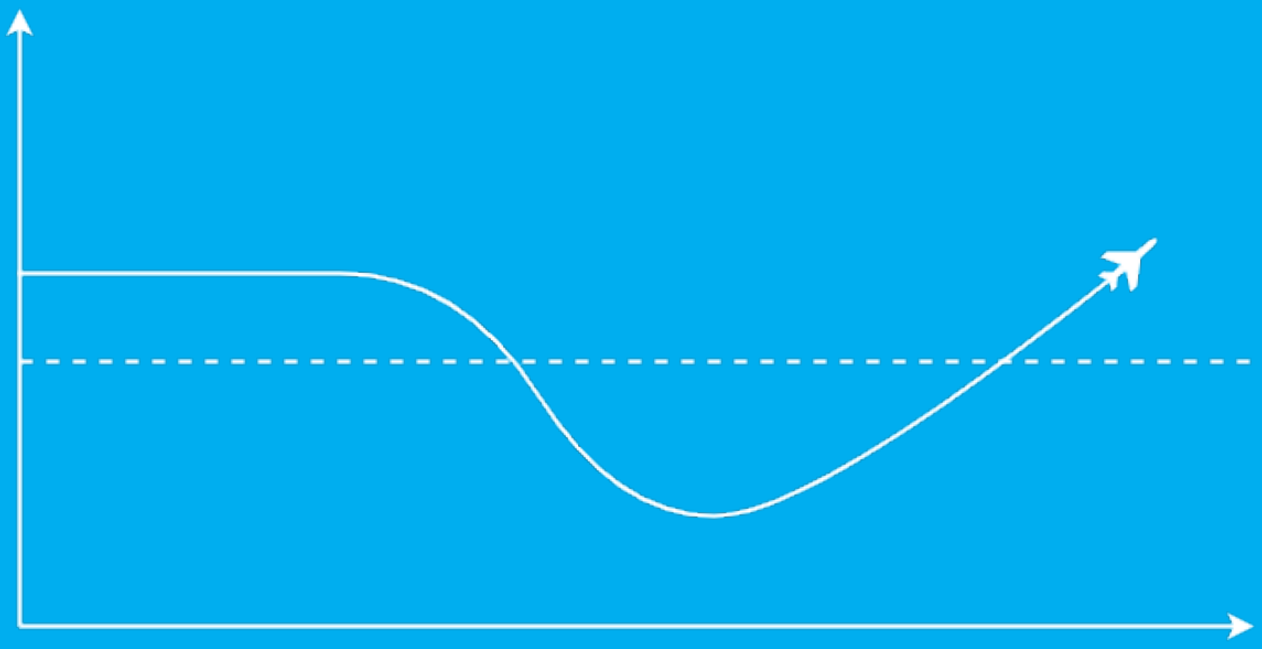
Technische Universiteit Delft, Airbus Defence & Space

Probabilistic Quantification of Airspace Resilience

Thesis report

Jakub Janowski



TU Delft

AIRBUS

Probabilistic Quantification of Airspace Resilience

Thesis report

by

Jakub Janowski

to obtain the degree of MSc in Aerospace Engineering
at the Delft University of Technology,
to be defended publicly on 16th January 2020 at 13:00

Student number:	4347900	
Supervisors:	Prof. dr. ir. JM. Hoekstra,	TU Delft, supervisor
	Dr. Ir. J. Ellerbroek,	TU Delft, co-supervisor
	Dr. H. Udluft,	Airbus, company supervisor
	Dr. G. Sirigu,	Airbus, company supervisor

TU Delft

AIRBUS

Preface

This thesis was made possible by the support and trust of Carina Schnitzenbaumer and Heiko Udluft from Airbus Airsense. They made it possible for me to conduct a thesis in co-operation with Airbus and utilise their resources and data.

Furthermore, I would like to sincerely thank Jacco Hoekstra, Joost Ellerbroek, and Giuseppe Sirigu for their guidance and critique that allowed me to develop the method proposed in this thesis.

Lastly, my whole education was made possible by my parents, who unconditionally supported me throughout this entire journey. Thank you.

Jakub Janowski
Delft, The Netherlands, January 2020

Report Outline

This report consists of two parts. The first is a Scientific Paper, which presents the final results of the thesis and the methodology used to arrive at them. The second part is the Preliminary Thesis report. This part contains the preliminary work performed up until the midterm review. This includes a literature study along with the first results.

The Scientific Paper is written in a format required by IEEE journal Transactions on Intelligent Transportation Systems. It consists of six sections: Introduction, Background Information, Methodology, Results, Conclusion, and Acknowledgements.

The Preliminary Thesis report contains sections on methodology and planning of the thesis, four chapters on the research relevant to the topic of this thesis, and, finally, the preliminary results. The Preliminary Thesis has been graded following a midterm thesis review.

The Scientific Paper Part is recommended to a reader interested only in the outcomes and concise methodology description. However, if a reader is interested in the full process taken to arrive at results, then reading the complete report is recommended. Chronologically, the Preliminary Thesis has been produced first. Hence, the Scientific Paper picks up on the work in the Preliminary Thesis and develops upon it.

Contents

I

Scientific Paper

Quantification of ATM System Resilience under uncertainty

Jakub Janowski
Supervisors: Jacco Hoekstra[1], Joost Ellerbroek[1], Giuseppe Sirigu[2], Heiko Udluft[2]

***Abstract*— The resilience of the Air Traffic Management (ATM) system to disturbances is required to maintain high operation performance. Before it can be improved, the resilience of the ATM system must be quantified. The measurement of resilience requires knowledge of a system reference state. This paper proposes a novel methodology to detect disruptions without a pre-specified reference state and to quantify airspace resilience to disturbances. The method utilises residual-based anomaly detection to model a reference state based on historical values and detect deviations from it. The method has been tested in assessing the resilience of arrival time (airspace state) to high winds (disturbance) in 9 airports worldwide for a year. The results have shown that the method is capable of detecting disruptions as well as airports experiencing high wind conditions tend to be more resilient to them.**

I. Introduction

The Air Traffic Management (ATM) system is being put consistently under more pressure. The number of Instrumental Flight Rule (IFR) operations is increasing year to year. In Europe, it is forecast to increase by more than 50% between the years 2017 and 2050 [5]. With this growth comes the need to optimise air traffic operations such that the limited resources are used to their full potential. An area for potential improvement is the ATM system resilience, which is understood as the ability to respond to disturbances such as weather or technical faults. If severe enough, disruptions due to, for example, the weather may negatively impact operations by increasing delays. In the US, 78% of delays occur in the near-airport phase of flight, while in Europe, this number is 48.4% [9]. In both cases, weather constitutes at least half of the delays.

Before the Resilience of a system can be improved, it must be defined as well as made quantifiable through a metric. An essential step in measurement of resilience is the detection of system disruptions understood as a deviation of a system state from a reference state. However, the reference state may be ambiguous and related to the overall context in which the system is operating in a given moment. Hence, the research problem addressed in this paper is the ability to define resilience when the nominal state, i.e., disruption-free state, is unknown.

The proposed solution to the research problem utilises a bayesian linear regression model that relates a Key Performance Indicator (KPI) of interest to variables that describe the airspace context. This method is de facto, a residual-based anomaly detection method known from the field of data mining [20]. This approach allows for detecting KPI deviations when the exact nominal range of KPI values is unknown. Furthermore, a Resilience metric is proposed, which is based on measuring absorptive and restorative capacities of airspace.

The method has been tested on deviations of arrival time KPI in response to high wind speeds. The results have shown that the method can identify deviations due to this disturbance. The method has been employed on a year of Automated Dependent Surveillance (ADS-B) data for nine major airports, and for each, a Resilience score has been computed.

The results show that there is a tendency for airports to increase the resilience of their operations to the wind if they face this disturbance frequently. This is done through the placement of runways in multiple directions, such as in the case of Amsterdam Schiphol. The complete method can be utilised to track the effectiveness of resilience improvement initiatives.

The structure of the paper is as follows. First resilience, ATM KPIs, and time-series analysis shall be introduced in section II. Next, the methodology shall be presented in section III. In section IV, the results of the deviation detection and resilience metric are presented. Lastly, the paper is concluded in section V.

II. Background information

Research of airspace response to disruption requires the employment of analytical tools and field domain knowledge covering airspace KPIs and resilience. To set a theoretical basis for later discussion, the following subsections shall introduce key concepts for this research.

A. Resilience

The understanding of resilience varies between domains [27]. As such, different properties are attributed to the resilience of a system. Three capacities can be attributed to resilience: absorptive, restorative, and adaptive [28]. The absorptive capacity allows the system to **withstand** perturbations without changing the system state in reaction to the perturbation. The restorative capacity is responsible for **return** to a system nominal condition following a perturbation. The adaptive capacity allows the system to modify its internal characteristics to **adjust** to a perturbation.

Table I presents the three capacities concerning system properties of robustness, dependability, and resilience. Resilience in this schematic is an overarching property encompassing robustness and dependability. However, while restorative and absorptive capacities can be decoupled from

[1] Delft University of Technology
[2] Airbus Defence & Space

each other, they cannot be decoupled from the adaptive capacity. This is because the adaptation of the system will change its restorative and absorptive capacities.

This is important to consider as measurements of the two capacities will be partially influenced by adaptability. In the context of ATM, this can be understood through airport operational adaptation to a change in wind direction. If the airport has favoured directions due to, for example, runway positioning or approach procedures, then the restorative and absorptive capacities concerning wind speed will be affected by the runway configuration utilised.

TABLE I: Resilience capacities [30].

Related system properties	Resilience capacities		
	Absorptive	Restorative	Adaptive
Robustness	+	-	-
Dependability	+	+	-
Resilience	+	+	+

Various metrics for measuring resilience have been proposed. One of the first measures of resilience originates from the field of earthquake engineering. In it, a value Q_t represents the quality of performance of a system, with 100% meaning that the system is functioning fully. As such, resilience is defined as in equation 1. R is the resilience value, while t_0 is the start time of disruption, and t_1 is the time of full recovery. The highest value resilience may take is 100; in this case, the disruption does not affect system performance quality.

$$R = \int_{t_0}^{t_1} (100 - Q(t))dt \tag{1}$$

Figure 1 presents the disruption and recovery process. It can be seen that at the moment t_0 the quality of service experiences a drop due to a disrupting event such as an earthquake. From that time on, the quality of service begins to recover to reestablish the full performance. This concept is known as the resilience triangle due to its shape.

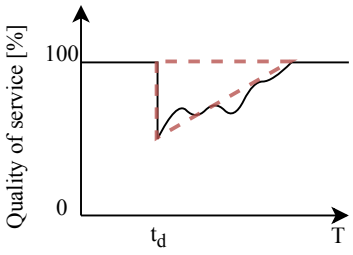


Fig. 1: Resilience triangle concept. t_d indicates time of disruption.

The method, as mentioned earlier, includes the assumption that only a single disruption occurs. Zobel has proposed an extension that overcomes this assumption. Et al., in the context of the resilience of cyberinfrastructure [34]. In the proposed method, the area above the quality curve $Q(t)$ is measured as a sum of within and between event losses. Within-event losses are those that occur due to slow-onset disruptions.

Within the field of ATM, resilience is known mainly through the mostly qualitative Resilience Engineering framework [31]. However, a quantitative measurement framework has been proposed, which allows for the measurement of resilience [32]. The performance-based approach proposed by Gluchshenko and Foerster [32] defines resilience based on the return of a system state to its reference level following a disruption. In this framework, resilience is considered high if the time of deviation is higher than the time of recovery.

The proposed resilience measurement methodology focuses on the restorative capacity from table I. In this definition, the time of deviation is the time between the beginning of disruption and the peak of deviation from the reference state. The time of recovery is then the time between the peak, as mentioned earlier, and the end of disruption. Figure 2 presents a system response to a disturbance. The time of deviation is much larger than the time of recovery; hence, the system has high resilience.

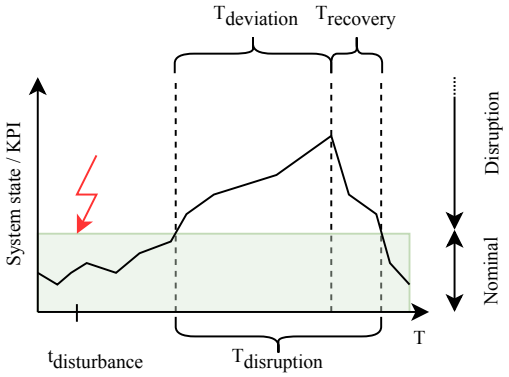


Fig. 2: High resilience according to performance based ATM resilience approach. (adapted from Gluchshenko et al. [31])

When discussing concepts such as deviation, it is essential to note that a deviation occurs with respect to a reference state. In the resilience triangle-based definitions, the reference state is 100% - maximum quality of service. In the definition by Gluchshenko et al., this is a set of performance indicator values. The determination of an adequate reference state is critical in measuring resilience. If the reference state is set incorrectly, deviations may not be detected or will be detected too frequently.

B. ATM system performance metrics

In the study of resilience, system states are observed for disruptions. Hence, there exists a need to quantify the state by some metric. Such metrics can then be utilised to assess the effects of system modifications. However, the ATM system performance can be considered from different perspectives.

The International Civil Aviation Association (ICAO) and the Civil Air Navigation Services Organisation (CANSO) created KPIs that are used to evaluate the performance of the ATM system. A 2009 ICAO report [8] has defined

eleven Key Performance Areas (KPAs). The KPAs that can be derived based on observed traffic include Capacity and Efficiency.

CANSO also proposes Efficiency and Capacity KPIs [7]. Some of the Capacity KPIs include landing and departure capacities as well as airport throughput, which is the sum of the two previous mentioned capacities. They can be considered as declared capacity or utilised capacity where the former is the theoretical maximum of the capacity while the latter is the actual usage of capacity.

The efficiency KPIs proposed by CANSO are divided per phase of flight and include Taxi, Departure, En-route, and Arrival KPIs. Efficiency KPIs can be utilised to, for example, compare the arrival time with a nominal arrival time or an ideal arrival time. Furthermore, distance flown can also be considered in comparison to an ideal value such as a great circle distance.

Another set of KPIs are predictability KPIs [7]. These are utilised to measure the variability of the ATM system performance. Those can include travel time variation, flight plan variation, or capacity variation.

Additionally, more sophisticated metrics are utilised in research to study airspace congestion. These can include flow-based, geometrical, or dynamical system metrics [6].

Overall, arrival time is a metric that is simple to measure as only entry and exit times are needed to obtain it. With distance, integration must be performed over the complete flight path. Furthermore, abnormal arrival time the airspace is directly related to delays, hence it serves as a relevant KPI.

C. Time-series analysis

One can notice that the metrics based on which resilience is measured in section II-A occur in the form of a time-series. A uni-variate time-series can be defined as an ordered set of scalar values $F_{uv} = \{f_1, f_2, ..., fn\}$ where each value f_i corresponds to the i^{th} observation of parameter f [23], [24].

However, complex systems, such as the ATM system, are described by multiple parameters simultaneously. In this case, the concept of the multi-variate time-series becomes relevant. Thus, a multi-variate time-series can be defined as an order set of vectors $F_{mv} = \{\vec{f_1}, \vec{f_2}, ..., \vec{f_n}\}$. For example, in the context of ATM, this vector could consist of aircraft count and landing rate.

The uni-variate time-series analysis has to consider temporal auto-correlations. In the case of multi-variate time-series, this becomes more complex as cross-correlation and inter variable causality can occur, e.g., landing rate affects aircraft count in the airspace. In this regard, utilising expert knowledge can be beneficial to predetermine the directions of causality as well as choosing suitable variables to include in the time-series for modeling of the airspace state.

1) Anomaly detection for deviation detection: An additional observation can be made based on the resilience concepts from section II-A, that inherently resilience deals with system response during **non-nominal events** caused by disturbances. In this regard, a parallel can be drawn to the outlier detection field of data mining. In outlier detection, the goal is to determine data that deviates significantly from other data [22], [21].

As an example, consider figure 3, which presents the concept of an anomaly in the context of time-series. Outside of the red dotted region, the time-series appears to be governed by nominal conditions. However, within the dotted box, the behaviour of the time-series suggests the presence of another, anomalous mechanisms.

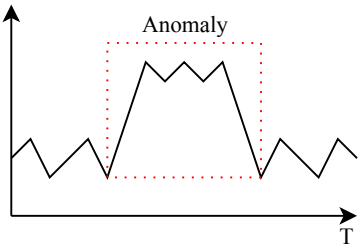


Fig. 3: Anomalous data example. The anomaly seems as if it was caused by another mechanism.

The nominal state of a complex system such as the ATM system can be non-obvious. This is because the norm of a particular state can be potentially stochastic as well as dependent upon other variables. In this regard, anomaly detection methods can be utilised to define the nominal state as well as to detect deviations from it. In the data mining terminology, the deviation would be the anomaly.

Overall, three outlier types can be identified [21]:

- Global: data object varies significantly from the rest of the dataset.
- Contextual: data object is different from the dataset considering its context.
- Collective: some data objects collectively, but not individually, differ from the dataset.

Contextual outliers are relevant to developing models of nominal airport states. In this regard, drops in performance due to poor weather or other disruptions can be considered outliers in the context of nominal operations, which must be defined concerning the disruptions being considered. I.e., the nominal state when studying disruptions due to weather is the state that is present when no poor weather occurs.

Another essential consideration in anomaly detection is whether a labelled set of anomalies is available. If a sufficient amount of anomalies have been previously captured, anomaly detection can be posed as a supervised learning problem. In this case, the algorithm would be taught to detect anomalies based on labelled data. However, labelled data is not easily obtainable for the resilience measurement problem at hand. Hence, an alternative is unsupervised learning.

In the unsupervised learning approach, the algorithm is trained without labelled data. In the context of resilience measurement, this means that no pre-labelled disruptions are needed. Hence, the unsupervised approach is preferred for the problem at hand.

The challenge involved with detecting contextual outliers is the definition of context. In some applications, the context

is easy to specify. For example, when measuring temperature, the context is the geographic location where it is measured. However, in other cases, it may be problematic as the context, if defined by multiple variables that relate to the variable being measured, is non-trivial.

2) Residual based anomaly detection: The solution to this challenge is to model the nominal state based on the contextual factors. The outlier detection itself is performed by measuring the variable deviation from the nominal state as determined by the variable's contextual factors. Consider a deterministic variable x and its corresponding context $\mathbf{c}$. Equation 2 presents the function that models the nominal state x_N based on $\mathbf{c}$.

$$x_N = f_N(\mathbf{c}) \tag{2}$$

Given the nominal state function, the residual between the function output and the variable x can be computed as the difference between the two values.

$$\epsilon = x - f_N(\mathbf{c}) \tag{3}$$

If the residual ϵ is above a set threshold α then the variable x is considered an outlier. Due to the utilisation of the model residual in outlier assessment, this method is referred to as residual-based anomaly detection [20].

While the method, as mentioned earlier, presented the nominal state as deterministic, it can as well be modelled as a probability distribution:

$$X_N \sim F(\mathbf{c}) \tag{4}$$

This approach is beneficial as it allows for the soft classification of an outlier, i.e., a probability of deviation is attributed to a data point instead of a hard classification. For this, the probability density function of the nominal state can be utilised:

$$f(x_n \mid \mathbf{c}) = F(\mathbf{c}) \tag{5}$$

The training of function $F(\mathbf{c})$ can be treated as a supervised machine learning problem with outputs being x values and c context values being the input. In this scenario, the model training must be performed on non-outlier data to teach it the nominal state.

Regression is utilised in order to perform residual-based anomaly detection. Regression is a method of mapping a relationship between dependent and independent variables. A variety of regression methods exist, some with high approximation power such as neural networks that have been utilised for sophisticated tasks such as computer vision and natural language processing [19].

However, the most common method of regression is linear regression [18]. This method remains popular due to its simplicity and, thus, also interpretability. The model parameters allow for investigation of how the independent variables affect the dependent variables.

Linear regression can be treated from the Bayesian perspective. Within this approach, the problem changes from determining the best line to determining the best probability distribution fit. This method has the added benefit of automatically obtaining information about the likelihoods of outputs values given the input values.

Equation 6 presents linear regression. y is a single output value, $\boldsymbol{w}$ are the model parameters, $\mathbf{x}$ are the input values and $\boldsymbol{\epsilon}$ is the error term.

$$y = \boldsymbol{w}^T\mathbf{x} + \boldsymbol{\epsilon} \tag{6}$$

Consider equation 7 which no longer presents the value y as a scalar value, but rather a Gaussian probability distribution with the mean of $\mu = \boldsymbol{w}^T x$ and variance of α. This representation is equivalent to equation 6 as the first term of that equation is deterministic whilst the error term ϵ is normally distributed with variance α and mean $\mu = 0$.

$$P(y|\boldsymbol{w}, \boldsymbol{x}, \alpha) = \mathcal{N}(\boldsymbol{w}^T\mathbf{x}, \alpha) \tag{7}$$

The Bayesian linear regression approach attempts to obtain a probability distribution of the model parameters w. This is done by introducing a prior probability distribution on the model parameters [18]. The choice of a Gaussian distribution as a prior of w will have the same result as the introduction of ℓ_2 regularisation in ridge regression. In practice, the regularisation is introduced to prevent model overfitting by penalizing large parameter values.

Bayesian regression as implemented in Scikit-Learn uses the Bayesian Ridge variant [16]. This method, introduces a spherical Gaussian prior on parameters $\mathbf{w}$ as in equation 8. This results in all the model parameters $\mathbf{w}$ having the same variance. Parameters λ from equation 8 and α from 7 have a Gamma distribution priors, each with two hyperparameters.

$$p(\mathbf{w}|\lambda) = \mathcal{N}(\mathbf{w}|0, \lambda^{-1}\mathbf{I}_p) \tag{8}$$

The proposed method of utilising Bayesian regression uses supervised learning to model the nominal state of arrival time based on the context. The model is trained to predict x from $\mathbf{c}$ for days where wind speed is below a strength threshold. Such days are considered nominal as the wind does not cause disruptions due to its low strength. The anomalies can be detected by comparing the model's output to the actual arrival time. If the difference between the two values is above a threshold, then an anomaly may be present.

3) K-means clustering: An alternative approach to outlier detection is to utilise clustering. Clustering is a process of grouping data objects such that objects similar to each other are in the same groups, while objects that differ are in different groups. In this context, three outlier types exist [21]. The first type of an outlier corresponds to an object that does not belong to any cluster. The second type of an outlier corresponds to objects for which the distance to the cluster centre is large. The last outlier type corresponds to objects belonging to small or sparse clusters.

One of the popular clustering algorithms due to its simplicity and interpretability is K-means clustering [18]. This approach aims to minimise the intra-cluster distance of points

belonging to a cluster and the clusters centre (mean). It does this iteratively by calculating cluster means based on the points belonging to a cluster and then reassigning the points to their nearest cluster centre. This is repeated until a prescribed stopping criterion or convergence. The clustering goal is to minimize the expression in equation 9.

$$\underset{\mathbf{S}}{arg\,min} \sum_{i=1}^{k} \sum_{x \in S_i} \|\mathbf{x} - \boldsymbol{\mu_i}\|^2 \tag{9}$$

k is the preset number of clusters, S_i is the i^{th} cluster of set $\mathbf{S} = \{S_1, S_2, ..., S_k\}$. μ_i is the mean of the i^{th} cluster and $\mathbf{x}$ is a data point belonging to cluster S_i.

The expression in equation 9 describes the sum of the intra-cluster distances, which are distances between objects belonging to the same cluster. The minimisation of this expression results in a set of clusters S, where the similarity between objects in a particular cluster is high. I.e., if the distance is minimised, then the similarity is maximised.

The requirement to preset the number of clusters k is the critical drawback of this method [15]. In multidimensional problems ($n > 2$), it is problematic to determine beforehand the representative amount of clusters to utilise if no prior information is known on the mechanics involved in generating the data points. For this, a selection criterion must be utilised in order not to cluster based on an arbitrary number of clusters.

Two commonly used criteria for model selection are the Akaike information criterion (AIC) and Bayesian Information Criterion (BIC). In general, both criterion attempt to weigh the goodness of fit of a model versus model complexity. That means a model with high approximation power and a large number of parameters may be able to fit the data very well. However, this model would be prone to overfitting, i.e., it will fit the training data well; however, it will perform poorly on data it has not seen. The equations for AIC and BIC can be seen in equations 10a and 10b respectively.

$$AIC = 2k - 2ln(\hat{L}) \tag{10a}$$

$$BIC = ln(n)k - 2ln(\hat{L}) \tag{10b}$$

In equations 10, k is the number of parameters in the model. This is also the number of clusters in k-means. n is the number of samples. $\hat{L}$ is the maximum value of the likelihood function of the model.

The proposed method of utilising k-means clustering differs from Bayesian regression as the former does not explicitly model the nominal state. Additionally, no context is utilised in the clustering.

III. Methodology

Having presented the background information, the methods utilised in this paper are specified in this section. The resilience analysis process consists of the steps below. This method has been applied to the particular case of airport resilience assessment to disruption in the form of high winds. The data processing flow can be seen in figure 4. While the approach is general, this paper shall focus on the impact of high winds on arrival time.

1) Derive Airspace Metrics
2) Train nominal conditions model
3) Detect deviations due to disruptions
4) Quantify resilience based on deviations

The ADS-B data is obtained from the Airbus Airsense global ADS-B database [4]. The database utilised for this research contains ADS-B messages captured by a global ground-based receiver network for the year 2018. Each ADS-B message is stored in a decoded format along with additional metadata per message consisting of UNIX timestamp, unique flight id, arrival, and departure airports. The data is stored in a cloud, and airspace metric derivation has been performed utilising Apache Spark big data computing engine running on a cloud-based cluster [33].

Wind information for corresponding airports has been obtained from METAR messages. The METAR data has been queried from the Iowa State University database, which collects historical information from the Automated Surface Observing System network [3].

The details of the process follow.

A. Metric derivation

The metrics are derived for a circular region with a radius of 100 Nautical Miles around the airport being investigated. There also exists the possibility to precisely model the airport Terminal Manoeuvring Area (TMA) based on a three-dimensional polygon of that airspace. However, that approach would not create a possibility for benchmarking airports as some of the metric values would be incomparable. For example, an airport with a smaller TMA will have, most likely, shorter nominal arrival times than an airport with a large TMA.

Thus, a more general representation in the form of a circular region is preferred. The radius of 100 NM is chosen based on the practice of Air Navigation Service Providers in approximating the arrival airspace [14]. The filtering for aircraft positions within the 100 NM distance of the airport can be performed with the use of Haversine formula between the aircraft position and the airport reference point.

The metrics to be derived from ADS-B data can include continuous values such as aircraft count or event-based values such as landing rate. The latter are referred to as event-based as the value can only be changed once the event occurs. The former can be measured continuously as the aircraft are present in the observed airspace at every moment.

Event-based metrics are a collection of objects with corresponding times of occurrence. As the general trend in the arrival airspace is being considered, these events can be transformed into a time-series metric using windowed aggregation.

Figure 5 presents an example of how time-based window aggregation is performed. The event, landing, is being captured. From this event, two possible metrics are derived. First is the landing count in a time window. Second, the mean

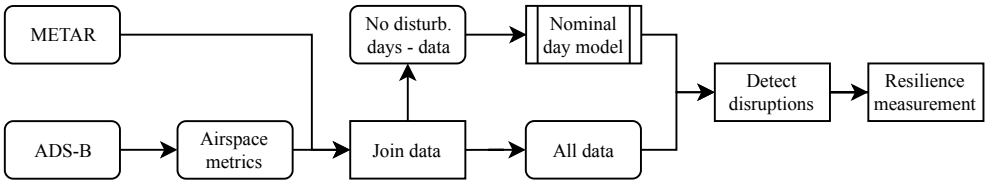


Fig. 4: Resilience measurement process, flow of data.

Time	Time-in	Landing
1	**25**	**True**
2	None	False
3	**35**	**True**
4	**30**	**True**
5	None	False
6	None	False
7	None	False
8	**20**	**True**
9	**20**	**True**
10	None	False

Window start time	Time-in (mean)	Landing (count)
1	30	3
6	20	2

Fig. 5: Window aggregation of events for time-series creation

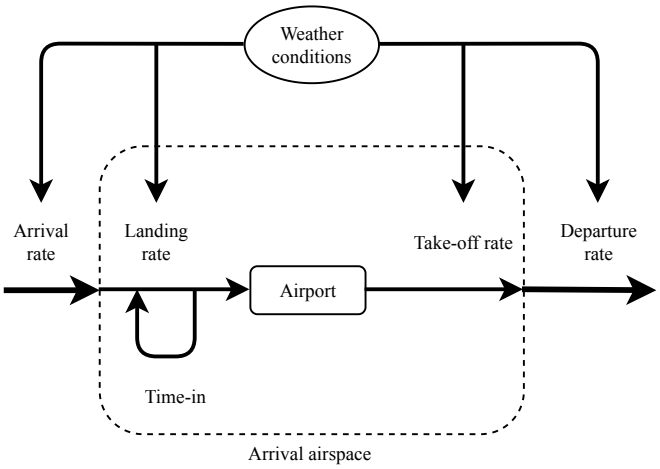


Fig. 6: Airspace context for time-in metric.

time-in sector can be derived, assuming that the entry time into the airspace is also known.

A key aspect in the development of context-based outlier detection modelling is the proper selection of features representing the context. As the time-in sector metric is being considered as the target value, the features must adequately describe critical parameters in the mechanism that affect this metric. The context for this is metric is presented in figure 6.

Figure 6 presents the relations and factors affecting the Time-in metric. Overall, the system boundary is defined on the physical boundary of the abstracted arrival airspace. The system has two parameters that directly impact the time-in metric the arrival and landing rates. An imbalance between the two results in an increase of the time-in metric because the arriving aircraft will need to remain longer in the airspace before they can land.

Due to the limited throughput of airport runways, the take-off rate affects the landing rate. Air Traffic Controller (ATC) may choose to increase the landing rate at the expense of the take-off rate or vice versa, depending on the needs. As such, the take-off rate will indirectly influence the time-in metric.

All of the throughput related rates are affected by weather conditions. A direct causal relationship between weather conditions and airport throughput exists. Poor weather in the form of high winds, low visibility, or other conditions will reduce airport throughput. An indirect relation is existent between arrival and departure rate as those two may be reduced if the poor weather conditions are forecasted and information is communicated to aircraft scheduled to arrive or depart at the airport. In such a scenario, the aircraft may hold for the weather conditions to improve.

There exist other factors affecting the time-in sector, such as the number of staffed ATCs, controller decisions, TMA structure, runway closures, and deicing. Some of these factors will indirectly affect the time-in metric through other metrics. For example, a runway closure might not be directly present as a feature in the data; however, it presents itself through the landing rate.

B. Data description

The **response variable considered within this paper is the median arrival time in a 15-minute window**. The arrival time is defined as the time between crossing into 100 NM radius and last descent below 1200 feet before landing. The features (explanatory variables) are derived based on context values of figure 6. However, feature engineering must be performed on the context values to improve model performance. This constitutes performing transformations and combinations of input values based on domain knowledge.

The data consists of airspace metrics at 15-minute intervals. The sampling rate value has been chosen based on the 30 minute METAR issuing period such that if a weather change is observed, the effect on the metrics can be captured before the next METAR message [2].

Two sets of features (explanatory variables) were investigated. The first includes the basic features, while the second adds lagged values of features. The two sets shall be compared to select the feature set that more accurately models the nominal conditions during low wind conditions. The basic set of features includes:

- **Arrivals** count in the 15 minute time window
- **Landings** count in the 15 minute time window
- **Departures** count in the 15 minute time window
- **Take-offs** count in the 15 minute time window
- **Landing & take-off** difference
- **Arrival & landing** difference
- **Departure & take-off** difference
- **Wind speed** in knots (as reported by METAR)
- **Wind speed directional components**

The direction of the wind plays a significant role in effecting an airport. For example, a strong wind along a runway direction may not cause a disruption. However, the wind of the same magnitude but perpendicular to the runway could severely impact operations due to the need to perform cross-wind landings. Hence, the Wind Speed Directional Components are a method to account for wind speed as well as direction. They are defined as in equation 11. c_ϕ is the wind speed directional component, w_{speed} is the wind speed in knots, w_{dir} is the wind direction, and ϕ is the centre of the half-sinusoid pulse function while $\Delta\phi$ is the half-width of the pulse.

$$c_\phi = w_{speed}\, f_{hs}(w_{dir}, \phi, \Delta\phi) \tag{11}$$

The half-cycle sinusoid pulse function is defined as in equation 12.

$$f_{hs} = \begin{cases} \cos((w_{dir} - \phi)\frac{360}{4\Delta\phi}) & \text{if } w_{dir} \in [\phi - \Delta\phi, \phi + \Delta\phi] \\ 0 & \text{else} \end{cases} \tag{12}$$

The characteristic of the function f_{hs} is that it is zero outside of the interval $[\phi - \Delta\phi, \phi + \Delta\phi]$ while in the interval it takes the value of the positive half of the sinusoid cycle. The term $\frac{360}{4\Delta\phi}$ modifies the frequency of the sinusoid so that it fits the interval.

Eighteen features are created using the f_{hs} function. Each one is separated 20 degrees and has a half-width of 20 degrees, such that there is an overlap between each function. Hence, $\phi \in \{0, 20, ..., 340\}$ and $\Delta\phi = 20$.

The second set of features includes all the previously mentioned features plus lagged values of selected features. The lagged values are considered as there may be a delay in the effect of some variables. I.e., a decreased landing capacity will only cause the time-in value to increase after an hour. The lagged features include: landing and take-off counts, landing & take-off difference, arrival & landing difference, departure & take-off difference, and wind speed. The time lags utilised are 0.5, 1, 1.5, and 2 hours.

C. Nominal conditions modelling

With the metrics derived and processed into a tabular form that allows for utilisation in algorithms, the nominal state can be modelled. Two methods shall be utilised for this task. First is K-means clustering, which will be used in a context-free setting - the second being Bayesian Ridge Regression, which will utilise the knowledge of the context to model the nominal state.

In the nominal states in both context-aware and context-free settings is derived based on days where the wind speed has not crossed a threshold of 15 knots. The threshold is selected based on Eurocontrol's ATM Airport Performance (ATMAP) weather algorithm [13]. The algorithm is utilised to score the weather conditions at airports based on visibility, wind, and other conditions. The higher the score, the worst the weather conditions. In this algorithm, wind becomes contributing to the score only once it crosses the value of 15 knots.

1) K-means metric clustering: The K-means algorithm, described in section II-C.3, is used only on the response variable (arrival time) to model the nominal state in a context-free setting. The k-means algorithm accepts inputs in a tabular format with rows being samples and columns being features. The input format is shown in figure 7. The m parameter is dependent upon the sampling chosen for the metric. For an entire day, it will be equal to the length of a day over the sampling interval time.

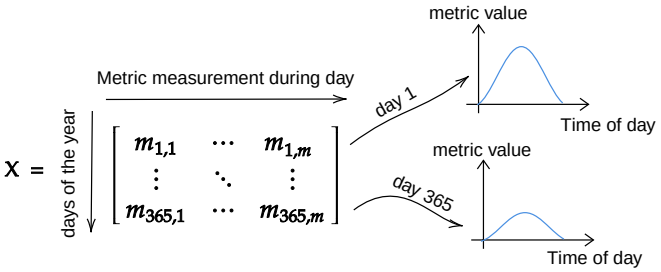


Fig. 7: Input format for K-means clustering algorithm.

2) Bayesian Ridge Regression: For context-aware nominal state modelling, Bayesian Ridge Regression is utilised using the Scikit-learn implementation using the default hyperparameter values [16]. In this method, each sample corresponds to one metric measurement during the day, unlike in the k-means method, where a full day constitutes a sample. The output of the regression will be the mean and variance of a Gaussian distribution that is the posterior probability distribution for a target metric value given its context.

The bayesian ridge regression variant, as implemented in Scikit-learn, was chosen because of the high number of features, particularly in the feature set with lagged features. Ridge regression, as such, will prevent over-fitting the model. Additionally, by using the Scikit-learn implementation, the reproducibility becomes simpler as this package is well known within the machine learning community.

D. Deviation detection

The detection of deviation is inherently an outlier detection problem. The Bayesian Ridge Regression method directly returns a probability distribution, which is utilised for quantification of the deviation likelihood. In the case of K-means, the distribution can be estimated utilising Gaussian Mixture Models under the assumption that the values are normally distributed around the cluster centre.

Consider that the nominal state has been modelled. The probability distribution for a metric measurement i is given by equation 13. μ_i is the mean whilst σ_i^2 is the variance for the i^{th} measurement.

$$M_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{13}$$

Note that in the case of Bayesian Ridge Regression, and more generally, context-based methods, the values of μ_i and σ_i^2 are dependent upon the context $\mathbf{C}_i$. In context-free methods like k-means, the values above are derived without consideration of the context.

As such, the deviation probability can be quantified based on the Cumulative Distribution Function (CDF) of the distribution from equation 13. If the deviations are considered to be values larger than usual, then the value of the CDF can be used directly, as shown in equation 14.

$$F_{M_i}(m_i) = \int_{-\infty}^{m_i} \mathcal{N}(\mu_i, \sigma_i^2) \tag{14}$$

The CDF returns a value between 0 and 1, which indicates the likelihood of a value being lower than m_i for the given probability distribution. It is important to note that the CDF value for a $m_i = \mu_i$ will be 0.5. Hence if a disruption results in a reduction of a metrics value, such as in the case of landing rate, then the Survival Function can be used instead, which is defined in equation 15. For the time-in metric, disruptions result in a positive deviation; hence, the CDF is appropriate.

$$S_{M_i}(m_i) = 1 - F_{M_i}(m_i) \tag{15}$$

The CDF and Survival Function return a probability for a single measurement. However, deviations, especially those caused by weather, will possibly occur over multiple measurements. As such, there exists a likelihood of spurious, single measurement deviations that are not caused by a disruption.

The solution to the above is to utilise a filter to smooth the values in time. The simplest method for this is a rolling average filter. The concept of a rolling window is presented in figure 8.

Within each window, an operation can be performed on the previously obtained probability values. In this research, the mean is utilised. Based on the resultant value, an assessment can be made whether a deviation has occurred or not. In the case of a mean, a confidence level value can be used, such as 0.95. However, the mean value is no longer a statistical value but rather a score. Hence, it should not be treated as a probability or likelihood.

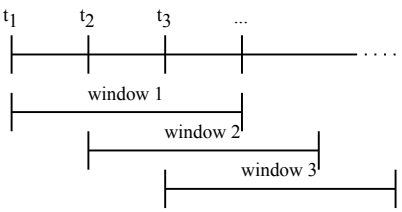


Fig. 8: Rolling window

Another consideration is the blurring introduced by the window, which will result in the deviation start and end being below the detection threshold. A proposed solution is to capture a period half of the window length before and after the disruption to capture the complete extent of the deviation. Alternatively, a more sophisticated window such as the exponential window can be utilised. However, the exponential window will not be utilised in this paper.

E. Resilience quantification

Once the deviations are captured, the resilience can be computed on a per airport basis. As mentioned in section II-A, resilience consists of three capacities: absorptive, restorative, and adaptive. The first two cannot be measured independently of the adaptive capacity if the latter is present.

The absorptive capacity allows the airport to withstand a disturbance without experiencing stress, i.e., without a metric deviation. Hence, if an airport does not experience deviations due to a frequently occurring disturbance, then it can be considered robust or possessing a strong absorptive capacity.

However, airports will experience varying amounts of a particular disturbance. For example, airports may experience particular weather conditions more frequently and at a higher intensity than other airports. Thus, robustness should be dependent upon the intensity of disruptions experienced as well. Thus, equation 16 presents the absorptive capacity metric.

$$M_{abs.} = \left(1 - \frac{n_d}{m_{high\ wind}}\right)\left(1 - \frac{15}{w_{mean}}\right) \tag{16}$$

n_d is the number of deviations during which the wind speed exceeded the 15-knot limit. $m_{high\ wind}$ is the number of days during the year where the wind speed exceeded the threshold mentioned above. w_{mean} is the mean wind speed during the disruptions of n_d.

The absorptive capacity defined in 16 is limited to values between 0 and 1 for no absorptivity and full absorptivity. The first term defines absorptivity in terms of how frequently a disruption occurs in relation to the frequency of disruptive conditions. The second term accounts for the magnitude of the disturbances in relation to the 15-knot threshold of wind speed.

The second capacity of resilience is the restorative capacity, which allows the airport to recover from a disruption.

For this capacity, the definition created by Gluchshenko et al. for resilience shall be utilised. In this definition, a system is resilient if the recovery time is short compared to the time between the start of the deviation and its maximum magnitude.

As such, this capacity can be quantified based on the ratio in equation 17.

$$M_{rest.} = \frac{t_{dev}}{t_{total}} \tag{17}$$

t_{dev} is the time from the start of the deviation until the maximum deviation magnitude. t_{total} is the total time of deviation. This metric is also limited to values between 0 and 1. In the perfect scenario, the system recovers immediately following the deviation, and the value of t_{dev} will be equal to the time t_{total}.

An assumption is made that there exists an underlying probability distribution dictating the behaviour of time-in metric in response to wind speed. As such, the deviations observed throughout a year are only a sample of data obtained from a population. In this regard, the sample may inappropriately describe the properties of the entire population. In order to address this, bootstrapping can be used to recalculate values of absorptive and restorative capacities from equations 16 and 17.

Bootstrapping is a technique which is based on sampling numerous times from the original sample [11]. The sampling is done with replacement in order not to modify the probability distribution of the original sample. Hence, the original sample is treated as the population, while the resamples are subsets of the population, i.e., statistical samples.

In the context of capacity metrics, every time a resample is performed, the metric can be recomputed. Thus, if a large number of resamples are performed, e.g., 1000, then a median value of the metric values computed for each of the resamples can be used. The result will be a more statistically accurate value of the metric for an airport.

For decision-makers, it is important to present the performance of an ATM system concisely. As such, a resilience metric is proposed that combines the absorptive and restorative capacities. The concept of the combined metric is shown in figure 10.

Resilience, according to this definition, becomes a vector in an absorption-restoration space. Thus, the larger the values of the two capacities, the higher the resilience of the system. In this general definition, the metrics of absorption and restoration can be modified to suit the needs of the decision-makers. However, the values need to be scaled to the same range for the resilience value to be meaningful.

In the above definition, the adaptive capacity is not included. If adaptation is made quantifiable, the definition can be expanded into three dimensional space to include it. However, ATM system adaptation through actions such as runway configuration change can be observed in absorption and restoration metrics.

IV. Results

In this section, the results are presented, which have been obtained based utilising data for nine airports. They are Amsterdam Schiphol (AMS), Copenhagen Kastrup (CPH), New York JFK (JFK), Chicago O'hare (ORD), Zurich (ZRH), London Heathrow (LHR), Hong Kong (HKG), Rome Fiumicino (FCO) and Munich (MUC). The selection of airports is such that a variety of wind conditions and geographical location is present. For example, Schiphol and Kastrup are known to be affected strongly by windy conditions, while Munich and Zurich are not [10].

The metrics have a sampling period of 15 minutes. To transform the events into a time series, a median has been utilised as an aggregation method for continuous values such as time-in and wind speed. I.e., the median of time-in values is taken for the 15 minute period. The metrics are extracted for the year 2018, excluding May 29$^{\text{th}}$ due to data corruption for that day.

A. Deviation from nominal

Two methods are considered for the deviation detection, the non-contextual K-means clustering, and contextual Bayesian Ridge Regression. The results of these methods are presented in the following subsections.

1) K-means clustering: The K-means clustering is performed on day-long metric samples. Hence, with a 15-minute sampling interval, there are 96 metric measurements during the day. Each measurement at a particular time thus becomes a feature. There are 362 samples are three days of the year have been removed due to data corruption.

As mentioned in subsection II-C.3, the k parameter in k-means clustering has to be preset. A proposed approach is to utilise the Akaike information criterion (AIC) and Bayesian Information Criterion (BIC). The k value should be the local minimum with the lowest k parameter. Furthermore, both AIC and BIC should suggest a similar optimal k parameter. The normalised AIC and BIC scores for clustering performed on the arrival time metric can be seen in figure 9.

In figure 9, the x-axis is the number of clusters, while the y-axis is the normalised criterion values. Normalised values are utilised as the criterion value in of itself is not essential. Instead, features like minimums and maximums are crucial. On the example of the first subplot for AMS, the AIC value suggests an optimal cluster number between 10 and 20, which is the location of the minimum. However, there is no local minimum present in BIC as the value is continuously increasing for increasing cluster number.

In the majority of the cases, the AIC value has a minimum in the range of k between 10 and 20. Some airports, such as ORD and ZRH, do not have a well presented local minimum. Furthermore, the BIC criterion features no local minimum for all airports.

Based on the computed values of AIC and BIC, an optimal amount of clusters cannot be well defined. Additionally, some airports like ORD or JFK do not have a clear minimum for both criteria. As such, the utilisation of clustering for

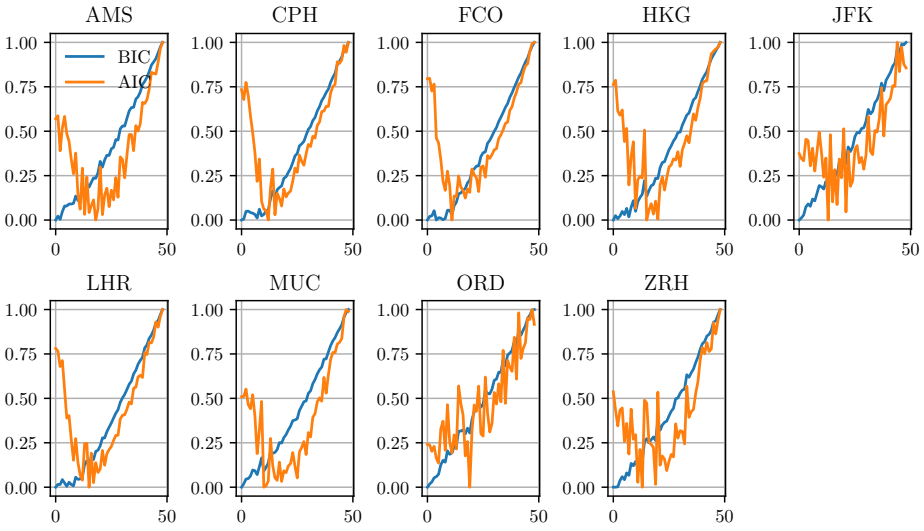


Fig. 9: AIC and BIC normalised metrics for all 9 airports. X axis is the number of clusters whilst Y axis is the normalised metric value.

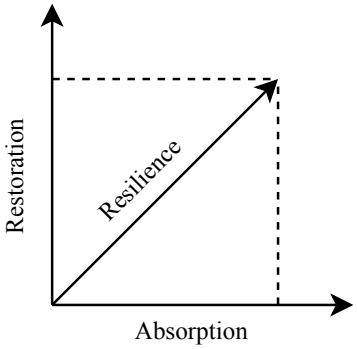


Fig. 10: Concept of Resilience Vector.

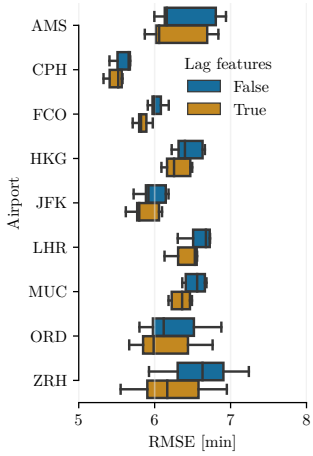


Fig. 11: Root Mean Square Errors for models trained on airport and feature combinations.

deviation detection would be ambiguous as no justification for a particular cluster number can be performed.

2) Bayesian Ridge Regression: To validate the model performance, grouped k-fold cross-validation has been performed on the two models, one with the lag features and the other without. The grouping is required in order not to train and test the model on the same days. If not performed, leakage would occur from the training set to the testing set. Ten folds have been utilised, and for each fold, a Root Mean Square Error (RMSE) has been computed. The results are reported in the box plot of figure 11. The error is computed between the model output and the actual time-in values in the testing set.

The RMSE values for all airports are between 5 and 7 minutes. Models for some airports display a more considerable variance of performance than others. For example, AMS has a broader range of RMSE values than CPH. This would suggest that CPH has conditions varying less throughout the year as the model performs similarly for each fold. As can be seen, the introduction of lag features results in a decrease of the RMSE value for all airports. Hence, the expanded feature set shall be utilised for deviation detection.

The dataset for cross-validation consists of days where maximum wind speed did not exceed 15 knots. The same is true for the overall model training set. This is done, so the model parameters reflect the system behaviour under nominal wind conditions. The predictions from such a model can be interpreted as the expected nominal values given the context.

With the model trained on days with wind speed below the threshold, the detection of deviations can be performed. Figure 12 presents a deviation due to wind high winds. The blue line is the actual arrival time value that occurred. The red line is the expected arrival time value, while the red area indicates one standard deviation on that expected value. Wind speed is shown as a green line to show a temporal correlation between the wind speed magnitude and deviation.

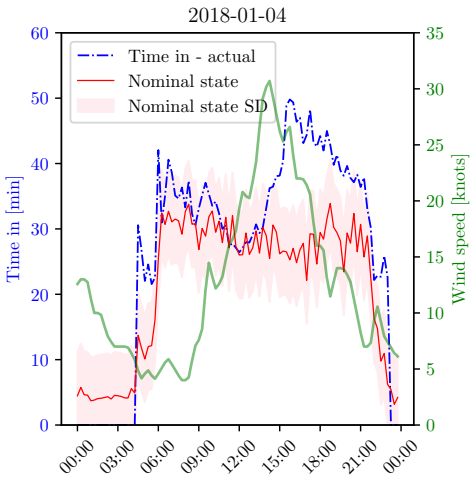


Fig. 12: London Heathrow arrival time metric deviation due to high wind disruption.

Note that a value of time-in of zero indicates that no landings are occurring.

As can be seen in figure 12, the increase in wind speed results in a deviation of the arrival time value (blue) from the expected nominal value (red). The expected nominal value is considered the centre of the distribution with a standard deviation. Given the previous, the actual arrival time value can be converted to a CDF value as in equation 14. The CDF values are averaged using a two-hour rolling window to detect deviations. As mentioned in subsection III-D, an hour is added before the start and end of the deviation to account for blurring.

It is important to note that the method does not perform well in periods when little or no operations are performed. For example, LHR has a night curfew between 23:30 and 6:00 where little or no flights depart or arrive at the airport. Despite this, it can be seen that in figure 12, the nominal state for these times is non-zero.

As mentioned previously, there exists a temporal correlation between arrival time value and wind speed. This relation is non-linear, and a lag is present between the increase of wind speed and arrival time value. The study of such dynamics could result in better arrival time predictions and capacity estimation given a weather forecast.

B. Aggregate airport statistics

The airport aggregate statistics have been obtained utilising bootstrapping using 1000 resamples. The result for the mean value of absorptive capacity can be seen in figure 13b. Note that the results are shown in a box plot form to show the range of mean values obtained from bootstrap.

From figure 13b, it can be seen that Amsterdam Schiphol airport posses the highest absorptive capacity followed by New York JFK and Rome Fiumicino. The lowest absorptive capacities are displayed by Hong Kong International, Munich, and London Heathrow.

The performance of Schiphol can be potentially explained by the airport design adjusted to the windy conditions, which are frequently occurring in Amsterdam. The airport utilises eight different runway configurations over ten runway directions. The three airports with the lowest absorptive capacity have a parallel runway configuration.

From the worst performers, Munich airport experiences relatively little windy conditions [13]. Thus, it is hypothesised that its relatively weak absorptive capacity is resultant from a lack of system adaptation to such conditions. Furthermore, Hong Kong airport experiences severe wind shear, which is likely to contribute to its low absorptive capacity [1].

Next, the restorative capacity is presented in figure 13a. The restorative capacity is the highest in Amsterdam Schiphol, followed by Copenhagen Kastrup. The worst performance is seen in Munich, followed by Rome Fiumicino. There is a tendency for airports that frequently experience strong winds to have a better restorative capacity to such conditions than airports not experiencing these disruptions.

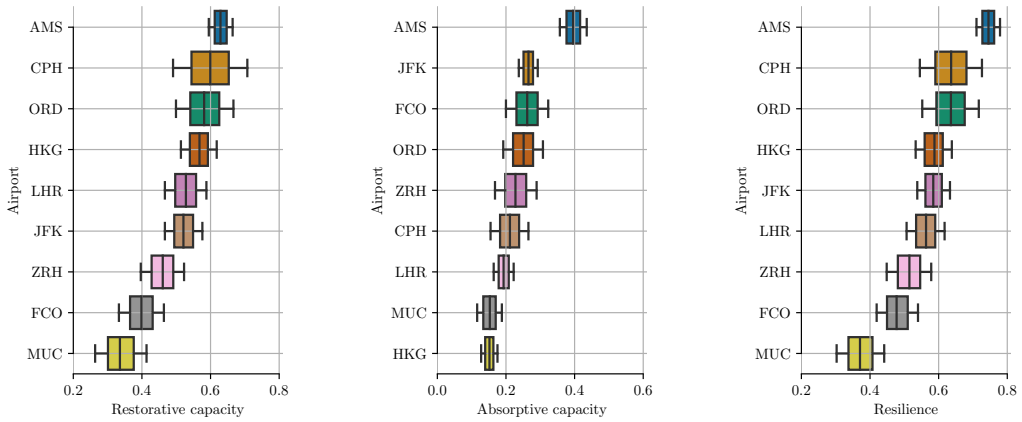
Lastly, resilience as defined in figure 10 is show in 13c. Amsterdam and Copenhagen airports top the list. Notably, Amsterdam has the best absorptive and restorative capacities for the airports considered, which makes it also the most resilient airport to high winds. However, Copenhagen performs better in restoration than absorption. This indicates that Copenhagen is relatively easily disturbed by high winds. However, once a disruption occurs, it is quickly dealt with.

The airports with lower resilience tend to have a parallel runway configuration except for Rome. Airports with high resilience have runways facing different directions. This difference is expected to increase resilience to high winds as the airport has more choice to adapt to varying wind direction and thus avoid cross-wind landings.

V. Conclusion

A method for detecting KPI metric deviation due to high wind speeds has been proposed. The method utilises Bayesian Ridge Regression to asses the nominal KPI metric state given current airspace conditions. The use of Bayesian Ridge regression provides distribution variance, which is utilised to assess the likelihood of a deviation being present. K-means clustering has been proven unusable for day-long metric samples due to the inability to specify the k parameter unambiguously. The Bayesian Ridge Regression method has been applied to traffic in the year 2018 for nine airports. The proposed resilience measurement method based on absorptive and restorative capacities has shown that airports facing disruptions from high winds tend to be more resilient to them than those that do not. Further research must be performed on the utility of this method to different KPI and disruptions.

This research presents two key contributions. The first of which is the utilisation of context-based anomaly detection as a means of defining a nominal state of a system state quantified through a metric such as a system KPI. This method enables measuring system state disruptions, and hence also resilience, when the nominal state of the system is unknown or not deterministically definable. The second



(a) Restorative capacity (b) Absorptive capacity (c) Resilience

Fig. 13: Capacity and resilience scores of all the airports. Scores obtained by bootstrapping with $n = 1000$.

contribution is the introduction of resilience as a vector in the space of absorptive and restorative capacities. Furthermore, absorptive and restorative capacities are defined through a metric that accounts for the magnitude and frequency of occurrence of disturbances.

The contributions, as mentioned earlier, can be utilised by aviation stakeholders to detect the presence of disruptions that would affect operations. For example, the anomaly detection algorithm could be adapted to operate in real-time and alert airline operation centres when airports to which their aircraft are flying to are experiencing arrival time disruptions. Additionally, the resilience metric can be utilised to assess the impact of changes introduced by airports or Air Navigation Service Providers.

The methodology has been tested only on arrival time metric as system state and wind speed as the disturbance factor. Furthermore, there exists a possibility that a deviation detected during high wind weather may have been caused by a third factor, thus causing a spurious relationship. Furthermore, only ADS-B and METAR data have been utilised. Lastly, only airports with high traffic volumes have been considered.

Two paths for further research are identified. First focuses on the system dynamic response to disturbance, which includes more precisely understanding the relationship between the disturbance properties and the airspace response. Individual aircraft trajectories under varying disturbance conditions could be studied utilising more sophisticated models such as Long Short Term Memory neural networks or Gradient Boosting. Additionally, the application of the method for metrics other than arrival time should be tested

VI. Acknowledgements

This research would not be possible without the support of Airbus Defence & Space. I want to thank Carina Schnitzenbaumer sincerely, Airbus Airsense Project Lead and Heiko Udluft, Airbus Airsense Product Owner, for making it possible for me to conduct my master thesis with Airbus. Furthermore, I thank Jacco Hoekstra, Joost Ellerbroek, and Giuseppe Sirigu for their academic and technical guidance.

References

[1] P. W. Chan, Severe wind shear at Hong Kong International Airport: climatology and case studies, Meteorological Applications, vol. 24, no. 3, pp. 397403, 2017.

[2] Decoding Aviation Weather Report (METAR/SPECI). [Online]. Available: $https : //www.hko.gov.hk/aviat/decode_metar_e.htm.$ [Accessed: 14-Nov-2019].

[3] Iowa State University, Iowa Environmental Mesonet of Iowa State University. [Online]. Available: https://mesonet.agron.iastate.edu/request/download.phtml. [Accessed: 06-Nov-2019].

[4] AirSense - Trusted insights for smarter decisions, AirSense. [Online]. Available: https://www.airsense.airbus.com/. [Accessed: 04-Nov-2019].

[5] Long-term Forecast of Annual Numbers of IFR Flights up to 2040, Eurcontrol, Sep. 2018.

[6] D. Delahaye, S. Puechmorel, and S. Puechmorel, Modeling and Optimization of Air Traffic - Airspace Congestion Metrics, in Modeling and Optimization of Air Traffic, Somerset, UNITED STATES: John Wiley & Sons, Incorporated, 2013.

[7] Civil Air Navigation Services Organisation, Recommended Key Performance Indicators for Measuring ANSP Operational Performance, Civil Air Navigation Services Organisation, Mar. 2015.

[8] International Civil Aviation Organization, Manual on global performance of the Air Navigation System, International Civil Aviation Organization, Montral, Quebec, 2009.

[9] Comparison Of Air Traffic Management-related Operational Performance U.S./Europe, Eurocontrol, FAA, Mar. 2019.

[10] M. Schultz, S. Lorenz, R. Schmitz, and L. Delgado, Weather Impact on Airport Performance, Aerospace, vol. 5, p. 109, Oct. 2018.

[11] C. F. Mooney, C. L. Mooney, C. Z. Mooney, R. D. Duval, and R. Duvall, Bootstrapping: A Nonparametric Approach to Statistical Inference. SAGE, 1993.

[12] A. C. Bovik and S. T. Acton, Chapter 10 - Basic Linear Filtering with Application to Image Enhancement, in The Essential Guide to Image Processing, A. Bovik, Ed. Boston: Academic Press, 2009, pp. 225239.

[13] Algorithm to describe weather conditions at European airports, Eurocontrol, Brussels, Belgium, 2011.

[14] Civil Air Navigation Services Organisation, Recommended Key Performance Indicators for Measuring ANSP Operational Performance, Civil Air Navigation Services Organisation, Mar. 2015.

[15] D. T. Pham, S. S. Dimov, and C. D. Nguyen, Selection of K in K-means clustering, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 219, no. 1, pp. 103119, 2005.

[16] F. Pedregosa et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, vol. 12, pp. 28252830, 2011.

[17] K. P. Murphy, Machine learning: a probabilistic perspective. MIT press, 2012.
[18] C. M. Bishop, Pattern recognition and machine learning. Springer, 2006.
[19] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. The MIT Press, 2016.
[20] F. Xue and W. Yan, Parametric model-based anomaly detection for locomotive subsystems, in 2007 International Joint Conference on Neural Networks, 2007, pp. 30743079.
[21] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Saint Louis, UNITED STATES: Elsevier Science & Technology, 2011.
[22] A. Zimek and E. Schubert, Outlier Detection, in Encyclopedia of Database Systems, L. Liu and M. T. zsu, Eds. New York, NY: Springer New York, 2017, pp. 15.
[23] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, Time-series clustering A decade review, Information Systems, vol. 53, pp. 1638, Oct. 2015.
[24] J. Lin, E. Keogh, S. Lonardi, and P. Patel, Finding Motifs in Time Series, Proceedings of the Second Workshop on Temporal Data Mining, Oct. 2002.
[25] D. Insua, F. Ruggeri, and M. Wiper, Bayesian Analysis of Stochastic Process Models. Hoboken, United Kingdom: John Wiley & Sons, Incorporated, 2012.
[26] A. Stevenson, Ed., Oxford Dictionary of English. Oxford University Press, 2010.
[27] S. Hosseini, K. Barker, and J. E. Ramirez-Marquez, A review of definitions and measures of system resilience, Reliability Engineering & System Safety, vol. 145, pp. 4761, Jan. 2016.
[28] R. Francis and B. Bekera, A metric and frameworks for resilience analysis of engineered and infrastructure systems, Reliability Engineering & System Safety, vol. 121, pp. 90103, Jan. 2014.
[29] M. Bruneau et al., A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities, Earthquake Spectra, vol. 19, no. 4, pp. 733752, Nov. 2003.
[30] A. Cook and D. Rivas, Complexity Science in Air Traffic Management. London, United Kingdom: Routledge, 2016.
[31] Jrg Leonhardt, Erik Hollnagel, and Luigi Macchi, A White Paper on Resilience Engineering for ATM, Eurocontrol, White Paper, Sep. 2009.
[32] O. Gluchshenko and P. Foerster, Performance based approach to investigate resilience and robustness of an ATM System, p. 7, 2013.
[33] M. Zaharia et al., Apache Spark: A Unified Engine for Big Data Processing, Commun. ACM, vol. 59, no. 11, pp. 5665, Oct. 2016.
[34] C. W. Zobel and L. Khansa, Quantifying Cyberinfrastructure Resilience against Multi-Event Attacks, Decision Sciences, vol. 43, no. 4, pp. 687710, 2012.

II

Preliminary Thesis [graded]

Nomenclature

Abbreviations

ADS-B	Automatic dependent surveillance – broadcast
ANSP	Air Navigation Service Provider
API	Application Programming Interface
ATCo	Air Traffic Controller
ATM	Air Traffic Management
CANSO	Civil Air Navigation Services Organization
DLR	German Aerospace Center
DTW	Dynamic Time Warping
FAA	Federal Aviation Administration
ft	Feet
GMM	Gaussian Mixture Model
ICAO	International Civil Aviation Organization
IFR	Instrumental Flight Rules
KPA	Key Performance Areas
KPI	Key Performance Indicators
kts	Knots
LHR	London Heathrow
NM	Nautical Miless
PC	Principal Component
PCA	Principal Component Analysis
TMA	Terminal Manoeuvring Area
TMI	Traffic Management Initiative
UTC	Coordinated Universal Time
XOR	Exclusive OR operator

List of Figures

List of Tables

Introduction

The Air Traffic Management system is consistently put under more pressure. From year to year the number of flights increases. In 2008 there were 800 million passengers transported in Europe. In 2017, this number grew to over a billion [22]. Whilst major airports are planning expansions, the increase in capacity will not be enough as Eurocontrol predicts a seven-fold increase of delays between 1-2 hours [19].

The increased demand requires a higher optimisation of airport and airspace usage. Whilst initiatives like Airport Collaborative Decision Making are proving useful in the planning of operations, the delays continue growing. As such, over the last two decades, interest has been growing in the field of resilience. Resilience, defined as the ability to recover quickly from difficulties is crucial to maintaining operational capacity through disruptive events. Colloquially speaking, the field of resilience has attracted attention as there is a desire to not only operate optimally but also operate through as many conditions as possible.

Originally, the field of resilience has been associated purely with safety management in the ATM field. As such, the majority of the work focuses on qualitative frameworks for safety analysis [41]. However, over the last several years it became apparent that resilience perspective can also be used to look at issues of capacity. As such, airspace and airport can be analysed and designed with regard to their ability to recover from disturbances.

Before airspace resilience can be designed for, it must be measurable. As such, the goal of this research project is to develop metrics for measurement of resilience in the case of adverse weather conditions including high winds and low visibility. In particular, the metrics should be obtainable from ADS-B data at around airports as this is an easily accessible data source that is present worldwide. As such, it allows for performance comparison between airports. Previous research has already been performed on resilience metrics however not utilising ADS-B data.

The challenge of measuring resilience in real systems is determining the nominal state and detection of deviations from it. Air traffic is extremely complex and day to day operations vary vastly. These high variations can depend on the day of the week, month, season, schedule etc. The scale of the variations is large enough that it can cover up deviations due to disturbing events. Thus, it is difficult to determine what is considered nominal and what is an anomaly caused by weather. To visualise this, consider figure 1, which presents aircraft counts within London Terminal Manoeuvring Area during some days in the year 2018. One of those days has been affected by severe winds, which caused flight cancellations. However, which day is it?

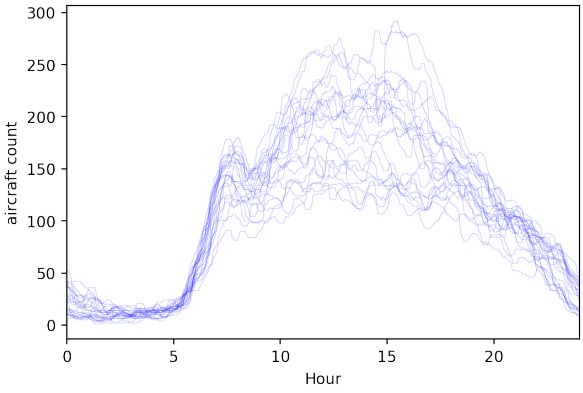


Figure 1: Aircraft count within London TMA for a random set of days. One day has had severe winds, which day is it?

This preliminary thesis report starts with the description of the overall project methodology. Chapter 1 presents the trends in ATM delays and their causes. Next, chapter 2 presents the concept of resilience and research performed on this topic. Chapter 3 gives an overview of developed ATM metrics. Afterwards, chapter 4 introduces clustering for time-series analysis. Lastly, the preliminary results which include metric extraction, Principal Component Analysis and clustering are presented in chapter 5.

Methodology & planning

This chapter presents the planning and methodology of the thesis work. The thesis goal is to develop a method for quantifying airspace resilience by utilising ADS-B data. Before actual research work can commence the essence of the research must be determined. For this, the research questions to be answered are required. The main research question to be answered is ***whether airspace resilience can be quantified using aircraft positional data***. The sub-questions are formulated below.

- What are the metrics that describe the Airspace State?
  - Do the metrics correlate with weather conditions in the Terminal Manoeuvring Area?
  - Do the metrics have a causal relationship with weather conditions in the Terminal Manoeuvring Area?
- Which resilience measurement method is most appropriate?
  - Stochastic vs. Deterministic measure?
  - What is considered a nominal state?
  - What is considered a perturbed state?
- Can clustering on ADS-B derived data be used to identify days during which disruptions have occurred?
  - What clustering algorithm is most appropriate?
    - Which timescale?
    - How to deal with different dimensional data?
    - Cluster including weather data?
    - What distance metric?
  - Can results of the clustering be used to derive nominal airspace state?

Overall, the project consists of a preliminary thesis phase and a thesis phase. The preliminary phase consists of literature study where previous research and technical documentation is analysed for information useful to the thesis work. Additionally, the preliminary phase consists of data pre-processing phase where data obtained from Airbus Airsense ADS-B database is converted into raw airspace metrics. The preliminary phase shall finish with initial prototyping work on clustering of metrics.

The final thesis phase begins with performing clustering on multiple airspace metrics to obtain a prototype of a nominal airspace state which will be then used to derive airspace resilience by comparing non-nominal days recovery to a nominal state. The results of this phase will be validated by comparing two airports. The detailed workflow can be seen in figure 2 whilst the scheduling is presented in the Gantt chart in figure 3.

The thesis phases serve as delimiters for reporting purposes. I.e the preliminary thesis phase shall end with a presentation of progress and a preliminary thesis report shall be handed in for grading. The final thesis phase shall conclude with the thesis defence and a final thesis report.

Aside from the division into phases, the implementation aspects of the project are divided into work packages. The work packages serve as a means to discretize the work into tangible, linearly related goals. As such, consider the first work package:

Data pre-processing The goal of this work package is to utilise the ADS-B positional data along with the knowledge from literature study to obtain filtered and equally sampled airspace metrics for a Terminal Manoeuvring Area around a European airport.

Data analysis The continuation of data pre-processing. After obtaining workable data from the previous phase it is possible to perform analytics on it. The result of this phase should be clusters of days derived based on all the metrics.

Resilience derivation In resilience measurement, a key notion is the nominal state to which deviations are compared. As such, in this work package, stochastic description of the nominal airspace state is derived from the clusters (i.e obtaining cluster prototypes and adding probability bounds) and from this resilience is measured using a probabilistic method and weather data.

Validation Finally, to prove the applicability of this method to different airspaces other airports resilience will be derived with the same process and it will be compared to the first airport.

However, it is important to note that the work packages are not the only tasks to be realised in this project. The literature study is performed in conduction with raw metric extraction in the initial phases and the reporting is also done in parallel throughout the project. The schedule is planned such that the whole project will be defended in December 2019.

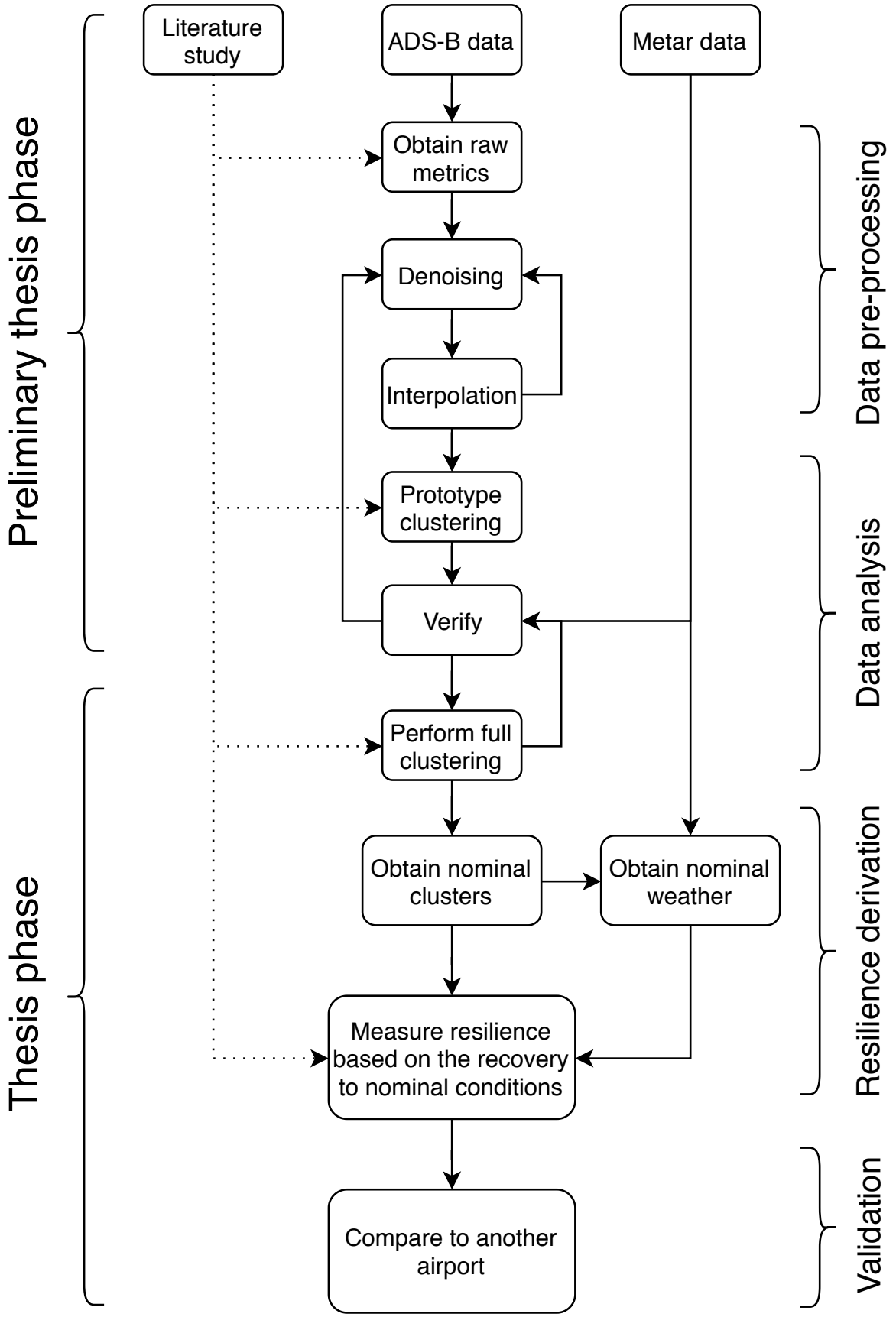


Figure 2: Thesis phases and work flow.

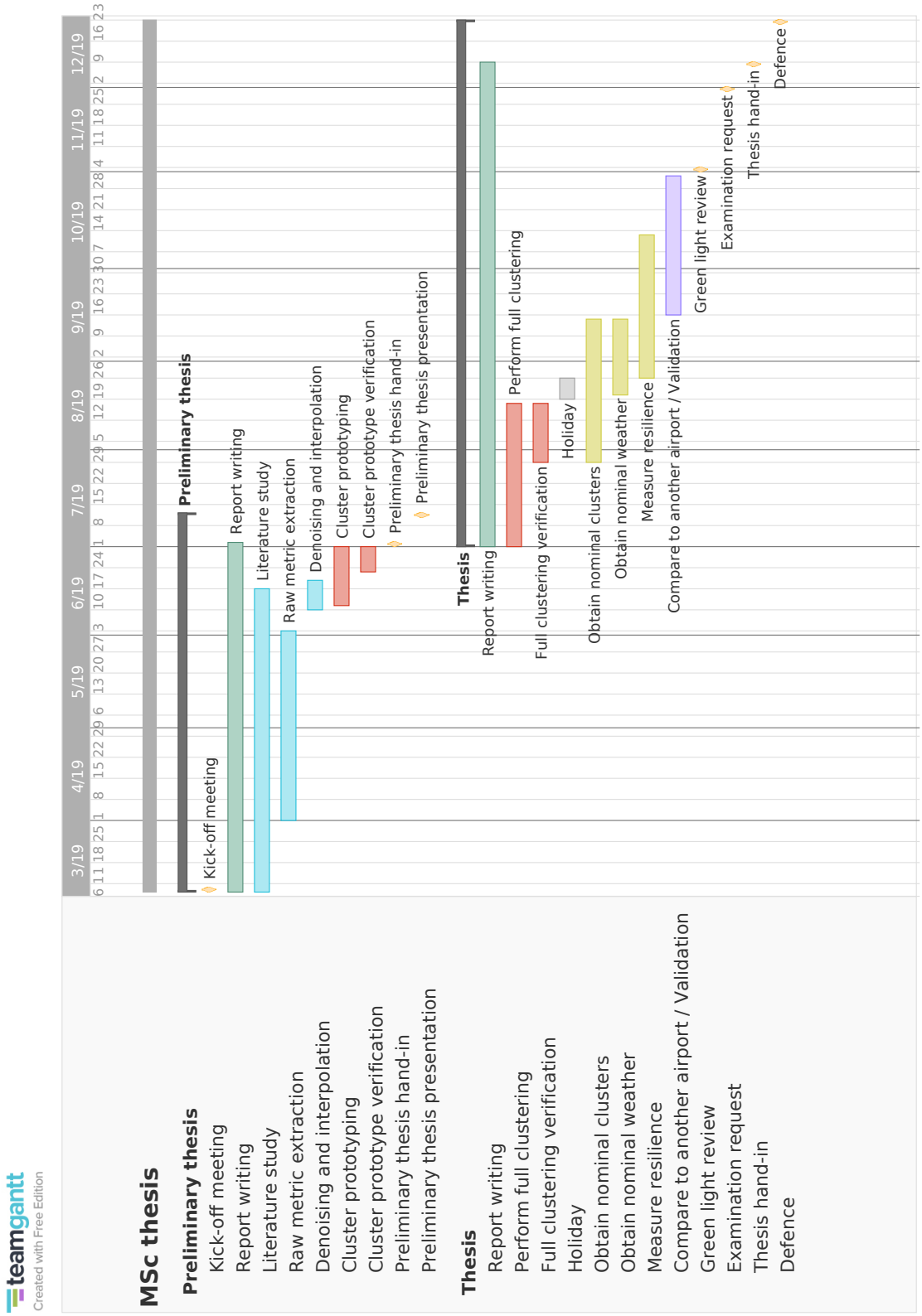


Figure 3: Gantt chart of thesis tasks. Tasks coloured by work package. Blue - preprocessing, Red - data analysis, Yellow - resilience measurement, Purple - validation, Green - report writing.

1

ATM system disruptions

The air traffic system is under increasing pressure as it's utilisation increases through growing passenger counts. The US-EU comparison report published by FAA and Eurocontrol informs that the passenger numbers in the US have increased by 7% between 2015 and 2017 whilst in Europe by 11.4%. The increases in IFR movements per airport have not been as fast given 2.4% and 5.2% increases in the US and Europe respectively. The disparity results from the fact passenger numbers per flight have also been increasing in both regions. Unfortunately, also delays are on the rise as shown in figure 1.1 [21].

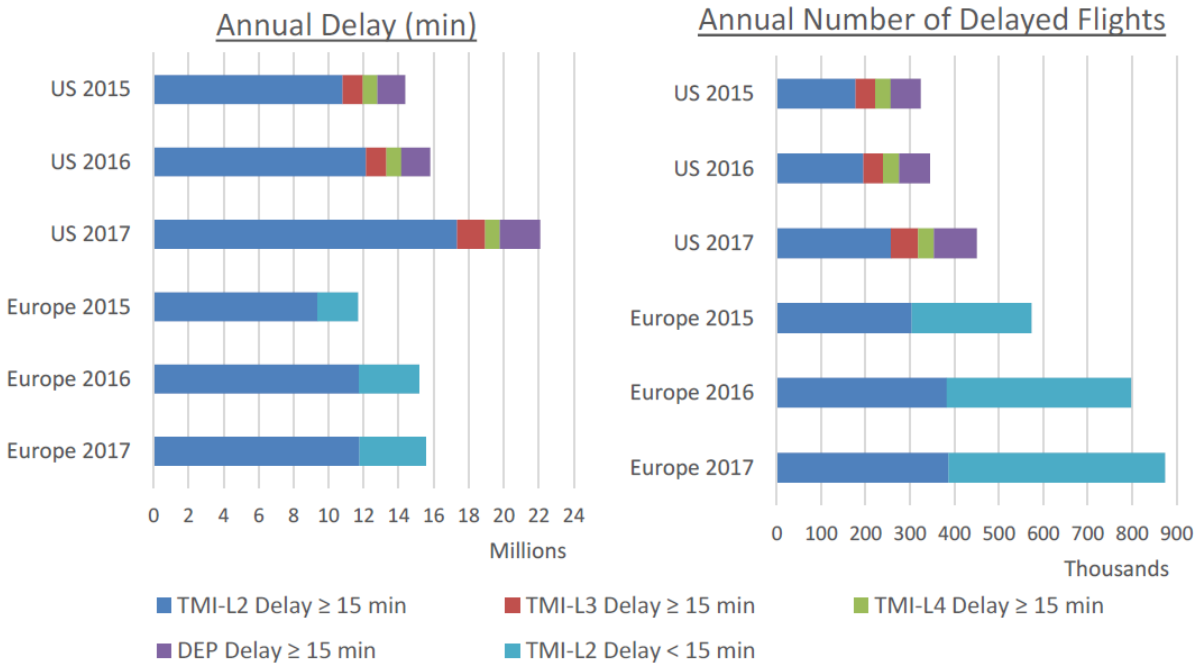


Figure 1.1: ATM Delay comparison US-Europe. TMI-L1: scheduling initiatives, TMI-L2: initiatives on day of flight but before pushback, TMI-L3: sequencing measures after pusback but before take-off, TMI-L4: initiatives after take-off, DEP: flow restrictions causing departure delay. Figure from US-EU comparison report [21].

From the above, it can be seen that the US experiences less delayed flights, once delayed the delay is larger than in Europe. Furthermore, the number of delayed flights, as well as delay length, are increasing in Europe and the US. The above shows that qualities like adaptability, robustness and resilience are becoming of more relevance to ATM as the disruptions are on the rise. Hence, the increased need for abilities to deal with them.

Whilst investigating delays, it is important to consider their causes. Figure 1.2 shows the taxonomy of delays in Europe and the US. As can be seen, in both region delays experienced at airports make up a significant proportion of the delays. Within Europe the delays are more often caused by traffic volume however, weather attributes to just over 40% of delays overall. Around airports, weather causes more than half of the delay.

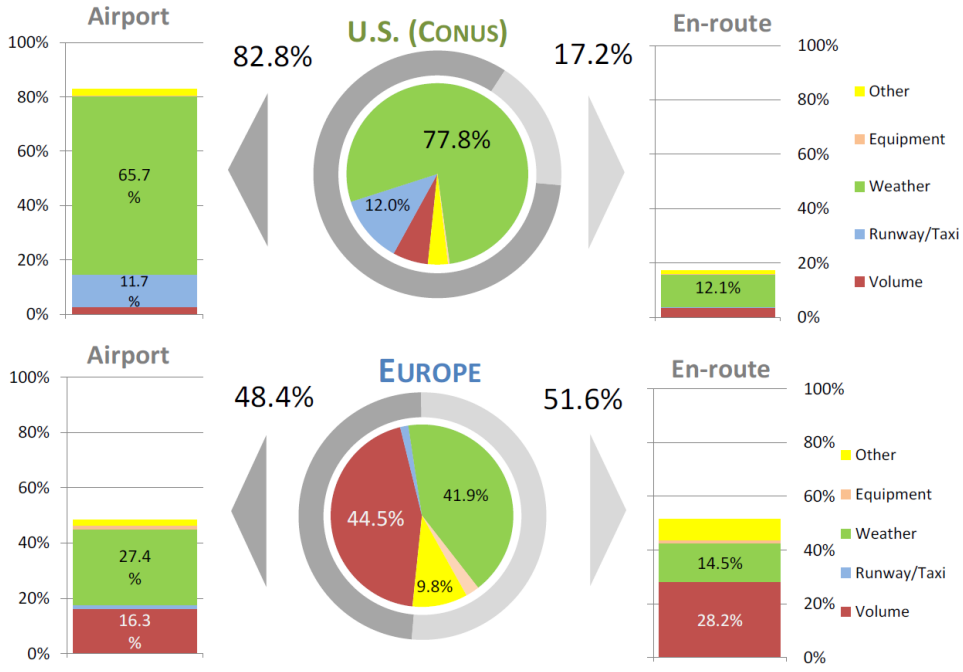


Figure 1.2: ATM delay taxonomy by cause and place. Delays above 15 minutes. Figure from US-EU comparison report [21].

The 2018 Network Operations by Eurocontrol reports that the European airports that contributed the most to delay were Amsterdam Schiphol, Barcelona El Prat, Lisbon and London Heathrow. From those, Lisboa was the strongest affected by capacity issues. The rest were facing mostly weather-related delays. In London Heathrow, the weather-related delays accounted for 75% of all delays whereas Amsterdam had around 60% of its ATM delays caused by weather. The weather-related delays were particularly dominant in the winter season. Considering this, Schiphol and Heathrow are well-suited airports for analysis of weather-related disruptions [20].

An important question for analysis of airspace response to weather is what is considered to be severe weather. Focusing on airports which are in the scope of this research, Eurocontrol suggests that Capacity Limited Weather should be defined as weather which contains any of the below.[64] These values will be utilised in further analytic work.

- Storms
- Convective activity
- Winds above 25kts or gusts above 30kts
- Visibility below 500m or cloud level below 200ft
- Severe precipitation

Furthermore, With accelerated climate change in progress, the weather conditions will be changing over the coming decades. The exact effects of climate changes on conditions such as local winds, fogs and other aviation affecting phenomena are yet to be determined [45]. However, more frequent occurrences of such events would increase the need for resilience.

2

Resilience

The definition of Resilience as defined by the Oxford Dictionary is [55]:

1. The capacity to recover quickly from difficulties; toughness
2. The ability of a substance or object to spring back into shape; elasticity

The commonality of these two definitions is the ability of an object to return to a natural steady state after a disturbing event occurred. The term originates from the field of material sciences where resilience is defined as *"the ability of a metal to absorb energy when elastically deformed and then to release it upon unloading"* [34]. According to this definition, a brittle object like a ceramic pot will not be considered resilient as it will shatter quickly under tensile stress.

Indeed, in the common language, the term resilience is understood in the context of attitude towards stress and adversity. In particular, the ability to recover. It brings on the thoughts of a sportsman recovering from a defeat, a boxer standing up after a hard blow, a runner continuing a marathon after a fall.

Since the 1980s, resilience, along with robustness, has been gaining popularity as a topic. This can be seen in the figure 2.1 generated with Google Ngram viewer [29].

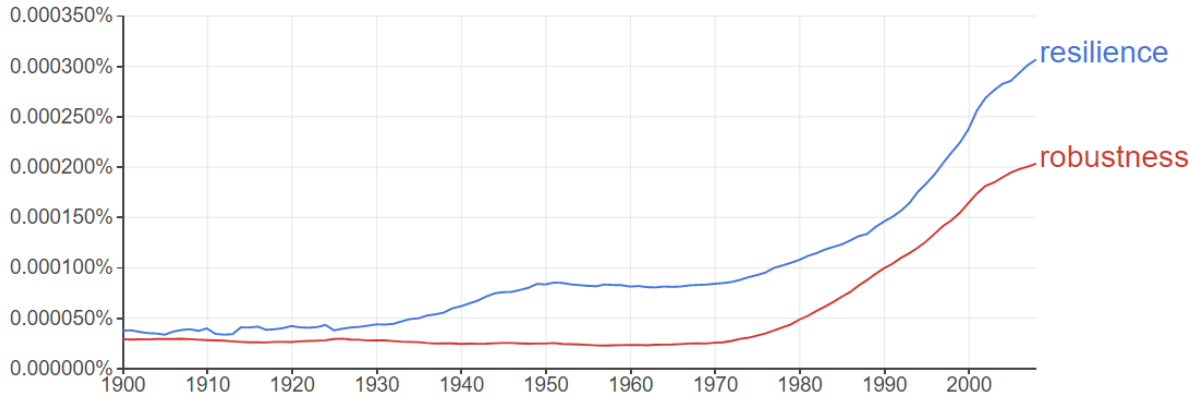


Figure 2.1: Frequency of appearance "robustness" and "resilience" terms in the English corpus of books.

The increase in popularity can be possibly attributed to a better awareness of operational issues caused by disruptions to operations. Now, not only do we require our systems to be highly optimised for their nominal operating conditions, but we also need them to be able to perform under adverse conditions. This increase led to the term becoming a buzzword. [1]

[1] https://medium.com/usaid-frontlines/insights-tom-staal-5a7ab307d818

The goal of this chapter is to introduce to the reader the concept of resilience in the scientific literature. In particular, the perspectives from which it is seen as well as the contexts in which it appears. Furthermore, a special focus will be put on its use in the field of Air Traffic Management.

2.1. Resilience capacities

In 2014 Francis and Bekera performed an analysis on definitions of resilience used throughout different domains. Based on the results, they developed the paradigm of resilience capacities [26]. They have discovered that the definitions include one or more of the following qualities: absorption, adaptation and restoration.

The absorptive capacity is the ability of a system to maintain the same operational performance despite adversities and disturbances. In practice, absorptive capacity is implemented through buffers and other mitigation strategies. An example of a buffer in aviation might be adding extra time to the expected turn around period of an aircraft to better absorb any disruptions.

The adaptive capacity is the ability of a system to rearrange it's functioning to accommodate for undesirable operating conditions. Francis and Bekera highlight that the difference between the adaptive and the absorptive capacity is the former results in a change of a system whilst the latter doesn't. An example of the adaptive capacity of the ATM system can be the redistribution of airspace sectors between controllers based on the number of aircraft within those sectors. Such an approach allows the system to adapt to changing conditions.

Lastly, the restorative capacity is the ability of a system to return to its nominal state following a disruption. In the case of ATM, it can be seen in the case of snowfall at an airport. Initially, during the snowfall, the airport capacity might become constrained or even become null. However, once the weather conditions improve to a sufficient level, the ATM system will begin an attempt to restore its nominal capacity.

The three aforementioned capacities make up, as named by Francis and Bekera, the triangle of resilience [26]. All of these capacities are required for a system to be called resilient. However, robustness and dependability can also be defined as based on those capacities. Blom and Boufara in the book Complexity science in air traffic management have introduced the relation between these terms and resilience capacities [8]. Table 2.1 presents this relation.

Related system properties	Resilience capacities		
	Absorptive	Restorative	Adaptive
Robustness	+	-	-
Dependability	+	+	-
Resilience	+	+	+

Table 2.1: Relation of resilience capacities to system properties. Table adjusted from Complexity science in air traffic management [8].

In the table above the plus sign indicates that the capacity is present in the system property. As such, robustness contains the absorptive capacity. Dependability expands robustness with the addition of restorative capacity. Resilience encompasses dependability but also includes the adaptive capacity. As such, resilience is an enriched system property as compared to dependability and robustness. The triangle of resilience with robustness and dependability is shown in figure 2.2

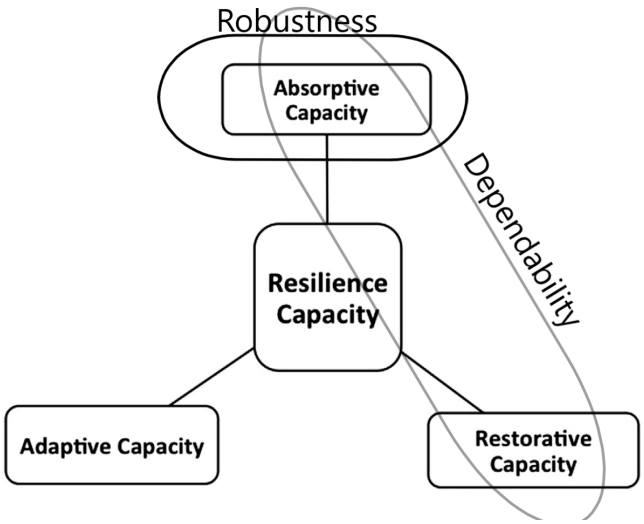


Figure 2.2: Resilience triangle with robustness and dependability marked. Figure adjusted from Francis and Bekera 2014. [26]

When considering a system through the resilience capacities paradigm it is important to note the relation of adaptability with restorative and absorptive capacities. In the case of a dependable system, absorption and restoration capacities can be measured directly. However, once adaptability is introduced into the system, measurement of absorption and restoration alone becomes impossible. This is because adaptability modifies the absorptive and restorative properties of the system. In the context of this research project, it is important to note that **any measurement of absorptive and restorative capacities will be affected by the adaptability of the system**. To capture a true state of these capacities individually, two measurements would have to be performed under the same conditions with the adaptive capacity present and missing.

2.2. Resilience engineering

When considering the safety of systems, the Swiss Cheese Model is common, a metaphorical approach introduced by James Reason in his influential paper *Human error: models and management* [53]. Namely, pieces of cheese represent safety barriers such as redundant subsystems. The holes in the cheese, are the deficiencies in the safety barrier. Thus, an alignment of the holes of multiple pieces of cheese allows an error to penetrate the defences and result in a failure.

An earlier model dating back to the first half of the 20th century is the Heinrich Domino Theory of accident causation [37]. In this metaphor, causes are domino's, which when pushed over result in a cascade of events which in the results in a failure. A very common theme in detective TV series.

The previously mentioned theories serve as a good way to "visualise" how failures occur, however, they are deficient in performing safety analysis of existing systems. This is noted in a report for Eurocontrol by Erik Hollnagel and James Reason himself in which they convey that the Swiss Cheese Model serves as means of communication. It is not able to answer the questions as to why holes appear and how do they change with time [54].

To address the critiques of existing safety management methodologies, Resilience Engineering has been introduced as an alternative in the early 2000s with the book *Resilience Engineering: Concepts and Precepts* by Woods et al [65]. The study of resilience looks at safety and performance more holistically. Thus, according to Erik Hollnagel, resilience is not just a lack of accidents and failures [17]. Furthermore, it is not only the study of what went wrong, but also what went right. Within Resilience Engineering a belief is present that inherently failures occur in the same manner as successes. In that regard, resilience is also about utilising opportunities as well as reacting to threats. Erik Hollnagel in his website calls this a transition *"... from protective safety to productive safety (...) Resilience is about*

how systems perform, not just about how they remain safe" [17].

The definition of resilience in the 2006 edition of *Resilience Engineering: Concepts and Precepts* is:

> *"The essence of resilience is therefore the intrinsic ability of an organisation (system) to maintain or regain a dynamically stable state, which allows it to continue operations after a major mishap and/or in the presence of a continuous stress [65]."*

The 2010 edition of *Resilience Engineering: Concepts and Precepts* changes the focus of the above definition from maintaining a stable state to sustaining required operations.

> *"The intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions."*

Erik Hollnagel suggests that the evolution of definitions is to broaden the meaning of Resilience. That is, to encompass also the performance under a variety of conditions as proof of resilience. Thus, he suggests a newer version definition of resilience:

> *"A system is resilient if it can adjust its functioning prior to, during, or following events (changes, disturbances, and opportunities), and thereby sustain required operations under both expected and unexpected conditions [17]."*

In this field, resilience encompasses values such as flexibility, where a system can hastily respond to disturbances to accommodate the changing conditions. However, the definition goes further than that and also includes the ability to forecast adverse events and adapt accordingly to maintain safety. Importantly, resilience is a dynamic process in which failure is a means of adapting to a situation rather than a malfunction [31].

2.3. Ecological resilience

Resilience has been defined in ecology since the 1970s when Holling published his highly cited paper *Resilience and stability of ecological systems* [35]. The ecological resilience definition is considered to be one of the most highly used definitions. A literature study conducted by Francis and Bekera found that it is common in various domains [25]. Furthermore, note the increase in usage of term "resilience" in figure 2.1 starting from 1970. He defines resilience as:

> *"The persistence of relationships within a system and is a measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist [35]."*

Hollings contribution has been also to define stability and denote its difference from resilience. From a systems theory point of view, stability relates to the fluctuations of state variables. More precisely, with the speed of recovery and fluctuations of the state variables after a disturbance. As such, a system can be resilient but not stable and vice versa. He also noted that stability and resilience tend to be conflicting in ecological systems. In general, populations with low stability tend to be more resilient in the face of extreme, fluctuating climatic conditions [35].

In 1995 Holling updated his definition of resilience to include adaptability [30]. This definition is below. Looking through the perspective of resilience capacities, the above definition includes absorptivity and adaptivity. However, it is not full resilience as it does not include the restorative capacity which is more similar to the definition of stability.

> *"Resilience is the buffer capacity or the ability of a system to absorb perturbations,or the magnitude of disturbance that can be absorbed before a system changes its structure by changing the variables and processes that control behaviour [30]."*

2.4. Resilience contexts

However, the term resilience appears in a variety of contexts. A literature review performed by Righi et al. has analysed 237 studies including the term "resilience engineering" [56]. Furthermore, domains in which resilience research was performed were numerous. Some examples are fishing, meteorology, military, nuclear power plants, health care and aviation.

Within aviation, of the articles analysed by Righi et al. almost half concerned themselves with the theory of resilience. Only 2 out of 34 dealt with Identification and classification of resilience which hints that the research in resilience quantification is needed due to the low number of studies concerning it. Furthermore, it is also noted that quantitative studies are not common to observe [56].

Aside from a variety of contexts, also a variety of properties are assigned to resilience. Usually, the properties are domain related as shown by Royce and Bekera who conducted a definition review in their article [25]. For example, within organisational systems domain resilience has properties such as: recognising unanticipated perturbations, capacity to recognise threats and ability to sustain a shock. Within Economic systems, resilience is considered to be the ability to recover, ability to withstand or capacity to survive.

As mentioned previously, aviation is also a domain that interests itself in resilience. Aviation domain resilience articles analysed by Righi et al. have been the most numerous, the understanding of resilience within ATM shall be described in the next section [56].

2.5. Resilience in Air Traffic Management

As seen from the previous sections, resilience has found it's way to a variety of domains and is perceived through different perspectives. Of course, the main interest of this literature study is to find the past and present usage in the ATM domain.

Commercial aviation overall is a safe mode of transport. The year 2014 had 7 accidents with fatalities out of 37.4 million flights flown that year [1]. This results in about 1 in 5 million chance for a flight to be fatal. This can be attributed to the safety culture present in the aviation industry.

It should come as no surprise then that the ATM industry became interested in Resilience Engineering soon after it became recognised. In 2007 a project was started by Eurocontrol to determine the role Resilience Engineering can play in ATM. A result of this project was a White Paper was published with Erik Hollnagel as the coauthor [41]. In this report, Hollnagel et al. recommend using a Functional Resonance Assessment Method for safety analysis of ATM systems.

Whilst the work of Hollnagel on Resilience Engineering is influential and important, it has the drawback of being purely qualitative. Furthermore, most of the research on resilience in ATM has been thus far performed from the safety perspective. A need exists for a quantitative framework that will allow dealing with non-safety related perspectives.

Such a need was noted by Olga Gluchshenko in a DLR report published in 2012 which was in 2013 expanded into an article [27, 28]. As Gluchshenkos work is highly relevant to the topic of this literature study, her work will be presented in-depth.

Gluchshenko et al. introduce another definition of resilience with the focus on non-safety related disturbances:

> *"The ability of a system to respond on a **disturbance** within a time horizon by **transient perturbation**, i.e. the system is resilient against the disturbance over the considered time horizon; is relative to the specified **reference state** of the system and to a particular disturbance [28]."*

To understand this definition, the words in bold font have to be described. As such, consider a tree standing upright when there is no wind present. It can be assumed, that the upright position of the tree is the **reference state** of the system. This is because this is the state of the system when no **disturbance** is present. In this metaphor, the wind is the disturbance.

Now imagine, the wind picks up which results in the disturbance being present. If the wind is strong enough the tree will sway from its vertical position. This sway is the **perturbation**,

a departure from the reference state. If the wind is not very strong, the tree will return to its vertical position after the disturbance has subsided. This is called a **transient perturbation**. However, a **permanent perturbation** would occur, if instead, the wind was strong and the tree has undergone what could be considered a plastic deformation. As such, a new reference state would be achieved that is no longer upright.

Gluschenko et al. also introduces a term closely linked with perturbation which is **stress**. While perturbation is referred to as the ”action of a system”, the stress is the ”reaction of a system” [28]. It is the departure from a reference state to a disturbed state. In the authors' opinion, stress according to Gluschenkos definition is the measure of perturbation. As such, in the metaphor of the tree in wind, the stress will be the deviation of the tree from the vertical measured in meters. Such stress is called **survival stress** as the system manages to maintain it's inherent properties. However, if the tree was to brake, changing the system properties, it would have experienced **lethal stress**.

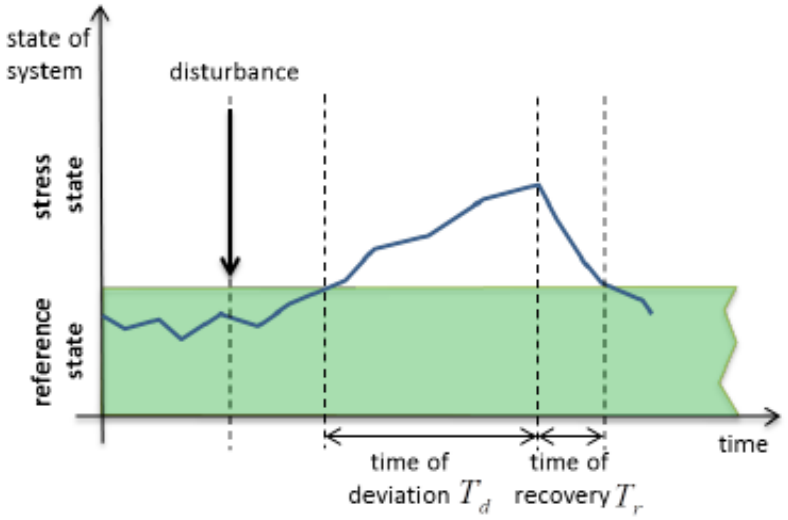
Additionally, the topic of resilience versus robustness is tackled by Gluschenko et al. A 2015 literature review of publications containing the term ”resilience engineering” shows that the term resilience frequently occurs with robustness [56].In the aforementioned metaphor, a tree is more robust if it can experience a higher wind before it begins to sway. Thus, the definition of robustness is:

> *”The ability of a system to experience no stress since a disturbance had occurred, i.e. the system is robust against the disturbance over the considered time horizon; is relative to the specified reference state of the system and to a particular disturbance [28].”*

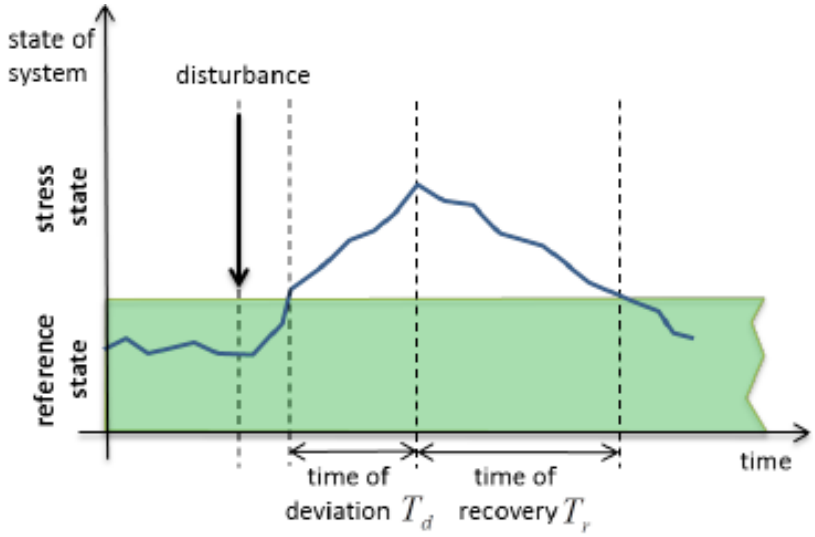
Gluschenko suggests measuring resilience by the recovery time from disruption. In particular, three categories of resilience are suggested. High resilience where the time of recovery is shorter than the time of disruption, this is presented in figure 2.3a. Medium resilience where the time of recovery and time of disruption are approximately equal, figure 2.3b. Lastly, low resilience when recovery time is longer than the time of disruption, figure 2.3c.

While the resilience categorisation proposed by Gluschenko et al. is simple to understand it lacks consideration for the magnitude of deviation. This can be understood by two scenarios relating to airport operations during winds. The system is described by the number of flight cancellations at an airport. According to the aforementioned categorisation and definition of resilience, two airports will have the same resilience if their relation between time of deviation and time of recovery is the same. However, one of the airports might have fewer cancellations compared to its reference state, overall, handling the high winds better than the other airport.

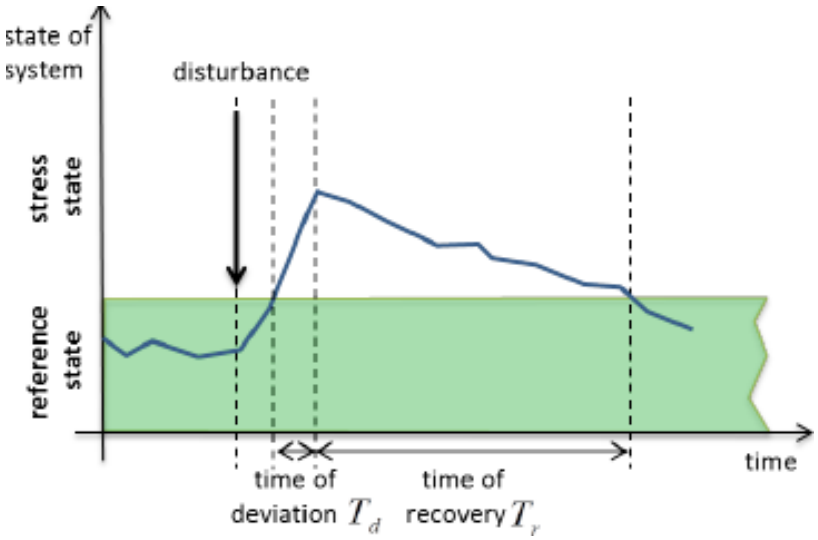
Despite the deficiencies, this work presents an intelligible abstraction of ATM resilience. Gluschenko et al. present a methodology for measuring resilience and robustness which can be seen in figure 2.4 [28]. From figure 2.4 indicators, time horizon and reference state arise as critical elements in measuring resilience. Indicators should be able to capture the state of the system appropriately. The time horizon has to be specified as the system response to long term events such as the winter period will be different than short term events like high winds. Lastly, the reference state is of importance as resilience and robustness can only be measured when the indicators are compared to a value representative of the nominal state.



(a) High resilience, recovery time is significantly longer than time of deviation.



(b) Medium resilience, recovery time is similar to time of deviation.



(c) Low resilience, recovery time is significantly longer than time of deviation.

Figure 2.3: Resilience types. Figures from [28].

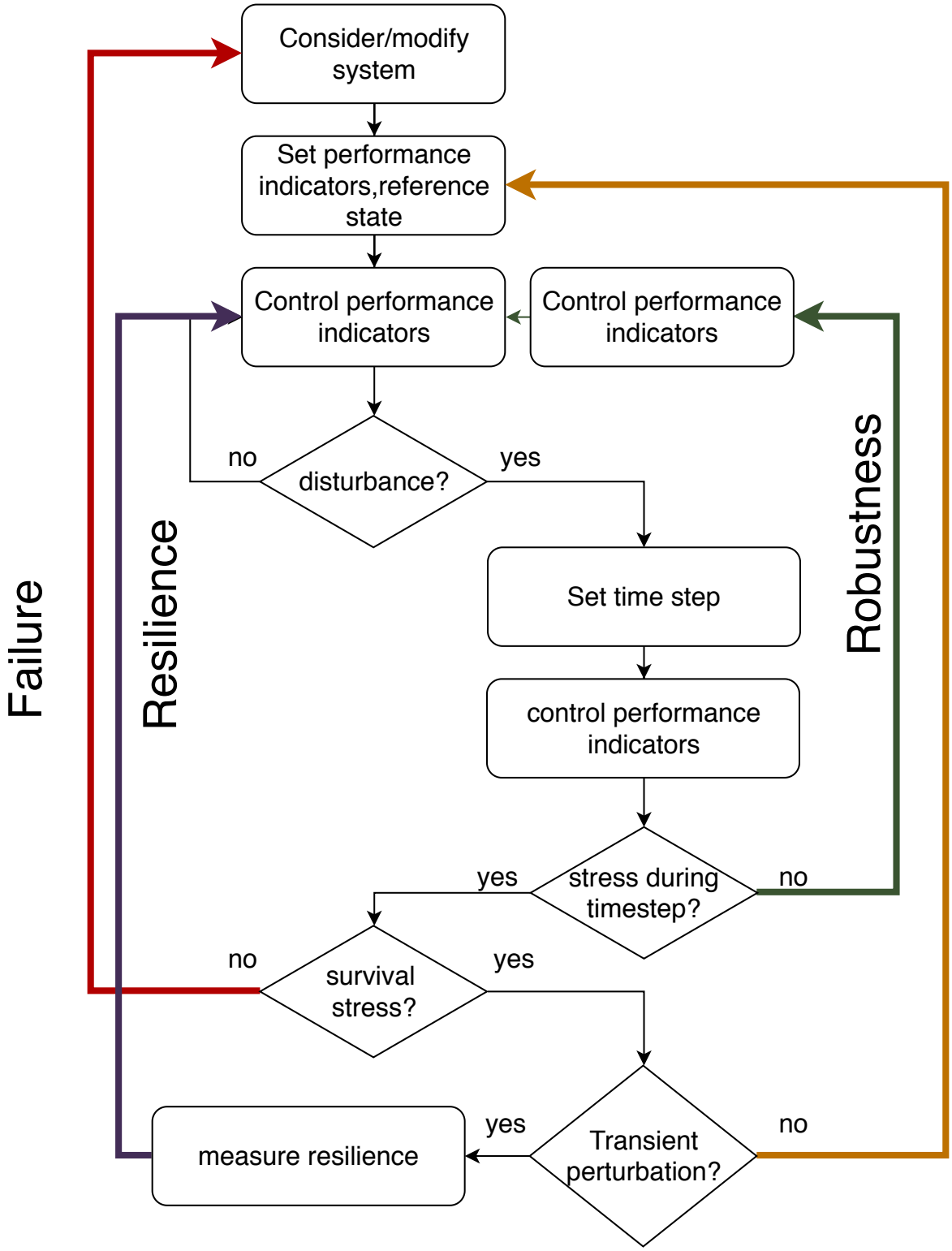


Figure 2.4: Methodology for measuring resilience and robustness. Figure adapted from Gluschenko et al. [28].

3
Metrics

With the deregulation and corporatisation of the aviation industry more importance has been placed on performance. As such, performance-based air navigation became of interest to the industry. The main characteristic of this approach is the a result oriented approach. In particular, through the use of objectives. Furthermore, the achievement of the objectives is done through decision-making based data with the focus on obtaining the objectives. The 2009 *Manual on global performance of the Air Navigation System* published by ICAO states that the performance based approach can be used throughout the whole spectrum of activities within aviation [38]. Thus, also including ATM.

As such, before attempting to improve airspace resilience, there must be a methodology for measuring it. Thus, there is a need for metrics. A variety of metrics, sometimes referred to as Key Performance Indicators (KPI) have been developed for measurement of different aspects of the ATM system. The goal of this chapter is to introduce the reader to metrics developed up to date. The first section will focus on metrics in ATM which include complexity metrics and throughput metrics. Lastly, resilience metrics, not necessarily from ATM domain, will be shown.

3.1. ATM KPIs - ICAO

The 2005 report by ICAO *Global Air Traffic Management Operational Concept* defined eleven Key Performance Areas (KPAs) of Air Traffic Management [50] . Subsequently, the aforementioned 2009 Manual also by ICAO included a comparison between KPIs used by two ATM organisations [38]. Each KPI considered in the report pertained to one of the eleven KPAs. The KPAs are: Access and Equity, Capacity, Cost Effectiveness, Efficiency, Environment, Flexibility, Global Interoperability, Participation by the ATM community, Predictability, Safety and Security.

Overall 2009 comparison indicated a lack of cohesion in indicators used. The discrepancies occurred due to multiple factors. One, different filtering criteria was used to obtain the metrics. For example, measurements would be taken at specific locations or specific times Two, values were normalised in different manners. Three, statistical operations were applied such as windowing, averaging etc. Lastly, the definitions of terms varied between organisations.

The above discrepancies yield a useful conclusion for the research project. Due to the available ADS-B data, a comparison between multiple airspaces can be conducted using the exact same metrics and methodology. Allowing for a useful comparison that would otherwise be unavailable based purely on ANSPs data.

Of the eleven KPAs, Capacity and Efficiency are of most importance to this research project. This is mainly due to the available data and the context of the other areas. For example, Access and Equity, Global Interoperability and Participation by the ATM community are business related. Additionally, safety and flexibility KPAs are considered on a higher level. I.e. safety is defined through the number of accidents which cannot be easily derived

from ADS-B data.

3.1.1. Capacity KPA

The Capacity KPA is divided into three categories, System-wide, Airspace and Airport capacities [38]. The first, can be the number of flights accommodated within the entire system. However, such approach is problematic as the estimation of maximum capacity is complex. Furthermore, the scalability of this approach is problematic as the same methodology would need to be applied to different ATM systems. A suggested alternative is to measure the actual number of flights flown, distance covered etc. However, the critique of such approach is that it does not measure the actual system capacity but rather the conditions in which it operates. As such, in economic terms what is being measured is the supply.

The airspace level metrics include the number of IFR flights that can be handled at one time. However, similar issue arises as to system-wide metrics because it is hard to judge objectively how many aircraft can enter an airspace volume simultaneously.

The airport indicators have the most commonality between organisations. Most of the indicators include IFR movements defined as number of arrivals plus departures. However, aggregations are also used to for example compute movements for whole metropolitan areas. Furthermore, they values can also be presented based on meteorological conditions.

3.1.2. Efficiency KPA

The KPAs of efficiency are mostly computed through ”Temporal Efficiency” [38]. That is, consideration is given mostly to delays. As such, metrics include fraction of flights departing on time, average departure delays and etc. The main issues arising in the case of efficiency indicators are the thresholds as to what is considered a delay. Furthermore, certain delays are omitted such as delays due to weather.

3.2. ATM KPIs - CANSO

Another perspective on ATM KPIs is offered by the Civil Air Navigation Services Organisation (CANSO). This approach is based upon ICAO KPAs: capacity, efficiency and predictability [14]. Whilst the aforementioned ICAO report *Global Air Traffic Management Operational Concept* serves as a high level guideline providing no specific solutions, the CANSO report *Recommended Key Performance Indicators for Measuring ANSP Operational Performance* in turn provides KPIs for each KPA.

3.2.1. Capacity KPIs

Declared capacity is the theoretical maximum that can be handled by an airspace sector or airport. Factors that affect the declared capacity include: airline schedules, aircraft mix, delay tolerance, ground infrastructure, staffing, airspace restrictions, weather, passenger infrastructure. Operationally, declared capacity may be utilised as a fixed value for the airspace. However, an adaptive approach is also possible whereas the ANSP declares the daily maximum based on known operation factors [14].

Knowledge of the declared capacity is important as it allows the ANSP to benchmark it’s operations. However, the computation of this value is nontrivial and depends upon multiple factors. Furthermore, in the context of this research project the declared capacity cannot be derived with certainty from collected ADS-B data. However, for airports close to operational limit such as Amsterdam Schiphol an approximation of declared capacity can be derived with the assumption that the airport operates at it’s highest operational capacity. Thus, when observing the capacity over a long period of time a potential upper limit may be detectable.

Capacity utilisation is a measure of effectiveness of capacity usage by the ANSP. This KPI requires two elements, one is the aforementioned declared capacity and the second is the actual capacity usage at a given time. As such, it also suffers from the non-trivial determination of declared capacity [14].

However, the actual capacity usage can be easily computed with access to ADS-B data. This can be done by counting unique aircraft IDs within a geographically filtered dataset corresponding to the sector of interest. As mentioned previously, the declared capacity could

be estimated using historical values.

Delay Attributed to Capacity KPI is a measure of how over demand or, inversely, under supply of airspace capacity impacts delays. The computation of this KPI requires knowledge of aircraft schedules as well as association of delay with excess demand [14].

A potential method of computing this KPI may be possible by comparing a nominal approach/descent time with actual approach/descent times. However, this is exclusive for Terminal Control Areas and results in a pseudo delay.

3.2.2. Efficiency KPIs

Efficiency KPIs are used to determine the efficiency of five different flight phases: taxi-out, departure, en-route, arrival, taxi-in. Whilst the report also mentions gate departure and gate arrival as flight phases, they do not fall under the jurisdiction of the ANSP but rather the airline itself and ground handlers. As such, they will not be considered further in this report [14].

A common method to demarcate those flight phases is by setting a radius around an airport to indicate areas in which arrivals/departures occur. As such, flights are considered departing if they are within 40 nautical miles of the airport whilst arriving flights are set to be within 100 nautical miles of the airport. However, some ANSPs chose to set both of those limits to 40 nautical miles [14].

Taxi-out and taxi-in KPIs both take a similar form of comparing a nominal taxi time to the actual values experienced by the aircraft. Factors that play a role in those KPIs include aerodrome layout, airport staffing, current ATM situation and availability of parking slots. It must be noted that the taxi-in process is less affected by the ATM situation. Furthermore, availability of parking slots is not a factor for taxi-out time [14].

The KPIs for these flight phases may include: number of aircraft delayed, delay time, distribution of delays and etc. However, a simplification can be used whereas not the delay is considered but simply the actual taxi time. Furthermore, a zero-velocity time as part of total taxi time can also be an indicator of inefficiencies in the taxi process.

Theoretically, these values can be easily and accurately derived from ADS-B data. However, in practice the difference in Standard Operating Procedures of airlines and individual pilot choices will impact taxi time measurement from ADS-B. This is because there is no standard on when to turn on the ADS-B transponder. For example, some aircraft may first begin transmitting their position just before the runway whilst others already during push back.

Departure KPIs include two separate segments: take-off and departure. For the former, a comparison can be made between the Calculated Take-off Time and the actual take-off time. For the latter, multiple options are possible. They include: average delay within departure zone, average excess horizontal distance per flight, average excess time per flight.

The required data for the computation of this KPI include the departure time, aircraft trajectory and crossing time of ring of 40NM radius. Furthermore, a baseline must be set for computation of excess times and distances. Aside from the baseline, these values can be easily derived per flight utilising ADS-B data as aircraft are usually transmitting whilst on the runway. However, the baseline values would need to be either preset based on unattainable ideal conditions such as ideal horizontal distance being equal to 40NM. Historical analysis can also be used to determine nominal conditions although data for an appropriately long amount of time must be available.

En-route KPIs are defined in terms of comparison of actual distance flown between Terminal Control Area exit and arrival points and the ideal distance. For short routes, the great circle distance is a good approximation for ideal distance. However, numerical optimisation studies have shown that wind-optimised routes tend to be a better solution for longer flights [39]. Additionally, the excess time can also be computed if an assumption is made that the aircraft would travel the ideal route at the same average velocity as the actual route.

The ADS-B data is well suited for computation of this KPI considering that the update time of ADS-B is 0.5 seconds and thus at least an order of magnitude lower than radar [67]. Additionally, ADS-B data tends to be more accurate as compared to radar [67]. However, for longer routes the ideal distance is non-trivial and thus hard to obtain. For shorter routes

however the great circle approximation can be used. Furthermore, excess distance capture by this KPI might result from airspace restrictions such as military zones.

The arrival KPIs include both the descent phase as well as the final approach and landing. A metric for the latter can be the runway occupancy time computed as the difference between wheels on ground and runway exit. This will be influenced by aircraft category as well as presence of high-speed exit ramps. For the descent phase two metrics can be derived, first, ground track inefficiency which could be defined as ground track distance starting from 100 NM away from an airport till touch down, divided by 100 NM. Second, excess distance and time can be derived when comparing the actual flown path to an ideal value.

CANSO has developed a methodology for computing such ideal arrival routes based on clustering of flights by aircraft type and conditions [13]. Overall, this method is similar to the aforementioned derivation of nominal conditions based on historical data. Both, the computation of the nominal approach path as well as the actual approach path can be performed using ADS-B.

3.2.3. Predictability KPIs

Predictability, as defined by ICAO, is the *"ability of airspace users and ANSPs to provide consistent and dependable levels of performance"* [50]. As such, the predictability KPIs deal primarily with variability in services provided by the ANSP. This is of importance for airspace users as it allows for accurate planing of operations without unnecessary buffers to account for uncertainty. CANSO defines three predictability indicators [14].

Capacity Variation KPIs measures the variability in capacity. They can be implemented as a standard deviation of a capacity for a desired time period. Furthermore, this metric will give an indication of sector or airports with high delays. This metric can be implemented based on ADS-B data as aircraft/operation counts can be derived from it [14].

Travel Time Variation KPIs can be defined based on the time scale considered. On a strategic scale they can be derived based on the schedules submitted by the airlines. Such a measure would give an insight about airlines subjective judgement on the predictability of air facilities. For this KPI access to airline schedules must be obtained. Thus, ADS-B cannot be used to derive it.

However, on the tactical scale, the time variability per flight-phase can computed. For example, for arriving flights the time between crossing the 100NM ring and landing may be considered for variability. If total flight time between airport pair is considered, this measure can give an indication of routes where buffers are applied to schedule to account for possible delays. Furthermore, the per flight-phase variability can be computed utilising ADS-B data.

Flight Plan variation KPIs can be derived from flight plans and optionally actual trajectories flown. If considering only the flight plans between an airport pair, the last pre-departure flight plan can be compared to the the other flights in other to determine the variability. When combining flight plans with trajectories the so called "filed versus flown" measures can be derived. However, all of those metrics cannot be derived purely based on ADS-B.

3.3. Airspace Congestion Metrics

The purpose of Airspace Congestion Metrics is to quantify the difficulty of an air traffic situation as perceived by an Air Traffic Controller (ATCo) supervising the sector. Being able to determine this difficulty is critical in assessing how many aircraft can a ATCo manage safely. If a situation is simple then an ATCo is able to handle additional aircraft in the sector. Conversely, if the situation is difficult then aircraft may need to be prevented from entering the sector.

However, this difficulty is not only dependent upon the number of aircraft airborne. Consider figure 3.1 which presents easy, average and difficult traffic situations with each having the same amount of aircraft. Intuitively, a reader with no air traffic management experience will be able to imagine that handling the right most situation will be more difficult than handling the left most.

Thus, the distribution of aircraft positions, velocities and headings plays a significant role. Overall, from figure 3.1 it is seen that the larger the chaos of the situation, i.e. less uniformity

of the aforementioned factors, the more difficult the traffic situation. As such, more advanced complexity metrics attempt to include such factors as well.

Delahaye and Puechmorel in their book *Modeling and Optimization of Air Traffic - Airspace Congestion Metrics* suggest three approaches for modelling traffic situation complexity [16]. They are: flow-based, speed vector distribution based and dynamic system based metrics.

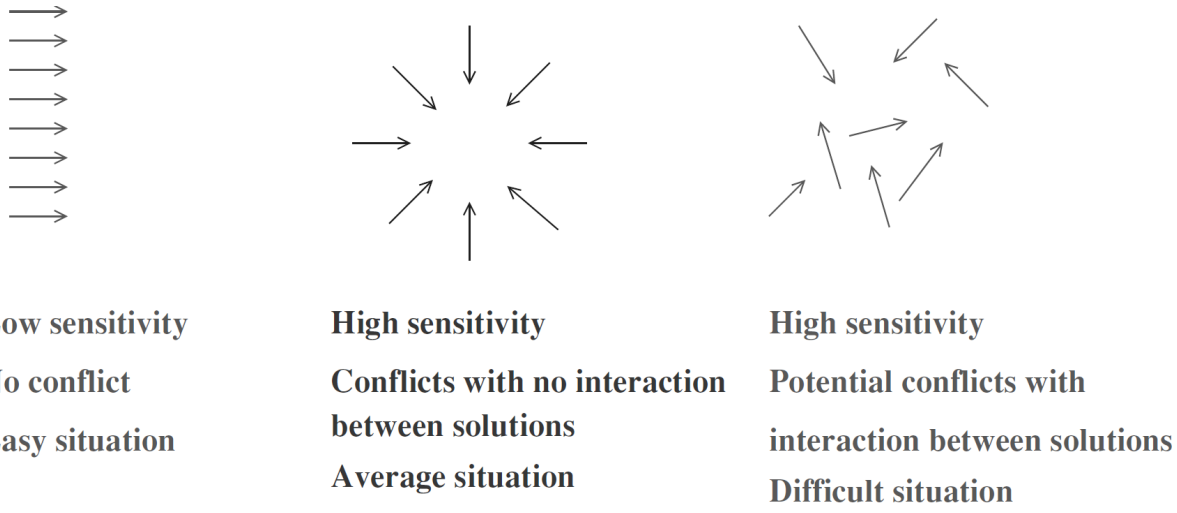


Figure 3.1: Three Air Traffic situations. From left: Easy, Average, Difficult situations [16].

3.3.1. Flow-based metrics

Flow-based metrics work by considering the air transport system through a network perspective. In this abstraction, airways become the edges of a graph whilst airports and airway junctions become the nodes. Based on a survey Delahaye and Puechmorel suggest three types of workloads: conflict workload, coordination workload and monitoring workload [16].

The **conflict workload** is based on the need for the ATCo to resolve conflicts at nodes of the graph. In other words, at the intersections of aircraft paths. An implicit assumption of the method proposed is that the aircraft arriving at a node follow a Poisson Process. The Poisson Process models a discrete set of events where the timing between events is independent of the previous occurrence of events. Furthermore, this model assumes that only the average time between events is known [44].

Consider figure 3.2 which presents a schematic drawing of airways ij and lj intersecting at node j. f_{ij} and f_{lj} are aircraft flows on airways ij and lj respectively. θ_{ilj} is the angle between two airways crossing at Node j.

Using the aforementioned Poisson Process model, equation 3.1 can be derived which presents the average number of conflicts at the node. Variables V_{lj} and V_{ij} are the average velocities of aircraft travelling along the airways. N_s is described by Delahaye and Puechmorel as a standard separation norm [15]. Whilst no further description is given, the author believes that this is a separation between aircraft in units of time. I.e. 2 minute, 180 seconds etc.. Thus, the unit of f is expected to be aircraft per unit time. As such N_c will be a value of conflicts per unit time.

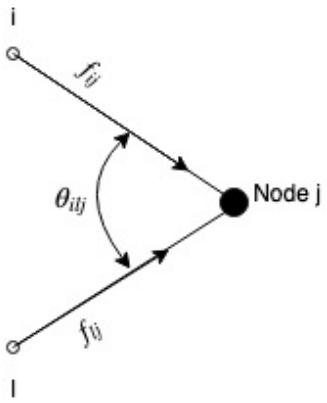


Figure 3.2: Crossing of airways schematic [16].

$$N_c = \frac{2N_s\sqrt{V_{lj}^2 - 2V_{ij}V_{lj}\cos\theta_{ijl} + V_{ij}^2}}{V_{ij}V_{lj}\sin\theta_{ijl}} f_{ij}f_{lj} \tag{3.1}$$

It can be assumed that the conflict number is proportional to controller workload. Furthermore, if multiple airways are crossing at the junction, the workload is considered to

be the sum of workloads of all pairs of airway combinations. In equation 3.3 $C_{cf}(j)$ is the controller workload whilst α_{ijl} is defined in equation 3.2.

$$\alpha_{ijl} = \frac{2N_s\sqrt{V_{lj}^2 - 2V_{ij}V_{lj}\cos\theta_{ijl} + V_{ij}^2}}{V_{ij}V_{lj}\sin\theta_{ijl}} \tag{3.2}$$

$$C_{cf}(j) = \frac{1}{2}\sum_{\substack{i \in \mathcal{N} \\ i \neq j}} \sum_{\substack{l \in \mathcal{N} \\ l \neq i; l \neq j}} \alpha_{ijl} f_{ij} f_{lj} \tag{3.3}$$

However, an ATCo might have more nodes in his sector than just one. Thus, the total workload is considered to be the sum of nodes in sector S_k. The set of nodes in sector S_k is represented by $\mathcal{N}_k$. Equation 3.4 describes this mathematically.

$$C_{cf}(S_k) = \frac{1}{2}\sum_{j \in \mathcal{N}_k} \sum_{\substack{i \in \mathcal{N} \\ i \neq j}} \sum_{\substack{l \in \mathcal{N} \\ l \neq i; l \neq j}} \alpha_{ijl} f_{ij} f_{lj} \tag{3.4}$$

The **coordination workload** is associated with handling of aircraft crossing between sectors. In such a situation, the ATCos of the neighbouring sectors must coordinate the transfer by agreeing whether the transfer can actually happen and how it should happen. To visualise this, consider figure 3.3 which presents a schematic scenario of airways and an airspace sector [16].

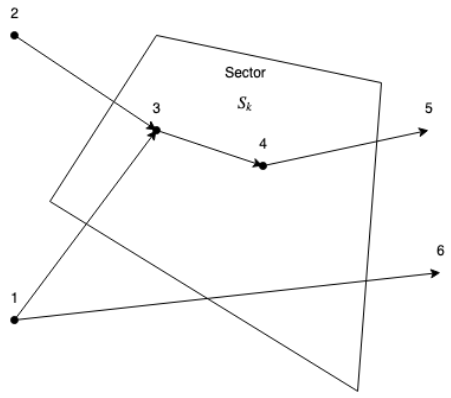


Figure 3.3: Visualisation of airspace borders and airways.

Three possibilities of intersection between an airway and sector exist. One, the airway is fully within the sector, in figure 3.3 this is airway 3-4. Two, one of the airway ends is within the sector whilst the other end is not, eg. airway 4-5. Lastly, both of the ends of an airway can be outside of a sector however part of the airway still crosses within a sector such as airway 1-6.

For the coordinator workload only cases two and three result in a need to perform handovers. As such, only those cases generate workload. Thus, the coordination workload can be describe by equation 3.5. $\mathcal{A}_k$ is a set of airways with both ends outside of the sector but with a segment within. $\mathcal{N}_k$ are again the nodes within the sector.

$$C_{co}(S_k) = \sum_{i \oplus j \in \mathcal{N}_k} \beta_{ij} f_{ij} + \sum_{\substack{i \notin \mathcal{N}_k \\ j \notin \mathcal{N}_k \\ (i,j) \notin \mathcal{A}_k}} 2\beta_{ij} f_{ij} \tag{3.5}$$

The first term represents the workload from airways with one end within the sector. The mathematical symbol $\oplus$ represents exclusive OR (XOR) operator. f_{ij} is the same as in conflict workload representing aircraft flow. Term β_{ij} is a weighing coefficient used to adjust the coordination workload with respect to other workloads.

The second term includes the workload due to airways with only a segment crossing through the sector. The multiplier ”2” is added to represent a need for a double handover, while going into the sector and while going out.

Lastly, the **monitoring workload** is the effort of the ATCo to maintain watch over aircraft not currently posing any conflicts or not entering/leaving the sector. This effort can be describe with equation 3.6.

$$C_{mo}(S_k) = \eta \sum_{(i,j)\in\mathcal{L}_k} \frac{l_{ij}}{V_{ij}} f_{ij} \tag{3.6}$$

$mathcall L_k$ represents the set of airways within or intersecting with the sector S_k. l_{ij} is the length of the airway within the sector. η is a coefficient to control the relative importance of monitoring effort.

The three aforementioned workloads can be summed together to achieve the total sector workload. Whilst relatively simple, the flow based method has it's deficiencies. Firstly, it is not simple to extract aircraft flows at particular airways from ADS-B data. This requires a process of matching ADS-B messages to airways. Furthermore, N_s can only be estimated without knowing the exact procedures. Lastly, the network based approach requires knowledge of airway and waypoint positions. As such, this method will not work in Free Route Airspace which is becoming more common within Europe.

3.3.2. Geometrical

Geometrical approaches rely on quantifying the air traffic complexity by considering the geometrical distribution of aircraft and their velocity vectors. Such approach is beneficial as compared to flow based method as knowledge of airway structure is not necessary. Furthermore, in some scenarios aircraft do not follow airways.

Traffic density

Hilburn notes that the most common metric in this approach is Traffic density [33]. Density is simply computed as the number of aircraft within an airspace over the volume of the airspace. For equal volumes, when for example simplifying an airspace as a cylinder the division by volume can be skipped in order to utilise the aircraft count only.

Track inefficiency

A metric recommended by the Civil Air Navigation Services Organisation is the ground track inefficiency [14]. This metric quantifies the efficiency of aircraft track by integrating the path to obtain ground distance covered. This distance is then normalised by a reference track distance. This reference could be the nominal distance or an ideal distance such as a great circle between airspace edge and airport.

The track inefficiency can be calculated as in equation 3.7

$$\frac{\sum_{i=1}^{n} dist(\vec{x}_i, \vec{x}_{i-1})}{100nm} \tag{3.7}$$

where $100nm$ is the reference distance and $dist$ is a numerical distance computation function between two points such as euclidean distance or Haversine distance as in equation 3.8.

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_i - \varphi_{i-1}}{2}\right) + \cos(\varphi_i)\cos(\varphi_{i-1})\sin^2\left(\frac{\lambda_i - \lambda_{i-1}}{2}\right)}\right) \tag{3.8}$$

Proximity

With proper scaling aircraft proximity can also be quantified [16]. The proximity metric is based on a distance vector defined in 3.9. The goal of the proximity metric is to give insight into the clustering of aircraft which the simple traffic density does not give.

$$d_{ij}^{a,h} = \left\|\overrightarrow{p_i} - \overrightarrow{p_j}\right\|_{a,h} = \sqrt{\frac{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}{a^2} + \frac{\left(z_i - z_j\right)^2}{h^2}} \tag{3.9}$$

The horizontal and vertical relative distances are normalised by distances characteristic for those dimensions. For example, a can be set to an aircraft spacing distance within the Terminal Manoeuvring Area and h can be set to vertical spacing. The normalised distance is computed between each pair of aircraft and is then summed as in equation 3.10.

$$P(i) = \sum_{j=1}^{N} e^{-\alpha d_{ij}^2} \tag{3.10}$$

α is a scaling factor to control importance of distances.

Convergence

The convergence metric goes further in quantification of aircraft clustering by including also relative aircraft velocity vectors to the proximity. This means that the metric can capture the organisation of the relative aircraft movements. The variation in relative distance is defined as:

$$r_{ij} = \frac{\partial}{\partial t} \left\| \vec{p}_{ij} \right\|_2 = \frac{\partial}{\partial t} \sqrt{\vec{p}_{ij} \cdot \vec{p}_{ij}} = \frac{\vec{p}_{ij} \cdot \vec{v}_{ij}}{d_{ij}} \tag{3.11}$$

where $\vec{p}_{ij}$ is the relative position vector and $\vec{v}_{ij}$ is the relative velocity vector. d_{ij} is the distance from equation 3.9. As in the case of proximity, the convergence values are computed per aircraft pair and summed. However, only pair with negative relative distance are used in the summing as those are the aircraft converging. The summing is presented in equation 3.12 and as in the proximity equation a scaling factor α_c is used.

$$Cv(i) = \lambda_c \sum_{j/r_{ij} \leq 0} -r_{ij} \cdot e^{-\frac{1}{2}\left(\alpha_c \cdot d_{ij}\right)^2} \tag{3.12}$$

3.3.3. Dynamic systems

A third type of airspace congestion metrics are based on dynamic systems theory [16]. In such methods, the airspace is modelled as a dynamic system which has a state and a set of inputs affecting the state. The general case can is described by equation 3.13. Vector $\vec{X}$ is a positional vector of form $[x, y, z]^T$.

$$\dot{\vec{X}} = \vec{f}(\vec{X}, t) \tag{3.13}$$

Equation 3.13 describes a model which is non-linear and time variant. It can be further simplified on time-invariant cases respectively in equations 3.14 and 3.15.

$$\dot{\vec{X}} = \vec{f}(\vec{X}) \tag{3.14}$$

$$\dot{\vec{X}} = \mathbf{A} \cdot \vec{X} + \vec{B} \tag{3.15}$$

Given multiple aircraft positions and their velocity vectors the guiding equations $\vec{f}$ or in the linear case matrices $\mathbf{A}$ and vector $\vec{B}$ can be derived by minimising the error between the aircraft velocity vectors and the state equations. The advantage of such methods is that they are better suitable to trajectories not just instantaneous vectors. Their disadvantage is their perceptual complexity and difficulty of implementation. As such, the exact guiding equations shall not be mentioned in this report however they are available in literature [16].

3.4. Resilience Metrics

Quantitative methods of resilience assessment can be categorised into general measures and structure-based models. The former method is based on measuring system performance prior to and following a disturbance. The advantage of this method is that it does not require knowledge of the system structure. Contrarily, the structural-based method requires the knowledge of systems characteristics. The advantage of structural-based methods is the ability to study how the structure affects resilience [36].

General measures can be categorised into probabilistic and deterministic approaches. Structural-based models consists of optimisation, simulation and fuzzy logic models. The schematic drawing of this categorisation can be seen in figure 2.2.

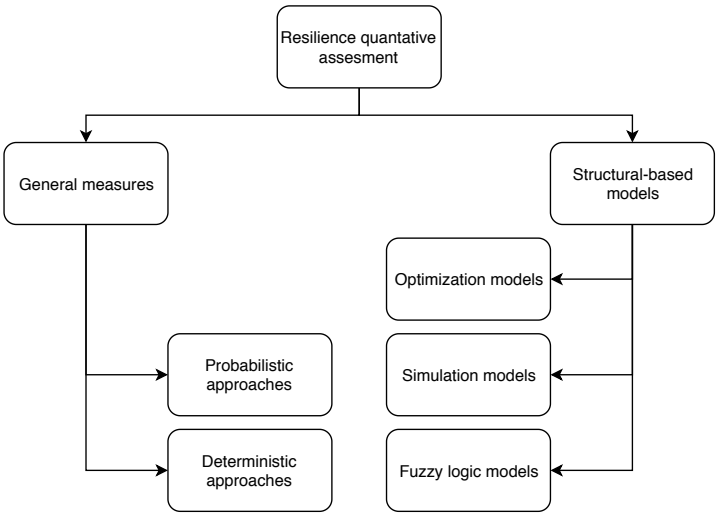


Figure 3.4: Categorisation of quantitative resilience assessment methods. Figure adapted from *A review of definitions and measures of system resilience* by Hosseini et al. [36].

3.4.1. General measures

Unlike the probabilistic method, the deterministic one does not measure the stochastisity in the system behaviour. I.e it ignores uncertainties related in the system response. This may be considered an advantage considering that probabilistic data is not required for their computation. Conversely, the probabilistic aspect provides more insight into complex systems which frequently behave stochastically [36]. Furthermore, Hosseini et al. note that metrics can capture dynamic or static system behaviour. Where dynamic behaviour includes the temporal aspect of response while the static approach is free of it.

Deterministic Approaches

Bruneau et al. proposed a static resilience metric based on the concept of quality $Q(t)$ [9]. Quality can represent performance metrics which in the ATM context could be for example airspace capacity. In this method, resilience is defined by equation 3.16. A larger value of RL, resilience loss, indicates lower resilience. This can be graphically seen in figure 3.5.

$$RL = \int_{t_0}^{t_1} [100 - Q(t)]dt \qquad (3.16)$$

In the above equation t_0 is the time when disruption occurs whilst t_1 is the time when

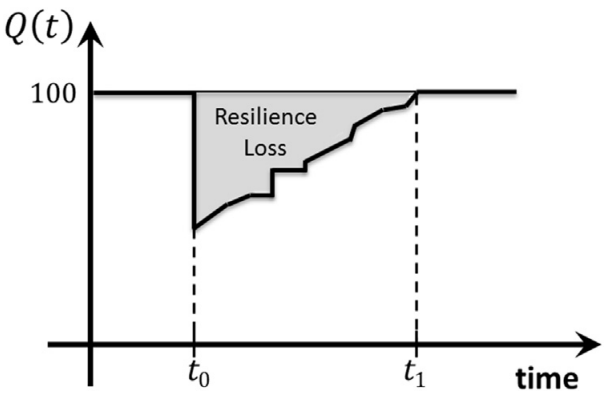


Figure 3.5: Resilience loss after a disruption. Figure from [36].

Q returns to it's pre-disruption state. This method has two assumptions, one, that the quality of the system before the disruption occurred is 100. Two, a single disruption is present. These assumptions come from the context of the metric which is earthquake impact upon infrastructure. The advantage of the above method is it's simplicity and general applicability. Furthermore, the metric due to abstract units can be hard to comprehend.

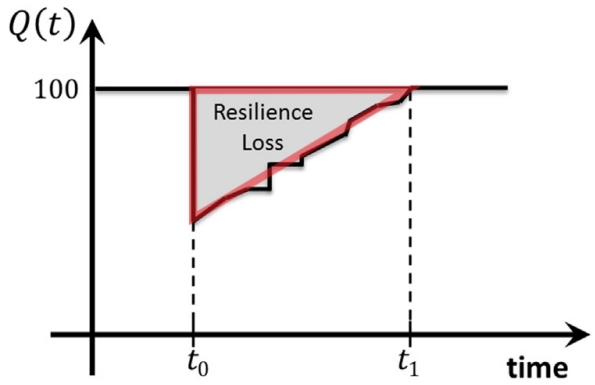


Figure 3.6: Resilience loss approximation, marked by red triangle. Figure adapted from [36].

Zobel offers a different method based on resilience triangles and predicted maximum loss in order to set a comparative baseline for disruptive events. The resilience triangle is an approximation of resilience loss suggested by Bruneau et al. and can be seen in figure 3.6 [9, 68].

The resilience in this method is described by equation 3.17. X is the maximum loss of quality, T is the recovery time whilst T^* is the maximum predicted time the recovery could take. T^* is used as a baseline in this case. The concept of maximum resilience loss is presented in figure 3.7 by the dark blue shaded polygon between t_0 and $t_0 + T^*$. This approach relies on the assumption that recovery of the system begins directly after the disruption has occurred.

$$R(X,T) = \frac{T^* - \frac{XT}{2}}{T^*} \tag{3.17}$$

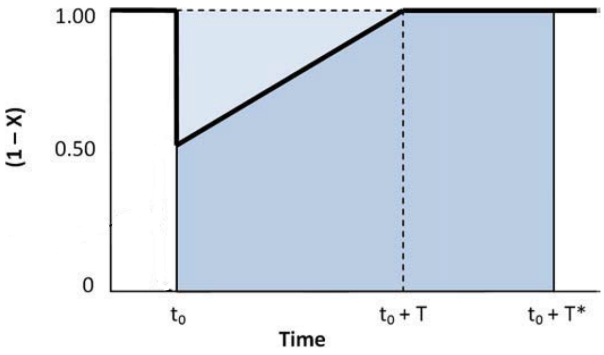


Figure 3.7: Resilience triangle and polygon of maximum resilience loss. Figure from [68]

A dynamic metric was proposed by Henry and Ramirez-Marquez which quantifies resilience at time t by comparing the recovery to the loss experienced for a given disrupting event e_j [32]. Resilience $R_F(t|e_j)$ for performance metric F is described in equation 3.18. Intuitively, this resilience measures the proportion of recovered state to the initial loss.

$$R_F(t|e_j) = \frac{F(t|e_j) - F(t_d|e_j)}{F(t_0) - F(t_d|e_j)} \; where \; t \in (t_d, t_f) \tag{3.18}$$

The notations in equation 3.18 are explained in figure 3.8. t_e is the time at which the disruptive event occurs, t_d is the time at which the system is in the final disrupted state, t_s is when recovery beings to take place and tf is the time at which the systems enters a post recovery equilibrium. F is a performance metric.

The method from equation 3.18 in it's proposed form does not give information about the rate of recovery but rather how much has the recovery progressed.

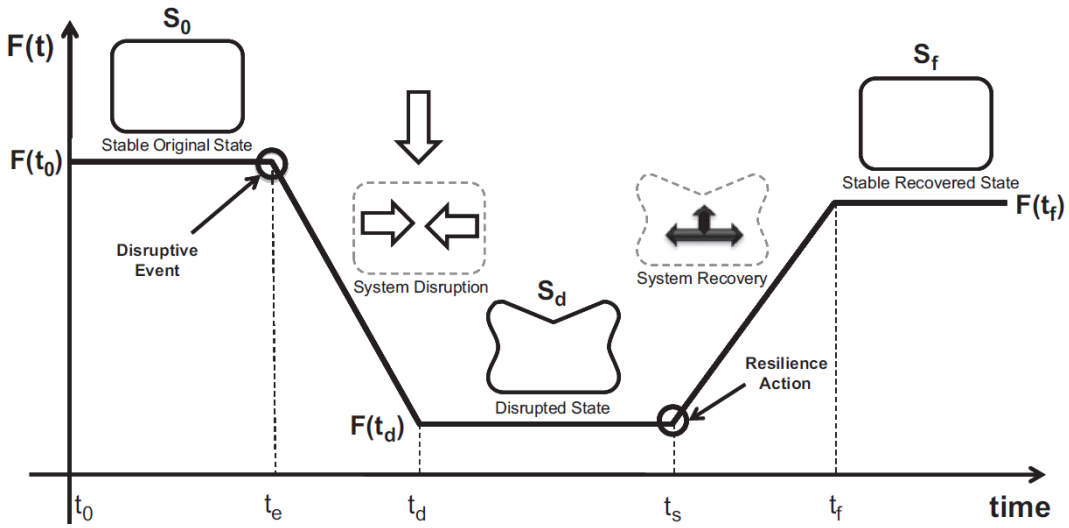


Figure 3.8: Disruption and recovery process.[32].

Probabilistic Approaches

A simple probabilistic resilience metric proposed by Chang and Shinozuka measures the probability of a system experiencing performance loss lesser than an acceptable level and a recovery time shorter than an acceptable level [12]. The equation below presents this concept. r^* and t^* are respectively the maximum acceptable loss of system performance and maximum acceptable recovery time from disruption i.

$$R = P(A|i) = P\left(r_0 < r^* \text{ and } t_1 < t^*\right) \tag{3.19}$$

Another stochastic approach is to compare the actual performance to target performance over a certain time period T. This is described by equation 3.20. $P(t)$ is the actual performance curve whilst $TP(t)$ is the target performance curve. The target performance curve can be considered as deterministic or stochastic.

$$AR = E\left[\frac{\int_0^T P(t)dt}{\int_0^T TP(t)dt}\right] \tag{3.20}$$

3.4.2. Structural-based models

Structural-based models differ from general measures by the need to model internal mechanics of systems which resilience is being analysed. As such, these approaches allow for conducting simulations or optimisation.

Optimisation models

Optimisation models are developed in order to plan actions for increasing resilience. In these approaches the resilience measurement is not a direct goal but rather it's maximisation for a given problem. An example of this is the problem of allocating resources for maintaining operations of an airports runway and taxiway network [24]. It can be noted that the optimisation models are dominant in scenarios where boundaries of the problem can be clearly defined and a model of the system can be constructed. As such, this method is mostly present in network problems with solution found using linear programming [4, 23, 40].

Simulation models

Simulation models require the modelling of actors and elements of a system. Unlike general measures, this allows to explore resilience in scenarios which have not occurred. This method of measuring system resilience has been applied in various domains such as logistics, infrastructure networks and disaster response [3, 11, 59].

Fuzzy logic models

Lastly, fuzzy logic models attempt to capture the probabilistic nature of processes by modelling them using many-valued logic where the truth values range from 0 to 1. This approach allows for accommodation of partial truths which is helpful in modelling non-deterministic processes. The approach has been utilised in fields like organisational and infrastructural resilience [47, 60].

4

Time-series clustering

Clustering is the process of grouping objects such that objects of similar characteristics are in the same group whilst those that differ are in other groups. Cluster analysis finds use in various domains such as banking where it can be used to segment clients and present them with personalised product offers. Furthermore, clustering is useful in data exploration as it is an unsupervised method i.e. it does not require labelled data. As such, it can be used to find previously unknown relationships useful for the development of prediction models. Clustering can be considered good when the intra-group similarity is high whilst the similarity between objects from different groups is low.

Time-series clustering is the specific use of clustering for time-series which can be defined as a set of observations each performed at a specific time. Time series can be defined as in Definition 4.0.1. An example use of time-series clustering is credit card fraud detection. In this use case, time-series of transactions are classified to detect whether it is an anomaly [57].

Definition 4.0.1. Time-series clustering *Given a dataset of n time-series data $D = \{F_1, F_2, .., F_n\}$, the process of unsupervised partitioning of D into $C = \{C_1, C_2, .., C_k\}$, in such a way that homogeneous time-series are grouped together based on a certain similarity measure, is called time-series clustering. Then C_i is called a cluster, where $D = \cup_{i=1}^{k} C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$ [2].*

In the context of Air Traffic Management system Resilience, clustering can be of use to determine nominal and non-nominal airspace conditions. The states of the airspace which consist of parameters like aircraft count, trajectory efficiency, convergence, etc. are all time-series. I.e. they are all measurements taken at a particular time and as such can be clustered to search for patterns which can be later utilised to quantify resilience.

4.1. General process

Time-series clustering methods contain four components: time-series representation, similarity/distance measure, clustering prototypes and finally time-series clustering [2]. These components occur in the order listed however methods may not have all the components. An example of this is clustering based on euclidean distance between time-series which would not require a time-series representation as the method works directly on the series itself. The schematic in figure 4.1 presents the four components.

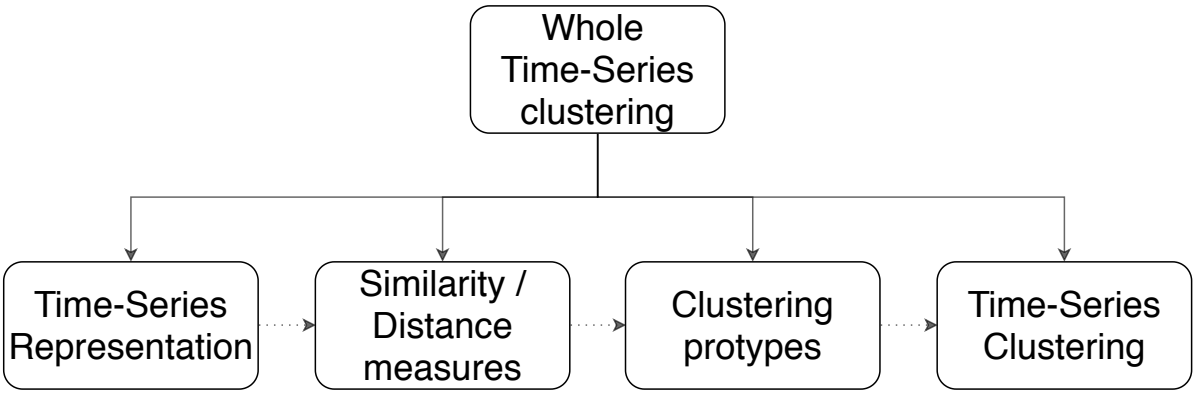


Figure 4.1: Time-series clustering components

4.1.1. Time-Series Representation

Time-Series Representation is the process of transforming a raw time-series single into a difference representation space such that the transformed representation in space R^m_{trans} is of less dimensions than the raw space R^n_{raw} i.e. $m < n$. The formal definition is introduced in 4.1.1.

Definition 4.1.1. Time-series representation *Given a time-series vector $F = \{f_1, f_2, .., f_m\}$, representation is transformation of the time-series to another dimensionality reduced vector $F_{reduced} = \{f_1, f_2, .., f_n\}$* where $n < m$ and such that the degree of similarity between different time-series is maintained after the transformation [2].

Aghabozorgi et al. suggest representation is performed for three reasons. First, for large time-series data, it reduces the computational effort required to perform distance measures. Measurements from high-frequency sensors like accelerometers can become memory intensive when accumulated for longer time-periods and thus the second reason is to reduce memory requirements. Lastly, real-world time-series are influenced by varying noise and bias such that the underlying process driving the time-series may not be visible. As such, performing clustering on such series would result in a grouping being performed not necessarily on the driving signal characteristics [2].

Four types of representation methods exist. First, are the **data adaptive methods** which attempt to minimise global reconstruction errors by using techniques such as Piecewise Polynomials or Singular Value Decomposition [62]. Second are **non-data-adaptive methods** which allow for easier comparison of multiple time-series and work on approximating local features rather than minimising the total reconstruction error as in the previous type. The methods include Wavelets and Chebyshev Polynomials [62]. The third representation type is **model-based** such as Markov Models or Auto-Regressive Moving Average. Lastly, there are **data dictated** representations which are derived directly based on the data. An example of a data dictated method is a bit representation of a time-series where each bit indicates whether an element of the time-series is above or below the average [5].

Aghabozorgi et al. note that time-series representation research is limited for discrete-valued time-series. Furthermore, a majority of research focuses on evenly-spaced time-series. Importantly for this research, the authors note that in the papers review for the 2015

Time-series clustering literature review no articles are addressing multivariate time series data of different lengths [2].

4.1.2. Dissimilarity measures

Once a time-series representation method is chosen, the similarity/dissimilarity between time-series can be measured. The most simple similarity measure can be computed when a time-series is considered to be univariate. That is *"the time-series is a sequence of real numbers sampled regularly in time where each number presents a value"* [2]. Under such a scenario, the distance between two time-series can be defined as in 4.1.2.

Definition 4.1.2. Time-series distance *Let* $F = \{f_1, f_2, .., f_n\}$ *be a time-series of length* T*. If the distance between two time-series is defined across all time points,then* $dist(F_i, F_j)$ *is the sum of the distance between individual points* $\text{dist}(F_i, F_j) = \sum_{t=1}^{T} \text{dist}(f_{it}, f_{jt})$ *[2].*

The challenge of distance measurement results from the inherent problems facing real time-series data which include: amplitude and time scaling, noise, linear drifts, temporal drifts and discontinuities. Furthermore, multiple factors affect the choice of distance measure, this is schematically shown in figure 4.2.

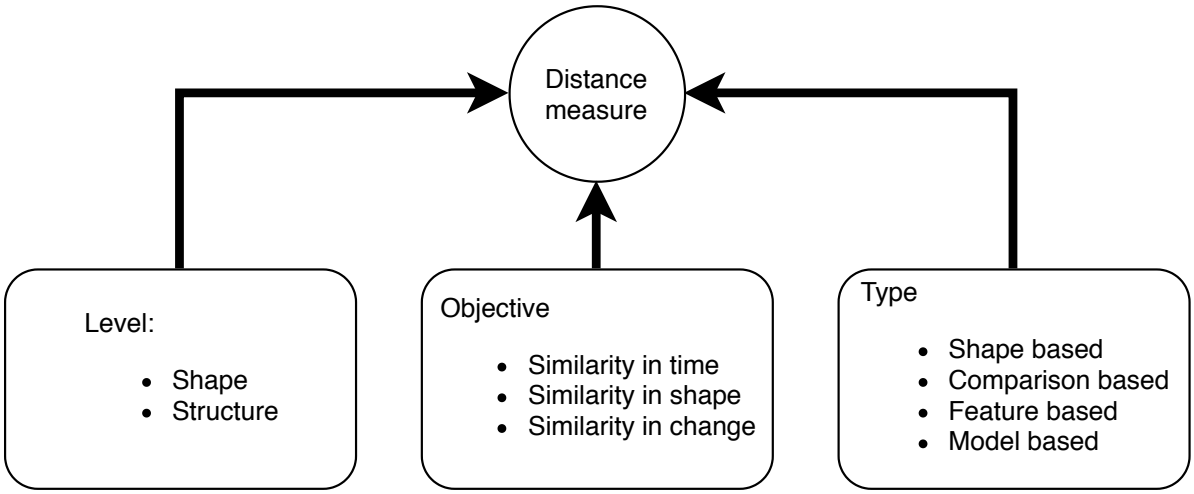


Figure 4.2: Factor in choosing the dissimilarity measure. Figure adapted from *Time-series clustering – A decade review* [2].

Aghabozorgi et al. distinguish three objectives that can be achieved, first being similarity in time in which the time of the occurrence of a pattern is of importance. For this objective, euclidean distance methods are useful [43]. For finding similarity in shape the exact time of occurrence is not of importance. This objective is useful for short time-series. For long time-series similarity in change is of more importance which is also referred to as similarity in structure. For this objective modelling approaches are used. Once model parameters are obtained, they are used to find similarities [2].

One of the most common dissimilarity measures for time-series is Dynamic Time Warping. This measure attempts to resolve the deficiency of performing direct euclidean distance computation on two equal length time series. The deficiency is that features in one time-series may appear in another time-series however with a time shift or amplitude scaling applied. As such, direct euclidean distance computation would result in a high distance despite actual similarities being present.

The goal of Dynamic Time Warping is to shift points of the time-series with respect to another time series such that the points align with minimum distance [48]. Figure 4.3 presents this concept graphically. The part of the time-series below is shifted with respect to the time-series above, however, part of it remains the same. A naive euclidean computation would return a high distance despite the two series being the same in their structure.

Whilst DTW is beneficial for distance computation, it suffers from poor computational complexity of $O(N^2)$ [48]. As such, it becomes problematic to use for long time-series such

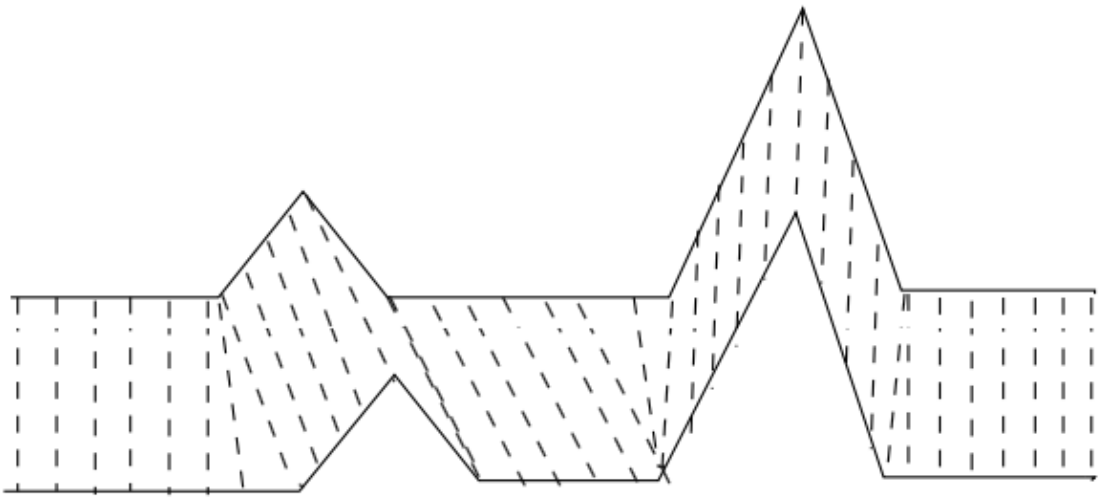


Figure 4.3: Conceptual representation of dynamic time warping algorithm [51].

as airspace metrics throughout the day.

4.2. Clustering prototypes

A cluster prototype is a representative value in the original data-space characteristic of a cluster. For example, if clustering would be performed on an aircraft approach paths to a runway, a cluster prototype would be a characteristic approach path to a runway. The simplest form of obtaining a prototype is by averaging. Given a set of time-series within a cluster the values of those time-series are averaged to obtain a representative value. However, this approach has multiple drawbacks. First, the time-series must be equal length. Second, averaging a time-series with similarities in shape rather than time causes the same issue as performing Euclidean distance computation. The resultant prototype time-series may not contain characteristic features of the cluster elements as some features may have experienced shifting in time or scaling.

A commonly used prototyping method is the medoid [2]. Whilst average prototype was obtained through calculating the mean, the medoid is obtained by computing the median time-series out of those in the cluster set. In practice, this is performed by computing distances between all the time-series combinations with Dynamic Time Warping or Euclidean distance. Then, the time-series which has the smallest cumulative distance to all other time-series becomes the medoid.

4.3. Time-Series clustering

Various categories of clustering methods exist. A common method is hierarchical clustering [42]. This method attempts to create a cluster hierarchy by a bottom-up or top-down approach. The former is based on starting with each sample having its own cluster and then joining the clusters upwards into bigger clusters. The latter does the opposite with all the samples in one cluster which are then divided up into smaller and smaller groupings.

The benefit of hierarchical clustering is that it does not require a specified amount of clusters as an algorithm input unlike k-Means, k-Medoids and so on. This a significant strength as with time-series signals of complex systems like airspace an initial amount of clusters is hard, if not impossible, to judge. An additional benefit of this method is the ability to accept unequal length time-series. The main drawback of this method is the poor computational complexity as the frequently utilises DTW as a distance metric.

The aforementioned k-Means algorithm is part of partitioning clustering methods. The algorithms in this category attempt to partition into k clusters such that the total distance of all elements in a cluster to each other is minimised. In the k-Means method, the cluster prototype is the mean of all elements in the cluster. A similar method is k-Medoids which utilises the medoid instead of the mean.

The partitioning clustering methods described so far can be referred to as crispy - that is they cluster samples into discrete clusters. However, there exist also soft methods which rather than directly assigning elements to clusters place a probability of them belonging to the particular clusters. Such clustering methods include Fuzzy c-Means and Fuzzy c-Medioids [2]. The downside of both soft and crispy partitioning based clustering methods is the need to define a number of clusters to partition into. Furthermore, unlike hierarchical clustering, there is a need to calculate a prototype for each cluster on which the accuracy of clustering dependant.

Another common clustering method outside of time-series analysis field is DBSCAN which is part of Density-based clustering methods [18]. The density-based methods create clusters in spaces of high density which are separated by spaces of low density. It has to be noted that such methods are not applied in time-series analysis [2].

Various other methods like grid-based clustering and multi-step clustering exist. However, they are complex or do not apply to time-series clustering [2].

4.4. Gaussian Mixture Models for cluster modelling

An important part of measuring resilience is the definition of the nominal state. As mentioned in the introduction, to measure resilience a reference state is needed. This reference state in case of noisy signals can be ambiguous. For this reason, there exists a need for modelling the reference state as a probability distribution. In this, Gaussian Mixture Models (GMM) can form a useful combination with clustering.

Once a cluster is obtained, it's elements are time-series of some metric (or metric in multivariate case) during a particular day at a Terminal Manoeuvring Area. Based on those time-series a nominal state for a time step can be derived using GMMs.

A GMM in of itself is a linear superposition of multiple Gaussian distributions with potentially different parameters, hence the mixture in the name [7]. GMMs are useful as they allow modelling of distributions of real variables which frequently do not adhere to a particular distribution such as Gaussian or Pareto. Mathematically the probability distribution resultant from the superposition is described as in equation 4.1.

$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(x|\mu_k, \Sigma_k\right) \tag{4.1}$$

Majority of the above equation is a multivariate Gaussian distribution. The novelty is the introduction of π_k term which is referred to as the mixing coefficient. If the Gaussian distribution terms are normalised then the sum of mixing coefficients is one. Hence, the mixing coefficient is, in fact, a probability of picking the kth distribution. Thus, all the mixing coefficients are bound between values of 0 and 1.

As such, GMMs can serve as a means of fuzzyfing otherwise crispy clusters. Additionally, GMMs can be utilised to generate extra data when required. However, for this research GMMs may prove useful to determine the nominal state as well as quantifying the likelihood of non-nominal state. A common GMM implementation in Python comes as part of scikit-learn package [10].

5

Preliminary results

The goal of the thesis is to quantify airspace resilience based on ADS-B data. To perform this, airspace metrics describing various aspects of ATM operations shall be derived based on the aircraft positional data obtained from ADS-B. These metrics along with weather information will allow for quantifying recovery from weather-related disruptions which include high winds and low visibility. However, the nature of ADS-B and Air Traffic Management are complex and result in the data being noisy and thus the driving phenomena remain uncertain.

For this reason, to derive airspace resilience a method must be developed for quantifying the likelihood of a particular day being non-nominal. This is the case because airspace operations vary greatly from day to day and as thus so do the metrics. An anomaly in the metric data may or may not stem disturbing events such as fog or high winds. Thus a probabilistic description is required to assess the results of such disruptions.

To quantify the resilience, first, the metrics have to be derived from ADS-B. The raw metrics then should be interpolated such that all metrics are available at the same time intervals. Lastly, the data should be filtered to remove the noise and expose the underlying trends. Next, the feature space of the metric time series should be reduced and then clustered.

In this section, the results of the preliminary results of the thesis shall be presented. These include: raw metric derivation in section 5.1, metric interpolation and filtering in section 5.2 and finally Principal Component Analysis in 5.3. These sections correspond to work packages: data processing and part of data analysis from figure 2.

5.1. Raw metric extraction

The airspace metrics to be utilised need to describe various aspects of the airspace to more completely describe the situation. Some events such as a closed runway might cause effects that are not visible in, for example, airborne aircraft count. However, they will be visible in the ground track inefficiency because the airborne aircraft shall be holding. As such, the following metrics shall be utilised:

- Airborne aircraft count
- Ground track inefficiency
- Proximity
- Convergence
- Landing count
- Take-off count

However, before the metrics can be extracted, the meaning of airspace has to be defined. Whilst it is possible to use the exact airspace volumes obtained from navigational databases,

such an approach would be more complex and allow less of comparison between airspaces. As such, the airspace will be defined as a cylindrical region of airspace around an airport with a radius of 100 NM.

For efficient filtering of aircraft within this cylinder, a formula has been developed to determine whether a latitude-longitude point (ϕ, λ) lies within a great circle distance d_{gc} of another point $(\phi_{ref}, \lambda_{ref})$ on earth that can represent an airport. This formula is presented in equation 5.1.

$$a\cos(\phi)\cos(\lambda) + b\cos(\phi)\sin(\lambda) + c\sin(\phi) > \cos(\frac{d_{gc}}{R_{earth}}) \tag{5.1}$$

where factors a, b, and c are:

$$\begin{aligned} a &= cos(\phi_{ref})cos(\lambda_{ref}) \\ b &= cos(\phi_{ref})sin(\lambda_{ref}) \\ c &= sin(\phi_{ref}) \end{aligned} \tag{5.2}$$

Based on the above formulas, the point (ϕ, λ) is within a great circle distance d_{gc} of reference point $(\phi_{ref}, \lambda_{ref})$ if expression in equation 5.1 is true. The resultant aircraft positions for a week of traffic within 100 NM of London Heathrow Airport can be seen in figure 5.1.

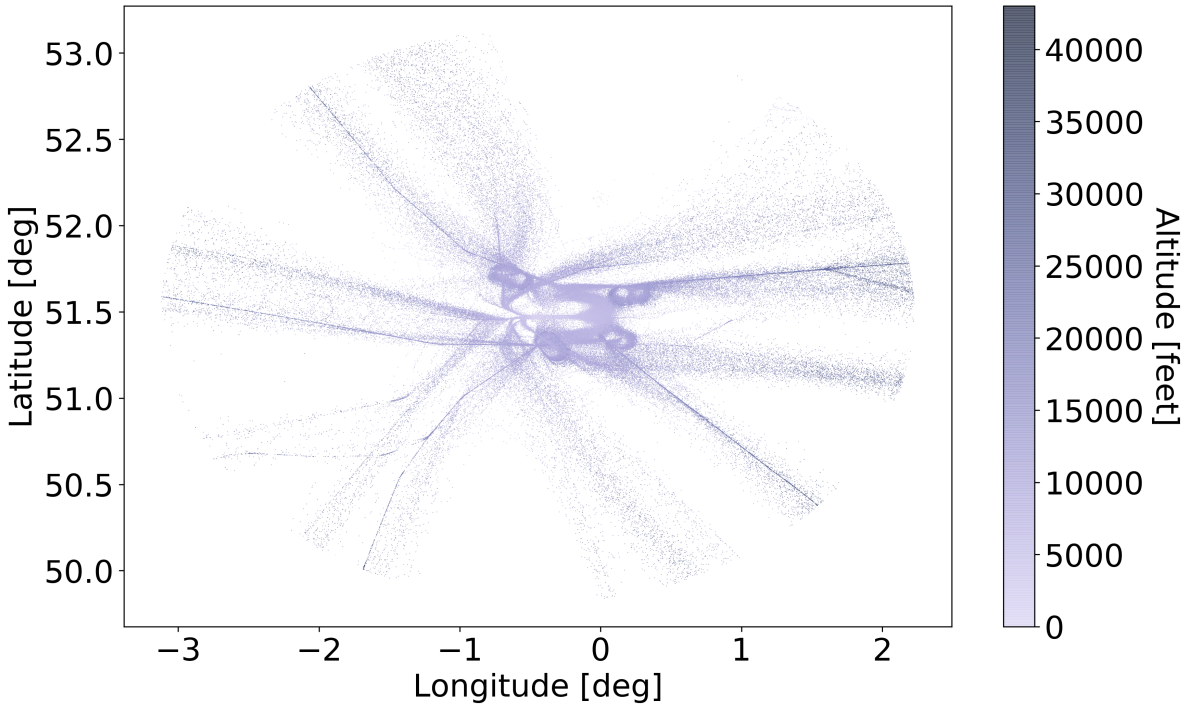


Figure 5.1: Traffic positions within 100 NM of London Heathrow airport during week starting 22nd January 2018.

The available ADS-B data is available in a decoded columnar format. The data is obtained from a ground-based receiver network and fused with operations schedule information to include additional fields such as departure airport, scheduled arrival airport and real arrival airport. Additional fields utilised throughout the metric extraction include message reception time in UNIX time as well as unique flight identification not to be confused with flight numbers.

The data is of reduced size to lower the required storage volume. The size reduction has been performed by the elimination of messages during unchanging flight conditions. For example, during cruise messages from a particular aircraft may be available in sixty-second intervals however during terminal manoeuvring this is reduced to one message every 5 seconds. An assumption within this research is that the aircraft equipped with ADS-B are representative of all aircraft within an airspace.

The discrete nature of ADS-B messages presents a challenge for deriving metrics that are continuous nature. For example, aircraft count can be performed continuously however if messages from different aircraft are received at different discrete times then the aircraft count may be inaccurate as for a given discrete-time messages for every aircraft may not have been received. This is further complicated by the aforementioned additional data reduction. To overcome this, the message reception times shall be rounded to the lowest value possible for a metric to remain representative.

The following sections present the metrics. The metric extraction is implemented using Apache Spark Python API whilst the analysis is performed locally using Pandas. [46, 66]

Aircraft count

The aircraft count is defined as the number of airborne aircraft within a radius of 100 NM of the airport. The aircraft count is obtained by first, **rounding the ADS-B message reception time to the nearest sixty seconds**. Then, for each minute a count of distinct transmitting aircraft is performed. The messages themselves do not serve as an indication of the amount of aircraft in the airspace as multiple messages per aircraft might have been received within a sixty-second period. If no messages have been received within a sixty-second period the aircraft count is set to zero. The process is described by a flow chart in figure 5.2

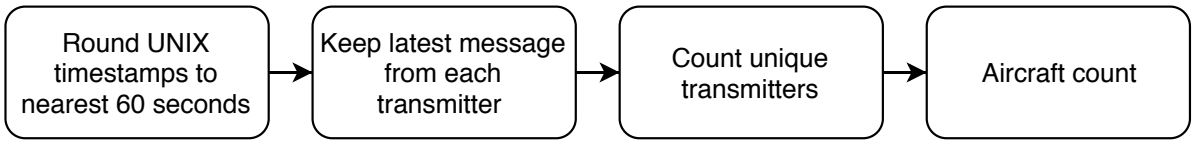


Figure 5.2: Flow chart of aircraft count acquisition from ADS-B data obtained around an airport.

This aircraft count has been performed for the complete year 2018 around London Heathrow. The result can be seen in figure 5.3. Each line in the plot represents one day.

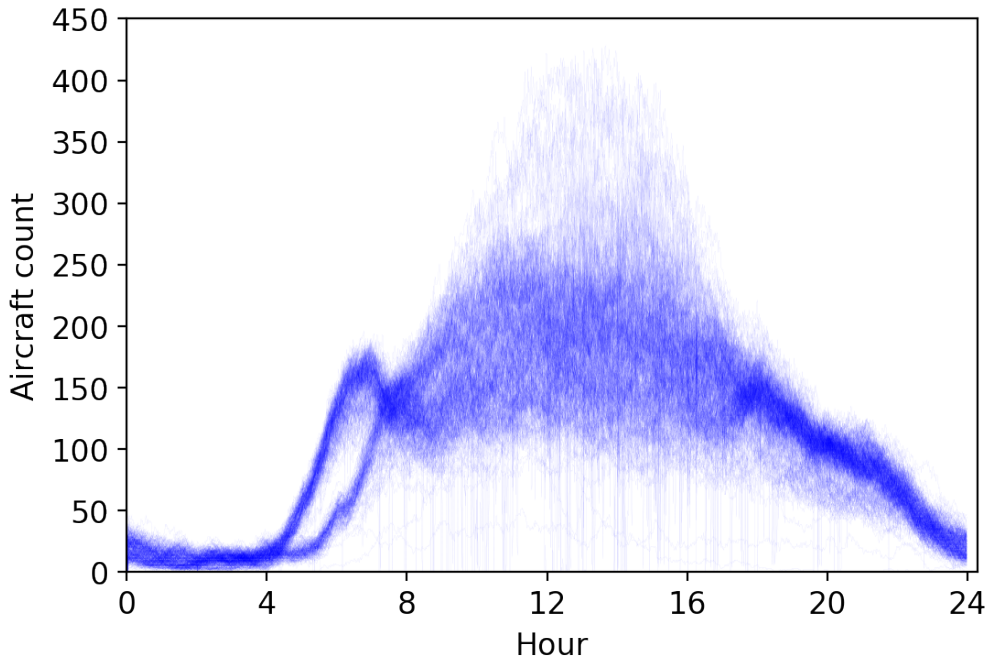


Figure 5.3: Plot of aircraft counts within 100NM radius of London Heathrow during year 2018. Each line presents one day.

Ground track inefficiency

Ground track inefficiency is defined as the ratio of ground track distance covered by aircraft between crossing the 100 NM radius around an airport and landing over 100NM. To account for the discrete nature of ADS-B and coverage limitations the inefficiency shall be implemented as a discrete integration of ADS-B positions up to the last position before the aircraft descending below 1200 feet. The cut off altitude of 1200 feet is used as the last received position altitude varies between aircraft and thus creates an inconsistency in the final position used to compute the inefficiency. Figure 5.4 presents quantile plot of aircraft altitude at the last ADS-B message before landing. As can be seen from it, the majority of last aircraft positions occur below 1200 feet. Figure 5.5 presents the process of computation this metric.

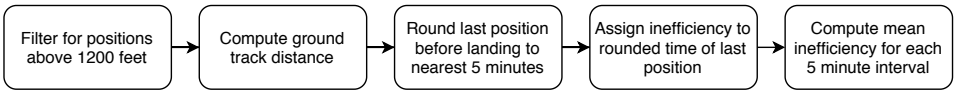


Figure 5.5: Flow chart of approach ground track inefficiency computation.

The inefficiency itself is presented in equation 5.3, the distance measure between two points that is used is Haversine distance.

$$\eta = \frac{\sum_{i=1}^{n} dist(\vec{x}_i, \vec{x}_{i-1})}{100NM} \tag{5.3}$$

The metric is equal to one if the aircraft would fly from the edge of the radius to landing in a straight path. Values above 1 show an inefficiency, an inefficiency of 0 is used to indicate that no landings were occurring in the given period. Whilst values below 1 are theoretically impossible, they do occur in practice as the first ADS-B message after crossing the radius may be some distance within the radius and the ground track integration will yield a value below 100 NM for efficient approaches. Inefficiencies for the year 2018 of aircraft arriving at LHR are present in figure 5.6.

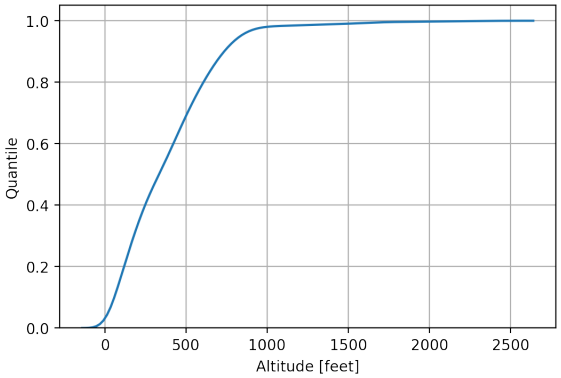


Figure 5.4: Quantile plot of last transmitted altitudes before landing.

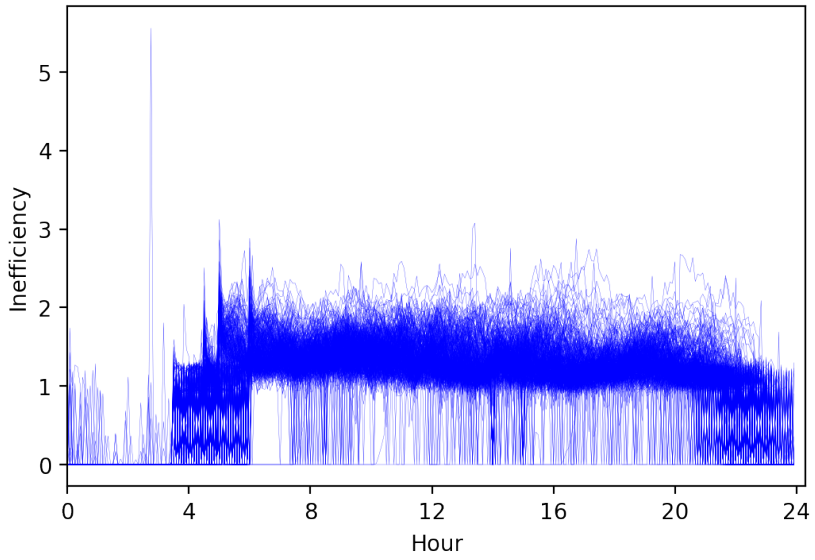


Figure 5.6: Arriving aircraft ground track inefficiency within 100NM radius of London Heathrow during year 2018. Each line presents one day. 0 presents no landing aircraft.

Proximity

The goal of the proximity metric is to characterise the level of closeness between aircraft in the airspace by prioritising the aircraft closer to each other. The metric to be used has been developed by Delahaye et al. and is presented in equations 5.4 and 5.5.[16] a and h are 5 and 0.3 km respectively and are considered characteristic distances for horizontal and vertical spacing. $\vec{p}$ is the position vector whilst x, y and z are the position coordinates.

$$d_{ij}^{a,h} = \|\vec{p_i} - \vec{p_j}\|_{a,h} = \sqrt{\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{a^2} + \frac{(z_i - z_j)^2}{h^2}} \tag{5.4}$$

$$P(i) = \sum_{j=1}^{N} e^{-\alpha d_{ij}^2} \tag{5.5}$$

The α factor in equation 5.5 is a scaling factor for distances and a value of 0.1 is used in the computation. Equation 5.5 is performed on a per aircraft basis. The airspace proximity metric is taken to be the mean of all the aircraft values. The process is summarised by a flow chart in figure 5.7. As with the previous metrics, the aircraft timestamps are rounded to a value which in this case is the nearest 15 seconds. The aircraft positions are then used to compute proximity between aircraft pairs which are then aggregated by taking the mean to obtain a value for the airspace.

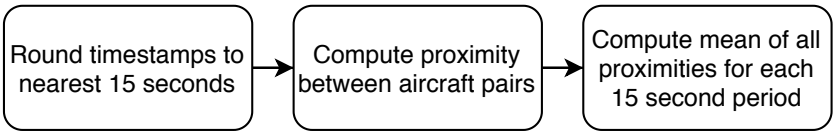


Figure 5.7: Computation process for proximity metric.

The resultant proximity values for the year 2018 around Heathrow airport can be seen in figure 5.8.

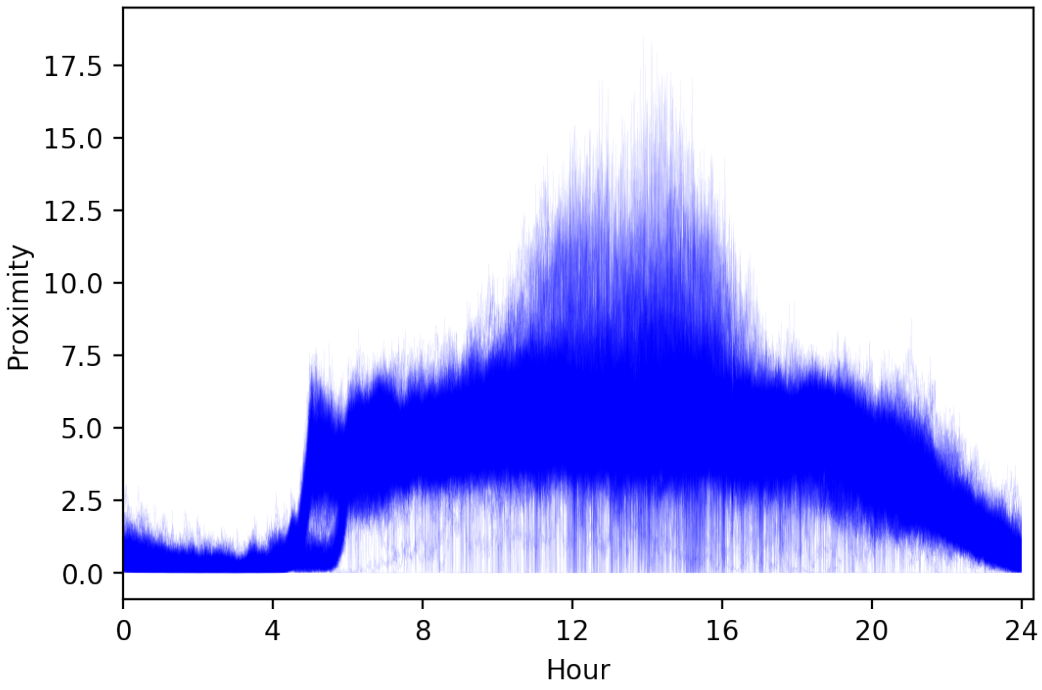


Figure 5.8: Proximity metric values throughout year 2018 around Heathrow airport. Each line presents one day.

Convergence

The convergence metric is the derivative of the distance between aircraft. I.e. it represents the nearing rate between aircraft and was developed by Delahaye et al [16]. Intuitively the value of the metric increases as the aircraft are closer together and moving faster in each other's direction. This nearing rate is defined in equation 5.6. $\vec{p}_{ij}$, $\vec{v}_{ij}$ and d_{ij} are respectively: the relative position vector, relative velocity vector and distance scalar.

$$r_{ij} = \frac{\partial}{\partial t} \left\| \vec{p}_{ij} \right\|_2 = \frac{\partial}{\partial t} \sqrt{\vec{p}_{ij} \cdot \vec{p}_{ij}} = \frac{\vec{p}_{ij} \cdot \vec{v}_{ij}}{d_{ij}} \tag{5.6}$$

The convergence itself is presented in equation 5.7. Like proximity, it is computed on per aircraft pair basis however only converging aircraft are used in the summation hence $r_{ij} \leq 0$. The two parameters include λ_c and α_c. The former is set to 1 whilst the latter to 0.1.

$$Cv(i) = \lambda_c \sum_{j/r_{ij} \leq 0} -r_{ij} \cdot e^{-\frac{1}{2}\left(\alpha_c \cdot d_{ij}\right)^2} \tag{5.7}$$

The process of computation is similar to proximity and is presented in figure 5.10. Figure 5.10 presents the resultant convergence values for a whole year.

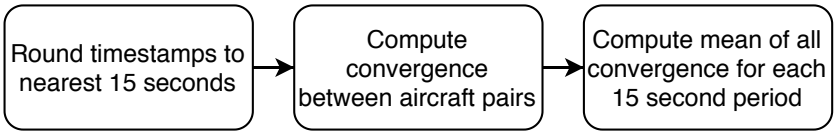


Figure 5.9: Convergence computation process.

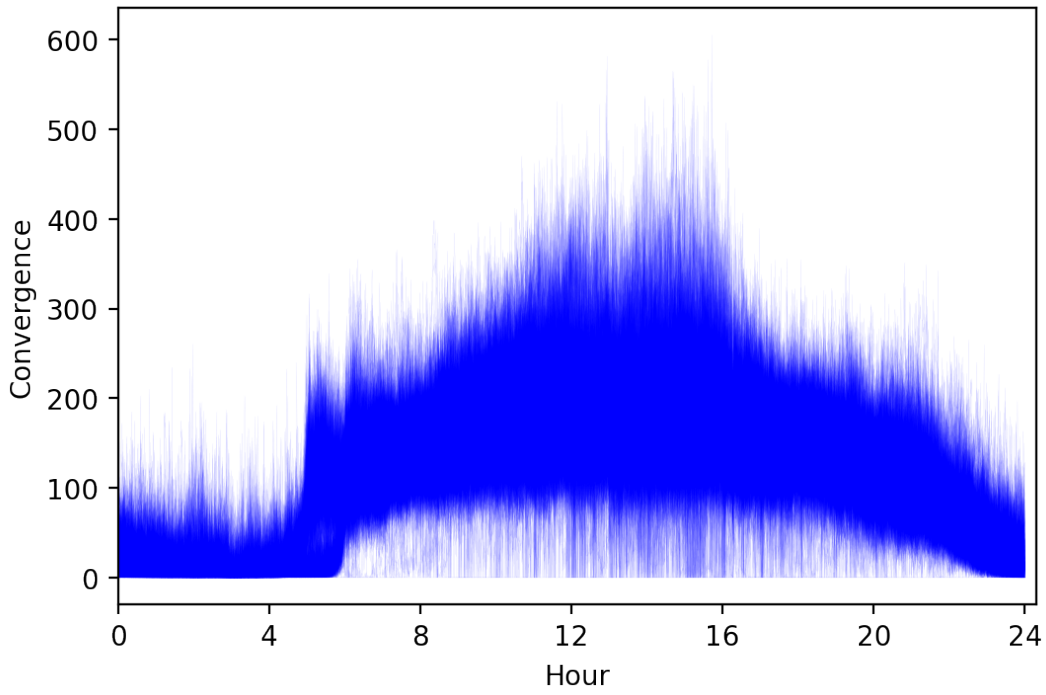


Figure 5.10: Convergence metric values throughout year 2018 around Heathrow airport. Each line presents one day.

Landing & Take-off count

The landing and take-off counts are performed by capturing first aircraft positions on the ground for landing or first airborne positions for take-off. However, this unlike the previously mentioned metrics does not provide a value but rather an event and it's time. To obtain actual aircraft take-off and landing counts a rolling window summation is performed over a 5 minute period for landings and take-offs separately. The obtained value is interpreted as the number of landings and take-offs in a 5 minute period. Figure 5.11 presents the computation process whilst figures 5.12a and 5.12b show the landing and take-off counts respectively. Unlike the figures for other metrics, figures 5.12 show the metric only for morning hours of one day. This has been done for figure clarity.

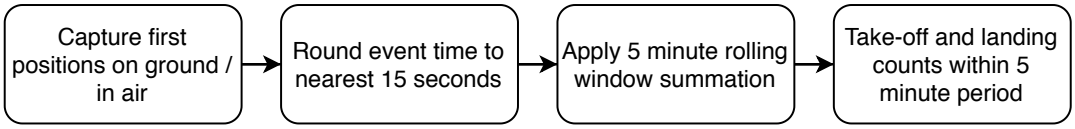
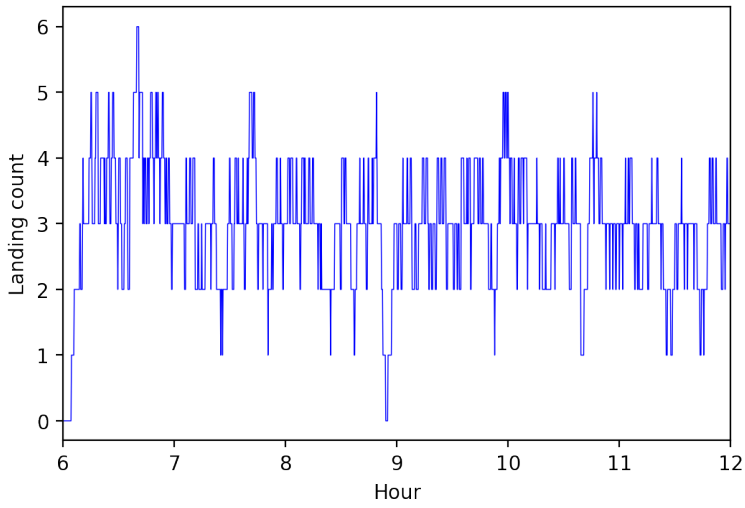
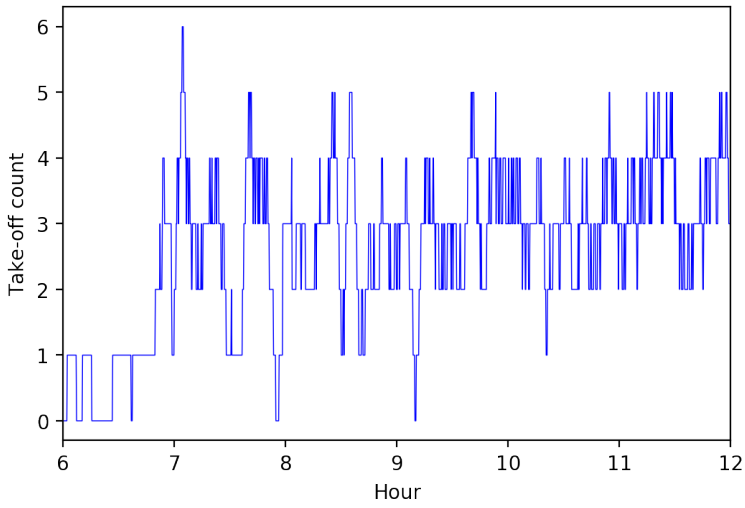


Figure 5.11: Operations computation process.



(a) Landing counts.



(b) Take-off counts.

Figure 5.12: Operation plots for 1st January 2018. 5 minute rolling sum. Morning hours. Single day presented for clarity.

5.2. Interpolation and filtering

From the figures in section 5.1 it can be seen that the data is noisy and the underlying process is not visible. Furthermore, the data is sampled at different intervals for the most accurate computation of the given metric given the discrete nature of ADS-B. Table 5.1 presents the characteristics of the derived metrics.

	Aircraft count	**Inefficiency**	**Proximity**	**Convergence**	**Operations**
Time interval [s]	60	300	15	15	Variable
Discrete	Yes	No	No	No	Yes

Table 5.1: Metric characteristics.

Interpolation

Majority of clustering and analytics methods such as Principal Component Analysis require vectors of equal lengths [52]. As such it is desirable to have values of each metric variable at the same time stamps. However, as can be seen from table 5.1 this is not the case for the raw metrics. As such, interpolation is used to reduce the spacing of metrics to the lowest time interval of 15 seconds of Proximity and Convergence. To obtain equal spacing and consistent vector length linear interpolation formula 5.8 is used.

$$y_n = y_{last}\left(1 - \frac{t_n - t_{last}}{t_{next} - t_{last}}\right) + y_{next}\left(\frac{t_n - t_{last}}{t_{next} - t_{last}}\right) \tag{5.8}$$

In the above formula, n indicates the number of the interval. For example, there are four 15s intervals in 60s. y is the metric corresponding to the metric sampled at a lower frequency than the one that is being interpolated to. y_{last} is the last available value whilst y_{next} is the next available value. t_{next} and t_{last} correspond to the times of y_{next} and y_{last}. The t_n presents the time of start of the nth interval. For a Δt second interval it can be defined as in equation 5.9. n must be discrete thus, $t_{next} - t_{last}$ must be divisible by Δt.

$$t_n = n\Delta t + t_0 \tag{5.9}$$

Filtering

The above equations are thus used for interpolation of metrics sampled at a lower frequency to a higher frequency. However, interpolation of noisy data introduces extra noise to the data. The noise in the data is problematic as it hides the underlying phenomena however its removal must be performed cautiously as the true underlying signal may be interfered with and important characteristics of it may be removed.

A common filtering method is utilising rolling windows. A rolling window is a method of dividing up a long series into multiple overlapping windows in which a determined operation is performed such as summation, averaging or another smoothing. The size of the window along with the operation to be performed in the window will affect the results of the filtering. The window size must be chosen to be wide enough to perform smoothing whilst not being too wide to hide the underlying phenomena. For the window size, no exact methodology is available. However, knowledge of the signal characteristic along with domain knowledge allows for selecting an appropriate size window. Consider figure 5.13 which presents the Fast Fourier Transform of the Aircraft Count metric for the complete year 2018. The x-axis, instead of units of frequency has units of time interval in hours to aid in readability.

It can be seen from the Fourier transform that the dominating time interval corresponds to just over 20 hours. This is reasonable as it corresponds to the daily fluctuation of aircraft during the day, low numbers at night whilst high during the day. Other peaks are visible on the scale between an hour and 10 hours for which no simple explanation is directly available. However, this information is useful as it indicates that no strong cyclical component is present on a scale of below an hour. If it existed, windowing filtering on such scale would have to be performed more carefully to not remove information.

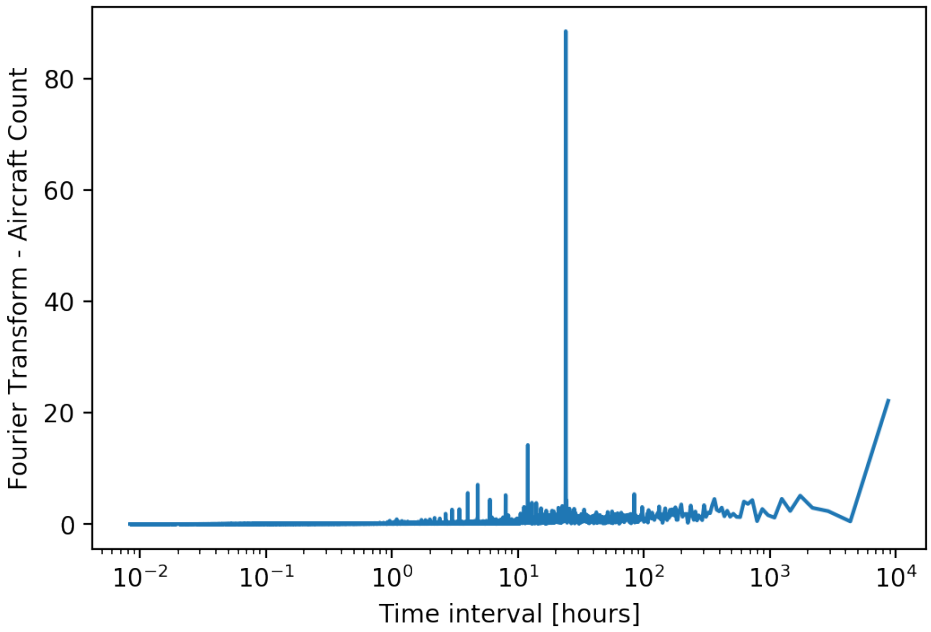


Figure 5.13: Fast Fourier Transform of Aircraft Count throughout year 2018 for London Heathrow. x-axis in the units of hours unlike usual frequency.

As such, for window filtering on a time scale of 25 minutes is deemed appropriate. Given the window size, a filtering method can be chosen. Here, one of the most common linear filtering methods is Window Averaging [63]. However, this method suffers from the drawback of being sensitive to outliers. As such, a better candidate is a Moving Window Median. Whilst more robust to outliers, it is more computationally expensive which is an acceptable deficiency. The resultant filtered metrics for 2nd May 2018 can be seen in figures 5.14 through 5.18.

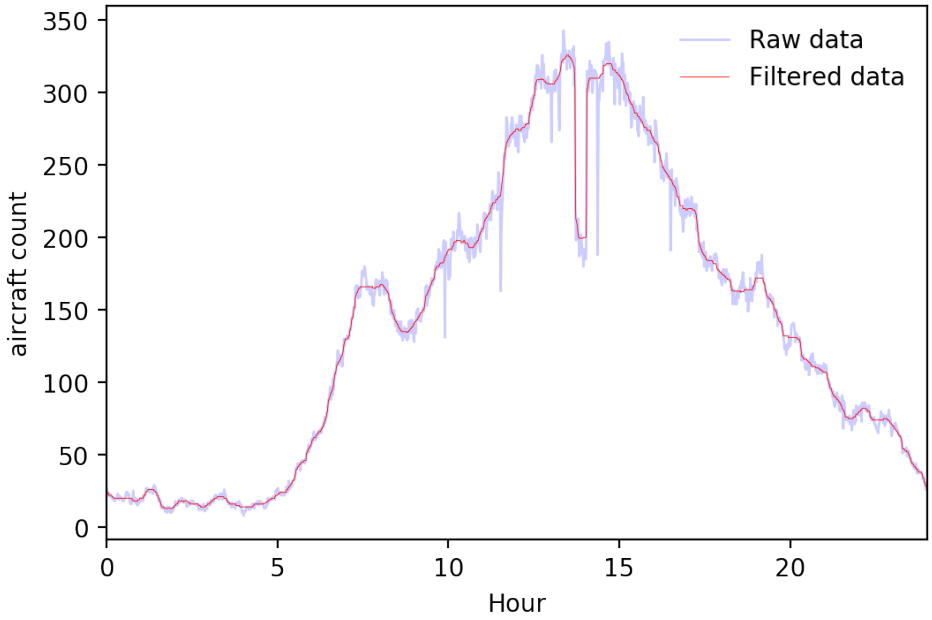


Figure 5.14: Aircraft count, filtered and nonfiltered comparison for Wednesday, 2nd May 2018.

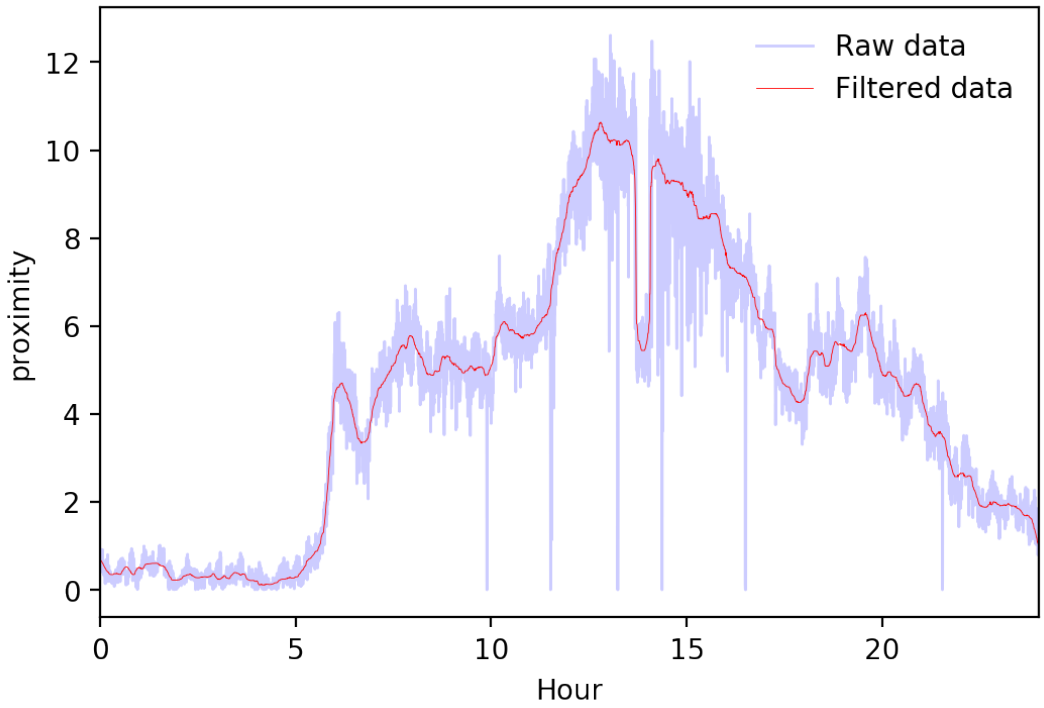


Figure 5.15: Proximity, filtered and nonfiltered comparison for Wednesday, 2nd May 2018.

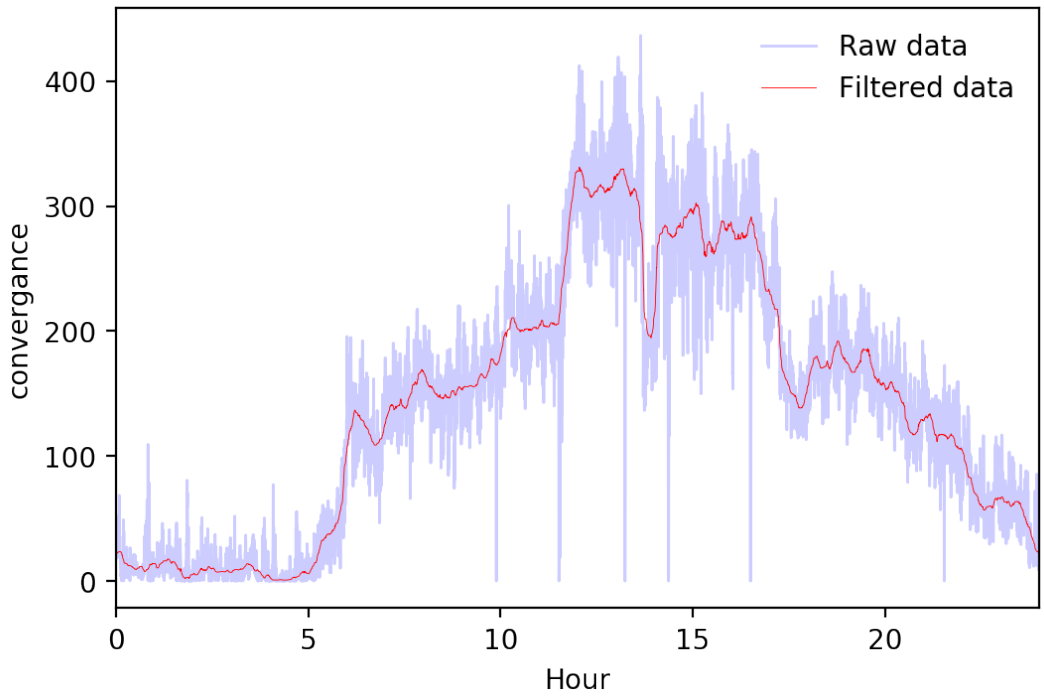


Figure 5.16: Convergence, filtered and nonfiltered comparison for Wednesday, 2nd May 2018.

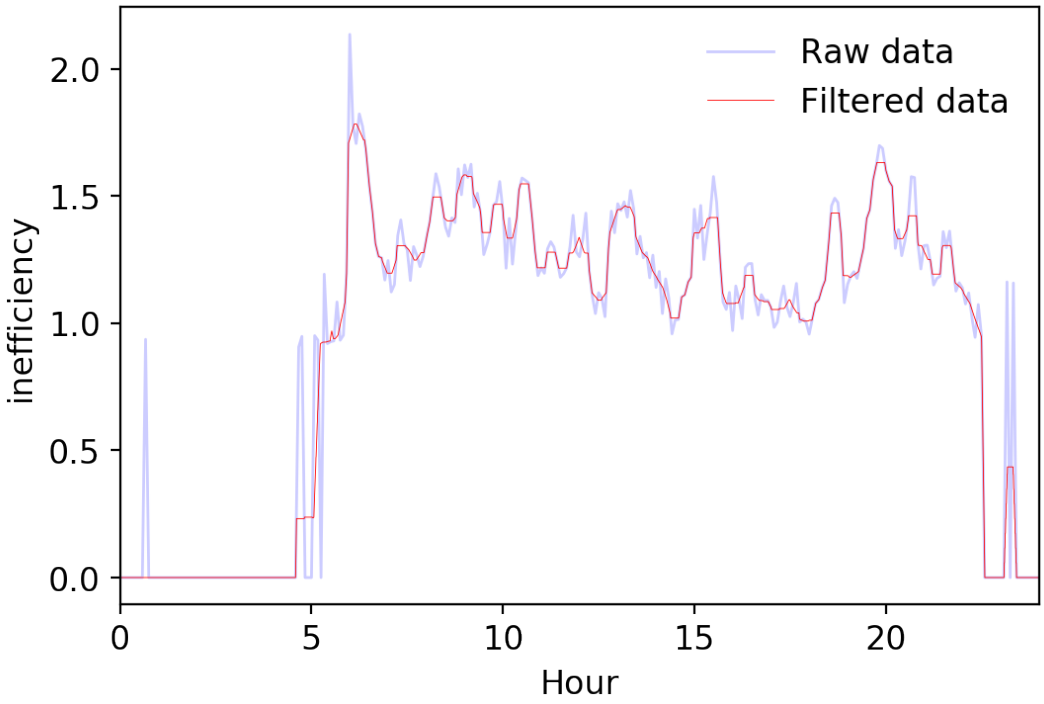


Figure 5.17: Inefficiency, filtered and nonfiltered comparison for Wednesday, 2nd May 2018.

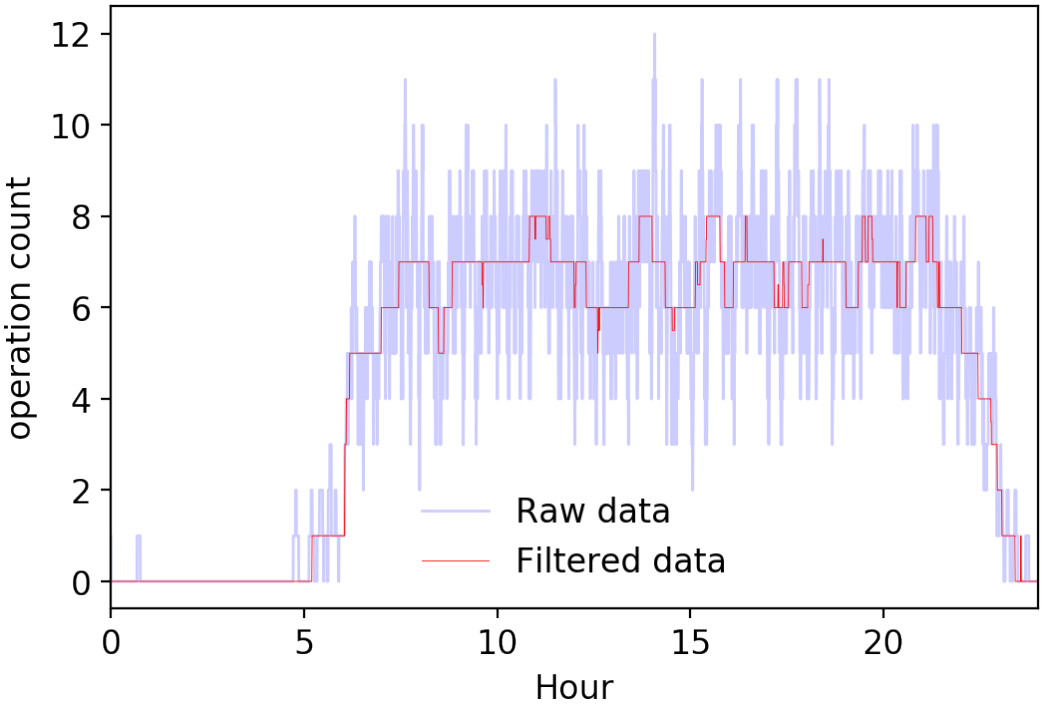


Figure 5.18: Operations - landings and take-offs, filtered and nonfiltered comparison for Wednesday, 2nd May 2018.

Daylight savings adjustment

London Heathrow flight schedules change by an hour with respect to the Coordinated Universal Time (UTC). This is the reason why in figure 5.3 two slopes are visible in the early part of the day. Whilst this seems like an actual variation by e.g. an airport having a different opening time on some days it only an effect present due to the time change. As compared to the local time the airport begins operation at the same times.

To correct for this, all the UNIX timestamps for the daylight savings days between 2018-03-25 01:00:00 and 2018-10-28 01:00:00 UTC are shifted forward by an hour by adding 3600 to them. The gap that this causes for the change forward in March is corrected for by placing 0s in for the metric values during the ”non-existent” hour during that day. The overlap caused at the end of the period is dealt with by removing the duplicate timestamps and leaving only the latest. Whilst the latter removes data, it is acceptable as the data removed is of less relevance as it from the night curfew period where no activity takes place. Furthermore, it is only an hour which is insignificant compared to the whole year 2018.

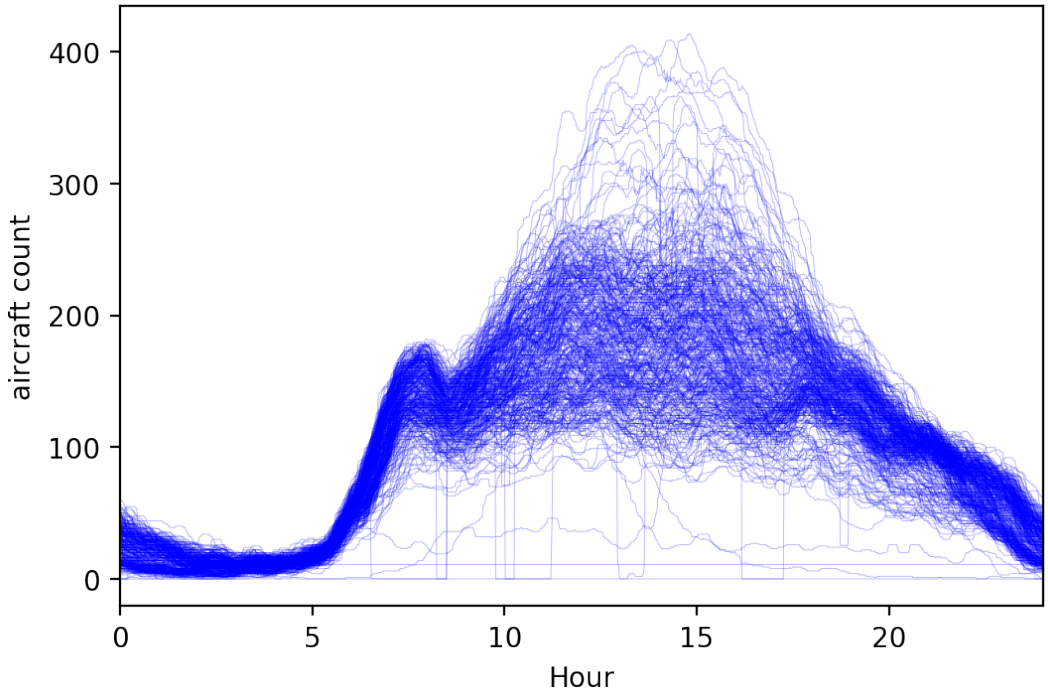


Figure 5.19: Aircraft counts for year 2018 corrected for day light savings. Note the lack of second slope in the morning hours as compared to plot 5.3.

5.3. Clustering

The goal of clustering in the scope of this research is to obtain groups of time-series which will best describe characteristic nominal airspace states. However, this is, in fact, a balancing act. Consider calculating nominal conditions for a complete year - i.e. clustering all days in a single cluster. The resultant nominal conditions will be ”one fit all” however, all the days of the year can be used to compute the confidence interval for this nominal state.

The solution to this is performing nominal state calculation more specifically, for example per quarter of the year or other feature. However, as mentioned earlier, this is a balancing act. Now consider a case where every day of the year is considered its cluster. Whilst this would theoretically be most accurate, as for example, the Friday before Christmas, is a unique day in a year due to heavier than usual traffic and thus it would make sense to compute a nominal ”Friday before Christmas”. The issue that arises as more nominal conditions are to be calculated the smaller the amount samples per nominal condition are available. This is related to the curse of dimensionality [6].

As such, the goal is to find an optimal amount of nominal states through clustering such that enough samples per feature are present but the nominal states are specific enough. For this to be performed, the metric data must be explored. For this, Principal Component Analysis has been used.

Principal Component Analysis - data exploration

Principal Component Analysis is a data exploration method which allows for reducing high dimensional data into lower dimensions for visualisation or feature reduction [58]. Mathematically speaking, PCA performs a transformation on a matrix $\mathbf{X}$ to a new coordinate system. It does this by eigendecomposition of a covariance matrix Σ of the features in matrix $\mathbf{X}$. Within this new system, the axis of it represent the directions of the maximal variance of the original matrix and are ordered from most varying to least varying. The axis, which are eigenvectors, of matrix Σ, represent the new axis and are called Principal Components.

To understand this method consider matrix in 5.20. In this matrix, the value represents a value of some metric such as aircraft count, inefficiency, etc. In the diagram, aircraft count is taken as an example.

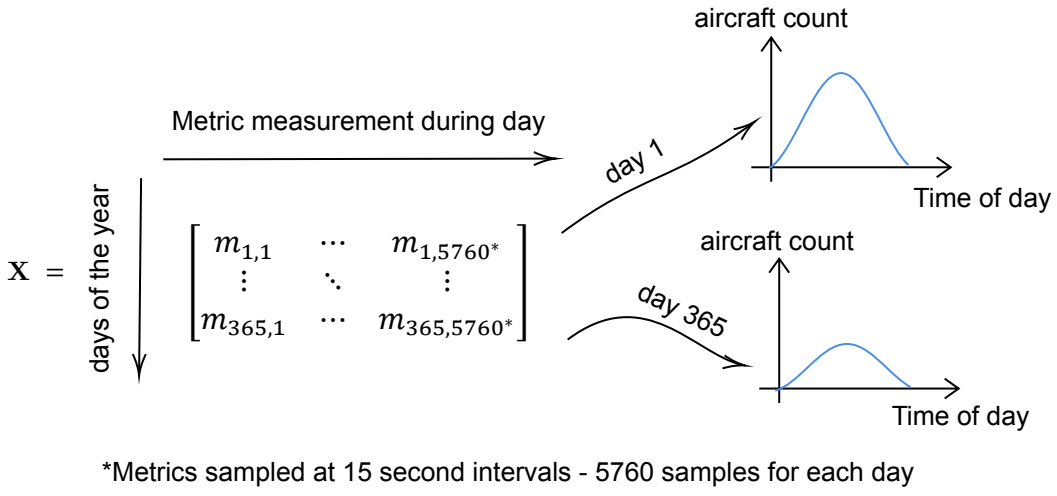


Figure 5.20: Description of matrix $\mathbf{X}$ used in Principal component analysis

Figure 5.20 shows the composition of matrix X used in Principal Component Analysis as an input. The matrix rows are days of the year whilst the columns are measurements taken during the day. Hence, each row corresponds to one curve for a particular day.

As mentioned before, PCA performs an eigendecomposition of the covariance of matrix $\mathbf{X}$. Because $\mathbf{X}$ is, in fact, a dataset with rows representing samples and columns variables (measurement at each timestep is considered a variable) the covariance Σ can be estimated

with equation 5.10 where n is the number of samples which in the case of the PCA being performed it is equal to 365 - number of days in a year.

$$\Sigma = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n-1} \tag{5.10}$$

Thus, the Principal Components are the eigenvectors of the covariance matrix Σ. The eigenvectors with the largest corresponding eigenvalues are the most influential in terms of the variance of data along them. In intuitive terms, they describe the most variance in the data from all principal components. Consider figure 5.21 which presents a toy example of PCA. In this image the two dimensional normally distributed scatter data has two principal components along directions of most variance.

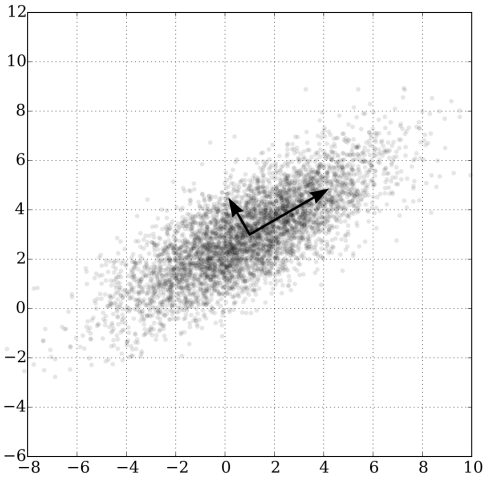


Figure 5.21: Toy example of PCA on two dimensional normally distributed data. Note the arrows indicate the directions of most variance - they are the Principal Components [49].

Before the PCA can be analysed, it is important to consider how much information do the two Principal Components have about the variance of data. As mentioned earlier, the principal components are eigenvectors with corresponding eigenvalues. The magnitude of the eigenvalue indicates how ”strong” a particular Principal Component is. Thus, the eigenvalue can be used to obtain information on how much variance does each Principal Component explain. This is done by simply by dividing the eigenvalue by the sum of all eigenvalues as in equation 5.11.

$$\sigma^2_{\lambda_k, frac} = \frac{\lambda_k}{\sum_{i=1}^{n} \lambda_i} \tag{5.11}$$

The result of this operation for the first 10 components can be seen in figure 5.22. As can be seen, the first Principal Component is dominant with almost 80% of variance explained by it. The second component explains only 10% with the next values having importance on the variance. Together, the two first Principal Components explain close to 90% of the variance.

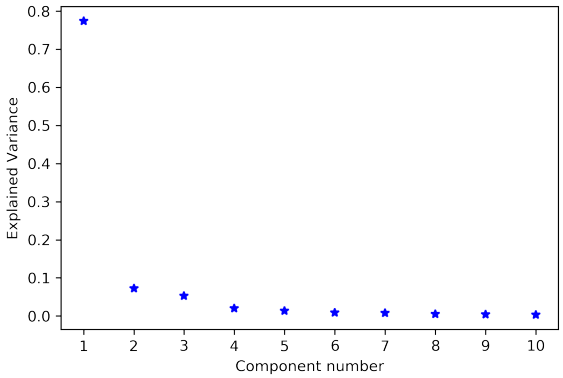


Figure 5.22: Explained variance of the first 10 Principal Components of PCA performed on aircraft count throughout year 2018.

For this research, PCA has been used to visualise the time-series data representing days of the year. Such visualisation allows to determine the overall structure of the time-series on a per-sample basis and give insight as to how to proceed with further operations such as clustering. The disadvantage of PCA is that it is not directly able to work on multivariate time-series data with multiple samples and as such it has to be performed on per metric basis. The result of the PCA for aircraft count with colouring by month and day of the week can be seen in figures 5.23 and 5.24 respectively. **Each dot in the PCA plots represents one day.** The PCA implementation used comes from scikit-learn and the inputs were standardised before inputting them into the PCA function [10].

In the aforementioned figures, one day is marked with an orange triangle. During this day the gathered data has been corrupted and very limited activity was observed in the airspace. An investigation into the cause has resulted in no events of importance occurring within the

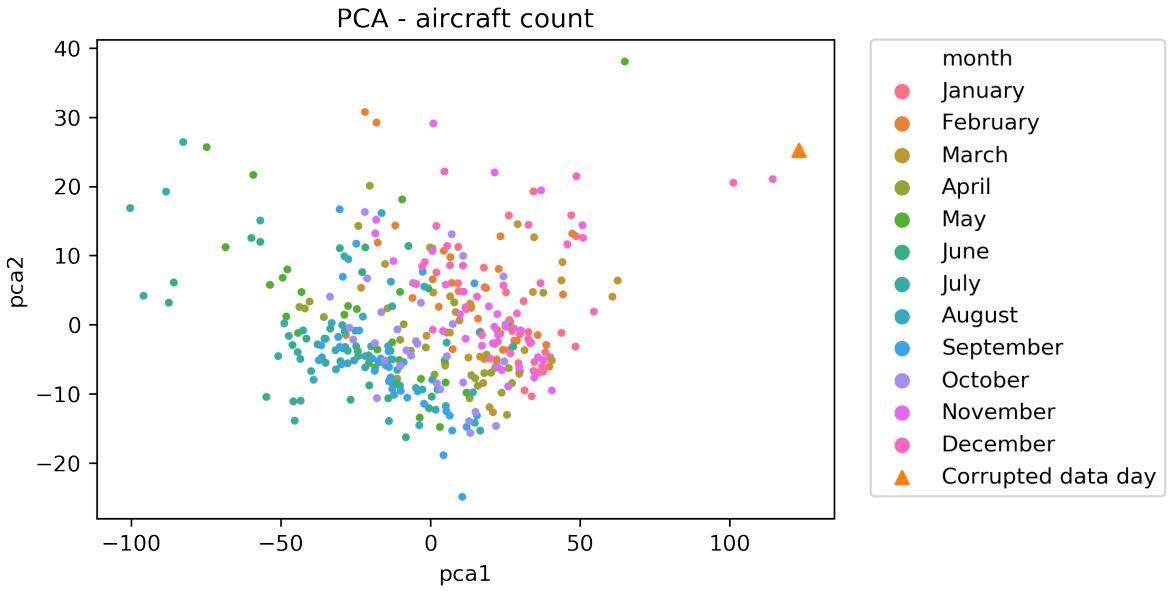


Figure 5.23: Principal Component Analysis of aircraft count with each point representing a day. Colouring by month. Triangle presents a day for which data has been corrupted and low amount of flights was detected.

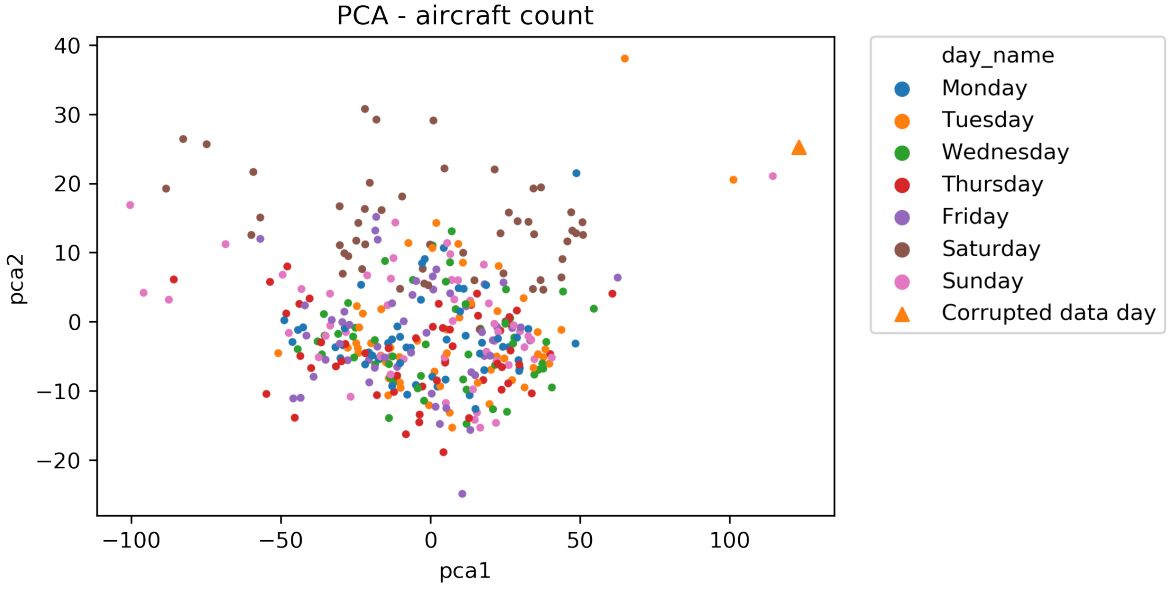


Figure 5.24: Principal Component Analysis of aircraft count with each point representing a day. Colouring by day of week. Triangle presents a day for which data has been corrupted and low amount of flights was detected.

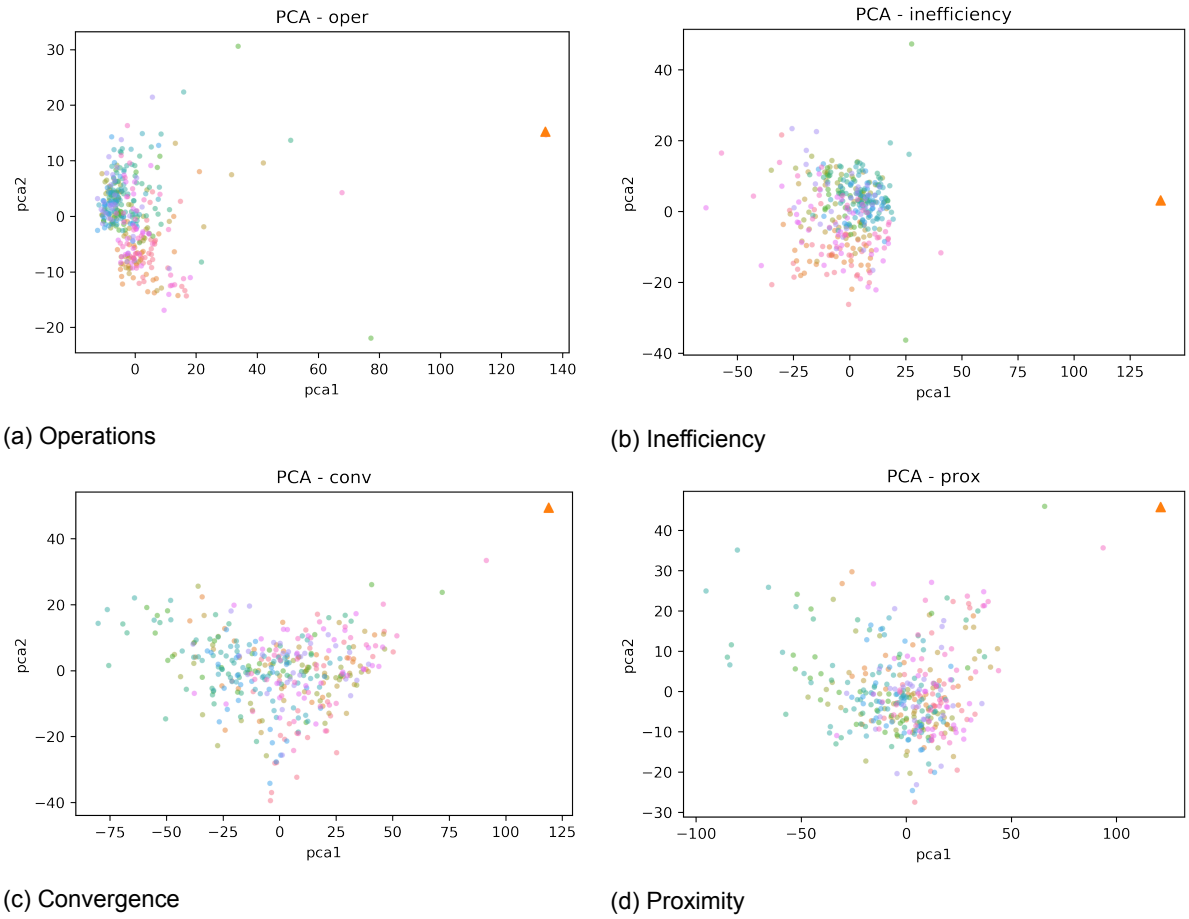
airspace and the data was deemed to be corrupt. However, this day does remain useful as it allows to verify whether the outliers in the PCA are unusual days. By analysis of the figure, it can be seen that the point is one of the furthest from the main cluster.

From the two PCA plots, it can be seen that no multiple clusters can be distinguished for all days. Majority of the points lies within one central cluster. It can be seen however that months tend to be a more important factor in the daily aircraft count rather than the day of the week. This is because some clustering can be observed when considering the month as seen in figure 5.23 where the winter months are different from summer months. This is not the case for days of the week in figure 5.24 where no similarities can be distinguished.

The above is of importance to further feature engineering as it highlights the importance of the seasonality in the daily trends. Thus, it appears that the month of a year is a better feature for deducing a days nominality as compared to the day of the week. Figures 5.25 present a similar pattern for the other metrics. Table 5.2 presents the variances of the first two components along with their sum for each metric.

	PC 1	PC 2	SUM
aircraft count	0.77	0.07	0.84
convergence	0.58	0.1	0.68
proximity	0.55	0.12	0.67
inefficiency	0.32	0.1	0.42
landing count	0.24	0.08	0.32
take-off count	0.22	0.08	0.3
operations	0.37	0.09	0.46

Table 5.2: Principal component variances for all metrics



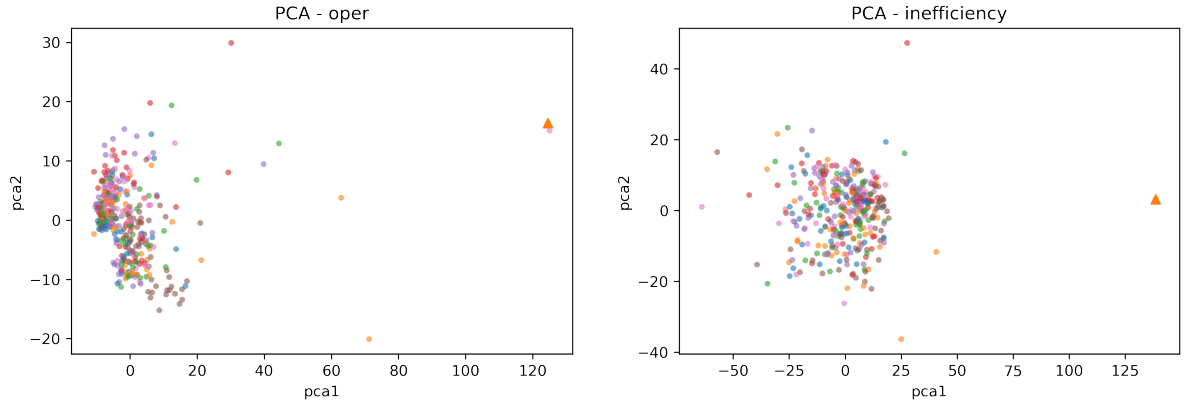
(a) Operations

(b) Inefficiency

(c) Convergence

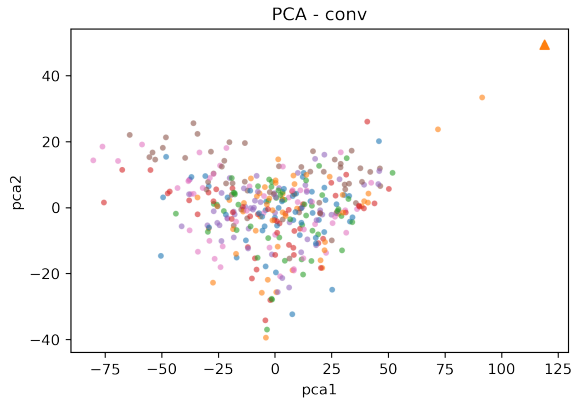
(d) Proximity

Figure 5.25: Principal Component Analysis of metrics with each point representing a day. Colouring by month. Triangle presents a day for which data has been corrupted and the low amount of flights was detected. Legend as in figure 5.23.

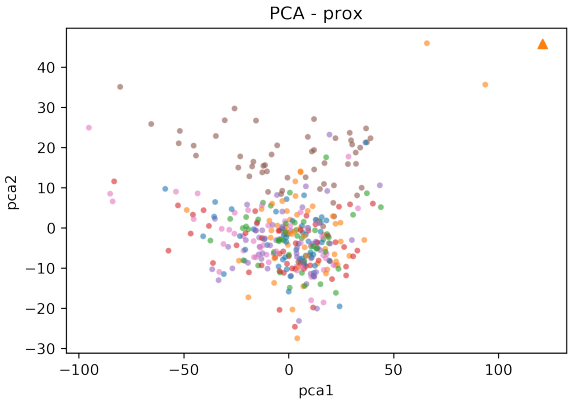


(a) Operations

(b) Inefficiency



(c) Convergence



(d) Proximity

Figure 5.26: Principal Component Analysis of metrics with each point representing a day. Colouring by day of the week. Triangle presents a day for which data has been corrupted and the low amount of flights was detected. Legend as in figure 5.24.

K-means clustering

K-means is a common method of clustering based on minimization of variance within a cluster [7]. This is described by equation 5.12. S_i is a cluster that is part of a set $S_1, S_2, ..., S_k$ where k is the desired number of clusters. $\mathbf{x}$ and $\mathbf{y}$ are observations belonging to a particular cluster. Note that $k \leq n$.

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \frac{1}{2\,|S_i|} \sum_{\mathbf{x},\mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2 \tag{5.12}$$

In practice, the clusters are computed iteratively starting from an initial random assignment of cluster centres. First, the observations are assigned to a cluster based on the smallest Euclidean distance or alternative distance measure. Next, the cluster centre is updated by computing the mean of points within the cluster. The process repeats until the change in the cluster centre is below a tolerance.

The K-means clustering algorithm has been tested on inefficiency metric. The inefficiency metric has been chosen as this metric is most likely to show the effects of disruptions. This is because if weather conditions become severe enough to constrain operations the aircraft already inbound to the airport will require to hold or go-around. As such, this metric should rise in poor weather conditions. The K-means implementation comes from Python library tslearn which features various functionalities for operations on time-series data [61].

The goal of clustering within the scope of this research is to create nominal conditions clusters. As such, the fitting of clusters in the K-means has been performed only with days during which weather was not severe. A day is considered to consist of severe weather if any of the following is true: wind speed over 25 knots, gusts over 30 knots, visibility below 0.5 NM. The clustering has been performed on per day basis just as the PCA. Hence, the input matrix is of shape: $(n_{days}, 5760)$ just as in figure 5.20. Figure 5.27 presents the results as the colouring of points on the PC plot of inefficiency metric. This result includes severe weather days which have been classified based on clusters derived from good weather days.

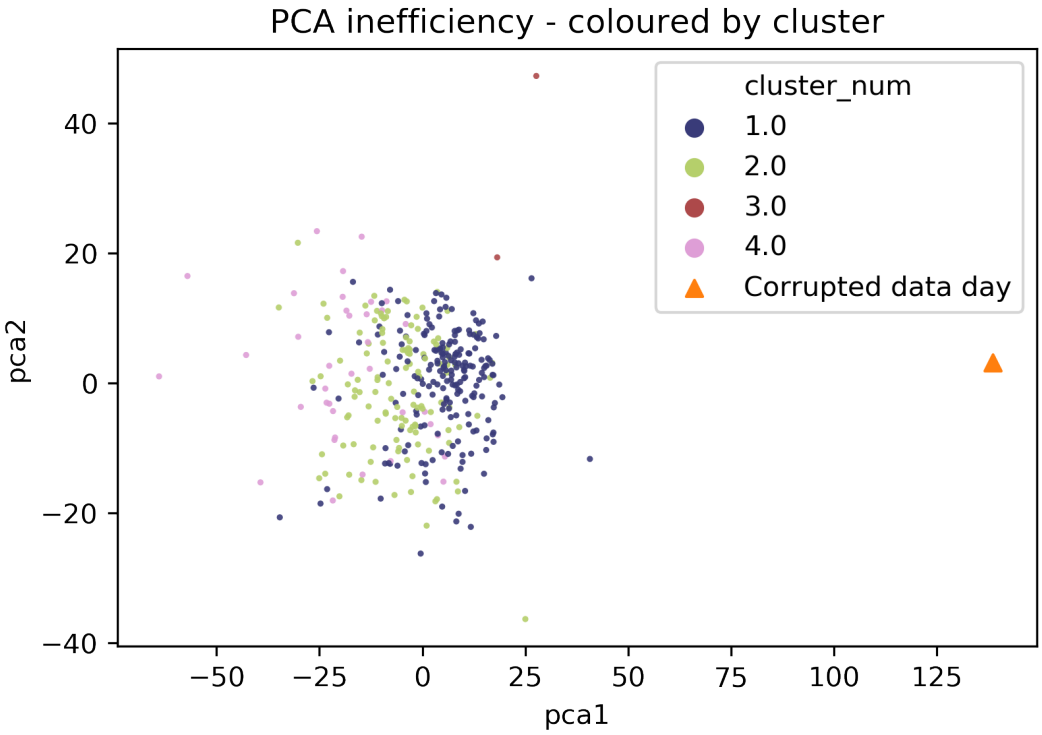


Figure 5.27: K-means clustering results displayed on a Principle Component Plot of airspace inefficiency metric. Colouring by cluster.

The K-means clustering has been performed with **4 clusters** based on the results from data exploration using PCA. As mentioned in the previous subsection, seasonality is of importance in metric patterns for London Heathrow more so than the day of the week. As such, 4 clusters represent four distinct seasons.

Euclidean distance serves as the distance measure. Experiments were performed using Dynamic Time Warping however the resultant cluster centres could not be explained and the computation required significantly downsampling the metric time-series due to the high computational complexity mentioned in chapter 4. As such, Euclidean distance proved to be robust and understandable. Importantly, the computation could be performed without downsampling the time-series data. As the data has been adjusted for daylight savings no large scale time lag will be present to disrupt the distance metric.

As can be seen from figure 5.27, the direct clustering on time-series data has resulted in clusters being formed in the transformed coordinate system of Principal Components. Cluster 1 (blue) is most numerous and also least varied. Next is cluster 4 (pink) which lies close to cluster 1 however it is more varied. Clusters 2 (green) and 3 (red) are the least numerous with cluster 3 representing numerous outliers. Note that the corrupted data day values have not been used in the clustering process.

Additionally, K-means utilises a random initial cluster centre placement. As such, every time the algorithm is performed the result is different. To prevent this, K-means clustering has been ran 100 times and the best clustering defined as the minimum distance of samples to the cluster centre.

To verify the results of clustering the cluster centres (non-severe weather only) have been plotted. They can be seen in figure 5.28. The colouring of the line corresponds to the colouring from figure 5.27.

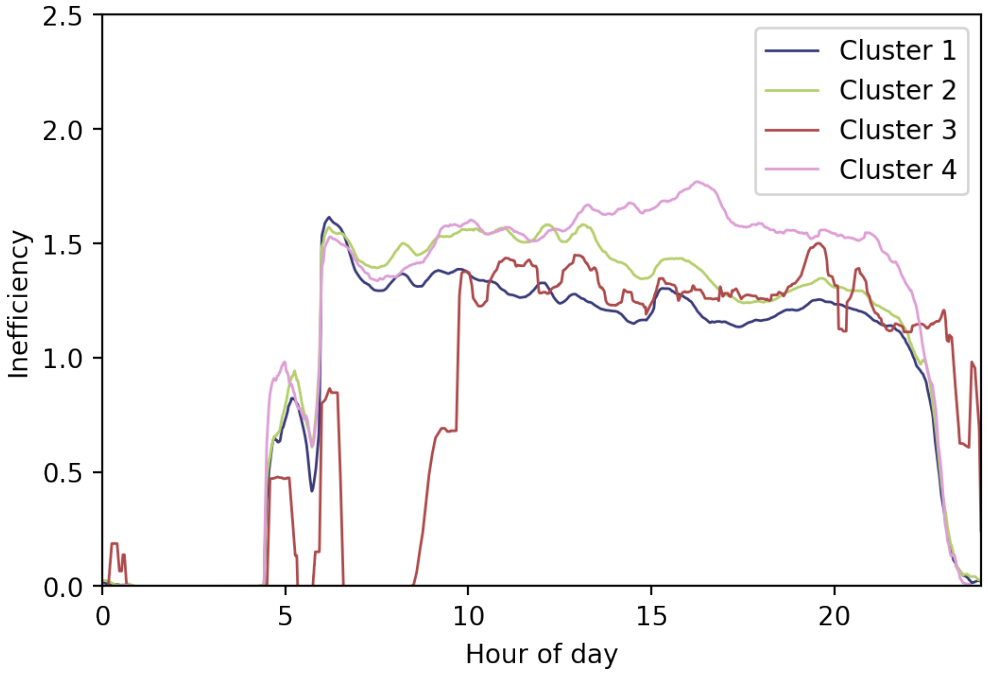
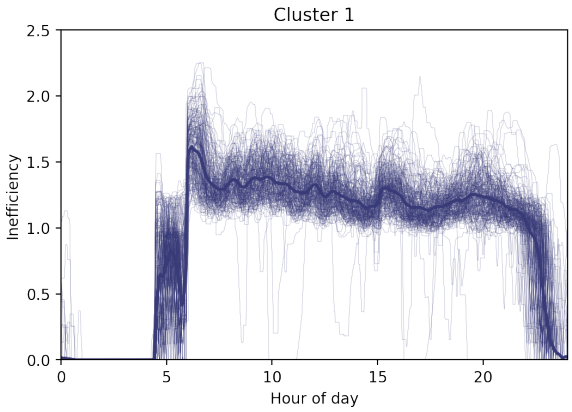


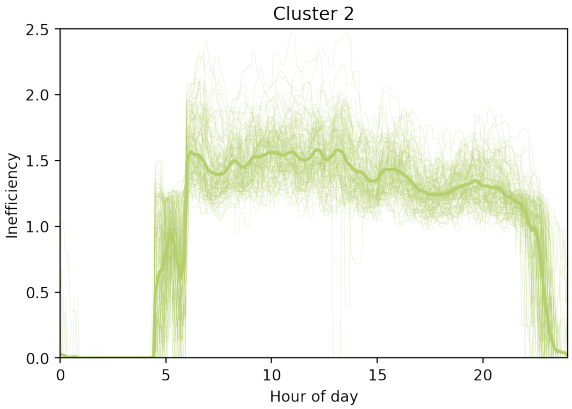
Figure 5.28: Cluster centres for the four clusters computed using K-means. Colouring the same as in figure 5.27.

The above plot shows that clusters remain quite similar around morning and evening hours where traffic increases and decreases respectively. Majority of variation occurs during the day. For cluster 1, after the morning peak caused by a concentration of arrivals the values begin to decrease throughout the rest of the day. Cluster 2 is similar to cluster 4 due to an increase in inefficiency following the morning rush. However, within cluster 2 the values begin to subside around noon. Cluster 3 is an anomalous cluster. Cluster 4 is characterised by a continuous increase in inefficiency throughout the day. Only during the evening hours

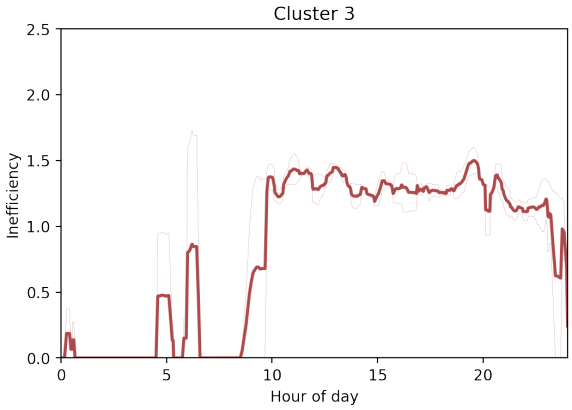
does it begin to subside.



(a) Cluster 1 - decrease after morning peak.



(b) Cluster 2 - increase till noon, decrease throughout rest of day.



(c) Cluster 3 - anomalies.

(d) Cluster 4 - increase throughout day.

Figure 5.29: Cluster centres along with the days belonging to a particular cluster.

Figure 5.29 presents the cluster centres from figure 5.28 with days belonging to the cluster plotted on top. The anomalous cluster 3 contains only two days for which there appear to be no flights in the morning hours. The other clusters centres appear to well match the distribution of the samples within them.

After the K-means clustering has been fitted on good weather days, the remaining bad weather days were classified based on the obtained clusters. Table 5.3 displays the total number of days in a cluster as well as the share of good and bad weather days within them.

	Bad Weather	**Good Weather**	**Total**
Cluster 1	15	184	199
Cluster 2	15	107	122
Cluster 3	0	2	2
Cluster 4	4	34	38

Table 5.3: Good and bad weather day counts in clusters. Sum of totals is not 365 due to removed days.

Table 5.3 presents the information regarding weather in clusters. Overall it can be seen that clusters 1 and 2 are most numerous. Overall, the bad weather days are relatively evenly distributed amongst clusters. Cluster 1 consists of 7.5% bad weather days whilst cluster 2 and 3 of 12% and 11% respectively. The anomalous cluster contains none. This balance between the dominant clusters suggests that no cluster is representative of disruptions.

Conclusion

In this preliminary thesis report, a literature study has been performed which has highlighted the need to explore the concept of resilience. ATM delays have been on the rise in both, the United States as well as Europe. About a quarter of near airport ATM delays are caused by weather conditions.

To lessen the impact of the aforementioned weather conditions on air traffic, the ATM system resilience to such disruptions should be considered. Interest in Resilience has been on the rise in the past decades due to desire for systems to maintain their operational capacity throughout and following disruptions. Within ATM community Resilience is understood as the ability to recover from a disruption through transient perturbation to a reference state. The challenge in measuring resilience in real systems is the ambiguity of the reference state as well as the effects of the disruption.

The transient perturbation and reference state have to be quantified. Various metrics exist that quantify various aspects of the ATM system. Most numerous are airspace complexity metric which serve to define the complexity of an air traffic situation as perceived by an Air Traffic Controller. These metrics appear to be useful when considering airspace also in the spatial dimensions, i.e. on a map. However, aggregated values for complete airspace varying only in time are heavily correlated with aircraft count in the airspace, as such they do not provide much more information than this simple metric. Key Performance Indicators such as the number of operations, efficiency, time in the airspace, etc. can also serve as airspace metrics. In particular, efficiency metrics may contain information on bad weather conditions as those are related to holdings and go-arounds.

As mentioned previously, the airspace reference state and deviations from it are ambiguous. Each day in the airspace is different and what is considered nominal is not obvious and as such, neither is the deviation from it. For this reason, time-series clustering can become useful to estimate a set of nominal conditions in the airspace. This clustering can be later utilised to match bad weather day metrics to it's most likely nominal condition and estimate the probability of deviation through means such as Gaussian Mixture Models.

The preliminary work performed included extracting airspace metrics for London Terminal Manoeuvring Area from ADS-B data supplied by Airbus Airsense for the year 2018. These metrics have also been refined through interpolation and filtering to achieve evenly spaced time-series for clustering purposes. Principal Component Analysis of airspace metrics has yielded that no distinctive clusters exist when projecting the metrics into two-dimensional space. However, the clustering performed using K-means algorithm on inefficiency metric has yielded three types of nominal days: declining inefficiency throughout the day, increasing inefficiency till evening and third, increase till noon with a drop afterwards.

Overall, the clustering to define nominal days has yielded promising results with clear trends in data. However, the remaining issue within the process of resilience measurement is the measurement of deviation from nominal. The complexity metrics of proximity and convergence do not seem to add much value over aircraft count due to the strong correlation between them.

Recommendations

Within the context of the complete thesis, this preliminary report objective is to document research performed thus far. The results obtained hint at the feasibility of using clustering in the process of resilience measurement. However, additional work remains to be done. Below are highlighted the key points for the next thesis stage:

- Consider time inside of airspace on arrival as a metric due to possibly smaller noise as compared to inefficiency metric.
- Perform feature engineering to determine if a combination of metrics has a higher relation to weather conditions.
- Consider lower sampling rate.
- Perform analysis on a time-scale shorter than a day to explore relations between weather conditions during the day and the metrics.
- Utilise, Gaussian Mixture Models for modelling of cluster time-series elements.
- Achieve greater stability in clustering results through picking clustering results based on intra-cluster and inter-cluster distances.
- Utilise Gaussian Mixture Model results to measure the probability of deviation for bad and good weather days. Compare the results.
- Perform the complete clustering and deviation analysis process for Amsterdam Schiphol and compare results to London Heathrow.

Bibliography

[1] Adrian Young. Aviation safety review: facts and improvements, December 2015. URL https://to70.com/aviation-safety-review-facts-and-improvements/.

[2] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, and Teh Ying Wah. Time-series clustering – a decade review. *Information Systems*, 53:16–38, Oct 2015. ISSN 0306-4379. doi: 10.1016/j.is.2015.04.007.

[3] Pavel Albores and Duncan Shaw. Government preparedness: Using simulation to prepare for a terrorist attack. *Computers & Operations Research*, 35(6):1924–1943, Jun 2008. ISSN 0305-0548. doi: 10.1016/j.cor.2006.09.021.

[4] A. Azadeh, V. Salehi, B. Ashjari, and M. Saberi. Performance evaluation of integrated resilience engineering factors by data envelopment analysis: The case of a petrochemical plant. *Process Safety and Environmental Protection*, 92(3):231–241, May 2014. ISSN 0957-5820. doi: 10.1016/j.psep.2013.03.002.

[5] Anthony Bagnall, Chotirat "Ann" Ratanamahatana, Eamonn Keogh, Stefano Lonardi, and Gareth Janacek. A bit level representation for time series data mining with shape based similarity. *Data Mining and Knowledge Discovery*, 13(1):11–40, Jul 2006. ISSN 1573-756X. doi: 10.1007/s10618-005-0028-0.

[6] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1957. ISBN 978-0-691-07951-6. Google-Books-ID: wdtoPwAACAAJ.

[7] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, Apr 2011. ISBN 978-0-387-31073-2.

[8] Henk A P Blom and Soufiane Bouarfa. Resilience. In *Complexity science in air traffic management*, chapter Chapter 5. Routledge, 2016.

[9] Michel Bruneau, Stephanie E. Chang, Ronald T. Eguchi, George C. Lee, Thomas D. O'Rourke, Andrei M. Reinhorn, Masanobu Shinozuka, Kathleen Tierney, William A. Wallace, and Detlof von Winterfeldt. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra*, 19(4):733–752, November 2003. ISSN 8755-2930. doi: 10.1193/1.1623497. URL https://earthquakespectra.org/doi/10.1193/1.1623497.

[10] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[11] Helena Carvalho, Ana P. Barroso, Virgínia H. Machado, Susana Azevedo, and V. Cruz-Machado. Supply chain redesign for resilience using simulation. *Computers & Industrial Engineering*, 62(1):329–341, Feb 2012. ISSN 0360-8352. doi: 10.1016/j.cie.2011.10.003.

[12] Stephanie E. Chang and Masanobu Shinozuka. Measuring improvements in the disaster resilience of communities. *Earthquake Spectra*, 20(3):739–755, Aug 2004. ISSN 8755-2930. doi: 10.1193/1.1775796.

[13] Civil Air Navigation Services Organisation. Methodologies for Calculating Delays/Improvement Opportunity Pools By Phase of Flight. Technical report, Civil Air Navigation Services Organisation, May 2013. URL https://www.canso.org/sites/default/files/Methodologies%20for%20Calculating%20Delays.pdf.

[14] Civil Air Navigation Services Organisation. Recommended Key Performance Indicators for Measuring ANSP Operational Performance. Technical report, Civil Air Navigation Services Organisation, March 2015. URL https://www.canso.org/sites/default/files/RecommendedKPIforMeasuringANSOOperationalPerformance.pdf.

[15] Daniel Delahaye and Stéphane Puechmorel. Airspace Congestion Metrics. *Modeling and Optimization of Air Traffic*, pages 243–289, 2013. doi: 10.1002/9781118743805.ch7.

[16] Daniel Delahaye, Stéphane Puechmorel, and Stephane Puechmorel. Modeling and Optimization of Air Traffic - Airspace Congestion Metrics. In *Modeling and Optimization of Air Traffic*. John Wiley & Sons, Incorporated, Somerset, UNITED STATES, 2013. ISBN 978-1-118-74376-8. URL http://ebookcentral.proquest.com/lib/delft/detail.action?docID=1251410.

[17] Erik Hollnagel. Resilience Engineering, 2016. URL http://erikhollnagel.com/ideas/resilience-engineering.html.

[18] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD-96 Proceedings*, KDD'96, page 226–231. AAAI Press, 1996. URL http://dl.acm.org/citation.cfm?id=3001460.3001507.

[19] Eurocontrol. European Aviation in 2040. Technical report, Eurocontrol, 2018. URL https://www.eurocontrol.int/sites/default/files/content/documents/official-documents/reports/challenges-of-growth-2018.pdf.

[20] Eurocontrol. *Network Operations Report 2018*. Eurocontrol, April 2019.

[21] FAA Eurocontrol. *Comparison Of Air Traffic Management-related Operational Performance U.S. - Europe*. Eurocontrol, March 2019.

[22] Eurostat. Main tables - Eurostat, 2019. URL https://ec.europa.eu/eurostat/web/transport/data/main-tables.

[23] Reza Faturechi and Elise Miller-Hooks. Travel time resilience of roadway networks under disaster. *Transportation Research Part B: Methodological*, 70:47–64, Dec 2014. ISSN 0191-2615. doi: 10.1016/j.trb.2014.08.007.

[24] Reza Faturechi, Eyal Levenberg, and Elise Miller-Hooks. Evaluating and optimizing resilience of airport pavement networks. *Computers & Operations Research*, 43:335–348, Mar 2014. ISSN 0305-0548. doi: 10.1016/j.cor.2013.10.009.

[25] Royce Francis and Behailu Bekera. A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering and System Safety*, 121:90–103, 2014. ISSN 09518320. doi: 10.1016/j.ress.2013.07.004. URL http://dx.doi.org/10.1016/j.ress.2013.07.004.

[26] Royce Francis and Behailu Bekera. A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety*, 121:90–103, January 2014. ISSN 09518320. doi: 10.1016/j.ress.2013.07.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832013002147.

[27] Olga Gluchshenko. Definitions of Disturbance, Resilience and Robustness in ATM Context. Technical report, DLR, 2012.

[28] Olga Gluchshenko and Peter Foerster. Performance based approach to investigate resilience and robustness of an ATM System. *Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)*, 2013. URL http://atmseminarus.org/seminarContent/seminar10/papers/277-Gluchshenko{_}0127130117-Final-Paper-4-8-13.pdf.

[29] Google. Google ngram viewer - robustness, resilience terms, 2019. URL https://books.google.com/ngrams.

[30] Lance H. Gunderson, C. S. Holling, and Stephen S. Light. *Barriers and Bridges to the Renewal of Regional Ecosystems*. Columbia University Press, June 1995. ISBN 978-0-231-51598-6.

[31] Andrew Hale and Tom Heijer. Defining Resilience. In *Resilience engineering : concepts and precepts*, page 8. Crc Press, 2006. ISBN 0-7546-4904-0.

[32] Devanandham Henry and Jose Emmanuel Ramirez-Marquez. Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering & System Safety*, 99:114–122, March 2012. ISSN 0951-8320. doi: 10.1016/j.ress.2011.09.002. URL http://www.sciencedirect.com/science/article/pii/S0951832011001748.

[33] Brian Hilburn. Cognitive complexity in air traffic control: a literature review. Technical report, Centre for Human Performance Research, January 2004.

[34] R. M. Hoffman. A Generalized Concept of Resilience. *Textile Research Journal*, 18(3): 141–148, 1948.

[35] C S Holling. Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics*, 4(1):1–23, November 1973. ISSN 0066-4162. doi: 10.1146/annurev.es.04.110173.000245. URL http://www.annualreviews.org/doi/10.1146/annurev.es.04.110173.000245.

[36] Seyedmohsen Hosseini, Kash Barker, and Jose E. Ramirez-Marquez. A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, 145:47–61, January 2016. ISSN 0951-8320. doi: 10.1016/j.ress.2015.08.006. URL http://www.sciencedirect.com/science/article/pii/S0951832015002483.

[37] Seyyed Shahab Hosseinian and Zahra Jabbarani Torghabeh. AJOR THEORIES OF CONSTRUCTION ACCIDENT CAUSATION MODELS : A LITERATURE REVIEW. *International Journal of Advances in Engineering & Technology*, 2012.

[38] International Civil Aviation Organization. Manual on global performance of the Air Navigation System. Technical report, International Civil Aviation Organization, Montréal, Quebec, 2009. OCLC: 751572818.

[39] Matthew R. Jardin and Arthur E. Bryson. Methods for Computing Minimum-Time Paths in Strong Winds. *Journal of Guidance, Control, and Dynamics*, 35(1):165–171, 2012. ISSN 0731-5090. doi: 10.2514/1.53614. URL https://doi.org/10.2514/1.53614.

[40] Jian Gang Jin, Loon Ching Tang, Lijun Sun, and Der-Horng Lee. Enhancing metro network resilience via localized integration with bus services. *Transportation Research Part E: Logistics and Transportation Review*, 63:17–30, Mar 2014. ISSN 1366-5545. doi: 10.1016/j.tre.2014.01.002.

[41] Jörg Leonhardt, Erik Hollnagel, and Luigi Macchi. A White Paper on Resilience Engineering for ATM. White Paper, Eurocontrol, September 2009.

[42] Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Sep 2009. ISBN 978-0-470-31748-8. Google-Books-ID: YeFQHiikNo0C.

[43] Eamonn Keogh and Shruti Kasetty. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371, Oct 2003. ISSN 1573-756X. doi: 10.1023/A:1024988512476.

[44] Will Koehrsen. The Poisson Distribution and Poisson Process Explained, January 2019. URL https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459.

[45] Piet Rietveld Mark J.Koetse. The impact of climate change and weather on transport: An overview of empirical findings. *Transportation Research Part D: Transport and Environment*, 14:205–221, May 2009. ISSN 1361-9209. doi: 10.1016/j.trd.2008.12.004.

[46] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[47] George Muller. Fuzzy architecture assessment for critical infrastructure resilience. *Procedia Computer Science*, 12:367–372, Jan 2012. ISSN 1877-0509. doi: 10.1016/j.procs.2012.09.086.

[48] Meinard Müller. *Dynamic Time Warping*, page 69–84. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-74048-3. doi: 10.1007/978-3-540-74048-3_4. URL https://doi.org/10.1007/978-3-540-74048-3_4.

[49] Nicoguaro. *PCA example*. Nicoguaro, Feb 2016. URL https://commons.wikimedia.org/wiki/File:GaussianScatterPCA.svg.

[50] International Civil Aviation Organization. Global Air Traffic Management Operational Concept. Technical report, International Civil Aviation Organization, 2005. URL https://www.icao.int/Meetings/anconf12/Document%20Archive/9854_cons_en%5B1%5D.pdf.

[51] Programminglinguist. *English: Dynamic Time Warping*. Programminglinguist, Jul 2015. URL https://commons.wikimedia.org/wiki/File:Dynamic_time_warping.png.

[52] Markku Rahiala. *Panel Data*, page 1047–1048. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_438. URL https://doi.org/10.1007/978-3-642-04898-2_438.

[53] James Reason. Human error: models and management. *BMJ*, 320(7237):768–770, March 2000. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.320.7237.768. URL https://www.bmj.com/content/320/7237/768.

[54] James Reason and Erik Hollnagel. European Organisation for the Safety of Air Navigation (EUROCONTROL). Technical report, Eurocontrol Agency, 2019. URL https://referenceworks.brillonline.com/entries/international-year-book-and-statesmens-who-s-who/*-SIM_org_39214.

[55] Resilience. *Oxford Dictionary of English*. Oxford University Press, 2019. URL https://en.oxforddictionaries.com/definition/resilience.

[56] Angela Weber Righi, Tarcisio Abreu Saurin, and Priscila Wachs. A systematic literature review of resilience engineering: Research areas and a research agenda proposal. *Reliability Engineering & System Safety*, 141:142–152, September 2015. ISSN 09518320. doi: 10.1016/j.ress.2015.03.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832015000654.

[57] L. Seyedhossein and M. R. Hashemi. Mining information from credit card time series for timelier fraud detection. In *2010 5th International Symposium on Telecommunications*, page 619–624, Dec 2010. doi: 10.1109/IS□.2010.5734099.

[58] Jonathon Shlens. A tutorial on principal component analysis. *arXiv:1404.1100 [cs, stat]*, Apr 2014. URL http://arxiv.org/abs/1404.1100. arXiv: 1404.1100.

[59] James P. G. Sterbenz, David Hutchison, Egemen K. Çetinkaya, Abdul Jabbar, Justin P. Rohrer, Marcus Schöller, and Paul Smith. Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. *Computer Networks*, 54(8): 1245–1265, Jun 2010. ISSN 1389-1286. doi: 10.1016/j.comnet.2010.03.005.

[60] Danijela Tadić, Aleksandar Aleksić, Miladin Stefanović, and Slavko Arsovski. Evaluation and ranking of organizational resilience factors by using a two-step fuzzy ahp and fuzzy topsis, 2014. URL https://www.hindawi.com/journals/mpe/2014/418085/.

[61] Romain Tavenard. tslearn: A machine learning toolkit dedicated to time-series data, 2017. https://github.com/rtavenar/tslearn.

[62] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, Mar 2013. ISSN 1573-756X. doi: 10.1007/s10618-012-0250-5.

[63] Eric W. Weisstein. Moving average, 2019. URL http://mathworld.wolfram.com/MovingAverage.html.

[64] David White. Network severe weather programme, 2012. URL https://www.eurocontrol.int/sites/default/files/content/documents/events/Presentations/120925-fmp-exchange-severe-weather-procedure-white.pdf.

[65] David D. Woods, Nancy Leveson, Professor, and Erik Hollnagel, Professor. *Resilience Engineering: Concepts and Precepts*. Chapman & Hall/CRC Press, Abingdon, UNITED KINGDOM, 2006. ISBN 978-0-7546-8136-6. URL http://ebookcentral.proquest.com/lib/delft/detail.action?docID=429564.

[66] Matei Zaharia, Reynold S. Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J. Franklin, and et al. Apache spark: A unified engine for big data processing. *Commun. ACM*, 59 (11):56–65, Oct 2016. ISSN 0001-0782. doi: 10.1145/2934664.

[67] Jun Zhang, Wei Liu, and Yanbo Zhu. Study of ADS-B Data Evaluation. *Chinese Journal of Aeronautics*, 24(4):461–466, August 2011. ISSN 1000-9361. doi: 10.1016/S1000-9361(11)60053-8. URL http://www.sciencedirect.com/science/article/pii/S1000936111600538.

[68] Christopher W. Zobel. Representing perceived tradeoffs in defining disaster resilience. *Decision Support Systems*, 50(2):394–403, January 2011. ISSN 0167-9236. doi: 10.1016/j.dss.2010.10.001. URL http://www.sciencedirect.com/science/article/pii/S0167923610001764.