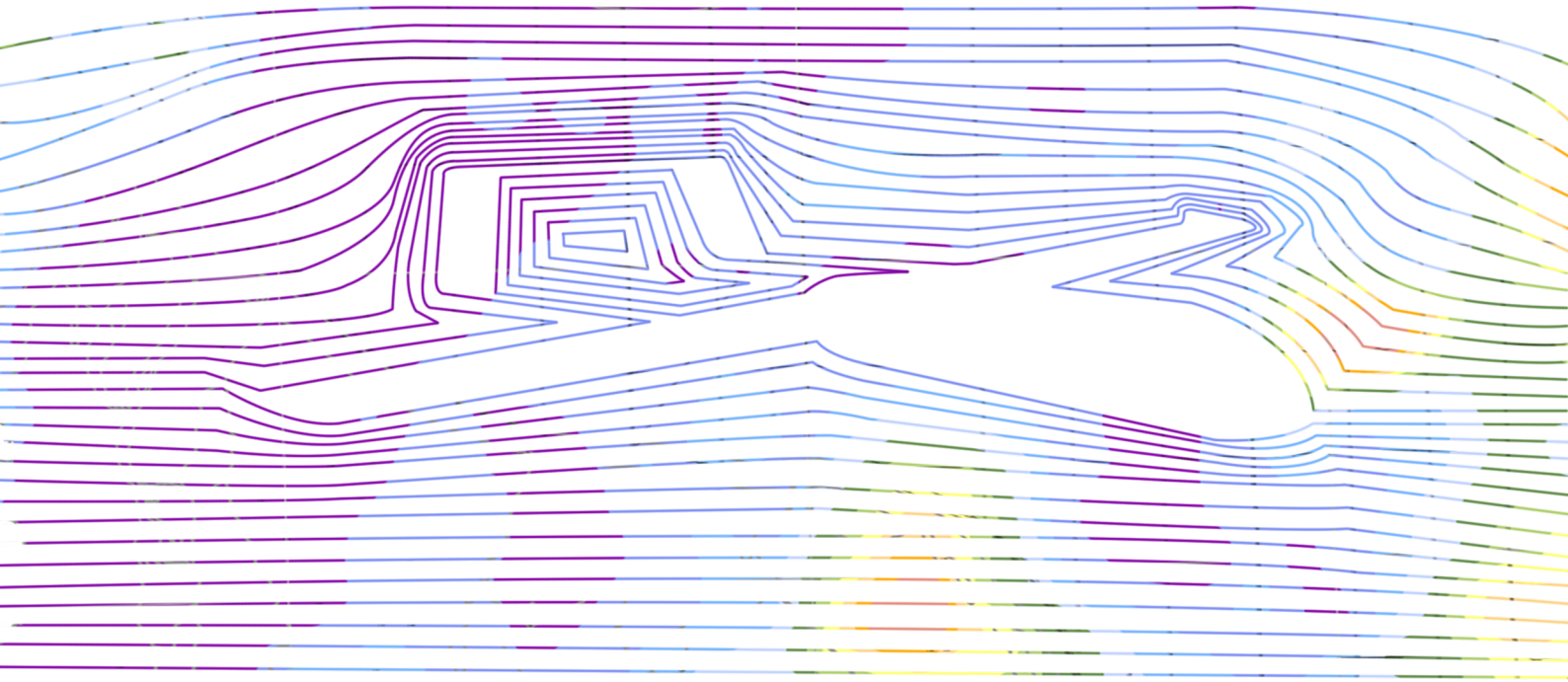
Reference-Free Aeromagnetic Platform Noise Compensation for Magnetic Aided Navigation

Fedde van der Meer



TU Delft Delft University of Technology

TNO

Reference-Free Aeromagnetic Platform Noise Compensation for Magnetic Aided Navigation

by

Fedde van der Meer

to obtain the degree of Master of Science

at Delft University of Technology,

to be defended publicly on Monday September 8, 2025 at 10:00 AM.

Student number:	4749693	
Project duration:	October 23, 2024 – September 8, 2025	
Thesis committee:	Dr. M. Jukić,	TNO, daily supervisor
	Dr. E. S. A. M. Lepelaars,	TNO, supervisor
	Dr. M. Kok,	TU Delft, supervisor
	Dr. K. M. Dowling,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

The work in this thesis was supported by TNO. Their cooperation is hereby gratefully acknowledged.

Delft Center for
Systems and Control

Summary

Magnetic-aided navigation using unmanned aerial vehicles (UAVs) is a promising method in case traditional navigation methods fail, but aeromagnetic platform noise from electric motors, electronics, and actuators can mask subtle geological signals. Traditional compensation methods, such as the Tolles-Lawson (TL) model, assume linear relationships between platform orientation and platform noise, failing to capture complex, time-varying disturbances from dynamic onboard systems. Existing machine learning approaches typically require noise-free reference measurements or known anomaly maps, resources often unavailable in practical surveying scenarios.

This thesis develops data-driven, reference-free methods for compensating platform noise in aeromagnetic measurements. The research addresses two key questions: whether deep learning methods can effectively predict and compensate platform noise without ground-truth references, and which drone subsystems contribute most significantly to platform noise.

A validation apporoach was implemented using flight data where crustal anomalies are naturally attenuated, enabling reference-free performance assessment. Comprehensive data from a fixed-wing UAV equipped with scalar and vector magnetometers, logging 273 platform-related input signals during figure-of-merit manoeuvres over an 800 m $\times$ 800 m survey area was used.

The compensation approach employed hierarchical modelling: Extended Tolles-Lawson (ETL) compensation incorporating drone inputs projected onto the magnetic field direction, followed by multilayer perceptron (MLP) neural networks trained to predict residuals. SHAP (SHapley Additive exPlanations) analysis provided model-agnostic feature importance assessment to identify the most influential platform inputs.

Results demonstrate substantial improvements over traditional methods. ETL compensation achieved improvement ratios averaging 7.31 for scalar magnetometers and 18.76 for vector magnetometers, compared to 4.90 and 10.19 respectively for standard TL compensation. The combined ETLNN approach (ETL + neural network) further enhanced performance to average improvement ratios of 8.87 on total field magnetometer and 22.22 on vector magnetometer data, a significant improvement over traditional TL methods.

SHAP analysis revealed that engine-related parameters (battery current, throttle commands), inertial measurement data (accelerations, gyroscopic rates, vibration), and attitude information (roll, pitch, yaw) are the primary contributors to platform noise. Features projected onto the magnetic field direction consistently outperformed raw inputs, validating the physical basis for this transformation, while derivative features contributed minimally whilst increasing overfitting.

The primary limitation is the inability to validate performance on data containing actual magnetic anomalies, as the high-altitude validation approach deliberately suppressed geological signals. Future work should prioritise validation using artificial magnetic sources or reference magnetometer configurations to assess preservation of genuine geological signals.

Preface

This thesis started with nothing more than a general idea of what I wanted to do with my time. I liked the idea of doing something academic, but did not see myself surviving nine months working on something purely theoretical. In TNO I found a great place with such interesting research, often practical but with an academic mindset. Having a practical application is hugely motivating, because you feel you are making a contribution, albeit very small, to developments that will have an impact in the not so distant future. After my first couple of meetings with Mia and Eugene I felt very lucky to have stumbled into a topic that suited me, in a workplace that contains some of the smartest people I have met who also really love what they are doing.

Still, being lucky about finding such an interesting topic in a great environment did not mean completing this thesis was easy. On the contrary, it has been one of the hardest things I have ever done. Part of the reason that I managed to finish it is my supervisor Mia, who was not only enjoyable to talk to when things were going well, but also delivered some great speeches and pep-talks when I was struggling. I really think you would make an excellent motivational speaker. I also want to thank Eugene, who who in every meeting seemed to impart a sliver of wisdom. And thanks to Manon, who even though and perhaps because we spoke less often, always managed to put the finger on the sore spot, which I think is one of the most valuable attributes you can have as a supervisor.

Most of all I want to thank Mayra, who is always there for me. Of everything that happened this year, spending time with you was my favourite part. Next to that I want to acknowledge you for creating the beautiful cover art.

Fedde van der Meer
Delft, August 2025

Contents

1

Introduction

Aerial navigation is a critical capability for both manned and unmanned aircraft, enabling precise movement and positioning [1]. Navigation systems allow aircraft to determine their position, velocity, altitude, and heading to safely traverse routes and avoid obstacles. Traditionally, aerial navigation is achieved using the global navigation satellite system (GNSS) together with an inertial navigation system (INS) [2]. GNSS is a satellite-based technology providing real-time geolocation and time information to users across the globe. GNSS has become ubiquitous in navigation due to its accuracy, reliability, and availability, making it the primary method for aerial, maritime, and terrestrial navigation [1].

However, the reliance on GNSS presents vulnerabilities that can jeopardise navigation accuracy and safety. GNSS signals, which are inherently weak and transmitted from satellites orbiting thousands of kilometres above Earth, are susceptible to interference [3], [4]. In recent years there has been a significant increase in GNSS jamming and spoofing, most notably around regions of conflict [5], [6]. GNSS jamming, using a signal to block or overwhelm the GNSS frequency, can prevent accurate positioning [3]. GNSS spoofing can manipulate the signal, misleading the navigation system with false data [3]. Both techniques have been demonstrated to disrupt or even completely deceive GNSS-dependent systems, posing significant risks, particularly for aerial navigation where precise positioning is essential for safety [3].

In response to these vulnerabilities, there is a growing interest in developing alternative navigation methods that are resistant to interference [6]. Alternative navigation systems aim to provide reliable positioning data without reliance on satellite signals, using techniques such as vision-based navigation [7], dead reckoning based on inertial measurement units (IMUs) [2], radio-based [1] and magnetic anomaly based localisation systems [8]. These technologies can potentially enhance the resilience of aerial navigation, enabling aircraft to operate in GNSS-denied environments and safeguarding them from intentional or accidental signal interference.A method for localisation that has showed promise recently is magnetic-aided navigation, also called magnetic anomaly navigation or MagNav. In the aerial domain the Earth's magnetic anomalies are used, which is called magnetic anomaly navigation. Magnetic anomaly navigation can be used to aid an inertial navigation system (INS). This method is especially valuable in GNSS-denied environments, such as during military operations or in remote regions where satellite signals are unreliable. Magnetic anomaly navigation, as demonstrated in the work of Canciani [9], offers an unjammable alternative to GNSS by utilising the Earth's magnetic field. The Earth field consists of the core field and the crustal field. The core field stems from convection currents in the liquid core of the Earth [10]. The much smaller crustal field stems from the magnetic rock in the lithosphere of the Earth [10]. The core field can be accurately modelled by, for instance, the international geomagnetic reference field (IGRF) model [11] which is the scientific standard [12]. The crustal field is usually not directly modelled. Instead, the anomaly field is used, which is the difference between the core field model and the Earth field measured by surveys. The anomaly field is used to generate magnetic anomaly maps, like the world digital magnetic anomaly map (WDMAM) [13]. These anomaly maps have varying degrees of accuracy in different locations depending on the quality and quantity of surveys at those locations.

The property that makes aerial navigation using magnetic anomaly maps so attractive is that it is practically unjammable. This is due to the fact that at sufficient distance, any magnetic field source will behave as a magnetic dipole, for which the magnitude decreases by the cube of the distance. Mathematically,

$$B \propto \frac{1}{r^3}, \tag{1.1}$$

where B is the magnitude of the magnetic field and r is the distance to the centre of the magnetic field [14].

One of the first to show aerial magnetic anomaly navigation was Canciani in [9]. Canciani's method relies on taking very accurate scalar measurements of the magnetic field and comparing them to a map. Canciani showed navigation with an accuracy in the range of 10's to 100's of metres distance root-mean square (DRMS). The error increases with altitude, because at higher altitude, less detail in the magnetic anomalies can be detected [9]. The navigation accuracy also depends on the accuracy of the magnetic anomaly map.

Magnetic-aided navigation has been done in other environments as well. An example of work that has been done in magnetic-aided navigation in an indoor environment is [15], which shows simultaneous-localisation-and-mapping (SLAM) using magnetic field maps. Mapping is done using the magnetic anomalies stemming from the ferromagnetic materials in the building. This approach has been directly applied by Lee and Canciani [16] to aerial navigation. The work by Lee and Canciani shows that there is benefit in surveying research from different domains for MagNav in the aerial domain.

1.1. Challenges in Platform Noise Compensation

The practical implementation of aerial magnetic anomaly navigation encounters significant challenges due to the corruption of magnetic measurements by platform-induced noise [9]. Magnetometers mounted on aircraft measure a combination of the Earth's magnetic field and magnetic interference generated by the aircraft itself, including engines, electronics, and moving parts. This platform noise blends with the Earth's natural magnetic signals, making accurate field recovery difficult. Canciani underscored the critical importance of effective platform noise compensation to address this issue [9], as limited platform noise compensation accuracy directly constrains navigation accuracy [17].

In Canciani's work, platform noise was assumed negligible after being effectively compensated using traditional methods. This assumption was valid because the magnetometer was mounted on a tail-stinger, isolating it from time-varying platform noise sources. The aircraft used by Canciani with a mounted tail-stinger can be seen in Figure 1.1. Under these conditions, traditional linear time-invariant compensation methods proved effective. However, when magnetometers cannot be positioned on tail-stingers, particularly on small UAV platforms, it becomes necessary to address time-varying platform noise from engines, electronics, and moving components [17], [18].

Traditional compensation methods, such as the Tolles-Lawson (TL) model, assume linear and time-invariant relationships between platform orientation and platform noise. While effective for steady-state effects, these methods fail to capture the complex, non-linear dynamics present in real aircraft systems. For example, control surface actuators draw current that varies non-linearly with wind resistance and angular position, and vibrations can cause current-carrying components to move, creating position-dependent magnetic fields. As demonstrated by Hezel [18], relationships exist between control surface inputs and the resulting platform noise. The system complexity, involving numerous interacting components with unknown dynamics, motivates the use of methods capable of modeling non-linear, time-varying relationships.

Recent advances in deep learning have shown promise for platform noise compensation by modeling these complex relationships between platform dynamics and platform noise [17]. However, existing deep learning approaches typically rely on the availability of noise-free reference measurements or accurate magnetic anomaly maps for training and validation. In practical applications, such ground truth data is often unavailable, particularly for surveys conducted over unmapped regions or using platforms without dedicated reference sensors.

Figure 1.1: Geosurvey aircraft similar to the one used by Canciani. Note the tail-stinger at the rear of the aircraft. At the tip of the tail-stinger there are magnetometers. Source: [19].

1.2. Problem Statement

To investigate platform noise compensation in a reference-free scenario, the Netherlands Organisation for Applied Scientific Research (TNO) performed experiments using a fixed-wing drone platform operating at 250 m altitude. This resulted in a dataset that presents representative challenges: it contains aeromagnetic measurements with no reliable anomaly map or reference magnetometer available. At this altitude, short-wavelength magnetic anomalies are significantly attenuated, resulting in a nearly constant background field. Consequently, compensation must be performed without ground truth references, and evaluation depends on assumptions about the local smoothness of the Earth's field.

The drone simultaneously records 273 inputs, including battery currents, control surface positions, inertial measurements, and attitude information. Jukić showed these inputs are correlated with magnetic measurements [20]. Furthermore, Jukić improved platform noise compensation by extending the traditional linear compensation model to include various drone inputs.

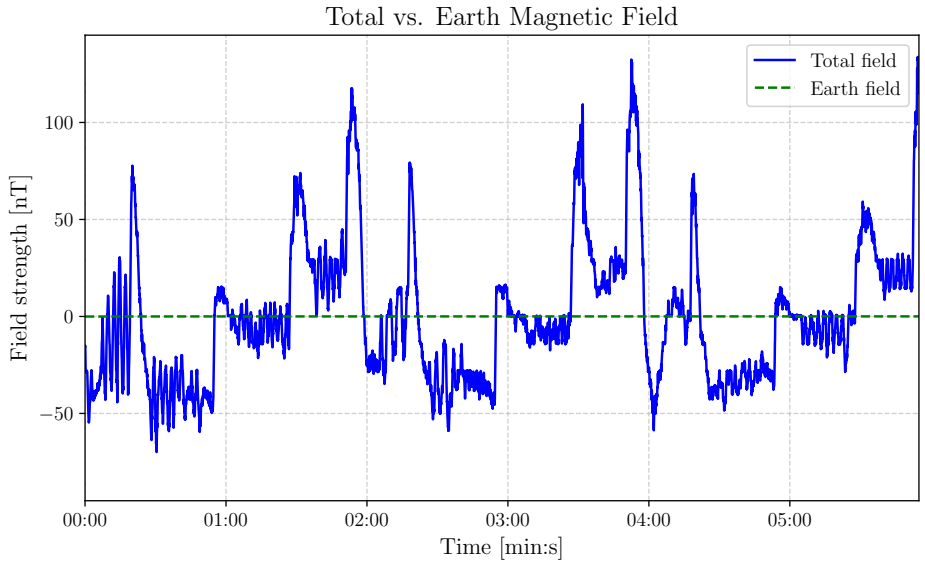


Figure 1.2: Snippet of the TNO measurements showing total magnetic field measured by the OPM magnetometer and the theoretical Earth field as given by the IGRF model at the GNSS position. Note that the magnetic field data has been demeaned.

To summarise, the goal of aeromagnetic noise compensation is to remove platform noise from measurements, like the total field in Figure 1.2, so that the Earth's magnetic field, shown in the same figure, can be recovered. This thesis addresses two key limitations of existing machine learning-based compensation methods:

- **Dependence on ground truth**: Supervised learning methods typically rely on noise-free references or prior anomaly maps for training and validation. This is unrealistic for many practical surveying operations, especially over unmapped terrain or on UAV platforms without dual-magnetometer configurations.
- **Input relevance**: With hundreds of available drone inputs, it is difficult to determine which signals significantly contribute to platform-induced noise. Existing models do not provide a systematic way to identify or quantify the importance of individual inputs across different flight conditions.

1.2.1. Research Questions

Given these challenges, this thesis investigates the following research questions:

- **RQ1: Reference-free noise compensation**
  Can deep learning models effectively predict and compensate aeromagnetic platform noise in the absence of a noise-free reference signal or accurate anomaly map?
- **RQ2: Input relevance and feature selection**
  Which drone inputs are most relevant for predicting platform noise, and how can their contribution to the compensation model be quantified across different manoeuvres and flights?

1.2.2. Thesis Outline

This thesis is divided into five chapters. Chapter 2 covers background information on Earth's magnetic field, platform noise modelling, platform noise estimation methods, and feature selection techniques. Chapter 3 defines the problem and analyses the TNO dataset. Chapter 4 presents the theoretical basis for approximations used in literature, the methodology for platform noise estimation and feature selection, and the main results. Chapter 5 discusses the results, draws conclusions, and suggests future work.

2

Background

This chapter covers the basic concepts and methods used in traditional magnetic-aided aerial navigation, focusing on how platform noise is modelled and reduced. We start by looking at Earth's magnetic field and how its different parts (the core field, crustal field, and diurnal field) create the magnetic environment that aircraft encounter.

Next, we develop the framework for measuring magnetic fields and detecting anomalies, introducing the scalar magnetometer approximations used in most navigation systems. The key part of this discussion is the Tolles-Lawson model, which has been basis for compensating platform noise since the 1950s. We explain how this model works and how to estimate its coefficients, including the problems that come up like multicollinearity and the need for specific flight patterns to get good data.

We then look at extensions to the basic Tolles-Lawson model. These include ways to account for onboard electronics using Biot-Savart law calculations, and linear methods that try to link leftover magnetic signals with aircraft control inputs. We discuss both how these extended models work and where they fall short.

The final part of the chapter introduces ideas from physics-informed machine learning and feature engineering that will be useful for the more advanced methods in later chapters. Feature selection techniques like correlation analysis and SHAP (Shapley Additive Explanations) values will help to understand which input variables matter most for predicting platform noise.

2.1. The Earth's Magnetic Field

The Earth's magnetic field is a complex superposition of several sources, each with distinct spatial and temporal characteristics. It is typically decomposed into three main components: the core field, the crustal field, and external fields such as the diurnal variation.

2.1.1. Core Field

The core field is the dominant contribution to the Earth's magnetic field and originates from the motion of molten iron within the outer core, a process known as the geodynamo. This field is largely dipolar at the Earth's surface [11]. It is relatively stable over short timescales but undergoes slow changes over years to centuries. The core field is around 20000 to 70000 nT in magnitude. Global models such as the International Geomagnetic Reference Field (IGRF) model the large-scale structures of the Earth field, which are dominated by the core field [11]. Since magnetic-aided navigation uses small-scale structures, the approximation is made that the IGRF model can be used to describe and predict core field.

2.1.2. Crustal Field and Magnetic Anomalies

Superimposed on the core field is the crustal field, which arises from the magnetization of rocks in the Earth's lithosphere. This field reflects the geological composition and structural features of the crust and

varies over much shorter spatial scales, ranging from a few meters to hundreds of kilometers. Local concentrations of ferromagnetic minerals can produce magnetic anomalies that deviate significantly from the smoother core field.

These deviations are often visualised in magnetic anomaly maps. An anomaly is defined as the difference between the observed magnetic field magnitude and the core field magnitude predicted by a model, usually the IGRF. The order of magnitude of magnetic anomalies is around 10's to 100's of nT. This magnitude depends both on the strength of the anomaly field and the distance from the surface of the Earth. Figure 2.1 shows such a map for the Netherlands and the North Sea, where magnetic anomalies reveal underlying crustal structures.

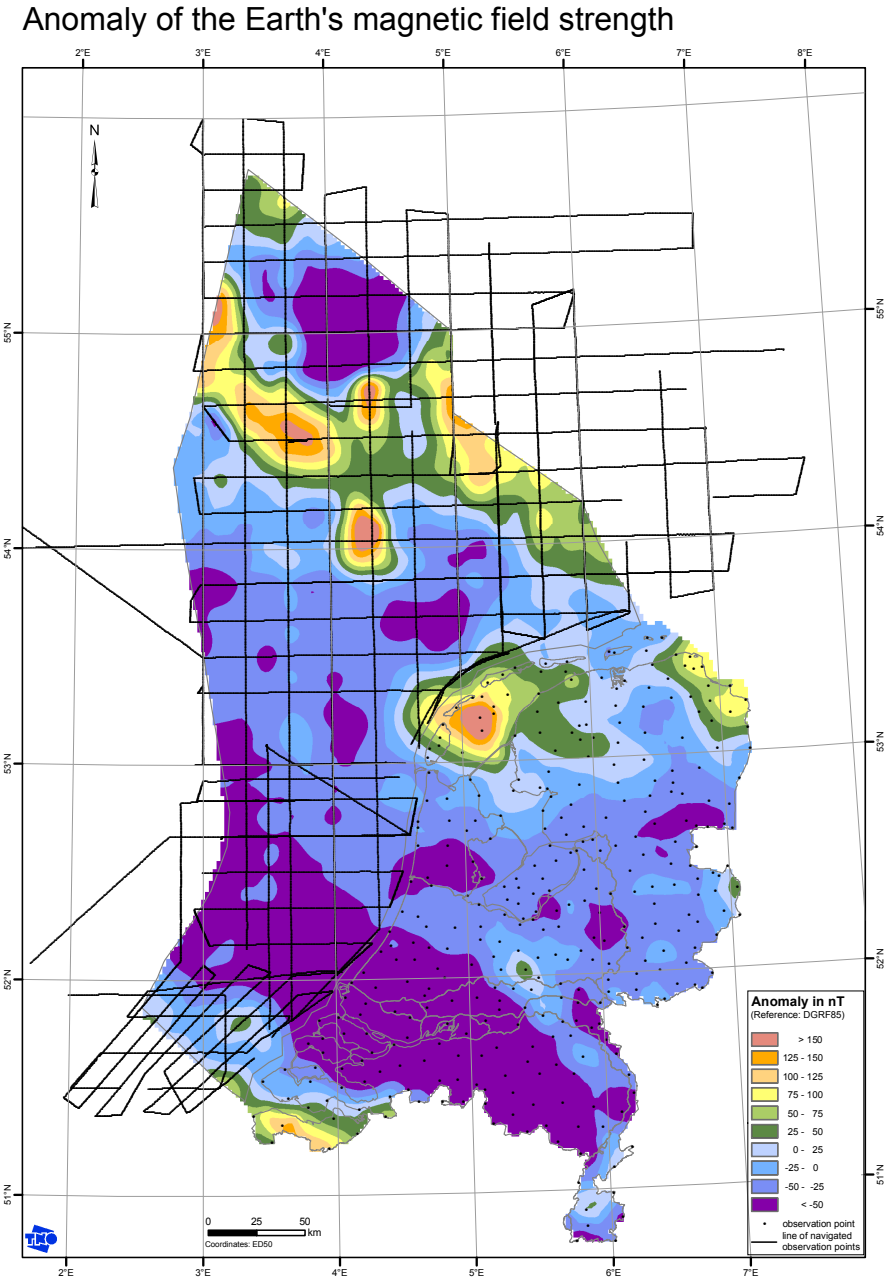


Figure 2.1: Magnetic anomaly in the Netherlands and North Sea. From **nlog_magnetic_gravity**.

2.1.3. Diurnal Variation

In addition to internal sources, the Earth's magnetic field is influenced by external currents in the ionosphere and magnetosphere, which vary on timescales ranging from seconds to days. The most regular of these is the diurnal variation, caused by the interaction between solar radiation and the Earth's atmosphere. Solar heating drives ionospheric currents that induce magnetic variations at the Earth's surface with a period of approximately 24 hours. While typically small in magnitude (on the order of tens of nanotesla), these variations can be significant in some cases. [21]. An extensive discussion about the diurnal variation can be found in [9].

2.1.4. Magnetic field model

The magnetic field at a fixed point in the platform reference frame at time t can be described as

$$\mathbf{B}_t(t) = \mathbf{B}_c(t) + \mathbf{B}_l(t) + \mathbf{B}_p(t), \tag{2.1}$$

where $\mathbf{B}_{\{c,l,p,t\}} \in \mathbb{R}^3$ represent the core, crustal, platform, and total fields respectively. Note that all these variables are defined in the platform reference frame and the sensor is stationary with respect to this reference frame. The diurnal variation is assumed to be negligible. Diurnal variation can be assumed to be negligible in the case the duration of the measurements is smaller than the timescale at which diurnal variation takes place. For longer duration or more precise measurements, diurnal variation needs to be taken into account.

The platform reference frame is a coordinate system fixed to the platform on which the magnetic field sensor is mounted, in this case a type of fixed-wing drone. This frame moves rigidly with the platform, meaning that any point within the frame remains stationary relative to the platform itself. The axes of the platform reference frame are aligned with the physical orientation of the platform (e.g., forward, lateral, and vertical directions). Since the sensor is considered stationary with respect to this frame, all magnetic field measurements ($\mathbf{B}_c$, $\mathbf{B}_l$, $\mathbf{B}_p$) are expressed in this reference frame, ensuring consistency in the description of the magnetic field dynamics.

For simplicity of notation, the core field $\mathbf{B}_c$ and the crustal field $\mathbf{B}_l$ are combined into a single variable referred to as the Earth field

$$\mathbf{B}_e(t) = \mathbf{B}_c(t) + \mathbf{B}_l(t). \tag{2.2}$$

By substituting $\mathbf{B}_e(t)$ from (2.2) into (2.1), the total magnetic field becomes

$$\mathbf{B}_t(t) = \mathbf{B}_e(t) + \mathbf{B}_p(t), \tag{2.3}$$

$$\tag{2.4}$$

where $\mathbf{B}_t(t)$ is expressed as the sum of the Earth field $\mathbf{B}_e(t)$ and the platform field $\mathbf{B}_p(t)$.

In literature most commonly used for magnetic navigation are scalar magnetic field values [9], [22]. Next to scalar versus vector magnetic field sensor sensitivity, there are some other factors that make scalar magnetic fields more suitable for magnetic navigation which are discussed in [22]. That is why the quantity that is given in anomaly field maps is not a vector field, but a scalar field. The anomaly field, like mapped in Figure 2.1, is the magnitude of the Earth field minus the magnitude of the core field,

$$B_a = |\mathbf{B}_e| - |\mathbf{B}_c|. \tag{2.5}$$

Note that B_a is not the same as the magnitude of the crustal field,

$$|\mathbf{B}_l| = |\mathbf{B}_e - \mathbf{B}_c| \neq B_a = |\mathbf{B}_e| - |\mathbf{B}_c|. \tag{2.6}$$

As is explained by Canciani in [9], this quantity B_a is an approximation of the projection of the crustal field onto the Earth field. Canciani notes this approximation is only valid because $B_e \gg B_l$.

The core field magnitude $|\mathbf{B}_c|$ can be found using either a high-pass filter or values obtained from the IGRF model [9], [11]. Note that values of the IGRF model can only be obtained if the location is known. However, since the core field changes slowly across space, it can be approximated well by using an approximate position [9]. After the core field is accounted for, the only variable left unaccounted for in Equation (2.1) is the platform field $\mathbf{B}_p$.

Estimating the platform field $\mathbf{B}_p$ in (2.3) has traditionally been done using the Tolles-Lawson model and extensions thereof. In these models, which will be referred to as the 'traditional' platform noise compensation, certain assumptions and approximations are (often implicitly) made. In this thesis an effort will be made to state all these assumptions and approximations explicitly and investigate their efficacy. For ease of notation the dependence on time t will not always be explicitly noted.

2.1.5. Scalar Magnetometer Measurements and Approximations

As mentioned in Section 2.1.4, scalar magnetometer measurements are used for calculating the anomaly field. However, as will be seen, the vector direction of the magnetic field is still necessary and so the vector magnetic field $\mathbf{B}_t$ is measured using the less sensitive vector magnetometer. The total magnetic field is modelled as

$$B_t = |\mathbf{B}_t| = |\mathbf{B}_e + \mathbf{B}_p|. \tag{2.7}$$

For ease of notation the magnitude of a vector $\mathbf{B}$, given by $|\mathbf{B}|$, will also be given by B. Using the fact that for a vector $\boldsymbol{v}$, $v = |\boldsymbol{v}| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$, (2.7) can be written as

$$\begin{aligned}
|\mathbf{B}_t| &= \sqrt{(\mathbf{B}_e + \mathbf{B}_p) \cdot (\mathbf{B}_e + \mathbf{B}_p)} \\
&= \sqrt{B_e^2 + B_p^2 + 2\mathbf{B}_e \cdot \mathbf{B}_p} \\
&= B_e \sqrt{1 + \frac{B_p^2}{B_e^2} + 2\frac{\mathbf{B}_p \cdot \mathbf{B}_e}{B_e^2}} \\
&= B_e \sqrt{1 + \left(\frac{B_p}{B_e}\right)^2 + 2\frac{B_p}{B_e}\cos\theta}.
\end{aligned} \tag{2.8}$$

Where θ is the angle between $\mathbf{B}_e$ and $\mathbf{B}_p$. Let $u = \frac{B_p}{B_e}$. Now because $B_e \gg B_p$, (2.8) is approximated by an expansion around $u = 0$. This approximation is done by a first-order Maclaurin series expansion $M(u)$ of $f(u)$, where

$$f(u) = B_t = B_e \sqrt{1 + u^2 + 2u\cos\theta},$$

as follows,

$$\begin{aligned}
M(u) &= f(0) + \frac{d}{du} f(u)\bigg|_{u=0} u + \epsilon_{HO} \\
&= B_e + \frac{B_e\,(u + \cos\theta)}{\sqrt{1 + u^2 + 2u\cos\theta}}\bigg|_{u=0} u + \epsilon_{HO} \\
&= B_e + uB_e\cos\theta + \epsilon_{HO} \\
&= B_e + \frac{B_p}{B_e} B_e \cos\theta + \epsilon_{HO} \\
&= B_e + \mathbf{B}_p \cdot \frac{\mathbf{B}_e}{B_e} + \epsilon_{HO}.
\end{aligned}$$

Assuming the higher order terms are negligible, i.e. $\epsilon_{HO} \approx 0$,

$$B_t \approx B_e + \mathbf{B}_p \cdot \frac{\mathbf{B}_e}{B_e}. \tag{2.9}$$

This expression is useful as it can be used to determine the magnitude of the Earth field B_e, and consequently the anomaly field B_a [9]. But before (2.9) can be used for determining the anomaly field B_a, first the term $\mathbf{B}_p \cdot \frac{\mathbf{B}_e}{B_e}$ must be determined. As $\frac{\mathbf{B}_e}{B_e}$ is unknown, some further steps are required.

The direction of the Earth field, $\frac{\mathbf{B}_e}{B_e}$, must be approximated as it is unknown. Because $B_e \gg B_p$, the total field is approximately equal to the Earth field, $\mathbf{B}_e \approx \mathbf{B}_t$. The direction of the Earth field is therefore approximated by the direction of the total field,

$$\frac{\mathbf{B}_e}{B_e} \approx \frac{\mathbf{B}_t}{B_t}. \tag{2.10}$$

Filling (2.10) into (2.9) yields

$$B_e \approx B_t - \mathbf{B}_p \cdot \frac{\mathbf{B}_t}{B_t}, \tag{2.11}$$

and in turn the anomaly field is

$$B_a \approx B_t - B_c - \mathbf{B}_p \cdot \frac{\mathbf{B}_t}{B_t}. \tag{2.12}$$

Even though in (2.11) B_t is measured by the scalar magnetometer, to determine B_e we still require knowledge of the vector direction of the total field $\frac{\mathbf{B}_t}{B_t}$. That is why a vector magnetometer is used to measure $\frac{\mathbf{B}_t}{B_t}$. The estimation of B_e using both the scalar and vector magnetometer will be discussed in section 2.2.1.

For completeness, another way of obtaining (2.11) is included here. This is the method in [23]. This derivation is done by rewriting (2.8) as

$$g(v) = B_e = B_t\sqrt{1 + \left(\frac{B_p}{B_t}\right)^2 - 2\frac{B_p}{B_t}\cos\phi},$$

where ϕ is the angle between $\mathbf{B}_t$ and $\mathbf{B}_p$ and performing a Maclaurin series expansion of $g(v)$, where $v = \frac{B_p}{B_t}$. This expansion around $v = 0$ is done because if $B_e \gg B_p$, then $|B_e + B_p| \approx B_e$, and $v \approx \frac{B_p}{B_e} \approx 0$.

2.2. Platform Noise Modeling with the Tolles-Lawson Model

Now that we have a model that can be used to estimate the magnitude of the Earth field (2.11), and consequently anomaly field B_a in (2.12), our attention turns to the modelling and estimation of the platform noise $\mathbf{B}_p$. From the Earth field magnitude B_e the anomaly field B_a can be calculated as in (2.5).

Research on modelling aeromagnetic platform noise dates back to the 1940s, starting with the model by Tolles and Lawson [24]. The Tolles-Lawson model is the basis of traditional aeromagnetic platform noise compensation techniques. It represents the platform field as the sum of three sources:

$$\mathbf{B}_p = \mathbf{B}_{perm} + \mathbf{B}_{ind} + \mathbf{B}_{eddy}, \tag{2.13}$$

where $\mathbf{B}_{perm}$, $\mathbf{B}_{ind}$, and $\mathbf{B}_{eddy}$ represent the permanent, induced, and eddy current magnetic fields, respectively.

A derivation of the permanent, induced and eddy-current magnetic fields is given in [23]. Here, an intuitive explanation of each term is given.

Permanent Magnetic Field ($\mathbf{B}_{perm}$) This field arises from the permanent magnetization of ferromagnetic materials present in the platform. These materials have magnetic fields that remain fixed after becoming magnetised, even without an external magnetic field. This creates a constant magnetic contribution to the platform's overall field.

Induced Magnetic Field ($\mathbf{B}_{ind}$) The induced magnetic field is generated when ferromagnetic materials in the platform interact with the Earth's magnetic field. These materials become temporarily magnetised in alignment with the Earth field, producing an additional magnetic field that is linear with the Earth field as defined relative to the platform.

Eddy-Current Magnetic Field ($\mathbf{B}_{eddy}$) This field is a result of Faraday's law of induction. When the platform moves through or experiences a changing magnetic field, it induces electric currents (eddy currents) within the conductive materials of the platform. These circulating currents, in turn, generate a magnetic field that opposes the change, as dictated by Lenz's law.

As done in [23], the Tolles-Lawson model in (2.13) can be parameterized as

$$\mathbf{B}_p = \boldsymbol{a} + \mathbf{b}\mathbf{B}_e + \mathbf{c}\dot{\mathbf{B}}_e, \tag{2.14}$$

where $\boldsymbol{a}$, $\mathbf{b}\mathbf{B}_e$ and $\mathbf{c}\dot{\mathbf{B}}_e$ correspond to the permanent, induced, and eddy-current terms respectively. The notation $\dot{(\cdot)}$ represents the time-derivative.

Since $\mathbf{B}_e$ is not known this is approximated by $\mathbf{B}_t$ [25], leading to

$$\mathbf{B}_p \approx \boldsymbol{a} + \mathbf{b}\mathbf{B}_t + \mathbf{c}\dot{\mathbf{B}}_t. \tag{2.15}$$

The coefficients $\boldsymbol{a}$, $\mathbf{b}$, and $\mathbf{c}$ are defined as follows,

$$\boldsymbol{a} = \begin{bmatrix}\beta_1 & \beta_2 & \beta_3\end{bmatrix}^T, \quad \mathbf{b} = \frac{1}{2}\begin{bmatrix}2\beta_4 & \beta_5 & \beta_6 \\ \beta_5 & 2\beta_7 & \beta_8 \\ \beta_6 & \beta_8 & 2\beta_9\end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix}\beta_{10} & \beta_{11} & \beta_{12} \\ \beta_{13} & \beta_{14} & \beta_{15} \\ \beta_{16} & \beta_{17} & \beta_{18}\end{bmatrix}.$$

The elements in $\mathbf{b}$ are chosen for convenience in notation later on.

Filling in the above (2.15) in (2.11) results in

$$B_e \approx B_t - (\boldsymbol{a} + \mathbf{b}\mathbf{B}_t + \mathbf{c}\dot{\mathbf{B}}_t) \cdot \frac{\mathbf{B}_t}{B_t}, \tag{2.16}$$

which can be rewritten as

$$B_e \approx B_t - (\boldsymbol{a}^T\frac{\mathbf{B}_t}{B_t} + \mathbf{B}_t^T\mathbf{b}^T\frac{\mathbf{B}_t}{B_t} + \dot{\mathbf{B}}_t^T\mathbf{c}^T\frac{\mathbf{B}_t}{B_t}). \tag{2.17}$$

This can then be written as a linear system of equations

$$B_t - B_e \approx \boldsymbol{\delta}^T\boldsymbol{\beta}, \tag{2.18}$$

where

$$\boldsymbol{\beta} = \begin{bmatrix}\beta_1 & \dots & \beta_{18}\end{bmatrix}^T,$$

$$\boldsymbol{\delta} = \frac{1}{B_t}\begin{bmatrix}\mathbf{B}_t^T & \text{vec}(\text{utri}(\mathbf{B}_t\mathbf{B}_t^T))^T & \text{vec}(\mathbf{B}_t\dot{\mathbf{B}}_t^T)^T\end{bmatrix}^T.$$

Here vec($\mathbf{B}\mathbf{B}^T$) indicates a vectorization of matrix $\mathbf{B}\mathbf{B}^T$ and utri($\mathbf{B}\mathbf{B}^T$) selects only the upper triangular part of matrix $\mathbf{B}\mathbf{B}^T$. Now if the coefficient vector $\boldsymbol{\beta}$ is known, the Earth field magnitude can be estimated as

$$B_e \approx B_t - \boldsymbol{\delta}^T\boldsymbol{\beta}. \tag{2.19}$$

2.2.1. Estimation of Tolles-Lawson Coefficients

Using the TL model, we can try to estimate the TL coefficients β. In [26] Leliak determined a way of calculating the TL coefficients [26]. Let us consider the magnetic fields sampled across time from sample $n = 1, \dots, n = N$, where N is the number of samples. The platform coefficients $\boldsymbol{\beta}$ can be calculated by removing the low-frequency B_e component from the data by high-pass filtering. There is also some high-frequency random noise in B_t [17] that can be removed with a low-pass filter. The high- and low-pass filters are combined in a band-pass filter, bpf(). Let

$$\boldsymbol{y} = \begin{bmatrix}|\mathbf{B}_t(1)| & |\mathbf{B}_t(2)| & \dots & |\mathbf{B}_t(N)|\end{bmatrix}^T, \tag{2.20a}$$

$$\mathbf{c}_{\mathbf{B}_e} = \begin{bmatrix}|\mathbf{B}_e(1)| & |\mathbf{B}_e(2)| & \dots & |\mathbf{B}_e(N)|\end{bmatrix}^T, \tag{2.20b}$$

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{\delta}(1) & \boldsymbol{\delta}(2) & \dots & \boldsymbol{\delta}(N)\end{bmatrix}^T. \tag{2.20c}$$

Because the bandpass filter removes all low frequency components and the Earth field contains only low-frequency components, $\text{bpf}(\mathbf{c}_{\mathbf{B}_e}) = \mathbf{0}$. Then

$$\text{bpf}(\boldsymbol{y} - \mathbf{c}_{\mathbf{B}_e}) = \text{bpf}(\boldsymbol{A})\boldsymbol{\beta}, \tag{2.21}$$

$$\text{bpf}(\boldsymbol{y}) = \text{bpf}(\boldsymbol{A})\boldsymbol{\beta}. \tag{2.22}$$

To find the coefficients a least-squares problem can be posed as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||\text{bpf}(\boldsymbol{y}) - \text{bpf}(\boldsymbol{A})\boldsymbol{\beta}||_2^2, \tag{2.23}$$

where the least-squares optimal solution for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \text{bpf}(\boldsymbol{A})^{\dagger}\text{bpf}(\boldsymbol{y}), \tag{2.24}$$

with $(\cdot)^{\dagger}$ denoting the Moore-Penrose pseudo-inverse.

2.2.2. Determining TL Coefficients using the Core Field Model

If the platform is flown at high enough altitude, the anomaly field becomes negligible $B_a \approx 0$. This is due to the cubic decay with distance in (1.1). Filling in $B_a \approx 0$ in (2.5), the Earth field can be approximated by the core field $\mathbf{B}_e \approx \mathbf{B}_c$. Then this result can be used in (2.14) which results in

$$\mathbf{B}_p = \boldsymbol{a} + \mathbf{b}\mathbf{B}_c + \mathbf{c}\dot{\mathbf{B}}_c. \tag{2.25}$$

The core field $\mathbf{B}_c$ can be determined by using the IGRF [11] model for the core field. Then by numerical approximation, $\dot{\mathbf{B}}_c$ can be calculated. Using this information (2.17) becomes

$$|\mathbf{B}_c| \approx B_t - (\boldsymbol{a}^T\frac{\mathbf{B}_t}{B_t} + \mathbf{B}_c^T\mathbf{b}^T\frac{\mathbf{B}_t}{B_t} + \dot{\mathbf{B}}_c^T\mathbf{c}^T\frac{\mathbf{B}_t}{B_t}). \tag{2.26}$$

Then by parameterizing in the same way as in subsection 2.2.1, a least-squares solution can be found as

$$\hat{\boldsymbol{\beta}}_c = \boldsymbol{A}_c^{\dagger}(\boldsymbol{y} - \boldsymbol{c}_{\mathbf{B}_c}). \tag{2.27}$$

With

$$\boldsymbol{c}_{\mathbf{B}_c} = \begin{bmatrix} |\mathbf{B}_c(1)| & |\mathbf{B}_c(2)| & \dots & |\mathbf{B}_c(N)| \end{bmatrix}^T, \tag{2.28}$$

$$\boldsymbol{A}_c = \begin{bmatrix} \boldsymbol{\delta}_c(1) & \boldsymbol{\delta}_c(2) & \dots & \boldsymbol{\delta}_c(N) \end{bmatrix}^T, \tag{2.29}$$

$$\boldsymbol{\delta}_c = \frac{1}{B_t}\begin{bmatrix} \mathbf{B}_t^T & \text{vec}(\mathbf{B}_c\mathbf{B}_t^T)^T & \text{vec}(\dot{\mathbf{B}}_c\mathbf{B}_t^T)^T \end{bmatrix}^T, \tag{2.30}$$

and $\boldsymbol{y}$ defined as in (2.20a). This method will be referred to as map-based estimatation of the TL coefficients, since one needs a map of the core field to apply it. Note that now there are 21 TL coefficients due to $\mathbf{B}_t\mathbf{B}_t^T$ in (2.17), which is symmetric, being replaced with $\mathbf{B}_c\mathbf{B}_t^T$ which is not symmetric.

2.2.3. Multicollinearity

In [26] Leliak found that there is high multicollinearity of the matrix $\boldsymbol{A}$ in Equation (2.20c). This means that the columns are very nearly linearly dependent. This gives a problem when solving the least-squares problem [27]. The solution involves the inverse of $\boldsymbol{A}^T\boldsymbol{A}$. If the columns of $\boldsymbol{A}$ are linearly dependent, this inversion is not possible. If the columns are nearly linearly dependent, then the result is subject to large numerical errors. This causes the TL coefficients which are found by this least-squares solution to vary heavily under small perturbations of the columns of $\boldsymbol{A}$, like when there is a slight measurement error.

There are several ways to cope with this ill-conditioned matrix $\boldsymbol{A}$. Leach investigates multiple methods of dealing with multicollinearity [27]. A common way in literature is to use ridge regression. Ridge

regression, also called Tikhonov regularization, adjusts the least-squares problem in (2.23) by adding a regularization term,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||\text{bpf}(\boldsymbol{y}) - \text{bpf}(\boldsymbol{A})\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_2^2. \tag{2.31}$$

The optimal solution to this problem is given by

$$\hat{\boldsymbol{\beta}} = (\text{bpf}(\boldsymbol{A})^T\text{bpf}(\boldsymbol{A}) + \lambda\boldsymbol{I})^{-1}\text{bpf}(\boldsymbol{A})^T\text{bpf}(\boldsymbol{y}). \tag{2.32}$$

This kind of regularization introduces an extra hyperparameter λ.

The matrix $\boldsymbol{A}$ may contain features with very different scales and physical units, which can affect the behavior of the regularization term. Specifically, features with smaller numerical values tend to produce larger regression coefficients, leading to disproportionate penalization. To mitigate this, the columns of $\boldsymbol{A}$ are scaled by a vector of constants, resulting in a scaled matrix

$$\boldsymbol{A}_s = \frac{\boldsymbol{A}}{\mathbf{s}},$$

where the division is performed elementwise across columns, and $\mathbf{s}$ is a vector of scaling constants. These constants are chosen to have the inverse units of their respective features, rendering the scaled matrix $\boldsymbol{A}_s$ unitless. As a result, the coefficients $\boldsymbol{\beta}_s$ estimated from the scaled system will have the same units as the output variable, which in this case is nanotesla (nT).

The ridge regression problem is then solved on the scaled system

$$\boldsymbol{\beta}_s = \left(\text{bpf}(\boldsymbol{A}_s)^T\text{bpf}(\boldsymbol{A}_s) + \lambda\boldsymbol{I}\right)^{-1}\text{bpf}(\boldsymbol{A}_s)^T\boldsymbol{y},$$

yielding the predicted output

$$\hat{\boldsymbol{y}} = \boldsymbol{A}_s\boldsymbol{\beta}_s.$$

To recover physically meaningful coefficients, the scaling is reversed. Noting that

$$\hat{\boldsymbol{y}} = \boldsymbol{A}_s\boldsymbol{\beta}_s = \left(\frac{\boldsymbol{A}}{\mathbf{s}}\right)\boldsymbol{\beta}_s = \boldsymbol{A}\left(\frac{\boldsymbol{\beta}_s}{\mathbf{s}}\right),$$

we define the final coefficients as:

$$\boldsymbol{\beta} = \frac{\boldsymbol{\beta}_s}{\mathbf{s}},$$

with elementwise division. These final coefficients $\boldsymbol{\beta}$ have units such that when multiplied with the unscaled features in $\boldsymbol{A}$, the result is in units of nanoTesla.

2.2.4. Flight Patterns for System Excitation

Accurate estimation of the Tolles-Lawson (TL) model coefficients relies heavily on obtaining sufficiently rich data to ensure parameter identifiability. This can be achieved through deliberate design of flight patterns that excite the system across a broad range of its operating conditions. Such flight patterns are analogous to input signals in system identification, where a sufficiently diverse input spectrum enhances the estimation of unknown parameters [28].

In the context of aeromagnetic surveys, platform manoeuvres are designed to induce measurable variations in the permanent, induced, and eddy-current magnetic field components as described in (2.14). The objective is to expose the platform to changes in the Earth's magnetic field $\mathbf{B}_e$, its time derivative $\dot{\mathbf{B}}_e$, and the platform's relative orientation. Specifically:

- Permanent Field Excitation ($\boldsymbol{a}$): To observe the static contribution of the permanent magnetic field, the platform should fly a trajectory with varying orientations. This ensures that the platform's magnetic moment interacts differently with the Earth's field over time.

- Induced Field Excitation ($\mathbf{bB}_e$): The induced magnetic field depends on the Earth's field strength and its projection onto the platform frame. To excite this term, the platform must perform manoeuvres such as pitch, roll, and yaw, which alter the relative alignment of the Earth's field vector $\mathbf{B}_e$ with the platform frame.
- Eddy-Current Field Excitation ($\mathbf{c\dot{B}}_e$): The eddy-current field is a dynamic effect arising from time-varying magnetic flux. Changes in roll, pitch and yaw, which cause changes in $\dot{\mathbf{B}}_e$, are required to sufficiently excite this component.

In practice, well-designed flight patterns include periodic manoeuvres, such as oscillatory pitch, roll, or heading changes, to provide persistent excitation of the system. These manoeuvres ensure that the platform's magnetic response is rich enough such that $\boldsymbol{\delta}$ in (2.15) contains linearly independent observations, mitigating issues related to multicollinearity discussed in Section 2.2.3. On the other hand, a poor flight pattern, lacking sufficient variation in platform motion, results in near-collinear regressors. This leads to ill-conditioned least-squares problems and unreliable coefficient estimates [27].

Therefore, the design of flight patterns should balance two competing objectives: (1) ensuring sufficient excitation for parameter estimation, and (2) adhering to operational constraints, such as altitude stability, safety, and survey area coverage [26].

Leliak first describes a way to excite the system using a flight pattern. These flight patterns consist of sinusoidal variation in roll, pitch and yaw. These manoeuvres should be done at specific frequencies, such that they are not removed during bandpass filtering [26]. Having sufficient variation in the roll, pitch and yaw is essential for the identifiability of the different elements of the TL model [26].

Commonly used as an input for system identification and validation is the "Figure-of-Merit" (FoM) flight pattern [29]. In each of the four legs of the square, a sinusoidal pitch, roll and yaw manoeuvre with amplitudes of $20°$, $10°$ and $10°$ respectively is flown. This must be done in an area of low geomagnetic gradient, so the Earth field can be properly filtered out. This square flight pattern can be seen in Figure 2.2. The FoM pattern is used in geosurveying to calculate the Figure-of-Merit, a number that gives some indication of how well the platform noise is compensated. This calculation is described by Noriega in [29]. Noriega also states that the FoM pattern "does not completely represent the range of aircraft manoeuvring normally anticipated". In system identification terms, it can be questioned whether the FoM pattern persistently excites the system. An input to a system is persistently exciting if the knowledge of the response of the system to that input, allows prediction of the response of the system to any input [28].

A flight pattern designed to validate that the identified system has no "heading effect" is the clover-leaf flight pattern [30], which can be seen in Figure 2.3. This pattern flies over the same point from four orthogonal headings. This allows for validation of the platform noise compensation model, because at this point the Earth field B_e should have the same value every time, regardless of heading [30].

2.3. Extensions of the Tolles-Lawson Model

This section explores improvements to the standard Tolles-Lawson (TL) model to better handle platform noise, particularly in cases involving onboard electronics and additional inputs.

The first approach extends the TL model to include noise generated by onboard electronics, modeled using the Biot-Savart law. This accounts for magnetic fields induced by currents, which in turn can also generate their own eddy-current fields.

Further refinements use correlations between residual magnetic signals and drone control inputs to improve compensation. These methods expand the model to incorporate more variables, addressing limitations of the standard TL approach.

Finally, pre-processing techniques and adjustments to the loss function are introduced to mitigate issues like multicollinearity and low-frequency bias, enabling more robust noise reduction.

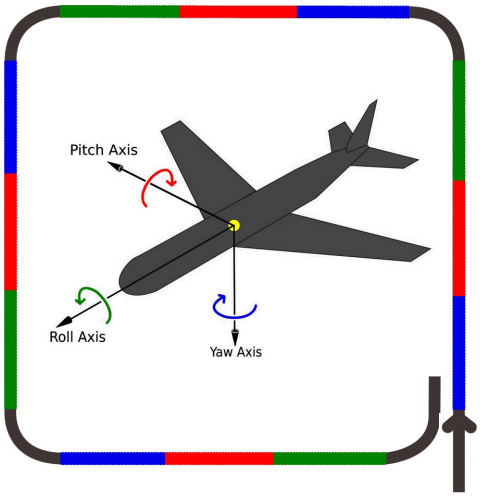


Figure 2.2: The "Figure-of-Merit" flight pattern. In this pattern the platform performs sinusoidal roll, pitch and yaw movements along a square path. Image adapted from [31].

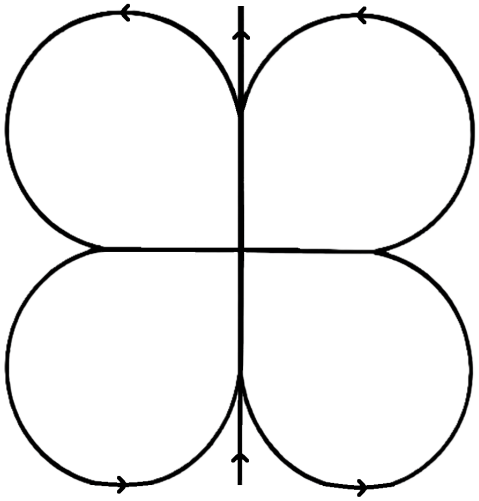


Figure 2.3: The cloverleaf flight pattern. In this pattern the platform traverses a single point from four different directions.

2.3.1. Modeling On-board Electronics

The standard TL model does not take into account the on-board electronics (OBE). These electronics introduce magnetic noise that can have a large effect on the measurements, depending on how far the measurements are taken from the electronics. In [32] the situation in which the magnetometers are closer to the platform is examined. The on-board electronics are modeled based on the Biot-Savart law, which describes the magnetic field generated by an electric current. Small distance between electronics and magnetometers can cause current-induced magnetic field, $\mathbf{B}_I$ in the same order of magnitude as the Earth magnetic field, $\mathbf{B}_e$. The OBE interference from input k due to current I_k is modeled as

$$\mathbf{B}_{OBE,k} = \mathbf{B}_{I_k} + \mathbf{B}_{I_k,ind} + \mathbf{B}_{I_k,eddy}, \tag{2.33}$$

where $\mathbf{B}_{I_k,ind}$ and $\mathbf{B}_{I_k,eddy}$ are the induced and eddy-current field due to $\mathbf{B}_{I_k}$. Since $\mathbf{B}_{I_k,ind}$ and $\mathbf{B}_{I_k}$ are both linear w.r.t I_k and $\mathbf{B}_{I_k,eddy}$ is linear w.r.t. $\dot{I}_k$. Then the OBE interference of the scalar magnetometer can be represented by

$$\mathbf{B}_{OBE,k} = I_k \boldsymbol{d}_k + \dot{I}_k \boldsymbol{e}_k, \tag{2.34}$$

where $\boldsymbol{d}_k = \begin{bmatrix}\gamma_{6k-5} & \gamma_{6k-4} & \gamma_{6k-3}\end{bmatrix}^T$ and $\boldsymbol{e}_k = \begin{bmatrix}\gamma_{6k-2} & \gamma_{6k-1} & \gamma_{6k}\end{bmatrix}^T$ with γ_i scalar real values.
The Tolles-Lawson model given in (2.14) is then extended with the OBE magnetic field model in (2.34) for each input k as

$$\mathbf{B}_{p'} = \mathbf{B}_p + \sum_{k=1}^{K} \mathbf{B}_{OBE,k}.$$

This can then be used to predict platform noise similarly to Section 2.2.1 by defining

$$\boldsymbol{A}' = \begin{bmatrix} \boldsymbol{\delta}(1) & \boldsymbol{\delta}(2) & \dots & \boldsymbol{\delta}(N) \\ \boldsymbol{\rho}_1(1) & \boldsymbol{\rho}_1(2) & \dots & \boldsymbol{\rho}_1(N) \\ \boldsymbol{\rho}_2(1) & \boldsymbol{\rho}_2(2) & \dots & \boldsymbol{\rho}_2(N) \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\rho}_K(1) & \boldsymbol{\rho}_K(2) & \dots & \boldsymbol{\rho}_K(N) \end{bmatrix}^T, \tag{2.35}$$

$$\boldsymbol{\beta}' = \begin{bmatrix}\boldsymbol{\beta}^T & \gamma_1 & \dots & \gamma_{6K}\end{bmatrix}^T,$$

$$\boldsymbol{\rho}_k = \frac{1}{B_t}\begin{bmatrix} I_k \mathbf{B}_t^T & \dot{I}_k \mathbf{B}_t^T \end{bmatrix}^T.$$

Then,

$$\text{bpf}(\boldsymbol{y}) = \text{bpf}(\boldsymbol{A}')\boldsymbol{\beta}', \tag{2.36}$$

where the least-squares optimal solution for $\boldsymbol{\beta}'$ is

$$\hat{\boldsymbol{\beta}}' = \text{bpf}(\boldsymbol{A}')^{\dagger}\text{bpf}(\boldsymbol{y}). \tag{2.37}$$

2.3.2. Linear Model of Inputs

Jukić et al. extend the approach in [32] with the heuristic assumption that the magnetic field is linearly dependent on more inputs than just current [20]. First, the standard TL model is used to find the TL coefficients as in (2.24). Then, the remaining signal that is not accounted for by the TL model is correlated with the drone inputs. Inputs with high correlations are then used in solving a linear system,

$$\begin{aligned}
\boldsymbol{y}_r &= \boldsymbol{A}_r\boldsymbol{\beta}_r, \\
\boldsymbol{A}_r &= \begin{bmatrix} \boldsymbol{\rho}_{r,1}(1) & \boldsymbol{\rho}_{r,1}(2) & \dots & \boldsymbol{\rho}_{r,1}(N) \\ \boldsymbol{\rho}_{r,2}(1) & \boldsymbol{\rho}_{r,2}(2) & \dots & \boldsymbol{\rho}_{r,2}(N) \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\rho}_{r,K_r}(1) & \boldsymbol{\rho}_{r,K_r}(2) & \dots & \boldsymbol{\rho}_{r,K_r}(N) \end{bmatrix}^T, \\
\boldsymbol{\rho}_{r,k} &= \begin{bmatrix} I_k \frac{\mathbf{B}_t^T}{B_t} & \dot{I}_k \frac{\mathbf{B}_t^T}{B_t} & I_k \end{bmatrix}^T
\end{aligned} \tag{2.38}$$

where $\boldsymbol{y}_r$ is a vector containing the residual magnetic field over time and I_k is the k^{th} drone input where $k \in \{1, \dots, K_r\}$. Note that another heuristic assumption is made, namely that the drone inputs directly have an effect on the Earth-frame magnetic field independently of the direction of B_t. This is captured by the last element of $\boldsymbol{\rho}_{r,k}$.

2.4. Evaluation Metrics

To quantify the amount of platform noise on an uncompensated total field signal, the standard deviation is commonly used [29],

$$\sigma_u = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(B_t(k) - \bar{B}_t\right)^2}, \tag{2.39}$$

where $\bar{B}_t$ is the mean of the uncompensated signal. To quantify the performance of compensation methods the Improvement Ratio (IR), as discussed in [29], is used,

$$IR = \frac{\sigma_u}{\sigma_c}. \tag{2.40}$$

Here σ_u is the standard deviation of the uncompensated signal B_t and σ_c is the standard deviation of the compensated signal B_e^{est}, which estimates the Earth field.

Sometimes we are only interested in a certain frequency band of a signal. In that case the performance of the compensation of that signal is evaluated in that frequency band. This is done using the bandpass-filtered improvement ratio,

$$IR_{bpf} = \frac{\sigma_{u,bpf}}{\sigma_{c,bpf}}. \tag{2.41}$$

2.5. Physics-Informed Machine Learning

Often when studying physical systems, there is knowledge of the physical laws that govern the system. This knowledge can be incorporated into the neural network to increase accuracy and interpretability. This approach is called physics-informed machine learning (PIML) and is further discussed in [33]. Here some principles of introducing physics-information into the machine learning algorithm are introduced, namely:

- Observational bias
- Inductive bias
- Learning bias

These biases are added into the algorithm to guide it into finding physically consistent solutions. Observational bias amounts to ensuring the data the algorithm uses to learn reflect the underlying physics. Inductive biases can be added by mathematically imposing physical laws or hard constraints on the system. A drawback of this method is that it may be complex and hard to scale. Learning biases come from choosing a loss function such that during the training phase, the algorithm leans toward a solution that complies with the underlying physics. Each of these biases comes with their own benefits and drawbacks and can be combined to reach the best outcome.

2.6. Feature Engineering and Selection

Feature engineering is the process of transforming raw input data into informative features that improve model performance. Feature selection helps identify which inputs are most influential in predicting magnetic noise. We first explore simple statistical relationships using correlation. Next, SHAP values are introduced to interpret both the ETL and NN models, guiding feature refinement and model comparison.

2.6.1. Input Feature Engineering

The design of input features significantly influences model performance in platform noise compensation applications. Effective feature engineering incorporates domain knowledge about the underlying physics while providing the network with informative representations.

Raw sensor inputs: Direct use of accelerometer, magnetometer, and angular rate measurements provides the most flexibility but may require the network to learn complex transformations. Proper normalization and scaling are crucial for stable training.

Derived features: Physics-inspired features such as magnetic field derivatives, aircraft attitude angles, and manoeuvre-specific parameters can significantly improve performance by providing the network with relevant inductive biases.

Temporal features: Including temporal context through sliding windows, finite differences, or explicit time embeddings can help capture dynamic effects and time-varying magnetic field relationships.

2.6.2. Correlation

In [32] and [20], features were selected based on linear correlation, defined as follows,

$$r = \frac{\sum_{i=1}^{N}(B_i - \bar{B})(I_i - \bar{I})}{\sqrt{\sum_{i=1}^{N}(B_i - \bar{B})^2} \cdot \sqrt{\sum_{i=1}^{N}(I_i - \bar{I})^2}},$$

where B denotes the magnetic field and I the input. This formula represents the Pearson correlation coefficient, which measures the strength and direction of the linear relationship between two variables. Values of r close to 1 or -1 indicate a strong positive or negative linear correlation, respectively, while values near 0 suggest weak or no linear relationship.

Because the methods used in [32] and [20] assume linear dependencies, using linear correlation for feature selection is appropriate in those contexts. However, relationships between inputs and platform noise may be nonlinear.

To account for such cases, a common alternative is the Spearman rank correlation coefficient, which measures the strength and direction of monotonic relationships (not necessarily linear) by evaluating the linear correlation of the ranked variables,

$$r_s = \frac{\sum_{i=1}^{N}(R_{B_i} - \overline{R_B})(R_{I_i} - \overline{R_I})}{\sqrt{\sum_{i=1}^{N}(R_{B_i} - \overline{R_B})^2} \cdot \sqrt{\sum_{i=1}^{N}(R_{I_i} - \overline{R_I})^2}},$$

where R_{B_i} and R_{I_i} are the ranks of the magnetic field and input, respectively. Here rank is the numerical position of a value within the ordered list of data, with 1 assigned to the smallest value. The Spearman correlation is more robust to outliers and nonlinearities, and can uncover consistent increasing or decreasing trends between variables even when the relationship is not linear.

2.6.3. SHAP Values

To assess the contribution of each input feature to the prediction is to compute SHAP (SHapley Additive exPlanations) values [34]. SHAP values are grounded in cooperative game theory and provide a unified framework to explain the output of any machine learning model.

Each SHAP value quantifies the marginal contribution of a feature to a specific prediction, averaged over all possible combinations of input features. Formally, for a prediction function p and a set of input features x, the SHAP value ϕ_i for feature i represents

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[p(S \cup \{i\}) - p(S)\right],$$

where F is the set of all features and S is a subset excluding i. While computing exact SHAP values is computationally expensive, efficient approximations are available for tree-based models, deep learning models, and others [34].

SHAP values provide several key advantages:

- They offer local explanations, attributing a portion of an individual prediction to each input feature.
- By aggregating SHAP values over multiple samples, one can obtain global feature importance rankings.
- SHAP values are model-agnostic and ensure consistency: if a model changes such that a feature contributes more, its SHAP value will not decrease.

In this work, SHAP values are used to identify which drone inputs most significantly contribute to magnetic field estimation errors. This provides a more interpretable and nuanced understanding of feature relevance than correlation-based metrics alone.

2.6.4. SHAP for Deep Learning

To interpret the neural network models used for magnetic noise compensation, we apply the *Gradient-Explainer* from the SHAP library, which is an implementation of *Deep SHAP* [34]. This method extends the concept of Shapley values to deep learning models.

Deep SHAP approximates SHAP values by integrating the gradients of the model output with respect to the input features along a straight path from a baseline input x' to the actual input x:

$$\phi_i = (x_i - x_i') \cdot \int_0^1 \frac{\partial p\left(x' + \alpha(x - x')\right)}{\partial x_i}\, d\alpha \tag{2.42}$$

Here, ϕ_i represents the attribution assigned to input feature i, p is the prediction function, and α parametrises the interpolation between x' and x. The integral is computed numerically via sampling. We use randomly sampled inputs of the training data as the baseline x'.

While *GradientExplainer* is computationally efficient and interpretable, it has several limitations:

- Attributions are sensitive to the choice of baseline input.
- The straight-line integration path may traverse unrealistic regions of input space.
- It does not fully separate interaction effects between features.
- The method assumes the model is differentiable with respect to its inputs.

In this thesis, we use *GradientExplainer* to generate SHAP values for each feature input to our neural network. This allows us to assess which inputs (e.g., pitch, airspeed, control signals) contribute most to the residual magnetic noise prediction and how these contributions vary across flight manoeuvres.

3

Dataset Description and Preliminary Analysis

This chapter introduces the dataset collected by TNO to investigate aeromagnetic plaform noise on a fixed-wing drone. Note that for confidentiality reasons, all magnetic field data in this thesis has been demeaned. The chapter begins with a discussion on the ground truth used in this dataset versus that used in the literature. It then outlines the characteristics of the drone platform, the magnetometers, and the drone input data used for analysis. Additionally, the chapter describes the process of data collection, including both ground and flight tests, and highlights several practical challenges encountered during and after data acquisition. These challenges set constraints which are taken into consideration during the modelling approaches discussed in later chapters.

3.1. Ground Truth

Flying at high altitude naturally filters out short-wavelength magnetic anomalies, leaving primarily the core field. This low-pass filter like effect is described in Appendix A. This creates a reference-free ground truth where any significant variation is likely due to platform noise rather than the anomaly field. We can estimate the presence of magnetic anomalies in our data as follows.

We start with the following assumptions satisfied by the TNO data:

- The survey is flown at an altitude of $z = 250\,\mathrm{m}$
- The survey area is $800\,\mathrm{m} \times 800\,\mathrm{m}$
- The aircraft velocity does not exceed $100\,\mathrm{km/h} = 27.8\,\mathrm{m/s}$
- Crustal anomalies are modelled as sinusoidal signals with spatial wavelength λ

In the Fourier domain, upward continuation of the magnetic field to height z introduces an exponential decay factor [35], which is described in Appendix A. We can use this to estimate the decay of anomalies of certain wavelength, shown in Table 3.1.

Table 3.1: Attenuation of sinusoidal magnetic components at 250 m altitude

Wavelength λ	$\mathcal{F}[\psi] = e^{-2\pi \cdot 250\mathrm{m}/\lambda}$	Attenuation (%)
50 m	$e^{-31.4} \approx 2.4 \times 10^{-14}$	∼100 %
100 m	$e^{-15.7} \approx 1.5 \times 10^{-7}$	∼99.99999 %
200 m	$e^{-7.85} \approx 3.9 \times 10^{-4}$	99.96 %
400 m	$e^{-3.93} \approx 0.020$	98 %
800 m	$e^{-1.96} \approx 0.14$	86 %

The significance of the anomalies also depends not only on the wavelength, but also on the magnitude.

Assuming a typical anomaly in this region does not exceed 50 nT, 400 m wavelength anomalies will be attenuated to below 1 nT. Anomalies with wavelength of 800 m will be attenuated to 7 nT. This amount could still be significant.

To understand how this attenuation affects the temporal signal recorded by the magnetometer, we relate the spatial wavelength to a temporal frequency using the aircraft's velocity $v = 27.8\,\mathrm{m/s}$:

$$f_t = \frac{v}{\lambda}.$$

This leads to the correspondence shown in Table 3.2.

Table 3.2: Temporal frequencies corresponding to spatial wavelengths

Wavelength λ	Temporal Frequency f_t (Hz)
800 m	0.035
400 m	0.070
200 m	0.140
100 m	0.278
50 m	0.556

The results in Table 3.2 imply that the signal measured at 250 m altitude is dominated by very low-frequency components, with most short-wavelength features absent due to strong attenuation. In practice, this means that the recorded magnetic field will appear smooth and slowly varying if not constant. This high-altitude signal is therefore used as ground truth and any variation in the signal is likely due to platform noise rather than geological anomalies.

3.2. TNO Dataset Overview

The measurement platform is an electrically driven, 3600 mm wingspan Mugin fixed-wing drone [36], shown in Figure 3.1. Scalar and vector magnetometers were attached near the wing tips below the wing on both sides, as shown in Figure 3.2. This position was chosen due to the distance from the expected main noise source: the electrically driven engine, which draws up to 120 A from the battery [36]. In addition to being a source of current, which induces a magnetic field, the electric motor also contains permanent magnets and ferromagnetic material such as iron.

Figure 3.1: Mugin fixed-wing drone.

Figure 3.2: Close-up of the Mugin fixed-wing drone right side wing tip. The MicroSAM OPM magnetometer is labelled with '0' and the VMR vector magnetometer is labelled with '1'.

3.2.1. Magnetometers

Both scalar and vector magnetometers are used in traditional noise compensation methods, which were described in Section 2.2. Scalar magnetometers are selected for their accuracy in determining field magnitude. Vector magnetometers are less accurate in absolute terms, but can still be reliably used to measure the vector direction of the magnetic field [21]. The specifications of the magnetometers are given in Table 3.3. The scalar magnetometer used is an optically pumped magnetometer

(OPM), for which the operating principle is discussed in [37]. The vector magnetometer used is a VMR magnetometer, for which the operating principle is discussed in [38].

Table 3.3: Specifications of the magnetometers used [39], [40]

Property	Scalar Magnetometer	Vector Magnetometer
Type	Total field	Vector field
Model	Twinleaf MicroSAM	Twinleaf VMR
Sensitivity	20 pT/$\sqrt{\text{Hz}}$	300 pT/$\sqrt{\text{Hz}}$
Sample Frequency	250 Hz	200 Hz
Field Range	10,000 to 100,000 nT	-100,000 to 100,000 nT

Even though the sampling frequencies of these sensors are 200 to 250 Hz, using all data from these frequencies is not desired. Frequency-dependent noise is a factor that limits the accuracy of our models and of the reconstructed signal. The signal of interest, being the anomaly field, will only appear at frequencies that are low compared to the sampling frequency. The frequency of the anomaly field as measured depends on platform velocity and altitude. In [9], the formula for the maximum frequency due to the crustal field is given as $f_{\max} = \frac{v}{h}$, where v is the velocity and h is the altitude. Given that the maximum velocity of the drone is $160\,\mathrm{km/h} \approx 45\,\mathrm{m/s}$ and assuming an altitude greater than $50\,\mathrm{m}$, it was estimated that the anomaly field will not exceed 0.9 Hz. Some higher frequency components could help in predicting the platform noise. However, too high a sampling frequency could be redundant and computationally expensive; therefore, the magnetometers were resampled to $20\,\mathrm{Hz}$.

Sampling at $20\,\mathrm{Hz}$ results in a noise standard deviation of $20\,\mathrm{pT}/\sqrt{\mathrm{Hz}} \times \sqrt{10\,\mathrm{Hz}} \approx 63\,\mathrm{pT}$ for the OPM and $300\,\mathrm{pT}/\sqrt{\mathrm{Hz}} \times \sqrt{10\,\mathrm{Hz}} \approx 949\,\mathrm{pT}$ for the VMR. This gives us a noise floor and a limit to the accuracy that can be achieved by compensating the deterministic noise.

Frequency-dependent noise is not the only factor determining the usefulness of the sensors. The VMR sensor displays a large amount of drift across time. Furthermore, both sensors need a 'warm-up' time where the sensor needs to be turned on for an extended period of time, up to 45 minutes, before the readings settle. Moreover, both sensors, but especially the VMR, vary depending on temperature. The VMR also appears to be sensitive to movement, showing large spikes during the manoeuvres.

The drift over time and sensitivity to movement of the VMR sensor is clear in Figure 3.3. The VMR data appears to vary much more than the OPM during manoeuvres during manoeuvres, but similar or smaller in the ground tests. In the figure, the VMR data also drifts upward over time, while the OPM data varies around a steady value. These characteristics make the OPM more reliable for magnitude measurements, whilst the VMR provides useful directional information despite its limitations.

3.2.2. Drone Inputs

The drone inputs used to predict the platform noise come from data collected automatically by the drone. They consist of a wide range of data, including but not limited to aileron, elevator and rudder control signals, battery current, roll, pitch and yaw angles, quaternions and airspeed. There are a total of 273 features. Many of these are highly or perfectly correlated, representing the same signal at different stages of the flight control pipeline. Initial feature selection for further examination was motivated by empirical reasoning and a qualitative understanding of the drone. The full list of drone features can be found in Appendix B. The traditional methods discussed in Chapter 2 are not equipped to deal with varying sample rates; therefore, all drone features were resampled to 20 Hz, the resampled sampling rate of the magnetometers.

3.2.3. Data Collection

Two pairs of vector and scalar magnetometers were placed on the wing tips of the platform, as can be seen in Figure 3.2. First, ground tests were conducted to assess the impact of drone features on the magnetic field. Next, in-flight data was collected. Multiple flights were conducted over the same area, on different days spanning a year. Flight patterns were pre-programmed and flown by the Ardupilot automatic pilot.

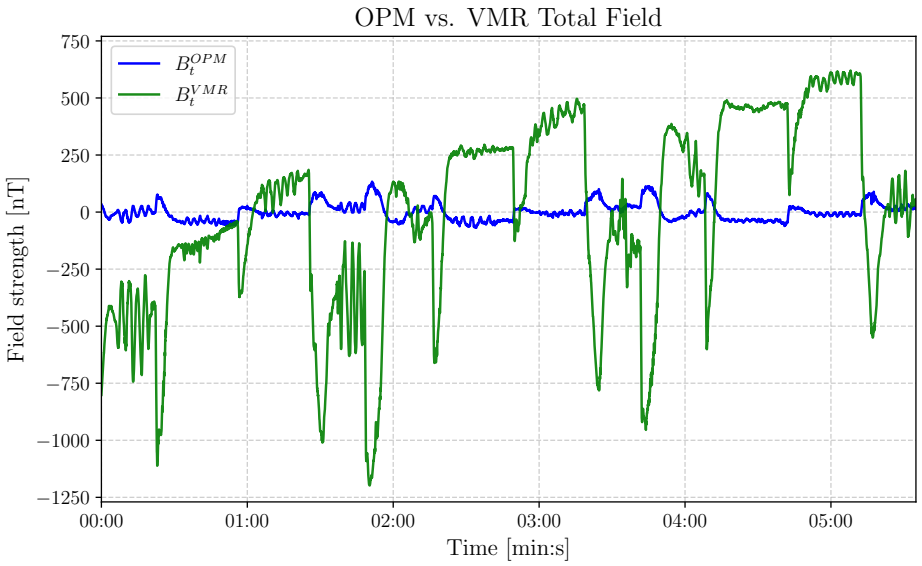


Figure 3.3: Comparison of the magnetic field measured by the OPM versus VMR magnetometers. The VMR shows greater variability and nonlinear response during manoeuvres compared to the more stable OPM measurements.

Ground Test

Magnetometer data was captured on-ground. Three drone inputs were tested, consisting of activating the LED lights, steering left and right, and throttling the engine. The magnetic field measured during the entire ground test can be seen in Figure 3.4a. The largest effect was during the section that had the LED light blinking, shown close-up in Figure 3.4b, which disturbed the magnetic field by about 50 nT. In this figure it is also clear that there are some ringing artifacts. These are caused by the low-pass filtering of the blinking LED input, which represents an impulse. The steering, shown in close-up in Figure 3.4c had an effect of around 10 nT. Throttling the engine, shown in Figure 3.4d, varied the magnetic field by around 5 nT. These value of platform noise are significant for magnetic navigation, since they are of similar magnitude that anomalies can be.

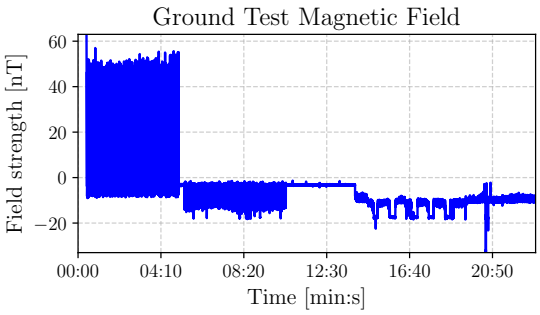
The effect of the 150 mA LED compared to the 120 A engine is a good example of the extremely quick falloff of magnetic field strength due to distance, because the LED wires run close to the sensor, while the engine is relatively far away. So even though the engine has 800 times the current, the effect is 10 times less. Assuming the magnetic field strength is related to current I and distance r as $B \propto \frac{I}{r^3}$, the magnetic field strength due to the LED current relative to the throttle current is

$$\frac{I_{\text{thr}}/B_{\text{thr}}}{I_{\text{LED}}/B_{\text{LED}}} = \frac{120\,\text{A}/5\,\text{nT}}{0.150\,\text{A}/50\,\text{nT}} = 8000,$$

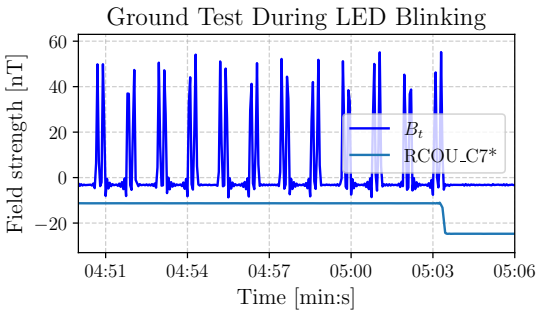
then the relative distance is $\sqrt[3]{8000} = 20$, meaning the distance from the throttle source is 20 times greater than the distance from the LED source. So even though the throttle source has an 800 times larger current, the effect is 10 times smaller. Note that this proportionality does not hold exactly for wires closeby and depends on the exact geometry of wires, but serves the purpose of showing the effect of cubic dropoff. This quick decay can mean that any small change in platform noise can have a large effect on the magnetic field if it occurs close to the magnetometer.

Flight Overview

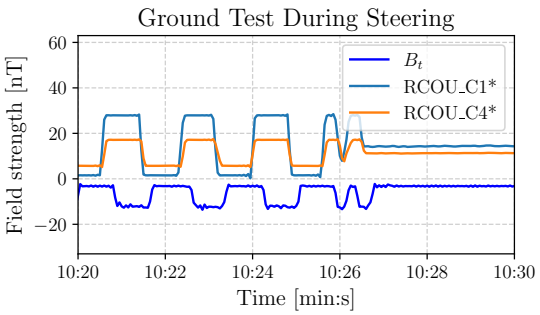
In the TNO dataset, the pattern flown is a Figure-of-Merit (FoM) (discussed in Section 2.2.4), as is shown in Figure 3.5. Each leg of the FoM square was flown in a cardinal direction. Three squares were flown per pattern, one for each roll, pitch and yaw phase. The resulting magnetic field during a FoM can be seen in Figure 3.6. Each background color in Figure 3.6 represents one phase of the FoM. The roughly sinusoidal variation in the roll, pitch and yaw angles can be seen to translate to variation in the magnetic field, as expected. The largest changes in the magnetic field occur due to heading-related variation. This happens when the platform turns 90° in yaw angle.



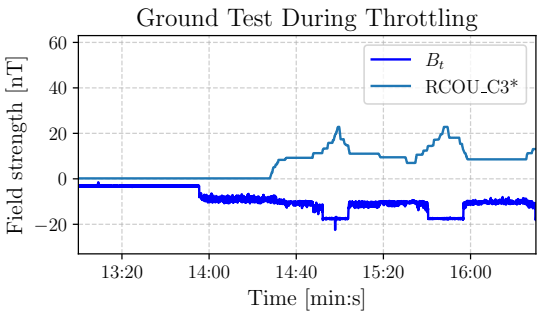
(a) Ground data magnetic field during several drone operations. Zoomed in parts of this figure are shown in b to d.



(b) Ground test data during LED operation. RCOU_C7 is the LED control signal. When high, the LEDs repeatedly blink twice rapidly followed by a pause. When low, the LEDs are off. Here, the control signal transitions from high to low, at which point the magnetic field also stops spiking.



(c) Ground test data during steering operation. RCOU_C1 is the aileron control, RCOU_C4 is the rudder control. During steering left and right repeatedly the aileron and rudder are actuated.



(d) Ground test data during throttle operation. Note RCOU_C3 is the throttle control.

Figure 3.4: Ground test data during different drone input actuations. *Drone inputs rescaled.

For this thesis, out of the TNO dataset a selection of 5 flights was made. During each flight one or two FoM patterns were flown. This was done to perform calibration on one FoM, and validation on another. In some cases the FoMs were followed by a 'clean' manoeuvre, consisting of a completion of a single square path without the sinusoidal attitude changes. This clean manoeuvre is done to more accurately represent regular flight and is used to validate the calibration done during a FoM. In Table 3.4 there is an overview of which flights have FoMs and which have clean manoeuvres. The flight numbers and manoeuvre numbers will be referenced later to make the distinction between the different flights and manoeuvres. During all flights both OPM and VMR data was collected. However, only some flights contain both left- (LHS) and right-hand side (RHS) sensors and others only contain LHS sensor data, as is indicated in Table 3.4.

Date	Flight #	Manoeuvre #	Manoeuvre Type	Sensor
20-Jul-2023	2	1	FoM	L+R
20-Jul-2023	2	1C	Clean	L+R
20-Jul-2023	2	2	FoM	L+R
05-Sep-2023	7	3	FoM	L+R
14-Sep-2023	9	4	FoM	L+R
11-Jul-2024	10	5	FoM	L
11-Jul-2024	10	6	FoM	L
11-Jul-2024	10	6C	Clean	L
11-Jul-2024	11	7	FoM	L
11-Jul-2024	11	8	FoM	L
11-Jul-2024	11	8C	Clean	L

Table 3.4: Overview of flights by date, flight number, and manoeuvre number, as well as whether the flight collected data on the left wing (L) or on both the left and right wing (L+R).

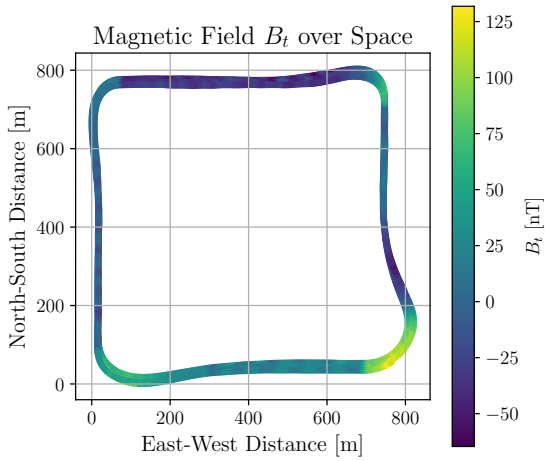


Figure 3.5: Top view of a manoeuvre flight path. The color bar represents magnetic field magnitude in nT.

3.2.4. Quantization Effect

A significant challenge comes from the data captured using the scalar magnetometer. There appears to be quantization or staircasing, where the data 'prefers' some specific values over most other values. In Figure 3.7, the left sensor (blue line) shows some values which occur very frequently.

Quantization is not limited to data from this specific flight but occurs in all flights. In Figure 3.8, all OPM data is plotted as a histogram. Sharp, evenly spaced peaks can be seen, indicating that some specific, evenly spaced magnetic field values occur far more often than most values. This is not expected, because during flight, the magnetic field should continuously change. The fact that these peaks occur on different flights, different days and even on different sensors at the same magnetic field values indicates that this is not physical behaviour.

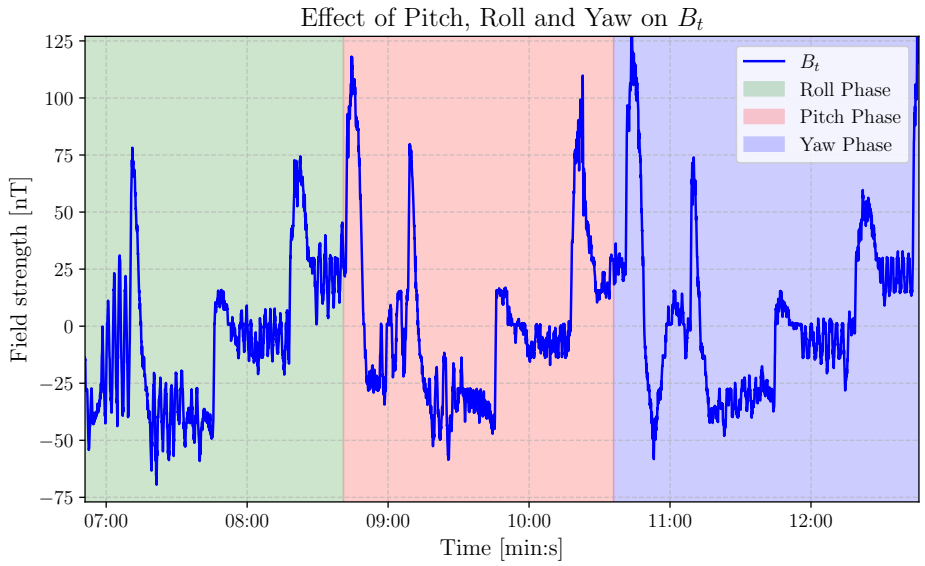
Determining the cause of quantization is not trivial. It was determined that the quantization already occurs in the raw data, before preprocessing. This means that the quantization happens in one of the following processes:

- Sensor measurement
- Analogue-to-digital conversion (ADC)
- Writing data to hard drive
- Reading data for processing

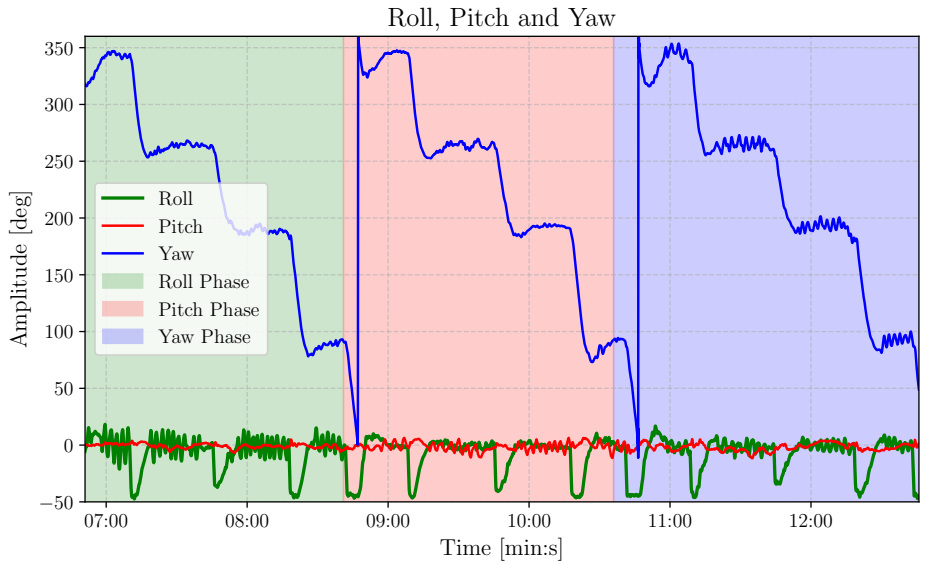
When looking at a narrow band around a single peak in the histogram of the raw data, it can be seen that the data does not perfectly match one exact value. There is still some very narrow distribution around the peaks. This means that after the quantization occurs, some small amount of noise is still added. After the ADC, the data is in digital form, which should not be susceptible to noise. This is evidence that the quantization happens before ADC, meaning there is an error in the actual measurement of the magnetic field.

One way to deal with quantization is to simply remove the quantized values from the data and assume the rest of the data is accurate. The downside is that the quantized data represents over 30% of all data, resulting in significant reduction in the size of the dataset. To assess whether quantized data impacts prediction accuracy, models should be trained on both the data including the quantized values and the data with the quantized values removed. Since the quantization is most significant in the left-side sensor, this is the one that is further examined. The histogram of the de-quantized data is shown in Figure 3.9.

Using the algorithm introduced in Section 2.2.2, testing shows improvement in the error metric in some



(a) Magnetic field B_t



(b) Drone orientation angles (roll/pitch/yaw)

Figure 3.6: Magnetic field and orientation during the manoeuvres of a single FoM.

flights and worsening in others. Since there is no consistent improvement in the error metric between the de-quantized data and the quantized data, all subsequent analyses in this thesis include the quantized data. This decision preserves the full dataset whilst acknowledging that approximately 30% of the OPM measurement data may contain measurement artefacts.

3.3. Chapter Summary

Having described the dataset and platform, this research is distinguished from previous work by several constraints. The data is reference-free, as the TNO dataset provides no noise-free reference measurements unlike commonly used datasets such as the MIT aeromagnetic dataset [25], requiring validation from the assumption about the Earth field being constant. A challenge is posed by quantization effects on the OPM data. Finally, due to the large feature dimensionality a systematic feature selection

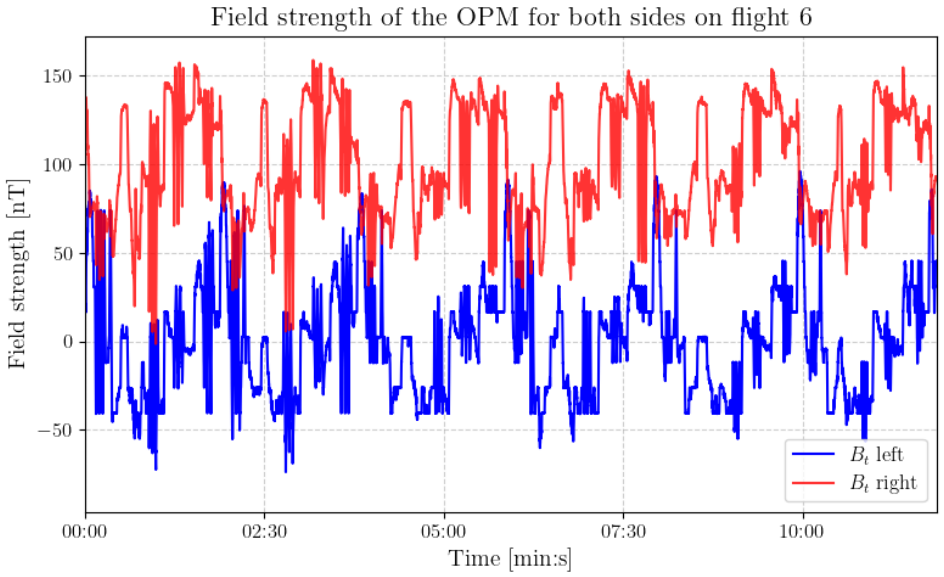


Figure 3.7: Field strength against time snippet. The blue line (left OPM) has significant quantization.

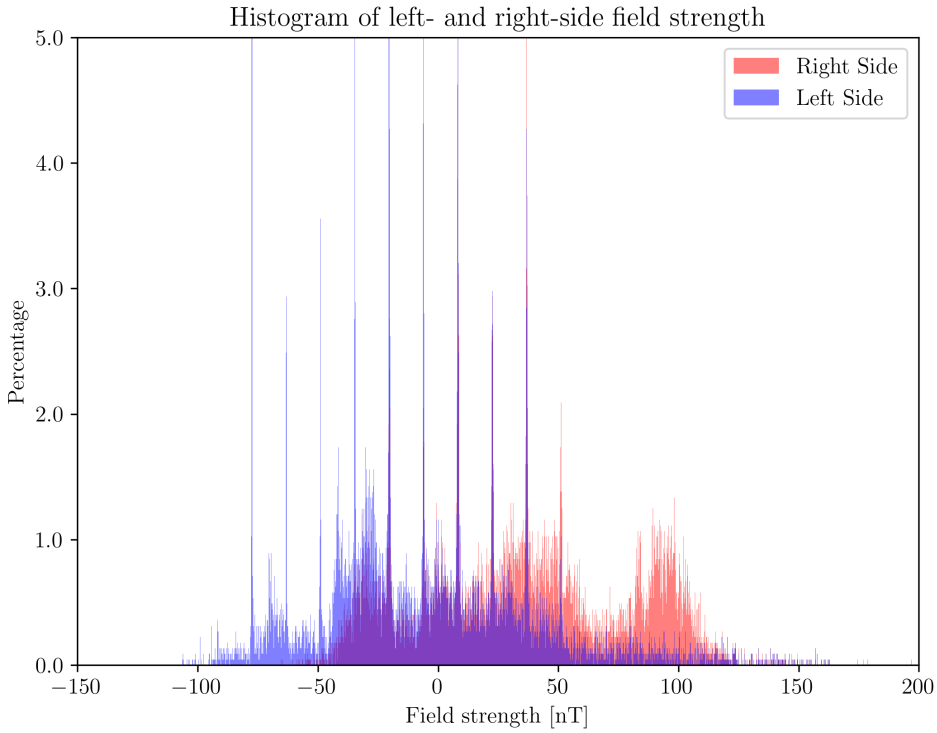


Figure 3.8: Histogram of all scalar magnetometer data, left and right side. The left side shows significant quantization.

approach is necessary. These characteristics and constraints will inform the modelling approaches presented in subsequent chapters.

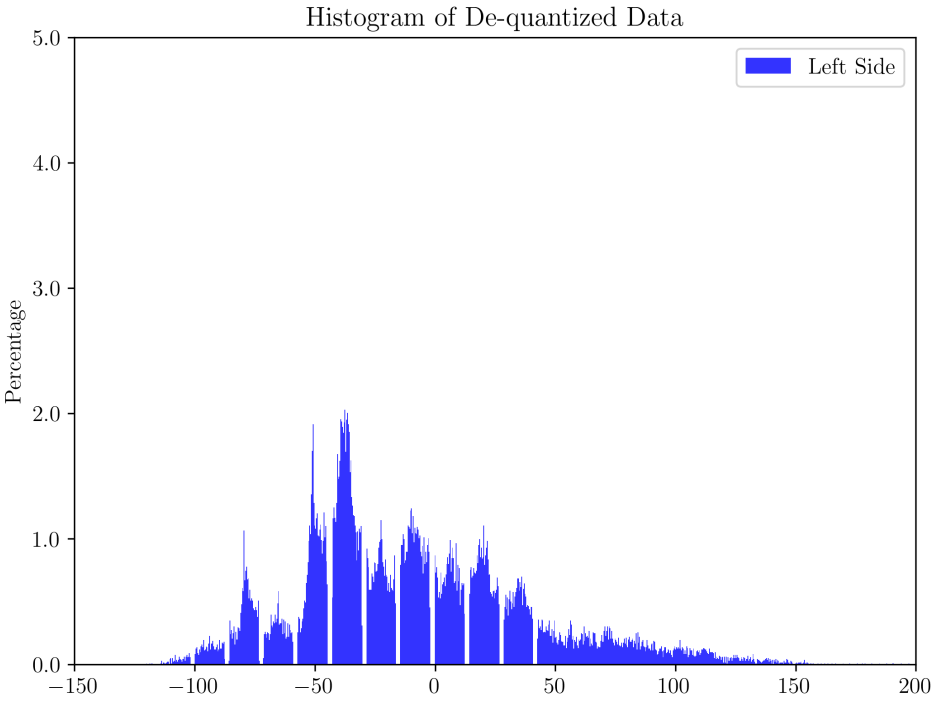


Figure 3.9: De-quantized left side sensor data histogram.

4

Methods and Results

This chapter presents the implementation and evaluation of magnetic field compensation methods applied to the TNO dataset. The work progresses systematically from validating fundamental assumptions to developing and comparing different compensation approaches.

The chapter begins by examining the validity of the approximation underlying the Tolles-Lawson model, an assumption that has not been validated in prior literature despite its common use. This analysis establishes the conditions under which the approximation remains accurate and gives bounds on errors.

Following this validation, the standard Tolles-Lawson compensation method is implemented and analysed. A previously unreported trivial solution is given. Then the necessity of regularization techniques is shown. The analysis extends to the Extended Tolles-Lawson (ETL) approach that incorporates drone inputs alongside traditional magnetic field measurements.

Next, deep learning approaches are investigated, implementing multilayer perceptron (MLP) neural networks to model residual platform noise after initial compensation. These neural network methods are applied both to supplement Tolles-Lawson compensation (TLNN) and to enhance the Extended Tolles-Lawson approach (ETLNN), with systematic evaluation of their performance on both optically pumped magnetometer (OPM) and vector magnetometer (VMR) data.

A feature analysis using SHAP (SHapley Additive exPlanations) values provides insights into which drone inputs most significantly contribute to platform noise prediction. This analysis guides the development of feature-informed models that balance compensation performance with model complexity to counteract overfitting.

The chapter concludes with the implementation of feature-selected neural network architectures designed to mitigate overfitting while maintaining compensation effectiveness. Results are evaluated using the improvement ratio (IR) and overfitting ratio (OR), providing a measure of each method's performance across multiple flight manoeuvres.

4.1. Validity of Approximation

Here the validity of the approximation of the Earth field magnitude by the difference between the total field magnitude and the platform field projection, given in (2.11), is checked. This validation is necessary as the approximation might introduce significant error in the estimation of the Earth field magnitude B_e and consequently in estimation of the anomaly $\tilde{B}_a$. Furthermore, the author is unaware of a validation of assumption (2.11) done elsewhere in literature, even though this approximation is done any time the Tolles-Lawson model is used. Different values of $\mathbf{B}_p$ and $\mathbf{B}_e$ will be used to see when the approximation is valid and when not. The approximation (2.11) is similar to the approximation of the projection of the anomaly field onto the core field in (2.5). Consequently the analysis of the approximation in (2.11) is similar to the analysis done by Canciani in [9] of the approximation in (2.5).

In Figure 4.1 and 4.2 it can be seen how the angles between the vectors and the value of B_p relative

to B_e influence the accuracy of the approximation in (2.11). The approximation is worst when $\theta = 90°$. This is because then the magnitude of the projection of the platform field onto the total field is zero. Note that these figures do not consider realistic values, that is why the error is quite large. Next, realistic values will be considered for validation of the approximation.

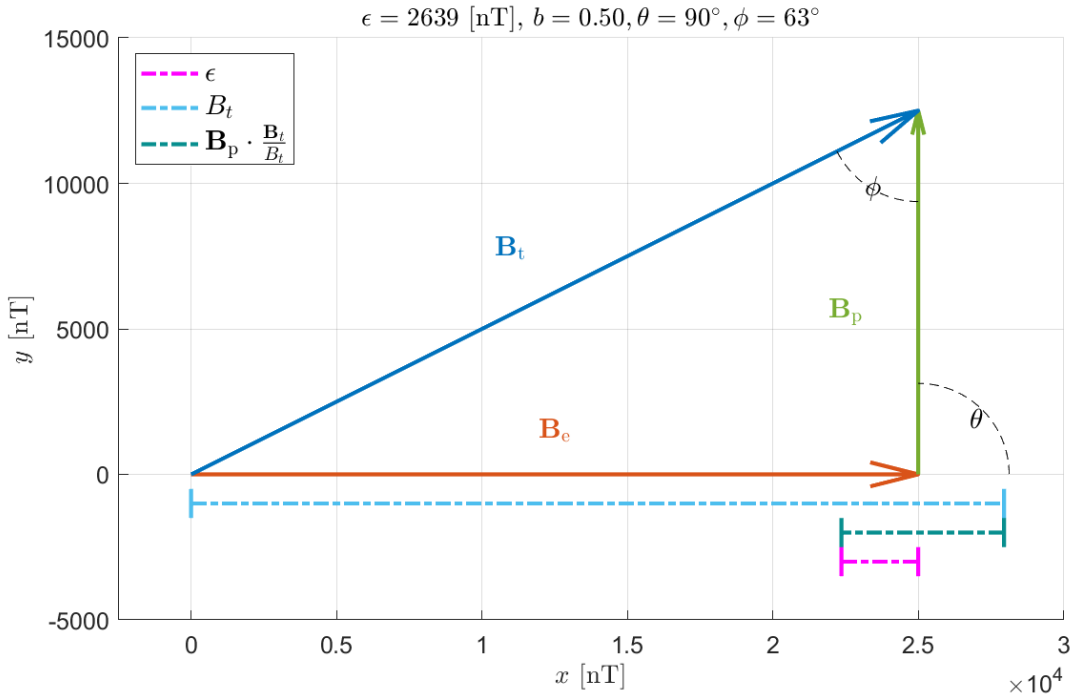


Figure 4.1: Interaction of the total field $\mathbf{B}_t$, the Earth field $\mathbf{B}_e$, and the platform noise $\mathbf{B}_p$. The error made in the approximation in (2.11) depends on the angle θ and the relative magnitude $\frac{B_p}{B_e}$.

In [41] a realistic amount of platform noise is given as 100 nT for a surveillance aircraft and up to 1000 nT for an F16. According to the IGRF [11] the Earth field strength ranges from about 25000 nT to 70000 nT. Based on the IGRF, values can conservatively be chosen as $B_e \geq 25000$ nT and $B_p \leq 1000[nT]$. From this a scaling factor can be calculated as $\frac{B_p}{B_e} \leq \frac{1}{25}$. It can be shown that (2.11) achieves the worst absolute error,

$$\epsilon = \left| B_e - \left(B_t - \mathbf{B}_p \cdot \frac{\mathbf{B}_e}{B_e} \right) \right|,$$

when $\mathbf{B}_t$ and $\mathbf{B}_p$ are perpendicular. In the case $B_p \ll B_e$, this can be approximated by $\mathbf{B}_e$ and $\mathbf{B}_p$ being perpendicular. This worst case angle, with a realistic scaling factor, can be seen in Figure 4.3.

In the perpendicular case the relative error, calculated as $\epsilon_r = \frac{\epsilon}{B_e}$, as a function of the relative magnitude, $b = \frac{B_p}{B_e}$, is shown in Figure 4.4.

In the worst case assumption of $\frac{B_p}{B_e} = \frac{1}{25}$ the relative error as a function of the angle between θ can be seen in Figure 4.5. It can be seen that under these conservative estimates for B_e and B_t, the relative error does not exceed 0.08 %. This equates to an absolute error of 40 nT, which is in the same order of magnitude as the crustal field.

The previous scenario is based on the worst case angle. In general a component of the platform field will be induced by the Earth field. This induced field will largely point in the same direction as the Earth field. Therefore the platform field will have a large component pointing in the same direction, meaning $-180° \ll \theta \ll 180°$. Then the relative error will drop significantly, as can be seen in Figure 4.5.

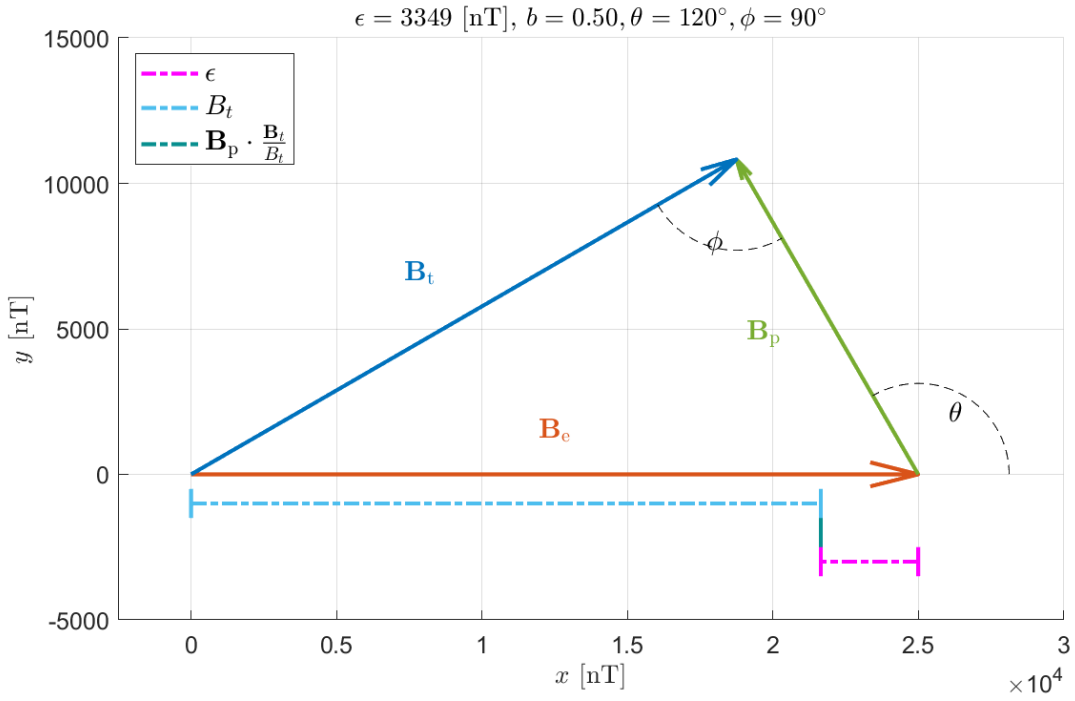


Figure 4.2: The approximation error in (2.11) is maximal when $\phi = 90^\circ$. In that case, $\mathbf{B}_p \cdot \frac{\mathbf{B}_t}{B_t} = 0$.

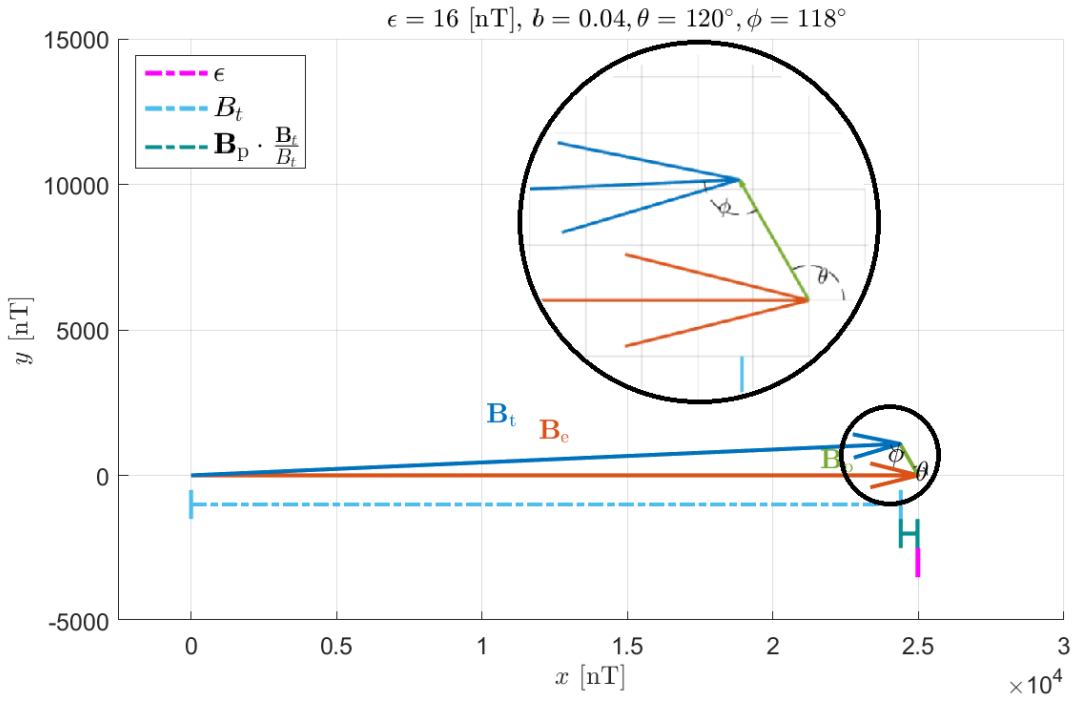


Figure 4.3: Interaction of vectors $\mathbf{B}_t$, $\mathbf{B}_e$ and B_p. When $B_e \gg B_p$, $B_t \approx B_e$ and the error becomes very small.

It can be shown that for small relative magnitude, the relative error grows quadratically. This is shown in Appendix C. It is also shown that for small relative magnitude, the relative error is upper bounded by

$$\epsilon_r \leq \frac{b^2}{2} \quad \text{for } b \ll 1.$$

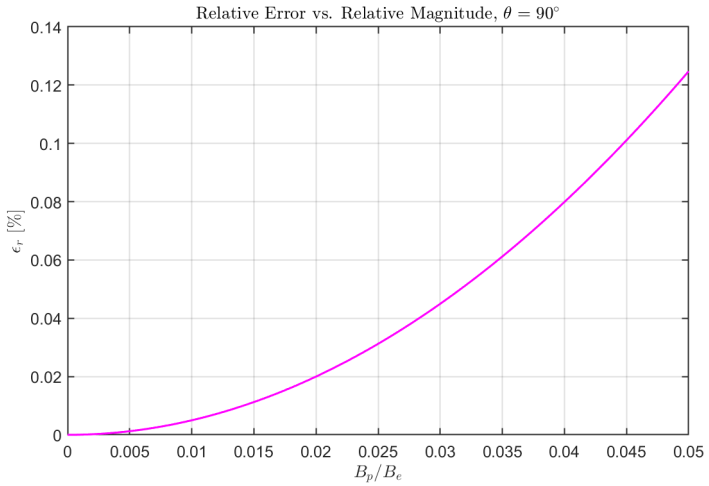


Figure 4.4: Relative error ϵ_r versus relative magnitude $\frac{B_p}{B_e}$. The relative error increases with the square of the relative magnitude.

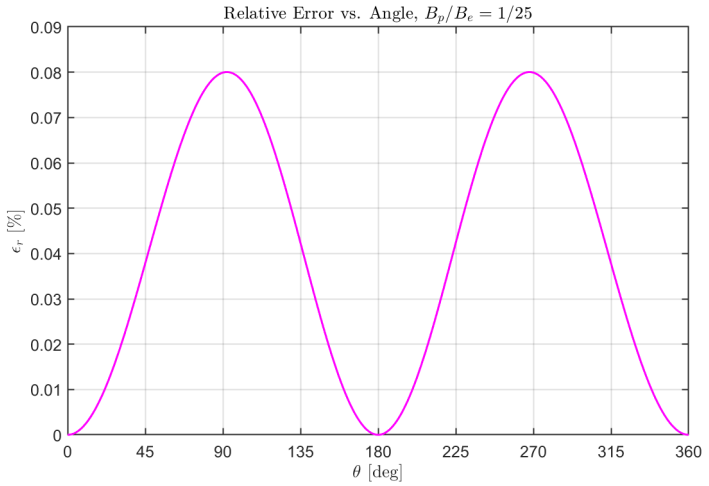


Figure 4.5: The relative error is plotted against the angle. When the angle between the Earth field $\mathbf{B}_e$ and the platform noise $\mathbf{B}_p$ is small, $\theta \approx 0$ (equivalently $\theta \approx 360$), the relative error is small as well. In practice this is the case due to the induced field, which has the same direction of $\mathbf{B}_e$, being a large part of $\mathbf{B}_p$[9].

From this, it can be determined how large the approximation error will be if the platform noise and Earth field magnitude are known. If this upper bound is less than the noise after compensation, the approximation can be ruled out as cause of the remaining noise. In the TNO data, platform noise is maximally 200 nT, while the Earth field is around 50000 nT. This results in a relative magnitude of $b = 1/250$, a maximal relative error due to approximation of about 0.0008%, and a maximal absolute error of 0.4 nT.

4.2. Tolles-Lawson Compensation

In this section a trivial solution is shown to exist in the frequency-based method for estimating the Tolles-Lawson coefficients. Then, the results of TL compensation applied to data is shown and the effect of regularization is investigated.

4.2.1. Trivial Solution

In literature there does not appear to be any mention of a trivial solution to (2.21). Here this trivial solution is derived. Applying bandpass filtering to both sides of (2.19) yields

$$\text{bpf}(|\mathbf{B}_e(t)|) = \text{bpf}\left(|\mathbf{B}_t(t)| - \left(\boldsymbol{a}^T \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|} + \mathbf{B}_t(t)^T \mathbf{b} \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|} + \dot{\mathbf{B}}_t(t)^T \mathbf{c} \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|}\right)\right), \tag{4.1}$$

$$0 = \text{bpf}\left(|\mathbf{B}_t(t)| - \left(\boldsymbol{a}^T \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|} + \mathbf{B}_t(t)^T \mathbf{b} \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|} + \dot{\mathbf{B}}_t(t)^T \mathbf{c} \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|}\right)\right). \tag{4.2}$$

Where the dependency on time has been explicitly written because the bandpass operation is applied to a continuous time signal.

Then setting the TL coefficients to $\boldsymbol{a} = \mathbf{0}_{3\times1}$, $\mathbf{b} = \boldsymbol{I}_{3\times3}$ and $\mathbf{c} = \mathbf{0}_{3\times3}$, (4.1) reduces to

$$0 = \text{bpf}\left(|\mathbf{B}_t(t)| - \mathbf{B}_t(t)^T \frac{\mathbf{B}_t(t)}{|\mathbf{B}_t(t)|}\right), \tag{4.3}$$

$$0 = \text{bpf}(|\mathbf{B}_t(t)| - |\mathbf{B}_t(t)|), \tag{4.4}$$

$$0 = 0, \tag{4.5}$$

which is a perfectly valid solution and will be the optimal value of the least-squares problem if $\text{bpf}(|\mathbf{B}_e(t)|) = 0$. These TL coefficients reduce (2.16) to $B_e \approx 0$, which demonstrates that this is not a physically correct solution. This trivial solution highlights the need for careful validation of the Tolles-Lawson coefficients $\boldsymbol{\beta}$.

The source of this trivial solution can be traced back to the assumption that (2.15) can be used as a definition for $\mathbf{B}_p$. If this assumption is not made the solution given in (4.3) does not exist. However, this assumption is necessary because otherwise the problem is that knowledge of $\mathbf{B}_e$ is required to calculate B_e.

Typically regularization is used when estimating TL coefficients [27], [42], [43], as discussed in Section 2.2.3. It can be shown that regularization can prevent the trivial solution from occurring. In the case of ridge regression, the trivial solution will have a loss value of $\lambda(1^2 + 1^2 + 1^2) = 3\lambda$. Because the trivial solution has zero error and the non-trivial solution is not perfect, the non-trivial solution will have a larger error. With this regularization, the non-trivial solution can then still be more optimal by having smaller β coefficients. This is the case as long as the loss is below 3λ.

Using the map-based TL from (2.27), discussed in Section 2.2.2, also avoids this trivial solution.

4.2.2. Tolles-Lawson Compensation

The TL compensation method, shown in Figure 4.6 is investigated. The vector magnetometer (VMR) inputs the normalized magnetic field vector $\hat{\mathbf{B}}_t$, while the total field magnetometer (OPM) gives the magnetic field magnitude B_t. The results for each manoeuvre can be seen in Table 4.1.

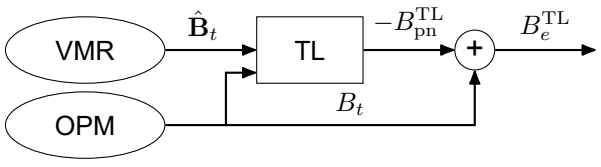


Figure 4.6: Diagram of Tolles-Lawson estimation.

In Figure 4.7 the estimate of the Earth field after compensation with the TL model can be seen. There is already significant improvement with respect to the original uncompensated signal. The heading related field change is removed quite well. Most of the sinusoidal variation is compensated for, although during the pitch phase, around `02:00` to `03:00`, some sinusoidal oscillation remains.

Table 4.1: Standard deviations and Improvement Ratios for TL compensation without regularization.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	37.96	14.40	2.64	31.52	14.05	2.24
6	5	43.15	8.41	5.13	-	-	-
8	7	35.32	6.30	5.61	-	-	-
1C	1	33.95	4.35	7.80	23.38	5.58	4.19
6C	5	45.94	11.06	4.15	-	-	-
8C	7	30.76	6.86	4.49	-	-	-
	Mean	37.85	8.56	4.97	27.45	9.81	3.22

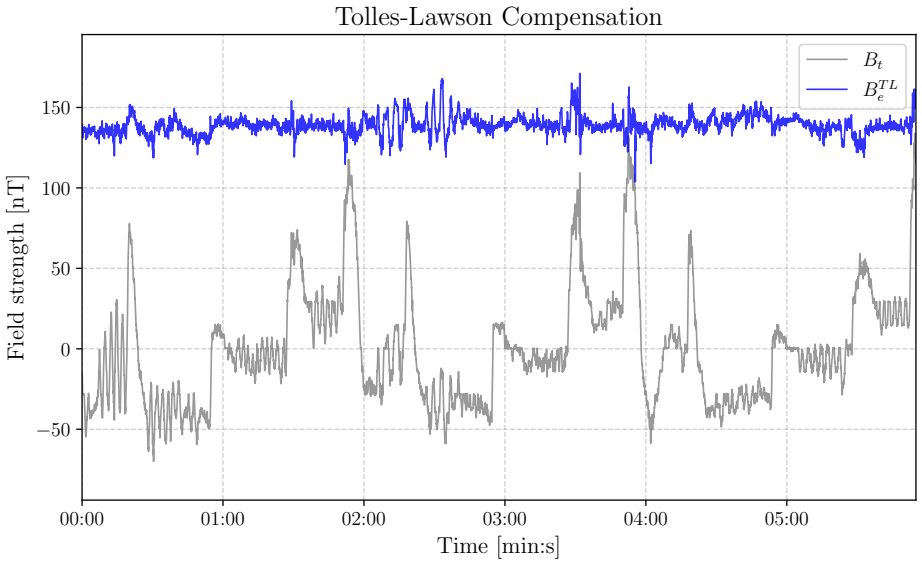


Figure 4.7: Tolles-Lawson compensation of manoeuvre 8.

In Figure 4.9 the 21 TL coefficients can be seen of the both left-hand side (LHS) and the right-hand side (RHS) for each manoeuvre. Shapes are used to indicate which manoeuvre each coefficient belongs to. The manoeuvres with the same shape but different coloring, being either filled or empty, belong to the same flight. From this figure it appears that these coefficients vary wildly between flights. This could indicate high multicollinearity as discussed in Section 2.2.3.

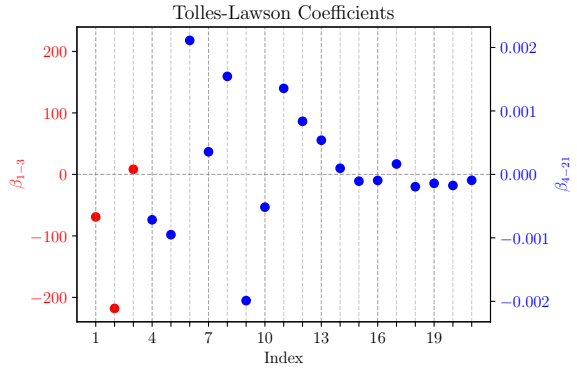
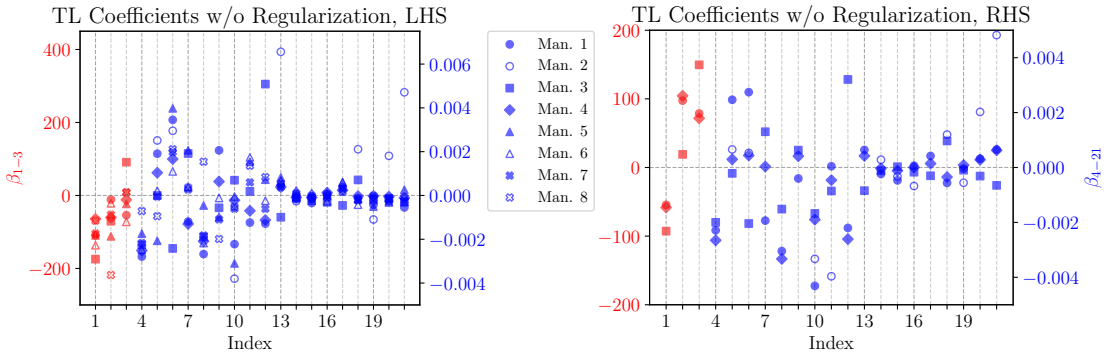


Figure 4.8: TL coefficients of manoeuvre 8.

To counter this large variation, regularization by ridge regression is employed as in (2.32). Due to the

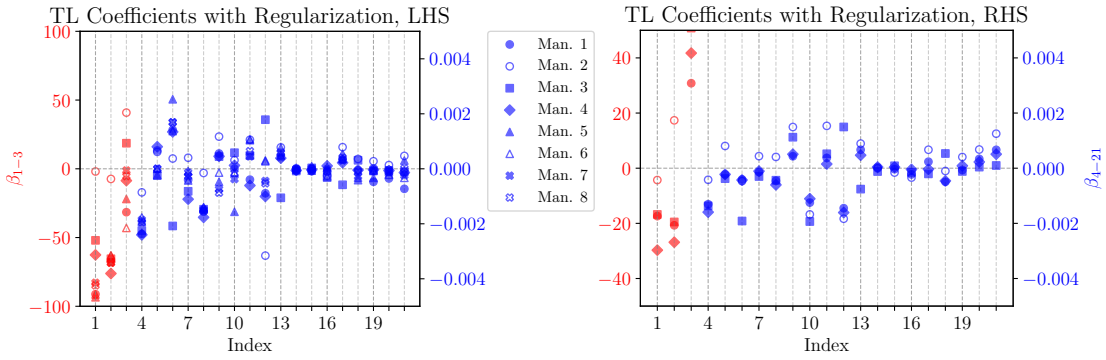
large difference in magnitude of columns, some columns of matrix $\mathbf{A}$ from (2.20c) are much more subject to regularization than others. To prevent this, an intermediate step was done where the columns were scaled. The scaling factors were found empirically by taking the mean of each column of matrix $\mathbf{A}$ from a representative manoeuvre. The regularization parameter λ was found by using the L-curve, as is explained in [44]. Ridge regression is a trade-off between minimizing the unregularized cost and minimizing the l_2-norm of the TL coefficients. The plot of these two variables is called an L-curve. Determining a good trade-off point is done by finding the λ that maximizes the curvature of the L-curve [44].



(a) TL coefficients of the left side for all flights.

(b) TL coefficients of the right side for all flights.

Figure 4.9: Tolles-Lawson coefficients for all flights, separated by left and right sides.



(a) TL coefficients of the left side for all flights.

(b) TL coefficients of the right side for all flights.

Figure 4.10: Tolles-Lawson coefficients for all flights, separated by left and right sides.

Looking at Figure 4.10, we can see the TL coefficients with regularization. Comparing Figure 4.10 to the unregularized coefficients in Figure 4.9 it becomes clear that the TL coefficients are not only much smaller valued, but more consistent between manoeuvres and between flights. This increased consistency within a flight and also between flights suggests that these coefficients might be closer the 'true' TL coefficients. However, there is still some variation between flights. It is unknown where the variation in TL coefficients between flights comes from. In [45] it is speculated that the variation stems from vibration of the ferromagnetic material during landing, which alters the ferromagnetic properties of the platform.

Another possible benefit of regularizing the TL coefficients is decreased overfitting. To test this, coefficients estimated during the first FoM of a flight were used to compensate for platform noise on the second FoM of a flight. Furthermore, the 'clean' manoeuvres were also compensated for using coefficients found on the first FoM of the same flight. In other words, compensation was tested on unseen data. The results of this test can be seen in Table 4.2. Here '**Man. #**' refers to the manoeuvre that was compensated and '**Cal. #**' refers to the manoeuvre that was used to estimate the TL coefficients.

Table 4.2: Improvement Ratios (IR) for TL compensation of the OPM sensor with and without regularization.

Man. #	Cal. #	IR LHS w/o Reg.	IR LHS w/ Reg.	IR RHS w/o Reg.	IR RHS w/ Reg.
2	1	2.64	2.61	2.24	2.10
6	5	5.13	5.18	-	-
8	7	5.61	5.54	-	-
1C	1	7.80	7.50	4.19	5.95
6C	5	4.15	4.15	-	-
8C	7	4.49	4.41	-	-
	Mean	4.97	4.90	3.22	4.03

Table 4.3: Standard deviations and Improvement Ratios for TL compensation with regularization of VMR signal.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	296.87	24.29	12.22	200.46	24.05	8.34
6	5	247.87	34.41	7.20	-	-	-
8	7	231.68	28.23	8.21	-	-	-
1C	1	264.43	14.34	18.44	160.09	18.46	8.67
6C	5	248.85	34.73	7.16	-	-	-
8C	7	222.24	28.16	7.89	-	-	-
	Mean	251.99	27.36	10.19	180.28	21.25	8.51

Next, the IR of the compensated signals with and without regularization were compared for the left- and right-hand side of every test manoeuvre. From the table, it can be seen that in most cases there is not a significant difference in IR on the left-hand side, giving either a slight increase or slight decrease. This suggests that the unregularized left-hand side TL coefficients did not overfit much to begin with. On the right-hand side there is an improvement, which could be because of decreased overfitting.

As discussed in Section 3.2.4, the OPM magnetic field sensor may be limited by an unusual quantization error. Therefore, another set of experiments was done where compensation is applied to the VMR instead. The results of the regularized TL applied to VMR data can be seen in Table 4.3. An example from manoeuvre 8 is shown in Figure 4.12.

Compared to compensation of the OPM sensor, the improvement ratio is much larger for the VMR compensated data. However, the standard deviation for the VMR data, both compensated and uncompensated is much larger. This is because the VMR has much larger variation due to the platform noise than the OPM. Still, this confirms that the VMR sensor can be used to investigate the efficacy of different noise compensation methods.

4.3. Extended Tolles-Lawson

In this work another variant is introduced where similarly to section 2.3.2 all drone inputs are considered for to be used to predict the platform noise. However the estimation of the coefficients belonging to the drone features happens simultaneously with the estimation of the TL coefficients, similarly to section 2.3.2. This should give a better fit to the data when solving the standard least squares problem. Similarly to Section 2.2.2 a scaling step is applied to adjust for unit correctness and scale of the columns of $\boldsymbol{A}_{ETL}$.

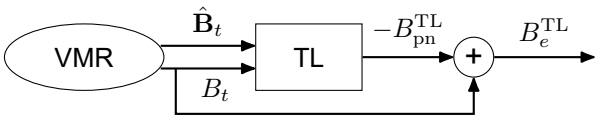


Figure 4.11: Diagram of Tolles-Lawson estimation using VMR.

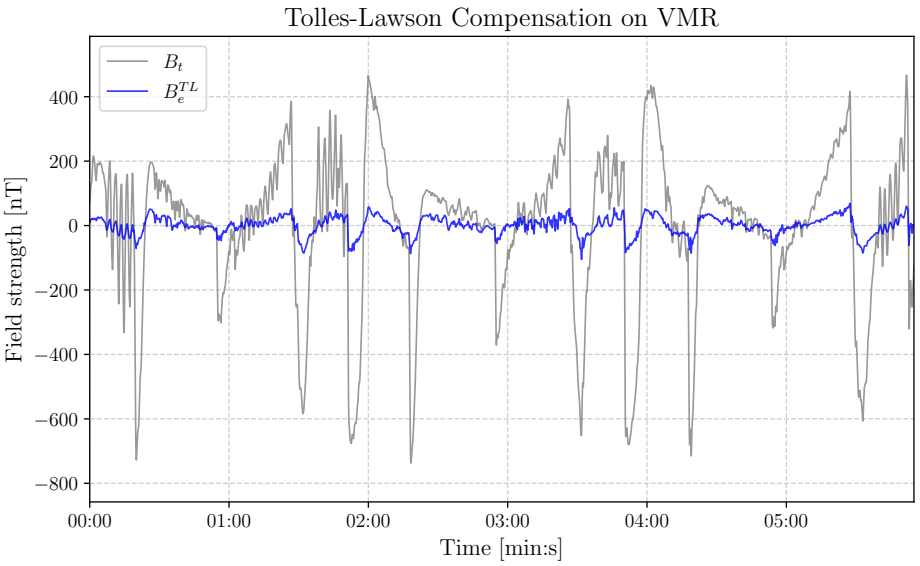


Figure 4.12: TL compensation of the VMR signal.

The variables in question are defined as

$$\boldsymbol{A}_{ETL} = \begin{bmatrix} \boldsymbol{A}_c & \boldsymbol{A}_r \end{bmatrix}, \tag{4.6}$$

for which the solution to the least-squares problem is

$$\hat{\boldsymbol{\beta}}_{ETL} = \boldsymbol{A}_{ETL}^{\dagger}(\boldsymbol{y} - \boldsymbol{c}_{\boldsymbol{B}_c}). \tag{4.7}$$

(4.8)

Now the ETL compensation method introduced in Section 4.3 is applied to the TNO dataset. This is done to validate the implementations and create a baseline against which other compensation methods can be compared. The diagram for this method is shown in Figure 4.13. In this method inputs are used as well. Drone inputs were selected based on the most linearly correlated signals, as discussed in Section 4.5. The same regularization approach was used as in the ridge regression regularized TL method. In Figure 4.14 the ETL compensated signal can be seen compared to the TL compensated signal. The ETL compensated signal appears flatter than the TL compensated signal. The pitch related oscillation in the center of the figure is also mitigated to a certain extent. Similar improvements were seen in other manoeuvres. In all except one case improvement in IR relative to the IR of the TL compensated manoeuvres is seen, as shown in Table 4.4.

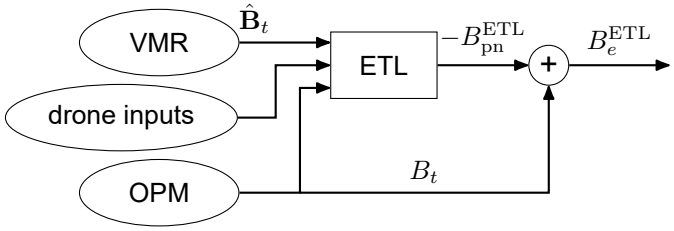


Figure 4.13: Diagram of Extended Tolles-Lawson estimation.

Now let us look at the results of ETL applied to the VMR sensor. These results can be found in Table 4.5, which also includes the TL results for comparison.

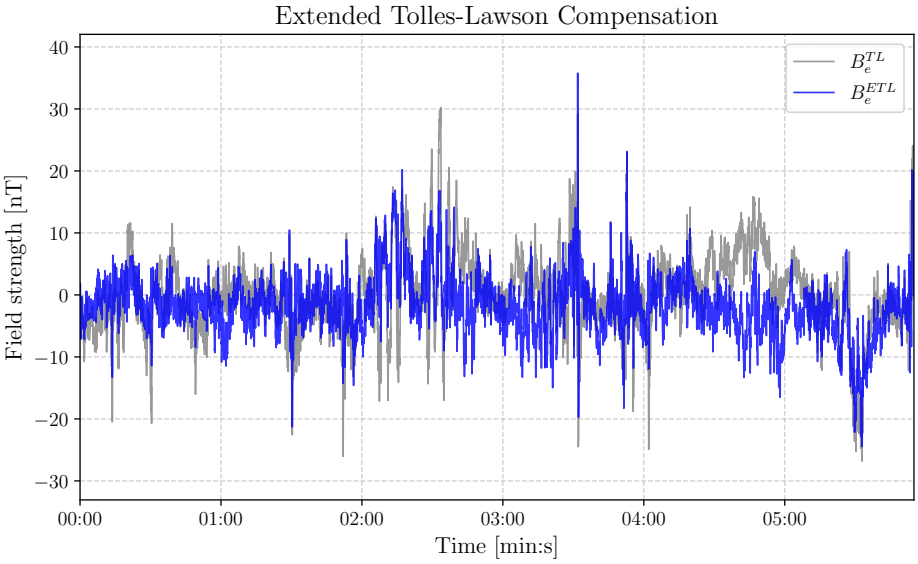


Figure 4.14: Extended Tolles-Lawson compared with regular Tolles-Lawson compensation.

Table 4.4: Improvement Ratios (IR) for TL and ETL compensation of the OPM signal.

Man. #	Cal. #	IR LHS TL	IR LHS ETL	IR RHS TL	IR RHS ETL
2	1	2.61	4.62	2.24	3.22
6	5	5.18	8.52	-	-
8	7	5.54	7.39	-	-
1C	1	7.50	6.09	4.19	4.63
6C	5	4.15	11.48	-	-
8C	7	4.41	5.79	-	-
	Mean	4.90	7.31	3.22	3.92

From this table it can be seen that ETL generally has a much larger IR than TL. What this increase in IR looks like can be seen in Figure 4.15. The increase of ETL IR relative to TL IR is also greater than in the OPM results. This could indicate that the VMR is closer to magnetic field sources on the platform. Another explanation is that, since the IR in the VMR results are calculated in a specific frequency range, the bandpassed magnetic field signal contains relatively more platform noise and therefore will benefit more from compensation. In any case, the compensation results here show an improvement when taking into account drone inputs, and provide a baseline to compare more advanced methods against.

4.4. Deep Learning

In this section, an MLP is designed to compensate for the platform noise in TNO data. First, the MLP itself is designed. Then, it is used to predict the residual after TL compensation. Next, it is used to predict the residual after ETL compensation. This is done for both OPM and VMR sensors. Finally, the results are analysed to guide further model design and feature selection in the sections that follow.

4.4.1. Network Design

To model the residual platform noise, a multilayer perceptron (MLP) was chosen as the neural network architecture. The MLP is a natural starting point due to its simplicity and general applicability to regression problems. While limited in its ability to capture temporal dependencies, the MLP has shown good performance in literature [17], [46].

The activation function used throughout the network is the sigmoid-weighted linear unit (SiLU), selected based on the findings of Gnadt et al. [47], who demonstrated its effectiveness in a similar context. Input

Table 4.5: Improvement Ratios (IR) for TL and ETL compensation of VMR signal.

Man. #	Cal. #	IR LHS TL	IR LHS ETL	IR RHS TL	IR RHS ETL
2	1	12.22	16.07	8.34	14.88
6	5	7.20	21.32	-	-
8	7	8.21	21.86	-	-
1C	1	18.44	22.74	8.67	20.31
6C	5	7.16	12.75	-	-
8C	7	7.89	17.84	-	-
	Mean	10.19	18.76	8.51	17.60

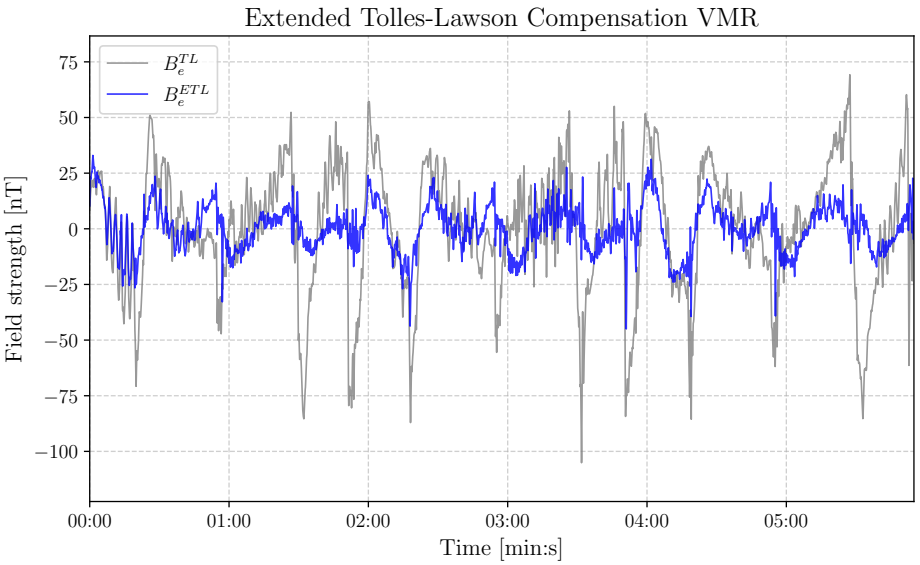


Figure 4.15: ETL vs. TL compensation of the VMR signal during manoeuvre 8.

features were normalised by removing the mean and scaling to unit variance, a standard preprocessing step that improves optimization stability and convergence speed.

The network is trained using the Adam optimizer, which uses adaptive learning rates and momentum [48]. Hyperparameters such as the number of layers, number of neurons per layer, learning rate, batch size, and number of epochs were tuned using the Optuna framework [49], which performs efficient hyperparameter optimization through Bayesian search. It was found that a single hidden layer with 11 neurons provided the best performance, and this architecture was used consistently for both the OPM and VMR compensation models.

The target for the neural network is the residual platform noise after TL compensation. The diagram in Figure 4.16 shows that the MLP uses drone inputs to predict the residual of the TL platform noise. This is then added to the TL predicted platform noise to calculate the total platform noise, which is then subtracted from the total magnetic field to get the Earth field. Later on we will see that this can also be done for the ETL model.

The results of the TLNN model are reported in Table 4.6, expressed in terms of improvement ratio (IR) and compared to the TL method. The neural network outperforms the standard TL model, which is not a surprise. Compared to the ETL results however, the results are similar for the left-hand side, and better for the right-hand side.

The TLNN method was also applied to the VMR, shown in Figure 4.17. As the VMR data is prone to drift, we should be careful about defining our target. Therefore we will use the bandpassed uncompensated signal as a reference and try to decrease the standard deviation of that. The frequency band was chosen to be 0.02-2Hz. The aim for this is to remove the low frequency drift, while still being low

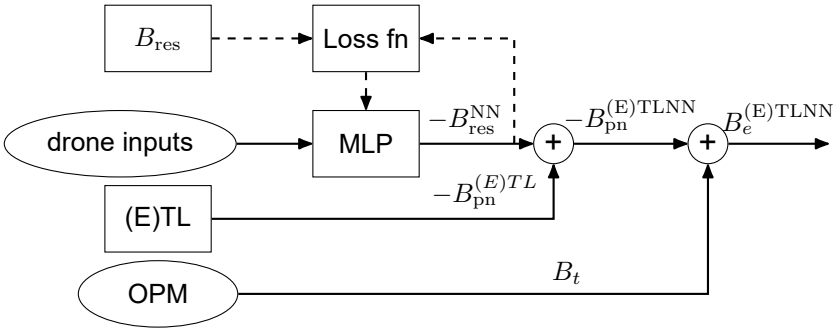


Figure 4.16: Diagram of (E)TLNN estimation. The dashed lines are only included during training.

Table 4.6: Improvement Ratios (IR) for TL and TLNN compensation of the OPM signal.

Man. #	**Cal. #**	IR LHS TL	**IR LHS TLNN**	**IR RHS TL**	**IR RHS TLNN**
2	1	2.61	6.30	2.24	6.03
6	5	5.18	8.51	-	-
8	7	5.54	8.67	-	-
1C	1	7.50	4.98	4.19	8.27
6C	5	4.15	8.89	-	-
8C	7	4.41	6.38	-	-
	Mean	4.90	7.29	3.22	7.15

enough to capture possible anomaly variation, and selecting a high enough frequency that the anomaly variation presumably does not reach.

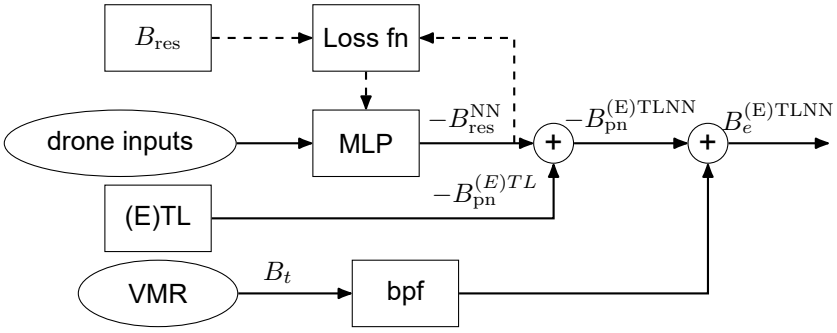


Figure 4.17: Diagram of (E)TLNN estimation for the VMR sensor. The dashed lines are only included during training.

Similar results were found for the TLNN method as in the case using OPM data. The NN improves compared to the TL but is generally worse than ETL. Furthermore, the NN seemed to to fit much better to the training data opposed to the test data.

To further investigate this overfitting, we can look at the learning curve. The learning curve is the epoch number of the training and validation data plotted against the loss. A representative example for one manoeuvre is given in Figure 4.18. If both training and validation curve decrease in unison, this indicates a good learning of the true underlying system. If the validation loss increases while the training loss decreases this indicates the model is overfitting to the training data. In Figure 4.18 it can be seen that even for NN the learning loss quickly decreases at the start and then continues to decrease. The validation loss however, quickly decreases and then almost no improvement is made. So there is still overfitting happening, but it is not necessarily decreasing performance on the validation set.

Experimenting with Optuna shows that increasing the number of layers decreases performance and increases overfitting. To improve our model further without overfitting we can do a number of things:

- Early stopping

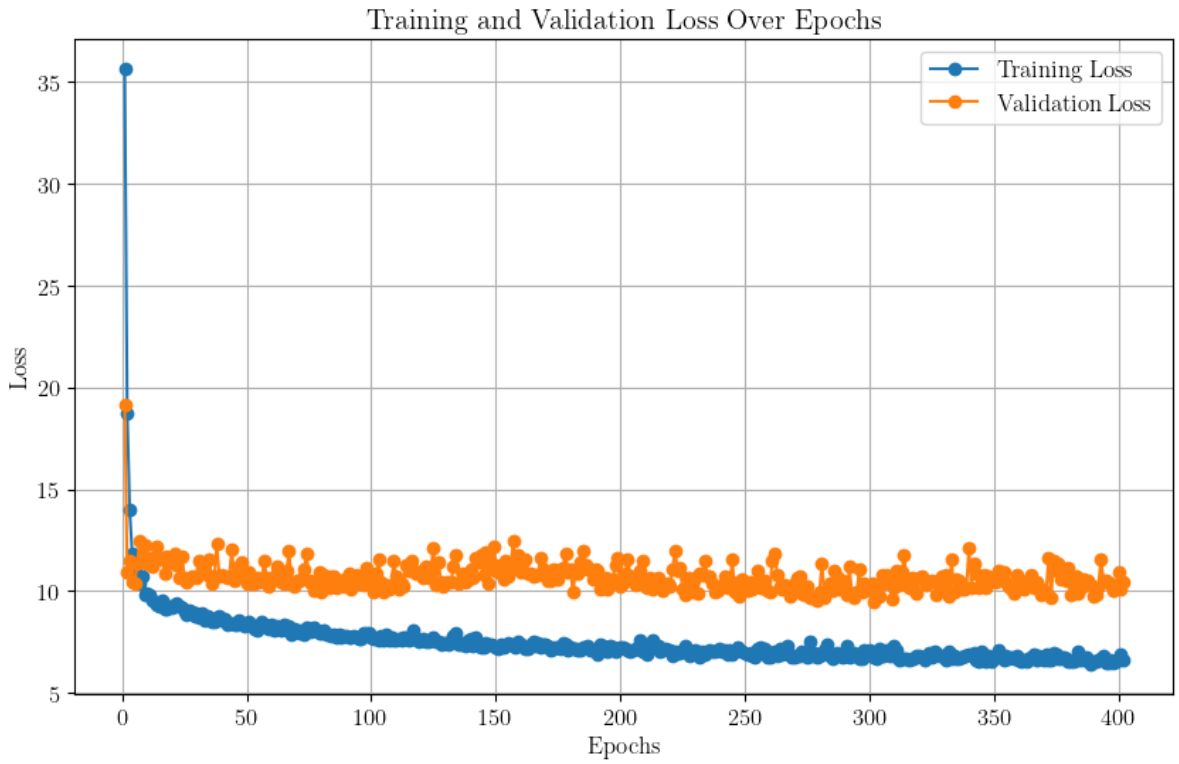


Figure 4.18: Learning curve for the NN model of manoeuvre 7.

- Increase training data
- Decrease model size
- Regularization by dropout
- Decrease number of features
- Make features more representative

A way to increase training data is to include data from different flights. The reason a model was trained for each manoeuvre is that the TL coefficient change between flights. However, there might still be some things that stay the same even between the flights. Then using the model trained on data from all flights simultaneously can be further refined by training on data from only one flight.

4.4.2. ETLNN

Another method we can consider is trying to fit the neural network to the residual after ETL compensation. Since ETL already provides a large improvement with respect to TL, this residual could be simpler to learn than the TL residual.
The results of this ETLNN method can be found in Table 4.7. From this it is clear ETLNN generally outperforms ETL. Especially on the right-hand side the increase in IR is significant. The ETLNN compensated signal next to the uncompensated signal can be seen in Figure 4.19.

As before, the ETLNN method is also applied to the VMR data. The result for this can be seen table 4.8. The improvement with respect to the ETL method is similar for the VMR and OPM for the left-hand side. The right-hand side sees similar improvement as the left-hand side.

In the model tested on manoeuvre 1C, there was a very large IR of 31.18, which comes down to a standard deviation of only 8.5 nT. This is quite significant, compared to the uncompensated standard deviation of 264 nT. This comparison is made more clear in Figure 4.20.

Table 4.7: Improvement Ratios (IR) for ETL and ETLNN compensation of the OPM signal.

Man. #	Cal. #	IR LHS ETL	IR LHS ETLNN	IR RHS ETL	IR RHS ETLNN
2	1	4.62	6.85	3.22	6.80
6	5	8.52	9.51	-	-
8	7	7.39	10.00	-	-
1C	1	6.09	7.34	4.63	10.07
6C	5	11.48	12.72	-	-
8C	7	5.79	6.81	-	-
	Mean	7.31	8.87	3.92	8.44

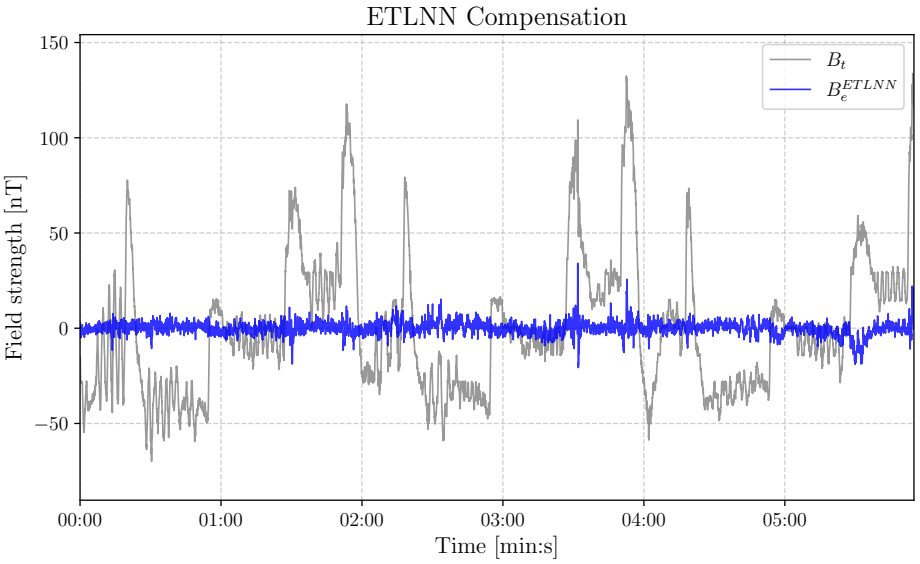


Figure 4.19: ETLNN compensated magnetic field versus the uncompensated magnetic field of manoeuvre 8.

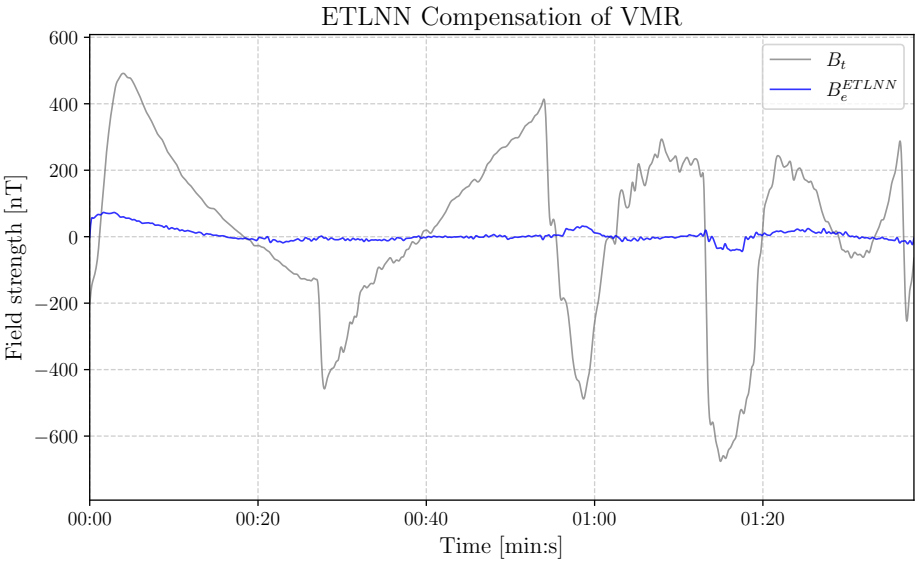


Figure 4.20: ETLNN compensated magnetic field versus the uncompensated magnetic field of manoeuvre 1C.

4.4.3. Analysis

The average IR from the TL compensation of the OPM signal is 4.48, the average IR from the ETLNN compensation of the OPM signal is 8.77. So the ETLNN models improve on the TL models with a factor

Table 4.8: Improvement Ratios (IR) for ETL and ETLNN compensation of VMR signal.

Man. #	Cal. #	IR LHS ETL	IR LHS ETLNN	IR RHS ETL	IR RHS ETLNN
2	1	16.07	19.37	14.88	18.87
6	5	21.32	23.67	-	-
8	7	21.86	27.25	-	-
1C	1	22.74	31.18	20.31	22.95
6C	5	12.75	12.65	-	-
8C	7	17.84	19.18	-	-
	Mean	18.76	22.22	17.60	20.91

1.95 on average. The average IR from the TL compensation of the VMR signal is 9.77, the average IR from the ETLNN compensation of the VMR signal is 21.89. So the ETLNN models outperform the TL models by a factor 2.24 on average when applied to the VMR.

The hyperparameters were chosen using Optuna. This search included a broad search area searching learning rate, number of epochs, batch size, layer width and number of layers. The results suggest that the model is prone to overfit as the best network was the one with lowest number of epochs, small batch size, and only a single layer with 11 nodes. This could explain why training end-to-end with an MLP, i.e. without the (extended) Tolles-Lawson model in between did not perform well.

To counteract overfitting some experiments were done using dropout layers, which during training randomly eliminate connections between features and neurons. However, this did not improve or degrade performance significantly.

The good performance increase when adding a neural network to predict the residual of a base linear model suggests that perhaps some kind of hierarchical model might be suited for this task, where models are successively trained on the residual of the previous model. The ETLNN model is already a kind of hierarchical model with first the ETL model predicting the noise and the NN predicting the residual. A step further would be to add another model, perhaps another NN, which predicts the residual of the first NN.

Another observation is that ETLNN increases more w.r.t. TL on the VMR data than the OPM data. This could be because the VMR platform noise is only predicted in a frequency range between 0.02 and 2 Hertz. This could make the signal less complex and easier to predict.

4.5. Feature Selection

Understanding which drone inputs are most relevant for predicting platform noise is essential for building effective compensation models. A common first step in the literature is to compute correlations between inputs and the magnetic signal. While this provides initial insight, it does not account for how these signals project onto the external magnetic field or how they interact in multivariate models. This section begins with a correlation analysis of the raw inputs, then extends the analysis by projecting these inputs onto the direction of the magnetic field, better reflecting the physics underlying the platform noise. Finally, SHAP values are used to assess the true importance of each feature in both linear and nonlinear models, providing a more complete picture for informed feature selection.

The result of the correlations between every drone input and the OPM magnetic field signal can be seen in Figure 4.21. Only inputs which which were in the top 20 highest correlations for either manoeuvre are shown. The result of the Spearman correlations can be seen in Figure 4.22. This result resembles the linear correlation plot of Figure 4.21 very closely.

As was shown in (2.9), the platform noise is equal to the dot product between the platform field and the total field direction. In other words, the platform noise is the magnitude of the projection of the platform field onto the total field. Now if we consider that the drone inputs stay in the reference frame of the platform, we should also project the drone inputs onto the total field. This means we can multiply the drone inputs by the direction cosines, i.e. the direction of the total field, we take into account the angle between the total field and the platform. The result of this can be seen in Figures 4.23 and 4.24. An

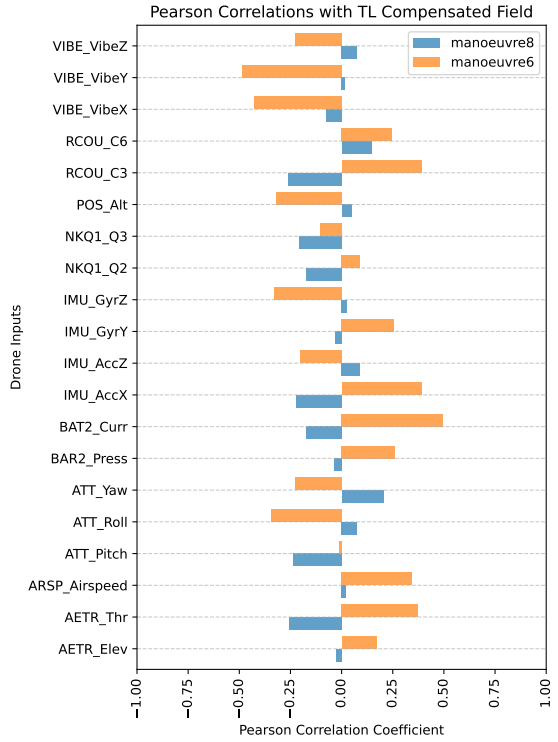


Figure 4.21: Pearson correlation coefficients of various drone inputs with the magnetic field captured on two flights.

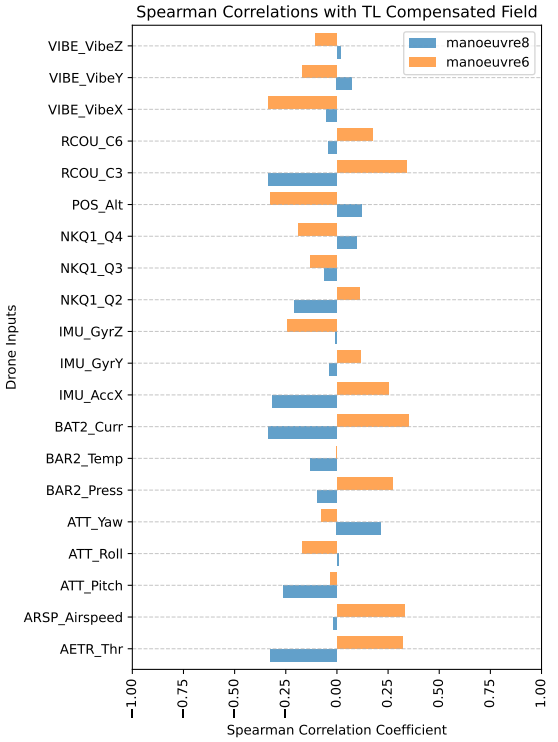


Figure 4.22: Spearman correlation coefficients of various drone inputs with the magnetic field captured on two flights.

unprocessed input has suffix `_1`, an input multiplied by the first, second or third direction cosine has suffix `_2`, `_3` and `_4`. Furthermore, since the ETL model also takes into account drone input derivatives multiplied by direction cosines, these features are also taken into account and are given suffix `_5`, `_6` and `_7`. From these figures it can be seen that the top 20 features for both type of correlation consist mostly of features related to:

- Engine current and throttle (`Batt2_Curr`, `RCOU_C3`, `AETR_Thr`)
- IMU data (`GyrX/Y/Z`, `AccX/Y/Z`, `VIBE_VibeX/Y`)
- Attitude (`ATT_Roll`, `ATT_Pitch`)

The highest correlated feature is the acceleration in the X direction. This acceleration is caused by an increase in engine power which is related to the also highly correlated battery current.In the 20 highest correlated features the drone input derivatives are absent. This is evidence that the derivative terms might be superfluous. It should be noted that some features might be purely becoming highly correlated due to being multiplied by direction cosine terms, which could themselves be correlated.

These correlations give an indication how much each feature is linearly or monotonically related to the platform noise, but they do not give the complete picture. This is because the features are not necessarily linearly independent. Furthermore, the features could be influencing platform noise at the same time, which the correlations cannot account for. Finally, they do not show what the different models actually use to predict the platform noise. Neither can they account for nonlinear behaviour, like that of neural networks.

4.5.1. SHAP for TL

A measure of how much each feature influences a model is the SHAP value, which is introduced in Chapter 2. We begin by examining the contribution of the standard TL features to the model output using SHAP values in order to investigate the effect of the standard TL features on model output. This shows the use of SHAP values for feature selection. Previously, the TL coefficients were shown.

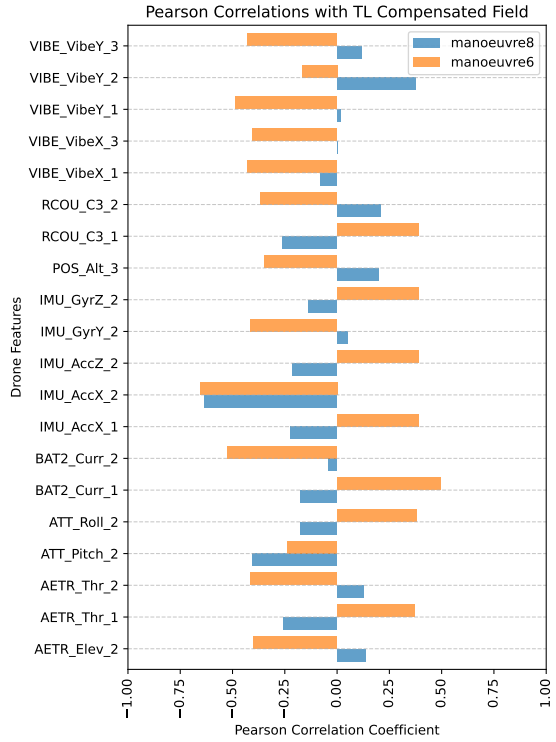


Figure 4.23: Top Pearson correlation coefficients of various drone features with the magnetic field captured on two flights.

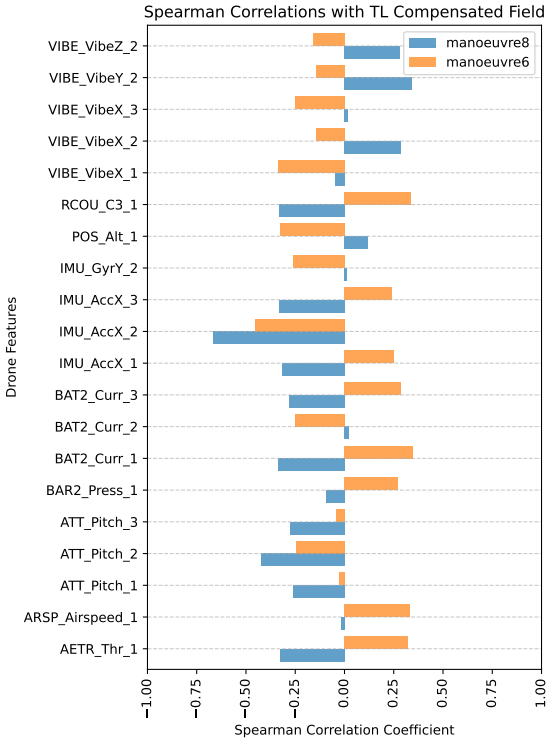


Figure 4.24: Top Spearman correlation coefficients of various drone features with the magnetic field captured on two flights.

However, this obscures the true effect on the output of the model. SHAP gives a better measure for this, as it can show how many nanoTesla specific features have changed the outcome of the prediction. In Figure 4.25 this effect is shown for every feature in the form of a distribution of SHAP values. The features are sorted by the mean absolute SHAP value, so the largest average impact on the model prediction. The feature values, indicated by color, can be seen to change linearly with the SHAP value. This is because we are using a linear model. We can also see that `TL20` and `TL21` have some outliers. This could be because the numerical derivative introduces some error. It looks like the top features are all occupied by either permanent or induced TL features, `TL1` – `TL3` and `TL4` – `TL12` respectively.

Instead of looking at each individual TL feature, we can look at the permanent, induced and eddy-current groupings in Figure 4.26. From this we can see that the biggest terms are usually the permanent features. The induced features also have a relatively large effect. The eddy-current features have very little effect compared to the permanent and induced features. The small effect of the eddy-current features could have multiple causes. First of all, since eddy currents are induced by a changing magnetic field in a conductive material, it could be that there is not a significant conductive loop near the sensor. In this case we should also not expect large eddy-currents induced by drone inputs, since they too require a conductive loop near the sensor to have an effect. Another possible cause of the small eddy-current feature effects has to do with the derivative which is necessary to calculate the eddy-current features. Numerically calculating the derivative is subject to large influence of high-frequency noise. A fast changing measurement due to random noise will have a large effect on the derivative even though this is not causes by a magnetic field.

4.5.2. SHAP for ETL

The top features for the ETL method, shown in Figure 4.27, beside the standard TL terms, consist mostly of terms relating to the attitude (quaternions `AHR2_Q1/Q4`, yaw angle `ATT_YAW`) and engine throttle (battery curent `BAT2_Curr` and `AETR_Thr`). Notably there are few inputs that directly affect the model, those with suffix `_1`. Furthermore there are features related to the derivative of the drone inputs, i.e. those with suffix `_5`, `_6` or `_7`. Most of the features relate to the drone inputs multiplied by the direction

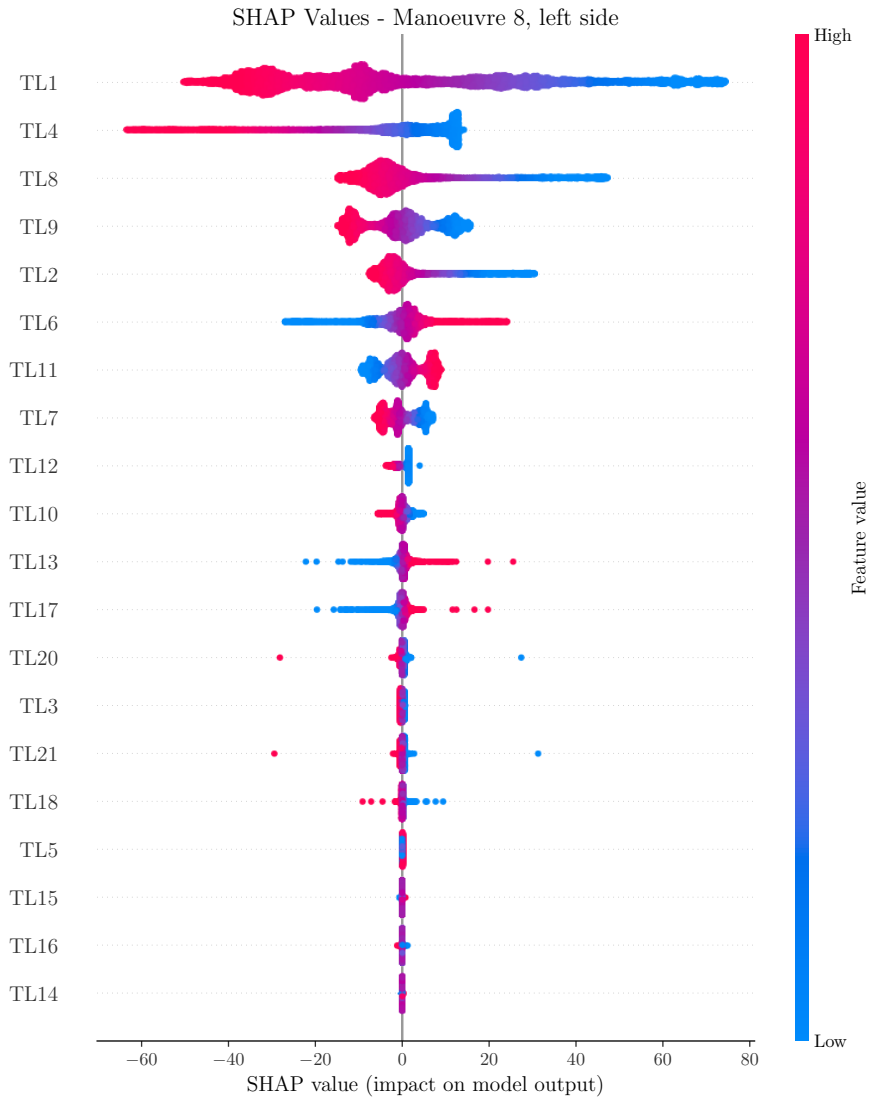


Figure 4.25: SHAP values for manoeuvre 8.

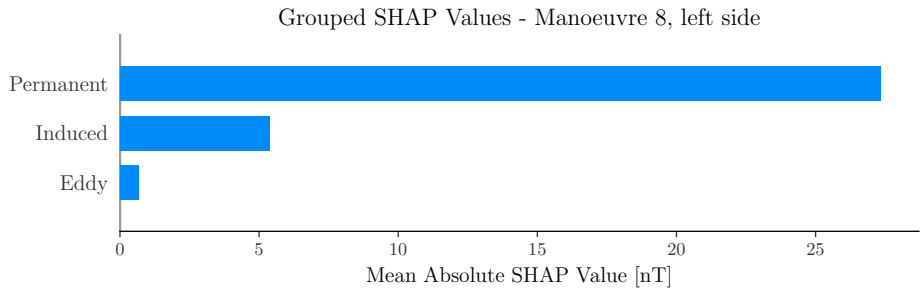


Figure 4.26: Mean absolute SHAP values for the grouped terms for manoeuvre 8.

cosines, which are those with suffix _2, _3, or _4.

To further investigate the effect of the different sets of features we can group them by the TL features, raw drone inputs (group 1), drone inputs multiplied by direction cosines (group 2), and derivative drone inputs multiplied by direction cosines (group 3). This is shown in Figure 4.28. From this it is clear the

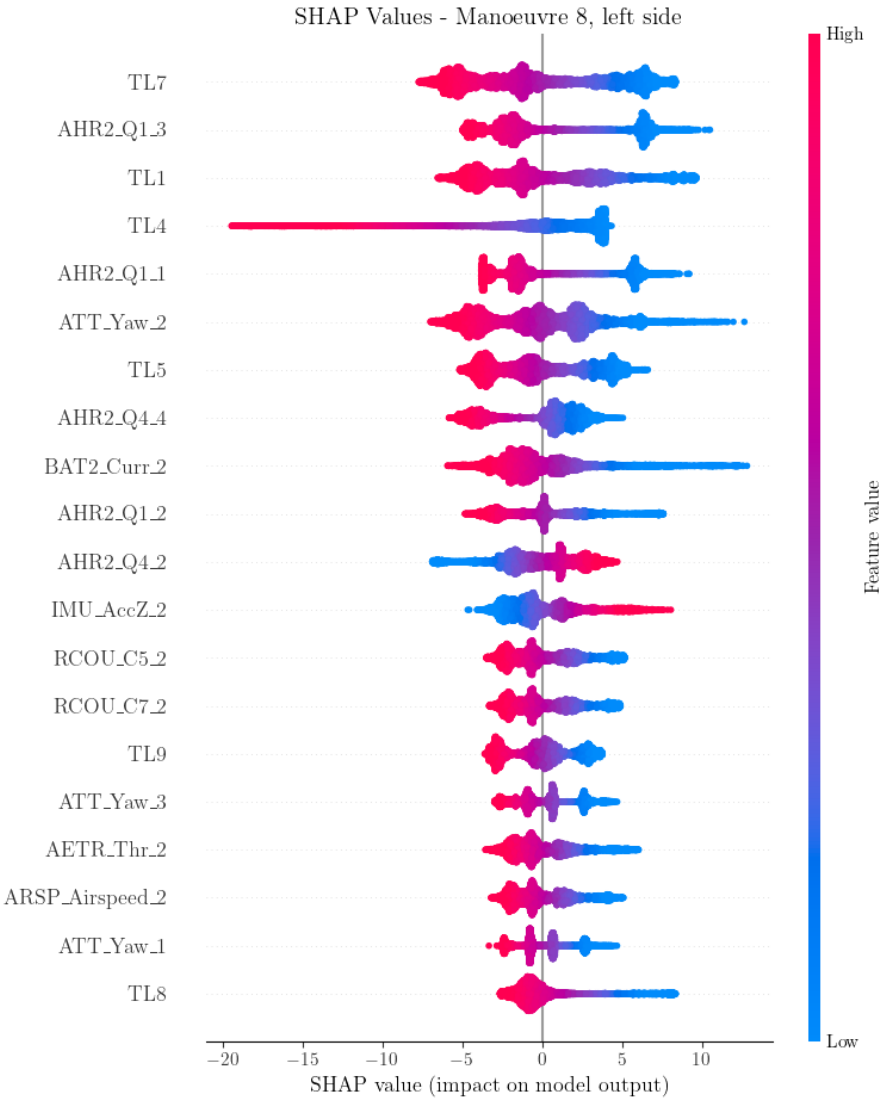


Figure 4.27: SHAP value distribution of ETL model trained on manoeuvre 7 applied to manoeuvre 8.

group 3 features have very little effect. This is similar to what we saw when analyzing the different TL components. Next there are the group 1 components. The effect of this group is caused mainly by the angle features. From a physical perspective, it makes sense that this group has little effect, as the fact that drone inputs are in the platform reference frame is not taken into account. Furthermore, the angle features could have an effect because they describe a difference between the platform reference frame and the Earth reference frame. We should then expect the group 2 features to have a much larger effect, since they are in the correct reference frame, which is indeed the case. The group 2 features, representing drone inputs projected onto the field direction, exhibit a greater influence than the TL features, suggesting that they may capture similar or addition physical effects more effectively. Furthermore, the effect of the TL features is diminished with respect to the effect of the TL components in the standard TL approach. This could be explained by the group 2 features taking over some of the impact of the TL features, especially the permanent TL features which are just the direction cosines. If there is a drone input which is approximately constant throughout and is then multiplied by the direction cosines to produce features, these features will be linear with the permanent TL features.

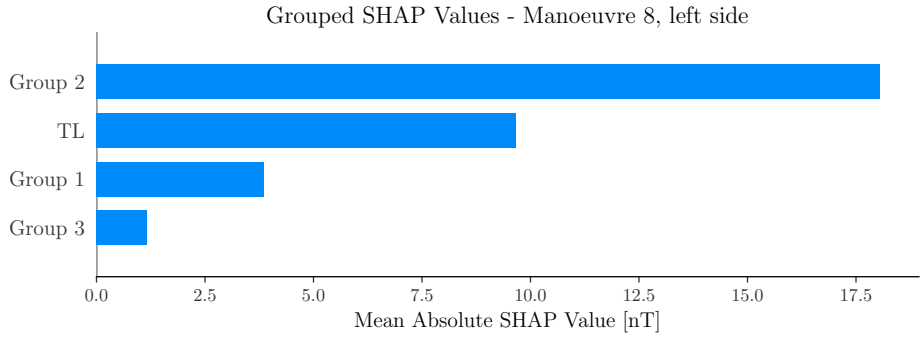


Figure 4.28: Mean absolute SHAP values for the grouped terms for manoeuvre 8.

4.5.3. SHAP for (E)TLNN

In Figure 4.29 the SHAP values for each NN feature can be seen. To make conclusions about the importance of these features, the features are grouped and their SHAP values are summed and averaged. The grouping by suffix is shown in Figure 4.30. Grouping by drone input is shown in Figure 4.31. Only the SHAP values for manoeuvre 8 are shown. This is mostly representative of the models used on the other manoeuvres. The set of features with the highest average absolute SHAP value was always those in group 2. In some models those with in group 1 also had a significant contribution. This contribution consisted mostly of features having to do with drone attitude. The group 3 features, which are drone input derivatives multiplied by direction cosines, consistently had the lowest SHAP values.

In Figure 4.32 the SHAP values for the ETLNN model are shown. The SHAP values grouped by drone input and suffix are shown in Figure 4.33 and 4.34 respectively. Interestingly, in the SHAP values for the ETLNN model, the group 1 features become more important. Furthermore, in the SHAP values for manoeuvre 8 there is a large importance of the yaw features. In general the derivative terms are again the least important, while attitude terms, especially yaw, have large influence.

4.5.4. Feature selection

The SHAP values discussed for both ETL and NN will inform the selection of features for further models. In both ETL and NN we saw low importance of derivative drone features. This motivates to leave out these derivatives. Furthermore, for both ETL and NN there was some importance of direct drone inputs (group 1 features), but they came mostly from attitude related features. This motivates the choice to use the attitude information directly in the form of direction cosines. This is already an input used by Tolles-Lawson, however a non-linear relation could exist which could be captured by the neural network. Such non-linear relationship might also exist for other TL components, like the induced and eddy-current components. That is why the induced TL features are also added as a feature for this next model. The eddy-current components showed little importance in TL so this will be not used as a feature in the NN.

To summarise, feature analysis reveals which drone inputs contribute most to magnetic field prediction, motivating a reduced feature set. This refined input will be used in subsequent retraining of both ETL and NN models for improved performance and interpretability.

4.6. Feature-Informed NN

Overfitting can negatively impact model performance. One reason for overfitting could be having too many parameters which cause the model to be highly flexible and fit to random noise, instead of fitting to true platform noise caused by drone inputs. That is why here a feature selection is used to train models which will be compared in terms of performance and amount of overfitting. Models fit to OPM data and models fit to VMR data will be examined separately. Four different feature sets were evaluated to study the impact of feature selection on model performance and overfitting. Feature selection will be done for the best performing model, which is ETLNN. The base feature set, FS0, included all features

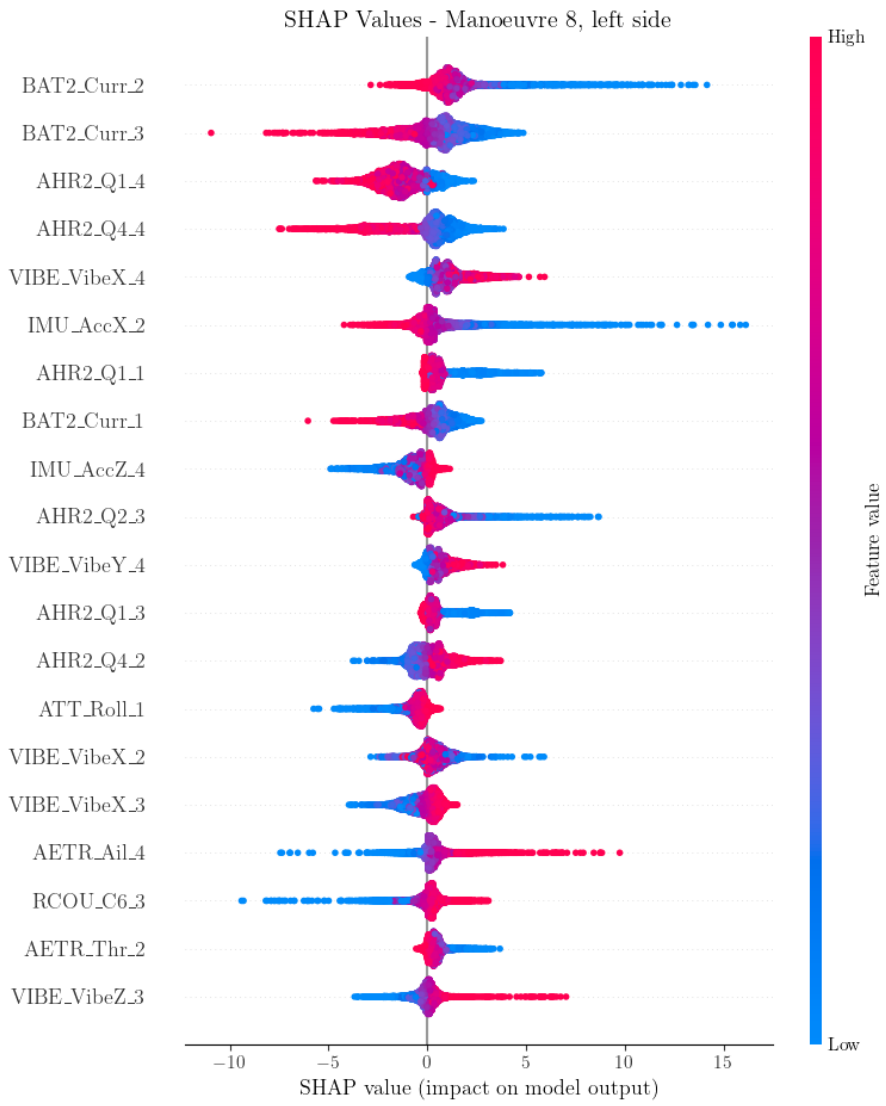


Figure 4.29: SHAP value distribution for top feature in manoeuvre 8 for the TLNN model. The top features consist of group 1 and group 2 features.

previously discussed, including both raw inputs, inputs multiplied with direction cosines and derivatives. To assess the influence of derivative features specifically, FS1 was created by removing all derivative-based features from FS0. Building on FS1, SHAP values were computed, and two reduced subsets were created: FStop50, which contains the top 50% most important features, and FStop10, which contains only the top 10%. The top 10% of features were `AETR_Thr_1`, `IMU_AccZ_1`, `ARSP_Airspeed_1`, `ATT_Yaw_1`, `IMU_AccZ_3`, `IMU_AccX_1`, `ATT_Pitch_2`, `AETR_Thr_3`, `BAT2_Curr_1` and `IMU_AccX_2`. Both FStop sets exclude derivatives and represent increasingly aggressive levels of pruning. These feature sets enable a comparison between full-feature and pruned models, isolating the effects of feature reduction on compensation and overfitting.

4.6.1. Feature-Informed ETLNN for OPM data

First the group 3 features were removed. The tables with the full results of each model trained for each combination of manoeuvre, left and right side, and sensor type is given in Appendix D. To evaluate the compensation performance for the OPM sensor, average improvement ratios (IR) were computed for both the left and right sensor sides across various feature sets, as can be seen in Table 4.9. Feature set FStop50 achieved the highest improvement on the left-hand side (IR = 8.95), slightly outperforming

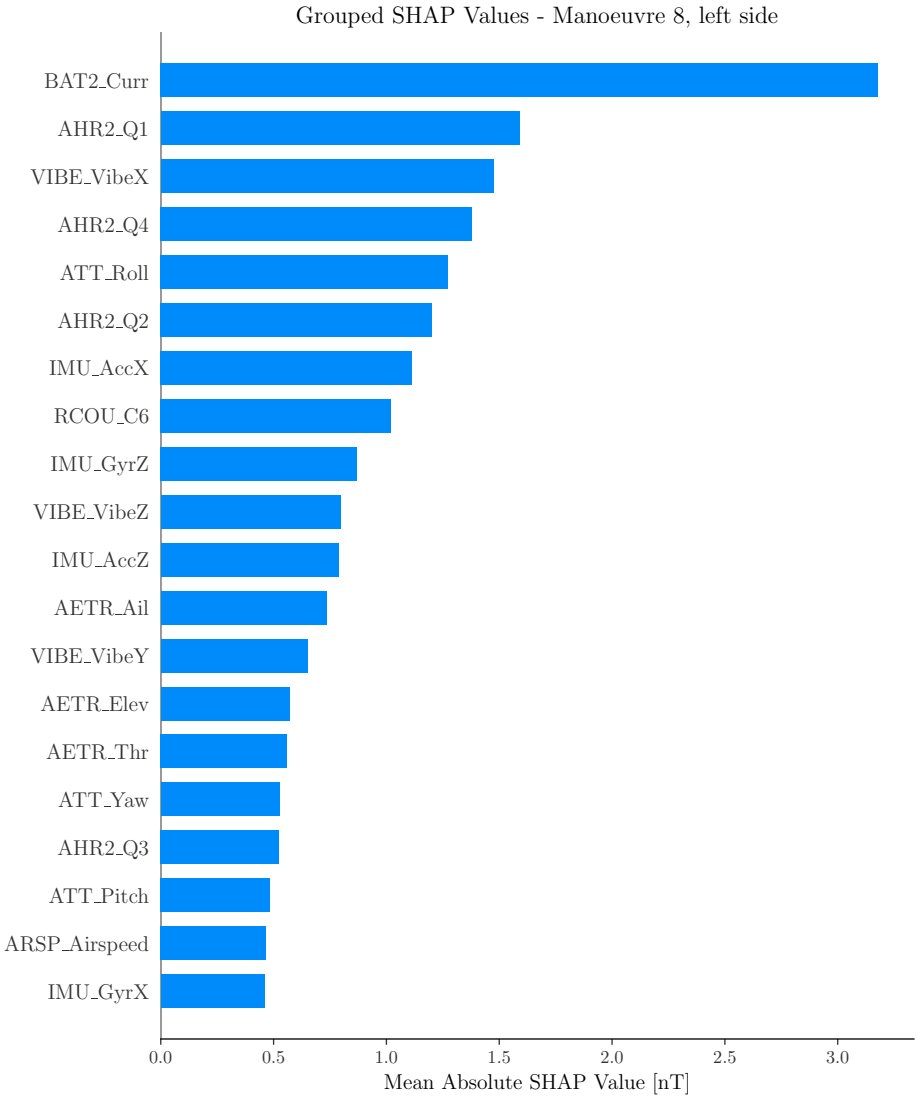


Figure 4.30: Average absolute SHAP value for top features in manoeuvre 8 for the TLNN model, grouped by drone input.

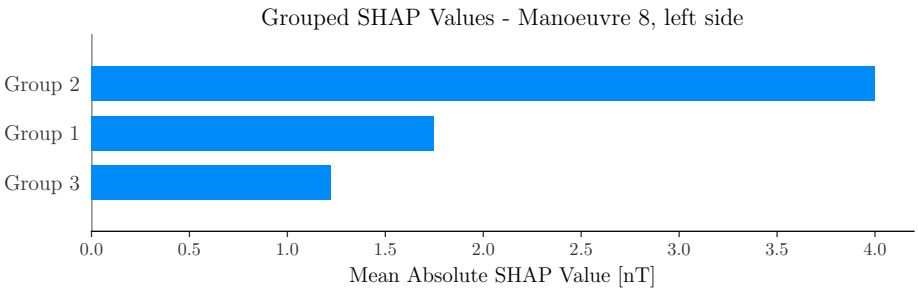


Figure 4.31: Average absolute SHAP value for top features in manoeuvre 8 for the TLNN model, grouped by suffix group. The derivative features of group 3 has the lowest contribution on average.

FS0 and FStop50. However, the difference is not significant. The FStop10 set showed a notably lower improvement (IR = 7.72), suggesting that reducing the number of features too aggressively can hurt model performance. For the right-hand side, FS0 performed best (IR = 8.44), while FStop10 again performed the worst (IR = 5.31). These results suggest that derivative features contribute positively to

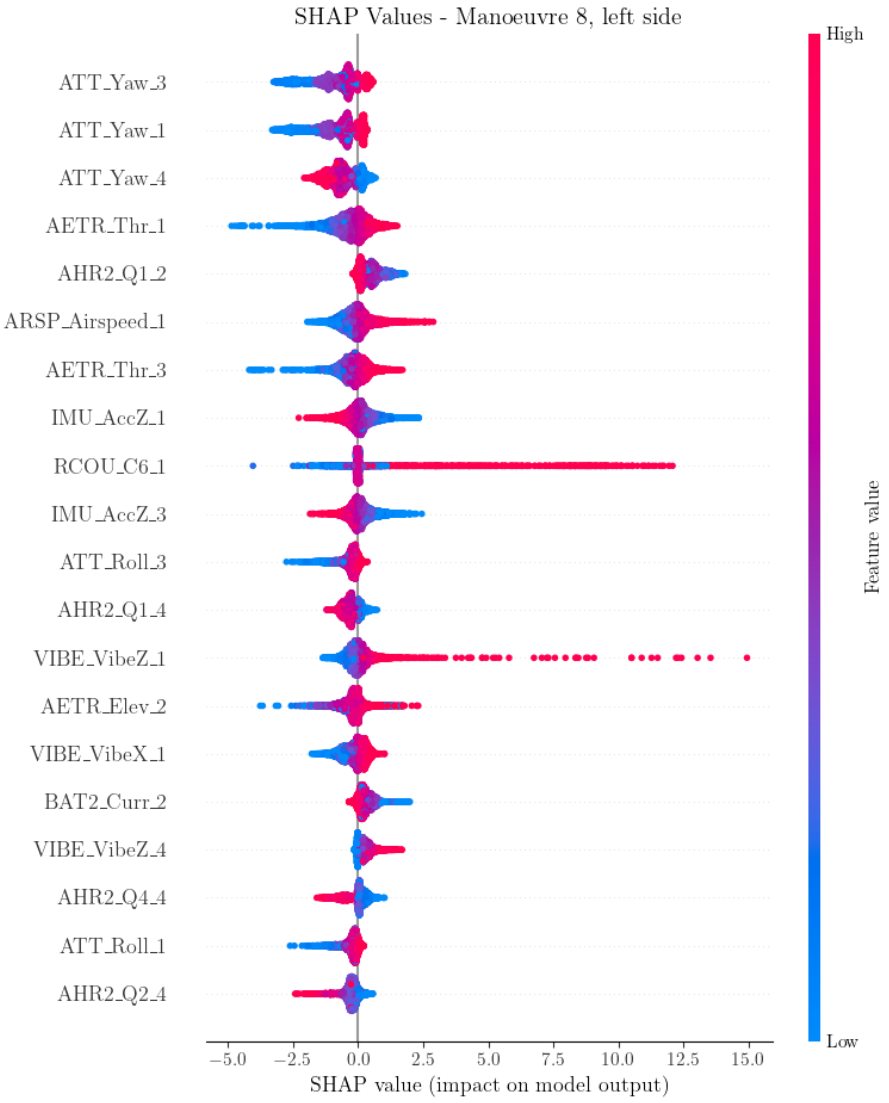


Figure 4.32

the model's effectiveness on the right-hand side but did not contribute on the left-hand side. Since the right-hand side only has data for manoeuvre 1 and 2, the variance in model performance is expected to be larger, which could explain the decrease in IR on the right side but not the left side.

To assess model generalization and evaluate our intermediate goal of reducing overfitting, the overfit ratio (OR) is introduced, which is the ratio of training to test set standard deviations (OR),

$$OR = \frac{\sigma_{test}}{\sigma_{train}}, \tag{4.9}$$

where higher OR indicates more overfitting. The OR for various feature sets can be seen in Table 4.10. For the left-hand side, all feature sets resulted in relatively mild overfitting, with OR values ranging between 1.22 and 1.33. The lowest overfitting occurred with FStop50 (OR = 1.22), while FS1 showed the highest (OR = 1.33). This suggests that feature pruning can help slightly reduce overfitting. On the right-hand side, which was only evaluated for one manoeuvre pair, FS1 showed the highest OR (1.75), while FStop10 had the lowest (1.55). Although the differences are modest, they suggest that smaller feature sets may help mitigate overfitting, but this may come at the cost of performance.

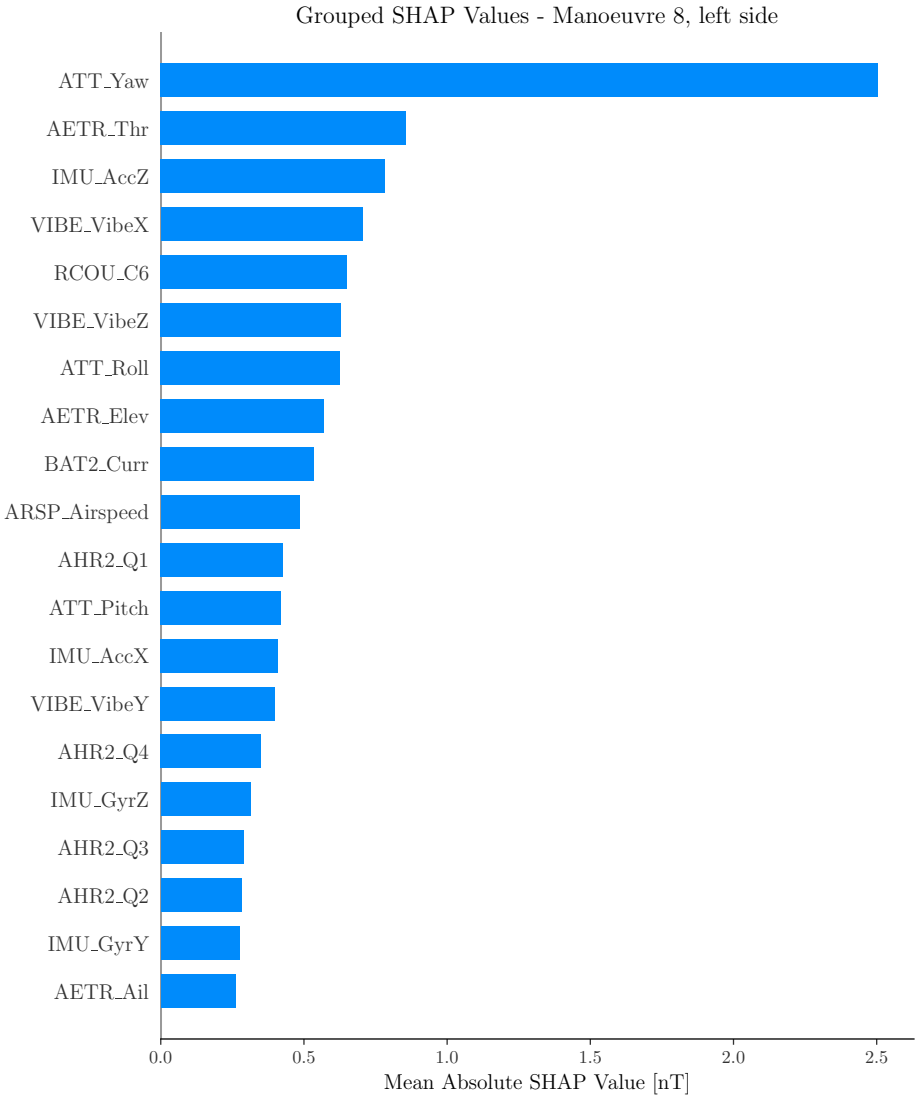


Figure 4.33

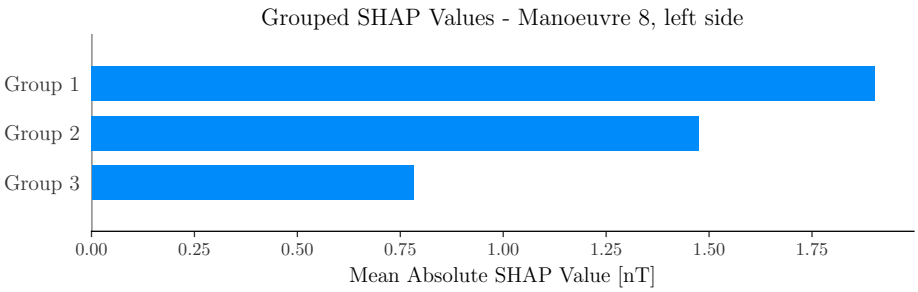


Figure 4.34

4.6.2. Feature-Informed ETLNN for VMR data

The VMR sensor showed substantially higher improvement ratios (IR) than the OPM sensor across all feature sets, reflecting the stronger signal and greater potential for compensation. As shown in Table 4.11, FS1 yielded the highest IR on the left-hand side (22.58), followed closely by FS0 (22.22) and FStop50 (22.19). FStop10 resulted in a slightly lower IR of 22.00, but the differences between

Table 4.9: Average Improvement Ratios (IR) over all manoeuvres for Left and Right sensor sides, Feature Sets, Sensor Type: OPM

Feature Set	Avg IR LHS	Avg IR RHS
FS0	8.87	8.44
FS1	8.91	7.28
FStop50	8.95	7.21
FStop10	7.72	5.31

Table 4.10: Average Overfitting Ratios (OR) for Left sensor side and Right sensor side, Feature Sets. Sensor Type: OPM

Feature Set	Avg OR LHS	OR RHS
FS0	1.32	1.61
FS1	1.33	1.75
FStop50	1.22	1.67
FStop10	1.23	1.55

feature sets were relatively small. On the right-hand side, FS0 achieved the highest improvement (IR = 20.91), while FStop50 had the lowest (IR = 19.91). These results suggest that the VMR signal can be compensated effectively even with reduced feature sets, although performance does not seem to be enhanced by decreasing the number of features.

The full results for the VMR sensor feature selected models can be found in Appendix D. Removing Group 3 features had similar effect on IR and OR as in the OPM case, as can be seen in Table 4.11. Performance was not changed much and OR decreased very slightly. Similarly to OPM, more features are removed to further reduce overfitting.

Table 4.11: Average Improvement Ratios (IR) over all manoeuvres for Left and Right sensor sides, Feature Sets, Sensor Type: VMR

Feature Set	Avg IR LHS	Avg IR RHS
FS0	22.22	20.91
FS1	22.58	20.05
FStop50	22.19	19.91
FStop10	22.00	20.60

Overfitting was more pronounced for the VMR sensor than for the OPM sensor, with training-to-test standard deviation ratios exceeding 2.0 in most cases. As shown in Table 4.12, the left-hand side overfitting ratio was highest for FS0 (OR = 2.48), followed closely by FS1 (2.27). Pruning the feature set reduced overfitting noticeably: FStop50 achieved an OR of 2.00, and FStop10 reduced it further to 1.66. On the right-hand side, which was only evaluated for one manoeuvre pair, the trend was similar. Feature set FS0 exhibited the highest OR (2.31), while FStop10 achieved the lowest (1.50). These results suggest that aggressive feature reduction is effective at mitigating overfitting in models trained on the VMR signal, without significantly reducing performance in terms of IR.

The top 50% and top 10% features have very similar performance to the previous models with all features and the model without derivative features. However, the overfitting ratio has decreased significantly. This is evidence for the 90% of features that were not removed mainly contributing to overfitting and not test set performance. The features left are `AHR2_Q2_3`, `AHR2_Q4_3`, `AHR2_Q4_1`, `AHR2_Q2_1`, `IMU_AccX_2`, `AHR2_Q1_4`, `AHR2_Q2_4`, `ATT_Pitch_1`, `ATT_Yaw_2`, `AHR2_Q4_2`. Except for `IMU_AccX_2`, these features all have to do with orientation of the platform. Quaternions are estimated using an extended Kalman filter (EKF). Roll, pitch and yaw are calculated from the estimated quaternions. According to the documentation of the flight system which was used, called Ardupilot, a magnetometer is used to correct for heading drift, influencing the yaw angle. This raises the possibility of magnetic field disturbances influencing quaternion estimates, a form of data leakage. As a result, features derived

Table 4.12: Average Overfitting Ratios (OR) for Left sensor side and Right sensor side, Feature Sets. Sensor Type: VMR

Feature Set	Avg OR LHS	OR RHS
FS0	2.48	2.31
FS1	2.27	1.92
FStop50	2.00	1.87
FStop10	1.66	1.50

from these quaternions, especially those sensitive to yaw, can implicitly contain information about the magnetic field including both platform noise and Earth field. This could explain why a small subset of orientation features still achieves high performance while reducing overfitting.

Looking more closely into the influence of the onboard magnetometers onto quaternion estimation, from documentation [50] [51] it seems that the specific quaternion estimation algorithm used in our features includes magnetic field magnitude information. The way in which this is incorporated is as follows. An EKF estimates the quaternions. Next, the direction cosine matrix (DCM), which is a rotation matrix, is calculated from the quaternions. Then the platform-frame magnetic field is calculated by multiplying the IGRF model with the DCM. This vector is compared with onboard vector magnetometer values, and the EKF estimation of the quaternions is adjusted based on how much the two vectors differ. To further investigate the possible effect we can look at the magnetometers used by the drone.

The drone magnetometers can be identified as BMM350 magnetometers. Given the BMM350's noise characteristics, low resolution[52] of 100nT, and the fusion behavior of the ArduPilot EKF, it seems infeasible for platform noise in the 100–300 nT range to reliably influence fused quaternion outputs. Any influence would be highly attenuated, non-linear, and inconsistent, making data leakage through quaternions or derived orientation features negligible for the purposes of magnetic compensation modeling. However, it is difficult to rule out data leakage for sure without validating with test data gathered in an area with known magnetic anomalies.

4.6.3. Pre-training MLP

One other way of reducing overfitting is to use more data to train the model. By including data from all training manoeuvres, even those from different flights, the amount of training data is increased. Furthermore, by covering multiple flights, the model could be more general than the other models which are trained for a specific flight. However, since models do appear to benefit from training for manoeuvres done on the same flight, performance is expected to be worse for the model trained on data from multiple flights. That is why this more general model is used as a kind of 'pre-trained' model, which can then be further trained on flight-specific data.

After training and testing a pre-trained model, it was found that it did not predict the residual of ETL well at all. After further examination, it does make sense that it is difficult for the MLP to predict the residual. Each manoeuvre has its own ETL model for noise compensation. The platform noise that is left after ETL compensation is the residual the pretrained model tries to predict. However, since each ETL model has different coefficients, the way it uses the features is different and the pretrained model cannot generalise to all flights.

5

Discussion and Conclusion

This research addressed two questions about reference-free magnetic compensation. The following sections try to answer those questions based on experimental results.

The first research question asked: *Can deep learning models effectively predict and compensate aeromagnetic platform noise in the absence of a noise-free reference signal or accurate anomaly map?* The answer is yes, but with limitations. Neural network approaches significantly outperformed traditional methods. The ETLNN method achieved improvement ratios of 8.87 (OPM) and 22.22 (VMR), compared to standard Tolles-Lawson ratios of 4.90 and 10.19. This represents approximately a twofold increase in performance. Moreover, the best OPM results achieved 2.3 nT standard deviation. Since geological anomalies vary by tens to hundreds of nanoteslas, this noise level suggests practical viability for magnetic-aided navigation. Therefore, this thesis showed reference-free validation by exploiting altitude filtering effects. At 250 m survey height, anomalies shorter than 400 m wavelength are attenuated below 1 nT, allowing assumption of constant background field. This removes the need for reference magnetometers during validation. However, the critical limitation is that compensation performance was only evaluated on anomaly-free data. The ability to preserve actual anomaly signals remains unverified.

The second question asked: *Which drone inputs are most relevant for predicting platform noise, and how can their contribution be quantified across different manoeuvres and flights?* SHAP analysis suggests the following drone inputs are important:

- Engine-related parameters (e.g., battery current, throttle)
- Inertial measurement unit (IMU) data (e.g., accelerations, vibrations)
- Platform attitude information (e.g., roll, pitch, yaw angles, quaternions)
- To a lesser extent, aileron and elevator control signals

The feature analysis revealed that features projected onto the magnetic field direction (Group 2) consistently outperformed raw inputs (Group 1), validating the Extended Tolles-Lawson approach. Furthermore, derivative features added little value while increasing overfitting. Next, using only the top 10% of SHAP-identified features reduced overfitting (from >2.0 to 1.66 for VMR) without hurting performance.

However, there are some important limitations. The method of determining SHAP value for the (E)TLNN was only an approximation, as determining the exact SHAP value would be too computationally expensive. Furthermore, SHAP values are not a perfect measure of the actual effect a drone input has on the platform noise. It only tries to quantify the effect on a specific model. This is another reason why it is important to prevent overfitting, as a model which has overfit can falsely assign importance to unimportant drone inputs. Another limitation is that the analysis was limited to one platform (Mugin drone), and Tolles-Lawson coefficients varied across flights, requiring per-flight calibration. This limits generalization to across flights, meaning the MLP needs retraining for every new flight.

A methodological contribution was made by exploiting the low-pass filtering effect of altitude (250 m). By showing attenuation of certain wavelengths at that altitude on crustal magnetic anomalies, it was established that anomalies with wavelengths shorter than 400 m are attenuated significantly. This enabled the assumption of a near-constant background field for validation and removed the need for a reference magnetometer. Furthermore, a kind of hierarchical model was used effectively. Linear ETL compensation addresses primary platform effects, while a neural network models nonlinear residuals. This architecture outperformed direct neural network models, suggesting that physics-based preprocessing improves learning efficiency. Finally, the feature study highlighted the importance of projecting drone inputs onto the magnetic field direction, grounded in physical principles. Features transformed using direction cosines (Group 2) consistently showed higher SHAP values than raw inputs (Group 1).

5.1. Limitations and Challenges

Several limitations and challenges emerged during the development and evaluation of the proposed compensation framework. One key issue was the quality of the magnetometer data. Quantisation artefacts in the OPM measurements affected over 30% of samples. This highlights the importance of thorough sensor characterisation and preprocessing.

Another challenge was model overfitting, particularly in neural network models trained on VMR sensor data, where overfit ratios exceeded 2.0. Reducing the input space to the top 10% of SHAP-identified features helped mitigate overfitting (e.g., OR = 1.66 for VMR) without significant degradation in overall performance. Reducing the feature set size resulted in less overfitting and decreased overfitting up to a point for OPM as well. Nonetheless, ensuring generalisation across varied datasets remains a concern, especially given the absence of a noise-free reference.

The analysis was also constrained by its reliance on a single UAV platform (the Mugin fixed-wing drone) and a specific sensor configuration. While the proposed methods could generalise to other platforms, this assumption was not empirically validated. Furthermore, because (Extended) Tolles–Lawson coefficients exhibited noticeable variation across flights, calibration needs to be done every flight.

A final and significant limitation lies in the validation approach. Since compensation performance was evaluated on anomaly free data, the ability of the model to preserve magnetic anomalies remains unverified. Although compensation effectively reduced measurement variance, it is unclear whether this came at the cost of suppressing subtle geological signals.

5.2. Future Work

Future work should focus on validating the compensation methods with data that contains known magnetic anomalies. This could be done by generating artificial anomalies with large coils, using reference magnetometers placed far from the drone, or flying over areas where geological anomalies are already mapped. These tests would help to check whether real signals are preserved during compensation.

There is also room to improve the algorithms. Temporal models such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs), or temporal convolutional neural networks (CNNs) could be useful for capturing time-dependent patterns in the data. It may also be worth trying models that include simple physical constraints, or using loss functions such as Huber or mean absolute error (MAE), which are less sensitive to outliers.

A recent study [53] presented a differentiable architecture search-guided physics-informed neural network (DARTS-PINN) for aeromagnetic compensation. This method adds the physics-based Tolles–Lawson model to the loss function of a PINN, while the differentiable architecture search selects the network structure. The approach reduces the amount of training data needed, lowers the standard deviation of residuals, and adapts to different scenarios by finding suitable architectures for different inputs. Although the study used a dataset with reference measurements, the method could also work in a reference-free setting, as it performs well with limited data. Its adaptability makes it a promising option for experiments on other platforms.

Larger datasets could further improve results by reducing overfitting and allowing the models to learn more complex patterns. Data collected from multiple flights would capture differences in conditions. Ground tests with each drone input actuated separately could also help in understanding platform

noise and refining feature selection. In Figure-of-Merit flight patterns there is a risk of data leakage due to the platform performing the same movements in the same position for each square. To prevent data leakage, cloverleaf flight patterns, as described in Section 2.2.4, could be used provide a way to test compensation methods. This flight pattern crosses a single point from four different headings, which makes it well suited for checking whether the Earth's magnetic field is preserved. Since the true field at the crossing point should be the same regardless of direction, any differences in compensated results would indicate limited accuracy.

The assumption of rigid platform movement, outlined in Section 2.1.4, could also be examined further. In practice, wings and drone structures bend during flight, changing the relative positions of magnetometers, ferromagnetic materials, and other magnetic sources. Future work could develop models that account for this flexibility.

The numerical derivative calculations discussed in Section 4.5.1 also leave room for improvement. More advanced differentiation methods could be tested to help reduce the noise amplification that comes with numerical differentiation. Furthermore, the uniform 20 Hz resampling described in Section 3.2.2 may not be the best choice for all inputs, especially high-frequency inputs. Perhaps there is a method better suited for those inputs.

The fact that ETLNN models perform better on VMR data than on OPM data, possibly because of the restricted 0.02–2 Hz prediction band, suggests that frequency-domain methods could be useful. Future work could look at splitting magnetic signals into different frequency bands and training separate models for each. Modelling by frequency band may capture different levels of complexity across ranges and improve performance compared with traditional models.

Finally, the pre-training experiments described in Section 4.6.3 showed that creating models that generalise across flights remains difficult. Future work could explore transfer learning to capture flight-specific differences while keeping some generality. Possible solutions include hierarchical models with both common and flight-specific components, or online adaptation methods that adjust quickly to new conditions while making use of knowledge from earlier flights.

5.3. Conclusion

This research demonstrated that MLP models in combination with linear ETL models can effectively compensate for aeromagnetic platform noise in the absence of reference magnetometers. The ETLNN approach yielded approximately 2 times the improvement of standard Tolles-Lawson methods.

SHAP-based feature analysis provided interpretable insights into platform dynamics, highlighting the influence of engine-related parameters, inertial measurements, and attitude data.

However, the lack of validation on magnetic anomaly data remains the main limitation. While the reference-free validation confirmed noise reduction, it did not confirm the preservation of true geological signals.

Overall, the combination of hierarchical modelling and interpretable feature selection provides a useful approach for future development. With appropriate anomaly validation, these methods can improve magnetic-aided navigation in reference-free situations.

Bibliography

[1] M. Albéri et al., "Accuracy of Flight Altitude Measured with Low-Cost GNSS, Radar and Barometer Sensors: Implications for Airborne Radiometric Surveys," en, *Sensors*, vol. 17, no. 8, p. 1889, Aug. 2017, Number: 8 Publisher: Multidisciplinary Digital Publishing Institute. DOI: 10.3390/s17081889

[2] D. A. Grejner-Brzezinska, R. Da, and C. Toth, "GPS error modeling and OTF ambiguity resolution for high-accuracy GPS/INS integrated system," en, *Journal of Geodesy*, vol. 72, no. 11, pp. 626–638, Nov. 1998. DOI: 10.1007/s001900050202

[3] K. C. Zeng, Y. Shu, S. Liu, Y. Dou, and Y. Yang, "A Practical GPS Location Spoofing Attack in Road Navigation Scenario," en, in *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*, Sonoma CA USA: ACM, Feb. 2017, pp. 85–90. DOI: 10.1145/3032970.3032983

[4] E. Shafiee, M. R. Mosavi, and M. Moazedi, "Detection of Spoofing Attack using Machine Learning based on Multi-Layer Neural Network in Single-Frequency GPS Receivers," en, *The Journal of Navigation*, vol. 71, no. 1, pp. 169–188, Jan. 2018. DOI: 10.1017/S0373463317000558

[5] *Global Navigation Satellite System (GNSS) Outages and Alterations | EASA*, en.

[6] M. Felux, P. Fol, B. Figuet, M. Waltert, and X. Olive, "Impacts of Global Navigation Satellite System Jamming on Aviation," en, *NAVIGATION: Journal of the Institute of Navigation*, vol. 71, no. 3, navi.657, 2024. DOI: 10.33012/navi.657

[7] H. Li, M. Liu, and F. Zhang, "Geomagnetic Navigation of Autonomous Underwater Vehicle Based on Multi-objective Evolutionary Algorithm," English, *Frontiers in Neurorobotics*, vol. 11, Jul. 2017, Publisher: Frontiers. DOI: 10.3389/fnbot.2017.00034

[8] A. Canciani and K. Brink, "Improved magnetic anomaly navigation accuracy through cooperative navigation," in *Proceedings of the ION 2017 Pacific PNT Meeting*, 2017, pp. 239–262.

[9] A. J. Canciani, "AFIT Scholar Theses and Dissertations Student Graduate Works Absolute Positioning Using the Earth' s Magnetic Anomaly Field," Tech. Rep., 2016.

[10] J. Anderson and R. Coates, *The lithosphere: Geochemistry, geology and geophysics*. Jan. 2009, Pages: 383.

[11] P. Alken, "International Geomagnetic Reference Field: The thirteenth generation," *Earth, Planets and Space*, vol. 73, no. 1, p. 49, Feb. 2021. DOI: 10.1186/s40623-020-01288-x

[12] National Centers for Environmental Information (NCEI). "Geomagnetism frequently asked questions," Accessed: Dec. 14, 2024. [Online]. Available: https://www.ncei.noaa.gov/products/geomagnetism-frequently-asked-questions

[13] V. Lesur, M. Hamoudi, Y. Choi, J. Dyment, and E. Thébault, "Building the second version of the World Digital Magnetic Anomaly Map (WDMAM)," *Earth, Planets and Space*, vol. 68, no. 1, p. 27, Feb. 2016. DOI: 10.1186/s40623-016-0404-6

[14] D. J. Griffiths, *Introduction to electrodynamics* (Always learning), en, 4. ed., international ed. Boston: Pearson, 2013.

[15] M. Kok and A. Solin, "Scalable Magnetic Field SLAM in 3D Using Gaussian Process Maps," in *2018 21st International Conference on Information Fusion (FUSION)*, Jul. 2018, pp. 1353–1360. DOI: 10.23919/ICIF.2018.8455789

[16] T. N. Lee and A. J. Canciani, "MagSLAM: Aerial simultaneous localization and mapping using Earth's magnetic anomaly field," en, *NAVIGATION*, vol. 67, no. 1, pp. 95–107, Mar. 2020. DOI: 10.1002/navi.352

[17] A. R. Gnadt, "Advanced Aeromagnetic Compensation Models for Airborne Magnetic Anomaly Navigation," AAI30330933, phd, Massachusetts Institute of Technology, USA, 2022.

[18] M. C. Hezel, “AFIT Scholar AFIT Scholar Improving Aeromagnetic Calibration Using Artificial Neural Improving Aeromagnetic Calibration Using Artificial Neural Networks Networks,” Tech. Rep.

[19] J. Davies. “Cessna 208b grand caravan, sander geophysics.” Licensed under GFDL 1.2, via Wikimedia Commons, CYOW Airport Watch, Accessed: Aug. 13, 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File:Cessna_208B_Grand_Caravan,_Sander_Geophysics_AN0326961.jpg

[20] M. Jukić, L. Visseren, A. Vijn, and R. Tan, “Applications of quantum sensing to aerial magnetic navigation,” in *Quantum Technologies for Defence and Security*, vol. 13202, SPIE, Nov. 2024, pp. 98–116. DOI: 10.1117/12.3031550

[21] A. Canciani and J. Raquet, “Absolute Positioning Using the Earth’s Magnetic Anomaly Field: Magnetic Anomaly Navigation,” en, *Navigation*, vol. 63, no. 2, pp. 111–126, Jun. 2016. DOI: 10.1002/navi.138

[22] A. J. Canciani and C. J. Brennan, “An analysis of the benefits and difficulties of aerial magnetic vector navigation,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 6, pp. 4161–4176, 2020, Publisher: IEEE.

[23] A. R. Gnadt, A. B. Wollaber, and A. P. Nielsen, “Derivation and Extensions of the Tolles-Lawson Model for Aeromagnetic Compensation,” Dec. 2022, arXiv: 2212.09899.

[24] W. Tolles and J. Lawson, *Magnetic compensation of mad equipped aircraft*, 1950.

[25] A. R. Gnadt et al., “Signal Enhancement for Magnetic Navigation Challenge Problem,” 2023, arXiv: 2007.12158v2.

[26] P. Leliak, “IRE TRANSACTIONS ON AEROSPACE AND NAVIGATIONAL ELECTRONICS Identification and Evaluation of Magnetic-Field Sources of Magnetic Airborne Detector Equipped Aircraft*,” Tech. Rep.

[27] B. W. Leach, “Automatic aeromagnetic compensation,” en, National Research Council of Canada, Tech. Rep., 1979, Artwork Size: 132 p. : ill. ISSN: 0384-3157 Publication Title: Laboratory Technical Report (National Research Council of Canada. Flight Research Laboratory) Volume: LTR-FR-69. DOI: 10.4224/40003373

[28] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 2007.

[29] G. Noriega, “Performance measures in aeromagnetic compensation,” en, *The Leading Edge*, vol. 30, no. 10, pp. 1122–1127, Oct. 2011. DOI: 10.1190/1.3657070

[30] C. Reeves, *Aeromagnetic Surveys: Principles, Practice and Interpretation*. Geosoft, 2005.

[31] Auawise and Jrvz, *Pitch roll yaw*, https://w.wiki/_nvVw, Diagram showing yaw, pitch, and roll movements of aircraft in flight. Original by Auawise, modifications by Jrvz. Licensed under CC BY-SA 3.0., 2010.

[32] Y. Liu, W. Li, D. Wei, and G. Shen, “A Modified Aeromagnetic Compensation Method Robust to In-Cabin OBE Interferences,” 2024. DOI: 10.20944/preprints202401.1460.v1

[33] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” en, *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, Jun. 2021, Publisher: Nature Publishing Group. DOI: 10.1038/s42254-021-00314-5

[34] S. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, arXiv:1705.07874 [cs], Nov. 2017. DOI: 10.48550/arXiv.1705.07874

[35] R. J. Blakely, *Potential Theory in Gravity and Magnetic Applications*, en. Cambridge University Press, Sep. 1996.

[36] Mugin UAV. “Mugin product catalog 2020.” Accessed: 2025-06-02. [Online]. Available: https://www.muginuav.com/wp-content/uploads/2020/04/Mugin-Product-Catalog-2020.pdf

[37] T. M. Tierney et al., “Optically pumped magnetometers: From quantum origins to multi-channel magnetoencephalography,” en, *NeuroImage*, vol. 199, pp. 598–608, Oct. 2019. DOI: 10.1016/j.neuroimage.2019.05.063

[38] N. Hadjigeorgiou, K. Asimakopoulos, K. Papafotis, and P. P. Sotiriadis, “Vector Magnetic Field Sensors: Operating Principles, Calibration, and Applications,” *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12 531–12 544, Jun. 2021. DOI: 10.1109/JSEN.2020.3045660

[39] Twinleaf. “Microsam scalar magnetometer,” Accessed: Apr. 30, 2025. [Online]. Available: https://twinleaf.com/magnetometers/microSAM/

[40] Twinleaf. “Vmr vector magnetometer,” Accessed: Apr. 30, 2025. [Online]. Available: https://twinleaf.com/magnetometers/VMR/

[41] A. J. Canciani and J. F. Raquet, “Magnetic anomaly navigation accuracy with respect to map quality and altitude,” in *Proceedings of the 2016 International Technical Meeting of The Institute of Navigation*, 2016, pp. 110–116.

[42] F. Mengyin et al., “Aeromagnetic Compensation Based on Tikhonov Regularization with Limited L-Curve Parameter-Choice Algorithm,” *Chinese Control Conference, CCC*, vol. 2018-July, pp. 1834–1838, Oct. 2018, Publisher: IEEE Computer Society ISBN: 9789881563941. DOI: 10.23919/CHICC.2018.8483472

[43] B. Gu, Q. Li, and H. Liu, “Aeromagnetic compensation based on truncated singular value decomposition with an improved parameter-choice algorithm,” *Proceedings of the 2013 6th International Congress on Image and Signal Processing, CISP 2013*, vol. 3, pp. 1545–1551, 2013, ISBN: 9781479927647. DOI: 10.1109/CISP.2013.6743921

[44] P. C. Hansen, “The L-curve and its use in the numerical treatment of inverse problems,” en,

[45] S. H. Bickel, “Small signal compensation of magnetic fields resulting from aircraft maneuvers,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-15, no. 4, pp. 518–525, 1979. DOI: 10.1109/TAES.1979.308736

[46] N. Laoué, A. Lepers, L. Deletraz, and C. Faure, “Neural Network Calibration of Airborne Magnetometers,” in *2023 IEEE 10th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, IEEE, 2023, pp. 37–42.

[47] A. R. Gnadt, “Advanced Aeromagnetic Compensation Models for Airborne Magnetic Anomaly Navigation,” en, Accepted: 2022-08-29T16:35:48Z, Thesis, Massachusetts Institute of Technology, May 2022.

[48] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980 [cs], Jan. 2017. DOI: 10.48550/arXiv.1412.6980

[49] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[50] x-io Technologies. “Quaternion from magnetometer.” Accessed: 2025-07-11, x-io Technologies. [Online]. Available: https://ahrs.readthedocs.io/en/latest/filters/aqua.html#quaternion-from-magnetometer

[51] ArduPilot Developers. “ArduPilot: ArduPlane 4.0.7.” version 4.0.7. Accessed: 2025-07-11, ArduPilot Project. [Online]. Available: https://github.com/ArduPilot/ardupilot/tree/ArduPlane-4.0.7

[52] Bosch Sensortec. “Bmm350: High-precision 3-axis magnetometer.” Accessed: 2025-07-10, Bosch Sensortec GmbH. [Online]. Available: https://www.bosch-sensortec.com/products/motion-sensors/magnetometers/bmm350/#technical

[53] Z. Jiang, T. Zhao, M. Wang, J. Zhou, Z. Deng, and X. Lin, “An Aeromagnetic Compensation Method Based on Differentiable Architecture Search-Guided Physics-Informed Neural Network,” *IEEE Geoscience and Remote Sensing Letters*, pp. 1–1, 2025. DOI: 10.1109/LGRS.2025.3583559

[54] ArduPilot Developers. “Log messages — ArduPilot documentation,” Accessed: May 7, 2025. [Online]. Available: https://ardupilot.org/plane/docs/logmessages.html

A

Low-Pass Filter Effect of Altitude

Crustal magnetic anomalies, measured by airborne magnetometers, become progressively smoother and more uniform as altitude increases. This behavior can be understood by examining upward continuation in the spatial frequency domain. According to [35], the Fourier transform of the magnetic field at an elevated plane $z + \Delta z$ is given by

$$\mathcal{F}[B_{z+\Delta z}] = \mathcal{F}[B_z]\,\mathcal{F}[\psi],$$

where the upward continuation operator is defined as

$$\mathcal{F}[\psi] = e^{-\Delta z|\mathbf{k}|}, \quad \text{with} \quad |\mathbf{k}| = \sqrt{k_x^2 + k_y^2},$$

where k is the wavenumber of the magnetic anomaly [9].

This exponential decay term shows that components with large wavenumbers, corresponding to small-scale, high-frequency variations in the magnetic field, are rapidly attenuated with height. As a result, the signal at higher altitudes is dominated by low-wavenumber components, leading to a smoother, more slowly varying magnetic field. This low-pass filtering effect of upward continuation is useful for noise compensation when trying to isolate the platform noise during calibration.

B

Drone Inputs

ATT_DesPitch	IMU2_AccY	NKF2_AZbias	NKF1_VE	IOMC_NPkt	IMU_GHz
ATT_DesRoll	IMU2_AccZ	NKF2_GSX	NKF1_VN	IOMC_TS	IMU_GyrX
ATT_ErrRP	IMU2_GHz	NKF2_GSY	NKF1_Yaw	NKF6_GX	IMU_GyrY
ATT_ErrYaw	IMU2_GyrX	NKF2_GSZ	NKF1_dPD	NKF6_GY	IMU_GyrZ
ATT_Pitch	IMU2_GyrY	NKF2_MD	VIBE_VibeX	NKF6_GZ	IMU_T
ATT_Roll	IMU2_GyrZ	NKF2_ME	VIBE_VibeY	NKF6_OH	POWR_Flags
ATT_Yaw	IMU2_T	NKF2_MI	VIBE_VibeZ	NKF6_PD	POWR_Safety
NKF5_HAGL	BAT2_Curr	NKF2_MN	BAT_Curr	NKF6_PE	POWR_VServo
NKF5_eAng	BAT2_CurrTot	NKF2_MX	BAT_CurrTot	NKF6_PN	POWR_Vcc
NKF5_ePos	BAT2_EnrgTot	NKF2_MY	BAT_EnrgTot	NKF6_Pitch	PIDR_Act
NKF5_eVel	BAT2_Res	NKF2_MZ	BAT_Res	NKF6_Roll	PIDR_D
NKF5_offset	BAT2_Volt	NKF2_VWE	BAT_Volt	NKF6_VD	PIDR_P
POS_Alt	BAT2_VoltR	NKF2_VWN	BAT_VoltR	NKF6_VE	PIDR_Tar
POS_Lat	IMU3_AHz	ARSP_Airspeed	CTUN_Aspd	NKF6_VN	NKT2_AngMax
POS_Lng	IMU3_AccX	ARSP_DiffPress	CTUN_NavPitch	NKF6_Yaw	NKT2_AngMin
POS_RelHome	IMU3_AccY	ARSP_RawPress	CTUN_NavRoll	NKF6_dPD	NKT2_Cnt
POS_RelOrigin	IMU3_AccZ	ARSP_Temp	CTUN_Pitch	DSF_Blk	NKT2_EKFMax
PIDP_Act	IMU3_GHz	MODE_Mode	CTUN_RdrOut	DSF_Bytes	NKT2_EKFMin
PIDP_D	IMU3_GyrX	MODE_ModeN	CTUN_Roll	DSF_FAv	NKT2_VMax
PIDP_P	IMU3_GyrY	PM_I2CC	CTUN_ThrDem	DSF_FMn	NKT2_VMin
PIDP_Tar	IMU3_GyrZ	PM_I2CI	CTUN_ThrOut	DSF_FMx	RCOU_C1
NKF9_SH	IMU3_T	PM_Load	STAT_Armed	MAG_MagX	RCOU_C2
NKF9_SM	AHR2_Alt	PM_MaxT	STAT_Still	MAG_MagY	RCOU_C3
NKF9_SP	AHR2_Lat	PM_NLoop	NKQ2_Q1	MAG_MagZ	RCOU_C4
NKF9_SV	AHR2_Lng	PM_SPIC	NKQ2_Q2	MAG_S	RCOU_C5
NKF9_errRP	AHR2_Pitch	RAD_Noise	NKQ2_Q3	BAR2_Alt	RCOU_C6
GPA_Delta	AHR2_Q1	RAD_RSSI	NKQ2_Q4	BAR2_CRt	RCOU_C7
GPA_HAcc	AHR2_Q2	RAD_RemNoise	NKF3_IPD	BAR2_GndTemp	FMTU_FmtType
GPA_SAcc	AHR2_Q3	RAD_RemRSSI	NKF3_IPE	BAR2_Press	PARM_Value
GPA_SMS	AHR2_Q4	RAD_TxBuf	NKF3_IPN	BAR2_SMS	MAV_1_rxp
GPA_VAcc	AHR2_Roll	BARO_Alt	NKF3_IVD	BAR2_Temp	MAV_1_txp
GPA_VDop	AHR2_Yaw	BARO_CRt	NKF3_IVE	RCIN_C1	NKF8_IPD
NKF7_AZbias	MAV_2_txp	BARO_GndTemp	NKF3_IVN	RCIN_C2	NKF8_IPE
NKF7_GSX	GPS_Alt	BARO_Press	NKF3_IYAW	RCIN_C3	NKF8_IPN
NKF7_GSY	GPS_GCrs	BARO_SMS	NKQ1_Q1	RCIN_C4	NKF8_IVD
NKF7_GSZ	GPS_GMS	BARO_Temp	NKQ1_Q2	RCIN_C6	NKF8_IVE
TSYN_RTT	GPS_HDop	NKF1_GX	NKQ1_Q3	RCIN_C7	NKF8_IVN
NKT1_AngMax	GPS_Lat	NKF1_GY	NKQ1_Q4	RCIN_C8	NKF8_IYAW
NKT1_AngMin	GPS_Lng	NKF1_GZ	NKF4_SH	HEAT_I	MAG2_MagX
NKT1_Cnt	GPS_NSats	NKF1_OH	NKF4_SM	HEAT_Out	MAG2_MagY
NKT1_EKFMax	GPS_Spd	NKF1_PD	NKF4_SP	HEAT_P	MAG2_MagZ
NKT1_EKFMin	GPS_VZ	NKF1_PE	NKF4_SV	HEAT_Temp	MAG2_S
NKT1_VMax	AETR_Ail	NKF1_PN	NKF4_errRP	IMU_AHz	
NKT1_VMin	AETR_Elev	NKF1_Pitch	NTUN_AltErr	IMU_AccX	
IMU2_AHz	AETR_Rudd	NKF1_Roll	NTUN_AspdE	IMU_AccY	
IMU2_AccX	AETR_Thr	NKF1_VD	NTUN_Dist	IMU_AccZ	

Figure B.1: List of drone inputs available for model input. A description of each input can be found in [54].

C

Relative Error Derivation

Let the Earth magnetic field vector be

$$\mathbf{B}_e = \begin{bmatrix} B_e \\ 0 \end{bmatrix}, \quad \text{and the platform noise vector be} \quad \mathbf{B}_p = B_e b \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix},$$

where $b = \frac{\|\mathbf{B}_p\|}{\|\mathbf{B}_e\|}$ is the relative magnitude, and θ is the angle between $\mathbf{B}_p$ and the x-axis.

The total field vector is then:

$$\mathbf{B}_t = \mathbf{B}_e + \mathbf{B}_p = B_e \begin{bmatrix} 1 + b\cos\theta \\ b\sin\theta \end{bmatrix}.$$

The magnitude of $\mathbf{B}_t$ is:

$$\|\mathbf{B}_t\| = B_e\sqrt{(1 + b\cos\theta)^2 + (b\sin\theta)^2} = B_e\sqrt{1 + 2b\cos\theta + b^2}.$$

The projection of $\mathbf{B}_p$ onto $\mathbf{B}_t$ is:

$$\mathbf{B}_p^\top \mathbf{B}_t = B_e^2 b\left[\cos\theta(1 + b\cos\theta) + \sin\theta(b\sin\theta)\right] = B_e^2 b(\cos\theta + b).$$

The estimated magnitude of $\|\mathbf{B}_e\|$ is:

$$\widehat{B}_e = \|\mathbf{B}_t\| - \frac{\mathbf{B}_p^\top \mathbf{B}_t}{\|\mathbf{B}_t\|} = B_e\left(\sqrt{1 + 2b\cos\theta + b^2} - \frac{b(\cos\theta + b)}{\sqrt{1 + 2b\cos\theta + b^2}}\right).$$

The relative error is defined as:

$$\epsilon_r(\theta, b) = \frac{\epsilon}{B_e} = \frac{\left|B_e - \widehat{B_e}\right|}{B_e} = \left|1 - \frac{\widehat{B}_e}{B_e}\right|.$$

Substituting in the expression for $\widehat{B}_e$, we get:

$$\epsilon_r(\theta, b) = \left|1 - \left[\sqrt{1 + 2b\cos\theta + b^2} - \frac{b(\cos\theta + b)}{\sqrt{1 + 2b\cos\theta + b^2}}\right]\right|.$$

Finally, simplifying:

$$\boxed{\epsilon_r(\theta, b) = \left|1 + \frac{b\cos\theta + b^2}{\sqrt{1 + 2b\cos\theta + b^2}} - \sqrt{1 + 2b\cos\theta + b^2}\right|}$$

This function is shown in Figure C.1. In this figure, there is a peculiar jump from at relative magnitude equal to 1 and angle equal to 180 i.e. π rad.

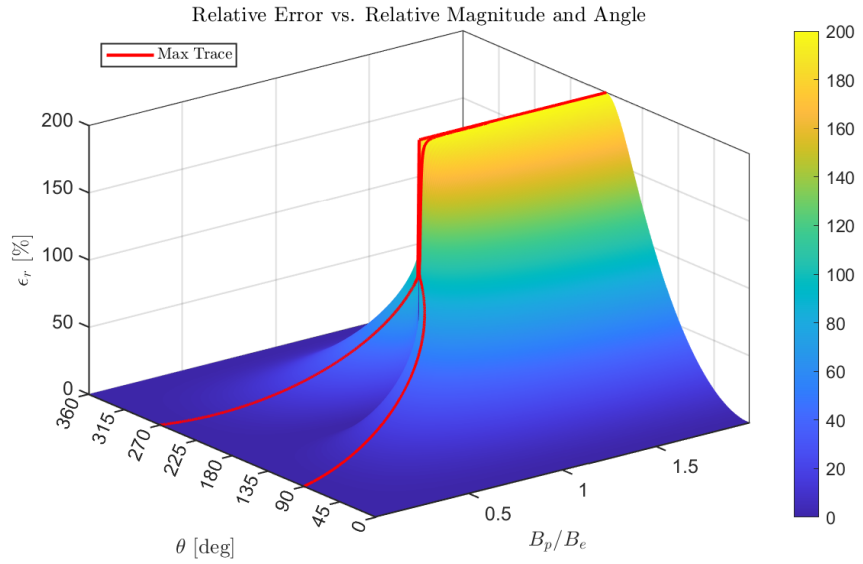


Figure C.1: Relative error surface $\epsilon_r(\theta, b)$ plotted as a function of perturbation direction θ and relative magnitude $b = B_p/B_e$.

Special Case: $\theta = \pi$

At $\theta = \pi$, we have $\cos\theta = -1$. The expression inside the square root becomes:

$$\sqrt{1 + 2b\cos\theta + b^2} = \sqrt{1 - 2b + b^2} = \sqrt{(1-b)^2} = |1 - b|.$$

The numerator of the second term becomes

$$b\cos\theta + b^2 = -b + b^2 = b(b-1).$$

Substitute into the expression for ϵ_r,

$$\epsilon_r(\pi, b) = 1 + \frac{b(b-1)}{|1-b|} - |1-b|.$$

Now, consider the three cases:

Case 1: $b < 1$
Then $|1 - b| = 1 - b$ and $b(b-1) = -b(1-b)$, so:

$$\epsilon_r(\pi, b) = 1 - b - (1-b) = 0.$$

Case 2: $b = 1$
Then $|1 - b| = 0$ and $b(b-1) = 0$, so:

$$\epsilon_r(\pi, 1) = 1 + \frac{0}{0} - 0.$$

This expression is indeterminate, but taking the limit as $b \to 1$ yields

$$\lim_{b\to 1} \epsilon_r(\pi, b) = 1.$$

Case 3: $b > 1$
Then $|1 - b| = b - 1$ and $b(b-1)$ remains positive,

$$\epsilon_r(\pi, b) = 1 + \frac{b(b-1)}{b-1} - (b-1) = (1 + b - (b-1)) = 2.$$

Thus, the relative error at $\theta = \pi$ evaluates to

$$\epsilon_r(\pi, b) = \begin{cases} 0 & \text{if } b < 1 \\ 1 & \text{if } b = 1 \\ 2 & \text{if } b > 1 \end{cases}$$

Small-b Approximation and Upper Bound of the Relative Error

When the relative magnitude $b = \|\mathbf{B}_p\|/\|\mathbf{B}_e\|$ is small (i.e., $b \ll 1$), the relative error can be approximated by expanding the expression

$$\epsilon_r(\theta, b) = \left(1 + \frac{b\cos\theta + b^2}{\sqrt{1 + 2b\cos\theta + b^2}} - \sqrt{1 + 2b\cos\theta + b^2}\right)$$

using a second-order Taylor expansion of the square root,

$$\sqrt{1 + 2b\cos\theta + b^2} \approx 1 + b\cos\theta + \frac{1}{2}b^2\sin^2\theta.$$

Substituting this into the original expression yields the approximation:

$$\epsilon_r(\theta, b) \approx \frac{1}{2}b^2\sin^2\theta.$$

This indicates that, for small b, the relative error increases quadratically with b, and reaches its maximum when $\mathbf{B}_p$ is perpendicular to $\mathbf{B}_e$ (i.e., $\theta = \pi/2$ or $3\pi/2$). The error vanishes when the vectors are aligned or anti-aligned (i.e., $\theta = 0$ or π).

In the worst-case scenario where $\theta = \pi/2$, we obtain a simple upper bound,

$$\epsilon_r \approx \left|1 - \sqrt{1 + b^2}\right| \approx \frac{b^2}{2},$$

so that

$$\epsilon_r \leq \frac{b^2}{2} \quad \text{for } b \ll 1$$

and the absolute error is

$$\epsilon \leq \frac{\|\mathbf{B}_p\|^2}{2\|\mathbf{B}_e\|} \quad \text{for } b \ll 1.$$

D

Extended Results

Here are results from all ETLNN models that were trained using different feature sets. This includes standard deviations (std), improvement ratios (IR) and overfit ratios (OR). Tables D.1 to D.8 are based on OPM data. Tables D.9 to D.16 are based on VMR data.

Table D.1: Standard deviations and Improvement Ratios for ETL + NN compensation.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	37.96	5.54	6.85	31.52	4.63	6.80
6	5	43.15	4.54	9.51	-	-	-
8	7	35.32	3.53	10.00	-	-	-
1C	1	33.95	4.63	7.34	23.38	2.32	10.07
6C	5	45.94	3.61	12.72	-	-	-
8C	7	30.76	4.52	6.81	-	-	-
	Mean	37.85	4.39	8.87	27.45	3.48	8.44

Table D.2: Ratio of training to test set standard deviations for ETLNN compensation using FS0 (all features).

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	1.34	1.61
Man. 5	Man. 6	1.31	1.40
Man. 7	Man. 8	1.32	1.34

Table D.3: Standard deviations and Improvement Ratios for ETLNN, FS1 compensation.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	37.96	5.62	6.75	31.52	4.98	6.33
6	5	43.15	4.10	10.52	-	-	-
8	7	35.32	3.82	9.25	-	-	-
1C	1	33.95	5.20	6.53	23.38	2.84	8.23
6C	5	45.94	3.28	13.99	-	-	-
8C	7	30.76	4.80	6.41	-	-	-
	Mean	37.85	4.47	8.91	27.45	3.91	7.28

Table D.4: Ratio of training to test set standard deviations for ETLNN compensation using FS1 (no derivative features).

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	1.35	1.75
Man. 5	Man. 6	1.29	1.25
Man. 7	Man. 8	1.35	1.27

Table D.5: Standard deviations and Improvement Ratios for ETLNN compensation with FStop50 features.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	37.96	5.40	7.04	31.52	5.19	6.07
6	5	43.15	4.32	9.99	-	-	-
8	7	35.32	3.51	10.06	-	-	-
1C	1	33.95	5.18	6.55	23.38	2.80	8.34
6C	5	45.94	3.43	13.39	-	-	-
8C	7	30.76	4.59	6.70	-	-	-
	Mean	37.85	4.41	8.95	27.45	4.00	7.21

Table D.6: Ratio of training to test set standard deviations for ETLNN compensation using FStop50.

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	1.24	1.67
Man. 5	Man. 6	1.23	1.18
Man. 7	Man. 8	1.20	1.25

Table D.7: Standard deviations and Improvement Ratios for ETLNN compensation with FStop10 features.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	37.96	6.21	6.12	31.52	7.43	4.24
6	5	43.15	5.01	8.61	-	-	-
8	7	35.32	4.57	7.73	-	-	-
1C	1	33.95	4.84	7.02	23.38	3.67	6.37
6C	5	45.94	4.08	11.25	-	-	-
8C	7	30.76	5.52	5.57	-	-	-
	Mean	37.85	5.04	7.72	27.45	5.55	5.31

Table D.8: Ratio of training to test set standard deviations for ETLNN compensation using FStop10.

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	1.24	1.55
Man. 5	Man. 6	1.20	1.20
Man. 7	Man. 8	1.25	1.22

Table D.9: Standard deviations and Improvement Ratios for ETLNN compensation of VMR signal with FS0 features.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	296.87	15.33	19.37	200.46	10.62	18.87
6	5	247.87	10.47	23.67	-	-	-
8	7	231.68	8.50	27.25	-	-	-
1C	1	264.43	8.48	31.18	160.09	6.98	22.95
6C	5	248.85	19.67	12.65	-	-	-
8C	7	222.24	11.58	19.18	-	-	-
	Mean	251.99	12.34	22.22	180.28	8.80	20.91

Table D.10: Ratio of training to test set standard deviations for ETLNN compensation using FS0 (all features).

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	2.94	2.31
Man. 5	Man. 6	2.47	2.36
Man. 7	Man. 8	2.04	1.91

Table D.11: Standard deviations and Improvement Ratios for ETLNN compensation of VMR signal with FS1 features.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	296.87	15.73	18.88	200.46	10.77	18.62
6	5	247.87	10.43	23.77	-	-	-
8	7	231.68	8.37	27.67	-	-	-
1C	1	264.43	8.36	31.63	160.09	7.45	21.49
6C	5	248.85	18.94	13.14	-	-	-
8C	7	222.24	10.91	20.37	-	-	-
	Mean	251.99	12.12	22.58	180.28	9.11	20.05

Table D.12: Ratio of training to test set standard deviations for ETLNN compensation using FS1 (no derivative features).

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	2.56	1.92
Man. 5	Man. 6	2.23	2.29
Man. 7	Man. 8	2.02	1.90

Table D.13: Standard deviations and Improvement Ratios for ETLNN compensation of VMR signal with FStop50 features.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	296.87	15.61	19.01	200.46	10.89	18.40
6	5	247.87	10.93	22.68	-	-	-
8	7	231.68	7.99	28.98	-	-	-
1C	1	264.43	9.09	29.10	160.09	7.48	21.41
6C	5	248.85	19.49	12.77	-	-	-
8C	7	222.24	10.79	20.59	-	-	-
	Mean	251.99	12.32	22.19	180.28	9.19	19.91

Table D.14: Ratio of training to test set standard deviations for ETLNN compensation using FStop50.

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	2.12	1.87
Man. 5	Man. 6	2.15	2.22
Man. 7	Man. 8	1.73	1.66

Table D.15: Standard deviations and Improvement Ratios for ETLNN compensation of VMR signal with FStop10 features.

Man. #	Cal. #	std uncomp LHS [nT]	std comp LHS [nT]	IR LHS	std uncomp RHS [nT]	std comp RHS [nT]	IR RHS
2	1	296.87	16.47	18.03	200.46	10.83	18.51
6	5	247.87	10.86	22.83	-	-	-
8	7	231.68	8.40	27.59	-	-	-
1C	1	264.43	9.81	26.96	160.09	7.06	22.69
6C	5	248.85	18.32	13.58	-	-	-
8C	7	222.24	9.67	22.99	-	-	-
	Mean	251.99	12.25	22.00	180.28	8.94	20.60

Table D.16: Ratio of training to test set standard deviations for ETLNN compensation using FStop10.

Train	Test	OR LHS	OR RHS
Man. 1	Man. 2	1.81	1.50
Man. 5	Man. 6	1.87	1.81
Man. 7	Man. 8	1.30	1.30