Crowds and clouds - the potential of using data from smartphones in numerical weather prediction

Kasper Stener Hintz
kah@dmi.dk

Danish Meteorological Institute &
The Niels Bohr Institute, UCPH

Outline

- **Type of data:** Wind and Pressure measurements from smartphones
- Size of the crowd / Number of observations
- Strategy for data validation
- Strategi for handling crowdsourced data

Wind measurements from smartphones

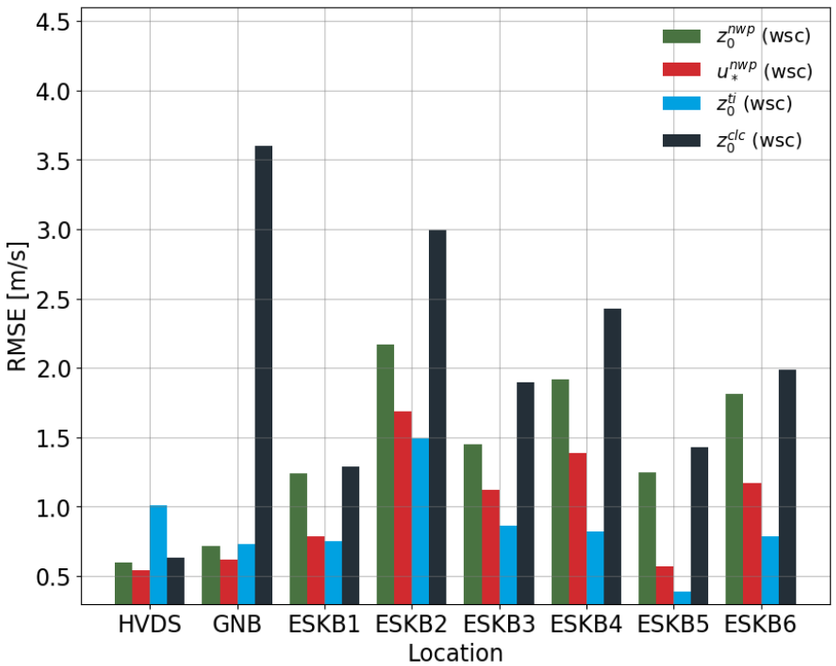
Relative accuracy: 0.25 m/s

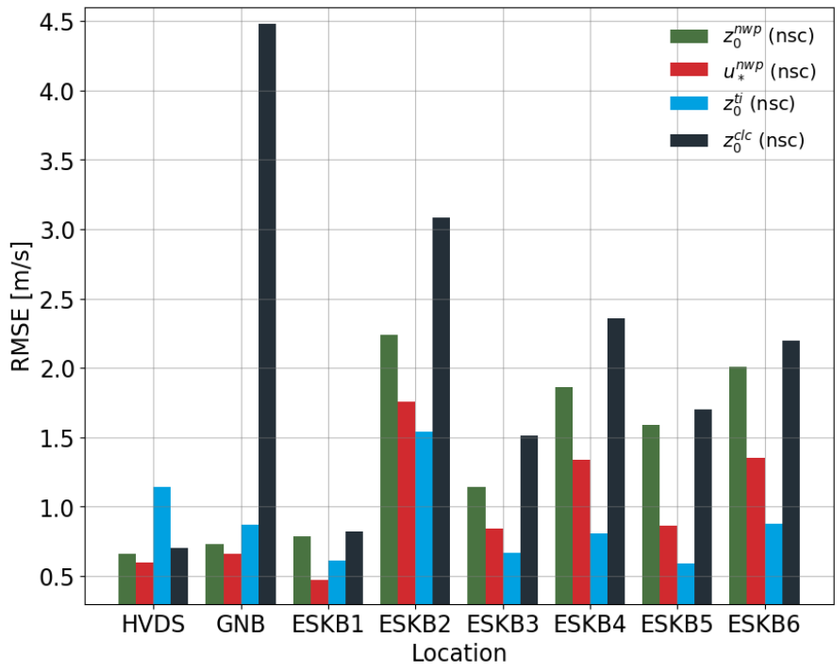
Data refresh rate: 4 Hz

Wind range: 2-40 m/s

Vaavud Sleipnir
(vaavud.com)

It is problematic to use the wind measurement itself due to turbulence near the surface. It was possible to derive the roughness length from the wind measurement. Important for surface wind forecasts and surface fluxes.



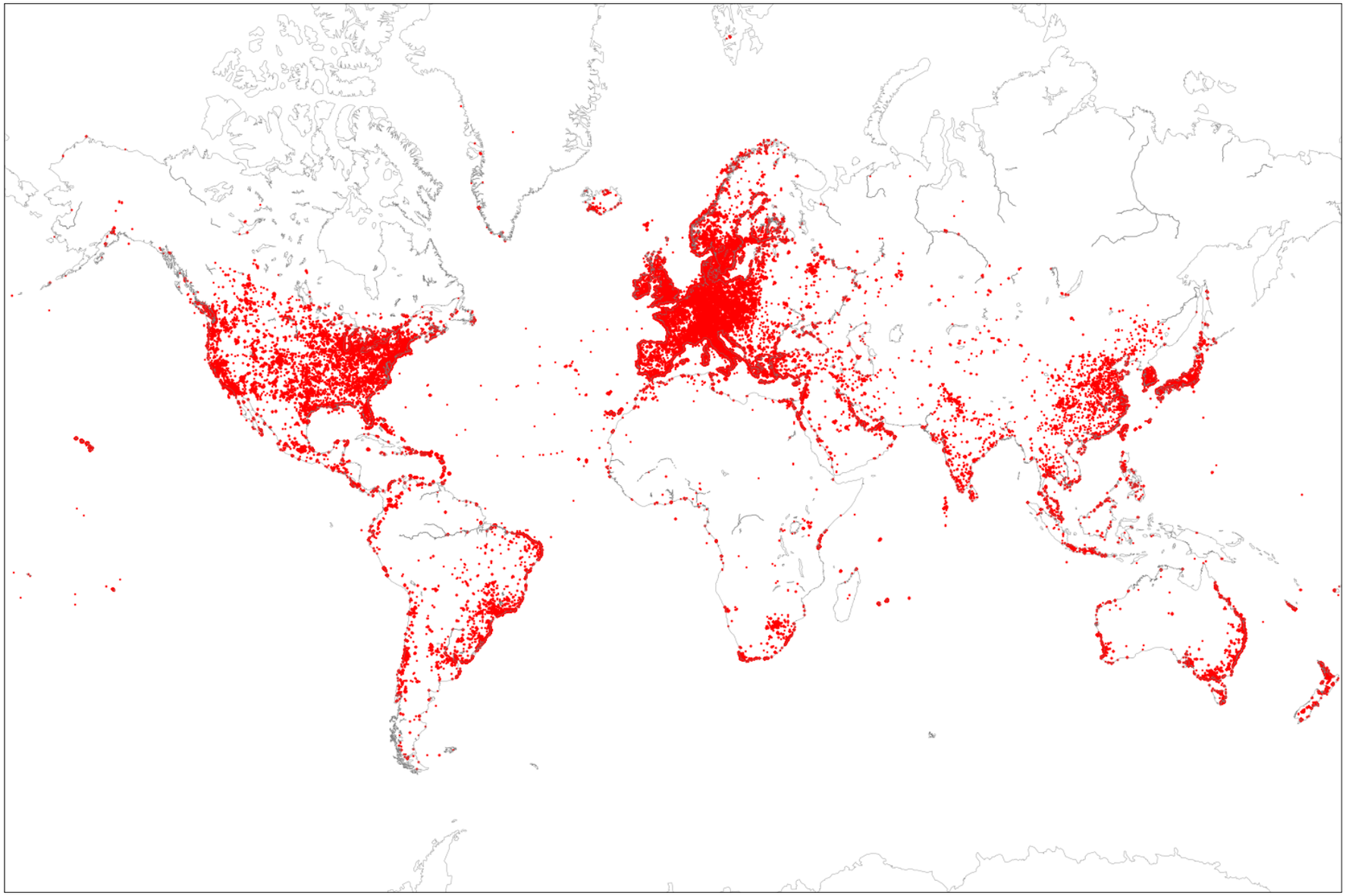


RMSE for extrapolation of wind measurements using sources of roughness length of different origin.

z_0^{ti} Is derived from the smartphone wind measurement and has in general the lowest error.

Wind measurements from smartphones

The map shows the location of measurements taken in the period 2013-2014 from Vaavud users.



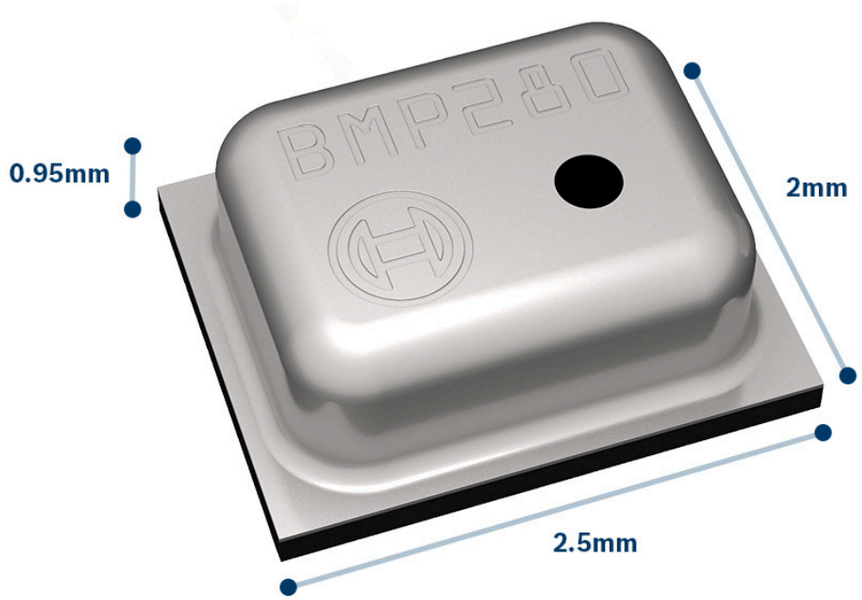
Wind measurements are stored for later use. Because of the roughness length changes slowly, the derived roughness length can be used to improve the table values of roughness length in an NWP model and thereby improve the forecast of surface wind.

Pressure from smartphones

Absolute accuracy: 1 hPa

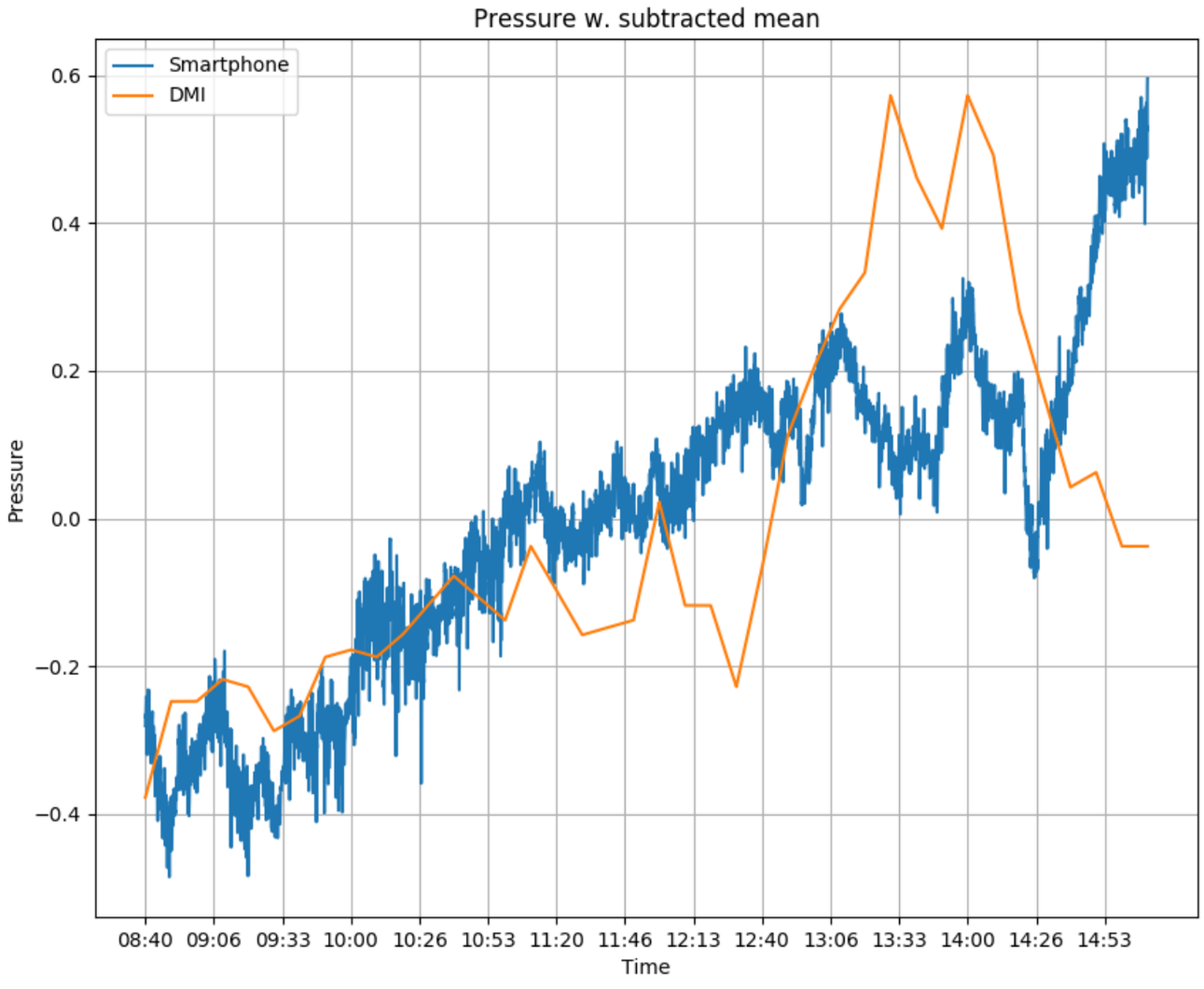
Relative accuracy: 0.12 hPa

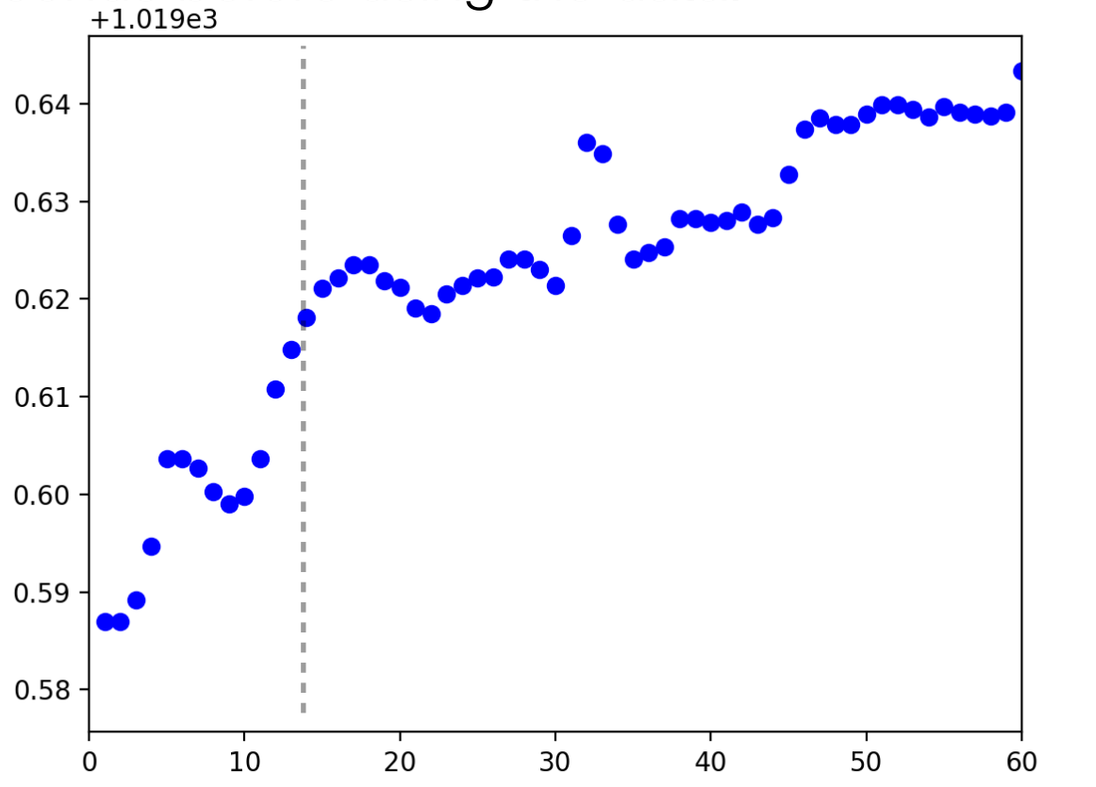
Data refresh rate: 1 Hz



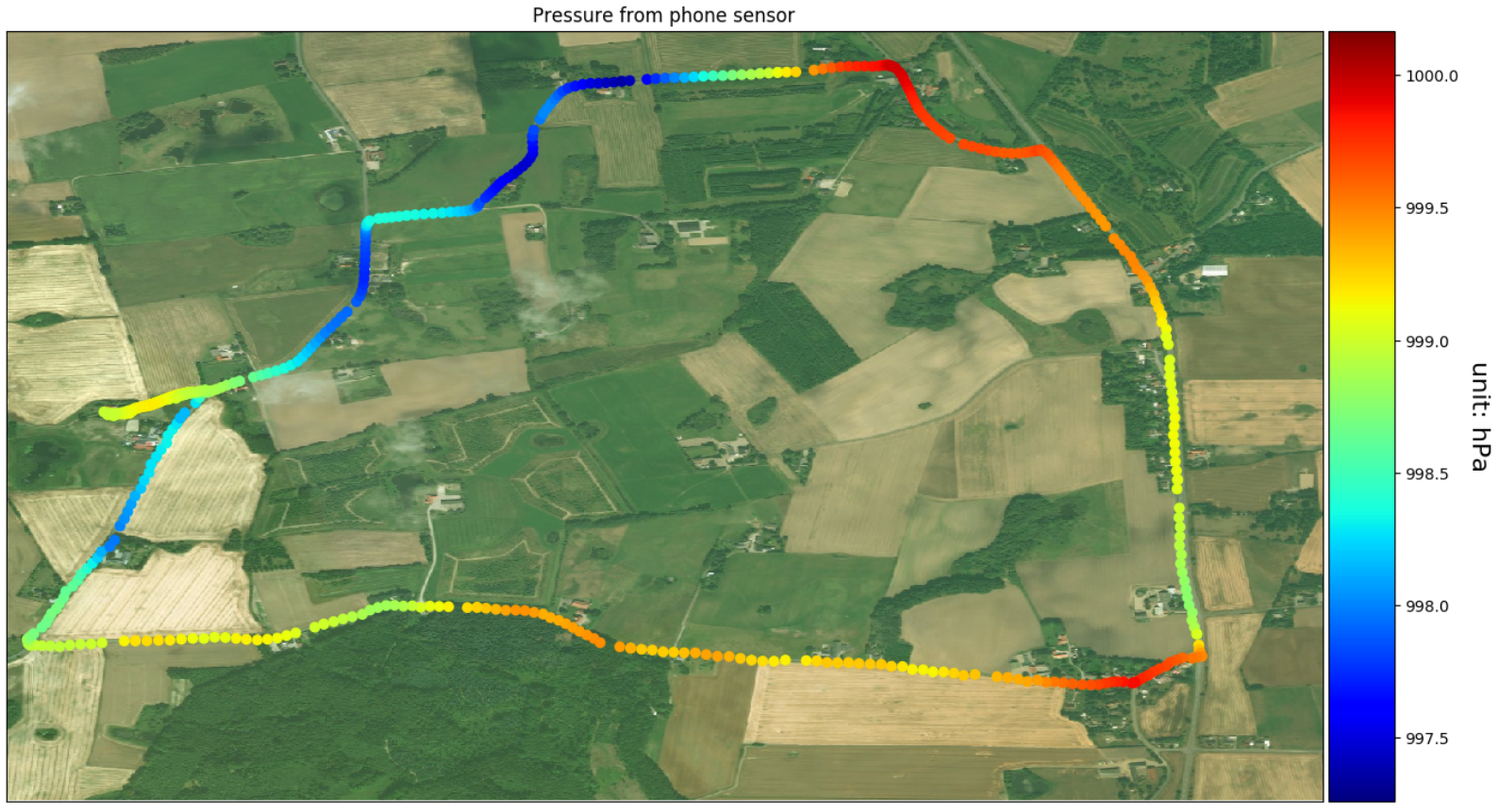
Bosch BMP280 Pressure sensor
(https://www.bosch-sensortec.com/bst/products/all_products/bmp280)

The sensor is switched off by default, so it needs to run for a few seconds before using the data.





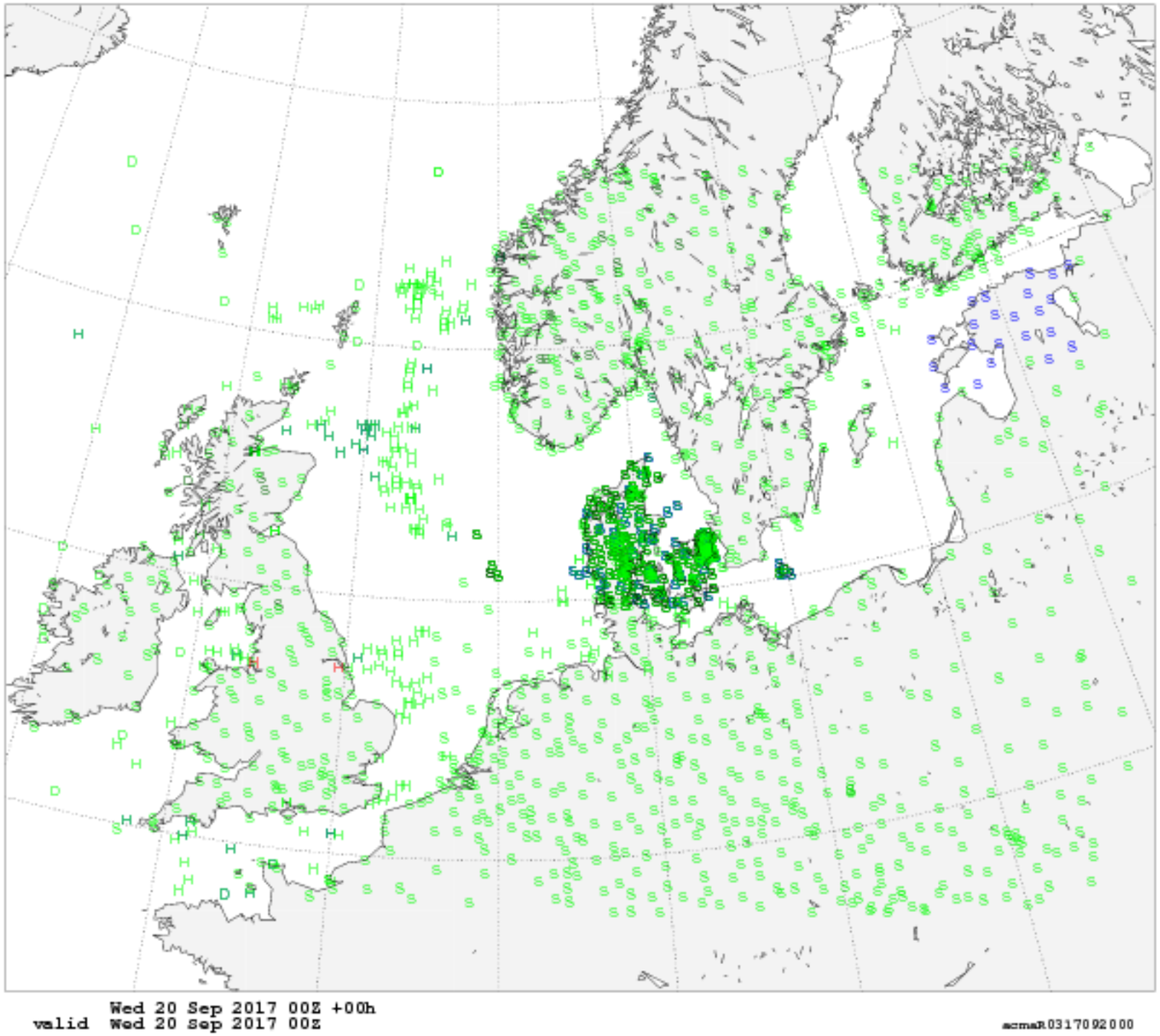
Raw pressure data output from an iPhone 7 smartphone



Logging data while biking

Logging onboard of a train

Used SYNOPs in DMI R03 (HIRLAM)



total	synop	:	3036
active	synop	:	1266
rejected	synop	:	0
redundant	synop	:	323
nodata	synop	:	1445
notused	synop	:	0
total	ship	:	252
active	ship	:	218
rejected	ship	:	2
redundant	ship	:	32
nodata	ship	:	0
notused	ship	:	0
total	dribu	:	20
active	dribu	:	17
rejected	dribu	:	0
redundant	dribu	:	2
nodata	dribu	:	1
notused	dribu	:	0

Total synop: 3026

Active synop: 1266

Size of the crowd

A rough estimate from the most popular weather apps in Denmark.

≈ 120.000 daily users*

Approx. 4000% increase in Denmark.

If this number is scalable based on population density, the European Union will in total give approximately **10 million daily users**. Letting the software run in the background once every hour will give us 240 million extra pressure observations every day.

With a data refresh rate of 1 Hz, and assuming a measurement length of 30-60 seconds gives 7.2 - 14.4 billion measurement points per day.

*2% conversion rate in Denmark

Strategy for data validation (of pressure)

Far from all observations are of high-quality. Bad observations includes:

- Altitudes away from the surface (GPS altitude is not very accurate)
- Measuring while moving (on a slope)
- Measuring in non-weather related turbulent conditions (Driving through a tunnel)

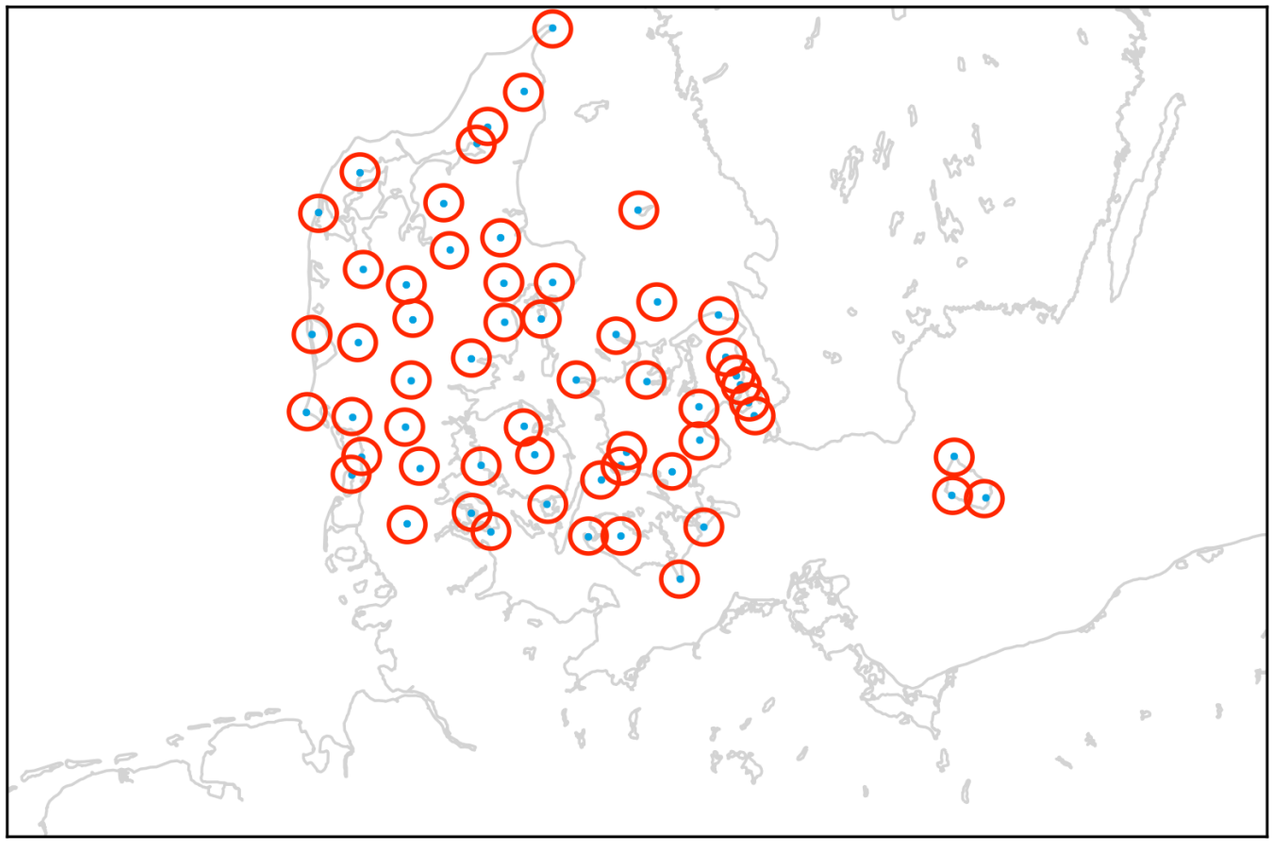
There is no need to send observations that:

- Have a large standard deviation
- Fail a range check and a buddy check
- Deviate too much from a background

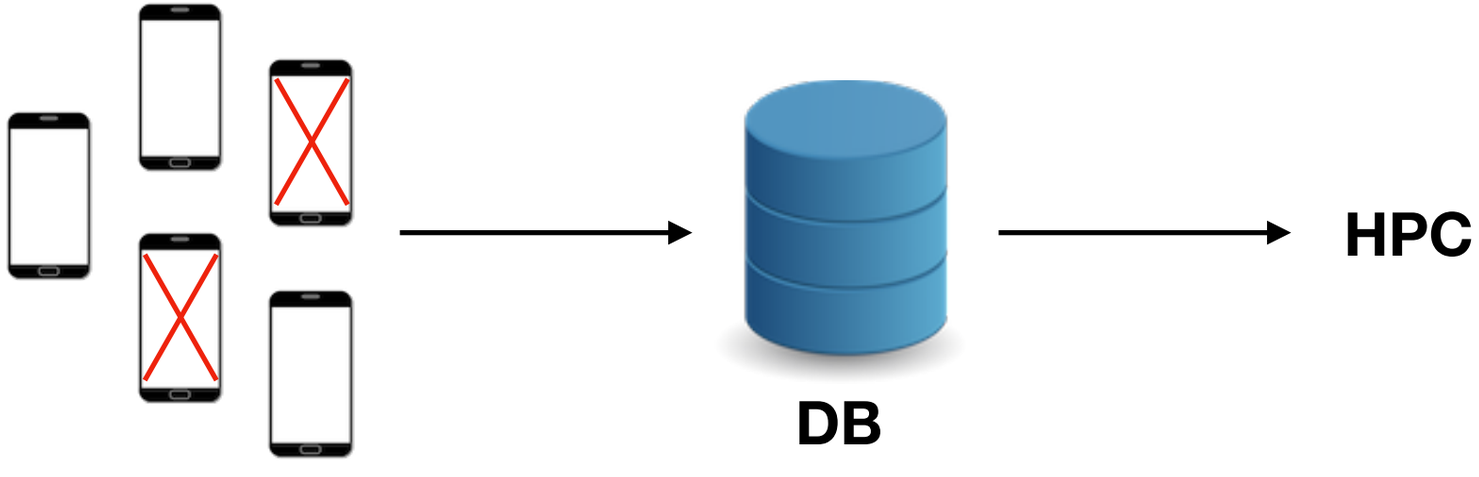
- Currently, work is done on how to predict and correct the bias of each device. Using regression with decision trees, with input from all the phone sensors are one possibility that is being explored.

Strategy for data validation (of pressure)

- Bias can also be calculated and updated for each device when a user is near a SYNOP station (blue dots), which can be tracked with geofence (red circles).

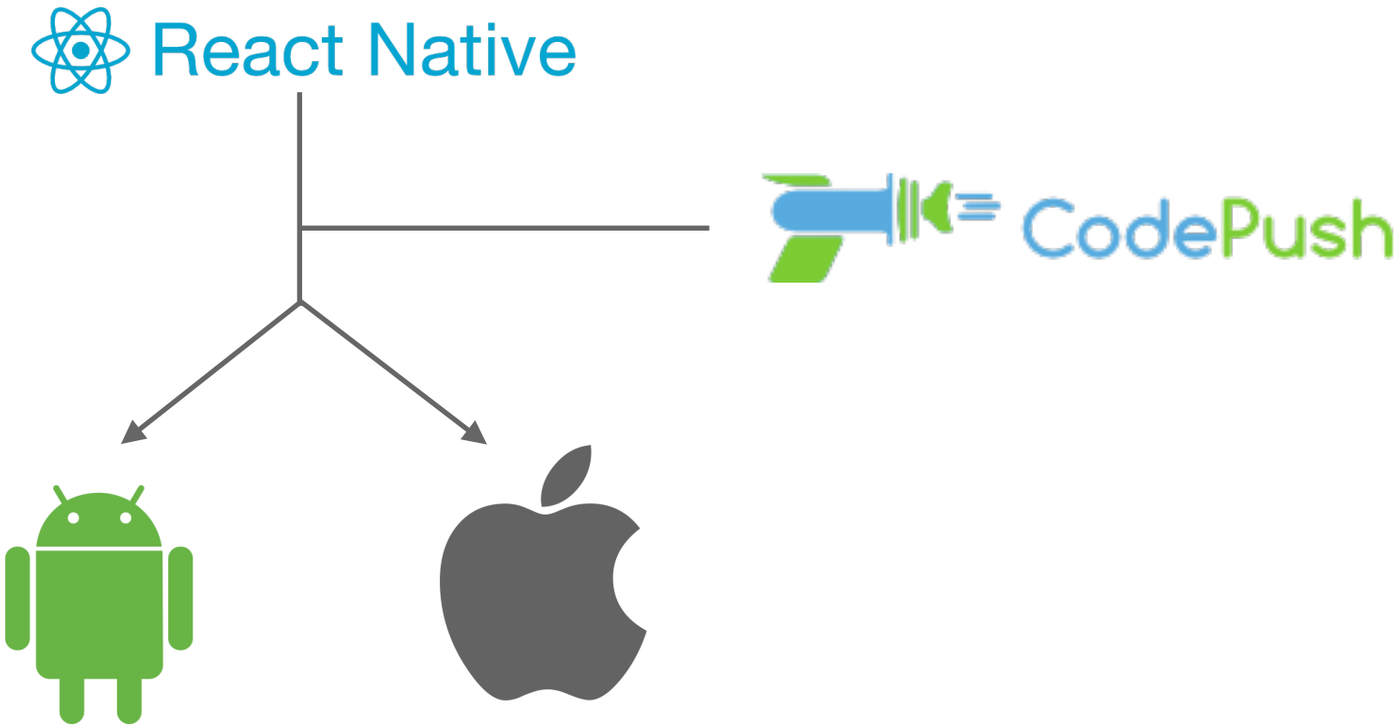


- Instead of sending all observations to an HPC system for validation, **we aim to process as much data as possible "online" on the individual devices**, to take advantage of the computational resources the smartphones have.



Future strategi for handling crowdsourced data

Ideally we will provide an SDK to remove the ‘black box’ of smartphone observations.



React Native is a framework for building Android and iOS apps using only JavaScript

We can focus on maintaining one JavaScript codebase, instead of one for Android (Java) and one for iOS (objective-C/Swift)

We can make sure essential parts of the data processing is up-to-date on all devices using CodePush which lets us skip cumbersome review processes and makes sure all users have the updated version.

Summary

- Measurements from smartphones
  - Roughness-length from wind measurements
  - Pressure measurements

- Size of the crowd
  - ~10 million daily users for Europe

- Strategy for data validation
  - Geofencing
  - Traditional measures
  - Predict bias using machine learning?

- Strategi for handling crowdsourced data
  - SDK
  - Framework for multiple platforms

Kasper Stener Hintz
kah@dmi.dk