Forschungsbericht 2024-10

An exploratory research on European ATM Network resilience through supervised learning

Rasoul Sanaei

Deutsches Zentrum für Luft- und Raumfahrt
Institut für Luftverkehr
Hamburg

DLR Deutsches Zentrum für Luft- und Raumfahrt

Forschungsbericht 2024-10

An exploratory research on European ATM Network resilience through supervised learning

Rasoul Sanaei

Deutsches Zentrum für Luft- und Raumfahrt
Institut für Luftverkehr
Hamburg

187 Seiten
64 Bilder
46 Tabellen
156 Literaturstellen

Deutsches Zentrum
DLR für Luft- und Raumfahrt

Herausgeber:

Deutsches Zentrum
für Luft- und Raumfahrt e. V.
Wissenschaftliche Information
Linder Höhe
D-51147 Köln

ISSN 1434-8454
ISRN DLR-FB-2024-10
Erscheinungsjahr 2024
DOI: 10.57676/wn5j-e845

Luftverkehrsflusssteuerung, Resilienz, Robustheit, Systemresilienz, Überwachtes Maschinelles Lernen, Konvolutionale Neuronale Netzwerke (CNN), Rekurrentes Neuronales Netzweke (RNN), Europäisches ATM-Netzwerk (EATMN), Netzwerkstatus, ATFM Verspätung

Rasoul SANAEI
DLR, Institut für Luftverkehr, Hamburg

Eine explorative Untersuchung zur Resilienz des europäischen ATM-Netzwerks mittels Supervised Learning Methoden
Dissertation, Technische Universität Hamburg

Air Traffic Management (ATM) ist ein komplexes System, das durch verschiedene Aspekte wie mehrschichtige Management-Subsysteme, verschiedene Geschäftsmodelle für Fluggesellschaften und ein dynamisches Umfeld herausgefordert wird. Diese Aspekte setzen den geplanten Betrieb täglichen Störungen aus, die zu verspäteten Passagieren führen. Störungen einerseits und Systemüberlastung andererseits deuten darauf hin, dass ATM resilient werden muss. Durch einen datengestützten Ansatz arbeitet diese Dissertation an der Konzeptualisierung der Resilienz des europäischen ATM-Netzes. Da die Systemresilienz mit dem Situationsbewusstsein korreliert, schlägt diese Studie eine Netzwerkzustandsdefinition vor und untersucht dann Lernmethoden, um bessere Vorhersagen zu extrahieren (Verspätung und verspätete Flüge).

Air Traffic Flow Management (ATFM), Resilience, Robustness, System resiliency, Supervised Machine Learning, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), European ATM Network (EATMN), Network State, ATFM Delay

(Published in English)

Rasoul SANAEI
German Aerospace Center (DLR), Institute of Air Transport, Hamburg

An exploratory research on European ATM Network resilience through supervised learning

Doctoral Thesis, Hamburg University of Technology

Air Traffic Management (ATM) is a complex system challenged by different aspects such as layered management subsystems, various business models for airlines, and dynamic environment. These aspects expose planned operations to daily disruptions leading to delayed passengers. Disruptions on one hand and system saturation on the other hand, suggest that ATM needs to become more resilient. Through a data-driven approach, this thesis works on conceptualizing the European ATM network resilience. Since system resiliency is correlated with situational awareness, this study proposes a network state definition and then explores learning methods to extract better predictions (delay and delayed flights).

An exploratory research on European ATM Network resilience through supervised learning

Dissertation approved by the Doctoral Degree Committee of
Hamburg University of Technology

in pursuit of the academic degree of
Doctor of Engineering (Dr.-Ing.)

written by
Rasoul Sanaei

from
Kazeroon, Iran

2024

List of reviewers

Chair of the exam Committee	Prof. Dr.-Ing. Ralf God
1. Examiner:	Prof. Dr.-Ing. Volker Gollnick
2. Examiner:	Assoc. Prof. Bruno F. Santos (TUDelft)

Date of the oral examination: 02.05.2024

To my beloved family: Lotfollah, Zohreh, Reza & Roya

Acknowledgment

This manuscript is a memoire of a unique journey of my life that was a chapter of numerous experiences many of them along the support found on the way for which I am grateful.

To my Brother, Dr. Reza Sanaee, for being a light house from the beginning to the end, showing how potentials can be materialized into achievements.

To DLR that proved science is above all geographical and political circumstances by paving the way for international scholars to conduct their researches in an interactive, learning environment.

I would also like to cherish the key decisions and moments of trust from my supervisory panel, Prof. Volker Gollnick, Dr. Florian Linke and Dr. Alexander Lau that motivated and allowed me to pursue new ideas, no matter how challenging they appear to be.

And to my friends, fellow DLR researchers and experts of EUROCONTROL for contributing to my learning path to grow from an industrial engineer to become a researcher in aviation.

Summary

Air transportation is managed to accommodate more flights every day with solutions such as new business models for airlines, but at cost of rising congestion. Solutions like low-cost carriers stimulated the traffic demand to a higher growth rate. Such consequences exacerbate the gap between demand and airspace capacity despite planning procedures. Congestion adds to the complexity of air traffic management (ATM), that challenges planning phases of air traffic flow and capacity management (ATFCM). Factors such as dynamic capacities and built-in flexibilities expose the ATM system to emergent Demand-Capacity Balancing (DCB) issues. In an exploratory attempt this study considers resilience as a systematic solution to cope with emergent dynamics. A resilient system basically accepts the dynamic environment and tries to manage the raised complexities with performance variability.

Resilience is intertwined with situational awareness. Thus, after conceptually modeling the European ATM network (EATMN) resilience, a proposed methodology determines the network state (based on large scale disruptions) and then the thesis delivers a prediction method to assess reviving solutions against emergent imbalances. Most of emergent disruptions (DCB issues) are currently managed by simulation-based assessments that require high computational power and access to different data bases. In comparison, this study is a data driven approach with statistical evaluations and supervised learning algorithms focused on disruptions rather than constant monitoring of demand (traffic) and capacity.

Throughout the methodology chapter, network state is defined based on statistical inferences and predictability of disruptions is improved by supervised learning. More specifically, situational awareness is improved by daily network predictions from a deep Convolutional Neural Network (CNN). The model exploits characteristics of disruptions such as their spatio-temporal dimension. Furthermore. the resilient path to revive the network is an accumulation of individual corrective actions (i.e. capacity regulations). Therefore, a Recurrent Neural Network (RNN) is proposed to predicts the impact (delay) of corrective actions, because at a higher granularity temporal dimension of data is more informative.

The conceptual achievements of the thesis support the operational need to declare solid cases of network anomaly based on performance indicators while authorities such as the European Aviation Crisis Coordination Cell (EACCC) rely on safety metrics. This paradigm shift is on one hand evaluated by 2018 use cases and received expert feedbacks from an EU-SESAR project. On the other hand, RNN results are evaluated against results from the deep CNN model. In fact, the post-operational dataset (regulations from 2015 to 2018) show that a network-wide prediction that is accumulated from RNN predictions has an accuracy of 97 percent for cumulative daily delays. This actively demonstrates that even if such a high precision cannot be realized throughout operations, still the proposed approach not only delivers an improved situational awareness but also enables the network manager to foresee the network impact of submitted list of corrective regulations from local authorities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

A-CDM	Airport Collaborative Decision Making
ACC	Area Control Centre. The part of ATC that is concerned with en-route traffic coming from or going to adjacent centers or APP. It is a unit established to provide air traffic control service to controlled flights in control areas under its jurisdiction.
AD	Aerodrome
ADP	ATFM Daily Plan
ADR	Airport Departure Rate
ADS-B	Automatic Dependent Surveillance-Broadcast
AFP	Airspace Flow Program (US)
AIM	ATFCM Information Message (Europe)
AIRAC	Aeronautical Information, Regulation and Control
AIS	Aeronautical Information Service
ANM	ATFCM Notification Message (Europe)
ANS	Air Navigation Service. A generic term describing the totality of services provided in order to ensure the safety, regularity and efficiency of air navigation and the appropriate functioning of the air navigation system.
ANSP	Air Navigation Services Provider
AO	Aircraft Operator also referred as AU
APP	Approach Control Unit
AR	Alternative routing scenario
ARTCC	Air Route Traffic Control Center, the equivalent of an ACC in Europe.
ASM	Airspace Management
ASMA	Arrival Sequencing and Metering Area. The volume around an aerodrome taken as a reference for measuring the efficiency in handling the arrival flow. Typically, it is a cylinder of 40 NM radius.
ASP	Arrival Spacing (US)
ASPM	Similar to European NMIR, FAA Aviation System Performance Metrics is an online access system (https://aspm.faa.gov) that provides data on flights to and from the ASPM airports and all flights by ASPM carriers, including flights by those carriers to international and domestic non-ASPM airports. All IFR and some VFR flights are included.
ATC	Air Traffic Control. A service operated by the appropriate authority to promote the safe, orderly and expeditious flow of air traffic.
ATCO	Air Traffic Control Officer/Air Traffic Controller
ATCSCC	Air Traffic Control System Command Centre (US) is a facility dedicated to balancing the air traffic demand with system capacity (similar to DCB operations in Europe).
ATFCM	Air Traffic Flow and Capacity Management, extends ATFM to include the optimization of traffic patterns and capacity management. Through managing the balance of capacity and demand, the aim of ATFCM is to enable flight punctuality and efficiency, according to the available resources with the emphasis on optimizing the network capacity through the collaborative decision-making processes.
ATFM	Air Traffic Flow Management, is established to support ATC in ensuring an optimum flow of traffic to, from, through or within defined areas during times when demand exceeds, or is expected to exceed, the available capacity of the ATC system, including relevant aerodromes. In contrast to ATFCM, ATFM considers capacity as an input constraint.
ATFM delay (CFMU)	The duration between the latest requested (by AO) take-off time and the take-off slot given by the CFMU. More specifically it is the difference between Calculated Take Off Time and Estimated Take Off Time (CTOT-ETOT).

ATFM Regulation	When traffic demand is anticipated to exceed the declared capacity in en-route control centers or at the departure/arrival airports, ATC units may call for an "ATFM regulations". It is a requested time window with reduced entry rates for a reference airspace.
ATM	Air Traffic Management. A system consisting of a ground part and an airborne part, both of which are needed to ensure the safe and efficient movement of aircraft during all phases of operation. ATM is comprised of functionalities such as of Air Traffic Services (ATS), Airspace Management (ASM) and Air Traffic Flow Management (ATFM).
ATO	Air Traffic Organization (US), is the operational arm of the FAA (similar to functionality of Network Manager in EUROCONTROL).
ATS	Air Traffic Service. A generic term meaning variously, flight information service, alerting service, air traffic advisory service and air traffic control service.
AU	Airspace User also referred as Aircraft operator (AO).
Bad weather	For the purpose of this thesis, "bad weather" is defined as any adverse weather condition (e.g. strong wind, low visibility, snow) which causes a significant drop in the available airport capacity.
CAA	Civil Aviation Authority
CAS	Complex Adaptive Systems
CASA	Computer Assisted Slot Allocation (CASA) system is a part of the Enhanced Tactical Flow Management System (ETFMS) which provides automatic message exchange in the form of Slot Allocation Messages and other Air Traffic Flow and Capacity Management (ATFCM) messages. CASA is triggered by activating a regulation.
CBA	Cross-Border Area, is an airspace restriction or reservation established over international borders for specific operational requirements.
CCF	Combined Control Facility (US): An air traffic control facility that provides approach control services for one or more airports as well as en-route air traffic control (center control) for a large area of airspace. Some may provide tower services along with approach control and en-route services. CCF also includes Combined Center Radar Approach (CERAP) facilities.
CCSD	Collaborative Constraint Situation Display (US)
CDF	Cumulative Distribution Function
CDM	Collaborative Decision Making
CDR	Coded Departure Route (US)
CFMU	Central Flow Management Unit established in 1995 (See NMOC)
CNN	Convolutional Neural Network
CODA	EUROCONTROL Central Office for Delay Analysis
CONUS	Continental United States, see US CONUS
CTOP	Collaborative Trajectory Options Program (US)
CTOT	Calculated take-off Time
DCB	Demand Capacity Balancing
DCNN	Deep Convolutional Neural Network, or Deep CNN
DDR2	Demand Data Repository
DOF	Date Of Flight, A date of flight shall be included in all messages (esp.in item 18 of submitted flight plan) where the estimated off-blocks time is more than 24 hours in advance, but not more than 120 hours (5 days) in advance the time at that message is processed by the IFPS.
DRR	Disaster Risk Reduction
DSNA	Direction des Services de la Navigation Aerienne, DSNA is the ANSP of France.
DSP	Departure Spacing (US)
DTW	Departure Tolerance Window
EATMN	European ATM Network
EC	European Commission
ECAC	European Civil Aviation Conference

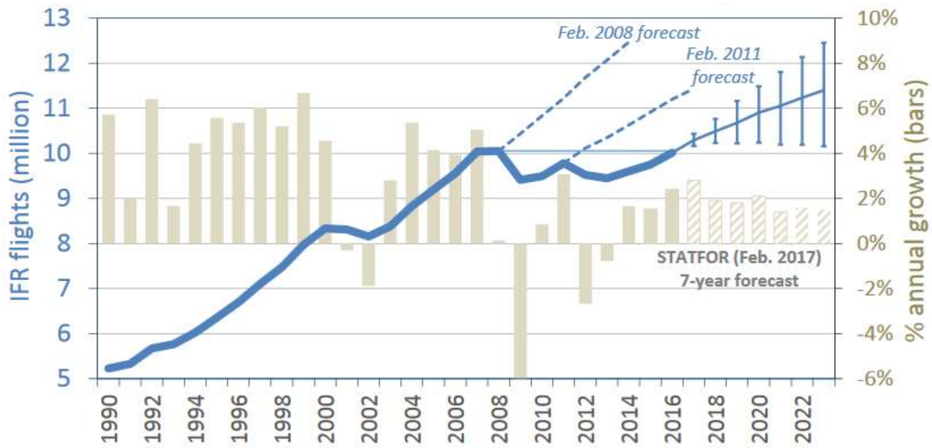
EDCT	Estimate Departure Clearance Time. EDCT is a long-term Ground Delay Program (GDP), in which the Command Centre (ATCSCC) selects certain flights heading to a capacity limited destination airport and assigns an EDCT to each flight, with a 15-minute time window.
EEC	EUROCONTROL Experimental Centre in Brétigny-sur-Orge, France
EOBT	Estimated Off Block Time, the estimated time at which the flight starts to be pushed back from the gate and start to taxi.
ENAIRE	ENAIRE is the air navigation and aeronautical information service provider in Spain.
ENAV	ENAV SPA, is the Italian ANSP with four Area Control Centers (ACC).
ETA	Estimated Time of Arrival
ETFMS	Enhanced Tactical Flow Management System (Europe) provides enhanced tactical data to all operational stakeholders, regardless of national boundaries, language, or equipment. ETFMS facilitates improvements in flight management from the pre-planning stage to the arrival of the flight. It maximizes the updating of flight-related data and thus improves the real picture of a given flight.
EU	Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxemburg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and United Kingdom. All these 28 States are also Members of the ECAC.
EUROCONTROL	The European Organization for the Safety of Air Navigation. It comprises Member States and the Agency.
EUROCONTROL Member States (2023)	Since 1963: Belgium, France, Germany, Luxembourg, Netherlands, United Kingdom; Ireland (1965), Portugal (1986), Greece (1988), Malta (1989), Turkey (1989), Cyprus (1991), Albania, Armenia, Austria, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Georgia, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, Moldova, Monaco, Montenegro, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia and Ukraine. Comprehensive agreement states: Israel (2016), Morroco (2016).
FAA	US Federal Aviation Administration
FAA-ATO	US Federal Aviation Administration - Air Traffic Organization
FAB	Functional Airspace Block (Europe) means an airspace block based on operational requirements and established regardless of state boundaries, where the provision of air navigation services and related functions are performance-driven.
FCA	Flow Constrained Area (US)
FCFS	First Come First Serve principle
FDP	Flight data processing
FEA	Flow Evaluation Area (US)
FEI	Flight Efficiency Initiative
FIR	Flight Information Region. An airspace of defined dimensions within which flight information service and alerting service are provided.
FL	Flight Level; the altitude above sea level in 100-foot units measured according to a standard atmosphere. Basically, a flight level is an indication of pressure, not of altitude. Flight levels are used mainly above the transition level (e.g. FL135) and to indicate altitude below the transition level, feet are used (e.g. 4000 ft).
FL	Level capping scenario; this means that flights that meet certain conditions would be subject to a restriction, e.g. all flights departing from ZZZZ must be at FLXXX or below over point ENTRY. This is the most commonly used STAM.
FMP	Flow Management Position (also referred to as LTM: Local Traffic Manager). The FMP's role is, in partnership with the NM, to act in such a manner so as to provide the most effective ATFCM service to ATC and AOs. Each FMP area of responsibility is normally limited to the area for which the parent ACC is responsible including the area(s) of responsibility of associated Air Traffic Services (ATS) units as defined in the

	NM Agreement. However, depending on the internal organization within a State, some FMPs may cover the area of responsibility of several ACCs, either for all ATFCM phases or only for part of them. The size of individual FMPs will vary according to the demands and complexities of the area served. [1]
FMS	Flight Management System
FOQA	Flight Operational Quality Assurance data
FRA	Free Route Airspace
FTS	Fast Time Simulation (FTS) is a technique to estimate the capacity of each ATC sector.
GDP	Ground Delay Program (US)
General Aviation	All civil aviation operations other than scheduled air services and non-scheduled air transport operations.
GS	Ground Stop (US)
IATA	International Air Transport Association (www.iata.org)
ICAO	International Civil Aviation Organization
ICR	Integrated Collaborative Rerouting (US)
IFR	Instrument Flight Rules; one of the two types of regulations that apply to flights (IFR and VFR). Visual Flight Rules, is mostly for general aviation and small sized aircraft such as Cessna Skyhawk. IFR flights include commercial flights and Cargo flights.
KPI	Key Performance Indicator
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean squared error
NAS	National Airspace System
NATS	National Air Traffic Control Services and Civil Aviation Authority of United Kingdom
NextGen	The Next Generation Air Transportation System (NextGen) is the name given to a new NAS due for implementation across the United States in stages between 2012 and 2025.
NLP	Natural language processing, is a field of research that studies the capabilities of learning algorithms to enable a computer to "understand" the contents of documents.
NM	Network Manager (EUROCONTROL) or Nautical Mile (1.852 km)
NMIR	The Network Manager Interactive Reporting
NMOC	EUROCONTROL's Network Management Operations Centre located in Brussels (formerly CFMU).
NN	Neural Network
NOP	Network Operations Plan or Network Operations Portal
NRP	North American Route Program (US & Canada)
OBT	Off-Block Time is the time defined in the flight plan at which the flight leaves its parking position with a push back.
OD	Origin Destination, also referred as city pairs
OPSNET	The Operations Network is the official source of NAS air traffic operations and delay data. The data is used to analyze the performance of the FAA's air traffic control facilities.
Percentile	A percentile is the value of a variable below which a certain percent of observations fall. For example, the 80th percentile is the value below which 80 percent of the observations may be found.
PRC	Performance Review Commission
PRU	Performance Review Unit (Europe) which is in charge of performance review report (PRR).
Punctuality	On-time performance with respect to published departure and arrival times.
RAD	Route availability document
RE	Resilience Engineering
ReLU	Rectified Linear Unit
RF	Random Forest
RFR	Random Forest Regression

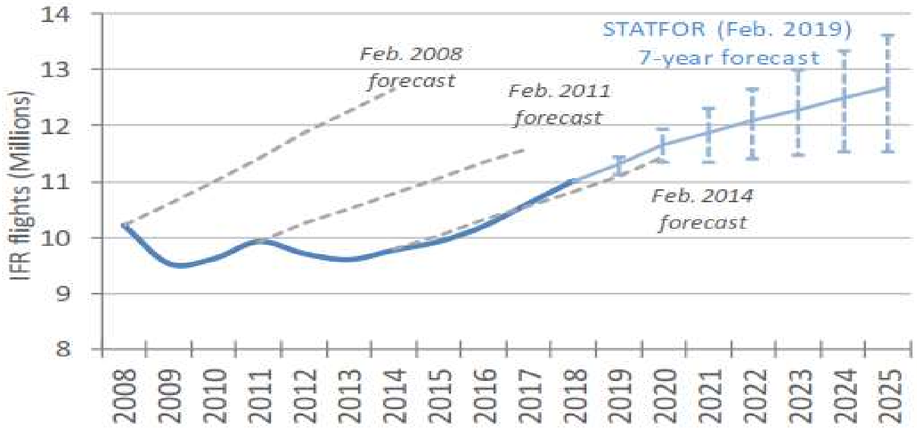
RL	Reinforcement learning & Reference Location
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RR	Rerouting scenario (Europe) & Required Reroutes TMI (US)
R^2	R squared; coefficient of determination
RTS	Real Time Simulation is a method of improving the ATC capacity estimation.
Separation minima	The minimum required distance between aircraft. Vertically usually 1,000 ft below flight level 290, 2,000 ft. above flight level 290. Horizontally, depending on the radar, 3 NM or more. In the absence of radar, horizontal separation is achieved through time separation (e.g. 15 minutes between passing a certain navigation point).
SESAR	Single European Sky ATM Research (SESAR) project was set up in 2004 as the technological pillar of the Single European Sky initiative. SESAR is founded by the European Union and EUROCONTROL.
SGD	Stochastic Gradient Descent
SNN	Sequential Neural Network
Slot (ATFM)	A take-off time window assigned to an IFR flight for ATFM purposes
STAM	Short Term ATFCM Measure
STATFOR	Statistics & Forecasts Service
STD	Scheduled Time of Departure
STW	Slot Tolerance Window
Summer season	IATA Summer schedule - begins on the last Sunday of March and ends on the last Saturday of October.
SVM	Support Vector Machines
SVR	Support Vector Regression
SWAP	Severe Weather Avoidance Plan (US)
Taxi-in	The time from touch-down to arrival block time.
Taxi-out	The time from off-block to take-off, including eventual holding before take-off.
Thales	Thales Group is a French multinational company that designs and builds electrical systems and provides services for the different industrial sections including aviation.
TFMS	Traffic Flow Management System (US)
TMA	Terminal Maneuvering Area
TMI	Traffic Management Initiative (US)
TOS	Trajectory Option Set (US)
TSA	Temporary Segregated Area
TSD	Traffic Situation Display (US)
TV or TFV	Traffic Volume (Europe) A computer code used to identify the number of flights over an airspace, point, aerodrome or set of aerodromes in order that they can be monitored or regulated within the tactical/pre-tactical ATFCM.
UAV	Unmanned Aerial Vehicle
UIR	Upper Information Region
US	United States of America
US CONUS	The 48 contiguous States located on the North American continent south of the border with Canada, plus the District of Columbia, excluding Alaska, Hawaii and oceanic areas
VFR	Visual Flight Rules

1. Introduction

Air transportation is constantly changing by adopting new technologies and hosting new business models. In general, the demand for aviation services is increasing, and its persistent growth is predicted by long-term forecasts (e.g. from EUROCONTROL [2] in Figure 1-1). Although the actual data deviate from prediction, yet in normal environment demand is growing at different rates. These predictions include factors such as economic growth, fuel prices, load factors, high-speed rail network development and airline schedules[1] in three scenarios (global growth, regulation & growth, fragmenting world).



a) actual and forecasts for 1990-2022 (10 million IFR flights in 2016; i.e. +2.4% compared to 2015)



b) actual and forecasts for 2008-2025 (11 million IFR flights in 2018; i.e. +3.8% compared to 2017)

Figure 1-1 Growth of European IFR flights, *[2]*

The pessimistic scenario of fragmenting world addresses major safety issues such as volcanic eruption in 2010 and COVID-19 pandemic. But COVID aftermath requires more studies (as in [3]) since fragmented world is no more considered as a pessimistic scenario with regard to realities such as changed mentality in business models (e.g. digitalization and home-office). Nevertheless, the risk of eventually facing a saturated Air Traffic Management (ATM)

[1] The forecast ignores Emissions Trading Scheme (ETS) and Carbon Offsetting & Reduction Scheme for International Aviation (CORSIA). DLR’s institute of air transport is studying the impact of emissions in different aspects which are relevant to such predictions.

network is inevitable even in case of a pandemic, in which both traffic demand and airspace capacity will be degraded simultaneously. In fact, during COVID-crisis, staff management at Area Control Centers (ACCs) has been proved to limit capacity because staff availability for different roles and working positions (such as executive and planner controllers) at any control center is a key factor in airspace capacity.

Saturation risk can also be tracked in industrial forecasts such as Global Market Forecast (GMF) from Airbus [4]. Despite having less relevance to ATM topics, industrial forecasts provide a picture of fleet expansion. For instance, in Figure 1-2, Airbus predicts to deliver more than 39 thousand new aircrafts in next twenty years. In the same time window, Boeing also forecasts [5] to deliver almost 25 thousand new airplanes. A total of added 64 thousand airliners will push ATM services to reach a much higher level of efficiency.

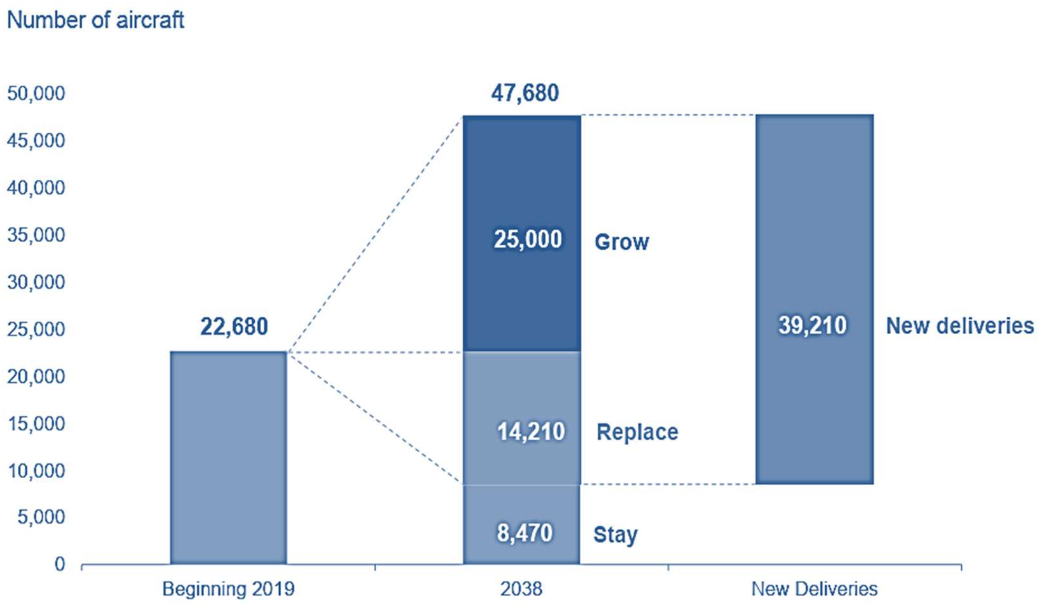


Figure 1-2 Demand for new Airbus aircraft delivery by 2038, *[4]*

Similarly, ICAO (International Civil Aviation Organization) predicts that global revenue in aviation will continue to grow annually at 4.1% rate (Figure 1-3) and this forecasted increase is accompanied by a 3.9% increase rate for freight traffic from 2015 to 2035.

Although every long-term prediction serves a specific objective but in general there are some factors that are missing such as Technology Readiness Level (TRL) or saturation limit. Yet, hints of such aspects can be spotted in annual reports rather than predictions. For instance, EUROCONTROL [6] reported some effects of congestion in 2018. It was observed that in top 30 busiest European airports, departure management was a challenge that led to a general increase of the additional taxi-out times and ATC pre-departure delays. What intensifies this is that such a degradation has been observed despite of solutions like Airport Collaborative Decision Making (A-CDM) concept that supports Air Traffic Flow and Capacity Management (ATFCM) to reduce delays, improve predictability and optimize the utilization of resources [7].

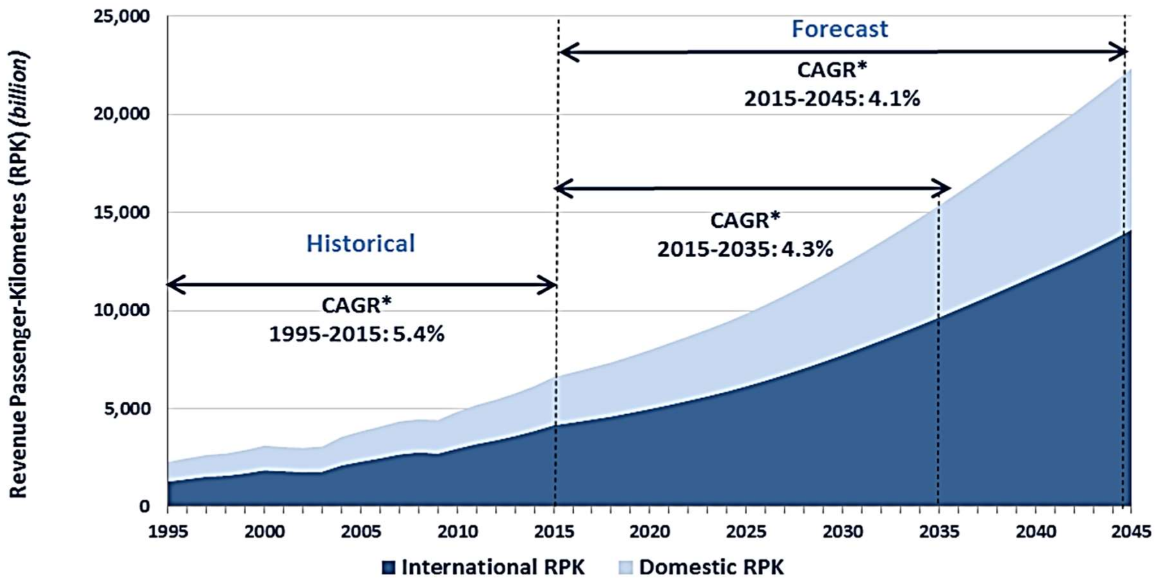


Figure 1-3 World Total revenue in aviation: history and forecasts, *[8]*. *CAGR: Compound Annual Growth Rate

Upgrading industrial solutions are less likely to provide a total solution for congestion. This claim is reinforced by Performance Review Report [6], that captured the highest inefficiencies in 2018 for flights arriving at airport with modern solutions such as Frankfurt (FRA), Paris Charles-de-Gaulle (CDG), London (LHR) and Paris-Orly (ORY).

Such observations remind that the rising pattern of demand in aviation cannot be efficiently accommodated by relying only on different industrial solutions for different stakeholders. But a systematic approach is needed to address efficiency in a saturated environment. This thesis is an endeavor to explore a new perspective (i.e. resilience) to pave the way in resolving the saturation problem. The methodology is based on a data driven approach and learning models. The thesis claims that resilience is a sound choice with regard to complexity of European ATM network. This chapter elaborates more on this claim in four sections: firstly, statistics and figures are provided to review the dimension of increasing demand that leads to growing delays; secondly, the major ATM procedures against delays are discussed to orient the thesis approach and discuss obstacles and limitations. Next, the research question is formulated with regard to current European research program, ATM resiliency and realization of ATM as a system. This chapter is then concluded by providing the study outline.

1.1. Motivation

As mentioned the emerging problems of reaching the saturation level at ATM add to the importance of efficiency. COVID experience implied that congestion is not necessarily a result of excessive demand but also capacity shortages (e.g. ATC staffing issues due to infected controllers) may still cause the same challenges that lead to longer delays. This section shortly addresses the general growth of demand and delay prior to enumerating current traffic flow management procedures in the following section.

1.1.1 Growing demand

Demand figures are showing an increasing trend in most congested airspaces in United States and Europe. Such similarities are more evident in comparative reports that have been published by a mutual effort from FAA and EUROCONTROL since 2009.

Comparative figures [1, 9] do imply that Europe has a bigger increase in its traffic demand. Figure 1-4 takes the year 2000 as the baseline and shows the detrimental impact of 2008 economic crisis on both traffic situations. Europe experienced a faster recovery due to factors of being an aggregated airspace with different economies compared to FAA in US that provides services on national environment (Figure 1-5).

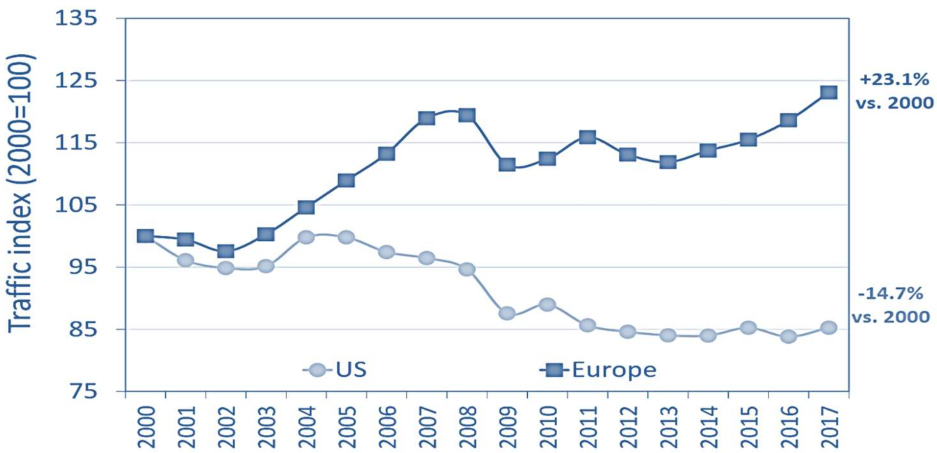


Figure 1-4 Evolution of IFR traffic US vs Europe, *[9]*

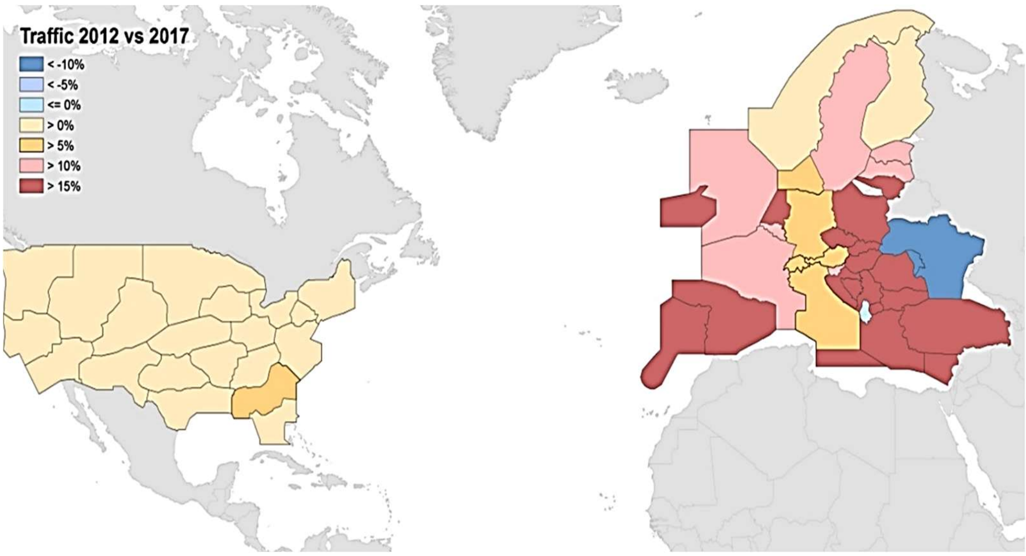


Figure 1-5 Faster IFR traffic growth in Europe with different states and economies, *[9]*

Such different geo-economic factors make FAA and EUROCONTROL to have different planning horizons to accommodate the traffic demand. Compared to American ATM system, ATFCM in Europe is managed in four phases [10]: strategic, pre-tactical, tactical and post-operations. These phases and their definition contribute to standardization of European ATM. The reference day for these rolling phases is the target date (i.e. day of operations) at which the actual flights take place. EUROCONTROL describes planning phases as follows:

- ***Strategic Flow Management*** includes research, planning and coordination activities that are concluded seven days or more prior to the day of operations. This phase considers procedures and measures toward early identification of major demand/capacity imbalances (e.g. traffic axis management). The output of this phase is the Network Operations Plan (NOP).
- ***Pre-Tactical Flow Management*** is applied during the six days prior to the day of operations and consists of planning and coordination activities to study the demand on target date, comparison against predicted capacity, and making necessary adjustments to the strategic plan (e.g., sector configuration management). Apart from coordination activities based on predictions and available capacity a wide range of appropriate ATFCM measures is proposed in form of ATFCM Daily Plan (ADP)[1].
- ***Tactical Flow Management*** takes place on the day of operations and involves adaptation and implementation of ADP into flight operations. The objective is to ensure that strategic and pre-tactical corrective measures are the minimum required to solve the DCB issues. The provision of accurate information is of vital importance, since it feeds short-term forecasts that reveal the impact of events.
- ***Post Operational Analysis*** is the final step of analysis that investigates and reports on operational processes. This phase compares the anticipated outcome against the actual measured outcome, generally in terms of delay and route extension with respect to performance targets.

These planning horizons are the actual procedures that eventually meet the expected demand in discussed long-term predictions. Among all, tactical phase links plans and predictions to records of performance. The last-minute demand predictions are done at this phase, with the challenge to consider the cumulative uncertainty at a much higher granularity than the whole European airspace. In fact, at tactical phase the load in each Traffic Volume[2] (TV) with a limited capacity is predicted (Figure 1-6).

[1] These measures are offered to all stakeholders by different services such as ATFCM Notification Messages (ANMs) and Initial Network Plan (INP).
[2] TV is a commonly used expression for number of flights over an airspace or a reference location (e.g. an airports) in a specific time window.

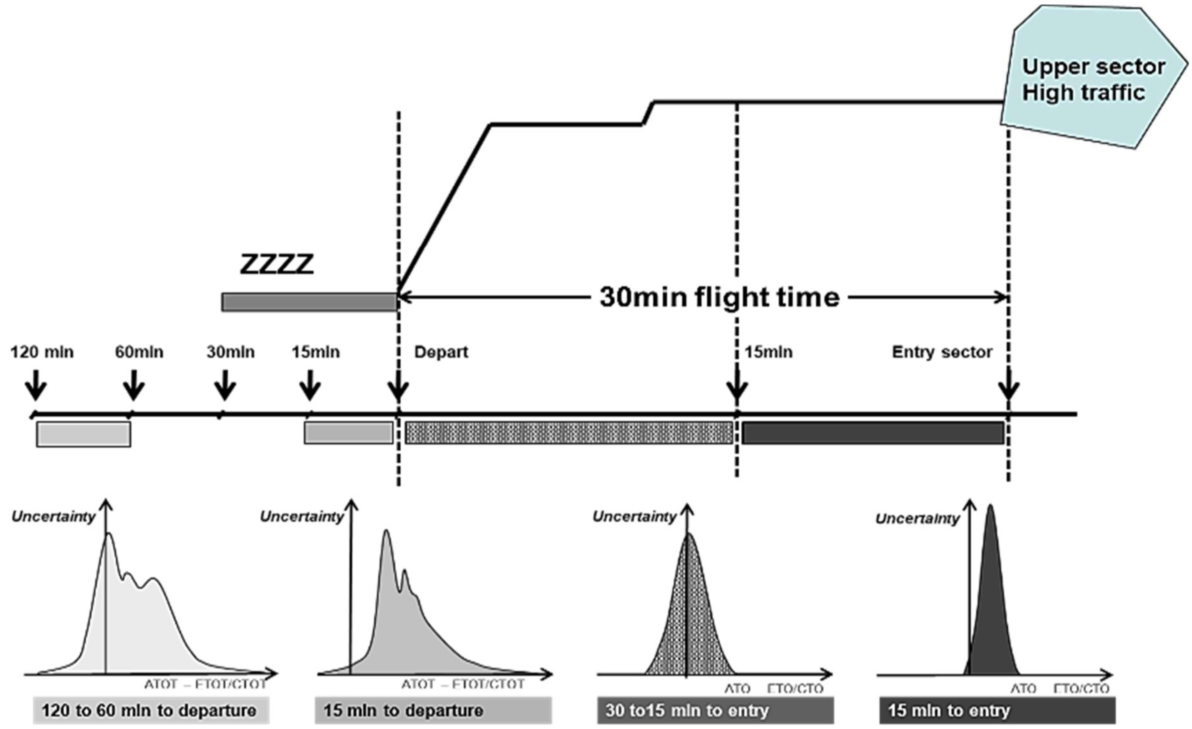


Figure 1-6 Distribution of demand uncertainty over a flight profile to enter a traffic volume, i.e. entry time, **[11]**

Figure 1-6 elucidates the different uncertainty distributions for a flight through different prediction intervals. The distribution curves have different shapes before and after Actual Take-Off Time (ATOT). This figure illustrates the dynamic uncertainty in a typical demand prediction for a target flight in a specific TV (note that each flight trajectory connects multiple TVs to reach the destination. In this approach, the cumulative value of predicted demand/load at a target TV has different shapes based on selected prediction horizon. The uncertainty at each TV is at much higher magnitude since a TV hosts multiple flights from different traffic flows and entry/exit points.

The described uncertainty to foresee the demand for a couple of hours in tactical phase is only a fraction compared to uncertainty of delivered plans from strategic and pre-tactical phases when it comes to making decisions about delaying or rerouting a flight at day of operations. Along with predicted saturation in long-term, the uncertainty of demand prediction in tactical phase pushes the ATM community to seek innovative approaches to control resonance of prediction errors in corrective measures that can trigger secondary problems with delay.

1.1.2 Growing delay

The discussed growing demand and efforts to support real-time decision making are the frontier of ATM evolution. Despite endeavors to modernize ATM in Europe and US, the recorded data trends show a degraded delay figure (Figure 1-7). Comparative reports [9] on different management methodologies almost outline similar challenges regarding delay.

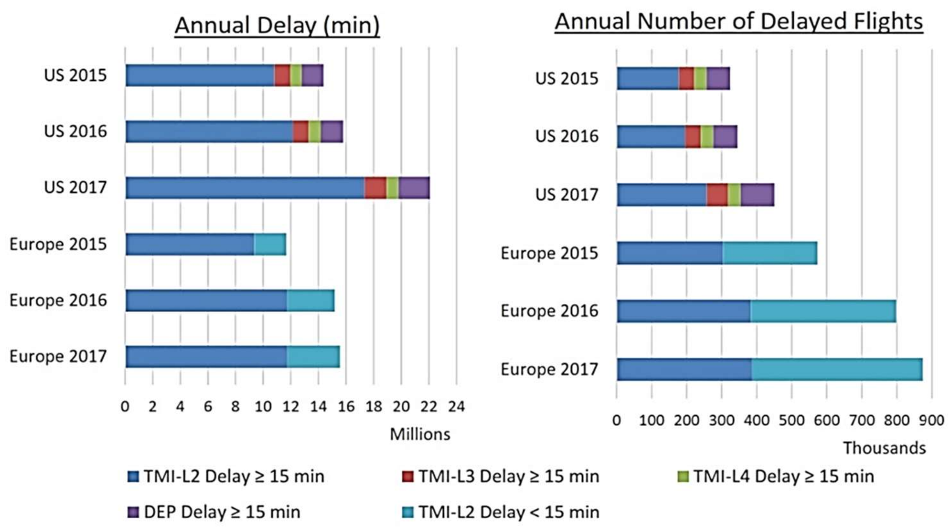


Figure 1-7 Reported annual delay and delayed flights in US and Europe, **[9]**

Figure 1-7 shows the increasing delays with a significant portion of delays to take more than 15 minutes. Such delays are either ground delay (US terminology) or ATFM[1] delay (European term). Although delay in US is higher but the number of delayed flights is much more in Europe that might be due to different strategies. ATM authorities in US and Europe decompose [9] the overall delay and number of delayed flights in Figure 1-7 as:

"in U.S.:

- *10% of the recorded delay is departure delay, that accounts for a bigger proportion (20%) of the delayed flights;*
- *10-15% of the recorded delay is ATC-related (TMI-L3 and TMI-L4); which is imposed on 25% of the delayed flights.*
- *75-80% of the recorded delay is ATFM-related (TMI-L2); and more than half (55%) of the delayed flights are affected by this type of delay and*

in Europe:

- *75-80% of total recorded delay (≥ 15 minutes) is from approximately half (45-55%) of the delayed flights. The other half of the delayed flights experiences only small delays.*
- *Despite the large number of affected flights, the 'small delays' account for only 20-25% of the total annual delay.*

In both regions, if a flight is delayed, the cause most likely (75-80%) is an ATFM issue. In Europe when traffic demand is anticipated to exceed the available capacity (in en-route sectors or at airports) Air Traffic Control (ATC) units may contact the local Flow Management Position (FMP) to initiate an ATFM measure or regulation. Flights that cross these areas receive an ATFM delay with a new departure time from EUROCONTROL as the Network Manager (NM). Basically [7]:

[1] Air Traffic Flow Management (ATFM)

> *ATFM delay is defined as the duration between the last Estimated Take-Off Time (ETOT) and the Calculated Take-Off Time (CTOT) allocated by the Network Manager. ATFM delay comprises both Airport ATFM delay and En route ATFM delay.*

ETOT is the airline's requested departure time that is driven by a set of airline constraints. These constraints depend on passengers, airline schedule, fleet management, or operational limitations. ETOT has its own line of research that addresses concepts like punctuality.

CTOT is calculated by a mathematical model that considers active regulations. This mathematical model is called Computer Assisted Slot Allocation (CASA) in Europe. CASA indirectly takes in the request from airline in form of a filed flight plan (defining ETOT) and with respect to active ATFM regulations assigns a departure slot (CTOT) as its output [10].

CASA algorithm is part of the Enhanced Tactical Flow Management System (ETFMS) that generates CTOTs. EFTMS provides tactical data and has two main functions:

1. calculating traffic demand and occupancy counts based on the information from Initial Flight Planning System (IFPS), and
2. balancing demand with regard to capacity and sequencing the flights by CASA.

ETFMS in European ATM system (similar to TFMS in US) performs a number of key activities: flight and pre-flight data collection, flight activation monitoring, entry and sector occupancy counting, flight profile calculations and data distribution. However, despite of these centralized tasks the cost of delay to airlines is rising each year. As in Figure 1-8 (left) while Air Navigation Service (ANS) provision costs remained almost at the same level for airlines, en-route costs grow each year. Also, the increasing gap between annual en-route and airport ATFM delay, Figure 1-8 (right), can be a sign of saturated traffic system that significantly suffers from en-route capacity issues.

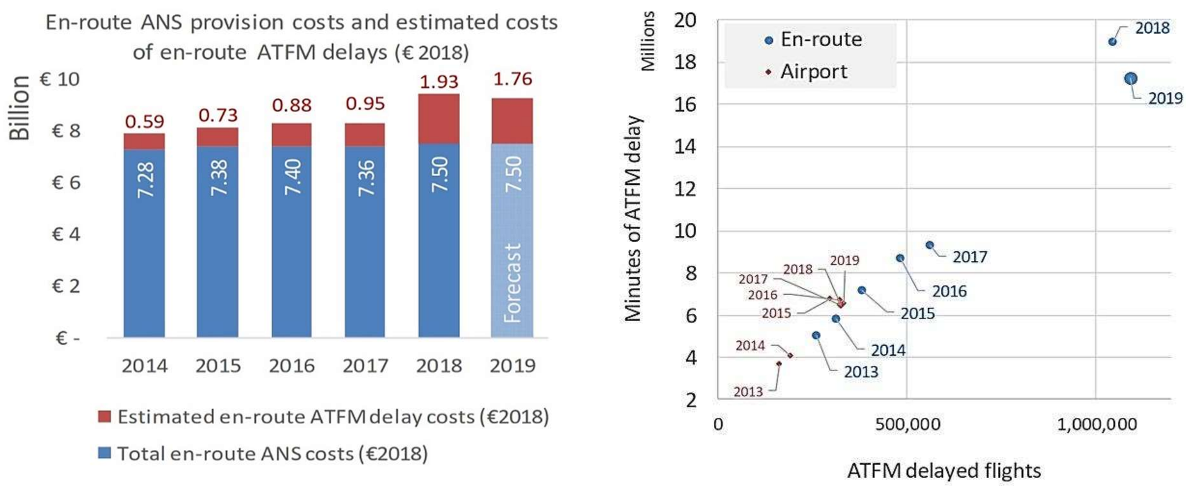


Figure 1-8 Rise of en-route ATFM delay despite controlled ANS provision costs, **[12]**

Even in absence of Corona crisis, EUROCONTROL in Performance Review Reports[1] [12] estimated that in 2019 twenty two percent of flights will be regulated (and most likely delayed). Among which 5 percent are airport driven and 17% are en-route driven ATFM delays. This leads to a total estimated 9.29 billion-euro cost for delays in 2019.

ETFMS in European ATM saves almost 80% of flights from delay. Conversely, the costs of delay and its growing demand pattern warns a systematic challenge that 2018 traffic data exposed (compare 2017 and 2018 delayed flights in Figure 1-8). Knowing that EFTMS has been improved for years of its service through known performance areas (e.g. punctuality, capacity and safety), the unsatisfactory results of accommodating flights in 2018 were off the charts.

In this thesis, system resiliency is explored with a distinction to consider resilience as a performance topic rather than a safety aspect. This proposed approach requires better justification of the scope. So far, data and reports from both US and European ATM are reviewed. The next step is to clarify which region should be the focus of the study and why.

1.2. ATM procedures in Europe and US

Europe and US are the two busiest airspaces; yet their ATM system evolved differently. Table1-1, shows that while there are similarities in e.g. area and number of airports, European ATM is facing more challenges compared to US, especially in terms of congestion [9].

Table 1-1 Comparison of European and American ATM dimension, [10]

Factor	Europe	US	Comments
Area (million Km^2)	11.5	10.4	
Service Provider	37	1	
En-route facilities	62	20	
Airports with ATC services	406	517	
Highly Congested airports (IATA Level 3)	~ 100	1	US: JFK
Congested airports (IATA Level 2)	~ 70	6	US: EWR, LAX, MCO, ORD, SEA, & SFO
Average daily flights	28 475	41 874	
Share of general aviation (IFR)	3.5%	19%	

Figure 1-9 [9] shows major airports in both US (Continental United States - CONUS) and Europe (European Civil Aviation Conference – ECAC). CONUS is on average accommodating more flights while having significantly fewer congested airports that are scattered across the country in contrast to Europe that hosts a central cluster of congested airports. Conversely, both regions are using relatively similar approaches in ATM. In strategic phase, for instance ECAC benefits from considering following aspects:

- Flexibility: implementing Free Route Airspace (FRA) to allow airlines plan their routes directly, without adhering to published route network.

[1] PRR reports are post operational annual reports (e.g. PRR 2019 was published on 18. June 2020).

- Airspace structure: regular RAD (Route Availability Document) updates that include related references such as policies, procedures and route network.
- Operational planning: offering various scenarios of measures to combine airspace organization, route flow restrictions, sector configuration plan, capacity plan, rerouting plan and/or regulation plan.
- Event Management: defining temporary plans for south-west and north-east axis flows, the ski season traffic flows as well as major sport events such as Olympic games and military events.

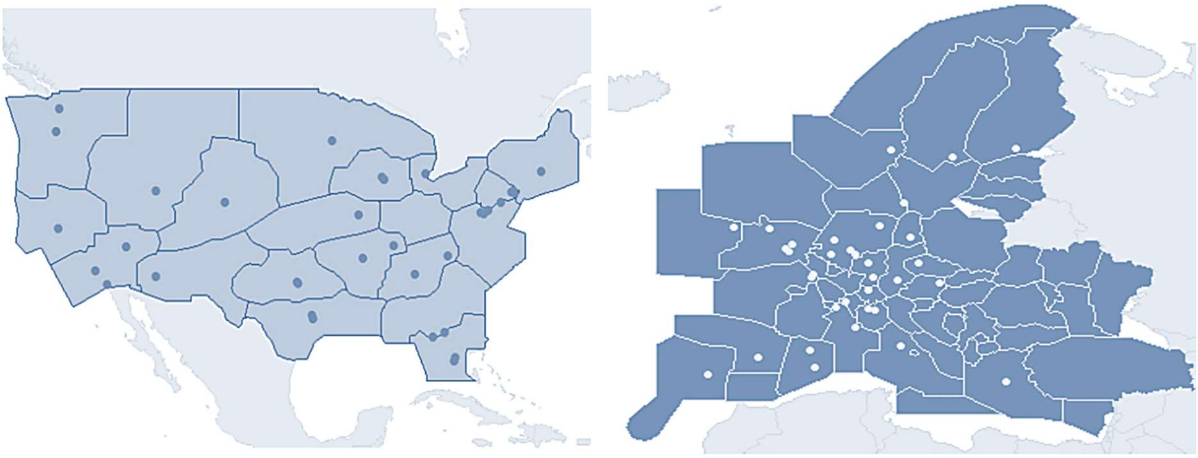


Figure 1-9 FAA/ATO (CONUS) and EUROCONTROL covered area (ECAC), *[9]*

US on the other hand applies another set of procedures and solutions in strategic planning that some can be listed as:

- North American Route Program (NRP) is agreed between US and Canada for upper airspace (flight level 290 and above) is similar to FRA and allows flights to choose flexible routes in the cruise phase (from 200 nautical mile after departure until 200 nautical mile distance to arrival airport).
- Pre-defined routes are validated and coordinated solutions such as Severe Weather Avoidance Plans (SWAPs) and Coded-Departure Routes (CDRs).
- Altitude segregation is the code name for deconflicting traffic flows by capping and tunneling. Capping means assigning a lower than requested flight level and tunneling is advising a flight to descend prior to the normal descent point.

Despite similarities in strategic phase, tactical phase is managed by procedures that are more customized to regional patterns. There are characteristics that urge different systematic perspectives to study ATM in each region. These drivers are for instance, meteorological patterns, passenger demand, route network, airline business models and number of service providers [9].

Next rather than discussing mentioned drivers, applied managing processes for tactical DCB are addressed based on comparison reports such as [9]. In US, the Air Traffic Control System Command Center (ATCSCC) manages the flow of air traffic and minimize delays while in Europe most of such procedures are administered by Network Manager Operation Center (NMOC). On a local level such measures are implemented by either Combined Control Centers (CCF) in US or ACCs in Europe. Examples of major procedures are categorized to airport and en-route constraints, minimal adaptations, flow management and weather prediction to provide an overview that is required for scoping this exploratory thesis and definition the research question.

1.2.1 Airport constraints

- **US, CONUS:** inbound traffic to airports are managed either by Arrival Spacing (ASP), Ground Delay Program (GDP) or Ground Stop (GS). GDPs are mostly triggered by sustained airport capacity loss (due to e.g. severe weather as in Severe Weather Avoidance Plan - SWAP). Compared to GDPs, GS are not supposed to exceed more than 30 minutes.

- **EU**, ECAC: ATFM regulations manage airport traffic flows. Airport ATFM regulations can be applied to a single aerodrome (AD) or to a set of aerodromes (AZ) as Reference Location (RL). In most cases only arrivals are restricted. Airport ATFM regulations with a non-zero rate (flight entering rate) are similar to a GDP and those with a zero rate are same as GS (closed RL). In some cases, an airport ATFM regulation starts off with a zero rate, that eventually increases to accept a limited amount of traffic (low-rate). This is the equivalent of a combined GS and GDP.

1.2.2 En-route constraints

- US, CONUS:
    - Departure stop, similar to a GS that is being for instance, assigned to an airway, fix, departure gate or sector;
    - Airspace Flow Program (AFP), is a type of Traffic Management Initiative (TMI) that is defined with similar parameters as of a GDP but AFP is applied to a volume of airspace (referred to as Flow Constrained Areas-FCAs);
    - Flow Evaluation Areas (FEA) are 3-dimentional airspaces defined for a period of time, with a filter for flights to evaluate the demand in monitored airspace (in ECAC this is referred as a Hotspot). Note that the airspace is not restricted but closely monitored. Both FEAs and FCAs

provide reroutes to flights and are visible through e.g. Traffic Situation Display (TSD) or collaborative constraint situation display (CCSD).

- Required Reroutes (RR) is another TMI coupled with a delay program and they are issued by departure, arrival or FCA entry time;

- Collaborative Trajectory Options Program (CTOP), is a relatively new procedure for DCB that automatically assigns delay and/or reroutes flights to avoid FCAs. CTOP considers the preferences of airlines (Trajectory Option Set, TOS) by taking a set of alternative routes (AR) from airlines;

- Integrated Collaborative Rerouting (ICR) is based on the FCAs and allows the airlines to revise their trajectory preferences according to the FCA and finally if the imbalance is not resolved, the traffic managers will decide on the next action that can be recommended routes, RRs and AFPs.

- EU, ECAC:

    - En-route ATFM regulations that can be applied on a specific airspace volume (AS) or special point (SP) as the Reference Location (RL). Such a regulation can limit all or a set of traffic crossing the RL (Referred to as a TV). En-route regulations can be similar to AFPs if they impose delay or in case of rerouting it can be in form of:

        - Flight level capping (imposing vertical limitations),
        - Required reroutes (RRs), or
        - Alternative rerouting (AR) that opens a low rate through airspace which normally is not accessible to the traffic flow.
    - Flight Efficiency Initiative (FEI) enables airlines to revise their flight plans in search of more efficient trajectory. The cost of each trajectory can be evaluated based on a criterion (cost): flight time, fuel, cost of delay.

    - Airspace Users Fleet Priorities and Preferences Processes (UDPP), similar to CTOP in US, considers the priorities and preferences of airlines in both en-route and airport collaborative processes.

1.2.3 Minimal adaptation

- US, CONUS: there is also the possibility of exchanging (subbing) the departure time slots. The substitution process provides a way for airspace users to manage their flights during a GDP, GS or AFP. Airlines can, for example, swap slots between a high priority flight and a less important flight, reducing the delay on one at the cost of increasing the delay for another flight of their own.

- EU, ECAC: the same possibility is referred as slot swapping that also allows slot extension. Airlines only request swaps concerning flights for which they operate or where there is a formal agreement between two different airlines for swaping.

The differences between mentioned aspects of ATM in each region is also a result of different traffic patterns (Figure 1-10). Not only the seasonal patterns are different both in their shape and density but traffic flows and axes are different too. Figure 1-10 depicts the annual flight hours per square kilometers to identify the traffic density [9]. Most of the congested airports are located in central Europe while in US, they are mainly located at the coast lines and US has less challenges to deal with weekly traffic patterns. Europe also has crossing traffic axes in contrast to converging flows in US (Figure 1-11).

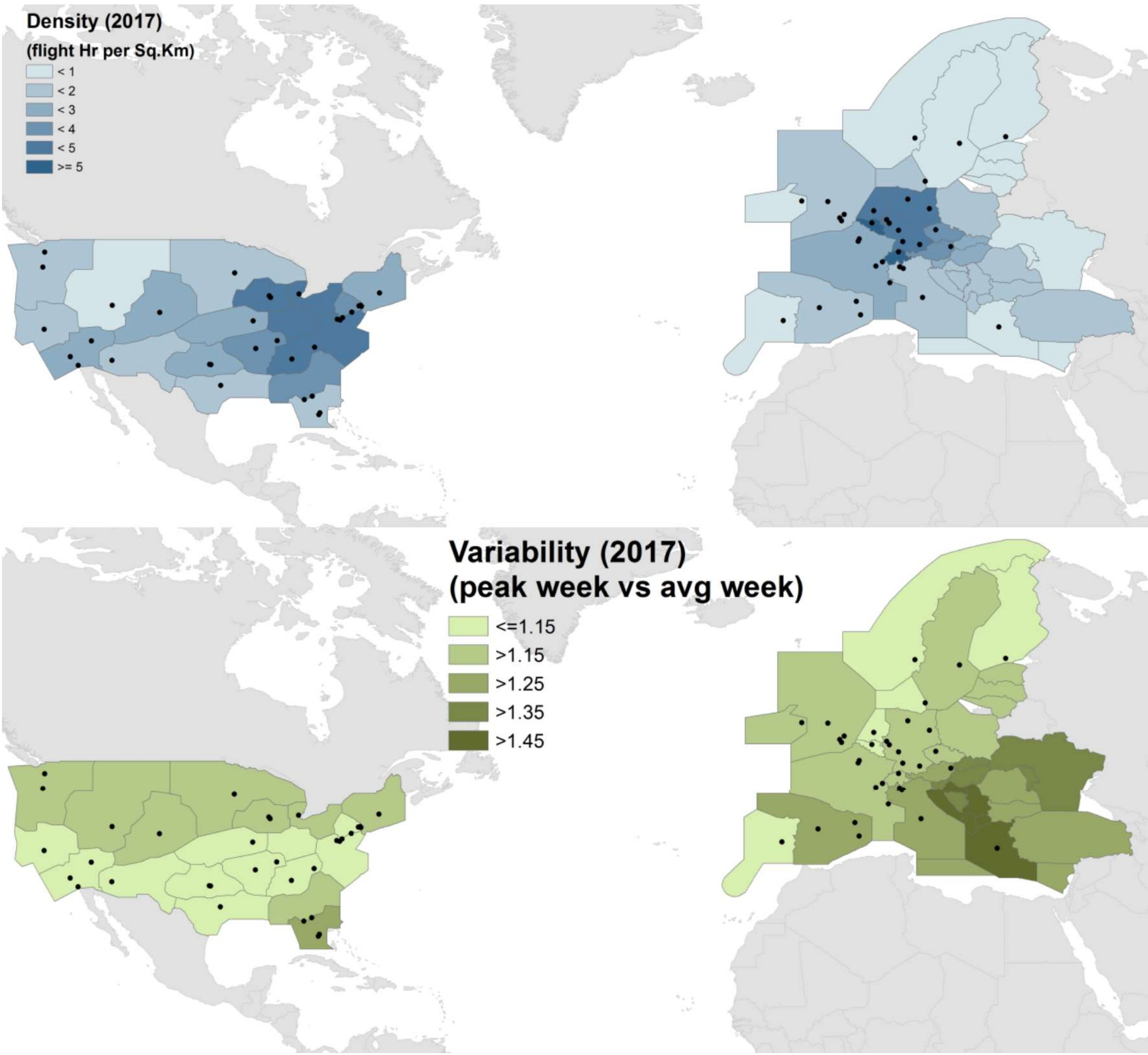


Figure 1-10 Comparison of US and Europe traffic density and weekly variability, *[9]*

Annually, ICAO reports the traffic flow chart for all movements across the world and Figure 1-11 cuts the European and US section of the global map [13]. The cross-traffic flows over Europe and pointed converging flows to north-eastern coast of United States are evident.

This crossing pattern of traffic flow in Europe and high number of service providers require EUROCONTROL role as the Network Manager. NMOC is the *'former Central Flow Management Unit -CFMU operations room that manages one single flow management system over Europe together with its partners, the airlines, airport authorities and air navigation service providers'* [7]. The role of NMOC is more transparent in ATFM and delay management that is discussed next.

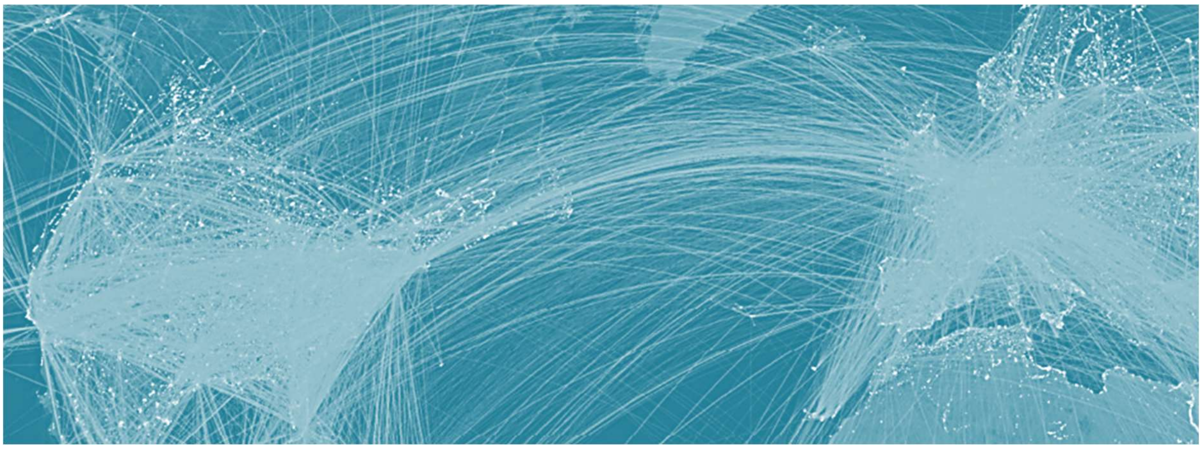


Figure 1-11 ICAO traffic flow map of 2018, showing cross shape traffic axis in Europe compared to rather converging pattern in United States, *[13]*

1.2.4 Flow management: ATFCM regulation vs. TMI (Europe vs. US)

NMOC mission is to optimize traffic flows through DCB procedures. As provided by Figure 1-12, the DCB framework in different planning phases [14] benefits from constant updates on both demand and capacity estimations and the uncertainty of factors such as weather and staffing are persistence and eventually will affect the planned operation at tactical phase.

Some levels of uncertainty are intentionally built in to allow the required flexibility of operations. A good example is the flight plan submission allowance for airlines up to 3 hours before departure (or more specifically 3 hours to Estimated Off-Bock Time - EOBT).

The cost of such relaxations is more traffic complexity at tactical phase. For instance, the implementation of FRA in European airspace improves the fuel efficiently but consequently leads to unpredicted excessive demand for ATC capacity. In other words, since FRA allows pilots to fly direct routes which were not originally filed in flight plans; in some cases, the pilot request results in earlier entry times over adjacent sectors that tags them as *intruders* since they are unexpected for local ATC units.

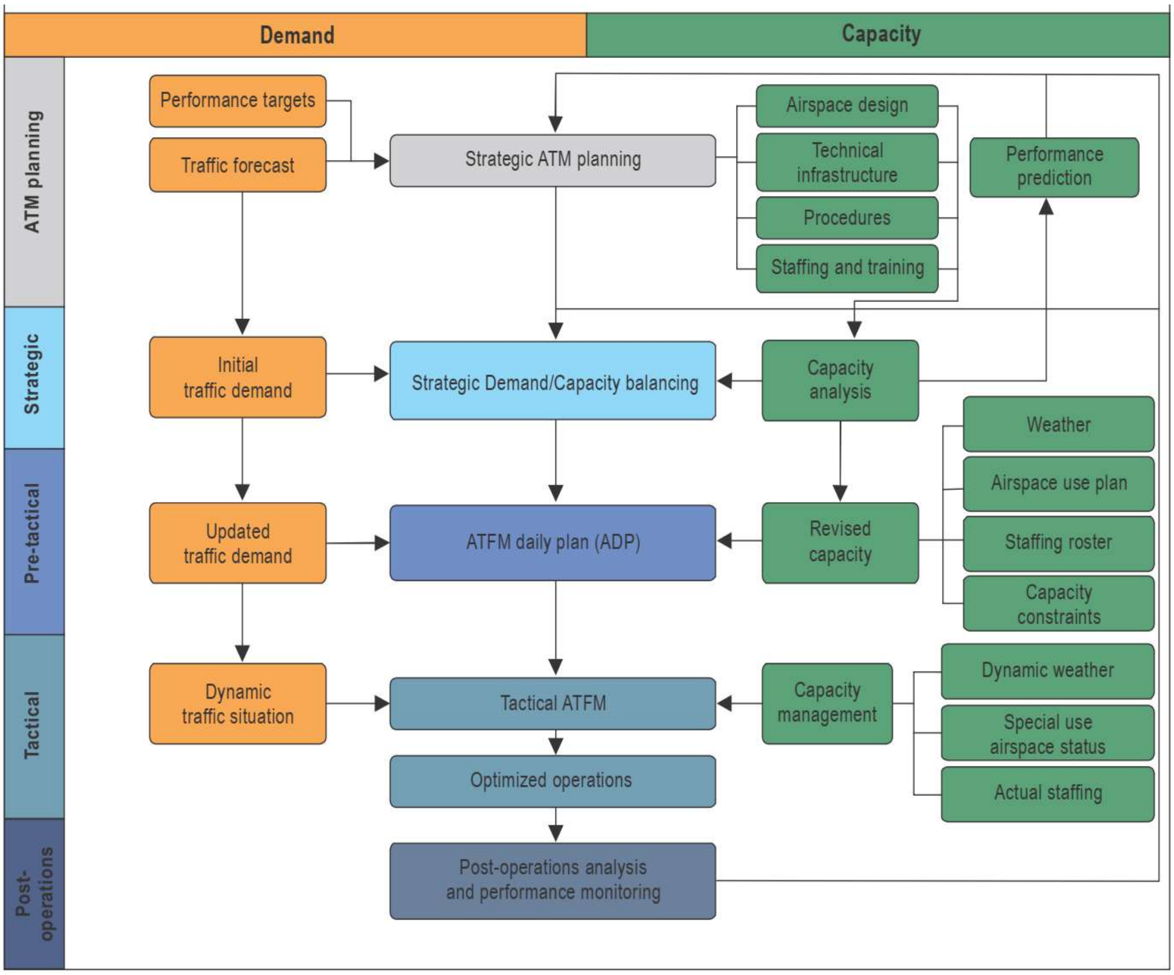


Figure 1-12 ATM planning and ATFM Phases, *[14]*

Figure 1-13 maps the dynamic recursive cause and effect chain in tactical phase that is triggered by demand-capacity imbalances (e. g. intruders). Both overloads and unused capacity costs to airlines and ANSPs are managed through ATFCM measures in Europe and TMIs in United States.

Such ATFCM or TMI measures are applied and requested by air traffic control centers in US and in Europe by FMPs and applied after being authorized from NMOC. They can be separated depending on the impact, whether it is affecting the airborne flights or penalizes them before departure time. While similar to assigning a CTOT in Europe, in US an updated Estimated Departure Clearance Time (EDCT) delays a flight; each region has a different tolerance time window for assigning a delay [9]: the EDCT window is ±5 minutes and the CTOT Slot Tolerance Window (STW) is -5 to +10 minutes.

However, not every flight is restricted with a CTOT or EDCT. As an example, in Europe for flights without an ATFM slot, the Departure Tolerance Window (DTW) for Actual Take-Off Time (ATOT) is normally 30 minutes, from -15 to +15 minutes of Estimated Take Off Time (ETOT) that can be extended in adverse conditions to 45 minutes (-15/+30min). In US at New York area for example, similar controlling is applied through Departure Spacing Program (DSP) that is planned to be replaced by 2026.

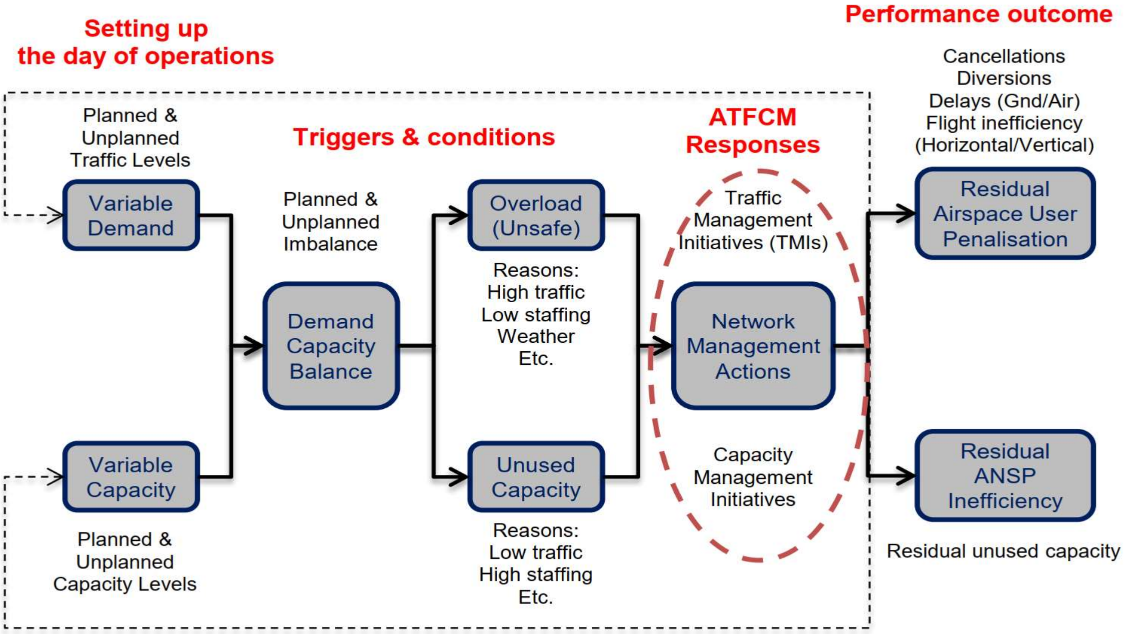


Figure 1-13 DCB triggers and outcomes in tactical phase of ATFM, *[9]*

As reviewed there are various devised tactical procedures for DCB issues that each may be caused by operational uncertainties. More specifically, weather induced uncertainties has a different category both in terms of its nature and impact. Next section is dedicated to review this aspect since despite the improved weather prediction models, incidents such as closure of Istanbul airport due to adverse conditions on 24. Jan.2022 have and will significantly interrupt the flight operations across Europe.

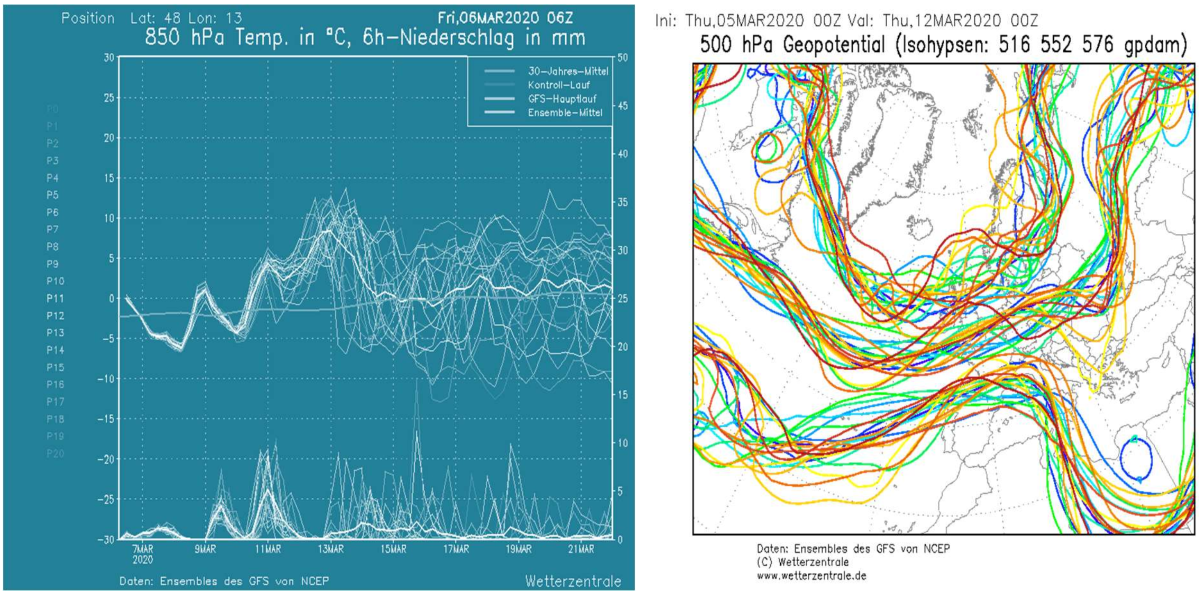
1.2.5 Weather predictions

The limit at which day-to-day weather can be predicted is one of the drivers of demand-capacity imbalance. Experts of meteorology and atmospheric sciences argue this limit on average, is about two weeks and large scale high-impact events such as hurricane tracks can be predicted with an accuracy of 150 km up to 4 days in advance [15]. Such numerical weather predictions (NWP) are an asset in pre-tactical rather than tactical phase of ATFM. For instance, temporary bad weather situations are one of reasons that makes pilots request alternative cruise flight levels, leading to unexpected demand and increased complexity of flight trajectory at the day of operations.

But there are also other weather-related issues that can neither be predicted by NWPs nor be detected and considered by delay assignment algorithms such as CASA. Factors such as stability of airport facilities and management experience makes it a challenge to estimate the duration of an airport closure (e.g. in case of a heavy snowfall). At Istanbul incident (on Jan 24th, 2022), the wrong initial assessment of airport suspension period, caused a lot of delays. Although the airport was opened in 2019, a cargo terminal roof was collapsed because of heavy snow, runways were blocked and airports ground services couldn't be supported by the local authorities since the access roads to the airport was also blocked.

Nevertheless, even the precision of NWPs is highly dependent on initial conditions. Sensitive dependence on initial conditions or Butterfly effect is pointed out by Lorenz [16]. His investigations into predictability of the atmosphere led to introduction of chaos theory, strange attractors and *chaotic solutions* that usually appear in nonlinear systems as of weather. Such fundamental difficulties limit the precision of weather predictions and despite constant improvements even same models provide different forecasts at different runs.

As an example, Figure 1-14a shows the results of ensemble Global Forecast System (GFS) at different runs (P1 to P20) for temperature and the precipitation level as target values. Despite of fixed geographical position and reference date, the divergent pattern for different runs is evident for both target values. This uncertainty does not only concern the values of predicted parameter but also it has a significant deviation in geographical span as shown in so called Spaghetti figures (figure 1-14b).



a. Predicted Temperature (at 850 hPa Pressure) and precipitation level for 6th March as the reference date

b. Spaghetti graph of prediction runs that shows the geographical divergence for different runs of the model

Figure 1-14 Divergence pattern of different runs (P1-P20 shown by different colors) for ensemble GFS model from NCEP, *[17]*

Such uncertainties in weather forecasts in general are one of the reasons for flexibility in the tactical phase of ATFM. Weather conditions are affecting both the airborne and ground capacities. For instance, airborne flights may alter the flight trajectories to avoid weather conditions. Such modified trajectories increase the workload for controllers and in some cases creates secondary problems in other ATC units (adjacent sectors) that provide surveillance to the same flight.

***

As discussed from different perspectives, there are many devised planning steps for a smooth traffic flow. However, the rise of demand still challenges the saturated capacity and no procedure is in place to resolve or monitor network issues other than extreme cases of a pandemic for example. In Europe the risk of ignoring network awareness is much more tangible since it is integrated from different states with various evaluation models in terms of demand, capacity or weather.

The risks and simultaneously the potential benefits of current traffic patterns can be realized by monitoring locally optimal ATFM measures and capture their impact on the network (e.g. European ATM Network-EATMN). Such functionality is mainly projected on top level authorities i.e. FAA/ATO (Air Traffic Organization is the operational arm of the FAA) and NMOC. These operational units are the most relevant stakeholders to address and study systematic improvements. Specially in Europe the exposure to secondary problems is higher because 37 service providers (Table 1-1) request and apply numerous local optimal solutions at tactical phase in absence of a network situational awareness. Throughout previous five sub-sections, these solutions were investigated from different aspects and the conclusions are offered below:

- Why such local solutions need to be reviewed? Reported rise in demand figures and its discussed uncertainties remind that the current system is reaching its saturation level and it is time to reach out for revolutionary ideas for air traffic management. But first a solid understanding of active solutions is needed to spot potential directions of improvement.
- What is the cost of a saturated network? The rise of delay despite collaborated processes in European (ECAC) and American (CONUS) sky, is a sign that delay management approaches such as CASA algorithm are not designed to be consistent with new concepts (e.g. FRA) and need to be revised. This is observed specially in summer 2018 that despite timely raised alarms about excessive demand, delay figures significantly degraded beyond control;
- Why not invest on improving local solutions with current methods? Complexity of dynamic traffic management is more significant in tactical phase. Both in ECAC and CONUS there are controlled (e.g. flight plan modification) and uncontrolled (e.g. weather) uncertainties that cannot be fully realized by analytical and numerical methods.

1.3. Problem Definition

In spite of exploratory nature of the thesis, the holistic problem of providing systematic approach against European airspace saturation needs to be formulated more specifically. Therefore, this section begins with describing ATM as a system. Consequently, life cycle stages are considered to locate the phase at which the saturation problem needs to be addressed. Saturation of a system is directly related to its adaptation capacity. System resiliency is the concept that addresses the saturation in this regard and is interrelated to situational awareness. Furthermore, the perspective from European research program is adopted to identify and map system's basic data flows. Such an understanding leads to a situational awareness based on a data driven approach that avoids ATM complexity.

1.3.1 ATM as a system

In system engineering, ISO 15288 standard [18] man-made systems are designed to provide stakeholders, services/products within defined environments. Therefore, Air Traffic Management (ATM) is a standard system since it is the aggregation of airborne and ground-based services (air traffic services-ATS, airspace management-ASM and air traffic flow management-ATFM) [19] that are provided to four main categories of stakeholders (in Europe):

- *NMOC: Network Manager Operations Center* operated by EUROCONTROL,
- *FMP: Flow Management Position* representing the ANSP (air navigation service provider) stakeholders that can be also designated by ACCs,
- *APOC: Airport Operations Center* representing the airport perspective since each center is the core organizational unit responsible for airside operations, and
- *FOC: Flight Operation Center* (also known as *Airline Operation Center-AOC*) represents the airline interests and hosts required functions for flight operations.

These categories are considered as a system-of-interest with defined roles and authorities in ATM and all four are linked by different flows of information. In systems-of-interest, humans play different roles in each group. For instance, FMP coordinators interact with Collaborative Decision Making (CDM) processes in ATM.

Nevertheless, ATM as a system is getting closer to saturation level and retirement stage. In absence of radical changes, ever growing capacity costs (e.g. improving navigation infrastructures and building new airports) will change the opportunity of the rising demand into an obstacle. Generally, a system has six stages in its life cycle (Figure 1-15) and retirement is the last [20]. A resilient ATM system should constantly be engaged in support and concept stages. The system provides sustained capabilities at support stage and any needs/requirement for a new system-of-interest or modifications are recognized at concept stage. Establishment of EUROCONTROL in 1960 [21] serve as a good example of such changes at concept stage.

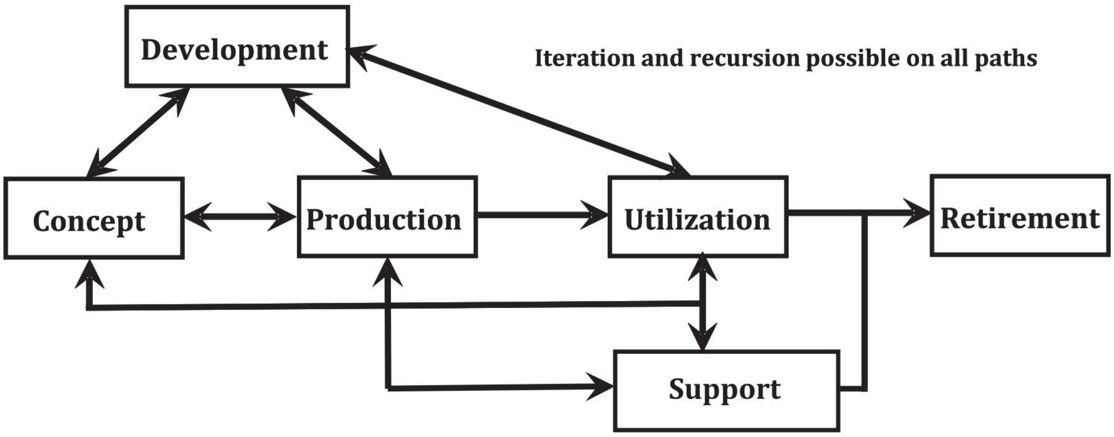


Figure 1-15 Stages of a system's life cycle, *[20]*

EUROCONTROL started to acquire data, assess potentials and set strategies to support European aviation (stage: development). More specifically to assist traffic management, Network Manager (NM) role was established (stage: concept) that later started to operate NMOC (stage: production). This center optimizes traffic flows by monitoring demand and capacity. But while the numbers show a steady trend in traffic and delay figures (stage: utilization), still the EATMN suffers from saturation specially in summer season (March-October). This draws the attention to support stage.

According to ISO standard [20], "support stage includes processes related to providing services that support utilization of the system-of -interest. This stage also includes processes to use and monitor the support system itself, including the identification, classification, and reporting of anomalies, deficiencies and failures of the support system and services". The given diagram in Figure 1-15 is showing the connection of support stage that is directional to retirement stage but interacting with other stages except development. Because development stage considers strategic needs while support stage deals with inspected issues at hand.

Saturation can result from unproportionate assessment of system load at concept stage. In EATMN, firstly it was unrealistic to draw prediction figures at concept stage for upcoming 20 years knowing that it remains in service for much longer than 20 years. ATM should be considered as a rolling system, which accommodates more and more flights every year and provides ranges of solutions (i.e. vertical separation minima) as time goes by. Secondly, predicting the load of the system means gathering tons of data that is both an acquisition challenge and an analysis issue since there are different types of data with different granularities.

Such challenges in Europe are addressed by international cooperative research programs such as SESAR (Single European Sky ATM Research). In next section the contribution of SESAR in EATMN is discussed to realize any conceptual approach against saturation.

1.3.2 EATMN and SESAR

SESAR was launched in 2004 to not only define the challenges but also to develop and deploy solutions to support EATMN performance. In 2007, SESAR Joint Undertaking (SJU) was established to be responsible for modernization of EATMN. SJU started the first program of research in 2008. SESAR-I continued until 2016 with 400 projects that took 20 million hours to ensure the quality of deliverables to fit the operational needs [22]. SESAR-I successfully delivered numerous industrial prototypes, operational and technical solutions but more importantly a wide range of new questions, potentials and challenges were inspected.

SESAR 2020 was launched as the follower until 2024 with a budget of 1.6 billion Euros. European Union and EUROCONTROL and other 19 members work together in a setting that gathers regulatory bodies, airspace users, airports, ANSPs, manufacturing industry and scientific community such as German Aerospace Center (DLR). Its target was to deliver a 'modular and automated' ATM based on digital and virtual technologies in 4 key areas:

- Airport operations,
- Network operations,
- Air traffic services, and
- Technology enablers.

These key areas are planned as pipelines that transfer ideas to industrial solutions, in three strands of Exploratory Research (ER), Industrial Research (IR), or validation & very large-scale demonstrations. To the benefit of this thesis, the author as a concept expert have joined project PJ09: DCB or 'advanced demand capacity balancing'. The project supported the European ATM master plan [23] that aims at providing an interoperable concept of European ATM in which operations are built around a continuous sharing of data between actors, i.e. ANSPs, airspace users (AUs), airports and NMOC. The focus towards performance ambitions of the master plan to further develop DCB processes is addressed by improving collaborative processes and mutual situational awareness. In essence one of work packages of PJ09 (network performance) and this thesis share ideas such as focus on data flows and situational awareness. Global thinking was focused through situational awareness at regional levels.

PJ09 focus was on performance driven DCB in a collaborative environment among actors. Actors (systems-of-interest) are communicating through dataflows among each other. While each dataflow serves a specific purpose, none is dedicated for a mutual situational awareness at EATMN level. Figure 1-16, summarizes current main data flows in tactical phase. Among all dataflows, ATFCM regulations represent results of collaborative decision-making processes among all actors. Regulation data is constantly updated and published for all actors in the course of tactical phase.

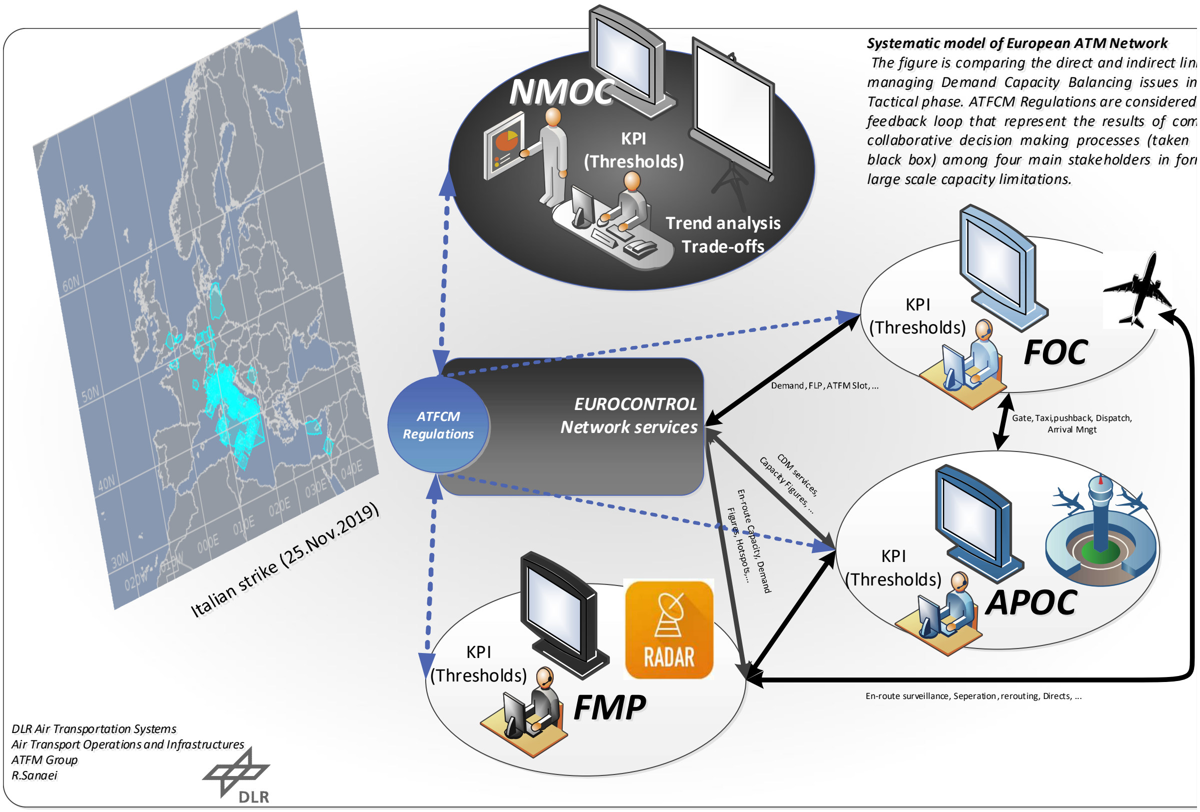


Figure 1-16 Proposed European ATM network as a system. Such a model takes ATFCM regulations as feedback loop that connects all stakeholders. The model avoids complexity of numerous direct data flows between stakeholders.

ATFCM Regulations or measures are introduced in section 1.1.2 (growing delay) as a method of matching traffic demand to available capacity by limiting the number of flights planned to enter an airspace or aerodrome by ATFM delays [7].

PJ09 provided the chance to explore network performance improvements. Having EUROCONTROL as a partner in this project assured that a better network wide situational awareness is operationally motivated since NMOC experts in practice realized the need for a conceptual improvement against saturation.

1.3.3 ATM resiliency

The constant update of regulations at tactical phase is a result of emergent behavior of European ATM system. Emergent behavior is a complementary expression to resultant behavior. Resultant behaviors are those dynamics that happen as a consequence of known causal links. In contrast, emergent behavior is a response to unpredicted issues in large scale systems usually with sub-systems as of European ATM.

Emergent disruptions have a direct relationship to complexity. Pariès [24] describes the emergence relative to the size and complexity of a system. EATMN is considered as both a largescale and complex system with inevitable disruptions. These disruptions can be decomposed to components of the system for analytics but the coping mechanism is bounded to following questions:

- How to realize an emergent disruption?
- Which level of decomposition is needed to revive the system?

The first question is a bottom-up and the second one is a top-down challenge. Each system has an identical answer based on available dataflows and provided level of control. In EATMN, NMOC has the authority to control network issues and in tactical phase this can be mostly realized through ATFCM regulations.

While both NMOC roles and Regulation data are active in daily operations, emergent disruptions are not covered by resilience. Prior to this study, resilience was solely a safety (rather than performance) topic. Instead robustness of operations was at focus, e.g. by providing the mentioned flexibilities in flight plan submission. However, robustness is addressing the predictable disruptions (resultant) in the planning phase and resilience is more focused on system functionality and emergent disruptions (mostly in tactical phase). In fact, a resilient system accepts the inevitable challenges of its dynamic states and adapts itself by changing operational processes to maintain its core functionality.

In EATMN, the four mentioned systems-of-interest NMOC, ANSPs, airports and Airspace Users (AUs) are interacting with each other through eight systems [25]:

1. Systems and procedures for airspace management.
2. Systems and procedures for air traffic flow management.
3. Systems and procedures for air traffic services, in particular flight data processing (FDP), surveillance, data processing and human-machine interface systems.
4. Communications systems and procedures for ground-to-ground, air-to-ground and air-to-air communications.
5. Navigation systems and procedures.
6. Surveillance systems and procedures.
7. Systems and procedures for aeronautical information services (AIS).
8. Systems and procedures for the use of meteorological information.

Extent of these systems defines, EATMN (here after also referred as network) high level of complexity. The systems and procedures for ATFM is the most relevant in tactical phase (Figure 1-16 is a simplified role base model of this subsystem). This part of network is the most interactive systems in tactical phase and therefore taken as the frontier for detection and resolving emergent disruptions.

In conclusion, in order to improve the tactical operations, as an exploratory research this thesis addresses the gap in coping mechanism against emergent disruptions by conceptualization of network resiliency. The following statement reflects the aim of the study:

> *Considering the NMOC role and regulations data, the study firstly tries to model the network as a resilient system(conceptually), then it proposes a mechanism to detect network emergent disruptions and finally investigates the required level of decomposition in reviving the resilient network. In other words, the following research objectives are considered:*
>
> 1. *Demonstrate the idea of EATMN resiliency by a conceptual model;*
> 2. *Since the initial requirement of monitoring resilience is the situational awareness, propose a mechanism to detect the network state that serves for both current and reference states;*
> 3. *In order to revive the network from emergent disruptions, investigate the required level of decomposition for corrective measures (so viel wie nötig, so wenig wie möglich: as much as necessary, as less as possible).*

The data driven methodology of the thesis explores an alternative approach based on statistical (Objective 2) and learning methods (Objective 3) in comparison to current simulation-based approach in ATFM. As of today, DCB issues are declared as hotspots by demand prediction models and tactical simulations. Hotspots are locally identified problems across the network. By conceptualizing the EATMN as a resilient system, this work demonstrates the benefit of inspecting 'network' disruptions rather than 'local' hotspots.

1.4. Study outline

The following chapters define the road map to realize the objective of the thesis: Chapter two goes through resilience and its advantages over robustness, and is concluded by system state and assumptions of the thesis. Third chapter, methodology, starts with objective 1 to conceptually model the EATMN state. Next, the selection of regulations as a feedback loop is described in realizing the second objective of the work. More specifically, a twostep statistical analysis to detects tactical EATMN state is provided at first and then with regard to data driven approach, two intermediate sub-problems on learning algorithms are devised to associate second objective to the third:

Sub-problem I: a feasibility study on different learning techniques to estimate network ATFM daily parameters from regulation data.

Sub-problem II: base on the result of sub-problem I, devise a solid supervised learning algorithm that extracts the spatiotemporal dimension of regulations (Deep Convolutional Neural Network- DeepCNN).

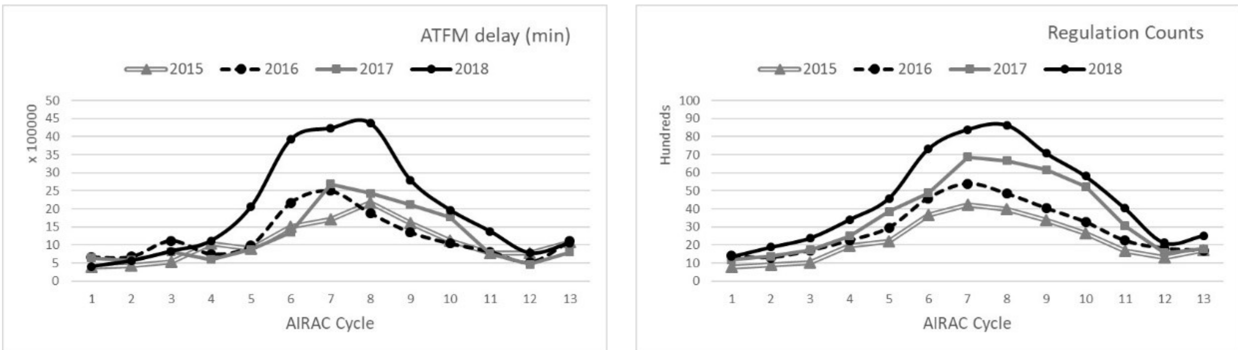
Chapter four (Resilient Path) along with an understanding of Complex Adaptive Systems (CAS) addresses the third objective of the study on achieving the required granularity on predicting corrective actions (i.e. regulations) by providing a Recurrent Neural Network (RNN) architecture that predicts ATFM parameters of each regulation (max granularity). These predicted values are accumulated to deliver daily predictions at a network level so that at any given time the Network Manager (NM) is able to evaluate a list of regulations. Chapter five, provide results from all three objectives which are further discussed in chapter six. The thesis is concluded by providing a brief overview of COVID pandemic and stating possibilities for future works.

2. Literature Review

Limited studies on resilience as a performance topic rather than a safety aspect is a reason that this thesis is categorized as an exploratory research. This chapter reviews resilience and system state in general and then narrows down to address EATMN.

In fact, EATMN as a complex system with eight subsystems is highly exposed to emergent disruptions. As given in Figure 2-1, despite expected systematic disruptions, there is also a seasonal pattern which is also reflected in delay figures in DCB issues. Generally, the rise of delay is either a ramification of major disruptive events (e.g. strikes, unpredicted demand) or a wave of airspace capacity limitations (e.g. prolonged weather issues). Studies on controlling such disruptions tend to propose solutions for strategic or pre-tactical phases. For instance, campanelli et al. [26] compared US and European air traffic networks by analyzing the propagation of delays due to disturbances. Through agent-based models, their work on different delay management systems (flight sequencing) conclude that a priority system in ATFM is more efficient in avoiding congestion compared to a first-come first-served managing protocol for flights. Similar studies in EATMN are less likely to address resilience in tactical phase through a data driven approach.

But prior to discuss EATMN resiliency in tactical phase, it is crucial to locate resilient performance against safety resilience and then differentiate robustness from resilience (structural versus tactical solutions).



(**a**) (**b**)

Figure 2-1 Statistical survey of regulations per AIRAC cycle to compare (**a**) total ATFM delay and (**b**) regulation count. Apart from seasonal patterns, the figure shows that an increase in number of regulations does not necessarily imply an increase in the ATFM delay (see annex B.1. for AIRAC cycles; Data source: EUROCONTROL, NMIR reports).

2.1. Resilience

The term *resilience* is described by a number of definitions, concepts and approaches in different disciplines [27]. It has a strong bond with other concepts such as *robustness*, *flexibility* and *agility.* The extended scope of resilience has evolved and became more mature through specific stages. Hoffman [28] introduced resilience as '*a high percentage of recovery after, but not necessarily immediately after, a deformation*'. His study also provided a definition for resilience as a capability of a substance to return to its original state at some time after removal of the deforming stress. Reviving time is a key aspect of both definitions, hence if the damage or deformation is so extreme and abrupt that the object/system seizes to exist/operate, resilience is not a topic anymore.

Another definition is introduced by Holling [29] that compared resiliency and stability to develop the *concept of resilience*. It is coupled with the definition of disruption and based on deformation time, the concept determines multiple *states* for disrupted systems [30]. A system that faces (internal or external) forces is considered as resilient if its core functionalities are not lost in a disrupted state. This perspective suits the dynamic nature of the EATMN with its required operational flexibility for efficient services.

Moreover, resilience engineering (RE) was introduced by Hollnagel et al. [31] and charted studies from different disciplines. RE is a focused paradigm on how safety managers can be empowered to handle complexity under pressure. Its approach relies on safety as a dynamic process of systems. Resilience engineering invests on strength of a system to compensate effects of a disruption [32]. Such a system is considered to have control over its *performance variability*. In other words, RE respects the performance variability by assessing both sets of different system outputs: failures (extracting disruptions) and non-failures (detecting system strengths).

Resilience engineering contributes to safety by improving performance in contrast to lowering risks through applying constraints. In fact, traditional methods of accident analysis and risk assessments are being compromised by more complex technologies in dynamic systems. Therefore, current methods combine the technical aspects and human factors to improve safety. In this regard, RE is using the principle of *resonance* to explain how the variability of normal performance can (in dynamic ways) lead to disproportionate disruptions. EUROCONTROL describes the resonance principle as:

> *'Resonance: A principle that explains how disproportionate large consequences can arise from seemingly small variations in performance and conditions [33].'*

For instance, RE is providing the base for methods such as Functional Resonance Assessment Method (FRAM)[1] that has four principles:

- **Success and failure equivalency:** none of the processes in a system is meant to produce failures. In other words, failures and successful outputs are generated from the same system.
- **Approximate adjustments:** every planned activity needs some levels of adjustment since resources, time and in general the actual situation is not the same as assumed conditions in planning phase.
- **Emergence:** performance variability may build up on unexpected results, that are disproportionally large and disturbing the whole system. An outcome is emergent if it neither can be attributed to nor explained by (mal)functions of the system.
- **Functional resonance:** that is an alternative to linear causality. It represents the detectable signal that emerges from the unintended combination of the variability of many signals. This explains how the variability of a number of functions can reinforce each other, leading to excessive disturbance in downstream functions. The consequences may spread through the system by means of tight couplings rather than easily identifiable cause-effect links.

Among these principles, functional resonance is closer to systems like EATMN with numerous local solutions, i.e. ATFM measures (capacity regulations) that might trigger secondary problems.

Moreover, a study by Francis and Bekera [27] provided categories of resilience definitions in different settings. They concluded that "resilience is a conceptual framework composed of multiple dimensions. Absorptive, adaptive, and restorative capacities are at the center of what a system needs to do and how it needs to respond to perceived or real shocks". Considering EATMN settings, these capacities/levels of resilience (Cook et al. [34]) can be modified as in Table 2-1 to locate challenges at each ATFCM phase [35].

For instance, reliability and robustness of a system should be considered within the strategic phase, since a complex system is able to implement structural solutions through strategic plans. Generally, systems are more vulnerable at the strategic phase because of possible broad consequences. A structural failure can shatter other resilience levels severely (low rate but drastic impact).

[1] To read more about FRAM refer to [155] and for a ATM study consider [156].

Table 2-1 Resilience Levels

Level	Features	ATFCM Phase
Absorptive	Robustness, Reliability e.g. Air Traffic Flow Management (ATFM) Procedural Contingency Plan	strategic
Adaptive	Consideration of adverse impacts, Anticipation of disruption, Recognition of unanticipated events e.g. Reaccommodation of network flows during an ATC strike	pre-tactical
Restorative	Control measures, Conflict handling, Cost estimation e.g. STAM[a] measures	Tactical

a. Short-term ATFCM measures (STAMs) include a set of automated support tools at the network level which detect hotspots and disseminate the information to flow management positions in the ACCs.

Table 2-1 suggests that each type of DCB solutions can be improved according to a corresponding level of resilience. Such a classification helps for an efficient selection of corrective measures in different disrupted situations (or non-nominal states). Every system has iterative processes at different intervals, therefore assuring a level of resilience for each process varies in terms of effort, cost and impact domain.

Most recently, resilience has been defined through the European research project, 'Resilience 2050' as *the capacity to recover [quickly] from difficulties; toughness*. In fact, ATM resilience is defined as the capacity of the aviation system to behave as scheduled in spite of incidences, so that flights arrive on time whenever they encounter a difficulty.

Along with the project Resilience 2050, DLR (German Aerospace Center) has also focused on terminology of resilience in different disciplines. For instance, DLR [36] has addressed following associated terms with resilience:

- **Reference State**: in order to be able to measure the resilience of a system there is a need to identify deviations at first place. Rationally the planned status (also referred as *Nominal situation*) of the system is considered as the Reference state.

- **Current State**: the status of the system which is captured by indicators at a given time. It is compared against reference state to measure either resilience or robustness of the system.

  These states do not imply a static mode of the system but rather a domain in which the system is considered to be functional. For example, a reference state is possible to be defined by either single values of performance indicators, intervals or acceptable range where performance indicators can vary.

- **Disruptions and disturbances:** despite the importance of clarifying disruptions and disturbances, there is no general definition for them due to technical complexity of each system. Nevertheless, a *disruption* can be considered as a state, where the deviation from the plan is sufficiently large to impose a substantial change [37].

Note that the disruption does not have to be always negative but a resilient system is considered to be able to capture opportunities too. In some cases, a disrupted state needs to be realized so that the system can gain more or better outputs.

Similarly, a disruption can be expressed by thresholds. Adopted from available DLR literature [38], disruptions can be defined using the approach introduced in ecology in which a disturbance is defined as the cause (not the state) of stress and perturbation.

- **Stress:** the reactionary state of the system or the consequences of the disruption on the system functionality that can be divided to:
    - *Survival:* if the effect of the disruption is not severe and the system can respond and damp the consequences through modifications.
    - *Lethal:* if the system does/should not respond to the consequences of disruption. In this case there is always a call for largescale modifications and generally a lethal stress is regarded as a Crisis.
- **Perturbation**: this term is referred to the reaction of the system to imposed changes. Based on the impact and severity of disturbance, two scenarios can be defined: either system is partially engaged or every subsystem is affected. Nevertheless, in case of a survival stress, perturbation can be:
    - *Transient:* temporary solutions which are able to revive the system over limited time and absorb the effects of disruption in system, and
    - *Permanent:* the consequences are so severe that the system is forced to set a new reference state through fixed modification solutions. A permanent perturbation pushes the system into a new reference state to cope with a lethal stress.

In summary, despite of provided definitions, classifications and approaches for system resiliency, only a few attempts such as Resilience 2050 project was dedicated to ATM resiliency. The project was executed by six academic and research institution in the absence of industrial partners or European aviation authorities. As expected, the results are more contributing to absorptive and adaptive rather than restorative level (Table 2-1) of ATM resiliency. Consequently, a larger scope of transportation systems should be reviewed regarding measurement of resilience. IEEE published a comprehensive review paper of such studies in 2019 [39] that is addressed next.

2

2.1.1 Resilience measuring methods

As provided in Figure 2-2, studies on transportation systems resiliency has been increased over the past years. Among which the majority of the studies (44%) are related to resilience of road networks. Other domains of research include freight transports, railway, maritime networks, air traffic networks and multimodal transportations.

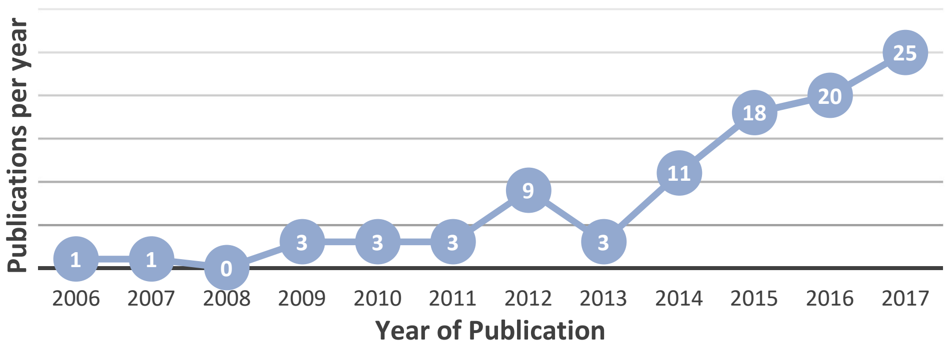


Figure 2-2 Publications on resilient transportation systems per year, [39]

Share of studies on air transportation is reported [39] to be only 8% despite the availability of multiple standard data types. But even with the rise in published papers from 2015 onwards, it seems that other modes of transport hosted more studies compared to air transportation due to its complexity and extent.

Nevertheless, resilience metrics in transportation systems are either topological metrics (mostly based on network graph theory) or those that consider attributes and performance of the system. Metrics that evaluate the build-in ability of the system to act resilient are referred as *attribute metrics* while those that measure dynamic reaction of the system (e.g. throughput or network flow) are called as *performance metrics*.

As an example, in aviation, Yoo and Yeo [40], take an attribute-based metric to measure the resilience of air transportation network in US. The metric addressed the adaptive capacity of resilience to measure the ability of the network to replace a disrupted node with an adjacent node.

Furthermore, in a network graph study, Janić [41] proposed a resilient metric based on the airport importance (relative to share of each airport in accommodating flights) and operated flights by an airline. Janić [42] later continued his work to assess resilience from another set of indicators in a freight transportation systems, including: flights, profits, time of transport, and the inventory cost at airports.

Another example of attribute metrics in topological models, is a commonly used metric: size of giant component. It is calculated by determining the proportion of nodes that act as of a cluster. In aviation domain [43, 40] it has been used to realize the impact of resilient strategies to maintain the network connectivity since the metric measures the network's proportion that can be travelled by air routes.

Critical component analysis also uses attribute metrics in resilient air transportation studies [44]. This technique is important with regard to limited available resources in time of a disrupted network. It sets priorities to assign resources to most crucial nodes of a network either in strategic pre-cautionary planning or in post-disaster mitigation strategies.

It is important to notice that in almost all of the mentioned references, metrics do not identify if a system is disrupted or not but rather assign a score to system resiliency. In fact, the system vulnerability can be classified into two topics: the probability of having disruption in the system and the magnitude of system disruption. Most cited works focus on the latter that is also recommended by some classic studies [45, 46] on vulnerability. These studies claim that measuring the consequences should be the primary objective. This objective is referred as conditional vulnerability and most of the mentioned studies on resilience follow this line of research and not the definition of disrupted system. As a result, this thesis is relying on performance metrics rather than attribute metrics since the purpose of the work includes network state definition.

Apart from type of resilient metrics, the contribution of this thesis in using system resiliency becomes clearer by better clarification of system vulnerability, especially in terms of flexibility and robustness. Therefore, next section is dedicated to distinguish robustness from resilience.

2.1.2 Resilience and robustness

Although to some extent resilience, robustness, stability and flexibility are used in the same context but each is technically different. Resilience is a comprehensive term for the ability of a system to handle changes, while robustness is more focused on the absorptive level of resilience, as inherent resistance against stresses beyond normal system functionalities. Thus, robustness is less likely to support performance variability in the tactical ATFCM phase. EATMN belongs to complex networks that are counted as robust if basic functionalities remain operational under the failure of sub-components. However, in order to provide an exclusive and inclusive definition for resilience and robustness, the following general definitions are provided:

- **Robust System**: A system is identified to be Robust, if it has the ability to continue functioning in the presence of internal and external challenges without fundamental changes to the system [47]; In other words, a Robust System is *designed to prevent possible failures*.
- **Resilient System**: A resilient system accepts the inevitable challenges of its dynamic structure and in case of a disruption, adapts itself by changing its operational processes while continue its core functionality. This is also addressed

in early definition of the word 'Resilience' in Psychology [48] which is coming from the Latin root, i.e. 'resilire', meaning "to jump back" or "to recoil." Hence, a resilient system is based on early discovery and fast recovery *from unpredicted (or emergent) disruptions.*

In other words, robustness indicates a system design to cover more uncertainties. But, the improvement of system performance in order to empower system compensation during a disruption is the key feature of a resilient system. Robustness in ATM is more investigated from topological aspects and indicators. Such studies help to realize important local nodes in a modeled ATM network [49].

In transportation networks, systems are subject to disruptions. Therefore, robustness and resilience of such systems can be improved by means of increasing the redundancy as an example. However, such measures and their associated investments can be very expensive. In transportation networks, appealing sustainable and feasible solutions are generally based on more effective management techniques. These techniques are highly relied on remodeling and optimization of underlying complexity of the system.

Here system modeling is a technique that helps to realize system vulnerability and recovery. Such a model should address different states of the system. States can be defined according to many aspects such as life-cycle or performance levels. Next section discusses, the system state that is defined by levels of performance.

2.2. System state

Resilience is a concept that deals with the functionality of a disrupted system. But with regard to conditional vulnerability, the first step toward resilience is to realize the state of the system. Devoe [50] provide a theoretic notion of system state in his book:

> *"at each instant of time, the system is in some definite state that we may describe with values of the macroscopic properties we consider to be relevant for our purposes. The values of these properties at any given instant define the state at that instant".*

This definition reminds that a system state can be expressed by a set of variables (i.e. key indicators). But in case of a dynamic system, such a definition can lead to indefinite number of states. Therefore, phase transition is a better alternative for complex systems such as EATMN. These systems intentionally have a built-in degree of flexibility in its internal environment.

Devoe also mentions an "equilibrium state" as a state that remains unchanged indefinitely unless some external forces violates its internal environment. Theoretically, a sealed system with zero interactions with the surroundings is named as an isolated system. In such a system if any change happens, the system will experience a series of reactions that concludes with a new equilibrium state. In conclusion, "steady states" is noted by him, that is different from an equilibrium state.

A steady system is regarded as constant for a period of time that it exchanges matters or energy with its environment. As an example, one can consider a thermometer. Once the thermometer is in contact with a cold or hot object (exchanging heat energy), it measures the temperature accordingly (remain stable) without malfunction or (tangible) heat exchange with the person using the thermometer.

Such a definition is more consistent to EATMN state definition since EATMN interacts with different level of internal (e.g. passenger demand) and external forces (e.g. weather uncertainties) without the need of major airspace closure. Consequently, in this thesis addressed EATMN state is assumed to be under the category of steady states. In resilient studies a key assumption is that the system at study has at least one steady state (disrupted vs. nominal state).

The connection of steady states and resilience is better discussed in Disaster Risk Reduction (DRR) studies. In fact, the combination of resilience and stability in general, motivated the subject of disaster resilience, especially after the disastrous Tsunami of 2004 in Indian Ocean.

Coetzee et al. [51] mention the discussion of DRR and resilience. They also noted that the definition of system resiliency as the ability to recoil (to bounce back) needs careful consideration since a complete reset leaves the system vulnerable to similar disruptions in future. The statement is not relevant to all levels of resilience (Table 2-1) but mainly a key aspect in absorptive level. In restorative level of resilience, the resilient system is able to maintain its core functionality while reducing the effects of disruption.

In other words, an improved steady state is not the goal at restorative level of resilience. As illustrated by Figure 2-3, the initial state (S_0) and the final state (S_f) are not necessarily at the same level of performance (F_t) but both are significantly improved states compared to the disrupted state (S_d).

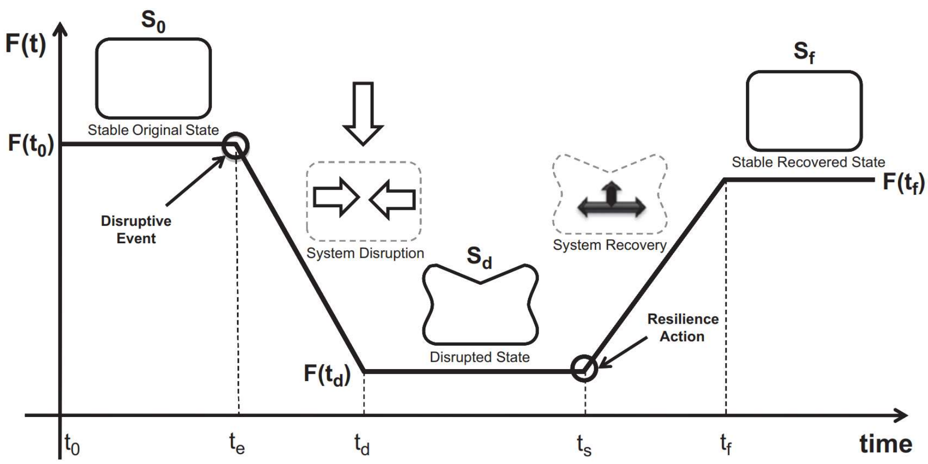


Figure 2-3 System states and Resilience adopted from, *[30]*

A review on system studies suggests that some distinguished the state of a system from its mode. This difference does not apply to all systems. But as an example, in aviation, Baduel et al. [52] focused their research on the definition of states and modes in a system engineering context. The study contributed to the specification and design of system behavior. Based on their analysis system states have the following aspects:

- Each state is able to characterize the system;
- each state is valid or considered at a given point of time or at a time window; and
- states are providing a specific kind of information, knowledge domain or system design (e.g. operations, level of readiness, energy).

Similarly, one can conclude the following points about the "mode" of a system based on their study:

- A mode also characterizes the behavior of a system;
- modes are defined for a set of conditions that is not necessarily a temporal condition (e.g., specific states of the system); and
- modes express a behavior regarding a set of capabilities, functions or actions (e.g., moving forward or backward, performing flight maneuver, etc.).

> *In other words, the state of a system is mainly a variable that can be measured and quantified while the mode of a system is the label that tags a set of system behaviors that are either activated intentionally or triggered by disturbances of the system.*

Similar to the selection of a correct network definition, the description of non-nominal states is relying on correct network impact description. For example, Cook et al. [53] described non-nominal states as a phase transition. It refers to high number of locally interacting elements causing a collective phase change. It is concluded that unlike other traffic systems (e.g. road traffic networks) phase-transition behavior in air transportation systems requires more solid definitions.

Nevertheless, in terms of network states, nominal and non-nominal conditions are not similarly discussed in the literature. In their study on a passenger oriented and event-driven model, Cook et al. [54] considered stochastic growth of average departure delay as an indicated disruption against nominal conditions. In another attempt to study resilience of air transportation system as an optimization problem, Filippone et al. [55] examined non-nominal conditions in order to find the resilience path that is the most valuable chain of processes to push the system back to nominal conditions. This microscale[1] study described a model for non-nominal scenarios based on different quantifications of given key performance indicators (KPIs).

[1] Microscale, *Mesoscale and Macroscale studies are explained in next section.*

Notation of states is also studied in other domains of aviation such as aircraft maintenance schedules. But in comparison to network states, there are significant differences in definition of states and the dynamics. As an example, a recent study by Andrade et al. [56] applied reinforcement learning (RL) for maintenance scheduling. The methodology takes a transition function to relate different states at a cost (as of a Markov Decision Process). But in a network state, there is no global set that contains all possible states. Even if such a set is given, the complexity of parameters and unknown dynamic of intertwined operations at such a scale pose a severe challenge in defining a transition function. In a similar study at delft university [57], maintenance states are addressed by optimization. However, instead of a transition function, the stochastic framework takes a two-phase state transition based on known probabilities.

2.3. Assumptions and scope

To summarize, reviewed studies on resilience and air transportation system reveal some gaps and guidelines in the literature. A summary of these potential research directions is offered below:

- despite advantages of organizational efforts such as establishment of EUROCONTROL in Europe, the air transportation system needs to consider structural changes with respect to less cultivated concepts such as resilience;
- incidents such as delay peak in 2018 cannot be modeled by classic comparative studies. Current European CASA algorithm based on first-come-first-serve (FCFS) is less efficient in case of a saturated network;
- resilience engineering with methods such as FRAM is providing better principles for dynamics of air traffic flow management. For instance, functional resonance is more relevant to emerging disruptions in ATFM;
- mapping of resilience levels and ATFCM phases, acts as a reference to locate different mitigation approaches. It also helps to assess the impact and costs of modifications at each resilience level;
- Importance of system state definition is discussed and supported by a review on resilience measuring techniques. In aviation domain most studies on resilience are dedicated to measuring resilience (by attribute metrics) as a mean to evaluate imposed costs but the contribution of resilience as a detecting mechanism (by performance metrics) is not addressed. Most studies focus on major disruptive events and no study is dedicated to emergent disruptions;

- lastly, the review of studies on system state and its mode, provided the visibility that performance variability needs to be addressed in a system state context rather than developing recovery scenarios in different system modes.

As discussed, resilience is a concept that is defined on the system level. EATMN is considered as a system and this study aims at understanding its resilience. Therefore, the goal of this exploratory study is to work on conceptualizing the EATMN resiliency as an attempt to investigate the possibility and mechanism of addressing resilience in an ATFM context. Secondly, the thesis is set to enhance performance of the network from a network manager (NM, i.e. EUROCONTROL) perspective. The best practice as of today is the use of simulations to predict demand-capacity imbalances and to evaluate the efficiency of capacity regulations. This research is providing an alternative by machine learning based predictions instead of simulations. Demand-capacity imbalances are addressed as network disruptions and capacity regulations are regarded as reviving measures in a resilient ATM system while performance metrics are used to measure network resiliency.

In general, the complexity and dimensions of the EATMN make it challenging to detect disrupted network situations by monitoring procedures and operations. Thus, the term emergent is used rather than the term resultant to describe such situations. Understanding network states provides a better opportunity to investigate emergent disruptions rather than resultant failures. Therefore, one of the objectives of the thesis is to propose a methodology to capture emergent disruptions as a result of dynamic interactions among DCB actors in tactical phase of operations.

Emergent forces in a network are more likely to happen in mesoscale or macroscale of the ATM system. Cook et al. [53] defined three scales to investigate emergent interactions: *microscale*, that only considers a single flight; *mesoscale* as an intermediate scale covering a given airspace with many flights following a given set of rules e.g. as in a Terminal Maneuvering Area (TMA) or in Air Traffic Control (ATC) sectors; and the largest scale is *macroscale*.

A macroscale air transportation network can be considered at the level of regional, national and supra-national networks, or even at the level of the global ATM system. As an example of macroscale studies, impact of major external disruptions on an ATM network has been studied by Lau et al. [58, 59]. Hosted by DLR-air transportation systems, they analyzed weather-induced network disruptions that generally have adverse effects on network performance. This study led to better understanding on interactions of ATM subsystems. But more importantly it has intensified the necessity of implementing systematic concepts such as resilience to an ATM network that supports stabilized functionality and performance levels.

Another known factor of macroscale studies on air transportation modeling and resilience is the underlying data [60] and subsequent limitations. Therefore, this thesis has some basic assumptions toward understanding the mode (behavior) of the EATMN as a system. These collected assumptions enable achieving higher levels of control over EATMN state:

- while in most studies on system resiliency, the key assumption is that the system is already disrupted; this study is aimed at realizing if a system (EATMN) is suffering from such a wide disruption;
- the scope is not covering sources of uncertainties in air transportation system such as prediction of adverse weather situations or large-scale disasters (e.g. volcano eruptions). The assumption is that such uncertainties are ultimately reflected in demand-capacity imbalances;
- demand prediction topics and related indicators are avoided since the idea is to quantify network state that eventually serves as a baseline to standardize common ATM performance indicators;
- the size of available ATM data and their update rates challenge the consolidation of relevant data from different stakeholders. Therefore, selected basic descriptive statistics at relevant stages of the study are considered to avoid excessive complications;
- since the initial survey (Figure 2-1) verified seasonal patterns in ATFM delay, historical data has been compared to understand current network state. This approach is selected based on the fact that, although the network is always impacted by different sources of uncertainties, imbalances are part of a finite set of possibilities -which can be considered as recursive scenarios.
- as a proposed rule, emergent non-nominal states are declared based on control intervals that assume network states as nominal in 99.3% (corresponding to 2.7σ) of the cases. In fact, it is possible to investigate less or more severe disruptions relative to control intervals by modifying this assumption.

  Here major safety issues or significant performance losses are in focus to address network resiliency. This reminds the different perspectives of resilience and robustness, since a robust system is hardly designed to prevent circumstances with occurrence probabilities of 0.7 percent and below. In other words, the control interval assumes that benefiting from strategic and pre-tactical plans, the ATM network has the flexibility to cope with disruptions up to 1.5 times of estimated imbalances (± 1.5 Interquartile Range, i.e. 2.7σ).

3. Methodology

This chapter provides an overview on the methodological selections and the design of the research. The main aspect in driving such decisions is the research problem. More, specifically, it explains why and how the study is started with an exploratory design and extended to assessments. Further, this chapter gives the procedures to selection of ATFCM regulations, collect relevant data, analyze them and understand the European ATM network resiliency [61, 35, 62, 63, 64, 65].

Moreover, the close collaboration with international research organizations in the SESAR Solution PJ09.01 “Network Prediction and Performance” fostered the methodology and its operational benefits. Project PJ09 had 29 partners ranging from research organizations (e.g. DLR, Thales), airports (e.g. Heathrow), major European ANSPs (e.g. NATS, DSNA, ENAIRE and ENAV), to airlines (e.g. Air France). Such a wide range of audience, enriched the operational understanding of the different understanding of resilience in European aviation community. Resilience is more perceived as a safety topic, but resilience (as studied in this thesis) is also about performance. Performance in aviation is highly connected to indicators such as delay which is further discussed in this chapter that explains the research design, describes the EATMN state, predicts disruptions by learning methods, explains data collection process and concludes by describing developed tools.

3.1. Research design

Considering three forms of research design: exploratory, descriptive and explanatory, this thesis is following an exploratory theme since the concept of resilience is mainly described from a safety engineering perspective [33]. Subsequently, resilience engineering [66] changed the focus to performance management rather than safety concerns. However, in practice the aviation industry still categorizes the resilience concept as a safety topic (also observed through early brainstorming sessions of the SESAR project PJ09). In contrast to most studies that consider resilience as a safety measure with attribute metrics, this thesis is set to be an exploratory study on European ATM resiliency through performance metrics.

Exploratory studies are those conducted during the early stages of a research, mostly in conceptualizing an idea or doing feasibility studies (as of framing the ATM resiliency). In comparison, descriptive researches focus on well-established problems such as delay in air traffic flow management. Explanatory researches explain why a particular phenomenon exists to provide answers to its causality. The behavior of EATMN is resilient to some extent and the first section of this chapter works on network state to illustrate this resilient behavior through tactical situational awareness.

3

3.2. EATMN state

EATMN is unique in its complexity due to numerous stakeholders, airspace configurations and accommodated traffic volumes. As discussed previously, such a complexity challenges both tactical visibility and the network resiliency in general. Therefore, in this section the principle of performance variability used to determine network states. To this end, performance metrics and indicators are required. However, due to the absence of a solid baseline (or reference state), network state is proposed to be bound to consolidated. The methodology is based on capturing the emergent disruptions as drivers of performance variability. A key assumption and claim is that emergent disruptions across network are revealed through capacity regulations as restorative mechanisms for tactical ATFM.

After describing the regulation data, this section addresses the network state definition at two divided levels of macro and micro analysis. Macro analysis serves as a constant monitoring scheme while micro analysis is only focused on disrupted states.

Based on capacity regulations the results show that proposed statistical approach is even able to distinguish non-nominal disruptions to either crisis or critical states. The proposed approach is then demonstrated by a data sample covering six months. Furthermore, to assess the severity of non-nominal states, the probability distributions of different regulation types are estimated. This section is then concluded by offering insights on long term network resiliency based on estimated probability distributions.

The general overview of how resilience is bound to state of the system is emphasized in Figure 3-1. Knowing the level of system performance at each given time (P(t)) the state can be monitored based on the extent of disruption. The general assumption is that in the design phase, the realistic assessment of system internal and external forces (in strategic and pre-tactical phases) enables EATMN to maintain its functionality for most of its life cycle (remain nominal).

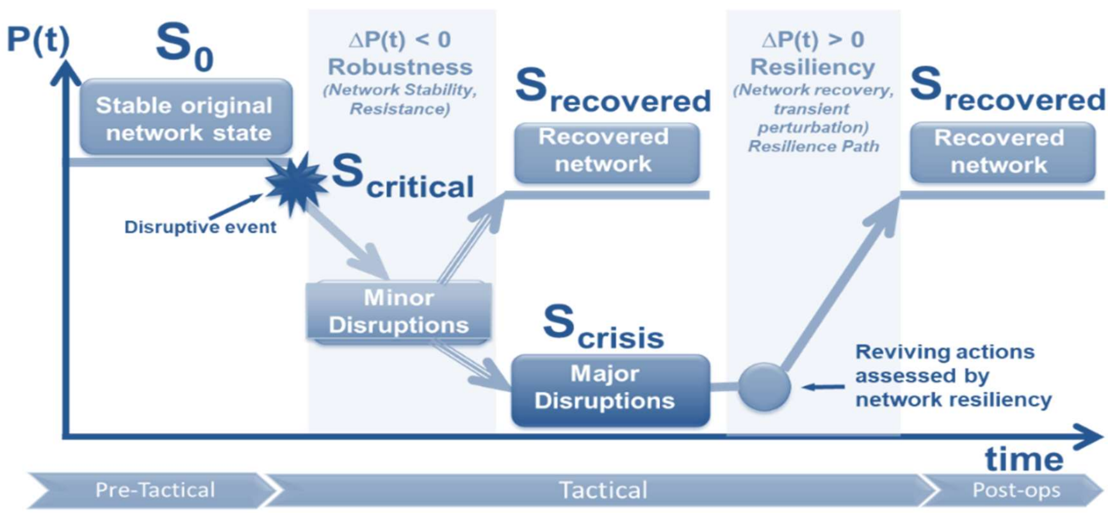


Figure 3-1 Symbolic model showing the difference between robustness and resilience of a system and their connection to system state. S represents the state of the system and P is the level of its performance.

Figure 3-1 shows that a resilient network accepts the odds of performance loss beyond its robustness (critical/crisis), yet still is able to revive its core functionalities through resilient capacities to reach a new steady state ($S_{recovered}$). In order to realize such a resilient EATMN, the proposed approach is to (I) find most contributing data that describes and contributes to tactical situational awareness (measures levels of S), and then (II) make statistical inferences for resilient decision-making (is the network degrading to a critical state or a crisis?), and (III) develop monitoring measures, i.e. thresholds for network state identification (i.e. micro analysis that informs on magnitude of disruption).

An unusual ATFCM situation or even crisis might be triggered by imbalances between capacity and demand as a result of major capacity losses. Another trigger may be major failure of information flow in at least one sub-system of the network [67]. Such situations are well-planned and managed by NMOC. Guidelines are also available for local contingency planning for national ANSPs in the event of failure or disruption of related services. These definitions and such procedures are contributing to safety-I perspective [68, 66]. Safety-I is more concerned on failures or adverse outcomes and tries to enhance preventive mechanisms or containing the consequences. Conversely, safety-II perspective is more bounded to performance levels since it is focused on successful outcomes (safe and efficient performance) rather than mitigation plans. Safety-II considers performance as a variable, that leads to study system characteristics to understand successful safety mechanisms. System resiliency and safety-II perspective are more aligned as both rely on constant performance monitoring and are not only focused on degradations.

In general, the complexity and dimensions of the EATMN make it challenging to detect disrupted network situations through monitoring numerous procedures and operations. Thus, the term emergent is used rather than the term resultant to remind that causal links are not at focus. Understanding network states provides a better opportunity to investigate emergent disruptions rather than resultant failures. Therefore, following sections provide a proposed methodology for capturing emergent disruptions as a result of dynamic interactions among DCB actors in the tactical phase.

In search of most contributing data-type different databases were compared with respect to certain criteria (section 3.2.1). The acquired data are firstly used to provide statistical inferences about network state in general (section 3.2.2) and secondly to provide a more detailed overview on characteristics of identified non-nominal state (section 3.2.3).

3.2.1 Selection of regulation data (ANM data)

The consolidation challenge of relevant data from different stakeholders and update rates is considered to be managed by basic descriptive statistics and several data-types and databases were compared with regard to the following criteria:

- Update rate: the database must be published throughout the tactical and pre-tactical phase to be more relevant to decision making processes on the day of operation,
- Granularity: selected data should be able to provide required precision to understand types of disruption including spatial and temporal dimensions,
- Coverage: selected data shall be relevant and accessible by all layers of decision makers across the European Civil Aviation Conference (ECAC) area.

According to the mentioned characteristics, delay statistics including reports from the Central Office for Delay Analysis (CODA), statistics and forecasts (STATFOR) and those that are published in the post-operational phase are not considered. Likewise, databases including National Performance Reports (NPR) and ATFCM Statistics and Network Operations Reports are only published for authorized users and cannot fulfill the coverage criterion. Among all capacity (ATFCM) regulations denote the results of collaborative decision making. In fact, regulation is a method of matching traffic demand to available capacity by limiting the number of flights planned to enter an airspace or aerodrome [7].

ATFCM regulations are initiated based on the evaluation of ATFCM Daily Plans (ADP) from the pre-tactical phase while being updated constantly in tactical phase. Regulations correspond to network states in the restorative level of resilience and are accessible through ATFCM Notification Messages (ANM) that are published by NMOC before the day of operation. In contrast to ADP, ANMs are updated throughout the tactical phase. Moreover, it is considered as official medium for the notification of ATFCM measures (regulations) to all actors [69]. These messages are offered to provide a summary of planned measures and to promulgate any specific instructions on them to represent each ATFCM regulation. Finally, as ANMs fulfill all three criteria it has been selected to represent ATFCM regulations.

The evaluated amount of data at this stage of the study covers six months, from May to October 2017. The investigated period of year is chosen as previous studies [70] on European air transportation system revealed that network centrality measures for both air navigation route network and airport network are significantly fluctuating from AIRAC[1] sixth cycle (May) up to the end of eleventh cycle (October).

3.2.2 Macro analysis

To identify potential non-nominal states, a macro analysis is performed that is focused on regulation counts (Step 1) and durations of active regulations in the tactical phase (Step 2). Since the interest lies in emergent characteristics of non-nominal network states, the approach is depending on the size of the assessed data sample. In other words, a benefit of statistical

[1] Aeronautical Information Regulation And Control (AIRAC), see annex B.1.

inferences is that based on sample size, statistics get different values with the same confidence levels (i.e. different patterns can be monitored with different thresholds). This dependency enables the realization of temporary patterns as well. This means that a non-nominal state is able to be compared against various time frames of network performance. Here, ANM data are analyzed in two different time frames: six-month (seasonal patterns) and single month (weekly patterns).

As discussed in previously, resilience is about network behavior in disrupted conditions and not only in degraded conditions. Therefore, in step 1 (counts) a two-sided control interval is used to detect *outliers* because the intension is to monitor both negative and positive deviations. Such incidents (positive disruptions: absence of regulations) may indicate an impact of other factors (e.g. airline strikes). They may also provide the opportunity to update performance baselines in terms of accommodated traffic demand.

In step 2 regulation durations are additionally evaluated as they provide more details on the severity of a network disruption. Despite the relevance of number of affected flights to the network state, the concrete number of affected flights per ANM is only available to all actors in post-operational databases. Therefore, the only tactically available data are regulations, from which their magnitude can be assessed by their duration. Consequently, outliers are identified with respect to calculated mean and standard deviation as descriptive statistics of regulations' duration per day.

The mean duration of published regulations represents the overall severity of the disrupted network condition while standard deviation indicates the dispersion of the problem. Depending on different combinations of mean and standard deviation values, non-nominal states are classified to critical and crisis states (Figure 3-2) with the following definitions:

- Critical state: Regulations show large mean values (exceeding control intervals) accompanied by significant large standard deviations. Such a condition emerges when severe but local disruptions are affecting network operations. Hence such states need NM support in collaboration among local actors of both ANSP and airport networks to handle traffic flows.

  Based on this definition and added dimension of activation time for regulations, both nominal and critical states are broken down into more specific types of states when the methodology is implemented into developed tools (see NetRes in annex C.2).

- Crisis state: Regulations show large mean values with rather small standard deviations. The network is facing a wave of prolonged impacts restricting safe operations. In such situations the loss of airspace capacity is so severe that the number of available restorative measures is very limited. Accordingly, NMOC is the main actor in handling the situation which is no longer a regional issue.

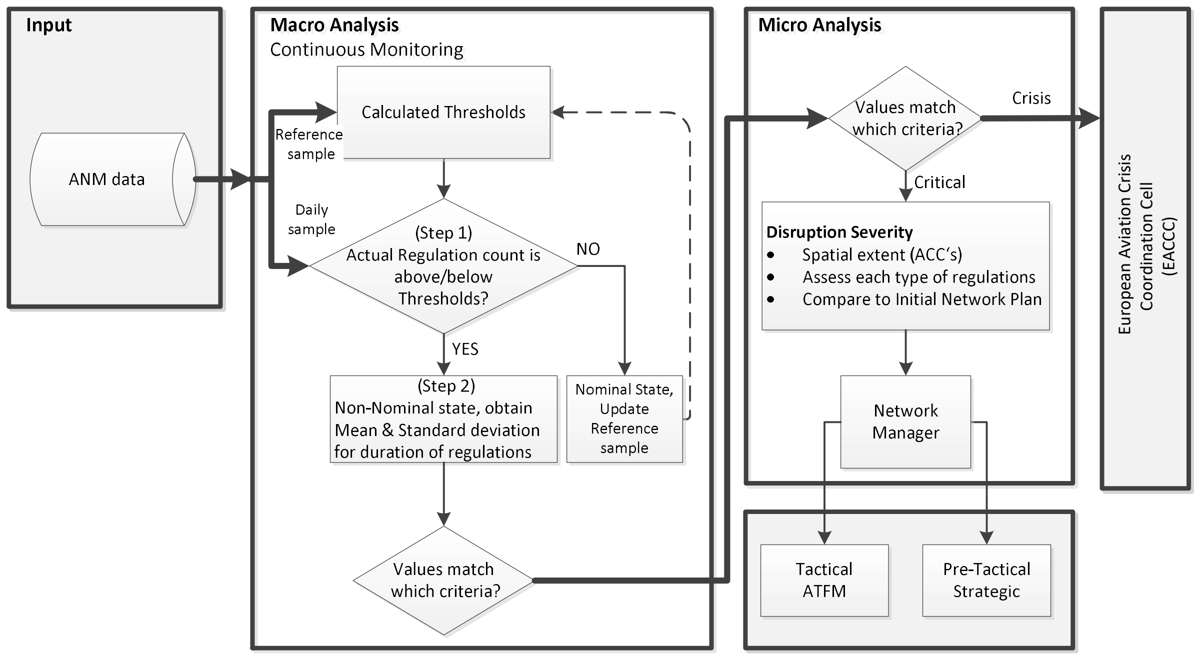


Figure 3-2 Two-step model of network state detection. In macro analysis each day (a daily sample) is compared against thresholds corresponding to selected reference sample (e.g. a month). Note the different procedures of reviving a critical state and a crisis in EATMN.

In strategic planning horizon, different stakeholders are more concerned on reserving the capacity and scheduling their resources in advance. In pre-tactical phase there is a need to re-allocate the tasks and resources in accordance with latest updates of other stakeholders. Such updates (Events, serious prolonged weather conditions, strikes or major technical problems in resources such as maintenance issues) or disturbances may interact with each other and potentially lead to adverse impacts that may span over multiple spatial and time scales [55].

As provided in Figure 3-2, the devised algorithm proposes a mechanism to capture network states and only in case of a non-nominal situation, the algorithm proceeds to micro analysis that improves the restorative level of resilience and provides analytics for absorptive and adaptive levels of resilience. This model was also documented in PJ09 project [11] after being presented to project partners (refer to section C.2.3 and Figure C-16 in annex).

3.2.3 Micro analysis

The micro analysis considers type of regulations and network states. Results also contribute to overall robustness of an ATM network by realizing critical airspaces. For such an analysis, the following challenges are identified and addressed:

- Data type: at tactical phase, publicly available data are ANM, ADP and Initial Network Plan (INP) but the structure of ANM data is different from the ADP and Initial Network Plan (INP) in terms of terminology and format.

- Data precision: each ANM record (i.e. regulation) can be evaluated in terms of the number of affected flights. In fact, the operational flexibilities on flight plans and delay assignment algorithms result in different counts for regulated flights. However, such details are published in the post operations reports.
- Temporal dimension: ANM data are updated in tactical phase by push messages and accordingly contain 'change' and 'cancellation' records. But ADP is presented as a reference document that serves as an input for tactical phase.
- Spatial dimension: ANM data are referenced to traffic volumes. A traffic volume can be referred to an airspace, point, aerodrome or set of aerodromes, i.e. they can be assigned to both ANSP and airport networks. Diversity in visualization of data is covered by considering related ACC
- that leads to homogenous set of reference locations.

With respect to mentioned challenges, ANM data is divided based on the regulation reason or cause. As stated in the ATFCM user's manual [10], causes of regulations are divided into 14 different categories. Major five types are considered including ATC capacity, ATC routings, ATC staffing, aerodrome capacity and weather. The remaining nine categories are integrated into a single type, named 'Others'.

To provide secondary inferences on a critical state, initially the distribution of data is tested by quantile-quantile plotting to understand the dispersion of regulation types and to realize any similar distributional patterns among different regulation types. Figure 3-3 depicts the comparison of each regulation type against normal distribution through estimated normal cumulative distribution functions (CDF) on sample data. The advantage of the given plot is that it shows different statistical aspects including the shift in scale or location, presence of outliers and changes in data symmetry. The figure also shows the expected significant deviation from fitted normal distribution of the integrated category of 'Others'. Upon declassification of mixed 9 regulation types, it has been realized that the regulation type 'ATC industrial action' (mostly strikes) is the main driver of this deviation. Likewise, weather regulations are proved to be far from best fitted normal distribution. Knowing the significance of weather impact on ATM network, weather regulations are selected to demonstrate the methodology in estimating the Probability Density Function (PDF). As the most challenging type, weather regulations are proved to be from a skewed and heavy-tailed probability distribution (Figure 3-4).

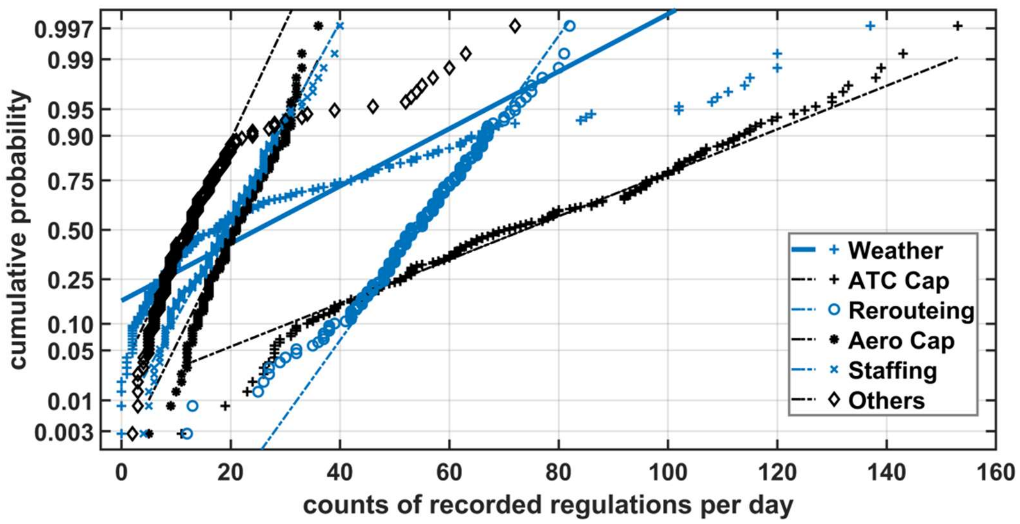


Figure 3-3 Simultaneous comparison of ANM data against estimated normal distributions. Note the extended distribution of Weather and ATC Capacity compared to other regulation types. Such regulations are expected to have a heavy-tailed distribution.

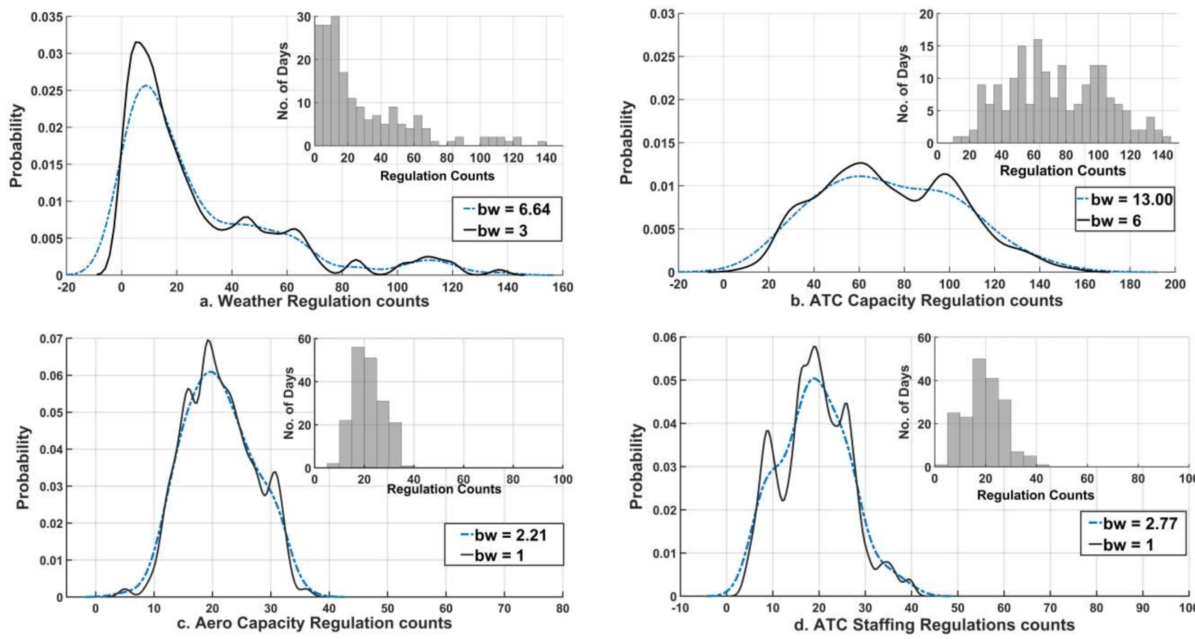


Figure 3-4 Sample histogram and estimated probability density function of regulations' count from ANM data (2017: 01.May-31.Oct).

For micro analysis, the counts of weather regulations in ANM data are plotted separately. Results of initial sets of curve-fitting proved that the probability distribution curve is considerably skewed and has an asymmetric multi-modal shape as given in Figure 3-4a. Such distributions cannot be represented by parametric distributions like Poisson or Gamma. Therefore, the use of Kernel density estimation with a normal smoother function is proposed. It is a method for estimating PDF of samples from an unknown distribution. Kernel estimation actually computes the probability of data by dividing the domain into intervals and then estimating PDFs for each. The final PDF is provided by merging them (Figure 3-4). Kernel estimation relies on a probability function and a fixed bandwidth (denoted by bw). Dashed curves are based on the best bandwidth that statistically minimizes the errors while black curves show the bandwidths from the proposed correction method as estimations for: (a.) weather (heavily skewed) and (b.) ATC capacity regulations (multi-modal). Note that for other types of regulation there is no need to correct the bandwidth (as given in (c.) and (d.)).

Kernel density smoothing methods mainly differ from each other based on the kernel function (K) used. In the case of using the PDF of normal standard distribution (zero mean and unit variance), the smoothing is called normal kernel. Other common kernel smoothing functions are called box, triangle and epanechnikov [71]. The final distribution is estimated by cumulating the probabilities based on multiple estimated normal distributions for each interval.

Normal kernel smoothing is used with the formulae given in equations (1), (2) and (3) that are evaluated over each data point within the interval [xi – (bw/2), xi + (bw/2)]. As a normal kernel, the error function (erf in equation 3) is used as it denotes the probability of observing a random value in the interval [-x, x]. The equation given in (3) is needed for the computation of CDF when they are transferred from a discrete into a continuous distribution. Cumulative functions are later required to set thresholds for each regulation type.

$$Y_{bw}(x) = \frac{1}{n.bw} \sum_{i=1}^{n} \left[K\left(\frac{x - X_i}{bw}\right) \right] \quad (1)$$

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{-2}} \quad (2)$$

$$\mathrm{Erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \exp(-t^2)\ dt \quad (3)$$

$$\int \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = \frac{1}{2} \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) + \mathrm{C} \quad (4)$$

The distributions for weather and ATC capacity regulations as given in Figure 3-4 are with regard to described normal kernel smoothing method. The right bandwidth is obtained from minimization of errors. This statistical approach assures the best fit to the data but results should be refined to be consistence with the operational understanding of the data.

From a pure mathematical perspective, it is well established to use minimization of errors to select the best bandwidths. However operational understanding of the data reminds the necessity of assuring the values of calculated bandwidth for each regulation type in our study.

As given in Figure 3-4, statistically optimal values of bandwidths are specified by dashed curves, but these bandwidths need to be verified as negative values cannot represent the number of regulations on a given day (Figure 3-4a and 3-4b). The corrected bandwidth for such type of regulations is actually the greatest integer less than (floor or round down) half of the optimal bandwidth. Characteristics of such a corrected curve are more contributing to identification of thresholds for detecting outliers in shuffled data. Also, less smoothing decreases the loss of precision due to the probability of observing negative values which has no added value.

It is also realized that for larger datasets the recommended bandwidth is even smaller since the data size has a negative correlation with bandwidth. For other regulation types, no correction was applied and the statistically optimal bandwidth is considered for estimating the probability functions. Nevertheless, the screening of the results with both methods was also considered for every type of regulation (Figure 3-4).

Once the probability density curve is extracted from the data, a second set of thresholds according to reference confidence levels is calculated within the micro analysis. In the results section values based on estimated CDFs are given (Table 5-1).

ANM messages can also provide spatial patterns of regulation types (Figure 3-5). Also, potential relationships among types of regulations are addressed based on estimations in micro analysis (see results section Figure 5-2).

Figure 3-5 projects regulation data on airspace volumes considering ACC areas of responsibility. The borders for each ACC (relative to Cross Border Area- CBA) are gathered from the EUROCONTROL's Demand Data Repository (DDR2). On top of quantitative results of the micro analysis, such figures provide geographical perspective for disrupted areas. According to the given guidelines on regulation [10], each type can be interpreted differently with regard to consequences it implies. ATC capacity regulations may include flights in departure, arrival and en-route phases. Hence, the potential efficiency of such regulations is expected to be high in resolving imbalances between demand and capacity (Figure 3-5a). Moreover, weather regulations (Figure 3-5b) denote a reduction of planned capacity. These regulations also affect departure, arrival and en-route segments but the prediction uncertainty and available measures to counteract are quite different than ATC capacity.

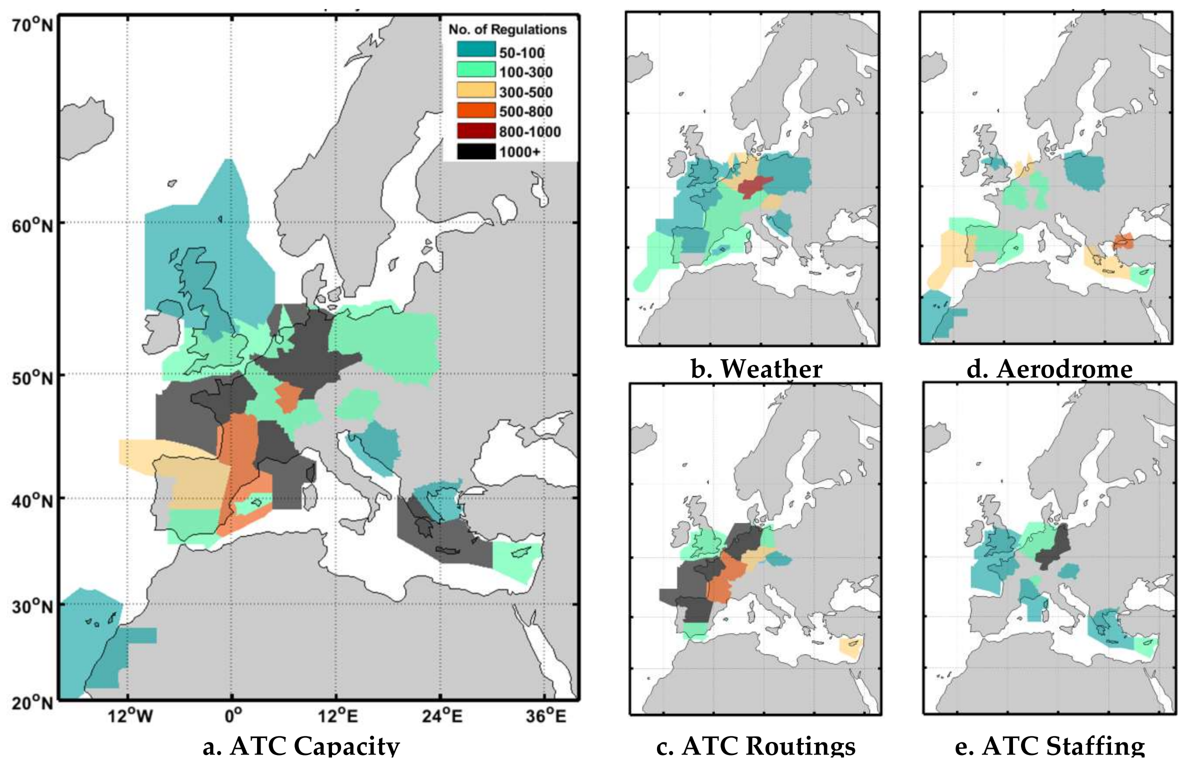


Figure 3-5 ACCs and count of different ATFCM regulation types across Europe (May to Oct. 2017) based on ANM data. Geographical patterns for five major **regulation types** are provided as heatmaps.

There are more specific regulations like ATC routing regulations (Figure 3-5c) that are only applied to en-route flights. Similarly, aerodrome capacity regulations (Figure 3-5d) only affect departures and arrivals, which are more contributing in understanding airport network disruptions. Finally, ATC staffing regulations (Figure 3-5e) are less frequent across the network. This category includes unplanned staff shortage and has signs of correlation with other type of regulations. In the result section a closer look to potential causal relationship among different types of regulations is provided. Generally, the study shows that regulation types reveal even more detailed input for other levels of resilience with regard to different affected flight phases and driving factors.

3.3. Disruption management procedures

As declared before, robustness of a system is the ability to avoid majority of failures. This means that there are a number of cases that can be detected based on the information that are used in design phase of the system. For instance, the underlying processes such as flight scheduling, sequencing processes and optimization models, each have defined risk management plans and contingency strategies to maintain functionality or minimize the impact of possible failures. A good example of this case was recorded on 3rd of April, 2018 at EUROCONTROL.

EUROCONTROL reported [72] that the flight plan data in the Network Manager's (NM) IFPS and Enhanced Tactical Flow Management System (ETFMS) was accidentally deleted on the day. The tactical operations were disrupted for a total of 12 hours and 40 minutes. The ATFM procedural contingency plan (a robustness plan) was activated which included precautionary reductions in ATC capacities and reduction of airport departure rates (ADR). As the backup solution, EUROCONTROL activated the contingency site at the EUROCONTROL Experimental Centre (EEC) in France to contain the impact. The robustness of the system and designed risk assessments and contingency plans allowed the European airspace to cancel only a few flights at the day and the negative impact was only captured by excessive delays.

Similarly, for airlines, robustness is conceptualized in different ways so that for this stakeholder majority of the deviations are contained in normal operations. One of them is the designed buffer in scheduling process. Buffers are also referred as scheduling contingencies. Table 3-1, provides an example to better understand different type of allowances in airline scheduling. The "Off-block" buffer is allocated to the aircraft in getting from gate A to gate B. The purpose of this buffer is to absorb off-block delays such as taxi, line-up, runway sequencing (in ASMA- Arrival Sequencing and Metering Area) and airborne delays (such as arrival management delays). Note that this type of buffer in Europe is consistent with the duration of a departure slot (15 minutes). The "At-gate buffer" is considered to cover delays incurred at destination to secure a punctual departure for the next leg of the aircraft. This type of buffer is designed for ground delays and possible recovery between rotations of airplane.

Table 3-1 Timetable for an airplane to illustrate buffers

Leg	Scheduled departure	Off-block buffer	Scheduled arrival	Turn-around time (min)	Slack time	At-gate buffer
1	dep. A: 0730	15	arr. B: 0930	60	0 min	15 mins
2	dep. B: 1045	15	arr. A: 1300	65	0 mins	10 mins
3	dep. A: 1415	15	arr. B: 1615	60	10 mins	5 mins
4	dep. B: 1730	15	arr. A: 1945	65	10 mins	0 mins
5	dep. A: 2100	15	arr. B: 2300	(out-stationed overnight)		

Adopted from [73]

Nevertheless, it might be inevitable for a flight to wait for connecting passengers or crew rotations before continuing to the next leg. Also, there might be an issue over the available departure slots at the airport. "Slack time" is the built-in flexibility to absorb such discrepancies for airlines.

Similarly, other stakeholders ensure some degrees of freedom in strategic planning phase to deal with the concept of robustness. As mentioned before, this capacity of handling such sorts of deviations is noted as absorption in Resilience Engineering.

However, terminology of industry does not clearly distinct robustness from resilience. a quick glance at industrial solutions from related businesses confirms this. In Europe, Lufthansa systems (one of the leading providers of IT services in the airline industry) offers a number of different scheduling solutions to enhance airline abilities in their planning. These services often follow other objectives than what academia pursuits in resilience. To present the bridge between business objectives and robustness/resilience objectives, a short review of a tool from Lufthansa systems is given next.

Lufthansa in partnership with airlines offers NetLine/Plan tool based on 20 years of experience in managing planning challenges. The tool's primary goal is to maximize *profitability* of an airline's schedule. At the very first stage, airlines need to optimize their route network. This is done by consideration of both market demand and other airspace users (airlines). In fact, there is a need to not only optimize the connectivity of hubs but also to monitor the route network[1]. In general, such solutions (designed to optimize airlines flight planning procedures) are based on the ATM statistics of strategic phase (up to 18 months before the day of operations) and market analysis in coping with following challenges:

- *Connectivity of Hubs:* slot constraints in managing the passenger streams in hubs are one of the main issues for airlines, especially airlines that rely on more than one hub in their operations. Consequently, solutions are offered in terms of decision support systems to deal with raised issues on airline schedules.
- *Route network:* each airline needs a calculated visibility over flight schedules in case of planning new routes. For instance, flight connections are managed and scheduled by consideration of factors such as local traffic requirements, crow rotation plan, and fueling options.

Most of the described aspects are challenges studied in the strategic phase of ATFM[2]. Strategic solutions offer robust flight planning procedures to maximize profit and prevent failures. All robust solutions, specifically in ATM have an intrinsic flexibility which is essential for a smooth traffic flow on the day of operation, hence slight deviations from plans and schedules are anticipated in business solutions. The challenge is to control such flexibilities by acceptable deviation tolerance levels for flight plans. Punctuality (not robustness or resilience) is the dedicated topic in ATM that investigates adherence to these tolerance levels. Nevertheless, Figure 3-6 shows the decreasing pattern of punctuality in both US and EU.

[1] This phase is covered by two solutions from Lufthansa systems: NetLine/Plan Hub optimizer and NetLine/Plan Route optimizer

[2] Strategic phase covers plans from several month up to 7 days before the day of operations.

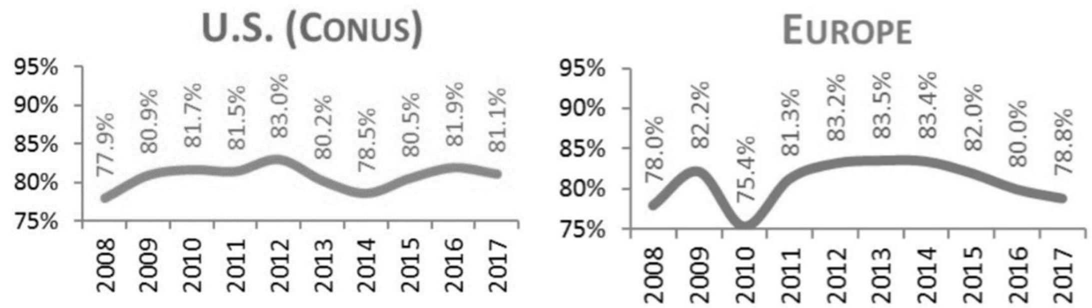


Figure 3-6 Arrival punctuality for main 34 airports. The vertical axis is showing the percentage of arrivals delayed by less than or equal to 15 minutes vs published schedule, [9]

In Europe the decreasing pattern is mainly driven by factors such as en-route ATC capacity shortage, adverse weather and ATC strikes (all regarded as disruptions in resilience). These tactical factors cannot be fully absorbed by strategic robust planning because of required flexibility. Implicitly, deterioration of such tactical issues may lead to network wide disruptions. Therefore, resilience as a broader concept that accounts for system dynamics is more capable than robust solutions in achieving higher level of control on EATMN states.

With regard to different levels of resilience, on top of effective strategic disruption management procedures (absorptive level), the interdependencies should be modeled to reach restorative and adaptive levels of resilience. From the Netline/Plan tool it is showed that the leading industrial solutions use state of the art methodologies but to profitability and not system resilience. This demonstrates that the announcement of detected EATMN states is less likely to be properly addressed by industrial tools and solutions.

Therefore, the thesis is continued by taking a data driven learning approach to navigate its exploratory research with trending methodologies across industry to keep the pace with stakeholder needs in Europe. To model raised interdependencies from EATMN complexity and its extent, this study investigates holistic learning methods based on enriched data flows.

3.3.1 Feasibility of machine learning approach

Perceiving EATMN in terms of resilience requires a closer look at its behavior. At network level, the operational procedures that structure ATFM are fully established and constantly updated through extensive research programs such SESAR. But such researches mainly contribute to absorptive level of resilience rather than restorative level at tactical phase.

At restorative level, reaction time is limited and active situational awareness is crucial. Compared to other complex systems, data availability in ATM enables Machine Learning (ML) to contribute to required situational awareness. ML in regression and classification problems is most effective when either the causal link cannot be defined or emergent behavior needs to be monitored. In EATMN, both the use of numerous procedures by different ACCs and impact of emergent disruptions such as weather conditions, motivate the use of ML to understand the dynamics in each network state.

Indeed a few studies are dedicated to take ML in addressing challenges of restorative level of resilience, especially among American academia. For instance, Gorripaty et al. [74] considered airport demand figures, capacity estimations and weather situation (METeorological Aerodrome Report- METAR data) to find the most similar day to day-of-operations. The methodology is based on a random survival forest model, that is a learning method based on a feature selection mechanism to manage missing data or process noisy features [75]. Their study takes data from 2011 to 2015 to offer a decision support tool for only one airport at tactical ATFCM phase. In contrast to one airport, this section is focused on much wider geographical span that contains 70 congested airports. At such a scale, complex dynamics challenge the required computational power. ML methods benefit from huge datasets on a process to learn from it. But required data fusion from different data types (i.e. demand, capacity, weather, etc.) on top of numerous processes at tactical phase is less likely to guide feasibility study at current scope. Instead, metrics such as delay seem to be more instructive.

ML models for delay

The study of air traffic delay (i.e. a performance metric of resilience) is a live topic and in US, the literature is more extended in predicting departure delay (in Europe equivalent to airport delay as part of ATFM delay). For instance, Rebollo and Balakrishnan [76] used the random forest algorithm to predict departure delay with the help of data from National Airspace System (NAS). Their study estimated the network related delay on a certain Origin-Destination (OD) pair. In a similar study by Kim et al. [77], Recurrent Neural Network (RNN) was applied as deep learning method to predict aggregated delay.

In a joint study by SESAR (Europe) and NextGen (USA), Kravaris et al. [78] studied arrival delays in a multi-agent system setting. Three different methods of alternative multi-agent reinforcement learning were implemented. The work was further extended by experimental results to study the significance of methods in a follow up paper [79]. Arrival delay was also addressed as a predictability estimator by Montes et al. [80] in a ML study. Similarly, OD pairs are initially clustered by a classification with density-based clustering algorithm. Then regression models were applied for each cluster to predict delays. However, none of the mentioned methodologies predicted delay on a network level but mostly on specific OD pairs.

In Europe, among unified databases and standard definitions, specific data types such as ATFCM regulations are effective assets to resolve complexity of delay prediction on a network scale. Despite the following advantages of regulation data, they are less explored in this regard;

- Regulations are already classified to fourteen different types, meaning that the need for clustering and a classification problem is already covered by available features of ATFCM regulations;

- ATFCM regulations can be planned both in tactical and pre-tactical phase of ATFM. Therefore, application of ML methods can be extended to tactical phase depending on the model inputs;
- Each regulation is defined for a reference location, which is not an OD pair or a specific route but they are valid for an airspace block (i.e. a traffic volume). In fact the FMP which proposes the regulations to resolve DCB problems is defining the reference location. This aspect alone opens a new opportunity to set a comparative delay prediction study on an agent-based model taking FMPs as agents rather than flights.

Subproblem-I definition

Learning models are investigated from two defined sub-problems on feasibility (I) and performance/prediction quality (II). Subproblem-I investigates the applicability of ML methods in predicting ATFM delay with capacity (ATFCM) regulation data. As candidates of two different families of learning mechanisms, a Sequential Neural Network and a Random Forest Regression were applied to pre-processed data. Generally, neural networks belong to pattern recognition models while random forests are an extension of decision trees in which different features of data are handled as decision points.

In pre-processing some statistical features are calculated on reference locations (i.e. ACC) from capacity regulation types. These calculated features were considered as added features to input data in predicting normalized mean ATFM delay. Both models are coded in Python 3.6.8 environment by Keras library and Scikit-learn module.

Sequential Neural Network (SNN)

Sequential neural networks [81] build high-level features through their successive layers. SNNs are linear stack of layers without any arbitrary graphs of layers such as parallel or branching architectures.

Denoyer and Gallinari [81] denoted the structure of a SNN in comparison to NNs to illustrate the advantage of SNN in using a sequence of transformation functions rather than a global one in neural networks. These models have a Directed Acyclic Graph (DAG) structure defined as follow:

- Each node n is in $\{n_1, ..., n_N\}$, where N is the total number of nodes of the DAG;
- n_1 represents the root node (without any parent node);
- $c_{n,i}$ corresponds to the i^{th} child of node n;
- *leaf(n)* is true if it is a node without children;
- Each node is associated to a particular representation space and act as layers in classical neural networks
    - the dimension of the root node is the dimension of the input layer,

- ○ if leaf(n) = true then the dimension of the leaf nodes equals the output layer;

- Mapping functions ($f_{n,m} \in F$) transforms the input x in the node n to adjacent node of m. The output produced by the model is a sequence of f-transformation applied to the input like in a neural network; and

- Every node is also associated with a selection function, which is a probability function denoted by P_n that assigns a score for each child of node n. This function defines a probability distribution (z) over the children of a given node.

The learning algorithm in case of a gradient decent, tries to minimize the error of expected values from mapping and selection functions. Each chain of transformation functions from the root node to a leaf is denoted by H in Equation 5. This equation evaluates the performance of the SNN architecture (*J*) through the expected value (*E*) for a given θ and γ as of parameters for mapping (*F*) and probability (*p*) functions. Same parameters also serve in learning procedure, that is formulated as an optimization problem on gradients (∇) of output vectors (Equation 6).

$$J(\theta,\gamma) = E_{P(x,H,y)}[\Delta(F(x,H),y)] \tag{5}$$

$$\nabla_{\theta,\gamma}J(\theta,\gamma) = \int \nabla_{\theta,\gamma}\big(P(H|x)\big)\Delta(F(x,H),y))P(x,y)dHdxdy \tag{6}$$

From the same logic the proposed architecture is built with an input layer feeding four hidden layers to converge into the output layer. The proposed SNN architecture allowed experimenting on different activation functions at each layer and for the case of regulation data, Rectified Linear Unit (ReLU) as the activation function for all layers led to better results. The NN is then compiled by Adam [82] optimizer because it converges faster and requires little memory requirements compared to normal SGD (stochastic gradient descent optimizer). Adam is also a suggested algorithm for noisy gradients that is the case with regulation data since delay can variate significantly due to temporary factors such as weather.

One of the reasons that Adam algorithm is efficient with noisy data is its choice of step size. Step size (Δ_t) at each iteration (t) is calculated based on the learning rate (α), exponential moving averages of the gradient (momentum or m_t) and squared gradient (v_t) as given in Equation 7.

$$\Delta_t = \alpha.m_t/\sqrt{v_t} \tag{7}$$

$$|\Delta_t| \leq \begin{cases} \alpha.(1-\beta_{1)})/\sqrt{1-\beta_2}, & (1-\beta_1) > \sqrt{1-\beta_2} \\ \alpha \quad , & (1-\beta_1) \leq \sqrt{1-\beta_2} \end{cases} \tag{8}$$

Step size is bounded by two upper thresholds based on the chosen hyper parameters of Adam (α, β_1 and β_2). Inequality 8, provides these thresholds based on β_1 and β_2 , i.e. exponential decay rates for the moving averages (m_t& v_t). In other words, if the algorithm at a time step reaches a gradient that has been zero at all previous time steps (severe sparsity), Adam continues with larger step size than the specified learning rate. To illustrate, in case of a $\beta_1 = 0{,}9$ & $\beta_2 = 0{,}999$ the step size can jump to 3 times of the specified learning rate:

$$\frac{1-\beta_1}{\sqrt{1-\beta_2}} = \frac{0{,}1}{\sqrt{0{,}001}} = \frac{0.1}{0{,}032} \approx 3.125 \ .$$

Such an advantage is an efficient asset in processing regulation data with important outliers. As an instance Istanbul airport (LTFM) had a closure in Jan. 2022 due to heavy snowfall and collapse of a cargo terminal roof. Airport authorities had problems in estimating the required time for retrieving operations. Because LTFM is a busy hub in EATMN, many flights had to be heavily delayed by network manager (11900 minutes on January 26th in Figure 3-7). Such incidents happen rarely but cannot be ruled out as an outlier specially if one considers that high ATFM delays are recorded for the following day as well (10806 min for 27th).

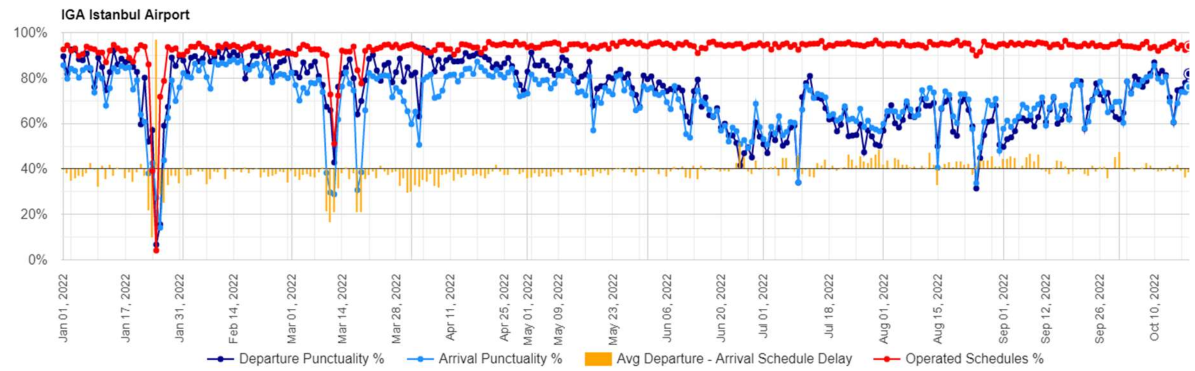


Figure 3-7 Istanbul airport (LTFM) daily delay on 27. Jan.2022 (EUROCONTROL- Aviation intelligence)

Relying on the advantages of Adam over described characteristic of regulation data, the algorithm's performance is tested at three levels of learning rates (0.1, 0.01 and 0.001). Results of experiments on fixed loss function (Mean Squared Error), indicated that a 0.01 rate delivers better results in terms of computational time and calculated Root Mean Squared Error (RMSE).

Random Forest Regression (RFR)

To compare results from SNN (from pattern recognition models), RFR is explored (from decision tree models). Despite stablished ATM procedures across Europe, that assure acceptable level of performance for daily operations (robustness), emergent disruptions pose unforeseen downfalls. A data science perspective translates this aspect as a system that produces low-biased output but with a meaningful variance. Bagging (i.e. bootstrap aggregation) is an ensemble learning technique to reduce variance within a noisy dataset. The name bagging comes from the sampling technique such that for every learning model (also referred as weak learners) a subset of data is independently sampled (a data point can be selected for more than one sample).

As an extension of bagging in ensemble learning, Tim Kam Ho [83] proposed a method for extending decision tree-based classifiers. The method of random forests is based on building multiple de-correlated trees (i.e. weak learners) in a randomly selected feature space. Breiman [84] extended this method in machine learning knowing that the generalization error converges as the number of trees in a forest grows. In his study it is proved that random forests do not overfit and therefore they are promising in predictions (both in classification and regression problems).

This can be statistically expressed by knowing that trees of a forest are identically distributed, therefore every tree in a forest predicts with the same variance (σ^2). The average of B trees has a σ^2/B variance and since the trees are identically distributed (but not necessarily independent) with positive pairwise correlation (ρ), the average variance is calculated by (Equation 9)

$$\overline{Var} = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2. \tag{9}$$

When the forest grows (increased B) the second term approaches zero, i.e. the average variance can be decreased by reducing the correlation among trees (ρ). In a random forest, this is realized by random selection of input variables at each decision tree.

Random Forests (RF), compared to other classification and regression models, such as logistic regression boosting and linear regression, deliver a superior performance [85]. Moreover, it has been applied in prediction of air traffic delays by Rebollo and Balakrishnan [86]. Their work noted the advantages of RF as:

- automatic generation of variable importance,
- low sensitivity to outliers in the training data,
- efficient in cases that number of variables is large compared to number of samples.

With the intended small scale at feasibility check, i.e. limited number of available data points for a given Area Control Center (ACC/FMP) over a year, and with consideration of above-mentioned advantages, RFR is chosen to compare the results from SNN.

Data

At this early stage EUROCONTROL's post-operational data on ATFCM regulations is acquired. A more comprehensive overview of regulation data is provided at section 3.4 (data collection process). As the use-cases, capacity regulations from Langen FMP (EDGG) for the years of 2016 and 2017 are selected. For each regulation twenty different parameters are recorded in the data structure. Since the purpose here is to only consider one FMP, the balance between number of parameters (i.e. data features) and count of data points is less proportionate and needed to be managed. Consequently, the number of parameters is reduced by removal of less contributing features and some are expressed in form of indicators. As a result, the set of features per regulation has been characterized as of Table 3-2. This reduced set does not include any categorical features such as type of regulations. In other words, no label encoding was required in data pre-processing. Also, each feature is normalized to assure balanced learning for the estimators by avoiding any feature dominance. Daily average ATFM delay is predicted value from the first five rows of Table 3-2 that form features per data point in pre-processing (sum and average values are used in (11 and 12) since each FMP can have multiple incidents of a regulation type on a day).

Table 3-2 Data Preparation (Features for feasibility study)

Code	Formula	Description	Type/Class
T1M	$\frac{\mu_{\text{regulated traffic}} * \mu_{\text{Regulation Duration (min)}}}{\mu_{\text{Activation Notice (min)}}}$ (10)	Magnitude score of each FMP based on mean direct impact of regulated flights & Reg duration together with inverse relation with activation notice[a]	Magnitude idicator/input
T2M	$\frac{\sum \text{Regulated traffic} * \sum \text{Regulation Duration (min)}}{\sum \text{Activation Notice (min)}}$ (11)	Cumulative version of T1M	Magnitude idicator/input
T2RA	$\frac{\sum \text{MP Regulated traffic}}{\sum \text{Activation Notice (min)}}$ (12)	Time to recover (Applied) : measuring the direct impact of Most Penalized Regulated Traffic and inverse role of Activation notice per regulation	Time to Recover indicator/input
D	$\mu_{\text{Regulation Duration (min)}}$	Duration of ATFCM Regulation	Time to Recover parameter/input
LT	$\mu_{\text{Activation Notice (min)}}$	Lag-Time (Activation Notice)	Time to Recover parameter/input
ATFM	$\mu_{\text{ATFM delay (min)}}$	ATFM delay (airport and en-route)	predicted variable /output

[a.] Activation Notice or Lag-Time is the time between the publication of a regulation and the time that the regulation becomes effective

Along with main parameters of ATFCM regulations, indicators are proposed in terms of resilience. To avoid multiple features for limited data points, a feature selection is conducted (among indicators and parameters) to focus on dominant features and Table 3-2 only provides the selected set. The balance between input vector and number of data points result in less computational effort and lower risks of overfitting.

The column 'type/class' in Table 3-2, identifies type of inputs the reduces set belongs to. Both magnitude and time-to-recover, represent different aspects of DCB disruption's severity and will contribute to better situational awareness (esp. in case of a preferred weighted input vector). Training and testing datasets are separated by a fixed rate (70% - 30%) for every case of Table 3-3. Among different types of regulations, weather regulations (WX) are intentionally separated in different cases since its nature is from higher levels of uncertainty.

Results

The two methods of RFR and SNN are applied on twelve identical cases in terms of size of input vector and respective year of data. Filtered data provide larger dataset in 2017 compared to 2016 (Table 3-3). Nevertheless, SNN is verified to deliver better predictions in every case (note the RMSE scores). Figure 3-8 and Figure 3-9 provide regression charts for two selected cases (best SNN results). The recorded delay per capacity regulation is fluctuating even after being normalized. Sequential plots in these figures show that SNN is less likely to be affected by short term patterns. However, both RFR and SNN methods proved to be capable of predicting extreme chaotic behavior of data points. This is expected from SNN, because data is shuffled to eliminate effects of such data characteristics. But, RFR (regardless of data shuffling) has higher prediction errors. Moreover, figures reveal that RFR is less efficient in predicting high values of delay (sorted values) while SNN seems robust.

Table 3-3 Comparison of applied methods on different cases

Case ID	WX	Input feature					RFR score (RMSE)	SNN Score (RMSE)	Train points	Test points
		T1M	*T2M*	*T2RA*	*D*	*LT*				
2017-10	Yes				*	*	0.1255	0.0787	281	121
2017-11	Yes			*	*	*	0.1282	0.0762	281	121
2017-12	Yes		*	*	*	*	0.1262	0.0290	281	121
2017-13	Yes	*	*	*	*	*	0.1265	0.0208	281	121
2016-10	Yes				*	*	0.1265	0.0981	206	89
2016-11	Yes			*	*	*	0.1271	0.0907	206	89
2016-12	Yes		*	*	*	*	0.1282	0.0647	206	89
2016-13	Yes	*	*	*	*	*	0.1283	0.0354	206	89
2016-00	No				*	*	0.0929	0.0914	168	72
2016-01	No			*	*	*	0.0942	0.0890	168	72
2016-02	No		*	*	*	*	0.0724	0.0241	168	72
2016-03	No	*	*	*	*	*	0.0718	0.0368	168	72

The feature importance vectors (Table 3-4) reveal the dominance of Lag-Time (LT) in prediction of ATFM delay. But calculated RMSEs suggest to use all features to gain best predictions. From a tactical point of view this dominance implies the importance of LT (time difference between announcement of regulation and start of regulation). Basically, the system is less resilient to sudden disruptions and this is actively reflected in delay as a consequence. In comparison, the duration of a regulation (D in Table 3-4) is less contributing to prediction. This reminds that such cases hint that current systems are not resilient to disruption as a resilient system should suffer more from duration of a disruption. Also, such findings support the claim that regulations are able to represent a network behavior even when being studied on a single FMP (here EDGG).

Table 3-4 Feature importance (2017 cases)

Case ID	T1M	T2M	T2RA	D	LT
2017-10	-	-	-	0.0654	0.9345
2017-11	-	-	0.0859	0.0592	0.8548
2017-12	-	0.0618	0.0548	0.0413	0.8419
2017-13	0.0037	0.0603	0.0516	0.0394	0.8448

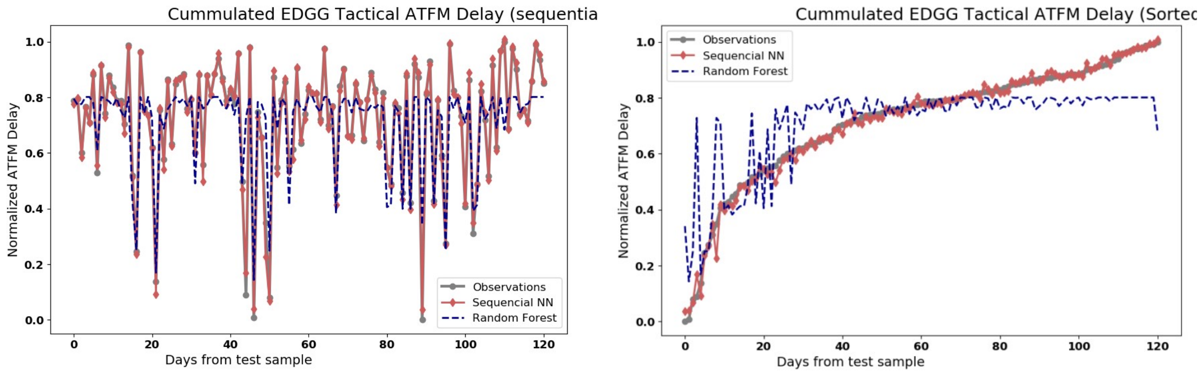


Figure 3-8 Regression chart (case 2017-13) best performance achieved by SNN among all cases on 2017 (including weather regulations).

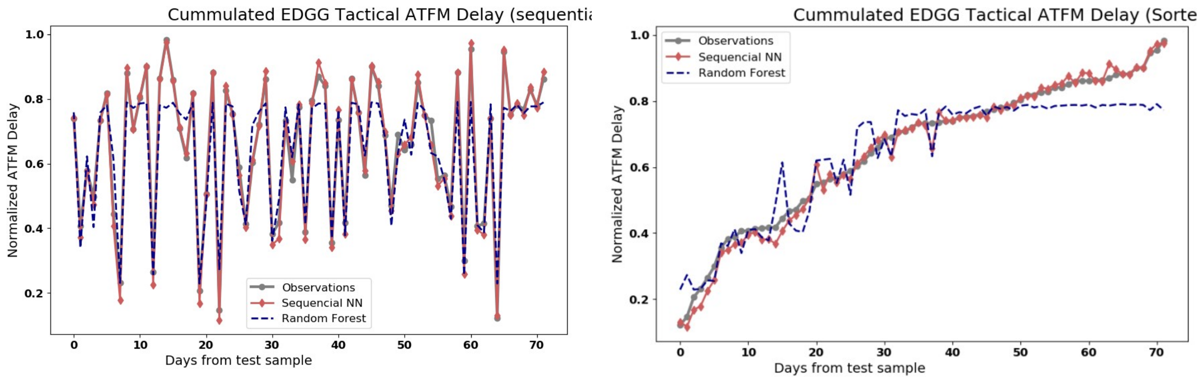
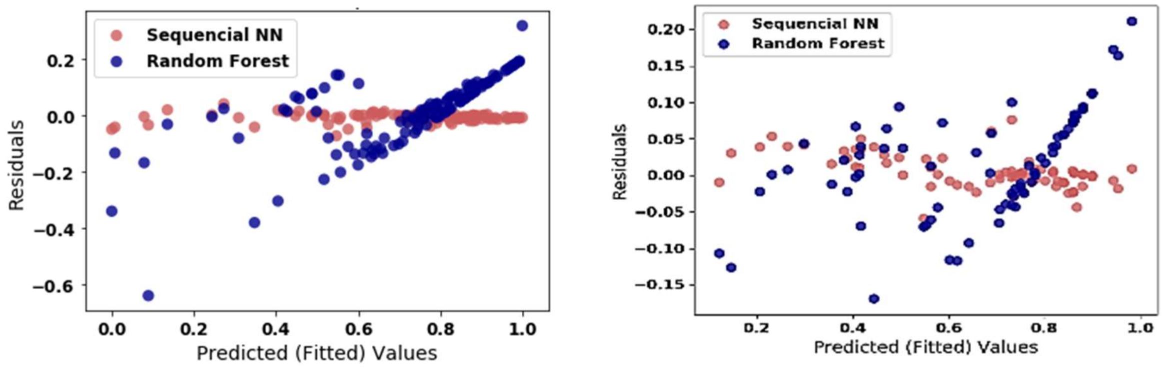


Figure 3-9 Regression chart (case 2016-02) best performance achieved by SNN among all cases on 2016 (excluding weather regulations).

Moreover, the extraction of weather induced regulations is observed not to be much effective for SNN (compared to RFR); probably due to resulted reduction of training sets. But in absence of weather induced regulations, both RFR and SNN tend to response better to more input features. With weather induced regulations, SNN reaches better precision (esp. for 2016 cases) by considering both magnitude indicators (T1M & T2M). Despite similarities of these two indicators, RMSE improved significantly by adding T1M as an input. In contrast RFR delivers same quality of predictions which is a reminder of fundamental differences of pattern recognition models against decision trees. This difference is also evident of better performance of SNN for 2017 cases with more data points (compared to 2016) while RFR tend to be neutral against size of training and testing sets.

Better performance of SNN is also illustrated by residual plots in Figure 3-10. Plotted predicted values against residuals show a smooth dispersion for SNN compared to RFR. This figure also implies that RFR suffers from increasing residuals for estimating bigger values of delay. Experiments on different cases also justified that reduced feature space is proportionate to size of data set since the achieved improvement after including forth feature is minimal.



2017-13 case (included weather regulations) 2016-02 case (excluded weather regulations)

Figure 3-10 Residual plots, better performance of SNN compared to RFR.

Conclusion

In subproblem-I, different approaches in estimating the delay by some studies for US and Europe were briefly reviewed. Two different learning techniques are tested to check the feasibility of learning methods in delay prediction, on a controlled database (only EDGG). RFR and SNN are chosen since each represent different classes of learning methodologies (decision tree vs. pattern recognition). In contrast to previous studies that select OD pairs, experimenting on 24 cases verified that capacity regulations not only can be used to predict delay but feature importance values (Table 3-4) reveal that regulations represent same network behavioral patterns even at smaller geographical span. Moreover, these preliminary set of learning experiments [35], implied firstly the possibility of delay prediction without modeling causal relationships, and secondly SNN's reliability that grows as the number of daily regulations increase.

3.3.2 Customized learning model

In previous section, the feasibility of both learning models and use of regulation data has been studied by preliminary experiments. This section invests more on extracting features of regulation data at first and then seeks different learning approaches to find a baseline model. Lastly, based on the data input and performance of investigated models a customized learning architecture is designed in response to second aspect (prediction quality) of third objective.

> *Objective 3: In order to revive the network from emergent disruptions, investigate the required level of decomposition for corrective measures.*

In subproblem-I, regulations proved to represent network behavioral patterns even at a geographically smaller scale compared to EATMN. However, compared to other tactical DCB solutions (such as cherry picking) capacity regulations represent largescale corrective measures. Initial findings such as importance of lag-time in learning process, motivated the study to invest more on regulations toward resilience at ATFM subsystems (EATMN has 8 subsystems [25]). Getting an overview of the most relevant procedures at ATFM is an active research topic in the literature [87]. Therefore, regulation data are set to be investigated to capture dynamics of network behaviors rather than modeling the entire complex ATFM.

Since the 3rd objective is focused on emergent disruptions, prediction of delay and delayed traffic are desired. Both indicators are categorized as performance metrics of resilience (and not attribute metrics). Delay (ATFM delay) has a dedicated line of research. For instance, Ivanov et al. [88] in an effort to resolve the en-route DCB problem, used a layered mixed-integer optimization model to minimize delay across Europe. From an airline perspective, their study considered delay propagation despite flight schedule buffers. Optimization techniques are widely implemented for delay. Various techniques such as multi-objective optimization [89], integer programming [90], and stochastic integer programming [91] are explored under the category of delay assignment. There are also studies to minimize (ground) delay by alternatives such as airborne delay [92]. Likewise, reducing cruise speed is proved to reduce ground delay by up to 15% [93].

Such studies are focused on minimizing the ATFM and ground delays which is raised as Flow Management Problem (FMP) by Odoni [94] in 1986. On a network level, cost benefits in Europe are measured in 2007 [95] to be 80 million euros. From 2007 (793 million passengers [96]) to 2018 (almost 1 billion passengers [97]) despite observed ATC productivity gains [98], ATFM delay reached 25 million minutes with a substantial yearly increase (+64.5%). Statistics of 2018 demonstrated that a network with optimized delay cannot necessarily be considered as a resilient network since weather, staff shortages and ETFMS outage conspired to take delays to the extreme. Alternatively, this study is aimed at predicting (not optimizing) delay and delayed traffic to offer more situational awareness against emergent disruptions.

With regard to aviation advantage in data availability compared to other means of transportation, ML has been used for delay prediction [99, 100, 101]. Some studies designed learning architectures that combine different methods for delay prediction. For instance, Gui et al. [102] merged Long Short-Term Memory (LSTM) and decision trees to enable their approach in integrating different datasets (ADS-B[1], weather, airport info). They reached a 90.2% accuracy by a random forest-based model (for a binary prediction). In a similar study [103], LSTM and Support Vector Regression (SVR) were used to calculate the air traffic flow instead of delay. They concluded that the LSTM architecture outperforms SVR, especially in case of abnormal traffic flows. Their methodology was applied on selected air routes (OD pairs).

However, regulation data have not been investigated specially in understanding emergent behavior of EATMN. In subproblem-I the feasibility of a learning model based on regulation

[1] Automatic Dependent Surveillance-Broadcast (ADS-B)

data, encourage this study to pursue more complex models that can learn better from different aspects of regulations.

Subproblem-II definition

Based on results of subproblem-I, next steps are revised to be twofold: firstly, to exploit regulations as a rich datatype that encodes multiple interactions between subsystems of network, and secondly, to predict the network performance in presence of large-scale capacity regulations with a learning method. The proposed model is required to be developed such that it can predict two network indicators: total ATFM delay and the number of delayed flights (indicators that represent magnitude of disruption).

More specifically subproblem-II is intended to answer the following aspects:

1. To handle the modeling challenge of network; that is addressed by supervised learning to avoid complexities of interaction in network's subsystems;
2. To select the most relevant data; that is answered by capacity regulations since each record encodes the result of different coordinated planning processes to deal with a DCB issue at the day of operations (*tactical phase*). In this phase, network resiliency is highly vulnerable to disruptions;
3. To capture the spatiotemporal dimension of network dynamics, that is managed by a proposed deep convolutional neural network architecture.

Regulations are mainly studied in DCB and ATFM optimization approaches [104, 105, 106]. Data on regulations are available both at post-operational and tactical phases. Therefore, different supervised models are tested next to: a) provide a baseline to assess quality of results, and b) to select the best potential model for further development. This part of thesis is fostered through a master thesis [107] and is separately published [62]. Primary results are concluded with selection of random forest as the baseline since its accuracy is directly linked to forest size (i.e. accuracy can be increased even up to overfitting). Furthermore, neural networks is chosen because of its superior performance and intrinsic flexibility in learning from regulations (reminding the results of subproblem-I). Such an approach guided the study to Convolutional Neural Networks (CNN) which is further improved to propose a deep CNN with higher prediction quality. The aforementioned steps (data preparation, setting a Random Forest (RF) model as the baseline, and the design of the proposed deep CNN) are described in more details in the following sections.

Subproblem-II variables

Subproblem-I provided a better picture on choosing the right data range. The annual growth of delay and regulation counts in presence of persistent seasonal patterns reminds that model training set shall be limited to most recent years. 2018 stands out with the highest number of regulations and the highest amount of delay. However, data from 2018 and 2017 are combined to construct a dataset with adequate data points for train/test sets (knowing that supervised learning methods account for generalization of trained model). Every tested model is using this dataset to enable comparison of the results explored models. Section 3.4, describes the data collection approach, data structure, and its characteristics in detail.

Input features

In subproblem-I (feasibility study), normalized mean ATFM delay was predicted. But the main intention is to predict daily target values at day of operation (d_{op}). Such a predictability matters most to network state definition based on pre-tactical regulations; in which the learning algorithm tries to learn the dynamics of tactical phase as a black box. The described methodology so far worked on the claim that regulations encapsulate these dynamics and NMIR is a EUROCONTROL database that offers a post operational dataset on regulations. Tactical regulations are published in form of ANM messages that has different structure but with some common attributes as of NMIR.

From NMIR, a cut of desired data for 2018 and 2017 is acquired (let N be the number of days in this cut), then regulations for each day are filtered out to only those that are being activated before 06:00 UTC. Such a list of regulations is taken as the pre-tactical (d_{op} -1) regulations. Pre-tactical regulations for each d_{op}, are selected by filtering the attribute of "*regulation activation date*". Although the resolution of the dataset allows to break down to hours, it has been refrained since with a coarse resolution, the pattern may disappear. In a final step each day is reconstructed with daily aggregated attributes, specified weekdays and respective AIRAC cycles (the input vector and its features for a given d_{op}):

- N_{Reg}**:** Number of active regulations for each d_{op} (i.e. regulations that their start time is from pre-tactical phase (d_{op}-1) up until 6:00 UTC in the tactical phase);
- $\overline{D}_{Reg}$**:** Average duration of all activated regulations at d_{op};
- N_{Tact}**:** Tactical regulations count (i.e. number of regulations with a start time from 0:00 up until 6:00 UTC);
- N_{ACC}**:** Represents number of ACC with regulations from pre-tactical phase up until 6:00 UTC at d_{op};
- $Rtype_i$**:** A family of 14 features (that each correspond to a regulation type as of Table A-2 in annex). $Rtype_i$ is the total number of type i regulations across the network (active regulations from pre-tactical up until 6:00 UTC at d_{op});
- ***AIRAC:*** The AIRAC cycle (1 to 13) to which each d_{op} belongs. This feature is mapped to NMIR data from a reference table (Table B-1);

In context of learning models, the AIRAC cycle should be considered as categorical data. This is because AIRAC13 is not greater than AIRAC1, or vice versa in any sense. Therefore, this feature has to be encoded such that learning model can use it without giving numerical significance to the AIRAC number. The one-hot encoding of Scikit-learn [108] pre-processing module is used for this purpose. With such an encoding, any

AIRAC is represented by a binary vector of length 13 and only one of the items in the vector will have a binary high.

- ***Weekday:*** Similar to AIRAC, the seven weekdays are one-hot encoded resulting in a binary vector of length seven. This feature is added with consideration of a study from Sun et al. [70], that captured a weekday variation in the European air transportation network connectivity.

Predicted values

The total daily ATFM delay and delayed traffic (also referred as Most Penalized (MP) delayed traffic) are considered as the predicted values (i.e. labels) for the supervised learning model. These values constitute the volume (delay) and extent (count of delayed flights) of network disruption. More specifically target values for a given d_{op} are total values at the end of the day (24:00 UTC) as:

- ***Delay (min):*** Total daily ATFM delay in the network; and
- ***Delayed Traffic (flights):*** total number of delayed flights or daily MP delayed traffic. A flight can be subject to more than one regulation on its route and in such cases, only the most penalizing regulation is considered to impose a delay, i.e. other regulations on the flight route are ignored.

Train-test split

In learning models, the size of the train and test sets needs to be proportionate since a relatively large training set would increase the risk of overfitting while a small training set challenges the generalization of the model and the prediction error will rise, especially in absence of evident patterns in a scattered dataset. However, such a choice is not a point of concern for this study, because of the persistent seasonal trend of regulations. The acquired dataset (2017 & 2018) includes a total of 730 days (with more than 118 thousand regulations) and is splitted by a 70-30 ratio for train and test sets.

Baseline: supervised learning models

ML is a suitable approach for dynamics of EATMN as a system with complex non-linear structures that data acquisition is much more convinient than modeling the system. In comparison to deterministic optimization models, ML applications are mainly about generalization. These models consist of a combination of optimization cores and statistical analysis in their algorithms.

According to the SESAR publications, the application of ML has gained more interest since 2017. Among different supervised learning approaches, applications of Neural Networks cover more topics of ATM [109, 110]. NNs were used in different aspect, for instance to predict the flight trajectories [111] and flight levels [112]. Along with NNs, decision tree based models such as Gradient Boost Machines (GBM) are used to predict the runway occupancy count for a single airport [113]. Gradient Boost and Recurrent Neural Network (RNN) are also addressed in predicting take-off times [114].

RF (another example of decision tree models) has also been used for ATM topics such as predicting the flight efficiency [115]. However, the prediction of daily delay and delayed traffic at network level by learning techniques is rather remained as a gap. The closest work is in [116], where a simple decision tree model was used to find the delay variations in a small group of sectors instead of whole European airspace.

In order to foster the modeling approach and to select a baseline model, four different supervised learning methods are applied on regulation data with the described features. The selected baseline model will serve to assess the efficiency of final proposed model. These explored methods are: RF, Linear Regression (LR), Support Vector Regression (SVR), and Neural Network (NN). As these models are intended for comparison, similar performance metrics are required.

Performance metrics

For regression problems, standard metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) are more common. However, the extended range of variations for delay and delayed traffic calls for a customized metric. In selected dataset, daily delay varies from 1 958 to 327 795 minutes. Similarly, delayed traffic can be as small as 117 flights and reach up to 10 812 flights. These metrics treat the deviations equally, but ignore the operational understanding of data. For example, a deviation of 50 000 minutes is not acceptable for an actual target value of 25 000 minutes, but is considered as a decent error when the target value is 350 000. Therefore, it is risky to rely on such metrics to evaluate the overall performance of the model. This aspect is answered by following two solutions:

- Mean Absolute Percentage Error (MAPE)
  Similar to MAE, this metric is the average value of errors that is expressed in percentages. Suppose y_i is the actual value for which the prediction is $\hat{y}_i$, then the MAPE is calculated as:

$$MAPE = \frac{100}{N} \sum_{N} \frac{|y_i - \hat{y}_i|}{\hat{y}_i} \ . \tag{13}$$

- Evaluation per delay category
  Delay records of 2017 and 2018 can be ordered to three categories of: low (first quartile), moderate (2^{nd} and 3^{rd} quarlile), and high (last quartile) as given in Table 3-5. To achieve a better insight on performance quality, evaluations for each model is calculated for per category and overall values.

Table 3-5 Categorization ranges for model performance evaluation

Category	Delay (min)	Delayed traffic (flights)
Low	[0, 20 000)	[0, 1 250)
Moderate	[20 000, 80 000]	[1 250, 4 650]
High	(80 000, ∞)	(4 650, ∞)

Baseline models

The baseline model has been selected after comparison of individually tuned and evaluated explored models (LR, SVR, RF and a NN architecture). Although trained individually, simultaneous prediction of two target values leads to poor performance since delay and traffic values have different ranges. If not tailored properly, this aspect imposes optimization challenges, because the model through the general loss function mostly in a backward propagation, tries to minimize the calculated prediction error. A multi-variate prediction misleads the optimization model in favor of one of the predicted values. The particular reason is that the correlation between delay and traffic is not intended to be provided to the learning model, and the purpose here is to minimize the prediction error instead of understanding the correlation.

The following options are considered to control the different scale of delay and traffic:

- Scaling the predicted values, with the cost of losing the operational understanding of both delay and delayed traffic. Specially during an exploratory phase, it makes the results to be less intuitive and more theoretical;
- weighted loss function, with weights that are required to be either pre-assigned or learned. In absence of solid correlation, if these weights are set to be learned, it leads to excessive complexity and more data points will be required to control the relative error;
- or training separate models for each variable. The key advantage of this option is that model can detect and learn different dependencies on input vector (features). For instance, it might be the case that a specific type of regulation leads to more delayed flights while another type is more persistent and cause longer delays. Although the need for more datapoints is less crucial with this option, but the cost is higher computational effort.

The 3rd option serves the best to the limitation on dataset size and the interest of the methodology, therefore for each learning method, two independent models are being trained and tested. However, identical performance metrics and model design (e.g. cost functions or activation functions) are used for both predicted values.

Linear Regression (LR)

It is a basic prediction method to estimate a linear function of independent variables. In general regression models take a response variable (Y) and search for an approximation function on predictor variables (X). Approximation function can have different forms; a linear regression assumes the function to be linear as in:

$$Y = \alpha_0 + \alpha_1 X + \varepsilon \ , \tag{14}$$

where α_0 and α_1 are constant coefficients or weights, and ε is a random disturbance or error. In a LR learning model, the gradient descent optimization technique is typically used to find the optimal coefficients that minimizes the error.

LR is used here to check a model with two key assumptions: linearity and normal distribution of prediction error. Though it is expected that regulation features and the target values are less likely to be in a linear relation but the model also considers ε as an independent random variable with standard normal distribution [117]. The data are prepared, scaled and splitted as described earlier and using the Scikit-learn library the model is trained on 511 days. Table 3-6 (delay) and Table B-4 (delayed traffic) provide the model performance on train and test sets.

Table 3-6 Performance of applied LR to predict delay.

Category	Train				Test			
	Days	MAPE[a]	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	127	77.59	-5.81	8 744	55	92.51	-5.54	8 562
Nominal	261	34.85	-0.14	13 753	111	36.1	-0.33	14 834
High	123	18.02	0.56	24 329	53	22.53	0.26	27 884
Overall	511	41.47	0.82	15 054	219	46.98	0.77	16 417

[a] in percentage, [b] minutes.

Similar performance in both train and test sets, demonstrate a stable model but in general the performance is not convenient enough to account for a linear relationship between variables. But the benefit of MAPE over other metrics is evident in Table 3-6. Even when categories are ignored (overall), it still shows the poor quality of predictions, in contrast R^2 (goodness-of-fit) indicate a relatively good prediction (i.e. 0.82 in training). Moreover, smaller absolute errors in low category led to smaller MAE values exposing the risk of being interpreted as a better prediction quality. This pattern which is observed also in Table B-4 is emerged from the use of absolute error as the cost function. In other words, errors are penalized similarly in different ranges of predicted values that lead to worst MAE values to be at high category.

Support Vector Regression (SVR)

Results from LR actively demonstrated that there is no linearity. Consequently, SVR (a derivative of Support Vector Machines (SVMs) is explored next to assess the non-linearity. SVRs are recommended for small and medium size datasets in presence of outliers. Assisted by tunable hyper parameters, they search for the best prediction within symmetric thresholds.

Generally, SVR tries to find a function (hyperplane) that is surrounded by an error tube. The idea of an error tube formulates the optimization problem to search for the flattest tube that best approximates the hyperplane that contains most of the training data points (refer to chapter 4 of [118]).

SVR hyper parameters are: kernel, C, epsilon, and gamma as briefly described below:

- ***Epsilon:*** defines the size of the tube in which the training loss function is equal to zero. Value of epsilon controls the generalization of the model;
- ***C:*** the regularization parameter that defines the extent to which the outliers are to be penalized in fitting the model. A large penalization on outliers may result in over-fitting and poor generalization;
- ***Kernel:*** a transformation function (a kernel) is used instead of a hyperplane in case of an assumed nonlinearity between input features and the response variable. A kernel can be either precomputed or linear, polynomial, sigmoid, and Radial Basis Function (RBF). RBF is basically an exponential function;
- ***Gamma:*** is the assigned coefficient in case of a polynomial or exponential (RBF) kernel. It is a positive value that defines the influence of each training sample (i.e. curvature weight of the decision boundary). Higher values of gamma lead to a more complex kernel and increase chances of over-fitting.

During the training phase, a grid search is performed to tune hyper parameters (in Scikit-learn library). The following values constitute different combinations for the grid search (for both delay and delayed traffic):

- Epsilon: 0.1, 0.5, 1.5, 2, 2.5;
- C: 1, 100, 5 000, 8 000 and 10 000;
- Kernel: 'Linear', 'Poly' and 'RBF';
- Gamma: 0.01, 0.1, 1, 'auto'.

Despite same reference sets, using separate models for delay and delayed traffic lead to different hyperparameters in grid search (Table 3-7). For instance, the significance of nonlinearity for delay compared to delayed traffic is once more identified by the selection of a polynomial kernel as the best kernel. In fact, the nonlinearity of delayed traffic is so complex that SVR performs better with a linear kernel and maximum errors (C=10 000) compared to either a polynomial or RBF case.

Table 3-7 Best hyper-parameters for SVR.

Response value	Epsilon	C	Kernel	Gamma
Delay	2.5	5 000	Poly	1
Delayed traffic	2	10 000	Linear	0.1

Performance metrics of tuned SVR model (Table 3-8 and B-5) confirms similar pattern as observed by LR along the lines of quality degradation over low category. In general, SVR outperforms LR but the model seems to be rather overfitted for delay compared its homogeneous behavior for delayed traffic. The values for hyperparameters, i.e. lower regularization parameter (C), polynomial kernel and bigger gamma provide a plausible reason for a more complex model with overfitting.

Table 3-8 Performance of applied SVR to predict delay.

Category	Train				Test			
	Day	MAPE[a]	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	127	30.87	-0.06	2 553	55	71.64	-4.0	7 361
Nominal	261	11.31	0.7	4 759	111	29.87	0.13	12 068
High	123	12.42	0.57	18 115	53	23.41	0.17	31 753
Overall	511	16.44	0.88	7 426	219	38.8	0.78	15 649

[a] in percentage, [b] minutes.

Despite better performance of SVR for delayed traffic (Table B-5: MAPE metric is 26.46% for overall category in test set), the purpose of finding an approach with acceptable performance for both delay and delayed traffic is yet to be fulfilled.

Random Forest Regression (RF)

As explained before, RF is a typical decision tree learning method that employs multiple learners (weak learners) to generate a weighted prediction (strong learner) as the final result. Ensemble learning is known to provide better generalization ability and more accurate prediction [118].

Random Forests provides an average predicted value based on a set of noisy predictors (trees) with relatively low bias. Generally, such a model reduces the variance of predictions and can fit perfectly on training set by either unlimited depth (feature exploitation), or unconstrained minimum samples for each split. Similar to other decision tree methods, RF recursively selects a variable (split-point) to grow a forest of trees. However, in absence of tuned hyperparameters, performance on test set is less likely to be satisfactory. The important hyper-parameters and their significance are explained below:

- ***Number of trees***: defines the number of estimators in a forest. Number of estimators is in direct relation with generalization;

- ***Maximum depth:*** controls the extent of splitting at each tree. Smaller depth avoids chances of overfitting since it leads to low bias;
- ***Maximum features***: sets the maximum number of features in splitting, because selecting only a subset of features for building a regression tree minimizes the over fitting risk;
- ***Bootstrap:*** allows creating random sub-samples of the main dataset with replacement (same value can be used multiple times). It is a powerful statistical technique for estimating a quantity from a data sample. RF is a bootstrap aggregation (bagging) algorithm. In Scikit-learn library, this is a boolean variable that if set to false, the whole dataset is used to build each tree without resampling.

RF is less sensitive to type of features since it aggregates the output from a number of weak estimators. This understanding helps to balance the size of input vector and categorial features such as AIRAC cycle and weekday are not required to be encoded. Based on such a reduced input vector, a grid search is performed to find the best combination of hyperparameters from:

- Number of trees: 50, 70, 100, 130;
- Max_features: 2, 4, 6, 8, 10;
- Max_depth: 5, 10, 20, 25, 50;
- bootstrap: True, False.

First impression from Table 3-9 is that Maximum depth is selected to be at highest. Higher depth lead to better prediction, however such a choice exploited the training set (Table B-2 & B-3 shows the overfitted model). Therefore, lower values for this parameter has been separately tested and a value of 12 seems to avoid overfitting and delivers the best results on test set (values below 12 are underfitted models).

Table 3-9 Best hyper-parameters for RF.

Response value	Tree counts	Max_features	Max_depth	Bootstrap
Delay	70	6	50	False
Delayed traffic	70	8	50	False

Table 3-10 and Table B-6 provide the performance metrics (max_depth is 12) for delay and delayed traffic, respectively. The results imply that tuned RF model outperforms previous models (SVR and LR).

Table 3-10 Performance of applied RF to predict delay.

Category	Train				Test			
	Day	MAPE[a]	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	127	6.76	0.95	746	55	74.15	-3.49	7 741
Nominal	261	1.96	1.0	776	111	26.2	0.24	10 612
High	123	0.68	1.0	792	53	17.65	0.47	24 263
Overall	511	2.85	1.0	772	219	36.18	0.85	13 195

[a] in percentage, [b] minutes.

Table 3-10 demonstrates that RF as a decision-tree approach can fit on training set as if the model is overfitted but the metrics on the test set assures that the model is not overfitted. provide best performance so far baseline because of its tree-based approach. This was already anticipated by results of feasibility study (subproblem-I) that verified the potentials of a tuned RF model. Those results also manifested the nonlinearity and superior performance of a sequential NN against random forest regression [35]. Therefore, as the last model for this phase, same data structure is fed to a candidate NN.

Neural Networks (NNs)

In general, a neural network learns in a hierarchical order and their structure involves multiple levels of abstraction for knowledge representation. NNs accumulate propagated information through higher levels in a sequential order such that learning at each layer is based on statistical learning procedures at the previous layers (refer to chapter 7 of [118]).

In current discussion, it is clear that the prediction problem at hand is rather nonlinear which NNs basically manage by activation functions. A network can have different activation functions at each layer in comparison to random forests that trees are identically distributed. Furthermore, prediction errors are evaluated by cost (loss) function and through iterations, optimization function pushes the network toward minimizing the errors. Each iteration is performed on batches that are subsets of the training set. Once a batch is processed, each node of every layer gets a new weight (learning). An epoch is completed when all the batches of a training set are fed as inputs.

To implement a fully connected sequential NN with three hidden layers, Keras [119] (an open source deep learning library of python) is used. The input layer has 38 neurons that matches the length of input vector (features). The three hidden layers converge from 100 to 50, and 25 neurons. A single neuron at the output predicts the delay or the delayed traffic for the two separate models. Each layer uses Rectified Linear Unit (ReLU) as the activation function. The model is trained with MAE cost function and Adam optimizer for 500 epochs with a batch size of 30.

Considering the model performance in Table 3-11 (and Table B-7), the tested architecture is not considered to be overfitted, since the metrics report similar quality of prediction for train and test sets. Over the test set, even such a basic network remarkably performs with almost same quality as in RF. But compared to RF, NN has not exploited the training set and offers a consistent model.

Table 3-11 Performance of applied NN to predict delay.

Category	Train				Test			
	Days	**MAPE**[a]	**R^2**	**MAE**[b]	**Days**	**MAPE**	**R^2**	**MAE**
Low	127	54.96	-2.52	5 621	55	59.58	-2.98	5 417
Nominal	261	25.33	0.26	10 687	111	30.33	0.08	12 334
High	123	21.01	0.36	28 693	53	23.94	0.18	30 975
Overall	511	31.65	0.81	13 762	219	36.13	0.79	15 108

[a] in percentage, [b] minutes.

Table 3-12 sums up all experiments that has been stated to this point. Performance of four different regression approaches is expressed by MAPE for both delay and delayed traffic. It is evident that all models had challenges in predicting lower category and precision improves for bigger values. It is intuitive that higher target values benefit from a richer input vector because of more regulations that are encapsulated in daily features. Nonlinear models such as NN outperform linear models and Figure 3-11 confirms this claim by visualized dispersion of predictions. Moreover, the optimization of RF hyperparameters not only led to higher precision but also the scatter plot shows a steady narrow prediction error for both delay and delayed traffic.

Table 3-12 Performance of explored learning models over test set.

Category	Delay [a]				Delayed traffic [a]			
	LR	**SVR**	**RF**	**NN**	**LR**	**SVR**	**RF**	**NN**
Low	92.51	71.64	74.15	59.58	64.28	54.79	55.95	47.95
Nominal	36.1	29.87	26.2	30.33	22.0	20.76	17.31	23.13
High	22.53	23.41	17.65	23.94	9.9	11.1	11.64	9.75
Overall	46.98	38.8	36.18	36.13	29.05	26.46	25.09	25.73

[a] measured by MAPE metric.

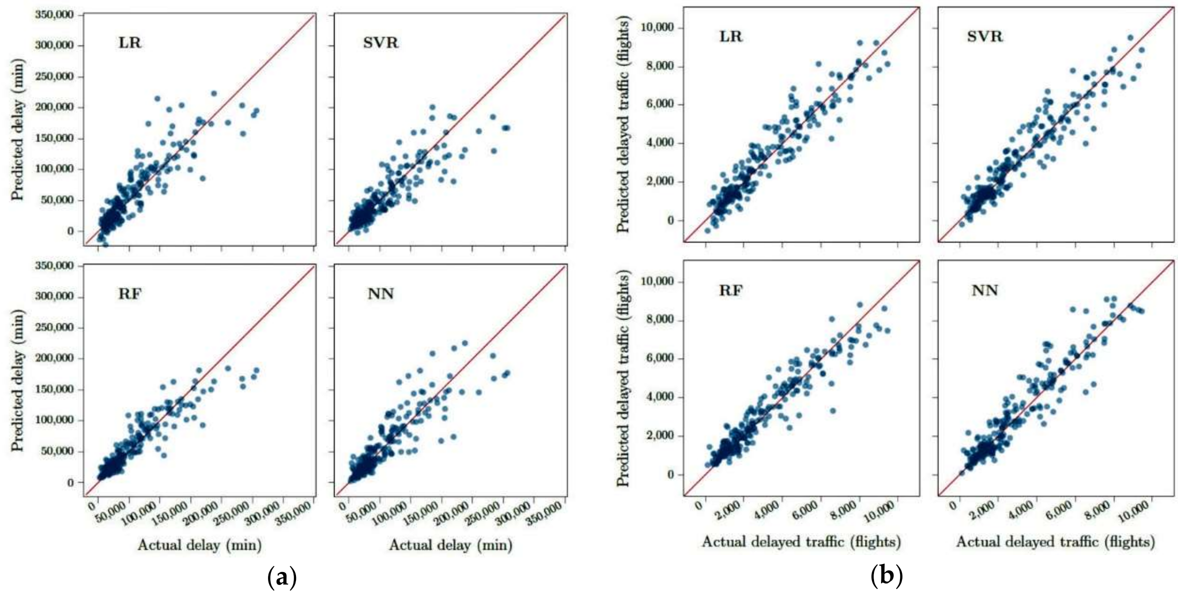


Figure 3-11 Scatter plots for prediction quality of learning models on test set. (a) delay, (b) delayed traffic. Explored models perform better on delayed traffic due to its smaller range compared to delay. RF (Random Forest) provides minimum errors with symmetrical low dispersion.

Clearly the result show that RF and NN deliver higher quality of prediction. Apart from poor performance, LR is a linear model and SVR has less flexibility to be evolved to a more complex architecture. Therefore, the study is guided to invest more on NNs respecting a diverse range of architectures while RF is chosen as the baseline model knowing that with tuned hyperparameters it gets closer to overfitting boundaries.

Proposed Deep Convolutional Neural Network (DCNN)

In the previous section, the pre-processing on data were arranged so that it provides an aggregation of data features that can be fed to different models. The aggregation of data ignores the spatiotemporal features to a great extent. But, the traffic flows connect separate ACCs across Europe and regulated traffic volumes may lead to secondary effects on other traffic volumes in adjacent airspaces. The propagation of this consequential impact is known as *network effect* in ATFCM [116]. Such secondary effects can be perceived better by CNNs since they are designed to capture different features of data (spatial ant temporal) through convolutional layers.

CNNs are mainly employed for classification problems, especially in image processing, where learning is about spatial characters (as of curves or sharp edges). Relatively few studies try to extract spatiotemporal features by CNN. For instance, in intelligent transportation systems, Bilong et al. [120] proposed a deep 3-dimensional CNN to extract the spatial and temporal correlations. They evaluated the model with a database on taxi trajectories in New York city. Similarly, a recurrent CNN is developed in a study by Wang et al. [121] to predict the traffic speed and congestion. Their model integrated the spatiotemporal traffic speeds of contiguous road segments as the input matrix.

The architecture of such deep networks is identical in each study because deep networks have higher degrees of freedom compared to other learning methods. In fact, apart from hyper-parameters of CNN such as kernel size and stride, the model design can also be different in selection of activation functions, optimization methods, etc. Before structuring a Deep CNN to consider network effect by extracting deep characteristics of regulation data, one should understand CNNs in general.

Why CNN? The obstacles of modeling the European ATM network has been addressed in previous chapters. This thesis seeks intuition against deduction to learn the behavior of EATMN as a measure to deal with dynamic complexities of intertwined ATFM operations. In search of intuitive inference technique for prediction, the proposed methodology invests on representation of regulation data as daily records. In fact, each daily data point that encapsulates regulations can be considered like a daily image captured by a traffic camera at an intersection. In this analogy a datapoint can have 14 different type of regulations just like cars with different colors in an RGB image. Similarly, different ACCs can be mapped as of different brands of cars, or large-scale weather conditions can be considered as presence of pedestrians at intersection, etc.

Majority of applications of CNNs are dedicated to image processing applications with highlights such as AlexNet (2012), VGGNet (2014), GoogLeNet (or inception V1, 2014). Success of CNNs compare to SNNs are attributed to key advantages of parameter sharing and sparse connections.

- Parameter sharing is the benefit of filters in convolutional layers. The assigned value to each node is calculated upon a neighborhood (depending on the filer size) and distinct values of the filter. Each convolutional layer has one filter (i.e. kernel) to produce all nodes of the next layer. In an image processing task this allows to consider features bigger than one pixel (e.g. one filter for curves and another for sharp edges).

- Sparsity of connections means that not every neuron is connected to all neurons of the previous layer. In fact, instead of having a dense layer (each node is connected to all previous nodes) a filter is applied to a specific neighborhood (extracting regional patterns rather than processing the whole input at once). In other words. If the input is an image of 100x100 pixels, then the input vector has 10 thousand neurons and a dense layer will have 100 million (10^4x10^4) weighted connections (learning parameters). In comparison a convolutional layer uses same filter for all nodes of the next layer. Input of size 100x100 and a 25x25 filter creates an output of only 76x76 (assuming that hyperparameters stride is 1 and pad is 0).

Moreover, CNN as a neural network follows principally same algebra on loss function (forward pass) and back propagation (updating weights). Each filter at a convolution layer acts as a dot product of a filter (w) and a subset of input vector (x), plus the bias term (Equation 15). The difference for CNN at this level is that output of each convolution layer is produced by filter (f) and generated subsets of the input vector (function g in Equation 16).

$$Z = w^T x + b \ , \tag{15}$$

$$Z = f \otimes \sum^{input\ Vector} g \ , \tag{16}$$

f: applied filter at each convolutional layer,
g: the function that produces a chunk of input vector (signal).

In back-propagation of the loss function (L), same considerations are regarded in calculation of the derivatives. In convolution layers derivatives of the loss function (L) from the previous layer are achieved based on Equation 17 (Back-propagation).

$$\begin{Bmatrix} \frac{\partial L}{\partial f} = \frac{\partial L}{\partial Z} * \frac{\partial Z}{\partial f} \\ \frac{\partial L}{\partial g} = \frac{\partial L}{\partial Z} * \frac{\partial Z}{\partial g} \end{Bmatrix}. \tag{17}$$

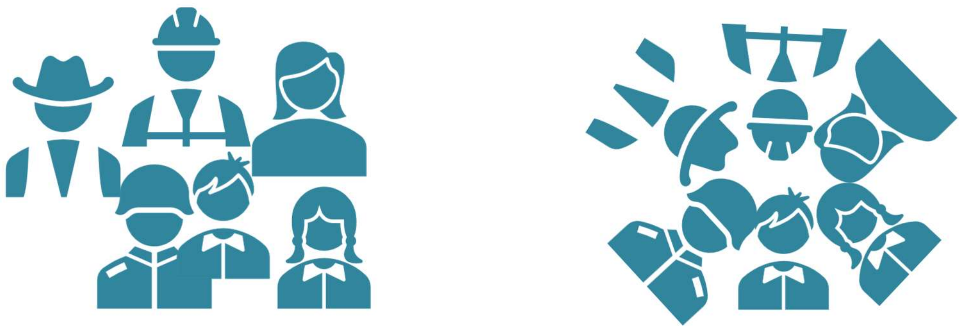
The loss from previous layer that needs to be backpropagated to other layers is denoted by ($\frac{\partial L}{\partial Z}$) since input of each layer is the output of the previous layer. Filter's group of partial derivatives ($\frac{\partial L}{\partial f}$) is required for updating (learn) the filter values according to the learning rate (Equation 18).

$$F_i = F_{i-1} - \alpha \frac{\partial L}{\partial F_{i-1}} \ , \tag{18}$$

α: learning rate
i: iterations for a layer

Activation map is produced as the result of applying a filter on the input. It can be considered as a matrix with values connecting a small region of the input to the filter. In fact, each activation map is a compressed transformation that shares same parameters. The stack of activation maps can be fed to a pooling layer. A pooling layer makes the representation more abstract and manageable (down-sampling). There are different types of pooling layers such as: maximum, minimum, average and adaptive pooling.

Pooling layers in deep learning offer translational invariance. This feature (mostly in classification applications) allows the model to detect patterns regardless of their positions. For instance, if a model is designed to count faces in a photo, it might encounter problems in detecting faces in a selfie with faces from different orientations (Figure 3-12).



a. Same orientation for faces b. Radial face orientation

Figure 3-12 Translational invariance by pooling layers in CNN; pooling allows the detection of patterns regardless of their position, without pooling layers, the model is more efficient in face detection in a rather than b.

However, in case of enhancing EATMN situational awareness, both geographical vicinity and time sequence are valuable and pooling layers are avoided to protect such inferences (esp. that the model is a regression model and not a classifier).

In accordance to described aspects of convolutional networks, the input of the model needs to be restructured but same database is processed such that results can be compared with baseline model. Once the input data is structured to be fed into a convolutional network, other adaptations were made to maximize the efficiency of the features because the architecture of proposed DCNN benefits from deep learning.

Data preparation

From the same span of data (2017 and 2018), daily features representing each day are added to include more features in the model. These are the same daywise features that were used for baseline models (Subproblem-II variables) except regulation types that are included as channels in DCNN. This results in a feature vector with a length of 24 to represent each day:

- N_{Reg}, $\overline{D}_{Reg}$, N_{Tact}, and N_{ACC};
- AIRAC cycles that adds 13 encoded features, and
- weekdays that are converted into 7 encoded features.

Spatiotemporal feature map

In convolution layers, either Traffic Volumes (TVs) or ACCs can be used to construct the spatial bins for feature map. However on a EATMN scale, TVs are at a lower granularity compared to ACCs and buills up on the model’s complexity with no significant benefit. Even, a division of the data over TVs limits the number of data points for learning. But taking ACCs is a better compromise since it avoids detailed granularity while preserving the spatial patterns of regulations in bins. Therefore, ACCs are extracted from TVS Id instead of TV Id (Table 3-13).

Table 3-13 Regulation's dataset structure (NMIR)

Field	Sample Entry	Field	Sample Entry
TVS Id	EDYYFMP	Reg Activation Notice [a]	98
TV Id	MASBWST	Reg Duration [a]	42
Reg Id	YBWST01	Reg Window Width [a]	10
Protected Location Id	EDYYBWST	MP Regulated Traffic [b]	90
Protected Location Type	Airspace	Regulated Traffic [b]	93
Reg Start Time	01.01.2018 20:00:00	ATFM Delay [a]	259
Reg Truncated Start	01.01.2018	MP Delayed Traffic [b]	24
Reg End Date	01.01.2018 21:40:00	Avg Delay per Regulated Traffic [a]	2.8
Reg Cancel Status	Cancelled	Reg Reason Name	S - ATC Staffing
Reg Cancel Date	01.01.2018 20:42:21	Reg Description	(text)
Reg Activation Date	01.01.2018 18:22:19	Day of the Week	Monday

[a] in minutes, [b] flight count.

For 88 different ACCs across ECAC area, each day is divided with a bin size of one hour to also make the temporal bins (i.e. hour of the day and respective ACC are regarded as vertical and horizontal position of a pixel in an image for CNN). Moreover, instead of merging all regulations for each ACC at each time bin, 6 channels are set for different regulation types as of Table 3-14. These definitions for spatiotemporal bins build a *N*x*C*x*H*x*W* matrix that can be taken for a 2D convolution in Pytorch [122]. *N* corresponds to the number of days, *C* to the number of channels, *H* is the time bins and *W* is the spatial bins.

Table 3-14 Defined channels based on regulation types.

Channel	Regulation Type
1	C-ATC Capacity
2	S-ATC Staffing
3	G-Aerodrome Capacity
4	W-Weather
5	I-ATC Ind Action
6	M-Airspace Management, O-Other, P-Special Event V-Environmental Issues, E-Aerodrome Services, T-ATC Equipment, R-ATC Routings, A-Accident/Incident, N-Ind Action non-ATC

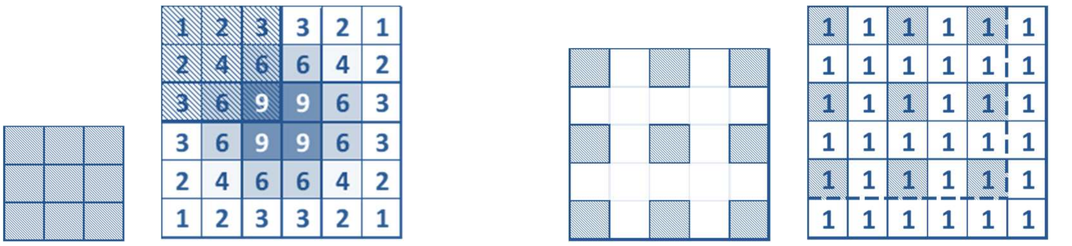
For the activation function, a variation of ReLU function known as Leaky ReLU is taken in activation maps, since it supports generalization for deep NNs [123]. Furthermore, the proposed model uses Weighted Mean Absolute Error (WMAE) as the cost function to improve predictions for low category of target values. These weights are calculated from a negative exponential function (Equation 19), which delivers higher magnitude for low target values (y_i) and flattens out for medium to large values.

$$w_i = 25 * y_i^{-1/3} . \tag{19}$$

Model architecture

The proposed model was designed using an iterative process and is inspired by the model in [120]. In their model, the authors have not explicitly reasoned why a large number of filters (kernels) were used but it is clear that they used their model to process numerous frames of a video in a pixel-wise video prediction task. Since a large number of filters significantly increases the computational effort an initial architecture with few filters and layers is implemented at first. Based on the performance of the model on test set, the filters and the model architecture were iteratively improved to achieve the final proposed architecture.

For instance, dilation as a feature of deep learning proved to be efficient in improving the results. A convolutional layer without dilation, applies the filter (or kernel) less on the edges and more on the middle values of input matrix. This actively demonstrates that the model is less sensitive to early hours of the day and the sequence of 88 ACCs, asserts less importance on the first and the last ACC; i.e. these hours and ACCs are less exposed to learning. Dilation rate mitigates this risk by specifying a spacing between values of a filter. For instance, in a 2D space, a 3x3 filter takes 9 adjacent pixels of input image, but a filter with a dilation rate of 2, takes 9 pixels out of 5x5 region as of a 5x5 filter that ignores every second column and row (Figure 3-13).



Left: 3x3 filter, dilation: zero; right: 6x6 input matrix   Left: 3x3 filter, dilation: two; right: 6x6 input matrix

Figure 3-13 Dilation in convolutional layers; numbers inside each input matrix shows times that each cell has been scanned by filter. Dilation assures fair usage of all values of input matrix.

The finalized architecture, (Figure 3-14 and Table 3-15) has two blocks of convolution filters that are applied to six input channels (spatiotemporal feature maps). Each block has two independent temporal and spatial filters, which are followed by a spatiotemporal filter to check for correlated patterns. Second block has bigger filters that considers a longer time span and a larger geographical area, therefore dilation is used to improve fair calculations. The output of second block (which extracts deeper features) is then aggregated with a unit size kernel to get a single channel.

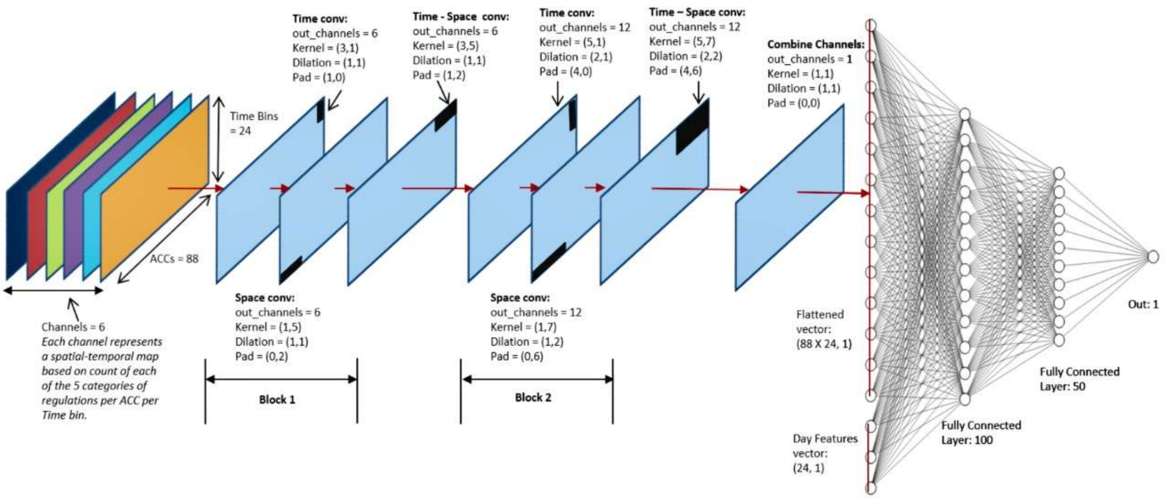


Figure 3-14 Proposed architecture for DCNN. Channels are set based on the regulation types and two blocks of convolutional layers, which learn the spatiotemporal characters of regulations, *[62]*.

The first block, that deals with temporal dimension is less responsive to different dilations since there are no major activities at early morning hours as well as late hours of the night (silent edges versus busy hours of the day). In contrast, second block is intentionally explored by bigger filters and different dilations in search of correlated airspaces (ACCs). Moreover, each block is creating a defined set of activation maps (6 and 12 for first and second block respectively). The result of convolutional layers is flattened to a vector (by a 1x1 filter with zero padding) and concatenated with daywise feature vector. This vector is processed by a sequential neural network (SNN) with two fully connected layers (100 and 50 neurons). The output of the model is a single neuron that predicts either delay or the delayed traffic.

Table 3-15 Outline of convolution layers in proposed DCNN architecture.

Layer	Kernel Size	Dilation	Padding	Type
1	(3,1)	(1,1)	(1,0)	time
2	(1,5)	(1,1)	(0,2)	ACC
3	(3,5)	(1,1)	(1,2)	spatiotemporal
4	(5,1)	(2,1)	(4,0)	time
5	(1,7)	(1,2)	(0,6)	ACC
6	(5,7)	(2,2)	(4,6)	spatiotemporal
7	(1,1)	(1,1)	(0,0)	aggregation

3.4. Data collection process

This section describes the data domain, acquisition and sampling technique that formed the input data for different part of the study. As mentioned before, the regulation data are selected since they represent an encapsulated information about how the daily demand-capacity imbalances are being managed at a large scale.

3.4.1 ATFCM regulations & flight plans

Each ANSP is closely monitoring its capacity to accommodate the flight demand. The demand itself is calculated based on the submitted flight plans from airlines. A flight plan defines many characteristics of a flight but most importantly it declares the requested flight route and the scheduled time of departure (STD) and estimated arrival (ETA) (flight profile). The flight profile is a representation of the four-dimensional path that a flight is expected to follow between departure and arrival aerodromes. The profile calculation is required to validate the flight route, to determine the ANSP address list for message distribution that further assist demand forecast.

In Europe the capacity is defined as the maximum number of aircrafts that can safely enter an air traffic control sector in a specified period [124]. Each sector is a defined airspace region for which an associated controller (or controllers) has ATC responsibility. ANSPs mostly use fast time simulations (FTS) to estimate the en-route capacity which is actually a computer modeling of controller workload. The results of simulations are then post-processed to formulate the relationship between the number of entering aircrafts and controller workload over a given period of time.

Knowing the demand (from flight plans) and the capacity (From FTS) enables the automatic balancing between demand and capacity. But, on the day of operations (tactical phase) both demand and capacity figures change. Because, FTS are not covering important factors such as the complexity of the traffic, structure and geometry of the sector and their interactions; hence 'Real-time simulations' or RTS are being used to improve capacity estimations.

Both FTS and RTS are done in strategic phase of ATFM. The latter considers the human related elements such as cognition, thinking and judgement by accounting for operational environment which pushes the process to be expensive and requires personnel training, specific infrastructures and significant simulation time. Consequently, RTS is not an option for all of the en-route capacity estimations. Another solution for capacity declaration is air traffic controller workload model, for example:

- Sector Design and Analysis Tool (SDAT), developed by FAA, is an analytic model used in United States. The model is focused on routine tasks, probabilistic conflict resolution, sector scanning and planning. SDAT addresses both Planning Controller (PC) and Tactical Controller (TC).
- Total Airspace and Airport Modeler (TAAM), developed by Preston Group (Boeing), is a simulation model which is used in Germany and Switzerland. The focus is on routine tasks, deterministic conflict detection and resolution. This method accounts for TC only.
- Reorganized ATC Mathematical Simulator (RAMS), developed by EUROCONTROL is based on controller (PC & TC) observation, and is focused

on same aspects as TAAM. It has been applied in numerous European airspaces and proved to be flexible and easy to use.

- Performance and Usability Modeling in ATM (PUMA), developed by NATS, is also a simulation model which incorporates observation task analysis and cognitive debrief. Yet unlike other models, it assumes that controllers can handle more than one task simultaneously.

However, tactical phase is more dynamic to be fully modeled by different demand and capacity prediction models. For instance, the challenges to capacity predictions are [73]:

- individual sector capacities,
- flexibility & adaptability of airspace (configuration and sectorization),
- staffing,
- tactical configuration management,
- exogenous factors (e.g. weather).

As a response to such challenges, in Europe ATFCM Regulations (also referred as Regulations and Flow Regulations) are measures that are available to cope with demand-capacity imbalances. A regulation is basically a restriction over the rate of flights being authorized to enter a monitored sector. For instance, a regulation with zero-rate is in fact a closed airspace (e.g. Temporary Segregated Area, TSA). In reference documents [7], it is defined as:

> *Regulation is a method of matching traffic demand to available capacity by limiting the number of flights planned to enter an airspace or aerodrome, achieved by the issuing of departure slots (CTOTs).*

A regulation is a potential cause for ATFM delay. ATFM delay can be assigned to a flight based on the submitted flight plan. In Europe under the authority of the EUROCONTROL, Network Manager (NM) is using a centralized flight plan processing service to organize flights [125]. The service is provided by the integrated IFPS.

Each airline is required to submit flight plans to the IFPS for processing at least three hours before the EOBT (Estimated Off-Block Time) where possible. The option to submit flight plans in such a time window is assuring the required level of flexibility in tactical phase. IFPS accepts flight plans that are filed even less than three hours ahead of departure time if operational reasons restrict the normal submission. In general, IFPS accepts a submission up to a maximum of 120 hours (Figure 3-15), ahead of its EOBT (item 18 of flight plan should include Date Of Flight (DOF) in case of a flight plan for a future date).

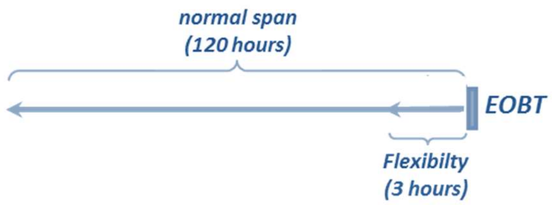


Figure 3-15 Allowance time to submit flight plans to IFPS.

When a message is submitted to the IFPS, a copy of the flight plan is sent by IFPS to the Enhanced Tactical Flow Management System (ETFMS) where the flight information is being analyzed with regard to all active flow regulations. If the flight is penalized with one of the active regulations in its route, origin or destination, ETFMS assigns a delay to the flight by issuing a CTOT to the flight. This delay is referred as ATFM delay.

> *The IFPS and ETFMS are separate systems; any message submitted to the IFPS must be acknowledged before it is transmitted to the ETFMS, where any relevant flow regulations may then be applied, thus the IFPS as the interface to airlines is not able to assess the impact of flow regulations on flights [126].*

Moreover, if the flight is close to its departure by the time that the regulation is activated (or updated), ETFMS is not assigning a new delay to the flight. In fact, if the Off-Block Time (OBT) is within the next 30 minutes this rule applies in ETFMS to protect the airport startup sequence and avoids last minute change massages to the airline.

The final decision on the regulation implementation is for the responsible ANSP (also referred to as respective ACC or FMP). However, the details on the regulation itself should be coordinated with the NM. Sometimes a network measure (e.g. level capping, rerouting) impacts more than one FMP and NM will be the responsible decision maker for the regulation.

After coordination with the FMP, the NM decides to activate regulations in those locations where it is necessary. In ETFMS, regulations include the start and the end times, the description of the location, the entering flow rate and similar parameters.

3.4.2 Covered domain

Since the thesis is addressing the resilience of EATMN, referred regulation data should cover the geographic ECAC area. This indicates that all flights which are flying from IFPS zone (IFPZ, see annex A.1) are considered for all active regulations. All submitted flight plans to IFPS which are typically IFR flights are considered to define impacted flights. Also, each regulation can be active for different durations (from hours up to several days). Therefore, the reference time stamp for deciding on tactical regulations is the regulation start time and regulation publication time. Lastly, full year cut of data is acquired because of known seasonal demand patterns.

3.4.3 Databases (NOP & NMIR)

FAA offers Aviation System Performance Metrics (ASPM) as an online system that collects data from sources such as: Traffic Flow Management System (TFMS), Airline Service Quality Performance (ASQP), CountOps and Flight Schedule Data System (FSDS). EUROCONTROL also offers various data sources that in this study are considered against the following criteria:

- update rate: the database must be published throughout the tactical and pre-tactical phase to be more relevant to decision making processes on the day of operation,
- granularity: selected data should provide required precision to understand types of disruption including spatial and temporal dimensions,
- coverage and accessibility: selected data should not only be accessible by all layers of decision makers across the European Civil Aviation Conference (ECAC) area but also collect data from ECAC to represent EATMN.

According to the mentioned characteristics, delay statistics including reports from CODA (Europe), OPSNET (US), STATFOR and those that are published in the post-operational phase are not considered. Likewise, databases including NPR and ATFCM statistics and network operations reports are only published for authorized users and cannot fulfill the accessibility criterion. However, a focus on capacity (ATFCM) regulations is promising since they denote the results of collaborative decision making. In fact, a regulation is a method of matching traffic demand to available capacity by limiting the number of flights planned to enter an airspace or aerodrome (EUROCONTROL 2014).

In tactical phase, ATFCM regulations are initiated based on the evaluation of ATFCM Daily Plans (ADP) from the pre-tactical phase and they are subject to constant updates if required. Regulations correspond to network states in the restorative level of resilience.

ADP conveys the results of pre-tactical planning processes to the tactical phase of operation. This plan is promulgated by means of INP and ANM [10] :

- INP (Initial Network Plan): informs the ANSPs and AUs about the congested areas and suggests alternatives to avoid heavy delays. ANSPs act and organize accordingly to maximize the airspace utilization and airlines consider suggested routes or flight levels in filing their flight plans to optimize their operations.
- ANM (ATFCM Notification Message) is a message issued publicly to notify all concerned of any ATFCM regulations. Some ANMs reflect regulations from the ATFCM Daily Plan from pre-tactical phase. But the list of ANMs is constantly being updated in the tactical phase to notify any new, changed or cancelled regulation.

ANMs are generated in pre-tactical and tactical phase. Post-operation records are stored separately and is provided by Network Manager Interactive Reporting dashboard (NMIR) but only to authorized users.

> *The Network Manager Interactive Reporting (NMIR) is a web interface allowing users of NM systems to access a wide range of reports and statistical data on European ATFCM archived data [127].*

Tactically, ANM records are available both on NOP portal and CHMI (Collaboration Human Machine Interface that offers authorized users real time information). Same online tools also provide ATFCM Information Message (AIM) which is intended to notify NMOC daily operations including possible disruptions. In this thesis AIMs are not used because of the update rate and their format as text messages that serve as a description with much less structured technical data.

There is no record of ANM messages in post operations but NMIR collects final regulation data with complementary details. NMIR is offered to all ATM authorized actors, but because universities and research institutes such as DLR are not typically considered as ATM actors, an access tight is acquired for the thesis that can be used for follow up studies by DLR too.

3.4.4 Data characteristics

The purpose of ANMs is to provide the information related to implementation of ATFCM measures and they are published by the EUROCONTROL's network operations portal [12]. A sample of an ANM record in NOP is given in Figure 3-16 and Table 3-16. The list is updated by push messages when a new regulation is activated, a parameter of one gets changed, or regulation is cancelled in the Enhanced Tactical Flow Management System (ETFMS). The tag for type of push-message is stored in the field of 'state'.

Table 3-16 ANM data structure

Field	Sample Entry	Field	Sample Entry
Seq no	009	State	New
FMP	LFFFAD	Published	12/03/2021 06:00
Regulation Id	LFPNVD12	WEF [c]	12/03/2021 08:00
Flight Level [a]	ALL	UNT [d]	12/03/2021 12:00
Reason	ATC Capacity		
RMK[b]	Calibration flight LFPN + LFPV Departures		

[a] Flight Level can also be e.g. "145-" to indicate that regulation applies to all levels below 14 500 feet,

[b] Remark, [c] With Effect From, [d] UNTil.

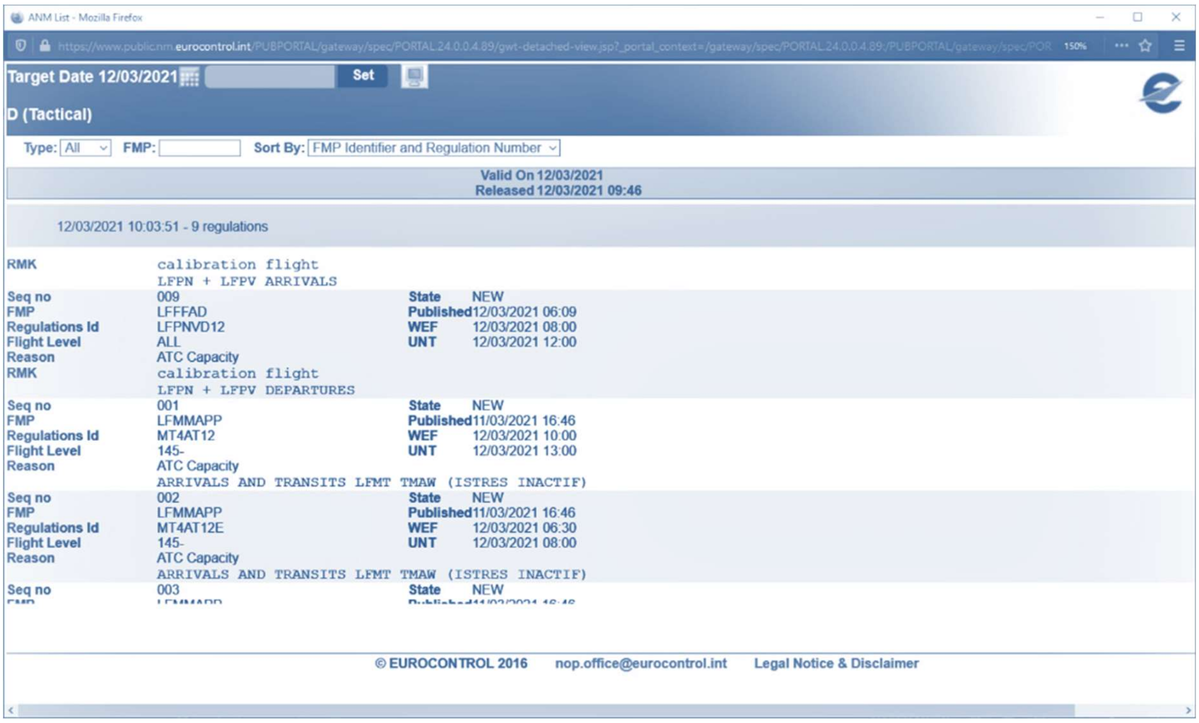


Figure 3-16 Snapshot of ANM list

The offered data structure in NMIR is extended as a post-operational data base (Table 3-13). In this research both ANM and NMIR data on regulations are used with different purposes. ANMs contribute to network state definition and NMIR records are used for learning models.

In contrast to NOP/CHMI, NMIR allows to download all regulations for different years in different formats (granularity is per day). The vacancy of ANM messages and their updates are not offered as a data base. Therefore, a sampling technique is implemented to save public ANM messages for each day in two granularities, daily and 10-min snapshots.

Daily records can be downloaded as limited post-operation records. Only the last status of ANM list is stored temporary on the NOP portal (input for ANMStat tool). The 10-min snapshots are acquired by writing an executable file (ANM Capture.exe). As provided in Figure 3-17, once the written code is executed it will take a copy of active ANM list from NOP portal every 10 minutes and save it in an excel sheet. Therefore, for each day there is an excel file with sheets that each contains a copy of ANM list at a specific time. This allows to capture any update with a granularity of 10 minutes at day of operation.

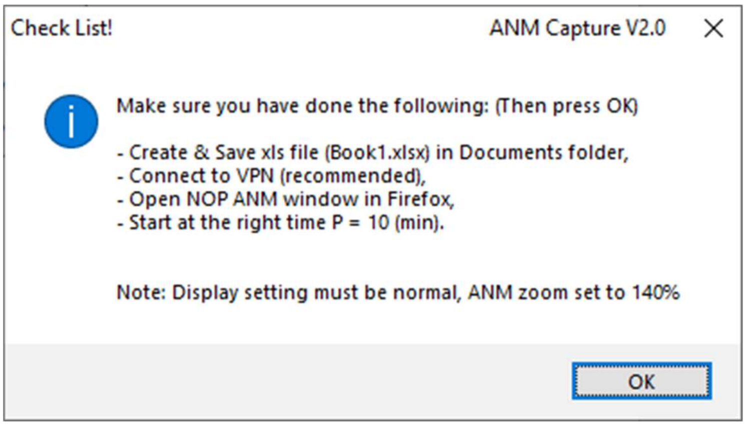


Figure 3-17 ANM Capture, executable file

4. Resilient Path (RNN)

4.1. Complex Adaptive Systems (CAS)

The concept of CAS involves a group of systems that basically possess two main characters: evolving structure and moving target [128]. A typical CAS, evolves to adapt according to changes of its environment and this will reflect on iterative update of targets. This demonstrates that an optimal performance at some point does not necessarily lead to a global optimal performance. As reported by MIT [129, 130], the topic of CAS is introduced back in 1980's with an emphasis on crossing of traditional disciplinary boundaries. CAS provides an alternative to the linear reductionist methodology in modeling systems, that relies on fixed assumptions to simplify the modeling task. CAS benefits from computer aided simulations instead. ATM network as described earlier is a complexity that results from the inter-relationship, inter-action and inter-connectivity of its elements and environment.

In fact, EATMN is a CAS since it hosts complex disruptions that emerge as a result of dynamic and nonlinear spatio-temporal interactions among its subsystems. The adaptability of CAS is totally consistent with what has been conceptualized as a resilient behavior. A MIT study [130] defines seven attributes for CAS and Table 4-1 provides the matching characteristics of EATMN.

Table 4-1 EATMN as a Complex Adaptive System

CAS attribute	EATMN characteristic
Distributed Control	Four main Decision Makers (NM, ANSP, Airport, AU).
Emergent Order	Emergent behavior of the network and performance variability are discussed in section 2.1 (resilience).
Connectivity	Flights and infrastructures such as route network and dataflows link elements of EATMN.
Co-evolution	rising demand or lost capacity trigger all actors to adapt their resources to deliver services.
State of Paradox	Operations in ATM has built in levels of flexibility which reminds the idea of bounded instability in CAS.
Sensitive Dependence on Initial Conditions	Despite of strategic and pre-tatcical plans, unpredictability of the network is a fact that is observed during volcanic ash in 2010 and ETFMS failure in 2018.
Far From Equilibrium	This attribute is observed with daily traffic pattern adaptations. Not only flights are rescheduled but also re-routed to avoid low weather conditions, restricted areas, costly charging zones and etc.

The idea of CAS is mainly discussed in American academics. Donohue from George Mason university presented air transportation as CAS in 2003 [131] . Computer simulation brought more attention to this topic; in 2009 a study from Purdue university [132] addressed the network transition problem in air transportation by agent-based modeling to assess two scenarios. Supported by increased computational capacity, researchers from Arizona state university [133], were able to provide a more mathematical model to address the emergent disruptions from the interoperability of system components.

In the same context, a PhD dissertation [134] considered the evolution of the airline route network and its impact on air traffic delay through machine learning. At early 2016, DARPA[1] focused assessing and predicting the complex emergent behaviors that constantly change across time and space especially in aviation. The Complex Adaptive System Composition And Design Environment (CASCADE) program tries to develop mathematical foundations of system design to enable real-time resilient response in dynamic environments. Adaptive and resilient urban infrastructure is one of the key focus areas in this program.

Other studies from George Mason university reviewed resilience in CAS. Roberts et al. [135] used statistical approach to quantify risk of emergent behavior to reduce probability of excessive costs. Punpuni-Lenss et al. [136] also studied the CAS resiliency but their approach includes agent-based modeling rather than statistical approaches.

Most recently, Ordoukhanian and Madni [137], explored the resilience of a Multi-Unmanned Aerial Vehicle (UAV) system in face of disrupting events. In their study, real-time evaluations of resilience alternatives showed that multi-UAV system tactics dynamically change the priorities of the system based on the system state. Similarly, in this research, the state of the network is determined to capture the dynamic environment of EATMN but next, instead of computer simulations or agent-based models, the focus is on ML predictions on regulation’s impact that eventually enables real-time evaluation of restorative measures.

4.2. Recurrent Neural Network (RNN)

CAS benefits from evaluation of corrective measures and since EATMN can be considered as a CAS, it is well established that any measure (i.e. capacity regulations), should be in alignment with defined dynamic environment (network state). Therefore, similar to the approach for UAV systems [137], capacity regulations in EATMN also need to be evaluated in real-time but not by simulations.

This thesis addresses the need to evaluate corrective measures by predicting the impact of ATFCM regulations through RNNs. In previous sections, some supervised machine learning methods (especially CNN) have been studied. But the focus is now shifted to predict individual regulations by RNNs because not only primary results dedicated the degree of

[1] Defense Advanced Research Projects Agency

tactical network evolution but also RNNs tend to be more capable of capturing the sequential nature of regulations as time series.

Another important factor in deciding to continue with RNN is reference location of each regulation (spatial dimension of data). Despite the demonstrated advantage of CNN in predicting network parameters, CNN is less likely to outperform RNN in predicting individual regulations. Each regulation is assigned to a specific TV which has a parent reference location (i.e, either an elementary or a collapsed sector). The count of available reference locations (Figure 4-1) is so high that will challenge CNN in extraction of spatial patterns.

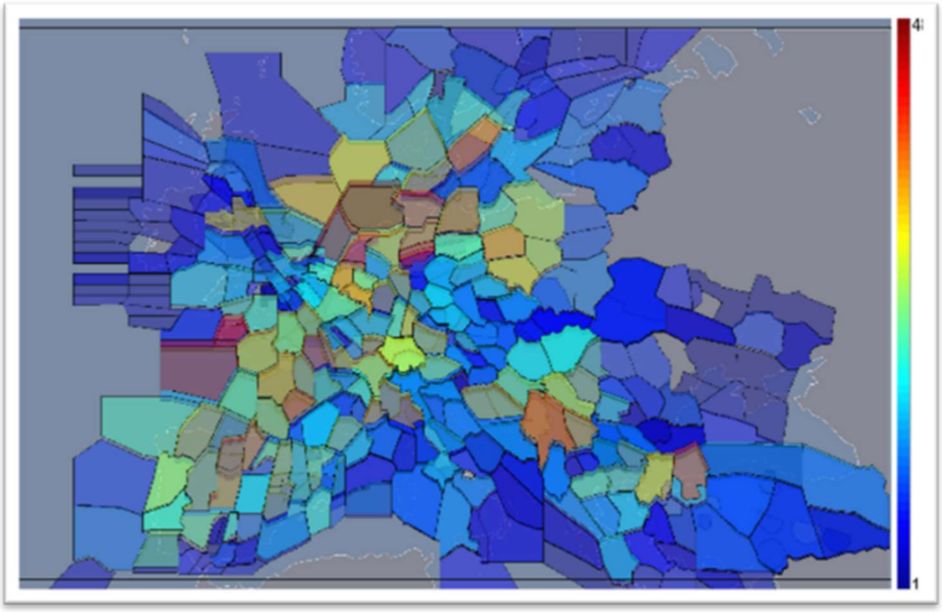


Figure 4-1 Sectors in European Sky: DLR-institute of air transport, NFE tool

RNN experiments on sequential data cover a wide range from word processing applications such as translations [138] to anomaly detection in aircraft data. A study from George Mason university [139] proposed a RNN model with LSTM cells to detect canonical anomalies from Flight Operational Quality Assurance (FOQA) data. Therefore, before designing the RNN it is important to locate other methodologies in the recent literature. Cited works in Table 4-2, consider post operational data including flight schedules and trajectories (ADS-B data). But the size of the datasets is not comparable to learning applications in Natural Language Processing (NLP). NLP is the research area that most RNN applications are focused on. Nevertheless, the studies continue with learning models and report significant capabilities in delay predictions.

Also, it can be observed that some studies focus on flight delays over specific routes [140] or a single airport [141]. The spatial domain is less likely to includes multi-national airspaces. For instance, all cited works in Table 4-2 are applied to either US or china airspace.

Table 4-2 Recent studies to predict delay with learning methods.

Author/Year	Method(s)	Target	Data	Notes
Gopalakrishnan & Balakrishnan [140]/2017	ANN, CART[a], MJLS[b]	OD delay	2011-2012	7 features, trained on 2011, tested on 2012 data, delay threshold: above 60 minutes.
Thiagarajan et al. [142]/2017	NN, RF, GBoost[c]	Arrival & departure delay	2012-2016	Classification and regression models are explored; 21 features for departure delay & 36 for arrival delay; data filtered to selected 15 airports.
Manna et al. [143]/2017	GBoost	Arrival & departure delay	Apr-Oct 2013	8 features, data is filtered out by size of delay (25 to 75 percentile).
Yu et al. [141] /2019	Deep Belief N, KNN, SVM, LR	Departure delay	Jan-Mar 2018	Data included 528 471 domestic flights from a single airport, model also encoded SVR.
Lin et al. [144]/2019	ConvLSTM	Traffic flow matrix	Jul-Aug 2014	Focus on both spatial and temporal dimension; total number of data samples: 89 280; ADS-B data.
Gui et al. [102] /2020	LSTM, RF	Departure delay	Dec 2018-May 2019	15 features including weather info; both classification and regression; ADS-B data.

[a] Classification And Regression Tree, [b] Markov Jump Linear System, [c] Gradient Boost.

Modeling approaches is another key difference in studies. A study from MIT [140] took the data from National Airspace System (NAS) of US to create a delay network for main 30 airports. Their study only considered delays bigger than one hour and concludes the superior performance of ANN in classification (if a delay will occur or not). Thiagarajan et al. [142] added the weather information as input features. Their prediction model considered 36 features for arrival and 21 features for departure delay. Despite using different feature selection methods, they reported only a minor improvement in prediction accuracy (only 0.22) but similar to this thesis, their study declared random forest and gradient boost method to be efficient as classic models for both regression and classification.

Gradient boost was also applied by Manna et.al [143], but their approach in filtering the data, significantly affected the quality of results. They cut the data for each day to be between Q1−1.5∗IQR and Q3+1.5∗IQR (Inter-Quartile Range), where Q1 and Q3 are the 25 and 75 percentiles. Therefore, as given in Table 4-3, despite having a small dataset (less than a year) and less features, their gradient boos method delivered significantly better results than [142].

Table 4-3 Departure delay regression results by Gradient Boost method

Author	MSE	R^2
Thiagarajan et al. [142]/2017	1 218.75	0.055
Manna et al. [143]/2017	67.027	0.948

Size of the dataset is not regarded as an issue in contrast with what is generally expected from supervised learning methods. In another study by Yu et al. [141] data is obtained from one airport but the design of the model overcome this limitation. In their model, Deep Belief Network is acting as a feature selection process that filters the input to a SVR that predicts delays. The results showed that 99.3% of predictions have a tolerance of ± 25 minutes from the observed values.

Some studies [102, 144] rely on ADS-B data that provide a dataset with higher granularity, especially on flight trajectory. Lin et al. [144], proposed a model (ConvLSTM) to capture tempo-spatial correlation. After a transformation, almost 90 thousand data point constitute the input dataset. However, the model predicts traffic instead of delay. But Gui et al. [102] predicted flight delay by combining ADS-B data with weather, airport information, and flight schedule. Their model was also based on the LSTM units that generally have four control gates, i.e., input gate, forget gate, cell, and output gate. Conversely, the proposed LSTM [102] had only limited dataset as input (max 1542 input sequences) and predicts delay as a classification task.

Compared to aforementioned studies, the intention here is to finalize a RNN with LSTM units to predict delay as a regression task and instead of ADS-B data, NMIR data are used to cover EATMN for all airports/airlines. The required steps to build the RNN is rich enough that this part of the thesis is fostered through a master thesis [145].

4.2.1 Data preparation

In this section, preliminary steps on NMIR data is described in different steps including data transformations, feature encoding and splitting. Regulation of 2015 until 2018 are obtained from NMIR to be processed for this section. Despite the time stamps that allows the regulations to be studied in different sizes, the preference here is to have yearly samples rather than monthly samples. This allows the model to learn seasonal patterns too. 2017 is selected as the training set in comparison to 2018 that is rather risky with high traffic and delay figures.

Since RNN is a neural network the split ratio for training and testing needs to respect the risk of over and underfitting similar to described CNN. Therefore, data is splitted to 70%, train set and 30% for the test set. To split the data set Scikit-learn from python programming is used. The reference parameter for splitting the data is the target value (delay) that is ranging from 0 to 32,795. The result of dividing the data into train and test subsets is given in Figure 4-2. This figure shows that the splitting process assures similar distribution of subsets.

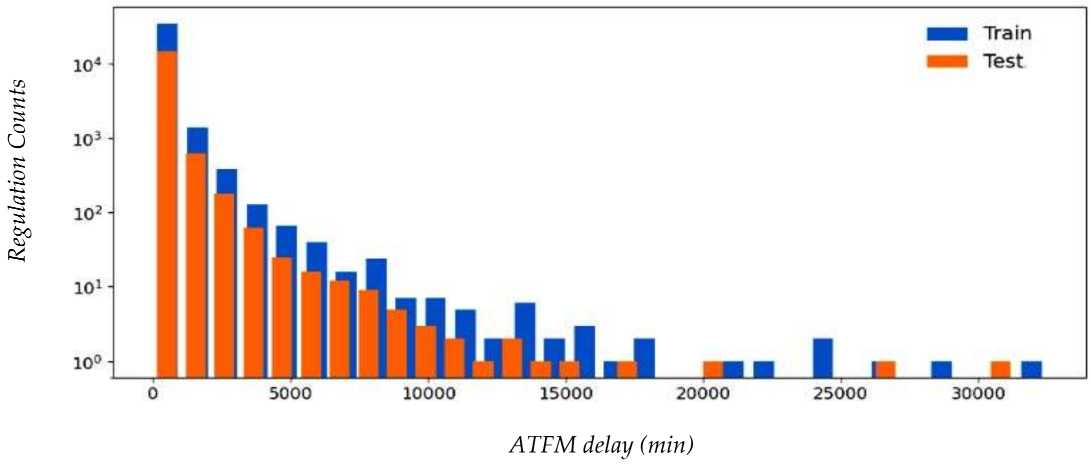


Figure 4-2 Train and test set after splitting based on delay values, *[145]*

After splitting the data, the input vector needs to be prepared with regard to different types of features. Although the data structure is same as Table 3-13, pre-processing considers three types of features that are described in Table 4-4 and discussed individually.

Table 4-4 Different types of Feature in pre-processing

Feature Type	Features
Scalable	Regulation Duration (min); Regulation Window Width (min); Regulation Activation Notice (min); Regulation Counter [a]
Recursive	Regulation Activation Date; Regulation Start Time; Regulation End Time; Regulation Cancel Date; Weekday; AIRAC [a]
Categorial	TVS Id / FMP; Regulation Reason; Regulation Id, Protected Location ID; Protected Location Type; TV Id; Regulation Cancel Status; Regulation Activation Date; Regulation Start Time; Regulation End Time; Weekday; AIRAC [a].

[a] AIRAC cycles and regulation counter are not provided in NMIR dataset.

Note: some features belong to more than one category since different encodings have been tested.

- Scalable features: those that require scaling, otherwise their different ranges push the neural network to be biased in favor of features with bigger values. These features are scaled by scikit-learn preprocessing library. Among different provided functions, MinMaxScaler and StandardScaler are used in this section.

The former transforms each value by its distance to reference point (minimum value) and then scales it according to the range (max-min). The latter scaler, takes the feature and transforms them to a standard normal distribution (μ = 0 & σ = 1).

Regulation Counter is not a given feature in NMIR data structure. This feature is designed to represent the number of active regulations in the network according to algorithm 1.

Algorithm 1: Calculate regulation counts feature

Input: NMIR data
Output: Regulation Counter
Select the required features (regulation Activation Notice, TVS ID)
Sort the regulation according to activation notice time
Create empty columns for each scenario (3hrs, 4hrs, 12hrs and 24hrs)
n ⇦ total number of regulations
for *i* = 1 to *n* **do**
step 1. read the activation time (e.g. 13:00)
step 2. define relative time intervals for each reference scenario
(e.g. 10:00 to 13:00 for 3hrs scenario)
step 3. count number of active regulations for each time window of step 2
(e.g. 36 active regulations in last 3hrs)
step 4. Count only regulations with the same corresponding TVS ID for each time window
(e.g. 4 active regulations in same TVS ID in last 3hrs)
Step 5. Save calculated numbers in step 4 to corresponding column for each scenario
End

- Categorial features: these features are encoded with the same library with one-hot encoder from the same library of scikit-learn. As described by algorithm 2, it assigns a binary array with the length of number of possible categories.

Algorithm 2: One-Hot Encoding for categorial feature

Input: NMIR data
Output: transformed categorial feature
x ⇦ Select the categorial feature (e.g. regulation reason)
C ⇦ Set of all possible values of *x*
n ⇦ Number of all possible categories (e.g. 14 categories for regulation reas
E ⇦ Create an empty binary array with the length *n*
for *i* = 1 to *n* **do**
if $x == C_{(i)}$ ***then***
$E_{(i)}=1$
else
$E_{(i)}=0$
end
return *E* (e.g. for a reason of second category, *E* = [0 1 0 0 0 0 0 0 0 0 0 0 0 0])

- Recursive features: these features deal with time, such as weekdays and hours. Such features are sequential and are different to categorial features. It is necessary to make the neural network realize this aspect. For example, a categorial encoding of weekdays is ignoring the fact that Wednesday is 4th day of the week and is closer to Tuesday compared to Saturday.

  This cyclic nature of the features can be considered by assigning values of a periodic function such as sine or cosine functions [145]. Figure 4-3 is showing the assigned values to 1440 (24x60) values each representing a minute of 24 hours.

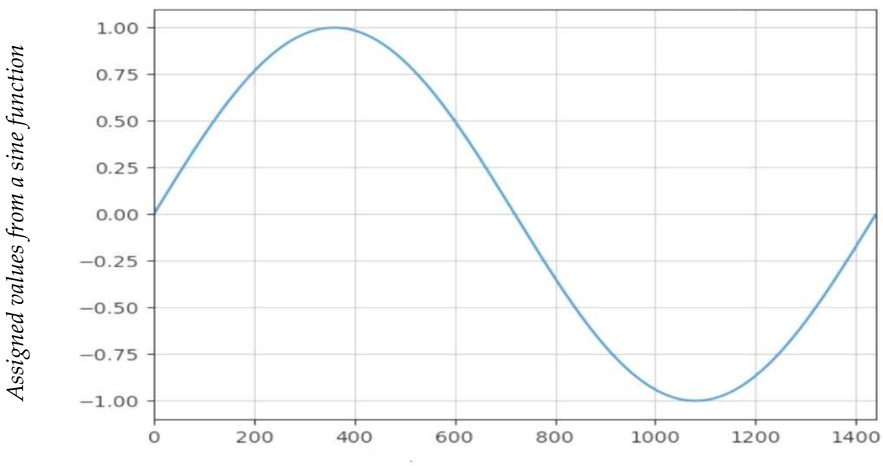


Figure 4-3 Sample of transformation for recursive features (sine function for regulation Start time), *[145]*

The figure, shows that this transformation is for example assigning values close to zero for both 23:55 and 00:05 which original values are at a big distance to each other (i.e. 1435 and 5). The transformation is done by the following algorithm in which as an example a sine function is used to create the reference array:

Algorithm 3: Cyclic transformation for recursive feature

Input: NMIR data

Output: transformation matrix for recursive feature

x ⇦ Select the recursive feature (e.g. regulation start time)

C ⇦ Set of all possible values of x

(e.g. time of the day as discrete values from 0 to 1440)

n ⇦ length of set C

E ⇦ Create an empty transformation reference matrix sized n x 2

for $i = 1$ to n **do**

Convert the feature value ($C_{i,1}$) with Sin function

(i.e. $E_{i,1} = \sin\left(2 * \pi * \frac{C_i}{1440}\right)$)

Convert the feature value ($C_{i,2}$) with Cos function

(i.e. $E_{i,2} = \cos\left(2 * \pi * \frac{C_i}{1440}\right)$)

end

return E

In similar studies, datasets are sanitized by removing outliers [143, 140]. However, this thesis is taking the network resilience as a core value, and outliers are therefore not treated as noise. Also, in order to avoid vanishing or exploding gradient problem, the target value (delay) has also been scaled by MinMaxScaler to a range of [-1,1]. The basic rule of backpropagation in neural networks is that once the prediction is done all weights in layers will be updated proportionate to the associated gradient. A large gradient will exponentially increase in updating the weights of each layer (exploding). Similarly, a small gradient will exponentially decrease (vanishing) as the model propagates back at each layer.

4.2.2 RNN architecture design

The model which presented here is finalized after an iterative design process in which different aspects were considered. Since there is no practiced approach in calculating the design parameters (such as number of neurons and layers), it is inevitable to start from an arbitrary selection and then start an iterative process to improve the architecture. Table 4-5, charts the design aspects that we [145] explored to improve the model.

Table 4-5 RNN model architecture design aspects

Design aspects	Tested alternatives	Result
Feature encoding	Categorial vs. Recursive	Recursive
Learning rate	Constant vs. Dynamic	Dynamic
Neurons at each layer	Max 577	Range of neurons: 71 - 279
Number of layers	Max 12	3 Dense & 3 LSTM layer
Optimizer	Adam, SGD, Adadelta, Adagrad, Adamax, Nadam	Adam

Despite the fact that more layers (deeper network) is increasing the model generalization, the depth of the model needs to be proportionate to size of dataset. NMIR data set on regulations provides data for different years, but in aviation there are yearly and seasonal traffic patterns that is from a different nature than e.g. available texts (data) for a RNN application in NLP. Therefore, it is unrealistic to build a model with numerous layers and here up to 12 layers has been checked. In fact, the main layout of the model is adopted from the deep RNN in a Georgia institute of technology study [77] that offers four ways of forming a deep RNN as:

- Deep input-to-hidden architecture that reduces the effect of non-linear dimensionality that breaks the original input layers such that the underlying factors of variation will be revealed;
- Deep hidden-to-output architecture is considered to be effective to extract variation factors in the hidden state, that ultimately leads to easier output prediction;

- Stack of hidden LSTM layers empowers a model to realize state transitions in different timescales. This is the main advantage especially when the focus is on the sequential patterns of the data, and

- Deep hidden-to-hidden transition architecture enables the RNN to learn a highly nonlinear and non-trivial transition between the consecutive hidden states.

According to mentioned methods, different trials lead us [145] to choose the stack of hidden LSTM layers to make the deep RNN model. The three LSTM layers are designed to converge toward the output layer. In other words, the input layer has 140 neurons, then a dense layer with 279 neurons is between the input layer and the stack of LSTM layers to learn from non-sequential features. The core of the model to predict regulations delay are the 3 LSTM layers. More LSTM layers slightly improved the quality of the predictions at a great cost of computational time. These layers have 279, 140 and 71 LSTM units.

The attempt to add up to 3 more dense layers before the output layer (hidden-to-output) but regardless of number of neurons at each layer no significant improvement was achieved. As another alternative use of dropout layers are tested since they improve the model generalization through preventing the model to overfit. However, results were not promising enough and we changed the focus to investigate other design aspects such as learning rate, optimizer function and encoding technique.

As expected, recursive encoding was proved to be more efficient (Table 4-6) since, apart from loss of sequence, one-hot encoding of features leads to much higher dimensions for input vector. For instance, 'day of the week' feature can be represented by adding only two (by Sin and Cos transformation in Algorithm 3), compare to 7 extras by one-hot encoding (detailed discussion is provided in [145]).

Table 4-6 Comparison of Cyclic vs. One-hot encoding transformations

Set	Cyclic			One-hot encoding		
	MAPE[a]	RMSE[b]	R^2	MAPE[a]	RMSE[b]	R^2
Train	10.20	127.92	0,973	8.47	121.51	0.975
Test	12.80	149.38	0.987	13.69	131.99	0.985

[a] in percentage, [b] minutes.

Furthermore, two optimizers of Adam and Nadam are compared for chances of RNN architecture improvement. Nadam is a variant of Adam optimizer that benefits from nesterov technique [146] to improve the momentum component of the Adam algorithm. More specifically, as Adam searches gradients in each iteration, the new value (θ or new update) for the optimization function ($f(\theta)$) is calculated based on a momentum component and the adaptive learning rate. Adam recalculates the momentum based on previous gradients ($g_t =$

$\nabla_{\theta_{t-1}} f_t(\theta_{t-1})$) solely which is modified to be a decaying sum over previous updates (θ instead of g) in Nadam. Basically, momentum is calculated based on both previous momentum and current gradient ($m_t = \mu_{m_{t-1}} + \alpha_t g_t$).

At a higher computational time, Nadam showed a faster learning process in early epochs as expected but as number of epochs grew, the algorithm showed stabilization problem into reaching the minimum and the cost function started to oscillate and avoid further convergence.

The proposed DCNN in Section 3.3.2, benefitted from Leaky ReLU (activation function) and weighted Mean Absolute Error (WMAE) as the cost function. Here after realizing the superior performance of Adam optimizer in RNN design with a constant learning rate of 0.001, the focus is to improve model performance through experiments on learning rate.

Typically, a learning curve is expected to flatten by iterating over epochs until it reaches the stop criteria. If it is determined to make sure that all epochs provide equal learning chances, learning rate should decrease throughout the learning process. Two of the methods to achieve a dynamic learning rates are, step decaying rate and time decaying rate. The former is basically assigning a decrease rate at every selected number of epochs, while the latter continually reduces the learning rate at each epoch. Keras [119] offers a dynamic decaying rate with two control variables: count of epochs and decay rate (k). Figure 4-4 shows the used reduction of learning rate with k=5 and the initial value of 0.001.

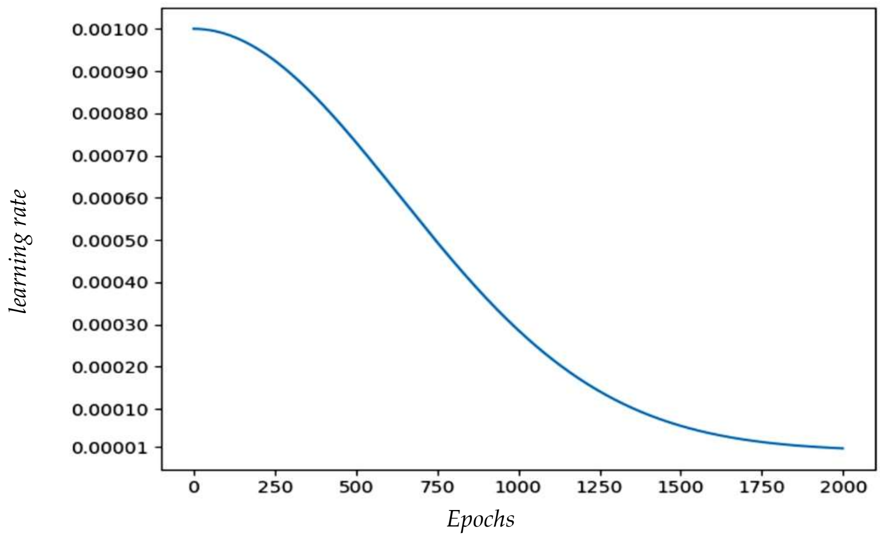


Figure 4-4 Time decaying learning rate, [145]

Experiments with such a learning rate improved the performance of the model to the extent that the delay per regulation can be predicted with an average tolerance of 6.73 minutes.

In order to capture the sequential aspect of regulation not only RNN is used but also previous active regulations in the network is encoded to the input vector. To this end as described by algorithm 1, four different scenarios (3, 4, 12 and 24 hrs) have been considered as additional input feature (i.e. regulation counter). Results (Table 4-7), show that the consideration of regulations in past 12 hours are most effective in predicting the delay.

The precision of the results seems to become better as the input vector considers more previous regulation (from 3 to 12 hrs) but consideration of the past 24 hours reduces the accuracy of the predictions. This is operationally understandable. The pre-tactical planning is putting corrective measures for the day of operation (24 hrs.) and it is true that the traffic patterns in the afternoon is always highly dependent on morning situation (12 hrs.). This is due to the airline schedules and plans for each aircraft that need to be realized by the end of the day.

Table 4-7 RNN model: Comparison of MAPE[a] values for delay prediction in different Scenarios

Set	**Scenario**			
	3hrs	**4hrs**	**12hrs**	**24hrs**
Train	10.20	9.70	8.43	10.53
Test	12.80	12.53	12.18	12.84

[a] in percentage.

5. Results

Firstly, the implemented methodology to declare network state is given based on a use case from selected subset of 2017 [61]. Then the results and discussions on measuring the network disruption through learning methods is provided (with two divisions of feasibility study [64, 35] and DCNN implementation [62, 107]). The chapter is then concluded by results of RNN model and its evaluation against results of DCNN [145]. Basically, this chapter demonstrates the path to reach the goal of this exploratory research and provide added values by discussing the findings toward a new network resiliency concept to enhance tactical situational awareness. The following objectives are set to reach this goal:

- to propose the methodology of capturing emergent disruptions as a result of dynamic interactions among DCB actors in the tactical phase,
- to define network state based on regulations as comprehensive data that present emergent disruptions since regulations encode multiple interactions between subsystems of network,
- to conduct a feasibility study on use of different learning methods in network predictions based on regulation data,
- to propose a new learning architecture designed for predicting network disruptions in terms of delay and delayed flights.

5.1. Network state

The described methodology in section 3.2 was applied to a subset of acquired ANM data from 2017. Since the calculation is based on confidence levels, the two-sided thresholds are estimated for different time horizons. Table 5-1, provides the upper and lower thresholds regarding the confidence level (99.3%) for the day of operations.

Table 5-1 Calculated thresholds of initiated regulations and identified non-nominal states based on ANM data (2017)

Time frame	Thresholds Upper	Thresholds Lower	Interval width	Outliers -	Outliers +
May	137	66	71	0	0
June	166	76	90	0	0
July	191	117	74	0	0
Aug.	168	110	58	1[a]	0
Sept.	150	98	52	1[b]	1[c]
Oct.	131	53	78	2[d]	0
May-Oct.	184	53	131	1[e]	0

[a] 12. Aug. [b] 03. Sept. [c] 27. Sept. [d] 10. Oct. & 18. Oct. [e] 22. July

Listed incidents of outliers in Table 5-1, elaborates on the reasoning behind use of different time-horizons. Although the application of larger time frames with identical confidence levels results in wider control intervals, it is intended to account for short-term patterns. For instance, the summer traffic pattern in July made the single month control interval unable to detect July 22nd as an outlier but it is indeed a disrupted day which is captured by the six-month time frame. In fact, 36% of the 31161 flights were regulated and 22% were delayed with a total of 138818 minutes of delay [147].

In Table 5-1 positive and negative outliers are separated since negative outliers (degradation) are days that exceed upper thresholds and positive outliers (excellence) are those below lower thresholds. Positive outliers are subject to further investigation to set new performance goals with respect to performance variability as the essence of Safety-II. This allows detecting any kind of disruption including cases that the network successfully accommodated to the planned traffic with only minor imbalances. Detected September 27th, 2017 is a good example of a positive outlier in which from 33535 flights, only 9.6% were delayed and total delay was 57425 minutes [147].

As described in Section 3.2.2 for each given daily set of regulations the mean and standard deviation values are compared against reference cumulative values and if the calculated values match any of the critical and crisis state definitions, the ATM network is statistically considered to face a non-nominal state (Figure 3-2). In such cases, the second type of thresholds are evaluated in micro analysis. This way the severity and characteristics of an identified non-nominal state are assessed and the affected ACCs in the network are located.

In aforementioned use case, probability functions are estimated (Section 3.2.3) for all regulation types based on the 99.3% confidence level (as of Figure 3-4) and the corrected bandwidths are considered only for weather and ATC capacity regulations. For other types of regulation, the bandwidths that minimize the estimation error are used.

Table 5-2 presents two days from different kinds of time frames to verify the result of macro analysis and to show the benefits of the micro analysis in distinguishing types of critical states. In both days, significant traffic demand was sufficiently large to push almost 30 percent of flights to be regulated and the given definition of non-nominal states tags both days as critical. In contrast to a crisis mode, the measured loss of capacity is reflected by various regulations as a result of numerous local restrictions. Quantitatively, the 99.3% control interval is [101.8, 315.8] for the standard deviation and measured sigma values fall in the middle of the control interval and declare a critical state. The significant degradation in EATMN at both incidents is confirmed by post-operational report in Table 5-2.

Table 5-2 Statistics of detected non-nominal states

Statistics	Dates	
	12. Aug [b]	**22. Jul** [c]
Mean duration (min)	182.67	157.9
sigma (min)	180.5	175
Traffic demand (flights) [a]	31 343	31 161
Percent of flights Regulated [a, d]	29%	36%
Mean delay of all flights(min) [a]	2.9	4.5

a. Adopted from NM ATFCM Daily Summary (Post-ops).
b. Detected date based on thresholds for Aug.
c. Detected date based on thresholds for May-Oct.
d. Part of traffic demand passing through one or more regulations.

This table also shows similar demand rates at both dates, but July 22nd suffered from more delays than August 12th. Micro analysis investigates such differences as in Table 5-3. Estimated Cumulative Distribution Functions (CDFs) describe totally different types of critical states.

On August 12th, along with ATC capacity and ATC rerouting regulations, the largest number of aerodrome capacity regulations (for the observed six months of 2017) was implemented to accommodate demand. This led the day to be a complex example of a non-nominal state with significant loss of capacity. However, July 22nd was different as high traffic demand was impacted by major weather conditions pushing the total number of regulations to 348 - almost three times of what was initially expected in pre-tactical phase. In fact, the day constitutes a verified case of performance degradation as a result of adverse weather impact. In other words, one is a case of excessive demand (airport network) while the other is reduced capacity (en-route challenge).

Such cases contribute to understanding EATMN resiliency since they emphasize the network manager's role in actively managing unforeseen disruptions and their knock-on effect throughout the tactical phase.

Table 5-3 Statistics of active regulations for detected non-nominal States

Dates	Total (Pre-tactical)	Total (Tactical)	Weather	ATC Capacity	ATC Routing	Aerodrome Capacity	Others
22. Jul:							
Counts	123	348	72	143	71	21	41
CDF (x)[a]	***	***	0.925	0.989	0.937	0.514	0.710
12. Aug:							
Counts	128	285	6	130	80	32	37
CDF (x)	***	***	0.206	0.962	0.987	0.957	0.587

a. CDF (Cumulative Distribution Function) of observed number of regulations (or counts). Each regulation type has an identical CDF function derived from kernel density estimation (micro analysis).

Estimated distribution functions (Figure 5-1) is an asset in comparing different regulation types but also reveal characteristics of regulation types. Aerodrome capacity and ATC staffing regulations are less frequent at network level compared to ATC capacity, which proved to be the dominant regulation type. Strategic restrictions on resources and infrastructures led to a unimodal function for routing, staffing and aerodrome (airport) capacity regulations. In contrast, more stochastic tactical imbalances reflect in multimodal distributions of weather and ATC capacity regulations. Such figures are an asset to evaluate and compare resilient changes at strategic level. For instance, it is evident that the current network planning is more resilient toward ATC capacity regulations and least against weather induced disruptions because the curve for ATC capacity is the flattest and the estimated curve for weather is multimodal and skewed. A resilient long-term planning will ideally flatten the estimated curve for weather too. Needless to say, the probability values are small at first glance, but this is mathematically expected as the area below the curves has to be equal to one in probability functions.

Furthermore, possible correlations between different regulation types is investigated by comparative heatmaps of pairwise combinations as in Figure 5-2. The results opened new discussions; for example, the stretched increasing pattern in Figure 5-2a suggests the chance of a causal relationship as if a certain range of staffing regulations induces ATC capacity regulations over the network. However further studies are needed to evaluate such a causal relationship between any pair of regulation types. For instance, the geographical locations and traffic volumes at which these regulations are activated should be investigated since it possibly reveals potential bottlenecks or traffic flow knock-on effects.

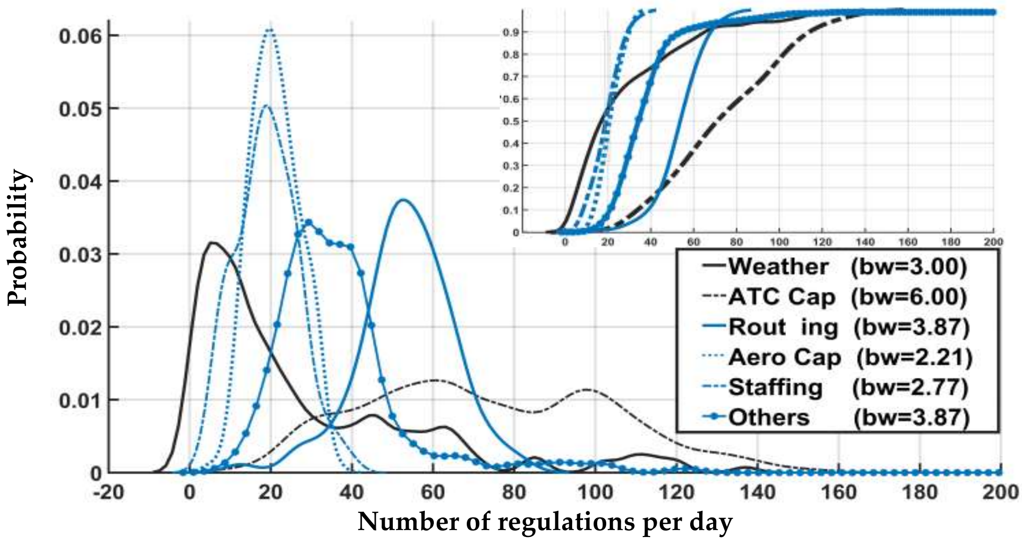
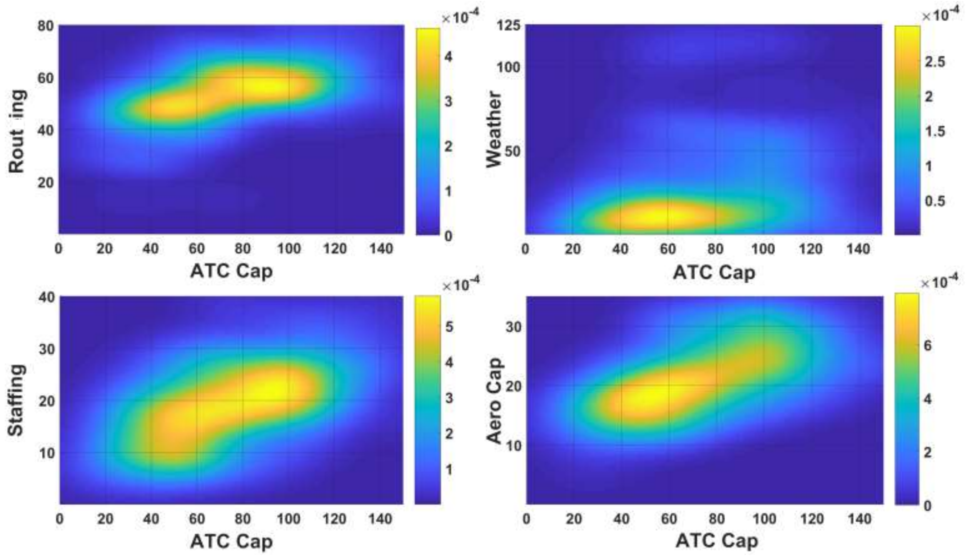
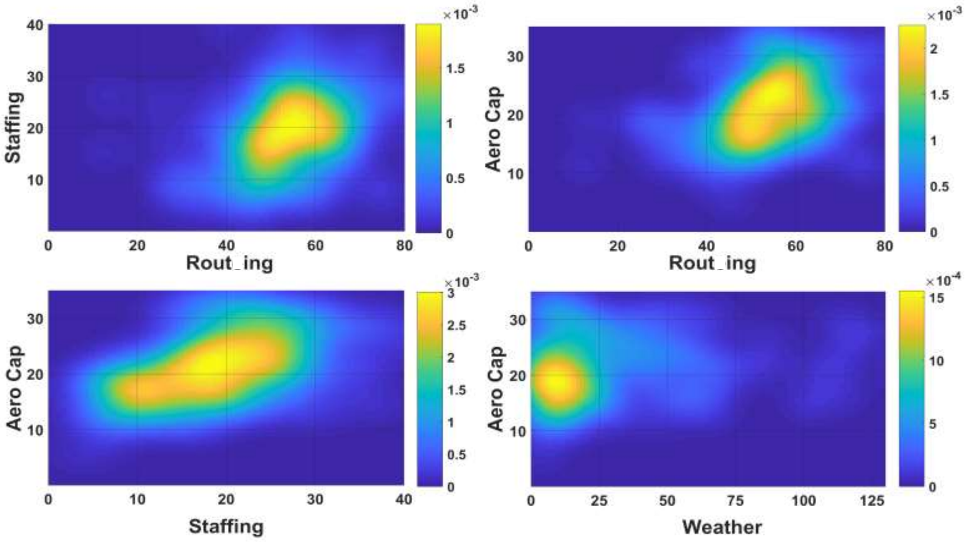


Figure 5-1 Estimated probability density functions with selected bandwidth for each regulation type (Cumulative Distribution Functions are plotted as inset with same dimensions on base plot axes). The figure shows identical characteristics of each regulation type. As mentioned before, the bandwidth parameter is tested and this figure shows the final set of curves that is divided into corrected (black) and optimal (blue) curves *[61]*.



a. selected comparison of regulations against ATC Capacity regulations



b. selected pairwise comparison of regulations – Dominant values

Figure 5-2 Pairwise comparison of EATMN regulations based on their type (Probability density estimate of bivariate data from May to Oct. 2017). a. the heatmaps suggest direct correlation between ATC capacity and Staffing compared to significantly small correlation between weather and ATC capacity regulations. b. Most dense probabilities from all possible pairwise-combinations; Low variance of aerodrome capacity and ATC staffing regulations result in higher estimated probability densities. The effect of different data-ranges is also magnified in weather vs. Aerodrome capacity regulations [61].

A comparative review of regulations contributes to a better understanding of network behavior. For instance, the chances of having simultaneous weather issues and ATC capacity regulations are so rare as if weather-induced ATFCM measures also remedy ATC capacity limitations (Figure 5-2a).

At the network level, ATC capacity regulations are a big part of tactically implemented measures with different influences on other regulation types. For instance, an increase in the number of ATC capacity regulations is most likely accompanied by a rise of ATC staffing regulations (at relative scales). This is consistent with the general rule that ATC capacity regulations are implemented when demand exceeds or complexity reduces expected capacity and although one traffic controller will have reduced complexity but such a regulation affect the entry rates for other segments of flight routes too.

In other words, the increased complexity can intensify unplanned staff shortages in adjacent ACCs (cause of ATC staffing regulations). Likewise, staff shortages may reduce the expected capacity within an ACC.

In contrast a rising pattern in the number of ATC capacity regulations is not expected to be observed with a constant increase of aerodrome capacity regulations. This is aligned with the fact that flights, as the key element of traffic flows, bind the en-route and airport demands. The implication is that imposed limitations on en-route capacity will affect demand for aerodrome capacities, especially when the large scale of the network is considered (that also include arrival airports).

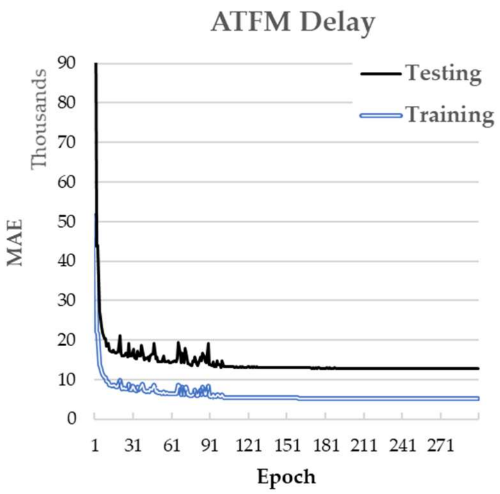
Another observation for less frequent types of regulation is the effect of limited span and low variance of their data (Figures 5-2b and 5-1) that result in higher combined probability densities (from Figure 5-2a to 5-2b, the scale of density bar moves from 10^{-4} to 10^{-3}). Nevertheless, same speculating on causal relationships can be made as in the heatmap for ATC staffing and aerodrome capacity regulations. However, in case of large difference in data-ranges, this type of comparison is less productive (e.g. the heatmap for weather vs. aerodrome capacity regulations).

5.2. Disruption prediction (DCNN)

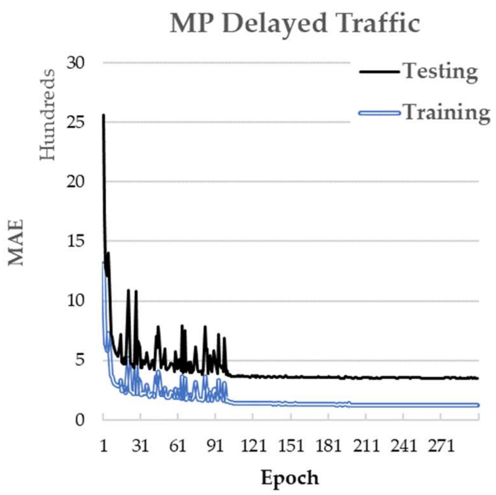
Section 3.3.1 (Feasibility of machine learning approach) presented the results of applying SNN and RFR on regulation data. Moving forward from predicting ATFM delay per ACC, the thesis scaled up to a more comprehensive study by considering different supervised learning models on network scale. The experiments led to selecting a RF model as the baseline model and a deep convolutional neural network (DCNN) was proposed by benefitting from both CNN and SNN layers.

Both baseline model and DCNN trials were trained on NMIR data with similar pre-processing steps. The convergence of DCNN model is depicted by Figure 5-3. The model performs and converges regardless of size of data input (similar trends for both train and test sets). Along with expected shorter learning time and higher errors for test set, it has been observed that after 100 epochs the learning curve has been flattened.

In comparison to RF (Table 5-4) as baseline, DCNN delivers a significantly improve performance. Predictions for the low category of target values gained maximum benefit. MAPE improved 70% for delayed traffic and 60% for delay. The efficiency of the model specially in low category demonstrates the advantage of introduced weighting method. In general, the proposed architecture successfully improved the results by 50% (overall category for delay).



(a)



(b)

Figure 5-3 Learning curve of DCNN through training and testing phase. (a) Delay, (b) delayed traffic *[62]*.

Table 5-4 Performance of DCNN model vs RF in testing phase.

Category	Delay [a]		Delayed traffic [a]	
	DCNN	RF	DCNN	RF
Low	28.3	74.15	17.21	55.95
Nominal	15.28	26.2	9.18	17.31
High	12.56	17.65	5.04	11.64
Overall	17.89	36.18	10.06	25.09

[a] measured by MAPE metric.

Since MAPE is calculated based on relative percentage values, absolute errors are plotted to investigate any patterns in wide range of target values. Figure 5-4 provides precision scatter plots for both delay and delayed traffic in two columns. DCNN outperforms RF and smoothly predicts the target values regardless of their category. RF scatter plot shows more dispersion as the target values grow but DCNN delivers a steady quality of predictions.

Such a performance can be contributed to many factors. A RF model with optimized hyperparameters has a less complex structure and learn based on a decision tree with pure probabilities of an accurate prediction. On the other side, DCNN has the advantages of convolution layers and SNN in its architecture. While CNN layers work on spatiotemporal features, the added stack of SNN layers enable the model to expect disruptive dynamics in tactical phase (compared to a typical CNN model). These layers are reinforced by focus on temporal characteristics of regulations by added vector of daily features. Since the architecture intentionally focus more on temporal dimension, a validation on different data samples is performed next.

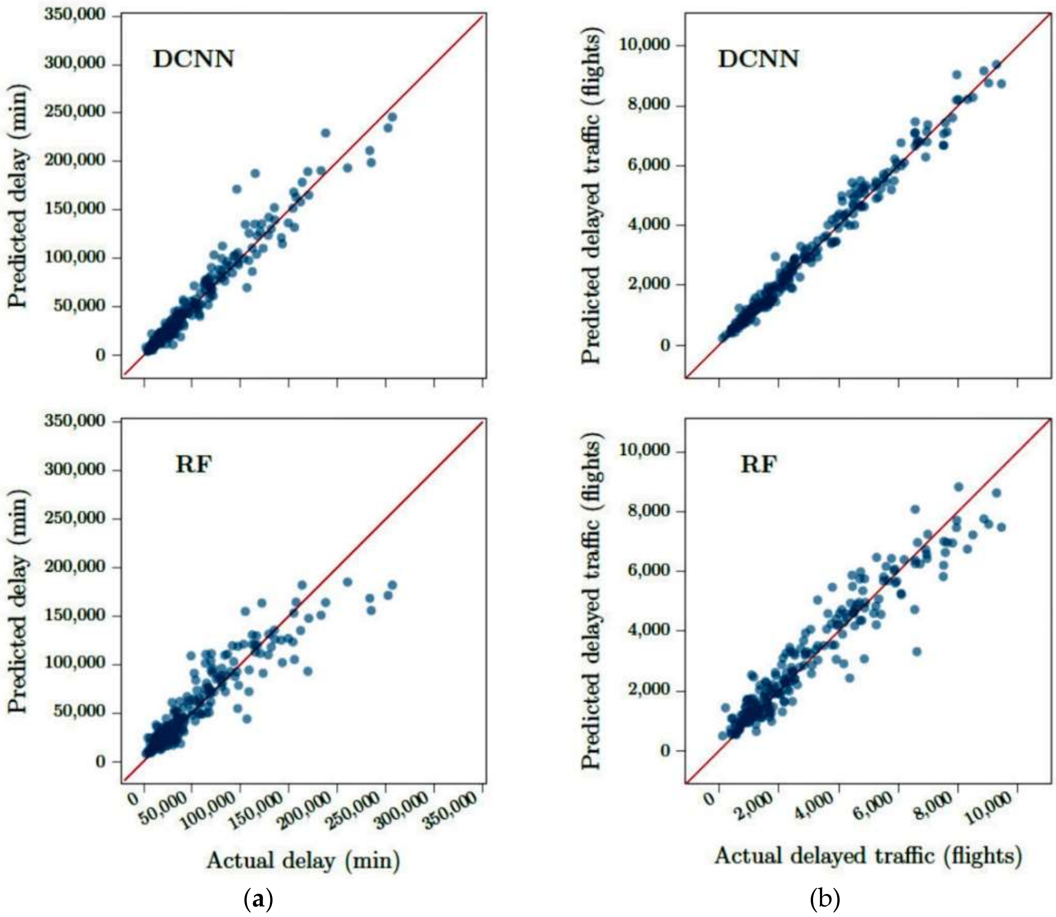


Figure 5-4 Comparative scatter plots of prediction quality. Plots on top are from DCNN and the bottom plots are from baseline model (i.e., random forest): (a) delay, (b) delayed traffic; *[62]*

Validation and discussion

The operational understanding on sequential nature of network disruptions is evident. But the proposed data driven approach rely on big data sets. Therefore, the model was trained on post-operational data of two consecutive years to predict the next year. Metrics in Table 5-5, imply that the model performed better in predicting 2018. Although 2018 is reported to be the highest figure of delay in recent years, such a behavior can be interpreted as a result of model dependency on input vector, i.e. number of pre-tactical regulations. In 2018 on average, more pre-tactical regulations were implemented per day and more traffic volumes (more ACCs) were engaged. These characteristics enrich the input data specially for CNN part of DCNN.

Table 5-5 Validation results of DCNN model, *[107]*

Train set	Target	Delay			Delayed traffic		
		MAPE[a]	R^2	MAE[b]	MAPE[a]	R^2	MAE[c]
2015-2016 (70%)	2017	34.06	0.72	13 273	16.56	0.89	400
2016-2017 (70%)	2018	21.47	0.91	11 139	13.47	0.93	438

[a] in percentage, [b] minutes, [c] flights.

In order to further investigate the impact of pre-tactical regulation counts on model, model performance is plotted in different AIRAC cycles. Figure 5-5 provides the average daily MAPE values for both delay and delayed traffic predictions. Despite high load of traffic and delay over the summer season (AIRAC 5 to 10), DCNN is more accurate over these periods (scatter plots for errors are given in Figure A-1). The descending pattern of MAPE over summer suggests that the number of regulations is a key driver in prediction accuracy specially in absence of traffic data. This is expected since proposed architecture is based on NNs to capture nonlinearity and more regulations indicates more data points from dynamic disruptions. The inverse pattern of prediction accuracy and actual delay (inset in Figure 5-5) also confirms that DCNN is more affected by number of regulations and perform better in summer. Another observation from Figure 5-5 is that such an impact (regulation counts) seems to dominate the effect of different ranges of delayed traffic and delay. During summer in 2018, high number of regulations canceled out the gap between the quality of predictions for delayed traffic and delay.

Other external factors can also affect the quality of predictions. For instance, the observed errors in prediction of summer 2017 might be an effect of a change in delay calculations, which was implemented by EUROCONTROL from April 2016 onwards [148].

Nevertheless, the enhanced prediction capability of DCNN model compared to RF is clear and seasonal patterns are aligned with dynamics of EATMN in different AIRAC cycles. For operational use cases, the prediction is most relevant for daily values and not a full year, hence the model is expected to generate smaller errors when compared to cumulative values of MAPE.

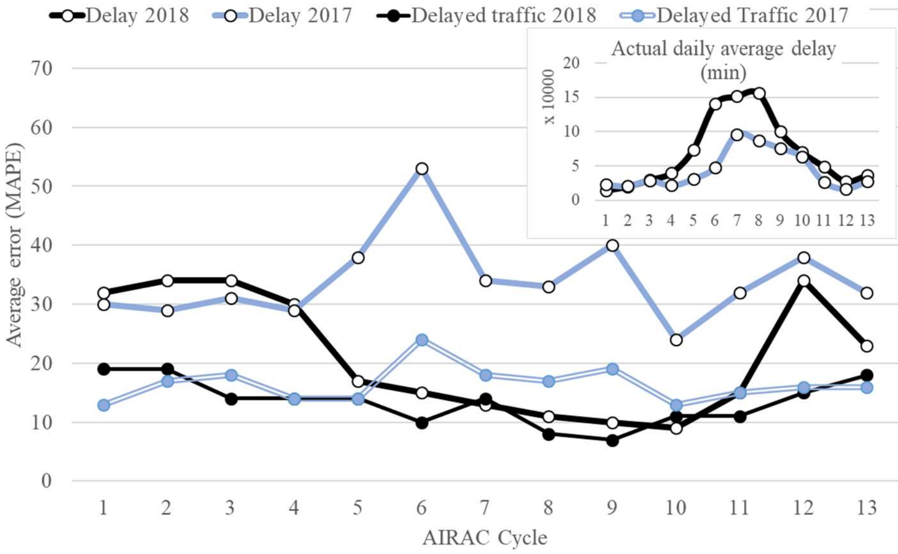


Figure 5-5 DCNN validation: prediction errors in different AIRAC cycles, *[62]*. More regulations in 2018 (especially during summer season) provide better prediction quality regardless of the expected high values for both delay and delayed traffic

5.3. RNN validation and comparison with DCNN

Results from previous sections laid out the proposed mechanism for modeling EATMN as a resilient system and the procedure in which capacity regulations were given to DCNN model to predict daily ATFM delay (and daily delayed traffic). In essence, DCNN model is measuring the magnitude of network disruptions.

DCNN took the pre-tactical regulations and predict end of the day situation. However, there are many regulations that are implemented at the tactical phase upon request to revive back from unexpected DCB issues. Therefore, the capability of NMOC to predict the impact of each proposed tactical regulation is crucial in setting the operations on the resilient path.

Since the tactical regulations are from an interactive nature, the proposed RNN (which is ideal for time series) was fostered through different aspects. These include cyclic transformations, added AIRAC cycles that enrich NMIR raw data. Also, through dynamic learning rates, Adam optimizer and different activation functions, the proposed RNN architecture was reinforced in its prediction task.

Table 5-6 summarizes the performance of the customized RNN on training set from yearly datasets. The quality degradation on validation set is not a surprise since training and validation set are different in their sizes and 2018 is the recorded year with the most regulations and scattered incidents of high delay regulations. unaccustomed to having 10 to 20 percent of same dataset to be used for validation purposes, RNN model is pushed to its limits with 2018 regulations that is almost double in size. But still RNN shows rather a stable performance with a MAPE of 35.76%.

Table 5-6 Validation results of RNN model (2017 for 2018), *[145]*

Set	Size	MAPE [a]	MAE [b]	RMSE [b]
Train 2017 (70%)	36 491	10.20	14.60	127.92
Validation 2018	66 136	35.76	34.05	168.47

[a] in percentage, [b] minutes.

The scatterplot of this trial in Figure 5-6, reveals the extended range of delays and hints about optimal range of target values. In other words, RNN model performs best on regulations with expected delays from 5000 to 12000 minutes. Early findings of the thesis showed that specific types of regulations (out of 14 different type) are more penalizing to the traffic (e.g. weather and ATC-Capacity regulations) and since RNN lacks the capabilities of a CNN in feature exploitation, the quality of predictions is decreasing for less frequent regulations with smaller expected values of delay. Therefore, it is a challenge to the model to relate different regulations (e.g. different locations, types, dates) with the same target value of delay.

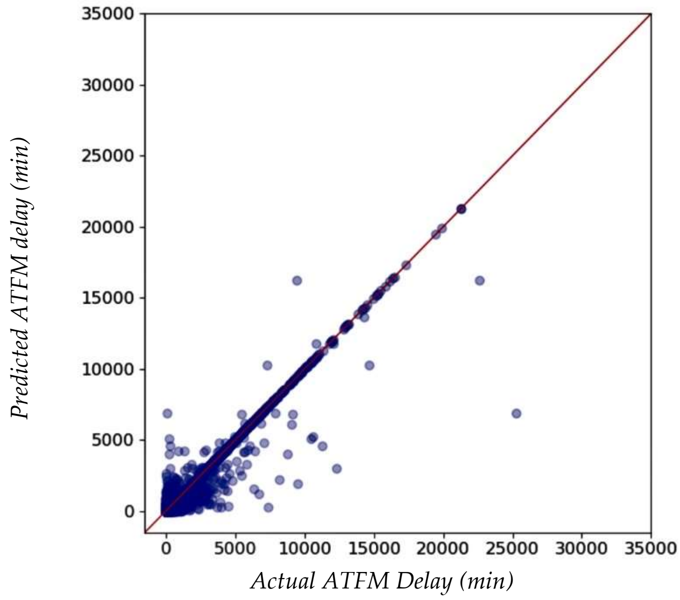


Figure 5-6 Scatter plot of prediction quality in validation phase, *[145]*

In the second trial, more training data is used to control the effect of different regulation types. But because of increasing yearly patterns of network demand and relatively constant figures of capacity, only data from past two consecutive years are considered. Such a limitation is absent in other use cases of RNN on time series such as temperature predictions.

Table 5-7 completes previous table and firstly offers model performance on a homogeneous 30% validation set with lowest MAPE. For larger validation sets, the approach of using larger training sets proved to be efficient. In these experiments, both validation and training sets are intentionally kept at similar sizes to avoid relaxing the model in prediction task such that calculated MAPE are more comparable (for an average of 60 thousand predictions).The improvement of predictions is evident for 2018 with a MAPE of 25.49 (-10%).

Table 5-7 Validation: RNN performance in predicting ATFM delay per regulation [145]

Train set	Size	Validation set	size	MAPE (%)
2017 (70%)	36 491	2017 (30%)	15 640	12.80
2017 (70%)	36 491	2018 (100%)	66 136	35.76
2015-2016 (70%)	52 330	2017 (100%)	52 132	16.95
2016-2017 (70%)	66 082	2018 (100%)	66 136	25.49

There is an intrinsic complexity with data from 2018 that makes the prediction task for the RNN to be more challenging compared to other years. In fact, Regulation statistics (Figure 2-1 & Table 5-8) show that there is a significant increase in number of regulations, delay and delayed traffic. It should also be reminded that RNN is predicting individual regulations (resilient path) compared to daily values for DCNN model (that contributes to network state detection).

Table 5-8 Summarized NMIR statistics on regulations (2015-2019)

Year	Counts	Reg Duration (min)	ATFM Delay (min)	MP Delayed Traffic
2019	62 798	10 203 414	24 132 723	1 444 527
2018	66 136	10 201 611	25 623 133	1 382 176
2017	52 131	7 462 949	15 886 900	891 374
2016	42 270	5 827 706	15 576 691	814 678
2015	32 484	4 706 664	14 065 108	680 981

Hence, the result of RNN are required to be expressed in cumulative daily values when compared to DCNN results. In Table 5-9, MAPE values are recalculated for RNN and the achieved accuracy of 97% is not a result of a trained model on transformed daily values. In fact, the positive and negative residual errors for each predicted regulation cancel each other on daily values and a portion of RNN's higher levels of accuracy is a result of this calculation.

Table 5-9 Validation: RNN vs. DCNN performance in predicting daily ATFM delay

Train set	Validation set	Model	Accuracy (%)
2015-2016 (70%)	2017 (100%)	RNN	97.59
2015-2016 (70%)	2017 (100%)	DCNN	65.94
2016-2017 (70%)	2018 (100%)	RNN	97.88
2016-2017 (70%)	2018 (100%)	DCNN	78.53

Another factor is the granularity of the data which was much higher for RNN model against the cumulative input for DCNN. Practically, more data points were available to RNN compared to DCNN and since both models are derivatives of neural networks, RNN has an advantage in feedforward learning as learned weights are tailored by more iterations. This claim is even stronger for predicted days of summer season in which each day has on average of more than 100 regulations (1 data point for DCNN vs. 100 data points for RNN); a fact that also leads to lower variance in predicted daily values.

6. Conclusion and outlook

In sections 3.2 and 5.1 (methodology and results on network state), two main topics have been covered. Firstly, the concept of resilience was customized in domain of European Air Traffic Management (ATM) with a focus on the boundaries of resilience and robustness. Secondly, in an attempt to improve tactical network resiliency based on performance variability, the possibilities to define an ATM network state by a deeper understanding of its performance dynamics were explored. This was achieved by network state definition according to implemented local DCB solutions (i.e. capacity regulations). Definition of network states expands our understanding of the enriched regulation data as a feedback loop on tactical plans and intensifies the importance of NMOC role as the network manager that monitors all European ACCs.

The selection of regulations was tailored by a survey on different databases. ATFCM Notification Messages (ANM) data were selected since they reflect the result of corrective DCB measures in tactical phase with the advantage of being published by push messages for all ATM actors. To evaluate the quality of information an analytical tool (i.e. ANMStat, see Annex C.1.) was developed in MATLAB. It enabled preliminary statistical evaluation of ANM messages. Derived reports include outlier detection (days with extreme conditions), or estimated probability distributions of different regulation types. ANMStat is capable to check various aspects of ANM messages to understand the prominent data features (statistical, temporal and geographical) and decide on target time horizon and most active FMPs for further analysis and case study designs.

The methodology to define network state is comprised of a macro and a micro analysis based on quantitative measures. Results were investigated further to characterize network states with regard to regulation types and potential geographical patterns. Similarly, with acquired knowledge from ANMStat, results were further evaluated according to different data spans. From an operational standpoint, there are known monthly patterns (as in summer season) that ATM actors anticipate. Therefore, different reference thresholds are tested to understand short-term trends in traffic demand and to capture possible effects of different sample sizes.

In order to establish control thresholds, probability density functions (PDFs) of major regulation types was estimated by normal kernel smoothing method. These thresholds were evaluated by comparison of a use case with official reports. More specifically two detected non-nominal dates were checked against both ATFCM daily summary reports and published Initial Network Plans in micro analysis.

In order to address other levels of resilience, spatial mapping plots were then provided to indicate a geographical reference for different type of regulations. Since the ATM network is more flexible in pre-tactical and strategic phases such dedicated maps contribute to customized mitigation plans to improve network resiliency. Moreover, estimated PDFs were tested by pairwise-comparison plots (Figure 5-1) in search of potential causal relationships. Observations showed potential correlations among specific type of regulations (staffing and ATC capacity for instance). More deterministic findings are an asset to network manager role to anticipate secondary regulations at network level induced by large scale requests for a specific type of regulation at local levels. However, dedicated studies are needed to monitor the consistency of such relationships in different network states.

Guided by expert opinions (from SESAR project), the research then focused on reviving measures (resilient path). In section 3.3.1 (feasibility of machine learning approach), the aim was set on predicting impact of correction measures (i.e. regulation) rather than understanding causes of disruptions and preventive mechanisms. The thesis invests on learning algorithms for this task as it also contributes to achieving accurate situational awareness. Consequently, different supervised learning methods are applied to regulation data, in contrast with other studies that select city pairs and rely on traffic and demand figures.

On a network level, since the casual relationships between regulation is parked for future studies, learning methods are chosen to benefit from prediction based on training data rather than analytical methods or simulations with higher model complexities. For the feasibility task, complexity of the prediction problem was reduced in two aspects. On one hand, only a selected FMP (rather than whole network) is chosen because it is the key decision maker in coping with DCB issues by capacity regulations. On the other hand, data is pre-processed to modify input vectors with a combination of extracted indicators and parameters based on operational understanding of ATFM disruptions' severity. Another advantage of learning approach with regulation data is its structure that eliminates the need for classification methods and provides a straight forward approach on regression methods.

Even at a reduced dimension of a single FMP (Langen) with the historical data from 2016 and 2017, results [35] confirmed the added value of delay prediction by neural networks and random forest regression.

In order to demonstrate the added value of results to network manager position, a tool was coded in python (NetRes) that offers extracted indicators, network state mechanism and some intuitive graphs in a stand-alone interface. NetRes (annex C.2) takes regulation data from two sources: ANM messages and NMIR. Both databases almost offer similar data structures but NMIR is post-operation and ANMs are tactical. NMIR only offers daily logs of regulations while ANM is being updated through the day of operations with push messages. NetRes is able to process each type of data for different tasks, NMIR to calculate thresholds and ANMs for state definition. The tool is presented and delivered to SESAR community in PJ09 project (Advanced Demand Capacity Balancing) and is documented under 'Solution 1 – Network prediction and performance' [149].

The study is then continued to invest more on learning algorithms with application of different supervised learning methods at network scale. The prediction task at this stage consider EATMN (and not only one FMP) and data is being fully considered without omitting any feature of its structure (i.e. no cumulative indicators). After multiple design iterations, a deep architecture based on convolutional and sequential neural networks (i.e. DCNN) is finalized to serve as the disruption prediction model.

Through a deep learning process both spatial and temporal dimension of regulation data is extracted by the model. The proposed model proved to be efficient in predicting both delay and delayed traffic as the two consequences of applying capacity regulations. DCNN significantly improved the prediction quality in comparison to an optimized RF model as the baseline model. The data driven approach to predict daily delay, gives DCNN the advantage to perform better in more dynamic situations, since a busy day with more pre-tactical regulations provide more data points as the model input.

Finally, in search of evaluating the impact of individual regulations as corrective actions (section 4.2), a recurrent neural network is designed with a focus on sequence of regulations. The designed RNN, predicts regulation compared to DCNN which predicts daily values at the end of the day. In other words, proposed RNN enables the impact assessment for each regulation on demand. After understanding the network state and evaluating the level of disruption at a network level by DCNN, the resilient path to revive the network requires the RNN capability to predict the impact of each regulation as a corrective measure to cope with DCB disruptions.

In summary, this exploratory research served as an attempt toward EATMN resiliency. Conceptually, the study demonstrates that EATMN can be modeled as a resilient system. Operationally, the thesis offers an alternative network prediction based on capacity regulations and supervised learning methods. The results of the thesis are offered to both academic experts (through peer reviewed publications) and industrial partners (in course of a SESAR project).

6.1. COVID-19 pandemic

This study aimed for ATM resiliency by providing better networkwide situational awareness in Demand Capacity Balancing (DCB). Incidents such as the volcano eruption in 2010 and COVID-19 pandemic in 2020 are challenges to different levels of resilience (Table 6-1). The volcano eruption was a safety risk in the pre-tactical phase (adaptive level) and COVID-19 is a large-scale issue in strategic phase (absorptive level). Similar incidents are mostly regarded as safety challenges while the scope of current research is mainly on network performance and DCB disruptions (restorative level). Perhaps, challenges like snowfalls in march 2013 are better examples to elaborate on the restorative level. Heavy snow hit Chicago airport with 9,2 inches (23 cm) of snow that showered on one day (5th of march) [150] . Only some days after, on 12th of march, snowfalls made Frankfurt airport to shut down its operations. Same weather system continued to disrupt flights across northwestern Europe leading to up to 50% cancellation rate at some airports [151].

Table 6-1 Resilience Levels

Level	ATFCM Phase
Absorptive	strategic
Adaptive	pre-tactical
Restorative	Tactical

Moreover, safety disruptions such as volcano eruption and pandemics have different propagation and transition mechanisms. The volcanic eruption started on April 14th and persisted for six weeks (end of May 2010). Volcanic ash is considered as a known hazard to aviation and at the time of incident there were already mitigation plans based on satellite measurements and advanced dispersion models. Yet at the first day of the eruption 8200 flights were cancelled. Although, the sudden eruption was recognized at the restorative level, the impacts persisted and escalated to adaptive level and even at absorptive level led to some corrective procedures in mitigation plans. Almost a year later another volcanic eruption from 23rd to 25th May 2011 happened (the Grimsvötn crisis, [152]) and the strategic corrections efficiently limited the cancellations to 900 flights (out of a total of 90,000 expected flights).

Generally, a volcanic ash is a flight safety issue that in DCB terms, translates to an unexpected loss of capacity (airspace closure) which can be mitigated by flight cancellation, rerouting and capacity regulations. Therefore, such incidents are much closer to the scope of this study compared to COVID-19 pandemic.

COVID pandemic was a passenger safety risk and unlike a volcanic eruption, no technical/operational issues or capacity degradations were triggered. In contrast to emergent disruptions it did not directly affected the European aviation at a specific date and didn’t start at the restorative level. In fact, instead of arising from restorative to absorptive level with strategic changes in case of a volcanic ash; the pandemic took the opposite direction and disseminated to aviation sector from strategic decisions by European Commission (EC) and head of states (limiting the spread of the virus through means of transportation). Decided policies caused a reduction in airport operations and route restrictions. Line of such decisions continued even a year after by network manager (EUROCONTROL) through 6-week mitigation plans [153], i.e. strategic phase and adaptive level of resilience.

Moreover, ML approaches depend on sufficient data to learn. The aforementioned cases are exceptions and there is no previous situation that learning models can learn from. In both cases, the system is not suffering from delay but mostly flight cancelation, which is not included in the scope of this work.

6.2. Future works

As expected from an exploratory research, the thesis exposes a number of challenges that are required to be studied in future. The follow up studies can be categorized into different approaches:

- Concept: this thesis demonstrated the realization of performance variability is a key aspect for a resilient network. The next step can be to further investigate control mechanism on top of monitoring measures. Control mechanisms can be better investigated by an agent base model that considers FMPs as controlling agents. Such a study can still rely solely on regulation data (even in absence of traffic data) because regulations are mainly proposed by FMPs and they are not applied to OD pairs but on target traffic volumes;

- Method: the customized method of DeepCNN has two CNN and SNN components. This architecture can be improved further by integrating convolutional layers to a RNN. The resulting architecture is highly expected to improve prediction quality especially by relying on achieved results from RNN model in chapter 4. Current state of study shows that regulation data can be considered for LSTM as time-series;

- Scope: since the results of pre-tactical phase of ATFCM are published as Initial Network Plan (INP) it is possible to extend the model to adaptive level of resilience. INP can be merged with traffic data from flight plans in pre-tactical phase. Flight plans can be acquired from EUROCONTROL's Demand Data Repository (DDR2);

- Data: include and merge more datasets on capacity and traffic situations into the learning models. These datasets can be acquired from Demand data repository (DDR). More specifically planned flight plans and actual flight trajectories can complement the regulation data from NMIR database. Such an approach incorporates ideas from multimodal data fusion techniques such as early or late fusion and are more beneficiary in network state identification.

References

[1] FAA, EC and EUROCONTROL, "2015 Comparison of ATM-related performance: U.S. - Europe," EUROCONTROL Performance Review Unit and FAA-ATO, 2016.

[2] EUROCONTROL, "EUROCONTROL Seven-Year Forecast," 2019.

[3] V. Gollnick and C. M. Weder, "Die Covid-19-Pandemie–Anlass oder Ursache der Konsolidierungsphase in der Luftfahrt," *Internationales Verkehrswesen,* vol. 73, no. 1, pp. 24-29, 2021.

[4] Airbus S.A.S, Global Market Forecast, Cities, Airport & Aircraft 2019-2038, Lavaur: Art & Caractère, 2019, p. 8.

[5] Boeing, "Commercial Market Outlook 2019 – 2038," Boeing , Arlington, 2019.

[6] EUROCONTROL Performance Review Commission (PRC), "Performance Review Report: An Assessment of Air Traffic Management in Europe during the Calendar Year 2017," EUROCONTROL, Brussels, 2018.

[7] EUROCONTROL, "EUROCONTROL ATM Lexicon," [Online]. Available: www.eurocontrol.int/atmlexicon. [Accessed 12 Feb 2020].

[8] ICAO, "Long-Term Traffic Forecasts: Passenger and Cargo," International Civil Aviation Organization, Montreal, 2018.

[9] FAA & EUROCONTROL, 2017 Comparison of Traffic Management-related Operational Performance U.S./Europe, EUROCONTROL Performance Review Unit and FAA-ATO, 2019.

[10] EUROCONTROL, ATFCM USERS MANUAL, 22.1 ed., Brussels: EUROCONTROL, 2018.

[11] EUROCONTROL, SESAR PJ09 OSED-SPR-INTEROP – Part I, 00.02.01 ed., Brussels: EUROCONTROL, 2019.

[12] EUROCONTROL Performance Review Commission (PRC), "Performance Review Report: An Assessment of Air Traffic Management in Europe during the Calendar Year 2019," EUROCONTROL , Brussels, 2020.

[13] GIS-ICAO, "Air Traffic Flow Chart 2018," [Online]. Available: https://gis.icao.int/sgisweb/TRAFFICCOLOR/JPEG2/FLOWCHART2018greenblackOP.pdf. [Accessed 06 September 2019].

[14] ICAO, Manual On Collaborative Air Traffic Flow Management (Doc 9971), 3rd ed., Montreal: International Civil Aviation Organization, 2018.

[15] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari and S. Stadtler, "Can deep learning beat numerical weather prediction?," *Philosophical Transactions of the Royal Society A,* vol. 379, no. 2194, 2021.

[16] E. M. Lorenz, "Deterministic Nonperiodic Flow," *Journal of the Atmospheric Sciences,* vol. 20, pp. 130-141, 1963.

[17] Wetterzentrale. [Online]. Available: https://www.wetterzentrale.de/. [Accessed 12 10 2020].

[18] ISO\IEC\IEEE, Systems and software engineering - System life cycle processes, 1st ed., Geneva: Institute of Electrical and Electronics Engineers, Inc., 2015.

[19] European Union, *Regulation (EC) No 549/2004 of the European Parliament and of the Council of 10 March 2004 laying down the framework for the creation of the single European sky,* Official Journal of the European Union, 2004.

[20] ISO\IEC\IEEE, Systems and software engineering — Life cycle management, 1st ed., Geneva: Institute of Electrical and Electronics Engineers, Inc., 2018.

[21] EUROCONTROL, "about us," [Online]. Available: https://www.eurocontrol.int/about-us. [Accessed 13 05 2020].

[22] Single European Sky ATM Research Joint Undertaking, "About SESAR - History," [Online]. Available: https://www.sesarju.eu/discover-sesar/history. [Accessed 14 03 2020].

[23] Single European Sky ATM Research Joint Undertaking, *European ATM Master Plan,* Brussels: SESAR Joint Undertaking, 2015.

[24] J. Pariès, "Chapter 4: Complexity, Emergence, Resilience...," in *Resilience Engineering: Concepts and Precepts*, Ashgate, 2006, pp. 43-53.

[25] European Union, *Regulation (EC) No 552/2004 of the European Parliament and of the Council of 10 March 2004 on the interoperability of the European Air Traffic Management network (the interoperability Regulation),* Official Journal of the European Union, 2004.

[26] B. Campanelli, P. Fleurquin, A. Arranz, I. Etxebarria, C. Ciruelos, V. M. Eguíluz and J. J. Ramasco, "Comparing the modeling of delay propagation in the US and European air traffic networks," *Journal of Air Transport Management,* vol. 56, pp. 12-18, 2016.

[27] R. Francis and B. Bekara, "A metric and frameworks for resilience analysis of engineered and infrastructure systems," *Reliability Engineering and System Safety,* vol. 121, pp. 90-103, 2014.

[28] R. M. Hoffman, "A Generalized Concept of Resilience," *Textile Research Journal,* vol. 18, pp. 141-148, 1948.

[29] C. S. Holling, "Resilience and Stability of Ecological Systems," *Annual Review of Ecology and Systematics,* vol. 4, pp. 1-23, 1973.

[30] D. Henry and J. E. Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering and System Safety,* vol. 99, pp. 114-122, 2012.

[31] E. Hollnagel, D. D. Woods and N. Leveson, Resilience Engineering: Concepts and Precepts, Aldershof: Ashgate, 2006.

[32] D. Furniss, J. Back, A. Blandford, M. Hildebrandt and H. Broberg, "A resilience markers framework for small teams," *Reliability Engineering and System Safety,* vol. 96, pp. 2-10, 2011.

[33] EUROCONTROL, "A White Paper on Resilience Engineering for ATM," EUROCONTROL, Brussels, 2009.

[34] A. Cook, L. Delgado, S. Cristobal and A. Blanch, "What Cost Resilience?," in *Forth SESAR Innovation Days,* Madrid, Spain, 25-27 November, 2014.

[35] R. Sanaei, A. Lau, F. Linke and V. Gollnick, "Machine Learning Application in Network Resiliency based on Capacity Regulations," in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, San Diego, USA, 8-12 September, 2019.

[36] O. Gluchshenko, "Definitions of Disturbance, Resilience and Robustness in ATM Contex," DLR, Institute of Flight Guidance, Braunschweig, 2012.

[37] J. Clausen, A. Larsen and J. Larsen, "Disruption Management in the Airline Industry - Concepts, Models and Methods," Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Lyngby, 2005.

[38] E. J. Rykiel, "Towards a definition of ecological disturbance," *Australian Journal of Ecology,* vol. 10, no. 3, pp. 361-365, 1985.

[39] Y. Zhou, J. Wang and H. Yang, "Resilience of Transportation Systems: Concepts and Comprehensive Review," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 12, pp. 4262-4276, 2019.

[40] S. Yoo and H. Yeo, "Evaluation of the resilience of air transportation network with adaptive capacity," *International Journal of Urban Sciences,* vol. 20, no. sup1, pp. 38-49, 2016.

[41] M. Janić, "Reprint of “Modelling the resilience, friability and costs of an air transport network affected by a large-scale disruptive event”," *Transportation Research Part A: Policy and Practice,* vol. 81, pp. 77-92, 2015.

[42] M. Janić, "Modeling the resilience of an airline cargo transport network affected by a large scale disruptive event," *Transportation Research Part D: Transport and Environment,* vol. 77, pp. 425-448, 2019.

[43] S. Dunn and S. M. Wilkinson, "Increasing the resilience of air traffic networks using a network graph theory approach," *Transportation Research Part E: Logistics and Transportation Review,* vol. 90, pp. 39-50, 2016.

[44] A. Voltes-Dorta, H. Rodríguez-Déniz and P. Suau-Sanchez, "Vulnerability of the European air transport network to major airport closures from the perspective of passenger delays: Ranking the most critical airports," *Transportation Research Part A: Policy and Practice,*, vol. 96, pp. 119-145, 2017.

[45] G. M. D'este and M. A. Taylor, "Network vulnerability: an approach to reliability analysis at the level of national strategic transport networks," *The Network Reliability of Transport,* pp. 23-44., 13 Jan 2003 (reprint 2017 by Emerald Group Publishing Limited).

[46] K. Berdica, "An introduction to road vulnerability: what has been done, is done and should be done," *Transport Policy,* vol. 9, no. 2, pp. 117-127, 2002.

[47] G. W. Klau and R. Weiskircher, "Chapter 15 Robustness and Resilience," in *Network Analysis: Methodological Foundations*, Berlin-Heidelberg, Springer, 2005, pp. 417-437.

[48] R. F. Vernon, "A Brief History of Resilience," in *Community Planning to Foster Resilience in Children*, Boston, Springer, 2004, pp. 13-26.

[49] K. C. Pien, K. Han, W. Shang, A. Majumdar and W. Ochieng, "Robustness analysis of the European air traffic network," *Transportmetrica A: Transport Science,* vol. 11, no. 9, pp. 772-792, 2015.

[50] H. Devoe, Thermodynamics and Chemistry, 2nd ed., Maryland: University of Maryland, 2020.

[51] C. Coetzee, D. Van Niekerk and E. Raju, "Disaster resilience and complex adaptive systems theory: Finding common grounds for risk reduction," *Disaster Prevention and Management,* vol. 25, no. 2, pp. 196-211, 2016.

[52] R. Baduel, J. M. Bruel, I. Ober and E. Doba, "Definition of states and modes as general concepts for system design and validation.," in *12th International Conference on Modeling, Optimization and Simulation - MOSIM18*, Toulouse, France, 27-29 June, 2018.

[53] A. Cook, H. A. Blom, F. Lillo, R. N. Mantegna, S. Micciche, D. Rivas, R. Vazquez and M. Zanin, "Applying complexity science to air traffic management," *Journal of Air Transport Management,* vol. 42, pp. 149-158, 2015.

[54] A. Cook, L. Delgado, G. Tanner and S. Cristobal, "Measuring the cost of resilience," *Journal of Air Transport Management,* vol. 56, pp. 38-47, 2016.

[55] E. Filippone, F. Gargiulo, A. Errico, V. Di Vito and D. Pascarella, "Resilience management problem in ATM systems as a shortest path problem," *Journal of Air Transport Management,* vol. 56, pp. 57-65, 2016.

[56] P. Andrade, C. Silva, B. Ribeiro and B. F. Santos, "Aircraft maintenance check scheduling using reinforcement learning," *Aerospace,* vol. 4, no. 8, p. 113, 2021.

[57] Q. Deng and B. F. Santos, "Lookahead approximate dynamic programming for stochastic aircraft maintenance check scheduling optimization," *European Journal of Operational Research,* vol. 299, no. 3, pp. 814-833, 2021.

[58] A. Lau, J. Berling, F. Linke and V. Gollnick, "Large-Scale Network Slot Allocation with Dynamic Time Horizons," in *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*, Lisbon, Portugal, 23-26 June, 2015.

[59] A. Lau, "Air Traffic Flow Management in the Presence of Convective Nowcasting Information (Doctoral Dissertation)," German Aerospace Center (DLR), Cologne, 2017.

[60] X. Sun and S. Wandelt, "Chapter 9 - Network Modelling and Resilience Analysis of Air Transportation: A Data-Driven, Open-Source Approach," in *Data-Driven Solutions to Transportation Problems*, Amsterdam, Elsevier, 2019, pp. 227-245.

[61] R. Sanaei, A. Lau and V. Gollnick, "A study of Capacity Regulations to define European Air Traffic Management Network States," *Transportation Planning and Technology,* vol. 44, no. 4, 2021.

[62] R. Sanaei, B. A. Pinto and V. Gollnick, "Toward ATM Resiliency: A Deep CNN to Predict Number of Delayed Flights and ATFM Delay," *Aerospace,* vol. 8, no. 2, p. 28, 2021.

[63] R. Sanaei, "Exploring Supervised Learning to Predict Air Traffic Delay in ATM Network Resiliency," in *DLR WAW Machine Learning V*, Wessling, Germany, 03-05 December. 2019.

[64] R. Sanaei, "Improving ATM Network Resilience with Machine Learning," in *DLR WAW Machine Learning IV*, Cologne, Germany, 06-07 June, 2019.

[65] R. Sanaei, "Resilient Systems or Being Cool!," in *NextGen Conference*, Potsdam, Germany, 25-27 July, 2018.

[66] E. Hollnagel, Safety-I and Safety-II: The Past and Future of Safety Management, Farnham: Ashgate, 2014.

[67] European Union, *COMMISSION REGULATION (EU) No 255/2010 of 25 March 2010 laying down common rules on air traffic flow management,* Official Journal of the European Union, 2010.

[68] E. Hollnagel, J. Leonhardt, T. Licu and S. Shorrock, *From Safety-I to Safety-II: A White Paper,* EUROCONTROL, 2013.

[69] EUROCONTROL, "NOP portal," [Online]. Available: https://www.public.nm.eurocontrol.int/PUBPORTAL/gateway/spec/index.html. [Accessed 15 Jan 2017].

[70] X. Sun, S. Wandelt and F. Linke, "Temporal evolution analysis of the european air transportation system: air navigation route network and airport network," *Transportmetrica B: Transport Dynamics,* vol. 3, no. 2, p. 153–168, 2015.

[71] E. Lehman, Elements of Large Sample Theory, New York: Springer, 1999.

[72] EUROCONTROL, "Network Manager outage on 3 April 2018: Summary brief," EUROCONTROL, Brussels, 2018.

[73] A. Cook, European Air Traffic Management: Principles, Practice and Research, Aldershot: Ashgate, 2007.

[74] G. Sreeta, M. Hansen and A. Pozdnukhov, "Decision support framework to assist air traffic management," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sacramento, USA, 25-29 September, 2016.

[75] H. Ishwaran, U. B. Kogalur, E. H. Blackstone and M. S. Lauer, "Random survival forests," *Annals of Applied Statistics,* vol. 2, no. 3, pp. 841-860, 2008.

[76] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation Research Part C: Emerging Technologies,* vol. 44, pp. 231-241, 2014.

[77] Y. J. Kim, S. Choi, S. Briceno and D. Mavris, "A Deep Learning Approach to Flight Delay Prediction," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sacramento, USA, 25-29 September, 2016.

[78] T. Kravaris, G. A. Vouros, C. Spatharis, K. Blekas, G. Chalkiadakis and J. M. C. Garcia, "Learning Policies for Resolving Demand-Capacity Imbalances During Pre-tactical Air Traffic Management," in *German Conference on Multiagent System Technologies*, Leipzig, Germany, 23-26 August, 2017.

[79] C. Spatharis, T. Kravaris, K. Blekas, G. A. Vouros and J. M. C. Garcia, "Multiagent Reinforcement Learning Methods for Resolving Demand - Capacity Imbalances," in *IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, London, UK, 23-27 September, 2018.

[80] R. B. Montes, A. D. Rojas, V. F. G. Comendador, R. A. Valdés and L. P. Sanz, "A Novel Predictability Performance Metric and Its Forecast Using Machine Learning Techniques," in *IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, London, UK, 23-27 September, 2018.

[81] L. Denoyer and P. Gallinari, *Deep sequential neural network,* arXiv preprint arXiv:1410.0510 , 2014.

[82] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, USA, 7-9 May, 2015.

[83] T. K. Ho, "Random decision forests," in *3rd International Conference on Document Analysis and Recognition (ICDAR)*, Montreal, Canada, 14-16 August, 1995.

[84] L. Breiman , "Random Forests," *Machine Learning,* vol. 45, p. 5–32, 2001.

[85] T. Hastie, J. Friedman and R. Tibshirani, The Elements of Statistical Learning, second ed., New York: Springer, 2009.

[86] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation Research Part C: Emerging Technologies,* vol. 44, pp. 231-241, 2014.

[87] T. Kistan, A. Gardi, R. Sabatini, S. Ramasamy and E. Batuwangala, "An evolutionary outlook of air traffic flow management techniques," *Progress in Aerospace Sciences,* vol. 88, pp. 15-42, 2017.

[88] N. Ivanov, N. F. J. R. S. Starita and A. Strauss, "Air Traffic Flow Management slot allocation to minimize propagated delay and improve airport slot adherence," *Transportation Research Part A: Policy and Practice,* vol. 95, pp. 183-197, 2017.

[89] A. Montlaur and L. Delgado, "Flight and passenger efficiency-fairness trade-off for ATFM delay assignment," *Journal of Air Transport Management,* vol. 83, p. 101758, 2020.

[90] T. Bolić, L. Castelli, L. Corolli and D. Rigonat, "Reducing ATFM delays through strategic flight planning," *Transportation Research Part E: Logistics and Transportation Review,* vol. 98, pp. 42-59, 2017.

[91] Y.-H. Chang, S. Solak, J.-P. B. Clarke and E. L. Johnson, "Models for single-sector stochastic air traffic flow management under reduced airspace capacity," *Journal of the Operational Research Society,* vol. 67, no. 1, pp. 54-67, 2016.

[92] L. Delgado and X. Prats, "En Route Speed Reduction Concept for Absorbing Air Traffic Flow Management Delays," *Journal of Aircraft,* vol. 49, no. 1, pp. 214-224, 2012.

[93] X. Prats and M. Hansen, "Green delay programs, absorbing ATFM delay by flying at minimum fuel speed," in *Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011)*, Berlin, Germany, 14-17 June, 2011.

[94] A. R. Odoni, "The flow management problem in air traffic control," in *NATO advanced research workshop on flow control of congested networks*, Capri, Italy, 12-18 October, 1986.

[95] S. Carlier, I. de Lépinay, J.-C. Hustache and F. Jelinek, "Environmental Impact of Air Traffic Flow Management Delays," in *ATM Seminar*, Barcelona, Spain, 02 - 05 July, 2007.

[96] L. D. L. F. LAYOS, "Statistics in focus - Transport - Air passenger transport in Europe in 2007," Eurostat (European Commission), 2009.

[97] Eurostat, Energy, transport and environment statistics, Luxembourg: Publications Office of the European Union, 2020.

[98] Performance Review Unit, "Technical note on ANSPs Productivity, ATFM delays and ATCOs working hours," EUROCONTROL, Brussels, 2020.

[99] M. Bardach, E. Gringinger, M. Schrefl and C. G. Schuetz, "Predicting Flight Delay Risk Using a Random Forest Classifier Based on Air Traffic Scenarios and Environmental Conditions," in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, San Antonio, USA, 11-16 October, 2020.

[100] Z. Guo, G. Mei, S. Liu, L. Pan, L. Bian, H. Tang and D. Wang, "SGDAN-A Spatio-Temporal Graph Dual-Attention Neural Network for Quantified Flight Delay Prediction," *Sensors,* vol. 20, no. 22, 2020.

[101] M. Zanin, Y. Zhu, R. Yan, P. Dong, X. Sun and S. Wandelt, "Characterization and Prediction of Air Transport Delays in China," *Applied Sciences,* vol. 10, no. 18, 2020.

[102] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou and D. Zhao, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning,," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 1, pp. 140-150, 2020.

[103] G. Gui, Z. Zhou, J. Wang, F. Liu and J. Sun, "Machine Learning Aided Air Traffic Flow Analysis Based on on Aviation Big Data," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 5, pp. 4817-4826, 2020.

[104] M. Melgosa, X. Prats, Y. Xu and L. Delgado, "Enhanced Demand and Capacity Balancing based on Alternative Trajectory Options and Traffic Volume Hotspot Detection," in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, San Diego, USA, 8-12 September, 2019.

[105] Y. Xu, X. Prats and D. Delahaye, "Synchronised demand-capacity balancing in collaborative air traffic flow management," *Transportation Research Part C: Emerging Technologies,* vol. 114, pp. 359-376, 2020.

[106] D. Bertsimas, G. Lulli and A. Odoni, "An Integer Optimization Approach to Large-Scale Air Traffic Flow Management," *Operations Research,* vol. 59, no. 1, pp. 211-227, 2011.

[107] B. A. Pinto, Air Traffic Management (ATM) Network Awareness through Synchronization Analysis (Master Thesis), Hamburg: Hamburg University of Technology (TUHH), 2019.

[108] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, A. Passos, D. Cournapeau and M. Brucher, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011.

[109] R. Mori and D. Delahaye, "Simulation-Free Runway Balancing Optimization Under Uncertainty Using Neural Network," in *Ninth SESAR Innovation Days,* Athens, Greece, 2-6 December, 2019.

[110] A. Fernandez, D. Martınez, P. Hernandez, S. Cristobal, F. Schwaiger and J. M. Nunez, "Flight Data Monitoring (FDM) Unknown Hazards detection during Approach Phase using Clustering

Techniques and AutoEncoders," in *Ninth SESAR Innovation Days*, Athens, Greece, 2-6 December, 2019.

[111] Z. Wang, M. Liang and D. Delahaye, "Short-term 4D Trajectory Prediction Using Machine Learning Methods," in *Sixth SESAR Innovation Days*, Belgrade, Serbia, 28-30 November, 2017.

[112] M. Poppe, R. Scharff, J. Buxbaum and D. Fieberg, "Flight Level Prediction with a Deep Feedforward Network," in *Eighth SESAR Innovation Days*, Salzburg, Austria, 3-7 December, 2018.

[113] D. Martinez, S. Belkoura, S. Cristobal, F. Herrema and P. Wächter, "A Boosted Tree Framework for Runway Occupancy and Exit," in *Eighth SESAR Innovation Days*, Salzburg, Austria, 3-7 December, 2018.

[114] R. Dalmau, F. Ballerini, H. Naessens, S. Belkoura and S. Wangnick, "Improving the Predictability of Take-off Times with Machine Learning," in *Ninth SESAR Innovation Days*, Athens, Greece, 2-6 December, 2019.

[115] R. Marcos, R. Herranz, R. R. Vázquez, P. García-Albertos and O. G. Cantú Ros, "Application of Machine Learning for ATM Performance Assessment–Identification of Sources of En-Route Flight Inefficiency," in *Eighth SESAR Innovation Days*, Salzburg, Austria, 3-7 December, 2018.

[116] B. PEŠIĆ LE FOLL, "Network Effect : A Possible Model to Highlight Interdependencies between Flow Management Regulations," EUROCONTROL, Brétigny-sur-Orge, 2005.

[117] S. Chatterjee and A. S. H. Hadi, Regression Analysis by Example, 5th ed., Hoboken, New Jersey: John Wiley & Sons, Inc., 2015.

[118] M. Awad and R. Khanna, Efficient learning machines, 1st ed., New York: Apress, 2015.

[119] F. Chollet and others, "Keras," 2015. [Online]. Available: https://keras.io. [Accessed 19 February 2020].

[120] B. Shen, X. Liang, Y. Ouyang, M. Liu, W. Zheng and K. Carley, "Stepdeep: a novel spatial-temporal mobility event prediction framework based on deep neural network," in *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, UK, 19-23 August, 2018.

[121] J. Wang, Q. Gu, J. Wu, G. Liu and Z. Xiong, "Traffic speed prediction and congestion source exploration: A deep learning method," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, Spain, 12-15 December, 2016.

[122] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga and A. Lerer, "Automatic differentiation in PyTorch," in *NIPS 2017 Autodiff Workshop*, Long Beach, USA, 9 December, 2017.

R

[123] A. L. Maas, A. Y. Hannun and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *International Conference on Machine Learning (ICML)*, Atlanta, USA, 16-21 June, 2013.

[124] L. Sillard, F. Vergne and B. Desart, "TAAM Operational Evaluation (project report," EUROCONTROL, Paris (Bretigny-Sur-Orge), 2000.

[125] EUROCONTROL (Network Manager), IFPS Users Manual, 26.0 ed., Brussels: EUROCONTROL, 2022.

[126] EUROCONTROL (Network Manager), IFPS Users Manual, 21.1 ed., Brussels: EUROCONTROL, 2017.

[127] Network Manager Directorate (NMD), NMIR User Guide, 1st ed., Brussels: EUROCONTROL, 2018.

[128] J. H. Holland, "Complex adaptive systems," *Daedalus,* vol. 121, no. 1, pp. 17-30, 1992.

[129] R. Dodder and R. Dare, "Complex Adaptive Systems and Complexity Theory: Inter-related Knowledge Domains," in *ESD.83: Research Seminar in Engineering Systems*, Cambridge, USA, 31 Oct 2000.

[130] S. Chan, "Complex Adaptive Systems," in *ESD.83 Research Seminar in Engineering Systems*, Cambridge, USA, 31 October-6 November, 2001.

[131] G. L. Donohue, "Air Transportation is a Complex Adaptive System: not an aircraft design," in *AIAA/ICAS International Air and Space Symposium and Exposition: The Next 100 Years*, Dayton, USA, 14-17 July, 2003.

[132] D. A. DeLaurentis and S. Ayyalasomayajula, "Exploring the synergy between Industrial Ecology and System of Systems to understand complexity: a case study in air transportation," *Journal of Industrial Ecology,* vol. 13, no. 2, pp. 247-263, 2009.

[133] M. Haghnevis and R. G. Askin, "A modeling framework for engineered complex adaptive systems," *IEEE Systems Journal,* vol. 6, no. 3, pp. 520-530, 2012.

[134] T. Kotegawa, Analyzing the Evolutionary Mechanisms of the Air Transportation System-of-Systems using Network Theory and Machine Learning Algorithms (Doctoral Dissertation), West Lafayette: ProQuest LLC Purdue University, 2012.

[135] B. Roberts, T. Mazzuchi and S. Sarkani, "Engineered Resilience for Complex Systems as a Predictor for Cost Overruns," *Systems Engineering,* vol. 19, no. 2, pp. 111-131, 2016.

[136] G. Pumpuni-Lenss, T. Blackburn and A. Garstenauer, "Resilience in Complex Systems: An Agent-Based Approach," *Systems Engineering* , vol. 20, no. 2, 2017.

[137] E. Ordoukhanian and A. M. Madni, "Model-Based Approach to Engineering Resilience in Multi-UAV Systems," *Systems,* vol. 7, no. 1, p. 11, 2019.

[138] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 25–29 October, 2014.

[139] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)," in *2016 Integrated Communications Navigation and Surveillance (ICNS)*, Herndon, USA, 19–21 April, 2016.

[140] K. Gopalakrishnan and H. Balakrishnan, "A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks," in *12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, Seattle, USA, 26-30 June, 2017.

[141] B. Yu, Z. Guo, S. Asian, H. Wang and G. Chen, "Flight delay prediction for commercial air transport: A deep learning approach," *Transportation Research Part E: Logistics and Transportation Review,* vol. 125, pp. 203-221, 2019.

[142] B. Thiagarajan, L. Srinivasan, A. V. Sharma, D. Sreekanthan and V. Vijayaraghavan, "A machine learning approach for prediction of on-time performance of flights," in *IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, St. Petersburg, USA, 17-21 September, 2017.

[143] S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta and S. Barman, "A statistical approach to predict flight delay using gradient boosted decision tree," in *International Conference on Computational Intelligence in Data Science (ICCIDS). IEEE*, Chennai, India., 2-3 June, 2017.

[144] Y. Lin, J.-w. Zhang and H. Liu, "Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation," *Aerospace Science and Technology,* vol. 93, pp. 105-113, 2019.

[145] A. Lichtenwald, Prediction of European ATM Network disruptions by Recurrent Neural Networks (Master Thesis), Hamburg: Technical University of Hamburg (TUHH), 2021.

[146] T. Dozat, "Incorporating nesterov momentum into adam," in *International Conference on Learning Representations (ICLR 2016 - Workshop Track)*, San Juan, Puerto Rico, 2-4 May, 2016.

[147] EUROCONTROL, "NMIR (Network Manager Interactive Reporting)," [Online]. Available: https://www.eurocontrol.int/dashboard/network-manager-interactive-reporting-dashboard.

[148] EUROCONTROL (Network Manager), "Presentation of NM 20.0," EUROCONTROL, Brussels, 2016.

[149] EUROCONTROL, "SESAR Solution PJ09-1: Validation Report (VALR) for V2," European Commission: Community Research and Development Information Service (CORDIS), Brussels, 2019.

[150] US Dept of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, "A Look Back at the March 5th, 2013 Snowstorm and Historical Perspective," Chicago, IL Weather Forecast Office, 2013. [Online]. Available: https://www.weather.gov/lot/2013mar05. [Accessed 05 August 2021].

[151] M. Schultz, S. Lorenz, R. Schmitz and L. Delgado, "Weather Impact on Airport Performance," *Aerospace,* vol. 5, no. 4, p. 109, 2018.

[152] European Commission, "Volcano Grimsvötn: how is the European response different to the Eyjafjallajökull eruption last year? Frequently Asked Questions," 28 May 2011. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_11_346. [Accessed 04 July 2021].

[153] EUROCONTROL (Network Management Directorate), "European Network Operations Plan 2021," EUROCONTROL, Brussels, 2021.

[154] ICAO, "AIRAC," International Civil Aviation Organization, [Online]. Available: https://www.icao.int/Safety/information-management/Pages/AIRACAdherence.aspx. [Accessed 20 September 2020].

[155] E. Hollnagel, FRAM - The Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems, Farnham: Ashgate., 2012.

[156] R. Woltjer and E. Hollnagel, "Modelling and evaluation of air traffic management automation using the functional resonance analysis method," in *International Symposium of the Australian Aviation Psychology Association AAvPA*, Sydney, Australia, 18 August, 2008.

Annex A. Reference charts and lists

A.1. IFPS Zone (IFPZ)

IFPS Zone is the geographical zone in Europe that is used by Initial Flight Planning System (IFPS). The system is operated by ECTL (NM) and process and distributes flight plans [125]. IFPZ covers the same area as ICAO Europe region as provided in following table. ICAO codes in this table is same designation for ACCs which is a key feature of input vector in proposed learning models.

Table A-1: IFPS message distribution (IFPS Zone) *[125]*

State	Country Code	IFPZ	FIR/UIR	ICAO
Albania	LA	Yes	Tirana	LAAA
Armenia	UD	Yes	Yerevan	UDDD
Austria	LO	Yes	Vienna	LOVV
Azerbaijan	UB	Yes	Baku	UBBA
Belarus	UM	Copy Only	Minsk	UMMV
Belgium	EB	Yes	Brussels	EBBU/EBUR
Bosnia and Herzegovina	LQ	Yes	Sarajevo	LQSB
Bulgaria	LB	Yes	Sofia	LBSR
Croatia	LD	Yes	Zagreb	LDZO
Cyprus	LC	Yes	Nicosia	LCCC
Czech Republic	LK	Yes	Prague	LKAA
Denmark	EK	Yes	Copenhagen	EKDK
Estonia	EE	Yes	Tallinn	EETT
Finland	EF	Yes	Finland	EFIN
France	LF	Yes	Paris	LFFF
			Reims	LFEE
			Brest	LFRR
			Bordeaux	LFBB
			Marseille	LFMM
Georgia	UG	Yes	Tbilisi	UGGG
Germany	ED	Yes	Bremen	EDWW
			Langen	EDGG
			Munich	EDMM
			Rhein	EDUU
			Hanover	EDVV
Greece	LG	Yes	Athens	LGGG
Hungary	LH	Yes	Budapest	LHCC
Ireland	EI	Yes	Shannon	EISN
			SOTA	EISN
Israel	LL	Yes	Tel-Aviv	LLLL
Italy	LI	Yes	Rome	LIRR
			Brindisi	LIBB
			Milan	LIMM
Latvia	EV	Yes	Riga	EVRR
Lithuania	EY	Yes	Vilnius	EYVL
Luxembourg	EL	Yes	Brussels	EBBU/EBUR

State	Country Code	IFPZ	FIR/UIR	ICAO
North Macedonia	LW	Yes	Skopje	LWSS
Malta	LM	Yes	Malta	LMMM
Republic of Moldova	LU	Yes	Chisinau	LUUU
Monaco (Marseille)	LN	Yes	Marseille	LFMM
Morocco	GM	Yes	Casablanca	GMMM
Netherlands	EH	Yes	Amsterdam	EHAA
Norway	EN	Yes	Norway	ENOR
			Bodo Oceanic	ENOB
Poland	EP	Yes	Warsaw	EPWW
Portugal	LP	Yes	Lisbon	LPPC
			Santa Maria	LPPO
Romania	LR	Yes	Bucharest	LRBB
Rostov FIR (Russian Federation)	URR	Copy Only		
Kaliningrad FIR (Russian Federation)	UMK	Copy Only		
Slovak Republic	LZ	Yes	Bratislava	LZBB
Slovenia	LJ	Yes	Ljubljana	LJLA
Spain	LE	Yes	Barcelona	LECB
			Madrid	LECM
			Canaries	GCCC
Sweden	ES	Yes	Sweden	ESAA
Switzerland	LS	Yes	Switzerland	LSAS
Turkey	LT	Yes	Ankara	LTAA
			Istanbul	LTBB
Ukraine	UK	Yes	L'Viv	UKLV
			Kyiv	UKBV
			Dnipropetrovsk	UKDV
			Odessa	UKOV
			Simferopol	UKFV
United Kingdom	EG	Yes	London	EGTT
			Scottish	EGPX
			Shanwick	EGGX
Serbia and Montenegro	LY	Yes	Belgrade	LYBA

A.2. Regulation causes

Regulation types are also referred to as regulation causes in related documents such as the ATFCM user manual [10]. Each regulation can be implemented based on a set of provided guidelines. The coding also provides details on regulation location that declares the phase of the delayed flight. In this study, these classes are only used for the learning models without considering the flight phase.

Table A-2: Description of different ATFCM Regulation types *[10]*

Regulation Cause	Code	Regulation location [a]	Guidelines
Accident/incident	A	D, A	Reduction of expected ATC capacity due to an aircraft accident /incident.
ATC capacity	C	D, E, A	En Route: Demand exceeds or complexity reduces declared or expected ATC capacity; Airport: Demand exceeds declared or expected ATC capacity.
Aerodrome Services	E	D, A	Reduced capacity due to the degradation or non-availability of support equipment at an airport e.g. Fire Service, De-icing / snow removal equipment or other ground handling equipment.
Aerodrome capacity	G	D, A	Reduction in declared or expected capacity due to the degradation or non-availability of infrastructure at an airport. e.g. Work in Progress, shortage of aircraft stands, etc. Or when demand exceeds expected aerodrome capacity.
ATC industrial action	I	D, E, A	Reduction in any capacity due to industrial action by ATC staff.
Airspace management	M	D, E, A	Reduction in declared or expected capacity following changes in airspace / route availability due to small scale military active.
Industrial action NON-ATC	N	D, E	A reduction in expected / planned capacity due to industrial action by non-ATC personnel.
Other	O	D, E, A	This should only be used in exceptional circumstances when no other category is sufficient.
Special event	P	D, E, A	Reduction in planned, declared or expected capacity or when demand exceeds the above capacities as a result of a major sporting, governmental or social event. It may also be used for ATM system upgrades and transitions. Large multinational military exercises may also use this reason.
ATC routings	R	E	Network solutions / scenarios used to balance demand and capacity.
ATC staffing	S	D. E, A	Unplanned staff shortage reducing expected capacity.
ATC equipment	T	D, E, A	Reduction of expected or declared capacity due to the non-availability or degradation of equipment used to provide an ATC service.
Environmental issue	V	D, E, A	Reduction in any capacity or when demand exceeds any capacity due to agreed local noise, runway usage or similar procedures. This category should only be used with prior agreement in the planning process.
Weather	W	D, E, A	Reduction in expected capacity due to any weather phenomena. This includes where weather impacts airport infrastructure capacity but where aerodrome services are operating as planned / expected.

[a] *Regulation Location code D: Departures, E: En-route and A: Arrivals*

A

Annex B. Supplement of DCNN

B.1. AIRAC cycles from 2015 to 2018

Aeronautical Information Regulation And Control (AIRAC) cycles' effective dates are obtained from the International Civil Aviation Organization (ICAO) website [154] and compiled as below:

Table B-1. Schedule of AIRAC effective dates, 2015-2018.

2015	2016	2017	2018
08 Jan.	07 Jan.	05 Jan.	04 Jan.
05 Feb.	04 Feb.	02 Feb.	01 Feb.
05 Mar.	03 Mar.	02 Mar.	01 Mar.
02 Apr.	31 Mar.	30 Mar.	29 Mar.
30 Apr.	28 Apr.	27 Apr.	26 Apr.
28 May	26 May	25 May	24 May
25 June	23 June	22 June	21 June
23 July	21 July	20 July	19 July
20 Aug.	18 Aug.	17 Aug.	16 Aug.
17 Sept.	15 Sept.	14 Sept.	13 Sept.
15 Oct.	13 Oct.	12 Oct.	11 Oct.
12 Nov.	10 Nov.	09 Nov.	08 Nov.
10 Dec.	08 Dec.	07 Dec.	06 Dec.

B.2. Overfitted RF model

Table B-2. Performance of applied RF to predict delay (max_depth = 50).

Category	Train				Test			
	Days	MAPE[a]	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	127	0.0	1.0	0	55	77.04	-4.16	8 100
Nominal	261	0.0	1.0	0	111	27.82	0.22	11 208
High	123	0.0	1.0	0	53	18.45	0.47	24 829
Overall	511	0.0	1.0	0	219	37.91	0.85	13 724

[a] in percentage, [b] minutes.

Table B-3. Performance of applied RF to predict delayed traffic (max_depth = 50).

Category	Train				Test			
	Days	MAPE[a]	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	129	0.0	1.0	0	52	56.20	-2.02	349
Nominal	256	0.0	1.0	0	113	17.84	0.66	437
High	126	0.0	1.0	0	54	11.86	0.44	765
Overall	511	0.0	1.0	0	219	25.47	0.91	497

[a] in percentage, [b]flights.

B.3. Learning performance (delayed traffic)

Table B-4. Performance of applied LR.

Category	Train				Test			
	Day	MAPE	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	129	47.98	-2.63	360	52	64.28	-2.07	370
Nominal	256	21.06	0.59	503	113	22.00	0.46	569
High	126	11.42	0.56	727	54	9.90	0.62	617
Overall	511	25.48	0.91	522	219	29.05	0.9	533

[a] in percentage, [b] flights.

Table B-5. Performance of applied SVR.

Category	Train				Test			
	Day	MAPE	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	129	41.38	-1.86	308	52	54.79	-1.47	319
Nominal	256	18.69	0.62	457	113	20.76	0.5	536
High	126	12.82	0.37	835	54	11.10	0.52	697
Overall	511	22.97	0.9	513	219	26.46	0.9	524

[a] in percentage, [b*]flights.

Table B-6. Performance of applied RF.

Category	Train				Test			
	Day	MAPE	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	129	0.0	1.0	0	52	55.95	-1.96	346
Nominal	256	0.0	1.0	0	113	17.31	0.68	427
High	126	0.0	1.0	0	54	11.64	0.46	749
Overall	511	0.0	1.0	0	219	25.09	0.91	487

[a] in percentage, [b]flights.

Table B-7. Performance of applied NN.

Category	Train				Test			
	Day	MAPE	R^2	MAE[b]	Days	MAPE	R^2	MAE
Low	129	31.02	-0.8	229	52	47.95	-1.46	286
Nominal	256	16.13	0.68	411	113	23.13	0.41	599
High	126	11.12	0.56	696	54	9.75	0.55	608
Overall	511	18.65	0.92	435	219	25.73	0.89	527

[a] in percentage, [b]flights.

B

B.4. DCNN prediction performance

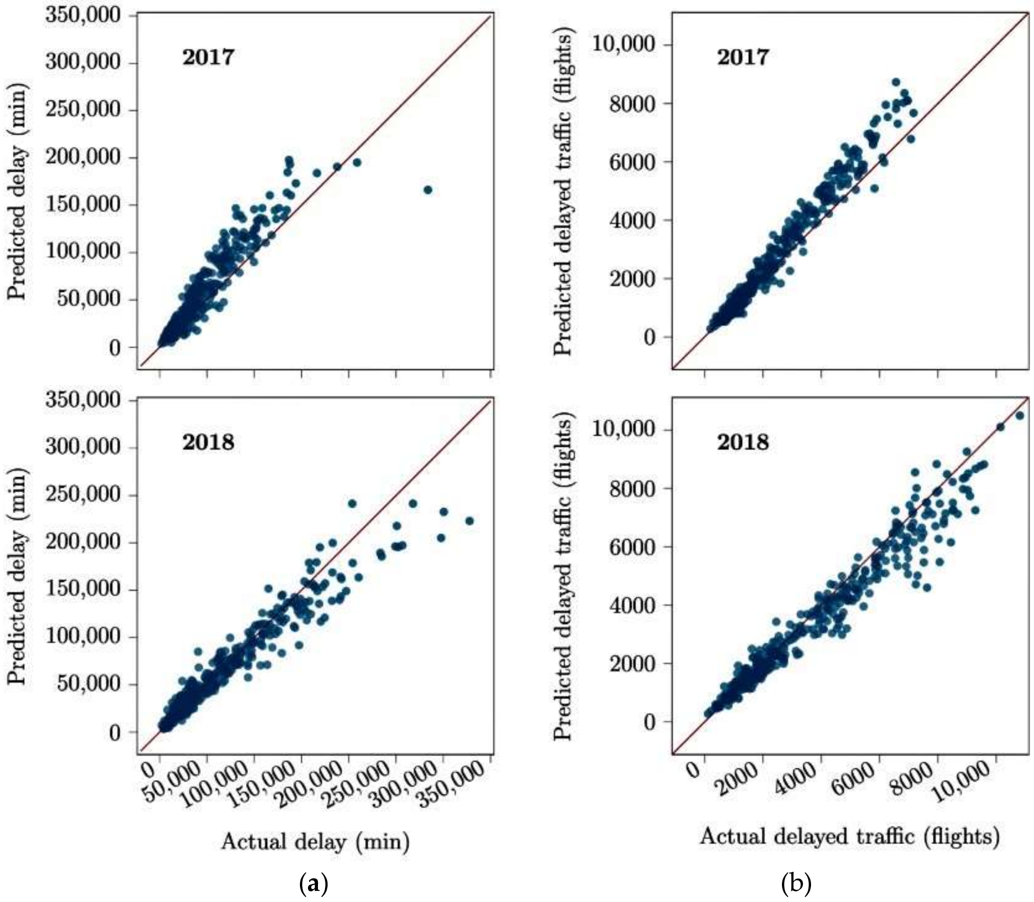


Figure A-1 Comparative scatter plots for prediction quality of DCNN in different years. (a) Delay, (b) delayed traffic. Model is more precise in predicting lower values. CNN validation: prediction errors in different AIRAC cycles. More regulations in 2018 (especially during summer season) provide better prediction quality regardless of the expected high values for both delay and delayed traffic.

B

B

Annex C. Developed tools

In the course of thesis, two main tools have been coded for analysis of results and visualization of the methodology. Such tools are required because of the exploratory nature of the study that makes visual interfaces an asset in demonstrating the added values. MATLAB is used to code the ANMStat and Python to code NETRES. Each of these tools are presented in the following sections.

C.1. ANMStat

As the initial step to understand the regulations it is required to analyze the data from a statistical point of view. Using MATLAB, the ANM messages are studied from different aspects. ANM Stat is developed to facilitate the recognition of different characteristics of regulations, specifically with the details provided in ANM records. As an intuitive tool, the objective is to generate multiple analytical plots.

C.1.1. Purpose

The tool is designed to parse captured data from the ANM list. It parses daily ANM lists from excel sheets and files into MATLAB (.mat) format, to start monthly analysis of the data. Regulations can be studied by 26 different plots to provide insight on any monthly and seasonal trends. Trends are observed by Survey-plots and analyzed by Net-vision plots (Figure C-1). The latter include both statistical predictions as well as geographical representation. The details of this extensive overview on all aspects of the ANM data is provided next.

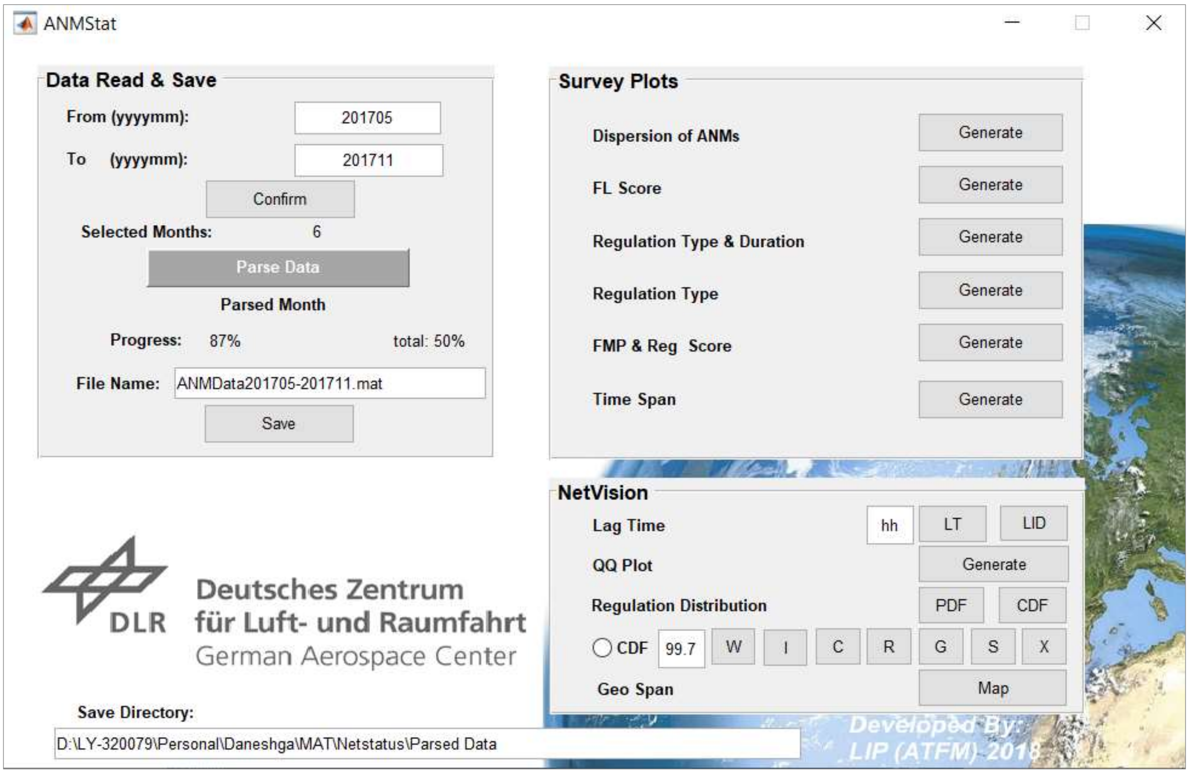


Figure C-1 ANM Stat interface

C.1.2. User interface

- Data read & save: user selects the span of the data in years and month. Parsing of the data triggers the calculation of parameters needed to the other two sections of the tool. All of the calculated parameters are saved as a mat file at the selected directory.

- Survey Plots: once parsing is 100%, the tool is able to generate 6 types of descriptive plots based on the different aspects of the regulation data in ANM records with a daily precision. These plots include:

    - Dispersion of ANMs: gives a total overview of different ANM counts (Figure C-2), so that the load of regulations can be monitored. Range of regulation counts are presented by box-plots and outliers are marked (both for each month and for total parsed data).

C

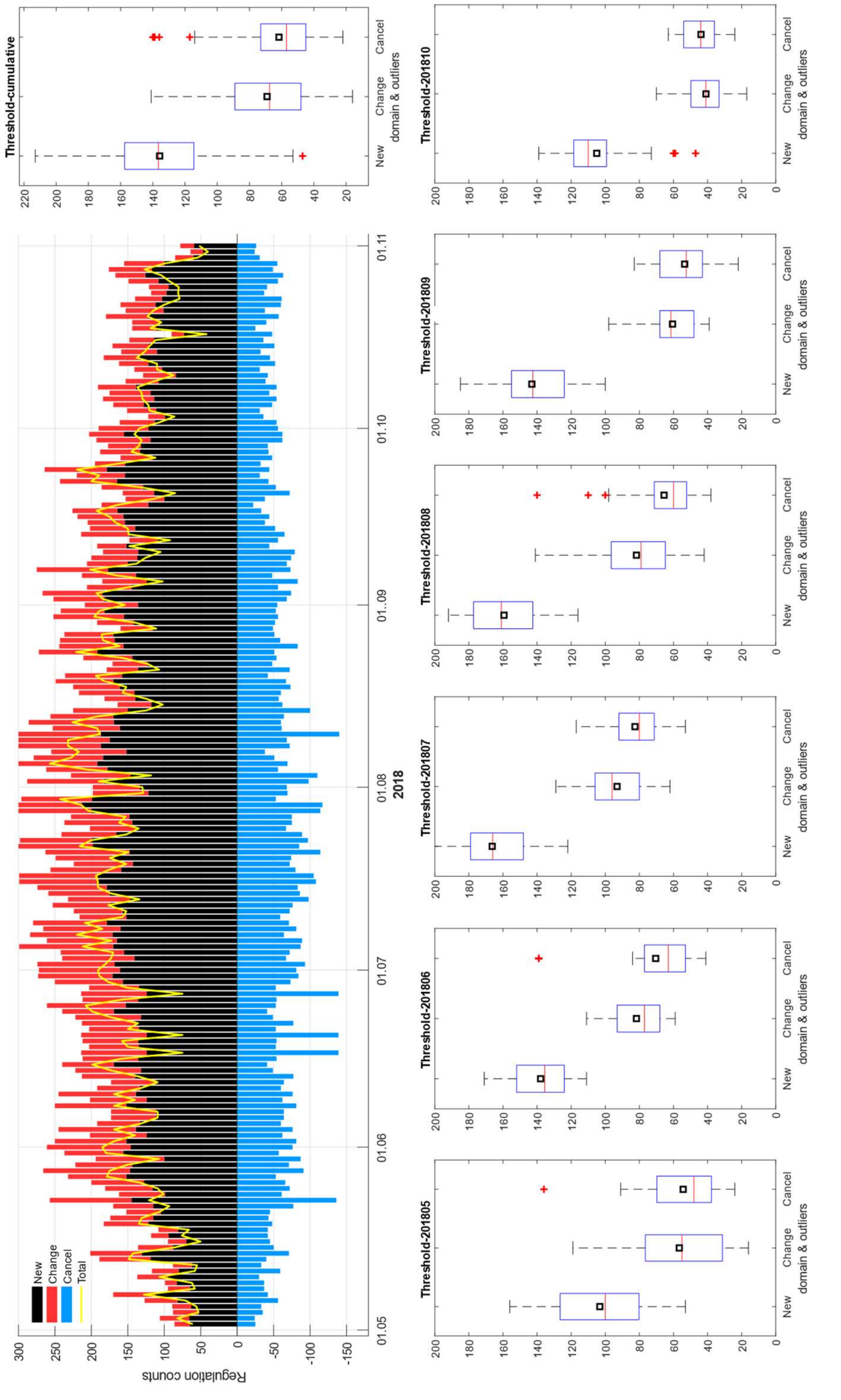


Figure C-2 ANM Stat, Dispersion Plot

- FL Score: provides an overview that clarifies which flight levels are mostly blocked in which days of the input data. Four plots are generated that two are dedicated to calculated severity scores and two heatmaps show blocked flight levels for each day (e.g. Figure C-3)

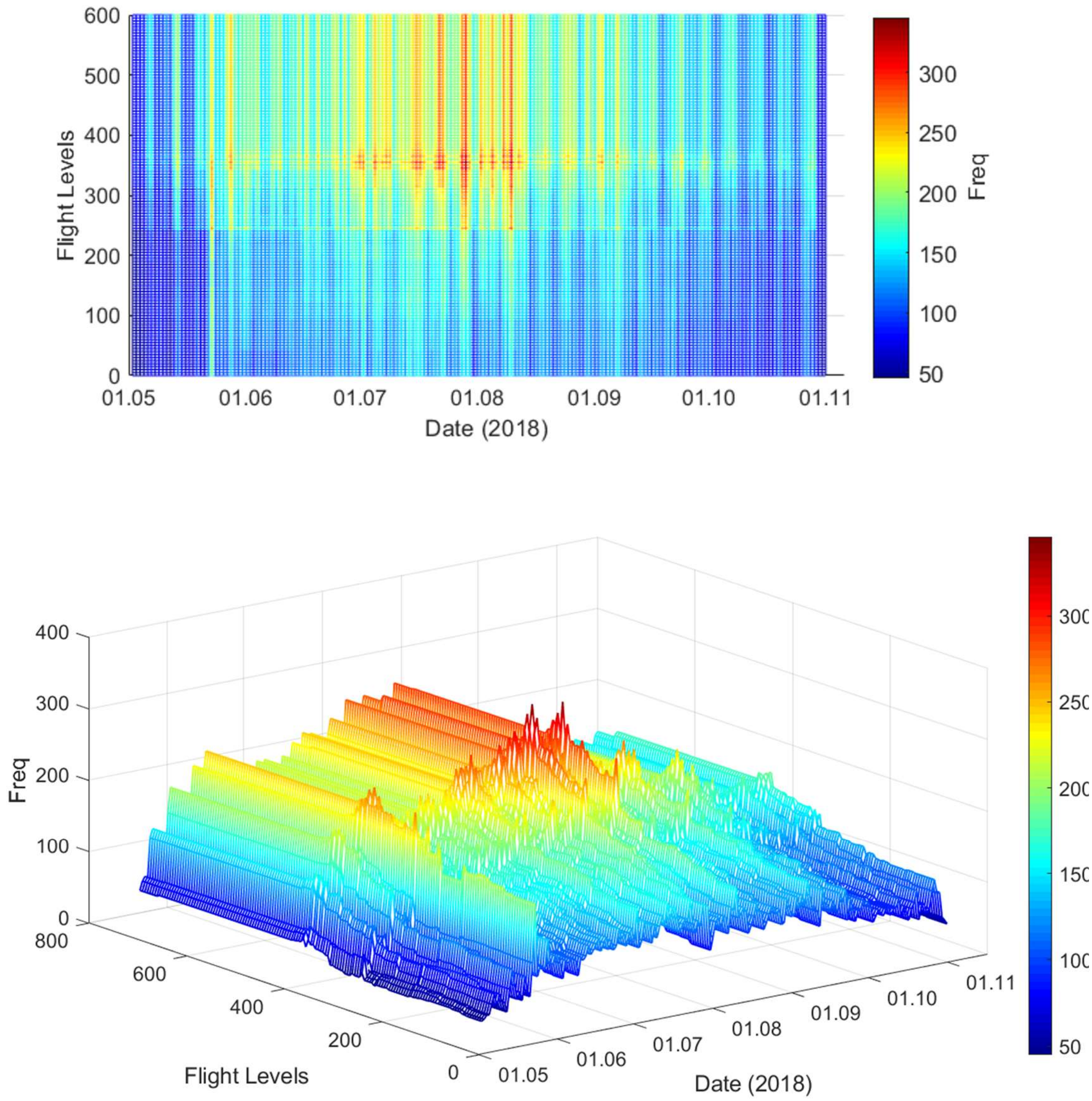


Figure C-3 ANM Stat, Flight Level Heatmap

- Regulation Type & Duration: provides four plots for "ATC Capacity, ATC Rerouting, Weather and Other" categories. Each plot shows both the duration and count of each type of regulation over days.
- FMP & Regulation Score: measured scores for different FMPs are plotted. For instance, Figure C-4 shows number of regulations for five groups of FMPs. This allows monitoring of different FMPs based on their saturation level. More over this plot shows which groups of FMPs are comparable. Such a figure checks if there are extractable information at Functional Airspace Blocks (FAB) level. FABs are airspaces that are formed regardless of states boundaries.

- A-Class represents busiest FMPs:
  - EGTT is the Upper Flight Information Region (UIR) for UK and is managed by NATS,
  - EDGG is the busiest of three german FIRs (Langen FMP) and is managed by DFS;
- B-Class includes five next busiest FMPs combined: LECB (Barcelona), LECM (Madrid), LFFF (Paris), LFRR (Brest) and EHAA (Amsterdam);
- C-Class is a group of other 19 FMPs based on their activities
- All other remaining FMPs in the ECAC area (54 FIRs and UIRs)

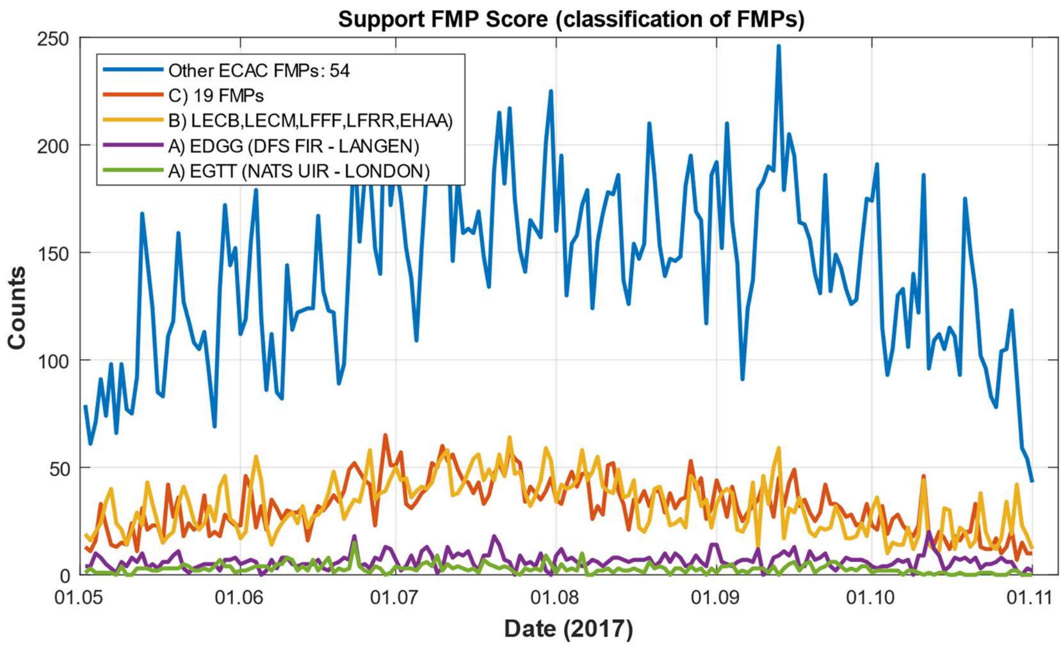


Figure C-4 ANM Stat, FMP Scores

- Time Span: generates plots to identify both outliers and standard ranges for regulation duration. Moreover, it enables the detection of days with maximum regulation duration and respective severity (standard deviation) at a network level (Figure C-5).

C

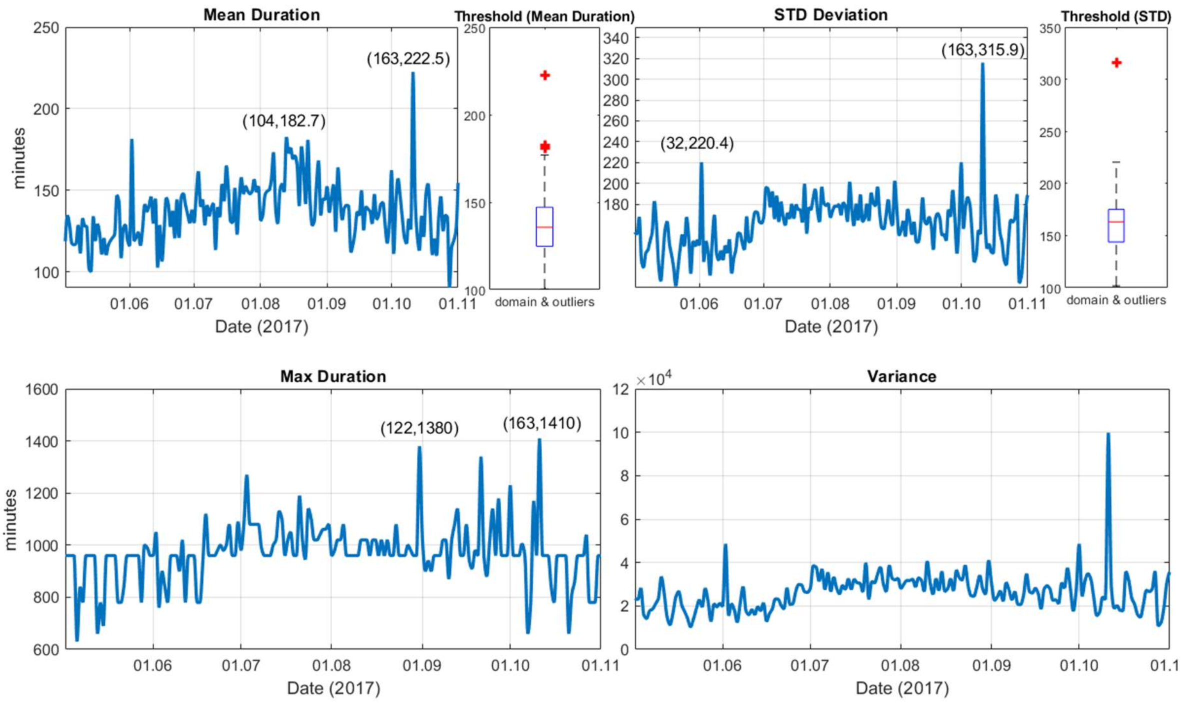


Figure C-5 ANM Stat, Time Span

- NetVision: this panel of the tool (Figure C-6) is analyzing the data to provide distribution curves and is also providing a geographical map to identify spatial characteristics of blocked airspace.

Figure C-6 ANM Stat, NetVision

  - Lag time: two sets of plots provide a network level insight based on activation notice, regulation duration, counts, and blockage time. These plots assist the visualization of network resiliency and enables the user to compare evolution of regulations in different time frames with a focus of comparing planned and implemented regulations. Such plots are also an asset in understanding the impact of different type of regulation types.

- LT button: four plots are provided to monitor the evolution of planned network situation in pre-tactical phase and any other cuts (tool allows to select different cuts of the day, e.g. -12 hrs). The Lead Time is the activation notice as the difference between regulation publish and start time (Figure C-7). Negative values indicate updated regulations. if the update extends a regulation, less adaptation time will impact airlines (resilience: time-to-recover). Even if the update, relaxes a regulation it is still a cost of lost capacity which may be challenging to retrieve.

- LID button: 3 plots are generated for stacked regulation activation notice, count of regulations and blockage duration. This option helps to realize which type of regulation is planned earlier and which ones cause longer blockage times and if there is a direct ratio between count of regulations and blockage time per regulation type (Figure C-8). Furthermore, such outputs provide multi-dimensional means to spot network disrupted dates and to better form statements on network behavior on different regulation types. For instance, routing regulations tend to be published earlier than other types of regulations in contrast to weather induced regulations with minimum anticipation rate.

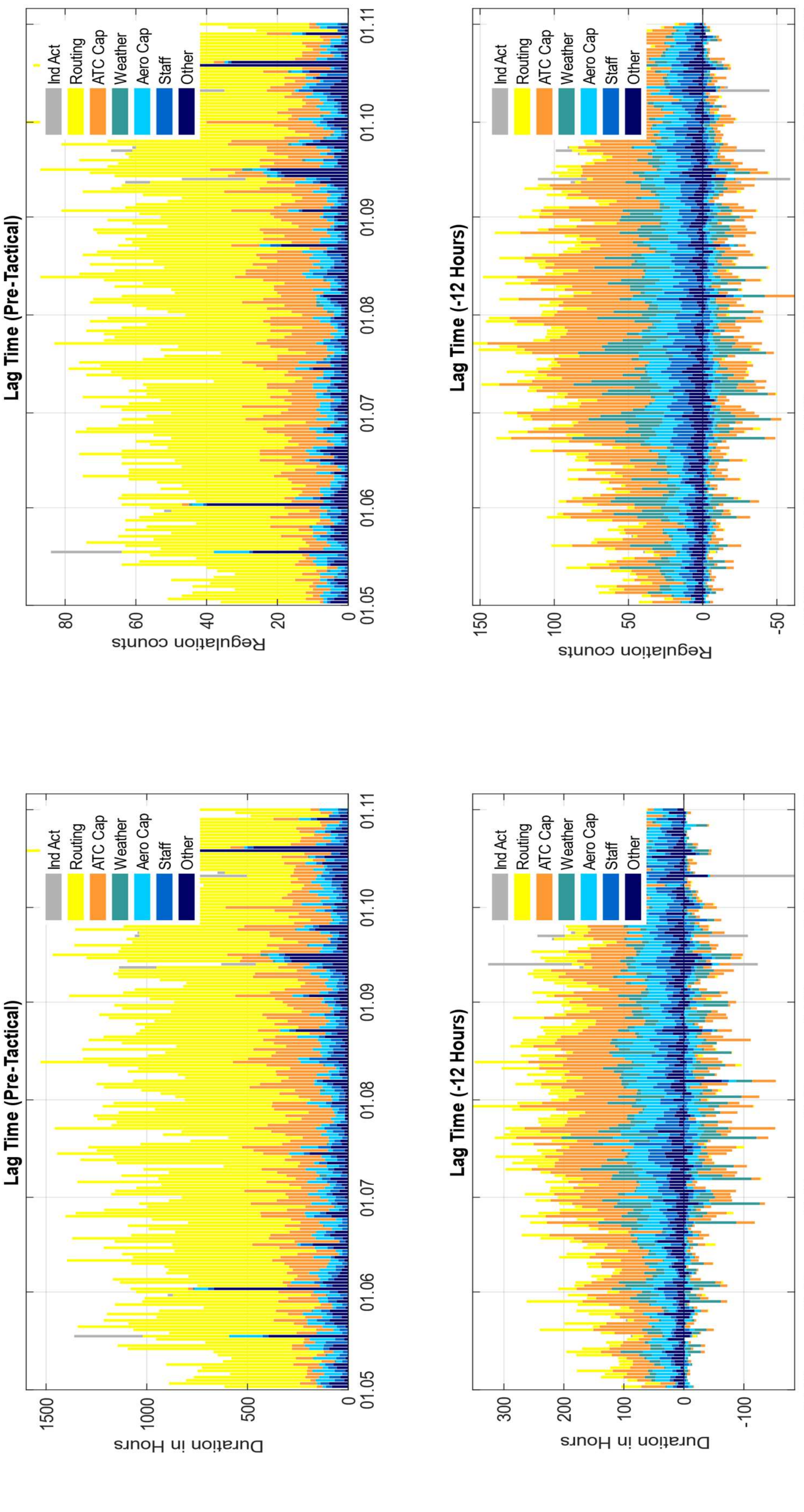


Figure C-7 ANM Stat, Lead Time

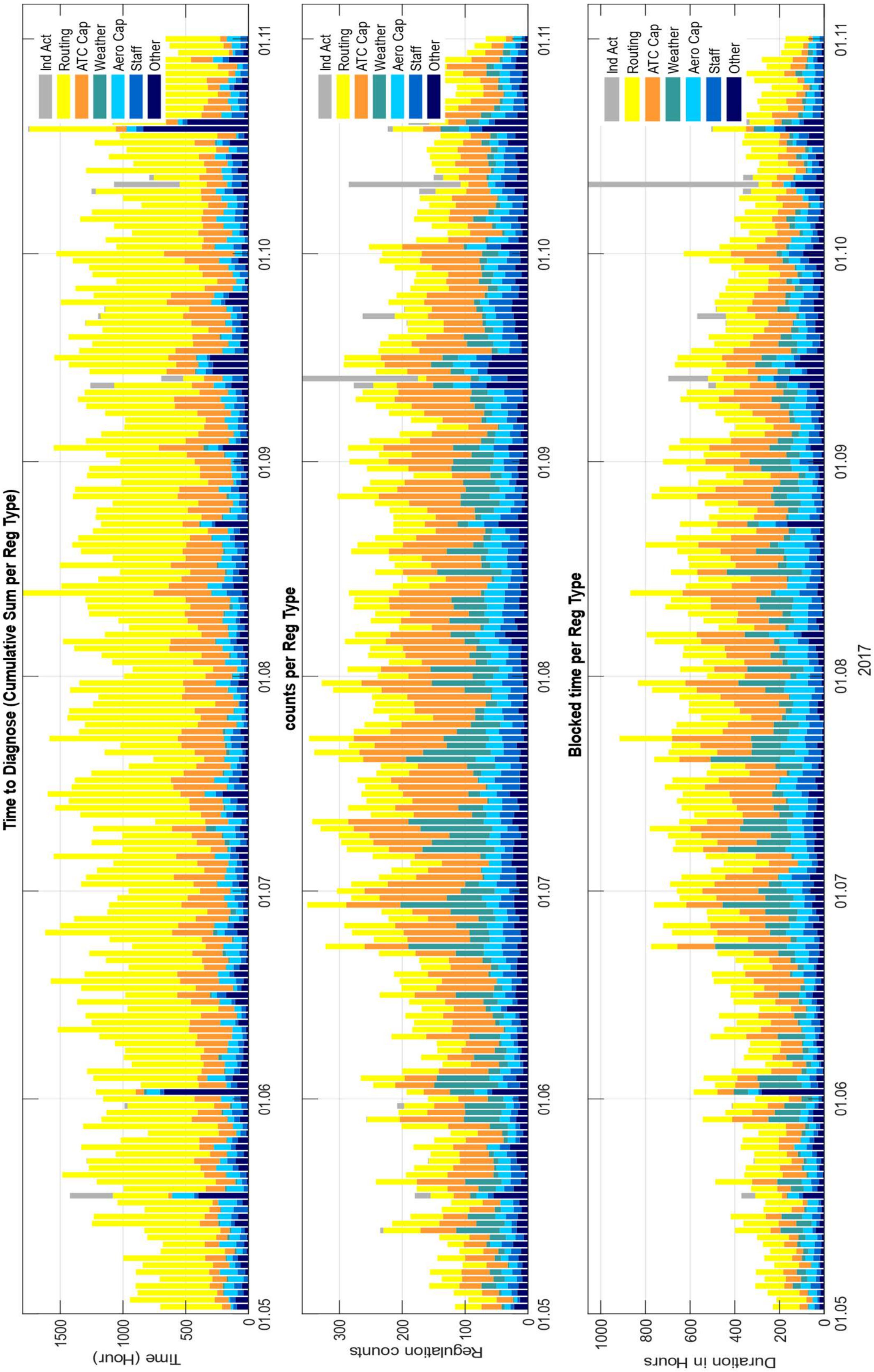


Figure C-8 ANM Stat, Threefold Comparison

- QQ Plot: provides Quantile-Quantile plots that is a statistical method to compare fitted probability distributions against normal distributions. User will be able to intuitively analyze estimated distributions for different regulation types and presume if a specific regulation type is more likely to have a Chi-squared distribution or a Weibull, Gamma or Beta (Figure C-9). Other statistics such as Level of dispersion and domain are aids to pick desired control thresholds.

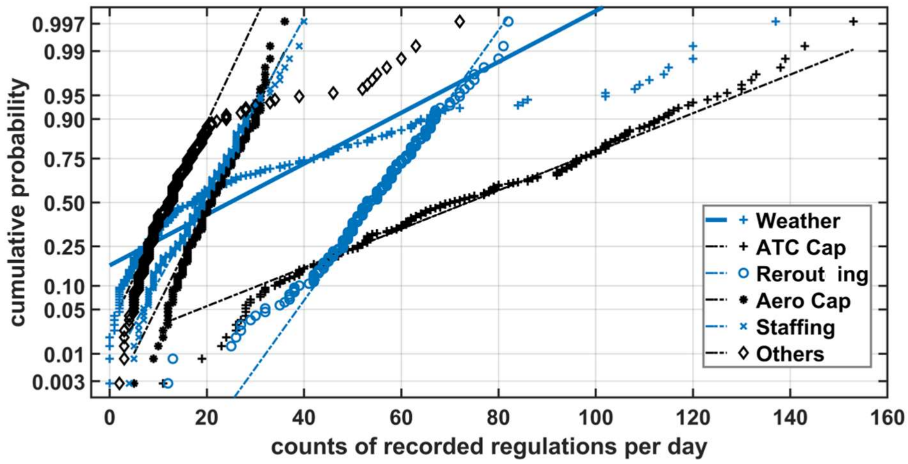


Figure C-9 ANM Stat, QQ plot *[61]*

- Regulation Distributions: a kernel estimation of probability distributions is provided by these plots for different type of regulations (both probability and cumulative distribution functions, i.e. PDF and CDF). User is able to set different threshold for each plot (default value is 99.7%). All regulation types are able to be plot either separately or combined (e.g. Figure C-10, inset is the CDF with same axes): Weather (W), Industrial action (e.g. strike), ATC Capacity (C), Rerouting (R), Aerodrome Capacity (G), Staffing (S) and the rest of regulations are summed up in others category (X).

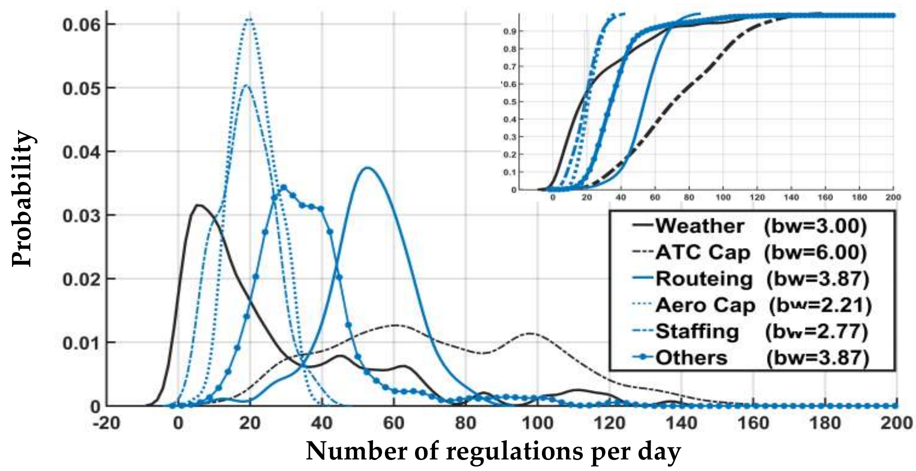


Figure C-10 ANM Stat, Regulation Distribution (dimensions of inset are same as base plot), *[61]*

- Geo Span: plots the input data on ECAC map to identifies airspaces with different color-codes with regard to regulation counts (Figure C-11). This enables identifying bottle-necks for different regulation types.

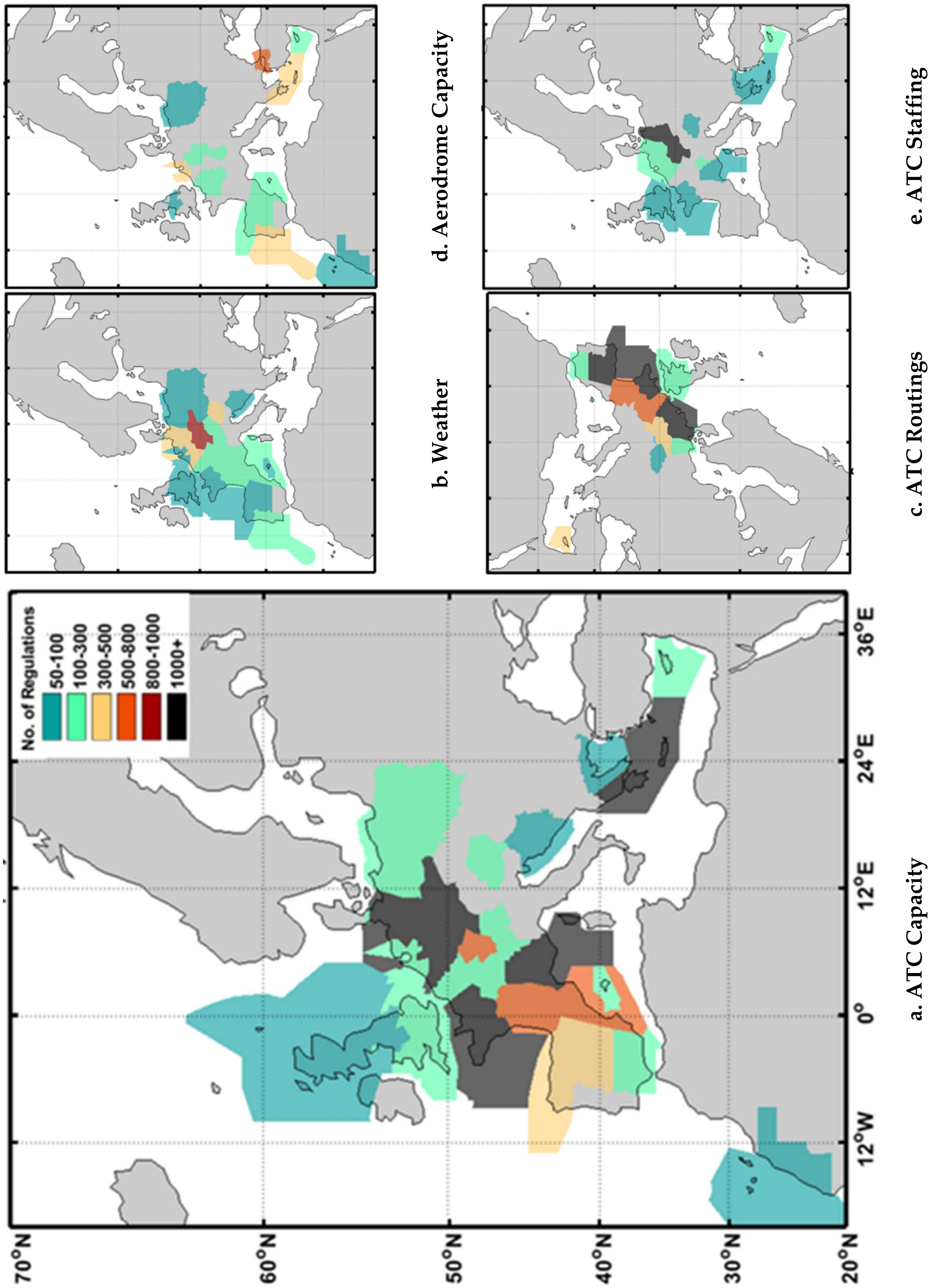


Figure C-11 *ANM Stat, Geo Span, [61]*

C.1.3. Added value

As described in previous section, ANM Stat provides a much deeper understanding of regulations. But more importantly it estimates the probability distributions and detects outliers. Also, in terms of resilience, it provides the network visibility in post operational phase. It allows to cluster different FMPs into comparable groups. Besides, on the post operational analysis it avoids extraction of risky assumptions. For instance, in Figure C-12 it is evident that planned regulations do not necessary lead to less blocked airspace.

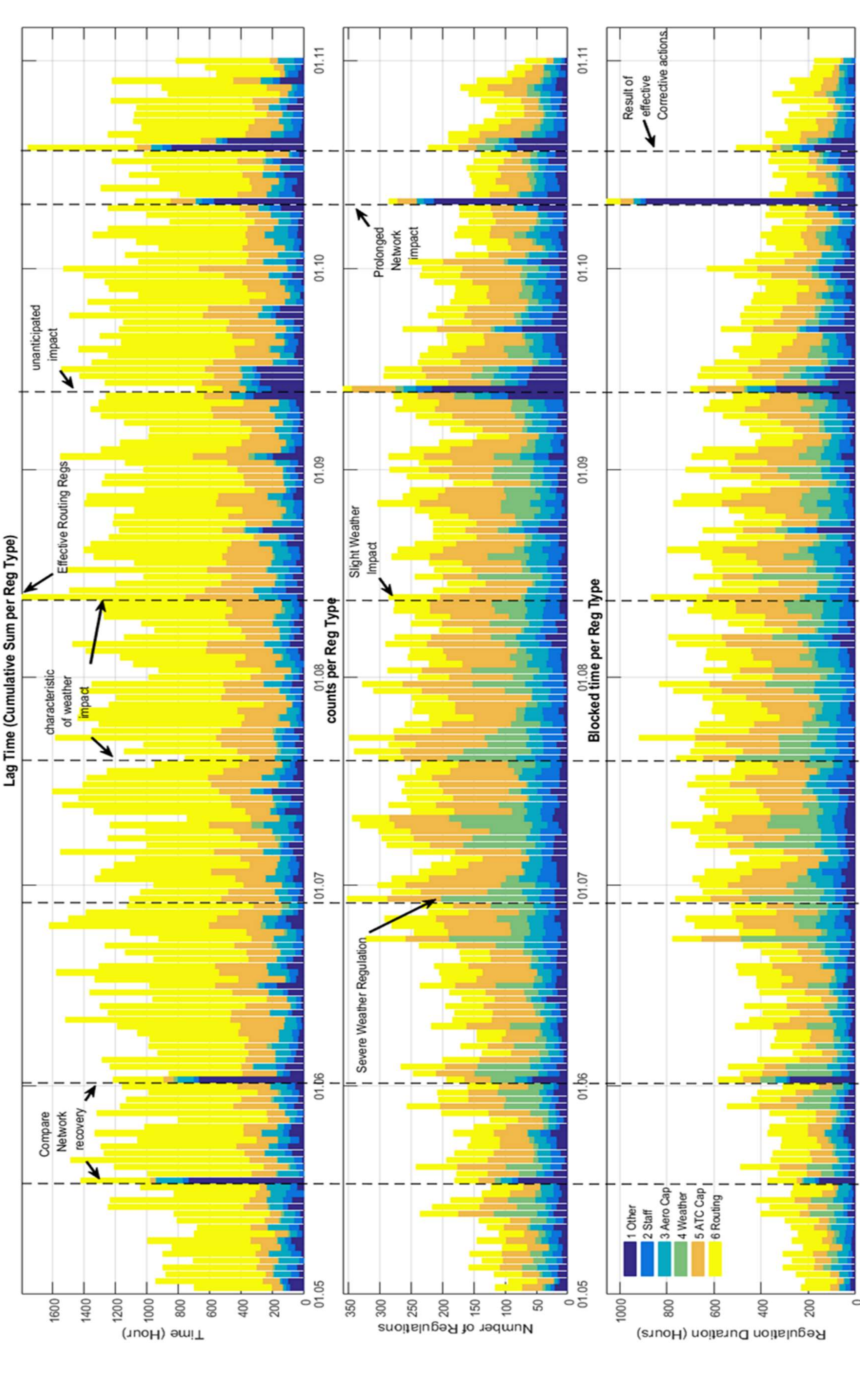


Figure C-12 Benefit of threefold comparison

C.2. NETRES

In the course of the SESAR PJ09 project, the essential role of demonstrative tools became more evident. The ATM experts in the project considered resilience as a safety topic (safety-I). However, this thesis is set as an exploratory research on Safety-II resiliency that is more focused on ATM network performance. The design of NETRES is proposed as an intuitive tool (Figure C-13) to demonstrate relatively new exertion of network resiliency. The idea is appreciated by experts in the SESAR project and NETRES proved its effective role in data visualization. Also, it offers a unique advantage in describing the benefits of the research objectives and the added value to ATM stakeholders, specially Network Manager role.

C.2.1. Purpose

Since the design of the NETRES foresaw some requirements, it is coded in python. These requirements are; built-in availability for further development, ability to connect to non-academic tools (with respect to close collaboration in SESAR) and compatibility in use of machine learning platforms.

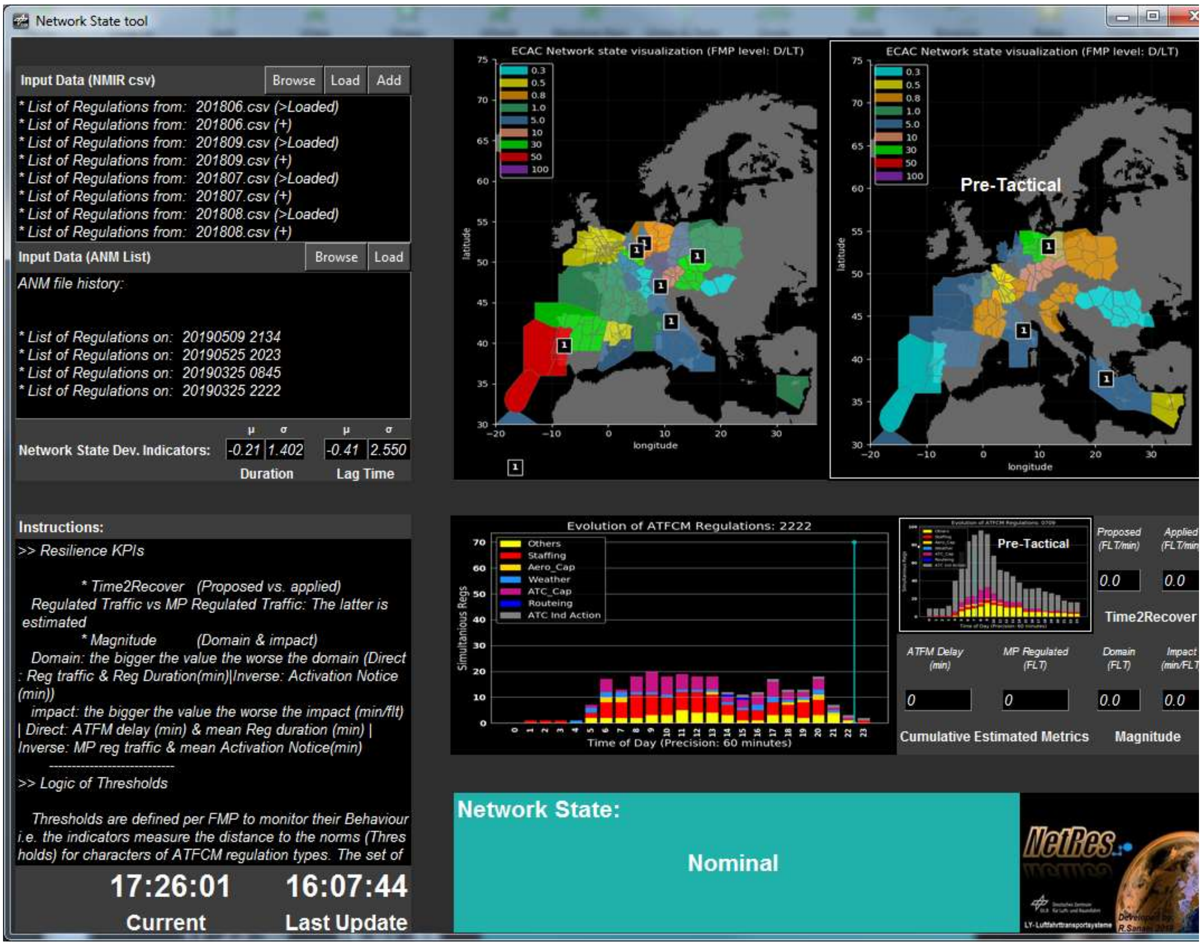


Figure C-13 NETRES, Demo screenshot

The tool is designed to provide a tactical network situational awareness by its interface. Based on the regulations, it is able to measure some performance indicators and provide both temporal and special views on ECAC area. The stand-alone tool shows the network state and allows to intuitively monitor the evolution of network disruptions during tactical phase. It is designed as a follow-up from ANM Stat tool that brought the network evolution into focus (Figure C-12).

C.2.2. User interface

- User input, NMIR: user selects the post-operational data from the NMIR database. This set of data is used to derive thresholds for network state definition and for calculation of resilience indicators. The selected datasets are parsed as reference for impact evaluation of tactical regulations.

- User input, ANM: allows the evaluation of different ANM lists from tactical phase. Once loaded, all of the values and figures in the tool is updated accordingly.

- Intuitive display, ECAC Network state: this part of the interface provides two comparative maps for pre-tactical (Figure C-14 right) and tactical phase (Figure C-14 left snapshot for 25.03.2019 at 20:33). Boxed numbers on the plots show the number of times that a regulation has been updated. Depending on the loaded list of ANMs and ACC sectorization maps, each regulated sector is colored. Regulation color-codes represent the relative reaction time (larger number represents a worse case):

$$\frac{\mathrm{Duration}}{\mathrm{WEF-PUB}} \tag{20}$$

LagTime (WEF-PUB) is the time window from ANM publishing time until start of regulation (The bigger the better).

C

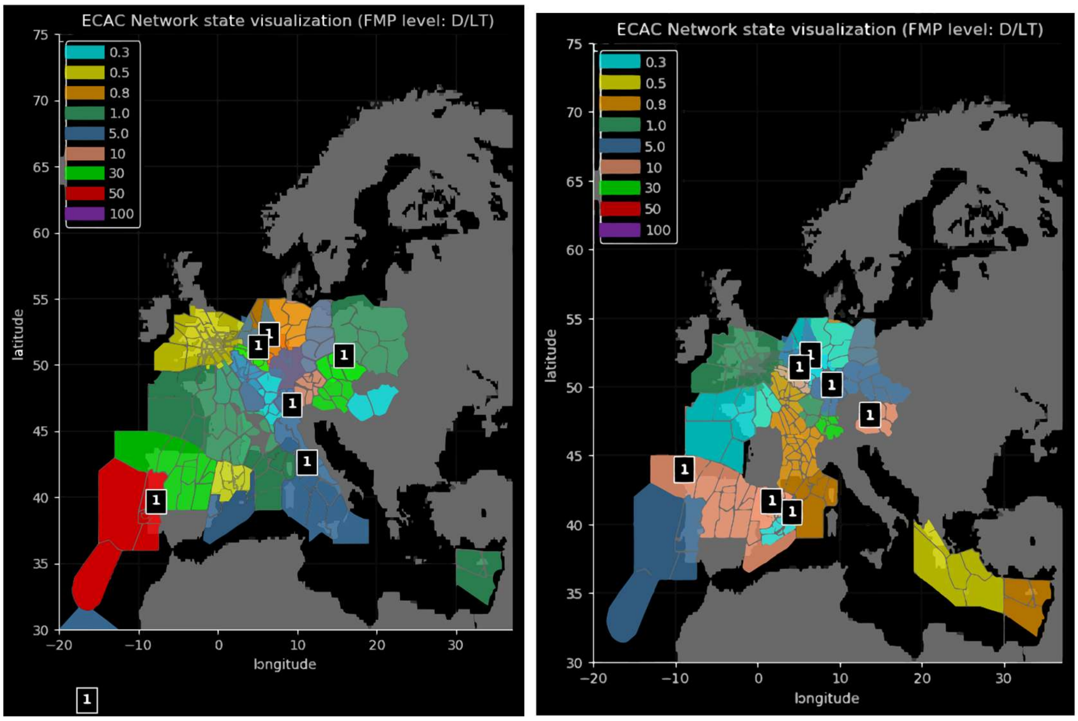


Figure C-14 NETRES, ECAC network state visualization

- Intuitive display, network evolution: similar to heatmaps, and with respect to daytime, bar charts show the evolution of regulations from planned regulations in pre-tactical phase to loaded tactical phase. Each bar represents number of different active regulation types to better anticipate and observe disruption peak times (Figure C-15). For instance, the tool shows time of expansion/resolution of a weather situation and ATC strikes (marked as ATC Ind Action).

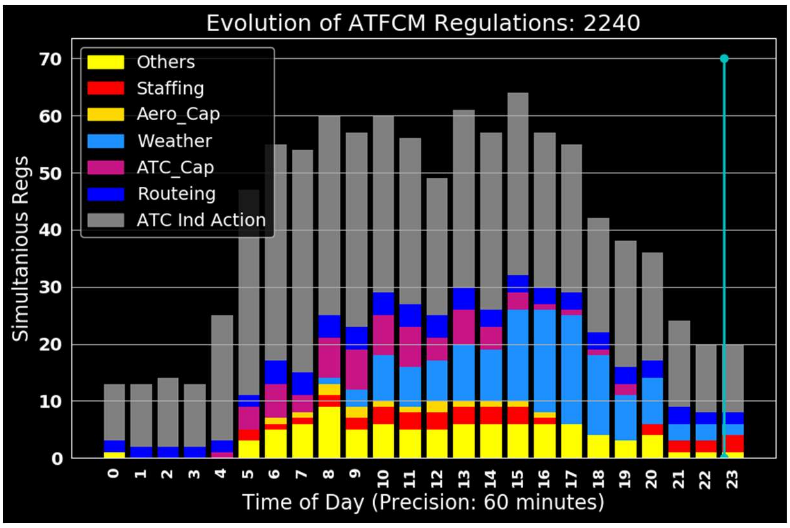


Figure C-15 NETRES, regulation evolution bar chart

- Quantitative display, NetState indicators: These indicators are expressed by mean and standard deviation values. Each one is calculated per FMP and for different type of regulations. Four indicators are calculated for both duration and lag time of regulations. For instance, the formula for mean duration is:

$$\mu_D = {\left\{\sum_{FMP}\sum_{RegType}\left(\frac{U_{RF}-x_t}{U_{RF}-L_{RF}}\right)\right\}}\Big/{N}\,, \tag{21}$$

U: *upper bound of calculated threshold per FMP for each regulation type,*
L: *lower bound of calculated threshold per FMP for each regulation type,*
x: *current mean for each FMP per regulation type,*
N: *number of regulation types for all active FMPs.*
U and L are derived from post-ops (NMIR), x and N are derived from tactical updates (ANM list), and each FMP may simultaneously have multiple regulations with same regulation type.

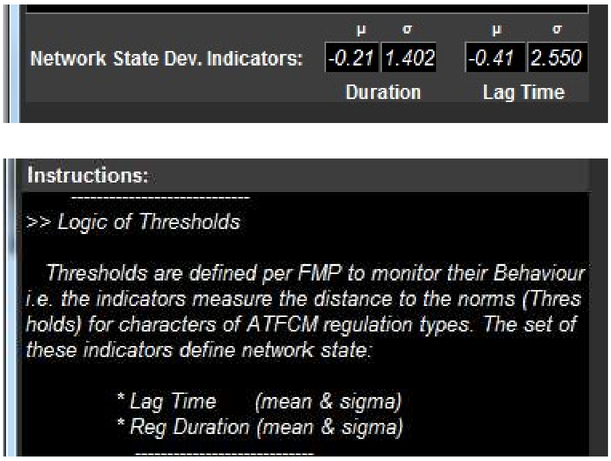


Figure C-16 NETRES, network state indicators

- Quantitative display, resilience indicators: these indicators (also mentioned partly in Table 3-2) are measuring the network disruption (Figure C-13) and expressed in two classes of magnitude and time-to-recover in Table C-1.

Table C-1 Resilience indicators that measure disruption

Name	Code	Formula	
Magnitude (domain)	MAGD	$\frac{\sum Regulated\ Traffic * \sum Reg\ Duration}{\sum Activation\ Notice}$.	(22)
Magnitude (impact)	MAGI	$\frac{\sum \text{ATFM Delay}}{\sum \text{MP regulated Traffic}} * \frac{\mu_{\text{Reg Duration}}}{\mu_{\text{Activation Notice}}}$.	(23)
Time-to-recover (imposed)	T2RP	$\frac{\sum Regulated\ Traffic}{\sum Activation\ Notice}$.	(24)
Time-to-recover (applied)	T2RA	$\frac{\sum MP\ Regulated\ Traffic}{\sum Activation\ Notice}$.	(25)

- Intuitive Display, network state: derived from the values of cumulative Netstate indicators, eight cases of network situations are classified into three major states: nominal, critical and crisis. Each of eight cases (four nominals, three critical and one crisis) is unique with respect to reviving strategy and its interpretation (Table C-2).

Table C-2 Different network states

Case	State	Description
N1	Nominal	Set of prolonged regulations are planned but corrective actions seen on some ACCs (Regional issue). $(\mu_D\colon High,\ \sigma_D\colon low,\ \mu_{LT}\colon High,\ \sigma_{LT}\colon High) *$
N2	Nominal	Set of prolonged Regulations are planned (risk of secondary impacts). $(\mu_D\colon High,\ \sigma_D\colon low,\ \mu_{LT}\colon High,\ \sigma_{LT}\colon low)$
N3	Nominal	Major performance loss is expected in pre-tactical phase (e.g. standby for severe weather or other predictable traffic flow management issue. $(\mu_D\colon High,\ \sigma_D\colon High,\ \mu_{LT}\colon High,\ \sigma_{LT}\colon low)$
N4	Nominal	Major performance loss is expected in pre-tactical phase (e.g. multiple local bottlenecks having flow management issues). $(\mu_D\colon High,\ \sigma_D\colon High,\ \mu_{LT}\colon High,\ \sigma_{LT}\colon High)$
C1	Critical	Set of prolonged Regulations may lead to significant loss of performance. $(\mu_D\colon High,\ \sigma_D\colon low,\ \mu_{LT}\colon low,\ \sigma_{LT}\colon High)$
C2	Critical	Performance loss can be significant based on the dispersion of regulation across the network. $(\mu_D\colon High,\ \sigma_D\colon High,\ \mu_{LT}\colon low,\ \sigma_{LT}\colon High)$
C3	Critical	Parts of the network suffer from significant performance loss (e.g. multiple local bottlenecks). $(\mu_D\colon High,\ \sigma_D\colon High,\ \mu_{LT}\colon low,\ \sigma_{LT}\colon low)$
CR	Crisis	Set of prolonged regulations limit network operations in upcoming hours. $(\mu_D\colon High,\ \sigma_D\colon low,\ \mu_{LT}\colon low,\ \sigma_{LT}\colon low)$

$*\ \mu_D$: mean disruption time, σ_D: standard deviation of blockage times for ACCs,
μ_{LT}: mean adaptation time, σ_{LT}: standard deviation of adaptation times for ACCs.

C.2.3. Added value

The design and demonstration of NETRES is recorded as an ATM operational and technical content development produced by SESAR 2020 Project. Figure C-17, provides the European Air Traffic Management Architecture (EATMA) model for network resilience [11] based on NETRES tool. Moreover, the benefits of data analytics and use of machine learning techniques in this tool draw the attention of NMOC to invest more on ML based solutions for managing network performance.

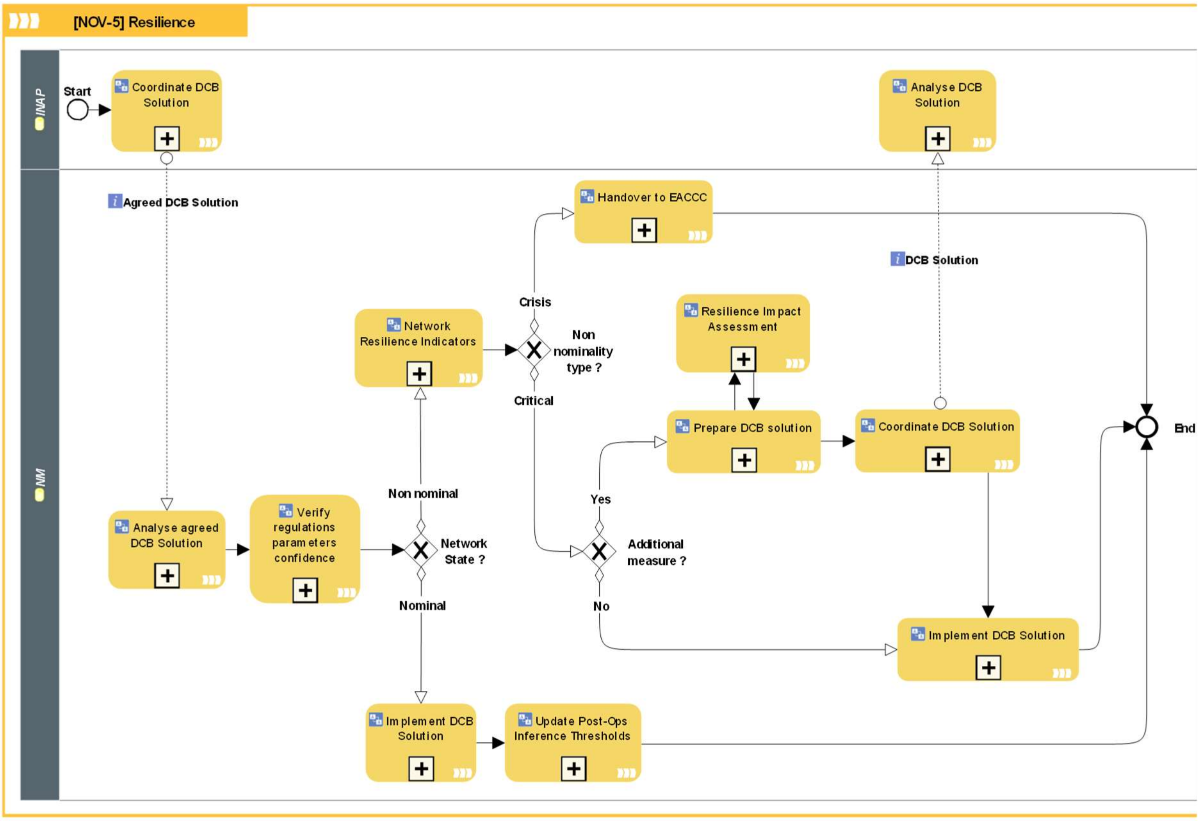


Figure C-17 SESAR EATMA model: network resilience *[11]*

More importantly, this design is allowing the realization of tradeoff analysis between different ANM lists, i.e. different corrective actions in form of capacity regulations. Also, as this study serves as an exploratory research, NETRES is making the conceptualization of network resiliency more transparent and provides the chance to build road-maps for such a concept.

On a technical level, the logs of the tool can be further analyzed in term of calculated indicators and network behavior in different network situations. Every time a new list is selected by the user the tool records a csv log file that include:

- Date and time; corresponding to the evaluated ANM list as input,
- Net_D_Mean, Net_D_Std, Net_L_Mean, and Net_L_Std; calculated values for NetState indicators according to evaluated ANM list,
- T2RP, T2RA, MAGD and MAGI; resilience indicators of magnitude and time to recover,
- MPR and ATFMD; predicted most penalized regulated flights and ATFM delay at the end of the day which ANM list is from,

- State and Comment; state records the assigned network state and the comment field records the reference month of NMIR post-operational data that are loaded in the tool to update all the values.