To be published in the SPACEOPS 2023 Conference Proceedings, The 17th International Conference on Space Operations, 6 – 10 March 2023, Dubai, United Arab Emirates

Innovative Multi-carrier Broadband Waveforms Classification Using Machine Learning for Future GNSS Applications

Jose Mendez-Villanueva[1], Gustavo Sopena[2], Tien M. Nguyen[3], Charles H. Lee, and Sam Behseta
California State University, Fullerton, California, USA

Dan Shen, Genshe Chen, John Nguyen and Xiwen Kang
Intelligent Fusion Technology, Germantown, Maryland, USA

Khanh D. Pham
Air Force Research Laboratory, Space Vehicles Directorate, Kirtland AFB, New Mexico, USA

Abstract

This paper studies an innovative concept that uses machine learning for onboard satellite systems' operational environment prediction and signal classification. The study focuses on a newly proposed set of Global Navigation Satellite Systems using Multi-Carrier Broadband Waveforms (GNSS-MCBBW) with existing GPS waveforms (L1, L2, and L5). The Operational Environmental Predictor (OEP) and GNSS-MCBBW signal classifier under this study include (i) deterministic and Bayesian OEP prediction algorithms, (ii) signal features selection, a cumulant calculator, offline supervised machine learning module, along with a signal classification module using Support Vector Machine processing algorithm, respectively. The Operational Environmental Predictor (OEP) is proposed to enhance the GNSS-MCBBW signal classifier performance by reducing the uncertainties due to unknown operating conditions, such as satellite operating temperature and high-power amplifier (HPA) input back-off (IPBO). This paper aims to describe the design and MATLAB implementation of the proposed OEP and GNSS-MCBBW signal classifier. This can also be used to detect and identify the Most-Evil Waveform (MEWF) caused by non-ideal system components, along with presenting related modeling and simulation results to demonstrate the newly proposed concept.

1. Background and Introduction

The IFT-CSUF research team has recently developed a new GNSS-MCBBW by employing two onboard processing techniques (OBPT), namely, Hilbert Transform (HT) and bandpass filtering (BPF). These processing techniques can generate single sideband GNSS-MCBBW (SSB-GNSS-MCBBW) waveforms that fit the spectrum within the L-band bandwidth of 500MHz. The team has developed an end-to-end SSB-GNSS-MCBBW emulator with both ideal and non-ideal system component models to assess the proposed waveform performance. Specifically, performance under various non-ideal operations and unknown operational environmental conditions [1]. Our team observed that under non-ideal system operation conditions, such as an imperfect clock, unbalanced subcarrier modulator, non-ideal phase modulator, and non-linear HPA operation, the satellite system could generate undesired GNSS-MCBBW "Most Evil Waveform" (MEWF). Using the MEWF definition discussed in [2-4], the GNSS-MCBBW MEWF are GNSS-MCBBW waveforms that will produce the largest differential pseudo-range error (PRE) for a particular user terminal while also appearing entirely undetectable to the station monitor receiver. As discussed in [1], our team has proposed an advanced waveform adaptation module (AWAM) to identify the GNSS-MCBBW signal components for quick waveform adaption to the operational environment. This

[1] Graduate Student in Applied Mathematics, Lead developer, and MATLAB implementer of GNSS-MCBBW Classifier, California State University in Fullerton (CSUF).

[2] Graduate Student in Applied Mathematics, Lead MATLAB implementer of GNSS-MCBBW OEP, CSUF.

[3] Dr. Tien M. Nguyen is also with The Aerospace Corporation. He is a retired Engineering Fellow, Raytheon.

includes recognizing GNSS-MCBBW signal components and associated patterns, HPA operational temperature, HPA operation point, and other related system parameters for resolving the MEWF challenge. The AWAM includes a newly proposed OEP, an innovative GNSS-MCBBW signal classifier, and a Game engine processing module. This paper focuses on the design and MATLAB implementation of the OEP and signal classifier to detect and identify the GNSS-MCBBW MEWF caused by non-ideal system components.

At the beginning of the 2022 academic year, the Industrial Project for Graduate Program in Applied Mathematics (IPGPAM) initiated a collaboration project between CSUF and Intelligent Fusion Technology (IFT). The CSUF group consisted of eight graduate students and three CSUF faculty members to collaborate with the IFT team through this joint industry collaboration project. In early 2022, they addressed the challenges associated with the proposed OEP and signal classifier for future GNSS applications. This IPGPAM project focused on the advanced mathematical modeling and MATLAB simulation aspect of the signal classifier and OEP.

As discussed in [5], our previous work has been on the development of the next-generation GNSS-MCBBW emulator that can generate the MCBBW waveforms with non-ideal operating conditions such as (i) balanced modulator, (ii) phase modulator, (iii) reference satellite clock, (iv) intermediate frequency-to-radio frequency (IF-to-RF) converter, and (v) IF and RF filters. As discussed in [1], the GNSS-MCBBW waveforms can have different combinations of GPS L1, L2, and L5 waveforms and PCM/PSK/PM-sinewave subcarriers. This paper assumes the presence of GPS L1, L2, and L5 signal components simultaneously. The primary objective of this paper is to discuss the design and MATLAB implementation of the signal classifier and OEP using machine learning and artificial intelligence (ML-AI) technology for onboard prediction of operational environment conditions and signal classification of the newly proposed SSB-MCBBW signal components. Furthermore, simulation results for both HT and BPF techniques under non-ideal system operating conditions will also be presented.

The paper is organized as follows: (i) Section 2 provides an overview of the proposed GNSS-MCBBW satellite simulator; (ii) Section 3 discusses the proposed onboard GNSS-MCBBW waveform adaptation processing module (WAPM) with an emphasis on OEP, and GNSS-MCBBW signal classifier – The game engine processing module will be provided for the sake of completeness; (iii) Section 4 describes the MATLAB implementation of GNSS-MCBBW signal classifier and OEP; (iv) Section 5 presents modeling and simulation results; (v) Section 6 discusses the simulation results and potential future work; and (Vi) Section 7 concludes with a summary of the work done on the modeling and simulation of the non-OEP and signal classifier and our way forward.

2. Overview of the GNSS-MCBBW Satellite System Emulator

Figure 2.1 illustrates the MATLAB satellite system emulator. The emulator can generate GNSS-MCBBW signal using either Hilbert Transform (HT) or bandpass filter (BPF) processing technique. The user can select the HT technique by choosing the switch on the blue path connected to the HT node. The red path connected to the BPF node indicates the BPF technique. The proposed HT technique generates an upper single sideband (SSB) GNSS-MCBBW signal that fits the existing allocated L-band spectrum reserved for the GPS satellite system. The SSB BPF technique also generates an SSB GNSS-MCBBW signal that fits the specified L-Band spectrum. In addition, the satellite system emulator also includes: (i) an HPA pre-distorter (PD) using multi-objective reinforcement learning and adaptive neural network (MORL-ANN), (ii) an HPA model using actual L-band HPA AM-AM/AM-PM data collected from an industry HPA supplier, and (iii) waveform adaptive processing module (WAPM).

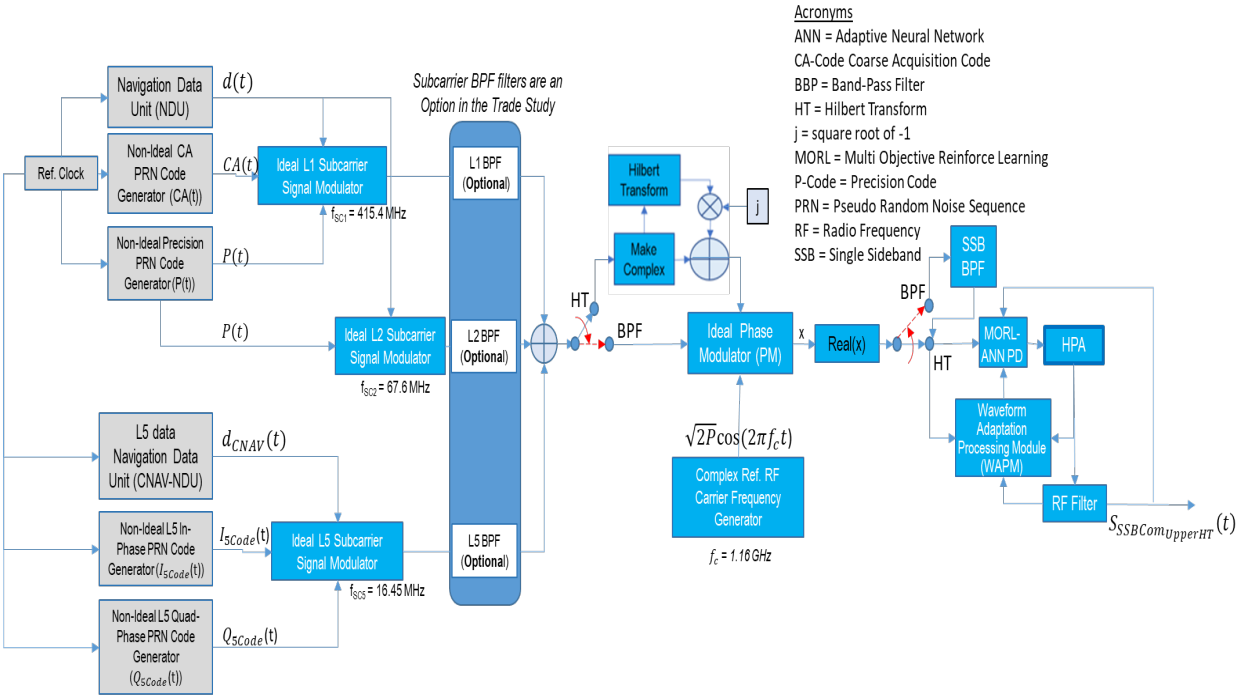

Figure 2.1: GNSS-MBBW Satellite System Emulator

3. Proposed Onboard GNSS-MCBBW Waveform Adaptation Processing Module (WAPM)

An overview of the waveform adaptation processing module (WAPM) is shown in Figure 3.1. The inputs to WAPM include the output of the RF filter and the HPA, as described in Figure 2.1. The WAPM consists of five models: the Operational Environment Predictor (OEP), Cumulant Calculator, Supervised Machine Learning model, Signal Classifier, and a game engine. The objective of the OEP is to predict the unknown operating conditions within the satellite systems, including system operating temperature, HPA operating point, and associated Input Power Back-Off (IPBO). The cumulant calculator is used as a feature to identify the impacts of non-ideal system components on L11, L12, L2, L51, and L52 signal components. The supervised machine learning model provides non-real-time training capability using available data to extract features and recognize the cumulant patterns of the GNSS-MCBBW signal components. The signal classifier provides real-time feature extraction of the received GNSS-MCBBW signal and input from the supervised machine learning process to recognize the cumulant patterns and accurately identify the GNSS-MCBBW signal components according to their cumulant features. The game engine model is used to optimize the transmitted signal using a payoff and cost function (PCF) subject to desired system performance objective. The following section provides additional details on these five models.

The signal classifier can also be designed in conjunction with the OEP, and the output of the signal classifier can be used to enhance the OEP performance. Moreover, the OEP provides a set of initial predicted environment conditions to the HPA PD that can help the pre-distorter converge faster.

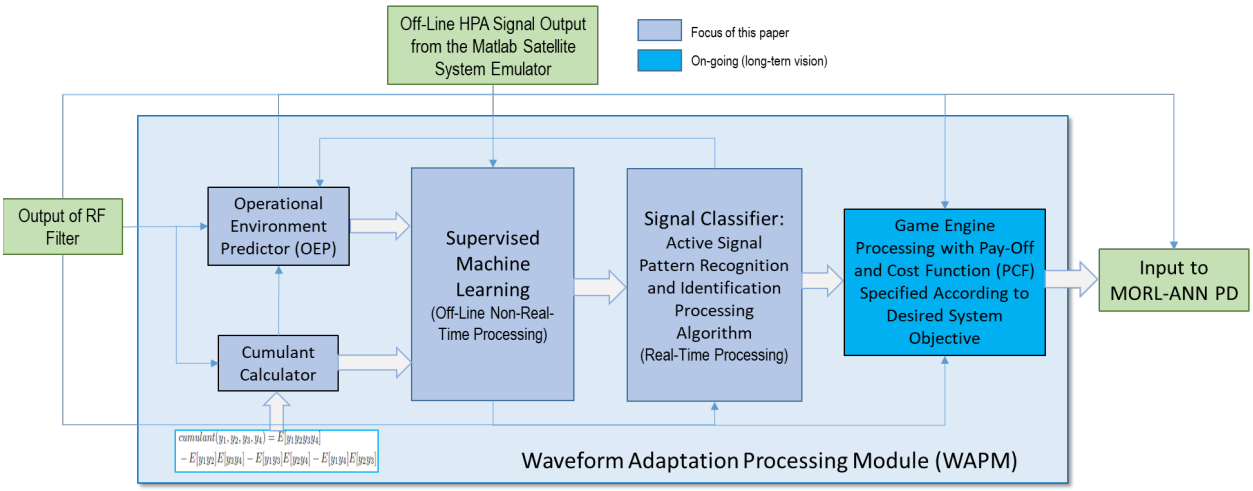

Figure 3.1: Waveform Adaptation Processing Module (WAPM)

3.1 Signal Features Selection

The purpose of the signal features selection is to identify specific signal features associated with the proposed GNSS-MCBBW signal components. The proposed GNSS-MCBBW signal consists of four key signal components, namely, the residual carrier signal component at the RF carrier frequency, f_C, the GPS L1 signal sitting at $f_C + f_{SC1}$, GPS L2 signal at $f_C + f_{SC2}$, and the L5 signal at $f_C + f_{SC5}$. The frequencies f_{SC1}, f_{SC2}, and f_{SC5} denote the subcarrier frequencies for L1, L2, and L5 signal components, respectively. Furthermore, the L1 signal has two signal components with the same subcarrier frequency f_{SC1} with the in-phase (I) L11 component and the quadrature (Q) L12, L2 has only the I component with subcarrier f_{SC2}, and L5 has the I and Q components like L1 signal with subcarrier f_{SC5}. By examining the proposed GNSS-MCBBW signal characteristics, the following features are selected to incorporate into the OEP and signal classifier processing algorithms:

- Modulation Technique: L1, L2, and L5 signals use QPSK, BPSK, and QPSK modulation formats, respectively;
- Signal Carrier Frequency, f_C: The carrier signal frequency is uniquely selected to produce specific L1, L2, and L5 frequencies at 1575.42 MHz, 1227.6 MHz, and 1176 MHz;
- Signal Subcarrier Frequency: The subcarrier frequencies f_{SC1}, f_{SC2}, and f_{SC5} for L1, L2, and L5 signal components, respectively, are selected based on the selected f_C and the L1, L2, and L5 frequencies;
- Signal Bandwidth: The bandwidth associated with each signal component L11, L12, L2, L51, and L52 are dependent on the PNR code rates. The L11 component has a CA code with null-to-null bandwidth (BW) of 1.023 MHz, and the L12 has a P-code with a BW of 10.23 MHz. L2 uses P-code with a BW of 10.23 MHz. The L51 and L52 both use 10.23 Mcps PRN codes with BW 10.23 MHz.

The use of these signal features will help to zoom in on the correct signal component with high probability. The remaining processing functions presented below will describe a process for accurate signal classification of the GNSS-MCBBW signal components in the presence of non-ideal signal component operations.

3.2 Cumulant Calculator

The MCBBW signal components are processed using a cumulant calculator after passing through the HPA. Here we use cumulant values, alternatives to the moments of continuous/discrete random variables as defined in Equation 1.

$$K_Y(t) = \ln E\left[e^{tY}\right] \quad (1)$$

In our simulations, we consider the signal as a discrete random variable. We obtain the second, fourth, and sixth-order cumulants as part of our cumulant calculator. We will denote this as C_{pq}, where the first subscript p indicates the order of the cumulant, and the second subscript q indicates the number of conjugates that are used in the calculation. Our signal input will be defined as Y.

$$C_{20} = C(Y_1, Y_2) = E[Y^2] \quad (2)$$

$$C_{40} = C(Y_1, Y_2, Y_3, Y_4) = E[Y^4] - 3C_{20}^2 \quad (3)$$

$$C_{60} = C(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6) = E[Y^6] - 15C_{20}C_{40} + 30C_{20}^3$$

We denote Equation 2, which is for the 2nd order cumulant value as C20, Equation 3, which is for the 4th order cumulant value as C40, and Equation 4, which is for the 6th order cumulant value as C60. Note that our cumulant values do not have any conjugates since the second subscript value is 0. We do not have any need for conjugates as we take apart the I/Q signal and use either the In-Phase signal or the Quadrature signal. This depends on which signal component we are looking at (i.e., signal component L11 uses an In-Phase signal). An important detail to highlight is that we are only taking in the numerical values of the In-Phase or Quadrature signal. Since we are taking in the numerical values of either the In-Phase or Quadrature signal, there is no need to use conjugates since we are removing 'imaginary' numbers. For instance, C_{40} turns out to be the same as C_{41}, C_{42}, and so forth due to only taking in the numerical values of either the In-Phase signal or Quadrature signal, meaning that using conjugates is not necessary. It should be noted that this is true when the system components are ideal, i.e., all system components are assumed to be operated within system specifications (a.k.a. in-spec) and in an ideal environment.

3.3 Offline Supervised Machine Learning Processing Model

As discussed in [1], the offline supervised machine learning processing model is described in Figure 3.2. This model uses the satellite system emulator described in [5] to generate and store training data. The three types of training data are (i) unscaled cumulants, (ii) scaled cumulants, and (iii) the decision boundaries associated with the scaled cumulants for all possible combinations of satellite system temperature, HPA IPBO, and non-ideal system operational conditions. The non-ideal system operational conditions consist of In-Spec and Out-of-Spec system components, where the Out-of-Spec case is considered the case where the PNT performance degradation is seriously affected. Each set of stored training data is tagged with a specific temperature, HPA IPBO, and system component operating condition. The stored training data can be used for both active real-time OEP and Signal classifiers. The signal features are described in Section 3.1. Each set of cumulants data is scaled according to features associated with it. Performance features such as S-curve bias, coherent autocorrelation function (CAF), correlation loss, and ambiguity function can be used as potential features. These performance features are not a focus of this paper.

3.4 GNSS-MCBB Signal Classifier Using Support Vector Machine (SVM)

To demonstrate the signal classifier, the supervised offline training processing model is configured and tuned for learning to identify five classes of GNSS-MCBBW signal components, including L11, L12, L2, L51, and L52. Using the signal features such as subcarrier frequencies and orthogonality of I-Q Components features; the offline training model can separate L11, L12, L2, L51, and L52 signal components and calculate the corresponding cumulants. For each satellite MCBBW signal component (e.g., L11), the

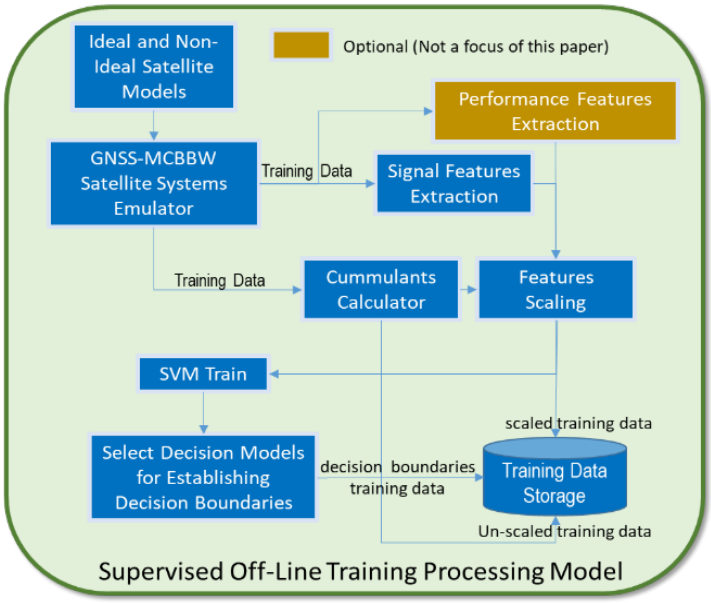

Figure 3.2: Architecture Framework of the Supervised Offline Training Processing Model

GNSS-MCBBW satellite system simulator generates the training data for all 18 possible cases for each satellite system setting of the non-ideal system parameters. As mentioned earlier, the three system temperatures [25 27 30] °C and six HPA IPBO are [0 2 4 6 8 10] dB, and the two imperfect system parameters setting are the In-spec (good case), and Out-of-Spec (bad case). Thus, offline learning processing requires learning all 18 x 2 system settings x 5 class-of-signal = 180 possible signal combinations and related Cumulant features, including C20, C40, and C60. The data was stored in the training data storage of the supervised offline training processing model. The training data are used by the signal classifier. Figure 3.3 illustrates an architecture framework of the proposed active signal classifier. The classifier extracts features and performs signal scaling using an optimum scaling technique to separate cumulant features. The signal classifier selects the features set and conducts a five-fold SVM configuration to determine the optimum decision surfaces associated with each class of signal (e.g., L11) and identify associated temperature, HPA IPBO, and imperfection conditions associated with the test data set.

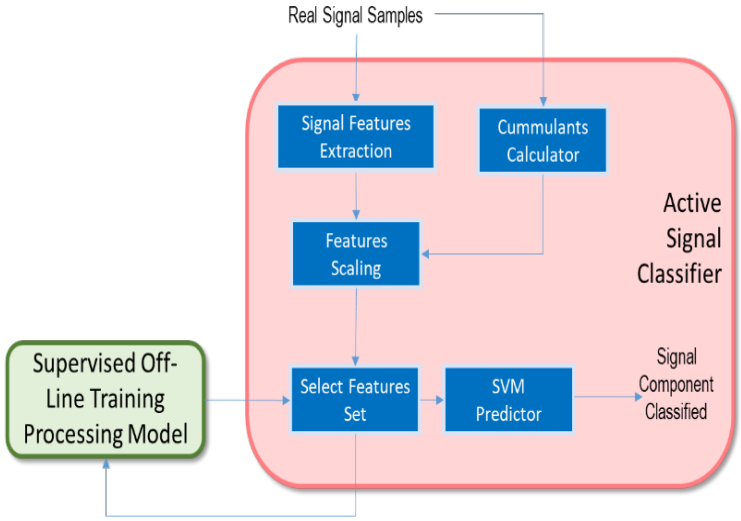

Figure 3.3: Architecture Framework of the Signal Classifier

Sections 4.1 and 5.1 describe the signal classifier MATLAB implementation model and related simulation results, respectively.

3.5 Operational Environment Predictor (OEP)

As a proof of concept, we consider a deterministic approach and the Bayesian approach to predict the satellite operating conditions [1, 6]. Figure 3.2 shows an OEP architecture framework using cumulants as the key feature for predicting the unknown temperature and HPA IPBO. The feature scaling is used to separate the cumulants cluster to effortlessly select a particular system operating feature set from the offline supervised machine learning processing model, thereby predicting the unknown temperature and IPBO. The unknown satellite operating conditions include temperature and IPBO of the HPA. The temperature assumes three operating temperatures 25°C, 27°C, and 30°C. The IPBO is assumed to have six possible even IPBO values in [0, 10]. These assumptions provide 18 possible temperature and IPBO combinations that are required for the data training collection. Thus, for each imperfect (non-ideal) system component setting (e.g., unbalanced amplitude and unbalanced phase associated with an imperfect balanced modulator), we must collect 18 possible combinations of temperature and IPBO. As an example, for 8 possible settings of the imperfect system components, we need to collect 144 (18x8) sets of cumulants as our training data. The training data is assumed to be stored in the data storage using the offline supervised machine learning model. The set of all possible temperature and IPBO combinations will be generated and stored in the training data storage provided by the offline training processing model, as depicted in Figure 3.2.

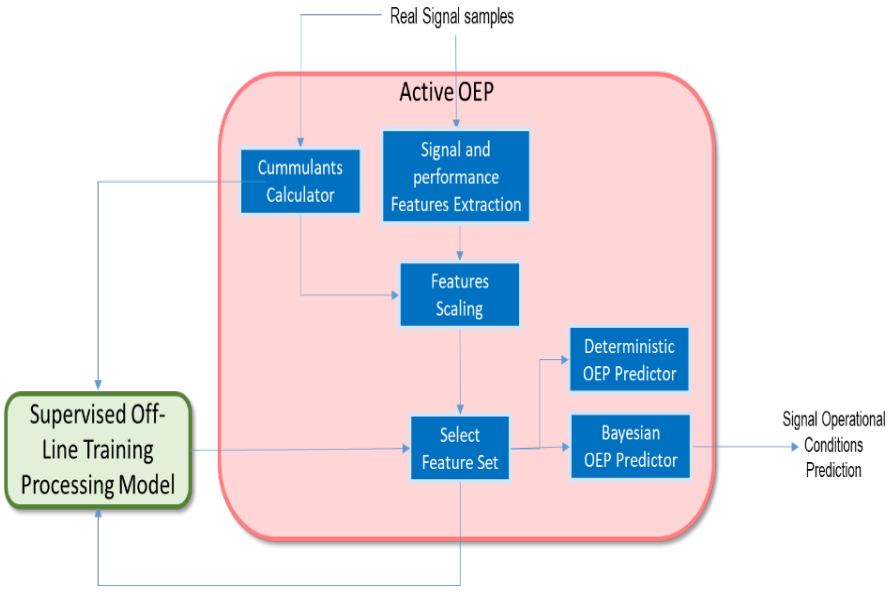

Figure 3.4: OEP Architecture Framework

For demonstration purposes of the Deterministic approach, we will assume that only the second-order cumulant, C_{20}, is used. The deterministic approach simply calculates the mean-square error (MSE) between real-time cumulant C_{20} of the signal of interest (SOI) and compares it with the stored C_{20} data. The set of stored C_{20} data with the temperature and IPBO identical to the SOI's C_{20} will be identified as the predicted satellite temperature and IPBO. Other cumulant orders can be used for predicting the satellite system operating condition. This is similar to [6], but we use cumulants as a feature instead of Bit Error Rate (BER) and bit signal-to-noise power spectral density ratio (EBNO).

The Bayesian approach employs a naive Bayes classifier like [1, 6] using cumulant as features instead of the BER and EBNO. This classification approach uses simple probabilistic classifiers based on Bayes' theorem with strong but naïve, independent assumptions between the features of the data. The prediction objective remains the same, i.e., to classify the 18 possible temperature and IPBO combinations. In the MATLAB model implementation described below, we fit a Normal Naïve Bayes classifier model, effectively training a model by passing in our cumulant values of C20, C40, and C60. We pass in all System Types of an Ideal System, such as Non-Ideal Systems with In-Spec operating conditions, Non-Ideal Systems with Out-of-Spec operating conditions, and Non-Ideal Systems combined with both In-Spec and Out-of-Spec operating conditions. Note that the In-Spec operational condition can be considered an ideal environment since the PNT performance degradation is negligible when the system components operate within the specification. For Out-of-Spec operating conditions, the system components operate outside the system specification causing potential PNT performance degradation. Such degradation can range from mild but acceptable to severe and unacceptable degradation conditions. We will consider the Out-of-Spec case that can cause severe PNT performance degradation for the demonstration. Sections 4.2 and 5.2 describe the OEP MATLAB implementation model and related simulation results.

3.6 Game Engine Processing Module

The game engine model is not the focus of this paper. It is currently an ongoing research task. The goal of the game engine is to develop a pay-off-and-cost function (PCF) that can be used to (i) optimize the power division among the RF residual carrier, L1, L2, and L5 signal components, and (ii) mitigate the MEWF caused by the imperfections associated the non-ideal system components.

4. MATLAB Implementation of the GNSS-MCBBW Signal Classifier and OEP

This section describes the MATLAB implementation approach for the signal classifier and OEP models. Figure 4.1 showcases the general MATLAB workflow for the Signal Component Classification that uses a Support Vector Machine (SVM), a Deterministic OEP, and a Bayesian approach to OEP. Figure 4.1 shows that there are two processes, one that is offline, and the other that is online. The offline process accesses stored training data and respective processes. The online process generates new data that then passes through its respective 'trained' model to predict signal components and operating conditions. It is worth mentioning that the Deterministic OEP does not utilize the signal component classifier as the Bayesian OEP, this will be implemented in future work. It should be noted that the training cumulant data is assumed to be generated offline and stored in the green database box labeled "Training Cumulant Data," as shown in Figure 4.1. The training data generates all desired combinations of IPBOs, HPA operating temperature, and non-ideal system components operations.

4.1 GNSS-MCBBW Signal Classifier MATLAB Model

As shown in Figure 4.1, the MATLAB implementation of the proposed Signal Classifier fits into the MATLAB workflow describing the signal processing flow for identifying and classifying each GNSS-MCBBW signal component, including L11, L12, L2, L51, and L52. It is worth mentioning that a scaling scheme as this significantly impacts the accuracy of classification. Further detailed in [7], we see how the scaling impacted the spread of data for both the signal classifier and OEP. Finding an optimal scaling scheme will help with classification as well as convergence for the proposed machine learning model. Several scaling and transformation schemes were tested to optimize classification. Approaches that were taken were frequency scaling, chip-rate scaling, sine/cosine transformations, and leveraging certain I-Q aspects. Going into more detail, before the signal went through the cumulant calculator, we attempted scaling the signal by the desired frequency factor. For example, we would scale by either the L1, L2, or L5

frequency to determine if this would somehow lead to a better separation of cumulant values. The same idea was considered for the chip rate scaling factor. Any scaling before the cumulant calculator did not improve the classification performance. Thus we tried affecting the data after the cumulant calculations by using typical functions (i.e., sine/cosine transformations) and even normalizing the data to have a mean of unity. The last failed approach involved the I and Q channels of the complex signals. We attempted to somehow leverage the I-Q channels for the differentiation of cumulant values. The normalization scheme of Min-Max scaling was believed to best separate the cumulants. We then used the following Min-Max normalization scheme of Equation 5, but that was not helpful in cumulant value separation.

$$\frac{(X - X_{Min})}{X_{Max} - X_{Min}} \tag{5}$$

where $X = Cumulant\ Value$

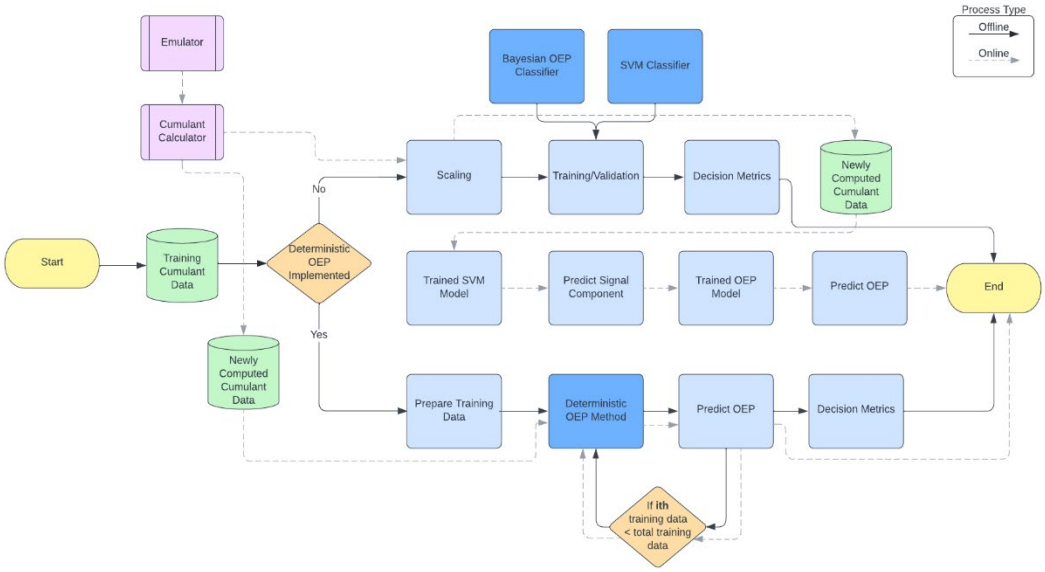

Figure 4.1: MATLAB GNSS-MCBBW Signal Classifier & OEP Workflow

This led to a Min-Max scaling within a specific boundary of choice (i.e., 0 to 2) but again there was no change from the original Min-Max scheme. This led to tweaking this Min-Max scaling of Equation 6,

$$a + \frac{(X - X_{Min})(b - a)}{X_{Max} - X_{Min}} \tag{6}$$

where $X = Cumulant\ Value$, $a = Lower\ Bound$, and $b = Upper\ Bound$.

to eventually get to Equation 7,

$$a + \frac{X - X_{Min}\,(b - a)}{X_{Max} - X_{Min}} \tag{7}$$

which finally gave us the separation that was desired in [7]. The validation scheme that is implemented is a K-Fold Cross Validation where K=5. This support vector machine is interested in classifying signal components L11, L12, L2, L51, and L52. Meaning that we have a multi-class classifier since we are interested in classifying a total of 5 different signal components. Support vector machines are natively binary classifiers meaning we implement this multi-class support vector machine using a One vs. One

scheme. Hence, we split the multi-class data into binary comparisons. Figure 4.2 shows notional boundaries for L11, L12, L2, L51, and L52. The righthand side of Figure 4.2 shows the One vs. One scheme in which it compares one signal to another signal until all respective signals have been compared against each other; each signal is depicted by different colored surfaces. Note in Figure 4.2 that signals that are not compared are colored as grey surfaces. Choosing an optimal hyperplane in which the margin between classes is the largest is important. From this, using a Gaussian Kernel Scheme gave the highest level of accuracy and thus was the kernel of choice.

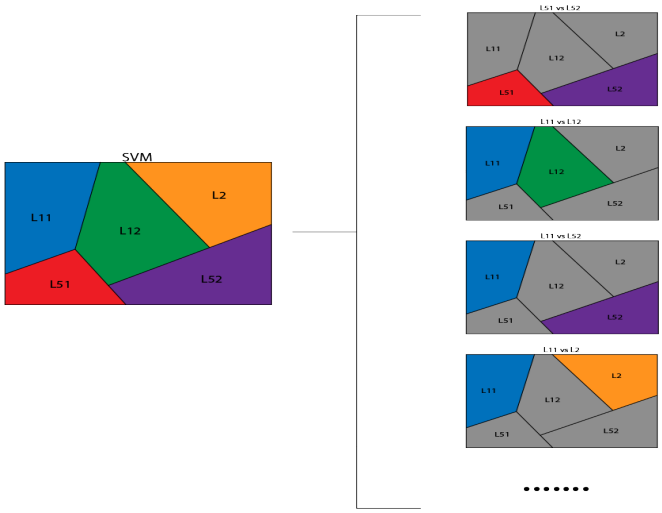

Figure 4.2: SVM One vs. One Binary Scheme with Arbitrary Decision Surfaces for L11, L12, L2, L51, L52.

While the above focuses on specifications on how to optimize the classification due to data scaling and classifier schemes, we now mention several cases where the signal classification will take as input. Several cases for both the Bandpass Filter Onboarding Processing Technique (BPFOBPT) architecture and Hilbert Transform Onboarding Processing Technique (HTOBPT) architecture were investigated. Two parameter settings within the BPFOBPT and HTOBPT architectures were specifically investigated; here, the amplitude imbalance and phase imbalance are changed to make the alpha parameter (coefficient of unsuppressed carrier component) either less than -20dB (In-Spec) or greater than -20dB (Out-of-Spec) [5]. Both onboard processing technique architectures were tested for using several different System Types that consist of an Ideal System, a Non-Ideal System with In-Spec operating conditions, a Non-Ideal System with Out-of-Spec operating conditions, and a Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions.

4.2 OEP MATLAB Implementation Model

This section describes two different MATLAB model approaches used to classify the operating conditions. As mentioned in section 3.2, we use signal features, behaviors of performance features, and cumulant patterns of the MCBBW distortions to predict the onboard operational conditions.

4.2.1 Deterministic Approach

It is important to note that we are unable to obtain the Bit Error Rate (BER) and Energy Per Bit to Noise Power Ratio (EBNO) as featured. Working within the satellite system provides us full control of signal transmitted power without propagation loss and noise in the environment while onboard the satellite. Hence, these values are replaced with the cumulant orders of interest. A similar approach to [6] was constructed to

accommodate the cumulant values as predictors: a squared difference of values. The satellite emulator was tuned to work with three temperatures, 25°C, 27°C, and 30°C, as well as six IPBO values, 0 dB, 2 dB, 4 dB, 6 dB, 8 dB, and 10 dB. The emulator was also configured to three of four System Types: Ideal, In-Spec operating conditions, and Out-of-Spec operating conditions. Once cumulant data has been obtained, the model modifies this data by finding the mean of the data per temperature and IPBO combinations. This means that we have 18 classes that the model can predict. Training data for our model is composed of the 18 mean data points across combinations of the three temperatures and six IPBO values. Testing data for our model is a cumulant value for the specified signal component (an observation). In Figure 4.3, the testing value can be visualized across the IPBO values since we would not know which IPBO it is associated with (the red line in Figure 4.3).

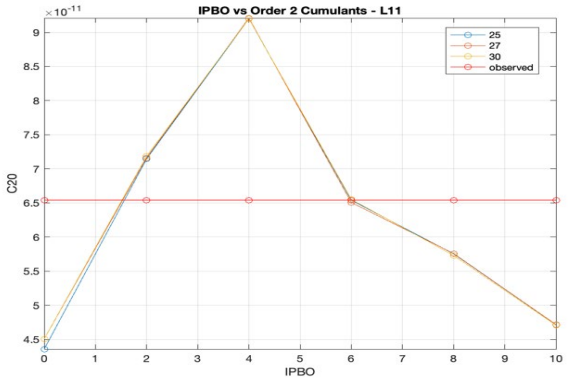

Figure 4.3: IPBO versus modified 2nd order cumulant values for signal component L11 (Ideal System)

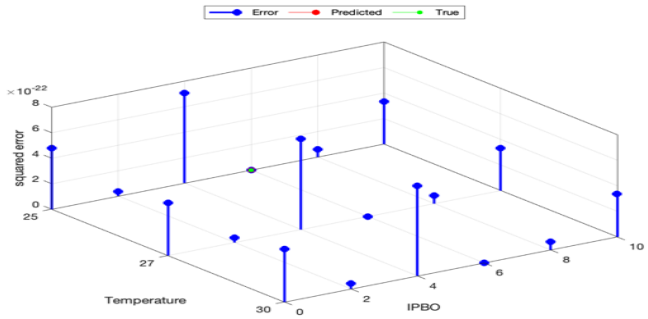

Figure 4.4: Squared Errors used for temperature and IPBO prediction for 2nd order cumulant values for signal component L11 (Ideal System)

To obtain a prediction for the temperature and IPBO, the model sums up the square differences between each point of the red line, along with the points of the other lines (one for each of the temperature values). The prediction for the temperature and IPBO from our model would be the smallest squared error. Figure 4.4 shows the squared differences, denoted as squared errors. We can see that most squared errors roughly share the same altitude from the temperature and IPBO plane. In Figure 4.4, the classification is chosen to be the pair: temperature 25°C, IPBO 6 dB.

4.2.2 Bayesian approach

A MATLAB function file is created to store the Bayesian classifier to be used as a call-up function in the driver file. In this driver file, the data is pulled from the database and then scaled, as was mentioned in

section 4.1. After the model is trained/cross-validated, analyzed, and eventually used for online predictions. Once both the signal component classifier and the Bayesian classifier have been trained, we can transfer information into our MATLAB prediction file. Once the signal component has been classified, we use that classification to determine the signal component. It is important to mention this as we will have a total of 5 different OEP models, one for each signal component. Meaning we have an L11 OEP, L12 OEP, L2 OEP, L51 OEP, and L52 OEP. In the prediction file, once the signal component has been classified from new cumulant data, we then use that classification to know which OEP model to use to predict the operational temperature and IPBO.

5. MATLAB Simulation Results

This section presents the simulation results for OEP and signal classifiers under various operational operating conditions and imperfect system components.

5.1 GNSS-MCBBW Signal Classifier MATLAB Simulation Results

We now analyze our results of both the BPFOBPT and HTOBPT architectures that were investigated. Table 5.1 shows that for Cases 1 and 2, the signal classification accuracies are 100% and 99.8% for ideal system components using BPFOBPT and HTOBPT, respectively.

Table 5.1: Signal Component Classification Results

Case	OBPT Architecture	System Type	Accuracy
1	Bandpass Filter	Ideal	100%
2	Bandpass Filter	Non-Ideal (In-Spec & Out-of-Spec)	84%
2.1	Bandpass Filter	Non-Ideal (In-Spec)	77%
2.2	Bandpass Filter	Non-Ideal (Out-of-Spec)	87%
3	Hilbert Transform	Ideal	99.8%
4	Hilbert Transform	Non-Ideal (In-Spec & Out-of-Spec)	87.2%
4.1	Hilbert Transform	Non-Ideal (In-Spec)	86.4%
4.2	Hilbert Transform	Non-Ideal (Out-of-Spec)	77%

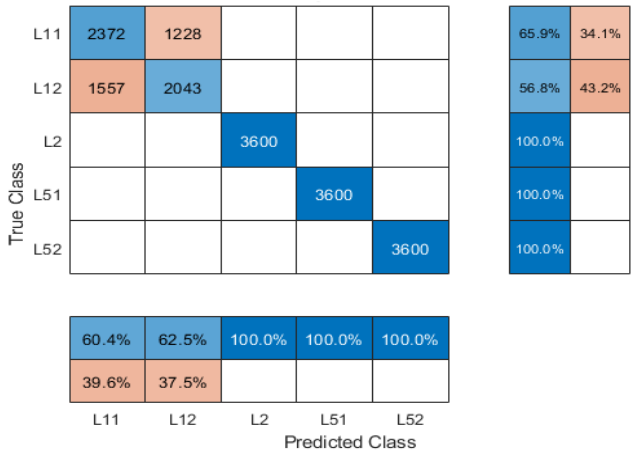

Figure 5.1: Case 2- BPF Non-Ideal System Confusion Matrix

Case 2 (BPF with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions) and Case 4 (HT with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions)

in Table 5.1 will be analyzed in further detail. The following analysis will be performed by looking at Case 2 (BPF with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions).

Figure 5.1 shows that by dealing with a non-ideal system, we are still able to classify most signal components correctly. Analyzing the percentages on the right-hand side, we see that these are True positive rates versus False positive rates. However, looking at the percentages at the bottom, these are positive predicted rates versus false predicted rates. When looking at the True positive rates versus False positive rates from Figure 5.1, we have a large percentage of misclassification of signal component L11 with signal component L12 and vice versa. Overall, this provides an excellent classification for this case. Even though signal component L11 and signal component L12 are misclassified against each other, we are getting 65.9% of the correct classification for signal component L11 and 56.8% correct classification for signal component L12.

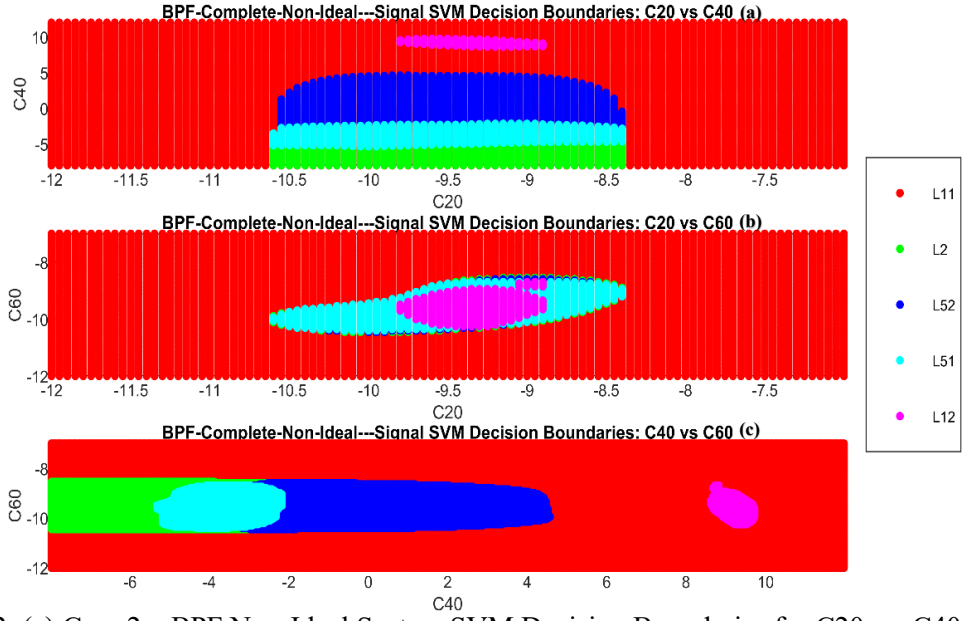

Figure 5.2: (a) Case 2 – BPF Non-Ideal System SVM Decision Boundaries for C20 vs. C40 where the title text 'Complete-Non-Ideal' means that both subcases 2.1 and 2.2 are included. (b) Case 2 – BPF Non-Ideal System SVM boundaries for C40 vs. C60 (c) Case 2 – BPF Non-Ideal System SVM boundaries for C40 vs. C60

Figure 5.2 displays the decision surfaces of our signal component classification made from a 3D mesh grid of the minimum and maximum cumulant values (C20, C40, & C60) with padding to both the upper and lower bounds. Using the 3D mesh grid, this is then passed into the trained model to predict, and the results are displayed in Figure 5.2. In Figure 5.2, there are three different decision surfaces; in Figure 5.2a, we have C20 vs. C40; in Figure 5.2b, we have C20 vs. C60; and in Figure 5.2c, we have C40 vs. C60. Although only one decision surface (i.e., C20 vs. C60) could be considered for analysis, it is of great value to acknowledge all three different decision surfaces. Using a 3D mesh grid means that we have a 3D decision surface, but not all cumulant decision boundaries can be seen from one decision surface. This is shown in Figure 5.2b. Here we do not see all the relevant information that is needed as there is an overlap between the decision boundaries. Here we also notice that C20 and C60 are highly correlated. Whereas if we look at Figure 5.2c, we are able to see the decision boundaries very clearly. This will clearly help in knowing which cumulant values (C20, C40, & C60) will correspond to which signal component, assuming the accuracy of classification of the trained model is high enough to utilize. Furthermore, from these decision surfaces, we can see how our model will predict and function on different cumulant values. From our confusion matrix above we know which classes will give us the correct boundaries. This implies that one cannot rely on the boundaries of signal components L11 or L12 when it comes to looking at these decision boundaries. This is

due to the accuracy not being high enough to completely believe in those decision boundaries. Regardless of this, we can see where the boundaries between different signal component cumulant values are and where potential confusion might be present.

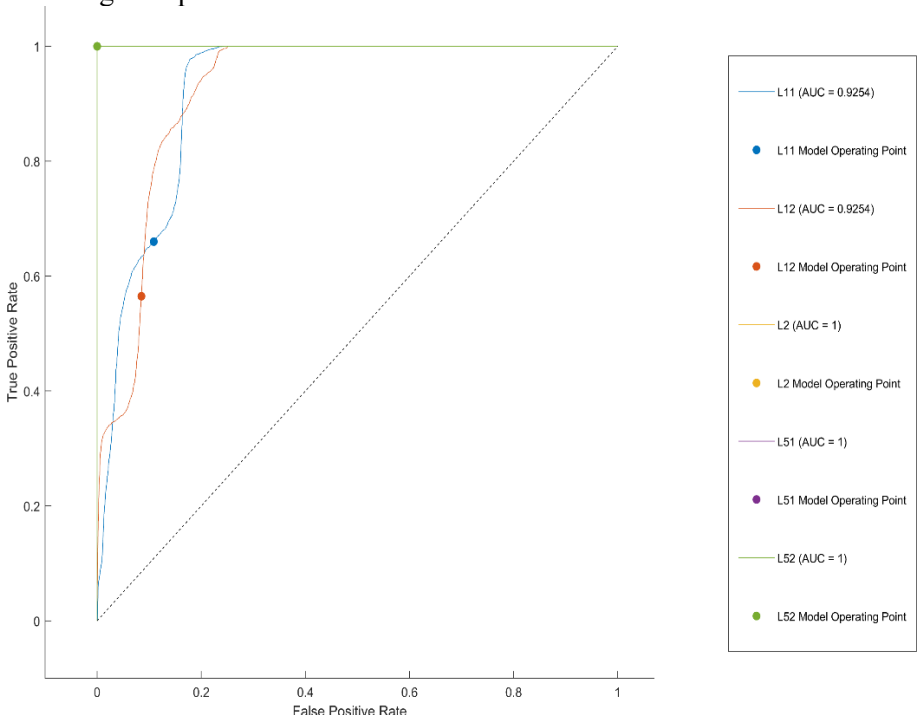

Figure 5.3: Case 2 – BPF Non-Ideal System ROC Curve

Figure 5.3 shows a ROC (Receiver Operating Characteristic) curve and the curves corresponding to the Area Under Curve (AUC) value. Visually one can see three curves when in fact, there are five curves. This is due to the perfect classification of the signal components L2, L51, and L52. When analyzing the corresponding AUC values, one can see that all three have AUC values of 1. This means that they are being classified perfectly. While the AUC value is important, it does not fully define the true nature of the classification, as we will see. When we shift our focus to the operating points, we see that for signal components L2, L51, and L52, the operating point is at the top left corner, thereby being classified perfectly. Now if we look at signal component L11, we see that the AUC is high with a value of 0.92, but if we look at the operating point, we can determine that it is around 0.6, more specifically when looking at the Y-axis (True Positive Rate). This shows how the classification is functioning overall and at different decision thresholds.

The following analysis examines Case 4 (HT with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions). Figure 5.4 shows us great improvement in classification between signal components L12 and L11 compared to the BPFOBPT approach. Presented is a near-perfect classification for signal component L12. The issue seems to be that the classifier cannot differentiate between signal component L11 and signal component L12. While this approach might be perceived to significantly perform better than the Bandpass Filter approach, due to the individual classification, we only see an increase of 3% between Case 2 (BPF with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions) and Case 4 (HT with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions).

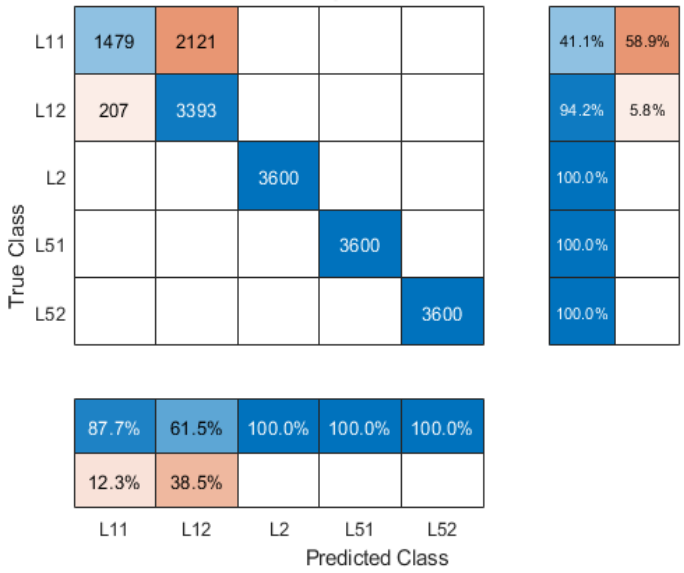

Figure 5.4: Case 4 – HT Non-Ideal System Confusion Matrix

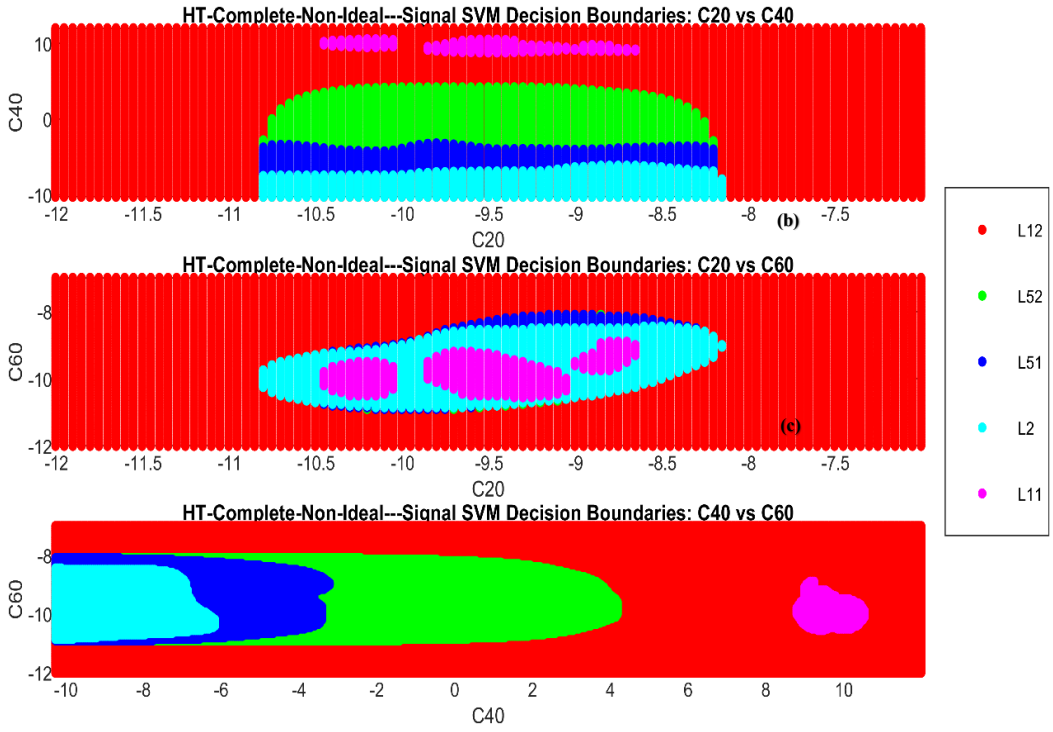

Figure 5.5: (a) Case 4 – HT Non-Ideal System SVM Decision Boundaries for C20 vs. C40 where the title text 'Complete-Non-Ideal' means that both subcases 4.1 and 4.2 are included. (b) Case 4 – HT Non-Ideal SVM boundaries for C40 vs. C60 (c) Case 4 – HT Non-Ideal System SVM boundaries for C40 vs. C60

In Figure 5.5, we note that signal components L12, L2, L51, & L52 are reliable decision surfaces in which we know the cumulant values that these signal components will have. We take note that the classifier misclassifies signal component L11 and signal component L12 more than 50% of the time, suggesting that we should disregard this decision boundary.

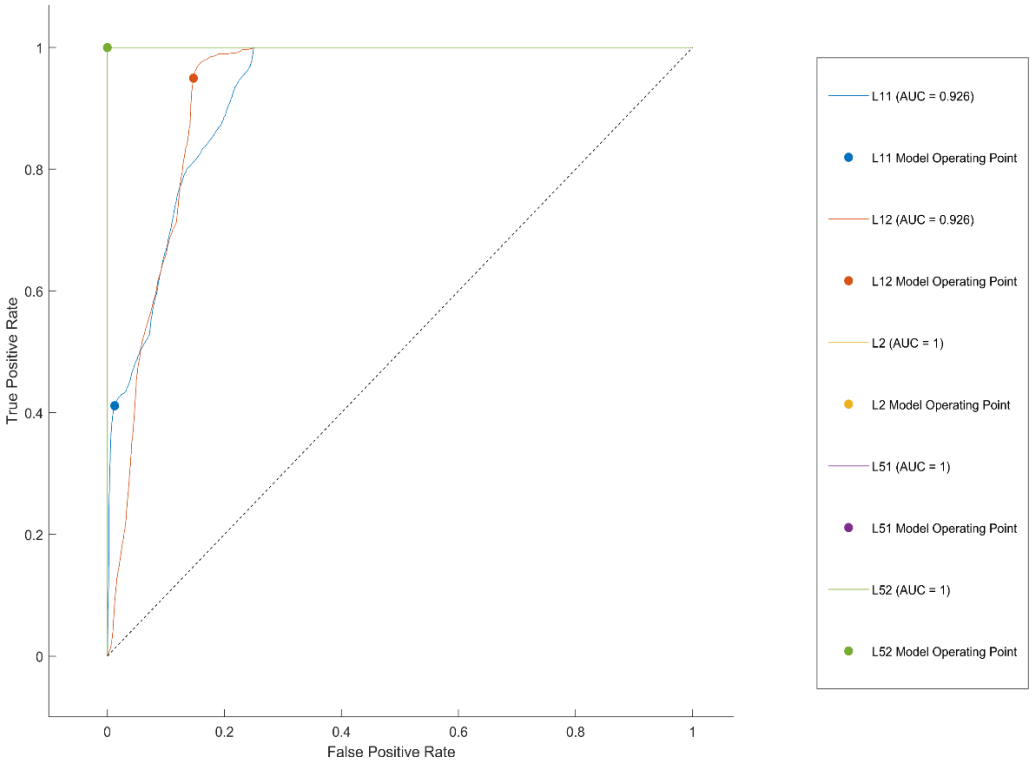

Figure 5.6: Case 4 – HT Non-Ideal System ROC Curve

Figure 5.6 shows us similar results as in Figure 5.3 for signal components L2, L51, and L52. The differences are the operating points for signal components L11 and L12. We see that signal component L12 is near perfect as it is very close to the top left corner. Whereas signal component L11 is very low and just above .4 when looking at the Y-axis (True Positive Rate).

Overall, both the BPFOBPT architecture and the HTOBPT architecture provide accuracies above 80% but have difficulty classifying between signal components L11 and L12. There are a few differences in how both architectures classify signal components L11 and L12, but they classify signal components L2, L51, and L52 with ease.

5.2 OEP MATLAB Simulation Results

This section shows the simulation results from the OEP MATLAB models described in Section 4.2 under various system imperfections and operating conditions.

5.2.1 Deterministic Approach

To demonstrate proof of concept, we were only able to apply our deterministic approach to the second-order cumulant values for signal component L11 for the BPF Onboard Processing Technique (BPFOBPT). In addition, as part of the training data, we consider different System Types for an Ideal System, Non-Ideal System with In-Spec operating conditions, and Non-Ideal System with Out-of-Spec operating conditions. Moreover, as part of the testing data, the model took as input the same data as the training data. Since we have considered three System Types for training and testing data, then we will have nine combinations of testing and training data from these System Types. The pairings are as follows: Ideal system cumulants alongside modified versions of these cumulants, Ideal system cumulants alongside modified Non-Ideal system cumulants (In-Spec), Ideal system cumulants alongside modified Non-Ideal system cumulants (Out-

of-Spec), Non-Ideal system cumulants (In-Spec) alongside modified Ideal system cumulants, Non-Ideal system cumulants (In-Spec) alongside modified versions of these cumulants, Non-Ideal system cumulants (In-Spec) alongside modified Non-Ideal system cumulants (Out-of-Spec), Non-Ideal system cumulants (Out-of-Spec) alongside modified Ideal system cumulants, Non-Ideal system cumulants (Out-of-Spec) alongside modified Non-Ideal system cumulants (In-Spec), and Non-Ideal system cumulants (Out-of-Spec) alongside modified versions of these cumulants. Table 5.2 illustrates the combinations and the accuracies of our method resulting from the data sets.

Table 5.2: OEP Deterministic Method Results for Several Combinations of System Types with The BPFOBPT architecture

	Ideal Train	**Non-Ideal Train (In-Spec)**	**Non-Ideal Train (Out-of-Spec)**
Ideal Test	34.94%	5.55%	5.55%
Non-Ideal Test (In-Spec)	5.55%	18.77%	8.16%
Non-Ideal Test (Out-of-Spec)	5.55%	7.77%	18.05%

It is determined that the lowest accuracy is 5.55% for four of the nine off-diagonal cases in Table 5.2. The following accuracies are 7.77% and 8.16% in the remaining off-diagonal cases in the table. The highest accuracies occur on the diagonal, which obtain values of 34.94%, 18.77%, and 18.05%. These cases show that when the training data and testing data inputted into the model are the same, there is increased accuracy.

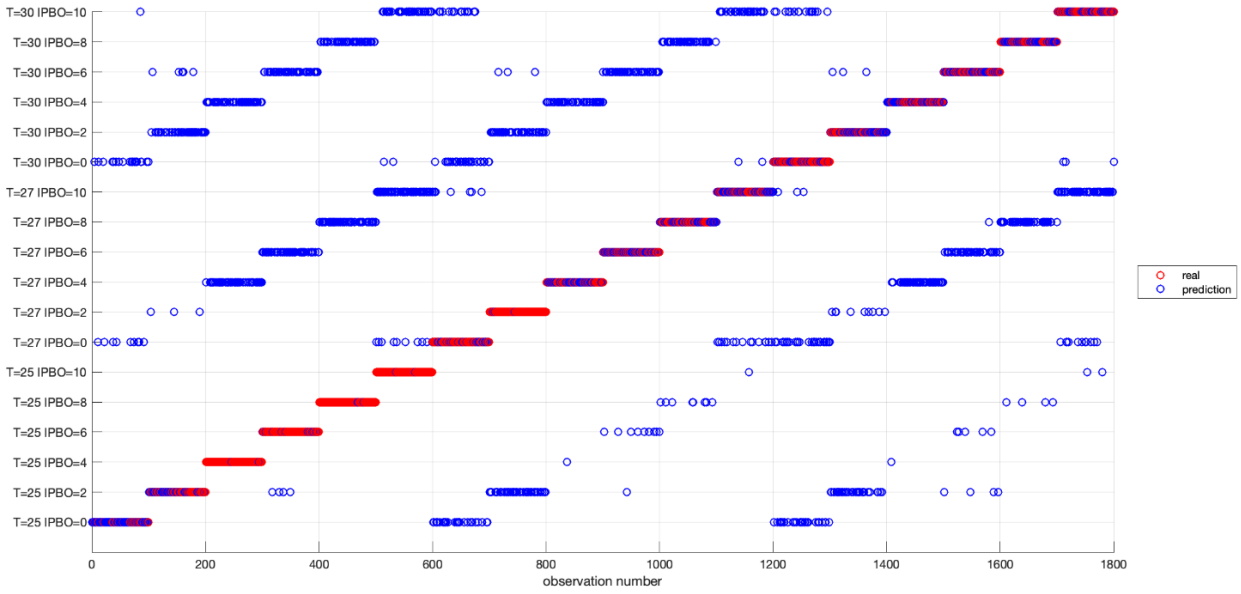

Figure 5.7: Case 1 – BPF Ideal System L11 Classification Tally

We elaborate on how we obtained the 34.94% accuracy measure from the model. As the minimum squared differences are calculated, the model keeps a tally of the number of correct classifications. Figure 5.7 shows how the model compares to the true and predicted temperature and IPBO combination (y-axis) for each of the observations from the testing data (x-axis). Each observation has a blue mark and a red mark, where a blue mark indicates a prediction from the model, and a red mark indicates the true classification for the

given cumulant value. In Figure 5.7, we see the case when the training and testing data are the same, namely, the data from an ideal satellite emulation system with the BPFOBPT architecture. Overall, out of the 1800 observations, we obtain 629 correct classifications. By extension, the other percentages in Table 5.2 are found in a similar manner.

5.2.2 Bayesian approach

Before looking at the results of Case 2 (BPF with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions) and Case 4 (HT with Non-Ideal System combined with both In-Spec and Out-of-Spec operating conditions), it is important to take note of the spread of cumulant data points for given temperatures and IPBO's. When analyzing these data points, it is evident that there are minor to no changes between the given temperature and IPBO for said signal component (i.e., signal component L11). Therefore, it is expected to receive low accuracies of correct classification in identifying both the correct temperature and IPBO.

Table 5.3: OEP Case 2 - BPF Non-Ideal System Simulation Results

Case	OBPT Architecture	System Type
2	Bandpass Filter	Non-Ideal (In-Spec & Out-of-Spec)

Signal Component	Accuracy
L11	13%
L12	10%
L2	7%
L51	7%
L52	8%

Looking at Table 5.3, we see that the signal component L11 has the highest accuracy of classification in terms of both temperature and IPBO. We focus on one specific signal component to do further analysis of the Operational Environment Predictor.

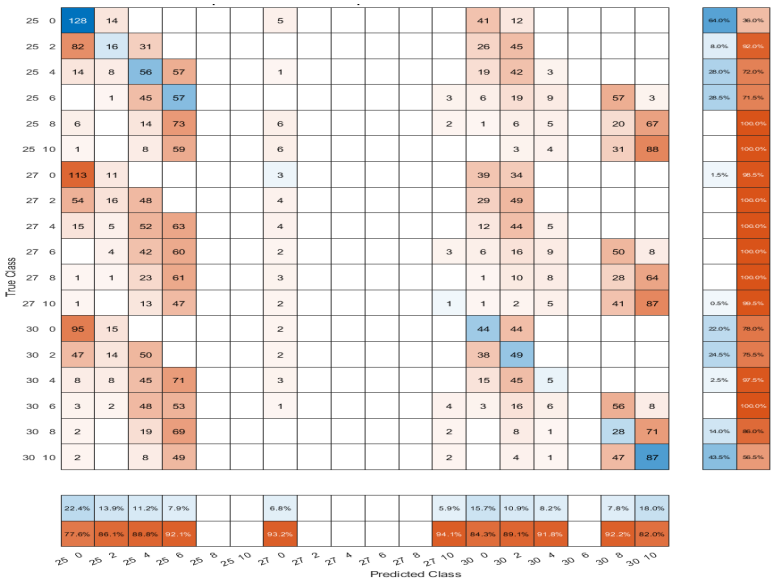

Figure 5.8: Case 2 – BPF Non-Ideal System L11 Confusion Matrix

Figure 5.8 shows that only temperature 25°C IPBO 0 dB gives an accuracy of correct classification higher than 50%, whereas all the other classes are severely misclassified. This is expected as the cumulant data points for different temperature and IPBO combinations are similar.

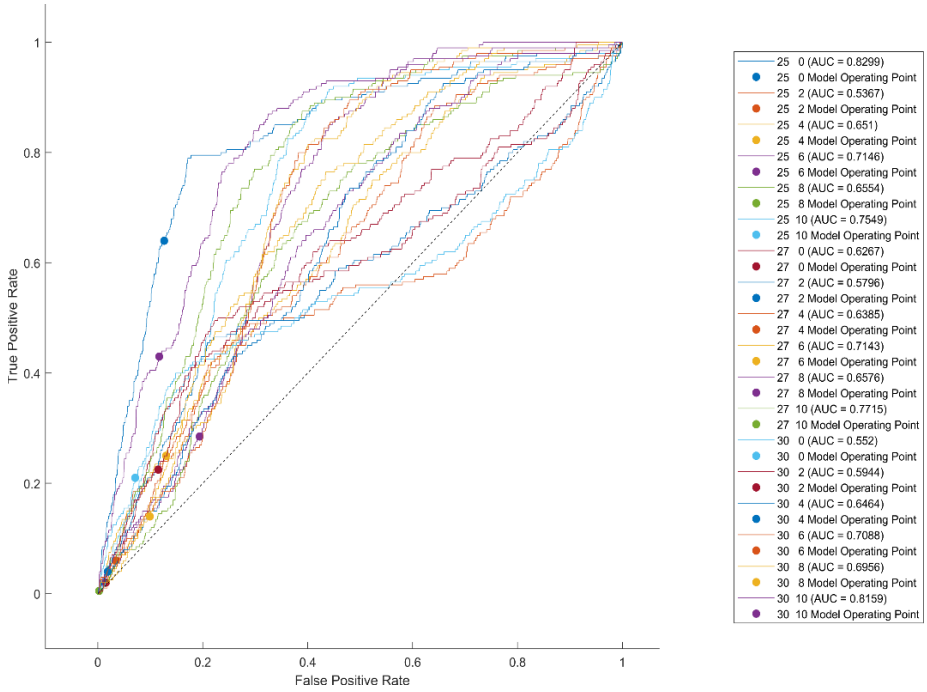

Figure 5.9: Case 2 - BPF Non-Ideal System L11 ROC Curve

Figure 5.9 reiterates that each class is poorly classified. Thus, we deem this as our curves are very close to being linear and on the diagonal dotted line. Several curves are found to even dip below the diagonal dotted line which is not a desirable characteristic that we want in our ROC Curve. Lastly, when looking at the operating points we see that most fall below 0.2 on the y-axis (True Positive Rate).

Table 5.4: OEP Case 4 - HT Non-Ideal System Simulation Results

Case	**OBPT Architecture**	**System Type**
4	Hilbert Transform	Non-Ideal (In-Spec & Out-of-Spec)

Signal Component	**Accuracy**
L11	9%
L12	9%
L2	9%
L51	11%
L52	14%

Examining Table 5.4, we see that our accuracy for L11 is worse than in Table 5.3. Here the signal component with the highest accuracy is L52, with a 14% accuracy of classification for combinations of temperature and IPBO. Next, we analyze the same signal component L11, but this time for Case 4.

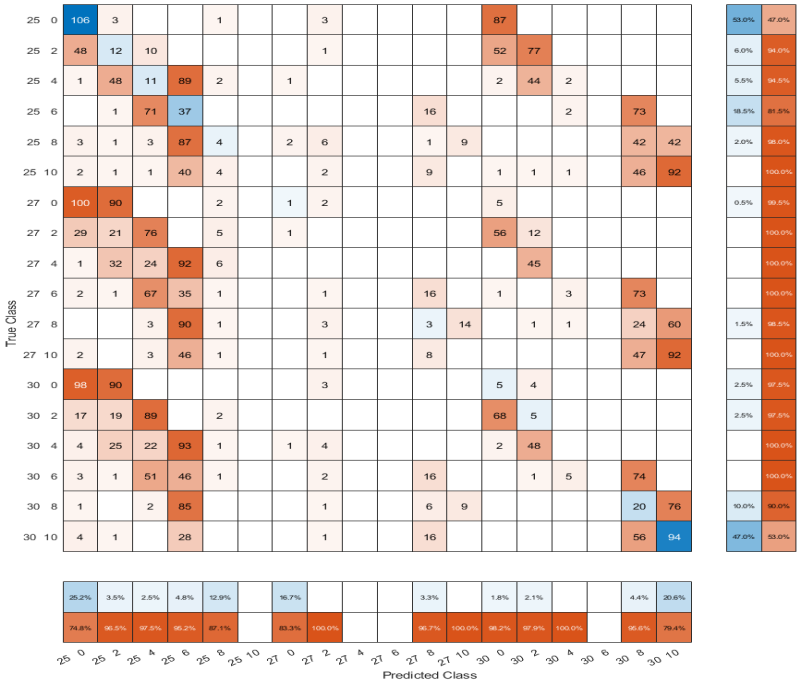

Figure 5.10: Case 4 – HT Non-Ideal System L11 Confusion Matrix

Figure 5.10 confirms that the accuracy of signal component L11 is 9%. We can see that only two configurations of temperature and IPBO give accuracies of a correct classification close to 50%. These two configurations are for temperature 25°C IPBO 0 dB and temperature 30°C IPBO 10 dB. Once again, this was expected as the cumulant data points for different combinations of temperature and IPBO are almost identical.

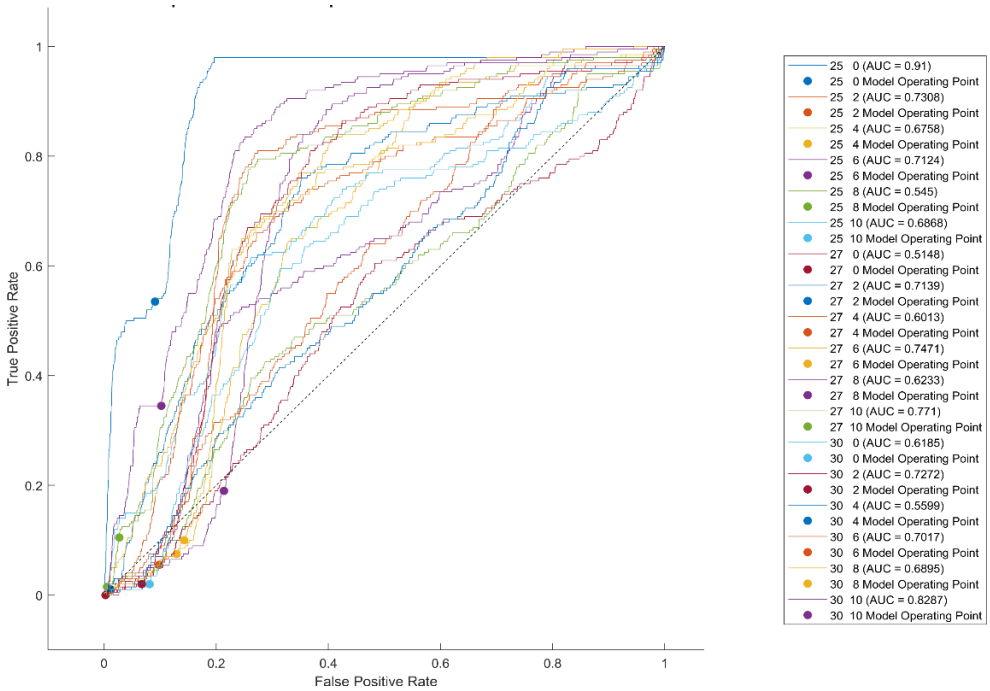

Figure 5.11: Case 4 – HT Non-Ideal System L11 ROC Curve

Figure 5.11 displays curves that are almost as linear as the dotted diagonal line signifying poor classification results across all decision thresholds. This is confirmed by focusing on the operating points, as well as the low AUC values.

6. Discussion and Future Work

The Signal Classification performed very well using the Support Vector Machine with a Gaussian Kernel and One vs. One scheme. The innovative scaling technique was applied uniformly to all cases, but it is worth noting that scaling should vary between cases to use an optimal scaling for that given data. The reason is that for signal components L11 and L12, there was very close clustering in most cases, thus resulting in these signal components being highly misclassified. This misclassification happens due to the I-Channel (In-Phase) and Q-Channel (Quadrature) of L1 being too similar, and as we know, L11 is the I-Channel of L1, and L12 is the Q-Channel of L1. While L11 uses Random CA-Code PRN Chips and L12 uses Random Precision-Code PRN Chips, we still receive cumulant values that are too similar. There are several approaches to address this issue. One of the approaches is to leverage [8] as a potential solution for the high misclassification of these signal components. This solution includes using hybrid images where we layer the In-Phase, Quadrature, and Power Spectral Density (PSD) images on top of one another and pass it through a Convolutional Neural Network (CNN). This image will be passed in as an RGB image, where each component of the image will correspond to either the R-Channel, G-Channel, or B-Channel. Meaning that the In-Phase image will be part of the R-Channel, the Quadrature image will be the G-Channel, and the PSD image will be the B-Channel. This will be particularly helpful because we can extract features from each channel separately. In terms of how this will affect our existing architecture, we plan on using this if one of the signal components goes under a specified accuracy threshold. We would call upon a CNN to help improve this particular signal component classification accuracy. The other option will be to leverage this hybrid image as a feature in our SVM classifier.

The other approaches are to exploit (i) the complex conjugates of the cumulants, (ii) S-Curve bias as a new feature along with those cumulants that closely cluster, and (iii) Correlation Loss as a new feature.

In other scaling schemes, we noted that signal components L2 and L51 were clustered together, meaning we need to implement a systematic approach to always have these signal component data values separated. An approach that can be taken would be to always calculate the Euclidean distance between signal components and find the minimum distance to separate this data to get good classification accuracy. Further improvements can be made by changing the SMV's kernel or binary comparison scheme.

Our modifications to the model for the deterministic approach to calculating the predictions of temperature and IPBO values did not perform well. We note that we did not use the innovative scaling technique for this approach. The unscaled data is likely the key factor for the poor classification. At the time of writing this paper, we were unable to check the performance of the model against the other signal components. However, we believe that they are likely to also result in low accuracies. Moreover, higher accuracies obtained from this approach are likely due to using the same training and testing datasets. As such, the observations closely matched the means derived for each temperature and IPBO combination. In addition, our Normal Naïve Bayes classifier model did not perform as expected. We will likely have to incorporate several different scaling techniques to find an optimal scaling technique for each given case and model of the OEP. Potentially leveraging hybrid images as a feature to our Naïve Bayes classifier can improve accuracy, or altogether approaching the task of predicting operational conditions with a CNN as mentioned above might be ideal for this task. Further improvements for this classifier can be made by changing the kernel of the Naïve Bayes classifier and, as mentioned above, incorporating correlation loss as a feature.

Another idea regarding optimal scaling for both the signal component classification and the OEP includes scaling before the cumulant calculator. It is possible to look at several different positions within our process which can be deemed as the optimal scaling position to investigate for the best results. Not only this, but we can use our entire I/Q signal before passing it into the cumulant calculator which would allow us to use conjugates within the cumulant calculator. These future works will help improve our performance results.

The simulation results have demonstrated the use of ML-AI for onboard signal component classification. It is important to note that by monitoring and observing the identified signal components, we can identify the waveform distortions and predict the corresponding sources of distortion. The steps corresponding to "identify waveform distortion" and "predict the source of distortion" are referred to as "fault diagnosis" and "fault prognosis," respectively. Furthermore, by monitoring the identified source of distortions, we can "detect distortion exceeds threshold" and "adapt the waveform for distortion compensation." These steps correspond to "fault detection" and "fault recovery," respectively.

7. Conclusion

Overall, our signal component classification performed relatively well except for signal components L11 and L12. As Section 6 mentioned, it is possible to increase the correct classification of signal components L11 and L12 via many different methods. While we hoped that the OEP would perform better, there were several reasons that we did not achieve acceptable performance. A couple of these reasons include needing more features and optimal scaling for each combination of temperature and IPBO for each signal component. Continuing, our ideas for future work include optimization of the scaling factor to improve the classification/prediction performance and use this new scaling as a feature to detect and identify the source of imperfection. While the OEP models did not perform as expected for combinations of temperature and IPBO due to the cumulant values being too similar for all different combinations, it was noted that for subcases of the given onboard signal processing architecture (i.e., Cases 2.1 and 2.2), the cumulant values were distinguishable between In-Spec and Out-of-Spec. Therefore, we can truly identify the source of imperfection. These implementations will significantly help the signal classification, the OEP, and extended work such as onboard fault diagnosis, prognosis, detection, and recovery for future GNSS applications.

Correspondence
Dr. Tien M. Nguyen (CSUF POC), 800 N State College Blvd, Fullerton, CA 92831, e-mail: tmnguyen57@fullerton.edu.
Dr. Charles H. Lee, (CSUF Industrial Project for Graduate Program in Applied Mathematics, Director), 800 N State College Blvd, Fullerton, CA 92831, e-mail: charlesHLee@fullerton.edu.
Dr. Genshe Chen (Contract PM POC), 20410 Century Blvd, Germantown, MD 20876, e-mail: gchen@intfusiontech.com.
Dr. Dan Shen (Contract Chief Engineer, Technical POC), 20271 Goldenrod Ln, Germantown, MD 20876, e-mail: dshen@intfusiontech.com.
Dr. Khanh D. Pham (AFRL PM POC), 3550 Aberdeen Avenue S.E., Kirtland AFB, NM 87117, e-mail: Khanh. Pham.1@ spaceforce.mil.

Acknowledgment

The authors would like to thank Ms. Maria Rios, Technical Communication Specialist, IFT, for her support during the preparation of this paper.

References

[1] Tien M. Nguyen, Charles H. Lee, Yinwei Chen, Sam Behseta, Dan Shen, Genshe Chen, John Nguyen, Xiwen Kang, Khanh D. Pham, "Innovative Multicarrier Broadband Waveforms for Future GNSS Applications – A System Overview," Accepted for presentation and publication in the 202 IEEE/ION Position Location and Navigation Symposium (PLANS) Proceedings, April 24 – 27, 2023, Monterey, California.

[2] Alexander M. Mitelman, Jaewoo Jung, Per K. Enge, "LAAS Monitoring For A Most Evil Satellite Failure," 1999 ION National Technical Meeting, Proceedings of the 1999 National Technical Meeting, 1999, pp 129-134, San Diego, California.

[3] Per K. Enge, Robert Eric Phelts, Alexander M. Mitelman, "Detecting Anomalous Signals from GPS Satellites," ICAO, GNSS/P, Toulouse, France, 1999.

[4] Robert Eric Phelts, "Multicorrelator Techniques for Robust Mitigation of Threats to GPS Signal Quality," Sections 1.6, 1.7.1 3.3.2, and 3.4, Ph.D. Dissertation, June 2001, Standford University.

[5] Josselyn Romero, Leo Amador, Tien M. Nguyen, Charles H. Lee, Yinwei Chen, Sam Behseta, Dan Shen, Genshe Chen, John Nguyen, Xiwen Kang, Khanh D. Pham, "GNSS-Multicarrier-Broadband-Waveform Satellite System Emulator," Accepted for presentation and publication in the SPACEOPS 2023, The 17th International Conference on Space Operations, 6 – 10 March 2023, Dubai, United Arab Emirates.

[6] Tien M. Nguyen, Charles H. Lee, Sean Cantarini, Xuanyu Huang, Jennifer Gudgel, Chanel Lee, Cristal Gonzalez, Ground-Based HPA Pre-Distorter using Machine Learning and Artificial Intelligent for Satellite Communication Applications, a Book Chapter, submitted to the Book titled "Data and Decision Sciences - Recent Advanced Applications," to be published by IntechOpen Publisher in early 2023.

[7] Charles Lee, Jose Mendez-Villanueva, Gustavo Sopena, Tien Nguyen, Genshe Chen, Dan Shen, YinWei Chen, "Operational Environment Prediction and Support Vector Machine for MCBBW Subcarrier Components," Final Report from IFT-CSUF Research Team as a requirement for the IPGPAM

[8] Hilal Elyousseph, Majid L Altamimi, "Deep Learning Radio Frequency Signal Classification with Hybrid Images," arXiv:2105:09063 [cs], May 2021