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

Hintz, Kasper S.; Vedel, Henrik; Kaas, Eigil

Published in:
Meteorological Applications

DOI:
10.1002/met.1805

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
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...
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 Mbii, 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 QCMOB, 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. QCMOB 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:

\[
\bar{x}_i = \frac{i-1}{i} x_{i-1} + \frac{1}{i} x_i \quad (1)
\]
and

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

where $\bar{x}_i$ is the average over $i = N$ measurements and $x_i$ is the $i$th measurement. $\sigma_i$ 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 $1.6 \times 10^6$ 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):

$$p_{\text{msl}} = (p - 0.3) \left(1 + k_1 \frac{h}{(p - 0.3) k_2}\right)^{1/k_2}$$ (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 $p_{\text{msl}}$ is the MSLP in hectopascals. Then a climatological check is performed, whether $p_{\text{msl}}$ is within the range 850–1050 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 $\pm 1$ 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 $0.2^\circ$ and a threshold of 1.0 hPa were used. In EXP_MEDV2 a distance radius of $0.5^\circ$ 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.
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,
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

**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

<table>
<thead>
<tr>
<th>Experiment name</th>
<th>OPR</th>
<th>REF</th>
<th>EXP</th>
<th>EXP_MEDV1</th>
<th>EXP_MEDV2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYNOP DK</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SPO</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>QCMOB</td>
<td>-</td>
<td>-</td>
<td>Simple</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>QCMOB type</td>
<td>-</td>
<td>-</td>
<td>No background, median, SYNOP, orography check</td>
<td>OBSA loose median check</td>
<td>OBSA strict median check</td>
</tr>
</tbody>
</table>

See text for details about QCMOB and QCMOB type.
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 SYNOPs were done. In addition, two runs with no ASPOs were done, one without the Danish SYNOPs (REF) and one with the Danish SYNOPs (OPR). In all runs SYNOP 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 SYNOPs.

The root mean square error (RMSE) and the bias of surface pressure were computed using the DMI SYNOP 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:

\[
\text{FSS} = 1 - \frac{\text{MSE}}{\text{MSE}_{\text{REF}}} \tag{4}
\]

Here MSE is the mean squared error of a model field given as:

\[
\text{MSE} = \frac{1}{N_xN_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (O_{i,j} - M_{i,j})^2 \tag{5}
\]

where \(O_{i,j}\) and \(M_{i,j}\) are a binary observation field and the model field respectively and \(N\) represents the spatial scale. \(\text{MSE}_{\text{REF}}\) is defined as the largest possible MSE that can be obtained from the model and observation fields:

\[
\text{MSE}_{\text{REF}} = \frac{1}{N_xN_y} \left( \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M_{i,j}^2 \right) \tag{6}
\]

FSS = 1 is a perfect score and 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 ordinates, lines) and bias (right ordinates, 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>
<thead>
<tr>
<th>Experiment</th>
<th>REF</th>
<th>EXP</th>
<th>EXP_MEDV1</th>
<th>EXP_MEDV2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASPO input to DA</td>
<td>0</td>
<td>8,385</td>
<td>1,566</td>
<td>872</td>
</tr>
<tr>
<td>ASPO rejected by DA</td>
<td>0</td>
<td>1,040 (12.4%)</td>
<td>49 (3.1%)</td>
<td>20 (2.3%)</td>
</tr>
<tr>
<td>Total ASPO</td>
<td>28,757</td>
<td>27,177</td>
<td>28,708</td>
<td>28,737</td>
</tr>
<tr>
<td>Total ASPO + rejected</td>
<td>28,757</td>
<td>28,757</td>
<td>28,757</td>
<td>28,757</td>
</tr>
</tbody>
</table>

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.
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 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).
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
Reading: European Centre for Medium-range Weather Forecasts. https://doi.org/10.21957/uv295rfmx5.


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;26:733–746. https://doi.org/10.1002/met.1805