Modelling of punctuality at Frankfurt Airport

Von der Fakultät für Mathematik und Physik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktor der Naturwissenschaften
Dr. rer. nat.
genehmigte Dissertation
von
Dipl.-Met. Peer Röhner
geboren am 13.05.1977 in Eisenach

2009

Referent: Prof. Dr. Thomas Hauf
Korreferent: Prof. Dr. Dieter Etling
Tag der Promotion: 4.12.2009

Abstract

The objective of this study is to develop a punctuality model for both diagnosis and prognosis. This necessarily implies a better understanding of the weather impact on air traffic delays and punctualities and a classification of significant actuating variables.

Taking Frankfurt Airport as study airport, it is exemplarily shown, how much of the variability of daily punctuality can be explained through a mathematical model. For this purpose, a hybrid model, based on multivariate linear regression, was developed via several expansion stages. It additionally comprises an autoregressive term. A complex regression tree correction algorithm cares for model adjustment in the low punctuality domain.

The model introduced in this work is a continuation of the work previously done by Spehr (2003). Suggestions for improvement were seized and developed further along with new ideas. In that respect it is analysed, which model enhancements are meaningful and how they are reflected in improvements of model quality. It is shown that by means of enhanced punctuality models, an R^2 exceeding 0.6 can be realised. This is a significant improvement compared to model results obtained by Spehr, where days with e.g. strikes or system failures were excluded from the analyses beforehand. Moreover, the weather impact on punctuality is, in the present study, quantified using a new method. Previous approaches often drew upon an interpretation of delay codes. By means of the punctuality models at hand, more than 45 % of the variability in daily punctuality can be explained through local weather. Thus, the hypothesised strong weather impact on air traffic delays is to a high degree verified. In that respect, weather acts as a governing descriptor regarding the generation and development of delays and, consequently, not all single processes need to be modelled individually.

In a follow-up study, the developed punctuality models were analysed focusing on their potential for a punctuality forecast. The option of true punctuality forecasting on a daily basis for one or several days into the future, with the aid of weather- and traffic forecasts, is a valuable source of information for airport divisions planning in the medium term. It is shown that good results can be achieved using punctuality models based on predictable input variables, only. By means of independent data, an R^2 of almost 0.6 was realised. Based on these results, the operational application of a punctuality forecast model seems not only possible but also reasonable. Therefore, an optimisation of the forecast of significant predictor variables should be aimed for in follow-up studies.

Keywords: punctuality, air traffic delay, modelling

Zusammenfassung

Ziel dieser Arbeit ist die Entwicklung eines Pünktlichkeitsmodells sowohl für die Diagnose als auch die Vorhersage. Dies impliziert ein besseres Verständnis der Größe des Wettereinflusses auf die Verspätungen und die Pünktlichkeit im Flugverkehr, sowie eine Benennung und Einordnung der signifikanten Einflussgrößen.

Anhand des Flughafens Frankfurt wird exemplarisch gezeigt, welcher Variabilitätsanteil der Tagespünktlichkeit mit einem mathematischen Modell erklärt werden kann. Dazu wurde in mehreren Ausbaustufen ein hybrides Modell auf Basis von multivariater linearer Regression entwickelt. Zur zusätzlichen Verbesserung wurden Zeitreiheninformationen ausgewertet, die in einem autoregressiven Term im Pünktlichkeitsmodell Anwendung finden. Ein komplexer Regressionsbaum-Korrekturalgorithmus sorgt für Modellanpassungen im Niedrigpünktlichkeitsbereich.

Die in dieser Arbeit vorgestellten Modelle sind eine Fortführung der Arbeit von Spehr (2003). Verbesserungsvorschläge werden aufgegriffen und mit neuen Ideen weiterentwickelt. In diesem Zusammenhang wird untersucht, welche Modellerweiterungen sinnvoll sind und wie sich diese in einer Verbesserung der Modellgüte niederschlagen. Es wird gezeigt, dass mit Hilfe erweiterter Pünktlichkeitsmodelle ein R^2 von mehr als 0.6 erreicht werden kann. Dies ist eine deutliche Verbesserung gegenüber den Modellergebnissen von Spehr, bei denen z.B. durch Streik oder Systemausfälle vorbelastete Tage für die Untersuchungen ausgeschlossen wurden. Darüber hinaus wird in der vorliegenden Arbeit der Wettereinfluss auf die Pünktlichkeit auf eine neue Art und Weise quantifiziert. Bisherige Ansätze stützten sich in der Regel auf die Auswertung von Delay-Codes. Mit Hilfe der vorliegenden Pünktlichkeitsmodelle kann mehr als 45 % der Variabilität der Tagespünktlichkeit mit lokalem Wetter erklärt werden. Damit wird der angenommene starke Wettereinfluss auf Verspätungen im Flugverkehr bestätigt. Wetter zeigt sich in diesem Zusammenhang bei der Generierung und Entwicklung von Verspätungen als übergeordneter Deskriptor, wodurch auf eine individuelle Modellierung von Einzelprozessen verzichtet werden kann.

In einer sich anschließenden Analyse werden die entwickelten Pünktlichkeitsmodelle auf ihr Potential zur Pünktlichkeitsvorhersage hin untersucht. Die Möglichkeit einer echten Pünktlichkeitsvorhersage auf Tagesbasis für einen oder gar mehrere Tage in der Zukunft, unter Zuhilfenahme von Wetter- und Verkehrsprognosen, ist eine wertvolle Option für mittelfristig planende Flughafenabteilungen. Es wird gezeigt, dass Pünktlichkeitsmodelle auf Basis prognostizierbarer Prädiktoren gute Resultate erzielen. Anhand unabhängiger Daten wurde ein R^2 von nahezu 0.6 erreicht. Auf Grundlage dieser Ergebnisse scheint der operative Einsatz eines Pünktlichkeitsvorhersagemodells nicht nur möglich sondern auch sinnvoll. In Anschlussstudien sollte daher die Optimierung der Vorhersage signifikanter Prädiktoren anvisiert werden.

Schlagworte: Pünktlichkeit, Flugverspätungen, Modellierung

Contents

List of Figures

List of Tables

Abbreviations

ACARS	Aircraft Communications Addressing and Reporting System
AMDAR	Aircraft Meteorological Data Relay
AR	Auto Regressive
ATC	Air Traffic Control
ATFM	Air Traffic Flow Management
ATFCM	Air Traffic Flow and Capacity Management
CAT	Approach Category
CFMU	Central Flow Management Unit
CODA	Central Office for Delay Analysis
CPS	Capacity Prognosis Schiphol
DH	Decision Height
DFS	Deutsche Flugsicherung
DLH	Deutsche Lufthansa
DWD	Deutscher Wetterdienst
EDDF	Frankfurt Airport (ICAO code)
FAA	Federal Aviation Administration
FAR	False Alarm Ratio
FFF	Future for FRA
FL	Flight Level
FRA	Frankfurt Airport (IATA code)
FRAPORT	Frankfurt Airport Company
GDP	Ground Delay Program
IATA	International Air Transport Association
ICAO	International Civil Aviation Organisation
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
INBFL	Inbound Flight Movements
INBP	Inbound Punctuality
KLM	Koninklijke Luchtvaart Maatschappij (Royal Dutch Airlines)
KNMI	Koninklijk Nederlands Meteorologisch Instituut (Royal Netherlands Meteorological Institute)
MIT	Massachusetts Institute of Technology

MLR	Multivariate Linear Regression
MLRCL	Multivariate Linear Regression Correction Limit
MOS	Model Output Statistics
NAS	National Airspace System
NWFM	Numerical Weather Forecasting Model
NWP	Numerical Weather Prediction
OLS	Ordinary Least Squares
OTBFL	Outbound Flight Movements
OTBP	Outbound Punctuality
POD	Probability Of Detection
PRS	Parallel RWY System
RTCL	Regression Tree Correction Limit
RVR	Runway Visual Range
RWY	Runway
SAS	Scandinavian Airlines System
THR	Total Hit Rate
TMA	Terminal Manoeuvring Area
TOTFL	Total Flight Movements
TOTP	Total Punctuality
VHF	Very High Frequency
VMC	Visual Meteorological Conditions
WITI	Weather Impacted Traffic Index
WMO	World Meteorological Organization

Chapter 1

Introduction

> *"Air transport delays in Europe are a major concern for the industry and a relentless source of complaints from the passengers, as often verified in the media. Not only is it a painful inconvenience for the actors, but delays also induce large costs, for the airlines, their customers and the community as a whole."*

The above citation taken from ITA (2000) concisely describes the significance of air transport delays as a key indicator for air traffic performance. In that respect, delay, and strongly connected to it, punctuality as one of its measures, is just one among many performance parameters for air transportation. In the same breath, other factors such as capacity, safety, security, cost-effectiveness, environmental sustainability, flexibility, predictability, access and equity, participation and interoperability have to be considered (EUROCONTROL, 2008).

Certainly, there is no other industry that is more sensitive to weather than the aeronautical industry. *"In spite of our improved ability to observe and forecast the weather to a greater degree of accuracy than ever before, adverse meteorological conditions continue to severely impact the operational safety and efficiency, as well as the system's capacity."* This citation from Sprinkle and Macleod (1991) – though published almost 20 years ago – still reflects the current status of weather within the aviation industry. Weather was and still is one of the root causes of schedule disruptions and air traffic delays. However, delay and, consequently, a decrease in punctuality does not only arise from weather. Other factors are e.g. high volume of traffic, insufficient adoption of aviation infrastructure to enhanced safety and security standards or just non-optimised ground operations. In a tightly integrated air traffic system, these primary delays are multiplied and spread in time and space. Exemplarily for 2008, Figure 1.1 gives an overview over the on-time performance of the most important European airports. Displayed is the fraction of flights being more than 15 minutes late (upper part of the graph) or early (lower part of the graph). Looking at the upper chart, the

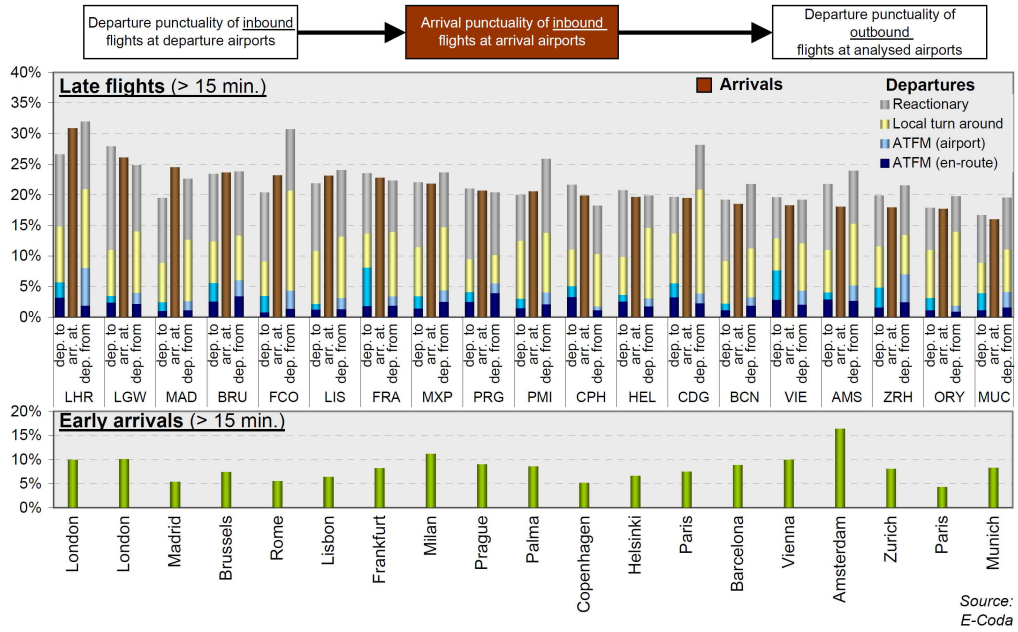


Figure 1.1: On-time performance at European airports in 2008 (EUROCONTROL, 2009b).

left bars show the fraction of late off-blocks (using the 15 minutes threshold) at the various airports of origin by traffic inbound to the respective airport of interest. The centre bars give the fraction of late arrivals at the airport of interest, the right bars the fraction of late departures. Additionally, a simple breakdown into four delay causes is given for departures: reactionary delay, delay produced at local turnaround, ATFM delay produced at the airport and ATFM delay produced en-route. The factor "weather" is not directly covered in the this course classification but it is hidden in the four other factors. Apparently, the on-time performance as well as the impact of imported delays and the translation of arrival into departure delays varies among the displayed airports. The reasons for this are manifold and further discussed in this work. Figure 1.2 shows the development of the average on-time performance in Europe between 2002 and 2008 as defined above, subdivided into arrivals and departures being more than 15 minutes late and arrivals being more than 15 minutes early. Whereas early arrivals are at a constant level of roughly 7 %, the fraction of late departures and arrivals is subject to variation over time at a level of roughly 20 %.

For Frankfurt Airport, the airport of investigation within this study, the distribution of delay causes for 2008 was such that, according to DFS (2008), 47 % of all departure delays were attributed to airline internal reasons, followed by 20 % for airport reasons, 11 % both for ATM and security, respectively, and weather with 10 %. At first glance, the relative importance of weather seems rather low. It is, however, well known that airports operat-

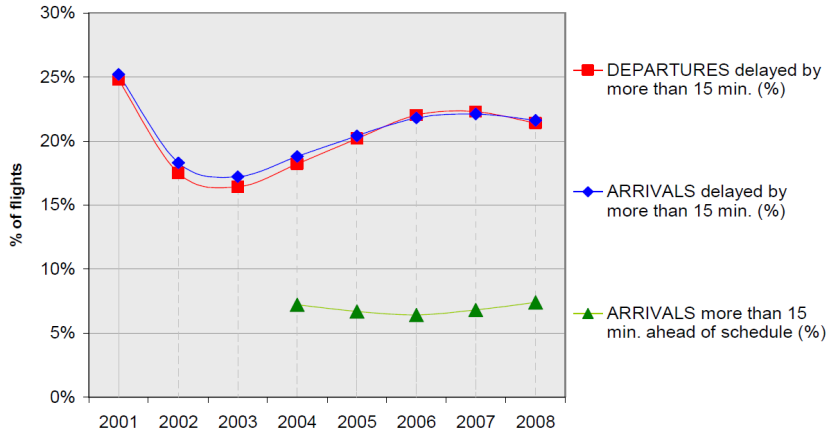


Figure 1.2: On-time performance in Europe between 2001 and 2008 (EUROCONTROL, 2009b).

ing close to their nominal capacity experience an amplifying effect on delays trough adverse weather conditions at times of high work load – without weather being officially allocated as the primary cause for the delays produced. One of the objectives of the present study hence is to take a closer look at the weather impact on delay and punctuality.

Another major issue of delay is the corresponding costs. Air traffic delays induce financial and economic consequences on airlines, on their clients and on the community (ITA, 2000). Airlines suffer from additional costs on fleet and are obliged to compensate passengers for their discomfort. SAS, for example, estimates airline costs associated with one minute of delay to range between $150 and $300 (Wu and Caves, 2003b). A better understanding of delay – the governing processes and the complex interrelationships leading to delay propagation and multiplication – is thus likely to generate significant economic benefits through optimised, pro-active operations and adaptations, as well as adequate and timely reactions.

1.1 Definition of Delay Measures

There are many different performance indicators used to compare airlines' or airports' performance regarding air traffic delays. The simplest is just the recording of delay minutes produced within in a certain time frame with regard to a reference, be it published schedules, estimated on-/off block times or just average flight/movement times in a certain sector respectively from reference point A to reference point B, or for a certain procedure.

An internationally accepted performance indicator for the operational performance of airlines and airports is *punctuality* (EUROCONTROL, 2005), quantifying delays as compared to published schedules. Punctual-

ity, is defined as the fraction of punctual flights to the total number of flights within a certain time frame. In that regard, a flight is punctual if its delay compared to schedule is no larger than 15 minutes. Generally, the reference point for a flight is its on- and off-block time, respectively. Thus, for arrival punctuality, scheduled on-block time is compared to actual on-block time. For departure punctuality, the same accounts for off-block times. Often, arrival and departure punctuality are referred to as inbound ($INBP$) and outbound ($OTBP$) punctuality. Total punctuality is simply the fraction of all flights being punctual, be it arrivals or departures. For punctuality modelling in the present work, the focus is on total daily punctuality ($TOTP$) which is defined accordingly as:

$$TOTP = \frac{\text{daily punctual flights}}{\text{total daily flight movements}} \tag{1.1}$$

$TOTP$ thus ranges between 0 and 1. Alternatively, it is sometimes given in percent. The definition of punctuality has clear advantages over other delay performance indicators, but also drawbacks not to be dismissed. Using the 15 minutes envelope, small deviations from a schedule are not accounted for. The reasoning here is that in air traffic small unpunctualities are generally accepted, still making smooth operations possible. Also, as there exists no international definition on delay reporting, but only a set of conventions (Theusner and Röhner, 2008), some airlines/airports measure delays at different time thresholds (Guest, 2007), thus making cross-comparisons difficult. However, delay is generally measured at least from 15 minutes upward, again speaking for the use of punctuality as delay measure for comparisons. With regard to large delays, punctuality is an unbiased delay measure and not over-sensitive to extensively delayed flights, which do not reflect the delay charging for the majority of flights (see also Spehr, 2003).

Focusing on drawbacks, on-/off-block references are at least to be discussed. For airport purposes, it might be appropriate to use this reference as delays resulting from e.g. taxi-out hold-up, remote de-icing queues or departure queues at runway heads are not accounted for and do not show up in the performance statistics. From an airline or passenger viewpoint, on the other hand, it is not only important that aircraft go off-block on time, but also that they are punctual at their destination. Any delay produced after gate departure is thus potentially critical.

Moreover, as punctuality is defined with published schedules as a reference, it is only to a limited amount appropriate for an evaluation of true system performance as compared to optimum performance. Airlines tend to apply schedule padding, especially at peak times[1], when there are systematic

[1]For further information on schedule padding and resulting problems at peak times please refer to Frank et al. (2005)

deviations of actual block times[2] from scheduled block times to compensate for these systematic delays and to reduce their knock-on effects. Thus, flight schedules are adapted using schedule buffers, i.e. extra time to absorb arrival delays, unexpected departure delays due to ground handling disruptions and to accommodate inevitable time gaps in flight schedules, thus maintaining a good on-time performance (EUROCONTROL, 2005).

If the focus is on an evaluation of true system performance, schedules should thus be validated against optimal operations which assume ideal operational conditions. When interpreting punctuality over time, the application of schedule buffers should be kept in mind. An improvement of punctuality from one year to another might in that respect not come from improved or more efficient operations but just from airlines using more conservative schedule buffers (see also Theusner and Röhner, 2008). The interpretation of punctuality trends is thus rather difficult. For a more advanced approach to delay accounting see Thrasher and Weiss (2001).

Furthermore, not only the amount of delay produced is of interest, but also its origin. Delays were thus classified according to a delay code system introduced by IATA (see Appendix E), with roughly 80 different delay codes covering all types of causes which may trigger a delay. Delay codes generally have to be handled with care. Often, there is not a single reason for a delay, but it is an integrated effect of many reasons. If a delay can then only be assigned one delay code, the complexity of reasons is lost and the reality is strongly simplified. In addition, throughout the daily rotation of an aircraft with often several flight legs, the true reason of the original delay might get lost. Imagine a flight waiting for crew, passengers and/or baggage from another flight, which is delayed due to constricted operations through heavy snowfall at its departure airport. This flight is likely to be assigned a reactionary delay code (see Appendix E). However, the true reason for this reactionary delay was adverse weather at another airport. This simple example impressively shows that delay codes as used today can only to a limited amount reflect the true complexity of delays, heavily depending on if the focus is on original or immediate delay. In general, there is too much weight on the downstream processes and an underestimation of upstream influences (Niehues et al., 2001). Often, only the last and most obvious disturbance or event in the process is reported as the cause of the delay. Statistics presenting the amount of delay attributed to certain cause groups of reasons (see e.g. EUROCONTROL, 2007, 2009a) can thus be strongly biased.

Last but not least it should be emphasised that punctuality does not

[2]The *block time* is the time between off-block at the departure airport and on-block at the destination airport. It includes the taxi-out time at the departure airport, the airborne time and the taxi-in time at the arrival airport. Thus, it is often referred to as *gate-to-gate time*.

account for flights being early. Thus, such flights are not accounted for in any bonus system. Likewise, punctuality does not allow for the negative effects of flights being early, such as still occupied gates, short term gate changes or built-up of holdings due to increased demand exceeding nominal capacity, to name just a few.

1.2 Low Punctuality: A Typical Example

There are several causal reasons for low punctualities. Looking at Frankfurt Airport, days exhibiting large delays are usually days with adverse weather conditions. As an example for a typical low punctuality day, 6 October 2006 is chosen. Middle-Europe, at that time, was in a westerly to southwesterly flow. Frontal systems connected to the storm system *ex Isaac* were to cross Germany from West to East (Verein Berliner Wetterkarte, 2006a,b). The storm system was supposed to intensify and the German Weather Service (*DWD*) issued a wind warning for Frankfurt Airport, valid from 9 to 16 local time.

Figure 1.3 gives background information on the development of traffic, runway occupation and punctuality at Frankfurt Airport on 6 October 2006. The data are taken from the FRAPORT *On-Time Navigator* system. Displayed is the number of arrivals, departures and total movements, against the background of nominal arrival and departure capacity, which are both variable over time (FHKD, 2008). Runway occupancy is shown in the lower part of the graph. In the upper part, additional information is given on punctuality and flight movements, subdivided into total figures and figures for day- and nighttime. More detailed information on the development of arrival, departure and total punctuality is given in Figure 1.4.

Punctuality graphs in both figures indicate that departure operations were rather smooth until shortly after 7 hours local time. Arrivals, on the contrary, experienced low punctualities around 50 % already at that early time of day. With increasing traffic during the first arrival bank, holdings were built up due to high winds and, consequently, reduced capacity through conservative approach staggering. The German ATC (DFS) had to limit arrivals to below 40 per hour in the morning hours. This is, though not significantly below nominal capacity (see Section 2.1.1), a crucial capacity cut during the time-critical first arrival push.

After eight o'clock, arrival punctuality dropped dramatically to values between 10 and 20 % and never recovered from that level before the late evening hours. A significant amount of this delay was produced in the Frankfurt Terminal Manoeuvring Area (TMA). As a consequence of low arrival punctuality, also departure punctuality dropped significantly. From around nine o'clock, values were in the range of 20 to 30 % and thus only slightly better than the

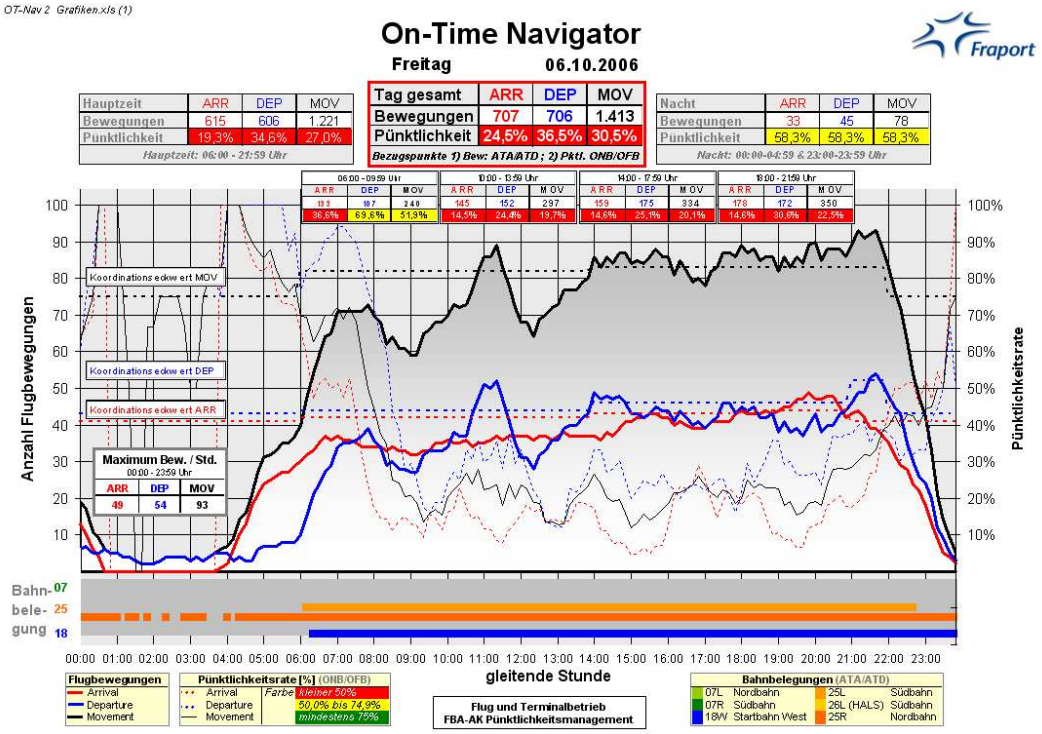


Figure 1.3: Traffic (solid lines), runway occupation (at the bottom of the graph) and punctuality (dashed thin lines) at Frankfurt Airport on 6 October 2006. Arrival/departure figures are given in red/blue and figures referring to total movements are given in black. Thick dashed lines indicate the current nominal capacity.

arrival figures.

From 9:30 local time, holdings could successively be reduced again. At 9:55 local time, the DWD cancelled the wind warning for Frankfurt Airport. However, up to this point, schedules had already been disturbed to a degree not allowing for a recovery to normal operations for the rest of the day. This is remarkable, since from afternoon hours, DFS could turn back to normal operations with unconstrained arrival capacity.

Analyses of prevalent weather at Frankfurt Airport revealed that wind speeds, both upper level and surface, were high on 6 October, but not exceptionally high. Runways could continuously be used for operations without cross- or tailwind-components exceeding critical thresholds, thus implying partial or total runway closures. Still, the local wind conditions were non-optimal at a critical time, thus at least triggering the successive built-up of delays. The weather impact on that day was, however, not limited to the Frankfurt area. Quite the reverse, weather had an impact on larger regions in Europe. From 11 local time, a significant fraction of arrivals at Frankfurt Airport was delayed due to en-route regulations through the CFMU. These prevented operations from recovery when local conditions would have allowed for it. At the end of that respective day, average arrival punctu-

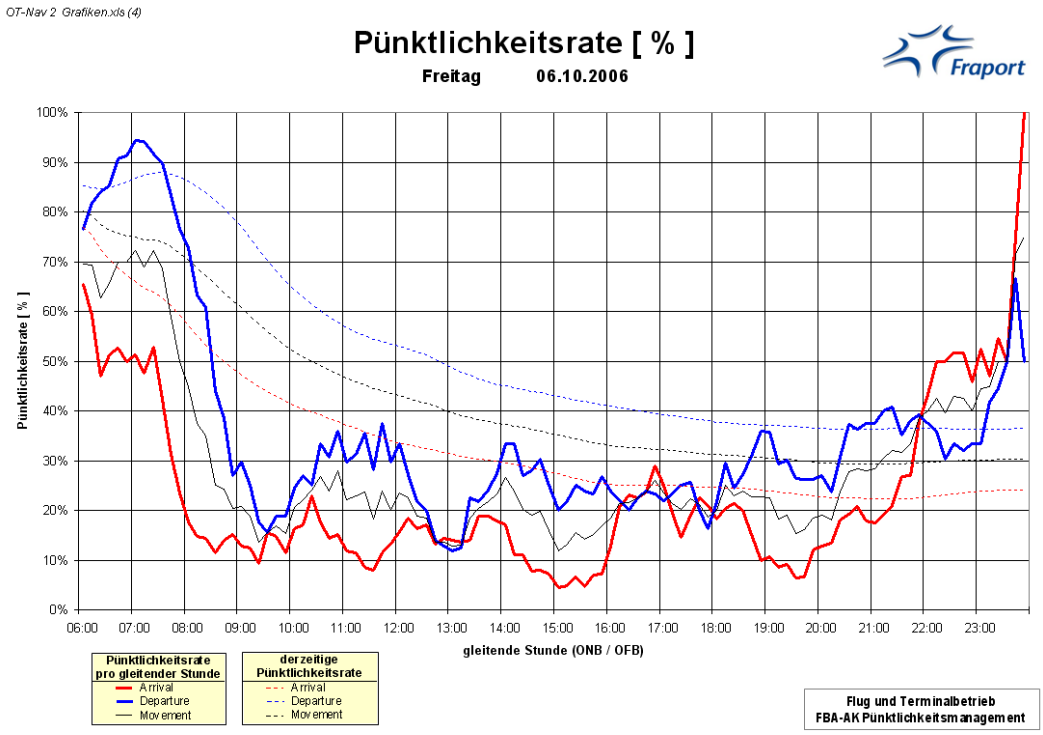


Figure 1.4: Current (solid lines) and integral (dashed lines, for the current day up to the time of reading) punctuality at Frankfurt Airport on 6 October 2006, subdivided into arrival, departure and total punctuality.

ality at Frankfurt Airport was at 24.4 %. Departure punctuality at 36.5 % was a little higher, leading to a total punctuality of 30.5 % for 6 October. Altogether, 1413 movements were operated, 30 flights were canceled.

This example impressively shows how delays are easily built up at an airport with tight schedules and operating at its nominal capacity. The role of weather in that process is complex and adverse weather conditions, both local and outside the airport TMA, are likely to have a large effect on operations and, consequently, punctuality. An analysis of the size of the impact is within the scope of this study.

1.3 Review of Previous Studies on Air Traffic Delay and Delay Modelling

Modelling of air traffic delay is a rather new area of research. The background of studies and initiatives is manifold, but mostly economic reasons were and are the drivers of research and development within this sector. Air traffic delays are cost intensive and in times of growing traffic and limited capacity, air traffic stakeholders are looking for ground-breaking ideas, supporting

them in the understanding of the complex processes leading to delays. Along with this gain of insight, resources are expected to be used more efficiently. Even more, new strategies and products are likely to help in the monitoring of processes and the evaluation and quantification of expected or realised benefits.

In the recent years, several scientific studies on the weather impact on air traffic were accomplished. Whereas most of these were performed in the U.S., only a limited number of comparable studies is known for Europe. In a pilot study on the weather impact on air traffic, Sasse (2000) and Sasse and Hauf (2003) focused on a single weather impact factor and analysed the effect of thunderstorms in the Frankfurt TMA on landing traffic. It was found that average hourly arrival delay on thunderstorm days was up to 10 times higher than on optimum weather days. Arrival and departure delays at Vienna International Airport in 2002 were analysed by Peer (2003) and Peer et al. (2008) with special focus on the impact of different weather factors. Snowfall and low visibility conditions proved to be the dominating weather-related causes for delays during the winter season. In the summer period, thunderstorms and strong precipitation rank first. Röhner (2004) and Röhner and Hauf (2008) concentrated on the impact of winter weather on airport operations and punctualities. They found that at Frankfurt and Munich Airports, most delays are produced when snowplough and runway treatment operations are conducted. These necessitate the partial or total closure of runways and taxiways, which produces large schedule disruptions, especially during arrival or departure peaks. A cumulative effect comes from aircraft de-icings, which are generally inevitable under these conditions. A major conclusion from this study was that the size of the impact of winter weather conditions on airport operations and, consequently, delays depends heavily on the operational infrastructure at the airport of interest. This implies that findings for one airport might not directly apply to other airports, even when the same impact factors are considered.

The current best source of information on delays and delay causes in Europe, including analyses of weather and non-weather related sources, is EUROCONTROL. A comprehensive introduction in the subject of delay, considering its definition, nature and categorisation, its causes, its measurement, and its mitigation, is given by the manager of the EUROCONTROL Central Office for Delay Analysis (CODA), Tim Guest (Guest, 2007). CODA as an organisational unit of EUROCONTROL produces regular and ad-hoc air traffic delay reports, available on their website (EUROCONTROL, 2009a). Delays are analysed based on the IATA delay code system (see Appendix E). Various statistic are compiled and provided on a monthly basis. These allow for a rudimentary overview over the weather-related share of air traffic delay. In-depth analyses are, however, not published. The IATA delay code system does not allow for a detailed breakdown into weather factors. Raw

data are not provided for independent investigations. General analyses of the performance of the European Air Traffic Management system are presented in EUROCONTROL's annual Performance Review Reports (see e.g. EUROCONTROL, 2009b). Aside from the delay and punctuality situation in Europe, these reports focus on the key performance areas of safety, capacity, flight efficiency, cost-effectiveness and environmental impact. Another valuable source of information on delays and punctuality in Europe and the determining factors is the Performance Review Commission of EUROCONTROL. Within a report on punctuality drivers at major European airports (EUROCONTROL, 2005), airport related details on the weather impact on operations and, especially, on capacity and delay, are given. There a three major factors for capacity reduction during bad weather identified: strong winds/thunderstorms, reduced visibility and quality of meteorological forecasts and their integration in the ATFM/ATC decision-making process. In that respect, a major problem is that at certain airports a significant gap between good and bad weather capacity exists. This gap depends strongly on an airport's assumed service quality criterion, the given airport layout and its usage and the general preparedness for bad weather in terms of equipment and processes.

Already in the nineties, studies on the weather impact on air traffic were carried out in the U.S., where the weather regime is somewhat different from that in Europe. On the one hand, weather phenomena like hurricanes or tornadoes are less prevalent or absent in Europe or, when thinking of e.g. thunderstorms, just have a less strong impact. Additionally, there exist significant differences in the organisation and configuration of U.S. and European airspace and airports[3]. In the context of delay generation, the most striking difference is that in Europe, declared capacity for an airport is based on IMC (Instrument Meteorological Conditions) operations, as compared to VMC (Visual Meteorological Conditions) operations in the U.S. Thus, adverse weather conditions are likely to have a larger impact on air traffic operations in the U.S., from the outset.

As a support for NASA's Terminal Area Productivity Program, Chin et al. (1997) investigated U.S. airport surface delays and causes, accessing several databases, reports and information sources. They found that weather frequently has an impact on surface movements even when the airport of interest is not directly affected by adverse weather. Nevertheless, delays are generated through ground holds due to adverse weather at the airport of destination or en-route. Chin et al. emphasised the importance of visibility for surface movements. In that context, it should be noted that ground visibility is often still sufficient for normal ground operations even when IFR are in effect. A major problem Chin et al. faced with respect to

[3]For more information, refer to Liang et al. (2000).

a detailed quantification of the weather impact on air traffic operations was the limitation of delay databases, in particular with focus on delay causes. Weather as one among many delay generating factors was hard to isolate. Accessing the FAA's Air Traffic Activity and Delay Report and considering delays greater than 15 minutes only, Chin et al. found that, from 1984 to 1994, about 65 % of delays were due to weather. Other studies from Lockheed Martin and from MIT Lincoln Laboratory, where a simple Aviation Weather Delay Model, based on average delay on clear days and additional delay on days with a particular type of weather in effect, was developed, estimated that weather accounts for roughly 40 % of all delays, considering all delay durations. Results obtained from the Aviation Weather Delay Model also revealed that the fraction of weather-related delays varies significantly among U.S. airports. Whereas, for 1994, only 22.7 % of all delays at John F. Kennedy International Airport were due to weather, Detroit Metropolitan Wayne County Airport and Los Angeles International Airport experienced more than 60 % of weather-related delays.

Dillingham (2005) analysed U.S. federal government, airline and airport initiatives to reduce flight delays and enhance capacity. In particular, he focused on remaining challenges against the background of adverse weather and limited resources. Dillingham emphasised that 70 % of the U.S. flight delays from 2000 to 2004 were, according to FAA figures, related to weather, followed by the general lack of capacity. Large potential for improvement with focus on an optimisation of airspace utilisation through maximisation of system throughput and optimisation of traffic flow is expected from new technologies such as the Integrated Terminal Weather System or the Collaborative Convective Forecast Product, which is dedicated to support strategic planning and management of air traffic during severe, mostly convective weather.

Both in the U.S. and in Europe, several studies were conducted in which the costs related to delays and the benefits that might be achieved through reduction of these delays were analysed. Robinson (1989) investigated the impact of various types of weather on operations for one airline operating at Atlanta Hartsfield International Airport. He found that fog and thunderstorms, only, involved annual delay costs of more than $6 million for that respective airline. Robinson emphasised that better forecasts could potentially reduce the costs through adapted flight planning. ITA (2000) concluded that for 1999, delay costs in Europe were between 6.6 and 11.5 billion Euros. These estimates included airline costs, passenger costs and costs for non-optimal scheduling. Within a report prepared by the Performance Review Commission of EUROCONTROL, European airlines' costs of one minute of airborne or ground delay were evaluated (EUROCONTROL, 2004). Extensive information is presented for different cost scenarios, aircraft types and lengths of delay. It is emphasised that tactical costs of delay de-

pend strongly on these key factors. One major finding was that passenger delay costs incurred by airlines are found to be at 0.3 Euros per average passenger, per average delay minute and per average delayed flight.

With focus on modelling, there exist several studies with wide-spread objectives. In the U.S., initiatives to estimate benefits of planned or executed investments induced the development of models that, among other things and often just as a spin-off product, allowed for a quantitative analysis of the weather impact on air traffic operations, with special focus on system performance and efficiency. Hansen and Wei (1999), for example, estimated the benefit of a major capacity expansion at Dallas-Fort Worth International Airport, employing multivariate statistical methods. They related flight times to demand, weather, origin airport congestion and the expansion itself. In their a posteriori study, they found that through this NAS investment, the daily average flight time for arrivals could be significantly reduced, especially on low visibility days. Hansen and Wei, however, pointed out that benefit estimates were difficult to assess since other effects, such as increased demand or worse weather, had partially offset the primary benefits from the expansion. In a study on arrivals at Los Angeles International Airport, Hansen and Bolic (2001) again used the approach described in Hansen and Wei (1999) in order to investigate and isolate the effect of two system enhancements. Running the developed models, about 75 % of the day-to-day variation in their daily flight time index, which basically comprised a weighted average of daily arrivals' flight times, could be explained, with origin airport congestion being the most and demand being the least important source of variation. Weather factors associated with temperature, wind or visibility were found to be highly relevant. Additionally allowing for quadratic and interaction terms, even an R^2 of 0.82 could be achieved.

Callaham et al. (2001) addressed the problem of weather normalisation, which is necessary when airspace system performance measures are compared over different time intervals and effects of system improvements are to be quantified. They developed a scalar Weather Impacted Traffic Index (WITI) that classifies each day according to occurred weather, both en-route and within terminal areas. In a second approach, Callaham et al. analysed a weather and traffic based day-type clustering approach in order to determine the contribution of weather to airspace system performance. In both approaches, regression modelling was applied, using an airspace system performance measure such as e.g. average arrival delay greater than 15 minutes as the predictant. Results obtained were promising. Taking e.g. the WITI approach, an R^2 of 0.71 could be realised. A potentially valuable spin-off application of both methods is their usage for a forecast of airspace performance for a day in the future, given that traffic and weather forecast figures are provided, thus aiding strategic ATFM planning and decision making.

Against the background of improved weather forecasting, which is ex-

pected in the future, Hoffmann et al. (2004) analysed the weather impact on airport ground delay program (GDP) procedures through a linear regression modelling approach. Chicago O'Hare International Airport was chosen as the airport of investigation. Traffic and simple, reflectivity-based weather information was used to determine the relationship between weather and enroute delays. Based on this, a projection of arrival delay for arriving flights is provided. This information can finally be used to determine optimal GDP procedures thus minimising delays. Abdelghany et al. (2004) focused on flight delay propagation within a single airline schedule. They constructed a model which allows for a projection of down-line flight delays during irregular operations, especially when GDPs are issued. The model additionally allows for a distinction of delay reasons. For example, it is found that about 42 % of all delayed flights are delayed due to assigned aircraft being not timely ready. The major benefit of the delay projection system is that controllers can take pro-active recovery action to recover or even avoid delays. For one U.S. airline only, the benefit of the flight delay projection amounted to $1.6 million in the first quarter of 2004.

For Europe, comparable modelling approaches in many cases evolved from more theoretical reasoning. Often, pure scientific questions were the drivers for developments and tools that, nevertheless, are potentially qualified to be used in operational environments. One of the first studies that concentrated on the modelling of punctualities in Europe was published by Spehr (2003). Using a two-year dataset from Frankfurt and Munich Airports, Spehr applied multivariate linear regression in order to explain variability in daily punctualities through local weather and traffic. Diagnostic model results achieved were promising, exhibiting R^2 values between 0.4 and 0.5. However, exceptional days, e.g. days with strikes or system failures, were a priori excluded from the analysis. Besides this drawback, several potential model improvements were suggested and are subject of this thesis.

Rehm (2003) and Rehm and Klawonn (2005) took up the modelling approach introduced by Spehr (2003) and developed it further in another direction. Instead of analysing punctualities, Rehm and Rehm and Klawonn concentrated on the modelling of schedule-independent travel times from entrance into an airport's TMA until landing. Indeed, these times are less interesting from a passenger point of view, however, they constitute an important quality measure for air traffic services responsible for approach and landing. By applying this approach, only local effects are considered. Delays produced outside the TMA are not taken into account and do thus not bias computations. Rehm and Rehm and Klawonn evaluated several data mining techniques such as linear regression, regression trees, fuzzy clustering and neural networks. Good modelling results could be achieved through application of linear regression, where diagnostic R^2 values realised were in the range of 0.6. Solf (2005) analysed several factors having an effect on

the variability of aircraft approach times within an airport TMA. Exemplary calculations of approach times were made with the aid of a trajectory calculator. Solf found that blocking of runways, failures of landing systems and adverse weather have the greatest influence. With focus on weather, thunderstorms, solid precipitation and strong winds rank first.

In several studies, Wu and Caves (2000, 2002, 2003a,b, 2004) and Wu (2005) did significant research in airline network and aircraft rotation modelling as well as in the field of schedule optimisation. They developed both an aircraft turnaround and an en-route model to simulate ground operations and processes as well as aircraft rotation throughout a whole day. These models, for example, allow for an in-depth analysis of the relationship between flight schedule punctuality and aircraft turnaround efficiency at airports and can thus be used to minimise system operational costs while maintaining a required level of schedule punctuality. As a major finding of the studies by Wu and Caves, departure punctuality of a turnaround aircraft was found to be mainly determined by the arrival punctuality of inbound aircraft, the operational efficiency of aircraft ground services and the length of the scheduled turnaround time. In that context, longer turnaround times at hubs including conservative buffer times turned out to improve the punctuality performance in aircraft rotations. In a study issued by EUROCONTROL (2003), the propagation of flight delay was investigated using local arrival and departure models on French flight and airport data. These models were based on predictors of the form "day of the week" or "load on arrival". Weather was not directly covered but indirectly captured by the predictor "capacity". It could be shown that arrival delay depends heavily on departure delay produced at the previous station. For departure delay, a strong correlation to the previous arrival delay is found. Nevertheless, departure delay for an aircraft can generally be predicted by departure plane load, in case the aircraft experienced no significant arrival delay at the station of interest. By the use of a general model for delay formation along aircraft trajectories, it could also be demonstrated that long delays are generally produced through ATFM regulations or non-traffic related events.

Eventually, there are some industry developments outside the frame of scientific studies and modelling. Most of these developments arose from specific and often location-dependent air traffic stakeholder needs and are, in most cases, also limited in their application. System prototypes, tools and even fully operational systems are mostly customised to account for local or regional environmental, instrumentational and organisational constraints. At Frankfurt Airport, for example, the *CAPMAN* tool was developed in the frame of the K-ATM project (FRAPORT, 2007; Brozat, 2007). The initial purpose of CAPMAN is a forecasting of available airport capacity for the ongoing day. As a spin-off product, a punctuality module allows for a derivation of punctuality estimates from demand and predicted capacity. CAPMAN,

in that sense, is to be understood as a nowcasting tool, intended to be implemented in FRAPORT's operational infrastructure. A similar tool is also known for Amsterdam Schiphol Airport. The Capacity Prognosis Schiphol Tool *CPS*, a joint project of Amsterdam Schiphol Airport, the local ATC, KNMI and KLM, provides short- and long-term capacity forecasts, based on a special probability forecast of weather elements by the Royal Netherlands Meteorological Institute (KNMI). In particular, only wind, visibility and snowfall are considered (KLM, 2009). The tool is to be understood as support for decision making in the case of anticipated severe weather and runway maintenance.

These developments show that there is a clear need for both an improved understanding of delay-generating processes and, consequently, strategies and tools aimed at a reduction of air traffic delays and a more efficient use of limited resources. All studies introduced confirm that there is great potential for saving costs.

1.4 Motivation and Objectives

As introduced in the previous sections, there exists only a limited number of studies where mathematical models were used to investigate delays and related implications. On the contrary, many analyses on air traffic delay draw upon delay codes or related delay specification sources with all their drawbacks. Many authors have criticised this approach. Therefore, Niehues et al. (2001) proposed three lines of action out of this misery:

1. process monitoring and sampling
2. simulation
3. conventional analytical methods

The present work is based on the latter approach. The starting point for the analysis at hand is the groundwork on punctuality modelling done by Spehr (2003). Her work paved the way by setting a principal punctuality modelling frame, discussing the mathematical background and presenting first results and limitations of the underlying modelling approach.

The present work picks up on suggestions for improvement made by Spehr. For her investigations, Spehr used two limited 2-year data sets from Frankfurt and Munich Airport. Naturally, a first recommendation was an enlargement of the underlying database in order to investigate stability and generalisability of punctuality models and results obtained. Secondly, Spehr proposed the inclusion of predictor variables related to upper level wind, as an in-depth analysis of low-punctuality days provided a clear indication of the strong impact of upper level winds on current airport capacity and thus punctuality. A third area for improvement is a higher temporal

resolution of predictor variables. That way, also different impact weighting during the course of day can be realised. In order to better reflect the actual demand, information on scheduled traffic or another capacity-related predictor should be additionally included. Last but not least, a better integration of operational thresholds and a translation of these into meaningful predictor variables was recommended. Together with these suggestions for improvement, new ideas are developed and tested in different enhanced punctuality models.

The general question evolving for the present work is: To what extent can modelling results be improved? In order to answer this question, a working hypothesis is formulated. It is assumed and it will later be shown that a major part of the variability of daily punctuality and thus air traffic delays is attributable to weather. In that respect, the present work abstains from an analysis of en-route weather and concentrates on weather at the airport and within its terminal area. The reasoning behind this assumption is that weather observed in that respective area is representative for a much larger catchment area. To be more precise, weather observed at Frankfurt Airport or its vicinity is often correlated with weather observed at other German or even Middle-European airports. For example, frontal systems moving in from the Atlantic and affecting Frankfurt Airport are likely to have already passed Amsterdam Schiphol Airport or Paris Airport and are also likely to later affect e.g. Munich or Vienna Airport. This is of course not always true and often certain weather events are clearly local, especially when they are of convective nature. But in general, it can be assumed that adverse weather is locally correlated within Europe to a certain extent. When focusing on weather on a daily basis, e.g. through formulation of day statistics, a weather signal recorded at Frankfurt Airport is then to be interpreted as partly representative for a certain correlation radius, with decreasing correlation when moving away from its centre.

In the approach at hand, the complex processes generating delays are not considered separately. Rather, punctuality is modelled through an integrated multivariate approach, using local parameters only. This allows for an analysis of combined effects without abstaining from special analyses on single factor impacts. Single processes, however, need not to be individually modelled. Local weather is, in that respect, considered as a governing factor, covering or representing many delay sources. For example, crews or passengers may arrive late at the airport due to weather-induced traffic holdups, or ground operations such as e.g. baggage or bus transport may be impeded by contaminated apron surfaces. Thus, weather acts on airport operations in many ways, be it directly and indirectly. Based on this, four major questions are derived as the cornerstones of this work:

1. To what extent can punctuality be modelled, i.e. how much of the variability of daily punctuality can be explained through the approach chosen?

2. How large is the weather impact on punctuality?
3. What is the weighting of certain impact factors, be it weather- or non-weather related?
4. Is the chosen approach qualified for a punctuality forecast?

The stated questions are to be answered using Frankfurt Airport as the airport of investigation. It is assumed that underlying findings can be generalised or at least transferred to other major European hub airports. It has to be pointed out that the present approach does neither take into account imported delays[4] nor delay generating ground operation processes at aircraft turnaround[5]. Nevertheless, these are indirectly reflected in the daily punctuality statistics used for model calibration. The modelling approach itself, however, remains purely independent of these.

Taking all this into account, the present work is ground-breaking in the modelling of punctuality and thus makes a significant contribution to the understanding of the true weather impact on air traffic operations and delays. Moreover, the developed punctuality forecast model, potentially allowing for a punctuality forecast on a daily level, offers new planning possibilities and is thus of great value as a decision support for air traffic stakeholders depending on performance parameters for the planning of daily operations.

[4]See e.g. Wu and Caves (2003b) or EUROCONTROL (2003) for more details on the development of knock-on delays through aircraft rotation and Beatty et al. (1998) or Boswell and Evans (1997) for details on flight delay propagation.

[5]See e.g. Wu and Caves (2000) or Wu and Caves (2004) for more details on the impact of aircraft turnaround processes on punctuality.

Chapter 2

Methodology and Data

2.1 The Study Airport Frankfurt

2.1.1 General Description

Frankfurt Airport (ICAO code *EDDF*, IATA code *FRA*) is an airport in the state of Hesse, Germany at 111 m a.m.s.l. It is the largest German airport and it is equipped with three runways (see Figure 2.1). The parallel runway system (RWY07/25) is operated as a dependent dual runway system and is used for both take-offs and landings. Runway 18 is used for departures only. It is solely operated in southward direction.

Figure 2.1: Frankfurt Airport with its three runways.

The development of declared airport capacity at Frankfurt Airport is shown in Figure 2.2. The current declared airport capacity for Frankfurt Airport is 75-83 total movements per hour (FHKD, 2008), depending on the time of day. Within this total frame, 41-44 arrivals and 43-52 departures

can be operated per hour. During peak hours, these hourly values can be exceeded for short periods in fair weather conditions.

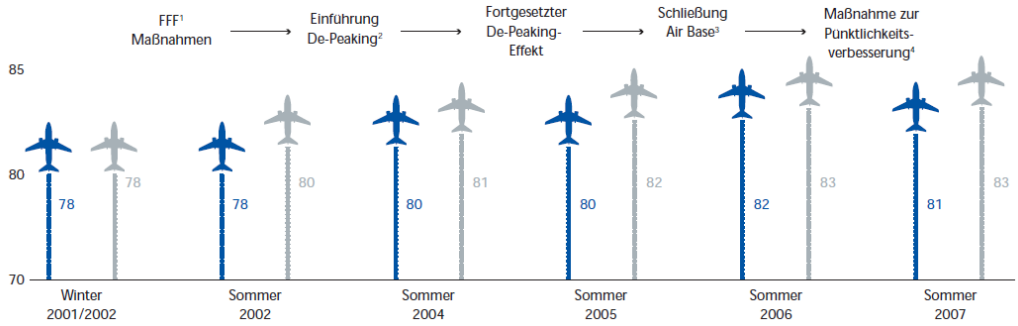


Figure 2.2: Declared capacity at Frankfurt Airport (FRAPORT, 2006).
[1)]Future for FRA (cooperation between DLH, DFS and FRA)
[2)]more evenly slot distribution per hour regarding arrivals and departures
[3)]closure U.S. airbase
[4)]cutting of declared capacity during busy morning hours (focus on punctuality), blue: before noon, grey: afternoon.

2.1.2 Traffic and Passenger Volume at Frankfurt Airport

Frankfurt Airport has an annual traffic volume of almost 500,000 flight movements and a passenger volume of more than 50 Million passengers per year. It is thus the second largest airport in Europe behind Paris Charles de Gaulle with regard to traffic volume and the third largest behind London Heathrow and Paris Charles de Gaulle considering passenger volume. Figure 2.3 shows the development of traffic (upper chart) and passenger (lower chart) volume for the period of investigation. Traffic has increased from 456,000 flight movements in 2001 to 490,000 in 2006. Inbound and outbound flight movements are naturally balanced. Passenger volume has increased from 50 Million in 2001 to 53 Million in 2006. The 9/11 event likely explains the slight decrease in passenger volume in 2002 and 2003. Flight movements remained almost constant in these two years.

For modelling purpose (also see Sections 3.2.6 and 3.2.8), information on scheduled as well as actual traffic was available for Frankfurt Airport, separated into inbound, outbound and total movements per day. Daily scheduled traffic at Frankfurt Airport has continuously increased in the period of investigation. In 2001, there were an average of 1276 scheduled movements per day. In 2006 the respective figure was 1366. Figure 2.4 shows both the total scheduled daily movements ($TOTFL_{scheduled}$) and the trend-corrected daily residuals. The red line in the upper graph depicts the linear traffic trend. It

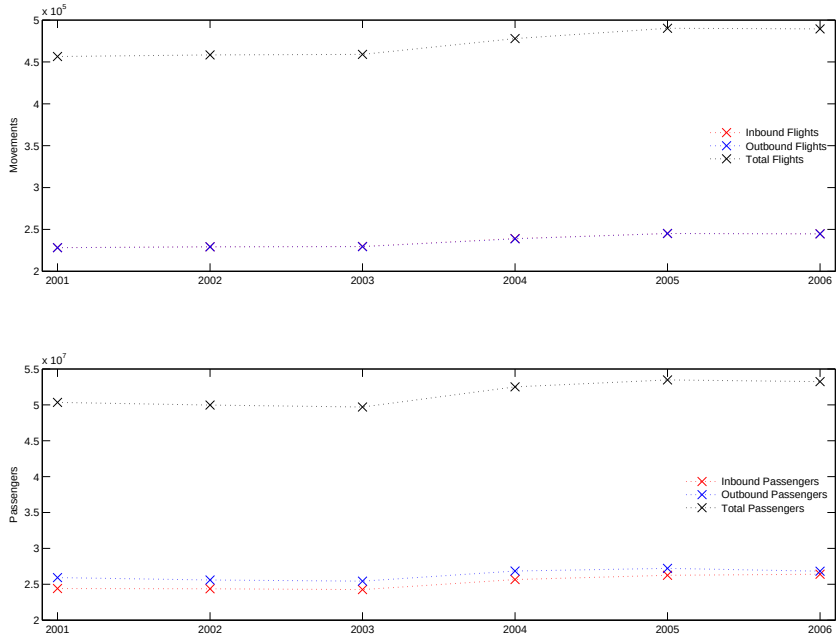


Figure 2.3: Actual traffic (upper chart) and passenger (lower chart) volume at Frankfurt Airport from 2001-2006. Red: Inbound, Blue: Outbound, Black: Total.

shows that $TOTFL_{scheduled}$ generally has a yearly minimum on New Year's Eve and New Year's Day and a maximum in the summer holiday season.

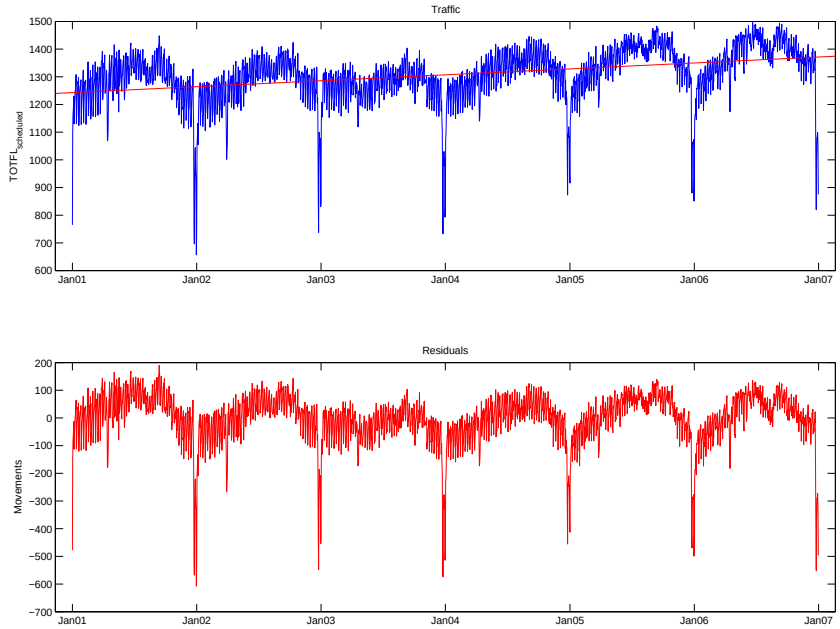
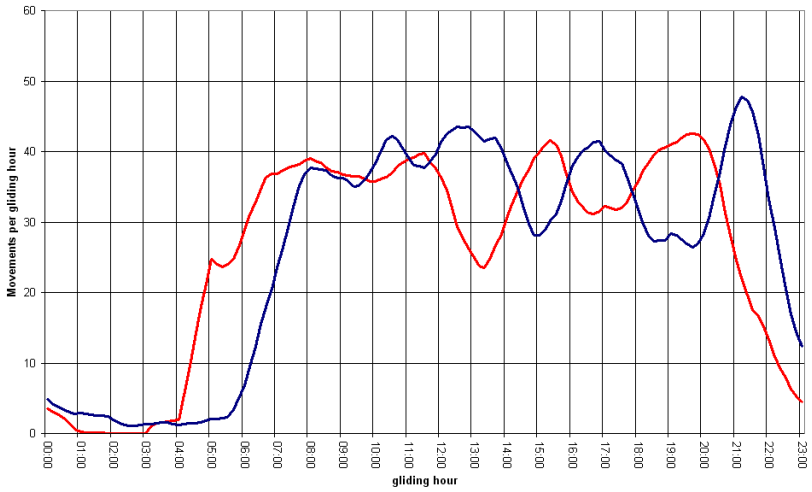
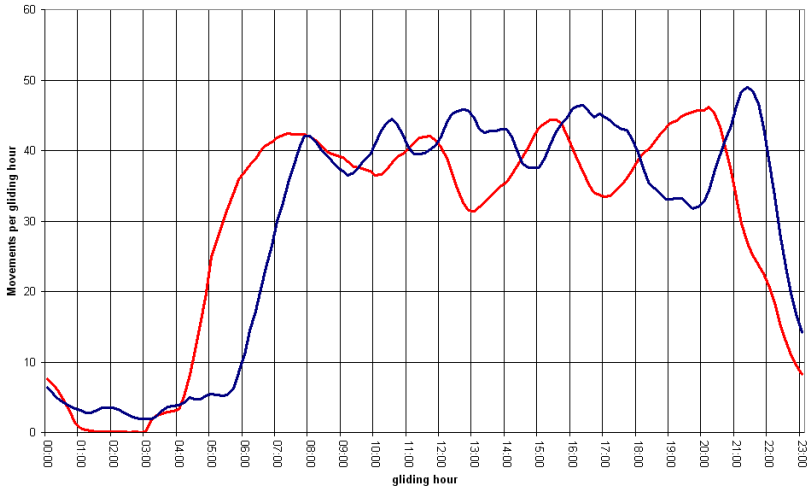


Figure 2.4: Scheduled daily traffic ($TOTFL_{scheduled}$) at Frankfurt Airport between January 2001 and December 2006. The red line in the upper graph depicts the trend in the 6-year period. The lower graph shows the trend-corrected daily residuals.



(a)



(b)

Figure 2.5: Arrivals (red) and departures (blue) at Frankfurt Airport in 2008 per gliding hour, (a) winter (November-March), (b) summer (April-October).

$TOTFL_{scheduled}$ is generally a good representative for scheduled flight movements. The correlation with scheduled inbound traffic ($INBFL_{scheduled}$) is high ($r = 0.995$), just as with scheduled outbound traffic ($OTBFL_{scheduled}$, $r = 0.995$). On most days, there is a small gap between scheduled and actual flight movements which is due to flight cancellations and diversions. On average, there are 15 flight movements less per day than actually scheduled with a maximum difference of 687 movements on the 3 March 2006. On that day, Frankfurt experienced heavy snowfall with a buildup of a snow cover of 17 cm at subzero temperatures. There are, however, also days with

more flights than actually scheduled. Generally, this is due to flight diversions to Frankfurt Airport, short term flight plan changes or recovering from disrupted previous days.

For reference purposes, Figure 2.5 shows average daily arrivals and departures (exemplarily for 2008), subdivided into summer (April-October) and winter (November-March). One can clearly distinguish four arrival and five departure banks, both in the summer and the winter season. Arrivals peak around 8 a.m., between 11 a.m. and 12 p.m., between 3 p.m. and 4 p.m. and around 8 p.m. Departure peaks are found at 8 a.m, between 10 and 11 a.m., around 1 p.m., between 4 and 5 p.m. and between 9 and 10 p.m. Especially the afternoon and evening arrival banks are clearly followed by departure banks. This is a typical pattern for a hub airport.

2.1.3 Operational Procedures

With regard to modelling punctuality at Frankfurt Airport, some operational procedures and thresholds have to be taken into account. As mentioned in Section 2.1.1, movements are officially limited to a maximum of 83 per hour. Within this frame, a maximum of 44 arrivals per hour can be handled. In low visibility and/or high wind conditions, the acceptance rate can decrease to as much as 35 arrivals per hour. Exact values are given in the arrival rate matrix (see Figure A.1 in Appendix A). It should be mentioned that at Frankfurt Airport regular operations are at CAT I level (see Table 2.1). Thus, deterioration of Runway Visual Range (*RVR*, corresponds to visibility) and Decision Height (*DC*, corresponds to ceiling) down to CAT II or III conditions imposes a lower acceptance rate. Regarding approach staggering, crucial wind levels are between FL30 and FL50 as shown in Figure A.1. Headwind thresholds inducing lower acceptance rates are at 15, 25 and 35 kts. Acceptance rate can, of course, decrease to 0 movements when runways and taxiways are temporarily closed during snowplough and runway treatment operations.

Runway 18 is generally closed when the tailwind component (hence nor-

Table 2.1: Definition of CAT stages, determined by minima of Runway Visual Range (RVR) and Decision Height (DH). In case RVR and DH apply to different CAT stages, the worse category is chosen.

CAT stage	**RVR minimum [m]**	**DH minimum [ft]**
CAT I	550	200
CAT II	300	100
CAT IIIa	200	50
CAT IIIb	75	<50
CAT IIIc	0	0

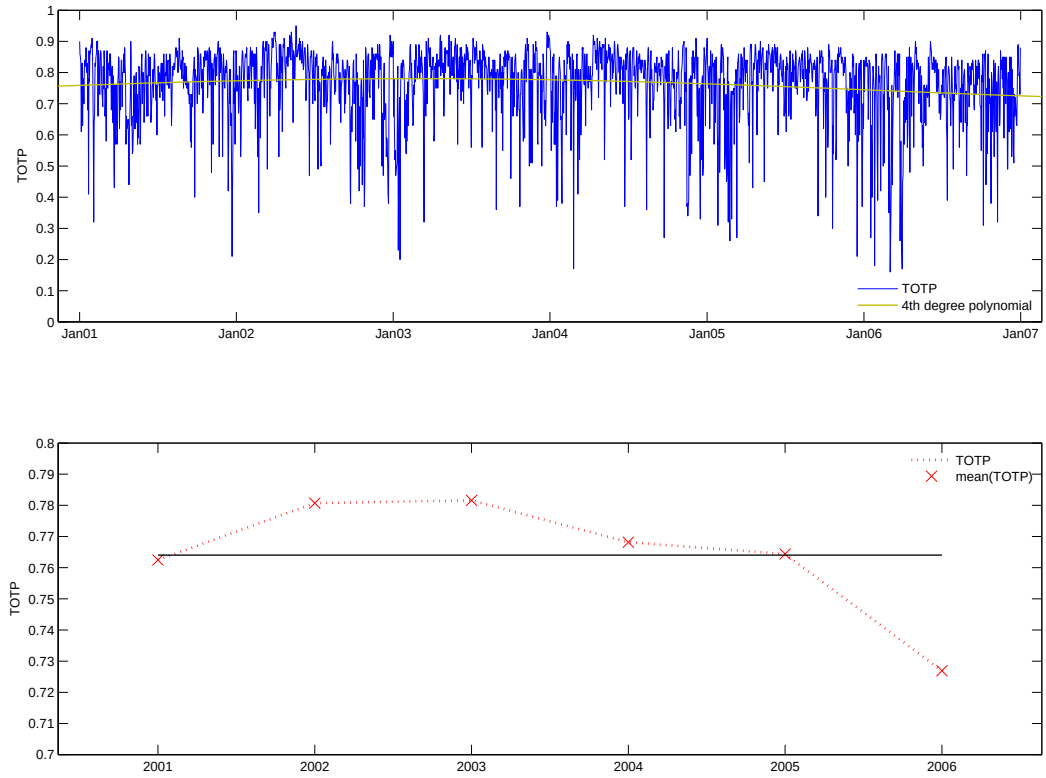


Figure 2.6: Daily punctuality *TOTP* at Frankfurt Airport between January 2001 and December 2006 (upper chart). Also shown is a 4th-polynomial fit. The lower chart shows annual means of TOTP and the mean of TOTP in 2001-2006.

therly wind) is greater than 15 kts. Operational constraints, however, already arise when the tailwind component exceeds 10 kts as some pilots consider this threshold as a crucial take off limit. There are no official thresholds for operations in heavy crosswind conditions. Runway configuration on the parallel runway system depends on the tailwind component. The preferred and also prevalent runway configuration is RWY25. A change of runway configuration is initiated when the tailwind component exceeds a limit of 5 kts.

2.1.4 Punctuality at Frankfurt Airport

Archived monitored punctualities at Frankfurt were available for investigations on a daily level. The upper chart of Figure 2.6 shows total daily punctuality (*TOTP*) for 2001-2006. The mean value for this period is 0.764. The lowest *TOTP* value was observed on 3 March, 2006 with 0.16. This was a snowfall day at Frankfurt Airport as described in Section 2.1.2. The highest value was recorded on 20 May 2002 with 0.95. Using a separation in summertime and wintertime period (daylight saving time as reference), an average punctuality value of 0.776 for the summer and 0.747 for the winter period is

Table 2.2: Annual statistics of *TOTP*. The last row gives the percentage of days with *TOTP*<0.5.

	2001	**2002**	**2003**	**2004**	**2005**	**2006**
min	0.21	0.35	0.20	0.17	0.21	0.16
mean	0.76	0.78	0.78	0.77	0.76	0.73
max	0.91	0.95	0.93	0.92	0.91	0.90
std	0.11	0.11	0.11	0.12	0.12	0.14
<0.5 (%)	3.01	3.84	2.19	4.64	4.93	7.40

found. Mean punctuality also varies among years. The lower chart of Figure 2.6 shows annual mean punctualities for 2001-2006. Punctualities lie roughly between 0.76 and 0.78, only 2006 sticks out with a mean of only 0.727. According to FRAPORT (2006) that drop in punctuality was due to weather, preparative reorganisation for the A380, short-term changes in parking positions, new security regulations and construction sites due to modernisation and fire prevention. Table 2.2 gives a summary of total punctuality statistics for each year in the investigation period.

Total punctuality correlates well with inbound punctuality (*INBP*) and outbound punctuality (*OTBP*). The first correlation is $r = 0.973$, the latter is $r = 0.965$. *INBP* and *OTBP* are correlated with $r = 0.881$. The autocorrelation of *TOTP* is shown in Figure 2.7. The lag-1 autocorrelation is $\rho = 0.42$ and thus significantly higher than for any other lags. Heavy disruptions on a certain day generally also have an impact on the following day. Hence, a day with a low *TOTP* is likely to be followed by a day with a suboptimal

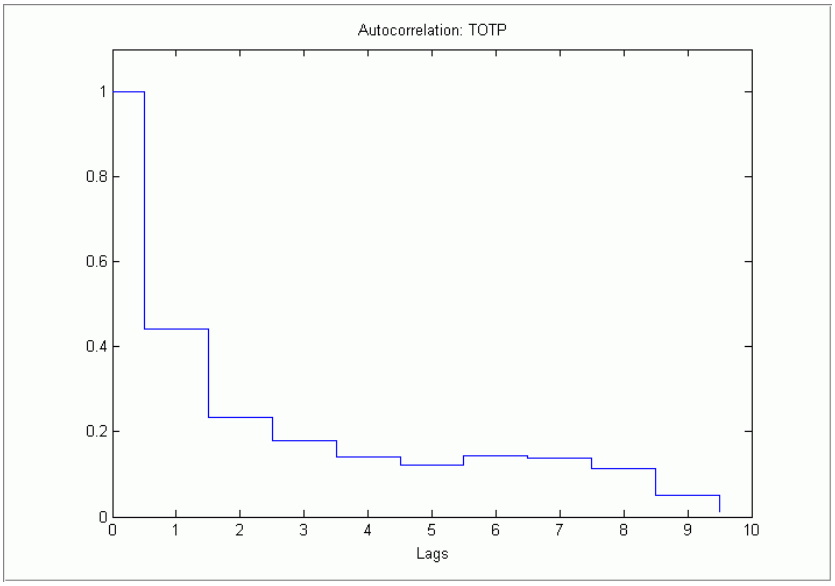


Figure 2.7: Autocorrelation of *TOTP* at Frankfurt Airport.

TOTP. There are several reasons for this effect, such as:

- aircraft scheduled for departure on the following day were not able to land on the disrupted day
- aircraft scheduled for departure on the disrupted day could not take off

On the day following the disrupted day, a situation is generated with certain aircraft not in place with regard to their regular schedule on the one hand and additional aircraft for departure on the other hand. At an airport operating at its nominal capacity, these deviations from the regular schedule are likely to cause suboptimal operations.

2.1.5 Weather at Frankfurt Airport

This section focuses on the variability of a selection of weather parameters at Frankfurt Airport within the 6-year investigation period. For longer climatologies, in-depth literature such as Heinemann (2008), Bartels et al. (1990) or Müller-Westermeier et al. (1999, 2001, 2003, 2005) is recommended. Data used are described in Section 2.2.2.

Table 2.3 shows the numbers of occurrence for a selection of weather parameters, relevant to airport operations, in 2001-2006. Numbers given represent the days of occurrence within each of the six years under investigation.

Table 2.3: Selection of weather parameters for Frankfurt Airport in 2001-2006. Numbers given are days of occurence in the respective years. *Min*, *mean* and *max* values are additionally shown.

parameter	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**sum**	**min**	**mean**	**max**
snow cover	27	20	16	31	30	27	151	16	25.2	31
frost	71	46	83	73	69	71	413	46	68.8	83
frost at ground	92	71	105	93	92	89	542	71	90.3	105
ground frozen	44	22	50	48	38	42	244	22	40.7	50
glaze	3	1	1	1	5	6	17	1	2.8	6
fog	18	10	13	17	13	26	97	10	16.2	26
thunderstorm	26	29	32	29	24	46	186	24	31.0	46
precip. any	243	230	195	241	237	241	1387	195	231.2	243
precip. any > 10 mm	17	24	7	13	13	15	89	7	14.8	24
precip. solid	41	21	29	54	51	40	236	21	39.3	54
precip. mod.[a]/str.[b]	129	118	77	92	98	103	617	77	102.8	129
precip. freezing	1	3	2	0	3	3	12	0	2.0	3
storm gust[c]	5	7	4	7	4	3	30	3	5.0	7
windlim RWY18[d]	39	49	43	33	36	30	230	30	38.3	49
$\text{wind}_{\text{ul}} \geq 15$ kt[e]	/	/	261	295	283	297	1136	261	284.0	297
$\text{wind}_{\text{ul}} \geq 25$ kt	/	/	150	163	140	170	623	140	155.8	170
$\text{wind}_{\text{ul}} > 35$ kt	/	/	64	65	61	86	276	61	69.0	86

[a] moderate
[b] strong
[c] ≥ 20.8 m/s
[d] 10 kt tailwind
[e] upper level wind (3000-5000 ft) in direction of RWY07/25

Table 2.4: Selection of weather parameters for Frankfurt Airport in 2001-2006. Numbers given are the *min*, *mean* and *max* number of days of occurrence in the respective months.

parameter		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
snow cover	min	5	1	0	0	0	0	0	0	0	0	0	0
	mean	10	6.2	3.2	0.3	0	0	0	0	0	0	1	4.5
	max	19	13	9	2	0	0	0	0	0	0	5	11
frost	min	12	5	5	1	0	0	0	0	0	0	1	12
	mean	17.8	15.3	11.8	2.3	0	0	0	0	0	1.2	4.8	15.5
	max	23	24	19	5	0	0	0	0	0	6	10	20
frost at ground	min	15	12	8	4	0	0	0	0	0	0	3	12
	mean	19.8	18.5	14.8	6.7	0.3	0	0	0	0	2.8	8.5	18.8
	max	24	27	21	11	2	0	0	0	0	11	15	23
ground frozen	min	0	4	2	0	0	0	0	0	0	0	0	7
	mean	9.5	10.5	6.7	0.3	0	0	0	0	0	0.5	2.8	10.3
	max	14	20	10	1	0	0	0	0	0	3	7	15
glaze	min	0	0	0	0	0	0	0	0	0	0	0	0
	mean	1.3	0.5	0.2	0	0	0	0	0	0	0	0	0.8
	max	5	1	1	0	0	0	0	0	0	0	0	3
fog	min	1	0	0	0	0	0	0	0	0	0	0	0
	mean	2	1.2	1.5	0.5	0.3	0.7	0.3	0.3	0.8	4.2	2.2	2.2
	max	3	3	5	1	1	1	1	1	4	6	4	5
thunderstorm	min	0	0	0	1	2	4	5	2	0	0	0	0
	mean	0	0.3	0.8	2.5	4.7	6.3	7.3	6.5	1	1.3	0.2	0
	max	0	1	2	4	8	9	13	13	2	3	1	0
precip. any	min	20	13	15	14	11	14	15	9	9	11	20	19
	mean	23	20.5	20.7	17.8	18.3	16.2	19	19.2	14.2	17.5	22	22.8
	max	26	25	26	26	23	20	25	28	27	24	25	27
precip. any >10 mm	min	0	0	0	0	1	0	0	0	0	0	0	0
	mean	1.2	0.7	0.8	0.7	3	0.7	1.2	2.2	1	1.3	1.5	0.7
	max	2	2	3	4	4	1	3	5	3	4	5	2
precip. solid	min	7	5	0	0	0	0	0	0	0	0	0	1
	mean	11.2	10.7	6.2	1.5	0	0.2	0.2	0	0.2	0	1.8	7.5
	max	18	18	14	5	0	1	1	0	1	0	6	16
precip. mod./str.	min	7	3	3	5	4	3	7	3	2	2	6	7
	mean	10.2	9.5	8.2	9.5	8.3	5.7	8	8.7	5.8	8.8	10.2	10
	max	13	16	18	16	14	9	10	16	12	15	14	13
precip. freezing	min	0	0	0	0	0	0	0	0	0	0	0	0
	mean	1	0	0	0	0	0	0	0	0	0	0	1
	max	2	0	0	0	0	0	0	0	0	0	0	2
storm gust	min	0	0	0	0	0	0	0	0	0	0	0	0
	mean	0.8	1	0.5	0.5	0.2	0.5	0.5	0.3	0	0.5	0	0.2
	max	2	3	2	1	1	1	1	2	0	2	0	1
windlim RWY18	min	0	2	1	4	1	0	0	0	0	0	0	0
	mean	2.5	4.5	4.7	6.3	5	2.7	2	2.2	2.5	1.3	1.5	3.2
	max	6	6	8	8	11	4	6	6	7	3	4	8
wind$_{ul} \geq 15$ kt	min	24	19	21	19	16	16	18	11	17	27	26	25
	mean	27.8	21.8	25.3	22.5	21.8	19	21.8	21	20	28	27.5	27.8
	max	30	24	30	26	27	23	25	26	25	29	29	31
wind$_{ul} \geq 25$ kt	min	13	10	6	9	3	4	5	4	5	16	14	13
	mean	19	14.3	12	10.8	10.8	7	9	9.5	8.5	19.8	18.3	17
	max	23	18	19	13	19	11	14	15	15	23	22	22
wind$_{ul} > 25$ kt	min	2	2	4	3	1	0	0	1	2	7	4	7
	mean	10	5.8	6.3	4.8	4.8	1.8	1.8	2.8	4	9	8.3	10
	max	17	9	11	7	10	3	3	5	7	12	16	16

For easier interpretation, the last four columns show total numbers as well as minimum, mean and maximum numbers. Table 2.4 gives more detailed information on the monthly distribution of the selected parameters. In the following, all parameters are discussed with focus on their impact on airport operations.

Tables 2.3 and 2.4 show parameters with and without distinct seasonal distribution at Frankfurt Airport. Typically, parameters can be classified into summer-, winter- and all-season parameters. Precipitation, for example, is found throughout the whole year. On average, Frankfurt Airport experienced 231 precipitation days per year. The highest average monthly precipitation day numbers were found in the winter season, with a maximum of 23 precipitation days in January, on average. However, August experienced 28 precipitation days at maximum and thus 2 days more than January with its maximum of 26 days. When looking at heavy precipitation days, here defined as days with more than 10 mm of total precipitation, and at days with moderate/strong precipitation, no distinct seasonal distribution can be extracted, either. On average, there were almost 15 heavy precipitation days per year. A maximum of 5 days was recorded in August and November, each. Moderate/strong precipitation was observed on 103 days per year, on average, with a maximum of 18 days in March 2001. The mean value for March, however, is only 8.2 days and thus lower than for e.g. February and April.

Strong winds, both at surface and in approach levels, have an impact on airport capacity, as described in Section 2.1.3 (also see Figure A.1). Storm is defined as wind exceeding 20.8 m/s. Hence, a storm gust is a gust satisfying this condition. On average, Frankfurt Airport experienced 5 days per year with storm gusts. Generally, those days were connected to the passage of low-pressure storm systems. Except from September and November, each month experienced such events at least once in the investigation period. A maximum of 3 events was observed in February 2002. As described in Section 2.1.3, 15 respectively 10 kts are critical tailwind limits with regard to the use of RWY18. On average, almost 40 days per year were observed exceeding the 10 kts limit. Following the arrival rate matrix for Frankfurt Airport (A.1), days were analysed, where winds in 3000-5000 ft exceeded defined wind thresholds. The limits were 15, 25 and 35 kts. The 25 kts limit was exceeded on 284 days per year, on average. Generally, more events were found during the winter seasons. The same pattern holds for the 25 and 35 kts thresholds. A limit of 35 kts was exceeded on 69 days per year, on average. January 2005 experienced as much as 17 days exceeding this threshold.

Low visibility conditions generally lead to a significant drop in airport capacity (see Figure A.1). Depending on Runway Visual Range (RVR) and Decision Height (DC), conditions are classified in categories CAT I to CAT III (see Table 2.1). CAT III itself is again subclassified into CAT IIIa to

CAT IIIc. Generally, RVR cannot be directly translated into meteorological visibility and vice versa as RVR is defined in RWY-direction only. Thus, Tables 2.3 and 2.4 give more general visibility information on the occurrence of fog instead. Fog is reported when visibility is reduced to less than 1 km. Frankfurt Airport experienced 16 fog days per year, on average. Most fog days were observed in October. From April to August, fog was a very rare event with one observation per month, at maximum.

Winterly weather conditions such as frost and ground frost, solid or freezing precipitation, glaze and the built-up of a snow cover have an impact on airport operations in many ways. Runways, taxiways and apron have to be cleared of snow and ice such that safe operations are guaranteed. Snowplough and runway treatment operations generally necessitate temporary closures of runways and taxiways and thus lead to disruptions in the scheduled traffic flow. Contamination of aircraft surfaces enforces de- and anti-icing procedures, potentially causing delays. For a detailed analysis of winter weather operations with focus on snow removal and aircraft de-icing operations please refer to RÖHNER (2004), RÖHNER and HAUF (2008) and ICAO (1993). In the investigation period, Frankfurt Airport experienced 25 days per year with snow cover, on average. Numbers lie between 16 and 31 days. The main season for the built-up of a snow cover was from December to March, with a distinct maximum in January. November and April also noticed a few snow cover events. Solid precipitation was observed from November to April with January and February being almost on the same level with an average of roughly 11 snowfall days. The events in June, July and September are connected to hail instead of snow. Glaze was a rather rare event at Frankfurt Airport with an average of roughly 3 events per year. However, variability was high with a minimum of one event and a maximum of 6 events per year. Glaze occurred between December and March with a maximum in January. The reason for glaze can either be water patches that freeze on the ground due to cooling, or freezing precipitation. The latter occurred twice per year, on average, with a minimum of zero occasions and a maximum of 3 occasions. Freezing precipitation was found in December and January, only. Frost and frost at ground were rather common events between November and March. Frost at ground occurred as late as in May and as early as in October. In accordance with long-time climatologies, January observes most frost/frost at ground days with an average number of roughly 18 days. Altogether, there were approximately 70 days per year with frost and 90 days per year with ground frost, on average. Ground frost, i.e. frozen soil, is related to frost at ground. It was observed on 40 days per year on average and between October and April. Most frequent occasions were between December and February with an average of approximately 10 days per month and a maximum of 20 days per month.

The only typical summer parameter among the selected weather param-

eters is *thunderstorm*. Thunderstorms are critical weather events and when they pass the airport directly or occur in the airport TMA, they can cause a significant drop in airport capacity down to a temporary closure of the runways in extreme cases. Within the investigation period, Frankfurt Airport experienced 31 thunderstorm days per year, on average. In 2006, a maximum of 46 days was observed. The main thunderstorm season at Frankfurt Airport was from May till August, with July and August being the peak months. At maximum, 13 thunderstorm days per month were observed in July and August, respectively. December and January experienced no thunderstorm events at all, and also in November and February thunderstorms were very rare with only one occasion per month at maximum.

2.2 Data

In the following sections, data used for punctuality modelling are described, based on the initial raw data via several pre-processing steps down to the point of the final predictor variables, which are prepared for feeding in the model equations.

2.2.1 Punctuality and Operational Data

Punctuality and operational data as reported in the daily airport logs were kindly provided by FRAPORT for 2001-2006 in daily resolution. Punctuality was recorded separately for inbounds and outbounds as well as for total daily movements (see columns 7-9 in Figure 2.8). For further analyses, only total daily punctuality (*TOTP*) was considered. As shown in Section 2.1.4, *TOTP* is highly correlated with *INBP* and *OTBP*.

Operational data was extracted from the daily airport logs. A short example on the structure and information contained in these logs is given in Figure 2.8. For further analyses, the following information was used and translated into predictor variables (given in italic) that are later used for

	Vorkommnisse	Flugbewegungen				Pünktlichkeit			Passagiere *INFO		
		Bahn	INB	OTB	TOT	INB	OTB	TOT	INB	OTB	TOT
...	...	...	...	...	...	...	...	...	...	...	...
26.01.2001		25	641	637	1278	87.21%	84.77%	85.99%	65,268	62,934	128,202
27.01.2001		Mix	558	555	1113	87.63%	87.93%	87.78%	55,803	52,037	107,840
28.01.2001		Mix	572	580	1152	88.29%	89.66%	88.98%	57,469	57,453	114,922
29.01.2001		Mix	609	592	1201	88.51%	87.84%	88.18%	56,463	60,157	116,620
30.01.2001		25	610	620	1230	91.15%	91.29%	91.22%	50,991	56,291	107,282
31.01.2001		07	625	618	1243	84.80%	87.54%	86.16%	55,341	62,644	117,985
01.02.2001		Mix	620	626	1246	77.90%	76.68%	77.29%	55,798	62,385	118,183
02.02.2001	Schneefall	25	642	626	1268	64.95%	57.51%	61.28%	58,597	63,245	121,842
03.02.2001	Schneefall	25	493	499	992	37.32%	26.25%	31.75%	47,595	55,375	102,970
04.02.2001	DFS Computerausfall	25	572	589	1161	61.36%	66.38%	63.91%	58,406	62,743	121,149
05.02.2001	starker Gegenwind; Schlechwetter Nord D	25	604	586	1190	58.11%	75.43%	66.64%	54,430	56,948	111,378
06.02.2001		25	624	632	1256	79.65%	85.92%	82.80%	51,599	55,993	107,592
07.02.2001		25	632	625	1257	78.96%	85.28%	82.10%	54,471	60,328	114,799
08.02.2001		25	625	633	1258	84.00%	86.89%	85.45%	53,585	61,236	114,821
09.02.2001		Mix	635	637	1272	83.78%	87.28%	85.53%	59,200	66,372	125,572
...	...	...	...	...	...	...	...	...	...	...	...

Figure 2.8: Extract from daily log at Frankfurt Airport.

punctuality modelling:

1. special events and incidents, e.g. strikes, system failures (extracted from the 2nd column "Vorkommnisse")
   ⇒ *DayIndex*

2. ATC regulations (extracted from the 2nd column "Vorkommnisse")
   ⇒ *DayIndex*

3. CAT stage > CAT I (extracted from the 2nd column "Vorkommnisse")
   ⇒ *DayIndex*

4. strong upper level winds (extracted from the 2nd column "Vorkommnisse")
   ⇒ *Hoehenwinde*

5. change of runway configuration (extracted from the 3rd column "Bahn")
   ⇒ *Mix*

6. actual traffic (extracted from the 6th column "TOT")
   ⇒ *TOTFL*

DayIndex is a boolean (dichotomous) variable being assigned the value *1* if either special events and/or incidents occurred, special ATC regulations were mentioned in the airport logs or a CAT stage worse than CAT I was declared. *DayIndex* is constructed that way that the three categories it is at maximum built of can separately be ignored at value assignment, thus enabling detailed analyses. The latter feature becomes especially important in the context of punctuality forecasting as only predictable input variables can be used in this respect.

In addition to the information drawn from the airport logs, numbers on scheduled traffic (*TOTFL_scheduled*) were provided by FRAPORT. If in the analyses both actual and scheduled flight movements were used, *TOTFL* was replaced by a new variable *TOTFL_scheduled_minus_TOTFL*, i.e. the difference of scheduled and actual flight movements. This new variable describes the effect of scheduled demand being greater than actual capacity, integrated over the whole day. *TOTFL_scheduled_minus_TOTFL* can both be positive or negative. Positive values are due to cancellations, e.g. because of weather or technical defects, and due to diversions (to other airports than Frankfurt Airport). Cancellations, in that respect, can either mean no movement at all, return from taxiway/runway or even return to airport after take off. Negative values are mostly due to short-term additional flights, diversions to Frankfurt Airport or additional flights on days following days with heavily disturbed operations.

2.2.2 SYNOP Weather Data

Archived weather observation data for Frankfurt Airport was obtained from the German Weather Service (DWD) for 2001-2006, mostly on an hourly basis, encoded on the basis of SYNOP-/FM12 coding. Some parameters were observed and reported on a 6-/12-/24-hour basis only (see Table 2.5). Altogether, 173 parameters describing weather at Frankfurt Airport were provided and initially read in. Since punctuality is modelled at daily level (see Section 2.3), weather data had to be pre-processed before being fed into the model equations. The pre-processing basically comprised three steps:

1. choice of raw data for further processing
2. correction of missing or defective raw data
3. creation of day-representatives

Table 2.5 shows all weather parameters chosen for further processing from the initial set of raw data. Units are already changed for further processing. Column 5 of Table 2.5 indicates the observation frequency per day. Parameters reported only once per day are generally reported at 6 UTC, parameters reported twice per day at 6 and 18 UTC and parameters reported four times per day at 0, 6, 12 and 18 UTC. The last column of Table 2.5 gives the replace index which is defined as follows:

0 = take previous value
2 = take the rounded average of the previous and the following value
3 = take the previous official value
4 = take the average of the previous and the following value
5 = replace with -1
6 = replace with 0

Missing parameter values were replaced according to these replace indices in order to produce complete data matrices for further analysis.

Before the adoption of the replacement rules for missing values as described above, missing values of *TG24* were replaced by the minimum of *TG* (if *TG*-data was available on these days) for the respective day. Missing values of *fx24* were replaced by values calculated from *ff* (if *ff*-data was available on these days) for the respective day using the mean ratio of *fx24* and max(*ff*) calculated for each possible day (both values needed) in the 6-year investigation period.

Starting from this corrected raw dataset, *max*-, *min*- and *mean*-values were calculated for each day and for each non-encoded weather parameter. Regarding precipitation, *RRR* was summed up to a total daily precipitation amount to replace *RR24*, which is the amount of precipitation within the

Table 2.5: Raw weather data chosen for further processing. 1st column: SYNOP code abbreviation, 2nd column: description, 3rd column: physical unit, 4th column: resolution, 5th column: number of observations per day, 6th column: replace index.

code	description	unit	res.	obs.	rep.
h	height of cloud base	m	10 m	24	4
VV	horizontal visibility	m	10 m	24	4
N	total cloud amount	eighth	eighth	24	2
dd	wind direction	deg	10 deg	24	0
ff	wind speed (10 min average)	m/s	0.1 m/s	24	4
fx24	maximum wind gust	m/s	0.1 m/s	1	4
TT	temperature (at 2 m)	°C	0.1 °C	24	4
TD	dew point temperature	°C	0.1 °C	24	4
P0	air pressure at station	hPa	0.1 hPa	24	4
ww	significant weather	coded	/	24	5
Nh	cloud amount, low clouds only	eighth	eighth	24	2
CL	cloud type, low clouds	coded	/	24	0
TE	extreme temperature (min/max)	°C	0.1 °C	2	3
TG	temperature at ground	°C	0.1 °C	24	4
TG24	minimum temperature at ground	°C	0.1 °C	1	4
RRR	precipitation amount	mm	0.1 mm	4	6
RR24	precipitation amount, 24 hours	mm	0.1 mm	1	6
RR1	precipitation amount, 1 hour	mm	0.1 mm	24	6
r1	precipitation duration last hour	min	10 min	24	6
E	state of ground without snow	coded	/	4	5
SH	height of snow	cm	1 cm	4	4
UU	relative humidity	%	1 %	24	4

last 24 hours, reported at 6 UTC. Additional wind parameters were created using the information on runway directions at Frankfurt Airport. Thus, mean head-, tail- and crosswind components were calculated for each runway using the following relationships:

$$\begin{aligned}
\textit{headwind}\ \text{RWYxx} &= \begin{cases} ff \cdot \cos(dd - \text{xx}) & \text{for sign}(\cos(dd - \text{xx})) > 0 \\ 0 & \text{otherwise} \end{cases} \\
\textit{tailwind}\ \text{RWYxx} &= \begin{cases} -ff \cdot \cos(dd - \text{xx}) & \text{for sign}(\cos(dd - \text{xx})) < 0 \\ 0 & \text{otherwise} \end{cases} \\
\textit{crosswind}\ \text{RWYxx} &= \text{abs}(ff \cdot \sin(dd - \text{xx}))
\end{aligned}$$

For the crosswind component it is assumed that crosswind from the left has the same impact as crosswind from the right. Keeping in mind that a change of runway use is initiated when the tailwind component exceeds 5 kts, the tangential wind component, i.e.:

$$\textit{tanwind}\ \text{RWYxx} = \text{abs}(ff \cdot \cos(dd - \text{xx}))$$

was additionally calculated. By doing that, it is implicitly assumed that runway configuration changes are automatically initiated when the threshold value for the tailwind component is exceeded. For an enhanced model version, *xx*, i.e. the physically fixed runway direction, was replaced by the used runway configuration *xx'*, i.e. *07* for arrivals/departures in direction *07* and and *25* for arrivals and departures in direction *25*. Daily airport logs (see Section 2.2.1), however, only contained one entry per day in the format *07*, *25* and *Mix*. That means, for days with a runway mix no decided information on the time of runway configuration change is provided. As *25* is the preferred runway configuration at Frankfurt Airport, *25* was applied for the whole day as an approximation for days with a mix of runway configurations. In order to accomodate the tailwind threshold for departures on RWY18, an additional boolean wind variable *Wt_b2_lim* / *WindInt_N_lim* (the first name is used in runway configuration independent mode, the latter in runway configuration dependent mode) was introduced. It is assigned the value *1*, when the tailwind component for RWY18 (i.e. northwind) exceeds 10 kts.

Variables with only one entry per day such as *fx24*, *TG24* for *RR24* were kept in this form not creating any statistics. The same holds for *TE*, extracting the minimum temperature from the 6 UTC entry and the maximum temperature from the 18 UTC entry.

Encoded variables were translated into boolean-type variables using the encoding tables given in Appendix D.1 to D.3. These variables were assigned the value *1*, if the respective weather event was reported and *0* if not. With regard to significant weather, six groups of weather events were defined:

1. *ww31b*: reduced visibility (codes 28, 41-49))
2. *ww32b*: thunderstorm (codes 13, 17, 29, 91-99)
3. *ww33b*: precipitation (codes 15-17, 20-27, 29, 50-75, 77, 79-99)
4. *ww34b*: solid precipitation (codes 22, 23, 26-27, 68-75, 77, 79 , 83-90, 93-94, 96, 99)
5. *ww35b*: moderate/strong precipitation (codes 52-55, 57, 59, 62-65, 67, 69, 72-75, 81-82, 84, 86, 88, 90-99)
6. *ww36b*: freezing precipitation (codes 24, 56-57, 66-67)

From the CL-group, four variables were created combining cloud types:

1. *CLc1*: cumulus mediocris or congestus (code 2), cumulus and multilevel stratocumulus (code 8)
2. *CLc2*: cumulonimbus (codes 3, 9)
3. *CLc3*: stratocumulus (codes 4, 5)
4. *CLc4*: stratus or cumulus fractus (codes 6, 7)

Regarding the state of ground without snow or measurable ice cover, two groups were formulated:

1. *E1*: ground frozen (code 4)
2. *E2*: glaze on ground (code 5)

As an alternative, boolean variables were additionally translated into non-boolean variables, using information on how often the respective event was reported on a given day. In this regard, a maximum value can be defined which is then assigned only if the respective event was reported at each of the

Table 2.6: Final set of weather variables for feed in the model equations.

variable	**description**	**range**
P0_mean	mean air pressure	950-1050 hPa
VV_min	minimum visibility	50-70,000 m
VV_mean	mean visibility	50-70,000 m
h_min	minimum height of cloud base	0-2500 m
h_mean	mean height of cloud base	0-2500 m
h_max	maximum height of cloud base	0-2500 m
N_mean	mean total cloud amount	0-8 eighths
N_max	maximum total cloud amount	0-8 eighths
CLc1	cumulus clouds observed	0/1
CLc2	cumulonimbus clouds observed	0/1
CLc3	stratocumulus clouds observed	0/1
CLc4	stratus or cumulus fractus clouds observed	0/1
U_max	maximum relative humidity	5-100 %
TT_mean	mean temperature at 2 m	−30-+40°C
Ws_a_mean	mean crosswind RWY07/25	0-40 m/s
Ws_b_mean	mean crosswind RWY18	0-40 m/s
Wt_a1_mean	mean headwind RWY07	0-40 m/s
Wt_a2_mean	mean tailwind RWY07	0-40 m/s
Wt_b1_mean	mean headwind RWY18	0-40 m/s
Wt_b2_mean	mean tailwind RWY18	0-40 m/s
ff_mean	mean wind speed	0-40 m/s
fx24_max	maximum wind gust	0-40 m/s
RRR_mean	precipitation amount 24 hours	0-150 mm
RR1_max	maximum precipitation amount in 1 hour	0-50 mm
r1_mean	mean precipitation duration per hour	0-60 min
E1	ground frozen	0/1
E2	glaze on ground	0/1
SH_max	maximum height of snow	0-100 cm
ww31b	reduced visibility reported	0/1
ww32b	thunderstorm observed	0/1
ww33b	precipitation reported	0/1
ww34b	solid precipitation reported	0/1
ww35b	moderate or strong precipitation reported	0/1
ww36b	freezing precipitation reported	0/1
Wt_b2_lim	tailwind limit of 10 kts exceeded on RWY18	0/1

Table 2.7: Alternative set of wind variables for feed in the model equations.

variable	description	range
Ws_PB_mean	mean crosswind PRS[a]	0-40 m/s
Ws_WB_mean	mean crosswind RWY18	0-40 m/s
Wt_head_PB_mean	mean headwind PRS[a]	0-40 m/s
Wt_tail_PB_mean	mean tailwind PRS[a]	0-40 m/s
Wt_N_mean	mean northwind	0-40 m/s
Wt_S_mean	mean southwind	0-40 m/s
Wt_N_lim	tailwind limit of 10 kts exceeded on RWY18	0/1

[a]Parallel RWY System

24 observation times. A value of *zero* is assigned when there was no event on that day. Interim values are evenly distributed. The results of this enhanced mode are discussed in Section 3.2.4.

At the expense of a considerably larger weather data base, all variables described above were also created on a six-hour basis, thus dividing each day into four blocks (0-5:59 UTC, 6-11:59 UTC, 12-17:59 UTC and 18-23:59 UTC). This approach, as already used by Hansen and Bolic (2001), accommodates the fact that certain weather events may have a different impact depending on the time of day. Results of this enhanced mode are discussed in Section 3.2.9.

After these steps, the obtained weather data base still exhibited a deficiency with regard to their application in a multivariate regression model. Correlation analysis of all variables showed strong correlations among the temperature variables. As multivariate regression models are highly sensitive to multicollinearity, the number of temperature related variables had to be reduced. Using the method of principal component analysis as described in Markovic et al. (2008), the daily mean temperature *TT_mean* was determined to represent best all temperature related variables. In a last step, some variables which were, based on meteorological reasoning, considered to have no obvious impact on punctuality, were a priori removed from the pool of potential predictor variables. The final set of weather variables for feed in the model equations is shown in Table 2.6. One should keep in mind that head- and tailwind components for RWY25 are identical with tail- and headwind components for RWY07, and hence redundant. For the enhanced model version with an application of the used runway configuration *xx'* instead of the physically fixed runway direction *xx*, runway related wind variables were replaced by the variables listed in Table 2.7.

The naming conventions as in Tables 2.6 and 2.7 are applied to all further analyses. With regard to 6-hour block data, endings *_1* to *_4* are appended to parameter names to specify the time reference. For example, cumulus clouds

observed at 8 UTC result in *CLc2_2* being assigned the value *1*. For easy distinguishing, the naming convention is such that *min*, *mean* and *max* terms are additionally replaced by *MIN*, *MEAN* and *MAX* terms. For example, the mean tailwind for RWY07 in the second time block (6-11:59 UTC) is labelled *Wt_a2_MEAN_2*.

2.2.3 AMDAR Wind Data

As an alternative to the boolean variable *Hoehenwinde* (see Section 2.2.1), and according to a recommendation taken from Spehr (2003), AMDAR based wind data was considered as an independent source of information on upper level winds, potentially replacing the information drawn from the manually managed daily airport logs. The AMDAR concept basically comprises the use of meteorological information from aircraft equipped with meteorological measurement systems. In the context of AMDAR, valuable information on the state of the troposphere, complementing remote sensing and radio sounding, is collected during flights through on-board measuring and recording and subsequent VHF data downlink through use of ACARS. Meteorological information is provided for temperature as well as wind speed and direction. Some aircraft even measure humidity, turbulence and icing. For our purpose, the focus is on wind information, only. For more information on AMDAR please refer to WMO (2004), Moninger et al. (2003), Drüe et al. (2008) or Frey (2006).

AMDAR data for investigation was obtained by the German Weather Service (DWD). Unlike SYNOP-weather information, archived AMDAR data was only available from 2003-2006. As the database contained global information, the amount of data provided had to be locally reduced. In a first conditioning step, data was filtered using a geographical window with borders 49.5°N, 50.5°N, 7.7°E and 9.3°E. This window is assumed to represent the area relevant to operations at Frankfurt Airport, or more precisely on arrivals at Frankfurt Airport's parallel runway system RWY07/25. Information then extracted was time and position of observation, i.e. latitude, longitude and altitude (calculated from measured pressure), wind speed and wind direction. In a second filtering step, observations outside the height intervall 3000-5000 ft were removed from the dataset as they are not relevant for approach staggering (see Appendix A). The remaining database contained approximately 450,000 entries for the the 4-year period. On average, there were 309 observations per day as a basis for the calculation of relevant predictor variables. Observations were, however, not evenly distributed over the day. During nighttimes, there were much less or even no observations as there were less flights and thus less aircraft reporting. The final set of predictor variables created from the AMDAR database is shown in Table 2.8. The boolean variables reflect the thresholds as used in the arrival rate matrix

Table 2.8: Final set of upper level wind variables for feed in the model equations.

variable	description	range
max_wind	maximum wind speed	0-60 m/s
max_Tangentialwind_a_head	maximum headwind RWY07	0-60 m/s
max_Tangentialwind_a_tail	maximum tailwind RWY07	0-60 m/s
max_wind_ge15	maximum wind speed $\geq$ 15 kts	0/1
max_wind_ge25	maximum wind speed $\geq$ 25 kts	0/1
max_wind_gt35	maximum wind speed $>$ 35 kts	0/1
max_Tangentialwind_a_head_ge15	maximum headwind RWY07 $\geq$ 15 kts	0/1
max_Tangentialwind_a_head_ge25	maximum headwind RWY07 $\geq$ 25 kts	0/1
max_Tangentialwind_a_head_gt35	maximum headwind RWY07 $>$ 35 kts	0/1
max_Tangentialwind_a_tail_ge15	maximum tailwind RWY07 $\geq$ 15 kts	0/1
max_Tangentialwind_a_tail_ge25	maximum tailwind RWY07 $\geq$ 25 kts	0/1
max_Tangentialwind_a_tail_gt35	maximum tailwind RWY07 $>$ 35 kts	0/1

(see Appendix A). Again, one should keep in mind that head- and tailwind components for RWY25 are identical with tail- and headwind components for RWY07 and hence redundant.

Like with the predictor variables created from the SYNOP database for an enhanced model stage, boolean variables were additionally translated into non-boolean variables, using information on how often the respective event was reported on a given day. The procedure is as described on page 35. Also, 6-hour block variables were created for an enhanced model stage as described in Section 2.2.2. For upper level wind variables the naming convention is such that suffixes *_1* to *_4* are appended, representing the four different time blocks. In that respect it should be mentioned that for the first time block (0-5:59 UTC), there is less data (on average 34 observations) available than for the other time blocks (on average 85-100 observations per block) and the data basis thus is very sparse. For 25 days, there was even no information available. Values were then linearly interpolated using data from the previous and the following day.

2.2.3.1 AMDAR Data versus Daily Logs: A short Analysis

Within this short section, the quality of AMDAR data is discussed with focus on its potential to replace information on upper level winds drawn from the daily airport logs. The advantage of AMDAR data as compared to the variable *Hoehenwinde*, drawn from the daily airport logs, clearly is that it is independent in that sense that it is not prestressed by a manual selection process deciding for or against putting a note in the daily logs. Additionally, AMDAR based predictor variables are created in such a way that prognostic input can potentially be drawn from numerical weather forecasting models (NWFM), possibly upgraded by model output statistics (MOS), and thus be used in a punctuality forecasting system. A disadvantage of AMDAR based

variables is the low number of observations it is based on. Compared to the daily number of movements at Frankfurt Airport, the number of AMDAR messages is low, considering that the average number of roughly 300 messages per day comes from only a fraction of aircraft arriving at or departing from Frankfurt Airport. This due to the fact that only a limited number of AMDAR equipped aircraft is chosen for further processing of meteorological data, satisfying special DWD needs with regard to data quality as well as spacial and temporal resolution. As an alternative to AMDAR data, wind profiler soundings could be used to obtain information on upper level winds.

In the following, AMDAR based variables are analysed and compared to information drawn from the daily airport logs. It is assumed that entries in the daily logs, i.e. a note on the occurrence of strong upper level winds, reflect the true state of the atmosphere with strong upper level winds prevalent on the respective day. When there is no log entry for a certain day, it is assumed that no strong upper level winds were prevalent. Of course it is difficult to say which data source really tells the truth – airport logs or AMDAR. However, entries in the daily logs are a rather reliable indicator that strong upper level winds in fact had a measurable impact on operations, be it through pilot reports or attestable interventions by local ATC.

It is now analysed how well variables drawn from the AMDAR data base do reflect days with and without strong upper level winds, provided that daily logs tell the truth. For investigation, the AMDAR based variables *max_wind* and *max_wind_gt35kt* were consulted. As quality measures, the probability of detection (POD), the false alarm rate (FAR), sometimes also referred to as false alarm ratio, and the total hit rate (THR) were used as defined in Wilks (1995). Using the following notation:

H:	Hit	$\Rightarrow$ here: log entry & AMDAR signal
F:	False	$\Rightarrow$ here: no log entry & AMDAR signal
M:	Missed	$\Rightarrow$ here: log entry & no AMDAR signal
Z:	Zero	$\Rightarrow$ here: no log entry & no AMDAR signal
N:	Total number of observation days ($H + F + M + Z$)	

, the measures are calculated as:

$$POD = \frac{H}{H + M} \tag{2.1}$$

$$FAR = \frac{F}{H + F} \tag{2.2}$$

$$THR = \frac{H + Z}{N} \tag{2.3}$$

Here, POD is the fraction of days, when AMDAR based wind variables indicated strong upper level winds and airport logs supported these events. Given

that log entries tell the truth, a POD value of *1* would be the best and a POD value of zero the worst scenario. FAR, by contrast, is the fraction of days, when AMDAR based wind variables indicated strong upper level winds, but airport logs did not support this claim. Optimal FAR values would thus be close to zero. THR is simply the fraction of events or nonevents supported by both daily logs and AMDAR based wind variables. A high THR is to be interpreted carefully as it credits H and Z equally and is thus strongly influenced by the more common category (ECMWF, 2007).

Using *max_wind_gt35kt* for the analysis, the POD as defined above is 0.70. That means, in 70 % of all cases when there was an entry in the daily logs on strong upper level winds, *max_wind_gt35kt* also gave a positive signal. THR is at 0.80. However, the high FAR value of 0.64 points out that upper level winds exceeding the 35 kts threshold do not necessarily result in operations being disturbed to an extent that it is to be reported in the daily logs. Obviously, operations can, under certain circumstances, also run rather smoothly when strong upper level winds are prevalent. It is, in that context, not assured that daily logs are exhaustive. It is thinkable that log incompleteness is to be blamed for the relatively bad agreement of the two databases, either due to unconsidered combinations of weather events beeing responsible for operation disruptions (e.g. high wind and low visibility) or simply due to heterogeneous log accounting.

Using *max_wind* as AMDAR based variable and screwing up the threshold to wind speeds larger than 35 kts, generally results in FAR values decreasing, however, at the expense of deteriorating POD values. Screwing down the threshold wind speed gives the opposite effects. For example, using a threshold wind speed of 15 kts results in a POD of more than 0.99, bought dearly with a FAR of 0.86. The findings of this analysis are discussed in Section 3.2.5 in the context of punctuality modelling.

2.3 Theoretical Approach

Within this section, the methods used for punctuality modelling are introduced. In that respect, the main focus is not on a detailed description of the mathematical methods themselves, as they are well known and illustrated in many textbooks, but on the successive composition of the hybrid model with focus on configuration and internal interaction.

2.3.1 Multivariate Linear Regression

Multivariate linear regression constitutes the core of the punctuality modelling approach. The method shall only be shortly described to make the reader familiar with the mathematical background. For a deeper insight in

multivariate regression, refer to e.g. Backhaus et al. (2003), Weisberg (1985) or Wilks (1995).

In our case of modelling total daily punctuality $\boldsymbol{TOTP}$, i.e. determining an estimate $\widehat{\boldsymbol{TOTP}}$, a simple multivariate linear regression model is described by the following equation:

$$\widehat{\boldsymbol{TOTP}} = \boldsymbol{X} \cdot \hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon} \quad , \tag{2.4}$$

with $\boldsymbol{X}$ being the $n \times (k+1)$ matrix of predictors as described in the previous sections (often referred to as "design matrix"), having k different predictor variables arranged in columns and n observations of these predictors arranged in rows, and $\hat{\boldsymbol{\beta}}$ being the $(k+1) \times 1$ vector of predictor coefficients to be determined. The matrix of predictors initially contains a first column of ones, representing the constant term, to accommodate the variation of $\boldsymbol{TOTP}$ around its mean value. Without good reason for doing so, this column of ones should not be removed from the model equation. The $n \times 1$ vector $\boldsymbol{\epsilon}$ represents the residuals $\boldsymbol{TOTP} - \widehat{\boldsymbol{TOTP}}$ for each observation, i.e. the difference between observed and estimated punctualities. Using ordinary least squares (OLS), $\hat{\boldsymbol{\beta}}$ is determined by:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{TOTP} \tag{2.5}$$

This OLS estimate minimizes the residual sum of squares $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}$. The difficulty now is to find the right set of predictor variables out of the pool of potential predictors. For reduction in the number of predictors, backward selection, respectively backward elimination was applied as described in Weisberg (1985), consulting t-statistics and F-statistics. Under the null-hypothesis H_0: $\beta_i = 0$, i.e. that the coefficient for predictor variable i, is zero, t-values (non-boolean variables) respectively F-values (boolean variables) were computed for each predictor variable. t-values respectively F-values could then be translated into p-values, giving the conditional probability of observing t-/F-values as large or even larger than the observed value, given that the H_0 is true. In that respect, it should be kept in mind that large t-/F-values would speak for a (in this case wrong) rejection of the null-hypothesis. Thus, small p-values provide evidence against the null-hypothesis H_0: $\beta_i = 0$ (Weisberg, 1985).

The backward elimination procedure starts with the complete set of predictors with initial coefficients β_i, stepwise eliminating the variable with the highest p-value associated with its t- respectively F-value. After each removal step, the β-coefficients are again estimated on the basis of the new model. The elimination procedure stops when all variables left in the model exhibit associated coefficient p-values smaller than a fixed significance level α. Generally used are significance levels of $\alpha = 0.05$ or $\alpha = 0.01$.

When the set of relevant predictors is found, hence the final design matrix $\boldsymbol{X}_{final}$ is determined, the significance of the relation between the set of predictor variables and the predictant ($TOTP$) is to be tested globally, i.e. it is tested if the model is statistically significant and thus generally valid. In that respect it is tested if all β_i are non-zero. Follwing Backhaus et al. (2003), the null-hypothesis claims that all β_i are zero. The empirical F-value is calculated as:

$$F_{emp} = \frac{R^2/k}{(1-R^2)/(N-(k+1))} \quad , \tag{2.6}$$

with N being the number of observations, k being the number of independent variables and k+1 being the number of coefficients to be determined (i.e. the number of predictants counting the constant). The empirical F-value can either be compared with F-tables giving the theoretical F-values for a given significance level or it can again be translated into a p-value. For example, a p-value of 0.05 means that the level of significance is 5 %, i.e. that in 5 % of all cases the null-hypothesis is misleadingly rejected (and thus a statistically significant model is assumed) when the null-hypothesis is actually true.

2.3.2 Regression Trees

This section is dedicated to regression trees as introduced by Breiman et al. (1984). The name *regression tree* originates from its visualisation using a tree structure with branches representing bifurcations and terminal nodes, called leaves, representing final value assignment. From a mathematical point, regression trees are simply a set of decision rules obtained by an iterative error minimizing procedure. Regression trees allow for the use of numerical as well as categorical variables in the matrix of predictors. They are easy to interpret, both for analysis and prediction purposes. For illustration, Figure 2.9 shows an example of a regression tree. In this example, moving down the tree, one exemplarily obtains an average $TOTP$ of 0.76 if snow height is larger than 5 cm, visibility higher than 1 km and wind speed smaller than 15 kts.

In the following, the CART-algorithm developed by Breiman et al. (1984) is shortly described. According to this work, which was groundbreaking in the field of regression tree modelling, there are three elements necessary to construct a regression tree:

1. a way to select a split at every intermediate node
2. a rule for determining when a node is terminal, i.e. when it becomes a leaf
3. a rule for assigning a value $\hat{y}$ to every terminal node

The difficulty in the construction of regression trees is to find the right set of splits, i.e. splits which most successfully separate the high predictant values

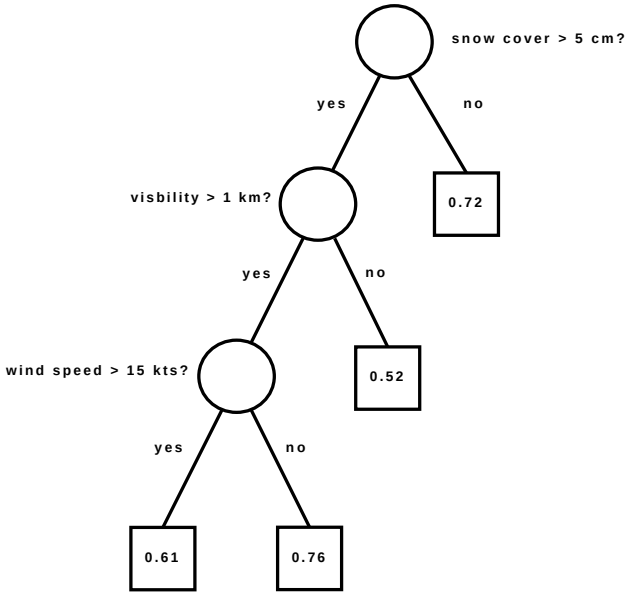


Figure 2.9: A simple example of a regression tree using weather variables as predictors and *TOTP* as predictant. Leaves are labelled with assigned punctualities, which are the average of punctualities for the respective leaf.

from the low ones, thus minimizing the total regression error rate. In that respect, a classical measure of accuracy for regression trees is the resubstitution estimate of the tree T:

$$R(T) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{t\in\ T}\ \sum_{i\,|\,\boldsymbol{X}_i\in\ t}(y_i - \bar{y}(t))^2 \tag{2.7}$$

The second alternative in Equation 2.7 makes uses of the definition:

$$\bar{y}(t) = \frac{1}{n(t)}\sum_{i\,|\,\boldsymbol{X}_i\in\ t} y_i \quad , \tag{2.8}$$

where t denotes a terminal node t and $\bar{y}$ is the average value of all cases falling into that node t. Accordingly, the resubstitution estimate for any node t is:

$$R(t) = \frac{1}{n}\sum_{i\,|\,\boldsymbol{X}_i\in\ t}(y_i - \bar{y}(t))^2 \tag{2.9}$$

Using this definition, the best split s^* of all splits s of a node t into a left node t_L and a right node t_R is the one maximizing:

$$\Delta R(s^*, t) = \max_{s\,\in\,S}\Delta R(s, t) \quad , \tag{2.10}$$

with S representing all possible splits and $\Delta R(s, t)$ defined as:

$$\Delta R(s, t) = R(t) - R(t_L) - R(t_R) \tag{2.11}$$

The splitting works such that at each node the tree algorithm searches through the set of predictor variables, starting with the first one and continuing up the the last one, finding the best split for each predictor. Out of the best split for each predictor, it finally selects the overall best split. The splitting procedure as described above continues until each terminal node contains at maximum N_{max} cases. BREIMAN et al. (1984) propose to initially grow large trees – at maximum to a point where terminal nodes are either pure or contain only one single case. We followed the latter approach.

Trees grown applying the above algorithm can become rather large. In the context of generalisability, this is of course not appropriate. Such trees are heavily overfitted and just able to reproduce their learning sample, albeit perfectly, but at the expense of poor performance when applied to independent data. Hence, it is inevitable to prune trees down to a point where predictability balances misclassification, respectively error rate. BREIMAN et al. (1984), therefore, introduced a minimal error-complexity pruning procedure. Introducing a complexity parameter α, error-complexity is then defined as:

$$R_\alpha(T) = R(T) + \alpha\,|\tilde{T}| \quad , \tag{2.12}$$

with $\tilde{T}$ being the set of terminal nodes and $|\tilde{T}|$ being the number of terminal nodes of the tree T. R_α thus is burdened with a loading depending on the size of the tree. The higher α becomes, the more the loading and the smaller the minimal error-complexity tree to be picked. For a detailed description of the whole pruning procedure refer to BREIMAN et al. (1984). For our purpose, it is important to note that error-complexity pruning starts with the maximum grown tree, successively pruning branches to terminal nodes on the basis of v-fold (generally 10-fold) cross-validation. A major reason for growing a maximum tree and then pruning it again, especially as an alternative to growing smaller trees with more cases per terminal node, at the outset, is that there may be nodes where splitting only leads to a small decrease of error rate, but where splits of descendant nodes might offer a significant decrease of error rate. Hence, by stopping the splitting procedure too early, potentially good splits might be unused.

In the present studies, basic construction of regression trees was accomplished using the Matlab Statistics toolbox algorithm *classregtree*, which is based on mathematical concepts formulated by BREIMAN et al. (1984). Pruning of maximum grown trees was done with Matlab's *test* and *prune* algorithms, forcing for a fixed final number of terminal nodes. This was necessary since it turned out that only using the built-in crossvalidation procedures led to extremely unstable pruning results with a maximum difference in proposed pruning steps of up to 15 steps, heavily depending on the random choice of samples used for crossvalidation. This would result in trees of varying size and leaf number. In terms of stability and reconstructability of trees, this was not acceptable. The use of a fixed final number of terminal nodes is thus

recommended. A limit of 28 terminal nodes proved to be most efficient and was thus used for tree pruning.

2.3.3 AR-Processes

AR models exploit time series information on variable autocorrelation to extrapolate a variable of interest. Following e.g. BAMBERG and BAUR (1998) or HIPEL and MCLEOD (1994), AR(p)-processes are described by the equation:

$$y_t - \mu = \Phi_1 (y_{t-1} - \mu) + \Phi_2 (y_{t-2} - \mu) + \ldots + \Phi_p (y_{t-p} - \mu) + \xi_t \quad , \quad (2.13)$$

with y being the variable of interest at time t, μ being the mean of y, Φ_i being AR-parameters to be determined and ξ_t being white noise at time t. In an enhanced model stage (see Section 3.2.11), the linear regression model is upgraded using an additional AR(1) modelling term. This combined model is referred to as linear statistical model with first-order autoregressive error correction (JUDGE et al., 1988).

In our case, not the dependent variable $TOTP$ itself but the error term $\boldsymbol{\epsilon}$, i.e. $\boldsymbol{TOTP - \widehat{TOTP}}$, is modelled using an AR model. In the following, we consider an AR(1) processes only, as we postulate that $TOTP$ on day i is only affected by $TOTP$ on day $i-1$. This postulation is supported by the autocorrelation of $TOTP$ as shown in Figure 2.7 and by the rationale given in Section 2.1.4. Keeping in mind that $\bar{\epsilon}$ is assumed to equal zero, equation 2.13 hence translates into:

$$\epsilon_t = \rho\, \epsilon_{t-1} + \xi_t \quad , \quad (2.14)$$

with $\rho = \Phi_1$ being the lag-1 autocorrelation coefficient of ϵ. The regression model with first-order autoregressive error then becomes:

$$\widehat{TOTP}_t = \boldsymbol{X}_t \cdot \hat{\boldsymbol{\beta}} + \rho\, \epsilon_{t-1} + \xi_t \quad (2.15)$$

2.3.4 Model Quality Measures

In the following, quality measures used for later evaluation of different model setups are introduced. Again, estimated values are labelled by a hat.

A commonly used quality measure in regression modelling is the coefficient of determination. It is defined as:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (2.16)$$

The response variable y is in our case given by $TOTP$. In the first case, R^2 is defined as explained variability to total variability, in the second case as one minus non-explained variability to total variability. An alternative definition

(see JUDGE et al., 1988) of the coefficient of determination is via the square of the multiple correlation coefficient r:

$$R^2 = r^2 = \left(\frac{\sum_{i=1}^{n} (y_i - \bar{y}) \left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} \left(\hat{y}_i - \bar{\hat{y}}\right)^2}} \right)^2 \tag{2.17}$$

R^2 thus always ranges between 0 and 1, with values close to 1 indicating a nearly perfect model and values close to 0 indicating a rather useless model. The latter definition of R^2 is preferable to Equation (2.16) as it gives correct R^2 values both for calibration and independent data. In order to accommodate the fact that adding more and more predictors to the model equations results in non-decreasing diagnostic R^2 values, a corrected R^2, called *adjusted* R^2, was introduced according to JUDGE et al. (1988):

$$R^2_{adj} = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-(k+1)} \quad , \tag{2.18}$$

with k+1 being the number of β-coefficients to be determined, including the one for the constant. Additional quality measures used are the mean absolute model error *MAE*:

$$MAE = \text{mean}(|y - \hat{y}|) \quad , \tag{2.19}$$

which gives the average deviation of the predicted values $\hat{y}$ from the true values y, the root mean square error *RMSE*:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2.20}$$

and the standard error of regression (see BACKHAUS et al., 2003):

$$SE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-(k+1)}} \tag{2.21}$$

which is closely related to RMSE and only scaled differently, using n and k as defined above. Both RMSE and SE are good measures of accuracy. They differ from MAE in that sense that outliers are more emphasised through the squaring of residuals.

2.3.5 Hard- and Software Environment for Implementation

Punctuality modelling as described in the previous sections was done on a Windows XP Intel Pentium IV system using Matlab R2008b, version

7.7.0.471 as programming environment. Reading and pre-processing of AMDAR data was done on a Linux (openSUSE 10.2-64 / SLES 9) dual core Intel Xeon machine using Fortran 90/95 and shell scripts. The set up of the AR-module and the implementation of the MLR-module into a Matlab-environment was done by Danijela Markovic in the frame of the European FLYSAFE project (see also Markovic et al., 2008; Theusner and Röhner, 2008).

Chapter 3

Results

In this chapter the results of the punctuality modelling are presented, starting from a baseline model to the point of the final hybrid model, detailing several enhancement steps. Mathematical methods applied are described in Section 2.3.

3.1 Preliminary Investigations

Before discussing the model results, a preliminary analysis of days with low punctualities is presented in this section. These investigations are later consulted for the construction of regression trees, which constitute an important component in the enhanced punctuality models introduced in Section 3.2.12.

As will be shown in Section 3.2, punctuality modelling using multivariate linear regression, optionally enhanced by an AR(1)-model, generally gives good results for days with punctualities higher than roughly 0.50. Days with a punctuality less than 0.50 are in the following referred to as *low punctuality days*. For general understanding and later construction of special regression trees, an analysis of prevalent weather on low punctuality days was done. Roughly 80 weather related criteria were created using the weather variables introduced in Sections 2.2.2 and 2.2.3. As an example, one of these criteria is e.g. related to a special combination of solid precipitation and sub-zero temperatures. A question to be posed would, accordingly, be. "Was solid precipitation observed and the mean temperature below 0°C on day x?" Cross table 3.1 presents the results of the analysis. Entries in the table are ordered by the average $TOTP$ on eventdays (2^{nd} column), i.e. days where criterion xy was fulfilled. The first column lists the criteria, which were chosen based on operational thresholds and meteorological experience, using the notation introduced in Sections 2.2.2 and 2.2.3. It is complemented by wind variables *Wt_x_mean* and *max_Tangentialwind_x_absolute_**, representing tangential winds in x-direction (i.e. not distinguishing between head- and tailwind), where "* " represents one of the suffixes introduced.

Table 3.1: Prevalent weather on low punctuality days. Notations introduced in Sections 2.2.2 and 2.2.3 are used. The second column gives the average *TOTP* on the "cases" (first column) eventdays. Column "% cases" gives the fraction of eventdays with $TOTP<0.5$ to the total number of eventdays. Column "% l-cases" gives the fraction of eventdays with $TOTP<0.5$ to the total number of days with $TOTP<0.5$.

weather criterion	**cases**	∅ ***TOTP***	**% cases**	**% l-cases**
h_mean<30 m	3	0.39	100.00	3.16
SH_max>8 cm	14	0.45	50.00	7.37
VV_mean<2000 m	5	0.45	60.00	3.16
Wt_b2_mean/Wt_N_mean>5 m/s	20	0.52	55.00	11.58
ww31b>0 & max_wind_gt35>0	13	0.53	46.15	8.57
Ws_a_mean/Ws_PB_mean>5 m/s	29	0.56	37.93	11.58
Wt_a1_mean>5 m/s	2	0.58	0.00	0.00
fx24_max>25 m/s	4	0.58	50.00	2.11
VV_min<300 m & h_min<30 m	20	0.58	25.00	5.26
VV_min<300 m	21	0.59	23.81	5.26
ff_mean>7 m/s & max_wind>25 m/s	27	0.59	29.63	11.43
ww36b>0 & TT_mean<0°C	8	0.60	25.00	2.11
VV_min<300 m & max_wind_gt35>0	3	0.60	33.33	1.43
ww34b>0 & TT_mean<0°C	98	0.62	25.51	26.32
ww36b>0	12	0.63	16.67	2.11
max_Tangentialwind_a_head_gt35>0	14	0.63	21.43	4.29
ff_mean>7 m/s	60	0.63	21.67	13.68
ww31b>0 & h_min<30 m	84	0.63	17.86	15.79
Wt_b_mean>5 m/s	121	0.63	20.66	26.32
ww31b>0	91	0.64	17.58	16.84
r1_mean>45 min	19	0.65	26.32	5.26
Wt_b1_mean/Wt_S_mean>5 m/s	98	0.66	14.29	14.74
max_Tangentialwind_a_tail>25 m/s	80	0.66	12.50	14.29
max_Tangentialwind_a_absolut>25 m/s	81	0.66	12.35	14.29
max_wind>25 m/s	109	0.66	14.68	22.86
ww34b>0	236	0.66	16.10	40.00
max_Tangentialwind_b_head_gt35>0	28	0.67	17.86	7.14
max_Tangentialwind_b_absolut_gt35>0	58	0.67	15.52	12.86
E2>0	17	0.68	17.65	3.16
RR1_max>5 mm	51	0.68	11.76	6.32
max_Tangentialwind_b_tail_gt35>0	30	0.68	13.33	5.71
TT_mean<0°C	185	0.68	16.22	31.58
max_Tangentialwind_a_absolut_gt35>0	276	0.68	13.04	51.43
h_min<30 m	168	0.69	13.10	23.16
max_Tangentialwind_a_tail_gt35>0	262	0.69	12.60	47.14
Wt_a_mean>5 m/s	96	0.69	12.50	12.63
Wt_head_PB_mean>5 m/s	96	0.69	12.50	12.63
Wt_a2_mean>5 m/s	94	0.69	12.77	12.63
TG_mean<0°C	230	0.69	14.35	34.74
Ws_b_mean/Ws_WB_mean>5 m/s	48	0.69	12.50	6.32
max_wind_gt35>0	363	0.70	11.29	58.57
P0_mean<990 hPa	125	0.70	7.20	9.47
max_Tangentialwind_b_tail_ge25>0	151	0.70	9.93	21.43

to be continued on next page

continued from last page

Wt_b2_lim/Wt_N_lim>0	230	0.71	10.00	24.21
max_Tangentialwind_a_head_ge25>0	120	0.71	8.33	14.29
ww35b>0	617	0.71	8.75	56.84
max_Tangentialwind_a_absolut_ge25>0	623	0.71	8.67	77.14
max_Tangentialwind_a_tail_ge25>0	506	0.71	8.70	62.86
TT_min<0°C	413	0.72	10.17	44.21
max_Tangentialwind_b_absolut_ge25>0	325	0.72	8.00	37.14
max_wind_ge25>0	871	0.73	7.23	90.00
ww32b>0	186	0.73	6.45	12.63
RRR>0 mm	1019	0.73	6.48	69.47
E1>0	244	0.73	8.61	22.11
max_Tangentialwind_b_head_ge25>0	183	0.73	7.65	20.00
max_Tangentialwind_b_tail_ge15>0	514	0.73	7.39	54.29
TG_min<0°C	542	0.74	8.30	47.37
CLc2>0	433	0.74	4.62	21.05
max_Tangentialwind_a_tail_ge15>0	800	0.74	6.63	75.71
TG24_min<0°C	605	0.74	7.60	48.42
ww33b>0	1387	0.74	5.84	85.26
max_Tangentialwind_b_absolut_ge15>0	982	0.74	6.11	85.71
max_Tangentialwind_a_absolut_ge15>0	1136	0.74	5.99	97.14
max_Tangentialwind_b_head_ge15>0	566	0.74	5.48	44.29
max_Tangentialwind_a_head_ge15>0	378	0.75	4.50	24.29
max_wind_ge15>0	1325	0.75	5.21	98.57
max_Tangentialwind_b_head>25 m/s	2	0.78	0.00	0.00
max_Tangentialwind_b_absolut>25 m/s	6	0.79	0.00	0.00
max_Tangentialwind_b_tail>25 m/s	4	0.80	0.00	0.00
ww31b>0 & Wt_b2_lim>0	2	0.81	0.00	0.00
max_Tangentialwind_a_head>25 m/s	1	0.87	0.00	0.00

end of table

Not all of the criteria listed shall be discussed in detail. Generally, looking at average punctualities connected to the events, it shows that there are only few events with an average *TOTP* below 0.5, giving strong evidence that they are at least involved in processes generating delays. However, many of the events with related average *TOTPs* larger than 0.5 might later be used to enhance the punctuality models by applying regression trees, building on these sub-datasets. Table 3.1 can be seen from two sides. Looking at column "% cases", one can see, how many eventdays in fact had a *TOTP*-value lower 0.5, compared to the total number of eventdays. High values indicate that the respective events are rather pure in the sense of low punctualities. For example, using a daily mean ceiling lower than 30 m as a criterion, all cases observed implied a *TOTP* of less than 0.5. As one might expect, also snow cover (*SH_max*>8 cm), low visibility conditions (*VV_mean*<2000 m, *VV_min*<300 m), strong winds (*Wt_b2_mean*/ *Wt_N_mean*>5 m/s, *Ws_a_mean*/ *Ws_PB_mean*>5 m/s, *fx24_max*>25 m/s) or combinations of low ceiling and visibility, visibility and strong winds or solid/freezing precipitation at

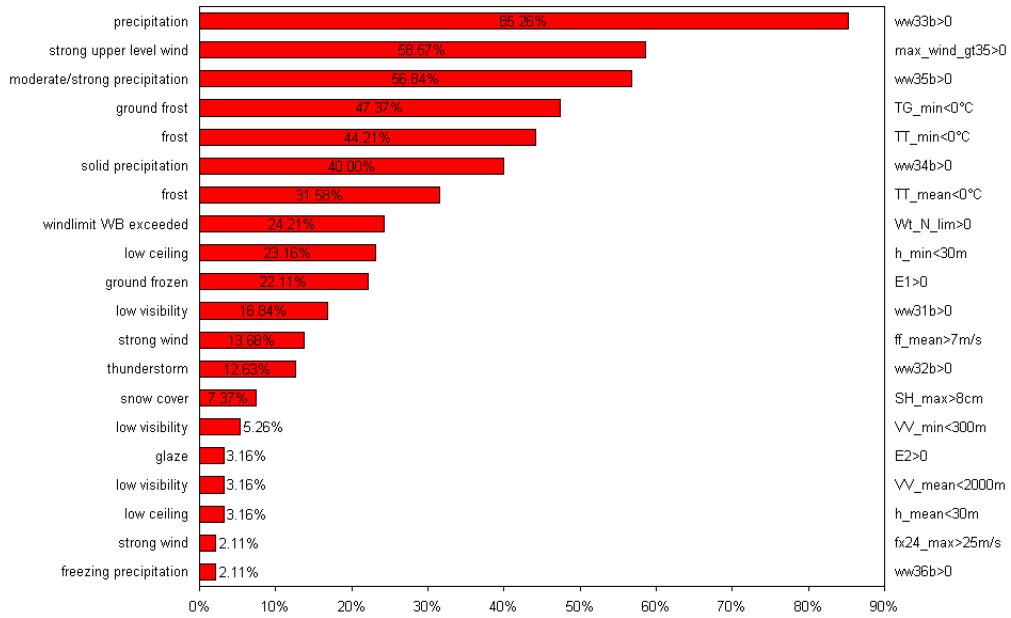


Figure 3.1: Prevalent weather on low punctuality days. The x-axis gives the percental fraction of eventdays with $TOTP<0.5$ to the total number of days with $TOTP<0.5$. The left-hand labelling gives the general weather category, the right-hand labelling the exact criterion to be fulfilled using the proposed notation.

sub zero temperatures have a strong impact.

The other view on Table 3.1 is when focus is on column "% l-cases". This column tells something about the weather prevalent on low punctuality days. For example, when again looking at days with a mean ceiling lower than 30 m, the value of 3.16 % means that 3.16 % of the 95 days with $TOTP<0.5$ (in 2001-2006) fulfilled the criterion "$h\_mean<30$ m". The values given in column "% l-cases" should be interpreted carefully and only in combination with column "% cases". Solely on the basis of the last column, one should not conclude that high values with respect to a certain event are an indicator for a strong impact on punctuality. Precipitation ($ww33b>0$), for example, exhibits an l-case value of 85,26 %, which means that on far more than three quarters of all days with $TOTP<0.5$ this criterion was fullfilled. However, on $1387 - 81 = 1306$ precipitation days, total punctuality was larger than 0.5. Thus, "$ww33b>0$" is fullfilled on more than 50 % of all days. This other side of the coin is reflected in a low "% cases" value of 5.85. Figure 3.1 gives a summarizing overview over a selection of basic weather events on low punctuality days. Again, it shows that precipitation is almost always present on low punctuality days, followed by strong winds and frost/ground frost. The reason why other, intuitively obvious impact events, such as thunderstorms or low visibility, only exhibit low "l-cases" values, is that they are simply rather seldom, compared to a general precipitation event.

3.2 Modelling Results

In the following sections, results from different model stages and setups are presented. Different model stages are in the following referred to as *Model x*. The work approach is such that each model stage is shortly introduced, based on the previous model setup where applicable. Results are discussed with special focus on performance on independent data. An overview table of model quality criteria for different model stages is given in Appendix B.3. A difficulty in terms of comparability of different model setups is that using the backward elimination scheme as described in Section 2.3.1 with different sets of potential predictor variables generally leads to a different choice of final predictors. Most notably, the number of final predictor variables is likely to vary, as well.

In order to make model results from different model stages comparable, a fixed set of predictor variables is determined for each model setup, setting a limit of 20 predictors. By claiming this requirement, another procedural deficiency is remedied. Only using one model run for the determination of the final set of predictors, one runs the risk of not finding the optimal set of predictors, but one that just works best for the calibration period. However, only changing the calibration period slightly may result in heavily differing predictor choices. Hence, as focus is on good model performance on independent data, it is essential to determine a quasi-fixed set of global predictor variables in order to increase model generalisability. Quasi-fixed is meant in the sense of "fixed for each model setup".

A set of predictors is in the following determined for each model setup, moving a window of 1 year, 3 years and 5 years over the 6-year data set. At each window step, the set of predictors is determined using the backward selection scheme with a requested significance level of $\alpha = 0.05$, as described above. Counting the number of selections for each potential predictor and calculating the respective percentage as compared to the maximum possible selections (i.e. for e.g. the 1-year window 1827 selections are at maximum possible in a 6-year period) generates a ranking among the predictors. The

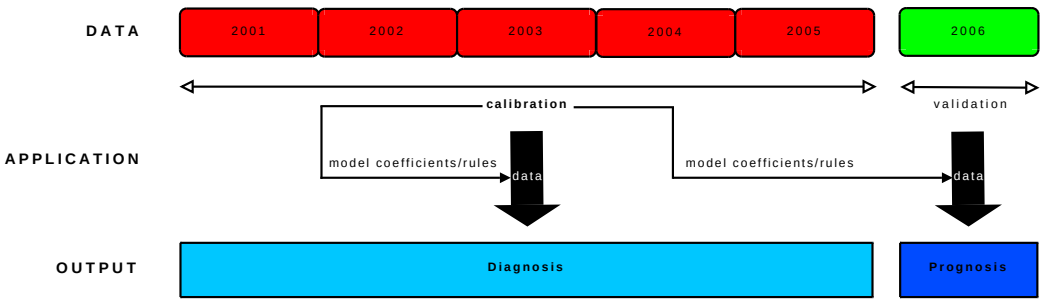


Figure 3.2: Process flow for model calibration and validation.

three rankings from the three different window lengths are finally merged and used to determine the 20 most selected predictors, which then constitute the fixed set of predictor variables for the respective model. This set of predictors is finally used to determine the diagnostic as well as the prognostic model performance. The terms " diagnosis" and "diagnostic" are in the following used for model application with calibration data. When a model is applied to independent data, the terms "prognosis" and "prognostic" are used, accordingly. This will, in the following, also be referred to as "validation". The term "forecast" is, in the present study, used in the context of a punctuality forecast for a day in the future. Punctuality forecast is discussed in Section 3.3. For general model development and visualisation of model improvements, data for the period 2001-2005 is used for model calibration. The year 2006 is reserved for model validation. Figure 3.2 illustrates the calibration and validation process flow.

3.2.1 Model 1 – Rudimentary Baseline Model

In this section, a rudimentary pre-baseline model is introduced. This model uses weather variables from ground observations, as described in Section 2.2.2, only. The mathematical basis is pure multivariate linear regression with normalised predictors. As described in Section 3.2, the calibration period is 2001-2005 and the year 2006 is used for validation. This model represents the minimum model and is only to be used as a reference. Interpretation of model coefficients is not advisable as important predictor variables are intentionally omitted and model coefficients are thus biased, bearing the weight of non-included predictors.

Table 3.2 lists the fixed set of predictors for this model, i.e. the 20 predictors chosen using the procedure described in Section 3.2. The numbers in brackets are the percentage of selections within the 1-year, 3-year and 5-year windows, respectively. The numbers given in bolt represent the maximum value from the three windows.

Table 3.3 compares experimental diagnostic model results, using the moving 1-year, 3-year and 5-year windows. Results are given for two modes: 1) automatic model calibration at each time step, using the backward selection scheme, 2) using the fixed set of predictors summarised in Table 3.2. In the first mode, there are large deviations in the number of selected predictors. Taking the 1-year window, the range is from 6 to 18 variables. Also R^2_{adj} exhibits strong variability with values between 0.274 and 0.555. Looking at the 3- and 5-year windows, one recognises a clear tendency of decreasing variability, i.e. both R^2_{adj} values and the number of predictors selected exhibit a more narrow range. This indicates that longer calibration periods generally result in more stable modelling results. The second important point is that using the fixed set of predictors leads to similar R^2_{adj} values as in automatic

Table 3.2: Fixed set of predictors for Model 1. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	CLc2 (32.4, 81.6, **100**)
ww31b (97.9, **100**, **100**)	Wt_b2_lim (27.9, 44.4, **100**)
SH_max (95.6, 99.7, **100**)	N_mean (7.1, 53.0, **99.2**)
E1 (75.0, **100**, **100**)	CLc1 (17.6, 41.7, **98.1**)
r1_mean (63.0, 90.4, **100**)	ww32b (29.6, 27.9, **93.7**)
Wt_a1_mean (52.9, 99.9, **100**)	VV_mean (25.1, 62.4, **93.2**)
ww34b (49.1, 87.4, **100**)	Ws_a_mean (**67.5**, 46.3, 0)
Ws_b_mean (41.2, 68.0, **100**)	TT_mean (27.9, 29.5, **67.5**)
ff_mean (36.1, 35.5, **100**)	Wt_a2_mean (38.2, **61.8**, 0)
h_mean (33.7, 98.7, **100**)	Wt_b1_mean (**37.5**, 32.8, 0)

Table 3.3: Diagnostic model results for Model 1 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.2.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.274	0.228	0.293	0.285	0.356	0.355
	N_{var}	6	19	12	20	16	20
mean	R^2_{adj}	0.418	0.404	0.391	0.389	0.378	0.377
	N_{var}	12	20	16	20	18	20
max	R^2_{adj}	0.555	0.553	0.473	0.467	0.389	0.389
	N_{var}	18	20	20	20	19	20

mode, where the backward selection scheme is applied at every time step. Noteworthy that the fixed-set results are achieved with a consistently higher number of predictors. The reason for having a minimum of only 19 predictors in the 1-year windows using the fixed set of predictors is that also here variables were skipped if there were less than 5 observations in the investigation period. The last note to make is that the tendency of R^2_{adj} values decreasing with larger windows should not be overrated. This is a typical diagnostic pattern as shorter time series can simply be better reproduced by the method.

Table 3.4 gives the final results of the model, both in diagnostic and in prognostic mode, using 2001-2005 as the calibration period and 2006 for validation. The mean absolute error *MAE* is lower than 0.1 for both modes, with smaller values in the diagnostic mode. The last pattern also holds for the *RMSE* and the *SE*. R^2 is around 0.37 and slightly higher in prognostic mode. The R^2_{adj} of 0.358 in diagnostic mode accounts for the inclusion of the 20 predictor variables and is thus slightly lower than R^2. The whole model

Table 3.4: Quality criteria for Model 1.

quality criterion	diagnostic	prognostic
MAE	0.066	0.081
$RMSE$	0.092	0.115
SE	0.093	0.118
$r_{multiple}$	0.604	0.610
R^2	0.365	0.372
R^2_{adj}	0.358	/

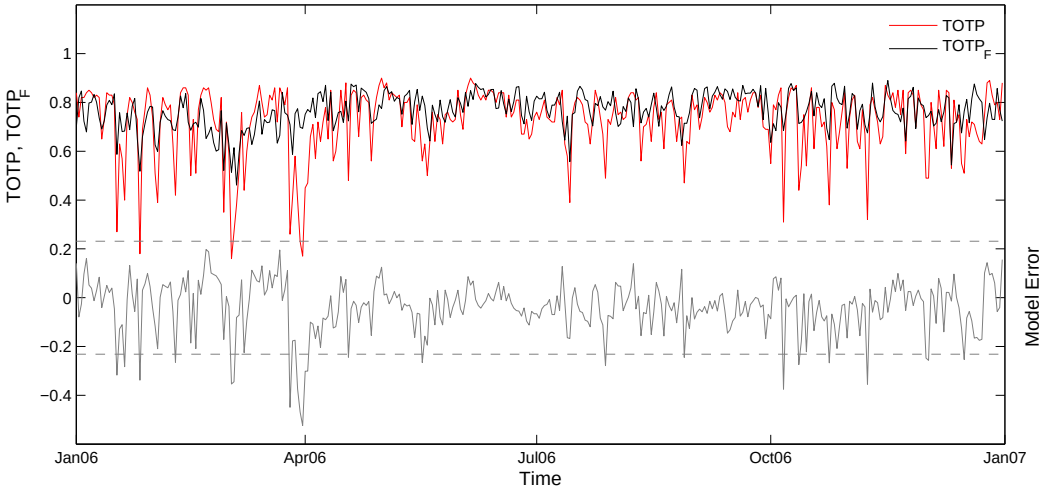


Figure 3.3: Time series of $TOTP$, $TOTP_F$ and the model error for Model 1. Dashed lines give twice the standard deviation of $TOTP$.

is significant on the 2.29×10^{-13} level, running a global F-test.

Figure 3.3 shows the time series of $TOTP$ and $TOTP_F$ for the validation period (2006). In the lower part of the graph the model error $TOTP - TOTP_F$ is shown. The dashed lines give twice the standard deviation of $TOTP$. Even this simple model can explain a lot of the variability in $TOTP$, as already shown in Table 3.4. Generally, low and high punctuality pattern are well recognised in $TOTP_F$ and trends reflected. The largest deficiency of this model is its inability to reflect extremely low punctualities. This is also found again in the scatterplot of $TOTP$ and $TOTP_F$, shown in Figure 3.4. There is a clear tendency of the model to overestimate punctuality on days with $TOTP < 0.6$. Days with high punctualities are, on the other hand, generally well reflected.

3.2.2 Model 2 – Variable Transformations

Often, a predictor and the predictant exhibit a nonlinear relationship. Horizontal visibility, for example, is likely to be related to $TOTP$ in a nonlinear way. Clearly, a decrease from 1.5 to 0.5 km has another impact on air traffic

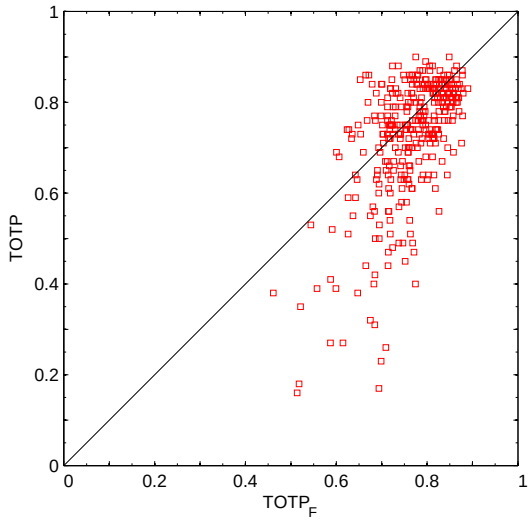


Figure 3.4: Scatterplot of $TOTP$ and $TOTP_F$ for Model 1.

Table 3.5: Nonlinear predictor variable transformations used (BACKHAUS et al., 2003).

transformation	**function** $f(x)$	**validity**	**predictor suffix**
logarithm	$\ln(x)$	$x > 0$	_log
exponential	$\exp(x)$	$-\infty$ - ∞	_exp
reciprocal	$1/x$	$x \neq 0$, here: $x > 0$	_1/x
radical	$\sqrt{(}x)$	$x \geq 0$	_sqrt
power 2	x^2	here: $x \geq 0$	_^2
power 3	x^3	here: $-\infty$ - ∞	_^3
power 4	x^4	here: $x \geq 0$	_^4

and airport operations than one from 52 to 51 km. A logarithmic or reciprocal transformation, for example, is hence likely to improve modelling results. Model 2, therefore, allows for the use of nonlinear predictor variable transformations, as described in BACKHAUS et al. (2003). This method accounts for nonlinear relationships between predictors and the predictant $TOTP$. Nonlinear transformations used in this approach are listed in Table 3.5. Suffixes appended to predictor denotation after transformation are also given. For example, transforming a hypothetic predictor x_i to $x'_i = f(x_i)$, using one of the proposed tranformations, Equation 2.4 becomes:

$$\widehat{TOTP}_t = \hat{\beta}_0 + x_{1,t} \cdot \hat{\beta}_1 + \ldots + x'_{i,t} \cdot \hat{\beta'}_i + \ldots + x_{k,t} \cdot \hat{\beta}_k + \epsilon_t \qquad (3.1)$$

, with $\hat{\beta'}_i$ beeing the coefficient of x'_i to be estimated. Equation 3.1 is still linear, but the nonlinear relationship between x_i and $TOTP$ is now accounted for.

Within this model approach, a transformed predictor x'_i is not just added to the design matrix, thus increasing the number of predictors. When in-

Table 3.6: Fixed set of predictors for Model 2. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	ww34b (32.6, 64.9, **100**)
SH_max (95.3, 98.0, **100**)	CLc2 (31.2, 72.7, **100**)
VV_mean_1/x (81.7, **100**, **100**)	Wt_a2_mean (14.1, 48.5, **100**)
E1 (71.0, **100**, **100**)	N_mean (13.5, 46.8, **99.5**)
ww31b (68.5, 90.8, **100**)	ww32b (23.8, 39.4, **94.8**)
Ws_a_mean_^2 (48.5, 71.2, **100**)	P0_mean (33.0, 34.5, **93.4**)
Wt_b2_mean_^4 (47.7, 92.6, **100**)	fx24_max (29.4, 54.9, **70.2**)
ff_mean (44.0, 62.4, **100**)	ww33b (19.7, **47.4**, 0.5)
r1_mean (41.8, 81.7, **100**)	Wt_a1_mean_^4 (**40.5**, 35.0, 0)
VV_min_log (33.8, 94.7, **100**)	Ws_b_mean (**37.8**, 35.6, 0)

dicated, it replaces the untransformed predictor x_i it originates from, instead. As a decision rule if and how to include a predictor in the design matrix – transformed or untransformed – the correlation coefficient of x_i, respectively x'_i, and $TOTP$ is calculated for each potential transformation (see Table 3.5) and each predictor, using the whole 6-year dataset. The transformed predictor $x'_{i,max}$ exhibiting the highest squared correlation coefficient r^2 with $TOTP$ is then chosen to replace x_i in the design matrix, if $r^2(x'_{i,max}, TOTP) - r^2(x_i, TOTP) > 0.01$. The latter threshold is claimed in order to use a transformed predictor only if it exhibits a potentially valuable improvement of model performance.

Table 3.6 shows the predictor selection for Model 2, now allowing for nonlinear variable transformations. Compared to Table 3.2, there is a shift in variable weighting, triggered by better implementation of predictors, such as e.g. horizontal visibility. Whereas some predictors exhibit almost the same ranking position as in Model 1, some predictors, such as VV_mean_1/x, VV_min_log, P0_mean or ww33b, feature a much higher significance. Other predictors, such as h_mean or CLc1 have dropped out of the top twenty predictors completely. There are also some other minor ranking shifts, especially among the wind variables.

Table 3.7 shows model statistics for the three defined moving windows. Compared to results from Model 2 (see Table 3.3), R^2_{adj}-values have consistently improved. Focusing on mean-values there is a significant rise in model quality. This is reconfirmed by the final results using the 2001-2005 calibration window. Model 2 then is significant on the 6.47×10^{-15} level. Table 3.8 lists the quality criteria for this model. Both in diagnostic and in prognostic mode, model results are better than for Model 1. Nonlinear variable transformation thus prove to be appropriate in the punctuality modelling approach.

Table 3.7: Diagnostic model results for Model 2 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.6.

		1 year		3 years		5 years	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.305	0.286	0.318	0.307	0.388	0.386
	N_{var}	6	19	12	20	16	20
mean	R^2_{adj}	0.449	0.435	0.424	0.423	0.409	0.408
	N_{var}	11	20	15	20	17	20
max	R^2_{adj}	0.590	0.586	0.509	0.504	0.418	0.418
	N_{var}	19	20	21	20	20	20

Table 3.8: Quality criteria for Model 2.

quality criterion	**diagnostic**	**prognostic**
MAE	0.065	0.081
RMSE	0.090	0.115
SE	0.091	0.118
$r_{multiple}$	0.628	0.629
R^2	0.394	0.396
R^2_{adj}	0.387	/

3.2.3 Model 3 – Runway-Related Wind Components

When creating wind-related predictors, it seems natural to define these variables against the background of the orientation of the runway system, as several operational thresholds are based on tailwind- and, at some airports, also crosswind-components. Model 3 accommodates this given fact. Thus, it is identical with Model 2, except for the wind-related predictors used. For Model 3, the alternative wind variable definitions as described in Section 2.2.2 were followed. Thus, the runway configuration independent wind variables as used in Model 2 were replaced by the runway dependent wind variables listed in Table 2.7. The variable selection is somewhat different to that in the previous model, as shown in Table 3.9. Worth mentioning is that all wind variables depending on the runway configuration were found to be significant.

Model 3 is significant on the 3.77×10^{-15} level. The quality criteria of this model are listed in Table 3.11. Compared to Model 2, there are only minor diagnostic improvements. In prognostic mode, *MAE*, *RMSE* and *SE* have slightly decreased, whereas R^2 has decreased from 0.396 to 0.389. Based on these ambiguous model results, the strategy of using runway configuration dependent wind-related predictor variables is not further pursued within these analyses. Model 2 will, therefore, be the basis for further model enhance-

Table 3.9: Fixed set of predictors for Model 3. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	Ws_PB_mean (38.3, 72.6, **100**)
SH_max (95.7, 98.0, **100**)	Wt_S_mean (36.8, 49.3, **100**)
Wt_head_PB_mean (83.1, **100**, **100**)	VV_min_log (30.7, 97.5, **100**)
Ws_PB_mean_^2 (82.1, 99.9, **100**)	ww34b (28.8, 70.5, **100**)
VV_mean_1/x (75.5, **100**, **100**)	CLc2 (27.2, 78.2, **100**)
E1 (71.6, **100**, **100**)	N_mean (21.1, 74.7, **100**)
Wt_tail_PB_mean (71.2, 94.1, **100**)	P0_mean (34.6, 57.2, **98.4**)
ww31b (68.7, 90.9, **100**)	fx24_max (39.2, 76.4, **97.5**)
Wt_N_mean_^4 (53.6, 88.6, **100**)	ww32b (20.7, 22.6, **93.9**)
r1_mean (44.2, 78.5, **100**)	h_mean_log (23.5, **57.5**, 43.7)

Table 3.10: Diagnostic model results for Model 3 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.9.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.294	0.286	0.321	0.317	0.392	0.390
	N_{var}	7	19	14	20	19	20
mean	R^2_{adj}	0.452	0.438	0.430	0.428	0.412	0.411
	N_{var}	13	20	18	20	20	20
max	R^2_{adj}	0.589	0.588	0.515	0.509	0.421	0.420
	N_{var}	19	20	22	20	22	20

ment steps. For later studies with more precise data on runway configuration changes it should be investigated if predictor variables based on this refined information are capable of improving model quality more significantly. For the sake of completeness, also for this approach diagnostic model results are summarised in Table 3.10.

3.2.4 Model 4 – Enhanced Boolean Predictor Variables

Focusing on boolean type predictor variables, describing e.g. a thunderstorm occurrence or the observation of a certain cloud type, the question of event frequency and duration comes to the fore. A thunderstorm passing the airport quickly, will, in the simple boolean variable scheme, be represented the same way as a sequence of several thunderstorm cells, hitting the airport during the course of day. The impact of the latter on air traffic operations is, however, likely to be considerably larger.

Based on model assumptions made for Model 2, Model 4, therefore, uses

Table 3.11: Quality criteria for Model 3.

quality criterion	diagnostic	prognostic
MAE	0.065	0.080
RMSE	0.090	0.114
SE	0.090	0.117
$r_{multiple}$	0.631	0.624
R^2	0.398	0.389
R^2_{adj}	0.392	/

Table 3.12: Fixed set of predictors for Model 4. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	RRR_mean_sqrt (39.2, 45.3, **99.7**)
SH_max (90.8, 94.9, **100**)	VV_mean_1/x (40.7, **99.5**, 96.2)
ww31b (81.4, **100**, **100**)	N_mean (22.1, 49.8, **98.1**)
ww34b (70.4, **100**, **100**)	TT_mean (48.4, 31.7, **94.0**)
E1 (63.5, **100**, **100**)	Wt_b2_lim (38.2, 69.5, **72.4**)
ff_mean (52.6, 86.7, **100**)	Ws_a_mean_^2 (44.2, **69.6**, 53.8)
CLc2 (41.3, 98.1, **100**)	CLc1 (29.4, **55.9**, 4.9)
ww33b (37.3, 41.9, **100**)	ww35b_sqrt (26.0, **43.6**, 0.3)
VV_min_log (23.8, 44.8, **100**)	fx24_max (31.7, **43.5**, 32.2)
Wt_a2_mean (20.1, 40.5, **100**)	Wt_b2_mean_^4 (32.6, 30.5, **41.8**)

enhanced boolean predictor variables. In Model 2, predictors from categories *ww3xb*, *CLcx*, *Ex* and *Wt_b2_lim* were purely boolean, thus only exhibiting values 0 or 1, depending on whether the respective event occurred or not. Model 4, in contrast, additionally uses information on the number of occurrences per day. In this model, each predictor from the above groups of predictors is allowed to take values between 0 and 10. A maximum value of 10 is assigned, if the respective event occurred at each official observation, i.e. 24 times per day. Likewise, a value of 0 is assigned, if the event did not occur at all. Values between 0 and 10 are assigned according to the daily number of occurrences. This way, boolean type variables are transformed to interval-scaled variables. Non-linear variable transformations as described in Section 3.2.2 can be applied to these enhanced boolean variables just like to any other non-boolean predictor variable, taking into account the limitations given in Table 3.5. Table 3.12 shows the selection of significant variables for Model 4. There are some major shifts in variable significance as compared to Model 2. Some predictors such as *r1_mean*, *ww32b* or *P0_mean* have completely dropped out of the list. Other previously not selected variables such as *RRR_mean_sqrt*, *TT_mean*, *Wt_b2_lim*, *CLc1* or *ww35b_sqrt* now proved to be of higher significance. Under this new constellation, there were also some shifts in variable importance with regard to wind related predictors.

Table 3.13: Diagnostic model results for Model 4 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.12.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.339	0.320	0.342	0.340	0.415	0.413
	N_{var}	6	19	11	20	14	20
mean	R^2_{adj}	0.476	0.459	0.450	0.447	0.434	0.433
	N_{var}	12	20	16	20	17	20
max	R^2_{adj}	0.581	0.578	0.525	0.522	0.444	0.443
	N_{var}	18	20	21	20	20	20

Table 3.14: Quality criteria for Model 4.

quality criterion	**diagnostic**	**prognostic**
MAE	0.063	0.077
RMSE	0.088	0.111
SE	0.089	0.114
$r_{multiple}$	0.648	0.653
R^2	0.420	0.426
R^2_{adj}	0.414	/

Diagnostic model results for the moving 1-year, 3-year and 5-year windows are shown in Table 3.13. Compared to the results shown in Table 3.7, the present approach exhibits a considerable improvement in R^2_{adj}. Model 4 calibrated with 2001-2005 data is significant on the 1.96×10^{-16} level. The introduction of enhanced boolean variables yields a noticeable improvement in model quality as shown in Table 3.14. Both in diagnostic and in prognostic mode, model errors have decreased as compared to Model 2. R^2_{adj} has increased from 0.387 to 0.414, in prognostic mode the improvement in R^2 is from 0.396 to 0.426. Against the background of these results, enhanced boolean predictor variables prove to be a reasonable model enhancement. They are thus used for later model stages.

3.2.5 Model 5 – Upper Level Wind

The introduction of upper level wind variables into the model, as proposed by Spehr (2003), is a another step forward to a model incorporating predictors based on operational thresholds. Unarguably, direction and strength of upper level winds are significant factors with regard to approach staggering (see arrival rate matrix in Appendix A) and, consequently, airport acceptance rate and punctuality. In the following, it is analysed, if the inclusion of upper level

Table 3.15: Fixed set of predictors for Model 5 using log information on upper level winds. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	RRR_mean_sqrt (33.2, 59.6, **100**)
SH_max (92.7, 97.3, **100**)	VV_min_log (24.6, 33.7, **100**)
ww31b (80.7, **100**, **100**)	Wt_a2_mean (20.5, 45.8, **100**)
ww34b (73.2, **100**, **100**)	ww33b (30.8, 18.7, **91.3**)
E1 (66.3, 99.6, **100**)	Wt_b2_lim (40.1, 69.5, **88.3**)
Hoehenwinde (65.8, **100**, **100**)	N_mean (17.0, 49.2, **87.7**)
TT_mean (56.8, 50.2, **100**)	CLc4 (27.0, 32.8, **67.5**)
CLc2 (53.4, 99.9, **100**)	CLc3 (14.6, 28.0, **67.5**)
ff_mean (47.0, 73.3, **100**)	Ws_a_mean_^2 (46.5, **63.0**, 36.6)
VV_mean_1/x (42.4, 98.7, **100**)	P0_mean (18.6, 35.3, **57.4**)

Table 3.16: Fixed set of predictors for Model 5 using AMDAR information on upper level winds. The numbers in brackets are the percentage of selections for the moving 1-year and 3-year window.

predictors	
const (**100**, **100**)	E1 (52.4, **98.4**)
ww31b (98.5, **100**)	RRR_mean_sqrt (42.4, **96.7**)
ww34b (93.3, **100**)	N_mean (15.2, **94.8**)
SH_max (86.9, **100**)	VV_min_log (29.4, **91.3**)
Wt_b2_lim (69.1, **100**)	h_mean_log (27.1, **84.2**)
Ws_a_mean_^2 (66.8, **100**)	CLc1 (37.4, **72.4**)
VV_mean_1/x (42.6, **100**)	max_Tangentialwind_a_tail_gt35 (29.2, **71.3**)
CLc2 (46.4, **99.5**)	max_Tangentialwind_a_head_ge15 (15.5, **65.6**)
Ws_b_mean (35.7, **99.5**)	P0_mean (24.9, **62.0**)
TT_mean (48.3, **98.9**)	ww33b (24.7, **60.9**)

wind variables results in the expected model improvement. Based on Model 4, there were two different upper level wind data sources available for Model 5, as described in Section 2.2.1 and 2.2.3 and discussed in Section 2.2.3.1: 1) daily logs, 2) AMDAR data. Notice that both the log-based boolean variable *Hoehenwinde* and AMDAR-based boolean variables remain purely boolean, i.e. they are not transformed like the SYNOP-based predictors as described in Section 3.2.4. This is because log information is only available in day resolution. AMDAR information, on the other hand, is not evenly distributed over the course of day with significantly less reports during nighttime. Hence, it is not reasonable to create enhanced boolean variables. In the following, results after inclusion of either log or AMDAR data are compared.

Table 3.15 shows the variable selection when upper level wind information is taken from the daily logs. The additional predictor *Hoehenwinde* is among the most selected predictors in all three windows. Table 3.16

Table 3.17: Diagnostic model results for Model 5 using a moving 1-year, 3-year and 5-year window and log information on upper level winds. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.15.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.340	0.318	0.357	0.351	0.437	0.436
	N_{var}	7	19	12	20	15	20
mean	R^2_{adj}	0.493	0.474	0.470	0.467	0.462	0.461
	N_{var}	13	20	17	20	19	20
max	R^2_{adj}	0.610	0.604	0.547	0.545	0.475	0.474
	N_{var}	19	20	22	20	21	20

Table 3.18: Diagnostic model results for Model 5 using a moving 1-year and 3-year and AMDAR information on upper level winds. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.16.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.383	0.335	0.491	0.485	/	/
	N_{var}	7	20	15	20	/	/
mean	R^2_{adj}	0.537	0.514	0.514	0.510	/	/
	N_{var}	14	20	20	20	/	/
max	R^2_{adj}	0.608	0.586	0.535	0.533	/	/
	N_{var}	21	20	23	20	/	/

shows the respective variable selection, when AMDAR information on upper level winds is included. Notice that AMDAR information was only available from 2003 to 2006, hence no 5-year window could be moved through the data. Two AMDAR-based predictors (*max_Tangentialwind_a_tail_gt35*, *max_Tangentialwind_a_head_ge15*) were found to be among the 20 most important predictors. Both are related to upper level winds tangential to the PRS orientation.

Tables 3.17 and 3.18 specify the diagnostic model results for both model versions. Comparisons with the reference Model 4 (see Table 3.13) show that the inclusion of either log or AMDAR based upper level wind variables improves diagnostic model performance. In direct comparison, AMDAR based predictors prove to be most efficient.

Tables 3.19 and 3.20 give the final model results for Model 5, including either log or AMDAR based predictors. Model 5 calibrated with 2001-

Table 3.19: Quality criteria for Model 5 using log information on upper level winds. Calibration periods 2001-2005 and 2003-2005 for comparison with Table 3.20.

	2001-2005		2003-2005	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.062	0.074	0.058	0.077
RMSE	0.086	0.105	0.083	0.107
SE	0.087	0.108	0.084	0.110
$r_{multiple}$	0.666	0.667	0.721	0.660
R^2	0.443	0.445	0.520	0.436
R^2_{adj}	0.437	/	0.511	/

Table 3.20: Quality criteria for Model 5 using AMDAR information on upper level winds.

quality criterion	**diagnostic**	**prognostic**
MAE	0.060	0.077
RMSE	0.084	0.109
SE	0.085	0.112
$r_{multiple}$	0.708	0.652
R^2	0.501	0.426
R^2_{adj}	0.492	/

2005/2003-2005 data and using log information on upper level winds is significant on the $7.23 \times 10^{-18}/1.16 \times 10^{-14}$ level. Model 5 calibrated with 2003-2005 data and using AMDAR information on upper level winds is significant on the 1.01×10^{-13} level. For comparison, Table 3.19 gives results both for the established calibration period 2001-2005 and for the calibration period 2003-2005.

As given in Tables 3.19 and 3.20, both versions of Model 5 exhibit better results, both diagnostic and prognostic, than the reference Model 4 (see Table 3.14). Unlike intermediate results shown in Tables 3.17 and 3.18, final quality criteria for the model with log based upper level wind information are better than those for the model with AMDAR based information included, both in diagnostic and prognostic mode. Furthermore, it shows that using the 3-year calibration window in the log-enhanced model leads to better diagnostic but worse prognostic results, compared to the model calibrated on 5-years. This effect is later discussed in Section 3.2.13.1.

In the following, taking the latter results and quality criteria into account, further model enhancements will emanate from the superior Model 5 with log information on upper level winds included. In particular, this choice is also justified since the long calibration period 2001-2005 can thus be further used, as log data is available for the whole 5-year period. For potential punctuality forecast models, which will be discussed in Section 3.3, AMDAR-based predictors are reconsidered. On one hand, they proved to raise model

Table 3.21: Fixed set of predictors for Model 6. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	h_mean_log (37.5, 85.1, **100**)
TOTFL_scheduled (**100**, **100**, **100**)	Wt_a2_mean (32.0, 36.7, **100**)
SH_max (98.2, 99.3, **100**)	CLc1 (31.7, 79.6, **100**)
ww31b (76.5, **100**, **100**)	fx24_max (23.0, 60.7, **100**)
Wt_b2_lim (72.7, 97.8, **100**)	h_max (9.4, 53.3, **97.0**)
E1 (68.5, **100**, **100**)	Ws_a_mean_^2 (41.4, 75.4, **96.2**)
ww34b (67.4, **100**, **100**)	N_mean (16.6, 35.0, **95.6**)
Hoehenwinde (61.1, **100**, **100**)	VV_min_log (22.1, 32.0, **92.3**)
ff_mean (54.3, 79.3, **100**)	RRR_mean_sqrt (35.6, 53.6, **88.0**)
ww33b (46.5, 87.4, **100**)	CLc2 (51.2, 76.1, **87.2**)

quality as compared to previous models. On the other hand, they can, unlike the log based predictor *Hoehenwinde*, be forecasted by NWP models and thus be generally used as an input for a punctuality forecast model.

3.2.6 Model 6 – Traffic

Up to this point, model improvements were achieved either through methodic advancements or the extension and refinement of the database of weather-related predictors. It is well known, that the amount of traffic an airport has to cope with in a given time frame, especially against the background of the airport's nominal capacity, is a crucial factor in terms of the generation and multiplication of delays. In Model 6, which is based on Model 5, information on scheduled traffic is, therefore, taken into account. Table 3.21 lists the selection of predictors for this enhanced model. Again, there are predictor variable shifts compared to the previous model. Most important, the newly

Table 3.22: Diagnostic model results for Model 6 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Table 3.21.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.465	0.437	0.485	0.480	0.515	0.512
	N_{var}	9	19	16	20	19	20
mean	R^2_{adj}	0.575	0.556	0.552	0.549	0.539	0.536
	N_{var}	14	20	19	20	22	20
max	R^2_{adj}	0.669	0.664	0.594	0.590	0.549	0.548
	N_{var}	18	20	24	20	24	20

Table 3.23: Quality criteria for Model 6.

quality criterion	diagnostic	prognostic
MAE	0.058	0.069
RMSE	0.080	0.095
SE	0.081	0.097
$r_{multiple}$	0.721	0.719
R^2	0.519	0.517
R^2_{adj}	0.514	/

introduced predictor *TOTFL_scheduled* is always selected within all three windows. Thus, it is a meaningful extension of the predictor base.

The diagnostic model results shown in Table 3.22 support the relevance of traffic information to be used as model input variable. R^2_{adj} values are consistently higher than in Model 5, both diagnostic and prognostic. The gain in mean R^2_{adj} is almost 10 percentage points. The final model results (see Table 3.23) show the same tendency. Both model errors and R^2 exhibit a significant improvement. Even prognostic R^2 values are now larger than 0.5. Model 6 calibrated with 2001-2005 data is significant on the 1.67×10^{-23} level and thus highly significant. Scheduled traffic will, therefore, be included in further model stages.

3.2.7 Model 7 – Weather Related Predictors

Model 7 further continues with the inclusion of potentially relevant weather-related predictors. In this model, two additional variables (also see Section 2.2.1) are introduced: *DayIndex* and *Mix*. *DayIndex* for this model comprises information on the CAT stage, only. ATC regulations and special events/incidents are not considered in Model 7. The simple boolean pre-

Table 3.24: Fixed set of predictors for Model 7. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	ww33b (50.1, 90.3, **100**)
TOTFL_scheduled (**100**, **100**, **100**)	Wt_a2_mean (34.6, 40.0, **100**)
SH_max (98.3, 99.6, **100**)	h_mean_log (33.3, 81.9, **100**)
Wt_b2_lim (70.7, 93.0, **100**)	CLc1 (31.9, 76.1, **100**)
E1 (70.6, **100**, **100**)	fx24_max (22.2, 55.4, **100**)
ww31b (70.3, **100**, **100**)	VV_mean_1/x (40.1, 90.9, **98.6**)
ww34b (67.0, **100**, **100**)	RRR_mean_sqrt (34.4, 70.1, **97.5**)
DayIndex (64.0, **100**, **100**)	h_max (6.2, 37.6, **82.5**)
Hoehenwinde (63.2, **100**, **100**)	Ws_a_mean_^2 (41.3, 76.1, **80.3**)
ff_mean (53.7, 82.5, **100**)	CLc2 (48.1, 75.3, **79.0**)

Table 3.25: Diagnostic model results for Model 7 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Tables 3.24.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.476	0.441	0.494	0.489	0.523	0.521
	N_{var}	9	18	16	20	20	20
mean	R^2_{adj}	0.584	0.566	0.563	0.559	0.545	0.543
	N_{var}	14	20	20	20	22	20
max	R^2_{adj}	0.682	0.671	0.602	0.600	0.555	0.552
	N_{var}	20	20	26	20	25	20

Table 3.26: Quality criteria for Model 7.

quality criterion	**diagnostic**	**prognostic**
MAE	0.057	0.070
RMSE	0.079	0.096
SE	0.080	0.099
$r_{multiple}$	0.728	0.706
R^2	0.530	0.499
R^2_{adj}	0.525	/

dictor *Mix*, as defined in Section 2.2.1, tells if runway configuration changes occurred on a given day. The number of changes per day is not accounted for. Both *DayIndex* as defined above and *Mix* are weather related and can potentially be predicted by NWP models, and a follow-up model, respectively. The CAT stage can, at least theoretically, be drawn from visibility and cloud information. Preconditions for the initiation of runway configuration changes are well specified. *Mix* can thus be deduced from wind data.

Table 3.24 lists the predictors selected for this model. The variable *Mix* was not among the 20 most important predictors. *DayIndex*, in contrast, is highly significant. Using this extended set of predictors exhibits slightly better diagnostic results than for Model 6, as shown in Tables 3.25 and 3.26. Model 7 calibrated with 2001-2005 data is significant on the 1.83×10^{-24} level and thus highly significant. In prognostic mode, however, model quality has decreased compared to the previous model, both in terms of model errors and R^2 values. In following models, this effect will be kept in mind and discussed again.

Table 3.27: Fixed set of predictors for Model 8. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	ww33b (47.7, 58.3, **100**)
DayIndex (**100**, **100**, **100**)	VV_mean_1/x (45.4, 94.6, **100**)
TOTFL_scheduled (**100**, **100**, **100**)	ff_mean (59.6, 80.4, **99.2**)
SH_max (91.1, 99.5, **100**)	Wt_a2_mean (31.2, 44.9, **93.7**)
TOTFL_scheduled_minus_TOTFL (82.8, **100**, **100**)	fx24_max (15.8, 41.8, **86.1**)
Hoehenwinde (66.7, **100**, **100**)	CLc1 (21.1, 59.6, **77.3**)
E1 (64.4, 99.5, **100**)	N_mean (17.7, 47.5, **77.3**)
ww31b (63.1, **100**, **100**)	Ws_a_mean_^2 (39.7, **68.3**, 60.7)
ww34b (62.5, **100**, **100**)	Wt_b2_mean_^4 (37.7, 33.9, **65.8**)
CLc2 (52.9, 82.3, **100**)	Wt_b2_lim (23.9, **63.5**, 34.2)

3.2.8 Model 8 – Non-Weather Related Predictors

Model 8 uses a new definition of the predictor *DayIndex* which now comprises information on the CAT stage, incidents and system failures as well as ATC regulations. This implies that *DayIndex* is no longer a purely weather based predictor. The advantage of this approach is that – in terms of unusual non-weather-related events – critical days do not have to be excluded from the analysis as done by Spehr (2003). Rather, those days are now flagged. In a thinkable punctuality forecast approach, this newly defined predictor *DayIndex* could then potentially be used as a switch to calculate two different scenarios for *TOTP*, either assuming critical incidents to arise or not.

Additionally, information on actual traffic is now included in the set of potential predictors. A new predictor *TOTFL_scheduled_minus_TOTFL* as described in Section 2.2.1 is then created, generally representing cancellations. Table 3.27 shows the predictor variable selection for Model 8. Again, *Mix* did not prove to be of high significance in this new, extended set of potential pre-

Table 3.28: Diagnostic model results for Model 8 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Tables 3.27.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.555	0.511	0.561	0.550	0.587	0.584
	N_{var}	10	19	16	20	18	20
mean	R^2_{adj}	0.643	0.623	0.622	0.614	0.605	0.599
	N_{var}	15	20	20	20	21	20
max	R^2_{adj}	0.733	0.724	0.657	0.651	0.614	0.607
	N_{var}	21	20	23	20	23	20

Table 3.29: Quality criteria for Model 8.

quality criterion	diagnostic	prognostic
MAE	0.053	0.067
RMSE	0.074	0.092
SE	0.074	0.095
$r_{multiple}$	0.768	0.735
R^2	0.590	0.541
R^2_{adj}	0.585	/

dictors. The newly introduced variable *TOTFL_scheduled_minus_TOTFL*, in contrast, was among the 5 most selected variables. Likewise, the extended *DayIndex* variable proved to be of high significance. It was always selected in all three windows.

Table 3.28 clearly shows that diagnostic model results have consistently improved compared to all previous models. Mean R^2 values are now around or higher than 0.6 in all three windows. Accordingly, diagnostic results for the final model have significantly improved, both model errors and R^2, as shown in Table 3.29. Also in prognostic mode, model errors decreased to $MAE = 0.067$ and $RMSE = 0.092$. R^2, on the other hand, increased to 0.54. Model 8 calibrated with 2001-2005 data is highly significant on the 1.68×10^{-30} level.

3.2.9 Model 9 – Higher Resolution Weather Variables

Undoubtedly, the impact of a certain weather event on airport operations is likely to vary depending on the time of day. The passage of a thunderstorm at midnight, for example, certainly affects less aircraft around midnight than at noon. In order to accommodate this fact, higher resolution weather variables

Table 3.30: Fixed set of predictors for Model 9. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window.

predictors	
const (**100**, **100**, **100**)	Ws_b_MEAN_1 (36.0, 54.2, **100**)
DayIndex (**100**, **100**, **100**)	RRR_2_sqrt (35.2, 63.0, **100**)
TOTFL_scheduled (99.9, **100**, **100**)	VV_MEAN_1_1/x (28.3, 81.1, **100**)
ww31b_2 (69.7, **100**, **100**)	ff_MEAN_3 (25.7, 58.8, **100**)
ww34b_2 (67.0, **100**, **100**)	Wt_b2_lim_1 (17.1, 31.0, **100**)
Hoehenwinde (66.4, **100**, **100**)	ww31b_1 (10.8, 79.9, **100**)
TOTFL_scheduled_minus_TOTFL	E1_3 (9.9, 43.5, **98.4**)
(65.5, **100**, **100**)	fx24_max (28.5, 45.9, **96.4**)
ww32b_3 (47.7, 79.9, **100**)	SH_MAX_3 (51.0, 37.6, **95.9**)
ff_mean (44.0, 66.4, **100**)	Wt_a1_MEAN_3_^4 (39.7, 71.1, **95.6**)
Wt_b1_MEAN_1 (37.1, 35.2, **100**)	

Table 3.31: Diagnostic model results for Model 9 using a moving 1-year, 2-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Tables 3.30.

		1 year		**2 years**		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.631	0.517	0.606	0.554	0.619	0.580	0.628	0.599
	N_{var}	17	17	23	19	25	20	32	20
mean	R^2_{adj}	0.724	0.632	0.683	0.621	0.673	0.622	0.646	0.610
	N_{var}	30	19	33	20	36	20	38	20
max	R^2_{adj}	0.817	0.715	0.750	0.677	0.706	0.658	0.654	0.616
	N_{var}	50	20	49	20	49	20	44	20

as introduced in Section 2.2.2 are added to the pool of predictors in Model 9. Blocks of six hours length are evenly distributed over the course of day and supposed to better satisfy the effect of weather events at special times of day. Using a mix of predictors on a 24 as well as 6 hour basis significantly increases the number of potential predictor variables. Table 3.30 shows the selection of final predictor variables taken from the extended pool of predictors. It is a well-balanced mix of predictors available at 6-hour resolution and predictors available at 24-hour resolution, whereof many proved to be highly significant in previous models.

Results of Model 9 are brought together in Table 3.31. Compared to Model 8, diagnostic results have again improved in all three windows. A 2-year window is added for comparisons with later models. Concentrating on the most important window, the 5-year window, mean R^2 has increased from 0.599 to 0.610. Values given using the automatic backward selection scheme should be carefully interpreted, as more variables (50 at maximum) proved to be significant on the $\alpha = 0.05$-level.

Model 9 calibrated with 2001-2005 data is highly significant on the 4.09×10^{-32} level. Model results are shown in Table 3.32. Diagnostic, and more

Table 3.32: Quality criteria for Model 9.

quality criterion	**diagnostic**	**prognostic**
MAE	0.052	0.066
RMSE	0.073	0.090
SE	0.073	0.093
$r_{multiple}$	0.777	0.748
R^2	0.604	0.559
R^2_{adj}	0.599	/

Table 3.33: Fixed set of predictors for the summer season, Model 10. The numbers in brackets are the percentage of selections for the moving 1-year, 2-year and 3-year window.

predictors	
const (**100**, **100**, **100**)	RR1_MAX_1_sqrt (42.1, 61.4, **100**)
DayIndex (**100**, **100**, **100**)	ww32b_3 (40.9, 65.5, **100**)
TOTFL_scheduled (94.5, **100**, **100**)	Wt_b2_MEAN_3_^4 (32.8, 91.8, **100**)
Hoehenwinde (78.0, **100**, **100**)	TT_MEAN_3 (31.2, 52.6, **100**)
TOTFL_scheduled_minus_TOTFL	TT_MEAN_4 (27.1, 64.0, **100**)
(66.1, 69.1, **100**)	VV_MEAN_4_log (18.8, 33.5, **100**)
Wt_b1_MEAN_3_^3 (60.5, **100**, **100**)	ww31b_2 (16.0, 53.8, **99.5**)
VV_MIN_2_log (56.2, 69.2, **100**)	Wt_b1_MEAN_2_^2 (37.1, 51.9, **97.9**)
RRR_1_sqrt (53.1, 75.8, **100**)	RRR_2_sqrt (45.0, 86.0, **96.4**)
VV_MEAN_3_1/x (43.8, 74.8, **100**)	Ws_b_MEAN_2 (38.4, 53.8, **95.4**)
CLc2_2 (42.9, 76.2, **100**)	

Table 3.34: Fixed set of predictors for the winter season, Model 10. The numbers in brackets are the percentage of selections for the moving 1-year and 2-year window.

predictors	
const (**100**, **100**)	ww31b_1 (27.3, **94.8**)
DayIndex (**100**, **100**)	VV_MIN_1_log (40.6, **92.0**)
TOTFL_scheduled (94.5, **100**)	Ws_a_MEAN_4 (78.7, **89.7**)
TOTFL_scheduled_minus_TOTFL	Wt_a1_MEAN_3^4 (60.7, **83.9**)
(97.8, **100**)	CLc2_4 (8.0, **82.2**)
ww34b_2 (85.9, **100**)	RR1_MAX_3_sqrt (56.0, **78.7**)
ww31b_2 (79.6, **100**)	ff_mean (44.9, **77.6**)
Hoehenwinde (69.4, **100**)	ww33b_3 (30.2, **77.6**)
VV_MEAN_1_1/x (66.2, **100**)	Wt_b2_lim_1 (19.7, **70.7**)
Ws_a_MEAN_3^3 (49.4, **100**)	ff_MEAN_3 (42.1, **66.7**)
r1_MEAN_3_sqrt (46.0, **100**)	

important, also prognostic model results have improved over Model 9. Prognostic R^2 is at 0.559, *SE* at 0.066 and *RMSE* at 0.090. Thus, 6-hour block weather predictors are considered for further enhanced models as an extension to the set of daily representatives.

3.2.10 Model 10 – Breakdown into Summer and Winter Season

Model 10 makes one last step towards the final set of predictor variables. In the following it is tested, if a breakdown into summer/winter season is appropriate in the model approach at hand. The division is geared to the procedural method established at Lufthansa, the most important airline at Frankfurt Airport. Lufthansa uses different flight plans for summer and winter season, mainly to accommodate different passenger demand and also

Table 3.35: Diagnostic model results for the summer season, Model 10, using a moving 1-year, 2-year and 3-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Tables 3.33.

		1 year		2 years		3 years	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.575	0.530	0.587	0.534	0.598	0.560
	N_{var}	11	19	27	19	27	20
mean	R^2_{adj}	0.691	0.590	0.637	0.572	0.608	0.567
	N_{var}	32	20	37	20	37	20
max	R^2_{adj}	0.802	0.672	0.691	0.626	0.619	0.575
	N_{var}	52	20	49	20	46	20

Table 3.36: Diagnostic model results for the winter season, Model 10, using a moving 1-year and 2-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Tables 3.34.

		1 year		2 years	
		auto	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.680	0.544	0.689	0.622
	N_{var}	18	19	22	20
mean	R^2_{adj}	0.740	0.645	0.701	0.637
	N_{var}	31	20	30	20
max	R^2_{adj}	0.791	0.707	0.714	0.653
	N_{var}	49	20	42	20

different weather and traffic conditions. In that respect, Lufthansa and many other airlines even use different schedule buffers in the summer and winter season.

In the following, data is subdivided into summer and winter data. In that regard, the summer season is identical with the time of year with daylight-saving time. Thus, the summer season is slightly longer than the winter season and, accordingly, there is less winter data for calibration. Tables 3.33 and 3.34 show the final set of predictors selected for the summer and winter season season. Note that no 5-year window could be moved through the summer data as there were only 1288 summer days available for calibration. Thus, a 2-year window was introduced in addition to the 1- and 3-year windows. For the winter season, only a 1-year and 2-year window could be applied as the dataset offered only 903 winter season days according to the above definition. Both predictor selections exhibit typical seasonal predictors. For

Table 3.37: Quality criteria for Model 10.

quality criterion	diagnostic	prognostic
MAE	0.053	0.065
RMSE	0.073	0.091
SE	0.073	0.094
$r_{multiple}$	0.778	0.744
R^2	0.606	0.554
R^2_{adj}	0.601	/

example, *34b_2*, i.e. solid precipitation, is higly significant in winter. In the summer variable set, *32b_2*, i.e. thunderstorms, proved to be highly significant. On the other hand, other typical seasonal variables, such as *SH_MAX_x* or *E1_x*/*E2_x*, are not among the most selected predictors. Altogether, 34 different predictor variables are selected, some both in the summer and in the winter set of predictor variables. As only one set of predictors is applied at a time, depending on the season, only 20 predictors are used for calculations at a time. Thus, performance comparisons with previous models, where a set of 20 predictors was used each time, are possible and reasonable.

Diagnostic model results for summer and winter seasons are given in Tables 3.35 and 3.36. Comparisons with results from the previous Model 9 show that there are no significant quality improvements when applying this enhanced model. Note that splitting data into summer and winter data results in less data to move the summer/winter windows through. Furthermore, the 5-year window, and in case of the winter season, also the 3-year window, could not be used. Comparing the 1- and 2-year windows with each other, it strikes that using only winter data for calibration produces much better R^2 values as compared to results obtained using summer data. For winter days only, mean R^2_{adj} is slightly higher than mean R^2_{adj} for the non-seasonal datasets in Model 9 (see Table 3.31). For summer days, on the other hand, there is a significant drop in R^2_{adj}. There is no obvious reason for this behaviour. Since Model 10 is separately optimised for summer and winter season, it has to be concluded that either other, non-weather-related factors more dominantly come into play in the summer season, or that the weather impact on punctuality is less in the summer period. It is, on the other hand, also thinkable that this impact is just more complex than in the winter season and less well captured through the model approach at hand.

Table 3.37 gives the quality criteria for Model 10. This model, calibrated with 2001-2005 data, is highly significant on the 4.22×10^{-32} level. In diagnostic mode, both model errors and R^2 values are nearly identical with the respective criteria from Model 9 (see Table 3.32). In prognostic mode, R^2 has slightly decreased, whereas model errors remained almost unchanged. The main reason for this somewhat surprising result is probably that calibra-

tion periods, using this breakdown into summer and winter season, are much shorter than for a non-seasonal calibration. Again, the thesis is supported that slightly longer calibration periods are favourable.

Based on these results – Model 9 performs comparably well in diagnostic mode and even better in prognostic mode at less complexity – it is not reasonable to further pursue the approach chosen in Model 10 of separating the year into one summer and one winter season for model calibration and application. Therefore, Model 9 will be the basis for further model enhancement steps.

3.2.11 Model 11 – AR(1) Extension

Table 2.7 in Section 2.1.4 revealed that *TOTP* is autocorrelated with a lag-1 autocorrelation coefficient of 0.42. After fixation of the final set of predictor variables (see Table 3.30), model performance shall thus be increased using an additional AR(1) component, as described in Section 2.3.3. This way, time series information, which was ignored in previous models, is evaluated and a correction term is added to the multiple linear regression equation in order to account for autocorrelation effects.

Table 3.38 shows the diagnostic model results for Model 11. Clearly, there is a significant improvement in diagnostic R^2 values in all three windows. Also final results for this model, brought together in Table 3.39, corroborate the value of including time series information in the model approach. For the AR-enhanced Model 11, which is significant on the 4.09×10^{-32} level, diagnostic R^2_{adj} is 0.648. *MAE* is at 0.049 and *RMSE* at 0.068, respectively. In prognostic mode, R^2 is as high as 0.587 and has thus increased another 3 percentage points, compared to Model 9. The mean absolute error has decreased from 0.066 to 0.063, *RMSE* from 0.090 to 0.088. Based on these results, the AR(1) term is included in all following models.

Table 3.38: Diagnostic model results for Model 11 using a moving 1-year, 3-year and 5-year window. Minimum, mean and maximum values of R^2_{adj} and the number of selected predictor variables (N_{var}) are given, both for automatic model calibration using the backward selection scheme (*auto*) and the fixed set of predictors taken from Tables 3.30.

		1 year		**3 years**		**5 years**	
		auto	**fixed**	**auto**	**fixed**	**auto**	**fixed**
min	R^2_{adj}	0.661	0.577	0.668	0.639	0.674	0.646
	N_{var}	17	17	25	20	32	20
mean	R^2_{adj}	0.740	0.664	0.706	0.665	0.688	0.652
	N_{var}	30	19	36	20	38	20
max	R^2_{adj}	0.826	0.746	0.728	0.692	0.697	0.658
	N_{var}	50	20	49	20	44	20

Table 3.39: Quality criteria for Model 11.

quality criterion	diagnostic	prognostic
MAE	0.049	0.063
$RMSE$	0.068	0.088
SE	0.069	0.090
$r_{multiple}$	0.807	0.766
R^2	0.652	0.587
R^2_{adj}	0.648	/

3.2.12 Model 12 – Regression Trees

In the previous models, the multivariate linear regression approach was gradually enhanced by mathematical refinements and procedural improvements. A final set of predictor variables was determined and used to evaluate diagnostic as well as prognostic model performance. Using pure linear regression produced good results for days with high punctualities. However, on days with lower punctualities, there is a tendency of overestimating punctuality. That means, modelled punctualities are generally higher than actual punctualities. This is due to nonlinear effects not captured by the linear model. The effect is most obvious for days with $TOTP < 0.5$. Therefore, endeavours are being made, to improve model results for low punctuality days as defined in Section 3.1.

3.2.12.1 A Pure Regression Tree Model

Regression trees, in combination with other methodical approaches when indicated and when intelligently constructed and applied, theoretically allow for a detailed consideration and modelling of low punctuality days. Pure multivariate linear regression modelling, on the contrary, leaves little room for this approach. In a first step, a comparison model is set up on the basis of pure regression tree modelling. For this and all consecutive models, new ground is broken to evaluate model performance.

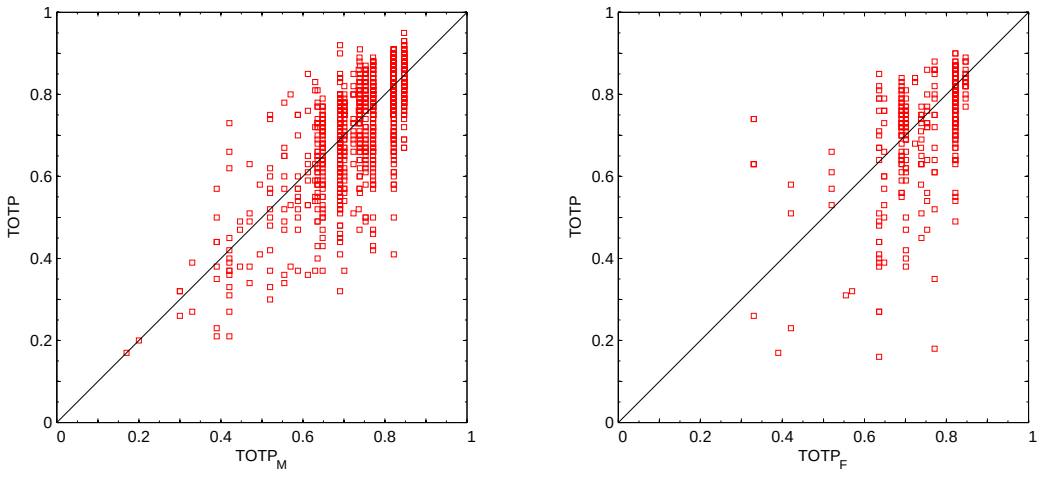
Regression trees, in general, can be tuned that way that diagnostic model results are close to or even perfect. Such an artificially tuned model is, however, totally overdetermined and only able to reproduce its calibration dataset, which is of little value. Applied to independent data, model results are generally bad. Therefore, there is no need to use the window method for model evaluation, as applied in the previous sections. Rather, model performance is evaluated on a prognostic basis, only.

Table 3.40 shows diagnostic (only for comparison with previous models and model stages) and prognostic model results for a pure regression tree model. Predictor variables are taken from Table 3.30. As described in Section 2.3.2, the tree is grown until each terminal node contains only one single

Table 3.40: Quality criteria for a pure regression tree model.

quality criterion	diagnostic	prognostic
MAE	0.053	0.078
$RMSE$	0.073	0.117
SE	0.074	0.120
$r_{multiple}$	0.774	0.563
R^2	0.599	0.316
R^2_{adj}	0.594	/

case. Afterwards, the tree is successively pruned until it contains 28 leaves, only. No AR(1) component is added. Both, diagnostic and prognostic R^2-values and model errors are worse than for the linear regression Model 9. Figures 3.5a and b show the scatterplots of $TOTP$ and $TOTP_M/TOTP_F$ for the pure regression tree model. While in diagnostic mode, variability is similar to variability obtained through linear regression, the scattering is unacceptably large in prognostic mode. The somewhat artificial look of both the scatterplots and the time series (see Figure 3.6) is due to the limitation to 28 terminal nodes. Generally, it is the crucial difficulty in the construction of regression trees to find a good balance between reproducibility and generalisability. Figure 3.6 impressively shows that the isolated regression tree approach at hand is not capable of correctly predicting low punctuality values. Also, it cannot reproduce daily variation of punctuality in a quality MLR is able to achieve. In Section 3.2.12.2, a more sophisticated approach will, therefore, be presented.



(a) Diagnosis (b) Prognosis

Figure 3.5: Scatterplot of $TOTP$ and (a) $TOTP_M$, (b) $TOTP_F$ for the pure regression tree model.

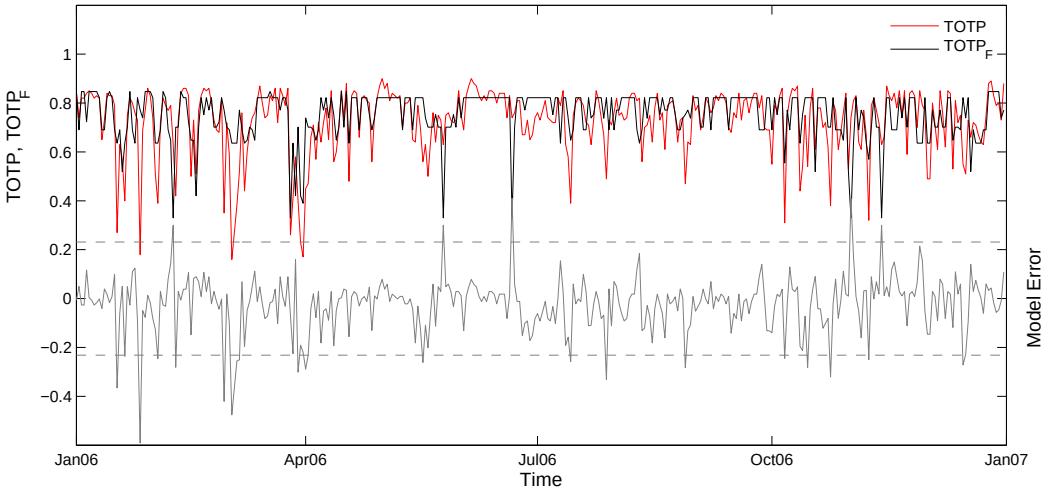


Figure 3.6: Time series of $TOTP$, $TOTP_F$ and the model error for the pure regression tree model. Dashed lines give twice the standard deviation of $TOTP$.

3.2.12.2 A Hybrid Model Approach

As shown in the previous section, an isolated regression tree approach is no good candidate for punctuality modelling. However, an intelligent combination of multivariate linear regression and regression trees, enhanced by an AR(1) component, might still improve model quality. The idea is to use multivariate linear regression as the basic mathematical approach. MLR proved to give reasonable modelling results for a wide range of punctualities. Only on low punctuality days, i.e. days with $TOTP$<0.5, MLR showed a tendency of overestimating punctualities. In the following, the focus is on an improvement of model results for days with punctualities lower than 0.5. In Section 3.2.12.1 it was found that an approach with one single regression tree will not lead to this desired improvement as low punctualities are generally badly modelled by a single tree.

Instead of using one tree, several trees are generated, using the background information collected and discussed in Section 3.1 as a starting point. Each of these special regression trees is constructed from a limited dataset which is specified below, only. The limitation criteria are extracted from Table 3.1 in the following way. A rank coefficient ψ_i is introduced for each criterion:

$$\psi_i = 100 \cdot \frac{N_{i,0.5}}{N_i} \cdot (1 - \text{mean}(TOTP_i)) \tag{3.2}$$

with mean($TOTP_i$) being the average punctuality on those days where criterion i was fulfilled, N_i the number of days where where criterion i was fulfilled, and $N_{i,0.5}$ the number of days where criterion i was fulfilled and punctuality was lower than 0.5. In the following, all criteria were ordered in decreasing order after their rank coefficients. The 15 criteria heading this

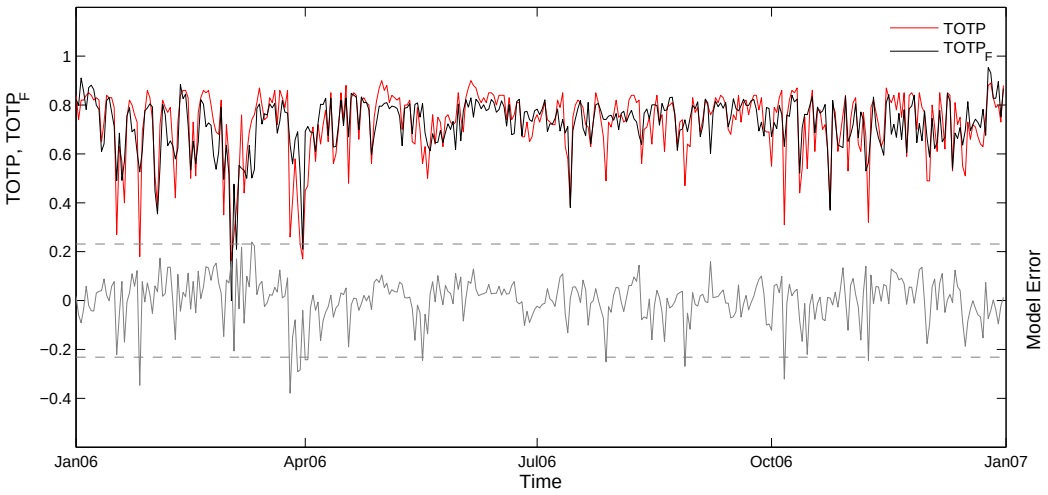


Figure 3.7: Time series of $TOTP$, $TOTP_F$ and the model error for Model 12a. Dashed lines give twice the standard deviation of $TOTP$.

ranking, representing weather events typical for low punctuality days, were selected for the construction of special trees. For each of the 15 special regression trees, only those days were used, where the respective criterion was fulfilled. This way, rather small and clear regression trees were obtained. The general construction procedure is described in Section 2.3.2. The resulting ranking is found in Appendix B.1. A respective ranking was also compiled for high resolution weather data. The associated ranking table is shown in Appendix B.2.

It should be noted that special regression trees solely will not provide a punctuality value for each day, but only for those days where one of the criteria listed in Tables B.1 or B.2, respectively, is fulfilled. Thus, special regression trees cannot be used in a standalone model. Rather, special regression trees may offer an alternative punctuality value, which might be better or worse than the corresponding MLR value. If more than one of the 15 special regression trees potentially offers an alternative $TOTP$ value, the topmost special regression tree in the ranking is considered, only.

The final hybrid punctuality model is now constructed the following way. As special regression trees shall only be used to potentially correct punctualities obtained by MLR, two limits and one additional decision rule are introduced. The first limit will in the following be referred to as *MLR Correction Limit* (*MLRCL*), the second as *Regression Tree Correction Limit* (*RTCL*). The *MLRCL* will be assigned the value 0.6, meaning that a regression tree value will only be considered as a potential replacement candidate for a punctuality value obtained by MLR, if that MLR value is smaller than 0.6. This is considered a "signal". In order to avoid upward misreplacements, i.e. upward corrections of rather reliable MLR values by less reliable regression tree values, it is claimed that, in order to be used, the regression tree value has to

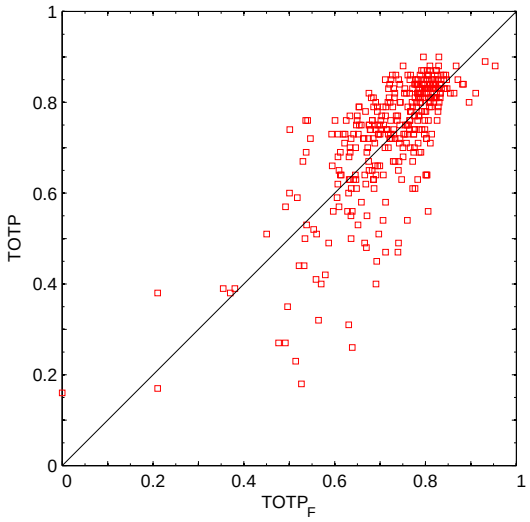


Figure 3.8: Scatterplot of $TOTP$ and $TOTP_F$ for Model 12a.

be smaller than the MLR value. Furthermore, the second limit, the $RTCL$, is consulted. The regression tree correction limit is assigned the value 0.5, i.e. an MLR value is only replaced if it is smaller than 0.6 and the alternative regression tree value is smaller than 0.5 and also smaller than the MLR value. This way it is guaranteed that the shortcomings of the two approaches – multivariate linear regression and regression trees – are eliminated as much as possible, but their potentials be combined in an intelligent way.

In order to demonstrate the efficiency of the hybrid model approach, it is applied to both low (Model 12a) and high (Model 12b) resolution predictor variables. An AR(1) component is initially not considered. Thus, comparison models are Models 8 and 9, respectively. Figure 3.7 shows the time series produced by Model 12a for the validation period, using the set of low resolution predictor variables brought together in Table 3.27. Figure 3.8 shows the corresponding scatterplot. Predictive model quality is generally good. Fluctuations in the timeseries are well reproduced, as well as minima and maxima. Still, there is a tendency of generally overestimating low punctuality days. However, a considerable amount of these low punctuality days are well reflected.

Altogether, the regression tree correction algorithm replaced four MLR based punctuality values, whereat four different special regression trees (see Table B.1) finally came into play. Obviously, the very selective correction algorithm is not dedicated to replace a large amount of values, but to take corrective action in a few cases. However, this is done in a very effective way. Table 3.41 shows the improvements that could be realised through the four replacements mentioned above, only. Compared to Model 8, R^2 has increased from 0.541 to 0.564. Also model errors have decreased remarkably. For completion, results are also given when an additional AR(1) component

Table 3.41: Prognostic quality criteria for Model 12a.

	Model 8		**Model 12a**	
	without AR(1)	**with AR(1)**	**without AR(1)**	**with AR(1)**
MAE	0.067	0.064	0.066	0.063
RMSE	0.092	0.089	0.090	0.089
SE	0.095	0.092	0.092	0.091
$r_{multiple}$	0.735	0.755	0.751	0.760
R^2	0.541	0.570	0.564	0.578

is used. Thus, Model 8 is enhanced the same way as Model 11 (see Section 3.2.11), where high resolution predictors are used. Results obtained by Model 12a with AR(1) component are even better than results obtained without the AR(1) component and almost of the same quality as Model 11, where high resolution predictors are used.

Taking the high resolution Model 9 as a starting point for Model 12b and falling back on Table B.2 as the guideline for special regression tree application leads to marginal improvements in diagnostic, but no further improvements in prognostic model quality as no value replacements were effected in prognostic mode. On the one hand, this effect is due to the stringent prerequisites for value replacements. On the other hand, there are many more potential predictor variables when high resolution predictors are used. Thus, the set of criteria for special regression trees is much more selective and prerequisites for the application of one of the special regression trees are less often met. In the present case of using 2006 as the validation period and falling back on the fixed set of predictor variables taken from Model 9, only Tree 9 (see Table B.2) comes into consideration at all. In that context, it is not advisable to extend the set of special regression trees in order to allow for more trees to potentially come into play, as trees further down the ranking are less representative for low punctuality days.

Despite the latter results it is convenient to stick to the hybrid model approach as described above. Changes in calibration or validation periods or the set of predictor variables, e.g. because of a change in model setup or availability/unavailability of data for certain predictors, might as well lead to quality improvements in the high resolution model, just as realised in the low resolution model. For example, when AMDAR based predictors are included in the set of potential predictor variables, the final set predictors might differ significantly from the set obtained for Model 9. As many of the special regression trees listed in Table B.2 are built on AMDAR based criteria, this might lead to more special regression trees to come into play.

3.2.13 Model 13 – The Final Hybrid Model

In this section, the final hybrid Model 13, which is based on Model 12b and which includes the regression tree correction algorithm, is specified and discussed. Up to this point, many model improvements have been implemented, but no decision has yet been made regarding the final number of predictor variables. So far, a number of 20 predictors was chosen for each intermediate model without further discussion. This section shall, therefore, start with a suggestion for the final number of variables to be included in the model.

In order to establish a basis for decision making, the 5-year calibration period 2001-2005 and 2006 as the validation period were used. The number of predictor variables was, based on predictors determined in Model 9, one by one reduced and model quality criteria were calculated. As a removal criterion, p-values associated to predictors (see Section 2.3.1) were adducted. The predictor exhibiting the highest p-value in the current, adapted model was considered for removal at the next step. The results obtained are brought together in Table 3.42. Not unexpectedly, weather predictors related to wind, (solid) precipitation and low visibility are among the last variables to be removed, indicating their high significance. With focus on prognostic performance, the model with 16 final predictors shows the best results with the

Table 3.42: Quality criteria for models with reduced numbers of predictors.

no	para removed	p-value	diagnosis		prognosis	
			R^2_{adj}	MAE	R^2	MAE
20	/	/	0.652	0.048	0.587	0.063
19	fx24_max	0.146	0.650	0.049	0.587	0.063
18	Wt_b1_MEAN_1	0.017	0.650	0.049	0.588	0.063
17	ff_MEAN_3	0.009	0.649	0.049	0.588	0.063
16	ww31b_1	1.39×10^{-4}	0.648	0.049	0.590	0.063
15	Ws_b_mean_1	6.33×10^{-5}	0.644	0.049	0.577	0.063
14	Wt_a1_mean_3_^4	2.61×10^{-6}	0.644	0.049	0.576	0.064
13	VV_MEAN_1_1/x	3.31×10^{-7}	0.635	0.050	0.559	0.065
12	E1_3	1.41×10^{-9}	0.631	0.050	0.566	0.065
11	TOTFL_scheduled _minus_TOTFL	2.01×10^{-12}	0.622	0.050	0.556	0.065
10	ww32b_3	4.41×10^{-14}	0.609	0.051	0.554	0.066
9	Wt_b2_lim_1	1.94×10^{-16}	0.593	0.052	0.545	0.066
8	RRR_2_sqrt	1.87×10^{-26}	0.565	0.053	0.527	0.067
7	ww31b_2	4.31×10^{-29}	0.530	0.055	0.467	0.071
6	TOTFL_scheduled	2.43×10^{-27}	0.505	0.056	0.435	0.072
5	SH_MAX_3	7.62×10^{-19}	0.476	0.057	0.406	0.074
4	ff_mean	2.49×10^{-34}	0.428	0.059	0.377	0.074
3	ww34b_2	1.57×10^{-51}	0.375	0.062	0.293	0.077
2	DayIndex	2.41×10^{-125}	0.220	0.071	0.218	0.082
1	Hoehenwinde	4.46×10^{-18}	0.200	0.073	0.154	0.086

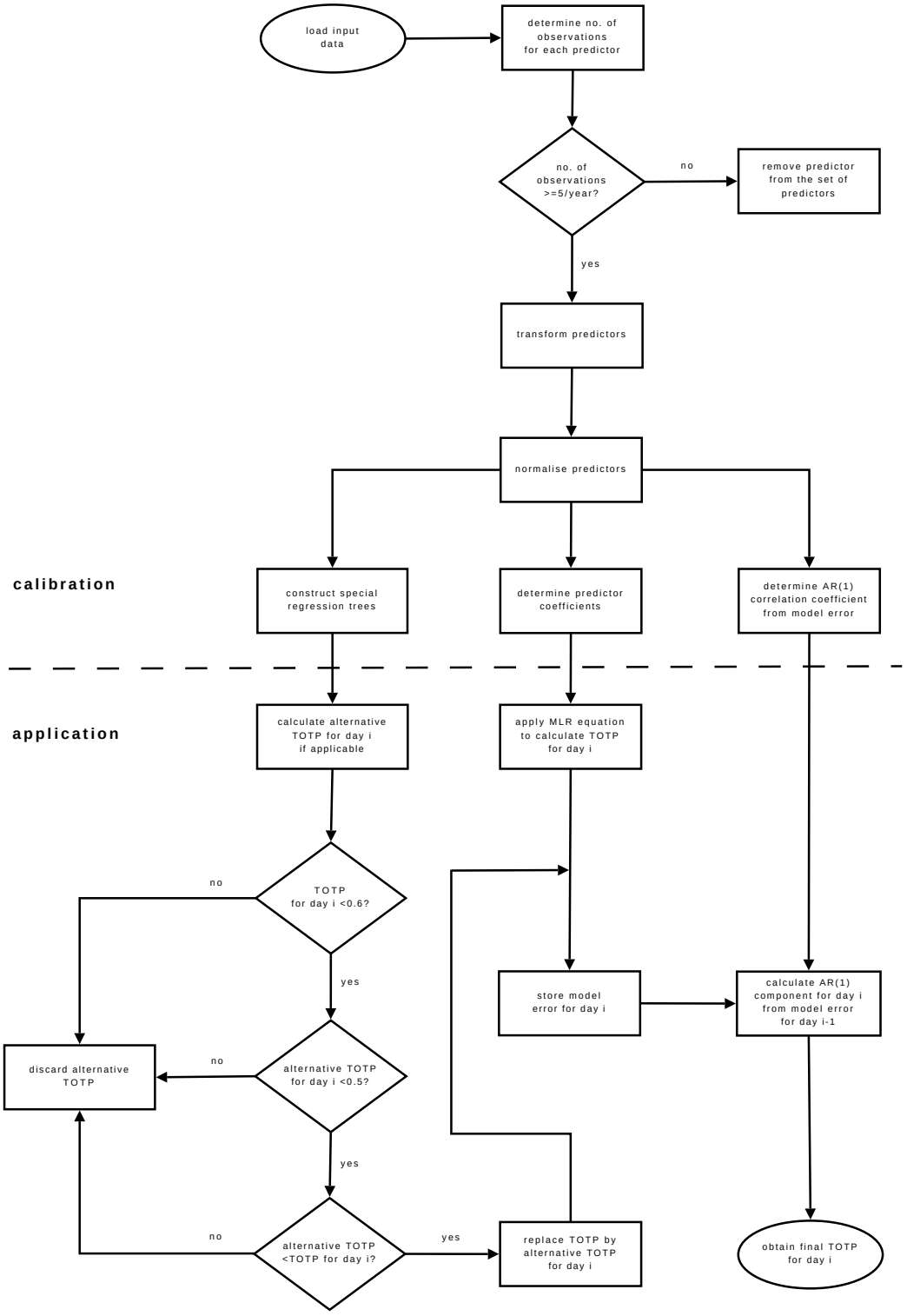


Figure 3.9: Flow diagram of the final hybrid punctuality Model 13.

Table 3.43: Fixed set of predictors for Model 13.

predictors	
const	ff_mean
DayIndex	Ws_b_MEAN_1
TOTFL_scheduled	RRR_2_sqrt
ww31b_2	VV_MEAN_1_1/x
ww34b_2	Wt_b2_lim_1
Hoehenwinde	E1_3
TOTFL_scheduled_minus_TOTFL	SH_MAX_3
ww32b_3	Wt_a1_MEAN_3_^4

highest R^2. Models with less predictors exhibit worse R^2- and MAE-values, with a significant decrease in model performance when only 7 or less predictors are used. Thus, it is reasonable to limit the number of predictors in the final Model 13 to the 16 variables listed in Table 3.43. This mix exhibits a good compromise between reproducibility and generalisability.

Figure 3.9 visualises the functionality and the interactions within the final Model 13 in a flow diagram. Note in the application domain the separate calculation of a $TOTP$ value applying MLR and, if applicable, applying regression trees (alternative $TOTP$). Figures 3.10 and 3.11 demonstrate the prognostic performance of the model through time series and scatterplots. Quality criteria are given in Table 3.44. Model 13 is significant on the 5.83×10^{-39} level. Still, there is a slight tendency of the model to better cope with higher punctualities, i.e. to give better modelling results when modelled punctualities are high. This effect, also referred to as *heteroscedasticity*, has already been discussed in the previous sections. It is visualised in Figure 3.13. Both for diagnosis and prognosis, the residuals, i.e. $TOTP\text{-}TOTP_{M,AR1}$ and $TOTP\text{-}TOTP_{F,AR1}$, respectively, are plotted against modelled punctuality. Obviously, in the high punctuality domain, residuals are smaller than in the low punctuality domain, both for diagnosis and prognosis. According to Backhaus et al. (2003), this is called heteroscedasticity of Type II. However, when splitting off modelled punctualities larger than 0.7, this effect diminishes and residual variance is almost constant. Thus, through several

Table 3.44: Quality criteria for Model 13.

quality criterion	diagnostic	prognostic
MAE	0.049	0.063
$RMSE$	0.068	0.087
SE	0.069	0.089
$r_{multiple}$	0.807	0.768
R^2	0.651	0.590
R^2_{adj}	0.648	/

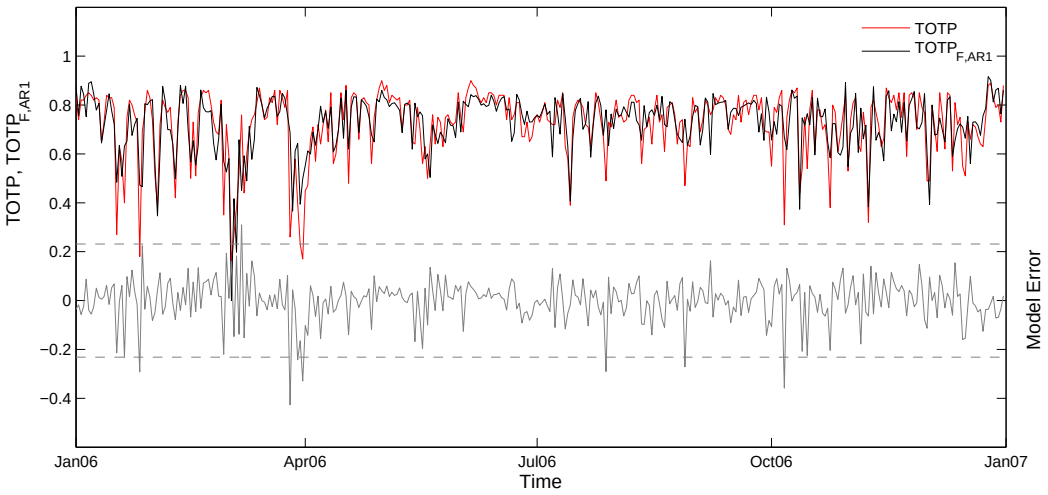


Figure 3.10: Time series of $TOTP$, $TOTP_F$ and the model error for Model 13. Dashed lines give twice the standard deviation of $TOTP$.

model enhancement steps, a significant improvement could be achieved in modelling medium to low punctuality days. High punctuality days, i.e. days with punctualities larger than 0.7-0.75 are generally very well modelled on a higher quality level than lower punctuality days.

Another important issue to look at are correlations among the predictors used in the final model. In order to achieve stable modelling results, perfect multicollinearity is to be avoided. For the detection of pairwise dependencies, the correlation matrix for all predictors is calculated and visualised in Figure 3.12. Note that the constant is excluded from the correlation matrix. The predictors in the correlation matrix are numbered and ordered

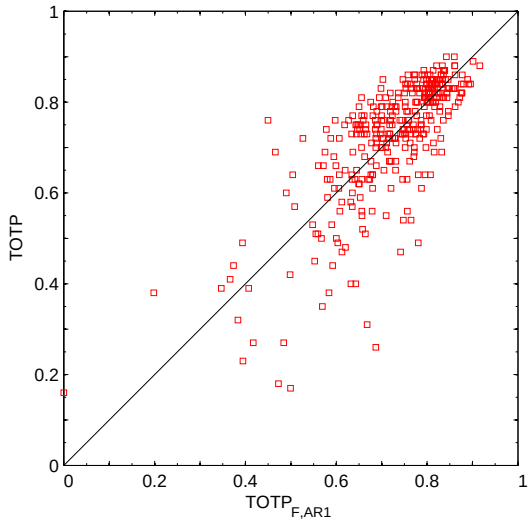


Figure 3.11: Scatterplot of $TOTP$ and $TOTP_F$ for Model 13.

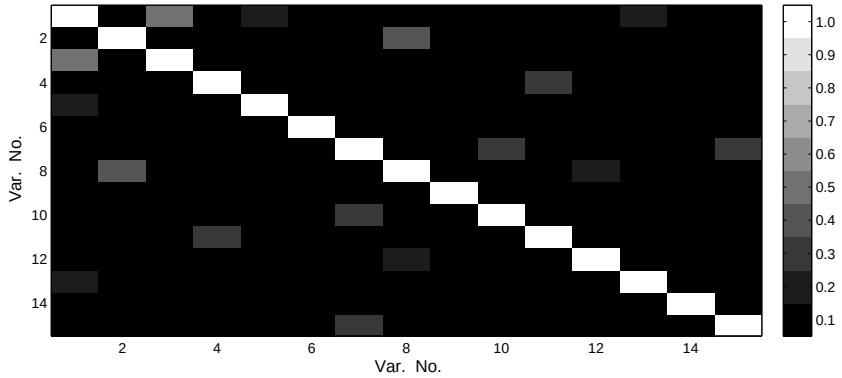
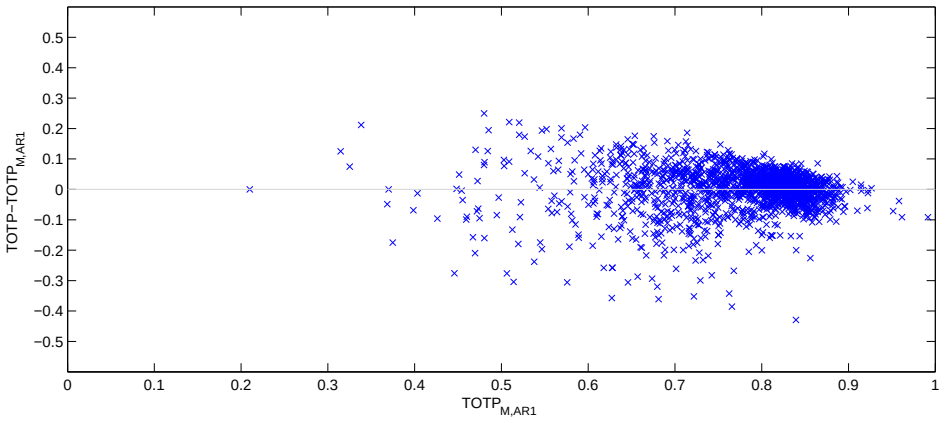


Figure 3.12: Visualisation of the predictors correlation matrix, Model 13.

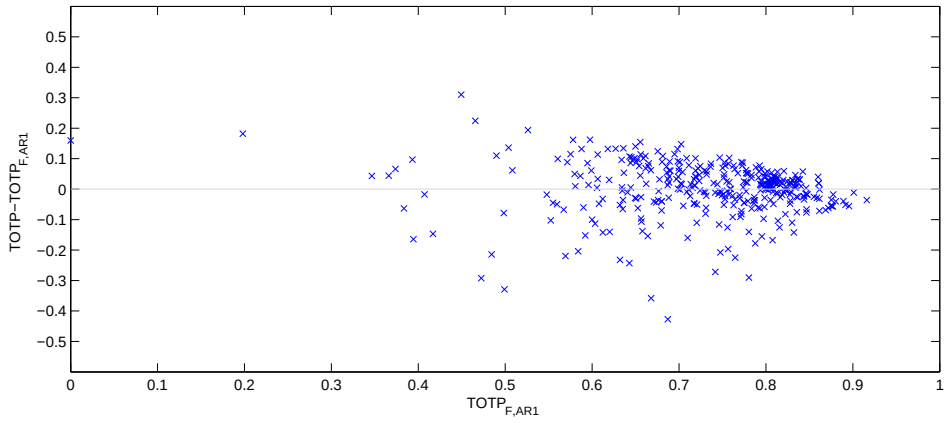
according to the labelling of the x-axis in Figure 3.14, ignoring the constant. Obviously, pairwise correlation is no issue among the 16 predictors. The highest correlation with $r = 0.52$ is between predictor 1 (*ff_mean*) and predictor 3 (*Ws_b_MEAN_1*). The remaining correlation coefficients are below 0.5. This is a good result, especially against the background that no perfect multicollinearity is generated through linear combination of predictor variables, either. Thus, multicollinearity is in general no issue for the punctuality model at hand.

A critical matter in the context of multivariate linear regression is the stability of predictor coefficients. Different calibration periods and lengths might in the worst case lead to very different β-coefficients, meaning that the coefficients are hard to interpret. Special attention has to be paid if coefficient sign changes occur under use of alternative calibration datasets. Sign changes are an indication that the corresponding predictor variable might rather be excluded from the set of predictors as its effect on punctuality is sometimes positive and sometimes negative. This is of course not desirable.

In order to determine coefficient stability, the resampling method of bootstrapping (see e. g. JUDGE et al., 1988) was applied to the dataset, using 1000 simulations and 65 % random days from the calibration dataset for each simulation run to determine the values of the beta coefficients. Figure 3.14 shows the results of these simulations for each predictor. Given in filled dark green circles are the β-coefficients calculated for Model 13. The unfilled light green circles are the mean of the β_i obtained by the simulations. The light green errorbars give $\pm$ twice the standard deviation of the β-coefficients from the simulations. Obviously, all predictor coefficients are very stable, not exhibiting much variation. Mean β_i from the simulations and β_i from Model 13 are almost identical for all variables. Moreover, there are no sign changes, indicating that there is an unambiguous effect of each predictor on punctuality. Beta coefficients are discussed in more detail in Section 3.2.13.2.



(a)



(b)

Figure 3.13: Residuals vs. modelled *TOTP* for identification of heteroscedasticity, Model 13. a) diagnostic mode, b) prognostic mode.

3.2.13.1 Modifications of Model 13

In this section, modifications of Model 13 are discussed, in order to demonstrate general model stability. In the following, two modifications are made:

1. A variation of the validation period,
2. A reduction of the calibration period.

First, the results through use of different validation periods shall be discussed. Table 3.45 lists model quality criteria for models with different calibration/validation datasets. The length of the calibration period is 5 years, each, the length of the validation period is 1 year, respectively. The year used for validation is given in the header of Table 3.45 and removed from the calibration dataset. Apparently, both diagnostic and prognostic model

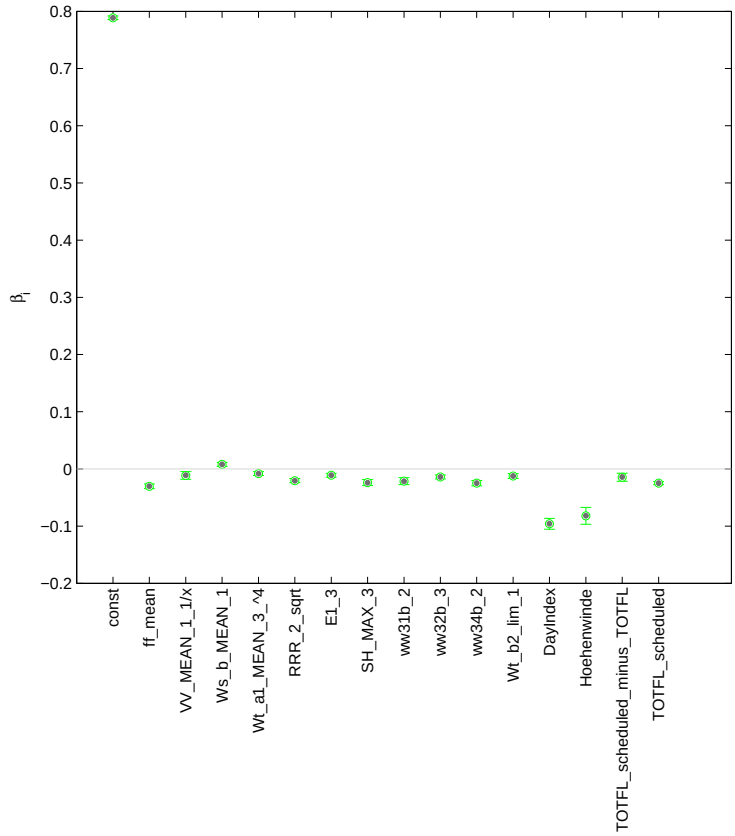


Figure 3.14: Stability of predictor coefficients, Model 13.

performance are subject to variability. Altogether, diagnostic model performance is by far less variable than prognostic model performance. Prognostic R^2-values are between 0.573 for 2001 and 0.682 for 2004. Most notably, for 2003 and 2004 as validation periods, respectively, prognostic model performance is better than diagnostic model performance for the remaining calibration dataset. Again, this is to be explained with the variability in *TOTP*. Especially 2003 exhibited rather good punctuality performance with only a small fraction of low punctuality days which are harder to model. In general, prognostic model performance for the punctuality model at hand is determined by the amount of low punctuality days in the validation dataset, on the one hand, and how well representative as well as exceptional events and their correlation to punctuality are captured through the calibration dataset, on the other hand. For later analyses, 2006 is in the following used for validation. With focus on prognostic performance, this is, according to Table 3.45, a conservative estimate of model performance without running the risk of giving overly optimistic results.

The second issue to analyse is to which extent the choice of the length of the calibration period influences the modelling results. For the present

Table 3.45: Quality criteria for models with different calibration/validation periods. The years in the column header are the years used for validation, remaining years from the period 2001-2006 are used for model calibration.

quality criterion		2001	2002	2003	2004	2005	2006
MAE	diagnostic	0.051	0.051	0.052	0.051	0.051	0.049
	prognostic	0.051	0.051	0.046	0.050	0.055	0.063
RMSE	diagnostic	0.072	0.071	0.073	0.072	0.071	0.068
	prognostic	0.072	0.072	0.065	0.070	0.078	0.087
SE	diagnostic	0.073	0.072	0.073	0.072	0.071	0.069
	prognostic	0.073	0.074	0.067	0.071	0.080	0.089
$r_{multiple}$	diagnostic	0.808	0.810	0.802	0.800	0.805	0.807
	prognostic	0.757	0.767	0.812	0.826	0.787	0.768
R^2	diagnostic	0.653	0.656	0.643	0.640	0.649	0.651
	prognostic	0.573	0.588	0.660	0.682	0.620	0.590
R^2_{adj}	diagnostic	0.650	0.654	0.641	0.637	0.645	0.648
	prognostic	/	/	/	/	/	/

punctuality modelling, we assume stationarity. In practice, this assumption is of course not perfectly satisfied as airport organisation and configuration is adjusted on occasion. Hence, there are basically two competing effects. On one hand, shorter and thus more actual calibration periods should better reflect the current airport configuration and the impact of certain events on punctuality. Longer calibration periods, in that context, might be less suited to cope with adaptations in the regulatory and operational framework. On the other hand, longer calibration periods should, in general, give more stable modelling results, as e.g. unusual weather events are more likely to be

Table 3.46: Quality criteria for models with reduced calibration periods (based on Model 13).

quality criterion		5 years	4 years	3 years	2 years	1 year
MAE	diagnostic	0.049	0.049	0.049	0.049	0.047
	prognostic	0.063	0.063	0.062	0.062	0.064
RMSE	diagnostic	0.068	0.069	0.068	0.069	0.068
	prognostic	0.087	0.087	0.086	0.087	0.089
SE	diagnostic	0.069	0.069	0.068	0.070	0.070
	prognostic	0.089	0.089	0.088	0.089	0.091
$r_{multiple}$	diagnostic	0.807	0.811	0.824	0.829	0.834
	prognostic	0.768	0.772	0.778	0.774	0.760
R^2	diagnostic	0.651	0.658	0.680	0.688	0.695
	prognostic	0.590	0.597	0.605	0.600	0.578
R^2_{adj}	diagnostic	0.648	0.654	0.675	0.680	0.680
	prognostic	/	/	/	/	/

part of the calibration dataset and their effect is thus reflected in the model coefficients. Moreover, the impact of events is set on a wider base.

Table 3.46 shows the changes in diagnostic and prognostic model performance when the model calibration period is successively shortened. The year 2006 is used for validation, years used for calibration are the latest years, always. Not surprisingly, the highest diagnostic R^2-values are achieved using short calibration periods. With focus on prognostic results, a calibration period of 3 years turned out to be most efficient with R^2 at 0.605 and MAE at 0.062. The reason for 3 years being the optimal calibration length under the given basic conditions is not further investigated in this study. According to these findings, the further use of a 3-year calibration period is suggested and it is in the following also used for the final punctuality model.

3.2.13.2 The Role of Weather

This section is dedicated to the analysis of the weather impact on punctuality. First, the importance of weather-related predictors among each other and in comparison with non-weather related predictors is discussed. As all variables are normalised, this can be done through comparison of beta coefficients, which for normalised data are also referred to as *beta weights*. According to Garson (2009), beta coefficients only say something about the unique contribution of each independent variable, but not the joint contributions, which are reflected in R^2 only. Thus, by solely looking at beta weights, one might underestimate the importance of a variable exhibiting strong joint contributions to R^2, but without strong unique contribution. To account for this effect, correlations between each predictor and $TOTP$ are to be calculated as well.

Furthermore, beta coefficients are only valid in the respective model at hand. That means, adding or removing predictors or changing the setup of the multivariate linear regression approach is likely to lead to different betas. Thus, predictor variable impact is only to be interpreted against the background of the respective model used. In order to determine the role of

Table 3.47: Fixed set of predictors for Model 13L.

predictors	
const	ww34b
DayIndex	CLc2
TOTFL_scheduled	ww33b
SH_max	VV_mean_1/x
TOTFL_scheduled_minus_TOTFL	ff_mean
Hoehenwinde	Wt_a2_mean
E1	fx24_max
ww31b	Wt_b2_lim

Table 3.48: Quality criteria for Model 13L, using both the latest 5 (2001-2005) and 3 years (2003-2005) for calibration and 2006 for validation.

	2001-2005		2003-2005	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.049	0.063	0.048	0.064
RMSE	0.069	0.088	0.067	0.087
SE	0.070	0.090	0.067	0.089
$r_{multiple}$	0.800	0.762	0.829	0.765
R^2	0.640	0.581	0.687	0.585
R^2_{adj}	0.637	/	0.682	/

weather, it is essential to also include all relevant and significant non-weather related predictors in the model. Otherwise, estimates of beta coefficients for weather related predictors are biased, bearing the weight of relevant predictors omitted. In that context, it should be kept in mind that neither information on reactionary delay is included in the model as a predictor, nor information on ground processes/operations having an impact on punctuality.

In the following, two different model versions are discussed: One low resolution model with weather-related predictor variables on a 24-hour basis, only (see deduction below), and the high resolution Model 13. The low resolution model, for simplification named Model 13L, is, based on Model 8, deduced the same way as Model 13 in the beginning of Section 3.2.13 and Section 3.2.13.1. Likewise, an additional AR(1) correction term is used. Taking Table 3.27 as a starting point and successively eliminating the 4 predictors exhibiting the worst p-values (*Wt_b2_mean_^4*, *N_mean*, *Ws_a_mean_^2*, *CLc1*), as introduced in Section 3.2.13, Table 3.47 is obtained, giving the set of the 16 final predictors for the low resolution model.

For comparison with Model 13 (see Table 3.46), Table 3.48 brings together model quality criteria for Model 13L, both for the original 5-year calibration

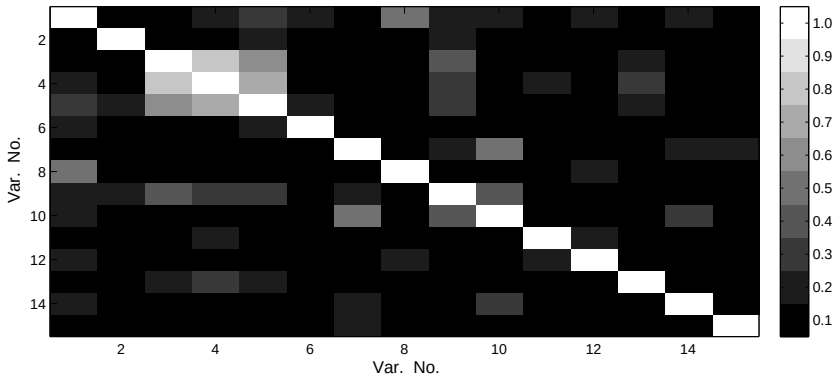


Figure 3.15: Visualisation of the predictors correlation matrix, Model 13L.

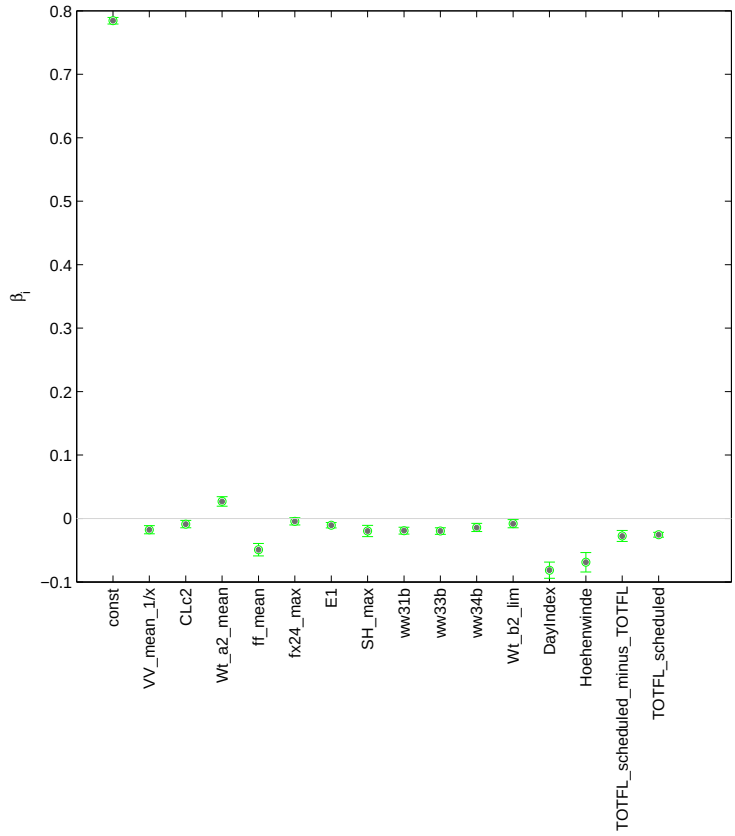


Figure 3.16: Stability of predictor coefficients, Model 13L.

period and the 3-year calibration period, found as an optimum for Model 13. Also for Model 13L, the 3-year calibration is preferable, exhibiting better model performance. The prognostic R^2 of 0.585 is only 2 percentage points lower than for the high resolution model. Based on these results, the low resolution model will, for the analysis of the weather impact, be calibrated using a 3-year calibration period.

Pairwise correlations among predictors used in Model 13L are unobjectionable with regard to multicollinearity. Highest correlations are between wind related predictors with a maximum of $r = 0.81$ between *ff_mean* and *Wt_a2_mean*. All correlations are visualised in Figure 3.15. Parameter coefficient stability for Model 13L is shown in Figure 3.16.

Table 3.49 lists all predictors used in Model 13L together with their corresponding p-values, β-coefficients and correlations with the predictant *TOTP*, in descending order of β_i. Additionally, mean values and standard deviations of the $\beta_{i,B}$, obtained from the bootstrapping algorithm, are given. The rather high p-value of 0.26 for *fx24_max* indicates that this predictor would be marked for removal through a significance test on an α-level of 0.05. Its direct impact on *TOTP* is, however, low, expressed by a low beta weight of

Table 3.49: Criteria for interpretation of predictor relevance, ordered according to β_i, Model 13L. Also given are p-values, the correlation coefficient r between the predictor and *TOTP* as well as mean and standard deviation of the $\beta_{i,B}$ obtained from bootstrapping.

predictor	**p-value**	r	β_i	$\varnothing\ \beta_{i,B}$	**std**($\beta_{i,B}$)
const	0	/	0.7844	0.7846	2.54×10^{-3}
Wt_a2_mean	7.54×10^{-8}	-0.194	0.0269	0.0267	3.86×10^{-3}
fx24_max	0.26	-0.238	-0.0044	-0.0044	2.89×10^{-3}
Wt_b2_lim	7.33×10^{-3}	-0.265	-0.0082	-0.0083	3.25×10^{-3}
CLc2	1.05×10^{-3}	-0.064	-0.0088	-0.0088	2.92×10^{-3}
E1	7.78×10^{-6}	-0.150	-0.0105	-0.0105	2.32×10^{-3}
ww34b	8.00×10^{-8}	-0.403	-0.0143	-0.0143	3.05×10^{-3}
VV_mean_1/x	4.57×10^{-8}	-0.340	-0.0178	-0.0178	3.25×10^{-3}
ww31b	9.18×10^{-10}	-0.255	-0.0191	-0.0192	2.66×10^{-3}
SH_max	2.59×10^{-8}	-0.294	-0.0196	-0.0196	4.11×10^{-3}
ww33b	3.20×10^{-10}	-0.399	-0.0197	-0.0195	2.61×10^{-3}
TOTFL_scheduled	9.13×10^{-23}	-0.047	-0.0256	-0.0255	2.00×10^{-3}
TOTFL_scheduled _minus_TOTFL	1.21×10^{-13}	-0.413	-0.0277	-0.0274	4.17×10^{-3}
ff_mean	7.27×10^{-16}	-0.364	-0.0491	-0.0489	4.95×10^{-3}
Hoehenwinde	8.32×10^{-241}	-0.247	-0.0689	-0.0691	7.67×10^{-3}
DayIndex	8.32×10^{-241}	-0.496	-0.0814	-0.0814	6.58×10^{-3}

-0.0044. Generally, all predictors exhibit a negative correlation with *TOTP*. Also, beta weights are, except for one predictor, negative. Solely *Wt_a2_mean* has a positive beta. That means, increasing the value of a *Wt_a2_mean* leads to a punctuality increase. Since normalised data is used, the beta weight is the average amount punctuality increases when the predictor corresponding to the respective beta, in this case *Wt_a2_mean*, increases one standard deviation and other predictors are held constant. The latter is important and helps to understand the, at first glance, unexpected sign of the beta weight for *Wt_a2_mean*. Generally, one would assume a negative sign, meaning that increasing winds lead to lower punctualities. However, there are five wind-related predictors in the set of predictor variables for this model. Thus, the wind impact on punctuality is always jointly reflected. Not unexpectedly, the remaining four wind-related predictors exhibit a negative beta. All other predictors, continuous, boolean oder enhanced boolean type, can logically be interpreted. Note that *VV_mean* is transformed and represented by its reciprocal. That means, higher visibility also leads to higher punctuality.

Generally, model betas and betas from the bootstrapping show no significant differences. Thus, model betas are used for further discussion. Standard deviations of the $\beta_{i,B}$ are between 2×10^{-3} and 5×10^{-3}, only *Hoehenwinde* and *DayIndex* stick out with higher values.

The highest correlations with *TOTP* are featured by *DayIndex*, *ww34b* and *TOTFL_minus_TOTFL_scheduled* with correlation coefficients higher

Table 3.50: Criteria for interpretation of predictor relevance, ordered according to β_i, Model 13. Also given are p-values, the correlation coefficient r between the predictor and *TOTP* as well as mean and standard deviation of the $\beta_{i,B}$ obtained from bootstrapping.

predictor	**p-value**	r	β_i	$\varnothing\ \beta_{i,B}$	**std**($\beta_{i,B}$)
const	0	/	0.7840	0.7841	2.51×10^{-3}
Ws_b_MEAN_1	1.02×10^{-4}	-0.080	0.0100	0.0100	2.31×10^{-3}
ww32b_3	8.24×10^{-6}	-0.036	-0.0097	-0.0097	2.14×10^{-3}
E1_3	3.20×10^{-7}	-0.114	-0.0113	-0.0112	2.19×10^{-3}
Wt_a1_MEAN_3_^4	5.42×10^{-5}	-0.136	-0.0138	-0.0138	3.88×10^{-3}
Wt_b2_lim_1	7.53×10^{-9}	-0.219	-0.0159	-0.0155	2.72×10^{-3}
VV_MEAN_1_1/x	6.26×10^{-8}	-0.287	-0.0166	-0.0173	4.07×10^{-3}
RRR_2_sqrt	4.91×10^{-14}	-0.227	-0.0191	-0.0189	2.68×10^{-3}
TOTFL_scheduled	7.80×10^{-16}	-0.047	-0.0199	-0.0200	1.73×10^{-3}
ww31b_2	3.58×10^{-13}	-0.267	-0.0199	-0.0197	3.75×10^{-3}
SH_MAX_3	2.05×10^{-11}	-0.298	-0.0233	-0.0234	4.09×10^{-3}
ww34b_2	3.61×10^{-24}	-0.384	-0.0243	-0.0243	2.68×10^{-3}
TOTFL_scheduled _minus_TOTFL	1.18×10^{-15}	-0.413	-0.0296	-0.0293	4.10×10^{-3}
ff_mean	1.40×10^{-29}	-0.364	-0.0335	-0.0335	2.89×10^{-3}
Hoehenwinde	1.22×10^{-245}	-0.247	-0.0793	-0.0792	7.36×10^{-3}
DayIndex	1.22×10^{-245}	-0.496	-0.0894	-0.0894	6.04×10^{-3}

than 0.4. However, interpretation of correlations of simple boolean predictors such as *Hoehenwinde* and *DayIndex* is difficult. Also *ww33b*, *ff_mean* and *VV_mean_1/x* exhibit a high correlation with *TOTP*. With regard to beta coefficients, highest absolut values, synonymous with the highest direct model impact on *TOTP*, are found for *DayIndex*, *Hoehenwinde*, *ff_mean*, *Wt_a2_mean* and the traffic-related predictors. Note again that *DayIndex* and *Hoehenwinde* are purely boolean and thus rather interpreted as an if/then-switch, indicating the effect of the presence or absence of the respective event.

To summarise, the most relevant predictors found are, besides *DayIndex*, those bearing information on traffic (*TOTFL_scheduled*, *TOTFL_minus _TOTFL_scheduled*), wind (*ff_mean*, *Wt_a2_mean*, *Hoehenwinde*), precipitation (*ww33b*, *ww34b*) and visibility (*ww31b*, *VV_mean_1/x*). Note that the joint effect of all windspeed-related predictors is a reduction of punctuality with increasing wind speeds. Valuable model input is also drawn from information on winterly weather conditions (*SH_max*, *ww34b*).

Table 3.50 gives results corresponding to Table 3.49, but for the high-resolution Model 13. Generally, results are very similar. However, through higher resolution of weather-related predictors, some predictor variables dropped out of the set of predictors and others were picked. For example, *ww33b_x* is not included for any time interval *x*. In the high resolution model, the precipitation impact is represented by *RRR_2_sqrt*. Amongst wind-related

predictors there were some replacements as well. Most notably, information on the presence of thunderstorms (*ww32b_3*) is now directly included.

Again, all predictor correlations with *TOTP* are negative. And again, there is one wind-related predictor (*Ws_b_mean_1*) with a positive beta weight. All remaining betas are negative. Coefficient stability is good, there are no significant deviations of β_i and mean($\beta_{i,B}$). Standard deviations of the $\beta_{i,B}$ are comparable with those obtained for the low resolution model. Besides *DayIndex*, *Hoehenwinde*, *ff_mean* and *TOTFL_minus_TOTFL_scheduled*, already discussed in the results for Model 13L, *ww31b_2*, *ww34b_2*, *SH_MAX_3* and *VV_mean_1_1/x* exhibit a high correlation with *TOTP*, closely followed by *RRR_2_sqrt* and *Wt_b2_lim_1*. Regarding beta weights, largest direct model impact is again through *DayIndex*, *Hoehenwinde*, *ff_mean* and the traffic-related predictors. Concerning high resolution weather predictors, solid precipitation (*ww34b_2*) and low visibility (*ww31b_2*) between 6 and 12 UTC, as well as snow height between 12 and 18 UTC exhibit high absolute betas. Also general precipitation (*RRR_2_sqrt*) between 6 and 12 UTC and low visibility in the early morning (*VV_mean_1_1/x*) have a large impact.

From the analyses of both the high and the low resolution models, the following can be summarised. For punctuality modeling, traffic is an important factor and should be part of the set of predictors. A *DayIndex* variable should be used as a scenario switch if unusual events occurred or are expected. This way, the modelling approach at hand also works for days with e.g. strikes or system failures and goes without the a-priori elimination of those days. With focus on weather related predictors, the following is found. Information on upper level winds emerged as non-negligible. In the set of predictors, there should be at least one variable representing general precipitation. In addition, a predictor related to solid precipitation or snow height proved to be of high relevance. Low visibility conditions have to be represented, as well, either through a fog related predictor or through general meteorological visibility or RVR, respectively. Information on frozen ground proved to be a valuable add on information. In contrast, a thunderstorm related predictor turned out to be of lower relevance. Obviously, critical weather phenomena connected to thunderstorms, like e.g. heavy precipitation or strong winds, are better separately represented by single predictors instead of a global thunderstorm predictor. Regarding wind-related variables, two predictors proved to be of high significance: the average daily wind speed (*ff_mean*) and the exceedance of the tailwind limit on RWY18 (*Wt_b2_lim*). Additional predictors describing head-/tailwind and crosswind relative to the runways proved to further increase model quality. Regarding cloud related predictors, only information on thunderstorm clouds (*CLc2*) turned out to be relevant.

After discussion of the significance of weather and non-weather-related predictors, it is shortly analysed how well a pure weather based model performs compared to a full model as above. In that context, the focus is on

Table 3.51: Fixed set of predictors for Model 13wL. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window. The stroke out predictors were eliminated through analysis of p-values, in order of the numbers attached.

predictors	
const (**100**, **100**, **100**)	RRR_mean_sqrt (34.7, 49.3, **100**)
SH_max (91.5, 98.0, **100**)	Wt_a2_mean (20.5, 48.5, **100**)
ww31b (75.2, **100**, **100**)	~~TT_mean~~^{2} (58.5, 61.2, **96.2**)
ww34b (73.3, **100**, **100**)	~~VV_min_log~~^{1} (23.5, 14.0, **89.9**)
E1 (68.0, **100**, **100**)	N_mean (17.8, 28.6, **80.9**)
Hoehenwinde (66.9, **100**, **100**)	Wt_b2_lim (35.9, 69.4, **77.3**)
DayIndex (59.7, **100**, **100**)	~~P0_mean~~^{4} (20.6, 35.1, **69.1**)
CLc2 (50.7, 99.5, **100**)	ww33b (32.7, 37.8, **63.1**)
VV_mean_1/x (49.8, **100**, **100**)	Ws_a_mean_^2 (46.7, **59.1**, 44.3)
ff_mean (45.6, 78.3, **100**)	~~Wt_b2_mean_^4~~^{3} (39.8, 30.6, **47.5**)

general model quality, e.g. represented by R^2, and not on single predictors and their individual impact on punctuality. The key question is: How much of the variability in *TOTP* can be explained by weather only? In order to answer this question, both Models 13 and 13L were reduced to finally include only weather-related predictors. Table 3.51 shows the final set of predictors for a low resolution weather-only model, named Model 13wL. Table 3.52 gives the respective set of predictors for a high resolution weather-only model (Model 13w). The procedure for variable selection is as described in the beginning of Section 3.2.13, with 20 predictors as a starting point and successive elimination of the four variables exhibiting the worst p-values (stroke out in Table 3.51). However, this time only a 3-year (2003-2005) calibration period was used at the outset, as this has already been found preferable (see Section 3.2.13.1). Note that *DayIndex* only includes information on the CAT stage, which, in turn, is based on visibility and cloud base.

Table 3.52: Fixed set of predictors for Model 13w. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window. The stroke out predictors were eliminated through analysis of p-values, in order of the numbers attached.

predictors	
const (**100**, **100**, **100**)	SH_MAX_1 (32.1, 40.9, **100**)
Hoehenwinde (70.7, **100**, **100**)	r1_MEAN_3_sqrt (28.2, 23.7, 99.5)
ww31b_2 (65.9, **100**, **100**)	Wt_a1_mean_3_^4 (45.8, 81.8, **97.0**)
ww34b_2 (62.1, **100**, **100**)	~~ff_MEAN_3~~^{1} (29.3, 46.6, **96.7**)
DayIndex (51.5, **100**, **100**)	RRR_1_sqrt (51.3, 53.7, **95.1**)
ff_mean (49.1, 77.3, **100**)	RR1_MAX_1_sqrt (43.0, 53.5, **95.1**)
ww32b_3 (45.5, 61.1, **100**)	Ws_b_MEAN_1 (19.7, 45.1, **94.0**)
E1_1 (42.6, 74.9, **100**)	VV_MEAN_1_1/x (27.4, 50.1, **93.7**)
~~RRR_2_sqrt~~^{4} (38.5, 49.6, **100**)	~~Wt_b1_MEAN_1~~^{2} (26.3, 42.2, **90.7**)
Wt_b2_lim_1 (34.7, 35.4, **100**)	~~ww31b_1~~^{3} (11.3, 22.3, **90.7**)

Table 3.53: Quality criteria for Model 13wL, without and with AR(1) component, using 2003-2005 for calibration and 2006 for validation.

	without AR(1)		with AR(1)	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.056	0.078	0.052	0.070
RMSE	0.081	0.107	0.075	0.098
SE	0.081	0.110	0.076	0.100
$r_{multiple}$	0.738	0.652	0.776	0.706
R^2	0.545	0.425	0.603	0.499
R^2_{adj}	0.538	/	0.596	/

Table 3.54: Quality criteria for Model 13w, without and with AR(1) component, using 2003-2005 for calibration and 2006 for validation.

	without AR(1)		with AR(1)	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.058	0.076	0.054	0.070
RMSE	0.082	0.105	0.077	0.096
SE	0.083	0.107	0.078	0.098
$r_{multiple}$	0.728	0.682	0.765	0.730
R^2	0.530	0.465	0.585	0.533
R^2_{adj}	0.522	/	0.578	/

The focus now is not on an interpretation of predictors chosen, but on model results that can be achieved using one of the above weather-only models. Using a 3-year calibration period (2003-2005), Model 13wL is significant on the 7.66×10^{-19} level. Quality criteria for this model are given in Table 3.53. Doing the same for the high resolution model, results shown in Table 3.54 are obtained. This model is significant on the 9.48×10^{-19} level. Comparing the results of both models, the low resolution model exhibits slightly better diagnostic performance. The high resolution model, however, performs much better in prognostic mode. In order to determine the variability in *TOTP* that can solely be explained by weather, the inclusion of an AR(1) correction algorithm is, in principle, not indicated, as general times series information in *TOTP* is evaluated. In the first place, this information is not directly related to weather. Thus, only the combination of multivariate linear regression and regression trees is initially used. In the broader sense, however, using an AR(1) correction term does not involve any further, non-weather related predictors. It is a pure mathematic correction approach which exploits the additional potential for explanation power, using only weather-related predictors and the predictant *TOTP* itself as an input. Thus, an AR-enhanced model is consulted for evaluations, as well.

Focusing on diagnosis results and falling back on the high resolution model, weather alone can explain more than 50 % of the variability in *TOTP*.

When focusing on the more realistic prognostic results, weather can still be attributed 46.5 % of the variability in *TOTP*. When also exploiting time series information through inclusion of an AR-term, a diagnostic R^2 of almost 0.6 and a prognostic R^2 of more than 0.5 can be achieved through an AR-enhanced weather-only model.

3.3 Punctuality Forecast

This section is dedicated to punctuality forecasting. The inherent question is: Can a modified punctuality model, based on the previously introduced models, be used for a punctuality forecast for a day in the future? Two issues evolve from this question. On one hand, it has to be shown, what quality level can be accomplished using a model in prognostic mode, when only predictable input variables are used. This is essential as, for a real forecast model, input for predictor variables has to be forecasted. With regard to weather-related predictor variables, this input can be obtained from NWP models, if necessary enhanced by a MOS system. Information on scheduled traffic is also available beforehand, as schedules are made on a long time basis and experience only slight short-term adaptations. Derivative predictors like *Mix* and *CAT*, as part of the predictor *DayIndex*, are based on weather-dependent thresholds and can thus potentially be forecasted, as well. The subsequent question is, how well weather events can actually be forecasted. In that respect, special focus is on critical event-driven predictors like *ww3xb* or problematic variables like e.g. visibility, where forecast quality is still unsatisfactory. The quality of NWP or MOS models, respectively, generally has a large impact on the quality of a punctuality forecast model. The worse the weather prediction and, accordingly, the larger the discrepancy between forecasted weather and weather occurred, the worse the forecast of punctuality. In that respect, weather forecast errors add to the model errors of the punctuality model itself.

Within this study, the question of weather forecast quality shall not be further discussed. Rather, the focus is on an analysis of the forecast potential of the methods introduced and the definition and specification of a final punctuality forecast model. This can then be used for a true forecast of *TOTP*, if input for predictor variables is provided for a day in the future. Within this section, the term ”forecast model” is in the following used for a punctuality model applied in prognostic mode using only predictable input variables as described above. The terms ”true forecast” and ”real forecast” are reserved for models which are fed with forecasted predictor variables, which are, in the case of weather-related predictors, taken from NWP/MOS output. This section focuses on forecast models. The enhancement step to a real forecast is only touched in Section 4.3.

Table 3.55: Fixed set of predictors for a low resolution Forecast Model not including predictors related to upper level wind. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window. The stroke out predictors were eliminated through analysis of p-values, in order of the numbers attached.

predictors	
const (**100**, **100**, **100**)	Ws_a_mean_^2 (39.3, 74.0, **100**)
TOTFL_scheduled (**100**, **100**, **100**)	Wt_a2_mean (34.7, 37.9, **100**)
SH_max (97.2, 99.5, **100**)	CLc1 (33.9, 71.4, **100**)
ww31b (72.8, **100**, **100**)	fx24_max (25.2, 49.2, **100**)
Wt_b2_lim (72.5, 97.9, **100**)	~~TT_mean~~[4] (30.7, 56.2, **98.4**)
E1 (69.6, **100**, **100**)	~~h_mean_log~~[2] (31.7, 62.7, **98.1**)
ww34b (65.2, **100**, **100**)	~~N_mean~~[3] (19.4, 56.2, **97.8**)
DayIndex (64.9, **100**, **100**)	RRR_mean_sqrt (33.6, 66.9, **95.6**)
ff_mean (55.6, 92.8, **100**)	VV_mean_1/x (36.1, **90.5**, 87.7)
ww33b (54.1, 94.0, **100**)	~~h_max~~[1] (3.1, 26.3, **86.3**)

In order to find the model best suited for a punctuality forecast, four model versions are discussed: two low and two high resolution models, with and without predictors related to upper level wind, respectively. Table 3.55 shows the final set of predictors for the low resolution model without upper level wind predictors. The procedure is as described in the previous sections, using the preferred 3-year calibration period 2003-2005 for general model calibration and for the elimination of the four predictors exhibiting the worst p-values, leaving 16 final predictor variables in each model. Note that in the forecast model, *DayIndex* only contains archived information on the CAT-stage as declared by local ATC. Table 3.56 gives the results for this model, with and without an AR(1) component, respectively. An AR-correction value can – theoretically – be calculated, if the model error is known for day *x-1* when forecasting punctuality for day x, using the lag-1 correlation coefficient obtained from the calibration procedure. The model at hand is significant on the 4.06×10^{-21} level.

Table 3.56: Quality criteria for a low resolution Forecast Model not including predictors related to upper level wind, without and with AR(1) component, using 2003-2005 for calibration and 2006 for validation.

	without AR(1)		**with AR(1)**	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.055	0.072	0.053	0.067
RMSE	0.078	0.101	0.074	0.093
SE	0.079	0.104	0.075	0.096
$r_{multiple}$	0.758	0.678	0.783	0.730
R^2	0.575	0.460	0.613	0.533
R^2_{adj}	0.568	/	0.607	/

Table 3.57: Fixed set of predictors for a low resolution Forecast Model including predictors related to upper level wind. The numbers in brackets are the percentage of selections for the moving 1-year, 2-year and 3-year window. The stroke out predictors were eliminated through analysis of p-values, in order of the numbers attached.

predictors	
const (**100**, **100**, **100**)	CLc2 (54.4, 80.2, **98.6**)
TOTFL_scheduled (**100**, **100**, **100**)	CLc1 (51.6, 81.8, **95.4**)
SH_max (99.5, **100**, **100**)	max_Tangentialwind_a_head_ge25
Wt_b2_lim (96.8, **100**, **100**)	(27.9, 49.4, **77.3**)
ww31b (85.0, 97.9, **100**)	ww33b (33.1, **71.4**, 56.8)
ww34b (76.6, **100**, **100**)	~~Ws_b_mean~~[4] (34.7, **71.3**, 44.5)
E1 (69.9, 99.9, **100**)	~~N_mean~~[2] (11.6, 19.2, **65.6**)
DayIndex (69.8, 94.1, **100**)	VV_mean_1/x (27.1, 55.4, **62.0**)
Ws_a_mean_^2 (68.1, 92.3, **99.7**)	max_Tangentialwind_a_head_ge15
~~h_mean_log~~[1] (55.1, 73.5, **98.9**)	(15.1, 40.1, **60.7**)
	max_Tangentialwind_a_tail_^2
	(35.2, 35.8, **59.3**)
	~~RRR_mean_sqrt~~[3] (23.5, 53.2, **54.6**)

Table 3.57 shows the fixed set of predictors for a low resolution forecast model, which contains predictors related to upper level wind. Table 3.58 gives the quality criteria for this model, which is significant on the 1.62×10^{-21} level. Both models, without and with information on upper level winds, achieve good results in prognostic mode. Diagnostic results are given for comparison only and not further discussed. Prognostic R^2-values are around 0.45 without an AR component and larger than 0.5 when an AR(1) correction term is included. Comparing the two low resolution models, the model without inclusion of predictors related to upper level wind performs better, consulting both model errors and R^2-values.

The two low resolution models are in the following compared with two high resolution forecast models. Table 3.59 lists the final set of predictors for a high resolution forecast model without predictors related to upper level

Table 3.58: Quality criteria for a low resolution Forecast Model including predictors related to upper level wind, without and with AR(1) component, using 2003-2005 for calibration and 2006 for validation.

	without AR(1)		**with AR(1)**	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.055	0.072	0.052	0.068
RMSE	0.077	0.102	0.074	0.095
SE	0.078	0.105	0.075	0.098
$r_{multiple}$	0.764	0.664	0.785	0.713
R^2	0.584	0.441	0.616	0.508
R^2_{adj}	0.577	/	0.610	/

Table 3.59: Fixed set of predictors for a high resolution Forecast Model not including predictors related to upper level wind. The numbers in brackets are the percentage of selections for the moving 1-year, 3-year and 5-year window. The stroke out predictors were eliminated through analysis of p-values, in order of the numbers attached.

predictors	
const (**100**, **100**, **100**)	RRR_2_sqrt (42.7, 55.3, **100**)
TOTFL_scheduled (**100**, **100**, **100**)	E1_1 (38.9, 91.2, **100**)
ww31b_2 (71.2, **100**, **100**)	E1_4 (33.1, 71.8, **100**)
ww34b_2 (65.0, **100**, **100**)	~~ww34b_4~~[2] (30.0, 44.3, **100**)
ff_mean (62.9, 80.7, **100**)	Wt_b2_lim_1 (17.4, 27.3, **100**)
~~RRR_1_sqrt~~[3] (58.7, 59.5, **100**)	SH_MAX_3 (54.8, 43.5, **99.7**)
DayIndex (55.9, 94.4, **100**)	VV_MIN_1_log (34.4, 57.5, **99.7**)
~~RR1_MAX_1_sqrt~~[4] (48.2, 59.1, **100**)	VV_MEAN_1_1/x (22.3, 56.2, **99.7**)
Wt_a1_MEAN_3_^4 (48.2, 85.5, **100**)	ww31b_1 (8.4, 41.9, **99.7**)
ww32b_3 (47.6, 74.8, **100**)	~~fx24_max~~[1] (31.7, 82.2, **98.9**)

wind. This model is significant on the 3.35×10^{-19} level. The results achieved with this model are given in Table 3.60. Prognostic performance of this high resolution model is better than for the respective low resolution model, independent of the use of an additional AR(1) correction component. Hence, for the purpose of punctuality forecast, the use of a high resolution model is recommended.

It remains to be analysed, if an implementation of predictors related to upper level wind leads to an additional increase of prognostic model performance. Table 3.62 shows the results for a high resolution forecast model including predictors on upper level wind, using the final set of predictors listed in Table 3.61. Note that two predictors describing upper level winds are part of this final set. This high resolution model is significant on the 4.17×10^{-19} level.

Unlike with the low resolution model, the inclusion of predictors related to upper level wind produces a significant improvement of prognostic model

Table 3.60: Quality criteria for a high resolution Forecast Model not including predictors related to upper level wind, without and with AR(1) component, using 2003-2005 for calibration and 2006 for validation.

	without AR(1)		**with AR(1)**	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.058	0.070	0.054	0.065
RMSE	0.081	0.099	0.077	0.091
SE	0.082	0.101	0.077	0.094
$r_{multiple}$	0.732	0.703	0.766	0.746
R^2	0.536	0.494	0.587	0.557
R^2_{adj}	0.529	/	0.581	/

Table 3.61: Fixed set of predictors for a high resolution Forecast Model including predictors related to upper level wind. The numbers in brackets are the percentage of selections for the moving 1-year, 2-year and 3-year window. The stroke out predictors were eliminated through analysis of p-values, in order of the numbers attached.

predictors	
const (**100**, **100**, **100**)	max_Tangentialwind_a_head_2_^4
TOTFL_scheduled (**100**, **100**, **100**)	(35.4, 32.4, **97.8**)
RRR_1_sqrt (89.1, 77.0, **100**)	E1_1 (38.1, 53.5, **93.2**)
ww31b_2 (77.0, **100**, **100**)	Wt_b2_lim_1 (38.2, 81.3, **92.9**)
ww34b_2 (69.4, 92.7, **100**)	~~CLc1_1~~[2] (26.5, 58.5, **92.9**)
RR1_MAX_1_sqrt (67.2, 84.8, **100**)	ww33b_3 (42.0, 76.2, **91.8**)
VV_MEAN_3_1/x (35.6, 95.1, **100**)	~~E1_4~~[4] (24.2, 55.4, **90.4**)
SH_MAX_3 (54.3, 55.5, **99.7**)	CLc2_2 (64.8, **89.1**, 76.5)
max_wind_gt35_3 (30.1, 59.6, **99.7**)	Wt_b2_lim_3 (20.5, 45.0, **87.7**)
~~ww34b_4~~[3] (40.8, 73.1, **97.8**)	VV_MIN_2_log (36.6, **87.6**, 82.2)
	~~Wt_a1_MEAN_1_^4~~[1] (8.8, 4.2, **81.4**)

performance. Even without the use of an AR(1)-correction, prognostic R^2 is at 0.531 and thus significantly higher than for both low resolution models. Using an additional AR(1)-correction term, prognostic R^2 experiences another significant step and reaches a value of almost 0.6, which is very close to results obtained for the full model introduced in Section 3.2.13. Diagnostic and prognostic results are, in that respect, on the same level. Also regarding prognostic model errors, this enhanced forecast model performs astonishingly well. *MAE* is at 0.064 and *RMSE* at 0.088.

Based on the results from the four forecast models, it is found that best prognostic performance is to be expected, when a high resolution model is used and information on upper level winds is included. Taking into account that the quality of the latter information could be further improved (also see analysis in Section 2.2.3.1), another increase in model performance can be assumed. Generally, the forecast of upper level winds is of high quality and

Table 3.62: Quality criteria for a high resolution Forecast Model including predictors related to upper level wind, without and with AR(1) component, using 2003-2005 for calibration and 2006 for validation.

	without AR(1)		**with AR(1)**	
quality criterion	**diagnostic**	**prognostic**	**diagnostic**	**prognostic**
MAE	0.058	0.069	0.054	0.064
RMSE	0.081	0.094	0.078	0.088
SE	0.082	0.096	0.078	0.090
$r_{multiple}$	0.733	0.729	0.761	0.766
R^2	0.537	0.531	0.579	0.586
R^2_{adj}	0.530	/	0.573	/

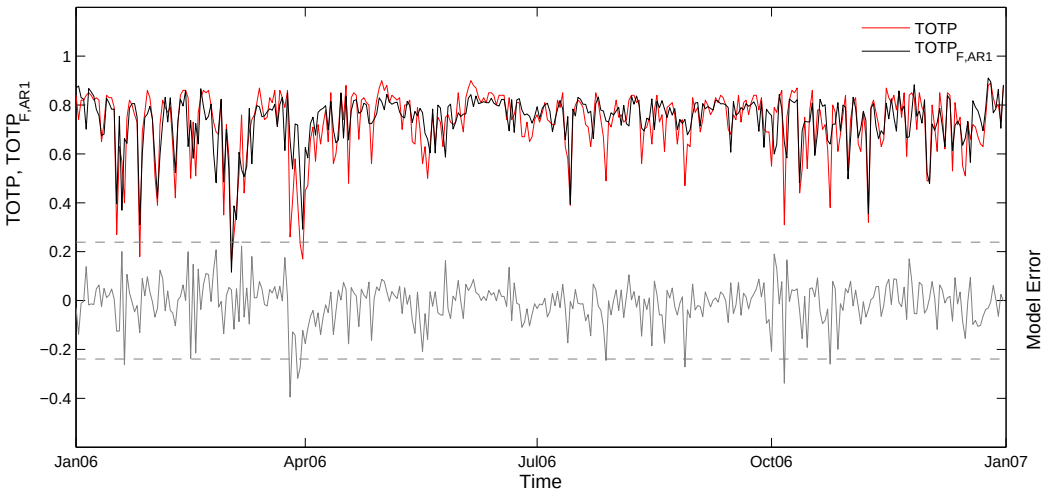


Figure 3.17: Time series of $TOTP$, $TOTP_F$ and the model error for the high resolution Forecast Model including predictors related to upper level wind. Dashed lines give twice the standard deviation of $TOTP$.

predictors can easily be extracted from NWP output.

To simplify matters, the final forecast model – the high resolution model with predictors related to upper level wind – is in the following referred to as *Forecast Model*. Figure 3.17 visualises the potential of the Forecast Model in a time series for 2006. Figure 3.18 shows even better that not only the high punctuality domain is well reproduced, but also in the low punctuality domain model results are on a good quality level.

Correlations among predictors, visualised in Figure 3.19, are generally

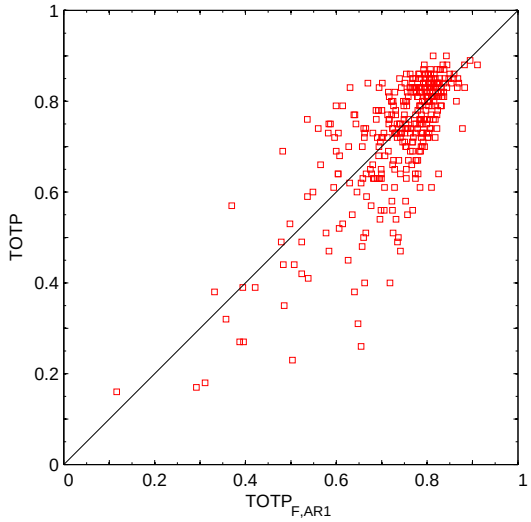


Figure 3.18: Scatterplot of $TOTP$ and $TOTP_F$ for the high resolution Forecast Model including predictors related to upper level wind.

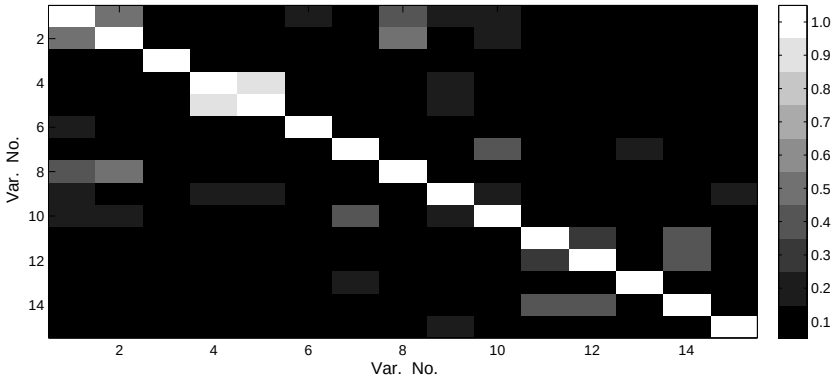


Figure 3.19: Visualisation of the predictors correlation matrix for the high resolution Forecast Model including predictors related to upper level wind.

unproblematic. Only predictors 4 (*RRR_1_sqrt*) and 5 (*RR_1_MAX_1_sqrt*) exhibit a high correlation with $r = 0.948$. Since both predictors are connected to the first time block, it is to be deliberate whether either of them is removed from the set of predictors. For our purpose, it is, however, not considered

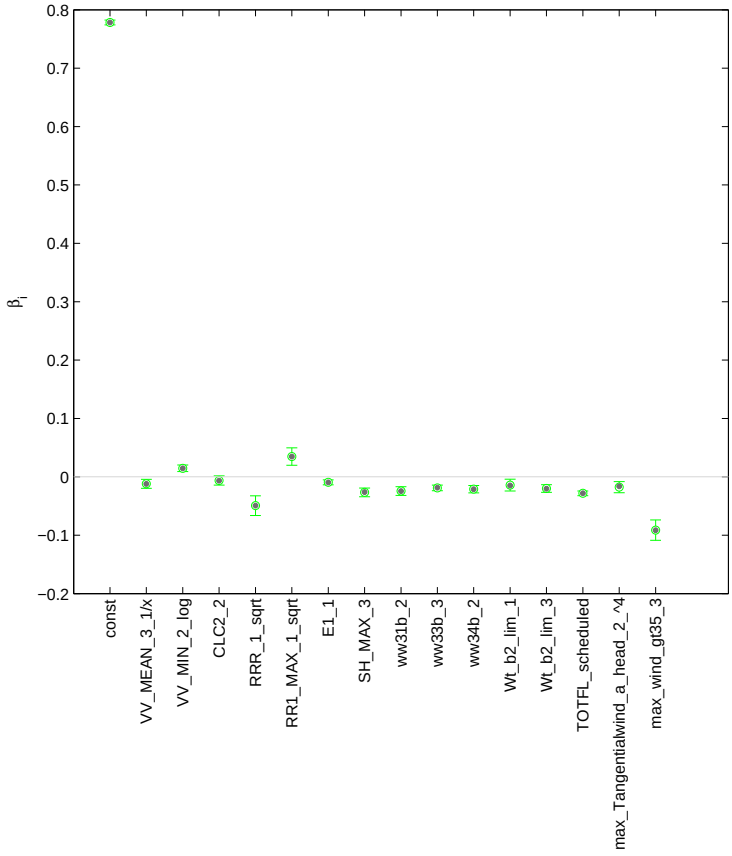
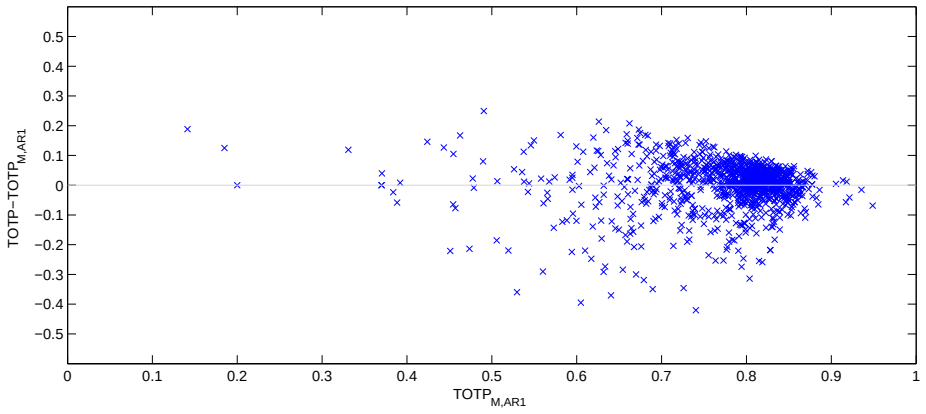
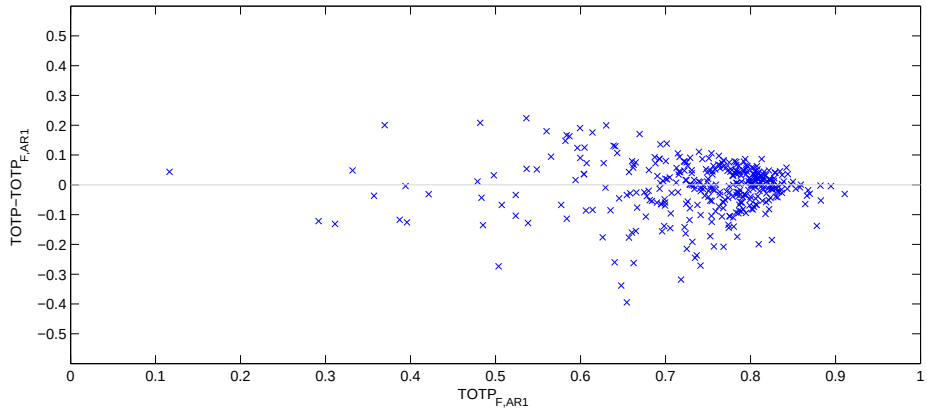


Figure 3.20: Stability of predictor coefficients for the high resolution Forecast Model including predictors related to upper level wind.



(a)



(b)

Figure 3.21: Residuals vs. modelled *TOTP* for identification of heteroscedasticity in the high resolution Forecast Model including predictors related to upper level wind. a) diagnostic mode, b) prognostic mode.

as both predictors exhibit high p-values of 1.37×10^{-8} and 2.42×10^{-5}, respectively.

Beta coefficients for the Forecast Model are stable, as shown in Figure 3.20. The underlying bootstrapping procedure is described in Section 3.2.13. Apart from *CLc2_2*, all betas point to a clear positive or negative impact on punctuality. For *CLc2_2*, model beta and the average beta obtained by bootstrapping coincide well, and the assumed effect on punctuality is clearly negative. Interestingly, there is a complex interaction of predictors related to precipitation in the first time block. As shown in the previous paragraph, those two predictors are well correlated. Thus, their effect on punctuality may not be interpreted by solely looking at their beta weights. Rather, they may only be considered together. Removing either of them is likely to result in a clearly negative beta weight for the remaining predictor.

The issue of heteroscedasticity has already been discussed in Section 3.2.13. Figure 3.21 shows the corresponding plots for the Forecast Model. Not surprisingly, heteroscedasticity of Type II is again found. However, general model quality is on a level, which allows for a meaningful punctuality forecast in all punctuality domains.

These encouraging results give rise to a successful implementation of an operational punctuality forecast system at major airports. For Frankfurt Airport, a graphical user interface has already been developed. It can be used for special punctuality analyses as well as for a true punctuality forecast – based on the Forecast Model introduced in this section – if forecasted input for predictor variables is provided.

Chapter 4

Conclusions, Limitations and Outlook

4.1 Conclusions

The present work deals with the subject of punctuality modelling. The statistical methods chosen are multivariate linear regression, regression trees and AR-models. Input information comprises weather, traffic, punctuality and operational data from Frankfurt Airport for a 5-year period. The study at hand represents a logical and consequent continuation of the work previously done in the area of punctuality modelling. It builds directly on ideas developed by Spehr (2003) and refines these to a point, where the quality of the models constructed finally allows for a true punctuality forecast. Suggestions for improvement made by Spehr were seized and, when proven meaningful, integrated in enhanced models. In this process, modelling was expanded to a larger database to test generalisability and stability.

The basic findings of the research done in this work can be summarised as follows. First of all, it was proven that total daily airport punctuality can be statistically modelled as a function of few, mostly meteorological predictor variables with an R^2 between 0.5 and 0.7. The final model is visualised in Figure 3.9. It is based on multivariate linear regression and uses an additional AR-correction term as well as a complex regression-tree correction algorithm. It was shown in Section 3.2.13 that modelling results are of good quality, independent of the time chosen for calibration and validation. In that respect, it could be abstained from the exclusion of certain days, e.g. days with non-weather related events like strikes or system failures, from the modelling data base, as done by Spehr. Rather, a complete modelling was aimed for and realised.

Through application of enhanced approaches, individually introduced in Sections 3.2.2-3.2.13, modelling results could be significantly improved over results achieved by Spehr. In prognostic mode, up to 60 % of the variabil-

ity of *TOTP* could be explained using a fixed set of only 16 high resolution predictor variables. These represent the most significant factors having an impact on punctuality at Frankfurt Airport. Likewise, other potential predictor variables were revealed as non-significant.

Furthermore, the relevance of weather with regard to punctuality was quantified without falling back on error-prone delay code systems. It was found that using weather-related predictor variables only, an R^2 of more than 0.45 can be achieved. Moreover, the weighting of impact factors, weather and non-weather related, was discussed against the background of the fixed set of predictors. As a result, it was found that at least five groups of predictors need to be represented when punctuality at Frankfurt Airport is to be modelled: precipitation, low visibility conditions, winds and upper level winds, solid precipitation/snow cover and traffic. Additional predictors describing weather in more detail lead to a higher model quality.

Based on these findings, the true forecast potential of the models developed was analysed. A conservative assessment revealed that almost 60 % of the variability in *TOTP* can be explained when only predictable variables are used as predictors. That means, model quality has for the first time reached a level where a true punctuality forecast is within the realms of possibility. In that regard it was found that a model calibration period of 3 years is favourable, exhibiting enough information for proper event representation, on the one hand, and timeliness with focus on schedule adaptations and operational process optimations, on the other hand.

4.2 Limitations

The present work represents a significant step in the understanding of the weather impact on air traffic delays. It avails oneself with independent analysis methods and is to be understood as a top-down approach to punctuality modelling. Unlike in a bottom-up approach where all individual processes are modelled, it does not draw upon either single process simulations, on the one hand, or evaluations based on delay codes, on the other hand. Compared to the bottom-up approach, the present work is of course a simplification, as processes are not looked at in detail. That means, in particular, that delay causing processes during aircraft turnaround are not directly captured. This is of course a weakness of the present approach as a significant delay impact factor – ground processes and operations – is not directly included.

A second drawback of the modelling approach at hand is to be attributed to the non-availability of information on imported delay. Imported delay, however, represents a significant fraction of total delays as during aircraft rotation, reactionary delay is produced due to the complexity and interconnectivity of air traffic schedules, especially at large hub airports. A delay

produced in the morning can thus be transported, spread and multiplied through schedules in the course of day. The second difficulty with regard to imported delay is the question on how to properly incorporate it into a punctuality modelling approach. Truly, delay produced outside an assumed weather correlation radius as described in Section 1.4 is likely to have nothing to do with conditions at Frankfurt Airport. However, also in that case there might be a causal relationship, as ATM in certain cases, e.g. when capacity at Frankfurt Airport is degraded due to adverse weather, may, through GDPs, take appropriate action and hold back aircraft with destination Frankfurt. In order to distinguish these cases from other cases, delay codes have to be considered. However, this makes analyses much more complicated and unreliable, as delay codes are, through inappropriate or incorrect use, a controversial means of information. Moreover, it is not foreseeable, what effect an incorporation of information on imported delay would have on modelling results. A quality rise is to be expected as other studies (see e.g. Hansen and Wei, 1999; Hansen and Bolic, 2001) pointed to origin airport congestion being a major source of delay variation. However, a significant effort has to be put in a proper model integration.

It has to be noted, as well, that the approach at hand is a purely local approach. All predictor variables created and used for punctuality modelling are local. Of course, and as already stated above, delays are not only locally produced. Besides ordinarily imported delay, delay can also be produced enroute, either through sector overload or simply by adverse weather. These factors are not directly incorporated into the present modelling. However, they are indirectly considered, as weather at point x is to a certain extent often correlated with weather at point y, with decreasing correlation the larger the distance between the two points is. When modelling punctuality on a daily level, using representative local predictors only is thus comparable with using information for a larger catchment area.

When a higher temporal model resolution is aimed for, the issue of deterministic delay comes into play. As discussed above, imported delays account for a major fraction of overall delays. This is all the more the case when punctuality is to be modelled on an hourly or even minute-by-minute level. The dominating stochastic nature of delay when analysed on a daily level is being gradually replaced by a deterministic nature on finer time scales. Then, delay rather has to be considered on a flight-by-flight level, where the current delay status of flights arriving or departing at the airport of interest within the next reference period is taken into consideration for punctuality assessment and projection.

With respect to the role of weather in the context of air traffic delays and punctuality, one remark has to be made. Generally, weather acts as a governing factor with respect to the production and development of delays. Though weather is sometimes not the primary source of a delay in the sense

of a causal relationship, there, nevertheless, often exists a correlation with other delay causing factors, which are not directly captured and discussed.

Last but not least, the mathematical approach in this work exhibits advantages and disadvantages and it is of course one among many possible approaches. The limitations with regard to the used input data and the mathematical constraints are well considered. Especially the problem of different punctuality domains with corresponding model results on different quality levels is noted and discussed. Yet, the quality of the achieved modelling results clearly indicates that there is a benefit from the chosen approach. It is as well thinkable to use e.g. cluster methods or neural networks for punctuality modelling. However, interpretation then becomes rather difficult. Many studies on similar research topics have, as well, revealed that alternative methods often offered at most comparable or even worse predictive skill than methods applied in the present study. Hansen and Bolic (2001), for example, emphasised that the content of the used delay model, i.e. the included predictors, is more important than the form of modelling. In the context of the hybrid approach chosen it has to be pointed out that interaction terms are intentionally not considered, either. These terms may improve modelling results, but at the expense of a far larger and more complex predictor database. Furthermore, interaction terms are particularly prone to multicollinearity. Rather, non-linear effects are tackled through non-linear variable transformations in combination with a comprehensible correction algorithm.

Altogether, there is still potential for model improvements, either through methodical enhancements or refinement and enlargement of the input database. It is well noted that the dataset at hand is just a restricted dataset of a limited length. A competing effect with regard to the assumption of stationarity, which is assumed in order to apply the statistical methods chosen in this study, is the existence of an underlying system variability as a consequence of an ever-changing airport infrastructure. Hence, no perfect punctuality modelling is ever to be expected as there always remains room for uncertainty. Finally, annual and interannual variability of determining factors will set a threshold for achievable model quality.

4.3 Summary and Outlook

This study is the first study known so far which, as a major result, comprises an analysis of the potential of reliable medium-term punctuality forecasting. If forecasts of weather and traffic are provided, the approach chosen allows for a prediction of total daily punctuality for several days in advance. It was shown that validating the Forecast Model on independent data produced good modelling results with an R^2 of almost 0.6. Further model im-

provements might be achieved through the incorporation of information on imported delays as described above. However, a significant effort is likely to be spent in a meaningful model modification, especially when focusing on forecasting punctuality, where input for predictor variables has to be known in advance. A further rise of model quality is also to be expected, when delay generating ground processes are reflected through inclusion of respective information. Last but not least, the introduction of en-route weather information or even weather data from destination airports and airports of origin for flights operating at Frankfurt Airport might as well lead to significant model quality improvements.

The next step in the development of an operational punctuality forecast system is the application of forecasted weather for validation. This has already been partly tried with rudimentary MOS forecast data used in a simplified Forecast Model. Results obtained through this preliminary approach were surprisingly good. The difficulty, in that regard, consists in an optimisation of predictions for critical weather related predictors such as e.g. low visibility or snowfall. Reliable and non-conservative predictions of these key predictors are essential for an acceptable model sensibility and skill.

In a further step it is thinkable to also calibrate the model on forecasted data. This requires the availability of archived weather forecasts, ideally for the past 3 years, as shown in Section 3.2.13.1. The effects of this approach are not foreseeable and the quality of results obtained through this approach will be strongly depending on the quality of NWP/MOS forecasts and their potential to actually reflect weather occurred.

Appendix A

Arrival Rate Matrix

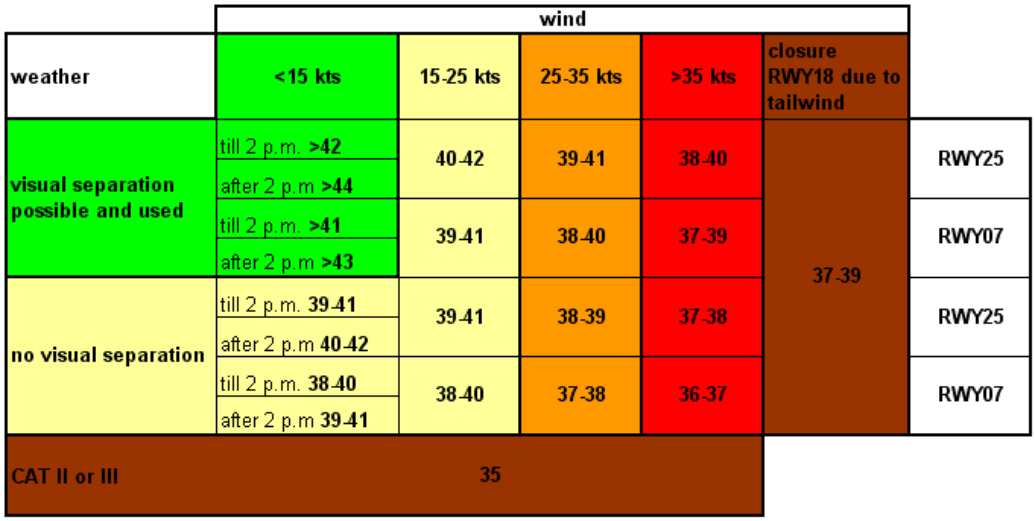


Figure A.1: Arrival Rate Matrix for Frankfurt Airport. Maximum number of arrivals per hour in low visibility conditions. Wind speeds given are head wind speeds in FL 30-FL 50 (3000-5000 ft). RWY07/RWY25 is the parallel runway system, RWY07: landing direction 70°, RWY25: landing direction 250° (kindly provided by FRAPORT).

Appendix B

Model Background Information

B.1 Regression Trees – 24h-Data

Table B.1: Special regression trees used for enhanced punctuality models (24h predictor resolution), ordered after column "rank coeff." (for a definition see Section 3.2.12.2). The second column gives the data selection rule for construction and application of the special tree. The fourth column gives the average *TOTP* on the "cases" (third column) event days.

tree	**weather criterium**	**cases**	**∅ *TOTP***	**rank coeff.**
1	h_mean<30 m	3	0.39	61.33
2	VV_mean<2000 m	5	0.45	33.12
3	SH_max>8 cm	14	0.45	27.71
4	Wt_b2_mean/Wt_N_mean>5 m/s	20	0.52	26.43
5	ww31b>0 & max_wind_gt35>0	13	0.53	21.69
6	fx24_max>25 m/s	4	0.58	21.00
7	Ws_a_mean/Ws_PB_mean>5 m/s	29	0.56	16.86
8	VV_min<300 m & max_wind_gt35>0	3	0.60	13.33
9	ff_mean>7 m/s & max_wind>25 m/s	27	0.59	12.07
10	VV_min<300 m & h_min<30 m	20	0.58	10.46
11	ww36b>0 & TT_mean<0°C	8	0.60	10.03
12	VV_min<300 m	21	0.59	9.78
13	ww34b>0 & TT_mean<0°C	98	0.62	9.73
14	r1_mean>45 min	19	0.65	9.10
15	ff_mean>7 m/s	60	0.63	8.00

B.2 Regression Trees – 6h-Data

Table B.2: Special regression trees used for enhanced punctuality models (6h predictor resolution), ordered after column "rank coeff." (for a definition see Section 3.2.12.2). The second column gives the data selection rule for construction and application of the special tree. The fourth column gives the average *TOTP* on the "cases" (third column) event days.

tree	**weather criterium**	**cases**	**∅ *TOTP***	**rank coeff.**
1	ww31b_4>0 & max_wind_gt35_4>0	1	0.20	80.00
2	max_Tangentialwind_a_head_gt35_2>0	2	0.32	68.00
3	h_mean<30 m	3	0.39	61.33
4	ww31b_2>0 & max_wind_gt35_2>0	2	0.41	59.50
5	SH_MAX_4>8 cm	11	0.42	37.02
6	VV_MEAN_3<2000 m	7	0.42	33.31
7	VV_mean<2000 m	5	0.45	33.12
8	h_MEAN_3<30 m	5	0.46	32.40
9	ww31b_1>0 & max_wind_gt35_1>0	2	0.42	29.25
10	E2_3>0	2	0.42	29.00
11	VV_MEAN_2<2000 m	22	0.47	28.76
12	Wt_b2_MEAN_1/Wt_N_MEAN_1>5 m/s	16	0.49	28.44
13	SH_MAX_3>8 cm	12	0.44	27.92
14	SH_max>8 cm	14	0.45	27.71
15	Wt_b2_mean/Wt_N_mean>5 m/s	20	0.52	26.43

B.3 Summary of Model Quality Criteria

Table B.3: Quality criteria for all model stages M 1-M 13. Figures for Model 5 are for the log-enhanced model using the 5-year calibration period. Enhancements in models labelled with '*' did not prove to be efficient and are not considered for later models. Results shown for Model 12 are for a high resolution model with regression tree correction and AR component. The Forecast Model, taking into account an AR component and predictors related to upper level wind, is labelled 'FM'.

		M 1	**M 2**	**M 3***	**M 4**	**M 5**	**M 6**	**M 7**	**M 8**	**M 9**	**M 10***	**M 11**	**M 12**	**M 13**	**FM**
MAE	diag	0.066	0.065	0.065	0.063	0.062	0.058	0.057	0.053	0.052	0.053	0.049	0.048	0.049	0.054
	prog	0.081	0.081	0.080	0.077	0.074	0.069	0.070	0.067	0.066	0.065	0.063	0.063	0.063	0.064
RMSE	diag	0.092	0.090	0.090	0.088	0.086	0.080	0.079	0.074	0.073	0.073	0.068	0.068	0.068	0.078
	prog	0.115	0.115	0.114	0.111	0.105	0.095	0.069	0.092	0.090	0.091	0.088	0.088	0.087	0.088
SE	diag	0.093	0.091	0.090	0.089	0.087	0.081	0.080	0.074	0.073	0.073	0.069	0.068	0.069	0.078
	prog	0.118	0.118	0.117	0.114	0.108	0.097	0.099	0.095	0.093	0.094	0.090	0.090	0.089	0.090
$r_{multiple}$	diag	0.604	0.628	0.631	0.648	0.666	0.072	0.728	0.768	0.777	0.778	0.807	0.810	0.807	0.761
	prog	0.610	0.629	0.624	0.653	0.667	0.719	0.706	0.735	0.748	0.744	0.766	0.766	0.768	0.766
R^2	diag	0.365	0.394	0.398	0.420	0.443	0.519	0.530	0.590	0.604	0.606	0.652	0.656	0.651	0.579
	prog	0.372	0.396	0.389	0.426	0.445	0.517	0.499	0.541	0.559	0.554	0.587	0.587	0.590	0.586
R^2_{adj}	diag	0.358	0.387	0.392	0.414	0.437	0.514	0.525	0.585	0.599	0.601	0.648	0.652	0.648	0.573
	prog	/	/	/	/	/	/	/	/	/	/	/	/	/	/

Appendix C

Forecast Model

C.1 Monthly Plots

In the following, monthly time series of $TOTP$, $TOTP_F$ and the model error are shown for the validation year 2006. The underlying model is the high resolution Forecast Model including predictors related to upper level wind as introduced in Section 3.3. The calibration period is 2003-2005.

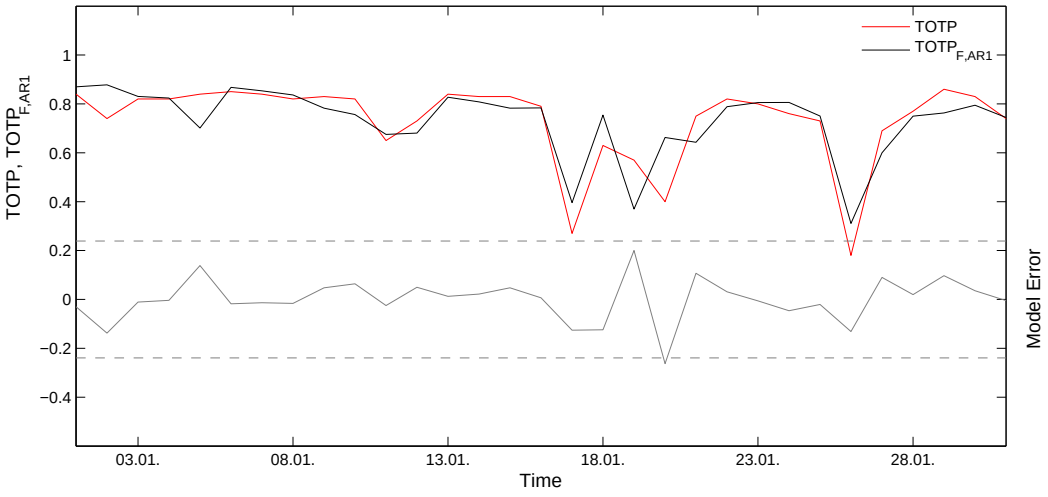


Figure C.1: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, January 2006. Dashed lines give twice the standard deviation of $TOTP$.

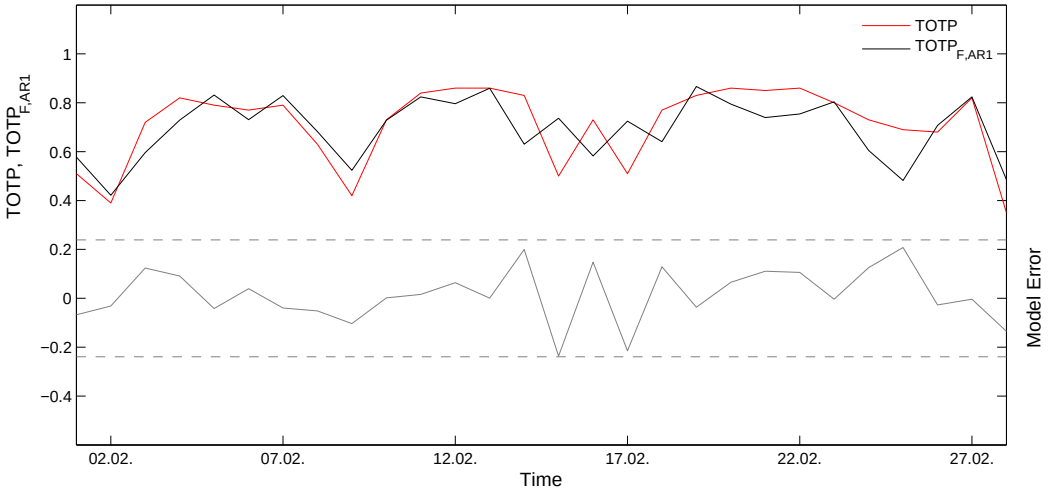


Figure C.2: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, February 2006. Dashed lines give twice the standard deviation of $TOTP$.

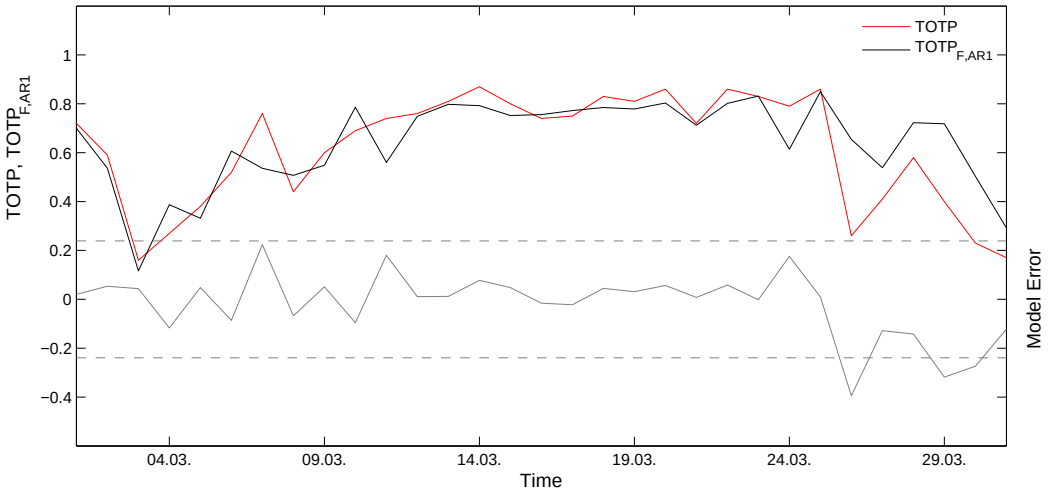


Figure C.3: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, March 2006. Dashed lines give twice the standard deviation of $TOTP$.

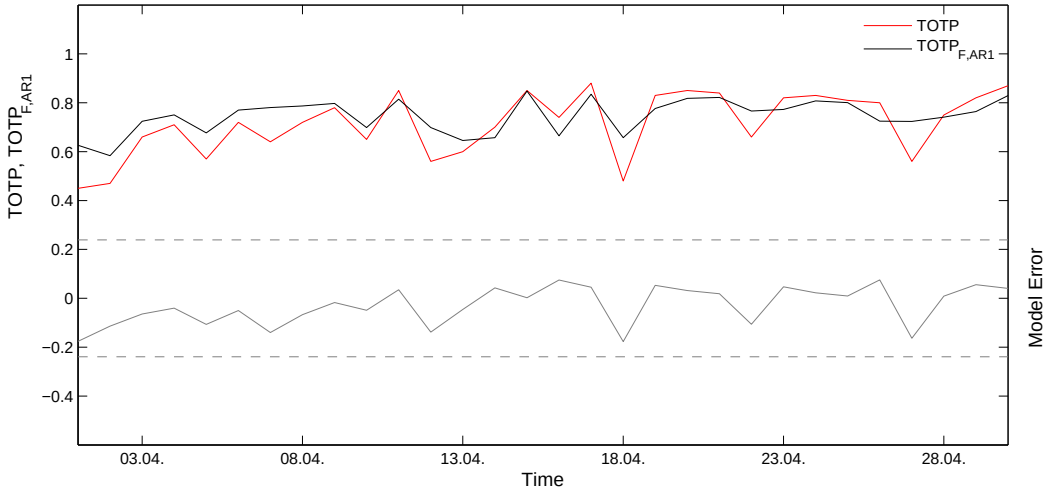


Figure C.4: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, April 2006. Dashed lines give twice the standard deviation of $TOTP$.

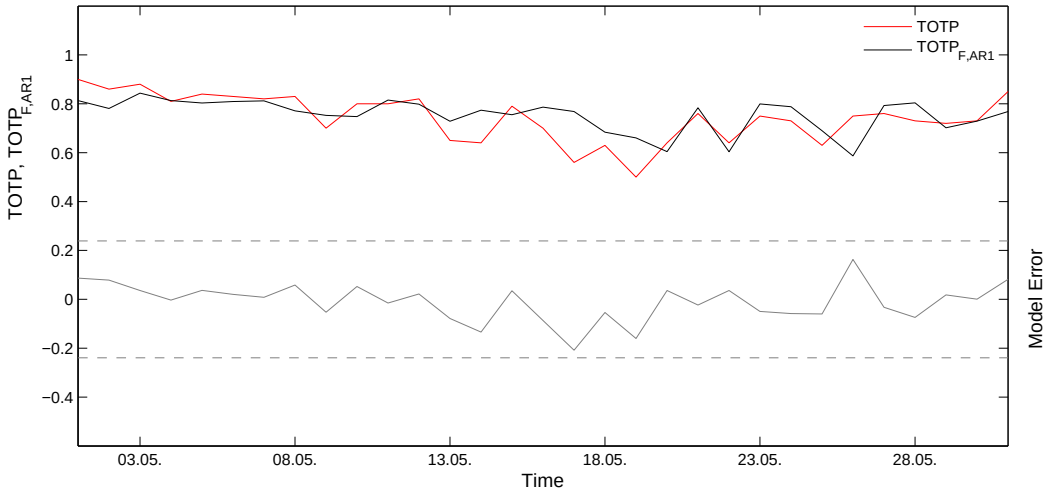


Figure C.5: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, May 2006. Dashed lines give twice the standard deviation of $TOTP$.

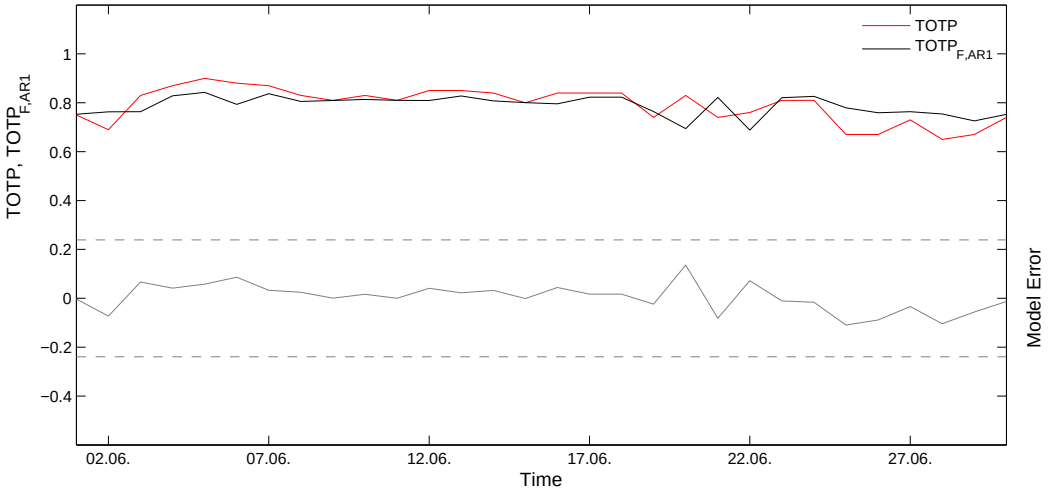


Figure C.6: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, June 2006. Dashed lines give twice the standard deviation of $TOTP$.

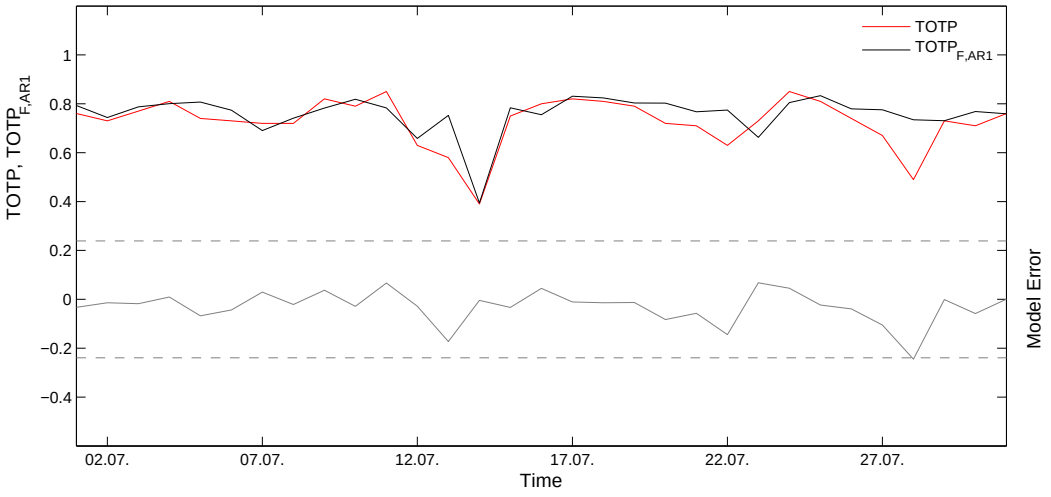


Figure C.7: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, July 2006. Dashed lines give twice the standard deviation of $TOTP$.

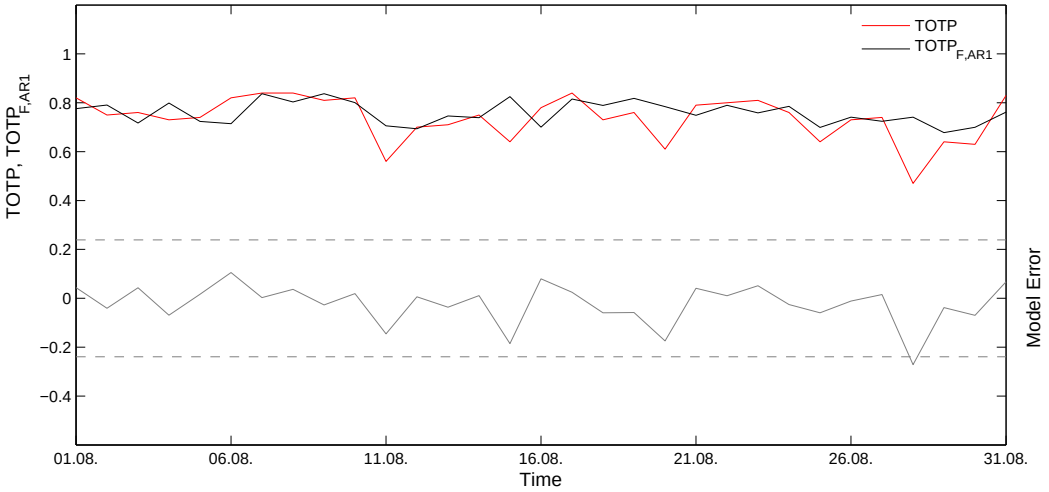


Figure C.8: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, August 2006. Dashed lines give twice the standard deviation of $TOTP$.

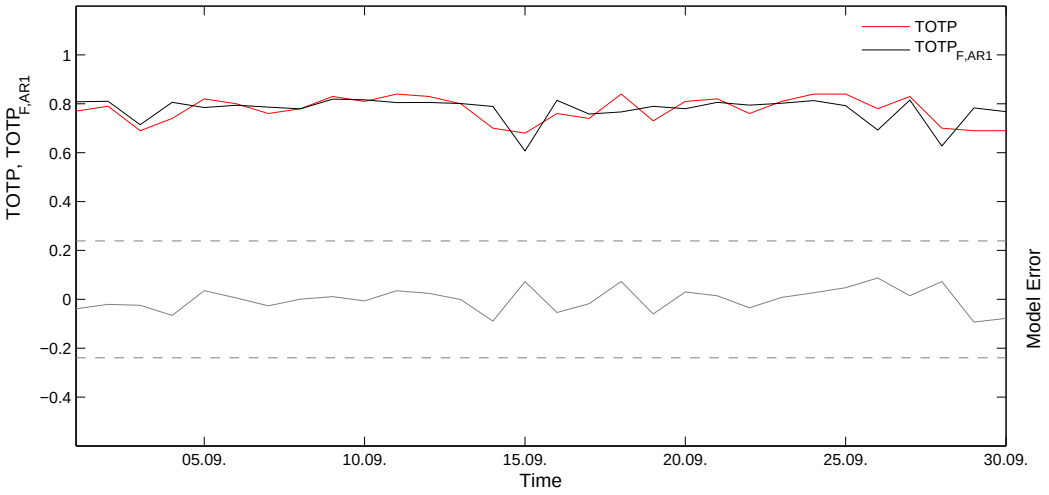


Figure C.9: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, September 2006. Dashed lines give twice the standard deviation of $TOTP$.

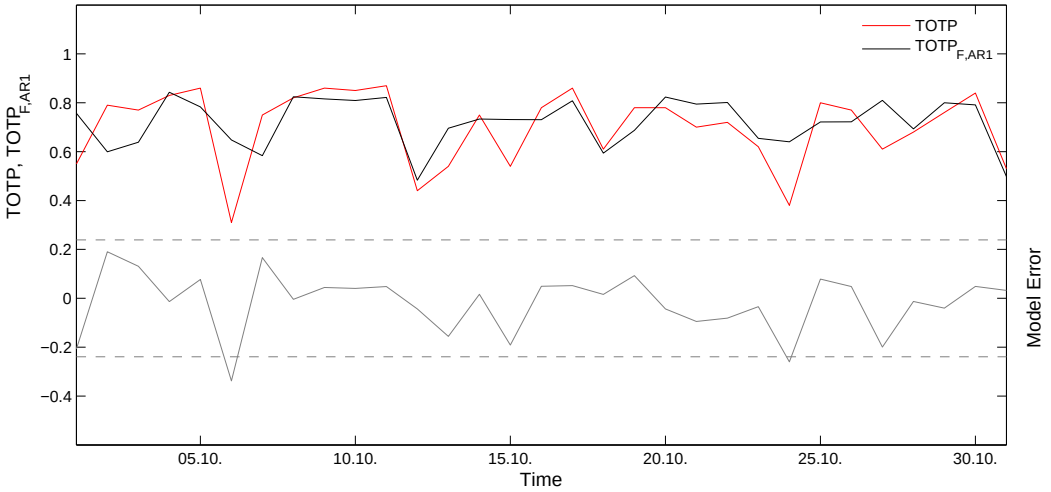


Figure C.10: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, October 2006. Dashed lines give twice the standard deviation of $TOTP$.

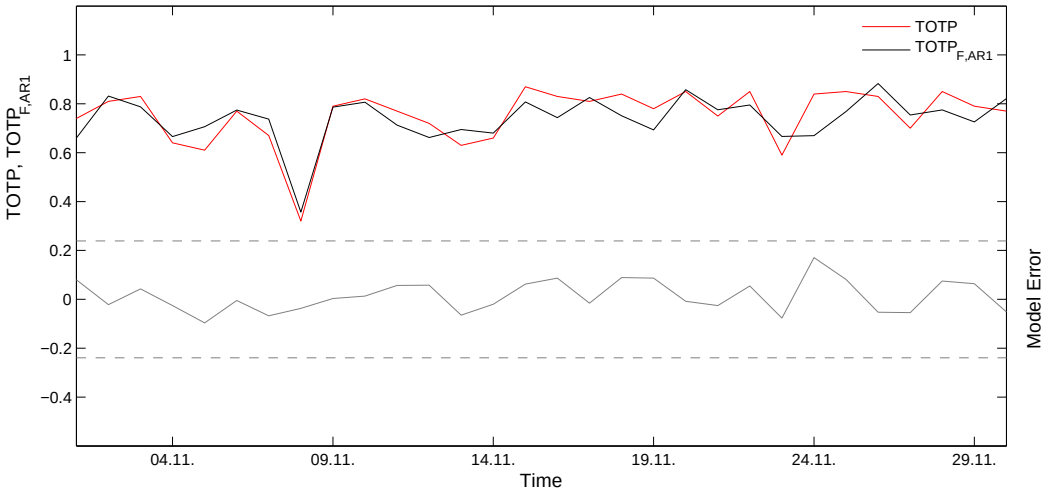


Figure C.11: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, November 2006. Dashed lines give twice the standard deviation of $TOTP$.

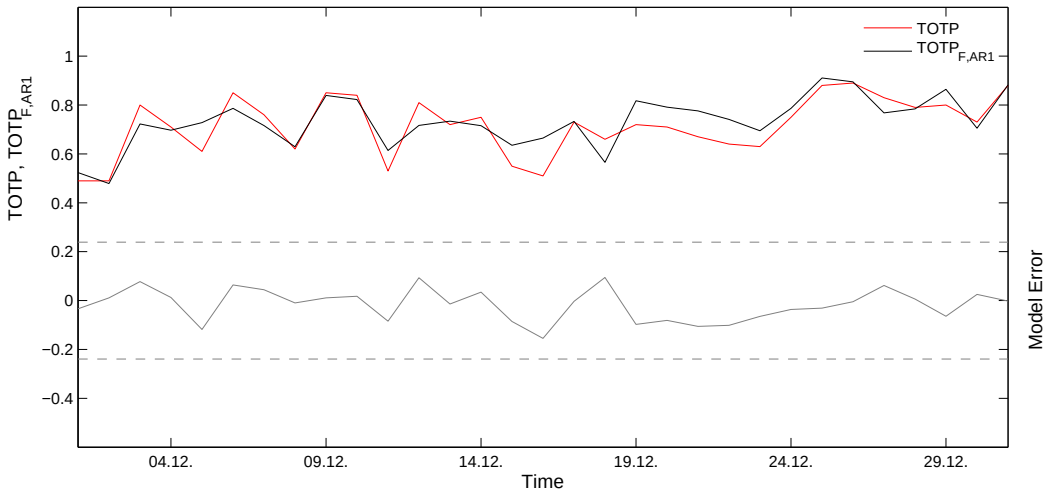


Figure C.12: Time series of $TOTP$, $TOTP_F$ and the model error for the Forecast Model, December 2006. Dashed lines give twice the standard deviation of $TOTP$.

Appendix D

SYNOP Format

D.1 ww-Encoding

00 clear skies
01 clouds dissolving
02 state of sky unchanged
03 clouds developing

Haze, smoke, dust or sand

04 visibility reduced by smoke
05 haze
06 widespread dust in suspension not raised by wind
07 dust or sand raised by wind
08 well developed dust or sand whirls
09 dust or sand storm within sight but not at station

Non-precipitation events

10 mist
11 patches of shallow fog
12 continuous shallow fog
13 lightning visible, no thunder heard
14 precipitation within sight but not hitting ground
15 distant precipitation but not falling at station
16 nearby precipitation but not falling at station
17 thunderstorm but no precipitation falling at station
18 squalls within sight but no precipitation falling at station
19 funnel clouds within sight

Precipitation within past hour but not at observation time

20 drizzle
21 rain
22 snow
23 rain and snow
24 freezing rain
25 rain showers
26 snow showers
27 hail showers
28 fog
29 thunderstorms

Duststorm, sandstorm, drifting or blowing snow

30 slight to moderate duststorm, decreasing in intensity
31 slight to moderate duststorm, no change
32 slight to moderate duststorm, increasing in intensity
33 severe duststorm, decreasing in intensity
34 severe duststorm, no change
35 severe duststorm, increasing in intensity
36 slight to moderate drifting snow, below eye level
37 heavy drifting snow, below eye level
38 slight to moderate drifting snow, above eye level
39 heavy drifting snow, above eye level

Fog or ice fog

40 Fog at a distance
41 patches of fog
42 fog, sky visible, thinning
43 fog, sky not visible, thinning
44 fog, sky visible, no change
45 fog, sky not visible, no change
46 fog, sky visible, becoming thicker
47 fog, sky not visible, becoming thicker
48 fog, depositing rime, sky visible
49 fog, depositing rime, sky not visible

Drizzle

50 intermittent light drizzle
51 continuous light drizzle
52 intermittent moderate drizzle
53 continuous moderate drizzle

54 intermittent heavy drizzle
55 continuous heavy drizzle
56 light freezing drizzle
57 moderate to heavy freezing drizzle
58 light drizzle and rain
59 moderate to heavy drizzle and rain

Rain

60 intermittent light rain
61 continuous light rain
62 intermittent moderate rain
63 continuous moderate rain
64 intermittent heavy rain
65 continuous heavy rain
66 light freezing rain
67 moderate to heavy freezing rain
68 light rain and snow
69 moderate to heavy rain and snow

Snow

70 intermittent light snow
71 continuous light snow
72 intermittent moderate snow
73 continuous moderate snow
74 intermittent heavy snow
75 continuous heavy snow
76 diamond dust
77 snow grains
78 snow crystals
79 ice pellets

Showers

80 light rain showers
81 moderate to heavy rain showers
82 violent rain showers
83 light rain and snow showers
84 moderate to heavy rain and snow showers
85 light snow showers
86 moderate to heavy snow showers
87 light snow/ice pellet showers
88 moderate to heavy snow/ice pellet showers
89 light hail showers

90 moderate to heavy hail showers

Thunderstorms

91 thunderstorm in past hour, currently only light rain
92 thunderstorm in past hour, currently only moderate to heavy rain
93 thunderstorm in past hour, currently only light snow or rain/snow mix
94 thunderstorm in past hour, currently only moderate to heavy snow or rain/snow mix
95 light to moderate thunderstorm
96 light to moderate thunderstorm with hail
97 heavy thunderstorm
98 heavy thunderstorm with duststorm
99 heavy thunderstorm with hail

Adopted from UNISYS (2009).

D.2 CL-Encoding

0 no low clouds
1 cumulus humulis or fractus (no vertical development)
2 cumulus mediocris or congestus (moderate vertical development)
3 cumulonimbus calvus (no outlines nor anvil)
4 stratocumulus cumulogenitus (formed by spreading of cumulus)
5 stratocumulus
6 stratus nebulosus (continuous sheet)
7 stratus or cumulus fractus (bad weather)
8 cumulus and stratocumulus (multilevel)
9 cumulonimbus with anvil
/ low clouds unobserved due to darkness or obscuration

Adopted from UNISYS (2009).

D.3 E-Encoding

0 ground dry (no cracks or appreciable amounts of dust/loose sand)
1 ground moist
2 ground wet (standing water in small or large pools on surface)
3 flooded
4 ground frozen
5 glaze on ground
6 loose dry dust or sand not covering ground completely
7 thin cover of loose dry dust or sand covering ground completely
8 mod/thick cover of loose dry dust/sand covering ground completely
9 extremely dry with cracks

Adopted from BADC (2009).

Appendix E

Standard IATA Delay Codes

Others

00-05	AIRLINE INTERNAL CODES
06 (OA)	NO GATE/STAND AVAILABILITY DUE TO OWN AIRLINE ACTIVITY
09 (SG)	SCHEDULED GROUND TIME LESS THAN DECLARED MINIMUM GROUND TIME

Passenger and Baggage

11 (PD)	LATE CHECK-IN, acceptance after deadline
12 (PL)	LATE CHECK-IN, congestions in check-in area
13 (PE)	CHECK-IN ERROR, passenger and baggage
14 (PO)	OVERSALES, booking error
15 (PH)	BOARDING, discrepancies and paging, missing checked-in passenger
16 (PS)	COMMERCIAL PUBLICITY/PASSENGER CONVENIENCE, VIP, press, ground meals and missing personal
17 (PC)	CATERING ORDER, late or incorrect order given to supplier
18 (PB)	BAGGAGE PROCESSING, sorting etc.

Cargo and Mail

21 (CD)	DOCUMENTATION, errors etc.
22 (CP)	LATE POSITIONING
23 (CC)	LATE ACCEPTANCE
24 (CI)	INADEQUATE PACKING
25 (CO)	OVERSALES, booking errors
26 (CU)	LATE PREPARATION IN WAREHOUSE
27 (CE)	DOCUMENTATION, PACKING etc. (mail only)
28 (CL)	LATE POSITIONING (mail only)
29 (CA)	LATE ACCEPTANCE (mail only)

Aircraft and Ramp Handling

31 (GD) AIRCRAFT DOCUMENTATION LATE/INACCURATE, weight and balance, general declaration, pax manifest, etc.
32 (GL) LOADING/UNLOADING, bulky, special load, cabin load, lack of loading staff
33 (GE) LOADING EQUIPMENT, lack of or breakdown, e.g. container pallet loader lack of staff
34 (GS) SERVICING EQUIPMENT, lack of or breakdown, lack of staff, e.g. steps
35 (GC) AIRCRAFT CLEANING
36 (GF) FUELLING/DEFUELLING, fuel supplier
37 (GB) CATERING, late delivery or loading
38 (GU) ULD, lack of or serviceability
39 (GT) TECHNICAL EQUIPMENT, lack of or breakdown, lack of staff, e.g. pushback

Technical and Aircraft Equipment

41 (TD) AIRCRAFT DEFECTS
42 (TM) SCHEDULED MAINTENANCE, late release
43 (TN) NON-SCHEDULED MAINTENANCE, special checks and/or additional works beyond normal maintenance schedule
44 (TS) SPARES AND MAINTENANCE EQUIPMENT, lack of or breakdown
45 (TA) AOG SPARES, to be carried to another station
46 (TC) AIRCRAFT CHANGE, for technical reason
47 (TL) STAND-BY AIRCRAFT, lack of planned stand-by aircraft for technical reasons
48 (TV) SCHEDULED CABIN CONFIGURATION/VERSION ADJUSTMENTS

Damage to Aircraft & EDP/Automated Equipment Failure

51 (DF) DAMAGE DURING FLIGHT OPERATIONS, bird or lightning strike, turbulence, heavy or overweight landing, collision during taxiing
52 (DG) DAMAGE DURING GROUND OPERATIONS, collisions (other than during taxiing), loading/off-loading damage, contamination, towing, extreme weather conditions
55 (ED) DEPARTURE CONTROL
56 (EC) CARGO PREPARATION/DOCUMENTATION
57 (EF) FLIGHT PLANS

Flight Operations and Crewing

61 (FP)	FLIGHT PLAN, late completion or change of, flight documentation
62 (FF)	OPERATIONAL REQUIREMENTS, fuel, load alteration
63 (FT)	LATE CREW BOARDING OR DEPARTURE PROCEDURES, other than connection and standby (flight deck or entire crew)
64 (FS)	FLIGHT DECK CREW SHORTAGE, sickness, awaiting standby, flight time limitations, crew meals, valid visa, health documents, etc.
65 (FR)	FLIGHT DECK CREW SPECIAL REQUEST, not within operational requirements
66 (FL)	LATE CABIN CREW BOARDING OR DEPARTURE PROCEDURES, other than connection and standby
67 (FC)	CABIN CREW SHORTAGE, sickness, awaiting standby, flight time limitations, crew meals, valid visa, health documents, etc.
68 (FA)	CABIN CREW ERROR OR SPECIAL REQUEST, not within operational requirements
69 (FB)	CAPTAIN REQUEST FOR SECURITY CHECK, extraordinary

Weather

71 (WO)	DEPARTURE STATION
72 (WT)	DESTINATION STATION
73 (WR)	EN ROUTE OR ALTERNATE
75 (WI)	DE-ICING OF AIRCRAFT, removal of ice and/or snow, frost prevention excluding unserviceability of equipment
76 (WS)	REMOVAL OF SNOW, ICE, WATER AND SAND FROM AIRPORT
77 (WG)	GROUND HANDLING IMPAIRED BY ADVERSE WEATHER CONDITIONS

Air Traffic Flow Management Restrictions

81 (AT)	ATFM DUE TO ATC EN-ROUTE DEMAND/CAPACITY, standard demand/capacity problems
82 (AX)	ATFM DUE TO ATC STAFF/EQUIPMENT EN-ROUTE, reduced capacity caused by industrial action or staff shortage, equipment failure, military exercise or extraordinary demand due to capacity reduction in neighbouring area
83 (AE)	ATFM DUE TO RESTRICTION AT DESTINATION AIRPORT, airport and/or runway closed due to obstruction, industrial action, staff shortage, political unrest, noise abatement, night curfew, special flights
84 (AW)	ATFM DUE TO WEATHER AT DESTINATION

Airport and Governmental Authorities

85 (AS) MANDATORY SECURITY
86 (AG) IMMIGRATION, CUSTOMS, HEALTH
87 (AF) AIRPORT FACILITIES, parking stands, ramp congestion, lightning, buildings, gate limitations, etc.
88 (AD) RESTRICTIONS AT AIRPORT OF DESTINATION, airport and/or runway closed due to obstruction, industrial action, staff shortage, political unrest, noise abatement, night curfew, special flights
89 (AM) RESTRICTIONS AT AIRPORT OF DEPARTURE WITH OR WITHOUT ATFM RESTRICTIONS, including Air Traffic Services, start-up and pushback, airport and/or runway closed due to obstruction or weather, industrial action, staff shortage, political unrest, noise abatement, night curfew, special flights

Reactionary

91 (RL) LOAD CONNECTION, awaiting load from another flight
92 (RT) THROUGH CHECK-IN ERROR, passenger and baggage
93 (RA) AIRCRAFT ROTATION, late arrival of aircraft from another flight or previous sector
94 (RS) CABIN CREW ROTATION, awaiting cabin crew from another flight
95 (RC) CREW ROTATION, awaiting crew from another flight (flight deck or entire crew)
96 (RO) OPERATIONS CONTROL, re-routing, diversion, consolidation, aircraft change for reasons other than technical

Miscellaneous

97 (MI) INDUSTRIAL ACTION WITH OWN AIRLINE
98 (MO) INDUSTRIAL ACTION OUTSIDE OWN AIRLINE, excluding ATS
99 (MX) OTHER REASON, not matching any code above

Adopted from Guest (2007).

Bibliography

Abdelghany, K. F., S. S. Shah, S. Raina and A. F. Abdelghany, 2004: A model for projecting flight delays during irregular operation conditions, *Journal of Air Transport Management*, **10**, pp. 385–394.

Backhaus, K., B. Erichson, W. Plinke and R. Weiber, 2003: Multivariate Analysemethoden, Springer Verlag, 10th edition, 818 pp.

BADC, 2009: WMO Meteorological codes, published in the internet, URL `http://badc.nerc.ac.uk/data/surface/code.html`, cited 2009 Mar 3.

Bamberg, G. and F. Baur, 1998: Statistik, Oldenbourgs Lehr- und Handbücher der Wirtschafts- und Sozialwissenschaften, Oldenbourg Verlag, 10th edition, 343 pp.

Bartels, H., F. M. Albrecht and J. Guttenberger, 1990: Starkniederschlagshöhen für die Bundesrepublik Deutschland, Teil 1: Niederschläge längerer Dauerstufen ($D \geq 24\,\mathrm{h}$), Zeitraum 1951-1980, Deutscher Wetterdienst.

Beatty, R., R. Hsu, L. Berry and J. Rome, 1998: Preliminary evaluation of flight delay propagation through an airline schedule, in 2nd USA/Europe Air Traffic Management R&D Seminar, Orlando.

Benker, H., 2000: Mathematik mit Matlab – Eine Einführung für Ingenieure und Naturwissenschaftler, Springer, 550 pp.

Boswell, S. B. and J. E. Evans, 1997: Analysis of downstream impacts of air traffic delay, *Project Report ATC-257*, MIT Lincoln Laboratory.

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone, 1984: Classification and Regression Trees, The Wadsworth Statistics/Probability Series, Wadsworth International Group, 358 pp.

Brozat, R., 2007: Abschlusspräsentation LUFO III, Verbundvorhaben K-ATM, Teilvorhaben KOPIM Airport, published in the internet, URL `http://www.dfs.de/dfs/internet_2008/module/forschung_und_entwicklung/deutsch/forschung_und_entwicklung/service/k_atm/`

abschlusspraesentation_ii/katm_capman_ataman_v1_00_fraport_20071129.pdf, cited 2009 May 14.

Callaham, M. B., J. S. DeArmon, A. M. Cooper, J. H. Goodfriend, D. Moch-Mooney and G. H. Solomos, 2001: Assessing NAS Performance: Normalizing for the Effects of Weather, in 4th USA/Europe Air Traffic Management R&D Symposium, Santa Fe, p. 11.

Chin, D. K., J. Goldberg and T. Tang, 1997: Airport Surface Delays and Causes – A Preliminary Analysis, *NASA Contractor Report 201721*, NASA, National Aeronautics and Space Administration, Langley Research Center, Hampton, Virginia 23681-0001.

DFS, 2008: Luftverkehr in Deutschland – Mobilitätsbericht 2008, *Jahresbericht*, Deutsche Flugsicherung.

Dillingham, G. L., 2005: National Airspace System – Initiatives to reduce flight delays and enhance capacity are ongoing but challenges remain, *GAO-05-755-T*, United States Government Accountability Office.

Drüe, C., W. Frey, A. Hoff and T. Hauf, 2008: Aircraft type-specific errors in AMDAR weather reports from commercial aircraft, *Quaterly Journal of the Royal Meteorological Society*, **134**, pp. 229–239.

ECMWF, 2007: Hit rate and false alarm rate, published in the internet, URL http://ecmwf.int/products/forecasts/guide/Hit_rate_and_False_alarm_rate.html, cited 2009 Mar 5.

EUROCONTROL, 2003: Flight Delay Propagation – Synthesis of the study, *Technical report*, EUROCONTROL Experimental Centre, EEC Note No 18/03.

EUROCONTROL, 2004: Evaluating the true cost to airlines of one minute of airborne or ground delay, *Final Report, commissioned by the Performance Review Commission*, Performance Review Commission, 96 Rue de la Fusée, B-1130 Brussels, Belgium.

EUROCONTROL, 2005: Report on Punctuality Drivers at Major European Airports, *Final Report, commissioned by the Performance Review Commission*, Performance Review Unit, 96 Rue de la Fusée, B-1130 Brussels, Belgium.

EUROCONTROL, 2007: Delays to Air Transport in Europe, Digest - Annual 2007, published in the internet, URL https://extranet.eurocontrol.int/http://prisme-web.hq.corp.eurocontrol.int/ecoda/coda/public/standard_page/codarep/2007/2007DIGEST.pdf, cited 2009 Feb 23.

EUROCONTROL, 2008: Global ATM interoperability, Focus on: SESAR & NextGen, *Skyway, The Eurocontrol Magazine*, **12** (49).

EUROCONTROL, 2009a: CODA – Public Reports, published in the internet, URL `https://extranet.eurocontrol.int/http://prisme-web.hq.corp.eurocontrol.int/ecoda/coda/public/standard_page/public_application.html`, cited 2009 May 14.

EUROCONTROL, 2009b: Performance Review Report – An Assessment of Air Traffic Management in Europe during the Calendar Year 2008, *Final Report, commissioned by the Performance Review Commission, PRR 2008*, Performance Review Commission, 96 Rue de la Fusée, B-1130 Brussels, Belgium.

FHKD, 2008: Flughafenkoordinator Deutschland, Koordinationseckwerte, published in the internet, URL `https://sws.fhkd.org/EWPS/pAirportParameters.input`, cited 2009 Feb 18.

Frank, M., M. Mederer, B. Stolz and T. Hanschke, 2005: Depeaking – Economic Optimization of Air Traffic Systems, *Aerospace, Science and Technology*, **9** (8), pp. 738–744, URL `http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VK2-4H391YR-1&_user=2148698&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000056361&_version=1&_urlVersion=0&_userid=2148698&md5=2ff5e829bf836635613fc469820fab21`, cited 2009 Mar 12.

FRAPORT, 2006: Kapazität managen, Wachstum sichern, *Geschäftsbericht 2006*, URL `http://www.fraport.de/cms/investor_relations/dokbin/235/235213.geschaeftsbericht_2006.pdf`, cited 2009 Feb 23.

FRAPORT, 2007: Optimum Capacity Utilization, CAPMAN – Capacity Manager, published in the internet, URL `http://www.fraport.com/cms/company/rubrik/12/12774.innovation_projects.htm`, cited 2009 May 14.

Frey, W., 2006: Investigation of type-specific errors in AMDAR weather reports of commercial aircraft, Diplomarbeit, Gottfried Wilhelm Leibniz Universität Hannover, Institut für Meteorologie und Klimatologie.

Garson, D., 2009: Multiple Regression, published in the internet, URL `http://faculty.chass.ncsu.edu/garson/PA765/regress.htm`, cited 2009 Apr 22.

Guest, T., 2007: A Matter of Time: Air Traffic Delay in Europe, *Trends in Air Traffic*, EUROCONTROL.

Hansen, M. and T. Bolic, 2001: Delay and Flight Time Normalization Procedures for Major Airports: LAX Case Study, *NEXTOR Research Report UCB-ITS-RR-2001-5*, Institute of Transportation Studies, University of California at Berkeley.

Hansen, M. M. and W. Wei, 1999: Multivariate Analysis of the Impacts of NAS Investments: A Case Study of a Major Capacity Expansion at Dallas-Forth Worth Airport, *Research Report UCB-ITS-RR-98-11*, Institute of Transportation Studies, University of California at Berkeley.

Heinemann, H.-J., 2008: Eine Winterchronik: Die Kälte der Winter in Deutschland von 1960/61 bis 2007/08, *Berichte des Deutschen Wetterdienstes*, (232), p. 62.

Hipel, K. W. and A. I. McLeod, 1994: Time Series Modelling of Water Resources and Environmental Systems, volume 45 of *Developments in Water Science*, Elsevier, 1st edition, 1013 pp.

Hoffmann, B., J. Krozel and R. Jakobavits, 2004: Potential benefits of fix-based ground delay programs to address weather constraints, in AIAA Guidance, Navigation, and Control Conf., Providence, Ri, American Institute of Aeronautics and Astronautics, p. 13.

ICAO, 1993: Airport Services Manual, Part 2, Pavement Surface Conditions, *Technical report*, ICAO, 2nd edition.

ITA, 2000: Costs of air transport delay in europe, *Final report*, Institut du Transport Aérien, sponsored by the EUROCONTROL Performance Review Unit.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl and T.-C. Lee, 1988: Introduction to the theory and practice of econometrics, John Wiley & Sons, 2nd edition, 1024 pp.

KLM, 2009: Capacity Forecast Schiphol (CPS), published in the internet, URL `http://92.254.52.180/cdm/cps/default.asp`, cited 2009 May 14.

Liang, D., W. Marnane and S. Bradford, 2000: Comparison of US and European Airports and Airspace to Support Concept Validation, in 3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Italy.

Markovic, D., T. Hauf, P. Röhner and U. Spehr, 2008: A statistical study of the weather impact on punctuality at Frankfurt Airport, *Meteorological Applications*, **15**, pp. 293–303.

Middleton, G. V., 200: Data Analysis in the Earth Sciences Using Matlab, Prentice Hall, 260 pp.

Müller-Westermeier, G., A. Kreis and E. Dittmann, 1999: Klimaatlas Bundesrepublik Deutschland, Teil 1: Temperatur, Niederschlagshöhe, Sonnenscheindauer, Deutscher Wetterdienst.

Müller-Westermeier, G., A. Kreis and E. Dittmann, 2001: Klimaatlas Bundesrepublik Deutschland, Teil 2: Verdunstung, Maximumtemperatur, Minimumtemperatur, Kontinentalitat, Deutscher Wetterdienst.

Müller-Westermeier, G., A. Kreis, E. Dittmann and W. R. Klemens Barfus, Gerhard Czeplak, 2003: Klimaatlas Bundesrepublik Deutschland, Teil 3: Bewölkung, Globalstrahlung, Anzahl der Tage klimatologischer Ereignisse, Phänologie, Deutscher Wetterdienst.

Müller-Westermeier, G., A. Walter and E. Dittmann, 2005: Klimaatlas Bundesrepublik Deutschland: Klimatische Wasserbilanz, Teil 4: Tägliche Temperaturschwankung, Windgeschwindigkeit, Dampfdruck, Schneedecke, Deutscher Wetterdienst.

Moninger, W. R., R. D. Mamrosh and P. M. Pauley, 2003: Automated Meteorological Reports from Commercial Aircraft, *Bulletin of the American Meteorological Society*, **84**, pp. 203–216.

Niehues, A., S. Belin, T. Hansson, R. Hauser, M. Mostajo and J. Richter, 2001: Punctuality: How airlines can improve on-time performance, *Airline and Aerospace*.

Peer, C., 2003: An Evaluation of Weather Parameters causing Aircraft Departure and Arrival Delays at Vienna International Airport, Diplomarbeit, Leopold Franzens Universität Innsbruck.

Peer, C., H. Pümpel and T. Hauf, 2008: A study on weather related aircraft departure and arrival delays at Vienna International Airport, *Berichte des Instituts für Meteorologie und Klimatologie der Leibniz Universität Hannover*, Institut für Meteorologie und Klimatologie.

Rehm, F., 2003: Data Mining Methoden zur Bestimmung des Einflusses von Wetterfaktoren auf Anflugverspätungen an Flughäfen, Master Thesis, Otto-von-Guericke-Universität Magdeburg.

Rehm, F. and F. Klawonn, 2005: Learning Methods for Air Traffic Management, in Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science, 8th European Conference, ECSQARU, Barcelona, Spain, July 6-8, 2005, Springer Verlag, pp. 992–1001.

Röhner, P., 2004: Meteorologische Analyse des Wettereinflusses auf den Flugbetrieb im Winter, Diplomarbeit, Universität Hannover, Institut für Meteorologie und Klimatologie.

Röhner, P. and T. Hauf, 2008: Some meteorological insights on winter weather operations at two major German airports, *Berichte des Instituts für Meteorologie und Klimatologie der Leibniz Universität Hannover*, Institut für Meteorologie und Klimatologie.

Robinson, P. J., 1989: The influence of weather on flight operations at the atlanta hartsfield international airport, *Weather and Forecasting*, **4** (4), pp. 461–468.

Sasse, M., 2000: Pilotstudie zum Wettereinfluss auf den Flugverkehr, Diplomarbeit, Universität Hannover, Institut für Meteorologie und Klimatologie.

Sasse, M. and T. Hauf, 2003: A study of thunderstorm-induced delays at Frankfurt Airport, Germany, *Meteorological Applications*, **10**, pp. 21–30.

Solf, M., 2005: Einflüsse auf die Variabilität von Anflugzeiten bei Verkehrsflugzeugen, Diplomarbeit, Technische Universität Braunschweig.

Spehr, U., 2003: Analyse des Wettereinflusses auf die Pünktlichkeit im Flugverkehr, Diplomarbeit, Universität Hannover, Institut für Meteorologie und Klimatologie.

Sprinkle, C. H. and K. J. Macleod, 1991: Impact of weather on aviation: A global view, in Preprints of 4th International Conference on Aviation Weather Systems, Paris, France, pp. 191–196, preprints of 4th International Conference on Aviation Weather Systems.

Theusner, M. and P. Röhner, 2008: Comprehensive Global Assessment on Weather Risks and Impact on Delays, *Deliverable D 2.1-2*, FLYSAFE.

Thrasher, T. and W. Weiss, 2001: A proposed method for measuring air traffic delay, in ATCA Conference.

UNISYS, 2009: SYNOP data format (FM-12), Surface Synoptic Observations, published in the internet, URL `http://weather.unisys.com/`, cited 2009 Mar 3.

Verein Berliner Wetterkarte, 2006a: Berliner Wetterkarte, 6.10.2006, **55** (194), URL `http://www.met.fu-berlin.de/wetter/wetterkarte`.

Verein Berliner Wetterkarte, 2006b: Berliner Wetterkarte, 7.10.2006, **55** (195), URL `http://www.met.fu-berlin.de/wetter/wetterkarte`.

Weisberg, S., 1985: Applied Linear Regression, John Wiley & Sons, 2nd edition, 324 pp.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences, volume 59 of *International Geophysics Series*, Academic Press, 24-28 Oval Road, London NW1 7DX, 1st edition, 467 pp.

WMO, 2004: WMO AMDAR Programme, published in the internet, URL `http://amdar.wmo.int/index.html`, cited 2009 Jun 5.

Wu, C.-L., 2005: Inherent delays and operational reliability of airline schedules, *Journal of Air Transport Management*, **11**, pp. 273–282.

Wu, C.-L. and R. E. Caves, 2000: Aircraft operational costs and turnaround effciency at airports, *Journal of Air Transport Management*, **6**, pp. 201–208.

Wu, C.-L. and R. E. Caves, 2002: Modelling of aircraft rotation in a multiple airport environment, *Transportation Research 38E: Logistics and Transportation Review*, **38** (3-4), pp. 265–277.

Wu, C.-L. and R. E. Caves, 2003a: Flight schedule punctuality control and management: A stochastic approach, *Transportation Planning and Technology*, **26** (4), pp. 313–330.

Wu, C.-L. and R. E. Caves, 2003b: The punctuality performance of aircraft rotations in a network of airports, *Transportation Planning and Technology*, **26** (5), pp. 417–436.

Wu, C.-L. and R. E. Caves, 2004: Modelling and optimization of aircraft turnaround time at an airport, *Transportation Planning and Technology*, **27** (1), pp. 47–66.

Acknowledgements

Zunächst gilt besonderer Dank Herrn Prof. Dr. Thomas Hauf als meinem Doktorvater für die Betreuung und Begutachtung dieser Arbeit. Über die Jahre des Studiums und der Promotion hinweg hat er immer mit gutem Rat zur Seite gestanden und neue Impulse und Denkanstöße gegeben.

Ein großer Dank geht an Herrn Beckmann und Herrn Streicher vom Flughafen Frankfurt. Ohne ihren Support mit Daten und Hintergrundinformationen wäre diese Arbeit so nicht möglich gewesen. In vielen Gesprächen vor Ort konnten neue Ideen entstehen und dank ihrer großartigen Unterstützung auch schnell umgesetzt werden.

Weiterhin gilt mein Dank meinen Kommilitionen und Kollegen, in deren Kreis Probleme aller Art diskutiert werden konnten, was nicht unerheblich zum Gelingen dieser Arbeit beigetragen hat. Hervorzuheben ist besonders Dr. Danijela Markovic, mit der der Grundstein zur vorliegenden Pünktlichkeitsmodellierung gelegt wurde und mit der ich in einer fruchtvollen Zusammenarbeit viele neue Ideen verwirklichen konnte. Des Weiteren möchte ich Ulrike Spehr nennen, die mit ihren Vorarbeiten auf dem Gebiet der Verspätungsanalysen den Weg für weitere Entwicklungen aufgezeigt hat. Nicht vergessen möchte ich die Kollegen in meinem Büro, Dr. Michael Theusner, Dr. Tanja Weusthoff, Katharina Koppe und Stefan Himmelsbach, mit denen ich nicht nur fachliche Probleme diskutieren konnte, sondern die auch sonst für eine angenehme Arbeitsathmosphäre gesorgt haben.

Bedanken möchte ich mich auch für die Unterstützung durch die Arbeitsgruppe „Verkehrsmeteorologie“. Besonders hervorheben möchte ich Dr. Clemens Drüe und Michael Köckritz, die mir in besonderem Maße bei der Prozessierung von AMDAR-Daten weitergeholfen haben.

Nicht zuletzt geht mein großer Dank an meine Familie, die mich während meines gesamten Studiums und meiner Promotion unterstützt hat. Ohne ihren Rückhalt wäre diese Promotion sicherlich nicht möglich gewesen.

Curriculum Vitae

PERSONAL DATA

Name: Peer Röhner

Nationality: German

Birth Date: 13/05/77

Birth Place: Eisenach

SCHOOL EDUCATION

09/83 – 08/89	Willy Settner Oberschule Eisenach (elementary school)
09/89 – 07/90	Orientierungsstufe Mellendorf (two-year middle school)
08/90 – 06/97	Gymnasium Mellendorf (grammar school)

MILITARY SERVICE

07/97 – 04/98	Panzergrenadierlehrbataillon 92, Munster

UNIVERSITY EDUCATION

10/98 – 07/02	studies in meteorology, Universität Hannover, Institut für Meteorologie und Klimatologie
08/02 – 06/03	special studies in *"Arctic Geophysics"* at UNIS, Longyearbyen/Svalbard (Norway), participation in field campaigns in the Fram Straight and fjords at the westcoast of Svalbard
07/03 – 06/04	diploma thesis in meteorology, subject: Meteorological analysis of the weather impact on winter service at German airports
09/04	diploma, title: Diplom-Meteorologe

PUBLICATIONS

RÖHNER, P. and T. HAUF, 2008: Some meteorological insights on winter weather operations at two major German airports, *Berichte des Instituts für Meteorologie und Klimatologie der Leibniz Universität Hannover*.

MARKOVIC, D., T. HAUF, P. RÖHNER and U. SPEHR, 2008: A statistical study of the weather impact on punctuality at Frankfurt Airport, *Meteorological Applications*, **15**, pp. 293-303.