

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY TO
THE PHD SCHOOL OF THE FACULTY OF SCIENCE,
UNIVERSITY OF COPENHAGEN

On The Usage of Crowdsourced Data in Numerical Weather Prediction

Kasper Stener HINTZ

Supervisors:

Professor Eigil KAAS (NBI)

Dr Henrik VEDEL (DMI)



 **novation Fund Denmark**



JULY 10, 2019

A Thesis Submitted for The Degree of Doctor of Philosophy

Title: On The Usage of Crowdsourced Data in Numerical Weather Prediction

Publication: July 2019

Support: This work has received funding from Innovation Fund Denmark, through the industrial researcher programme, grant 5189-00042B

Author: MSc, Kasper Stener Hintz

E-mail: kah@dmi.dk

Phone: +45 28 73 71 29

UCPH Id: kct206

University Supervisor: Professor Eigil Kaas

External Supervisor (DMI): Dr Henrik Vedel

External Supervisor (Vaavud): Dr Juan Muñoz-Gomez[†]

External Co-Supervisor (Vaavud): CEO Thomas Helms[†]

Industrial PhD partners: Danish Meteorological Institute
University of Copenhagen, Faculty of Science
Vaavud ApS[†]

Cite as: Hintz, K. S., On The Usage of Crowdsourced Data in Numerical Weather Prediction. Danish Meteorological Institute, 2019.

Keywords: Crowdsourcing, Numerical Weather Prediction, Technology, Nowcasting, Sensors, Assimilation

[†]A comment from the Author:

Unfortunately, 14 months into the project, the industrial PhD partner Vaavud ApS was unable to further fund the project and employ the PhD student. The Danish Meteorological Institute took over from September 1st, 2017. I greatly appreciate the time I had in Vaavud ApS and for the supervising from my former supervisors Dr Juan Muñoz-Gomez and CEO Thomas Helms.

Preface

This PhD study has two main branches due to unforeseen circumstances of which the reader should be aware. This project started in June 2016 as an industrial PhD study. Here, the candidate was employed at Vaavud ApS, a small start-up company which manufactured anemometers for smartphones and made related software, for example, to provide tailored weather forecasts.

Halfway into the PhD project, Vaavud ApS went bankrupt and had to terminate all contracts with employees, including the PhD candidate of this study. Uncertain months followed, with different solutions being worked on.

The Danish Meteorological Institute (DMI) took over the project in September 2017. The primary objective of the study and the overall subject changed from utilising crowdsourced wind measurements in meteorology to examining pressure observations from smartphones in Numerical Weather Prediction (NWP). Therefore the work of this study has been carried out in two phases. The shift of both focus and working environment has given the candidate good skills to adapt and utilise research in a new project quickly.

The first half of the study was done at Vaavud ApS in a consortium of partners which consisted of Vaavud ApS, University of Copenhagen, ConWX ApS and DMI. Here, the company supervisor was Dr Juan Muñoz-Gomez. The second part of the study was carried out at DMI in a consortium consisting of DMI and University of Copenhagen. Here, the company supervisor was Dr Henrik Vedel. Throughout the project, Professor Eigil Kaas has been the University supervisor.

At Vaavud, I was given the opportunity to work with people from a wide variety of fields, such as software programmers, data scientists and a large environment of innovative entrepreneurs. At DMI the work was done as part of the HIRLAM-C (High-Resolution Limited Area Model) programme, in which the primary purpose is to carry out research and development for short-range weather forecasting activities. For this purpose, an examination of smartphone pressure observations was an excellent opportunity for both the HIRLAM community and myself. As a part of this community, I have been blessed with the possibility to work closely with experts within NWP from all over Europe.

Four articles have been produced and are included in the appendices. Two of the four articles are published, one is currently in revision, and one is in preparation.

Abstract

Observations are a vital part of Numerical Weather Prediction (NWP) to give accurate forecasts of future weather. As the spatial resolution of NWP models increases, so does the need for more observations for use in data assimilation. Installing new professional meteorological observing equipment is costly and expensive to maintain. Crowdsourced data is a new potential data source for NWP that have emerged in recent years. Crowdsourced data is an overall term covering reports from users and data from equipment owned and operated by the public. Such data are often less accurate than traditional meteorological observations, but there are much more available. This PhD project studies the potential use of crowdsourced data in NWP. The main focus is on handheld wind measurements from smartphones, pressure observations from smartphones and on Personal Weather Stations (PWS). It is shown that handheld wind measurements can, in some cases, be more representative than traditional wind observations and a method of estimating the surface roughness length from a handheld measurement is presented. Software for collecting Smartphone Pressure Observations (SPOs) has been developed and have in one year successfully collected more than 60 million observations from Denmark. Also, a scheme for quality control of the SPOs has been developed. SPOs and pressure observations from PWS have been assimilated into the DMI HARMONIE NWP model using 3-Dimensional Variational (3D-Var) data assimilation. It is shown that SPOs can contribute positively to NWP, but it is also concluded that the full potential has not yet been reached. Other assimilation techniques than 3D-Var are likely to be more suitable for assimilating crowdsourced data. Observations from PWS are likely to be useful for NWP and observation based nowcasting, but a quality monitoring system needs to be developed before this can be applied. The developments and findings in this PhD project constitute fundamental contributions to advance the work on utilising crowdsourced data as it is shown that such data is of a quality that can contribute to improving weather forecasts from NWP models.

Abstract in Danish

Observationer er en vigtig del af numeriske vejrmodeller for at kunne give gode prognoser af det kommende vejr for dets brugere. Efterhånden som den rumlige opløsning af vejrmodellerne øges, kræves der også flere observationer til brug i dataassimilering. Installation og vedligehold af professionelle meteorologiske observationsstationer har en høj økonomisk omkostning. I de senere år er en ny potentiel datakilde til brug i numeriske vejrmodeller opstået, kaldet crowdsourced data. Crowdsourced data er en overordnet term der dækker over data ejet af privatpersoner. Disse data er typisk mindre nøjagtige end traditionelle meteorologiske observationer, til gengæld er der ofte mere data til rådighed. Dette PhD projekt undersøger potentialet af crowdsourced data til brug numeriske vejrmodeller. Det primære fokus er på håndholdte vindmålinger fra smartphones, trykobservationer fra smartphones samt observationer fra personlige vejrstationer. Det bliver vist at håndholdte vindmålinger kan i nogen tilfælde være mere repræsentative end traditionelle vindobservationer. Ydermere, præsenteres en metode til at udlede en overflade ruhedparameter fra håndholdte vindmålinger. Der er udviklet software til at indsamle trykobservationer fra smartphones and der er i løbet af et år, indsamlet mere end 60 millioner observationer. En algoritme til kvalitetskontrol af trykobservationer fra smartphones er også udviklet. Trykobservationer fra smartphones og personlige vejrstationer er blevet assimileret i DMI's HARMONIE vejrmodel ved hjælp af 3-Dimensionel Variationel (3D-Var) dataassimilering. Det bliver vist at trykmålinger fra smartphones kan bidrage positivt til nøjagtigheden af numeriske vejrmodeller, men det konkluderes også at det fulde potentiale af denne type data endnu ikke er opnået. Andre dataassimileringsmetoder end 3D-Var forventes at være bedre egnet til at assimilere crowdsourced data. Observationer fra personlige vejrstationer må formodes at være til gavn for numeriske vejrmodeller samt observationsbaseret straksvarsling, men der kræves et system der kan monitorere sådanne stationer før disse kan implementeres. Dette PhD projekt viser at crowdsourced data er af en kvalitet der kan udnyttes i sammenhæng med numeriske vejrmodeller og giver derved et fundamentalt bidrag til udviklingen af brugen af crowdsourced data.

Acknowledgements

Many people have had a part in this study throughout the years. Only with the support and guidance from my two superior supervisors, **Henrik Vedel** and **Eigil Kaas**, and my family, **Mette Hintz** and **Karl Hintz** were it possible to obtain such excellent results and learning outcomes.

I thank **Henrik Vedel** and **Eigil Kaas** for their commitment to the project and their lessons in academic life.

Also, **Thomas Helms** took a chance of starting an industrial PhD project at a start-up company because he believed in the possibilities of the project. I want to thank Thomas Helms, for, without him, this study would have never started nor completed. Also, my former supervisor at Vaavud **Juan Muñoz-Gomez** made an excellent effort for the project to succeed.

Research and technical staff at the Danish Meteorological Institute has been helpful in all matters. This is especially true for **Niels Woetmann Nielsen**, **Bent Hansen Sass**, **Xiaohua Yang**, **Mats Dahlbom**, **Ulrik Smith Korsholm**, **Claus Petersen**, **Ole Leth Krarup**, **Bjarne Amstrup**, **Keld Qvistgaard**, **Thomas Bøvith**, **Rashpal Gill** and **Kristian Pagh Nielsen**.

Søren Foged Skov made the **SFS Development** weather app available for this project. For this, I and DMI as a whole are grateful. Also, agriculturist **Poul Graae** provided open access to his farmland for meteorological experiments.

Former Vaavud staff has also provided both software solutions and technical assistance when needed. **Diego Galindo** has been an excellent mentor for me in software development for mobile platforms in both frontend and backend applications. **Nicolai Mølgaard Christensen** put a large effort into helping automating wind measurements in field experiments.

Constructive criticism and academic discussions with **Søren Borg Nielsen**, **Christian Holme**, **Mads Poulsen**, **Roger Randriamampianina**, **Conor McNicholas** and **Andreas Nyholm** has also been of great value.

Acronyms

Abbreviation	Description
3D-Var	3-Dimensional Variational data assimilation
4D-Var	4-Dimensional Variational data assimilation
ABL	Atmospheric Boundary Layer
AI	Artificial Intelligence
AMV	Atmospheric Motion Vector
API	Application Programming Interface
App	Application (Typical for Smartphones)
ARW	Advanced Research WRF
ASCII	American Standard Code for Information Interchange
AWS	Automated Weather Stations
BUFR	Binary Universal Form for the Representation of meteorological data
CART	Classification And Regression Trees
CCTV	Closed-Circuit Television
DA	Data Assimilation
DMI	Danish Meteorological Institute
DMPAR	Distributed-Memory Parallelism
DTM	Danish Terrain Model
DTU	Danish Technical University
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecast
EnKF	Ensemble Kalman Filter
EUMETNET	European Meteorological Services Network

Abbreviation	Description
FDE	Finite Difference Equations
FGAT	First Guess at Appropriate Time
FMI	Finnish Meteorological Institute
FSS	Fractional Skill Score
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
GFS	Global Forecasting System
GNSS	Global Navigation Satellite System
GTS	Global Telecommunication System
HARMONIE	Hirlam Aladin Regional Meso-scale Operational NWP in Europe
HIRLAM	High Resolution Limited Area Model
HPC	High-Performance Computing
HWM	Handheld Wind Measurement
IBL	Internal Boundary Layer
IFS	Integrated Forecasting System
IoT	Internet of Things
KF	Kalman Filter
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LAM	Limited Area Model
LETKF	Local Ensemble Transform Kalman Filter
ML	Machine Learning
NBI	Niels Bohr Institute
NWP	Numerical Weather Prediction
ODB	Observation Database
OI	Optimal Interpolation
OS	Operating System
PBL	Planetary Boundary Layer
PDE	Partial Differential Equations
PDF	Probability Density Function
PWS	Personal Weather Stations
RMI	Royal Meteorological Institute of Belgium

Abbreviation	Description
SDK	Software Development Kit
SLP	Sea Level Pressure
SMAPS	SMArtphone Pressure System
SPO	Smartphone Pressure Observation
SWO	Smartphone Wind Observation
SYNOP	surface synoptic observations
TCP	Transmission Control Protocol
TKE	Turbulent Kinetic Energy
UI	User Interface
UX	User Experience
WMO	World Meteorological Organization
WRF	Weather Research and Forecasting
ZAMG	Zentralanstalt für Meteorologie und Geodynamik

Nomenclature

Sign	Description	Unit
$()'$	Perturbed state of a variable	–
C	Constant	–
H	Forward model operator	–
J	Diabatic heating	J/s
K	Diffusion coefficient	m^2/s
L	Monin-Obukhov length	m
P	Precipitation intensity	$\text{kg}/\text{m}^2\text{s}$
R_d	Gas constant for dry air	J/kgK
T	Temperature	K
Z	Radar reflectivity factor	mm^6/m^3
H	Linearised forward model operator	–
P_b	Background error covariance	–
R	Observation error covariance	–
x	Model state of prognostic variables	–
y	Observation vector	–
ω	Vertical velocity	Pa/s
$\overline{()}$	Mean state of a variable	–
ρ	Density	kg/m^3
σ	Standard deviation	–
τ_0	Shear stress	N/m^2
θ	Potential temperature	K
c_p	Specific heat capacity at constant pressure	J/K/kg
k	von Kármán constant	–

Sign	Description	Unit
l	Mixing length	m
p	Pressure	hPa
r	Radial distance	m
t	Time	s
u^*	Friction velocity	m/s
u	Longitudinal wind speed	m/s
v	Latitudinal wind speed	m/s
w	Vertical velocity	m/s
z_0	Roughness length	m
z	Vertical coordinate	m

Glossary

API (Application Programming Interface) is an interface that allows software programs to interact. One example of an API could be an interface which returns a weather forecast given a position as latitude and longitude.

Client-Side Client-side refers to processes that occur locally on a users device such as a smartphone.

Crowdsourcing Crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.

IoT (Internet of Things) refers to devices communicating together via the internet and/or ethernet. A classic example is an alarm clock setting off in the morning and at the same time sending a message to the coffee machine to start brewing coffee.

NWP (Numerical Weather Prediction) is the process of creating weather forecasts using discretised models of physics and mathematics to be solved using computers.

SDK (Software Development Kit) is a set of tools, components and examples which can be used to develop software applications targeting specific platforms. An SDK can hold several APIs.

Server-Side Server-side is opposite to client-side processes that occur on remote servers that can be shared by many clients.

Contents

Preface	ii
Abstract	iii
Abstract in Danish	iv
Acknowledgements	v
Acronyms	vi
Nomenclature	ix
Glossary	xi
1 Introduction	1
1.1 Motivation	1
1.2 Primary objectives	2
1.3 Overview	2
1.4 Included papers	4
2 Data collection	6
2.1 Wind observations from smartphones	7
2.2 Pressure observations from smartphones	9
2.2.1 The Danish terrain model	13
2.3 Personal weather stations	15
2.4 Data management	16
2.4.1 European GDPR act	17

3	Introduction to numerical weather prediction	18
3.1	Vertical coordinates	20
3.2	Numerical methods	21
3.3	Verification	22
3.3.1	Fractional skill score for precipitation verification	23
4	The Vaavud system for numerical weather prediction	26
4.1	Script environment on Odin	27
4.2	WRF implementation on Frigg	27
5	Fundamentals of atmospheric turbulence	29
5.1	The Boussinesq approximation	30
5.2	Reynolds decomposition	31
5.3	The closure problem	33
6	The atmospheric boundary layer	34
6.1	Similarity theory	35
6.1.1	The fundamentals of similarity theory	35
6.1.2	The Monin-Obukhov similarity	36
6.2	The wind profile in the atmospheric boundary layer	36
6.2.1	The neutral surface layer	37
6.2.1.1	The classical approach	37
6.2.1.2	Mixing length	37
6.2.2	The wind profile in non-neutral surface layers	38
6.3	Extrapolating wind measurements	39
6.4	The roughness length	40
7	Traditional atmospheric observations and measurements	42
7.1	WMO guidelines for wind observations	42
7.2	Experimental setup of wind observations	43
7.2.1	Extendable transportable mast	43
7.2.2	Ultrasonic anemometer	43
7.3	WMO guidelines for atmospheric pressure observations	46
7.4	Validation of observations	46
8	Estimation of wind speed and roughness length using smartphones: Method and quality assessment - Publication A	48

9	Data assimilation	51
9.1	Least squares method	52
9.2	The framework of data assimilation	53
9.3	Optimal interpolation	55
9.3.1	The OI weight	57
9.3.2	The analysis covariance	58
9.4	3 dimensional variational data assimilation	59
9.4.1	The two temperature problem	62
9.4.2	Data assimilation time window and FGAT	62
9.5	4 dimensional variational data assimilation	63
9.6	Nudging	65
9.7	Kalman filtering	66
9.7.1	The Kalman algorithm	67
10	Data assimilation in the DMI HARMONIE NWP system	70
10.1	The HARMONIE NWP model	70
10.2	Observation pre-processing in HARMONIE	72
10.2.1	Local pre-processing	72
10.2.2	Oulan	73
10.2.3	Bator	74
10.2.4	Lamflag	75
10.2.5	Screening	75
10.3	Structure functions in practice	76
11	Use of big data and crowdsourced data in NWP	79
12	Machine learning	82
12.1	Fundamentals of machine learning	82
12.1.1	Classifiers	83
12.1.2	Regressors	84
12.2	Preprocessing of data	84
12.2.1	Choosing input variables	84
12.2.1.1	Multicollinearity	85
12.2.1.2	Curse of dimensionality	85
12.3	Machine learning models for predicting errors	86
12.3.1	k-Nearest-Neighbour	86
12.3.2	CART	86
12.3.2.1	Gradient boosting	88

13 Correcting smartphone pressure observations	89
13.1 Complementary results and discussion	90
14 Collecting and processing of Barometric Data from Smartphones for Potential use in NWP Data Assimilation - Publication B	97
15 Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018 - Publi- cation C	101
16 Evaluating pressure observations from private weather stations and smartphones - Publication D	104
17 Summary and conclusions	107
17.1 Summary	107
17.2 Conclusions	108
18 Future work and challenges	110
References	112
List of Figures	126
List of Tables	130
List of Blocks	131
List of Publications	132
Publications	133
Publication A	134
Publication B	173
Publication C	188
Publication D	197
Appendices	216
A Additional figures	217

B	Derivation of the Monin-Obukhov length	220
C	Bias, standard deviation and mean squared error	221
D	Namelist of the Frigg-WRF setup	223
E	Mathematical equivalence of OI and 3D-Var	227

1.1 Motivation

A primary motivation of this study is to improve nowcasting forecasts of hazardous weather, such as heavy fog, excessive rainfall and thunderstorms. Nowcasting methods can be based on Numerical Weather Prediction (NWP) models, or they can be purely based on observations. As the resolution of NWP models increase, whereas the traditional observing networks are not densifying at the same rate, so does the need for more observations to determine the initial conditions satisfactorily and validate results. The skill of a weather forecast depends heavily on the estimate of the initial state of the atmosphere. Therefore, to improve nowcasting forecasts, additional observations are needed.

A potential new source of observations is crowdsourced data. Crowdsourcing was originally defined by Howe (2006) as "*the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call*". An example that complies with Howe's definition is data from Personal Weather Stations (PWS). Today, many private homes have PWS installed (Muller et al., 2015), which can report observations for use in NWP (Netatmo SAS, 2018). However, in the atmospheric sciences, the term *crowdsourced data* is often used more broadly (Krennert et al., 2018). Here, data from vehicles and smartphones are also classified as crowdsourced data, even though such data are not used in operational NWP models yet. Definitions of terms within this research topic are still somewhat unclear but have been improved with the work of Muller et al. (2015) and Krennert et al. (2018).

It is not clarified how to deal with crowdsourced observations in order to include

them in an NWP model. If appropriately utilised, crowdsourced data can also potentially aid operational meteorologists in decision making. This PhD thesis provides a detailed study of crowdsourced data in NWP. Three different types of crowdsourced data types have been studied in detail, namely handheld wind observation from smartphones, pressure observations from smartphones and pressure observations from PWS.

1.2 Primary objectives

Prior to the initiation of the project, three primary objectives were defined. Because the use of crowdsourced data is a new broad topic within NWP, the objectives were deliberately formulated without too many specific details to be able to take on a broad perspective.

- Make fundamental research on the use of crowdsourced data in NWP.
- Develop a validation method which makes it possible to include crowdsourced data in short-term weather forecasting.
- Investigate the options for using crowdsourced data dynamically in the modelling processes of the atmosphere.

Additionally, three success criteria were defined:

- Crowdsourced data are included in the process of NWP
- The skill of weather forecasting is thereby improved
- The methods must be applied in commercial projects

1.3 Overview

The reader will quickly realise that there appear to be two branches of this PhD thesis. The project started at the company Vaavud ApS, which manufactured low-cost anemometers for smartphones and developed mobile software tailored for different use cases. Halfway into the project, Vaavud ApS went bankrupt, and all contracts with employees had to be terminated. The Danish Meteorological Institute (DMI) took over the project and employed the PhD candidate as a part of the HIRLAM-C project. The PhD project then shifted focus from utilising crowdsourced wind measurements to investigate the potential use of smartphone pressure observations in NWP. Due to the

mentioned reasons, this thesis consists of two branches, one branch for research done at Vaavud ApS in which the focus is mainly on handheld wind measurements from smartphones and one branch for research done at DMI where the main focus is on Smartphone Pressure Observation (SPO)s and PWS.

The thesis is structured as follows: Chapter 1 is a general introduction to the PhD project. Chapter 2 presents the work done on data collection in this study for wind and pressure observations from smartphones and data from PWS. Hereafter follows chapter 3 with a general introduction to NWP. Then, the focus is solely upon wind observations from smartphones performed during the first half of the project, starting with chapter 4 which gives an overview of the Vaavud NWP system which was set up in the very beginning of this PhD project.

Chapter 5 and 6 introduces the theory for turbulence and the atmospheric boundary layer. Atmospheric observations and measurements with a focus on traditional methods and the experimental setups made in this study are described in chapter 7. Here, the experimental setup for the validation of handheld wind measurements from smartphones is presented. The guidelines from World Meteorological Organization (WMO) for both wind and pressure are also given. Chapter 8 introduces the first article: "*Estimation of wind speed and roughness length using smartphones: Method and quality assessment*", with a summary and key findings.

Then, the main focus shift towards pressure observations from smartphones and the work carried out mainly at DMI, starting with an introduction to data assimilation in chapter 9, with a focus on 3-Dimensional Variational data assimilation (3D-Var) which is used in the Hirlam Aladin Regional Meso-scale Operational NWP in Europe (HARMONIE) model. Chapter 10 presents the practical implementation of the HARMONIE data assimilation environment and documents the changes made to the system in order to assimilate pressure observations from smartphones. Chapter 11 gives a general overview of potential data sources, other than the ones that are in focus in this study. Also, a discussion on the differences between the terms "big data" and "crowdsourced data" is given to avoid confusion between the two. Chapter 12 and chapter 13 gives an introduction to machine learning algorithms which has been used to predict the error of smartphone pressure observations and the initial results of these experiments, respectively.

Chapter 14 introduces the second article "*Collecting and processing of Barometric Data from Smartphones for Potential use in NWP Data Assimilation*". Chapter

15 introduces the third article, "*Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018*". Additionally, the fourth article is introduced in chapter 16. Finally, chapter 17 and 18 gives a summary, presents conclusions and discusses future challenges.

1.4 Included papers

Paper A:

Hintz, K. S., Vedel, H., Kaas, E., Nielsen, N. W., (2019a)

Estimation of wind speed and roughness length using smartphones: Method and quality assessment, (in-Revision) Journal of Atmospheric and Oceanic Technology.

Collection and utilising smartphone wind observations to estimate the roughness length of the surface and comparison to traditional wind observations.

Paper B:

Hintz, K. S., Vedel, H., Kaas, E., (2019b)

Collecting and processing of Barometric Data from Smartphones for Potential use in NWP Data Assimilation (published), Meteorological Applications, doi:10.1002/met.1805.

Collecting barometric data from smartphones and examining the quality of these. A screening method was developed, and preliminary data assimilation experiments are presented.

Paper C:

K. S. Hintz, K. O'Boyle, S. L. Dance, S. Al Ali, I. Ansper, D. Blaauboer, M. Clark, A.Cress, M. Dahoui, R. Darcy, J. Hyrkkänen, L. Isaksen, E. Kaas, U. S. Korsholm, M.Lavanant, G. Le Bloa, E. Mallet, C. McNicholas, J. Onvlee-Hooimeijer, B. Sass, V. Siirand, H. Vedel, J. A. Waller, X. Yang., (2019c)

Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018 (published) Atmospheric Science Letters, doi:10.1002/asl.921.

Summary of a workshop planned and hosted by the PhD candidate, bringing together experts in crowdsourced data in NWP from Europe and the United States of

America.

Paper D:

Hintz, K. S., Vedel, H., Kaas, E., (2019d)

Evaluating pressure observations from private weather stations and smartphones (In-Prep. for Meteorological Applications)

Collecting and examining pressure observations from personal weather stations for use in numerical weather prediction. Also, both pressure observations from smartphones and personal weather stations were assimilated into the HARMONIE model, for a period of nearly two months.

Chapter 2

Data collection

Data from smartphones can be separated into sensor data and user data. The former is data coming from sensors either built into or attached to the smartphone, and the latter is cases where the user interacts with the software to send some sort of data. One example of user data is to collect user reports of the current weather such as is done by for example Météo France, Finnish Meteorological Institute (FMI), Royal Meteorological Institute of Belgium (RMI) and Zentralanstalt für Meteorologie und Geodynamik (ZAMG) in Austria (FMI, 2018; Hintz et al., 2019c; Krennert et al., 2018).

In the atmospheric sciences, sensor data from smartphones is mostly focused on pressure observations as many modern smartphones can measure the pressure directly (see section 2.2). Temperature observations have also been extracted from the battery in smartphones by using a heat transfer model (Droste et al., 2017; Overeem et al., 2013), to quantify the urban heat island effect.

Even with the vast amount of potential sources of data coming available in recent years and the future (Evans, 2011), work based on robust data collection methods is still the foundation for the research going on within the topics of crowdsourced data, also within NWP. Hence, data collection has been a fundamental and essential part of this PhD project. Throughout the project, this single PhD study has collected data from more than 2 % of the entire Danish population. If the software which has been developed during this PhD project is upscaled, that number will only increase.

This chapter is dedicated to describe the data collection processes of the data used in this project, namely wind observations from smartphones (section 2.1), pressure observations from smartphones (section 2.2) and observations from PWS (section 2.3). Finally, a brief overview of data management is given in section 2.4, which also holds

general remarks on the General Data Protection Regulation (GDPR) act from the European Union (European Union, 2018).

2.1 Wind observations from smartphones

The motivation for collecting wind observations from smartphones is to provide better-tailored weather forecasts for the users of the anemometer, by using the observation in a post-processing routine of a NWP forecast. Also, as the number of observations increases a micro-climatological wind atlas for a variety of locations could be made.

All wind observations from smartphones used in this study come from the Danish private company Vaavud (Vaavud, 2018). Vaavud manufactured mobile anemometers for smartphones and developed tailored software for web and mobile platforms such as iOS and Android. Vaavud had users globally, but the primary markets were Denmark, United Kingdom and the Netherlands, from which most observations naturally come. The primary Vaavud products are shown in figure 2.1. In this project, wind measurements were obtained from the Vaavud Sleipnir anemometer (see figure 2.1).



Figure 2.1: The Vaavud anemometers, Mjolnir (left) and Sleipnir (middle and right). The Sleipnir anemometer also measures the wind direction due to its asymmetric rotor design.

The Sleipnir anemometer is attached to the smartphone through the jack stick. An infrared beam is sent from the outer surface of the anemometer to the centre, where a receiver is located. On the anemometer a uniformly spaced teeth block is attached, which interrupts the infrared beam as it rotates. When rotating, this creates a pattern of signal, no-signal, which is transferred back to the phone via the jack stick. The time between the signals translates to the rotation speed and from that the wind speed is calculated. The asymmetric rotor allows one to calculate the wind direction also. Due to the asymmetry the rotor accelerates and decelerates as the wind blows into the cup, from which the wind direction is computed.¹

When the anemometer is plugged into the smartphone, the user can via an app start a wind measurement. Each measurement consists of a 30 second period with a sampling rate of approximately 4 Hz. The observation, which is the average of the measurements are then uploaded to a database at Vaavud via the Vaavud Software Development Kit (SDK), together with the wind direction, the variance of the observation, position and a unique smartphone identifier. The Vaavud SDK was connected to the Vaavud app suite, which consisted of three separate apps:

- **Vaavud:** Official general purpose app
- **Vaavud Kitesurfing:** Dedicated app for kitesurfing purposes
- **Vaavud Sailing:** Dedicated app for sailing purposes

In all three apps, weather forecasts were also provided by Vaavuds own NWP system which was set up as a part of this PhD study at the beginning of the project (see chapter 4).

Wind measurements from smartphones are challenging to utilise. This is partly due to representativeness errors (see figure A1) but also due to a lower quantity compared to other crowdsourced data sources because users have to make an action with the existing methods. It is generally believed that the amount of data is less for applications where the user needs to perform an action (Hintz et al., 2019c). In this case, users have to attach an anemometer to the smartphone and manually start a measurement.

¹The technical description of Sleipnir originates from personal communication with the inventor of the device and former CTO of Vaavud, Andreas Okholm.

2.2 Pressure observations from smartphones

The focus on SPOs in this PhD project has been on collection methods, processing and usage of pressure observations in NWP and meteorology in general. The reasons for this focus is 1) barometers are often built into smartphones making data easily accessible and 2) pressure is a variable that is being assimilated in modern NWP assimilation systems already. Hence, the motivation for collecting pressure observations is to examine how a very dense dataset of a traditional parameter can be utilised in modern NWP.

Barometers started to be built into smartphones around 2011 to increase the speed of location fixes (Morris, 2011). Barometers are used to provide a first estimate of the altitude of the smartphone to retrieve a location fix faster, from the built-in Global Navigation Satellite System (GNSS). Also, it requires less power and is a more reliable way of tracking altitude changes of the device compared to using the GNSS. The standard deviation of the altitude assignment of a smartphone can be as high as 30 meters (Bauer, 2013). However, with the introduction of dual-frequency GNSS chipsets in smartphones the precision can get down to decameters (Robustelli et al., 2019).

The main Operating System (OS) (in Europe) for smartphones are ‘Android’ (Google) and ‘iOS’ (Apple Inc.) (StatCounter Global Stats, 2018). The native programming language for Android and iOS are Java and Swift/Objective-C, respectively. Apps developed from native code are referred to as native apps. Both Android and iOS provide SDKs to access sensor data through an Application Programming Interface (API), including barometric pressure. To collect pressure observations, it was necessary to develop software to the OS of interest.

Frameworks exist that makes it possible to build native apps from one codebase by cross-compiling. One such example is the React Native framework (Facebook Inc., 2017), an open-source framework for building native apps using javascript. The framework is promising, but are still in beta, and so it was decided to build the software using a native language. To begin with, a DMI-internal test-app was built for both Android and iOS, to examine sensor behaviour, sampling rates and the quality of the sensor readings. This also provided the PhD candidate with a subjective evaluation of uncertainties of measurements related to movements, such as driving in vehicles and biking (e.g. see figure A3) and external noise such as nearby doors opening and closing.

Earlier studies have taken the approach of developing an entire app (Kim et al.,

2015, 2016; Madaus and Mass, 2017; McNicholas and Mass, 2018), focusing both on User Experience (UX) and User Interface (UI). While such an approach represents a full solution, problems do exist due to the nature of crowdsourced data. Users do need a reason to download an app and keep using it for the researchers to obtain data. It is argued that, in general, research institutions are not in shape to advertise and maintaining an app to keep the retention and conversion rate high. Also, it makes applications dependent on SPOs from a single source, vulnerable if the source breaks down or users change to a new platform. Therefore, a different approach was used in this PhD project.

A software package² was developed to be installed on top of third-party apps to collect sensor data as a background process, when the app is in use. The software package is named SMARtphone Pressure System (SMAPS) and includes two sub-packages, ‘PMOB’, which works client-side, and ‘QCMOB’ which works server-side. SMAPS is described in detail in Hintz et al. (2019b). For testing purposes, a simple UI was built on top of PMOB; making it possible to see sensor data in real-time (see figure 2.2). PMOB was installed by the company SFS Development in their app ‘DMI Vejret’ (‘DMI Weather’) for both iOS (iOS App Store, 2019) and Android (Google Play, 2019) from where most of the SPOs used in this project originated.

Through the ‘DMI Vejret’ app, a large number of observations, in the range of 140,000 to 170,000, were obtained every day. In total, over one year, from 4th of April 2018 to 4th of April 2019, 61,728,672 observations were collected from 149,782 unique smartphones. That is approximately 2300 % more observations than what the traditional Danish SYNOP network delivers. August 2018 was the month with most observations with 9,928,301 reports, and June 2018 was the month with fewest observations with 3,131,730 reports. It is clear from these numbers that one has to take into account the natural variability which lies in crowdsourced data. One advantage of building the data collection software as a SDK is that the variability will decrease if the SDK is installed in multiple apps. Table 2.1 lists the total observations and unique devices. Note that the total of unique devices is not a sum as multiple devices can report in multiple months.

Block 2.1 shows a minimal example of initialising the iOS SDK for retrieving pressure observations using Swift 3.0.

²A better term for the software package would be SDK, but because the software is not yet operational and improvements might be needed in future it is chosen not to use that term.

Table 2.1: Overview of collected smartphone observations from 4th of April 2018 to 4th of April 2019. The decline in April 2019 is because only four days are included from this month.

	Total Observations	Unique Devices
April (2018)	4,256,983	35,974
May	3,155,072	37,672
June	3,131,730	34,142
July	6,857,070	56,450
August	9,928,301	66,143
September	6,179,692	57,692
October	4,534,645	54,607
November	3,146,321	47,072
December	3,748,830	46,219
January (2019)	5,343,834	54,570
February	4,027,405	52,628
March	6,814,125	54,336
April	604,664	22,904
Total	61,728,672	149,782

```

class PressureRx {
    private var pressureController: CMAltimeter? = { return
    CMAltimeter.isRelativeAltitudeAvailable() ? CMAltimeter()
    : nil }()
    static let shared = PressureRx()
    public var pressureCallback = PublishSubject<Double>()

    func initPressure() {
        pressureController?.startRelativeAltitudeUpdates(to: .
        main) { altitudeData, error in
            if let kpa = altitudeData?.pressure.doubleValue {
                self.pressureCallback.onNext(kpa*10)
            }
        }
    }
}

```

Block 2.1: Example of initialising the iOS SDK using Swift 3.0 for retrieving pressure observations. CMAltimeter is the class which is called from the iOS SDK. See text for details.

Figure 2.2 shows the sensor screens of the test-app, which is a UI on top of PMOB.

The left screen shows the numerical value from different sensors, such as the accelerometer. The right screen shows the time evolution of pressure. As seen in figure 2.2, not only pressure is collected. All variables that are retrieved and calculated are given in table 2.2.

Table 2.2: Overview of observed variables collected from the smartphones. The calculated variables are calculated either on the smartphone directly or server-side when the observation has been received. σ represents the standard deviation.

Observed Variable	Calculated Variable	Unit
Pressure	σ_p	hPa
Timestamp		ms since 1 Jan 1970
Latitude		Degrees (WGS94)
Longitude		Degrees (WGS94)
Altitude	σ_{Alt}	m
Horizontal Accuracy		m
Vertical Accuracy		m
Speed	σ_{Spd}	m/s
Acceleration \hat{x}	σ_{ax}	m/s ²
Acceleration \hat{y}	σ_{ay}	m/s ²
Acceleration \hat{z}	σ_{az}	m/s ²
	Terrain Model Height	m
User ID		

The standard deviation of pressure, altitude, speed and acceleration are calculated during the measurement and stored. The first five seconds of a measurement series is discarded due to a spin-up effect of the sensor (see McNicholas and Mass (2018) and Hintz et al. (2019b)).

Due to the uncertainties in altitude assignment, the metadata just described was collected to examine whether such could be used as a proxy for an altitude error. When an observation enters the database, the terrain altitude is calculated from the Danish terrain model (see section 2.2.1) and appended to the observation.

As have been stressed, one of the main motivations of this study is to improve the prediction of short-term severe weather events. However, data from smartphones can also have an application in climatology. Figure 2.3 shows a time-series of surface pressure during one month in December 2018 from a DMI operated SYNOP station close to Copenhagen (Kastrup) compared to observations from different smartphones that have

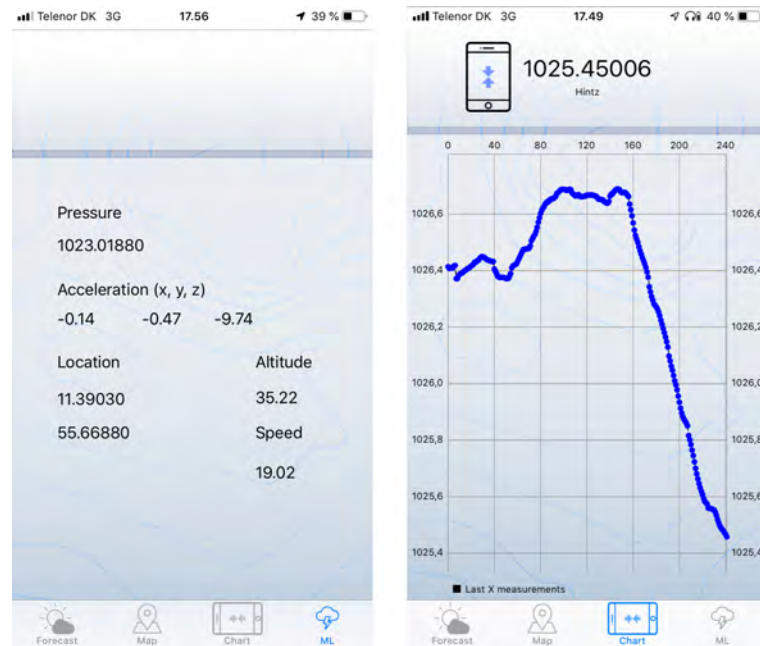


Figure 2.2: A UI on top of P MOB to access raw data in real time. Left: sensor values from barometer, accelerometer and location services. Right: Timeseries of pressure while the app is open.

been nearby (10 km) the location of the station. It is seen that the general tendency follows the SYNOP observation good. It is also seen clearly that there is a negative bias of the smartphone observations. The reason for this is that smartphones tend to be located more above the surface (e.g. tall buildings) than below the surface and hence a bias is introduced (Hintz et al., 2019b).

Figure 2.4 shows observations from two individual smartphones during one month. A pattern of movement is seen in both cases. Using more detailed maps, it would be possible to make it probable who owns the phone and hereafter track the movement of this person. This possibility makes the data personal data and handling of the data must then comply with the GDPR act from the European Union (European Union, 2018) (see section 2.4.1).

2.2.1 The Danish terrain model

To obtain an altitude of the terrain at the position of the smartphone, the Danish Terrain Model (DTM) (Danish Environmental Protection Agency, 2015) was used. The DTM version used here has a horizontal resolution of 10 m, but a data set with a

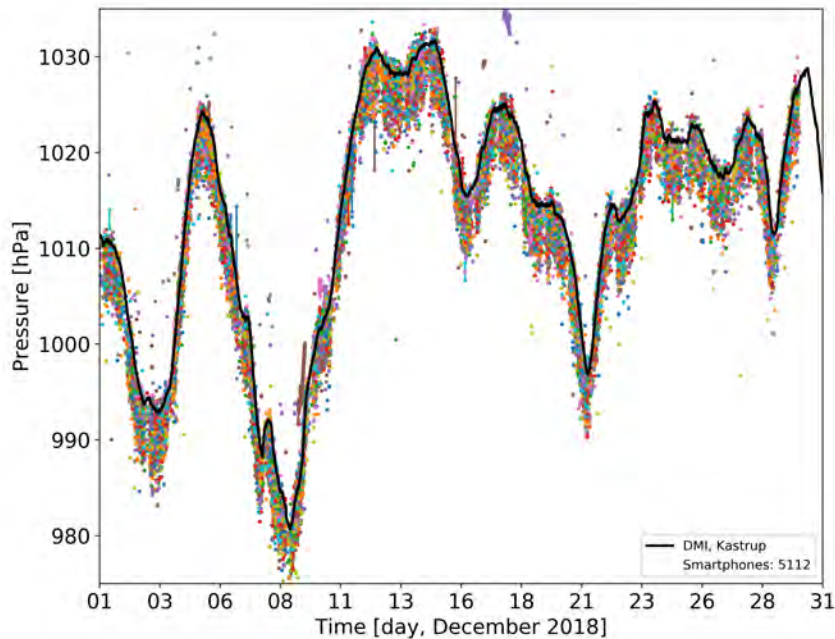


Figure 2.3: Timeseries of the pressure measured by the DMI operated SYNOP station in Kastrup, Denmark (Black solid line) compared to observations from different smartphones within 10 km from the station (coloured circles).

horizontal resolution of 0.4 m is also available. The Danish Environmental Protection Agency (2015) does not provide an API to fetch an altitude given a position. Therefore a module was made which return an altitude given a latitude and longitude as input. The module was written in Python and was used not only by this study but was also used by other research projects at DMI, e.g. for calculating altitude changes for vehicles in road weather forecasting. The module can read both a single input or a file containing a list of coordinates.

The dataset is separated into files each covering a tile of 10×10 km. When the module is initialised the first time, a list of corner-points of each file is generated. These are used to determine which file to open for a given coordinate. If a list of coordinates is given, the list is sorted to ensure files are only opened once to reduce Input/Output (I/O) overhead. The difference of terrain elevation for the DTM and an operational NWP model at DMI (see chapter 3) is shown in figure A2.

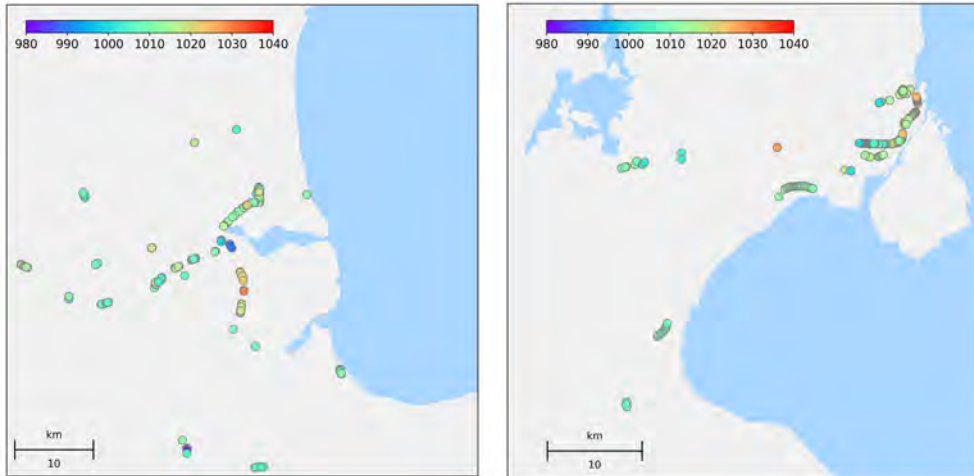


Figure 2.4: Pressure observations from two individual smartphones for one month. The ability to track individual phones are what makes the data personal. The details of the map are deliberately coarse to prevent identification possibilities. The colour bar shows the pressure in hPa.

2.3 Personal weather stations

PWS are a potentially useful source of observations both for direct and indirect measurements. Also, PWS are a potential source for both data assimilation and aiding of operational meteorologists. Clark et al. (2018) used observations from PWS to make an observational analysis of a hailstorm, in which a clear signal was seen which could not be seen in conventional observations due to a limited resolution. PWS has the advantage compared to mobile platforms (such as smartphones) that they are stationary. This makes bias-correction easier as a bias correction scheme does not have to be location-aware (Hintz et al., 2019c).

In Denmark, there are three main providers of PWS which connect easily to a network so that a user can see observations in near real-time, namely:

- Netatmo (netatmo.com)
- Lonobox (lonosoft.com)
- FieldSense (fieldsense.com)

Netatmo is the largest provider of these and is also available globally. Lonobox and FieldSense are currently mainly in the Danish market, and the latter is specialised for

agricultural purposes. In 2018 Netatmo partnered with YR, a collaboration between the Norwegian Meteorological Institute and NRK (a Norwegian TV-station), to provide more accurate forecasts (Netatmo SAS, 2018). Today, 2 m temperatures from Netatmo stations are used operationally in post-processing of the YR forecasts.

Netatmo provides a public available API to retrieve public observations (Netatmo SAS, 2019a). This study started collecting Netatmo data in February 2018 from Denmark. In March 2019, Dr T. Bøvith (DMI) extended the domain over which data were collected to cover most of north-western Europe (see figure A4).

2.4 Data management

Examining the efficiency of data storage systems is out of scope of this thesis, but a few notes based on the experiences gained during this work are given here.

The advantage of crowdsourced data is the potentially available quantity of observations. This, however, also creates challenges related to managing the data. Both in terms of legal issues (see section 2.4.1) and handling large data sets which in most cases is not possible on personal computers due to limited computing and storage resources.

Collection of data from smartphones was set up using the Google Firebase (Google Firebase, 2017) real-time database. Firebase was chosen because it is easy to connect with mobile applications. While Firebase is efficient and easy to use for applications providing user profiles, chats and other small separated datasets, Firebase is not well suited for handling big datasets. The Firebase real-time database can be considered as a single JSON object which can be modified in real-time. While the JSON format can be read fast enough for small datasets, it becomes time-consuming for large datasets. Reading in a binary format like NetCDF or GRIB is much faster.

To keep the Firebase database small, data are automatically downloaded daily and injected into a SQL database, where queries are much faster. In the same SQL database, observations from both smartphones and PWS are stored. Before applications of this study are made operational, it should be examined if other systems than Firebase are better suited.

2.4.1 European GDPR act

As noted in section 2.2, data collected from smartphones are in this particular case characterised as personal data, which is solely due to a unique identifier given to a device on the first install. The identifier is needed to be able to bias correct observations from a device. Could the bias-correction be done client-side there would be no need for an identifier and the legal issues would be much less. Unfortunately, this creates limits such as sharing data with other researchers. Before sharing, the data needs to be stripped from the unique id's. One example where this caused trouble was in Price et al. (2018). Here, observations were obtained from a third-party provider who could not provide device identifiers, and so Price et al. (2018) was not able to bias correct the observations.

Due to the unique identifier, data management must comply with the GDPR act from the European Union (European Union, 2018). Legal advice was requested from the law firm of 'Nørregaard Mieritz'. The PhD candidate was responsible for communication in this process. Seven documents³ were created, which in total covers the policy of collecting and using crowdsourced data at DMI in the HIRLAM-C project:

- Internal data policy
- Risk analysis
- List of contact details and data collected
- Declaration of consent
- Form of a request of the right of access
- External data policy
- Documentation sheet in case of a security breach

With these documents, the data collection from smartphones could be done without violating the GDPR act. The external data policy is what the user has to agree or disagree with, at the first use of the software. If the user disagrees, no data collection occurs. It is recommended here that legal advice is requested when starting a new work which involves crowdsourced data. Before this PhD project received legal advice from experts, there was a long, frustrating period with many weeks of work on legal issues that DMI and the PhD candidate of this study is not an expert.

³These documents are not included as a part of this thesis.

Chapter 3

Introduction to numerical weather prediction

Here a brief introduction to NWP is given before introducing the Vaavud NWP system in chapter 4 and the model which was used when assimilating SPOs, namely the HARMONIE model (Driesenaar, 2009) in chapter 10. Data Assimilation (DA) is treated separately in chapter 9 and chapter 10 with a general introduction and the DA system at DMI respectively.

In NWP a set of equations that describe the flow of the atmosphere is employed. The equations are solved numerically and include approximations and compromises of different complexity from one model to another. The equations form a system of non-linear coupled Partial Differential Equations (PDE), which can be solved to predict the future state of the atmosphere. The governing variables are the velocity $\mathbf{v} = (u, v, w)$, the pressure p , the temperature T , density ρ and the specific humidity q . Using Newton's laws of motion and conservation principles one can write a system of seven coupled equations with seven unknown variables (Holton, 2004a). First, the momentum equation for a rotating sphere can be written as

$$\frac{d\mathbf{v}}{dt} = \frac{\partial\mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla\mathbf{v} = -2\boldsymbol{\Omega} \times \mathbf{v} - \frac{1}{\rho}\nabla p + \mathbf{g} + \mathbf{F}, \quad (3.1)$$

where \mathbf{g} is the gravity including the centrifugal force, $\boldsymbol{\Omega} = (0, \Omega \sin \phi, \Omega \cos \phi)$ is the Earth's angular velocity where ϕ is the latitude and \mathbf{F} is the frictional force.

Conservation of mass is accounted for through the equation of continuity

$$\frac{d\rho_d}{dt} + \rho_d \nabla \cdot \mathbf{v} = 0, \quad (3.2)$$

where ρ_d is the density of dry air only. Continuity equations for various tracers, such as

water vapor and liquid water, are included explicitly on the same form (Kalnay, 2003), namely

$$\frac{d\rho_t}{dt} + \rho_t \nabla \cdot \mathbf{v} = S_t,$$

with S_t representing the sources and sinks for a given tracer. Equation (3.1) and equation (3.2) is often referred to as the Navier-Stokes equations.

Energy conservation is accounted for through the first law of thermodynamics

$$c_p \frac{dT}{dt} - \alpha \frac{dp}{dt} = J, \quad (3.3)$$

where c_p is the specific heat at constant pressure, $\alpha = \rho^{-1}$ and J is diabatic heating effects, such as latent heat release. The pressure, temperature and density are linked via the equation of state

$$p = \rho R_d T, \quad (3.4)$$

where $R_d = 287 \text{ J/K} \cdot \text{kg}$ is the universal gas constant for dry air. Finally the prognostic equation for specific humidity is

$$\frac{dq}{dt} = S_q, \quad (3.5)$$

with S_q representing all the sources and sinks for the specific humidity, q . Equation (3.1)-(3.5) are the governing exact equations, which can be said to be fundamental for all models. Often models have in practice more prognostic equations such as the French model, AROME, which have 12 prognostic variables (Seity et al., 2011).

In some models vertical accelerations are assumed to be small compared to the buoyancy such that the buoyancy force is balanced by the gravitational force¹. Then the vertical component of (3.1) reduces to the hydrostatic equation,

$$\frac{\partial p}{\partial z} = -\rho g. \quad (3.6)$$

The hydrostatic balance is valid for planetary and synoptic scales. When considering

¹Historically, this has been the case for most global models, but a non-hydrostatic dynamical core is expected to be implemented for the Global Forecasting System (GFS) model in December 2019 (NCEP/EMC Model Evaluation Group, 2019). Then, both the Integrated Forecasting System (IFS) model at ECMWF and GFS will have non-hydrostatic dynamical cores.

mesoscale phenomena, the hydrostatic approximation starts to be troublesome as there can be large vertical accelerations within convective systems and hence there is a grey-zone at which grid-size one must use a hydrostatic or a non-hydrostatic model (Petch et al., 2002). In general, though hydrostatic models are believed to be insufficient below a grid-size of 5 km (Field et al., 2017). However, large vertical velocities in hydrostatic models have been observed and are comparable to those generated by non-hydrostatic models given that the time step is not chosen too high (Hintz, 2015). The governing equations with the hydrostatic approximation (equation (3.6)) are often referred to as the primitive equations (Holton, 2004a) as these describe the fundamental variables one needs to predict the evolution of the atmosphere.

The derivation of the governing equations are given by Arya (2001); Kalnay (2003) and Holton (2004a).

3.1 Vertical coordinates

When solving (3.1)-(3.5) it is necessary to choose which coordinate system to use. One can have different criteria for selecting vertical coordinates. One common criterion is that the vertical coordinates should work well over both flat and steep terrain. Also, the pressure gradient force is a key parameter to represent well; therefore, p would be a better choice than using z as a vertical coordinate. Another advantage of using p over z is that the density, ρ , does not appear in isobaric coordinates. We have no way of measuring density directly, so this is a great advantage in the creation of an analysis. p does not, however, follow the terrain.

Phillips (1957) developed a vertical coordinate which follows the terrain based on the pressure. Phillips (1957) defined the σ -coordinate as $\sigma = p/p_s$, where p_s is the pressure at surface level. While the σ -coordinates do follow the terrain by definition, it also comes with a cost. Over steep terrain, the steepness of individual surfaces will extend throughout the atmosphere. Away from the surface, there are practical and numerical advantages of choosing a coordinate that follows isobaric surfaces (Mesinger et al., 1988; Simmons and Strüfing, 1981), as it improves the accuracy of the calculation of the pressure gradient force.

Eckermann (2009) noted that a vertical coordinate can take the functional form

$$\eta = h(p, p_s), \tag{3.7}$$

where h is a monotonic function of p . η is then defined to satisfy a lower and upper boundary condition, namely that $h(p_s, p_s) = 1$ and $h(p_t, p_s) = 0$ where p_t is the pressure at the highest model layer. The simplest choice of η would then be, following common normalisation,

$$\eta = \frac{p - p_t}{p_s - p_t} = \sigma, \quad (3.8)$$

which is a modified form of the σ -coordinates defined by Phillips (1957). Simmons and Strüfing (1981) presented the hybrid $\sigma - p$ coordinates through the relation

$$p(\eta, p_s) = A(\eta) + B(\eta)(p_s - p_t), \quad (3.9)$$

where $A(\eta)$ and $B(\eta)$ are two coefficients which controls the isobaric and terrain following properties (Eckermann, 2009). For $A(\eta) = 0$ and $B(\eta) = 1$ the coordinate is completely terrain following and for $A(\eta) = 1$ and $B(\eta) = 0$ the coordinate are completely isobaric (Laprise, 1992; Mesinger et al., 1988; Simmons and Strüfing, 1981). For a more general overview of vertical coordinates the reader is referred to Laprise (1992). The HARMONIE model, which is used in this PhD project, uses the hybrid $\sigma - p$ coordinates (Driesenaar, 2009; Seity et al., 2011), whereas the Weather Research and Forecasting (WRF) model uses the σ -coordinate (Skamarock et al., 2008).

3.2 Numerical methods

Numerical methods as a subject are out of the scope of this study; thus this section is a simple description of the problems arising when solving the governing equations in practice, to make a brief introduction to the unaware reader. For a detailed overview, the reader is referred to the literature cited in the following.

Equations such as the momentum equation (equation (3.1)) need to be solved by discrete approximations to the continuous equations. Computers do not perform calculus, so derivatives have to be approximated to be solved with arithmetic. Also, with no analytic solutions, approximations have to be found using numerical methods. Consider a simple PDE that describes the change of a variable ψ with respect to time, t . Numerically this can be approximated as a Finite Difference Equations (FDE) as (Durrant, 2010)

$$\frac{d\psi(t)}{dt} \approx \lim_{\Delta t \rightarrow 0} \frac{\psi(t_n + \Delta t) - \psi(t_n)}{\Delta t} \quad (3.10)$$

dt is approximated by Δt while assuring Δt is sufficiently small. For $t_n = 0$, $\psi(t_n)$ would be the model's initial conditions determined with DA (see chapter 9) and $\psi(t_{n+1})$ is the predicted state for the next timestep. Equation (3.10) can be solved by using the forward Euler method as

$$\phi_{t_{n+1}} = F(\phi_n, t_n)\Delta t + \phi_{t_n}, \quad (3.11)$$

where ϕ is the numerical approximation of ψ . Equation (3.11) is not used in NWP in practice because solutions of this method quickly becomes numerical unstable. More advanced schemes exist which are more stable and allow larger timestepping (Ascher et al., 1997; Durran, 2010; Kalnay, 2003; Kar, 2006; Tumolo and Bonaventura, 2015) schemes. It is desirable to be able to increase the timestep without loss of quality as this will decrease the computational costs of simulations. The introduction of semi-Lagrangian schemes allowed this (Robert, 1981) because such schemes are much more stable than traditional Euler based schemes such as used by the WRF model.

3.3 Verification

Verification is crucial to know if changes to a given NWP model is improving or worsening the forecast. Also, it is important to have standardised verification methods when comparing models to each other. The choice of verification scores naturally depends on what should be measured. However, in the case of using crowdsourced data where the quantity of surface observations is much higher than conventional observations (see chapter 2) a few general remarks can be given. Naturally, to gain most of the high quantity of crowdsourced data in NWP it is necessary to use high-resolution models, such as HARMONIE. One issue, though, that can occur when verifying high-resolution models is the double-penalty problem, i.e. objective verification scores may be better for low-resolution models than for high-resolution models (Mittermaier et al., 2013; Nurmi, 2003; Skok and Roberts, 2016). Figure 3.1 shows an illustration of the double penalty problem. The blue circles represent two rain gauge precipitation observations, and the red squares represent a model forecast at a horizontal resolution of 3 km (left) and 6 km (middle). If using standard point-verification, such as RMSE grid-point by grid-point, the high-resolution model (left) would receive a double penalty because the model misses the observed precipitation and falsely predicts precipitation where there are none observed.

The low-resolution model (middle) in figure 3.1 will get a higher score than the

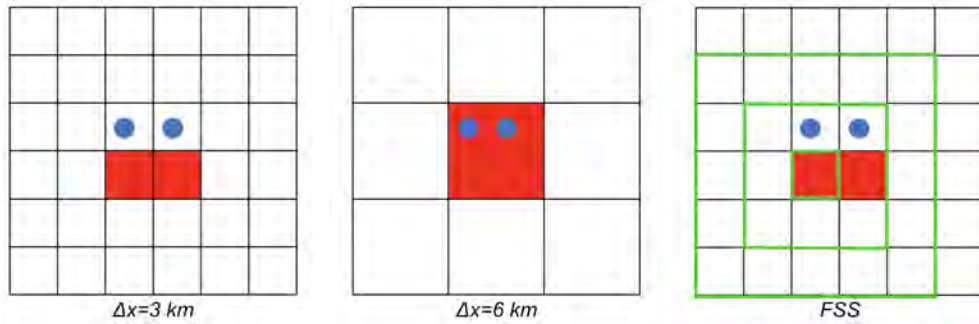


Figure 3.1: Illustration of the double penalty problem. Blue circles represent observations of precipitation. Red squares represent the model forecast of precipitation in grid-point space. Black squares represent the grid-boxes. The horizontal resolution is 3 km and 6 km for the left and middle figure, respectively. The right figure is an illustration of the Fractional Skill Score (FSS).

high-resolution model for this case because it correctly predicts the location of the precipitation in grid-point space. However, it is easy to imagine that the precipitation pattern, in this case, is better predicted by the high-resolution model, which offers more details on the structure of the event than the low-resolution model.

3.3.1 Fractional skill score for precipitation verification

One measure that can be used to avoid the double-penalty problem is Fractional Skill Score (FSS) (Roberts and Lean, 2008), which is a scale-aware score aimed for precipitation. A DMI FSS Python routine based on Roberts and Lean (2008) was developed as a part of Hintz (2015) prior to this project. During this study, a FSS Fortran routine was developed to speed up calculations when going towards large scales, which was used in Hintz et al. (2019b).

The method of FSS is shown in the right illustration of figure 3.1, where the neighbourhood is taken into account. First, a score is calculated based on grid-scale $N = 1$, such that a score between 0 and 1 is calculated. Then the size of the neighbourhood is increased to contain 9 grid-boxes ($N = 3$) illustrated by the middle green square in figure 3.1 so the forecast is smoothened. This continues up to the full scale of the domain.

Before doing so, the model field and the observational (precipitation) field needs

to be converted onto the same grid. In this PhD project, precipitation estimates are calculated from radar reflectivity assuming a Marshall-Palmer distribution (Marshall and Palmer, 1948) as

$$P = \left(\frac{Z}{a}\right)^{1/b}, \quad (3.12)$$

where P is the precipitation intensity, Z is the reflectivity factor and a and b are empirical constants. a and b are not well defined Fujiwara (1965), but typical values are $a = 200$ and $b = 1.6$ (Marshall and Palmer, 1948). Z is found by using the ‘radar equation’ which gives the returned energy to the radar, \overline{P}_r , as a function of the reflectivity factor as (Battan, 1973)

$$\overline{P}_r = \frac{CK_w^2}{r} Z. \quad (3.13)$$

Here K_w^2 is the dielectric constant for water, r is the distance from the radar, C is a radar constant describing radar characteristics and \overline{P}_r is the returned energy measured by the radar. Radar reflectivity is often given on a logarithmic scale as a ratio between Z and a reference value. At DMI the reference value is defined as the reflectivity for 1 m^3 with one droplet of a diameter of 1 mm ($Z = 1 \text{ mm}^6/\text{m}^3$). With this, radar reflectivity is given in dBZ as

$$\text{dBZ} = 10 \cdot \log_{10} \left(\frac{Z}{1}\right). \quad (3.14)$$

At DMI, dBZ spans from -30 dB to 70 dB , where $\text{dBZ} = 0$ implies that $Z = 1 \text{ mm}^6/\text{m}^3$. Then equation (3.12) can be integrated over time to obtain the precipitation amount. Finally, the precipitation amount is interpolated from the radar grid, which has a horizontal resolution of 500 m , to the model grid to be compared with the NWP precipitation forecast using FSS.

The details of the FSS equations are given by Roberts and Lean (2008) and a summary is found in Hintz et al. (2019b) (see chapter 14). However, it is highlighted that FSS is a score that lies in the range 0 to 1 for each size of the neighbourhood, where 0 is the worst possible score, and 1 is a perfect forecast. Both the model field and the observational field is converted to binary fields before computing the FSS. This can be done either by using a percentile or a threshold. The advantage of choosing a percentile is that the score will then be a measure of the placement of the highest precipitation

rates, whereas using a threshold will hold information about the precipitation amounts also (Olsen et al., 2015). Using a percentile also ensures that the score converges towards 1 as the size of the neighbourhood increases.

Chapter 4

The Vaavud system for numerical weather prediction

A system for NWP was set up at Vaavud for two purposes. First, to provide high-resolution weather forecasts for the users of Vaavud products (mainly kite-surfers, sailors, hunters and farmers) and secondly to obtain a numerical reference and a tool for research activities. The purpose of this chapter is to describe the practical implementation, and the long-term plans of the system had Vaavud continued business. Most importantly the model was used for the research to be described in chapter 8.

The model which was implemented at Vaavud was the open-source WRF (Skamarock et al., 2008, Version 3.8.1) model using the Advanced Research WRF (ARW) dynamic core. Due to computational limitations DA (see chapter 9) was not a part of the NWP system, but was instead initiated from an analysis from the GFS model (NCEP, 2018). However, the NWP system involves both pre-processing, computation, post-processing and data delivery to an API.

In practice, computations were done on computing facility services to avoid maintaining hardware. Two separate instances were created, named ‘Odin’ and ‘Frigg’, respectively. The Frigg instance is the compute instance which performs the WRF model computations, and Odin is a small server which serves as a helper for Frigg, such as handling data and initiation of new cycles via a scripting environment developed for the purpose. Odin was set up on the Amazon elastic compute service (Amazon, 2019) and Frigg was set up on the High-Performance Computing (HPC) system of Abacus at University of Southern Denmark (DeiC Abacus, 2018). A few of the technical details of Odin and Frigg are given in table 4.1.

Table 4.1: An overview of the two Amazon instances used at Vaavud. Frigg is the primary compute instance used to run WRF.

Name	Instance Type	Nodes	CPUs/Node	Cores/Node	Memory [gb]/Node	Provider
Odin	General Purpose	1	2	2	4	Amazon EC2
Frigg	Computing	16	2	12	64	Abacus HPC

4.1 Script environment on Odin

Odin is the primary data handler of the Vaavud NWP system and data handled by Odin has been used in the article introduced in chapter 8; thus it is relevant to mention the workflow of Odin briefly. The primary scripting language used on Odin is Python. Odin is written inside a Docker container (Docker, 2019), which can be thought of as an independent software package with its own libraries and configurations. Using the ‘container-approach’ makes it easy to develop client-side and deploy to a server when the software package is ready to go into operation. When the container is initialised using Docker, the model environment is operational, meaning that Odin starts to listen for new initial conditions to start Frigg. When all initial conditions are ready and prepared, Odin starts up Frigg which then runs WRF. Afterwards, Odin receives the model output data and delivers it to the Vaavud Weather API and stores relevant data for later use, such as is the case for the research purposes of this thesis. A monitor system was also created to monitor the operational model setup. If Frigg aborts unexpectedly, messages are sent to relevant people either via e-mail or via Slack (a communication platform). The latter has the advantage that it is possible to send push notifications to only those people who need to take action.

4.2 WRF implementation on Frigg

The WRF model implementation described here is the same as was used in the article described in chapter 8, here with details that were not possible to include in the article. The WRF model is primarily written in a combination of both Fortran77 and Fortran90, a HPC language suitable for NWP applications.

The Frigg-WRF model was set up with a domain focused on centring Denmark and the Netherlands. The initial conditions and boundary conditions were provided by the GFS model (NCEP, 2018). The Frigg-WRF model consists of two domains, an outer and inner domain (see figure 4.1). The outer domain has a horizontal resolution of 9 km, and the inner domain (inner red square in figure 4.1) has a horizontal resolution

of 3 km. The inner nest is set up as a two-way nest. For further details on the WRF model see Skamarock et al. (2008). Many settings for running WRF can be changed, such as which physics parameterisation schemes to use, timestep, vertical levels. The namelist for the Frigg-WRF setup is given in appendix D.

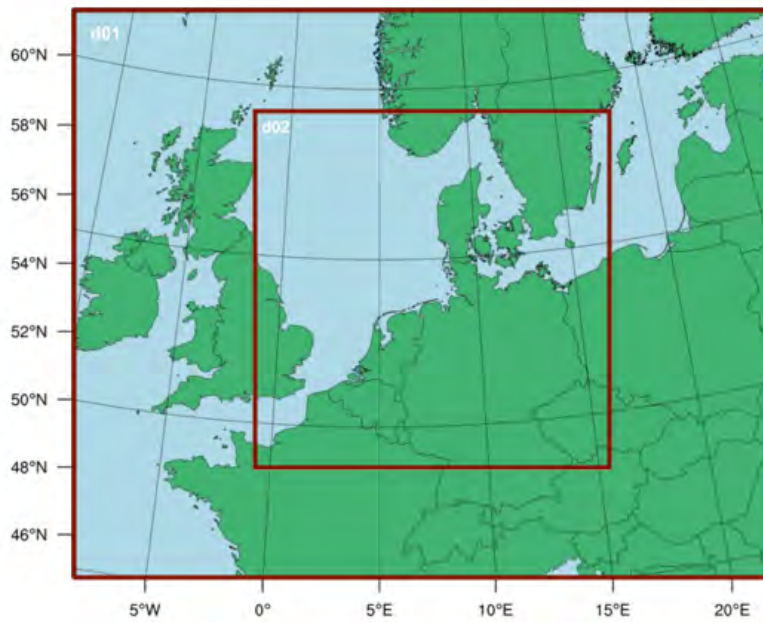


Figure 4.1: Model domain of the Vaavud Frigg-WRF model. The inner red square shows the inner domain.

Different compiler options were tested with a focus on speed and accuracy. Ultimately the Frigg-WRF model was compiled to Distributed-Memory Parallelism (DM-PAR) using Intel compilers. The GNU GCC compilers were also tested, but an increase of factor 1.6 was seen when changing to Intel compilers. When the Frigg-WRF computation is completed, post-processing is performed on the model output. The post-processing module computes standard diagnostics such as the Sea Level Pressure (SLP) and cloud cover.

Chapter 5

Fundamentals of atmospheric turbulence

The theory of the Atmospheric Boundary Layer (ABL) is essential for the work of Hintz et al. (2019a). However, before introducing the ABL in chapter 6, an introductory discussion on the fundamental theory of turbulence is given, which is used to build upon in chapter 6.

First, the Navier-Stokes equations are considered, which leads to the Boussinesq approximation under certain assumptions (section 5.1), which describes the flow in a shallow boundary layer. From these it will be shown how turbulence can be described in the ABL using Reynolds decomposition (section 5.2) and finally a brief note is given on the closure problem in section 5.3.

A fluid in a boundary layer has to be treated as a viscous fluid. Even though no fluids are completely inviscid, it is a useful approximation outside a boundary layer (Buresti, 2009). An inviscid fluid is non-turbulent and can pass objects without friction or drag. Friction and transfer of heat and momentum are essential processes in the ABL and therefore the fluid in the ABL has to be treated as a viscous fluid. Friction between adjacent fluid layers per unit area is called shearing stress, τ_0 , and is a function of the velocity gradient, and the dynamic viscosity, μ , so that $\tau_0 = \mu \partial u / \partial z$ (Arya, 2001). The shearing stress should be interpreted as a resistance force to shearing flows. Viscous flows can furthermore be divided into laminar and turbulent flows. As the name implies, the laminar flow has only little mixing, whereas turbulent flows are almost random and unpredictable and have a high degree of mixing (Wyngaard, 2010).

5.1 The Boussinesq approximation

Assume there exists a reference state of the atmosphere at rest, which is characterised by ρ_0 and p_0 which satisfies the hydrostatic equation ($dp_0/dz = -\rho_0 g$). Further, in the real atmosphere, small deviations from the reference exist such that

$$\rho = \rho_0(z) + \rho_1(x, y, z, t) \text{ and} \quad (5.1)$$

$$p = p_0(z) + p_1(x, y, z, t), \quad (5.2)$$

where ρ_1 and p_1 are assumed to be small compared to the reference state. Assuming that the density, ρ , is constant ($\rho = \rho_0$) everywhere in the ABL, except for the buoyancy force in the vertical momentum equation, one obtains the Boussinesq approximation in which the governing equations (equation (3.1) to (3.4)) reduces to

$$\frac{du}{dt} = -\frac{1}{\rho_0} \frac{\partial p_1}{\partial x} + fv + F_x, \quad (5.3)$$

$$\frac{dv}{dt} = -\frac{1}{\rho_0} \frac{\partial p_1}{\partial y} - fu + F_y, \text{ and} \quad (5.4)$$

$$\frac{dw}{dt} = -\frac{1}{\rho_0} \frac{\partial p_1}{\partial z} - \frac{\rho_1}{\rho_0} g + F_z. \quad (5.5)$$

in cartesian coordinates. Further, expanding the equation of state (equation (3.4)) and assuming that the pressure variations within the ABL is negligible one can write equation (5.5) in terms of the potential temperature, θ , as (Bluestein, 1993)

$$\frac{dw}{dt} = -\frac{1}{\rho_0} \frac{\partial p_1}{\partial z} + \frac{\theta_1}{\theta_0} g + F_z. \quad (5.6)$$

The frictional forces can be expressed in terms of the kinematic viscosity, ν , as (Holton, 2004b),

$$\begin{aligned} F_x &= \nu \left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right] = \nabla^2 u, \\ F_y &= \nu \left[\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right] = \nabla^2 v \text{ and} \\ F_z &= \nu \left[\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right] = \nabla^2 w. \end{aligned}$$

In a shallow atmosphere it is further assumed that the air is incompressible so the continuity equation (equation (3.2)) reduces to

$$\nabla_z \cdot \mathbf{v} = 0. \quad (5.7)$$

Expanding the potential temperature as $\theta = \theta_0(z) + \theta_1(x, y, z, t)$, in the first law of thermodynamics, the thermodynamic equation can be written as (Holton, 2004a)

$$c_p \frac{d \ln(\theta_0 + \theta_1)}{dt} = J/T, \quad (5.8)$$

where c_p is the specific heat capacity at constant pressure and J is the heating rate due to diabatic processes such as latent heat release and radiation processes. For adiabatic processes, $J = 0$.

Equation (5.3), (5.4), (5.6) and (5.7) are the Boussinesq approximation, valid for a shallow atmosphere, i.e. where no deep convection is present. It must be noted that the assumption of incompressibility would be a poor choice in a NWP model and are in practice never used (Kalnay, 2003).

5.2 Reynolds decomposition

Turbulence occurs on many scales in both time and space (Stull, 1988). Because turbulent flows behave as a stochastic process, statistical measures are used to express effects of turbulence in terms of the mean flow. To describe turbulence, Reynolds decomposition is used. Instantaneous variables are decomposed into a slowly varying mean part and a highly varying turbulent part as

$$\begin{aligned} u &= \bar{u} + u', \\ v &= \bar{v} + v', \\ w &= \bar{w} + w' \text{ and} \\ \theta &= \bar{\theta} + \theta', \end{aligned} \quad (5.9)$$

where $\bar{()}$ is the slowly varying large-scale part and $()'$ is the turbulent part. Usually, the average is calculated as a time mean where one makes a continuous record over time and taking the mean of all the samples. The average time should be chosen so that it does not blur real events such as diurnal heating but long enough to contain information of the largest eddies (Arya, 2001). Before substituting equation (5.9) into the Boussinesq

equations a note on the Reynolds conditions (Monin and Yaglom, 1971) is made. By definition, it is required that the average of the fast fluctuating part vanishes and that the average of an average is the average itself. Practically the Reynolds conditions are (Cotton et al., 2011)

$$\begin{aligned}
\overline{\bar{a}} &= \bar{a}, & \overline{a + b} &= \bar{a} + \bar{b}, \\
\overline{a'} &= 0, & \overline{a'b} &= 0, \\
\overline{c\bar{a}} &= c\bar{a}, & \overline{\frac{\partial a}{\partial t}} &= \frac{\partial \bar{a}}{\partial t} \\
\overline{\bar{a}b} &= \bar{a}\bar{b}
\end{aligned} \tag{5.10}$$

where c is a constant and a and b is dependent variables. Substituting equation (5.9) into equations (5.3) to (5.8) and using the rules of (5.10) yields the Reynolds averaged equations as (Arya, 2001)

$$\begin{aligned}
\frac{\partial \bar{u}}{\partial t} + \bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{v} \frac{\partial \bar{u}}{\partial y} + \bar{w} \frac{\partial \bar{u}}{\partial z} &= f\bar{v} - \frac{1}{\rho_0} \frac{\partial \bar{p}_1}{\partial x} + \nu \nabla^2 \bar{u} \\
&\quad - \left(\frac{\partial \overline{u'^2}}{\partial x} + \frac{\partial \overline{u'v'}}{\partial y} + \frac{\partial \overline{u'w'}}{\partial z} \right), \tag{5.11}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \bar{v}}{\partial t} + \bar{u} \frac{\partial \bar{v}}{\partial x} + \bar{v} \frac{\partial \bar{v}}{\partial y} + \bar{w} \frac{\partial \bar{v}}{\partial z} &= -f\bar{u} - \frac{1}{\rho_0} \frac{\partial \bar{p}_1}{\partial y} + \nu \nabla^2 \bar{v} \\
&\quad - \left(\frac{\partial \overline{u'v'}}{\partial x} + \frac{\partial \overline{v'^2}}{\partial y} + \frac{\partial \overline{v'w'}}{\partial z} \right), \tag{5.12}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \bar{w}}{\partial t} + \bar{u} \frac{\partial \bar{w}}{\partial x} + \bar{v} \frac{\partial \bar{w}}{\partial y} + \bar{w} \frac{\partial \bar{w}}{\partial z} &= g \frac{\bar{\theta}_1}{\theta_0} - \frac{1}{\rho_0} \frac{\partial \bar{p}_1}{\partial z} + \nu \nabla^2 \bar{w} \\
&\quad - \left(\frac{\partial \overline{w'u'}}{\partial x} + \frac{\partial \overline{w'v'}}{\partial y} + \frac{\partial \overline{w'^2}}{\partial z} \right), \tag{5.13}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \bar{\theta}}{\partial t} + \bar{u} \frac{\partial \bar{\theta}}{\partial x} + \bar{v} \frac{\partial \bar{\theta}}{\partial y} + \bar{w} \frac{\partial \bar{\theta}}{\partial z} &= \alpha_h \nabla^2 \bar{\theta} \\
&\quad - \left(\frac{\partial \overline{u'\theta'}}{\partial x} + \frac{\partial \overline{v'\theta'}}{\partial y} + \frac{\partial \overline{w'\theta'}}{\partial z} \right), \tag{5.14}
\end{aligned}$$

$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial z} = 0, \tag{5.15}$$

For the derivation the reader is referred to either of the following: Arya (2001); Cotton et al. (2011); Holton (2004a); Wyngaard (2010) and Bluestein (1993). The parentheses on the right-hand side are the turbulent flux divergence terms which appear due to the Reynolds averaging and contains unknown variances and covariances. An important note from equation (5.11) to (5.15) is that they contain more unknown variables than there are equations. Ultimately, the equations are therefore not closed and in principle

unsolvable. This problem is in the literature referred to as the *closure problem* and different assumptions made to solve them are called *closure assumptions* (see section 5.3).

The covariances are representing the turbulent fluxes and are very important for efficient mixing and diffusion in the ABL. For example, $\overline{u'v'}$ is a horizontal turbulent flux of momentum (if multiplied by the density). $\overline{u'w'}$ is the vertical flux of momentum and $\overline{w'\theta'}$ is the vertical turbulent heat flux. Kaimal and Wyngaard (1990) measured and estimated the turbulent fluxes and calculated the standard deviations in the 1968 Kansas experiment (Izumi, 1971). Kaimal and Wyngaard (1990) found a correlation coefficient for $\overline{w'\theta'}$ of 0.59, while the correlation coefficient for $\overline{u'\theta'}$ was found to be -0.16, indicating that w' and θ' are highly correlated, while u' and θ' are negatively weakly correlated.

Another measure of turbulence are the standard deviations, $\sigma_u = \overline{u'^2}^{1/2}$, $\sigma_v = \overline{v'^2}^{1/2}$ and $\sigma_w = \overline{w'^2}^{1/2}$ used to define turbulent intensities as $\sigma/|\mathbf{v}|$ (Counihan, 1975), where \mathbf{v} is the mean wind speed. Turbulent intensities are a key part of Hintz et al. (2019a) (see chapter 8) to estimate the surface roughness length (see section 6.4).

5.3 The closure problem

Detailed analysis and evaluation of solutions to the closure problem introduced in section 5.2 are out of scope of this work; thus, only some brief general notes are given. The closure problem can be solved using both first-order and higher-order models (Wyngaard, 2010). An example of a first-order model is a parameterisation of the vertical turbulent heat flux in the thermodynamic heat equation (equation (5.14)), expressed in terms of the mean flow as

$$-\overline{w'\theta'} = K_h \frac{\partial \bar{\theta}}{\partial z}, \quad (5.16)$$

where K_h is a coefficient of eddy diffusivity, hence by parameterising turbulent fluxes in terms of known variables the unknowns are reduced and equation (5.11) to (5.15) can then be solved under these assumptions. Another parameterisation is the mixing length hypothesis in which the vertical turbulent momentum flux is described in terms of the mean flow and a characteristic length scale (see section 6.2.1.2).

Chapter 6

The atmospheric boundary layer

The ABL is the part of our atmosphere closest to the surface, and since everyday-life takes place here, it is of great interest to have a good understanding of the energy budget, such as the exchange of heat between the surface and the atmosphere. The ABL is the layer of the atmosphere where most of the exchange of momentum and heat takes place (Arya, 2001). Therefore, sharp variations in properties of the flow are also found here, such as wind speed and temperature. The ABL is in some literature defined as the lowest 1000 m to 1500 m of the atmosphere (Counihan, 1975). This is an acceptable rule of thumb, but it is incorrect to define the height of the ABL with a constant. The height depends among others on the rate of cooling and heating of the surface, wind shear, large-scale vertical motion and roughness of the surface (Garratt, 1994). Therefore, the height of the ABL is not constant but changes both with time and space. However, the ABL can be defined as the layer where turbulent mixing is dominant. The ‘free atmosphere’ is above the ABL. Here the turbulence is weak so that the wind is approximately geostrophic.

The ABL itself is also divided into parts. In the lowest part, the surface layer (approximately the lowest 10 %) is found. The surface layer is also sometimes called the constant flux layer (Wyngaard, 2010) and the constant shear stress layer (Counihan, 1975). The surface layer is of particular interest to this project as the mean wind flow is often assumed to follow a logarithmic law in this layer (see section 6.2.1.1 and chapter 8). Between the surface and the surface layer, canopy sublayers (roughness sublayers) are found where the flow is disturbed by individual roughness elements. The canopy layer varies in height dependent on the surface characteristics. Over smooth surfaces, such as sea ice, the height can be a few millimetres, and over urban areas, the height can extend up to several decameters (Langleben, 1974; Tennekes, 1973).

Above the surface layer, the mixed layer is found¹. In the top of the ABL, an interfacial layer is found. It can be thought of as a boundary zone where information from the free atmosphere and the ABL is exchanged. This is a somewhat simplified description of the ABL, in practice, effects such as changes in surface characteristics creates an Internal Boundary Layer (IBL) which also needs to be taken into account.

The objective of the following sections in this chapter is to describe fundamental methods and variables of particular interest for the work presented in Hintz et al. (2019a) (see chapter 8).

6.1 Similarity theory

Similarity theory is used to a great extent in micrometeorology. By defining a fundamental set of dimensions (such as mass, length and time), one can express the dimensions of all quantities under consideration.

6.1.1 The fundamentals of similarity theory

Similarity theory is based on the Buckingham Pi theorem which can be interpreted as the optimal approach to determine a variable in a given problem. The theorem states that if one can identify the $m - 1$ parameters governing the dependent variable and n is the number of dimensions (such as length, time and mass) represented by the parameter, then the following are true (Wyngaard, 2010):

- If m quantities (Q_1, Q_2, \dots, Q_m), involving n fundamental dimensions form a dimensionally homogeneous equation, the the relationship can always be expressed as $m - n$ dimensionless groups ($\Pi_1, \Pi_2, \dots, \Pi_{m-n}$) made from the original m quantities. That is, $m - n$ independent dimensionless quantities can be formed. An independent quantity is one that cannot be made from the others.
- The $m - n$ independent dimensionless quantities are functionally related so that the dependent variable can be taken as a function of the governing parameters. That is, the dimensional relationship $f(Q_1, Q_2, \dots, Q_m) = 0$ is equivalent to the dimensionless relationship $F(\Pi_1, \Pi_2, \dots, \Pi_{m-n}) = 0$.

The Buckingham Pi theorem also indirectly states the benefits of considering dimensionless groups instead of dimensional variables. It follows from the theorem that the

¹Wyngaard (2010) emphasised that this should not be taken as a mixing layer, but a mixed layer.

non-dimensionalization always reduces the number of involved parameters. For example, if only one dimensionless group ($m - n = 1$) can be formed out of all the quantities, then that group must be a constant. It should be stressed that the Buckingham Pi theorem is only a mathematical theorem. The theorem does not come out with any physics of the problem. Therefore the theorem is only used as a part of the similarity theories. One must use physical intuition to develop a useful similarity theory.

6.1.2 The Monin-Obukhov similarity

The Monin-Obukhov similarity theory has provided much of the present understanding of the atmospheric surface layer. It was developed through the similarity theory described in section 6.1.1.

Monin and Obukhov (1954) argued that in a horizontally homogeneous surface layer the mean flow and turbulence characteristics depend only on four independent variables, namely, the height above the surface, z , the surface drag, τ_0/ρ , the kinematic surface heat flux, $H_0/\rho c_p$ and the buoyancy, g/θ_0 . Here, τ_0 is the wall shear stress and H_0 is the heat flux. For simplicity the definitions $Q_0 = H_0/\rho c_p$ and $u_* = \sqrt{\tau_0/\rho}$ are used. Under these assumptions the Monin-Obukhov length is given as (see Appendix B for the derivation),

$$L = -\frac{u_*^3 \theta_0}{kgQ_0}, \quad (6.1)$$

where $k = 0.4$ is the von Kármán constant. L was introduced as the characteristic height of the sublayer of dynamic turbulence. L may vary between $-\infty$ to ∞ . When L is negative, it corresponds to a positive surface heat flux ($Q_0 > 0$) which typically occurs in the daytime where the atmosphere receives heat from the surface. In nighttime L is usually positive since typically $Q_0 < 0$ here (the surface receives heat from the atmosphere). Over arctic surfaces, it can be seen that the cold surface receives heat from the atmosphere even during the day.

6.2 The wind profile in the atmospheric boundary layer

Many factors influence the wind profile in the ABL. Large-scale horizontal pressure gradients, such as highs and lows, are the main driving force for the flow in the ABL, together with horizontal temperature gradients. Other factors are advection of heat

and momentum, thermal stratification, entrainment from the free atmosphere, the earth rotation and the surface roughness (Garratt, 1994) (see section 6.4). In the remainder of this section the neutral surface layer (section 6.2.1) and the non-neutral surface layer (section 6.2.2) are treated separately.

6.2.1 The neutral surface layer

The wind profile in the neutral ABL can be derived with different approaches more or less sophisticated. Here we shall go through the two most common approaches.

6.2.1.1 The classical approach

The classical derivations involve dimensional analysis and similarity-theory. Any viscous sublayer is neglected, and the Coriolis force is neglected, and the momentum flux is assumed constant, independent of height. It is then assumed that for a horizontally homogeneous neutral boundary layer the mean wind shear is only dependent on the height above the surface, z , the surface drag, τ_0 , and the density of the fluid, ρ . The kinematic momentum flux is obtained by combining the surface drag and the fluid density into their ratio, $\tau_0/\rho \equiv u_*^2$. With these assumptions we look for a function $f(\frac{\partial U}{\partial z}, z, u_*) = 0$ (see section 6.1.1). This function has 3 quantities and 2 fundamental dimensions so that only 1 dimensionless group given as $kz/u_*\partial U/\partial z = 1$ can be formed. Integrating from z_0 to z and using that $U(z_0) = 0$ yields the logarithmic wind profile law for a neutral surface layer,

$$U(z) = \frac{u_*}{k} \ln \frac{z}{z_0}, \quad (6.2)$$

where z_0 is the roughness length (see section 6.4). This derivation gives only little information on the validity of the logarithmic wind profile; however, it has been verified by laboratory work, on which an overview is given by (Counihan, 1975) and (Arya, 2001).

6.2.1.2 Mixing length

The logarithmic wind profile can also be derived by using the mixing length hypothesis, proposed initially by L. Prandtl, which parameterises turbulent mixing by using a characteristic length scale, l , defined as the mixing length, the length at which a fluid parcel travels before being mixed with its surroundings. The mixing length hypothesis

states that the vertical momentum flux can be expressed as (Arya, 2001)

$$\overline{u'w'} \approx -\overline{l}^2 \left(\frac{\partial U}{\partial z} \right)^2. \quad (6.3)$$

By assuming that the mixing length can be written as $l = kz$ and assuming a horizontally homogeneous surface layer, such that $\overline{u'w'} = -\tau_0/\rho = -u_*^2$, equation (6.3) reduces to the logarithmic wind law.

6.2.2 The wind profile in non-neutral surface layers

For non-neutral surface layers equation (6.2) is invalid as stability effects are not taken into account. This is accounted for by the Monin-Obukhov similarity theory (see section 6.1.2), from which the similarity prediction of the mean wind shear is (Wyngaard, 2010)

$$\frac{kz}{u_*} \frac{\partial U}{\partial z} = \phi_m(z/L), \quad (6.4)$$

where $\phi_m(z/L)$ is an universal similarity function which have been determined empirically by numerous authors (Businger et al., 1971; Panofsky, 1963; Paulson, 1970a) on the form

$$\phi(z/L) = \begin{cases} 1 - \beta \frac{z}{L} & \text{if } z/L \geq 0 \\ (1 - \gamma z/L)^{-1/4} & \text{if } z/L < 0, \end{cases} \quad (6.5a)$$

$$(6.5b)$$

where β and γ are constants found empirically. It is seen that for $z/L = 0$ equation (6.4) reduces to the logarithmic wind profile law (equation (6.2)).

Adding and subtracting 1 from equation (6.4) (to follow the notation of Panofsky (1963)) and integrating from z_0 to z one obtains

$$u(z) = \frac{u_*}{k} \left[\ln \frac{z}{z_0} - \psi(z/L) \right], \quad (6.6)$$

where

$$\psi = \int_{z_0}^z (1 - \phi(z/L)) dz. \quad (6.7)$$

Panofsky (1963) and Paulson (1970a) gives the solution for ψ for the stable and unstable

case respectively as

$$\psi(z/L) = \begin{cases} -\beta \frac{z}{L} & \text{for } z/L \geq 0 \quad (6.8a) \\ \ln \left[\left(\frac{1+x}{2} \right)^2 \cdot \left(\frac{1+x^2}{2} \right) \right] - 2 \tan^{-1}(x) + \pi/2 & \text{for } z/L < 0, \quad (6.8b) \end{cases}$$

where $x = (1 - \gamma z/L)^{1/4}$. Equation (6.6) through equation (6.8a) and (6.8b) are also used in the work of Hintz et al. (2019a) (see chapter 8 for a resume).

6.3 Extrapolating wind measurements

Suppose one have an actual wind measurement and wants to use that to get information about the wind speed at higher elevations. Let $u(z_2)$ be the extrapolated wind measurement and let $u(z_1)$ be the actual wind measurement ($z_2 > z_1$), then one can solve for $u(z_2)$ by subtracting $u(z_1)$ from $u(z_2)$ using equation (6.6) to get

$$u(z_2) = \frac{u_*}{k} \left(\ln \frac{z_2}{z_1} + \Delta\psi \right) + u(z_1), \quad (6.9)$$

where

$$\Delta\psi = \begin{cases} \beta_1 \frac{z_2 - z_1}{L} & \text{for } z/L \geq 0 \quad (6.10a) \\ \ln \left[\frac{1+x_1^2}{1+x_2^2} \cdot \frac{(1+x_1)^2}{(1+x_2)^2} \right] - 2(\tan^{-1}(x_1) - \tan^{-1}(x_2)) & \text{for } z/L < 0, \quad (6.10b) \end{cases}$$

where $\Delta\psi = \psi_2 - \psi_1$. It is seen that z_0 is not explicitly a part of the system of equations. The surface roughness will, however, implicitly be a part of the wind measurement. L and u_* is the only unknowns of the system.

An alternative derivation can be made so that u_* is removed explicitly in which z_0 is used instead. This is done by using equation (6.6) to obtain an expression for both $u(z_1)$ and $u(z_2)$ and dividing them to get

$$u(z_2) = \frac{\ln z_2/z_0 - \psi_2(z_2/L)}{\ln z_1/z_0 - \psi_1(z_1/L)} u(z_1), \quad (6.11)$$

using equation (6.8a) and (6.8b) for the stable and unstable cases respectively. Equation (6.9) and (6.11) represents two different approaches to extrapolate a wind measurement in the surface layer, using u_* and z_0 respectively. Equation (6.11) are used to extrapo-

late wind measurements in Hintz et al. (2019a).

6.4 The roughness length

The surface roughness of the earth serves as a momentum sink for the atmospheric flow (Wieringa, 1993), and it is, therefore, important to determine this good if one wants to extract information about the wind speed at different heights or nearby an observation site. The surface roughness can be expressed in terms of a theoretical measure called the roughness length, z_0 . z_0 is traditionally determined via linear regression assuming neutral conditions (equation 6.6 with $\psi(z/L) = 0$) and hence z_0 is the height where the mean wind speed becomes zero, $u(z = z_0) = 0$. Bergeron and Abrahams (1992) notes that the regression should be done by linear regression of u on $\ln z$ because u is observed with much greater error than is $\ln z$. Using linear regression is a good way to determine z_0 ; however, in practical applications, it is nearly impossible to implement, especially for crowdsourced applications. Measurements in various heights are needed over a long period. What is done in most practical applications is that the surface roughness has been determined over different surfaces. The surface characteristics have then been classified, and similar surface characteristics in other areas are then associated with the same roughness length (Silva et al., 2007). In NWP, z_0 is often determined from a table of land use category with a given resolution and then upscaled to be an average of a given grid box. In WRF (see chapter 4) this is also the case, though, a summer and winter value are used to take into account the growth of crops and leaves on trees.

The roughness length can also be determined using the turbulence intensity following Counihan (1975) as

$$z_0^{ti} = z \cdot \exp(-u(z)/\sigma_u), \quad (6.12)$$

where $u(z)$ is the mean wind speed at height z and σ_u , is the standard deviation of $u(z)$. $\sigma_u/u(z)$ represents the horizontal turbulent intensity (Counihan, 1975). Equation (6.12) is derived from the logarithmic wind profile based on the assumption that $\sigma_u/u_* = 2.5$ (Arya, 1995).

Equation (6.12) is fundamental for the work presented in Hintz et al. (2019a) (see chapter 8), as it provides an apparent solution to determining the roughness length without expensive setups, which is ideal for crowdsourced data. The traditional method of determining z_0 would require tall masts and private people to maintain the equipment

themselves. To expect that to work is, of course, utopian. It should be noted for clarity that since equation (6.12) is derived from the logarithmic wind profile law (equation (6.2)), equation (6.12) too is only valid within the surface layer. This limits the validity of Handheld Wind Measurement (HWM)s to smooth surfaces, because these must be taken in the surface layer for equation (6.12) to be valid, and the altitude of these measurements are typically low.

Chapter 7

Traditional atmospheric observations and measurements

For an improved understanding of how crowdsourced data are different from traditional data sources, this chapter focus on traditional wind and surface pressure observations, allowing the reader to distinguish between traditional observations and crowdsourced observations. Here a measurement is defined as a unique single value coming from a sensor whereas an observation is defined as an average of measurements, following McNicholas and Mass (2018) and Hintz et al. (2019b). At the time of writing this thesis, the WMO guide to meteorological instruments and methods of observations (WMO, 2014) does not contain any information about crowdsourced data. However, that cannot be expected at the current stage of development of crowdsourced data in NWP.

7.1 WMO guidelines for wind observations

Traditional wind observations are those taken by Automated Weather Stations (AWS) certified by the WMO. The WMO guidelines (WMO, 2014) has published general recommendations for the design of a wind measuring system. First, wind measuring equipment should be placed at 10 m above ground surface level and on open ground. WMO (2014) defines open ground as a surface where obstacles are situated at a minimum distance equal to at least ten times their height. Furthermore, it is recommended that the wind is measured with a frequency of 4 Hz. Wind observations should be given as an average over 10 minutes and wind gust as an average over 3 s. For further details see WMO (2014).

7.2 Experimental setup of wind observations

To obtain reference observations of wind which satisfies the WMO recommendations, assets were needed to position the reference anemometer correctly. Therefore, an experimental setup was constructed in order to obtain reliable observations, to be compared with the crowdsourced data. This setup was used as a reference in Hintz et al. (2019a). It lies in the nature of crowdsourced data that it can be spontaneous in both space and time and so the setup was designed to be transportable. A mobile setup guarantees that it is always possible to quickly move to an area where observations are being generated by the crowd at a given time.

7.2.1 Extendable transportable mast

To position the anemometer in a height of 10 m as recommended by WMO (2014) a mast was needed. The mast is a telescopic mast from ClarkMasts (Clarkmasts, 2017). It consists of a tripod and can be extended to 8.8 m. An adaptor to the anemometer is attached on top of the main mast extending the mast to 9.4 m. The anemometer is attached on the adaptor and with the height of the anemometer mount base the altitude of the instrument reaches 10.0 m.

Figure 7.1 shows a picture of the mast when set up as it was used in Hintz et al. (2019a), however at a different location. On 1st of January 2019 the mast was set up at the location shown in figure 7.1. Here it will measure one year continuously with a frequency of 10 Hz. After one year the effect of growing crops on wind will be investigated by using the same methods described by Hintz et al. (2019a). The results of this study will be examined after this PhD project is finalised.

7.2.2 Ultrasonic anemometer

An ultrasonic anemometer (WindObserver 65) from Gill Instruments (Gill Instruments, 2017) was used due to its high precision and high sensitivity. The anemometer can output a sampling rate between 1 Hz and 10 Hz with a resolution of 0.01 m/s and an accuracy of 2 % at 12 m/s.

From the manufacturer, the anemometer is delivered with serial wiring. It is professional equipment and relatively expensive. Therefore, different methods of connecting and collecting data from the anemometer were investigated with a focus on a low cost.



Figure 7.1: Picture of the extendable transportable mast. On top the Gill Ultrasonic Anemometer is placed.

The anemometer outputs an RS422 serial signal with a baud rate of 4800 Bd, meaning that up to 4800 symbols can be sent per second. The anemometer requires a power source of 12 V DC and uses 40 mA. Two power sources were prepared. First, a 12 V DC battery with 48 Ampere-hours was retrieved, making it easy to power the anemometer if no other power sources are available nearby. Secondly, an inexpensive mobile-phone charger that converts 230 V AC to 12 V DC was modified to connect it to the anemometer so that the device can be powered from nearby power supplies if available.

Further to collect data, a datalogger able to read an RS422 signal was needed. A Sparkfun CAN-BUS electronic board (Sparkfun, 2017) was installed in a waterproof casing, together with a battery-pack, allowing the datalogger to operate for more than a week at a time. For periods of more than a week, another solution was required; one

solution investigated was to set up an old computer with an RS232 serial port at the mast and use an RS422-RS232 converter for data logging. RS232 was a standard port on old computers. Such a solution creates a new obstacle since the computer would then need power at the location of the mast. Instead, an RS422-USB converter was used together with a USB-Ethernet booster. The computer and the anemometer could then be connected via a 100 m long CAT6 ethernet cable, allowing the computer to be located in a dry environment with a power supply in many cases.

The ability to log data continuously via a computer opened up other uses of the data than was initially thought of. A web application presenting real-time wind measurements in 10 Hz was developed and made available for researchers at DMI (see figure 7.2). Not only was it then possible to store data continuously, but a visual presentation of the behaviour of the wind in high temporal resolution provided a possibility for all researchers to gain experience with the variations of the wind in different weather regimes.



Figure 7.2: Real-time wind measurements from the web-application during one minute 7th February 2018 at 20:32:33 UTC to 20:33:33 UTC. The white curve shows the measured wind speed with a frequency of 10 Hz, given on the left axis. The orange and blue line shows the mean over 3 s and 60 s respectively. The red dots show the wind direction given on the right axis. On the internal webpage the plot updates with a frequency of 10 Hz.

In practice, data was read with the computer using the PySerial (PySerial, 2018) framework for Python and uploaded to a Firebase database using the associated Python SDK (Google Firebase, 2017). Serverside, a web application fetching and visualising the data was made with Javascript and the Firebase Javascript SDK. Block 7.1 shows an

example of retrieving data from Firebase in real-time used to visualise wind measurements. The snapshot is triggered every time a new data point is detected.

```
1  var getObs = function () {
2    var dbtime = moment()
3    firebase.database().ref('/windobserver/'+dbtime.format('YYYY/MM/DD/HH
   /mm')).once('value').then(function(snapshot) {
4      fiobj = snapshot.val()
5      return fiobj
6    })
7  }
8  fiobj = getObs()
```

Block 7.1: Example of a function to retrieve data from Firebase using the JavaScript SDK. The 'snapshot' listen for changes and return new values to 'fiobj'.

7.3 WMO guidelines for atmospheric pressure observations

WMO (2014) suggests that a barometer should be able to measure the pressure in the range of 500 hPa to 1080 hPa with a resolution of 0.1 hPa and the uncertainty should be at most 0.1 hPa. Also, external effects such as wind, temperature and vibrations should be minimised. Further, barometers must be regularly calibrated against a reference with known errors.

At DMI, calibration is done against a Vaisala PTB330 Barometer (Vaisala, 2018). The Vaisala PTB330 uses three internal barometers independent from each other to improve reliability, which also improves long-term stability by preventing a bias drift. The reader is referred to WMO (2014, chapter 3) for an extensive discussion of observing atmospheric pressure.

7.4 Validation of observations

Validation of observations in the atmospheric sciences is a complex subject. In general, multiple quality checks are performed in DA systems (see chapter 9). One quality check can be a background check, meaning that observations are compared to a short-term forecast acting as a background. However, this approach can be somewhat contradictory as the observations are supposed to correct an erroneous forecast analysis, which is, on the other hand, used to check the quality of the observations. This is the reasoning for

using the least square approach in modern DA methods. Fast responding devices with no bias do, in theory, always measure ‘the truth’. A thermometer placed on a sunny side of a wall a sunny summer day will measure ‘the truth’ for that exact location if no bias is present.

Regarding representativeness, however, such a measurement is to a high degree useless. Observations need to be representative of the scales resolved by the model grid. That is one of the most significant issues to overcome for crowdsourced data. However, even traditional observations have similar problems. Observations that are not representative needs to be rejected by a DA system. An observational error can be divided into an instrument error and a representativeness error (Lahoz et al., 2010). Instrument error is an engineering matter associated with the functioning of the physical measurement itself, including deterioration of the instrument over time and manufacturing and materials details resulting in variations of readings under the same actual conditions. Representativeness error is introduced because an accurate observation does not represent the average value over the entire model grid box (for example, a surface observation showing the high pressure, strong winds, and cold air associated with a thunderstorm outflow).

Chapter 8

Estimation of wind speed and roughness length using smartphones: Method and quality assessment - Publication A

Fundamentally, the motivation of Hintz et al. (2019a) is to extract information from a series of wind measurements to provide an improved weather forecast to a user rather than only showing the observed wind speed. One original objective of this PhD project was to provide post-processed forecasts based upon the measured wind speeds, which was one of the reasons that the Vaavud NWP system (see chapter 4) was implemented.

An introductory examination of downscaling post-processed NWP winds to a much finer grid was started at the beginning of this PhD project. During this process, a collaboration between the department of wind energy at the Danish Technical University (DTU) and Vaavud started. DTU was able to provide a dataset of roughness lengths for different parts of Denmark, with a horizontal resolution of 200 m. This was used to downscale the NWP winds to the grid of that of the roughness lengths by extrapolating from the lowest model layer to 10 m using the logarithmic wind profile (see equation (6.6)). Figure 8.1 shows one example of this, where the effect of surface characteristics can be seen in the wind field (middle plot). For example, the forest on south-western Amager can be seen directly in the downscaled wind field.

Vaavud users consist mostly of kite- and windsurfers and other athletes within water sports, which is an advantage for future applications of such data, because these users tend to use the same locations and that the locations are close to water. Hence, observations will be taken at locations with a low roughness length in general, which is important as observations must be taken in the constant stress layer (see chapter 6) to be able to derive the roughness length from the observations. One stressed point in Hintz et al. (2019a) is that the method should not be used when $z_0 > 0.1$ m. Also, observations will tend to be focused on some specific locations which are often used for

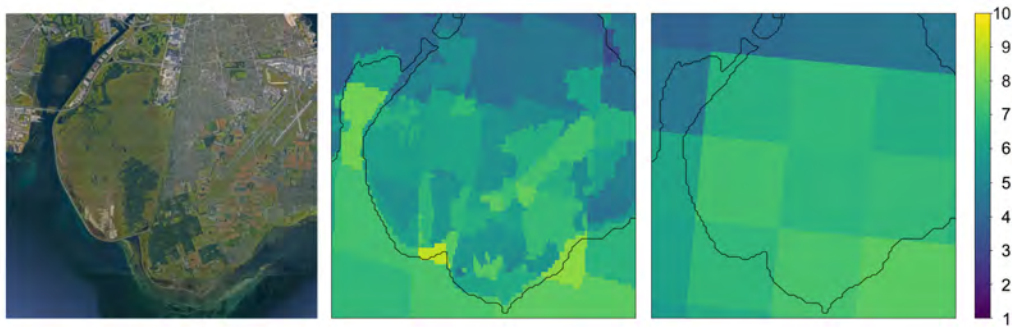


Figure 8.1: Left: Satellite imagery of most of the island of Amager. Centre: Adjusted wind speed forecast using roughness lengths from a wind atlas. Right: Direct diagnosed 10 m wind speed forecast from the Vaavud NWP system. The wind speed is given in m/s and is shown in the colorbar.

water sports. Therefore, there is a potential for creating spatial maps of the wind flow in ultra-high resolution (less than 100 m) at locations from where many observations are received. Hintz et al. (2019a) is the foundation of this objective. Another potential use case is for the agricultural sector. In situations with low wind speeds, it can be small differences in the wind speed, which determines if a field can be sprayed with pesticides or not. Making a HWM can help both the farmer and neighbouring farms in decision making, which ultimately can reduce the farmers costs and help the environment by using fewer pesticides.

Counihan (1975) provides a review and a comprehensive analysis of fully adiabatic boundary layers up to the year of 1975. Most fundamental micrometeorology research was established earlier. Also, Counihan (1975) describes a relationship between the longitudinal length scale of turbulence and surface roughness (z_0) (Counihan, 1975, Eq. 4) (see equation (6.12)) and gave one key remark that led to the studies presented in Hintz et al. (2019a), namely that: "*The scatter of these data (red. turbulent intensity) is considerably less than that of the mean velocity measurements: therefore this may be the best quantity to measure at a site and to use a reference quantity in determining other flow characteristics*". In Hintz et al. (2019a) it is investigated if the roughness length can be derived from measurements of the turbulent intensity from a handheld anemometer attached to a smartphone (see figure 2.1).

First, it is naturally of considerable interest to examine the quality of a HWM compared to traditional observations. There can be no doubt that in general HWMs does

have worse quality than traditional wind observations. However, there have been no studies examining the quality of HWMs from smartphones earlier, so it is vital to do so, before utilising HWMs. The quality of HWMs was examined by comparing to the reference wind measuring setup described in section 7.2 and nearby DMI SYNOP stations. It was found that in some cases even the raw HWM can be more representative than a nearby SYNOP station. However, it should be noted that the placement of the SYNOP stations is not ideal, which is also discussed in Hintz et al. (2019a).

Secondly, the roughness length is derived for three different locations using the HWMs, all with a relatively low roughness length ($z_0 < 0.1$ m). Extracting the roughness length from crowdsourced data could help to increase the quality of NWP forecasts as z_0 is more constant than the wind speed itself and are important when predicting surface winds and momentum fluxes (Nielsen, 2010).

Chapter 9

Data assimilation

DA is the process of determining the most likely state of a system given a set of data describing the system. In NWP the data is a set of observations from, e.g. satellites and meteorological ground-based stations and a background (first guess) of the system (the atmosphere). To start a forecast, a set of initial conditions at time $t = 0$ is needed which describes the initial state of the atmosphere. DA combines the available information to determine the most likely state of the atmosphere which acts as initial conditions for the NWP model. Given an analysis of the present state of the atmosphere, the model forecasts its progress.

DMI operates multiple operational models for both Denmark and Greenland. The main model is the HARMONIE based NEA-model. NEA has a spatial resolution of 2.5 km with 65 vertical layers with a grid mesh of 1200 longitudinal points and 1080 latitudinal points. HARMONIE uses 12 prognostic variables (Bénard et al., 2010; Bengtsson et al., 2017; Seity et al., 2011): momentum (u , v , w), temperature (T), water vapor (q_v), rain (q_r), snow (q_s), graupel (q_g), cloud droplets (q_c), ice crystal (q_i), Turbulent Kinetic Energy (TKE) and pressure (p). In total $1200 \times 1080 \times 65 \times 12 \approx 10^9$ variables are thus needed to be given an initial value.

There are not enough observations to overcome this demand and observations are not taken precisely at the location of the model grid points. Furthermore, observations are not evenly spread around the globe in either time or space. Most synoptic data come from land areas, especially from the mid-latitudes on the northern hemisphere (Zhang et al., 2018). During the last decades, however, satellite data and aircraft data have become increasingly available, so more data covering ocean areas and the southern hemisphere are available. On the other hand, as the resolution of NWP models continuously increases, more observations are needed. Vedel and Yang (1999) have described

the model's ability to advect information from data-rich areas to sparse data areas. It would be inadequate to represent the temperature field over the Atlantic Ocean by fitting a curve from temperature observations between North America and the UK. Therefore, a background field, \mathbf{x}_b , is used to estimate the initial conditions for the forecast. The background field is compared to observations, \mathbf{y} , and the DA algorithm changes the background field, given the observations, to improve the representation of the atmosphere to minimise the error of the analysis. This is the fundamental concept of DA.

9.1 Least squares method

Linear regression is used to find a relation between a predictor variable and a predictand. Using the regression parameters will give the linear relation between the two variables with the smallest error. This derivation is included to present the fundamental idea of DA, namely reducing the errors to find the most likely state of the atmosphere.

Traditionally the predictand is denoted as y and the predictor is denoted as x . One can think of y as the true state of a variable and x as observations of y with associated errors. Defining \hat{y} as the predicted state of the true state, y , one wants to find the regression parameters, a and b , in the linear relationship

$$\hat{y} = a + bx, \quad (9.1)$$

with the residuals, e_i , as

$$e_i = y_i - \hat{y}(x_i). \quad (9.2)$$

The regression equation is obtained by combining equation (9.1) and (9.2), such that,

$$y_i = a + bx_i + e_i, \quad (9.3)$$

which shows that the value of y is the sum of the predicted value and the residual. To find the regression parameters, the sum of the squared residuals is minimised by taking

the derivative and setting them equal to zero. This yield

$$\frac{\partial \sum_{i=1}^N e_i^2}{\partial a} = 0 \implies \sum_{i=1}^N y_i = na + b \sum_{i=1}^N x_i \quad (9.4)$$

$$\frac{\partial \sum_{i=1}^N e_i^2}{\partial b} = 0 \implies \sum_{i=1}^N x_i y_i = a \sum_{i=1}^N x_i + b \sum_{i=1}^N x_i^2. \quad (9.5)$$

Solving for a and b gives

$$a = \bar{y} - b\bar{x} \quad \text{and} \quad (9.6)$$

$$b = \frac{\sum_{i=1}^N ((x_i - \bar{x})(y_i - \bar{y}))}{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad (9.7)$$

where $\bar{(\cdot)}$ denotes the mean, $\bar{f} = 1/N \sum_{i=1}^N f_i$. In practice there exist more than one predictor so this example proves too simple for an NWP system. When multiple predictors exist one can use multiple linear regression instead, which is based on the same concepts.

9.2 The framework of data assimilation

As described in the introduction of this chapter the objective of DA is to combine a background state, \mathbf{x}_b , with observations, \mathbf{y} , to obtain the most likely analysis, \mathbf{x}_a . Here the fundamental framework of DA are described before introducing the Optimal Interpolation (OI) method and the 3D-Var method which is used at DMI and in Hintz et al. (2019b) (see chapter 14).

The framework for combining information from multiple sources luckily already exist in the form of Bayes theorem which states that (Taylor, 1997)

$$\underbrace{\text{pdf}(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} = \frac{\overbrace{\text{pdf}(\mathbf{y}|\mathbf{x})}^{\text{Likelihood}} \overbrace{\text{pdf}(\mathbf{x})}^{\text{Prior}}}{\text{pdf}(\mathbf{y})}, \quad (9.8)$$

where pdf is the Probability Density Function (PDF). The prior PDF is describing the prior information coming from the model, and the likelihood is describing the likeli-

hood of the observations, given the background. The \mathbf{x} that maximises the posterior, $\text{pdf}(\mathbf{x}|\mathbf{y})$, is the analysis. $\text{pdf}(\mathbf{y})$ acts in practice as a normalisation constant, so that is often not calculated explicit (Nichols, 2010).

To solve for the posterior, it is first assumed that the errors are Gaussian which is the fundamental assumption of most DA methods¹. The prior and the likelihood can then be described in terms of the mean and the covariance as

$$\text{pdf}(\mathbf{x}) = \frac{1}{2\pi^{N/2}|\mathbf{P}_b|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{P}_b^{-1}(\mathbf{x} - \mathbf{x}_b)\right) \text{ and} \quad (9.9)$$

$$\text{pdf}(\mathbf{y}|\mathbf{x}) = \frac{1}{2\pi^{N/2}|\mathbf{R}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - H(\mathbf{x}))\mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))\right). \quad (9.10)$$

respectively. Here \mathbf{P}_b is the background error covariance matrix and \mathbf{R} is the observation error covariance matrix. $H(\mathbf{x})$ is a forward model operator which interpolates and converts model variables to observation space, so $\mathbf{y} - H(\mathbf{x})$ is the difference between the observations and the models guess on the observations. The posterior can then be written as

$$\text{pdf}(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\underbrace{[(\mathbf{x} - \mathbf{x}_b)^T \mathbf{P}_b^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))]}_{\text{Cost-function, } J(\mathbf{x})}\right) \quad (9.11)$$

The term above the bracket is defined as the cost-function, $J(\mathbf{x})$. By minimising the cost-function, the posterior is maximised. The state, \mathbf{x} , that minimises the cost-function is defined as the analysis, \mathbf{x}_a . This minimisation problem can be solved in two different ways that lead to different approaches:

1) The mean and the covariance can be solved directly, by using the best linear unbiased estimate (BLUE) on the posterior (equation (9.11)) (Nichols, 2010). This approach is referred to as sequential DA and both the OI (section 9.3) and Kalman Filter (KF) methods are based on this.

2) Alternatively, one can solve for the mode of the posterior instead to find the \mathbf{x} that minimises the cost-function (maximises the posterior)

$$\mathbf{x}_a = \text{argmin}(J(\mathbf{x})). \quad (9.12)$$

¹Methods that do not assume that errors follow a Gaussian distribution includes Particle Filters, see, e.g. van Leeuwen (2009).

Methods based on this approach is referred to as variational methods and includes 3D-Var (see section 9.4) and 4-Dimensional Variational data assimilation (4D-Var). It is noted that if the posterior is truly Gaussian, the mean and the mode yield the same result (Wilks, 2011).

9.3 Optimal interpolation

OI is based on the sequential DA method. OI consist of vectors containing all the model variables, observations and errors and takes into account the errors of the observations and the errors of the background state.

Imagine that we want to know the true state, \mathbf{x}_t , of the atmosphere. We can estimate the true state by writing a linear combination of a background field, \mathbf{x}_b , and some observations, \mathbf{y}_o , as (Nichols, 2010)

$$\mathbf{x}_a = \mathbf{L}\mathbf{x}_b + \mathbf{W}\mathbf{y}_o, \quad (9.13)$$

where \mathbf{x}_a is the analysis, \mathbf{x}_b is the background field, typically a previous forecast, and \mathbf{y}_o is the observation vector. Equation (9.13) states that the best estimate (\mathbf{x}_a) of the true state is a linear combination of our background field (\mathbf{x}_b) and some observations (\mathbf{y}_o). The weight given to each term is determined by the weights \mathbf{L} and \mathbf{W} .

Equation (9.13) can be simplified by making a few assumptions. First of all it is noted that the errors of the background and the analysis can be written as

$$\begin{aligned} \mathbf{e}_b &= \mathbf{x}_b - \mathbf{x}_t \text{ and} \\ \mathbf{e}_a &= \mathbf{x}_a - \mathbf{x}_t, \end{aligned} \quad (9.14)$$

respectively, where \mathbf{x}_t is the (unknown) true state. Also, an observation process is defined as

$$\mathbf{y}_o = H(\mathbf{x}_t) + \mathbf{b}_o, \quad (9.15)$$

where H is the forward model operator as in section 9.2, and \mathbf{b}_o is the observational errors. Furthermore, it is assumed that the observation errors have zero mean and covariance \mathbf{R} ,

$$\mathbb{E}(\mathbf{b}_o) = 0, \quad (9.16)$$

$$\mathbb{E}(\mathbf{b}_o(\mathbf{b}_o^T)) = \mathbf{R}_k \delta_{kk'}, \quad (9.17)$$

where $\delta_{kk'}$ is the dirac-delta function, such that the observational errors are assumed to be uncorrelated. It is also assumed that the observation errors and model errors are uncorrelated,

$$\mathbb{E}(\mathbf{b}_t(\mathbf{b}_o)^T) = 0. \quad (9.18)$$

Subtracting \mathbf{x}_t from equation (9.13) and substituting with equation (9.14) and rearranging gives the error of the analysis

$$\mathbf{e}_a = \underbrace{\mathbf{L}\mathbf{e}_b}_{\text{Background error}} + \underbrace{\mathbf{W}\mathbf{b}_o}_{\text{Observational Error}} + \underbrace{(\mathbf{L} + \mathbf{W}H - \mathbf{I})\mathbf{x}_t}_{\text{Bias}}. \quad (9.19)$$

Assuming that the forecast error is unbiased ($\mathbb{E}(\mathbf{e}_b) = \mathbb{E}(\mathbf{x}_b - \mathbf{x}_t) = 0$), the condition $(\mathbf{L} + \mathbf{W}H - \mathbf{I})\mathbb{E}(\mathbf{x}_t) = 0$, must be met. In general however, $\mathbb{E}(\mathbf{x}_t) \neq 0$, so to obtain an unbiased analysis one can write the first weight in terms of the second weight as

$$\mathbf{L} = \mathbf{I} - \mathbf{W}H. \quad (9.20)$$

Substituting equation (9.20) into equation (9.13) and reducing gives the analysis equation used in OI,

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W} \underbrace{(\mathbf{y} - H(\mathbf{x}_b))}_{\text{Innovation}} = \mathbf{x}_b + \mathbf{W}\mathbf{d}, \quad (9.21)$$

where $\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b)$ is an innovation term. The weight \mathbf{L} has disappeared from our expression and so it is only necessary to determine one weight. The last term is called the innovation term and is determined by the difference between the observations, \mathbf{y}_o and the background \mathbf{x}_b and the weight \mathbf{W} .

H can both be linear and highly non-linear. In the simplest case, one can do a simple interpolation of temperatures, but it typically requires both an interpolation and conversion to the observed variable. The operator H can be linearized as

$$H(\mathbf{x} - \delta\mathbf{x}) = H(\mathbf{x}) + \mathbf{H}\delta\mathbf{x}, \quad (9.22)$$

where \mathbf{H} is a linear forward model operator that transforms between model space and observation space. Then the innovation vector, \mathbf{d} , can be written as

$$\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b) = \mathbf{y} - H(\mathbf{x}_t + (\mathbf{x}_b - \mathbf{x}_t)) \approx \mathbf{y} - H(\mathbf{x}_t) - \mathbf{H}(\mathbf{x}_b - \mathbf{x}_t). \quad (9.23)$$

Then the error of the analysis, \mathbf{e}_a , can be expanded by substituting with equation (9.23) and (9.15) to get

$$\mathbf{e}_a = \mathbf{x}_a - \mathbf{x}_t = (\mathbf{I} - \mathbf{WH})(\mathbf{x}_b - \mathbf{x}_t) + \mathbf{Wb}_0, \quad (9.24)$$

which is useful when determining the weight, \mathbf{W} , later (see section 9.3.1).

\mathbf{x}_a and \mathbf{x}_b are vectors of length n , the number of grid points times the number of variables. \mathbf{y} is a vector of length p , the number of observations and so the weight, \mathbf{W} is a matrix of dimension $n \times p$ and the forward operator, \mathbf{H} is a matrix of dimension $p \times n$.

9.3.1 The OI weight

The weight, \mathbf{W} , is chosen such that the variances of the analysis are minimised. To derive the weight consider first the error covariance of the analysis,

$$\mathbf{P}_a = \text{cov}[\mathbf{e}_a] = \begin{pmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n}^2 \\ \sigma_{2,1}^2 & \sigma_{2,2}^2 & \cdots & \sigma_{2,n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m,1}^2 & \sigma_{m,2}^2 & \cdots & \sigma_{m,n}^2 \end{pmatrix}. \quad (9.25)$$

Note that the variances are the trace of the covariance matrix, $\sigma_{1,1}^2, \sigma_{2,2}^2, \dots, \sigma_{n,n}^2$. Substituting with equation (9.24) and using that in general terms the covariance is defined as $C = \overline{(x - \bar{x})(x - \bar{x})^T}$, \mathbf{P}_a can be written as

$$\mathbf{P}_a = \overline{\mathbf{e}_a \mathbf{e}_a^T} = \overline{[(\mathbf{I} - \mathbf{WH})(\mathbf{x}_b - \mathbf{x}_t) + \mathbf{Wb}_0][(\mathbf{I} - \mathbf{WH})(\mathbf{x}_b - \mathbf{x}_t) + \mathbf{Wb}_0]^T}. \quad (9.26)$$

Expanding and assuming that the background and observations are uncorrelated, so that $\overline{\mathbf{b}_0 \mathbf{e}_b^T} = \overline{\mathbf{e}_b \mathbf{b}_0^T} = 0$, the analysis error covariance yields

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{WH})\mathbf{P}_b(\mathbf{I} - \mathbf{WH})^T + \mathbf{WRW}^T, \quad (9.27)$$

where $\mathbf{R} = \text{cov}(\mathbf{b}_o)$ and $\mathbf{P}_b = \text{cov}(\mathbf{x}_b - \mathbf{x}_t)$. To yield a simpler form, one can expand equation (9.27), using that $\mathbf{I} = \mathbf{I}^T$ and define $\mathbf{S} = \mathbf{HP}_b\mathbf{H}^T + \mathbf{R}$ to get

$$\mathbf{P}_a = \mathbf{P}_b - \mathbf{P}_b\mathbf{H}^T\mathbf{W}^T - \mathbf{WHP}_b + \mathbf{WSW}^T. \quad (9.28)$$

To obtain the weight that minimises the variances equation (9.27) one must take the derivative of the trace of \mathbf{P}_a and setting it equal to 0. Using that $\text{Tr}(\mathbf{A} + \mathbf{B}) = \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B})$ and using that covariance matrices are symmetric (e.g. $\mathbf{P}_a = \mathbf{P}_a^T$) the trace can be written as

$$\begin{aligned}\text{Tr}(\mathbf{P}_a) &= \text{Tr}(\mathbf{P}_b) - \text{Tr}(\mathbf{WHP}_b) - \text{Tr}(\mathbf{P}_b\mathbf{H}^T\mathbf{W}^T) + \text{Tr}(\mathbf{WSW}^T) \\ &= \text{Tr}(\mathbf{P}_b) + 2\text{Tr}(\mathbf{WHP}_b) + \text{Tr}(\mathbf{WSW}^T)\end{aligned}\quad (9.29)$$

Then, taking the derivative of \mathbf{P}_a with respect to \mathbf{W} yields

$$\frac{\partial \text{Tr}(\mathbf{P}_a)}{\partial \mathbf{W}} = -2\mathbf{P}_b\mathbf{H}^T + 2\mathbf{WS} \equiv 0. \quad (9.30)$$

Solving for \mathbf{W} yields the (optimal) weight,

$$\mathbf{W} = \mathbf{P}_b\mathbf{H}^T\mathbf{S}^{-1} = \frac{\mathbf{P}_b\mathbf{H}^T}{\mathbf{HP}_b\mathbf{H}^T + \mathbf{R}}. \quad (9.31)$$

Recall that \mathbf{R} is the observational error covariance matrix and \mathbf{P}_b is the background error covariance matrix. In these terms, it is easy to get an intuitive understanding of the weight. If the observational error is much larger than the model error the weight will go towards 0, making the innovation term small such that the observations are given a low weight. On the other hand, if the model error is much larger than the observational error, the weight will go towards 1, giving the innovation term a high weight when creating the analysis. In the Kalman Filter (section 9.7) the weight is called the Kalman Gain and is typically denoted \mathbf{K} , but in practice, it is the same weight.

9.3.2 The analysis covariance

An advantage of the OI method, is that one also can get an estimate of the uncertainty of the analysis through equation (9.28). It is however, easy to simplify the equation significantly. Equation (9.31) is multiplied with \mathbf{SW}^T to get

$$\mathbf{WSW}^T = \mathbf{P}_b\mathbf{H}^T\mathbf{W}^T. \quad (9.32)$$

Substituting this expression for \mathbf{WSW}^T into equation (9.28) and reducing one obtains a simpler equation for the error covariance matrix of the analysis,

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{WH})\mathbf{P}_b. \quad (9.33)$$

9.4 3 dimensional variational data assimilation

Recall the cost-function derived in section 9.2,

$$J(\mathbf{x}) = \frac{1}{2} \left[\underbrace{(\mathbf{x} - \mathbf{x}_b)^T \mathbf{P}_b^{-1} (\mathbf{x} - \mathbf{x}_b)}_{J_b} + \underbrace{(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}))}_{J_o} \right] \quad (9.34)$$

The first term on the right-hand side (J_b) measures the deviation between the state vector, \mathbf{x} , and the background \mathbf{x}_b , weighted by \mathbf{P}_b . The last term on the right-hand side (J_o) measures the deviation between the observations \mathbf{y} , and the prediction of the observations from the model $H(\mathbf{x})$, weighted by the observation error covariance matrix \mathbf{R} .

It is often assumed that observation errors made at different locations are uncorrelated so \mathbf{R} will be a diagonal matrix and thereby easier to handle. In some cases, this assumption can be valid, but not always. For example, a systematic error of a temperature measurement will not affect the measurements made by a station nearby. However, if a satellite measures a wrong radiance, it might be likely to do so in every measurement.

The model background error covariance matrix, \mathbf{P}_b , has a size of order $10^7 \times 10^7$, which is too huge to store on a modern computer. Per definition \mathbf{P}_b is defined as $\mathbf{P}_b = \text{cov}[\mathbf{x}_b - \mathbf{x}_t]$, where \mathbf{x}_t is the true state which is unknown. Therefore, to calculate \mathbf{P}_b a proxy is needed for \mathbf{x}_t . One proxy that is often used is *observation-minus-background* statistics. Here the NWP model is run for a long period and the average statistics compared to observations are calculated. With this approach \mathbf{P}_b is static in time.

The minimum of the cost-function is, as mentioned in section 9.2, obtained for $\mathbf{x} = \mathbf{x}_a$ which is the solution of

$$\nabla_x J(\mathbf{x}_a) = 0, \quad (9.35)$$

thus, the gradient of the cost-function must be solved. To solve for the gradient, some assumptions are made. First, assume that the analysis is close to the truth so only small increments are performed:

$$\mathbf{x} = \mathbf{x}_b + (\mathbf{x} - \mathbf{x}_b) = \mathbf{x}_b + \delta\mathbf{x}. \quad (9.36)$$

Futhermore, it requires many calculations to find $H(\mathbf{x})$ each time \mathbf{x} is changed, since in theory, the full NWP model must be run each time. Therefore second-order terms of H is neglected such that

$$H(\mathbf{x} + \delta\mathbf{x}) = H(\mathbf{x}) + \mathbf{H}\delta\mathbf{x}, \quad (9.37)$$

Using equation (9.36) and (9.37) the cost-function (equation (9.34)) can be written as

$$J(\mathbf{x}) = \frac{1}{2} \left[(\delta\mathbf{x})^T \mathbf{P}_b^{-1} \delta\mathbf{x} + [(\mathbf{y} - H(\mathbf{x}_b)) - \mathbf{H}(\mathbf{x} - \mathbf{x}_b)]^T \mathbf{R}^{-1} \right. \\ \left. (\mathbf{y} - H(\mathbf{x}_b)) - \mathbf{H}(\mathbf{x} - \mathbf{x}_b) \right]. \quad (9.38)$$

$H(\mathbf{x}_b)$ is known a priori, so the computational costs have been reduced with equation (9.38). To solve the gradient of the cost-function, it is beneficial if a quadratic equation can be obtained because the solution to these is known. Given a quadratic function $F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{d}^T \mathbf{x} + c$ the gradient are given by (Kalnay, 2003)

$$\nabla F(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{d}. \quad (9.39)$$

To obtain a quadratic function it is assumed that \mathbf{R} is symmetric so that $(\mathbf{H}\mathbf{R}^{-1})^T = \mathbf{R}^{-1}\mathbf{H}$. Using this and expanding equation (9.38) gives

$$J(\mathbf{x}) = \frac{1}{2} \left[(\delta\mathbf{x})^T \mathbf{P}_b^{-1} \delta\mathbf{x} + (\mathbf{x} - \mathbf{x}_b)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}_b) \right. \\ \left. - (\mathbf{y} - H(\mathbf{x}_b))^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}_b) - (\mathbf{x} - \mathbf{x}_b)^T \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}_b)) \right. \\ \left. + (\mathbf{y} - H(\mathbf{x}_b))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}_b)) \right]. \quad (9.40)$$

Then, the two first terms on the right hand side in equation (9.40) is combined and then reduced to obtain

$$\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2} \left[(\mathbf{x} - \mathbf{x}_b)^T [\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}] (\mathbf{x} - \mathbf{x}_b) \right. \\
&\quad - (\mathbf{y} - H(\mathbf{x}_b))^T \mathbf{R}^{-1} \mathbf{H} (\mathbf{x} - \mathbf{x}_b) \\
&\quad - (\mathbf{x} - \mathbf{x}_b)^T \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}_b)) \\
&\quad \left. + (\mathbf{y} - H(\mathbf{x}_b))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}_b)) \right] \\
&= \frac{1}{2} \left[(\mathbf{x} - \mathbf{x}_b)^T [\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}] (\mathbf{x} - \mathbf{x}_b) \right. \\
&\quad - 2(\mathbf{y} - H(\mathbf{x}_b))^T \mathbf{R}^{-1} \mathbf{H} (\mathbf{x} - \mathbf{x}_b) \\
&\quad \left. + (\mathbf{y} - H(\mathbf{x}_b))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}_b)) \right].
\end{aligned} \tag{9.41}$$

It is noted that the last term in equation (9.41) is independent of \mathbf{x} , and so can be discarded in the computation of the gradient of the cost-function.

Using equation (9.39) to compute the gradient and setting the expression equal to zero such that $\mathbf{x} = \mathbf{x}_a$ the analytic solution to cost-function is obtained as

$$\begin{aligned}
\mathbf{x}_a &= \mathbf{x}_b + [\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}_b)) \\
&= \mathbf{x}_b + \mathbf{Q} (\mathbf{y} - H(\mathbf{x}_b)),
\end{aligned} \tag{9.42}$$

where $\mathbf{Q} = [\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{R}^{-1}$. Unfortunately inverting such large matrices is practically impossible. In practice, a minimum is found by minimising the cost-function in an iterative process. In the iteration process \mathbf{x} is first chosen to be equal to \mathbf{x}_b , so the first term on the right-hand side in equation (9.34), J_b , does not contribute to $J(\mathbf{x})$, and the last term on the right-hand side, J_o , gives the deviation between the observations and the model expectation.

It can be shown that the weight, \mathbf{Q} , is mathematically equivalent to the weight obtained for OI (equation (9.31)), \mathbf{W} . This is shown in appendix E. The results from the two methods, however, are different because the methods of solutions differ. 3D-Var has several advantages compared to OI because the cost-function is minimised by using global minimisation algorithms. Therefore 3D-Var uses all available data simultaneously whereas in OI the solution is obtained "grid-point by grid-point". In practice, not all observations are allowed influence on a grid in OI. Observations too far away are ignored.

9.4.1 The two temperature problem

As an intuitive simple example of the cost-function, imagine a system described by two independent temperature observations T_1 and T_2 , each with an observational error, σ_1 and σ_2 . Assuming that the probability of obtaining a measurement T_1 , given the true temperature T , is a Gaussian distribution with an observational standard deviation σ_1 , one can write the probability to observe T_1 as

$$p(T_1|T) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(T_1-T)^2}{2\sigma_1^2}\right], \quad (9.43)$$

and likewise for T_2 . For independent measurements the probability for measuring T_1 and T_2 given T is their product so $L(T|T_1T_2) = L(T|T_1)L(T|T_2)$, where $L(x|y)$ denotes the likelihood of x given y (Wilks, 2011). The temperature, T , that maximises the likelihood is the most likely temperature, T_a , given an observation, T_1 , with error σ_1 .

As the logarithm is a monotonic function, one can take the logarithm and still obtain the same most likely temperature. Taking the logarithm and rearranging terms gives,

$$\ln[L(T|T_1T_2)] = \text{const} - \frac{1}{2} \left[\frac{(T - T_1)^2}{\sigma_1^2} + \frac{(T - T_2)^2}{\sigma_2^2} \right]. \quad (9.44)$$

The last term on the right-hand side is the cost function for this example. It is seen that by minimising the cost function, the most likely T is obtained.

9.4.2 Data assimilation time window and FGAT

An assimilation time window is defined around the analysis time, so if the analysis describes the atmosphere at time t_a^0 , the assimilation window could start at t_a^{-3} and end at t_a^{+3} , which is ± 3 hours giving an assimilation window of 6 hours. Observations taken in this time window are allowed to enter the assimilation system. However, assimilation time windows do not necessarily have to be symmetric around the assimilation time. In 3D-Var systems without First Guess at Appropriate Time (FGAT), it is assumed that all the observations are made at the analysis time, t_a^0 , which is valid only for synoptic observations. However, observations might be taken at t_a^{-2} and not t_a^0 , which leads to inaccuracies. Also, in 3D-Var systems only observations closest to t_a^0 is taken into account, such that only one observation per site is allowed throughout the assimilation time window.

3D-Var systems can be combined with FGAT to take in to account that observations

are taken at different times by splitting the assimilation window into N parts. By defining $x_b^{t_0}$ as the background field at the analysis time t_a^0 , and x_b^{t-1} as the background field at t_a^{-1} etc. Then observations are compared to the background field at the time which they are closest to. In practice this splits up $H(\mathbf{x}_b)$ into $H(\mathbf{x}_b^{t-1})$, $H(\mathbf{x}_b^{t_0})$, $H(\mathbf{x}_b^{t+1})$ etc., to be compared to observations at \mathbf{y}_{t-1} , \mathbf{y}_{t_0} , \mathbf{y}_{t+1} , which is equivalent to write a sum over the distance between observations, \mathbf{y} at time i and the model expectation $H(\mathbf{x}_b)$ at time i ,

$$\sum_{i=0}^N H(\mathbf{x}_{bi}) - \mathbf{y}_i.$$

This is an advantage as it is not assumed that all observations are valid at the analysis time t_a^0 . With this method, the departure of the background field from the observations are smaller, so the last term in equation (9.34) gets smaller, leading to a smaller value of the cost function, $J(\mathbf{x})$.

FGAT was used in the previous operational High Resolution Limited Area Model (HIRLAM) model at DMI, but it is not used in the current operational HARMONIE model.

9.5 4 dimensional variational data assimilation

While examining results from numerical experiments with 3D-Var in this study, multiple disadvantages of using 3D-Var to assimilate crowdsourced data was identified. First, as the temporal resolution of crowdsourced data is very high, one does not take full advantage of this by using 3D-Var as the time evolution is not taken into account. Also, the structure functions (\mathbf{P}_b) smoothens the information from the observations much. While these algorithms are necessary for keeping the model stable, it does, unfortunately, make the current 3D-Var system in HARMONIE less attractive for assimilating crowdsourced data (Hintz et al., 2019b).

3D-Var can be extended to take into account the development of the model field with time. This results in the 4D-Var assimilation technique. As in 3D-Var, the concept is to minimise a cost function. In 4D-Var observations are allowed to be taken at any time in the assimilation window, allowing asynoptic observations. Practically this is done by redefine the part of the cost-function which is containing the observations, the last term

in (9.34). So in 4D-Var the cost-function becomes (Talagrand, 2010)

$$\begin{aligned}
 J(\mathbf{x}) = & \frac{1}{2}[(\mathbf{x}_0 - \mathbf{x}_{b_0})^T \mathbf{P}_b^{-1}(\mathbf{x}_0 - \mathbf{x}_{b_0})] \\
 & + \frac{1}{2} \sum_{i=0}^N [\mathbf{y}_i - H(\mathbf{x}_i)]^T \mathbf{R}^{-1}[\mathbf{y}_i - H(\mathbf{x}_i)].
 \end{aligned}
 \tag{9.45}$$

Note that the first term on the right-hand side (J_b) is the same as in 3D-Var. The model expectation is found by running the model back and forth in time so in 4D-Var the forward operator, H , does not only spatially interpolates to observations and converts model variables to the observed quantities but it also propagates the model fields defined at t_0 to the time of the observations. This makes the 4D-Var method computationally very expensive.

4D-Var is not yet operational in HARMONIE, but it is an active development branch². As a continued work of Hintz et al. (2019b) (see chapter 14) it will be exciting to examine the data collected in this PhD project with 4D-Var in HARMONIE when ready.

One key finding of Hintz et al. (2019b) is that the accuracy of the vertical position of a SPO needs substantial improvement. However, this issue can be eliminated by assimilating pressure tendencies using 4D-Var rather than assimilating absolute pressure using 3D-Var, if one can assure that the smartphone has not moved between multiple observations. Furthermore, one can take better advantage of the high temporal resolution of crowdsourced data using 4D-Var, because multiple observations from the same source can be utilised. As an example, Netatmo PWS report observations every ten minutes through the Netatmo API (Netatmo SAS, 2019a). Here, dependent on the length of the assimilation windows, 4D-Var would be able to use many more observations than 3D-Var would.

Moreover, because the computation of 4D-Var is expensive, in practice the minimisation is often run with a more coarse spatial resolution than the linked NWP model meaning that the structure functions, \mathbf{P}_b , would also be based on a coarser resolution. One can imagine that information from individual observations would then be spread over larger distances; hence the high-spatial resolution of crowdsourced data would not be utilised to its full potential.

²Personal communication with Dr Xiaohua Yang, DMI.

To minimise the cost-function in 4D-Var, one has to consider the propagation of the model in time. The gradient is found by considering a perturbation to the initial state to linearize the model about the non-linear model trajectory. This ultimately introduces two new concepts which prove efficient when solving the gradient of the 4D-Var cost-function, namely the tangent linear model and the adjoint model (Talagrand, 2010). The derivation of the tangent linear model and the adjoint model are out of scope of the discussion of 4D-Var here, but the interested reader is referred to Kalnay (2003); Nichols (2010) and Talagrand (2010) for further details. It should be noted, though, that the linearization of the model around the initial state, \mathbf{x}_0 , is called the tangent linear model. The tangent linear model propagates an initial perturbation at time t_i to t_{i+1} . The transpose of the tangent linear model is called the adjoint model, which advances a perturbation backwards in time from t_N to t_0 . The introduction of these makes the back-and-forth computation less expensive, but it does require some effort to derive the tangent linear model and the adjoint model in practice.

9.6 Nudging

Another possible method to maximise the potential of crowdsourced data in NWP is nudging. No studies have yet looked at this subject in detail. Nudging is a continuous, four-dimensional, empirical DA technique used within the first period of a NWP forecast. At each timestep, observations are used to determine a nudging factor based on the differences between the model and the observations. When the differences are found, increments or decrements are added to a model field of interest, nudging it gradually closer to the observed weather (Olsen et al., 2015). This is done by adding forcing terms to prognostic equations in the NWP model with the hope that the model can then better describe the observed conditions. The forcing terms depend on whether the conditions are measured or derived. The forcing term is included at the time of the observation, and to assure that the forcing on the model dynamics is not too sudden a relaxation time is included, which is a weighting factor for the time dimension.

An example of how nudging can be applied in a prognostic equation can be shown with the zonal velocity forecast equation in isobaric coordinates (Kalnay, 2003)

$$\frac{\partial u}{\partial t} = -\mathbf{v} \cdot \nabla u + fv - \frac{\partial \phi}{\partial x} + \underbrace{\frac{u_{obs} - u}{\tau_u}}_{\text{Nudging term}}, \quad (9.46)$$

where the last term is the additional forcing term performing the nudging. In this example the nudging is weighted by the relaxation time, τ_u . The value of the relaxation time determines the impact of the forcing term which is found empirically.

Only prognostic variables can be nudged directly, though other parameters can be nudged through a series of model variables giving the desired result, such as is done with radar data in the DMI nowcasting system (Olsen et al., 2015) and the latent heat nudging method (Jones and Macpherson, 1997). One issue with the nudging technique is that there is no built-in quality check on the observations, which can allow observations with significant errors to be nudged into the model. Another possible issue would be that the process gives the wanted effect for one parameter but makes another one worse, e.g., improving the representation of precipitation, but worsening the wind field.

Nudging is an attractive method for use with crowdsourced data because observations can be included quickly, and the impacts are easy to measure. In the case of SPOs, the surface pressure can be nudged directly, or one can use nudging to include pressure tendencies before using, for example, 4D-Var. However, nudging might only be a suitable choice for nowcasting models which runs in rapid update cycles, because the effect of the nudging tends to decline with forecasting time (Olsen et al., 2015).

9.7 Kalman filtering

The KF (Kalman, 1960) is another compelling technique from the perspective of utilising crowdsourced data. A KF has the advantage over variational methods that it is, in general, easier to implement. Also, the analysis can be updated sequentially as observations come in. McNicholas and Mass (2018) used a Local Ensemble Transform Kalman Filter (LETKF) to obtain an analysis based on SPOs and was the first to assimilate SPOs using Kalman based methods, and showed promising results. Naturally, Kalman based methods should be considered for crowdsourced data based on the study of McNicholas and Mass (2018).

The KF has proven to be suitable for various areas, such as auto-piloting (Fossen and Perez, 2009), computer graphics (Kautz and Eskofier, 2015) and DA in NWP (Lorenz, 2003). The filter can run in near real-time as it is computationally efficient for many applications. This makes the filter suitable as a tool to include crowdsourced data in NWP. The filter works in methodologies very much as the OI method (section 9.3), but as it is a recursive filter, the notation is slightly different than that from OI. In KF the

analysis equation from OI (equation (9.21)) is called the update equation and is written as

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^f), \quad (9.47)$$

where k refers to an iteration in time and the weight, the Kalman Gain (\mathbf{K}_k) is updated for each iteration. \mathbf{K}_k has the same form as the OI weight (equation 9.31) and the derivation is identical. Here \mathbf{x}^f is used instead of \mathbf{x}_b in OI to refer to forecast and distinguish between the two.

The analysis covariance matrix is also similar to the one in OI (equation 9.33), but in the KF, \mathbf{P}_a is updated in each iteration by updating the Kalman gain in each iteration as,

$$\mathbf{K}_k = \frac{\mathbf{P}_k^f \mathbf{H}_k^T}{\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}}, \quad (9.48)$$

where it should also be noted that only the notation is different from the OI weight. Then the analysis covariance matrix is updated as

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f, \quad (9.49)$$

to get the uncertainty of the current analysis state vector, \mathbf{x}_a .

9.7.1 The Kalman algorithm

Because of the recursive nature of the KF, it would be natural to apply the filter to update a forecast as new observations enter the forecasting system. To evaluate the KF a simple theoretical implementation will be considered briefly.

The final algorithm for the KF consists of an update step and a prediction step. First, the analysis, \mathbf{x}_k^a , is updated in each iteration with the observations available at that time, \mathbf{y}_k , and after updating the analysis, the state in the next iteration is predicted, \mathbf{x}_k^f , based on the new analysis and covariance, \mathbf{P}_a . The algorithm can be summarised as follows:

The prediction step

$$\begin{aligned}\mathbf{x}_k^f &= \mathbf{M}\mathbf{x}_{k-1}^f + \mathbf{w} \\ \mathbf{P}_k^f &= \mathbf{M}\mathbf{P}_{k-1}^f\mathbf{M}^T + \mathbf{q}\end{aligned}$$

The update step

$$\begin{aligned}\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^f) \\ \mathbf{K}_k &= \mathbf{P}_k^f\mathbf{H}_k^T\mathbf{S}^{-1} = \frac{\mathbf{P}_k^f\mathbf{H}_k^T}{\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}} \\ \mathbf{P}_a &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f\end{aligned}$$

The prediction step

In the first iteration, the previous forecast, \mathbf{x}_{k-1}^f , is set equal to the background state from the previous forecast, \mathbf{x}_b . The prediction step assumes that the system can be modelled by a linear stochastic state transition equation as

$$\mathbf{x}_k^f = \mathbf{M}\mathbf{x}_{k-1}^f + \mathbf{w}, \quad (9.50)$$

where \mathbf{w} is a vector of random noise and \mathbf{M} is a transition matrix that progresses the state from the previous time to the current time step, so that \mathbf{M} is the tangent linear model and \mathbf{M}^T is the adjoint model (see section 9.5). For a remarkably simple model or if the system is entirely random, \mathbf{M} can be set equal to one, i.e. the predicted state is the same as the current state. That would, of course, be unreasonable for NWP (and most other) applications. In the prediction of the forecast error covariance, \mathbf{P}_k^f , \mathbf{q} is the forecast error covariance.

The update step

The analysis, \mathbf{x}_k^a , is updated with the predicted forecast from the prediction step, \mathbf{x}_k^f , and the new observations, \mathbf{y}_k weighted by the Kalman Gain, \mathbf{K}_k . The Kalman Gain itself is also updated to update the analysis error covariance, \mathbf{P}_a .

One important difference between OI and the KF is that the error covariance matrix is static in the case of OI, but dynamic in the case of the KF where it is evolved in time. Furthermore, variations of the KF exists. In NWP much focus are on Ensemble Kalman Filter (EnKF) (Houtekamer et al., 2005; Lorenc, 2003) and LETKF (Hunt et al., 2007), in which the filter holds multiple ensemble members. With these methods, covariance inflation is needed to ensure the spread of ensemble members (Houtekamer and Mitchell, 1998). Also, localisation is often used, which adjusts the error covariance to suppress the influence of observations far away to remove spurious long-distance correlations (Greybush et al., 2011). Also, this makes parallel implementation easier,

resulting in faster computations. KFs does have some interesting properties related to crowdsourced data. At the beginning of this PhD study, the process of implementing and testing a KF for crowdsourced data was started at Vaavud. Initial and preliminary results for updating a wind forecast from the Vaavud NWP system (see chapter 4) using observations from users of wind was obtained. Unfortunately, it was not possible to complete those studies due to the unexpected bankruptcy of Vaavud (see chapter 1). For further details on the KF, the reader is referred to the references mentioned above in the current section.

Chapter 10

Data assimilation in the DMI HARMONIE NWP system

Chapter 9 considered the theory of DA in general. This chapter is devoted to the DA system of the HARMONIE NWP system in a practical sense, especially with a focus on the parts that were restructured during this PhD project to create an analysis using crowdsourced data. First, a fundamental introduction to the DMI HARMONIE model is given.

10.1 The HARMONIE NWP model

The HARMONIE model is a non-hydrostatic model that is suited for predicting phenomena on mesoscale so that the assumption of hydrostatic balance is not required. HARMONIE is a result of a cooperation between two research groups, the ALADIN (Aire Limitée Adaptation dynamique Développement InterNational) consortium led by Météo-France and the HIRLAM-B consortium, where DMI is a member of the latter.

HARMONIE is at DMI runned every 3 hours (see figure 10.1) with a grid-spacing of 2.5 km and 65 vertical layers. Data assimilation is done every cycle using 3D-Var (see chapter 9). The configuration of HARMONIE can be done via a set of namelists and a scripting system.

At DMI different model domain configurations exists. One model domain covers only Denmark but has a horizontal resolution of 750 m¹. Other domains cover parts of Greenland. The main operational NWP model at DMI, named NEA, covers most of Northern Europe and parts of the Atlantic and has a horizontal resolution of 2.5 km.

¹This model, and a 500 m resolution model is currently not operational but parts of research projects. The sub-km models have been set up by Dr Xiaohua Yang (DMI).

The model domain used here is the DKA domain (see figure 10.2), also with a resolution of 2.5 km.

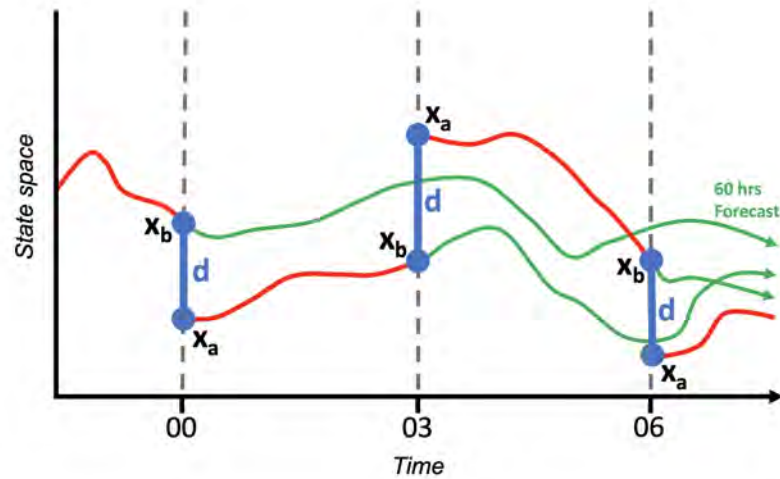


Figure 10.1: Forecast cycle of the DMI HARMONIE model. A new forecast is initiated every 3 hours from a new analysis \mathbf{x}_a which is produced by the previous forecast, \mathbf{x}_b and an innovation term \mathbf{d} (see chapter 9).

The dynamics and physics parameterisations have not been modified in this project. Particularly the DA system of HARMONIE has been changed (see chapter 10). The reader is referred to Seity et al. (2011) and Driesenaar (2009) for an extensive model description, but it should be noted that HARMONIE is a semi-Lagrangian model which allows larger time-stepping than Euler based models such as the WRF model.

SPOs was mainly collected from Denmark, and the focus has been on short time-scales, mainly the first six forecast hours. Therefore the model domain, DKA, shown in figure 10.2, was chosen to reduce the computational costs compared to a larger domain with more grid points. It is believed that the DKA model area has a large enough area outside Denmark to give boundary conditions a minimum impact on short time-scales. The DMI default settings for the model was used (Yang et al., 2017), with modifications to the DA system.

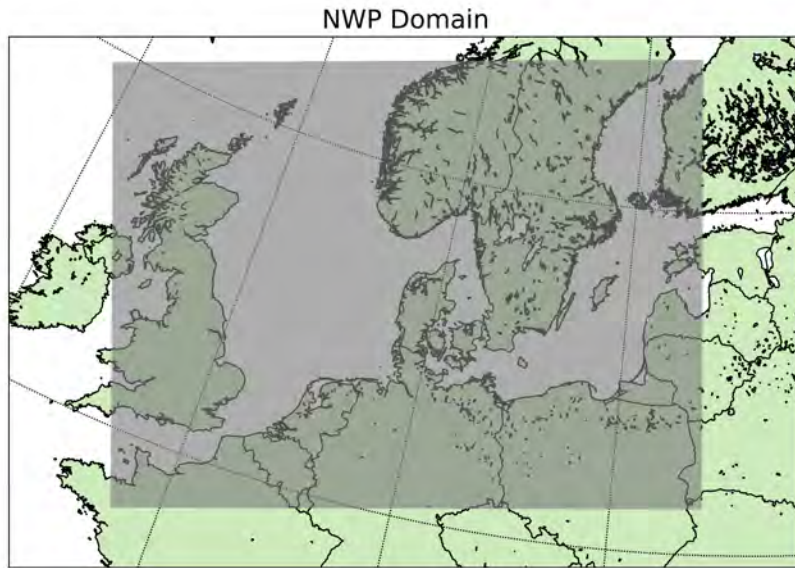


Figure 10.2: Model domain of the DKA model shown by the grey shaded area. This figure is also included as supplemental material to Hintz et al. (2019b).

10.2 Observation pre-processing in HARMONIE

Before any DA methods are performed, observations must be pre-processed, to remove observations of poor quality and to align the observation formatting. Pre-processing observations refers to a chain of methods from data collection to detection of outliers, to generate the observational state, \mathbf{y} , used by the 3D-Var algorithm in HARMONIE.

10.2.1 Local pre-processing

Observations come from multiple systems from global to local systems; thus, no pre-processing observation system is identical, but observations do enter the same kind of methods. In general, observations come from shared networks such as

- Global Telecommunication System (GTS)
Received in American Standard Code for Information Interchange (ASCII) or Binary Universal Form for the Representation of meteorological data (BUFR) format.
- Satellite Receiving stations

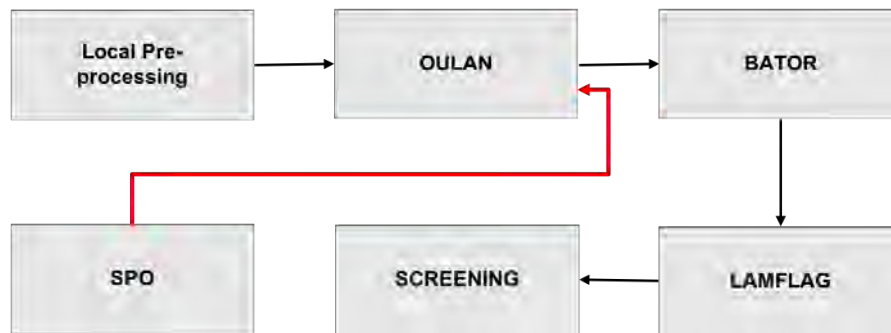


Figure 10.3: Outline of the observation pre-processing chain in HARMONIE. SPOs enters the system after the Oulan routine.

- File Transfer Protocol (FTP) Servers
- Transmission Control Protocol (TCP) lines (such as local ethernet, phone wires etc.)

The GTS is a generalized meteorological platform for sharing observations globally controlled by the WMO. The local pre-processing in HARMONIE receives data from all the individual sources and convert the data into an intermediate ASCII file which Oulan takes as an input.

10.2.2 Oulan

Oulan reads the data from the local pre-processing system and outputs an intermediate ASCII file named OBSOUL, which is input to the next module, Bator. In Oulan, data types are identified and read separately and controlled via namelists. In such manner it is easy to exclude different observation types in numerical experiments, such as excluding only aircraft data or surface synoptic observations (SYNOP) data.

OBSOUL is written in a specific format. The structure of OBSOUL is only covered briefly here and documents how SPOs was incorporated into the HARMONIE system. OBSOUL is divided into three sections. The *article*, containing date and hour, the *header*, which holds the length of the observation article, observation type, observation characteristics, latitude, longitude, station ID, date and hour of the observation, the altitude of the station, the number of bodies included, a quality flag and a site-dependent integer. Finally, the *body* (or *bodies* if more than one parameter is included) is defined, which holds the observation itself, such as the type of parameter and the observed value.

Block 10.1 shows a minimal working example of an OBSOUL file with a single observation of surface pressure.

```

1 20180510 09 # Article
2 17 1 10014011 54.97655 12.27767 '416789 ' 20180510 083005
3 34.5000000 1 1111 100000 # Header
4 1 -100721.6626 0.1699999976E+39 0.0000000000E+00 2064 # Body

```

Block 10.1: Minimal example of a working OBSOUL file with a single observation. Comments are written in red. See text for details.

The first number in the header of block 10.1 specifies the number of entries in the header and the body, including itself. The latitude, longitude and altitude of the station are set in the header together with the station ID in a string of length nine characters. The first number in the body is the parameter number for surface pressure, and the second value is the observed pressure. Surface pressure is always given as a negative value in OBSOUL. If block 10.1 was the output of Oulan only one observation would enter the DA system.

In this way SPOs could be included in the HARMONIE system. After local processing in SMAPS (see section 2.2 and Hintz et al. (2019b)) each SPO was converted to OBSOUL format and sent to HARMONIE via Oulan (see figure 10.3).

10.2.3 Bator

The input to Bator is the OBSOUL observation file generated by Oulan. The output of Bator is an Observation Database (ODB) which is input to the variational analysis scheme. While creating the ODB, Bator takes into account a namelist, LISTE_-NOIRE_DIAP², which contains information about blacklisted observations.

The blacklist was used in this project to force some observations to be removed when performing numerical experiments. Block 10.2 shows an example of three observations that are blacklisted. The first (SHIP) is blacklisted in the operational setup, while the two last (SYNOP) observations are only blacklisted for the sake of this example.

²'A black list' in French

```

1 # OBSTYPE CODETYPE VARNO STATIONID DATE
2 1 SHIP          24   58 LAJS6      01022007
3 1 SYNOP        14    1 06079     01032018 # Anholt
4 1 SYNOP        14    1 06093     01032018 # Vester Vedsted

```

Block 10.2: Minimal working example of a blacklisting file. Comments are written in red. See text for details.

The first two columns in block 10.2 is the observation type, the third column is the code type, and the fourth column is the variable number. For example ‘1 *Synop 14 1*’ is an automatic land SYNOP station reporting surface pressure. The last two columns are the station ID and the date at which the blacklisting starts. For further information about the ODB see ECMWF ODB (2018). Finally, Bator specifies the observation errors for the conventional observations before writing to the ODB.

10.2.4 Lamflag

Lamflag is an additional namelist for controlling the usage of observations. It can be used to only use specific types of observations or only observations from a specific region. Lamflag is not used in the setup at DMI.

10.2.5 Screening

The observation screening is the last frontier before observations enter the DA schemes. The observation screening is divided into six parts with different sub-programs as listed below.

1. Preliminary Check
 - 1.1 Missing elements of observation reports.
 - 1.2 Check coordinate system of the reports.
 - 1.3 Scanning for blacklisting.
2. Background quality control
 - 2.1 Background variance are estimated.
 - 2.2 If an observation deviates too much it is flagged.
3. Vertical consistency
 - 3.1 Duplicate levels from multi-level reports are removed.

4. Removal of duplicate reports
5. Redundancy check
 - 5.1 For land SYNOP the report from one station closest to the screening time window is kept.
 - 5.2 For ship SYNOP, reports are rejected if moving platforms are within a circle of one-degree radius.
6. Data thinning
 - 6.1 Horizontal thinning is performed for aircraft and satellite reports.
 - 6.2 Vertical thinning is performed for aircraft and Atmospheric Motion Vector (AMV) observations.
 - 6.3 A pre-described minimum horizontal distance between reports from the aforementioned observation types are enforced.

It is important to note that the data assimilation schemes assume that there is no correlation horizontally between observation errors, so the observational error covariance matrix, \mathbf{R} , is diagonal. Therefore, data thinning is necessary for most operational systems. In the current DA scheme in HARMONIE there is no thinning of surface data. Hence, it is necessary to do a manual thinning before allowing surface observations into the system. Potentially that can be a significant issue when assimilating crowdsourced data. Another possibility is to do 'superobbing', in which observations close in time and space are combined to create a single super-observation with a smaller error than the individual observations.

When screening is complete, the system continues to start the minimisation of the 3D-Var cost-function (see section 9.4) to obtain the analysis.

10.3 Structure functions in practice

Recall the background error covariance, \mathbf{P}_b , introduced in chapter 9. \mathbf{P}_b defines how errors spread throughout the model space, e.g. how do errors in pressure affects the wind? Ideally, \mathbf{P}_b should be dynamic, to take into account different weather regimes that distribute errors differently. For example, in very stable boundary layers, errors are more likely not to be advected upwards. However, as mentioned in chapter 9, \mathbf{P}_b is too large to handle directly in practice, so approximations are needed. Figure 10.4

visualise the effect of \mathbf{P}_b , shown by the difference in two analyses of surface pressure valid at the same time, in which the only difference is a single surface pressure observation included from Copenhagen, Denmark. It is seen how information from the single observation spread relatively far from the observation site to northern Germany and opposite changes over the UK. The spread is determined by \mathbf{P}_b .

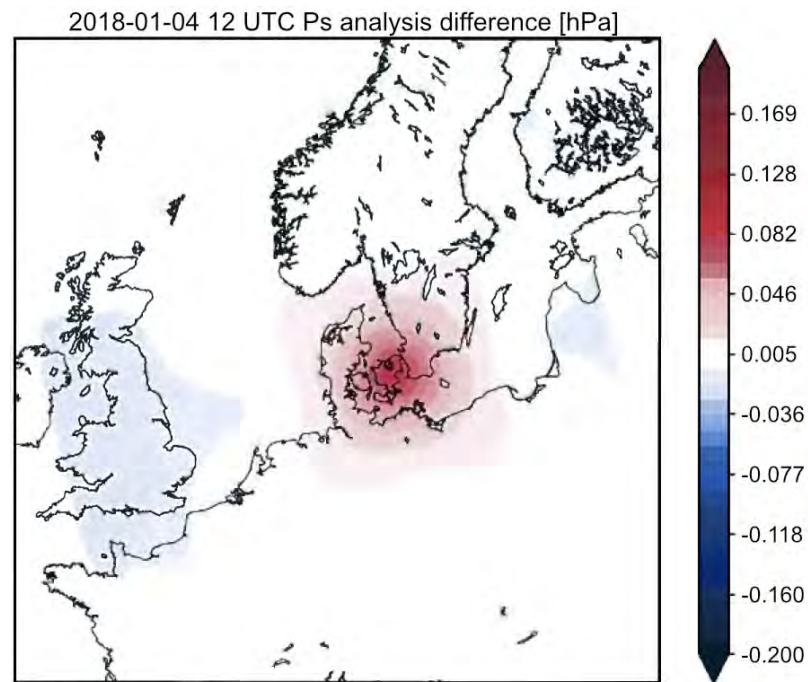


Figure 10.4: Single Observation experiment: The difference between two analyses valid at 2018-01-04 12 UTC is shown. In one analysis, a single additional pressure observation was included. Colorbar is in units of hPa.

To determine \mathbf{P}_b , one option is to run the NWP model for an extended period to obtain samples of forecasts to estimate the error correlations in space and between variables. As described, any change in one variable may change another if the variables are correlated. Such cross-covariances can be estimated by linear regression (Descombes et al., 2015). Since \mathbf{P}_b spreads errors, it acts as a smoother on the model field (Montmerle et al., 2010). The smoothing helps to ensure that the model field is in balance. However, for crowdsourced data, where the observation density is usually very high, too much smoothing may result in loss of high-resolution information. This should be investigated in more detail in future studies. Another possibility for crowdsourced data is to do ‘superobbing’ which combines multiple observations close in space and time to

a single super-observation with a smaller error than the individual observations.

Another method to determine, \mathbf{P}_b , is to generate a set of ensemble members and perform the statistics based on these. This approach is used in the HARMONIE community. Four ensembles are run twice daily (00 UTC and 12 UTC) for a longer period, usually one winter month and one summer month to obtain information both about seasonal variations and diurnal variations³. Structure functions have not been generated as part of this study, so practical details are omitted.

³Information from the internal Harmonie System Documentation at Hirlam.org

Chapter 11

Use of big data and crowdsourced data in NWP

‘Big Data’ and ‘Crowdsourced Data’ are in some manner closely linked; however, there are some important differences between the two terms that will be elaborated in this chapter. Roughly, it can be said that crowdsourced data can be a part of big data, but not always. Crowdsourced data for use in NWP is not restricted to SPOs and PWS. Other data sources also exist. An overview of other data sources is also given in this chapter.

Crowdsourced data can indeed be a source of big datasets. However, crowdsourced data also includes data from applications which do not produce much data. Muller et al. (2015) categorised crowdsourced data into two sub-categories, namely ‘inanimate’ and ‘animate’ crowdsourced data. Inanimate crowdsourcing includes those applications that do not require human interaction, one example being data from PWS (Fieldsense A/S, 2019; Lonobox, 2019; Netatmo SAS, 2018). Animate crowdsourcing, on the other hand, does require some human interaction. For example, the mPing app (NSSL, 2014) collects user reports of the weather and is one of the first of its kind. Later, FMI (FMI, 2018) developed a mobile application in which users can report current weather conditions for use by operational meteorologists. A downside of animate crowdsourced data is that the data quantity is often much less than what can be obtained via inanimate methods. One may argue that animate methods lead to better quality, but this question is still open for discussion.

‘Big Data’ is a term with no clear definition. However, it is clear that the term is not static but changes over time. For example, large datasets ten years ago might be considered relatively small today. What does the term cover, how it is used, and how it is not used? Big data should not necessarily only be defined based on the size of the dataset. Schonberger and Cukier (2013) argues that there are three main characteristics

of big data. First, the ability to handle and analyse the entire dataset is crucial instead of only using sub-samples of the dataset. Secondly, the messy structure of big data makes it easier to look for correlation instead of causality; hence statistically models are often preferred over physical models when it comes to big data as it is easier to handle big data with such. Finally, Schonberger and Cukier (2013) argues the necessity to accept the inexact and messy structure of the data. One has to accept an often large variability in the data. This comes as a somewhat natural consequence of the nature of Internet of Things (IoT).

IoT are an umbrella term covering sensors connected to the internet which communicates in real-time or near real-time. A few examples could be a smartwatch making an emergency call if it detects a sudden fall in heartbeat rates, or a households heating system gearing down when it detects that the main door gets locked. There are numerous of such examples; however, it is out of scope of this work to give a brainstorm on such things. Instead, focus is on current techniques and developing ideas within the atmospheric sciences. The IoT era started around 2008, where the number of devices connected to the internet exceeded the number of humans alive (Evans, 2011). Evans (2011) also predicts that in 2050 there will be 50 billion IoT devices. According to a blog post from the tech company Hewlett Packard (Patric, 2014), 44 zettabytes¹ will be shared between devices by 2020.

Common used examples of IoT devices are smartphones and vehicles from which one can retrieve information about, for example, the temperature and pressure. Less obvious examples are drying launders, animals in agriculture and streetlights. Such data have many potentials, but one must consider the quality that can be obtained from an analysis of the data. One may ask, based on the data, if X .AND. Y what is the probability for Z ? Considering the entire dataset, there might be billions of hypotheses to consider. Many of those will fail, but there can also be good results purely by chance, which have no meaning in real life applications. Therefore, it is good practice to consider which variables that should be used (see chapter 12) and test the results before upscaling any application. Having said this, methods for dealing with big data are still evolving.

Other crowdsourced data types that have not been considered in detail in this thesis includes data from vehicles. Anderson et al. (2012) collected over 239.000 temperature and pressure observations from nine vehicles in a test facility and examined these by

¹1 zettabyte=10¹² gigabytes

comparing to a weather station. They found good skill of the temperature measurements but concluded that more work is needed to improve the quality of pressure observations from vehicles. Also, Météo France and Deutscher Wetterdienst (DWD) have ongoing projects with the private industry to utilise observations from vehicles (Hintz et al., 2019c). Another potential useful new source of data can come from mobile telecommunicating networks. Zinevich et al. (2009) used the signal strength between mobile masts to quantify rainfall amounts. Suryana (2017) investigated the effect of rainfall intensity on the 5G mobile signal to develop a filter that could reduce the signal loss. Suryana (2017) found rainfall to be the most important factor in the reduction of the mobile signal. Therefore, it is possible that the 5G-network can also be used to quantify rainfall intensity in future. Closed-Circuit Television (CCTV) is another potential data source, which can be used to improve road weather forecasting or to detect fog (Choi et al., 2018).

Dense datasets are mainly useful for nowcasting and short-term NWP forecasting because the observation networks have a much higher spatial resolution than the model resolution. Within one square kilometre there can in some cases be more than ten PWS (Hintz et al., 2019d). The high-resolution information from such dense observation networks is first being smoothed when assimilated on to a coarser grid. Pre-processing needs to be done to ensure that the high spatial resolution of the observations is also representative of the model grid. Further, when the NWP forecast is being integrated forward in time, the information about the small-scale structures that might be in the observations slowly vanishes, which is especially true for regions that are upstream data-sparse regions. To solve this, a rapid-update cycle of the NWP model can be a solution. Also, observation-based nowcasting by interpolation to a grid can help to aid operational meteorologists in decision making (Clark et al., 2018).

Chapter 12

Machine learning

Big data (chapter 11) and Machine Learning (ML) are often used in the same context, but there are big fundamental differences. Big data is about how to handle massive data volumes, whereas ML is about how to use statistical algorithms for predictions, often with large datasets. ML does not necessarily relate to big data or vice versa. With that said, ML proves very useful for finding ‘hidden’ structures in big datasets which is otherwise difficult to find and describe.

In this PhD project, ML has been used to predict the associated error of an SPO, by training a model based on O-B (observation minus background) statistics. Before describing the method and presenting results in chapter 13, the theory is described in this chapter.

12.1 Fundamentals of machine learning

Before going into details with error prediction of SPOs using ML, the fundamentals of ML is briefly discussed to establish a baseline. In popular science especially, terms such as neural networks, big data, and ML are often used commonly and are sometimes misinterpreted. Therefore, the purpose of this section is to ‘unbox’ the fundamentals of ML. ML covers numerous topics, such as classification and decision algorithms. Given some input, x , we want to predict some output y . Traditionally this is done by mapping a function on to x to predict y , such that $f(x) = y$. This can, for example, be done with linear regression if we expect a linear relationship between x and y . However, some problems cannot be solved with traditional methods. Image classification is a good example. If someone was given a task of writing a program that could predict the type of a bird given an image of a bird. For simplicity, assume only two birds are considered, a parrot and a crow. One could write a function that counts the ratio of blue pixels in

the picture. A high ratio could indicate that it is an image of a parrot. However, an image of a crow could be one where there is also blue sky leading to a wrong prediction. One would have to write extensive amounts of if-statements and if one were given the task of predicting a new type of bird, one would need to start completely over again. Instead, one could make use of an algorithm which creates the rules itself, making the solution much faster and more general. That is where ML proves useful.

ML deals with supervised learning, unsupervised learning, and semi-supervised learning. In supervised learning there is a set of training data, meaning that one knows both the input and the output; thus, the algorithm knows the target values and can be optimised by minimising the error of the predictions. This approach is used for both classification and regression problems. In unsupervised learning, there is as the name implies no training data. It is mainly used to find patterns in large datasets, such as if a customer buys *A*, they also tend to buy *B*. In unsupervised learning, it is possible to tell an algorithm that it was wrong after an event has occurred. In this way, it can be trained in long-term. This is partly how self-driving cars are trained (Kendall et al., 2018). Semi-supervised learning is a combination of the two where one has some a priori knowledge of the output for only some of the input.

It is possible to write a general pipeline for all supervised ML problems. First, one must collect training data and split it up into test data and target data. The test data is used after the algorithm has been trained to test its accuracy, which is essential as it is often difficult to know if the best algorithm was chosen. After data collection has been accomplished, one must choose good features. Features are the actual input. If one wants to predict the weight of a person, a good feature would probably be that person's height in meters. A bad feature, on the other hand, would probably be the eye colour or the person's height in centimetres as it is perfectly correlated with the first feature of height in meters. Experience is important when selecting good features. Without good features, it is not realistic to train a model to give good results.

12.1.1 Classifiers

It is outside the scope of this thesis to cover classifiers in general, and so only the general terms will be described. Classifiers all work the same on a high-level. Given some features x they predict an output y . A classifier predicts a discrete class label. For example, the smartphone app *PlantSnap* (Plantsnap, 2019) uses a photo as input to predict a classification label with the name of the photographed plant. Examples

classification algorithms include k-Nearest-Neighbour search, decision trees, and neural networks; thus, using the terms ‘big data’, ‘ML’ and ‘neural networks’ interchangeable are not only dangerous but also wrong. Even though neural networks are one of the most sophisticated classifiers, it works on a much lower level than the terms ‘big data’ and ‘ML’.

12.1.2 Regressors

Regressors predicts a continuous quantity given some input features, x . Ukkonen et al. (2017) used artificial neural networks for identifying occurrences of thunderstorms in Finland and to evaluate thunderstorm predictors. In this PhD project, different regressors have been tested to predict the error of an SPO to be able to correct the observations (see chapter 13).

12.2 Preprocessing of data

In order to predict y as good as possible, pre-processing of data is needed. If one of the input variables explains by far most of the variance of the dataset, it is likely to be the most significant input variable to the model. It might be that one wants to predict the age of a person, based on that person’s annual income, the person’s height and the number of places the person has lived. The dataset might contain heights ranging from 1.5 m to 2.2 m, places lived from 1 to 10, but the annual income might vary from 0 \$ to 100.000 \$. Clearly, almost all of the variance in the dataset is explained by annual income. To solve this issue, the data needs to be scaled appropriately before entering any model. This is can for example be done with either normalization (12.1) as

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (12.1)$$

which scales all numeric data in the range 0-1, or standardization (12.2)

$$x_{\text{scaled}} = \frac{x - \bar{x}}{\sigma}, \quad (12.2)$$

which transform the data to have zero mean ($\bar{x}_{\text{scaled}} = 0$) and unit variance.

12.2.1 Choosing input variables

It is tempting to gather as much data as possible and to let some pre-made ML software do its ‘magic’. However, by having too many variables one risk overfitting the model,

creating a model unable to predict future events. Secondly, this approach introduces a black box approach. Given a model, $y = m_1x_1 + m_2x_2$, where y is the predictand and x are predictors (input data) and m_1 and m_2 are the model parameters. As scientists, we are generally not interested in y but more in the model (m_1 and m_2) explaining the phenomena. Surely, y is interesting and is used for validation, but the physics lies in the model. When working with data models it is therefore important to consider multicollinearity as this can alter the explanations of the system widely.

12.2.1.1 Multicollinearity

Multicollinearity is a term used when dealing with highly correlated predictors. One does not need to consider this effect if the only interest is in the value of the predictand and not the model parameters. However, it is essential to check for multicollinearity if one wants to gain insight into a model. Imagine a purely imaginary dataset consisting of x_1 and x_2 , where the correlation ($r(x_1, x_2)$) is high. One example could be the height and weight of a person, claiming higher persons tends to weigh more. If one wants to predict y (which in this example could be shoe size) using x_1 via linear regression one would obtain a value of the slope m_1 . With m_1 it can be argued that for every unit increase in x_1 , y will increase with the value of m_1 . If x_1 and x_2 is highly correlated, the prediction, y , will not necessarily change by introducing x_2 to our model, but our conclusions might change. By introducing x_2 , the value of m_1 can change, so our conclusion is now different for a unit increase of x_1 . This happens when x_1 and x_2 are correlated and shows that x_2 is indirectly included in the model via x_1 even when x_2 has not been directly specified and hence one risks making wrong conclusions.

12.2.1.2 Curse of dimensionality

The curse of dimensionality describes the effect of increasing the dimensions of the input data first described by Bellman (1961). In ML terms, this is when one increase the number of features. Assume that a dataset $\mathbf{x}^T = \{x_1, x_2, \dots, x_N\}$, exist with $N = 100$ observations. If one were to create a classifier to predict y and one dimension ($D = 1$) exist, one could separate the domain into a number of evenly sized regions. For example, with a uniform distribution, one could choose to create ten regions such that we expect to find $N/10 = 10$ observations within each region on average. Now, if there exist two dimensions ($D = 2$) our dataset would have the form $\mathbf{x}^T = \{(x_{11}, x_{12}), \dots, (x_{N1}, x_{N2})\}$ and instead of ten evenly spaced regions we would have $10^D = 100$ regions, but with $N = 100$ we see that the observation density has decreased from 10 observations per region in average to 1 observation per region in average. Adding a third dimension

results in 1000 regions and an average observation density of only 0.1 observation per region.

From the nearest neighbour (see section 12.3.1) point of view, we increase the Euclidian distance between observations for each dimension we add, making the model less accurate. If reducing the number of dimensions is not an option, the only option to make the model accurate enough is to add more training data. To have the same observation density in three dimensions as in one dimension, one would need $N = 100^3 = 10^6$ observations instead of $N = 100$ for one dimension. Therefore, simply adding extra features is not necessarily the best answer to complicated problems. It quickly becomes difficult to meet the requirement of enough training data for high dimensions. Conversely, having a dataset with many features, one can increase the observation density by reducing the number of dimensions by examining if there exist features that are not strictly needed to predict the target.

12.3 Machine learning models for predicting errors

As was mentioned in section 12.1.2, this PhD project has tested different regressors for predicting errors of SPOs. The results and discussion of these tests are given in chapter 13. First, the models used are presented and described in this section.

12.3.1 k-Nearest-Neighbour

The k -Nearest-Neighbour algorithm is a simple, intuitive model. It has been widely used for classification within satellite imagery and handwritten text recognition (Hastie et al., 2009a). For each data point in feature space, the k nearest neighbours are found from which the target is predicted by linear interpolation (Buitinck et al., 2013). In this study, all neighbours are weighted uniformly, not taking the distance to each into account. Due to the complication with dimensionality (see section 12.2.1.2) it has to be prioritised not to choose too many features.

12.3.2 CART

A gradient boosting regression was also used to predict the error of an SPO. Gradient boosting is a sub-category of Classification And Regression Trees (CART); thus, a brief introduction to CART is given before describing the gradient boosting method. Tree-based methods or decision trees are relatively simple and intuitive to understand. Also,

decision trees have the advantage that it is possible to trace back decisions made based on the input data and thereby examine the feature importance (Hastie et al., 2009b). Figure 12.1 illustrates the workflow of a small, grown tree. Suppose a set of features exists, x_1, x_2 . The feature space is partitioned into sub-regions, in this case, until a leaf node is created, represented by the circles in figure 12.1. At each branch, a split is applied as a boolean condition. For example, on the top level in figure 12.1 it is asked if $x_1 < c_1$, where c_1 is a constant. If *true*, the left branch is followed and if *false* the right branch is followed. This process continues until one of the leaf nodes of the tree is reached.

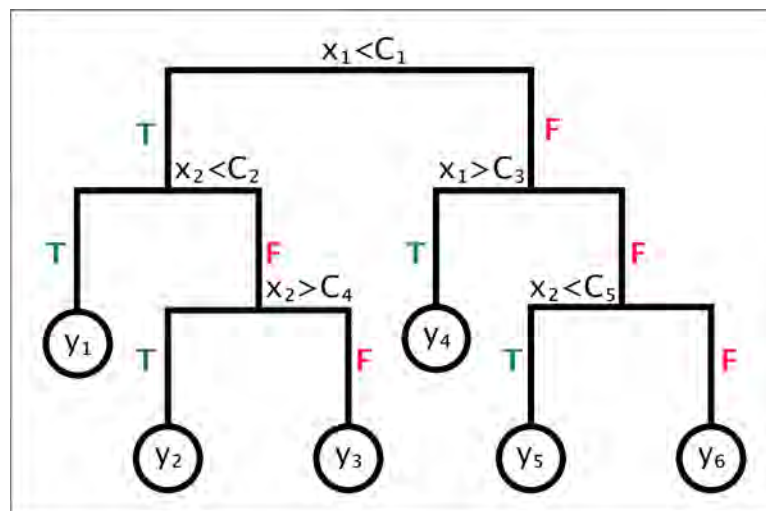


Figure 12.1: Conceptual sketch of a decision tree. T and F stands for *true* and *false* respectively. The black circles represents the leaf nodes of the tree.

When a leaf node is reached, the target, $y_{i=1..6}$, is returned. In reality, y_i represents a region in feature space, and the returned value from the model can be defined as whatever is found suitable for each case. For example, the response value can be defined as the average of all data points within y_i . The reader is referred to Hastie et al. (2009b) and Bishop (2009) for theory on how to grow a tree. This work used the software provided through the scikit-learn API for Python (Buitinck et al., 2013).

Some disadvantages of decision trees are that they can overfit the data if the tree becomes too large and they are not robust to new data, so if new unseen data enters the system, the predictive skill is usually poor. A ‘random forest’ is the development of a

decision tree to decrease the tendency of overfitting. A random forest consists of many decision trees and is a method in which many uncorrelated decision trees are trained. The response will be the average or median of all individual decision trees. The features used to grow each tree is selected randomly from the full feature set.

12.3.2.1 Gradient boosting

The Gradient Boosting method is a method which combines multiple so-called ‘weak learners’ to obtain a single strong learner. A weak learner is only vaguely defined; however, in general, it is a model that scores only slightly better than random guesses (Bishop, 2009). In this sense, the gradient boosting works much like the random forest just explained in section 12.3.2, by creating an ensemble of weak learners. The collection of weak learners can be single decision trees, but it can be any type of model. In the software from Buitinck et al. (2013), used here, decision trees are used as the weak learners; thus, in this case, the random forest method and the gradient boosting method is very similar. However, where a random forest selects the subset of data used to grow the tree randomly, the gradient boosting selects data in an iterative manner (Hastie et al., 2009c). A loss function, which measures the error of the model, is minimised so more weight is given to data points that are difficult to predict in terms of a large error.

Chapter 13

Correcting smartphone pressure observations

Chapter 12 introduced the fundamentals of ML and the algorithms which have been used in this PhD project. ML is used to predict the error of an SPO, which is then corrected by subtracting the predicted error. This chapter describes the methods and presents complementary results from these experiments.

Smartphone observations from the period April 2018 to December 2018 was collected (see table 2.2 for a list of all the collected variables). In total this dataset consist of 44.992.644 observations (see table 2.1). Then two datasets were created for each smartphone individually, a training and test dataset, denoted \mathbf{D}_T , and a verification dataset, denoted \mathbf{D}_V .

\mathbf{D}_T holds all observations from the period April 2018 to November 2018 and \mathbf{D}_V holds all observations from December 2018. Furthermore, \mathbf{D}_T is randomly divided into a training set, \mathbf{D}_T^{tr} , and a test set, \mathbf{D}_T^{te} , using a fraction of 0.7 for \mathbf{D}_T^{tr} and 0.3 for \mathbf{D}_T^{te} . It is then possible to train a model based on \mathbf{D}_T^{tr} and calculate skill scores for the trained models based on \mathbf{D}_T^{te} . \mathbf{D}_V works as an independent dataset, to verify how the trained models perform when presented for data from a period different from that in the training data. The *target* is the error of an SPO given a set of *features*. The error is defined as the observation minus background, where the background is a short-term HARMONIE forecast. Each SPO is then corrected by subtracting the predicted error (target).

First, a baseline denoted BL_NC (‘Baseline No Correction’), is defined in which the SPOs has not been corrected in any way, before calculating the difference of each SPO and a short-term forecast. Also, a reference is defined as the difference of each SPO subtracted by the average error for each smartphone and the short-term forecast and

is denoted BL_MC (‘Baseline Mean Correction’). Then two different ML models were trained for each smartphone present in \mathbf{D}_T^{tr} , to predict the error. The two models is a k-Nearest-Neighbour model and a Gradient-Boosting-Regression model (see chapter 12), in the following abbreviated as ML_KN and ML_GB respectively.

The features for both ML_KN and ML_GB is *latitude, longitude, observed pressure, the standard deviation of the pressure and the hour of the day*. Better choices may exist and are something that should be tested going forward. The reasoning of choosing these features is that the error is dependent on location (latitude, longitude) since observations are taken at different altitudes at, e.g. home and work. If the altitude is precisely the same for different locations, there can still be a varying difference between the terrain of the NWP model and the physical world (see figure A2). The observed pressure is used to include potential systematic errors which can depend on the pressure itself. The standard deviation holds information about the change in altitude while a device is measuring. Standard deviations of the same order of magnitude could, therefore, hold information about the magnitude of the error. The hour of the day is included partly due to the same reasoning of the position, namely that the error can be different at, e.g. 9 UTC and 18 UTC. However, the hour of the day is more natural to normalise than latitude and longitude and the hope is therefore that including the hour of the day as a feature the model will be more robust to new unseen locations. All features was normalised (see equation (12.1)) using the data contained in \mathbf{D}_T^{tr} . Thus, \mathbf{D}_V is also normalised using the minimum and maximum of \mathbf{D}_T^{tr} .

13.1 Complementary results and discussion

In addition to the method described in the previous section, KN was also run without normalising the features and not including the hour of the day as a feature. Figure 13.1 shows a spatial comparison of SPO background deviations using a short-term HARMONIE forecast as background (see chapter 10). All circles represent a single SPO during December 2018, and the colours represent the background deviation. The left figure shows the unprocessed SPOs (BL_NC). In the right figure, the SPOs has been corrected with the predicted error using KN with $N = 3$. The features have not been normalised nor standardised; thus, not surprisingly, there is room for improvements. The RMSE decreased from 2.3 hPa to 1.3 hPa after the correction. For the remaining results in this section, all features have been normalised as described in the previous section.

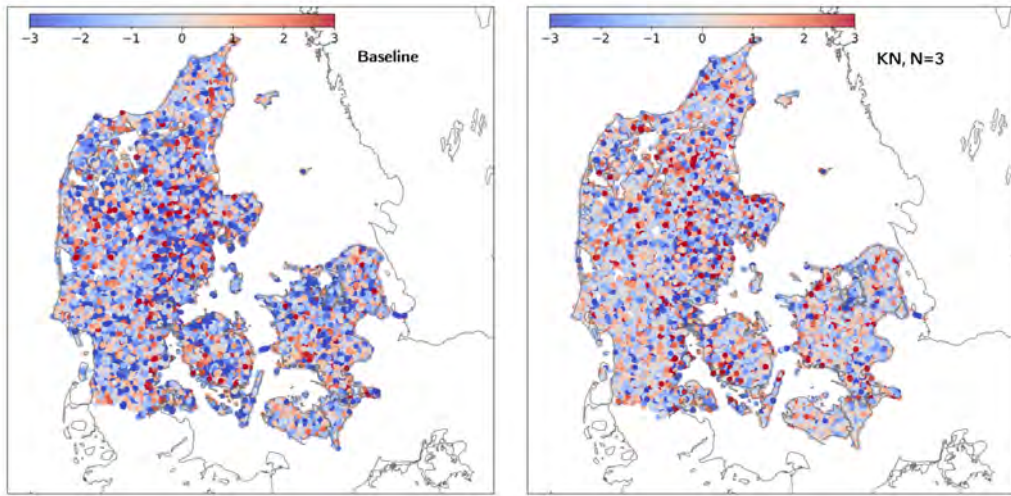


Figure 13.1: SPOs from December 2018. Colours show the background deviation. Left: Untreated SPOs (RMSE=2.3 hPa). Right: SPOs corrected by a prediction from the nearest neighbour model (N=3), using latitude, longitude, pressure and the standard deviation of the pressure as features (RMSE=1.3 hPa).

Figure 13.2 shows the root mean square error of the predicted errors of the SPOs in the test batch, \mathbf{D}_T^{te} . BL_NC scores worst, however, that is not surprising. BL_MC reduces the average error with 0.77 hPa, and the median has decreased with 0.80 hPa. Even larger improvements are seen for both ML_KN and ML_GB, where both the average and median of the errors have decreased by more than 1 hPa.

Figure 13.3 shows the distribution of the bias for the test batch, \mathbf{D}_T^{te} . First, it is noted that BL_NC has a large spread, and it is difficult, if not impossible, to interpret any shape of the distribution in figure 13.3. Again, using the average error as the predicted error, BL_MC improves the bias significantly. In terms of the average, median and standard deviation of the distribution, BL_MC, ML_KN and ML_GB are similar, but with a lower standard deviation of about 0.3 hPa for ML_KN and ML_GB compared to BL_MC.

Figure 13.4 shows the distribution of the standard deviation for the test batch, \mathbf{D}_T^{te} . The distribution of BL_NC and BL_MC is nearly identical. The only difference is rounding errors that make the RMSE, bias and standard deviation of the distributions

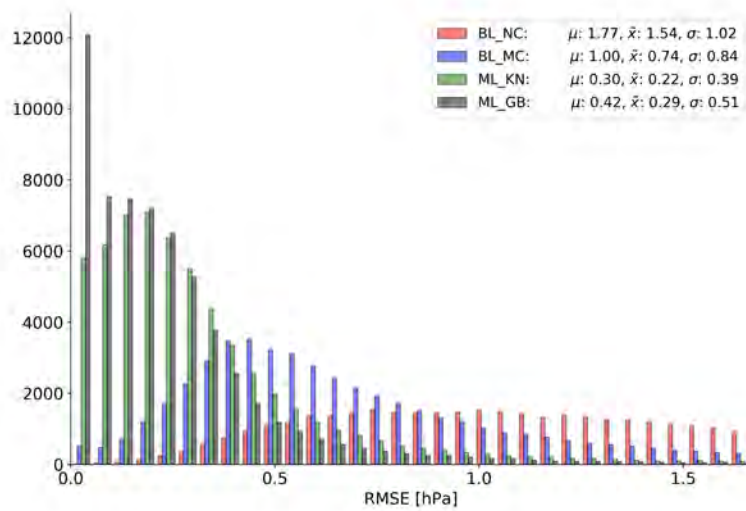


Figure 13.2: Root Mean Square Error of SPO error predictions for the test batch of BL_NC (red), BL_MC (blue), ML_KN (green) and ML_GB (black). μ is the average, \tilde{x} is the median and σ is the standard deviation of the distribution. See text for details.

slightly different. That is because of the way the standard deviation is defined (see equation (C.8)), so instead of subtracting the mean error (bias) before calculating the RMSE, bias and standard deviation, the bias is subtracted when calculating the standard deviation and therefore BL_NC and BL_MC are the same in this case. Also, the distribution of ML_KN and ML_GB for the standard deviation are very similar to the distribution of RMSE, because the bias of both distributions is close to zero.

Figure 13.5 and figure 13.6 shows an example of corrected SPOs contained in \mathbf{D}_V for a period of 45 minutes for ML_KN and ML_GB respectively, plotted as the deviation from the background value. Each row is 15 minutes, and the left column shows the uncorrected SPOs, and the right column shows the SPOs corrected by the predicted error. For both ML_KN and ML_GB, a substantial improvement is seen, indicating that the methods are indeed promising. There are still a few large differences, but it is expected that these can either be removed with improved methods or a suitable screening procedure. It is hard to see a difference between ML_KN and ML_GB by eye. This is expected with figure 13.2, 13.3 and 13.4 in mind. However, small differences are seen on the island of Sealand, where the colours are in general slightly less green (lower

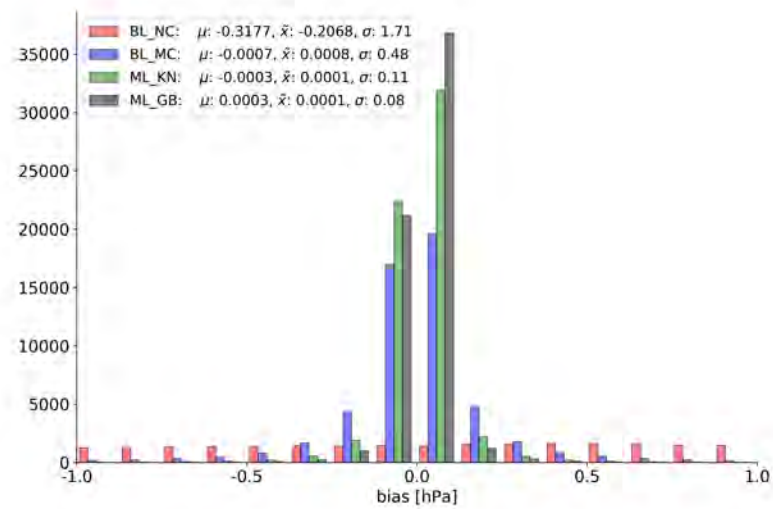


Figure 13.3: Bias of SPO error predictions for the test batch of BL_NC (red), BL_MC (blue), ML_KN (green) and ML_GB (black). μ is the average, \tilde{x} is the median and σ is the standard deviation of the distribution. See text for details.

errors) for ML_GB compared to ML_KN.

The methods that have been presented in chapter 12 and this chapter have shown promising results. It has been shown how two different ML models give better results than correcting observations with simple bias correction. This is mostly because the ML models can take into account that the devices are non-stationary. Therefore, one must expect different bias from the same device. Training a ML model for each smartphone is a rather heavy computational process. However, in operational systems, the training can be done in near-real time as new observations enter a system. Alternatively, it is possible to take advantage of recent advances within the operating systems of mobile devices. Both Android and iOS have implemented ML-kits into each of their developer SDKs (Apple Inc., 2019; Google Developers, 2019a). With these frameworks, it is possible to take advantage of the computational power of each smartphone by training the model directly on the device itself. Then, instead of sending an observation with relevant metadata to a database, the corrected pressure observation can be sent directly to the observing system.

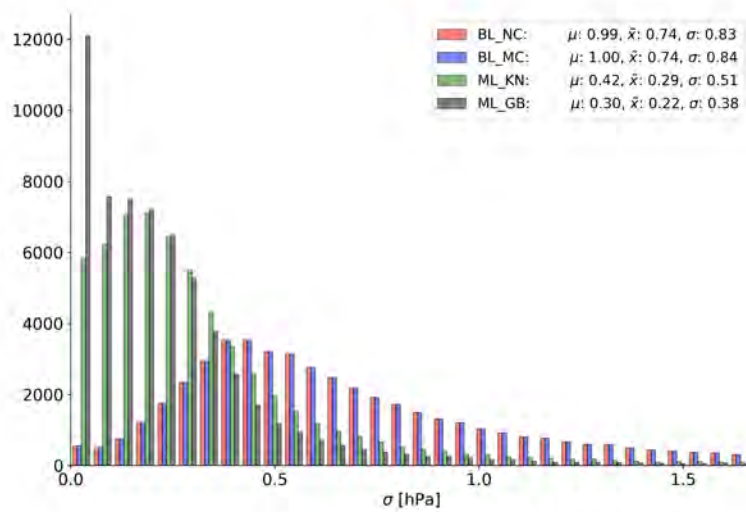


Figure 13.4: Standard deviation of SPO error predictions for the test batch of BL_NC (red), BL_MC (blue), ML_KN (green) and ML_GB (black). μ is the average, \tilde{x} is the median and σ is the standard deviation of the distribution. See text for details.

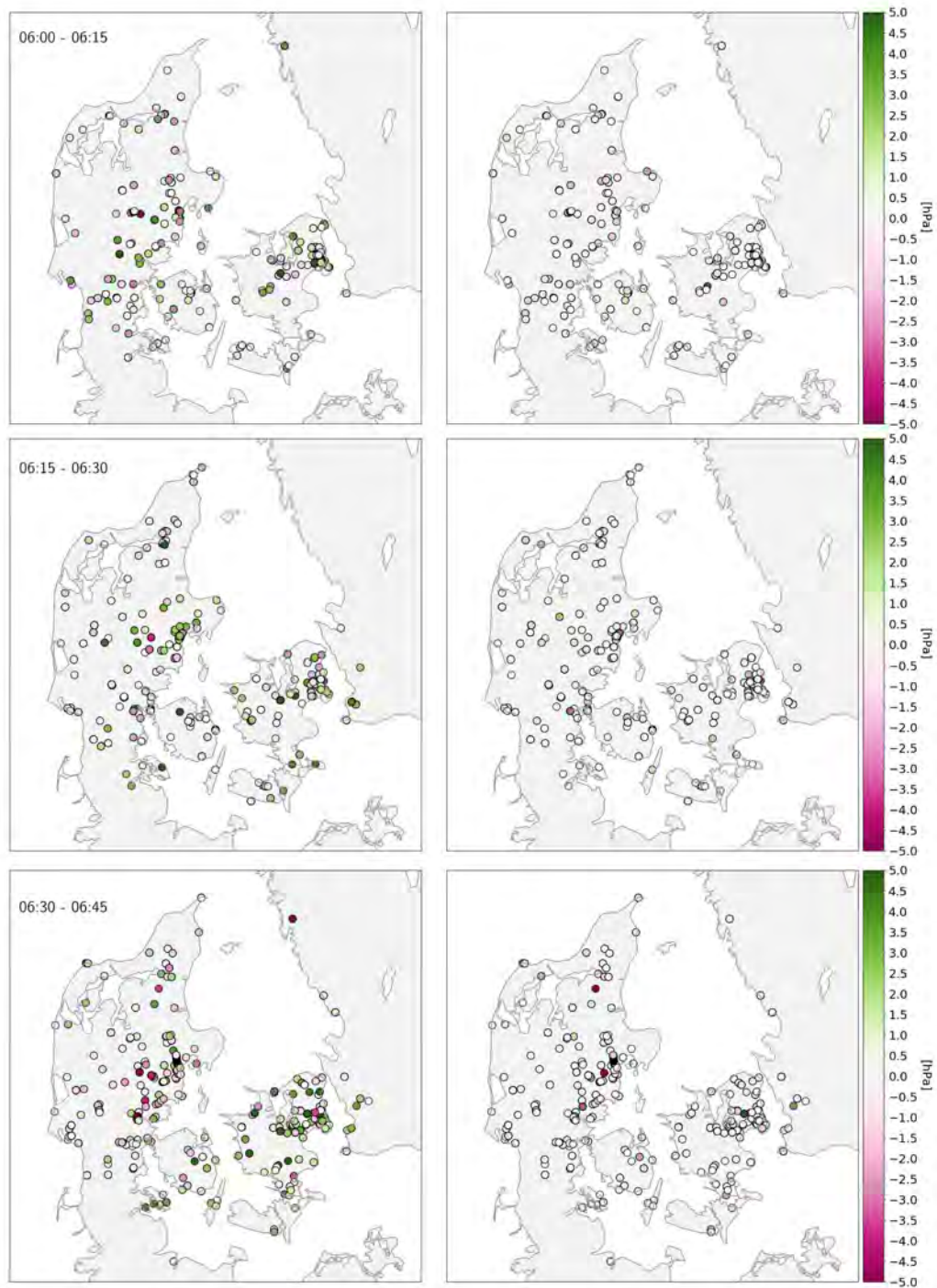


Figure 13.5: Difference between SPOs and NWP surface pressure. Left column: Raw SPOs. Right Column: SPOs corrected by the predicted error using K-Nearest-Neighbour method. Each row is 15 minutes. The time interval is seen in left-column plots. The date is 12 Dec 2018. Only smartphones with over 50 observations in total has been used.

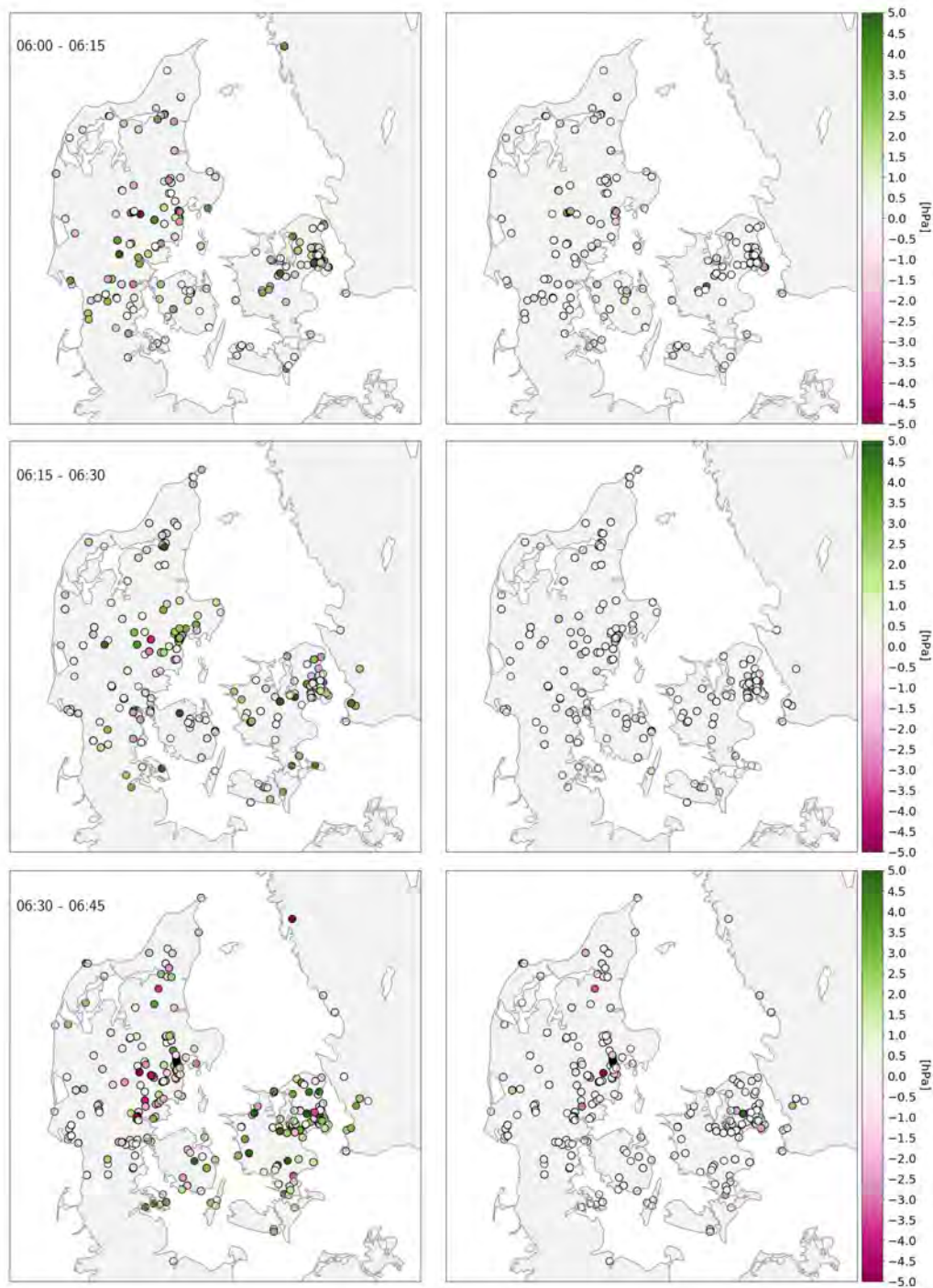


Figure 13.6: Difference between SPOs and NWP surface pressure. Left column: Raw SPOs. Right Column: SPOs corrected by the predicted error using a Gradient Boost model. Each row is 15 minutes. The time interval is seen in left-column plots. The date is 12 Dec 2018. Only smartphones with over 50 observations in total has been used.

Chapter 14

Collecting and processing of Barometric Data from Smartphones for Potential use in NWP Data Assimilation - Publication B

Utilising SPOs and correctly assimilate these into modern NWP models are a new field of research and relatively unexplored, but with vast potential. Few studies exist which have collected SPOs (Kim et al., 2015, 2016; Madaus and Mass, 2017; McNicholas and Mass, 2018; Price et al., 2018). Kim et al. (2016) was the first to apply machine learning algorithms for correcting errors of SPOs. Madaus and Mass (2017) assimilated SPOs into the WRF model (Skamarock et al., 2008) (see also chapter 4) using an Ensemble Adjustment Kalman Filter (EAKF). However, SPOs was not included as an integral part of the DA-system. In a continuing study McNicholas and Mass (2018) improved the assimilation and used a LETKF scheme to assimilate SPOs into the WRF model. Also, McNicholas and Mass (2018) presented a promising approach of predicting errors of SPOs using a random forest machine learning algorithm. With this, McNicholas and Mass (2018) was able to get useful information from moving sensors and sensors away from the surface. While Kim et al. (2015), Kim et al. (2016) and McNicholas and Mass (2018) collected SPOs via a self-developed app, Price et al. (2018) obtained data from a third-party provider. Due to privacy constraints, it was not possible to obtain a unique identifier for each device, and hence, Price et al. (2018) was not able to bias correct the observations. Also, as already described (see chapter 2), making a full single app makes the data source potentially unstable.

Hintz et al. (2019b) presents an improved method to collect SPOs from several sources by integrating a SDK (see section 2.2), which can be used in an unlimited number of third-party apps; thus, there is no need to spend time on designing UI and UX. Hintz et al. (2019b) is the first study to assimilate SPOs using variational DA. The article is divided into three parts. First laboratory examinations of smartphone barometric sensors are performed to clarify which orders of magnitude of bias that can

be expected and to investigate if the bias drifts over time. Secondly, the method of data collection, processing of data client-side and a method of screening SPOs for poor quality are presented. Thirdly, the NWP system and how the SPOs was assimilated is presented.

Processing of data client-side needs to be done carefully. Figure 14.1 shows 50 individual measurement series of pressure from a smartphone barometer over 180 s with a sampling rate of 1 Hz subtracted by the mean of each measurement. The solid black line shows the average between all 50 measurements. It is seen that there is a spin-up period at the beginning of a measurement. This is due to the internal IIR (Infinite Impulse Response) filter as described by McNicholas and Mass (2018). Therefore, the first few samples should always be skipped when the barometer is accessed. Here the first 5 s is always ignored independent of OS. iOS returns a sampling rate of 1 Hz, while Android returns a sampling rate dependent on the specifications of the barometric sensor. For a Samsung Galaxy S7, the sampling rate is approximately 4 Hz. Due to the difference of sampling rate, it was decided to skip the first 5 s, instead of the first five samples, to be consistent over different OS. Hereafter an average is computed iteratively before sending data to the server.

The iOS SDK does not allow apps to run continuously in the background by default¹, which causes observations to enter the database irregular in time. Figure 14.2 shows the total distribution of observations entering the database as a function of time of day in the period 4th of April 2018 to 24th of May 2018. A sharp rise in incoming observations is seen in the morning (the black vertical dashed line shows the approximate time of sunrise in Denmark), whereafter it declines and levels out. In the evening a rapid decline is seen with relatively low numbers throughout the night.

The uneven distribution in time makes it problematic to obtain a reliable bias for unique devices fast, as a minimum number of reports are required to compute the bias. Price et al. (2018) found that the systematic bias does not change significantly over time; thus, if the bias of individual smartphones can be found, and if the observations are either corrected to the surface or known to be taken at the surface, the SPOs can be assimilated directly.

Running operations in background mode is done differently on different OS such as

¹Activities in the background can be enabled by setting certain background modes. For example, an app can be activated by sending push notifications. See Apple Inc. (2017) for details.

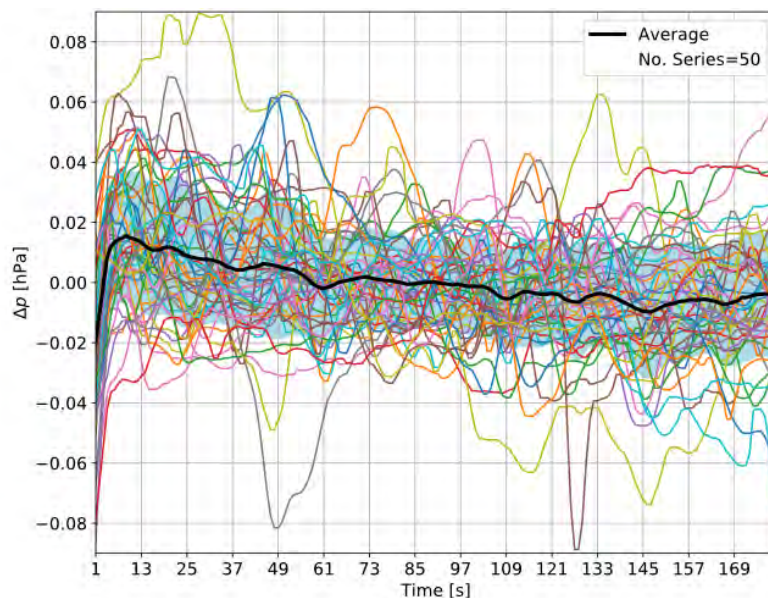


Figure 14.1: Spin-up by the barometric sensor in an iPhone 7 (Bosch BMP280). Each coloured curve is a single measurement series over 180 s, subtracted by its average value. The full black line shows the average of 50 measurements series.

iOS and Android. In general, the Android OS is more open for background executions than the iOS OS. However, with the introduction of the Android OS version 8 in August 2017, background executions have also been more restricted here (Google Developers, 2019b). It is still possible to obtain more frequent observations than shown in figure 14.2, on both iOS and Android; however, if that is also the case in future is unknown. Screening methods for crowdsourced data should therefore not be based on assumptions of observations coming regularly in time.

A key finding of Hintz et al. (2019b) is that the vertical position of individual smartphones is still too inaccurate to utilise SPOs directly. Some preprocessing must be done before the observations can be assimilated. However, this issue is likely to reduce in future as the accuracy of vertical positioning increases (Robustelli et al., 2019). Until then, screening of SPOs is essential. The second part of Hintz et al. (2019b) presents a method for screening SPOs. The screening method is divided into separate blocks that can be turned on and off. First, a ‘flagging’ vector containing penalties for all observations are allocated. Each block can then add a penalty to the relevant observation. With such setup, it can easily be tested how each block affects the screening procedure

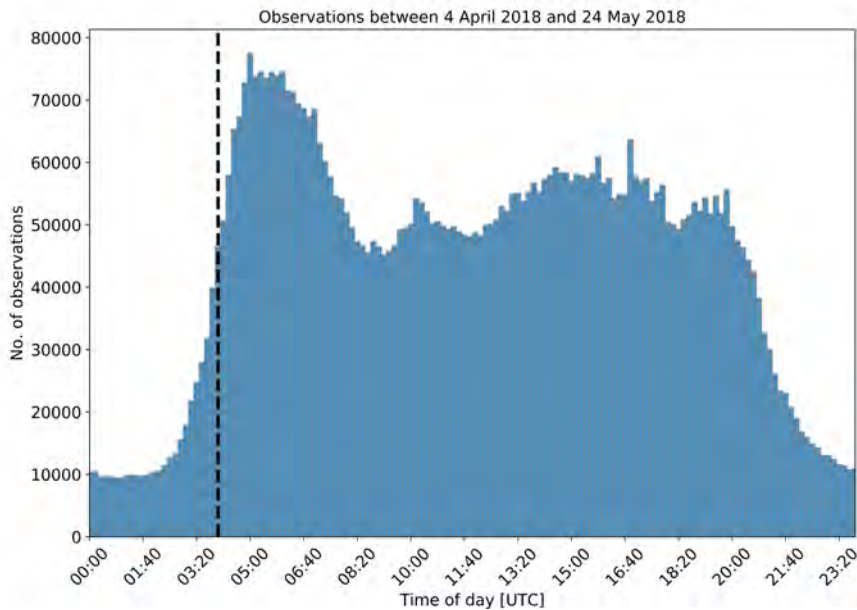


Figure 14.2: Distribution of SPOs entering the database as a function of time of day during the period 4th of April 2018 to 24th of May 2018. The black vertical dashed line shows the approximate time of sunrise in Denmark.

and no observations are discarded but rather stored with an associated penalty value. In the end, the SPOs which have a penalty less than a chosen threshold can be written to a HARMONIE compliant observation file (see the OBSOUL format described in chapter 10).

The third and final part of Hintz et al. (2019b) examines assimilation of SPOs in HARMONIE. The DA system was modified to allow SPOs to enter the assimilation system (see chapter 10). Four different runs over a period of 5 days with 3-hour assimilation cycles were done to examine the impact of different settings of the screening procedure: a reference with no SPOs, a run with all SPOs (no screening), a run with relaxed screening and finally a run with strict settings of the screening. Hintz et al. (2019b) finds that the HARMONIE DA system is not prepared for receiving SPOs without any prior screening. It is also shown that the bias decreases relative to the reference in both cases when screening is used, without a change in the root mean square error. The article contributes positively towards assimilating crowdsourced data in NWP operationally.

Chapter 15

Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018 - Publication C

Crowdsourced data in NWP is still a rather new field of research; thus, communities have not yet been established in which sharing of knowledge and ideas can take place efficiently. Further, there has been a lack of awareness about ongoing activities at other institutions and meteorological services. DMI has come a long way already with data collection of crowdsourced meteorological observations, assimilation of SPOs in NWP and legal issues regarding user privacy. Still, challenges exist on which it is believed that the meteorological community should collaborate closer to solve. This includes methods of data collection, data processing, data storage and usage. To improve on the previously mentioned points, a workshop at DMI was planned with two main aims: 1) to gather experts on crowdsourced data focused on NWP, to start a network of people working on the subject and 2) producing a white paper directing the research community towards best practices and guidelines on the subject. Hintz et al. (2019c) is the resulting paper and the main product of the workshop.

The workshop was planned and lead by the PhD candidate and was funded by DMI. Planning started in August 2018 and invitation letters were sent out during September, and October 2018 to institutions and individuals who were thought could benefit from attending the workshop. Great interest was seen from many national meteorological institutes and universities. Ultimately the following institutions were represented at the workshop:

- European Centre for Medium-range Weather Forecasts (ECMWF)
- Estonian Environment Agency (ESTE A)
- European Meteorological Services Network (EUMETNET)

- Danish Meteorological Institute (DMI)
- Deutsche Wetterdienst (DWD)
- Finnish Meteorological Institute (FMI)
- Met Éireann
- Météo France
- Met Office, UK
- Koninklijk Nederlands Meteorologisch Instituut (KNMI) (Royal Netherlands Meteorological Institute)
- University of Copenhagen, Denmark
- University of Reading, UK
- University of Washington, USA

The workshop spanned two days with the first day allocated for presentations and the second day allocated for discussions.

Prior to the workshop, European Meteorological Services Network (EUMETNET) had arranged a meeting on crowdsourced data with a more generic focus, hosted by the FMI. The EUMETNET made an excellent starting point; however, there was a prominent focus on image processing, legal issues and data formats. Crowdsourcing is a complex subject which is hard to reach general conclusions for. Therefore, the DMI workshop narrowed the subject to crowdsourcing in NWP. Still, many topics were represented at the workshop, such as data from vehicles, user reports of current weather, PWS, data management, legal issues and data from smartphones.

The first day was allocated primarily for presentations by the participants on current activities and challenges followed by a short discussion and questions for each presenter. Day one ended with an open discussion facilitated by the PhD candidate, and a networking dinner funded by the DMI.

The second day of the workshop started with a single presentation by the PhD candidate on the subject of legal issues on which DMI has gathered experience (see section 2.4.1). Afterwards, two discussions groups were formed, each with a different

15. Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018 - Publication C 103

set of questions based on the discussions from day one. Finally, the workshop ended with a conclusive discussion where recommendations and conclusions for a wider community were agreed on. Hintz et al. (2019c) gives a summary of the presentations to provide the community with an overview of ongoing activities and to share the key findings from the workshop with the public.

Chapter 16

Evaluating pressure observations from private weather stations and smartphones - Publication D

During the work of the three first articles (Hintz et al., 2019a,b,c) it has become clear that one of the most challenging parts of observations from smartphones, is the fact that these are non-stationary, which makes bias-correction and screening more difficult. Therefore, it was decided to investigate data from PWS further as another potential source of data for NWP. Because weather stations from Netatmo (Netatmo SAS, 2019b) are popular in Denmark and the rest of Europe, these became the main data source for this study. However, other manufacturers do exist (Fieldsense A/S, 2019; Lonobox, 2019; Rainwise Inc., 2019). A Netatmo weather station has by default one indoor and one outdoor module, which both measures the temperature and relative humidity. The indoor module also measures air pressure. The indoor module is connected to WiFi and is powered by cable, while batteries power the outdoor module. The indoor and outdoor modules communicate wirelessly via Bluetooth. The weather station can be expanded with a rain gauge and an ultrasonic anemometer, but data from these are not considered here; however, rainfall observations from rain gauges are being considered by the DMI radar group as a source of data, useful for verification of precipitation (see figure A4).

Observations from the Netatmo API (Netatmo SAS, 2019a) was continuously retrieved since 1st of February 2018 with a sampling rate of 30 minutes. Not all retrievals were successful, either because the API was under maintenance or stations not reporting. Furthermore, Netatmo SAS (2019a) defines a fair-use policy so that responses are being limited if one makes too many queries. To query observation data, a rectangle is defined, and at most 1.000 stations within the area defined by the rectangle are being returned. One can, of course, make many small rectangles and query observations from each of these. However, that would conflict with the fair-use policy. The Netatmo API

(Netatmo SAS, 2019a) returns the observed outdoor temperature and relative humidity and the (indoor) air pressure. Other indoor observations are not available through the public API.

For three individual Netatmo stations, the authors got access to all observation data from 1 April 2018 to 31 March 2019, with a temporal resolution of one hour. The bias of the three stations is examined in detail in Hintz et al. (2019d), in which a varying bias throughout the year is found. A conclusion on the reasons for this is yet to be determined; however, a high correlation with the observed indoor temperature is found, which can indicate that there is a temperature dependency which is not adequately accounted for. To investigate this further, the authors acquired two Netatmo stations to perform controlled experiments. These experiments are currently being designed and are planned to start in July 2019. Therefore, these are not yet a part of Hintz et al. (2019d), but are essential experiments needed to finalise the article. However, conclusions based on public data can already be made.

Furthermore, it has been a strong wish of the authors of Hintz et al. (2019b) to extend the simulation period of five days of the NWP run in which SPOs was assimilated. Therefore, both SPOs and Netatmo pressure observations has been assimilated into the DMI HARMONIE NWP model using 3D-Var in two separate runs spanning nearly two months. The methods used to handle SPOs follows Hintz et al. (2019b). Pressure observations from Netatmo were bias-corrected using two months of short-term HARMONIE NWP forecasts prior to the model run. It was believed that data-thinning of surface observations was implemented in the HARMONIE model. However, when examining the observation usage, it became clear that it was not the case, meaning that all observations that passed the observation screening were used in the minimisation routine. Therefore, an extra run will be performed, which does include data thinning of the Netatmo pressure observations.

Another aspect of using observations from PWS is data tampering. The size of such an issue is yet to be determined. However, different examples were revealed during this study. In one case, a Netatmo user placed a station at open sea. While that is a case that is easy to detect, it raises the question about how many stations are positioned wrong, either because the user has chosen a wrong location deliberately or by accident. Also, station owners have the possibility to ‘calibrate’ their pressure observations by adding an offset to their observations, and there is no information of calibration offsets in the response of the Netatmo API. Bell et al. (2015) presented a year-long field study

in which seven PWS was compared to professional equipment and concluded that the bias can be corrected for if the bias is considered for individual stations. However, to fully utilise observations from PWS, monitoring systems will need to be developed, which takes into account rapid changes in observed variables and movement of the station both horizontally and vertically. Bell et al. (2015) considered only the quality of the observations, not the behaviour of data coming from real applications. To do so, a unique device identifier is necessary. The Netatmo API (Netatmo SAS, 2019a) returns the device MAC address, but that may be different for each service provider. Hintz et al. (2019d) is the foundation for developing such a monitoring system.

17.1 Summary

This PhD project has been investigating the potential use of crowdsourced data in Numerical Weather Prediction (NWP) with a focus on data from smartphones and Personal Weather Stations (PWS). The project spans a variety of subjects necessary for working with crowdsourced data. First of all, the main subject has been on the use of crowdsourced meteorological variables for use in NWP and weather forecasting. To study such data, it is, of course, necessary to acquire the data. Therefore, a significant part of this PhD project has been on methods of collecting data and developing software for this purpose. A consequence of being one of the first to collect observations owned by the public, in the same period that the European General Data Protection Regulation (GDPR) act was implemented, meant that much effort also had to be put into legal issues regarding collection, storage and handling of the data. Chapter 2 described how data was collected from both smartphones and PWS and how the project was made GDPR compliant.

Chapter 3 and chapter 4 gave a general introduction to NWP and the forecasting system implemented at Vaavud, respectively. The Vaavud NWP system was used in the work of the first article (introduced in chapter 8), in which a new method of estimating the surface roughness using handheld wind measurements from smartphones was presented. Chapter 5 and chapter 6 introduced the background theory for this method. Chapter 7 gave a short introduction to traditional observations of wind and pressure.

After introducing the first article in chapter 8, the focus changed to pressure observations from smartphones and PWS. Chapter 9 and chapter 10 introduced the background theory for relevant data assimilation schemes and the assimilation system of the HAR-

MONIE NWP model respectively. Other relevant schemes were also described briefly. Then a clarification of the terms ‘big data’ and ‘crowdsourced data’ were provided in chapter 11 together with a brief discussion on other potential crowdsourced data sources which has not been examined in this PhD project. Then, promising preliminary studies in which SPOs were corrected by predicting the errors using machine learning models were presented, first by introducing fundamental concepts of machine learning in chapter 12 and then presenting complementary results to the study as a whole in chapter 13.

Hereafter, an introduction to the last three articles was given in chapter 14, 15 and 16, followed by this summary and conclusions.

17.2 Conclusions

It is deemed that this PhD project has given a positive contribution to the focus on crowdsourced data in NWP as the first and only study which have investigated Handheld Wind Measurements (HWMs) from smartphones, and the first study that has actively collected Smartphone Pressure Observations (SPOs) in Europe. It has been shown that crowdsourced data is valuable for NWP when dealt with appropriately, which can be seen as the primary conclusion of this PhD project.

It was shown that HWMs from smartphones could in some cases be more representative for an area than nearby official weather stations and one conclusion is that HWMs can potentially contribute to give operational meteorologists a better overview of current weather conditions.

An efficient, yet simple method to collect pressure observations from smartphones has been presented. A screening scheme for SPOs has been developed, which improves a short-term forecast when used in HARMONIE. Thus, a full pre-operational production line for SPOs has been developed as a part of this project, which therefore can be made operational with little effort. However, it is also recommended to examine the impact of SPOs when assimilated with 4D-Var in future, as this is expected to perform better than 3D-Var. Potential improvements to the screening scheme have been suggested. It is also concluded that the quality of SPOs increases in future, as the built-in navigation systems in smartphones improves, thereby reducing the error on the retrieved altitude of the smartphone. Also, it is concluded that screening methods and monitoring software for PWS needs to be improved with a focus on the inexact nature of crowdsourced data.

Predicting the errors of SPOs using two different machine learning models and using

these to correct the observations proved to decrease the errors of the SPOs compared to standard mean bias correction.

It is concluded that all of the objectives of this PhD study have been met (see section 1.2). For all the pre-defined success criteria to be met, the methods still need to be applied in one or more commercial projects. However, this criterion was higher prioritised when the project acted as a part of Vaavud ApS. Furthermore, the commercial criterion is met if the findings of this study are operationalised. Also, the part of the study which has investigated data collection from smartphones is a part of multiple commercial smartphone applications already.

Chapter 18

Future work and challenges

There are still challenges to solve before all types of crowdsourced data can be used in NWP. The use of other types of data assimilation schemes needs to be investigated in future studies. A few examples of potential schemes have already been given in chapter 9. Assimilating SPOs with different methods such as 4D-Var, nudging and KF to study observation impact would be one intriguing study. It is expected that these methods are better suited to take into account the high temporal and spatial resolution that crowdsourced data has. Also, it is crucial that data assimilation methods, which include crowdsourced data, are scale-aware, as the observation density varies dramatically. For example, in urban regions, there is usually many orders of magnitude more observations than there is from outside these regions.

This study has collected a large amount of data in total. However, only a fraction of the potential data has been collected. The methods used in this study to handle the massive amounts of data is likely not sufficient in the future. This calls for an interdisciplinary research study between data scientists and meteorological scientists on how to handle and process such data. Also, a collaboration between national meteorological services has to be worked further on. It will benefit all countries if crowdsourced data collection is a united process.

For future work, tests of crowdsourced observation-based nowcasting should be made. For example, if SPOs can be used to derive reliable pressure tendencies, the vertically integrated divergence can be calculated. Plots of either deviation of pressure tendencies or divergence from model expectations can be used by operational meteorologist to detect fast developing convective systems in near real-time.

Other, less difficult challenges, which can be solved more quickly and with fewer

resources and are listed below:

- The SDK for collecting SPOs, should be improved by adding a possibility of collecting observations in background mode if a user grants access. Running these tasks in background mode complicates the software significantly and are mostly a programming challenge.
- Experiments with machine learning methods to predict SPO errors should be continued. The results from preliminary studies are promising, and the fundamental programming environment for carrying out these experiments are ready for use. These methods can potentially improve the screening scheme developed in this PhD project.
- A new bias correction scheme for PWS is a high priority for future studies. A key part of a future scheme is a monitoring system which detects if stations are moved or experience a sudden shift in bias. A long-term investigation of bias under a controlled environment is the initial state of developing a new scheme which was started at DMI in July 2019.

References

- Amazon (2019), Amazon Web Services Elastic Compute Service, aws.amazon.com/ec2/, [Online; Accessed 8 January 2019].
- Anderson, A. R. S., M. Chapman, S. D. Drobot, A. Tadesse, B. Lambi, G. Wiener, and P. Pisano (2012), Quality of Mobile Air Temperature and Atmospheric Pressure Observations from the 2010 Development Test Environment Experiment, *Journal of Applied Meteorology and Climatology*, 51(4), 691–701, doi:10.1175/JAMC-D-11-0126.1.
- Apple Inc. (2017), Apple Developer Archive, <https://developer.apple.com/library/archive/documentation/iPhone/Conceptual/iPhoneOSProgrammingGuide/BackgroundExecution/BackgroundExecution.html>, [Online; Accessed 1 May 2019].
- Apple Inc. (2019), Core ML - Integrate machine learning models into your app, <https://developer.apple.com/documentation/coreml>, [Online; Accessed 12 June 2019].
- Arya, S. P. (1995), *Atmospheric Boundary Layer and Its Parameterization*, pp. 41–66, Springer Netherlands, Dordrecht, doi:10.1007/978-94-017-3686-2_3.
- Arya, S. P. (2001), *Introduction to micrometeorology*, Volume 79 in the International geophysics series, 2nd ed., Academic Press, San Diego.
- Ascher, U. M., S. J. Ruuth, and R. J. Spiteri (1997), Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations, *Applied Numerical Mathematics*, 25(2-3), 151–167, doi:10.1016/S0168-9274(97)00056-1.
- Battan, L. J. (1973), Radar observation of the atmosphere, *Quarterly Journal of the Royal Meteorological Society*, 99(422), 793–793, doi:10.1002/qj.49709942229.

- Bauer, C. (2013), On the (In-)Accuracy of GPS Measures of Smartphones: A Study of Running Tracking Applications, in *Proceedings of International Conference on Advances in Mobile Computing & Multimedia - MoMM '13*, pp. 335–341, ACM Press, doi:10.1145/2536853.2536893.
- Bell, S., D. Cornford, and L. Bastin (2015), How good are citizen weather stations? Addressing a biased opinion, *Weather*, 70(3), 75–84, doi:10.1002/wea.2316.
- Bellman, R. (1961), *Adaptive Control Processes: A Guided Tour*, Princeton Legacy Library, Princeton University Press.
- Bénard, P., J. Vivoda, J. Mašek, P. Smolřková, K. Yessad, C. Smith, R. Brořková, and J.-F. Geleyn (2010), Dynamical kernel of the Aladin–NH spectral limited-area model: Revised formulation and sensitivity experiments, *Quarterly Journal of the Royal Meteorological Society*, 136(646), 155–169, doi:10.1002/qj.522.
- Bengtsson, L., U. Andrae, T. Aspelién, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, K.-I. Ivarsson, G. Lenderink, S. Niemelä, K. P. Nielsen, J. Onvlee, L. Rontu, P. Samuelsson, D. S. Muñoz, A. Subias, S. Tijn, V. Toll, X. Yang, and M. Ø. Kølřzow (2017), The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System, *Monthly Weather Review*, 145(5), 1919–1935, doi:10.1175/MWR-D-16-0417.1.
- Bergeron, N. E., and A. D. Abrahams (1992), Estimating Shear Velocity and Roughness Length From Velocity Profiles, *Water Resour. Res.*, 28, 2155–2158, doi:10.1029/92WR00897.
- Bishop, C. M. (2009), *Pattern recognition and machine learning*, Information science and statistics, 8th ed., Springer, New York, NY.
- Blackadar, A. K., and H. Tennekes (1968), Asymptotic Similarity in Neutral Barotropic Planetary Boundary Layers, *Journal of the Atmospheric Sciences*, 25(6), 1015–1020, doi:10.1175/1520-0469(1968)025<1015:ASINBP>2.0.CO;2.
- Bluestein, H. (1993), *Synoptic-Dynamic Meteorology in Midlatitudes*, vol. 2, Oxford University Press.
- Buitinck, L., G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux (2013), API design for machine learning software: experiences from the scikit-learn project, in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.

- Buresti, G. (2009), Notes on the role of viscosity, vorticity and dissipation in incompressible flows, *Meccanica*, *44*(4), 469–487, doi:10.1007/s11012-009-9216-0.
- Businger, J. A., J. C. Wyngaard, and Y. I. and E. F. Bradley (1971), Flux-Profile Relationships in the Atmospheric Surface Layer, *J. Atmos. Sci.*, *28*, 181–189, doi:10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2.
- Carl, D. M., T. C. Tarbell, and H. A. Panofsky (1973), Profiles of Wind and Temperature from Towers over Homogeneous Terrain, *Journal of the Atmospheric Sciences*, *30*(5), 788–794, doi:10.1175/1520-0469(1973)030<0788:POWATF>2.0.CO;2.
- Choi, Y., H.-G. Choe, J. Y. Choi, K. T. Kim, J.-B. Kim, and N.-I. Kim (2018), Automatic Sea Fog Detection and Estimation of Visibility Distance on CCTV, *Journal of Coastal Research*, *85*, 881–885, doi:10.2112/SI85-177.1.
- Clark, M. R., J. D. C. Webb, and P. J. Kirk (2018), Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations, *Meteorological Applications*, *25*(3), 472–492, doi:10.1002/met.1715.
- Clarkmasts (2017), PU Clarkmasts, <http://www.clarkmasts.com/products/military-masts/pu-mast/>, [Online; Accessed 15 January 2017].
- Cotton, W. R., G. H. Bryan, and S. C. Van den Heever (2011), *Storm and cloud dynamics: the dynamics of clouds and precipitating mesoscale systems*, no. 99 in International geophysics series, 2nd ed., Acad. Press.
- Counihan, J. (1975), Adiabatic Atmospheric Boundary Layers: A Review and Analysis of the Data From the Period 1880-1972, *Atmospheric Environment*, *9*, 871–905, doi:10.1016/0004-6981(75)90088-8.
- Danish Environmental Protection Agency (2015), Danmarks Højdemodel, DHM/Terræn, <https://sdfe.dk/hent-data/danmarks-hoejdemodel/>, [Online; Accessed 20 November 2018].
- DeiC Abacus (2018), DeiC National HPC Center, University of Southern Denmark, www.deic.dk/en/HPCCenter_SDU, [Online; Accessed 8 December 2018].
- Descombes, G., T. Auligné, F. Vandenberghe, D. M. Barker, and J. Barré (2015), Generalized background error covariance matrix model (GEN_BE v2.0), *Geoscientific Model Development*, *8*(3), 669–696, doi:10.5194/gmd-8-669-2015.

- Docker (2019), Docker Container Documentation, <https://www.docker.com/resources/what-container>, [Online; Accessed 21 June 2019].
- Driesenaar, T. (2009), General description of the HARMONIE-AROME model, <http://hirlam.org/index.php/hirlam-programme-53/general-model-description/mesoscale-harmonie>, [Online; Accessed 21 June 2018].
- Droste, A. M., J. J. Pape, A. Overeem, H. Leijnse, G. J. Steeneveld, A. J. Van Delden, and R. Uijlenhoet (2017), Crowdsourcing Urban Air Temperatures through Smartphone Battery Temperatures in Sao Paulo, Brazil, *Journal of Atmospheric and Oceanic Technology*, *34*(9), 1853–1866, doi:10.1175/JTECH-D-16-0150.1.
- Durran, D. R. (2010), *Numerical Methods for Fluid Dynamics, Texts in Applied Mathematics*, vol. 32, Springer New York, doi:10.1007/978-1-4419-6412-0.
- Earth System Research Laboratory (2019), WRF NAMELIST.INPUT FILE DESCRIPTION, esrl.noaa.gov/gsd/wrfportal/namelist_input_options.html, [Online; Accessed 20 February 2019].
- Eckermann, S. (2009), Hybrid $\sigma - p$ Coordinate Choices for a Global Model, *Monthly Weather Review*, *137*(1), 224–245, doi:10.1175/2008MWR2537.1.
- ECMWF (2016), Changes in ECMWF model, <http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model>, [Online; accessed 7-July-2016].
- ECMWF ODB (2018), ODB Database, <http://apps.ecmwf.int/odbgov/obstype/>, [Online; Accessed 2 August 2018].
- Edward, A. (1992), *Likelihood - Expanded Edition*, The Johns Hopkins University Press 1992.
- European Union (2018), REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679, [Online; Accessed 6 October 2018].
- Evans, D. (2011), How the Next Evolution of the Internet Is Changing Everything, *Tech. rep.*, Cisco.
- Facebook Inc. (2017), React Native, <https://facebook.github.io/react-native/>, [Online; Accessed 12 September 2017].

- Field, P. R., R. Brožková, M. Chen, J. Dudhia, C. Lac, T. Hara, R. Honnert, J. Olson, P. Siebesma, S. de Roode, L. Tomassini, A. Hill, and R. McTaggart-Cowan (2017), Exploring the convective grey zone with regional simulations of a cold air outbreak, *Quarterly Journal of the Royal Meteorological Society*, *143*(707), 2537–2555, doi:10.1002/qj.3105.
- Fieldsense A/S (2019), Fieldsense Weather Stations, <https://site.fieldsenseapp.com/en/>, [Online; Accessed 24 May 2019].
- FMI (2018), FMI App, en.ilmatieteenlaitos.fi/smartphones, [Online; Accessed 20 November 2018].
- Fossen, T. I., and T. Perez (2009), Kalman filtering for positioning and heading control of ships and offshore rigs, *IEEE Control Systems*, *29*(6), 32–46, doi:10.1109/MCS.2009.934408.
- Fujiwara, M. (1965), Raindrop-size Distribution from Individual Storms, *Journal of the Atmospheric Sciences*, *22*(5), 585–591, doi:10.1175/1520-0469(1965)022<0585:RSDFIS>2.0.CO;2.
- Garratt, J. (1994), Review: the atmospheric boundary layer, *Earth-Science Reviews*, *37*(1-2), 89–134, doi:10.1016/0012-8252(94)90026-4.
- Gill Instruments (2017), WindObserver 65, <http://gillinstruments.com/products/anemometer/windobserver-65.html>, [Online; Accessed 10 May 2017].
- Google Developers (2019a), Machine learning for mobile developers, <https://developers.google.com/ml-kit/>, [Online; Accessed 12 June 2019].
- Google Developers (2019b), Background Execution Limits, <https://developer.android.com/about/versions/oreo/background.html#services>, [Online; Accessed 20 May 2019].
- Google Firebase (2017), Firebase Documentation, <https://firebase.google.com/docs/>, [Online; Accessed 8 January 2019].
- Google Play (2019), Weather from DMI and YR on Android, <https://play.google.com/store/apps/details?id=dmi.byvejr.vejret&hl=en>, [Online; Accessed 23 June 2019].
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt (2011), Balance and Ensemble Kalman Filter Localization Techniques, *Monthly Weather Review*, *139*(2), 511–522, doi:10.1175/2010MWR3328.1.

- Hastie, T., R. Tibshirani, and J. Friedman (2009a), Prototype Methods and Nearest-Neighbors, in *The Elements of Statistical Learning*, pp. 1–25, Springer New York, New York, NY, doi:10.1007/b94608_13.
- Hastie, T., R. Tibshirani, and J. Friedman (2009b), Additive Models, Trees, and Related Methods, in *The Elements of Statistical Learning*, pp. 1–42, Springer New York, New York, NY, doi:10.1007/b94608_9.
- Hastie, T., R. Tibshirani, and J. Friedman (2009c), Boosting and Additive Trees, in *The Elements of Statistical Learning*, pp. 1–51, Springer New York, New York, NY, doi:10.1007/b94608_10.
- Hintz, K. S. (2015), The Sensitivity of Extreme Precipitation in Relation to Changing Time Step in the HIRLAM NWP Model, *Tech. rep.*, Danish Meteorological Institute.
- Hintz, K. S., H. Vedel, E. Kaas, and N. W. Nielsen (2019a), Estimation of wind speed and roughness length using smartphones: Method and quality assessment, *Journal of Atmospheric and Oceanic Technology*, (In-Revision).
- Hintz, K. S., H. Vedel, and E. Kaas (2019b), Collecting and Processing of Barometric Data from Smartphones for Potential Use in NWP Data Assimilation, *Meteorological Applications*, doi:10.1002/met.1805.
- Hintz, K. S., K. O’Boyle, S. L. Dance, S. Al-Ali, I. Ansper, D. Blaauboer, M. Clark, A. Cress, M. Dahoui, R. Darcy, J. Hyrkkänen, L. Isaksen, E. Kaas, U. S. Korsholm, M. Lavanant, G. Le Bloa, E. Mallet, C. McNicholas, J. Onvlee-Hooimeijer, B. Sass, V. Siirand, H. Vedel, J. A. Waller, and X. Yang (2019c), Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4–5 December 2018, *Atmospheric Science Letters*, 20(7), e921, doi:10.1002/asl.921.
- Hintz, K. S., H. Vedel, and E. Kaas (2019d), Evaluating pressure observations from private weather stations and smartphones, *Meteorological Applications*, (In-Prep.).
- Holton, J. R. (2004a), *An Introduction to Dynamic Meteorology*, vol. 88, 4th ed., Elsevier Academic Press.
- Holton, J. R. (2004b), Chapter 1 Introduction, in *An Introduction to Dynamic Meteorology*, *International Geophysics*, vol. 88, pp. 1–27, Academic Press, doi:10.1016/S0074-6142(04)80035-4.

- Houtekamer, P. L., and H. L. Mitchell (1998), Data Assimilation Using an Ensemble Kalman Filter Technique, *Monthly Weather Review*, *126*(3), 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen (2005), Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations, *Monthly Weather Review*, *133*(3), 604–620, doi:10.1175/MWR-2864.1.
- Howe, J. (2006), Crowdsourcing: A Definition, http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html, [Online; Accessed 8 August 2018].
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh (2007), Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D: Nonlinear Phenomena*, *230*(1-2), 112–126, doi:10.1016/j.physd.2006.11.008.
- iOS App Store (2019), Weather from DMI and YR on iOS, <https://apps.apple.com/dk/app/vejret-fra-dmi-og-yr/id987253904?l=en>, [Online; Accessed 23 June 2019].
- Izumi, Y. (1971), Kansas 1968 Field Program Data Report, *Tech. Rep. AFCRL-TR-72-0041*, Air Force Geophysics Laboratory.
- Jones, C. D., and B. Macpherson (1997), A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model, *Meteorological Applications*, *4*(3), 269–277, doi:10.1017/S1350482797000522.
- Kaimal, J. C., and J. C. Wyngaard (1990), The Kansas and Minnesota experiments, *Boundary-Layer Meteorology*, *50*(1-4), 31–47, doi:10.1007/BF00120517.
- Kalman, R. E. (1960), *A New Approach to Linear Filtering and Prediction Problems*, vol. 82, 35-45 pp., Transactions of the ASME–Journal of Basic Engineering.
- Kalnay, E. (2003), *Atmospheric Modelling, Data Assimilation and Predictability*, Cambridge University Press.
- Kar, S. K. (2006), A Semi-Implicit Runge–Kutta Time-Difference Scheme for the Two-Dimensional Shallow-Water Equations, *Monthly Weather Review*, *134*(10), 2916–2926, doi:10.1175/MWR3214.1.
- Kautz, T., and B. Eskofier (2015), A Robust Kalman Framework with Resampling and Optimal Smoothing, *MDPI Sensors*, *15*(3), 4975–4995, doi:10.3390/s150304975.

- Kendall, A., J. Hawke, D. Janz, P. Mazur, D. Reda, J.-M. Allen, V.-D. Lam, A. Bewley, and A. Shah (2018), Learning to Drive in a Day, *arXiv:1807.00412 [cs, stat]*, arXiv: 1807.00412.
- Kim, N.-Y., Y.-H. Kim, Y. Yoon, H.-H. Im, R. K. Y. Choi, and Y. H. Lee (2015), Correcting Air-Pressure Data Collected by MEMS Sensors in Smartphones, *Journal of Sensors*, 2015, 0–10, doi:10.1155/2015/245498.
- Kim, Y.-H., J.-H. Ha, Y. Yoon, N.-Y. Kim, H.-H. Im, S. Sim, and K. Y. R. Choi (2016), Improved Correction of Atmospheric Pressure Data Obtained by Smartphones through Machine Learning, *Journal of Sensors*, 2016, 0–12, doi:10.1155/2016/9467878.
- Kouznetsov, R. D., and S. S. Zilitinkevich (2010), On the Velocity Gradient in Stably Stratified Sheared Flows. Part 2: Observations and Models., *Boundary-Layer Meteorol*, 135, 513–517, doi:10.1007/s10546-010-9487-y.
- Krennert, T., G. Pistotnik, R. Kaltenberger, and C. Csekits (2018), Crowdsourcing of weather observations at national meteorological and hydrological services in Europe, *Advances in Science and Research*, 15, 71–76, doi:10.5194/asr-15-71-2018.
- Lahoz, W., and Q. Errera (2010), Constituent Assimilation, in *Data Assimilation*, pp. 449–490, Springer Berlin Heidelberg, doi:10.1007/978-3-540-74703-1.
- Lahoz, W., B. Khatatov, and R. Menard (Eds.) (2010), *Data Assimilation*, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-540-74703-1.
- Langleben, M. P. (1974), On wind profiles over sea ice, *Geophysical Research Letters*, 1(2), 82–85, doi:10.1029/GL001i002p00082.
- Laprise, R. (1992), The Euler Equations of Motion with Hydrostatic Pressure as an Independent Variable, *Monthly Weather Review*, 120(1), 197–207, doi:10.1175/1520-0493(1992)120<0197:TEEOMW>2.0.CO;2.
- Lonobox (2019), Lonobox, <http://lonobox.com/>, [Online; Accessed 24 May 2019].
- Lorenc, A. C. (2003), The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, 129(595), 3183–3203, doi:10.1256/qj.02.132.
- Madaus, L. E., and C. F. Mass (2017), Evaluating Smartphone Pressure Observations for Mesoscale Analyses and Forecasts, *Weather and Forecasting*, 32(2), 511–531, doi: 10.1175/WAF-D-16-0135.1.

- Marshall, J. S., and W. M. K. Palmer (1948), THE DISTRIBUTION OF RAINDROPS WITH SIZE, *Journal of Meteorology*, 5(4), 165–166, doi:10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.
- McNicholas, C., and C. Mass (2018), Smartphone Pressure Collection and Bias Correction Using Machine Learning, *J. Atmos. Oceanic Technol.*, 35, 523–540, doi:10.1175/JTECH-D-17-0096.1.
- Mesinger, F., Z. I. Janjić, S. Ničković, D. Gavrilov, and D. G. Deaven (1988), The Step-Mountain Coordinate: Model Description and Performance for Cases of Alpine Lee Cyclogenesis and for a Case of an Appalachian Redevelopment, *Monthly Weather Review*, 116(7), 1493–1518, doi:10.1175/1520-0493(1988)116<1493:TSMCMD>2.0.CO;2.
- Mittermaier, M., N. Roberts, and S. A. Thompson (2013), A long-term assessment of precipitation forecast skill using the Fractions Skill Score, *Meteorological Applications*, 20(2), 176–186, doi:10.1002/met.296.
- Monin, A. S., and A. M. Obukhov (1954), Dimensionless characteristics of turbulence in the surface layer, *Doklady Akademicheskikh Nauk SSSR*, 93, 223–226.
- Monin, A. S., and A. M. Yaglom (1971), *Statistical fluid mechanics: mechanics of turbulence*, english ed. updated, augmented and rev ed., MIT Press, Cambridge, Mass.
- Montmerle, T., Y. Michel, and B. Ménétrier (2010), Modelling of background error covariances for the analysis of clouds and precipitation, *ECMWF JCSDA Workshop*, Météo-France, CNRM-GAME, Toulouse, France.
- Morril, D. (2011), Dan Morill on GPS in smartphones, <https://www.hardwarezone.com.sg/tech-news-why-there-barometer-galaxy-nexus-google-answers>, [Online; Accessed 05 April 2019].
- Muller, C., L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R. Leigh (2015), Crowdsourcing for climate and atmospheric sciences: current status and future potential, *International Journal of Climatology*, 35(11), 3185–3203, doi:10.1002/joc.4210.
- NCEP (2018), NCEP GFS, www.emc.ncep.noaa.gov/GFS/.php, [Online; Accessed 8 December 2018].

- NCEP/EMC Model Evaluation Group (2019), FV3GFS OFFICIAL EVALUATION, <https://www.emc.ncep.noaa.gov/users/meg/fv3gfs/>, [Online; Accessed 24 June 2019].
- Netatmo SAS (2018), Netatmo Blog, <https://www.netatmo.com/blog/en/weather/understanding>, [Online; Accessed 10 September 2018].
- Netatmo SAS (2019a), Netatmo Developer Weather API, <https://dev.netatmo.com/resources/technical/reference/weatherapi/getpublicdata>, [Online; Accessed 26 March 2019].
- Netatmo SAS (2019b), Netatmo Weather Stations, <https://www.netatmo.com/weather>, [Online; Accessed 17 June 2019].
- Nichols, N. K. (2010), Mathematical Concepts of Data Assimilation, in *Data Assimilation*, pp. 13–40, Springer Berlin Heidelberg, doi:10.1007/978-3-540-74703-1.
- Nielsen, N. W. (2010), A new scheme for parameterization of turbulent surface fluxes in the stably stratified atmospheric boundary layer, *Scientific Report 10-01*, Danish Meteorological Institute.
- NSSL (2014), mPING, <https://mping.nssl.noaa.gov/>, [Online; Accessed 14 September 2018].
- Nurmi, P. (2003), Recommendations on the verification of local weather forecasts., *ECMWF Technical Memoranda*, (430), 19, doi:10.21957/y1z1thg5l.
- Olsen, B., U. Korsholm, C. Petersen, N. Nielsen, B. Sass, and H. Vedel (2015), On the performance of the new NWP nowcasting system at the Danish Meteorological Institute during a heavy rain period, *Meteorol. Atmospheric Phys*, (127), 519–535, doi:10.1007/s00703-015-0388-y.
- Overeem, A., J. C. R. Robinson, H. Leijnse, G. J. Steeneveld, B. K. P. Horn, and R. Uijlenhoet (2013), Crowdsourcing urban air temperatures from smartphone battery temperatures, *Geophysical Research Letters*, 40(15), 4081–4085, doi:10.1002/grl.50786.
- Panofsky, H. A. (1963), Determination of stress from wind and temperature measurements, *Q.J.R. Meteorol. Soc.*, 89, 85–94, doi:10.1002/qj.49708937906.
- Patric, B. (2014), HP scaling out for the Internet of Things, <https://community.hpe.com/t5/Grouped-in-the-Cloud/HP-scaling-out-for-the-Internet-of-Things/ba-p/6636688#.WZWmA3cjGjR>, [Online; accessed 17 August 2017].

- Paulson, C. A. (1970a), The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer, *J. Appl. Meteor.*, *9*, 857–861, doi:10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2.
- Paulson, C. A. (1970b), The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer, *J. Appl. Meteor.*, *9*, 857–861, doi:10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2.
- Petch, J. C., A. R. Brown, and M. E. B. Gray (2002), The impact of horizontal resolution on the simulations of convective development over land, *Quarterly Journal of the Royal Meteorological Society*, *128*(584), 2031–2044, doi:10.1256/003590002320603511.
- Phillips, N. A. (1957), A COORDINATE SYSTEM HAVING SOME SPECIAL ADVANTAGES FOR NUMERICAL FORECASTING, *Journal of Meteorology*, *14*(2), 184–185, doi:10.1175/1520-0469(1957)014<0184:ACSHSS>2.0.CO;2.
- Plantsnap (2019), Plantsnap website, <https://www.plantsnap.com>, [Online; Accessed 7 June 2019].
- Price, C., R. Maor, and H. Shachaf (2018), Using smartphones for monitoring atmospheric tides, *Journal of Atmospheric and Solar-Terrestrial Physics*, *174*, 1–4, doi:10.1016/j.jastp.2018.04.015.
- PySerial (2018), PySerial Documentation, <https://pythonhosted.org/pyserial/>, [Online; accessed 10-November-2018].
- Rainwise Inc. (2019), Website of Rainwise, <https://rainwise.com/>, [Online; Accessed 13 June 2019].
- Robert, A. (1981), A stable numerical integration scheme for the primitive meteorological equations, *Atmosphere-Ocean*, *19*(1), 35–46, doi:10.1080/07055900.1981.9649098.
- Roberts, N. M., and H. W. Lean (2008), Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, *Monthly Weather Review*, *136*(1), 78–97, doi:10.1175/2007MWR2123.1.
- Robustelli, U., V. Baiocchi, and G. Pugliano (2019), Assessment of Dual Frequency GNSS Observations from a Xiaomi Mi 8 Android Smartphone and Positioning Performance Analysis, *Electronics*, *8*(1), 91, doi:10.3390/electronics8010091.
- Sass, B. H. (2002), A research version of the STRACO cloud scheme, *Tech. rep.*, Danish Meteorological Institute.

- Schonberger, V., and K. Cukier (2013), *Big Data: A Revolution That Will Transform How We Live, Work and Think*, Hodder Export.
- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson (2011), The AROME-France Convective-Scale Operational Model, *Monthly Weather Review*, 139(3), 976–991, doi:10.1175/2010MWR3425.1.
- Silva, J., C. Ribeiro, and R. Guedes (2007), Roughness Length Classification of CORINE Land Cover Classes, *Tech. rep.*, MEGAJOULE-Consulting, Portugal.
- Simmons, A., and R. Strüfing (1981), An energy and angular momentum conserving finite-difference scheme hybrid coordinates and medium range weather forecasts, (28), 68.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers (2008), A Description of the Advanced Research WRF Version 3, *NCAR Tech. Note NCAR/TN-475+STR*, p. 113, doi:10.5065/D68S4MVH.
- Skok, G., and N. Roberts (2016), Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts, *Quarterly Journal of the Royal Meteorological Society*, 142(700), 2599–2610, doi:10.1002/qj.2849.
- Sparkfun (2017), Sparkfun CAN-BUS shield, <https://www.sparkfun.com/products/13262>, [Online; Accessed 15 October 2017].
- StatCounter Global Stats (2018), Mobile OS Marketshare, <http://gs.statcounter.com/os-market-share/mobile/europe/2018>, [Online; Accessed 05 April 2019].
- Statistics Denmark (2017), HOME APPLIANCES, <http://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/home-appliances>, [Online; Accessed 22 August 2017].
- Stull, R. B. (Ed.) (1988), *An Introduction to Boundary Layer Meteorology*, Springer Netherlands, doi:10.1007/978-94-009-3027-8.
- Suryana, J. (2017), The Rainfall Intensity Effects on 1–13 GHz UWB-Based 5G System for Outdoor Applications, *Wireless Communications and Mobile Computing*, 2017, 1–13, doi:10.1155/2017/6495145.
- Talagrand, O. (2010), Variational Assimilation, in *Data Assimilation*, pp. 41–67, Springer Berlin Heidelberg, doi:10.1007/978-3-540-74703-1.

- Taylor, J. R. (1997), *An introduction to error analysis: the study of uncertainties in physical measurements*, 2nd ed., University Science Books, Sausalito, Calif.
- Tennekes, H. (1973), The Logarithmic Wind Profile, *Journal of the Atmospheric Sciences*, 30(2), 234–238, doi:10.1175/1520-0469(1973)030<0234:TLWP>2.0.CO;2.
- Tumolo, G., and L. Bonaventura (2015), A semi-implicit, semi-Lagrangian discontinuous Galerkin framework for adaptive numerical weather prediction: SISL-DG Framework for Adaptive NWP, *Quarterly Journal of the Royal Meteorological Society*, 141(692), 2582–2601, doi:10.1002/qj.2544.
- Ukkonen, P., A. Manzato, and A. Mäkelä (2017), Evaluation of Thunderstorm Predictors for Finland Using Reanalyses and Neural Networks, *Journal of Applied Meteorology and Climatology*, 56(8), 2335–2352, doi:10.1175/JAMC-D-16-0361.1.
- Vaavud (2018), Vaavud Website, <http://vaavud.com/>, [Online; Accessed 2 December 2018].
- Vaisala (2018), BAROCAP Digital Barometer PTB330, <https://www.vaisala.com/sites/default/files/documents/PTB330-Datasheet-B210708EN-F.pdf>, [Online; Accessed 25 June 2018].
- van Leeuwen, P. J. (2009), Particle Filtering in Geophysical Systems, *Monthly Weather Review*, 137(12), 4089–4114, doi:10.1175/2009MWR2835.1.
- Vedel, H., and X. Yang (1999), HIRVDA About the New Data Assimilation System for HIRLAM, *Tech. rep.*, Danish Meteorological Institute.
- Wieringa, J. (1993), Representative Roughness Parameters for Homogeneous Terrain, *Boundary-Layer Meteorol*, 63, 323–363, doi:10.1007/BF00705357.
- Wilks, D. (2011), Parametric Probability Distributions, in *International Geophysics*, vol. 100, pp. 71–131, Elsevier, doi:10.1016/B978-0-12-385022-5.00004-X.
- WMO (2014), *Guide to Meteorological Instruments and Methods of Observation*, 8th ed., World Meteorological Organization.
- Wyngaard, J. C. (2010), *Turbulence in the atmosphere*, Cambridge, UK ; New York.
- Yang, X., B. Stig, M. Dahldom, B. Sass, S. Zhuang, B. Amstrup, C. Petersen, K. Nielsen, N. W. Nielsen, and A. Mahura (2017), NEA, the Operational Implementation of HARMONIE 40h1.1 at DMI, pp. 104–111.

- Zhang, L., J. Gong, and R. Wang (2018), Diagnostic Analysis of Various Observation Impacts in the 3DVAR Assimilation System of Global GRAPES, *Monthly Weather Review*, *146*(10), 3125–3142, doi:10.1175/MWR-D-17-0182.1.
- Zinevich, A., H. Messer, and P. Alpert (2009), Frontal Rainfall Observation by a Commercial Microwave Communication Network, *Journal of Applied Meteorology and Climatology*, *48*(7), 1317–1334, doi:10.1175/2008JAMC2014.1.

List of Figures

2.1	The Vaavud anemometers, Mjolnir (left) and Sleipnir (middle and right). The Sleipnir anemometer also measures the wind direction due to its asymmetric rotor design.	7
2.2	A UI on top of PMOB to access raw data in real time. Left: sensor values from barometer, accelerometer and location services. Right: Timeseries of pressure while the app is open.	13
2.3	Timeseries of the pressure measured by the DMI operated SYNOP station in Kastруп, Denmark (Black solid line) compared to observations from different smartphones within 10 km from the station (coloured circles).	14
2.4	Pressure observations from two individual smartphones for one month. The ability to track individual phones are what makes the data personal. The details of the map are deliberately coarse to prevent identification possibilities. The colour bar shows the pressure in hPa.	15
3.1	Illustration of the double penalty problem. Blue circles represent observations of precipitation. Red squares represent the model forecast of precipitation in grid-point space. Black squares represent the grid-boxes. The horizontal resolution is 3 km and 6 km for the left and middle figure, respectively. The right figure is an illustration of the Fractional Skill Score (FSS).	23
4.1	Model domain of the Vaavud Frigg-WRF model. The inner red square shows the inner domain.	28
7.1	Picture of the extendable transportable mast. On top the Gill Ultrasonic Anemometer is placed.	44

7.2	Real-time wind measurements from the web-application during one minute 7th February 2018 at 20:32:33 UTC to 20:33:33 UTC. The white curve shows the measured wind speed with a frequency of 10 Hz, given on the left axis. The orange and blue line shows the mean over 3 s and 60 s respectively. The red dots show the wind direction given on the right axis. On the internal webpage the plot updates with a frequency of 10 Hz.	45
8.1	Left: Satellite imagery of most of the island of Amager. Centre: Adjusted wind speed forecast using roughness lengths from a wind atlas. Right: Direct diagnosed 10 m wind speed forecast from the Vaavud NWP system. The wind speed is given in m/s and is shown in the colorbar.	49
10.1	Forecast cycle of the DMI HARMONIE model. A new forecast is initiated every 3 hours from a new analysis \mathbf{x}_a which is produced by the previous forecast, \mathbf{x}_b and an innovation term \mathbf{d} (see chapter 9).	71
10.2	Model domain of the DKA model shown by the grey shaded area. This figure is also included as supplemental material to Hintz et al. (2019b).	72
10.3	Outline of the observation pre-processing chain in HARMONIE. SPOs enters the system after the Oulan routine.	73
10.4	Single Observation experiment: The difference between to analyses valid at 2018-01-04 12 UTC is shown. In one analysis, a single additional pressure observation was included. Colorbar is in units of hPa.	77
12.1	Conceptual sketch of a decision tree. T and F stands for <i>true</i> and <i>false</i> respectively. The black circles represents the leaf nodes of the tree.	87
13.1	SPOs from December 2018. Colours show the background deviation. Left: Untreated SPOs (RMSE=2.3 hPa). Right: SPOs corrected by a prediction from the nearest neighbour model (N=3), using latitude, longitude, pressure and the standard deviation of the pressure as features (RMSE=1.3 hPa).	91
13.2	Root Mean Square Error of SPO error predictions for the test batch of BL_NC (red), BL_MC (blue), ML_KN (green) and ML_GB (black). μ is the average, \tilde{x} is the median and σ is the standard deviation of the distribution. See text for details.	92

13.3	Bias of SPO error predictions for the test batch of BL_NC (red), BL_MC (blue), ML_KN (green) and ML_GB (black). μ is the average, \tilde{x} is the median and σ is the standard deviation of the distribution. See text for details.	93
13.4	Standard deviation of SPO error predictions for the test batch of BL_NC (red), BL_MC (blue), ML_KN (green) and ML_GB (black). μ is the average, \tilde{x} is the median and σ is the standard deviation of the distribution. See text for details.	94
13.5	Difference between SPOs and NWP surface pressure. Left column: Raw SPOs. Right Column: SPOs corrected by the predicted error using K-Nearest-Neighbour method. Each row is 15 minutes. The time interval is seen in left-column plots. The date is 12 Dec 2018. Only smartphones with over 50 observations in total has been used.	95
13.6	Difference between SPOs and NWP surface pressure. Left column: Raw SPOs. Right Column: SPOs corrected by the predicted error using a Gradient Boost model. Each row is 15 minutes. The time interval is seen in left-column plots. The date is 12 Dec 2018. Only smartphones with over 50 observations in total has been used.	96
14.1	Spin-up by the barometric sensor in an iPhone 7 (Bosch BMP280). Each coloured curve is a single measurement series over 180 s, subtracted by its average value. The full black line shows the average of 50 measurements series.	99
14.2	Distribution of SPOs entering the database as a function of time of day during the period 4th of April 2018 to 24th of May 2018. The black vertical dashed line shows the approximate time of sunrise in Denmark.	100
A1	Smartphone wind observations taken near Brøndby stadium in Copenhagen, Denmark. While the difference in wind speeds (red circles) at the stadium can be interesting for athletes, the observations represent a very high representativeness error, which makes utilizing the observations in numerical weather prediction difficult.	217
A2	Absolute terrain height difference between the Danish Terrain Model and the Terrain of the DMI HARMONIE model. The resolution of the terrain model is upscaled to that of the DMI HARMONIE model. Units are in meters. The DMI HARMONIE model has an overall bias of 1.11 m compared to the terrain model.	218

-
- A3 Example of a data stream of pressure measured by a smartphone on board on a train. The green arrow shows the position of a tunnel where it is seen that the position of the smartphone is not well defined. In general, changes in altitude is clearly seen. The colours are given in hPa in the colorbar. 219
- A4 Precipitation 15th of March 2019 at 19 UTC. Left: Radar Reflectivity over Europe from the OPERA project. Right: Netatmo stations with a pluviometer installed. Colours represent rain over one hour. White stations (white circles) have not measured rain. 219

List of Tables

2.1	Overview of collected smartphone observations from 4th of April 2018 to 4th of April 2019. The decline in April 2019 is because only four days are included from this month.	11
2.2	Overview of observed variables collected from the smartphones. The calculated variables are calculated either on the smartphone directly or server-side when the observation has been received. σ represents the standard deviation.	12
4.1	An overview of the two Amazon instances used at Vaavud. Frigg is the primary compute instance used to run WRF.	27

List of blocks

2.1	Example of initialising the iOS SDK using Swift 3.0 for retrieving pressure observations. CMAltimeter is the class which is called from the iOS SDK. See text for details.	11
7.1	Example of a function to retrieve data from Firebase using the JavaScript SDK. The ‘snapshot’ listen for changes and return new values to ‘firobj’.	46
10.1	Minimal example of a working OBSOUL file with a single observation. Comments are written in red. See text for details.	74
10.2	Minimal working example of a blacklisting file. Comments are written in red. See text for details.	75
D.1	The Fortran namelist file (namelist.input) for the Frigg-WRF setup.	223

List of publications

Hintz et al. (2019a)

Hintz, K. S., Vedel, H., Kaas, E., Nielsen, N. W., (2019a, in-Revision), Estimation of wind speed and roughness length using smartphones: Method and quality assessment. *Journal of Atmospheric and Oceanic Technology of the American Meteorological Society*.

Hintz et al. (2019b)

Hintz, K. S., Vedel, H., Kaas, E., (2019b) Collecting and Processing of Barometric Data from Smartphones for Potential Use in NWP Data Assimilation. *Meteorological Applications*, doi:10.1002/met.1805.

Hintz et al. (2019c)

K. S. Hintz, K. O'Boyle, S. L. Dance, S. Al Ali, I. Ansper, D. Blaauboer, M. Clark, A. Cress, M. Dahoui, R. Darcy, J. Hyrkkänen, L. Isaksen, E. Kaas, U. S. Korsholm, M. Lavanant, G. Le Bloa, E. Mallet, C. McNicholas, J. Onvlee-Hooimeijer, B. Sass, V. Siirand, H. Vedel, J. A. Waller, X. Yang., (2019c), Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018. *Atmospheric Science Letters of the Royal Meteorological Society*, doi:10.1002/asl.921.

Hintz et al. (2019d)

K. S. Hintz, H. Vedel, E. Kaas. (2019d, In-Prep.), Evaluating pressure observations from private weather stations and smartphones. *In-Prep. for Meteorological Applications*.

Publications

Publication A

Title: Estimation of wind speed and roughness length using smartphones: Method and quality assessment.

Submitted: March 2019.

Authors: K. S. Hintz, H. Vedel, E. Kaas and N. W. Nielsen.

Journal: Journal of Atmospheric and Oceanic Technology (JTECH).

Journal Website: <https://journals.ametsoc.org/toc/atot/current>

Status: In Revision

1 **Estimation of wind speed and roughness length using smartphones: Method**
2 **and quality assessment**

3 K. S. Hintz*

4 *Danish Meteorological Institute, Copenhagen, Denmark*

5 H. Vedel

6 *Danish Meteorological Institute, Copenhagen, Denmark*

7 E. Kaas

8 *The Niels Bohr Institute, University of Copenhagen, Denmark*

9 N. W. Nielsen

10 *Danish Meteorological Institute, Copenhagen, Denmark*

11 *Corresponding author address: K. S. Hintz, Danish Meteorological Institute, Lyngbyvej 100, 2100

12 Copenhagen, Denmark.

13 E-mail: kah@dmi.dk

ABSTRACT

14 Crowdsourced data is now seen as a potential source of high-resolution ob-
15 servations in the atmospheric sciences and operational meteorology. In this
16 paper, the potential applications of wind observations obtained with handheld
17 anemometers connected to smartphones are investigated. The quality of hand-
18 held wind measurements relative to traditional wind measurements is exam-
19 ined. Comparisons to professional grade SYNOP stations are performed. It
20 is shown that even raw handheld wind measurements, before extrapolating to
21 10 m height, can in some cases be more representative than SYNOP stations
22 only about a kilometer away. Secondly, a method of determining roughness
23 lengths based on the turbulent intensity derived from the high-frequency raw
24 signal of handheld smartphone wind measurements is examined. Roughness
25 length is an essential parameter in numerical weather prediction for the de-
26 termination of surface winds and surface fluxes, but they are often poorly
27 determined. Further, roughness lengths are necessary when correcting near-
28 surface wind observations for height offsets. A series of field experiments
29 were carried out, after which the handheld wind measurements were extrap-
30 olated with height to 10 m, using roughness lengths of different origin as a
31 parameter. This enabled a comparison of the quality of the derived rough-
32 ness length to traditional sources. Under certain circumstances, the roughness
33 lengths obtained with the approach presented here give better results than us-
34 ing traditional sources. This study opens up new possibilities for the use of
35 crowdsourced data in meteorology.

36 **1. Introduction**

37 The spatial resolution of Numerical Weather Prediction (NWP) models increases steadily,
38 whereas the conventional meteorological observing network is not densifying at the same rate,
39 if at all. This results in a lack of observations to validate, fine-tune, and initialize NWP models.
40 A potential new source of observational data is crowdsourced data from devices owned by the
41 public. To cite Howe (2006), crowdsourced data can be defined broadly as *'the act of taking a*
42 *job traditionally performed by a designated agent (usually an employee) and outsourcing it to an*
43 *undefined, generally large group of people in the form of an open call.'* In addition to the usage
44 in NWP, crowdsourced data can provide an increase of the information available about the state
45 of the atmosphere to operational meteorologists (Clark et al. 2018).

46
47 In recent years the number of low-cost personal devices connected to the internet and mobile
48 network has increased substantially (Evans 2011), potentially providing new observations for use
49 in the atmospheric sciences. These sensors can either run fully automated, such as barometric
50 sensors in smartphones (Kim et al. 2015), or they need to be operated by humans, such as in
51 the Meteorological Phenomena Identification Near the Ground (mPING) (NSSL 2016) project.
52 Barometric pressure obtained via smartphones has been suggested by Mass and Madaus (2014)
53 as a data source to improve forecasts of mesoscale phenomena. The Weather Observation
54 Website (WOW) project is looking into collecting data from Personal Weather Stations (PWS)
55 as additional data (UK Met Office 2017). The privately owned company Netatmo SAS (2017)
56 manufactures personal weather stations and collects data from all over the world and have
57 partnered with the Norwegian meteorological service, YR (Netatmo SAS 2018).

58

59 Obtaining Handheld Wind Measurements (HWMs) has recently become possible with low-cost
60 anemometers that can be attached to smartphones. In this study, we investigate the usefulness
61 of such observations, made with a particular cup anemometer for smartphones manufactured by
62 Vaavud ApS. Other manufacturers exist, such as WeatherFlow Inc. The Vaavud anemometer
63 measures both wind speed and wind direction, due to its asymmetric rotor design. It is a low-cost
64 anemometer, easily affordable for ‘the crowd’. There are apparent pitfalls of these data which
65 must be addressed, related both to the quality of the measurement devices themselves, and related
66 to the non-ideal behavior of the user, who might stand next to an obstacle, not hold it at the right
67 angle, or even have fun blowing on the anemometer. Such problems are common to all types
68 of crowdsourced data, but if these problems can be solved, there will be access to an enormous
69 amount of additional meteorological observations at a low cost.

70
71 The objective of this study is two-fold. First, and most importantly, to quantify the quality of
72 HWMs coming directly from the handheld anemometer compared to traditional wind observations
73 following the requirements of WMO (2008, ch. 5). As more and more observations from smart-
74 phones can be sent in real or near real time, it must be considered if such observations can be used
75 to increase the information about specific weather phenomena for operational meteorologists. To
76 do so, it is critical first to investigate how HWMs compares to traditional wind observations. It
77 is also examined if extrapolating the HWMs with height to 10 m gives better results compared
78 to the WMO (2008, ch. 5) standards. This requires knowledge of z_0 , the roughness length of
79 the surface, thus, the second objective of the present study is therefore to investigate whether
80 HWMs can be used to estimate local roughness lengths by utilizing the turbulent intensity as
81 suggested by Counihan (1975). Counihan (1975) gives a relation between the roughness length
82 and the horizontal turbulent intensity by which high-frequency smartphone wind measurements

83 potentially might be used to improve the roughness lengths used in NWP. Such an approach is
84 tested by using the derived roughness length as a parameter in an extrapolation of the HWM with
85 height to compare to a professional high-precision reference measurement at 10 m. Further, by
86 using this approach we focus on methods that can be easily implemented in practice so there will
87 no need for professional equipment, which is usually too expensive for ‘the crowd’. In this way, it
88 is the hope that the method presented in this study is scalable and easy to implement in practice;
89 thus, opening for possibilities for new observation types for operational weather forecasting and
90 NWP models.

91
92 For this study the majority of the measurements were performed by ourselves. In a future study
93 more measurements made by the public will be considered. Some of the measurement campaigns
94 were carried out not far from two SYNOP stations operated by the Danish Meteorological Institute
95 (DMI) in order to compare with the official DMI wind measurements. An additional reason for the
96 focus on roughness length is that it is of vital importance to NWP since roughness length is closely
97 related to the estimation of near-surface winds and surface fluxes of sensible heat, moisture, and
98 momentum (Arya 2001).

99 **2. Method**

100 The basis of this study is a set of measurement sessions made at different locations (see section
101 2.B). In each session, two wind measuring devices, a handheld anemometer and a reference high
102 precision anemometer, were used to measure wind speed and wind direction during the same
103 period and in the near vicinity of one another. The reference measurement measured in 10 m
104 and served two purposes. First of all, and most importantly, as a traditional, continuous wind
105 measurement allowing a direct comparison to both the HWM and nearby DMI SYNOP stations.

106 Secondly, to investigate whether the use of roughness lengths derived from the HWMs leads to
107 more accurate HWM based estimates of 10 m wind speeds than using roughness lengths obtained
108 via other sources. Extrapolating the HWM to the height of the reference observation thus gives
109 an estimate of the quality of the derived roughness length from the HWM. The extrapolation
110 is performed from about 2 m to 10 m both With Stability Correction (WSC) and No Stability
111 Correction (NSC) using stability parameters from a short-term NWP model forecast. This is
112 elaborated further on in section 3. The methods are below divided into four sub-sections, the
113 wind measurement equipment (section 2.A), the location of measurement sites (section 2.B), the
114 NWP reference (section 2.C) and the surface characteristics (section 2.D) represented by different
115 sources of roughness lengths.

116

117 *A. Wind measurement equipment*

118 The HWMs was obtained with a Vaavud Sleipnir anemometer (Vaavud 2015) mounted on
119 a smartphone together with the Vaavud app. The anemometer measures with a frequency of
120 approximately 4 Hz, and is designed to minimize the influence of flow distortion from the
121 smartphone¹. The 4 Hz output from the Vaavud app is considered the measurement signal in this
122 study. According to Vaavud, the anemometer is specified to work in the wind speed range of 2
123 to 40 m/s. Therefore it should be used with caution in low wind speeds. Latitude and longitude
124 are stored, based on the built-in smartphone Global Navigation Satellite System (GNSS) receiver
125 position estimate. The GNSS receiver also returns altitude, but a comparison of 237 altitude
126 estimates obtained in one session with the smartphone at a fixed height, showed that GNSS
127 altitude is currently not accurate enough for our purpose (the estimates have a standard deviation

¹Personal communication with the inventor of the device, Andreas Okholm, former CTO of Vaavud ApS.

128 of 20 m). Instead, the height of the device above the ground was determined with a tape measure.
129 It is noted though, that the accuracy of GNSS altitudes is believed to improve significantly in
130 future (Robustelli et al. 2019). The standard duration of the HWM used here is 30 s, and the
131 time of the measurement is defined at the time at the beginning of the measurement. An HWM
132 was obtained approximately every 2 min. Afterwards, the individual HWMs were binned into
133 intervals of 10 minutes for which the mean was computed.

134

135 The reference measurements were obtained with a high precision ultrasonic anemometer (Gill
136 Instruments 2017) placed 10 m above ground, on a thin mast stabilized against bending and
137 undulation by three wires running from the top of the mast to spears in the ground. 10 m is the
138 height that WMO (2008, ch. 5) suggest for traditional wind observations. The sampling frequency
139 of the ultrasonic device was 1 Hz. The accuracy is given by Gill Instruments (2017) as $\pm 2\%$ at 12
140 m/s. In one experiment, the ultrasonic anemometer was placed at the same height as the HWM
141 instrument to compare the two anemometers directly.

142

143 The minimum amount of available metadata from the smartphone for an HWM to be considered
144 valid was in this study defined as

- 145 1. The duration and measurement time of the HWM
- 146 2. The latitude, longitude and the measurement height above ground
- 147 3. A device identification number

148 Furthermore, additionally metadata was collected to be used for comparison to the roughness
149 length derived from the HWMs and examination of the effects of stability correction:

- 150 1. The characteristics of the surface, represented by the roughness lengths from two different
151 sources, used for comparison to the roughness length determined via the HWMs. In addition
152 the surface was characterised subjectively by the authors.
- 153 2. The lower tropospheric atmospheric stability (obtained here via an NWP model, see section
154 2.C), used to examine the effects of stability corrections when extrapolating the HWM with
155 height.

156 *B. Location of measurement sites*

157 In total, eight measurement sessions were carried out at three different locations, namely at
158 Hvide Sande (HVDS), Griben (GNB) and Eskebjerg (ESKB), all in Denmark. Figure 1 shows
159 satellite imagery of each location, plotted with the location of the reference measurement (red
160 star) and the DMI SYNOP station (blue star). The locations with coordinates are listed in table
161 1. Also, a summary of the prevailing weather conditions for each measurement session is given
162 in table 2. HVDS is a beach location in western Denmark at the North Sea. The north sea is to
163 the west of HVDS and to the east there is a fjord. The DMI station (HVDS DMI) is categorized
164 as a coastal station (Cappelen 2012). However, HVDS DMI is located approximately 1.7 km
165 inland, downstream of the prevailing wind direction (westerly onshore winds) in a small city.
166 The reference measurement was performed on the beach and upstream of the city, in westerly,
167 south-westerly winds. GNB is a coastal location in central Denmark, also categorized as a coastal
168 station (Cappelen 2012). The associated DMI SYNOP station (GNB DMI) is located about
169 1 km North-west of GNB, 200 m from the coastline on a narrow, long peninsula, in a fenced
170 military facility. The HWMs were done within a few meters of the coastline in conditions with
171 north-easterly onshore winds. GNB DMI is not placed optimally. Only 16 m south of the station is
172 a lattice tower installed with a height of 54 m. To the east is a row of 2-floor buildings, the closest

173 one 44 m away, and to the North, is a row of small buildings. There are no obstacles to the west.
174 However, there is a steep hillside with a height of ≈ 30 m leading up to the GNB DMI station about
175 80 m away. There is a slope all around the station with the most prominent to the west. ESKB is
176 an inland location in eastern Denmark, about 4 km from the nearest coastline. ESKB can be con-
177 sidered open, agricultural terrain. The measurements were performed in wintertime when there
178 were no crops in the fields. The nearest SYNOP station to ESKB is Holbaek, 20 km to the east
179 and therefore the Holbaek DMI SYNOP cannot be seen in figure 1. The Holbaek DMI SYNOP is
180 set up at a good location in open terrain with no disturbing objects close to the wind measurement.
181 We are considering the area of ESKB semi-smooth for most wind directions, with a few isolated
182 obstacles. ESKB is a generally open area with mainly agricultural fields and a few farms. 1.2
183 km to the west there are a few houses, and there is 3.6 km to the closest coastline to the north-west.

184

185 In all sessions but one, all smartphone measurements were performed by the first author of
186 the present paper to control the measurement setup better. In the HVDS session, measurements
187 from kitesurfers standing on the beach nearby the reference anemometer are included. The height
188 above ground of the surfer measurements is assumed to be 2 m. The DMI SYNOP data are used
189 for comparison and as an indicative, subjective measure of representativeness, an essential aspect
190 when validating NWP models.

191

192 *C. NWP reference*

193 To examine if extrapolation of the HWMs was improved by using stability correction or not (see
194 section 3) stability parameters were obtained from a short-term NWP model forecast using the

195 WRF model (Skamarock et al. 2008, WRF V.3.9.1). From the WRF model stability parameters
196 such as the kinematic heat flux H and the Monin-Obukhov length L (see section 3) was obtained.

197 The model was run with a horizontal resolution of 3 km, and 30 vertical levels run over the
198 period of the measurements. The model was initiated from analysis data from NCEP (NCEP/FNL
199 2000) every 6 hours. The surface layer scheme used was the MM5 similarity scheme based on the
200 Monin-Obukhov similarity theory (Jiménez et al. 2012). The model was initially stabilized with
201 a digital filter (Lynch and Huang 1993) to prevent excitation of fast inertia-gravity waves and to
202 reduce spin-up time.

203 *D. Surface characteristics*

204 The characteristics of the surface were determined both quantitatively and subjectively by the
205 authors. Most importantly, the roughness length is determined from the HWM using the measured
206 horizontal turbulence intensity. The roughness lengths derived from the HWMs are denoted z_0^{hwm} .
207 The theory and uncertainties behind z_0^{hwm} are elaborated in section 3.

208 The NWP model is also used as a source of roughness lengths and is denoted z_0^{nwp} . In the
209 NWP model z_0^{nwp} is derived from a table of land use category, with a resolution of 30 arcseconds
210 (approximately 1 km) with a summer and winter value, upscaled to the grid-box resolution.

211 The third and last source of roughness lengths is obtained from the work of Silva et al. (2007),
212 who converted the Corine Land Cover (CLC) 2000, Version 18.5 (Buttner and Kosztra 2001)
213 dataset, with a resolution of 100 m, to a dataset of roughness lengths, here denoted z_0^{clc} .

214
215 The extrapolated HWMs are denoted based on the source of roughness length which have been
216 used as u_{nwp}^{10m} , u_{clc}^{10m} and u_{hwm}^{10m} . The HWMs that have not been extrapolated are denoted u_{hwm}^s using
217 s for surface. The reference is denoted u_{ref}^{10m} and the wind observations from the DMI SYNOP

218 stations are denoted u_{dmi}^{10m} . The raw HWMs that have not been averaged over time are denoted
219 u_{raw} .

220 3. Theory

221 The roughness length can be estimated from wind measurements by considering the horizontal
222 turbulent intensity (Counihan 1975). The idea is that for a given wind speed, a rough surface
223 will generate more turbulence near the surface than will a smooth surface. The roughness length
224 derived by this approach is defined as

$$z_0^{hwm} = A \cdot z \cdot \exp(-u(z)/\sigma_u), \quad (1)$$

225 following Counihan (1975). Here σ_u is the standard deviation of the wind measurement, $u(z)$ is
226 the mean wind speed over an appropriate averaging time at height z and $A = \sigma_u/u_* \cdot k$, where u_* is
227 the friction velocity and k is the von-Kàrmàn constant ($k = 0.4$). $\sigma_u/u(z)$ represents the horizontal
228 turbulent intensity. Equation (1) is derived from the logarithmic wind profile (see equation (4))
229 based on the assumption that $\sigma_u/u_* = 2.5$ (Arya 1995) such that $A = 1$. The measurement height
230 of a handheld anemometer of about 2 meters will often be influenced by upstream local effects,
231 so this approach can only be assumed valid for smooth terrain ($z_0 < 0.1$ m). Also, equation (1)
232 implies that for a constant roughness length, the ratio $u(z)/\sigma_u$ must also be constant since A and
233 the measurement height z are constants. The ratio $u(z)/\sigma_u$ has been smoothed by convolving the
234 time series with a Gaussian kernel using a standard deviation of 5 so that weights are decreased
235 further away from the target value.

236

237 As previously described, to compare the HWMs to the reference and to validate the derived
238 roughness lengths, the HWMs were extrapolated to a height of 10 m for comparison with the refer-

239 ence 10 m measurements. The wind speed increases considerably with height near the surface due
 240 to surface friction and hence z_0 is an essential parameter for extrapolation of wind measurements.
 241 Therefore, by extrapolating the HWMs, one can test the quality of the roughness lengths. To ex-
 242 trapolate the HWMs two approaches were used based on the logarithmic wind profile law with and
 243 without stability correction (named WSC and NSC respectively). Based on the Monin-Obukhov
 244 similarity theory (Monin and Obukhov 1954), it follows that the non-dimensional vertical mean
 245 wind speed gradient ($\phi(z/L)$) is given by

$$\phi\left(\frac{z}{L}\right) = \frac{kz}{u_*} \frac{\partial u}{\partial z}, \quad (2)$$

246 where z is the height above the surface and L is the Monin-Obukhov length given by $L =$
 247 $-u_*^3 c_p \rho T / kgH$, where $H/\rho c_p$ is the kinematic heat flux, g is the gravitational acceleration, and T
 248 is the absolute temperature. L and H was determined from the NWP model described in section 2.
 249 $\phi(z/L)$ has been determined empirically for unstable ($z/L < 0$) and stable ($z/L > 0$) conditions as
 250 (Businger et al. 1971)

$$\phi(z/L) = \begin{cases} 1 + \beta_1 \frac{z}{L} & \text{if } z/L \geq 0 \\ (1 - \gamma_1 z/L)^{-1/4} & \text{if } z/L < 0, \end{cases} \quad (3)$$

251 where β_1 and γ_1 are empirical constants. Discrepancies do exist in the chosen values for these
 252 constants. Here $\beta_1 = 5$ is used based on the argumentation of supercritical flux Richardson
 253 numbers by Kouznetsov and Zilitinkevich (2010), and $\gamma_1 = 15$ following Arya (2001).

254

255 Integrating equation (2) from z_0 to z , using $u(z_0) = 0$, yields (Panofsky 1963)

$$u(z) = \frac{u_*}{k} \left[\ln \frac{z}{z_0} - \psi(z/L) \right], \quad (4)$$

256 where

$$\psi(z/L) = \int_{z_0}^z (1 - \phi(z/L)) dz.$$

257 For $z/L = 0$, which is neutral conditions, equation (4) reduces to the logarithmic wind profile law
 258 ($\phi(z/L) = 1$, e.g. the case of NSC), which can typically be obtained in cases with strong winds and
 259 overcast skies. Panofsky (1963) and Paulson (1970) gives the solution for $\psi(z/L)$ for the stable
 260 ($z/L \geq 0$) and unstable ($z/L < 0$) case respectively as

$$\psi(z/L) = \begin{cases} -\beta_1 \frac{z}{L} & \text{for } z/L \geq 0 \\ \ln \left[\left(\frac{1+x}{2} \right)^2 \left(\frac{1+x^2}{2} \right) \right] - 2 \tan^{-1}(x) + \pi/2 & \text{for } z/L < 0, \end{cases} \quad (5)$$

261 where $x = (1 - \gamma_1 z/L)^{1/4}$. Now let $u(z_2)$ be the extrapolated wind measurement and let $u(z_1)$ be
 262 the HWM, where $z_2 > z_1$. Assuming that z_2 and z_1 are located in the same Internal Boundary
 263 Layer (IBL) then one can solve for $u(z_2)$ by using equation (4) to obtain an expression for $u(z_1)$
 264 and $u(z_2)$ and dividing them with each other to obtain

$$u(z_2) = \frac{\ln z_2/z_0 - \psi_2(z_2/L)}{\ln z_1/z_0 - \psi_1(z_1/L)} u(z_1), \quad (6)$$

265 using equation (5) for the stable and unstable cases respectively. Equation (6) then represents an
 266 approach to extrapolate a wind measurement in the surface layer, correcting for stability (WSC),
 267 using z_0 as a parameter. Setting $\psi = 0$ corresponds to the case of NSC.

268

269 **4. Results**

270 First, as mentioned in section 2, the handheld anemometer and the ultrasonic anemometer was
 271 in one experiment placed at the same height to compare the two anemometers directly. They were
 272 placed in 2.2 m, 1 m apart in the horizontal at the ESKB location. Both devices were measuring
 273 continuous for 8 hours. The mean wind speed in the period from the ultrasonic anemometer
 274 was 4.00 m/s. A net positive bias of 0.05 m/s was found in that case relative to the ultrasonic
 275 anemometer used as a reference measurement. Also, z_0^{hwm} was in this case calculated based

276 on both the handheld anemometer and the ultrasonic reference measurement. For the handheld
277 device, $z_0^{hwm}=1.5 \cdot 10^{-2} \text{ m} \pm 0.3 \cdot 10^{-2} \text{ m}$, and for the reference measurement, $z_0^{hwm}=1.0 \cdot 10^{-2} \text{ m}$
278 $\pm 0.2 \cdot 10^{-2} \text{ m}$ where the uncertainty is given as one standard deviation of the mean.

279

280 Figure 2 shows the time series of wind measurements for ESKB4 starting on March 3rd, 05:30
281 UTC. As described, the mean HWMs was extrapolated to 10 m using roughness lengths of
282 different origin, z_0^{nwp} , z_0^{clc} and z_0^{hwm} , all with NSC. For this case $z_0^{nwp} = 0.06 \text{ m}$, $z_0^{clc} = 0.10 \text{ m}$ and
283 $z_0^{hwm} = 3.3 \cdot 10^{-3} \text{ m}$ (see Table 3). Note that z_0^{clc} is constant for each location. The bias of the
284 extrapolated wind speed relative to the reference anemometer is lowest for u_{hwm}^{10m} NSC with 0.27
285 m/s. The difference in bias between using stability correction (WSC) or not (NSC) for u_{hwm}^{10m} is
286 only 0.05 m/s for this case. For u_{nwp}^{10m} NSC the net bias is 1.57 m/s and for u_{clc}^{10m} NSC the net bias is
287 2.08 m/s. In this case, u_{hwm}^{10m} performs best both regarding bias and RMSE for both NSC and WSC.

288

289 One example where a large difference of using stability correction (WSC) or not (NSC) is
290 shown in figure 3 for u_{nwp}^{10m} in the case of ESKB5. The extrapolations both used $z_0^{nwp} = 0.05 \text{ m}$.
291 The RMSE for u_{nwp}^{10m} WSC and NSC was 1.03 m/s and 1.43 m/s respectively. However, no clear
292 trend in the data of RMSE with regard of using NSC or WSC can be seen in general, other than
293 that the difference of using stability correction or not is relatively small (see table 3). In the cases
294 of ESKB1, ESKB3 and ESKB4 the RMSE is higher WSC than NSC for u_{nwp}^{10m} . The same is true
295 for the bias.

296

297 Figure 4 shows the measurement session for GNB. It is seen that u_{clc}^{10m} is performing worse than
298 both u_{hwm}^{10m} and u_{nwp}^{10m} . RMSE and bias for u_{hwm}^{10m} improves WSC compared to NSC, decreasing from
299 0.87 m/s and 0.43 m/s respectively to 0.74 m/s and 0.13 m/s (see table 3). In general, all errors

300 decreases when correcting for stability in the case of GNB.

301

302 Figure 5 shows observations from HVDS in connection with a kitesurfing competition from
303 September 2016. Here also measurements from ordinary users of the handheld anemometer have
304 been included, such that measurements from many different smartphones and different Vaavud
305 anemometers are mixed. One notice is that the HWMs appears more irregular compared to
306 figure 2 and figure 3. All measurements are taken on-shore spreading 200 m north and south
307 of the reference 10 m measurement, within 50 meters from the sea, and with onshore wind. In
308 this case, u_{hwm}^{10m} has the highest RMSE in total with 1.49 m/s with NSC. The DMI observations
309 are also shown in figure 5. The RMSE of u_{dmi}^{10m} is 3.00 m/s almost 2 m/s higher than for the HWMs.

310

311 It was shown in section 3 that for z_0^{hwm} to be constant the ratio $u(z)/\sigma_u$ must also be constant.
312 $u(z)/\sigma_u$ is plotted with wind speed and wind direction in figure 6 for ESKB5. This plot was
313 chosen because during this session the biggest changes in wind conditions was seen. The wind
314 speed increased from about 3 m/s to 7 m/s; also a small change in wind direction is seen. The
315 ratio $u(z)/\sigma_u$ is relatively constant during the increase and decrease in wind speed. However,
316 there is a large spread in especially the handheld measurements. Convolving the time series with
317 a Gaussian kernel (black lines in figure 6) helps to define a general trend in both the case of the
318 reference and the handheld measurements. The standard deviation for all $u(z)/\sigma_u$ -ratios is shown
319 in the right-most column in table 3.

320

321 The RMSE and bias relative to the 10 m reference measurement, u_{ref}^{10m} , are calculated for each
322 method on each location. The results are shown in figure 7 and in table 3. The biases of u_{hwm}^s
323 are all negative. This is because u_{hwm}^s is measured closer to the surface than the reference, hence

324 the wind speed is in general lower. The difference in RMSE and bias between HWMs and
 325 the extrapolations using roughness lengths of different origin as a parameter are all due to the
 326 extrapolation to the reference height of 10 m. u_{dmi}^{10m} in figure 7 (red bars) shows the RMSE and
 327 bias from the nearest DMI SYNOP station relative to u_{ref}^{10m} . Recall that for the ESKB location the
 328 nearest DMI station is about 20 km away, whereas for GNB and HVDS they are of the order 1
 329 km away. For both GNB and HVDS, the RMSE and bias for both the extrapolated handheld 10
 330 m wind estimates and u_{hwm}^s , are all lower than for u_{dmi}^{10m} . Also, it is seen that u_{clc}^{10m} WSC and NSC
 331 gives worse results than u_{hwm}^s and u_{dmi}^{10m} . Overall, u_{hwm}^{10m} has both the lowest RMSE and bias.

332

333 5. Discussion

334 It is found that the stability correction does not play a vital role in any of the cases studied here,
 335 and it has no clear trends when considering the RMSE of NSC and WSC in table 3. For the cases
 336 considered here, it thus shows that the error one makes in assuming neutral conditions is minor.
 337 It must be emphasized that the stability correction is prone to errors of the numerical forecast
 338 model, which can alter the interpretation of the results. Because of the small net bias between the
 339 HWMs and the reference measurement presented in the results, the bias between the HWMs and
 340 the reference measurement is not considered a significant source of error.

341

342 For the cases of HVDS and GNB, it was found that HWMs are likely to contribute positively
 343 to the information of the state of the surface wind field. For HVDS, the difference between u_{dmi}^{10m}
 344 and u_{ref}^{10m} are seen in figure 5. Here the HWMs are found to be more representative for the coast
 345 than the DMI measurement, even for u_{hwm}^s . This is seen directly in table 3, where u_{hwm}^s have both
 346 a lower RMSE and bias than the DMI measurement. It is very likely that these differences come

347 from the fact that the measurements done at the beach are taken in the maritime IBL, while the
348 DMI observation is measuring in a growing IBL of the land and the city.

349

350 It is seen from figure 4 that there are also differences for the GNB session between the HWMs
351 and DMI. u_{dmi}^{10m} is overall measuring wind speeds lower than the reference measurement, u_{ref}^{10m} .
352 The DMI station and the reference measurement are approximately 1 km apart (see figure 1). The
353 DMI SYNOP station is 100 m away from the coastline, whereas the reference measurement was
354 only 5 m away from the coastline, although we note here, that these measurements were likely
355 impacted by beach ridges. For easterly wind as in this case, it is seen that u_{dmi}^{10m} is measuring lower
356 wind speeds in general. We claim, based on the RMSE and bias given in table 3, that the HWMs
357 (both at 2 m and extrapolated to 10 m) are more representative of the general wind conditions in
358 the area than u_{dmi}^{10m} in this situation.

359

360 Due to the non-complex terrain of HVDS and GNB, it is safe to assume the surface roughness
361 of water is the correct value in these cases. z_0^{hwm} is a little too high for HVDS, but in this case, real
362 user data was included which will introduce a source of unknown errors. However, it is noted that
363 HVDS is one of the best locations for including real user data when westerly winds are present,
364 as measurements are likely to be done in the maritime boundary layer. For HVDS the effect of the
365 upstream city is clearly seen in the u_{dmi}^{10m} measurement. It ought to bring up some considerations
366 at DMI if the HVDS station should be considered as a coastal station in the future.

367

368 It is seen that the RMSE of u_{clc}^{10m} is performing worst for the sites investigated. This is most
369 noticeable in figure 4, where u_{clc}^{10m} is overshooting the reference measurement by 2-4 m/s in
370 general. The values of z_0 for GNB (table 3) explain this though. z_0^{nwp} is using a roughness

371 for water which is almost correct in this case, while z_0^{clc} is using a typical value of farmland
 372 ($z_0^{clc} = 0.2$ m) (Arya 2001, fig. 10.5). The resolution of the dataset of z_0^{clc} is much higher than
 373 z_0^{nwp} . This also makes it more vulnerable in areas with discontinuities in surface roughness
 374 if the upstream surface roughness is not accounted for. At GNB, the measurements were
 375 performed at the beach, so close to the coastline that the measurements will not be far from
 376 representing the IBL of the sea. This can be accounted for by taking the wind direction into
 377 account and hence using the upwind surface roughness. This must also take the upwind wind fetch
 378 into account (Wieringa 1993). An upwind value from z_0^{clc} gives 10^{-4} m comparable to that of z_0^{nwp} .

379
 380 z_0^{nwp} cannot necessarily be expected to be representative of the location where a measurement
 381 takes place as it is upscaled to the resolution of the model (see section 2), neither is wind direction
 382 taken into account when determining z_0^{nwp} . This might explain why u_{nwp}^{10m} , in general, performs
 383 worse than u_{hwm}^{10m} for the ESKB location. This is both the case with NSC and WSC.

384
 385 It is emphasized that the logarithmic wind profile is not valid down to $z = 0$ m, but only in an
 386 inertial sublayer where $z/z_0 \rightarrow \infty$ and $zf/u_* \rightarrow 0$ (where f is the Coriolis parameter) (Tennekes
 387 1973). Wieringa (1993) argues that the lower limit of validity is $z \approx 20 \cdot z_0$ to $50 \cdot z_0$ and the
 388 upper bound is $z \approx 0.1h$ to $0.2h$, where h is the boundary layer height. For ESKB using a value
 389 of $z_0^{clc} = 0.1$ m, the lower limit is $z \approx 2$ m to 5 m. Using the lowest value of z_0^{nwp} for ESKB,
 390 $z_0^{nwp} = 5.0 \cdot 10^{-2}$ m, one gets the lower limit as $z \approx 1.0$ m to 2.5 m. It is evident from this that
 391 estimating the roughness length from HWMs faces two main challenges. First, the measurement
 392 height will most often be in the interval 2.0 m to 2.2 m, making it clear that using HWMs for
 393 quantifying the current wind conditions and for estimating roughness length should only be done
 394 with utmost caution for areas for which $z_0 > 0.1$ m. Therefore, a background field is needed to

395 determine if an HWM can be trusted or not. Such a background field could potentially come from
396 a land surface model. This limits the validity of HWMs in general to smooth, open areas. The
397 general user of the handheld anemometer used in this study is sailors, kite surfers, and farmers, so
398 most measurements will meet this requirement. Secondly, the duration of an HWM will rarely be
399 10 min as suggested by WMO (2008). Standing with an arm raised for 10 min with a handheld
400 anemometer is not realistic. However, users can be encouraged to take several measurements
401 within 10 min as was done in this study.

402

403 Considering the required wind fetch given by Wieringa (1993, eq.6) for ESKB, the fetch for
404 $z = 10$ m is 1.2 km to 1.9 km using the lowest and highest value of z_0 for ESKB respectively. For
405 $z = 2$ m the corresponding values are 0.2 km and 0.3 km. It is, however, only to the west we find
406 an area that can be considered a significant change in effective roughness length within this range
407 (based on the CLC dataset (Silva et al. 2007)).

408 z_0 can also be determined by linear regression of u on $\ln z$. The order matters as the errors of
409 $\ln z$ are usually much lower than those of u (Bergeron and Abrahams 1992). This approach was
410 not used in this work though. We expect such roughness lengths to be unreliable in this study for
411 several reasons. First, only two levels are considered here. While assuming a neutral stratification
412 is enough to perform a linear regression, even small errors in wind measurements can result in
413 substantial errors in u_* of about $\pm 15\%$ when only using two levels (Langleben 1974), and the
414 errors in z_0 is a factor of ku/u_* larger (Wieringa 1993). Furthermore, it has no clear, practical
415 implications, and so has little value for the objective of this study.

416

417 The ratio $u(z)/\sigma_u$ must be constant for the roughness length to be constant, as shown with
418 equation (1). The ratio in the case of ESKB5 was plotted in figure 6, together with the wind

419 speed and wind direction. The spread of the ratio from the individual HWMs are quite large, as
420 one would suspect. However, when convoluted with a Gaussian kernel the spread of both the
421 reference measurements and the HWMs becomes much smaller. It can not be argued that the
422 ratios are constant, but based on figure 6 one can easily imagine how more HWMs can contribute
423 to finding the true roughness length. A change in wind direction is also seen in figure 6 which
424 unavoidably will add some noise to the measured ratio except if the new upstream surface area
425 is uniform within the wind directions experienced. Table 3 gives the average of the ratio and the
426 standard deviation of the ratio $u(z)/\sigma_u$ for each case. Not surprisingly, the standard deviation for
427 u_{hwm}^s are higher than u_{ref}^{10m} . The absolute values of $u(z)/\sigma_u$ for both u_{hwm}^s and u_{ref}^{10m} in the ESKB
428 cases are comparable too.

429
430 The RMSE and bias for each case and location are shown in figure 7. Is it expected that for
431 ESKB, the bias for the DMI SYNOP station is in some cases high in absolute terms because the
432 DMI station is about 20 km away. The bias of HWMs is generally lower, indicating that for this
433 area HWMs can be considered to be more representative than the better equipt, but more distant
434 DMI station. This is also seen from the RMSE. It is worth noticing that an HWM can be more
435 representative than a traditional SYNOP observation, given that it is taken in an open area. Based
436 on the results shown in section 4 it is not recommended to use the Silva et al. (2007) dataset for
437 roughness lengths as both u_{clc}^{10m} WSC and NSC gives higher RMSE and bias than u_{hwm}^s .

438
439 The use of smartphones can contribute to a dataset of surface roughness. As previously seen
440 with z_0^{clc} , the roughness length does in reality depends on conditions in an area, in particular the
441 upstream area, making z_0 vary with wind direction. Therefore, handheld devices that cannot mea-
442 sure the wind direction should not be used for this purpose; or only used for homogeneous terrain,

443 where the wind direction can be estimated from numerical models or traditional measurements.

444

445 More importantly, HWMs can be of a good enough quality to be incorporated into an overview
446 of the state of the atmosphere. This does not necessarily mean assimilating such observations into
447 NWP models, but they can provide complementary aid for the operational meteorologists. This is
448 especially true if the HWMs are extrapolated using a source of z_0 of good quality, as both RMSE
449 and bias decreases. This can only be considered valid for smooth ($z_0 < 0.1$ m) surfaces. Correction
450 for stability was not found to be the highest concern for the cases considered here; thus, there will
451 in practical applications be little or no need for NWP model output making the application much
452 simpler to implement. Filtering of poor observations from real user data will probably be the most
453 challenging task onwards.

454 **6. Conclusion**

455 It has been shown that Handheld Wind Measurements (HWMs) obtained with smartphones are
456 in some cases more representative than wind measurements from even relatively nearby official
457 weather stations. Hence, utilizing HWMs has the potential to aid operational meteorologists in
458 obtaining a near real-time overview of the state of the atmosphere and bias-correcting of NWP
459 surface winds. We find smaller errors in wind speeds obtained from HWMs, even before applying
460 an adjustment for the height offsets, than those from official weather stations about a kilometer
461 away. One advantage of HWMs is that they are not limited by the need for continuous electricity
462 supply and on-going maintenance. Extrapolating the HWMs using roughness lengths of different
463 origin to the standard height of 10 m generally yields smaller errors than the raw HWM, both
464 regarding RMSE and bias. It is shown that using the roughness length estimated from the signal
465 from a handheld anemometer can be used to extrapolate an HWM from about 2 m to a height of

466 10 m, and these generally give better results than using roughness lengths from an NWP model
467 and a look-up table, respectively. Furthermore, it is concluded that HWMs should not be used to
468 extract information from areas with roughness lengths above 0.1 m, and thus a background field
469 is required. Utilizing HWMs does indeed have a future potential to be applied at meteorological
470 centers, though it will most likely not happen in the near future, as work with validation of real
471 user data is required before such can be applied. Future work will include real, distributed user
472 data to get a mapping of surface roughness and wind atlas, to be compared with an accepted source
473 of data.

474 *Acknowledgments.* We would like to thank Innovation Fund Denmark for funding of this study
475 and agriculturalist Poul Graae for unlimited access to his farmland, and Dr. Juan Munoz-Gomez
476 for valuable technical advice.

477 **References**

- 478 Arya, S. P., 1995: *Atmospheric Boundary Layer and Its Parameterization*, 41–66. Springer Nether-
479 lands, Dordrecht, doi:10.1007/978-94-017-3686-2_3.
- 480 Arya, S. P., 2001: *Introduction to Micrometeorology*. 2nd ed., Cambridge University Press.
- 481 Bergeron, N. E., and A. D. Abrahams, 1992: Estimating shear velocity and roughness length from
482 velocity profiles. *Water Resour. Res.*, **28**, 2155—2158, doi:10.1029/92WR00897.
- 483 Businger, J. A., J. C. Wyngaard, and Y. I. and E. F. Bradley, 1971: Flux-profile relationships
484 in the atmospheric surface layer. *J. Atmos. Sci.*, **28**, 181—189, doi:10.1175/1520-0469(1971)
485 028<0181:FPRITA>2.0.CO;2.
- 486 Buttner, G., and B. Kosztra, 2001: *Manual of CORINE Land Cover changes*. European Environ-
487 ment Agency.

488 Cappelen, J., 2012: Vindstatistik for danske kyststationer 2001-2010 - hyppighed af stiv kuling og
489 derover. *Technical Report, Danish Meteorological Institute*, 6–7.

490 Clark, M. R., J. D. C. Webb, and P. J. Kirk, 2018: Fine-scale analysis of a severe hailstorm using
491 crowd-sourced and conventional observations. *Meteorological Applications*, **25** (3), 472–492,
492 doi:10.1002/met.1715, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1715>.

493 Counihan, J., 1975: Adiabatic atmospheric boundary layers: A review and analysis of the data
494 from the period 1880-1972. *Atmospheric Environment*, **9**, 871–905, doi:10.1016/0004-6981(75)
495 90088-8.

496 Evans, D., 2011: How the Next Evolution of the Internet Is Changing Everything. Tech. rep.,
497 Cisco.

498 Gill Instruments, 2017: Windobserver 65. [Online; Accessed 10 May 2017], <http://gillinstruments.com/products/anemometer/windobserver-65.html>.

500 Howe, J., 2006: Crowdsourcing: A definition. [Online; Accessed 7 October 2016], http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html.

502 Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-
503 Bustamante, 2012: A Revised Scheme for the WRF Surface Layer Formulation. *Monthly*
504 *Weather Review*, **140** (3), 898–918, doi:10.1175/MWR-D-11-00056.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-11-00056.1>.

506 Kim, N.-Y., Y.-H. Kim, Y. Yoon, H.-H. Im, R. K. Y. Choi, and Y. H. Lee, 2015: Correcting
507 air-pressure data collected by mems sensors in smartphones. *Journal of Sensors*, **2015**, 0–10,
508 doi:10.1155/2015/245498.

509 Kouznetsov, R. D., and S. S. Zilitinkevich, 2010: On the velocity gradient in stably stratified
510 sheared flows. part 2: Observations and models. *Boundary-Layer Meteorol*, **135**, 513–517, doi:
511 10.1007/s10546-010-9487-y.

512 Langleben, M. P., 1974: On wind profiles over sea ice. *Geophys. Res. Lett.*, **1**, 82–85, doi:10.1029/
513 GL001i002p00082.

514 Lynch, P., and X. Huang, 1993: Diabatic initialization using recursive filters. *Tellus A*, **46**, 583–
515 597, doi:10.1034/j.1600-0870.1994.t01-4-00003.x.

516 Mass, C. F., and L. E. Madaus, 2014: Surface pressure observations from smartphones: A potential
517 revolution for high-resolution weather prediction? *Bull. Amer. Meteor. Soc.*, **95**, 1343–1349,
518 doi:10.1175/BAMS-D-13-00188.1.

519 Monin, A. S., and A. M. Obukhov, 1954: Dimensionless characteristics of turbulence in the surface
520 layer. *Doklady Akademicheskikh Nauk SSSR*, **93**, 223–226.

521 NCEP/FNL, 2000: Ncep fnl operational model global tropospheric analyses. ncar computational
522 and information systems laboratory research data archive. [Online; Accessed 23 January 2017],
523 doi:10.5065/D6M043C6.

524 Netatmo SAS, 2017: Netatmo weathermap. [Online; Accessed 10 September 2017], <https://weathermap.netatmo.com/>.
525

526 Netatmo SAS, 2018: Netatmo blog. URL [https://www.netatmo.com/blog/en/weather/](https://www.netatmo.com/blog/en/weather/understanding)
527 [understanding](https://www.netatmo.com/blog/en/weather/understanding), [Online; Accessed 10 September 2018], [https://www.netatmo.com/blog/](https://www.netatmo.com/blog/en/weather/understanding)
528 [en/weather/understanding](https://www.netatmo.com/blog/en/weather/understanding).

529 NSSL, 2016: mping crowdsourcing weather reports. [Online; Accessed 20 March 2017], <http://mping.nssl.noaa.gov>.
530

- 531 Panofsky, H. A., 1963: Determination of stress from wind and temperature measurements. *Q.J.R.*
532 *Meteorol. Soc.*, **89**, 85–94, doi:10.1002/qj.49708937906.
- 533 Paulson, C. A., 1970: The mathematical representation of wind speed and temperature pro-
534 files in the unstable atmospheric surface layer. *J. Appl. Meteor.*, **9**, 857–861, doi:10.1175/
535 1520-0450(1970)009<0857:TMROWS>2.0.CO;2.
- 536 Robustelli, U., V. Baiocchi, and G. Pugliano, 2019: Assessment of Dual Frequency GNSS Ob-
537 servations from a Xiaomi Mi 8 Android Smartphone and Positioning Performance Analysis.
538 *Electronics*, **8** (1), 91, doi:10.3390/electronics8010091.
- 539 Silva, J., C. Ribeiro, and R. Guedes, 2007: Roughness length classification of corine land cover
540 classes. *MEGAJoule-Consulting: Portugal*.
- 541 Skamarock, W. C., and Coauthors, 2008: A description of the advanced research wrf version 3.
542 *NCAR Tech. Note NCARTN-475+STR*, 113, doi:10.5065/D68S4MVH.
- 543 Tennekes, H. H., 1973: The logarithmic wind profile. *J. Atmos. Sci.*, **30**, 234–238, doi:10.1175/
544 1520-0469(1973)030<0234:TLWP>2.0.CO;2.
- 545 UK Met Office, 2017: Ukmo weather observations website. [Online; Accessed 20 August 2017],
546 <https://www.metoffice.gov.uk/>.
- 547 Vaavud, 2015: Vaavud sleipnir anemometer. [Online; Accessed 1 July 2019], [http://vaavud.com/
548 blog/tech/so-how-exactly-do-they-work](http://vaavud.com/blog/tech/so-how-exactly-do-they-work).
- 549 Wieringa, J., 1993: Representative roughness parameters for homogeneous terrain. *Boundary-*
550 *Layer Meteorol*, **63**, 323–363, doi:10.1007/BF00705357.
- 551 WMO, 2008: *Guide to Meteorological Instruments and Methods of Observation*. 7th ed., World
552 Meteorological Organization.

553 **LIST OF TABLES**

554 **Table 1.** Location of DMI SYNOP stations and sessions. 27

555 **Table 2.** Summarised weather conditions for each of the eight measurement sessions.
556 T2m is the 2 meter temperature from the nearest DMI SYNOP station. The
557 wind conditions are summarised from the reference measurement. Cloud cover
558 is given in okta when relevant. 28

559 **Table 3.** Table of Root Mean Square (RMS) errors and bias relative to the reference
560 observation at 10 m. HVDS: Hvide Sande, GNB: Gniben, ESKB: Eskebjerg.
561 WSC: With Stability correction. NSC: No Stability correction. All numbers
562 are in units of m/s unless otherwise is stated. z_0 of u_{ref}^{10m} is the roughness length
563 obtained from the reference observation using the same method as with z_0^{hvm} 29

TABLE 1. Location of DMI SYNOP stations and sessions.

Location	Latitude	Longitude	Number of sessions
Hvide Sande (HVDS)	55.9951	8.1205	1
Gniben (GNB)	56.0004	11.2884	1
Eskebjerg (ESKB)	55.7037	11.2945	6
HVDS DMI station	56.0078	8.1413	-
GNB DMI station	56.0072	11.2800	-
Holbaek DMI station	55.7358	11.6035	-

564 TABLE 2. Summarised weather conditions for each of the eight measurement sessions. T2m is the 2 meter
 565 temperature from the nearest DMI SYNOP station. The wind conditions are summarised from the reference
 566 measurement. Cloud cover is given in okta when relevant.

	Wind speed	Wind direction	T2m	Conditions
ESKB1	2-5 m/s	S-SW	7-9 °C	Haze, then clear
ESKB2	6-11 m/s	NW	5-6 °C	Scattered low clouds, 4/8
ESKB3	4-6 m/s	SW	3-4 °C	Few low clouds, 1/8
ESKB4	4-8 m/s	SW	4-6 °C	Scattered clouds, 3/8
ESKB5	4-7 m/s	NE-NEE	6-9 °C	Clear
ESKB6	4-8 m/s	SW-W	5-12 °C	Fog, then clear
GNB	8-11 m/s	E-SE	10-11 °C	Overcast, 8/8, Drizzle
HVDS	6-11 m/s	SW	16-17 °C	Few clouds, 2/8

567 TABLE 3. Table of Root Mean Square (RMS) errors and bias relative to the reference observation at 10
568 m. HVDS: Hvide Sande, GNB: Gniben, ESKB: Eskebjerg. WSC: With Stability correction. NSC: No Stability
569 correction. All numbers are in units of m/s unless otherwise is stated. z_0 of u_{ref}^{10m} is the roughness length obtained
570 from the reference observation using the same method as with z_0^{hwm} .

ESKB1	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]		RMS	BIAS	$\overline{(u/\sigma_u)}$	$\sigma(u/\sigma_u)$
u_{nvp}^{10m}	0.78	1.18	0.29	0.84	0.086	u_{dmi}^{10m}	0.73	-0.47	-	-
u_{clc}^{10m}	0.82	1.25	0.35	0.90	0.100	u_{hwm}^s	0.89	-0.73	7.57	1.28
u_{hwm}^{10m}	0.61	0.68	-0.15	0.19	0.008	u_{ref}^{10m}	-	-	6.48	0.26
u_{ref}^{10m}	-	-	-	-	0.031					
ESKB2	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]		RMS	BIAS	$\overline{(u/\sigma_u)}$	$\sigma(u/\sigma_u)$
u_{nvp}^{10m}	2.06	1.97	1.66	1.57	0.050	u_{dmi}^{10m}	1.96	1.68	-	-
u_{clc}^{10m}	2.92	2.80	2.58	2.48	0.100	u_{hwm}^s	2.05	-1.78	4.97	0.52
u_{hwm}^{10m}	1.48	1.42	0.97	0.90	0.022	u_{ref}^{10m}	-	-	5.97	0.35
u_{ref}^{10m}	-	-	-	-	0.038					
ESKB3	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]		RMS	BIAS	$\overline{(u/\sigma_u)}$	$\sigma(u/\sigma_u)$
u_{nvp}^{10m}	1.05	1.3	0.76	1.03	0.050	u_{dmi}^{10m}	4.42	4.26	-	-
u_{clc}^{10m}	1.51	1.84	1.26	1.59	0.100	u_{hwm}^s	1.24	-1.14	6.51	0.52
u_{hwm}^{10m}	0.67	0.82	0.26	0.47	0.017	u_{ref}^{10m}	-	-	6.20	0.47
u_{ref}^{10m}	-	-	-	-	0.026					
ESKB4	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]		RMS	BIAS	$\overline{(u/\sigma_u)}$	$\sigma(u/\sigma_u)$
u_{nvp}^{10m}	1.86	1.92	1.57	1.66	0.060	u_{dmi}^{10m}	0.79	0.01	-	-
u_{clc}^{10m}	2.36	2.43	2.08	2.18	0.100	u_{hwm}^s	1.27	-1.12	7.37	0.40
u_{hwm}^{10m}	0.81	0.82	0.27	0.32	0.033	u_{ref}^{10m}	-	-	5.79	0.20
u_{ref}^{10m}	-	-	-	-	0.035					
ESKB5	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]		RMS	BIAS	$\overline{(u/\sigma_u)}$	$\sigma(u/\sigma_u)$
u_{nvp}^{10m}	1.43	1.03	1.21	0.88	0.050	u_{dmi}^{10m}	1.96	-1.7	-	-
u_{clc}^{10m}	1.78	1.34	1.56	1.19	0.100	u_{hwm}^s	0.68	-0.61	9.81	1.02
u_{hwm}^{10m}	0.59	0.38	0.34	0.14	0.020	u_{ref}^{10m}	-	-	8.09	0.91
u_{ref}^{10m}	-	-	-	-	0.015					

ESKB6	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]	RMS	BIAS	$\overline{(\mathbf{u}/\sigma_{\mathbf{u}})}$	$\sigma(\mathbf{u}/\sigma_{\mathbf{u}})$	
u_{nwp}^{10m}	2.01	1.79	1.80	1.59	0.080	u_{dmi}^{10m}	4.57	4.38	-	-
u_{clc}^{10m}	2.20	1.97	1.99	1.76	0.100	u_{hwm}^s	0.83	-0.69	8.02	0.49
u_{hwm}^{10m}	0.88	0.78	0.57	0.45	0.003	u_{ref}^{10m}	-	-	5.95	0.35
u_{ref}^{10m}	-	-	-	-	0.034					
GNB	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]	RMS	BIAS	$\overline{(\mathbf{u}/\sigma_{\mathbf{u}})}$	$\sigma(\mathbf{u}/\sigma_{\mathbf{u}})$	
u_{nwp}^{10m}	0.73	0.71	0.12	-0.13	0.0003	u_{dmi}^{10m}	2.58	-2.03	-	-
u_{clc}^{10m}	4.48	3.68	4.23	3.44	0.200	u_{hwm}^s	1.51	-1.36	9.28	1.26
u_{hwm}^{10m}	0.87	0.74	0.43	0.13	0.001	u_{ref}^{10m}	-	-	10.75	0.40
u_{ref}^{10m}	-	-	-	-	0.0001					
HVDS	RMS NSC	RMS WSC	BIAS NSC	BIAS WSC	z_0 [m]	RMS	BIAS	$\overline{(\mathbf{u}/\sigma_{\mathbf{u}})}$	$\sigma(\mathbf{u}/\sigma_{\mathbf{u}})$	
u_{nwp}^{10m}	0.67	0.60	0.38	0.25	0.0001	u_{dmi}^{10m}	3.00	-2.42	-	-
u_{clc}^{10m}	0.71	0.64	0.45	0.32	0.0003	u_{hwm}^s	1.06	-0.91	9.97	0.80
u_{hwm}^{10m}	1.49	1.31	1.34	1.13	0.010	u_{ref}^{10m}	-	-	11.76	0.63
u_{ref}^{10m}	-	-	-	-	0.0009					

571 **LIST OF FIGURES**

572 **Fig. 1.** Satellite imagery of the locations of the measurements sessions. Left: Hvide Sande (HVDS),
573 Center: Gniben (GNB), Right: Eskebjerg (ESKB). The red star shows the location of the
574 reference measurement and the blue star shows the location of the DMI SYNOP stations. 32

575 **Fig. 2.** Time series of wind measurements at Eskebjerg starting 2017 March 3rd, 05:30 UTC
576 [ESKB4]. Abscissa: Hours since the beginning of measurements. Ordinate: Wind speed
577 [m/s]. Black crosses shows individual handheld measurements, u_{raw} and solid black circles
578 shows the average within 10 minutes, u_{hwm}^s . The grey shaded area shows the standard deviation
579 of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Holbaek. Purple broken curve: u_{clc}^{10m}
580 NSC. Green broken curve: u_{hwm}^{10m} NSC. Black broken curve: u_{nwp}^{10m} NSC. 33

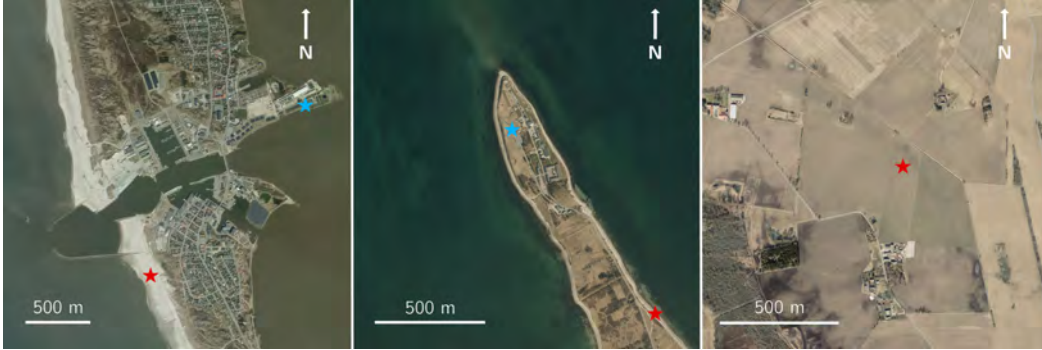
581 **Fig. 3.** Time series of wind measurements at Eskebjerg starting at 2017 March 23rd, 08:20 UTC
582 [ESKB5]. Abscissa: Hours since the beginning of measurements. Ordinate: Wind speed
583 [m/s]. Black crosses shows individual handheld measurements, u_{raw} and solid black circles
584 shows the average within 10 minutes, u_{hwm}^s . The grey shaded area shows the standard deviation
585 of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Holbaek. Black curve: u_{ref}^{10m} . Black
586 dashed curves: u_{nwp}^{10m} NSC and WSC (see legend). 34

587 **Fig. 4.** Time series of wind measurements at Gniben starting 2016 October 13th, 08:00 UTC
588 [GNB]. Abscissa: Hours since the beginning of measurements. Ordinate: Wind speed [m/s].
589 Black crosses shows individual handheld measurements, u_{raw} and solid black circles shows
590 the average within 10 minutes, u_{hwm}^s . The grey shaded area shows the standard deviation of
591 u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Gniben. Purple broken curve: u_{clc}^{10m} NSC.
592 Green broken curve: u_{hwm}^{10m} NSC. Black broken curve: u_{nwp}^{10m} NSC. 35

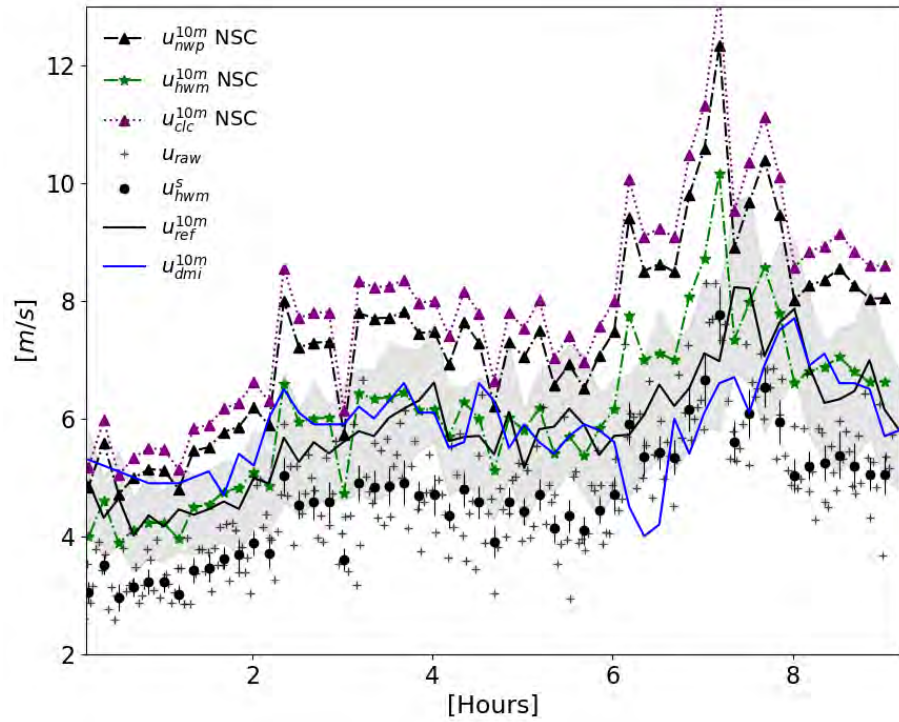
593 **Fig. 5.** Time series of wind measurements at Hvide Sande starting at 2016 September 3rd, 07:10
594 UTC [HVDS]. Abscissa: Hours since the beginning of measurements. Ordinate: Wind
595 speed [m/s]. Black crosses shows individual handheld measurements, u_{raw} and solid black
596 circles shows the average within 10 minutes, u_{hwm}^s . The grey shaded area shows the standard
597 deviation of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Hvide Sande. Purple broken
598 curve: u_{clc}^{10m} NSC. Green broken curve: u_{hwm}^{10m} NSC. Black broken curve: u_{nwp}^{10m} NSC. 36

599 **Fig. 6.** Time series of wind measurements at Eskebjerg starting at 2017 March 23rd, 08:20 UTC
600 [ESKB5]. Abscissa: Hours since the beginning of measurements. Ordinate left: Ratio of
601 $u(z)/\sigma_u$. Ordinate right: Wind speed (blue) and wind direction (red) from 10 m reference
602 observations. Black dots and crosses shows the ratio $u(z)/\sigma_u$ from the reference measure-
603 ments and the handheld measurements respectively. The full black line and dashed black
604 line shows the convolved measurements using a Gaussian filter, from the reference and the
605 handheld measurements respectively. 37

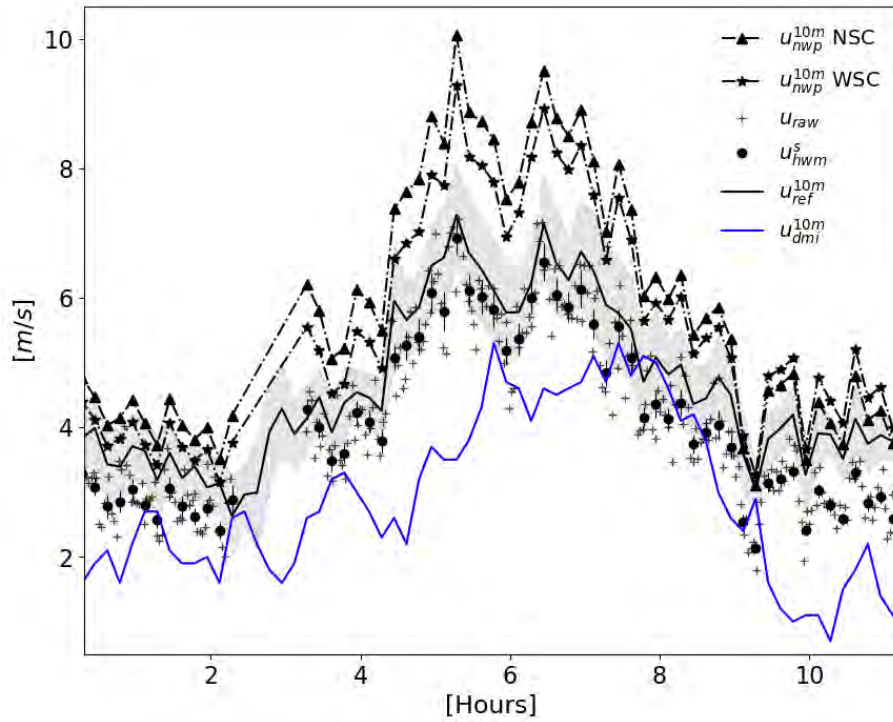
606 **Fig. 7.** Abscissa: Name of location. Lower ordinate: RMSE, upper ordinate: Bias. The ‘truth’
607 is the reference observation at 10 m, u_{ref}^{10m} . u_{dmi}^{10m} (blue) and u_{hwm}^s (orange) are the nearest
608 SYNOP station and the handheld observations, respectively. The rest, named in the legend
609 on top, is the extrapolated winds using roughness lengths of different origin. 38



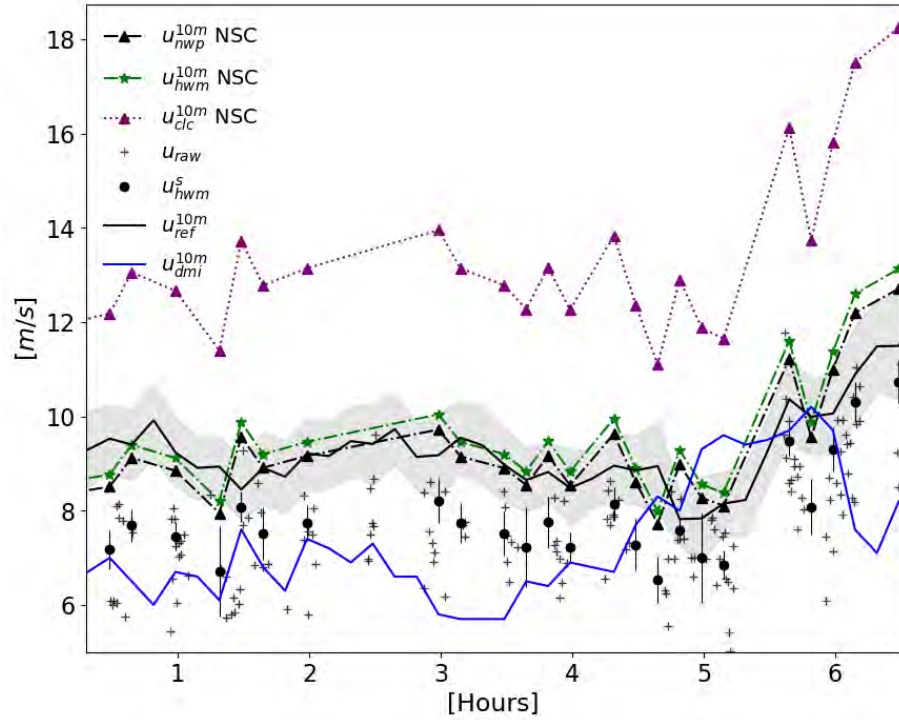
610 FIG. 1. Satellite imagery of the locations of the measurements sessions. Left: Hvide Sande (HVDS), Center:
611 Griben (GNB), Right: Eskebjerg (ESKB). The red star shows the location of the reference measurement and the
612 blue star shows the location of the DMI SYNOP stations.



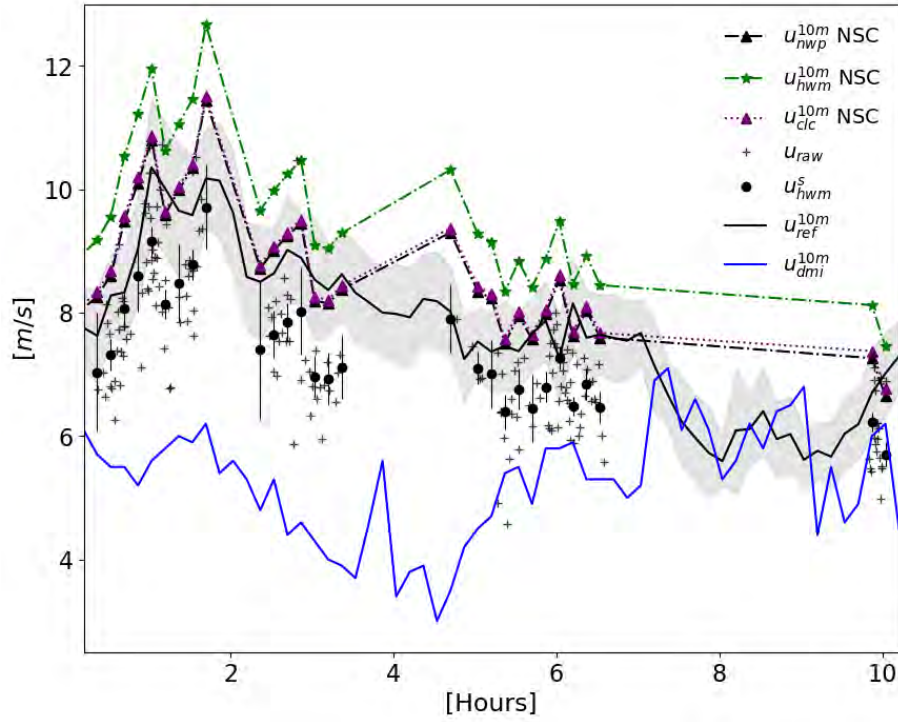
613 FIG. 2. Time series of wind measurements at Eskebjerg starting 2017 March 3rd, 05:30 UTC [ESKB4].
 614 Abscissa: Hours since the beginning of measurements. Ordinate: Wind speed [m/s]. Black crosses shows
 615 individual handheld measurements, u_{raw} and solid black circles shows the average within 10 minutes, u_{hwm}^s . The
 616 grey shaded area shows the standard deviation of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Holbaek.
 617 Purple broken curve: u_{clc}^{10m} NSC. Green broken curve: u_{hwm}^{10m} NSC. Black broken curve: u_{nwp}^{10m} NSC.



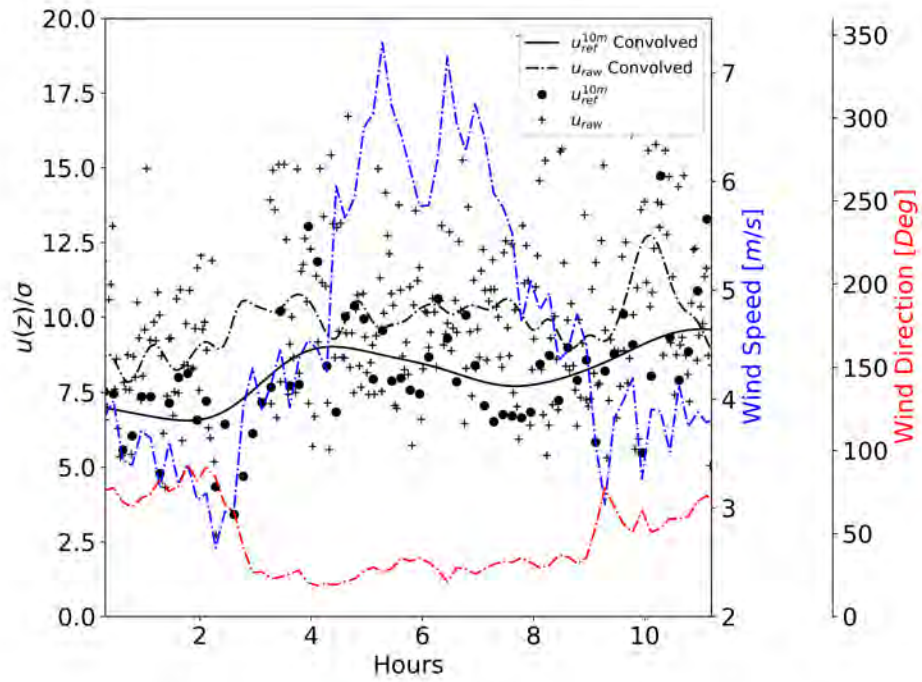
618 FIG. 3. Time series of wind measurements at Eskebjerg starting at 2017 March 23rd, 08:20 UTC [ESKB5].
 619 Abscissa: Hours since the beginning of measurements. Ordinate: Wind speed [m/s]. Black crosses shows
 620 individual handheld measurements, u_{raw} and solid black circles shows the average within 10 minutes, u_{hwm}^s . The
 621 grey shaded area shows the standard deviation of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Holbaek.
 622 Black curve: u_{ref}^{10m} . Black dashed curves: u_{nwp}^{10m} NSC and WSC (see legend).



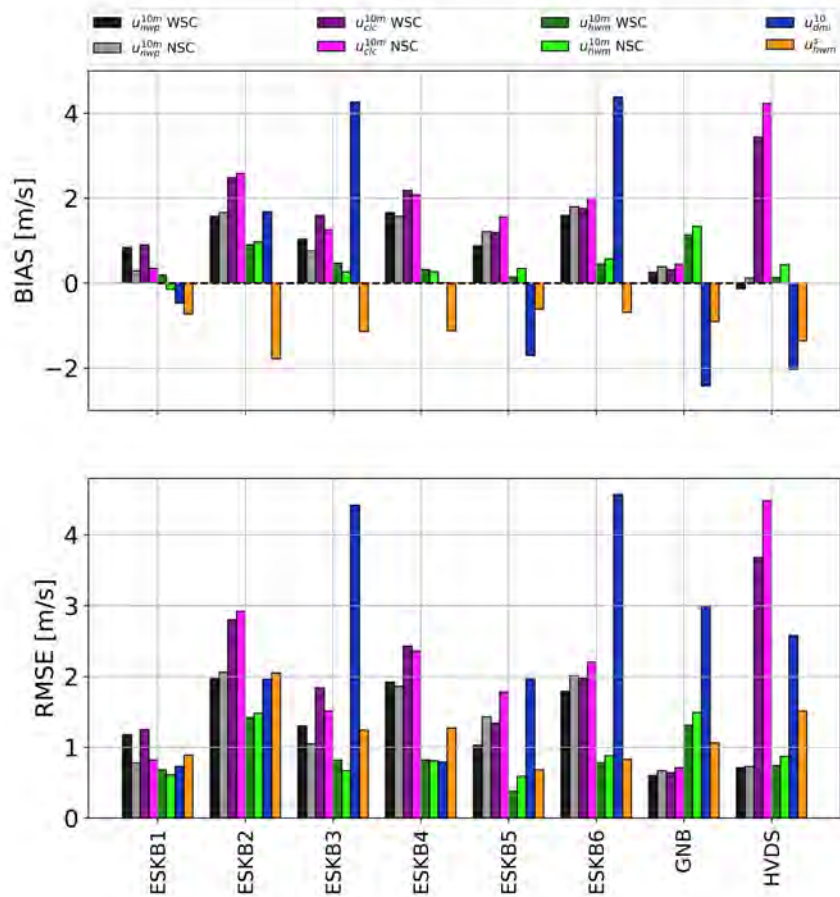
623 FIG. 4. Time series of wind measurements at Griben starting 2016 October 13th, 08:00 UTC [GNB]. Ab-
 624 scissa: Hours since the beginning of measurements. Ordinate: Wind speed [m/s]. Black crosses shows individ-
 625 ual handheld measurements, u_{raw} and solid black circles shows the average within 10 minutes, u_{hwm}^s . The grey
 626 shaded area shows the standard deviation of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from Griben. Purple
 627 broken curve: u_{clc}^{10m} NSC. Green broken curve: u_{hwm}^{10m} NSC. Black broken curve: u_{nwp}^{10m} NSC.



628 FIG. 5. Time series of wind measurements at Hvide Sande starting at 2016 September 3rd, 07:10 UTC
 629 [HVDS]. Abscissa: Hours since the beginning of measurements. Ordinate: Wind speed [m/s]. Black crosses
 630 shows individual handheld measurements, u_{raw} and solid black circles shows the average within 10 minutes,
 631 u_{hwm}^s . The grey shaded area shows the standard deviation of u_{ref}^{10m} . Black curve: u_{ref}^{10m} . Blue curve: u_{dmi}^{10m} from
 632 Hvide Sande. Purple broken curve: u_{clc}^{10m} NSC. Green broken curve: u_{hwm}^{10m} NSC. Black broken curve: u_{nwp}^{10m} NSC.



633 FIG. 6. Time series of wind measurements at Eskebjerg starting at 2017 March 23rd, 08:20 UTC [ESKB5].
 634 Abscissa: Hours since the beginning of measurements. Ordinate left: Ratio of $u(z)/\sigma_u$. Ordinate right: Wind
 635 speed (blue) and wind direction (red) from 10 m reference observations. Black dots and crosses shows the
 636 ratio $u(z)/\sigma_u$ from the reference measurements and the handheld measurements respectively. The full black line
 637 and dashed black line shows the convolved measurements using a Gaussian filter, from the reference and the
 638 handheld measurements respectively.



639 FIG. 7. Abscissa: Name of location. Lower ordinate: RMSE, upper ordinate: Bias. The ‘truth’ is the
 640 reference observation at 10 m, u_{ref}^{10m} . u_{dmi}^{10m} (blue) and u_{hwm}^s (orange) are the nearest SYNOP station and the
 641 handheld observations, respectively. The rest, named in the legend on top, is the extrapolated winds using
 642 roughness lengths of different origin.

Publication B

Title: Collecting and processing of Barometric Data from Smartphones for Potential use in NWP Data Assimilation.

Submitted: September 2018.

Authors: K. S. Hintz, H. Vedel and E. Kaas.

Journal: Meteorological Applications.

Journal Website: <https://rmets.onlinelibrary.wiley.com/journal/14698080>

Status: Published

Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation

Kasper S. Hintz¹  | Henrik Vedel¹ | Eigil Kaas²

¹Research and Development, Danish Meteorological Institute, Copenhagen, Denmark

²Climate and Computational Geophysics, University of Copenhagen, Niels Bohr Institute, Copenhagen, Denmark

Correspondence

Kasper S. Hintz, Research and Development, Danish Meteorological Institute, Lyngbyvej 100, 2100 Copenhagen, Denmark.
Email: kah@dmi.dk

Funding information

This work has received partial funding from Innovation Fund Denmark, through the industrial researcher programme, grant 5189-00042B.

Abstract

The potential for use of crowd-sourced data in the atmospheric sciences is vastly expanding, including observations from smartphones with barometric sensors. Smartphone pressure observations can potentially help improve numerical weather prediction and aid forecasters. In this contribution a method of collecting data from smartphones is presented, other methods are discussed and guidelines are derived from the experience. Quality control is vital when using crowd-sourced data. Screening methods aimed at smartphone pressure observations are presented. Results from previous studies, showing a substantial but long-term stable bias in combination with high relative accuracy, are confirmed. The collection of Danish smartphone pressure observations has been very successful, with over 6 million observations during a 7 week period. Case studies show that distinct weather patterns can be seen in unprocessed data. The screening method developed reduces the observational noise but filters out the majority of observations. Assimilating smartphone pressure observations in a single case study, using the 3D variational data assimilation system of the HARMONIE numerical weather prediction system, proved to decrease the bias of surface pressure in the model without increasing the root mean square error and the skill of accumulated precipitation increased. It is found that the altitude assignment of smartphones needs improvement.

KEYWORDS

crowdsourcing, observations, smartphones, surface pressure

1 | INTRODUCTION

The number of observations available for use in the geosciences, including numerical weather prediction (NWP), has vastly expanded due to advances in technology. Increasingly, private weather stations and mobile devices are connected to the internet, creating a dense observational network (e.g. Lonobox, 2018; Netatmo SAS, 2018;

Wunderground, 2018) which can be used to obtain information about the state of the atmosphere. Data originating from individuals are often referred to as “crowd-sourced data” (Howe, 2006). Muller *et al.* (2015) give an extensive overview of the potential of crowd-sourced data in the atmospheric sciences. Currently only a few preliminary studies on the use of crowd-sourced data in meteorology exist. Chapman *et al.* (2017) used Netatmo data to quantify the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Meteorological Applications published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

urban heat island effect of London, Overeem *et al.* (2013) quantified urban air temperatures using battery temperatures from smartphones and Clark *et al.* (2018) used private weather stations to do a fine-scale analysis of a severe hailstorm.

Regarding pressure, today's smartphones are often equipped with a built-in barometer (to track the user's change in altitude), which has created much interest from meteorological communities. Potentially, smartphone pressure observations (SPOs) can have a positive impact in NWP in regions devoid of SYNOPS (surface synoptic observations), which is the case in many places of the world. As the number of smartphones increases, as in sub-Saharan Africa (Aker and Mbiti, 2010), SPOs can potentially increase forecast quality. With a tendency towards increasing resolution in NWP without an increase in the national SYNOPS network, crowd-sourced data could be a valuable data source for NWP in the future.

Price *et al.* (2018) used smartphones for monitoring atmospheric tides and found that the bias of the observations is nearly constant over time. Madaus and Mass (2017) used the high resolution rapid refresh model forecasts from the (American) National Center for Atmospheric Research as the first guess and boundary condition. On top, they assimilated SPOs using an ensemble adjustment Kalman filter for creating initial conditions for the Weather Research and Forecasting model (Skamarock *et al.*, 2008). Even though Madaus and Mass (2017) did not include SPOs as an integrated part of the data assimilation system they obtained promising results with regard to forecasting a mesoscale convective system. In further work McNicholas and Mass (2018) improved the bias correction of SPOs using a random forest regressor, which proved to reduce the errors of the observations significantly. The random forest builds on top of classification and regression trees (see, for example, Hsieh, 2009, sec. 9.2).

Kim *et al.* (2015) focused on data collection and did the first experiments where SPOs were corrected using machine learning methods. In a continuing work, Kim *et al.* (2016) improved the methods for bias correction and data collection.

In earlier studies new mobile apps were developed and people were encouraged to download and use the apps (Kim *et al.*, 2015, 2016; McNicholas and Mass, 2018). One downside of this approach is the large effort going into advertising and maintenance of the app's codebase to keep the retention rate high; the risk in this approach is that it will only give a fraction of the potential number of available observations compared to the retrieval of pressure observations from inclusion of separate software in existing apps. Furthermore, individual apps may have a short lifetime which would cause the lifetime of individual data sources to be short.

Other studies have collected data *via* third-party applications (McNicholas and Mass, 2018; Price *et al.*, 2018), in which there are two challenging factors: first, the introduction of a blackbox approach of data collection, as one often does not know how the data have been processed; second, user privacy is an important aspect, especially with the new General Data Protection Regulation (GDPR) act from the European Union (European Union, 2018), which makes it difficult to combine the requirements for owner anonymity with the need to be able to identify each device in order to do bias correction based on data acquired over an extended timespan (Price *et al.*, 2018).

Based on initial idealized laboratory studies, the present study developed software for obtaining SPOs. Methods for collecting data and for quality assurance of the observations are presented, with the long-term aim of using them in an NWP data assimilation system. The remainder of this paper contains four main parts. Section 2 presents methods and results for idealized studies of SPOs. Section 3 covers methods for data collection and observation quality assurance and presents results. Section 4 presents a test of assimilation of SPOs using the 3D variational (3DVar) data assimilation system in the HARMONIE NWP system. Last, the interpretation of the results presented in Sections 1, 2 and 3 is discussed in Section 5.

2 | IDEALIZED STUDIES

2.1 | Methodology

First, in this study, a measurement is defined as a single value coming from the barometric sensor. An observation is defined as an average of measurements over a given time interval, following McNicholas and Mass (2018).

When starting a pressure measurement, it was observed that the short-term variance of pressure measurements on a phone at rest is larger at the beginning of an observation compared to the rest of the observation. Thus, when a measurement is started there is a short spin-up time, during which the pressure should not be logged. The spin-up time is due to a sensor internal infinite impulse response filter described by McNicholas and Mass (2018). The spin-up period was investigated by analysing 50 time series of 180 s of the pressure using an Apple iPhone 6 which contains a Bosch BMP280 pressure sensor (Bosch Sensortech, 2018). The BMP280 sensor is the same as that used in, for example, Apple iPhone 7, Huawei Nexus 6P and Samsung Galaxy S7 Edge. The absolute and relative accuracy of the BMP280 sensor is 1 and 0.12 hPa respectively (Bosch Sensortech, 2018). The average measurement interval of BMP280 is 5.5 ms. The sampling frequency returned by the iOS operating system is 1 Hz, which is used here. However, it is noted

that the Android operating system can return higher sampling rates. The time between each time series obtained was in all cases more than 2 hr.

The magnitude of the bias and the change of bias over time were examined by comparing measurements from an Apple iPhone 6 and a Samsung Galaxy A5 with a professional reference, a Vaisala PTB330 barometer (Vaisala, 2018). The absolute accuracy of the PTB330 barometer is 0.2 hPa (Vaisala, 2018). All devices were located at the same height, within a few metres horizontally, in a locked testing facility at the Danish Meteorological Institute (DMI). Seven individual measurement sessions were performed during 1 month.

2.2 | Results

An approximate spin-up time of 5 s was found on average for the 50 time series. The standard deviation over all experiments was 0.02 hPa. Figure 1 shows a comparison of the smartphones and the DMI reference barometer. The red shaded region shows the absolute accuracy of the barometer (± 0.2 hPa) (Vaisala, 2018). During the measurement period of Figure 1, the iPhone 6 had a bias of 1.0 hPa and the Galaxy A5 had a bias of -2.0 hPa. Even though the smartphones have a bias, it is seen that the variabilities of the curves are highly correlated.

In total, seven sessions, similar to Figure 1, were performed over a period of 1 month. The Apple iPhone 6 had a bias of 0.96 ± 0.06 hPa and the Samsung Galaxy A5 had a bias of -2.06 ± 0.09 hPa. Downsampling the smartphone

pressure time series to the same frequency as the DMI pressure time series, which has a frequency of 1/600 Hz, yielded a Pearson correlation co-efficient of 0.977 and 0.965 for the Apple iPhone 6 and Samsung Galaxy A5 measurements respectively relative to the DMI measurement, for all sessions.

In Figure 1 consistent short-term variability of the phone pressures can be seen, not resolved by the reference. The correlation co-efficient of the two smartphone pressure series based on 1 Hz data is 0.994.

3 | COLLECTION OF SMARTPHONE DATA

3.1 | Methodology

Data from smartphones operated by ordinary users in Denmark were collected over a period of nearly 2 months, from April 5, 2018, to May 24, 2018, to investigate methods of data collection and quality control. The data collection continues today aiming for future studies.

To work with the SPOs a testbed for data collection and observation control was made. The testbed system is referred to as SMAPS (Smartphone Pressure System). The SMAPS consists of several packages and functions for logging data from smartphones and quality assurance of the collected data. The SMAPS contains two main sub-packages: PMOB, which is installed client-side for data collection; and QCMB, which is installed server-side for quality assurance. PMOB is a software package written for iOS and Android which can be implemented in apps as a separate sub-program. PMOB logs and uploads data from the smartphone to a database. QCMB does further processing and quality assurance.

To collect observations from “the crowd,” PMOB was integrated into the app “DMI Vejret” (DMI Weather) developed and maintained by the private company SFS Development. It is a popular weather app in Denmark, based on meteorological products from both DMI (dmi.dk) and YR (yr.no).

3.1.1 | Data handling on the smartphones (PMOB)

PMOB uses the iOS and Android software development kit to access data from the barometer. Auxiliary data are collected from additional sensors if they are available. All variables uploaded are observations, averages calculated iteratively as:

$$\bar{x}_i = \frac{i-1}{i} \bar{x}_i - 1 + \frac{1}{i} x_i \quad (1)$$

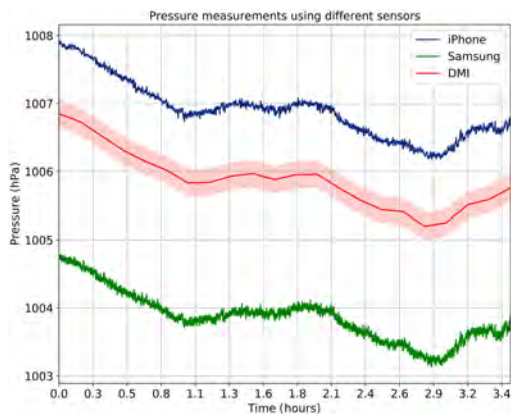


FIGURE 1 Comparison of two smartphone pressure measurements over time to a Vaisala PTB330 barometer (red/middle); blue (top), Apple iPhone 6; green (bottom), Samsung Galaxy A5 (2017). The shaded area shows the accuracy of the Vaisala PTB330 barometer of 0.2 hPa as stated by the manufacturer

and

$$\sigma_i = \sqrt{\frac{i-1}{i} \sigma_{i-1}^2 + \frac{1}{i-1} (x_i - \bar{x}_i)^2} \quad (2)$$

where \bar{x}_i is the average over $i = N$ measurements and x_i is the i th measurement. σ_i in Equation 2 is the variance computed over N measurements. In this way, one does not need to store time series in the memory of the phone or, most importantly, send all data to an online database, which could drain the battery and use more bandwidth. Due to the spin-up time mentioned in Section 2 PMOB skips the measurements for the first 5 s. The following observations are derived from measurements obtained in 7 s periods. This means $N \geq 7$ per SPO are sent by the app, depending on the sampling frequency (where the iOS 1 Hz is the lowest encountered), and that a minimum of 12 s is needed to obtain any SPO when the app is opened. The effect of using other periods than 7 s has not been tested, but it is noted that on average the app is open for 26 s based on 1.6×10^6 sessions recorded in April 2018.

The auxiliary data collected, if available, from each smartphone are acceleration in 3D, geo-location and the speed of the device. In addition, a smartphone id (uid) is always collected. In all cases data were collected only with the user's acceptance and knowledge and with clear communication. The smartphone identifier is created by PMOB on the first installation. It uniquely identifies the phone, necessary for bias correction. The ability from long sequences of data to identify the owner makes the collected data "personal," according to the GDPR act (European Union, 2018). Hence, legal advice regarding handling and security of data was required before data collection could start.

3.1.2 | Quality control of data received from the smartphones (QCMOB)

When the SPOs have been retrieved the observations enter the quality assurance component QCMOB. The workflow of QCMOB is illustrated in Figure 2.

First, the background departure of the pressure observation is calculated. In this study a short-term NWP forecast (0–2 h) from the DMI's operational model HARMONIE cycle 40 h1 (Driesenaar, 2009) is used as background. The background pressure is found by bilinear inverse distance weighting interpolation to the location of the observation. The model data are stored in 1 hr intervals and the model hour nearest the observation time is used. The background departure is then stored with the observation. A lookup table of biases with all reporting smartphones is then updated. If no bias exists, meaning a device reports for the first time, the background departure is stored with the unique identifier as a key. If the device exists, then the bias is recalculated with the new observation included.

Second, an altitude of the terrain at the latitude–longitude position of the observation is retrieved from the Danish Terrain Model (DTM), which has a horizontal resolution of 10 m and a vertical resolution of 0.05 m. The reference used in the DTM is the DVR90 geoid model (Danish Environmental Protection Agency, 2015), which has a mean deviation of 0.05 m and a standard deviation of 0.34 m compared to the World Meteorological Organization recommended Earth Gravitational Model 1996 (EGM96) geoid model (WMO, 2014). The reason for using the terrain model is that the Global Navigation Satellite System (GNSS) derived altitudes from the smartphones are not of sufficient quality at present for use in NWP. However, the GNSS altitudes and DTM altitudes are compared in a screening check, which

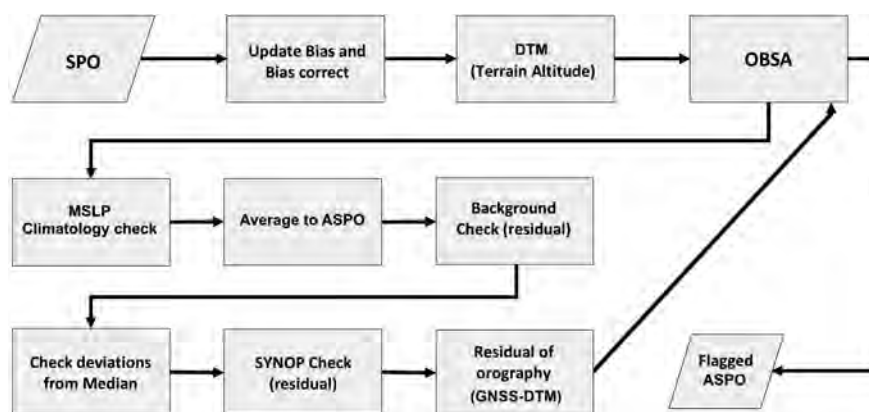


FIGURE 2 Workflow of QCMOB. Smartphone pressure observations (SPOs) are prepared for Observation Assurance (OBSA) by updating the bias for each observation and finding the terrain height in the Danish Terrain Model (DTM). See text for details

will be described later. GNSS derived altitudes are likely to improve significantly in the future (Robustelli *et al.*, 2019), when more smartphones support dual-frequency GNSS. Today, vertical inaccuracy can be of the order of tens of metres (Bauer, 2013) but is expected to reach sub-metre level. Third, the observations enter the initial Observation Assurance (OBSA, see Figure 2). Each observation is allocated with a *flag*, initially set to zero. In each check, an observation can have a unitless penalty added to the *flag* value. The size of the penalties can be changed for each check *via* a namelist.

In OBSA the mean sea level pressure (MSLP) is computed following Madaus and Mass (2017):

$$p_{\text{msl}} = (p - 0.3) \left(1 + k_1 \frac{h}{(p - 0.3)^{k_2}} \right)^{1/k_2} \quad (3)$$

Here $k_1 = 8.4228807 \times 10^{-5}$ and $k_2 = 0.190284$. h is the altitude above mean sea level, obtained by adding 1 m to the altitude from the DTM, assuming that the smartphone is in the hand of a person, p is the SPO in hectopascals and p_{msl} is the MSLP in hectopascals. Then a climatological check is performed, whether p_{msl} is within the range 850–1,050 hPa. For observations outside this range a penalty of 10 is added. Then it is checked whether observations close in time from the same device exist, to reduce the number of observations entering the further processing by making time averages. The time interval adopted here is 5 min. The averaged SPOs are named ASPOs. The ASPOs are then returned to OBSA. It is the ASPOs which are entering the remaining checks. An ASPO can be considered a time average of the SPOs obtained during one app session as the typical session time is 26 s. Penalties from the previous check are averaged. Hereafter, a background check is performed using the background departure found in the first step before entering the OBSA. Departures greater than ± 1 hPa are given a penalty of 10. It is noted that the standard deviation of surface pressure in the NWP model is of the order of 0.3–0.4 hPa, and that 1 hPa corresponds to approximately 8 m altitude difference.

ASPOs within a predefined distance from each other and obtained within 10 min in time are then used to compute a median value on a grid in the median check. To do this ASPOs are first converted to MSLP and then corrected by the horizontal pressure gradient obtained from the NWP background. If an ASPO deviates more than a predefined threshold from the median, it is given a penalty and excluded from the next iteration median computation. The search for ASPOs that deviate too much continues until no more outliers are found. Typically, four to six loops are needed for this to happen. The median values are only used to derive penalties; the medians themselves are not stored.

Two settings were used for the distance and threshold in the median check. In experiment EXP_MEDV1 a distance radius of 0.2° and a threshold of 1.0 hPa were used. In EXP_MEDV2 a distance radius of 0.5° and a threshold of 0.2 hPa were used. The former is referred to as a “loose” median check and the latter a “strict” median check.

Hereafter, a check against SYNOP stations managed by the DMI is performed. The resolution in time for the SYNOP data is 10 min. Inverse distance weighting interpolation from the four nearest SYNOP stations is performed to the observation point. Altitude differences are corrected for by comparing the MSLP. The ASPO is given a penalty with a magnitude of the absolute value of the residual in units of hectopascals. Finally, the deviations between the terrain model, DTM, and the GNSS altitudes of each ASPO are computed. A substantial deviation from the GNSS altitude could suggest that a user is not located near the surface but in a tall building, making the height of an observation inaccurate. Deviations greater than 3 m are given a penalty of 10.

Based on all penalty calculations OBSA writes an output in which each ASPO is associated with a penalty which is the sum of all penalties given to each ASPO in the OBSA filtering.

3.2 | Results

During the period considered, April 5 to May 24, 2018, 6,336,475 observations were obtained, from 45,506 individual smartphones. The observations are not uniformly distributed in time throughout the day. A sharp rise in the number of observations was seen in the morning followed by a more gradual decrease in the early evening. The ratio of the minimum and maximum number of observations *per* 10 min is about 7.

Table 1 summarizes the checks and filtering performed in OBSA. A flagged ASPO is defined as an observation which has a penalty of more than 1. Two setups have been focused on: the loose and the strict median check. Bold numbers in parentheses refer to the loose median check and bold numbers refer to the strict median check. NWP experiments EXP_MEDV1 and EXP_MEDV2 are presented in Section 4.1.

The background check flags 80% of the SPOs. In total 88.6% and 90.6% of all ASPOs are flagged in EXP_MEDV1 and EXP_MEDV2, respectively. One may argue that these numbers indicate that the background check is too strict, with the danger of removing important observations with a significant background departure. Whether that is the case depends on the intended use. The 3DVar data assimilation system used here incorporates a data thinning procedure, requiring a certain minimum distance between observations

TABLE 1 Number of flagged observations by each check in OBSA using two different setups of the OBSA, compared to the total number of observations

Type of check	Total (6,336,475)	Per cent
MSLP and climatology of MSLP	552,383	8.7
ASPOs (reduced observations)	1,421,642	22.4
Background check (comparison to NWP)	1,137,128	80.0
Median check (comparison to median within area)	(272,046) / 796,541	(19.1) / 56.0
SYNOP check (comparison to DMI stations)	750,677	52.8
Orography (deviation from terrain model)	929,219	65.4
Total flagged ASPOs	(1,264,544) / (88.6) / 90.6	1,287,982
Total number of ASPOs	1,421,642	

ASPOs, averaged smartphone pressure observations; DMI, Danish Meteorological Institute; MSLP, mean sea level pressure; NWP, numerical weather prediction; OBSA, Observation Assurance; SYNOP, surface synoptic observation. Bold numbers in parentheses refer to experiment EXP_MEDV1 and bold numbers refer to EXP_MEDV2. Numbers in normal formatting are identical for the two. In total (88.6%)/90.6% of all ASPOs were flagged (penalty above 1). The total number of observations (6,336,475) is reduced by averaging to ASPOs (1,421,642) to which values hereafter are compared.

actively included, which means that the ASPOs would be heavily filtered anyway. The threshold is a balance of reducing noise and keeping important observations. This is further discussed in Section 5.

Figure 3 shows all ASPOs within 1 hr on May 10, 2018 (left) and the observations that were given a penalty less than

1 (right) using settings of EXP_MEDV1. The overall pressure tendency is seen to be clearer after the OBSA routine has been done. However, it appears that many good observations have also been removed, indicating that the screening method might be more strict than necessary.

Figure 4 shows ASPOs during a meteorological event on April 30, 2018. A small low-pressure system moved northeast towards Denmark over western Europe. The occluded point moved over northern Jutland giving rise to local, high precipitation rates. The coloured circles show the pressure tendency, deduced from individual smartphones that have provided observations about an hour apart. All such ASPO pairs available were used to produce Figure 4. The contours in Figure 4 show 1 hr accumulated radar-derived precipitation following the methods described by Olsen *et al.* (2015).

Figure 5 shows ASPOs during a meteorological event on May 10, 2018. During the day a surface cold front with embedded convection moved across Denmark from southwest to northeast. A general positive pressure tendency was observed with a magnitude of about 0.2 to 0.4 hPa/hr. Wind observations (not shown) show in general that the frontal zone at the surface was advancing 30–50 km in front of the rainband.

4 | SMARTPHONE PRESSURE IN NWP DATA ASSIMILATION

4.1 | Methodology

In total, five numerical simulations were performed, all initiated at May 5, 2018, 0000 UTC and running to May 10, 2018, 0900 UTC in cycles of 3 hr (see Table 2). The HARMONIE cycle 40 h1 was used for all NWP model runs,

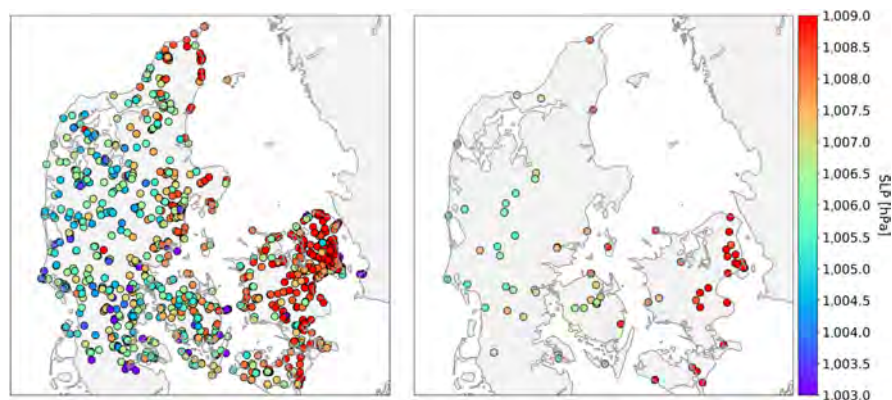


FIGURE 3 Averaged smartphone pressure observations (ASPOs) between May 10, 0830 UTC and 0930 UTC. Left: All observations. Right: Observations with a penalty below 1. Colour presents the ASPOs reduced to mean sea level pressure (MSLP) in hPa given by the colour bar

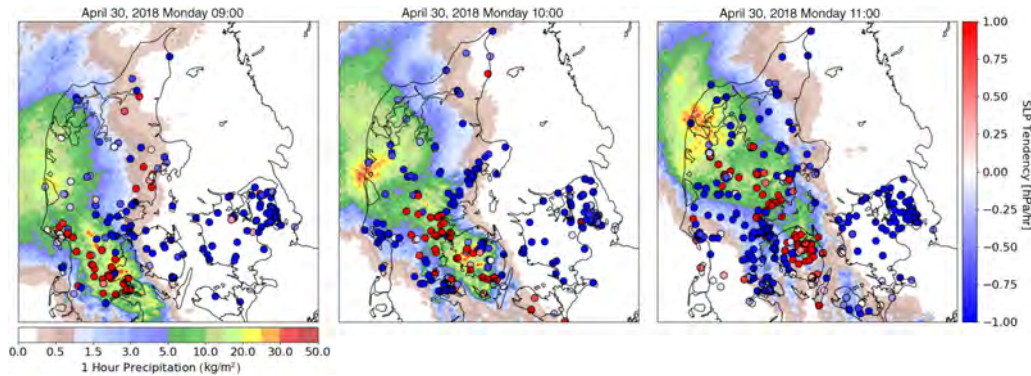


FIGURE 4 Sea level pressure (SLP) tendency (hPa/hr) over 1 hr and radar estimated accumulated precipitation over 1 hr over Denmark. Coloured circles show the pressure tendency measured by individual smartphones and contours show radar-derived hourly accumulated precipitation. All observations from individual devices are shown

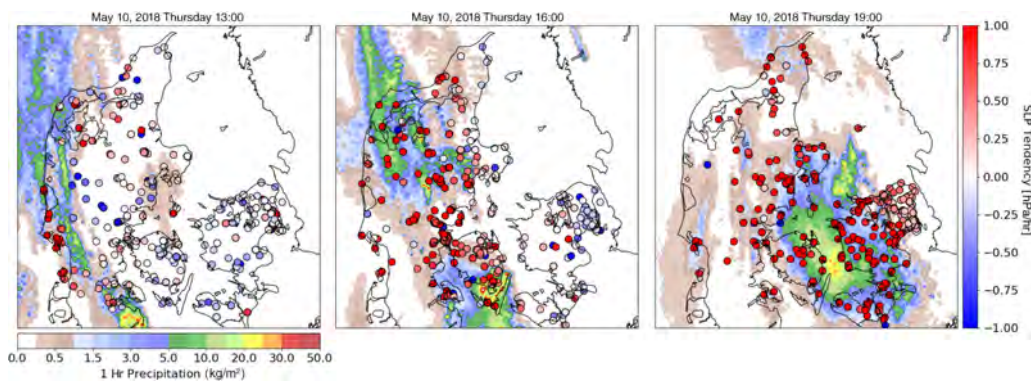


FIGURE 5 Sea level pressure (SLP) tendency (hPa/hr) over 1 hr and radar estimated accumulated precipitation over 1 hr over Denmark. Coloured circles show the pressure tendency measured by individual smartphones and contours show radar-derived hourly accumulated precipitation. All observations from individual devices are shown

TABLE 2 An overview of the numerical simulations performed in this study

Experiment name	OPR	REF	EXP	EXP_MEDV1	EXP_MEDV2
SYNOPSIS DK	Yes	No	No	No	No
SPO	No	No	Yes	Yes	Yes
QCMOB	-	-	Simple	Yes	Yes
QCMOB type	-	-	No background, median, SYNOPSIS, orography check	OBSA loose median check	OBSA strict median check

See text for details about QCMOB and QCMOB type.

using the DMI default settings for the model (Yang *et al.*, 2017). This includes assimilation of all observations used in the main operational NWP model with the exceptions mentioned below. The model area includes Denmark, UK, Belgium, Netherlands, northern France, Germany, Poland, the

southern half of Sweden and Norway, western parts of the Baltic States and Finland. See Yang *et al.* (2017) and Figure S1 for details. The observation handling was modified to control which pressure observations enter the preprocessing system of HARMONIE. Only ASPOs passing

QCMOB are allowed to enter the preprocessing system; hence, ASPOs have to pass both SPO specific quality control (QCMOB) and the general quality control in HARMONIE to be allowed into the 3DVar data assimilation system.

Three experimental runs with ASPOs and no Danish SYNOPSIS were done. In addition, two runs with no ASPOs were done, one without the Danish SYNOPSIS (REF) and one with the Danish SYNOPSIS (OPR). In all runs SYNOPSIS pressure from outside Denmark was included. The OPR run corresponds to a normal operational DMI HARMONIE forecast. An overview of the simulations is listed in Table 2.

In the three runs with ASPOs the filtering of SPOs is increasingly stringent. The goal is to assess whether ASPOs can have a positive impact in a region devoid of SYNOPSIS.

The root mean square error (RMSE) and the bias of surface pressure were computed using the DMI SYNOPSIS stations as a reference. The NWP value of surface pressure was computed by bilinear inverse distance weighting interpolation to the location of the observation. For verification of precipitation the fractional skill score (FSS) (Roberts and Lean, 2008) was used. The FSS is a field-based verification score given by:

$$\text{FSS} = 1 - \frac{\text{MSE}}{\text{MSE}_{\text{REF}}} \quad (4)$$

Here MSE is the mean squared error of a model field given as:

$$\text{MSE} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (O_{i,j} - M_{i,j})^2 \quad (5)$$

where $O_{i,j}$ and $M_{i,j}$ are a binary observation field and the model field respectively and N represents the spatial scale. MSE_{REF} is defined as the largest possible MSE that can be obtained from the model and observation fields:

$$\text{MSE}_{\text{REF}} = \frac{1}{N_x N_y} \left(\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M_{i,j}^2 \right) \quad (6)$$

$\text{FSS} = 1$ is a perfect score and $\text{FSS} = 0$ is the worst possible score.

In this study the observation field is 6 hr accumulated precipitation estimated by radar data following the methods described by Olsen *et al.* (2015). Only the area covered by the DMI radar network was used for verification of precipitation. Both the observation and model fields of 6 hr accumulated precipitation were converted into binary fields using a percentile of 95%, such that values greater than the 95th percentile are given a value of 1 and 0 otherwise. By using a

percentile, the FSS score is assured to converge towards 1 as N increases. It makes the FSS sensitive to location of precipitation but not sensitive to precipitation amounts. FSS is used to avoid the double penalty problem one can risk when validating precipitation from high resolution models against rain gauge point measurements (Nurmi, 2003).

4.2 | Results

The numbers of surface pressure observations used in the data assimilation system of HARMONIE are listed in Table 3, as a sum of all observations over all data assimilation cycles in the NWP simulation period. Only observations of surface pressure are shown. The ASPO inputs to the data assimilation system are those observations that have passed QCMOB in each experiment. In all experiments 28,757 traditional pressure observations from outside Denmark were used. In addition, traditional observations (TEMP, aircraft, AMSU etc.) from within the model area were used. The assimilation system of HARMONIE rejects observations based on a first guess check and a required minimum distance between sites. As expected, the rejection rate is reduced when the filtering of ASPOs becomes tighter, seen by the decreasing rejection fractions of EXP_MEDV1 and EXP_MEDV2 compared to EXP with 3.1, 2.3 and 12.4% respectively. Note that these numbers are from the data assimilation system of HARMONIE and hence the fractions are relative to the total number of ASPOs that passed QCMOB, thus indicating that observations of higher quality entered the system and less filtering occurred.

Figure 6 shows the RMSE (left ordinate, lines) and bias (right ordinate, bars) as a function of lead-time for surface pressure in the NWP experiment period 6 May 0000 UTC to 10 May 0900 UTC. The first day (5 May) has been excluded to allow for spin-up for the model. It is seen that EXP has the highest RMSE for all lead-times and is the only run with

TABLE 3 Use of surface pressure in the HARMONIE data assimilation (DA) system in the period May 5, 0000 UTC, to May 10, 0900 UTC, for each experiment

Experiment	REF	EXP	EXP MEDV1	EXP MEDV2
ASPO input to DA	0	8,385	1,566	872
ASPO rejected by DA	0	1,040 (12.4%)	49 (3.1%)	20 (2.3%)
Total ASPO	28,757	27,717	28,708	28,737
Total ASPO + rejected	28,757	28,757	28,757	28,757

ASPO, averaged smartphone pressure observation.

ASPO input to DA counts only observations that passed QCMOB. Numbers given in parentheses are the rejected ASPOs in per cent relative to the total.

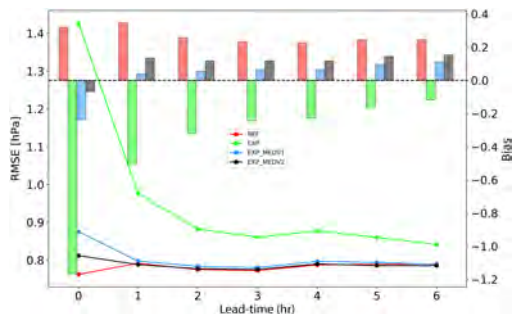


FIGURE 6 Root mean square error (RMSE) and bias for surface pressure as a function of lead-time in the range from May 6 to May 10, 2018. Lines show RMSE on the left ordinate and bars show the bias on the right ordinate. Note that the first day has been excluded to allow for spin-up time for the model

only negative bias. For EXP_MEDV1 and EXP_MEDV2 the RMSE is close to that of REF but the bias has decreased. OPR is not included in Figure 6 as the SYNOP stations used for verification are included in that run. The ASPOs improve the forecast with respect to REF in EXP_MEDV1 and EXP_MEDV2.

Figure 7 shows the RMSE and bias as in Figure 6 but for 10 m wind speed. EXP performs poorly, concerning both RMSE and bias. The RMSE values for EXP_MEDV1 and EXP_MEDV2 are a little higher than REF. During the first few hours, the bias is lower compared to REF but at 4 hr and onwards the bias increases; it is noted that the bias is still low compared to RMSE.

Figure 8 shows the FSS for the 95th percentile (left) and a threshold of 24 kg/m² (right) of 6 hr accumulated precipitation between 10 May 0900 UTC and 1500 UTC when a frontal zone passed over Denmark (see Figure 5). The

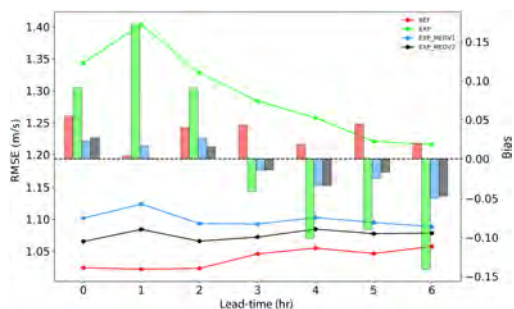


FIGURE 7 Root mean square error (RMSE) and bias for 10 m wind speed as a function of lead-time in the range from May 6 to May 10, 2018. Lines show RMSE on the left ordinate and bars show the bias on the right ordinate. Note that the first day has been excluded to allow for spin-up time for the model

threshold 24 kg/m² is a warning criterion at the DMI. Again, for the percentile EXP has the lowest score and is below the random score (FSS Random) until a scale of 25 km, meaning that a random forecast would perform better. EXP_MEDV1 shows substantial improvement but is still performing more poorly than REF.

Figure 9 shows the 6 hr accumulated precipitation for the same period as Figure 8 binned into different intervals. From Figure 9 it is seen that REF declines most rapidly and does not have the same observed high precipitation intensities as the other runs. EXP is in general too high. EXP_MEDV1 and EXP_MEDV2 are closer to the observed values with no clear positive or negative bias.

5 | DISCUSSION

Considering Figure 1 it is evident that a pressure bias must be determined for each phone individually. Examining the behaviour of the biases over a period of 1 month showed that the biases do not fluctuate much over time, with a standard deviation of only 0.06 hPa and 0.09 hPa for the Apple iPhone 6 and the Samsung Galaxy A5 phone respectively. These results are consistent with the findings of Price *et al.* (2018) who monitored the biases for 3–12 months. Considering the apparent long-term stability of the bias, corrected values will be of a quality that is promising for future application. It cannot be ruled out that bias may drift over the course of years. In applications where the pressure tendency can be used, rather than the absolute values, many problems related to SPOs will be removed if the tendency is based on SPOs from individual devices that are not moving. Considering Figure 1, one can see similarities in small scale variations in the phone pressures, e.g. a small decrease at 2.1 hr and a small increase at 0.3 and 3.4 hr. Accordingly the correlation of the two phone pressure series was found to be higher (0.994) than against the reference (0.977 and 0.965). It appears to resolve small scale fluctuations not resolved by the reference barometer.

Due to the lack of proper calibration of the smartphone barometers, it is necessary to apply an individual bias correction to SPO data from each phone. In this work a single bias correction is applied to each smartphone. This is suboptimal if many observations from a given smartphone come from a few different locations (e.g. home and work). McNicholas and Mass (2018) allow different bias corrections at different locations, which is more optimal but requires access to background processes to retrieve more observations. In our study the bias correction is done using NWP data. Potentially that is dangerous; bias correction of observations against an NWP model that has its own errors and later assimilates the corrected data has in some cases in the past led to NWP model drift (Vasiljevic *et al.*, 2006). See also Eyre (2016).

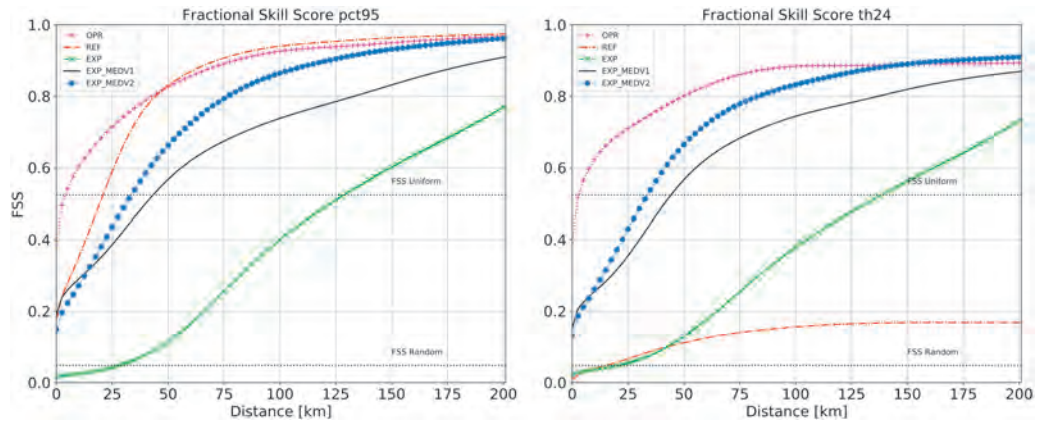


FIGURE 8 Fractional skill score (FSS) for 6 hr accumulated precipitation valid at 1500 UTC, May 10, 2018. Left: Using the 95th percentile. Right: Using a threshold of 24 kg/m^2 . FSS Random denotes the FSS from a random forecast with the same fractional coverage as the observations

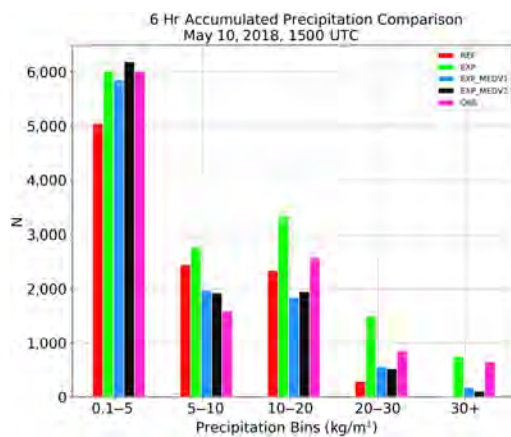


FIGURE 9 Six hour accumulated precipitation bins valid at 1500 UTC, May 10, 2018. Observations (OBS, magenta) are estimated from radar products

However, in the case of future SPO use in operational NWP one will still assimilate surface pressures from professional platforms (SYNOps from land, ships and buoys). At the DMI these are never bias corrected, which will anchor the model pressure, preventing drift. In addition, a comparison to SYNOP data is part of the quality control of the ASPOs, which will stabilize the selection of ASPOs and prevent model drift. In the study presented in this paper, the SPO bias corrections were derived using the operational DMI NWP model, which does not assimilate any SPO data but is otherwise similar to the model used in this study. Further, the simulations include assimilation of available SYNOP

surface pressures from SYNOps in the whole model area except Denmark.

One reviewer noted that: “Rejecting 80% of the SPOs based on the background check seems much too high to me. This reduces the independence of the SPOs, and in cases with large background error the accepted SPOs will tend to reinforce the error and less weight will be given to independent SYNOP reports.” For the reasons stated in the previous paragraph, this is not thought to be an issue. As was mentioned in Section 3.2, the 3DVar system used here includes a data thinning procedure, which causes the ASPOs to be heavily filtered in any case. Due to the check against SYNOps, these are implicitly given a higher weight than ASPOs. Here, it has been necessary to filter out observations with a poor altitude assignment through a strict background check, but this is something that should be improved in future to allow larger, potentially significant, background departures. As mentioned, a background departure of 1 hPa was used as a threshold for flagging observations (see Figures S2 and S3). The standard deviation of surface pressure from the operational DMI HARMONIE model is of the order of 0.3–0.4 hPa. 1 hPa corresponds to about three standard deviations. It is agreed that rejecting 80% of the SPOs is much too high for other uses, such as observational based nowcasting, and that the screening methods presented here can be improved. Experiments to relax the background check and to implement gradual penalty functions rather than threshold-based penalties are planned to be carried out in the future.

As stated in Section 3.2, more than 6 million observations were collected and the frequency during the day is not uniform. This is not surprising as SPOs are only collected when the app is in use. One disadvantage of this is that observation statistics cannot be based on regular time-spaced

observations. The non-uniform frequency reflects the diurnal pattern of people's lives. Further, one can imagine a device sending observations primarily from locations high above the ground (i.e. tall buildings; see also Figure S4). Bias correcting such a device will overshoot the true bias significantly, which is an issue future work will consider.

The DMI is operating 66 SYNOP stations measuring surface pressure (Scharling and Rajakumar, 2003), while on average SPOs were obtained from 7,342 unique smartphones *per day* during this study, more than a factor of 100 more. It must be stressed, though, that the qualitative value of a single SPO is lower than that of a SYNOP station and the SPOs are not evenly distributed in space and time. This is evident in Figure 3 where noise is widespread and only a few observations remain after QCMOB.

An improvement to the orography check in QCMOB (see Section 3.1) would be to allow greater residuals between the GNSS derived altitudes and the DTM, as the GNSS has a high uncertainty. This is also indicated in Table 1, from which it is seen that 65.4% of all ASPOs are flagged by this check.

The studies of Kim *et al.* (2015) and McNicholas and Mass (2018) both obtained SPOs through a dedicated app for the purpose. The great advantage of their approach is the comfort of being able to tune parameters. However, one disadvantage is that it is time-consuming and, with a few exceptions, the scientific community is in general not prepared for advertising its own apps and keeping the conversion rate and retention rate high. This is most evident when comparing Kim *et al.* (2015) where only 11,000 observations *per day* on average were collected over 240 days. Here, the SMAPS as described was included in an existing widely used app, resulting in a high number of pressure observations. A high number is necessary as the lower quality of the SPOs with respect to SYNOPs requires heavy filtering. It is important to recall that the standard error of the mean decreases with the square root of the number of individual observations.

To obtain the observations before the users turn off the app, the measurement period must be short. It is possible to run the software in the background on Android, and thereby obtain long sampling periods and many observations, but it is strongly advised against for two reasons. First, an app consumes more power when it is continuously active, especially when using the GNSS of the smartphone. Second, keeping user privacy in mind, it is best practice only to collect data when a user uses the app actively. On iOS, running software in background mode for an extended time is in general not allowed (Apple Inc., 2017).

In this study, measurements from a period of 7 s were used to calculate the mean, which was then sent to the database. It is not clear whether Kim *et al.* (2015) did any

averaging. Madaus and Mass (2017) used 15–40 s. However, in their case data were collected in the background every hour by default, while in our case the barometer was only accessed when the app was in use.

The few studies that consider data collection from smartphones indicate that the best approach is a collaboration with the industry, using an external app as a platform for data collection, such as “The Weather Channel” (McNicholas and Mass, 2018). It is advised that the data collection is kept in-house so that no blackbox of data processing is introduced. This will help ensure that the meteorological community can develop common standard methods for processing smartphone data as work progresses. One way to achieve this is to build a software development kit in a collaborative framework. If the bias correction can be done client-side the need for a unique ID can be removed, although other problems may then arise, e.g. lack of end-user version control as software updates are controlled by the phone owner.

Figure 5 shows that pressure tendencies obtained from smartphones without prior filtering can depict current weather. A general and coherent negative pressure tendency in front of the frontal zone is seen at both 1300 UTC and 1600 UTC at Funen and Zealand and eastern Zealand respectively. During the frontal passage a sharp rise in surface pressure is observed, as one would expect. Obviously, there is some noise, but a forecaster would get the overall picture. As the tendencies are derived from individual smartphones there is in this case no need for a bias correction. A more sophisticated approach is needed to derive the pressure tendency using observations from different smartphones, which would result in many more data.

Figure 6 shows that, as expected, it is not valuable to assimilate unfiltered SPOs into a 3DVar system, seen by the high increase of RMSE in the EXP run. Even though a minor increase of RMSE for EXP_MEDV1 and EXP_MEDV2 relative to REF at $t = 0$ is seen, it is noteworthy that the bias of EXP_MEDV1 and EXP_MEDV2 is lower than that of REF with no cost in RMSE. Also, the bias for EXP_MEDV1 is lower than that of EXP_MEDV2 except for $t = 0$. This can be an indication that the filtering of EXP_MEDV2 is too strict, flagging too many observations. One improvement to QCMOB would be to replace the step-wise functions for flagging observations with more sophisticated smooth functions to avoid discontinuities. It is also noted that the bias changes sign in the case of EXP_MEDV1 and EXP_MEDV2. It has not been possible to identify a conclusive reason for this.

Minor increases in RMSE were seen for 10 m wind speed in Figure 7 and small changes in bias, not considering EXP which has both a high RMSE and bias throughout the forecast. For the first forecast hours, the bias decreased slightly

comparing EXP_MEDV1 and EXP_MEDV2 to REF. However, from the fourth forecast hour, the biases increase with opposite sign relative to REF. It is seen in both Figures 7 and 8 that the bias starts with an opposite sign at the beginning of a forecast compared to the end of the forecast. The causes of this are at present unknown, but it is noted that the biases are very small relative to RMSE. It is seen from Figure 8 that OPR was the best for the particular case of May 10, 2018, between 0900 and 1500 UTC. Considering the 95th percentile REF overall had a better score than the experiments. One advantage of using percentiles to compute the FSS is that the scores converge towards 1 as the scale increases. However, this is at the cost of losing information about the precipitation amounts. One cannot be sure if a model creates enough precipitation which is evident when considering the threshold in Figure 8. A threshold of 24 kg/m^2 over 6 hr was used as this is a criterion of issuing a warning of heavy rainfall in Denmark. It is then seen that REF performs poorly. This is also seen from Figure 9, where REF has no observations in the highest bin of 30 kg/m^2 or more. Here, EXP scores better than REF but, considering the previous results, EXP cannot be argued to perform well overall.

EXP_MEDV1 and EXP_MEDV2 have a distribution that is closer to the observed distribution. However, imbalance in the initial conditions could be imagined causing moisture spin-up effects, which could be part of the reason for higher precipitation amounts in the experiments. Overall EXP_MEDV2 is argued to perform better than EXP_MEDV1 with respect to precipitation due to a better FSS score, when using both the 95th percentile and the threshold in the FSS. FSS Uniform represents the FSS obtained at the grid-scale from a forecast fraction equal to FSS Random at every point (Roberts and Lean, 2008).

It is clear by comparing EXP to REF in general that the assimilation system of HARMONIE is not well suited for receiving SPOs directly without any prior screening, also considering Table 3 where the rejection fraction of SPOs in EXP is about 9% higher compared to EXP_MEDV1 and EXP_MEDV2. It is also clear from the figures that the final word has not been said with respect to the optimal filtering of SPOs before assimilation.

6 | CONCLUSION

This study discusses the collection of Smartphone Pressure Observations (SPOs) *via* software installed in a weather app operated by ordinary citizens in Denmark and considers the usefulness of SPOs in professional meteorology.

A comparison between two smartphones and a reference barometer determined that the relative smartphone pressure is reliable, while biases of the order of 1 hPa exist. The biases were found to be stable over periods of

at least a month. This compares well with previous studies. The smartphone barometers studied in detail had a spin-up time of about 5 s, during which measurements should not be recorded. The short-term variability of the pressure measurements calls for averaging over a period. Typical measurement frequencies are high enough, 1 Hz or higher, to provide SPOs based on several in-phone measurements. During a 2 month period, more than 6.3 million SPOs were obtained from 45,506 unique smartphones in Denmark, more than 5,200 *per* hour but down to about 200 *per* hour overnight. It is demonstrated that the SPOs contain information about current, active weather and that with various filters one can obtain high quality pressure observations, at the expense of reducing the number of observations.

Finally, it is demonstrated that, when SPOs were incorporated into a numerical weather prediction (NWP) model with variational data assimilation, the forecasts were improved in a region (Denmark) artificially devoid of SYNOP pressure observations in the reference run, demonstrating that the filtered SPOs have quality on a level useful in NWP. Also, it was found that the HARMONIE data assimilation system is not well suited for receiving SPOs directly without any prior screening.

The results and methodologies presented in this study advance the use of crowd-sourced data both regarding the collection of such data and how to process the data. However, it is clear that future studies are required, regarding both the settings on each smartphone to obtain an SPO and the optimal filtering of SPOs at meteorological institutes to remove poor observations. The prime difficulty is to assign altitudes to the SPOs properly. Altitude corrections can be done *via* bias correcting; however, it is then of utmost importance that the bias correction can take into account that the device is mobile. This problem will disappear as in-phone derived altitudes improve, but it will take several years before that happens.

ACKNOWLEDGEMENTS

The authors wish to thank the company SFS Development for making their app available as a testbed for science. Also, partial funding from Innovation Fund Denmark is much appreciated. The authors are very grateful for the detailed and valuable comments provided by two anonymous reviewers, which led to substantial improvement of the paper.

ORCID

Kasper S. Hintz  <https://orcid.org/0000-0002-6835-8733>

REFERENCES

- Aker, J.C. and Mbiti, I.M. (2010) Mobile phones and economic development in Africa. *The Journal of Economic Perspectives*, 24, 207–232. <https://doi.org/10.1257/jep.24.3.207>.
- Apple Inc. (2017) *Background execution in iOS*. Available at: <https://developer.apple.com/library/archive/documentation/iPhone/Conceptual/iPhoneOSProgrammingGuide/BackgroundExecution/BackgroundExecution.html> [Accessed 21st June 2018].
- Bauer, C. (2013) On the (in-)accuracy of GPS measures of smartphones: a study of running tracking applications, In: *Proceedings of International Conference on Advances in Mobile Computing and Multimedia – MoMM '2013*. Presented at the International Conference, ACM Press, Vienna, Austria, pp. 335–341. <https://doi.org/10.1145/2536853.2536893>.
- Bosch Sensortec. (2018) *Bosch BMP280 pressure sensor*. Available at: https://www.bosch-sensortec.com/bst/products/all_products/bmp280 [Accessed 13th December 2018].
- Chapman, L., Bell, C. and Bell, S. (2017) Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations: CROWDSOURCING THE LONDON UHI. *International Journal of Climatology*, 37, 3597–3605. <https://doi.org/10.1002/joc.4940>.
- Clark, M.R., Webb, J.D.C. and Kirk, P.J. (2018) Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations: fine-scale analysis of a hailstorm. *Meteorological Applications*, 25, 472–492. <https://doi.org/10.1002/met.1715>.
- Danish Environmental Protection Agency. (2015) *Danmarks Højdemodel, DHM/Terræn* (Technical Report No. Spec_DHM/T_V2). Aalborg: Danish Geodata Agency.
- Driesenaar, T. (2009) *General description of the HARMONIE-AROME model*. Available at: <http://hirlam.org/index.php/hirlam-programme-53/general-model-description/mesoscale-harmonie> [Accessed 20th June 2018].
- European Union. (2018) Regulation (EU) 2016/679 of the European Parliament and of the Council. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> [Accessed 13th June 2018].
- Eyre, J.R. (2016) Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties: observation bias correction schemes in data assimilation systems. *Quarterly Journal of the Royal Meteorological Society*, 142, 2284–2291. <https://doi.org/10.1002/qj.2819>.
- Howe, J. (2006) *Crowdsourcing: A definition*. Available at: http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html [Accessed 8th August 2018].
- Hsieh, W.W. (2009) *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge; New York, NY: Cambridge University Press.
- Kim, N.-Y., Kim, Y.-H., Yoon, Y., Im, H.-H., Choi, R.K.Y. and Lee, Y.H. (2015) Correcting air-pressure data collected by MEMS sensors in smartphones. *Journal of Sensors*, 2015, 1–10. <https://doi.org/10.1155/2015/245498>.
- Kim, Y.-H., Ha, J.-H., Yoon, Y., Kim, N.-Y., Im, H.-H., Sim, S. and Choi, R.K.Y. (2016) Improved correction of atmospheric pressure data obtained by smartphones through machine learning. *Computational Intelligence and Neuroscience*, 2016, 1–12. <https://doi.org/10.1155/2016/9467878>.
- Lonobox. (2018) *Lonobox*. Available at: <http://lonobox.com/> [Accessed 12th December 2018].
- Madaus, L.E. and Mass, C.F. (2017) Evaluating smartphone pressure observations for mesoscale analyses and forecasts. *Weather and Forecasting*, 32, 511–531. <https://doi.org/10.1175/WAF-D-16-0135.1>.
- McNicholas, C. and Mass, C.F. (2018) Smartphone pressure collection and bias correction using machine learning. *Journal of Atmospheric and Oceanic Technology*, 35, 523–540. <https://doi.org/10.1175/JTECH-D-17-0096.1>.
- Muller, C.L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R.R. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35, 3185–3203. <https://doi.org/10.1002/joc.4210>.
- Netatmo SAS. (2018) *Netatmo weathermap. Netatmo*. Available at: <https://weathermap.netatmo.com/> [Accessed 18th June 2018].
- Nurmi, P. (2003) *Recommendations on the verification of local weather forecasts*. Reading: European Centre for Medium-Range Weather Forecasts. <https://doi.org/10.21957/y1z1thg51>.
- Olsen, B.T., Korsholm, U.S., Petersen, C., Nielsen, N.W., Sass, B.H. and Vedel, H. (2015) On the performance of the new NWP nowcasting system at the Danish Meteorological Institute during a heavy rain period. *Meteorology and Atmospheric Physics*, 127, 519–535. <https://doi.org/10.1007/s00703-015-0388-y>.
- Overeem, A., R., Robinson, J.C., Leijnse, H., Steeneveld, G.J., P. Horn, B.K. and Uijlenhoet, R. (2013) Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40, 4081–4085. <https://doi.org/10.1002/grl.50786>.
- Price, C., Maor, R. and Shachaf, H. (2018) Using smartphones for monitoring atmospheric tides. *Journal of Atmospheric and Solar-Terrestrial Physics*, 174, 1–4. <https://doi.org/10.1016/j.jastp.2018.04.015>.
- Roberts, N.M. and Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97. <https://doi.org/10.1175/2007MWR2123.1>.
- Robustelli, U., Baiocchi, V. and Pugliano, G. (2019) Assessment of dual frequency GNSS observations from a Xiaomi Mi 8 Android smartphone and positioning performance analysis. *Electronics*, 8, 91. <https://doi.org/10.3390/electronics8010091>.
- Scharling, M. and Rajakumar, K. (2003) *Teknisk rapport 03-32, Stationsoversigter* (Technical Report No. 03-32). Copenhagen: Danish Meteorological Institute.
- Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., Huang, X.-Y. and Duda, M. (2008) A description of the advanced research WRF version 3. Boulder, CO: University Corporation for Atmospheric Research/National Center for Atmospheric Research. <https://doi.org/10.5065/D68S4MVH>.
- Vaisala (2018) *BAROCAP Digital Barometer PTB330*. Available at: <https://www.vaisala.com/sites/default/files/documents/PTB330-Datashet-B210708EN-F.pdf> [Accessed 25 May 2018].
- Vasiljevic, D., Andersson, E., Isaksen, L. and Garcia-Mendez, A. (2006) Surface pressure bias correction in data assimilation.

Reading: European Centre for Medium-range Weather Forecasts.
<https://doi.org/10.21957/uv295rfmx5>.

WMO. (2014) *Guide to Meteorological Instruments and Methods of Observation*. Geneva: World Meteorological Organization.

Wunderground (2018) *Wunderground PWS*. Available at: <https://www.wunderground.com/weatherstation/overview.asp> [Accessed 12th December 2018]

Yang, X., Stig, B., Dahldom, M., Sass, B., Zhuang, S., Amstrup, B., Petersen, C., Nielsen, K., Nielsen, N. W. and Mahura, A. (2017) NEA, the Operational Implementation of HARMONIE 40h1.1 at DMI 104-111. *Aladin-HIRLAM Newsletter*, No. 8, pp. 104–111.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Hintz KS, Vedel H, Kaas E. Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation. *Meteorol Appl*. 2019; 1–14. <https://doi.org/10.1002/met.1805>

Publication C

Title: Collecting and utilising crowdsourced data for NWP: Propositions from the meeting held in Copenhagen, 4-5 December 2018.

Submitted: February 2019.


Authors: K. S. Hintz, K. O'Boyle, S. L. Dance, S. Al Ali, I. Ansper, D. Blaauboer, M. Clark, A.Cress, M. Dahoui, R. Darcy, J. Hyrkkänen, L. Isaksen, E. Kaas, U. S. Korsholm, M.Lavanant, G. Le Bloa, E. Mallet, C. McNicholas, J. Onvlee-Hooimeijer, B. Sass, V. Siirand, H. Vedel, J. A. Waller, X. Yang.

Journal: Atmospheric Science Letters.

Journal Website: <https://rmets.onlinelibrary.wiley.com/journal/1530261X>

Status: Published

Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4-5 December 2018

Kasper S. Hintz¹  | Katharine O'Boyle² | Sarah L. Dance⁴ | Saja Al-Ali⁸ | Ivar Ansper⁹ | Dick Blaauboer⁵ | Matthew Clark² | Alexander Cress⁷ | Mohamed Dahoui⁶ | Rónán Darcy¹¹ | Juhana Hyrkkänen¹⁰ | Lars Isaksen⁶ | Eigil Kaas¹² | Ulrik S. Korsholm¹ | Marion Lavanant⁸ | Gwenaelle Le Bloa⁸ | Emilie Mallet⁸ | Conor McNicholas³ | Jeanette Onvlee-Hooimeijer⁵ | Bent Sass¹ | Valeria Siirand⁹ | Henrik Vedel¹ | Joanne A. Waller⁴ | Xiaohua Yang¹

¹Danish Meteorological Institute, Copenhagen, Denmark

²Met Office, Exeter, UK

³Department of Atmospheric Sciences, University of Washington, Seattle, Washington

⁴University of Reading, Reading, UK

⁵Royal Netherlands Meteorological Institute, De Bilt, the Netherlands

⁶European Centre for Medium-Range Weather Forecasts, Reading, UK

⁷Deutsche Wetterdienst, Offenbach, Germany

⁸Météo-France, Paris, France

⁹Estonian Environment Agency, Tallinn, Estonia

¹⁰Finnish Meteorological Institute, Helsinki, Finland

¹¹Met Éireann, Dublin, Ireland

¹²University of Copenhagen, Niels Bohr Institute, Copenhagen, Denmark

Correspondence

Kasper S. Hintz, Danish Meteorological Institute, Copenhagen, Denmark.
Email: kah@dmi.dk

Funding information

Innovation Fund Denmark, Grant/Award Number: 5189-00042B; UK EPSRC, Grant/Award Number: EP/P002331/1

Abstract

In December 2018, the Danish Meteorological Institute organised an international meeting on the subject of crowdsourced data in numerical weather prediction (NWP) and weather forecasting. The meeting, spanning 2 days, gathered experts on crowdsourced data from both meteorological institutes and universities from Europe and the United States. Scientific presentations highlighted a vast array of possibilities and progress being made globally. Subjects include data from vehicles, smartphones, and private weather stations. Two groups were created to discuss open questions regarding the collection and use of crowdsourced data from different observing platforms. Common challenges were identified and potential solutions were discussed. While most of the work presented was preliminary, the results shared suggested that crowdsourced observations have the potential to enhance NWP. A common platform for sharing expertise, data, and results would help crowdsourced data realise this potential.

KEYWORDS

citizen science, crowdsourcing, data collection, opportunistic data, quality control, third-party data

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Atmospheric Science Letters* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

1 | INTRODUCTION

Within the atmospheric sciences, “crowdsourced” data is a relatively new term. While the term crowdsourcing was initially defined by Howe (2006) as outsourcing an act to the general public, this definition is no longer restricted to traditional tasks being outsourced. Today, crowdsourcing is more than outsourcing data collection to the general public. Instead, crowdsourcing embraces new data sources, data storage, quality control and utilisation, which requires standard methods and a common terminology.

Direct and indirect observations from non-conventional sources are being investigated for use in the atmospheric sciences. Examples of data sources include Personal Weather Stations (PWSs) (Bell *et al.*, 2013, 2015; Clark *et al.*, 2018), smartphones (Kim *et al.*, 2015; McNicholas and Mass, 2018; Price *et al.*, 2018; Hintz *et al.*, 2019), vehicles (Anderson *et al.*, 2012; Mahoney and O'Sullivan, 2013) and communication networks (Zinevich *et al.*, 2009).

Muller *et al.* (2015) provided a comprehensive review of “crowdsourcing” efforts in the atmospheric sciences. Since this review was published, new advancements have been made with crowdsourced datasets. Some of the most recent advancements include the collection and quality-control of atmospheric pressure observations from smartphones (Kim *et al.*, 2015, 2016; Madaus and Mass, 2017; McNicholas and Mass, 2018; Price *et al.*, 2018; Hintz *et al.*, 2019). Kim *et al.* (2016) was the first to apply machine-learning methods to bias correct smartphone pressure observations (SPOs). McNicholas and Mass (2018) demonstrated an efficient machine-learning approach to SPO bias correction that benefited from non-meteorological smartphone sensor data. Clark *et al.* (2018) examined the use of PWSs and made considerable progress in the quality control and use of such data. Examples of successful assimilation of such observations into operational Numerical Weather Prediction (NWP) models are currently few and far between. The integration of observations from PWSs into the NOAA Meteorological Assimilation Data Ingest System (NCEP, 2019) dataset is an early example. Also, in the U.S., the utility of PWSs has been of increasing interest for forecasts of severe convection (Madaus *et al.*, 2014; Carlaw *et al.*, 2015; Sobash and Stensrud, 2015; Gasperoni *et al.*, 2018).

A meeting on the use of crowdsourced data in NWP and weather forecasting was held in Copenhagen 4-5th December 2018 at the Danish Meteorological Institute (DMI), with two main purposes. First, to gather experts within the topics of crowdsourcing and create a network of people working on the subject, and second to discuss common issues encountered with crowdsourced data and how these can be addressed. Researchers from both universities and meteorological institutes attended the meeting, whose

experience spanned a variety of subjects, including SPOs, PWSs, vehicular data, and citizen weather reports. The first day was allocated for presentations from the participants, followed by plenary discussion. The second day was allocated for discussions, starting with a sketch of ongoing activities at Institutions and Universities. Two working groups were created who reviewed current research topics for various data sources and data formats. The purpose of this article is to document the propositions and recommendations from the meeting and to inform peers of ongoing activities.

2 | SCIENTIFIC PRESENTATIONS

C. McNicholas (University of Washington) discussed how measurements of atmospheric pressure could be efficiently retrieved from smartphones and subsequently bias-corrected. Results from a testbed Android app, uWx, revealed that inaccuracies in smartphone location and sensor internal filtering contributed to poor data quality. Correcting these issues facilitated the retrieval of pressure change without the need for post-processing/quality control. Using a machine learning approach, smartphone pressures were bias-corrected to account for large uncertainties in smartphone elevation (McNicholas and Mass, 2018). For each smartphone, a random forest was trained on auxiliary sensor/GPS data to predict and correct pressure errors. On average, bias correction reduced pressure errors by ~ 80%. During post-processing, fewer than 20% of smartphone pressure were discarded. In a real-world case-study bias-corrected smartphone pressures improved analyses and 1-hour forecasts of altimeter setting, 2-minute temperature, and 2-minute dewpoint.

K. S. Hintz (DMI) first presented a study on wind measurements from smartphones, in which the surface roughness length was estimated from the measured horizontal turbulence. In another work, more than 6 million SPOs were collected over 7 weeks through a software development kit installed in a third-party mobile app. These observations were quality controlled and assimilated with 3D-Var in the DMI HARMONIE NWP system (Hintz *et al.*, 2019; Yang *et al.*, 2017). A decrease of bias and no change in root mean squared error was found for a simulation period of nearly 2 months. Examples showing that raw observations can depict current weather was given.

X. Yang (DMI) presented the construction idea behind the operational COntinuous Mesoscale Ensemble Prediction System (COMEPS) (Yang *et al.*, 2017a, 2017b) at DMI used for a routine weather forecasts, which generates a 2.5 km grid resolution, 25 member, Rapid Update Cycle (RUC) like EPS forecast with an hourly update using time lagging. Currently, a proto-type ensemble nowcasting system applying the COMECS approach is in development, targeting sub-

hourly cycling of the high-resolution nowcasting system assimilating high frequency observation data such as radar and crowdsourced data. One of the novel system components in COMEPS is the time-lagged 3D-Var analysis on overlapped observation windows, which appears especially beneficial in nowcasting applications with variational data assimilation, as the setup appears to have better potential to address observation error correlation in time and space, as well as the issue of model spin-up in connection with frequent assimilation cycling.

A. Cress (Deutscher Wetterdienst, DWD), presented the activities of DWD concerning crowdsourced data applications and their use in the local DWD data assimilation system. Within the Fleet Weather Maps Project FloWKar, a collaboration between DWD and the German car manufacturer AUDI AG has been established, to investigate to what extent future environmental observations from vehicle sensors can be combined with existing data sources to improve nowcasting and warnings and therefore make a contribution to the security of future autonomous driving. A complete real-time weather conceptual framework has been established, focusing on the flow and processing of high resolution measurements and weather products and the development of corresponding forecasts. A fast data exchange is followed by quality control according to weather service standards and smart aggregation strategies, integrating all available data into a real-time weather map. Aiming for fast weather forecasting, a data assimilation cycle with a 5-minute update rate is necessary; therefore, an ultra-rapid data assimilation method is proposed. A real-world application employs the high resolution project observations in a 5-minute assimilation cycle for the regional operational weather model COSMO-D2, focusing on the model performance optimisation near the surface and its predictions along road sections in Germany, where the current observation network is not dense enough. First results, comparing car measurements, nearby weather stations and model analysis and forecasts were presented.

E. Mallet and S. Al Ali (Météo France) first gave a brief overlook of crowdsourcing activities at Météo-France. Those activities focus on the use of human observations from “expert” non-professional observers and “citizen” observers, and automated observations collected from PWSs, agricultural networks and connected vehicles. Then, the presentation focused on two ongoing projects: (1) The first action concerns the crowdsourcing module in Météo-France’s mobile application that allows mobile users to report the observed weather and to post pictures of the sky. The module provides, without access restriction, a simple entry of almost twenty phenomena to the users, who in their turn will select the observed phenomena and report the observed weather condition on a regular basis. Based on this module,

more than 10,000 observations are collected daily, and more than 40,000 in high-risk situations. Visualisation of crowdsourced data is already available to forecasters, and the next step is to feed it to operational databases in order to expand its possible uses. (2) The second action concerns the potential use of vehicle observations for meteorological applications which is the subject of a partnership between Météo-France and Continental. The aim is to infer weather (precipitation and low visibility) and road conditions (dry, wet, slick) at a particular location in time, through the analysis of vehicle data elements (temperature, wiper and headlight statuses, velocity, and the activation of ABS and ESP systems). The experimental campaign started in November 2016 and is still ongoing. The fleet consists of hundreds of vehicles, transmitting data through a connected dongle. Data filtering and quality checking routines were developed, and vehicle observations were evaluated against meteorological data. Machine learning classification algorithms were developed, using data from meteorological observation merging products as references for hydro-meteor discrimination and visibility. The preliminary results were promising and also showed the need to combine multiple parameters in order to successfully derive weather observations.

K. O’Boyle (Met Office) presented how The Met Office view crowdsourcing as distinct from citizen science (see section 4). There is a long history of citizen science at the Met Office. The Weather Observations Website (WOW) (wow.metoffice.gov.uk) is the Met Office citizen science portal. WOW has global reach, and is a platform for anybody to submit, share and display their weather observations, either manually or by connecting a PWS using APIs. WOW data is being trialled in nowcasting applications, but is not yet assimilated into NWP. Investigations into other opportunistic observations are ongoing at the Met Office, including collecting data from vehicles.

M. Clark (Met Office) presented on an automated quality control and gridding process for citizen science data. There has been a focus on Met Office WOW data from PWSs to create high resolution surface analyses. Parameter values from each WOW site are constrained to have the same long-term mean as neighbouring official sites, but are otherwise allowed to vary freely, as it is assumed that shorter-term, temporary deviations are the signature of genuine small scale features which are worth retaining in the analysis. A series of case studies have shown that there is value in this approach.

S.L. Dance (University of Reading) gave an overview of the DARE: Data Assimilation for the REsilient City project. This is a UK Engineering and Physical Sciences Research Council (EPSRC) Senior Fellowship in Digital Technology for Living with Environmental Change. The vision for the project is to use “datasets of opportunity”, such as CCTV

and vehicle observations, alongside scientific observing networks, such as satellite data (Mason *et al.*, 2018) to improve predictions of urban natural hazards such as flooding and high impact weather. There are many potential benefits of such data, including the availability of large numbers of inexpensive observations, in areas where there are people but there may be few sources of scientific observation data. For example, (1) air traffic management reports have potential to provide observations of temperature inversions in the boundary layer (Mirza *et al.*, 2016, 2019). (2) In many locations around the world, the population has access to smartphones, but ground-based scientific observations are sparse. Furthermore, there are a number of issues in collecting ‘datasets of opportunity’ for use in assimilation. These include the need for metadata such as time and location in order to carry out the assimilation, versus data protection for the data provider, who may be a private individual. Other issues include data ownership, intermittency, heterogeneity, data provenance and large data volumes. In order to use such observations in data assimilation, there needs to be an understanding of natural variability in urban areas (where many of these data originate) and the variability that can be resolved by a prediction model (e.g., Waller *et al.*, 2014; Janjić *et al.*, 2017). This was discussed further in the next talk by J.A. Waller.

J. A. Waller (University of Reading) presented on the potential to measure temperatures in urban areas using vehicles. Issues related to the assimilation of crowdsourced data were discussed; in particular, the need to understand the data inhomogeneity and natural variability of observation urban areas in order to understand the observation uncertainties. Collaborative work with the UK Met Office, is assessing the potential of temperature observations recorded by vehicles. The preliminary findings showed that the data collection method was not reliable for collecting large temperature data sets. Furthermore, for the initial data sets collected, it was shown that temperature measurements had a negative correlation with the speed of the vehicle. It was concluded that a new data collection technique was required, and a more detailed study was vital before the benefits of assimilating vehicle temperatures could be assessed.

D. Blaauboer (KNMI and EUMETNET) presented shortly the KNMI-activities in the domain of crowdsourcing. These include participation in the WOW project of UK Met Office, application of car data (temperature sensor, wiper data), smartphone data, damage reporting app (to report weather impacts by the public), wind data from hot air balloons. EUMETNET, the grouping of 31 European National MetServices, recognised the emerging availability and application opportunities of crowdsourced data and the Internet of Things among many of its members. Therefore EUMETNET has organised a few dedicated workshops on

this subject with the aim to bring experts in this field together, to foster networking and possibly create a platform or programme in near future to develop common applications to the benefit of all.

M. Dahoui (ECMWF) presented an overview of the importance on in-situ data in global NWP. It was shown that there are data gaps in the surface observations received at ECMWF and the potential and challenges for using crowdsourced data to fill these gaps were described. Also, it was stressed that crowdsourced data can be important for verification purposes. A denser network is useful to detect small scale features and rapid changes of the atmosphere, so observations have also the potential to improve the forecast verification aspects leading to a better understanding of model performance. The usage of crowdsourced observations is however very challenging. It was suggested that data collection and pre-processing needs a collaborative effort between NWP centres through coordination of the WMO, the industry and the private sector to improve and unify standards and to agree on best practices. A common and shared use of operationally managed data hubs (such as the MetOffice Weather Observation Website) is a cost-effective solution to manage the diversity of data sources and formats. A good understanding of the error characteristics of the observations is necessary to allow proper data selection and error specification. This requires a comprehensive and standardised description of metadata. Quality control, bias correction and blacklist management require unique identification of a reporting station which makes anonymous reports of less interest to NWP data assimilation unless technological solutions are available to anonymously identify the data or perform most of the quality control and bias correction near the data origin. Legal aspects related to privacy and data usage are also essential to clarify before the operational use of such observations.

3 | OVERVIEW OF ACTIVITIES

During the meeting, it became clear that there are many activities on-going, with opportunities for collaboration. Table 1 list activities, status and considerations for participating institutions together with ZAMG and Met Norway who agreed to share their current activities. It is seen that especially work with data from private weather stations is an active field of research at many institutions.

4 | CHALLENGES AND SOLUTIONS

The presentations and discussions identified several common challenges, and some solutions were proposed during the discussion sessions. These follow below:

TABLE 1 Overview of ongoing and considered activities at each participating institute and institutes that was not present but approved to be included

Institution	Current activities	Activity status	Considerations	Contact persons
KNMI	WOW-NL	Research, operational	Pollution measurements	Marijn De Haij
Met Office	WOW-UK, social media, cars, voluntary observations	Research, operational	User reports, 5G network	Katharine O'Boyle
DMI	SPO, PWS	Research	User reports, webcam	Kasper Hintz
FMI	User reports	Operational	SPO	Juhana Hyrkkänen
Météo France	Cars, PWS, user reports	Research		Émilie mallet
University of Reading	Cars, CCTV, WOW	Research	Buses	Sarah Dance
ECMWF	Monitoring Progress	Research		Mohamed Dahoui
DWD	Cars, PWS	Research	SPO, user reports	Alexander cress
Met Éireann	Voluntary observations	Research, operational	PWS	Ronan Darcy
ESTEIA	App in development			Ivar Anspér
University of Washington	SPO	Research		Conor McNicholas
ZAMG	Trusted spotter network, Austrian weather observer	Operational		Thomas Krennert
Met Norway	PWS (Netatmo)	Research, operational		Roger Randriamampianina

The activities at University of Reading are only including activities within the DARE (data assimilation for the REsilient city) project.

- I. Terminology is not agreed upon in the community. A common vocabulary needs to be established to facilitate future collaboration. The term “crowdsourced data” is used differently within the community, and there is no agreement what this term covers and what not. Often crowdsourced data is used as a collective term, for example, citizen-science and third-party data, which is how the term will be treated in this report, though with a recognition that a more precise definition is desirable.
- i. The Met Office suggested a terminology that clearly separates citizen-science data and crowdsourced data, and also attempts to define associated terms:
 - a. Citizen-science data: Information obtained from a group of people who are invited to participate in a data collection process.
 - b. Crowdsourced data: Information derived from a group of people without their explicit involvement in the data collection process.
 - c. Opportunistic data: Information derived from non-meteorological sensors or weather sensitivities.
 - d. Third-party data: Data collected by a third-party organisation using meteorological sensors.
- However, some similarities are expected between third-party data and the other groups. For example PWS observations might be classified as both third-party data and crowdsourced data.
- ii. ECMWF proposed four main categories of ‘crowdsourced’ data; private and third party, automated amateur weather stations, smart connected devices (mobile phones and vehicles), and human reporting of the current weather, relating each of these to the ease of utility in NWP.
- In the terminology proposed by the Met Office (i), there is a clear separation between citizen-science data and ‘crowdsourced’ data, wherein the ECMWF proposal (ii) the term ‘crowdsourced’ data is a collective term. It is

recommended that authors define their usage of these terms.

- II. Obtaining useful crowdsourced data may involve collaboration with commercial entities, such as manufacturers of PWSs or vehicles. In some cases, collaborations of this kind mark a step change in the way universities and meteorological institutes have previously operated. For professional use, crowdsourced data needs to be as unprocessed as possible when received. Working in collaboration with manufacturers may enable this. Some of the workshop participants have built successful collaborations with commercial entities, taking a “virtuous circle” approach, whereby data is provided by a manufacturer, and in return the meteorological institution provides forecast data or quality controlled observational data. It is crucial that intellectual property rights and data ownership are clear and agreed upon before starting collaborations.
- III. Law based restrictions on storage of personal data lead to a need to de-personalise crowdsourced data, which can lead to “black boxes”. Metadata can be used to help characterise the error of crowdsourced observations, and for bias correction, but the legal constraints regarding privacy and personal data can limit the collection of such metadata. Hence, metadata vs privacy is one issue that must be considered when collecting observations. DMI have invested in legal expertise and are open to sharing the information obtained with the community. This is mainly related to the European GDPR regulation (European Union, 2018).
- IV. New data sources can potentially produce more observations than current NWP models can realistically handle. New methods, such as those suggested by Dr. X. Yang (DMI), will need to be considered. Tendencies of parameters are not commonly assimilated into NWP; a change in approach may be required to extract maximum value from crowdsourced observations.
 - i. It was discussed that data streaming could be a way of handle the amount of observations in future, such that, in operational systems, observations that come in are utilised and then thrown away. This may seem somewhat provocative to some as the NWP community are often used to store data for an extended time. However, it was agreed that data streaming could perhaps be only realistic solution currently to overcome issues with data volume. Also near-real communication could perhaps be easier to implement with a streaming approach.

- ii. The scale of crowdsourced observations, any reference network, and NWP models will all be different. To make them comparable, methods to deal with multiscale comparisons are required for example, filtering or superobbing.

Further, other themes seemed to be well established. There was a general agreement that crowdsourced data can provide useful observations in areas otherwise devoid of observations. It was discussed whether stationary platforms (e.g., PWS) are easier to implement in existing systems than moving platforms (e.g., vehicles, SPOs). In general, stationary platforms are believed to be easier to bias-correct than moving platforms. Also, new data sources should be seen to supplement conventional observation networks rather than a replacement, as trusted observations are required as a reference for new data sources. A nested platform of reference may be a good way of organising networks in the future, for example, SYNOPs used as a reference for the quality control of PWS data, and PWS then used as a more dense reference dataset for observations from mobile platforms.

5 | CONCLUSIONS AND RECOMMENDATIONS

Much of the work presented at the workshop was at an early, exploratory stage, and many questions remain unanswered. However, a general set of conclusions were drawn from the discussion. Crowdsourced observations are potentially useful for NWP, and are undoubtedly useful for verification and forecasting. Use of crowdsourced observations in nowcasting, or post-processing, is perceived to be easier and less demanding than in NWP data assimilation. There is still much work to do before crowdsourced observations can widely be ingested into NWP models.

It was agreed upon that there is a sliding scale between ‘crowdsourced’ or ‘passive’ data collection, where an individual’s involvement is limited, and ‘citizen science’ or ‘active’ data collection where the individual is explicitly involved. It is generally thought that the lesser degree of interaction required by the participant the higher the volume of data that can be collected. It is not clear if either of the two are of superior quality.

Further, the following recommendations are made. An organised community of those involved in crowdsourcing activities would be beneficial. EUMETNET would provide a good forum for this, however, such a forum should not be restricted to European countries. This forum could be a simple, independent, platform accessible via a website. Regarding vocabulary, it would be beneficial for the community to agree on common terminology related to crowdsourcing. To realise the full potential of crowdsourced data for NWP, issues of data quality, privacy, and availability will need to

be addressed. Data quality could be enhanced by prioritising the collection of accurate metadata. Privacy issues should be addressed to determine if, how, and when unique identifiers can be retrieved for quality control purposes. Lastly, efforts to expand crowdsourced datasets by disseminating data operationally and working with private industry should be encouraged.

ACKNOWLEDGEMENTS

The authors wish to thank DMI meteorologist Ole Kristensen for giving an operational aspect on the use of crowdsourced data. The contribution of S.L. Dance and J.A. Waller was funded in part by UK EPSRC Grant No. EP/P002331/1 Data Assimilation for the Resilient City (DARE).

ORCID

Kasper S. Hintz  <https://orcid.org/0000-0002-6835-8733>

REFERENCES

- Anderson, A.R.S., Chapman, M., Drobot, S.D., Tadesse, A., Lambi, B., Wiener, G. and Pisano, P. (2012) Quality of mobile air temperature and atmospheric pressure observations from the 2010 development test environment experiment. *Journal of Applied Meteorology and Climatology*, 51(4), 691–701.
- Bell, S., Cornford, D. and Bastin, L. (2013) The state of automated amateur weather observations. *Weather*, 68(2), 36–41.
- Bell, S., Cornford, D. and Bastin, L. (2015) How good are citizen weather stations? Addressing a biased opinion. *Weather*, 70(3), 75–84.
- Carlaw, L.B., Brotzge, J.A. and Carr, F.H. (2015) Investigating the impacts of assimilating surface observations on high-resolution forecasts of the may 15, 2013 tornado event. *Electronic Journal of Severe Storms*, 10(2), 1–34.
- Clark, M.R., Webb, J.D.C. and Kirk, P.J. (2018) Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorological Applications*, 25(3), 472–492.
- European Union (2018). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>. (accessed June 16, 2018).
- Gasparoni, N.A., Wang, X., Brewster, K.A. and Carr, F.H. (2018) Assessing impacts of the high-frequency assimilation of surface observations for the forecast of convection initiation on April 3, 2014 within the dallasfort worth test bed. *Monthly Weather Review*, 146(11), 3845–3872.
- Hintz, K.S., Vedel, H. and Kaas, E. (2019) Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation. *Meteorological Applications*, 2019, 1–14. <https://doi.org/10.1002/met.1805>.
- Howe, J. (2006). Crowdsourcing: A definition. http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html. (accessed September 22, 2018).
- Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A. and Weston, P. (2017) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144 (713), 1257–1278.
- Kim, N.-Y., Kim, Y.-H., Yoon, Y., Im, H.-H., Choi, R.K.Y. and Lee, Y.H. (2015) Correcting air-pressure data collected by mems sensors in smartphones. *Journal of Sensors*, 2015, 1–10.
- Kim, Y.-H., Ha, J.-H., Yoon, Y., Kim, N.-Y., Im, H.-H., Sim, S. and Choi, K.Y.R. (2016) Improved correction of atmospheric pressure data obtained by smartphones through machine learning. *Computational Intelligence and Neuroscience*, 2016, 12. <https://doi.org/10.1155/2016/9467878>.
- Madaus, L.E., Hakim, G.J. and Mass, C.F. (2014) Utility of dense pressure observations for improving mesoscale analyses and forecasts. *Monthly Weather Review*, 142(7), 2398–2413.
- Madaus, L.E. and Mass, C.F. (2017) Evaluating smartphone pressure observations for mesoscale analyses and forecasts. *Weather and Forecasting*, 32(2), 511–531.
- Mahoney, W.P. and O'Sullivan, J.M. (2013) Realizing the potential of vehicle-based observations. *Bulletin of the American Meteorological Society*, 94(7), 1007–1018.
- Mason, D.C., Dance, S.L., Vetra-Carvalho, S. and Cloke, H.L. (2018) Robust algorithm for detecting floodwater in urban areas using synthetic aperture radar images. *Journal of Applied Remote Sensing*, 12, 12–20.
- McNicholas, C. and Mass, C. (2018) Smartphone pressure collection and bias correction using machine learning. *Journal of Atmospheric and Oceanic Technology*, 35, 523–540.
- Mirza, A.K., Ballard, S.P., Dance, S.L., Maisey, P., Rooney, G.G. and Stone, E.K. (2016) Comparison of aircraft-derived observations with in situ research aircraft measurements. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 2949–2967.
- Mirza, A.K., Ballard, S.P., Dance, S.L., Rooney, G.G. and Stone, E.K. (2019) Towards operational use of aircraft-derived observations: a case study at London heathrow airport. *Meteorological Applications*, 2019, 1–14. <https://doi.org/10.1002/met.1782>.
- Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35(11), 3185–3203.
- NCEP (2019). Madis. <https://madis.ncep.noaa.gov>. (accessed January 17, 2019).
- Price, C., Maor, R. and Shachaf, H. (2018) Using smartphones for monitoring atmospheric tides. *Journal of Atmospheric and Solar-Terrestrial Physics*, 174, 1–4.
- Sobash, R.A. and Stensrud, D.J. (2015) Assimilating surface mesonet observations with the enkf to improve ensemble forecasts of convection initiation on may 29, 2012. *Monthly Weather Review*, 143 (9), 3700–3725.
- Waller, J.A., Dance, S.L., Lawless, A.S., Nichols, N.K. and Eyre, J.R. (2014) Representativity error for temperature and humidity using the met Office high-resolution model. *Quarterly Journal of the Royal Meteorological Society*, 140(681), 1189–1197.
- Yang, X., Feddersen, H., Sass, B. and Sattler, K. (2017a) *Construction of a Continuous Mesoscale Eps with Time Lagging and Assimilation on Overlapping Windows*, ALADIN-HIRLAM Newsletter, Vol 8, pp. 112–118.

- Yang, X., Stig, B., Dahldom, M., Sass, B., Zhuang, S., Amstrup, B., Petersen, C., Nielsen, K., Nielsen, N.W. and Mahura, A. (2017b) *NEA, the Operational Implementation of HARMONIE 40h1.1 at DMI*, ALADIN-HIRLAM Newsletter, Vol 8, pp. 104–111.
- Zinevich, A., Messer, H. and Alpert, P. (2009) Frontal rainfall observation by a commercial microwave communication network. *Journal of Applied Meteorology and Climatology*, 48(7), 1317–1334.

How to cite this article: Hintz KS, O'Boyle K, Dance SL, *et al.* Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4-5 December 2018. *Atmos Sci Lett.* 2019;e921. <https://doi.org/10.1002/asl.921>

Publication D

Title: Evaluating pressure observations from private weather stations and smartphones.

Submitted: In Prep.

Authors: K. S. Hintz, H. Vedel, E. Kaas.

Journal: Meteorological Applications.

Journal Website: <https://rmets.onlinelibrary.wiley.com/journal/14698080>

Status: In-Prep. for Meteorological Applications.

Evaluating pressure observations from private weather stations and smartphones

K. S. Hintz^{a,*}, E. Kaas^b, H. Vedel^a

^a*Danish Meteorological Institute, Copenhagen, Denmark*

^b*University of Copenhagen, Niels Bohr Institute, Denmark*

Abstract

The amount of publicly available data from personal weather stations has increased significantly in recent years. One popular weather station in Europe is sold by the private company Netatmo. Pressure observations from Netatmo stations have been collected, and the quality of these has been examined. For three individual stations, a detailed examination of data spanning a year was carried out. A non-constant bias was found for all three stations, following the same pattern. Further, the observations were assimilated with 3D-Var using the HARMONIE numerical weather prediction model. A numerical weather prediction model has been run for a period of two months, in which both pressure observations from Netatmo and smartphones were assimilated using 3D-Var. It is recommended that high-density spatial surface observations are averaged to 'super-observations' or that a data thinning method is applied to surface observations when necessary.

Keywords: Crowdsourcing, Citizen Science, Quality Control, Data Collection, Assimilation

*Corresponding author: K.S. Hintz, kah@dmi.dk

1 **1. Introduction**

2 Within recent years, it has become increasingly easy for people to connect devices to the inter-
3 net (Patric, 2014). Such devices include smartphones and Personal Weather Stations (PWS). Many
4 smartphones can measure the atmospheric pressure, which is of interest for data assimilation in Nu-
5 merical Weather Prediction (NWP) (Madaus and Mass, 2017; McNicholas and Mass, 2018; Hintz
6 et al., 2019a). Also, PWS offers a new source of high spatial resolution of traditional meteorological
7 surface variables (Clark et al., 2018), such as pressure and temperature. Many other activities are
8 currently ongoing with other types of data (Muller et al., 2015; Hintz et al., 2019b). One popular
9 internet-connected PWS is Netatmo weather stations (Netatmo SAS, 2018). Other PWS includes
10 stations from e.g. Lonobox (2019) and Fieldsense A/S (2019). But many more exists (Clark et al.,
11 2018; Bell et al., 2015).

12
13 This study examines the quality of pressure observation from PWSs using stations by Netatmo
14 SAS (2018) as a source. Figure 1 shows an example of the high spatial resolution that these stations
15 can offer. In this case, temperature observations are shown. Such high-resolution data can be to
16 interest for nowcasting purposes and urban meteorology (Meier et al., 2017). Pressure observations
17 from PWS are assimilated into the HARMONIE NWP model for a period of two months, using
18 3D-Var, to study the impact of these observations. Also, pressure observations from smartphones
19 are assimilated in another experiment covering the same period, extending the simulation period
20 of an earlier study, Hintz et al. (2019a), which was five days. Hintz et al. (2019a) collected about
21 150.000 SPOs per day via a developed Software Development Kit (SDK) integrated into third-party
22 apps, which was assimilated into the HARMONIE NWP model (Yang et al., 2017). A vital issue of
23 SPOs is that the accuracy of the vertical position from the GNSS-system is poor; however, it is an
24 issue that is likely to reduce in future (Robustelli et al., 2019). An advantage of PWS over SPOs is
25 that they are stationary and thereby easier to bias correct (Hintz et al., 2019b).

26
27 In section 2, the methods for collecting and processing the Netatmo pressure observations are
28 presented. The collection and processing of SPOs are identical to those methods presented in Hintz
29 et al. (2019a) and are thus not described in detail here. Results are presented in section 3 and

30 discussed in section 4. Finally, conclusions and final remarks are given in section 5.

31 **2. Method**

32 First, the quality of Netatmo pressure observations was examined by comparing three individual
33 stations, from which the authors had got access, to SYNOP observations. A default Netatmo
34 weather station consists of two modules, one indoor module and one outdoor module. Both modules
35 measure the temperature and relative humidity. The outdoor module is powered by batteries and
36 is connected to the indoor module via Bluetooth. The indoor module is connected to WiFi and is
37 responsible for sending data to a Netatmo server where the user can access the data via an user
38 interface. The atmospheric pressure is measured indoors by the indoor module because the module
39 is powered via a power cable. The user does not see the observed pressure, but sees the Sea Level
40 Pressure (SLP) converted from the observed pressure. A rough conversion to sea level is made using

$$41 \quad p_s = p_0 \cdot \left(1 - \frac{\Gamma h}{T_0}\right)^{\frac{gM}{R_0}}, \quad (1)$$

42

43 where p_s is the surface pressure and p_0 is the SLP. Γ is the lapse rate for dry air (≈ 0.0065 K/m).
44 T_0 is a constant standard temperature of 288.15 K. g is the gravitational constant, M is the molar
45 mass of dry air (≈ 0.02897 kg/mol) and R_0 is the universal gas constant. h is the height above sea
46 level of the station. The altitude is defined by Netatmo via the horizontal position using a unknown
47 terrain model. Equation (1) was referenced by Netatmo customer support (Netatmo SAS, 2019)
48 via email.

49
50 An examination of individual stations was also performed. Pressure observations from three
51 individual Netatmo indoor modules was obtained over a period of one year, 1 April 2018 to 31
52 March 2019, with a temporal resolution of one hour. Each station was compared to observations
53 from the nearest DMI SYNOP station respectively to examine the bias of the Netatmo pressure
54 observations. To compare the Netatmo pressure observations to the SYNOP stations the SLP from
55 Netatmo was first converted to surface pressure using equation (1) using the elevation from the
56 Netatmo SAS (2018) Application Programming Interface (API). Then the pressure from both the

57 Netatmo and SYNOP station was converted to SLP using (WMO, 2010)

$$58 \quad p_0 = p_s \cdot \exp\left(\frac{\frac{g}{R} H_p}{T_s + \frac{\gamma H_p}{2} + e_s \cdot C_h}\right), \quad (2)$$

59

60 where H_p is the station elevation in gpm, T_s is the station temperature in K, e_s is the vapour
61 pressure at the station in hPa and C_h is a coefficient of 0.12 K/hPa. For stations with an elevation
62 of less than 50 m, a different correction is used following WMO (2010). A reduction coefficient is
63 added to the surface pressure as

$$64 \quad p_0 = p_s + \frac{p_s H_p}{29.27 T_v}, \quad (3)$$

65

66 where T_v is the virtual temperature at the station. However, one can not use the observed indoor
67 temperature and humidity from the Netatmo station to find the SLP. Instead the temperature and
68 humidity was obtained from a HARMONIE short term forecast using the nearest grid point to
69 the station. This 'back-and-forth' correction of Netatmo observations is done to ensure the same
70 methods has been applied and that any bias is not due to differences in layer thickness throughout
71 the year between the stations due to different station elevation.

72

73 Secondly, Netatmo data was obtained from Denmark, northern Germany and southern Sweden
74 via the Netatmo SAS (2018) API from 1st of April 2018 and onwards with a sampling frequency of
75 about 30 minutes, with the purpose of assimilation into the HARMONIE NWP model. Also, SPOs
76 was collected for the same period using the methods described by Hintz et al. (2019a), in which
77 an external screening method was applied before assimilating SPOs. Hintz et al. (2019a) obtained
78 about 120.000 pressure observations from smartphones per day, thus a strict filtering was necessary.

79

80 Four numerical simulations were performed with the DMI HARMONIE NWP model (Yang
81 et al., 2017), listed below

- 82 • OPR,
- 83 • REF,
- 84 • SMAPS,

85 • NETATMO,

86 where OPR is equivalent to the operational model at DMI. REF is a reference run in which pressure
87 observations from Danish SYNOP stations were excluded from the assimilation model to have an
88 independent reference of observations. In SMAPS, SPOs has been assimilated into HARMONIE
89 (similar to EXP_MEDV2 described in Hintz et al. (2019a) but for an extended period). In the
90 NETATMO run, NPOs has been assimilated into HARMONIE. In both SMAPS and NETATMO,
91 there are no pressure observations from Danish SYNOP stations as in REF. The four simulations
92 was run from 4 April 2018 to 25 May 2018, as six-hour forecasts with three hour data assimilation
93 cycles using 3 Dimensional Variational (3D-Var) assimilation. In all cases the surface pressure was
94 assimilated.

95

96 NPOs was bias-corrected using two months of NWP short-term forecasts prior to the simulation
97 period. If new stations entered the system which was not found in the bias-corrected stations it
98 was not used. Otherwise, no prior screening to the data assimilation system was done. However,
99 screening of observations are a part of the data assimilation system of HARMONIE (Driesenaar,
100 2009). Only observations within 30 minutes of the analysis time was used in the model and only
101 one observation per station.

102

103 3. Results

104 Figure 2 shows one year of bias of the SLP from an individual Netatmo station (black line). The
105 Netatmo station was compared to the nearest DMI SYNOP station (Holbaek) 15 km away. The
106 solid black line shows the data resampled over one month. An apparent yearly cycle is seen in the
107 bias. Also, the temperature (red line) and relative humidity (blue line) from the PWS is plotted.
108 Note that all observed variables are measured indoor. It is relevant to examine the temperature
109 and relative humidity indoor as it is also the indoor module that measures the pressure. A Pearson
110 correlation coefficient of 0.96 and 0.67 between the resampled surface pressure and temperature and
111 surface pressure and relative humidity were found respectively (see table 1).

112

113 The SYNOP stations was discarded as a source to the yearly cycle by examining the bias of
114 each SYNOP station to other SYNOP stations and NWP model data, in which no varying bias
115 was found (not shown). Also, the Netatmo station was compared to NWP data, in which the same
116 trend was seen. Two other Netatmo stations was also examined in which a yearly cycle in bias was
117 also seen, however, less visible. Station name, location and correlation coefficients for all stations
118 are given in table 1.

119

120 From table 1, it is seen that there is a generally strong correlation between the bias of resampled
121 surface pressure and the resampled temperature, σ_T^{RES} . Especially for the station of Snuderup, with
122 a positive correlation coefficient of 0.91. For the two other stations the correlations are negative.
123 Correlations are less for relative humidity, σ_H^{RES} . For the raw observations, σ_H^{RAW} , the correlations
124 are smaller and in no cases larger than 0.3 in absolute value.

125

126 Figure 3 shows the distribution of the differences between the observed pressure between one
127 Netatmo station (Snuderup) and the SYNOP station. Netatmo uses a simple algorithm to convert
128 the observed pressure to SLP (see equation (1)) while the observations from the SYNOP station is
129 corrected to SLP by taking both temperature and humidity into account (see equation (2)). Both
130 the bias in p_s and SLP follows a normal distribution which have been fitted to each distribution.
131 The difference of the shape between the two distributions are due to differences in the conversion to
132 SLP as described in section 2. The observations of p_s is coming from the elevation above sea level
133 from both the Netatmo station (62 m) and the SYNOP station (13 m), which makes the bias of sur-
134 face pressure larger than that of SLP, as no correction is done for the elevation difference in this case.

135

136 Figure 4 shows the number of surface pressure observations used in the minimisation of the
137 3D-Var system for each run for OPR, SMAPS and NETATMO from Denmark. REF is not shown
138 as all Danish surface pressure observations were excluded here. It is seen that OPR lies nearly
139 constant around about 50 SYNOP stations. A strict filtering of SPOs was applied in the SMAPS
140 run prior to the data assimilation, which is why a relatively low number of observations is used
141 in the minimisation. The high variability of crowdsourced data compared to SYNOP data is also
142 seen as a relatively large fluctuation in the numbers of observations that enters the minimisation in

143 the case of the SMAPS and the NETATMO runs. Data thinning is only applied to satellite data
144 in the 3D-Var system of HARMONIE and not surface data. Therefore a large number of Netatmo
145 pressure observations compared to OPR and REF are seen.

146

147 Figure 5 shows the RMSE, bias, and standard deviation averaged over all runs for each forecast
148 hour for surface pressure using SYNOP observations as reference. OPR and REF are nearly identi-
149 cal as expected, but it is noted that the REF does have a slightly lower bias than OPR throughout
150 the forecast, which is notable since Danish SYNOP pressure observations are used in the OPR run;
151 however, the difference is minor. Both the SMAPS and NETATMO run starts with a negative bias.
152 SMAPS changes sign and converges towards OPR and REF when the model runs freely. NETATMO
153 has the same tendency but starts with a large bias about four magnitudes larger than OPR with
154 opposite sign. Also, the RMSE of NETATMO is the largest throughout the forecast length. The
155 RMSE of SMAPS is in slightly lower than those of OPR and REF, but the standard deviation is
156 also slightly higher.

157

158 Figure 6 shows the RMSE, bias and standard deviation as figure 5 but for 10 m wind speed using
159 SYNOPs as reference. Again OPR and REF are nearly identical. Both SMAPS and NETATMO has
160 lower bias than OPR and REF but higher RMSE and standard deviation. The most extreme bias
161 difference is seen at the first forecast hour between NETATMO and REF, where NETATMO has a
162 lower bias of about 0.1 m/s. However, NETATMO has a RMSE about 0.2 m/s higher than OPR and
163 REF throughout the forecast, while the difference in bias is less than 0.1 m/s in general. SMAPS
164 also has a lower bias than both OPR and REF, but also has a higher RMSE and standard deviation.

165

166 Figure 7 shows RMSE, bias and standard deviation as figure 5 and figure 6 but for 2 m tempera-
167 ture. SMAPS scores best overall comparing the improvement in bias with a minor difference in both
168 RMSE and standard deviation. The RMSE of SMAPS is has decreased a little, while the standard
169 deviation has increased less than 0.05 K. The standard deviation of OPR, REF and NETATMO
170 are nearly identical and difficult to distinguish on the figure.

171 **4. Discussion**

172 A yearly cycle in the bias of p_0 was identified (see figure 2) for three individual PWS from
173 Netatmo. The variation in bias is likely related to internal processes of the hardware itself since
174 there is a clear correlation with the indoor temperature measured by the same device. It was seen
175 from table 1 that the correlation with temperature for the three stations was both positive and
176 negative. A possible explanation for this is the difference in elevation for the PWS compared to
177 the SYNOPs. For the Snuderup Netatmo station, the station lies higher (49 m) than the Holbaek
178 SYNOP station. Both Kgs. Lyngby and Skelkaeret (Netatmo stations) lie lower than the SYNOP
179 stations with a difference of 39 m and 12 m respectively. It is well known that most barometers
180 do have a temperature dependence (Meulen, 1992), however much can be done to minimise such
181 effects, for example, by using materials that are less affected by temperature changes. It has not
182 been possible to find documentation on what has been done by Netatmo to minimise such effects.

183

184 If a provider of observations, such as Netatmo, delivers SLP and not surface pressure, it is essen-
185 tial to know which method that has been used to convert the observation to sea level. It was shown
186 in figure 3 that the spread of the bias of SLP is larger than that of p_s , for one specific station. The
187 bias of SLP used the Holbaek SYNOP as a reference which was converted to SLP using equation
188 (2) where the Netatmo SLP has been using the more primitive equation (1). A smaller spread is
189 obtained when comparing p_s observed by the SYNOP station to the Netatmo observation, which is
190 converted back to the surface, using equation (1).

191

192 Comparing figure 4 to figure 5 one could suspect that the model has been oversaturated with
193 surface observations. Before the numerical runs, it was believed that automatic data thinning of
194 surface observations were already implemented; however, data thinning is only applied to satellite
195 data in the HARMONIE system. Data thinning is applied to ensure that the assumption of no
196 observation error correlations is true in the 3D-Var system. Because the Netatmo observations have
197 been bias-corrected with two months of NWP data prior to the simulation period, there can be
198 observational error covariances introduced via the NWP model. This can be one reason for the
199 large RMSE of p_s for the NETATMO run seen in figure 5. A decrease in the bias of 10 m wind was

200 shown in figure 6, which could suggest, un-intuitively, that an increase in the magnitude of pressure
201 bias can decrease the bias of 10 m wind. However, this is not the case as the SMAPS run shows a
202 decrease in the bias of both surface pressure and 10 m wind.

203

204 The NETATMO run have in general much larger errors than OPR, REF and SMAPS. Hintz
205 et al. (2019b) describes how mobile devices is more difficult to utilise than stationary devices. There-
206 fore, it was expected to see better results when assimilating Netatmo pressure observations. Further
207 investigations with a difference in the the degree of data thinning should be carried out in future.
208 Such a study will reveal the importance of data thinning in the 3D-Var system and better reflect
209 the impact of data from PWS. However, with the high temporal resolution of crowdsourced data
210 in general, 3D-Var may not be the best suited method. Other methods such as ensemble Kalman
211 filters (Houtekamer et al., 2005) or 4D ensemble variational (4D-EnVar) methods (Lorenc, 2003)
212 may be better suited (Madaus and Mass, 2017; McNicholas and Mass, 2018). A 4D-Var assimilation
213 scheme is under development for HARMONIE (Driesenaar, 2009), which can potentially improve
214 the value of data considered in this study significantly. One temporary solution could be to bias
215 correct the Netatmo stations continuously using the latest one to two months of observations as
216 the bias does not change much in this time interval (see figure 2). Another approach is to gener-
217 ate a set of fewer observations by averaging the observations to generate a set of 'super-observations'.

218

219 As described in section 2, Netatmo pressure observations was bias corrected using NWP data.
220 This can potentially lead to model bias drift (Vasiljevic et al., 2006). However, at DMI, SYNOP
221 pressure observations are never bias corrected, which anchors the model. Even though Danish
222 SYNOP pressure observations were excluded in the REF, SMAPS and NETATMO run, there are
223 still SYNOPS included from outside Denmark, which delivers about 95 % of the total surface pres-
224 sure observations.

225

226 SMAPS look much like the results obtained in Hintz et al. (2019a), in which the period only
227 covered five days. With the increased simulation length and similar results, SPOs must be said to
228 have a positive impact on NWP. This is only likely to increase in future as methods for correcting
229 and screening crowdsourced data improves.

230 **5. Conclusions**

231 A non-constant bias of pressure of three individual Personal Weather Stations (PWS) from
232 Netatmo was found over one year. To conclude on the reasons for the varying bias, further studies
233 are required. References used to determine the bias (SYNOP stations and NWP data) was discarded
234 as a possible source of the observed variation. Also, differences in height between isobaric surface
235 layers due to yearly cycles in temperature were discarded as the source of the variation. Furthermore,
236 the importance of correcting surface pressure to sea level pressure correctly are illustrated. One issue
237 with pressure observation from PWS was a crude algorithm to convert between surface pressure
238 and sea level pressure. This was, however, easily corrected.

239 Pressure observations from PWS were successfully assimilated into the DMI HARMONIE NWP
240 model using 3D-Var in which potential issues of the current assimilation system was identified,
241 such as missing data thinning for high-resolution surface observations. Four numerical experiments
242 using HARMONIE with 3 hourly data assimilation cycles was computed over a period of nearly
243 two months. These computations showed worse results when PWS pressure observations were
244 assimilated compared to the operational model. It is expected that improvements in the observation
245 preprocessing can increase the skill of the run with PWS observations significantly.

246 A Previous study is confirmed in that smartphone pressure observations do contribute positively
247 to the DMI HARMONIE forecasts when assimilated with 3D-Var. However, it is expected that more
248 impact will be seen, if methods that take the high temporal resolution of these data into account
249 are used.

250 **Acknowledgements**

251 Partial funding from Innovation Fund Denmark helped to fund this study, which is much ap-
252 preciated.

253 **References**

254 Bell, S., D. Cornford, and L. Bastin (2015), How good are citizen weather stations? Addressing a
255 biased opinion, *Weather*, 70(3), 75–84, doi:10.1002/wea.2316.

256 Clark, M. R., J. D. C. Webb, and P. J. Kirk (2018), Fine-scale analysis of a severe hailstorm using
257 crowd-sourced and conventional observations: Fine-scale analysis of a hailstorm, *Meteorological*
258 *Applications*, 25(3), 472–492, doi:10.1002/met.1715.

259 Driesenaar, T. (2009), General description of the HARMONIE-AROME model, [http://hirlam.](http://hirlam.org/index.php/hirlam-programme-53/general-model-description/mesoscale-harmonie)
260 [org/index.php/hirlam-programme-53/general-model-description/mesoscale-harmonie,](http://hirlam.org/index.php/hirlam-programme-53/general-model-description/mesoscale-harmonie)
261 [Online; Accessed 21 June 2018].

262 Fieldsense A/S (2019), Fieldsense Weather Stations, <https://site.fieldsenseapp.com/en/>, [On-
263 line; Accessed 24 May 2019].

264 Hintz, K. S., H. Vedel, and E. Kaas (2019a), Collecting and Processing of Barometric Data from
265 Smartphones for Potential Use in NWP Data Assimilation, *Meteorological Applications*, doi:10.
266 1002/met.1805.

267 Hintz, K. S., K. O’Boyle, S. L. Dance, S. A. Ali, I. Ansper, D. Blaauboer, M. Clark, A. Cress,
268 M. Dahoui, R. Darcy, J. Hyrkkänen, L. Isaksen, E. Kaas, M. Lavanant, G. Lebloa, E. Mallet,
269 C. McNicholas, J. Onvlee-Hooimeijer, B. Sass, V. Siirand, H. Vedel, J. A. Waller, and X. Yang
270 (2019b), Collecting and utilising crowdsourced data for numerical weather prediction: Proposi-
271 tions from the meeting held in Copenhagen, 4-5 December 2018, *Atmospheric Science Letters*,
272 doi:10.1002/ASL2.921.

273 Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen
274 (2005), Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real
275 Observations, *Monthly Weather Review*, 133(3), 604–620, doi:10.1175/MWR-2864.1.

276 Lonobox (2019), Lonobox, <http://lonobox.com/>, [Online; Accessed 24 May 2019].

277 Lorenc, A. C. (2003), The potential of the ensemble Kalman filter for NWP—a comparison with
278 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, 129(595), 3183–3203, doi:10.1256/
279 qj.02.132.

280 Madaus, L. E., and C. F. Mass (2017), Evaluating Smartphone Pressure Observations for Mesoscale
281 Analyses and Forecasts, *Weather and Forecasting*, 32(2), 511–531, doi:10.1175/WAF-D-16-0135.
282 1.

283 McNicholas, C., and C. Mass (2018), Smartphone Pressure Collection and Bias Correction Using
284 Machine Learning, *J. Atmos. Oceanic Technol.*, *35*, 523–540, doi:10.1175/JTECH-D-17-0096.1.

285 Meier, F., D. Fenner, T. Grassmann, M. Otto, and D. Scherer (2017), Crowdsourcing air tem-
286 perature from citizen weather stations for urban climate research, *Urban Climate*, *19*, 170–191,
287 doi:10.1016/j.uclim.2017.01.006.

288 Meulen, J. v. d. (1992), The WMO Automatic Digital Barometer Intercomparison, *Tech. Rep. 49*,
289 World Meteorological Organization, de Bilt Netherlands.

290 Muller, C., L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R. Leigh
291 (2015), Crowdsourcing for climate and atmospheric sciences: current status and future potential,
292 *International Journal of Climatology*, *35*(11), 3185–3203, doi:10.1002/joc.4210.

293 Netatmo SAS (2018), Netatmo Blog, [https://www.netatmo.com/blog/en/weather/
294 understanding](https://www.netatmo.com/blog/en/weather/understanding), [Online; Accessed 10 September 2018].

295 Netatmo SAS (2019), Netatmo Support, [https://helpcenter.netatmo.com/en-us/
296 smart-home-weather-station-and-accessories/top](https://helpcenter.netatmo.com/en-us/smart-home-weather-station-and-accessories/top), [Online; Accessed 28 May 2016].

297 Patric, B. (2014), HP scaling out for the Internet of Things, [https://community.hpe.com/t5/
298 Grounded-in-the-Cloud/HP-scaling-out-for-the-Internet-of-Things/ba-p/6636688#
299 .WZWmA3cjGjR](https://community.hpe.com/t5/Grounded-in-the-Cloud/HP-scaling-out-for-the-Internet-of-Things/ba-p/6636688#.WZWmA3cjGjR), [Online; Accessed 17-August-2017].

300 Robustelli, U., V. Baiocchi, and G. Pugliano (2019), Assessment of Dual Frequency GNSS Observa-
301 tions from a Xiaomi Mi 8 Android Smartphone and Positioning Performance Analysis, *Electronics*,
302 *8*(1), 91, doi:10.3390/electronics8010091.

303 Vasiljevic, D., E. Andersson, L. Isaksen, and A. Garcia-Mendez (2006), Surface pressure bias cor-
304 rection in data assimilation, *ECMWF*, doi:10.21957/uv295rfmx5.

305 WMO (2010), Commission for instruments and methods of observation, *Tech. Rep. CIMO/ET-
306 Stand-1/Doc.10*, World Meteorological Organization.

307 Yang, X., B. Stig, M. Dahldom, B. Sass, S. Zhuang, B. Amstrup, C. Petersen, K. Nielsen, N. W.

308 Nielsen, and A. Mahura (2017), NEA, the Operational Implementation of HARMONIE 40h1.1
309 at DMI, pp. 104–111.

310 **Figures**

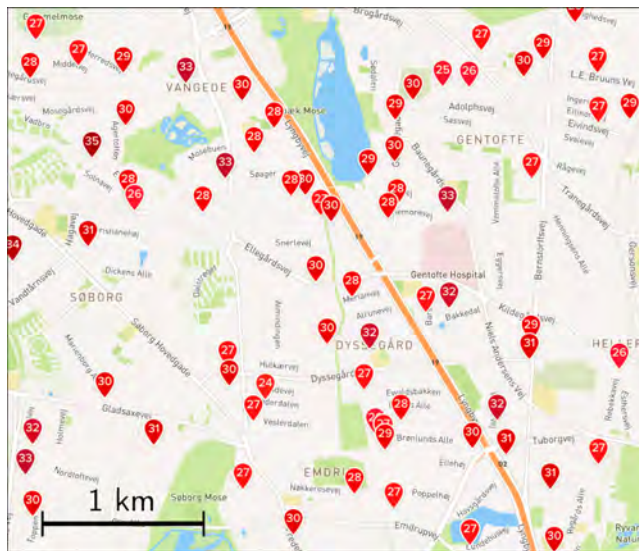


Figure 1: Map showing an example of the high spatial resolution of PWS temperature observations from Netatmo from a region north of Copenhagen, Denmark. The scale is shown in the lower left corner.
Credits: Netatmo, weathermap.netatmo.com

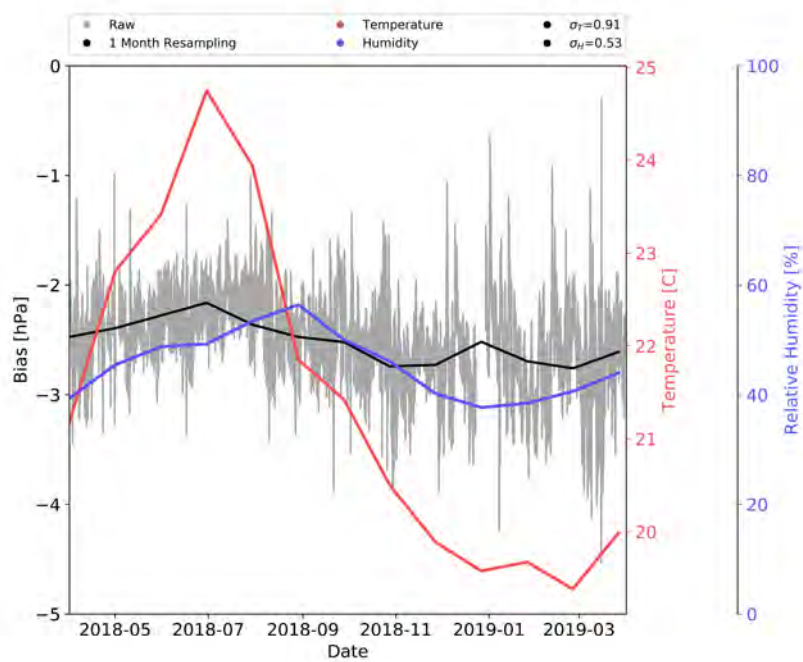


Figure 2: Station name: Snuderup. Bias of sea level pressure from an individual Netatmo station compared to the nearest DMI SYNOP station during one year. Grey: Raw data, as the difference between the PWS and the SYNOP station, which is resampled to 1 month to find the bias for each month (black line). Red: Observed indoor temperature by the PWS. Blue: Observed relative humidity. σ_T is the Pearson correlation coefficient between the resampled pressure bias and temperature. σ_H is the correlation between the resampled pressure bias and relative humidity.

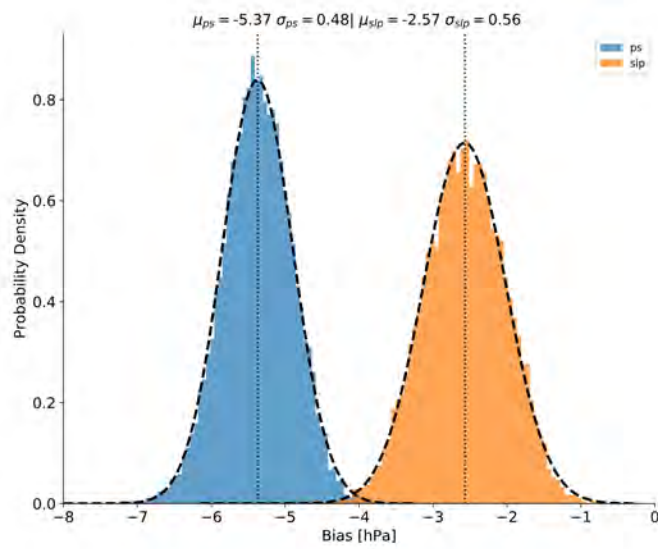


Figure 3: Station Name: Snuderup. Distribution of difference in pressure between a single Netatmo station and a DMI SYNOP station during a period of one year. Blue (left) shows the distribution of the surface pressure and the orange (right) shows the distribution of the sea level pressure. Both distributions have been fitted with a Gaussian distribution using the mean and standard deviation from each distribution shown in the top.



Figure 4: Observation usage of surface pressure observation in HARMONIE from 4 April 2018 to 25 May 2018 for Denmark only. REF is not shown as there are no surface pressure observations used from Denmark.

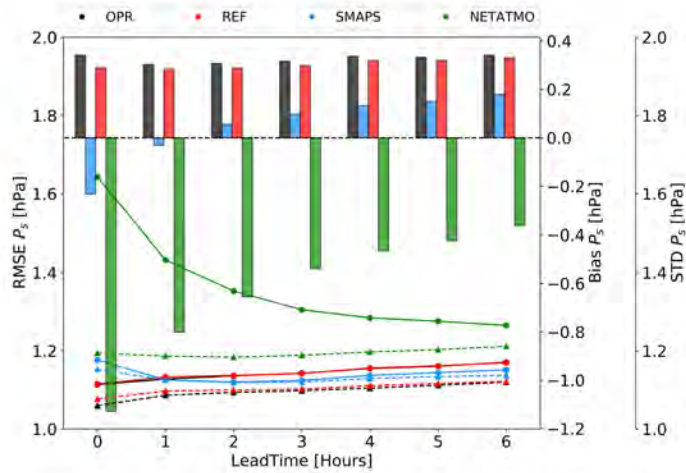


Figure 5: Root Mean Square Error (solid lines), Bias (bars) and Standard Deviation (broken lines) of surface pressure in average over two months using an assimilation cycle of three hours.

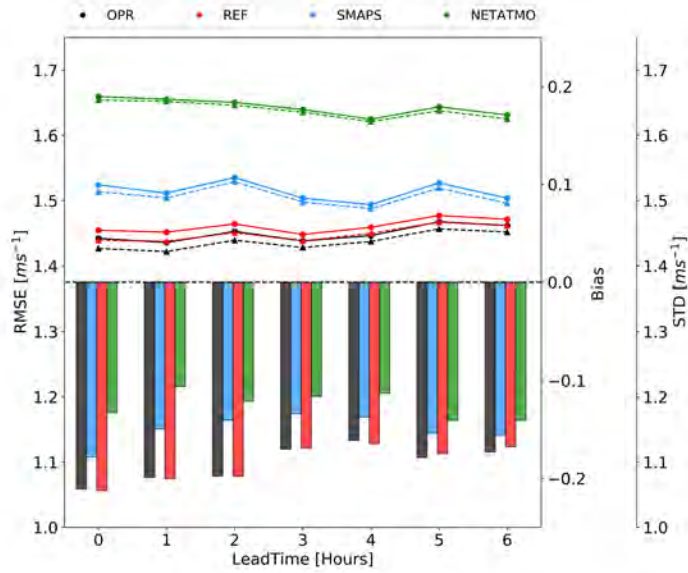


Figure 6: Root Mean Square Error (solid lines), Bias (bars) and Standard Deviation (broken lines) of 10 m wind in average over two months using an assimilation cycle of three hours.

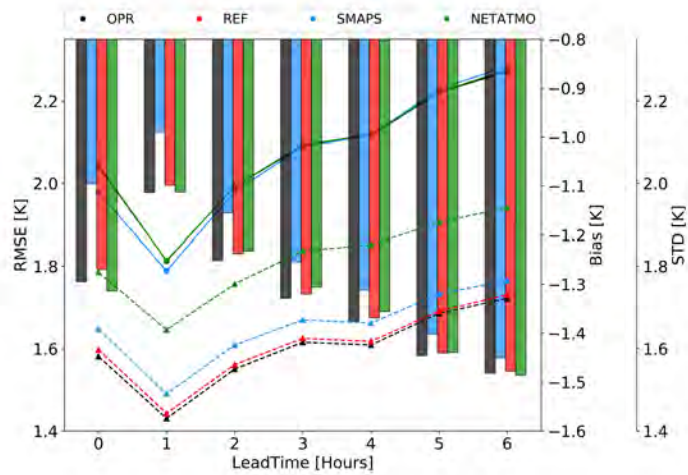


Figure 7: Root Mean Square Error (solid lines), Bias (bars) and Standard Deviation (broken lines) of 2 m temperature in average over two months using an assimilation cycle of three hours.

Table 1: Correlations coefficients between surface pressure and temperature (σ_T) and between surface pressure and relative humidity (σ_H). σ^{RAW} is the correlation coefficient for the raw data and σ^{RES} is the correlation coefficient for the resampled data. *Alt* specifies the altitude in meters relative to mean sea level from the Danish Terrain Model and the numbers in parentheses are the terrain elevation from the NWP model. Station names in parentheses refers to which stations that have been compared.

Name	Lat	Lon	Alt	σ_T^{RES}	σ_H^{RES}	σ_T^{RAW}	σ_H^{RAW}
Snuderup	55.690	11.370	62 (37)	0.91	0.53	0.44	0.20
Kgs. Lyngby	55.768	12.512	30 (32)	-0.50	0.17	-0.13	-0.02
Skelkaeret	56.006	12.045	4 (13)	-0.30	-0.70	0.07	-0.27
Holbaek	55.736	11.604	13 (9)	(Snuderup)			
Jaegersborg	55.766	12.527	42 (28)	(Kgs. Lyngby)			
Gilleleje	56.120	12.343	43 (8)	(Skelkaeret)			

Appendices

Chapter A

Additional figures



Figure A1: Smartphone wind observations taken near Brøndby stadium in Copenhagen, Denmark. While the difference in wind speeds (red circles) at the stadium can be interesting for athletes, the observations represent a very high representativeness error, which makes utilizing the observations in numerical weather prediction difficult.

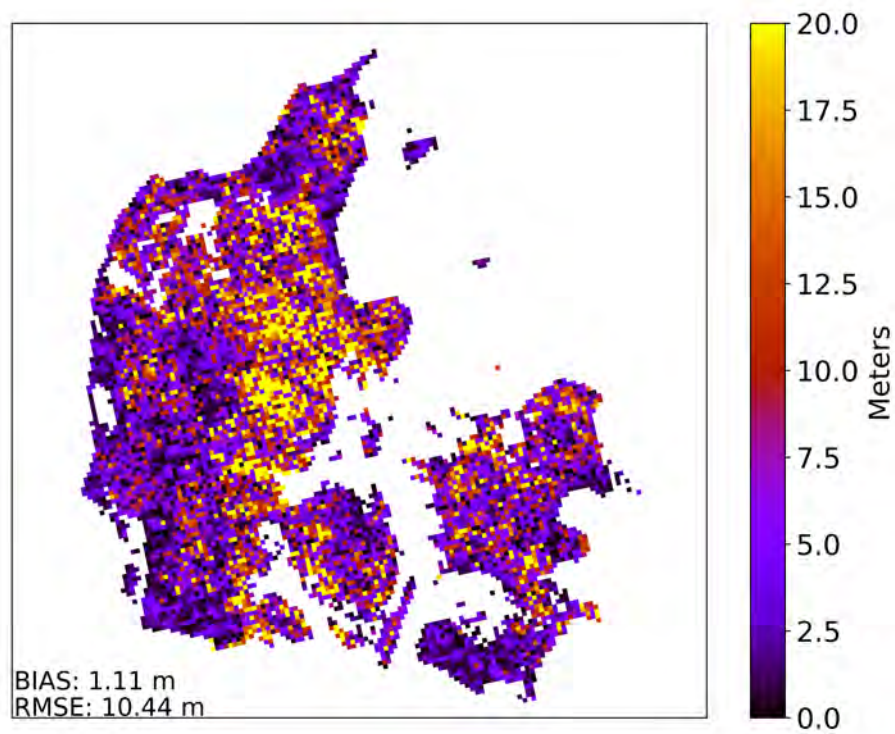


Figure A2: Absolute terrain height difference between the Danish Terrain Model and the Terrain of the DMI HARMONIE model. The resolution of the terrain model is upscaled to that of the DMI HARMONIE model. Units are in meters. The DMI HARMONIE model has an overall bias of 1.11 m compared to the terrain model.

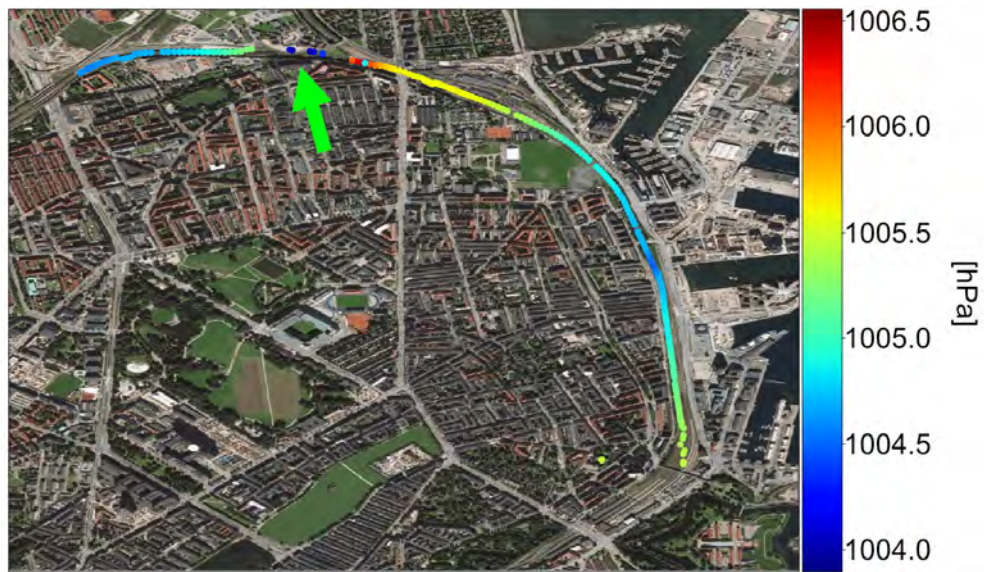


Figure A3: Example of a data stream of pressure measured by a smartphone on board on a train. The green arrow shows the position of a tunnel where it is seen that the position of the smartphone is not well defined. In general, changes in altitude is clearly seen. The colours are given in hPa in the colorbar.

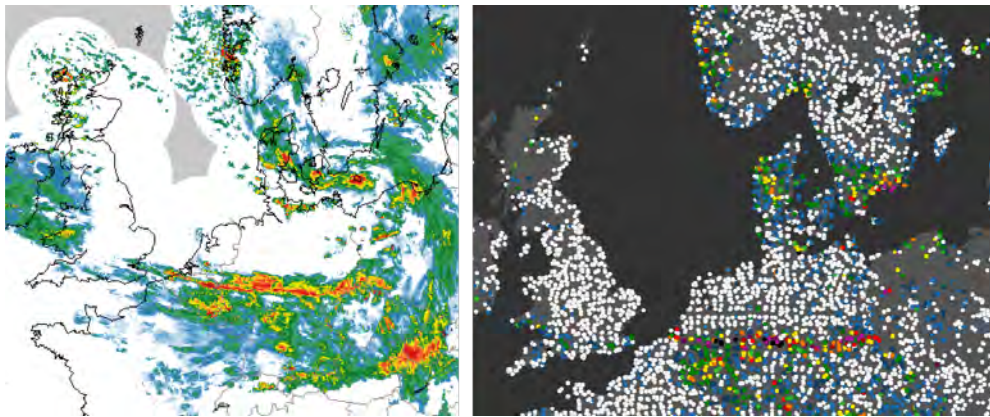


Figure A4: Precipitation 15th of March 2019 at 19 UTC. Left: Radar Reflectivity over Europe from the OPERA project. Right: Netatmo stations with a pluviometer installed. Colours represent rain over one hour. White stations (white circles) have not measured rain.

Credits: Dr Thomas Bøvith, DMI.

Chapter B

Derivation of the Monin-Obukhov length

Assuming that in a horizontally homogeneous surface layer the mean flow and turbulence characteristics depend only on four independent variables, namely, the height above the surface, z , the friction velocity, u_* , the kinematic surface heat flux, Q_0 and the buoyancy, g/θ_0 , as done by Monin and Obukhov (1954), one can derive the Monin-Obukhov length.

From the Buckingham Pi theorem (see section 6.1.1) it follows that there exist $m - n = 3$ dimensionless groups, because there exist $m = 4$ quantities ($z, u_*, Q_0, g/\theta_0$) with $n = 3$ fundamental dimensions (m, s, K). Recall the units of $[z] = m$, $[u_*] = m/s$, $[Q_0] = \text{Km}/s$ and $[g/\theta_0] = m/Ks^2$. One can form a dimensionless group by using the m quantities as

$$\left[\frac{z}{\frac{u_*^3 \theta_0}{Q_0 g}} \right] = \frac{m^3 s^{-1}}{m^3 s^{-1}} = 1. \quad (\text{B.1})$$

The dimensionless group can be written as z/L where L is the Monin-Obukhov length defined as

$$L = -\frac{u_*^3 \theta_0}{kgQ_0}, \quad (\text{B.2})$$

where k is the von Kármán constant (dimensionless) introduced for the sake of convenience, and the negative sign is introduced to have the same sign as the Richardson number.

Chapter C

Bias, standard deviation and mean squared error

To avoid any misunderstanding this appendix gives a brief definition and derivation of the widely used statistical terms, bias, standard deviation and mean squared error as used in this thesis.

Consider a Gaussian distribution $\mathcal{G}(x, \sigma)$, where from a value x_i has been drawn randomly N times. The mean is defined as

$$\bar{x} = \frac{\sum x_i}{N}. \quad (\text{C.1})$$

The standard deviation is denoted by σ and is given by

$$\begin{aligned} \sigma &= \left(\frac{\sum (x_i - \bar{x})^2}{N - 1} \right)^{1/2} \\ &= \left(\frac{\sum x_i^2 - N\bar{x}^2}{N - 1} \right)^{1/2}, \end{aligned} \quad (\text{C.2})$$

where the following entity has been used

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= \sum x_i^2 - 2 \sum x_i \bar{x} + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + N\bar{x}^2 \\ &= \sum x_i^2 - 2N\bar{x}^2 + N\bar{x}^2 \\ &= \sum x_i^2 - N\bar{x}^2. \end{aligned} \quad (\text{C.3})$$

The Root Mean Square (RMS) is simply defined as

$$\text{RMS} = \left(\frac{\sum x_i^2}{N} \right)^{1/2}. \quad (\text{C.4})$$

Often (this work is no excuse) one compares an observation to some reference or background field treated as the truth. Consider a residual as an observation minus a background given as $(O - B) = \text{OmB}$. Then the bias, root mean square error and standard deviation is given by

$$\text{BIAS} = \frac{\sum \text{OmB}_i}{N}, \quad (\text{C.5})$$

$$\text{RMSE} = \left(\frac{\sum \text{OmB}_i^2}{N} \right)^{1/2} \quad (\text{C.6})$$

and

$$\text{STD} = \left(\frac{\sum \text{OmB}_i^2 - N \overline{\text{OmB}}^2}{N - 1} \right)^{1/2} \quad (\text{C.7})$$

respectively. Considering equation (C.5), (C.6) and (C.7) it is seen that bias, RMSE and standard deviation are connected as

$$\text{STD} = \left(\frac{N \cdot \text{RMSE}^2 - N \cdot \text{BIAS}^2}{N - 1} \right)^{1/2}. \quad (\text{C.8})$$

Chapter D

Namelist of the Frigg-WRF setup

Block D.1 shows the namelist file used for the Frigg-WRF setup described in chapter 4. Using this namelist makes it possible to reproduce the results. The settings are specified via the numbers on the right. Earth System Research Laboratory (2019) gives a description of the options.

Block D.1: The Fortran namelist file (namelist.input) for the Frigg-WRF setup.

```
1 &time_control
2 run_days           = 0,
3 run_hours         = 72,
4 run_minutes       = 0,
5 run_seconds       = 0,
6 start_year        = 2016, 2016,
7 start_month       = 09, 09,
8 start_day         = 30, 30,
9 start_hour        = 06, 06,
10 start_minute     = 00, 00,
11 start_second     = 00, 00,
12 end_year         = 2016, 2016,
13 end_month        = 10, 10,
14 end_day          = 03, 03,
15 end_hour         = 06, 06,
16 end_minute       = 00, 00,
17 end_second       = 00, 00,
18 interval_seconds = 10800,
19 input_from_file  = .true.,.true.,
```

```
20 history_interval      = 60, 60,
21 frames_per_outfile    = 1000, 1000,
22 restart               = .false.,
23 restart_interval      = 5000,
24 io_form_history        = 2
25 io_form_restart       = 2
26 io_form_input         = 2
27 io_form_boundary      = 2
28 debug_level           = 0
29 /
30
31 &domains
32 time_step              = 54,
33 time_step_fract_num    = 0,
34 time_step_fract_den    = 1,
35 max_dom                = 2,
36 e_we                  = 253, 385,
37 e_sn                  = 208, 385,
38 e_vert                = 30, 30,
39 p_top_requested        = 5000,
40 num_metgrid_levels    = 32,
41 num_metgrid_soil_levels = 4,
42 dx                    = 9000, 3000,
43 dy                    = 9000, 3000,
44 grid_id                = 1, 2,
45 parent_id              = 0, 1,
46 i_parent_start         = 1, 65,
47 j_parent_start         = 1, 45,
48 parent_grid_ratio      = 1, 3,
49 parent_time_step_ratio = 1, 3,
50 feedback               = 1,
51 smooth_option          = 0,
52 sfcfcp_to_sfcfcp      = .true.
53 /
54
55 &physics
56 mp_physics              = 3, 3,
```

```

57 ra_lw_physics      = 1,      1,
58 ra_sw_physics      = 1,      1,
59 radt               = 30,     30,
60 sf_sfclay_physics  = 1,      1,
61 sf_surface_physics = 2,      2,
62 bl_pbl_physics     = 5,      5,
63 bldt               = 0,      0,
64 cu_physics         = 1,      1,
65 cudt               = 5,      5,
66 isfflx             = 1,
67 ifsnow             = 1,
68 icloud             = 1,
69 surface_input_source = 3,
70 num_soil_layers    = 4,
71 num_land_cat       = 21,
72 sf_urban_physics   = 0,      0,
73 /
74
75 &dynamics
76 w_damping          = 0,
77 diff_opt           = 1,      1,
78 km_opt             = 4,      4,
79 diff_6th_opt       = 0,      0,
80 diff_6th_factor    = 0.12,   0.12,
81 base_temp          = 290.,
82 damp_opt           = 0,
83 zdamp              = 5000.,   5000.,
84 dampcoef           = 0.2,     0.2,
85 khdif              = 0,      0,
86 kvdif              = 0,      0,
87 non_hydrostatic    = .true., .true.,
88 moist_adv_opt      = 1,      1,
89 scalar_adv_opt     = 1,      1,
90 /
91
92 &bdy_control
93 spec_bdy_width     = 5,

```

```
94 spec_zone           = 1,  
95 relax_zone         = 4,  
96 specified           = .true., .false.,  
97 nested              = .false., .true.,  
98 /  
99  
100 &namelist_quilt  
101 nio_tasks_per_group = 0,  
102 nio_groups           = 1,  
103 /
```

Chapter E

Mathematical equivalence of OI and 3D-Var

It can be shown that the weight obtained in 3D-Var,

$$\mathbf{Q} = [\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{R}^{-1},$$

is mathematically equivalent to the weight obtained for OI,

$$\mathbf{W} = \frac{\mathbf{P}_b \mathbf{H}^T}{\mathbf{H} \mathbf{P}_b \mathbf{H}^T + \mathbf{R}}.$$

For an explanation of mathematical symbols and their behaviour, the reader is referred to chapter 9. To proof the mathematical equivalence take the inverse of the 3D-Var weight and use that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$

$$\begin{aligned} \mathbf{Q}^{-1} &= [(\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}]^{-1} = (\mathbf{H}^T \mathbf{R}^{-1})^{-1} (\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \\ &= \mathbf{R} \mathbf{H}^{-T} (\mathbf{P}_b^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \\ &= \mathbf{R} \mathbf{H}^{-T} \mathbf{P}_b^{-1} + \mathbf{R} \mathbf{H}^{-T} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \\ &= \mathbf{R} \mathbf{H}^{-T} \mathbf{P}_b^{-1} + \mathbf{H} \\ &= (\mathbf{R} \mathbf{H}^{-T} \mathbf{P}_b^{-1} + \mathbf{H}) \mathbf{P}_b \mathbf{H}^T \mathbf{H}^{-T} \mathbf{P}_b^{-1} \\ &= (\mathbf{R} + \mathbf{H} \mathbf{P}_b \mathbf{H}^T) \mathbf{H}^{-T} \mathbf{P}_b^{-1}. \end{aligned} \tag{E.1}$$

Taking the inverse of equation (E.1) yields the OI weight

$$\frac{1}{\mathbf{Q}^{-1}} = \mathbf{Q} = \frac{\mathbf{P}_b \mathbf{H}^T}{\mathbf{H} \mathbf{P}_b \mathbf{H}^T + \mathbf{R}} = \mathbf{W} \tag{E.2}$$

Thus 3D-Var and OI are mathematical equivalent.