DataOps for Cyber-Physical Systems Governance: The Airport Passengers Flow Case

MARTIN GARRIGA, Tilburg University - JADS, The Netherlands
KOEN AARNS, Tilburg University - JADS, The Netherlands
CHRISTOS TSIGKANOS, TU Wien, Austria
DAMIAN A. TAMBURRI, Eindhoven University of Technology - JADS, The Netherlands
WILLEM-JAN VAN DEN HEUVEL, Tilburg University - JADS, The Netherlands

Recent advancements in information technology have ushered a new wave of systems integrating Internet technology with sensing, wireless communication, and computational resources over existing infrastructures. As a result, myriad complex, non-traditional Cyber-Physical Systems (CPS) have emerged, characterized by interaction among people, physical facilities, and embedded sensors and computers, all generating vast amounts of complex data. Such a case is encountered within a contemporary airport hall setting; passengers roaming, information systems governing various functions, and data being generated and processed by cameras, phones, sensors and other internet-of-things technology. This setting has considerable potential of contributing to goals entertained by the CPS operators, such as airlines, airport operators/owners, technicians, users and more. We model the airport setting as an instance of such a complex, data-intensive CPS where multiple actors and data sources interact, and generalize a methodology to support it and other similar systems. Furthermore, this paper instantiates the methodology and pipeline for predictive analytics for passengers flow, as a characteristic manifestation of such systems requiring a tailored approach. Our methodology also draws from DataOps principles, using multi-modal and real-life data to predict the underlying distribution of the passenger flow on a flight level basis (improving existing day-level predictions), anticipating when and how the passengers enter the airport and move through the check-in and baggage drop-off process. This allows to plan airport resources more efficiently, while improving customer experience by avoiding passenger clumping at check-in and security. We demonstrate results obtained over a case from a major international airport in the Netherlands, improving up to 60% upon predictions of daily passenger flow currently in place.

ACM Reference Format:
Martin Garriga, Koen Aarns, Christos Tsigkanos, Damian A. Tamburri, and Willem-Jan Van Den Heuvel. 2021. DataOps for Cyber-Physical Systems Governance: The Airport Passengers Flow Case . *ACM Trans. Internet Technol.* 1, 1 (May 2021), 25 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Recent developments in information technology, have fostered a new generation of enterprise applications that amalgamate powerful Internet technology with wireless networks, sensors, and, mobile devices [15]. As a consequence, Cyber-Physical Systems (CPS) have emerged as non-traditional systems characterized by a congruent interaction among people, physical facilities, and

Authors' addresses: Martin Garriga, Tilburg University - JADS, Den Bosch, The Netherlands; Koen Aarns, Tilburg University - JADS, Den Bosch, The Netherlands; Christos Tsigkanos, TU Wien, Vienna, Austria; Damian A. Tamburri, Eindhoven University of Technology - JADS, Den Bosch, The Netherlands; Willem-Jan Van Den Heuvel, Tilburg University - JADS, Den Bosch, The Netherlands.

1533-5399/2021/5-ART $15.00
https://doi.org/0000001.0000001

embedded sensors and computing resources. These systems generate vast amounts of data—also referred to as Big Data—with a huge potential to contribute to key goals in industry such as increased transparency, personalisation, and improved decision making [14]. However, the Big Data - CPS tandem has only caught traction recently and remains still underdeveloped. The complexity of CPS calls for systematic, proven yet innovative software and data engineering approaches—collectively referred to as *DataOps* [8]—that fluidly and continuously assume the specific traits of such systems [7]. This paper introduces a methodology to cater for on-the-fly analytics adopting DataOps as the paradigm of choice, while leveraging Big Data stemming from CPS.

Our target domain is one of the most complex and influential industries to date – the air transport industry. This reflects the intersection of CPS – think of the many sensors and devices active in any airport hall – and the Big Data generated therein. On the one hand, the air transport industry has experienced a consistent and sheer exponential growth over the past years. In 2017, a record of 4.1 billion passengers were transported world-wide. This growth will increase up to 5% per year over the next 20 years [1], bringing along operational challenges for airlines and airports. On the other hand, post-pandemic traveling requires novel insights regarding safety, security and health, to evolve key operational processes as check-in and baggage drop-off for departing passengers.

Recent literature has modeled the passenger flow at airports using various data, such as Wi-Fi indoor location [11; 16], security checkpoints information [19], or simulations [19]. Further, these approaches predict passenger flow applying deep learning [16; 17], Long-Short Term Memory networks [19], and Bayesian networks [11]. However, the state-of-practice entails informal forecasts by experienced employees, using rule-based systems with little automated or data-driven support. This stems from the Grounded Theory based in our in-site observations at the Europe's third busiest airport, and our interviews with managers, operators and staff from the airport and a major worldwide airline company (detailed in Section 3). Current methods in the aviation business do not exploit contemporary advancements, and neglect large volumes of sensor and device generated data. This results in losses for airports, airlines and passengers. In particular, for airports' landside operations, an adequate prediction of the passenger flow is crucial to guarantee security, safety, health, and overall wellness of passengers and airport staff. As an example, sometimes queues become long and chaotic, due to badly provisioned resources: less staff than actually needed at check in and security, poor spatial distribution, and little to no automated devices such as self check-in, baggage drop-off, or biometric e-gates. Meanwhile, at a different time of the day, or another sector of the airport, desk staff, security agents and technology remain idle.

The airport setting may be seen as a complex, data-intensive Cyber-Physical System (CPS), an integration of people, computation, data, and physical processes. Embedded devices monitor and control the physical processes, with feedback loops where physical processes affect computation and vice versa. In particular, we focus on passenger flow analysis, as it demands a tailored approach based on predictive analytics. Our methodology is based on DataOps and machine learning, combining and exploiting passenger profiling and location data from multiple sources (sensors, Wi-Fi, mobile apps), as input to a deep neural network, with the goal of predicting passenger flow more accurately. Our approach works on a fine-grained flight level basis, while state-of-practice predictions are daily. It also improves accuracy – predicting when, how, and how many passengers enter the airport and move through the check-in and baggage drop-off process. The resulting data pipeline blends within the CPS to then improve and optimize physical processes, closing the feedback loop.

Specifically, this paper features four main contributions:

(1) We model the target domain as a data-intensive CPS, featuring an interplay of multiple data sources, people, physical aspects and cyber/computational aspects interacting within a complex space.

(2) We propose a DataOps pipeline and methodology to bridge the Cyber- and Physical world, and explore bringing predictive analytics for the passenger flow forecasting problem.
(3) We instantiate our prediction pipeline in a prototype featuring four different predictive models (viz.: the existing Baseline Model, Simple and Extended Linear Regression, and Neural Networks).
(4) We evaluate predictive models within our pipeline over one year of real operational data from a major airport and airline[1], and demonstrate how our models substantially reduce prediction error with regards to the one currently used in practice.

The remainder of this paper is structured as follows. Section 2 presents an overview of the problem domain and its treatment as a CPS, along with the research questions. Section 3 then describes the methodology followed to design the solution, featuring domain and requirements analysis and data sources identification and analysis. Section 4 details data pipeline and predictive models for passenger flow. Section 5 presents the evaluation of our approach with a real case of a major airline and airport operations. Section 6 discusses results and limitations of our approach. Section 7 presents related work, and finally Section 8 concludes the paper.

2 AIRPORT PASSENGER FLOW AS A CYBER-PHYSICAL SYSTEM

CPS embody physical and engineered systems whose operations are monitored, coordinated, controlled and integrated by a computing and communication core [22]. By definition, CPS interact with the physical world, and are often required to operate dependably, safely, securely, and efficiently; modern CPS have to deal effectively with dynamic environments [33], and control their changing behavior in a resilient fashion. Figure 1 illustrates an airport as a complex CPS, where multiple actors and data sources interact with intertwined computational and physical aspects [34]. Essential features where the physical-digital domain 'interconnect' including (i) digital sensing that hooks into the physical realm and quantifies the physical world, such as location data from roaming passengers; (ii) DataOps – i.e., data engineering, quality and integration – intertwines through data the processes, goals and resources to perform analytics, and, (iii) actuation facilities that enforce controller-provided changes in the physical world, such as actions that change the passenger flow as a result of data-driven control decisions.

The physical aspect encompasses hardware, humans, rules, regulations, and business processes associated with aircraft and air transport systems. Additionally, performance goals, such as safety, capacity, efficiency, and aviation security, along with air transport system level performance metrics, such as flight delays, passenger flow rates, and air traffic densities. Of course it also includes aircraft management aspects and components [26], although those remain outside the scope of this work.

Conversely, computational aspects include mobile networking and information technology, embedded in airport systems. This notably pertains to performance goals, such as computer-software- and network- reliability and availability, information, computer and network security, data and location privacy. This has an operational footprint in the physical world, since it typically relies on additional physical components (e.g., networking equipment, mobile devices) and physical resources to function.

2.1 Research questions

We have observed that predicting the passenger flow in an airport is a characteristic problem that can be addressed from a new viewpoint – considering and modelling the airport as a complex CPS. We pose the following Research Questions (RQs), following a *model-predict-explain* pattern [30]:

[1] Some data are omitted for confidentiality reasons.

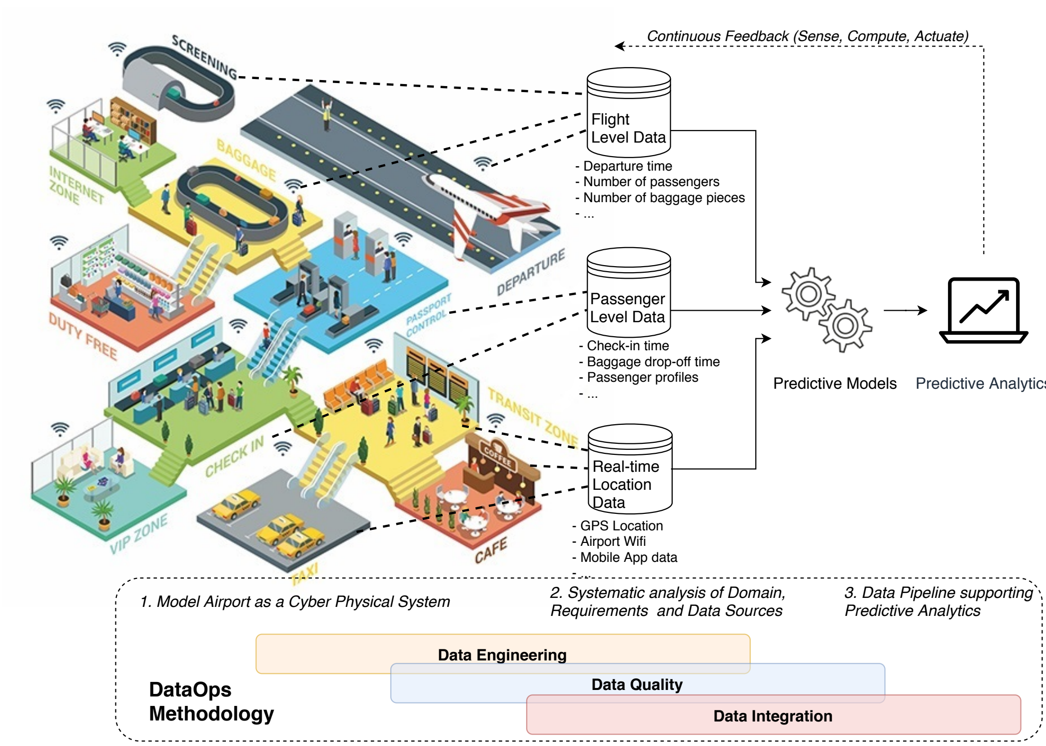


Fig. 1. Emerging Cyber-Physical System view from the airport setting, supported by a DataOps methodology for predictive analytics.

RQ1: How to *model* the problem domain by means of DataOps and Cyber-Physical Systems? First we need to understand the underlying business processes, composite system and domain. To this end, we carried out stakeholder interviews and in-site observations, driven by Grounded Theory as the systematic methodology to develop (and validate) the system model from domain knowledge.

RQ2: How to *predict* passengers flow more accurately? Afterwards, we illustrate the application of DataOps for CPS by designing a data pipeline for the problem of passenger flow prediction. The pipeline captures and combines multi-modal data streams from different heterogeneous and distributed sources to improve existing predictions in terms of accuracy, time and consistency. We develop several predictive models for passenger arrival behavior, and experimentally validate our proposal over a real case, comparing against the baseline model currently employed in practice.

RQ3: How to *explain* the prediction results in terms of the most important features? Within predictive models, there is a trade-off between explainability and performance [25; 41; 42], where more complex models such as neural networks usually outperform better interpretable models such as linear regressions. We explain the prediction results in terms of the most relevant features.

Figure 1 presents an overall view of our approach in the context of DataOps (bottom part), and shows how data engineering, quality and integration help to address the problem at hand. In the previous section we discussed how to model the problem as a complex CPS – Step 1 in Figure 1. In the following section (Section 3 we perform a systematic domain analysis and identify the available relevant data sources – Step 2 in Figure 1. Then, we model a pipeline and architecture (Section 4) to support predictions. Finally we evaluate a prototype of the approach, comparing against existing predictions (Section 5) – Step 3 in Figure 1.

3 METHODOLOGY

We followed the Design Science Research Methodology (DSRM) [21], whose goal it is to develop knowledge that the professionals of the domain – being data scientists, process owners, data engineers and decision makers on airports and airlines – can use to design solutions for their field problems. DSRM prescribes the following six key steps: (1) problem identification and motivation, (2) definition of the objectives for a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication.

Problem identification and motivation. Passengers arrive at distinct times prior to departure to the airport, posing a challenge to act upon the crowds and potential congestion during a day of operation. Lack of information regarding arrival times and service demands for decision makers in the check-in process is the main underlying problem. This problem is two-edged: firstly, long queues (at check-in, baggage drop, and security) and poor passenger experiences are undesirable; secondly, they result in idling employees, sub-optimal usage of resources and associated costs.

Objective of the solution. Development of a DataOps methodology to provide more accurate insights in the CPS at hand. We will focus on passenger flow at the departure section in an airport. This means better insights in the number of passengers that enter the check-in process throughout a day of operation, with the second-order goal to improve resource staff planning. The solution should incorporate passenger-specific parameters that possibly affect their arrival time, and novel data sources as model features, such as the travel motive (e.g., provided by a revenue management department), as well as mobile device and sensor data with arrival behavior (e.g., provided by a digital application team).

Design and development. We designed and developed a data pipeline that provides end-to-end support for predicting passenger flow. Particularly, it extracts, merges and transforms data from heterogeneous data sources throughout the CPS. Then, it feeds the curated and integrated data to a predictive model and presents the results.

Demonstration. A proof-of-concept prototype fed with real data demonstrates the potential of our approach. We implemented several predictive models to assess the feasibility, applicability and accuracy of the approach. The prototype could be potentially deployed and integrated into the processes of any airport or airline.

Evaluation. The proposed models are then compared showing substantial improvements against an existing baseline, both at flight level and daily level. We show that the quality-interpretability trade-off of machine learning models can be influenced in favor of more complex models by using SHapley Additive exPlanations (SHAP [18]). SHAP explains predictions in terms of the contribution of each feature to the final result, with the final goal of easy interpretation and actionability.

Communication. We communicate results in an agile manner, involving managers and decision-makers in periodic feedback sessions. Results are presented on different levels of technical detail – including, but not limited to, this paper – to achieve highly actionable and interpretable results. For example, we calculated the Mean Absolute Error of the predictions in terms of number of passengers every 15 minutes, which is meaningful for the domain experts.

The remainder of this section covers the methodology we followed to distill relevant requirements from the problem domain. The broad set-up was inspired by the Cross Industry Standard Process

Table 1. Key stakeholders interviewed for constructing the grounded theory (names and companies omitted for confidentiality reasons).

Position	Description
Airline R&D Manager	Senior Manager of R&D prototyping environment within the Airline, in collaboration with universities and research groups.
Airline Analyst Functional Programmer Airport Officer	Senior and junior users of the legacy passenger prediction model
Chief Data Officer	Senior officer of the Operations Decision Support department at the Airline.
Tactical Planner	Junior airline employee estimating the workload for a specific period.
Product Owner of ERP tool	Semi-senior user of the Enterprise Resource Planning tool taking as input the graphs made by a tactical planner and producing a plan that assigns workers to tasks.
Landside Manager (HR Planner)	Senior manager, end-responsible for the departure hall, who decides on the amount of personnel to hire for a specific day.
Operational Assignment Planner (HR Planner)	Semi-senior employee that performs swaps of employees during a day of operation when certain check-in desks are more busy than others.
Unit Manager Passenger Flow Control	Senior manager, responsible for a specific unit within the Passenger Services department. Main responsibility is to guarantee a (close-to) continuous flow of passengers.
Manager of Planning Passage	Senior manager, directly responsible of the operational assignment employees (see above).
Shiftleader	Senior manager, direct responsible of the employees and operational part in the departure hall.
Duty Manager	Senior manager of the shift leaders, oversees a large part of the passenger operations.

for Data Mining Model (CRISP-DM [39]). Section 3.1 presents domain and requirements analysis whilst Section 3.2 presents data sources analysis. These steps lay out a basis for a methodology that can be followed to implement predictive analytics in a CPS within any organization.

3.1 Domain & Requirements Analysis

We follow a Grounded Theory (GT [4]) systematic approach to understand critical processes such as passenger planning, staff scheduling and luggage handling. Constructing a GT starts with data collection, usually in the form of observations and interviews, afterwards a more general theory emerges from the analysis. First, we conducted unstructured interviews with key stakeholders, and observation of their daily tasks, summarized in Table 1 (names and companies are omitted for confidentiality reasons). These include but are not limited to HR planning and managers.

We discovered several attractive points both at landside and airside, including passenger flow baggage processing and runway operations. All of these have typically dedicated operational teams both at the airline and airport. In particular, operations in baggage handling and passenger flow are highly correlated: modelling such processes as CPS paves the way for applying Data Science methodologies and generate fitted prescriptive models.

All in all, focusing on passenger flow, the resulting theory is conceptualized in Figure 2. This confirms that a smooth passenger flow is essential for the operations in the departure hall, customer satisfaction and lately for safety and health of both passengers and employees.

Then, the planning process of ground agents was elicited, which is summarized in Figure 3. The predictions and flight schedule are used as input for the tactical planning department to create long-term plans and contract (external and flexible) resources. Next, the planning system generates

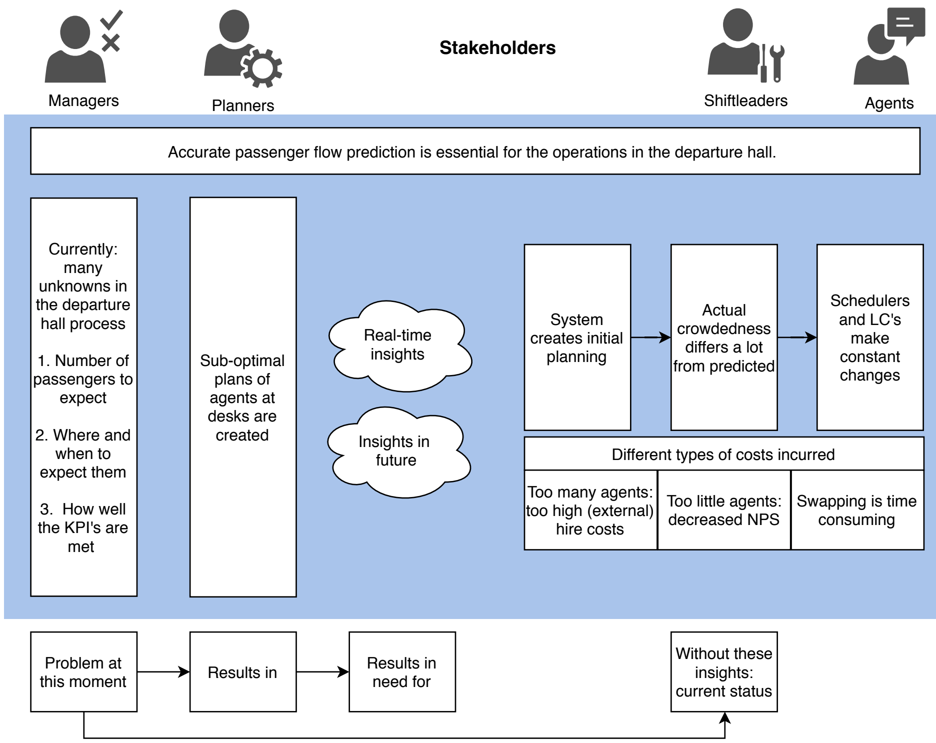


Fig. 2. Summary of the results of the Grounded Theory (names, companies and numbers omitted for confidentiality reasons). *Legenda: NPS stands for Net Promoter Score, LC for Landside Coordinator.*

a plan for a specific day with the number of personnel needed, two weeks in advance. Finally, the operational assignment department assigns employees scheduled to work during a working day, up to four days in advance. After that, no more changes can be done without incurring extra costs. However, the following three problems occur:

(1) The number of passengers that are predicted is *too large*. In this case, agents end up being idle resulting in resources over-provisioning – something not desirable, since the majority of the workforce is externally hired at relatively expensive rates;
(2) *Too low numbers of* outbound passengers are predicted. In this case, agents at check-in desks cannot cope with the demand – i.e., resources under-provisioning – causing queues and passenger waiting times to increase. This negative effect on customer satisfaction directly translates to a decrease in revenues and health, safety and security risks.
(3) Passengers enter the process *at different times* than initially anticipated. This incurs expensive swaps between agents that have to be performed during ongoing daytime operations.

We followed up with airport operations stakeholders to obtain more supporting data, in particular regarding arrival profiles of passengers. Figures 4a and 4b confirm that parameters such as the number of days a passenger booked her ticket in advance and the length of stay of a passenger influence the arrival behavior. In addition to passenger specific features, flight specific characteristics influence passenger behavior. These *gut feelings* from the different stakeholders are confirmed by data, summarized in Figures 4c and 4d. The hour of the day a flight leaves and its destination also influence the arrival behavior of passengers. The identified problems can be mitigated by gaining

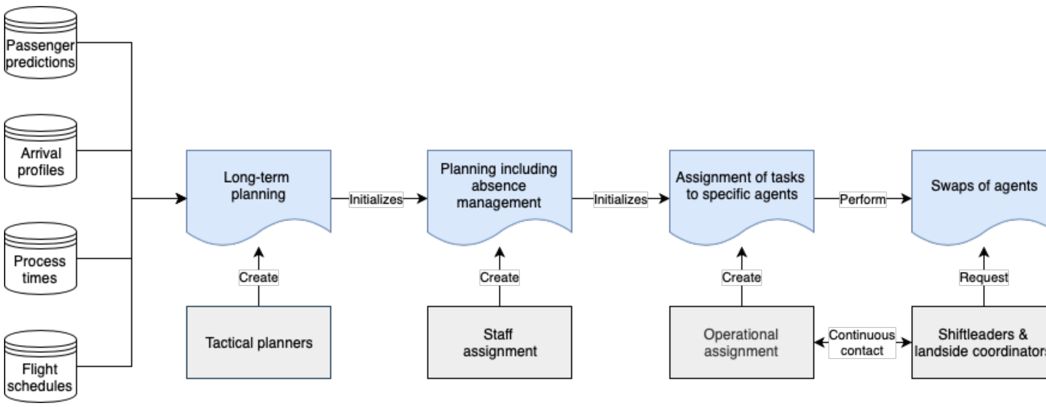
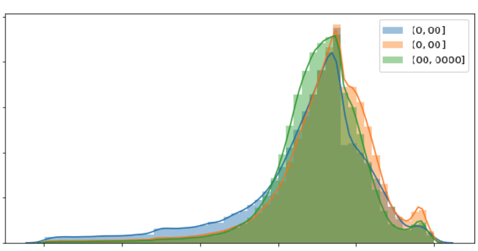
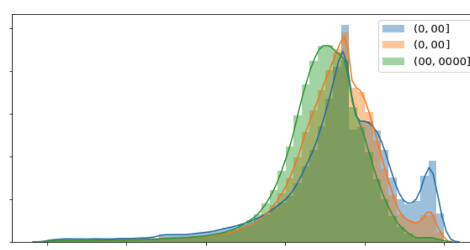


Fig. 3. The current planning process of the departure hall.

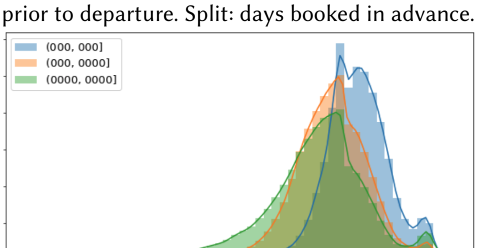
insights in the type and number of passengers expected during specific times of the day, based on the available data generated within the CPS.



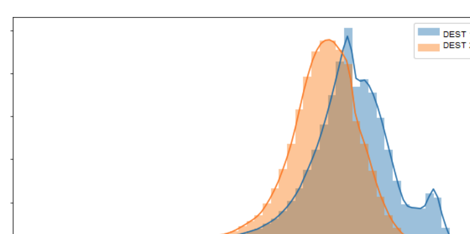
(a) Arrival profiles. y-axis: probability. x-axis: time prior to departure. Split: days booked in advance.



(b) Arrival profiles. y-axis: probability. x-axis: time prior to departure. Split: length of stay.



(c) Arrival profiles. y-axis: probability. x-axis: time prior to departure. Split: scheduled departure time.



(d) Arrival profiles. Split: Continental vs Intercontinental.

Fig. 4. Probabilities of arriving to the first touch point at the airport, based on different factors (actual values omitted for confidentiality).

3.2 Data Sources and Analysis

To understand available information in the airport setting, first we interviewed managers and senior employees both from the airline and airport operations teams. Then, we traced back the available data sources in use (or in lack). Two major categories emerged: (i) flight level and (ii) passenger level data.

Flight level data. Those capture basic flight details, such as departure gate, type of traveler (e.g., transfer and intercontinental passengers) carrier, departure and arrival airport, and, configuration of the plane. Secondary data pertains to time-related flight data, providing opportunities for analysis: departure times, delay times, boarding start times, and actual and estimated turnaround times. Furthermore, boarding intelligence data includes the number of passengers expected/present in the flight, predicted boarding times, current boarding speed, and target boarding speed. Similarly, baggage intelligence contains data on offloaded bags and the number of bags checked in for the flight. Finally, booking information such as number of bookings, split by class and status (frequent flyers, regrading, etc).

Passenger departure data. The most relevant are timestamps obtained from boarding pass scans. The major ones record the first baggage drop-off time and the check-in time. These can be superimposed with passenger profiles, such as cabin class, gender, nationality, check-in channel type, length of stay, and number of passengers in the booking. Other fields within departure data are passenger status, passenger-level predictions (if any), connection and security information. Finally, travel motive is likely to influence passenger behaviour, as those travelling for business reasons have a fixed (business) destination with fixed timing, whereas leisure passengers typically have more fluid schedules. The actual formula used for this business-leisure segmentation is typically confidential and internal to airlines, as it entails price segmentation – however, the main drivers are destination and length of stay.

Passengers location. is critical to embark on (more) accurate prediction activities. However, such information is not part of the in-place prediction models. This data could be obtained through sensors within mobile devices – passengers that use airlines' mobile applications may be tracked by means of GPS and WiFi beacons. Indoor positioning literature has produced a vast number of approaches for inferring positions and trajectories of devices roaming inside a spatial environment, such as an airport [3; 35]. Those can be instantiated for this type of data source.

Customer satisfaction data. Airlines have customer satisfaction departments that periodically gather such information through standardized passenger-oriented questionnaires. Their main goal is to determine the *Net Promoter Score* (NPS) [24]; an industry-wide key performance indicator (KPI) that measures how likely a passenger is to recommend the airline to others. Therefore, customer satisfaction data encompasses background information on the passenger, such as personal details and flight information. Then, the questionnaires split into departure, on board, and after flight segments. Note that NPS is an airline-oriented KPI. For airport-oriented benchmarks, one could use for example ASQ (Airport Service Quality)[2], as another primary data source of passenger satisfaction evaluation.

At this point, we can combine the insights gained by modeling the business domain as a CPS (recall Figure 1) with the key processes and data sources identified and analyzed through Grounded Theory. As an elaboration, Figure 5 presents the actual passenger flowchart which captures such knowledge. The rectangles represent a *touch point*, where a passenger interacts with an employee or a system. Aligned with self-service concepts, automated devices for the boarding procedures – such as self-check-in devices and biometric e-gates – remarkably shorten the standby time of passengers, and improve the overall flow. Diamonds represent splits in the passenger streams, allowing to distinguish distinct cohorts of passengers, e.g., whether a passenger checked in at home

[2]https://aci.aero/customer-experience-asq/

or at the airport, as this affects service needs. Paths and decision nodes in the flowchart can be annotated with concrete values based on data for each case (omitted here due to confidentiality).

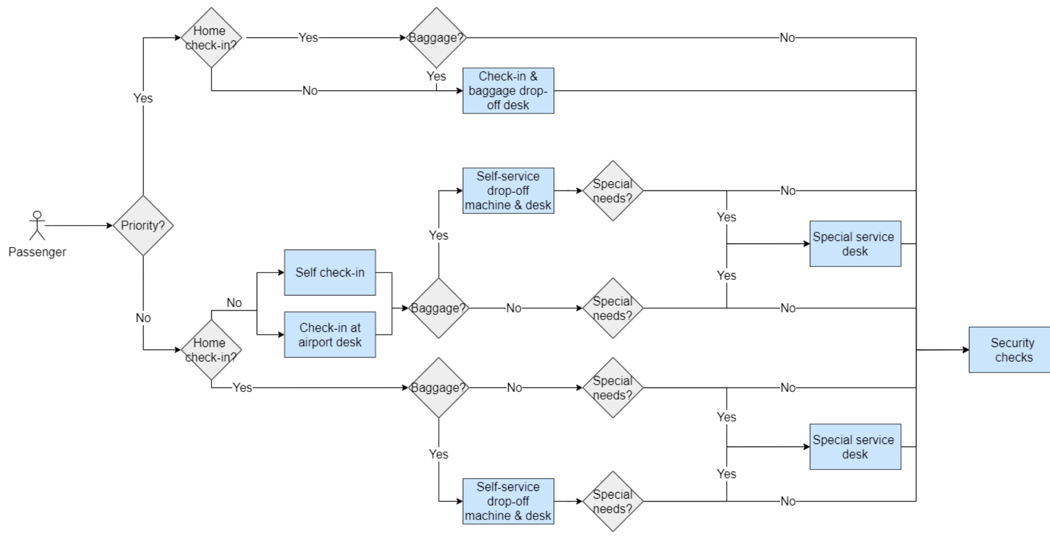


Fig. 5. The different passenger streams present in the departure hall.

4 DATA PIPELINE AND MODELS FOR PASSENGER FLOW PREDICTION

A data pipeline typically provide end-to-end support, from data identification, extraction, curation to their integration, analysis and visualization [8]. Figure 6 depicts a generic, end-to-end architecture of a DataOps pipeline that will cater for CPS-aware predictions. Notably, it can be dynamically adapted to detect novel data sources on-the-fly, and accommodate (close-to) run-time data integration. We designed and developed a prototypical implementation of this pipeline architecture for experimentation, exploration and validation in the context of this work.

In particular, the exploratory, prototypical DataOps pipeline for passenger flow prediction fuses the data sources as described in Section 3.2 into one, unifying data lake. The data lake stores conventional structured data such as, flight data and passenger data, along with data stemming from mobile devices and sensor networks, e.g., passenger locations data tapping into mobile apps and Wi-Fi. In this way, prediction models may exploit for example more accurate data about the time between a passenger entering the departure hall and the scheduled time of departure.

In the next stage of the DataOps pipeline, the data sets are merged and curated, based on syntactically and semantically reconciled passenger identifiers. Here one can apply relevant filters (e.g., first-leg departure station) and data cleaning steps (e.g., handling missing or inconsistent values), as preparation for analytics. This master data set may then be effectively used for training the predictive models (last stage of the pipeline), make predictions, and finally visualize results.

4.1 Predictive Models

This section outlines the predictive analytics models adopted for CPS-enabled passenger forecasting. We first introduce a baseline rule-based model before discussing two linear regression models and a neural network.

Baseline Model. Deterministic and stochastic passenger prediction in the airline industry is a well-established practice. During the domain analysis (see Section 2) we elicited a rule-based

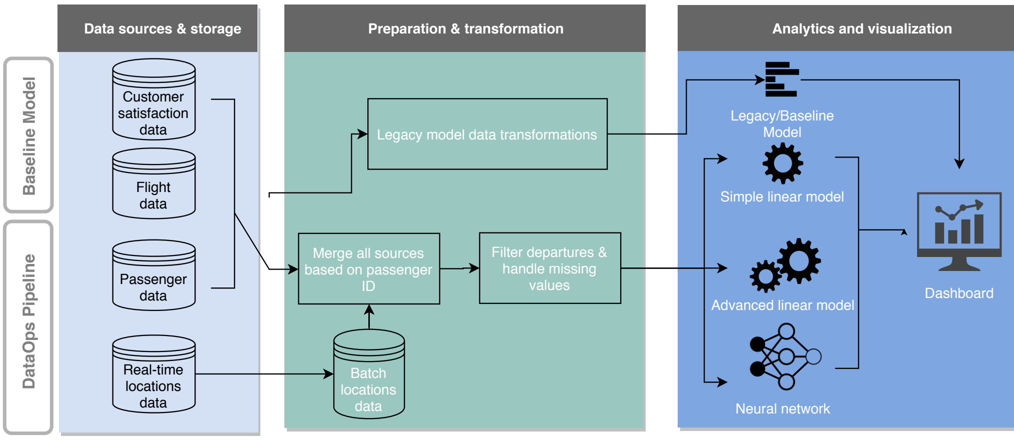


Fig. 6. Data pipeline for passengers flow prediction, integrating several data sources (actual names of the components omitted for confidentiality)

model developed by a specific airline, currently in operation. This model serves as the baseline, demonstrating what advanced CPS-based analytics can bring. The model comprises three key steps, from booking predictions to passenger arrival predictions. First, the model predicts the amount of passengers that will board each flight. This exploits data about available bookings for future flights and historical bookings of the past three years. Second, it computes several booking trend scenarios that determine the impact of increases/decreases of booking times. Finally, the model selects the *Best Booking Curve* by minimizing the *cost* function: the difference between the actual bookings and historical bookings. This is formalized in a formula that comprises the historical booking curves, which is confidential. Sample prediction and booking curves are shown in Figure 7 (actual curves and numbers are omitted for confidentiality). The closer to departure, the more bookings are made. The middle line represents the actual bookings made up until now. The dotted line represents the predictions that the model makes for a given flight, following the booking curve that lies closest to the current bookings.

Nowadays, static arrival profiles capture the ratio of passengers arriving at a given time before a flight takes off. These profiles take into account the departure and destination of the flight, as well as the travel class, among other indicators. These profiles are combined with scheduled flights, resulting in a prediction of the number of passengers during a day of operation. These predicted passengers are then distributed according to required services, such as check-in or baggage drop-off at the airport – as shown in Figure 5. Currently, the distribution is based on historical averages. The target variable of the model is the amount of passengers that need service during each 15-minute period for a day. Upon these predictions, managers and HR planners generate schedules to determine the staff (number and type) that should be allocated, and potentially hired externally for that specific day.

Towards Dynamic Prediction. The baseline method as presented in Figure 7 is static, and can be regarded as a rule-based model. It uses historical data and forecasts by senior employees to generate the static profiles that determine the prediction. In contrast, we propose a CPS-enabled, dynamic and data-driven methodology in which the target variable is predicted for each flight on-the-fly, rather than on a deterministic, daily basis. Our predictions are dynamic rather than static, as more features can be added to the model to test and improve its performance. The models proposed in this work directly predict the amount of passengers to arrive in the departure hall

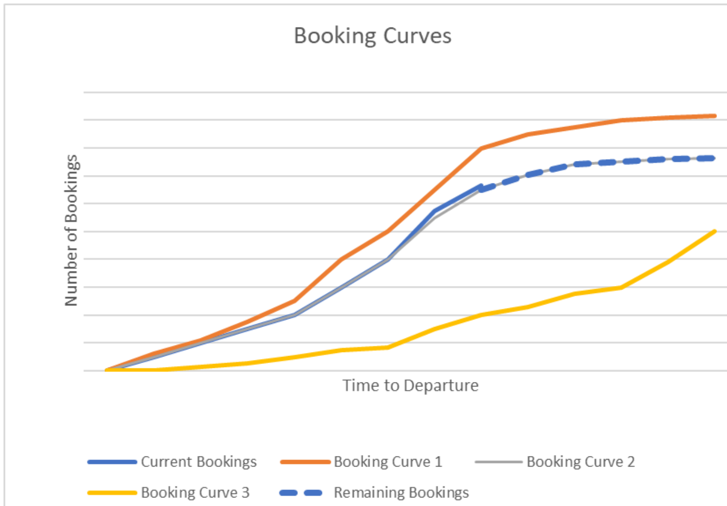


Fig. 7. Sample Passenger predictions through the *booking curve* method (real curves and values are omitted for confidentiality).

at 15-minute time intervals before the flight departure, and can then be aggregated when several flights depart within the same period. We propose here three models: (i) simple linear regression, (ii) extended linear regression, and, (iii) neural network. In the following, we briefly discuss these three models along with their complexity and prediction performance.

Linear Regression. The first model is a simple linear regression, which works well for mapping numeric inputs to numeric outputs, fitting an optimal correlation to the data points [37]. As its main advantage, the linear regression model is easy to understand and extrapolate, but its strong shortcoming is that it departs from the strong assumption that the features exhibit a (very close to) linear relationship with the target variable. The *simple model* predicts the amount of passengers to be expected in the departure hall every 15 minutes for each flight. Only the following basic features are factored in: `destination, scheduled time of departure, time interval`. A more complex model was also considered, adding more features. In particular, the *extended model* builds on top on the simple linear regression one, by adding more generic and aggregate features, including all features except the booking numbers.

Neural networks. Neural networks capture non-linear relationships between features and target variables, by means of more complex prediction structures [40]. Neural networks consume the same features as the extended linear regression, plus the actual booking numbers for different query moments. As such, they include all 23 features from the four different data sources. Different neural network architectures can yield different results. We investigated several configurations to determine the right trade-off between several architecture decisions – e.g., base architecture such as Deep Neural Network (DNN) or Long-Short Term Memory (LSTM), internal layer count, neurons count, etc. These experiments determined that a three-layer DNN was the best trade-off between accuracy and performance with respect to available infrastructure at the deployment site (which is *on premise* rather than cloud). A model complex enough to capture the structure of the data was required, but without becoming a time bottleneck that would take too long to make predictions. The relatively straightforward nature of the data does not require many layers, but only enough to capture the complex relationships between the features. Specifically, the input layer contains 23 neurons, corresponding to the number of trainable features. Then, three fully connected *hidden* layers with varying dimensions are added. We also experimented with three different shapes, given by the amount of neurons per layer: (i), a network with equal amount of neurons per layer, (ii) a diamond-shaped structure with more neurons in the middle hidden layer, and (iii) a decreasing amount of neurons per layer – as shown in Figure 8. In addition to the shapes of the layers, we experimented with the ratio of neurons between each layer: 2x or 4x number of neurons more/less between consecutive layer. All layers implement a ReLu (Rectified Linear Unit) activation function [6]. ReLu is a linear function that outputs the input directly if it is positive, otherwise zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance than sigmoid (S-shaped) functions [13].

The lowest number of neurons in the intermediate layers was set on 64. Then the amount of neurons was increased with a factor of 2 or 4. The output layer contains a linear activation function and one neuron, to predict the amount of passengers directly. The loss function of the model is optimized by the Adam algorithm [12]. We used a default batch size of 256 input samples per time (before the model is updated) and 500 epochs (training steps). An early stopping method was configured, to automatically stop training after the validation loss had not been improved for three training epochs. This allowed us to save training time and computational resources.

Six models were evaluated: *equal* shape (64 or 128 neurons), *diamond* shape (2x or 4x neuron factor per layer) and decreasing shape (2x or 4x less neurons per layer); at five relevant query moments: 14, 7 and 4 days, 24 hrs and 2 hrs in advance. This generates a total of *30* experimental configurations *(model, query-time)*. In the following, we report on the results of evaluating the above-explained prediction models and configurations with multi-sourced data from more than one year of operations of a major airline at a major airport.

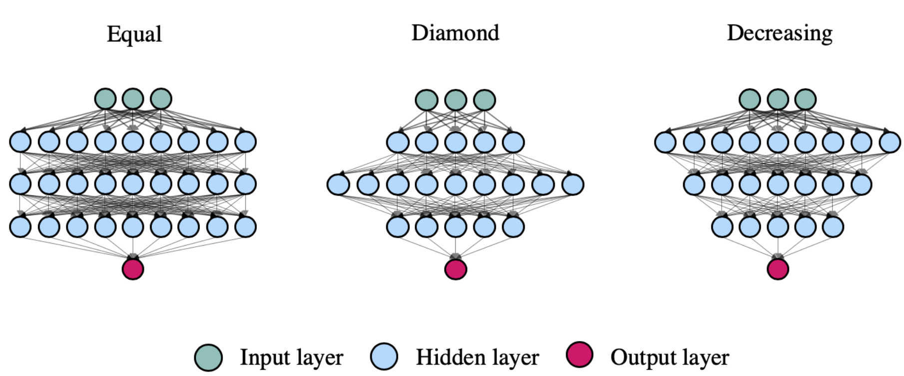


Fig. 8. Three different neural network architectures validated for the prototype.

5 EVALUATION

In this section, we evaluate the DataOps pipeline and CPS-enabled prediction models introduced above to a real case study on a major airline. Specifically, we assess the *accuracy* with regards to a baseline model currently used in the aviation industry. Accuracy will be measured in terms of the *error*, or the difference between actual and predicted number of passengers. To this end, we trained the models with more tha one year of operations data from the airline. This data was not restricted to conventional structured repositories, but also included CPS data such as GPS, sensors and devices. First, we introduce the datasets and experimental setup. Subsequently, we compare the different models in terms of prediction error. Finally, we discuss the major contributions and limitations of our methodology.

5.1 Experimental Data and Configuration

We populated the data sources that have been introduced in Section 3.2; the resulting dataset we exploit is in the order of millions of data points, and a time span of 16 months. Due to privacy reasons regarding the companies involved, we cannot disclose the dataset. The key features in the data can broadly be divided into three categories: generic features, time-based aggregates, and flight-based aggregates. We further generate two types of aggregated, auxiliary features. In particular, for each flight (ID), the mean values of the aggregate features are computed. Similarly, the mean values of the aggregate features are computed for 15-minute time brackets prior to a flight departure.

The goal of the models is to dynamically predict the amount of passengers that need service in the departure hall processes. Services in this process can be two-fold: (i) check-in at the airport and (ii) baggage drop-off. To model passenger arrival behavior in the departure hall, so-called arrival profiles are created. An arrival profile, shown in Figure 9, is the distribution of passengers that checks-in and/or drops off their baggage at a certain number of minutes prior to flight departure. By considering both check-in and baggage drop-off time the coverage of the target variable increases, as shown in Figure 10. Hence, the target variable defined in this experiments is the amount of passengers that need service in a 15-minute time interval for each flight.

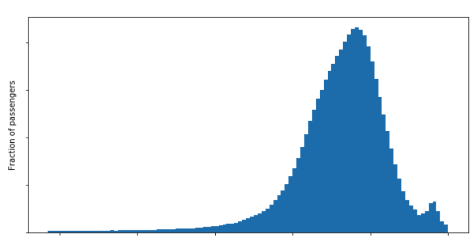


Fig. 9. Arrival profile. Target: baggage drop-off and airport check-in time (values omitted for confidentiality).

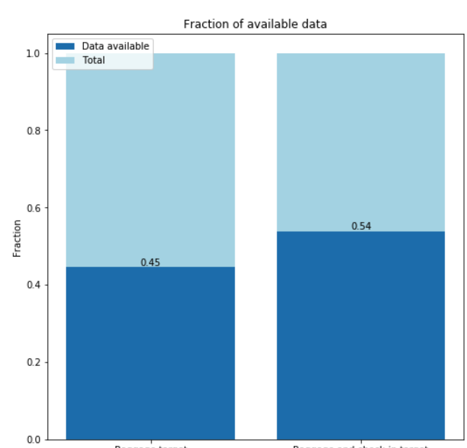


Fig. 10. Coverage of the two target variables over the whole set of passengers (values changed for confidentiality).

Figure 11 illustrates the daily number and distribution of relevant passengers in the data set. The upper line represents the total, the middle line represents those that need service, and the bottom line designates the ones tracked by the mobile app. Note that the latter are only available from July 2019 as CPS-related data collection and analysis from the mobile app is part of the contributions of this work, and consequently not considered before for the baseline model. All models have been trained on one year of *training* data, from 01-05-2018 to 30-04-2019. The *test* data for all models encompasses from 01-05-2019 to 10-09-2019. The five relevant query moments include 14 days, 7 days, 4 days, 24 hours, and 2 hours prior to departure.

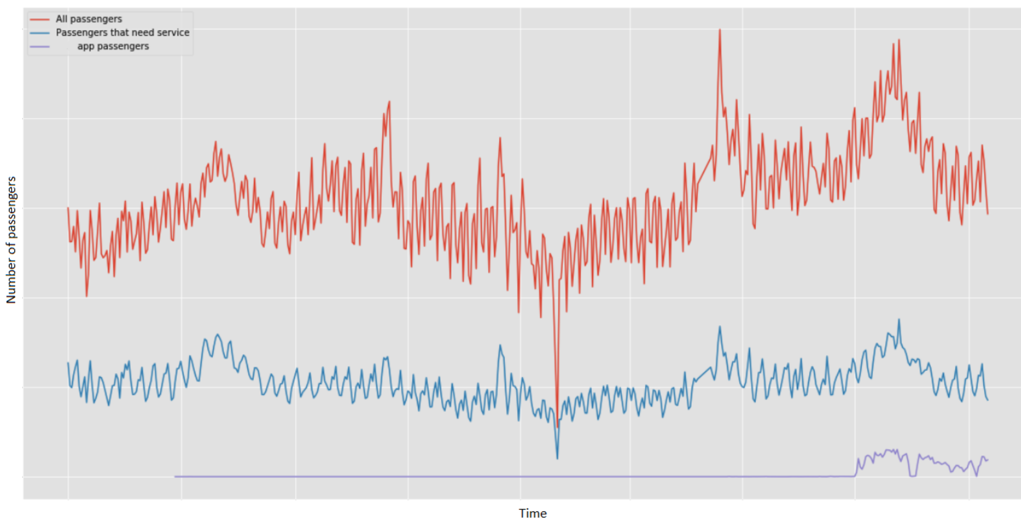


Fig. 11. Daily variation of passengers in the dataset (values and mobile app name omitted for confidentiality).

The performance of all proposed models was compared by the Mean Square Error (MSE) metric. MSE measures the average of the squares of the errors between the predicted and actual values and penalizes outliers, i.e., big deviations of the predictions from the actual data [38]. Outliers represent moments when current predictions fail excessively, harming the operations processes.

5.2 Results

We first compare the different combinations of neural network architectures and query times, to find and select the best performing one – as reported in Table 2. Then the selected neural network is compared against the other models on flight level and daily level. For doing so, we use the predictions obtained for query moment *4 days in advance*. This is the most representative moment, since personnel assignment can be adjusted up to four days in advance – recall Section 3.1.

Table 2. Mean Square Error (MSE) for test data on different neural network models and query moments. Best results in bold.

	Query moment				
	14 days	7 days	4 days	24 hours	2 hours
64 neurons per layer	11.83	**11.06**	11.36	11.19	11.07
128 neurons per layer	11.66	11.48	11.09	11.36	11.99
Diamond-shaped network: 2x factor	**11.09**	11.41	**11.08**	**11.03**	11.3
Diamond-shaped network: 4x factor	11.2	11.14	11.22	11.23	**10.82**
Decreasing no. of neurons: 2x factor	11.32	11.31	11.3	11.05	11.57
Decreasing no. of neurons: 4x factor	11.63	11.57	11.29	11.2	11.35

Table 2 compares the different permutations correlating neural network models versus query timing (varying between 14 days and 2 hours). The diamond model with a 2x neurons difference per layer performs best for the query moment of 4 days ahead – also for 14 days and 24 hours. As such, the results of this model will be further compared and analyzed against the other models on different levels of granularity. The training and validation values across epochs for the best performing neural network is visualized in Figure 12. Note that the test loss is stable and train loss keeps decreasing, which indicates no over-/under-fitting.

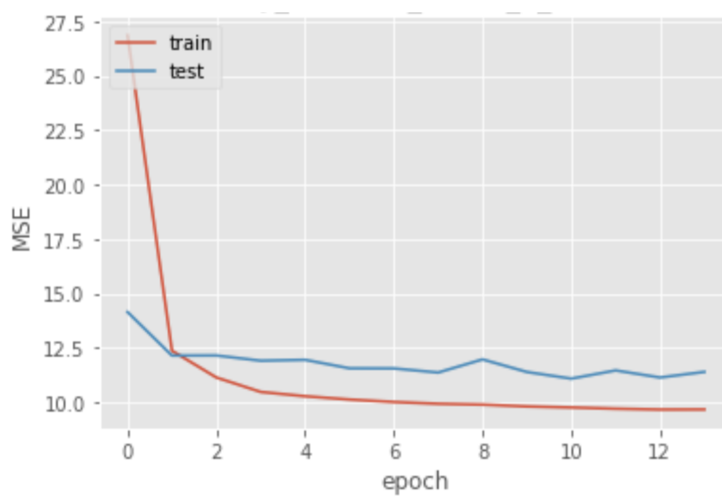
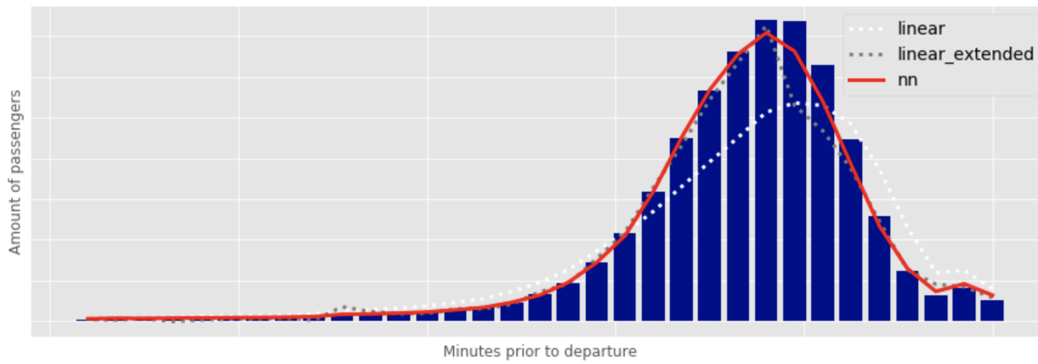


Fig. 12. Training and validation loss of the best performing neural network with query moment *4 days in advance.*

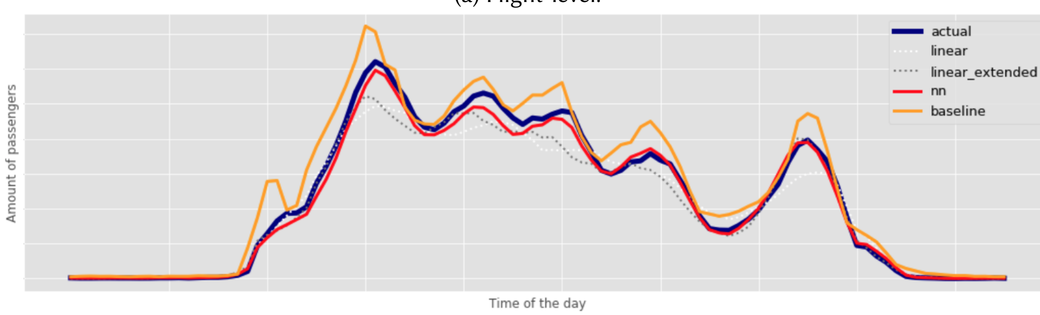
Flight-level results. The results of the best performing neural network, simple linear regression, and extended linear regression are presented in Table 3a. The models are evaluated on the MSE metric (defined in previous section), and also the Kolmogorov-Smirnov test (KS), a non-parametric test to compare a sample with a reference probability distribution. It measures how well the predicted distribution suits the underlying actual distribution [23]. That is, in Figure 13a, the fit between the lines (predictions) and the bars (actual distribution) through the entire test period. Note that the neural network captures the distribution much better than the linear models, confirming the insights from Table 3a.

Day-level results. In addition to the flight level results presented before, Figure 13b shows the daily aggregate predictions as used during airport operation. The dark line represents the actual amount of passengers that need service. The other lines represent predictions of our CPS-enabled model predictions and the baseline (legacy prediction). Note that the figure shows the averaged daily predictions for the entire test period. To compare the models against the baseline, the Mean Absolute Error (MAE) of the predictions was evaluated over the entire test period. MAE is straightforward to calculate and interpret, by summing the magnitudes (absolute values) of the errors to obtain the 'total error' and then dividing by the number of observations [38]. This metric reflects model accuracy, representing the difference between predicted and actual passengers every 15 minutes. The MAE is particularly fitting to communicate to the stakeholders in operations (who are non-experts) how well the models perform and the practical implications therein [38].

These results indicate that all three proposed models outperform the baseline, and that the neural network is the best performing one. It has a mean absolute error of 15.81, suggesting that it predicts around 16 passengers too many or too few per each 15 minutes period, throughout a day



(a) Flight-level.



(b) Day-level.

Fig. 13. Predictions on test data. (a) flight-level aggregations and (b) day-level aggregations (actual values omitted for confidentiality).

of operation. Considering that the baseline error was 25 passengers every 15 minutes, this means a relative error reduction of 40%. Considering around 8000 passengers/hour (2000 every 15 minutes), this represents a reduction of the absolute error from .12 to .7 of the total number of passengers (actual numbers are omitted for confidentiality).

Table 3. Overview of the results of all models (best results in bold).

Model	MSE	KS
Simple linear regression	19.31	0.45
Extended linear regression	16.48	0.33
Neural network	**11.08**	**0.34**

(a) Mean square error (in number of passengers) on a flight level basis.

Model	MAE
Baseline	24.36
Simple linear regression	22.02
Extended linear regression	19.19
Neural network	**15.81**

(b) Mean absolute error (in number of passengers) on a daily level basis.

6 DISCUSSION

This section first recaps the research questions introduced in Section 2, discussing what we have been able to resolve with our CPS-enabled, DataOps-driven methodology throughout the paper; subsequently, we outline the key limitations of our methodology.

RQ1. How to *model* the problem domain as an exemplary instance of CPS and Big Data? Section 3 elaborated on the airport setting as a complex Cyber-Physical System. This allows to capture the complex interactions between multiple actors such as managers, human resource planners, shiftl-leaders and desk agents, both from the airlines and airport operations, and passengers themselves; but also data sources ranging from structured and static, to unstructured and streaming in nature. These operate and interact with intertwined computational and physical artifacts, including mobile phones/apps, monitors, sensor networks, self check-in devices and biometric e-gates and office automation systems.

We then focused particularly on the passengers flow problem, as it is a recurring and representative problem in CPSs such as airport systems. We inferred the existing end-to-end planning process stemming from interviews and observations by means of the application of Grounded Theory in conjunction with a Data Science Research Methodology, which led to a thorough understanding of the underlying domain. From a DataOps perspective, we also identified relevant data parameters and their origin, used during the legacy planning process, such as destination and departure time at flight level and baggage and check-in time at passenger level. We identified that the legacy processes still in place do not exploit contemporary advancements in the field, and neglect large volumes of sensor- and device- generated multivariate data. As a result, airports, airlines and passengers miss on the potential that advanced techniques can bring, generating extra costs for airlines and airport operators, reducing customer satisfaction, and increasing risks for security, safety and health of passengers and staff. As a concrete contribution to the existing planning process, we brought novel data sources and parameters, captured from mobile devices and sensors – e.g., location data stemming from mobile applications – and connected those to a data pipeline and data lake end-points, for further use in predictions.

RQ2: How to *predict* passengers flow more accurately? We designed and implemented an end-to-end data pipeline for departing passenger predictions, as presented in Section 4. This pipeline provides effective support from data sources through data preparation, cu ration, transformation and fusion, to prediction, real-time analytics and visualization. In addition, we have explored and calibrated several techniques with regards to predictive analytics that outperform the classical approaches for modelling arrival behavior of passengers.

In particular, we conducted a comparative analysis between three different prediction models, ranging from simple/extended linear regression to neural networks. We optimized the accuracy of the neural networks experimenting with up to 30 different shapes and configurations. To effectively evaluate their performance, we compared them against existing day-level predictions as a baseline. From the results presented in Section 5, the performance of the neural network approach is superior to all other models, both on flight and day level, reducing error up to 60%, from 25 to 15 passengers every 15 minutes interval. The prediction obtained in the amount of passengers is better than the baseline at all relevant query times. The closer one gets to the moment of departure, the better the model performs, thanks to the multiple data sources considered. This allows to plan airport resources more efficiently, while improving customer experience by avoiding passengers clumping at check-in and security.

However, there is certainly a trade-off in predictive models between explainability and performance, where complex models such as neural networks usually outperform more interpretable models such as linear regressions, as demonstrated by results in Section 5. Neural networks are architectures featuring several hidden layers. As these become more complex and introduce increased nonlinearity, their structure become less transparent and it is harder to understand what operations or input information lead to specific decisions [29]. However, it is critical to understand, trust, and *explain* the rationale behind decisions. To support this, it is possible to increase model interpretability for former "black-box" models – as addressed in the following.

RQ3: How to *explain* the prediction results in terms of the considered features?
Explaining a model means providing extra qualitative information about why the model reached to a specific decision regarding the input features [29]. In an effort to explain black-box deep-learning models, Lundberg and Lee proposed SHapley Additive exPlanations (SHAP [18]). The goal of SHAP is to explain predictions in terms of the contribution of each feature. Shapley values tell how to fairly distribute prediction among features, helping to interpret how a neural network arrived to a certain prediction. The total set of features we analyze counts over ˜150 quantities and reflects multi-dimensional sensors data arranged around the cyber-physical system including, among others: (a) traffic sensors across the hall, (b) sensors in the passengers' mobile device, (c) sensors mounted on the entrance door, (d) online databases (e.g., for timestamps, weather, day of the week, etc.). The feature importance illustrated in Figure 14 shows the top 14 features with statistical significance. Features reported in the figure reflect:

(1) **Left-Bound-Drop-off.** These are passengers who have already checked in electronically and required baggage drop-off only; they typically transit in the hall for a short amount of time and are normally directly bound to the drop-off kiosk.
(2) **Flight Days Booked in Advance.** This is the time delta between the day of the flight and the day a passenger booked the ticket for that flight.
(3) **Scheduled Time of Departure (STD).** This reflects the time of the day for the flight departure, which may affect the passengers arrival – e.g., for early flights it tends to be closer to departure.
(4) **Time App Difference.** This reflects the mean time delta between the STD and the moment in which a passenger entered the departure hall, captured through the mobile app GPS data.
(5) **Total Economy-Class Bookings.** This quantity identifies how many passengers are currently booked to travel economy-class and are circulating in the hallway.
(6) **Weekday.** This is an indicator (between 1-7) that reflects the day of the week, set to start at 1 == "Monday".
(7) **Days of Ahead-Booking.** This quantity identifies the mean number of days of ahead-booking (i.e., how many days before the flight was the flight booked) for passengers currently across the hall.
(8) **Flight Length of Stay.** This quantity identifies the mean of how many days are passengers bound to stay in the location they are travelling to.
(9) **Flight PAX in Booking.** This quantity identifies how many people are bundled together in the same group and flying together.
(10) **Flight App Difference.** This reflects the mean time delta between the STD and the moment in which a passenger entered the departure hall, captured through the mobile app GPS data, averaged over the destination of the flight.
(11) **Time Length of Stay.** This quantity identifies the mean of how many days are passengers bound to stay in the location they are travelling to, averaged over the 15-minute time bracket prior to departure.
(12) **Month.** This identifies the month of the year in the current datapoint, retrieved according to the world clock.

(13) **Time PAX in Booking.** This quantity identifies how many people are bundled together in the same booking, averaged over the 15-minute time bracket prior to departure.
(14) **Total Business Class Bookings.** This quantity reflects the total amount of business class ticketed passengers currently roaming the hall.

The left bound drop-off of the time interval is the most important feature, affecting the predicted number of passengers by 7% on average. Also, the mean days booked in advance and scheduled time of departure of a flight turn out to be important features in this model. In addition, locations data is statistically significant, changing the predicted value by $\tilde{3}$%. The feature importance provides a first direction in explaining the model. For more informative insights, Figure 15, combines feature importance with positive/negative feature effects. Each point on the summary plot is a Shapley value for a feature and an instance. The features are ordered according to their statistical significance. Observations on the left of the vertical line indicate a negative effect on the target variable, whereas observations on the right indicate a positive effect. We note a negative effect for most observations in the number of days booked in advance, and a mostly positive effect for the locations data from the mobile app. One should note, however, that all effects describe the model behavior and are not necessarily causal.

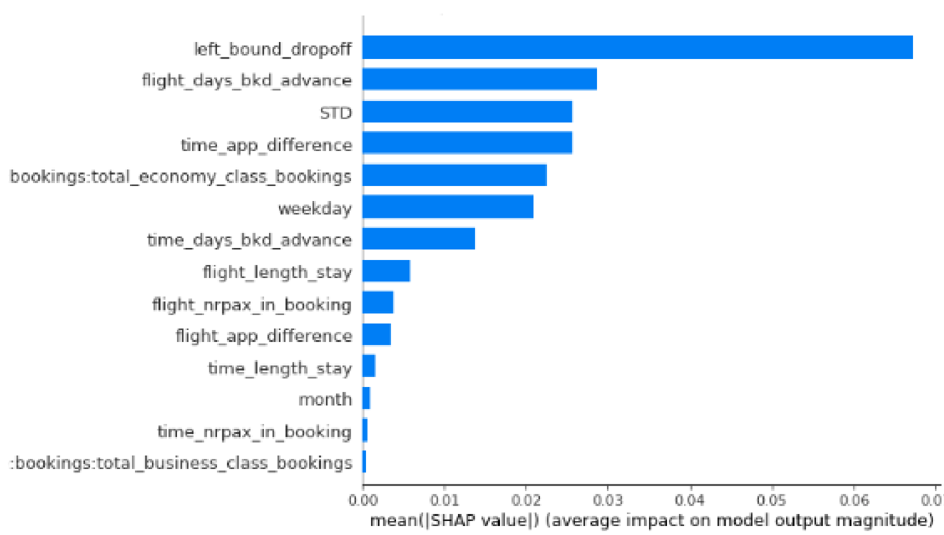


Fig. 14. SHAP feature importance measured as mean absolute Shapley values.

6.1 Limitations and Future Work

Here we report the main limitations in our approach and how these will be addressed in future work. The CPS-enabled prediction model grounded on the DataOps methodology as presented, can be further optimized to increase accuracy. For example, by including other data features to assess and predict flight delays such as *weather conditions*, traceable from API sources of weather information available online. Although not considered in this work weather as a delay component is partially accounted in the historical data, within seasonal and monthly average delays. An explicit treatment of weather conditions is a matter of future work. From a technical standpoint, tuning the hyper-parameters of the neural network and experimenting with the architecture of the network,

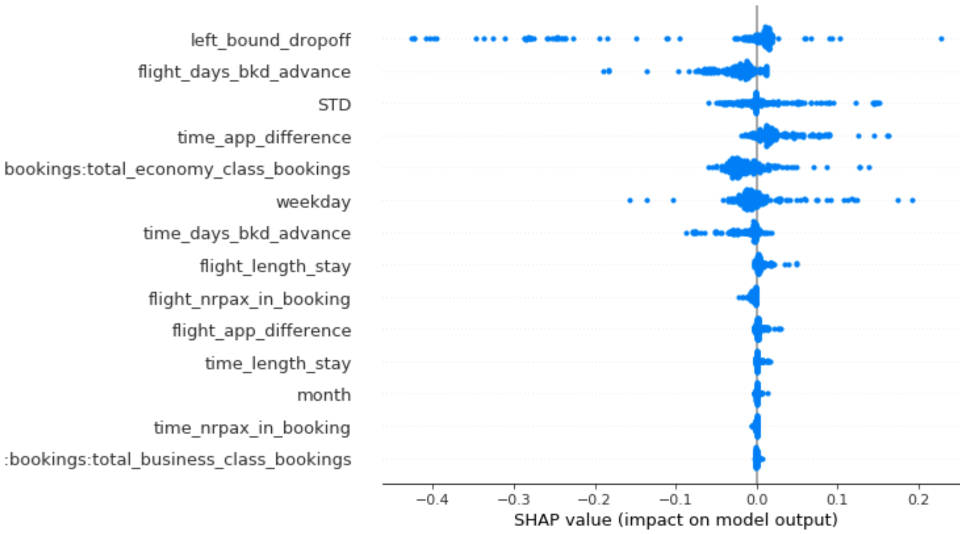


Fig. 15. SHAP summary plot with positive and negative influence per feature.

such as the number of layers and the number of neurons per layer. Also, directly predicting the distribution instead of 15-minute brackets may improve the current model.

Another possible improvement towards adoption of the model is to fit it for the specific use case of the departure hall. Specifically, the target variable can be split into passengers that have a bag drop-off at the airport and passengers that need to check-in at the airport. With this additional split, the agents can be directly scheduled at the desks in which they are needed most. In addition, the model can be extended by including data on partner airlines which share the same ground services facilities. Needless to say, this implies a tradeoff with the generalizability of the approach that should be carefully considered.

The quality of data used for prediction can be further analyzed and improved, e.g., by semantically enriching the data. For example, the arrival profile presented in Figure 9 suggests a distribution with two peaks, in which the second (smaller) peak is not yet addressed. The assumption made in the scope of this paper is that this group exists – in the sense that a small portion of the passengers arrives rather close to the flight departure. It can be, however, that this represents a group of passengers whose flight was delayed. In addition, the amount and diversity of data at hand could be further extended to improve validity, e.g., by including open data or social media data. Our experimental data represented one year of training data and four months of test data. This should be sufficient for capturing the different types of seasonality, but a longer period (a longitudinal study) would build more confidence in the results.

The consideration of the airport as a CPS for governance – a central tenet of the present paper – although not immediately obvious to the evaluation goals has an indirect effect to the metrics and features employed. The overall setting consists of decision-makers (i.e., stakeholders or airline operators), users (i.e., passengers) along with user-facing devices (i.e., sensors, mobile/web applications). Additionally, its computational aspect includes overall management, prediction and analytics functionalities, which constitute monitoring and control activities. Recall that (i) the processes identified through requirements elicitation upon the business domain and (ii) data sources within the system are consolidated in the passenger flow chart illustrated in Fig. 5. The departure hall – within the physical realm – is central to passenger flow. As such, flow within the hall distills available knowledge by making explicit the interactions of passengers with various system endpoints, which are essentially the interfaces of the CPS between the physical and the computational world. Furthermore, beyond the aforementioned purely technical aspects, it shall be remarked that the governance dimension in our endeavour focused primarily on end-user governance goals (e.g., increasing customer service times and levels, etc.) but several stringent requirements (especially so in the air travel service segment) remain explicitly unaddressed in our formalisation and approach.

More specifically, the aforementioned service segment typically requires very stringent privacy and security aspects to be addressed by any deployable solution—either featuring AI software or otherwise—which meets and processes end-user or observable circumstantial data. Although in the scope of this paper—and the deployed AI-software we prototyped as part of this study—we leave such aspects to be addressed by related work either in cybersecurity for CPS (e.g., see the overviews in [2; 28]) or *privacy-by-design* for data-intensive computing (e.g., see [9; 31; 32]) We highlight that both security and privacy issues and challenges present in the case-study were deliberately not addressed in the modelling exercise since they could indeed compromise its applicability altogether. Consider, for example, even the simple impossibility of using open customer data without explicit consent might make any predictive model inapplicable for passenger hall traffic prediction. In the future, research should be applied to automated and anti-tampering anonymization and de-anonymization middleware able to refine data-intensive pipelines such as the one we proposed and evaluated in this paper so as to address the aforementioned security and privacy requirements.

Finally, we acknowledge that the experiment is limited to data from one airport and airline, considering that regulations, passenger characteristics, airport environments or boarding procedures may differ among countries and thus results may not be directly generalizable. Nevertheless, these can be easily extrapolated for any international airport, and also adjusted to fit other scenarios around the world.

7 RELATED WORK

DataOps entails a set of practices, processes and technologies that combines an integrated and process-oriented perspective on data with automation and methods from agile software engineering to improve quality, speed, and collaboration and promote a culture of continuous improvement [7]. A DataOps methodology combines and interconnects data engineering, data integration, data quality, and data security/privacy [20] to deliver data from its source to the person, system, or application that can turn it into business value [8].

Modeling airports and their data flows as Cyber-Physical Systems has been present in the literature to address safety and performance of overall air transportation systems under increased throughput. For example, conflict detection and resolution in surface traffic, by means of information flow analysis [27]. Tight cyber–physical integration, within aircraft and between aircraft and off-board systems, warrant a surgical consideration of cyber–physical interactions and potential performance risks from cyber and physical threats, calling for a foundational understanding of properties, behavior, and performance of aviation cyber–physical integration [26]. Rich integration leads to improved quality and performance of gate-to-gate flight of each individual aircraft and travel experience of each individual passenger and crew. Undoubtedly, Airports and airlines can benefit from the insights and solutions of modeling key processes as CPS.

Recent literature models the passenger flow at airports using various data, such as Wi-Fi indoor location [11; 16], security checkpoints information [19], and/or simulations [5; 19]. Our approach leverages more data sources and features (e.g., mobile app data), and continuous predictive analytics, offering a dynamic model that can be fine-tuned over time. Additionally, some approaches use applied machine learning to predict passenger flow [10], e.g., by means of deep learning [16; 17], Long-Short Term Memory networks [19], and Bayesian networks [11]. Similarly, our approach accounts a high number and variety of features to solve a related problem. Moreover, we also model the airport and passenger flow as a CPS, and the construction and consolidation of the overall data pipeline, gaining valuable insights in the process.

8 CONCLUSIONS

This work introduces a CPS-enabled analytics methodology and data pipeline for bringing continuous predictive analytics to airlines and airports operations. The first part of the methodology is based on Design Science Research in tandem with Grounded Theory to understand and conceptualize the business context and domain as a Cyber-Physical System. In particular, for our scenario we interviewed several key stakeholders and observed their daily tasks to boil down the key processes and interactions in the passengers flow and resource allocation at airports. The resulting CPS is complex, unpredictable and dynamic as it features several actors and intertwined computation and physical aspects, while exhibiting emergent behavior. As such, our approach targets the critical system requirements phase and illustrates how to reason on requirements and data sources in the early stages of design. Additionally, such high-level reasoning can aid in analysis of system behavior after deployment, used for bootstrapping corrective actions.

We constructed a data pipeline, following the DataOps approach to deliver business value [36]. In particular, we exemplified how to capture, extract and combine data from at least four different sources, including three pre-existing ones – flight level data, passenger level data, and customer satisfaction data – as well as a novel data source stemming from a mobile app. Data governance is in place to enact data access controls in support of a "data marketplace". Then, data scientists can build and publish prediction models, often through experimentation and iteration. Our data pipeline consolidates the information for analysis using novel predictive models, capable of performing predictions at a flight-level basis, instead of existing day-level basis ones. Furthermore, DevOps teams build end-to-end pipelines that include and deploy models in the application logic. We illustrated this by means of a prototype of the CPS-enabled data pipeline and data lake, and empirically explored three predictive models: simple and extended linear regression, and neural networks, with up to 23 input features and 30 different configurations and query times. We compared our models against the existing baseline, showing substantial improvements. The best performing model is a neural network, with up to 60% reduction of the error in terms of amount of passengers predicted for every 15 minutes slot. Moreover, we explain the results of the neural network in terms of the most important features by means of SHAP, showing high importance of the novel features stemming from the mobile app.

As practical implications, having more accurate predictions provides business value at different, critical points. In the resource planning axis, this allows to plan airport resources more efficiently. Better insights in passenger peaks allow to employ enough agents, reducing waiting times and increasing flow throughout the airport. Second, better insight in the troughs allows to employ less agents during those moments. This obviously aids reducing costs. From the Grounded Theory, a better planning could result in a 10% reduction in temporary workforce, resulting in a yearly saving of up to 1 million euros (actual number not revealed for confidentiality) only for one airline and airport. Finally and in a general context, this proposal improves passengers experience, safety and health. Avoiding passengers' clumping at check-in desks and security checks reduces close contacts and helps social distancing, something vital for overall safety and public health especially under new regulations for post-pandemic travel.

ACKNOWLEDGEMENTS

We thank the interviewees from the airline and airport operations for their helpful insights. We also thank the anonymous reviewers for their comments and constructive feedback on the paper. Research partially supported by FWF Austria M 2778-N (EDENSPACE), the project ANPCyT PICT-2017-1725, the European Commission grants no. 0421 (Interreg ICT), Werkinzicht, no. 787061 (ANITA) no. 825040 (RADON), and no. 825480 (SODALITE).

REFERENCES

[1] Airbus. 2018. Global Market Forecast. https://www.airbus.com/content/dam/corporate-topics/publications/media-day/GMF-2018-2037.pdf

[2] Saqib Ali, Taiseera Al-Balushi, Zia Nadir, and Omar Khadeer Hussain. 2018. *Cyber Security for Cyber Physical Systems*. Studies in Computational Intelligence, Vol. 768. Springer.

[3] L. Baresi, D. F. Mendonça, M. Garriga, S. Guinea, and G. Quattrocchi. 2019. A Unified Model for the Mobile-Edge-Cloud Continuum. *ACM Trans. Internet Technol.* 19, 2, Article Article 29 (April 2019), 21 pages. https://doi.org/10.1145/3226644

[4] Kathy Charmaz and Linda Liska Belgrave. 2007. Grounded theory. *The Blackwell encyclopedia of sociology* (2007).

[5] Duilio Curcio, Francesco Longo, Giovanni Mirabelli, and Enrico Pappoff. 2007. Passengers' flow analysis and security issues in airport terminals using modeling & simulation. In *European Conference on Modeling & Simulation, Praga-Repubblica Ceca*. 4–6.

[6] Konstantin Eckle and Johannes Schmidt-Hieber. 2019. A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Networks* 110 (2019), 232 – 242. https://doi.org/10.1016/j.neunet.2018.11.005

[7] Julian Ereth. 2018. DataOps-Towards a Definition. (2018), 104–112.

[8] Jarah Euston. 2018. What is DataOps?: Platform for the Machine Learning Age. https://www.nexla.com/define-dataops

[9] Michele Guerriero, Damian Andrew Tamburri, and Elisabetta Di Nitto. 2018. Defining, enforcing and checking privacy policies in data-intensive applications.. In *SEAMS@ICSE*, Jesper Andersson and Danny Weyns (Eds.). ACM, 172–182. http://dblp.uni-trier.de/db/conf/icse/seams2018.html#GuerrieroTN18

[10] Xiaojia Guo, Yael Grushka-Cockayne, and Bert De Reyck. 2019. Forecasting Airport Transfer Passenger Flow Using Real-Time Data and Machine Learning. *Available at SSRN 3245609* (2019).

[11] Weixin Huang, Yuming Lin, Borong Lin, and Liang Zhao. 2019. Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data. *Energy and Buildings* 203 (2019), 109439. https://doi.org/10.1016/j.enbuild.2019.109439

[12] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[14] Jay Lee, Behrad Bagheri, and Hung-An Kao. 2015. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing letters* 3 (2015), 18–23.

[15] Jay Lee, Edzel Lapira, Behrad Bagheri, and Hung-an Kao. 2013. Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing letters* 1, 1 (2013), 38–41.

[16] Lijuan Liu and Rung-Ching Chen. 2017. A novel passenger flow prediction model using deep learning methods. *Transportation Research Part C: Emerging Technologies* 84 (2017), 74 – 91. https://doi.org/10.1016/j.trc.2017.08.001

[17] Xiaochen Liu, Lingshan Li, Xiaohua Liu, Tao Zhang, Xiangyang Rong, Ling Yang, and Dizhan Xiong. 2018. Field investigation on characteristics of passenger flow in a Chinese hub airport terminal. *Building and Environment* 133 (2018), 51 – 61. https://doi.org/10.1016/j.buildenv.2018.02.009

[18] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*. 4765–4774.

[19] P. Monmousseau, G. Jarry, F. Bertosio, D. Delahaye, and M. Houalla. 2020. Predicting Passenger Flow at Charles De Gaulle Airport Security Checkpoints. In *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)*. 1–9.

[20] A Palmer. 2015. From DevOps to DataOps. https://www.tamr.com/blog/from-devops-to-dataops-by-andy-palmer/.

[21] Ken Peffers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. 2007. A design science research methodology for information systems research. *Journal of management information systems* 24, 3 (2007), 45–77.

[22] Ragunathan Raj Rajkumar, Insup Lee, Lui Sha, and John Stankovic. 2010. Cyber-physical systems: the next computing revolution. In *Proc. of the 47th Design Automation Conference*. ACM, 731–736.

[23] Nornadiah Mohd Razali, Yap Bee Wah, et al. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics* 2, 1 (2011), 21–33.

[24] Frederick F Reichheld. 2003. The one number you need to grow. *Harvard business review* 81, 12 (2003), 46–55. https://hbr.org/2003/12/the-one-number-you-need-to-grow.
[25] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 1135–1144.
[26] Krishna Sampigethaya and Radha Poovendran. 2013. Aviation cyber–physical systems: Foundations for future aircraft and air transport. *Proc. IEEE* 101, 8 (2013), 1834–1855.
[27] Frank Saunders, Jason Rife, Sai Vaddi, and Victor Cheng. 2013. Information flow diagram analysis of a model cyber-physical system: Conflict detection and resolution for airport surface traffic. *IEEE Aerospace and Electronic Systems Magazine* 28, 12 (2013), 26–35.
[28] Dimitrios N. Serpanos, Howard E. Shrobe, and Muhammad Taimoor Khan. 2016. Behaviour-Based Security for Cyber-Physical Systems. *ERCIM News* 2016, 107 (2016). http://dblp.uni-trier.de/db/journals/ercim/ercim2016.html#SerpanosSK16
[29] Atefeh Shahroudnejad, Parnian Afshar, Konstantinos N Plataniotis, and Arash Mohammadi. 2018. Improved explainability of capsule networks: Relevance path by agreement. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 549–553.
[30] Galit Shmueli et al. 2010. To explain or to predict? *Statistical science* 25, 3 (2010), 289–310.
[31] Damian A. Tamburri. 2020. Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation. *Inf. Syst.* 91 (2020), 101469. http://dblp.uni-trier.de/db/journals/is/is91.html#Tamburri20
[32] Christos Tsigkanos, Cosmin Avasalcai, and Schahram Dustdar. 2019. Architectural Considerations for Privacy on the Edge. *IEEE Internet Comput.* 23, 4 (2019), 76–83.
[33] Christos Tsigkanos, Timo Kehrer, and Carlo Ghezzi. 2016. Architecting dynamic cyber-physical spaces. *Computing* 98, 10 (2016), 1011–1040.
[34] Christos Tsigkanos, Timo Kehrer, and Carlo Ghezzi. 2017. Modeling and verification of evolving cyber-physical spaces. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2017, 2017*. 38–48.
[35] C. Tsigkanos, L. Nenzi, M. Loreti, M. Garriga, S. Dustdar, and C. Ghezzi. 2019. Inferring Analyzable Models from Trajectories of Spatially-Distributed Internet of Things. In *2019 IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*. 100–106. https://doi.org/10.1109/SEAMS.2019.00021
[36] Crystal Valentine and William Merchan. 2018. DataOps: An Agile Methodology for Data-Driven Organizations. Oracle whitepaper. https://www.oracle.com/a/ocom/docs/oracle-ds-data-ops-map-r.pdf.
[37] Sanford Weisberg. 2005. *Applied linear regression*. Vol. 528. John Wiley & Sons.
[38] Cort J Willmott and Kenji Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research* 30, 1 (2005), 79–82.
[39] Rüdiger Wirth and Jochen Hipp. 2000. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. Springer-Verlag London, UK, 29–39.
[40] Bayya Yegnanarayana. 2009. *Artificial neural networks*. PHI Learning Pvt. Ltd.
[41] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. 2016. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530* (2016).
[42] Quanshi Zhang, Ying Nian Wu, and Song-Chun Zhu. 2018. Interpretable convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8827–8836.