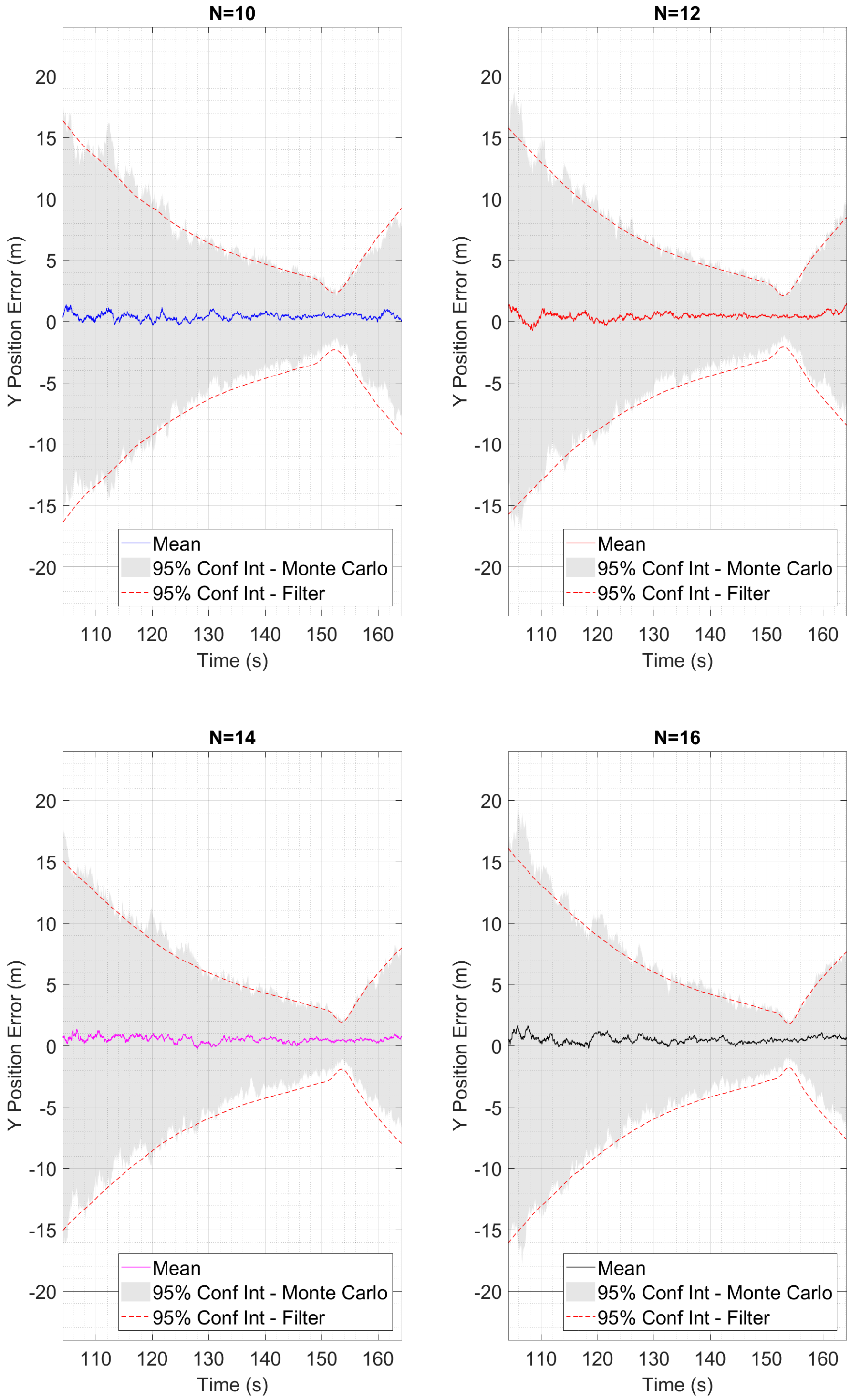

Figure 5.26: Wedge Buoy Field Deployments ($N = 10, \ldots, 16$) - Track and Intercept Time Intervals - Y Position Error

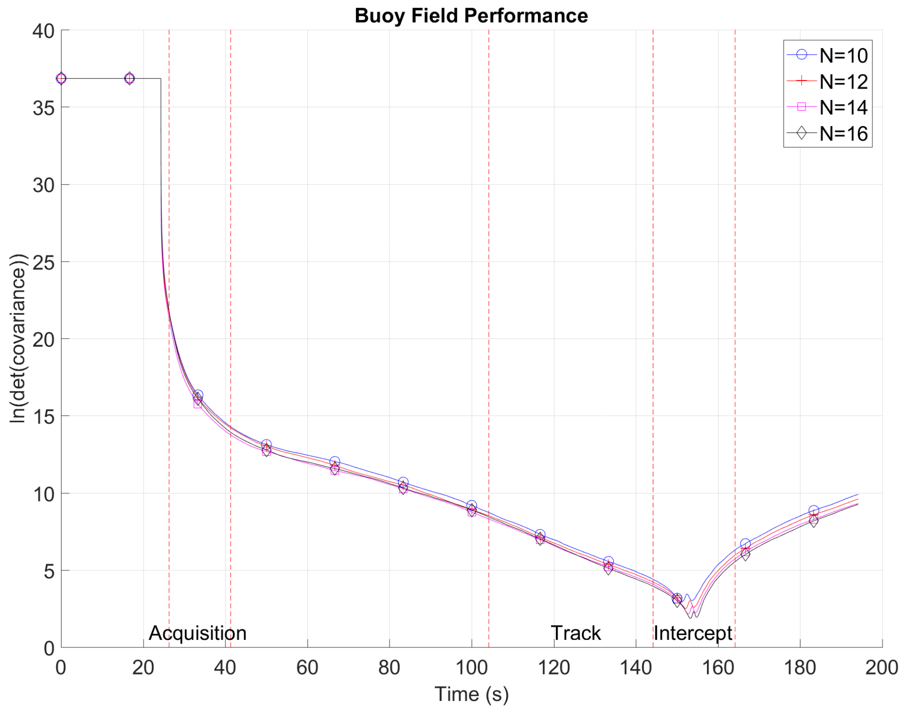

Figure 5.27: Wedge Buoy Field Deployments ($N = 10, \ldots, 16$) - Covariance vs. Time

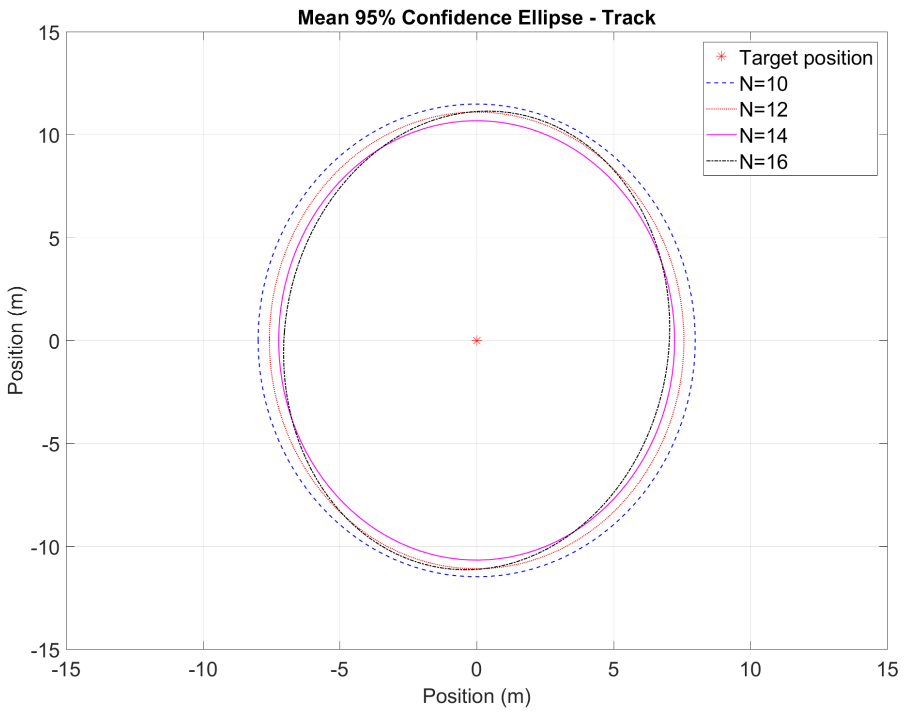

Figure 5.28: Wedge Buoy Field Deployments ($N = 10, \ldots, 16$) - Mean 95% Confidence Ellipse - Track Time Interval

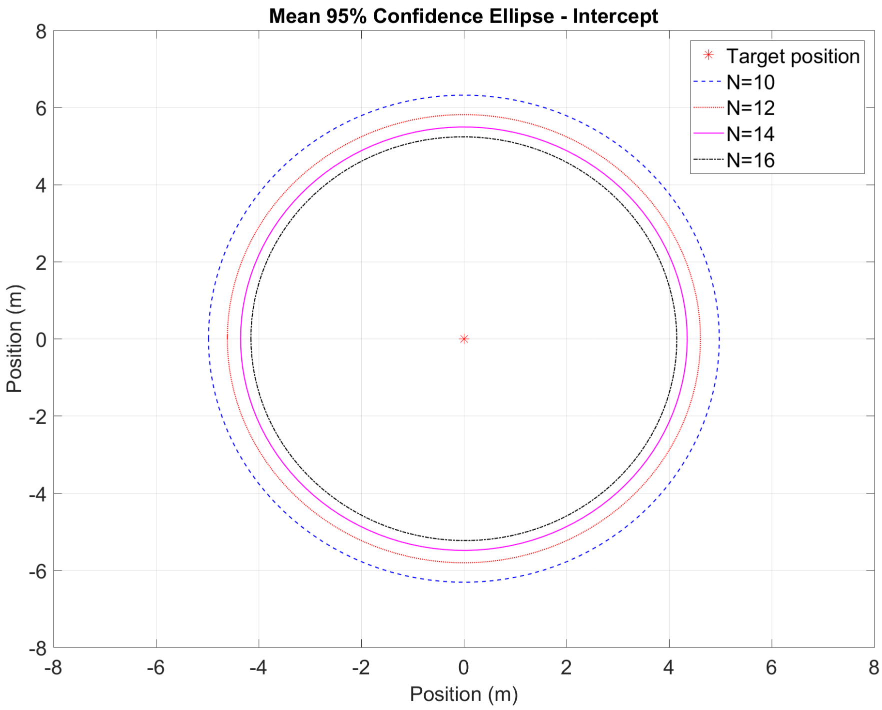

Figure 5.29: Wedge Buoy Field Deployments (N = 10, . . . , 16) - Mean 95% Confidence Ellipse - Intercept Time Intercept

5.4 Finding an Optimal Buoy Field using a Genetic Algorithm

Having explored buoy field performance for various sensor deployment configurations, we now look to optimise a buoy field. Utilising the simulation constructed in Section 4.2, we used a genetic algorithm (using MATLAB's `ga` function) to optimise buoy field deployments against the two cost functions defined in Section 5.1.

A genetic algorithm emulates an evolutionary process, where a population of candidate solutions to an optimisation problem are evaluated against a cost function. Each candidate solution has a set of properties that can be mutated and altered through a natural selection process, using the cost function as the fitness of the population set.

The concept of evolutionary algorithms were theorised and explored in the 1950's as computer scientists recognised evolutionary systems could be used to solve optimisation problems. Box, Friedman, Bledsoe, Bremermann, Reed, Toombs, and Baricelli all contributed to the early development of evolutionary algorithms [127]. Genetic algorithms, as we are familiar with them today, were first proposed by John Holland in the 1960s [128]. For a treatise on both the history and topic of genetic algorithms, refer to Mitchell's book "An Introduction to Genetic Algorithms (Complex Adaptive Systems)" [127].

Using the cost functions defined in Section 5.1, we constructed a genetic algorithm to solve the optimisation problem. Constraints were placed on the genetic algorithm, first constraining the buoy field positions to whole integers, and applying the following limits to the buoy field deployment.

$$\begin{Bmatrix} -500 \\ 0 \end{Bmatrix} \leq B_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \leq \begin{Bmatrix} 500 \\ 1200 \end{Bmatrix} \tag{5.4.1}$$

Figure 5.30 shows the output of one of the genetic algorithm optimisation runs for a buoy field deployment. Here we can see improvement in the population over the generations of the buoy field evolution. As the generations reach about 140 we begin to see stagnation of the population, with the genetic algorithm being unable to mutate and evolve the buoy field deployment into a better configuration.

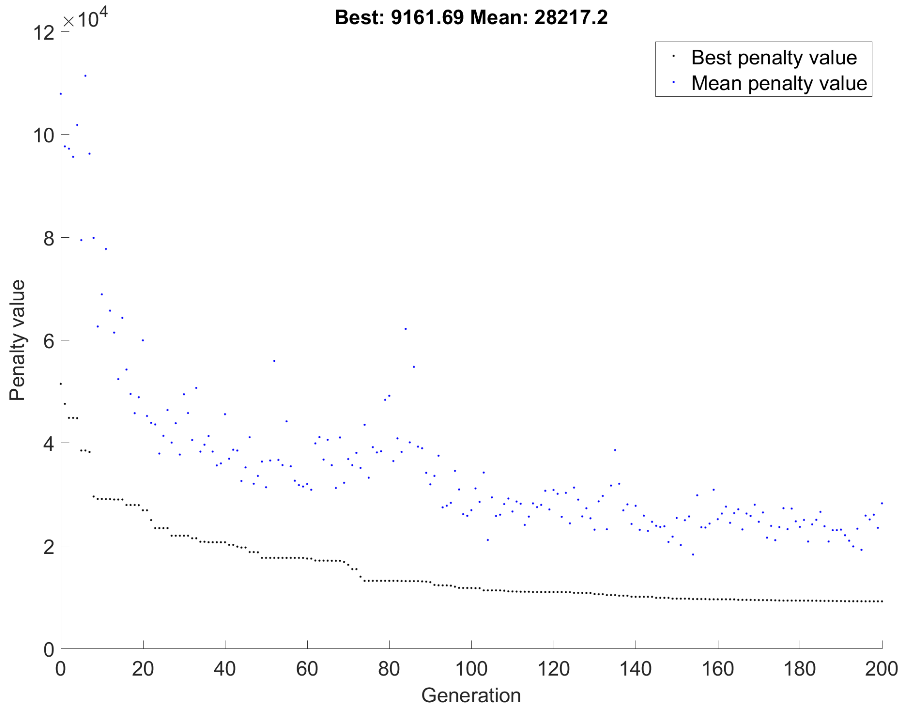

Figure 5.30: Genetic Algorithm Generation Performance

5.4.1 Optimisation using $\Phi_1(S(t_{s...e}, B_{1...N}))$

Recalling the objective function,

$$\Phi_1(S(t_{s...e}, B_{1...N})) = \sum_{t=t_s}^{t_e} (\det(P(t)) - D)^2$$

where

P = covariance matrix from simulation using buoy field $B_{1...N}$
D = desired determinant value per time instant

We executed the genetic algorithm to optimise a buoy field for N=3,4,6,8,10,12,16 and 20 against objective function Φ_1. Figure 5.31 shows the EKF's self perceived performance for the best candidate from each of the genetic algorithms for N=3,4,6 and 8. Similarly, Figure 5.32 shows the EKF's self perceived performance for the best candidate from each of the genetic algorithms for N=10,12,16 and 20. Figures 5.33 to 5.36 show the Y position error from the Monte Carlo simulation for each of the buoy field configurations. Overlayed on these figures are the EKF's self perceived accuracy which was found to be consistent with the actual performance calculated from the Monte Carlo simulation. Like previous buoy field configurations, over the acquisition time interval there is an initial error bias that converges towards zero as the target approaches the buoy field.

Reviewing the results we can assess, that for $N \leq 10$, the genetic algorithm struggles to to

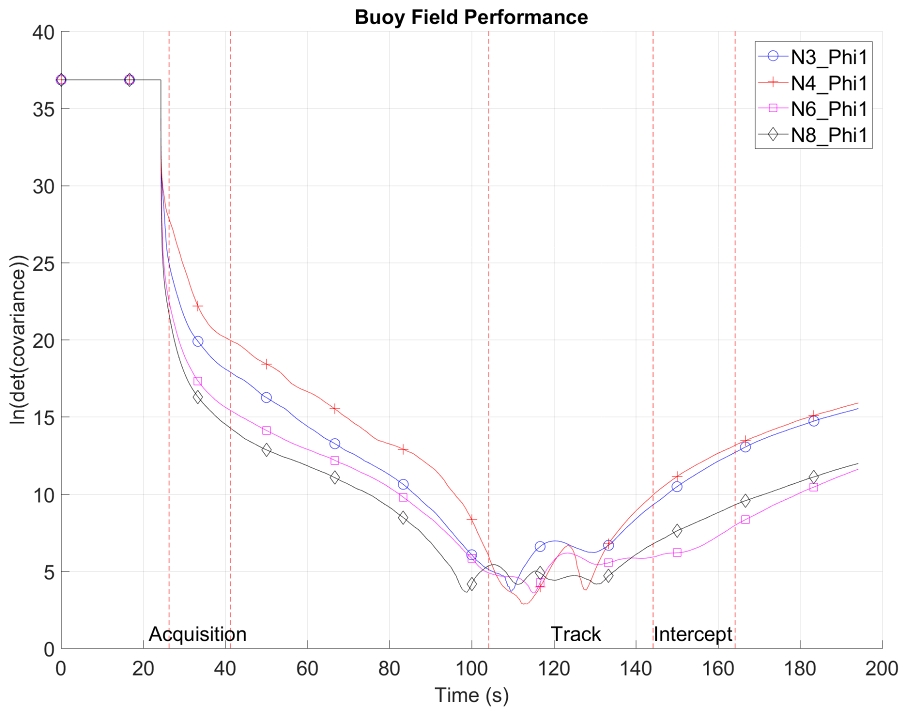

Figure 5.31: Φ_1 Genetic Algorithm Buoy Field Deployments - Covariance vs. Time

optimise against our objective function for the entirety of the track interval ($t_{trk\_s...trk\_e}$) and intercept interval ($t_{int\_s...int\_e}$). Additionally, the buoy field performance is irregular across these key time intervals, with peaks in performance arising from buoys placed in close proximity to the target trajectory.

Looking at the $N = 10$, $N = 16$ and $N = 20$ cases, we see that the genetic algorithm was able to adequately optimise the buoy field deployment against our objective function and all offer similar performance. The $N = 20$ case offers the smoothest estimator performance over the track and intercept time intervals. We note that for $N = 12$, the genetic algorithm failed to out perform the $N = 10$ case, instead stagnating at a local minimum as opposed to the superior deployment found in the $N = 10$ case.

Figures 5.37 and 5.38 show the buoy field deployments for $N = 10$ and $N = 20$, respectively, from the genetic algorithm using the Φ_1 objective function. Here we observe the structure that emerges as a relatively even distribution of buoys either side along the target trajectory. The differences between the $N = 10$ and $N = 20$ case is the amount of lateral coverage the buoy field provides, with the $N = 20$ case providing a larger distribution of buoys spread out laterally relative to the target trajectory.

These results agree with our findings gained from our prior analysis, where sustained optimal performance of a target tracking algorithm can be achieved by a sensor network that follows the target trajectory, with close proximity.

Using the Φ_1 cost function, the $N = 10$ buoy field deployment (shown in Figure 5.37) was the preferred deployment as it provided the best performance for minimising the target tracking confidence across the key time intervals with the least amount of buoys required.

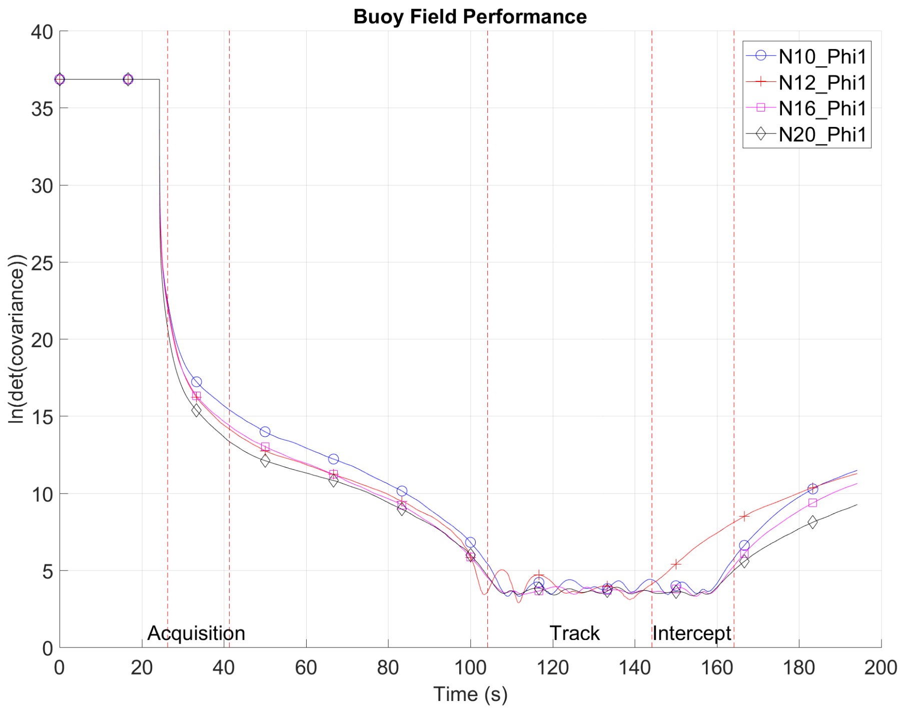

Figure 5.32: Φ_1 Genetic Algorithm Buoy Field Deployments - Covariance vs. Time

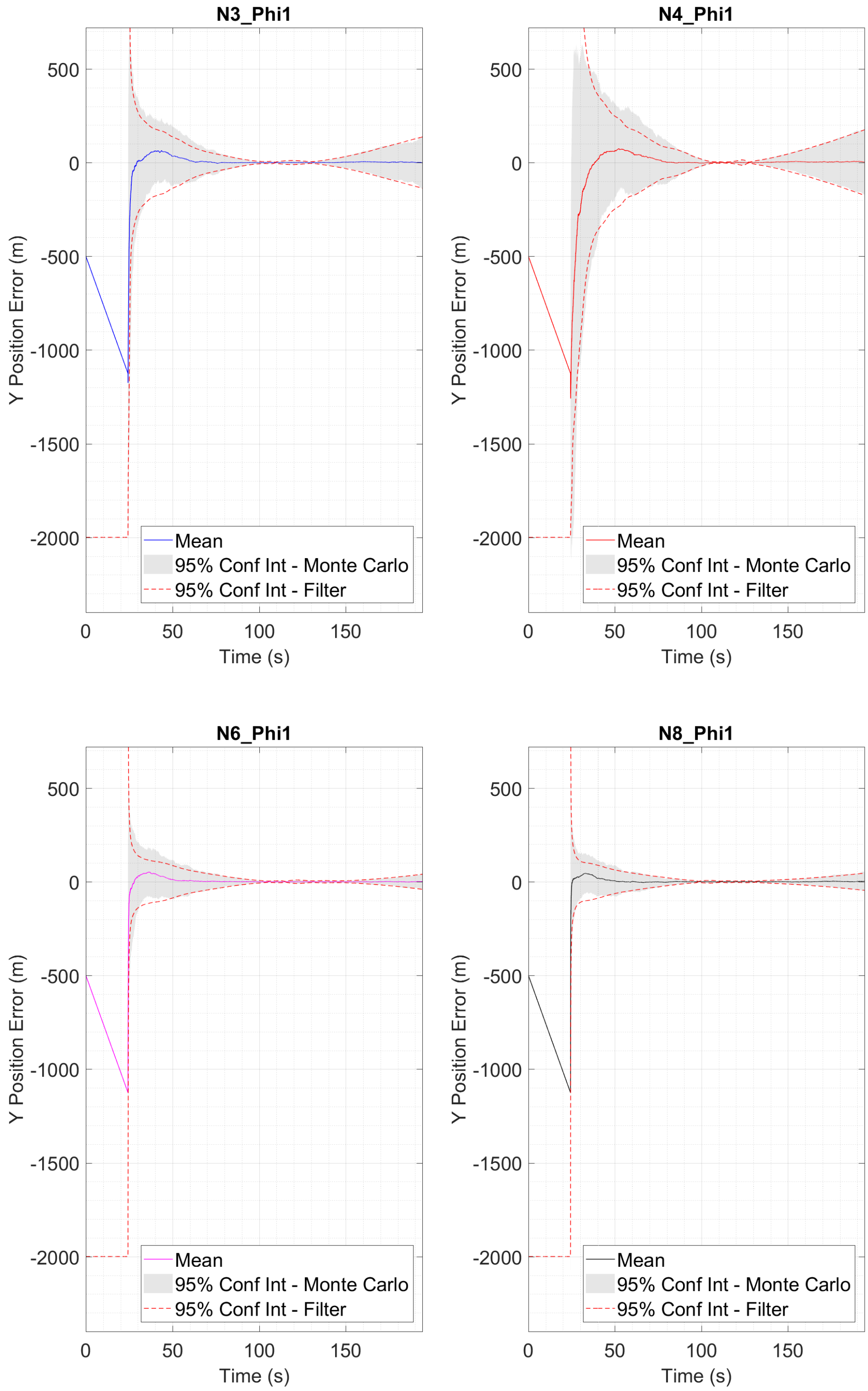

Figure 5.33: Φ_1 Genetic Algorithm Buoy Field Deployments - Y Position Error

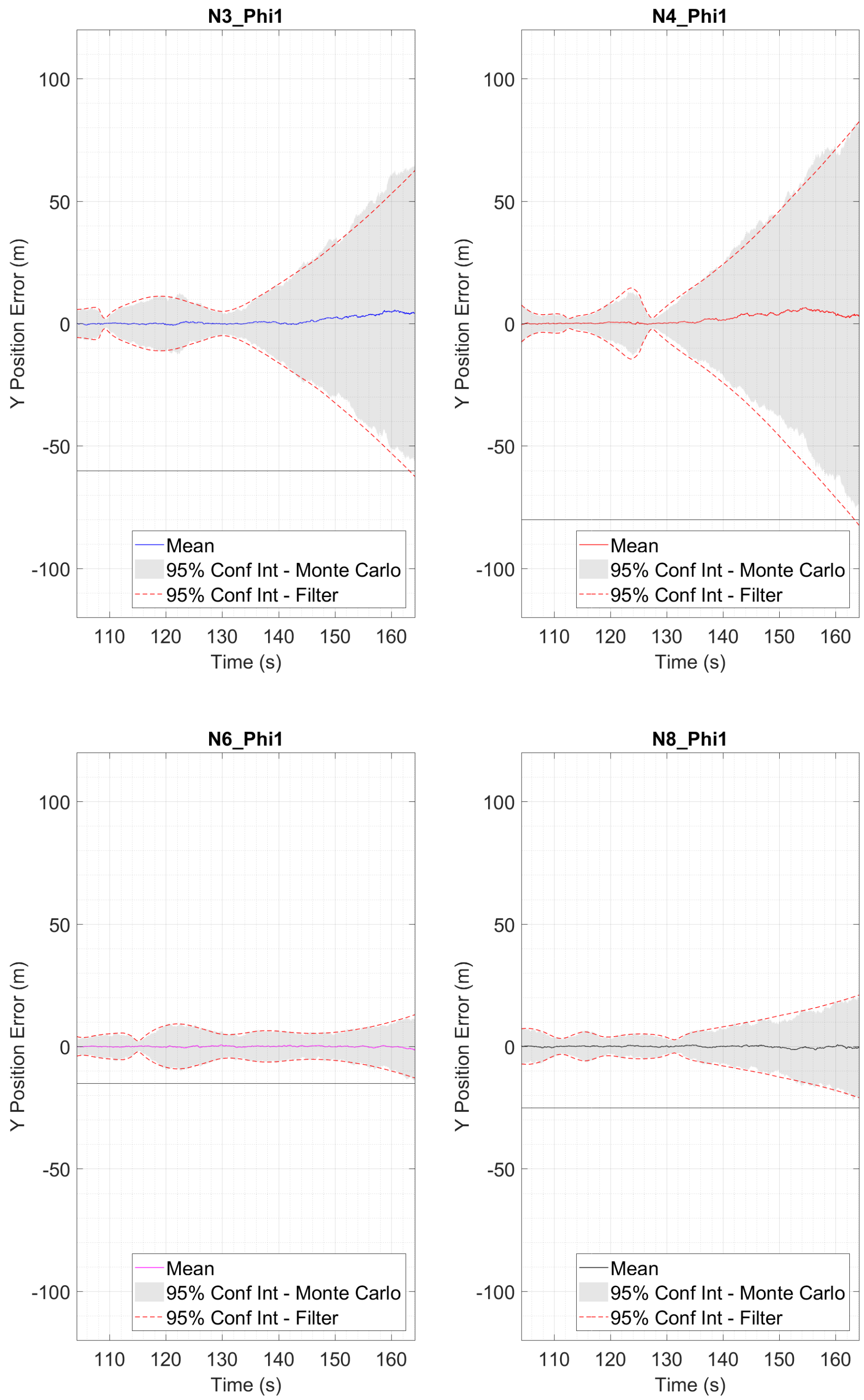

Figure 5.34: Φ_1 Genetic Algorithm Buoy Field Deployments - Track and Intercept Time Intervals - Y Position Error

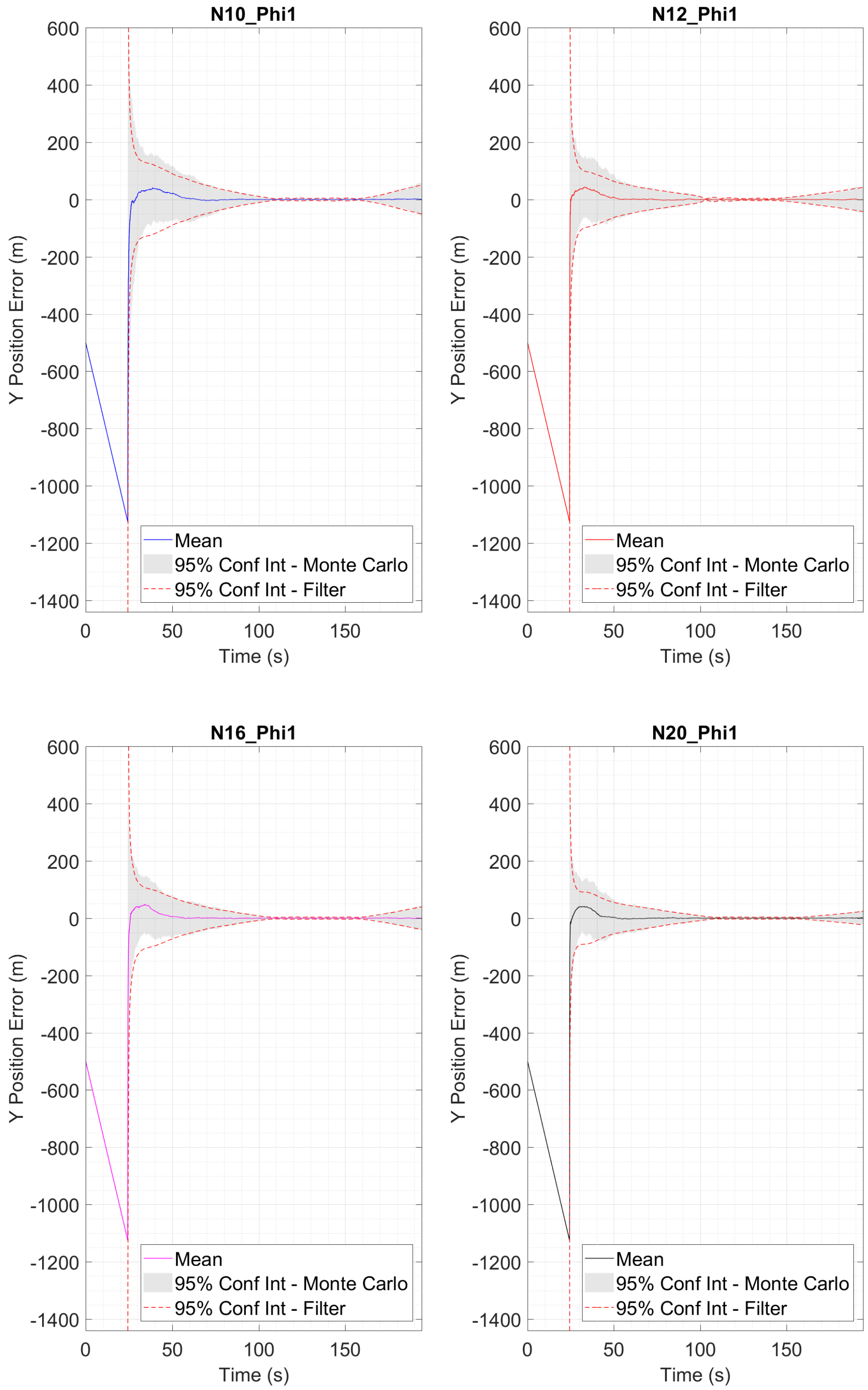

Figure 5.35: Φ_1 Genetic Algorithm Buoy Field Deployments - Y Position Error

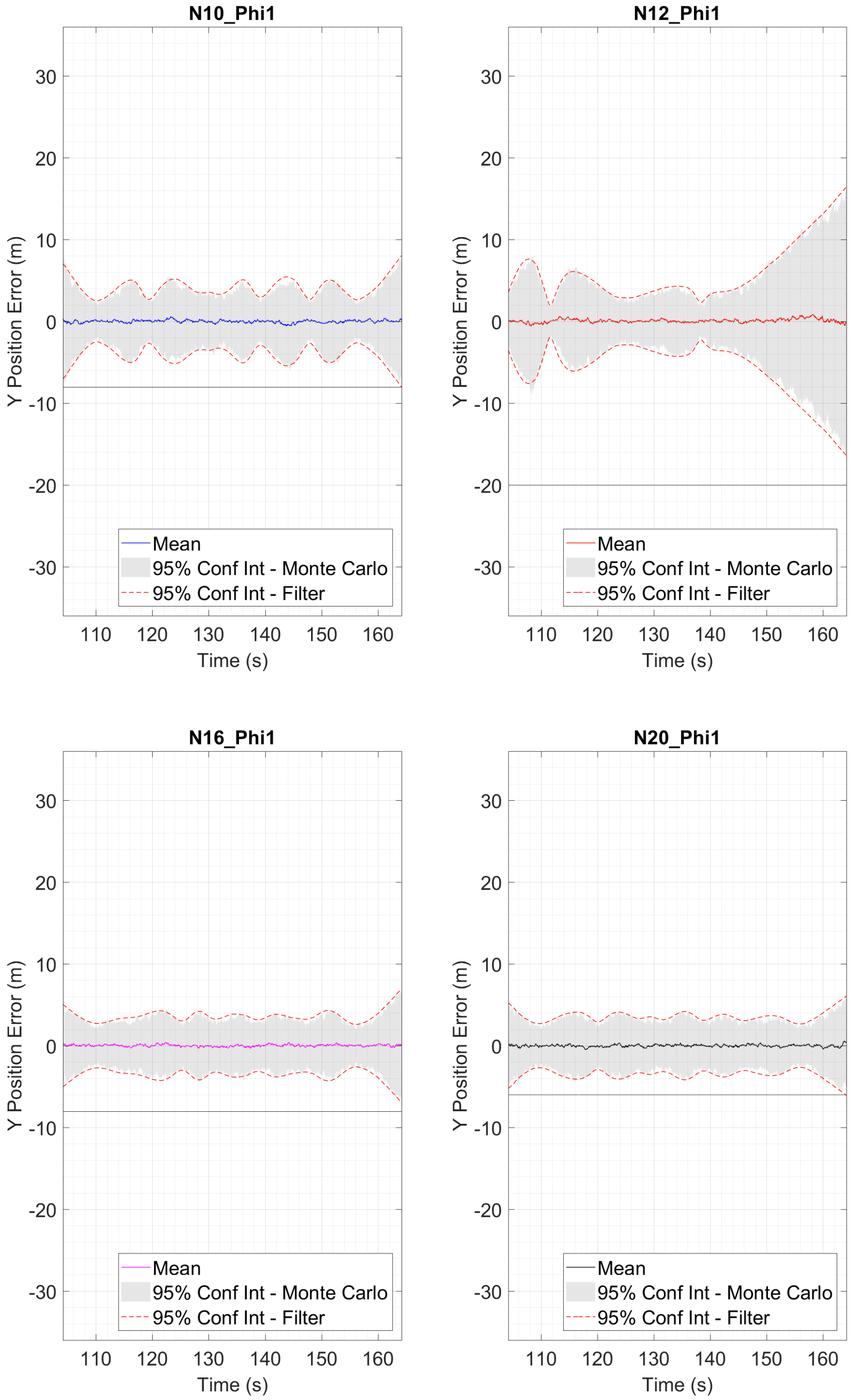

Figure 5.36: Φ_1 Genetic Algorithm Buoy Field Deployments - Track and Intercept Time Intervals - Y Position Error

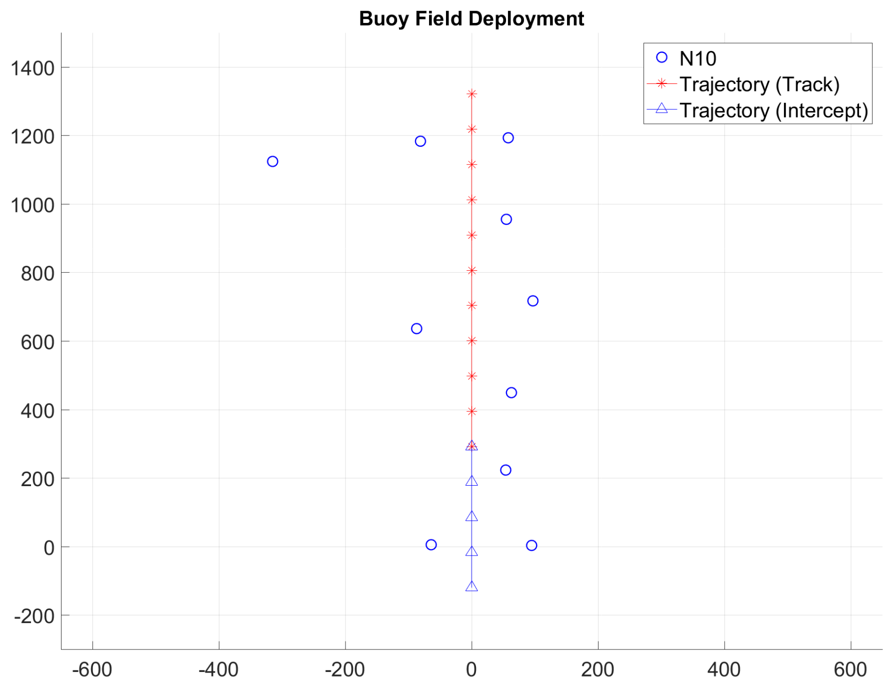

Figure 5.37: Φ_1 Genetic Algorithm Buoy Field Deployments - N=10

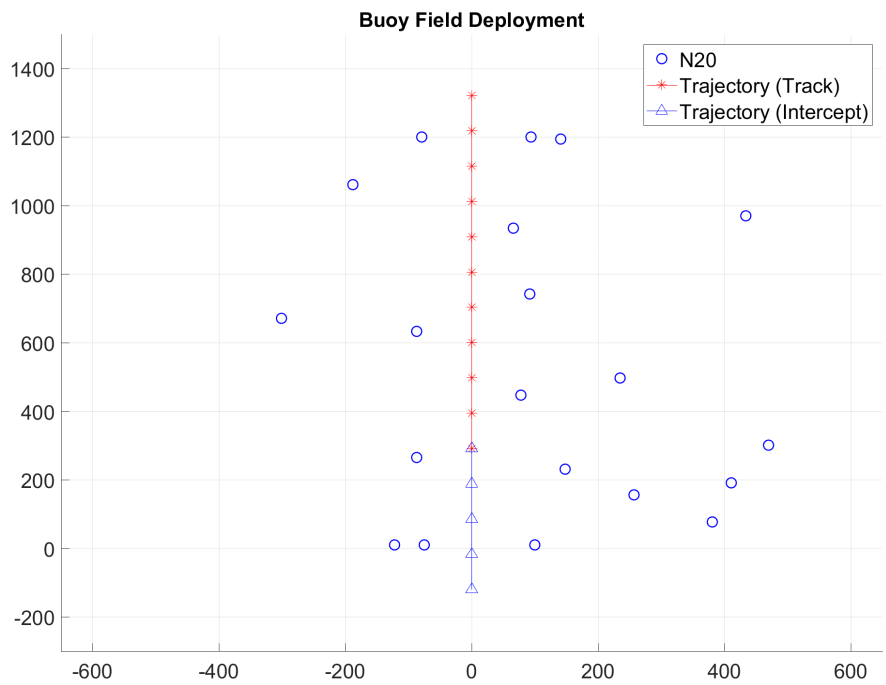

Figure 5.38: Φ_1 Genetic Algorithm Buoy Field Deployments - N=20

5.4.2 Optimisation using $\Phi_2(S(t_{s \ldots e}, B_{1 \ldots N}))$

Recalling the objective function,

$$\Phi_2(S(t_{s \ldots e}, B_{1 \ldots N})) = \sum_{t=t_{trk\_s}}^{t_{trk\_e}} \left(\det\left(P(t)\right) - D_{trk}\right)^2 + \sum_{t=t_{int\_s}}^{t_{int\_e}} \left(\det\left(P(t)\right) - D_{int}\right)^2$$

where

$$\begin{aligned} P &= \text{covariance matrix from simulation using buoy field } B_{1 \ldots N} \\ D_{trk} &= \text{desired determinant value per track time instant} \\ D_{int} &= \text{desired determinant value per intercept time instant} \end{aligned}$$

We executed the genetic algorithm to optimise a buoy field for N=3,4,6,8,10,12,16 and 20 against objective function Φ_2. Figure 5.39 shows the EKF's self perceived performance for the best candidate from each of the genetic algorithms for N=3,4,6 and 8. Figure 5.40 shows the EKF's self perceived performance for the best candidate from each of the genetic algorithms for N=10,12,16 and 20.

Figures 5.41 to 5.44 show the Y position error from the Monte Carlo simulation for each of the buoy field configurations. As before, overlayed on these figures are the EKF's self perceived accuracy which was found to be consistent with the actual performance calculated from the Monte Carlo simulation. We continue to observe over the acquisition time interval there is an initial error bias that converges towards zero as the target approaches the buoy field.

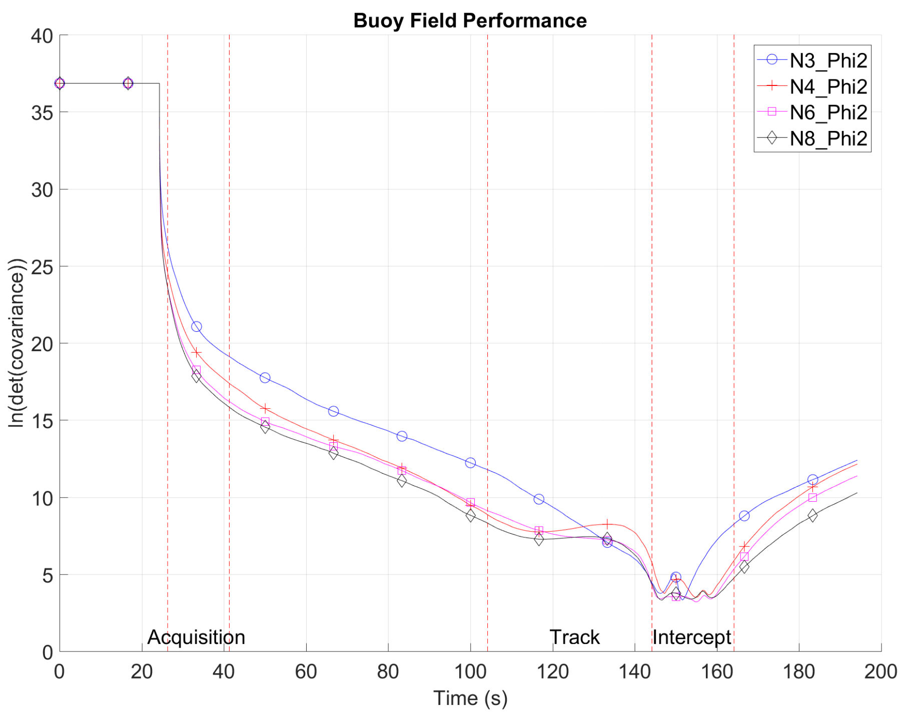

Figure 5.39: Φ_2 Genetic Algorithm Buoy Field Deployments - Covariance vs. Time

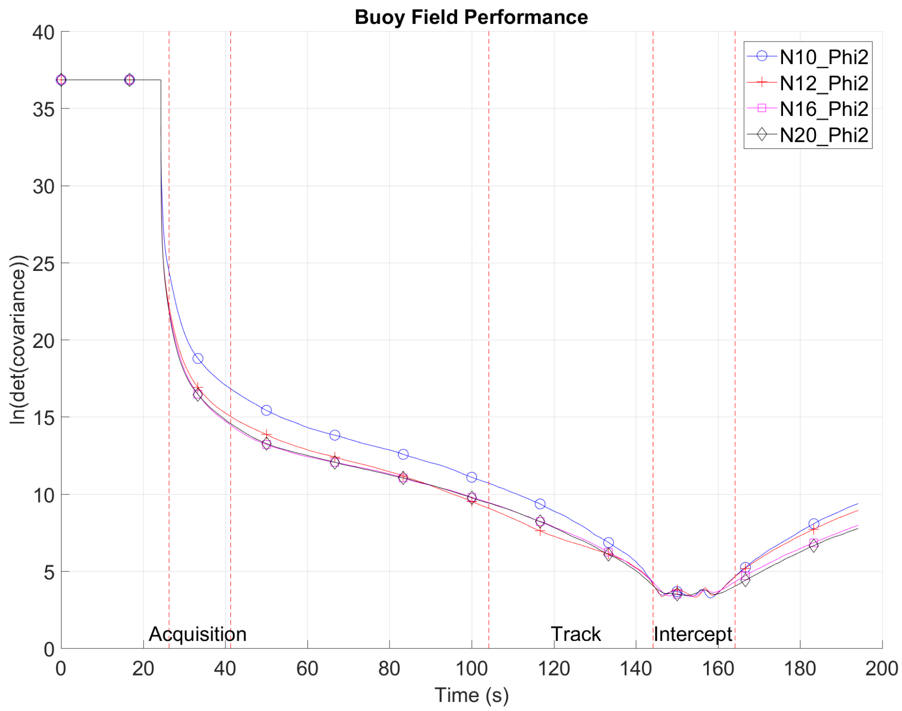

Figure 5.40: Φ_2 Genetic Algorithm Buoy Field Deployments - Covariance vs. Time

Reviewing the results we can assess that $N = 4$, $N = 6$ and $N = 8$ gave superior performance against the objective function Φ_2. $N \geq 10$ failed to provide any noticeable performance benefit over the track phase ($t_{trk\_s...trk\_e}$).

Figures 5.45, 5.46 and 5.47 show the buoy field deployments for $N = 4$, $N = 8$ and $N = 16$, respectively, from the genetic algorithm using the Φ_2 objective function.

Here we observe an interesting geometric pattern in the $N = 4$ and $N = 8$ deployments where we observed superior performance. The resulting deployments favours a single buoy deployed at extreme distance, with a cluster of buoys arranged around the ideal intercept point.

Reviewing Figure 5.47, we observe that the genetic algorithm has failed to make adequate use of additional buoys, likely due to a local minima through the initial benefits provided by a wide lateral spread of buoys. This resulted in an inability for the population to evolve in having a buoy outlier like we observer in $N = 4$ and $N = 8$, which would have provided similar benefits to the deployments for $N \geq 10$. Interestingly, these results demonstrate the utility of a single sensor at extreme distance providing overwatch of the target and superior angular diversity in the triangulation. Once again, target proximity to the sensor network plays an overriding factor in the performance of the estimator.

Adjustments to the genetic algorithm parameters to entice larger mutations in the deployment patterns, to attempt to ensure the possibility of convergence toward patterns such as $N4$ and $N8$, was not explored in this thesis. We also note that as $N > 10$, the genetic algorithm performance decreases leaving redundant sensor clustering, especially using the Φ_2 objective function.

Using the insight gained from the $N = 4$ through to $N = 8$ buoy field deployments, we've shifted one of the Buoys for the $N = 16$ case to an extreme distance from the sensor network (shown in Figure 5.48), providing overwatch of the incoming target. As expected, this has caused a significant improvement in performance over the track interval as shown in Figures 5.49.

We have seen the benefit of a single sensor providing overwatch of the target, greatly improving the overall performance of the state estimator by providing angular diversity and a more consistent angular rate of change over the target trajectory to facilitate triangulation. Continuing with this insight, we explored the effects of having two sensors provide overwatch of the target. In this instance we shifted one of the buoys that provided the least angular diversity in the $N = 6$ deployment. The resulting buoy fields are shown in Figure 5.50.

Reviewing the performance of the alternate $N = 6$ buoy field deployment as compared to the original, given in Figures 5.51, 5.52 and 5.53 we observe a vast improvement in the confidence on target position from the Acquisition time interval and all the way through the Track time interval. We also observe a small drop in confidence on target position over the intercept phase with the new buoy field configuration.

Reviewing the mean 95% confidence ellipses over the three key time intervals, given in Figures 5.54, 5.55 and 5.56, we also observe a reduction in eccentricity of the ellipse over the acquisition and track time intervals using this alternate deployment.

Using the Φ_2 cost function, the $N = 6$ alternate buoy field deployment (shown in Figure 5.50) was the preferred deployment as it provided the best performance for a tiered target tracking approach across the key time intervals with the least amount of buoys required.

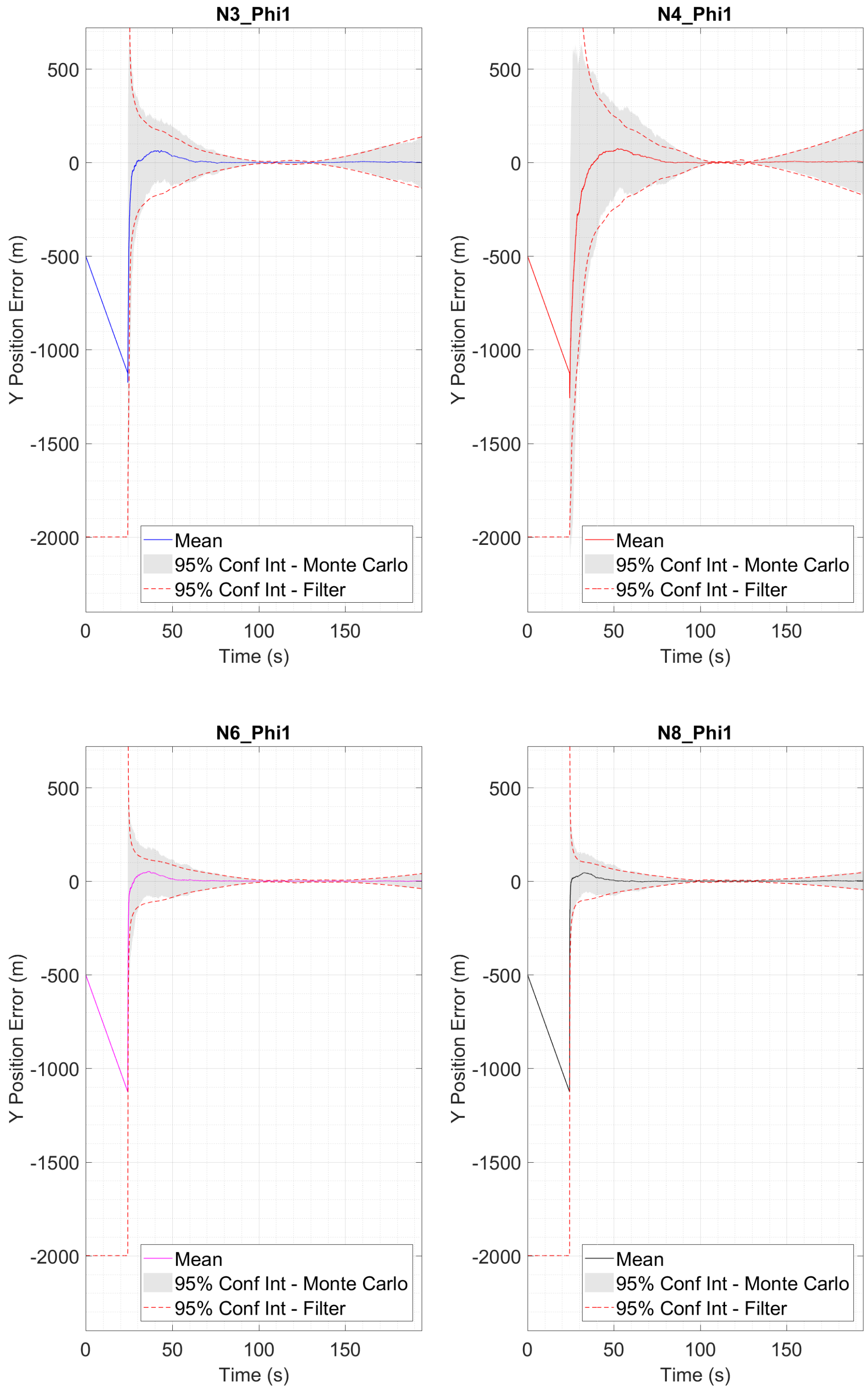

Figure 5.41: Φ_2 Genetic Algorithm Buoy Field Deployments - Y Position Error

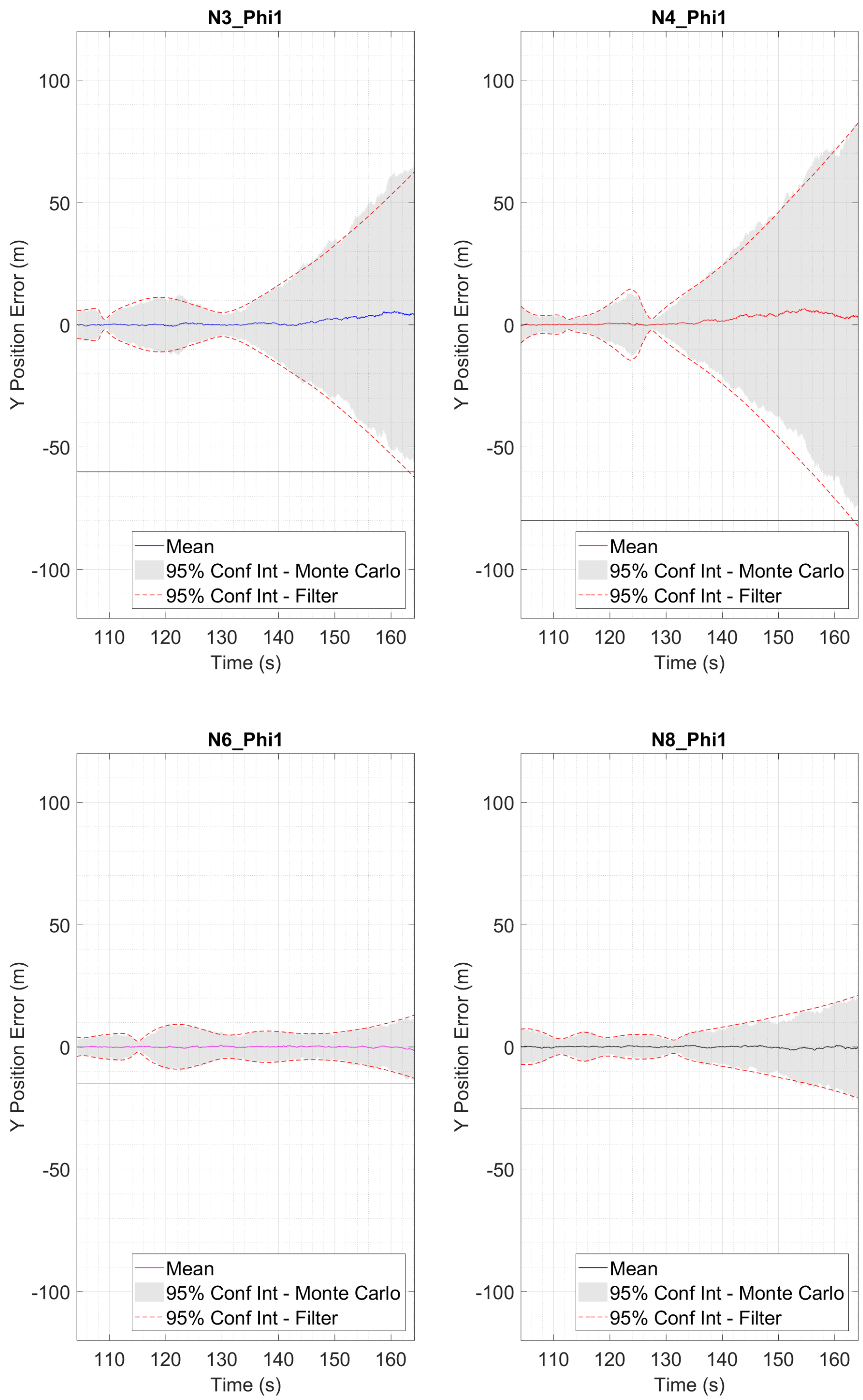

Figure 5.42: Φ_2 Genetic Algorithm Buoy Field Deployments - Track and Intercept Time Intervals - Y Position Error

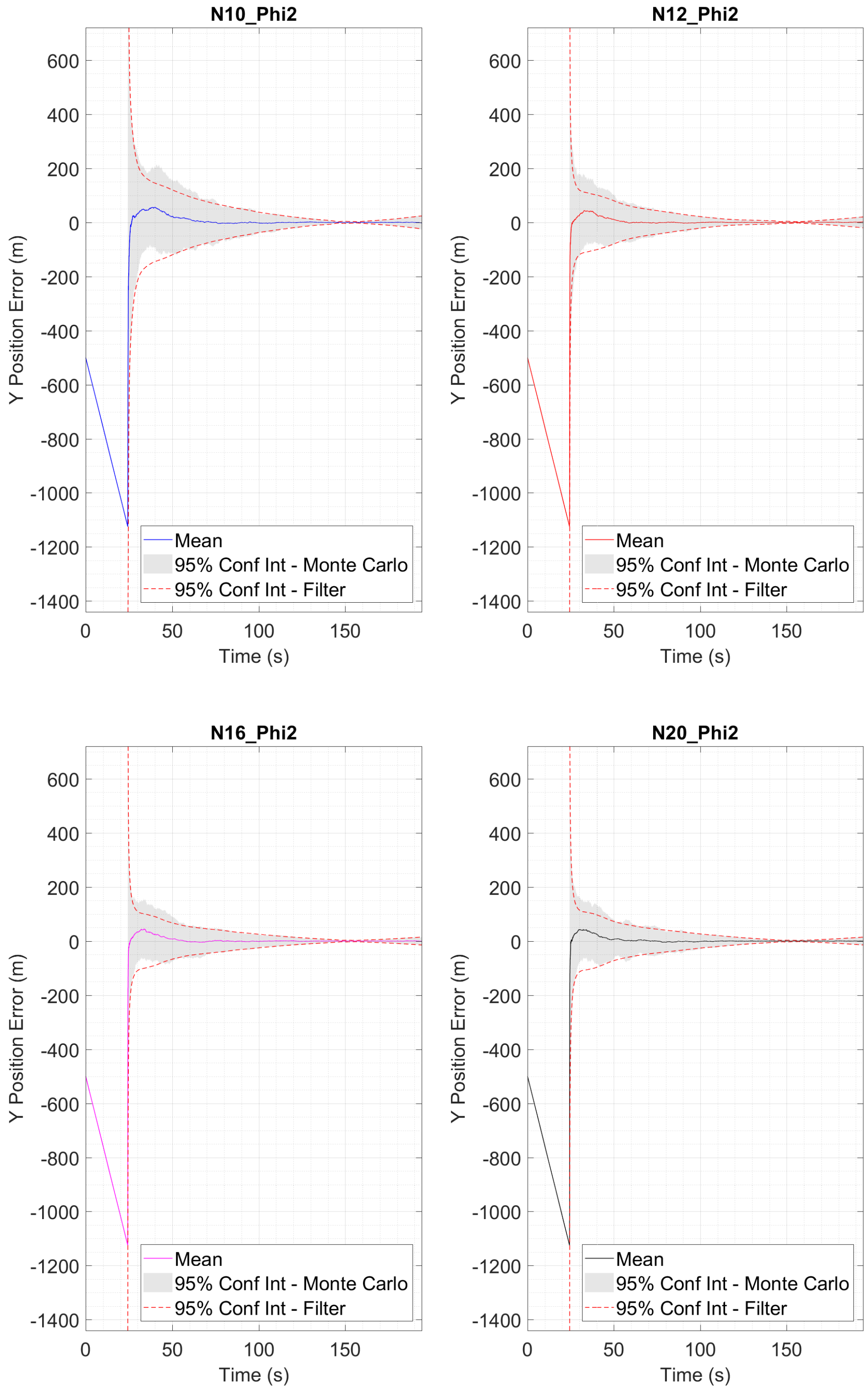

Figure 5.43: Φ_2 Genetic Algorithm Buoy Field Deployments - Y Position Error

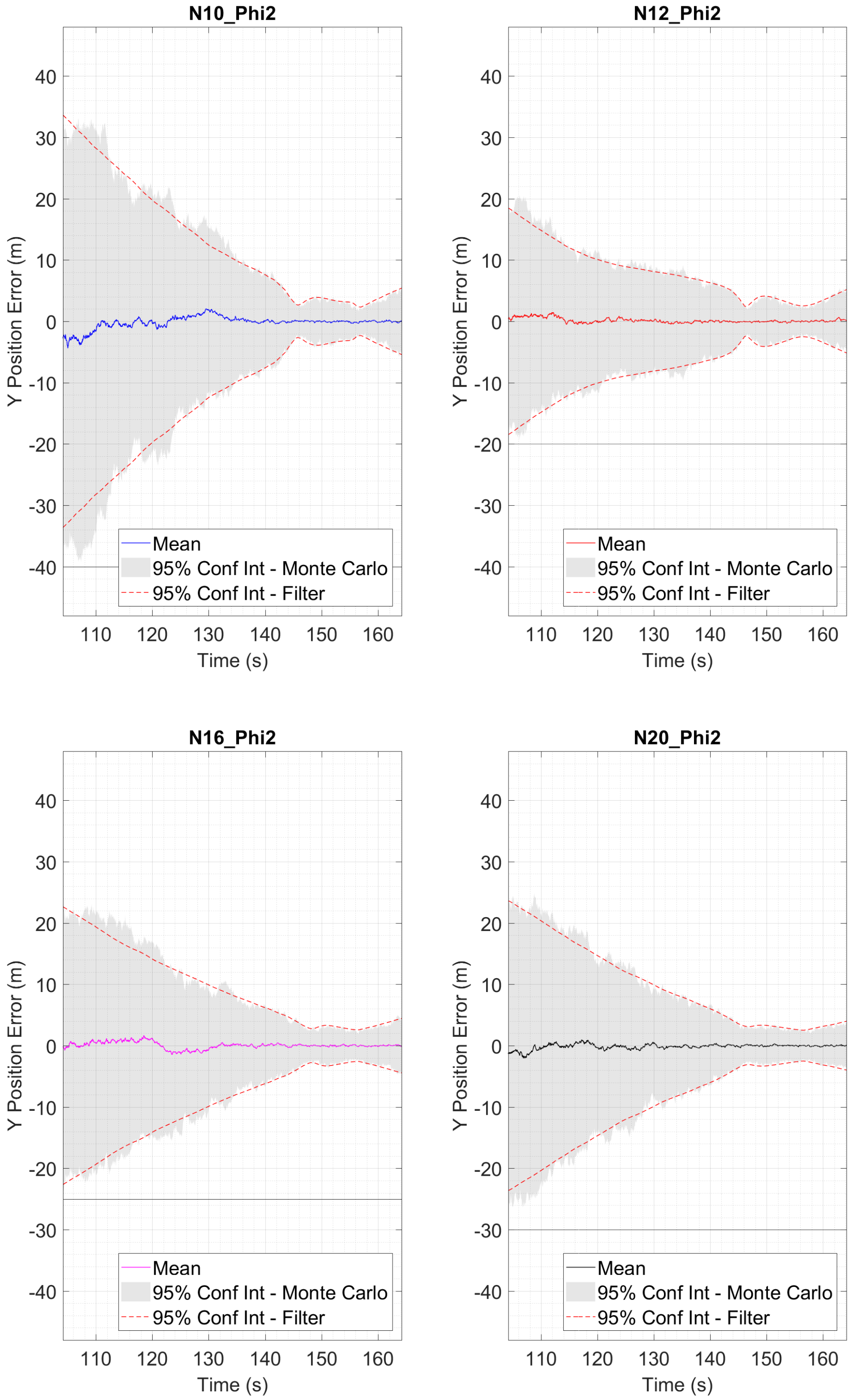

Figure 5.44: Φ_2 Genetic Algorithm Buoy Field Deployments - Track and Intercept Time Intervals - Y Position Error

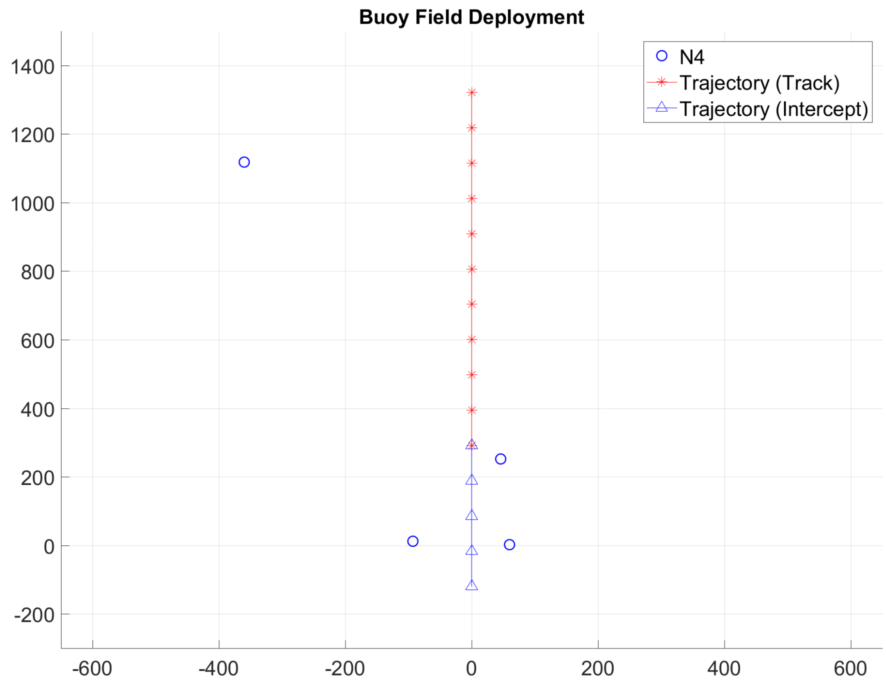

Figure 5.45: Φ_2 Genetic Algorithm Buoy Field Deployments - N=4

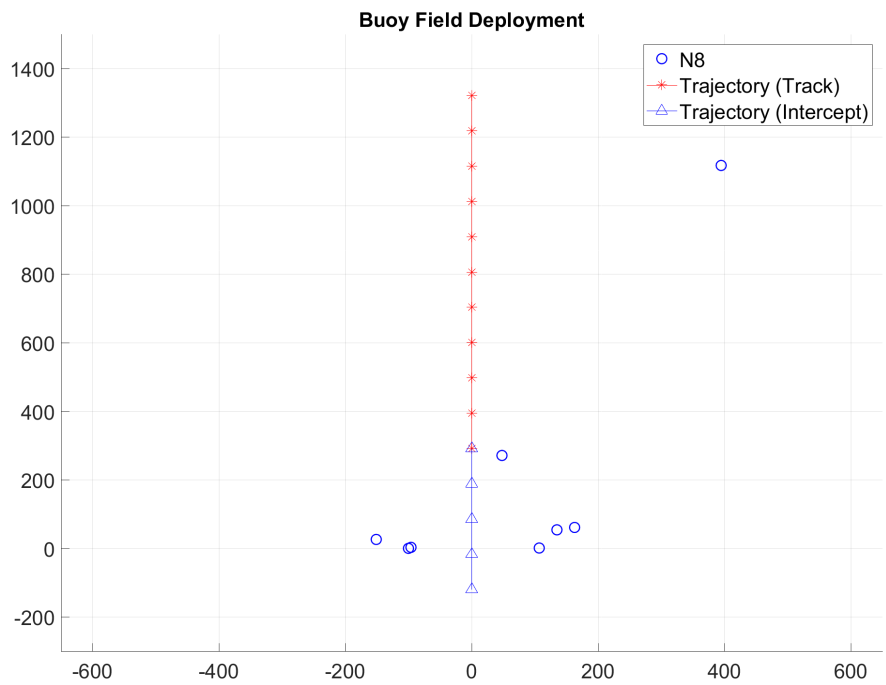

Figure 5.46: Φ_2 Genetic Algorithm Buoy Field Deployments - N=8

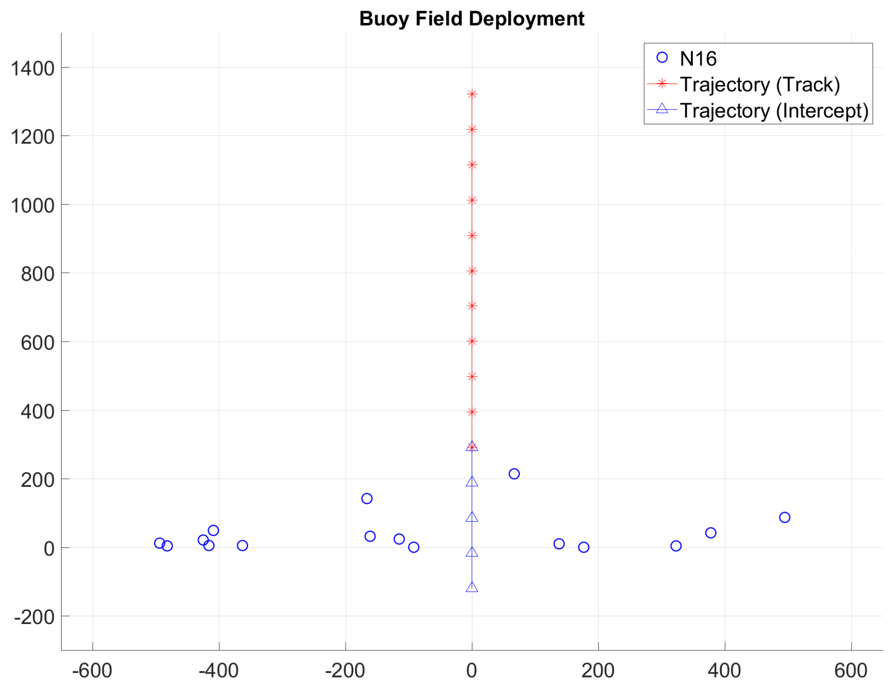

Figure 5.47: Φ_2 Genetic Algorithm Buoy Field Deployments - N=16

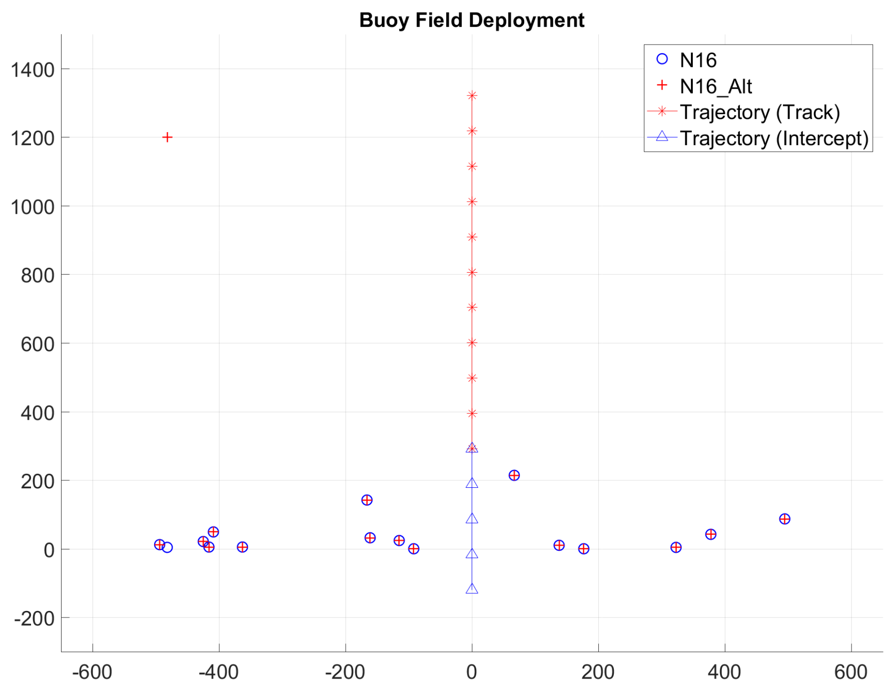

Figure 5.48: Φ_2 Genetic Algorithm Buoy Field Deployments - N=16 with shifted Buoy

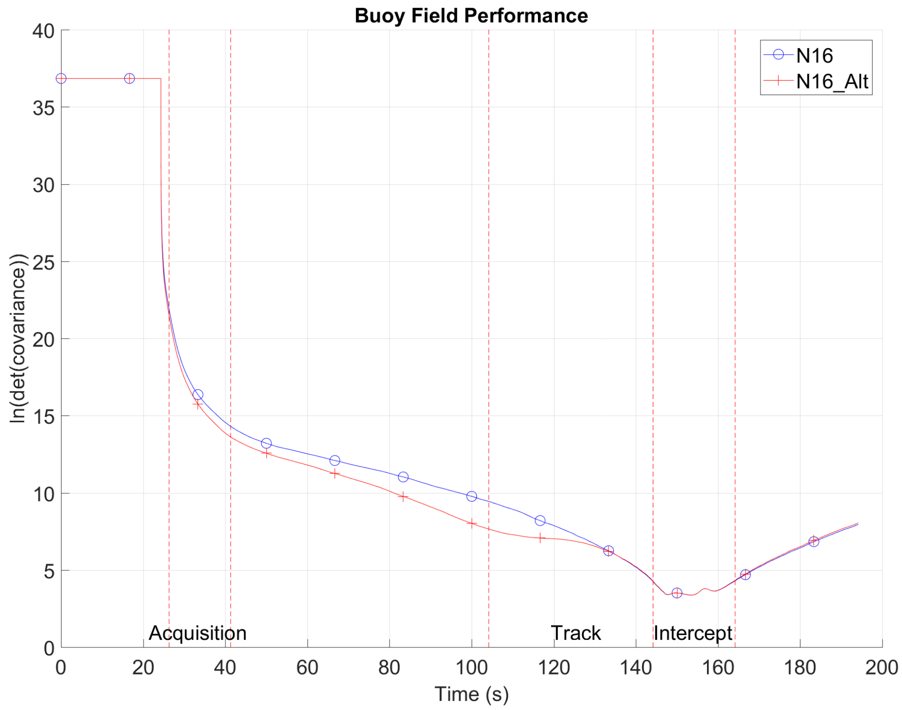

Figure 5.49: Φ_2 Genetic Algorithm Buoy Field Deployments - Covariance vs. Time - N=16 with shifted Buoy

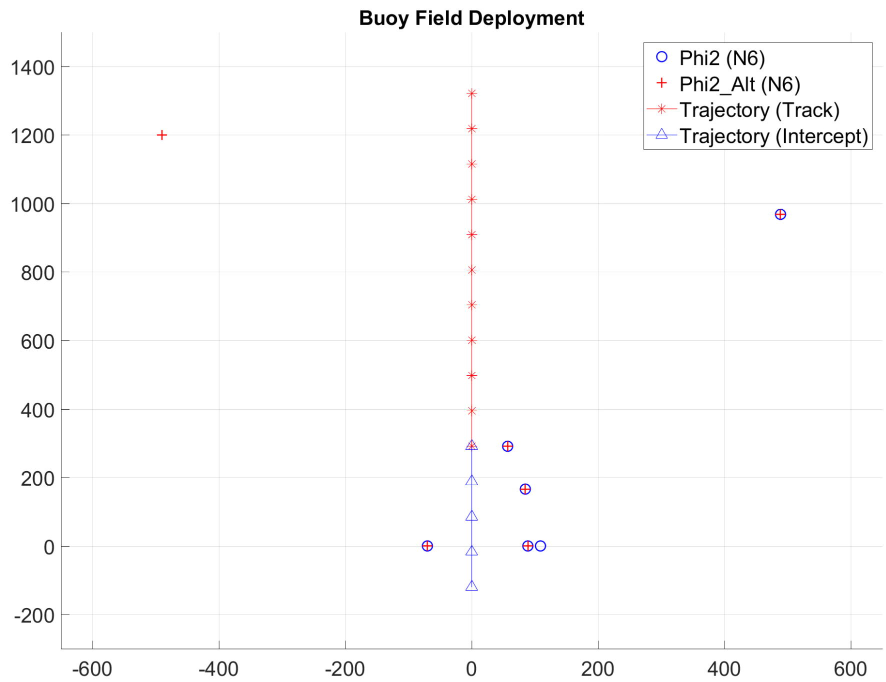

Figure 5.50: Φ_2 Genetic Algorithm Buoy Field Deployments - N=6 with shifted Buoy

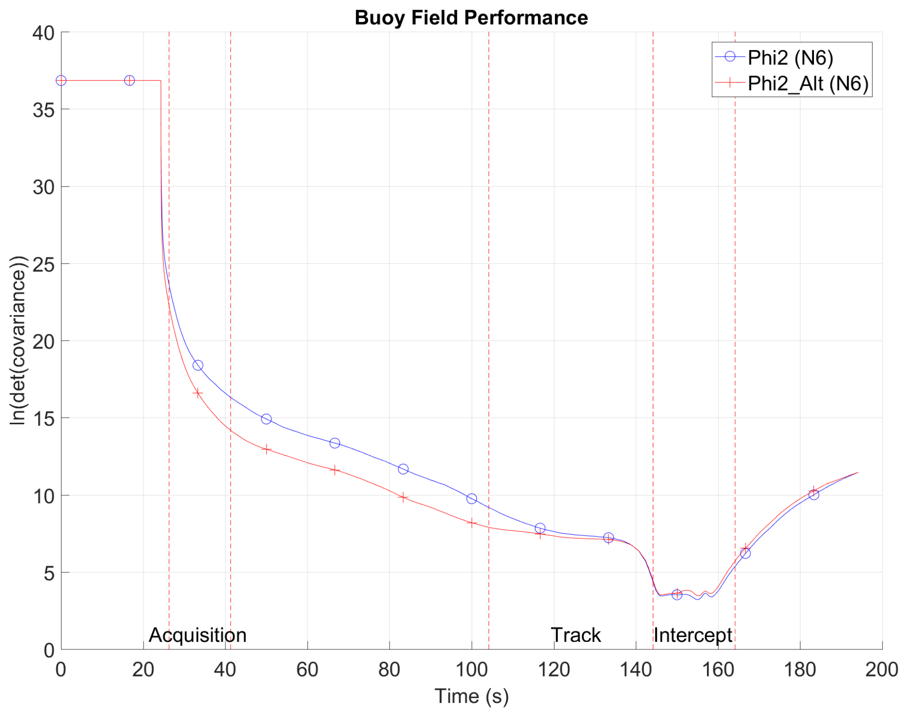

Figure 5.51: Φ_2 Genetic Algorithm Buoy Field Deployments - Covariance vs. Time - N=6 with shifted Buoy

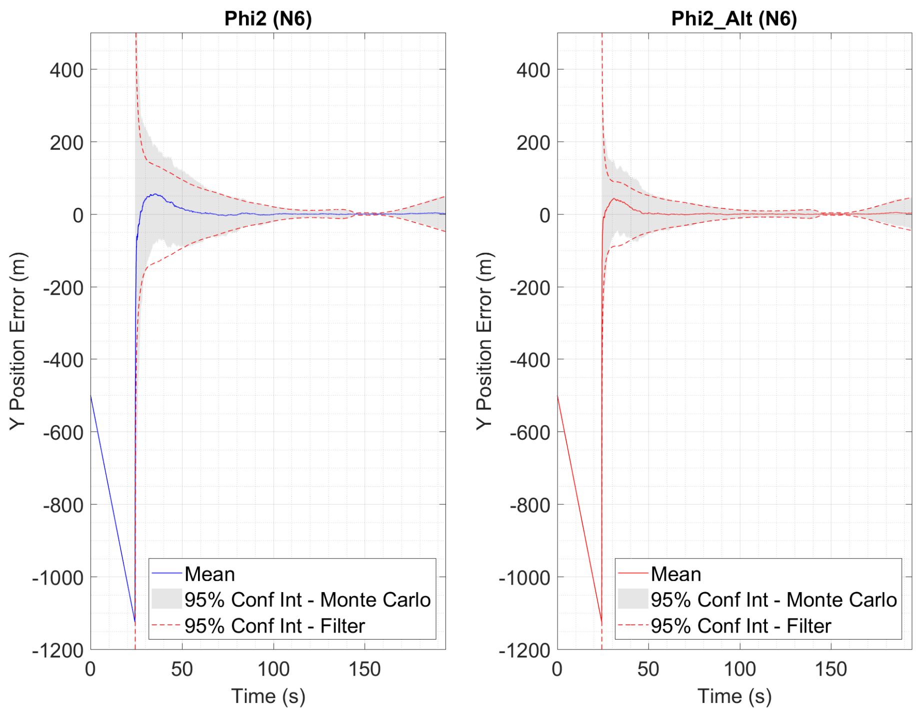

Figure 5.52: Φ_2 Genetic Algorithm Buoy Field Deployments - Y Position Error - N=6 with shifted Buoy

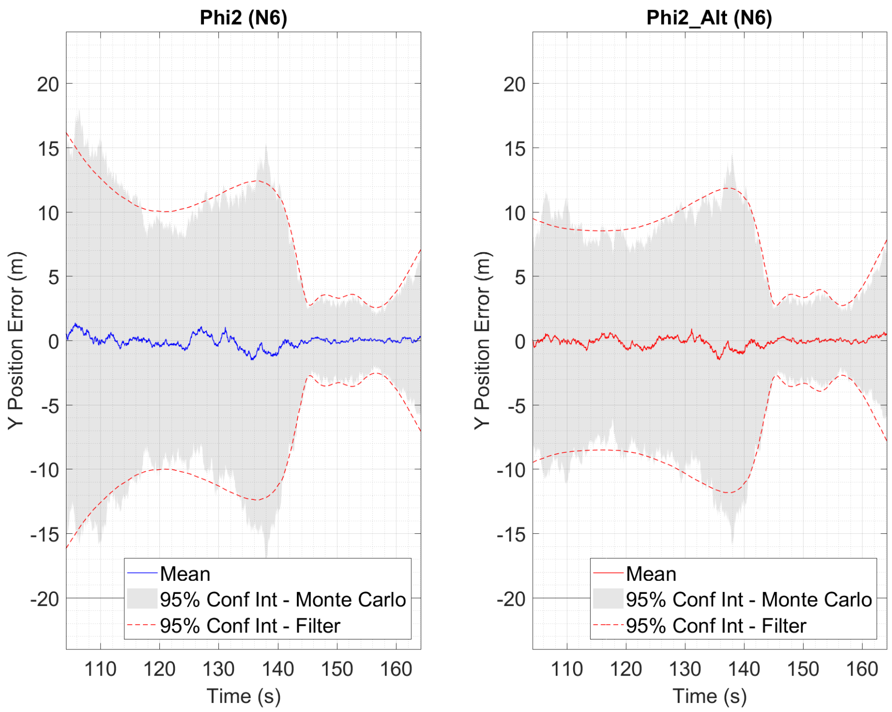

Figure 5.53: Φ_2 Genetic Algorithm Buoy Field Deployments - Track and Intercept Time Intervals - Y Position Error - N=6 with shifted Buoy

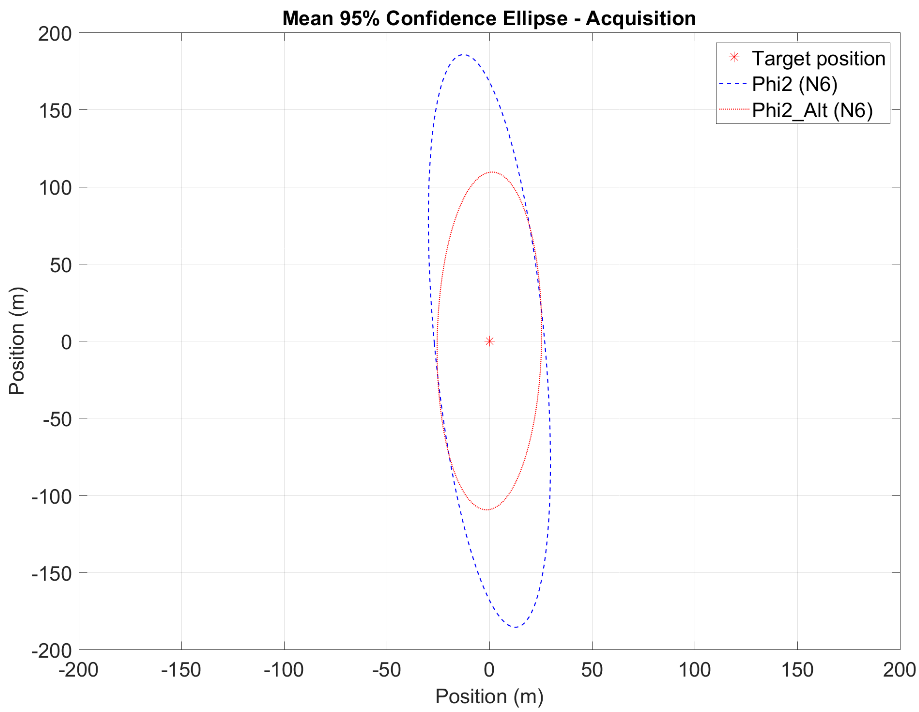

Figure 5.54: Φ_2 Genetic Algorithm Buoy Field Deployments - Mean 95% Confidence Ellipse - Acquisition Time Interval - N=6 with shifted Buoy

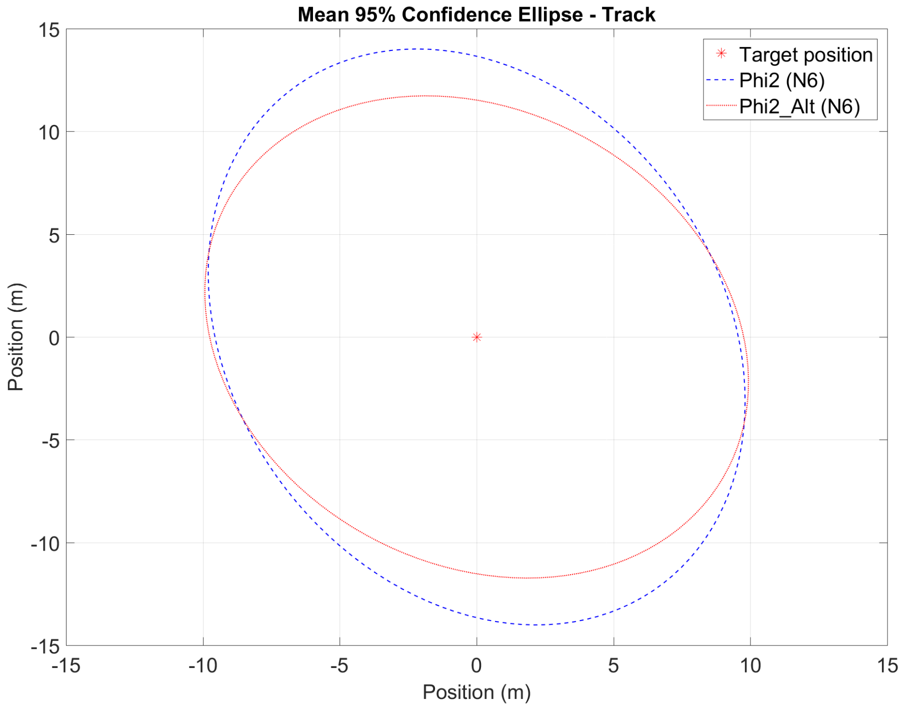

Figure 5.55: Φ_2 Genetic Algorithm Buoy Field Deployments - Mean 95% Confidence Ellipse - Track Time Interval - N=6 with shifted Buoy

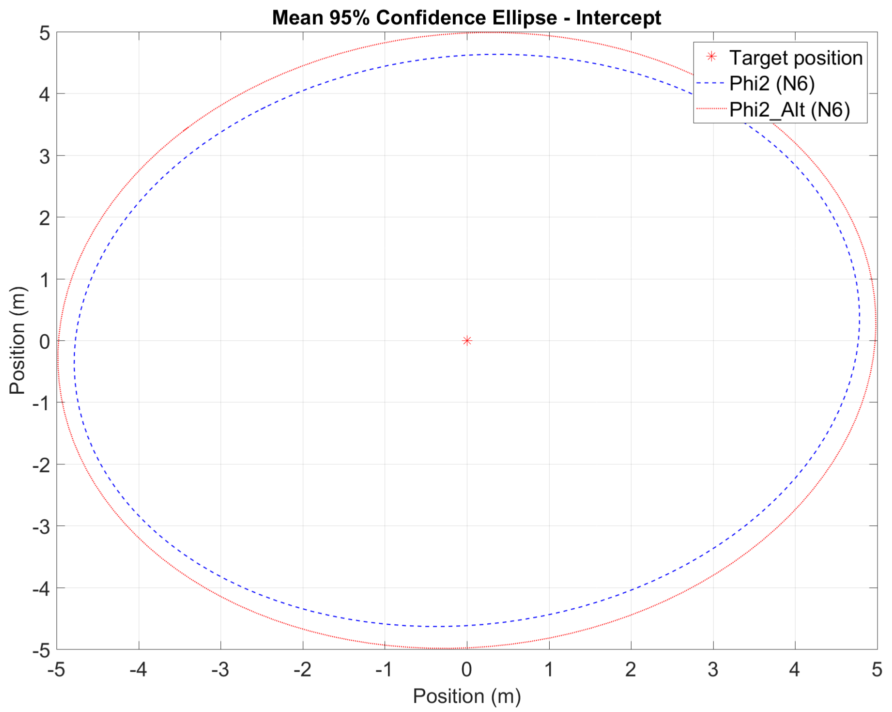

Figure 5.56: Φ_2 Genetic Algorithm Buoy Field Deployments - Mean 95% Confidence Ellipse - Intercept Time Interval - N=6 with shifted Buoy

5.5 Comparison

We now compare the performance of the best deployment candidate from each of the genetic algorithms against the typical buoy field deployments introduced in Section 5.2. We also provide commentary on the geometries that arose from the genetic algorithms and their performance against the two key time intervals Track ($t_{trk\_s...trk\_e}$) and Intercept ($t_{int\_s...int\_e}$).

5.5.1 Φ_1 Candidate vs. Typical Deployments

For the comparison between the candidate buoy field deployment resultant from the genetic algorithm using the Φ_1 objective function against the typical deployments given in Section 5.2, we've chosen the $N = 10$ buoy field. This deployment, from Section 5.4.1, provided the best tracking performance over the key time intervals using the least number of Buoys. For the comparison, a Monte Carlo simulation of 100 runs for each of the buoy field configurations was performed.

The buoy field deployments compared are given in Figure 5.57. We note that the structure of the Φ_1 buoy field deployment consists of column of buoys encompassing the target trajectory. There is also some lateral spread relative to the target trajectory towards the top of the of the buoy field.

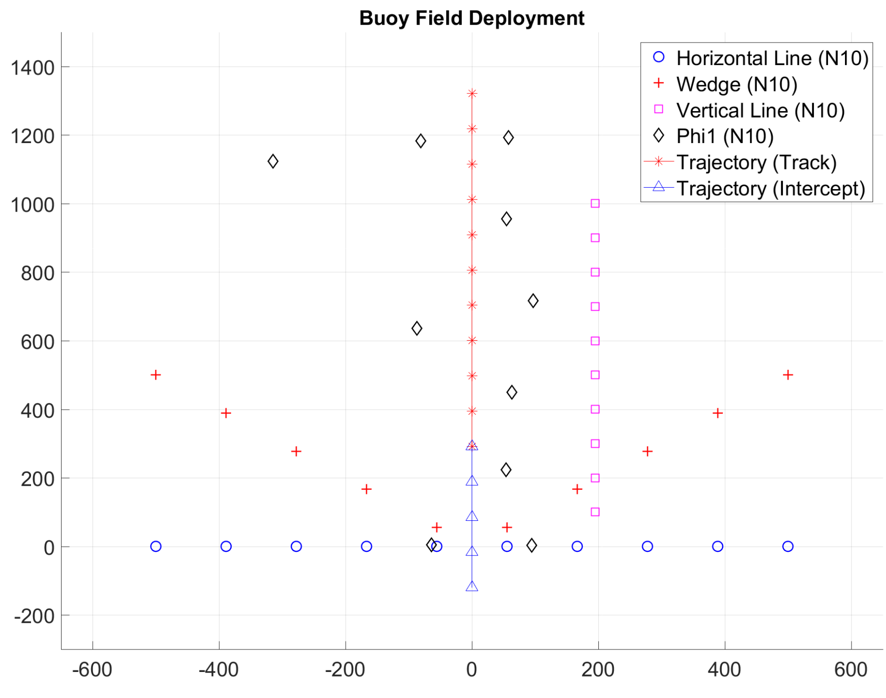

Figure 5.57: Φ_1 vs. Standard Buoy Field Deployments (N=10)

We observe that the state estimation position error, Figures 5.58, 5.59, 5.60 and 5.61, shows comparable and then superior performance for the Φ_1 deployment from $t = 65s$ onwards. Inferior performance prior to $t = 65s$ is a result of the buoy field having a poor lateral distribution relative to the target trajectory. As the target approaches, the distance between the buoy field and the target decreases and as a result the relative lateral distribution increases. As the target then passes through the buoy field, the column arrangement provides superior tracking of the target over the key time intervals Track ($t_{trk\_s...trk\_e}$) and Intercept ($t_{int\_s...int\_e}$) due to a maximisation of angular diversity and high rates of change in the angular measurements. This can also be observed in the covariance vs. time curve given in Figure 5.62.

We observe a spike in estimator performance for the wedge and horizontal line buoy fields at $t = 151.4s$ and $t = 155.8s$ respectively, but the Φ_1 buoy field deployment maintains a more overall consistent average of confidence on target position. The Φ_1 buoy field deployment has considerably superior performance over the track phase, as shown in Figure 5.61. We also observe the performance over the intercept phase being comparable to the wedge and horizontal line deployments.

Reviewing the mean 95% confidence ellipses over the three key time intervals from one of the simulation runs, given in Figures 5.63, 5.64 and 5.65, we can observe that the Φ_1 buoy field deployment provides reduced eccentricity in the 95% confidence ellipse compared with the other deployments.

Overall, it was shown that the $N = 10$ deployment resultant from the Φ_1 genetic algorithm optimisation outperformed the typical deployments introduced in Section 5.2. The $N = 10$ Φ_1 deployment provided a consistently lower target tracking uncertainty over the key time intervals, Track ($t_{trk\_s...trk\_e}$) and Intercept ($t_{int\_s...int\_e}$).

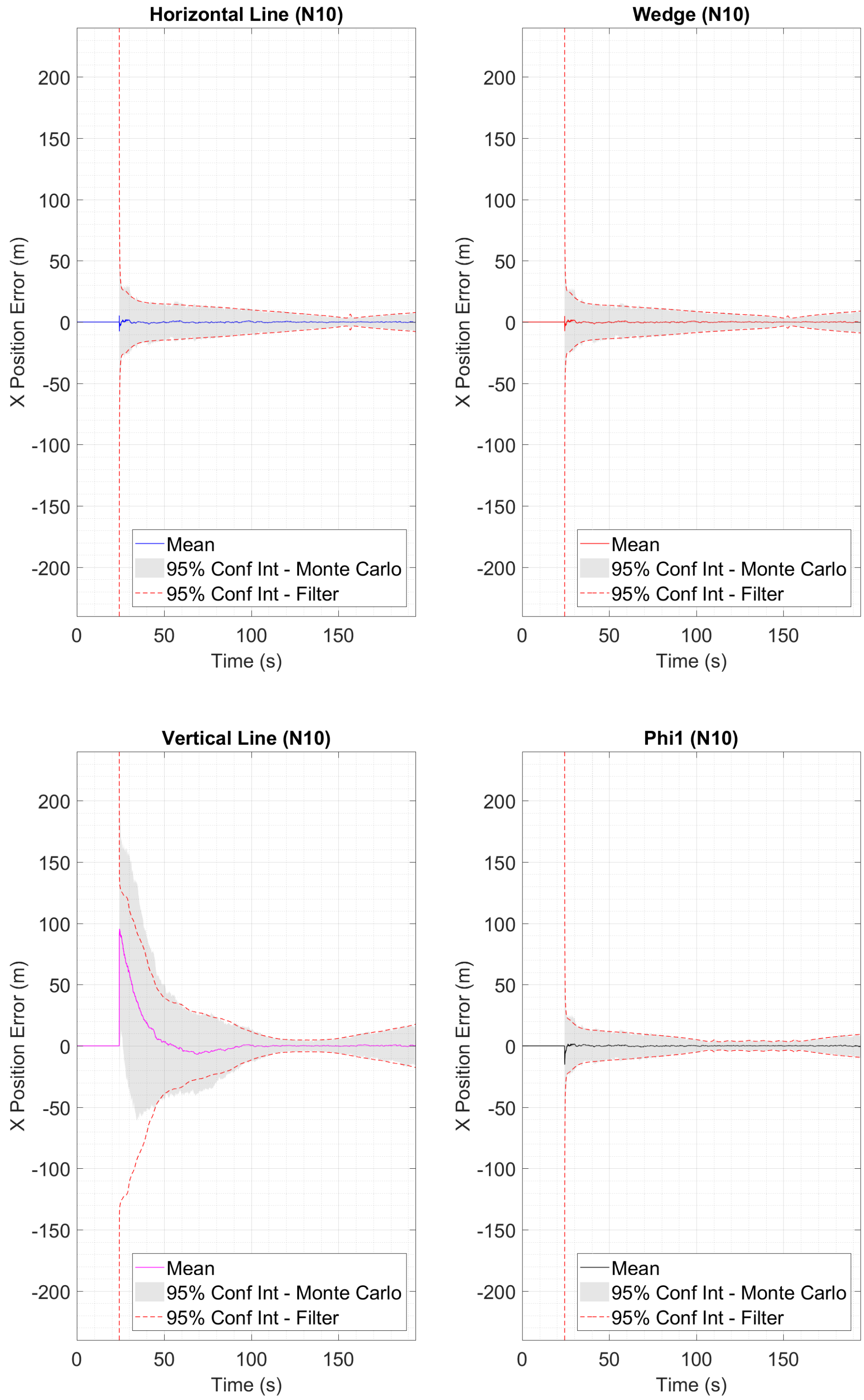

Figure 5.58: Φ_1 vs. Standard Buoy Field Deployments (N=10) - X Position Error

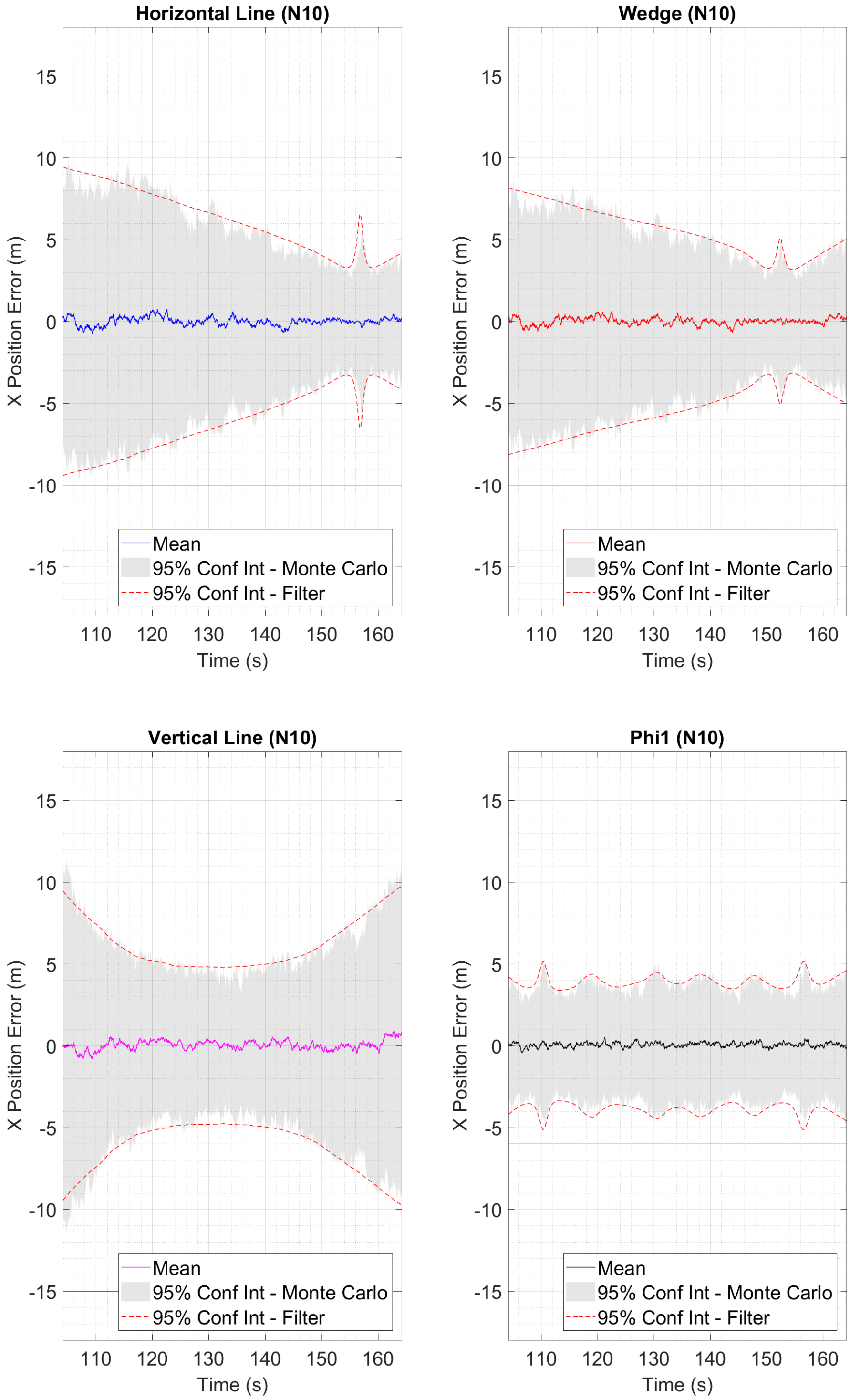

Figure 5.59: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Track and Intercept Time Intervals - X Position Error

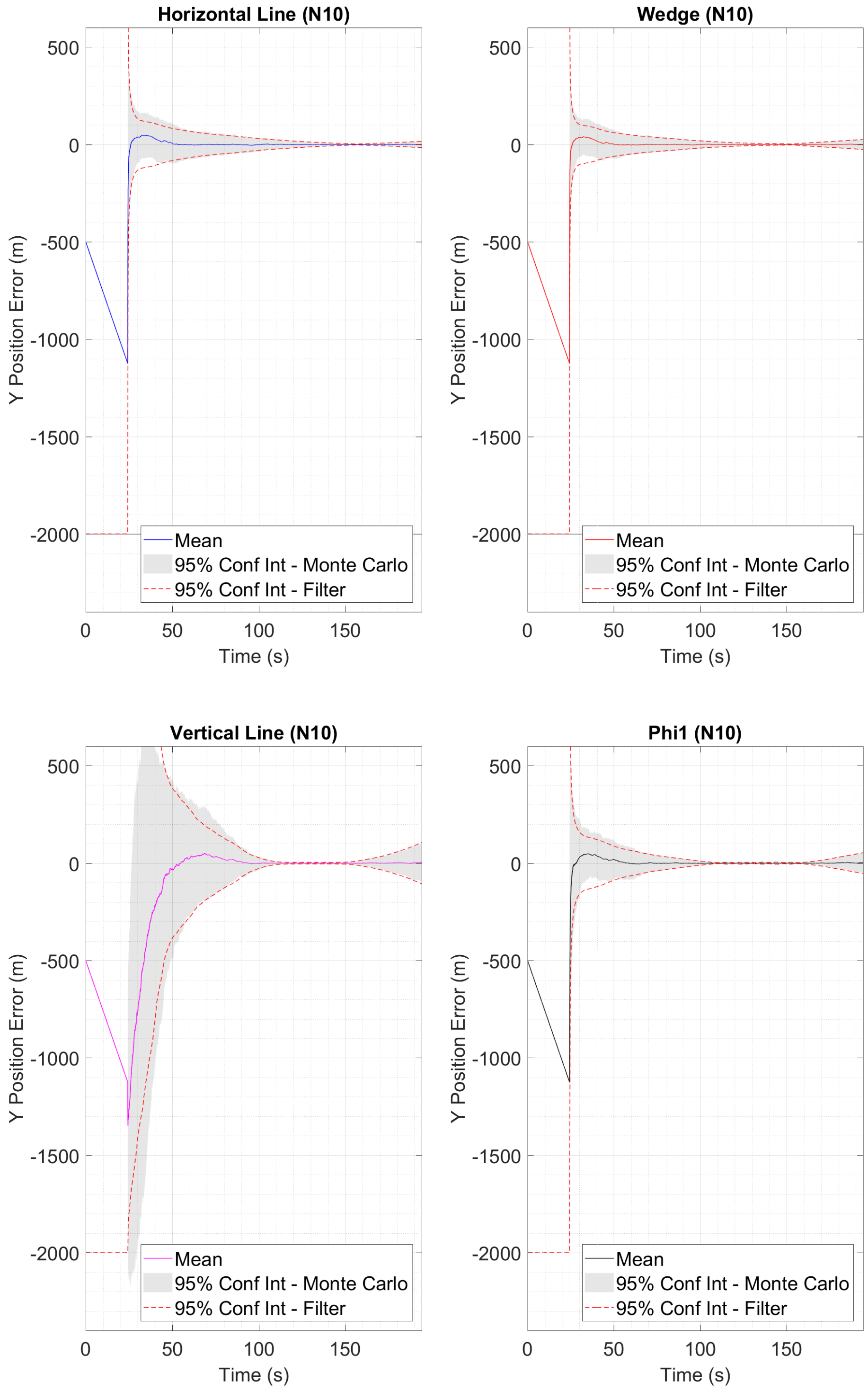

Figure 5.60: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Y Position Error

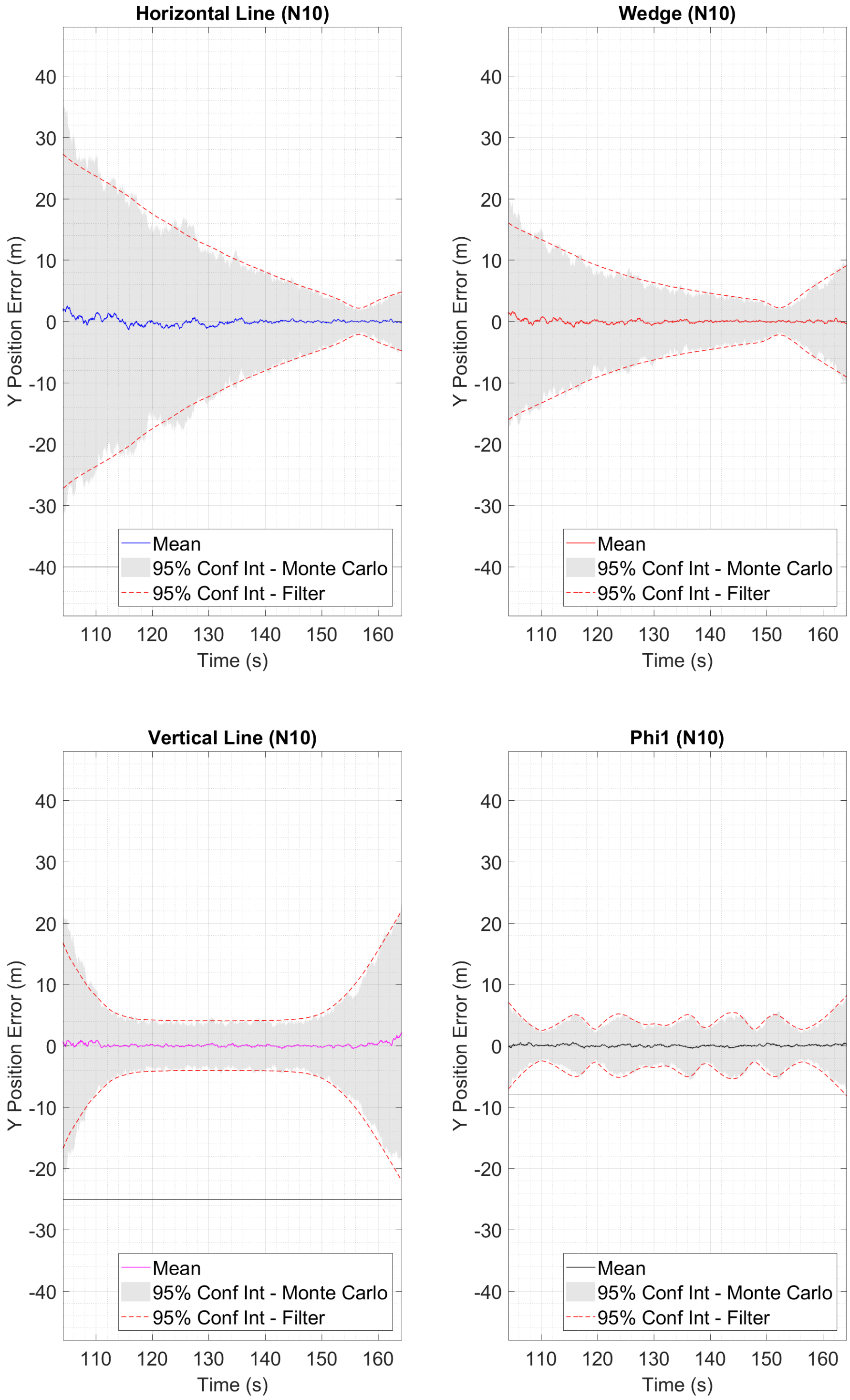

Figure 5.61: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Track and Intercept Time Intervals - Y Position Error

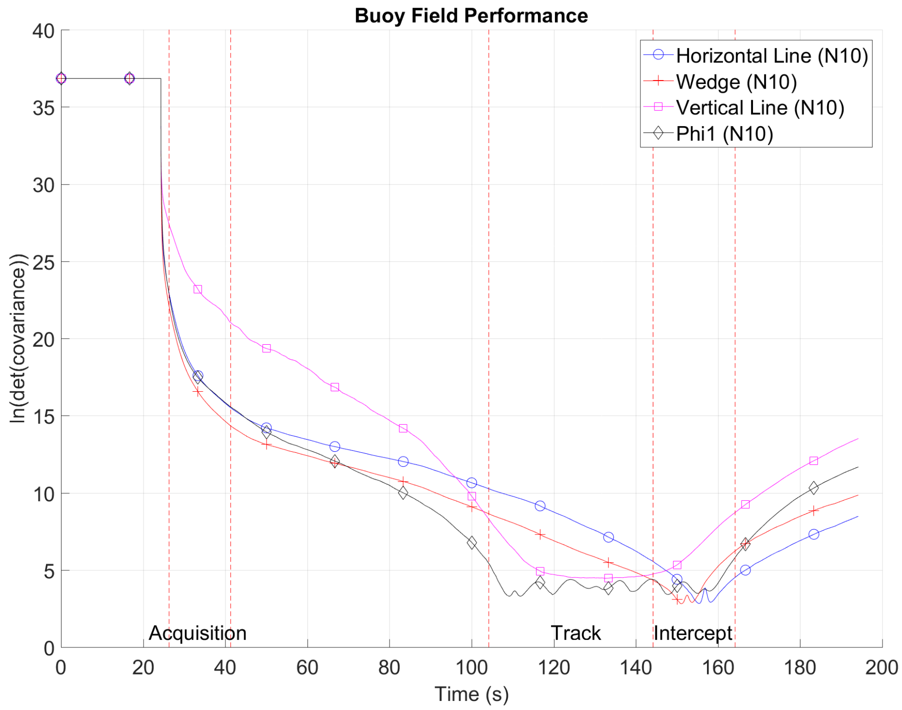

Figure 5.62: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Covariance vs. Time

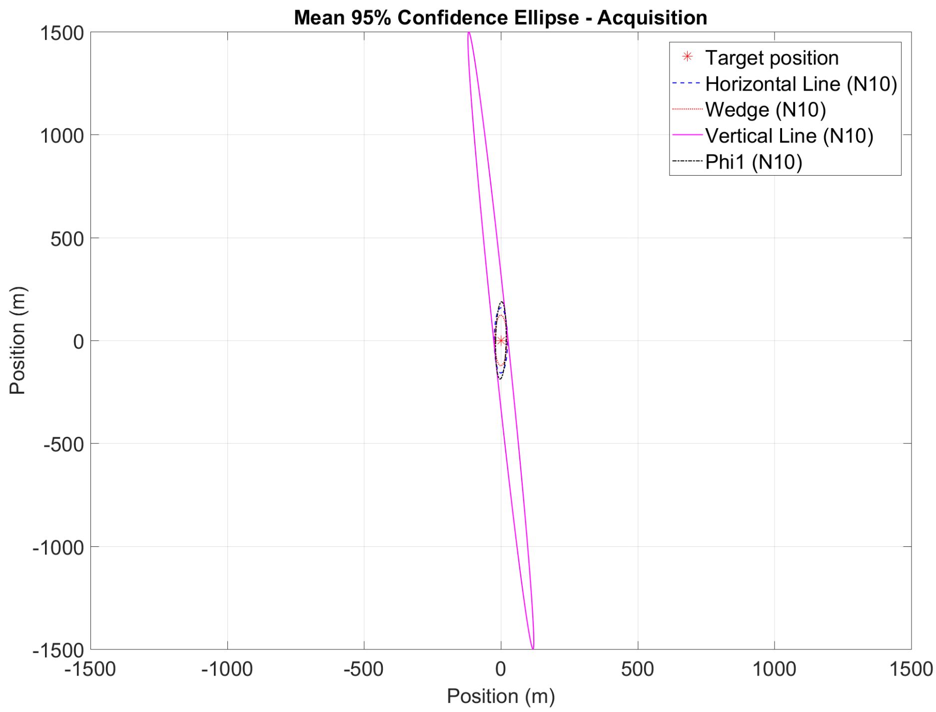

Figure 5.63: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Mean 95% Confidence Ellipse - Acquisition Time Interval

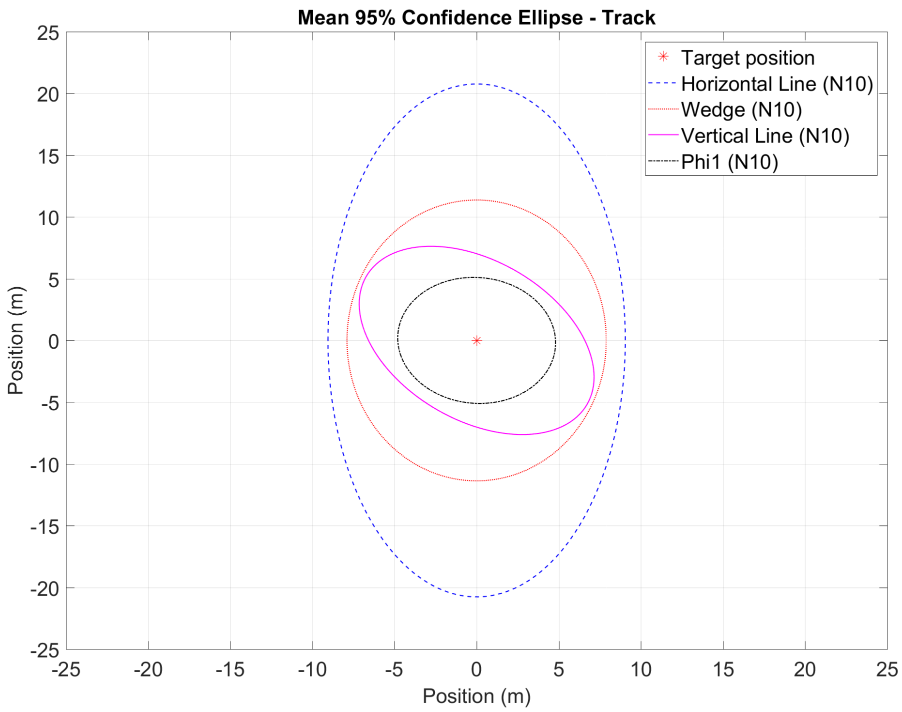

Figure 5.64: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Mean 95% Confidence Ellipse - Track Time Interval

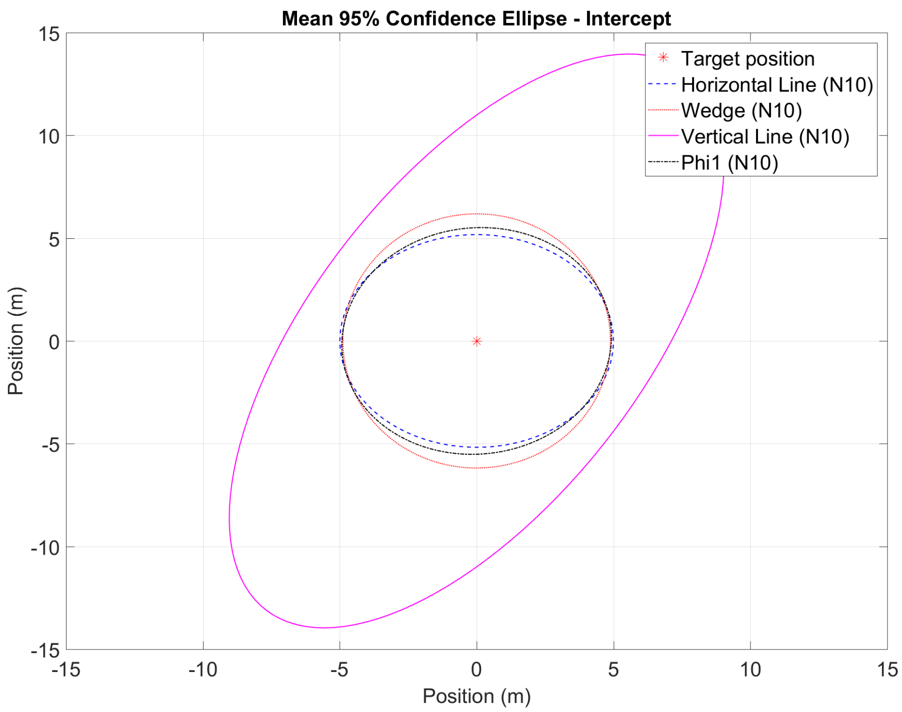

Figure 5.65: Φ_1 vs. Standard Buoy Field Deployments (N=10) - Mean 95% Confidence Ellipse - Intercept Time Interval

5.5.2 Φ_2 Candidate vs. Typical Deployments

For the comparison between the candidate buoy field deployment resultant from the genetic algorithm using the Φ_2 objective function against the typical deployments given in Section 5.2, we've chosen the alternate $N = 6$ buoy field deployment from Section 5.4.2. This deployment provided the best tracking performance over the key time intervals, in accordance with the objective function, using the least number of Buoys. For the comparison, a Monte Carlo simulation of 100 runs for each of the buoy field configurations was performed.

We note that the structure of the Φ_2 buoy field deployment consists of two buoys at extreme distance from the ideal intercept point providing overwatch of the target. The remaining buoys are then clustered around the ideal intercept point, contributing a degree of high confidence in the target position for intercept.

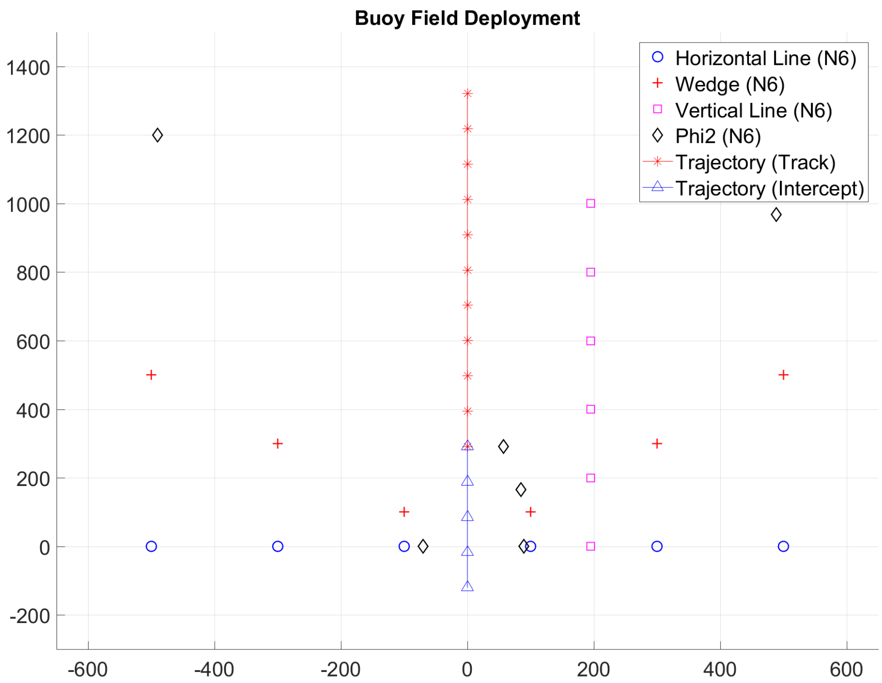

Figure 5.66: Φ_2 vs. Standard Buoy Field Deployments (N=6)

Reviewing the state estimation position error, given in Figures 5.67, 5.68, 5.69 and 5.70, we observe that the Φ_2 buoy field has comparable performance and converges to a lower estimation error at a similar rate as the wedge buoy field. We also observe a stepped reduction in the positional accuracy over the intercept time interval. These results are consistent with the EKF's self perceived performance given by the covariance vs. time curve from one of the simulated runs in Figure 5.71.

We now review the mean 95% confidence ellipses plots from one of the simulated runs for each of the key time intervals, given in Figures 5.72, 5.73 and 5.74. Here we observe the Φ_2 buoy field deployment provides reduced eccentricity in the confidence ellipse over the acquisition time interval, and has reduced eccentricity, comparable to the Wedge deployment, over the track and intercept time intervals.

Overall, it was shown that the alternate $N = 6$ deployment resultant from the Φ_2 genetic algorithm optimisation outperformed the typical deployments introduced in Section 5.2 against the desired target tracking performance. The Φ_2 deployment provided a balanced approach for target tracking, having a tiered approach for target tracking confidence over the key time intervals. It provided similar mean performance to the wedge deployment over the Track time interval ($t_{trk\_s \ldots trk\_e}$), and then significantly superior performance over the Intercept time interval ($t_{int\_s \ldots int\_e}$). This reduces the buoys required for target tracking, so long as a tiered approach to target tracking confidence is acceptable.

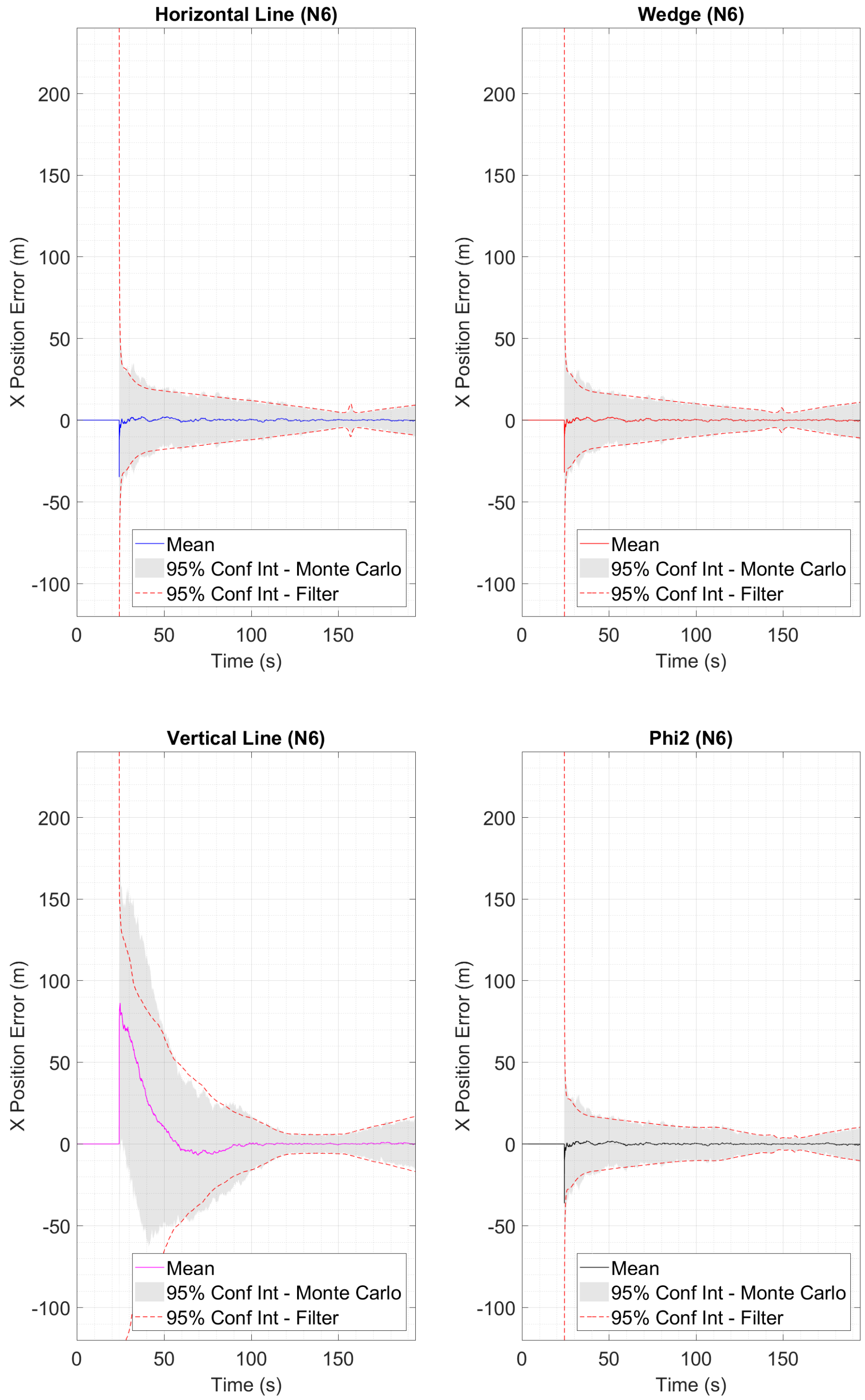

Figure 5.67: Φ_2 vs. Standard Buoy Field Deployments (N=6) - X Position Error

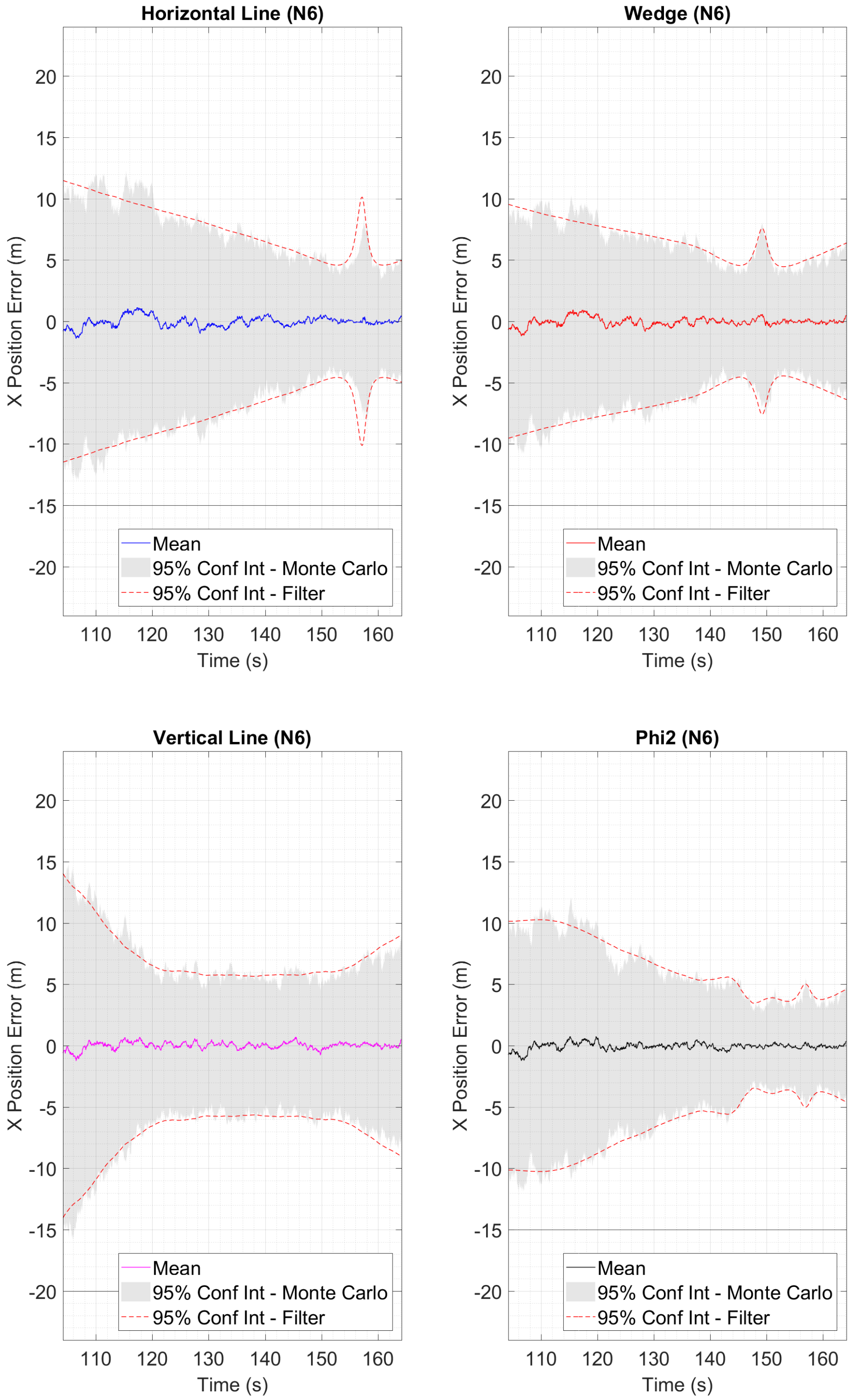

Figure 5.68: Φ_2 vs. Standard Buoy Field Deployments (N=6) - Track and Intercept Time Intervals - X Position Error

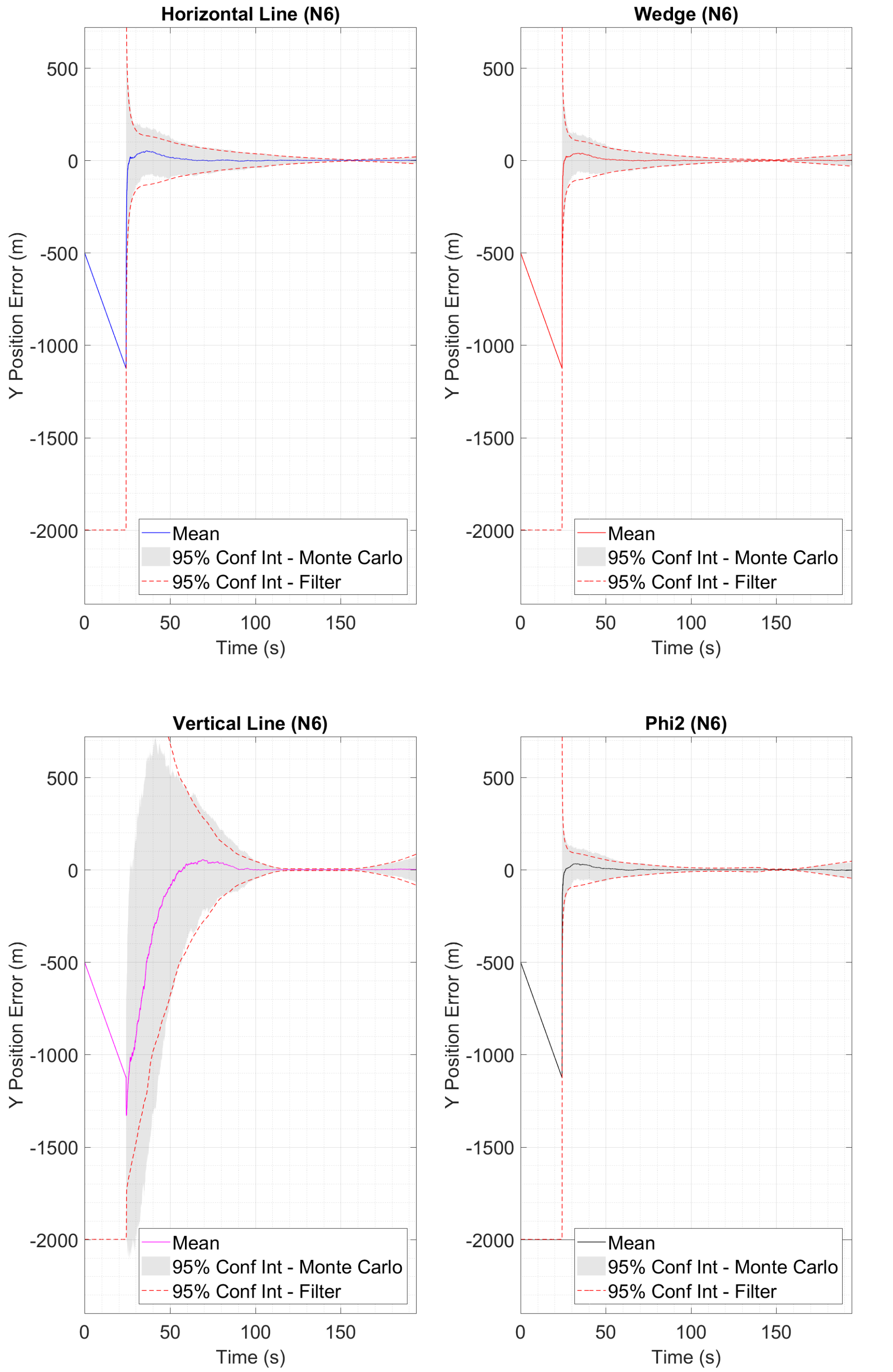

Figure 5.69: Φ_2 vs. Standard Buoy Field Deployments (N=6) - Y Position Error

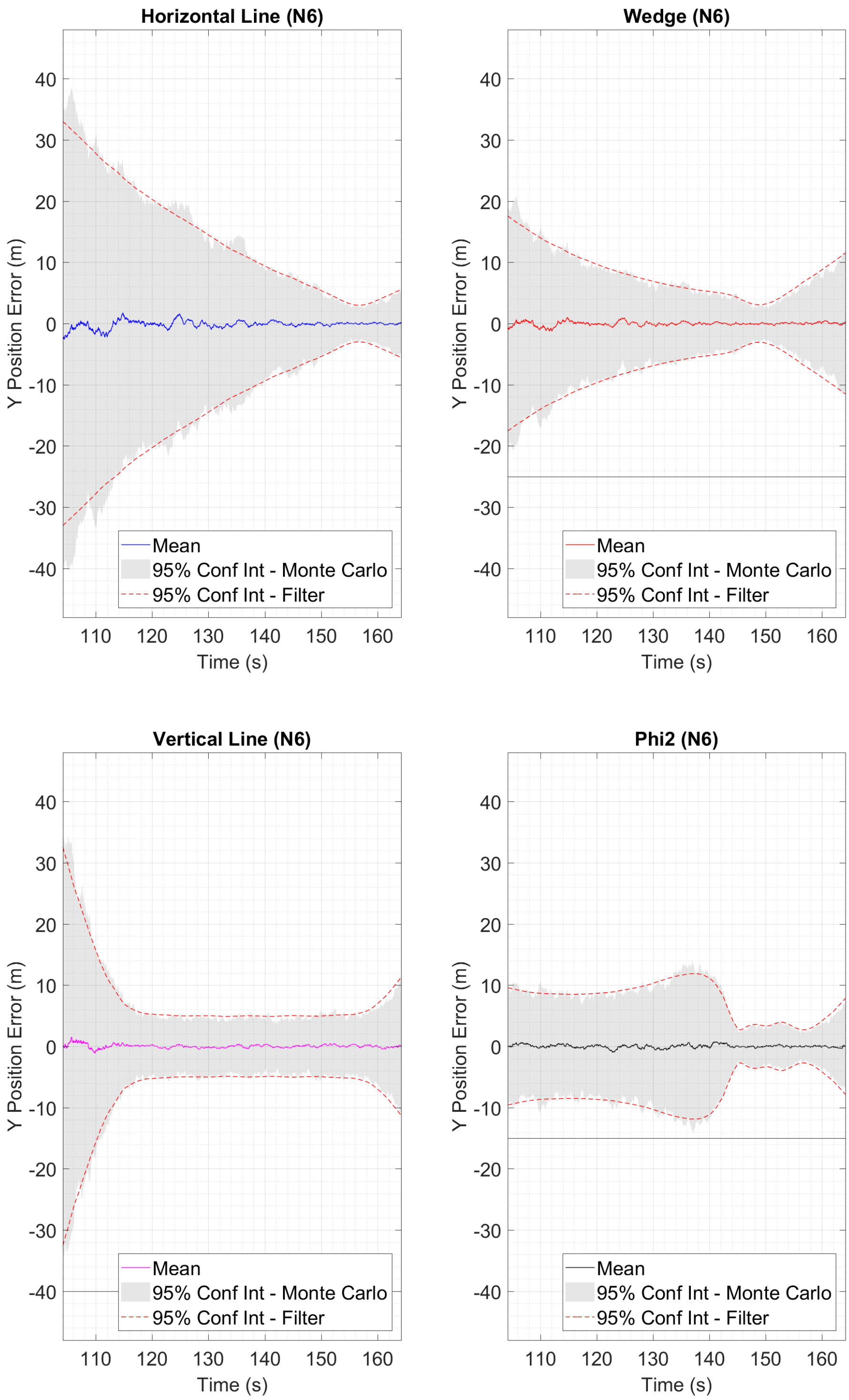

Figure 5.70: Φ_2 vs. Standard Buoy Field Deployments (N=6) - Track and Intercept Time Intervals - Y Position Error

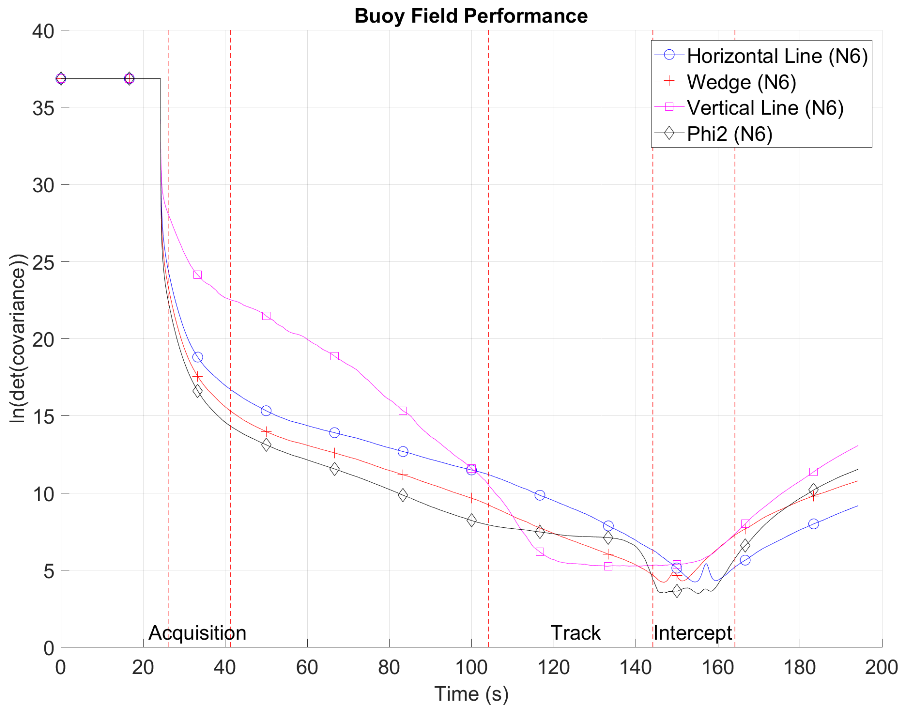

Figure 5.71: Φ_2 vs. Standard Buoy Field Deployments ($N = 6$) - Covariance vs. Time

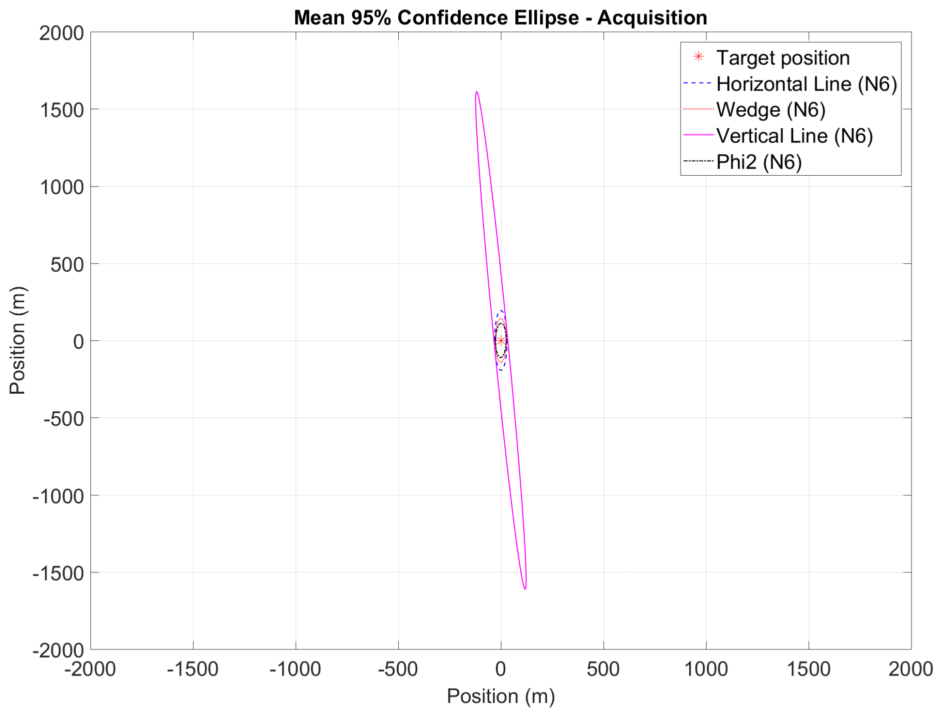

Figure 5.72: Φ_2 vs. Standard Buoy Field Deployments (N=6) - Mean 95% Confidence Ellipse - Acquisition Time Interval

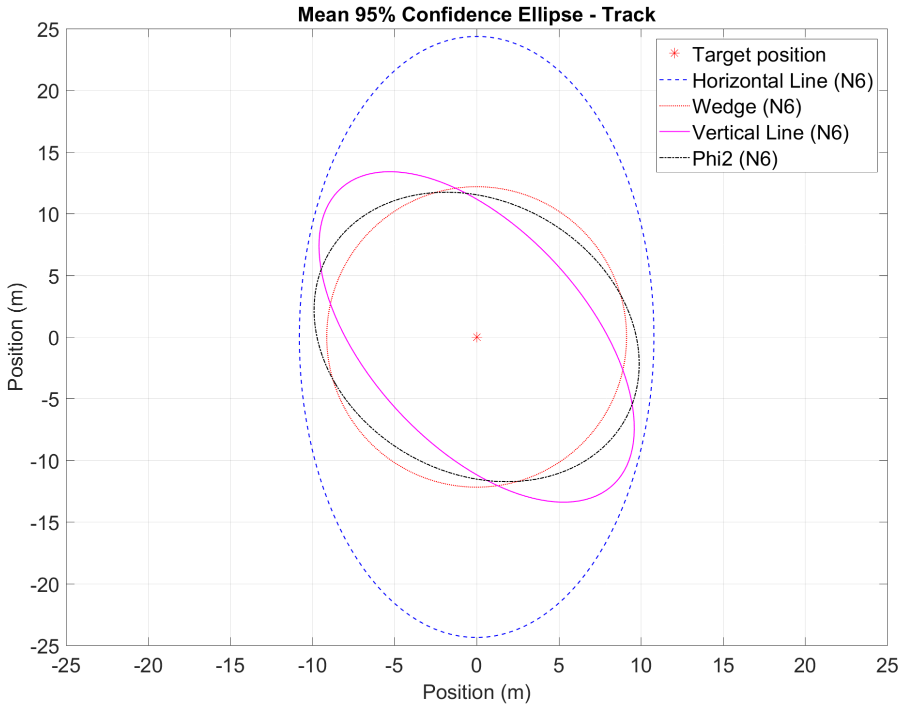

Figure 5.73: Φ_2 vs. Standard Buoy Field Deployments (N=6) - Mean 95% Confidence Ellipse - Track Time Interval

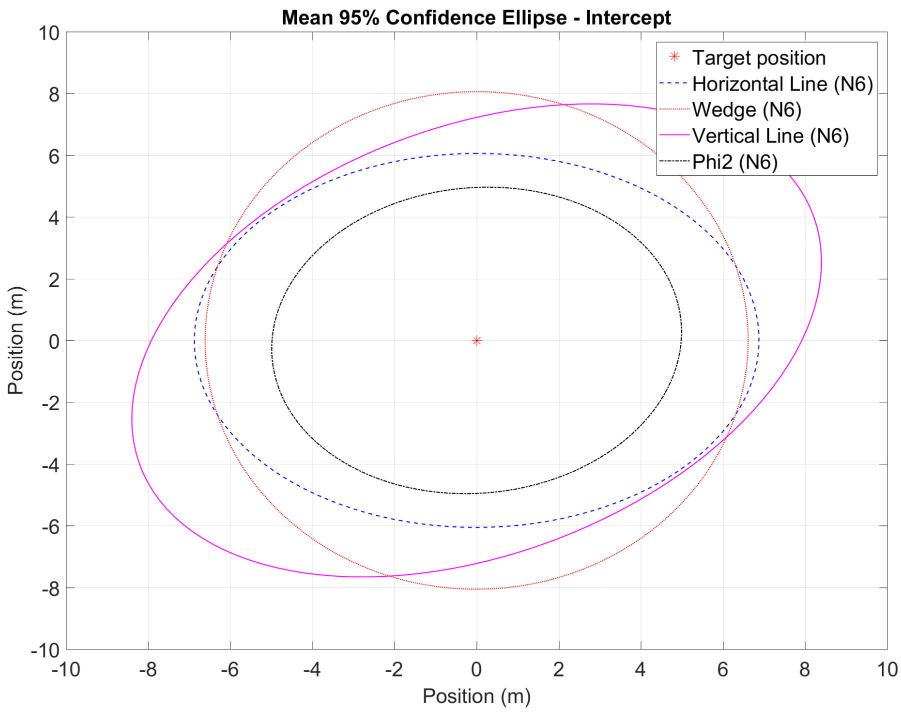

Figure 5.74: Φ_2 vs. Standard Buoy Field Deployments (N=6) - Mean 95% Confidence Ellipse - Intercept Time Interval

5.6 Conclusions and Future Work

Here we summarise the key results from the work presented under Chapters 4 and 5. Chapter 4 established two nonlinear state estimators, an EKF approach and a UKF approach, for the purpose of target tracking in a bearing only buoy field sensor network. A comparison was then made between these two approaches and the impact lateral distribution of buoy fields relative to the target trajectory had on estimator performance. Chapter 5 concerned itself with the challenge of defining an optimal buoy field deployment for target tracking. Some typical deployments were first analysed and initial conclusions were drawn about buoy field geometries. Following, two objective functions Φ_1 and Φ_2 were defined for target tracking performance, and buoy fields with increasing numbers of buoys N were optimised using a genetic algorithm against these objective functions. The results of these optimisations were then compared against the typical buoy field deployments.

It was shown, refer to Section 4.3, that the UKF provides superior state estimation performance in instances where the lateral distribution of the buoy field relative to the target trajectory is small and the angular diversity in the buoy field is extremely poor. Otherwise EKF and UKF provide equivalent performance for target tracking using a bearing-only buoy field sensor network.

It was observed, refer to Section 5.2, that the proximity of the bearing only sensor to the target plays an important role on target tracking performance. This corroborates Meng's observation [124] that for an angular sensor, the quality of the measurement degrades as the distance to the target increases. It was inferred that the performance of a buoy field was dependant on the rates of change and diversity in angular measurements from the sensors to ensure adequate triangulation of the target.

To facilitate both optimisation and the subsequent analysis of an optimal buoy field for target tracking, three key time intervals were defined. These were the acquisition time interval, the track time interval and the intercept time interval (refer to Section 5.1). The two objective functions for optimising buoy fields, Φ_1 and Φ_2, were also defined in this section.

The Φ_1 objective function sought to minimise the target tracking uncertainty of a buoy field equally over the track and intercept time intervals. The Φ_2 objective function sought to minimise the target tracking uncertainty of a buoy field using a tiered approach, accepting a lesser certainty over the track time interval, then minimising the uncertainty over the Intercept time interval. Two candidate buoy fields were obtained, one resultant from the Φ_1 objective function and the other from the Φ_2 objective function, using a genetic algorithm (refer to Section 5.4).

These candidate buoy fields were then compared against the typical buoy field deployments given in Section 5.2, and both shown to provide superior target tracking performance.

The Φ_1 buoy field deployment, refer to Section 5.5.1, demonstrated the benefit of encompassing the target trajectory with a column of sensors, which provided the best tracking performance. This does however introduce a risk if the target trajectory, or likely trajectory, cannot be ascertained prior to deployment of the buoy field. In the scenario explored in this thesis we have assumed a level of prior knowledge and certainty in the target approach direction and trajectory. The deployment of the column buoy field and how tight the column can encompass the target trajectory then depends on the level of confidence in the prior knowledge of the target approach direction and likely trajectory. Such a buoy field deployment can be practically achieved by launching pairs of buoys to the left and right of the ship at set intervals and distances as it sails away from the detected threat. This would iteratively deploy the desired buoy field whilst providing target state information that grows with accuracy and certainty as the buoys are deployed.

The comparison of the Φ_2 buoy field deployment, refer to Section 5.5.2, demonstrated the utility of having a small number of buoys providing overwatch of the target's approach with a larger number of buoys clustered around an ideal intercept point. This deployment strategy provided a balanced approach for target tracking and minimised the amount of buoys required for high target tracking confidence over the interception time interval. Such a deployment method could be achieved practically by use of a projectile deployment method for the buoys. First the overwatch buoys would be deployed at the greatest distance capable of the deployment mechanism. Then a third buoy could be deployed close to the ship to provide measurement diversity of the target leading to improved triangulation. With this initial target state information ascertained and a prediction of the target trajectory, the ship can then plan an ideal intercept point and using the same projectile deployment mechanism deploy the intercept buoy field array. The specific challenges and design of such a deployment mechanism and strategy is out of scope in this thesis.

It is also noted that there is a trade off to be made in buoy field deployment and the level of confidence of the target's likely trajectory. If the target trajectory does not adequately intersect the buoy field, the tracking performance of the state estimators won't be realised. The assumptions made in this thesis is that there is a high certainty of the target approach direction and trajectory given the target will attempt to intercept the ship which deployed the buoy field. This was the scenario explored in the simulation (refer to Section 4.2) in this thesis.

Future work would relax the simplifications made in the simulation and iteratively introduce complexity to explore its impact on the results achieved in this thesis. Additionally, the exploration of a static CRLB approach to optimising a buoy field over geometric areas of high likelihood the target's trajectory will intersect is of interest. This approach would look to achieve an assurance on a minimum level of estimator performance so long as the target passes through the defined areas. Results from this optimisation approach can then be validated through a dynamic simulation. This would provide an avenue for defining sensor field geometries and the subsequent optimal buoy field deployment to guarantee a minimum level of estimator performance.

Chapter 6

Concluding Remarks

This thesis explored the application of nonlinear state estimation within a Defence context. Advances in nonlinear state estimation provides an important technology to enable improved effectiveness within the six key capability streams, highlighted in the 2016 Australian Defence White Paper, required to deliver Australia's Strategic Defence Objectives. The motivation for the application of nonlinear state estimation within Defence is given in Section 1.1.

Two nonlinear state estimation application areas were explored in this thesis. The first application area explored was Fault Detection and Isolation (FDI) as applied to a military vehicle battery system. The application of FDI systems is relevant for the Defence capability stream "Key enablers to support the operation and sustainment of Defence". Chapters 2 and 3 detail this first application where a fault detection system was developed and evaluated for a military vehicle switched battery system. It was shown that this fault detection system was capable of detecting faults in the switched battery system. Two nonlinear state estimation methods were explored for the fault detection system, an Extended Kalman Filter (EKF) and a Moving Horizon Estimator (MHE). Performance between these two methods was shown to be equivalent, but there were distinctions and merits to both methods. The key results and recommendations for future work for fault detection as applied to a switched battery system are given in Section 3.6. These results are also summarised and presented in a conference paper by the Author [41].

The second nonlinear state estimation application area explored was target tracking using a bearing only buoy field sensor network, relevant for the Defence capability stream "Intelligence, Surveillance, Reconnaissance, Space, Electronic Warfare and Cyber Capabilities to ensure superior Situational Awareness". Chapters 4 and 5 detail this second application, where an implementation of a nonlinear state estimator for target tracking using a bearing only buoy field sensor network was explored. An analysis on buoy field geometry and how that impacted the performance of the state estimator was then conducted. Finally two buoy fields were optimised using a genetic algorithm against two different cost functions and it was shown that these buoy fields out performed typical deployment geometries. The key results and recommendations for future work relevant for target tracking using a bearing only buoy field sensor network explored in this thesis are given in Section 5.6.

The overall contributions made by this thesis are summarised in Section 1.3.

From the nonlinear state estimation applications explored in this thesis, it can be seen how this technology can enable the key capability streams that will facilitate the achievement of Australia's Defence Strategic Objectives. Nonlinear state estimation continues to be a heavily researched field and advances in the methods and enabling technologies, such as computers and sensors, will empower these methods to deliver improved mission effectiveness leading to a more capable, agile and potent Australian Defence Force (ADF).

Chapter 7

Bibliography

[1] P. Jennings, J. Goldrick, R. Medcalf, M. Kalms, A. Davies, and S. Fruhling, *Australian Defence White Paper*, 2016.

[2] C. Trammel, G. Vossler, and M. Feldmann, "UK Ministry of Defence generic health and usage monitoring system (GenHUMS)," *Aircraft Engineering and Aerospace Technology*, vol. 69, no. 5, pp. 414–422, 1997.

[3] A. Draper, "The operational benefits of health and usage monitoring systems in UK military helicopters," in *Third International Conference on Health and Usage Monitoring-HUMS2003*, 2002, p. 71.

[4] M. Mallick, B.-N. Vo, T. Kirubarajan, and S. Arulampalam, "Introduction to the issue on multitarget tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 373–375, 2013.

[5] T. B. Sch, F. Lindsten, J. Dahlin, W. Johan, C. A. Naesseth, A. Svensson, and L. Dai, "Sequential Monte Carlo Methods for System Identification," 2015.

[6] D. Simon, *Optimal State Estimation.* Hoboken, NJ, USA: John Wiley & Sons, Inc., may 2006.

[7] M. S. Grewal and A. P. Andrews, "Applications of Kalman Filtering in Aerospace 1960 to the Present," *IEEE Control Systems*, vol. 30, no. 3, pp. 69–78, 2010.

[8] S. M. Stigler, "Gauss and the Invention of Least Squares," *The Annals of Statistics*, vol. 9, no. 3, pp. 465–474, 1981.

[9] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series : With Engineering Applications.* MIT Press Ltd, 1949.

[10] A. Kolmogorov, "Interpolation and Extrapolation," *Bulletin de l'Academie des Sciences de U.S.S.R*, vol. Series Mat, pp. 3–14, 1941.

[11] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *Journal of Applied Physics*, vol. 21, no. 7, pp. 645–655, 1950.

[12] L. Zadeh and J. Ragazzini, "Optimum Filters for the Detection of Signals in Noise," *Proceedings of the IRE*, vol. 40, no. 10, pp. 1223–1231, oct 1952.

[13] K. Johnson, "Optimum, linear, discrete filtering of signals containing a nonrandom component," *IEEE Transactions on Information Theory*, vol. 2, no. 2, pp. 49–55, jun 1956.

[14] F. J. Beutler, "Prediction and filtering for random parameter systems," *IRE Transactions on Information Theory*, vol. 4, no. 4, pp. 166–171, dec 1958.

[15] C. Bissell, "A history of automatic control," *Springer Handbook of Automation*, pp. 53–69, 2009.

[16] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[17] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, vol. 83, no. 1, p. 95, 1961.

[18] L. A. Mcgee and S. F. Schmidt, "Discovery of the Kalman filter as a practical tool for aerospace and industry," no. November, 1985.

[19] G. Zames, "Feedback and optimal sensitivity: model reference transformation, multiplicative semiforms and approximate inverse," *IEEE Transactions on automatic control*, vol. 26N, no. 2, pp. 301–320, 1981.

[20] M. J. Grimble and D. Biss, "Selection of optimal control weighting functions to achieve good H-infinity robust designs," *Control, 1988. CONTROL 88., International Conference on*, pp. 683–688, 1988.

[21] M. Grimble, "H-infinity fixed-lag smoothing filter for scalar systems," *IEEE Transactions on Signal Processing*, vol. 39, no. 9, pp. 1955–1963, 1991.

[22] R. Banavar and J. Speyer, "A linear-quadratic game approach to estimation and smoothing," *American Control Conference, 1991*, pp. 2818–2822, 1991.

[23] J. Doyle, K. Glover, P. Khargonekar, and B. Francis, "State-space solutions to standard H-2 and H-infinity control problems," *IEEE Transactions on Automatic Control*, vol. 34, no. 8, pp. 831–847, 1989.

[24] J. Ball and J. Helton, "H-infinity control for nonlinear plants: connections with differential games," in *Proceedings of the 28th IEEE Conference on Decision and Control.* IEEE, 1989, pp. 956–962.

[25] J. Ball, J. Helton, and M. Walker, "H-infinity control for nonlinear systems with output feedback," *IEEE Transactions on Automatic Control*, vol. 38, no. 4, pp. 546–559, apr 1993.

[26] T. B. Schön, "Estimation of Nonlinear Dynamic Systems Theory and Applications," Ph.D. dissertation, Linkoping Universitet, 2006.

[27] G. Kitagawa, "A nonlinear smoothing method for time series analysis." *Statistica Sinica*, vol. 1, no. 2, pp. 371–388, 1991.

[28] A. H. Jazwinski, *Stochastic Processes and Filtering Theory.* Dover Publications, 1970.

[29] D. L. Alspach and H. W. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[30] S. J. Julier, J. K. Uhlmann, and H. Durrant-Whyte, "A new approach for filtering nonlinear systems," *Proceedings of 1995 American Control Conference - ACC'95*, vol. 3, no. June, pp. 0–4, 1995.

[31] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, mar 2000.

[32] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[33] G. Li, F. Sun, and N. Cheng, "Performance Analysis of UKF for Nonlinear Problems," in *2009 Third International Symposium on Intelligent Information Technology Application*, vol. 2. IEEE, 2009, pp. 209–212.

[34] J.-B. Lacambre, M. Narozny, and J.-M. Louge, "Limitations of the unscented Kalman filter for the attitude determination on an inertial navigation system," in *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*. IEEE, aug 2013, pp. 187–192.

[35] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, p. 107, 1993.

[36] X. L. Hu, T. B. Schön, and L. Ljung, "A basic convergence result for particle filtering," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 7, no. PART 1, pp. 288–293, 2007.

[37] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing : fiteen years later," *Handbook of nonlinear filtering*, 2009.

[38] F. Lindsten, A. M. Johansen, C. A. Naesseth, B. Kirkpatrick, T. B. Schön, J. A. Aston, and A. Bouchard-Côté, "Divide-and-Conquer With Sequential Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 26, no. 2, pp. 445–458, 2017.

[39] J. B. Rawlings and D. Q. Mayne, "Model Predictive Control : Theory and Design," pp. 1–11, 2012.

[40] C. V. Rao, J. B. Rawlings, and J. H. Lee, "Constrained linear state estimationa moving horizon approach," *Automatica*, vol. 37, no. 10, pp. 1619–1628, oct 2001.

[41] J. Mitchell and I. Manchester, "Fault detection for a switched battery system via constrained nonlinear state estimation," in *IEEE Conference on Industrial Electronics and Applications*, Wuhan, jun 2018.

[42] A. Avizienis, "Fault-tolerance: The survival attribute of digital systems," *Proceedings of the IEEE*, vol. 66, no. 10, pp. 1109–1125, 1978.

[43] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, no. 6, pp. 601–611, 1976.

[44] R. V. Beard, "Failure accomodation in linear systems through self-reorganization," Ph.D. dissertation, 1971.

[45] H. L. Jones, "Failure detection in linear systems," Ph.D. dissertation, 1973.

[46] F. Kimmich, A. Schwarte, and R. Isermann, "Fault detection for modern Diesel engines using signal- and process model-based methods," *Control Engineering Practice*, vol. 13, no. 2, pp. 189–203, 2005.

[47] D. M. Himmelblau, *Fault detection and diagnosis in chemical and petrochemical processes.* Elsevier Scientific Publishing Company, 1978.

[48] E. Chow and A. S. Willsky, "Analytical redundancy and the design of robust failure detection systems," *Automatic Control, IEEE Transactions on*, vol. 29, no. 7, pp. 603–614, 1984.

[49] J. Gertler, "Fault detection and isolation using parity relations," *Control Engineering Practice*, vol. 5, no. 5, pp. 653–661, 1997.

[50] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of Process Control*, vol. 7, no. 6, pp. 403–424, 1997.

[51] J. Gertler, "Analytical redundancy methods in fault detection and isolation," in *Proceedings of IFAC/IAMCS symposium on safe process*, vol. 1, 1991, pp. 9–21.

[52] R. Isermann, "Process fault detection based on modeling and estimation methodsA survey," *Automatica*, vol. 20, no. 4, pp. 387–404, 1984.

[53] T. Sproesser and G. L. Gissinger, "A method for fault detection using parameter and state estimation," *Annual Review in Automatic Programming*, vol. 17, no. 0, pp. 241–246, 1992.

[54] A. Pouliezos, G. Stavrakakis, and C. Lefas, "Fault detection using parameter estimation," *Quality and Reliability Engineering International*, vol. 5, no. 4, pp. 283–290, 1989.

[55] L. Ljung, *System Identification: Theory for the User.* Prentice Hall PTR, 1999.

[56] S. Töpfer, A. Wolfram, and R. Isermann, "Semi-physical modelling of nonlinear processes by means of local model approaches," in *Proc. 15th IFAC World Congress*, 2002.

[57] S. Ding, *Model-based fault diagnosis techniques.* Springer, 2008.

[58] R. Isermann and P. Ballé, "Trends in the application of model-based fault detection and diagnosis of technical processes," *Control Engineering Practice*, vol. 5, no. 5, pp. 709–719, 1997.

[59] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods An introduction," *Control Engineering Practice*, vol. 5, no. 5, pp. 639–652, 1997.

[60] ——, "Model-based fault-detection and diagnosis status and applications," *Annual Reviews in Control*, vol. 29, pp. 71–85, 2005.

[61] H. Soltani, S. B. H. A. Naoui, A. Aitouche, R. E. Harabi, and M. Abdelkrim, "Robust Simultaneous Fault Detection and Control Approach for Time-Delay Systems," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 1244–1249, 2015.

[62] M. T. Raza, A. Q. Khan, G. Mustafa, and M. Abid, "Robust Fault Detection and Isolation for Uncertain Switched Systems under Asynchronous Switching with Adaptive Threshold," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 131–136, 2015.

[63] P. Baldi, M. Blanke, P. Castaldi, N. Mimmo, and S. Simani, "Combined Geometric and Neural Network Approach to Generic Fault Diagnosis in Satellite Reaction Wheels," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 194–199, 2015.

[64] P. Witczak, K. Patan, M. Witczak, V. Puig, and J. Korbicz, "A neural network-based robust unknown input observer design: Application to wind turbine," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 263–270, 2015.

[65] Q. Zhang, "Stochastic Hybrid System Actuator Fault Diagnosis by Adaptive Estimation," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 150–155, 2015.

[66] W. Li, F. Jiang, Z. Wang, G. Zhou, and Z. Zhu, "Fault detection of markov jumping linear systems," *Mathematical Problems in Engineering*, vol. 2012, pp. 1–27, 2012.

[67] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[68] E. Alcorta García and P. M. Frank, "Deterministic nonlinear observer-based approaches to fault diagnosis: A survey," *Control Engineering Practice*, vol. 5, no. 5, pp. 663–670, 1997.

[69] J. Marzat, H. Piet-Lahanier, F. Damongeot, and E. Walter, "Model-based fault diagnosis for aerospace systems: a survey," Sage UK: London, England, pp. 1329–1360, 2012.

[70] L.-F. Pau, "Failure diagnosis and performance monitoring," 1981.

[71] R. J. Patton, P. M. Frank, and R. N. Clarke, *Fault diagnosis in dynamic systems: theory and application.* Prentice-Hall, Inc., 1989.

[72] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems.* Taylor \& Francis, 1998.

[73] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems.* Springer Publishing Company, Incorporated, 1999.

[74] E. L. R. L. R. D. B. L.H.Chiang, *Fault Detection and Diagnosis in Industrial Systems.* Springer, 2001.

[75] A. Radke and G. Zhiqiang, "A survey of state and disturbance observers for practitioners," in *American Control Conference, 2006*, 2006, p. 6 pp.

[76] D. G. Luenberger, "An Introduction to Observers," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 596–602, 1971.

[77] D. Graupe, *Principles of Artificial Neural Networks*, 2013.

[78] M. H. Terra and R. Tinós, "Fault detection and isolation in robotic manipulators via neural networks: A comparison among three architectures for residual analysis," *Journal of Robotic Systems*, vol. 18, no. 7, pp. 357–374, jul 2001.

[79] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010.

[80] F. Gustafsson, *Adaptive Filtering and Change Detection.* Chichester, UK: John Wiley & Sons, Ltd, oct 2001, vol. 15, no. 9.

[81] R. Barlow and P. Chatterjee, "Introduction to Fault Tree Analysis," Tech. Rep., 1973.

[82] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[83] T. J. Ross, *Fuzzy Logic with Engineering Applications*, 2004.

[84] A. Genovesi, J. Harmand, and J.-p. Steyer, "A fuzzy logic based diagnosis system for the on-line supervision of an anaerobic digestor pilot-plant," *Biochemical Engineering Journal*, vol. 3, no. 3, pp. 171–183, jun 1999.

[85] R. Patton, J. Chen, and C. Lopez-Toribio, "Fuzzy observers for nonlinear dynamic systems fault diagnosis," in *Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171)*, vol. 1, no. December. IEEE, 1998, pp. 84–89.

[86] R. Isermann, "Fault diagnosis of diesel engines," *Mechanical Engineering*, vol. 135, no. 12, 2013.

[87] B. Larder, H. Azzam, C. Trammel, and G. Vossler, "Smith Industries HUMS: changing the M from monitoring to management," in *Aerospace Conference Proceedings, 2000 IEEE*, vol. 6, 2000, pp. 449–455 vol.6.

[88] Y. Zhang, L. Zhang, and H. Zhang, "Fault detection for industrial processes," *Mathematical Problems in Engineering*, vol. 2012, 2012.

[89] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion battery," *Journal of Power Sources*, vol. 196, no. 15, pp. 6007–6014, 2011.

[90] "Future Land Warfare Report," Modernisation and Strategic Planning Division, Commonwealth of Australia, Tech. Rep., 2014.

[91] M. A. Casacca and Z. M. Salameh, "Determination of lead-acid battery capacity via mathematical modeling techniques," *IEEE Transactions on Energy Conversion*, vol. 7, no. 3, pp. 442–446, 1992.

[92] Z. M. Salameh, M. A. Casacca, and W. A. Lynch, "A mathematical model for lead-acid batteries," *IEEE Transactions on Energy Conversion*, vol. 7, no. 1, pp. 93–98, 1992.

[93] W. Chen, W.-T. Chen, M. Saif, M.-F. Li, and H. Wu, "Simultaneous Fault Isolation and Estimation of Lithium-Ion Batteries via Synthesized Design of Luenberger and Learning Observers," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 290–298, 2014.

[94] J. Zhang and C. Xia, "State-of-charge estimation of valve regulated lead acid battery based on multi-state Unscented Kalman Filter," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 3, pp. 472–476, 2011.

[95] C. R. Gould, C. M. Bingham, D. A. Stone, and P. Bentley, "New Battery Model and State-of-Health Determination Through Subspace Parameter Estimation and State-Observer Techniques," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 8, pp. 3905–3916, 2009.

[96] H. Ham, K. Han, and H. Lee, "Battery System Modeling for a Military Electric Propulsion Vehicle with a Fault Simulation," *Energies*, vol. 6, no. 10, pp. 5168–5181, 2013.

[97] D. Linden and T. Reddy, *Linden's Handbook of Batteries*, 3rd ed. McGraw-Hill Education, 2010.

[98] A. Jossen, "Fundamentals of battery dynamics," *Journal of Power Sources*, vol. 154, no. 2, pp. 530–538, 2006.

[99] N. Perez, Ed., *Electrochemistry and Corrosion Science.* Boston: Kluwer Academic Publishers, 2004.

[100] H. He, R. Xiong, and J. Fan, "Evaluation of Lithium-Ion Battery Equivalent Circuit Models for State of Charge Estimation by an Experimental Approach," *Energies*, vol. 4, no. 4, pp. 582–598, 2011.

[101] J. Chiasson and B. Vairamohan, "Estimating the state of charge of a battery," *IEEE Transactions on Control Systems Technology*, vol. 13, no. 3, pp. 465–470, 2005.

[102] V. Johnson, "Battery performance models in ADVISOR," *Journal of Power Sources*, vol. 110, no. 2, pp. 321–329, 2002.

[103] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background," *Journal of Power Sources*, vol. 134, no. 2, pp. 252–261, 2004.

[104] ——, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification," *Journal of Power Sources*, vol. 134, no. 2, pp. 262–276, 2004.

[105] ——, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation," *Journal of Power Sources*, vol. 134, no. 2, pp. 277–292, 2004.

[106] S. Piller, M. Perrin, and A. Jossen, "Methods for state-of-charge determination and their applications," *Journal of Power Sources*, vol. 96, no. 1, pp. 113–120, 2001.

[107] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, p. 1303, 2010.

[108] J. Lofberg, "YALMIP : a toolbox for modeling and optimization in MATLAB," *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, no. May, pp. 284–289, 2014.

[109] S. Cheong and I. R. Manchester, "Input design for discrimination between classes of LTI models," *Automatica*, vol. 53, pp. 103–110, 2015.

[110] R. James, "A history of radar," *IEE Review*, vol. 35, no. 9, p. 343, 1989.

[111] S. S. Blackman, *Multiple-Target Tracking With Radar Applications.* ARTECH House, Inc., 1986.

[112] J. K. Uhlmann, "Algorithms for Multiple-Target Tracking," *American Scientist*, vol. 80, no. 2, pp. 128–141, 1992.

[113] N. Wax, "Signal-to-noise improvement and the statistics of track populations," *Journal of Applied Physics*, vol. 26, no. 5, pp. 586–595, 1955.

[114] R. W. Sittler, "An Optimal Data Association Problem in Surveillance Theory," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 125–139, 1964.

[115] R. Singer and J. Stein, "An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems," *1971 IEEE Conference on Decision and Control*, pp. 171–175, 1971.

[116] A. D'Amico and R. Pittenger, "A brief history of active sonar," *Aquatic Mammals*, vol. 35, no. 4, pp. 426–434, 2009.

[117] R. J. Vaccaro, "The past, present, and the future of underwater acoustic signal processing," *Signal Processing Magazine, IEEE*, vol. 15, no. 4, pp. 21–51, 1998.

[118] S. Nardone and A. Lindgren, "Fundamental properties and performance of conventional bearings-only target motion analysis," *IEEE Transactions on Automatic Control*, vol. 29, no. 9, pp. 775–787, 1984.

[119] S. Nardone and M. Graham, "A Closed-Form Solution to Bearings-Only Target Motion Analysis," *IEEE Journal of Oceanic Engineering*, vol. 22, no. 1, pp. 168–178, 1997.

[120] X. R. Li and V. Jilkov, "Survey of maneuvering targettracking . part I: dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.

[121] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, I. Kadar, Ed., vol. 3068, jul 1997, p. 182.

[122] C. Yang, L. Kaplan, E. Blasch, and M. Bakich, "Optimal placement of heterogeneous sensors for targets with gaussian priors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1637–1653, 2013.

[123] A. N. Bishop, B. Fidan, B. D. Anderson, K. Doançay, and P. N. Pathirana, "Optimality analysis of sensor-target localization geometries," *Automatica*, vol. 46, no. 3, pp. 479–492, 2010.

[124] W. Meng, L. Xie, and W. Xiao, "Optimality analysis of sensor-source geometries in heterogeneous sensor networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1958–1967, 2013.

[125] O. Tekdas and V. Isler, "Sensor placement for triangulation-based localization," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 3, pp. 681–685, 2010.

[126] G. K. Rao and C. Radhakrishna, "Indian Statistical Institute," *The Indian Journal of Statistics*, vol. 15, pp. 331–342, 1955.

[127] M. Mitchell, "An Introduction to Genetic Algorithms (Complex Adaptive Systems)," *The MIT Press*, p. 221, 1998.

[128] J. H. Holland, "Outline for a Logical Theory of Adaptive Systems," *Journal of the ACM*, vol. 9, no. 3, pp. 297–314, 1962.

[129] A. Alessandri, M. Baglietto, G. Battistelli, and V. Zavala, "Advances in moving horizon estimation for nonlinear systems," *49th IEEE Conference on Decision and Control (CDC)*, pp. 5681–5688, 2010.

[130] S. Simani, "Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques," Ph.D. dissertation, University of Modena and Reggio Emilia, 1999.

Appendix A

Nonlinear State Estimation Equations

A.1 Extended Kalman Filter

Following is a standard implementation of the EKF [6]. Assuming a system and measurement model as follows:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k \\ y_k &= h(x_k, u_k) + v_k \\ v_k &\sim \mathcal{N}(0, Q) \\ w_k &\sim \mathcal{N}(0, R) \end{aligned} \tag{A.1.1}$$

1. Initialise the EKF

$$\begin{aligned} \hat{x}_0^+ &= \mathbb{E}[x_0] \\ P_0^+ &= \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \tag{A.1.2}$$

2. For $k = 0, 1, 2, ...$ iterate as follows:

(a) Calculate the process partial derivative matrices:

$$\begin{aligned} F_k &= \left.\frac{\partial f}{\partial x}\right|_{\hat{x}_k^+} \\ V_k &= \left.\frac{\partial f}{\partial w}\right|_{\hat{x}_k^+} \end{aligned} \tag{A.1.3}$$

(b) Update state estimate and covariance according to the process model:

$$\begin{aligned} P_{k+1}^- &= F_k P_k^+ F_k^T + V_k Q_k V_k^T \\ \hat{x}_{k+1}^- &= f(\hat{x}_k^+, u_k) \end{aligned} \tag{A.1.4}$$

(c) Calculate the measurement partial derivative matrices:

$$\begin{aligned} H_{k+1} &= \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_{k+1}^-} \\ W_{k+1} &= \left.\frac{\partial h}{\partial v}\right|_{\hat{x}_{k+1}^-} \end{aligned} \tag{A.1.5}$$

(d) Update the state estimate and covariance according to the measurement:

$$\begin{aligned} K &= P_{k+1}^- H_{k+1}^T (H_{k+1} P_{k+1}^- H_{k+1}^T + W_{k+1} R_{k+1} W_{k+1}^T)^{-1} \\ \hat{x}_{k+1}^+ &= \hat{x}_{k+1}^- + K(y_{k+1} - h(\hat{x}_{k+1}^-, u_{k+1})) \\ P_{k+1}^+ &= (I - K H_{k+1}) P_{k+1}^- \end{aligned} \tag{A.1.6}$$

A.2 Unscented Kalman Filter

Following is a standard implementation of the Unscented Kalman Filter (UKF) [32]. Assuming a system and measurement model as follows:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k \\ y_k &= h(x_k, u_k) + v_k \\ v_k &\sim \mathcal{N}(0, Q) \\ w_k &\sim \mathcal{N}(0, R) \end{aligned} \tag{A.2.1}$$

1. Initialise the UKF

$$\begin{aligned} \hat{x}_0^+ &= \mathbb{E}[x_0] \\ P_0^+ &= \mathbb{E}[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \tag{A.2.2}$$

2. For $k = 0, 1, 2, ...$ iterate as follows:

(a) Calculate Sigma points based on previous distribution knowledge

$$\begin{aligned} \hat{x}_k^i &= \hat{x}_k^+ + \tilde{x}^i \qquad i = 1...2n \\ \tilde{x}^i &= \left(\sqrt{nP_k^+}\right)_i^T \qquad i = 1...2n \\ \tilde{x}^{n+i} &= -\left(\sqrt{nP_k^+}\right)_i^T \qquad i = 1...2n \end{aligned} \tag{A.2.3}$$

Where $\sqrt{nP_k^+}$ is the square root of the covariance matrix P_k^+ weighted by n such that $\left(\sqrt{nP_k^+}\right)^T\left(\sqrt{nP_k^+}\right) = nP$. The notation $\left(\sqrt{nP_k^+}\right)_i$ indicates the ith row of $\sqrt{nP_k^+}$.

(b) Propagate Sigma points through nonlinear process model

$$\hat{x}_{k+1}^i = f(\hat{x}_k^i, u_k) \qquad i = 1...2n \tag{A.2.4}$$

(c) Calculate state mean and covariance from transformed Sigma points

$$\begin{aligned}
\hat{x}_{k+1}^- &= \frac{1}{2n}\sum_{i=1}^{2n}\hat{x}_{k+1}^i \\
P_{k+1}^- &= \frac{1}{2n}\sum_{i=1}^{2n}(\hat{x}_{k+1}^i - \hat{x}_{k+1}^-)(\hat{x}_{k+1}^i - \hat{x}_{k+1}^-)^T + Q_{k+1}
\end{aligned} \tag{A.2.5}$$

(d) Recalculate Sigma points based on updated Gaussian distribution step (b + c)

$$\begin{aligned}
\hat{x}_{k+1}^i &= \hat{x}_{k+1}^- + \tilde{x}^i \qquad i = 1...2n \\
\tilde{x}^i &= \left(\sqrt{nP_{k+1}^-}\right)_i^T \qquad i = 1...2n \\
\tilde{x}^{n+i} &= -\left(\sqrt{nP_{k+1}^-}\right)_i^T \qquad i = 1...2n
\end{aligned} \tag{A.2.6}$$

(e) Propagate Sigma points through nonlinear measurement model

$$\hat{y}_{k+1}^i = h(\hat{x}_{k+1}^i, u_{k+1}) \qquad i = 1...2n \tag{A.2.7}$$

(f) Calculate state mean and covariance from Sigma points and cross covariance between $\hat{x}_{k+1}^-$ and $\hat{y}_{k+1}^i$

$$\begin{aligned}
\hat{y}_{k+1} &= \frac{1}{2n}\sum_{i=1}^{2n}\hat{y}_{k+1}^i \\
P_y &= \frac{1}{2n}\sum_{i=1}^{2n}(\hat{y}_{k+1}^i - \hat{y}_{k+1})(\hat{y}_{k+1}^i - \hat{y}_{k+1})^T + R_{k+1} \\
P_{xy} &= \frac{1}{2n}\sum_{i=1}^{2n}(\hat{x}_{k+1}^i - \hat{x}_{k+1}^-)(\hat{y}_{k+1}^i - \hat{y}_{k+1})^T
\end{aligned} \tag{A.2.8}$$

(g) Calculate Kalman gain and update the state estimate

$$\begin{aligned} K &= P_{xy}P_y^{-1} \\ \hat{x}_{k+1}^{+} &= \hat{x}_{k+1}^{-} + K(y_{k+1} - \hat{y}_{k+1}) \\ P_{k+1}^{+} &= P_{k+1}^{-} - KP_yK^T \end{aligned} \tag{A.2.9}$$

A.3 Moving Horizon Estimator

Following is a standard implementation of the MHE with a sliding window of size N [129]. Assuming a system and measurement model as follows:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) \\ y_k &= h(x_k, u_k) \end{aligned} \tag{A.3.1}$$

1. If $k < N$, for k = 1, 2, 3, ..., N-1 iterate as follows:

$$\text{Solve} \quad \arg\min_{\hat{x}_{\{1:k\}}} \Phi(\bar{x}_0, y_{\{1:k\}}) \tag{A.3.2}$$

$$\begin{aligned} \Phi = ||\hat{x}_0 - \bar{x}_0||_P^2 + \sum_{i=0}^{k-1} ||\hat{x}_{i+1} - f(\hat{x}_i, u_i)||_Q^2 \\ + \sum_{i=0}^{k} ||y_i - h(\hat{x}_i, u_i)||_R^2 \end{aligned}$$

2. Else for k = N, N+1, N+2, ... iterate as follows:

$$\text{Solve} \quad \arg\min_{\hat{x}_{\{k-N:k\}}} \Phi(\bar{x}_{k-N}, y_{\{k-N:k\}}) \tag{A.3.3}$$

$$\begin{aligned} \Phi = ||\hat{x}_{k-N} - \bar{x}_{k-N}||_P^2 + \sum_{i=k-N}^{k-1} ||\hat{x}_{i+1} - f(\hat{x}_i, u_i)||_Q^2 \\ + \sum_{i=k-N}^{k} ||y_i - h(\hat{x}_i, u_i)||_R^2 \end{aligned}$$

Appendix B

Fault Detection and Isolation Nomenclature

Published by Isermann and Ball [58] — with the support of the IFAC SAFEPROCESS Technical Committee — the nomenclature for the field of supervision, fault detection and diagnosis has been defined as follows:

B.1 States and Signals

Fault: An unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable / usual / standard condition.

Failure: A permanent interruption of a system's ability to perform a required function under specified operating conditions.

Malfunction: An intermittent irregularity in the fulfilment of a system's desired function.

Error: A deviation between a measured or computed value (of an output variable) and the true, specified or theoretically correct value.

Disturbance: An unknown (and uncontrolled) input acting on a system.

Perturbation: An input acting on a system, which results in a temporary departure from the current state.

Residual: A fault indicator, based on a deviation between measurements and model-equation-based computations.

Symptom: A change of an observable quantity from normal behaviour.

B.2 Functions

Fault detection: Determination of the faults present in a system and the time of detection.

Fault isolation: Determination of the kind, location and time of detection of a fault. Follows fault detection.

Fault identification: Determination of the size and time-variant behaviour of a fault. Follows fault isolation.

Fault diagnosis: Determination of the kind, size, location and time of detection of a fault. Follows fault detection. Includes fault isolation and identification.

Monitoring: A continuous real-time task of determining the conditions of a physical system, by recording information, recognising and indicating anomalies in the behaviour.

Supervision: Monitoring a physical system and taking appropriate actions to maintain the operation in the case of faults.

Protection: means by which a potentially dangerous behaviour of the system is suppressed if possible, or means by which the consequences of a dangerous behaviour are avoided.

B.3 Models

Quantitative model: Use of static and dynamic relations among system variables and parameters in order to describe a system's behaviour in quantitative mathematical terms.

Qualitative model: Use of static and dynamic relations among system variables and parameters in order to describe a system's behaviour in qualitative terms such as causalities or if-then rules.

Diagnostic model: A set of static or dynamic relations which link specific input variables — the symptoms — to specific output variables — the faults.

Analytical redundancy: Use of two or more (but not necessarily identical) ways to determine a variable, where one way uses a mathematical process model in analytical form.

Robustness: A model is robust if it remains accurate in the presence of bounded model uncertainties and noise.

B.4 System Properties

Reliability: Ability of a system to perform a required function under state conditions, within a given scope, during a given period of time. Measure: MTBF = Mean Time BEtween Failures. MTBF $= \frac{1}{\lambda}$; λ : rate of failure (e.g. failures per year).

Safety: Ability of a system not to cause danger to persons or equipment or the environment.

Availability: Probability that a system or equipment will operate satisfactorily and effectively at any point of time. Measure:

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

MTTR: Mean Time To Repair
MTTR $= \frac{1}{\mu}$; μ: rate of repair.

Dependability: A form of availability that has the property of always being available when required. It is the degree to which a system is operable and capable of performing its required function at any randomly chosen time during its specified operating time, provided that the item is available at the start of that period.

$$D = \frac{\text{Time available}}{\text{Time available} + \text{Time required}}$$

B.5 Time dependency of faults

Additionally, Simani [130] provides descriptions on the typical types faults that are investigated with Fault Detection and Isolation methods:

Abrupt fault: Fault modeled as stepwise function. It represents bias in the monitored signal.

Incipient fault: Fault modeled by using ramp signals. It represents drift of the monitored signal.

Intermittent fault: Combination of impulses with varying amplitudes.

B.6 Fault typology

Additive fault: Influences a variable by an addition of the fault itself to the sensor output.

Multiplicative fault: Are represented by the product of a variable with the fault itself. They typically appear as parameter changes within a process.

B.7 Redundancy

Analytical redundancy was decomposed into the following definitions established by Chow and Willsky [48]

Direct redundancy: The relationship among instantaneous outputs of sensors. (i.e. algebraic relationships between sensor outputs)

Temporal redundancy: The relationship among the histories of sensor outputs and actuator inputs. (i.e. differential or difference relationships between sensor inputs and outputs)

Appendix C

Acronyms

ACFR Australian Centre for Field Robotics

ADF Australian Defence Force

ANN Artificial Neural Network

ASDIC Allied Submarine Detection Investigation Committee

C4ISR Command Control Communications Computers Intelligence Surveillance and Reconnaissance

CBM Condition Based Maintenance

CoA Commonwealth of Australia

CRLB Cramer-Rao Lower Bound

ECM Equivalent Circuit Model

EKF Extended Kalman Filter

FDI Fault Detection and Isolation

FTA Fault Tree Analysis

FTC Fault Tolerant Control

HUMS Health Usage and Monitoring System

IFAC International Federation of Automatic Control

ISR Intelligence Surveillance and Reconnaissance

JSEA Job Safety and Environmental Analysis

MHE Moving Horizon Estimator

MoD Ministry of Defence

MPC Model Predictive Control

MTT Multi-Target Tracking

NCW Network Centric Warfare

NSSM Nonlinear State Space Model

ODE Ordinary Differential Equation

PDF Probability Density Function

PEM Parameter Estimation Method

PPI Plan Position Indicator

SoC State of Charge

SoH State of Health

TLS Through Life Support

TMA Target Motion Analysis

UKF Unscented Kalman Filter