Workshop on Visualisation in Environmental Sciences (EnvirVis) (2021)
S. Dutta and K. Feige and K. Rink and D. Zeckzer (Editors)

Air Quality Temporal Analyser: Interactive temporal analyses with visual predictive assessments

S. Harbola †[1,2] S. Koch[1] & T. Ertl[1] & V. Coors[2]

[1]University of Stuttgart, Germany
[2]University of Applied sciences, Stuttgart, Germany

Abstract

This work presents Air Quality Temporal Analyser (AQTA), an interactive system to support visual analyses of air quality data with time. This interactive AQTA allows the seamless integration of predictive models and detailed patterns analyses. While previous approaches lack predictive air quality options, this interface provides back-and-forth dialogue with the designed multiple Machine Learning (ML) models and comparisons for better visual predictive assessments. These models can be dynamically selected in real-time, and the user could visually compare the results in different time conditions for chosen parameters. Moreover, AQTA provides data selection, display, visualisation of past, present, future (prediction) and correlation structure among air parameters, highlighting the predictive models effectiveness. AQTA has been evaluated using Stuttgart (Germany) city air pollutants, i.e., Particular Matter (PM) PM_{10}, Nitrogen Oxide (NO), Nitrogen Dioxide (NO_2), and Ozone (O_3) and meteorological parameters like pressure, temperature, wind and humidity. The initial findings are presented that corroborate the city's COVID lockdown (year 2020) conditions and sudden changes in patterns, highlighting the improvements in the pollutants concentrations. AQTA, thus, successfully discovers temporal relationships among complex air quality data, interactively in different time frames, by harnessing the user's knowledge of factors influencing the past, present and future behavior, with the aid of ML models. Further, this study also reveals that the decrease in the concentration of one pollutant does not ensure that the surrounding air quality would improve as other factors are interrelated.

Keywords: Time series, environmental visualisation, user interfaces, visual prediction, machine learning, meteorological data, city planning, visual analytics, air pollutants

1. Introduction and Related work

Temporal datasets are essential and measured across almost all the domains including environmental, healthcare, scientific and financial. Visual analytics (VA) supported with Scientific or Information Visualisation (Sci-Vis or Info-Vis) techniques are in demand and also crucial for analysing these time-series datasets patterns [WSHC11]. The characteristics of data, its size, multi-dimensionality, and distribution contribute to make situation assessment one of the most demanding tasks, both for the user and the platform [TC05, IHK*17]. Visual data exploration often follows Shniderman's mantra [Shn96]. The work related to visual prediction, time series visualisation and temporal analytical approaches which matches the keywords of the proposed work were explored. Recent techniques [KPN16, BZS*16] on visualising the time series data supported with mathematical and statistical metrics enable the user to build reasoning about the considered temporal datasets interactively. Visualisation techniques, highlighting the anomalies and underlying trends correlations, through an undirected interactive search [SSK*16] were developed. Moreover, time series visualisation were explored by providing examples of simple charts including stacked graphs, index charts, horizon graphs for visualising time-series datasets. The representations of time series data become more contextual with the support of cluster, calendar-based and, spiral visualisations [WAM11]. More detailed and aggregated representations, using multi-resolution layouts for handling over-plotting in large time series datasets were developed [HJM*11, Fu11]. Moreover they also reviewed the data mining method for classification, pattern exploration, segmentation and representation of time series data. Hochheiser and Shneiderman, invented dynamic query tools for time series dataset interactive explorations with user demand detailing [HS04]. Chronolenses were proposed for time series data visual exploration and correlation analysis [ZCB11b, ZCB11a]. Anomaly detection for modelling multiple time series [CM05], clustering and classification [Lia05] techniques to identify the similarity of data patterns among time series dataset using weighted dynamic time warping [JJO11], distance metrics and agglomerative clustering have been developed

† Corresponding author

DOI: 10.2312/envirvis.20211083

[HC20]. Inter parameters relationships definition rules are revolutionised by Hetland and Saetrom [HS05] with rule mining concept for time series database. The scientific temporal data visualisations are frequently used in support of interactive visual analytics and are well-accepted within the disciplines [AA03, NOV*20]. Moreover, for understanding the temporal datasets and its trends, predicting future and patterns remains a very challenging task with a few interactive visual models and user explorations behaviour support. Predicting the time series data using statistical methodologies like regression analysis, and computational machine learning approaches like neural networks, multilayer perceptron, fuzzy logic and self organising maps have been successfully applied for the existing studies [Lor86, Gui07, VSP11, JHZ11]. Visual prediction approaches in the act of visually predicting a time-series variable by observing the predictions from a computational model, shown alongside with the time-series representations for social media and financial datasets were designed. [HJM*11, LKT*14, BZS*16]. Furthermore, interaction techniques with engaging the user in an efficient dialogue in the contribution by people and computers to solve the task together *i.e.*, mixed-initiative interaction techniques have also been proposed [Hor10, Hor07, KLTH10, EFN12]. Data driven forecasting in visual predictions for time series dataset visualisation with highlighting the sequence and pattern in support of approaches to explore correlations in multivariate spatiotemporal data have been proposed by [HJM*11, MMJ*12].

However, the increased usage of the environmental monitoring system and sensors installation on a day-to-day basis has provided more information in monitoring the current environmental conditions. Sensor networking advancement with quality and quantity for air parameters, has given rise to an increase in techniques and methodologies supporting temporal data interactive visualisation analyses [Har06, Bog08]. Moreover, there exists a gap between the environment as observed and its digital representation in the user govern time frame for temporal data interactive analysis. Visualisation of meteorological and pollution data history and context plays an essential role in visual data mining, especially in exploring the large and complex datasets. and environmental conditions. Including the context and historical information in the visualisation could improve user understanding of the environmental dataset exploration process and enhancing the re-usability of mining and managing techniques and parameters analysis to achieve the required insides. Although, traditional approaches cannot fully support the visual exploration of future trends in complex multivariate time series datasets such as weather, and healthcare, mainly due to their lack of consideration of inter-variable relationships (*e.g.*, if PM_{10} increases, NO_2 decreases). Exploring these relationships through "what-if" questions (*e.g.*, what if PM_{10} increases?) could help the user to better judge the future environmental conditions than blindly trusting computational models that lack contextual information.

Thus, there is still a gap the user likely needs to bridge for comprehending the situation. The proposed work overcomes these dissociations by proposing an Air Quality Temporal Analyser (AQTA), an interactive system-user interface for visual prediction of multivariate time series through deep learning models as well as interactive visualisation techniques for air quality parameters. Following are the contributions of the current work, (i) interactive temporal visualisation of historical, present and future data through various charts, to support the user in the interpretation of the data that may be useful for further stages of the mining process such as cluster identifications, important feature and pattern detection, (ii) predicting the air quality standards for the desired temporal frame (dynamic) with five designed deep learning models, thereby highlighting the respective model's success and failure for inference data along with supporting the arguments with easy graphical support and suggesting best option to choose, (iii) visual preservation of context and historical information in all these user interactions. These contributions combine together to form three phases (1-3 shown in Figure 1) of interactive AQTA with back-and-forth dialogues between user and AQTA. This interactive dialogue between the AQTA and the user continues until the user finds sufficient information to come to a conclusion. This would infer smart decisions for air quality planning, which in turn would help in proficient management and development of the city's resources. AQTA is validated for Stuttgart, Germany as a used case study. The remaining paper is organised as follows: system and datasets used and proposed approaches are discussed in section 2 and section 3, respectively, section 4 discusses the results, followed by conclusion in section 5.

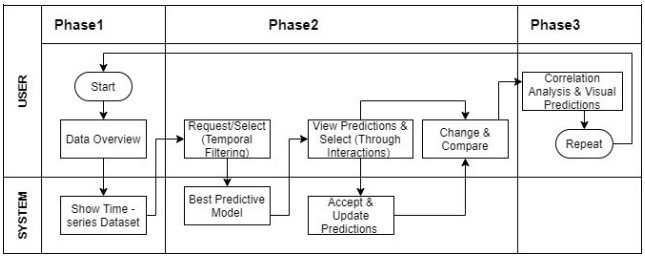
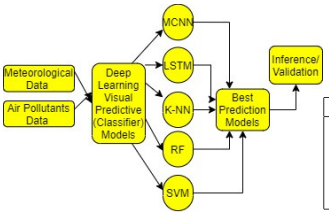


Figure 1: *AQTA workflow maintains an interactive dialogue between user and the system for visual prediction and in-depth analysis including correlation.*



(a)

Class	Lower Range	Upper Range
calm	$min(parameter)$	$\mu - k_1\sigma$
light	$\mu - k_1\sigma$	$\mu - k_2\sigma$
mild	$\mu - k_2\sigma$	$\mu + k_2\sigma$
strong	$\mu + k_2\sigma$	$\mu + k_1\sigma$
strongest	$\mu + k_1\sigma$	$max(parameter)$

(b)

Figure 2: *(a) Predictive models analysis flowchart, and (b) Various classes designed ranges.*

2. System and Datasets

The temporal air quality datasets that are used and analysed in this study ("luftdaten selber messen" http://www.luftdaten.info) provide city sensors measurements at several locations in Stuttgart. Historical dataset from 2016 to 2020 measured at total 8 city centre

locations with the wind (speed and directions), temperature, pressure and humidity along with NO, NO_2, O_3, PM_{10}, with temporal information attached in a 30-minute time interval http://www.stadtklima-stuttgart.de). The areas dataset were organised separately into individual years for each parameter, using time information with past data first, followed by current data. This helps to perform an in-depth study of air parameters. AQTA is implemented as a web-based application using D3.js, Streamlit, Keras library [Cho17] with TensorFlow in the backend in Python and executed on Intel® Core ™ i7- 4770 CPU @3.40 GHz having four cores. Each designed air quality predictor (ML) model was executed separately for selected time series data for predicting the (dominating) class magnitudes and analysing air nature. The following section 3 explains the proposed system architecture comprising of models, graphs and database at system side and interactive visualisation interface at the user side. Result in section 4 analyses and validates the outcome of sensor located at Stuttgart's city centre, and similar results were obtained for the other sensors as well. AQTA web deployment along with detailed figures are available in GitHub http://www.github.com/shharbola/EnvirVis_AQTA.

3. Approaches

The proposed work combines different visual analysis of air quality parameters, integrated into AQTA platform. Figure 2(a) provides an overview of the workflow and highlights the motivation behind the comparative analysis of different models. Here, the time series air quality datasets comprises of pollutants *i.e.*, PM_{10}, Nitrogen Oxide (NO), and Nitrogen Di-oxide (NO_2), and Ozone (O_3) and meteorological parameters like wind (speed and direction), pressure, temperature and humidity, with temporal resolution T and T_w ($w \rightarrow 1$ to m) denotes value of the selected parameter (above mentioned) at time w, where 1 and m are the first and last values in the dataset, respectively.

3.1. Air quality predictor

Multiple samples are designed using the dataset for training and testing the proposed prediction algorithms. A sample consists of a feature vector as an input with a corresponding output class. $Real_{V_b}$ (a scalar) consecutive values of considered parameter, from T_w to $T_{w+Real_{V_b}}$ form a feature vector of dimension $Real_{V_b} \times 1$ which is the input of the sample. $Real_{V_f}$ (a scalar) successive values of selected parameter after the last value in the input *i.e.*, $T_{w+Real_{V_b}}$, are used to define the sample's output class. Mean (μ), and standard deviation (σ) of the parameter of the entire dataset are calculated. Various class boundaries are designed using μ and σ as shown in Figure 2(b). Among $Real_{V_f}$, count of values occurring in each class in Figure 2(b) is noted, and the class that has a maximum count *i.e.*, dominant, is assigned to the sample. Similarly, multiple samples based on the selected parameter are created by taking $Real_{V_b}$ values in the corresponding input from T_w to $T_{w+Real_{V_b}}$ by varying w from 1 to m - $Real_{V_f}$, at an increment of 1. The outputs of these samples are designed as discussed above. Likewise, samples based on other parameters (each independently) are created for each dynamically selected parameter as discussed above. Thus, at this stage, for $Real_{V_b}$ values in the input from T_w to $T_{w+Real_{V_b}}$, there would be nine sets of samples, based on PM_{10}, NO, NO_2, O_3 and wind (speed and direction), pressure, temperature, and humidity. Here in this analysis the size of $Real_{V_b}$ and $Real_{V_f}$ are kept equal with four user options, (a) 12 representing 6 hours as temporal resolution of considered dataset is 30 minutes, (b) 24 representing 12 hours, (c) 36 representing 18 hours, and (d) 48 representing 24 hours. These conditions ensured comprehensive and accurate analysis of the data with respect to independent and different user selections.

The first proposed air quality predictor ML model is Multi-Convolutional Neural Network (MCNN) that has five single CNN, say (CNN_1, CNN_2, CNN_3, CNN_4, CNN_5). Each of these CNN_i ($i \rightarrow 1$ to 5) has its own input layer, three consecutive 1D convolutional layers and last convolutional layer of each CNN connects to a common fully connected layer which is followed by another fully connected layer and an output layer. The architecture is explained in detail in [HC19b]. The output layer is a softmax layer [SMKLm15], with the number of neurons same as the number of the classes. There are five classes in the present study as shown in Figure 2(b). The MCNN is trained and tested separately for the prediction of dominant temporal nature of the selected parameter (PM_{10}, NO, NO_2, O_3, wind, pressure, temperature and humidity). Therefore, for an inference sample, the MCNN could predict the air quality parameters classes separately and visually highlight time series data recurring motif.

The developed Long Short-Term Memory (LSTM) model (second) is a special kind of Recurrent Neural Networks (RNN) capable of learning long term dependencies with a chain-like structure. This has an input layer, four neural layers ($NL1$, $NL2$, $NL3$, $NL4$), *i.e.*, three sigmoid layers supported with two tanh layers and an output layer. The architecture is explained in detail in [HC19a]. The input layer is One Dimensional (1D) of the size of $Real_{V_b}$. The output layer is a softmax layer, having the number of neurons the same as the number of the classes *i.e.*, five.

The third proposed time series prediction model uses K-Nearest Neighbors (KNN) which is a supervised classification algorithm. KNN based method makes predictions on the fly by calculating the similarity between an input observation and values in the dataset, with respect to time. Here K value is decided empirically and kept fixed in all parameter analysis. The designed SVM based predictive fourth model classifies the data by finding the best hyper-plane that separates all data points of one class from those of the other class. The best hyper-plane signifies the one with the largest margin between the classes. Similarly the last proposed Random Forest (RF) based model uses a decision tree as a decision support tool for classification. When the RF is given a training sample, it formulates a set of rules which are used to perform predictions. Moreover, RF uses sufficient decision trees, to ensure the classifier does not overfit the model. The advantage of the RF as a classifier is that it can handle missing values, and the classifier could be modeled for categorical values. Therefore LSTM, MCNN, SVM, K-NN and RF (five deep learning models) are used to predict meteorological and pollution parameters separately.

During training, the sample's feature vector of dimension $RealV_b \times 1$, forms the input of the designed models, while the sample's output class forms the output of these models. The objective behind using a variety of supervised prediction models is to provide a

possible option of selecting models based on best (compare) accuracy with respect to the various date, time, and parameters conditions. The previous paragraphs discusses the various proposed models of temporal air quality prediction. Besides prediction, the detailed analysis of historical air quality parameters are also performed in this work. Temporal filtering along with Pearson correlation method analysis help to derive the relationships with highlighting interconnections between the meteorological and air pollutants. The user could select the parameters over the desired time frame and compare the patterns interactively in AQTA, thus, making the analysis more diverse and refined.

3.2. Visual interaction design

AQTA besides being air quality predictor, also provides tooltiping, brushing and linking for maintaining the transparency and combining different visualisation methods between user-computer dialogue efficiently and preserving the working memory of the user during interactions [Shn96, Hor10]. Figure 1 provides an overview of AQTA workflow (phases 1-3), with highlighting the system-user interfaces of visual predictions comparative analysis. **System:** consists of historical air quality temporal database, trained ML models, structure of various graphs and charts, and accepts user queries. **User:** interacts with this system in various ways. The user selects, inspects and views the states of the parameters with past present and future (predictions) information. The user could also choose among different ML models with analysing the performance of each selected model (MCNN, LSTM, RF, K-NN and SVM) in terms of total accuracy and difference metrics incorporated with the interactive display through various graphs and charts. The user could change the time step allowing for a different prediction duration, and compare the results with the time series dataset and the outcome of each model. This allows the user to decide which prediction algorithms are the best and provides sufficient information to make a decision.

The system works as per user desires with additional information of revelling the correlation among the selected parameters answering "what if" questions of nine parameters dependency with each other within selected time frame. Furthermore, detailed analysis of the patterns in the dataset in the three phases of AQTA are carried out using additional charts, heat-map, time histogram, that are explained below.

3.2.1. Inspecting data history visualisation (Phase 1)

The phase 1 visualisation of AQTA uses time series stack chart with calendar heat-map to provide interaction with the air quality datasets visually. The inspecting overview shows the overall patterns for multiple parameters of interest selected from air quality parameters list available in the interface (Figure 3). The time dataset overview design contains horizon graphs. The effective discrimination option in horizon graphs makes it more desirable [JME10]. This is accompanied with stacked chart to provide a detailed time series data inspection of parameters magnitudes with calendar heat-map view option in order to compare the trends among air quality parameters based on the months during a year. The user could select each year and then even explore in detail for each day with 30 minutes (here the sensors' data temporal resolution) for air quality parameters temporal analyses. This phase provides a detailed understanding of the air quality data history and preset with highlighting the patterns which are actually present and measured by the sensors (here no smoothing or data cleaning performed *i.e.*, real original datasets).

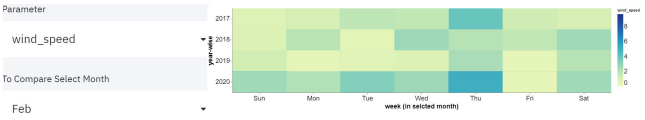


Figure 3: *Data inspection (a part of Phase 1) week-wise over the years for selected parameter*

3.2.2. Prediction visualisation (Phase 2)

The phase 2 visualisation consists of square-time charts, and temporal circle mark chart coupled with histogram highlighting the predicted value (Figure 4). Predicted outcome with respect to time frame (6hr, 12hr, 18hr, 24hr) choices are displayed with the help of square-time chart with tooltip highlighting the class assigned and color encoding makes it easy to distinguish in detail the classes with respect to each predicted value in the time frame. Each class is assigned dynamic color encoding according to predicted class range. The comparison and preforming the analysis of predicted versus the actual values is shown with the help of time series square-time graph with the color encoding representing the difference of actual and predicted (Figure 4 (a)), that occur in the range *i.e.*, (-4, -3, -2, -1, 0, 1, 2, 3, 4) calculated by assigning 1 = calm, 2 = light, 3 = mild, 4 = strong, 5 = strongest as in Figure 2(b). Tooltiping are also added to this representation to make it easier for user to understand the actual and predicted values along with their respective difference in the time frame. In order to provide a detailed comparison and more easy interaction by double encoding, mark circle with integrated histogram graph is designed (Figure 4 (b)). Here the circle radius is governed by the class ranges and color according to the assigned class with respect to time. The histogram shows the count of the records estimated or predicted each day and binned according to the assigned class patterns. Both actual (Figure 4 (b) left) and predicted (Figure 4 (b) right) values are compared in this interface with clearly highlighting the pattern of meteorological and pollution parameters in time frame, which helps user to make advance and comparative estimation of the environment and its pattern with model's success information.

3.2.3. Correlation visualisation (Phase 3)

The phase 3 visualisation of AQTA is implemented as an air quality parameters' correlation structure detailed analyses. The time series exploratory analysis of meteorological and pollution parameters also requires supporting, identifying the correlation and how these parameters are controlling and effecting each other's nature in interaction. Pearson correlation method is used for analysing these relationships among the parameters. The correlation graph (Figure 5) explores the correlation structure of the meteorological and pollution parameters dataset using two connected subplots: an interactive correlation heat-map (Figure 5, left image) and a 2D

Figure 4: *Phase 2 inference (comparing actual versus predicted output)*

histogram showing the density of values (Figure 5, right image). Clicking on a cell in the correlation heat-map shows correlation coefficient value for that particular cell, (shown in pink highlight in Figure 5, left image), where parameter1 (on X-axis) and parameter2 (on Y-axis) represent associated air quality parameters of the selected cell on X-axis and Y-axis. Selection linked binning on the fly is performed for the selected cell generating a 2D histogram (detailed bins) between parameter1 and parameter2, with the advantage of highlighting overlapping values leading to a higher density of values (frequency) in the darker color bins, making it clear to the observer that there are more similar range values in the selection. Thus, the correlation heat-map shows the parameters and value ranges (all) associated with that particular cell and the corresponding data in the 2D histogram. This enables user to quickly see the pattern in correlations using the heat-map, and allows to zoom in on the dataset underlying those correlations in the 2D histogram.

All the graphs and subsections integrating with visual predictions help the user to provide a more clarity of the time series and environmental conditions. AQTA (1-3 phases) tries to bring all the information together that could be derived from air quality parameters condition in the considered city (Stuttgart, Germany) in order to derive the time series patterns and correlations.

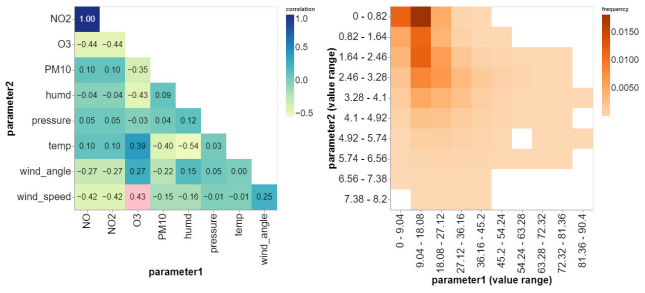


Figure 5: *Temporal visual correlation analysis using correlation heat-map (left) linked with 2D histogram (right).*

4. Results: Use case

AQTA was used for visual analysis of Stuttgart's COVID lockdown air quality situation (in year 2020) to facilitate visual exploration of prediction models outcome and reality conditions that occurred during this sudden pandemic. AQTA results were compared with real world measurements to support analyser inference outcomes and interaction in subsections 4.1 and 4.2 respectively, followed by 4.3 for discussion.

4.1. Inference

Several samples, each having input and corresponding output, were created as described in section 3.1. Values of k_1 and k_2 (Figure 2(b)) were empirically taken as 0.80 and 0.50 respectively (same for all parameters), so that a sufficient number of samples occur in each class. Moreover, Synthetic Minority Oversampling Technique (SMOTE) was used to do up-sampling of the classes having less number of samples. Total samples for a given year were randomly split into training and testing with 35% of the total samples as the testing samples. The designed models were trained and tested on these samples. When samples were prepared for the inference (validation) for year (2020), the samples were created similar to model training and testing phase (as mentioned in section 3.1). The models had never seen the dataset which were used in inference therefore the pattern and class predicted dynamically, were predicted based on the designed models achieved accuracy. The obtained accuracies for five designed models are approximately between 90% to 95% [HC19b, HC19a].

These classification outputs are shown in the supplemental material on Github. These outcomes represent AQTA phase 2 of Figure 1. The classes square chart uses diverse color coding to highlight the model's predicted classes assigned with respect to selected timeframe of (6hrs, 12hrs, 18hrs and 24hrs) future prediction. Class specific color coding provides more distinguish representation irrespective of the selected timeframe (small or large), that helps in quick user understanding and assessment of lot of predicted information at one go. The graphs (Figure 4) comparing the actual and predicted results difference, highlight the success and failure of the selected predicted models in the selected timeframe (as shown in Figure 4 (c)). The difference (actual - predicted) of the selected model classification outcome is shown with square chart (Figure 4 (a)), here sequential single-hue schemes (blues) encoding shows the difference values *i.e.*, (-4 (light blue), < 0 = model success, < 4 (dark blue)) attached with tooltip information. Another graph, circle mark charts, represents actual and predicted classes separately (Figure 4 (b)). In these circle mark charts, the radius encodes the ranges of the assigned classes (calm < light < mild < strong < strongest). Integrated histograms at the bottom of these graphs denote each day's (overall) predicted and actual record of classification outcome, with colors and conditional selection are linked with the above circle mark charts. The together build selection between circle mark charts and corresponding histograms, gives the user option to filter the outcome as per the requirements. This helps in detailed analysis of the actual and predicted classification outcomes and model's success-failure overview, in each selected time frame and arriving at a conclusion to pick the best model.

4.2. Interactions

The data inspection *i.e.*, phase 1 is used to provide user the freedom to visually analyse all the historical data (available in database) with graphs by temporal queries. The options available for user are either to compare all the years with respect to month, day for the desired parameter or to explore in depth each year independently with querying based on week (Figure 3), date and time with overall option palette available to change, update the selection, process new one, save the results and return. Therefore, users can use controls, which provide zooming, selection, tooltiping and saving outcome (image format) options, to view the models classification distributions at different timeframe. The user can use several available options on the screen, to get back to the default views, change the selection, reset the main phase view, the phase details views, or all the views. The output of phase 3 (Figure 5) is the temporal correlation analysis of meteorological and pollution parameters with yearly selection option available to user. Creating one cohesive interactive plot using correlation heat-map linked with 2D histogram (showing the density of values), helps to answer queries related to parameters interrelationships and how their dependency fluctuates in time with comparison option. Binning on the fly, with user parameters selections and displaying the correlation (heat-map with yellow green blue sequential multi-hue schemes) and frequency (2D histogram with oranges sequential single-hue schemes) allow to have details of an individual correlation as shown in Figure 5. The interactive chart enables to quickly distinguish pattern in correlations using the heat-map, and allows to zoom in on the meteorological and air pollution data underlying those correlations in the 2D histogram. This indicates that the correlation leans heavily on the tail of the data and vice versa. Visual correlation analysis queries would help to understand the data and temporal dependencies more clearly with interactive charts which makes understanding very easy and less time taking, making environmental planning more comprehensive and interesting.

4.3. Discussion

ML based prediction algorithms used in AQTA are based on [HC18,HC19a,HC19b]. These approaches with good prediction results are applied in phase 2 of AQTA to achieve an interactive visual prediction, and pattern analysis platform. This aides user to understand easily the inside of data, complexity of the parameters, trends and details, and air quality impact. AQTA focuses on integrating and linking the simple charts representation to discover complex air quality parameters interactively in various timeframes, with options to have a visual data overview (history and present in phase 1), predicting future with model success, failure comparison (phase 2), and a correlation structure of their interrelationships (phase 3). The proposed framework is successfully implemented for the Stuttgart city central location. However, it could be applied to any number of sensors for any given location (area) with some ML tuning and training of the respective datasets. The air pollution from predominantly non-traffic-related pollutants (*e.g.*, dust deposits) has decreased significantly in recent years. The traffic-related pollutants (*e.g.*,NO, NO_2, PM_{10}, O_3) remain at a high level in the city [sta21]. The city's air quality is controlled and not deteriorating further, due to the strong monitoring and control measures by the state governments, city's policymakers and increase environmental awareness among people. But still the AQTA analysis shows that during summer and autumn of the year 2019, PM_{10} trends are alike as in the previous years 2017 to 2018 with a few reductions. Furthermore, there is depletion in PM_{10} concentrations during the summer and autumn of the year 2020 probably due to the strict lockdown and movements restrictions. However, the decrease in the concentration of one parameter and increase in others' does not ensure that the overall air quality is improved for *e.g.*, PM_{10} is observed reduced in Oct 2020, while O_3 concentration is higher. The reasons behind these relationships and trends are more evident with the correlation structure integrated with this analysis, highlighting that PM_{10} is positively correlated with NO and NO_2, while negatively correlated with O_3. Similarly, NO and NO_2 are negatively correlated with O_3. Thus AQTA allows the actual data to convey itself and used to upgrade the user's hypothesis with the best understanding.

PM_{10} concentrations were predicted for 22-29 March 2020 when there were strict COVID lockdown restrictions during these days. The MLS models predicted the air quality parameters with good accuracy during these conditions. Thus, the proposed AQTA framework has good potential for visual analytics along with prediction in different conditions. While analysing parameters from 20 April 2020 to 1 May 2020 and taking 6 hours time in future, LSTM model predicted the NO and NO_2 concentration to be strongest on days 22, 28, and 29 April 2020. It was between calm and mild for rest of the days. The comparison of predicted values with the real data showed approximately 95% accuracy of the model. When the relaxation in the lockdown was given one month later, at that time also, model gave good results. Further, the model also predicted the pressure range from calm to mild on 21 April, strongest class on 23 April, and calm on 28 April. The predicted and actual values matched for 23, 28 April but there was a mismatch for 21 April. As pressure and PM_{10} are positively correlated their spikes and patterns show similarities with their effects over the days which also cross validates the correlation with reference to data range trends. PM_{10} concentrations were observed to be higher specially on Fridays (apart from other weekdays) in February of year 2020, as well as on Fridays and Saturdays in April and on Saturdays in May of 2020. Similar trends were observed for NO and NO_2 concentrations during the same timeframes (correlation discussed above). Usually these trends were also similar to previous years, weeks, and days patterns, with only fluctuation in concentrations ranges (calm to strongest). These patterns could be because people might be using public transports and shared cabs on working days. Transportation emits more than half of NO and NO_2 in the air. During the weekend, people travel to their homes, have family outings besides other important travel plans, thereby contributing to higher pollutants concentrations. PM_{10} concentrations predicted using LSTM for new year eve's on 31.12.2020 to 01.01.2021 from 12:00 pm to 12:00 pm was between mild to strong, to strongest (12:00 am to 2:00 am) and then mild ranges which matches with the published report of Stadtklima Stuttgart on 01.01.2021 on PM_{10} concentrations in Stuttgart on New Year's Eve 2020-2021 [ss21].

The day wise analysis of wind speed was also performed for February 2020. It was observed that on Thursday the magnitude of wind speed occurred mainly between strong and strongest classes. This trend was also noticed for the previous years as well. Simi-

larly also observed for temperature on Monday (strongest), Saturday (strong), Sunday (strong), and Friday (mild) classes patterns occurred within the selected timeframes. Such analysis helps in proper utilisation and planning of renewable sources like wind. Moreover, wind speed, pressure and temperature are positively correlated to each other, while wind direction and speed are in positive interrelationship with O_3. Therefore, it was also observed that the local wind could often develop that do not cause high magnitude winds, but play an essential role in local ventilation of the city areas and determines the spread of air pollutants (as found from correlations insight discussed above). The Stuttgart region is one of the areas with the lowest rainfall in Germany, mainly due to the lee location (Black Forest, Swabian Alb) and precipitation conditions playing a significant role in cleaning the atmosphere through the wet deposition. Moreover, the humidity of an area is highly controlled by the wind directions as they are positively correlated. In year 2020 April, May and August months, the measured humidity is lower (on average) while in comparison to same months in 2019. These trends also matches with the changing wind directions occurred during the same year and months patterns. Due to the high temperatures trends in recent years, combined with the existing humidity patterns, Stuttgart is one of the areas with increased heat load (approx. 30 days), with occasional cold fillips and this infer seems coherent with the state climate published annual report [ss21]. Hence, AQTA provides an add-on to the existing literature in terms of air quality multiple time series datasets dynamic visual predictions along with its detailed analyses comparisons and validation with reality.

5. Conclusions and Future Work

Visual analytics of air quality parameters using interactive graphs as well as to understand the data and temporal dependencies with interactive AQTA tool has been the objective of this work. AQTA provides a quick facts-crosscheck supporting the present alarming air quality situation in the city and requirement of probable control measures. The interactive platform for visual prediction of air quality parameters would help to plan the future with more green policies. Designed platform in this work could be further improved with the ensemble of advance visualisation approaches. The future focus for the authors would be to improve the visual analysis and utilising more advance deep learning models. Meanwhile, the devised work has the potential for creating environmental awareness among humankind, and moreover, provide a foreknowledge for better city planning.

Acknowledgement The Stuttgart dataset is downloaded from the city of Stuttgart DEPUC website and luftdaten.info-Feinstaub selber messen.

References

[AA03] ANDRIENKO N., ANDRIENKO G.: Coordinated views for informed spatial decision making. In *In: Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization - CMV 2003* (2003). 2

[Bog08] BOGUE R.: Environmental sensing: strategies, technologies and applications. *Earth-Science Reviews 28* (2008), 275–282. 2

[BZS*16] BADAM K., ZHAO J., SEN S., ELMQVIST N., DAVID E.: Timefork: Interactive prediction of time series. In *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), vol. 123, p. 52–55. 1, 2

[Cho17] CHOLLET F.: *Deep Learning with Python*. Manning Publications Co., Greenwich, CT, USA ©2017, 2017. 3

[CM05] CHAN P., MOHONEY M.: Semantic interaction for visual text analytics. In *proceedings of the IEEE conference on data mining* (2005), vol. 123, p. 52–55. 1

[EFN12] ENDERT A., FIAUX P., NORTH C.: Semantic interaction for visual text analytics. In *CHI '12: Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems* (2012), vol. 123, p. 52–55. 2

[Fu11] FU T. C.: A review on time series dagta mining. *Engineering Applications of Artificial Intelligence* (2011). 1

[Gui07] GUILHERME A. B.: Time series prediction with the self-organising map. *Springer, In perspective of Neural-symbol integration* (2007), 135–158. 2

[Har06] HART J. K.: Environmental sensor networks: a revolution in the earth system science? *Earth-Science Reviews 78* (2006). 2

[HC18] HARBOLA S., COORS V.: Geo-visualisation and visual analytics for smart cities: A survey. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42* (2018). 6

[HC19a] HARBOLA S., COORS V.: Comparative analysis of lstm, rf and svm architectures for predicting wind nature for smart city planning. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci* (2019), 65–70. 3, 5, 6

[HC19b] HARBOLA S., COORS V.: One dimensional convolutional neural network architectures for wind prediction. *Energy Conversion and Management 195* (2019), 70–75. 3, 5, 6

[HC20] HARBOLA S., COORS V.: Seasonality deduction platform: For pm 10, pm 2.5, no, no 2 and o 3 in relationship with wind speed and humidity. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci* (2020), 65–70. 2

[HJM*11] HAO C. M., JANETZKO H., MITTELSTAEDT S., HILL W., DAYAL U., KEIM D., MARWAH M., SHARMA R.: A visual analytic approach for peak preserving predictions of large seasonal time series. In *In computer graphics forum* (2011), vol. 123, p. 52–55. 1, 2

[Hor07] HORVITZ E.: Reflections on challenges and promises of mixed intiative interaction. In *In AI Magazine* (2007), vol. 53, pp. 59–67. 2

[Hor10] HORVITZ E.: Principles of mixed initiative user interfaces. In *In Proceedings of the ACM Conference on Human Factors in Computing Systems* (2010), vol. 53, pp. 59–67. 2, 4

[HS04] HOCHHEISER H., SHNEIDERMAN B.: Dynamic query tools for time series datasets:timebox widgets for interactive exploration. *Inforamtion Visualisation* (2004), 1–18. 1

[HS05] HETLAND M., SAETROM P.: Evolutionary rule mining in time series databases. *Machine learning* (2005), 135–158. 2

[IHK*17] ISENBERG P., HEIMERL F., KOCH S., ISENBERG T., XU P., STOLPER C. D., SEDLMAIR M., CHEN J., MOLLER T., STASKO J.: vispubdata.org: A metadata collection about ieee visualization (vis) publications. *EEE TRANSACTIONS OF VISUALIZATION AND COMPUTER GRAPHICS* (2017). 1

[JHZ11] JOHAN B., HUINA M., ZENG X.: Twitter mood predicts the stock market. *Journal of Computational Science* (2011), 135–158. 2

[JJO11] JEONG Y., JEONG M., OMITAAOMU O.: Weighted dynamic time warping for time series classification. *Pattern recognition* (2011). 1

[JME10] JAVED W., MCDONNEL B., ELMQVIST N.: Graphical perception of multiple time series. *IEEE TRANSACTIONS OF VISUALIZATION AND COMPUTER GRAPHICS* (2010). 4

[KLTH10] KAPOOR A., LEE B., TAN D., HORVITZ E.: Performance and preferences: interactive refinement of machine learning procedures. In *In Proceedings of the AAAI Conference on AI* (2010), vol. 53, pp. 59–67. 2

[KPN16] KRAUSE J., PERER A., NG K.: Integrating with predictions: Visual inspection of black-box machine learning models. *Association for Computing Machinery ACM* (2016). 1

[Lia05] LIAO T. W.: Clustering of time series data-a survey. *Pattern recognition* (2005). 1

[LKT*14] LU Y., KRUEGER R., THOM D., KOCH S., ERTL T., MACIEJEWSKI R.: Integrating predictive analytics and social media. In *IEEE symposium on visual analytics science and technology* (2014), vol. 123, p. 52–55. 2

[Lor86] LORENC A. C.: Analysis methods for numerical weather prediction. *Royal meteorological society quarterly journal* (1986). 2

[MMJ*12] MALIK A., MACIEJEWSKI R., JANG Y., HUANG W., ELMQUIVST N., EBERT D.: A correlative analysis process in a visual analytic environment. In *IEEE symposium on information visualisation* (2012), vol. 123, p. 52–55. 2

[NOV*20] NAVARRA C., OPACH T., VROTSOU K., JOLING A., WILK1 J., NESET T.: Visual exploration of climate-related volunteered geographic information. In *Workshop on Visualisation in Environmental Sciences (EnvirVis) (2020)* (2020), vol. 123, p. 52–55. 2

[Shn96] SHNEIDERMAN B.: The eye have it: a task by data type taxonomy for information visualization. In *In: Proceedings of IEEE Visual Languages, College Park, Maryland, pp. 336–343 (1996)* (1996), vol. 9, pp. 3–12. 1, 4

[SMKLm15] SU H., MAJI S., KALOGERAKIS E., LEARNED-MILLER E.: Multi-view convolutional neural networks for 3d shape recognition. In *International Conference on Computer Vision, Proceedings of the IEEE* (2015), pp. 945–953. doi:10.1109/ICCV.2015.114. 3

[ss21] STADTKLIMA STUTTGART.DE: stadtklima stuttgart news, 2021. Last accessed 10 Feb 2021. URL: www.stadtklima-stuttgart.de/index.php?info_news. 6, 7

[SSK*16] SACHA D., SENARATNE H., KWON C., ELLIS G., KEIM A.: The role of uncertainty, awareness and trust in visual analytics. *IEEE TRANSACTIONS OF VISUALIZATION AND COMPUTER GRAPHICS* (2016). 1

[sta21] STADTENTWICKLUNG.BERLIN.DE: Traffic-related air pollution along streets 2015 (edition 2017), 2021. Last accessed 10 Feb 2021. URL: https://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/ede311_01.htm. 6

[TC05] THOMAS J. J., COOK K. A.: Illuminating the path: the research and development agenda for visual analytics. *IEEE 54* (2005). 1

[VSP11] VENUGOPAL K. R., SRINIVASA G. K., PATNAIK L. M.: Fuzzy based neuro genetic algorithm for stock market prediction. In *In soft computing for data mining applications* (2011), vol. 123, p. 52–55. 2

[WAM11] WEBER M., ALEXA M., MULLER W.: Visualising time series on spirals. In *IEEE symposium on information visualisation* (2011), vol. 123, p. 52–55. 1

[WSHC11] WOLFGANG A., SILVIA M., HEIDRUN S., CHRISTIAN T.: *Visualization of Time-Oriented Data*. Springer, London, Springer-Verlag London Limited, 2011. 1

[ZCB11a] ZHAO J., CHEVALIER F., BALAKRISHNAN R.: Exploratory analysis of time-series with chronolenses. *IEEE TRANSACTIONS OF VISUALIZATION AND COMPUTER GRAPHICS* (2011). 1

[ZCB11b] ZHAO J., CHEVALIER F., BALAKRISHNAN R.: Kronominer: using multi-foci navigation for the visual exploration of time-series data. In *CHI '11: Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems* (2011), vol. 123, p. 52–55. 1