







Research Paper

Interpretable Faraday complexity classification

M. J. Alger^{1,2} , J. D. Livingston¹ , N. M. McClure-Griffiths¹ , J. L. Nabaglo¹ , O. I. Wong^{3,4,5}  and C. S. Ong^{2,6} 

¹Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia, ²Data61, CSIRO, Canberra, ACT 2601, Australia, ³CSIRO Astronomy & Space Science, PO Box 1130, Bentley, WA 6102, Australia, ⁴ICRAR-M468, University of Western Australia, Crawley, WA 6009, Australia, ⁵ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia and ⁶Research School of Computer Science, The Australian National University, Canberra, ACT 2601, Australia

Abstract

Faraday complexity describes whether a spectropolarimetric observation has simple or complex magnetic structure. Quickly determining the Faraday complexity of a spectropolarimetric observation is important for processing large, polarised radio surveys. Finding simple sources lets us build rotation measure grids, and finding complex sources lets us follow these sources up with slower analysis techniques or further observations. We introduce five features that can be used to train simple, interpretable machine learning classifiers for estimating Faraday complexity. We train logistic regression and extreme gradient boosted tree classifiers on simulated polarised spectra using our features, analyse their behaviour, and demonstrate our features are effective for both simulated and real data. This is the first application of machine learning methods to real spectropolarimetry data. With 95% accuracy on simulated ASKAP data and 90% accuracy on simulated ATCA data, our method performs comparably to state-of-the-art convolutional neural networks while being simpler and easier to interpret. Logistic regression trained with our features behaves sensibly on real data and its outputs are useful for sorting polarised sources by apparent Faraday complexity.

Keywords: astrostatistics – classification – radio astronomy – radio spectroscopy – spectropolarimetry

(Received 2 December 2020; revised 10 February 2021; accepted 12 February 2021)

1. Introduction

As polarised radiation from distant galaxies makes its way to us, magnetised plasma along the way can cause the polarisation angle to change due to the Faraday effect. The amount of rotation depends on the squared wavelength of the radiation, and the rotation per squared wavelength is called the Faraday depth. Multiple Faraday depths may exist along one line-of-sight, and if a polarised source is observed at multiple wavelengths then these multiple depths can be disentangled. This can provide insight into the polarised structure of the source or the intervening medium.

Faraday rotation measure synthesis (RM synthesis) is a technique for decomposing a spectropolarimetric observation into flux at its Faraday depths ϕ , the resulting distribution of depths being called a ‘Faraday dispersion function’ (FDF) or a ‘Faraday spectrum’. It was introduced by Brentjens & de Bruyn (2005) as a way to rapidly and reliably analyse the polarisation structure of complex and high-Faraday depth polarised observations.

A ‘Faraday simple’ observation is one for which there is only one Faraday depth, and in this simple case, the Faraday depth is also known as a ‘rotation measure’ (RM). All Faraday simple observations can be modelled as a polarised source with a thermal plasma of constant electron density and magnetic field (a ‘Faraday screen’; Brentjens & de Bruyn 2005; Anderson et al. 2015) between the observer and the source. A ‘Faraday complex’ observation is one which is not Faraday simple, and may differ from

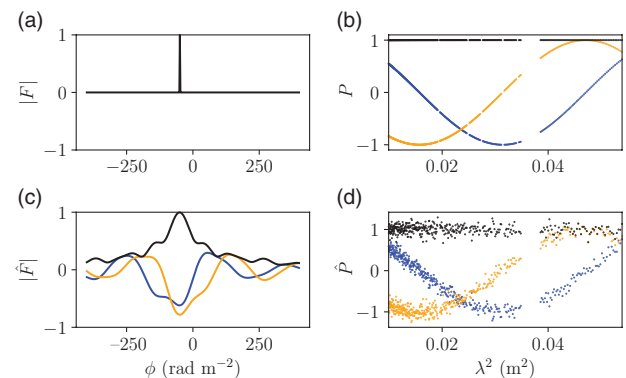


Figure 1. A simple FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components, respectively.

a Faraday simple source due to plasma emission or composition of multiple screens (Brentjens & de Bruyn 2005). The complexity of a source tells us important details about the polarised structure of the source and along the line-of-sight, such as whether the intervening medium emits polarised radiation, or whether there are turbulent magnetic fields or different electron densities in the neighbourhood. The complexity of nearby sources taken together can tell us about the magneto-ionic structure of the galactic and intergalactic medium between the sources and us as observers. O’Sullivan et al. (2017) show examples of simple and complex sources, and Figures 1 and 2 show an example of a simulated simple and complex FDF, respectively.

Author for correspondence: M. J. Alger, E-mail: matthew.alger@gmail.com

Cite this article: Alger MJ, Livingston JD, McClure-Griffiths NM, Nabaglo JL, Wong OI and Ong CS. (2021) Interpretable Faraday complexity classification. *Publications of the Astronomical Society of Australia* 38, e022, 1–11. <https://doi.org/10.1017/pasa.2021.10>

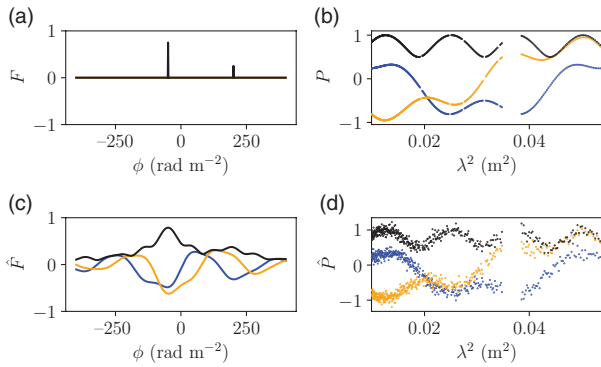


Figure 2. A complex FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components, respectively.

Identifying when an observation is Faraday complex is an important problem in polarised surveys (Sun *et al.* 2015), and with current surveys such as the Polarised Sky Survey of the Universe’s Magnetism (POSSUM) larger than ever before, methods that can quickly characterise Faraday complexity en masse are increasingly useful. Being able to identify which sources are simple lets us produce a reliable rotation measure grid from background sources, and being able to identify which sources might be complex allows us to find sources to follow-up with slower polarisation analysis methods that may require manual oversight, such as QU-fitting (as seen in, e.g. Miyashita *et al.* 2019; O’Sullivan *et al.* 2017). In this paper, we introduce five simple, interpretable features representing polarised spectra, use these features to train machine learning classifiers to identify Faraday complexity, and demonstrate their effectiveness on real and simulated data. We construct our features by comparing observed polarised sources to idealised polarised sources. The features are intuitive and can be estimated from real FDFs.

Section 2 provides a background to our work, including a summary of prior work and our assumptions on FDFs. Section 3 describes our approach to the Faraday complexity problem. Section 4 explains how we trained and evaluated our method. Finally, Section 5 discusses these results.

2. Faraday complexity

Faraday complexity is an observational property of a source: if multiple Faraday depths are observed within the same apparent source (e.g. due to multiple lines-of-sight being combined within a beam), then the source is complex. A source composed of multiple Faraday screens may produce observations consistent with many models (Sun *et al.* 2015), including simple sources, so there is some overlap between simple and complex sources. Faraday thickness is also a source of Faraday complexity: when the intervening medium between a polarised source and the observer also emits polarised light, the FDF cannot be characterised by a simple Faraday screen. As discussed in section 2.2, we defer Faraday thick sources to future work. In this section, we summarise existing methods of Faraday complexity estimation and explain our assumptions and model of simple and complex polarised FDFs.

2.1. Prior work

There are multiple ways to estimate Faraday complexity, including detecting non-linearity in $\chi(\lambda^2)$ (Goldstein & Reed 1984), change

in fractional polarisation as a function of frequency (Farnes, Gaensler, & Carretti 2014), non-sinusoidal variation in fractional polarisation in Stokes Q and U (O’Sullivan *et al.* 2012), counting components in the FDF (Law *et al.* 2011), minimising the Bayesian information criterion (BIC) over a range of simple and complex models (called ‘QU-fitting’; O’Sullivan *et al.* 2017), the method of Faraday moments (Anderson *et al.* 2015; Brown 2011), and deep convolutional neural network classifiers (CNNs; Brown *et al.* 2018). See Sun *et al.* (2015) for a comparison of these methods.

The most common approaches to estimating complexity are QU-fitting (e.g. O’Sullivan *et al.* 2017) and Faraday moments (e.g. Anderson *et al.* 2015). To our knowledge, there is currently no literature examining the accuracy of QU-fitting when applied to complexity classification specifically, though Miyashita *et al.* (2019) analyse its effectiveness on identifying the structure of two-component sources. Brown (2011) suggested Faraday moments as a method to identify complexity, a method later used by Farnes *et al.* (2014) and Anderson *et al.* (2015), but again no literature examines the accuracy. CNNs are the current state-of-the-art with an accuracy of 94.9% (Brown *et al.* 2018) on simulated ASKAP Band 1 and 3 data, and we will compare our results to this method.

2.2. Assumptions on Faraday dispersion functions

Before we can classify FDFs as Faraday complex or Faraday simple, we need to define FDFs and any assumptions we make about them. An FDF is a function that maps Faraday depth ϕ to complex polarisation. It is the distribution of Faraday depths in an observed polarisation spectrum. For a given observation, we assume there is a true, noise-free FDF F composed of at most two Faraday screens. This accounts for most actual sources (Anderson *et al.* 2015) and extension to three screens would cover most of the remainder—O’Sullivan *et al.* (2017) found that 89% of their sources were best explained by two or less screens, while the remainder were best explained by three screens. We model the screens by Dirac delta distributions:

$$F(\phi) = A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1). \quad (1)$$

A_0 and A_1 are the polarised flux of each Faraday screen, and ϕ_0 and ϕ_1 are the Faraday depths of the respective screens. With this model, a Faraday simple source is one which has $A_0 = 0$, $A_1 = 0$, or $\phi_0 = \phi_1$. By using delta distributions to model each screen, we are assuming that there is no internal Faraday dispersion (which is typically associated with diffuse emission rather than the mostly compact sources we expect to find in wide-area polarised surveys). F generates a polarised spectrum of the form shown in Equation (2):

$$P(\lambda^2) = A_0e^{2i\phi_0\lambda^2} + A_1e^{2i\phi_1\lambda^2}. \quad (2)$$

Such a spectrum would be observed as noisy samples from a number of squared wavelengths λ_j^2 , $j \in [1, \dots, D]$. We model this noise as a complex Gaussian with standard deviation σ and call the noisy observed spectrum \hat{P} :

$$\hat{P}(\lambda_j^2) \sim \mathcal{N}(P(\lambda_j^2), \sigma^2). \quad (3)$$

The constant variance of the noise is a simplifying assumption which may not hold for real data, and exploring this is a topic for future work. By performing RM synthesis (Brentjens & de Bruyn 2005) on \hat{P} with uniform weighting we arrive at an observed FDF:

$$\hat{F}(\phi) = \frac{1}{D} \sum_{j=1}^D \hat{P}(\lambda_j^2) e^{-2i\phi\lambda_j^2}. \quad (4)$$

Examples of F , \hat{F} , P , and \hat{P} for simple and complex observations are shown in [Figures 1 and 2](#), respectively. Note that there are two reasons that the observed FDF \hat{F} does not match the groundtruth FDF F . The first is the noise in \hat{P} . The second arises from the incomplete sampling of \hat{P} .

We do not consider external or internal Faraday dispersion in this work. External Faraday dispersion would broaden the delta functions of [Equation \(1\)](#) into peaks, and internal Faraday dispersion would broaden them into top-hat functions. All sources have at least a small amount of dispersion as the Faraday depth is a bulk property of the intervening medium and is subject to noise, but the assumption we make is that this dispersion is sufficiently small that the groundtruth FDFs are well-modelled with delta functions. Faraday thick sources would also invalidate our assumptions, and we assume that there are none in our data as Faraday thickness can be consistent with a two-component model depending on the wavelength sampling (e.g. [Ma et al. 2019](#); [Brentjens & de Bruyn 2005](#)). Nevertheless some external Faraday dispersion would be covered by our model, as depending on observing parameters Faraday thick sources may appear as two screens ([Van Eck et al. 2017](#)).

To simulate observed FDFs we follow the method of [Brown et al. \(2018\)](#), which we describe in [Appendix E](#).

3. Classification approach

The Faraday complexity classification problem is as follows: Given an FDF \hat{F} , is it Faraday complex or Faraday simple? In this section we describe the features that we have developed to address this problem, which can be used in any standard machine learning classifier. We trained two classifiers on these features, which we describe here also.

3.1. Features

Our features are based on a simple idea: all simple FDFs look essentially the same, up to scaling and translation, while complex FDFs may deviate. A noise-free peak-normalised simple FDF \hat{F}_{simple} has the form

$$\hat{F}_{\text{simple}}(\phi; \phi_s) = R(\phi - \phi_s), \tag{5}$$

where R is the rotation measure spread function (RMSF), the Fourier transform of the wavelength sampling function which is 1 at all observed wavelengths and 0 otherwise. ϕ_s traces out a curve in the space of all possible FDFs. In other words, \hat{F}_{simple} is a manifold parametrised by ϕ_s . Our features are derived from relating an observed FDF to the manifold of simple FDFs (the ‘simple manifold’). We measure the distance of an observed FDF to the simple manifold using distance measure D_f , that take all values of the FDF into account:

$$\zeta_f(\hat{F}) = \min_{\phi_s \in \mathbb{R}} D_f(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_s)). \tag{6}$$

We propose two distances that have nice properties:

- invariant over changes in complex phase,
- translationally invariant in Faraday depth,
- zero for Faraday simple sources (i.e. when $A_0 = 0$, $A_1 = 0$, or $\phi_0 = \phi_1$) when there is no noise,
- symmetric in components (i.e. swapping $A_0 \leftrightarrow A_1$ and $\phi_0 \leftrightarrow \phi_1$ should not change the distance),

- increasing as A_0 and A_1 become closer to each other, and
- increasing as screen separation $|\phi_0 - \phi_1|$ increases over a large range.

Our features are constructed from this distance and its minimiser. In other words we look for the simple FDF \hat{F}_{simple} that is ‘closest’ to the observed FDF \hat{F} . The minimiser ϕ_s is the Faraday depth of the simple FDF.

While we could choose any distance that operates on functions, we used the 2-Wasserstein (W_2) distance and the Euclidean distance. The W_2 distance operates on probability distributions and can be thought of as the minimum cost to ‘move’ one probability distribution to the other, where the cost of moving one unit of probability mass is the squared distance it is moved. Under W_2 distance, the minimiser ϕ_s in [Equation \(6\)](#) can be interpreted as the Faraday depth that the FDF \hat{F} would be observed to have if its complexity was unresolved (i.e. the weighted mean of its components). The Euclidean distance is the square root of the least squares loss which is often used for fitting \hat{F}_{simple} to the FDF \hat{F} . Under Euclidean distance, the minimiser ϕ_s is equivalent to the depth of the best-fitting single component under assumption of Gaussian noise in \hat{F} . We calculated the W_2 distance using Python Optimal Transport ([Flamary & Courty 2017](#)), and we calculated the Euclidean distance using `scipy.spatial.distance.euclidean` ([Virtanen et al. 2020](#)). Further intuition about the two distances is provided in [section 3.2](#).

We denote by ϕ_w and ϕ_e , the Faraday depth of the simple FDF that minimises the respective distances (2-Wasserstein and Euclidean).

$$\begin{aligned} \phi_w &= \operatorname{argmin}_{\phi_w} D_{W_2}(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_w)), \\ \phi_e &= \operatorname{argmin}_{\phi_e} D_E(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_e)). \end{aligned}$$

These features are depicted on an example FDF in [Figure 3](#). For simple observed FDFs, the fitted Faraday depths ϕ_w and ϕ_e both tend to be close to the peak of the observed FDF. However, for complex observed FDFs, ϕ_w tends to be at the average depth between the two major peaks of the observed FDF, being closer to the higher peak. For notation convenience, we denote the Faraday depth of the observed FDF that has largest magnitude as ϕ_a , i.e.

$$\phi_a = \operatorname{argmax}_{\phi_a} |\hat{F}(\phi_a)|.$$

Note that in practice $\phi_a \approx \phi_e$. For complex observed FDFs, the values of Faraday depths ϕ_w and ϕ_a tend to differ (essentially by a proportion of the location of the second screen). The difference between ϕ_w and ϕ_a therefore provides useful information to identify complex FDFs. When the observed FDF is simple, the 2-Wasserstein fit will overlap significantly, hence the observed magnitudes $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$ will be similar. However, for complex FDFs ϕ_w and ϕ_a are at different depths, leading to different values of $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$. Therefore, the magnitudes of the observed FDFs at the depths ϕ_w and ϕ_a indicate how different the observed FDF is from a simple FDF.

In summary, we provide the following features to the classifier:

- $\log |\phi_w - \phi_a|$,
- $\log \hat{F}(\phi_w)$,

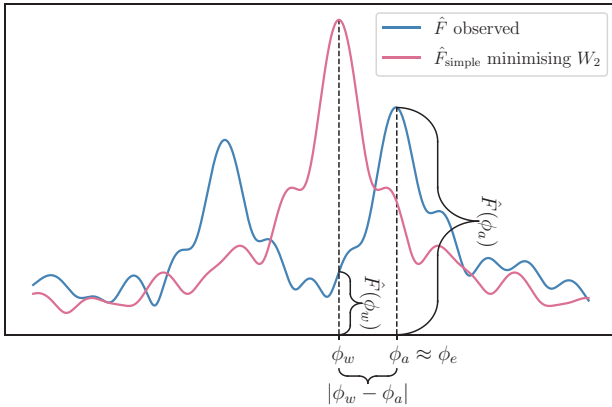


Figure 3. An example of how an observed FDF \hat{F} relates to our features. ϕ_w is the W_2 -minimising Faraday depth, and ϕ_a is the \hat{F} -maximising Faraday depth (approximately equal to the Euclidean-minimising Faraday depth). The remaining two features are the W_2 and Euclidean distances between the depicted FDFs.

- $\log \hat{F}(\phi_a)$,
- $\log D_{W_2}(\hat{F}(\phi) \| \hat{F}_{\text{simple}}(\phi; \phi_w))$,
- $\log D_E(\hat{F}(\phi) \| \hat{F}_{\text{simple}}(\phi; \phi_e))$,

where D_E is the Euclidean distance, D_{W_2} is the W_2 distance, ϕ_a is the Faraday depth of the FDF peak, ϕ_w is the minimiser for W_2 distance, and ϕ_e is the minimiser for Euclidean distance.

3.2. Interpreting distances

Interestingly, in the case where there is no RMSF, Equation (6) with W_2 distance reduces to the Faraday moment already in common use:

$$\zeta_{W_2}(F) = \min_{\phi_w \in \mathbb{R}} D_{W_2}(F(\phi) \| F_{\text{simple}}(\phi; \phi_w)) \quad (7)$$

$$= \left(\frac{A_0 A_1}{(A_0 + A_1)^2} (\phi_0 - \phi_1)^2 \right)^{1/2}. \quad (8)$$

See Appendix A for the corresponding calculation. In this sense, the W_2 distance can be thought of as a generalised Faraday moment, and conversely an interpretation of Faraday moments as a distance from the simple manifold in the case where there is no RMSF. Euclidean distance behaves quite differently in this case, and the resulting distance measure is totally independent of Faraday depth:

$$\zeta_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \| F_{\text{simple}}(\phi; \phi_e)) \quad (9)$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (10)$$

See Appendix B for the corresponding calculation.

3.3. Classifiers

We trained two classifiers on simulated observations using these features: logistic regression (LR) and extreme gradient boosted trees (XGB). These classifiers are useful together for understanding Faraday complexity classification. LR is a linear classifier that is readily interpretable by examining the weights it applies to each feature, and is one of the simplest possible classifiers. XGB is a powerful off-the-shelf non-linear ensemble classifier, and is an example of a decision tree ensemble which are widely used in

astronomy (e.g. Machado Poletti Valle et al. 2020; Hložek et al. 2020). We used the scikit-learn implementation of LR and we use the XGBoost library for XGB. We optimised hyperparameters for XGB using a fork of xgboost-tuner^a as utilised by Zhu, Ong, & Huttley (2020). We used 1 000 iterations of randomised parameter tuning and the hyperparameters we found are tabulated in Table C.1. We optimised hyperparameters for LR using a fivefold cross-validation grid search implemented in sklearn.model_selection.GridSearchCV. The resulting hyperparameters are tabulated in Table D.1 in Appendix C.

4. Experimental method and results

We applied our classifiers to classify simulated (Sections 4.2 and 4.3) and real (Section 4.4) FDFs. We replicated the experimental setup of Brown et al. (2018) for comparison with the state-of-the-art CNN classification method, and we also applied our method to 142 real FDFs observed with the Australia Telescope Compact Array (ATCA) from Livingston et al. (2021) and O’Sullivan et al. (2017).

4.1. Data

4.1.1. Simulated training and validation data

Our classifiers were trained and validated on simulated FDFs. We produced two sets of simulated FDFs, one for comparison with the state-of-the-art method in the literature and one for application to our observed FDFs (described in Section 4.1.2). We refer to the former as the ‘ASKAP’ dataset as it uses frequencies from the Australian Square Kilometre Array Pathfinder 12-antenna early science configuration. These frequencies included 900 channels from 700 to 1 300 and 1 500 to 1 800 MHz and were used to generate simulated training and validation data by Brown et al. (2018). We refer to the latter as the ‘ATCA’ dataset as it uses frequencies from the 1 to 3 GHz configuration of the ATCA. These frequencies included 394 channels from 1.29 to 3.02 GHz and match our real data. We simulated Faraday depths from -50 to 50 rad m^{-2} for the ‘ASKAP’ dataset (matching Brown) and -500 to 500 for the ‘ATCA’ dataset.

For each dataset, we simulated 100 000 FDFs, approximately half simple and half complex. We randomly allocated half of these FDFs to a training set and reserved the remaining half for validation. Each FDF had complex Gaussian noise added to the corresponding polarisation spectrum. For the ‘ASKAP’ dataset, we sampled the standard deviation of the noise uniformly between 0 and $\sigma_{\text{max}} = 0.333$, matching the dataset of Brown et al. (2018). For the ‘ATCA’ dataset, we fit a log-normal distribution to the standard deviations of O’Sullivan’s data (O’Sullivan et al. 2017) from which we sampled our values of σ :

$$\sigma \sim \frac{1}{0.63\sqrt{2\pi}\sigma} \exp\left(-\frac{\log(50\sigma - 0.5)^2}{2 \times 0.63^2}\right). \quad (11)$$

4.1.2. Observational data

We used two real datasets containing a total of 142 sources: 42 polarised spectra from Livingston et al. (2021) and 100 polarised spectra from O’Sullivan et al. (2017). These datasets were observed in similar frequency ranges on the same telescope (with different

^a<https://github.com/chengsoonong/xgboost-tuner>.

binning), but are in different parts of the sky. The Livingston data were taken near the Galactic Centre, and the O’Sullivan data were taken away from the plane of the Galaxy. There are more Faraday complex sources near the Galactic Centre compared to more Faraday simple sources away from the plane of the Galaxy (Livingston et al. 2021). The similar frequency channels used in the two datasets result in almost identical RMSFs over the Faraday depth range we considered (-500 to 500 rad m^{-2}), so we expected that the classifiers would work equally well on both datasets with no need to retrain. We discarded the 26 Livingston sources with modelled Faraday depths outside of this Faraday depth range, which we do not expect to affect the applicability of our methods to wide-area surveys because these fairly high depths are not common.

Livingston et al. (2021) used RM-CLEAN (Heald 2008) to identify significant components in their FDFs. Some of these components had very high Faraday depths up to 2000 rad m^{-2} , but we chose to ignore these components in this paper as they are much larger than might be expected in a wide-area survey like POSSUM. They used the second Faraday moment (Brown 2011) to estimate Faraday complexity, with Faraday depths determined using `scipy.signal.findpeaks` on the cleaned FDFs, with a cut-off of seven times the noise of the polarised spectrum. Using this method, they estimated that 89% of their sources were Faraday complex, i.e. had a Faraday moment greater than 0.

O’Sullivan et al. (2017) used the QU-fitting and model selection technique as described in O’Sullivan et al. (2012). The QU-fitting models contained up to three Faraday screen components as well as a term for internal and external Faraday dispersion. We ignore the Faraday thickness and dispersion for the purposes of this paper, as most sources were not found to have Faraday thickness and dispersion is beyond the scope of our current work. Thirty-seven sources had just 1 component, 52 had 2, and the remaining 11 had 3.

4.2. Results on ‘ASKAP’ dataset

The accuracy of the LR and XGB classifiers on the ‘ASKAP’ testing set was 94.4 and 95.1%, respectively. The rates of true and false identifications are summarised in Table 1. These results are very close to the CNN presented by Brown et al. (2018), with a slightly higher true negative rate and a slightly lower true positive rate (recalling that positive sources are complex, and negative sources are simple). The accuracy of the CNN was 94.9, slightly lower than our XGB classifier and slightly higher than our LR classifier. Both of our classifiers therefore produce similar classification performance to the CNN, with faster training time and easier interpretation.

4.3. Results on ‘ATCA’ dataset

The accuracy of the LR and XGB classifiers on the ‘ATCA’ dataset was 89.2 and 90.5%, respectively. The major differences between the ‘ATCA’ and the ‘ASKAP’ experiments are the range of the simulated Faraday depths and the distribution of noise levels. The ‘ASKAP’ dataset, to match past CNN work, only included depths from -50 to 50 rad m^{-2} , while the ‘ATCA’ dataset includes depths from -500 to 500 rad m^{-2} . The rates of true and false identifications are again shown in Table 1.

As we know the true Faraday depths of the components in our simulation, we can investigate the behaviour of these classifiers as a function of physical properties. Figure 4 shows the mean classifier prediction as a function of component depth separation

Table 1. Confusion matrix entries for LR and XGB on ‘ASKAP’ and ‘ATCA’ simulated datasets, and the CNN confusion matrix entries are adapted from Brown et al. (2018)

	‘ASKAP’			‘ATCA’	
	LR	XGB	CNN	LR	XGB
True negative rate	0.99	0.99	0.97	0.92	0.91
False positive rate	0.01	0.01	0.03	0.08	0.09
False negative rate	0.10	0.09	0.07	0.16	0.10
True positive rate	0.90	0.91	0.93	0.84	0.90

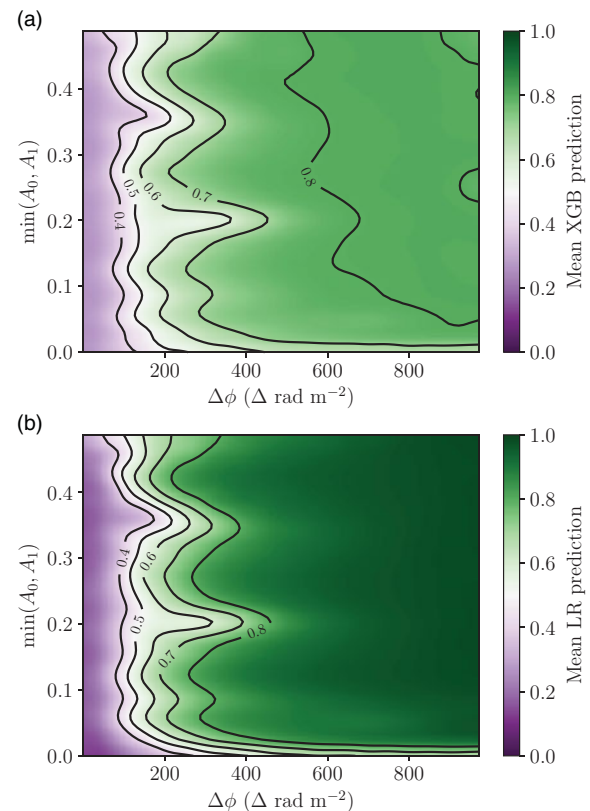


Figure 4. Mean prediction as a function of component depth separation and minimum component amplitude for (a) XGB and (b) LR.

and minimum component amplitude. This is tightly related to the mean accuracy, as the entire plot domain contains complex spectra besides the left and bottom edge: by thresholding the classifier prediction to a certain value, the accuracy will be 100% on the non-edge for all sources with higher prediction values.

4.4. Results on observed FDFs

We used the LR and XGB classifiers, which were trained on the ‘ATCA’ dataset to estimate the probability that our 142 observed FDFs (Section 4.1.2) were Faraday complex. As these classifiers were trained on simulated data, they face the issue of the ‘domain gap’: the distribution of samples from a simulation differs from the distribution of real sources, and this affects performance on real data. Solving this issue is called ‘domain adaptation’ and how to do this is an open research question in machine learning (Zhang 2019; Pan & Yang 2010). Nevertheless, the features of our observations

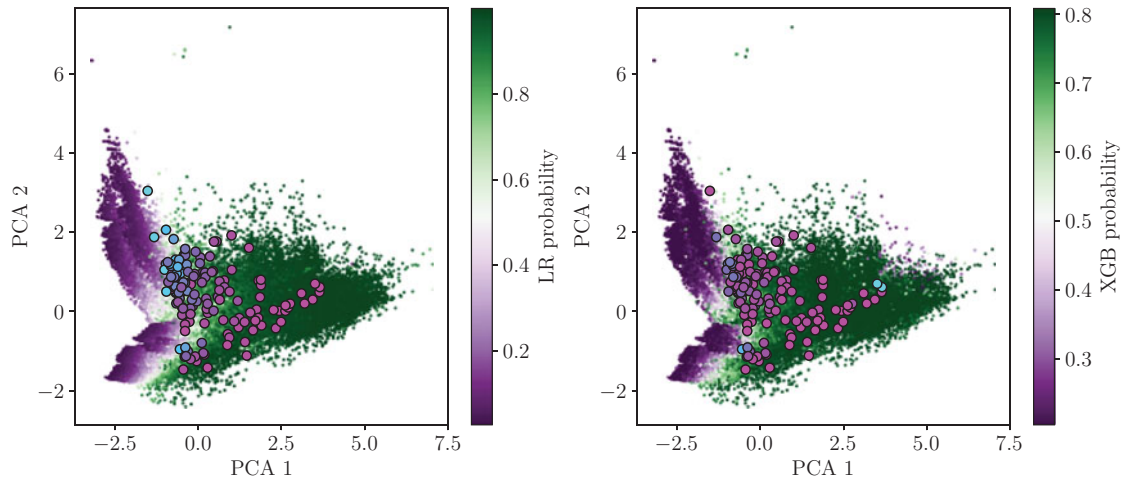


Figure 5. Principal component analysis for simulated data (coloured dots) with observations overlaid (black-edged circles). Observations are coloured by their XGB- or LR-estimated probability of being complex, with blue indicating ‘most simple’ and pink indicating ‘most complex’.

mostly fall in the same region of feature space as the simulations (Figure 5) and so we expect reasonably good domain transfer.

Two apparently complex sources in the Livingston sample are classified as simple with high probability by XGB. These outliers are on the very edge of the training sample (Figure 5) and the underdensity of training data here is likely the cause of this issue. LR does not suffer the same issue, producing plausible predictions for the entire dataset, and these sources are instead classified as complex with high probability.

With a threshold of 0.5, LR predicted that 96 and 83% of the Livingston and O’Sullivan sources were complex, respectively. This is in line with expectations that the Livingston data should have more Faraday complex sources than the O’Sullivan data due to their location near the Galactic Centre. XGB predicted that 93 and 100% of the Livingston and O’Sullivan sources were complex, respectively. Livingston *et al.* (2021) found that 90% of their sources were complex, and O’Sullivan *et al.* (2017) found that 64% of their sources were complex. This suggests that our classifiers are overestimating complexity, though it could also be the case that the methods used by Livingston and O’Sullivan underestimate complexity. Modifying the prediction threshold from 0.5 changes the estimated rate of Faraday complexity, and we show the estimated rates against threshold for both classifiers in Figure 6. We suggest that this result is indicative of our probabilities being uncalibrated, and a higher threshold should be chosen in practice. We chose to keep the threshold at 0.5 as this had the highest accuracy on the simulated validation data. The very high complexity rates of XGB and two outlying classifications indicate that the XGB classifier may be overfitting to the simulation and that it is unable to generalise across the domain gap.

Figures D.1 and D.2 show every observed FDF ordered by estimated Faraday complexity, alongside the models predicted by Livingston and O’Sullivan *et al.* (2017), for LR and XGB, respectively. There is a clear visual trend of increasingly complex sources with increasing predicted probability of being complex.

5. Discussion

On simulated data (Section 4.3), we achieve state-of-the-art accuracy. Our results on observed FDFs show that our classifiers

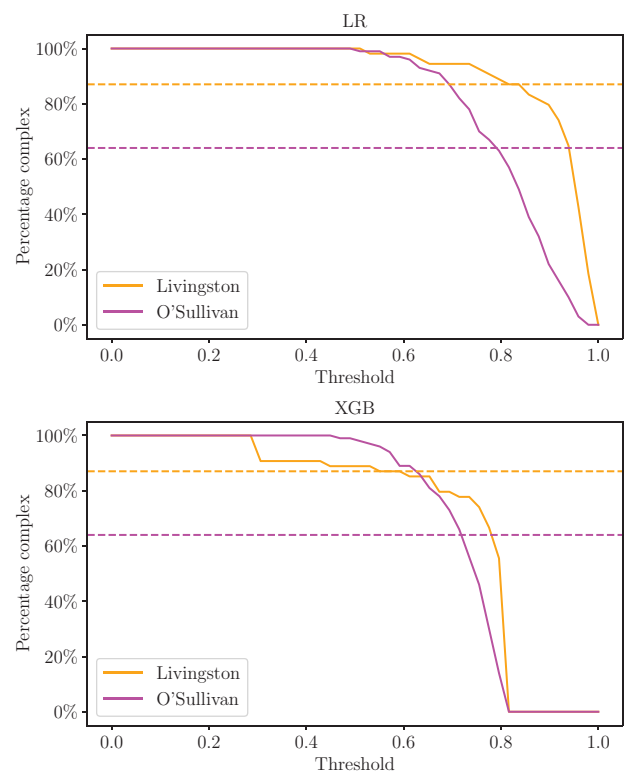


Figure 6. Estimated rates of Faraday complexity for the Livingston and O’Sullivan datasets as functions of threshold. The horizontal lines indicate the rates of Faraday complexity estimated by Livingston and O’Sullivan respectively.

produce plausible results, with Figures D.1 and D.2 showing a clear trend of apparent complexity. Some issues remain: we discuss the intrinsic overlap between simple and complex FDFs in Section 5.1 and the limitations of our method in Section 5.2.

5.1. Complexity and seeming ‘not simple’

Through this work, we found our methods limited by the significant overlap between complex and simple FDFs. Complex FDFs

can be consistent with simple FDFs due to close Faraday components or very small amplitudes on the secondary component, and vice versa due to noise.

The main failure mode of our classifiers is misclassifying a complex source as simple (Table 1). Whether sources with close components or small amplitudes should be considered complex is not clear, since for practical purposes, they can be treated as simple: assuming the source is simple yields a very similar RM to the RM of the primary component, and thus would not negatively impact further data products such as an RM grid. The scenarios where we would want a Faraday complexity classifier rather than a polarisation structure model—large-scale analysis and wide-area surveys—do not seem to be disadvantaged by considering such sources simple. Additional sources similar to these are likely hidden in presumably ‘simple’ FDFs by the frequency range and spacing of the observations, just as how these complex sources would be hidden in lower resolution observations. Note also that misidentification of complex sources as simple is intrinsically a problem with complexity estimation even for models not well-represented by a simple FDF, as complex sources may conspire to appear as a wide range of viable models including simple (Sun et al. 2015).

Conversely, high-noise simple FDFs may be consistent with complex FDFs. One key question is how Faraday complexity estimators should behave as the noise increases: should high noise result in a complex prediction or a simple prediction, given that a complex or simple FDF would both be consistent with a noisy FDF? Occam’s razor suggests that we should choose the simplest suitable model, and so increasing noise should lead to predictions of less complexity. This is not how our classifiers operate, however: high-noise FDFs are different to the model simple FDFs and so are predicted to be ‘not simple’. In some sense our classifiers are not looking for complex sources, but are rather looking for ‘not simple’ sources.

5.2. Limitations

Our main limitations are our simplifying assumptions on FDFs and the domain gap between simulated and real observations. However, our proposed features (Section 3.1) can be applied to future improved simulations.

It is unclear what the effect of our simplifying assumptions are on the effectiveness of our simulation. The three main simplifications that may negatively affect our simulations are (1) limiting to two components, (2) assuming no external Faraday dispersion, and (3) assuming no internal Faraday dispersion (Faraday thickness). Future work will explore removing these simplifying assumptions, but will need to account for the increased difficulty in characterising the simulation with more components and no longer having Faraday screens as components. Additionally, more work will be required to make sure that the rates of internal and external Faraday dispersion match what might be expected from real sources, or risk making a simulation that has too large a range of consistent models for a given source: for example, a two-component source could also be explained as a sufficiently wide or resolved-out Faraday thick source or a three-component source with a small third component. This greatly complicates the classification task.

Previous machine learning work (e.g. Brown et al. 2018) has not been run before on real FDF data, so this paper is the

first example of the domain gap arising in Faraday complexity classification. This is a problem that requires further research to solve. We have no good way to ensure that our simulation matches reality, so some amount of domain adaptation will always be necessary to train classifiers on simulated data and then apply these classifiers to real data. But with the low source counts in polarisation science (high-resolution spectropolarimetric data currently numbers in the few hundreds) any machine learning method will need to be trained on simulations. This is not just a problem in Faraday complexity estimation, and domain adaptation is also an issue faced in the wider astroinformatics community: large quantities of labelled data are hard to come by, and some sources are very rare (e.g. gravitational wave detections or fast radio bursts; Zevin et al. 2017; Gebhard et al. 2019; Agarwal et al. 2020). LR seems to handle the domain adaptation better than XGB, with only a slightly lower accuracy on simulated data. Our results are plausible and the distribution of our simulation well overlaps the distribution of our real data (Figure 5).

6. Conclusion

We developed a simple, interpretable machine learning method for estimating Faraday complexity. Our interpretable features were derived by comparing observed FDFs to idealised simple FDFs, which we could determine both for simulated and real observations. We demonstrated the effectiveness of our method on both simulated and real data. Using simulated data, we found that our classifiers were 95% accurate, with near perfect recall (specificity) of Faraday simple sources. On simulated data that matched existing observations, our classifiers obtained an accuracy of 90%. Evaluating our classifiers on real data gave the plausible results shown in Figure D.1, and marks the first application of machine learning to observed FDFs. Future work will need to narrow the domain gap to improve transfer of classifiers trained on simulations to real, observed data.

Acknowledgements. This research was conducted in Canberra, on land for which the Ngunnawal and Ngambri people are the traditional and ongoing custodians. M. J. A. and J. D. L. were supported by the Australian Government Research Training Program. M. J. A. was supported by the Astronomical Society of Australia. The Australia Telescope Compact Array is part of the Australia Telescope National Facility which is funded by the Australian Government for operation as a National Facility managed by CSIRO. We acknowledge the Gomeroi people as the traditional owners of the Observatory site. We thank the anonymous referee for their comments on this work.

References

- Agarwal, D., Aggarwal, K., Burke-Spolaor, S., Lorimer, D. R., & Garver-Daniels, N. 2020, *MNRAS*,
- Anderson, C. S., Gaensler, B. M., Feain, I. J., & Franzen, T. M. O. 2015, *ApJ*, 815, 49
- Brentjens, M. A., & de Bruyn, A. G. 2005, *A&A*, 441, 1217
- Brown, S. 2011, *Assess the Complexity of an RM Synthesis Spectrum*. No. 9 in POSSUM REPORT
- Brown, S., et al. 2018, *MNRAS*,
- Farnes, J. S., Gaensler, B. M., & Carretti, E. 2014, *ApJS*, 212, 15
- Flamary, R., & Courty, N. 2017, POT Python Optimal Transport library, <https://github.com/rflamary/POT>
- Gebhard, T. D., Kilbertus, N., Harry, I., & Schölkopf, B. 2019, *Phys. Rev. D*, 100, 063015
- Goldstein, S. J. J., & Reed, J. A. 1984, *ApJ*, 283, 540

- Heald, G. 2008, *PIAU*, 4, 591
 Hložek, R., et al. 2020, arXiv e-prints, 2012, [arXiv:2012.12392](https://arxiv.org/abs/2012.12392)
 Law, C. J., et al. 2011, *ApJ*, 728, 57
 Livingston, J. D., McClure-Griffiths, N. M., Gaensler, B. M., Seta, A., & Alger, M. J. 2021, *MNRAS*, 502, 3814
 Ma, Y. K., Mao, S. A., Stil, J., Basu, A., West, J., Heiles, C., Hill, A. S., & Betti, S. K. 2019, *MNRAS*, 487, 3432
 Machado Poletti Valle, L. F., Avestruz, C., Barnes, D. J., Farahi, A., Lau, E. T., & Nagai, D. 2020, arXiv e-prints, 2011, [arXiv:2011.12987](https://arxiv.org/abs/2011.12987)
 Miyashita, Y., Ideguchi, S., Nakagawa, S., Akahori, T., & Takahashi, K. 2019, *MNRAS*, 482, 2739
 O'Sullivan, S. P., et al. 2012, *MNRAS*, 421, 3300
 O'Sullivan, S. P., Purcell, C. R., Anderson, C. S., Farnes, J. S., Sun, X. H., & Gaensler, B. M. 2017, *MNRAS*, 469, 4034
 Pan, S. J., & Yang, Q. 2010, *IEEE TKDE*, 22, 1345
 Sun, X. H., et al. 2015, *AJ*, 149, 60
 Van Eck, C. L., et al. 2017, *A&A*, 597, A98
 Virtanen, P., et al. 2020, *NM*, 17, 261
 Zevin, M., et al. 2017, *CQG*, 34, 064003
 Zhang, L. 2019, arXiv preprint [arXiv:1903.04687](https://arxiv.org/abs/1903.04687)
 Zhu, Y., Ong, C. S., & Huttley, G. A. 2020, *Genetics*, 215, 25

A. 2-Wasserstein begets Faraday moments

Minimising the 2-Wasserstein distance between a model FDF and the simple manifold gives the second Faraday moment of that FDF. Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the sum-normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1}, \quad (\text{A1})$$

$$\tilde{S}(\phi; \phi_w) = \delta(\phi - \phi_w). \quad (\text{A2})$$

The W_2 distance, usually defined on probability distributions, can be extended to one-dimensional complex functions A and B by normalising them:

$$D_{W_2}(A \parallel B)^2 = \inf_{\gamma \in \Gamma(A, B)} \iint_{\phi_{\min}^{\phi_{\max}}} |x - y|^2 d\gamma(x, y), \quad (\text{A3})$$

$$\tilde{A}(\phi) = \frac{|A(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |A(\theta)| d\theta}, \quad (\text{A4})$$

$$\tilde{B}(\phi) = \frac{|B(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |B(\theta)| d\theta}, \quad (\text{A5})$$

where $\Gamma(A, B)$ is the set of couplings of A and B , i.e. the set of joint probability distributions that marginalise to A and B ; and $\inf_{\gamma \in \Gamma(A, B)}$ is the infimum over $\Gamma(A, B)$. This can be interpreted as the minimum cost to ‘move’ one probability distribution to the other, where the cost of moving one unit of probability mass is the squared distance it is moved.

The set of couplings $\Gamma(\tilde{F}, \tilde{S})$ is the set of all joint probability distributions γ such that

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\phi = \tilde{S}(\varphi; \phi_w), \quad (\text{A6})$$

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\varphi = \tilde{F}(\phi). \quad (\text{A7})$$

The coupling that minimises the integral in Equation (A3) will be the optimal transport plan between \tilde{F} and \tilde{S} . Since \tilde{F} and \tilde{S} are defined in terms of delta functions, the optimal transport problem

reduces to a discrete optimal transport problem and the optimal transport plan is

$$\gamma(\phi, \varphi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1} \delta(\varphi - \phi_w). \quad (\text{A8})$$

In other words, to move the probability mass of \tilde{S} to \tilde{F} , a fraction $A_0/(A_0 + A_1)$ is moved from ϕ_w to ϕ_0 and the complementary fraction $A_1/(A_0 + A_1)$ is moved from ϕ_w to ϕ_1 . Then:

$$D_{W_2}(\tilde{F} \parallel \tilde{S})^2 = \iint_{\phi_{\min}}^{\phi_{\max}} |\phi - \varphi|^2 d\gamma(\phi, \varphi) \quad (\text{A9})$$

$$= \frac{A_0(\phi_0 - \phi_w)^2 + A_1(\phi_1 - \phi_w)^2}{A_0 + A_1}. \quad (\text{A10})$$

To obtain the W_2 distance to the simple manifold, we need to minimise this over ϕ_w . Differentiate with respect to ϕ_w and set equal to zero to find

$$\phi_w = \frac{A_0 \phi_0 + A_1 \phi_1}{A_0 + A_1}. \quad (\text{A11})$$

Substituting this back in, we find

$$\zeta_{W_2}(F)^2 = \frac{A_0 A_1}{A_0 + A_1} (\phi_0 - \phi_1)^2 \quad (\text{A12})$$

which is the Faraday moment.

B. Euclidean distance in the no-RMSF case

In this section, we calculate the minimised Euclidean distance evaluated on a model FDF (Equation (1)). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1}, \quad (\text{B1})$$

$$\tilde{S}(\phi; \phi_e) = \delta(\phi - \phi_e). \quad (\text{B2})$$

The Euclidean distance between \tilde{F} and \tilde{S} is then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (\text{B3})$$

$$= \int_{\phi_{\min}}^{\phi_{\max}} |\tilde{F}(\phi) - \delta(\phi - \phi_e)|^2 d\phi. \quad (\text{B4})$$

Assume $\phi_0 \neq \phi_1$ (otherwise, D_E will always be either 0 or 2). If $\phi_e = \phi_0$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (\text{B5})$$

$$= \frac{1}{(A_0 + A_1)^2} \int_{\phi_{\min}}^{\phi_{\max}} A_1^2 |\delta(\phi - \phi_1) - \delta(\phi - \phi_0)|^2 d\phi \quad (\text{B6})$$

$$= \frac{2A_1^2}{(A_0 + A_1)^2} \quad (\text{B7})$$

and similarly for $\phi_e = \phi_1$. If $\phi_e \neq \phi_0$ and $\phi_e \neq \phi_1$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 = \frac{A_0^2 + A_1^2 + 1}{(A_0 + A_1)^2}. \quad (\text{B8})$$

The minimised Euclidean distance when $\phi_0 \neq \phi_1$ is therefore

$$\zeta_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \quad (\text{B9})$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (\text{B10})$$

If $\phi_0 = \phi_1$, then the minimised Euclidean distance is 0.

Table C.1. XGB hyperparameters for the ‘ATCA’ dataset.

Parameter	Value
colsample_bytree	0.912
gamma	0.532
learning_rate	0.1
max_depth	7
min_child_weight	2
scale_pos_weight	1
subsample	0.557
n_estimators	135
reg_alpha	0.968
reg_lambda	1.420

C. Hyperparameters for LR and XGB

This section contains tables of the hyperparameters that we used for our classifiers. Tables C.1 and D.1 tabulate the hyperparameters for XGB and LR, respectively, for the ‘ATCA’ dataset. Tables D.2 and D.3 tabulate the hyperparameters for XGB and LR, respectively, for the ‘ASKAP’ dataset.

D. Predictions on real data

This section contains Figures D.1 and D.2, which shows the predicted probability of being Faraday complex for all real data used in this paper, drawn from Livingston et al. (2021) and O’Sullivan et al. (2017).

E. Simulating observed FDFs

We simulated FDFs by approximating them by arrays of complex numbers. An FDF F is approximated on the domain $[-\phi_{\max}, \phi_{\max}]$ by a vector $F \in \mathbb{R}^d$:

$$F_j = \sum_{k=0}^1 A_k \delta(-\phi_{\max} + j\delta\phi - \phi_k), \quad (E1)$$

where $\delta\phi = (\phi_{\max} - \phi_{\min})/d$ and d is the number of Faraday depth samples in the FDF. F is sampled by uniformly sampling its parameters:

$$\phi_k \in [\phi_{\min}, \phi_{\min} + \delta\phi, \dots, \phi_{\max}], \quad (E2)$$

$$A_k \sim \mathcal{U}(0, 1). \quad (E3)$$

We then generate a vector polarisation spectrum $P \in \mathbb{R}^m$ from F using Equation (E4):

$$P_\ell = \sum_{j=0}^j F_j e^{2i(\phi_{\min} + j\delta\phi)\lambda_\ell^2} d\phi. \quad (E4)$$

Table C.2. LR hyperparameters for the ‘ATCA’ dataset.

Parameter	Value
penalty	L1
C	1.668

Table C.3. XGB hyperparameters for the ‘ASKAP’ dataset.

Parameter	Value
colsample_bytree	0.865
gamma	0.256
learning_rate	0.1
max_depth	6
min_child_weight	1
scale_pos_weight	1
subsample	0.819
n_estimators	108
reg_alpha	0.049
reg_lambda	0.454

Table C.4. LR hyperparameters for the ‘ASKAP’ dataset.

Parameter	Value
penalty	L2
C	0.464

λ_ℓ^2 is the discretised value of λ^2 at the ℓ th index of P . This requires a set of λ^2 values, which depends on the dataset being simulated. These values can be treated as the channel wavelengths at which the polarisation spectrum was observed. We then add Gaussian noise with variance σ^2 to each element of P to obtain a discretised noisy observation \hat{P} . Finally, we perform RM synthesis using the Canadian Initiative for Radio Astronomy Data Analysis RM package^b, which is a Python module that implements a discrete version of RM synthesis:

$$\hat{F}_j = m^{-1} \sum_{\ell=1}^m \hat{P}_\ell e^{-2i(\phi_{\min} + j\delta\phi)\lambda_\ell^2}. \quad (E5)$$

^b<https://github.com/CIRADA-Tools/RM>.

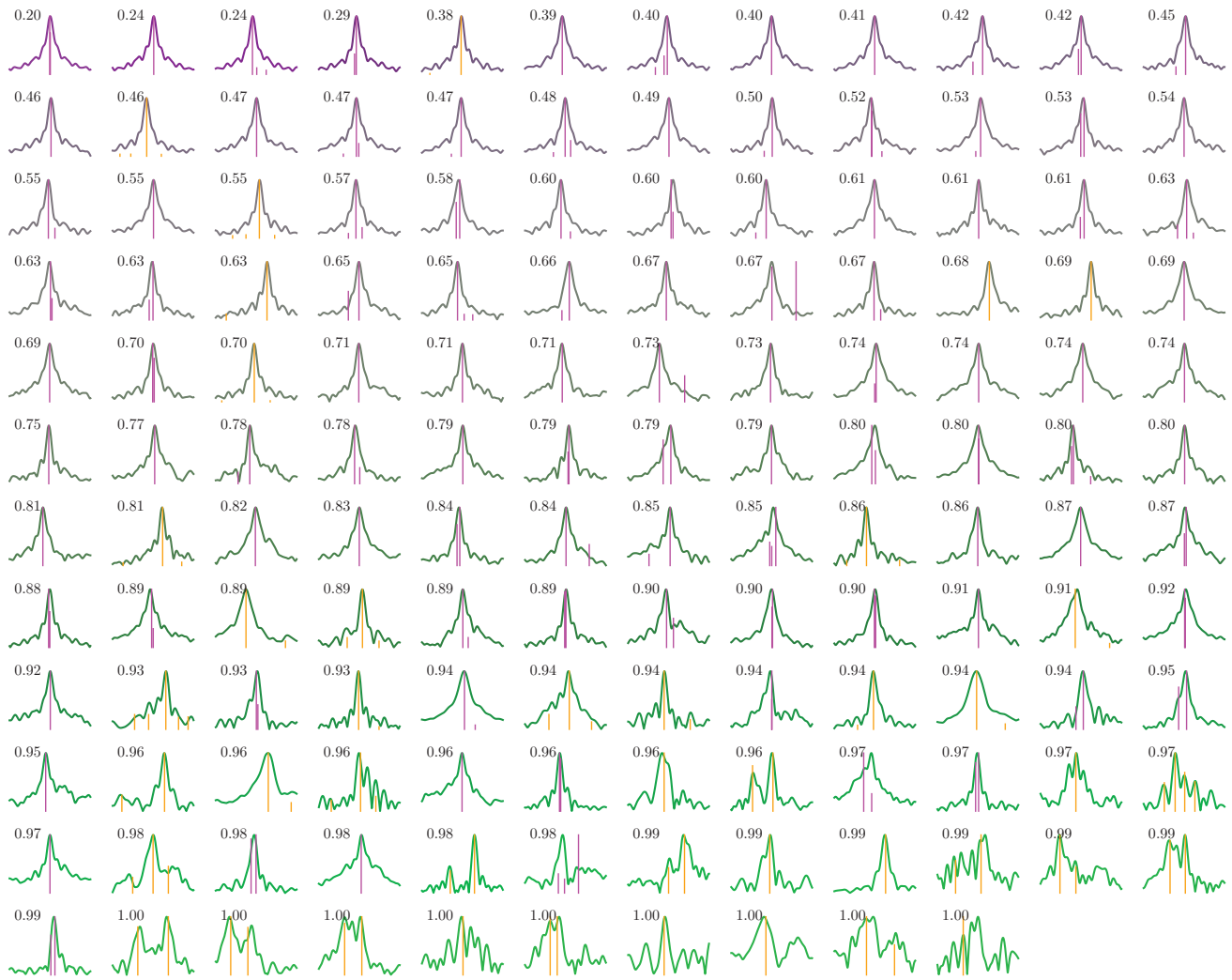


Figure D.1. The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the LR estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.

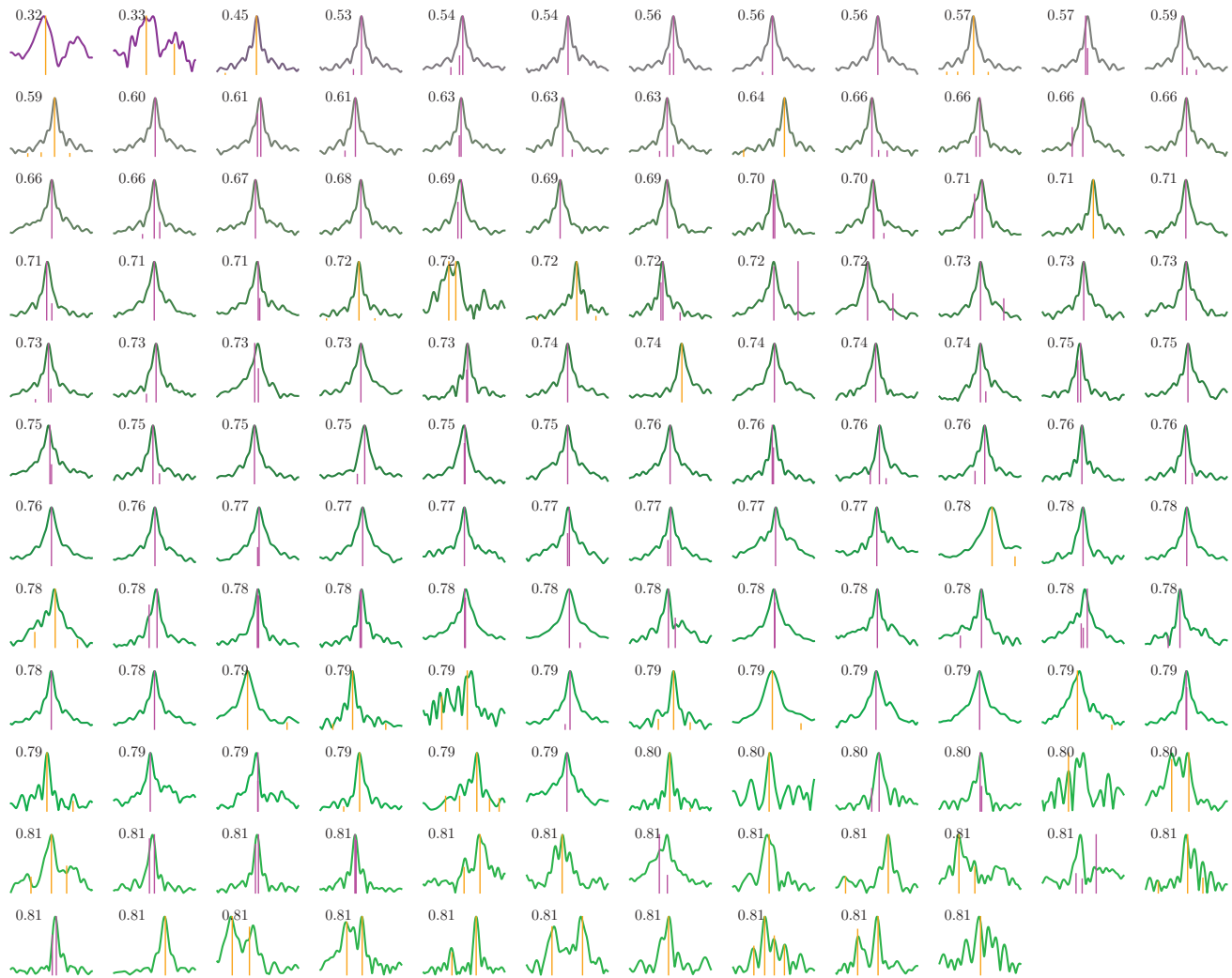


Figure D.2. The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the XGB estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.