SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Spatiotemporal Forecasting of Wireless Coverage and Frequency Availability with Sparse Geo-location Spectrum Databases

Vladimir II Christian R. Ocampo[1,*], Lawrence Materum[1,2]

[1] De La Salle University, Manila, 1004 Metro Manila, Philippines
[2] Tokyo City University, Setagaya City, Tokyo 158-8557, Japan

ARTICLE INFO

Article history:
Received
Received in revised form
Accepted
Available online

Keywords:

Channel availability; Forecasting; Geolocation; Database; TVWS

ABSTRACT

The television spectrum can contain unused frequencies or channels called white spaces. The white spaces can be managed to provide internet access in coordination with surrounding TV channels to avoid interference. Geo-location databases are helpful for dynamically shared frequencies when updated and complete. In real life, the spectrum availability for a secondary user lacks numerous information; hence, it is sparse. This paper aims to forecast wireless coverage and frequency availability in such sparse geolocation databases. Spatiotemporal models are formulated in this paper to forecast wireless coverage and frequency availability in sparse geo-location spectrum databases. Eight channels are explored by this study, and the data used are gathered from TV program guides sourced online and to the best knowledge of the researcher. The forecasting models are evaluated using accuracy, precision, recall, and F1 score. Spatiotemporal logistic models had a decent accuracy of at least 84%. The linear VAR models have a decent accuracy except for predicting time. The formulated spatiotemporal logistic VAR model attained the highest accuracy of at least 94%.

1. Introduction

Television (TV) broadcast is characterized by transmitting audio and video through electromagnetic (EM) waves. Carrying information on EM waves occupies a part of the EM spectrum. The spectrum can be divided into parts called channels. The unused portions of the EM spectrum are called white spaces.

The white space can be utilized for non-broadcasting purposes such as internet services and emergency communications. Due to the rise of network usage demand and the need for communication access in remote areas, white space usage has been researched. Television white space (TVWS) is a growing technology that allows the dynamic use of frequency spectrum, which can benefit the rural and underserved areas of developing countries. In other words, rural areas are left out. TVWS can serve as a super-high-speed Wi-Fi called Wi-Fi 2.0, which can help provide

* *Corresponding author.*
E-mail address: vladimir_ii_ocampo@dlsu.edu.ph

https://doi.org/10.37934/araset.62.1.117

communication access to rural and underserved areas. The applications of TVWS include but are not limited to emergency tracking and environment monitoring.

The frequency spectrum is dynamically shared across space and time. The disadvantage is that TVWS is not protected from interference, especially from primary incumbents like TV stations. A way to coordinate the white space devices must be conceived. One method is spectrum sensing, where the device listens and determines unoccupied frequencies. However, its design and performance are challenging according to the tests conducted by the Federal Communications Committee (FCC) [1]. A better option is to use a geo-location database that contains what frequencies are occupied at a given place and time together with other information. A white space device (WSD) can query the database and receive information such as frequency availability from the geo-location database. However, geo-location databases are sparse, and forecasting using sparse TVWS geo-location databases has not been undertaken.

Different studies have investigated this research topic. Studies have explored forecasting using geolocation databases, such as using deep learning [2]. Other studies have used time series and vector autoregression models for forecasting. However, there are no studies that have explored incorporating sparsity in forecasting, especially using a geolocation database. The study aims to forecast wireless coverage and frequency availability given sparse geolocation spectrum databases. In particular, the study formulated a sparse forecasting model that predicts wireless coverage and frequency availability, implemented programs for the forecasting model, and evaluated the predictive accuracy of the model.

1.1 Literature Review

Several studies have explored different methods of spectrum management to achieve dynamic spectrum sharing between primary and secondary users. Moreover, some studies dealt with the TVWS geo-location database. Shawel *et al.,* [2] used deep learning methods for making predictions using a geo-location database. The spectrum can also be managed through the real-time secondary spectrum market (RTSSM) and game theory [3]. The General Enhanced Detection Algorithm (GEDA) was conceived by Martin *et al.,* [4] as an enhancement for the detection of primary users and optimization of the secondary user's Quality-of-Service (QoS). In addition, studies have explored predictions using models in specific applications such as an Artificial Neural Network (ANN) on predicting bubble pressure on Sudanese oil field [5], a tNavigator model to predict gas coning [6], and Anand's model to predict solder inelasticity [7]. In addition, a study utilized the "You Only Look Once" (YOLO) algorithm to detect Personal Protective Equipment workplace violations [8]. However, sparse geo-location data was not explored by any studies.

Different studies have touched on the concept of sparsity. A study made a comparison of various sparsity measures [9]. Benarabi *et al.,* [10] sparsity measure takes the ratio of the number of non-zero elements to the total number of elements. Goswami *et al.,* [11] examined the sparsity of a network graph using the Gini index. There has been no exploration of the integration of sparsity in predictive analytics yet.

Time series models were some of the forecasting models used. The Adversarial Sparse Transform (AST) was derived from Generative Adversarial Networks (GANs), as referred by the study of Guerra-Montenegro *et al.,* [12]. The study of Ardia *et al.,* [13] predicts high-dimensional data from computed textual sentiment using time series aggregation. Moreover, spatiotemporal methods, such as predicting crimes in real time [14] or urban traffic flow [15], were applied to forecasting. In addition, vector autoregression (VAR) models are also used for forecasting. A VAR model is typically applied in forecasting economics since it captures interactions between multiple time series. For instance, COVID-19 infections in the United Arab Emirates (UAE), Saudi Arabia, and Kuwait were modelled based on VAR. The predictions are accurate enough with a very low Mean Absolute Percentage Error (MAPE) [16]. Katris [17] used a modified VAR model to see the impact of COVID-19 on the unemployment rate in Greece. They compared it with Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) univariate models. The study shows that the VAR and ARIMA models perform better than ANN. However, no studies examine sparse forecasting capabilities for geo-location databases, contributing to this study.

2. TV White Space

A television traditionally uses an antenna to pick up signals embedded in EM waves transmitted by broadcasters, referred to as primary users (PU) in this context. These PUs transmit signals in the air at specific frequencies. A television tuned to that frequency picks up that signal and extracts the information present. Primary users occupy a tiny portion of the wide frequency spectrum. This situation leaves most of the spectrum unused, which could be an opportunity for communication. Those unoccupied frequencies are called white spaces.

Regulators worldwide have considered the unlicensed use of TVWS. The FCC was the first to recommend this utilization for more effective use of TV spectrum resources. They designated two types of devices: fixed devices and portable devices. Afterward, other countries around the globe followed, from Europe to Asia and the Pacific. Ofcom in the UK issued WSD specifications in 2009. Europe's Electronic Communications Committee (ECC) released technical and operational requirements in 2011. Infocomm Development Authority (IDA) in Singapore established the white space pilot group (SWSPG) in 2012. Radio Spectrum Management of New Zealand settled on creating a licensing scheme that allowed the usage of TVWS devices. White space specifications, such as frequencies, power limits, bandwidth, device type, and seeking method, depend on regulators worldwide.

In the Philippines, the Department of Information and Communication Technology (DICT), through its attached agency, the National Telecommunications Commission (NTC), released a memorandum circular that regulates the unused TV broadcast channels for the utilization of TVWS technology [18]. TVWS pilot testing was done in Bohol and Leyte, Philippines, in 2014, which proved beneficial for communications and services of disaster and relief efforts after the Bohol earthquake and super typhoon Yolanda [19]. Moreover, companies that specialize in TVWS were set up around the world. Whizpace Pte. Ltd offers white space technology such as Whizmesh and Whizrange [20]. Carlson Wireless Technologies, Inc. in California, USA, offers RuralConnect Gen3TM systems such as the Trailblazer and LongHaul Time Division Multiplexing (TDM) [21].

In addition, there are several bodies developed for TVWS. IEEE 802.22 is the approved ISO standard for Cognitive Based Wireless Regional Area Networks that use beaconing as interference protection. Data rates vary from 22 to 29 Mbps per TV channel without using multiple-input multiple-output (MIMO). IEEE 802.11af is a standard conceived from 802.11ac and operates on 6, 7, or 8 MHz bands or multiples.

An application of TVWS is Super Wi-Fi, as it has a more extended range than traditional Wi-Fi. Super Wi-Fi uses the TV broadcast spectrum as Wi-Fi instead of 2.4 GHz or 5 GHz and can penetrate through walls, making it feasible for remote areas. Moreover, Super Wi-Fi is easy to deploy cable-free in different environments and has variable and high data rates of up to 54 Mbps. Figure 1 compares data rates between traditional Wi-Fi and Super Wi-Fi. It shows that TVWS Super Wi-Fi covers many kilometres, consequentially increasing bandwidth, lowering network costs, and reducing power consumption.

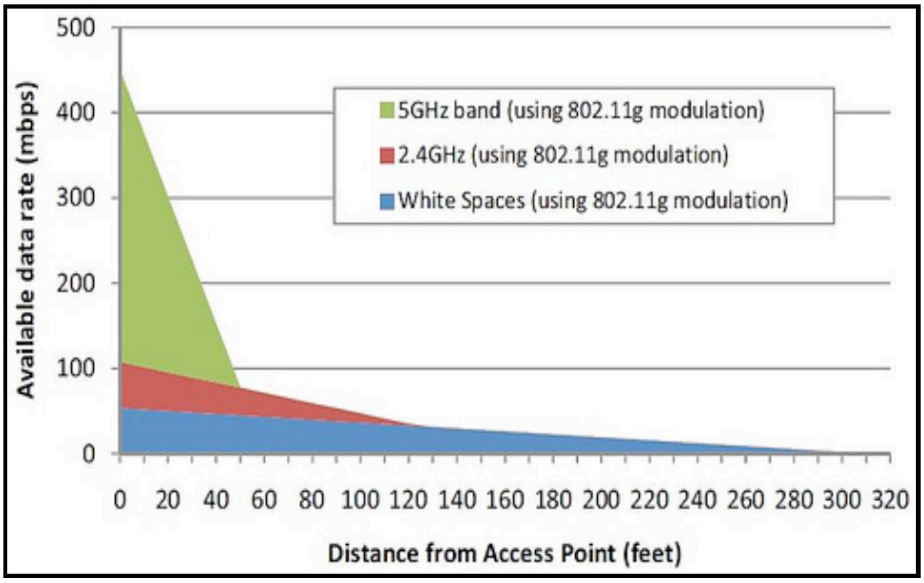


Fig. 1. Comparison of data rates between traditional Wi-Fi and Super Wi-Fi [22]

The dynamic spectrum sharing of TVWS means that WSDs and PUs take turns in using the channel. WSDs can use the channel when it is available. This availability opens the potential for internet access and smart monitoring, especially in rural and remote areas.

Table 1
TVWS Standardization Bodies and Working Groups

Standardization Body	Working Group
IEEE 802	802.22 (Wi-FAR) 802.11af (Wi-Fi) 802.15.4m (ZigBee, Wi-SUN)
IEEE DySPAN Standards Committee	1900.7 1900.4a 1900.4.1

3. Service Contour

Service contours are a geographical limit where the median electric field strength reaches a particular value. That value depends on the type of contour. The two types commonly used are Grade A and Grade B. A sample map of Grade A and Grade B service contours is shown in Figure 2.

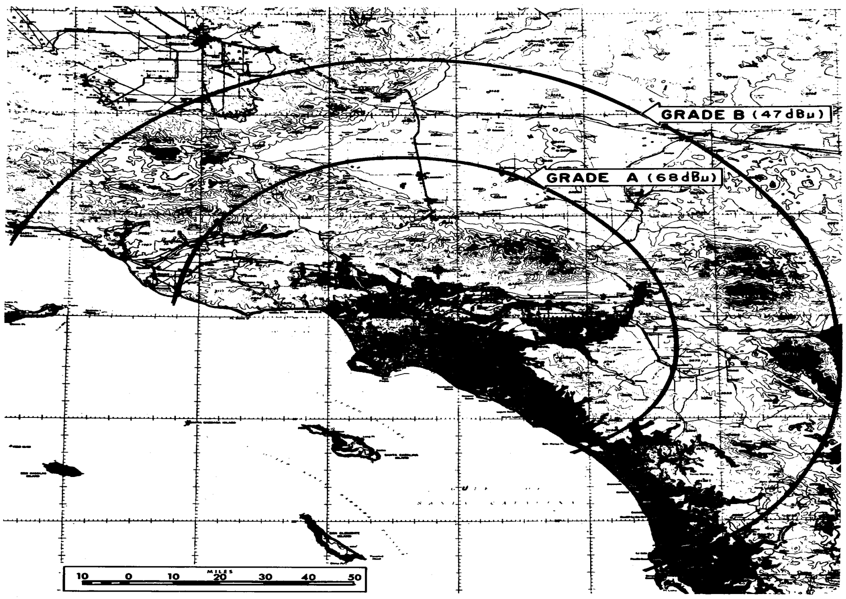


Fig. 2. Grade A and Grade B contours predicted for KNXT, Channel 2, Los Angeles, CA [23]

The grade A value is defined as "ambient median field strength existing 30 feet above ground which is deemed to be sufficiently strong, in the absence of interference from other stations, but with due consideration given to man-made noise typical of urban areas, to provide a picture which the median observer would classify as of 'acceptable' quality, assuming a receiving installation considered to be typical of suburban or not too distant areas." [23]. On the other hand, the Grade B value is defined as "ambient median field strength existing 30 feet above ground which is deemed to be sufficiently strong, in the absence of man-made noise or interference from other stations, to provide a picture which the median observer would classify as of 'acceptable' quality, assuming a receiving installation considered to be typical of outlying or near-fringe areas." [23]. These studies are referred by Bhattarai, Park and Lehr [24]. Simply put, the signal has sufficient strength to provide a picture at least 90% of the time at the best 70% of receiving locations for Grade A and at the best 50% of receiving locations for Grade B. Hence, Grade A service contours are the geographical limits where the medial field strength is at least the Grade A value, while Grade B service contours are the geographical limits where the medial field strength is at least the Grade B value. These service contours are derived from $F(50, 50)$. Table 2 lists the service contours expressed in decibels (dB) above 1 µV/m (dBu).

Table 2
Service Contours defined by the Code of Federal Regulations (CFR) [25]

Channels	Grade A (dBu)	Grade B (dBu)
2-6	68	47
7-13	71	56
14-69	74	64

Service contours are not the limits of actual service. TV stations can be received well outside the contour, where chances of interference-free reception decrease at further distances. Service contours define the areas where interference is protected by the FCC [26].

4. Databases

A database is an organized collection of logically related information or data usually stored in a computer and controlled by a Database Management System (DMS). Databases are accessed electronically from a computer system. They should be accessible, managed, and updated. The database studied in this research is a TVWS geo-location database. It coordinates white space devices to avoid interference with protected primary users as it holds the location and available frequencies. The TVWS geo-location database has several interfaces, such as incumbent, system administrator, regulator, geography, white space devices, operator, other data sources, and other TVWS databases.

In general, a protocol was followed enforcing interference avoidance with primary users. This avoidance is done by designating three zones: Exclusion, Restriction, and Protection [27]. A WSD should be able to communicate with a geo-location database in which their location is being sent, and the geo-location database sends the operating frequency and power level. Pakzad *et al.,* [28] used a database containing primary user information such as the company name, channel, lower and upper frequencies, frequency of transmission, location (descriptive), longitude, latitude, site elevation, structure height, transmission power, and effective radiated power (ERP).

This study gathered updated data for the primary user information. Channel availabilities are surveyed from different websites based on a daily schedule from 00:00H (12:00 AM) to 23:59H (11:59 PM) in 30-minute intervals. Available channels are represented by a value of 1, and unavailable channels are represented by a 0. The study discarded channels with incomplete information. This database information was then packaged and summarized.

5. Vector Autoregression

Vector Autoregression (VAR) models assess the dynamic relationships between variables that interact with one another. It is a multivariate time series model that relates current observations of a variable with the past observations of itself or other variables and is usually used in econometric predictions. VAR models are multiple time series generalizations of autoregressive models applied in forecasting. A univariate autoregression uses a linear model in which the current value of a variable is explained by its own lagged values. A VAR, on the other hand, uses n equations in n variables in which each variable is explained by its own lagged values, plus current and past values of the remaining $n-1$ variables. Furthermore, VAR models have three classifications: reduced-form, recursive, and structural [29]. This study uses reduced-form VAR models in predicting channel availability and coverage. Sometimes, predictor variables are included in the VAR model. These predictors are called exogenous variables. A $\mathrm{VAR}(p)$ model with exogenous variables is often denoted as $\mathrm{VARX}(p)$. The $\mathrm{VARX}(p)$ model has a form shown in Eq. (1).

$$\boldsymbol{y}_t = \boldsymbol{c}_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \mathbf{B}\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \tag{1}$$

In the VAR model, $\boldsymbol{y}_t \in \mathbb{R}^N$ is the response at time t, $\boldsymbol{x}_t \in \mathbb{R}^K$ are the predictors at time t (exogenous data), and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^N$ are the uncorrelated serial errors. The predictor parameters are denoted by the $N \times K$ matrix $\mathbf{B}$, and the autoregressive parameters for each of the lagged values are denoted by the $N \times N$ matrices $\boldsymbol{\Phi}_1$, $\boldsymbol{\Phi}_2$, …, $\boldsymbol{\Phi}_p$.

6. Methodology

Given such sparse and limited information, the study has postulated forecasting models to predict channel availability. The model determines the channel availability, which can be available or unavailable. The model also predicts whether a secondary user at a specific location can broadcast on a particular channel. It does this by consulting the geo-location database, if using such a channel would cause interference with a primary user. This process is summarized in Figure 3. For this study, the white space device has fixed properties. The forecasts from the models are then evaluated.

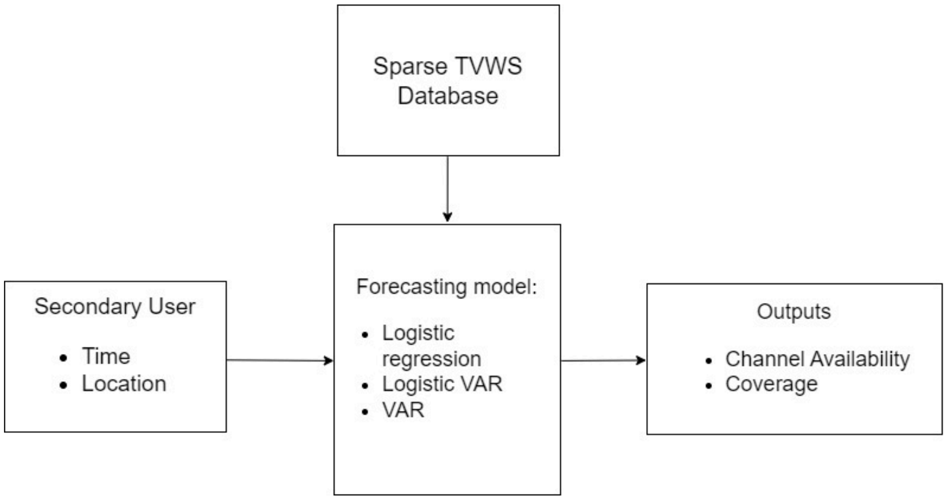


Fig. 3. Diagram of input, process, and output

6.1 Data Gathering and Pre-Processing

The TVWS Geo-location database contains information about the primary users, such as the frequency spectrum (channel), location (latitude and longitude), and transmission. The study explored eight channels, which are 4, 5, 7, 9, 11, 13, 27, and 37. It consists of two tables. The information table holds primary user information that is constant across time, while the availability table contains the channel availability through time. The availability table serves as the temporal data. The information table dictates the channels in the availability table. The channels were examined from 12:00 AM to 11:59 PM for their availability. In addition, the database was updated to the current state of the channels based on TV program guides gathered from online sources and to the best knowledge of the researcher, such as the exclusion of Channel 2 since it was closed. Because the TV program guides determine the channel availability, it is safe to assume that there is no missing data.

The days of the week are categorical and are pre-processed through dummy variables. Dummy variable encoding is like one-hot encoding; the categories are encoded as binary. In dummy variable encoding, the baseline variable is not included. This study's baseline day is Monday, although that depends on the implementing program. In addition, time is expressed in hours past 12 AM (00:00H). For example, 0.5 denotes 12:30 AM or 00:30H. Table 3 summarizes the features of the information table, while Table 4 summarizes the features of the availability table.

Table 3
Features Of TVWS Database Information Table

Feature	Description
CHANNEL	Channel number (CH + number)
COMPANY_NAME	Name of the company
DIGITAL	Indicator if it is using digital transmission (0 or 1)
CALLSIGN	Company callsign
LATITUDE	Location of the user; latitude in degrees
LONGITUDE	Location of the user; longitude in degrees
LOWER_BAND	Lower limit of the frequency (MHz)
UPPER_BAND	Upper limit of the frequency (MHz)
TX_FREQ	Transmission frequency (MHz)
ERP	Effective Radiated Power (kW)
TX_PWR	Transmission Power (kW)
SITE_ELEV	Site Elevation (m)
STRUCTURE_HEIGHT	Height of the antenna structure (m)

Table 4
Features Of the TVWS Database Availability Table

Feature	Description
DAY	Day of the week
TIME	Time in 24-hour format
CHX	Availability of channel X; denoted as 1 when available and 0 when unavailable

Moreover, the secondary user in a specific location must know if it can use a particular channel without interfering with the primary users. The spatial broadcast data is generated using service contours computed by the Longley-Rice model. The service contours of the primary and secondary users are approximated using radials drawn from the transmitter location. Eight is the typical number of radials used. An intersection of the primary user and secondary user contour indicates an interference, and the secondary user cannot broadcast on that channel. A value of 1 denotes if a secondary user can broadcast; otherwise, 0. Secondary user locations are sampled for generating spatial broadcast data. One hundred locations were selected for spatial broadcast availability, which was then incorporated into the database along with temporal availability.

6.2 Forecasting Models

A dynamic and sparse database makes forecasting difficult when relying solely on patterns. Hence, the study uses models to predict whether a channel is available or unavailable and whether the secondary user can use a specific channel in a particular location. The sparsity of the data is measured prior to model training. Sparsity could refer to either having a small amount of information or a small number of coefficients containing a large proportion of energy [3]. As sparsity is an abstract concept, there are different measures of sparsity, such as ℓ^0 norm, Hoyer index, and Gini index. Spatiotemporal logistic models are explored to predict channel availability and coverage. A logistic model is a generalized linear model used when the response uses discrete values [3]. Logistic regression is done using maximum likelihood. The model relates the log-odds of the probability of desirable outcome to a linear combination of inputs, as indicated in Eq. (2), where $\boldsymbol{X} \in \mathbb{R}^{N\times(K+1)}$ is the predictor matrix with K predictors and $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ are the coefficients.

$$\mathrm{logit}(\pi_i) = (\boldsymbol{X}\boldsymbol{\beta})_i \tag{2}$$

The inputs of the spatiotemporal logistic model are the day of the week, time of day, latitude, and longitude, which also serve as regression parameters. Moreover, a penalized spatiotemporal logistic model that incorporated lasso (ℓ^1 norm) and elastic net (ℓ^1 and ℓ^2 norms) is formulated.

In addition, the study formulated the spatiotemporal logistic VAR, a combination of the logistic and VAR models. This is done by replacing the response variable with the log odds of the probability. The logistic VAR model is expressed in Eq. (3), where $\boldsymbol{y}_t \in \mathbb{R}^N$ is the response data (channel availability data), $\boldsymbol{\Phi}_p \in \mathbb{R}^{N\times N}$ are the autoregressive coefficients at p previous time points, $\boldsymbol{x}_t \in \mathbb{R}^K$ is the exogenous data, $\mathbf{B} \in \mathbb{R}^{N\times K}$ are the coefficients for the exogenous data, $\pi_{t,i}$ represents the probability of success (availability of a channel), and $\varepsilon_{t,i}$ are the uncorrelated serial errors. When classifying, probabilities at or above 0.5 are deemed available, while those below 0.5 are deemed unavailable.

$$\operatorname{logit}\left(\pi_{t,i}\right) = c_{0,i} + (\boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1})_i + \cdots + \left(\boldsymbol{\Phi}_p \boldsymbol{y}_{t-p}\right)_i + (\mathbf{B}\boldsymbol{x}_t)_i + \varepsilon_{t,i} \tag{3}$$

Furthermore, a spatiotemporal linear VAR model (linear VAR) was explored for predicting channel availability and spatial broadcast coverage. The data set is transformed to suit the linear VAR model. Responses with a 1 are replaced with the input (day, time, and location). Thus, the linear VAR consists of nine components: six dummy-variable encoded days, time, latitude, and longitude. The study opted for thresholding the predicted value for classification. A thresholding coefficient η, where $0 \leq \eta \leq 1$, is defined to determine the threshold of classification y_{thr}.

Values below y_{thr} are classified as 0. Otherwise, it is classified as 1. In Eq. (4), the minimum non-zero response value determines the classification threshold. The forecasting models were implemented by computer programs using MATLAB. Comma Separated Values (CSV) files were used to host the databases.

$$y_{\text{thr},t} = \begin{cases} \eta y_t, & y_t > 0 \\ \eta y_{\text{MIN}}, & y_t = 0 \end{cases} \tag{4}$$

6.3 Evaluation

A portion of the data set is used to evaluate the models. A typical partition of the data sets is 80% for training and 20% for testing. Data points for training and testing are randomly selected. The metrics used to evaluate the forecast are accuracy, precision, and recall.

A hit or true positive (TP) is defined when a channel was forecasted as available and was observed as available according to the test data. A miss or false negative (FN) is when a channel was predicted as unavailable but was observed as available. A false alarm or false positive (FP) is when a channel is predicted as available when observed as unavailable. Lastly, a correct rejection or true negative (TN) is when a channel is predicted as unavailable and is observed as unavailable. Correct predictions include hits and correct rejections.

Accuracy is defined as the number of correct rejections over the total samples in the test data as mathematically defined in Eq. (5). Both temporal and spatiotemporal models are evaluated using this metric.

$$\text{Accuracy} = \frac{\text{TP+FP}}{\text{TP+FP+FN+TN}} \tag{5}$$

Precision (Eq. (6)) is the ratio of true positives to true positives and false positives. In other words, the precision tells how many channels are available compared to all channels predicted to be available.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{6}$$

Recall (Eq. (7)) is the ratio of true positives to true positives and false negatives. In other words, the recall tells how many available channels are predicted as available.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

The F1 score (Eq. (8)) combines precision and recall through the harmonic mean. The F1 score is preferred for imbalanced data. If both precision and recall are high, the F1 score is high.

$$\text{F1} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{8}$$

7. Results and Discussion

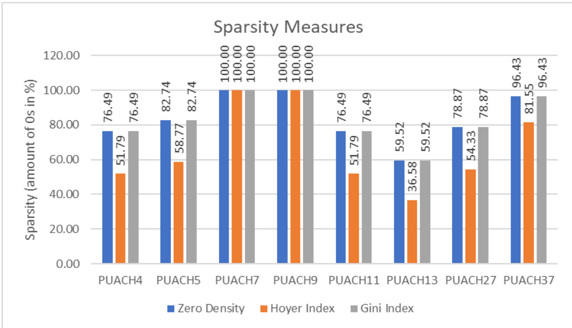
7.1 Forecasting Models

This study formulated spatiotemporal forecasting models to predict channel availability and coverage. These models predict the availability of a channel and whether the secondary user at a specific time and location can broadcast on a channel. The data is binary in nature, as channel availability is either available or unavailable. Hence, the study formulated logistic models. Lasso and elastic net are added to the logistic models to account for sparsity. Because prediction was done on both spatial and temporal aspects, both linear and logistic VAR models were formulated. The logistic VAR without exogenous data was considered, but the researcher found that including exogenous data (logistic VARX) had better training and performance. Logistic models and logistic VAR models work on all data. Linear VAR models, however, do not work on channels with continuous broadcasts and few available slots. In other words, the linear VAR model does not forecast on channels 7, 9, and 37. Channels 7 and 9 continuously broadcast, and channel 37 has high sparsity.

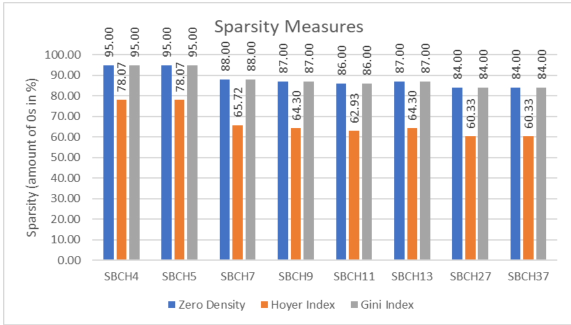
7.2 Forecasting Evaluation Results

Before discussing results regarding spatiotemporal predictions, there are notations to discuss first. PUACHX denotes temporal channel availability, and SBCHX denotes, where X is the channel number. The sparsities of the data are measured using the density of zeros (ℓ^0 norm-based), Hoyer Index, and Gini Index.

The sparsity measures for the temporal and spatial data are shown in Figure 4. For temporal availability, channels that broadcast all the time have a 100% sparsity, as the data are composed of zeros, and channels with fewer off-air times have higher sparsity measures. For spatial availability, fewer available locations meant a higher sparsity measure. The Gini Index is closer to the density of zeros than the Hoyer index.



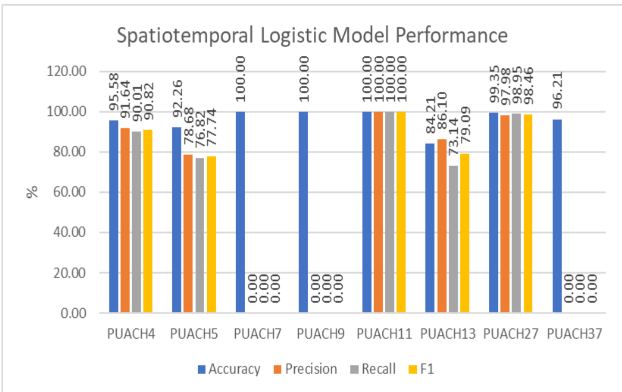
(a)



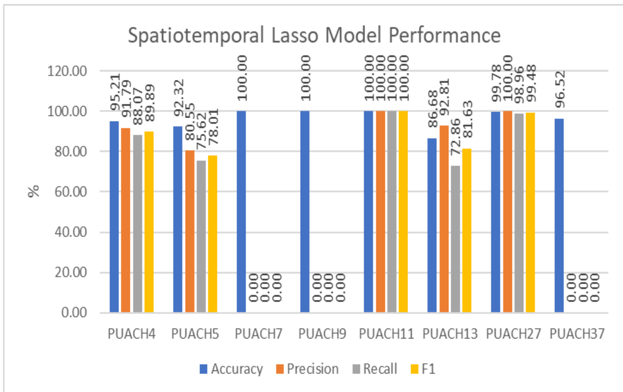
(b)

Fig. 4. Sparsity Measures of Temporal Data (a) and Spatial Data (b)

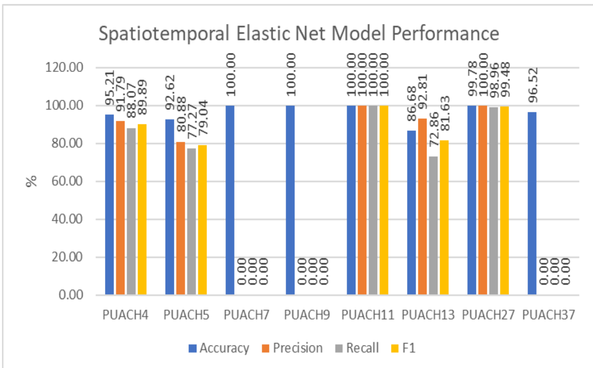
The study first explored unregularized and regularized spatiotemporal models, whose performances were evaluated using accuracy, precision, recall, and F1 score. The parameters for the lasso and elastic net models were automatically generated, and the optimal one was selected by cross-validation. The implementing program performs this process automatically. The researcher settled on using the penalty parameter one standard error (1SE) away from the optimal. Figure 5 and Figure 6 compare the accuracy, precision, recall, and F1 score of the spatiotemporal logistic models in temporal and spatial aspects, respectively. The unregularized and regularized spatiotemporal logistic models have an accuracy of at least 84% for both the temporal and spatial predictions.



(a)



(b)



(c)

Fig. 5. Performances of Spatiotemporal Unregularized (a) and Regularized Logistic Models (b and c) in Temporal Predictions

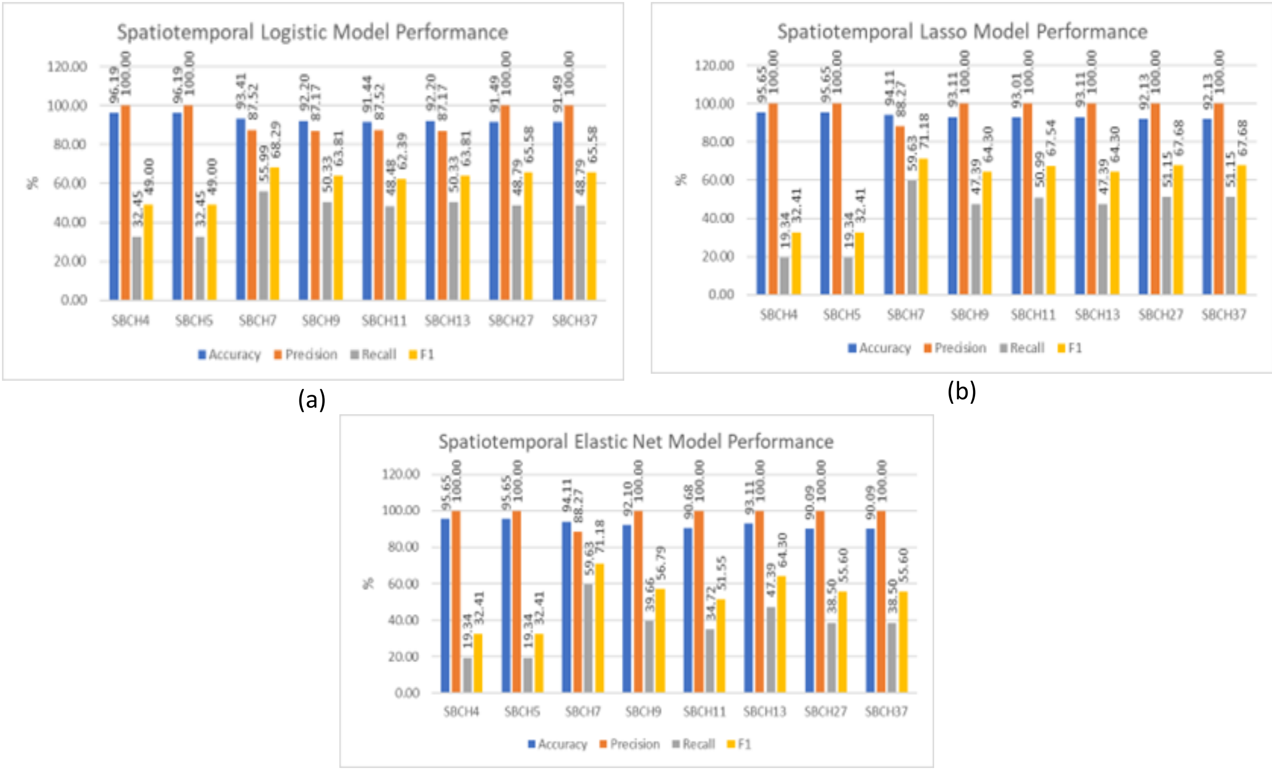


Fig. 6. Performances of Spatiotemporal Unregularized (a) and Regularized Logistic Models (b and c) in Spatial Predictions

The study then explored spatiotemporal linear VAR models because both spatial and temporal predictions were made simultaneously. The models' performances were evaluated using the same metrics. The study considered both linear VAR with and without exogenous variables (linear VARX) and found that adding exogenous variables makes better predictions. The predictions of the linear VARX model were graphed together with test data, resulting in the one shown in Figure 7.

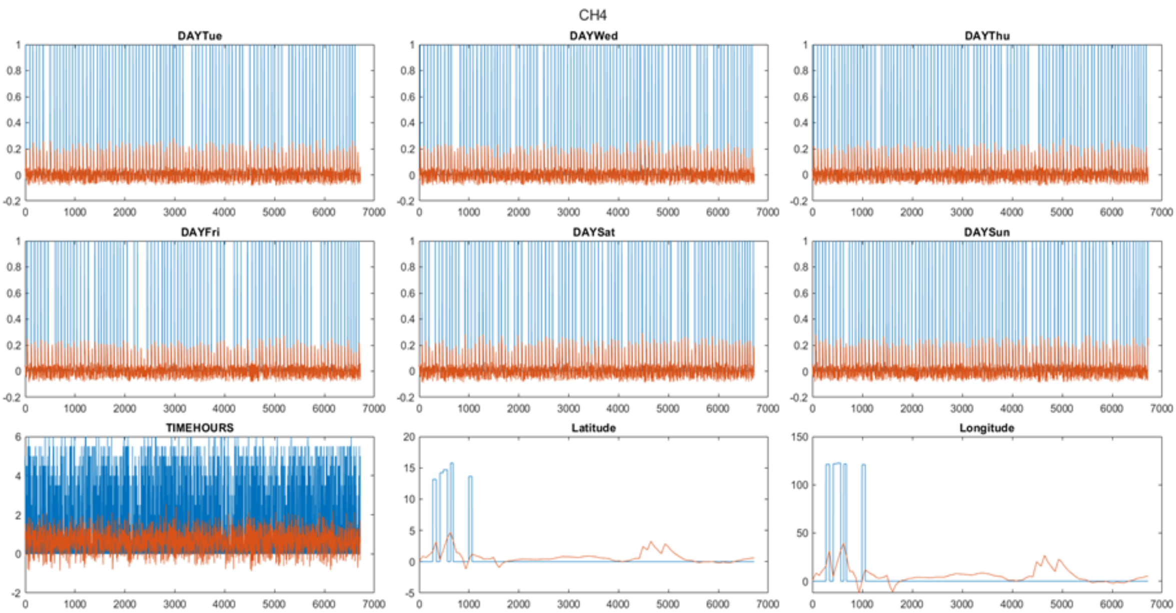


Fig. 7. Linear VARX(4) predictions (orange) and test data (blue)

One and four lags were examined. The linear VARX model predicts well except for time, as shown in Figure 8. The predictions could be a result of the linear VARX model not being appropriate for the given nature of data.

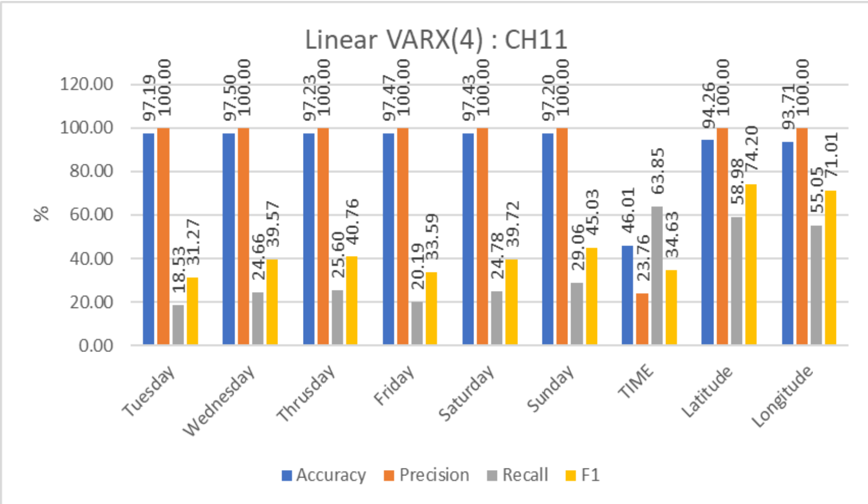
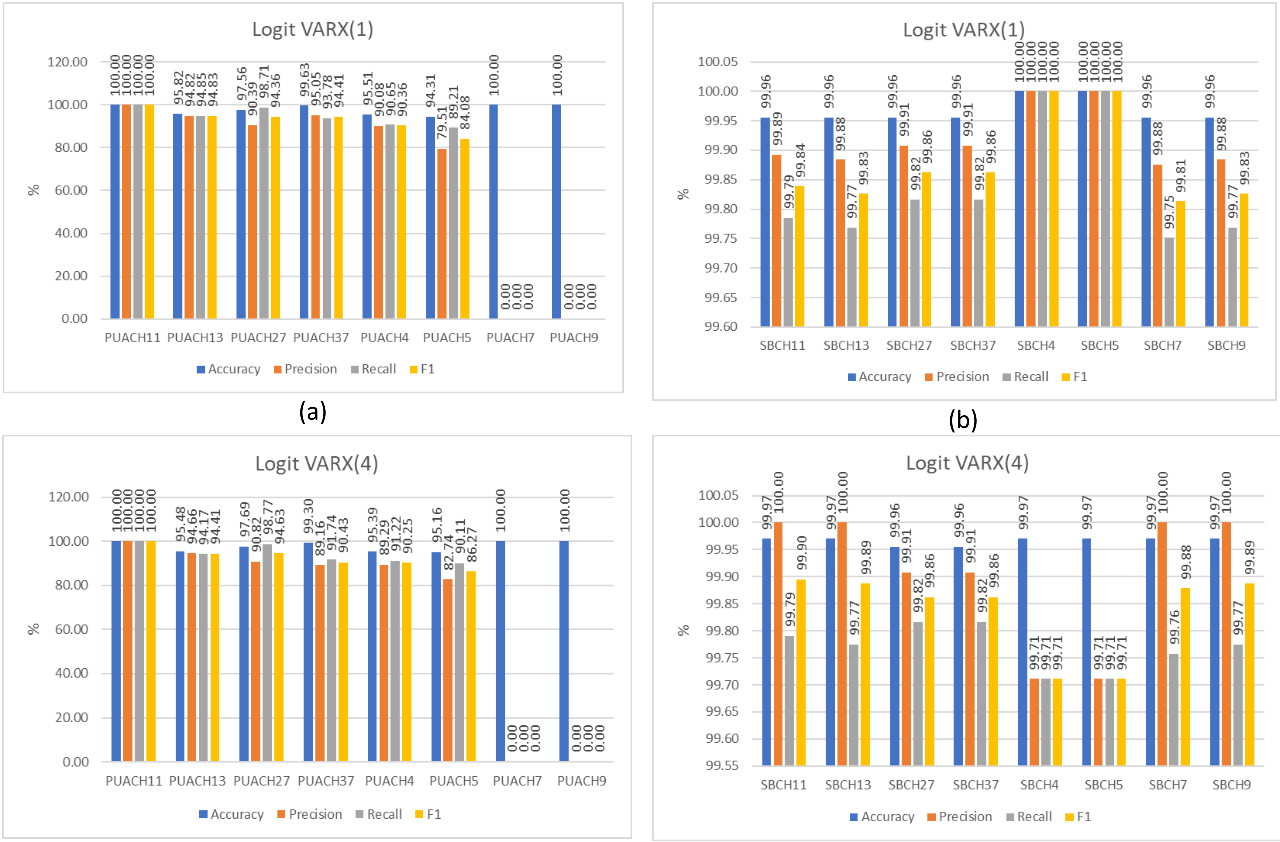


Fig. 8. Performance of the Linear VARX(4) model for Channel 11 (blue)

Hence, the study formulated spatiotemporal logistic VAR models, which were then implemented. One and four lags were considered, and models with and without exogenous data were explored. However, the model with exogenous data (logistic VARX) had better training with improved accuracy, precision, recall, and F1 score. The performances of the logistic VARX models are summarized in Figure 9. The logistic VARX models improved the spatial forecasting accuracy from at least 84% to at least 99% compared to spatiotemporal logistic and linear VARX models. The precision, recall, and F1 score of the spatiotemporal logistic VARX models are all at least 99% for spatial availability. Temporal prediction accuracy for the spatiotemporal logistic VARX model improved from at least 84% to at least 94%, and the F1 score is at least 84%.



(a) (b) (c) (d)

Fig. 9. Performances of the Logistic VARX Models

Comparisons were made between the spatiotemporal logistic models, spatiotemporal logistic VARX models, and supervised learning models. The study considered shallow and deep neural networks for supervised learning models. The shallow neural network has one hidden layer, while the deep neural network has three hidden layers. These comparisons are summarized in Table 5, Table 6, Table 7 and Table 8.

Table 5
Comparison of Accuracies for the Forecasting Models

Model		Accuracy (%) CH4	CH5	CH7	CH9	CH11	CH13	CH27	CH37
Spatiotemporal Logit	Temporal	95.58	92.26	100.00	100.00	100.00	84.21	99.35	96.21
	Spatial	96.19	96.19	93.41	92.20	91.44	92.20	91.49	91.49
Spatiotemporal Lasso	Temporal	95.21	92.32	100.00	100.00	100.00	86.68	99.78	96.52
	Spatial	95.65	95.65	94.11	93.11	93.01	93.11	92.13	92.13
Spatiotemporal Elastic Net	Temporal	95.21	92.62	100.00	100.00	100.00	86.68	99.78	96.52
	Spatial	95.65	95.65	94.11	92.10	90.68	93.11	90.09	90.09
Shallow NN	Temporal	75.71	99.17	100.00	100.00	76.15	70.10	99.35	100.00
	Spatial	100.00	94.36	87.32	97.69	85.36	86.34	83.38	83.38
Deep NN	Temporal	75.71	99.33	100.00	100.00	76.15	59.17	78.82	96.21
	Spatial	98.14	98.05	90.37	92.10	97.99	91.29	83.38	83.38

Table 6
Comparison of Precisions for the Forecasting Models

Model		Precision (%) CH4	CH5	CH7	CH9	CH11	CH13	CH27	CH37
Spatiotemporal	Temporal	91.64	78.68	0.00	0.00	100.00	86.10	97.98	0.00
Logit	Spatial	100.00	100.00	87.52	87.17	87.52	87.17	100.00	100.00
Spatiotemporal	Temporal	91.79	80.55	0.00	0.00	100.00	92.81	100.00	0.00
Lasso	Spatial	100.00	100.00	88.27	100.00	100.00	100.00	100.00	100.00
Spatiotemporal	Temporal	91.79	80.88	0.00	0.00	100.00	92.81	100.00	0.00
Elastic Net	Spatial	100.00	100.00	88.27	100.00	100.00	100.00	100.00	100.00
Shallow NN	Temporal	0.00	97.79	0.00	0.00	0.00	99.73	97.98	100.00
	Spatial	100.00	0.00	0.00	93.40	0.00	0.00	0.00	0.00
Deep NN	Temporal	0.00	98.06	0.00	0.00	0.00	0.00	0.00	0.00
	Spatial	82.23	100.00	100.00	85.25	95.40	76.73	0.00	0.00

Table 7
Comparison of Recalls for the Forecasting Models

Model		Recall (%) CH4	CH5	CH7	CH9	CH11	CH13	CH27	CH37
Spatiotemporal Logit	Temporal	90.01	76.82	0.00	0.00	100.00	73.14	98.95	0.00
	Spatial	32.45	32.45	55.99	50.33	48.48	50.33	48.79	48.79
Spatiotemporal Lasso	Temporal	88.07	75.62	0.00	0.00	100.00	72.86	98.96	0.00
	Spatial	19.34	19.34	59.63	47.39	50.99	47.39	51.15	51.15
Spatiotemporal Elastic Net	Temporal	88.07	77.27	0.00	0.00	100.00	72.86	98.96	0.00
	Spatial	19.34	19.34	59.63	39.66	34.72	47.39	38.50	38.50
Shallow NN	Temporal	0.00	97.46	0.00	0.00	0.00	26.86	98.95	100.00
	Spatial	100.00	0.00	0.00	89.43	0.00	0.00	0.00	0.00
Deep NN	Temporal	0.00	98.14	0.00	0.00	0.00	0.00	0.00	0.00
	Spatial	85.49	65.44	24.06	50.98	90.65	52.07	0.00	0.00

Table 8
Comparison of F1 Scores for the Forecasting Models

Model		F1 Score (%) CH4	CH5	CH7	CH9	CH11	CH13	CH27	CH37
Spatiotemporal Logit	Temporal	90.82	77.74	0.00	0.00	100.00	79.09	98.46	0.00
	Spatial	49.00	49.00	68.29	63.81	62.39	63.81	65.58	65.58
Spatiotemporal Lasso	Temporal	89.89	78.01	0.00	0.00	100.00	81.63	99.48	0.00
	Spatial	89.89	78.01	0.00	0.00	100.00	81.63	99.48	0.00
Spatiotemporal Elastic Net	Temporal	89.89	79.04	0.00	0.00	100.00	81.63	99.48	0.00
	Spatial	32.41	32.41	71.18	56.79	51.55	64.30	55.60	55.60
Shallow NN	Temporal	0.00	97.63	0.00	0.00	0.00	42.32	98.46	100.00
	Spatial	100.00	0.00	0.00	91.37	0.00	0.00	0.00	0.00
Deep NN	Temporal	0.00	98.10	0.00	0.00	0.00	0.00	0.00	0.00
	Spatial	83.83	79.11	38.79	63.80	92.97	62.04	0.00	0.00

8. Conclusion

The study used sparse geo-location databases to forecast wireless coverage and frequency by formulating spatiotemporal forecasting models. The spatiotemporal models include unregularized and regularized logistic models, linear VARX models, and logistic VARX models. The models' predictions were at least 70% accurate, except for only one under linear VARX. The spatiotemporal unregularized and regularized logistic models, as well as the spatiotemporal logistic VARX, performed better than the shallow NN and deep NN. The spatiotemporal logistic VARX models had the best accuracy, prediction, recall, and F1 score among all the spatiotemporal models. This study considered white space devices with fixed properties like transmission power. This study recommends having secondary users with varying properties and using different thresholding, data transform, or a different model to account for the low accuracy of the spatiotemporal linear VAR. The work could be extended to cover primary users beyond the Greater Manila Area.

Acknowledgement
The authors would like to thank Philippine Department of Science and Technology Engineering for Research and Technology (DOST-ERDT) Program and De La Salle University (DLSU) for funding this research.

References

[1] Zhang, Wenjie, Jingmin Yang, Guanglin Zhang, Liwei Yang, and Chai Kiat Yeo. "TV white space and its applications in future wireless networks and communications: A survey." *IET Communications* 12, no. 20 (2018): 2521-2532. https://doi.org/10.1049/iet-com.2018.5009

[2] Shawel, Bethelhem S., Dereje H. Woldegebreal, and Sofie Pollin. "Convolutional LSTM-based long-term spectrum prediction for dynamic spectrum access." In *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1-5. IEEE, 2019. https://doi.org/10.23919/EUSIPCO.2019.8902956

[3] Ocampo, Vladimir Iı Christian, and Lawrence Materum. "Wireless Channel Availability Forecasting with a Sparse Geolocation Spectrum Database by Penalty-Regularization Logistic Models." *The Eurasia Proceedings of Science Technology Engineering and Mathematics* 21 (2022): 39-45. https://doi.org/10.55549/epstem.1224555

[4] Martin, John Hugh, Laurence Sean Dooley, and Kam Cheung Patrick Wong. "New dynamic spectrum access algorithm for TV white space cognitive radio networks." *IET Communications* 10, no. 18 (2016): 2591-2597. https://doi.org/10.1049/iet-com.2016.0213

[5] Yagoub, Sami Abdelrahman Musa, Gregorius Eldwin Pradipta, and Ebrahim Mohammed Yahya. "Prediction of bubble point pressure for Sudan crude oil using Artificial Neural Network (ANN) technique." *Progress in Energy and Environment* (2021): 31-39.

[6] Kanaani, Osamah Othman, Sami Abdelrahman Musa Yagoub, Shabir Habib, Akmal Aulia, and Bonavian Hasiholan. "Prediction of gas coning in hydrocarbon reservoir using tNavigator." *Progress in Energy and Environment* (2021): 1-22. https://doi.org/10.37934/progee.18.1.122

[7] Rizaman, Mohd Syafiq Azfar, Ahmad Sufian Abdullah, and Aliff Farhan Mohd Yamin. "Determination of the Anand Parameters for SAC405 Solders Through the Use of Stress-Strain Data." *Journal of Advanced Research in Applied Mechanics* 113, no. 1 (2024): 162-175. https://doi.org/10.37934/aram.113.1.162175

[8] Arip, Afifuddin Arif Shihabuddin, Norazlianie Sazali, Kumaran Kadirgama, Ahmad Shahir Jamaludin, Faiz Mohd Turan, and Norhaida Ab Razak. "Object Detection for Safety Attire Using YOLO (You Only Look Once)." *Journal of Advanced Research in Applied Mechanics* 113, no. 1 (2024): 37-51. https://doi.org/10.37934/aram.113.1.3751

[9] Hou, Bingchang, Dong Wang, Tangbin Xia, Lifeng Xi, Zhike Peng, and Kwok-Leung Tsui. "Generalized Gini indices: Complementary sparsity measures to Box-Cox sparsity measures for machine condition monitoring." *Mechanical Systems and Signal Processing* 169 (2022): 108751. https://doi.org/10.1016/j.ymssp.2021.108751

[10] Benarabi, Tarek, Mourad Adnane, and Moufid Mansour. "Energy and sparse coding coefficients as sufficient measures for VEBs classification." *Biomedical Signal Processing and Control* 67 (2021): 102493. https://doi.org/10.1016/j.bspc.2021.102493

[11] Goswami, Swati, C. A. Murthy, and Asit K. Das. "Sparsity measure of a network graph: Gini index." *Information Sciences* 462 (2018): 16-39. https://doi.org/10.1016/j.ins.2018.05.044

[12] Guerra-Montenegro, Juan, Javier Sanchez-Medina, Ibai Laña, David Sanchez-Rodriguez, Itziar Alonso-Gonzalez, and Javier Del Ser. "Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges." *Applied Soft Computing* 102 (2021): 107082. https://doi.org/10.1016/j.asoc.2021.107082

[13] Ardia, David, Keven Bluteau, and Kris Boudt. "Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values." *International Journal of Forecasting* 35, no. 4 (2019): 1370-1386. https://doi.org/10.1016/j.ijforecast.2018.10.010

[14] Flaxman, Seth, Michael Chirico, Pau Pereira, and Charles Loeffler. "Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the NIJ “Real-Time Crime Forecasting Challenge”." (2019): 2564-2585. https://doi.org/10.1214/19-AOAS1284

[15] Zheng, Zengwei, Lifei Shi, Lin Sun, and Junjie Du. "Short-term traffic flow prediction based on sparse regression and spatio-temporal data fusion." *IEEE Access* 8 (2020): 142111-142119. https://doi.org/10.1109/ACCESS.2020.3013010

[16] Rajab, Khairan, Firuz Kamalov, and Aswani Kumar Cherukuri. "Forecasting COVID-19: vector autoregression-based model." *Arabian journal for science and engineering* (2022): 1-10. https://doi.org/10.1007/s13369-021-06526-2

[17] Katris, Christos. "Unemployment and Covid-19 impact in Greece: A vector autoregression (VAR) data analysis." *Engineering Proceedings* 5, no. 1 (2021): 41. https://doi.org/10.3390/engproc2021005041

[18] Ante, Marc Gelian, James Agustin Molina, Emmanuel Trinidad, and Lawrence Materum. "A survey and comparison of TV white space implementations in Japan, the Philippines, Singapore, the United Kingdom, and the United States." *International Journal of Advanced Technology and Engineering Exploration* 8, no. 80 (2021): 780. https://doi.org/10.19101/IJATEE.2021.874106

[19] Islam, Md Zobaer, John F. O'Hara, Dylan Shadoan, Mostafa Ibrahim, and Sabit Ekin. "TV white space based wireless broadband internet connectivity: A case study with implementation details and performance analysis." *IEEE Open Journal of the Communications Society* 2 (2021): 2449-2462. https://doi.org/10.1109/OJCOMS.2021.3123939

[20] Whizpace. "Whizpace." *Whizpace*. https://www.whizpace.com

[21] Carlson Wireless Technologies. "Carlson Wireless Technologies – Leaders in TV White Space." *Carlson Wireless Technologies*. https://carlsonwireless.com

[22] Montejo, Antonio, Alberto Bañacia, Hirokazu Sawada, Kentaro Ishizu, Takeshi Matsumura, Kazuo Ibuka, and Fumihide Kojima. "Performance Evaluation of an IEEE 802.11 af Prototype in a Suburban Environment." In *2018 Asia-Pacific Microwave Conference (APMC)*, pp. 741-743. IEEE, 2018. https://doi.org/10.23919/APMC.2018.8617422

[23] O'Connor, Robert A. "Understanding television's grade A and grade B service contours." *IEEE transactions on broadcasting* 47, no. 3 (2001): 309-314. https://doi.org/10.1109/11.969381

[24] Bhattarai, Sudeep, Jung-Min Park, and William Lehr. "Dynamic exclusion zones for protecting primary users in database-driven spectrum sharing." *IEEE/ACM Transactions on Networking* 28, no. 4 (2020): 1506-1519. https://doi.org/10.1109/TNET.2020.2986410

[25] ECFR. "47 CFR 73.683 -- Field Strength Contours and Presumptive Determination of Field Strength at Individual Locations." https://www.ecfr.gov/current/title-47/chapter-I/subchapter-C/part-73/subpart-E/section-73.683

[26] Federal Communications Commission. "General Information about FM and TV Service Contour Maps and Map Data." Federal Communications Commission. (2015). https://www.fcc.gov/media/radio/general-info-fm-tv-maps-data

[27] Stirling, Andrew. "TV white space developments in the UK." In *TV White Space Communications and Networks*, pp. 1-25. Woodhead Publishing, 2018. https://doi.org/10.1016/B978-0-08-100611-5.00002-3

[28] Pakzad, Armie E., Raine Mattheus Manuel, Jerrick Spencer Uy, Xavier Francis Asuncion, Joshua Vincent Ligayo, and Lawrence Materum. "Reinforcement Learning-Based Television White Space Database." *Baghdad Science Journal* 18, no. 2 (Suppl.) (2021): 0947-0947. https://doi.org/10.21123/bsj.2021.18.2(Suppl.).0947

[29] Cruz, Rafaela C., Pedro Reis Costa, Susana Vinga, Ludwig Krippahl, and Marta B. Lopes. "A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination." *Journal of Marine Science and Engineering* 9, no. 3 (2021): 283. https://doi.org/10.3390/jmse9030283