# A COVID-19 forecasting system using adaptive neuro-fuzzy inference

Kim Tien Ly [1]

*School of Computer Science, University of Nottingham, Nottingham, United Kingdom*

## ABSTRACT

This article proposes an Adaptive Neuro-Fuzzy Inference System (ANFIS) to forecast the number of COVID-19 cases in the United Kingdom. With the combination of artificial neural network and fuzzy logic structure, the model is trained based on collected data. The study examines various factors of ANFIS to come up with an effective time series prediction model. The result indicates that Spain and Italy data can strengthen the predictive power of COVID-19 cases in the UK. It is suggested that the policymakers should adopt Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict contagion effect during the COVID-19 pandemic.

## 1. Introduction

On March 11, 2020, the World Health Organization (WHO) announced that the novel coronavirus (COVID-19) could be characterized as a pandemic. The virus then caused a global crisis that affects not only human health and well-being but also the world economy. The COVID-19 crisis has caused high volatility in the global financial markets. According to The Guardian (2020), the main UK index, FTSE100, suffered the worst quarter since 1987, recording shocking losses of share values of 24.8% due to the outbreak of COVID-19. The finance literature during the COVID-19 pandemic era has been growing rapidly, such as stock market volatility (Li et al., 2020), financial contagion between China and G7 countries (Akhtaruzzaman et al., 2020), economic uncertainty (Choi, 2020), investor behavior (Arias, 2020, Ortmann et al., 2020), systemic risk (Rizwan et al., 2020).

Since the situation is getting out of control, the need to predict the number of cases, deaths, or recovered patients became extremely important to come up with an appropriate strategy to respond in time. However, to the best of knowledge, no studies have examined financial contagion indirectly driven by the number of cases in the UK. Therefore, this article aims to create a COVID-19 forecasting system to understand the condition, control the spread, and manage the situation in time.

This study proposes a novel time-series forecasting system to predict the future number of COVID-19 cases in the United Kingdom. The system estimates the number of cases of the next day using past data. Our work illustrates all technical factors of an Adaptive Neuro-Fuzzy Inference System (ANFIS) system as well as examines data for training a powerful model. The paper complements the research on UK stock market volatility and contagion effect. Therefore, it could help the regulators and policymakers to build their own predictive model to cope with the financial contagion and stabilize the financial system.

The rest of the paper is organized as follows. Section 2 summarizes related work. Section 3 presents data and methodology. Section

---

4 provides a description of Alternative ANFIS models and settings, while Section 5 provides the findings. Section 6 concludes.

## 2. Related work

Time series prediction refers to the concept of estimating future states based on current or past data, which have various applications in finance (Kim, 2003), weather (Taylor et al. 2009), environmental (Leone, 1987) forecasting, etc. One common approach is to use support vector machines (SVM). In 2009, Sapankevych & Sankar reported that SVM suits non-linear time series prediction, however, challenges have been placed in selecting kernel function, parameters or optimization techniques during implementation. Especially, tremendous noise and complex dimensionality are inherent in stock market data (Kim 2003). Another popular scheme is using neural networks (Frank et al., 2001). This approach generally performs well in long term forecasting after sufficient training.

Finance market risks have increased around the world due to the global outbreak of COVID-19 (Zhang et al., 2020). A sharp increase in systemic risk was realized during the COVID-19 period due to the interconnectedness among financial institutions (Rizwan et al., 2020). Akhtaruzzaman et al. (2020) find that the spillovers between Chinese and G7 stock returns increased significantly during the COVID-19 crisis, indicating that there is a similar pattern between financial contagion and virus contagion. Therefore, the role of an appropriate model in forecasting contagion effect is an urgent need during the global pandemic of COVID-19.

Fuzzy modelling refers to the method of manipulating values with vagueness. Fuzzy inference system (FIS) implements fuzzy modelling on an input-output system with a set of if-then rules, which has advantages over other traditional mathematical approaches due to its robustness to environmental changes. However, it is argued that there is no standard method to define fuzzy rules and tune membership functions (MFs). In 1992, Jang introduced an adaptive network-based fuzzy inference system (ANFIS) that integrates artificial neural network (ANN) to train FIS based on data. By combining neural networks and fuzzy logic, ANFIS has proven its optimal ability to deal with uncertainty and has been widely applied on forecasting systems. Alasha'ary et al. (2009) built a neuro-fuzzy model to estimate room temperatures in different residential buildings, which showed the reliability of prediction with a satisfying level of accuracy. In 2007, Singh, Sinha, & Singh proved that ANFIS has the best prediction accuracy among other neural networks. In 2008, Ying & Pan introduced an ANFIS to forecast electric loads, which proved to generate better results than other models, including regression, ANN, SVM, and hybrid ellipsoidal fuzzy system.

## 3. Data and methodology

### 3.1. Data

The training data is downloaded from the Github of the Center for Systems Science and Engineering (CSSE) of Johns Hopkins University (Github of The Center for Systems Science and Engineering (CSSE), 2020). The sample covers a period between 21 March 2020 and 16 May 2020.[2]

To investigate the financial contagion due to COVID-19 outbreak, this article attempts to analyze the trends of different countries for better training data, thus, the data on the number of cases from 10 countries is obtained, including Belgium, China, France, Germany, Iran, Italy, Spain, Switzerland, United Kingdom, United States. Arias (2020) find that COVID-19 pandemic increased herding behavior across the UK, France, Germany, Italy, and Spain. Zhang et al. (2020) highlight that Iran experienced rapid growth of COVID-19 cases from zero cases to over 1,000 only within 12 days. Just and Echaust, (2020) emphasize a simultaneous course of events in terms of a high correlation of the number of deaths in Italy and Switzerland as well as an increasing implied correlation as reported COVID-19 deaths of Spain and Belgium.

### 3.2. Evaluation criteria

To compare FIS models, this study analyzes the overall performance using Unscaled Mean Bounded Relative Absolute Error (UMBRAE). UMBRAE (Chen et al. 2017) is a developed measurement that resolves common challenges but still keeps the best features of other existing measures, which is reported to suit the time series forecasting application. Accordingly, when UMBRAE < 1, the proposed model performs (1-UMBRAE)*100% better than the benchmark method. On the other hand, when UMBRAE > 1, the system is roughly (UMBRAE-1)*100% worse than the benchmark method. This performance evaluation can be done via a handy accuracy function provided in the FuzzyR package. In our approach, the naive method is used as the benchmark for UMBRAE. The performance evaluation will focus on choosing the lowest validation error.

## 4. Description of Alternative ANFIS Models/Settings

### 4.1. Different inputs

Since the study aims to predict the number of future cases in the UK, the first trial (ANFIS1) examines the data of UK cases collected over five months from January to May 2020. We split the data into 86% for training, 7% for validation, and 7% for testing. This set of

---

[2] The UK ordered the first lockdown in March 2020.
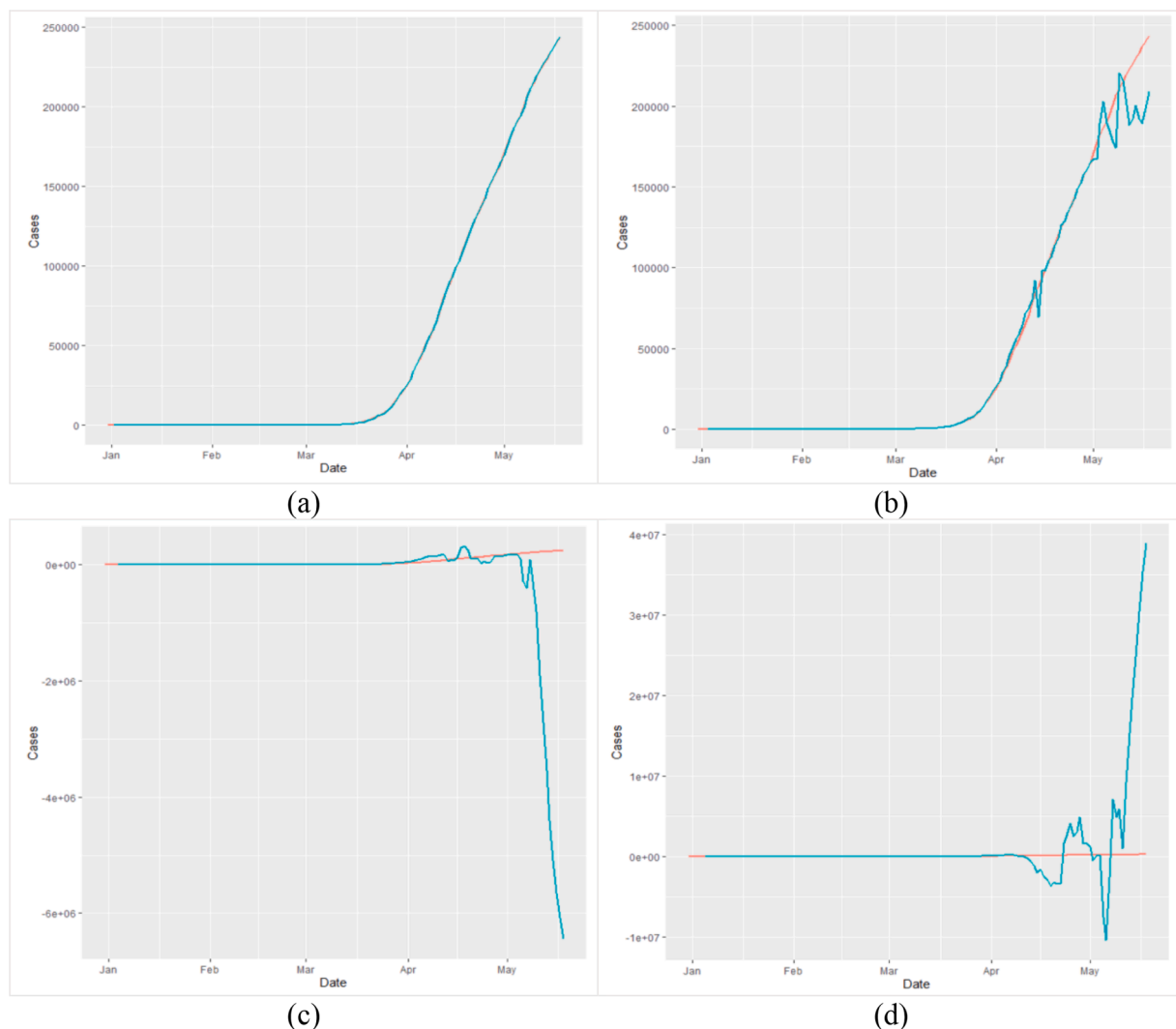
**Figure 1.** Comparison of ANFIS using (a) two inputs (ANFIS1), (b) three inputs (ANFIS2), (c) four inputs, (d) five inputs (2, 3, 4, and 5 days before the day to predict respectively). The red and blue lines represent the actual and predicted number of cases respectively.

samples are chosen based on the best result over trials. This FIS receives two inputs: $X_{t-1}$, $X_t$ and calculates $X_{t+1}$. $X_t$ refers to the number of cases for day t, where $X_{t+1}$ is the number of cases for the next day of day t; this convention is used throughout the report. The initial FIS is generated with the fis.builder function in the FuzzyR package Chen et al. (2020) with a full ruleset.

To explore the characteristic of time series, the number of inputs is increased to get more information about the past number of COVID-19 cases. ANFIS2 is created with three inputs: $X_{t-2}$, $X_{t-1}$, $X_t$ and calculates $X_{t+1}$. As this model produces worse performance, further tests in having more inputs are carried out to confirm the idea, which is illustrated in Figure 1.

In any input-output system, choosing the right input information for the forecasting system is important to generate the expected outcome. To compare with the above models, this research designs an ANFIS3 with two inputs: $X_{t-3}$, $X_{t-1}$. This model aims at evaluating the day gap variation in the time sequence of inputs.

Using ANFIS1, the number of MFs is changed for inputs to analyze the difference in performance. Because both two inputs of the targeted FIS present the number of COVID-19 cases, ANFIS4 changes the number to 3 MFs for the pair of inputs. More experiments taken prove that increasing the number of MFs will produce unstable outputs, which are shown clearly in Figure 2.

### 4.2. Different rules

Examine the rules settings, this study takes ANFIS1 and change the number of rules to 2 for ANFIS5, 3 for ANFIS6. Since the scenario explores time series, ANFIS5 takes the pairs of rules that experience changes in the two inputs: from low to high and from high to low number of cases. ANFIS6 adds in ANFIS5 a case of high value for both inputs. It is also noticed that for manually defined FIS, changing any rule's output is observed to make no changes in the accuracy measurement.
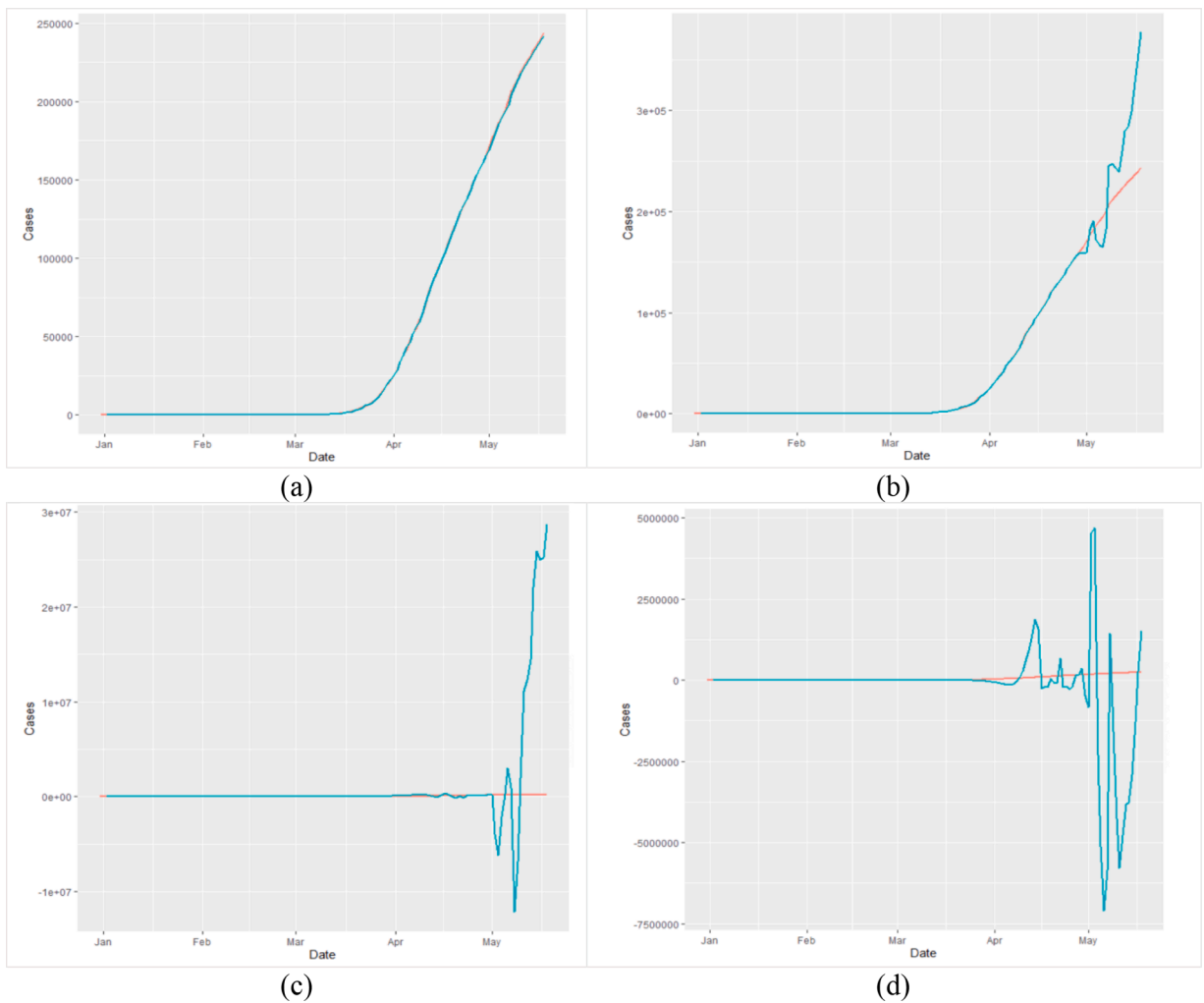
**Figure 2.** Comparison of ANFIS using (a) 2 MFs (ANFIS1), (b) 3 MFs (ANFIS4), (c) 4 MFs, (d) 5 MFs for two inputs. The red and blue lines represent the actual and predicted number of cases respectively.

### 4.3. Different FIS/ANFIS definition

For ANFIS7, this article creates a model using the same setting as ANFIS1, but define the initial FIS manually. The FIS is a TSK model with *gbellmf* for both two inputs $X_{t-1}$ and $X_t$ (Figure 3). The result points out that choosing the appropriate set of input parameters brings better performance than ANFIS1. It is also observed that there will be no big difference in changing either the number of MFs for the output or the parameters for each MF of the output.

Apart from the traditional type-1 fuzzy system, interval type-2 (it2) fuzzy sets are reported to be more efficient when dealing with uncertainty by defining fuzzy MFs. ANFIS8 implements it2 to compare the performance of these two types of fuzzy sets.

### 4.4. Different COVID-19 dataset

From another perspective, ANFIS9 explores the difference in choosing the dataset. By giving more pairs for training purposes: 90% for training, 5% for both validation and testing, we notice a gradual increase (Figure 4). It is concluded that only UK data is not sufficient for training a good forecasting system.

With the collected data from 10 countries, Figure 5 illustrates the changes over time to analyze the similarity among nations. The
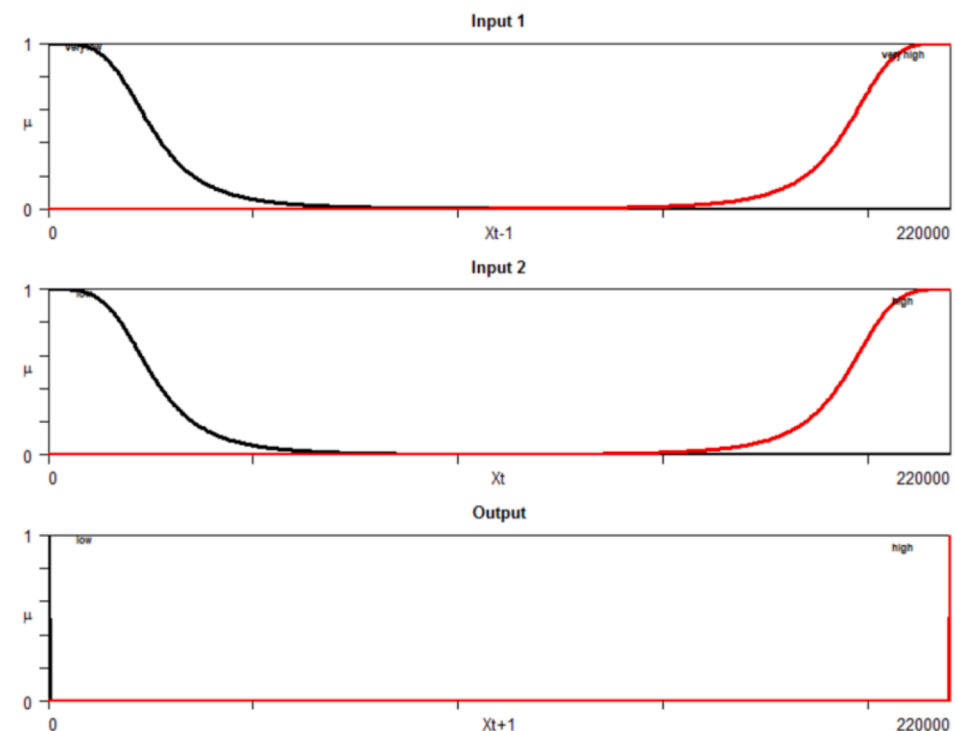
**Figure 3.** Defined MFs for ANFIS7.

growing trend of the UK is observed to be parallel to that of Italy, where the UK is two weeks behind Italy in the number of COVID-19 cases. For this reason, the next model will combine these two sets of numbers into a full data set for ANFIS10. Noted that for the best observation, validation and testing data only consider data from the UK.

As a further attempt to increase the dataset, Spain is also partly similar to the UK in the growing number of COVID-19 cases. Two approaches are taken into account in combining data. The first one is to blend all UK, Italy, and Spain data (ANFIS11). It is found that the trend of the UK is comparable to Italy for the lower number of cases and more identical to Spain for the rest. Therefore, ANFIS12 is designed to take all UK data, the first part of Italy data below 130000 cases, and Spain data above 130000 cases. The purpose of ANFIS12 is to not only get more data but also choose the closest information to what is going on in the UK.

## 5. Result and comparisons

This section first analyzes the training, validation, and testing errors of alternative ANFIS models, which are illustrated in Table 1. A short summary and conclusion of what this study targets to compare in each ANFIS is also provided. To avoid confusion, apart from what is listed in the note, other settings remain the same as the mentioned ANFIS for comparison.

From the recorded errors shown in Table 1, this research designs a final model with the better method among compared elements of an ANFIS. Accordingly, the initial FIS is created manually with fuzzy MFs specifically for an interval type-2 fuzzy set. The model takes two inputs: $X_{t-3}$, $X_{t-1}$ and calculates $X_{t+1}$ to forecast the next day's number of COVID-19 cases in the UK. There are 2 MFs for inputs and 2 MFs for output, as shown in Figure 6. Details of the membership functions:

- $X_{t-3}$, $X_{t-1}$: range [0; 220000].
- ■ Low: *it2gbellmf*, c(19000,26000, 2, 0)
- ■ High: *it2gbellmf*, c(19000,26000, 2, 220000)
- $X_{t+1}$: range [0; 220000].
- ■ Low: *linearmf*, c(0, 0, 0)
- ■ High: *linearmf*, c(220000, 0, 0)
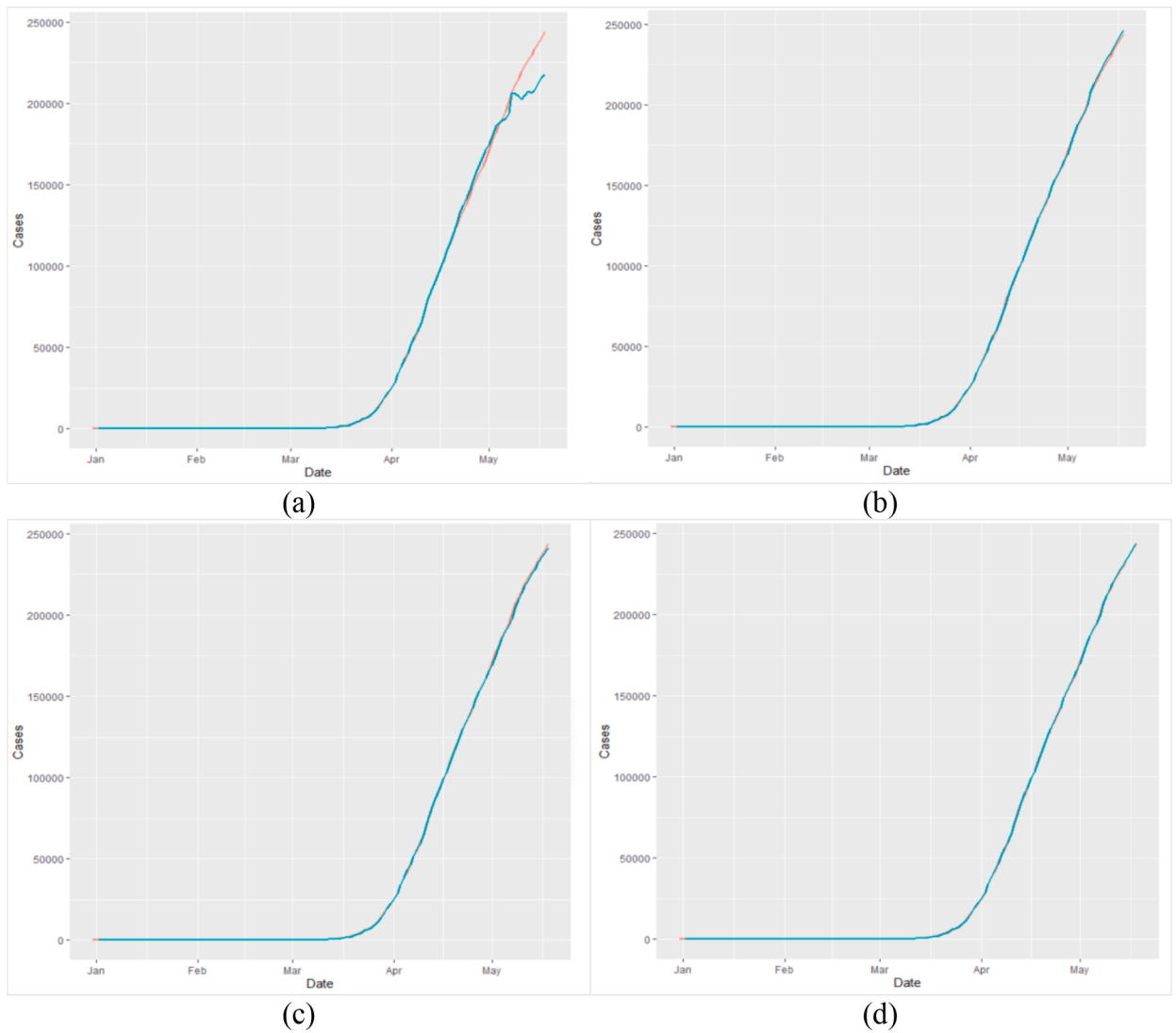
**Figure 4.** Comparison of ANFIS using (a) 80%, (b) 84%, (c) 88%, (d) 90% (ANFIS9) for training. The red and blue lines represent the actual and predicted number of cases respectively.
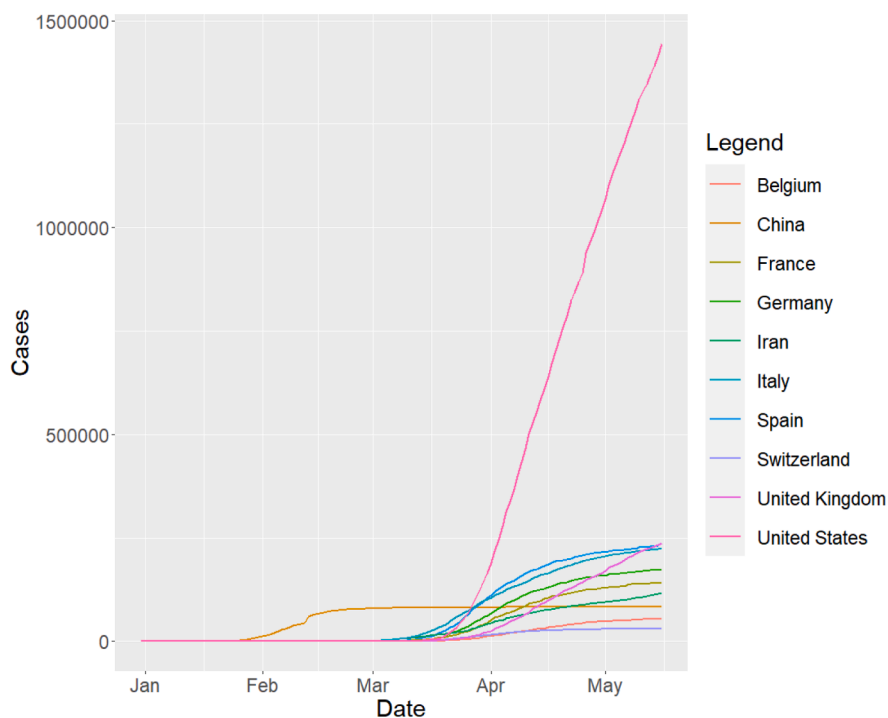
**Figure 5.** Time series of the number of COVID-19 cases in 10 countries.

On the other hand, Table 2 describes the 2 chosen rules for the model. The dataset collectively combines the UK, Italy (below 130000 cases) and Spain (above 130000 cases) data. Accordingly, 88% of all data is used for training, 6% for validation and 6% for testing.

The UMBRAE measure for our final ANFIS presents 1.2709837 for training, 0.07338731 for validation, and 0.4634651 for testing. Figure 7 shows the time series outcome of the proposed model.

## 6. Discussion

Most ANFIS models perform better than the defined benchmark method. Among these alternative models, the final ANFIS produces the best validation error. All design decisions were taken with various trials to make sure that the results have enough objectivity, along with throughout understandings of a good ANFIS. The surface of the initial FIS is shown in Figure 8. With the same $X_{t-3}$, increasing the $X_{t-1}$ will result in a rise in $X_{t+1}$. When $X_{t-3}$ is higher than $X_{t-1}$, $X_{t+1}$ remains unchanged. The outcomes suit the general convention and intuition of a time series system, therefore, successfully produce our desired performance.

Since the real world changes constantly, the approach of data integration with other countries prevents the model from ignorance of the changes and ensures long-term performance. Taking solely the UK data may result in bias and fail to update the knowledge in the future. Especially, Italy and Spain not only have similar trends in COVID-19 cases to the UK but also went before the UK, which enables better prediction if combining these with UK data. Overall, for the best forecasting result, training data should be updated regularly.

As the concluded model creates initial FIS manually, the range of inputs and outputs will need to be changed over time, which can also affect the parameters in MFs and lead to performance variation. Furthermore, the collected data shows the number of people that are tested positive to coronavirus, whereas some people who are infectious may not turn up for testing or recover without going to the hospital. For future development, a model can be taken to estimate this missing information. Finally, although this system only examines the number of COVID-19 cases anticipation, similar approaches can be taken to forecast the number of deaths or recovered in the UK.

## CRediT authorship contribution statement

**Kim Tien Ly:** Methodology, Validation, Formal analysis, Investigation, Writing - review & editing, Project administration.

**Table 1**
Results of alternative ANFIS models.

| ANFIS | Training error | Validation error | Testing error | Conclusion |
|---|---|---|---|---|
| **ANFIS1:** | 1.4534485 | 0.2354300 | 0.4062940 | |
| • Type-1 fuzzy set | | | | |
| • 2 inputs: $X_{t-1}$, $X_t$; 2 MFs each | | | | |
| • Full rule set | | | | |
| • Automatically created initial FIS | | | | |
| • UK data, 86% for training | | | | |
| **ANFIS2:** | 1.5382697 | 1.356089 | 4.515143 | Model with 2 inputs is better than 3 |
| • 3 inputs: $X_{t-2}$, $X_{t-1}$, $X_t$; 2 MFs each | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS3:** | 1.29636590 | 0.1760187 | 1.973486 | 2 days gap in inputs is better than 1 |
| • 2 inputs: $X_{t-3}$, $X_{t-1}$; 2 MFs each | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS4:** | 1.39449704 | 2.058999 | 12.26301 | 2 MFs for each input is better than 3 |
| • 2 inputs: $X_{t-1}$, $X_t$, 3 MFs each | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS5:** | 1.4474273 | 0.1270205 | 0.4294762 | Set of 2 rules (2,3) is better than 3 or 4 (full) rules |
| • 2 rules: 2, 3 | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS6:** | 1.3880980 | 0.1565407 | 0.4207974 | |
| • 3 rules: 2, 3, 4 | | | | |
| • Compare with ANFIS1,5 | | | | |
| **ANFIS7:** | 1.3655231 | 0.1402058 | 0.5538689 | Creating FIS manually is better |
| • Manual created initial FIS | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS8:** | 1.2822973 | 0.1524152 | 1.929276 | IT2 is better than T1 |
| • Interval type-2 fuzzy set | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS9:** | 1.3287442 | 0.1246852 | 0.2232682 | Selectively combining UK, Italy and Spain data produces the best result |
| • UK data, 90% for training | | | | |
| • Compare with ANFIS1 | | | | |
| **ANFIS10:** | 1.1981300 | 0.1556331 | 0.4258049 | |
| • Combine UK and Italy data, 90% for training | | | | |
| • Compare with ANFIS9 | | | | |
| **ANFIS11:** | 1.1285839 | 0.1452932 | 0.4431160 | |
| • Combine UK, Italy and Spain data, 90% for training | | | | |
| • Compare with ANFIS9, 10 | | | | |
| **ANFIS12:** | 1.1645781 | 0.09887350 | 0.6810328 | |
| • Combine UK, Italy (below 130000 cases) and Spain (above 130000 cases) data, 90% for training | | | | |
| • Compare with ANFIS9, 10, 11 | | | | |

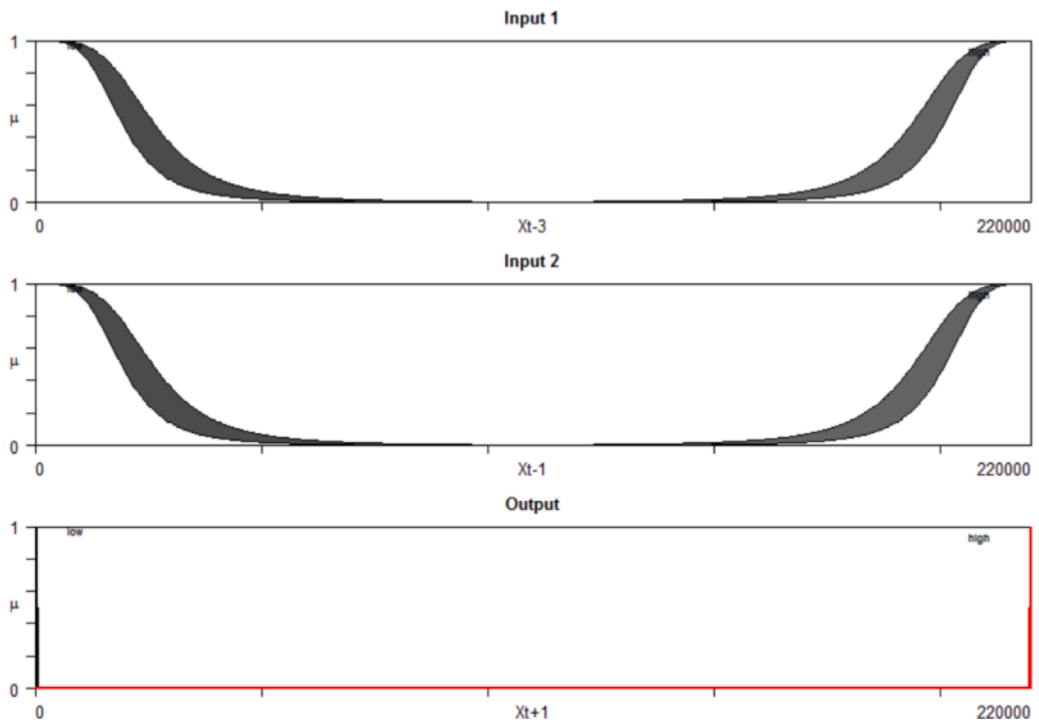**Figure 6.** MFs definition for the final ANFIS model.

**Table 2**
Rule set for the final ANFIS model.

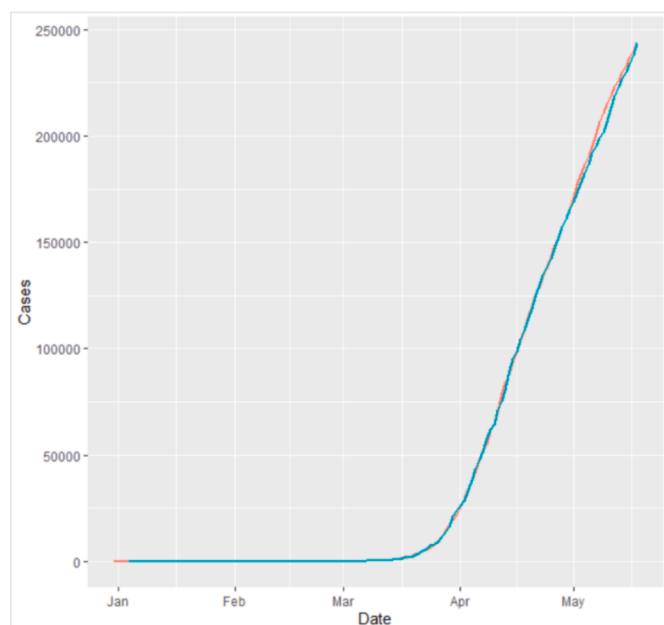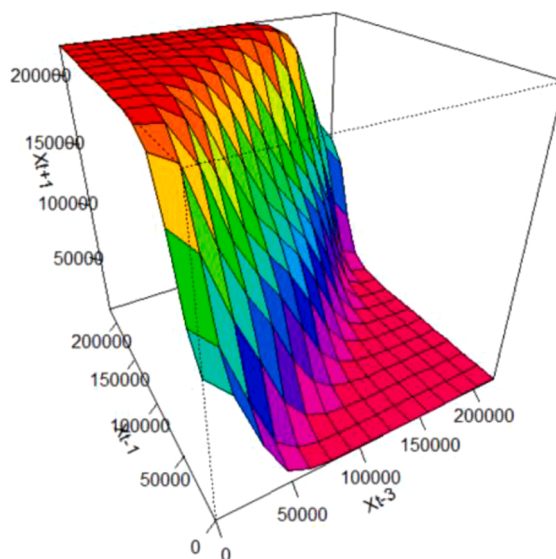| | | $X_{t-1}$ | |
|---|---|---|---|
| | | Low | High |
| $X_t$ | Low | | Low |
| | High | High | |



**Figure 7.** Time series of the final ANFIS model. The red and blue lines represent the actual and predicted number of cases respectively.

**Figure 8.** The surface generated by the initial FIS for the final model.

# References

Akhtaruzzaman, M., Boubaker, S., Sensoy, A., 2020. Financial contagion during COVID-19 crisis. Finance Research Letters forthcoming.

Alasha`ary, H., Moghtaderi, B., Page, A., Sugo, H., 2009. A neuro–fuzzy model for prediction of the indoor temperature in typical Australian residential buildings. Energy and Buildings 41 (7), 703–710.

Arias, J., 2020. Covid-19 effect on herding behaviour in European Capital Markets. Finance Research Letters forthcoming.

Chen, C., Garibaldi, J., Razak, T., 2020. FuzzyR: An Extended Fuzzy Logic Toolbox for the R Programming Language. IEEE International Conference on Fuzzy Systems.

Chen, C., Twycross, J., Garibaldi, J.M., 2017. A new accuracy measure based on bounded relative error for time series forecasting. PLoS ONE 12 (3).

Choi, S-Y., 2020. Industry volatility and economic uncertainty due to the Covid-19 pandemic: Evidence from wavelet coherence analysis. Finance Research Letters forthcoming.

Frank, R.J., Davey, N., Hunt, S.P., 2001. Time Series Prediction and Neural Networks. Journal of Intelligent and Robotic Systems 91–103.

Github of The Center for Systems Science and Engineering (CSSE). (2020). Retrieved from https://github.com/CSSEGISandData/COVID-19.

Just, M., Echaust, K., 2020. Stock Market Returns, Volatility, Correlation and Liquidity during the COVID-19 Crisis: Evidence from the Markov Switching Approach. Finance Research Letters forthcoming.

Kim, K.-j, 2003. Financial time series forecasting using support vector machines. Neurocomputing, 55 (1-2), 307–319.

Li, Y., Liang, C., Ma, F., Wang, J., 2020. The role of the IDEMV in predicting European stock market volatility during the Covid-19 pandemic. Finance Research Letters forthcoming.

Leone, R.P., 1987. Forecasting the effect of an environmental change on market performance: An intervention time-series approach. International Journal of Forecasting 3 (3-4), 463–478.

Ortmann, R., Pelster, M., Wengerek, S.T., 2020. Covid-19 and investor behavior. Finance Research Letters forthcoming.

Rizwan, M.S., Ahmad, G., Ashraf, D., 2020. Systemic risk: The impact of Covid-19. Finance Research Letters forthcoming.

The Guardian. (2020). FTSE 100 suffers worst quarter since 1987 as Covid-19 recession looms - as it happened.

Taylor, J.W., McSharry, P.E., Buizza, R., 2009. Wind Power Density Forecasting Using Ensemble Predictions and Time Series Models. IEEE Transactions on Energy Conversion 24 (3), 775–782.

Zhang, D., Hu, M., Ji, Q., 2020. Financial markets under the global pandemic of Covid-19. Finance Research Letters forthcoming.