sustainability

Article

A New Collaborative Multi-Agent Monte Carlo Simulation Model for Spatial Correlations Air Pollutions Global Risk Assessment

Mustafa Hamid Hassan [1,2], Salama A. Mostafa [1,*], Aida Mustapha [3], Mohd Zainuri Saringat [1], Bander Ali Saleh Al-rimy [4], Faisal Saeed [5], A. E. M. Eljial [6] and Mohammed Ahmed Jubair [1,2]

1 Faculty of Computer Science and Information Technology, Universiti Tun Hussin Onn Malaysia, Panchor 84600, Johor, Malaysia; mustafa.hamid.alani@gmail.com (M.H.H.); zainuri@uthm.edu.my (M.Z.S.); mohammed.a@sadiq.edu.iq (M.A.J.)
2 College of Information Technology, Imam Ja'afar Al-Sadiq University, Al-Muthanna, Iraq
3 Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Panchor 84600, Johor, Malaysia; aidam@uthm.edu.my
4 School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310, Johor, Malaysia; bander@utm.my
5 School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK faisal.saeed@bcu.ac.uk
6 Department of Information System, College of Computer Engineering & Sciences, Prince Sattam Bin Abdulaziz University, P.O. Box. 151, Alkharj 11942, Saudi Arabia; ae.mohammed@psau.edu.sa
* Correspondence: salama@uthm.edu.my

Abstract: Air pollution risk assessment is complex due to dynamic data change and pollution source distribution. Air quality index concentration level prediction is an effective method of protecting public health by providing the means for an early warning against harmful air pollution. However, air quality index-based prediction is challenging as it depends on several complicated factors resulting from dynamic nonlinear air quality time-series data, such as dynamic weather patterns and the verity and distribution of air pollution sources. Subsequently, some minimal models have incorporated time series-based predicting air quality index at a global level (for a particular city or various cities). These models require interaction between the multiple air pollution sensing sources and additional parameters like wind direction and wind speed. The existing methods in predicting air quality index cannot handle short-term dependencies. These methods also mostly neglect the spatial correlations between the different parameters. Moreover, the assumption of selecting the most recent part of the air quality time series is not valid considering that pollution is cyclic behavior according to various events and conditions due to the high possibility of falling into the trap of local minimum and poor generalization. Therefore, this paper proposes a new air pollution global risk assessment (APGRA) model for predicting spatial correlations air quality index risk assessment to address these issues. The APGRA model incorporates autoregressive integrated moving average (ARIMA), Monte-Carlo simulation, and collaborative multi-agent system, and prediction algorithm for reducing air quality index prediction error and processing time. The proposed APGRA model is evaluated based on Malaysia and China real-world air quality datasets. The proposed APGRA model improves the average root mean squared error by 41%, mean and absolute error by 47.10% compared with the conventional ARIMA and ANFIS models.

Keywords: air quality index; air pollution; risk assessment; autoregressive integrated moving average; Monte-Carlo simulation; multi-agent system

Citation: Hassan, M.H.; Mostafa, S.A.; Mustapha, A.; Saringat, M.Z.; Al-rimy, B.A.S.; Saeed, F.; Eljial, A.E.M.; Jubair, M.A. A New Collaborative Multi-Agent Monte Carlo Simulation Model for Spatial Correlations Air Pollutions Global Risk Assessment. *Sustainability* **2022**, *14*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Received: date
Accepted: date
Published: date

1. Introduction

Air quality has drawn much attention in recent years because it seriously affects people's health. Nowadays, monitoring stations in a city can provide real-time air quality measures [1]. Still, people strongly desire air quality prediction, which is challenging as it depends on several complicated factors, such as weather patterns and spatial-temporal dependencies of air quality. Air pollution risk assessment is complex due to its dynamic data and distributed pollution sources [2]. For instance, predicting air quality on weekdays and weekends may be different due to the difference in anthropic emissions [3]. Air Quality Index (AQI) effectively protects public health by communicating early warnings of harmful air pollutants. However, the prediction is challenging because it depends on several complex factors, such as weather patterns, nonlinear time series of air quality data, and distribution of air pollution sources [4,5]. The dynamic data and distributed air pollution risk assessment sources need to be estimated relying on two phases. The first phase is utilized to predict the AQI of a local area. The second phase is employed to assess the global risk level based on the AQIs of local areas [6,7].

Air quality prediction aims to predict the future state of air quality in a specified location based on existing data, like historical air quality and meteorological data. Many types of research have been conducted to tackle the problem of assessing the air pollution risk. Some examples are the works of [4,7,8]. However, each of them has mostly focused on assessing the concentration of a specific pollutant parameter such as $PM_{2.5}$, CO, and PM_{10} [7,10]. On the other hand, some approaches have been focused on predicting the level of certain parameters that directly impact the pollution state [6,9]. The literature provides a significant number of works in air pollution prediction models either for a specific location or specific variables. Feng et al. [11] combin air mass trajectory analysis and wavelet transformation with artificial neural network (ANN) to improve the prediction accuracy of daily average concentrations of $PM_{2.5}$. Tong et al. [8] deploy Monte-Carlo simulation (MCS) to estimate health risk conditions related to the concentration of dust-induced occupational.

Prasad et al. [12] used an adaptive neuro-fuzzy inference system (ANFIS) for predicting the concentration of several AQI parameters. However, these models predict the air pollution concentrations based on the most recent part of the time series. This mechanism requires larger data for producing proper prediction results. It is highly possible to fall into the trap of local minimum [4,10]. Moreover, the learning or training of these models with short-term prediction situations may never converge due to the training data/time insufficiency, which might cause the algorithms to be trapped in an infinite training situation [13]. Because of these constraints, the statistical approach represents the best option. One of the best statistical approaches that deal with short-term time series prediction is the autoregressive integrated moving average (ARIMA) algorithm, as it only requires prior data of a time series to generalize the prediction of the AQI model [4]. However, the ARIMA algorithm does not produce satisfactory results for certain air pollution parameters (i.e., PM_{10} and CO), even for a short prediction period [14–16].

On the other hand, many approaches have tackled the problem of weather variables prediction and forecasting and pollution estimation and alarming. Each one has concentrated on one aspect, while some strategies have focused on predicting the level of a certain variable that has a critical impact on the pollution state [6]. Others have dealt with the issue of lacking adequate measurement stations across the countries [9]. Moreover, some approaches have focused on building models for estimating air pollution more accurately based on feature selection or neural networks. Others have built one ahead forecasting models [11,14,15] and have used fuzzy models for alarming air pollution [4,7]. Hence, a set of neural networks and simple auto-regression forecasters can be used, such as in the work of Westerlund et al. [17] that is validated to be superior over a single forecaster. The literature has revealed that very limited models have been constructed to assess the global (interaction between the pollution sources) level of air pollution. However, the dynamic nature and high spatiotemporal variability of AQI represent a complex predicting problem. Hence, non-of the existing models are able to incorporate time-series data to provide dynamic forecasting of various weather variables. This can be achieved by tackling the global interaction of different locations using wind direction and speed for enabling contextual forecasting added to the mathematical model. Consequently, the research gap lies in the absence of interaction between air pollution parameters under investigation and finding the global prediction level of the dynamic and distributed air pollution risk.

This paper identifies that the global interaction among meteorological parameters like wind speed and wind direction at the different areas is essential in air pollution prediction and risk assessment due to the nature of dynamic weather and air pollution time series at various locations. The study in this paper aims to fill this gap by searching for ways to overcome the conditional heteroscedasticity problem. The contributions of this paper are represented by (i) To develop ARIMA-based MCS prediction algorithm that integrates ARIMA and MCS algorithms for reducing AQI prediction error (ii) To propose an air pollution global risk assessment (APGRA) model that incorporates the ARIMA-based MCS algorithm into a multi-agent system (MAS) for dynamic and distributed assessment of multiple sources of AQI risk levels, and (iii) To test and evaluate the performance of APGRA models in terms of prediction error and time by using China and Malaysia air quality datasets.

The remaining parts of the paper are organized as follows: Section 2 provides Materials and Methods. Section 3 describes the Monte-Carlo method to be used in accommodating the uncertainty of the forecast. Then it presents a proposed APGRA model. Section 4 discusses the results of the APGRA model, and Section 5 provides conclusions and future research.

2. Materials and Methods

This section covers the materials and methods which are used in this paper. At the same time, it first explains the air pollution datasets divided into two case studies: Malaysia and China. Second, it explains the prediction algorithms that are used in the local risk assessment, such as MCS and ARIMA. Third, it explains several performance measures such as root mean square error (RMSE) and mean absolute error (MAE) used to evaluate this work.

2.1. Air Pollution Datasets

This study utilizes two real-world air pollution datasets of Malaysia and China. The data layout is presented as a matrix $D = \{x_t^{i,j}\}$, where $i = 1,2 \ldots M, j = 1,2 \ldots N,$ M indicates the number of variables, and M indicates the number of cities, t represents the time, and it sampled as hourly. The data is fed into the framework for one of two goals, forecasting one of the time series within a certain defined time horizon or evaluating a given model configuration in terms of its forecasting accuracy in a certain time interval. Another element of the data is the map combined with longitude and latitude for all cities with their time series included in the matrix D. The datasets are described in the following subsections.

2.1.1. Malaysia Air Pollution Dataset

This paper applies the heterogeneous data set, including the one-dimension series data and the multi-dimension panel data. The one-dimension series data is composed of the value of AQI concentrations with the change of time. For the panel data, the "Sulphur Dioxide (SO_2)", "Nitrogen Dioxide (NO_2)", "Carbon Monoxide (CO)", "Sulfur Dioxide Ozone (O_3)" concentrations, temperature, relative humidity (RH), wind speed (WS), and PM_{10} concentrations of the previous hour are selected as the input variables. The AQI concentrations of the current hour are the output variable of the forecasting model. The Malaysia air quality monitoring network gathers the PM_{10}, SO_2, NO_2, CO, and O_3

concentrations data. The air quality monitoring stations include ten stations, as illustrated in Figure 1.

Figure 1. The geographic position of the Malaysia case study.

The data of air pollutant concentrations are collected from the different cities of Malaysia. Table 1 shows the locations of the included states in this study. Hourly air quality data have been collected from the eight air pollution monitoring stations during the ten years from 2016 to 2016 in Malaysia. These stations record data of some important AQI parameters such as the CO, NO_2, O_3, and "Particulate Matter (PM_{10})". These parameters are used to calculate the AQI. AQI is a commonly used indicator defined by the United States Environmental Protection Agency (EPA) to use air quality conditions. In order to calculate AQI for a location, an indicator value of AQI is calculated for each of the observed pollutant concentrations (CO, NO_2, O_3, and PM_{10}) using Equation (1) [18].

$$AQI = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} * (C \quad - C_{low}) + I_{low} \tag{1}$$

Table 1. Area of air quality data in Malaysia.

NO	Site State	Site ID	Location	Latitude	Longitude	Type
1	Johor	CAS 001	SM Pasir Gudang 2, Pasir Gudang, Johor	N01° 28.225	E103° 53.637	Residential
2	Terengganu	CAE 002	SRK Bukit Kuang, Teluk Kalung, Kemaman.	N04° 16.260	E103° 25.826	Residential
3	Pulau Pinang	CAN 003	Sek. Keb. Cenderawasih,Tmn. Inderawasih, Perai	N05° 23.470	E100° 23.213	Residential
4	Sarawak	CAK 004	Medical Store, Kuching, Sarawak	N01° 33.734	E110° 23.329	Residential
5	Melaka	CAS 006	Sek. Men. Keb. Bukit Rambai, Melaka	N02° 15.510	E102° 10.364	Residential
6	Pahang	CAE 007	Pej. Kajicuaca, Batu Embun, Jerantut, Pahang	N03° 58.238	E102° 20.863	Residential
7	Perak	CAN 008	SM Jalan Tasek, Ipoh, Perak	N04° 37.781	E101° 06.964	Residential
8	Pulau Pinang	CAN 009	SK Seberang Jaya II, Perai, Pulau Pinang	N05° 23.890	E100° 24.194	Residential
9	Negeri Sembilan	CAC 010	Taman Semarak (Phase II), Nilai, N.Sembilan	N02° 49.246	E101° 48.877	Residential

10	Selangor	CAC 011	SM(P) Raja Zarina, Klang, Selangor	N03° 00.620	E101° 24.484	Residential

2.1.2. China Air Pollution Dataset

The Beijing multi-site air-quality data dataset comprises AQI parameters for hourly from 10 measured air pollution monitoring locations countrywide [1]. The AQI data characterizes the Beijing public environmental areas for the 24-h care center. The climatological and meteorological data in the apiece AQI site are coordinated with China climatological management's adjacent climate. The historical time is from March 2013 to February 2017. Table 2 and Figure 2 show descriptive information related to the dataset.

Table 2. Area of air quality data in China.

Dataset Characteristics:	Multivariate, Time-Series
Number of Instances:	420,768
Area:	Physical
Number of Attributes:	18
Attribute Characteristics:	Integer, Real
Missing Values?	Yes
Associated Tasks:	Regression

The response AQI is classified into four categories: AQI $\le 35 \mu gm^{-3}$ (green), $35 \mu gm^{-3} < PM_{2.5} \le 75 \mu gm^{-3}$ (yellow), $75 \mu gm^{-3} <$ AQI $\le 150 \mu gm^{-3}$ (orange) and AQI $> 150 \mu gm^{-3}$ (red). The four numbers inside each colored node indicate the proportions of the AQI categories at each layer of the branch, and the percentage represents the marginal proportion of the sample at the node. Figure 2 shows the position of 36 air quality monitoring sites marked as purple and red circles and 15 metrological sites marked as blue triangles.

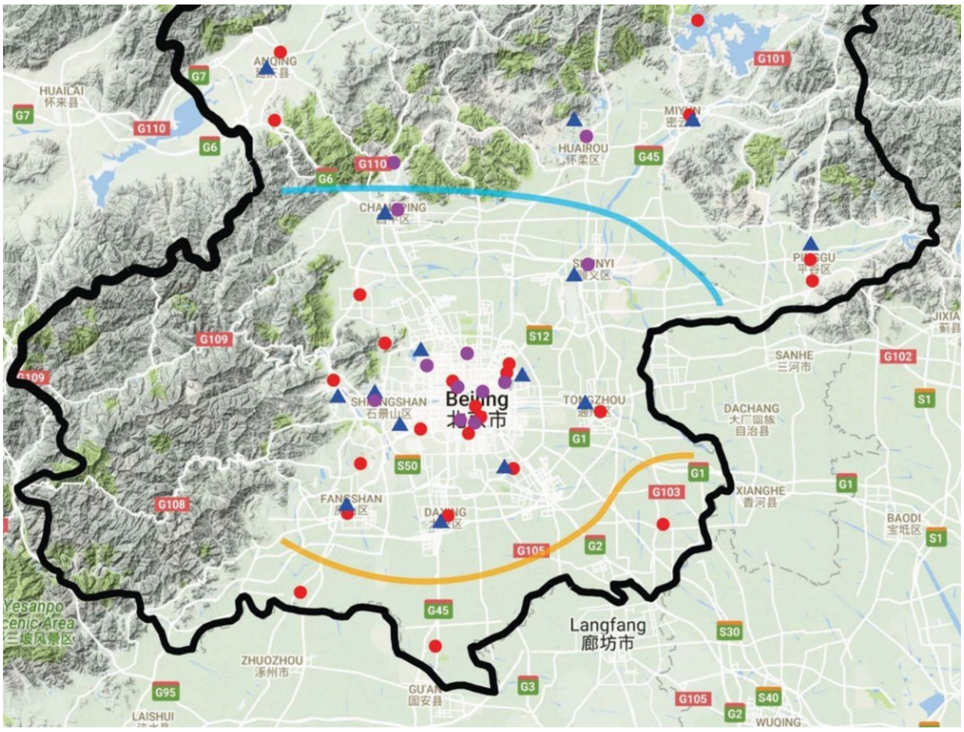


Figure 2. The geographic position of the China air pollution case study [1].

2.2. Prediction Methods

2.2.1. Descriptive Statistics

Descriptive statistics are used to quantitatively describe or summarize each monitoring station data's features for further explaining their implication. Mean and median are statistical terms introduced to understand the central tendency of the data. Minimum and maximum show the amplitude of the time series. Standard deviation is a measure for quantifying the amount of variation or dispersion of the data values. A low standard deviation indicates that the data points tend to be close to the data set's mean, while a high standard deviation indicates that the data points are spread out over a wider range of values. Skewness and kurtosis are applied to judge whether the sampling distribution is normal or not. Moreover, standard error of skewness (SES) and standard error of kurtosis (SEK) are presented to show the deviation between Skewness or Kurtosis's values.

2.2.2. ARIMA Algorithm

ARIMA, short for the auto-regressive integrated moving average, is actually a class of models that explains a given time series based on its own past values: its own lags and the lagged forecast errors so that Equation can be used to forecast future values. Any 'non-seasonal time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models [4]. An ARIMA model is characterized by three terms: p, d, q. Where p is the order of the AR term, q is the order of the MA term, d is the number of differences required to make the time series stationary. If a time series has seasonal patterns, then add seasonal terms and become SARIMA, short for seasonal ARIMA. The first step to building an ARIMA model is to make the time series stationary. Because the term "Auto Regressive" (AR) in ARIMA means it is a linear regression model that uses its own lags as predictors. Linear regression models work best when the predictors are not correlated and are independent of each other. The most common approach is to subtract the previous value from the current value. Sometimes, depending on the complexity of the series, differences might be needed. Therefore, the value of d is the minimum number of differences needed to make the series stationary. If the time series is already stationary, then d = 0. "p" is the order of the AR term. It refers to the number of lags of Y to be used as predictors. And "q" is the order of the "Moving Average" (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model. We adopt for forecasting the famous ARIMA model that is given by the Equation:

$$d^I x_t^{i,j} = \alpha_1 d^I x_{t-1}^{i,j} + \alpha_2 d^I x_{t-2}^{i,j} + \cdots \alpha_p d^I x_{t-p}^{i,j} + u_t + \beta_1 u_{t-1} + \cdots \beta_q u_{t-q} \quad (2)$$

2.2.3. Monte Carlo Simulation

Monte Carlo simulation (MCS) algorithms are mainly used in three problem classes: optimization, numerical integration, and generating draws from a probability distribution. MCS is one of the most common methods used to accommodate the uncertainties associated with many risk-related problems [8,19,20]. It has been recognized as a means of quantifying variability and uncertainty in risk assessments by the National Academy of Sciences and USEPA. This method provides a quantitative way to estimate the probability distributions for exposure risks and provides more information for making decisions related to risk protection. The widespread use of MCS in risk assessment promises a significant improvement in these assessments' scientific rigor [20]. The MCS method generally requires three main steps, which are intended as follows:

Step1: Construct a descriptive procedure to the probabilistic process:

- Built an appropriate probability model according to the simulated object's characteristics;
- Find a suitable distribution function to the desired solution;

Step2: Achieve sampling method from a known probability distribution:

- Generate a random variable (or random vector) with a known probability distribution;
- Generate a random variable of a sample;
- Establish the sampling method of the random distribution;

Step3: Establish various statistical estimators:

- Simulate a random variable as the solution to the object problem;
- Find the unbiased estimator.

Many statistics problems involve nested expectations and thus do not permit conventional MCS estimation. For such problems, a nest estimator, terms in an outer estimator, involve calculating a separate and nested expectation [19]. Nested expectations occur in a wide variety of portfolio risk management problems [21]. Tackling such problems requires some form of nested estimation scheme in the MCS. In this approach, MCS simulated as an interest in estimating quantities of the form:

$$\mathrm{E_Z} = [\mathrm{F}(\mathrm{E}(\mathrm{W}|\mathrm{Z})] \quad (3)$$

where Z represents deferent risk scenarios, and E[W|Z] represents exposure, conditional on the scenario.

2.3. Modeling Dynamic and Distributed Behavior

Dynamic and distributed problem solving is achieved by employing a MAS that has the behaviors and methods of interaction, communication, and collaboration [22,23]. The dynamic behaviors of the agent help the statistical methods such as ARIMA or MCS to perform dynamic prediction tasks and assess the risk with the availability of dynamic data sources [24,25]. The term “agent”, or software agent, has found its way into many technologies and has been widely used, for example, in artificial intelligence [26], data processing [25], operating systems [27], healthcare [28], and computer networks [28] literature. An agent can execute several behaviors concurrently. However, it is important to note that the scheduling of behaviors in an agent is collaborative rather than preemptive (as for running threads). This means when behavior is scheduled for execution, its action method is not called and runs until it returns but dynamically deliberate the selection of action options based on the agent and the environmental conditions [22,29].

On the other hand, the distributed risk assessment problem depends on the agent communication and collaboration features. Each agent represents a location or city in multiple cities environment in which the agents need to communicate with each other to assess the global risk of air pollution [23,30]. Agent communication is probably the most utilized feature of the Java Agent DEvelopment Framework (JADE) [31]. The communication paradigm is based on asynchronous message passing. Thus, each agent has a “mailbox” (the agent message queue) where the JADE run-time posts messages sent by other agents [30]. The receiving agent is notified whenever a message is posted in the mailbox message queue. However, the agent picking up the message from the queue for processing is a design choice of the agent programmer. This process is depicted in Figure 3.

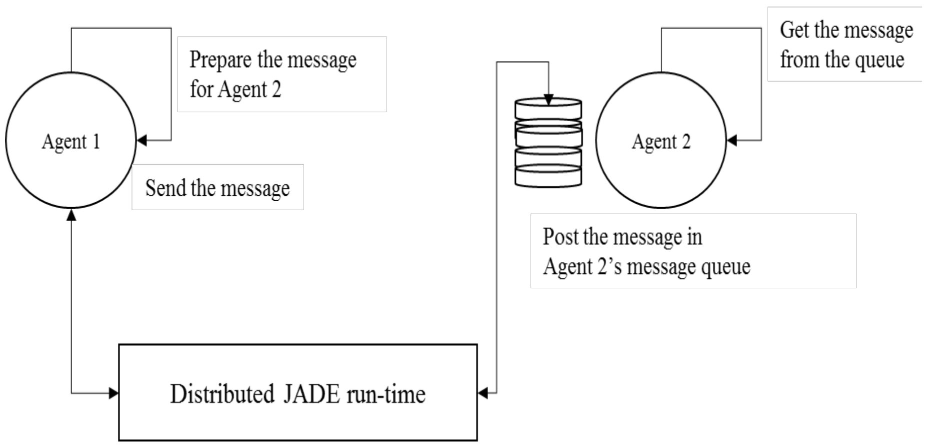


Figure 3. The asynchronous message passing in a MAS [31].

Each message includes the following fields: (i) the sender of the message, (ii) the list of receivers, and (iii) the communicative act (also called the "performative") indicating what the sender intends to achieve by sending the message. For instance, if the performative is REQUEST, the sender wants the receiver to act, if it is INFORM, the sender wants the receiver to be aware of a fact. The content containing the actual information to be exchanged by the message (e.g., the action to be performed in a REQUEST message, or the fact that the sender wants to disclose in an INFORM message, etc.). The content language indicates the syntax used to express the content. Both the sender and the receiver must be able to encode and parse expressions compliant with this syntax for the communication to be effective.

2.4. Evaluation Metrics

In order to evaluate the performance of a forecasting system, we use several model performance measures such as MAE, RMSE [25]. The formulas of the statistical measures used herein are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Yi - yi| \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - y_i)}{n}} \tag{7}$$

Y_i and y_i are the forecast value and the observed value, respectively. MAE and RMSE are applied as the performance criterion of the prediction model to quantify the errors of forecasting values. In general, the smaller the values, the better the prediction or the closer the estimator approaches the actually observed ones.

3. Air Pollution Global Risk Assessment (APGRA) Model

The Air Pollution Global Risk Assessment (APGRA) model consists of local air pollution risk assessment and Global air pollution risk assessment. The local air pollution risk assessment has an improved ARIMA-based MCS algorithm that performs local forecasting to the AQI risk of a particular area or city. Subsequently, in the Global AQI risk assessment, the APGRA model offers more accurate and global-oriented AQI forecasting through deploying MAS architecture in which agents are controlling the ARIMA-based MCS algorithm of cities. The APGRA model performs based on agents' interaction and processing of the AQIs' parameters.

3.1. Air Pollution Local Risk Assessment

This section explains the usage of the MCS to accommodate the uncertainty of the forecasting method presented in the ARIMA-based MCS algorithm. The concept of using MCS is to exploit the repeated sampling of the operation of the ARIMA outcomes to provide a more accurate description of the forecasting results of the local air pollution risk assessment. The ARIMA-based MCS algorithm defines a set of parameters that describe our usage of the MCS. Firstly, the algorithm selects the time interval that is used to fit the simulation model. *ARIMA(p,* I, q, i, j) function has the p, q, I, which are the same as presented earlier, and i, j represents the subject time series type and city that is under simulation. Secondly, the algorithm selects the time interval *T_past* that is used for fitting *ARIMA(p, I, q, i, j)*, and users assign it. Thirdly, the ARIMA-based MCS algorithm selects the time horizon *T_future* that the forecasting model uses. Fourthly, the algorithm selects the number of runs *N_runs* that MCS uses. The simulation results are two variables, namely, $\boldsymbol{y\_forecasted}$ and $\boldsymbol{\sigma\_forecasted}$.

Table 3 shows the ARIMA-based MCS. The main task of the MCS is to execute the forecasting of the ARMIA that is fitted in the requested $\boldsymbol{T\_past}$ for the requested $\boldsymbol{T\_future}$. This procedure is repeated several times equal to *N_runs*, representing the number of simulations. After accumulating all the forecasted time series, we calculated the random process summarized by the distribution of the predicted time series. Assuming that the distribution is normal, the ARIMA-based MCS algorithm offers two series, *y_forecasted*, which provides the forecasted time series, and *σ_forecasted*, which provides an indicator of the confidence or risk of the local AQI forecasting.

Table 3. Air pollution local risk assessment algorithm.

```
Input
    initial input< ARIMA (p, I, q, i, j); T_past; T_future; N_runs; y_history>;
    initial output< y_forecasted, sigma_forecasted>;
Output
    Y = [];
Start
    prediction model = fitARIMA (T_past, p, I, q, i, j);
    for t = 1 until No_runs do:
        y_forecasted = forecast (model, T_past, T_future);
        Y = add(y_forecasted);
    end
    sigma_forecasted=sqrt(variance(Y));
    y_forecasted=avg(Y);
End
```

3.2. Air Pollution Global Risk Assessment

The issue with the previous algorithm of the ARIMA-based MCS algorithm is its non-awareness of the global aspect regarding possible interaction between individual cities. In order to overcome this matter, the APGRA model is designed to assess a global value of the AQI that takes into account dynamic parameters reading at multiple, distributed local stations. The global model is developed based upon a MAS architecture consisting of many local agents representing a specific city. Local AQI forecast values are aggregated into the APGRA model based on the city's wind speed and direction under study. Figure 4 depicts the single-agent processing and how to be communicated with MAS.

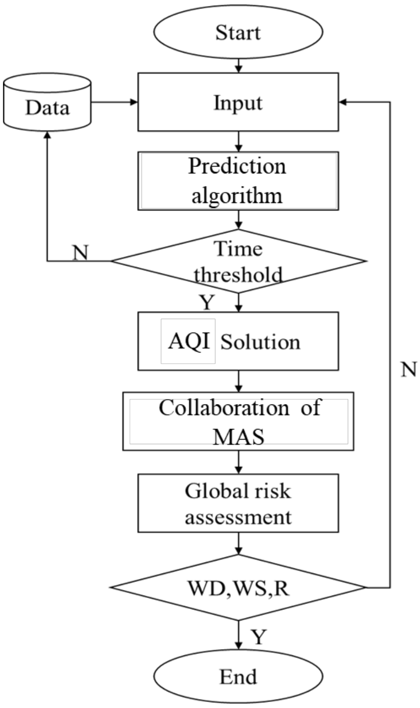


Figure 4. A single agent processing cycle.

The decisions affect the interval of training the agents' models and the horizon of forecasting and prediction. Each agent is equipped with an ARIMA-based MCS algorithm. The concept of collaborative MAS is essentially in estimating the global air pollution risk. In collaboration, agents work together to solve a complex problem of global risk while achieving their personal goals of local risk. The risk assessment visualization is a module used to present the risk assessment results for local and global risks. Based on Figure 4, the inputs of the APGRA model come from two sources: data provider and risk level, as shown in Figure 5. Firstly, the data provider is the historical time series data of various pollution and weather variables collected across all cities in the last T_y years. Secondly, the risk level is used to feed the data to the model by a mediator agent. The data fed into each city's computation layer is a combination of N agents, where N denotes the number of variables provided. Each agent is responsible for using the data to build the corresponding variable and the city's primary prediction algorithm. The agents are denoted as *Ai,j* where i = 1,2….M and j = 1,2…N indicates the number of variables, and M indicates the number of cities.

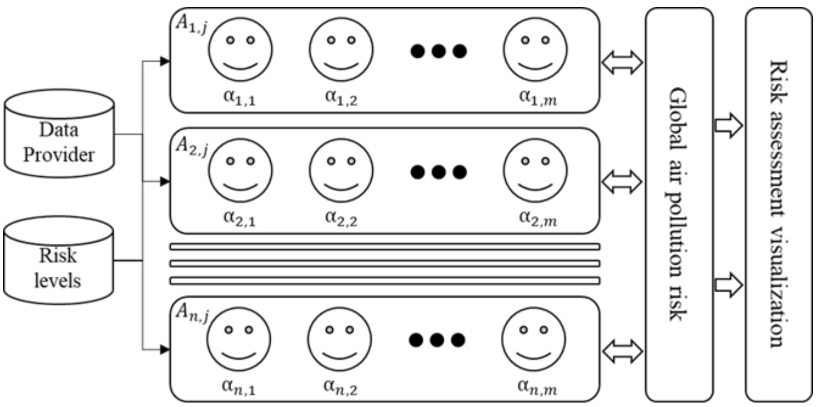


Figure 5. Air Pollution Global Risk Assessment (APGRA) model.

Figure 6 illustrates an example of local air pollution prediction and global air pollution assessment. The example consists of two neighboring cities that have an interaction effect. Each city has different local and global parameters with variable concentrations or values used to predict the local and global risk. The local pollution parameters are CO, NO_2, O_3, SO_2, and PM_{10}. In contrast, the global meteorological parameters are wind direction and wind speed. The local parameters are used to predict the local air pollution risk via utilizing the ARIMA-based MCS algorithm. Next, assess the global air pollution risk levels based on wind direction and wind speed by using the APGRA model.

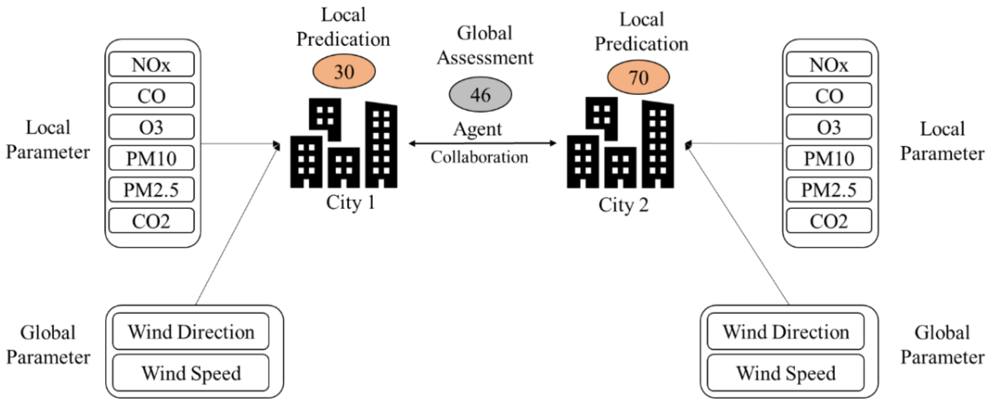


Figure 6. An example application of the APGRA model.

Agents communicate with each other, and the coordinator agent (mediator Agent) is responsible for interacting with the user and commanding two components; global AQI reading and local ARIMA-based MCS forecasts. This model includes the global risk assessment agent, which interacts with other agents to estimate the global risk as shown in Table 4.

Table 4. Air pollution global risk assessment algorithm.

```
Input
   A (i, j) //i=1, 2...n number of cities; j=1, 2...m number of time series //this represents the original
agent's models
   WS(i) //wind speed at city i
   WD(i) //wind direction at city i
   R    //Radius of interaction
   SpeedT //lower speed effect
Output
   AI (i, j)         //this represents the model after modifying with global interaction
Start
   for i=1: n   //to go through all cities
      cities=find Cities (i, R) //for each city we find influencing city
      for k=1: length(cities)
         if (WS(k)>SpeedT and WD(i) is toward location of city k)
            for j=1:m
               AI (i, j) =alpha*WS(k)*A (k, j) //to change all-time
               series to be affect by the source city
            end
            AI (i, j) =A (i, j) +AI (i, j)
         end
      end
   end
End
```

Based on Table 4, assuming that the $A_s(i,j)$ represents the agent that is responsible for forecasting. When a request for forecasting is given to $A_s(i,j)$, a circle with a radius R

will be created around the city i. Hence the surrounding cities will be taken as the source of effect to the subject model of $A_s(i,j)$. The effect source is represented as $SE(i,j) = (a_{j1}, a_{j2}, \ldots a_{je})$. Next, a vector of influence factors for each of the agents' $SE(i,j)$ is created based on the wind direction and speed described by the pseudocode.

This vector is called $fWE(i,j) = (w_{j1}, w_{j2}, \ldots w_{je})$. Then the forecasting model at the city i and the variable j will be as shown in Equation (4).

$$A_{gs}(i,j) = A_s(i,j) + w_{j1} \times a_{j1} + w_{j2} \times a_{j2} + .. w_{je} \times a_{je} = A_s(i,j) + SE.WE \quad (4)$$

The pseudo-code in Table 4 begins by scanning the cities one by one using the loop given in line number 1. Next, each city builds a circle around the city with radius R, and it checks the wind speed and direction of the subject city. If the wind speed is higher than a predefined threshold speed and the wind direction is towards the subject city, then it will be regarded as a source of effect to the subject city. Then the algorithm goes through the time series of the subject city one by one and changes them to include the effect of the corresponding time series of the subject city. A coefficient factor named alpha is used for adding the effect. After summing the effects of all source cities for a certain time series, it will be added to the subject city.

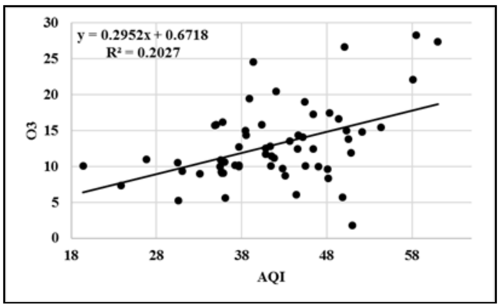
3.3. Risk Forecasting

The role of the APGRA is to issue an alarm when the AQI reaches a certain time series that indicate high risk. This alarm will be issued in a probabilistic way using the results of the ARIMA-based MCS and has been calculated using Equation 5.

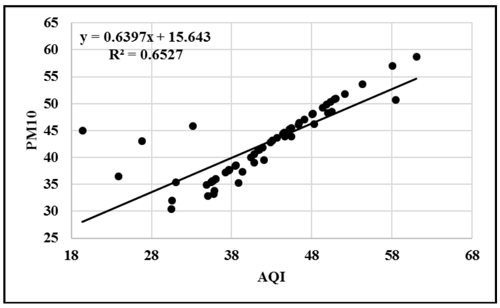
$$P_{L_i} = P(y(t) > L_i) = \frac{N_{L_i}}{N_s} \quad (5)$$

3.4. Correlation Analysis

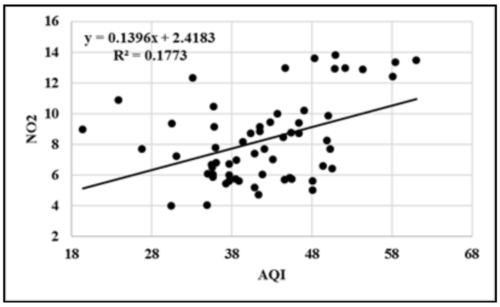
Correlation analysis is used to quantify the degree of relationship between two continuous variables, such as in between an independent and a dependent variable or between two independent variables. The correlation analysis is meant to prove or validate the correctness of the air pollution risk assessment operation. Figure 7a–e highlights the correlation between AQI reading and the concentrations of the parameters that affect air quality in the Malaysia case study, which are O_3, PM_{10}, NO_2, CO, and SO_2. The figures show that O_3, NO_2, and CO concentrations are around 0.2, with SO_2 having an even lover correlation of 0.08. These concentrations indicate a very low relationship with the AQI reading but will alert the prediction system if the concentrations increase. The highest correlation between the concentrations and the AQI is of PM10, which correlates with around 0.7. This indicates a high presence of particulate matter <10 μm in Malaysian air.



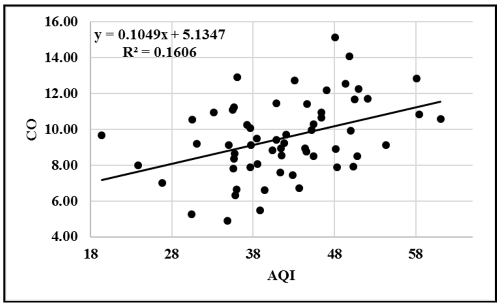
(**a**) O_3



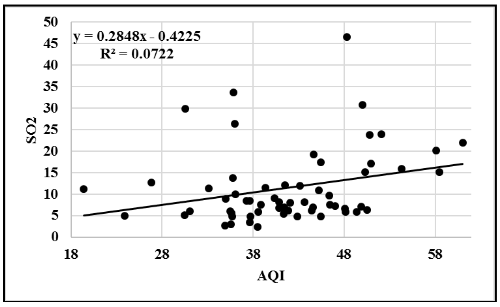
(**b**) PM_{10}



(**c**) NO_2



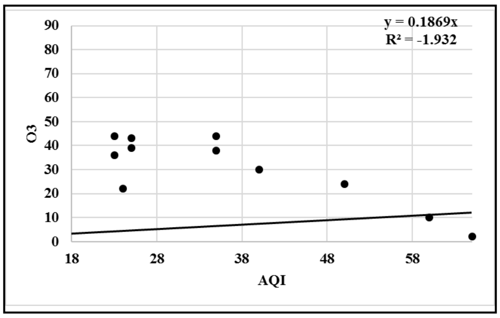
(**d**) CO



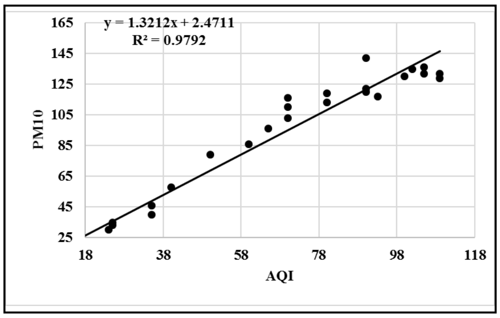
(**e**) SO_2

Figure 7. Correlation between AQI levels and all parameters in Malaysia.

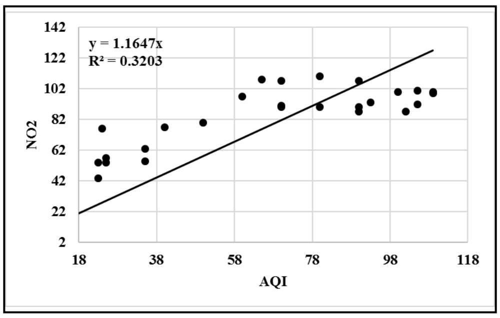
Next, Figure 8a–e highlights the correlation between AQI reading and all parameters that affect air quality in the China case study, which are O_3, PM_{10}, NO_2, SO_2, and $PM_{2.5}$. From the figures, it can be seen most of the concentrations correlations are higher in China as compared to Malaysia. Correlations of O_3 and NO_2 are low, around 0.2 and 0.3, respectively. Other parameters show a high correlation with AQI in China, with SO_2 around 0.6, while $PM_{2.5}$ and PM_{10} are both around 0.9.



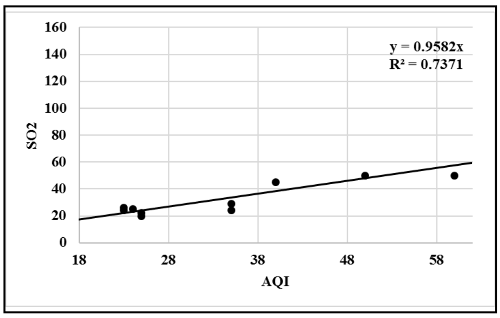
(**a**) O_3



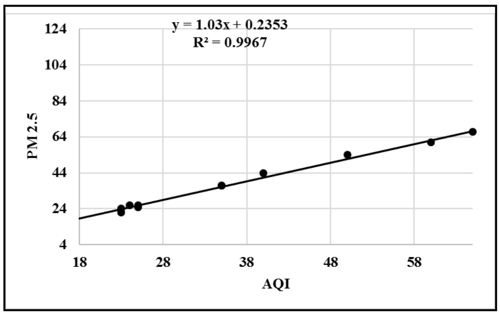
(**b**) PM_{10}



(**c**) NO_2



(**d**) SO_2



(**e**) $PM_{2.5}$

Figure 8. Correlation between AQI levels and all parameters in China.

For both case studies in Malaysia and China, the correlation analysis shows that the particulate matter, which is small enough to be suspended in the air, has a high degree of relationship with the AQI. In general, particle matter less than 10 μm in diameter can get deep into the lungs and some cases, into the bloodstream, which must be monitored closely by both countries. This implies that $PM_{2.5}$, tiny particles in the air that are two-, one-half microns or less in width, pose the greatest risk to health as compared to PM_{10}. Studies show that ambient $PM_{2.5}$ concentrations were significantly associated with influenza-like illness (ILI) risk in Beijing, China [11].

4. Results and Discussion

Prediction of the air pollutant concentrations represents a complex Spatio-temporal problem due to the dynamic nature and high Spatio-temporal variability in air pollution data. This section presents the results of the AQI prediction between three models; (1) the base model ARIMA, (2) the ANFIS model of Prasad et al. [12], (3) the improved ARI APGRA model by MA-based MCS algorithm. The prediction models' performance is evaluated based on accuracy based on MAE and RMSE as well as prediction ability based on the coefficient of determination R_2

4.1. Comparison between AQI Prediction Models

The experiments aim to examine the proposed models' effectiveness in predicting AQI concentrations for 1-day and 2-day in advance. The results are compared between the real AQI values against the base models of ARIMA, ANFIS, and the APGRA model for two separate case studies from Malaysia and China. Based on Table 5, the best prediction in the Malaysia case study is achieved for 1-day prediction is at City 10, and 2-day prediction is at City 7. The prediction of the first day yields a lower error than the second day since the prediction error for one day in advance is brought into the next day's prediction. The APGRA model produced the best results as compared to the direct approach in the APGRA and ANFIS models for both 1-day and 2-day predictions. This shows that the APGRA model plays an important role in obtaining good prediction results with approximately 41% improvement on RMSE. This is attributed to incorporating the uncertainty into the prediction, which allows for exploiting the repeated sampling of the operation that improves the accuracy of the forecasting. From the aspect of absolute errors, as measured by RMSE and MAE, the best prediction for 1-day is achieved by City 10 using MCS. The fact that City 10 produced lower absolute error than other cities indicates the importance of the local environment where a station locates. City 10 is located in the zone of "clean source". Therefore, the low variability of AQI concentrations makes it easier to predict than other cities. Nevertheless, the best R_2 is achieved for both 1-day and 2-day predictions at City 8, which indicates that the relative measure can objectively evaluate the prediction model in different backgrounds. The AQI prediction results showed a good R_2. Moreover, ARIMA produced better results than the APGRA model and ANFIS approach for both 1-day and 2-day prediction in terms of processing time. The result shows that the best algorithm among the three is the APGRA model in terms of RMSE and MAE as it achieves an average R_2 of 0.772, RMSE of 1.891, MAE of 1.642and time of 7.57. The basic ARIMA average R_2 of 0.571, RMSE of 3.22, MAE of 2.874, and time of 5.97. The ANFIS benchmark average R_2 of 0.48, RMSE of 3.7, MAE of 3.33 time of 10.37.

Table 5. Comparison between the three AQI prediction algorithms in Malaysia dataset.

	AQI 1-day advance prediction										
	Metric	**City 1**	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**	**City 7**	**City 8**	**City 9**	**City 10**
ARIMA	R_2	0.83	0.24	0.33	0.47	0.87	0.34	0.48	0.92	0.89	0.34
	RMSE	3.84	2.83	4.99	2.66	3.78	2.77	2.18	3.11	4.71	1.33
	MAE	3.45	2.69	4.40	2.32	3.40	2.26	2.04	2.78	4.30	1.10
	Time	7.20	6.30	5.20	6.30	6.50	4.70	3.80	6.60	6.90	6.20
MCS	R_2	0.91	0.75	0.50	0.64	0.89	0.88	0.73	0.94	0.92	0.56
	RMSE	2.15	1.08	2.05	1.24	3.08	1.92	1.02	2.47	3.10	0.80
	MAE	1.90	1.03	1.89	1.06	2.60	1.60	0.88	2.11	2.70	0.65
	Time	8.40	8.30	6.20	8.30	7.50	6.70	6.80	7.50	7.90	8.10
ANFIS	R_2	0.8	0.6	0.4	0.3	0.8	0.1	0.1	0.9	0.7	0.1
	RMSE	4.3	3.3	5	3.2	4.5	4	2.3	3.4	5	2
	MAE	4	3.1	4.4	2.7	4.1	3.5	2	3	4.8	1.7
	Time	10.40	12.30	9	11	9	9	9	11	12	11
	AQI 2-day advance prediction										
	Metric	**City 1**	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**	**City 7**	**City 8**	**City 9**	**City10**
ARIMA	R_2	0.12	0.20	0.30	0.39	0.01	0.02	0.25	0.82	0.80	0.10
	RMSE	19.40	9.20	5.90	22.30	14.80	8.10	3.60	6.76	12.60	16.20
	MAE	15.30	7.20	5.31	16.80	12.00	6.90	3.36	3.80	10.50	11.38
	Time	9.20	7.40	7.20	7.30	7.50	6.70	6.80	8.80	8.30	7.30
MCS	R_2	0.20	0.50	0.40	0.40	0.89	0.08	0.75	0.90	0.88	0.49
	RMSE	10.90	5.36	4.52	11.30	4.47	2.50	1.72	4.14	9.00	4.60
	MAE	7.96	3.75	3.77	8.51	3.54	1.95	1.42	2.66	7.00	2.80
	Time	9.80	9.40	9.80	9.50	9.20	8.70	8.30	9.80	9.30	9.30
ANFIS	R_2	0.1	0.4	0.6	0.4	0.1	0.1	0.3	0.9	0.8	0.2
	RMSE	19	9.3	3.9	22	14	8	2.7	3.2	12	16.4
	MAE	15.6	7.4	3.3	17	12	7	2.3	2.9	10	11.7
	Time	13	15	14	15	14	13	14	15	15	14

Based on Table 6, the best prediction in the China case study is achieved for 1-day in City 1, and 2-day is at City 3. The prediction values of the first-day yield lower error than the second day. This can be explained by the theory of error accumulation since the forecasting error for one day in advance is brought into the next day's prediction. The ARIMA-based MCS results better than the direct approach ARIMA and ANFIS models for both 1-day and 2-day forecasts. This confirms the ability of the Monte-Carlo simulation to accurately reproduce the sample, which boosts the predictive power of ARIMA. The results show that the ARIMA-based MCS algorithm plays an important role in obtaining good prediction results with approximately 47 % improvement on RMSE. As measured by RMSE and MAE, the best prediction is achieved for 1-day using the APGRA model from the aspect of absolute error. Nevertheless, the best R_2 is achieved for both 1-day and 2-day predictions, which indicates that the relative measure can evaluate the prediction model in different backgrounds.

Subsequently, the APGRA model provides the best solution among the three in terms of RMSE, MAE, and R_2. The model achieves an average R_2 of 0.852, RMSE of 7.509, MAE of 5.909, and time of 3.34. The basic ARIMA average R_2 of 0.718, RMSE of 14.14, MAE of

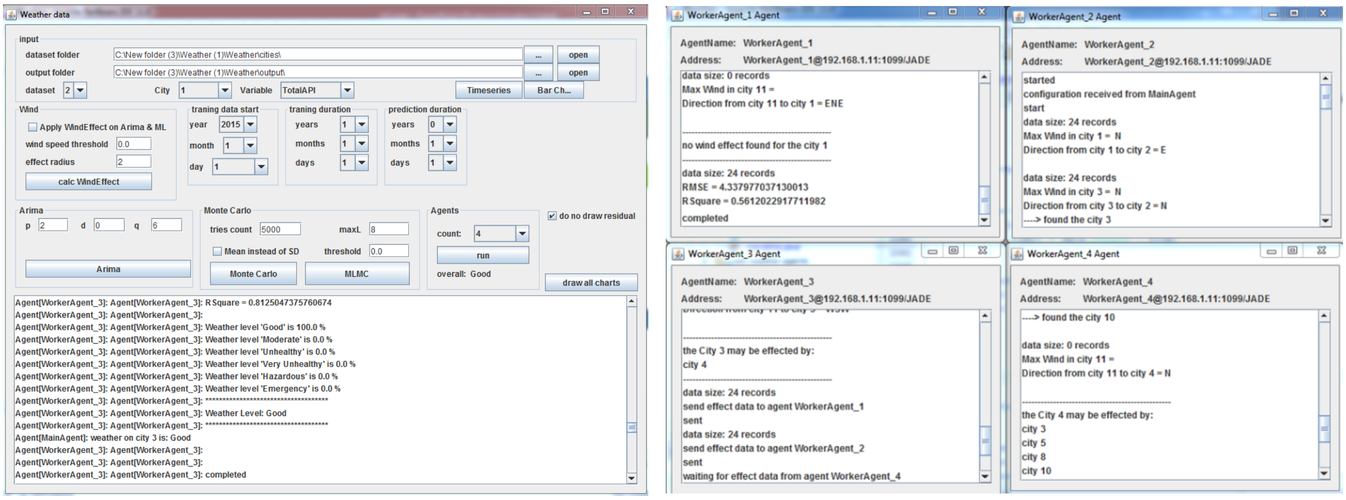
11.86, and time of 2.65. The ANFIS benchmark model achieves an average R_2 of 0.615, RMSE of 13.426, MAE of 11.4146, and time of 7.7.

Table 6. Comparison between the three AQI prediction algorithms in China dataset.

	AQI 1-day advance prediction										
	Metric	**City 1**	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**	**City 7**	**City 8**	**City 9**	**City 10**
ARIMA	R_2	0.97	0.94	0.55	0.89	0.58	0.45	0.66	0.40	0.81	0.93
	RMSE	4.45	9.78	6.14	10.0	5.46	20.2	21.4	30.4	19.1	14.24
	MAE	3.32	8.39	5.38	8.54	4.88	15.1	16.7	26.3	17.4	12.46
	Time	2.70	3.00	2.50	2.80	2.30	2.70	2.70	2.80	2.90	2.10
MCS	R_2	0.98	0.95	0.83	0.91	0.73	0.81	0.93	0.50	0.91	0.97
	RMSE	2.70	4.64	2.95	5.70	2.30	16.7	9.80	11.0	12.9	6.40
	MAE	1.80	3.75	2.40	4.54	2.00	11.7	7.10	9.80	10.7	5.30
	Time	3.10	3.90	3.10	3.90	3.10	3.10	3.90	3.10	3.10	3.10
ANFIS	R_2	0.8	0.8	0.40	0.79	0.63	0.25	0.69	0.14	0.73	0.92
	RMSE	4.43	10.3	6.09	8.99	6.12	13.3	19.7	30.0	20.5	14.7
	MAE	3.22	8.6	5.1	7.9	5.41	10.8	15.9	26	18.3	12.9
	Time	7	8	7	8	8	7	8	9	8	7
	AQI 2-day advance prediction										
	Metric	**City 1**	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**	**City 7**	**City 8**	**City 9**	**City 10**
ARIMA	R_2	0.90	0.86	0.69	0.64	0.89	0.74	0.52	0.30	0.58	0.817
	RMSE	15.8	22.1	11.1	15.0	13.2	30.3	25.2	38.7	21.6	18.98
	MAE	9.91	17.2	8.42	11.1	9.38	25.6	21.5	30.3	20.1	16.25
	Time	3.20	4.10	4.10	4.10	4.30	3.20	3.20	3.20	4.30	4.30
MCS	R_2	0.94	0.87	0.76	0.75	0.96	0.85	0.82	0.40	0.81	0.90
	RMSE	11.6	15.7	7.20	11.2	9.70	21.9	15.0	27.0	13.7	11.50
	MAE	6.80	10.6	4.70	7.40	6.00	17.9	11.8	17.0	12.1	9.00
	Time	4.10	4.90	4.80	4.90	4.80	4.80	4.80	4.10	4.80	4.10
ANFIS	R_2	0.9	0.8	0.5	0.6	0.8	0.7	0.5	0.5	0.5	0.8
	RMSE	16	22	11	15	13.5	30	24	39	23	19
	MAE	10	17	8	11	9.5	26	21	30	20	16
	Time	9	9	8	8	9.20	9.30	9.20	9.40	8.90	8.60

4.2. Results of Global Air Pollution Risk Assessment Model

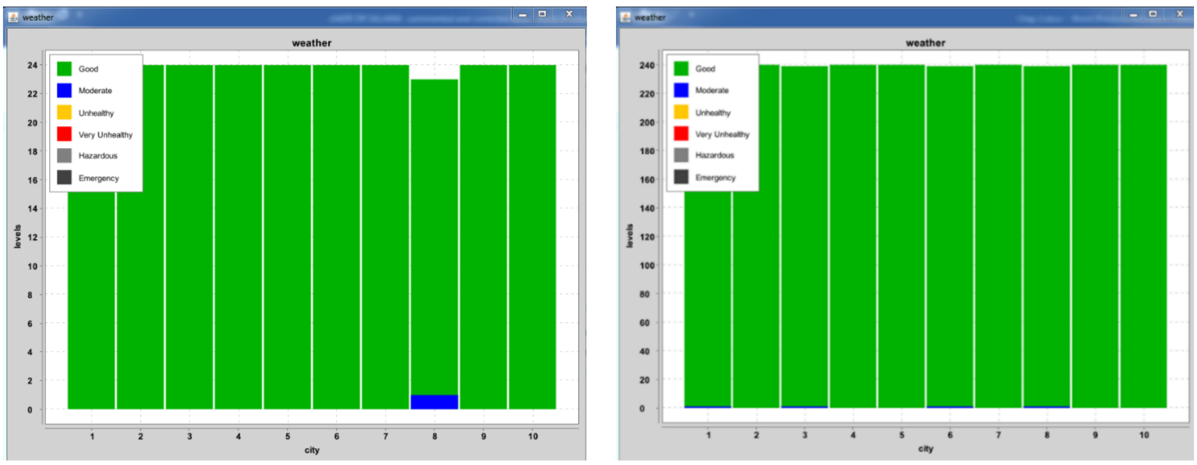
The APGRA model is the best prediction algorithm because it scores the lowest error from the earlier analysis. However, the AQI's prediction poses a distributed problem because the pollution risks are distributed in multiple places of cities. In turn, the process of risk as a prediction model is important to aggregate the risks from various local stations as represented in cities in both Malaysia and China case studies. Moreover, finding the risk level from AQI of various parameters is one of this research's main objectives. As a result, this paper proposes a Global Air Pollution Risk Assessment (APGRA) model based on a collaborative multi-agent architecture where each city is modeled as a collaborative agent. This model, therefore, aggregates risk input from multiple agents residing in distributed cities to produce a single global air pollution prediction value. An APGRA model is implemented in a system for testing and evaluation to achieve this. Figure 9a shows the mediator agent in APGRA with a dynamic selection of the number of agents to work with. The mediator agent in the APGRA model is responsible for decision-making depending on the information that comes from multiple agents (cities). The information that comes from the multi-agents includes the configuration of the main agent, the direction of the wind, the threshold for the error of prediction, the amount of data, the information that sent between each agent to others, the main city understudy, and other cities affected by the main city. The main agent aggregates all risk information calculated by the multi-agents depending on the individual AQI level in each city. The parameter of air pollutions in this research depends on the case study. Malaysia, for example, does not measure $PM_{2.5}$ in all cities. The mediator agent in APGRA monitors and visualizes the current range of data as determined by the user, along with options to filter data by year, month, and day. APGRA also provided the option to choose the model for calculating local AQI prediction, such as the ARIMA-based MCS. This model relies on cooperative multi-agents to produce the global assessment of air pollution. Figure 9b shows the cooperation process among the single-agent representing a single city. Assessment becomes more complex and challenging. Subsequently, a global air pollution risk



(**a**) The GUI of the mediator agent (**b**) Cooperation among agents

Figure 9. The implementation results of the APGRA model.

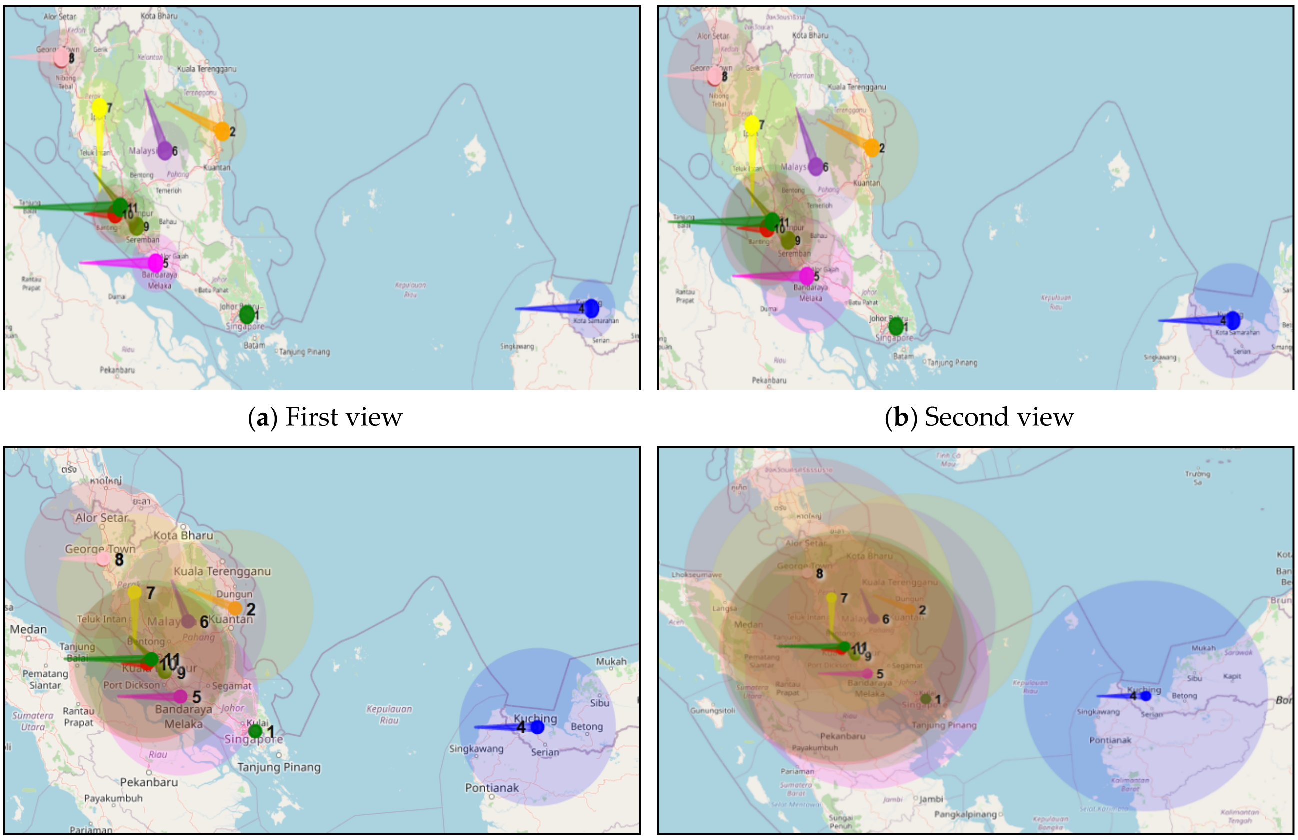
Figure 10a,b show the global AQI values from all cities in the different case studies (Malaysia and China). Note that the proposed ARIMA-based MCS algorithm conducted local prediction of AQI levels. Subsequently, the APGRA model under this multi-agent architecture produces a singular global air pollution risk prediction value.



(**a**) Global AQI level in Malaysia dataset (**b**) Global AQI levels in China dataset

Figure 10. The AQI levels of different cities.

The APGRA model depends on the wind data, which are the wind speed and wind directions, to produce the global prediction value. This is important to illustrate the dynamic changes of air pollution risks for a specific city concerning other cities. Wind direction determines the direction of pollution, while wind speed determines the zone pollution. There exists a direct correlation between the pollution zone and wind speed. When the wind speed increases, the zone of pollutions increases as well, and this relationship can be shown by the APGRA. Figure 11 explains how the APGRA model work relies on wind data in several cities in Malaysia with a ring of pollution zone between 0.5 km to 5 km.



(**a**) First view (**b**) Second view

(**c**) Third view (**d**) Fourth view

Figure 11. The radius of pollution.

Based on Figure 11, if the wind speed is normal at 5km/h, the zone pollution will be 0.5km (refer to Figure 11a). If the wind speed is between 10 to 20km/h, the zone pollution will be 1km (refer to Figure 11b). If the wind speed is between 20 to 30km/h, the zone pollution will be 2km (refer to Figure 11c). Finally, if the wind speed is more than 30km/h, the zone pollution will be 5km (refer to Figure 11d).

Table 7 explains the results of the global air pollution risk assessment model that depends on wind data to calculate the interaction pollution among cities for two cases study (Malaysia and China). The proposed model calculates the air pollution level for each city then calculates the global air pollution level for all cities. The wind speed response on the area of pollution zones, whereas the area of pollution zone increases proportionally to wind speed. The wind direction responds to the direction of pollution, which might also affect the other cities. The obtained results of the APGRA model are compared with the actual data of the ten cities in both case studies. The estimated risk levels of the 10 cities in both case studies have a full match. This indicates that the APGRA model correctly predicts the global risk levels. This is attributed to the ability of APGRA to assess the global value of the AQI and takes into account dynamic parameters reading at multiple, distributed local stations. Therefore, the prediction model becomes aware of the global aspect regarding possible interaction between individual cities, which improves the proposed APGRA.

Table 7. Sample of global air pollution risk levels.

AQI 1-day advance assessment of risk level in Malaysia dataset										
Results	**City 1**	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**	**City 7**	**City 8**	**City 9**	**City 10**
Risk level	good	good	moderate	good	good	good	good	moderate	moderate	good
Affected by	none	none	7,9	none	3,8	none	none	7,9	none	9
Effect on	none	none	5	none	none	none	3	5	7,3,10	none
Effect zone	1	2	0.5	1	1	0.5	2	2	2	1
R_2	0.9	0.6	0.5	0.8	0.7	0.6	0.6	0.7	0.8	0.6
RMSE	1.06	2.6	2.98	1.7	1.8	2	2.2	3.1	1.3	1.9
MAE	0.7	1.2	1.7	1.2	1.37	1.5	1.55	1.78	0.9	1.3
Time	3.20	5.30	4.20	3.30	3.20	4.90	4.80	3.20	3.90	5.10
AQI 1-day advance assessment of risk level in China dataset										
Results	**City 1**	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**	**City 7**	**City 8**	**City 9**	**City 10**
Risk level	moderate	good	moderate	good	good	moderate	good	moderate	good	good
Affected by	none	9	4	none	none	none	none	4	none	4
Effect on	none	none	none	8,3,10	none	none	none	none	2	none
Effect zone	0.5	0.5	1	1	0.5	0.5	0.5	2	0.5	0.5
R_2	0.6	0.6	0.8	0.8	0.9	0.5	0.7	0.7	0.8	0.9
RMSE	2.8	3.2	4.5	5	5.2	5.1	4.8	5.8	6.8	4
MAE	2.1	2.4	3.2	3.75	4.1	3.6	3.4	4	4.5	2.8
Time	3.70	3.20	3.10	3.20	3.20	3.20	3.20	3.70	3.90	3.20

Table 7 appointed the RMSE, MAE, R_2, and processing time. The APGRA model has a great matching with actual data that achieved low errors, good processing time, good ability, and flexibility, representing R_2. At the same time, the Malaysia case study's average performance metrics are R_2 of 0.7, RMSE of 2.064, MAE of 1.32, and time of 4.11. Likewise, China's case study average performance metrics are R_2 of 0.73, RMSE of 4.72, MAE of 3.385, and a time of 3.36.

5. Conclusions

This paper proposed a new air pollution global risk assessment (APGRA) model for predicting spatial correlations AQI risk assessment to address these issues. The APGRA model incorporates autoregressive integrated moving average (ARIMA), Monte-Carlo simulation (MCS), and collaborative multi-agent system (MAS), and prediction algorithm for reducing AQI prediction error and time. The proposed APGRA model is evaluated based on Malaysia and China's two real-world air quality datasets. The APGRA model improved the average Root Mean Squared Error (RMSE) by 41%, Mean and Absolute Error (MAE) by 47.10% when compared with the conventional ARIMA model and ANFIS model. The accuracy level of the ARIMA-based MCS algorithm was stably higher than that of ARIMA. Especially, RMSE and MAE of ARIMA-based MCS algorithm generate significant improvements, which helps to estimate the variation trend of the AQI concentrations. The proposed model provided the variance prediction in addition to AQI concentrations prediction expressing more information on the forecasting target. We analyzed and explained the AQI concentrations prediction with the ARIMA-based MCS algorithm, and the simulation results proved outstanding adaptively to the proposed model. The ARIMA-based MCS algorithm can be applied to other AQI forecasting if the model's appropriate input variables are selected. Some issues still need further investigation. This includes $PM_{2.5}$ emission data for the study area that were not available. The $PM_{2.5}$

with complex components exhibits a high correlation with the other AQIs. It is rather remarkable that the influence of $PM_{2.5}$ on AQI should be considered in the forecasting system. The $PM_{2.5}$ with complex components is another issue that exhibits a high correlation with the other AQIs. Therefore, it is rather remarkable that the influence of different AQIs on $PM_{2.5}$ should be considered in the forecasting system. As we mentioned before, the APGRA model solves the global pollution interaction between cities depending on a local ARIMA-based MCS algorithm developed in this paper and some additional parameters such as wind speed and wind direction. The issue with the ARIMA-based MCS algorithm is the cost of the simulation, resulting from the need to apply significant values of p and q that led to consuming processing time.

Author Contributions: Conceptualization, M.H.H. and S.A.M.; methodology, M.H.H., S.A.M. and A.M.; software, M.H.H. and M.A.J.; validation, M.H.H., S.A.M., and M.Z.S.; formal analysis, M.H.H., and B.A.S.A.; investigation, M.H.H., and S.A.M.; resources, M.Z.S., F.S. and A.E.M.E.; data curation, M.H.H.; writing—original draft preparation, M.H.H. S.A.M, and A.M.; writing—review and editing, M.H.H. S.A.M., and M.A.J.; visualization, M.H.H., F.S. and B.A.S.A.; supervision, S.A.M. and A.M.; project administration, M.Z.S., S.A.M. and A.M.; funding acquisition, M.Z.S., S.A.M., and A.E.M.E. All authors have read and agreed to the published version of the manuscript.

Funding: The authors express appreciation to the Malaysia Ministry of Higher Education (MoHE). This research was supported by the Fundamental Research Grant Scheme (FRGS/1/2019/ICT04/UTHM/03/1) grant vot number K209.

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement: The used dataset of this research is available online and has a proper citation within the paper contents.

Acknowledgments: The authors would like to thank the Department Of Environment (DOE) for providing the required data and assistance to this work. The authors also would like to thank the Center of Intelligent and Autonomous Systems (CIAS) at the Faculty of Computer Science and Information Technology (FSKTM), Universiti Tun Hussein Onn Malaysia (UTHM) for supporting this work.

Conflicts of Interest: The authors declare that they have no conflict of interest to be addressed related to this work.

References

[1] Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., & Chen, S. X. (2017). Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *473*(2205), 20170457.

[2] Liu, Y. S., Cao, Y., Hou, J. J., Zhang, J. T., Yang, Y. O., & Liu, L. C. (2020). Identifying common paths of CO2 and air pollutants emissions in China. *Journal of Cleaner Production*, *256*, 120599.

[3] Li, J., & Tartarini, F. (2020). Changes in air quality during the COVID-19 lockdown in Singapore and associations with human mobility trends. *Aerosol and Air Quality Research*, *20*(8), 1748-1758.

[4] Nyoni, T., & Mutongi, C. (2019). Modeling and forecasting carbon dioxide emissions in China using Autoregressive Integrated Moving Average (ARIMA) models, *Munich Pers. RePEc Arch.*, 2019.

[5] Bakhtavar, E., Hosseini, S., Hewage, K., & Sadiq, R. (2021). Air Pollution Risk Assessment Using a Hybrid Fuzzy Intelligent Probability-Based Approach: Mine Blasting Dust Impacts. *Natural Resources Research*, *30*(3), 2607-2627..

[6] Siwek, K., & Osowski, S. (2016). Data mining methods for prediction of air pollution. *International Journal of Applied Mathematics and Computer Science*, *26*(2), 467-478.

[7] Yang, Z., & Wang, J. (2017). A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction. *Environmental research*, *158*, 105-117.

[8] Tong, R., Cheng, M., Zhang, L., Liu, M., Yang, X., Li, X., & Yin, W. (2018). The construction dust-induced occupational health risk using Monte-Carlo simulation. *Journal of cleaner production*, *184*, 598-608.

[9] Yang, H., & O'Connell, J. F. (2020). Short-term carbon emissions forecast for aviation industry in Shanghai. *Journal of Cleaner Production*, *275*, 122734.

[10] Song, C., & Fu, X. (2020). Research on different weight combination in air quality forecasting models. *Journal of Cleaner Production*, *261*, 121169.

[11] Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of PM2. 5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, *107*, 118-128.

[12] Prasad, K., Gorai, A. K., & Goyal, P. (2016). Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmospheric environment*, *128*, 246-262.

[13] Zio, E. (2016). Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety*, *152*, 137-150.

[14] Wang, P., Zhang, H., Qin, Z., & Zhang, G. (2017). A novel hybrid-Garch model based on ARIMA and SVM for PM2. 5 concentrations forecasting. *Atmospheric Pollution Research*, *8*(5), 850-860.

[15] Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H. (2020). Forecasting of COVID19 per regions using ARIMA models and polynomial functions. *Applied soft computing*, *96*, 106610.

[16] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief*, *29*, 105340.

[17] Westerlund, J., Urbain, J. P., & Bonilla, J. (2014). Application of air quality combination forecasting to Bogota. *Atmospheric Environment*, *89*, 22-28.

[18] Mannshardt, E., Benedict, K., Jenkins, S., Keating, M., Mintz, D., Stone, S., & Wayland, R. (2017). Analysis of short-term ozone and PM2. 5 measurements: Characteristics and relationships for air sensor messaging. *Journal of the Air & Waste Management Association*, *67*(4), 462-474.

[19] Qazi, A., Shamayleh, A., El-Sayegh, S., & Formaneck, S. (2021). Prioritizing risks in sustainable construction projects using a risk matrix-based Monte Carlo Simulation approach. *Sustainable Cities and Society*, *65*, 102576.

[20] Zhao, L., Ji, Y., Yao, J., Long, S., Li, D., & Yang, Y. (2017). Quantifying the fate and risk assessment of different antibiotics during wastewater treatment using a Monte Carlo simulation. *Journal of Cleaner Production*, *168*, 626-631.

[21] Gordy, M. B., & Juneja, S. (2010). Nested simulation in portfolio risk measurement. *Management Science*, *56*(10), 1833-1848.

[22] Mostafa, S. A., Ahmad, M. S., Annamalai, M., Ahmad, A., & Gunasekaran, S. S. (2013). A dynamically adjustable autonomic agent framework. In *Advances in information systems and technologies* (pp. 631-642). Springer, Berlin, Heidelberg.

[23] Hassan, M. H., Mostafa, S. A., Mustapha, A., Abd Wahab, M. H., & Nor, D. M. (2018, August). A survey of multi-agent system approach in risk assessment. In *2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR)* (pp. 1-6). IEEE.

[24] Mostafa, S. A., Ahmad, M. S., Ahmad, A., & Annamalai, M. (2013, December). Formulating situation awareness for multi-agent systems. In *2013 International Conference on Advanced Computer Science Applications and Technologies* (pp. 48-53). IEEE.

[25] Kashinath, S. A., Mostafa, S. A., Mustapha, A., Mahdin, H., Lim, D., Mahmoud, M. A., ... & Yang, T. J. (2021). Review of Data Fusion Methods for Real-Time and Multi-Sensor Traffic Flow Analysis. *IEEE Access*.

[26] Mostafa, S. A., Hazeem, A. A., Khaleefahand, S. H., Mustapha, A., & Darman, R. (2018, November). A collaborative multi-agent system for oil palm pests and diseases global situation awareness. In *Proceedings of the Future Technologies Conference* (pp. 763-775). Springer, Cham.

[27] Mostafa, S. A., Mustapha, A., Gunasekaran, S. S., Ahmad, M. S., Mohammed, M. A., Parwekar, P., & Kadry, S. (2021). An agent architecture for autonomous UAV flight control in object classification and recognition missions. *Soft Computing*, 1-14.

[28] Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., Mahmoud, M. A., Al-Rimy, B. A. S., ... & Marks, A. (2021). An Adaptive Protection of Flooding Attacks Model for Complex Network Environments. *Security and Communication Networks*, *2021*.
[29] Mostafa, S. A., Gunasekaran, S. S., Ahmad, M. S., Ahmad, A., Annamalai, M., & Mustapha, A. (2014, June). Defining tasks and actions complexity-levels via their deliberation intensity measures in the layered adjustable autonomy model. In *2014 International Conference on Intelligent Environments* (pp. 52-55). IEEE.
[30] Mostafa, S. A., Mustapha, A., Mohammed, M. A., Ahmad, M. S., & Mahmoud, M. A. (2018). A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application. *International journal of medical informatics*, *112*, 173-184.
[31] Bellifemine, F. L., Caire, G., & Greenwood, D. (2007). *Developing multi-agent systems with JADE* (Vol. 7). John Wiley & Sons.